International Scholars Journals

Advanced Journal of Microbiology Research ISSN 2241-9837 Vol. 13 (5), pp. 001-010, May, 2019. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Predicting the catalytic sites of *Streptomyces clavuligerus* deacetylcephalosporin c synthase and clavaminate synthase 2

Hong Soon Chin[1]*, Khang Wen Goh[1], Kah Cheng Teo[1], Mun Yee Chan[1], Seong Wei Lee[2] and Lisa Gaik Ai Ong[1]

[1]Faculty of Science, Universiti Tunku Abdul Rahman, Kampar, Malaysia.

[2]Faculty of Agro Industry and Nature Resources, University Malaysia Kelantan, Kota Bharu, Malaysia.

Accepted 11 March, 2019

Enzymes categorized under the 2OG-Fe (II) oxygenase superfamily and Taurine catabolism dioxygenase TauD family demonstrated a striking structural conserveness even with low protein sequence homology. It is evident that these enzymes have an architecturally similar to 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. 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-CAS1 and IPNS-CAS1, a computational protocol for predicting the catalytic sites of proteins is now made available. This study shows that without the crystallized or NMR structures of DACS and CAS2, the plausible catalytic sites of protein can be forecasted using this structural bioinformatics approach.

Key words: Isopenicillin N synthase (IPNS), deacetoxycephalosporin C synthase (DAOCS), deacetylcephalosporin C synthase (DACS), clavaminate synthase (CAS), 2OG-Fe (II) oxygenase superfamily.

INTRODUCTION

Streptomyces clavuligerus is a Gram-positive filamentous bacterium which produces two major groups of β-lactam compounds. The sulfur-containing β-lactam compounds have antibiotic properties which include penicillins, cephalosporins, and cephamycins. The oxygen-containing β-lactam compounds are commonly referred as clavams which comprise clavulanic acid and several structurally related compounds. Clavulanic acid has potent β-lactamase inhibitory properties and is used clinically in combination with β-lactam antibiotics to battle infections caused by β-lactamase-producing micro-organisms (Paradkar and Jensen, 1995; Baldwin and Abraham, 1988).

In this study, the structural properties of five bacterial enzymes which involved in the synthesis of β-lactam compounds, namely IPNS, DAOCS, DACS, CAS1 and CAS2, were investigated. IPNS is known as a non-haem iron dependent oxidase which catalyses the reaction of a tripeptide, δ-(L-α-aminoadipoyl)-L-cysteinyl-D-valine (ACV) and dioxygen to form isopenicillin N (IPN) and two water molecules (Roach et al., 1997). Penicillin N is subsequently expanded to deacetoxycephalosporin C (DAOC) by DAOCS or expandase, an iron-dependent and αKG-requiring dioxygenase. DAOC is then hydroxylated to form deacetylcephalosporin C (DAC) by DACS or DAOC hydroxylase (Baldwin et al., 1992; Chin et al., 2003). In *Acremonium chrysogenum*, a single bifunctional DAOC/DACS encoded by *cef*EF gene

*Corresponding author. E-mail: chinhs@utar.edu.my. Tel: +60173656486. Fax: +6054667449.

Abbreviations: DAOCS, Deacetoxycephalosporin C synthase; **IPNS,** isopenicillin N synthase; **DACS,** deacetylcephalosporin C synthase; **CAS,** clavaminate synthase; **2OG,** 2-oxoglutarate, **αKG,** α-ketoglutarate; **ACV,** δ-(L-α- aminoadipoyl)-L-cysteinyl-D-valine; IPN, sopenicillin N; **DAOC,** deacetoxycephalosporin C; **DAC,** deacetylcephalosporin C; **PDB,** protein data bank; **Pen G,** Penicillin G.

catalyzes both reactions. The enzyme possesses both expandase and hydroxylase activities; thus, indicating its bifunctionality. However, in *S. clavuligerus*, the expansion of IPN to DAOC and hydroxylation of DAOC are catalyzed by DAOCS and DACS independently, where *cef*E gene encoding DAOCS and *cef*F gene encoding DACS (Chin et al., 2003). CAS catalyzes the conversion of proclavaminic acid (PCV) to clavaminic acid, and it exists as two isoenzymes, CAS1 and CAS2 which encoded by two separate but very similar genes which are cs1 and cs2 (Paradkar and Jensen, 1995).

Analyses of attainable primary amino acid sequences, assigned secondary structures and superimposed tertiary structures of IPNS related non-haem iron-dependent oxygenases and oxidases (designated as NHIDOX) has suggested that these enzymes having a low sequence homologies (~20%) but possess a conserved structural domains that fold into jelly-roll motifs. The secondary structures of IPNS consist of 10 helices and 16 β-strands. 8 of these β-strands fold into a jelly-roll motif. Crystallography study revealed that the active ligands of IPNS are buried within this jelly-roll motif and lined by hydrophobic residues, which possibly function in isolating the highly reactive intermediates from the environment. A similar structural architecture was observed in DAOCS (Roach et al., 1997; Sim et al., 2003). Five highly conserved residues, corresponding to H214, D216 and H270 in Aspergillus *nidulans* IPNS (aIPNS) for iron binding, R279 and S281 for substrate / co-substrate binding were demonstrated by Sim et al. (2003). These result suggested that IPNS, DAOCS, DACS and other NHIDOX enzymes shared a conserved structural framework in the catalytic centre. Hence, prediction model can be developed based on the structural similarity among these enzymes. Interesting, CAS which is categorized under the Taurine catabolism dioxygenase TauD family (Pfam classification) also showed a similar structural motif, possible due to the iron- and 2OG-binding properties of CAS. This observation suggested the possibility of the expanding this analysis pipeline to other iron-binding enzymes which is not categorize under the NHIDOX (Sim et al., 2003) or 2OG-Fe (II) oxygenase superfamily (Pfam classification).

The rapid progress of recombinant DNA technology especially site-directed mutagenesis has impelled the possibility to construct a novel protein with desired properties such as improved stability and activity or alteration in substrate specificity that opened an era of protein engineering or protein rationale designing. Biochemical analysis by Chin and Sim (2002), Chin et al. (2001, 2004) and Chua et al. (2008) has clearly showed that it is possible to use protein crystal structures as a framework for redesigning the properties of a specific protein. By exploiting the comparative biocomputational methods such as sequence-based analysis of primary, secondary and tertiary protein structures, we were able to envisage the structural relationship that is indigenous to this family of enzymes. A surprisingly comprehensive overview of the lesson learned from the studies can be acquired by focusing on specific representative member of the large family of enzyme, in this case IPNS, DAOCS and CAS1.

COMPUTATIONAL METHODS

Data retrieving

The amino acid sequences of known protein were retrieved from SwissProt Database. The accession number for IPNS, DAOCS, DACS, CAS1 and CAS2 of *S. clavuligerus* source are AAA26770, AAA26715, P42220, AAA26722 and AAA26723, respectively. The amino acid sequence similarity of these enzymes was calculated using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/). The structure coordinates were downloaded from the protein data bank (PDB). The PDB entries for IPNS, DAOCS and CAS1 are 1BLZ (IPNS from *A. nidulans* complexes with ACV, iron and nitric oxide), 1OUB (DAOCS from *S. clavuligerus* complexes with penicillin G, iron and 2OG) and 1DRT (CAS1 from *S. clavuligerus* complexes with PCV, iron and 2OG). 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 of DACS and CAS2 from *S. clavuligerus* source were assigned as MD-DACS and MD-CAS2, respectively. The final structure was 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) and RasMol 2.7.1.1 (Sayle and Milner-White, 1995). Protein structures were superimposed and computed for the 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 elements which distributed in close proximity to the authentic and imaginary core.

RESULTS AND DISCUSSION

The first crystal structure of IPNS was obtained from recombinant aIPNS at 2.5Å with the active site complexed with manganese instead of iron. The ligands that bind to manganese, namely H214, D216, H270 and E330 were suggested to be the endogenous ligands for iron binding. These results correspond exactly to the ligands identified by primary sequence analysis. Because of the instability of iron and ACV under aerobic conditions, the

Table 1. Percentage homologies and RMS values of protein pair**.**

Bacteria enzymes	DAOCS	IPNS	DACS	CAS1	CAS2
DAOCS		1.50	0.56	2.00	2.24
IPNS	14		2.15	1.60	1.59
DACS	58	19		1.34	1.39
CAS1	8	4	13		0.11
CAS2	8	6	10	80	

Table 2. Structure validation of MD-DACS and MD-CAS2.

PROCHECK program				
Simulated model	**MD-DACS**		**MD-CAS2**	
Plot statistics	**Residues**	**Percentage (%)**	**Residues**	**Percentage (%)**
Residues in most favored region	232	88.2	246	90.8
Residues in additional allowed regions	22	8.4	23	8.5
Residues in generously allowed regions	5	1.9	1	0.4
Residues in disallowed regions	4	1.5	1	0.4
G-factor*	Mean score		Mean score	
phi/ psi only	-0.25		-0.14	
All dihedral angle	-0.14		0.03	

*A G-factor of ≥ -0.5 is good.

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. The iron-binding ligands of aIPNS were verified and interactions of ACV at the iron centre were also resolved (Roach et al., 1997).

Although DAOCS and IPNS catalyze different reactions and share only 14% amino acid sequence identity (Table1), they possess obvious similarities in secondary and tertiary structures. Superimposition analysis of *S. clavuligerus* DAOCS (scDAOCS) and aIPNS revealed that their structures aligned well within 1.50 Å (Table 1). The successful crystallization of scDAOCS and aIPNS has allowed the elucidation of the spatial organization and function of substrate- and co-factors binding sites of these proteins (Roach et al, 1997; Valegard et al., 1998). Hence, in 2001, we are able to make use of computational tools to assist in inferring the candidate residues located in close proximity to the catalytic pocket of scDAOCS (Chin et al., 2001). In 2004, the X-ray structure of scDAOCS complexes with various penicillin analogues such as ampicillin and penicillin G (Pen G) were made available by Valegard et al. (2004). It has enabled us to re-evaluation our previous simulation model for predicting the substrate binding sites of scDAOCS (Chin et al., 2001).

The simulated MD-DACS structure was generated using DAOCS (PDB code 1e5iA) as modelling template. DACS and DAOCS shared a sequence identity of ~58%. The simulated MD-CAS1 structure were generated using CAS1 (PDB code1ds1A) as modelling template. CAS1 and CAS2 shared a sequence identity of ~80% (Table 1). Table 2 shows the structure statistics of MD - DACS and MD - CAS2 calculated using the PROCHECK program. A Raamachandran plot demonstrates that approximately 99% of the residues for MD-DACS and MD - CAS2 are found in the allowed region (Figure 2d, 2h, Table 2). The G-factors of these models are equivalent or great that -0.5; hence, the homology models are valid as calculated using PROCHECK program (Table 2).

IPNS, DAOCS, MD-DACS, CAS1 and MD-CAS2 shared a well-conserved iron-binding pocket and a reasonable similar 2OG-binding pocket (Figure 1b, c and d, 2c and g). Even though these protein shared a low similarity in amino acid sequence, as low as 4% between IPNS and CAS1 from *S. clavuligerus* source (Table 1), they share a well-conserved facial triad motif with consensus HisXAsp/Glu_XHis sequence (iron-binding sites) and RXS motif (2OG-binding sites).

It is suggested that IPNS, DAOCS and DACS belongs to enzyme family classified as isopenicillin N synthase related NHIDOX or described as iron/ascorbate oxido-reductase family under the Pfam accession number PF00671. These 2OG oxygenases are involved in many biosynthetic pathways including hose leading to β-lactam antibiotics, collagen, flavonoids and many modified amino

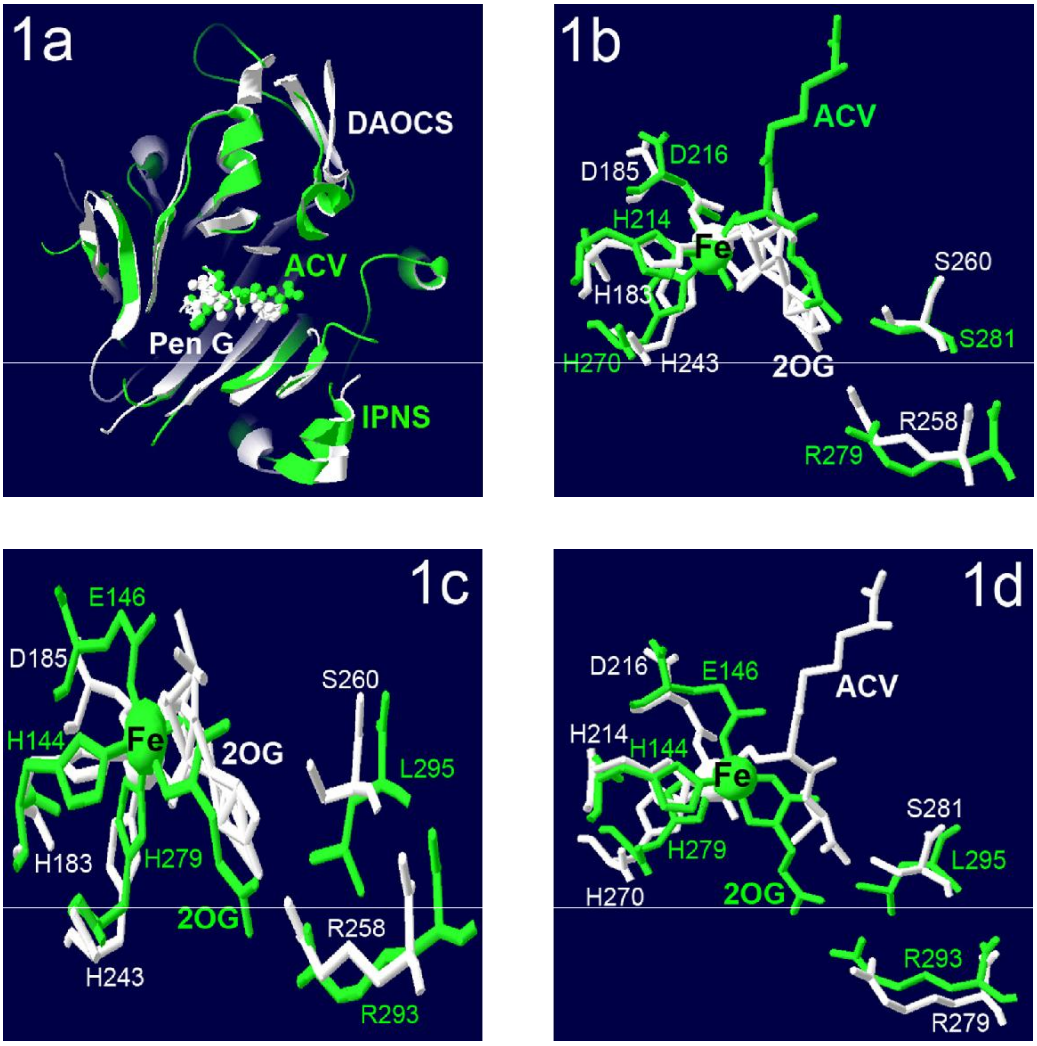

Figure 1. Superimpositions of Learning Models. 1a.Superimposition of DAOCS (white) and IPNS (green) docked with Pen G and ACV; 1b.Superimposition of DAOCS (white) and IPNS (green) highlighting the respective HisXAsp/Glu_XHis and RXS motif; 1c. Superimposition of DAOCS (white) and CAS1 (green) highlighting the respective HisXAsp/Glu_ XHis and RXS motif; 1d. Superimposition of IPNS (white) and CAS1 (green) highlighting the respective HisXAsp/Glu_XHis and RXS motif.

acids and peptide. Despite the low sequence homology (~20%), this family of enzymes showed conserved structural domains that fold into jelly-roll motifs that constitute a new structural family of enzymes. Several highly conserved residues within the NHIDOX were proposed to be crucial in stabilizing the active centre. It is also important to recognize the proposition of iron coordination in these proteins to its catalytic activity for potential exploitation. The absolute requirement of iron for the catalytic function in IPNS and DAOCS as well as the conservation of the histidine residues in the NHIDOX family strongly suggested the possible conservation of similar iron-binding sites in these enzymes (Sim et al., 2003). However, CAS1 and 2 were classified under Taurine catabolism dioxygenase TauD family with the Pfam accession of PF02668. TauD is referred to as group II of the αKG-dioxygenase family. A HisXAsp/Glu_$X_{23\text{-}26}$Thr/Ser$X_{114\text{-}183}$Arg motif is found in TauD, alkyl sulfate/RKG dioxygenase (AtsK), 2, 4-D/RKG dioxygenase (TfdA) and CASs. Structurally characterized representatives of the αKG-dioxygenase superfamilyinclude DAOCS, IPNS, CAS, proline 3-hydroxylase and anthocyanidin synthase. In general, these enzymes possess a core motif of 8 β-sheets that arranged in a jelly-roll manner and similarly arranged iron-binding ligands (Sim et al., 2003; Elkins et al., 2002). As these enzymes are likely to function via associated or analogous mechanisms, the observation made from one representative enzyme may allow accurate predictions for the other enzymes categories under the same family.

In Figure 1a, 2a, b, e, and f, we demonstrated the possibility of docking the imaginary substrates into other

protein structure using the superimposition features of Swiss_Pdb Viewer program. In Figure 1a, we showed that the authentic and imaginary substrates (pen G and ACV respectively) 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 superimposition model, whereby the authentic and imaginary substrate can be aligned according to the catalytic centre of these proteins.

In Figure 1b, c and d, we highlighted the iron - and 2OG-binding sites of DAOCS, IPNS and CAS1, respectively. By using the above mentioned procedure, we can identify the HisXAsp/Glu_XHis motif and RXS motif of DAOCS, IPNS and CAS1 straightforwardly and accurately. Since IPNS does not utilize 2OG as co-factor, ACV was docked into the catalytic pocket of the DAOCS-IPNS superimposition model for predicting the 2OG-binding sites of DAOCS. Figure 1b showed that the ferrous ion of DAOCS and IPNS overlapped in intimate proximity. 2OG from DAOCS and ACV from IPNS were also aligned closely. The iron-binding sites of scDAOCS namely H183, D185 and H243 superimposed exactly with the iron-binding sites of aIPNS namely H214, D216 and H270. The 2OG-binding site of DAOCS namely S260 and R258 superimposed accurately with R279 and S281 of IPNS, even though IPNS does not utilize this co-factor for catalysis. The iron-binding site of CAS1 can also be allocated straightforwardly using the DAOCS-CAS1 or IPNS-CAS1 superimposition model as illustrated in Figure 1c and d, respectively. Though CAS1 were classified under the Taurine catabolism dioxygenase TauD family (pfam classification) or Clavaminate synthase-like superfamily/Clavaminate synthase family (SCOP classification) , while DAOCS, IPNS and DACS that were classified under 2OG-Fe (II) oxygenase superfamily (pfam classification) or Clavaminate synthase-like superfamily / Penicillin synthase-like family (SCOP classification), the co-factors binding sites of these enzymes can be identify and aligned readily using our prediction pipeline. Our preliminary structural analysis has projected that most iron-binding oxidoreductase might shared a conserved iron-binding pocket (unpublished data).

Owing to the fact that CAS1 does no comprise a RXS motif in the primary amino acid sequence, some variation in the residues coordination was observed for the case of CASs. We cannot identify the RXS motif in CASs via structural nor primary sequence analysis. Nevertheless, the R293 of CAS1 aligned well with R258 of DAOCS and R279 of IPNS, respectively. L295 of CAS1 superimposed with the S260 of DAOCS and S281 of IPNS (Figure 1c and d). R258, R279, S260 and S281 are elements of the RXS motif. The RXL motif might assist in stabilizing the coordination of 2OG in CAS; however, the rationale for these observations yet to be proven. Similar result was also observed in MD-CAS2 which also does not pose a legitimate RXS motif in the primary sequence.

The plausible substrate binding sites of DAOCS were predicted using the DAOCS-IPNS superimposition model with ACV as the imaginary substrate. The tangible amino acids of DAOCS reside in close proximity to the authentic substrate (Pen G) were also tabulated to calculate the accuracy and reliability of the prediction. The accuracy and reliability of our prediction was calculated based on the DAOCS-IPNS and DAOCS-CAS1 superimposition model. The measurement of the accuracy of prediction is the ratio percentage of the number of elements in the intersection of Set A (actual elements reside at the circumference of authentic substrate or co-factors) and Set B (predicted elements reside at the circumference of imaginary substrate or co-factors) with the total number of elements in the union of Set A and B. The reliability of prediction is representing the ratio percentage of the number of success prediction elements with the number of total prediction. The percentage of matching is corresponding to the proportion of success prediction elements with the number of total actual elements.

We can predict the elements surrounding the imaginary substrate rather accurately (up to 80% reliability) using the DAOCS-IPNS superimposition model. For example, within 4Å radius of Pen G, 17 elements of DAOCS were identify, while 19 elements of DOACS were identify when ACV were docked into the catalytic centre of DAOCS. The accuracy and reliability of the prediction (at 4Å radius) were estimated at the level of 71% and 79%, respectively (Table 3). The iron-binding sites of DAOCS can be predicted fairly precisely using the DAOCS-IPNS and DAOCS-CAS1 superimposition model with an accuracy of 70 to 100%. However, the accuracy of 2OG-binding sites were relatively lower (~50%) but a reliability of ~60% can be achieved using our superimposition workflow (Table 3).

Using DAOCS-IPNS, DAOCS-CAS1 and IPNS-CAS1 superimposition model as learning template, we can forecast the plausible substrate and co-factors binding sites of protein without available 3D structures in the PDB. The protein sequences of DACS and CAS2 from *S. clavuligerus* source were retrieved. To date, no X-ray or NMR structures for DACS and CAS2 were deposited into PBD. The 3D structures of these proteins were generated via SWISS-MODEL program. We were able to dock in various imaginary substrates into the plausible catalytic pocket of MD-DACS and MD-CAS2 (Figure 2a, 2b, 2e and 2f). The possible iron- and 2OG-binding sites of MD-DACS and MD-CAS2 can be located relatively precisely as shown in Figure 2c and g, respectively. The predicted substrate-, iron- and 2OG-binding sites of MD-DACS and MD-CAS2 were tabulated in Table 4 and 5, respectively. The list of predicted sites can provide useful hints for future mutagenesis test in order to determine the functionality of these enzymes or to improve the catalysis prowess of the enzymes.

The accuracy of prediction is influence by the quality

Table 3. Calculation of accuracy and reliability of prediction.

	Substrate binding						Fe binding			2OG binding		
Distance from substrate or co-factor	**2 Å**		**3 Å**		**4 Å**		**4 Å**			**4 Å**		
Substrate or co-factor	**Pen G**	**ACV**	**Pen G**	**ACV**	**Pen G**	**ACV**	**Fe (DAOCS)**	**Fe (IPNS)**	**Fe (CAS1)**	**2OG (DAOCS)**	**ACV (IPNS)**	**2OG (CAS1)**
Amino acids or residue of DAOCS reside in proximity to the authentic and imaginary substrate or co-factor	R160 R162 2OG	L158 R160 2OG	M73 R160 R162 M180 Fe 2OG	M73 L158 R160 R162 S260 V262 R266 N304 Fe 2OG	M73 S102 L158 R160 R162 M180 I192 L204 V262 F264 I305 Fe 2OG	M73 R74 L158 R160 R162 F164 M180 I192 V262 F264 R266 N304 I305 Fe 2OG	M180 I305 Pen G 2OG	Pen G 2OG	M180 I305 Pen G 2OG	R162 F164 M180 L204 V245 V262 F264 I305 Fe Pen G	M73 R74 L158 R160 R162 F164 M180 I192 V262 F264 R266 N304 I305 Fe 2OG	M180 Q194 L204 V245 V262 Fe Pen G
Fe-binding sites	Nil	Nil	Nil	Nil	H183 D185 H243	H183 D185	H183 D185 H243	H183 D185 H243	H183 D185 H243	H183 D185 H243	H183 D185	H183
2OG-binding sites	Nil	Nil	Nil	Nil	S260	R258 S260	Nil	Nil	Nil	R258 S260	R258 S260	R258
Number of residues	**3**	3	**6**	10	**17**	19	**7**	5	7	**15**	19	9
Accuracy of Prediction	50%		45.5%		71.4%		N/A	71.4%	100%	N/A	47.8%	50%
Reliability of Prediction	66%		50%		78.9%		N/A	100%	100%	N/A	57.9%	88.9%
Matching Elements	2		5		15		N/A	5	7	N/A	11	8
Percentage of Matching	66%		83.3%		88.2%		N/A	71.4%	100%	N/A	73.3%	53.3%

of 3D structural fitting or root mean squared deviation (RMS) between 2 superimposed protein structures. When the RMS value of the super-imposition model is small the prediction will be more accurate. In most cases, optimal level of prediction accuracy occurs when the RMS value

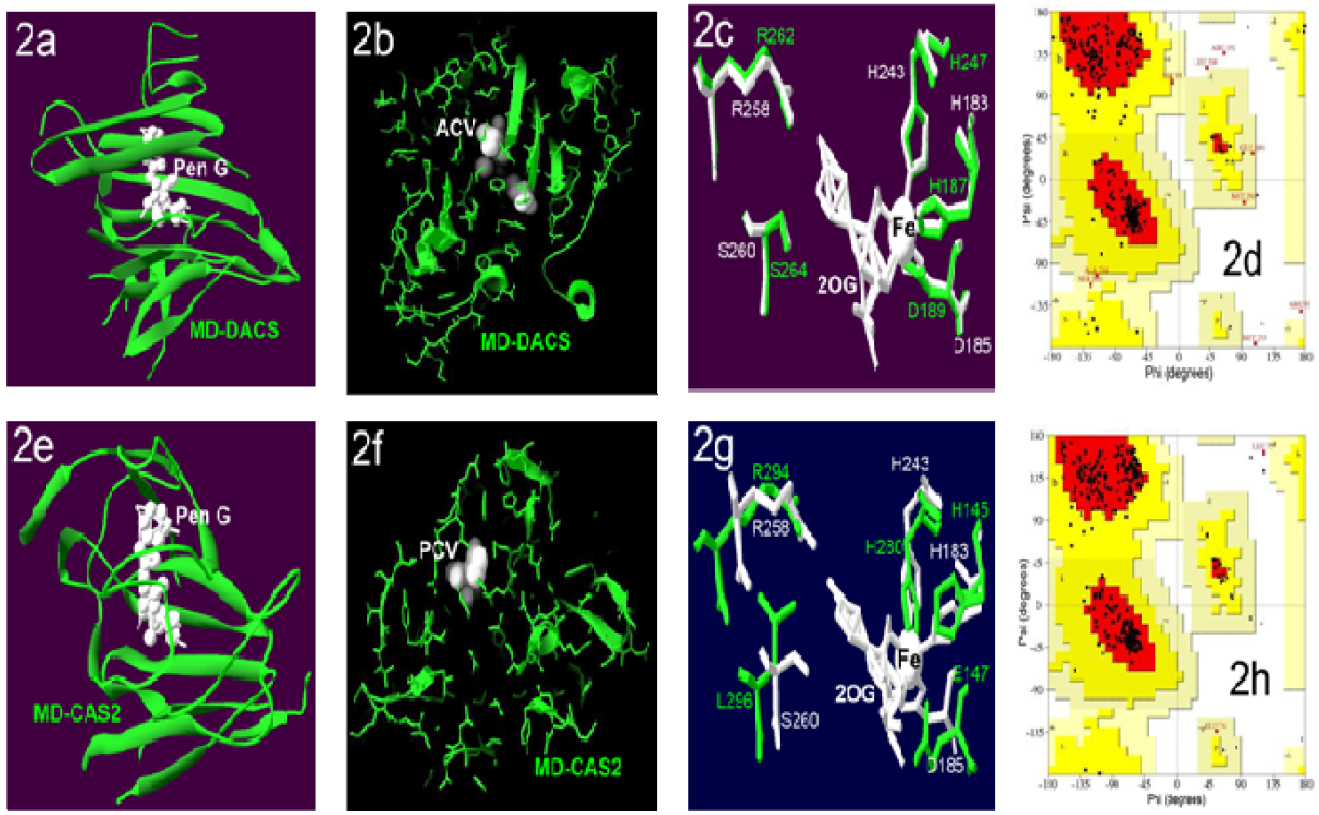

Figure 2. Prediction of catalytic center for MD-DACS and MD-CAS2. 2a. MD-DACS docked with Pen G; 2b. MD-DACS docked with ACV; 2c. Predicted iron- and 2OG-binding sites of MD-DACS (green) exploiting DAOCS (white) as modeling template; 2d. Ramachandran Plot for MD-DACS calculated via PROCHECK Program; 2e. MD-CAS2 docked with Pen G; 2f: MD-CAS2 docked with PCV; 2g. Predicted iron- and 2OG-binding sites of MD-CAS2 (green) exploiting DAOCS (white) as modeling template; 2h. Ramachandran Plot for MD-CAS2 calculated via PROCHECK Program.

is lesser than 2Å. It also showed that the prediction is more accurate when the spatial dimension of the docked element is small. For example, when iron is applied as the imaginary centre, the accuracy and reliability for predicting the iron-binding sites of DAOCS are greater than 70%. With larger spatial dimension such as ACV or Pen G, the accuracy and reliability of prediction will decrease (Table 3). However, these plausible substrate bindings still served as a useful guideline for mutagenesis purpose. Surprisingly, the homology of primary sequence has minimal effect on the outcome of prediction, with 4% homology of primary sequence (Table 1); it does not influence the accuracy and reliability of prediction.

Biochemical analysis has clearly showed that it is possible to use protein crystal structures as a framework for redesigning protein properties. It is also evident that by combining a clear scientific reasoning and accurate structural information, a

Table 4. Prediction of plausible substrate and co-factors binding sites of MD-DACS using DAOCS/IPNS/CAS1 and MD-DACS Superimposition model.

Prediction	**Substrate binding**			**Fe binding**			**2OG binding**	
Distance from substrate or co-factor	**2 Å**	**3 Å**	**4 Å**		**4 Å**		**3 Å**	**4 Å**
Imaginary substrate or co-factor	Pen G	Pen G	Pen G	Fe (DAOCS)	Fe (IPNS)	Fe (CAS1)	2OG (DAOCS)	2OG (CAS1)
Amino acids of MD-DACS reside in proximity to the imaginary substrate or co-factor	R164 R166	M74 S103 R164 R166 M184 I196	M74 S103 R164 R166 M184 I196 L208 M229 V249 V266 F268 T308 M309	Nil	Nil	M184 M309	R166	R166 F168 M184 Q198 L208 V249 V266
Plausible iron-binding sites	Nil	Nil	H187 D189 H247	H187 D189 H247	H187 D189 H247	H187 H247	Nil	H187
Plausible 2OG-binding sites	Nil	Nil	S264	Nil	Nil	Nil	R262 S264	R262 S264
Number of residues	2	6	17	3	3	4	3	10

useful advance could be made in tailoring the function and characteristic of a protein. It is envisaged that this approach might be applicable to other proteins superfamily which share conserved catalytic regions. To date, 5362 sequences classified under the 2OG-Fe (II) oxygenase superfamily and 2321 sequences classified under the Taurine catabolism dioxygenase TauD family has been deposited in the Pfam database. Only 132 proteins classified under these two families have structural data deposited in PDB till now. Structural templates of sites with know function can be employed as dataset or learning model against which structural templates of sites with unknown function. In perspective, this workflow which incorporated the process of protein simulation, model validation, structural superimposition and spatial computation can provide the basis for systematically and structural -based identification of the plausible active sites of an apo-enzyme, structurally resolved enzymes with no substrate or co-substrate complexes in the catalytic pocket, or even enzymes without any crystal or NMR structures information deposited.

Table 5. Prediction of plausible substrate and co-factor binding sites of MD-CAS2 using DAOCS/IPNS/CAS1 and MD-CAS2 Superimposition model.

Prediction	Substrate binding			Fe binding			2OG binding	
Distance from substrate or co-factor	3 Å	4 Å	5 Å		4 Å		4 Å	
Imaginary substrate or co-factor	PCV	PCV	PCV	Fe (DAOCS)	Fe (IPNS)	Fe (CAS1)	ACV (IPNS)	2OG (CAS1)
Amino acids of MD-CAS2 reside in proximity to the imaginary substrate or co-factor	L133 E147 D203 R298	L115 L133 S134 S135 M148 D203 R298 Y300	L115 R116 L133 S134 S135 L142 M148 Y150 D203 A205 F206 R298 Y300	I273	I273	Nil	Y112 L115 R116 H122 V124 S135 L142 V158 L160 R298 Y300	L142 L160 T173 R282 R298
Plausible iron-binding sites	Nil	H145 E147	H145 E147	H145 E147 H280	H145 E147 H280	H145 E147 H280	H145 E147 H280	H145 E147 H280
Plausible 2OG-binding sites	Nil	Nil	Nil	Nil	Nil	Nil	R294 L296	R294 L296
Number of residues	4	10	15	4	4	3	16	10

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).

REFERENCES

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

Baldwin JE, Goh KC, Schofield CJ (1992). Oxidation of deacetylcephalosporin C by deacetoxycephalosporin C/deacetylcephalosporin C synthase. J. Antibiot., 45:1378-1381.

Chin HS, Goo KS, Sim TS (2004). A complete library of amino acid alterations at N304 in *Streptomyces clavuligerus* deacetoxycephalosporin C synthase elucidates the basis for enhanced penicillin analogue conversion. Appl. Environ. Microbiol., 70:607-609.

Chin HS, Sim J, Seah KI, Sim TS (2003). Deacetoxycephalosporin C synthase isozymes exhibit diverse catalytic activity and substrate specificity. FEMS Microbiol. Lett., 218:251-257.

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., 287:507-513.

Chin HS, Sim TS (2002). C-terminus modification of *Streptomyces clavuligerus* deacetoxycephalosporin C synthase improves catalysis with an expanded substrate specificity. Biochem. Biophys. Res. Commun., 295:55-61.

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. Phytochem., 69:66-75.
Elkins JM, Ryle MJ, Clifton IJ, Dunning Hotopp JC, Lloyd JS, Burzlaff NI, Baldwin JE, Hausinger RP, Roach PL (2002). X-ray crystal structure of *Escherichia coli* taurine/α-ketoglutarate dioxygenase complexed to ferrous iron and substrates. Biochem., 41:5185-5192.
Guex N, Peitsch MC (1997). Swiss-Model and Swiss-Pdb Viewer: An environment for comparative protein modelling. Electroph., 18:2714-2723.
Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst., 26:283-291.
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., 177:1307-1314.
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, 387:827-830.
Sayle RA, Milner-White EJ (1995). RASMOL: biomolecular graphics for all. Trends Biochem. Sci., 20:374-376.
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., 23:17-27.
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., 11:95-101.
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, 394:805-809.
http://www.ebi.ac.uk/Tools/clustalw2/http://pfam.sanger.ac.uk/
http://www.ncbi.nlm.nih.gov/protein/
http://supfam.org/SUPERFAMILY/
http://www.expasy.ch/swissmod/SWISS-MODEL.html