ISSN 0975 - 6299 Vol 2 / Issue 3 / Jul - Sept 2011

International Journal of Pharma and Bio Sciences

RESEARCH ARTICLE BIO INFORMATICS

HOMOLOGY MODELLING OF POLYPHENOL OXIDASE FROM *SOLANUM* MELONGENA: SEQUENCE ANALYSIS AND STRUCTURAL VALIDATION STUDIES – *In Silico.*

MEDHAVI MALLICK[1], KOEL MUKHERJEE[1,*] AND NEETHA A.UDAYAKUMAR[2]

[1]Depertment of Biotechnology, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.
[2]Depertment of Biotechnology, Waljat College of Applied Sciences, Muscat, Sultanate of Oman.

KOEL MUKHERJEE
Depertment of Biotechnology, Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India.

*Corresponding author

ABSTRACT

Polyphenol oxidase (PPO) typically found in the chloroplasts of plants, is an enzyme that brings about browning in fruits and vegetables. This browning is a common phenomenon which leads to decreased market value and economic loss. In order to interpret the mechanism of process by which Polyphenol oxidases in Solanum melongena are making browning reaction, it is important to know its 3D structure. Homology modeling was done by Modeller and Geno3D with a template sequence of PPO of Vitis vinifera. The 3 D structure of the protein was evaluated and validated using PROCHECK and Verify_3D. The favored and unfavored regions of the amino acid residue were indicated by the Ramachandran plot. The results represented 224 numbers of hydrogen bonds, 15 helices, 11 strands and 50 numbers of turns. The modeled protein structure was subjected to *In silico* analysis using various bioinformatics tools. The significance of our study focuses on the 3 D structure prediction of this enzyme and *In silico* analysis of its secondary structure.

KEYWORDS

Polyphenol oxidase (PPO), Browning, Homology modeling, Motif, Domain.

INTRODUCTION

Browning is a chemical process which occurs in fruits and vegetables and results in brown pigments. It can be observed in fruits (apricots, pears, bananas, grapes)[1], vegetables (potatoes, mushrooms, eggplant) and also in seafood (shrimps, spiny lobsters and crabs). Browning reaction is detrimental to quality, particularly in post-harvest storage of fresh fruits, juices and some shellfish. It is responsible for up to 50% of all losses during fruit and vegetables production. Fruits and vegetables can turn brown through either enzymatic or non-enzymatic processes[2]. Enzymatic browning is also known as chemical browning where polyphenol oxidase (PPO) and peroxidase (POD) are two well known enzymes involved in the process [3,4]. Enzymatic browning is the main function of PPOs in fruits and vegetables, and responsible for unpleasant sensory qualities [5]. Alteration in the appearance like skin colour, texture significantly lowers the market value as appearance is the main feature for a consumer to appraise the quality of goods.

The phenolic oxidation reaction [6] is the main cause of browning. The damage of cellular compartments allows the phenolic substrates to be accessible to Polyphenol oxidase(PPO) which catalyzes the phenolic oxidation [7]. The activity of PPO is the major factor developing the browning problem. Polyphenol oxidases are a widespread group of enzymes found in the chloroplasts of plants. PPO is able to catalyze the transformation of an array of aromatic compounds that have two adjacent phenolic groups on them. The oxidation of phenolic groups produces a number of reactive oxygen molecules known as quinines. Such quinones are very reactive and can react with each other and surrounding proteins to generate a black pigment [8] called melanin. This causes dark spots on the fruit skin, frequently making the fruit or vegetable inedible. This browning causes the deterioration of fruits and vegetables, resulting in large economic losses.

PPO + Aromatic compound with two phenolic group ⟶ Quinines ⟶ Melanin

X-Ray Crystallography or NMR spectroscopy techniques are used to discover experimentally, the functionality of any protein using its 3D structure [9]. But these experimental techniques are very monotonous and prolonged. Nowadays the rate at which protein sequence and structural data is accumulated using different bioinformatics software is far speedier than the information we get from the weight lab and as a result creating a gap. Different *In silico* strategies like homology modeling [10], docking studies, structure-function activity can help in reducing this gap. The 3D structure of a protein or an enzyme and its validation can be done computationally within a very short period of time. Here, in this study, we studied the structure of oxidative enzyme PPO of *Solanum melongena* (brinjal). Inhibition of PPO activities can reduce the browning process. Three–dimensional models of brinjal PPO were constructed from the knowledge of template structure, analysis was done on secondary structure and binding site of

the enzyme. The aim of this study was to investigate the mystery of browning problem in *Solanum melongena* (brinjal) by sequence and structure analysis.

MATERIALS AND METHODS

(i) Sequence Retrieval & Selection of template:

The amino acid sequence of the target protein was obtained from NCBI database (http://www.ncbi.nlm.nih.gov/protein/CAA81798.1) with accession number ACR61399. This protein was reported to have 1 to 546 amino acid residues. The template selection was performed using BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) [11] with PDB database (http://www.rcsb.org/pdb/home/home.do). The protein sequence with a good query coverage and E-value of 3e-76 was selected as template. The template was of Vitis vinifera which has a 3D structure with PDB file name 2P3X.

(ii) Modeling of 3-D structure:

Homology modeling was done using MODELLER 9v7 (http://www.salilab.org/modeller/) [12]. To predict the unknown structure, the software needs one known protein structure. So, for the query sequence, we use the template sequence of *Vitis vinifera*. Parallely with MODELLER, attempt was also made to model the query sequence with Geno3D software (http://geno3d-pbil.ibcp.fr), which is freely available.

(iii) Evaluation and validation of the 3-D structure:

All the predicted models were evaluated using ProCheck (http://www.ebi.ac.uk/thorntonsrv/software/PROCHECK/), Verify_3D (http://nihserver.mbi.ucla.edu/Verify_3D/). Models generated by MODELLER and Geno3D were ranked by Q-Mean server (http://swissmodel.expasy.org/qmean/cgi/index.cgi).

(iv) 2D Structure and Domain Analysis:

2D structure of polyphenol oxidase was predicted by PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/), a highly accurate method for protein secondary structure prediction. Domain analysis carried out on "InterPro Domain Scan" [13] revealed presence of Di-copper centre containing domain, a polyphenol oxidase containing centre domain and a tyrosinase.

(v) Motif Identification:

For motif identification of PPO sequence of brinjal we have done the De novo Motif Discovery by MEME (http://meme.nbcr.net/meme4_6_1/intro.html). Total six sequences were given from the result of ClustalX [14] to generate the location of motifs and their length depending upon the similarities of these sequences.

RESULTS

The template sequence was selected using BLASTp and Phi-BLAST [11] before proceeding to homology modeling to obtain the accurate sequence. BLASTp find many numbers of homologous sequences for our query but the template, having the highest query coverage to PPO was chosen. 2P3X was the template, the sequence of PPO of *Vitis vinifera (*grape*)* which had a fairly good sequence similarity with *Solanum melongena.* The template selected for the query sequence was reported to have an E-value of 3e-76. To obtain the accurate template PSI-BLAST was done in parallel with 100 iteration and the selected template was the same one. Modeller 9v7 was utilized to align query sequence to template structures, followed by generation of model. At the same time Geno3D server was also used to generate models using

same template . Four models were allowed to build from the same template (Table 1).

Comparison of quality of models produced by Modeller and Geno3D server

Method	Model Name	Template	Model Energy (kcal\mol)	Residue in Core region in Ramachandan's Plot	Residue in Disallowed region in Ramachandan's Plot
Geno3D	Model 1.pdb	2P3X	-13950.10	71.4%	1.7%
	Model 2.pdb	-do-	4395.32	68.7%	2.7%
	Model 3.pdb	-do-	-14067.50	71.0%	3.0%
Modeler	query.B99990001.pdb	-do-	-45896.324	85.8%	1.9%

Table 1

All the models (three for Geno3D and one for Modeller) are having the same template as 2P3X. Different columns are representing the different energy values in Kcal/mol, percentage of residues which are in core and disallowed region in Ramachandans Plot. The energy values indicate the reliability and the precision of the structure.

From the above table it was clear that the homology model done by Modeler was the perfect among the others, so we proceed for *In Silico* analysis of sequence and structure with this query.B99990001.pdb. A superimposed image (Fig 1) of the brinjal PPO homology model (query.B99990001.pdb) and the grape PPO model was also done to find its resemblance. The two structures seemed to be very much similar to each other. The superimposed structure clearly showed the helices, strands and coils. This homology model represents 224 numbers of hydrogen bonds, 15 helices, 11 strands and 50 numbers of turns. The analyzing of the binding site led to the detection of presence of clefts and cavities in the protein. These clefts and cavities highlight the key residues in the protein, mainly with respect to its functional site. During superimposition, the stability of the structure was also determined.

Superimposition of the brinjal polyphenol oxidase and template

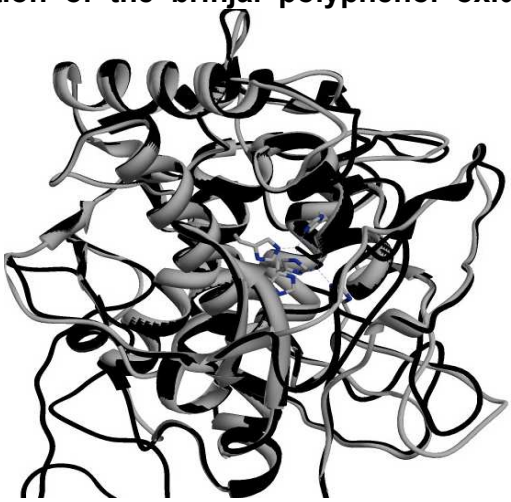

Figure 1

Cartoon representation of superimposition of homology model (query.B99990001.pdb) in grey colour and grape polyphenol oxidase (template, 2P3X.pdb) in black colour using chimera.

Following superimposition, the sequence was compared with non-redundant set of PDB structures, where the z-score was recorded. The graph was plotted between Q Mean score and the size of the nucleotide (Fig.2). Results of z-scores and QMEAN (Table 2) specify the reliability of the model.

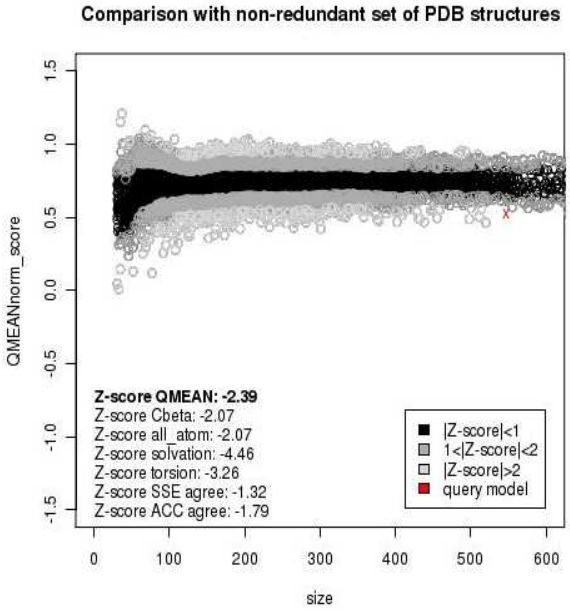

Figure 2

QMEAN score is a composite score consisting of a linear combination of 6 terms. The pseudo-energies of the contributing terms are given above together with their Z-scores with respect to scores obtained for high-resolution experimental structures of similar size solved by X-ray crystallography. The Red Cross is the target model structure whose Z score is -2.39.

Depiction of total QMEAN value

Model name	query.B99990001.pdb
C_beta interaction energy	-26.34 (Z-score: -2.07)
All-atom pairwise energy	-4849.07 (Z-score: -2.07)
Solvation energy	22.09 (Z-score: -4.46)
Torsion angle energy	-34.61 (Z-score: -3.26)
Secondary structure agreement	72.7% (Z-score: -1.32)
Solvent accessibility agreement	71.6% (Z-score: -1.79)
Total QMEAN-score	**0.527** (Z-score: -2.39) (estimated model reliability between 0-1)

Table 2

The table depicts the total QMEAN score, 0.529 of the model structure. Different energies are given here.

The Ramachandran plot was carried out to analyze the residues in the allowed and disallowed regions. The Psi/Phi distribution of the eggplant polyphenol

oxidase homology model (using 2P3X as template) was produced by PROCHECK [15]. It was observed that the residues in the allowed regions were much higher in number, in contrast to the residues in the disallowed regions (Fig 3). The plot of brinjal PPO revealed that the number of non-glycine and non-proline residues was 479. Out of these 411 (85.80%) were in the most favoured regions for brinjal PPO.

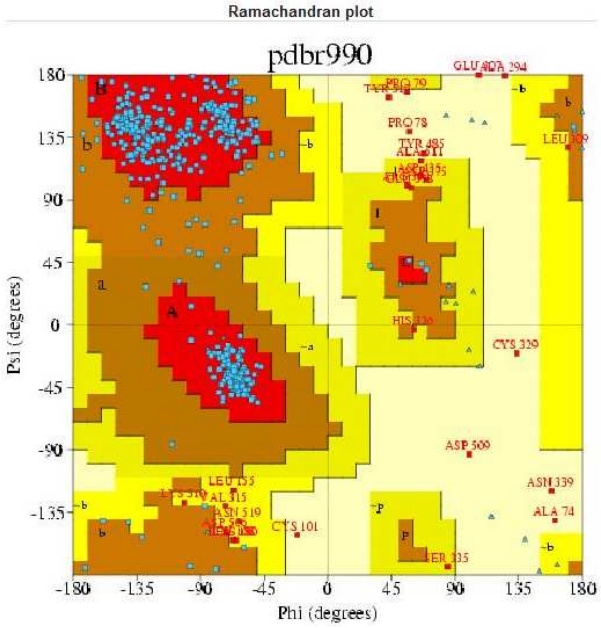

Figure 3

Ramachandran plot of the Psi/Phi distribution of the brinjal Polyphenol Oxidase homology model (using 2P3X as template) produced by PROCHECK. The favored and most favored regions are yellow and red, respectively. Pale yellow is the generally allowed region and disallowed region is white.

After the construction of homology model and the validation of that model was completed, the secondary structure, domain analysis and motif prediction were done. To predict the 2D structure JPRED was carried out. The result shows the number of helices, strands and coils with their location in the sequence and their length (Fig 4).

```
# PSIPRED HFORMAT (PSIPRED V3.0)

Conf: 9999866344443467765446999987446788885568999888986368770799999
Pred:
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEECCCCCC
 AA: MASLPWSLTTSTAIANTTNISAFPPSPLFQRASHVPVARNRSRRFAPSKVSCNSANGDPN
          10        20        30        40        50        60

Conf: 999876434789974511466531124443047889998877888789999988867789
Pred:
CCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
 AA: SDSTSDVRETSSGKLDRRNVLLGIGGLYGAAGGLGATKPLAFGAPIQAPDISKCGTATVP
          70        80        90       100       110       120

Conf: 999998657999767878779999998710456324999999999999999971999998
Pred:
CCCCCCCCCCCCCCCCCCCCCCCCCCCCEECCCCCCCCHHHHHHHHHHHHHHHCCCCCCC
 AA: DGVTPTNCCPPVTTKIIDFQLPSSGSPMRTRPAAHLVSKEYLAKYKKAIELQKALPDDDP
         130       140       150       160       170       180

Conf: 107788765045567887656843420001685354468999999999999985089998
Pred:
CHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCC
 AA: RSFKQQANVHCTYCQGAYDQVGYTDLELQVHASWLFLPFHRYYLYFNERILAKLIDDPTF
         190       200       210       220       230       240

Conf: 000223443048997532358867778989999986766565801115578899999988
Pred:
CCHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEECCCCCCCCCCCCC
 AA: NLAIMYKQIVSGATTPKLFLGYPYRAGDAIDPGAGTLEHAPHNIVHKWTGLADKPSEDMG
         310       320       330       340       350       360

Conf: 833578885133333446779999987299999888888889983223357875432322
Pred:
CCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
 AA: NFYTAGRDPIFFGHHANVDRMWNIWKTIGGKNRKDFTDTDWLDATFVFYDENKQLVKVKV
         370       380       390       400       410       420

Conf: 435998844667767787431799934799988099999935999984869990477519
Pred:
CEEEEEEEECCCCCCCCCCCCEEEEEEECHHHHHHHCCCCCCCEEEEEEEEECCCCCEEEEECEE
 AA: EFAGSFVNVPHKHMKEMKTKTNLRFAINELLEDLGAEDDESVIVTIVPRAGGDDVTIGGI
         550       560       570       580       590       600

Conf: 9998039
Pred: EEEECCC
 AA: EIEFVSD
```

Figure 4

Snapshots of secondary structure of brinjal PPO predicted by PSIPRED. Some regions were selected to show the helixes, coils and strands with H representing an alpha helix, C representing a coil and E representing a Beta-strand.

Domain is a unit in the protein structure that can act independently and individually. Domain analysis has had a deep impact on the study of individual protein as it can be the most useful level where the structure can be correlated with its function [16]. The complete sequence of brinjal PPO was pasted to 'InterPro Domain Scan' to reveal the nature of the domains and number of domains. Mainly PFAM (www.sanger.ac.uk/resources/databases/**pfa**

m.html) and SUPERFAMILY [17] predict the Tyrocinase, Polyphenol oxidase a central domain, Polyphenol oxidase in C-terminal and Di-Copper domain centre-containing domains (Fig 5).

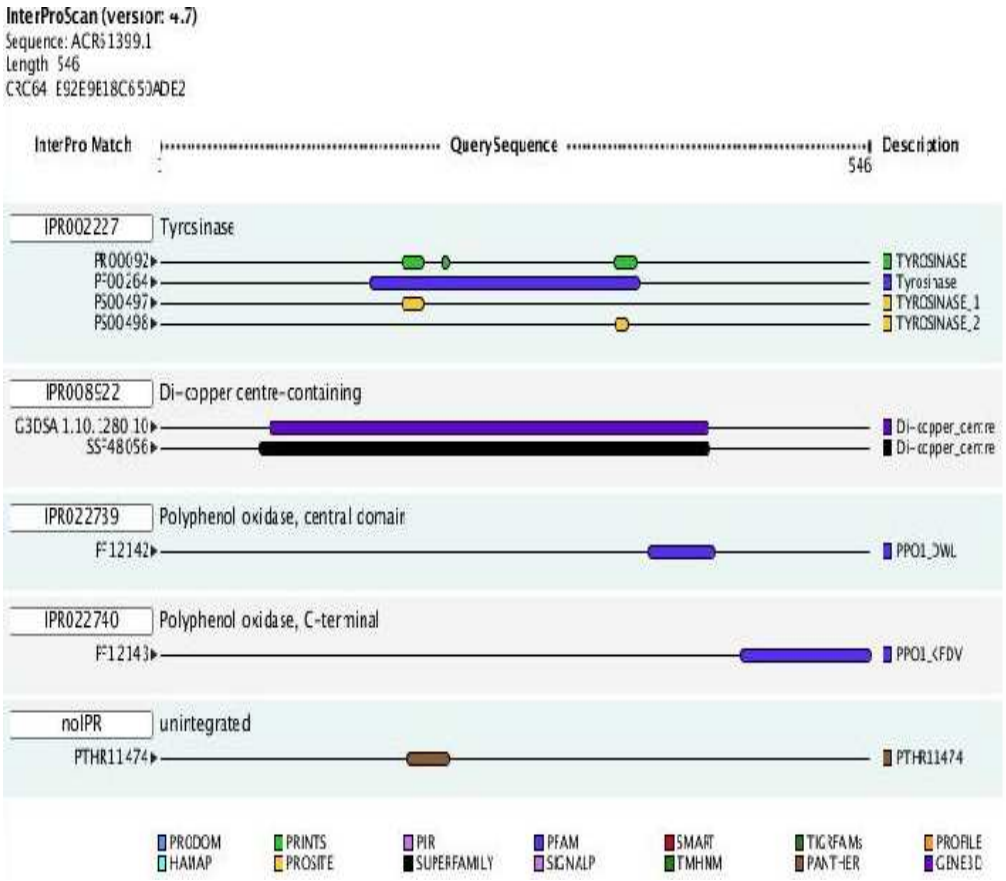

Figure 5

Schematic diagram of different domain analysed by InterProScan (V 4.7). Domains were generated using different databases. The PFAM and SUPERFAMILY results showing the violet and black bar within the query sequence to identify the types of domains.

After domain prediction, motif was analysed by MEME. In MEME submission, the minimum width was 2 and the maximum was given 50. Total 10 motifs were predicted with their E-value, Width, Sites, Log Likelihood Ratio and Information Content [18]. The first motif was having the best result and it is found in the sequence range of 175 to 230, around 50 width. Next good motif was number 2 with width more than 50 and from position 370 to 425. These two motifs were shown as sequence logo (Fig6) with their starting site and the E value.

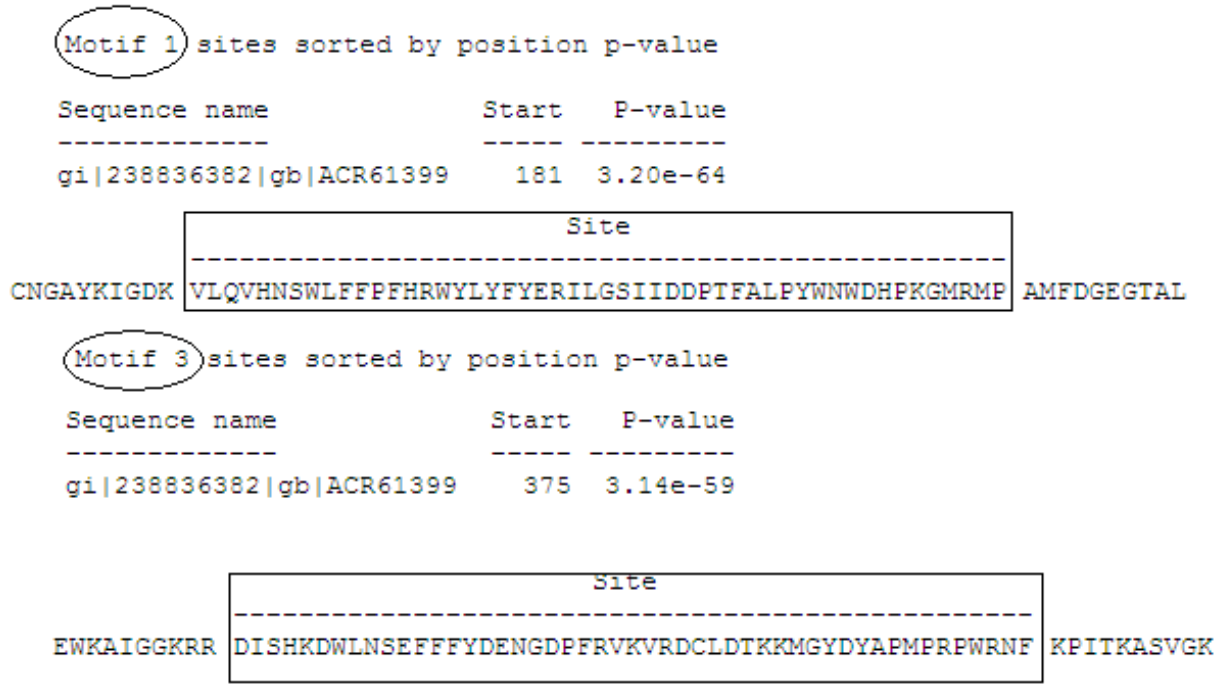

Figure 6

MEME motifs are represented by position-specific probability matrices that specify the probability of each possible letter appearing at each possible position in an occurrence of the motif. These are displayed as "sequence LOGOS", containing stacks of letters at each position in the motif. Here two motifs were shown with highest sequence length of modelled PPO.

CONCLUSION

It is evident from Table1 and Figure3 that the best model created using template 2P3X employing Modeller program (query.B99990001.pdb) has 86% of its Non-Proline and Non-Glycine amino acids residues in core region, 9% in allowed region, 3% in generously allowed region and 1.5% in disallowed region. The Modeller program was only giving the homology structure of PPO with 546 sequence length as compared to Geno3D which was giving three models but with less sequence length. Superimposition (Figure 1) was also done to see the matches and mismatches regions of the structure of brinjal PPO and the template. To carry on the *In silico* studies with the modelled structure the structure validation was important. The QMEAN report reveals that our model was within the range of the score normally found for proteins of comparable size. The constructed models were checked for accuracy using VERIFY 3D. In secondary structure PSIPRED analysis shows many helices and strands which was confirmed by the result of Modeller. The position of helices and strands were important for domain and motif analysis. Mainly three domains were identified by 'InterProScan' and ten motifs were found by MEME. Three domains were identified, Tyrocinase (200-325), Di-Copper (100-400) and Polyphenol Oxidase (325-546). Three largest motif with width 50, Motif 1 (180-230) and Motif 3 (374-424), Motif 6 (458-507) were within the Di-copper and PPO domain region. The JPRED result also revealed that first two motifs are having a good length of helices which make the protein more compact and the third motif build up with strands. After analysing the 2D and 3D structure of PPO we can go forward for identifying the ligand which can bind to the active site of the PPO and inhibit the action of browning.

ACKNOWLEDGMENT

The authors acknowledge BTISnet, Department of Biotechnology, Government of India, New Delhi (No.BT/BI/04/065/04), and the Department of Biotechnology, Birla Institute of Technology, Mesra for providing Infrastructure facility for Bioinformatics Research.

REFERENCES

1. G.W. Cheng, C.H. Crisosto, Browning Potential, Phenolic Composition, and Polyphenoloxidase Activity of Buffer Extracts of Peach and Nectarine Skin Tissue, J. Amer. SOC. Hort. Sci., 120 (5): 835-838, (1995).
2. G.M. Sapers, Browning of foods: Control by sulfites, antioxidants, and other means, Food Technol., 68:5–84, (1993).
3. A.F. Tomas-Barberan, J.C. Espin, Phenolic compounds and related enzymes as determinants of quality in fruits and vegetables, J. Sci. Food Agric., 81(9):853–876, (2001).
4. A. Yemenicioglu, M. Ozkan, B. Cemeroglu, Some characteristics of polyphenol oxidase and peroxidase from taro (Colocasia antiquorum), Tr. J. Agric. Forestry., 23:425–430, (1999).
5. A. Sanchez-Amat, F. Solano, A pluripotent polyphenol oxidase from the melanogenic marine *alteromonas* shares catalytic capabilities, Biochem Biophys Res. Comm., 240:787-792, (1997).
6. A.G. Mathew, H.A.B. Parpia, Food browning as a polyphenol reaction, Adv. Food Res., 119:75-145, (1997).
7. A.M. Mayer, E. Harel, Polyphenol oxidases in plants, Phytochemistry, 18:193-215, (1979).
8. V. Falguera, J. Pagan, S. Garza, A. Garvin, A. Ibarz, Inactivation of polyphenol oxidase by UV-VIS irradiation, protective effect of melanins from Tyrosine, International conference on food Inovation, 25-29th October, (2010).
9. B.R. Szymczyna, R.E. Taurog, M.J. Young, J.C. Snyder, J.E. Johnson, J.R. Williamson, Synergy of NMR, Computation, and X-Ray Crystallography for Structural Biology. Structure, 17 (4): 499–507, (2009).
10. P. Nokthai, V.L. Sanghiran, L. Shank. Molecular Modeling of Peroxidase and Polyphenol Oxidase: Substrate Specificity and Active Site Comparison, Int. J. Mol. Sci., 11 (9): 3266-3276, (2010).
11. S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res., 25 (17): 3389-402, (1997).
12. N. Eswar, M.A. Marti-Renom, B .Webb, M.S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali, Comparative Protein Structure Modeling With MODELLER, Current Protocols in Bioinformatics, 15 : 5.6.1-5.6.30, (2006).
13. E.M. Zdobnov, R. Apweiler, InterProScan - an integration platform for the signature-recognition methods in InterPro, Bioinformatics, 17 (9): 847-848, (2001).
14. R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson Multiple sequence alignment with the Clustal series of programs, Nucleic Acids Res., 31 (13): 3497-500, (2003).
15. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structures, J Appl Cryst., 26 (2): 283-291, (1993).
16. R.R. Copley, T. Doerks, I. Letunic, P. Bork, Protein domain analysis in the era of complete genomes. FEBS Letters, 513 (1): 129-134, (2002).
17. J. Gough, The SUPERFAMILY database in structural genomics, Acta Cryst., 58 (11): 1897-1900, (2002).
18. L.T. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, AAAI Press, 28-36, (1994).