

# EpiQuest-IM Manual



## Content:

[What is EpiQuest-IM?](#)

[Species specificity](#)

[Entering the data](#)

[Settings for Immunogenicity analysis](#)

[Viewing and saving the Results](#)

[Demo Sequences](#)

## What is EpiQuest-IM?

Whether the immune system will respond to an antigen, and if this response will be strong, is defined by a whole spectrum of factors:

- 1) antigenicity of the protein domains
- 2) accessibility of antigenic domains, to restimulate the response via BCR
- 3) copy number, which may compensate for lower antigenicity
- 4) presence of strong T-helper epitopes

and other factors.

When it comes to linear B-epitopes (they represent the overwhelming majority of all epitopes) we should discriminate between their **antigenicity**, as the *potential* of the epitope to elicit strong specific antibodies when presented to the immune system, and **immunogenicity**, which is the *true ability* of an epitope to elicit such response.

For a peptide antigen that is coupled to a carrier, antigenicity and immunogenicity will be identical. However, when it comes to a folded native protein, the factor of epitope accessibility becomes critical. This distinction is especially important for the epitopes that will be a target of immune response against viral proteins, complex allergens etc. When the goal is to predict *in silico* the epitopes that will detect an immune response against the target, preferably - immunodominant epitopes across a whole population of individuals/animals that were exposed to the target protein.

While EpiQuest-B builds an antigenicity profile for a protein sequence, allowing the discovery of *potentially* immunogenic linear B-epitopes, the new program - EpiQuest-IM - predicts the *immunogenicity* of such epitopes in real contact of the target protein with the immune system. It compensates for epitopes being hidden or semi-hidden in the mature protein.

## Species specificity

The models tested so far suggest applicability of the program to predict immunogenicity (immunodominance) of epitopes, identified in the analysed protein in humans, mice, rabbits, swine.

The best epitopes do perform as immunodominant when these species are immunized or otherwise contact (i.e. virus infection) with the immune system of the recipient from the listed species. This has been retrospectively analysed using sequences with known immunodominant epitopes.

We should indicate that, so far, we have such confirmation for geese and ducks, although the responses in chickens seem to be predicted correctly.

## Entering the Data

The form contains the following fields and controls:

- 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. These 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 original amino acid number will be preserved: the numbering starts from the first amino acid of the original sequence (not of the fragment)/ Thus, if the first aa of the fragment was 101, it will be indicated as 101 in all further data in 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 form includes the following controls:

- K =**
- Only surface-exposed sequences
- Show only positive fragments
- Antigenicity** (checked):
  - Frame size:  ∈ [3, 15]
  - Gap size:  ∈ [0, 5]
  - Threshold:  ∈ [-3, 3]
  - Peptide size:   higher
  - Sort by:
- Accessibility** (checked):
  - Frame size:  ∈ [3, 15]
  - Gap size:  ∈ [0, 5]
  - Threshold:  ∈ [-3, 3]
- Default** buttons for both Antigenicity and Accessibility settings.

The settings include defining parameters for analysis of **Antigenicity** and for surface **Accessibility**. Please see the explanation for **Frame size**, **Gap size** and **Threshold** in the manual for the EpiQuest-B.

**Unless you are an experienced user, we recommend keeping all the settings at Default values.** Current Default parameters were developed by testing the program against the data from models where the antibody response against linear epitopes was tested using a series of overlapping peptides covering the entire sequence of the target protein.

**K** – the parameter defining to what extent the accessibility of the sequence will influence the response to the epitope. Normally keep the default value. For large copy number antigens (i.e. such as spike proteins of some viruses) presented by 1000 or more copies per particle, you may increase it to 1. If you want to exclude the effect of accessibility on immunogenicity, switch it to 0.

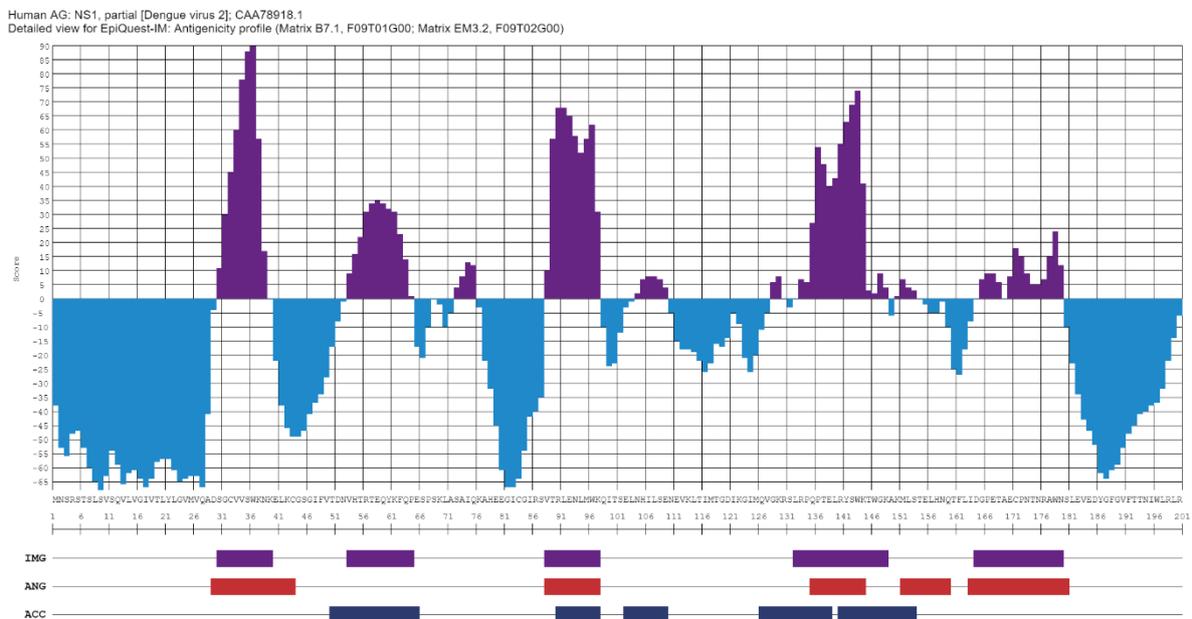
Selected by default  **Only surface-expose sequence**” cuts off all non-accessible and hidden sequences from the results (which are highly unlikely be the target of the antibody response against the native protein). You can limit your analysis to only the most accessible sequences by additionally increasing the **Threshold** to 1 or 2 in settings for **accessibility**.

By default, the program produces all sequences in the Result, positive and negative by their immunogenicity. You can disregard the negative results by clicking  **Show only positive fragments**.

**Sort by** defines how the results for positive fragments will be presented in the report tables. They can be sorted according to the order of the fragments in the original sequence (**Start**), according to their **Length** (longest goes first), cumulative potential **immunogenicity** of the fragment **IMI** (immunogenicity index) or per amino-acid residue of the fragment (**IPB**).

## Viewing and saving the Results

While the **Graph** in Results represents the values for **Immunogenicity** for every amino acid of the sequence, demonstrating the value distribution within the positive and negative fragments, the **Bar** shows only the **positive fragments** that are equal to or above the length of a sequence defined by **Peptide Size**. There are 3 bars indicating the positive fragments equal or longer than the selected Peptide size for Immunogenicity, Antigenicity and Accessibility. The positive fragments for Immunogenicity are automatically calculated and defined.



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)

Save diagram

Human AG: NS1, partial [Dengue virus 2]; CAA78918.1  
Detailed view for EpiQuest-IM: Antigenicity profile (Matrix B7.1, F09T01G00; Matrix EM3.2, F09T02G00)

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

View Save Copy **Report for "Antigenicity":**

Matrix B7.1, F09T01G00; Matrix EM3.2, F09T02G00; K=1.0; P=8+

Start	End	Length	IMI	IPB	Sequence
132	148	17	545.00	32.06	LRPQPTELRYSWKTWGK
88	97	10	528.00	52.80	VTRLENLMWK
30	39	10	476.00	47.60	SGCVVSWKNK
53	64	12	282.00	23.50	VHTRTEQYKFQP
164	179	16	149.00	9.31	DGPETAECPTNRAWN
154	164	11	-101.00	-9.18	TELHNQTFID
110	127	18	-290.00	-16.11	NEVKLTIMTGDIKIMQV
39	52	14	-417.00	-29.79	KELKCGSGIFVTDN
76	87	12	-532.00	-44.33	KAHEEGICGIRS
180	200	21	-831.00	-39.57	SLEVEDYGFGVFTTNIWLRLR
1	29	29	-1648.00	-56.83	MNSRSTLSVSVQLVGVITLYLGVMVQAD

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 Antigenicity profile (EpiQuest-IM)

**Date & Time:** 10.11.2020 13:39:30

**Protein Name:** Human AG: NS1, partial [Dengue virus 2]

**Accession Number:** CAA78918.1

**Sequence:**  
MNSRSTLSVSVQLVGVITLYLGVMVQADSGCVVSWKNKELKCGSGIFVTDNVHTRTEQYKF  
QPESPSKLSAIAIQAHEEGICGIRSVTRLENLMWKQITSELNHIHILSENEVKLTIMTGDIKGI  
MQVGRSLRPQTELRYSWKTWKGAKMLSTELHNQTFIDGPETAECPTNRAWNSLEVEDY  
FGVFTTNIWLRLREKQDAFCDSKLSAAIKDNRAVHADMGYHIESALNDTWKIEKASFIEV  
KSCHWPKSHLWSNGVLESEMVIPKNFAGPKSQHNWRPGYHTQTAGPWHLGLKLEMDDFDFCEG  
TTVVVTEDCGNRGLRRTTASGKLTIEWCCRSCTLPPLRYRGEDGCWYGMIEIRPLKEEEN LVSSLVTA

**Fragment:** 1 - 200

**Matrix:** B7.1

**Frame size:** 9

**Gap:** 0

**Threshold:** 1

**Peptide size:** 8 higher

**Sorted by:** IMI

Start	End	Length	IMI	IPB	Sequence
132	148	17	545.00	32.06	LRPQPTELRYSWKTWGK
88	97	10	528.00	52.80	VTRLENLMWK
30	39	10	476.00	47.60	SGCVVSWKNK
53	64	12	282.00	23.50	VHTRTEQYKFQP
164	179	16	149.00	9.31	DGPETAECPTNRAWN
154	164	11	-101.00	-9.18	TELHNQTFID
110	127	18	-290.00	-16.11	NEVKLTIMTGDIKIMQV
39	52	14	-417.00	-29.79	KELKCGSGIFVTDN
76	87	12	-532.00	-44.33	KAHEEGICGIRS
180	200	21	-831.00	-39.57	SLEVEDYGFGVFTTNIWLRLR
1	29	29	-1648.00	-56.83	MNSRSTLSVSVQLVGVITLYLGVMVQAD

## 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-**.

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