International Journal of Research and Review
www.ijrrjournal.com E-ISSN: 2349-9788; P-ISSN: 2454-2237

Review Article

The Architectural Network for Protein Secondary Structure Prediction

Anindya Sundar Panja, Shiboprosad Mandal, Bidyut Bandyopadhyay, Smarajit Maity

Oriental Institute of Science and Technology, Vidyasagar University, Midnapore, West Bengal-721102, India.

Corresponding Author: Anindya Sundar Panja

Received: 21/05/2016 *Revised: 24/05/2016* *Accepted: 26/05/2016*

ABSTRACT

Over the past 25 years, the accuracy of proteins secondary structure prediction has improved substantially. Recently evolutionary information taken from the deviation of proteins in some structural family have again enhance prediction accuracy for all these residues predicted correctly is in one of the three sates helix, strands and others. The new methods developed over the past few years may be interesting in context of improvements which is achieved through combination of the existing methods. Evolutionary divergences profile posses' adequate information to improve protein secondary structure prediction accuracy. These profiles can also able to correctly predict long stretches of identical residues in other secondary structure. This sequence structure relationship may help to help to developed tool which can efficiently predict the protein secondary structure from its amino acid sequence.

Keywords: Secondary structure, Evolution, Algorithm, Tool, programme

INTRODUCTION

Secondary structure is significantly determined by the hydrogen bonds made by biopolymers. That biopolymer was resolved by an atomic resolution structure. The building blocks of the secondary structure predicted from the information present into the amino acids sequences and analyzed through molecular modeling simulation. Proteins performed the major key role in almost all biometabolic process and their functional properties depend upon their structural folds (Yen-Ru Chen et al. 2008). Protein secondary structure prediction helps to determine a meaningful analysis of biological function. The linear sequence of amino acid means the primary structure is the basic information of a protein from which four state secondary structures can be predicted.Genome sequencing technologies now a day's widely used because it is relatively chief accurate and fast comparison with protein structure.

Early approaches

Linus Pauling accurately assumed the structural configuration of helixes and strands (Pauling L., and Corey R. B. 1951; Pauling L. et al. 1951). The theoretical concept of Pauling was verified with the first Xray structure published (Kendrew J.C. et al. 1960; Perutz M.F. et al. 1960). The Ramachandran angles present in the polypeptide chain and the rotation of the polypeptide backbone phi and psi bonds are described, which is present around the polypeptide. To determine the distribution of the Ramachandran angles or torsion angles of a protein Ramachandran Plot is very much useful (Ramachandran G.N et.al. 1963). Szent and his group already designed

a method for the prediction of secondary structure from a primary sequences (Szent-Gyorgyi A. G. and Cohen C. 1957) based on each twenty amino acids propensity values, the first generation productive method became very popular later the segment of amino acids residues are taken to calculate from the previously used propensity values (Rost B., and Sander C. 2000). Although the accuracy level reached just above 60% due to the imaginable algorithm applied to calculate the percentage of residue present in a protein which is helix strand and others. Prediction from genome sequencing data linear amino acid chain can be determined by using various computational tools (Adams PD et al. 2013). Most accurate protein secondary structure prediction is a necessary step for improve modeling of a protein fold (Pirovano W and Heringa J. 2010) and determination of its biological function also (Sleator RD. 2012) for the prediction of three dimensional design (Das R and Baker D .2008) and enzymatic function (Kiss G et al. 2013) as well as in drug design with development, we should model the secondary structure of a protein (winter C et al. 2012).

Impenetrability general protein structure prediction

The result of three states showed partially some error due to some restriction of local information. So, attempts have been made to introduce some global informatics parameter into the local ones (Dickerson R. E et al. 1976). Multiple sequence alignment information can improve the secondary structure prediction level (Zvelebil M. J, 1987). The third generation method for this prediction is designed from this multiple sequence alignment information and this concept is applied into an automatic prediction method increases the accuracy level upto 70% but this alignment method require a large number of dataset with more advance time management algorithm (Rost B., and Sander C. 2000 and 1993).

Exploring evolutionary information

The evolutionary significant data are the key component of the data set which is used in this new method. Sequence contain more than 35% pair wise identical residues with more than hundred align residues have similar structure isolated from natural source (Rost B. 1999). The natural mutational process shows sequence divergence increases the stability against the environmental hazards. Most of the mutations result in proteins that will not protect against environment only by the formation of globular structure. Substitution with lower number of residues shows adaptation against the extreme condition of environment. Exchange of amino acids shows specificity means position specific profiling gives important and crucial information about structure. This evolutionary divergence data set was the major informative key password for the prediction of secondary structure of protein in third generation.

Improving accuracy level

The most successful logic applies for secondary structure prediction apply machine learning algorithm which maximize the relationship between the primary sequence between the protein and their corresponding secondary structure (Kabsch W and Sander C. 1983). The DSSP programmes successfully predict and improve the accuracy level above 80% but it depends upon the sample of protein sequences and their coordinate data set (Rost B. 2001). The coordinate data set of few proteins was stored into protein data bank and their corresponding secondary structure prediction began. In 1980 the first deposited protein structure in protein data bank data base was in membrane protein, which contain membrane helix as well as β strand (Westbrook JD et al. 2003; Engelman DM et al. 1986). Later another way became very popular named as homology modeling which can precisely predict both secondary and tertiary structure (Jones DT et al. 1992). Homology modeling accurately predicts the fold of corresponding structure by comparing closely related sequential data set deposited into the protein data bank (Sutcliffe MJ. 1987). Later in the 1990 the

concept of neural network and hidden Markov models were designed to improve the accuracy level of secondary structure predicted by homology modeling. Later it can be concluded that homology modeling which is based on the logical and theoretical concept of neural network and hidden Markov models (Rost B. 1997; Rost B. 2001; Eyrich V A. 2001), partially solving the protein fold that gradually increases the accuracy level of protein secondary structure prediction. The prediction accuracy directly or indirectly affect on how protein are to be analyzed and annotated to specify proteome analysis (Cozzetto D et al. 2005; Rost B et al. 2004). Critical assessment of structure prediction (CASP) method which shows more accuracy than the other structure prediction model, which directly can predict secondary structure from primary sequences (Westbrook JD et al. 2003).

Incorporating structural information

The structural information of secondary structure helps to predict three dimensional models as well as many proteins represent numerous important clues, this information of amino acids chain used to determine their corresponding three dimensional structures. The explosive growth of sequence structural relationship information results the numerous growth of *de novo* prediction from sequence (Anfinsen C.B. 1973 and 1962). To determine secondary structure of a protein the *de novo* folding based approaches has been taken, called state-of-the-art, which is based on the sequence structural similarity present in to the structural data base and the similar fragments are assembled by using empirical intermolecular force fields. Such logical approaches have worked favorably in cases for smaller peptides (Bradley P et al. 2005; Raman S. et al. 2010; Lange O F. et al 2012).

Promising aspect for future direction

The present effort for the better structural prediction enabled us for a clear assumption about the proteins' structure-function relationship. Attempts have been made to generate some software tools which can efficiently predict the stress withstanding abilities in a protein form its amino acid sequence. Scientists also made an extensive effort to develop software for the secondary structure prediction from the amino acid sequence of a protein. Conclusively, studies on proteins structure can generate a strong base of understanding the organismal behavior/existence and speciation/species proliferation on course of long evolutionary period as well as upcoming period.

REFERENCES

- Adams PD, Baker D, Brunger AT, Das R, DiMaio F, et al.. (2013) Advances, interactions, and future developments in the cns, phenix, and rosetta structural biology software systems. Biophysics 42.
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* 181, 223-230
- Anfinsen, C.B. Some observations on the basic principles of design in protein molecules. *Comp. Biochem. Physiol.* 4, 229-240 (1962).
- Bradley, P., Misura, K.M. & Baker, D. (2005), Toward high-resolution de novo structure prediction for small proteins. *Science* 309, 1868-1871.
- Chou PY, Fasman GD: Prediction of protein conformation. *Biochemistry* (1974), 13: 222-245. 10.1021/bi00699a002.
- Cozzetto D, Di Matteo A, Tramontano A: Ten years of predictions ... and counting. *FEBS J* 2005, 272: 881-882.
- Das R, Baker D (2008), Macromolecular modeling with rosetta. Annual Review of Biochemistry 77: 363-382. doi: 10.1146/annurev.biochem.77.062906.171838
- Dickerson, R. E., Timkovich, R., and Almassy, R. J. (1976), The cytochrome fold and the evolution of bacterial energy metabolism, *J. Mol. Biol.* 100, 473-491.
- Engelman DM, Steitz TA, Goldman A: (1986), Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu*

Rev Biophys Biophys Chem, 15: 321-353. 10.1146/annurev.bb.15.060186. 001541.

- Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Fiser A, Pazos F, Valencia A, Sali A, Rost B: EVA: (2001), continuous automatic evaluation of protein structure predictionservers. *Bioinformatics*, 17: 1242-1243. 10. 1093/bioinformatics /17.12.1242
- Guzzo AV: (1965), the influence of amino acid sequence on protein structure. *Biophys J*, 5: 809-822.
- Jones DT, Taylor WR, Thornton JM: (1992), A new approach to protein fold recognition. *Nature*, 358: 86-89. 10.1038/358086a0.
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637. doi: 10.1002/bip. 360221211.
- Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. J., Davies, D. R., and Phillips, D. C. (1960) Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution, *Nature* 185, 422-427.
- Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk K (2013) Computational enzyme design. Angewandte Chemie International Edition 52: 5700-5725. Doi: 10.1002/anie.201204077.
- Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D. (2012), Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. *Proc. Natl. Acad. Sci. USA* 109, 10873-10878.
- Pauling L, Corey RB, Branson HR: (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain.*Proc Natl Acad Sci USA*, 37: 205-234. 10.1073/pnas. 37.4.205.
- Pauling, L., and Corey, R. B. (1951) Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets, *Proc. Natl. Acad. Sci. USA* 37, 729-740.
- Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins: Two hydrogen- bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. USA* 37, 205–234.
- Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, G., Will, G., and North, A. T. (1960) Structure of haemoglobin: A three-dimensional Fourier synthesis at 5.5 Å resolutions, obtained by X-ray analysis, *Nature* 185, 416-422.
- Pirovano W, Heringa J (2010) Protein secondary structure prediction. In: Data Mining Techniques for the Life Sciences, Springer. pp. 327-348.
- Ramachandran G.N, Ramakrishnan C, Sasisekharan V., (1963), J Mol Biol., 7:95-99.
- Raman S, Lange OF, Rossi P, Tyka M, Wang X, Aramini J, Liu G, Ramelot TA, Eletsky A, Szyperski T, Kennedy MA, Prestegard J, Montelione GT, Baker D. (2010) NMR structure determination for larger proteins using backbone-only data. *Science* 327, 1014-1018.
- Rost B, Yachdav G, Liu J: (2004) The PredictProtein server. *Nucleic Acids Res*, (32 Web Server): W321-326.
- Rost B (2001) Review: protein secondary structure prediction continues to rise. Journal of Structural Biology 134: 204-218. doi:10.1006/jsbi. 2001.4336.
- Rost B, Schneider R, Sander C (1997) Protein fold recognition by prediction-based threading. *J Mol Biol*, 270: 471-480. 10.1006/jmbi.1997. 1101.
- Rost, B., and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol.* 232, 584-599.
- Rost, B. (1999) Twilight zone of protein sequence alignments, *Protein Eng.* 12, 85-94.
- Rost, B., and Sander, C. (2000) Third generation prediction of secondary structure, *in* Webster, D. (Ed.), Protein Structure Prediction: Methods and Protocols, pp. 71-95, Humana Press, Clifton, NJ.

- Szent-Gyorgyi, A. G., and Cohen, C. (1957) Role of proline in polypeptide chain configuration of proteins, *Science* 126, 697.
- Sleator RD (2012) Prediction of protein functions. In: Functional Genomics, Springer. pp. 15-24.
- Sutcliffe MJ, Haneef I, Carney D, Blundell TL: Knowledge based modelling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng* 1987, 1: 377-384.
- Westbrook JD, Feng Z, Chen L, Yang H, Berman HM: (2003) The Protein Data Bank and structural genomics. *Nucleic Acids Res*, 31: 489-491. 10.1093.
- Winter C, Henschel A, Tuukkanen A, Schroeder M (2012) Protein interactions in 3d: From interface evolution to drug discovery. Journal of Structural Biology 179: 347-358. doi: 10.1016/j.jsb.2012. 04.009.
- Zvelebil, M. J., Barton, G. J., Taylor, W. R., and Sternberg, M. J. E. (1987) Prediction of protein secondary structure and active sites using alignment of homologous sequences, *J. Mol. Biol.* 195, 957-961.

How to cite this article: Panja AS, Mandal S, Bandyopadhyay B et al. The architectural network for protein secondary structure prediction. Int J Res Rev. 2016; 3(5):69-73.

****