African Journal of Biotechnology Vol. 11(75), pp. 13947-13953, 18 September, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.803
ISSN 1684-5315 ©2012 Academic Journals

Full Length Research Paper

Predicting the catalytic sites of isopenicillin N synthase (IPNS) related non-haem iron-dependent oxygenases and oxidases (NHIDOX) through a structural superimposition and molecular docking approach

Hong Soon Chin[1,2]*, Boon Hoor Lee[2], Yen Sen Liang[2], Yong Seng Wong[2], Shu Keong Yap[2], Mun Yee Chan[2], Khang Wen Goh[2] and Lisa Gaik Ai Ong[2]

[1]Cancer Research Initiative Foundation, Drug Discovery Laboratory, Subang Jaya, Malaysia.
[2]Faculty of Science, University Tunku Abdul Rahman, Kampar, Malaysia.

Accepted 22 June, 2012

Isopenicillin N synthase (IPNS) related Non-haem iron-dependent oxygenases and oxidases (NHIDOX) demonstrated a striking structural conservativeness, even with low protein sequence homology. It is evident that these enzymes have an architecturally similar catalytic centre with active ligands lining the reactive pocket. Deacetoxycephalosporin C synthase (DAOCS), isopenicillin N synthase (IPNS), deacetylcephalosporin C synthase (DACS), clavaminate synthase 1 and 2 (CAS1 and 2) are important bacterial enzymes that catalyze the formation of β-lactam antibiotics belonging to this enzyme family. Most plant enzyme members within this subfamily namely flavonol synthase (FLS), leucoanthocyanidin dioxygenase (LDOX), anthocyanidin synthase (ANS), 1-aminocyclopropane-1-carboxylic acid oxidase (ACCO), gibberellin 20-oxidase ($G_{20}O$), desacetoxyvindoline-4-hydroxylase (D4H), flavanone 3β-hydroxylase (F3H), and hyoscyamine 6β-hydroxylase (H6H) are involved in catalyzing the biosyntheses of plant secondary metabolites. With the advancement of protein structural analysis software, it is possible to predict the catalytic sites of protein that shared a structural resemblance. By exploiting the superimposition model of DAOCS-IPNS, DAOCS-IPNS-CAS, $G_{20}O$-LDOX, FLS-LDOX, ACCO-LDOX, D4H-LDOX, F3H-LDOX and H6H-LDOX model; a computational protocol for predicting the catalytic sites of proteins is now made available. This study shows that without the crystallized or nuclear magnetic resonance (NMR) structures of most NHIDOX enzyme, the plausible catalytic sites of protein can be forecasted using this structural bioinformatics approach.

Key words: Enzyme, catalytic sites, isopenicillin N synthase, ligands.

INTRODUCTION

Penicillins and cephalosporins are widely utilized β-lactam chemotherapeutics. However, the advantage of cephalosporins over penicillins is their resistance to penicillin β-lactamases. Therefore, cephalosporins were developed to defeat the disadvantages associated with penicillins. Isopenicillin N synthase (IPNS) catalyses the four-electron oxidation of L-d-(a-aminoadipyl)-L-cysteinyl-D-valine (ACV) to form isopenicillin N (Roach et al., 1997). Deacetoxycephalosporin C synthase (DAOCS) expanded the five-membered thiazolidine ring of the penicillin nucleus to form the six-membered dihydrothiazine ring of the cephalosporin nucleus. Deacetoxycephalosporin C (DAOC) is then hydroxylated to form deacetyl-cephalosporin C by deacetoxyvindoline C synthase (DACS) or DAOC hydroxylase (Baldwin et al., 1992). Clavaminate synthase (CAS) catalyzes the conversion of proclavaminic acid to clavaminic acid (Paradkar and

*Corresponding author. Email: hshschin@gmail.com. Tel: +60173656486. Fax: +6054667449.

Jensen, 1995; Baldwin and Abraham, 1988). The secondary structures of IPNS consist of 10 helices and 16 β-strands. 8 of these β-strands are then folded into a jelly-roll motif. Crystallography study revealed that the active binding sites of IPNS are buried within this jelly-roll motif and lined by hydrophobic residues that probably function in isolating the highly reactive intermediates from the external environment. A similar structural architecture is observed in DAOCS. Five highly conserved residues, corresponding to HIS214, ASP216 and HIS270 in *Aspergillus nidulans* IPNS (aIPNS) were assigned for Fe-binding, while ARG279 and SER281 were assigned for co-substrate (2-oxoglutarate, 2OG) binding. IPNS, DAOCS, DACS and other non-haem iron-dependent oxygenases and oxidases (NHIDOX) enzymes shared a conserved structural framework in the catalytic centre. Therefore, prediction model can be developed based on the structural similarity among these enzymes. Interestingly, CAS which is categorized under the Taurine catabolism dioxygenase TauD family (Pfam classification, PF02668), also showed a similar structural motif, possibly due to the Fe- and 2OG-binding properties of CASs. This observation suggested the likelihood of expanding this analysis pipeline to other iron-binding enzymes which is not categorized under the NHIDOX or 2OG-Fe (II) oxygenase superfamily (Pfam classification, PF03171).

Most plant enzyme members within NHIDOX family are involved in the catalyzing of plant secondary metabolites. These enzymes shared a well-conserved Fe-binding pocket and a reasonable similar 2OG-binding pocket. Flavanone 3β-hydroxylase (FLS), leucoanthocyanidin dioxyganase (LDOX) and anthocyanidin synthase (ANS) are involved in the biosynthesis of flavanoids.

Flavonoids have a wide array of physiological functions in plants. Most plants synthesize derivatives of one or more of the three major flavonols, which are quercetin, kaempferol or myricetin. The ratio of these flavonols varies substantially among different tissues and can be altered in response to environmental cues and serve as signaling molecules (Pelletier et al., 1997; Winkel-Shirley, 2002).

Thus, the studies of these enzymes are crucial with the potential of increasing crop productivity through controlling the plant hormone. 1-Aminocyclopropane-1-carboxylate oxidase (ACCO) catalyses the last step in the biosynthesis of ethylene (Adams and Yang, 1979). Ramassamy et al. (1998) showed that ACCO is the enzyme that catalyses the last step in the biosynthesis of the ethylene in plant. Both 1-aminocyclopropane-1-carboxylate synthase and ACCO exist as multi-gene families which are active under different physiological conditions; this implies that there is a need for the regulation of ethylene, which is due to its multiple roles played in the plant development (Zhang et al., 2004).

$G_{20}O$ is an oxidoreductase which plays a key role in gibberellin biosynthetic process. Gibberellin is a plant hormone involved in plant developmental process, such as fruits senescence and sex determination (Lange et al., 1994; Phillips et al., 1995). Desacetoxyvindoline-4-hydroxylase (D4H) is involved in the biosynthesis of vindolines (Vazquez-Flota et al., 1997). Vindoline is subsequently reacted with catharanthine to produce the cytotoxic dimeric alkaloids vinblastine and vincristine (Benoit et al., 1999). Hyoscyamine 6β-hydroxylase (H6H) is involved in synthesizing the plant alkanoid scopolamine. Scopolamine is a tropane alkaloid drug and it is part of the secondary metabolites of plants (Hakkinen et al., 2005). Flavanone 3β-hydroxylase (F3H) mainly catalyzes the flavonoid biosynthesis which is involved in the fruit ripening process. It is also involved in the formation of red-colored anthocyanins which is used as a target for RNase P-mediated gene disruption in maize cells (Rangarajan et al., 2004). In the analyses of primary amino acid sequences, secondary and tertiary structures of NHIDOX have shown that these enzymes shared low sequence homologies (~20%) but possess a remarkably conserved domain that fold into a jelly-roll motifs. As these enzymes are likely to function by means of associated or comparable mechanisms with the bacterial NHIDOX enzymes, the implication made from the representative bacterial enzyme may allow accurate predictions of catalytic sites for the plant enzymes.

With the rapid progress of recombinant DNA technology, particularly, site-directed mutagenesis and gene shuffling, has impelled the likelihood of re-engineering a protein with required properties such as improved thermo-stability, catalytic prowess or even alteration in substrate/co-substrate specificity.

Biochemical testing and analysis by our previous studies has evidently showed that it is possible to employ protein structures as a framework for redesigning its properties (Chin et al., 2001; Chua et al., 2008). By exploiting the comparative computational methods such as sequence-based analysis of protein structures, molecular superimposition and substrate docking, we were able to predict the structural relationship that is indigenous to NHIDOX and possibly expand this simple and economical bio-computational approach to other enzyme superfamily.

METHODS

Data retrieving

The amino acid sequences of known protein were retrieved from SwissProt Database. The amino acid sequence similarities of these enzymes were calculated using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/). The structure coordinates were downloaded from the Protein data bank (PDB). Protein superfamily classification was retrieved from the Pfam protein families' database (http://pfam.sanger.ac.uk/), National Center for Biotechnology Information Protein database (http://www.ncbi.nlm.nih.gov/protein/), SCOP Classification database (http://supfam.org/SUPERFAMILY/) and InterPro Protein database (http://www.ebi.ac.uk/interpro/IEntry?ac=IPR005123).

Tertiary structure analysis and protein simulation

SWISS-MODEL program (http://www.expasy.ch/swissmod/SWISS-MODEL.html) was used to generate tertiary structure information of proteins with yet undetermined structures (Guex and Peitsch, 1997). The simulated structures were evaluated using the PROCHECK program (Laskowski et al., 1993). Manipulation and viewing of 3D structures were performed using the Swiss_Pdb viewer program version 4.0 (Guex and Peitsch, 1997). Protein structures were superimposed and computed for plausible substrate or co-factors binding sites using the Swiss_Pdb viewer program. Superimpositions were carried out using the "Magic Fit" feature. Proteins with 3D structure in PDB were selected as reference layer. Subsequently, the superimposition models were refined using the "Explore Domain Alternative Fits" features and alternate alignments for all residues were performed. Lastly, the "Display Radius" feature was selected to forecast the amino acids or an element, which is distributed in close proximity to the authentic and virtual core.

Enzymes in lactam synthesis pathway

The first crystal structure of IPNS was obtained from recombinant aIPNS at 2.5 Å, with the active sites (HIS214, ASP216, HIS270 and GLU330) complexed with manganese. Due to the instability of iron and ACV under aerobic conditions, the crystal structure of aIPNS complexed with iron and ACV can only be obtained under anaerobic conditions. Later, crystal structure of aIPNS complexed with iron and ACV was resolved at a resolution of 1.3 Å. (Roach et al., 1997). The successful crystallization of *S. clavuligerus* DAOCS (scDAOCS, apo-enzyme) and aIPNS has enabled the clarification of the spatial organization and function of substrate- and co-factors binding sites of these proteins (Valegard et al., 1998). DAOCS and IPNS catalyze different reactions and share only 14% amino acid sequence identity. However, these enzymes possess an apparent similarity in terms of secondary and tertiary structures. Superimposition analysis of scDAOCS and aIPNS revealed that their structures aligned well within 1.50 Å. Consequently, we are able to predict the plausible catalytic residues, located in close proximity to the catalytic center of scDAOCS, using a bio-computational approach. In 2004, the X-ray structure of scDAOCS complexes with various penicillin analogues were made available by Valegard et al. (2004). It has enabled us to re-evaluate and calculate the precision of our models for prime and co-substrate binding sites prediction. IPNS, DAOCS, DACS, CAS1 and CAS2 shared an extremely low similarity in terms of amino acid sequence, as low as 4% between IPNS and CAS1 from *S. clavuligerus* source. Interestingly, they still share a well-conserved facial triad motif with consensus HisXAsp/Glu_XHis motif and ArgXSer (RXS) motif despite this low similarity. IPNS, DAOCS and DACS belong to iron/ascorbate oxido-reductase family but CAS was classified under Taurine catabolism dioxygenase TauD family. TauD is also referred as the group II of the αKG-dioxygenase family. A HisXAsp/Glu_X_{23-26}Thr/Ser$X_{114-183}$Arg motif is found in TauD, alkyl sulfate/RKG dioxygenase (AtsK), 2,4-D/RKG dioxygenase (TfdA) and CASs.

Using DAOCS-IPNS, DAOCS-CAS1 and IPNS-CAS1 superimposition model as learning template, the plausible substrate and co-factors binding sites of protein without readily accessible structures can be forecasted. We were able to dock in various virtual substrates into the plausible catalytic pocket of computational simulated structures. The virtual substrates of DOACS were allocated in extremely close proximity when examined, using the DAOCS-IPNS superimposition model. Similar outcome were obtained using the DAOCS-CAS1 and IPNS-CAS1 model. Also, the genuine and virtual substrate can be aligned accordingly in the catalytic centre of these proteins. The HisXAsp/Glu_XHis motif and RXS motif of DAOCS, IPNS and CAS1 can be spatially located accurately using this structural superimposition and molecular docking approach. The Fe-binding sites of scDAOCS, namely HIS183, ASP185 and HIS243 superimposed exactly with the Fe-binding sites of aIPNS, namely HIS214, ASP216 and HIS270. The 2OG-binding site of DAOCS, namely SER260 and ARG258 superimposed accurately with ARG279 and SER281 of IPNS, even though IPNS does not utilize this co-factor for catalysis. The Fe-binding site of CAS1 can also be allocated clearly using the DAOCS-CAS1 or IPNS-CAS1 superimposition model. The co-factors binding sites of these enzymes can also be identified and aligned readily, using our bio-computational protocol. The elements surrounding the virtual substrate were predicted rather accurately (up to 80% reliability) using the DAOCS-IPNS superimposition model (Chin et al., 2011). The predicted prime substrate binding sites can provide useful hints for future mutagenesis study in order to determine the functionality and biochemical properties of these enzymes.

$G_{20}O$-LDOX model

$G_{20}O$ and LDOX are both categorized NHIDOX family. Fe-binding motif (HIS246, ASP248 and HIS302) and 2OG-binding motif (TYR231, ARG312 and THR314) were found in primary sequence alignment of $G_{20}O$. Their secondary and tertiary protein structures appear comparable, despite the low homology in primary sequence (Rose et al., 2006). The conserved sequence such as co-factor binding motifs guided the proteins folding into a similar structure. *Arabidopsis thaliana* $G_{20}O$ structure was simulated using homology modeling method. ANS from *A. thaliana* (PDB ID: 1GP6) serve as the simulation template. The sequence similarity of ANS and $G_{20}O$ is 30.54%. The simulated $G_{20}O$ structure was well validated, since 100% of the amino acid residues dropped on the allowed region of Ramachandran plot. The G-factor value for the simulated structure is -0.24. A good structure was indicated by a G-factor which was larger than -0.05 (Laskowski et al., 1993; Mereghetti et al., 2008). The simulated $G_{20}O$ structure was then subjected to molecular superimposition and substrate docking procedure. The catalytic site of $G_{20}O$ was predicted using LDOX X-ray crystal structure as a reference model. The superimposition and docking were performed appropriately showing a Root Mean Squared Deviation (RMSD) value of 0.44 Å (Figure 1A). In 3 Å away from the docked Fe (II) molecule, the Fe-binding motif of $G_{20}O$, namely HIS246, ASP248 and HIS302 were identified (Figure 1B). This matched the H-X-D-X_n-H conserved sequence observed in the primary sequence alignment (Sim et al., 2003). Additionally, 2OG molecule also appeared in close proximity. This indicated that the superimposition and docking procedure were properly performed. Seo et al. (2004) showed that histidine (HIS302 and HIS246) and aspartate (ASP248) were involved in charge-charge interactions between Fe (II) molecule and protein. The co-factor, dioxygen molecule is also involved in the provision of extensive network of electrostatic interaction to optimized binding of Fe (II) molecule. At 3 Å away from the 2OG molecule, amino acid residues (TYR231, ARG312 and THR314), co-factor [Fe (II) molecules] and substrate (2R,3R)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydro-4H-chromen-4-one (DQH) appeared. The amino acid residues matched the ArgXThr (RXT) conserved sequence observed in primary sequence alignment. The ARG312 and THR314 are involved in hydrogen bonding with 2OG (Chua et al., 2008). 14 amino acids namely CYS320, PHE318, THR250, PRO249, ASP248, CYS247, HIS246, SER137, ARG126, SER131, SER153, ARG227, TRP349, and ILE225 were observed in *A. tahliana* $G_{20}O$ which resided 5 Å away from docked substrate (DQH). These residues served as potential mutagenesis candidate for enzymatic and functionally study.

FLS-LDOX model

FLSs and LDOX has a low homology similarity (30 ~ 40%). The

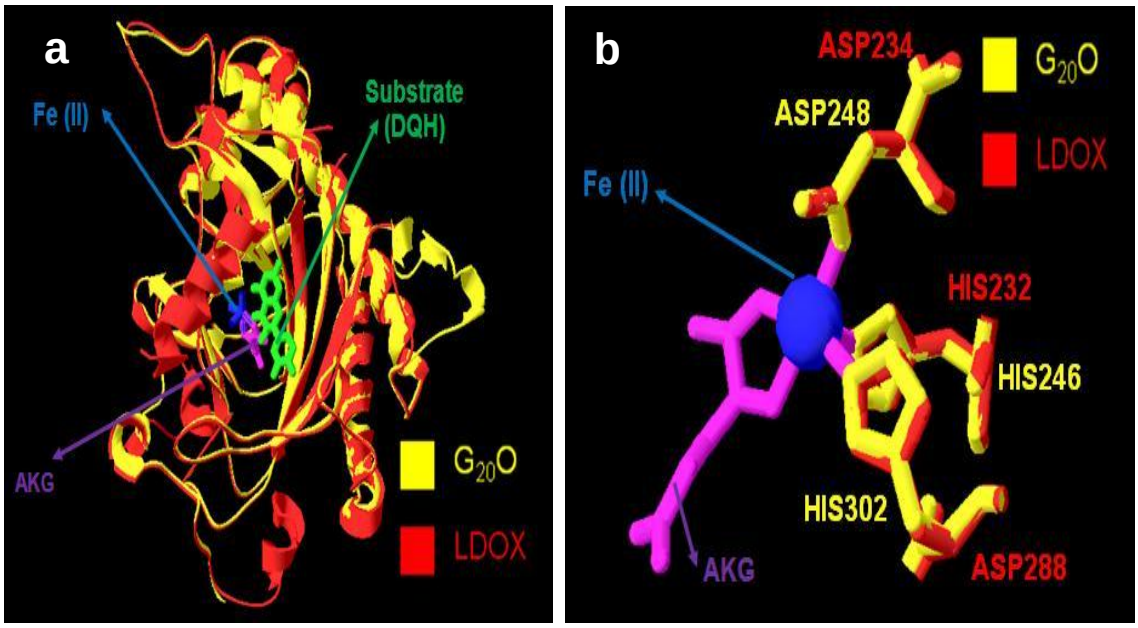

Figure 1. Superimpositions of learning models. (a) Superimposition of $G_{20}O$ (yellow) and LDOX (red) docked with DQH, Fe (II) and AKG; (b) Superimposition of $G_{20}O$ (yellow) and LDOX (red) highlighting the respective HisXAsp/Glu_XHis motif substrate (NAR) as the virtual substrate center.

percentage of similarity calculated was based on the primary sequence; therefore the sequence length will influence the result. All FLS sequences retrieved from the database were in different sequence length. Analysis through cladogram showed that FLS1 was closely related to the FLS6 as compared to other FLS isoforms. The result suggested that FLS1 have a relatively close evolutionary time to FLS6 with lesser series of duplication events. Comparing the secondary structure conformation of FLS1 and LDOX from *A. thaliana* shows that both proteins possess a similar number of α-helix (14 for both LDOX and FLS1), β-strand and β-turn. All FLS isoforms from *A. thaliana* [accession:. Q96330 (FLS 1); NP_201163 (FLS 2); NP_201164 (FLS 3); NP_680463 (FLS 4); NP_001032131 (FLS 5), and NP_680388 (FLS 6), respectively] were successfully simulated by SWISS-MODEL and validated. The amino acid residues involved in Fe (II) and 2OG can be identified using the superimpostion method, using LDOX as the reference layer. Within 3 Å from the virtual Fe residue, the H-X-D-X_n-H motif of all FLS isoform, except for FLS2, can be identified precisely. Only two amino acids involved in Fe binding can be identified for the FLS2 isoforms. The RXS motif for 2OG-binding can also be identified precisely using 2OG of LDOX as virtual co-substrate. Similarly, the RXS motif of most FLS enzymes, except for FLS2, can be identified precisely. The identification of co-factor binding site for FLS2 were not as precise, compared to that of other isoforms possibly due to the short amino acid length of FLS2 (250 amino acids). The plausible catalytic sites for FLSs are shown in Table 1. These residues are probably substrate binding site for the respective FLSs, and could potentially be subjected to mutagenesis and functionality study. This prediction used the prime substrate of LDOX (DHQ) as the virtual centre.

ACCO-ANS model

Primary sequences alignment of ACCO and ANS retrieved from different species indicate large differences between both enzymes. However, conserved motifs of Fe and 2OG are found in ACCO and ANS from different sources. The primary sequence similarity between ANS and ACCO is approximately 20 to 30%. Tertiary structure of ACCO from *Oryza sativa* (accession:. AAC05507) were generated and validated with a G-factor of -0.35. Ramachandran plot demonstrated that more than 99% residues of the model fall within the allowed region. The RMSD value obtained for superimposition of ANS and ACCO was at 1.36 Å, which indicated a proper superimposition. The conserved Fe-binding motif of ACCO can be identified within 4 Å away from the virtual Fe (ANS). The conserved 2OG-bindng motif of ACCO can be identified within 6 Å from the virtual 2OG (ANS). The Fe binding motif of ACCO, namely HIS177, ASP179, and HIS234 aligned well with the corresponding Fe-binding motif of ANS, namely HIS232, ASP234, and HIS288 (Wilmouth et al., 2002). The result from mutagenic studies strongly suggests the involvement of HIS177 and HIS234 in ligands binding at the active site of ACCO. HIS121 of ACCO suggests their involvement with either catalysis or maintaining the structure integrity of ACCO (Tayeh et al., 1999). 15 amino acids were observed in ACCO, which resided 5 Å away from docked substrate (naringenin).

D4H-LDOX, F3H-LDOX, H6H-LDOX model

D4H, F3H and H6H were selected for computational analysis. The accession number for primary sequence of D4H, F3H and H6H are AAC49826, AAA91227 and ABR15749, respectively. The 3D structures of these enzymes are highly similar to that of LDOX and ACCO. The simulated structure of D4H, F3H and H6H were generated and validated with PROCHECK program. LDOX was used as the reference layer for molecular superimposition of D4H, F3H and H6H. The RMSD value for D4H, F3H and H6H were 0.46, 0.41 and 1.4 Å, respectively. The Fe- and 2OG-binding site of D4H, F3H, and H6H can be predicted rather precisely, using this superimposition and docking procedure. The predicted Fe-binding site for D4H, F3H, and H6H are HIS193-ASP195-HIS249 (Table 2), HIS163-ASP165-HIS211, and HIS157-ASP159-HIS214, respectively.

Table 1. The amino acid residues lining the catalytic cavity of FLS 1-6 when using DQH as virtual substrate.

Distance from substrate	4 Å						5 Å					
Enzyme	**FLS1**	**FLS2**	**FLS3**	**FLS4**	**FLS5**	**FLS6**	**FLS1**	**FLS2**	**FLS3**	**FLS4**	**FLS5**	**FLS6**
Amino acids or residue in proximity to the authentic and virtual substrate	LYS105 LYS202 ASP223 ASN269 GLU295 LEU308 THR309 GLY310 ASP311 ASP312 ASN313 PRO314 PRO315	ARG176 ASP197 GLN237 ILE238 MET239 ALA240 OXT240	LYS76 ARG89 PHE105 ASP194 VAL195 ASN196 PHE264 GLU266 TYR294	LYS87 GLU98 ASP100 PHE110 LYS162 ASP183 PHE184 ASN185 HIS253 LYS255	LYS103 ASN105 HIS116 PHE118 VAL184 ARG186 HIS205 ASP207 MET208 GLY209 PHE278 VAL279 TYR307	ARG81 PHE92 LYS94 ILE107 ARG162 ASP180 PHE181 CYS182 LEU250 GLU252 PHE278	ILE112 LYS118 HIS132 PHE134 LYS202 HIS221 THR222 ASP223 LEU224 SER225 PHE293 GLU295 PHE322	LYS81 PHE88 THR92 ASN94 PHE108 VAL174 ARG176 HIS195 THR196 ASP197 ILE198 ILE199 GLN237 ILE238 MET239 ALA240 OXT240	LYS76 ILE83 ARG89 LYS89 PHE105 LEU171 LYS173 HIS192 THR193 ASP194 VAL195 ASN196 PHE264 GLU266 TYR294	LYS87 ILE94 GLU98 ASP100 ARG106 PHE110 LYS162 HIS181 THR182 ASP183 PHE184 ASN185 HIS253 LYS255	LYS103 ASN105 HIS116 PHE118 VAL184 ARG185 HIS205 SER206 ASP207 MET208 GLY209 PHE278 VAL279 ALA280 TYR307	ARG81 PHE92 LYS94 ILE107 ARG162 ASP180 PHE181 CYS182 LEU250 GLU252 PHE276 TYR284 ARG285 LYS286 ILE287 ILE288 OXT288
Number of residues	13	7	9	10	13	11	13	17	15	14	15	17

The plausible 2OG binding site for D4H, F3H, and H6H are ARG259-SER261 (Table 2), ARG231-SER233, and SER226, respectively. Unfortunately, we were unable to locate the ARG224 site for H6H using this docking approaches, which suggests the need for further optimization for the H6H-LDOX model. Likewise, the data for potential prime substrate binding sites were performed concurrently. 14, 15 and 10 amino acids were detected at 5 Å away from the virtual substrate (DQH of LDOX) which served as potential prime substrate binding site for D4H, F3H, and H6H, respectively. The plausible catalytic sites for D4H were tabulated in Table 2. 15 candidate amino acids namely MET51, LYS56, ILE60, SER62, ARG74, ILE76, THR78, HIS163, THR164, ASP165, PRO166, GLY167, PHE237, ASN239 and PHE165, which probably lined the catalytic center of F3H, were observed residing at 5 Å away from the docked substrate (DQH). 10 candidate amino acids (MET136, LEU138, HIS157, ASP159, GLY160, ASN161, THR164, LEU230, GLY232, and TYR259) probably lining the catalytic center of H3H were observed residing at 5 Å away from this virtual prime substrate (DQH).

Conclusion

Comparison of the protein sequences of NHIDOX enzymes reveals their linear relationships of having several highly conserved residues in the structure. However, analysis of their secondary and tertiary structure relationships of NHIDOX showed that these enzymes have evolved into common topological scaffolds, comprising of highly conserved anti-parallel running strands and bondings with certain residues, in order to form a jelly-roll motif structure for the active centers. NHIDOX enzymes possess functional residues that are highly conserved. These residues also involved in the modulation of enzymatic reactions. In most cases, optimal level of prediction accuracy occurs when the RMSD value of superimposed structures is less than 2 Å. Secondly, the spatial orientation of Fe- and 2OG-binding sites of reference and testing structure must be located in extreme close proximity (<1 Å). The accuracy of prediction can achieve up to a level of 80%, calculated using the DAOCS-IPNS superimposition model (Chin et al., 2011).

Biochemical analysis has demonstrated that it

Table 2. Prediction of plausible substrate and co-factor binding sites of D4H using D4H-LDOX superimposition model.

Distance from substrate or cofactor	Substrate binding (DQH)			Fe binding			Cofactor binding		
	3Å	4Å	5Å	3Å	4Å	5Å	3Å	4Å	5Å
Amino acid involved	Nil	ASN90 ASN96 GLU112 LEU174 ASP195 SER196 GLY197 ALA265	ASN90 SER94 ASN96 ARG108 GLN112 LEU174 HIS176 HIS193 SER194 ASP195 SER196 GLY197 ALA265 GLY267	HIS193 ASP195 HIS249	HIS193 ASP195 HIS249	HIS193 ASP195 HIS249	HIS176 LEU202 HIS249 SER261	HIS176 TYR178 HIS193 ASP195 LEU202 LEU210 HIS249 ARG259 SER261 ALA263	LEU174 HIS176 TYR178 THR190 HIS193 ASP195 LEU202 LEU210 HIS249 VAL251 ARG259 SER261 ALA263
Number of residues	0	8	14	3	3	3	4	10	13

is possible to exploit protein structures as a framework for re-engineering the properties of a protein.

To date (February 2012), 5411 sequences, classified under the 2OG-Fe (II) oxygenase superfamily and 4205 sequences classified under the Taurine catabolism dioxygenase TauD family, has been deposited in the Pfam database. However, there were only a total of 84 proteins (<1%) classified under these two families which have available structural data. This workflow, which incorporated the process of protein simulation, model validation, structural superimposition and spatial computation, can provide the basis for a universal, systematically structural- and functional-based identification of the plausible active sites of a protein, using an apo-enzyme or a DNA sequence from GenBank.

ACKNOWLEDGEMENT

This research was supported by grant from Fundamental Research Grant Scheme, Ministry of Higher Education Malaysia (Grant No. FRGS/1/10/ST/UTAR/03/2).

Abbreviations

ACCO, 1-Aminocyclopropane-1-carboxylate oxidase; **ACV,** L-d-(a-amonoadipyl)-L-cysteinyl-D-valine; **ANS,** anthocyanidin synthase; **CAS,** clavaminate synthase; **DACS,** deacetoxyvindoline C synthase; **DAOCS,** deacetoxycephalosporin C synthase; **D4H,** desacetoxyvindoline-4-hydroxylase; **DQH,** (2R,3R)-2-(3,4-dihydroxyphenyl)-3,5,7-trihydroxy-2,3-dihydro-4H chromen-4-one; **F3H,** flavanone 3β-hydroxylase; **FLS,** flavonol synthase; **$G_{20}O$,** gibberellins 20 oxidase; **H6H,** hyoscyamine 6β-hydroxylase; **IPNS,** isopenicillin N synthase; **AKG,** α-ketoglutarate; **LDOX,** leucoanthocyanidin dioxyganase; **2OG,** 2-oxoglutarate; **NHIDOX,** non-haem iron-dependent oxygenases and oxidases.

REFERENCES

Adams DO, Yang SF (1979). Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene. Proc. Natl. Acad. Sci. USA. 76(1):170-174.

Baldwin JE, Abraham EP (1988). The biosynthesis of penicillins and cephalosporins. Nat. Prod. Rep. 5(2): 129-145.

Baldwin JE, Goh KC, Schofield CJ (1992). Oxidation of deacetylcephalosporin C by deacetoxycephalosporin

C/deacetylcephalosporin C synthase. J. Antibiot.. (Tokyo) 45(8):1378-1381.
Benoit SP, Felipe AVF, Vincenzo DL (1999). Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. Plant Cell. 11(5):887-900.
Chin HS, Goh KW, Teo KC, Chan MY, Lee WS, Ong LGA (2011). Predicting the catalytic sites of *Streptomyces clavuligerus* deacetylcephalosporin C synthase and clavaminate synthase 2. Afr. J. Microbiol. Res. 5(21):3357-3366.
Chin HS, Sim J, Sim TS (2001). Mutation of N304 to leucine in *Streptomyces clavuligerus* deacetoxycephalosporin C synthase creates an enzyme with increased penicillin analogue conversion. Biochem. Biophys. Res. Commun. 287(2):507-513.
Chua CS, Biermann D, Goo KS, Sim TS (2008). Elucidation of active site residues of *Arabidopsis thaliana* flavonol synthase provides a molecular platform for engineering flavonols. Phytochemistry 69(1):66-75.
Guex N, Peitsch MC (1997). Swiss-Model and Swiss-Pdb Viewer: An environment for comparative protein modelling. Electrophoresis 18(15):2714-2723.
Hakkinen ST, Moyano E, Cusido RM, Palazo′n J, Pinol MT, Oksman-Caldentey KM (2005). Enhanced secretion of tropane alkaloids in *Nicotiana tabacum* hairy roots expressing heterologous hyoscyamine-6β-hydroxylase. J. Exp. Bot. 56(42):2611-2618.
Lange T, Hedden P, Graebe JE (1994). Expression cloning of a gibberellin 20-oxidase, a multifunctional enzyme involved in gibberellin biosynthesis. Proc. Natl. Acad. Sci. USA. 91(18):8552-8556.
Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26(Pt 2):283-291.
Mereghetti P, Ganadu ML, Papaleo E, Piercarlo F, Gioia LD (2008). Validation of protein models by a neural network approach. BMC Bioinformatics 9:66.
Pelletier MK, Murrell JR, Shirley BW (1997). Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis*. Further evidence for differential regulation of "early" and "late" genes. Plant Physiol. 113(4):1437-1445.
Paradkar AS, Jensen SE (1995). Functional analysis of the gene encoding the clavaminate synthase 2 isoenzyme involved in clavulanic acid biosynthesis in *Streptomyces clavuligerus*. J. Bacteriol. 177(5):1307-1314.
Phillips AL, Ward DA, Uknes S, Appleford NE, Lange T, Huttly AK, Gaskin P, Graebe JE, Hedden P (1995). Isolation and expression of three gibberellin 20-oxidase cDNA clones from *Arabidopsis*. Plant Physiol. 108(3):1049-1057.
Rangarajan S, Raj MLS, Hernandez JM, Grotewold E, Gopalan V (2004). The active site and substrate-binding mode of 1-aminocyclopropane-1-carboxylate oxidase determined by site-directed mutagenesis and comparative modeling studies. Biochem. J. 380(Pt 2):339-346.
Ramassamy S, Olmos E, Bouzayen M, Pech JC, Latche A (1998). 1-aminocyclopropane-1-carboxylate of apple fruit is periplasmic. J. Exp. Bot. 49(329):1909-1915.
Roach PL, Clinfton JC, Hensgens CMH, Shibata N, Schofield CJ, Hajdu J, Baldwin JE (1997). Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. Nature 387(6635):827-830.
Rose G, Fleming P, Banavar J, Maritan A (2006). A backbone-based theory of protein folding. Proc. Natl. Acad. Sci. USA. 103(45):16623-16633.
Seo YS, Yoo A, Jung J, Sung SK, Yang DR, Kim WT (2004). The active site and substrate binding mode of 1-aminocyclopropane-1-carboxylate oxidase determined by site-directed mutagenesis and comparative modeling studies. Biochem. J. 308(Pt 2):339-346.
Sim J, Wong E, Chin HS, Sim TS (2003). Conserved structural modules and bonding networks in isopenicillin N synthase related non-haem iron-dependent oxygenases and oxidases (NHIDOX). J. Mol. Catal. B: Enzym. 23(1):17-27.
Tayeh MA, Howe DL, Salleh HM, Sheflyan GY, Son JK, Woodard RW (1999). Kinetic and mutagenic evidence for the role of histidine residues in the *Lycopersicon esculentum* 1-aminocyclopropane-1-carboxylic acid oxidase. J. Protein Chem. 18(1):55-68.
Winkel-Shirley B (2002). Biosynthesis of flavonoids and effects of stress. Curr. Opin. Plant Biol. 5(3):218-223.
Valegard K, van Scheltinga ACT, Dubus A, Ranghino G, Öster LM, Hajdu J, Andersson I (2004). The structural basis of cephalosporin formalin in a mononuclear ferrous enzyme. Nat. Struct. Mol. Biol. 11(1):95-101.
Valegard K, van Scheltinga ACT, Lloyd MD, Hara T, Ramaswamy S, Perrakis A, Thompson A, Lee HJ, Baldwin JE, Schofield CJ, Hajdu J, Andersson I (1998). Structure of a cephalosporin synthase. Nature 394(6695): 805-809.
Vazquez-Flota F, De Carolis E, Alarco AM, De luca V (1997). Molecular cloning and characterization of desacetoxyvindoline 4-hydroxylase, a 2-oxoglutarate dependent dioxygenase involved in the biosynthesis of vindoline in *Catharanthus roseus* (L.) G. Don. Plant Mol. Biol. 34(6):935-948.
Wilmouth RC, Turnbull JJ, Welford RW, Clifton IJ, Prescott AG, Schofield CJ (2002). Structure and mechanism of anthocyanidin synthase from *Arabidopsis thaliana*. Structure 10(1):93-103.
Zhang ZH, Ren JS, Clifton lJ, Schofield CJ (2004). Crystal structure and mechanistic implications of 1-aminocyclopropane-1-carboxylic acid oxidase-The ethylene-forming enzyme. Chem. Biol. 11(10):1383-1394.