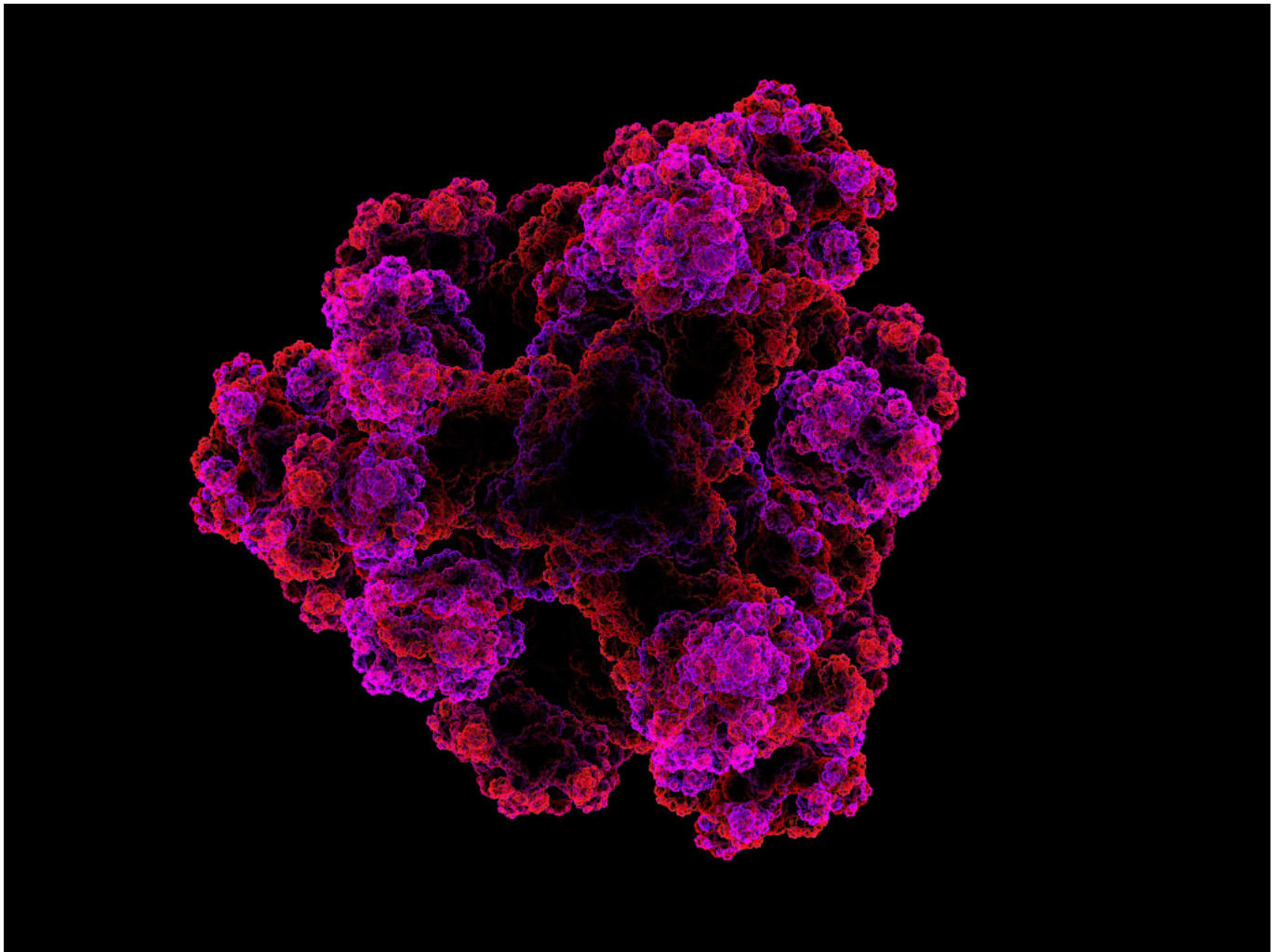


ab-initio prediction of protein structure: An introduction

Author : Muniba Faiza

Categories : [Algorithms](#), [Structural Bioinformatics](#), [Structure Prediction](#)

Date : December 10, 2017



We have heard a lot about the *ab-initio* term in Bioinformatics, which could be difficult to understand for newbies in the field of bioinformatics. Today, we will discuss in detail what ab-initio is and what are the applicable methods for it.

First of all, let's get familiar with the literal meaning of the term *ab-initio*, it means 'from the scratch'. This term is applied in the context of the protein structure prediction in bioinformatics, which is quite useful. Actually, *ab-initio* is one of the methods to predict a protein structure, which in case not available in protein data bank (PDB) [1]. There are basically three methods to predict a protein's structure:

a) homology modeling

b) *ab-initio*

c) threading

Homology modeling method is applied when there is a sufficient amount of similarity between the protein (structure to be predicted) and the template (whose structure have been determined already). But in the other case, when the similarity between the two is quite low, then the *ab-initio* method is applied. Although homology modeling aims to find a template protein which is evolutionary related to the query protein sequence. Threading is a little similar to homology modeling in the sense that it predicts the structure by recognizing the folds of the template and it aims to detect the evolutionary-related proteins and analogous folds, so we can say they are template-based methods. The homology modeling and threading both are capable of predicting protein structures with high-resolution folds based on the searched templates, but they suffer a few limitations that the native topology for the query sequence must have been solved, and new folds cannot be predicted using these two approaches.

The *ab-initio* method is often preferred for structure prediction when there is no or very low amount of similarity for the protein (let's say query protein sequence). It is the most difficult [2,3] and general approach where the query protein is folded with a random conformation. The *ab-initio* method is based on the thermodynamic hypothesis proposed by Anfinsen [4], according to which the native structure corresponds to the global free energy minimum under a given set of conditions.

There are several *ab-initio* structure prediction approaches available such as ROSETTA [5], TOUCHSTONE-II [6], and the most widely preferred I-Tasser [7,8]. These approaches are based on the Monte-Carlo algorithm [9,10]. It has been found that I-Tasser outperforms the ROSETTA and TOUCHSTONE-II approaches with a far lower CPU cost [11].

The *ab-initio* modeling is often termed as *de-novo* modeling [12], physics-based modeling [13], or free modeling [14]. The basic protocol followed by the *ab-initio* method of the protein structure prediction starts with the primary amino acid sequence which is searched for the different conformations leading to the prediction of native folds. After the folds have been recognized and predicted, the model assessment is performed to verify the quality of the predicted structure. ROSETTA and I-Tasser follow the enhanced methodology for *ab-initio* prediction of a protein.

ROSETTA prediction begins with the identification of small fragments (3mers and 9 mers) from the structure databases that have consistency with local sequence preferences. After that, all the fragments are assembled into models with global properties followed by the assessment of the models using a scoring function from decoy population [5]. The protocol followed by the I-Tasser includes threading along with the *ab-initio* method [6,7]. I-Tasser program is based on the secondary-structure enhanced Profile-Profile threading Alignment (PPA) [15] and the iterative implementation of the Threading ASSEMBly Refinement (TASSER) program [16]. The details of the I-Tasser program can be read [here](#).

We will be discussing other protein structure methods in detail in the upcoming articles.

References

1. Protein data bank (www.rcsb.org)

2. Lu, L., Lu, H., & Skolnick, J. (2002). MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. *Proteins: Structure, Function, and Bioinformatics*, 49(3), 350-364.
3. Floudas, C. A., Fung, H. K., McAllister, S. R., Mönnigmann, M., & Rajgaria, R. (2006). Advances in protein structure prediction and de novo protein design: A review. *Chemical Engineering Science*, 61(3), 966-988.
4. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, 181(4096), 223-230.
5. Simons, K. T., Bonneau, R., Ruczinski, I., & Baker, D. (1999). Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins: Structure, Function, and Bioinformatics*, 37(S3), 171-176.
6. Zhang, Y., Kolinski, A., & Skolnick, J. (2003). TOUCHSTONE II: a new approach to ab initio protein structure prediction. *Biophysical journal*, 85(2), 1145-1164.
7. Wu, S., Skolnick, J., & Zhang, Y. (2007). Ab initio modeling of small proteins by iterative TASSER simulations. *BMC biology*, 5(1), 17.
8. Zhang, Y. (2008). Progress and challenges in protein structure prediction. *Current opinion in structural biology*, 18(3), 342-348.
9. Simons, K. T., Kooperberg, C., Huang, E., & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *Journal of molecular biology*, 268(1), 209-225.
10. Hansmann, U. H., & Okamoto, Y. (1999). New Monte Carlo algorithms for protein folding. *Current opinion in structural biology*, 9(2), 177-183.
11. Wu, S., Skolnick, J., & Zhang, Y. (2007). Ab initio modeling of small proteins by iterative TASSER simulations. *BMC biology*, 5(1), 17.
12. Bradley, P., Misura, K. M., & Baker, D. (2005). Toward high-resolution de novo structure prediction for small proteins. *Science*, 309(5742), 1868-1871.
13. O'ndziej, S., Czaplewski, C., Liwo, A., Chinchio, M., Nanius, M., Vila, J. A., ... & Schafroth, H. D. (2005). Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. *Proceedings of the National Academy of Sciences of the United States of America*, 102(21), 7547-7552.
14. Jauch, R., Yeo, H. C., Kolatkar, P. R., & Clarke, N. D. (2007). Assessment of CASP7 structure predictions for template free targets. *Proteins: Structure, Function, and Bioinformatics*, 69(S8), 57-67.
15. Wu, S., & Zhang, Y. (2007). LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic acids research*, 35(10), 3375-3382.
16. Zhang, Y., & Skolnick, J. (2004). Automated structure prediction of weakly homologous proteins on a genomic scale. *Proceedings of the National Academy of Sciences of the United States of America*, 101(20), 7594-7599.

Sharing is caring. Spread the love!

- [Print](#)
- [Email](#)
- [LinkedIn](#)
- [Twitter](#)

- [Facebook](#)
- [Google](#)
-