

Protein secondary structure assignment using pc-polyline and convolutional neuron network

Lincong Wang¹, Chen Cao², and Shuxue Zuo¹

¹Jilin University

²University of Calgary

October 14, 2020

Abstract

The assignment of protein secondary structure elements (SSEs) underpins the structural analysis and prediction. The backbone of a protein could be adequately represented using a pc-polyline that passes through the centers of its peptide planes. One salient feature of pc-polyline representation is that the secondary structure of a protein becomes recognizable in a matrix whose elements are the pairwise distances between two peptide plane centers. Thus a pc-polyline could in turn be used to assign SSEs. Using convolutional neuron network (CNN) here we confirm that a pc-polyline indeed contains enough information for it to be used for the accurate assignments of six types of secondary structure elements: α -helix, β -sheet, β -bulge, 3₁₀-helix, turn and loop. The applications to three large data sets show that the assignments made by our CNN-based P2PSSE program agree very well with those by DSSP, STRIDE and quite well with those by five other programs. The analyses of the assignments by P2PSSE and those by other programs raise some general questions about the characterizations of protein secondary structure. In particular the analyses illustrate the difficulty with giving a quantitative and consistent definition for each of the six SSE types especially for 3₁₀-helix, β -bulge, turn or loop in terms of either backbone H-bond patterns, or backbone dihedral angles, or C α -polylines or pc-polylines. The difficulty suggests that the SSE space though being dominated by the regions for the six SSE types is to a certain degree continuous.

Hosted file

p2psse_lincong.pdf available at <https://authorea.com/users/367080/articles/486638-protein-secondary-structure-assignment-using-pc-polyline-and-convolutional-neuron-network>