



Contribution ID: 69

Type: not specified

What does artificial intelligence see in 3D protein structures?

Wednesday, 13 November 2019 10:15 (25 minutes)

Structural bioinformatics and structural biology in the last 40 years have been dominated by bottom-up approaches. Specifically, researchers have been trying to construct complex models of macromolecules starting from the first principles. These approaches require many approximations and very often turned out to be rough or even incorrect. For example, many classical methods are based on a dictionary of structural features determined by expert knowledge, such as protein secondary structure, electrostatic estimations, solvent accessibility, etc. However, the reality and underlying physics of proteins is much more complex than our current description of it. Therefore, more progress is needed in this field. Fortunately, deep learning has recently become a very powerful alternative to many classical methods, as it provides a robust machinery for the development of top-down techniques, where one can learn elementary laws from a number of high-level observations. Indeed, it allows constructing models using features and descriptors of raw input data that would be inaccessible otherwise. We have recently studied recurrent structural patterns in protein structures recognized by a deep neural network. We demonstrated that neural networks can learn a vast amount of chemo-structural features with only a very little amount of human supervision.

Our architecture learns atomic, amino acid, and also higher level molecular descriptors. Some of them are rather complex, but well understood from the biophysical point of view. These include atom partial charges, atom chemical elements, properties of amino acids, protein secondary structure, and atom solvent exposure. We also demonstrate that our network architecture learns novel structural features. For example, we discovered a structural pattern consisting of an arginine side-chain buried in a beta-sheet. Another pattern is a spatially proximate alanine and leucine residues located on the consecutive turns of an alpha helix. Overall, our study demonstrates the power of deep learning in the representation of protein structure. It provides rich information about atom and amino acid properties and also suggests novel structural features that can be used in future computational methods.

Presenter: Dr GRUDININ, Sergei (Inria/CNRS)

Session Classification: Morning 1