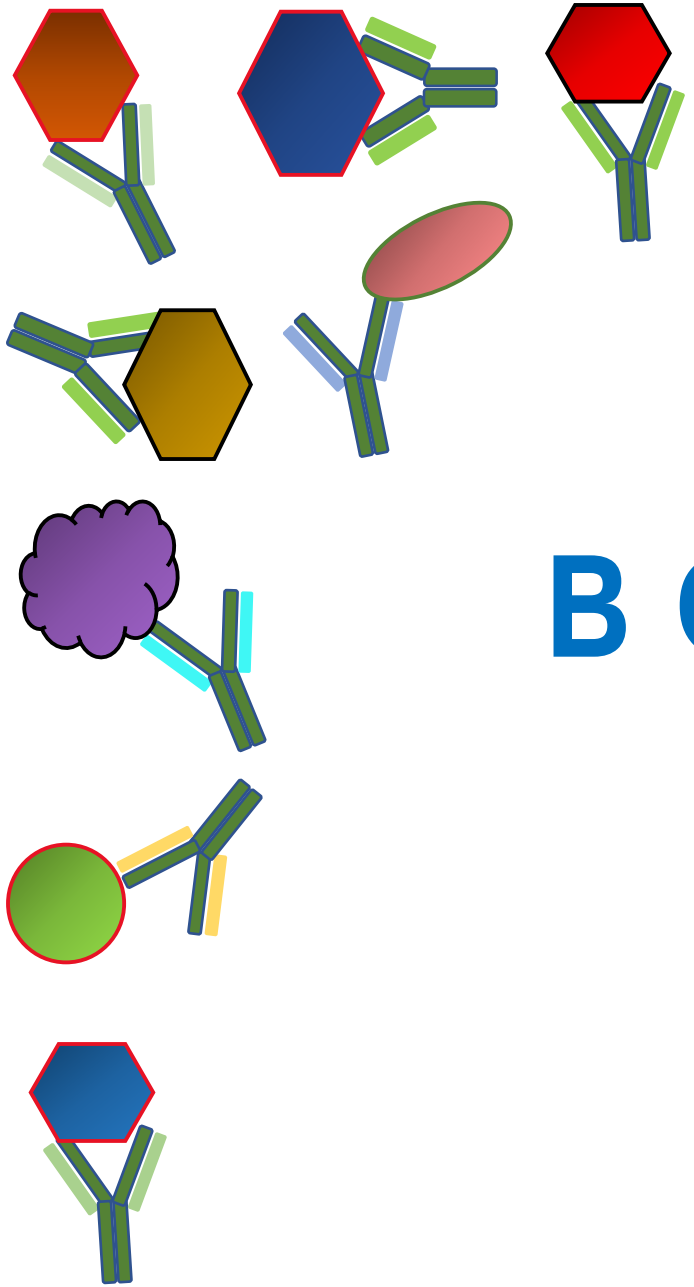


IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE

tools.iedb.org

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:

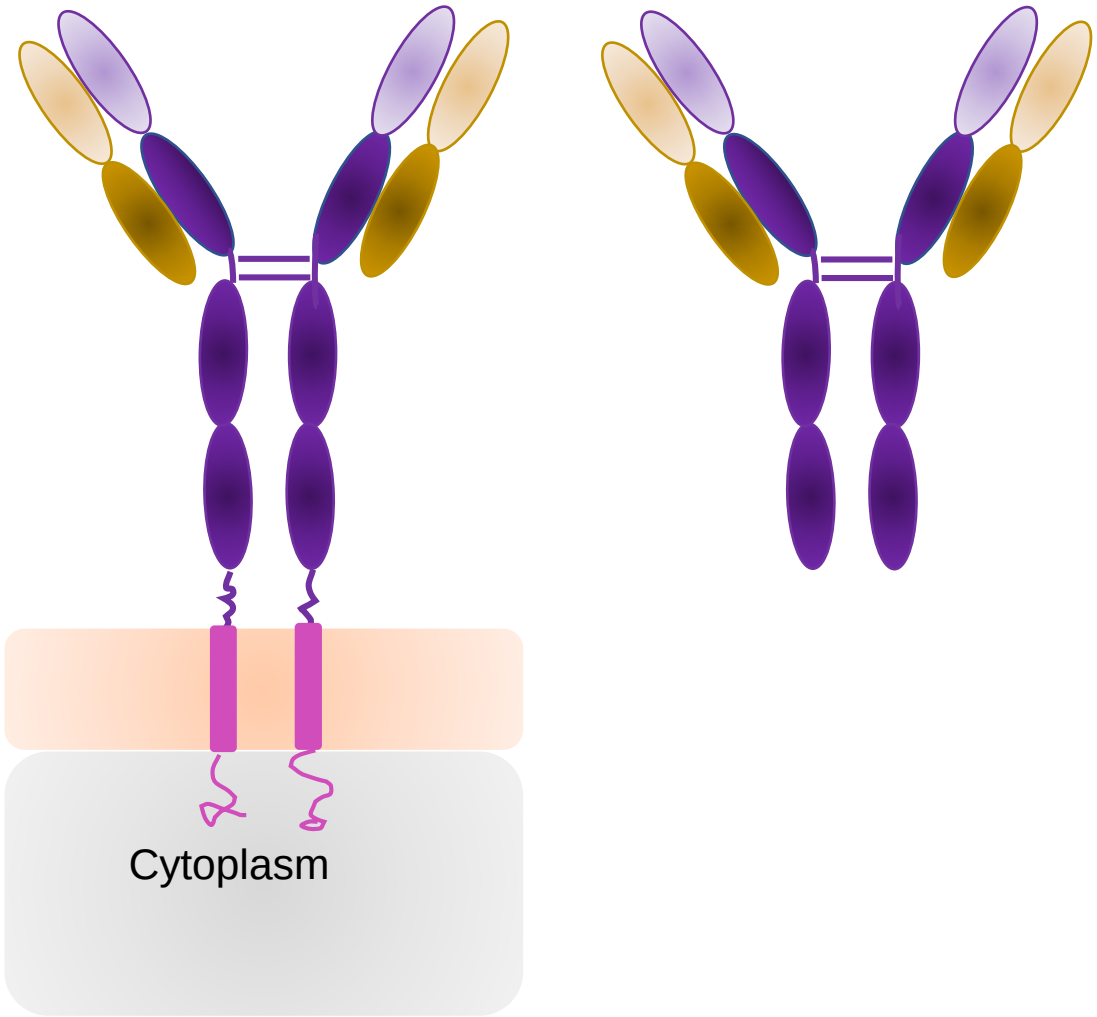
1. **B cell receptor (BCR) – antigen interactions: a brief introduction**
2. B cell epitope prediction methods in the IEDB
    A. Linear sequence-based prediction methods
    B. Methods for predicting discontinuous epitopes

An Introduction to BCR-Antigen Interactions

BCR – present on surface of B cells

Attached to B cells via transmembrane domain

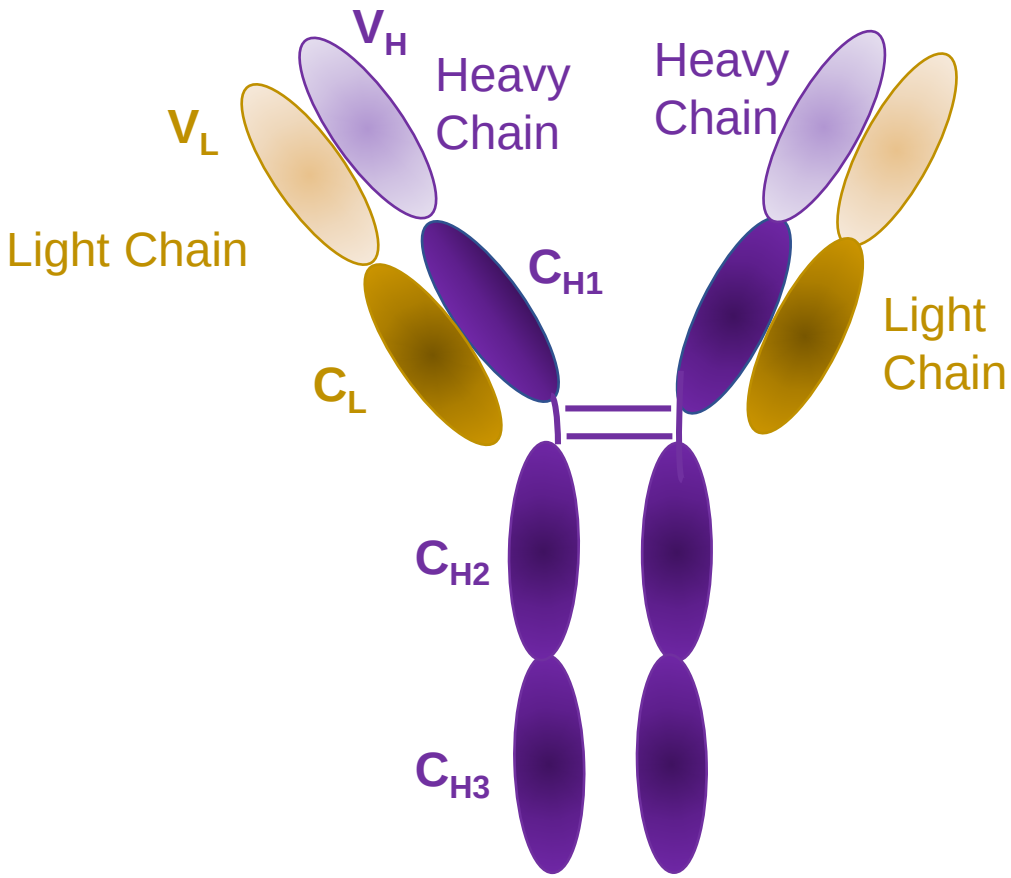
Antibodies – secreted forms of BCRs



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)



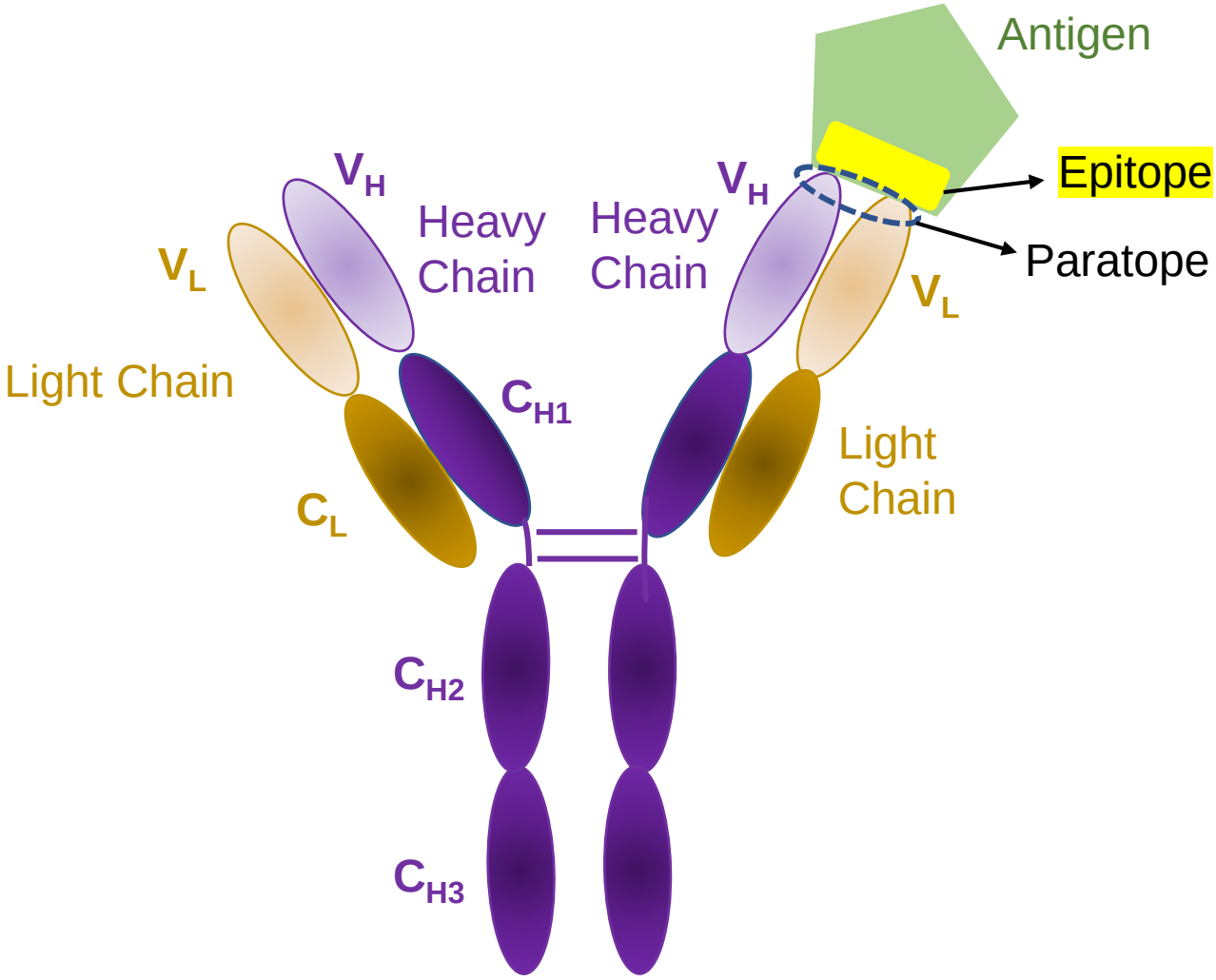
An Introduction to BCR-Antigen Interactions

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

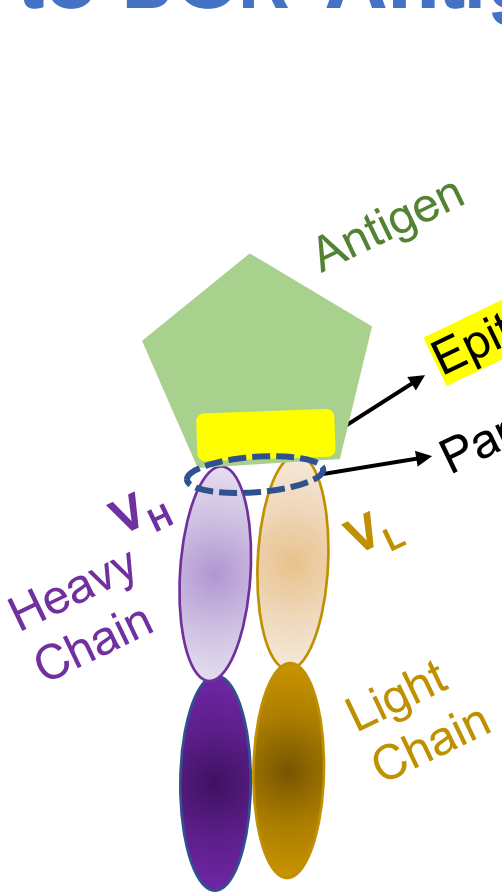


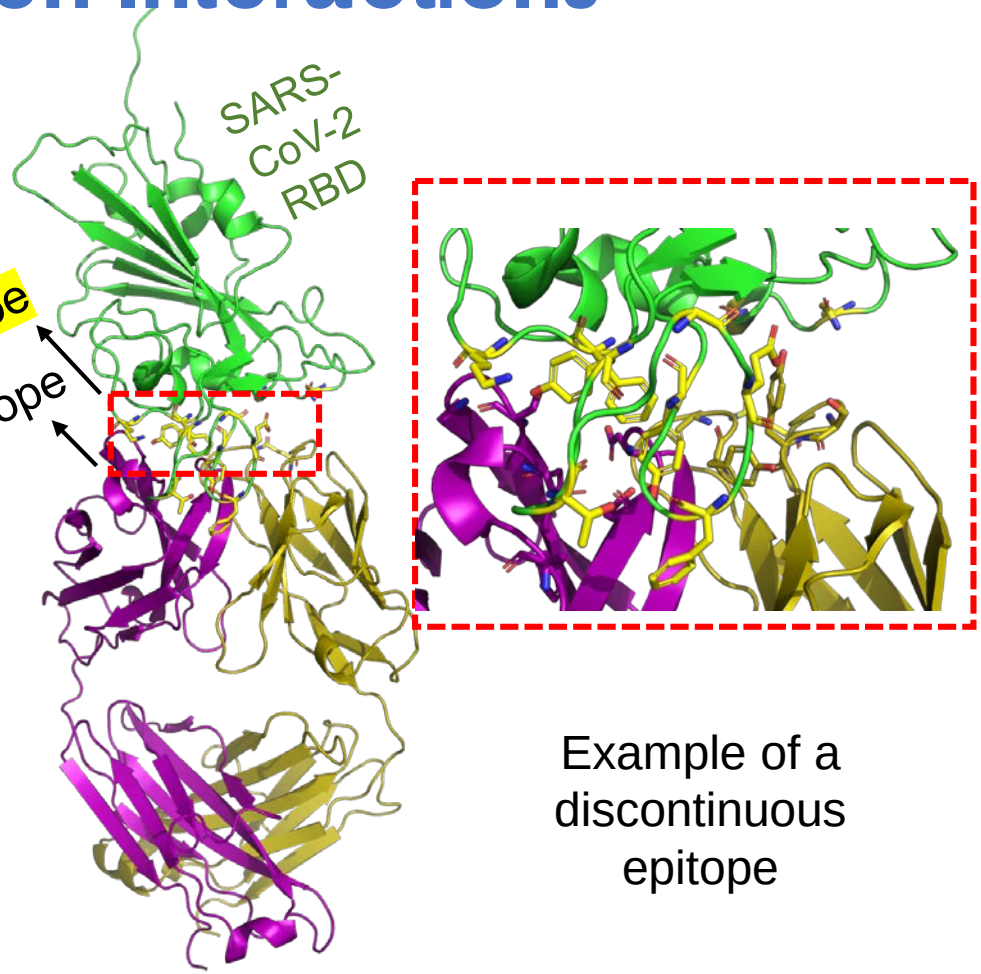
An Introduction to BCR-Antigen Interactions

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





Example of an antibody-antigen complex
PDB ID: 6XC4

Example of a discontinuous epitope

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

*Barlow et al, Nature. 1986.
*Van Regenmortel, Methods. 1996.

Outline of this presentation:

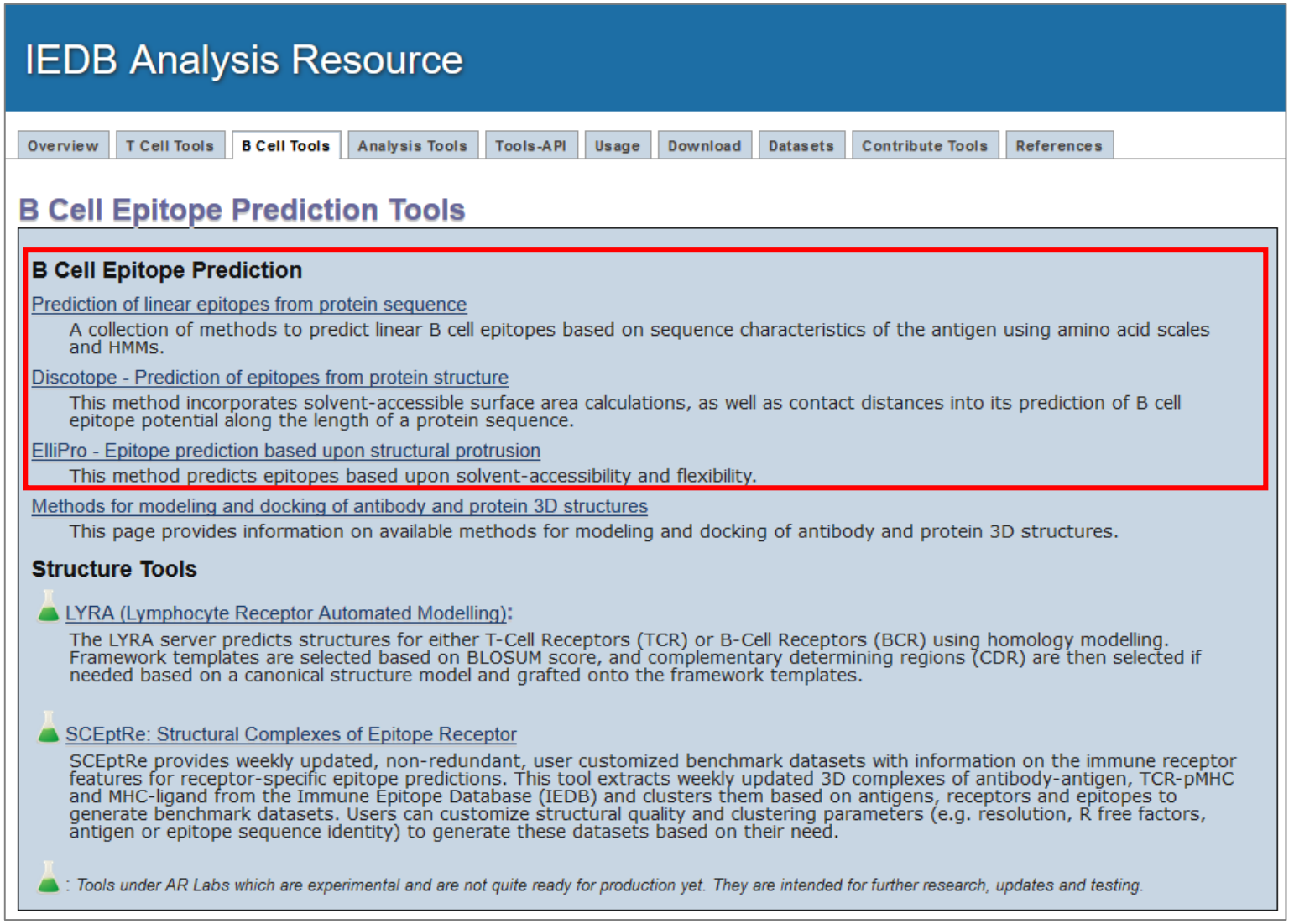
1. B cell receptor (BCR) – antigen interactions: a brief introduction

2. **B cell epitope prediction methods in the IEDB**

    A. Linear sequence-based prediction methods

    B. Methods for predicting discontinuous epitopes

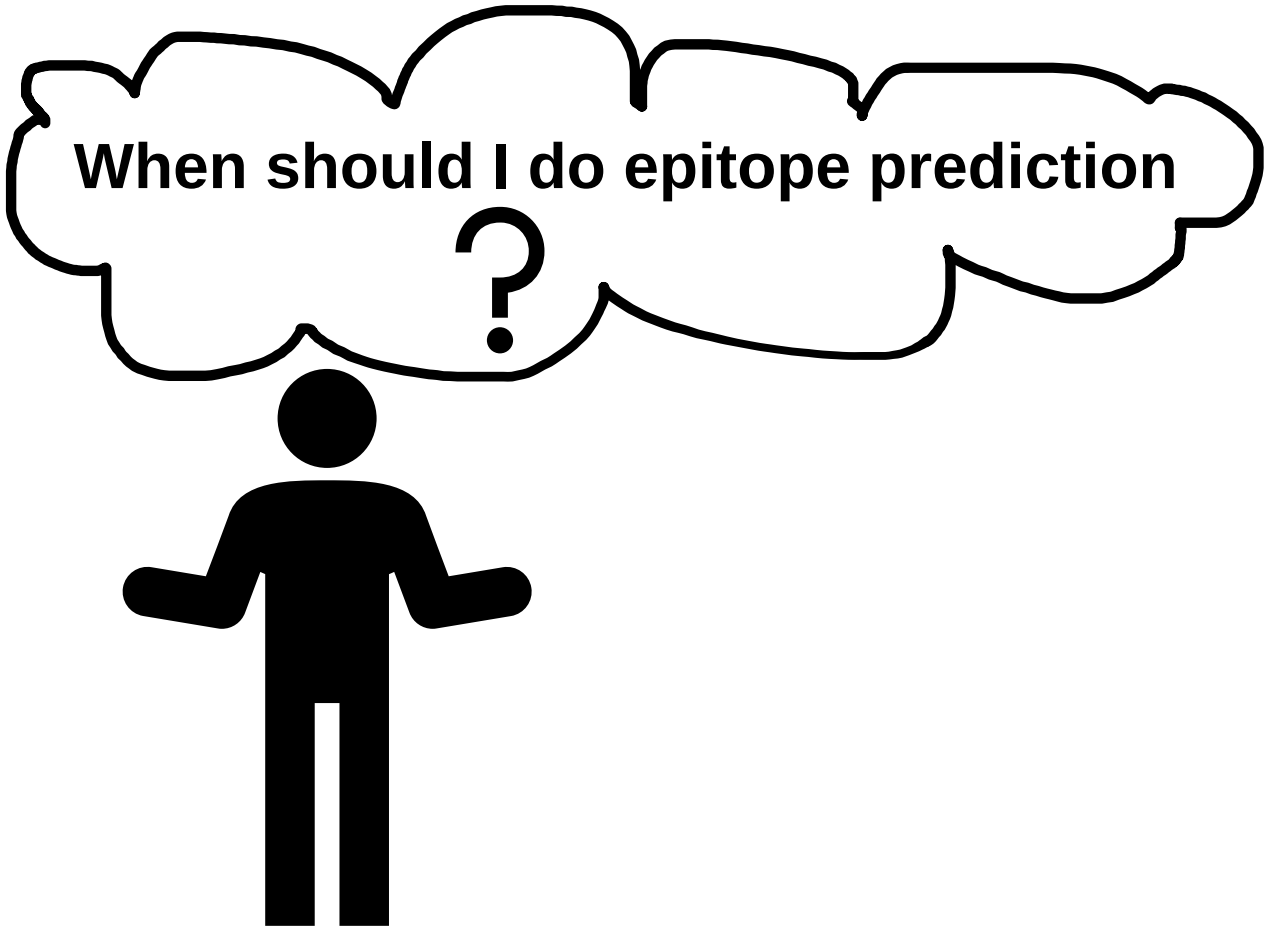
B Cell Epitope Prediction Tools in the IEDB



1. Go to http://tools.iedb.org/main/bcell/

2. Page provides an overview of B cell epitope prediction methods

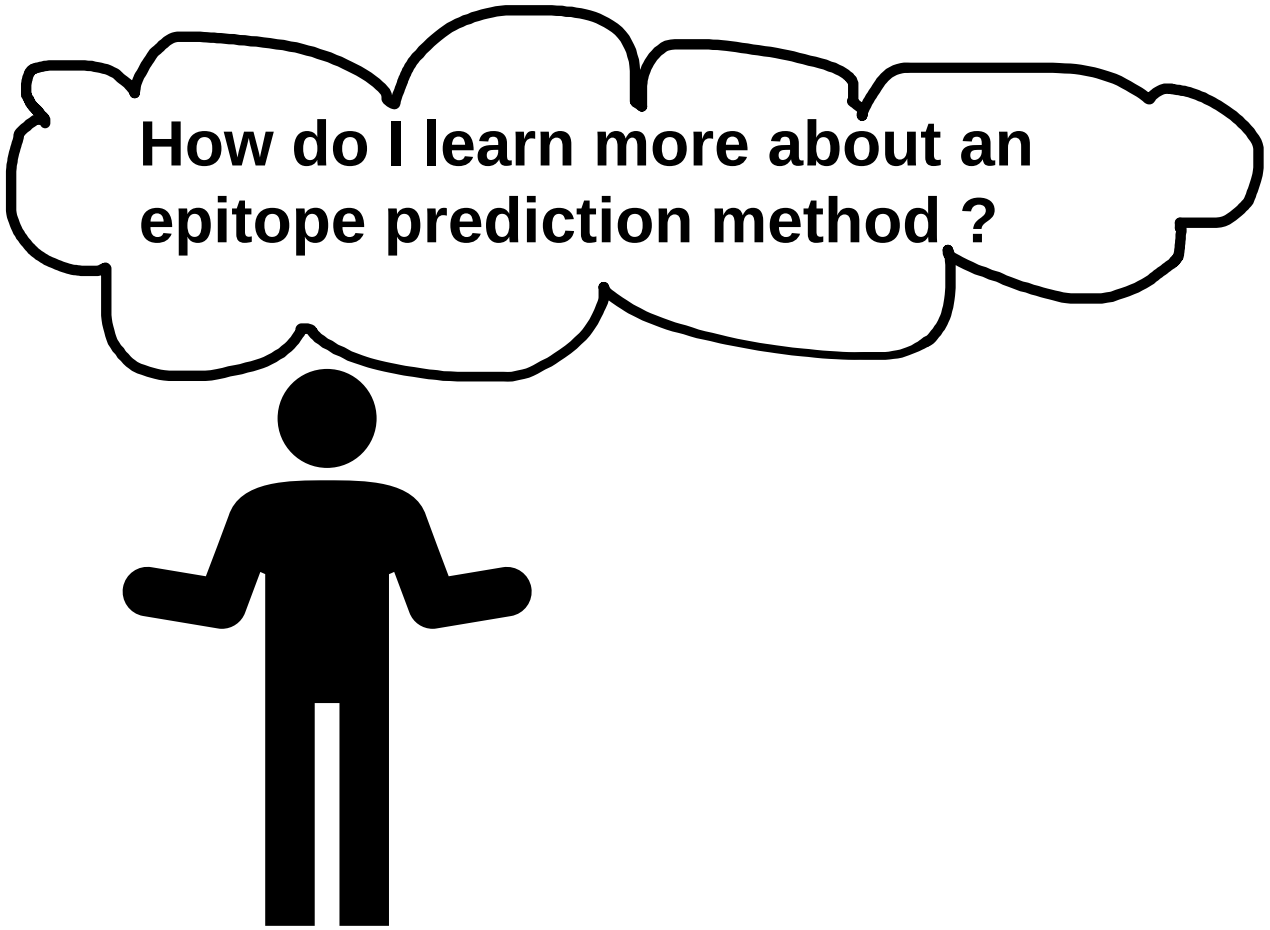
B Cell Epitope Prediction Tools in the IEDB



You have verified thoroughly that **no information is available in the IEDB** on the antigen of your interest

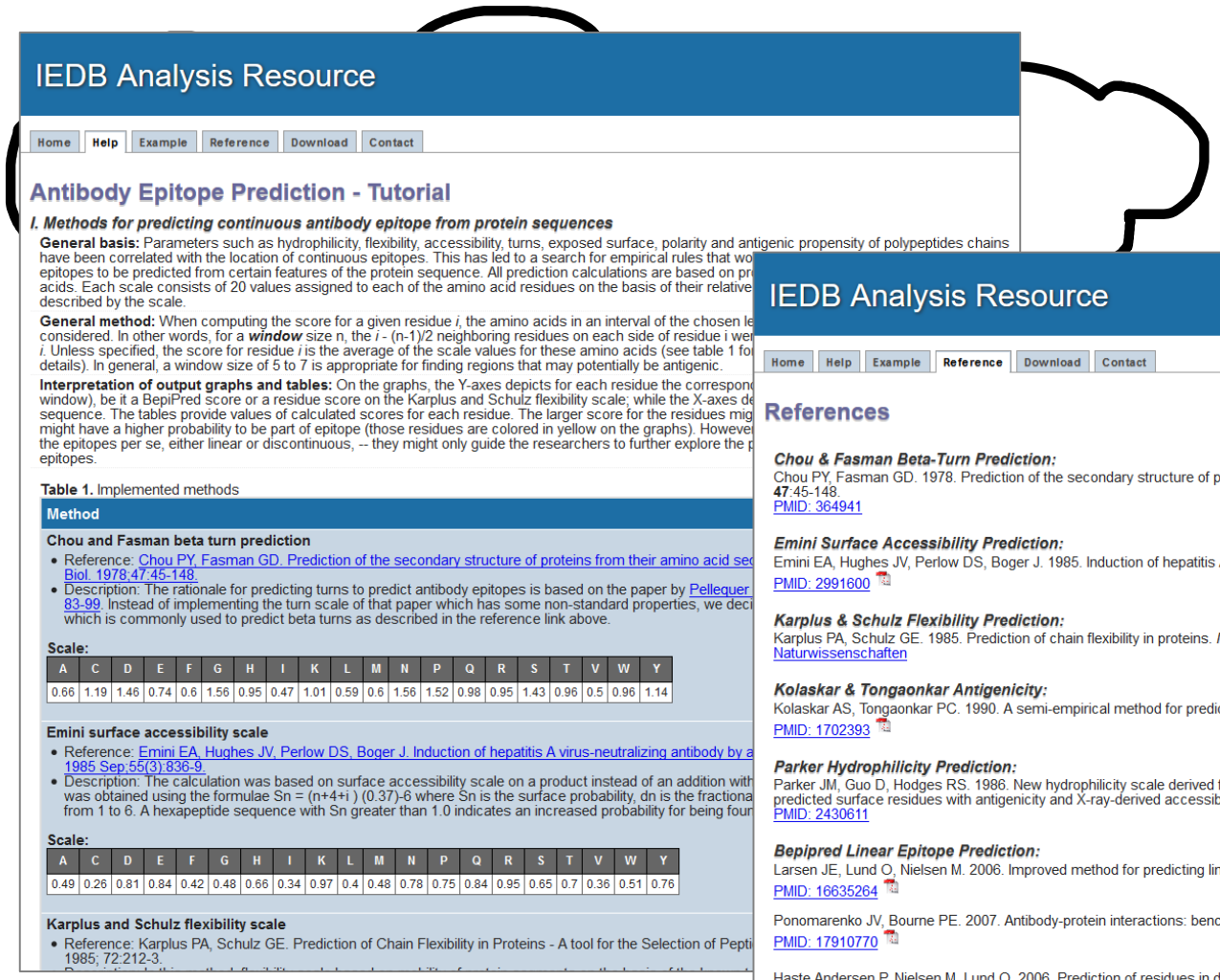
You **want to know all the candidate antigenic determinants** in an antigen of your interest other than epitopes provided in the IEDB

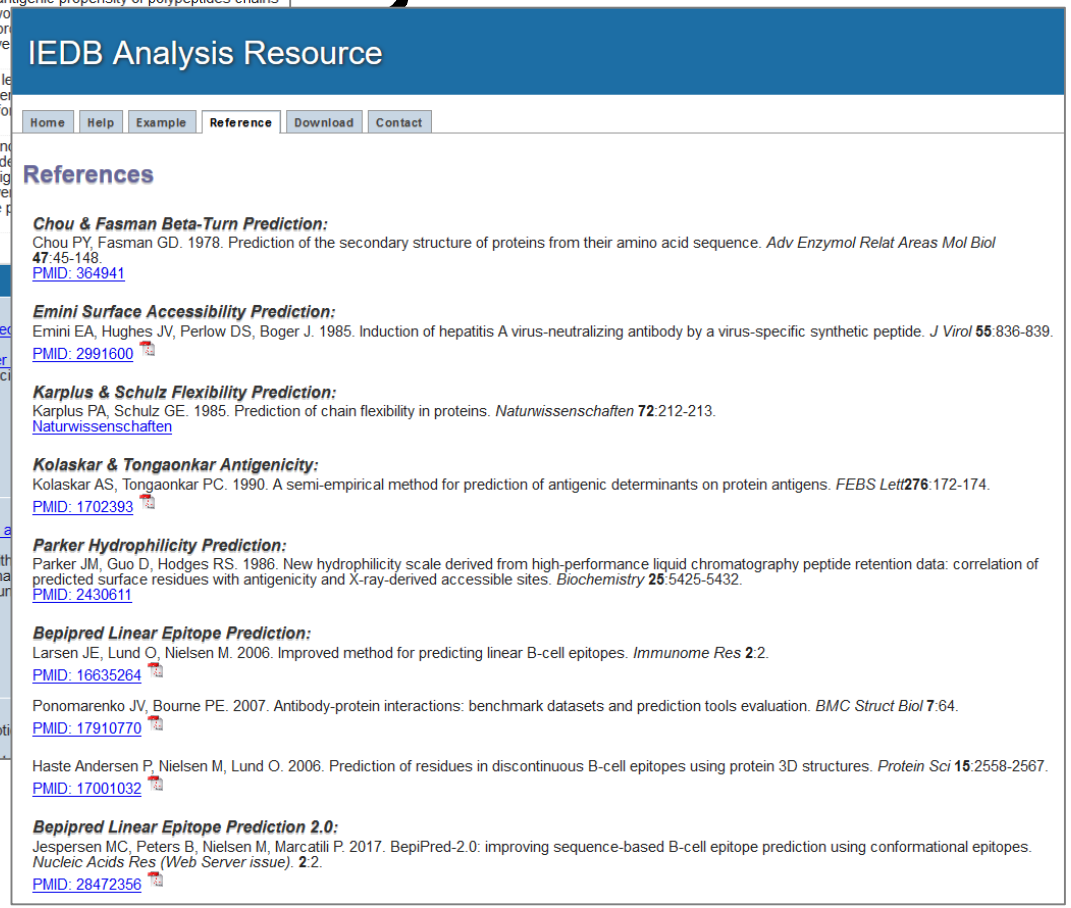
B Cell Epitope Prediction Tools in the IEDB



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

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

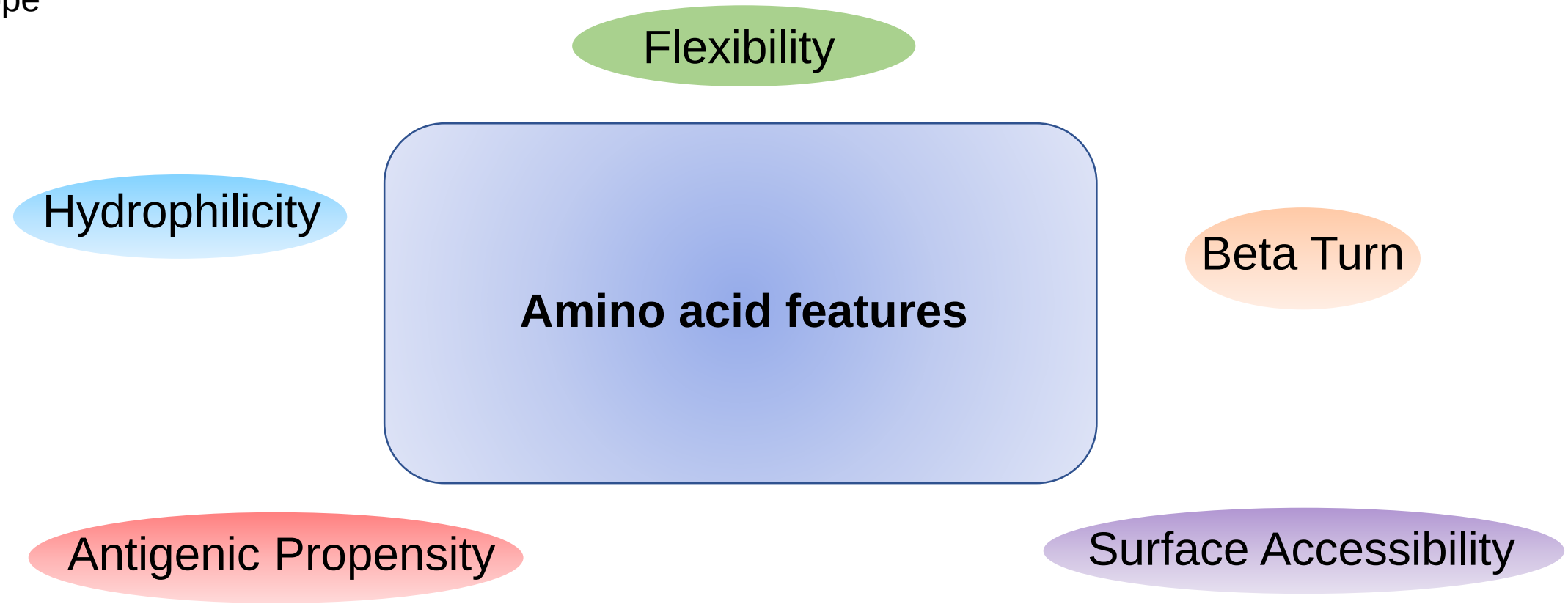
    B. Methods for predicting discontinuous epitopes

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

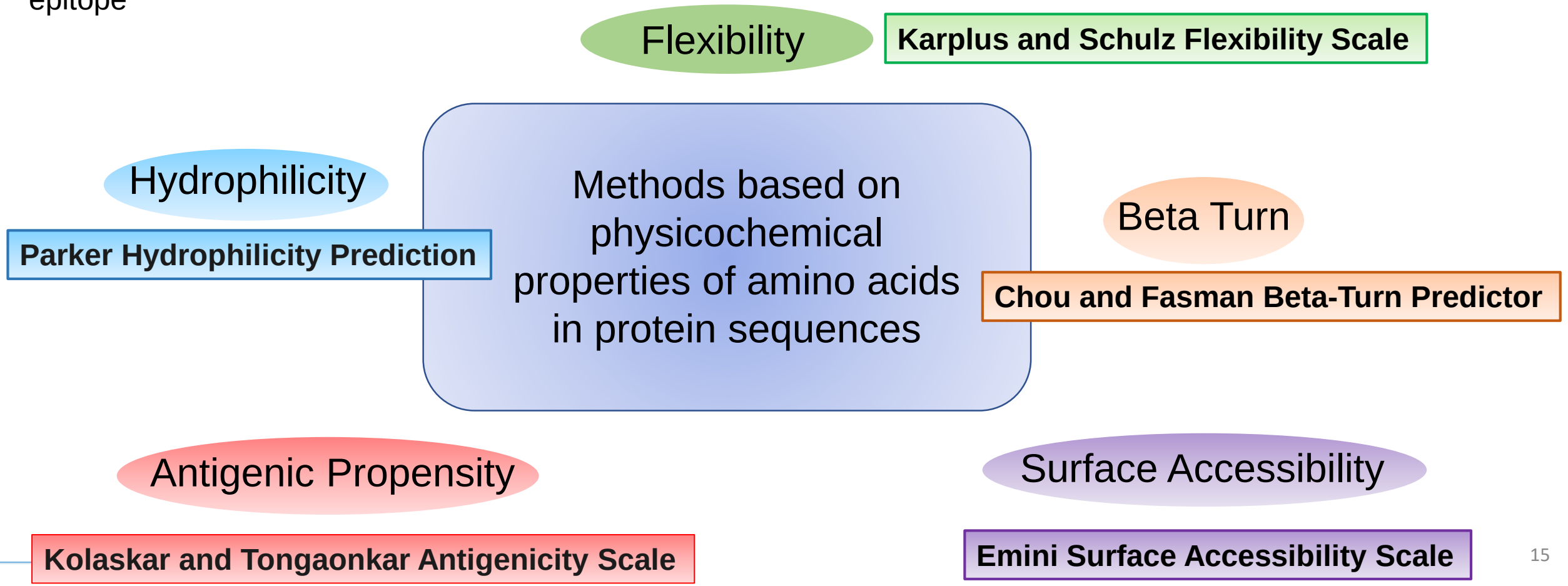


B Cell Epitope Prediction Tools in the IEDB

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



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

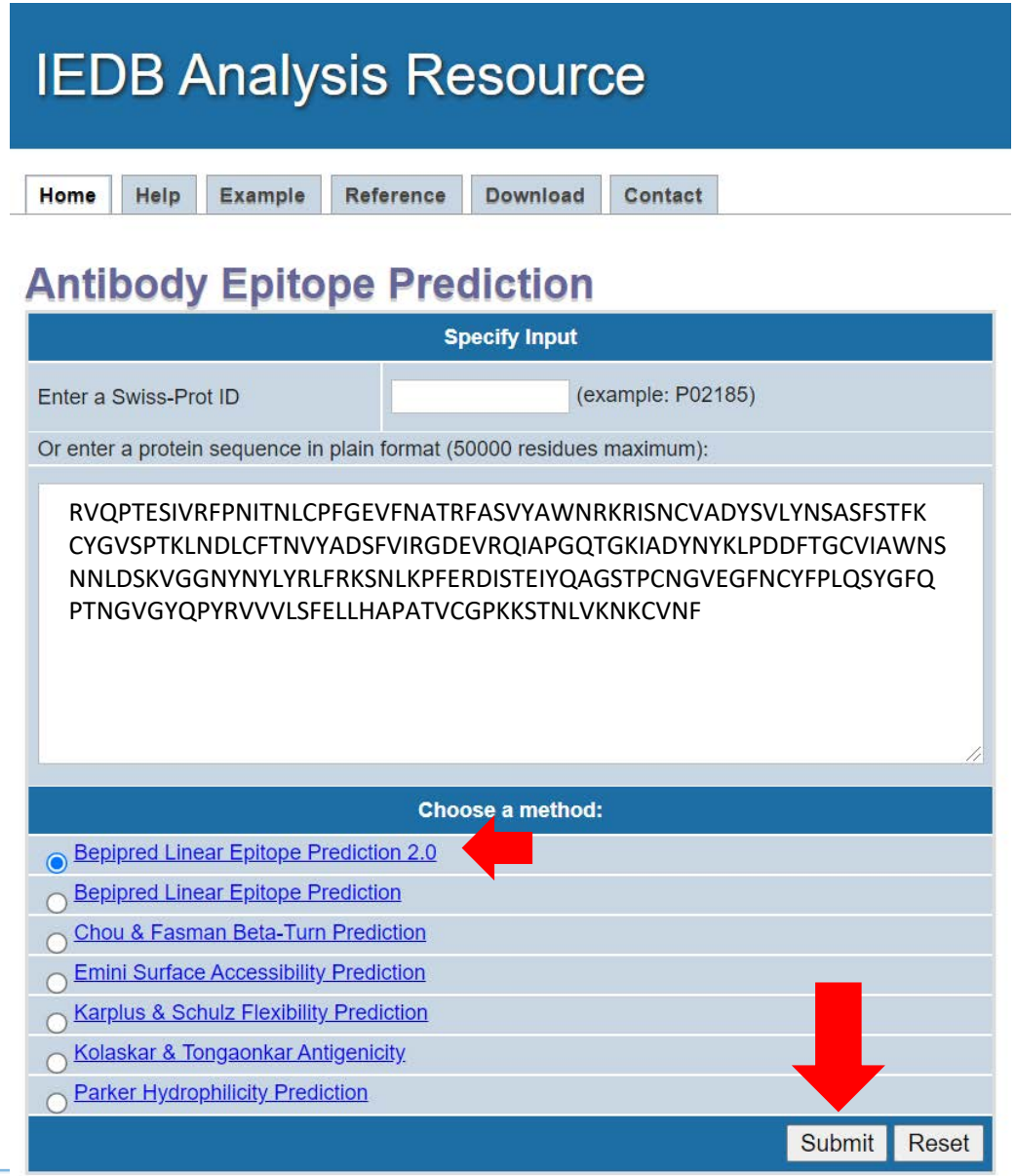
Uses Hidden Markov Models and propensity scale to predict location of linear B cell epitope

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



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

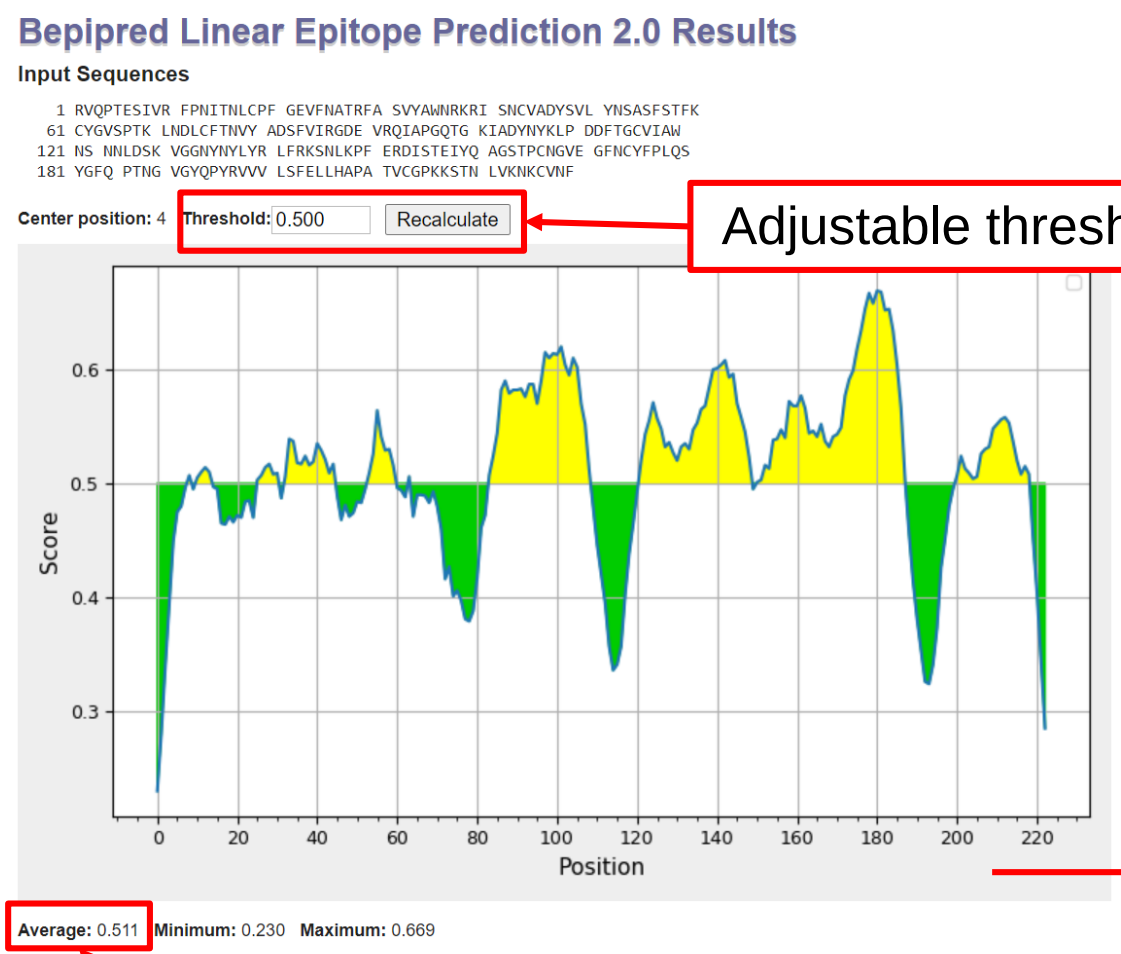
4. Click “Submit”

Linear/Continuous Epitope Prediction Method – An Example

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

Let us now see what linear epitopes the Bepipred-2.0 method predicted on RBD of SARS-CoV-2

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



Adjustable threshold value

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 score of a protein chosen as a threshold by default

Linear/Continuous Epitope Prediction Method – An Example

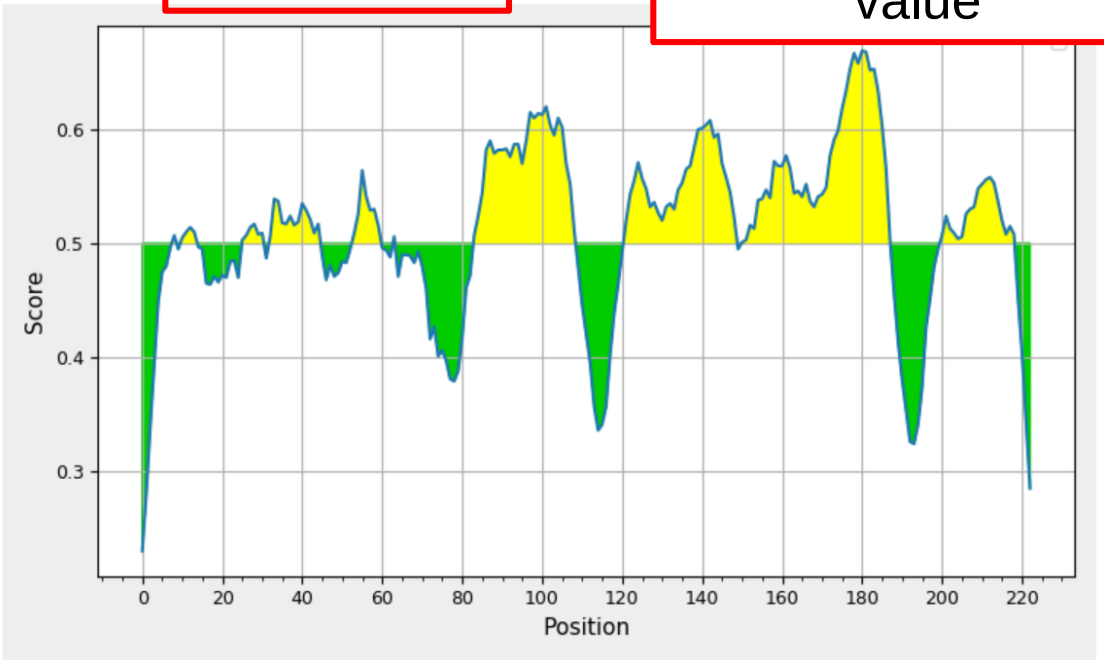
http://tools.iedb.org/bcell/

Bepipred Linear Epitope Prediction 2.0 Results

Input Sequences

```
  1 RVQPTESIVR FPNITNLCPF GEVFNATRFA SVYAWNRKRI SNCVADYSVL YNSASFSTFK
 61 CYGVSPTK LNDLCFTNVY ADSFVIRGDE VRQIAPGQTG KIADYNYKLP DDFTGCVIAW
121 NS NNLDSK VGGNYNYLYR LFRKSNLKPF ERDISTEIYQ AGSTPCNGVE GFNCYFP
181 YGFQ PTNG VGYQPYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
```

Center position: 4 **Threshold:** 0.500 Recalculate



Average: 0.511 **Minimum:** 0.230 **Maximum:** 0.669

Adjustable threshold value

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	YAWNRKRISNCVA	13
5	54	60	ASFSTFK	7
6	64	64	V	1
7	84	109	IRGDEVRQIAPGQTGKIADYNYKLPD	26
8	122	149	NLDSKVGGNYNYLYRLFRKSNLKPFERD	28
9	151	188	STEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQ	38
10	201	219	HAPATVCGPKKSTNLVKNK	19

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

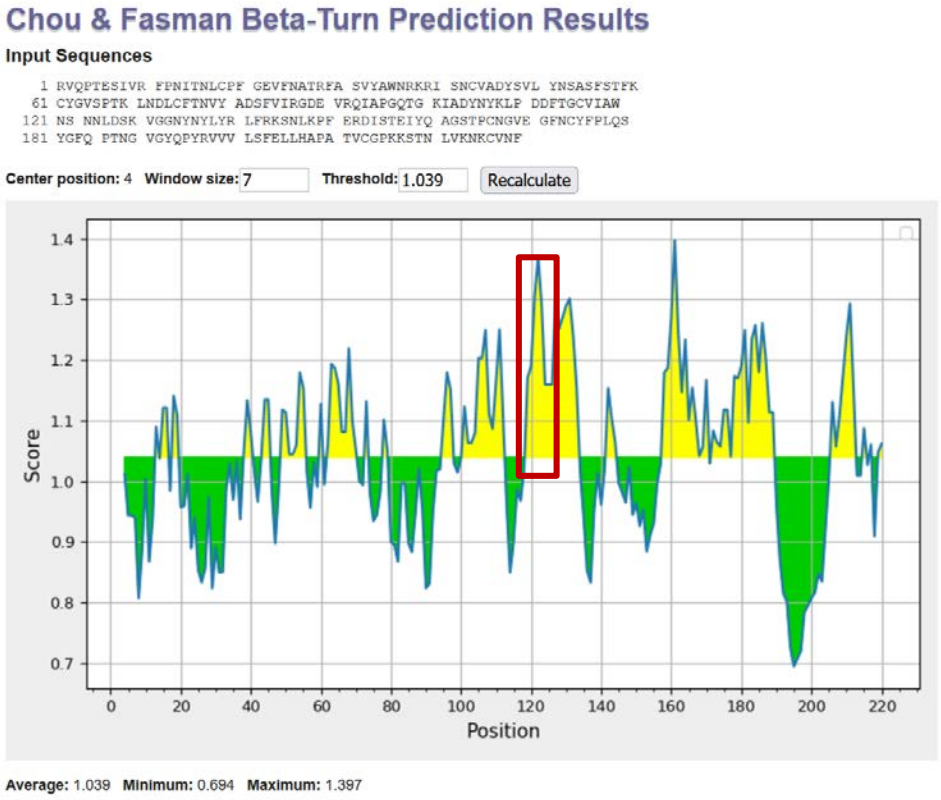
Predicted residue scores:

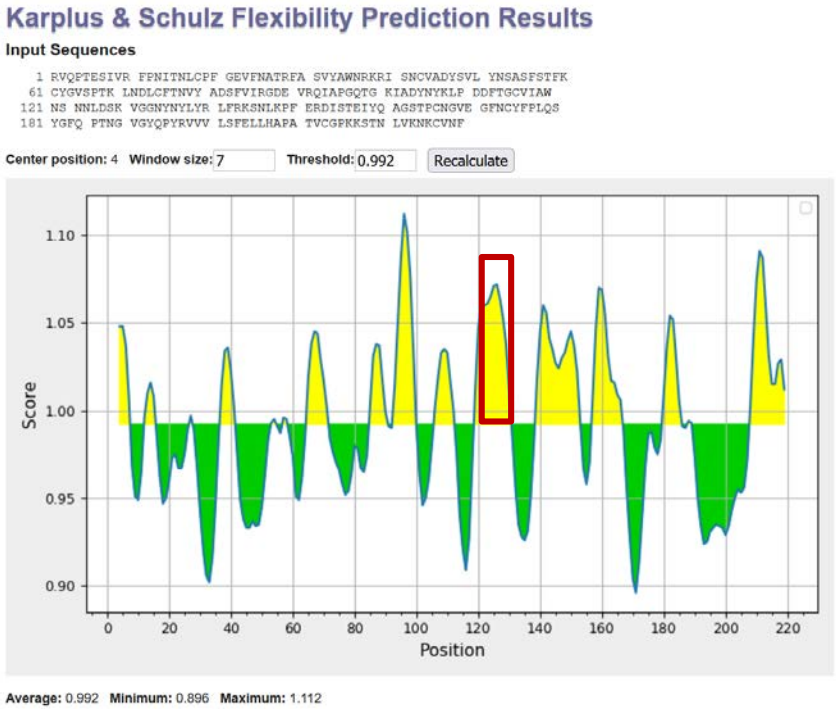
Position	Residue	Score	Assignment
0	R	0.230	.
1	V	0.280	.
2	Q	0.339	.
3	P	0.394	.
4	T	0.449	.
5	E	0.475	.
6	S	0.480	.
7	I	0.496	.
8	V	0.507	E
9	R	0.495	.
10	F	0.505	E
11	P	0.510	E
12	N	0.514	E
13	I	0.510	E
14	T	0.497	.
15	N	0.495	.
16	L	0.465	.
17	C	0.464	.
18	P	0.471	.
19	F	0.466	.

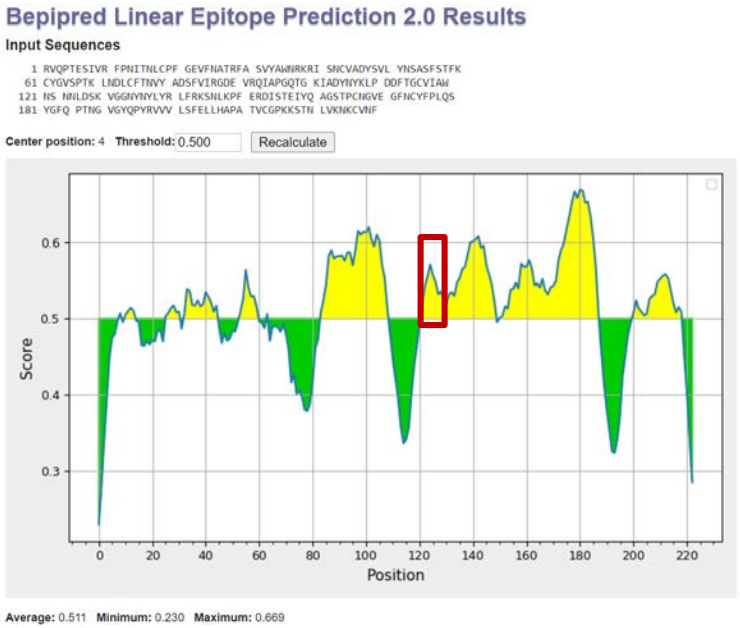
Results: a table of residues and their predicted scores
“E” = epitope

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







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

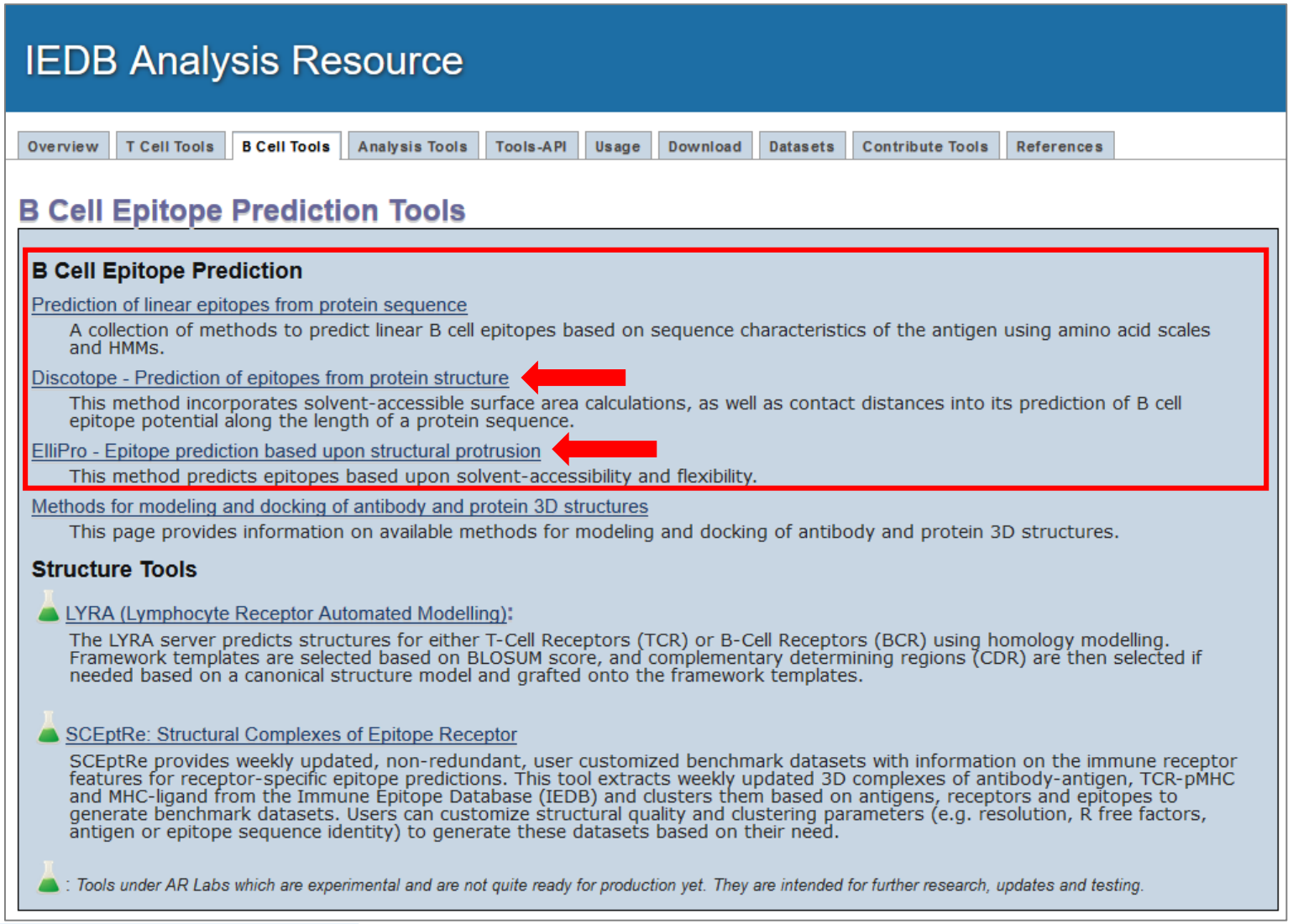
2. **B cell epitope prediction methods in the IEDB**

    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/

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

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[1], Nielsen M, Lund O.

PMID: 17001032 PMCID: PMC2242418 DOI: 10.1110/ps.062405906

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

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[1], 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

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

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.

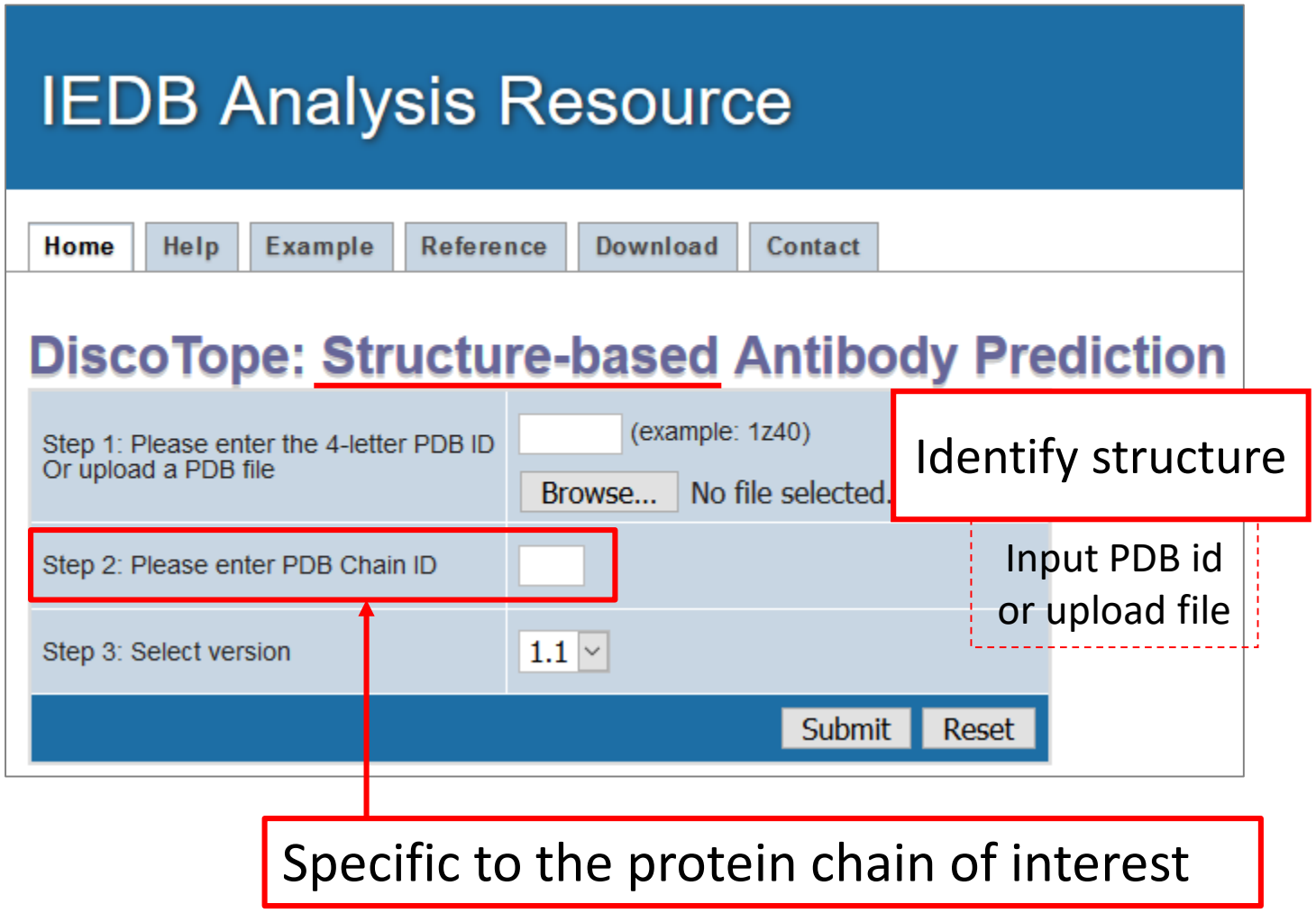
Kringelum JV[1], Lundegaard C, Lund O, Nielsen M.

PMID: 23300419 PMCID: PMC3531324 DOI: 10.1371/journal.pcbi.1002829

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



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

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

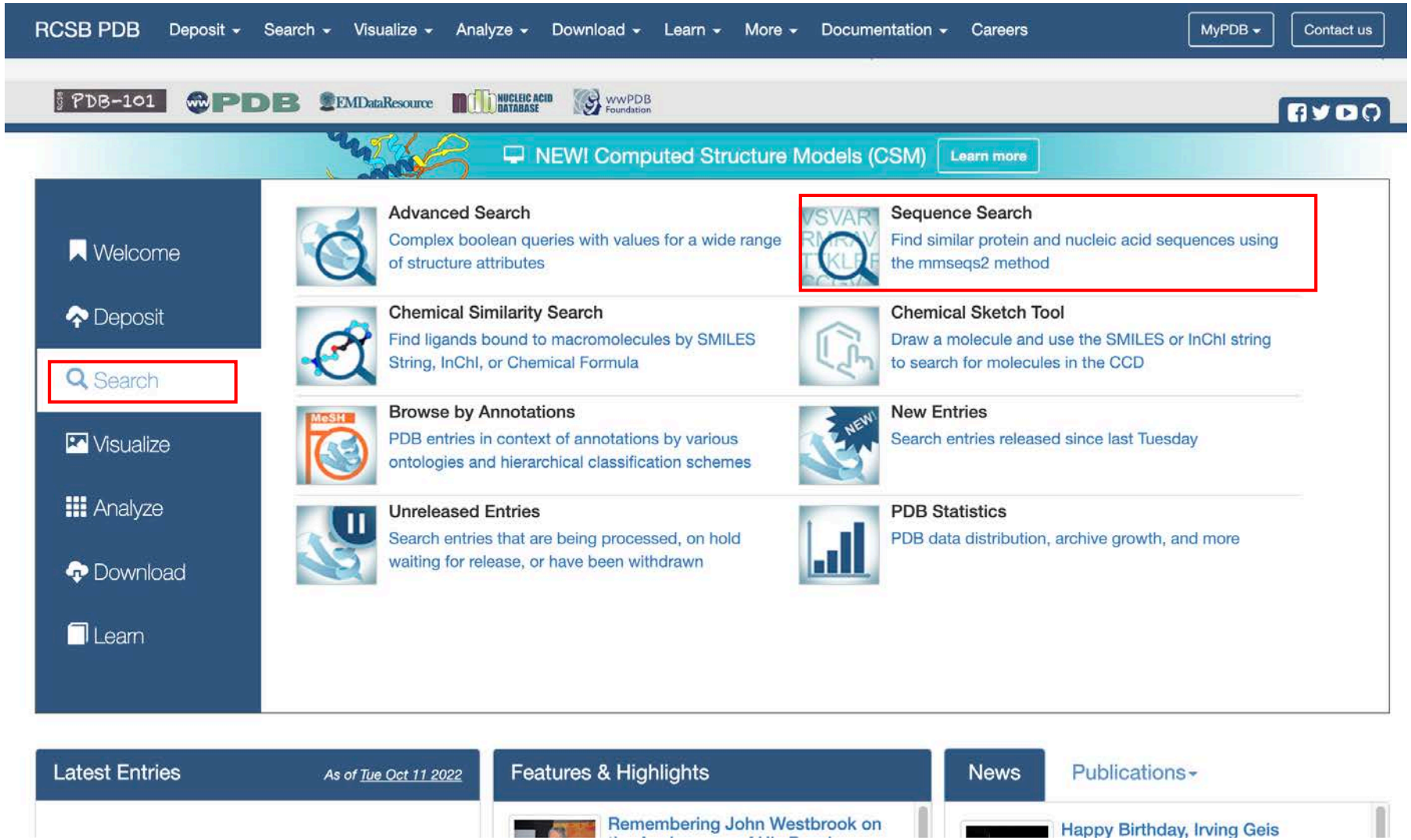
Using its PDB ID

OR

Uploading the PDB file

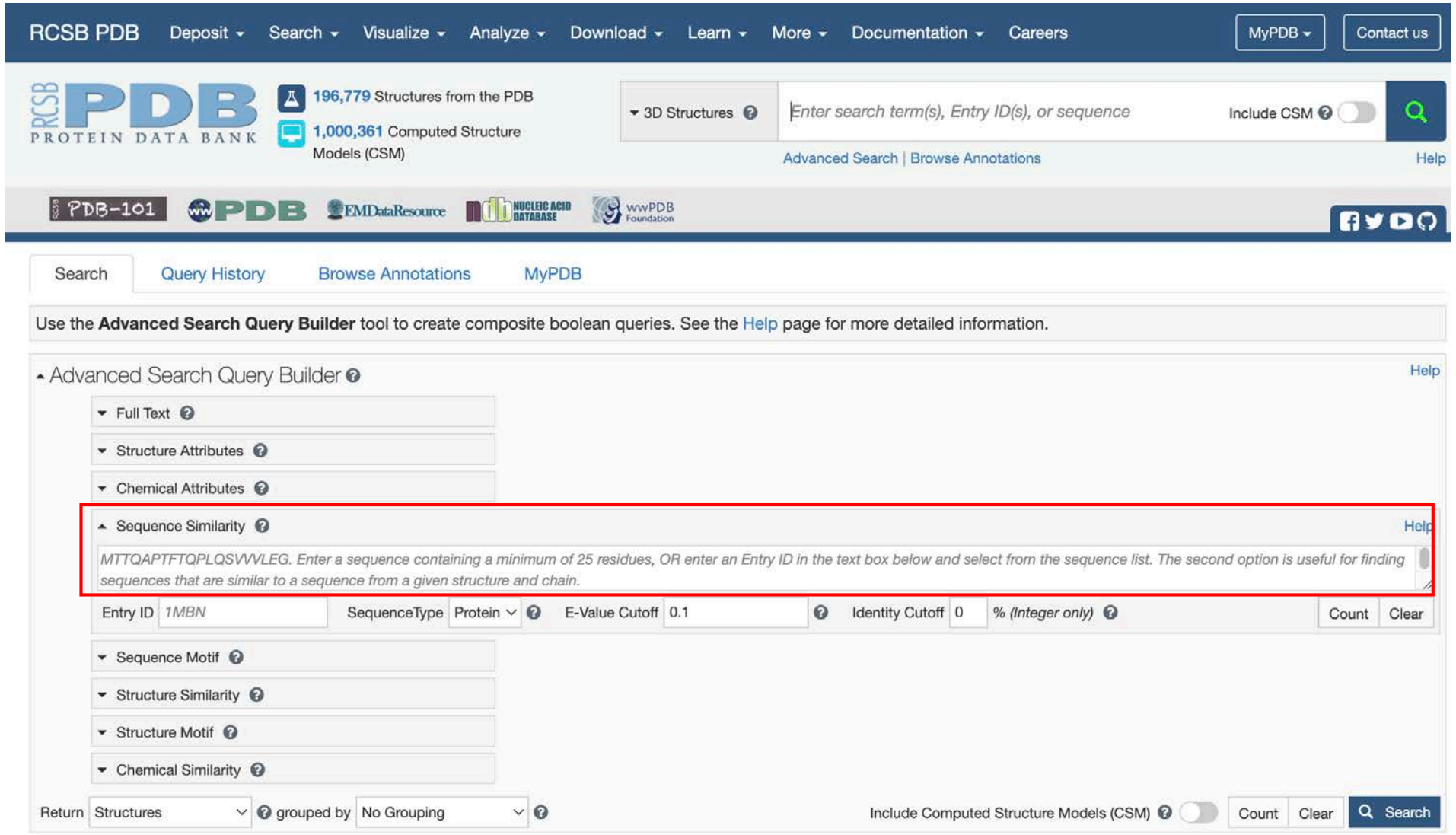
3. PDB structures have multiple chains. Choose a chain of interest by specifying its chain ID.

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



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

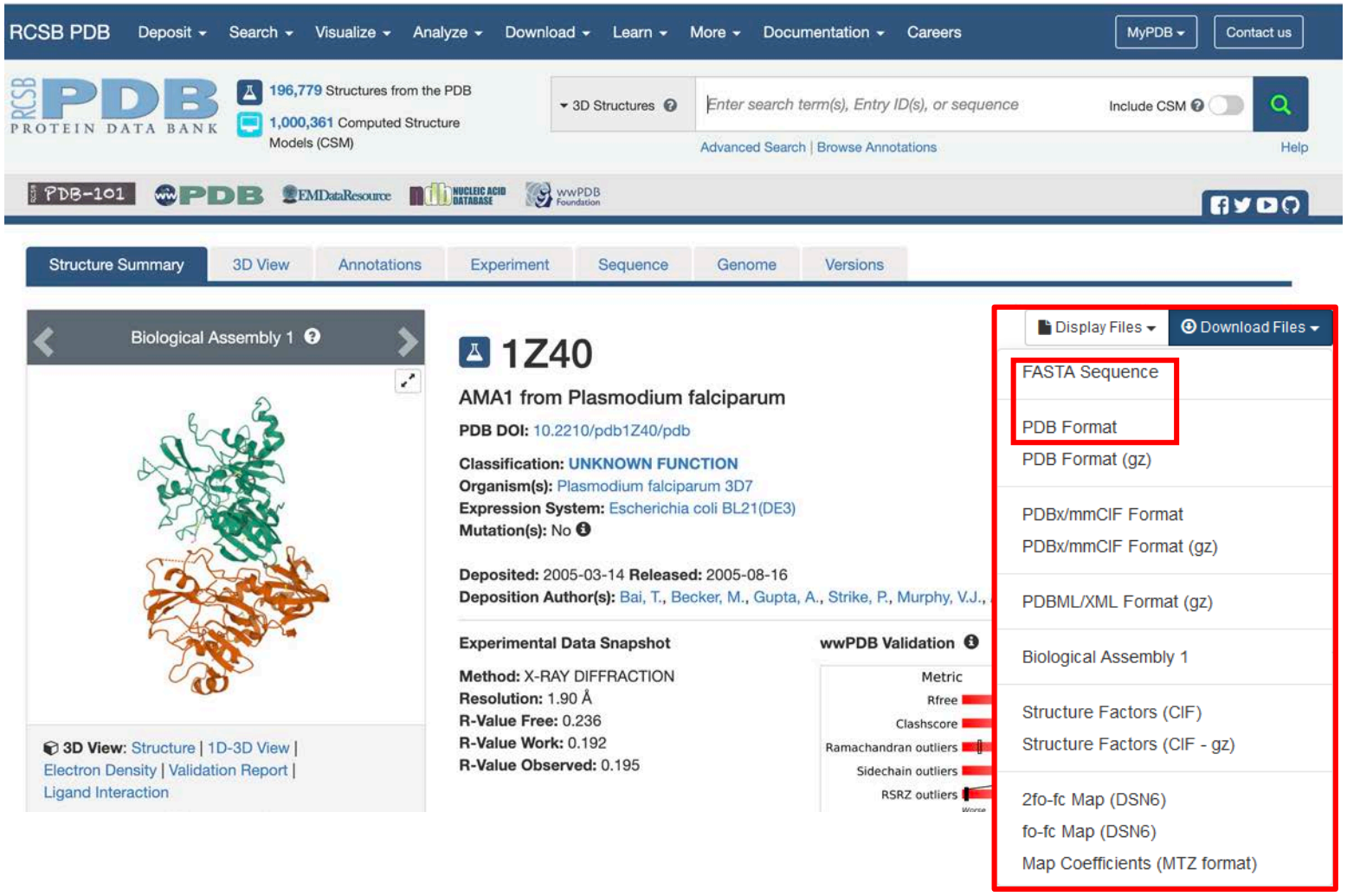
How to Search for 3D Structures in the Protein Data Bank?



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?



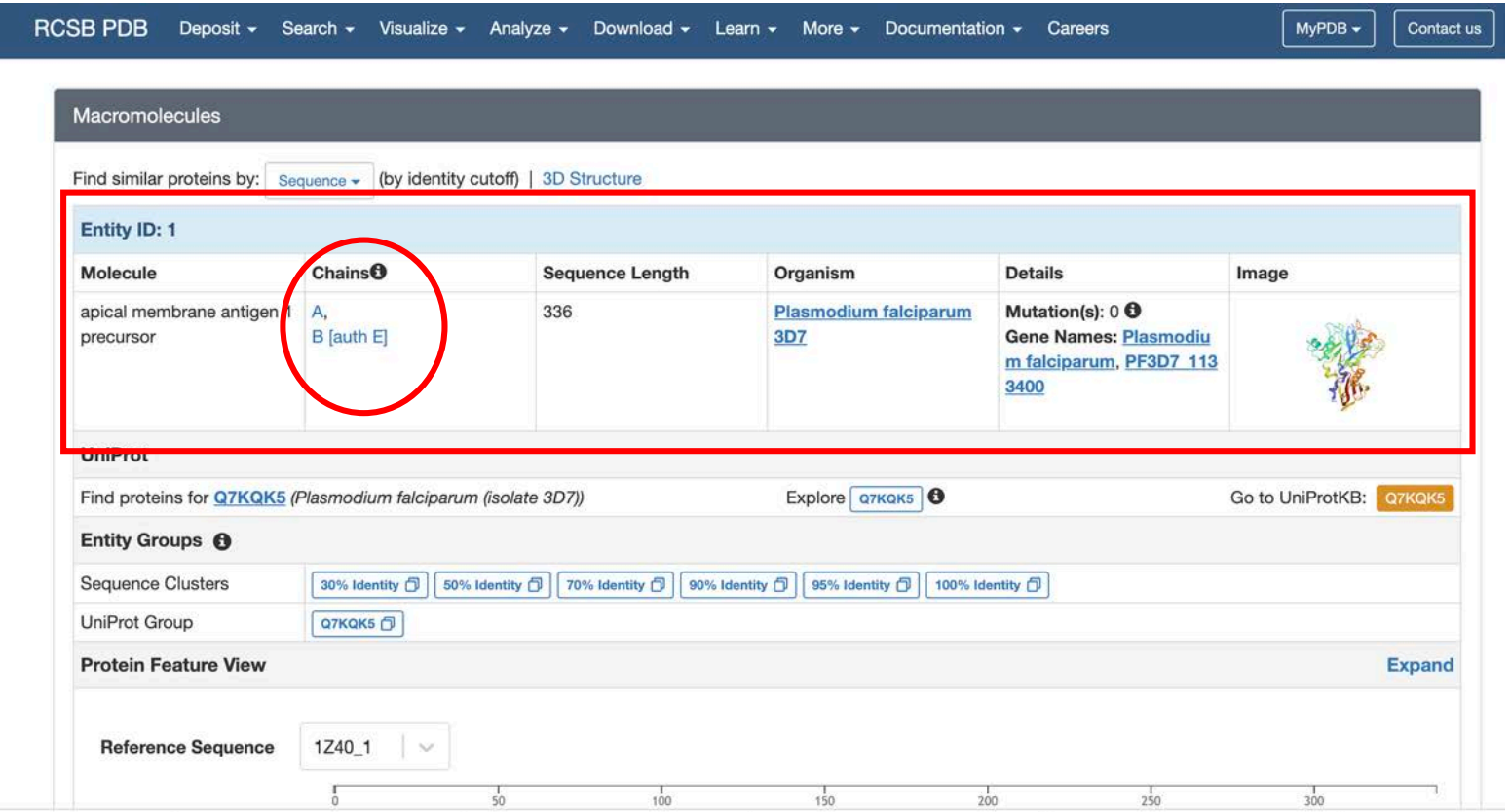
6. It will take us to a page containing details of the structure of the AMA1 protein
PDB ID : 1Z40

The 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?

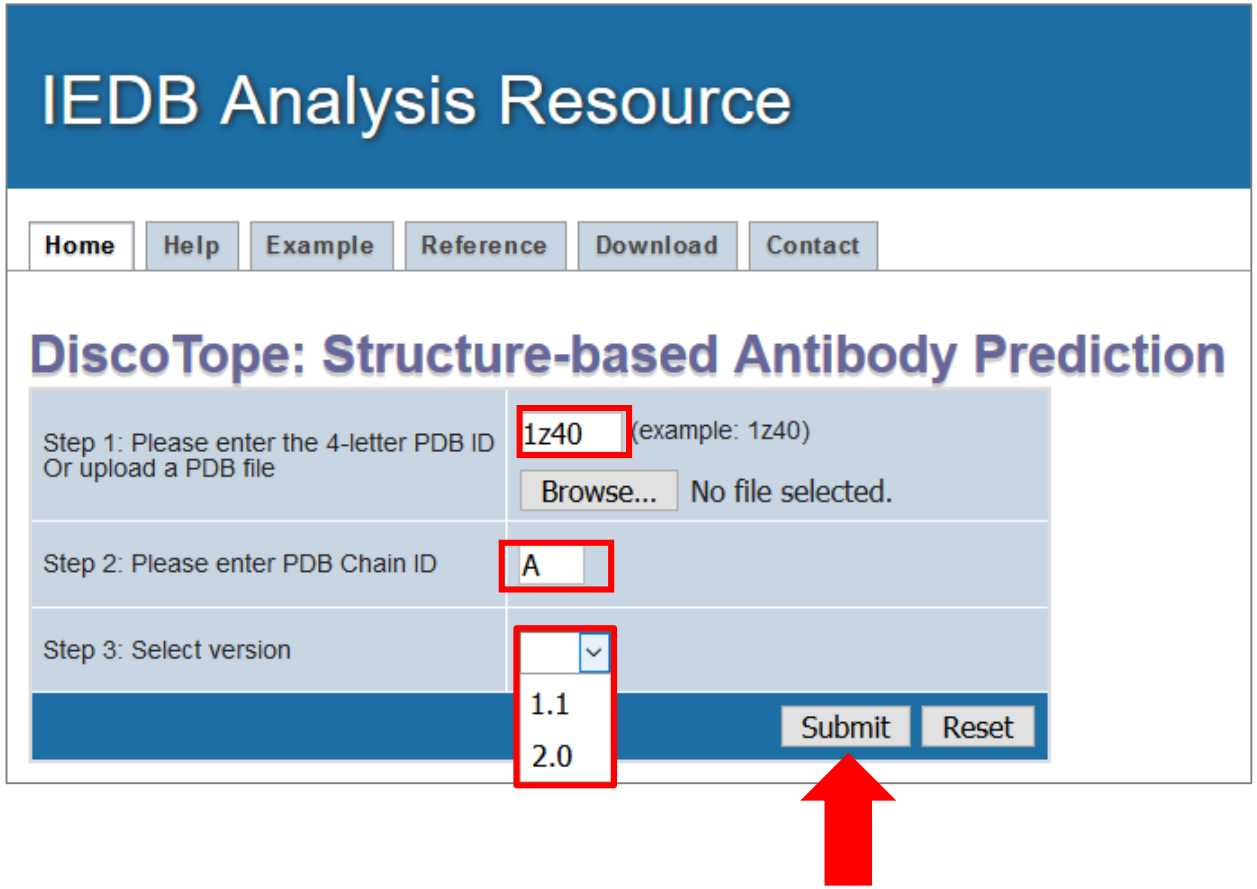


9. There are two units of the AMA1 protein in the crystal structure, one with chain id 'A' and the other with 'B/E"

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



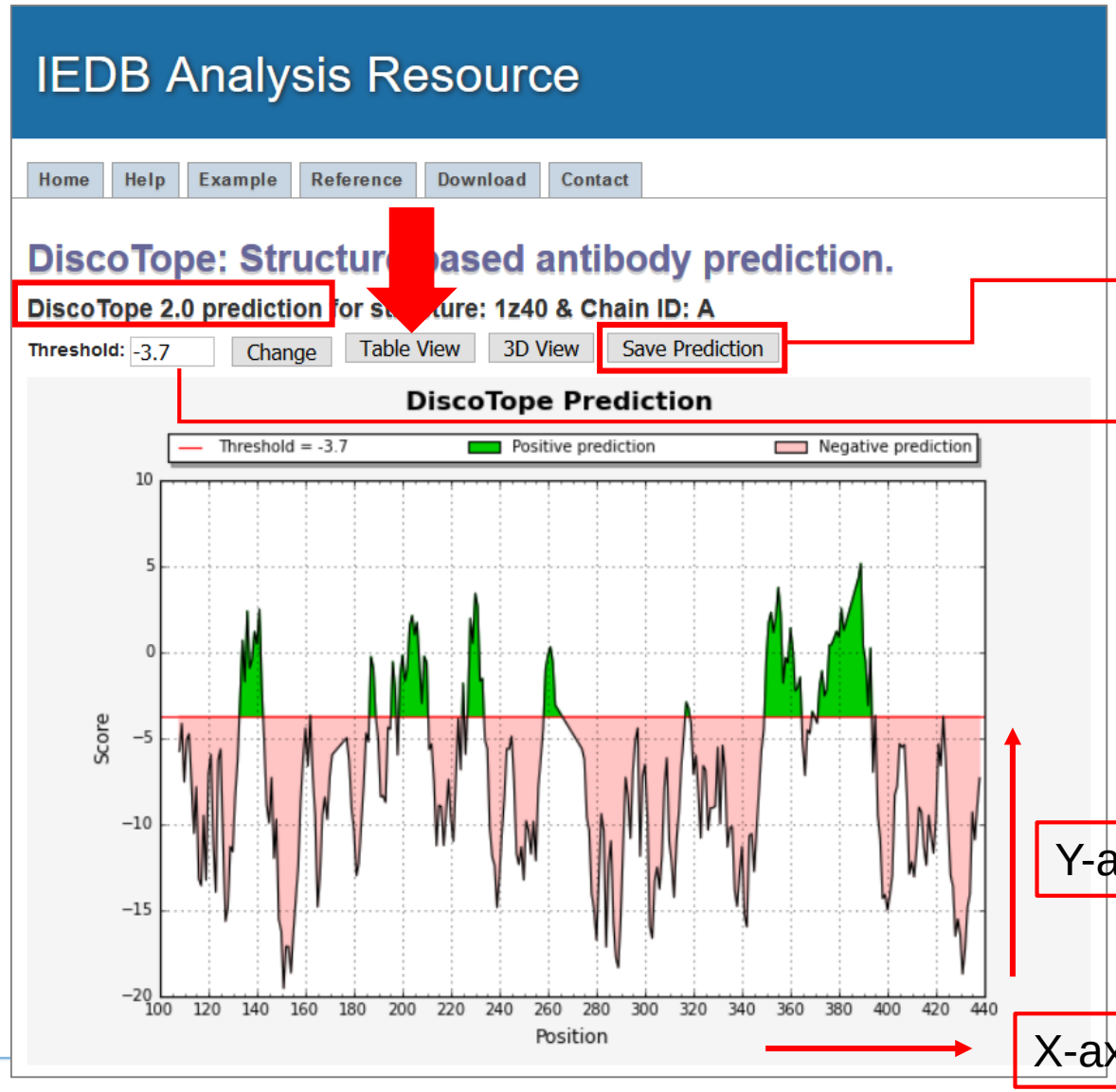
1. Enter in the PDB ID of AMA1 protein (1z40) and the chain ID ('A')

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

Y-axis: DiscoTope Score

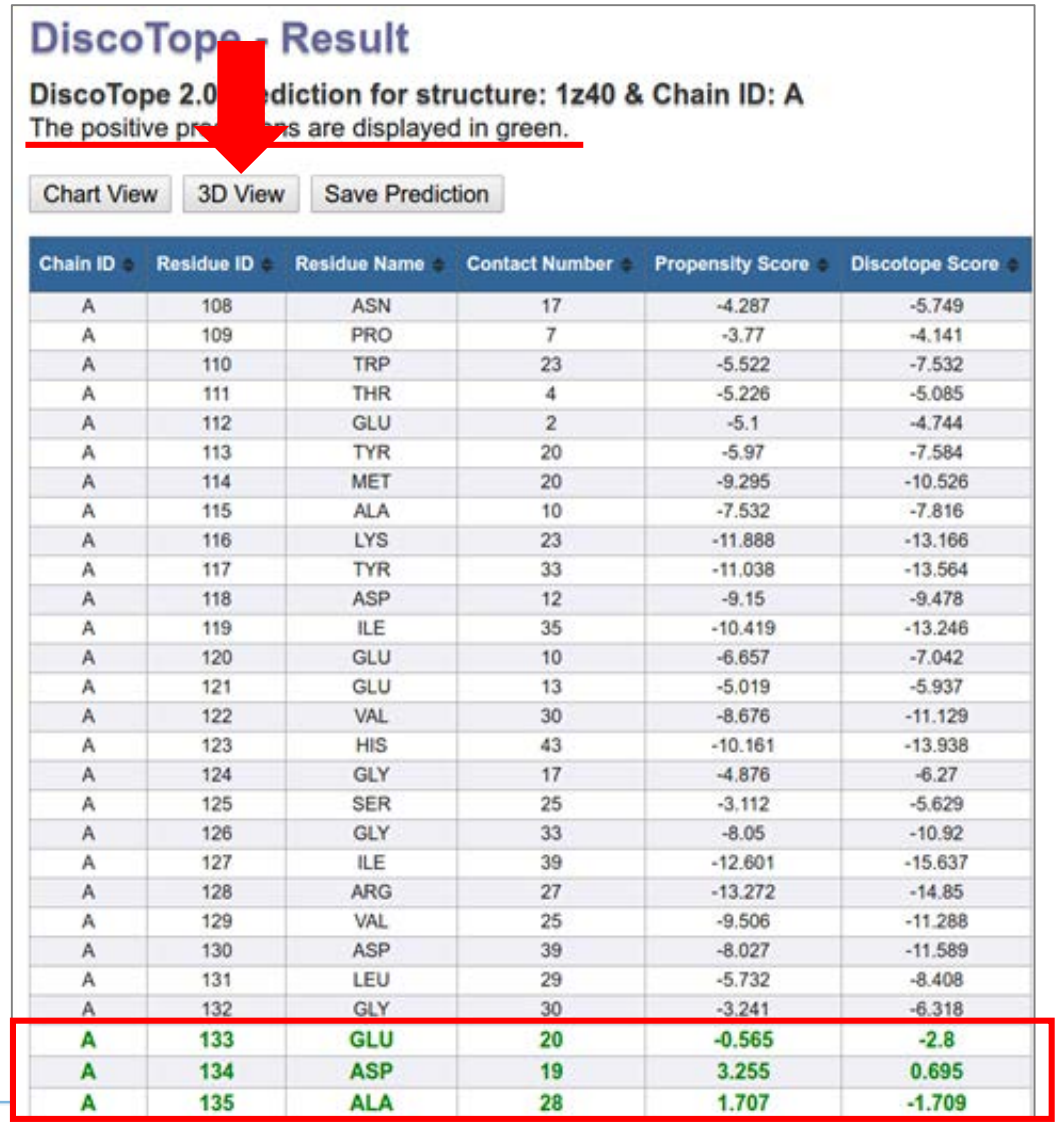
X-axis: Residue Position

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



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**

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

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

Predicts linear and discontinuous antibody epitopes based 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

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

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

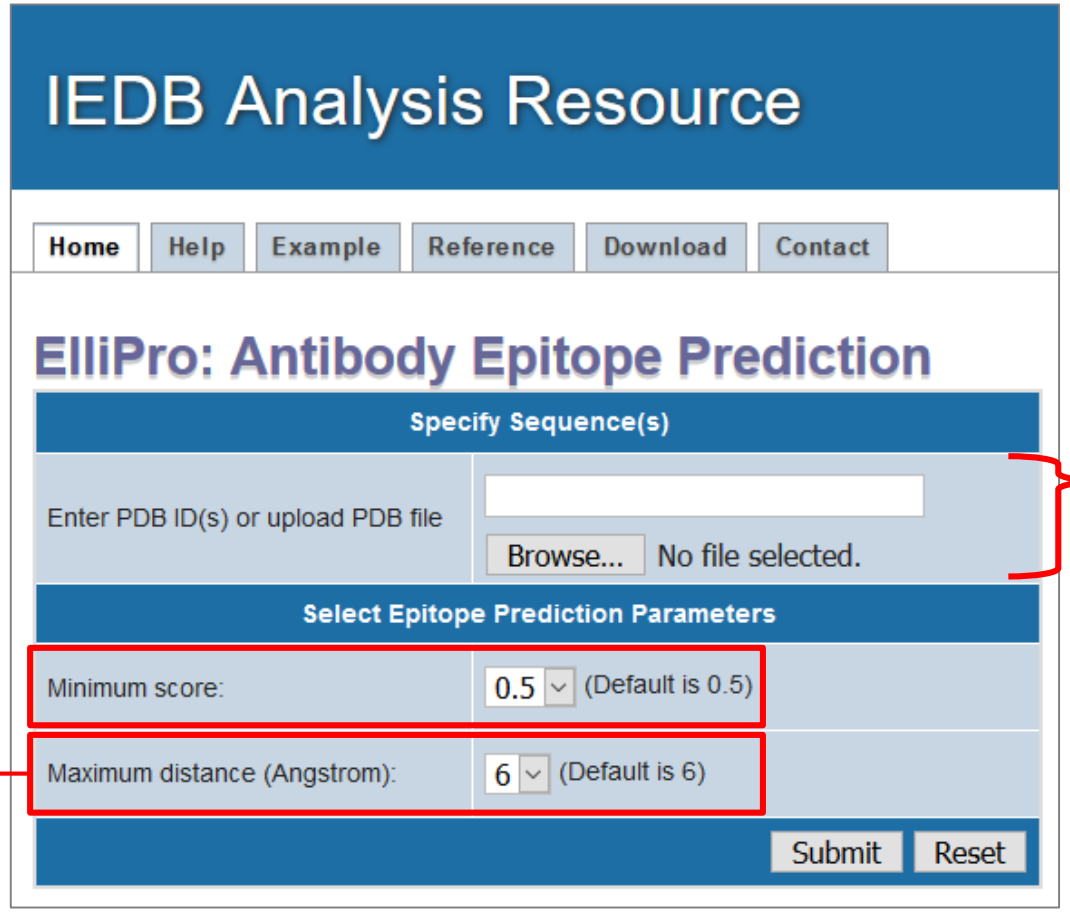
Location of 'continuous' antigenic determinants in the protruding regions of proteins.

Thornton JM, Edwards MS, Taylor WR, Barlow DJ.

PMID: 2423325 PMCID: PMC1166746

B Cell Epitope Prediction Tools in the IEDB

Using ElliPro to Predict Discontinuous Epitopes



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

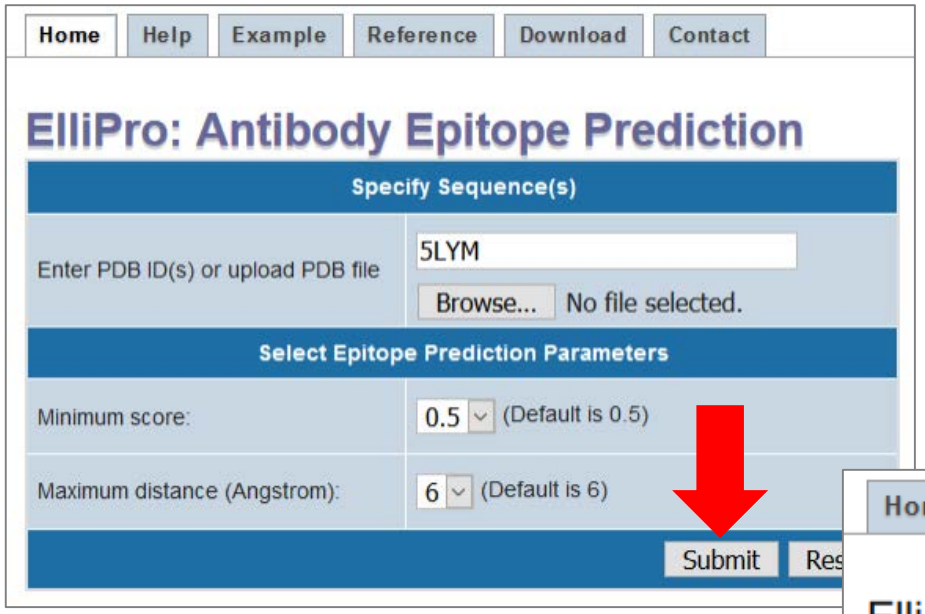
- Averaged over epitope residues
- Higher scores predict fewer epitopes

Specify max distance for predicting (grouping) discontinuous epitopes

Longer distances predict discontinuous epitopes spanning larger regions

B Cell Epitope Prediction Tools in the IEDB

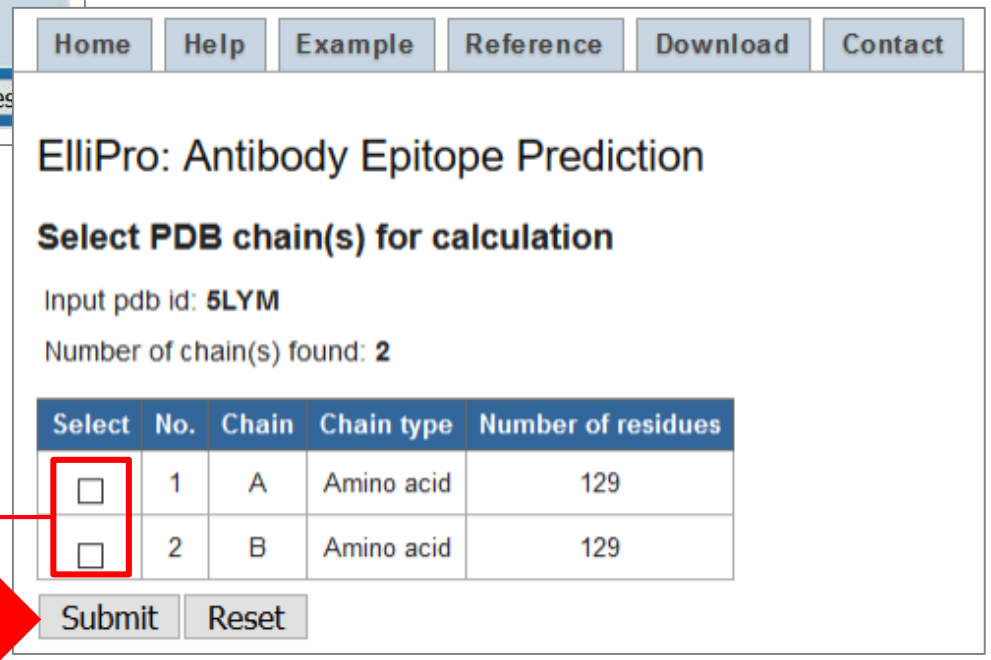
Using ElliPro to Predict Discontinuous Epitopes



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

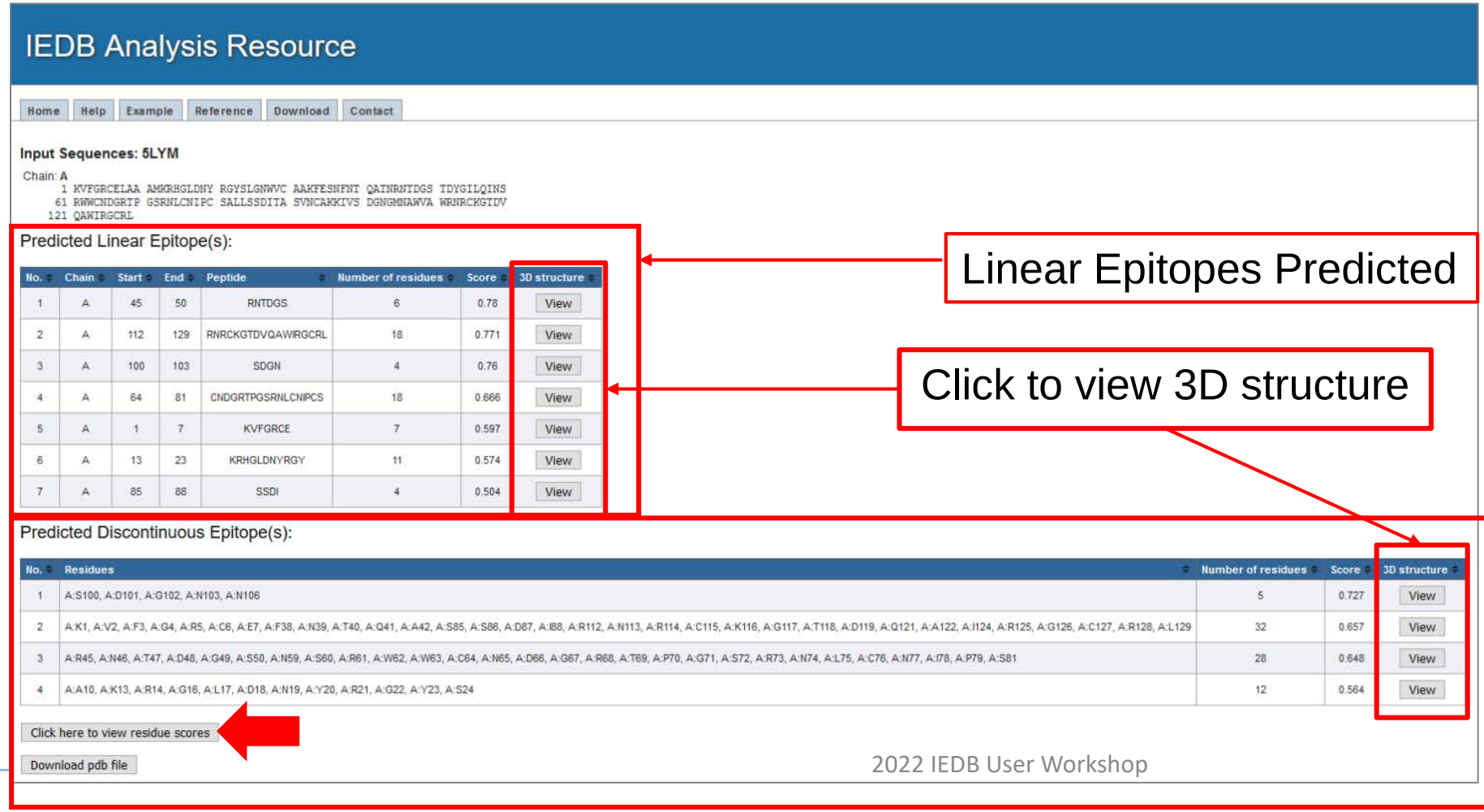


Select chain(s) of interest

B Cell Epitope Prediction Tools in the IEDB

Using ElliPro to Predict Discontinuous Epitopes

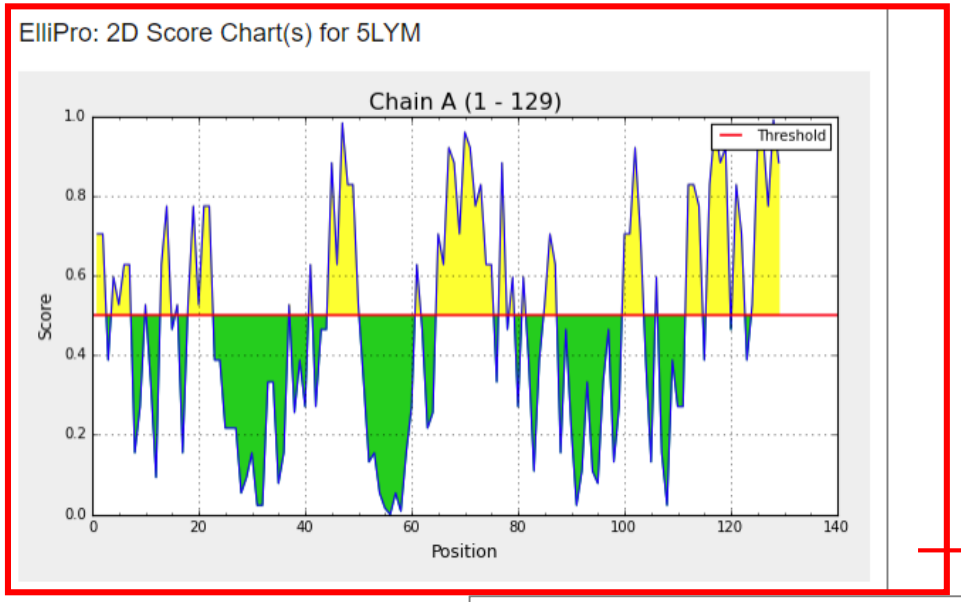
ElliPro predicts both linear and discontinuous epitopes



For each predicted epitope, ElliPro provides a score

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!