# **C-Scanner - Operation Manual**



Content: <u>What is C-Scanner?</u> <u>Species specificity</u> <u>Entering the data</u> <u>Settings for analysis of the relative Complexity of epitopes</u> <u>Viewing and saving the Results</u> <u>Demo Sequences</u>

# What is C-Scanner?

The C-Scanner allows you to rank the selected epitopes (or any other peptide sequences) according to their complexity.

When selecting epitope sequences using various software (EpiQuest-B, A or other platforms) you would like to end up with a *unique* epitope – whether you plan developing a specific antibody or using this peptide epitope as an antigen in antibody assays.

The C-Scanner allows you to scan a group of peptide sequences and determine the most unique ones. Uniqueness of epitopes greatly define the specificity of an antibody prepared against it i.e. the sequences with low complexity will likely be present in proteins other that the target one.

The software also allows you to define a shorter, more specific, sequence within the long epitope sequences, and thus ensure the specificity of epitope used in your antibody or assay development project.

#### **Species specificity**

The matrix C1.3 is based on the complexity of proteins in *higher mammals*. Sequences from i.e. lower invertebrates, fungi, etc. may receive a lower overall score in comparison to mammals, but still, the scanning allows you to establish the most unique sequences. Whether making an antibody or looking for epitopes recognised by natural antibodies, we aim to look at the immune response in higher mammals, so the uniqueness of the tested epitopes should be defined for these types of animals.

## **Entering the Data**

Project Na	me/Code		
Code	Sequence		
X292	GVITYSILNQEPKEPT		
X573	GTLVLNLLDVNDNGPFLEPQQESFCQKDPG		
X501	ATYTAQDPDKEQNQ		
X119	REGHRHRQDLFSGKHSHHPK		
X423	TDTGNIGLLKTVKGLDYE		
<b>X</b> 396	EDIEGTDAWNA		*
	CADH1 XENLA B-EPITOPES	~	Load demo

To test the program, select Demo sequences and click *Load demo*. In the window for sequences the two columns will appear: *Code* and *Sequence*.

To empty the window, click *Clear*.

To load your own data, it should be presented in *XLS* or *XLSL* format (MS-*Excel*) and to contain only *one dataset* with data per file. The data should be presented as two columns. The first should be named and contain *Code* (just your number or other identifier for the peptide; we do not recommend going above 15 signs, letters, numbers, and dash signs allowed) and *Sequence* (sequences of the peptides to be tested in a single letter code; we do not recommend going above 100 amino acids each,).

You may *Browse* for the location of the file and click on it to load.

### **Settings for Complexity analysis**

Settings for Analysis	is	/S	lv	а	n	Α	r	fo	S	q	n	tti	Se	
-----------------------	----	----	----	---	---	---	---	----	---	---	---	-----	----	--

Matrix: C1.3 V							
Frame size: 6 ∈ [3, 15]							
Peptide size: 9 $\in$ [6, 20]							
Threshold: $0 \in [-4, 4]$							
Sort by: CI BEST V							
Default							

Currently only one *Matrix* C1.3 – is available.

You can adjust your cut-off level for weak/functionally negative epitopes by changing the *Threshold* (positive numbers to decrease the sensitivity, negative – to increase). We do not recommend going beyond (-2) or (+2).

*Frame size* defines the size of the context analysed for every position in the sequence, we recommend keeping it at 6 (the default value).

**Peptide size** defines the size of epitope you are looking for in your peptides. If the tested sequences are of different length (i.e. your data set includes 10-mers, 20-mers, other), we recommend keeping it at 9, since very few epitopes and practically none of the specific ones are shorter than 9 amino acids. In case you wish to compare a set of peptides of different length and to select the best of them, set the size of the shortest one as the reported peptide size.

Sort by. You can select the data to be reported in several ways:

**Start:** the results will be sorted according to the order of the analysed peptides in your data file. If you analyse 19-20 mers and have defined the *Peptide size* as 9, your results will contain multiple 9-mers for every original peptide with AGI (antigenicity) defined for every 9-mer. In Results the data will be sorted according to the order of the peptide in the original sequence.

Code	Sequence	Start	End	Peptide	CI	CPB	
X292	GVITYSILNQEPKEPT	1	9	GVITYSILN	90	10	•
X292	GVITYSILNQEPKEPT	2	10	VITYSILNQ	57	6	
X292	GVITYSILNQEPKEPT	3	11	ITYSILNQE	21	2	
X292	GVITYSILNQEPKEPT	4	12	TYSILNQEP	-12	-1	
X292	GVITYSILNQEPKEPT	5	13	YSILNQEPK	-37	-4	
X292	GVITYSILNQEPKEPT	6	14	SILNQEPKE	-55	-6	
X292	GVITYSILNQEPKEPT	7	15	ILNQEPKEP	-63	-7	
X292	GVTTYSTI NOFPKEPT	8	16	I NOFPKEPT	-63	- 7	*

View Save Report for "Analysis of epitopes complexity":

*CI*: all results will be sorted according to the order of the analysed peptides in the data file, for every analysed sequence they will be presented from the best to the worst.

*CI ALL:* the results will be sorted according to the AGI values of the detected peptides of the defined peptide length (say, 9) irrespective of their origin from different original sequences, from the best to the worst.

	View Save Report for '	'Analys	is of e	pitopes complexity":			
Code	Sequence	Start	End	Peptide	CI	СРВ	
X669	NVTKLHITICQ	1	9	NVTKLHITI	162	18	
X669	NVTKLHITICQ	2	10	VTKLHITIC	155	17	
X852	ALNDWGPRFTKLADMYGGDED	1	9	ALNDWGPRF	146	16	
X852	ALNDWGPRFTKLADMYGGDED	2	10	LNDWGPRFT	139	15	
X669	NVTKLHITICQ	3	11	TKLHITICQ	137	15	
X852	ALNDWGPRFTKLADMYGGDED	3	11	NDWGPRFTK	135	15	
X852	ALNDWGPRFTKLADMYGGDED	4	12	DWGPRFTKL	125	13	
X852	ALNDWGPRETKLADMYGGDED	5	13	WGPRFTKI A	102	11	*

*CI BEST:* the data will be presented by *only* the best peptide form every sequence in the order from the highest to the lowest (here only 1 9-mer is shown for the original 4 peptides of various lengths.

	View Save Report for	"Analys	is of e	pitopes complexity":			
Code	Sequence	Start	End	Peptide	CI	CPB	
X669	NVTKLHITICQ	1	9	NVTKLHITI	162	18	
X852	ALNDWGPRFTKLADMYGGDED	1	9	ALNDWGPRF	146	16	
X292	GVITYSILNQEPKEPT	1	9	GVITYSILN	90	10	
X423	TDTGNIGLLKTVKGLDYE	8	16	LLKTVKGLD	82	9	
X634	AIVTGQSILELRP	1	9	AIVTGQSIL	60	6	
X158	SENEKGPFPKRI	4	12	EKGPFPKRI	59	6	
X119	REGHRHRQDLFSGKHSHHPK	3	11	GHRHRQDLF	57	6	
X396	FDTFGTDAWNA	1	9	FDTFGTDAW	57	6	*

#### Viewing and saving the Results

You may *View* all results in new window or *Save* them for your records as HTML file. You can always import such file into your spreadsheet program or simply copy the selected areas of interest in other file formats.

## Report: Relative complexity of epitopes (C-Scanner)

Date & Time: 05.11.2020 19:08:51 Project name: Program: EpiQuest C-Scanner v1.0.0.1 Matrix: C1.3 Peptide size: 9 Frame size: 6 Threshold: 0 Sorted by: CIBEST

Sequence	Start	End	Peptide	CI	CPB
NVTKLHITICQ	1	9	NVTKLHITI	162	18
ALNDWGPRFTKLADMYGGDED	1	9	ALNDWGPRF	146	16
GVITYSILNQEPKEPT	1	9	GVITYSILN	90	10
TDTGNIGLLKTVKGLDYE	8	16	LLKTVKGLD	82	9
AIVTGQSILELRP	1	9	AIVTGQSIL	60	6
SENEKGPFPKRI	4	12	EKGPFPKRI	59	6
REGHRHRQDLFSGKHSHHPK	3	11	GHRHRQDLF	57	6
EDIEGTDAWNA	1	9	EDIEGTDAW	57	6
PQYRPRPANPDEIGNFIDENLN	10	18	PDEIGNFID	50	5
GEEDQDFDLSQLHRGLDARP	4	12	DQDFDLSQL	43	4
YIKNPAKMKD	1	9	YIKNPAKMK	37	4
GTLVLNLLDVNDNGPFLEPQQESFCQKDPG	21	29	QESFCQKDP	20	2
VMDTNDNPPVFD	4	12	TNDNPPVFD	15	1
IIIKVQDQN	1	9	IIIKVQDQN	-13	-1
ATYTAQDPDKEQNQ	1	9	ATYTAQDPD	-15	-1
KKVVKEPLL	1	9	KKVVKEPLL	-21	-2
ASLSSLNSPNSDLDQDY	9	17	PNSDLDQDY	- 25	-2
	Sequence NVTKLHITICQ ALNDWGPRFTKLADHYGGDED GVITYSILNQEPKEPT TDTGNIGLLKTVKGLDYE AIVTGQSILELRP SENEKGPFPKRI REGHRHRQDLFSGKHSHHPK EDIEGTDAWNA PQYRPRANPDEIGNFIDENLN GEEDQDFDLSQLHRGLDARP YIKNPAKMKD GTLVLNLLDVNDNGPFLEPQQESFCQKDPG VMDTNDNPPVFD IIIKVQQN ATYTAQDPDKEQNQ KKVVKEPLL ASLSSLNSPNSDLDQDY	SequenceStartNVTKLHITICQ1ALNDWGPRFTKLADMYGGDED1GVITYSILNQEPKEPT1TDTGNIGLLKTVKGLDYE8AIVTGQSILELRP1SENEKGPFPKRI4REGHRHRQDLFSGKHSHHPK3EDIEGTDAWNA10GEEDQDFDLSQLHRGLDARP4YIKNPAKMKD1GTLVLNLLDVNDNGPFLEPQQESFCQKDPG21VMDTNDNPVFD4IIIKVQQN1ATYTAQDPDKEQNQ1KKVVKEPLL1ASLSSLNSPNSDLDQDY9	SequenceStartEndNVTKLHITICQ19ALNDWGPRFTKLADMYGGDED19GVITYSILNQEPKEPT19TDTGNIGLLKTVKGLDYE816ALVTGQSILELRP19SENEKGPFPKRI412REGHRHRQDLFSGKHSHHPK311EDIEGTDAWNA1018GEEDQDFDLSQLHRGLDARP412YIKNPAKMKD19GTLVLNLLDVNDNGPFLEPQQESFCQKDPG2129VMDTNDNPVFD412IIIKVQDQN19ATYTAQDPDKEQNQ19ASLSSLNSPNSDLDQDY917	SequenceStartEndPeptideNVTKLHITICQ19NVTKLHITIALNDWGPRFTKLADMYGGDED19ALNDWGPRFGVITYSILNQEPKEPT19GVITYSILNTDTGNIGLLKTVKGLDYE816LKTVKGLDAIVTGQSILELRP19AIVTGQSILSENEKGPFPKRI412EKGPFPKRIREGHRHRQDLFSGKHSHHPK311GHRHRQDLFEDIEGTDAWNA19EDIEGTDAWPQYRPRANPDEIGNFIDENLN1018PDEIGNFIDGEEDQDFDLSQLHRGLDARP412DQDFDLSQLYIKNPAKMKD19YIKNPAKMKGTLVLNLLDVNDNGPFLEPQQESFCQKDPG2129QESFCQKDPIIIKVQQN19IIIKVQQNATYTAQ0PDKEQNQ19ATYTAQ0PDKKVVKEPLL19KVVKEPLLASLSSLNSPNSDLDQDY917PNSDLDQDY	SequenceStartEndPeptideCINVTKLHITICQ19NVTKLHITI162ALNDWGPRFTKLADMYGGDED19ALNDWGPRF146GVITYSILNQEPKEPT19GVITYSILN90TDTGNIGLLKTVKGLDVE816LKTVKGLD82AIVTGQSILELRP19AIVTGQSIL60SENEKGPFPKRI412EKGPFPKRI59REGHRHRQDLFSGKHSHHPK311GHRHQDLF57EDIEGTDAWNA1018PDEIGNFID50GEEDQDFDLSQLHRGLDARP412DQDFDLSQL43YIKNPAKMKD19YIKNPAKMK37GTLVLNLLDVNDNGPFLEPQQESFCQKDPG2129QESFCQKDP20VMDTNDNPPVFD412TNDNPPVFD15IIIKVQQN19ATYTAQDPD-15KKVKEPLL19ATYTAQPD-25ALYTAQDPDKEQNQ1917PNSDLDQDY-25

#### **Demo Sequences**

For Demo, we supply several files with either tested selections of epitopes or selected sets of sequences for training.