B Cell Epitope Prediction

tools.iedb.org

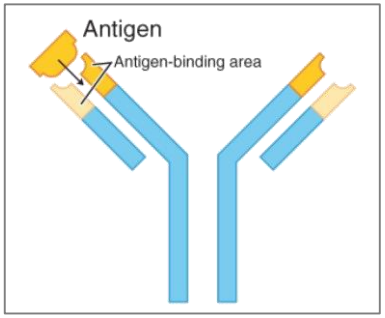
Presented by: Marcus Mendes, Postdoc Researcher

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



Antigen
Hemagglutinin HA1 Chain
from Influenza A Virus

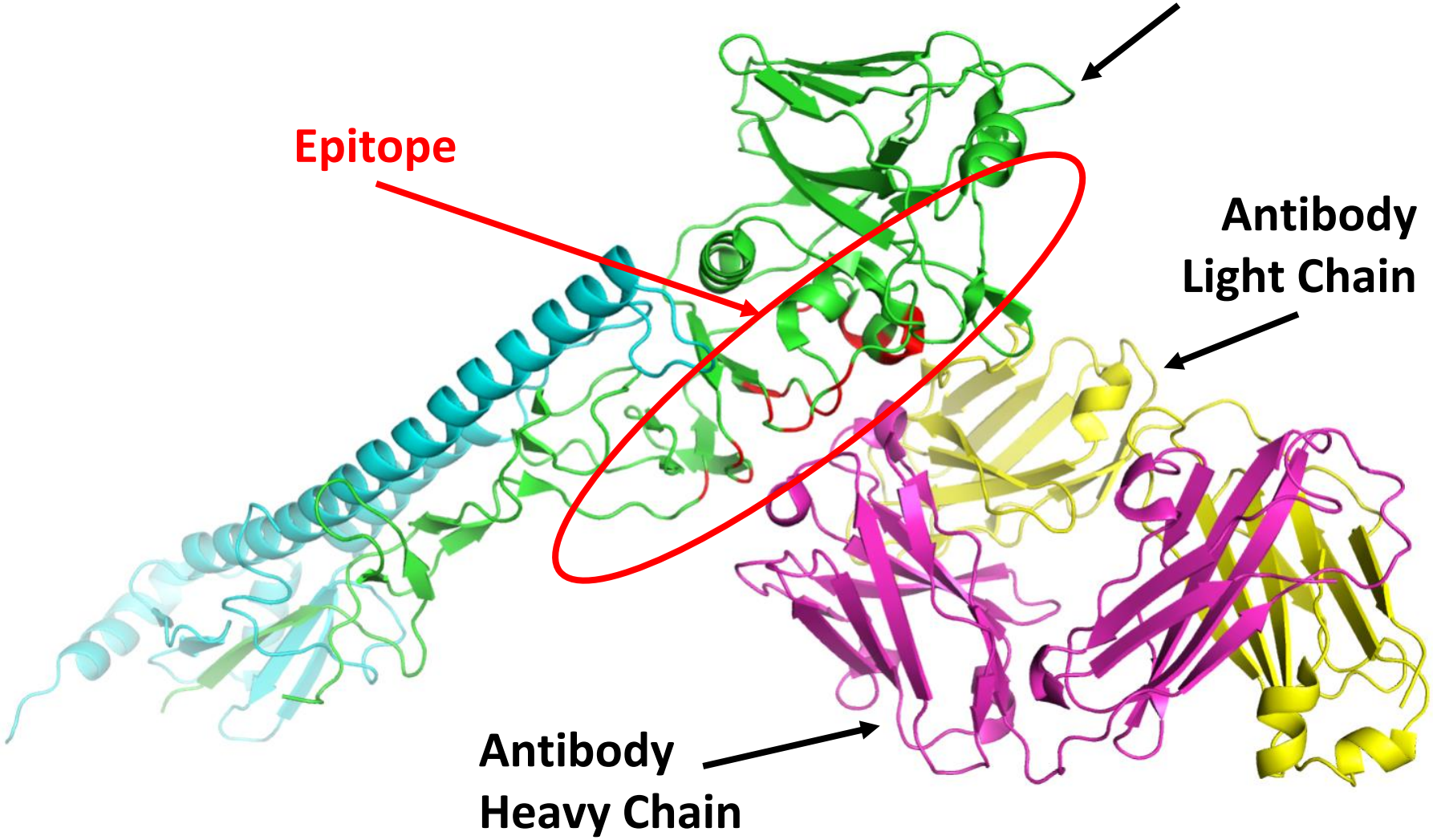
Antibody Light Chain

Hemagglutinin HA2 Chain

Antibody Heavy Chain

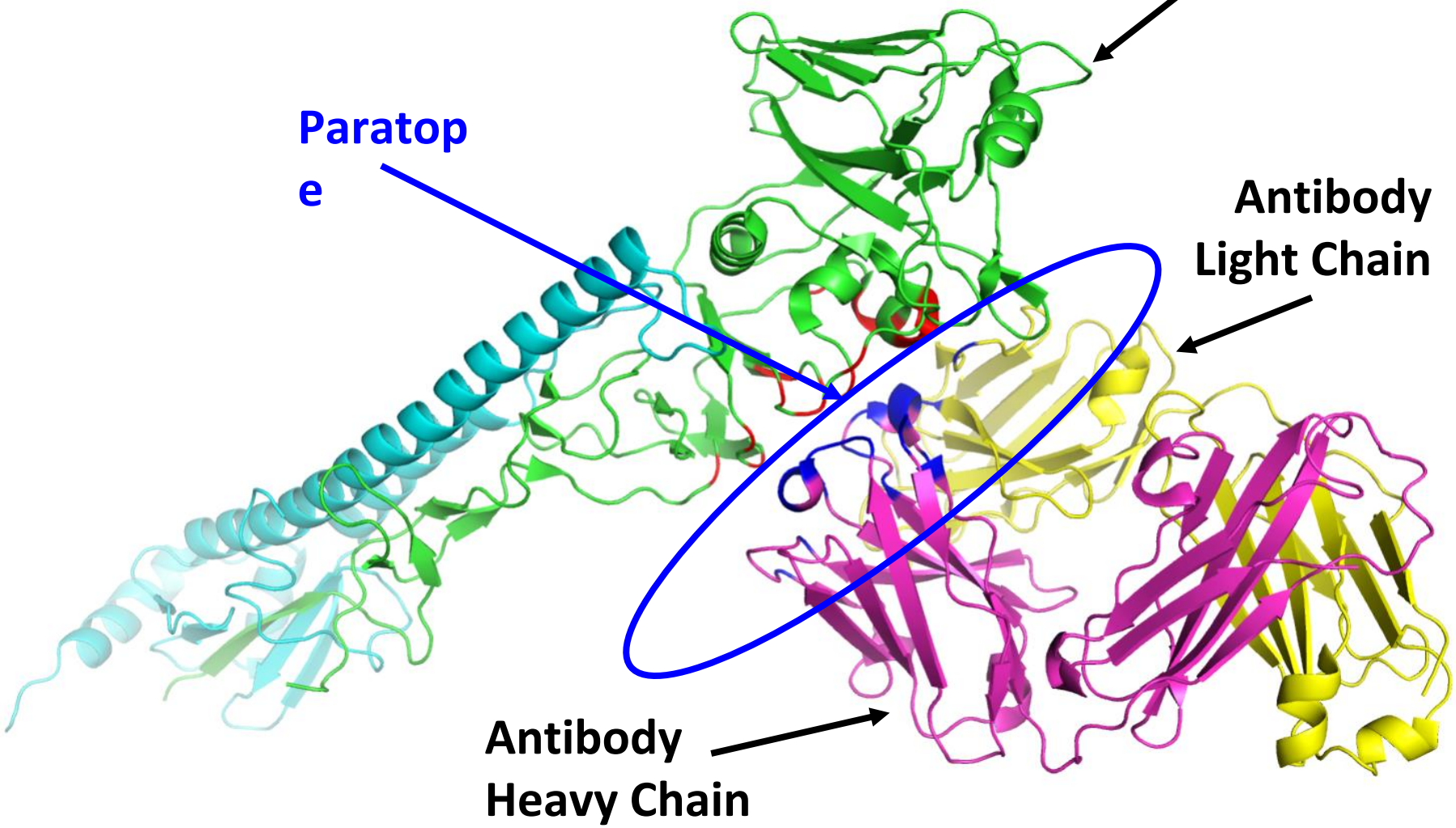
Discontinuous epitope

PDB ID: 1EO8



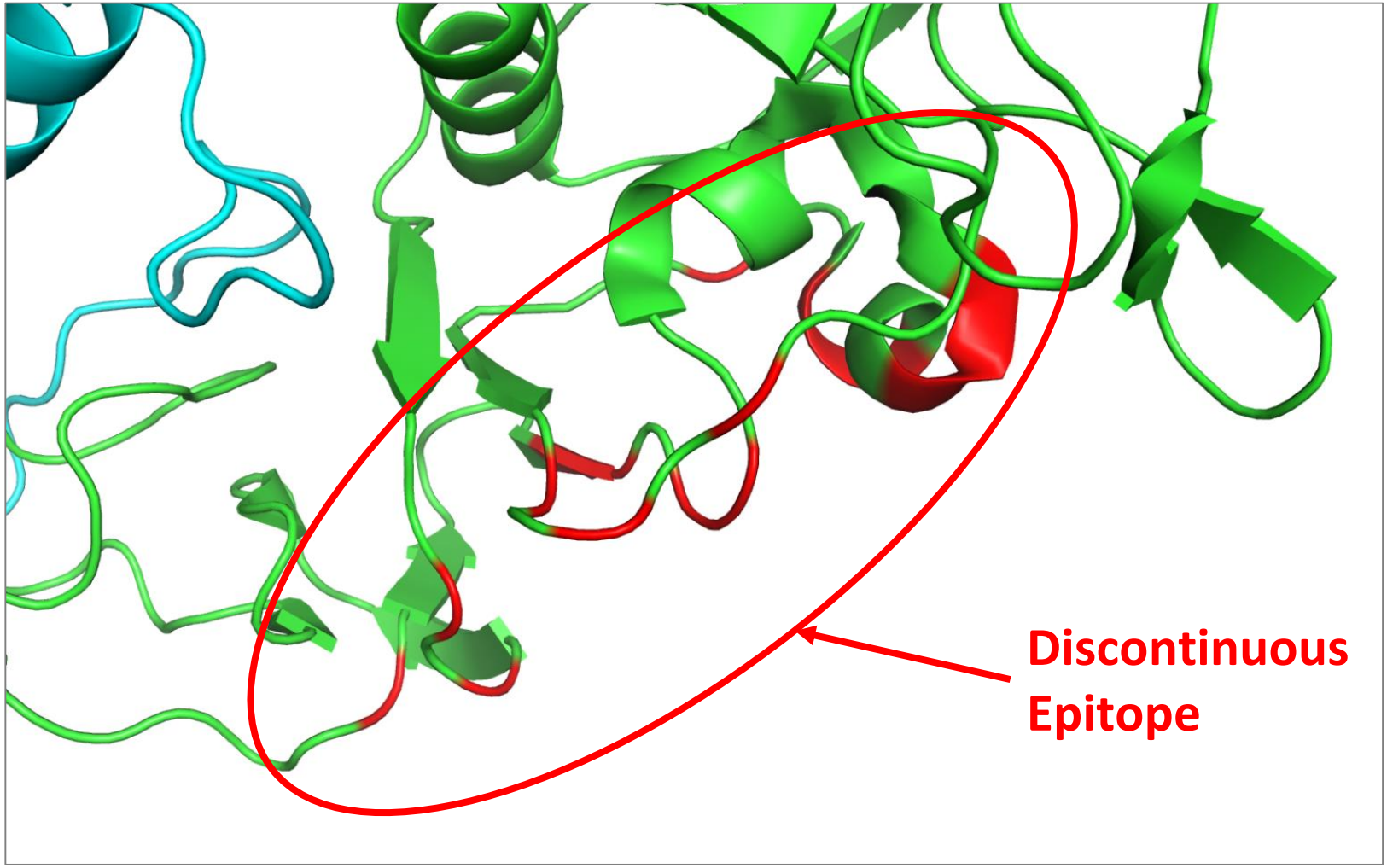
Discontinuous epitope

PDB ID: 1EO8



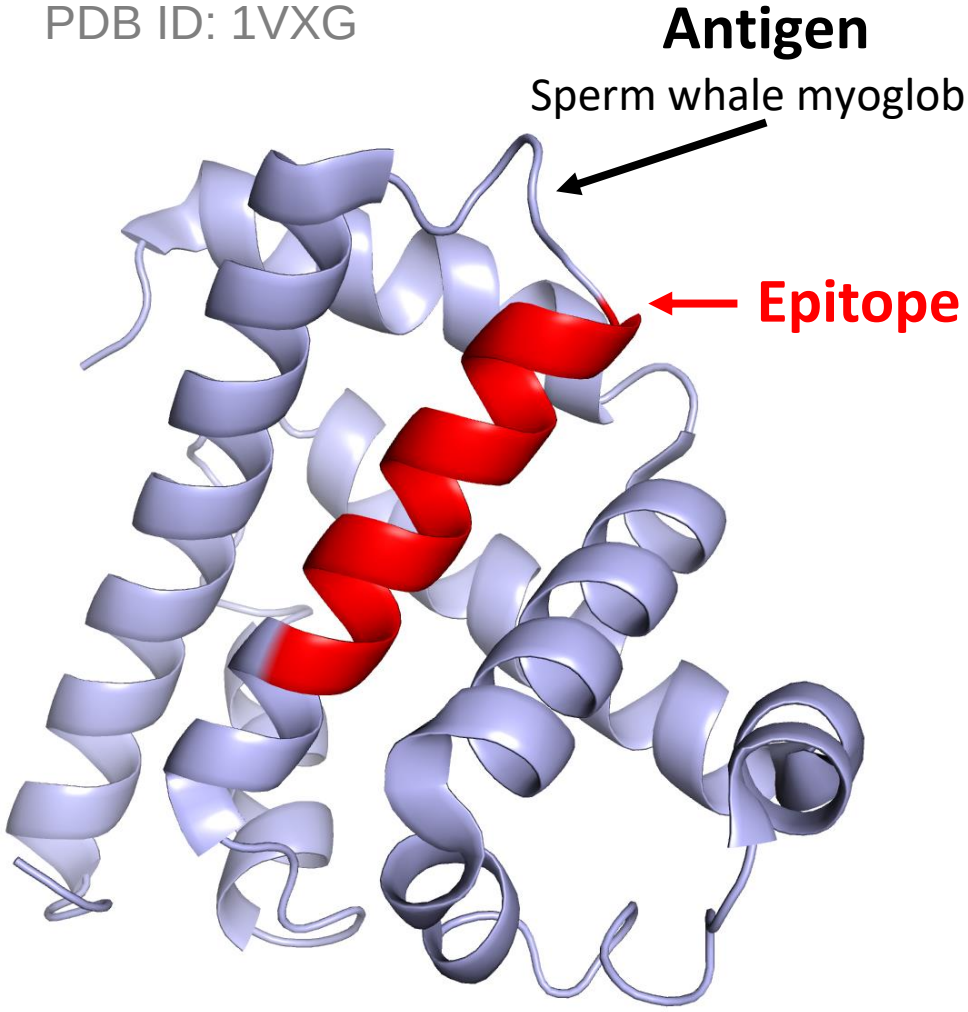
Discontinuous epitope

PDB ID: 1EO8



B cell epitopes

PDB ID: 1VXG

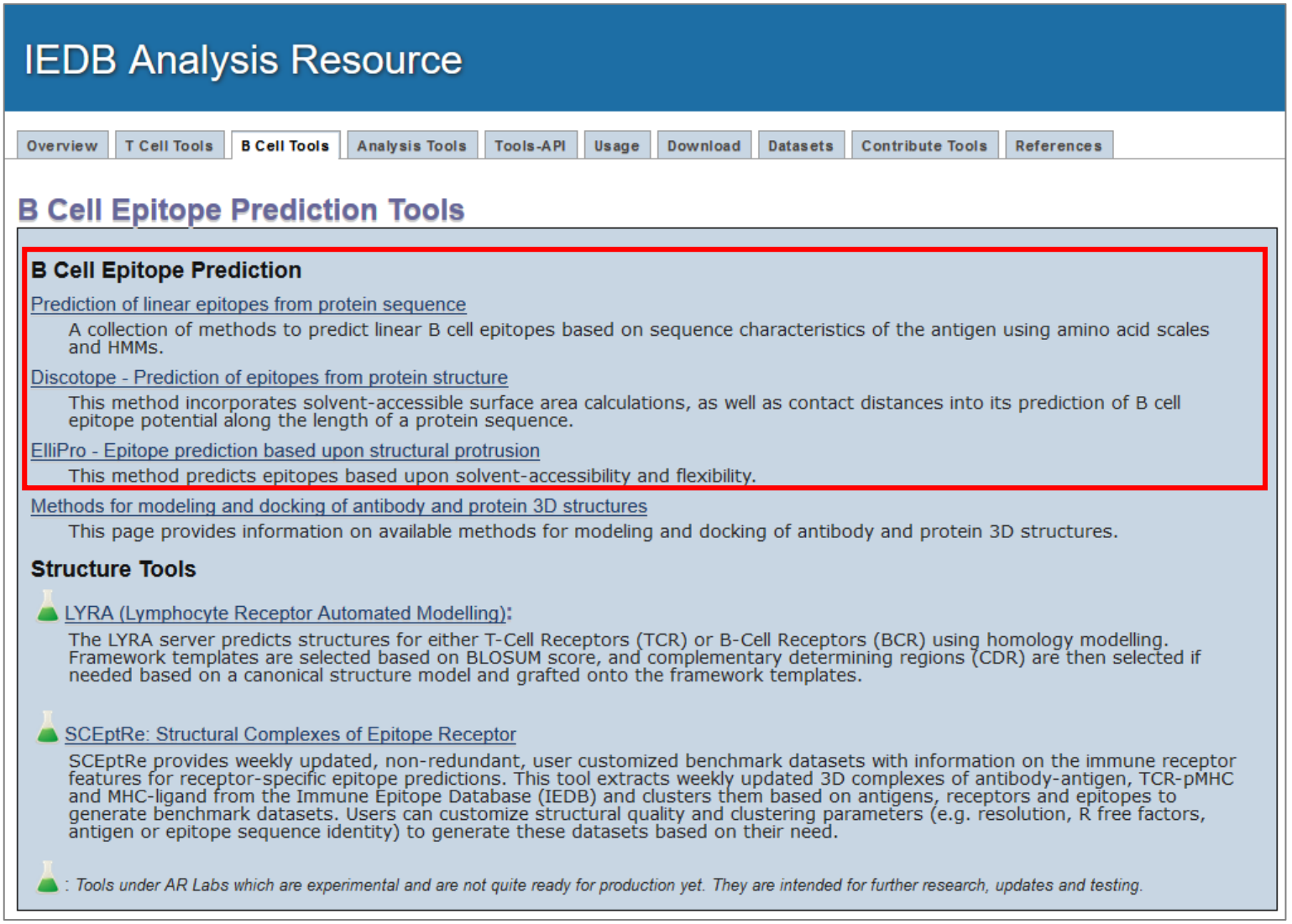


- Protein antigens usually contain both linear & discontinuous epitopes
  - "Linear" aka sequential or continuous
  - "Discontinuous" aka non-sequential or conformational epitopes
- More than 90% of the B cell epitopes are estimated to be discontinuous*

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

B cell prediction tools on IEDB

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



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

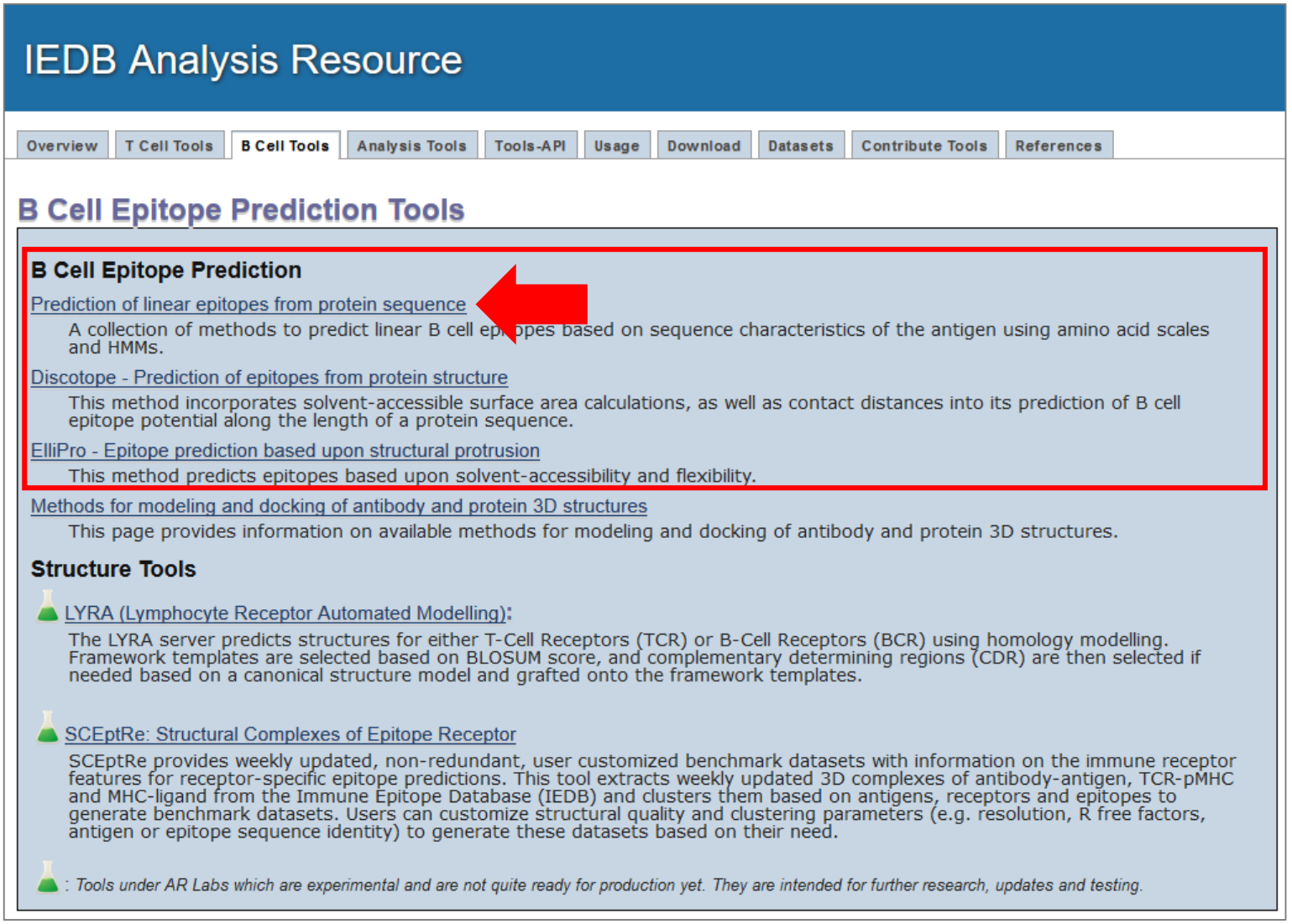
Linear epitope prediction

Linear epitope prediction methods:

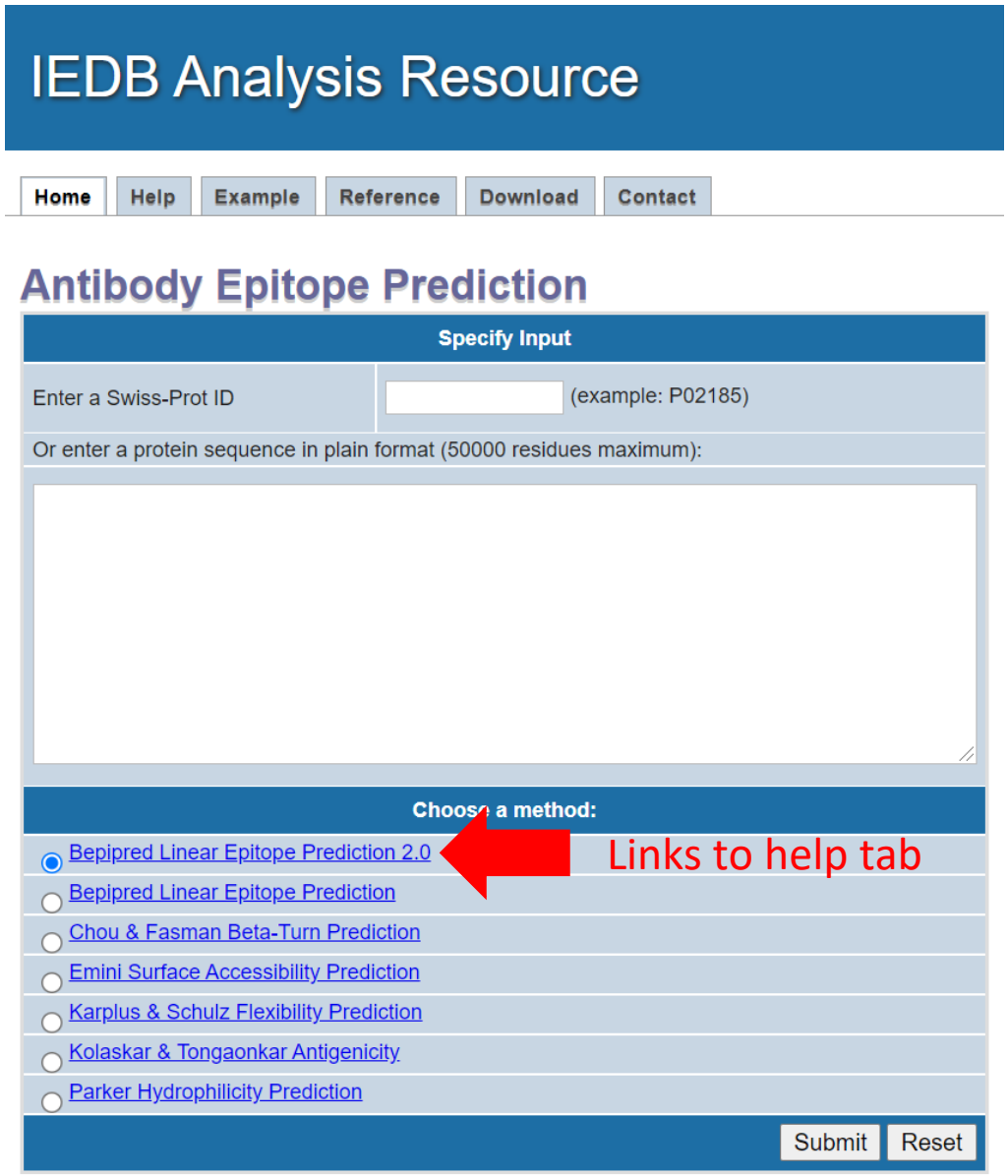
- Machine learning algorithms
  - Positive and negative training datasets are used
  - Combination of one or more amino acid scales are used as an input to one of the machine learning algorithms
    - Random Forest (BepiPred-2.0)
    - ANN: Artificial Neural Network (ABCpred)
    - SVM: Support Vector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

B cell prediction tools on IEDB

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



Linear B cell prediction



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

1. Input protein sequence

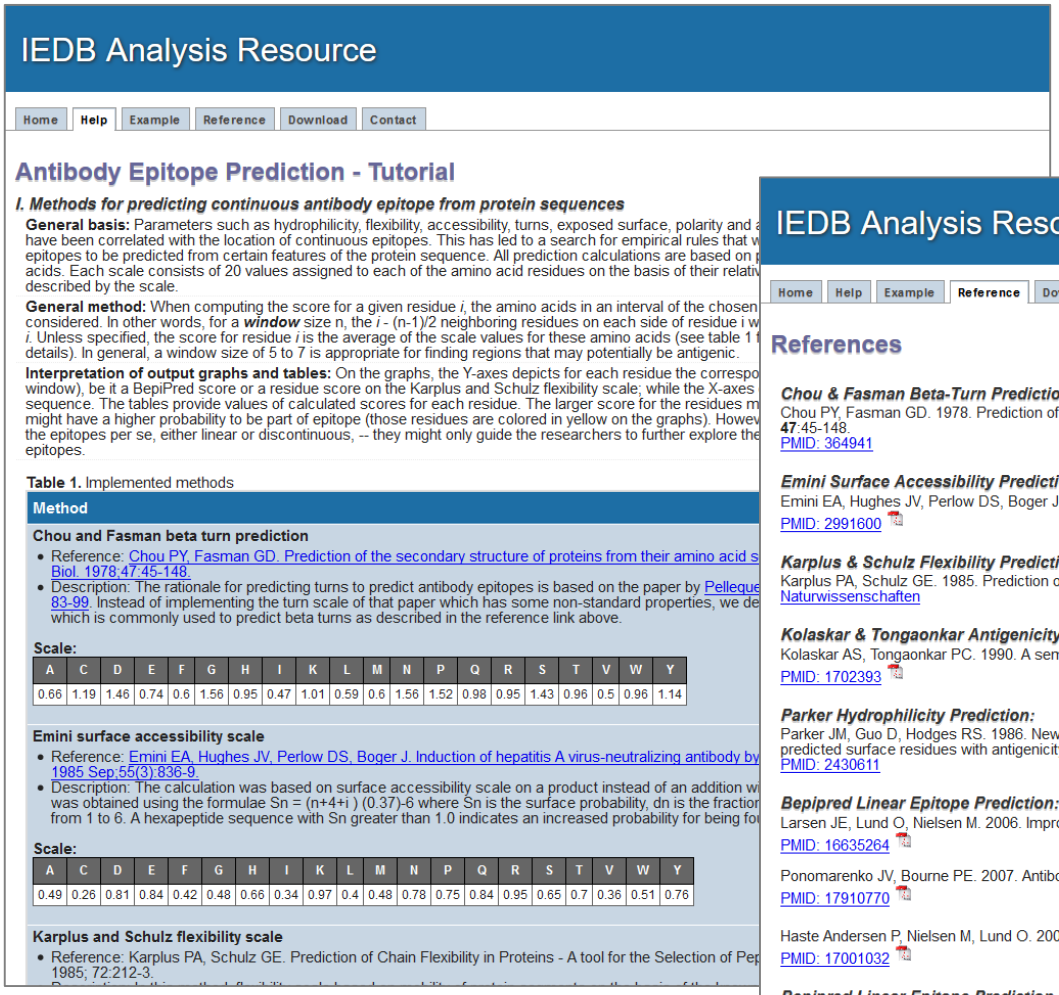
Entry allowed via Swiss-Prot ID or plain format

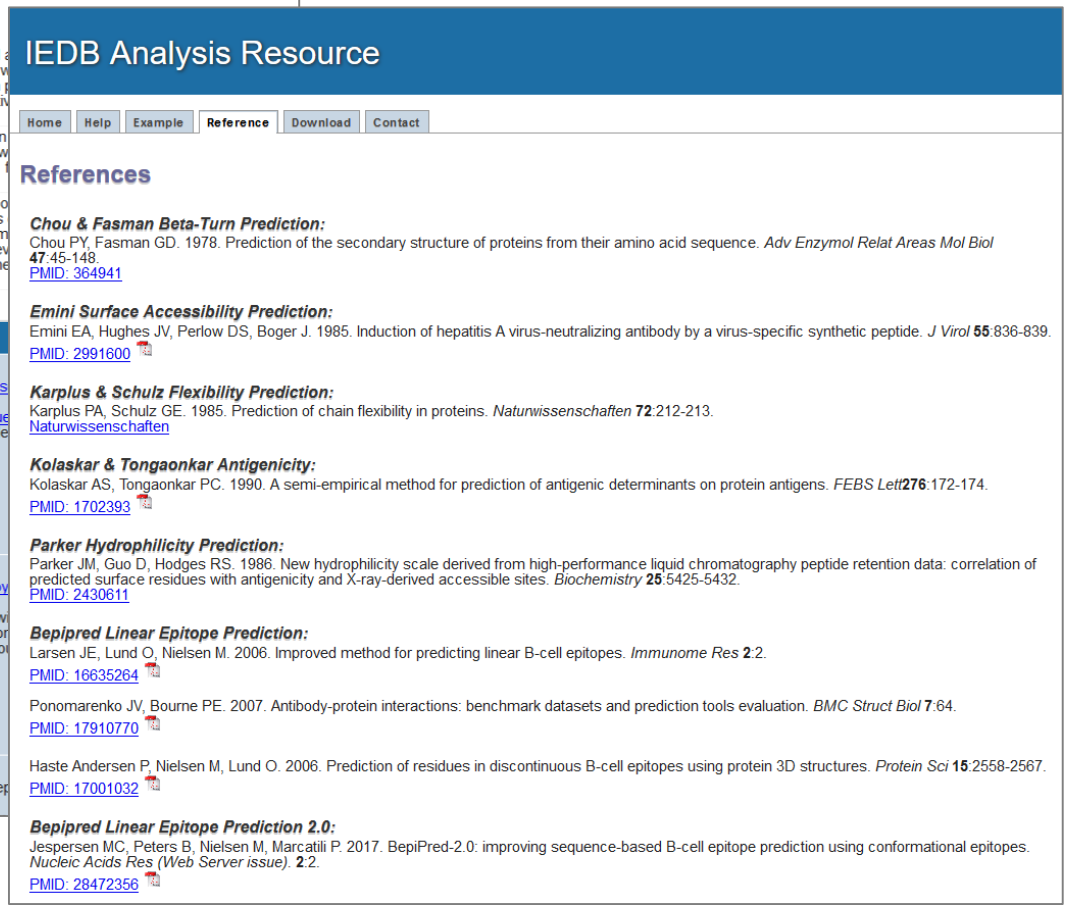
2. Select prediction method

BepiPred is the default & recommended method

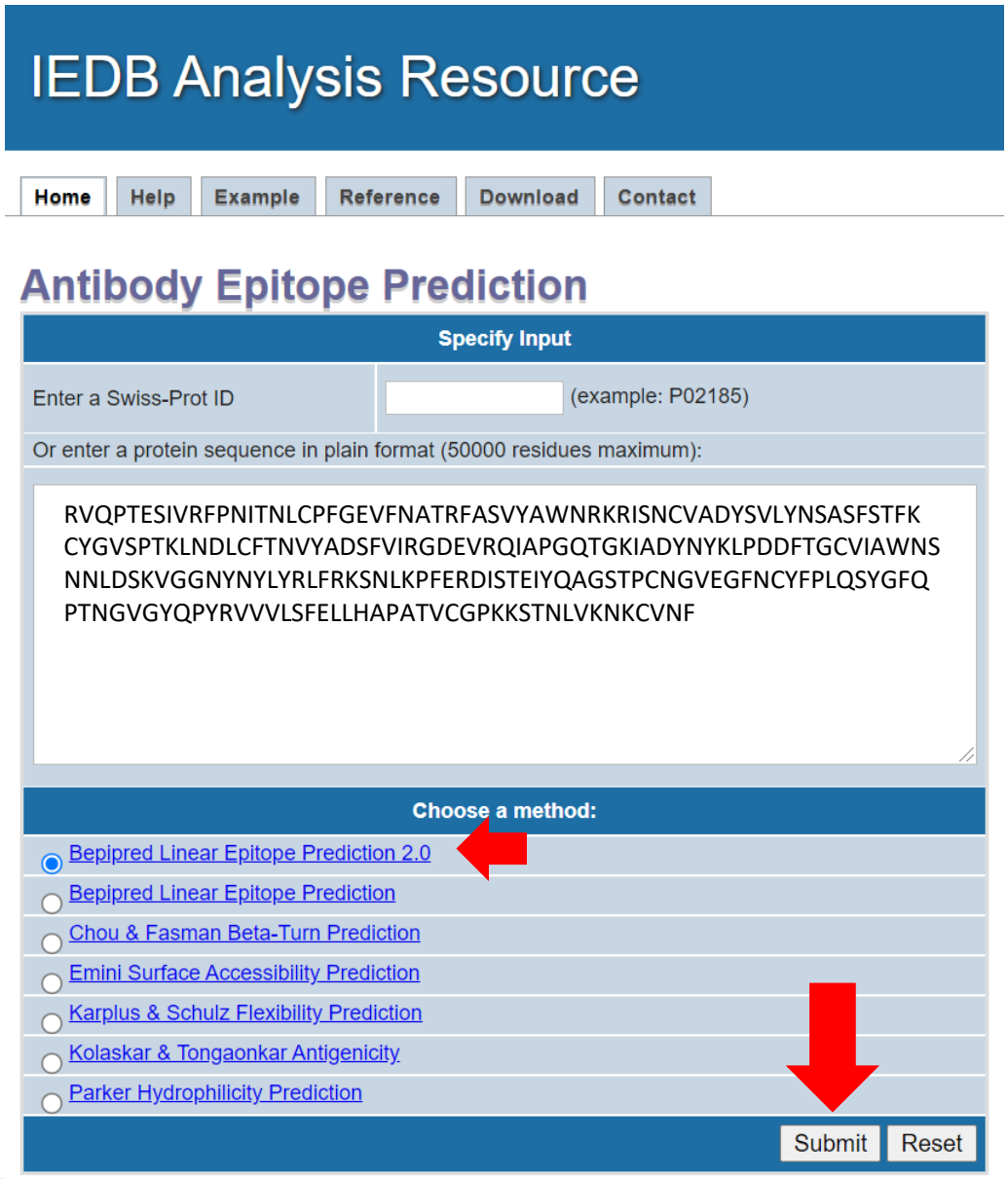
Visit Help & Reference tabs to learn about a prediction method

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





Linear B cell prediction - example



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

Example Sequence:

RBD region from SARS-Cov-2 Spike glycoprotein

Swiss-Prot ID: P0DTC2

Linear B cell prediction -example

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

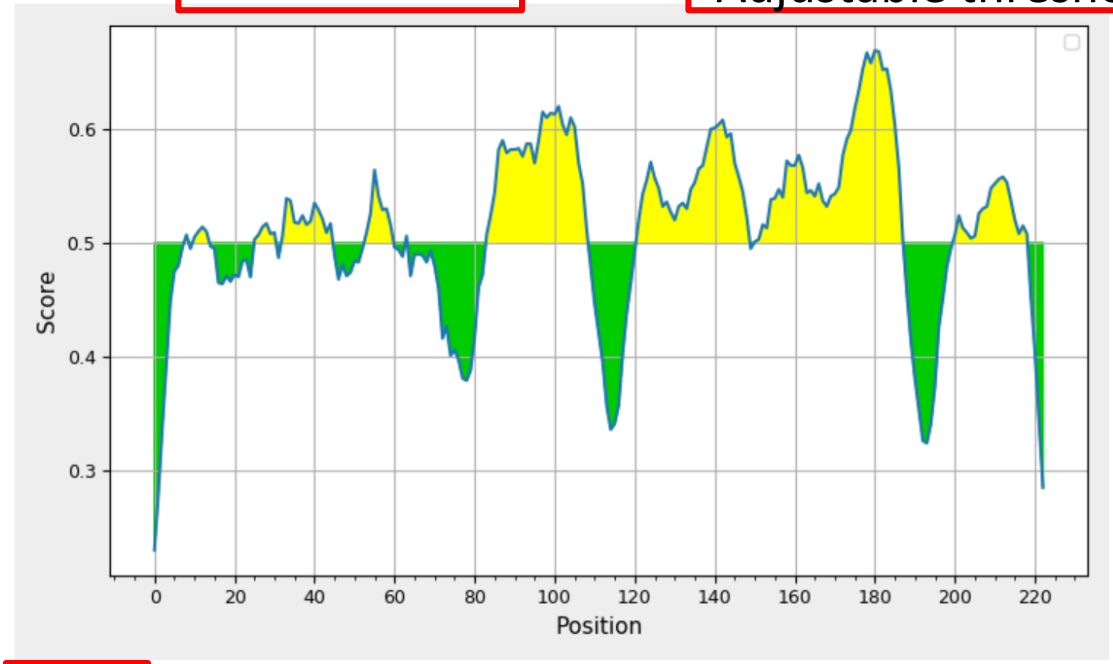
Bepipred Linear Epitope Prediction 2.0 Results

Input Sequences

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

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

Adjustable threshold



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	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

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	.

Method comparisons

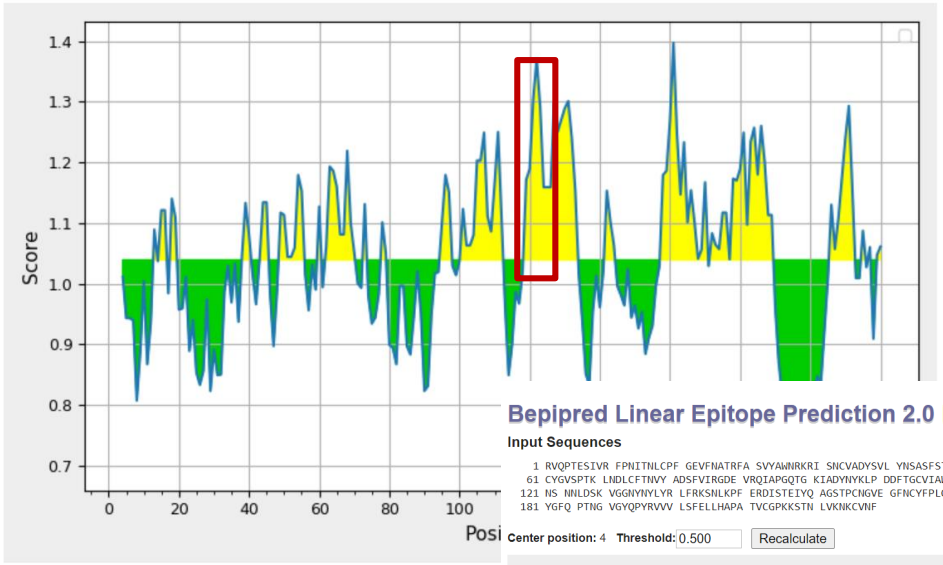
http://tools.iedb.org/bcell/

Chou & Fasman Beta-Turn Prediction Results

Input Sequences

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

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



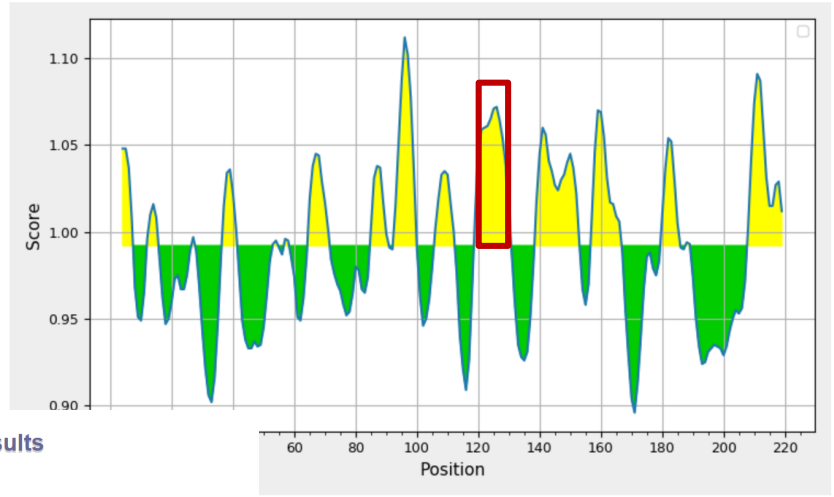
Average: 1.039 **Minimum:** 0.694 **Maximum:** 1.397

Karplus & Schulz Flexibility Prediction Results

Input Sequences

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

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



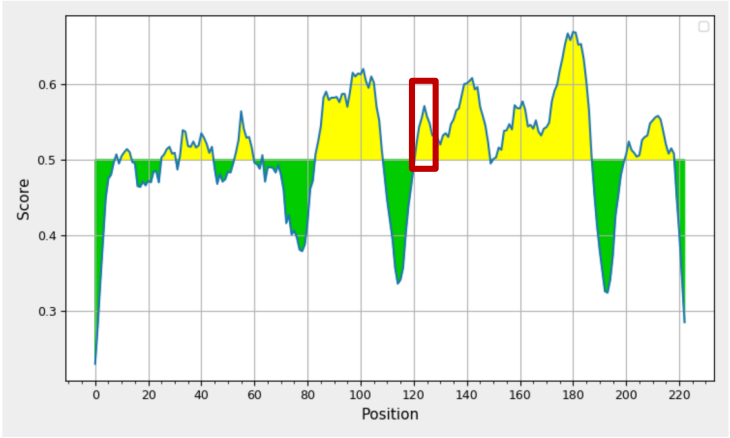
ıum: 1.112

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 GFNCYFPLQS
181 YGFQ PTNG VGYQPYRVVV LSFELLHAPA TVCGPKKSTN LVKNKCVNF
```

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



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

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 JA[1], 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

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 (2202 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)

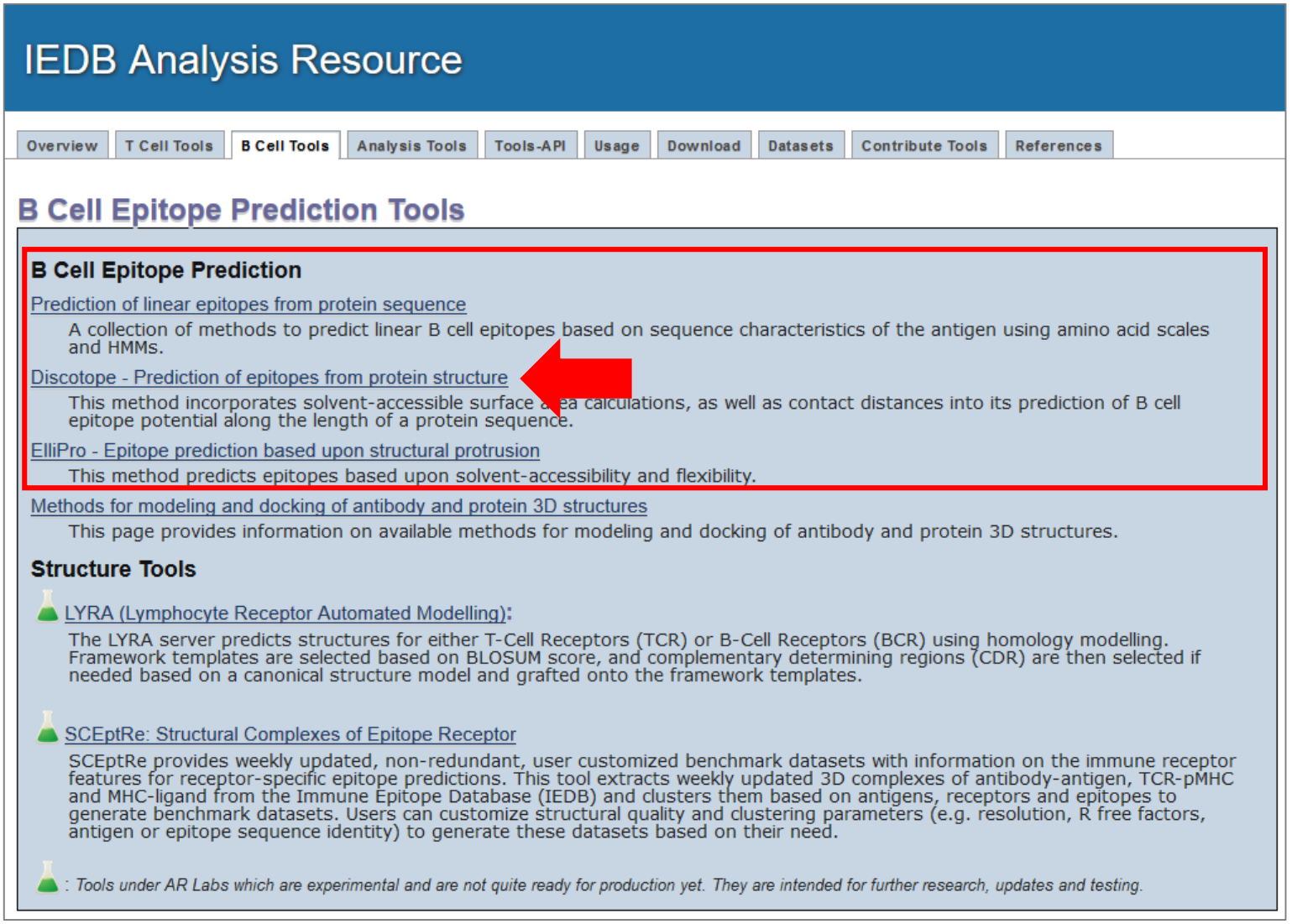
3D Structure-based epitope prediction

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

B cell prediction tools on IEDB

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



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 P[1], 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.

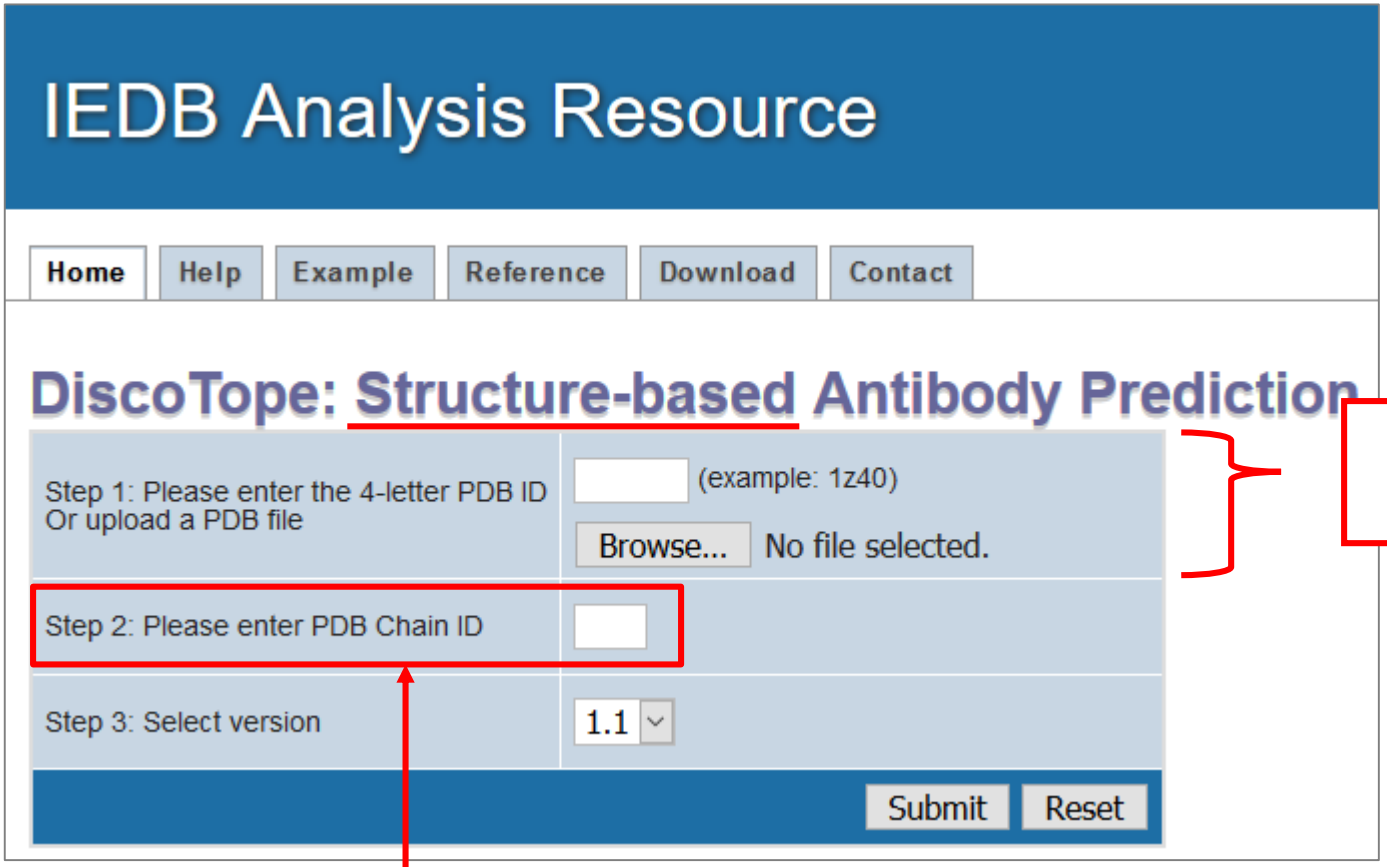
Reliable B cell epitope predictions: impacts of method development and improved benchmarking.

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

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

DiscoTope

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



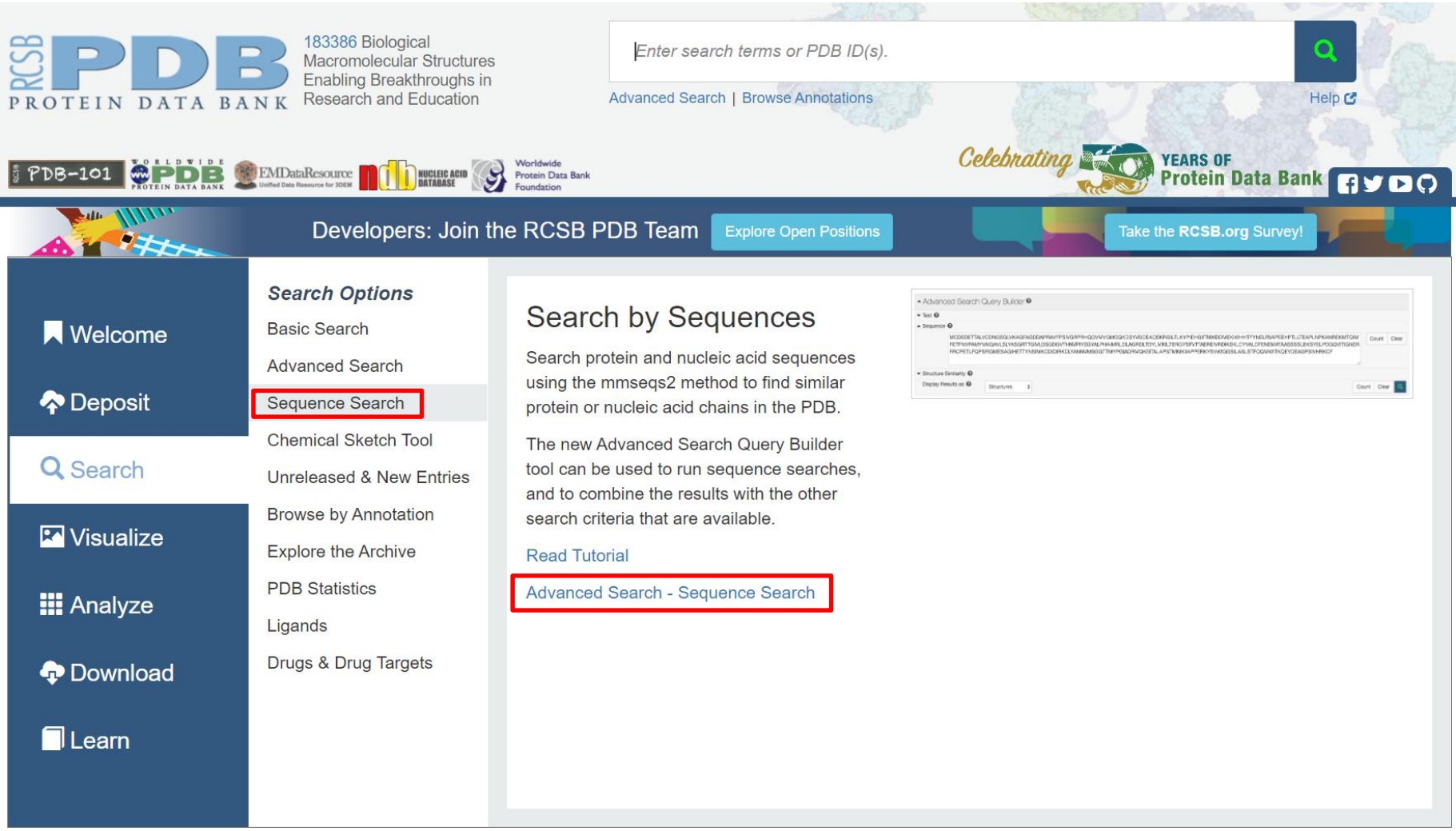
Identify structure

Input PDB id or upload file

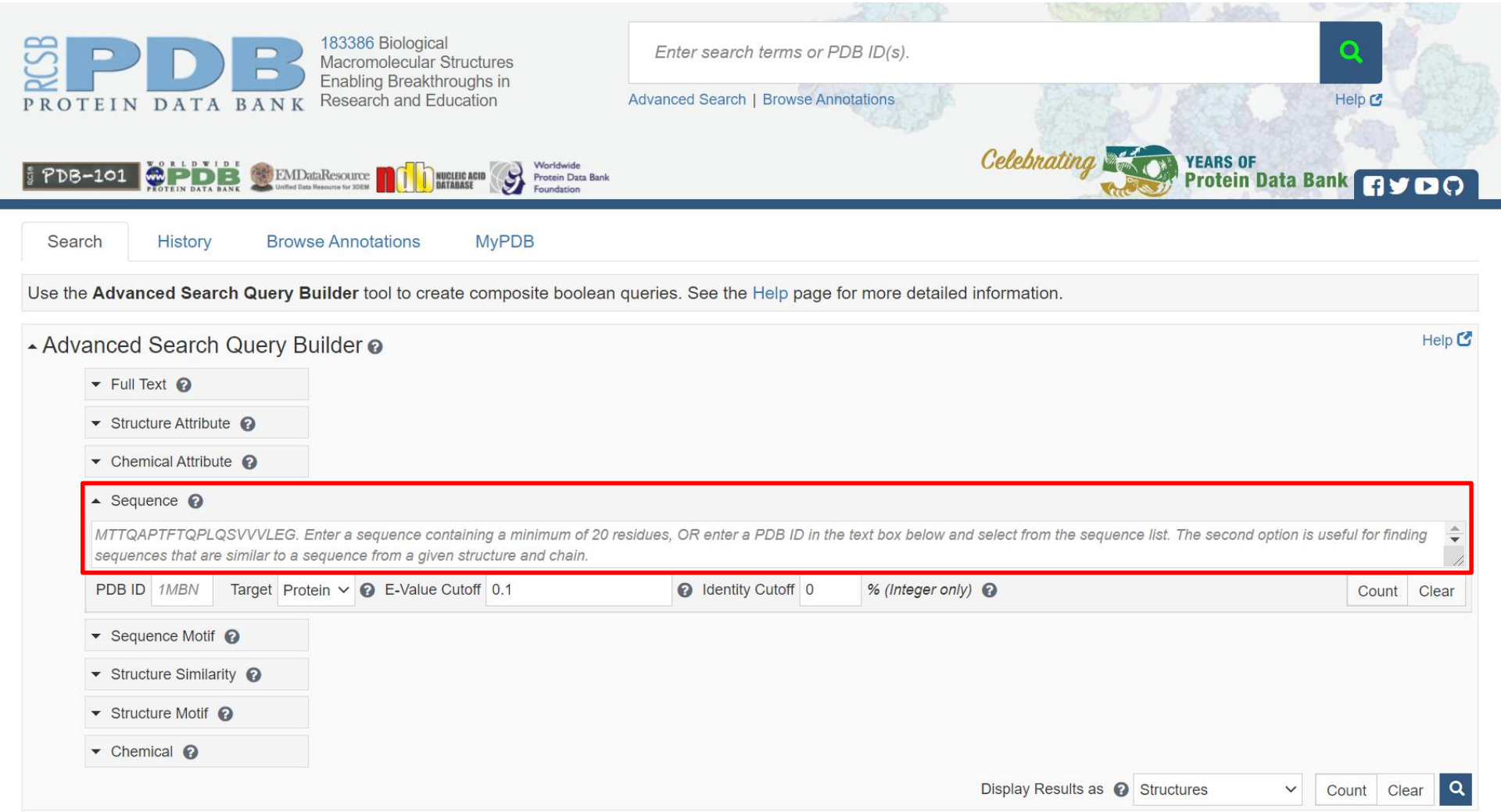
Specific to the protein chain of interest

Search in PDB to identify inputs

http://www.rcsb.org/



Search in PDB to identify inputs

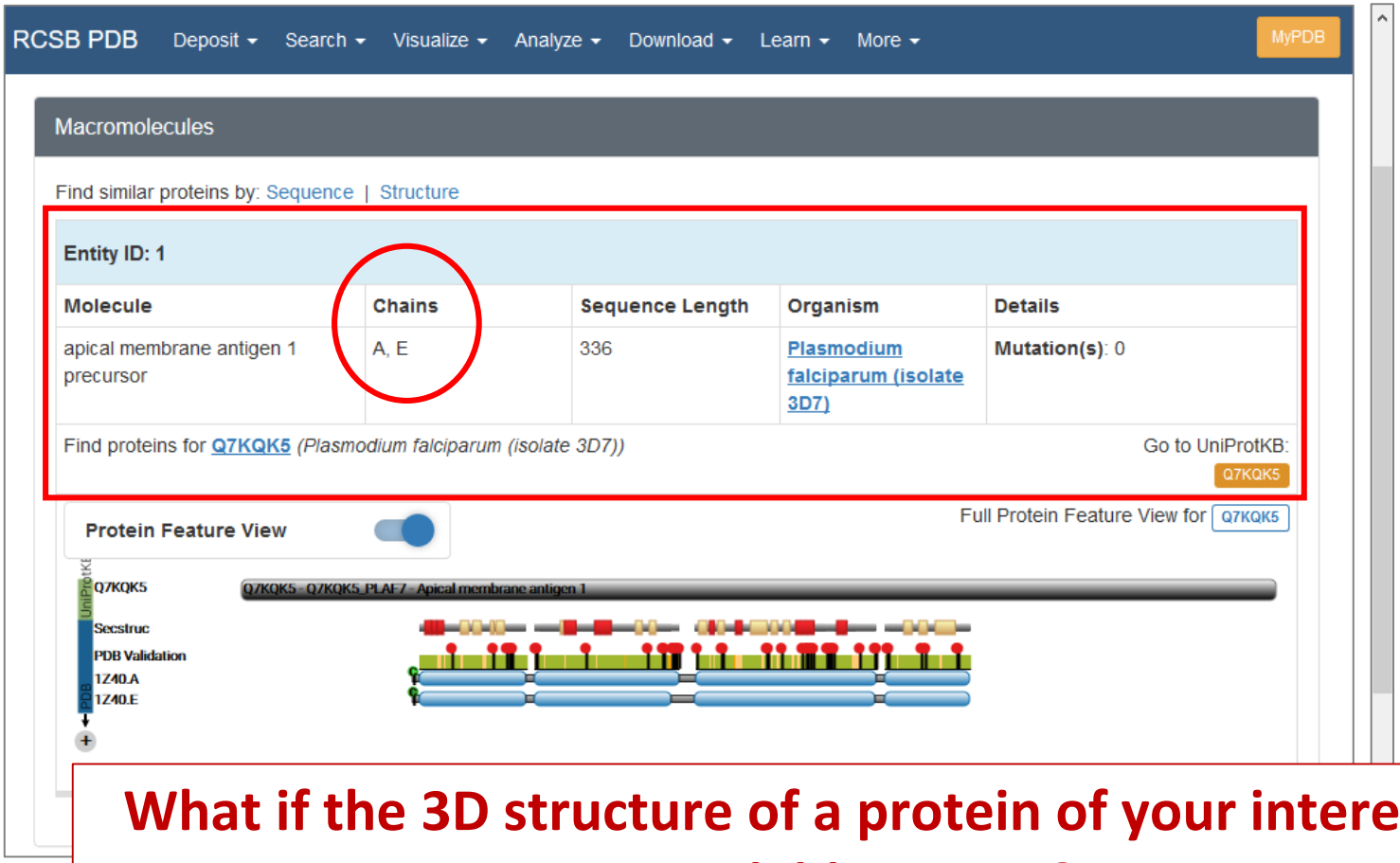


Search in PDB to identify inputs

http://www.rcsb.org/

Search in PDB to identify inputs

http://www.rcsb.org/

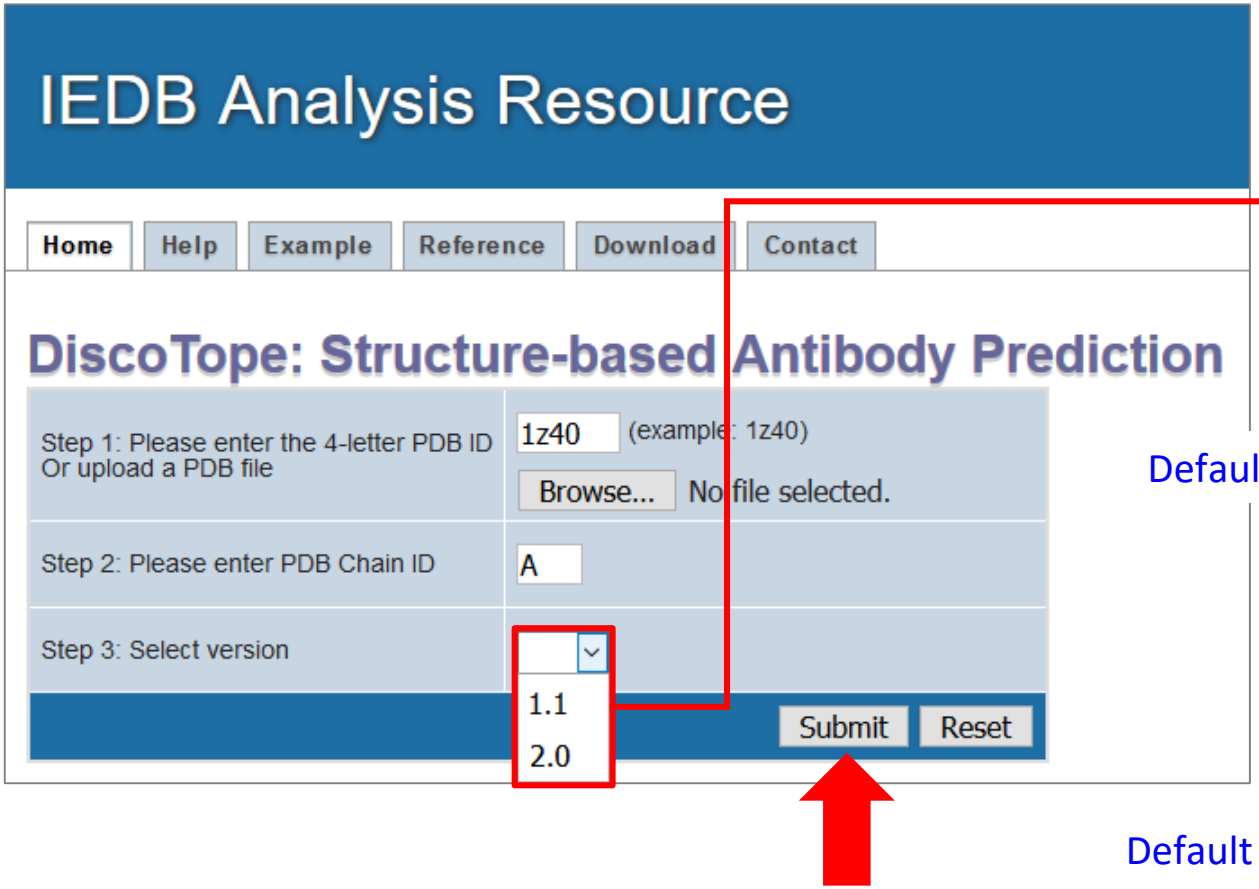


What if the 3D structure of a protein of your interest is not available in PDB?

> Homology or comparative modeling methods, servers and databases

DiscoTope - Example

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



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

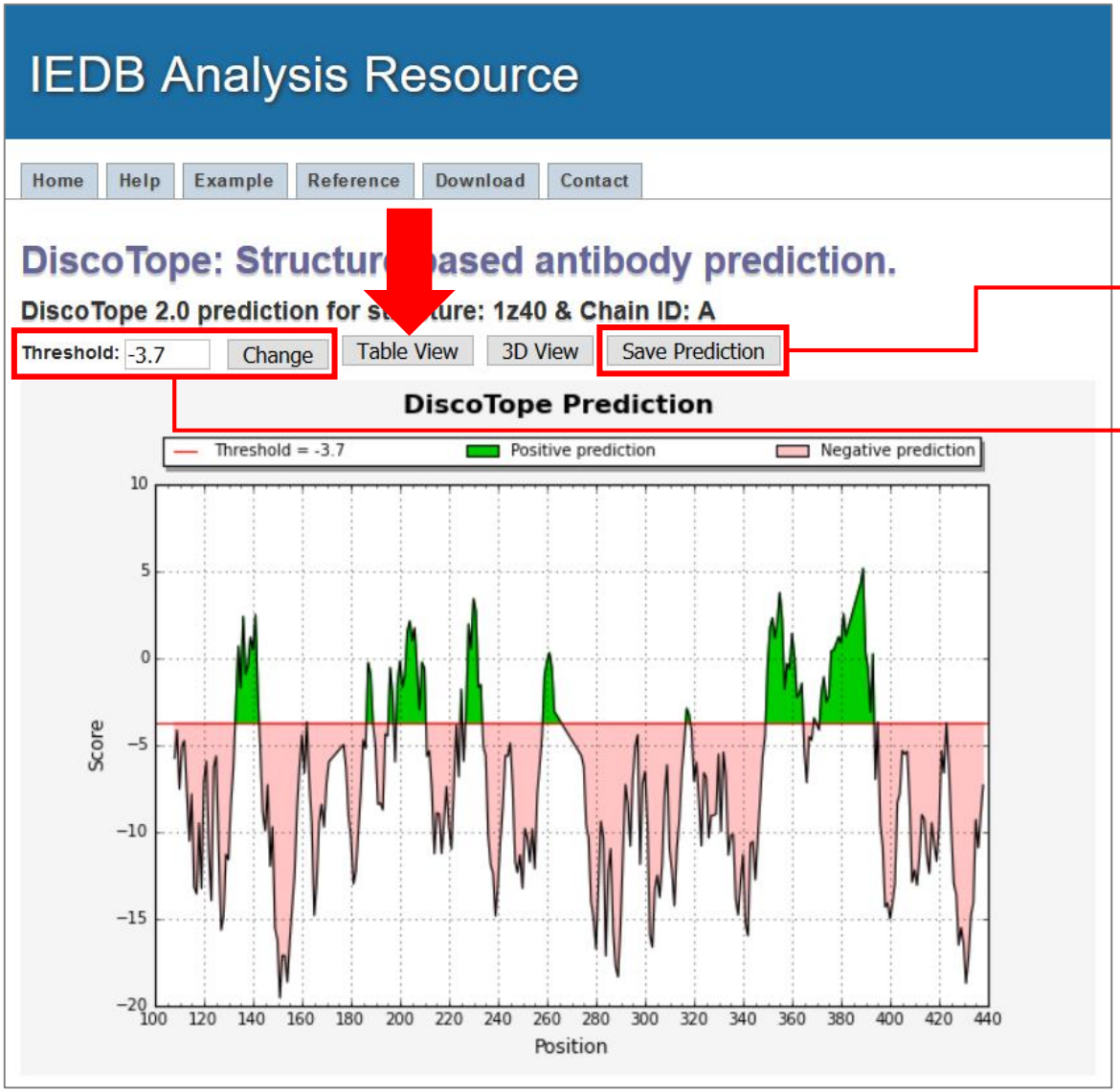
DiscoTope-2.0

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/



Downloads as csv

Adjustable threshold

DiscoTope - Example

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

DiscoTope - Result

DiscoTope 2.0 ediction for structure: 1z40 & Chain ID: A

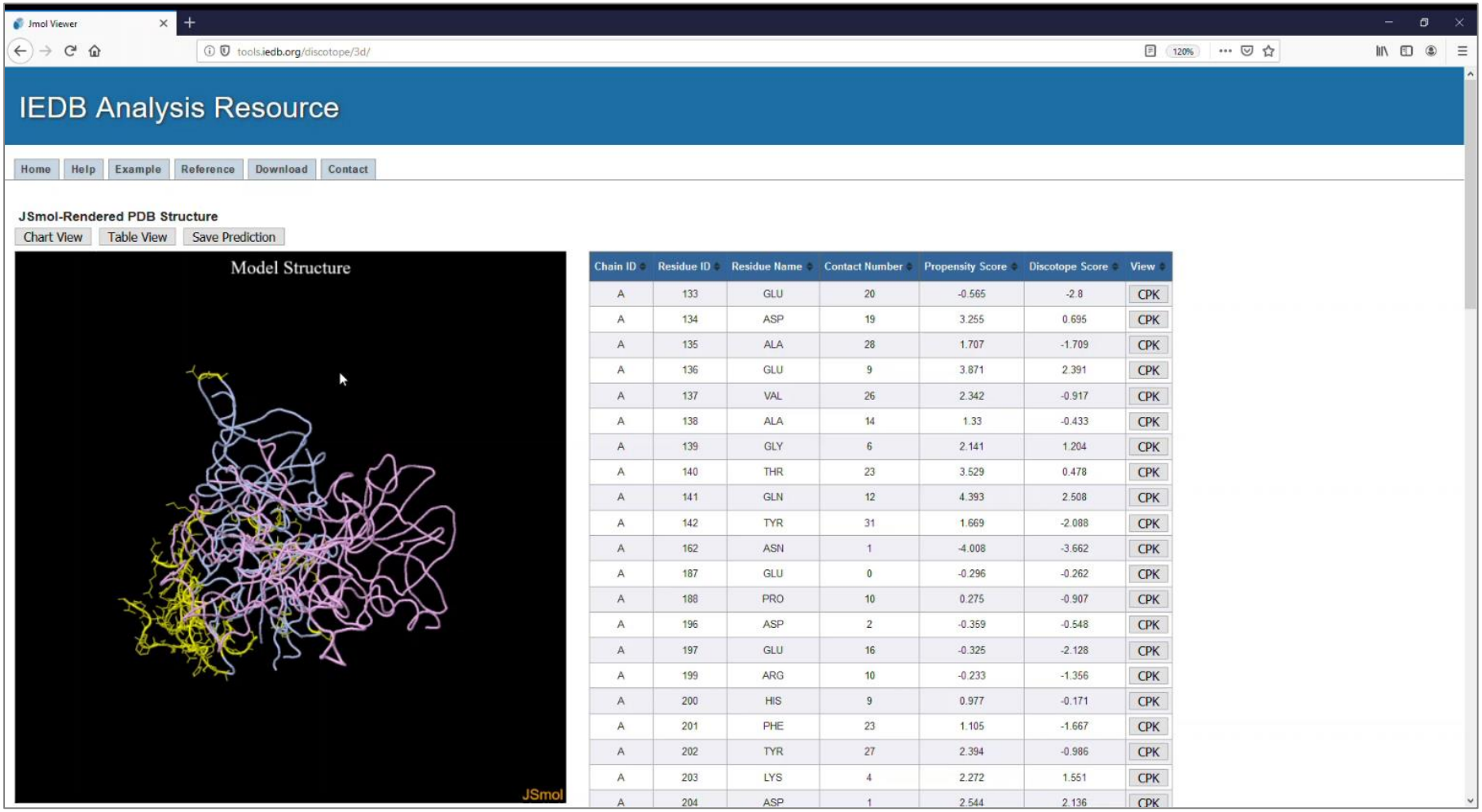
The positive pr ns are displayed in green.

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**

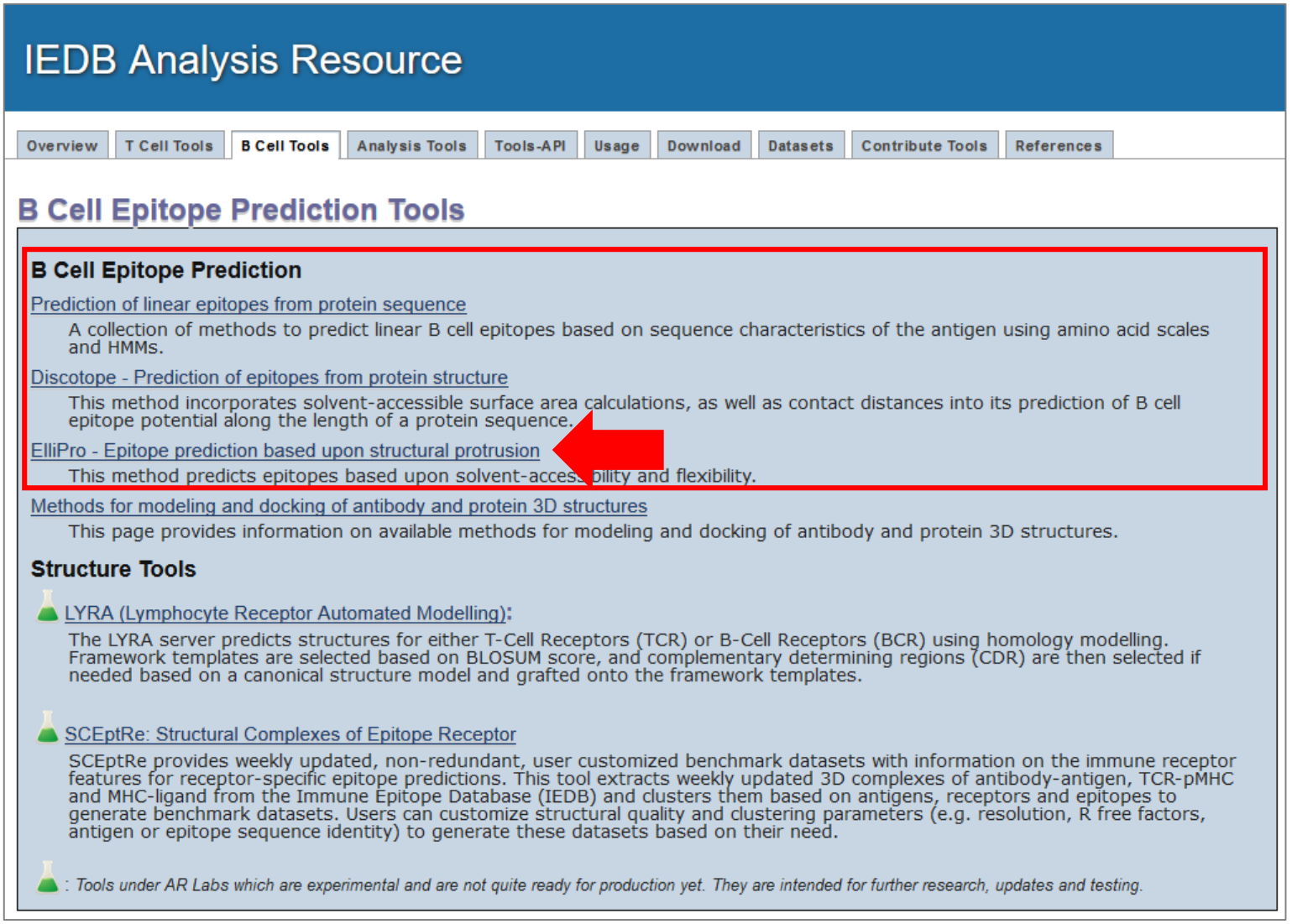
DiscoTope - Example

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



B cell prediction tools on IEDB

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



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.

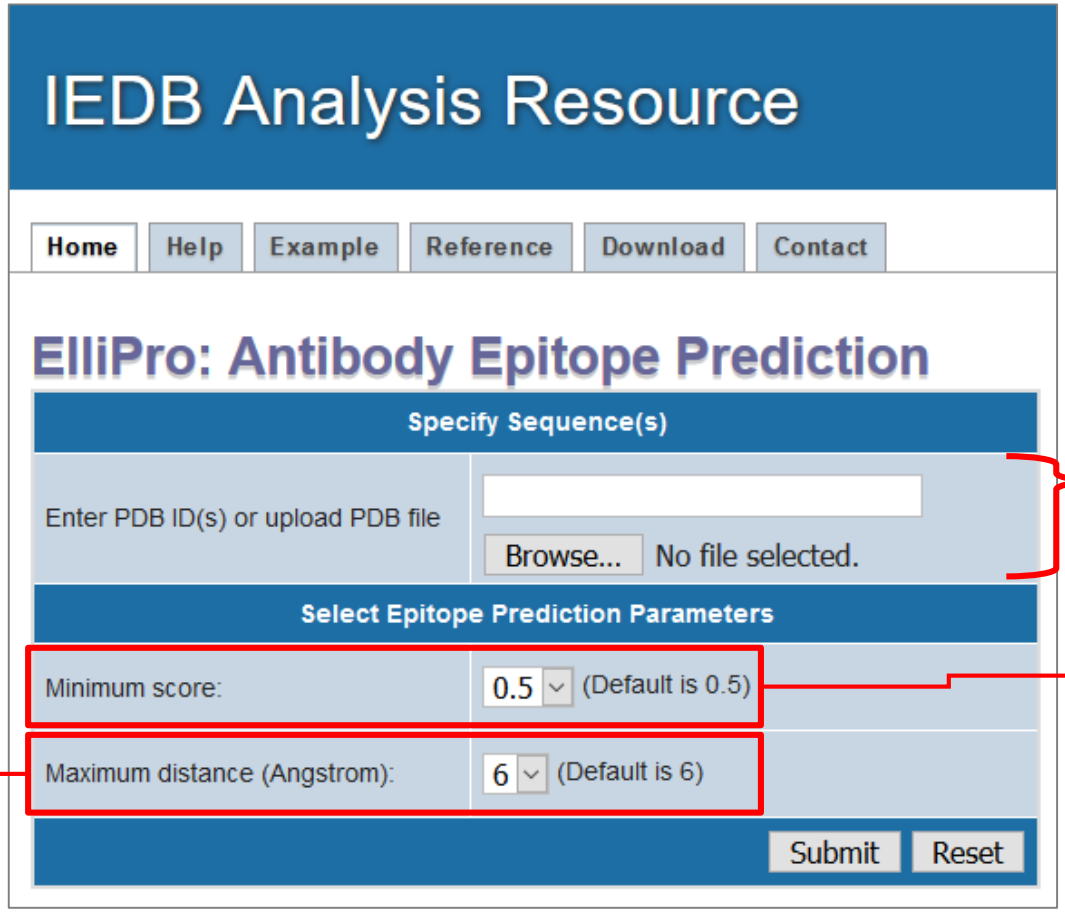
ElliPro: a new structure-based tool for the prediction of antibody epitopes.

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

ElliPro

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



Identify structure

Input PDB id or upload file

Select min. PI value

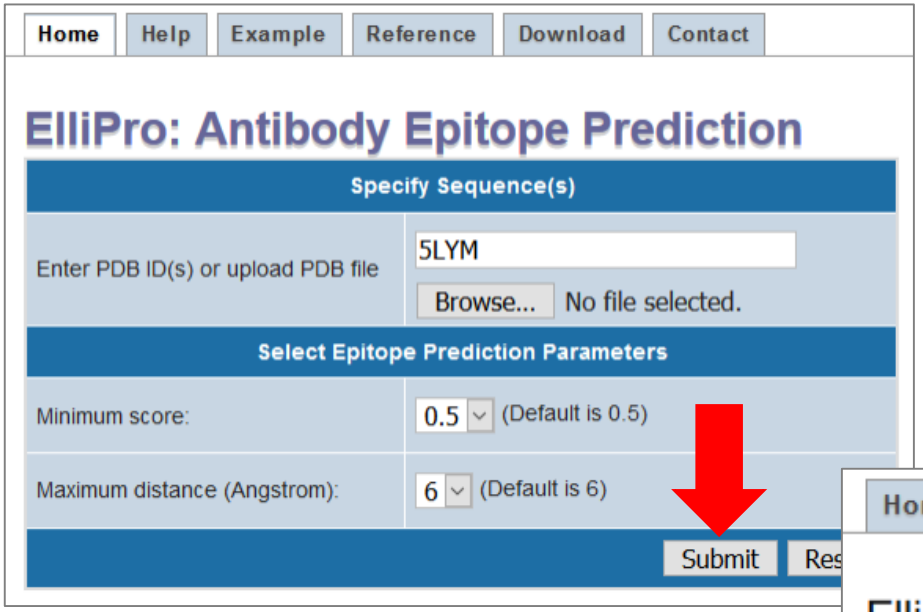
- Averaged over epitope residues
- Higher scores predict fewer epitopes

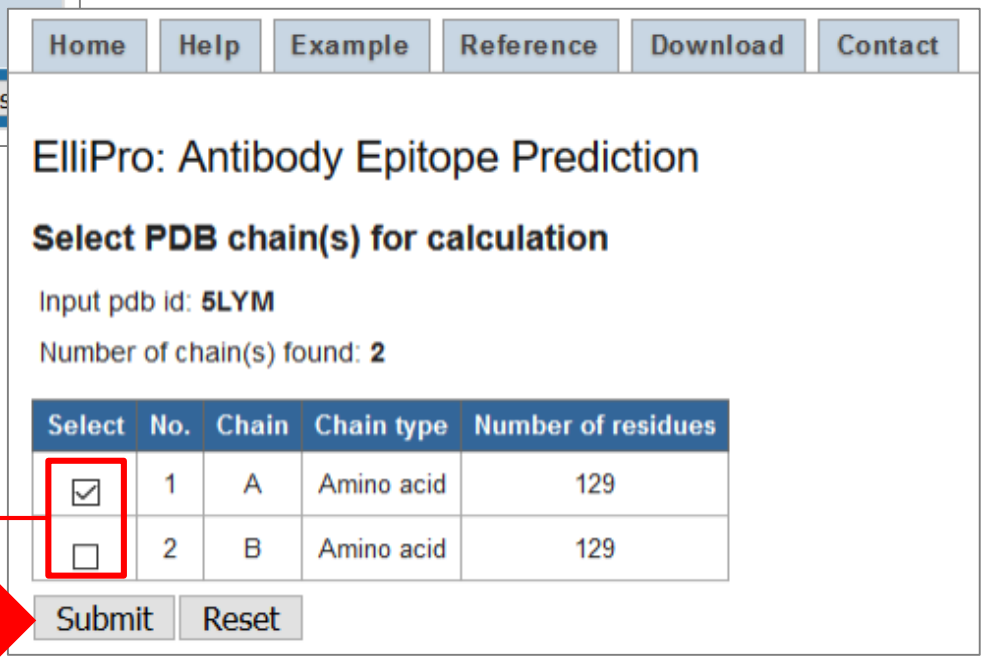
Specify max distance for predicting (grouping) discontinous epitopes

Longer distances predict discontinuous epitopes spanning larger regions

ElliPro -example

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

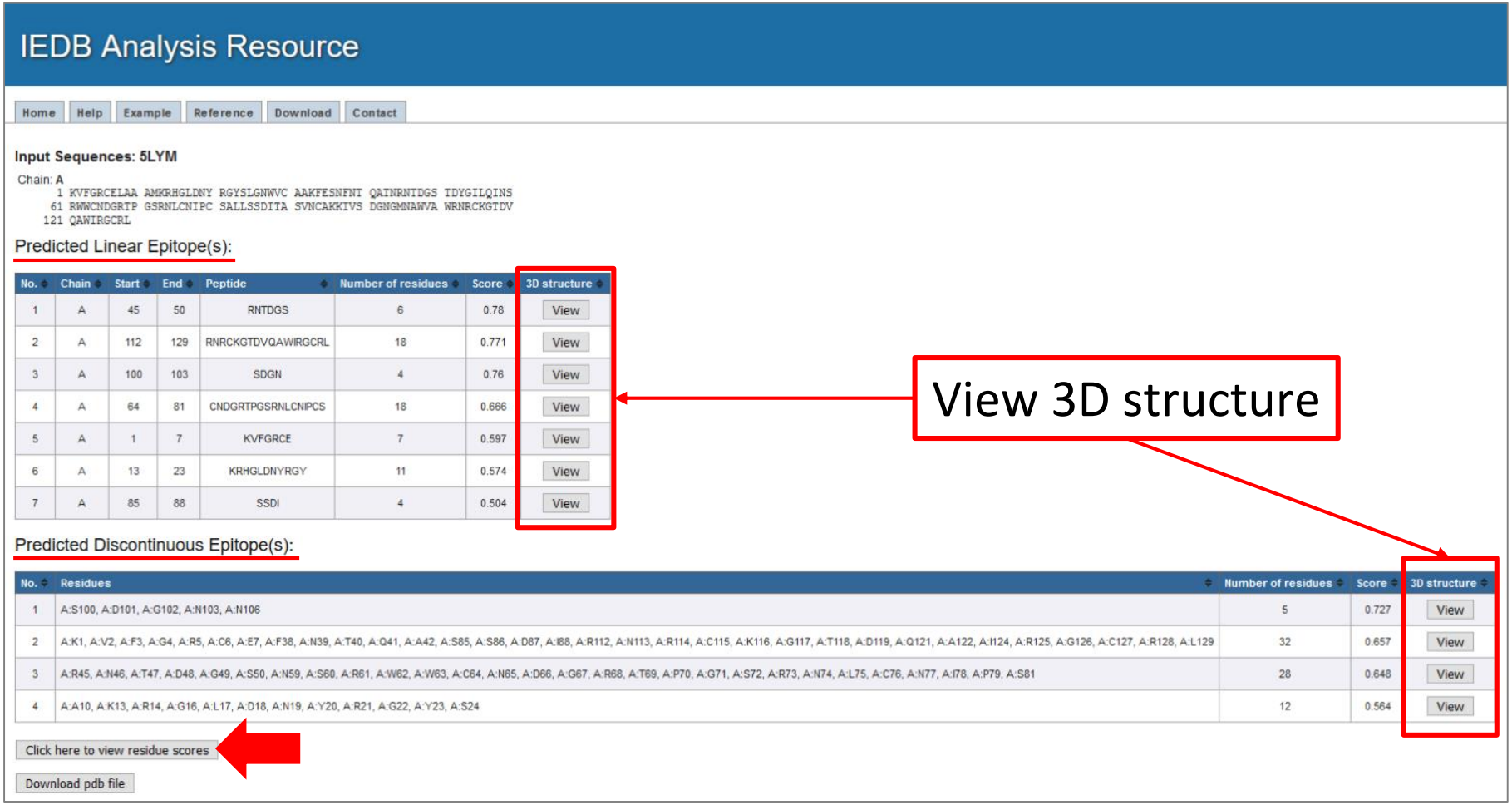




Select chain(s) of interest

ElliPro - Example

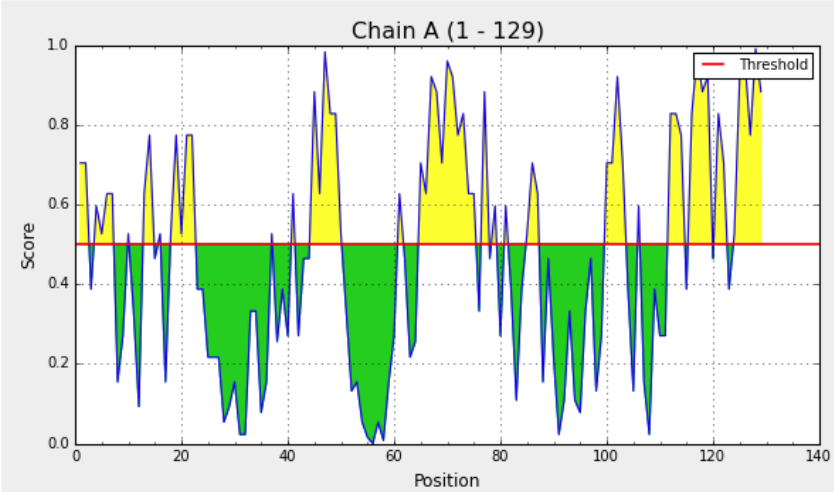
http://tools.iedb.org/ellipro/



ElliPro - Example

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

ElliPro: 2D Score Chart(s) for 5LYM



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

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.

IMMUNE EPITOPE DATABASE
AND ANALYSIS RESOURCE

THANK YOU!

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**)

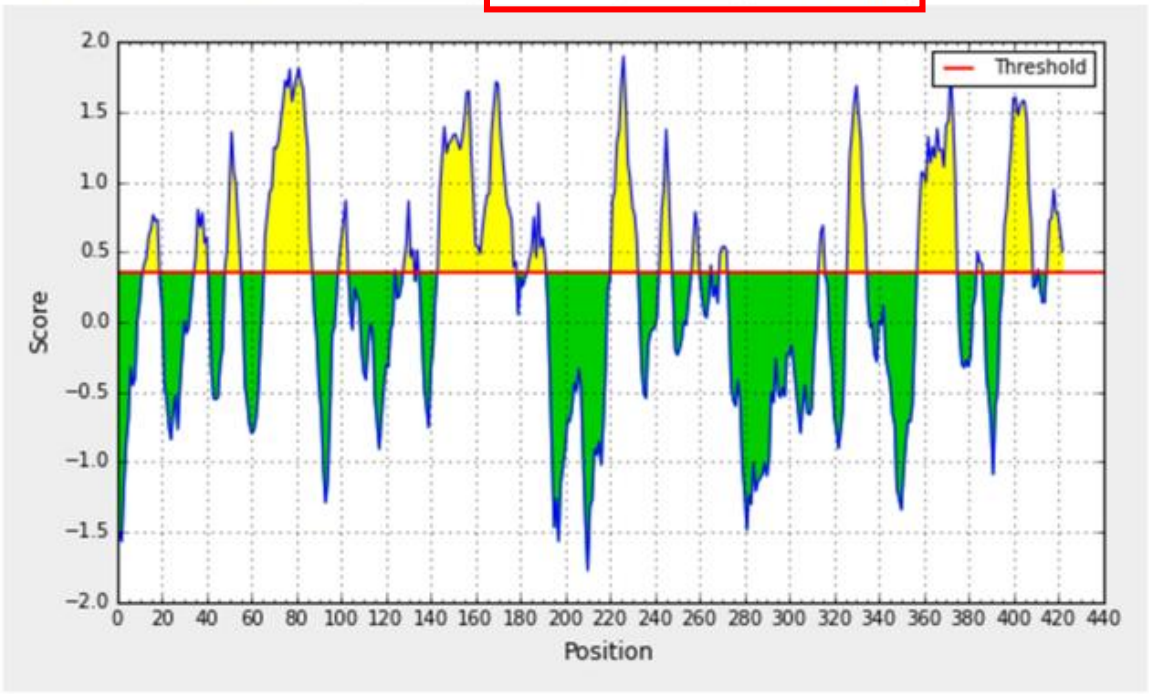
Practice Exercise

Bepipred Linear Epitope Prediction Results

Input Sequences

```
  1 MRCIGISNRD FVEGVSGGSW VDIVLEHGSC VTTMAKNKPT LDFELIKTEA KQPATLRKYC
 61 IEAKLTNTTT ESRCPTQGEP SLNEEQDK RFICKHSMVD RGWGNGCGLF 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 RVQYEGDGSP CKIPFEITDL EKRHVLGRLI
361 TVNPIVTEKD SPVNIEAEPP FGDSYIIVGV EPGQLKLNWL RPLESRGP FEGKPIPNPL
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?

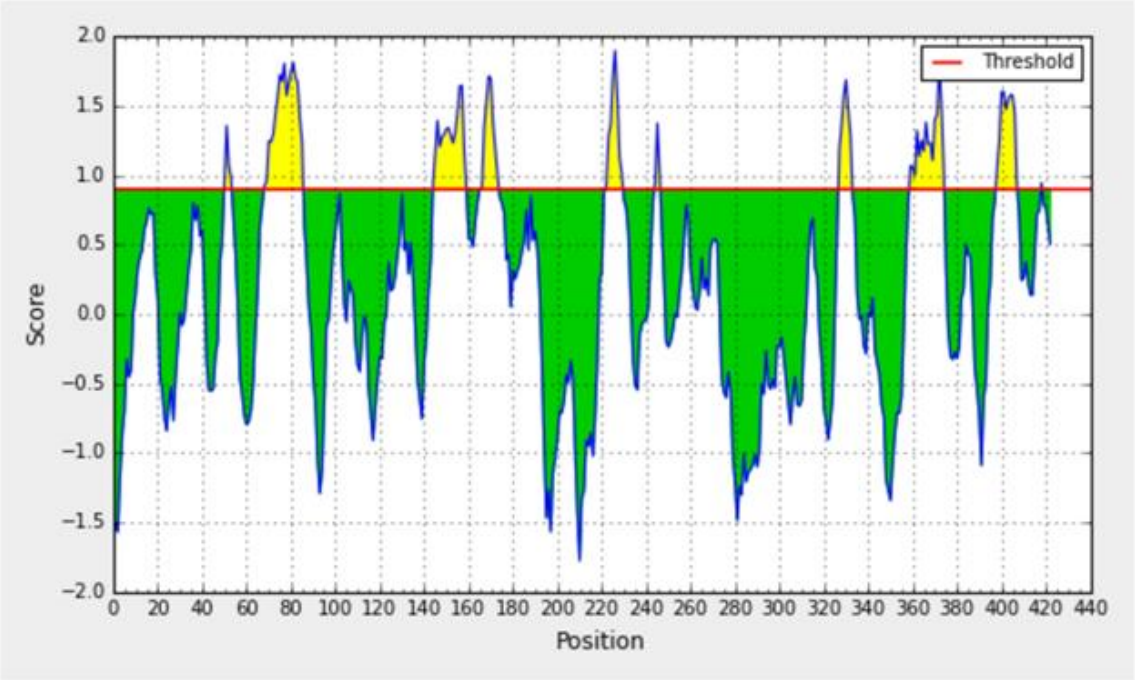
Practice Exercise

Bepipred Linear Epitope Prediction Results

Input Sequences

```
  1 MRCIGISNRD FVEGVSGGSW VDIVLEHGSC VTTMAKNKPT LDFELIKTEA KQPATLRKYC
 61 IEAKLTNTTT ESRCPTQGEP SLNEEQDK RFICKHSMVD RGWGNGCGLF 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 RVQYEGDGSP CKIPFEITDL EKRHVLGRLI
361 TVNPIVTEKD SPVNIEAEPP FGDSYIIVGV EPGQLKLNWL RPLESRGP FEGKPIPNPL
421 LGLDSTRTGH HH
```

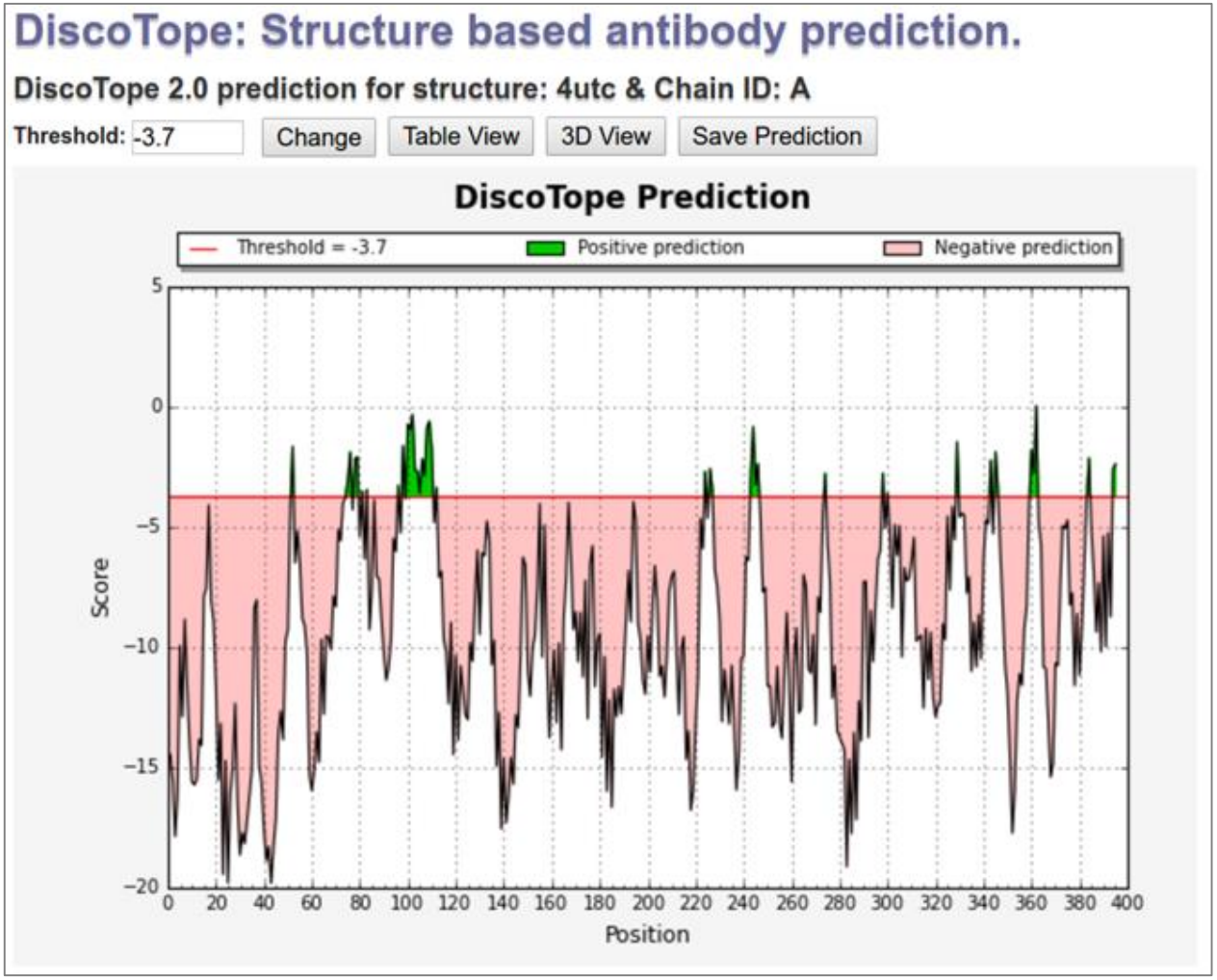
Center position: 4 **Window size:** 7 **Threshold:** 0.9 Recalculate



Average: 0.188 **Minimum:** -1.776 **Maximum:** 1.900

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

Practice Exercise



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

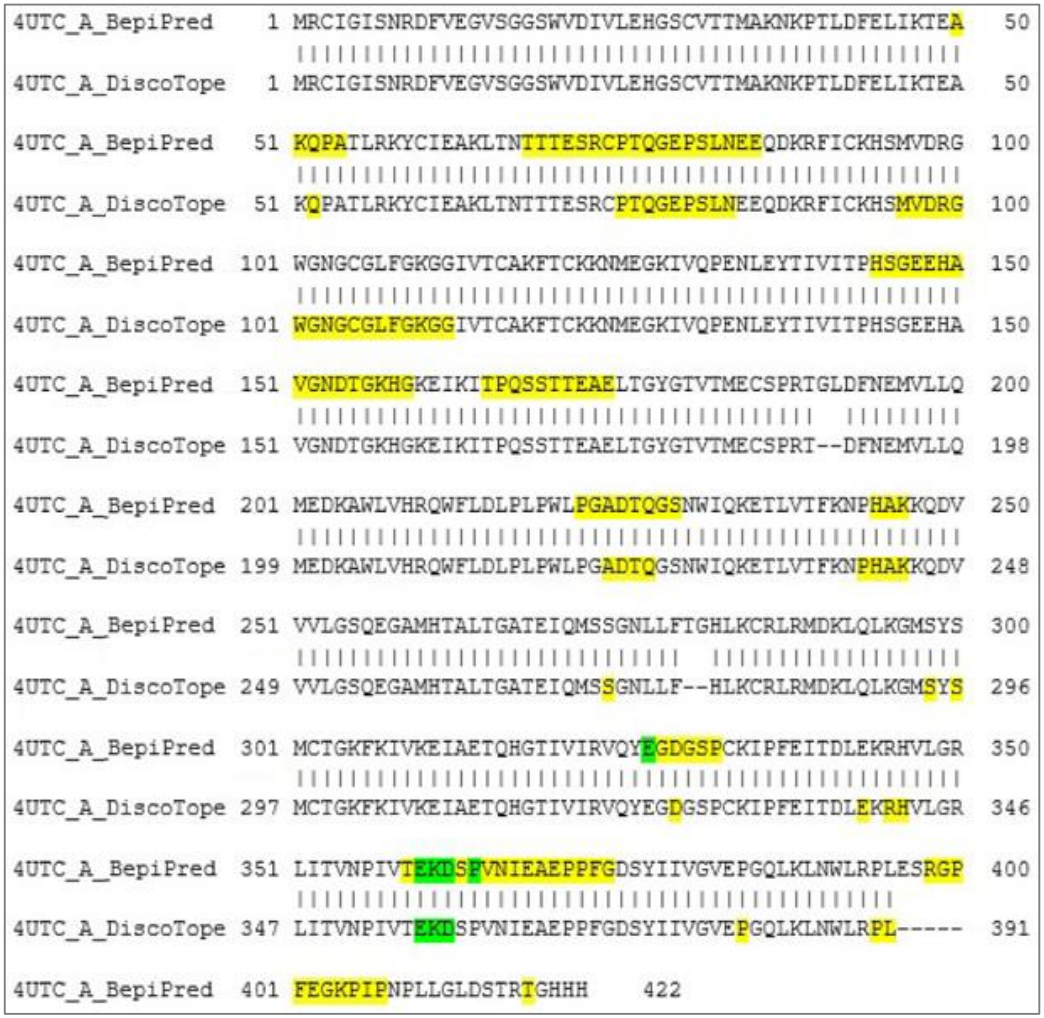
Practice Exercise

Protein alignment | Nucleotide alignment | Web services | Help & Documentation

```
#
#
#=====================================

4UTC_A_seqres      1 MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_atomse      1 MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA     50

4UTC_A_seqres     51 KQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_atomse     51 KQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG    100

4UTC_A_seqres    101 WGNGCGLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_atomse    101 WGNGCGLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA    150

4UTC_A_seqres    151 VGNDTGKHGKEIKITPQSSTTEAELTGYGTVTMECSPRTGLDFNEMVLLQ    200
                     ||||||||||||||||||||||||||||||||||||||||  |||||||||
4UTC_A_atomse    151 VGNDTGKHGKEIKITPQSSTTEAELTGYGTVTMECSPRT--DFNEMVLLQ    198

4UTC_A_seqres    201 MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_atomse    199 MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV    248

4UTC_A_seqres    251 VVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCRLRMDKLQLKGMSYS    300
                     ||||||||||||||||||||||||||||  ||||||||||||||||||||
4UTC_A_atomse    249 VVLGSQEGAMHTALTGATEIQMSSGNLLF--HLKCRLRMDKLQLKGMSYS    296

4UTC_A_seqres    301 MCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEITDLEKRHVLGR    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_atomse    297 MCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEITDLEKRHVLGR    346

4UTC_A_seqres    351 LITVNPIVTEKDSPVNIEAEPPFGDSYIIVGVEPGQLKLNWLRPLESRGP    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_atomse    347 LITVNPIVTEKDSPVNIEAEPPFGDSYIIVGVEPGQLKLNWLRPL-----    391

4UTC_A_seqres    401 FEGKPIPNPLLGLDSTRTGHHH    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

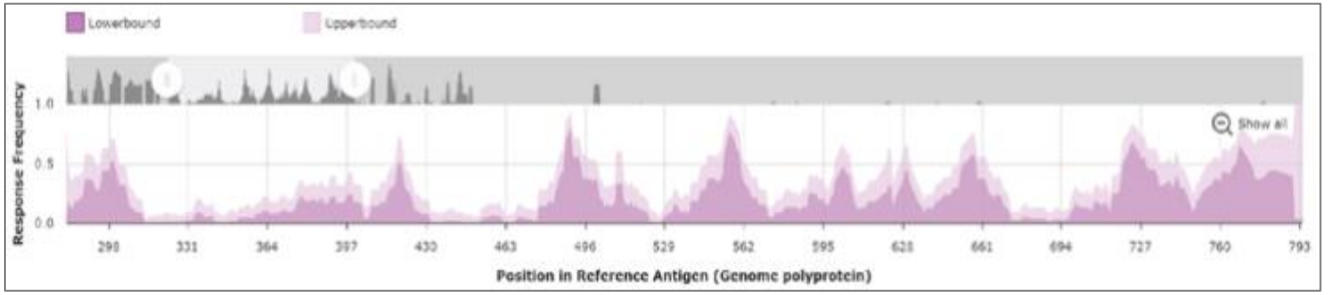


Predicted

Correctly predicted

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

Practice Exercise



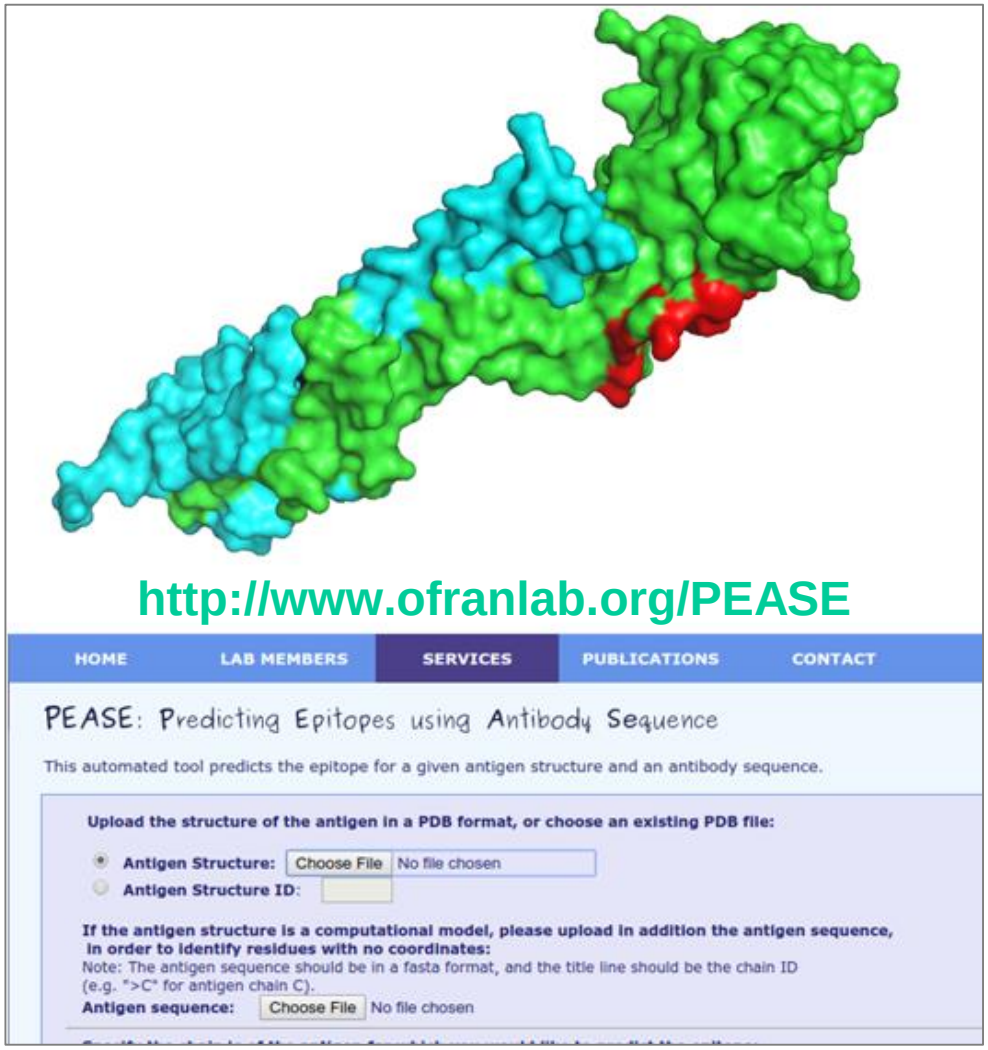

```
4UTC_A_BepiPred     1 MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA  50
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_DiscoTope    1 MRCIGISNRDFVEGVSGGSWVDIVLEHGSCVTTMAKNKPTLDFELIKTEA  50

4UTC_A_BepiPred    51 KQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG 100
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_DiscoTope   51 KQPATLRKYCIEAKLTNTTTESRCPTQGEPSLNEEQDKRFICKHSMVDRG 100

4UTC_A_BepiPred   101 WGNGCGLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA 150
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_DiscoTope  101 WGNGCGLFGKGGIVTCAKFTCKKNMEGKIVQPENLEYTIVITPHSGEEHA 150

4UTC_A_BepiPred   151 VGNDTGKHGKEIKITPQSSTTEAELTGYGTVTMECSPRTGLDFNEMVLLQ 200
                      |||||||||||||||||||||||||||||||||||  |||||||||
4UTC_A_DiscoTope  151 VGNDTGKHGKEIKITPQSSTTEAELTGYGTVTMECSPRT--DFNEMVLLQ 198

4UTC_A_BepiPred   201 MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV 250
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_DiscoTope  199 MEDKAWLVHRQWFLDLPLPWLPGADTQGSNWIQKETLVTFKNPHAKKQDV 248

4UTC_A_BepiPred   251 VVLGSQEGAMHTALTGATEIQMSSGNLLFTGHLKCRLRMDKLQLKGMSYS 300
                      |||||||||||||||||||||||||||  ||||||||||||||||||
4UTC_A_DiscoTope  249 VVLGSQEGAMHTALTGATEIQMSSGNLLF--HLKCRLRMDKLQLKGMSYS 296

4UTC_A_BepiPred   301 MCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEITDLEKRHVLGR 350
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
4UTC_A_DiscoTope  297 MCTGKFKIVKEIAETQHGTIVIRVQYEGDGSPCKIPFEITDLEKRHVLGR 346

4UTC_A_BepiPred   351 LITVNPIVTEKDSPVNIEAEPPFGDSYIIVGVEPGQLKLNWLRPLESRGP 400
                      |||||||||||||||||||||||||||||||||||||||||||
4UTC_A_DiscoTope  347 LITVNPIVTEKDSPVNIEAEPPFGDSYIIVGVEPGQLKLNWLRPL----- 391

4UTC_A_BepiPred   401 FEGKPIPNPLLGLDSTRTGHHH    422

4UTC_A_DiscoTope  392 ----------------------    391
```

Epitope residues **from the IEDB** in Dengue envelope glycoprotein

Predicted

Correctly predicted

3D Structure-based epitope prediction



- 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 I[1], Ashkenazi S[1], Peters B[1], Ofran Y[1].

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