

***IN SILICO* SEQUENCE ANALYSIS AND PREDICTION OF 3-D STRUCTURE OF SOYBEAN TRYPSIN INHIBITOR (SBTI) BY HOMOLOGY MODELLING**

P.G. Pohekar^{1,2}, N.J.Chikhale³

Department of Biotechnology, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India¹

Assistant Professor, Department of Biotechnology, Abasaheb Garware College, Pune, Maharashtra, India²

Principal, Shri Shivaji Agriculture College, Amravati, Maharashtra, India³

Abstract: Soybean trypsin inhibitor is the enzyme which inactivates the trypsin inhibitor and provides a serum free option when subculturing adherent cells. It acts as an inhibitor only when it is in its native state. The present study deals with the sequence analysis and homology modelling of Soybean trypsin inhibitor. The sequence analysis was performed with respect to secondary structure and antigenicity prediction. The secondary structure prediction reveals the extended strand contributing more to structure, where as the antigenicity prediction confirmed the presence of antigenic determinant. The 3D structure was predicted by using modeller. The resultant structure was evaluated by procheck, errat and verify 3D program showing the compatibility of the residues in the structure predicted.

Keywords: Soybean trypsin inhibitor, Homology modelling, 3-D structure

I. INTRODUCTION

Trypsin inhibitors are enzymes that reduce the availability of biologically active trypsin. There are various sources for trypsin inhibitors such as serum antitrypsin, lima beans, bovine pancreas, ovomucoids and soybeans. Soybean meal, obtained after oil extraction, is also widely used as the most cost-effective feed component for many aquaculture animals [1]. However, the nutritional value of soybean meal is much lower than expected, in spite of its protein content and amino acid profile of the proteins. This is largely attributed to the presence of antinutritional factors, such as protease inhibitors. Soybean trypsin inhibitor is one of the antinutritional protein present in the soybean which reduces the nutritional quality of soybean protein. Also their presence in the diet is known to reduce the growth rate of young monogastric animals [2]. Soybean trypsin inhibitor (SBTI) is a reversible competitive inhibitor of trypsin and other trypsin like proteases. Knowledge of 3-D structure of soybean trypsin inhibitor can help us to uncover more between the complex of trypsin-antitrypsin. So in order to uncover this, it is most necessary to know the 3-D structure of a protein, which may be achieved by means of X-ray crystallography or NMR spectroscopy. The experimental technique however are more time consuming and tedious also may not be always succeed in determining the structure of all the proteins, since large amount of data is generating day by day therefore creating a gap between available sequences and experimentally solved structure. Computational biology which deals with the computational approach to analyze biological data *in silico*, help in this regard to reduce the gap. The method like homology modelling [3] is one of them useful to resolve the structure. As the proteins are the result of previously existing one due to the process of continuous evolution of their ancestors therefore proteins can be grouped into families and the member of same family have same similar fold pattern. This fact utilized to predict the structure, if the structure of any of family protein is known, and the technique by means of which it is achieved is known as homology modelling. Modeller [4] is the program by means of which homology modelling is achieved based on the molecular probability density function. Study reveals that model predicted by modeller stand best with respect to SWISS-PROT, which predict the structure online [5].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

II. METHODOLOGY

For the prediction of 3D structure, the amino acid sequence of trypsin inhibitor was downloaded from the SWISS-PROT [6], having the accession number Q9LLX2. The sequence length reported to be 168 amino acids. A template selection search was performed using BLAST-P [7] against PDB [8] database from NCBI interface. Simultaneously “Template Identification Tool” at Swiss-Model interface [9] provided by Swiss Institute of Bioinformatics was also utilized for template selection. The BLAST output shows the 16 significant hits having the e-value almost equal to zero. Out of which the best model selected as a template was 1R8N from the species *Delonix regia* sharing about 81% of similarity with the query sequence. Finally the structure was predicted by using Modeller program.

III. SEQUENCE ANALYSIS

Sequence analysis was done for secondary structure prediction by using Self Optimized Prediction Method i.e. SOPM [10] and antigenicity prediction by using Tongaonkar and Kolaskar [11] method. The SOPM results for soybean trypsin inhibitor were, random coil (30.36 %) and extended strand (37.50 %) contributing more where as the alpha helix (16.67 %) and beta turn (15.48 %) contribute less. In whole sequence of Soybean trypsin inhibitor there are 6 antigenic determinants in, out of which determinant no. 1 and 5 having the peptide sequence SNCPVTVLQDYSE (13) and SIQKYKFGYKLVFCITGSGTCLD (23) respectively are potentially antigenic considering the highest antigenic propensity.

IV. EXPERIMENTAL RESULTS

For prediction of 3-D model of Soybean trypsin inhibitor by homology modeling, the Align2D command from Modeller was utilized to align query sequence to template structures, followed by generation of models by Loop-Model class. Overall five models were allowed to generate, after refinement the single model was selected on the basis of lowest molecular probability density function (mol-pdf) value. The mol pdf value for each structure is given in the table 1. The model so produced subject to evaluation programs like Procheck [12], ERRAT [13] and Verify 3D [14, 15]. The Procheck results shows that out of 168 amino acid in the sequence, only one amino acid was fallen in disallowed region, representing the degree of accuracy in the predicted structure (refer figure 1, table 2). The ERRAT result shows the overall quality factor 53.927. As shown in figure 2 the possibility to reject the region that exceeds the error value is limited and may not belong to the residues fall in the active site. The predicted structure was also confirmed by using verify 3D program, which shows the value of 0.2 or above for most of the residues indicating the compatibility of the residues with structure predicted. The predicted structure (figure 3a and b) submitted to Protein Model DataBase (PMDb), and it is available online having id: PM0079063.

Sr. no.	Loop model	Mol pdf value
1	Query.BL00010001.pdb	261.39584
2	Query.BL00010002.pdb	1836.04016
3	Query.BL00010003.pdb	1144.58972
4	Query.BL00010004.pdb	1602.32702
5	Query.BL00010005.pdb	610.19904

Table1: Summary of the successfully completed loop models

Sr. No.	Residues in most favoured regions		Residues in additional allowed regions		Residues in generously allowed regions		Residues in disallowed regions	
	%	number	%	number	%	number	%	number
1	87.2	116	8.3	11	3.8	5	0.8	1

Table2: Procheck results of Non Proline and Non Glycine residues for models produced using 1R8N as template

**International Journal of Innovative Research in Science,
Engineering and Technology**

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

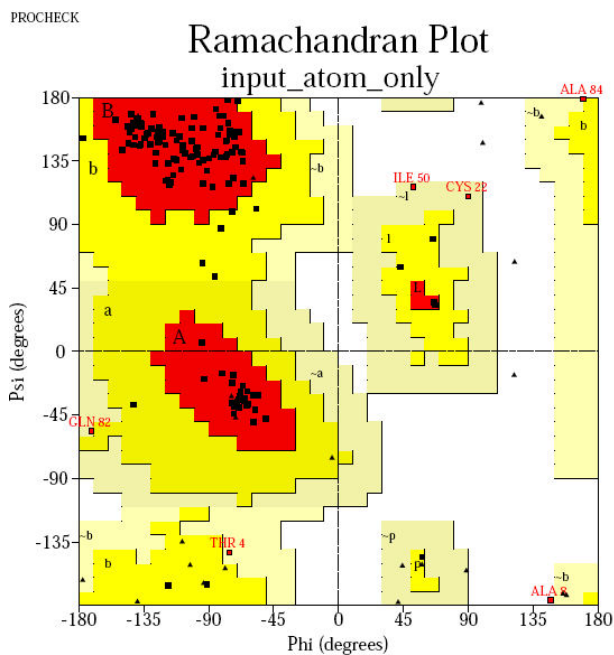


Figure1: Procheck result for best model created using modeller for SBTI

Program: ERRAT2
File: /var/www/html/Services/ERRAT/DATA/514864.pdb
Chain#:1
Overall quality factor*: 53.947

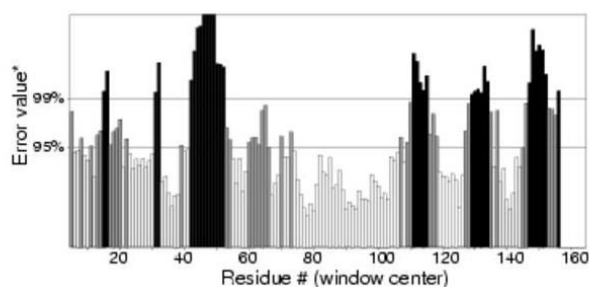


Figure 2: ERRAT result for the predicted model

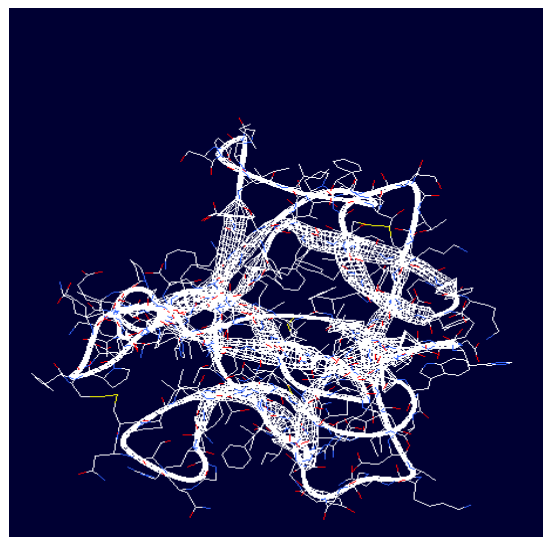
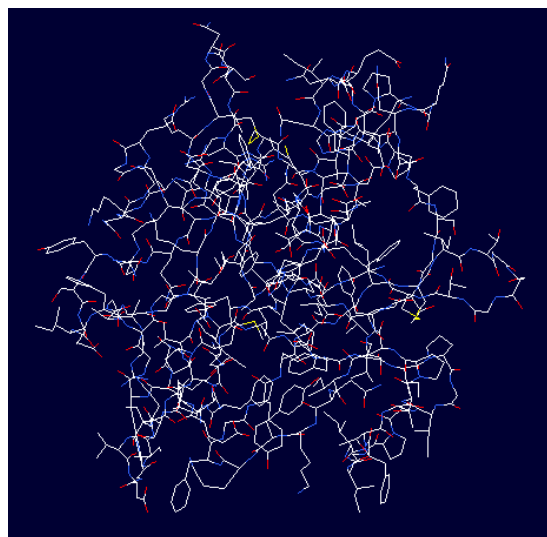


Figure 3 a) Model generated and selected after loop modelling

b) Ribbon form

V. CONCLUSION

Present study deals with the *in silico* sequence analysis and prediction of 3-D structure for soybean trypsin inhibitor. The antigenicity prediction shows the presence of antigenic determinants. The model predicted using modeller stand good on several aspects indicating the feasibility of the model generated. The predicted structure can also be used for molecular docking studies for more insight into the structure.

REFERENCES

- [1] Storebakken, T., Refstie, S. and Ruyter, B., "Soy products as fat and protein sources in fish feeds for intensive aquaculture". In: Drackley, J.K. (Ed), *Soy in Animal Nutrition*. Fed Anim. Sci. Sco., Savoy, II, USA, pp. 127-170, 2000.
- [2] Van Damme Els, J. M., Peumans, W. J., Pusztai, A. and Bardocz, S., "Handbook of plant lectins: Properties and biomedical applications" pp. 224. Chichester: Wiley.1997.
- [3] Krieger, E., Nabuurs, S. B. and Vriend, G., "Homology Modeling, in Structural Bioinformatics", Volume 44 (eds P. E. Bourne and H. Weissig), John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471721204.ch25,2005.
- [4] Sali A., et al. *J. Mol. Biol.* **234**: 779 [PMID: 8254673],1993.
- [5] Meshram R., J., et al, "Sequence analysis and homology modeling of laccase from *Pycnoporus cinnabarinus*" *Bioinformatics*. Vol. 5(4),2010.
- [6] Boeckmann B., et al. *Nucleic Acids Res* **31**:365 [PMID:12520024],2003.
- [7] Stephen F., et al. *Nucleic Acids Res.* **25**:3389 [PMID:9254694],1997.
- [8] Westbrook J., et al. *Nucleic Acids Res.* **31**: 489 [PMID:12520059],2003.
- [9] Schwede T. et al. *Nucleic Acids Res.* **31**:3381 [PMID:12824332],2003.
- [10] Geourjon C., Deleage G., "SOPM: a self-optimized method for protein secondary structure prediction." Vol. 7(2):157-164,1994.
- [11] Kolaskar, A. S., Tongaonkar, P. C., "A semi-empirical method for prediction of antigenic determinants on protein antigens." *FEBS Lett.* Vol.10;276(1-2):172-4. 1990.
- [12] Laskowski, R. A., et al. *J. Appl. Cryst.* **26**: 283,1993.
- [13] Colovos, C. and Yeates, T. O., "Verification of protein structures: patterns of nonbonded atomic interactions". *Protein Sci.* vol;2 (9):1511-9. 1993.
- [14] Bowie, J. U., Lüthy, R., and Eisenberg, D. A., "Method to identify protein sequences that fold into a known three-dimensional structure". *Science.* vol 12;253(5016):164-70.1991.
- [15] Lüthy, R., Bowie, J. U., and Eisenberg, D., "Assessment of protein models with three-dimensional profiles". *Nature.* 5;356(6364):83-5,1992.