



Signal peptide prediction: from plain neural networks to language models

Nielsen, Henrik

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Nielsen, H. (2021). *Signal peptide prediction: from plain neural networks to language models*. Abstract from Belgrade Bioinformatics Conference 2021, Vina, Serbia.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Signal peptide prediction: from plain neural networks to language models

Henrik Nielsen

Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Abstract:

The signal peptide is the most well-known protein sorting signal, signaling entry into the secretory pathway in all domains of life. Since it was launched in 1996, the program SignalP has been the most popular method for predicting signal peptides from amino acid sequences. In this talk, I will give a historical overview of the various versions of SignalP. Versions 1 to 4 were based on conventional “shallow” feed-forward neural networks, combined with a hidden Markov model in versions 2 and 3. In 2019, SignalP 5.0 was introduced based on deep recurrent neural networks combined with a conditional random field, and it is able to distinguish between several types of signal peptides in Bacteria and Archaea.

In the last part of my talk, I will present a preview of the not yet published SignalP 6.0, which is based on pretrained protein language models. This has made it possible, for the first time, to distinguish between all five known types of signal peptides in prokaryotes, including the rare Tat/SPII and Sec/SPIII types. Furthermore, SignalP 6.0 is able to assign the n-, h-, and c-regions within the signal peptides, enabling detailed analysis of signal peptide properties. In addition, SignalP 6.0 does not need information about the organism of origin, making it well suited for analysis of metagenomic datasets.

*Corresponding author, e-mail: henni@dtu.dk