ECRONICON OPEN ACCESS

EC PROTEOMICS AND BIOINFORMATICS
Opinion

Protein Structure Prediction: An Overview

Shaheera Rashwan*

Informatics Research Institute, City of SRTA, Alexandria, Egypt

***Corresponding Author**: Shaheera Rashwan, Informatics Research Institute, City of SRTA, Alexandria, Egypt.

Received: September 28, 2017; **Published:** October 12, 2017

A protein's structure determines its function. The prediction of the 3-D structure of polypeptides based only on the amino acid sequence (primary structure) is a problem that has, over the last decades, challenged biochemists, biologists, computer scientists and mathematicians. The Protein Structure Prediction Problem is one of the main research problems in Structural Bioinformatics.

Experimental protein structure determination is expensive, which has driven the search for methods that can predict protein structure from the amino acid sequence information [1]. The main challenge is to understand how the information encoded in the linear sequence of amino acid residues is translated into the 3-D structure, and from this knowledge, to develop computational methods that can accurately predict the native tertiary structure of a protein molecule.

There are two classes of protein structure prediction methods. The first class including threading and comparative modeling, rely on detecting at least one known structure. The second class of methods, *de novo* or *ab initio* methods, predict the structure from scratch [2]. About 50% of the known proteins are applicable to comparative modeling. For the remaining proteins, there are no satisfactory solution.

Most *de novo* (template-free) protein structure prediction methods rely on assembling proteins from short peptide fragments [3]. These fragments are taken from the database of known protein structures according to the similarity of their sequences. Thousands of trial structures are assessed by their expected stability based on energy functions derived from known structures. Only few models for proteins larger than 100 amino acids usually had the correct fold. Also, contrary to comparative modeling, it was hard to say if a model is reliable or not.

Progress in *de novo* prediction was slow until it was discovered that contacts between amino acids can be reliably predicted from large multiple sequence alignments. However, the contact predictions suffered from high false-positive rates. It was found that the source of false positives is from correlations arising through indirect chains of interactions. Similar problems had been solved in physics and statistics, and such methods could be applied to distinguish correlations from direct statistical couplings between the amino acids [4,5].

In the past 5 years, these methods have been improved and applied to predict the structures of many proteins [6] and even protein complexes [7]. One limitation is the requirement of the multiple sequence alignments containing hundreds to thousands of sequences for large-scale application.

Recent results from the latest blind Critical Assessment of Techniques for Structure Prediction (CASP12) [8] show that a few other methods perform almost as well or as well as the Rosetta server. A huge advantage of Rosetta is its free source code and a collaborative and open community of researchers.

Bibliography

1. Dill Ken A and Justin L MacCallum. "The protein-folding problem, 50 years on". *Science* 338.6110 (2012): 1042-1046.
2. Baker David and Andrej Sali. "Protein structure prediction and structural genomics". *Science* 294.5540 (2001): 93-96.

3. Kinch Lisa N., *et al*. "Evaluation of free modeling targets in CASP11 and ROLL". *Proteins: Structure, Function, and Bioinformatics* 84.1 (2016): 51-66.

4. Zhang Wenxuan., *et al*. "Integration of QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11". *Proteins: Structure, Function, and Bioinformatics* 84.1 (2016): 76-86.

5. Weigt, Martin, *et al*. "Identification of direct residue contacts in protein–protein interaction by message passing". *Proceedings of the National Academy of Sciences* 106.1 (2009): 67-72.

6. Marks Debora S., *et al*. "Protein 3D structure computed from evolutionary sequence variation". *PloS One* 6.12 (2011): e28766.

7. Nugent Timothy and David T Jones. "Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis". *Proceedings of the National Academy of Sciences* 109.24 (2012): E1540-E1547.

8. CASP 12.

Volume 1 Issue 2 October 2017