NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning

The ability to predict local structural features of a protein from the primary sequence is of paramount importance for unravelling its function in absence of experimental structural information. Two main factors affect the utility of potential prediction tools: their accuracy must enable extraction of reliable structural information on the proteins of interest, and their runtime must be low to keep pace with sequencing data being generated at a constantly increasing speed. Here, we present NetSurfP-2.0, a novel tool that can predict the most important local structural features with unprecedented accuracy and runtime. NetSurfP-2.0 is sequence-based and uses an architecture composed of convolutional and long short-term memory neural networks trained on solved protein structures. Using a single integrated model, NetSurfP-2.0 predicts solvent accessibility, secondary structure, structural disorder, and backbone dihedral angles for each residue of the input sequences. We assessed the accuracy of NetSurfP-2.0 on several independent test datasets and found it to consistently produce state-of-the-art predictions for each of its output features. We observe a correlation of 80% between predictions and experimental data for solvent accessibility, and a precision of 85% on secondary structure 3-class predictions. In addition to improved accuracy, the processing time has been optimized to allow predicting more than 1,000 proteins in less than 2 hours, and complete proteomes in less than 1 day. This article is protected by copyright. All rights reserved.

General information
Publication status: Published
Organisations: Novo Nordisk Foundation Center for Biosustainability, Bacterial Synthetic Biology, Immunoinformatics and Machine Learning, Bioinformatics, Department of Health Technology, Process engineering, National Centre for Nano Fabrication and Characterization, Copenhagen Center for Health Technology, Cognitive Systems, Department of Applied Mathematics and Computer Science, AI for Immunological Molecules, University of Copenhagen
Corresponding author: Marcatili, P.
Contributors: Klausen, M. S., Jespersen, M. C., Nielsen, H., Jensen, K. K., Jurtz, V. I., Senderby, C. K., Sommer, M. O. A., Winther, O., Nielsen, M., Petersen, B., Marcatili, P.
Number of pages: 8
Pages: 520-527
Publication date: 2019
Peer-reviewed: Yes

Publication information
Journal: Proteins: Structure, Function, and Bioinformatics
Volume: 87
Issue number: 6
ISSN (Print): 0887-3585
Ratings:
BFI (2019): BFI-level 1
Web of Science (2019): Indexed yes
Original language: English
Keywords: Deep learning, Disorder, Local structure prediction, Secondary structure, Solvent accessibility
DOIs:
10.1002/prot.25674
Source: FindIt
Source-ID: 2444112550
Research output: Contribution to journal › Journal article – Annual report year: 2019 › Research › peer-review