

Multi-head attention-based U-Nets for predicting protein domain boundaries using 1D sequence features and 2D distance maps

Sajid Mahmud¹, Zhiye Guo¹, Farhan Quadir¹, Jian Liu¹, and Jianlin Cheng¹

¹University of Missouri

April 20, 2022

Abstract

The information about the domain architecture of proteins is useful for studying protein structure and function. However, accurate prediction of protein domain boundaries (i.e., sequence regions separating two domains) from sequence remains a significant challenge. In this work, we develop a deep learning method based on multi-head U-Nets (called DistDom) to predict protein domain boundaries utilizing 1D sequence features and predicted 2D inter-residue distance map as input. The 1D features contain the evolutionary and physicochemical information of protein sequences, whereas the 2D distance map includes the structural information of proteins that was rarely used in domain boundary prediction before. The 1D and 2D features are processed by the 1D and 2D U-Nets respectively to generate hidden features. The hidden features are then used by the multi-head attention to predict the probability of each residue of a protein being in a domain boundary, leveraging both local and global information in the features. The residue-level domain boundary predictions can be used to classify proteins as single-domain or multi-domain proteins. It classifies the CASP14 single-domain and multi-domain targets at the accuracy of 72.7%, 8.02% more accurate than the state-of-the-art method. Tested on the CASP14 multi-domain protein targets with expert annotated domain boundaries, the average per-target F1 measure score of the domain boundary prediction by DistDom is 0.241, 18.72% higher than the state-of-the-art method.

Hosted file

DistDom_v5.docx available at <https://authorea.com/users/477406/articles/566029-multi-head-attention-based-u-nets-for-predicting-protein-domain-boundaries-using-1d-sequence-features-and-2d-distance-maps>

figures/Figure-1/Figure-1-eps-converted-to.pdf

figures/Figure-2/Figure-2-eps-converted-to.pdf

figures/Figure-3/Figure-3-eps-converted-to.pdf

figures/Figure-4/Figure-4-eps-converted-to.pdf

figures/Figure-5/Figure-5-eps-converted-to.pdf

figures/Figure-6/Figure-6-eps-converted-to.pdf

figures/Figure-7/Figure-7-eps-converted-to.pdf