

EPIBOT

User Guide

Table of Contents

1	What is EPIBOT	3
2	How to install EPIBOT	3
3	How to use EPIBOT	3
3.1	Menu Run/Setup.....	4
3.2	Menu Run/Prediction.....	8

1 What is EPIBOT

Several Web sites offer methods to help researchers to predict epitopes. The use of different predictors increases the epitope identification accuracy, mitigating the efforts and the financial costs in epitope identification. However, each method has a distinct user interface and each one show their results in a specific format, what makes the consolidation of those results a cumbersome task.

The EPIBOT software is a tool for automating the use of different epitope prediction algorithms to perform a more efficient epitope prediction. Using EPIBOT, users will easily work with the results from different algorithms and get a comprehensive result that consolidates all the epitopes found in a single list.

The current version of EPIBOT is configured to use the results of four epitope prediction algorithms: SYFPEITH, IEDB, NETMHC and BIMAS.

2 How to install EPIBOT

EPIBOT is made in Java so there is no need to compile or install anything. The only requirement is to have a Java Virtual Machine (JVM) installed in the computer.

To run EPIBOT the user must just unzip the installation file downloaded from the EPIBOT site and run the `epibot.jar` application.

3 How to use EPIBOT

After running EPIBOT, the user will get the windows shown in Figure 1. There are four options in the main menu.

The Main/Proxy Server option allows configuring a Proxy Server (Figure 2). It is helpful when EPIBOT will access the Internet thorough a proxy server in your local a network. The Main/About option has information about the EPIBOT team.



Figure 1: EPIBOT's initial window

EPIBOT can consolidate the result of different algorithms used to predict epitopes. This is made in two steps: qualify and predict. The Qualify step is used to make EPIBOT

assess the quality of a prediction algorithm and the Predict step is used to make EPIBOT compute the epitopes of a set of proteins selected by the user. Both steps are available in the Epitope menu.



Figure 2: Network Proxy setup

3.1 Menu Run/Setup

The Setup window is shown in Figure 3. In this option the user must inform to EPIBOT a set of proteins, each one with a known epitope. Then EPIBOT will query all preconfigured Web sites to evaluate how good they predict these known proteins. This set of proteins is named “Calibration Set”.

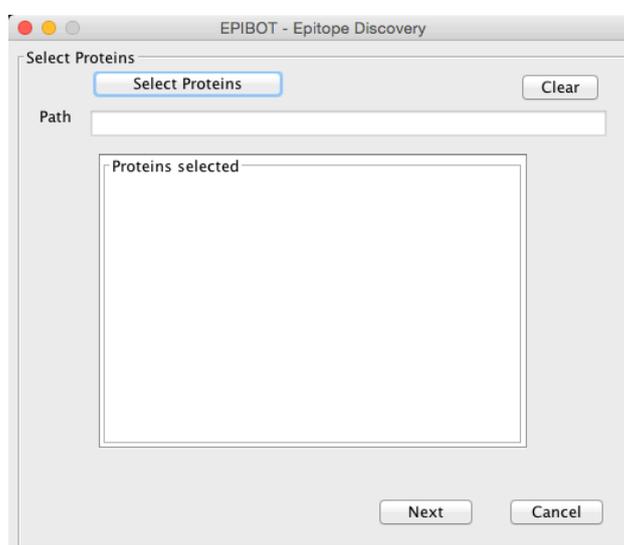


Figure 3: Setup window

The Calibration Set is informed using the option “Select Proteins. The user must use the next window to select a file in the FASTA format (Figure 4). The file may contain the set of proteins the user will use to calibrate the EPIBOT algorithm.

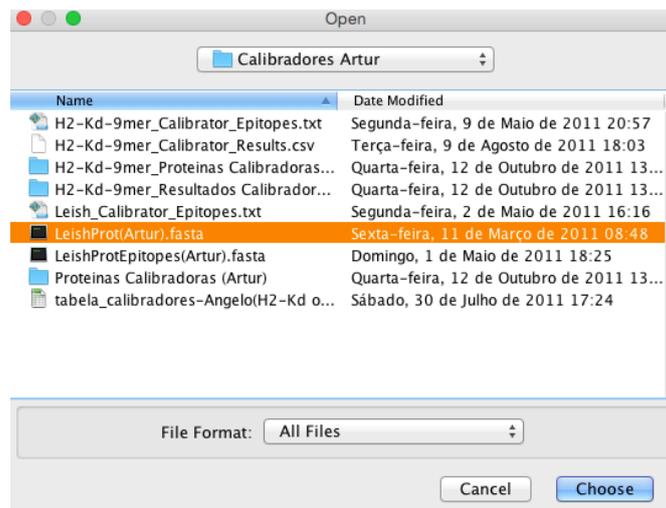


Figure 4: Window to select the FASTA file of the calibration proteins

If the file is not in FASTA format, EPIBOT will not recognize it and will show an error message (Figure 5).

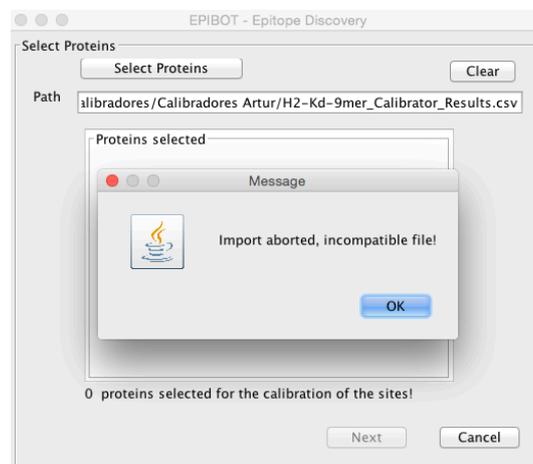


Figure 5: Error message if the file is not in FASTA format

After import the file, EPIBOT shows the ID of all proteins imported (Figure 6). The user must click Next to continue the calibration process.

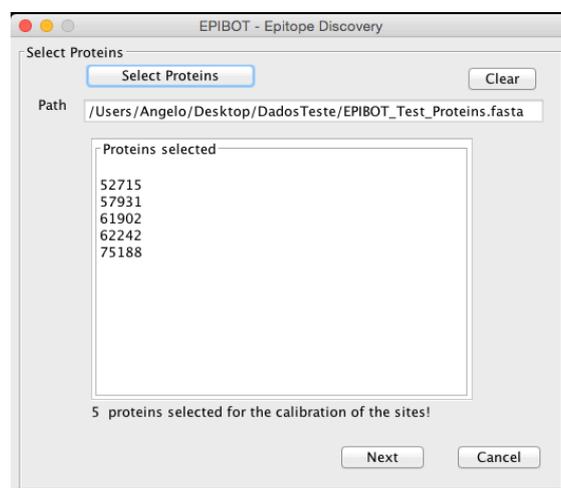


Figure 6: List of proteins imported

The next step is to inform to EPIBOT the epitopes presented in each protein of the calibration set. For this, the user must prepare a simple text file containing a list of proteins and their epitopes. Each line of the file must contain a pair **epitope,proteinID**, like shown in Figure 7. These proteins and their know epitopes are used by EPIBOT to calibrate its consolidation algorithm.

```
GGLPFSLL,75188
GIFCFRIL,62242
LGIFCFRIL,62242
LKFAFSMM,75188
SKINIEDDVI,62242
YSGYIFRDL,61902
TSICSLYQLE,52715
VHSQFNLSL,57931
WHFPFFIL,57931
```

Figure 7: A typical epitope,proteinID file

Each set of proteins and known epitope forms a calibration set that can be used as many time as needed for predict new epitopes. For this, EPIBOT allows that the user save each specific calibration. The user must ever give a name for the calibration before select the calibration file (Figure 8).

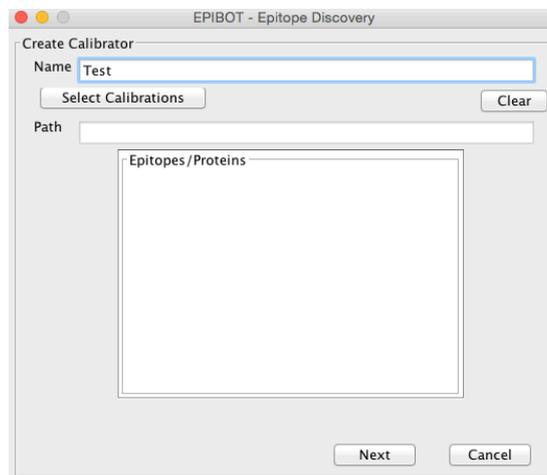


Figure 8: Window for input the calibration file. The option “Select Calibrations” is available only if there is a name in the Name field.

Once the file has been read, EPIBOT will show the list of proteins and their respective epitopes (Figure 9).

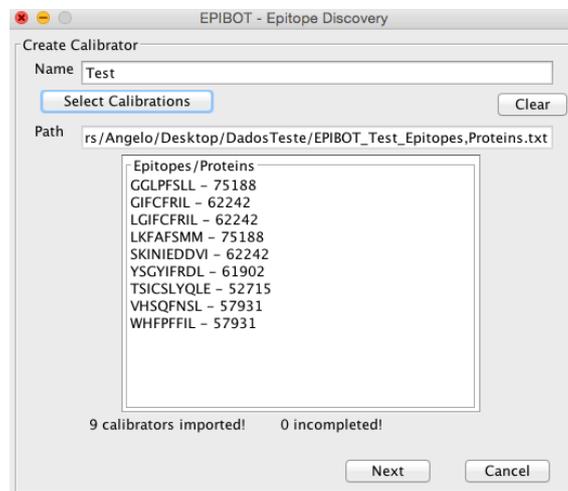
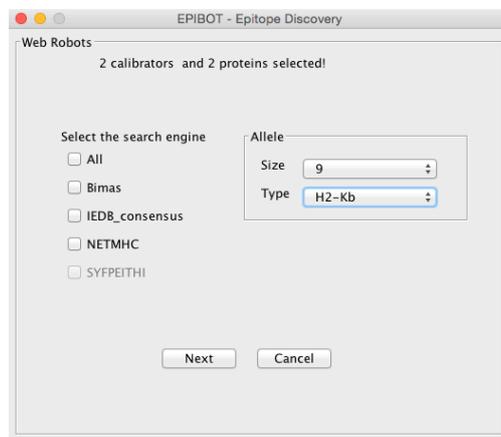


Figure 9: List of protein,epitope read from a calibration file.

In the next step, the user must select which algorithms will perform the calibration and the kind of allele (Figure 10). The available search engine will depend on the size/type option offered by the web sites. In Figure 10 one can note the the SYFPEITHI option is not available for a allele H2-Kb 9mer.

The best option is calibrate EPIBOT using all algorithms, however, since the Web sites used to access these algorithms are very dynamic and change from time to time, sometimes EPIBOT cannot connect to some site. Therefore, EPIBOT offers the possibility of selecting each algorithm individually.

**Figure 10: Window to configure the calibration procedure**

After this, EPIBOT will access all sites and perform the calibration of its algorithm. The total time for calibration varies according to the number of calibration epitopes and the number of sites selected. When the calibration process ends, EPIBOT will show a window with a score of each prediction algorithm, calculated using the EPIBOT internal algorithm (Figure 11).

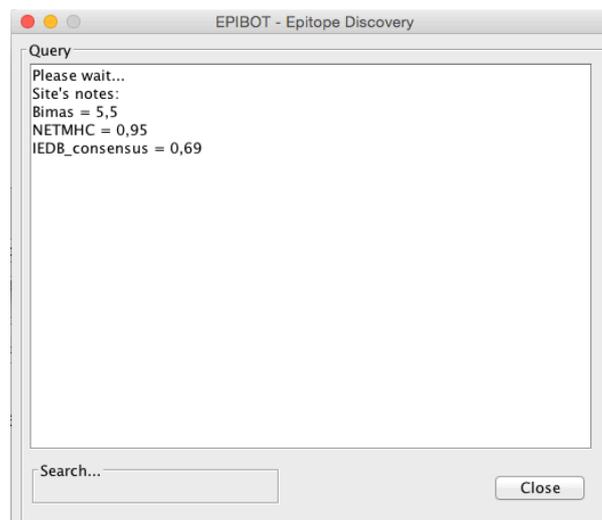


Figure 11: End of calibration process

3.2 Menu Run/Prediction

After perform the calibration, EPIBOT is ready to be used as an epitope's predictor, using the window shown Figure 11. The user may enter the protein sequence directly in the text field, or select a FASTA file as input using the option "Select File". The user also must use the "Select Calibrator" option to choose the calibrator set that will be used by EPIBOT to predict the epitopes. The "Information" box shows the kind of allele for the epitope search. Note that user cannot change the kind of allele since it was defined in the calibration process.

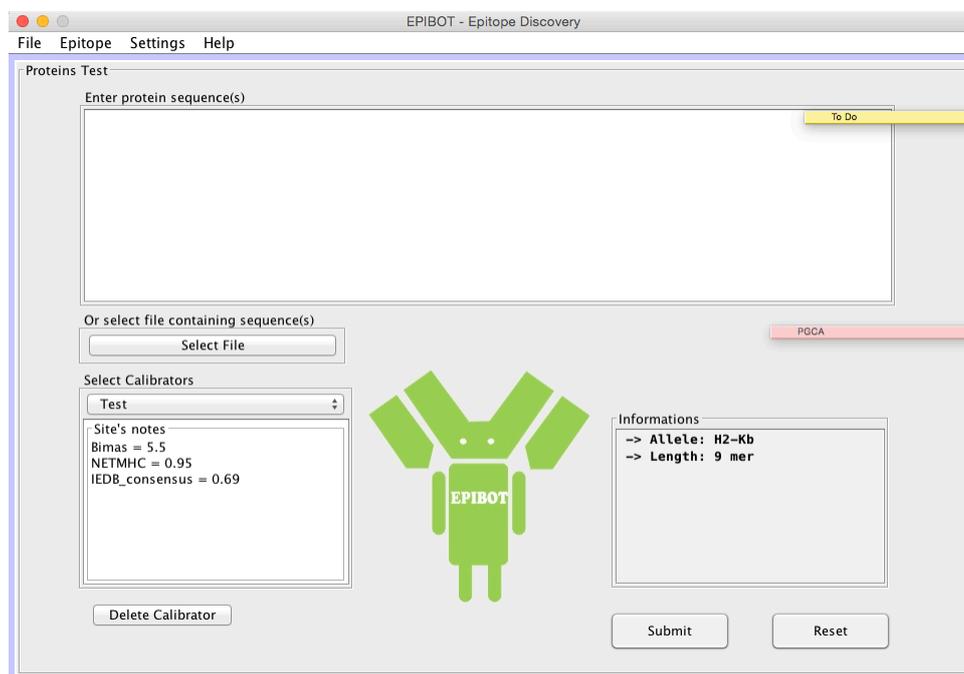


Figure 12: Prediction window

To option "Delete Calibrator" allows the user delete a calibration set if it will not be needed anymore.

After select the calibration set and the proteins of interest, the user must click "Submit" and EPIBOT will query the same prediction sites used to create the calibration set to get information about the epitopes in the proteins. The time required to conclude this procedure depends of the number of proteins and the response time of the servers of each site. At the end, EPIBOT will show a table with the rank of all predicted epitope in order of its prediction score (Figure 13).

RANK	EPITOPE	POSITION	GLOBAL_SCORE	ALLELE	SIZE	HEADER
1	SGTDKYYKL	552	5.956	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
2	DITTSYTYM	749	5.271	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
3	VSGHVYNAL	609	4.915	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
4	LFAQVNSLL	2	4.433	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
5	FDIQGKRM	254	4.425	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
6	TAPYYTKL	287	4.421	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
7	LYVRYNYS	517	3.47	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
8	SVHSQFNLS	17	3.068	H2-Kb	9	>gi 57931 emb CAA50285.1 prot...
9	SGLFPFSL	43	2.940	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
10	LQCFRLL	40	2.897	H2-Kb	9	>gi 62242 emb CAA40575.1 um...
11	RSAKEFATL	660	2.866	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
12	ITFLNYPFI	217	2.825	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
13	HTLSQNNM	797	2.815	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
14	LTYVGVNS	124	2.72	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
15	ALVMRFPL	2	2.691	H2-Kb	9	>gi 52715 emb CAA28433.1 pre...
16	FALLKTYL	14	2.612	H2-Kb	9	>gi 62242 emb CAA40575.1 um...
17	HRHFGRNM	107	2.388	H2-Kb	9	>gi 62242 emb CAA40575.1 um...
18	RFLNDFALL	9	2.363	H2-Kb	9	>gi 62242 emb CAA40575.1 um...
19	AFQMMFLSM	125	2.324	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
20	YSGYFRDL	585	2.243	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
21	IPHLTFGVL	716	2.204	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
22	NIHFFFFL	2	2.167	H2-Kb	9	>gi 57931 emb CAA50285.1 prot...
23	IRSEFLL	405	2.036	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
24	DDPNYFGL	695	1.897	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
25	ANYTYNLY	288	1.753	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
26	TSYDQCLQL	523	1.741	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
27	IEVYSEAL	155	1.717	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
28	DNLSLVDI	333	1.579	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
29	ERGFYTFM	45	1.51	H2-Kb	9	>gi 52715 emb CAA28433.1 pre...
30	GVLSYFKI	769	1.388	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
31	ETLGGYAVI	233	1.357	H2-Kb	9	>gi 62242 emb CAA40575.1 um...
32	LSNNIFRLI	693	1.33	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
33	TOGVFNQI	34	1.268	H2-Kb	9	>gi 57931 emb CAA50285.1 prot...
34	ILFLQSDL	22	1.256	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
35	LYVSRNN	429	1.2	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
36	LVMRFPLL	3	1.147	H2-Kb	9	>gi 52715 emb CAA28433.1 pre...
37	ELPISAKLL	390	1.117	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
38	LSFHSPPM	67	1.106	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
39	APSSIKAL	101	1.092	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
40	ITTSYTYML	750	1.06	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
41	NLDIAEYL	384	1.055	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
42	FIDQPRN	700	1.05	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
43	RMPFPHIL	792	1.049	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
44	TLQDQLTVL	667	1.029	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
45	RTPKGLFS	79	1.023	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
46	DPOVAQLEL	60	1.006	H2-Kb	9	>gi 52715 emb CAA28433.1 pre...
47	MVPDIDIL	730	0.998	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
48	QQPKNKIL	445	0.987	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
49	LLOSTLFL	17	0.966	H2-Kb	9	>gi 75188 pir QIIVE1 hypothetic...
50	KLQKSYIL	814	0.964	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
51	FFETWDPL	202	0.932	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...
52	RNESMIFEL	212	0.926	H2-Kb	9	>gi 61902 emb CAA34198.1 inn...

Figure 13: Rank Table

To save this result just select *File/Save* for choose the consolidated output folder (Figure 14).

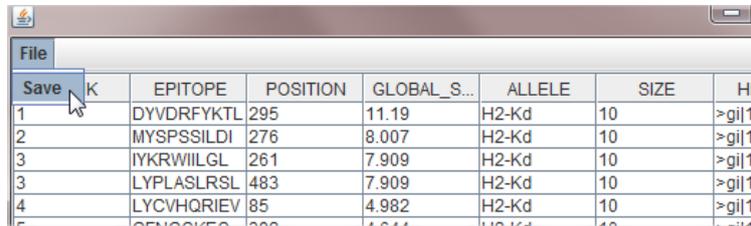


Figure 14: Saving the rank table

The result file is in CSV (comma separated) format. The first line is a header and the next lines are the predicted epitopes. A example of the first lines of a result file is shown in Figure 15

```
RANK, EPITOPE, POSITION, GLOBAL_SCORE, ALLELE, SIZE, HEADER
1,SGTDKYYKL,552,5.956,H2-Kb,9,>gi|61902|emb|CAA34198.1| inner core protein VP3 [Simian rotavirus A/SA11]
2,DITTSYTYM,749,5.271,H2-Kb,9,>gi|61902|emb|CAA34198.1| inner core protein VP3 [Simian rotavirus A/SA11]
3,VSGHVYNAL,609,4.915,H2-Kb,9,>gi|61902|emb|CAA34198.1| inner core protein VP3 [Simian rotavirus A/SA11]
4,LFAQVNSLL,2,4.433,H2-Kb,9,>gi|75188|pir|QIIVE1| hypothetical 18K protein - influenza A virus (strain A/PR/8/34)
5,FDIG
```

Figure 15: Example of a result file