Conformational B-cell epitope Prediction

Paolo Marcatili
Prediction of linear epitopes

**Pro**
- easily predicted computationally
- easily identified experimentally
- immunodominant epitopes in many cases
- do not need 3D structural information
- easy to produce and check binding activity experimentally

**Con**
- only ~10% of epitopes can be classified as “linear”
- weakly immunogenic in most cases
- most epitope peptides do not provide antigen-neutralizing immunity
- in many cases represent hypervariable regions
Sequence based prediction methods

- Linear methods for prediction of B cell epitopes have low performances.
- The problem is analogous to the problems of representing the surface of the earth on a two-dimensional map.
- Reduction of the dimensions leads to distortions of scales, directions, distances.
- The world of B-cell epitopes is 3 dimensional and therefore more sophisticated methods must be developed.
So what is more sophisticated?

- Use of the three dimensional structure of the pathogen protein
- Analyze the structure to find surface exposed regions
- Additional use of information about conformational changes, glycosylation and trans-membrane helices
Sources of three-dimensional structures

- Experimental determination
  - X-ray crystallography
  - NMR spectroscopy

- Both methods are time consuming and not easily done in a larger scale

- Structure prediction
  - Homology modeling
  - Fold recognition

- Less time consuming, but there is a possibility of incorrect predictions, specially in loop regions
Protein structure prediction

- **Homology/comparative modeling**
  - >25% sequence identity (seq 2 seq alignment)

- **Fold-recognition**
  - <25% sequence identity (Psi-blast search/ PSSM 2 seq alignment)

- **Ab initio structure prediction**
  - 0% sequence identity

Diagram showing a protein structure model transformation.
What does antibodies recognize in a protein?
# The binding interaction

<table>
<thead>
<tr>
<th>Epitope characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size</strong></td>
<td><strong>Van Regenmortel (2009)</strong></td>
</tr>
<tr>
<td>10-25 residues is involved in binding</td>
<td>Present study</td>
</tr>
<tr>
<td>15±4 residues is involved in binding</td>
<td>Sun et al., (2011)</td>
</tr>
<tr>
<td>22±8 residues is involved in binding</td>
<td>Rubinstein et al. (2008)</td>
</tr>
<tr>
<td>600-1000 Å² is buried upon binding</td>
<td>Sun et al., (2011)</td>
</tr>
<tr>
<td>847±279 Å² accessible surface area</td>
<td>Present study</td>
</tr>
<tr>
<td>The epitope plane (see results): 401±133 Å² when approximated by an ellipse</td>
<td>Present study</td>
</tr>
<tr>
<td>Thickness (see results): 8.2±2.0 Å</td>
<td>Present study</td>
</tr>
<tr>
<td><strong>Shape</strong></td>
<td><strong>Rubinstein et al. (2008)</strong></td>
</tr>
<tr>
<td>Flat rugged area</td>
<td>Present study</td>
</tr>
<tr>
<td>Flat oblong (ellipse) shaped area</td>
<td>Present study</td>
</tr>
<tr>
<td><strong>Segmentation</strong></td>
<td><strong>Rubinstein et al. (2008)</strong> and present study</td>
</tr>
<tr>
<td>Above 60% epitope residues exists in linear stretches of 3 or more residues</td>
<td>Sun et al., (2011) and present study</td>
</tr>
<tr>
<td>85% of epitopes has a linear stretch of 5 or more residues</td>
<td>Rubinstein et al. (2008) and Ofran et al. (2008)</td>
</tr>
<tr>
<td><strong>Secondary structure</strong></td>
<td><strong>Rubinstein et al. (2008)</strong></td>
</tr>
<tr>
<td>Enriched by loops</td>
<td>Present study</td>
</tr>
<tr>
<td>Depleted of strands and helixes</td>
<td>Present study</td>
</tr>
<tr>
<td><strong>Epitope position on the antigen</strong></td>
<td><strong>Andersen et al. (2006)</strong> and Rubinstein et al. (2008)</td>
</tr>
<tr>
<td>Epitopes are more surface exposed than the remaining antigen</td>
<td>Thornton et al. (1986)</td>
</tr>
<tr>
<td>Epitopes protrude from the antigen surface</td>
<td>Present study</td>
</tr>
<tr>
<td><strong>Orientation relative to the antibody</strong></td>
<td><strong>Present study</strong></td>
</tr>
<tr>
<td>Epitopes bind predominantly in a -30 to 60 degrees angle relative to the light to heavy antibody chain direction</td>
<td>Present study</td>
</tr>
<tr>
<td><strong>Amino acid composition</strong></td>
<td><strong>Andersen et al. (2006); Ofran et al. (2008); Rubinstein et al. (2008); Zhao and Li (2010) and Sun et al., (2011)</strong></td>
</tr>
<tr>
<td>Enriched by polar and charged amino acids and depleted of hydrophobic amino acids compared to non-epitope antigen residues, surface exposed antigen residues or general protein composition</td>
<td>Present study</td>
</tr>
<tr>
<td>No significant deviation from the non-epitope antigen surface, however a tendency for depletion of small hydrophobic amino acids is observed</td>
<td>Present study</td>
</tr>
<tr>
<td><strong>Amino acid coorporativeness</strong></td>
<td><strong>Rubinstein et al. (2008)</strong></td>
</tr>
<tr>
<td>Pairs of Tyr:Tyr, Cys:Pro, Asn:Tyr, Gly:Tyr, Asp:Pro, Thr:Tyr and Arg:Tyr are more frequently observed in epitopes compared to the remaining antigen surface</td>
<td>Sun et al., (2011)</td>
</tr>
<tr>
<td>Pairs of Asn:Tyr, His:Tyr and His:Met are more frequently observed in epitopes</td>
<td>Present study*</td>
</tr>
<tr>
<td><strong>Spatial amino acid composition</strong></td>
<td><strong>Present study</strong></td>
</tr>
<tr>
<td>Hydrophobic core flanked by charged amino acids</td>
<td>Present study</td>
</tr>
<tr>
<td>Preferable; hydrophobic amino acids closes to the antibody, then hydrophilic and furthest away positive charged amino acids</td>
<td>Present study</td>
</tr>
</tbody>
</table>
What does antibodies recognize in a protein?

- Antibody Fab fragment
- Protrusion index
- Surface Exposed
- Hydrophobic region
<table>
<thead>
<tr>
<th>Name</th>
<th>Input</th>
<th>Implemented epitope feature</th>
<th>LA</th>
<th>Performance Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM based prediction of linear B-cell epitopes (Linear)</td>
<td>Sequence</td>
<td>Amino acid composition. (Position specific amino acid composition using a 20 amino acid window of the primary sequence)</td>
<td>SVM</td>
<td>75% accuracy AUC = 0.84, Woe et al. 2010 [36]</td>
</tr>
<tr>
<td>Identification of conformational B-cell epitopes in an antigen from its primary sequence (Conformational)*</td>
<td>Sequence</td>
<td>Amino acid composition. (Pattern composition, which is the percentage of a given amino acid in a 15 amino acid stretch. See text)</td>
<td>SVM</td>
<td>96% accuracy MCC = 0.73, Ansari and Raghava 2010 [37]</td>
</tr>
<tr>
<td>EFSVR (Conformational)</td>
<td>Structure</td>
<td>Propensity scores based on: Epitope amino acid composition - Amino acid conservation - Side chain energy score - Contact number - Surface planarity score - Secondary structure composition.</td>
<td>SVR</td>
<td>AUC = 0.597, Liang et al., 2010 [34]</td>
</tr>
<tr>
<td>Mining for the antibody-antigen interacting associations that predict the B cell epitopes (Conformational)</td>
<td>Antigen-Sequence</td>
<td>Paratope-epitope co-occurring patterns of interacting residue pairs - Primary sequence cooperative of both paratope and epitope</td>
<td>None</td>
<td>AUC = 0.593, Zhao and Li 2009 [21]</td>
</tr>
<tr>
<td>Epitopia: a web-server for predicting B-cell epitopes</td>
<td>Sequence or Structure</td>
<td>Patch on surface of the protein equal to average epitope size - Amino acid composition - Frequency of helices - β-Strands, β-Sheets, α-Helices - Average curvature of atoms in epitopes - Proportion of patch atoms that reside within 4Å from a convex hull of the antigen</td>
<td>Naive Bayes Classifier</td>
<td>AUC = 0.625, Nielsen et al., 2008 [28]</td>
</tr>
<tr>
<td>COBBr: a novel system for predicting continuous B-cell epitopes (Linear)</td>
<td>Sequence</td>
<td>Measures of similarities between query peptides and a set of known epitope fragments: - Number of amino acids present in both sequence - Number of di-, tri-, tetrapeptides present in both sequences</td>
<td>SVM</td>
<td>AUC = 0.628, Svederalid and Balse 2009 [39]</td>
</tr>
<tr>
<td>Elipro: a new structure-based tool for the prediction of antibody epitopes (Conformational)</td>
<td>Structure</td>
<td>Prerogation Index (PI) Implements the Thornton method (see text) and clusters amino acid based on their PI value.</td>
<td>None</td>
<td>AUC = 0.732, Ponomare-Fko et al., 2008 [29]</td>
</tr>
<tr>
<td>Pepito: improved discontinuous B-cell epitope prediction using multiple threshold and half sphere exposure (Conformational)</td>
<td>Structure</td>
<td>Amino acid composition - Half-sphere exposure (Hamelryck 2005). Number of C-alpha atoms in 9Å lower and upper half-sphere individually.</td>
<td>None</td>
<td>AUC = 0.683, Svederalid and Balse 2009 [40]</td>
</tr>
<tr>
<td>Identification of discontinuous antigenic determinants on proteins based on shape complementarity (Conformational)</td>
<td>Structure</td>
<td>Paratope – epitope shape complementarity - Surface accessibility</td>
<td>None</td>
<td>AUC = 0.634, Rapesperger et al., 2007 [41]</td>
</tr>
<tr>
<td>Predicting B cell epitopes with network topology based amino acid indices (Linear)</td>
<td>Sequence</td>
<td>Relative connectivity propensity scores, based on network topology (connection between C-alpha atoms within 8Å), but similar to contact number to some extent half-sphere exposure.</td>
<td>None</td>
<td>AUC = 0.796, Huang et al., 2007 [42]</td>
</tr>
<tr>
<td>Prediction of linear B-cell epitopes using amino acid pair antigenicity scale (Linear)</td>
<td>Sequence</td>
<td>Amino acid cooperativity in epitopes in relation to non-epitopes</td>
<td>SVM</td>
<td>MCC = 0.37, Chen et al., 2007 [33]</td>
</tr>
<tr>
<td>DiscoTope: Prediction of residues in discontinuous B-cell epitopes using protein 3D structure (Conformational)</td>
<td>Structure</td>
<td>Amino acid composition - Residue contact number (number of C-alpha in a 18Å sphere)</td>
<td>None</td>
<td>AUC = 0.62, Andersen et al., 2006 [25]</td>
</tr>
</tbody>
</table>

* The author of this report has his doubts about this work. See text.
DiscoTope

- Prediction of residues in discontinuous B cell epitopes using protein 3D structures

Pernille Haste Andersen, Morten Nielsen and Ole Lund, Protein Science 2006
Predicting B-cell epitopes

DiscoTope 1.2 Server

DiscoTope 1.2 server predicts discontinuous B cell epitopes from protein three dimensional structures. The method utilizes calculation of surface accessibility (estimated in terms of contact numbers) and a novel epitope propensity amino acid score. The final scores are calculated by combining the propensity scores of residues in spatial proximity and the contact numbers.

Note: The DiscoTope server has been updated to improve user-friendliness. The server now predicts epitopes in complexes of multiple chains. Also, DiscoTope outputs are now easily downloaded and imported in spreadsheets. Furthermore, we have facilitated the visualization of prediction results.

SUBMISSION

Please choose one of the following three submission methods:

1. Chain(s) in an existing PDB entry. Use comma for separation of chain ids. If this box is unspecified, the prediction will be done using all chains in the pdb file.
   - PDB code:  
   - Chain(s):  

2. A file from your local disk containing a list of existing PDB entries with specified chain ID, one per line, in the format ‘entryname_chain’ e.g. 1zzs_B:
   - File name:  
   - Browse...

3. A file from your local disk containing your own structure in PDB format (not necessarily present in PDB):
The DiscoTope method
Some amino acids are preferred and disliked in the epitope.

Kringelum et al, 2011, manuscript in preparation
Some amino acids are preferred and disliked in the epitope.
Epitopes reside on the surface of the protein.

Kringelum et al., 2011
The DiscoTope method

• Predictions score for each residue are calculated by summing the epitope likelihood (propensity) of surrounding residues and subtracting the neighbor count.

\[ DS(r) = ps(r, 10\text{Å}) - 0.5 \times N_{neighbors} \]

\[ ps(r, k) = \sum_i P_{\text{epi}}(r_i) \text{ if } |r - r_i| < k \]

\[
DS(r, 10\text{Å}) = 2 \times P(\text{Tyr}) + P(\text{Phe}) + P(\text{Ale}) - 0.5 \times 4
\]

\[
DS(r, 10\text{Å}) = 2 \times 0.25 - 0.3 - 0.7 - 0.5 \times 4 = -2.5
\]

Andersen et al., 2006
The DiscoTope method

• Predictions score for each residue are calculated by summing the epitope likelihood (propensity) of surrounding residues and subtracting the neighbor count.

$$DS(r) = ps(r, 10\text{Å}) - 0.5 \times N_{neighbors}$$

$$ps(r, k) = \sum_i P_{epi}(r_i) \text{ if } |r - r_i| < k$$

Performance: $$A_{roc} = 0.700$$

On a dataset of 75 antigen-antibody complexes divided in 25 proteins

Andersen et al., 2006
The DiscoTope method

• Predictions score for each residue are calculated by summing the epitope likelihood (propensity) of surrounding residues and subtracting the neighbor count

\[ DS(r) = ps(r, 10\text{Å}) - 0.5 \times N_{neighbors} \]

\[ ps(r, k) = \sum_i P_{epi}(r_i) \text{ if } |r - r_i| < k \]

Performance: \( A_{roc} = 0.700 \)

On a dataset of 75 antigen-antibody complexes divided in 25 proteins

Andersen et al., 2006
However position matters

- Uneven spatial distribution of amino acid in epitopes

Kringelum et al, 2012,
Propensity score function

\[ PS(r) = \sum_i \beta_i \cdot ls(r_i) \]

\[ \beta_i = 0.8 \times (1 - (d_i / 21.6 \text{Å})) + 0.2 \]
Propensity score function

\[ PS(r) = \sum_i \beta_i \cdot ls(r_i) \]

\[ \beta_i = 0.8 \times (1 - (d_i/21.6\text{Å})) + 0.2 \]
Identify Neighbor residues within 21.6Å ($C_\alpha-C_\alpha$)

$PS(\text{THR256})$

$PS(r) = \sum_i \beta_i \cdot ls(r_i)$

$\beta_i = 0.8 \cdot (1 - (d_i/21.6\text{Å})) + 0.2$
Propensity score function

$PS(THR256)$

1) Identify Neighbor residues within 21.6Å ($C_{\alpha}-C_{\alpha}$)

\[ PS(r) = \sum_i \beta_i \cdot ls(r_i) \]

\[ \beta_i = 0.8 \times (1 - (d_i/21.6\text{Å})) + 0.2 \]
Propensity score function

1) Identify Neighbor residues within 21.6 Å (Cα-Cα)

2) Calculate summed propensity score:

\[
\begin{align*}
\text{ls}(\text{THR}) & = -0.23 \\
\beta_{256} & = 0.8 \cdot (1 - \frac{0.0}{21.6}) + 0.2 = 1.0 \\
\text{ls}(\text{THR}) \cdot \beta_{256} & = -0.23 \cdot 1.0 = -0.23
\end{align*}
\]

\[
PS(r) = \sum_{i} \beta_i \cdot \text{ls}(r_i)
\]

\[
\beta_i = 0.8 \cdot (1 - (d_i/21.6\text{Å})) + 0.2
\]

\[
d_{256} = 0.0 \text{Å}
\]
Propensity score function

1) Identify Neighbor residues within 21.6 Å (Cα - Cα)
2) Calculate summed propensity score:

\[
PS(THR_{256}) = \sum \beta_i \cdot ls(r_i)
\]

\[
\beta_i = 0.8 \times (1 - (d_i/21.6)) + 0.2
\]

\[
l_s(THR) = -0.23
\]

\[
\beta_{256} = 0.8 \times (1 - 0.0/21.6) + 0.2 = 1.0
\]

\[
l_s(THR) \times \beta_{256} = -0.23 \times 1.0 = -0.23
\]

\[
l_s(ASP) = 2.5
\]

\[
\beta_{255} = 0.8 \times (1 - 3.8/21.6) + 0.2 = 0.86
\]

\[
l_s(THR) \times \beta_{255} = 2.5 \times 0.86 = 2.15
\]

\[
d_{255} = 3.8\text{Å}
\]
Propensity score function

$$PS(r) = \sum_i \beta_i \cdot ls(r_i)$$

$$\beta_i = 0.8 \times (1 - (d_i/21.6\AA)) + 0.2$$

1) Identify Neighbor residues within 21.6Å ($C_{\alpha}-C_{\alpha}$)
2) Calculate summed propensity score:

<table>
<thead>
<tr>
<th>Residue</th>
<th>Propensity</th>
<th>$d_{254}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>THR</td>
<td>$-0.23\beta_{256} = 0.8 \times (1 - (0.0/21.6)) + 0.2$</td>
<td>6.1Å</td>
</tr>
<tr>
<td>THR</td>
<td>$-0.23\beta_{254} = 0.8 \times (1 - (6.1/21.6)) + 0.2$</td>
<td></td>
</tr>
<tr>
<td>ASP</td>
<td>$2.5\beta_{255} = 0.8 \times (1 - (3.8/21.6)) + 0.2$</td>
<td></td>
</tr>
</tbody>
</table>
**Propensity score function**

\[ PS(r) = \sum_i \beta_i \cdot ls(r_i) \]

\[ \beta_i = 0.8 \times (1 - (d_i/21.6\AA)) + 0.2 \]

<table>
<thead>
<tr>
<th>Residue</th>
<th>( ls(r_i) )</th>
<th>( \beta_i )</th>
<th>( ls(r_i) \times \beta_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>THR</td>
<td>-0.23</td>
<td>1.0</td>
<td>-0.23</td>
</tr>
<tr>
<td>ASP</td>
<td>2.5</td>
<td>0.86</td>
<td>2.15</td>
</tr>
<tr>
<td>THR</td>
<td>-0.23</td>
<td>0.77</td>
<td>-0.18</td>
</tr>
<tr>
<td>PRO</td>
<td>1.2</td>
<td>0.24</td>
<td>0.29</td>
</tr>
</tbody>
</table>

1) Identify Neighbor residues within 21.6Å (C\(_\alpha\)-C\(_\alpha\))
2) Calculate summed propensity score:

\[ ls(THR) = -0.23 \]
\[ \beta_{256} = 0.8 \times (1 - 0.0/21.6) + 0.2 = 1.0 \]
\[ ls(THR) \times \beta_{256} = -0.23 \times 1.0 = -0.23 \]

\[ ls(ASP) = 2.5 \]
\[ \beta_{255} = 0.8 \times (1 - 3.8/21.6) + 0.2 = 0.86 \]
\[ ls(ASP) \times \beta_{255} = 2.5 \times 0.86 = 2.15 \]

\[ ls(THR) = -0.23 \]
\[ \beta_{254} = 0.8 \times (1 - 6.1/21.6) + 0.2 = 0.77 \]
\[ ls(THR) \times \beta_{254} = -0.23 \times 0.77 = -0.18 \]

\[ ls(PRO) = 1.2 \]
\[ \beta_{254} = 0.8 \times (1 - 20.6/21.6) + 0.2 = 0.24 \]
\[ ls(PRO) \times \beta_{254} = 1.2 \times 0.24 = 0.29 \]

\[ d_{247} = 20.6\AA \]
Propensity score function

$PS(r) = \sum_i \beta_i \cdot ls(r_i)$

$\beta_i = 0.8 \times (1 - (d_i / 21.6\text{Å})) + 0.2$

<table>
<thead>
<tr>
<th>Residue</th>
<th>$ls(r_i)$</th>
<th>$\beta_i$</th>
<th>$ls(r_i)\beta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>THR</td>
<td>-0.23</td>
<td>1.0</td>
<td>-0.23</td>
</tr>
<tr>
<td>ASP</td>
<td>2.5</td>
<td>0.86</td>
<td>2.15</td>
</tr>
<tr>
<td>THR</td>
<td>-0.23</td>
<td>0.77</td>
<td>-0.18</td>
</tr>
<tr>
<td>PRO</td>
<td>1.2</td>
<td>0.24</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Summation of scores

$PS(THR256) = 0.39$
HS(THR256)

1) Create the upper half-sphere, which is the half-sphere where the residue side-chain is pointing
**Half-sphere exposure**

1. Create the upper half-sphere, which is the half-sphere where the residue side-chain is pointing.
2. Count neighbor residues within the half-sphere (nr of $C_\alpha$-atoms).
3. As high counts mean highly buried, the counts are multiplied by -1.

\[
\text{HS(THR256)} = -5
\]
The DiscoTope 2.0 Score

1) Calculate Propensity score
2) Calculate half-sphere exposure
3) The final score is a weighted sum of the Propensity score and half-sphere exposure

\[
\text{DS}(\text{THR256}) = (1 - \alpha) \times \text{PS}(\text{THR256}) + \alpha \times \text{HS}(\text{THR256})
\]

\[
\text{PS}(\text{THR256}) = 0.39
\]
\[
\text{HS}(\text{THR256}) = -5
\]
\[
\alpha = 0.115
\]

\[
\text{DS}(\text{THR256}) = (1 - 0.115) \times 0.39 + 0.115 \times (-5)
\]

\[
\text{DS}(\text{THR256}) = 0.885 \times 0.39 + 0.115 \times (-5)
\]

\[
\text{DS}(\text{THR256}) = -0.23
\]
Performance and limitations

Average AUC = 0.741
Performance and limitations

Glycosylated proteins
Performance and limitations

- Glycosylation effects predictions

Gp120

Hemagglutinin
Performance and limitations

Small fragments (<120 residues) of larger biological units
Performance and limitations

- Inclusion of biological units enhances performance

Potassium Channel

\[ A_{rec} = 0.737 \]  \[ A_{rec} = 0.880 \]  \[ A_{rec} = 0.946 \]
Performance and limitations

- External Benchmark Dataset
  - 52 antigen:antibody structures
  - 33 homology groups
- Performance: 0.731 AUC

<table>
<thead>
<tr>
<th># Residues</th>
<th>DiscoTope-2.0</th>
<th>DiscoTope-1.2</th>
<th>PEPITO</th>
<th>ElliPro</th>
<th>SEPPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV</td>
<td>0.178</td>
<td>0.176</td>
<td>0.169</td>
<td>0.134</td>
<td>0.142</td>
</tr>
<tr>
<td>Sens</td>
<td>0.168</td>
<td>0.150</td>
<td>0.147</td>
<td>0.134</td>
<td>0.135</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV</td>
<td>0.141</td>
<td>0.133</td>
<td>0.138</td>
<td>0.120</td>
<td>0.138</td>
</tr>
<tr>
<td>Sens</td>
<td>0.262</td>
<td>0.220</td>
<td>0.237</td>
<td>0.234</td>
<td>0.257</td>
</tr>
</tbody>
</table>
Conclusions

• DiscoTope V2.0 outperforms similar methods
  • High performance on 15/25
  • Medium performance on 7/25
  • Fail on 3/25

• Inclusion of surface measures does only slightly enhance predictions

• Use the entire biological unit, when possible
  • Small fragments (< 120 residues) have lower performance

• Glycosylation might course the prediction to fail
  • Check for clash between predicted epitopes and glycosylation sites
Rational vaccine design

PATHOGEN PROTEIN
KVFGRCLEAAAMKRHGLDNRYRGYS
LGNWVCAAKFESNF

Rational Vaccine Design
Rational B-cell epitope design

• Protein target choice

• Structural analysis of antigen

- Known structure or homology model
- Precise domain structure
- Physical annotation (flexibility, electrostatics, hydrophobicity)
- Functional annotation (sequence variations, active sites, binding sites, glycosylation sites, etc.)
Rational B-cell epitope design

• Protein target choice
• Structural annotation
• Epitope prediction and ranking

- Surface accessibility
- Protrusion index
- Conserved sequence
- Glycosylation status
Rational B-cell epitope design

- Protein target choice
- Structural annotation
- Epitope prediction and ranking
- Optimal Epitope presentation

- Fold minimization, or
- Design of structural mimics
- Choice of carrier (conjugates, DNA plasmids, virus like particles)
- Multiple chain protein engineering
Conclusions

• Rational vaccines can be designed to induce strong and epitope-specific B-cell responses
• Selection of protective B-cell epitopes involves structural, functional and immunogenic analysis of the pathogenic proteins
• When you can: Use protein structure for prediction
• Structural modeling tools are helpful in prediction of epitopes, design of epitope mimics and optimal epitope presentation