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 Chin¹,²*, Boon Hoor Lee², Yen Sen Liang², Yong Seng Wong², Shu Keong Yap², Mun Yee Chan², Khang Wen Goh² and Lisa Gaik Ai Ong²

¹Cancer Research Initiative Foundation, Drug Discovery Laboratory, Subang Jaya, Malaysia. ²Faculty 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), clavamate 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 (G₂O), 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, G₂O-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-aminoacyl)-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 deacetylcephalosporin C by deacetoxyvindoline C synthase (DACS) or DAOC hydroxylase (Baldwin et al., 1992). Clavamate synthase (CAS) catalyzes the conversion of proclavaminic acid to clavaminic acid (Paradkar and
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 dioxygenase (LDOX) and anthocyanidin synthase (ANS) are involved in the biosynthesis of flavonoids.

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).

G39O 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 alkaloid scopo-lamine. 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 superfamilies.

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/).
Tertiary structure analysis and protein simulation

SWISS-MODEL program (http://www.expasy.org/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, H1S270 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 faciial triad motif with consensus HisXAsp/Glu_XHHis motif and ArgXSer (RXS) motif despite this low similarity. IPNS, DAOCS and DACS belong to iron/ascorbate oxidoreductase family but CAS was classified under Taurine catalaldoxigenase TauD family. TauD is also referred as the group II of the αKG-dioxigenase family. A HisXAsp/Glu_XH_Thr/SerX_14_15_16_17_18Arg motif is found in TauD, alkyl sulfate/PIK glutathione (AsxK), 2,4-D/RIKG dioxygenase (TfdA) and CASs. Using DAOCS-IPNS, DAOCS-CAS1 and IPNS-CAS1 superposition 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_XHHis 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 H1S270. 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. 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.

G3O₄-LDOX model

G3O₄ and LDOX are both categorized NHDOX family. Fe-binding motif (H1S246, ASP248 and HIS302) and 2OG-binding motif (TYR231, ARG312 and THR314) were found in primary sequence alignment of G3O₄. 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 G3O₄ 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 G3O₄ is 30.54%. The simulated G3O₄ 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 G3O₄ structure was then subjected to molecular superimposition and substrate docking procedure. The catalytic site of G3O₄ 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 (Table 1A). In 3 Å away from the docked Fe (II) molecule, the Fe-binding motif of G3O₄, namely H1S246, ASP248 and HIS302 were identified (Figure 1B). This matched the H-X-D-X₄-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 (H1S302 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-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. thaliana G3O₄ 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
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 has 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 superimposition method, using LDOX as the reference layer. Within 3 Å from the virtual Fe residue, the H-X-D-X-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 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-binding 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 (Teyeh 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.

**Figure 1.** Superimpositions of learning models. (a) Superimposition of G2O (yellow) and LDOX (red) docked with DHQ, Fe (II) and AKG; (b) Superimposition of G2O (yellow) and LDOX (red) highlighting the respective HisXAsp/Glu_XHis motif substrate (NAR) as the virtual substrate center.
Table 1. The amino acid residues lining the catalytic cavity of FLS 1-6 when using DQH as virtual substrate.

<table>
<thead>
<tr>
<th>Distance from substrate</th>
<th>4 Å</th>
<th>5 Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme</td>
<td>FLS1</td>
<td>FLS2</td>
</tr>
<tr>
<td>LYS105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LYS202</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP223</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN269</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLU295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEU308</td>
<td></td>
<td></td>
</tr>
<tr>
<td>THR309</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLY310</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP311</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASN313</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO314</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRO315</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Amino acids or residue in proximity to the authentic and virtual substrate

| Number of residues | 13 | 7 | 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.

<table>
<thead>
<tr>
<th>Distance from substrate or cofactor</th>
<th>Substrate binding (DQH)</th>
<th>Fe binding</th>
<th>Cofactor binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>3Å</td>
<td>4Å</td>
<td>5Å</td>
<td>3Å</td>
</tr>
<tr>
<td>ASN90, SER94, ASN96, GLU112, LEU174, ASP195, SER196, GLY197, ALA265</td>
<td>ASN90, SER94, ASN96, GLN112, LEU174, HIS193, LEU202, HIS249, SER261, GLY267</td>
<td>HIS176, TYR178, HIS193, ASP195, SER261, ARG259, ALA263, VAL251</td>
<td>LEU174</td>
</tr>
</tbody>
</table>

Amino acid involved: Nil

Number of residues: 0 8 14 3 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-carboxylic acid oxidase; ACV, L-(d-(a-aminoadipyl)-l-cysteinyl-D-valine; ANS, anthocyanin 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; G20O, gibberellins 20 oxidase; H6H, hyoscyamine 6β-hydroxylase; IPNS, isopenicillin N synthase; AKG, α-ketoglutarate; LDOX, leucoanthocyanidin dioxygenase; 2OG, 2-oxoglutarate; NHIDOX, non-haem iron-dependent oxygenases and oxidases.

REFERENCES


