



B Cell Epitope Prediction

tools.iedb.org

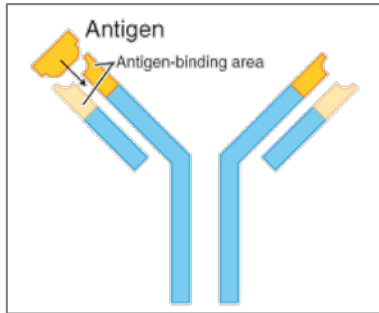
Presented by: Bjoern Peters, PI

Outline of topics

1. B cell epitope biology recap
2. Prediction tools on IEDB
3. Linear sequence-based epitope prediction methods
4. Discontinuous 3D structure-based epitope prediction methods
5. Computational antibody design
 - a. Antigen and Antibody structure modelling
 - b. Antibody-protein docking

Example: Ab binding HA1

PDB ID: 1EO8

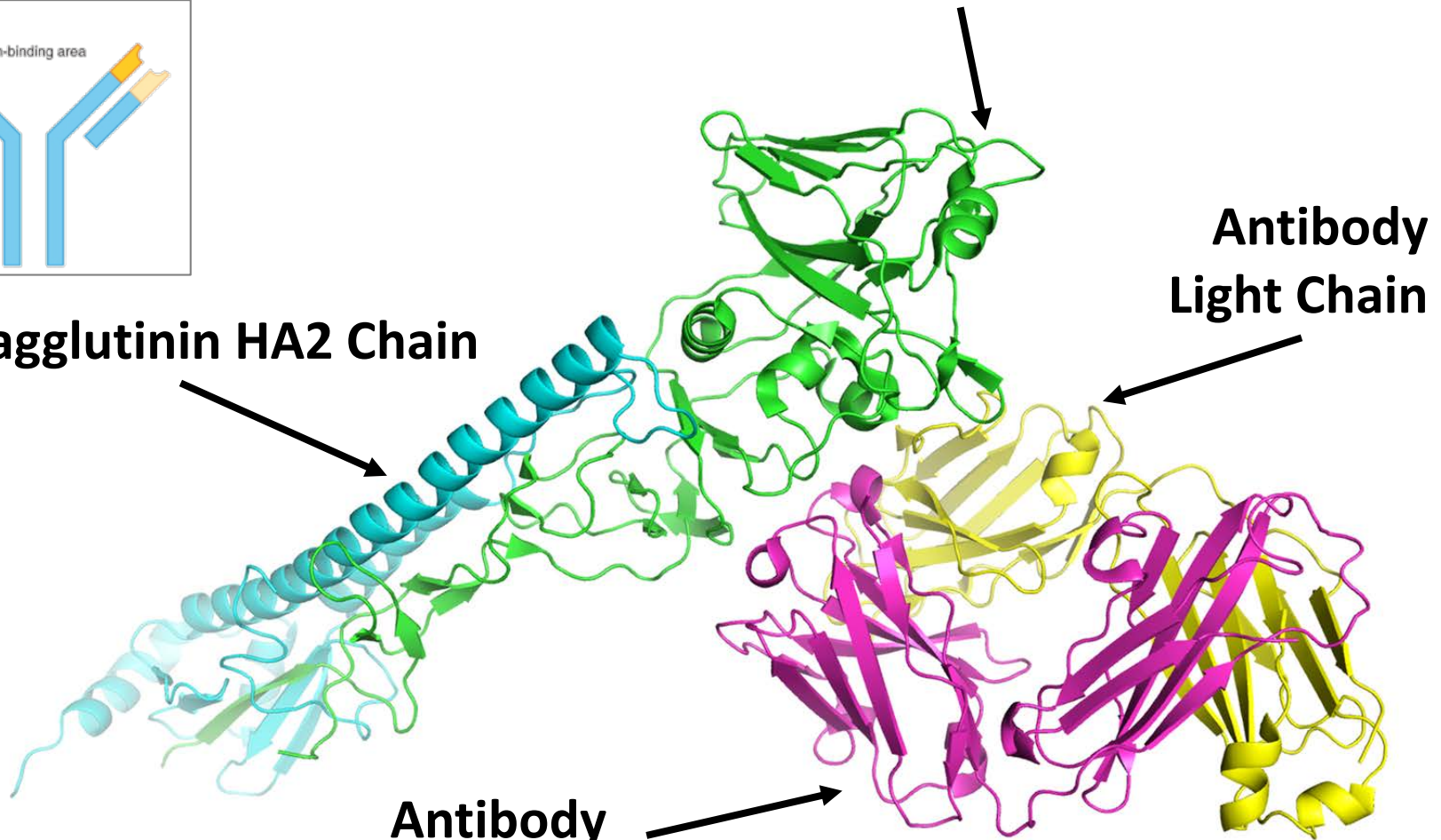


Hemagglutinin HA2 Chain

Antigen
Hemagglutinin HA1 Chain
from Influenza A Virus

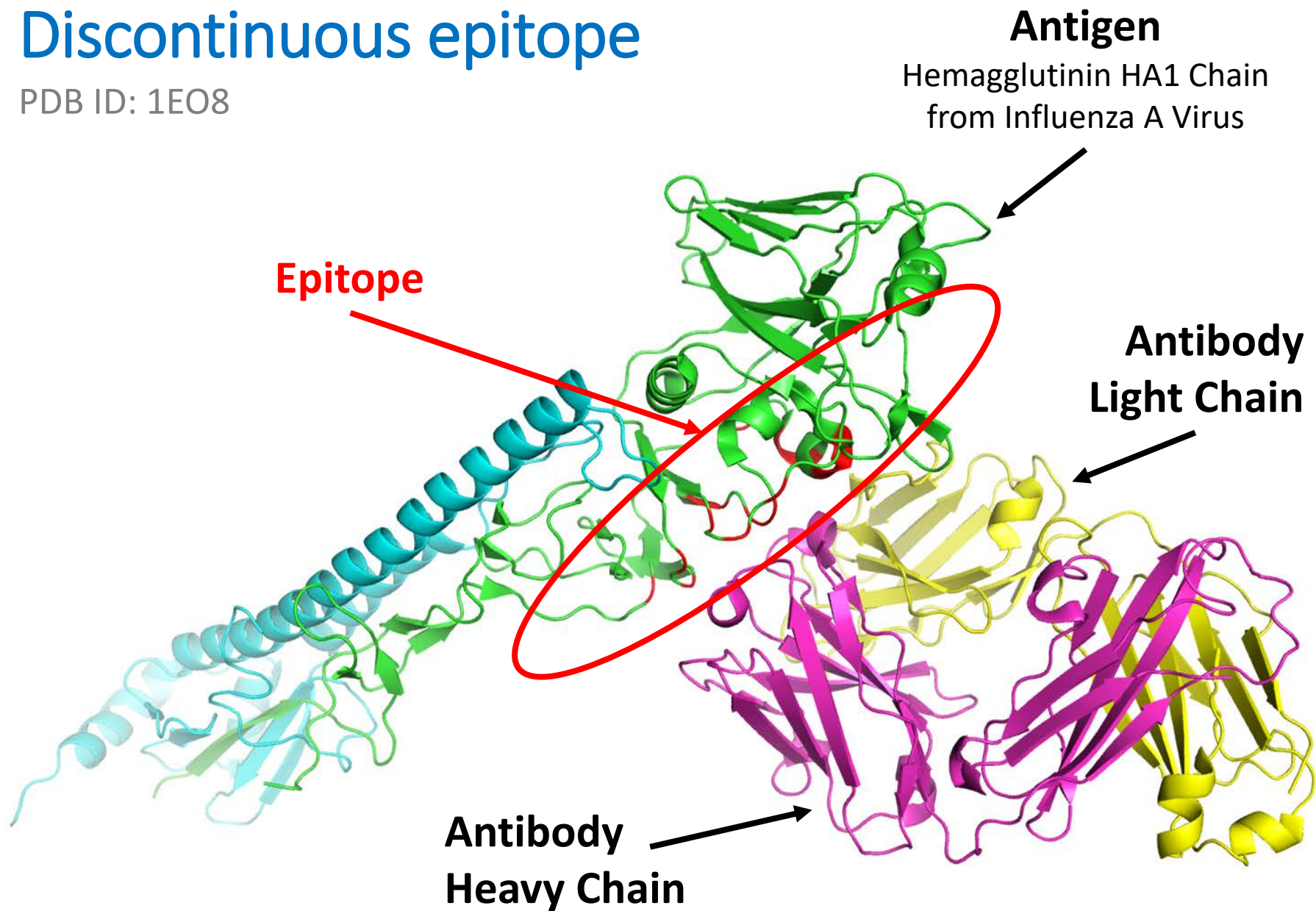
**Antibody
Light Chain**

**Antibody
Heavy Chain**



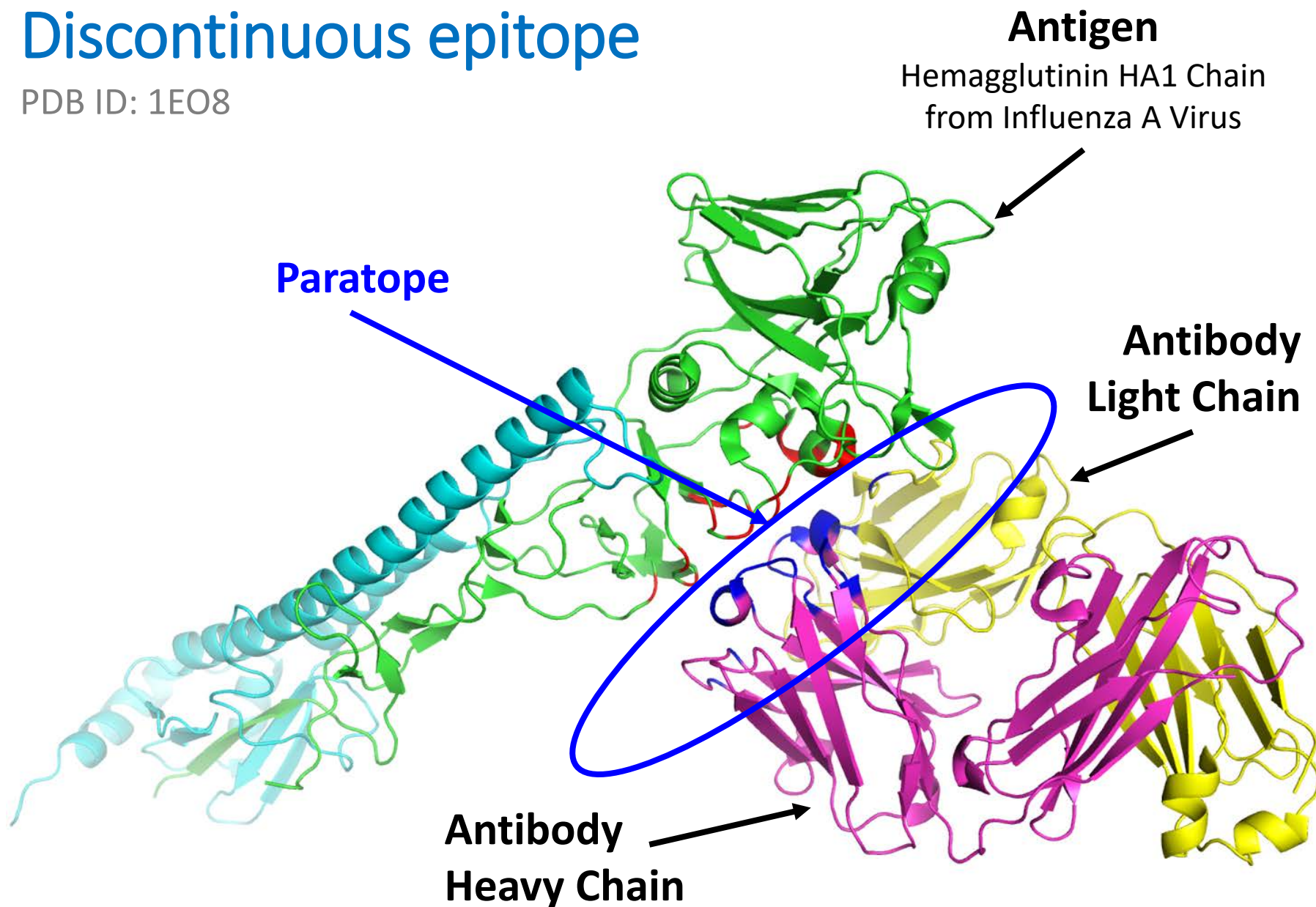
Discontinuous epitope

PDB ID: 1EO8



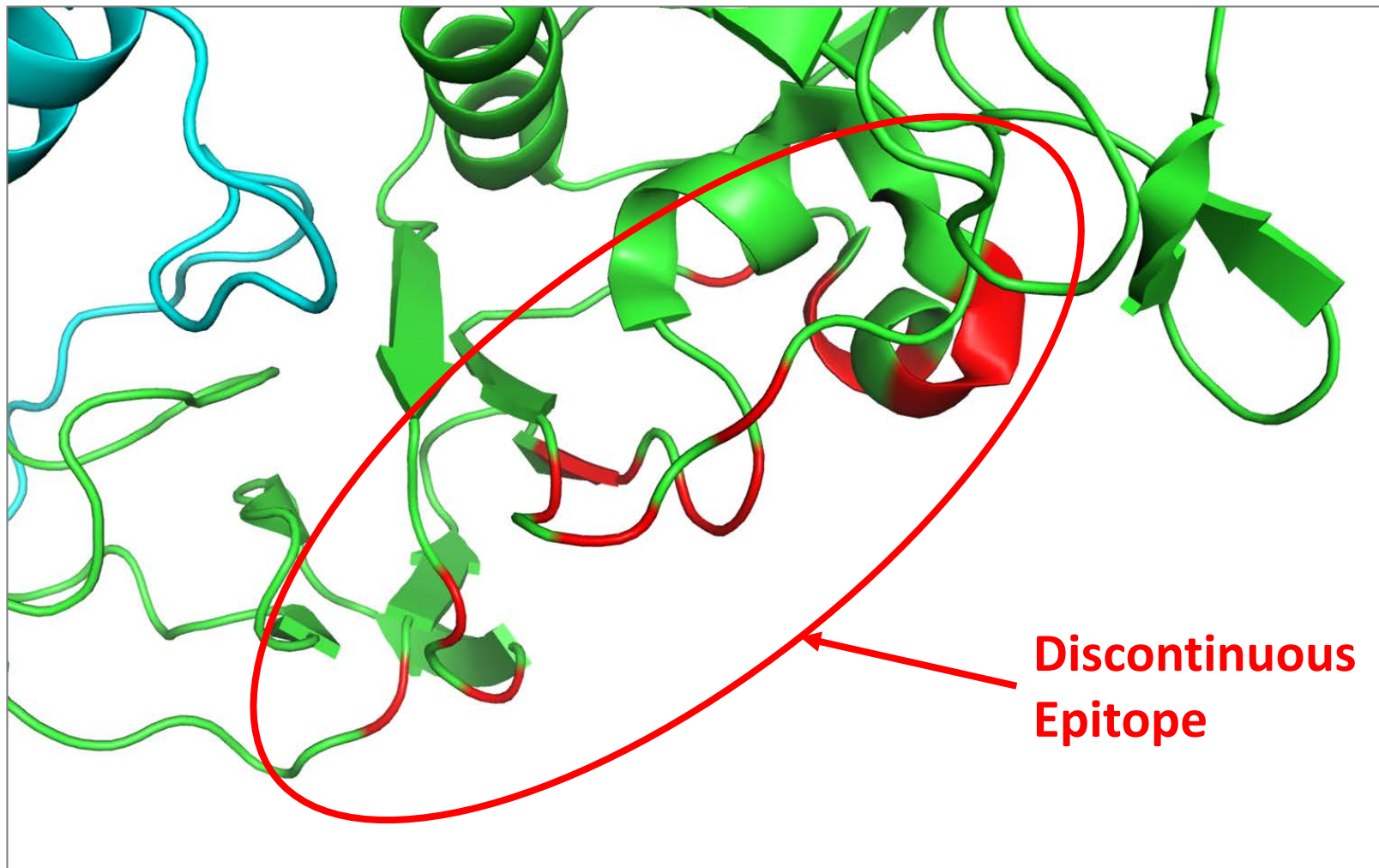
Discontinuous epitope

PDB ID: 1EO8



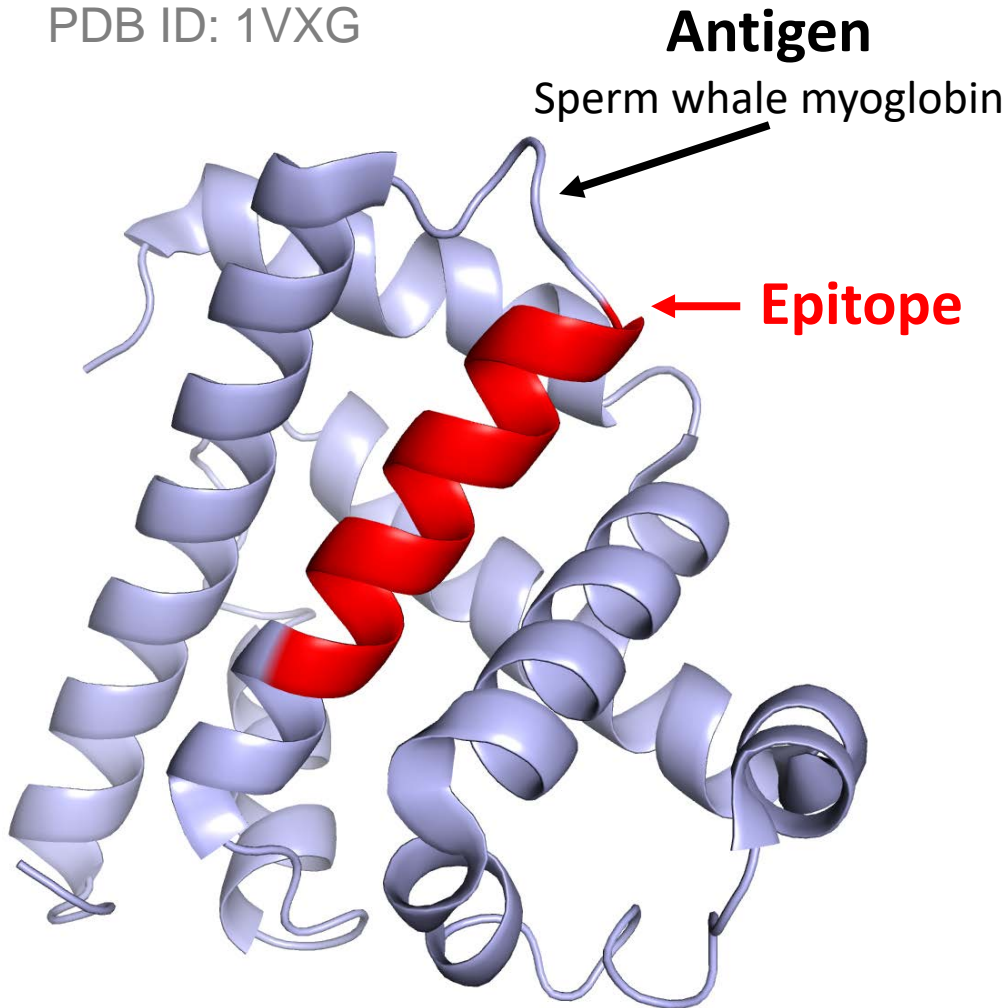
Discontinuous epitope

PDB ID: 1EO8



B cell epitopes

PDB ID: 1VXG



- Protein antigens usually contain both linear & discontinuous epitopes
 - “Linear” aka sequential or continuous
 - “Discontinuous” aka non-sequential or conformational epitopes
- More than 90% of the B cell epitopes are estimated to be discontinuous*

**Barlow et al, Nature. 1986.*

**Van Regenmortel, Methods. 1996.*

B cell prediction tools on IEDB

<http://tools.iedb.org/main/bcell/>

IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

B Cell Epitope Prediction Tools

B Cell Epitope Prediction

[Prediction of linear epitopes from protein sequence](#)

A collection of methods to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs.

[Discotope - Prediction of epitopes from protein structure](#)

This method incorporates solvent-accessible surface area calculations, as well as contact distances into its prediction of B cell epitope potential along the length of a protein sequence.

[ElliPro - Epitope prediction based upon structural protrusion](#)

This method predicts epitopes based upon solvent-accessibility and flexibility.

[Methods for modeling and docking of antibody and protein 3D structures](#)

This page provides information on available methods for modeling and docking of antibody and protein 3D structures.

Structure Tools



[LYRA \(Lymphocyte Receptor Automated Modelling\)*](#)

The LYRA server predicts structures for either T-Cell Receptors (TCR) or B-Cell Receptors (BCR) using homology modelling. Framework templates are selected based on BLOSUM score, and complementary determining regions (CDR) are then selected if needed based on a canonical structure model and grafted onto the framework templates.



[SCEptRe: Structural Complexes of Epitope Receptor](#)

SCEptRe provides weekly updated, non-redundant, user customized benchmark datasets with information on the immune receptor features for receptor-specific epitope predictions. This tool extracts weekly updated 3D complexes of antibody-antigen, TCR-pMHC and MHC-ligand from the Immune Epitope Database (IEDB) and clusters them based on antigens, receptors and epitopes to generate benchmark datasets. Users can customize structural quality and clustering parameters (e.g. resolution, R free factors, antigen or epitope sequence identity) to generate these datasets based on their need.



** Tools under AR Labs which are experimental and are not quite ready for production yet. They are intended for further research, updates and testing.*

Epitope prediction

When to use epitope prediction methods?

- You have verified thoroughly that no information is available in the IEDB on the antigen of your interest
- You want to know all the candidate antigenic determinants in an antigen of your interest other than epitopes provided in the IEDB

Sequence-based epitope prediction

Linear epitope prediction:

- Amino acid physicochemical properties-based methods
 - Features which have been correlated with the location of continuous epitopes
 - β -Turns (*Chou & Fasman*)
 - Surface Accessibility (*Emini*)
 - Flexibility (*Karplus & Schulz*)
 - Antigenic propensity: occurrence of residues in epitopes (*Kolaskar & Tongaonkar*)
 - Hydrophilicity (*Parker*)
- Only provide information on protein regions which are likely to be accessible for antibody binding

Linear epitope prediction

Linear epitope prediction methods:

- Machine learning algorithms
 - Positive and negative training datasets are used
 - Combination of one or more amino acid scales are used as an input to one of the machine learning algorithms
 - Random Forest (BepiPred-2.0)
 - ANN: Artificial Neural Network (ABCpred)
 - SVM: Support Vector Machine (BCpred, FBCpred)
 - Prediction accuracy is optimized

B cell prediction tools on IEDB

<http://tools.iedb.org/main/bcell/>

IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) **[B Cell Tools](#)** [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

B Cell Epitope Prediction Tools

B Cell Epitope Prediction

[Prediction of linear epitopes from protein sequence](#)

A collection of methods to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs.

[Discotope - Prediction of epitopes from protein structure](#)

This method incorporates solvent-accessible surface area calculations, as well as contact distances into its prediction of B cell epitope potential along the length of a protein sequence.

[ElliPro - Epitope prediction based upon structural protrusion](#)

This method predicts epitopes based upon solvent-accessibility and flexibility.

[Methods for modeling and docking of antibody and protein 3D structures](#)

This page provides information on available methods for modeling and docking of antibody and protein 3D structures.

Structure Tools



[LYRA \(Lymphocyte Receptor Automated Modelling\)*](#)

The LYRA server predicts structures for either T-Cell Receptors (TCR) or B-Cell Receptors (BCR) using homology modelling. Framework templates are selected based on BLOSUM score, and complementary determining regions (CDR) are then selected if needed based on a canonical structure model and grafted onto the framework templates.

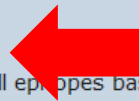


[SCEptRe: Structural Complexes of Epitope Receptor](#)

SCEptRe provides weekly updated, non-redundant, user customized benchmark datasets with information on the immune receptor features for receptor-specific epitope predictions. This tool extracts weekly updated 3D complexes of antibody-antigen, TCR-pMHC and MHC-ligand from the Immune Epitope Database (IEDB) and clusters them based on antigens, receptors and epitopes to generate benchmark datasets. Users can customize structural quality and clustering parameters (e.g. resolution, R free factors, antigen or epitope sequence identity) to generate these datasets based on their need.



* : Tools under AR Labs which are experimental and are not quite ready for production yet. They are intended for further research, updates and testing.



Linear B cell prediction

<http://tools.iedb.org/bcell/>

IEDB Analysis Resource

Home Help Example Reference Download Contact

Antibody Epitope Prediction

Specify Input

Enter a Swiss-Prot ID (example: P02185)

Or enter a protein sequence in plain format (50000 residues maximum):

Choose a method:

- [BepiPred Linear Epitope Prediction](#) ← Links to help tab
- [BepiPred Linear Epitope Prediction 2.0](#)
- [Chou & Fasman Beta-Turn Prediction](#)
- [Emini Surface Accessibility Prediction](#)
- [Karplus & Schulz Flexibility Prediction](#)
- [Kolaskar & Tongaonkar Antigenicity](#)
- [Parker Hydrophilicity Prediction](#)

Submit Reset

1. Input protein sequence

Entry allowed via Swiss-Prot ID or plain format

2. Select prediction method

BepiPred is the default & recommended method

Visit Help & Reference tabs to learn about a prediction method

<http://tools.iedb.org/bcell/>

IEDB Analysis Resource

[Home](#) [Help](#) [Example](#) [Reference](#) [Download](#) [Contact](#)

Antibody Epitope Prediction - Tutorial

I. Methods for predicting continuous antibody epitope from protein sequences

General basis: Parameters such as hydrophilicity, flexibility, accessibility, turns, exposed surface, polarity and charge have been correlated with the location of continuous epitopes. This has led to a search for empirical rules that will allow epitopes to be predicted from certain features of the protein sequence. All prediction calculations are based on amino acid scales. Each scale consists of 20 values assigned to each of the amino acid residues on the basis of their relative importance as described by the scale.

General method: When computing the score for a given residue i , the amino acids in an interval of the chosen window size n , centered on residue i , are considered. In other words, for a window size n , the $i - (n-1)/2$ neighboring residues on each side of residue i are considered. Unless specified, the score for residue i is the average of the scale values for these amino acids (see table 1 for details). In general, a window size of 5 to 7 is appropriate for finding regions that may potentially be antigenic.

Interpretation of output graphs and tables: On the graphs, the Y-axis depicts for each residue the corresponding score. On the tables, the X-axis depicts for each residue the corresponding score. The larger the score for the residues, the more likely they are to be part of an epitope (those residues are colored in yellow on the graphs). However, the scores per se, either linear or discontinuous, -- they might only guide the researchers to further explore the epitopes.

Table 1. Implemented methods

Method																																								
Chou and Fasman beta turn prediction																																								
<ul style="list-style-type: none">Reference: Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequences. <i>J Biol Chem</i> 1978;47:45-148.Description: The rationale for predicting turns to predict antibody epitopes is based on the paper by Pellegri et al. 1983-99. Instead of implementing the turn scale of that paper which has some non-standard properties, we use the scale which is commonly used to predict beta turns as described in the reference link above.																																								
Scale:																																								
<table border="1"><thead><tr><th>A</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th><th>H</th><th>I</th><th>K</th><th>L</th><th>M</th><th>N</th><th>P</th><th>Q</th><th>R</th><th>S</th><th>T</th><th>V</th><th>W</th><th>Y</th></tr></thead><tbody><tr><td>0.66</td><td>1.19</td><td>1.46</td><td>0.74</td><td>0.6</td><td>1.56</td><td>0.95</td><td>0.47</td><td>1.01</td><td>0.59</td><td>0.6</td><td>1.56</td><td>1.52</td><td>0.98</td><td>0.95</td><td>1.43</td><td>0.96</td><td>0.5</td><td>0.96</td><td>1.14</td></tr></tbody></table>	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	0.66	1.19	1.46	0.74	0.6	1.56	0.95	0.47	1.01	0.59	0.6	1.56	1.52	0.98	0.95	1.43	0.96	0.5	0.96	1.14
A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y																					
0.66	1.19	1.46	0.74	0.6	1.56	0.95	0.47	1.01	0.59	0.6	1.56	1.52	0.98	0.95	1.43	0.96	0.5	0.96	1.14																					
Emini surface accessibility scale																																								
<ul style="list-style-type: none">Reference: Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a synthetic peptide. <i>J Biol Chem</i> 1985 Sep;55(3):836-9.Description: The calculation was based on surface accessibility scale on a product instead of an addition which was obtained using the formulae $S_n = (n+1) \cdot (0.37)^{-n}$ where S_n is the surface probability, n is the fraction from 1 to 6. A hexapeptide sequence with S_n greater than 1.0 indicates an increased probability for being found in a surface.																																								
Scale:																																								
<table border="1"><thead><tr><th>A</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th><th>H</th><th>I</th><th>K</th><th>L</th><th>M</th><th>N</th><th>P</th><th>Q</th><th>R</th><th>S</th><th>T</th><th>V</th><th>W</th><th>Y</th></tr></thead><tbody><tr><td>0.49</td><td>0.26</td><td>0.81</td><td>0.84</td><td>0.42</td><td>0.48</td><td>0.66</td><td>0.34</td><td>0.97</td><td>0.4</td><td>0.48</td><td>0.78</td><td>0.75</td><td>0.84</td><td>0.95</td><td>0.65</td><td>0.7</td><td>0.36</td><td>0.51</td><td>0.76</td></tr></tbody></table>	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	0.49	0.26	0.81	0.84	0.42	0.48	0.66	0.34	0.97	0.4	0.48	0.78	0.75	0.84	0.95	0.65	0.7	0.36	0.51	0.76
A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y																					
0.49	0.26	0.81	0.84	0.42	0.48	0.66	0.34	0.97	0.4	0.48	0.78	0.75	0.84	0.95	0.65	0.7	0.36	0.51	0.76																					
Karplus and Schulz flexibility scale																																								
<ul style="list-style-type: none">Reference: Karplus PA, Schulz GE. Prediction of Chain Flexibility in Proteins - A tool for the Selection of Peptides for Synthesis. <i>J Biol Chem</i> 1985; 260:212-3.																																								

IEDB Analysis Resource

[Home](#) [Help](#) [Example](#) [Reference](#) [Download](#) [Contact](#)

References

Chou & Fasman Beta-Turn Prediction:

Chou PY, Fasman GD. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol* 47:45-148.
[PMID: 364941](#)

Emini Surface Accessibility Prediction:

Emini EA, Hughes JV, Perlow DS, Boger J. 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55:836-839.
[PMID: 2991600](#)

Karplus & Schulz Flexibility Prediction:

Karplus PA, Schulz GE. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72:212-213.
[Naturwissenschaften](#)

Kolaskar & Tongaonkar Antigenicity:

Kolaskar AS, Tongaonkar PC. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276:172-174.
[PMID: 1702393](#)

Parker Hydrophilicity Prediction:

Parker JM, Guo D, Hodges RS. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry* 25:5425-5432.
[PMID: 2430611](#)

Bepred Linear Epitope Prediction:

Larsen JE, Lund O, Nielsen M. 2006. Improved method for predicting linear B-cell epitopes. *Immunome Res* 2:2.
[PMID: 16635264](#)

Ponomarenko JV, Bourne PE. 2007. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol* 7:64.
[PMID: 17910770](#)

Haste Andersen P, Nielsen M, Lund O. 2006. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 15:2558-2567.
[PMID: 17001032](#)

Bepred Linear Epitope Prediction 2.0:

Jespersen MC, Peters B, Nielsen M, Marcotilli P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res (Web Server issue)* 2:2.
[PMID: 28472356](#)

Linear B cell prediction -example

<http://tools.iedb.org/bcell/>

IEDB Analysis Resource

Home Help Example Reference Download Contact

Antibody Epitope Prediction

Specify Input

Enter a Swiss-Prot ID (example: P02185)

Or enter a protein sequence in plain format (50000 residues maximum):

```
VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFD RFKHLKTEAEMKASE
DLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIKPIKYLEFISEAIIHVLHSRHPGN
FGADAGGAMNKALELFRKDIAAKYKELGYQG
```

Choose a method:

- [Bepipred Linear Epitope Prediction](#)
- [Bepipred Linear Epitope Prediction 2.0](#)
- [Chou & Fasman Beta-Turn Prediction](#)
- [Emini Surface Accessibility Prediction](#)
- [Karplus & Schulz Flexibility Prediction](#)
- [Kolaskar & Tongaonkar Antigenicity](#)
- [Parker Hydrophilicity Prediction](#)

Example Sequence
Sperm Whale Myoglobin
Swiss-Prot ID P02185



Linear B cell prediction -example

<http://tools.iedb.org/bcell/>

Bepipred Linear Epitope Prediction Results

Input Sequences

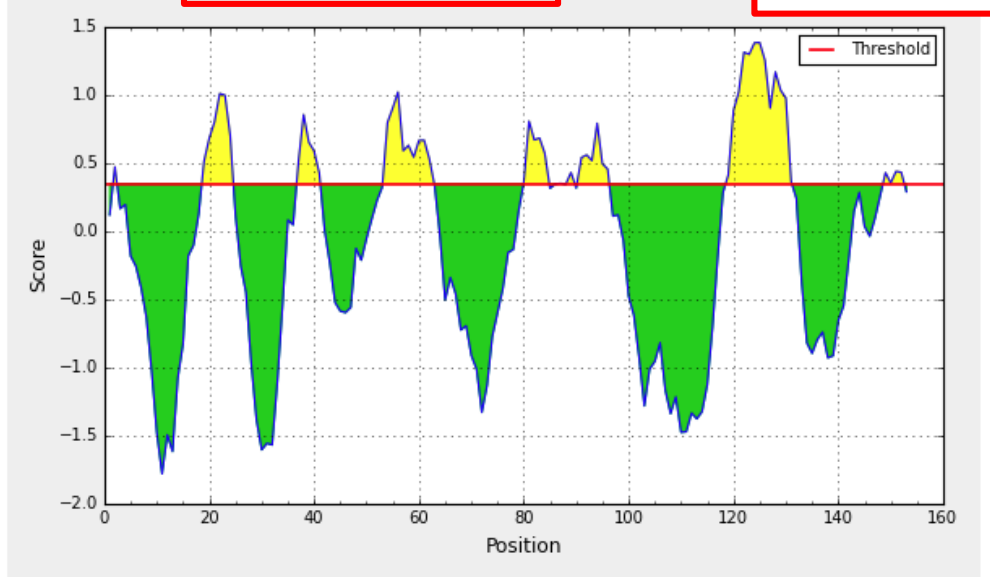
1 VLSEGEWQLV LHVAKVEAD VAGHGQDILI RLFKSHPETL EKDFRFKHLK TEAEMKASED
 61 LKKHGVTVLT ALGAILKKGK HHEAELKPLA QSHATKHKIP IKYLEFISEA IIVLHSRHP
 121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG

Center position: 4

Threshold: 0.350

Recalculate

Adjustable threshold



Average: -0.105 Minimum: -0.028 Maximum: 1.390

Average score of a protein chosen as a threshold by default

Predicted peptides:

No.	Start	End	Peptide	Length
1	2	2	L	1
2	19	24	ADVAGH	6
3	37	41	PETLE	5
4	54	62	EMKASEDLK	9
5	80	84	GHHEA	5
6	87	87	K	1
7	89	89	L	1
8	91	96	QSHATK	6
9	119	131	HPGNFGADAGGAM	13
10	149	152	LGYQ	4

Predicted residue scores:

Position	Residue	Score	Assignment
1	V	0.121	.
2	L	0.476	E
3	S	0.168	.
4	E	0.198	.
5	G	-0.180	.
6	E	-0.255	.
7	W	-0.412	.
8	Q	-0.631	.
9	L	-1.022	.
10	V	-1.482	.
11	L	-1.784	.
12	H	-1.496	.
13	V	-1.619	.
14	W	-1.079	.
15	A	-0.829	.
16	K	-0.179	.
17	V	-0.102	.
18	E	0.131	.
19	A	0.516	E
20	D	0.686	E
21	V	0.805	E
22	A	1.015	E
23	G	1.003	E
24	H	0.705	E

Method comparisons

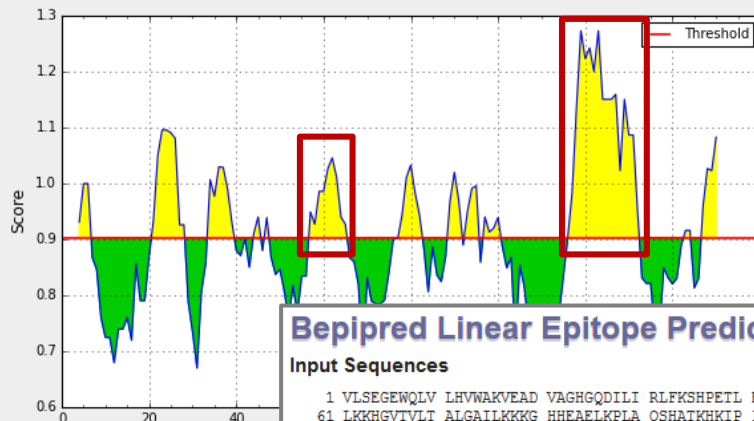
<http://tools.iedb.org/bcell/>

Chou & Fasman Beta-Turn Prediction Results

Input Sequences

```
1 VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLFKSHPETL EKFD RFKHLK TEAE MKASED
61 LKKHGVTVLT ALGAILKKGK HHEAELKPLA QSHATKHKIP IKYLEFISEA IIVLHRSRHP
121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG
```

Center position: 4 Window size: 7 Threshold: 0.903 Recalculate



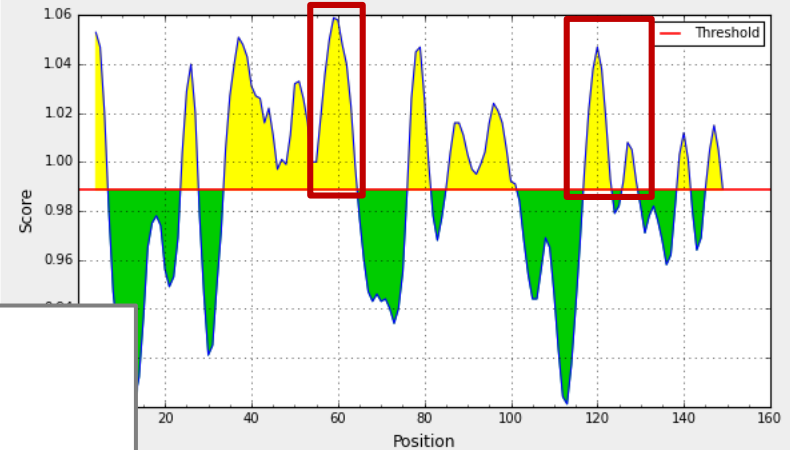
Average: 0.903 Minimum: 0.626 Max

Karplus & Schulz Flexibility Prediction Results

Input Sequences

```
1 VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLFKSHPETL EKFD RFKHLK TEAE MKASED
61 LKKHGVTVLT ALGAILKKGK HHEAELKPLA QSHATKHKIP IKYLEFISEA IIVLHRSRHP
121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG
```

Center position: 4 Window size: 7 Threshold: 0.988 Recalculate



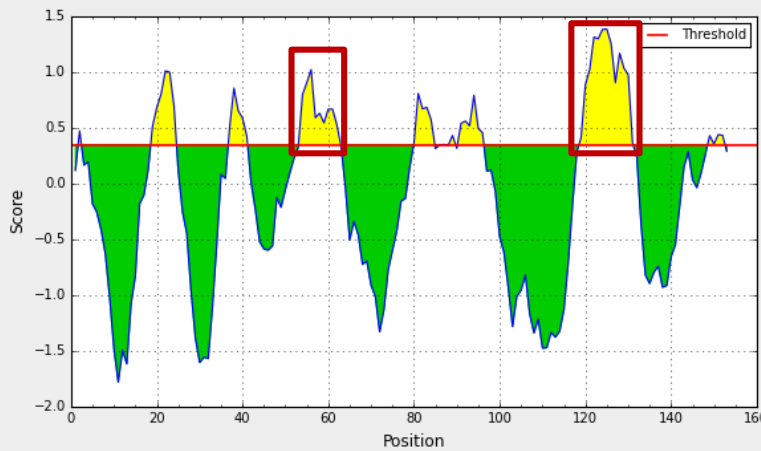
Minimum: 0.901 Maximum: 1.059

Bepipred Linear Epitope Prediction Results

Input Sequences

```
1 VLSEGEWQLV LHVWAKVEAD VAGHGQDILI RLFKSHPETL EKFD RFKHLK TEAE MKASED
61 LKKHGVTVLT ALGAILKKGK HHEAELKPLA QSHATKHKIP IKYLEFISEA IIVLHRSRHP
121 GNFGADAGGA MNKALELFRK DIAAKYKELG YQG
```

Center position: 4 Threshold: 0.350 Recalculate



Average: -0.105 Minimum: -0.028 Maximum: 1.390

It is better to use **consensus** of different methods rather than relying on a single method

Sequence-based epitope prediction

- Evaluation of amino acid scales: **no method gave AUC above 0.60**

J Mol Recognit. 2007 Mar-Apr;20(2):75-82.

Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools.

Greenbaum JA¹, Andersen PH, Blythe M, Bui HH, Cachau RE, Crowe J, Davies M, Kolaskar AS, Lund O, Morrison S, Mumey B, Ofran Y, Pellequer JL, Pinilla C, Ponomarenko JV, Raghava GP, van Regenmortel MH, Roggen EL, Sette A, Schlessinger A, Sollner J, Zand M, Peters B.

PMID: 17205610 DOI: [10.1002/jmr.815](https://doi.org/10.1002/jmr.815)

3D Structures of Ab-Ag complexes

Methods for 3D structure determination:

- **X-ray crystallography** (provides the most accurate identification of epitopes)
- **Nuclear magnetic resonance (NMR)**
- **Electron microscopy (EM)**

Where to get 3D Ab-Ag complexes??

- IEDB 3D export (1790 3D BCR assays)
→ SCEptRe (for annotation and redundancy removal)

Where to get 3D coordinates of proteins?

- Biomolecular 3D structural data is deposited into **PDB** (Protein Data Bank)

3D Structure-based epitope prediction

Discontinuous epitope prediction

- Structure-based epitope prediction using:
 - Geometrical properties combined with amino acid scales (**DiscoTope**, **ElliPro**, CEP)
 - Geometrical properties and amino acid scales used as input to machine learning approaches (EPSVR)
 - Protein-protein docking algorithms
- Sequence-based epitope prediction using
 - Machine learning approaches (CBTope)

B cell prediction tools on IEDB

<http://tools.iedb.org/main/bcell/>

IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

B Cell Epitope Prediction Tools

B Cell Epitope Prediction

[Prediction of linear epitopes from protein sequence](#)

A collection of methods to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs.

[Discotope - Prediction of epitopes from protein structure](#)

This method incorporates solvent-accessible surface area calculations, as well as contact distances into its prediction of B cell epitope potential along the length of a protein sequence.

[ElliPro - Epitope prediction based upon structural protrusion](#)

This method predicts epitopes based upon solvent-accessibility and flexibility.

[Methods for modeling and docking of antibody and protein 3D structures](#)

This page provides information on available methods for modeling and docking of antibody and protein 3D structures.

Structure Tools



[LYRA \(Lymphocyte Receptor Automated Modelling\)*](#)

The LYRA server predicts structures for either T-Cell Receptors (TCR) or B-Cell Receptors (BCR) using homology modelling. Framework templates are selected based on BLOSUM score, and complementary determining regions (CDR) are then selected if needed based on a canonical structure model and grafted onto the framework templates.



[SCEptRe: Structural Complexes of Epitope Receptor](#)

SCEptRe provides weekly updated, non-redundant, user customized benchmark datasets with information on the immune receptor features for receptor-specific epitope predictions. This tool extracts weekly updated 3D complexes of antibody-antigen, TCR-pMHC and MHC-ligand from the Immune Epitope Database (IEDB) and clusters them based on antigens, receptors and epitopes to generate benchmark datasets. Users can customize structural quality and clustering parameters (e.g. resolution, R free factors, antigen or epitope sequence identity) to generate these datasets based on their need.



* : Tools under AR Labs which are experimental and are not quite ready for production yet. They are intended for further research, updates and testing.



DiscoTope

- Trained on 75 X-ray structures of antibody-protein complexes
- DiscoTope 2 took into account multiple epitopes in an antigen
- Assigns each residue a score value calculated as a linear combination of normalized values
 - Parker's hydrophilicity scale
 - Amino acid occurrence
 - Number of contacts within 10Å
 - Area of relative solvent accessibility
- AUC 0.71 for DiscoTope 1 and 0.73 for DiscoTope 2

[Protein Sci.](#) 2006 Nov;15(11):2558-67. Epub 2006 Sep 25.

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

[Haste Andersen P¹](#), [Nielsen M](#), [Lund O](#).

PMID: 17001032 PMCID: [PMC2242418](#) DOI: [10.1110/ps.062405906](#)

DiscoTope 1

DiscoTope 2

[PLoS Comput Biol.](#) 2012;8(12):e1002829. doi: [10.1371/journal.pcbi.1002829](#). Epub 2012 Dec 27.

Reliable B cell epitope predictions: impacts of method development and improved benchmarking.

[Kringelum JV¹](#), [Lundegaard C](#), [Lund O](#), [Nielsen M](#).

PMID: 23300419 PMCID: [PMC3531324](#) DOI: [10.1371/journal.pcbi.1002829](#)

IEDB Analysis Resource

[Home](#) [Help](#) [Example](#) [Reference](#) [Download](#) [Contact](#)

DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID
Or upload a PDB file

(example: 1z40)

No file selected.

Step 2: Please enter PDB Chain ID

Step 3: Select version

1.1

Identify structure

Input PDB id
or upload file

Specific to the protein chain of interest

Search in PDB to identify inputs

<http://www.rcsb.org/>

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

RCSB PDB PROTEIN DATA BANK 156954 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or liga **Go**

Advanced Search | Browse by Annotations

PDB-101 WORLDWIDE PDB PROTEIN DATA BANK EMDatabank Unified Data Resource for 2018 NUCLEIC ACID DATABASE Worldwide Protein Data Bank Foundation

Search Options

- Welcome
- Deposit
- Search**
- Visualize
- Analyze
- Download
- Learn

Drill Down by Categories

- Advanced Search
- Sequences**
- Ligands
- Drugs & Drug Targets
- Unreleased & New Entries
- Browse by Annotation
- PDB Statistics

Search by Sequences

Choose **Option A** or **B** to search for protein and nucleic acid sequences. [Read Tutorial](#) **Advanced Sequence Searching**

NOTE Parameters: BLAST method, E-value cutoff: 10.0, Mask Low Complexity: On.

Option A: Use PDB Sequence

Select Associated Chain ...

or Option B: Paste Sequence

Run Sequence Search

Search in PDB to identify inputs

<http://www.rcsb.org/>

RCSB PDB Deposit Search Visualize Analyze Download Learn More MyPDB

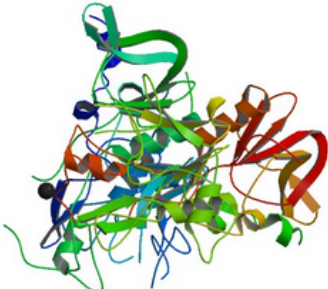
156954 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or liga Go

Advanced Search | Browse by Annotations

Structure Summary 3D View Annotations Sequence Sequence Similarity Structure Similarity Experiment

Biological Assembly 1



1Z40
AMA1 from Plasmodium falciparum
DOI: 10.2210/pdb1Z40/pdb
Classification: **UNKNOWN FUNCTION**
Organism(s): [Plasmodium falciparum \(isolate 3D7\)](#)
Expression System: [Escherichia coli BL21\(DE3\)](#)
Deposited: 2005-03-14 Released: 2005-08-16
Deposition Author(s): [Bai, T.](#), [Becker, M.](#), [Gupta, A.](#), [Batchelor, A.H.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.901 Å
R-Value Free: 0.236
R-Value Work: 0.192

wwPDB Validation

3D View: [Structure](#) | [Electron Density](#) | [Ligand Interaction](#)

Standalone Viewers
[Protein Workshop](#) | [Ligand Explorer](#)

Global Symmetry: Asymmetric - C1

Display Files Download Files

- FASTA Sequence
- PDB Format
- PDB Format (gz)
- PDBx/mmCIF Format
- PDBx/mmCIF Format (gz)
- PDBML/XML Format (gz)
- Biological Assembly 1
- Structure Factors (CIF)
- Structure Factors (CIF - gz)
- 2fo-fc Map (DSN6)
- fo-fc Map (DSN6)
- Map Coefficients (MTZ format)

Search in PDB to identify inputs

<http://www.rcsb.org/>

The screenshot shows the RCSB PDB website interface. At the top, there are navigation menus: 'RCSB PDB', 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', 'Learn', and 'More'. A 'MyPDB' button is on the right. Below the navigation is a 'Macromolecules' section with options to 'Find similar proteins by: Sequence | Structure'. A table lists search results for 'Entity ID: 1'. The table has columns for 'Molecule', 'Chains', 'Sequence Length', 'Organism', and 'Details'. The first row shows 'apical membrane antigen 1 precursor' with 'A, E' chains, a sequence length of 336, and the organism 'Plasmodium falciparum (isolate 3D7)'. Below the table, there is a search bar for 'Q7KQK5 (Plasmodium falciparum (isolate 3D7))' and a 'Go to UniProtKB: Q7KQK5' button. At the bottom, there is a 'Protein Feature View' section with a toggle switch and a diagram of the protein structure with various features highlighted.

Molecule	Chains	Sequence Length	Organism	Details
apical membrane antigen 1 precursor	A, E	336	Plasmodium falciparum (isolate 3D7)	Mutation(s): 0

Find proteins for [Q7KQK5](#) (*Plasmodium falciparum (isolate 3D7)*)

Go to UniProtKB: [Q7KQK5](#)

Protein Feature View Full Protein Feature View for [Q7KQK5](#)

What if the 3D structure of a protein of your interest is not available in PDB?

> Homology or comparative modeling methods, servers and databases

DiscoTope -example

<http://tools.iedb.org/discotope/>

The screenshot shows the IEDB Analysis Resource interface for DiscoTope. The main heading is "IEDB Analysis Resource" and "DiscoTope: Structure-based Antibody Prediction". The interface has a navigation bar with links: Home, Help, Example, Reference, Download, and Contact. The main form has three steps: Step 1: "Please enter the 4-letter PDB ID Or upload a PDB file" with input "1z40" (example: 1z40) and a "Browse..." button; Step 2: "Please enter PDB Chain ID" with input "A"; Step 3: "Select version" with a dropdown menu showing "1.1" and "2.0". A red box highlights the dropdown menu, and a red arrow points from it to the "DiscoTope-1.1" table. A red arrow also points from the "Submit" button to the "DiscoTope-2.0" table. Two blue ovals highlight the row with Score ">-7.7" in the "DiscoTope-1.1" table and the row with Score ">-3.7" in the "DiscoTope-2.0" table, both labeled "Default".

IEDB Analysis Resource

Home Help Example Reference Download Contact

DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID
Or upload a PDB file

1z40 (example: 1z40)

Browse... No file selected.

Step 2: Please enter PDB Chain ID

A

Step 3: Select version

1.1
2.0

Submit Reset

DiscoTope-1.1

Score	Sensitivity	Specificity
>-3.1	0.16	0.95
>-4.7	0.24	0.90
>-6.0	0.32	0.85
>-6.9	0.40	0.80
>-7.7	0.47	0.75

Default

DiscoTope-2.0

Score	Sensitivity	Specificity
>1.9	0.17	0.95
>0.5	0.23	0.90
>-1.0	0.30	0.85
>-2.5	0.39	0.80
>-3.7	0.47	0.75

Default

DiscoTope -example

<http://tools.iedb.org/discotope/>

IEDB Analysis Resource

Home Help Example Reference Download Contact

DiscoTope: Structure based antibody prediction.
DiscoTope 2.0 prediction for structure: 1z40 & Chain ID: A

Threshold:

DiscoTope Prediction

— Threshold = -3.7 ■ Positive prediction ■ Negative prediction

Score

Position

Downloads as csv

Adjustable threshold

DiscoTope -example

<http://tools.iedb.org/discotope/>

DiscoTope - Result

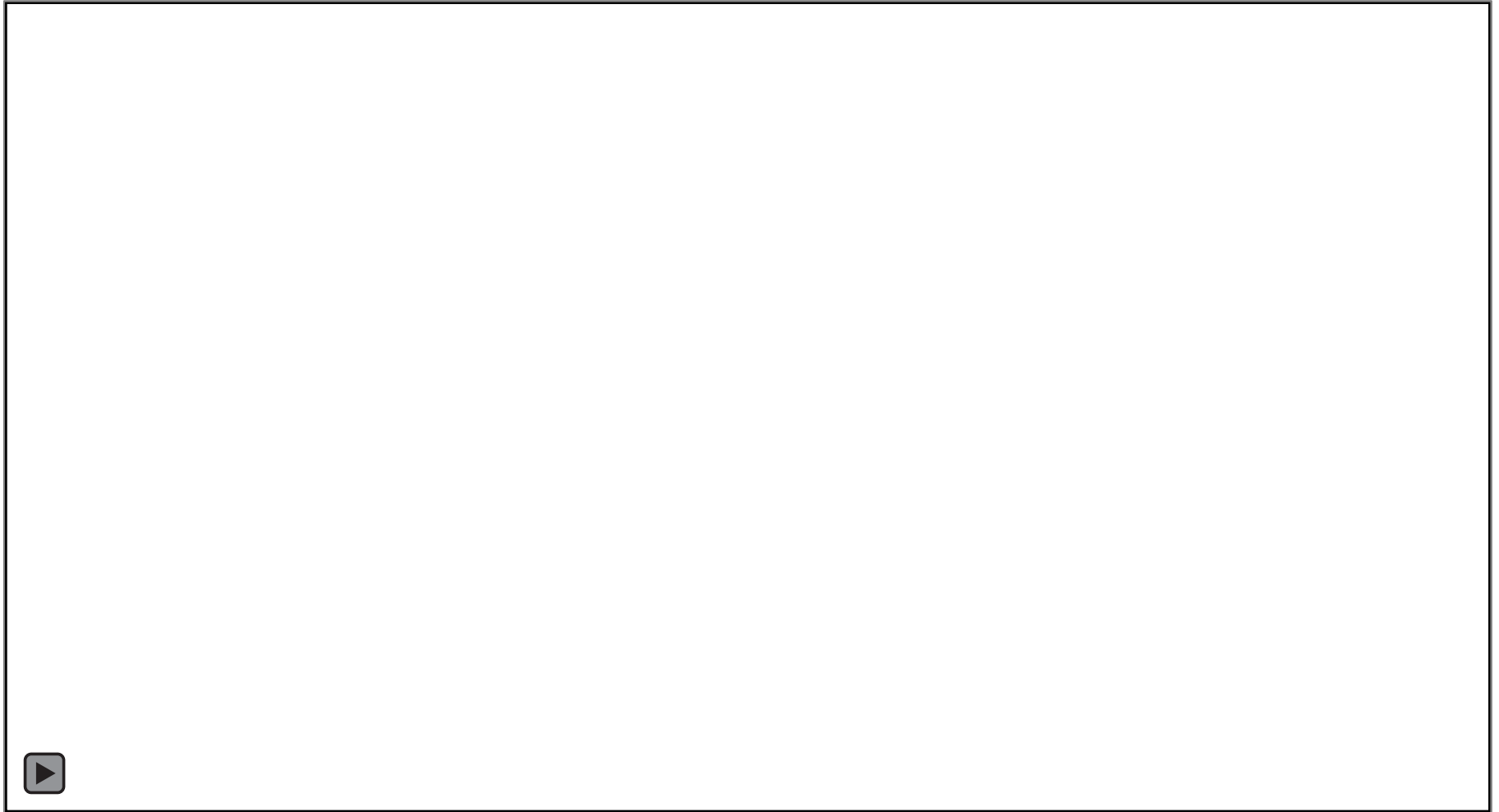
DiscoTope 2.0 Prediction for structure: 1z40 & Chain ID: A
The positive predictions are displayed in green.

Chart View 3D View Save Prediction

Chain ID	Residue ID	Residue Name	Contact Number	Propensity Score	Discotope Score
A	108	ASN	17	-4.287	-5.749
A	109	PRO	7	-3.77	-4.141
A	110	TRP	23	-5.522	-7.532
A	111	THR	4	-5.226	-5.085
A	112	GLU	2	-5.1	-4.744
A	113	TYR	20	-5.97	-7.584
A	114	MET	20	-9.295	-10.526
A	115	ALA	10	-7.532	-7.816
A	116	LYS	23	-11.888	-13.166
A	117	TYR	33	-11.038	-13.564
A	118	ASP	12	-9.15	-9.478
A	119	ILE	35	-10.419	-13.246
A	120	GLU	10	-6.657	-7.042
A	121	GLU	13	-5.019	-5.937
A	122	VAL	30	-8.676	-11.129
A	123	HIS	43	-10.161	-13.938
A	124	GLY	17	-4.876	-6.27
A	125	SER	25	-3.112	-5.629
A	126	GLY	33	-8.05	-10.92
A	127	ILE	39	-12.601	-15.637
A	128	ARG	27	-13.272	-14.85
A	129	VAL	25	-9.506	-11.288
A	130	ASP	39	-8.027	-11.589
A	131	LEU	29	-5.732	-8.408
A	132	GLY	30	-3.241	-6.318
A	133	GLU	20	-0.565	-2.8
A	134	ASP	19	3.255	0.695
A	135	ALA	28	1.707	-1.709

DiscoTope -example

<http://tools.iedb.org/discotope/>



B cell prediction tools on IEDB

<http://tools.iedb.org/main/bcell/>

IEDB Analysis Resource

[Overview](#) [T Cell Tools](#) [B Cell Tools](#) [Analysis Tools](#) [Tools-API](#) [Usage](#) [Download](#) [Datasets](#) [Contribute Tools](#) [References](#)

B Cell Epitope Prediction Tools

B Cell Epitope Prediction

[Prediction of linear epitopes from protein sequence](#)

A collection of methods to predict linear B cell epitopes based on sequence characteristics of the antigen using amino acid scales and HMMs.

[Discotope - Prediction of epitopes from protein structure](#)

This method incorporates solvent-accessible surface area calculations, as well as contact distances into its prediction of B cell epitope potential along the length of a protein sequence.

[ElliPro - Epitope prediction based upon structural protrusion](#)

This method predicts epitopes based upon solvent-accessibility and flexibility.

[Methods for modeling and docking of antibody and protein 3D structures](#)

This page provides information on available methods for modeling and docking of antibody and protein 3D structures.

Structure Tools



[LYRA \(Lymphocyte Receptor Automated Modelling\)*](#)

The LYRA server predicts structures for either T-Cell Receptors (TCR) or B-Cell Receptors (BCR) using homology modelling. Framework templates are selected based on BLOSUM score, and complementary determining regions (CDR) are then selected if needed based on a canonical structure model and grafted onto the framework templates.



[SCEptRe: Structural Complexes of Epitope Receptor](#)

SCEptRe provides weekly updated, non-redundant, user customized benchmark datasets with information on the immune receptor features for receptor-specific epitope predictions. This tool extracts weekly updated 3D complexes of antibody-antigen, TCR-pMHC and MHC-ligand from the Immune Epitope Database (IEDB) and clusters them based on antigens, receptors and epitopes to generate benchmark datasets. Users can customize structural quality and clustering parameters (e.g. resolution, R free factors, antigen or epitope sequence identity) to generate these datasets based on their need.



** Tools under AR Labs which are experimental and are not quite ready for production yet. They are intended for further research, updates and testing.*



ElliPro

- Predicts linear and discontinuous antibody epitopes based on the geometrical properties of protein structure
- Uses Thornton's Method
- Implements three algorithms:
 - Approximation of the protein shape as an ellipsoid
 - Calculation of the residue protrusion index (PI)
 - Clustering of neighboring residues based on PI values

[EMBO J.](#) 1986 Feb;5(2):409-13.

Location of 'continuous' antigenic determinants in the protruding regions of proteins.

[Thornton JM](#), [Edwards MS](#), [Taylor WR](#), [Barlow DJ](#).

PMID: 2423325 PMCID: [PMC1166746](#)

[BMC Bioinformatics.](#) 2008 Dec 2;9:514. doi: 10.1186/1471-2105-9-514.

ElliPro: a new structure-based tool for the prediction of antibody epitopes.

[Ponomarenko J](#)¹, [Bui HH](#), [Li W](#), [Fusseder N](#), [Bourne PE](#), [Sette A](#), [Peters B](#).

PMID: 19055730 PMCID: [PMC2607291](#) DOI: [10.1186/1471-2105-9-514](#)

IEADB Analysis Resource

Home Help Example Reference Download Contact

ElliPro: Antibody Epitope Prediction

Specify Sequence(s)

Enter PDB ID(s) or upload PDB file No file selected.

Select Epitope Prediction Parameters

Minimum score: (Default is 0.5)

Maximum distance (Angstrom): (Default is 6)

Identify structure

Input PDB id or upload file

Select min. PI value

- Averaged over epitope residues
- Higher scores predict fewer epitopes

Specify max distance for predicting (grouping) discontinuous epitopes

Longer distances predict discontinuous epitopes spanning larger regions

EliPro -example

<http://tools.iedb.org/ellipro/>

Home Help Example Reference Download Contact

EliPro: Antibody Epitope Prediction

Specify Sequence(s)

Enter PDB ID(s) or upload PDB file: No file selected.

Select Epitope Prediction Parameters

Minimum score: (Default is 0.5)

Maximum distance (Angstrom): (Default is 6)

Home Help Example Reference Download Contact

EliPro: Antibody Epitope Prediction

Select PDB chain(s) for calculation

Input pdb id: **5LYM**

Number of chain(s) found: **2**

Select	No.	Chain	Chain type	Number of residues
<input checked="" type="checkbox"/>	1	A	Amino acid	129
<input type="checkbox"/>	2	B	Amino acid	129

Select chain(s) of interest

EliPro -example

<http://tools.iedb.org/ellipro/>

IEDB Analysis Resource

[Home](#) [Help](#) [Example](#) [Reference](#) [Download](#) [Contact](#)

Input Sequences: 5LYM

Chain: A
1 KVFGRCELAA AMKRHGLDNY RGYSLGNWVC AAKFESNFNT QAINRNTDGS IDYGILQINS
61 RWWCNDGRTP GSRNLCNIPC SALLSSDITA SVNCARKTIVS DGNMGMAWVA WRNRCKGTDV
121 QAWIRGCRLL

Predicted Linear Epitope(s):

No.	Chain	Start	End	Peptide	Number of residues	Score	3D structure
1	A	45	50	RNTDGS	6	0.78	View
2	A	112	129	RNRCKGTDVQAWIRGCRLL	18	0.771	View
3	A	100	103	SDGN	4	0.76	View
4	A	64	81	CNDGRTPGSRNLCNIPCS	18	0.666	View
5	A	1	7	KVFGRCE	7	0.597	View
6	A	13	23	KRHGLDNYRGY	11	0.574	View
7	A	85	88	SSDI	4	0.504	View

View 3D structure

Predicted Discontinuous Epitope(s):

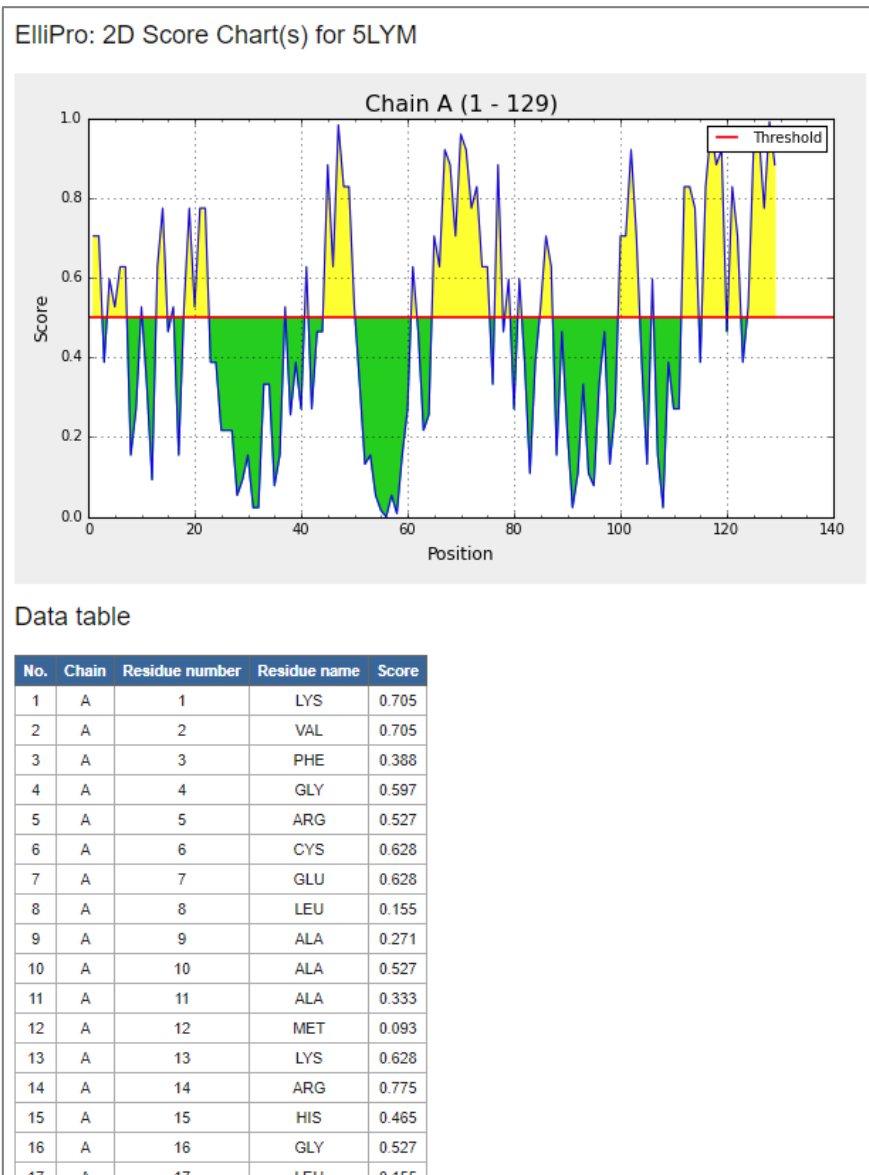
No.	Residues	Number of residues	Score	3D structure
1	A:S100, A:D101, A:G102, A:N103, A:N106	5	0.727	View
2	A:K1, A:V2, A:F3, A:G4, A:R5, A:C6, A:E7, A:F38, A:N39, A:T40, A:Q41, A:A42, A:S85, A:S86, A:D87, A:I88, A:R112, A:N113, A:R114, A:C115, A:K116, A:G117, A:T118, A:D119, A:Q121, A:A122, A:I124, A:R125, A:G126, A:C127, A:R128, A:L129	32	0.657	View
3	A:R45, A:N46, A:T47, A:D48, A:G49, A:S50, A:N59, A:S60, A:R61, A:W62, A:W63, A:C64, A:N65, A:D66, A:G67, A:R68, A:T69, A:P70, A:G71, A:S72, A:R73, A:N74, A:L75, A:C76, A:N77, A:I78, A:P79, A:S81	28	0.648	View
4	A:A10, A:K13, A:R14, A:G16, A:L17, A:D18, A:N19, A:Y20, A:R21, A:G22, A:Y23, A:S24	12	0.564	View

[Click here to view residue scores](#)

[Download pdb file](#)

EliPro -example

<http://tools.iedb.org/ellipro/>



Summary – B cell epitope predictions

- Linear and discontinuous (conformational) epitopes can be overlapping and depending on method of discovery
- Traditional B cell epitope prediction methods largely predict surface accessibility
- If a 3D structure of the antigen is available (or a reliable model thereof), predictions can be further improved.

Practice Exercise

- Use BepiPred and DiscoTope to predict B cell epitopes of dengue 2 virus envelope glycoprotein
- Download crystal structure and sequence of dengue 2 virus envelope glycoprotein from PDB (PDB ID: [4UTC](#))

Practice Exercise

Bepipred Linear Epitope Prediction Results

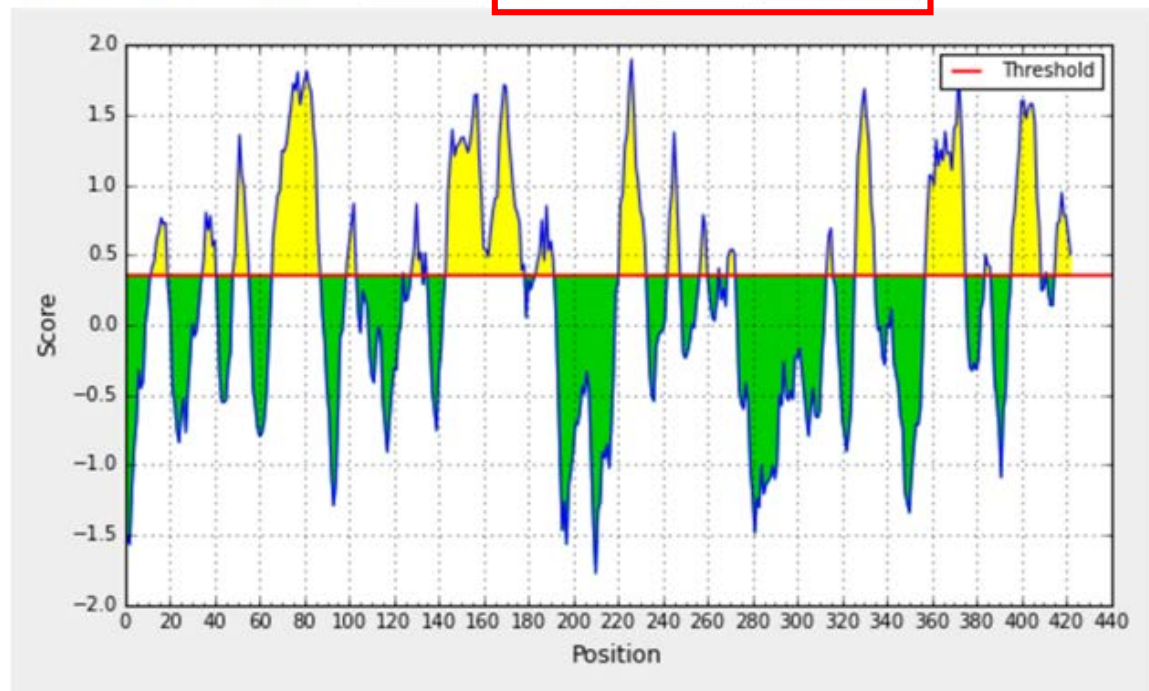
Input Sequences

```
1 MRCIGISNRD FVEGVSGGSW VDIVLEHGSC VTTMAKNKPT LDFELIKTEA KQPATLRKYC
61 IEAKLTNTTT ESRCPQTGEP SLNEEQDK RFICKHSMVD RGWNGCGGLF GKGGIVTCAK
121 FTCKKNMEGK IVQPENLEYT IVITPHSGEE HAVGNDTGKH GK EKITP QSSTTEAELT
181 YGTVTMECS PRTGLDFNEM VLLQMEDKAW LVHRQWFLDL PLPWLPGADT QGSNWIQET
241 LVTF KNPH AKKQDVVVLG SQEGAMHTAL TGATEIQMSS GNLLFTGHLK CRLRMDKLQL
301 KGMSYSMCTG KFKIVKEIAE TQHGTI VI RVQYEGDGSP CKIPFEITDL EKRHVLGRLI
361 TVNPIVTEKD SPVNIEAAPP FGDSYIIVGV EPGQLKLNWL RPLESRGP FEGKPIP NPL
421 LGLDSTRTGH HH
```

Center position: 4 Window size: 7

Threshold: 0.35

Recalculate



Average: 0.188 Minimum: -1.776 Maximum: 1.900

Too many epitope candidates?

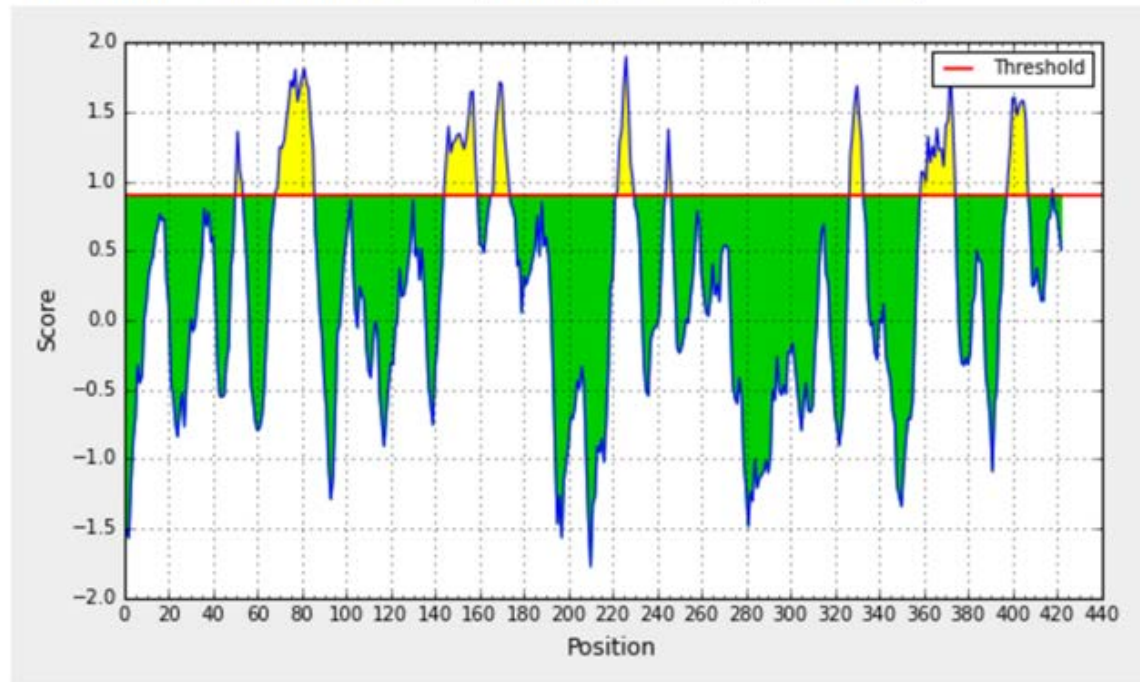
Practice Exercise

Bepipred Linear Epitope Prediction Results

Input Sequences

```
1 MRCIGISNRD FVEGVSGGSW VDIVLEHGSC VTTMAKNKPT LDFELIKTEA KQPATLRKYC
61 IEAKLTNTTT ESRCPQTQEP SLNEEQDK RFICKHSMVD RGWNGCGLF GKGGIVTCAK
121 FTCKKNMEGK IVQPENLEYT IVITPHSGEE HAVGNDTGKH GK EIKITP QSSTTEAELT
181 GYGTVTMECS PRTGLDFNEM VLLQMEDKAW LVHRQWFLDL PLPWLPGADT QGSNWIQKET
241 LVTF KNPH AKKQDVVVLG SQEGAMHTAL TGATEIQMSS GNLLFTGHLK CRLRMDKLQL
301 KGMSYSMCTG KFKIVKEIAE TQHGTI VI RVQYEGDGP CKIPFEITDL EKRHLVGLRI
361 TVNPIVTEKD SPVNIEAEPF FGDSYIIVGV EPGQLKLNWL RPLESRGP FEGKPIP NPL
421 LGLDSTRTGH HH
```

Center position: 4 Window size: Threshold:



Average: 0.188 Minimum: -1.776 Maximum: 1.900

Score **threshold of 0.9**
corresponds to **90%**
specificity

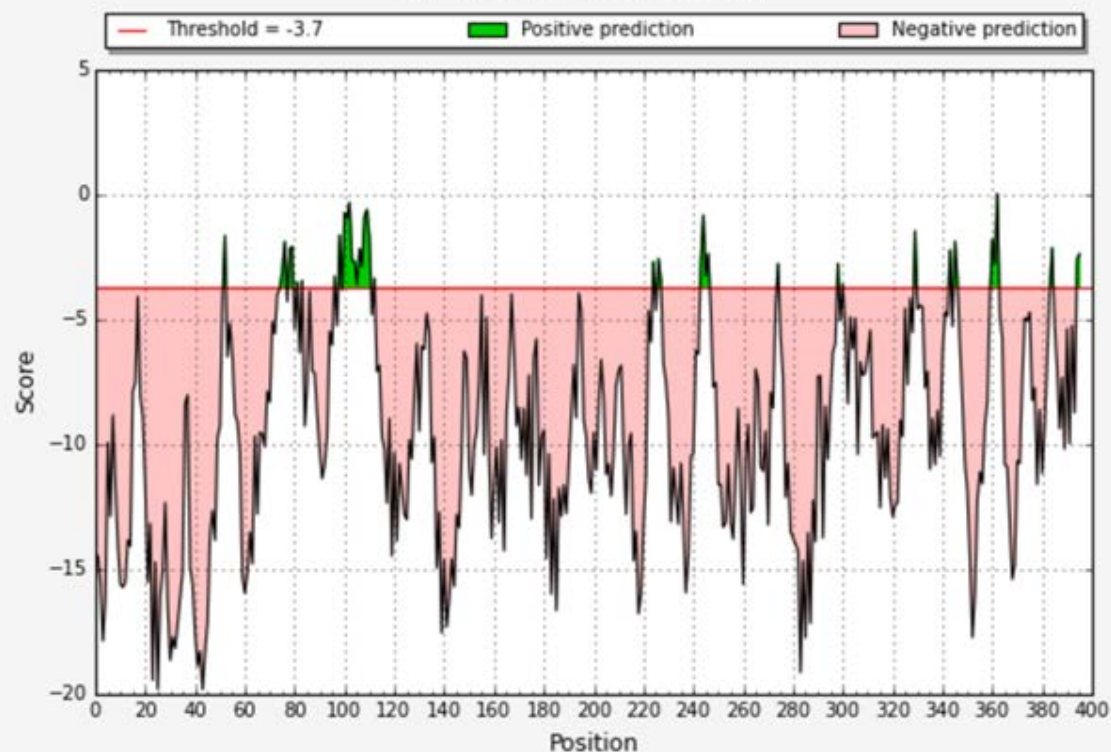
Practice Exercise

DiscoTope: Structure based antibody prediction.

DiscoTope 2.0 prediction for structure: 4utc & Chain ID: A

Threshold:

DiscoTope Prediction



Did you notice the length difference between **BepiPred** (length **422**) and **DiscoTope** (length **391**) outputs?

Practice Exercise

Protein alignment	Nucleotide alignment	Web services	Help & Documentation
#	#		
#-----			
4UTC_A_seqres	1 MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA		50
4UTC_A_atomse	1 MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA		50
4UTC_A_seqres	51 KQPATLRKYCIEAKLTNTTTERCPTQGEPSSLNEEQDKRFICKHSMVDRG		100
4UTC_A_atomse	51 KQPATLRKYCIEAKLTNTTTERCPTQGEPSSLNEEQDKRFICKHSMVDRG		100
4UTC_A_seqres	101 WGNCGCLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA		150
4UTC_A_atomse	101 WGNCGCLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA		150
4UTC_A_seqres	151 VGNDTGKHGKEIKITPQSSSTAEALTGYGTVTMECSPRTGLDFNEMVLLQ		200
4UTC_A_atomse	151 VGNDTGKHGKEIKITPQSSSTAEALTGYGTVTMECSPRT--DFNEMVLLQ		198
4UTC_A_seqres	201 MEDKAWLVHRQWFLDLPWPWPGADTQGSNWIQKETLVTFKNPHAKKQDV		250
4UTC_A_atomse	199 MEDKAWLVHRQWFLDLPWPWPGADTQGSNWIQKETLVTFKNPHAKKQDV		248
4UTC_A_seqres	251 VVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKRLRMDKLQKGMSSYS		300
4UTC_A_atomse	249 VVLGSQEGAMHTALTGATEIQMSSGNLLF--HLKRLRMDKLQKGMSSYS		296
4UTC_A_seqres	301 MCTGKFKIVKEIAETQHGTVIRVQYEGDGSCKIPFEITDLEKRHLVGR		350
4UTC_A_atomse	297 MCTGKFKIVKEIAETQHGTVIRVQYEGDGSCKIPFEITDLEKRHLVGR		346
4UTC_A_seqres	351 LITVNPIVTEKDSPVNIEAEPFGDSYIIVGVEPGQLKNWLRPLESRGP		400
4UTC_A_atomse	347 LITVNPIVTEKDSPVNIEAEPFGDSYIIVGVEPGQLKNWLRPL-----		391
4UTC_A_seqres	401 FEGKPIPPLLGLDSTRTGHHH	422	
4UTC_A_atomse	392 -----	391	
#-----			
#-----			

- There might be missing residues in PDB coordinate file compared to the sequence file provided by PDB
- These missing residues are not resolved properly in the structure
 - e.g. flexible loops

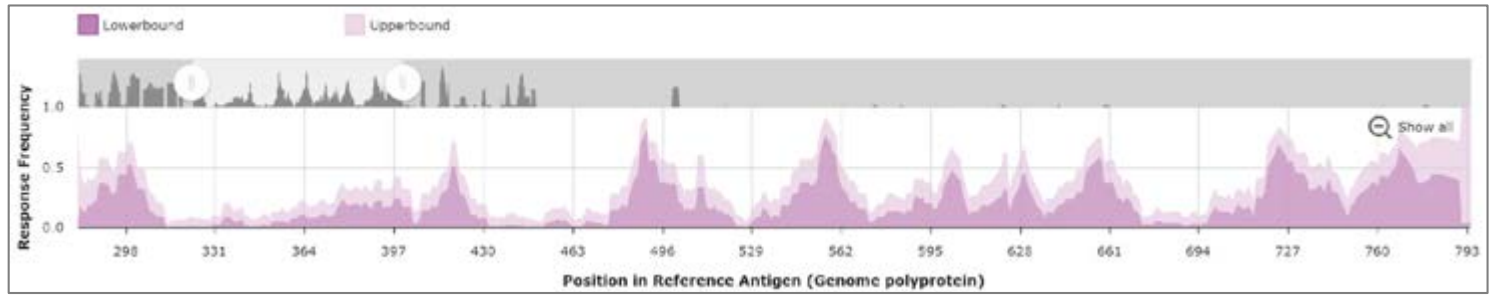
Practice Exercise

4UTC_A_BepiPred	1	MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA	50
4UTC_A_DiscoTope	1	MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA	50
4UTC_A_BepiPred	51	KQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG	100
4UTC_A_DiscoTope	51	KQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG	100
4UTC_A_BepiPred	101	WNGCGLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA	150
4UTC_A_DiscoTope	101	WNGCGLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA	150
4UTC_A_BepiPred	151	VGNDTGKHGKEIKITPQSSSTEAEELTGYGVTMECSPRTGLDFNEMVLLQ	200
4UTC_A_DiscoTope	151	VGNDTGKHGKEIKITPQSSSTEAEELTGYGVTMECSPRT--DFNEMVLLQ	198
4UTC_A_BepiPred	201	MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV	250
4UTC_A_DiscoTope	199	MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV	248
4UTC_A_BepiPred	251	VVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCLRMDKLQKGM SYS	300
4UTC_A_DiscoTope	249	VVLGSQEGAMHTALTGATEIQMSSGNLLF--HLKCLRMDKLQKGM SYS	296
4UTC_A_BepiPred	301	MCTGKFKIVKEIAETQHGTIVIRVQYEGDGS PCKIPFEITDLEKRHVLGR	350
4UTC_A_DiscoTope	297	MCTGKFKIVKEIAETQHGTIVIRVQYEGDGS PCKIPFEITDLEKRHVLGR	346
4UTC_A_BepiPred	351	LITVNPIVTEKISFVNIEAEPFPGDSYIIVGVEPGQLKLNWLRPLESRGP	400
4UTC_A_DiscoTope	347	LITVNPIVTEKISFVNIEAEPFPGDSYIIVGVEPGQLKLNWLRPL-----	391
4UTC_A_BepiPred	401	FEGKPIP NPLLGLDSTRIGHHH	422

- Predicted
- Correctly predicted

Epitope residues from 3D B cell [assay 3319631](#) (PDB ID: [2R69](#)) were mapped on Dengue envelope glycoprotein

Practice Exercise

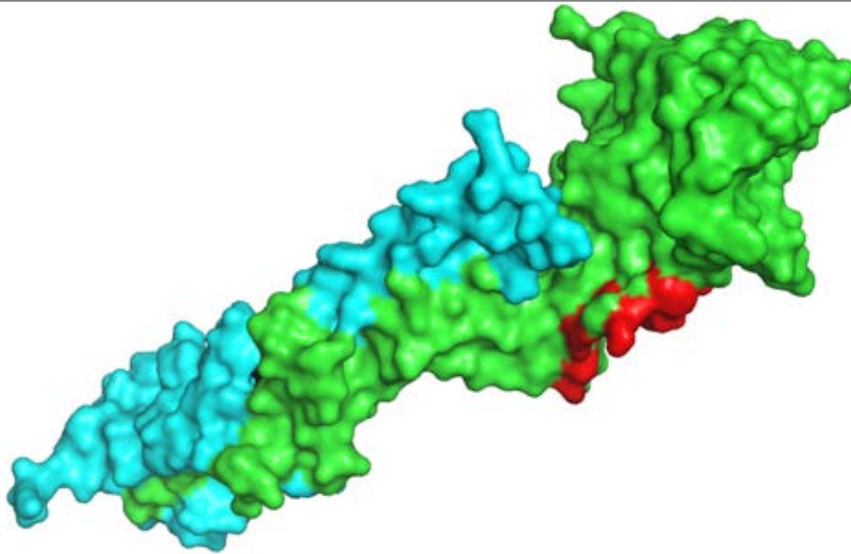


4UTC_A_BepiPred	1	MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVITMAKNKPTLDFELIKTE	50
4UTC_A_DiscoTope	1	MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVITMAKNKPTLDFELIKTEA	50
4UTC_A_BepiPred	51	KQPAILRKYCIEAKLNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG	100
4UTC_A_DiscoTope	51	KQPAILRKYCIEAKLNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG	100
4UTC_A_BepiPred	101	WNGGCGLFGKGGIVICAKFTCKKQMEGKIVQPENLEYTIVITPHSGEEHA	150
4UTC_A_DiscoTope	101	WNGGCGLFGKGGIVICAKFTCKKQMEGKIVQPENLEYTIVITPHSGEEHA	150
4UTC_A_BepiPred	151	VGNDIGKHGKEIKIIPQSSITEAELTGYGTVIMECSPRIGLDFNEMVLLQ	200
4UTC_A_DiscoTope	151	VGNDIGKHGKEIKIIPQSSITEAELTGYGTVIMECSPRIT--DFNEMVLLQ	198
4UTC_A_BepiPred	201	MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV	250
4UTC_A_DiscoTope	199	MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV	248
4UTC_A_BepiPred	251	VVLGSQEGAMHTALIGATEIQMSSGNLLFTGHLKCRLRMDKLQKGMSSYS	300
4UTC_A_DiscoTope	249	VVLGSQEGAMHTALIGATEIQMSSGNLLF--HLKCRLRMDKLQKGMSSYS	296
4UTC_A_BepiPred	301	MCTGKFKIVKEIAETQHGTVIRVQYEGDGSFCKIPFEITDLEKRHVLR	350
4UTC_A_DiscoTope	297	MCTGKFKIVKEIAETQHGTVIRVQYEGDGSFCKIPFEITDLEKRHVLR	346
4UTC_A_BepiPred	351	LITVNPVITKQDSEFNIEAEPFGDSYIIVGVEPGQLKLNWLRPLESRGP	400
4UTC_A_DiscoTope	347	LITVNPVITKQDSEFNIEAEPFGDSYIIVGVEPGQLKLNWLRPL-----	391
4UTC_A_BepiPred	401	FEGKPIPNPLGLDSTRIGHH	422
4UTC_A_DiscoTope	392	-----	391

Epitope residues from the IEDB in Dengue envelope glycoprotein

- Predicted
- Correctly predicted

3D Structure-based epitope prediction



<http://www.ofranlab.org/PEASE>

HOME LAB MEMBERS SERVICES PUBLICATIONS CONTACT

PEASE: Predicting Epitopes using Antibody Sequence

This automated tool predicts the epitope for a given antigen structure and an antibody sequence.

Upload the structure of the antigen in a PDB format, or choose an existing PDB file:

Antigen Structure: No file chosen

Antigen Structure ID:

If the antigen structure is a computational model, please upload in addition the antigen sequence, in order to identify residues with no coordinates:
Note: The antigen sequence should be in a fasta format, and the title line should be the chain ID (e.g. ">C" for antigen chain C).

Antigen sequence: No file chosen

- Theoretically, the whole exposed surface of an antigen can be targeted by different antibodies
- Antibody sequence based B cell epitope prediction method called PEASE was developed
 - Users must provide antigen structure and antibody sequence

Bioinformatics. 2015 Apr 15;31(8):1313-5. doi: 10.1093/bioinformatics/btu790. Epub 2014 Nov 27.

PEASE: predicting B-cell epitopes utilizing antibody sequence.

Sela-Culang I¹, Ashkenazi S¹, Peters B¹, Ofran Y¹.

PMID: 25432167 DOI: 10.1093/bioinformatics/btu790

Benchmark on 42 X-ray structures of Ab-protein complexes (Ponomarenko & Bourne, 2008)

Average AUC values

- **0.73 ElliPro** (Ponomarenko et al., 2008)
- **0.65 Epitopia** (Rubinstein et al., 2008)
- **0.63 PEPITO** (Sweredoski & Baldi, 2008)
- **0.60 DiscoTope 1** (Andersen et al, 2006)

- **0.59 DOT** (1st model, bound **Ab-protein docking**)
- **0.58 PatchDock** (1st model, bound **Ab-protein docking**)

Benchmark on 52 X-ray structures of Ab-protein complexes (Kringelum et al., 2012, PLoS Comp. Biol.)

Average AUC values

(* means p-value < 0.05 in comparison with DiscoTope 2)

- **0.73 DiscoTope 2** (Kringelum et al., 2012)
- **0.73 PEPITO** (Sweredoski & Baldi, 2008)
- **0.73 Epitopia** (Rubinstein et al., 2008)
- **0.72 SEPPA** (Sun et al., 2009)
- **0.71 DiscoTope 1** (Andersen et al., 2006)
- **0.69* ElliPro** (Ponomarenko et al., 2008)
- **0.65* EPCES** (Liang et al., 2009)
- **0.59* EPSVR** (Liang et al., 2010)

3D Structure-based epitope prediction

- The reason for the relatively poor performance is in the quality of the benchmark datasets
 - Structural information on the entire “biological unit” is often not available
 - Existence of well characterized epitopes from very few antigens