

Expanding signalHsmm using n-grams

Michał Burdukiewicz, Piotr Sobczyk

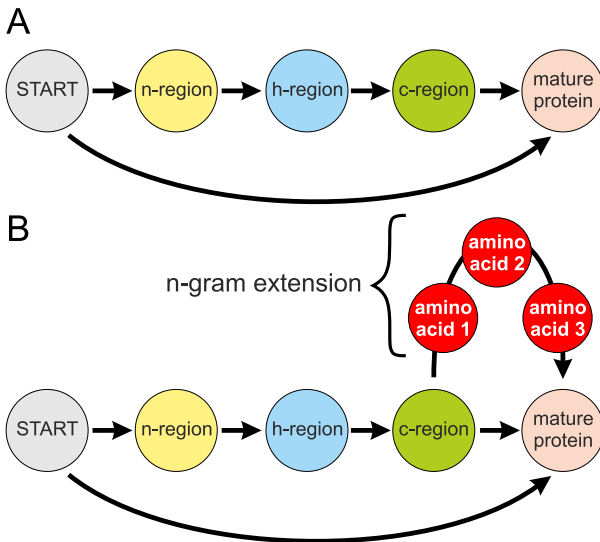
University of Wrocław, Department of Genomics, Poland

Outline

1 Motivation

2 Amino acid properties

To improve our model, we added few supplementary states representing specific motives that might occur in the proximity of cleavage site. The structure of cleavage sites, more conserved than other parts of signal peptide, may be reflected by n -grams (k-mers), short vectors of n characters derived from input sequences.



The diagram of simple (A) and extended version of signalHsmm with the n-gram cleavage site model (B).

Assumptions:

- 1 The optimal number of amino acid groups is 4.
- 2 Important properties are: size, polarity, charge, hydrophobicity and probability of being in the α -chain.
- 3 All properties above are equally important and distinguish signal peptides from mature proteins.

Outline

1 Motivation

2 Amino acid properties

bla