

# B Cell Epitope Prediction

**Presented by**

Mahita Jarjapu, PhD,  
Bioinformatics Postdoctoral Fellow

# Outline of this presentation:

1. B cell receptor (BCR) – antigen interactions: a brief introduction
2. B cell epitope prediction methods in the IEDB
  - A. Linear sequence-based prediction methods
  - B. Methods for predicting discontinuous epitopes

# Outline of this presentation:

1. **B cell receptor (BCR) – antigen interactions: a brief introduction**
2. B cell epitope prediction methods in the IEDB
  - A. Linear sequence-based prediction methods
  - B. Methods for predicting discontinuous epitopes

# An Introduction to BCR-Antigen Interactions

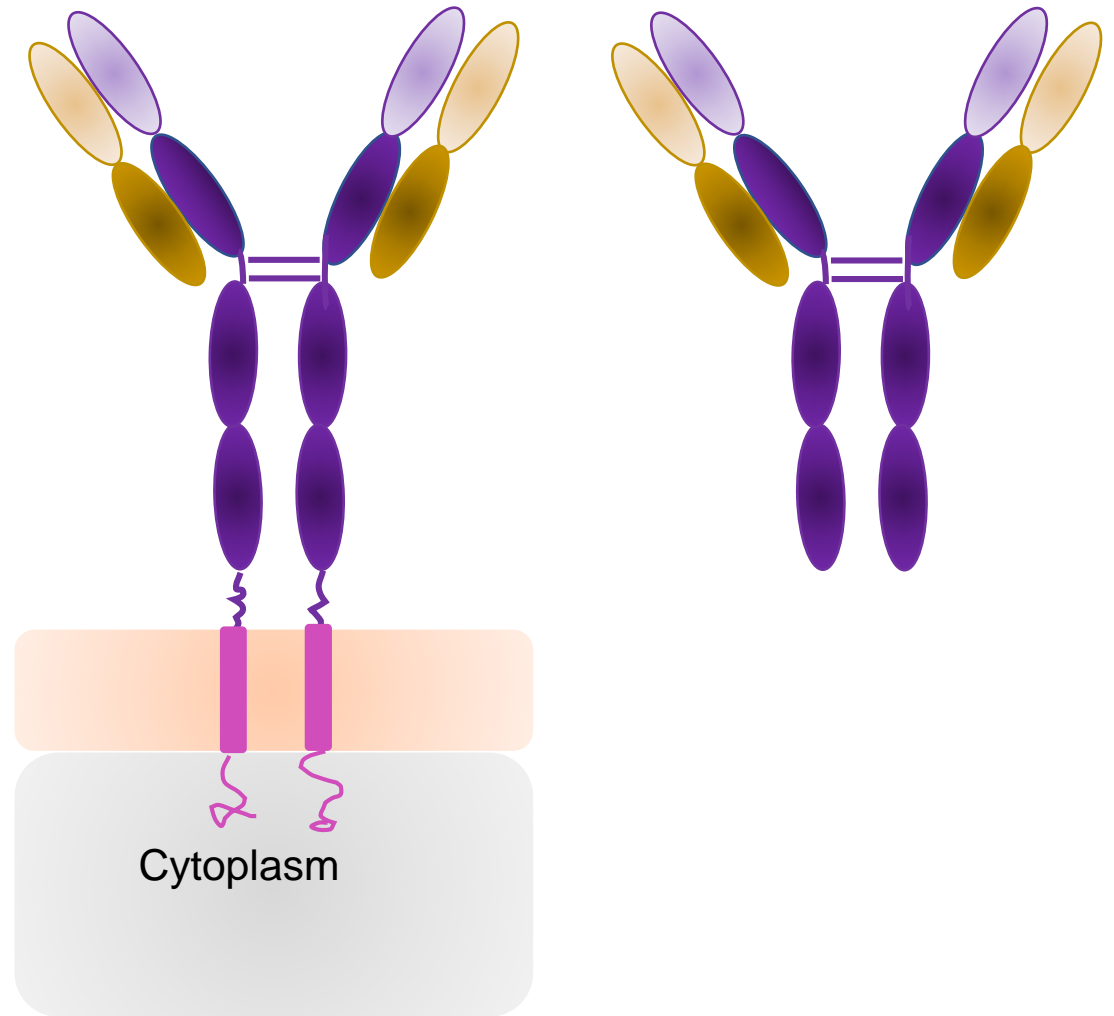
BCR – present on surface of B cells

Attached to B cells via transmembrane domain

Antibodies – secreted forms of BCRs

B-Cell plasma membrane

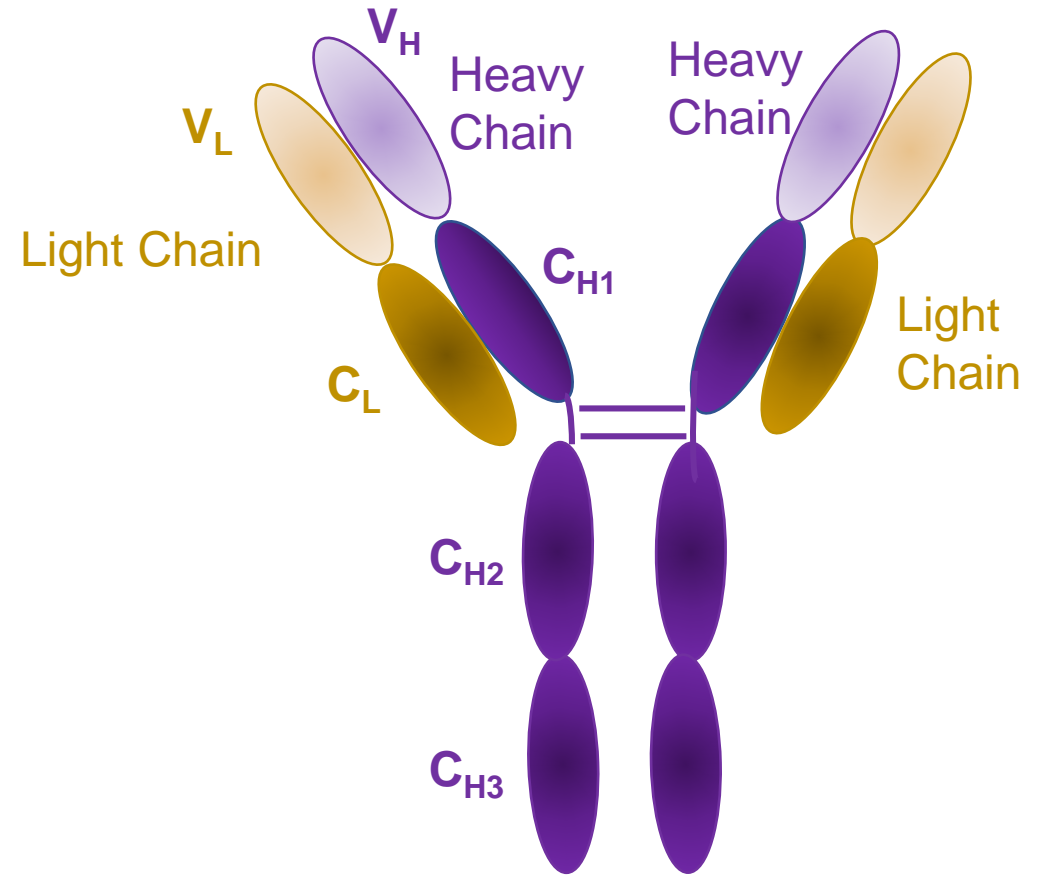
Cytoplasm



# An Introduction to BCR-Antigen Interactions

## Structure of BCR/antibody:

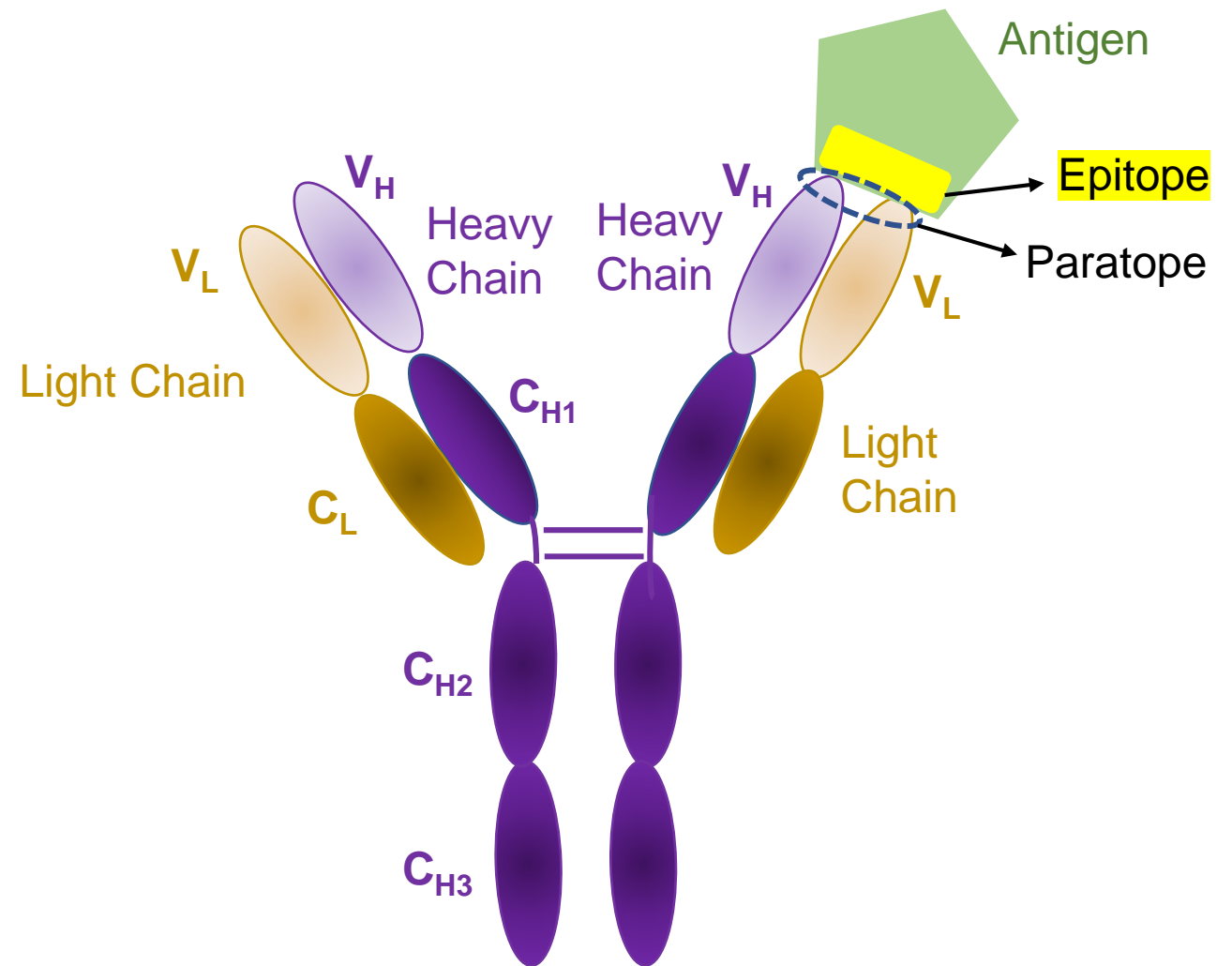
1. Y-shaped molecule, consists of 2 heavy chains and light chains
2. Each heavy chain is paired to a light chain
3. Heavy chain contains 1 variable domain ( $V_H$ ) and multiple constant domains ( $C_H$ )
4. Light chain contains 1 variable domain ( $V_L$ ) and one constant domain ( $C_L$ )



# An Introduction to BCR-Antigen Interactions

## BCR Interactions with Antigen:

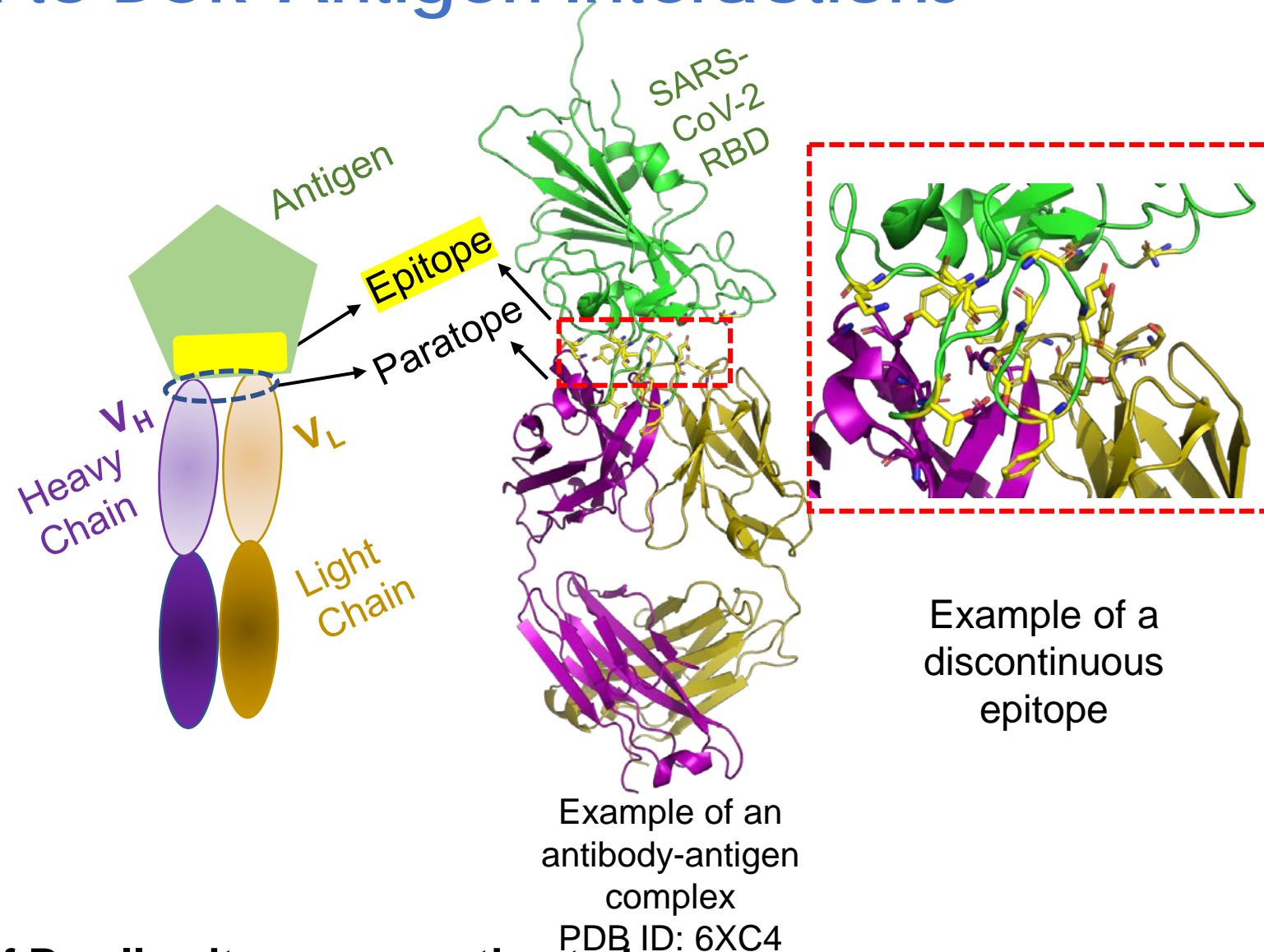
1. Variable domains of heavy and light chains interact with antigen
2. BCR epitope: Site on the antigen to which the BCR binds
3. BCR paratope: Site on the antibody that interacts with antigen



# An Introduction to BCR-Antigen Interactions

## Linear and Discontinuous Epitopes

- 1. Linear Epitope:** Epitope formed by a continuous stretch/sequence of amino acid residues
- 2. Discontinuous or Conformational Epitope:** Epitope formed by residues that are close together on the 3D structure but are non-sequential



**More than 90% of B cell epitopes are estimated to be discontinuous**

# Outline of this presentation:

1. B cell receptor (BCR) – antigen interactions: a brief introduction
- 2. B cell epitope prediction methods in the IEDB**
  - A. Linear sequence-based prediction methods
  - B. Methods for predicting discontinuous epitopes



# B Cell Epitope Prediction Tools in the IEDB

## 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.

1. Go to <http://tools.iedb.org/main/bcell/>
2. Page provides an overview of B cell epitope prediction methods

# B Cell Epitope Prediction Tools in the IEDB



When should I do epitope prediction?

?

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

# B Cell Epitope Prediction Tools in the IEDB



How do I learn more about an epitope prediction method ?

Visit the **Help** and **Reference** tabs to learn about a prediction method

# B Cell Epitope Prediction Tools in the IEDB

Visit the [Help](#) and [Reference](#) tabs to learn about a prediction method

**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 antigenic propensity of polypeptides chains have been correlated with the location of continuous epitopes. This has led to a search for empirical rules that would allow the prediction of epitopes to be predicted from certain features of the protein sequence. All prediction calculations are based on predicted values for each amino acid. Each scale consists of 20 values assigned to each of the amino acid residues on the basis of their relative antigenicity described by the scale.

**General method:** When computing the score for a given residue  $i$ , the amino acids in an interval of the chosen length  $n$  are considered. In other words, for a **window** size  $n$ , the  $i - (n-1)/2$  neighboring residues on each side of residue  $i$  were 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 BepiPred score or a residue score on the Karplus and Schulz flexibility scale, while the X-axis depicts the residue position in the protein sequence. The tables provide values of calculated scores for each residue. The larger score for the residues might indicate a higher probability to be part of epitope (those residues are colored in yellow on the graphs). However, the epitopes per se, either linear or discontinuous, -- they might only guide the researchers to further explore the protein sequence.

**Table 1. Implemented methods**

Method																																								
<b>Chou and Fasman beta turn prediction</b>																																								
<ul style="list-style-type: none"><li>Reference: <a href="#">Chou PY, Fasman GD. Prediction of the secondary structure of proteins from their amino acid sequence. <i>Biol. 1978;47:45-148.</i></a></li><li>Description: The rationale for predicting turns to predict antibody epitopes is based on the paper by <a href="#">Pelleguer 83-99</a>. Instead of implementing the turn scale of that paper which has some non-standard properties, we decided which is commonly used to predict beta turns as described in the reference link above.</li></ul>																																								
<b>Scale:</b>																																								
<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																					
<b>Emini surface accessibility scale</b>																																								
<ul style="list-style-type: none"><li>Reference: <a href="#">Emini EA, Hughes JV, Perlow DS, Boger J. Induction of hepatitis A virus-neutralizing antibody by a synthetic peptide. <i>J Virol</i> 1985 Sep;55(3):836-9.</a></li><li>Description: The calculation was based on surface accessibility scale on a product instead of an addition with was obtained using the formulae <math>S_n = (n+4+1) \cdot (0.37)^n - 6</math> where <math>S_n</math> is the surface probability, <math>n</math> is the fraction from 1 to 6. A hexapeptide sequence with <math>S_n</math> greater than 1.0 indicates an increased probability for being four residues accessible.</li></ul>																																								
<b>Scale:</b>																																								
<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																					
<b>Karplus and Schulz flexibility scale</b>																																								
<ul style="list-style-type: none"><li>Reference: <a href="#">Karplus PA, Schulz GE. Prediction of Chain Flexibility in Proteins - A tool for the Selection of Peptide Epitopes. <i>Nature</i> 1985; 72:212-3.</a></li></ul>																																								

## 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](#)

**BepiPred 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](#)

**BepiPred Linear Epitope Prediction 2.0:**  
Jespersen MC, Peters B, Nielsen M, Marcattili P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res (Web Server issue)*. 22.  
[PMID: 28472356](#)

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

# Outline of this presentation:

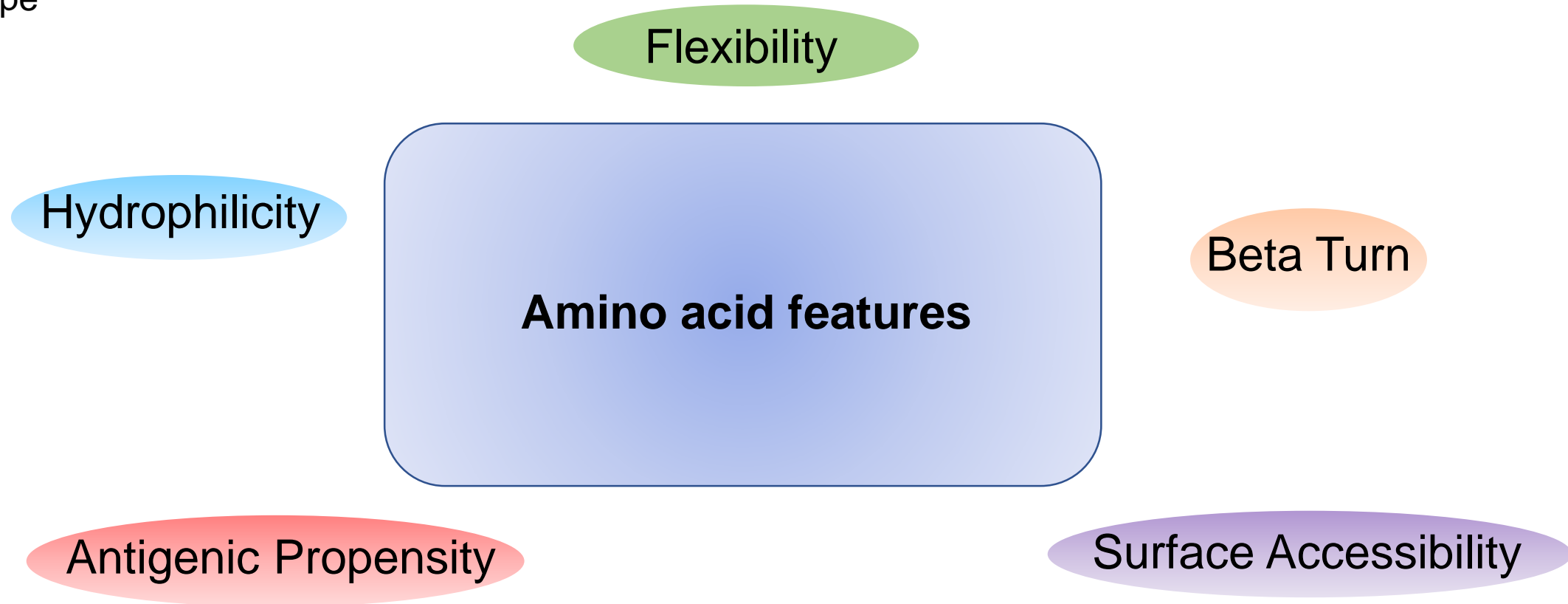
1. B cell receptor (BCR) – antigen interactions: a brief introduction
2. **B cell epitope prediction methods in the IEDB**
  - A. **Linear sequence-based prediction methods**
  - B. Methods for predicting discontinuous epitopes

# B Cell Epitope Prediction Tools in the IEDB

## Linear/Continuous Epitope Prediction Methods

Some amino acid features are correlated with location of continuous epitopes

These methods take advantage of this fact to predict if an amino acid on an antigen is part of a continuous epitope



# B Cell Epitope Prediction Tools in the IEDB

## Linear/Continuous Epitope Prediction Methods

Some amino acid features are correlated with location of continuous epitopes

These methods take advantage of this fact to predict if an amino acid on an antigen is part of a continuous epitope

Flexibility

Karplus and Schulz Flexibility Scale

Hydrophilicity

Parker Hydrophilicity Prediction

Methods based on physicochemical properties of amino acids in protein sequences

Beta Turn

Chou and Fasman Beta-Turn Predictor

Antigenic Propensity

Kolaskar and Tongaonkar Antigenicity Scale

Surface Accessibility

Emini Surface Accessibility Scale

# B Cell Epitope Prediction Tools in the IEDB

## Linear/Continuous Epitope Prediction Methods

Some amino acid features are correlated with location of continuous epitopes

These methods take advantage of this fact to predict if an amino acid on an antigen is part of a continuous epitope

### Bepipred-1.0

Uses Hidden Markov Models and propensity scale to predict location of linear B cell epitope

Methods based on physicochemical properties of amino acids in protein sequences  
**AND**  
machine learning

### Bepipred-2.0

Uses Random Forest algorithm trained on experimentally-verified epitopes and non-epitopes



# Linear/Continuous Epitope Prediction Method – An Example

## 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):

```
RVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRRKRISNCVADYSVLYNSASFSTFK
CYGVSPTKLNDLCFTNRYADSFVIRGDEVQRQIAPGQTGKIADYNYKLPDDFTGCVIAWNS
NNLDSKVGGNYNLYRFRKSNLKPFRDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQ
PTNGVGYQPYRVVVLSEFLLHAPATVCGPKKSTNLVKNKCVNF
```

**Choose a method:**

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

Let us now try to predict linear epitopes on an antigen using one of the methods

1. Go to <http://tools.iedb.org/bcell/>
2. Enter in an antigen protein sequence using either its Swiss-Prot ID or manually into the text box provided

### Example Sequence:

RBD region from SARS-Cov-2 Spike glycoprotein

Swiss-Prot ID: P0DTC2

3. Select any one of the prediction methods
4. Click “Submit”

# Linear/Continuous Epitope Prediction Method – An Example

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

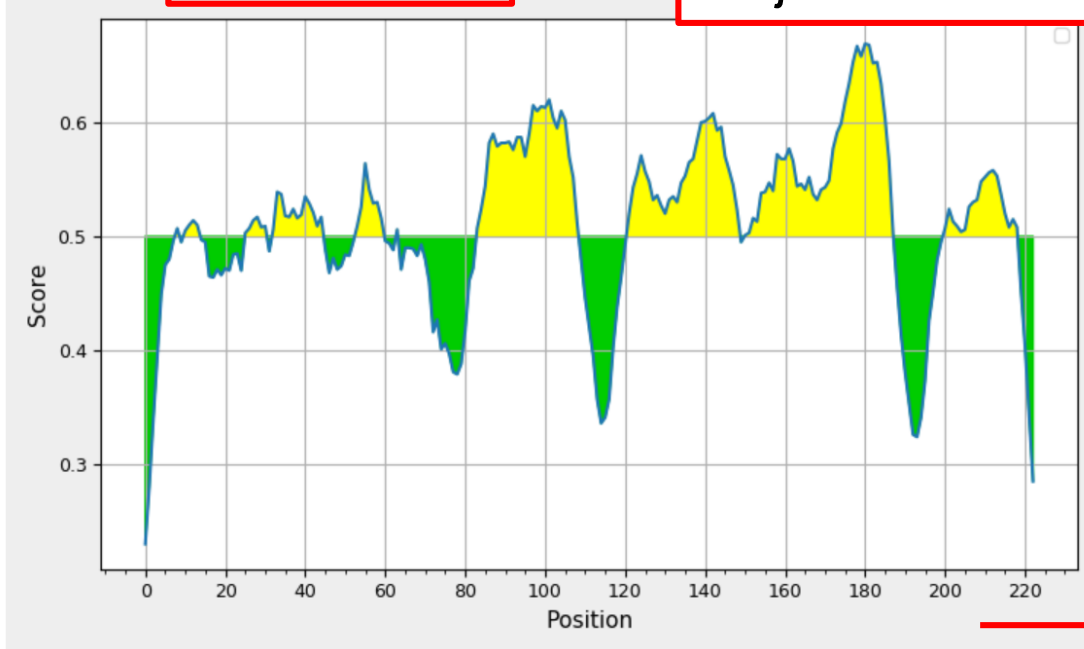
## Bepipred Linear Epitope Prediction 2.0 Results

### Input Sequences

```
1 RVQPTESIVR FPNITNLCPF GEVFNATRFA SVYAWNRKRI SNCVADYSVL YNSASFSTFK
61 CYGVSP TK LNDLCFTNVY ADSEVIRGDE VRQIAPGQTG KIADYNYKLP DDFTGCVIAW
121 NS NNLD SK VGGNYNYLYR LFRKSNLKP ERDISTEIQ AGSTPCNGVE GFNCYFPLQS
181 YGFQ PTNG VGYQPYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
```

Center position: 4 Threshold: 0.500 Recalculate

Adjustable threshold value



After submitting the sequence, it returns a graph shown on the left

Y-axis: Bepipred-2.0 score for each residue number  
Residue numbers having scores above a threshold value are predicted to be epitope residues

X-axis: Residue numbers of RBD protein sequence

Average: 0.511 Minimum: 0.230 Maximum: 0.669

Average score of a protein chosen as a threshold by default

# Linear/Continuous Epitope Prediction Method – An Example

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

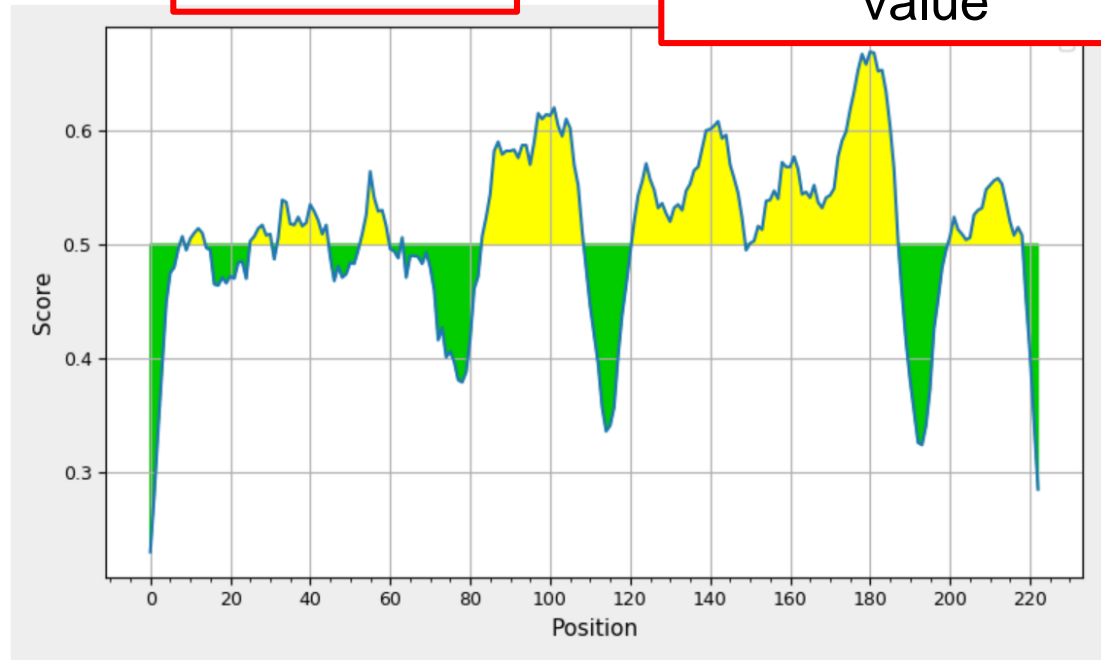
## Bepipred Linear Epitope Prediction 2.0 Results

### Input Sequences

```

1 RVQPTESIVR FPNITNLCPF GEVFNATRFA SVYAWNKRRI SNCVADYSVL YNSASFSTFK
61 CYGVSP TK LNDLCFTN VY ADSEFVIRGDE VRQIAPGQTG KIADYNYKLP DDFTGCVIAH
121 NS NNLD SK VGGNYNYLYR LFRKSNLKP ERDISTEIQ AGSTPCNGVE GFNCYFPI
181 YGFQ PTNG VGYQPYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
    
```

Center position: 4



Average: 0.511 Minimum: 0.230 Maximum: 0.669

Average score of a protein chosen as a threshold by default

### Predicted peptides:

No.	Start	End	Peptide	Length
1	9	9	V	1
2	11	14	FPNI	4
3	26	31	ATRFAS	6
4	33	45	YAWNKRIRISNCVA	13
5	54	60	ASFSTFK	7
6	64	64	V	1
7	84	109	IRGDEV RQIAPGQTGKIADYNYKLPD	26
8	122	149	NLDSKVGGNYNYLYRLFRKSNLKPFRD	28
9	151	188	STEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ	38
10	201	219	HAPATVCGPKKSTNLVKNK	19

Results: a table of predicted linear epitopes, their starting and ending positions, and length

### Predicted residue scores:

Position	Residue	Score	Assignment
0	R	0.230	.
1	V	0.280	.
2	Q	0.339	.
3	P	0.394	.
4	T	0.449	.
5	E	0.475	.
6	S	0.480	.
7	I	0.496	.
8	V	0.507	E
9	R	0.495	.
10	F	0.505	E
11	P	0.510	E
12	N	0.514	E
13	I	0.510	E
14	T	0.497	.
15	N	0.495	.
16	L	0.465	.
17	C	0.464	.
18	P	0.471	.
19	F	0.466	.

Results: a table of residues and their predicted scores "E" = epitope

# Do multiple methods give the same predictions?

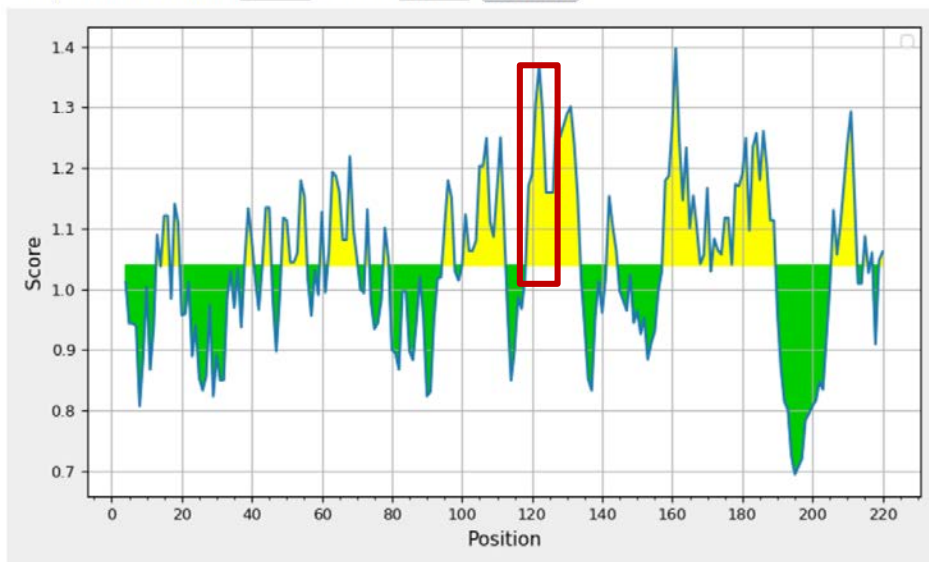
Let us check! As an example, we will compare predictions made on the RBD of SARS-CoV-2

## Chou & Fasman Beta-Turn Prediction Results

Input Sequences

```
1 RVQPTESIVR FPNITNLCPF GEVFNATRFA SVYAWNRKRI SNCVADYSVL YNSASFSTFK
61 CYGVSPFK LNDLCFTNVY ADSFVIRGDE VRQIAPGQTG KIADYNYKLP DDFTCGVIAW
121 NS NNLDK VGGNYNYLYR LFRKSNLKP ERDISTEIQ AGSTPCNGVE GFNCYFPLQS
181 YGFQ PTNG VGYQYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
```

Center position: 4 Window size: 7 Threshold: 1.039 Recalculate



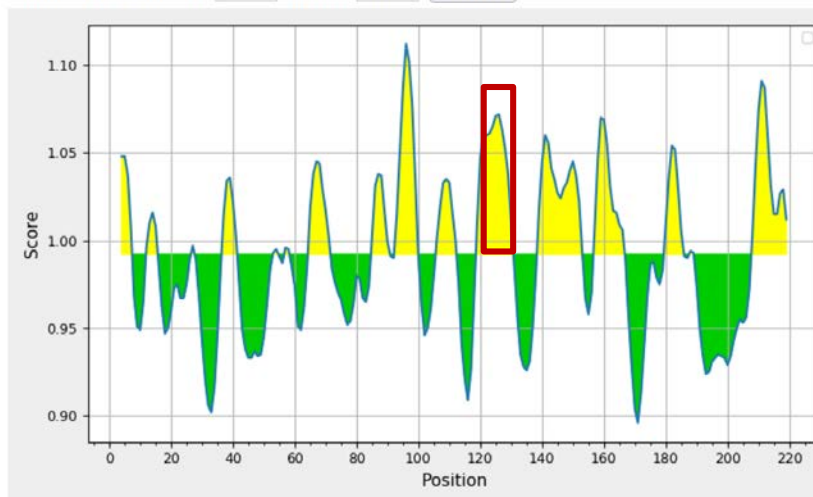
Average: 1.039 Minimum: 0.694 Maximum: 1.397

## Karplus & Schulz Flexibility Prediction Results

Input Sequences

```
1 RVQPTESIVR FPNITNLCPF GEVFNATRFA SVYAWNRKRI SNCVADYSVL YNSASFSTFK
61 CYGVSPFK LNDLCFTNVY ADSFVIRGDE VRQIAPGQTG KIADYNYKLP DDFTCGVIAW
121 NS NNLDK VGGNYNYLYR LFRKSNLKP ERDISTEIQ AGSTPCNGVE GFNCYFPLQS
181 YGFQ PTNG VGYQYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
```

Center position: 4 Window size: 7 Threshold: 0.992 Recalculate



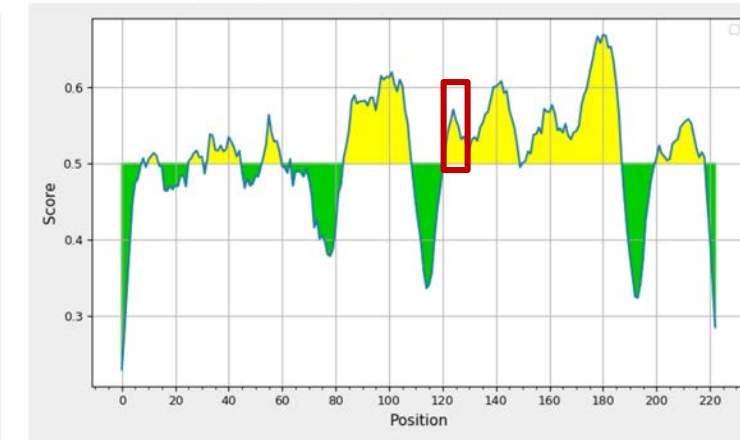
Average: 0.992 Minimum: 0.896 Maximum: 1.112

## Bepipred Linear Epitope Prediction 2.0 Results

Input Sequences

```
1 RVQPTESIVR FPNITNLCPF GEVFNATRFA SVYAWNRKRI SNCVADYSVL YNSASFSTFK
61 CYGVSPFK LNDLCFTNVY ADSFVIRGDE VRQIAPGQTG KIADYNYKLP DDFTCGVIAW
121 NS NNLDK VGGNYNYLYR LFRKSNLKP ERDISTEIQ AGSTPCNGVE GFNCYFPLQS
181 YGFQ PTNG VGYQYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
```

Center position: 4 Threshold: 0.500 Recalculate



Average: 0.511 Minimum: 0.230 Maximum: 0.669

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

# Outline of this presentation:

1. B cell receptor (BCR) – antigen interactions: a brief introduction
- 2. B cell epitope prediction methods in the IEDB**
  - A. Linear sequence-based prediction methods
  - B. Methods for predicting discontinuous epitopes**

# B Cell Epitope Prediction Tools in the IEDB

## Discontinuous/Conformational Epitope Prediction Methods

<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

ElliPro

# B Cell Epitope Prediction Tools in the IEDB

## Discontinuous/Conformational Epitope Prediction Methods

**Version 1:** Takes into account *statistics*, *spatial location* and *surface accessibility of amino acids* in protein structures to predict epitope residues

Logic: Discontinuous epitopes are present on protein surface.  
Certain amino acids are found more on the surface.  
Residues that are far apart in sequence may be close to one another in the 3D structure and could be part of the epitope

[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 P<sup>1</sup>](#), [Nielsen M.](#) [Lund O.](#)

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

# B Cell Epitope Prediction Tools in the IEDB

## Discontinuous/Conformational Epitope Prediction Methods

### DiscoTope

**Version 1:** Takes into account *statistics*, *spatial location* and *surface accessibility of amino acids* in protein structures to predict epitope residues

[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 P<sup>1</sup>](#), [Nielsen M](#), [Lund O](#).

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

**Version 2:** Improved over version 1  
Takes into account multiple epitopes on an antigen

[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 JV<sup>1</sup>](#), [Lundegaard C](#), [Lund O](#), [Nielsen M](#).

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

Assigns a score for each residue  
Score = linear combination of normalized values

Parker's hydrophilicity scale  
Amino acid occurrence  
Number of contacts within 10Å  
Area of relative solvent accessibility



# B Cell Epitope Prediction Tools in the IEDB

## Using DiscoTope to Predict Discontinuous Epitopes

IEDB Analysis Resource

Home Help Example Reference Download Contact

### DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID  (example: 1z40)  
Or upload a PDB file  No file selected.

Step 2: Please enter PDB Chain ID

Step 3: Select version

Identify structure

Input PDB id or upload file

Specific to the protein chain of interest

1. Go to <http://tools.iedb.org/discotope/>

2. Provide a 3D structure of the antigen by:

Using its PDB ID

OR

Uploading the PDB file

3. PDB structures have multiple chains. Choose a chain of interest by specifying its chain ID.

# How to Search for 3D Structures in the Protein Data Bank?

The screenshot shows the RCSB PDB website interface. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. Below this is a secondary bar with logos for PDB-101, PDB, EMDataResource, NUCLEIC ACID DATABASE, and wwPDB Foundation. A banner for 'NEW! Computed Structure Models (CSM)' is visible. On the left, a dark blue sidebar contains navigation options: Welcome, Deposit, Search (highlighted with a red box), Visualize, Analyze, Download, and Learn. The main content area features a grid of search and tool options: Advanced Search, Chemical Similarity Search, Browse by Annotations, Unreleased Entries, Sequence Search (highlighted with a red box), Chemical Sketch Tool, New Entries, and PDB Statistics. At the bottom, there are sections for 'Latest Entries' (dated Oct 11 2022), 'Features & Highlights', 'News', and 'Publications'.

1. Go to <http://www.rcsb.org/>
2. Click on “Search” on the left-hand side
3. Click on “Sequence Search”

# How to Search for 3D Structures in the Protein Data Bank?

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

RCSB PDB PROTEIN DATA BANK 196,779 Structures from the PDB 1,000,361 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM

Advanced Search | Browse Annotations Help

PDB-101 PDB EMDDataResource NUCLEIC ACID DATABASE wwPDB Foundation

Search Query History Browse Annotations MyPDB

Use the **Advanced Search Query Builder** tool to create composite boolean queries. See the [Help](#) page for more detailed information.

Advanced Search Query Builder Help

Full Text Structure Attributes Chemical Attributes

Sequence Similarity Help

MTTQAPTFTQPLQSVVLEG. Enter a sequence containing a minimum of 25 residues, OR enter an Entry ID in the text box below and select from the sequence list. The second option is useful for finding sequences that are similar to a sequence from a given structure and chain.

Entry ID 1MBN SequenceType Protein E-Value Cutoff 0.1 Identity Cutoff 0 % (Integer only) Count Clear

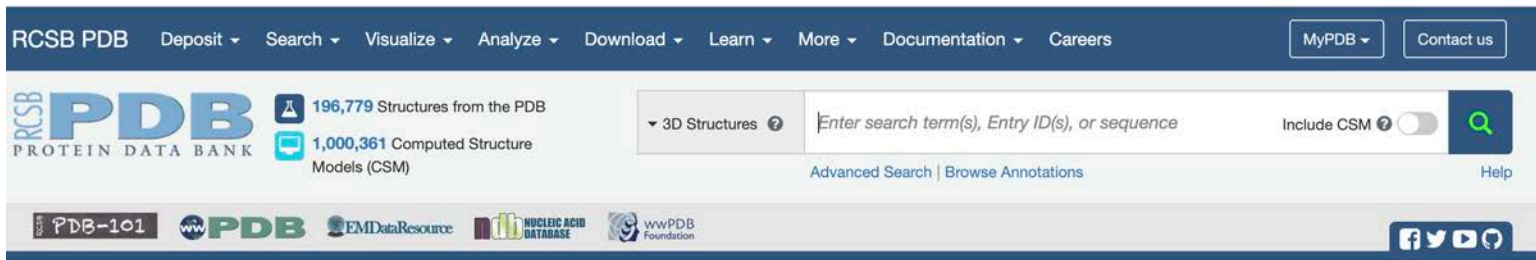
Sequence Motif Structure Similarity Structure Motif Chemical Similarity

Return Structures grouped by No Grouping Include Computed Structure Models (CSM) Count Clear Search

4. Paste the sequence of your protein antigen for which you want to predict epitopes and then click on “Search” in the bottom right corner of the page.

5. We will enter in the sequence of the AMA1 protein from *Plasmodium falciparum*

# How to Search for 3D Structures in the Protein Data Bank?



6. It will take us to a page containing details of the structure of the AMA1 protein  
PDB ID : 1Z40

The image shows the PDB entry page for 1Z40. On the left, there is a 3D ribbon diagram of the protein structure. The main content area contains the following information:

- 1Z40**
- AMA1 from Plasmodium falciparum**
- PDB DOI:** 10.2210/pdb1Z40/pdb
- Classification:** UNKNOWN FUNCTION
- Organism(s):** Plasmodium falciparum 3D7
- Expression System:** Escherichia coli BL21(DE3)
- Mutation(s):** No
- Deposited:** 2005-03-14 **Released:** 2005-08-16
- Deposition Author(s):** Bai, T., Becker, M., Gupta, A., Strike, P., Murphy, V.J.,

Below this information, there is a section for "Experimental Data Snapshot" with the following details:

- Method:** X-RAY DIFFRACTION
- Resolution:** 1.90 Å
- R-Value Free:** 0.236
- R-Value Work:** 0.192
- R-Value Observed:** 0.195

To the right of the experimental data, there is a "wwPDB Validation" section with a bar chart showing various metrics: Rfree, Clashscore, Ramachandran outliers, Sidechain outliers, and RSRZ outliers. A red box highlights a dropdown menu titled "Download Files" which contains the following options:

- 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)

The structure file can be downloaded in PDB format by clicking on “Download Files” and selecting “PDB Format” from the dropdown menu

7. We will choose this PDB ID to predict discontinuous epitopes on the AMA1 protein using Discotope


8. Scrolling down this page will give us information about Chain Ids in the AMA1 protein structure

# How to Search for 3D Structures in the Protein Data Bank?

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB Contact us

Macromolecules

Find similar proteins by: [Sequence](#) (by identity cutoff) | [3D Structure](#)

Entity ID: 1					
Molecule	Chains	Sequence Length	Organism	Details	Image
apical membrane antigen 1 precursor	A, B [auth E]	336	<a href="#">Plasmodium falciparum 3D7</a>	Mutation(s): 0 Gene Names: <a href="#">Plasmodium falciparum, PF3D7_1133400</a>	

UniProt

Find proteins for [Q7KQK5](#) (*Plasmodium falciparum* (isolate 3D7)) Explore [Q7KQK5](#) Go to UniProtKB: [Q7KQK5](#)


Entity Groups

Sequence Clusters: [30% Identity](#) [50% Identity](#) [70% Identity](#) [90% Identity](#) [95% Identity](#) [100% Identity](#)

UniProt Group: [Q7KQK5](#)

Protein Feature View [Expand](#)

Reference Sequence: [1Z40\\_1](#)



9. There are two units of the AMA1 protein in the crystal structure, one with chain id 'A' and the other with 'B/E'

10. We need to select only one chain ID

# B Cell Epitope Prediction Tools in the IEDB

## Using DiscoTope to Predict Discontinuous Epitopes in AMA1 Protein

**IEDB Analysis Resource**

Home Help Example Reference Download Contact

### DiscoTope: Structure-based Antibody Prediction

Step 1: Please enter the 4-letter PDB ID  (example: 1z40)  
Or upload a PDB file  No file selected.

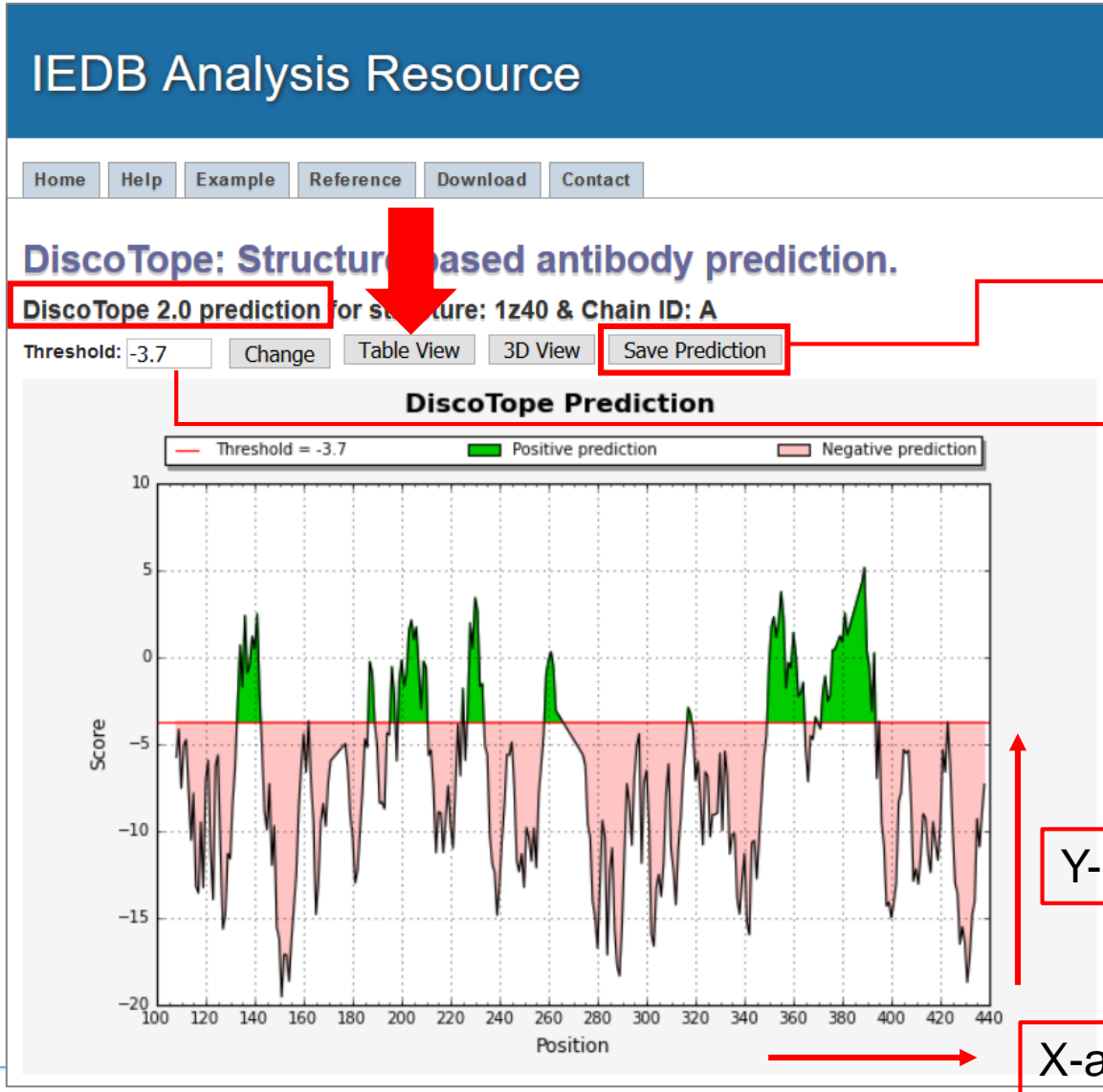
Step 2: Please enter PDB Chain ID

Step 3: Select version

1. Enter in the PDB ID of AMA1 protein (1z40) and the chain ID ('A')
2. Select the version of DiscoTope (1.1 or 2.0)
3. Click 'Submit'

# B Cell Epitope Prediction Tools in the IEDB

## Using DiscoTope to Predict Discontinuous Epitopes in AMA1 Protein



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

Let us see what epitopes DiscoTope predicts for the AMA1 protein

Downloads as csv

Adjustable threshold

1. Results shown as a plot
2. Residue positions colored in green (have values above a threshold value) predicted to be epitope residues

3. User has option to save prediction in CSV format

4. User can also view results in form of a table

# B Cell Epitope Prediction Tools in the IEDB

## Using DiscoTope to Predict Discontinuous Epitopes in AMA1 Protein

<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

1. Table view of DiscoTope predictions
2. DiscoTope scores for each residue
3. Residues predicted to be epitope residues (positive predictions) highlighted in green
4. User has option to view epitope residues on 3D structure by clicking on '3D View'



# B Cell Epitope Prediction Tools in the IEDB

## Discontinuous/Conformational Epitope Prediction Methods

ElliPro

[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](#)<sup>1</sup>, [Bui HH](#), [Li W](#), [Fusseder N](#), [Bourne PE](#), [Sette A](#), [Peters B](#).

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

[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](#)

Predicts linear and discontinuous antibody epitopes based on the geometrical properties of protein structure

Uses Thornton's Method

Performs the following steps to estimate a score for each predicted epitope:

- i. Protein shape approximated by multiple ellipsoids
- ii. Residue protrusion index (PI) calculated for each ellipsoid  
PI = 0.8: 80% of protein residues included in the ellipsoid, 20% outside the ellipsoid
- iii. Clustering of neighboring residues based on PI values

# B Cell Epitope Prediction Tools in the IEDB

## Using ElliPro to Predict Discontinuous Epitopes

IEDB 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)

1. Go to <http://tools.iedb.org/ellipro/>
2. Enter the PDB ID or upload PDB file of protein of interest

### Identify structure

Input PDB id or upload file

### Select min. PI value

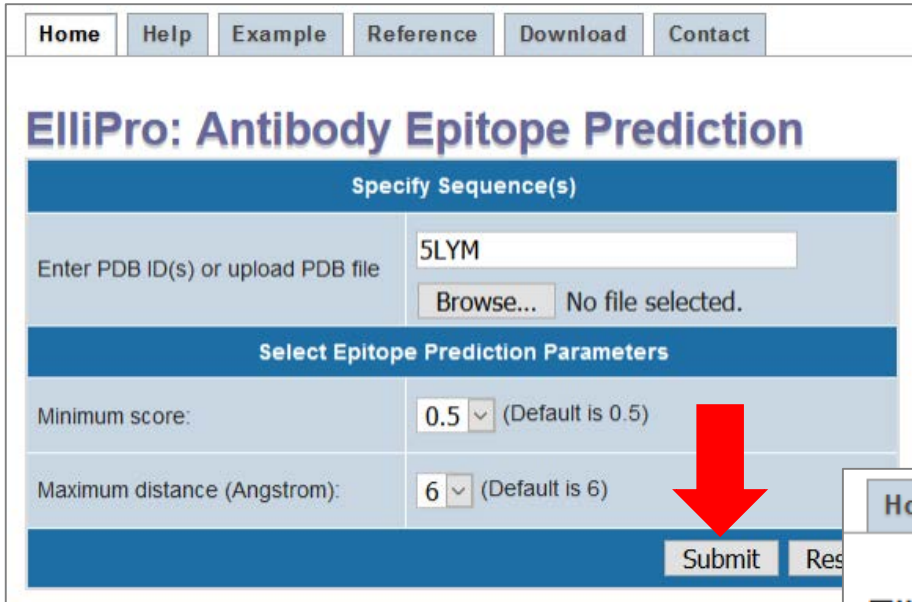
- Averaged over epitope residues
- Higher scores predict fewer epitopes

Specify max distance for predicting (grouping) discontinuous epitopes

Longer distances predict discontinuous epitopes spanning larger regions

# B Cell Epitope Prediction Tools in the IEDB

## Using ElliPro to Predict Discontinuous Epitopes



Home Help Example Reference Download Contact

### ElliPro: Antibody Epitope Prediction

**Specify Sequence(s)**

Enter PDB ID(s) or upload PDB file: 5LYM  
Browse... No file selected.

**Select Epitope Prediction Parameters**

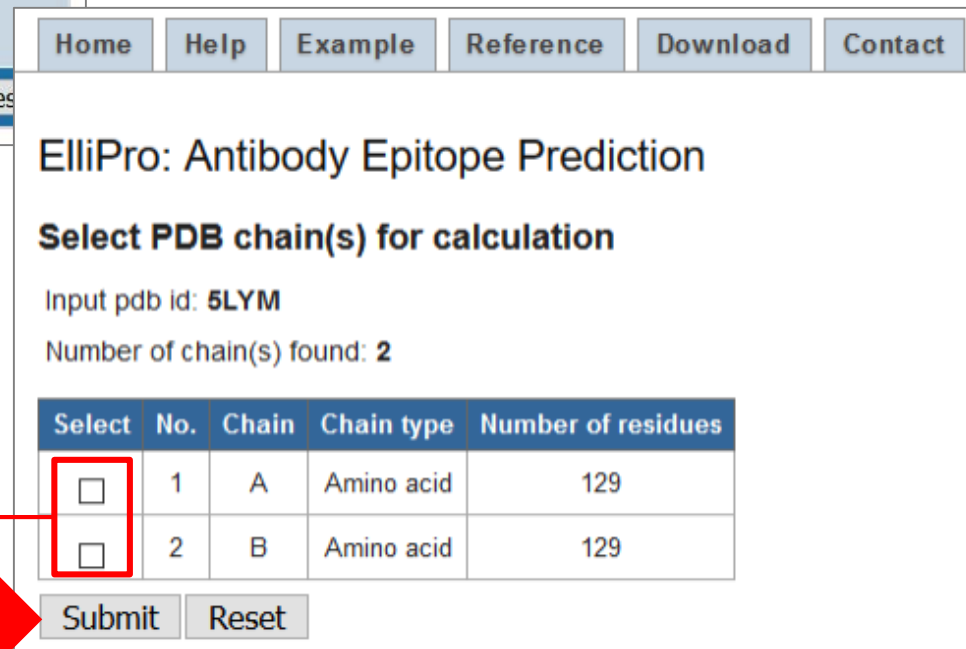
Minimum score: 0.5 (Default is 0.5)  
Maximum distance (Angstrom): 6 (Default is 6)

Submit Reset

3. Click 'Submit'

4. If there are multiple chains found in the PDB structure, user is asked to choose a chain ID

5. Click 'Submit' again



Home Help Example Reference Download Contact

### ElliPro: 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 type="checkbox"/>	1	A	Amino acid	129
<input type="checkbox"/>	2	B	Amino acid	129

Submit Reset

Select chain(s) of interest

# B Cell Epitope Prediction Tools in the IEDB

## Using ElliPro to Predict Discontinuous Epitopes

ElliPro predicts both linear and discontinuous epitopes

IEDB Analysis Resource

Home Help Example Reference Download Contact

Input Sequences: 5LYM

Chain: A  
1 KVFGRCELAA AMKRHGLDNY RGYSLGNWVC AAKFESHFNT QAINRNIDGS IDYGILQINS  
61 RWWCNDGRTP GSRNLCNIPC SALLSSDITA SVNCAKKIVS DGNGBNAWA WRNRCKGIDV  
121 QAWIRGRL

Predicted Linear Epitope(s):

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

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	<a href="#">View</a>
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	<a href="#">View</a>
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	<a href="#">View</a>
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	<a href="#">View</a>

[Click here to view residue scores](#)

[Download pdb file](#)

For each predicted epitope, ElliPro provides a score

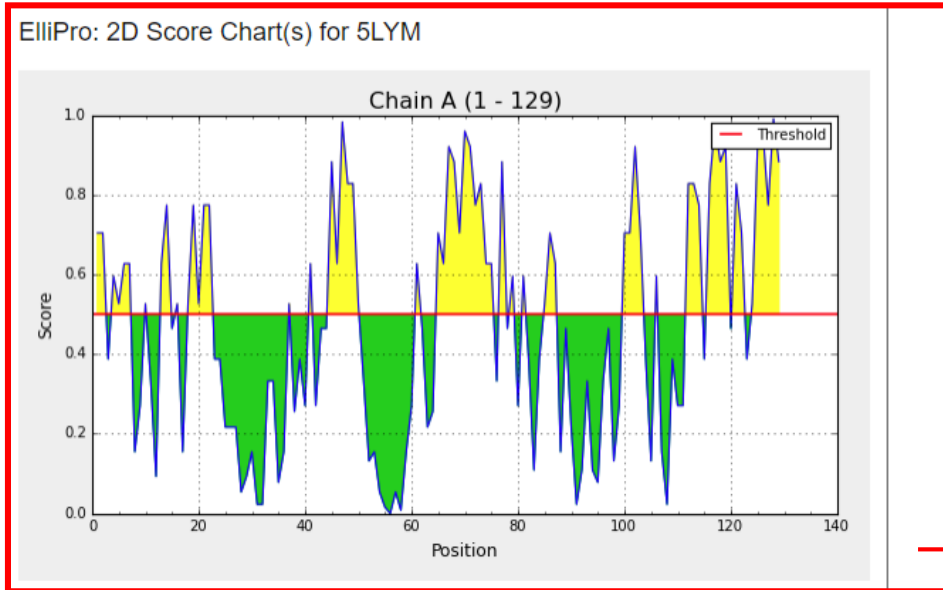
Linear Epitopes Predicted

Click to view 3D structure

Discontinuous Epitopes Predicted

# B Cell Epitope Prediction Tools in the IEDB

## Using ElliPro to Predict Discontinuous Epitopes



Graph View of Per-Residue Scores

Y-axis: ElliPro score  
Residues having ElliPro score above a threshold are predicted to be epitope residues

X-axis: Residue Position

Data table

No.	Chain	Residue number	Residue name	Score
1	A	1	LYS	0.705
2	A	2	VAL	0.705
3	A	3	PHE	0.388
4	A	4	GLY	0.597
5	A	5	ARG	0.527
6	A	6	CYS	0.628
7	A	7	GLU	0.628
8	A	8	LEU	0.155
9	A	9	ALA	0.271
10	A	10	ALA	0.527
11	A	11	ALA	0.333
12	A	12	MET	0.093
13	A	13	LYS	0.628
14	A	14	ARG	0.775
15	A	15	HIS	0.465
16	A	16	GLY	0.527
17	A	17	LEU	0.155

Table View

# B Cell Epitope Prediction Tools in the IEDB

## Summary

Linear and discontinuous (conformational) epitopes can be overlapping, 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.

**Thank You!**