Computational Biology and Chemistry 56 (2015) 19–29

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Molecular characterization, modeling and docking of CYP107CB2 from *Bacillus lehensis* G1, an alkaliphile

Swi See Ang [a,b], Abu Bakar Salleh [a,b,c,*], Adam Leow Thean Chor [a,b,d], Yahaya M. Normi [a,d], Bimo Ario Tejo [a,e], Mohd Basyaruddin Abdul Rahman [a,e]

[a] *Enzyme and Microbial Technology Research Center, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[b] *Laboratory of Enzyme Technology, Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[c] *Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[d] *Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[e] *Department of Chemistry, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

ARTICLE INFO

Article history:
Received 4 September 2014
Received in revised form 18 February 2015
Accepted 21 February 2015
Available online 25 February 2015

Keywords:
Cytochrome P450 monooxygenase
1α,25-dihydroxyvitamin D_3
Bacillus lehensis G1
Homology modeling
Molecular docking

ABSTRACT

Cytochrome P450s are a superfamily of heme monooxygenases which catalyze a wide range of biochemical reactions. The reactions involve the introduction of an oxygen atom into an inactivated carbon of a compound which is essential to produce an intermediate of a hydroxylated product. The diversity of chemical reactions catalyzed by cytochrome P450s has led to their increased demand in numerous industrial and biotechnology applications. A recent study showed that a gene sequence encoding a CYP was found in the genome of *Bacillus lehensis* G1, and this gene shared structural similarity with the bacterial vitamin D hydroxylase (Vdh) from *Pseudonocardia autotrophica*. The objectives of present study was to mine, for a novel CYP from a new isolate *B. lehensis* G1 alkaliphile and determine the biological properties and functionalities of CYP in this bacterium. Our study employed the usage of computational methods to search for the novel CYP from CYP structural databases to identify the conserved pattern, functional domain and sequence properties of the uncharacterized CYP from *B. lehensis* G1. A computational homology model of the protein's structure was generated and a docking analysis was performed to provide useful structural knowledge on the enzyme's possible substrate and their interaction. Sequence analysis indicated that the newly identified CYP, termed CYP107CB2, contained the fingerprint heme binding sequence motif FxxGxxxCxG at position 336-345 as well as other highly conserved motifs characteristic of cytochrome P450 proteins. Using docking studies, we identified Ser-79, Leu-81, Val-231, Val-279, Val-383, Ala-232, Thr-236 and Thr-283 as important active site residues capable of stabilizing interactions with several potential substrates, including vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3, in which all substrates docked proximally to the enzyme's heme center. Biochemical analysis indicated that CYP107CB2 is a biologically active protein to produce 1α,25-dihydroxyvitamin D_3 from 1α-hydroxyvitamin D_3. Based on these results, we conclude that the novel CYP107CB2 identified from *B. lehensis* G1 is a putative vitamin D hydroxylase which is possibly capable of catalyzing the bioconversion of parental vitamin D_3 to calcitriol, or related metabolic products.

1. Introduction

Cytochromes P450s are a superfamily of heme monooxygenases which are present in a wide variety of organisms in nature (Bernhardt, 2006; Zurek et al., 2006). They catalyze a broad range of biochemical reactions and play an essential role in the assimilation of carbon source. Prototypical reactions catalyzed by CYPs include fatty acid hydroxylation, secondary metabolite biosynthesis and drug metabolism (Schallmey et al., 2011; Sono et al., 1996).

* Corresponding author at: Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia. Tel.: +60 3 8947 1053; fax: +60 3 8946 7590.
E-mail address: abubakar@upm.edu.my (A.B. Salleh).

http://dx.doi.org/10.1016/j.compbiolchem.2015.02.015
1476-9271/

CYPs catalyze the molecular insertion of oxygen into a wide range of substrates, making them to be promising biocatalysts for the synthetic industry (Grogan, 2011; Urlacher and Eiben, 2006).

The biotransformation of vitamin D_3 to 1α,25-dihydroxyvitamin D_3 is one of the most successful applications of cytochrome P450 in biotechnology. The product has been used to treat numerous diseases such as osteoporosis, chronic renal failure and hypothyroidism (Sakaki, 2012; Yasutake et al., 2009). Vitamin D_3 is a biologically inactive compound that requires one or more CYPs to catalyze the formation of the most active form of vitamin D hormone, 1α,25-dihydroxyvitamin D_3 or calcitriol. The major function of calcitriol is to maintain safe levels of calcium and phosphorus in the blood, by regulating the absorption of these ions in the intestine, bone and kidney. In addition, calcitriol is also recognized as an important anti-proliferative factor for dividing cells and tissues, as vitamin D deficiency is now linked to over 20 forms of cancer (Guyton et al., 2003). The chemical synthesis of 1α,25-dihydroxyvitamin D_3 requires complex procedures, including almost 20 reactions steps with low production yields. Therefore, development of an efficient and simplified production process for calcitriol remains an important area of investigation (Sakaki et al., 2011; Zhu and Okamura, 1995). Only a few studies have reported on the usage of bacterial cytochrome P450s for the production of calcitriol in which vitamin D_3 was converted to 1α,25-dihydroxyvitamin via 25-hydroxyvitamin D_3.

Since, there is an increasing industrial demand to exploit cytochrome P450 as a valuable biocatalyst, considerable attempts have been devoted to search for novel enzymes with unique metabolic properties. This strategy was attempted in this study, where a novel CYP from the alkaliphilic bacteria, *Bacillus lehensis* G1, was mined. This bacterium dwells in soil with the capability to thrive at high pH up to 11. A survey on the complete *B. lehensis* G1 genome revealed a single candidate gene that potentially coded for a functional CYP enzyme.

Sequence analysis showed the protein sequence of the cytochrome P450 from *B. lehensis* G1 exhibited sequence identity of 44% with vitamin D_3 hydroxylase (Vdh) from *Pseudonocardia autotrophica* which had been loosely grouped into the CYP107 family of enzymes (Fujii et al., 2009; Yasutake et al., 2009). However, the lack of structural information for the putative CYP from *B. lehensis* G1 prevented a more detailed characterization of its biological role. Therefore, in this study, the conserved pattern, functional domain and sequence properties of cytochrome P450 from *B. lehensis* G1 were analyzed and subsequently, a structural model of the enzyme was constructed. Docking analysis was performed to provide valuable insight regarding the architecture of its active site and ligand binding interactions. All results pertaining to the possible structure and sequence properties of cytochrome P450 from *B. lehensis* G1, now referred to as CYP107CB2, are duly reported and discussed.

2. Materials and methods

2.1. Molecular analysis of P450

The search for gene sequence encoding cytochrome P450 from the genome of *B. lehensis* G1 was performed through the BLASTN program provided by National Center for Biotechnology Information (NCBI). The conserved domain of cytochrome P450 was determined through conserved domain database (CDD) provided by NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Protein molecular weight and theoretical isoeletric point were computed with ExPasy compute pI/Mw tool (http://web.expasy.org/compute_pi/). Gene ontology annotation was performed by using ScanProSite program (http://www.ebi.ac.uk/Tools/pfa/iprscan/). Protein family and functional domain were determined through Pfam server (http://pfam.sanger.ac.uk/).

2.2. Multiple sequence alignment

In order to investigate the conserved motif of CYP107CB2, three bacterial protein sequences from the CYP107 superfamily of enzymes include Vdh (PDB: 3A4G) from *P. autotrophica*; PikC (PDB: 2BVJ) from *Streptomyces venezuelae*; and EryF (PDB: 1EGY) from *Saccharopolyspora erythraea* were obtained from the Protein Data Bank (PDB) and aligned by using ClustalW (Thompson et al., 1994). Sequence identity was analyzed using PSI-BLAST (Position-Specific Iterated BLAST) against sequences in the PDB.

2.3. Homology modeling of CYP107CB2 from B. lehensis G1

The amino acid sequence of CYP107CB2 was submitted to NCBI BLAST server through PDB using PSI-BLAST for homologous template selection. The template having the closest match (highest similarity) for all retrieved query sequences was chosen for the construction of three dimensional structural model of cytochrome P450 of *B. lehensis* G1. An automated homology modeling program SwissModel (Arnold et al., 2006) was used to generate the model. Cytochrome P450 vitamin D_3 hydroxylase (Vdh) from *P. autotropica*'s (protein data bank code: 3A4G; resolution: 1.75 Å) was used as a template.

2.4. Validation of the CYP107CB2 model

The constructed model (CYP107CB2) was subjected to quality assessment and structure validation with respect to its geometry and energy aspect using PROCHECK (Laskowski et al., 1996) and ProSA (Wiederstein and Sippl, 2007). Root mean square deviation (RMSD) between the homologue model and its template was calculated by structural superimposition of both structures by using MUSTANG method (Konagurthu et al., 2006) in YASARA (Yet Another Scientific Artificial Reality Application) program Version 11.3.22 (Krieger et al., 2002).

2.5. Molecular docking of CYP107CB2 with vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3

To predict the potential substrates for CYP107CB2, the built model was subjected to *in silico* docking experiment using AutoDock and its functionality which is embedded into the YASARA Structure package. For this purpose, vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3 were focused and docked into the region of active site. The structures of the substrates (ligands) were obtained from NCBI PubChem. Lamarckian genetic algorithm using 0.2 Å translational step size; 5 Å orientational and torsional step size was applied in docking experiment (Morris et al., 1998). The grid size of $19 \times 21 \times 19$ Å^3 with grid point spacing 0.375 Å was set in the built model for ligands docking. The other docking parameters used in this analysis were: 25 docking runs; 25,000,000 energy evaluation; AMBER03 force field; 150 population size; the number of generation was 27,000. After docking simulation, the binding energy calculated by AutoDock was obtained from the summary of log file. The data were sorted by the positioned distances of oxidizable carbon atom and binding energy, where shorter distance and positive energy indicated stronger binding.

2.6. Expression of recombinant CYP107CB2 protein

Gene sequence of *CYP107CB2* was cloned into pET102/D-TOPO vector and the expression of recombinant cytochrome P450

protein was conducted using *Escherichia coli* Rosetta gami (DE3). The strains were grown at 37 °C with shaking (200 rpm) until the optical density of the culture at 600 nm (OD_{600}) reached the range of 0.5–0.8. Protein expression was induced by 0.8 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 0.5 mM heme precursor delta-aminolevulinic acid was added. Thereafter, the culture was further cultivated at a lowered temperature of 25 °C on a rotary shaker for 18 h. The cells were harvested at 12,000 × *g* for 30 min at 4 °C and the cell pellet was re-suspended in 10 mL, 50 mM phosphate buffer (pH 7.4). Cells were disrupted by sonication in ice (5 × 1 min, output control 4, duty cycler 40%) followed by centrifugation at 12,000 × *g* at 4 °C for 20 min to remove cell debris and insoluble proteins. The expressed protein in the soluble fraction was subjected to protein purification.

2.7. Purification of CYP107CB2

The recombinant CYP107CB2 protein was purified by using Nickel–Sepharose column. The cytosolic fraction containing soluble protein was applied to a 10 mL Nickel–Sepharose column (1.6 × 30 cm) previously equilibrated with 50 mM potassium phosphate buffer (pH 7.4) containing 40 mM imidazole and 0.5 M sodium chloride at a flow rate of 1 mL/min. The column-bound enzyme was washed with 50 mM potassium phosphate buffer (pH 7.4) containing 500 mM imidazole and 0.5 M sodium chloride. Elution was carried out by a linear gradient of imidazole (0–500 mM) in 50 mM potassium phosphate buffer (pH 7.4). The purified enzyme was subjected to hydroxylation activity measurement.

2.8. Measurement of CYP107CB2 hydroxylation activity

Hydroxylation activity of CYP107CB2 was measured according to a protocol as reported by Fujii et al. (2009) with slight modification. A reaction mixture (500 μL) containing 50 mM phosphate buffer (pH 7.4), 32 μg/mL spinach ferredoxin and 0.1 U/mL spinach ferredoxin–NADPH reductase, 3 U/mL glucose dehydrogenase, 60 mM D-glucose, 2 mM NADPH, and 0.78 mg purified CYP107CB2 was prepared. The reaction was incubated with 100 μM 1α-hydroxyvitamin D_3 at 25 °C for 3 h. The unreacted substrate and product were extracted once with 1 mL and once with 0.75 mL of ethyl acetate. The combined ethyl acetate phase

```
1     atgtcaattgtgctgaataaaaaagagtttcatcaagagccttacgaattttataaggag
1      M  S  I  V  L  N  K  K  E  F  H  Q  E  P  Y  E  F  Y  K  E
61    atccgtcctcatgatgcctttgcaaaggtgaaattagtgagtggtattcaggattcatgg
21     I  R  P  H  D  A  F  A  K  V  K  L  V  S  G  I  Q  D  S  W
121   gttgcgtttacatatgaagcggctgaagcagtcttaaaagatgaacgatttgtaaaaaat
41     V  A  F  T  Y  E  A  A  E  A  V  L  K  D  E  R  F  V  K  N
181   ccaagagcggtttttcccagacgtcagtgaacatgaattgatgccaattacccatagcatg
61     P  R  A  V  F  P  D  V  S  E  H  E  L  M  P  I  T  H  S  M
241   ttgtttgctgatccaccagatcatcgacggttgcgaagtcttgttcagcgaggttttaca
81     L  F  A  D  P  P  D  H  R  R  L  R  S  L  V  Q  R  G  F  T
301   ccgaaaatgattcaacgccttcaaggtcgaattgaagaaattgcaaagatacaagtcgaa
101    P  K  M  I  Q  R  L  Q  G  R  I  E  E  I  A  K  I  Q  V  E
361   caaatgaaagggaaggaaaccgttgatttaattgcagactatgcatttcctattcctatt
121    Q  M  K  G  K  E  T  V  D  L  I  A  D  Y  A  F  P  I  P  I
421   attgttatctgtgaattacttggtgtccctcctgaggaccgcttagactttcaaagatgg
141    I  V  I  C  E  L  L  G  V  P  P  E  D  R  L  D  F  Q  R  W
481   tctaacagtatggtagaaataaacgacgatccaagtttttatgagcaagttgaagcgcat
161    S  N  S  M  V  E  I  N  D  D  P  S  F  Y  E  Q  V  E  A  H
541   atgaaagaatttcagctttatattgaacaattgctggccgagaagcgaatccatcctcaa
181    M  K  E  F  Q  L  Y  I  E  Q  L  L  A  E  K  R  I  H  P  Q
601   gacgatttgctttcagagctgatccgtgcagaagaagatggggataagctatcggttcaa
201    D  D  L  L  S  E  L  I  R  A  E  E  D  G  D  K  L  S  V  Q
661   gaattatatggtgccattatgctcatgatcgttgccggacatgaaacgacagttaattta
221    E  L  Y  G  A  I  M  L  M  I  V  A  G  H  E  T  T  V  N  L
721   attgccaatggaatgctcgccttgtttacacatcctgagcagttaaaaaaattaaaagaa
241    I  A  N  G  M  L  A  L  F  T  H  P  E  Q  L  K  K  L  K  E
781   tcaccaagtttaattgatggagccattgaagagattttacgatttaatggtccagttgag
261    S  P  S  L  I  D  G  A  I  E  E  I  L  R  F  N  G  P  V  E
841   tttagtactgatcgctacgcaaaagaatcgtttacgttcatgggcaaacagctgcaaaaa
281    F  S  T  D  R  Y  A  K  E  S  F  T  F  M  G  K  Q  L  Q  K
901   ggagatcatgttctcgtttctcttgcgtcagcagatcatgacccggcagtttttcagag
301    G  D  H  V  L  V  S  L  A  S  A  D  H  D  P  A  V  F  S  E
961   ccggacaaacttgttatcactcgtgaaaaaagtccgcaccttgcgtttggaaaaggaatt
321    P  D  K  L  V  I  T  R  E  K  S  P  H  L  A  F  G  K  G  I
1021  cactattgcctaggtgcgccactcgcacgattagaagggaaaatagcgattcaaacattg
341    H  Y  C  L  G  A  P  L  A  R  L  E  G  K  I  A  I  Q  T  L
1081  cttaacaccttcccagagattcaaataaatacagcactcgctaatttggagtggcgacaa
361    L  N  T  F  P  E  I  Q  I  N  T  A  L  A  N  L  E  W  R  Q
1141  agcttcgtcattcgcggcttaaaagaactgccggtgaaattgaattaa
381    S  F  V  I  R  G  L  K  E  L  P  V  K  L  N  -
```

Fig. 1. Full length nucleotide and amino acid sequence of cytochrome P450 from *Bacillus lehensis* G1. The start (atg) and stop (taa) codon are in bold. Stop codon is denoted by dash. Three distinct characteristics of sequence motif used for identification and analysis of cytochrome P450 are indicated by the boxed amino acids; GxxT motif; ExxR motif; and cystein heme iron ligand (FxxGxxxCxG), respectively. ExPASy translate tool was used to translate nucleotide sequence to a protein sequence.

was allowed to dry and the remaining residue was dissolved in 100 μL of methanol. The methanol solution was analyzed by high performance liquid chromatography (HPLC) equipped with a SunFire™ C18 column, 5 μm (4.6 × 160 mm; Waters) with 1.0 mL/min flow rate. Samples were resolved on the column using a linear gradient of 50–100% acetonitrile for 12 min followed by an elution with 100% acetonitrile for 13 min. The metabolites were detected at 265 nm and the products were compared using authentic standards.

3. Results and discussion

3.1. Molecular analysis of cytochrome P450 gene (CYP107CB2) from alkaliphilic B. lehensis G1

A gene fragment with high similarity to cytochrome P450 listed on the NCBI database was identified from the *B. lehensis* G1 genomic database. The open reading frame contained 1188 base pairs (bp) or nucleotides that encoded a protein comprising

```
3A4G(Vdh)     1   ------MALTTTGTEQHDLFSGTFWQNPHPAYAALRAEDPVRKLALPDGP   44
Cyp107CB2     1   ------MSIVLNKKE--------FHQEPYEFYKEIRPHDAFAKVKLVSGI   36
2BVJ(PikC)    1   MRRTQQGTTASPPVLDLGALGQDFAADPYPTYARLRAEGPAHRVRTPEGD   50
1EGY(EryF)    1   ---------ATVPDLES----DSFHVDWYSTYAELRETAPVTPVR-FLGQ   36
                              .             *  : :  *  :*   .   :    *
3A4G(Vdh)    45   --VWLLTRYADVREAFVDPRLSKDWRHTLP--------ED-QRADMPATP   83
Cyp107CB2    37   QDSWVAFTYEAAEAVLKDERFVKNPRAVFP--------DVSEHELMPITH   78
2BVJ(PikC)   51   -EVWLVVGYDRARAVLADPRFSKDWRN--------STTPLTEAEAALN--   89
1EGY(EryF)   37   -DAWLVTGYDEAKAALSDLRLSSDPKKKYPGVEVEFPAYLGFPEDVRNYF   85
                     *:   *  .. .: * *: .: :
3A4G(Vdh)    84   TPMMILMDPPDHTRLRKLVGRSFTVRRMNELEPRITEIADGLLAGLPTD-  132
Cyp107CB2    79   S--MLFADPPDHRRLRSLVQRGFTPKMIQRLQGRIEEIAKIQVEQMKGK-  125
2BVJ(PikC)   90   -HNMLESDPPRHTRLRKLVAREFTMRRVELLRPRVQEIVDGLVDAMLAAP  138
1EGY(EryF)   86   ATNMGTSDPPTHTRLRKLVSQEFTVRRVEAMRPRVEQITAELLDEVGDS-  134
                     *   *** * ***.** : ** : :: :. *: :*.   :  :
3A4G(Vdh)   133   -GPVDLMREYAFQIPVQVICELLGVPAEDRDDFSAWSS--VLVDDSP--A  177
Cyp107CB2   126   -ETVDLIADYAFPIPIIVICELLGVPPEDRLDFQRWSNSMVEINDDPSFY  174
2BVJ(PikC)  139   DGRADLMESLAWPLPITVISELLGVPEPDRAAFRVWTDAFVFPDD----P  184
1EGY(EryF)  135   -GVVDIVDRFAHPLPIKVICELLGVDEAARGAFGRWSSEILVMDPER--A  181
                     .*::   *  :*: **.*****    *  *  *:.  :  :
3A4G(Vdh)   178   DDKNAAMGKLHGYLSDLLERKRTEPDDALLSSLLAVSDEDGDRLSQEELV  227
Cyp107CB2   175   EQVEAHMKEFQLYIEQLLAEKRIHPQDDLLSELIRAE-EDGDKLSVQELY  224
2BVJ(PikC)  185   AQAQTAMAEMSGYLSRLIDSKRGQDGEDLLSALVRTSDEDGSRLTSEELL  234
1EGY(EryF)  182   EQRGQAARFVVNFILDLVERRRTEPGDDLLSALISVQDDDDGRLSADELT  231
                     :      :.  ::  *:  :* .  : *** *: .. :*..:*: :**
                           GxxT                                  Exx
3A4G(Vdh)   228   AMAMLLLIAGHETTVNLIGNGVLALLTHPDQRKLLAEDPSLISSAVEEFL  277
Cyp107CB2   225   GAIMLMIVAGHETTVNLIANGMLALFTHPEQLKKLKESPSLIDGAIEEIL  274
2BVJ(PikC)  235   GMAHILLVAGHETTVNLIANGMYALLSHPDQLAALRADMTLLDGAVEEML  284
1EGY(EryF)  232   SIALVLLLAGFEASVSLIGIGTYLLLTHPDQLALVRADPSALPNAVEEIL  281
                     .   :::**.*::*.**. *   *::**:*   :  . : : .*:**:*
                     R
3A4G(Vdh)   278   RFDSPVSQAPIRFTAEDVTYSGVTIPAGEMVMLGLAAANRDADWMPEPDR  327
Cyp107CB2   275   RFNGPVEFSTDRYAKESFTFMGKQLQKGDHVLVSLASADHDPAVFSEPDK  324
2BVJ(PikC)  285   RYEGPVESATYRFPVEPVDLDGTVIPAGDTVLVVLADAHRTPERFPDPHR  334
1EGY(EryF)  282   RYIAPPETTT-RFAAEEVEIGGVAIPQYSTVLVANGAANRDPSQFPDPHR  330
                     *: .* . :. *:. * .   *  :   . *::  . *.: .  :.:*.:
                                FxxGxxxCxG
3A4G(Vdh)   328   LDITRDASGGVFFGHGIHFCLGAQLARLEGRVAIGRLFADRPELALAVGL  377
Cyp107CB2   325   LVITREKSPHLAFGKGIHYCLGAPLARLEGKIAIQTLLNTFPEIQINTAL  374
2BVJ(PikC)  335   FDIRRDTAGHLAFGHGIHFCIGAPLARLEARIAVRALLERCPDLALDVSP  384
1EGY(EryF)  331   FDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSLGIDA  380
                     : : *:    : **:***:*:*  **:**..:*:  *:   * : :
3A4G(Vdh)   378   DELVYRESTLVRGLSRMPVTMGPRSALE----                    405
Cyp107CB2   375   ANLEWRQSFVIRGLKELPVKLN----------                    396
2BVJ(PikC)  385   GELVWYPNPMIRGLKALPIRWRRGREAGRRTG                    416
1EGY(EryF)  381   DDVVWRRSLLLRGIDHLPVRLDG---------                    403
                     :: :  . ::**:. :*:
```

Fig. 2. Multiple sequence alignment of CYP107CB2 and CYP107 superfamily of protein sequences. Vdh (PDB: 3A4G) from *Pseudonocardia autotrophica*; PikC (PDB: 2BVJ) from *Streptomyces venezuelae*; EryF (PDB: 1EGY) from *Saccharopolyspora erythraea*. Protein sequences were aligned using ClustalW. Hyphens were introduced as a gap to provide better sequence alignments. Fully conserved residues are marked by a * in red, amino acid residues with conservation of strong group is indicated as in green, semi conserved residue is shown as in blue. The conserved motifs of cytochrome P450 are underlined. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

396 amino acids (Fig. 1). Sequence analysis using NCBI conserved domain database (CDD) showed the deduced amino acid sequence contained a highly conserved CYP family domain with an *E*-value of 4.69e-111. This analysis was in agreement with Pfam and InterProScan analyses which showed the presence of the functional domain of cytochrome P450 (accession number: PF00067 and IPR017972, respectively).

Furthermore, gene ontology (GO) annotation studies revealed that the novel CYP gene from *B. lehensis* G1 may be involved in oxidation-reduction processes (GO: 0055114), iron ion binding (GO: 0005506); electron carrier activity (GO: 0009055) and heme binding (GO: 0020037). Based on enzyme classification, this protein is classified as an oxidoreductase with E.C number: 1.14 that acts on paired donors, with insertion or reduction of molecular oxygen. All the above results thus strongly suggest that the gene sequence extracted from the genome of *B. lehensis* G1 indeed codes for cytochrome P450 with possible involvement in cellular oxidation processes.

The amino acid sequence has a predicted isoelectric point of 5.40 and molecular weight of 45113 Da. The most abundant amino acid residues in this protein was Leu (43 residues out of 395), followed by Glu (38 residues) and Ile (30 residues). The least abundant amino acid residues were Trp (3 residues) and Cys (2 residues). The protein seems to be of low polarity with 101 polar amino acid residues, 189 non polar residues and 105 charged residues.

Based on Nelson's P450 nomenclature (Nelson, 2006), the classification of cytochrome P450s into homologous families and superfamilies is mainly according to sequence similarity. If the sequence shares >40% amino acid identity, they are classified into the same family. The protein sequence of cytochrome P450 from *B. lehensis* G1 was submitted to P450 nomenclature committee. This sequence (assigned as CYP107CB2) was found to be 60% identical to *Bacillus clausii* CYP107CB1 and hence, duly classified into CYP107 family. The nucleotide sequence of CYP107CB2 was submitted to GenBank with the accession number KJ641923.

3.2. Multiple sequence alignments

Multiple sequence alignment of CYP107CB2 and CYP107 superfamily of protein sequences was performed in order to ensure the presence of conserved sequence motifs and functionally important amino acid residues that are highly conserved in the cytochrome P450 superfamily (Kumar, 2010). The sequence of CYP107CB2 and other members of the CYP107 proteins family (e.g., PDB: 3A4G, 2BVJ and 1EGY) were retrieved from Protein Data Bank (PDB) and used to align with CYP107CB2. The sequence identities between CYP107CB2 and 3A4G (Vdh), 2BVJ (PikC) and 1EGY (EryF) were 44%, 42% and 36%, respectively. As shown in Fig. 2, the sequence alignment indicated that the putative protein of CYP107CB2 revealed convincing homology to an authentic cytochrome P450 sequence. The cysteine heme-binding conserved motif FxxGxxxCxG that is classic to cytochrome P450 (Werck-Reichhart and Feyereisen, 2000) was found at position 336–345 in the deduced amino acid sequence. The sequence motif conforms to the PROSITE consensus pattern [FW]-[SGNH]-x-[GD]-{F}-[RKHPT]-{P}-C-[LIVMFAP]-[GAD] with PROSITE accession number PS00086. It is a general formulation for identifying cytochrome P450 sequence in a genome and this has allowed the identification of CYP107CB2 from the genome of *B. lehensis* G1.

The FxxGxxxCxG motif located in the C-terminal region of the heme binding domain constituted the cysteine pocket, and was found to be in proximity to the putative active site of CYP107CB2. The cysteine in this motif is essential for heme binding and thiolate-bond formation, and was determined to be at position 343 in CYP107CB2 (Gomaa et al., 2007). A second, highly-conserved, GxxT motif, responsible for oxygen binding (Schaller and Stintzi, 2009) was also noted at position 233-236, and a third motif, ExxR, responsible for salt bridge formation, (Graham-Lorence and Peterson, 1996) was observed at position 271-274 of the CYP107CB2 protein sequence.

3.3. Homology modeling of CYP107CB2

Homology modeling is one of the fastest and most attractive approaches for predicting three dimensional protein structures, outside of X-ray crystallography and NMR for insight of the structure–function relationship of an enzyme (Chang et al., 1997). For this purpose, the sequence of CYP107CB2 was used as input to identify the best template for three dimensional protein structure generation. NCBI BLAST database reveals that cytochrome P450 vitamin D_3 hydroxylase (Vdh) from *P. autotropica*'s (Protein data bank code: 3A4G; resolution: 1.75 Å) is a suitable template to build the model of CYP107CB2 with the highest sequence identity scores of 44%. This protein is responsible for vitamin D_3 hydroxylation by converting vitamin D_3 to 1α,25-hidroxyvitamin D_3 (Yasutake et al., 2010).

The homologue model of CYP107CB2 (Fig. 3A) shows structurally conserved modules of cytochrome P450 with 13 alpha-helices (A-L), 12 beta-strains (1-4), 1 meander loop and 1 cysteine pocket. The heme cofactor which served as the fifth ligand to cysteine (Zhang et al., 2011) was found to be buried in the cavity of the catalytic site (Fig. 3B). This is the site where carbon monoxide forms ferrous–CO complex and gives an absorbance peak at 450 nm (Munro et al., 2007; Peterson and Graham, 1998). Additionally, ferrous ion (from heme) was seen to be coordinated by the sulfur atom of Cys-343 with the predicted distance of 2.86 Å (Fig. 3C). The cysteine residue was located in a loop region, preceding L-helix. The heme group was bound to side chain of His-88, Arg-285 and His-341. As expected, the heme cofactor was located at I-helix and sandwiched with L-helix, where the other conserved motifs of cytochrome P450 including ExxR and GxxT were located at K-helix and I-helix, respectively. The location of key structural elements and the presence of conserved motifs indicate that CYP107CB2 satisfies the overall topology and general folding properties of a prototypical P450 enzyme (Mclean et al., 2011; Munro et al., 2007).

The constructed model of CYP107CB2 was subjected to structural quality assessment and the model was validated for *psi* and *phi* torsion positions using Ramachandra plot. The analysis revealed that 99.4% and 0.6% of the residues of CYP107CB2 built model were placed in the allowed and disallowed regions, respectively (Fig. 4A). Additionally, the value of goodness (*G*) factor from PROCHECK was obtained in acceptable range (−0.22) indicating the reliability of the model. The validation result shows the model is reasonably accurate in terms of stereochemistry. In comparison, ProSA evaluation shows the *Z*-score of −9.75 indicating a high quality of the predicted structure (Fig. 4B) and therefore considered reliable for subsequent *in silico* studies.

Superimposition of the polypeptide backbone of both CYP107CB2 model and its template (PDB: 3A4G) gave forth an RMSD value of 0.32 Å over 384 aligned residues with 42.71% sequence identity (Fig. 5). Despite the differences in the sources of CYP107CB2 and Vdh the main structure of CYP107CB2 including the active site architectures were very similar to each other. Hydrophobic residues at the active sites of CYP107CB2 and Vdh (such as Val–Leu and Ala) were also similar. Gotoh (1992) had proposed six locations that would be the potential substrate recognition sites (SRSs) in cytochrome P450s. One of the substrate recognition sites of cytochrome P450, a β4 turn located close to a

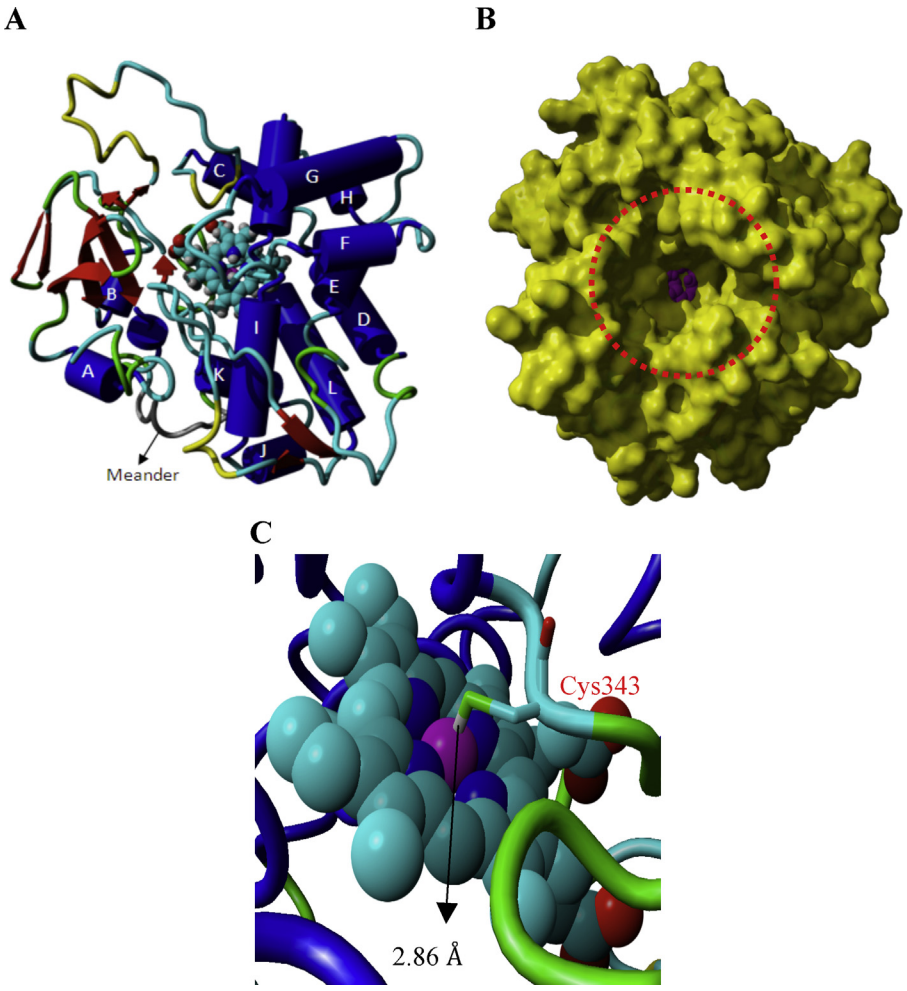

Fig. 3. (A) Homology model of CYP107CB2. Helices are shown as blue cylinder, β-sheets in red, meander loop in grey and heme in light blue. (B) Surface representation of cytochrome P450 with heme cofactor (magenta) located in hydrophobic cavity. The active site of CYP107CB2 is circled in red. (C) Stereoview of heme prosthetic group and cysteine residue of CYP107CB2. The distance between sulfur (from Cys-343) and ferrous ion (from the heme) had a calculated value of 2.86 Å. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cysteine pocket (Peterson and Graham, 1998) of CYP107CB2 is similar to Vdh. This indicates that structurally, the active site architecture of CYP107CB2 may be highly similar to Vdh.

Although the structure similarities between CYP107CB2 and Vdh were identified, the differences of both structures were also being observed. These were the non-conserved regions where the compositions of amino acids were different and varied among the superfamily of CYPs (Gomaa et al., 2007). The superimposition of CYP107CB2 homology model and Vdh shows that there were three parts of CYP107CB2 which deviated from Vdh. These structures

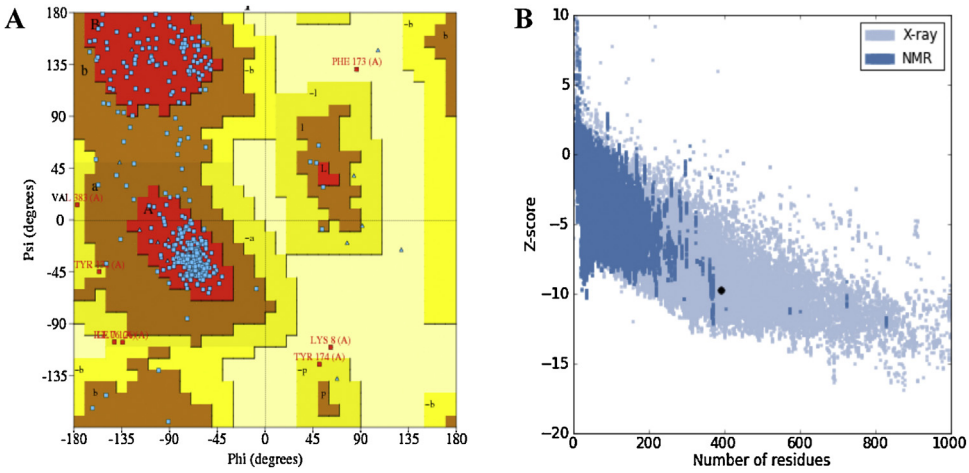

Fig. 4. Quality assessment of CYP107CB2 built model. (A) PROCHECK Ramachandran plot of CYP107CB2 model. (B) ProSA overall model quality of CYP107CB2.

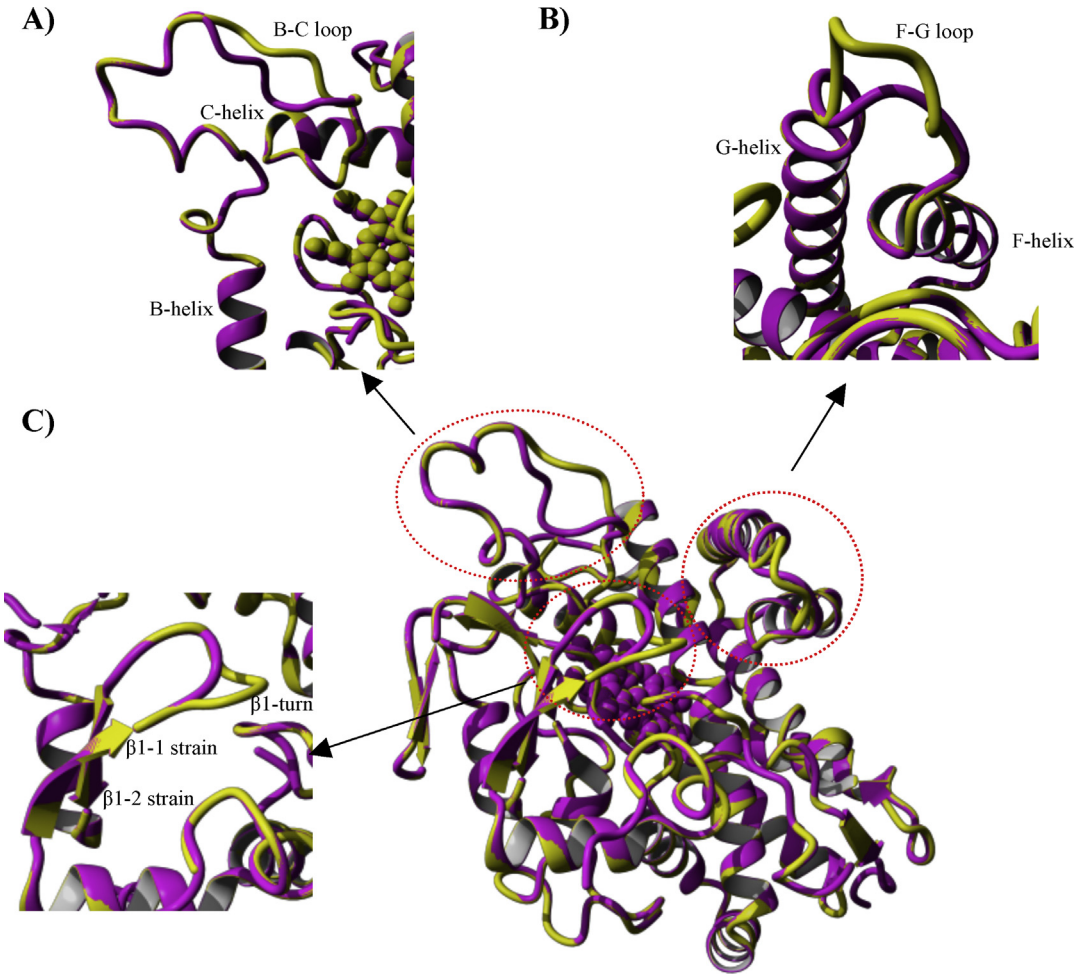

Fig. 5. Superimposed structure of homologue model CYP107CB2 (yellow) with the structure of cytochrome P450 Vdh (PDB: 3A4G) (magenta). Both structures had a calculated RMSD value of 0.32 Å over 384 aligned residues with 42.71% sequence identity. The flexible regions of CYP107CB2 which deviated from Vdh are indicated as (A) B-C loop; (B) F-G loop; (C) β1-turn. The overlapping structures display similar structural elements. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

displaying the deviations were mainly at the substrate entry site which is essential in substrate recognition. B-C loop (Fig. 5A) and F-G loop (Fig. 5B) are highly variable where these loops at CYP107CB2 are broadened and expanded as compared to Vdh. The hydrophobic amino acid residues Leu-73, Met-74, Pro-75 and Ile-76 constituted the B-C loop. In F-G loop of CYP107CB2, there is a longer and extended loop, as there is a difference in an extra six amino acids from Vdh. They are Asn-168, Asp-169, Asp170, Pro-171, Ser-172 and Phe-173 as compared to Ser-167 and Pro-168 in Vdh. Such a flexible region has been reported for many CYPs including CYP105A1 and CYP2R1 (Sugimoto et al., 2008; Strushkevich et al., 2008). They are one of the SRSs in CYPs and form a ceiling or lid which allows substrate binding (Graham and Peterson, 1999). The loop that connects between β-strain 1-1 and 1-2 was also found deviated from Vdh (Val-33, Ser-34 and Gly-35 from CYP107CB2 are different from Pro-33 in Vdh). These data presumed that the B-C loop, F-G loop and β1-turn of CYP107CB2 may be involved in substrate binding as they are located at the substrates entry point.

Table 1
Binding energy and predicted contacting residues of CYP107CB2 that interact with vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3.

Substrates	Binding energy (kcal/mol)	Contacting residues
Vitamin D_3 (C25)	8.73	Thr-77, Ser-79, Leu-81, Phe-82, Met-181, Met-227, Leu-228, Val-231, Ala-232, Glu-235, Thr-236, Val-279, Ser-282, Thr-283, Val-383, Ile-384, Heme
Vitamin D_3 (C1)	9.36	Thr-77, Ser-79, Leu-81, Phe-82, Met-181, Gly-224, Met-227, Leu-228, Val-231, Ala-232, Gly-233, Glu-235, Thr-236, Val-279, Thr-283, Val-383, Heme
25-hydroxyvitamin D_3 (C1)	8.97	Thr-77, Ser-79, Leu-81, Phe-82, Met-181, Gly-224, Met-227, Leu-228, Val-231, Ala-232, Glu-235, Thr-236, Val-279, Ser-282, Thr-283 Val-383, Heme
1α-hydroxyvitamin D_3 (C25)	9.37	Thr-77, Ser-79, Leu-81, Phe-82, Met-181, Met-227, Leu-228, Val-231, Ala-232, Thr-236, Val-279, Thr-283, Heme

3.4. Molecular docking of CYP107CB2

Protein three-dimensional structures provide important information for understanding of molecular function and allow the analysis of interaction between the homologue structure and its substrates (Beedkar et al., 2012). The three-dimensional structure of CYP107CB2 displayed a large pocket located above the heme iron. Based on the similarity of the active site of CYP107CB2 to that of Vdh this pocket was believed to fit vitamin D_3 or other related substrates. However, the large pocket of CYP107CB2 led to the accommodation of substrates in different orientations. Since the hydroxylation carbon atom of each substrate have been

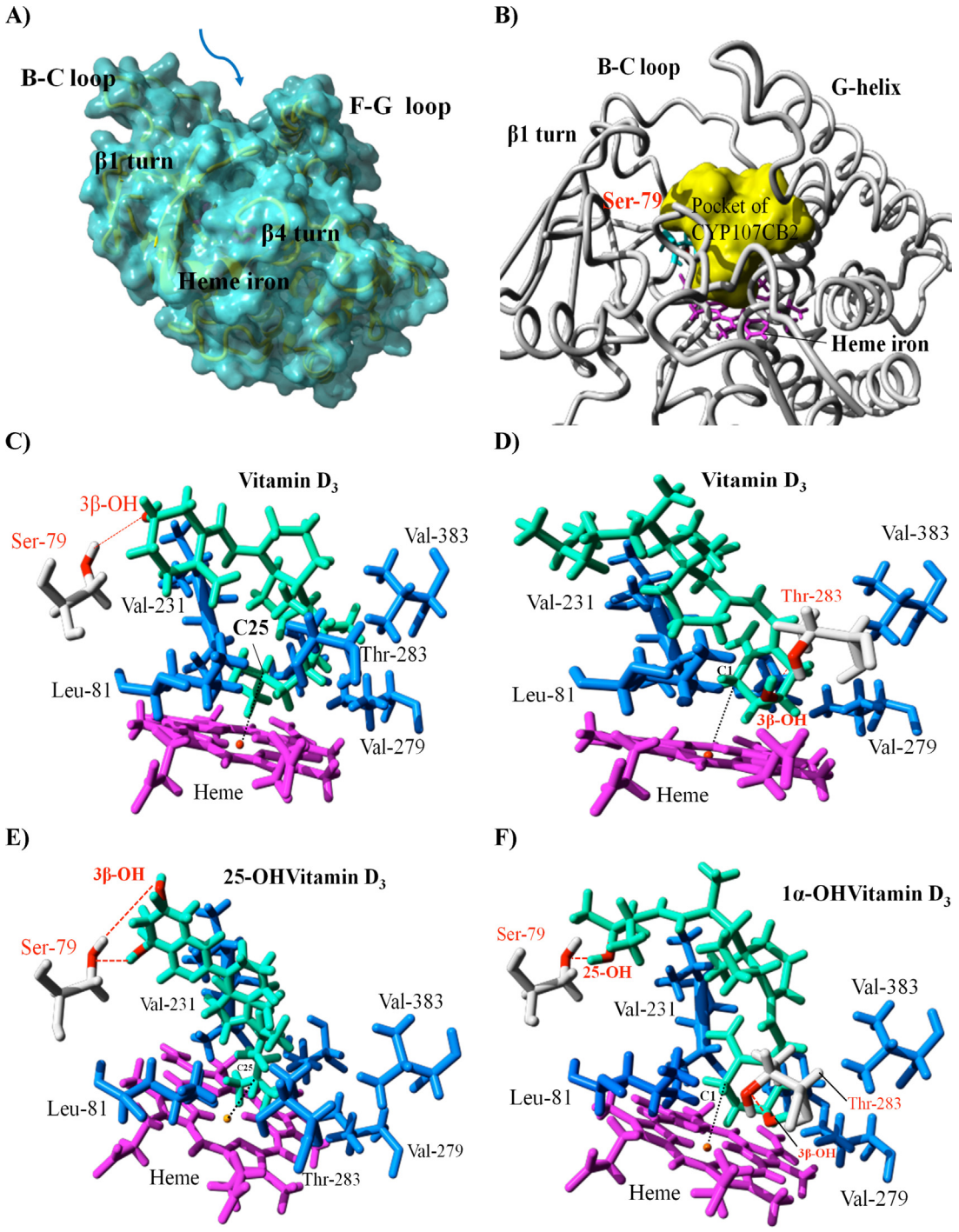

Fig. 6. Possible binding mode of CYP107CB2 and its candidate substrates for docking analysis. (A) Stereo view of substrate entrance and product exit channel. Arrow points the proposed access channel of CYP107CB2. (B) Active site of CYP107CB2 in the absence of substrate. B-C loop, F-G loop, β1 turn, β4 turn, and heme are shown in Fig. A and B. The substrate binding cavity is displayed in yellow. Docking models of CYP107CB2 with the potential substrates within its active site. (C) Vitamin D_3 for hydroxylation at C-25; (D) vitamin D_3 for hydroxylation at C-1; (E) 1α-hydroxyvitamin D_3 for hydroxylation at C-25; (F) 25-hydroxyvitamin D_3 for hydroxylation at C-1. The heme is shown as stick model in margenta; ferrous ion in brown; substrate in green; Ser-79 and Thr-283 that contributes to hydrogen bonding are represented in grey; similar contacting residues that interact with the predicted substrates are shown in blue. The distance between the predicted hydroxylation carbon atom with heme is indicated by the black dashed line. Hydrogen bonds are shown as red dotted line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

known, the best binding conformation showing the highest binding energy score calculated by YASARA and the appropriate hydroxylation site nearest to the heme were chosen and considered as the optimal conformation.

Blind docking of vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3 with CYP107CB2 was conducted in order to predict the possible substrates and ligand-binding residues of CYP107CB2. Results revealed that most of the residues that interacted with the substrates constituted the active site (Table 1). Seven residues including Leu-81, Val-231, Ala-232, Thr-236, Val-279, Thr-383 and Val-383 were found facing vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3 in the vertical orientation. Therefore, a simulation cell with the size of $19 \times 21 \times 19$ Å^3 was set and docking simulation was focused within these residues. In addition to these residues, a heme prosthetic group was also located within the P450 active site, where it played a central role in the P450's catalytic mechanism (Werck-Reichhart and Feyereisen, 2000). Therefore, the position of the heme molecule was taken into account for docking simulations as well. The most suitable binding mode for the substrate was selected from 25 poses that were generated.

Fig. 6A illustrates the surface view of CYP107CB2 that consists of B-C loop, F-G loop and β-1 turn. They form an "open gap" that binds specific substrates and believed to be substrate access channel starting from B-C loop. The channel directs the access of substrate into the substrate binding pocket (Fig. 6B). It has been reported that non-polar amino acids form a central hydrophobic cavity that defines much of the protein's substrate access channel and active site. Hydrophobic amino acid residues in this region function to stabilize the protein's tertiary structure and allow for hydrophobic interactions with substrates (Peterson and Graham, 1998). The presence of the hydrophobic cavity in CYP107CB2 was observed whereby the side chains of non-polar amino acid residues including Leu-81, Val-231, Ala-232, Val-279, and Val-383 were located proximately to the heme (Table 1). Hence, it can be assumed that CYP107CB2 most likely binds and metabolizes hydrophobic compounds such as vitamin D_3, 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3. The presence of hydrophobic residues in the catalytic pocket is common for CYP. For instance, vitamin D_3 is located in the catalytic pocket of CYP2R1 surrounded by hydrophobic side chain of amino acids facilitating hydrophobic interactions (Strushkevich et al., 2008; Yasutake et al., 2010).

Fig. 6C displays the docking model of CYP107CB2 which accommodated vitamin D_3 for 25-hydroxylation. Vitamin D_3 lies directly above the heme iron with the distance of 4.13 Å from C-25. The side chain of Ser-79 from B-helix located near to the entrance surface of CYP107CB2 is close to 3β-OH of vitamin D_3 with the distance of 3.86 Å. It is predicted that Ser-79 forms a hydrogen bond with 3β-OH group which allows the orientation of aliphatic side chain of vitamin D_3 to be placed above the heme-iron appropriate for 25-hydroxylation.

When the same substrate was docked into the active site, another substrate binding mode was observed (Fig. 6D).

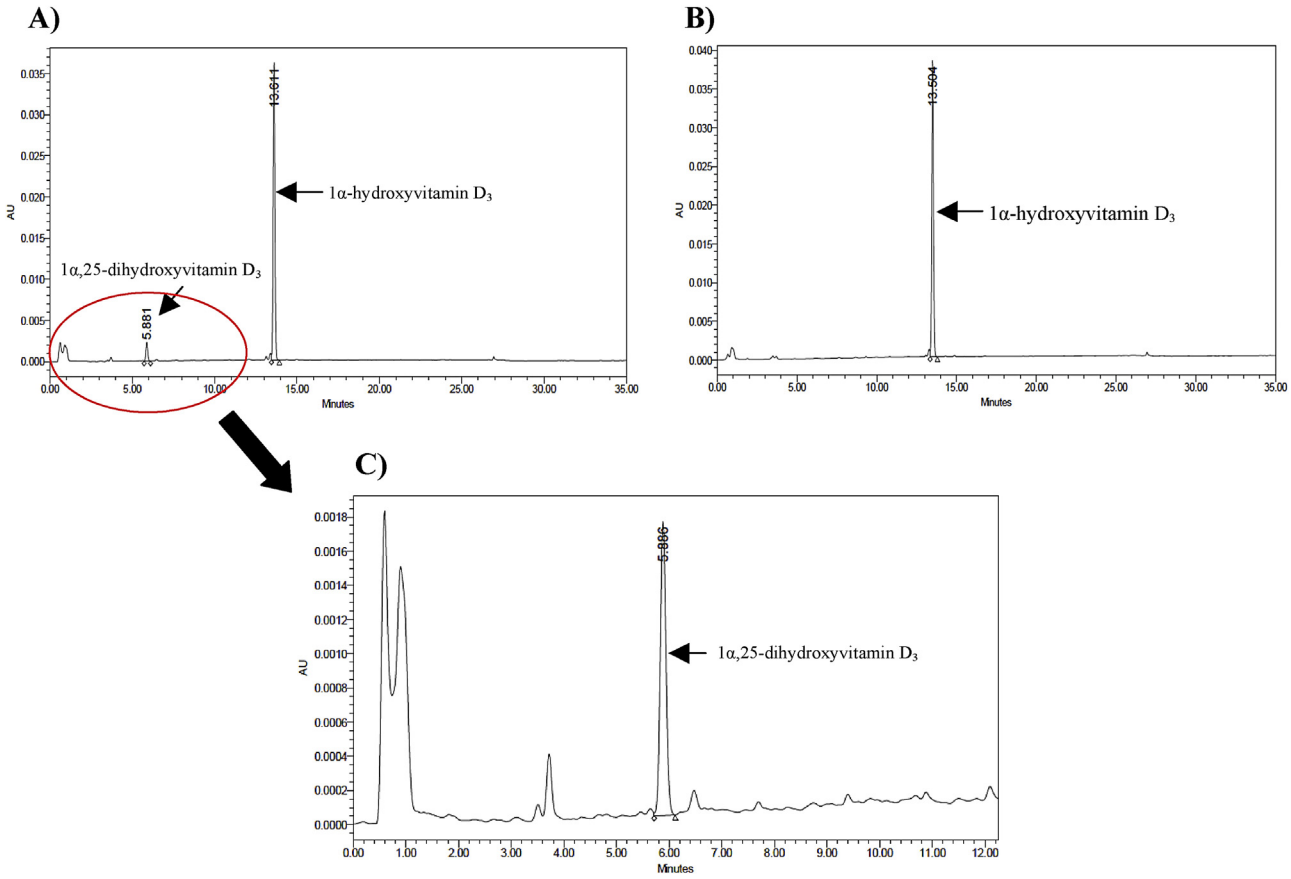

Fig. 7. HPLC analysis of 1α-hydroxyvitamin D_3 and its metabolite by CYP107CB2. The reaction mixtures in the (A) presence of CYP107CB2; (B) absence of CYP107CB2. The metabolites were extracted and analyzed by HPLC. The identified substrate and product peaks were compared with the retention time of the authentic standards: 1α-hydroxyvitamin D_3 and 1α,25-dihydroxyvitamin D_3, respectively. (C) The putative 1α,25-dihydroxyvitamin D_3 peak at RT = 5.88 min.

Thr-283 was observed to interact with 3β-OH of A-ring vitamin D_3 (distance 3.79 Å). The binding mode indicates that vitamin D_3 is in anti-parallel orientation where the A-ring was facing the heme iron. The carbon atom at position 1 is at closed proximity to the heme iron with the estimated distance of 4.19 Å. Therefore, 1α-hydroxylation is suspected to take place at this particular position.

Similar docking complex to vitamin D_3 (for 25-hydroxylation) was observed when 1α-hydroxyvitamin D_3 was docked into CYP107CB2 (Fig. 6E). The side chain of Ser-79 with a distance of 3.29 Å, is expected to form a hydrogen bond with 3β-OH group of 1α-hydroxyvitamin D_3. Additionally, the binding complex also shows that 1α-OH group of 1α-hydroxyvitamin D_3 may interact with Ser-79 as the position of 1α-OH is fairly close to OG1 atom (distance 1.99 Å). These interactions are suspected to orientate the molecule and lead the aliphatic side chain containing the carbon-25 closer to heme iron (distance of 3.48 Å) for 25-hydroxylation. In comparison to CYP105A1, hydroxylation process was reported to take place at C-25 of 1α-hydroxyvitamin D_3 for 1α,25-dihydroxyvitamin D_3 formation. The A-ring of 1α-hydroxyvitamin D_3 is anchored by the hydrogen bond between the 3β-OH group and Arg-193 from G-helix and thus, allow for hydroxylation as the C-25 was close to heme iron (Sugimoto et al., 2008; Sakaki et al., 2011).

The binding mode of the other potential substrate, 25-hydroxyvitamin D_3 into the active pocket of CYP107CB2 is illustrated in Fig. 6F. The Ser-79 and Thr-283 are the amino acids that form hydrogen bonds with 25-OH and 3β-OH group, respectively. The atom OG1 of Ser-79 is 1.95 Å from 25-OH group while OG1 of Thr-283 is 4.71 Å from 3β-OH of 25-hydroxyvitamin D_3. The A-ring of 25-hydroxyvitamin D_3 points towards the heme iron and the CD-ring is located in the hydrophobic pocket of CYP107CB2 and presumed to interact with the hydrophobic amino acids. Thus, CYP107CB2 is predicted to form 1α-hydroxylation as the carbon atom is located 4.57 Å from heme.

In Table 1, the binding energies of CYP107CB2 with vitamin D_3 (C25), vitamin D_3 (C1), 25-hydroxyvitamin D_3 and 1α-hydroxyvitamin D_3 were calculated to be 8.73, 9.36, 8.97 and 9.37 kcal/mol, respectively. These thermodynamic values signify that our homology model demonstrates catalytic potential for the selected substrates. Additionally, all three substrates were found to be well-accommodated within the enzyme's large, hydrophobic binding cavity, further suggesting that each compound is a putative substrate for CYP107CB2. They exhibited similar contacting residues although they were in different orientations. Similar observation was reported for CYP105A1 that hydroxylated 25-hydroxyvitamin D_3 and vitamin D_3, having common contacting residues bound to the same site in catalytic pocket, although they were in opposite direction as indicated by docking analysis (Sakaki et al., 2011; Sugimoto et al., 2008). In summary, these findings provide important ideas and clues in terms of CYP107CB2 substrates selectiveness and its downstream biochemical characterization to identify its functional properties.

3.5. Measurement of CYP107CB2 hydroxylation activity

The recombinant CYP107CB2 was expressed and purified to study its biological function. CYP107CB2 was reacted with 1α-hydroxyvitamin D_3 in reconstituted system as mentioned in Section 2. The hydroxylated product formed from 1α-hydroxyvitamin D_3 was extracted and analyzed by HPLC. The chromatogram shows that 1α-hydroxyvitamin D_3 was hydroxylated to 1α,25-dihydroxyvitamin D_3 (Fig. 7A) as a peak was clearly seen at retention time (RT) of 5.88 min besides the un-reacted substrate peak at RT = 13.61 min. This indicated that in the metabolism of 1α-hydroxyvitamin D_3, the hydroxylation process occurred at C-25. It can be observed that the hydroxylated product exhibited the same spectrum and retention time as the authentic standard which is 1α,25-dihydroxyvitamin D_3 at 5.88 min (Fig. 7C). Thus, confirming that the product formed from the reaction is indeed 1α,25-dihydroxyvitamin D_3. Negative control experiment which consisted of the reaction mixture per se without the presence or addition of CYP107CB2 did not give forth the peak which corresponds to 1α,25-dihydroxyvitamin D_3 (Fig. 7B). All the results above verified that CYP107CB2 has hydroxylation activity. The protein is biologically active and capable in performing 25-hydroxylation process and may act as a 25-hydroxylase.

4. Conclusion

In conclusion, a gene encoding for a novel cytochrome P450 was identified from the genome of *B. lehensis* G1. Using the predicted CYP107CB2 homologue model, the structural determinants of the substrate binding site and access channel were putatively identified. This information provides useful insights into the nature of substrate selection for CYP107CB2, which can guide future docking studies, as well as functional experiments. Furthermore, CYP107CB2 contains a hydrophobic pocket similar to Vdh suggesting a considerable functional similarity. By predicting the binding modes of substrates likely to interact with CYP107CB2, we have initiated the biochemical characterization of this novel bacterial P450. CYP107CB2 is believed to be a biologically active protein as it is able to interact with 1α-hydroxyvitamin D_3 based on the preliminary substrate screening analysis. Our current work is focused on optimizing the production of CYP107CB2, in preparation for an advanced biochemical analysis of the protein's *in vitro* structure/function.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The authors are grateful to Research University Grant (05-02-12-2185RU) from Universiti Putra Malaysia for financing this project. The authors also gratefully acknowledge Malaysia Genome Institute (MGI) for providing the bacteria strain and its genome data. Swi See Ang acknowledges Ministry of Education, Malaysia for MyPhD scholarship award.

References

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195–201. doi:http://dx.doi.org/10.1093/bioinformatics/bti770.

Beedkar, S.D., Khobragade, C.N., Bodade, R.G., Vinchurkar, A.S., 2012. Comparative structural modeling and docking studies of uricase: possible implication in enzyme supplementation therapy for hyperuricemic disorders. Comput. Biol. Med. 42, 657–666. doi:http://dx.doi.org/10.1016/j.compbiomed.2012.03.001.

Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. J. Biotechnol. 124, 128–145. doi:http://dx.doi.org/10.1016/j.jbiotec.2006.01.026.

Chang, Y.T., Stiffelman, O.B., Vakser, I.A., Loew, G.H., Bridges, A., Waskell, L., 1997. Construction of a 3D model of cytochrome P450 2B4. Protein Eng. 10, 119–129. doi:http://dx.doi.org/10.1093/protein/10.2.119.

Fujii, Y., Kabumoto, H., Nishimura, K., Fujii, T., Yanai, S., Takeda, K., Tamura, N., Arisawa, A., Tamura, T., 2009. Purification, characterization, and directed evolution study of a vitamin D3 hydroxylase from *Pseudonocardia autotrophica*. Biochem. Biophys. Res. Commun. 385, 170–175. doi:http://dx.doi.org/10.1016/j.bbrc.2009.05.033.

Gomaa, M.S., Simons, C., Brancale, A., 2007. Homology model of 1α, 25-dihydroxyvitamin D3 24-hydroxylase cytochrome P450 24A1 (CYP24A1): active site architecture and ligand binding. J. Steroid Biochem. Mol. Biol. 104, 53–60. doi:http://dx.doi.org/10.1016/j.jsbmb.2006.09.041.

Gotoh, O., 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J. Biol. Chem. 267, 83–90.
Graham-Lorence, S.E., Peterson, J.A., 1996. Cytochrome P450, part B, methods in enzymology. Methods in Enzymology. Elsevier doi:http://dx.doi.org/10.1016/S0076-6879(96) 72037-2.
Graham, S.E., Peterson, J.A., 1999. How similar are P450s and what can their differences teach us? Arch. Biochem. Biophys. 369, 24–29. doi:http://dx.doi.org/10.1006/abbi.1999.1350.
Grogan, G., 2011. Cytochromes P450: exploiting diversity and enabling application as biocatalysts. Curr. Opin. Chem. Biol. 15, 241–248. doi:http://dx.doi.org/10.1016/j.cbpa.2010.11.014.
Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J., Lesk, A.M., 2006. MUSTANG: a multiple structural alignment algorithm. Proteins 64, 559–574. doi:http://dx.doi.org/10.1002/prot.20921.
Krieger, E., Koraimann, G., Vriend, G., 2002. Increasing the precision of comparative models with YASARA NOVA – a self-parameterizing force field. Proteins 47, 393–402. doi:http://dx.doi.org/10.1002/prot.10104.
Kumar, S., 2010. Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. Expert Opin. Drug Metab. Toxicol. 6, 115–131. doi:http://dx.doi.org/10.1517/17425250903431040.
Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486.
Mclean, K.J., Girvan, H.M., Mason, A.E., Dunford, A.J., Munro, A.W., 2011. Iron-Containing Enzymes. Royal Society of Chemistry, Cambridge doi:http://dx.doi.org/10.1039/9781849732987.
Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662. doi:http://dx.doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
Munro, A.W., Girvan, H.M., McLean, K.J., 2007. Variations on a (t) heme—novel mechanisms, redox partners and catalytic functions in the cytochrome P450 superfamily. Nat. Prod. Rep. 24, 585–609. doi:http://dx.doi.org/10.1039/b604190f.
Nelson, D.R., 2006. Cytochrome P450 nomenclature, 2004. Methods Mol. Biol. 320, 1–10. doi:http://dx.doi.org/10.1385/1-59259-998-2:1.
Peterson, J.A., Graham, S.E., 1998. A close family resemblance: the importance of structure in understanding cytochromes P450. Structure 6, 1079–1085. doi:http://dx.doi.org/10.1016/S0969-2126(98)00109-9.
Sakaki, T., 2012. Practical application of cytochrome P450. Biol. Pharm. Bull. 35, 844–849.
Sakaki, T., Sugimoto, H., Hayashi, K., Yasuda, K., Munetsuna, E., Kamakura, M., Ikushiro, S., Shiro, Y., 2011. Bioconversion of vitamin D to its active form by bacterial or mammalian cytochrome P450. Biochim. Biophys. Acta 1814, 249–256. doi:http://dx.doi.org/10.1016/j.bbapap.2010.07.014.
Schaller, A., Stintzi, A., 2009. Enzymes in jasmonate biosynthesis – structure, function, regulation. Phytochemistry 70, 1532–1538. doi:http://dx.doi.org/10.1016/j.phytochem.2009.07.032.
Schallmey, A., den Besten, G., Teune, I.G.P., Kembaren, R.F., Janssen, D.B., 2011. Characterization of cytochrome P450 monooxygenase CYP154H1 from the thermophilic soil bacterium *Thermobifida fusca*. Appl. Microbiol. Biotechnol. 89, 1475–1485. doi:http://dx.doi.org/10.1007/s00253-010-2965-9.
Sono, M., Roach, M.P., Coulter, E.D., Dawson, J.H., 1996. Heme-containing oxygenases. Chem. Rev. 96, 2841–2888.
Strushkevich, N., Usanov, S.A., Plotnikov, A.N., Jones, G., Park, H.-W., 2008. Structural analysis of CYP2R1 in complex with vitamin D3. J. Mol. Biol. 380, 95–106. doi:http://dx.doi.org/10.1016/j.jmb.2008.03.065.
Sugimoto, H., Shinkyo, R., Hayashi, K., Yoneda, S., Yamada, M., Kamakura, M., Ikushiro, S., Shiro, Y., Sakaki, T., 2008. Crystal structure of CYP105A1 (P450SU-1) in complex with 1α,25-dihydroxyvitamin D3. Biochemistry 47, 4017–4027. doi:http://dx.doi.org/10.1021/bi7023767.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Urlacher, V.B., Eiben, S., 2006. Cytochrome P450 monooxygenases: perspectives for synthetic application. Trends Biotechnol. 24, 324–330. doi:http://dx.doi.org/10.1016/j.tibtech.2006.05.002.
Werck-Reichhart, D., Feyereisen, R., 2000. Cytochromes P 450: a success story. Genome Biol. 1 doi:http://dx.doi.org/10.1186/gb-2000-1-6-reviews3003 REVIEWS3003.
Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–W410. doi:http://dx.doi.org/10.1093/nar/gkm290.
Yasutake, Y., Fujii, Y., Cheon, W.K., Arisawa, A., Tamura, T., 2009. Crystallization and preliminary X-ray diffraction studies of vitamin D3 hydroxylase, a novel cytochrome P450 isolated from *Pseudonocardia autotrophica*. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 65, 372–375. doi:http://dx.doi.org/10.1107/S1744309109007829.
Yasutake, Y., Fujii, Y., Nishioka, T., Cheon, W.-K., Arisawa, A., Tamura, T., 2010. Structural evidence for enhancement of sequential vitamin D3 hydroxylation activities by directed evolution of cytochrome P450 vitamin D3 hydroxylase. J. Biol. Chem. 285 (31), 193–201. doi:http://dx.doi.org/10.1074/jbc.M110.147009.
Zhang, Y., Kulye, M., Yang, F., Xiao, L., Zhang, Y., Zeng, H., Wang, J., Liu, Z., 2011. Identification, characterization, and expression of a novel P450 gene encoding CYP6AE25 from the Asian corn borer, *Ostrinia furnacalis*. J. Insect Sci. 11, 37. doi:http://dx.doi.org/10.1673/031.011.0137.
Zhu, G.-D., Okamura, W.H., 1995. Synthesis of vitamin D (calciferol). Chem. Rev. 95, 1877–1952. doi:http://dx.doi.org/10.1021/cr00038a007.
Zurek, J., Foloppe, N., Harvey, J.N., Mulholland, A.J., 2006. Mechanisms of reaction in cytochrome P450: hydroxylation of camphor in P450cam. Org. Biomol. Chem. 4, 3931–3937. doi:http://dx.doi.org/10.1039/b611653a.