

# EpiQuest-C – Operation Manual



## Content:

[What is EpiQuest-C?](#)

[Species specificity](#)

[Entering the data](#)

[Settings for Complexity analysis](#)

[Settings for other tools](#)

[Viewing and saving the Results](#)

[Demo Sequences](#)

## What is EpiQuest-C?

EpiQuest-C creates a protein sequence profile reflecting the relative complexity of its various domains.

The Matrix C1.3 allows the identification of regions of high complexity that are (with high probability) unique for this protein and unlikely to be present in other proteins. This matrix is optimised, together with a unique algorithm, to evaluate the sequence's uniqueness/complexity from the perspective of the immune system so that antibodies recognising the selected high complexity epitope will bind to this protein only.

The uniqueness is evaluated from the perspective of amino acids and their combinations' presence in proteins of high mammals. Mainly developed to evaluate the range identified antigenic epitopes from the point of recognition by the immune system of humans, mice, rats, swine, and chickens.

Alternatively, it allows the identification of the most unique regions in proteins with a subsequent selection of the most immunogenic form of them.

## Species specificity

There is no species specificity for this calculation, however, it should be taken into account that the complexity of the regions is viewed from the point of frequency of amino acids and their combinations in the genome of higher mammals. Usually these are the species which will be immunised/exposed to (i.e. via infection) the analysed epitopes (antigen).

In other words, you can analyse protein sequences of any species origin; the results will reflect to what extent they will be viewed as unique by the immune system of high mammals (this is true for specificity to a given epitope of antibody response, and to some degree of any T-cell epitope, T-helper or CTL).

## Entering the Data

To analyse the sequence, it should be in one letter code; please simply **Paste** it into the window for **Sequence**. We recommend also entering the accession number of the sequence you use and the name of your sequence or project. This data will be present in all **Results** you save after the analysis.

**Protein Name**

**Accession Number**

**Sequence in FASTA**

Filter  Filter

Start:  End:

If your sequence contains gaps (may appear due to formatting) or contains abnormal signs (punctuation/letters that are not amino acid code) the analysis will be aborted. To avoid it, keep the “Filter” option selected.

To analyse only a fragment of the sequence, please enter the number of the first and the last amino acids of the fragment you wish to analyse. Please be aware that the numeration of the sequence starts from the first amino acid of the chosen sequence, not of the original one, and the order of numbering will be kept in all results even when you analyse a fragment.

## Settings for Complexity analysis

### Settings for Analysis

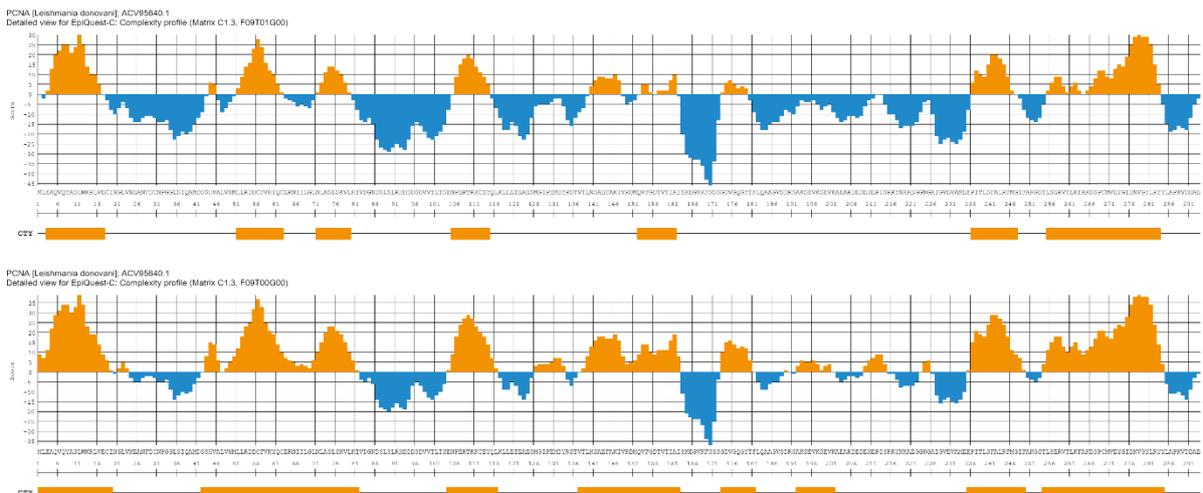
Frame size:  ∈ [3, 15]  
 Gap size:  ∈ [0, 3]  
 Threshold:  ∈ [-3, 3]  
 Peptide size:   higher  
 Results: show all  
 Sort by:   
 Chart type:

The Complexity analysis allows you to customise the settings. We recommend the **Default** settings (can be reset by the respective button) for most cases.

**Frame size** defines the size of the context sequence that will be reviewed for every position. As the most core epitope’s length is 9 amino acids, this is also the default setting for complexity analysis. For more precise mapping of the high complexity/unique regions you may decrease the Frame to 6.

**Gap size** Every position is scored as positive or negative. By default, it is 0. You may increase it up to 5 aa, in this case, the corresponding size gap of negative amino acids will be allowed between two positive fragments. In general, we would not recommend going above 2. The Gap size will affect the tabular results (somewhat longer fragments that otherwise would be divided by negative amino acids will be presented). It will have no effect on how the results will be presented in the Graphics.

**Threshold** To clearly identify the most unique regions of the proteins you may wish to raise the Threshold to 1 or 2, in this case multiple regions will be reduced to the most positive ones (see the Figure below).



This will also reduce the fragments reported in tabular form to a few shorter sequences with the highest complexity (compare the two tables below).

Threshold=0

[View](#) [Save](#) [Copy](#) **Report for "Complexity":**

Matrix = C1.3 / F = 9 / G = 0 / T = 0 / Peptide size = 9+

Start	End	Length	CI	CIR	Sequence
1	19	19	404.00	21.26	MLEAQVQYASLWKRLVECI
42	81	40	526.00	13.15	SSHVALVHMLLRDDCFVKYQCERNIILGLNLASLSKVLKI
104	116	13	219.00	16.85	ENPERTRKCEYQL
137	162	26	299.00	11.50	VTLNSAEFAKIVRDMQVFGDTVITAI
173	181	9	98.00	10.89	GDVGGQGYTF
192	201	10	37.00	3.70	AKSEVKSEVK
235	249	15	253.00	16.87	EPITLSFALRFMGIF
254	284	31	615.00	19.84	TLSERVTLKFAKDSPCMVEYGIDNVGYLRY

Threshold=1

[View](#) [Save](#) [Copy](#) **Report for "Complexity":**

Matrix = C1.3 / F = 9 / G = 0 / T = 1 / Peptide size = 9+

Start	End	Length	CI	CIR	Sequence
3	17	15	246.00	16.40	EAQVQYASLWKRLVE
51	62	12	161.00	13.42	LLRDDCFVKYQC
71	79	9	75.00	8.33	NLASLSKVL
105	114	10	119.00	11.90	NPERTRKCEY
152	161	10	34.00	3.40	QVFGDTVITIA
236	247	12	136.00	11.33	PITLSFALRFMG
255	283	29	346.00	11.93	LSERVTLKFAKDSPCMVEYGIDNVGYLRY

**Peptide size** defines whether your non-graphical results contain the set of whole positive fragments (irrespective of their length, i.e. **8 and above**) or will be presented as a set of positive N-a mers (depending on the chosen length of the fragments). If you are looking for the complexity of the selected 12-mer epitopes, use the peptide size 12, you will receive the data for all positive 12-mers.

**Sort by** defines how the results for positive fragments will be presented in the report tables: sorted according to the order of the fragments in the original sequence (*Start*), according to their *Length* (the longest goes first), cumulative complexity of the fragment *CI* (complexity index), or according to average complexity per amino-acid residue of the fragment (*CIR*)

**Chart Type.** The Chart for analysis will be presented as a graph or just bars indicating positions of antigenic fragments.

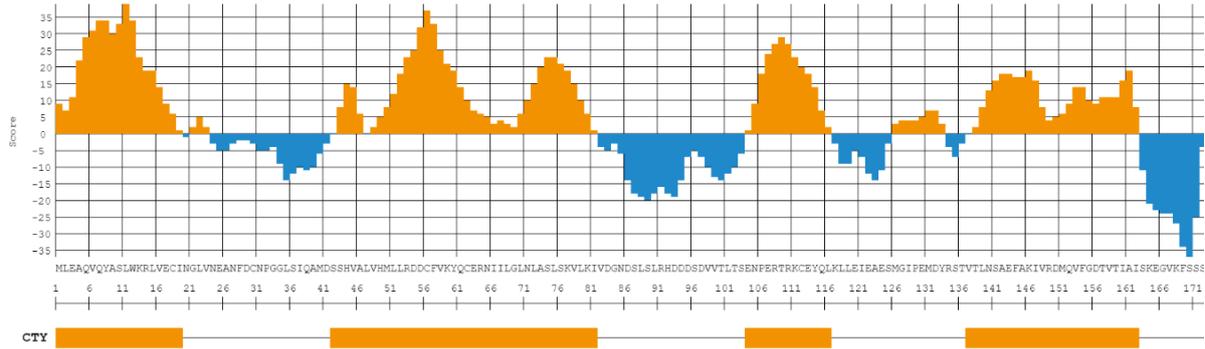
## Viewing and saving the Results

The results of the analysis are presented in 2 ways, graphical and tabular.

Graphical output will show full diagrams for Antigenicity and *bar* showing antigenic fragments according to the selected threshold level. Graphics of bars for other parameters will be shown, if the parameters were selected earlier.

The output contains the name of the sequence (name of the project), acc. number of the sequence, Matrix used, and the parameters you have chosen for the analysis of Antigenicity, where F stands for Frame, T- for Threshold and G for rge size of allowed gaps; in the emanple below F09T00G00 means that the graphic is obtained with F=9, N=0, G=0).

PCNA [Leishmania donovani]; ACV95640.1  
 Detailed view for EpiQuest-C: Complexity profile (Matrix C1.3, F09T00G00)



Please note that fragments shorter than the chosen **Peptide Size** (in this case 9 and above) are not shown in the Bar (in the figure above the peak 126-133).

You may save the entire diagram by clicking the Save diagram button above it. The image will be saved in a high resolution 96 ppi PNG format.

The **Tables** of results show the sequences for the detected antigenic fragments:

**Report for "Complexity":**

Matrix = C1.3 / F = 9 / G = 0 / T = 1 / Peptide size = 9+

Start	End	Length	CI	CIR	Sequence
3	17	15	246.00	16.40	EAQVQYASLWKRLVE
51	62	12	161.00	13.42	LLRDCCFVKYQC
71	79	9	75.00	8.33	NLASLSKVL
105	114	10	119.00	11.90	NPERTRKCEY
152	161	10	34.00	3.40	QVFGDTVIA
236	247	12	136.00	11.33	PITLSFALRFMG
255	283	29	346.00	11.93	LSERVTLKFAKDSPCMVEYGIDNVGYLRY

It contains the settings of the analysis, position of the fragment (Start-End), its length, cumulative Antigenicity index (AGI) and an index per amino-acid (AGR).

You may **Copy** the Table and just Paste it into an Excel sheet or Text (Word, .RTF etc).

You may **View** the Report in complete form, with the exact analysed sequence and print it into PDF using your browser Print function or **Save** it for your records as a standalone HTML document that may be further converted into other formats.

## Report: Analysis of Complexity profile (EpiQuest-C)

Date & Time: 26.10.2020 13:38:55

Protein Name: PCNA [Leishmania donovani]

Accession Number: ACV95640.1

MLEAQVQYASLWKRLVECIINGLVNEANFDCNPGGLSIQAMDSHVALVHMLLRDDCFVKYQC  
 ERNIIILGLNLASLSKVLKIVDGNDSLRLHDDSDVVTLTSENPERTRKCEYQLKLEIEAE  
**Sequence:** SMGIPENDYRSTVTLNSAEFAKIVRDMQVFGDVTIAISKEGVKFSSSGDVGQGYTFLQAAG  
 VSDRSKSEVKSEVKAEARDEDEHEPISRRYNKAEGNGAIGVEVAMEEPIITLSFALRFMGIF  
 FAKGSTLSERVTLKFAKDSPCMVEYGIDNVGYLRYYLAPKVVDAE

Fragment: 1 - 293

Matrix: C1.3

Frame size: 9

Gap: 0

Threshold: 0

Peptide size: 9 higher

Sorted by: Start

Start	End	Length	CI	CIR	Sequence
1	19	19	404.00	21.26	MLEAQVQYASLWKRLVECI
42	81	40	526.00	13.15	SSHVALVHMLLRDDCFVKYQCERNIIILGLNLASLSKVLKI
104	116	13	219.00	16.85	ENPERTRKCEYQL
137	162	26	299.00	11.50	VTLNSAEFAKIVRDMQVFGDVTIAI
173	181	9	98.00	10.89	GDVGQGYTF
192	201	10	37.00	3.70	AKSEVKSEVK
235	249	15	253.00	16.87	EPITLSFALRFMGIF
254	284	31	615.00	19.84	TLSERVTLKFAKDSPCMVEYGIDNVGYLRYYY

### Demo Sequences

For Demo and training purposes, we supply several sequences to try and test the program.

You may choose them by opening the drop-down list **-Select a demo sequence-**.

Clear

---Select a demo sequence---
 

---Select a demo sequence---  
**cadherin 1 [Xenopus laevis]; AAA61489.1**  
**DNA polymerase alpha [Plasmodium falciparum]; AAB28217.1**  
**keratin 18 [Xenopus laevis]; NP\_001089734.1**  
**PCNA [Leishmania donovani]; ACV95640.1**  
**ssDNA-binding protein [Yersinia pestis]; AAM84170.1**

Enter the

Start:  End:

Filter

You may clear the window for the sequence, name of the project and accession number by pressing the button **Clear**.

[In our blog](#) we further show how to use the program in combination with other ones to address different research tasks.