Structural Assignment of a Type II PHA Synthase and an Insight Into Its Catalytic Mechanism Using Human Gastric Lipase as the Modeling Template.

Nurul Bahiyah Ahmad Khairudin[1,2], Mohd Razip Samian[1,3], Nazalan Najimudin[1,3], Habibah A Wahab[1,2]

[1]*Laboratory of Biocrystallography and Structural Bioinformatics, Universiti Sains Malaysia, Penang, Malaysia*
[2]*School of Pharmaceutical Sciences. Universiti Sains Malaysia*
[3]*School of Biological Sciences, Universiti Sains Malaysia*
Email : habibahw@usm.my

ABSTRACT: A three dimensional (3D) model for the catalytic region of Type II *Pseudomonas* sp. USM 4-55 PHA synthase 1 (PhaC1$_{P.sp\ USM\ 4-55}$) from residue 267 to residue 484 was developed. Sequence analysis demonstrated that PhaC1$_{P.sp\ USM\ 4-55}$ lacked homology with all known structural databases. PSI-BLAST and HMM Superfamily analyses demonstrated that this enzyme belongs to the α/β hydrolase fold family. Threading approach revealed that the most suitable template to use was the Human gastric lipase (1HLG). The superimposition of the predicted PhaC1$_{P.sp\ USM\ 4-55}$ model with the 1HLG template structure covering 86.2% of the backbone atoms showed an RMSD of 1.15 Å. The catalytic residues comprising of Cys296, Asp451, His452 and His479 were found to be conserved and were located adjacent to each other. We proposed that the catalytic mechanism involved the formation of two tetrahedral intermediates.

1 INTRODUCION

PHA synthase is the key enzyme that plays the central catalytic role in polyhydroxyalkanoic acid (PHA) production. It uses coenzymeA (CoA) thioesters of hydroxyalkanoic acids (HA) as the main substrates and catalyzes the polymerization of HAs to yield PHA with the concomitant release of CoA [1-4]. This PHA will then associate to form PHA granules. Numerous studies have been carried out on these enzymes and they are well characterized at the molecular level [3-8]. This enzyme can be distinguished into four types based on their subunit composition and substrate specificities [4,9]. The Type I PHA synthase (e.g. in *Ralstonia eutropha*) is composed of a single PhaC subunit (62-73 kDa) and preferentially utilizes short carbon chain-length HA-CoA thioesters as the main substrate. The Type II PHA synthase which also composed of one type of subunit, PhaC, contains two PhaC genes *phaC1* and *phaC2* expressing PHA synthase 1 and PHA synthase 2 respectively. They are synthesized and accumulated in various pseudomonads and incorporate preferentially medium chain length monomer [10-14]. On the other hand, Type III PHA synthase (e.g. in *Chromatium vinosum*) consists of two different subunits, PhaC and PhaE of about 40 kDa each. The substrate specificity for Type III PHA synthase is similar to that of Type I PHA synthase although some medium-chain-length 3-hydroxy fatty acids can also be incorporated[15,16]. The last type which is Type IV is also composed of two different subunit (e.g. in *Bacillus megaterium*) which are PhaC subunit and PhaR subunit with molecular mass approximately 40 kDa and 20 kDa respectively[2,17,18]. To date, there is no experimentally determined structural information regarding this enzyme. Several studies demonstrated that this enzyme possesses the α/β-hydrolase fold domain[18,19]. The predicted 3D model of Type III PHA synthase was reported using lipase as the template[20]. Sequence alignment showed that the active site Cys149 of Type III *Chromatium vinosum* PHA synthase (PhaC$_{Cv}$) aligned with the active site Ser87 of the template *Pseudomonas cepacia* lipase with 19% pair-wise sequence identity. It was also found that His286, the essential member of the lipase catalytic triad, aligned with His331, a conserved residue among the whole PHA synthases protein family. For Type I and II PHA synthase enzymes, threading models had also been developed[21,22]. Type I *Ralstonia eutropha* PHA synthase (PhaC$_{Re}$) and Type II *Pseudomonas aeruginosa* PHA synthase (PhaC$_{Pa}$) were modeled employing the structure of lipase from *Burkholderia glumae* and mouse epoxide hydrolase as the template respectively. The catalytic residues for both Type I and II were located adjacent to each other and that these residues are conserved throughout the whole PHA synthases which is in agreement with the Type III PhaC$_{Cv}$ 3D structure prediction study[20]. It was also noted that Type II PHA synthase differs from Type I and III with respect to the catalytic residues[22]. In contrast to the Type I and III PHA synthase, the proposed catalytic triad for Type II PHA synthase (represented by PhaC$_{Pa}$) differs in which the corresponding catalytic histidine that acts as the general based catalyst was found to be His453 which is located next to the catalytic aspartic acid instead of the corresponding catalytic His480 that is conserved with Type I and III PHA synthase. Investigation of the role of the conserved His453 showed that this residue might be the main general base catalyst instead of His480. This was demonstrated by a massive impairment of the enzyme's *in vivo* and *in vitro* activities when His453 was mutated to Glu453. Nevertheless, His480 can still functionally replaced His453 as the general base catalyst. In the present study, the development of the 3D model for Type II PhaC1$_{P.sp\ USM\ 4-55}$ was presented employing the threading and molecular modeling approach. The choice of threading method over homology modeling was due to the very low sequence identity of this enzyme compared to available experimentally determined structures (< 30%)[23,24].

2 METHODS

2.1 Data mining and Sequence Analysis

The linear chain of PhaC1$_{P.sp\ USM\ 4-55}$ protein containing 559 residues[25] was subjected to various analysis on SWISS-PROT[26], PDB[27], GenBank[28] and PIR[29] using BLAST[30] and PSI-BLAST[31]. Pair-wise and multiple sequence alignment between PHA synthase Type I, II and III were carried out using LALIGN[32] and CLUSTALW[33] respectively. Superfamily HMM[34] and PSI-BLAST was used to identify any conserved domains or families found in the protein.

2.2 Template Selection, Model Development and Evaluation

Five secondary structure prediction methods were used in this work; PSIPRED[35], PHD[36], Prof[37] and SSpro[38] and Jnet[39] to obtain the secondary structure information of the enzyme. The amino acid sequence was then threaded to the library of known folds using mGenThreader[40], 3DPSSM[41], FUGUE[42] and SamT99[43]. The fold classification of these templates were checked using SCOP[44]. The sequence alignment were used as the input for the development of the 3D models employing MODELLER 6v2[45]. Further optimization was carried out using DISCOVER module in InsightII and MODELLER 6v2. The resulted models were evaluated using PROCHECK[46], Verify 3D[47], PROVE[48], WHAT_CHECK[49] and ERRAT[50]. The secondary structural assignment was carried out using DSSP[51].

3 RESULTS AND DISCUSSION

3.1 Data mining and Sequence Analysis

Pair-wise sequence alignment showed that PhaC1$_{P.sp\ USM}$ $_{4-55}$ shares 35.1% and 18.9% of sequence identity with both Type I (PhaC$_{Re}$) and Type III (PhaC$_{Cv}$) enzymes respectively and it shares 57.9% sequence identity with PhaC2$_{P.sp\ USM\ 4-55}$. As expected, results from BLAST and PSI-BLAST revealed that PhaC1$_{P.sp\ USM\ 4-55}$ did not show any significant homology with any solved three-dimensional structure from PDB (results not shown). Conserved domain search by Superfamily HMM demonstrated that this enzyme belongs to the α/β hydrolase fold family located from residue 194 to residue 530 of PhaC1$_{P.sp\ USM\ 4-55}$. A few conserved blocks were identified from the CLUSTALW multiple sequence alignment of various PHA synthases that might play important roles in the enzyme functional properties (results not shown). A putative active-site cysteine was identified within a consensus pentapeptide motif which made up of "Gly-Xaa-(nucleophile)-Xaa-Gly" frequently found in the active site of α/β hydrolase fold. This lipase-box like region (GACSG) is located at the residue 294 to residue 298 with Cys296 as the nucleophile. This region is one of the conserved region found in all PHA synthases and the presence of this consensus motif supports the fact that this enzyme conform to the paradigm of the α/β hydrolase fold[52]. Diverged from a common ancestor, the α/β

hydrolase superfamily of proteins is one of the largest known which includes synthases, esterases, lipases, transferases, thioesterases, haloperoxidases and many more. All of the enzymes of this family share a common fold and despite their differences in catalytic functions, these enzymes harbor a conserved catalytic amino acid sequence of the following format: nucleophile-acid-histidine, in which the nucleophile and the acid varies. The nucleophile can be cysteine, serine, aspartate etc., and the acid will be either aspartic acid or glutamic acid while histidine is strictly conserved. The multiple sequence alignment also showed that the conserved catalytic aspartic acid and the catalytic histidine are located at residues 451 and 479 respectively. From mutagenesis studies of Type II PhaC$_{Pa}$[22], it was proposed that two different catalytic histidines can functionally replaced each other as the general base catalyst for nucleophile activation and the replacement of the highly conserved cysteine, aspartate and histidine for alanine, asparagine and glutamate, respectively, did abolish the enzyme activity suggesting that these residues are indeed essential for this enzyme. Moreover, the importance of these residues had also been suggested by other site-directed mutagenesis studies on the corresponding residues from other types of PHA synthases[21,53-55]. Based on these findings, and due to the fact that there is still yet mutagenesis studies carried out for PhaC1$_{P.sp\ USM\ 4-55}$, we proposed that the catalytic residues for this enzyme may comprise of Cys296, Asp451, His479 and His452.

3.2 Secondary Structure Prediction

Results from secondary structure prediction methods showed that PhaC1$_{P.sp\ USM\ 4-55}$ was made up of 18 α-helices and 15 β-strands. This result was obtained by making a consensus between the results given by the 5 prediction methods. The consensus structures were shown in Figure 1. The core region which belonged to the α/β hydrolase fold family contained an alternate pattern of the helices and the strands. This behavior complemented with the family fold pattern.
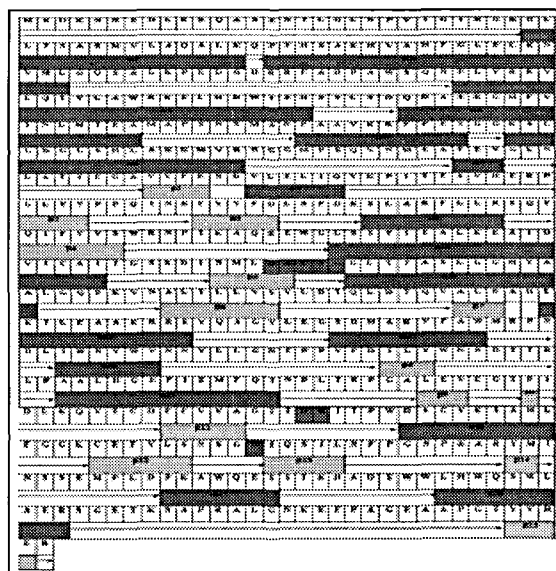

Figure 1: The consensus results of 5 different secondary structure prediction methods for all 559 residues. Red bars represent α-helices, blue bars are β-strands and arrows represent coils. It is predicted that PhaC1$_{P.sp\ USM\ 4-55}$ contains 18 α-helices and 15 β-strands. The purple colored boxes are the proposed catalytic machinery for this enzyme.

3.3 Template Selection

In the absence of high homology with all known structures, threading approach was used to predict the three-dimensional model of PhaCl$_{P.sp\,USM\,4-55}$. In comparative modeling technique, the best template is usually chosen based on the highest percentage of sequence identity between the target and the template proteins. On the contrary, selecting the best template for threading technique is not that straightforward. Sometimes, the proposed templates could turn out to be false positives. The best ranking template doesn't always necessarily reflect as the most suitable template. Due to this drawback as well as the low accuracy of this method, results from the threading programs were carefully examined and compared. The target sequence was scored for its compatibility against the proposed templates as summarized in Table I. However, only templates with significant high scores were tabulated. Seven out of the nine templates were found to be in the same family of α/β hydrolase fold. Among these templates, only human gastric lipase (1HLG[56]) was retained in all the results given by the three out of the four methods. It was either at the top rank or the third best. This could be happening by chance, but it was still worth to take note on this particular template protein compared to the others. Having only 11% sequence identity and 42% similarity, 1HLG outperformed the rest of the proposed templates in various ways. Coming from the same family, these two proteins not only shared the same catalytic residues; a nucleophile, an acid and histidine, but it was also found that from the alignment, these residues aligned very well with each other (Figure 2). However, this was not the case for the other proposed templates, in which not all of their catalytic residues aligned with the catalytic residues of the target protein (results not shown). Furthermore, it was also found that most of the sequence-structure alignments produced between PhaC1P.sp $_{USM\,4-55}$ and the other templates contained numerous insertions and deletions that would cause catastrophic problems in the 3D structure model. However, some templates did show higher sequence identity compared to 1HLG, for instance 1S80 with 14% identity. Unfortunately, the reliability of the alignment was pretty much doubtful with the reasons described former. Thus, 1HLG was chosen as the modeling template for PhaC1$_{P.sp\,USM\,4-55}$.

For further assessment on the reliability and reasonableness of the structure-sequence alignment between 1HLG and PhaC1$_{P.sp\,USM\,4-55}$, the consensus secondary structure prediction of PhaC1$_{P.sp\,USM\,4-55}$ was used to check the alignment. This was achieved by comparing the aligned secondary structures of 1HLG and the consensus predicted secondary structure of PhaC1$_{P.sp\,USM\,4-55}$ as shown in Figure 2. Surprisingly, the structure alignment managed to align 10 helices, 3 β-strands and 13 random coils altogether. The rest of the regions were seemed to have a few structural mismatched and gap containing segments that most likely to introduce errors in the predicted model.
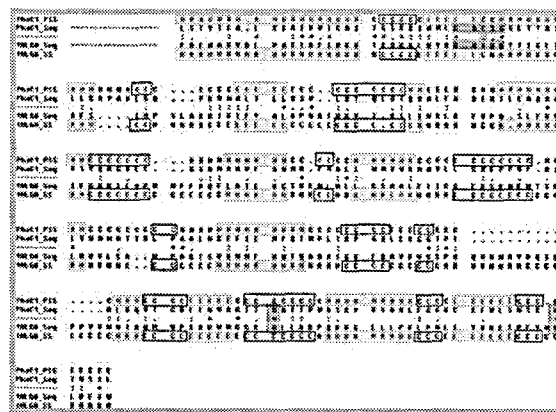

Figure 2: Sequence-structure alignment between PhaC1$_{P.sp\,USM\,4-55}$ and 1HLG generated by 3DPSSM. The filled vertical regions represent the catalytic site for both PhaC1$_{P.sp\,USM\,4-55}$ and 1HLG with their catalytic residues align with each other. The filled boxes represent the align secondary structural elements between the two proteins and the blank boxes showed the regions containing loop or random coils also aligned very well between the proteins. PhaC1 (PhaC1$_{P.sp\,USM\,4-55}$), 1HLGA (1HLG), pss/ss (secondary structure) and Seq (sequence).

3.4 3D Model Building

Since α/β hydrolase fold is regarded important as to elucidate the catalytic function of this protein, the 3D structure covering this region was built in this study with the structure sequence alignment between residue 267 to residue 484 of PhaC1$_{P.sp\,USM\,4-55}$ and residue 123 to residue 358 of 1HLG Chain A as shown in Figure 2. The alignment showed an expected result with the conserved catalytic residues of PhaC1$_{P.sp\,USM\,4-55}$ (Cys296, Asp451, His479) aligned with the catalytic triad of 1HLG (Ser153, Asp324, His353). However, the terminal regions of the PhaC1$_{P.sp\,USM\,4-55}$ (residue 1 till residue 266; residue 485 till residue 559) were not modeled due to unsatisfactory alignments. Thirty models comprising all non-hydrogen atoms were generated using MODELLER6v2. The coordinates of the structure 1HLG were assigned to the target sequence according to the structure sequence alignment in Figure 2. Model number 10 was selected as the best model based on the lowest objective function (OF) calculated by the program. Figure 3 showed the five models with the lowest objective function (kcal/mol). MODELLER works based on a satisfaction of spatial restraints. The highlight of the program is to satisfy all the restraints derived from the structure-sequence alignment. Models that produced high violations of the restraints were considered as poor, which in turn lead to higher objective functions, calculated by CHARMM-22 [57] forcefield. The

Method	Template	Fold	% id	Score
3DPSSM	Human Gastric Lipase (PDB id: 1HLG)	α/β hydrolase fold	11	E value[a] = 0.00888
	Human soluble epoxide hydrolase (PDB id: 1S80)	HAD-like	14	E value = 0.327
	Bromoperoxidase (PDB id: 1BRT)	α/β hydrolase fold	11	E value = 0.519
MgenThreader	Haloalkane Dehalogenase (PDB id: 1BN6:A)	α/β hydrolase fold	11	E value[b] = 3e-05
	Mouse Epoxide Hydrolase (PDB id: 1CR6:A)	HAD-like	8	E value = 3e-05
	Human Gastric Lipase (PDB id: 1HLG:A)	α/β hydrolase fold	11	E value = 3e-05
FUGUE	Human Gastric Lipase (PDB id: 1HLG:A)	α/β hydrolase fold	11	Z-SCORE[c] = 7.86
	Hydroxynitrile-Lyase (PDB id: 1QJ4)	α/β hydrolase fold	11	Z-SCORE = 6.31
	Proline Iminopeptidase (PDB id: 1AZW)	α/β hydrolase fold	12	Z-SCORE = 5.08

[a] E-value must be as low as possible
[b] E-value < 0.001; very significant and certain
[c] Recommended cutoff : Z-SCORE >= 6.0 (CERTAIN 99% confidence)

Table 1: Proposed folds obtained from three threading methods along with the calculated scores.

model was then subjected to optimization scheme to relieve all the bad contacts. Eighty models were generated and model number 76 was picked to represent the predicted structure of PhaCl$_{P.sp}$ USM 4-55 based on the satisfactory results obtained from evaluation analysis. The predicted 3D model of PhaCl$_{P.sp}$ USM 4-55 and the crystal structure of 1HLG are presented in Figure 5. The predicted model shows high similarity to the template structure. The superimposition of the predicted 3D PhaCl$_{P.sp}$ USM 4-55 model with the 1HLG structure yield an RMSD of 1.15 Å covering 86.2% of the backbone atoms.
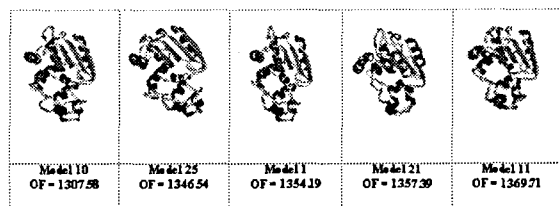

Figure 3: Five different 3D models generated using the threading approach with the objective functions given in kcal/mol
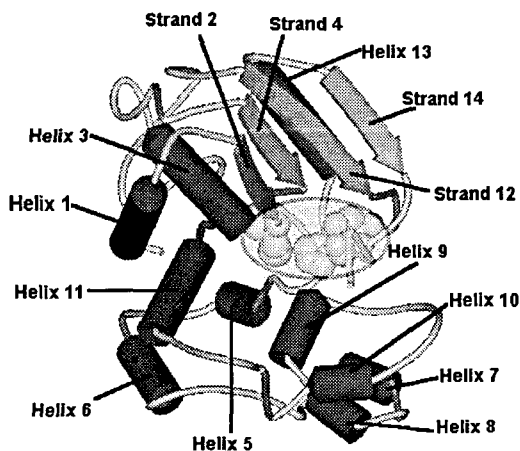

Figure 4: Overall topology of the predicted 3D model of PhaCl$_{P.sp}$ USM 4-55. The CPK model represents the catalytic residues, which comprise of CYS296, Asp451and His479. 10 α-helices and 4 β-strands were found to occur in the model. CYS296 was found to be located between strand 2 and helix 3.
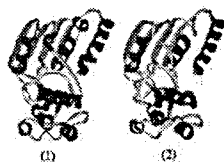

Figure 5: Comparison of the 3D structure between the template 1HLG (1) and the developed model of PhaCl$_{P.sp}$ USM 4-55 (residue 267- 484) (2). The blue ribbon represents β strands, the red ribbon represents α helices and the green region represents turns

Figure 6: Schematic representation of the 3D model of PhaCl$_{P.sp}$ USM 4-55 with the nucleophile Cys296 (green color) located at the sharp turn between β-strand (arrow) and α-helix (cylinder) that results in having unfavorable phi and psi angles. The mesh is colored based on electrostatic potential around the active site.

The nucleophile Cysteine was found to reside at the sharp elbow turn from β-strand to α-helix which is a conserved property of the α/β hydrolase fold family[19,58] as shown in Figure 6. This was also confirmed by the three other studies[20-22]. Figure 8 shows the specific location for the proposed active residues Cys296, Asp451, His452 and His479. The calculated distance between the Cαs of these

residues were as such that they had almost the same distance calculated between the Cαs of the catalytic residues of 1HLG as shown in Table 2. Again, this suggested that this threading model was reliable in terms of the accuracy of backbone threading from the template structure. The secondary structure from the generated model was calculated using DSSP with 41.3% of the residues forming α-helices, 12.4% of β-strands and 46.3% of coils (Figure 7). This ratio is almost in the same range with the calculated ratio obtained from the consensus secondary structure prediction methods that revealed 36.2% of α-helices, 16.5% of β-strands and 47.3% of coils. This demonstrates a good confidence level with the predicted model in terms of the secondary structure prediction.

Figure 7: Secondary structure assignment of the predicted three-dimensional model of PhaCl$_{P.sp}$ USM 4-55 from residue 267- 484 using DSSP. Red ribbon = helices, blue bars = coils and green arrow = strands.

	PhaCl$_{P.sp}$USM 4-55	1HLG
Nucleophile – Histidine	Cys296 – His479 = 7.95 Å	Ser153 – His353 = 7.41 Å
Histidine - Acid	His479 – Asp451 = 4.74 Å	His353 – Asp324 = 4.67 Å

Table 2. Comparison of the Cα distances for the catalytic residues between the template structure Human gastric lipase and the predicted structure of PhaCl$_{P.sp}$USM 4-55

Inspection of the predicted model revealed that Cys296, His479, His452 and Asp451 are located adjacent to each other at the core structure which is consistent with the work carried out by other studies[20-22]. Another important residue was Trp397, which was highly hydrophobic and was found to be located at the surface of the protein exposing to the solvent, it is thus highly probable that it might be involved in protein-protein dimerization process. This is in-line with the previous site-directed mutagenesis studies which proposed that the corresponding Trp397 for Type II PHA synthase might play important role in the protein-protein interaction during the dimerization of PhaC subunit since Type II PHA synthase exists in dimer when they are active[4,22,59]. However, further studies need to be carried out to confirm the exact function of this residue.
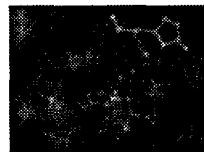

Figure 8: Surface representation of the active site containing the proposed catalytic residues shown as the ball and stick models colored according to the residue type (light blue = His452, red = Asp451, purple = His479, dark blue = Cys296). These residues are located adjacent to each other and the close proximity between His479 and Cys296 favors the charge relay mechanism.
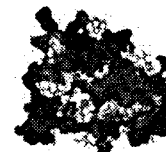
Figure 9: Surface presentation of the 3D model of PhaCl$_{P.sp}$USM 4-55. The surface represents the hydrophilic region of the protein. The hydrophobic residues represents by the red stick atom model.

3.5 Model Evaluation

In evaluating the optimized model, PROCHECK analyses showed that only 2 residues were located in the disallowed region of the Ramachandran Plot and 84.8% of the residues were located in the most favored region (red region from Figure 10). The other residues were found to reside in the additional and generously allowed regions. The two residues that were located in the disallowed region of the Ramachandran plot were Cys296 and Val326. The nucleophile Cys296 was expected to be in the disallowed region due to the sharp and abrupt turn from β-strand to α-helix which is a conserved and important feature of the α/β hydrolase fold family. The sharpness of this turn resulted in the phi and psi angles being in an unfavorable region on the Ramachandran plot. As expected, the residues after and before this nucleophile were residues with small side chains (glycine) to ease the packing of the β-strand against the α-helix and to prevent stearic conflicts. From VERIFY_3D analysis (Figure 11), it was found that 82.65% of the residues scored more than 0.2, meaning that 82.65% of the residues complemented with the 1D-3D profile. For the quality of the predicted model to be considered satisfactory, it is expected to have the Verify_3D score of more than 80%. Analysis of entire structure calculated from PROVE program gave Z-score RMS of 1.746. Z-score above 4.0 and below −4.0 represent the occurrence of many errors in the structure in terms of the packing of the buried atoms. Further analysis was done using the program WHAT_CHECK. It was found that residues Gln330, Asn377, Asn382 and Asn419 have different orientation for their side chains compared to the orientation based on hydrogen bond analysis with solved structures. This might be due to the fact that they could form more energetically favorable hydrogen bonds with their current configuration. Overall, the values obtained from evaluation programs however appeared reasonable, taking into account the resolution (3.0Å) of the template and the fact that the structure was obtained by threading with very low sequence identity (11%) and similarity (42%) between the template and the target.
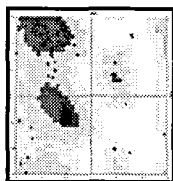

Figure 10: Ramachandran Plot of the 3D model of PhaCl$_{P.sp}$ USM 4-55. Red region represents the most favored region, yellow=allowed region, light yellow=generously allowed region, white= disallowed region.
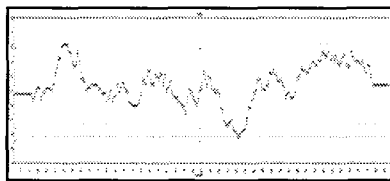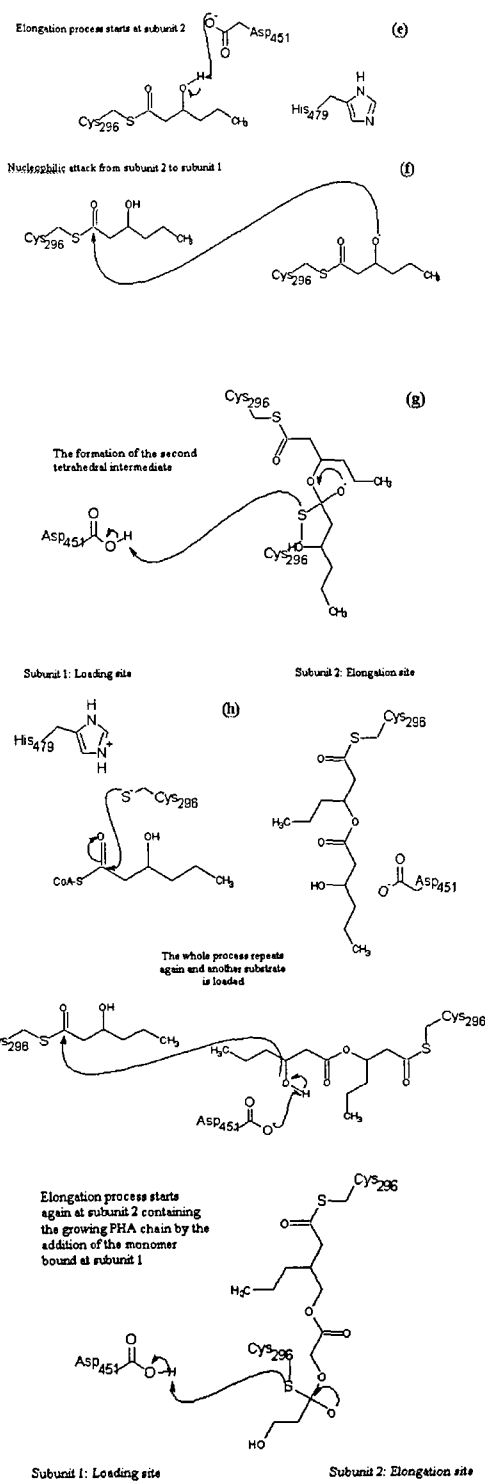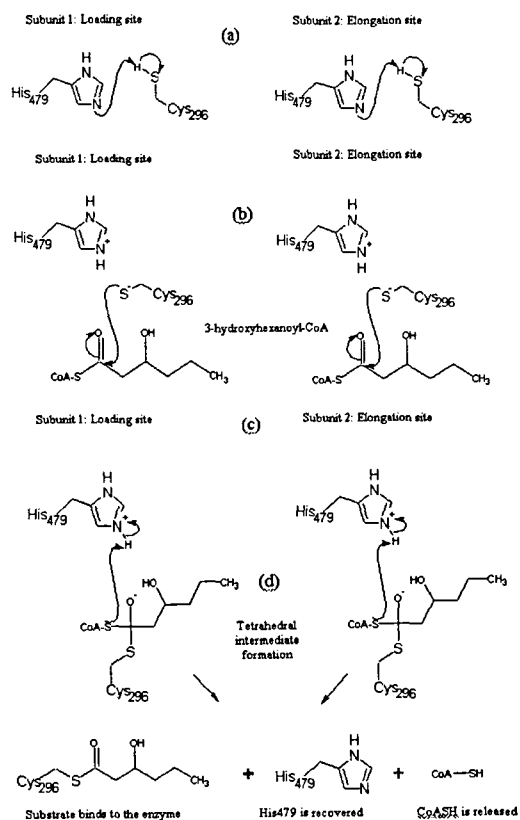
Figure 11: Verify-3D plot for the 3D model of PhaCl$_{P.sp}$ USM 4-55 with 82.65% of the residues had an averaged 3D-1D score > 0.2.

3.6 Proposed Catalytic Mechanism

While the 3D structure of PHA synthase remains elusive, the catalytic mechanism of this enzyme is also still poorly understood. In the present study, from the multiple sequence alignment, the proposed catalytic residues for PhaCl$_{P.sp}$ USM 4-55 consists of cystine as the nucleophile, aspartic acid as the conserved acid and histidine as the general base catalyst.

Hence, the catalytic mechanism for PhaCl$_{P.sp}$ USM 4-55 specifically and PHA synthase generally may follow the classical catalytic mechanism observed in cysteine proteases[60] since the nature and orientation of the catalytic groups are more or less similar. These groups of enzymes also belong to the same family containing an α/β hydrolase fold region. For lipases or serine proteases, the nucleophile serine is thought to be part of a charge-relay system along with a conserved histidine and aspartate by forming a catalytic triad. The negative charge from the carboxyl ion of aspartate is transferred to histidine and then to serine to enhance the power of the nucleophile[61,62]. However, in cysteine proteases and other related proteins, it has long been established that only histidine is sufficient for nucleophilic activation in cysteine deprotonation[20,63-67]. From previous study, it was found that aspartate did not involve in the initial priming step and was not part of the catalytic machinery in PhaC$_{Cv}$[20]. Hence, it can be assumed that PhaC$_{P.sp}$ 4-55 follows the same behavior by harboring catalytic dyad that excludes aspartate. However, this residue was still observed to be very important in PHA biosynthesis involving the elongation step which will be further explained. The two subunits of Type II PHA synthase were suggested to form a dimer when active[18,68]. This dimer will then attach to the surface of the PHA granule for polymerization to take place. It was proposed that during PHA biosynthesis, the thiol group from the first subunit will act as the loading site to load the substrate 3-hydroxyacyl-CoA to the enzyme. The other thiol group from the second subunit will act as the elongation site in which it will be responsible in PHA elongation[69]. Based on the previous postulated catalytic mechanism[20-22,68-71], we proposed an extension to the mechanism as shown in Figure 12. The close proximity between His479 and Cys296 of PhaCl$_{P.sp}$ USM 4-55 enables the former to extract a proton from the thiol group of the nucleophile Cys296 (Figure 12(a)). This activates the nucleophilic attack to the carbonyl carbon of the substrate (3-hydroxyhexanoyl-CoA) that will yield a tetrahedral intermediate as shown in Figure 12(b) and 12(c) respectively. This intermediate will then collapses once CoA is released to form CoASH (Figure 12(d)). For the elongation step which involved the second subunit, the same reaction will take place as above. The next step as shown in Figure 12(e), is the activation of the 3-hydroxy group of the bound substrate (nucleophile) at the second subunit by Asp451, which acts as the second general base catalyst[20-22]. This nucleophile will attack the acylated enzyme at the first subunit (Figure 12(f)). Consequently, a second tetrahedral intermediate will be formed which then collapses by the release of Cys296 as the preparation for the next substrate loading (Figure 12(g)). This whole process will repeat as elongation of the PHA chain proceeds (Figure 12(h)). In contrast to the previous postulated mechanism, the above mechanism accounts for the formation of two tetrahedral intermediates (sp^3 hybridization) during the PHA biosynthesis. These transition state intermediates have been widely agreed to be present in all cysteine proteases and lipases[60,72,73] but have not been suggested before in PHA biosynthesis. Another point to consider here is whether or not this enzyme contains another active site which known as the oxyanion hole. This is another conserved feature of cysteine proteases and the α/β hydrolase fold protein that is present during the stabilization of the tetrahedral intermediate state. For the enzyme to work as effectively as possible, this

unstable transition state needs some form of stabilization. An oxyanion hole is said to occur during the stabilization of the highly negatively charged oxyanion (O⁻) of the tetrahedral intermediate by two hydrogen bonds from the surrounding amide (NH) group as such that the NH groups are pointing towards the oxyanion[19,60,72,74]. Several studies have demonstrated that the presence of the oxyanion hole is significant in stabilizing the tetrahedral intermediate in various proteins[75-79]. For α/β hydrolase fold proteins, the mandatory feature of the occurrence of the oxyanion hole is that one of the residues that is involved in the hydrogen bonding with the oxyanion is always located following to the nucleophile[19] (Ser297 in PhaC1$_{P.sp}$ USM 4-55). By looking at the structure-sequence alignment in Figure 2, PhaC1$_{P.sp}$ USM 4-55 Ser297 was found to align with 1HLG Gln154 which has been agreed to contribute to the oxyanion hole formation for 1HLG. However, the second NH group that is involved in the stabilization of the tetrahedral intermediate in PhaC1$_{P.sp}$ USM 4-55 couldn't be identified. Perhaps knowledge of how the enzyme interacts with the substrate would probably shed some light into identifying the second residue that contributes to the formation of the oxyanion hole if this feature is indeed present in the PHA synthase catalytic mechanism.
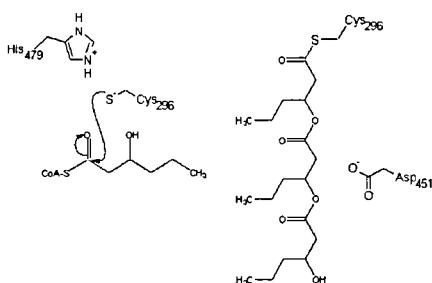
Elongation process starts at subunit 2   (e)

Nucleophilic attack from subunit 2 to subunit 1   (f)

The formation of the second tetrahedral intermediate   (g)

Subunit 1: Loading site   Subunit 2: Elongation site

(a)

Subunit 1: Loading site   Subunit 2: Elongation site

(b)

3-hydroxyhexanoyl-CoA

Subunit 1: Loading site   (c)   Subunit 2: Elongation site

(d)

Tetrahedral intermediate formation

Substrate binds to the enzyme   His479 is recovered   CoASH is released

Subunit 1: Loading site   Subunit 2: Elongation site

(h)

The whole process repeats again and another substrate is loaded

Elongation process starts again at subunit 2 containing the growing PHA chain by the addition of the monomer bound at subunit 1

Subunit 1: Loading site   Subunit 2: Elongation site

Figure 12: Proposed catalytic mechanism for PhaCl$_{Psp \ USM \ 4-55}$ involving Cys296, Asp451 and His479 with the protein dimer forming the loading and elongation site. In subunit 1 (the loading site), His479 activates Cys296 for nucleophilic attack to the carbonyl carbon of the substrate (3-hydroxyhexanoyl-CoA) yielding a tetrahedral intermediate which then collapses by the release of the CoA group. The elongation process proceeds at the second subunit by the activation of 3-OH group of the bound substrate for nucleophilic attack to the carbonyl carbon of the substrate covalently bound to the Cys296 of the first subunit to form an ester bond. This process repeats by adding more and more monomers to the growing PHA chain.

4 CONCLUSIONS

We have attempted to predict the 3D structure of PhaCl$_{Psp \ USM \ 4-55}$ using the threading approach due to very low homology to any available experimentally solved 3D protein structures. A series of molecular modeling and computational methods were combined in order to gain insight into the 3D structure of PhaCl$_{Psp \ USM \ 4-55}$ concentrating on the α/β hydrolase fold region. Human gastric lipase was used as the modeling template and from the developed model, it was shown that the proposed catalytic residues were located adjacent to each other. We also proposed an extension to the catalytic mechanism of this enzyme based on the cysteine protease catalytic mechanism. Two tetrahedral intermediates were postulated to occur during the PHA biosynthesis and we believe that the formation of an oxyanion hole is integral in the catalytic mechanism of this enzyme.

5 ACKNOWLEDGEMENTS

6 REFERENCES

[1]   B.H. Rehm and A. Steinbuchel.Biochemical and genetic analysis of PHA synthases and other proteins required for PHA synthesis, Int J Biol Macromol 25: 3-19,1999.

[2]   B.H.A. Rehm and A. Steinbüchel Biopolymer, in: Steinbüchel A and Doi Y (Eds.), Wiley-VCH, Heidelberg, 2001, pp. 199-200.

[3]   S. Taguchi , H. Nakamura, T. Kichise, T. Tsuge, I. Yamato and Y. Doi Production of Polyhydroxyalkanoate (PHA) from Renewable Carbon Sources in Recombinant *Ralstonia*

eutropha Using Mutants of Original PHA Synthase, Biochem. Eng. J 16: 107-113,2003.

[4]   B.H. Rehm, Q. Qi, B.B. Beermann, H.J. Hinz and A. Steinbuchel.Matrix-assisted in vitro refolding of *Pseudomonas aeruginosa* class II polyhydroxyalkanoate synthase from inclusion bodies produced in recombinant *Escherichia coli*, Biochem J 358: 263-268,2001.

[5]   S. Langenbach, B.H. Rehm and A. Steinbuchel. Functional expression of the PHA synthase gene phaC1 from *Pseudomonas aeruginosa* in *Escherichia coli* results in poly(3-hydroxyalkanoate) synthesis, FEMS Microbiol Lett 150: 303-309,1997.

[6]   S. Hein, J.R. Paletta and A. Steinbuchel. Cloning, characterization and comparison of the *Pseudomonas mendocina* polyhydroxyalkanoate synthases Phacl and PhaC2, Appl Microbiol Biotechnol 58: 229-236,2002.

[7]   T. Kichise, T. Fukui, Y. Yoshida and Y. Doi.Biosynthesis of Polyhydroxyalkanoates (PHA) by Recombinant *Ralstonia eutropha* and Effects of PHA Synthase Activity on in Vivo PHA Biosynthesis, Int. J. Biol. Macromol 25: 69-77,1999.

[8]   T. Hai, S. Hein and A. Steinbuchel.Multiple Evidence for Widespread and General Occurrence of Type III PHA Synthases in cyanobacteria and Molecular Characterization of The PHA Synthases from Two Thermophilic cyanobacteria: *Chlorogloeopsis fritschii* PCC 6912 and *Synechococcus sp.* Strain MA19, Microbiology 147: 3047-3060,2001.

[9]   A.A. Amara, A. Steinbüchel and B.H.A. Rehm.*In vivo* evolution of the *Aeromonas punctata* polyhydroxyalkanoate (PHA) synthase: Isolation and characterization of modified PHA synthases with enhanced activity, Appl. Microbiol. Biotechnol. 59: 477-482,2002.

[10]   H. Brandl, R.A. Gross, R.W. Lenz and R.C. Fuller.*Pseudomonas Oleovorans* as a Source of a New Poly(3-hydroxyalkanoates) for Potential Applications as Biodegradable Polyesters, Appl Env Microb 54: 1977-1982,1988.

[11]   G.N.M. Huijberts, G. Eggink, P. Waard, G.W. Huisman and B. Witholt.*Pseudomonas putida* KT2442 Cultivated on Glucose Accumulates poly(3-hydroxyalkanoates) Consisting of Saturated and Unsaturated Monomers, Appl Env Microb 58: 536-544,1992.

[12]   B. Witholt and B. Kessler. Perspectives of medium chain length poly 93-hydroxyalkanoates), a versatile set of bacterial bioplastics, Curr. Op. Biotechnol. 10: 279-285,1999.

[13]   Q. Qi, A. Steinbüchel and B.H.A. Rehm.*In vitro* Synthesis of poly (3-hydroxyalkanoate): Purification and Enzymatic Characterization of Type II Purification and Enzymatic Characterization of Type II Polyhydroxyalkanote Synthases PhaC1 and PhaC2 from *Pseudomonas*

earuginosa, Appl Microbiol Biotehnol 54: 37-43,2000.

[14] R.A. Gross, C. DeMello and R.W. Lenz.Biosynthesis and characterization of poly(-b-hydroxyalkanoates) produced by *Pseudomonas oleovorans*, Macromolecules 22: 1106-1115,1989.

[15] M. Liebergesell, B. Schmidt and A. Steinbüchel.Isolation and identification of granule-associated proteins relevant for poly(3-hydroxyalkanoic acid) biosynthesis in *Chromatium vinosum* D, FEMS Microbiol. 78: 227-232,1992.

[16] W. Yuan, Y. Jia, J. Tian, K.D. Snell, U. Muh, A.J. Sinskey, R.H. Lambalot, C.T. Walsh and J. Stubbe.Class I and III polyhydroxyalkanoate synthases from *Ralstonia eutropha* and *Allochromatium vinosum*: characterization and substrate specificity studies, Arch Biochem Biophys 394: 87-98,2001.

[17] G.J. McCool and M.C. Cannon.PhaC and PhaR are required for polyhydroxyalkanoic acid synthase activity in *Bacillus megaterium*, J Bacteriol. 183: 4235-4243,2001.

[18] B.H.A. Rehm.Review article, polyester synthases: natural catalysts for plastics, Biochem J 376: 15-33,2003.

[19] D.L. Ollis, E. Cheah, M. Cygler, B. Dijkstra , F. Frolow, S.M. Franken, M. Harel , S.J. Remington , I. Silman and J. Schrag The alpha/beta Hydrolase Fold, Protein Eng 5: 197-211,1992.

[20] Y. Jia, T.J. Kappock, T. Frick, A.J. Sinskey and J. Stubbe.Lipases provide a new mechanistic model for polyhydroxybutyrate (PHB) synthases: characterization of the functional residues in *Chromatium vinosum* PHB synthase, Biochemistry 39: 3927-3936,2000.

[21] B.H. Rehm, R.V. Antonio, P. Spiekermann, A.A. Amara and A. Steinbuchel.Molecular characterization of the poly(3-hydroxybutyrate) (PHB) synthase from *Ralstonia eutropha*: in vitro evolution, site-specific mutagenesis and development of a PHB synthase protein model, Biochim Biophys Acta 1594: 178-190,2002.

[22] A.A. Amara and B.H.A. Rehm.Replacement of the catalytic nucleophile cysteine-296 by serine in classII polyhydroxyalkanoate synthase from *Pseudomonas aeruginosa*-mediated synthesis of a new polyester: identification of catalytic residues, Biochem J 374: 413-421,2003.

[23] D. Eisenberg, R.M. Weiss and T.C. Terwilliger.The hydrophobic moment detects periodicity in protein hydrophobicity, Proc Natl Acad Sci U S A 81: 140-144,1984.

[24] C. Sander and R. Schneider.Database of homology-derived protein structures and the structural meaning of sequence alignment, Proteins 9: 56-68,1991.

[25] A. Baharuddin Cloning and characterization of polyhydroxyalkanoic acid (PHA) synthase gene from *Pseudomonas* sp. USM4-55, Biological Sciences Department, Universiti Sains Malaysia, Penang, 2001, pp. 47-66.

[26] A. Bairoch and B. Boekmann.The SWISS-PROT protein sequence data bank, Nuc Acids Res 20: 2019-2022,1992.

[27] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne.The Protein Data Bank, Nuc Acids Res 28: 235-242,2000.

[28] D.A. Benson, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell and D.L. Wheeler.GenBank: update, Nuc Acids Res 32,2004.

[29] C.H. Wu, L.L. Yeh, H. Huang and e. al.The protein information resource, Nuc Acids Res 31: 345-347,2003.

[30] S.F. Altschul, W. Gish, W. Miller, E.W. Myers and D.J. Lipman.Basic local alignment search tool, J. Mol. Biol. 215: 403- 410,1990.

[31] S.F. Altschul, T.L. Madden and A.A.e.a. Schaffer.Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs, Nuc. Acids Res 25: 3389-3402,1997.

[32] W.R. Pearson, T. Wood, Z. Zhang and W. Miller Comparison of DNA sequences with protein sequences, Genomics 46: 24-36,1997.

[33] J.D. Thompson, D.G. Higgins and T.J. Gibson.CLUSTALW: Improving The Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-specific Gap Penalties and Weight Matric Choice, Nuc. Acids Res 22: 4673-4680,1994.

[34] M. Madera, C. Vogel, S.K. Kummerfield, C. Chotia and J. Gough.The SUPERFAMILY Database in 2004: Additions and Improvements, Nuc. Acids Res 32: 235-239,2004.

[35] L.J. McGuffin, K. Bryson and D.T. Jones.The PSIPRED Protein Structure Prediction Server, Bioinformatics 15: 404-405,2000.

[36] B. Rost and J. Liu.The Predictprotein Server, Nucl. Acids Res 31: 3300-3304,2003.

[37] B. Rost.Predicting One-dimensional Protein Structure by Profile Based Neural Networks, Methods in Enzymology 266: 525-539,1996.

[38] G. Pollastri, D. Przybylski, B. Rost and P. Baldi.Improving the Prediction of Protein Secondary Structure in Three and Eight Classes Using Recurrent Neural Networks and Profiles, Proteins 47: 228-235,2002.

[39] J.A. Cuff and G.J. Barton.Application of Enhanced Multiple Sequence Alignment Profiles To Improve Protein Secondary Structure Prediction, Proteins 40: 502-511,1999.

[40] D.T. Jones.An Efficient And Reliable Protein Fold Recognition Method For Genomic Sequences, J. Mol. Biol. 287: 797-815,1999.

[41] L.A. Kelley, R.M. MacCallum and M.J.E. Sternberg.Enhanced Genome Annotation Using Structural Profiles in The Program 3DPSSM, J. Mol. Biol. 299: 499-520,2000.

[42] J. Shi, T. Blundell and K. Mizuguchi.Sequence-structure Homology Recognition Using Environment-specific Substitution Tables and Structure-dependent Gap Penalties, J Mol. Biol 310: 243-257,2001.

[43] A. Krogh, M. Brown, I.S. Mian, K. Sjolander and D. Haussler.Hidden Markov Models in Computational Biology: Applications To Protein Modeling, J Mol. Biol. 235: 1501-1531,1994.

[44] A.G. Murzin, S.E. Brenner, T. Hubbard and C. Chothia SCOP: A structural classification of proteins database for the investigation of sequences and structures, J. Mol. Biol. 247: 536-540,1995.

[45] A. Sali and T.L. Blundell Comparative protein modelling by satisfaction of spatial restraints, J Mol Biol 234: 779-815,1993.

[46] R.A. Laskowski, M.W. MacArthur, D.S. Moss and J.M. Thornton.PROCHECK: A Program To Check The Stereochemical Quality of Protein Strcutures, J. Appl. Cryst 26: 283-291,1993.

[47] Luthy, Bowie JU and Eisenberg D.Assessment of Protein Models With three-dimensional profiles, Nature 356: 83-85,1992.

[48] J. Pontius, J. Richelle and S.J. Wodak.Deviations from standard atomic volumes as a quality measure for protein crystal structures, J Mol Biol 264: 121-136,1996.

[49] R.W.W. Hooft, G. Vriend, C. Sander and E.E. Abola.Errors in protein structures, Nature 381: 272-272,1996.

[50] C. Colovos and T.O. Yeates.Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2: 1511-1519,1993.

[51] W. Kabsch and C. Sander.Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers: 2577-2637,1983.

[52] A. Steinbüchel , E. Hustede, M. Liebergesell, U. Pieper, A. Timm and H. Valentin.Molecular basis for biosynthesis and accumulation of polyhydroxyalkanoic acids in bacteria, FEMS Microbiol Rev 9: 217-230,1992.

[53] T.U. Gerngross, K.D. Snell, O.P. Peoples, A.J. Sinskey, E. Csuhai, S. Masamune and J. Stubbe.Overexpression and purification of the soluble polyhydroxyalkanoate synthase from Alcaligenes eutrophus: evidence for a required posttranslational modification for catalytic activity, Biochemistry 33: 9311-9320,1994.

[54] A. Hoppensack, B.H.A. Rehm and A. Steinbüchel.Analysis of 4-phosphopantetheinylation of polyhydroxybutyrate synthase from Ralstonia eutropha: Generation of beta-alanine auxotrophic Tn5 mutants and cloning of the panD gene region, J Bacteriol. 181,1999.

[55] S. Taguchi, A. Maehara, K. Takase, M. Nakahara, H. Nakamura and Y. Doi Analysis of mutational effects of polyhydroxybutyrate (PHB)

[56] polymerase on bacterial PHB accumulation using an in vivo assay system, FEMS Microb. Lett. 198: 65-71,2001.

[56] A. Roussel, S. Canaan, R. Verger and C. Cambillau.Crystal Structure of Human Gastric Lipase and Model of Lysosomal Acid Lipase, Two Lipolytic Enzymes of Medical Interest, J Biol Chem 274: 16995,1999.

[57] A.D.J. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao and e. al.All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, J. Phys. Chem. B 102: 3586-3616,1998.

[58] P. Heikinheimo, A. Goldman, C. Jeffries and D.L. Ollis.Of barn owls and bankers: a lush variety of a/b-hydrolases, Structure 7: 141-146,1999.

[59] J. Wodzinska, K.D. Snell, A. Rhomberg, A.J. Sinskey, K. Biemann and J. Stubbe.Polyhydroxybutyrate synthase: Evidence for covalent catalysis, J. Am. Chem. Soc. 118: 6319-6320,1996.

[60] R. Menard and A.C. Storer Oxyanion hole interactions in serine and cysteine proteases, Biol. Chem. Hoppe-Seyler 373: 393-400,1992.

[61] A.A. Kossiakoff and S.A. Spencer.Direct determination of the protonation state of aspartic acid-102 and histidine-57 in the tetrahedral intermediate of the serine proteases: neutron structure of trypsin, Biochemistry 20: 6462-6474,1981.

[62] A. Warshel, G. Naray-Szabo, F. Sussman and J.K. Hwang.How do serine proteases really work?, Biochemistry 28: 3629-3637,1989.

[63] A.C. Storer and R. Menard.Catalytic mechanism in papain family of cysteine peptidases, Methods Enzymology 244: 486-500,1994.

[64] G. Schmidt, J. Selzer, M. Lerm and K. Aktories.The Rho-deamidating Cytotoxic Necrotizing Factor 1 from Escherichia coli Possesses Transglutaminase Activity. CYSTEINE 866 AND HISTIDINE 881 ARE ESSENTIAL FOR ENZYME ACTIVITY, J Biol. Chem. 273: 13669-13674,1998.

[65] J.M. Chen, N.D. Rawlings, R.A.E. Stevens and A.J. Barrett.Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases, FEBS Lett. 441: 361-365,1998.

[66] X. Huang and F.M. Raushel.Deconstruction of the catalytic array within the amidotransferase subunit of carbamoyl phosphate synthetase., Biochemistry 38: 15909-15914,1999.

[67] T. Vernet, D.C. Tessier, J. Chatellier, C. Plouffe, T.S. Lee, D.Y. Thomas, A.C. Storer and Menard R.Structural and functional roles of asparagine 175 in the cysteine protease papain, J Biol Chem 270: 16645-16652,1995.

[68] Y. Jia, W. Yuan, J. Wodzinska, C. Park, A.J. Sinskey and J. Stubbe.Mechanistic studies on class I polyhydroxybutyrate (PHB) synthase from

Ralstonia eutropha: class I and III synthases share a similar catalytic mechanism, Biochemistry 40: 1011-1019,2001.

[69] U. Muh, A.J. Sinskey, D.P. Kirby , W.S. Lane and J. Stubbe.PHA synthase from *Chromatium vinosum*: Cysteine 149 is involved in covalent catalysis, Biochemistry 38: 826-837,1999.

[70] R. Griebel, Z. Smith and J.M. Merrick.Metabolism of poly-beta-hydroxybutyrate. 1. Purification composition and properties of native-poly-beta-hydroxybutyrate granules from *Bacillus megaterium*, Biochemistry 7: 3676-3681,1968.

[71] Y. Kawaguchi and Y. Doi Kinetics and mechanism of synthesis and degradation of poly(3-hydroxybutyrate) in *Alcaligenes eutrophus*, Macromolecules 25: 2324-2329,1992.

[72] P. Grochulski, Y. Li, J.D. Schrag, F. Bouthillier, P. Smith, D. Harrison, B. Rubin and M. Cygler.Insights into interfacial activation from an open structure of *Candida rugosa* lipase., J Biol. Chem. 268: 12843-12847,1993.

[73] H.V. Tilbeurgh, M. Egloff, C. Martinez, N. Rugani, R. Verger and C. Cambillau.Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by X-ray crystallography, Nature 362: 814-820,1993.

[74] R. Menard, C. Plouffe, P. Laflamme, T. Vernet , D.C. Tessier, D.Y. Thomas and A.C. Storer.Modification of the electrostatic environment is tolerated in the oxyanion hole of the cysteine protease papain., Biochemistry 34: 464-471,1995.

[75] K.M. Parnell, A.C. Seila and S.A. Strobel.Evidence against stabilization of the transition state oxyanion by a pKa-perturbed RNA base in the peptidyl trasferase center., Proc. Natl. Acad. Sci. USA 99: 11658-11663,2002.

[76] R. Menard, J. Carriere, P. Laflamme, C. Plouffe, H.E. Khouri , T. Vernet, D.C. Tessier, D.Y. Thomas and A.C. Storer.Contribution of the glutamine 19 side chain to transition-state stabilization in the oxyanion hole of papain., Biochemistry 30: 8924-8928,1991.

[77] Y. Zheng and T.C. Bruice.Is strong hydrogen bonding in the transition state enough to account for the observed rate acceleration in a mutant of papain?, Proc. Natl. Acad. Sci. USA 94: 4285-4288,1997.

[78] A. Ordentlich, D. Baraks, C. Kronman, N. Ariel, Y. Segall, B. Velan and A. Shafferman Functional characteristics of the oxyanion hole in human acetylcholinesterase, J Biol Chem 273: 19509-19517,1998.

[79] J. Cui, F. Marankan, W. Fu, D. Crich, A. Mesecar and M.E. Johnson.An oxyanion-hole selective serine protease inhibitor in complex with trypsin, Bioorganic & Medicinal Chem 10: 41-46,2002.