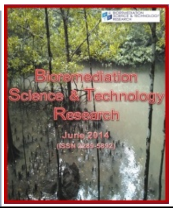
HIBISCUS

BIOREMEDIATION SCIENCE & TECHNOLOGY RESEARCH

Website: http://journal.hibiscuspublisher.com



3D Protein Structure Prediction of *Rhodococcus* UKMP-5M Phenol Hydroxylase Using Homology Modelling

Nor Suhaila, Y[1,2]., Hasdianty,A[3]., Rosfarizan, M[2,4]., Abdul Latif, I[1]. and Arbakariya, A[2,4].

[1]Institute of Bio-IT Selangor, Universiti Selangor, Jalan Zirkon A7/A, Seksyen 7, 40000 Shah Alam, Selangor
[2]Institute of Bioscience, Universiti Putra Malaysia 43400 Serdang, Selangor, Malaysia
[3]Faculty of Science and Biotechnology, Universiti Selangor, Jalan Timur Tambahan, 45600 Bestari Jaya, Selangor
[4]Department of Bioprocess Technology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Corresponding author: Nor Suhaila Yaacob; Tel no: +603-55223428; Fax no: +603-55139143; E-mail: sue_upm76@hotmail.com

History

Received: 21 January 2014
Received in revised form: 18 February2014
Accepted: 8 March 2014
Available online: 1 April 2014

Keywords

Rhodococcus UKMP-5M, phenol hydroxylase; Blastp; 3D structure; ProQ.

Abstract

The Swiss-prot data-base was used for the protein sequence of *Rhodococcus* UKMP-5M. BlastP was used to determine the suitable template for homology modelling. Swiss Model is the homology modelling software was used to determine the 3D structure which passed the ProQ quality test for further analysis. Validation result for the predicted structure of *Rhodococcus* UKMP-5M, in which the prediction structure has passed the validation test with 5.951 Lgscore. This is lies in the range of extremely good model and 0.514 MaxSub which is lies in the range of very good model.

INTRODUCTION

Phenol has been characterized as first priority pollutants by the European Union (EU) and the United State Environmental Protection Agency (US EPA) [1-3]. Ingestion of 1 g phenol contributes to human fatal and phenol concentration of higher than 50 part per billion (ppb) is toxic to some aquatic life [4]. The World Health Organization (WHO) recommends that the maximum permissible concentration of phenol in drinking water is 1 µg/L [5] as the concentration above this level may cause systemic poisoning [6].Phenols hydroxylase enzyme is the proteins that play a role in the degradation of phenol.

Proteins are biochemical compounds consisting of one or more polypeptides typically folded into a globular or fibrous form, facilitating a biological function. The function of protein is determined by its structure [7]. The 3D protein structure can be predicted through three main categories of methods: (1) homology or comparative modeling; (2) fold-recognition (predicting the global-fold of a protein); (3) *ab initio* (trying to model the 3D structure of proteins using only the sequence or the force field) [8].

Protein structure prediction using bioinformatics approach involved sequence similarity searches, multiple sequence alignments, identification and characterization of domains, secondary structure prediction and model validation [9] using PDB (Protein Data Bank) which is the only protein structure database [10]. Computational methods or algorithms including homology modelling which is based on alignment to know 3 dimensional protein structure which is derived experimentally (template) is most accurate method to replace X-ray crystallography or nuclear magnetic resonance (NMR) technique [10, 11]. The three critical steps involve in Critical Assessment of Techniques for Proteins Structural Prediction (CASP) are: (1) template selection; (2) target-template alignment step; (3) modeling of regions not present or significantly different from those in template and (4) modeling of side chains [12]. If the similarity of the sequence is more than 30%, it can be used as template. ESyPred3D and Swiss model is an automated homology modeling program. The method gets benefit of the increased in alignment performances of an alignment strategy. Alignments are obtained by combining, weighting and screening the results of several multiple alignment programs and the final 3D structure can be built using the modeling package MODELLER. EsyPred3D and Swiss model passed the targets in CASP4 and has been proved as the accurate method [10, 11]. The blast suite of programs is used to find local sequence similarities, which may lead to the clues about the structure and function of the query sequence [13]. Between 30% and 50% of identity, multiple alignments between target, template and similar proteins must be used. Pairwise alignments between target and template must be extracted from this multiple alignment [11].

ProQ is a neural network-based method to predict the quality of a protein model that extract structural features, such as frequency of atom-atom contacts and predicts the quality of a model which is measured either by LGscore [14] or MaxSub [15]. The difference between MaxSub and LGscore is the length of the target protein. Correct model should have LGscore more than 1.5 and MaxSub more than 0.1 [16]. ProFunc identifies the functions of a protein related to its 3D structure analyzing the protein's sequence and structure by identifying functional motifs [17]. Sequence motif is a nucleotide or amino-acid sequence pattern that widespread and has, or is assumption to have a biological significance (http://en.wikipedia.org/wiki/Sequence_motif). Motif is also a segment of DNA or structure that codes for functional unit of a protein.

MATERIALS AND METHODOLOGY

Microorganism

The bacterium, *Rhodococcus*UKMP-5M, was usedthroughout this study. This bacterium was isolated from apetroleum contaminated soil at an oil refinery in Malacca, Malaysia.

Determination of Amino Acid Sequence of Phenol Hydroxylase Gene

The DNA sequence of the phenol hydroxylase gene was obtained from the genome scaffold data of *Rhodococcus* UKMP-5M [18]. The sequence was then blast using NCBI, BLAST program against protein database using Blastx algorithm. The most similar protein sequence with 100% identity to the query sequence was retrieved in FASTA format.

Preparation of the Structure Templates

The sequence was again Blast against Protein Databank using Blastp algorithm to find the template for homology modeling. The sequence with similarity greater than 30% with known structure was selected as template.

Prediction of the 3D Structure using Swiss Model

The selected template was sent to Swiss model server [13, 19] to predict the 3D structure. The Swiss model server predicted the 3D structure model and the pdb file of the predicted 3D structure were then sent to the author's email.

Structure Validation

The structure predicted by the Swiss model server was then sent to ProQ-Quality Predictor [14] to check the quality of the model.

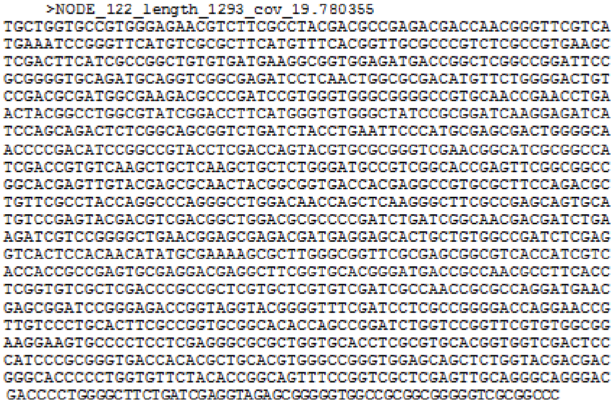
Function Predicting from Structure

The pdb file was loaded in the ProFunc online tool. Preliminary information was then entered before the result was sent to the email provided [17].

RESULTS AND DISCUSSION

The Amino Acid Sequence of Phenol Hydroxylase Gene

The phenol hydroxylase gene sequence (Figure 1) was obtained from the *Rhodococcus* UKMP-5M genome scaffold [18]. Sequence comparison is a major step in the prediction of protein structure from existing templates in the Protein Data Bank [20]. Figure 2 shows the results of Blastx analysis, which was performed against a protein database to find the most similar protein with the query sequence. The analysis has successfully generated informative result. The sequence with the highest score YP 119267.1 with 97% identity to the query is selected as the input sequence for BlastP analysis to search for the protein homolog (Figure 3).



```
    >NODE_122_length_1293_cov_19.780355
TGCTGGTGCCGTGGGAGAACGTCTTCGCCTACGACGCCGAGACGACCAACGGGTTCGTCA
TGAAATCCGGGTTCATGTCGCGCTTCATGTTTCACGGTTGCGCCCGTCTCGCCGTGAAGC
TCGACTTCATCGCCGGCTGTGTGATGAAGGCGGTGGAGATGACCGGCTCGGCCGGATTCC
GCGGGGTGCAGATGCAGGTCGGCGAGATCCTCAACTGGCGCGACATGTTCTGGGGACTGT
CCGACGCGATGGCGAAGACGCCCGATCCGTGGGTGGGCGGGGCCGTGCAACCGAACCTGA
ACTACGGCCTGGCGTATCGGACCTTCATGGGTGTGGGCTATCCGCGGATCAAGGAGATCA
TCCAGCAGACTCTCGGCAGCGGTCTGATCTACCTGAATTCCCATGCGAGCGACTGGGGCA
ACCCCGACATCCGGCCGTACCTCGACCAGTACGTGCGCGGGTCGAACGGCATCGCGGCCA
TCGACCGTGTCAAGCTGCTCAAGCTGCTCTGGGATGCCGTCGGCACCGAGTTCGGCGGCC
GGCACGAGTTGTACGAGCGCAACTACGGCGGTGACCACGAGGCCGTGCGCTTCCAGACGC
TGTTCGCCTACCAGGCCCAGGGCCTGGACAACCAGCTCAAGGGCTTCGCCGAGCAGTGCA
TGTCCGAGTACGACGTCGACGGCTGGACGCGCCCCGATCTGATCGGCAACGACGATCTGA
AGATCGTCCGGGGCTGAACGGAGCGAGACGATGAGGAGCACTGCTGTGGCCGATCTCGAG
GTCACTCCACAACATATGCGAAAAGCGCTTGGGCGGTTCGCGAGCGGCGTCACCATCGTC
ACCACCGCCGAGTGCGAGGACGAGGCTTCGGTGCACGGGATGACCGCCAACGCCTTCACC
TCGGTGTCGCTCGACCCGCCGCTCGTGCTCGTGTCGATCGCCAACCGCGCCAGGATGAAC
GAGCGGATCCGGGAGACCGGTAGGTACGGGGTTTCGATCCTCGCCGGGGACCAGGAACCG
TTGTCCCTGCACTTCGCCGGTGCGGCACACCAGCCGGATCTGGTCCGGTTCGTGTGGCGG
AAGGAAGTGCCCCTCCTCGAGGGCGCGCTGGTGCACCTCGCGTGCACGGTGGTCGACTCC
CATCCCGCGGGTGACCACACGCTGCACGTGGGCCGGGTGGAGCAGCTCTGGTACGACGAC
GGGCACCCCCTGGTGTTCTACACCGGCAGTTTCCGGTCGCTCGAGTTGCAGGGCAGGGAC
 GACCCCTGGGGCTTCTGATCGAGGTAGAGCGGGGGTGGCCGCGGCGGGGGTCGCGGCCC
```

Figure 1. Phenol hydroxylase gene sequence from *Rhodococcus* UKMP-5M genome scaffold [18]

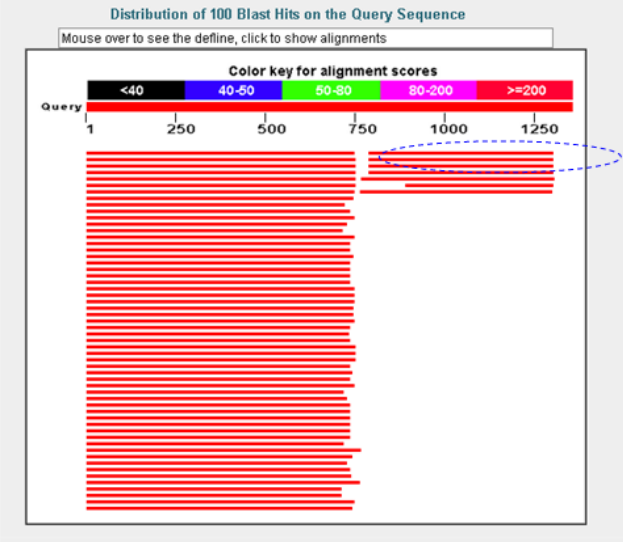
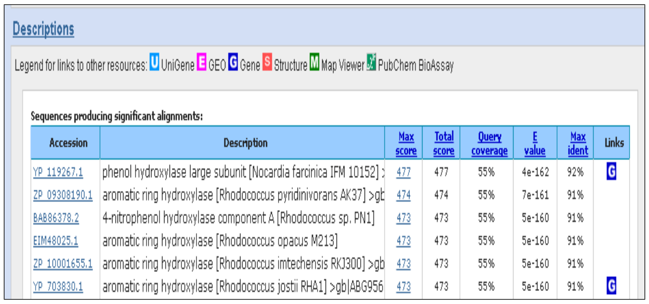


Figure 2. The blastx graphic display demonstrate high score of hits. The circled area showed the region of contig which has high identity with phenol hydroxylase in protein database.



Descriptions

Legend for links to other resources: U UniGene E GEO G Gene S Structure M Map Viewer PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
YP 119267.1	phenol hydroxylase large subunit [Nocardia farcinica IFM 10152] :	477	477	55%	4e-162	92%	G
ZP 09308190.1	aromatic ring hydroxylase [Rhodococcus pyridinivorans AK37] >gb	474	474	55%	7e-161	91%	
BAB86378.2	4-nitrophenol hydroxylase component A [Rhodococcus sp. PN1]	473	473	55%	5e-160	91%	
EIM48025.1	aromatic ring hydroxylase [Rhodococcus opacus M213]	473	473	55%	5e-160	91%	
ZP 10001655.1	aromatic ring hydroxylase [Rhodococcus imtechensis RKJ300] >gb	473	473	55%	5e-160	91%	
YP 703830.1	aromatic ring hydroxylase [Rhodococcus jostii RHA1] >gb|ABG956	473	473	55%	5e-160	91%	G

Figure 3. Blastx descriptions shows the total score of hits sequence. YP_119267.1 demostrates the highest score and identity, with the lowest e-value when aligned with the contig sequence.

Structure Template Preparation

Blastp analysis was used to find the template for homology modelling. From the analysis it can be described that sequence of 2YYG_A (crystal structure of the monooxygenase component (Hpab) of 4-hydroxyphenylacetate 3-Monooxygenase) resulted with 32% of identity and high score of 229 (Figures 4 and 5). 2YYG_A was then selected as the template for the 3D protein prediction.

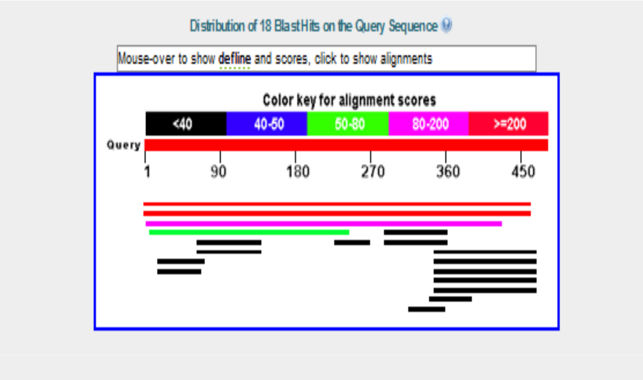
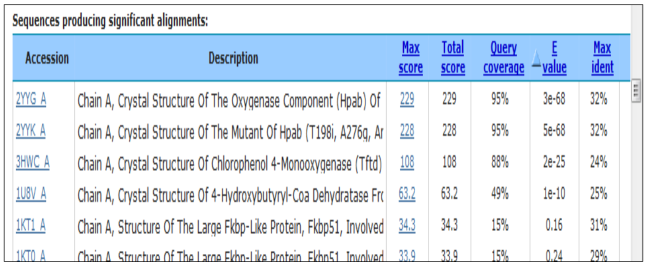


Figure 4. Blastp graphic shows only two hits with high score.



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
2YYG A	Chain A, Crystal Structure Of The Oxygenase Component (Hpab) Of	229	229	95%	3e-68	32%
2YYK A	Chain A, Crystal Structure Of The Mutant Of Hpab (T198i, A276q, Ar	228	228	95%	5e-68	32%
3HWC A	Chain A, Crystal Structure Of Chlorophenol 4-Monooxygenase (Tftd)	108	108	88%	2e-25	24%
1U8V A	Chain A, Crystal Structure Of 4-Hydroxybutyryl-Coa Dehydratase Fr	63.2	63.2	49%	1e-10	25%
1KT1 A	Chain A, Structure Of The Large Fkbp-Like Protein, Fkbp51, Involved	34.3	34.3	15%	0.16	31%
1KT0 A	Chain A, Structure Of The Large Fkbp-Like Protein, Fkbp51, Involved	33.9	33.9	15%	0.24	29%

Figure 5. Blastp descriptions give the total score, max identity and e-value. 2YYG_A is selected as the template sequence as it obtained the highest score with the max identity greater than 30% when aligned with the target sequence.

Prediction and Validation of the 3D Phenol Hydroxylase Structure

The 3D protein structure of the target protein was successfully predicted using Swiss Server. The predicted structure is shown in Figure 6.

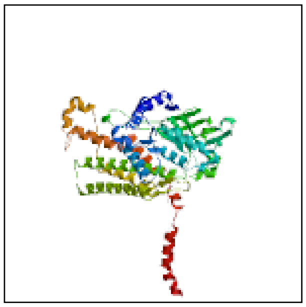
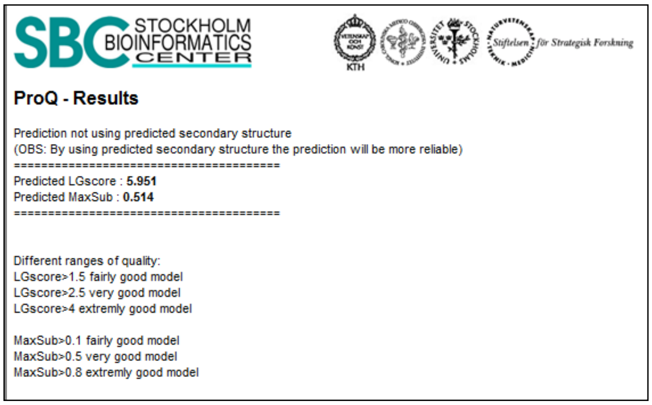


Figure 6. The 3D structure of phenol hydroxylase predicted by swiss model.

ProQ-Protein Quality Predictor was used to validate the 3D protein structure generated by Swiss model (Figure 7), which also presented the ranges of the scores. The predicted LGScore for the model is 5.951 indicating that this is an extremely good model of the predicted phenol hydroxylase structure. Meanwhile, the MaxSub predicted for the model is 0.514. This result shows a highly acceptable score which lies in the range of a good model. Hasdianty *et al*., (2012) reported that LG Score for the prediction of the 3D structure of phenol hydroxylase by *Rhodococcus* NAM81 is 5.936 [21].



ProQ - Results

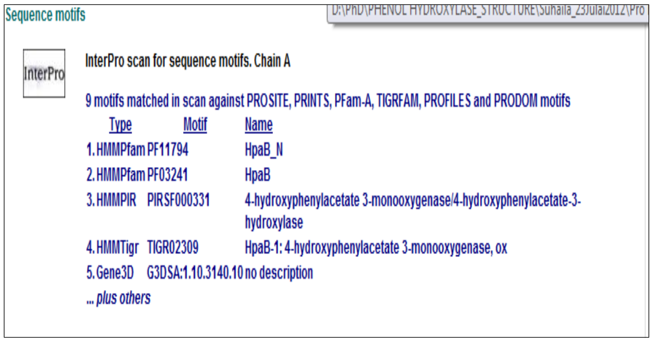
```
Prediction not using predicted secondary structure
(OBS: By using predicted secondary structure the prediction will be more reliable)
========================================
Predicted LGscore : 5.951
Predicted MaxSub : 0.514
========================================


Different ranges of quality:
LGscore>1.5 fairly good model
LGscore>2.5 very good model
LGscore>4 extremly good model

MaxSub>0.1 fairly good model
MaxSub>0.5 very good model
MaxSub>0.8 extremly good model
```

Figure 7. Pro-Q validation result for the predicted structure, in which the prediction structure has passed the validation test with 5.951 Lgscore (lies in the range of extremely good model) and 0.514 MaxSub (lies in the range of very good model).

Prediction of Function from Structure Analysis

Numerous new methods have been designed to predict the protein function from local structure similarities. One of the methods is the search for the derivation of libraries of motifs from known protein structures againts the new structures [22]. Result of sequence motif for SWISS model 3D structure, which showed the presence of phenol hydroxylase enzyme is shown in Figure 8.

Nine (9) motifs matches were observes in phenol hydroxylase structure. Motif is a segment of DNA or structure that codes for functional unit of a protein, which confirmed the function of phenol hydroxylase enzyme through the predicted structure.



Sequence motifs

D:\PhD\PHENOL HYDROXYLASE_STRUCTURE\Suhaila_23Julai2012\Pro

InterPro scan for sequence motifs. Chain A

9 motifs matched in scan against PROSITE, PRINTS, PFam-A, TIGRFAM, PROFILES and PRODOM motifs

Type	Motif	Name
1. HMMPfam	PF11794	HpaB_N
2. HMMPfam	PF03241	HpaB
3. HMMPIR	PIRSF000331	4-hydroxyphenylacetate 3-monooxygenase/4-hydroxyphenylacetate-3-hydroxylase
4. HMMTigr	TIGR02309	HpaB-1: 4-hydroxyphenylacetate 3-monooxygenase, ox
5. Gene3D	G3DSA:1.10.3140.10	no description

... plus others

Figure 8. ProFunc result shows 9 motif matched with other sequences in the databases.

CONCLUSION

The predicted structure of *Rhodococcus* UKMP-5M phenol hydroxylase shows the very good model of the predicted 3D structure through the scaffold gene sequence, where LG Score for the model is 5.951. Nine (9) motifs were also successfully matched with the phenol hydroxyalse enzyme from the predicted structure.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Science, Technology and Innovation (MOSTI), Malaysia under Vot no. 3090104000(G).

REFERENCES

[1] Ye FX, Shen DS. Acclimation of anaerobic sludge degrading chlorophenols and the biodegradation kinetics during acclimationperiod. Chemosphere. 2004; 54:1573–1580.

[2] Ruzgas T, Emneus J, Gorton L, Marko VG. The development of a peroxidase biosensor for monitoring phenol and related aromatic compounds.Anal. Chem. Acta. 1995; 31:245–253.

[3] Santos VLCD, Monteiro ADS, Braga DBT,Santoro MM. Phenol degradation by *Aureobasidium pullulans* FE13 isolated from industrial effluents. J Hazard. Mater. 2009; 161:1413–1420.

[4] Seetharam GB,Saville BA. Degradation of phenol using tyrosinase immobilized on siliceous supports. Water Resour. 2003; 37:436–440.

[5] Kumaran P, Paruchuri YL. Kinetics of Phenol Biotransformation. Water Res. 1997; 31(1):11–22.

[6] Basha KM, Rajendran A, Thangavelu V. Recent advances in the biodegradation of phenol: A review. Asian J. Exp. Biol. Sci. 2010; 1(2):219–234.

[7] Whisstock JC, Lesk AM. Prediction of protein function from sequence and structure. Q. Rev. Biophys. 2003; 36:307–340.

[8] Rost B, O'Donoghue S. Sisyphus and prediction of protein structure. Comput. Appl. Biosci. 1997; 13:345–356.

[9] Edwards YJK, Cottage A. Bioinformatics methods to predict protein structure and function. A practical approach. Mol. Biotechnol. 2003; 23(2):139–166.

[10] Brindha S, Sailo S, Chhakchhuak L, Kalita P, Gurusubramanian G, Kumar JNS. Protein 3D structure determination using homology modeling and structure analysis.Colloquium www.science.vision.in. 2011; 11(3):125–133.

[11] Lambert C, Leonard N, De Bolle X, Depiereux E. ESyPred3D: Prediction of proteins 3D structures. Bioinformatics. 2002; 18(9):1250–1256.

[12] Venclovas C, Zemla A, Fidelis K, Moult J. Some measures of comparative performance in the three CASPs. Protein. 1999; 3:231–237.

[13] Friedrich A, Ripp R, Garnier N, Bettler E, Deléage G, Poch O, Luc Moulinier L. Blast sampling for structural and functional analyses. BMC Bioinformatics. 2007; 8:62.

[14] Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A. A study of quality measures for protein threading models. BMC Bioinformatics. 2001; 2(1):5.

[15] Siew N, Elofsson A, Rychlewski L, Fischer, D. Maxsub: An automated measure to assess the quality of protein structure predictions. Bionformatics. 2000;16: 776–785.

[16] Wallner B, Elofsson A. Can correct protein models be identified? Protein Sci. 2003; 12:1073–1086.

[17] Laskowski RA, James DW, Janet MT. ProFunc: A server for predicting protein function from 3D structure. Nucleic Acid Res. 2005; 33:89–93.

[18] Nor Azlin M, Summers DK, Archer J, Abdul Latif I. Genome sequencing and assembly of *Rhodococcus rubber* UKMP-5M, a thermotolerant metabolically versatile actinomycete. Proceeding of 17th IAS Science Conference. 14-17 December 2009, Selangor Malaysia. 2009; pp 67–69.

[19] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. Bioinformatics. 2006; 22:195–201.

[20] Dunbrack Jr. RL. Sequence comparison and protein structure prediction. Curr. Opinion Struc. Biol. 2006; 16:374–384.

[21] Hasdianty A, Nor Suhaila Y, Mohd Fadzli A, Norazah, MN, Abdul Latif I. Preliminary 3D Protein Structure Prediction of Phenol Hydroxylase from *Rhodococcus* NAM81 Using Homology Modeling. In Proceedings of IAB women in Science International Symposium, 7 May 2012 Concorde Hotel Shah Alam, Selangor, Malaysia. 2012; pp. 66–68.

[22] Jones S, Thornton JM. Searching for functional sites in protein structures. Current Opinion in Chemical Biology. 2004; 8:3–7.