Development of Neighbors Atomic Coordinate Tables and Plot-Graphical Representation for Tertiary Structure in Proteins

[1]Hamed Fawareh, [2]Jehad Q. Odeh and [2]Nawal Al-Omoush
[1]Department of Software Engineering, Zarqa Private University, Jordan
[2]Department of Computer Science, Al al-Bayt University, Jordan

Abstract: Problem statement: Contact map prediction is an interesting method for its application in fold recognition and protein tertiary structure determination. **Approach:** This study developed a method to predict a protein tertiary structure from the amino acid sequence using contact map. The method showed prediction a protein in a binary space of exposure states. The approach produced several procedures to finds a set of three dimensional coordinates consistent with contact map of threshold t. **Results:** The approach used a Matlab tools for improve protein prediction accuracy. The developed tool selects a set of three dimensional coordinates for distance between nodes. Then, the node mapped into protein structure. **Conclusion:** Finally the approach produced a plot-graphical representation for tertiary structure in proteins.

Key words: Protein structure, contact map, algorithm

INTRODUCTION

Contact map used for a protein predictions. In view of this a direct random coordinate was desired, in terms of the contribution of atomic neighbors in making them exposed to predict a protein tertiary structure. Present study aims to prediction of a protein tertiary structure from contact map, how the random prediction of an atomic coordinate is affected by the change threshold t in a neighboring residue, or other positions of these residues. In addition, a graphical method of representing a protein tertiary structure was developed. For the past few years, several tools have been developed in order to help predicting a protein 3D structure to understand protein functionality. Vassura *et al*. (2007) produce a software tool for reconstructing a protein 3D structure form contact map. The tool based on distance geometry which, finds a set of three dimensional coordinates consistent with some given contact map of threshold t. Hu *et al*. (2002) present techniques describe how data mining can be used to extract valuable information from contact map. This tool used contact map to discover 3D structure by test each two amino acid to determine 3D distance by coordinate of α carbon atom. Pollastri and Baldi (2002) used a Neural Network to predict protein contact map and find its 3D structure. The tool focus in grained contact map prediction. The approach concentrates on find a 3D structure from liner sequences of protein. The major task in this approach is to propose and verify precise and robust adaptation rule

to predict contact map. Moré and Wu (1999) developed a tool based on Gaussian smoothing to develop an efficient and reliable code to solve the distance geometry problem in protein structure. The algorithm in this tool work with the sparse set of distance constraints while other algorithm work for distance geometry which tend to work with dense set of constraints.

MATERIALS AND METHODS

Predicting tertiary protein structure from it is primary structure is the greatest problem in the bioinformatics (Fariselli *et al*., 2001), solve this problem go from contact map to the protein structure an efficient and fast algorithm is needed, many of methods introduced to reconstruct contacts prediction in several way (Vassura *et al*., 2008a; 2008b; Hu *et al*., 2002; Moult, 1999; Gomes, 2006; Gutpa *et al*., 2005; Fariselli *et al*., 2001; Casbon, 2002).

Traditionally, contact map is Boolean matrix create from distance map used a pre assigned threshold value t. Distance map D is a N×N matrix where N is the number of residues in a protein and D[i,j] is the distance between coordinate of the α carbon in two residues i and j which measured in Angstroms A°. Two residues i and j in a protein are come in contact with each other if the tertiary distance D[i,j] is less than or equal to some threshold value.

Corresponding Author: Hamed Fawareh, Department of Software Engineering, Zarqa Private University, Jordan

This study developed a method to predict a protein structure from contact map with supported by Matlab tools. The approach taken contains several procedures to finds a set of three dimensional coordinates consistent with contact map of threshold t.

The research method contains three modules as shown in Fig. 1. The scanner module reads the protein ID from the list (extract from PDB). Then accepts the Contact Map (CM) of protein as an input and produces a New Contact Map (NCM) by Scanning method. This process based on prediction quality more than quantity of contacts. The producer module produces distance matrix procedure, which find a possible set of distance between nodes. Then compute a 3D point used nonlinear function from Matlab tools. New contact map is extract based on new coordinates and compare with native contact map to find number of differences, the final module corrector module generate a set of coordinates consistent with the given contact map. Then the module used Matlab plot 3D function to map a protein tertiary structure (Vassura *et al*., 2007).
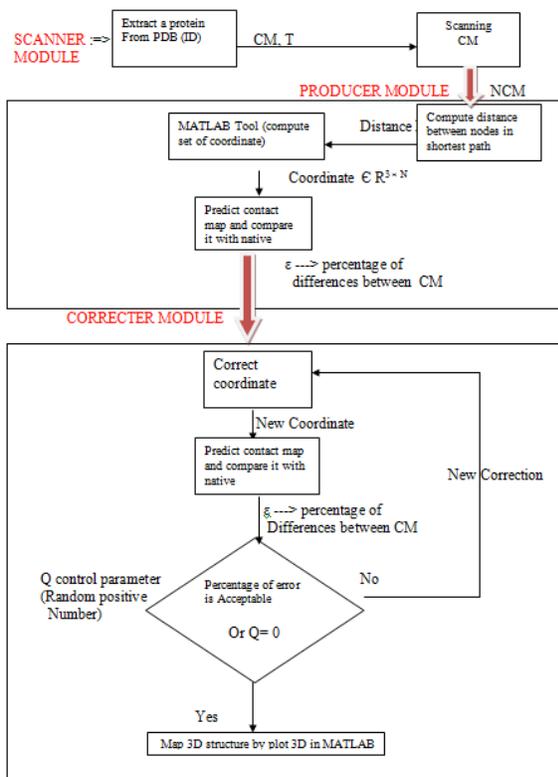

Fig. 1: Proposed method

RESULTS AND DISCUSSION

The experimental results shows that the efficiency of prediction method for a protein structure. We took the list of proteins of different lengths related to the most popular classes from the PDB.

For each protein in the selected list we generate different contact maps by changing the threshold value. The result is automatically analysis for prediction purposes. Based on the result, we scan the contact map with a pre assigned threshold to show the accuracy of extract contact map from dense area. Firstly, the result generated for a part of sequence instead of whole area. Also the method shows the effect of threshold on the tertiary structure of a protein. This study shows some experimental result for different proteins. In addition, the result has been analyzed and compared with the original proteins.

Table 1 shows error percentage of 12 different contact maps before and after the corrector module with average time. The contact threshold changed from 7 to 18 Angstrom. The result analysis shows that the corrector module decreases the error percentage when it applies iteratively. The producer module generate an accurate set of coordinates consist with the native contact map to predict protein tertiary structure. The result shows running time is increase and decrease depending on the threshold value and protein length.

The contact map computed with a threshold equal to 7 Angstrom does not contain enough global information of the protein structure. Also the result is similar to a huge number of protein contact map. The prediction structure is not clear compare with the native structure. Figure 2b shows the prediction of a 451C protein tertiary structure form its contact map at threshold of 16 Angstrom with error percentage 0.08. The prediction result recovered tertiary structure of a protein which is more similar to native tertiary structure as shown in Fig. 2a. The experimental result shows threshold equals to 12 angstrom or more gives more accurate for prediction of a protein tertiary structure.

Table 1: Recovery of tertiary structure from contact map

Threshold value	Error percentage	Error percentage after correct	Average time
7	0.11	0.07	13.30
8	0.10	0.08	17.40
9	0.11	0.09	21.14
10	0.12	0.11	20.50
11	0.13	0.11	23.16
12	0.11	0.09	28.43
13	0.12	0.10	25.30
14	0.12	0.10	22.15
15	0.11	0.08	27.00
16	0.09	0.08	24.10
17	0.09	0.09	20.55
18	0.10	0.08	18.25

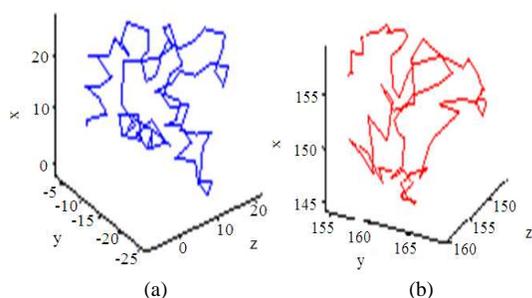(a)   (b)

Fig. 2: (a) 451C tertiary structure (native); (b) prediction of a 451C protein tertiary structure at threshold 16
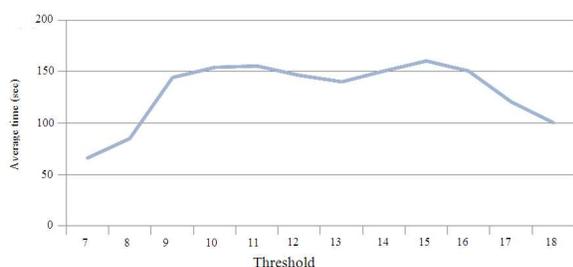

Fig. 3: Threshold and average Time for 6PTI

The experimental result show that the contact map computed using threshold values (12-18) Å allow better tertiary structure recovery than those computed at thresholds (7-9) Å. Also, correct procedure improve the coordinates to obtain the best set consists with native contact map in this approach, as shown in the Table 1.

Figure 3 shows that the running time of predicting protein structure is not effected by threshold value. The average time is decrease and increase in arbitrary way. The average time is changeable because the algorithm takes random number as starting points. The approach used nonlinear system with supported by FSOLVE function as a tools form Matlab.

CONCLUSION

Predicting a protein structure is one of the approaches that have been used in folding a protein 3D structure. For the past few years, several efforts have been developed in order to help predicting a protein 3D structure to understand protein functionality. These efforts used machine learning approaches such as neural network and support vector machine and distance geometric.

This study used contact map matrix to predict tertiary structure of a protein. Also, the result shows that the contact maps computed using threshold values

(12-18) Å allow better tertiary structure recovery than those computed at thresholds (7-9) Å.

The experimental results show that the scanning of contact map for a protein is much more reliable to predict the important areas of the protein structure. The study used the function FSOLVE form Matlab. The approach taken shows an efficient and fast way to predict a protein with a highly prediction accuracy.

REFERENCES

Casbon, J., 2002. Protein secondary structure prediction with support vector machine. M.Sc. Thesis, USA. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.19.2908&rep=rep1&type=pdf

Fariselli, P., O. Olmea, A. Valencia and R. Casadio, 2001. Prediction of contact maps with neural networks and correlated mutations. Protein Eng., 14: 835-543. PMID: 11742102

Gomes, M., 2006. Distance geometry, helix packing and contact map congruency advisors. Sciencestage Inc.
http://sciencestage.com/d/604753/distance-geometry-helix-packing-contact-map-congruency-advisors.html

Gutpa, N., N. Managal and S. Biswas, 2005. Evolution and similarity evaluation of protein structure in contact map space. Proteins Struct. Funct. Bioinform., 59: 196-204. PMID: 15726585

Hu, J., X. Shen, Y. Shao, C. Bystroff and M.J. Zaki, 2002. Mining protein contact map. Proceeding of the Workshop on Data Mining on Bioinformatics, (DMB'02), RPI Exploratory Seed Grant, USA., pp: 3-10.

Moult, J., 1999. Predicting protein three-dimensional structure. Curr. Opin. Biotechnol., 10: 583-588.

Moré, J.J. and Z. Wu, 1999. Distance geometry optimization for protein structures. J. Global Optim., 15: 219-234.

Pollastri, G. and P. Baldi, 2002. Prediction of contact maps by GIOHMMs and recurrent neural networks using lateral propagation from all four cardinal corners. Bioinformatics, 18: S62-S70.

Vassura, M., L. Margara, P. Di Lena, F. Medri and P. Fariselli *et al.*, 2007. Fault tolerance for large scale protein 3D reconstruction from contact maps. Lecture Notes Comput. Sci., 4645: 25-37.

Vassura, M., L. Margara, P. Di Lena, F. Medri and P. Fariselli *et al.*, 2008a. Reconstruction of 3D structures from protein contact maps. IEEE/ACM Trans. Comput. Biol. Bioinform., 5: 357-367.

Vassura, M., L. Margara, P. Di Lena, F. Medri and P. Fariselli *et al.*, 2008b. FT-COMAR: Fault tolerant three-dimensional structure reconstruction from protein contact maps. Bioinformatics, 24: 1313-1315. PMID: 18381401