

EpiQuest-A Manual



Content:

[What is EpiQuest-A?](#)

[Species specificity](#)

[Entering the data](#)

[Settings for Accessibility analysis](#)

[Viewing and saving the Results](#)

[Demo Sequences](#)

What is EpiQuest-A?

EpiQuest-B builds an antigenicity profile for a protein sequence, enabling the discovery of linear B-epitopes that may *potentially* be immunogenic. However, most of the sequences hidden from the surface of a folded protein molecule, even being antigenic when offered to the immune system as an isolated peptide sequence, will not induce a humoral immune response. Moreover, if antibodies to such “hidden” epitopes are prepared, they will not recognise the mature, folded protein and, at best, will react with proteins in western blot containing denatured proteins.

Additionally, only the epitopes well exposed at the surface of the molecule will perform as antigens in response to proteins *in vivo*. If you are looking for epitopes that will likely be a target of an immune response against a virus or bacteria (or another immunogen), especially for diagnostic detection of such immune response, you need to look to accessible “sequences”.

This is what EpiQuest-A was designed for.

The basic values for the relative likelihood of an amino acid being exposed at the surface of the molecule used are reported by *Janin et al* (1978) with corrections and advancements brought by othersⁱ. We made some additional modifications to potential for exposure on account of the environment (the protein context) in which a particular amino acid is located. The program also employs a quite different and unique algorithm that evaluates the probability of surface exposure for every amino acid position in the analysed sequence. In short, EpiQuest-A is quite dissimilar from programs based on *Emini et al* (1985).

This allows to discern well exposed sequences more clearly from hidden and semi-hidden, and the program is especially invaluable for detecting linear epitopes that elicit immune response against the native antigen.

Species specificity

We have seen no limitations of the program in analysis of proteins from different species.

Entering the Data

The interface consists of the following elements:

- Protein Name:** A text input field.
- Accession Number:** A text input field.
- Clear:** A button to reset the form.
- Select a demo sequence---**: A dropdown menu.
- Sequence in FASTA:** A large text area with the instruction "Enter the aminoacid sequence in 1-letter code (FASTA)".
- Filter:** A checkbox, currently checked, with a blue box highlighting it.
- Start:** A numeric input field with the value 0.
- End:** A numeric input field with the value 0.

To analyse the sequence, please simply **Paste** it into the window for **Sequence**. The sequence must be in a single-letter format.

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.

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 position 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 and will be kept during the analysis of the fragment. I.e. if the first aa of the fragment was 101, it will be indicated as 101 in all further analysis and Results.

However, if you paste into sequence window a fragment of some sequence, its first amino acid will be referred as in position 1, not the position in the original sequence - source of the fragment.

Settings for Accessibility analysis

Settings for Analysis

The settings are as follows:

- Frame size:** 9 (range: [3, 15])
- Gap size:** 0 (range: [0, 5])
- Threshold:** 0 (range: [-3, 3])
- Peptide size:** 8 (checkbox higher)
- Results:** show all
- Sort by:** Start
- Chart type:** Graph
- Default:** A button to reset settings.

EpiQuest-A allows you to customise the settings. We recommend keeping 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. For accessibility, we recommend using the setting 6, as this setting defines the exposed fragments with high precision. When you are looking for some grooves, where the sequence is half-hidden, we suggest using the setting 3. Frame 6 to 9 provides you secure information on whether the fragment is accessible at the surface of the folded protein.

Gap size: normally, your results will be reported as uninterrupted positive sequences of length equal to or exceeding the length defined in **Peptide Size** settings. Sometimes small gaps may interrupt the otherwise long fragment. This may be a result of occasional enrichment of the region with amino acids of a certain type but does not mean that the sequence is inaccessible. To avoid such interruption, you may use the settings for Gap Size from 1 to 3, in rare situations up to 5. The recommended default value is 1.

Threshold (Th): You may want to see the sequences that are very well exposed at the surface, the most accessible regions. In this case you may rise the Threshold to 1 or even 2. When looking for transmembrane regions, etc., keep Threshold =0.

Peptide size. While the **Graph** in Results represents the values for every amino acid of the sequence with the peak positive and negative values, the Bar and Tabular reports show only *positive fragments* that are equal or above the length of a sequence defined in Peptide Size. For that the **“and higher”** option should be selected. If it is not selected, the Table will show all positive peptides of the chosen parameters (i.e. 8-mers, overlapping, with at least one aa difference).

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* (longest goes first), *the cumulative probability* of surface exposure of the fragment *SEPI* (surface exposure index) or SEPI per amino-acid residue of the fragment (*SEPR*).

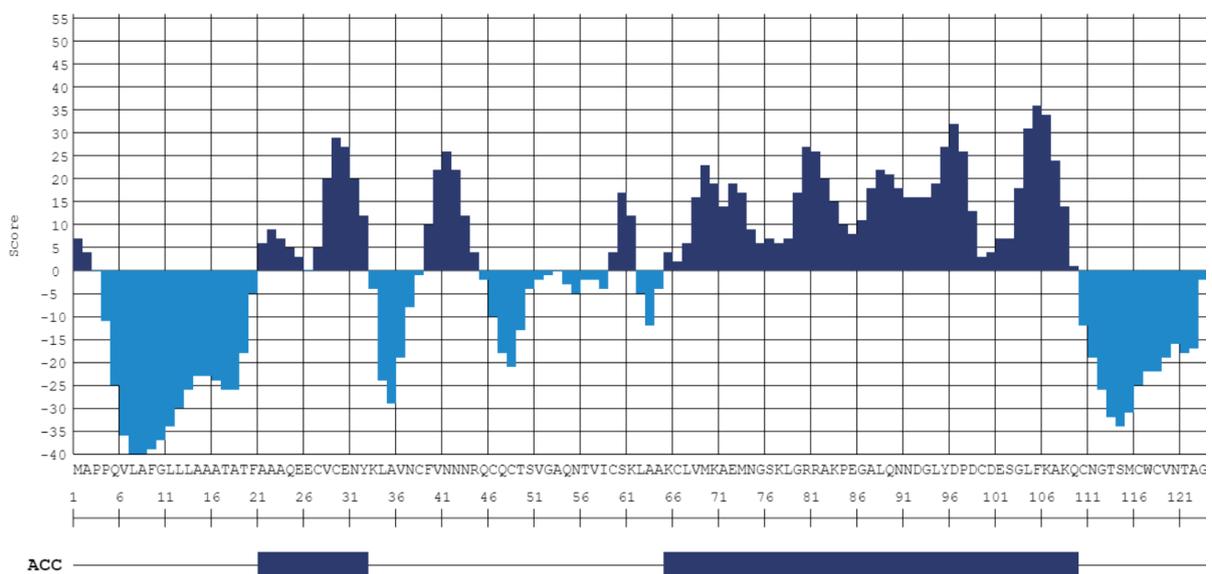
Chart Type: the Chart for analysis will be presented as a **Graph** (showing values for every aa of the sequence) or just **Bar** (the latter indicating only the positions of positive sequences *with length equal or above* the selected in **Peptide Size**). If you selected the Peptide size =9, the positive fragments of 8aa or shorter will not be shown in the Bar. In the stand-alone program both types of graphical output are presented in the results.

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 Accessibility (surface exposure), while *bar* shows only positive fragments of length equal or exceeding the chosen Peptide Size. In the picture below, for the N-terminal fragment of Ep-CAM, you may see that small positive peaks at positions 41 and 61 are not shown on the bar as they are shorter than 9 amino-acids or longer settings in **Peptide Size**.

Epithelial cell adhesion molecule (EPCAM); P16422.2
Detailed view for EpiQuest-A: Surface exposure/accessibility profile (Matrix EM3.2, F06T00G00)



The Graphical output contains the name of the sequence (name of the project), accessory number of the sequence, Matrix used, and the parameters you have chosen for the analysis, where F stands for Frame, T for Threshold and G for the sizes of allowed gaps (in the image shown F=6, T=0, G=0: F06T00G00)

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

The Tables of results show the sequences for the detected antigenic fragments. They will be reported in order of parameters chosen in Settings.

Report for "Accessibility":

Matrix = EM3.2 / F = 6 / G = 0 / T = 0 / Peptide size = 8+

Start	End	Length	SEPI	SEPR	Sequence
21	32	12	143.00	11.92	AAAEQEEVCENY
65	109	45	712.00	15.82	KCLVMKAEMNGSKLGRRAKPEGALQNNGLYDPDCDESGLFKAKQ
124	133	10	236.00	23.60	VRRTDKDTEI
144	160	17	641.00	37.71	IIIELKHKAREKPYDSK
163	178	16	359.00	22.44	RTALQKEITTRYQLDP
193	206	14	406.00	29.00	IDLVQNSSQKTQND
212	246	35	731.00	20.89	VAYYFEKDVKGESLFHSSKMDLTVNGEQLDLDPGQ
248	258	11	247.00	22.45	LIYYVDEKAPE
286	314	29	757.00	26.10	LVISRKKRMAKYEKAEIKEMGEMHRELNA

It contains the settings of the analysis, position of the fragment (Start-End), its length, cumulative Surface Exposure index (SEPI) and an index per amino-acid (SEPR).

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 Surface exposure/accessibility profile (EpiQuest-A)

Date & Time: 27.10.2020 13:53:40

Protein Name: Epithelial cell adhesion molecule (EPCAM)

Accession Number: P16422.2

Sequence: MAPPQVLAFLGLLAAATATFAAAEQEEVCENYKLVNCFVNNRQCQCTSVGAQNTVICSKL
AAKCLVMKAEMNGSKLGRRAKPEGALQNNGLYDPDCDESGLFKAKQCNCTSMCWCVNTAGV
RRTDKDTEITCSERVRTYWIIEELKHKAREKPYDSKSLRTALQKEITTRYQLDPKFITSILY
ENNVITIDLVQNSSQKTQNDVDIADVAYYFEKDVKGESLFHSSKMDLTVNGEQLDLDPGQTL
IYYVDEKAPEFSMQGLKAGVIAIVVVVIAVVAGIVVVLVISRKKRMAKYEKAEIKEMGEMHR ELNA

Fragment: 1 - 314

Matrix: EM3.2

Frame size: 6

Gap: 0

Threshold: 0

Peptide size: 8 higher

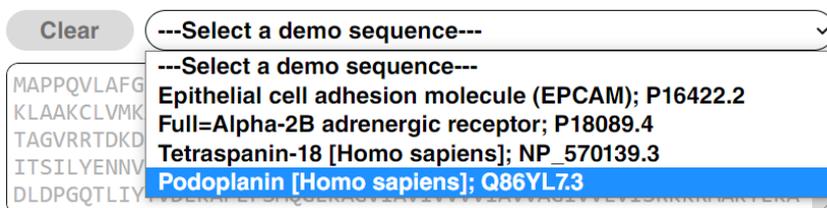
Sorted by: Start

Start	End	Length	SEPI	SEPR	Sequence
21	32	12	143.00	11.92	AAAEQEEVCENY
65	109	45	712.00	15.82	KCLVMKAEMNGSKLGRRAKPEGALQNNGLYDPDCDESGLFKAKQ
124	133	10	236.00	23.60	VRRTDKDTEI
144	160	17	641.00	37.71	IIIELKHKAREKPYDSK
163	178	16	359.00	22.44	RTALQKEITTRYQLDP
193	206	14	406.00	29.00	IDLVQNSSQKTQND
212	246	35	731.00	20.89	VAYYFEKDVKGESLFHSSKMDLTVNGEQLDLDPGQ
248	258	11	247.00	22.45	LIYYVDEKAPE
286	314	29	757.00	26.10	LVISRKKRMAKYEKAEIKEMGEMHRELNA

Demo Sequences

For Demo and training purposes, we supply several sequences with known with known accessible and inaccessible areas, including transmembrane domains and areas mapped by antibodies to various linear epitopes.

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



Clear ---Select a demo sequence---

---Select a demo sequence---

Epithelial cell adhesion molecule (EPCAM); P16422.2

Full=Alpha-2B adrenergic receptor; P18089.4

Tetraspanin-18 [Homo sapiens]; NP_570139.3

Podoplanin [Homo sapiens]; Q86YL7.3

MAPPQVLAFG
KLAAKCLVMK
TAGVRRTDKD
ITSILYENNV
DLDPGQTLIV

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

Related literature:

Emini, E.A., Hughes, J.V., Perlow, D.S., and Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology* 55, 836–839.

Janin, J., Wodak, S., Levitt, M., and Maigret, B. (1978). Conformation of amino acid side-chains in proteins. *Journal of Molecular Biology* 125, 357–386.

Koehler, J., Woetzel, N., Staritzbichler, R., Sanders, C.R., and Meiler, J. (2009). A Unified Hydrophobicity Scale for Multi-Span Membrane Proteins. *Proteins* 76, 13–29.

MacCallum, J.L., and Tieleman, D.P. (2011). Hydrophobicity scales: a thermodynamic looking glass into lipid–protein interactions. *Trends in Biochemical Sciences* 36, 653–662.

Moon, C.P., and Fleming, K.G. (2011). Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc Natl Acad Sci U S A* 108, 10174–10177.

Pace, C.N., Treviño, S., Prabhakaran, E., and Scholtz, J.M. (2004). Protein structure, stability and solubility in water and other solvents. *Philos Trans R Soc Lond B Biol Sci* 359, 1225–1235.

Peters, C., and Elofsson, A. (2014). Why is the biological hydrophobicity scale more accurate than earlier experimental hydrophobicity scales? *Proteins* 82, 2190–2198.

Simm, S., Einloft, J., Mirus, O., and Schleiff, E. (2016). 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biol Res* 49.

Wolfenden, R. (2007). Experimental Measures of Amino Acid Hydrophobicity and the Topology of Transmembrane and Globular Proteins. *J Gen Physiol* 129, 357–362.

Zhao, G., and London, E. (2006). An amino acid “transmembrane tendency” scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: Relationship to biological hydrophobicity. *Protein Sci* 15, 1987–2001.