EpiQuest-M – Operation Manual

Content:

What is EpiQuest-M?

EpiQuest-M analyses the linear protein sequence looking for clusters of amino acids/domains that are more/less prone to changes in a process of protein evolution.

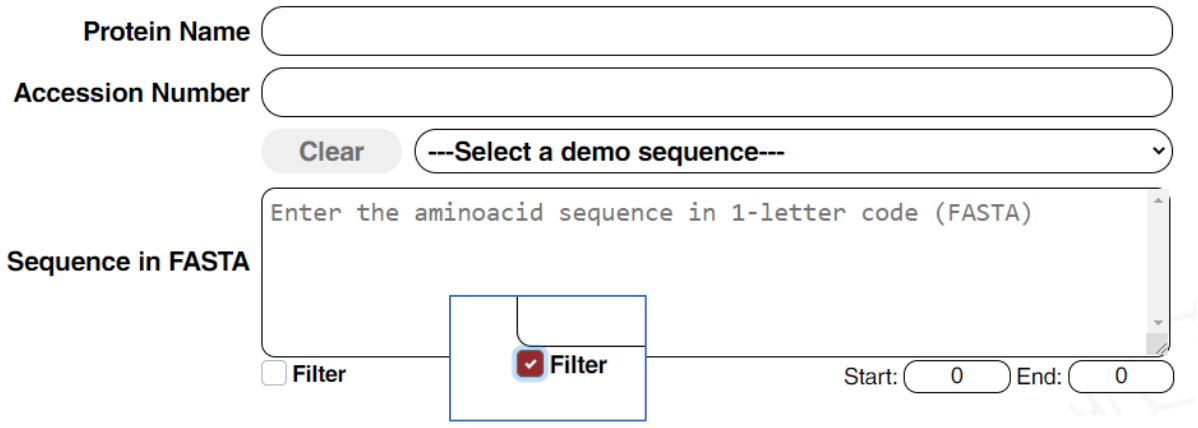
The Matrix is based on the relative probability of an amino acid being replaced by another (related or not) due to occasional point mutations (neutral theory of evolution, Kimura, 1991)[i]. It is based on the probabilities of various amino acid being replaced by another, reported for mitochondrial proteins by *Adachi & Hasegava* (1992)[i]. Therefore, it does not reflect a variability in the specific region of proteins caused by, for example, a selective pressure of immune system, but rather demonstrates overall potential for stability of variability.

In general, it is designed to select the most stable region of a particular viral or bacterial protein, or the region that is likely to be stable in a protein across several species.

Species specificity

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

Entering the Data



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.

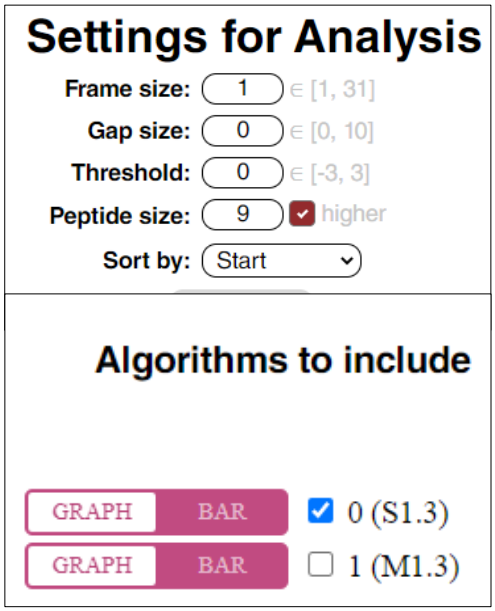
If your sequence contains gaps (may appear due to formatting) or contains abnormal signs (punctuation/letters that are not amino acid code) the analysis will be aborted. To avoid it, keep the ***"Filter"*** option selected.

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 original numeration of the

amino acids in protein will be preserved for the fragment. I.e. if the first amino acid of the fragment was in position 101 in the original sequence, it will be indicated as 101 in all further analysis and Results for the fragment.

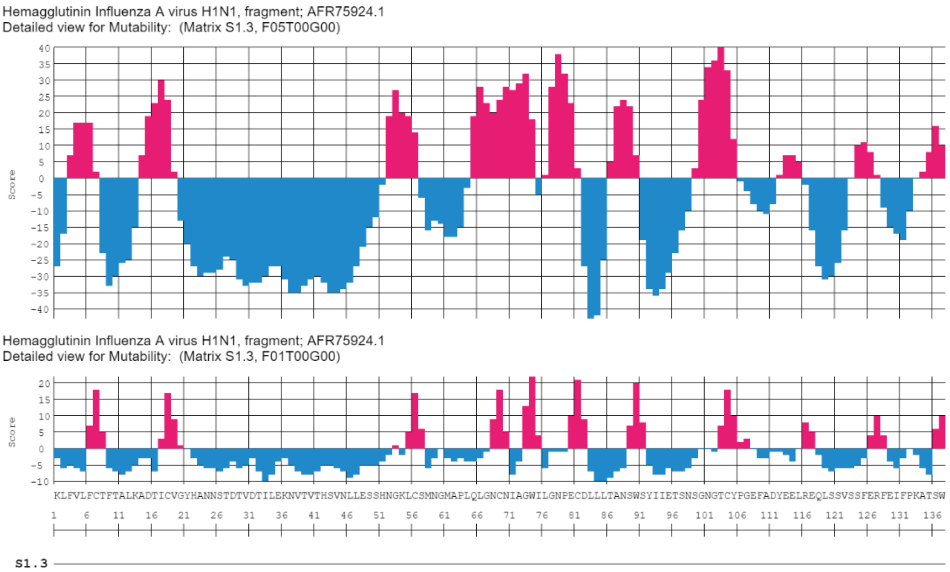
However, if you paste a fragment of a sequence into the sequence window its first amino acid will be referred as in position 1, not the position in the original sequence/source of the fragment.

Settings for Stability/Mutability analysis



EpiQuest-M allows you to customise the settings. We recommend keeping the ***Default*** settings (can be reset by the respective button) for most cases.

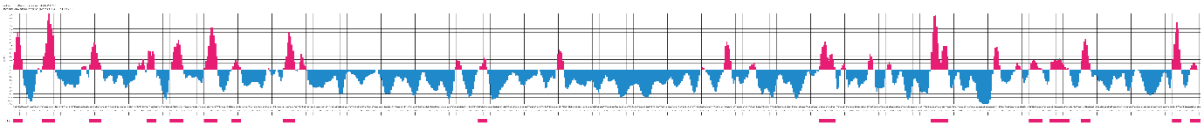
Frame size defines the size of the context sequence that will be reviewed for every position. The setting will define how general the image you receive in your analysis will be. With a small **Frame Size** (i.e. =1) you will see the potential for mutability/stability of every individual residue. We would recommend this setting when you are looking for the most stable regions in fast mutating virus proteins, where multiple cycles of replication lead to fast accumulation of amino acid replacements. See a fragment analysis do exactly that for hemagglutinin of influenza virus (below).



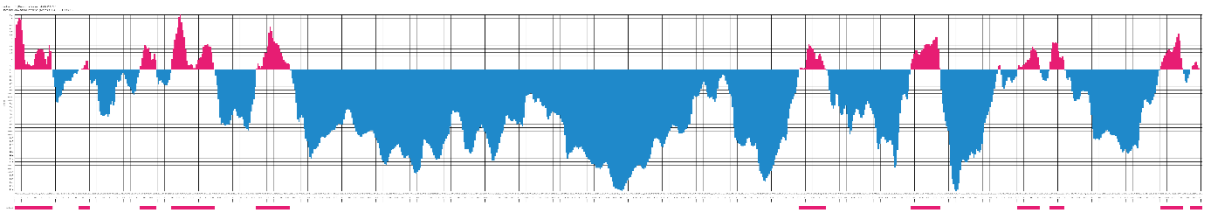
The analysis performed using frame 1 and frame 5 for the same fragment of influenza HA protein. Larger frame allows the detection of larger relatively stable areas.

When you are looking for evolutionary stable fragments in protein, use the large settings. Again, the smaller the ***Frame*** the shorter the domains but of relatively higher stability will be detected. Below we show an example of relative stability analysed for E-cadherin of Xenopus:

Frame=5 (below)



Frame=30 (below)



On bar only the fragments of 7aa or longer are shown in both graphs.

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 values 1-3 for ***Gap Size*** , in rare situations up to 5. The recommended default value is 1. In general, we would not recommend using a gap size larger than 1/5-1/10th of the ***Frame Size*** (i.e. for frame 10 use gap 1 to 2). The default **Gap Size**=0

Threshold (Th). You may want to see the highest peaks on the graph and isolate the most stable/prone to change regions. In this case increase **Th** to 1 or 2.

Peptide size. While the Graph in Results represents the values for every amino acid of the sequence, with peak positive and negative values, the Bar and Tabular reports show only *positive fragments* that are equal to 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), cumulative potential for change of the fragment *SI* (stability) or *SI* per amino-acid residue of the fragment (*SIR*).

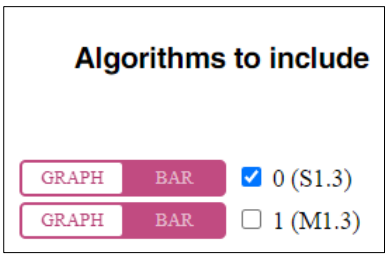


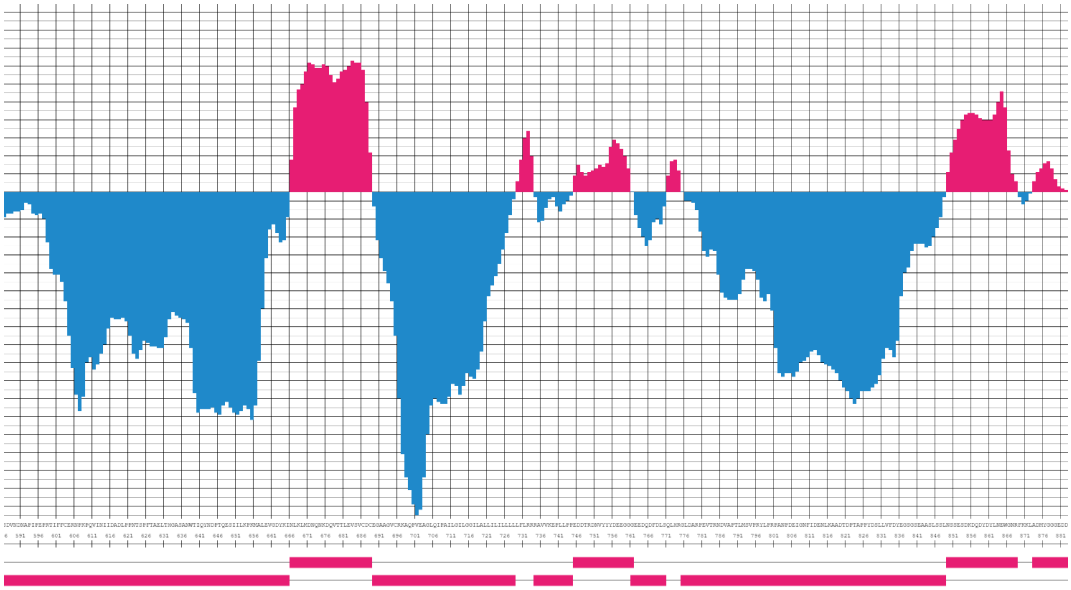
Chart Type

You may perform both analysis for Stability and for Mutability (which is a mirror of Stability) in the same run. The Chart for analysis will be presented as a Graph or just Bar indicating only the positions of positive sequences with length equal or above the selected in ***Peptide Size***.

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 Stability, while *bar* shows only positive fragments of length equal to or exceeding the chosen Peptide Size. If you want both Stability and mutability to be analysed, you may ask for the full diagram for one and only the Bar graph for the other. Below is shown a C-terminal chart of the analysis for E-cadherin of X. Levis, The diagram shows regions of high evolutionary stability, while the bars below show fragments of 9aa and longer (as defined by Peptide size):



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 ihe 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.

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.

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.

---

[i] **References:**

Adachi, J., and Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. *42*, 459–468.

Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. The Japanese Journal of Genetics *66*, 367–386.