Proposal for research stay

Michał Burdukiewicz

October 6, 2016

1 Workplan

Aim: improve performance of signalP for atypical signal peptides using a reduced amino acid alphabet.

Since signalP 5 is not yet available, I want to instead use signalP 4.1.

1. Adjust signalP 4.1 for reduced alphabets. I don't have an access to signalP 4.1 source code, but it is possible, that it has hardcoded alphabet of 20 amino acids, so a bit of programming might be necessary.

2. Benchmark modified signalP 4.1 on external data set(s) of atypical signal peptides and compare with normal signalP 4.1 and signalP 3.0.

If modified signalP 4.1 is as accurate as it was during benchmark on proteins belonging to other eukaryotes and it predicts atypical signal peptides better, it gives us a rationale to use reduced alphabets and justifies further search for more optimized reduced amino acid alphabet.

2 Concerns

It is possible that reduction of the alphabet may improve the general performance of signal peptide prediction, but lower accuracy of cleavage sites prediction. Cleavage sites seem to require more defined motifs than whole signal peptide and may need larger alphabets, possibly even the full amino acid alphabet.