Bioinformation Volume 5

open access

www.bioinformation.net Issue 4

Hypothesis

Sequence analysis and homology modeling of laccase from *Pycnoporus cinnabarinus*

Rohan J Meshram*, A J Gavhane, R B Gaikar, T S Bansode, A U Maskar, A K Gupta, S K Sohni, M A Patidar, T R Pandey, S N Jangle

Center for Biotechnology, Pravara Institute of Medical Sciences, Loni, Taluka: Rahata, District: Ahmednagar, Maharashtra, India; R J Meshram – Email: rohan_meshram@rediffmail.com, Phone:+919822725648; *Corresponding author

Received July 26, 2010; Accepted August 26, 2010; Published September 20, 2010

Abstract:
Industrial effluents of textile, paper, and leather industries contain various toxic dyes as one of the waste material. It imparts major impact on human health as well as environment. The white rot fungus *Pycnoporus cinnabarinus* Laccase is generally used to degrade these toxic dyes. In order to decipher the mechanism of process by which Laccase degrade dyes, it is essential to know its 3D structure. Homology modeling was performed in presented work, by satisfying Spatial restrains using Modeller Program, which is considered as standard in this field, to generate 3D structure of Laccase in unison, SWISS-MODEL web server was also utilized to generate and verify the alternative models. We observed that models created using Modeller stands better on structure evaluation tests. This study can further be used in molecular docking techniques, to understand the interaction of enzyme with its mediators like 2, 2-azinobis (3-ethylbenzthiazoline-6-sulfonate) (ABTS) and Vanillin that are known to enhance the Laccase activity.

Keywords: Homology modeling, Spatial restraints, Modeller, Laccase, QMEAN, Ml methods, Beale restart conjugate gradients method, Leap-frog verlet integrator.

Background:
In developing countries including India textile, leather and paper industries represent an important economic sector. Huge amount of capital and Human resource is engaged in these industries. These industries are one of the most important sources of Environmental pollution. Mills of these industries emancipate enormous amount of waste matter each year, that contain variety of chemicals such as formaldehyde, chlorine, heavy metals (such as lead and mercury) and toxic dyes, which lay noteworthy foundation of environmental degradation and human illnesses. Most of the dyes that are released from these industries are polymers possessing very complex structure and are very difficult to decompose biologically **[1]**. Many reactive dyes are not degraded in ordinary aerobic sewage treatment processes and that they can be discharged unaffected from the treatment plant **[2]**. An even very minute concentration of dyes in effluent is visible and is often carcinogenic **[3]**. Laccase belongs to group of enzyme named as large blue copper proteins or blue copper oxidases possessing polyphenol oxidase activity. It functions by generally reducing oxygen to water simultaneously oxidizing a polyphenolic substrate .Laccase has evolved with a remarkable property of non-specificity of its reducing substrates and encompass vast range of substrates oxidized **[4]**, making it a marvelous contrivance to oxidize toxic dyes which are generally polyphenols**[5]**. Knowledge of 3D structure of Laccase can aid us to uncover the mystery of how Laccase has attained such huge functional diversity. To experimentally discover functionality of any protein, the information of its 3D structure remains an indispensable fact, which is achieved using techniques like X-Ray Crystallography or NMR spectroscopy. Experimental techniques are very tedious and prolonged and not always succeed in determining structure for all proteins especially membrane proteins **[6]**. Moreover, the rate at which protein sequence data is accumulating is far more than the structural information available, thus creating a gap between available sequences and experimentally solved structures. Computational methods like homology modeling can help reduce this gap. It is known that existing proteins are result of continuous evolution of previously existing ones, thus proteins can be grouped into families. Members in same family are similar and thus have similar folds; this fact allows predicting the structure of other members of family if structure of single member is known and the technique by which this task is achieved is termed as Homology Modeling. Modeller **[7]** is stand-alone package for homology modeling that accomplishes the job by method called 'Satisfaction of spatial restraints' using a set of restraints derived from the alignment and expressed as Probability Density Functions, finally the model is obtained by minimizing the violations to these restraints. Studies have proved that Modeller outperforms most of other homology modeling suits, it's fast, reliable and freely available and hence we selected it in current study **[8]**.

Methodology:

Sequence Retrieval and Template selection:
The sequence of laccase enzyme of "Pycnoporus cinnabarinus" was retrieved from SWISS-PROT **[9]** database with accession number O59896. The complete sequence length of Laccase was reported as 518 amino acid residues which is synthesized as inactive precursor having first 21 residues as signaling peptide, the remaining 22 to 518 residues constitute the functional domains of the enzyme **[10]**. Hence while searching template, first 21 amino acid residues were removed. A template selection search was performed using BLAST-P **[11]** against PDB **[12]** database from NCBI interface simultaneously "Template Identification Tool" at Swiss-Model interface **[13]** provided by Swiss Institute of Bioinformatics was utilized for template selection. In results, 12 significant hits with E-value zero were observed, the best of these comprise of model 3FPX-A and 3DIV-A from species *Trametes hirsute* and *Cerrena maxima* sharing 84 % and 82% sequence identity with query sequence. The remaining three, of which two models 2QT6-B and 2VDZ-A are Laccase from *Lentinus tigrinus* and *Coriolopsis gallica* respectively showing 77 % sequence identity with query sequence, while template 1GYC-A from species *Tremetes versicolor* seems to have 78 % sequence identical to that of laccase from Pycnoporus cinnabarinus. Similar results were obtained from NCBI BLAST-P server.

Sequence Analysis:
Elementary domain analysis carried out on "InterPro Domain Scan"**[14]** revealed presence of one Multicopper oxidase domains of type I, II and III each from residue number 163 – 305, 365 – 492, 30 – 152 respectively, in addition one copper binding site is predicted around residue number 125 to 145. Phylogenetic analysis was performed using traditional approach i.e. Multiple sequence alignment (MSA) was constructed using Clustal-X

Bioinformation Volume 5

www.bioinformation.net Issue 4

open access

Hypothesis

version 1.81 **[15]** followed by tree reconstruction using Phylip Package **[16]**. 1000 datasets of query sequence were generated by resampling the original MSA using Seqboot followed by estimation of unrooted phylogeny for each dataset by using Protpars. Subsequently using family of consensus tree methods called the Ml methods implemented in Consense program, a consensus tree was drawn following majority rule consensus. Laccase from species *Coriolus zonatus* was used as an outgroup in tree reconstruction process. Finally, a rooted tree without using branch length was generated (see **Figure 1**) using Drawgram program. **Figure 1** clearly illustrates that Templates 3FPX and 3DIV are closest ancestor of query sequence and thus were used as templates for further modeling process. One interesting fact to note is that template 1GYC, which has higher sequence identity (78 %) to query, is distantly related as compared to templates 2VDZ and 2QT6 which shares less sequence identity (77 %).

Modeling of the structure:

The Align2D command from Modeller was utilized to align query sequence to template structures, followed by generation of models by Loop-Model class. Five models were allowed to be build from each template and single loopmodel was generated after refinement process. Thus five models from each template were generated. At the same time SWISS-MODEL server was also used to generate models using same template. Energy minimization was carried out by first implementing the Conjugate Gradient and subsequently Molecular Dynamics(MD) approach; Conjugate Gradient was allowed to perform 200 optimizing iterations using a modified version of the Beale restart conjugate gradients method, while Molecular Dynamics optimization was performed at 300 Kelvin for 500 MD steps using the leap-frog Verlet integrator.

Model Evaluation:

Models so produced were ranked on QMEAN server **[17]**. All the models created by means of Modeller using template 3FPX had Qmean score above 0.83 indicating good quality of each model produced, while only one model (Model 2.pdb) generated using template 3DIV had Qmean score above 0.83 (See **Table 1 in supplementary material**); also Models generated by using Swiss model Automated homology modeling server had score 0.85 and 0.84 for template 3FPX and 3DIV respectively. Models deduced by template 3DIV showed poor QMEAN score despite of the method used (Modeller or SwissModel server). Models were evaluated on basis of geometrical and stereochemical constraints using ProCheck **[18]** and factors like unfavorable atomic contacts, side chain planarity problems, connections to aromatic rings out of plane etc were assessed using WhatCheck (WhatIf) **[19]** utilities available at SIB server. Finally the models were visualized using Chimera software **[20]**.

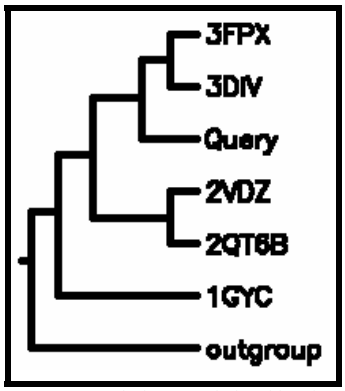

Figure 1: Evolutionary relationship of query sequence with templates

```
 _aln.pos         10        20        30        40        50        60
3FPXA     AVGPVADLTITDAAVSPDGFSRQAVVVNGVTPGPLVAGNIGDRFQLNVIDNLTNHTMLKSTSIHWHGF
query     AIGPVADLTLTNAAVSPDGFSREAVVVNGITPAPLIAGQKGDRFQLNVIDNLTNHTMLKTTSIHWHGF
 _consrvd * ******* * ********** ****** ** ** **  ****************** ********

 _aln.p   70        80        90        100       110       120       130
3FPXA     FQHGTNWADGPAFINQCPISPGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGPFVVYDPNDPHASR
query     FQHGTNWADGVSFVNQCPIASGHSFLYDFQVPDQAGTFWYHSHLSTQYCDGLRGPFVVYDPNDPQASL
 _consrvd **********  * *****  ****************************************** **

 _aln.pos  140       150       160       170       180       190       200
3FPXA     YDVDNDDTVITLADWYHTAAKLGPRFPGGADATLINGKGRAPSDSVAELSVIKVTKGKRYRFRLVSLS
query     YDIDNDDTVITLADWYHVAAKLGPRFPLGADATLINGLGRSPGTTTADLAVIKVTQGKRYRFRLVSLS
 _consrvd ** ************** ********* ********* ** *    * * ***** ************

 _aln.pos     210       220       230       240       250       260       270
3FPXA     CNPNHTFSIDGHNLTIIEVDSVNSQPLEVDSIQIFAAQRYSFVLDANQAVDNYWIRANPNFGNVGFDG
query     CDPNHTFSIDGHTMTVIEADSVNTQPLEVDSIQIFAAQRYSFVLDASQPVDNYWIRANPAFGNVGFAG
 _consrvd * **********  * ** **** ********************** * ********** ****** *

 _aln.pos      280       290       300       310       320       330       340
3FPXA     GINSAILRYDGAPAVEPTTNQTTSVKPLNEVDLHPLVSTPVPGSPSSGGVDKAINMAFNFNGSNFFIN
query     GINSAILRYDGAPEVEPTTTQTTSTKPLNEADLHPLTPMPVPGRPEAGGVDKPLNMVFNFNGTNFFIN
 _consrvd ************* ***** **** ***** *****    **** *  *****  ** ***** *****

 _aln.pos        350       360       370       380       390       400
3FPXA     GASFVPPTVPVLLQILSGAQTAQDLLPSGSVYVLPSNASIEISFPATAAAPGAPHPFHLHGHTFAVVR
query     NHSFVPPSVPVLLQILSGAQAAQDLVPDGSVYVLPSNSSIEISFPATANAPGTPHPFHLHGHTFAVVR
 _consrvd   ***** ************ **** * ********* ********** *** ***************

 _aln.p  410       420       430       440       450       460       470
3FPXA     SAGSTVYNYDNPIFRDVVSTGTPAAGDNVTIRFDTNNPGPWFLHCHIDFHLEGGFAVVMAEDTPDVKA
query     SAGSSEYNYDNPIFRDVVSTGQP--GDNVTIRFQTNNPGPWFLHCHIDFHLEAGFAVVLAEDTPDTAA
 _consrvd ****  **************** *  ******** ****************** ***** ******  *

 _aln.pos  480       490
3FPXA     VNPVPQAWSDLCPTYDALDPNDQ
query     VNPVPQSWSDLCPIYDALDPSDL
 _consrvd ****** ****** ****** *
```

Figure 2: Alignment of query sequence with 3FPX template

Bioinformation Volume 5

www.bioinformation.net Issue 4

open access

Hypothesis

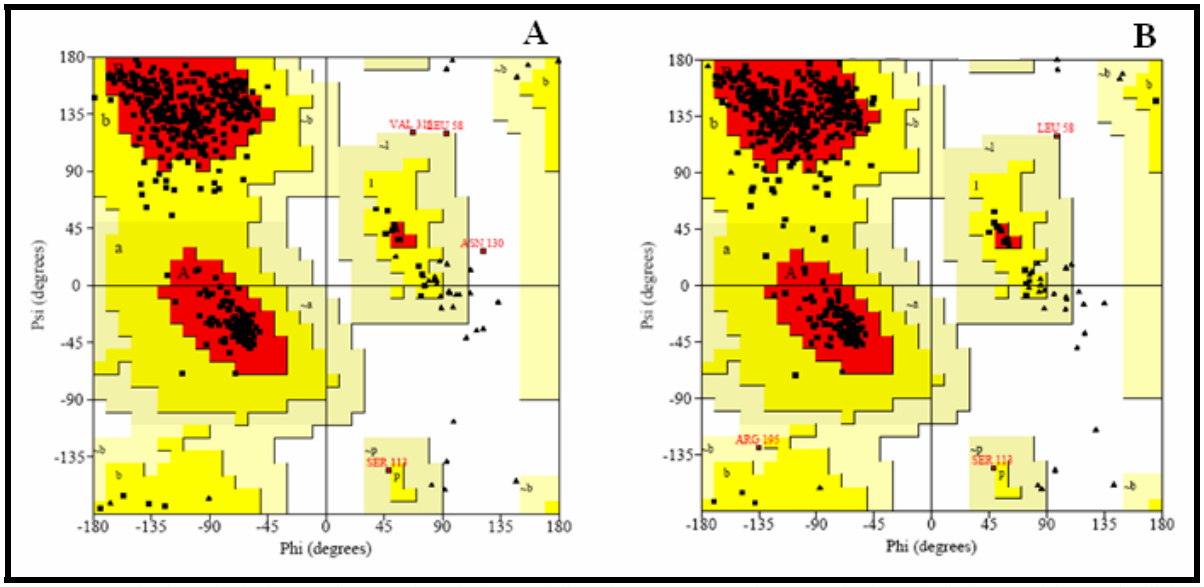

Figure 3 A Procheck results for best model created employing Modeller using 3FPX as template **B** Procheck results for best model created utilizing Modeller using 3DIV as template

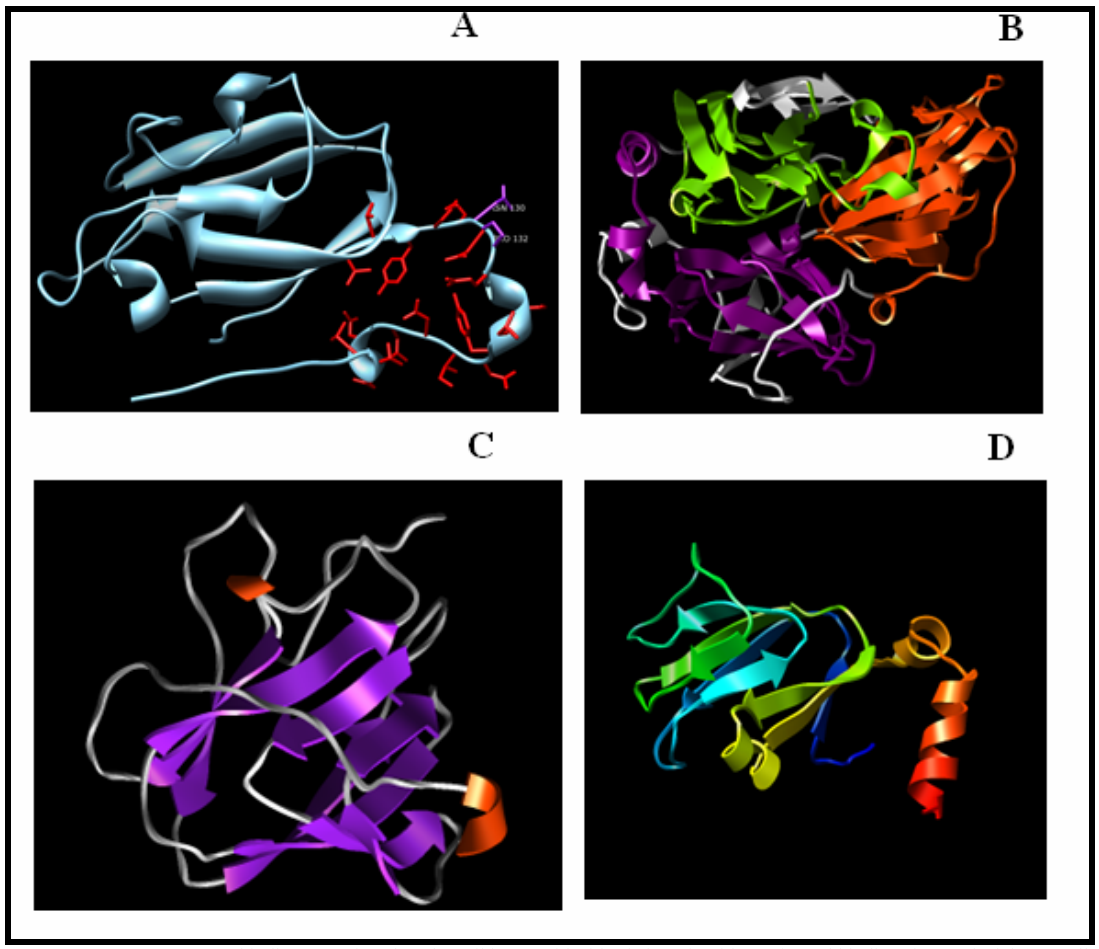

Figure 4: A Type III copper binding Domain (Sky Blue Ribbons) with side chains of residues of metal binding site (Red sticks); B All three domains with green coloured (Type III), orange coloured (Type I), dark magenta coloured (Type II); C Type I copper binding domain with β-sheets in Magenta, Helices in Orange and Loops in Grey; D Type II copper binding domain with rainbow colouring scheme.

Bioinformation **Volume 5**

open access

www.bioinformation.net Issue 4 **Hypothesis**

Discussion:

It is evident from **Table 2 (see supplementary material)** and **Figure 3A** that best model created using template 3FPX employing Modeller program (Model1.pdb) has more than 91 % of its Non Proline and Non Glycine amino acid residues in core region, 7.7 % in allowed region, 0.5 % in generously allowed regions and only 0.5 % in disallowed region as compared to models created using SWISSMODEL (Model 3FPX swissmodel.pdb) have only 82 % of its Non Proline and Non Glycine amino acid residues in core region, 16.4 % in allowed region, 0.5 % in generously allowed regions and none of its amino acid residue lie in disallowed region, in other words approximately only 8 % of residues created using template 3FPX employing Modeller lie outside Core region as compared to 17.2 % of residues created using template 3FPX employing SWISSMODEL remain outside Core region indicating that models created using Modeller are better in terms of geometrical and stereo chemical properties; Similarly in case of best model created using template 3DIV by means of Modeller(Model1.pdb) has around 89 % of its Non Proline and Non Glycine amino acid residues in core region, 9.6 % in allowed region, 0.7 % in generously allowed regions and none of its amino acid residue lie in disallowed region as compared to models created using SWISSMODEL(Model 3DIV swissmodel.pdb) that have 83% of its Non Proline and Non Glycine amino acid residues in core region, 15.4% in allowed region, 0.5 % in generously allowed regions and 0.7% in disallowed, in other words approximately only 10.3 % of residues created using template 3FPX employing Modeller lie outside Core region as compared to 16.8 % of residues created using template 3FPX employing SWISSMODEL region again demonstrating that models created using Modeller are better in terms of geometrical and stereo chemical properties (**Table 3 see supplementary material** and **Figure 3B**).

The WhatIf report reveals 5 errors regarding Side chain planarity problems in models created using 3FPX as template utilizing SWISSMODEL in contrast only two out of five models created using Modeller showed that error containing single residue possessing side chain planarity problem; none of model possessed error of Connections to aromatic rings out of plane in case of MODELLER while 6 same errors were identified in model created at SWISSMODEL server.

In case of models created using 3DIV as template utilizing Modeller reported single side chain planarity problem error in only one model out of five, in contrast 18 side chain planarity problem errors were observed in model created using SWISSMODEL. None of aromatic rings had connection out of plane in any model created using MODELLER while 8 residues had aromatic rings that had connection out of plane in model generated at SWISSMODEL server.

Domain Analysis:

As predicted in sequence, all the three domains are detected in structure. The type III copper binding domain containing Plastocyanin fold is shown in Ribbon representation in **Figure 4A** with copper binding site in Sticks representation among ASN 130 and PRO 132, the predicted copper binding residues in Magenta. Type I and Type II domains are depicted in **Figure 4C** and **Figure 4D** respectively, while **Figure 4B** represents the entire tertiary structure of Laccase with all the three domains.

Conclusion:

Finally to conclude, Model 1 constructed using template 3FPX employing MODELLER performs better in ProCheck and WhatIf structure validation test as compared to models made at SWISSMODEL server. The same model also outperforms the models created using 3DIV as template. Furthermore we can utilize the resulting models of this work in Molecular Docking studies to gain more insight of its interaction with its mediators like Vanillin and ABTS that are known to enhance its activity

References:

[1] B Neppolian *et al. J. Environ. Sci. health* **34:** 1829 (1999)
[2] CM Carliell *et al. Water* **22**:225 (1996)
[3] S Kim *et al. J. Biosci. Bioeng* **95**: 102 (2003) [PMID: 16233374]
[4] DA Wood., *J. Gen. Microbiol.* **117:** 327 (1980)
[5] PJ Kersten *et al. Biochem J.* **268**: 475 (1990) [PMID: 2163614]
[6] MS Johnson *et al. Crit Rev Biochem Mol Biol.* **29:** 1 (1994) [PMID: 8143488]
[7] A Sali *et al. J. Mol. Biol.* **234**: 779 (1993) [PMID: 8254673]
[8] B Wallner *et al. Protein Sci.* **14**: 1315 (2005) [PMID: 15840834]
[9] B Boeckmann *et al. Nucleic Acids Res* **31**:365 (2003) [PMID: 12520024]
[10] C Eggert *et al. Appl. Environ. Microbiol.* **62**:1151 (1996) [PMID: 8919775]
[11] F Stephen *et al. Nucleic Acids Res.* **25**:3389 (1997) [PMID: 9254694]
[12] J Westbrook *et al. Nucleic Acids Res.* **31**: 489 (2003) [PMID: 12520059]
[13] T Schwede *et al. Nucleic Acids Res.* **31**:3381 (2003) [PMID: 12824332]
[14] EM Zdobnov *et al. Bioinformatics.* **17**: 847 (2001) [PMID: 11590104]
[15] F Jeanmougin *et al. Trends Biochem Sci.* **23**:403 (1998) [PMID: 9810230]
[16] J Felsenstein, *Cladistics* **5**: 164 (1989)
[17] P Benkert *et al., Nucleic Acids Res.* **1**: 37 (2009) [PMID: 19429685]
[18] RA Laskowski *et al. J. Appl. Cryst.* **26**: 283 (1993)
[19] RW Hooft *et al. Nature* **381**: 272 (1996) [PMID: 8692262]
[20] EF Pettersen *et al. J. Comput. Chem.* **25**:1605 (2004) [PMID: 15264254]

Edited by P. Kangueane

Citation: Meshram *et al.* Bioinformation 5(4): 150-154 (2010)

Bioinformation Volume 5

open access

www.bioinformation.net Issue 4 **Hypothesis**

Supplementary material:

Table 1: Comparison of quality of models produced using Modeller and Swissmodel server.

Method	Templates	Models	QMEAN Score (0-1)
Modeller	3FPX	Model 1.pdb	0.864
		Model 2.pdb	0.852
		Model 3.pdb	0.847
		Model 4.pdb	0.867
		Model 5.pdb	0.839
	3DIV	Model 1.pdb	0.808
		Model 2.pdb	0.837
		Model 3.pdb	0.796
		Model 4.pdb	0.824
		Model 5.pdb	0.801
Swissmodel	3FPX	Model 3FPX swissmodel	0.85
	3DIV	Model 3DIV swissmodel	0.84

Table 2: Procheck results of Non Proline and Non Glycine residues for models produced using 3FPX as template

Model Name	Residues in Core region		Residues in additional allowed regions		Residues in generously allowed regions		Residues in disallowed regions	
	%	Number	%	Number	%	Number	%	Number
Model 1.pdb	91.3	379	7.7	32	0.5	2	0.5	2
Model 2.pdb	90.1	374	9.4	39	0.2	1	0.2	1
Model 3.pdb	90.4	375	9.2	38	0.2	1	0.2	1
Model 4.pdb	90.1	374	8.9	37	0.2	1	0.7	3
Model 5.pdb	89.2	370	9.9	41	0.5	2	0.5	2
Model 3FPX swissmodel	82.9	344	16.4	68	0.5	2	0.2	1

Table 3: Procheck results of Non Proline and Non Glycine residues for models produced using 3DIV as template

Model Name	Residues in core region		Residues in additional allowed regions		Residues in generously allowed regions		Residues in disallowed regions	
	%	Number	%	Number	%	Number	%	Number
Model 1.pdb	89.6	372	9.6	40	0.7	3	0	0
Model 2.pdb	88.9	369	9.9	41	0.7	3	0.5	2
Model 3.pdb	88.4	367	10.1	42	1.2	5	0.2	1
Model 4.pdb	88.2	366	10.6	44	1.0	4	0.2	1
Model 5.pdb	88.2	366	10.8	45	0.5	2	0.5	2
Model 3DIV swissmodel	83.1	345	15.4	64	0.7	3	0.7	3