CassOpt is a program for optimization of mini-genes which are translated to proteins with particular immunogenic peptides. The mini-genes are used in immunology to research an immune response to particular peptides binded to MHC (the major histocompatibility complex). To create an unbiased mini-gene sequence, the translation of the joined peptide coding subsequences should not originate unnecessary immunogenic peptides in addition to the peptides of interest. The main idea of CassOpt is to find the shortest combination of peptide flanks (the flanks are parts of the native proteins which originate the peptides of interest) that do not originate unnecessary peptides with ability to bind to MHC. To estimate peptide binding CassOpt uses predictions of netMHCpan program (http://www.cbs.dtu.dk/services/NetMHCpan/). Rearrangement of the initial subsequences in couple with variation of theirs flank lengths gives the best results for mini-gene sequence optimization (Figure 1).

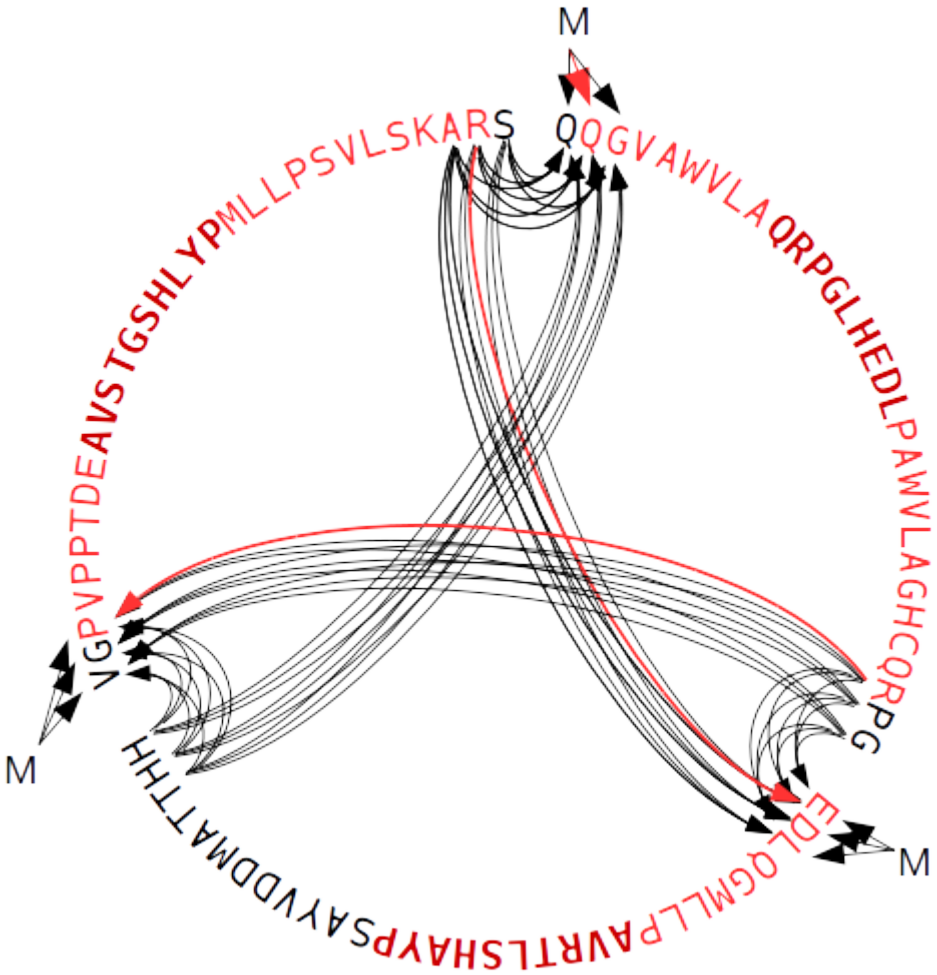

Figure 1. The diagram explains how the program arranges three peptides (a bold font) with flanks in order of the optimized protein sequence without immunogenic junctions (red sequence). Arrows represent possible paths of rearrangements, the optimal path is shown with red arrows. The program starts from each of the three start positions (“M”) and iterates over all combinations of the peptide arrangements using the shortest flanks. If two peptides can not be joined with a nonimmunogenic junction, the longer peptide flanks will be used. A junction is nonimmunogenic if there are no any immunogenic peptides which overlap the junction. Immunogenicity of the peptides is verified with netMHCpan program.

Dependencies

- Linux or MacOS
- python >= 3.5
- netMHCpan >= 4.0
- tcsh >= 6.18 (required for netMHCpan)

Installation

- install netMHCpan4 program which was described in the paper:

  J Immunol. 2017 Nov 1;199(9):3360-3368. doi: 10.4049/jimmunol.1700893.

  Epub 2017 Oct 4.

  (https://www.jimmunol.org/content/early/2017/10/04/jimmunol.1700893)
- add path of the program to $PATH environment variable:

  export PATH=$PATH:/path/to/netMHCpan
- download CassOpt using git:

  git clone https://github.com/open-projects/CassOpt
- test CassOpt:

  cd ./CassOpt

  CassOpt.py -f ./test/input_file.fa

Program usage

CassOpt.py [options] -f fasta_file [-o output_file]

-***f*** *fasta_file* – a file of immunogenic peptide sequences with flanks in FASTA file format (http://genetics.bwh.harvard.edu/pph/FASTA.html); the fasta headers have to be written in the next format:

>name (beg_pept_pos..end_pept_pos)

, where *beg_pept_pos* and *end_pept_pos* are start and end positions of the peptide of interest in the sequence

-***o*** *output_file* – the output file with the arranged immunogenic peptide sequences having the optimized flanks for each peptide;

options:

-***l*** *8 9 10 11* – lengths of tested peptides which overlap the flank junctions;

-***m*** *10* – the minimum length of flanks;

-***p*** */path/to/netMHCpan4* – you can specify the path to netMHCpan4 program;

-***t*** *./tmp* – the temporary directory;

License