International Journal of Drug Delivery 3 (2011) 386-396

http://www.arjournals.org/index.php/ijdd/index

Original Research Article

ISSN: 0975-0215

The Influence of Sequence Alignment Length of template on the Accuracy and Quality of Homology modelling: Human Cytochrome P450 2D6 (CYP2D6) case study

Elrashid Saleh Mahdi[1], Nurul Bahiyah Ahmad Khairudin[2], Habibah A. Wahab[1,3],

***Corresponding author:**

Habibah A. Wahab
[1]Department of Pharmaceutical Technology, School of Pharmaceutical Sciences Universiti Sains Malaysia
Minden, 11800 Penang, Malaysia.
elrasidm@yahoo.com,
[2]Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.
nurul@fkkksa.utm.my
[3]Division of Hit-to-Lead, Malaysian Institute of Pharmaceuticals and Nutraceuticals, Minden, 11800 Penang, Malaysia.
habibahw@usm.my.

Abstract

Homology modelling is one of the important alternative techniques to X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy in structural determination of protein. CYP2D6 drug metabolizing enzymes have extensively studied prior to the successful determination of its 3D structure using X-ray crystallography. This study is one of the earlier works being carried out before the crystal structure was solved. The enzyme is polymorphic and more than 80 different alleles have been identified. The aim of this study is to identify possible criteria to improve the quality of 2D6 model. Four mammalian crystal structures CYP2C8, CYP2B4, CYP2C9 and CYP2C5 were selected from protein databank as template for CYP2D6 model. Multiple sequence alignment between the selected target and template was performed. Secondary structure prediction was generated using CYP2D6 sequence and the 3D models built based on each template. The quality of the models was evaluated using Ramachandran's plot. The final models were superimposed with CYP2D6 crystal structure. The result shows that not only the sequence identity is important for the template selection but also the alignment length of sequence. The secondary structure prediction of CYP2D6 sequence was found significantly matched the secondary structure prediction of the 3D structures of CYP2D6 models. The stereochemical quality of CYP2D6 models were found adequately satisfied the Ramachandran.plot requirements and comparable to the stereochemical quality crystal structure proteins used as templates and CYP2D6 crystal structure. The criteria set shows that CYP2C8 is the better template for homology modelling of CYP2D6 since it has scored lower E-value and longer alignment length, high sequence identity provided by ClustalW . In addition the secondary structure prediction of the model better matched the consensus and satisfy the criteria of Ramachandran's plot as well it retained lowest RMSD value from the crystal structure. The study concludes to that the length of the sequence alignment is a critical factor in template selection it is also shows the important primary and secondary structure prediction.

Keywords: Homology modelling, CYP2D6, CYP2C8, CYP2C9, CYP2B4, CYP2C5.

Introduction

Homology modelling

Protein function is closely related to its structure. The structural determination of protein is one of the key steps in understanding molecular basis of protein function. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the only two experimental techniques used to determine the protein three dimensional structures, (3D). However, these techniques are not suitable to be applied to some proteins. Most proteins have poor solubility or too large for X-ray and NMR to solve the structure, respectively [1,2]. These two techniques are also laborious and time-consuming. As an alternative, many scientists have relied on computational protein structure prediction methods such as comparative modelling or homology modelling since this method is straightforward, fast and reliable [3]. Homology modelling is indeed a well-known and important technique to predict the 3D structures of proteins and have proven to be successful in the areas of pharmaceuticals as well as drug design and development [4,5]. The accuracy of this technique however, highly depends on the template selection, sequence alignment and the percentage of sequence identity between a target and a template protein [2,6].

Cytochrome P450s

Cytochrome P450s protein was firstly recognised in 1958 by Klingenberg Martin in a rat liver microsome's suspension) [7]. Cytochrome P450s are intracellular genetic protein term for a super-family (CYP's) consisting of heme-containing mono-oxygenases enzymes that catalyses oxidative metabolism of a wide variety of endogenous and exogenous compounds [8,9]. Cytochrome P450 2D6 (CYP2D6) is one of the most important drug metabolizing enzymes in this family. It is responsible for the metabolism of 20-30% of therapeutically used drugs [1,10,11]. The enzyme is highly polymorphic in nature and more than 80 different alleles have been identified [8,10]. Prior to the successful determination of its 3D structure using X-ray crystallography in 2006, [12] a lot of work was carried out in predicting the 3D structure of CYP2D6 using the method of homology modelling [13-15]. And this study was one of the earlier works being carried out before the crystal structure was solved.

The aim of this study was to evaluate the quality of the CYP2D6 models based of mammalian crystal structures with respect to the sequence identity, secondary structure prediction, streochemical quality and RMSD from CYP2D6 crystal in order to investigate possible criteria to improve the quality of the predicted models.

Methods

Primary structure prediction:

The sequence of CYP2D6 was obtained from SwissProt database [16] with the primary accession number of P10635 (497 amino acid residues). All protein sequences related to the CYP2D6 sequence were identified using pairwise sequence alignment PSI-BLAST [17]. From the PSI- BLAST search result, four mammalian crystal structures sequences were selected as the templates. They were human CYP2C9, (PDBID 1OG2) [18], human CYP2C8, (PDBID 1PQ2) [19], rabbit CYP2C5, (PDBID 1DT6) [20] and another rabbit enzyme CYP2B4, (PDBID 1PO5) [21].

Multiple sequence alignment between the target CYP2D6 and all the templates was generated with using CLUSTALW [22] .The CYP2D6 primary sequence was also analysed to predict transemembrane helices regions using Hidden Markov Models (TMHMM) [23] .

Secondary structure prediction

The secondary structures of CYP2D6 was predicted using the programme PROF [24], PSIPRED [25] and FI-Pred [26]. A consensus of predicted secondary structures was manually done based on the results obtained from these programmes.

Homology modelling

Four models of CYP2D6 were built using MODELLER 6 version 2 [27-30] with each model developed based on the templates CYP2C8, CYP2C9, CYP2C5 and CYP2B4. The stereochemistry of the models were checked using the programme PROCHECK [31]. The model that better satisfied the Ramachandran plot was selected as the raw model of CYP2D6 for further evaluation and optimization. The optimization was carried out using the MODELLER's loop refinement protocol in order to optimise the amino acid residues which are located in the disallowed regions of the Ramachandran plot. The secondary structure determination was performed using the programme STRIDE [32] which were then compared to the consensus of the predicted secondary structures. Finally, the CYP2D6 crystal structure (PDBID 2F9Q) [12] was used as a marker to evaluate the accuracy of the refined models based on the C-alpha RMSD.

Results and discussion

Sequence identity search results

From the results of PSI-BLAST, four mammalian proteins were selected as the possible templates for CYP2D6 as shown in Table 1. The results show that CYP2C8 has 362 hit score compared to CYP2B4, CYP2C9 and CYP2C5 which have the scores of 354, 339 and 338, respectively. This score is a scale of conservation obtained from the sequence alignment between the target and the possible template sequences. For each position in the derived scale, each residue assigns a score. For highly conserved residue at a particular position, will give a high positive score while at weakly conserved or unconserved position will give lower scores (near zero) [22].

The results also show that the all the sequence identities between CYP2D6 and all the templates (CYP2B4, CYP2C5, CYP2C8 and CYP2C9) are above 40% (Table1). The enzyme CYP2B4 shared the highest sequence identity (44%) with CYP2D6 sequence but shorter sequence alignment length. The alignment length of CYP2B4 is 440 amino acid residues, 196 of the amino acids are identical to CYP2D6 sequence. The second highest sequence identity is CYP2C5 with 42% sequence identity with alignment length of 469 amino acid residues. 197 of these residues were identical with CYP2D6 sequence. It is also shown that human CYP2C9 shared 40% sequence identity with CYP2D6 sequence with alignment length of 477 amino acid residues and 193 of the total amino acids were identical with CYP2D6 sequence. The last template which is CYP2C8 showed the sequence identity, 41% with alignment length of 205 amino acids. CYP2C8 showed the longest alignment length compared to CYP2B4, CYP2C5 and CYP2C9 sequences despite its lowest percentage of sequence identity. This result indicates that the percentage of sequence identity decreases with the increase in alignment length and was consistent with the findings from previous study [33] .

PSI-BLAST programme compares both the sequence identity and the alignment length using expectation value (E-value) [17]. The smaller the E-value the most identical the sequences are. The E-values obtained were almost zero (significant) as shown in Table 1. The results show high sequence similarity between all the templates and the CYP2D6 enzyme. It is also shown that CYP2C8 has lower E-value followed by CYP2B4, CYP2C9 and CYP2C5. Therefore, all the templates have E values low enough proving that the match did not occur by chance[17]. Thus, it could be concluded that these four templates satisfy the criteria for best template selection [33,34] and according to Sánchez and Šali (1997), accurate model can be built from any of the templates in Table 1.

Table 1 PSI-BLAST search results.

Protein name	Chain length	Alignment length	PSI BLAST score	% sequence identity	E-value
Human CYP2D6 (target)	497	497	969	(497/497) = 100	0.0
Human CYP2C8 CYPIIC8) (template)	490	495	362	(205/495) = 41.0	e^{-99}
Rabbit CYP2B4 (template)	491	440	354	(196/440) = 44.0	$3e^{-97}$
Human CYP2C9 (CYPIIC9) (template)	490	477	339	(193/477) = 40.0	$8e^{-93}$
Rabbit CYP2C5 (template)	487	469	338	(197/469) = 42.0	$2e^{-92}$

Table 2. Multiple sequence alignment scores obtained by CLUSTALW.

Protein name	Chain length	Alignment length	ClustalW score
Human CYP2D6 (target)	497	497	-
Human CYP2C8 (CYPIIC8) (template)	490	476	41
Rabbit CYP2B4	491	476	40
Human CYP2C9 (CYPIIC9)	490	475	39
Rabbit CYP2C5	487	473	39

The primary sequence analysis

The multiple sequence alignment between CYP2D6, CYP2C8, CYP2C9, CYP2C5 and CYP2B4 proteins results have identified regions of similarities, conserved and unconserved regions. The statistical part of ClustalW results [22] was shown in Table 2. This result shows the consequence functional, structural, relationships between the CYP2D6 sequence and the selected template. The results also show that all the template sequences having more than 30% of sequence identity with the target sequence with CYP2C8 shared high sequence identity (41%) followed by CYP2B4 (40%), CYP2C9 and CYP2C5 which have the same sequence identity (39%). The results of sequence alignment differ slightly from PSI-BLAST results.

The sequences of the crystal structures of the templates are edited prior model building. It is found that the first 27 (1-27) amino acid residues of the crystal structure of CYP2C8 are missing. These missing residues correspond to the first 31 (1-31) in the target sequence (CYP2D6). Therefore, the first 31 (1-31) amino acid residues of CYP2D6 sequence are deleted from CYP2D6 sequence and the alignment started from the amino acid residue ARG 32. While the missing residue of CYP2B4 are the first 30 correspond to the first 30 amino acid residues of CYP2D6 sequence which are also deleted and the model built based on CYP2B4 began from the amino acid residue ALA 31. Similarly, CYP2C9 and CYP2C5 lacks the first 29 (1-29) amino acid residues which correspond to the first 33 amino acids residues of CYP2D6 sequence and thus, the model built based on CYP2C5 began from amino acid PRO 34. In addition CYP2C5 crystal structure lacks other 10 residues from amino residue TYR 212 to LEU 222. The published crystal structure of CYP2D6 starts from amino acid residue PRO 34. Hence, the first 33 amino acid residues are absent in the crystal structure templates and experimental crystal CYP2D6. These missing residues are transmembrane segment, as confirmed by the result of the primary sequence analysis obtained from the Hidden Markov Models (TMHMM) [23] which show that the first 25 amino acid residues of CYP2D6 sequence are transmembrane protein (Figure 1) in red colour with a probability of 100%.

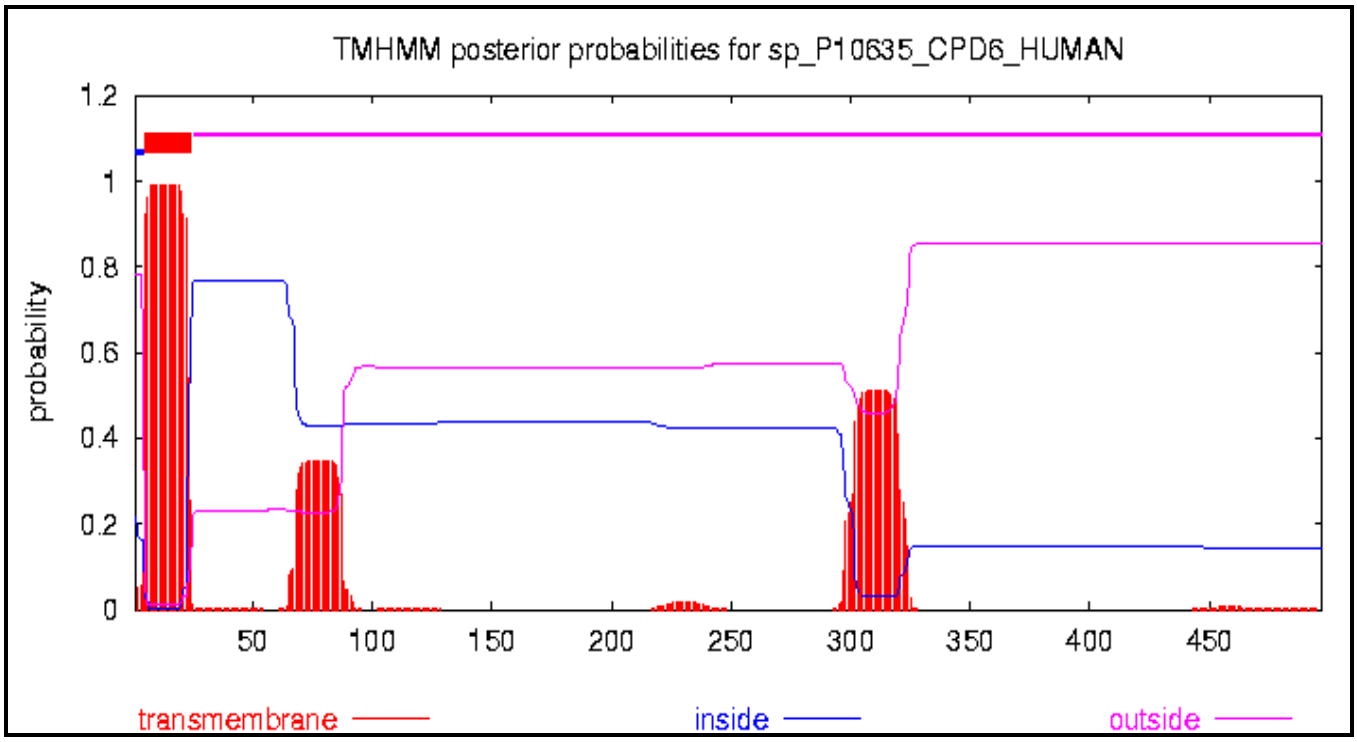


Figure 1: The TMHMM plot shows the predicted Transmembranes segments in CYP2D6 sequence, and the first segment was from 0 -25 amino acid residues Interpreted by red bar above the first highest peak.

Secondary structure prediction

The predicted transemembrane segment is further investigated using secondary structure prediction method. The secondary structures of CYP2D6 predicted by PROF, PsiPred and FI-FRED methods were presented in Table 3. Almost all programmes agreed in the identification of this region as the transmembrane segment. The results of PROF, PsiPred and FI-FRED of secondary structure prediction for the whole CYP2D6 are shown in Table 4.

Table 3. Secondary structure prediction of the first 25 amino acid residues using PROF, PsiPred and FI-Pred schemes.

Amino acid N0.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**1 0**	**11**	**12**	**13**
Amino acid	M	G	L	E	A	L	V	P	L	A	V	I	V
PROF	C	C	C	H	H	H	H	H	H	H	H	H	H
PsiPred	C	C	H	H	H	H	H	H	H	H	H	H	H
FI-Pred	C	C	H	H	H	H	H	H	H	H	H	H	H
Consensus	C	C	H	H	H	H	H	H	H	H	H	H	H
Amino acid S. N0.	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	
Amino acid	A	I	F	L	L	L	V	D	L	M	H	R	
PROF	H	H	H	H	H	H	H	H	H	H	H	C	
PsiPred	H	H	H	H	H	H	H	H	H	H	H	H	
FI-Pred	H	H	H	H	H	H	H	H	H	H	H	H	
Consensus	H	H	H	H	H	H	H	H	H	H	H	H	

Table 4. Summary of secondary structure prediction results of CYP2D6 using PROF, PsiPred and FI-Pred schemes.

Scheme	Coil loops(C)		Helixes(H)		β-sheets(E)	
	Freqequency	% age	Frequency	% age	Frequency.	%age
PROF	238	48.0	240	48.0	19	4.0
PsiPred	187	37.6	273	44.9	37	7.5
FI-Pred	215	43.0	236	48.0	46	9.0
Consensus	214	43.0	243	49.0	40	8.0

The results showed that PROF and FI-Pred almost agreed in predicting the coil regions and helices and mainly form the consensus. It is also noticed that long helices are the most common and composed of 15 -21 amino acid residues. In addition, the loops are short and distributed in fewer regions. The short region of loops is the characteristic of high resolution structure since the amino acids form tight turns (loops) between membrane helices [35,36]. The beta sheet regions in this result are interrupting the short coil regions. The three programmes agreed on the approximate locations, lengths and positions of alpha helices and the loop regions, but slightly disagreed in predicting the positions and lengths of beta sheets.

The secondary structure analysis of 3D models of CYP2D6 using STRIDE [32] is shown in Table 5. From the result of STRIDE, it is noticed that the secondary structures of the models mostly disagreed with the consensus prediction in coil regions. The percentage of the coils form the consensus predictions are higher than the STRIDE prediction. This might be due to the variations in predicting the location of secondary structures using sequence and the 3D structure of the model. This result (Table 5) is represent matched for all the models built based on the four templates shows that no significant variations to matched the consensus. Thus, the four templates are inspected for more investigation to select the best template for CYP2D6.

Table 5 Summary of secondary structure prediction results of 3D structures CYP2D6 models based STRIDE server compared to the consensus prediction based on CYP2D6 to its sequence.

Model	Freq. of (C)	% of (C)	Freq. of (H)	% of (H)	Freq. of (E)	% of (E)	% of consistent
Consensus	214	43.0	243	49.0	40	8.0	100
Model (CYP2C8)	139	29.8	215	46.1	25	5.4	81.3
Model (CYP2B4)	141	30.3	208	44.6	22	4.7	79.6
Model (CYP2C9)	145	31.1	220	47.2	9	1.9	80.2
Model (CYP2C5)	151	32.4	199	42.7	12	2.6	77.7

Homology modelling results:

The good stereochemistry properties is an indication of good comparative modelling method [37]. The overall quality of each model is evaluated in order to identify any unfavourable contacts between the amino acid residues. The results of model evaluations were listed in Table 6. The raw model of CYP2D6 built using the template CYP2C8 shows the highest stereochemical quality. The model shows that 91.4% of its residues are located in the most favoured region or the core region compared to other models built using other templates. [38]. The result of streochemical quality of CYP2D6 model based on CYP2C8 also shows that only one amino acid residue is found located in the disallowed region of the Ramachandran Plot. For the model built based on the template CYP2C9, there are around 89.8% of the residues in the core region and two amino acids residues are found located in the disallowed region. Similarly, for the models built based on CYP2B4 and CYP2C5, 88.9% and 85.5% of their amino acid residues are found located in the core regions, respectively. In the model of CYP2D6 built based on CYP2C5 there are five residues located in the disallowed region. The significantly lower stereochemical quality of CYP2D6 model built based on CYP2C5 might be due to the fact that CYP2C5 lacked ten amino acid residues between TYR 212 to LEU 222. Thus, this might affect the stereochemical quality and structural geometry of the model while doing the optimization step. For model built based on CYP2B4, no amino acid residue are found located inside the disallowed region of the Ramachandran plot, therefore no regional loop optimization were carried to this model.

The results of the loop refinement or optimisation step are shown in Table 6. The resulting models are referred to as the refined models. Subjection of the model built based on CYP2C5 to loop optimisation step did not significantly improved the stereochemical properties of the model as opposed to the models built based on CYP2C8 and CYP2C9. There is no change in the percentage of the amino acid residues in the core regions. The results of loop refinement of CYP2D6 model built based on CYP2C5 only pushed the amino acid residues in the disallowed region to the direction of allowed and generously allowed regions.

Table 6 The detailed scores of assessed stereochemical quality for CYP2D6 models built based on the selected templates using PROCHECK [31].

Template/CYP2D6 model/CYP2D6 crystal structure PDB	Core regions (%)	Allowed regions (%)	Generously allowed regions (%)	Disallowed regions (%)
Human CYP2C8 experimental	85.5	13.1	1.5	0.0
CYP2D6 (raw model)	91.4	7.3	1.0	0.3
CYP2D6 (refined model)	91.6	7.3	1.0	0.0
Human CYP2D6 experimental	66.4	25.3	5.3	2.9
Human CYP2C9 experimental	86.1	11.9	1.5	0.5
CYP2D6 (raw model)	89.8	7.1	2.5	0.5
CYP2D6 (refined model)	89.6	7.6	2.8	0.0
Human CYP2D6 experimental	66.4	25.3	5.3	2.9
Rabbit CYP2B4 experimental	85.5	12.2	2.0	0.2
CYP2D6 (raw model)	88.9	9.1	2.0	0.0
CYP2D6 (refined model)	88.9	9.1	2.0	0.0
Human CYP2D6 experimental	66.4	25.3	5.3	2.9
Rabbit CYP2C5 experimental	71.3	23.4	3.6	1.8
CYP2D6 (raw model)	85.5	9.9	3.3	1.3
CYP2D6 (refined model)	85.5	10.7	3.6	0.3
Human CYP2D6 experimental	66.4	25.3	5.3	2.9

It is interesting that the stereochemical quality of the refined models (Table 6.) are significantly better compared to the stereochemical quality of the crystal structure templates listed in Table 1. It is also found that the results of stereochemical qualities of the refined models are better compared to the CYP2D6 experimental crystal structure as shown in Table 6. Consequently, the quality of the CYP2D6 model built based on human crystal structures CYP2C8 and CYP2C9 are better compared to rabbit crystal structure CYP2B4 and CYP2C5. These results suggested that the human crystal structures are better templates for human CYP2D6 model compared to the rabbit crystal structures. Particularly, CYP2D6 model built based on the human CYP2C8 structure satisfied the stereochemical test better compared to CYP2D6 model based on CYP2C9 as shown by the backbone ϕ and ψ dihedral angles, of which 91.6% of the residues for the refined model located within the core regions, 7.3% with the allowed regions, 1% within generously allowed regions and no amino acid residue in the disallowed regions (Table 6). Despite the variation in the stereochemical quality built based on the selected templates, all models meet the requirements of good quality model based on the Ramachandran.plot [38] and the criteria of selection of good template for CYP2D6. Thus the selected templates are carried for further investigation to select best template for CYP2D6 model.

Superimposed of CYP2D6 models with the CYP2D6 crystal structure

The RMSD obtained from the superimposed of CYP2D6 models to it's solved crystal structure (2F9Q) was shown in Table 7. CYP2D6 model built using template CYP2C8 showed RMSD of 2.96 Å with respect to its crystal structure. Figure 2 showed the schematic superimposed between the CYP2D6 model built based on CYP2C8 template and the crystal structure. It is also shown in Figure 3 that the model developed shared high similarity with that of the crystal structure. This result suggested that the developed model is accurate. Model built based on the template CYP2B4 shows highest RMSD value (6.39 Å) compared to other models. Based on all these findings, human CYP2C8 crystal structure is found to be an adequate template for model development of CYP2D6 since it satisfied all the criteria of best template selection in which it produced the most adequate and accurate model representing CYP2D6 compared to using other templates. It could be concluded that the alignment length played the most important role in accurately predicting the structure of CYP2D6. as consistent with previous work[33]

Table 7 RMSD analysis of CYP2D6 model from the crystal structure based on C alpha.

Model of CYP2D6	RMSD (Å)
Based on human CYP2C8	2.96
Based on human CYP2C9	3.18
Based on rabbit CYP2C5	3.44
Based on rabbit CYP2B4	6.39

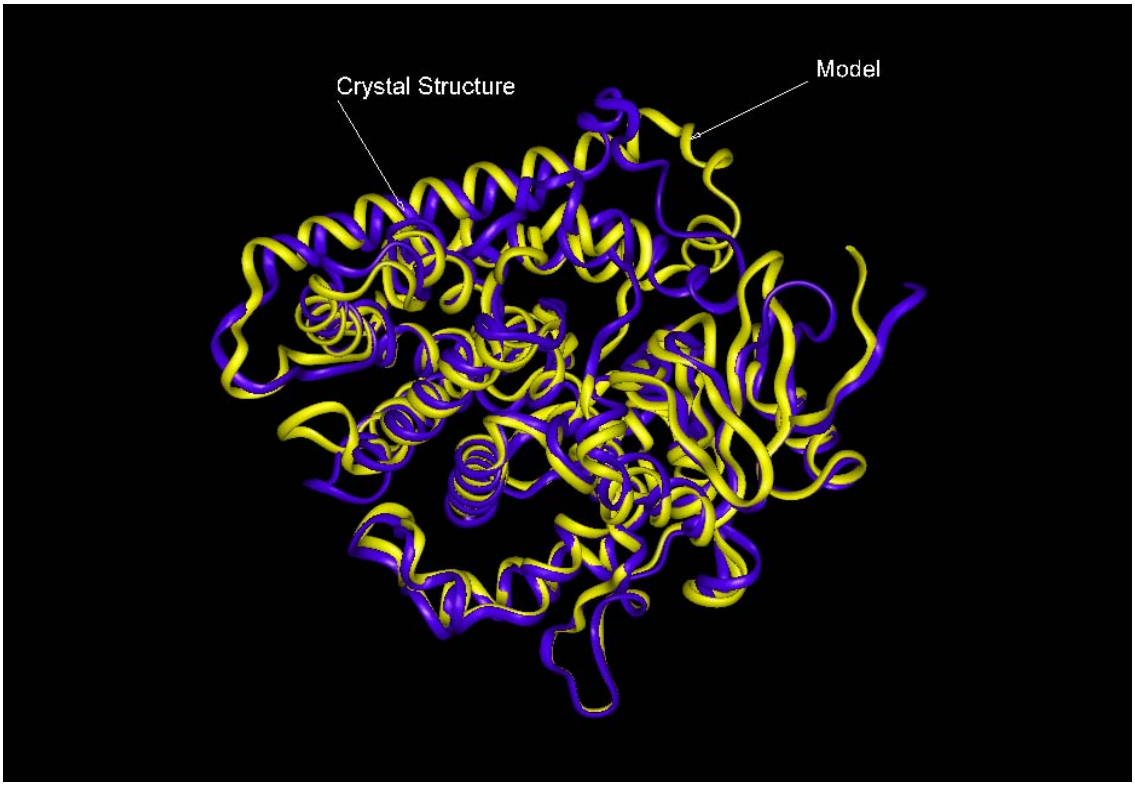
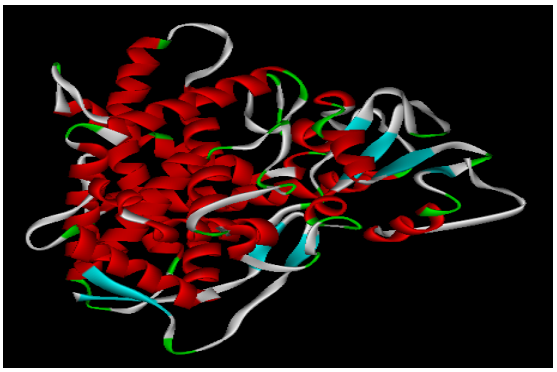
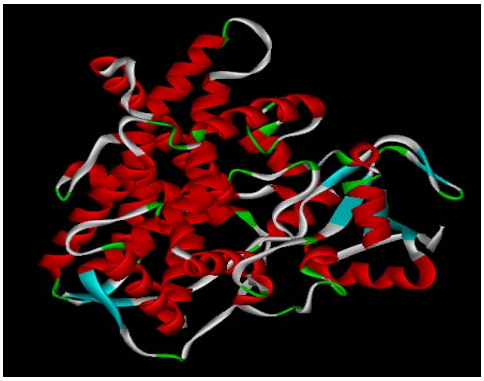


Figure 2 Ribbon representations showing the superimposed of the model of CYP2D6 built based on CYP2C8 (yellow) to the crystal structure of CYP2D6 (blue).



(a)



(b)

Figure 3 (a) Ribbon representations of the crystal structure and Figure 3 (b) the model of CYP2D6 based on CYP2C8.

Conclusion

As a conclusion, this study suggested that in order to accurately predict the 3D structure of proteins, not only the sequence identity and the accuracy of the sequence alignment are important, the length of the sequence alignment is also a critical factor in template selection for homology modelling. This study also showed the importance of the primary and the secondary structure predictions in homology modelling. It is also shown that the mammalian cytochrome P450 sub family templates used in this study are evolutionary and structurally similar to CYP2D6. Particularly, human CYP2C8 and CYP2C9 which are more suitable than the rabbit CYP2C5 and finally human CYP2C8 is found to be the best template for homology modelling of CYP2D6.

Acknowledgement
Financial support provided by Universiti Sains Malaysia for this research and for Elrashid is gratefully acknowledged.

References
1. Kemp C, Maréchal J, Sutcliffe M. Progress in cytochrome P450 active site modeling. Archives of Biochemistry and Biophysics. 2005;433:361-368.
2. Kirton S, Baxter C, Sutcliffe M. Comparative modelling of cytochromes P450. Advanced drug delivery reviews. 2002;54:385-406.
3. Sánchez R. Advances in comparative protein-structure modelling. Current opinion in structural biology. 1997;7:206-214.
4. Kopp J, Schwede T. Automated protein structure homology modeling: a progress report. pgs 2004;5:405-416.
5. Hillisch A, Pineda L, Hilgenfeld R. Utility of homology models in the drug discovery process. Drug discovery today. 2004;9:659-669.
6. Ali A. Comparative protein modeling by satisfaction of spatial restraints. Molecular medicine today. 1995;1:270-277.
7. Hasler J, Estabrook R, Murray M, et al. Human cytochromes P450. Molecular Aspects of Medicine. 1999;20:1-137.
8. Nebert D, Russell D. Clinical importance of the cytochromes P450. The Lancet. 2002;360:1155-1162.
9. Nebert D, Dalton T: The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. Nature Reviews Cancer. 2006;6:947-960.
10. Zanger U, Raimundo S, Eichelbaum M. Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. Naunyn-Schmiedeberg's archives of pharmacology. 2004;369:23-37.
11. Ingelman-Sundberg M. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. The Pharmacogenomics Journal. 2004;5:6-13.
12. Rowland P, Blaney FE, Smyth MG, et al. Crystal Structure of Human Cytochrome P450 2D6. Journal of Biological Chemistry. 2006;281:7614-7622.
13. de Groot M, Ackland M, Horne V, et al. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. J Med Chem. 1999;42:1515-1524.
14. Lewis D. Homology modelling of human CYP2 family enzymes based on the CYP2C5 crystal structure. Xenobiotica. 2002;32:305-323.
15. Ito Y, Kondo H, Goldfarb P, Lewis D. Analysis of CYP2D6 substrate interactions by computational methods. Journal of Molecular Graphics and Modelling. 2008;26:947-956.
16. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database: its relevance to human molecular medical research. Journal of molecular medicine. 1997;75:312-316.
17. Altschul S, Madden T, Schaffer A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997;25:3389.
18. Williams P, Cosme J, Ward A, et al. Crystal structure of human cytochrome P450 2C9 with bound warfarin. Nature 2003;424:464-468.
19. Schoch GA, Yano JK, Wester MR, et al. Structure of Human Microsomal Cytochrome P450 2C8. Journal of Biological Chemistry. 2004;279:9497-9503.
20. Williams P, Cosme J, Sridhar V, et al. Microsomal cytochrome P450 2C5: comparison to microbial P450s and unique features. Journal of Inorganic Biochemistry. 2000;81:183-190.
21. Scott EE, White MA, He YA, et al. Structure of Mammalian Cytochrome P450

2B4 Complexed with 4-(4-Chlorophenyl)imidazole at 1.9-Ã… Resolution. Journal of Biological Chemistry. 2004;279:27294-27301.
22. Thompson JD, Higgins DG, Gibson TJ, Clustal W. improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res. 1994;22:4673-4680.
23. Krogh A, Larsson B, von Heijne G, Sonnhammer E. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology. 2001;305:567-580.
24. Rost B. Review: protein secondary structure prediction continues to rise. Journal of Structural Biology. 2001;134:204-218.
25. Jones D: Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology. 1999;292:195-202.
26. Ceroni A, Frasconi P, Passerini A, Vullo A. A combination of support vector machines and bidirectional recurrent neural networks for protein secondary structure prediction. Lecture notes in computer science. 2003:142-153.
27. Fiser A, Do R, Šali A: Modeling of loops in protein structures. PRS 2000;9:1753-1773.
28. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Methods in enzymology. 2003;374:461-491.
29. John B, Sali A. Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic acids research. 2003;31:3982.
30. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. Journal of molecular biology. 1993;234:779-815.
31. Laskowski R, MacArthur M, Moss D, Thornton J. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography. 1993;26:283-291.
32. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. Proteins-Structure Function and Genetics. 1995;23:566-579.
33. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins: Structure, Function, and Genetics. 1991;9:56-68.
34. Notredame C, Higgins D, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology. 2000;302:205-217.
35. Monné M, Nilsson I, Elofsson A, von Heijne G. Turns in transmembrane helices: determination of the minimal length of a “helical hairpin” and derivation of a fine-grained turn propensity scale. Journal of molecular biology. 1999;293:807-814.
36. Monné M, von Heijne G. Effects of ‘hydrophobic mismatch’on the location of transmembrane helices in the ER membrane. FEBS letters. 2001;496:96-100.
37. Sánchez R, Šali A. Advances in comparative protein-structure modelling. Current opinion in structural biology. 1997;7:206-214.
38. Kleywegt G, Jones T. Phi/psi-chology: Ramachandran revisited. Structure. 1996;4:1395-1400.