The MKion program was designed to accept output form the Ion Torrent sequences of TCR sequences in Fasta (not FastaQ) format and to assign the Valpha and Jalpha present as well as the CDR3s within of each sequence. The program will run on any PC with Windows XP or earlier operating systems. It technically will run in a Windows 7 virtual Window XP window, but is VERY slow. We have included the uncompiled program text files, in case someone wants to modify the program to analyze other types of TCR or IG sequences. The code would then have to be recompiled with Turbo Pascal 7 or converted to some other programming platform.

The output of the Ion Torrent Fasta files are in Unix format and must be converted to Dos Format before input into MKion. We use the Vim editor, freely available on line, using the Command – “:set ff=DOS”. The Ion Torrent software removes barcodes, instrument specific primer sequences and very low quality sequences before its output to us. MKion assumes this has already occurred. As a DOS program MKion can only deal with filenames of less than 9 characters and it requires the “.fa” extension rather than the “.fasta” extension. Rename the input file accordingly.

To run MKion, place the fasta file, Valpha and Jalpha identifier files in the same folder. Execute the program and input the identifier files as instructed. MKion analyzes the data in several steps by menu.

First input the fasta file and choose option 1 to clean up the files. If the sequence is not the positive strand, it will be converted its complement. You define the shortest and longest sequence length allowed to eliminate worthless sequences and those that might crash the program. Generally pick 170bp and 600bp as the limit unless you know better. The program will then parse the sequences into 3 files SHORTSeq.fa, LONGSeq.fa and a file that you name for the cleaned up sequences. The first two files are rewritten every time the program is run, so rename them if you want to save them.

The fasta file with the cleaned up sequences will have a new header for each sequence with:

1. A unique sequence identifier.
2. Sample identifier,
3. Space for the TRAV number
4. Space for the TRAJ number
5. Space for the functional state of the sequence (P=productive, O=CDR3 out of frame, T=termination codon in the CDR3 and U= some portion of the sequence not identified.
6. The CDR3 protein sequence.

Next input the file of cleaned up sequences and pick option 2 to analyze the V, J and CDR3s. The header will be filled-in and the sequences output to a new file that you name.

You can then input this file and pick option 3 to output portions of the sequences to a new fasta file or option 4 to output all sequences to an Excel compatible text file.