

**EPdigger: an automatic dig tool of Epitope  
Peptide**

guideline

---

## catalogue

<b>1 Software Introduction.....</b>	<b>3</b>
<b>2 Operating Environment.....</b>	<b>3</b>
<b>2.1 Hardware Environment.....</b>	<b>3</b>
<b>2.2 Software Environment .....</b>	<b>3</b>
<b>3 Software Installation and Uninstallation .....</b>	<b>3</b>
<b>4 Instructions for Use.....</b>	<b>3</b>
<b>4.1 Opening and Closing the Software .....</b>	<b>3</b>
<b>4.2 Main Interface of the Software.....</b>	<b>3</b>
<b>4.3 Data Input.....</b>	<b>7</b>
<b>4.4 Sliding window length .....</b>	<b>7</b>
<b>4.5 Run.....</b>	<b>8</b>
<b>5 Technical Support .....</b>	<b>8</b>

## **1 Software Introduction**

This software uses a mathematical and computational modeling method to automatically mine epitope peptides.

## **2 Operating Environment**

### **2.1 Hardware Environment**

Ordinary PC (normal complete configuration is sufficient)

### **2.2 Software Environment**

Windows 7/8/10 platform

## **3 Software Installation and Uninstallation**

The software uses a no-installation mode, you can simply unzip the software package. To uninstall, just delete the software directory.

## **4 Instructions for Use**

### **4.1 Opening and Closing the Software**

To open the software: Double-click on the EPdigger.exe in the software directory to run and open the software.

To close the software: Click on the “x” close button in the upper right corner of the main interface.

### **4.2 Main Interface of the Software**

This software is an epitope peptide automatic mining software, and its main interface is shown in Figure 1 below.

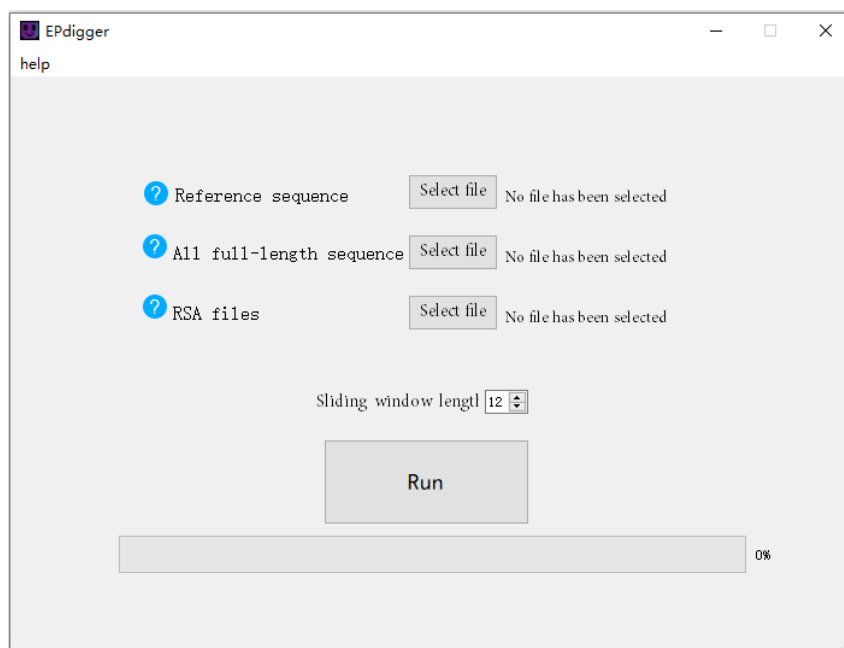


Figure 1

### 4.3 Data Preparation

To predict epitope peptides using EPdigger, you need to prepare three input files. The first file, Reference sequence, contains only one antigen protein sequence for epitope prediction and is saved in fasta format. The second file, all full-length sequences, contains more than 2 protein sequences that are highly homologous to the Reference sequence. This is used for conservative analysis of the predicted candidate epitope peptides and is also saved in fasta format. The last file, RSA file, is used to analyze the Relative Solvent Accessibility (RSA) of the predicted epitope peptides.

#### 1) Data Acquisition:

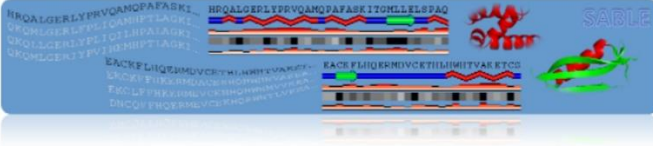
**Reference Sequence:** Save the target protein sequence (which you wish to use for peptide epitope prediction) in fasta format.

**All Full-Length Sequences:** Use the reference sequence as a query sequence, perform a Blastp search on the NCBI website, obtain all protein sequences that are homologous to the reference sequence and of equal length, and save them in fasta format. There should be no less than two sequences.

**Relative Solvent Accessibility:** Use the SABLE website (<http://sable.cchmc.org/>) as shown in Figure 2. At the location marked 1, upload the reference sequence, at the location marked 2, fill in your email. SABLE will send the analysis results to the user-specified email. Copy the data in red box from the “Relative solvent accessibility prediction (real values)” part of the analysis results in your email as shown in Figure 3, save it as a txt format file.

**SABLE** Accurate sequence-based prediction of relative Solvent Accessibility, secondary structures and transmembrane domains for proteins of unknown structure

[Documentation](#) | [Statistics](#) | [Credits](#) | [News](#) | [About us](#) | [Contacts](#)



**Downloads**

[Stand-alone version of SABLE](#)

[NCBI Pdb-BLAST](#)

[or and Swiss-Prot databases](#)

**Last request**

Tue Oct 10 02:19:00 2023 (EST)

Protein name: APA21932.1 capsid protein [Porcine circovirus 3]

Sequence length: 70

[Refresh](#)

**Last update**

An update of our HPC cluster is scheduled this Monday (April 19, 2021) through Tuesday (April 20, 2021). Any computational job requests via our web-servers will be rejected beginning 5pm (EST) today through 1pm (EST) tomorrow.

Posted by Alexey Porollo on Mon Apr 19 10:05:27 2021

Affected servers: MINNOU POLYVIEW POLYVIEW 3D POLYVIEW MM SABLE SCORPION SIFT SPIDER

[More news](#)

**Geography of users**

[Locations of visitors to this page](#)

**Target protein** [Submit](#)

Sequence name:

Amino acid sequence (using one letter codes):

Query sequence in the FASTA format file:  [选择文件](#) 未选择任何文件

[Submit](#) [Reset](#)

**Settings** [Submit](#)

Prediction goal: ☒ secondary structure ☒ solvent accessibility ☐ transmembrane domains

Server version: ☐ SABLE I ☒ SABLE II

Predictor type:









☒ Include real values in the results

☐ Send Pdb-Blast results by e-mail

☐ Results in CASP format

[Submit](#) [Reset](#)

**Our web services**

Developed by  
R. Adamczak  
A. Porollo  
J. Meller

Figure 2

```

SECTION_SA_PERCENT
Relative solvent accessibility prediction (real values)
0 -> fully Buried
100 -> fully Exposed

> 1 25
M R H R A I F R R R P R P R R R R R H R R R Y V R
47 47 30 31 21 12 13 31 32 27 44 35 40 39 29 27 29 18 18 17 36 44 31 19 44

> 26 50
R K L F I R R P T A G T Y Y T K K Y S T M N V I S
30 26 13 9 11 22 22 23 47 30 41 16 33 10 32 16 30 27 41 37 46 40 24 29 44

> 51 75
V G T P Q N N K P W H A N H F I T R L N E W E T A
20 46 35 43 52 55 35 32 33 17 14 8 17 15 5 12 5 27 8 44 39 14 42 35 38

> 76 100
I S F E Y Y K I L K M K V T L S P V I S P A Q Q T
12 20 8 23 15 6 25 6 31 25 5 17 8 20 8 16 10 16 26 24 32 15 23 38 28

> 101 125
K T M F G H T A I D L D G A W T T N T W L Q D D P
49 46 26 23 14 16 10 6 8 11 24 30 41 54 17 33 45 37 36 51 25 44 34 28 19

> 126 150
Y A E S S T R K V M T S K K K H S R Y F T P K P I
18 23 45 15 31 35 21 32 26 11 38 38 51 45 33 8 36 18 22 11 32 17 37 14 24

> 151 175
L A G T T S A H P G Q S L F F F S R P T P W L N T
8 43 36 30 43 45 26 22 33 35 22 22 24 15 17 37 32 36 26 34 20 8 8 30 18

> 176 200
Y D P T V Q W G A L L W S I Y V P E K T G M T D F
31 50 47 26 17 25 12 8 7 8 7 8 8 13 29 18 25 49 61 43 47 24 17 40 19

> 201 214
Y G T K E V W I R Y K S V L
12 16 12 29 11 4 24 8 32 7 31 41 13 13

END_SECTION

```

Figure 3

2) Example: Examples of three input files are placed in the “example” folder in the software directory, as shown in Figure 4.

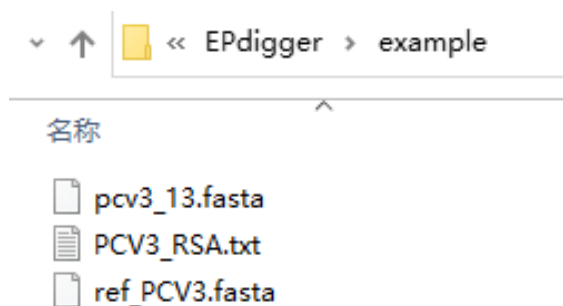


Figure 4

## 4.4 Data Input

For the first line “Reference sequence”, click the Browse button to select and upload the protein sequence file of the reference sequence. However, if there is an error in the file, a pop-up window will appear to remind you, as shown in Figure 3 below.

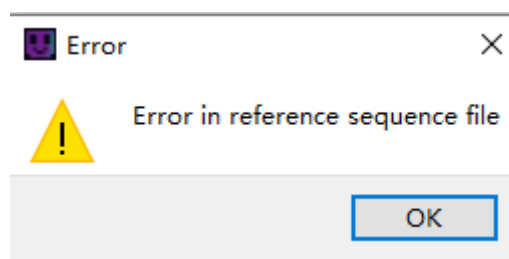


Figure 5

For the second line “All full-length sequence”, click the Browse button to select and upload all full-length protein sequence files that are the same length as the reference sequence (at least two sequences), which should not contain unknown amino acids. However, if there is an error in the file, a pop-up window will appear to remind you, as shown in Figure 4 below.

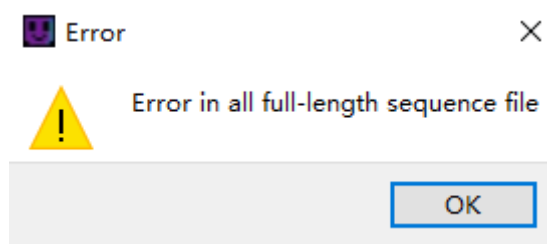


Figure 6

For the third line “RSA file”, click the Browse button to select and upload the relative solvent accessibility file of the reference sequence. However, if there is an error in the file, a pop-up window will appear to remind you, as shown in Figure 5 below.

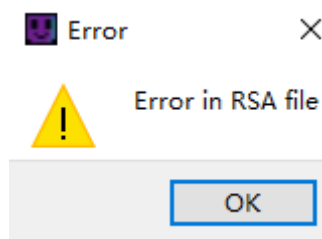


Figure 7

## 4.5 Sliding window length

You can adjust the length of the sliding window through the button, as shown in Figure 6 below.

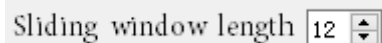
A user interface element for selecting a sliding window length. It consists of the text "Sliding window length" followed by a text box containing the number "12" and a small up/down arrow icon.

Figure 8

## 4.6 Run

After the data input and window length selection are completed, click the Run button and the software will run. During the software operation, the progress bar will run synchronously. When the program is finished, a result pop-up window will appear, as shown in Figure 6 below.

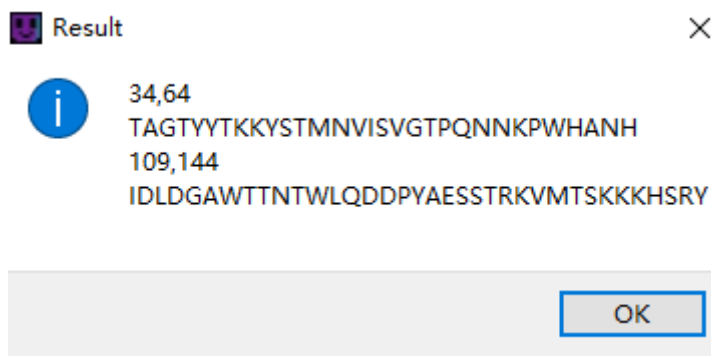


Figure 9

At the same time, the output result file will also be saved in the “result” folder under the software directory, and the file name is “result.xlsx”.

## 5 Technical Support

If you encounter any problems during the use of the system, you can describe the problems in the form of an email and send it to our mailbox. We will do our best to solve all your problems. If you have any suggestions, we are also very willing to discuss them together.