MOLECULAR DYNAMICS SIMULATION OF POLYHYDROXYALKANOATE SYNTHASE FROM *Chromobacterium* sp. USM2

by

NITHYANAN A/L ANNAMALAI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

May 2019

ACKNOWLEDGEMENT

This journey of my Master's studies is a wonderful episode in my life that I will never forget. First of all I would like to sincerely thank my supervisor, Dr. Hasni bin Arsad and my co-supervisor Prof. Razip bin Samian for giving me this opportunity to pursue my research study under their good guidance and supervision.

To my lovely Appa and Amma, Mr. Annamalai and Mrs. Santhi, I will never forget the sacrifices that both of you did to make sure I complete my study and achieve my dream. Both of you have always been there for me at every moments of my life. My sisters, Ms.Sasirehka and Ms. Bhanupriyah thank you being supportive for me always.

I indebted to my friends, Adebayo Ismail and Balogun Wasiu for their assistance since from the day I started my research. My special thanks and appreciation to the lecturers at the Integrative Medicine Cluster and the entire staff of IPPT.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3D	Three-dimensional
3HB	3-hydroxybutyrate
3HHx	3-hydroxyhexanoate
Cα	Alpha carbon
HA-CoA	Coenzyme A thioesters of hydroxyalkanaotes
MCL	Medium chain length
MM-PBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
P(3HB)	Poly(3-hydroxybutyrate)
PDB	Protein Data Bank
PHA	Polyhydroxyalkanaote
PhaC	Polyhydroxyalkanoate synthase
$PhaC_{Pa}$	Polyhydroxyalkanaote synthase from *Pseudomonas aeroginosa*
$PhaC_{Re/Cn}$	Polyhydroxyalkanoate synthase from *Cupriavidus necator*
$PhaEC_{Av}$	Polyhydroxyalkanoate synthase from *Allchromatium vinosum*
SASA	Solvent-accessible surface area
SCL	Short chain length
Rg	Radius of gyration
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation

LIST OF UNITS AND SYMBOLS

º	Degree
%	Percentage
Å	Angstrom
K	Kelvin
Na^+	Sodium ion
Cl^-	Chloride ion
bar^{-1}	per bar
fs	femtosecond
ps	picosecond
nm	nanometre
ns	nanosecond
nm	nanometre
nm^2	square nanometre

SIMULASI MOLEKUL DINAMIK POLIHIDROKSIALKANOAT SINTASE DARI *Chromobacterium* sp. USM2

ABSTRAK

Polihidroksialkanoat sintase daripada bakteria pencilan tempatan, *Chromobacterium* sp. (PhaC$_{Cs}$) telah dilaporkan sebagai enzim yang sangat aktif dalam bentuk semulajadi yang mampu menghasilkan monomer PHA yang diketahui mempunyai sifat mekanikal dan terma yang lebih baik berbanding homopolimer P (3HB) yang biasa dihasilkan. Dalam usaha untuk memahami pendimeran PhaC$_{Cs}$, satu penyelidikan telah dilakukan untuk mendapatkan model struktur monomer dan dimer PhaC$_{Cs}$.Hasil kajian mutagenesis terarah, telah mencadangkan bahawa mutasi residu Arg409Ala dan Arg490Ala pada PhaC$_{Cs}$ mengganggu pembentukan dimer PhaC$_{Cs}$. Walau bagaimanapun, setakat ini, belum dapat diketahui sepenuhnya bagaimana mutasi ini mempengaruhi struktur, kestabilan, dinamik, dan fungsi mengikat protein PhaC$_{Cs}$. Dalam kajian ini, pendekatan simulasi molekul dinamik telah digunakan untuk memahami isu – isu ini. Hasil kajian menunjukkan bahawa mutasi residu Arg409Ala dan Arg490Ala telah menggangu interaksi ikatan hidrogen antara His404 dan Gly480 serta interaksi jambatan garam antara Arg490 dan Asp488 yang mengakibatkan ketidakseimbangan kedudukan triad PhaC$_{Cs}$. Di samping itu, dapatan kajian ini juga mendedahkan bahawa Arg409 dan Arg490 tidak terlibat dalam pendimeran PhaC$_{Cs}$.

MOLECULAR DYNAMICS SIMULATION OF POLYHYDROXYALKANOATE SYNTHASE FROM *Chromobacterium* sp. USM2

ABSTRACT

Polyhydroxyalkanoate synthase from locally isolated organism, *Chromobacterium* sp. ($PhaC_{Cs}$) has been reported as a highly active enzyme in its natural form that is able to produce PHA that have better mechanical and thermal properties compared to the commonly produced P(3HB) homopolymer. In an attempt to understand the dimerization of $PhaC_{C,}$ a research been done to predict the monomer and dimer model the $PhaC_{Cs}$. Based on the result obtain via site-directed mutagenesis, it has been postulated that mutation of residue Arg409Ala and Arg490Ala of the $PhaC_{Cs}$ disrupt the formation of $PhaC_{Cs}$ dimer. However, it is completely unknown how this mutation affects the structure, stability, dynamics, and binding function of the $PhaC_{Cs}$ protein. In this research study, molecular dynamics simulation approach has been utilized to address these issues. From the simulation analysis of the native and proposed mutant structures of both monomer and dimer of $PhaC_{C,}$ it has been found that mutation of R409A and R490A has disrupted the hydrogen bond interaction of His404 and Gly480 and the salt bridge interaction between Arg490 and Asp488 respectively which leads to the disorientation of the catalytic triad position of $PhaC_{Cs}$. In addition, the results also demonstrated that there is no direct involvement of Arg409 and Arg490 in dimerization of $PhaC_{Cs}$.

CHAPTER 1 - INTRODUCTION

Polyhydroxyalkanoate synthase (PhaC) is an essential enzyme in the biosynthesis of polyhydroxyalkanoate (PHA), an eco-friendly and sustainable bioplastic synthesized by a wide range of bacteria and archaea. PHA is an excellent substituent for the traditional petroleum-based plastics that are currently available in the market. PhaC polymerises hydroxyalkanoates (HAs) monomers from coenzyme A thioesters of hydroxyalkanoates (HA-CoA). PhaC affects the monomer composition, substrate specificity and molecular weight of polymers produced that resulted in the synthesis of a variety of PHAs ranging from strong but brittle 3-hydroxybutyrate (3HB) homopolymer to the flexible copolymer of 3HB and 3-hydroxyhexanoate (3HHx). Thus, PhaC has been the subject of numerous research that focused on increasing the efficiency of substrate incorporation and production of PHAs with desirable properties for domestic and industrial use.

Polyhydroxyalkanoate synthases (PhaCs) were generally classified into four major classes based on their subunit compositions and substrate specificity. Class I and Class II PhaC were reported to exist in homodimeric form while Class III and Class IV form heterodimers. In terms of substrate specificity, Class I, III and IV favour short-chain-length (SCL) monomers comprising $C_3 – C_5$ carbon lengths, whereas Class II PhaC has a tendency towards medium-chain-length (MCL) monomers composed of $C_6 – C_{14}$ carbon lengths. Accordingly, Class I PhaC from *Cupriavidus necator* (also known as *Ralstonia eutropha*) ($PhaC_{Cn/Re}$), Class II PhaC from *Pseudomonas aeruginosa* ($PhaC_{Pa}$) and Class III PhaC from *Allochromatium vinosum* ($PhaEC_{Av}$) are the well studied and usually referred as prototypical synthases.

Interestingly, for the past 30 years, the quest for crystallization of the PhaC structure is still ongoing due the nature of the protein that is more than 40 kDa in size (Amara and Moawad, 2011, Liebergesell and Steinbűchel, 1992, Müh et al., 1999). The lack of a crystal structure of PhaCs has limited our understanding the detailed structural properties and mechanism of catalysis. However, implementation of bioinformatics approaches on predicting 3D structures of PhaCs has provided new insights on catalytic properties such as the existence an oxyanion hole in PhaC from Type II *Pseudomonas* sp. USM 4-55 and structural conformation of PhaC of *Pseudomonas* sp. LDC-25 that has never been reported using biochemical studies.

The application of molecular dynamics simulation studies using static three-dimensional protein structures has been reported to provide detailed understanding on the physical basis of dynamic trajectory of the complex macromolecule structures and their biological associations such as protein – protein, protein – ligand and protein – nucleic acid interactions. Molecular dynamics simulation technique is an excellent tool to facilitate the interpretation of experimental data and also to understand the outcome of an experimental work. Hence, this technique will be suitable to study PhaCs in order to gain more structural insight on the dimerization of the protein that almost has not been reported before.

Recently, Bhubalan et al (2011), has successfully identified highly active wild-type polyhydroxyalkanoate synthase from *Chromobacterium* sp. USM2 ($PhaC_{Cs}$) that has the ability to polymerize three different monomers (3-hydroxybutyrate, 3-hydroxyvalerate and 3-hydroxyhexanaote) with enzymatic activity 5-fold higher than prototypical PhaC from *Cupriavus necator* H16. This naturally occurring and highly

active PhaC with superior polymerizing ability makes it an excellent candidate for molecular dynamics simulation studies.

In a separate study, Jamil et.al (2013), had predicted the three dimensional model of PhaC$_{Cs}$ monomer and dimer structures using I-TASSER and Cluspro 2.0 servers respectively. Based on the interaction analysis of residues involved in the prediction of the dimer structure, it has been postulated that arginine residues at position 409 and 490 of the structure contributed to the dimer formation. These findings eventually lead to an experimental work, in which site-directed mutagenesis of arginine residue to alanine residue was performed to create mutants R409A, R490A, and R409A-R490A of PhaC$_{Cs}$. Subsequently, the *in-vitro* and *in-vivo* analysis showed that there is no accumulation of P(3HB) granules in all the mutants in contrast to wild-type PhaC$_{Cs}$.

In this work, the above mentioned mutations were introduced to the same predicted PhaC$_{Cs}$ monomer and dimer structures by *in-silico* mutation method and the effects were investigated via molecular dynamics simulation and MM-PBSA analysis. Analysis of the results obtained from the molecular dynamics simulations is mainly focused on the impact of mutations on the structural and dynamic behavior of monomer and dimer of PhaC$_{Cs}$. Although extensive biochemical studies have been reported in recent past on PhaCs, however, application of *in-silico* approaches is still limited. Hence, this study attempts to perform molecular dynamics simulation of the predicted monomer and dimer structures of phaC$_{Cs}$.

The objective of this study are:

i.) To study the dynamic of computationally predicted monomer and dimer structure of polyhydroxyalkanoate synthase from *Chromobacterium* sp. ($PhaC_{Cs}$)

ii.) To investigate the effect of the site-directed mutagenesis of proposed Arg409 and Arg490 residues on the dimerization of $PhaC_{Cs}$.

CHAPTER 2 - LITERATURE REVIEW

2.1 Polyhydroxyalkanoate (PHA)

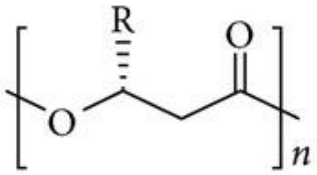
2.1.1 Introduction

Polyhydroxyalkanoic acids or commonly known as polyhydroxyalkanoates (PHAs) are eco-friendly polyesters that naturally synthesised by vast species of microorganisms (Suriyamongkol et al., 2007). PHAs are accumulated as hydrophobic granules inside the cell under the presence of excess carbon source due to the limited supply of other nutrients such as oxygen, phosphorus or nitrogen that are essential for the growth of the microorganism. The polyester granules function as carbon and energy storage inclusion that will be degraded by PHA-depolymerase enzymes to utilise the carbon as energy during carbon starvation (Laycock et al., 2013, Rehm, 2007, Willis et al., 2008). PHAs are reported to be synthesised by both Gram-positive and Gram-negative bacteria and more than 300 diverse microorganisms are known to produce and accumulate the PHAs intracellularly (Keshavarz and Roy, 2010).

Earliest studies on PHAs dates back to 20th century. French microbiologist, Maurice Lemoigne, is the pioneer who discovered the bacterial polyester granule composed of poly(3-hydroxybutyrate), P(3HB) in a Gram-positive bacterium, *Bacillus megaterium* (Lemoigne, 1926). In 1974, new constituents of PHAs, 3-hydroxyvalerate (3HV) and 3-hydroxyhexanoate (3-HHx) were found to exist in chloroform extracts of activated sewage sludge as reported by Wallen and Rohwedder (Wallen and Rohwedder, 1974). To date, however, there are approximately 150 different constituents of hydroxyalkanoic acids (HAs) has been discovered (Shrivastav et al., 2013). In general, PHAs are composed of the repeating unit of (R)-3-hydroxyalkanoic acid monomers that either homopolymer or heteropolymer by

forming an ester bond with the hydroxyl group of the adjacent monomers.The common structure of PHA is shown in Figure 2.1, where R represent the hydroxyl-substituted carbon atom during polymerization of the monomers. The side chain alkyl group at the C-3 atom that chemically known as the *β*-position can vary in length of carbon atom from methyl (C1) to tridecyl (C16) that determine the identity of the monomer that enables the classification of the types of PHAs. The side chain of the alkyl group could be made up of saturated, unsaturated, aliphatic or aromatic functional groups. Based on the number of carbon atoms in monomer units, the bacterial PHAs are ramified into two groups, namely short chain length PHAs (SCL-PHAs) and medium chain length PHAs (MCL-PHAs) that composed of three to five carbon atoms (C_3 to C_5) and six to sixteen carbon atoms (C_6 – C_{16}) in each monomer units respectively. Examples of SCL-PHAs are homopolymers, poly(3-hydroxybutyrate) and poly(3-hydroxyvalerate), P(3HV) whereas MCL-PHAs represented by homopolymers, poly(3-hydroxyhexanoate), P(3HHx), poly(3-hydroxyoctanoate), P(3HO) and heteropolymer, polyhydroxyhexanoate-*co*-polyhydroxyoctanoate, P(3HHx-*co*-3HO) (Amara, 2008, Castilho et al., 2009, Keshavarz and Roy, 2010, Madison and Huisman, 1999). Formation of the PHAs is highly influenced by the carbon source availability and substrate specificity of PHA synthase, the enzyme that polymerises hydroxyl acyl-CoA into the polymer (Somleva et al., 2013).

The physicochemical properties of the PHAs such as hydrophobicity, melting point, glass transition, temperature and degree of crystallinity vary based on the composition of the monomer chains in the polymer (Patel et al., 2005). For example, copolymer P(3HB-co-3HV), has lower crystallinity and decreased stiffness, being



Poly(3-hydroxyalkanoates) [PHA]

R group		
$-CH_3$	Poly(3-hydroxyalkanoates)	PHA
$-CH_2\text{-}CH_3$	Poly(3-hydroxyvalerate)	PHV
$-(CH_2)_2-CH_3$	Poly(3-hydroxyhexanoate)	PHHex
$-(CH_2)_4-CH_3$	Poly(3-hydroxyoctanoate)	PHO
$-(CH_2)_6-CH_3$	Poly(3-hydroxydecanoate)	PHD
$-CH_2-C_6H_5$	Poly(3-hydroxy-5-phenylvalerate)	PHPV

Figure 2.1 The common structure of polyhydroxyalkanoate (PHA) (adapted from Castilho et. al, 2009)

more flexible and tougher than the homopolymer, P(3HB) (Lee, 1996). The physical functionality of the PHAs can be improved by the addition of other material and additive such as bio-based plasticizers (Chen and Luo, 2009, Ha and Cho, 2002, Narayan, 2006, Yu et al., 2006). Biodegradability and biocompatibility are the most attractive features of the PHAs compared to the conventional plastics that are available in markets. This has made application of PHAs be diversified into various industrial fields such as fermentation, materials, medicine, and biofuels (Chen, 2010, Shrivastav et al., 2013).

Although, PHAs has more advantages compared to the petroleum-based plastics, however, high-cost production has been a limiting factor in the commercialization of the PHA-based plastics (Bernard, 2014, Chanprateep, 2010, Varsha and Savitha, 2012, Wang et al., 2014). The production cost of PHAs in industrial scale is still irrelevant in contrast to commodity prices of petroleum-based plastics such as polyethylene and polypropylene that are below the US $ 1 kg^{-1} (Chee et al., 2010). Hence, numerous research is still being performed to find new PHA producing bacterial strains and alternative carbon substrate to yield more PHAs. Furthermore, intense studies are focused on PHA synthase (PhaC) protein activity to reduce the PHA production cost (Jia et al., 2016, Numata et al., 2012).

2.1.2 Types of PHA

Generally, PHA can be divided into three main types based on the monomer that incorporated in the polymer (Lee, 1996b). Differences in PHA monomer composition can affect the properties and qualities of the polymer. Table 2.1 shows all types of bacterial PHA.

The first type is short chain length PHA (SCL-PHA) that consists of 3-HA monomer of 3-5 carbons. Example of SCL-PHA is poly-3-hydroxybutyrate, P(3HB) from *Wautersia eutropha* (Steinbuchel et al, 1998). SCL homoplomer like P(3HB) will form a stiff crystalline material and this would lead to brittleness of the polymer and low extension to break (Nomura et. al.,2004). Therefore, SCL-PHA can be used as thermoplastics due to its characteristics.

Second type of PHA is medium chain length PHA (MCL-PHA) that consists of 6-14 carbons in monomeric subunit, this PHA are semicrystalline, thermoplastic, elstomeric that could be further reinforced by addition of non-composite materials and fillers. MCL-PHA can be found in *Pseudomonas oleovorans* and *Pseudomonas putida* (Kim et. al, 2007).

Another type of PHA is short chain length-medium chain length PHA (SCL-MCL-PHA) (Nomura et. al.,2004). It consists of SCL and MCL subunits that consists of 4-14 carbons in single monomer. SCL-MCL PHA has wide range of properties that depend on the percentage of composition of different monomers incorporated into the polymer. It will be in more elastic form when there is low percentage of SCL monomer. On the other hand, when it has high level of SCL monomer, it will tend to be more flexible like polypropylene (Nomure et. al., 2004).

Table 2.1 Type of bacterial PHA

Type of PHA	Number of C atoms	Example	Representative bacterium
SCL	C_3-C_5	PHB	*Ralstonia eutropha*
MCL	C_6-C_{14}	PHO	*Pseudomonas sp.*
SCL-MCL	C_3-C_{14}	P(HB-*co*-HO)	*Pseudomonas sp.* 61-3

2.1.3 Metabolic pathways of P(3HB)

PHA biosynthetic pathway consist of three enzymatic reactions which are catalysed by the three distinct enzymes (Anderson and Dawes,1990; Madison and Huisman,1990). The enzymes are β-ketothiolase (phaA), acetoacyl-CoA reductase (phaB) and PHA synthase (phaC). These enzymes encoded by phaA, phaB, and phaC genes respectively in arrangement of phaCAB. The key enzyme for the production of PHA is PHA synthase (PhaC), that functions in polymerization of 3-HA monomer (Tsuge et. al, 2005). The first reaction in the biosynthesis of P(3HB) is the condensation of two acetyl coenzyme A (acetyl-CoA) by β-ketothialase to form acetoacyl-CoA. Acetoacyl-Coa can be synthesised in all living organisms. This is the reason for the wide spread presence of PHA in different environmental samples (Anderson & Huisman, 1999). The aectoacyl-Coa is then reduced to to generate (R)-hydroxylbutyryl-CoA by acetoacyl-CoA reductase. In the last step, the (R)-hydroxylbutyryl-CoA monomers are polymerized by PHA synthase to form homopolymer PHA. Figure 2.2 shows the illustration for biosynthesis pathway of P(3HB).

2.1.4 Physical and thermal properties of PHA

PHA can be found as small insoluble inclusions in cytoplasm of bacteria. Size of PHA is in the range of 0.2μm to 0.5 μm in diameter (Sudesh et. al,2000). It has various molecular weights, depending on various factors such as the type o bacterial strain, culture condition, and carbon source used in culture but the molecular weight is generally between 50,000 to 100,000 kilo Daltons (Madison and Huisman, 1999). PHA granules can be seen using phase contrast light microscope due to its high refractivity

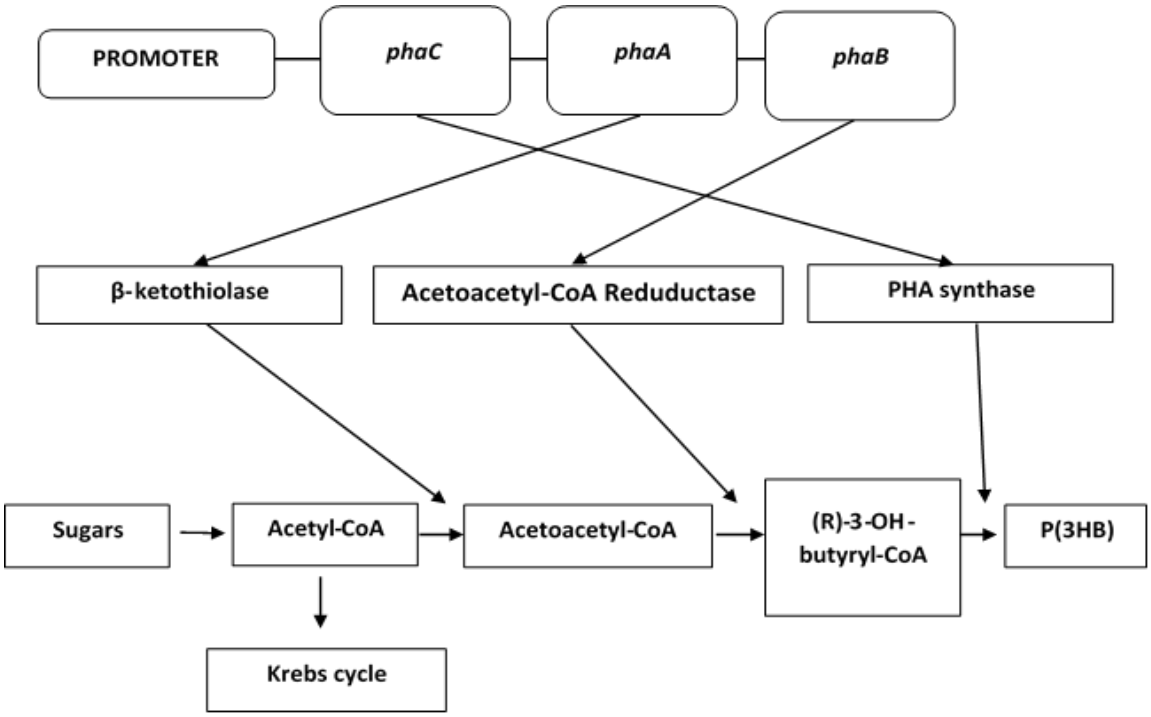


Figure 2.2 Biosynthethic pathway of P(3HB) polymer (Adapted from Madison and Huisman, 1990).

P(3HB) homopolymer has higher stiffness (3.5 GPa) and tensile strength (43MPa) as compared to conventional plastics such as polypropylene (1.7 GPa and 38 MPa respectively). Nevertheless, the percentage of elongation to break for polypropylene (400%) is higher than P(3HB) (5%). Other than that, P(3HB) have the same melting temperature with polypropylene, making this polymer behave as thermoplastic.

The addition of MCL 3-HA monomers into P(3HB) to form new copolymers will very much improved the physical properties of the polymer films (Khanna and Srivastava, 2005). Therefore, the properties of the resultant copolymer can be regulated by changing the monomer composition and content of a given P(3HB) monomer. This will results in copolymers that are stronger and stiffer or softer and more elastic.

2.1.5 Industrial application of PHA

PHAs are very useful in today's world because to their biodegradable and biocompatible properties. It can be used in wide range of applications. Some of the ares that meet the market needs are the packaging, medical, and disposable personal hygiene products (Poirier et. al, 1995; Rehm and Steibuchel, 1999).

To date, research shows that PHAs have potential in cosmetic and skin care applications and waste water treatment (Sudesh et. al, 2011). Facial oil blotting that are made from PHA cast films such as P(3HB), P(3HB-*co*-3HV) and P(3HB-*co*-3HHx) copolymer is able to absorb sebum on the face without the addition of additives.

Elastomeric MCL-PHAs are suitable for biomedical application due to their physical properties (Kim et. al., 2007). In 1996, Lee reported that the degradation product of PHB, D-(-)-3-hydroxybutyrate were detected relatively in large amount in human blood plasma (Lee, 1996a). Thus it shows that implanting PHB in mammalian tissues is applicable without any fatal risk (Ojumu et. al., 2004). Other possible application of bioplastic in medical and pharmaceutical fields are as biodegradable carriers for long term dosage of drugs inside the body, surgical pins, sutures, swabs, wound dressing, bone replacement and implant artificial blood vessels (Chen and Wu, 2005; Chen et. al, 2005; Verlinden et. al., 2007)

In plastic packaging industries, it is widely used to produce containers, bottles, and razors (Bucci et. al., 2005; Bucci et. al., 2007). One property of PHB films that make it possible to be used for food packaging is its relatively low oxygen diffusion. Plastic from PHA are proven to be degraded in both aerobic and anaerobic condition. P(3HB-HV) can be used in the production of films, bottels, and as hydrophobic coating on paper (Khanna and Srivastava, 2005).

In addition, PHA can be used in agricultural industries. It is proven that PHA can be degraded in soil. Therefore, the use of PHA is widely acceptable as biodegradable carrier for long term usage of insecticides, herbicides, fertilizers, seedling containers, and plastic sheaths protecting saplings and tubing for crop irrigation (Verlinden et. al., 2007)

2.2 PHA synthase

PHA synthase is the key enzyme in the PHA biosynthesis. It uses coenzyme A (CoA) thioesters of hydroxyalkanoic acids (HA) as substrates and catalyzes the polymerization of HA into PHA with the release of CoA (Rehm and Steinbuchel,

1999). Till now, there are over 88 PHA synthase genes that have been cloned from 68 bacteria species (Rehm, 2007).

2.2.1 Classification of PHA synthase

There are four classes of PHA synthase (PhaC) that have been distinguished. Class 1 PHA synthase has only one subunit. The size of PhaC is in range between 61 – 73 kilo Daltons and it utilizes SCL-3HA-CoA monomer to form SCL-PHA polymer. Polyhydroxyalkanaote synthase from *Ralstonia eutropha* is categorized as Class 1 synthase.

Class II consists of 2 similar subunit, PhaC1 and PhaC2. Both subunits are approximately around 60 kDa. Class II PHA synthase are preferentially active towards CoA thioesters of various MCL 3-HA to produce MCL-PHA polymers. It comprises of 6 to 14 carbon atoms. Class II PHA synthase often found in *Pseudomonas sp*

Class III and Class IV PHA synthase made up of two different subunits. Class III consists of PhaC and PhaE while class IV consist of PhaC and PhaR subunits (Sheu and Lee, 2004). The PhaC subunits in these classes have similarity around 21% to 28% towards class I and class II but the size is smaller than both class I and class II synthase which is only 40 kDa. The PhaE subunits from class III which is around 40kDa has no similarity to any other PHA synthases (Liebergesell et al., 1993) while PhaR subunit from class IV is the smallest subunits with the size around 22 kDa. Both classes are preferentially accumulates SCL-PHA polymer. Example of class III synthase is PHA synthase from *Thiocapsa pfennigii* and PHA synthase from *Bacillus megaterium* is categorized as class IV synthase (Rehm, 2003).

2.2.2 Biochemistry of PHA synthase

PHA synthase is made up of two main domains, N-terminal domain and C-terminal domain.The N-terminal of PHA synthase is approximately the first 100 amino acids relative to Class I PHA synthase (for PhaC from *Cupriavidus nector*). This terminal is highly variable among the PHA synthase from different classes and bacteria (Rehm and Steinbuchel, 1999).

The C-terminal domain is the catalytic domain whereby the polymerization of PHA mainly take place. It consists of α/β hydrolase fold that contains the catalytic triad. These regions is made up of highly conserved lipase-like box (GX[S/C]XG) that forms the active site of the protein. In PHA synthase, serine in the lipase-like box is replaced by Cysteine residue. Cysteine residue is located at the strand-elbow-helix motif that commonly known as nucleophilic elbow plays vital in catalytic activity of the synthase.(Gerngross et. al, 1994; Jia et. al, 2001).

2.2.3 Dimerization of PHA synthase

It has been proven by Zhang and his co-workers that PHA synthase is active in homodimeric form. (Zhang et. al., 2000). PHA synthase homodimer is formed when two monomers of were bind together at the specific sites on their surface to be activated and perform their function if polymerization of PHA polymer. In addition, Jia and colleagues also suggested that dimeric form of PHA synthase is involved in elongation phase of PHA polymer (Jia et. al., 2001). Figure 2.3 illustrates the mechanistic model of PHA synthase dimer model a proposed by Zhang for the polymerization of P(3HB) from PHB synthase from *Ralstonia eutropha* (Zhang et. al., 2000). This model can also be a reference for other Class I PHa synthases as well. From this model, it was proposed that two PHA synthase monomers bound together and form homodimer of

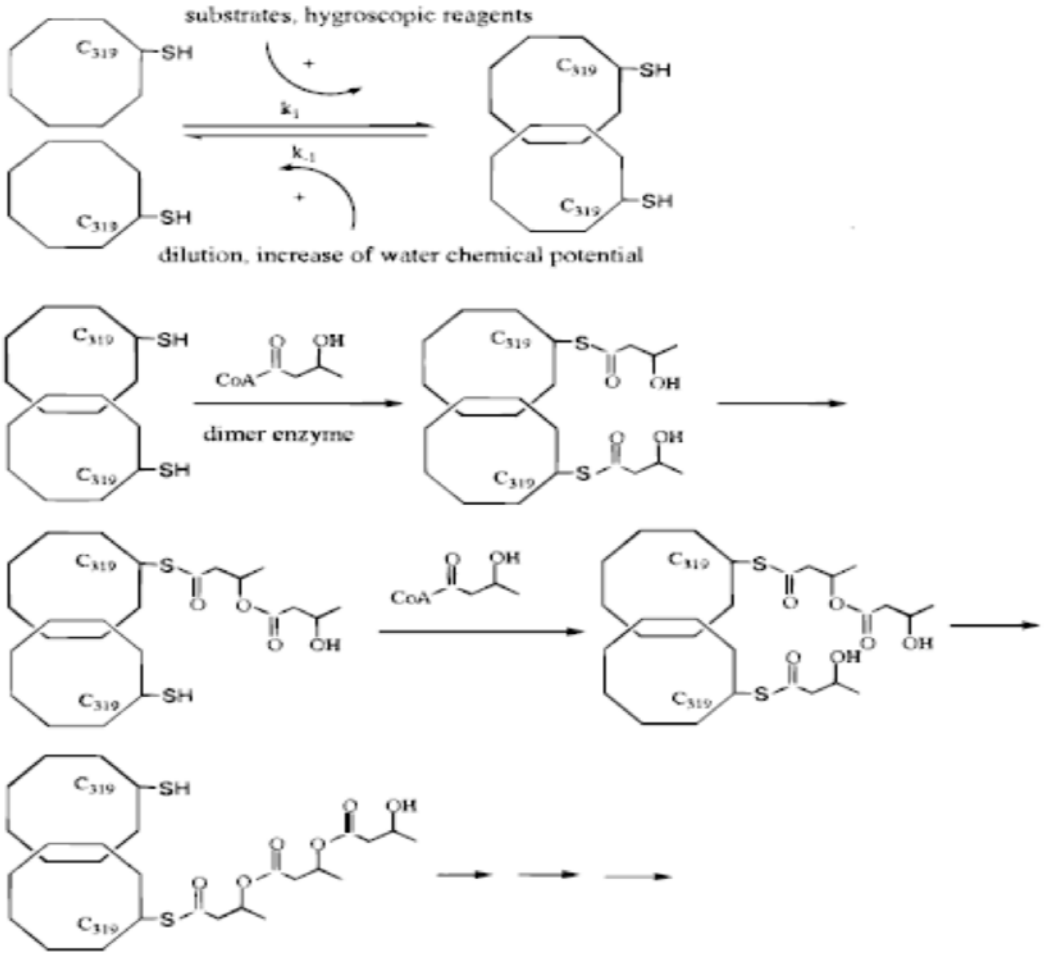


Figure 2.3 Mechanistic model proposed by Zhang et al for the polymerization of P(3HB) from PHB synthase from *Ralstonia eutropha*. (dapted from Zhang et. al., 2000).

PHA synthase consisting of two thiol group (Cys319). Subsequently, the addition of 3HB-CoA molecules initiates the polymerization of P(3HB). The mechanism starts with the initiation step that involves the binding of two 3HB-CoA substrates to two Cys319. Afterward, the first elongation reaction occurs by the formation of free thiol group via reaction of β-hydroxyl group on either substrates. The free thiol group then accepts another 3HB-CoA substrates and the same process continues. During the elongation step,only one binding site will be available in the homodimeric from of PHA synthase. This is because another binding site will be covalently bonded with the growing polymer chain.

2.2.4 Polyhydroxyalkanaote synthase from *Chromobacterium* sp. USM2

Polyhydroxyalakanote synthase from *Chromobacterium* sp. USM2 ($PhaC_{Cs}$) is highly active wild-type PhaC that capable of utilizing 3-hydroxybutyrate (3HB), 3-hydroxyvalerate (3HV), and 3-hydroxyhexanoate (3HHx) monomers in PHA biosynthesis. $PhaC_{Cs}$ exhibits superior polymerizing activity compared to the other proptotypical synthases and also genetically engineering mutant synthases. $PhaC_{Cs}$ has the ability to produce PHAs containing monomers of mixed chain lengths that highlights the potential of the synthase to be developed as a model synthase to compare the activities of other polyhydroxyalkanaote synthases (Bhubalan et al., 2011).

2.3 Molecular Dynamics Simulation

2.3.1 Introduction of Molecular dynamics (MD)

Molecular dynamics is the science of stimulating the motion of particles, can provide the ultimate details concerning individual particle motions as a function of time (Karplus and Petsko, 1990). It is a valuable tool for understanding the physical basis of structure and function of biological macromolecules. It can be used to address specific question about the properties of a model system that are difficult to access

experimentally (Karplus and McCammon, 2002). MD simulation typically start from a three-dimensional protein structure determined by X-ray crystallography, NMR or theoretical models and use an empirical forcefield to approximate the potential energy surface of a protein. Once the force field has been specified, the time evolution of the system is determined by Newton's equations of motion using a suitable integration scheme for a required number of steps in a particular ensemble. It has been possible to efficiently perform simulations on the timescale over which many physiologically relevant processes in the recent years and this has expanded the set of problems for which MD is applicable.

2.3.2 Force fields

Force field is a set of functional parameters that provide information on potential energy of a system of particles. These parameters are obtained from experimental and quantum mechanical studies of small molecules. Mainly, force field function consist of bonded and non-bonded interaction terms. Bonded interactions include energy of bond length, bond angles, and sometimes improper dihedrals and torsional dihedral angles, while Van der Waals interactions and electrostatic interaction contribute to non-bonded interactions. There are multiple force fields available, and selection of the most adequate force field for a specific application is essential in simulation studies. For biomolecular simulations widely used molecular force fields include AMBER96 (Kollman, 1996) , CHARMM27 (Foloppe and MacKerell, 2000), OPLS-AA (Kaminski et al., 2001), and GROMOS96 (Oostenbrink et al., 2004).

2.3.3 Essential simulation steps

2.3.3 (a) Simulation environment

Protein simulation is done to replicate the experimental conditions hence several parameters for the different physical conditions are considered (such as pressure, temperature). In general, protein simulation is done in canonical ensemble (NVT), particularly the initial equilibration steps, or isothermal-isobaric (NPT) ensemble. For simulation, the protein molecules should be kept in the unit cell and solvated with explicit solvent. There is a broad range of explicit water models available and most popular of these models include TIP3P, TIP4P (Jorgensen et al., 1983), TIP5P (Mahoney and Jorgensen, 2000), SPC, and SPC/E (Berendsen et al., 1987).

2.3.3 (b) Energy minimization

This involves finding global minimum energy with respect to the position of side chain atoms that represents the geometry of the particular arrangements of atoms in which the net attractive force on each atom reaches a maximum. Widely used method to compute the minimum energy are steepest descent and conjugate gradient.

2.3.3 (c) Heating the system and equilibration

Each atom of the system is assigned initial velocities at (0 K) during energy minimization and Newton's equations of motion that represent the time evolution of system are numerically integrated. At short predefined intervals, new velocities are assigned corresponding to a slightly higher temperature and the simulation is allowed to continue until desired temperature is achieved. Force constrains on different subdomains of the simulation system are gradually removed as structural tensions dissipate by heating. Equilibrium phase is used to equilibrate kinetic and potential energies. This functions to remove the constraints on the protein so that the whole system can evolve in time.

2.3.3 (d) Production phase

This the final phase of the simulation to remove the constraints on protein. It is carried out for desired time scale to generate trajectory of protein molecule. The timescale of the simulation can be varied from several hundred picoseconds to microseconds or more. The trajectories obtained from this stage is used for further analysis.

CHAPTER 3 - MATERIALS AND METHOD

3.1 PhaC$_{Cs}$ wild-type monomer and dimer structure

The PhaC$_{Cs}$ wild-type monomer and dimer model structure used in this research study were obtained from the M.Sc disertation of Jamil et. al (2013).

3.2 Hardware and softwares used

3.2.1 Hardware

Molecular dynamics simulation procedures were performed on Hypercube High Computing Server Clusters provided by Universiti Sains Malaysia. The server runs on CentOS 6.5 operating system with Intel(R)Xeon(R) CPU E6545@2.4 GHz. The visualization of protein models, results analysis, image rendering and graph plotting was carried out using Dell Workstation running on Ubuntu 10.04 LTS operating system with Intel(R)Xeon(R) E5620@2.4 GHz.

3.2.2 Softwares

The softwares used in this research are GROMACS 4.6.5 simulation package to run the simulations, InterProSurf web server for identification of protein dimerization interface residues on monomer structure, Protein Interaction Calculator for analysis of hydrophobicity interaction in dimer structure,Caver pluin to analyse substrate channel and g_mmpbsa tool for binding energy calculation for dimer structures,UCSF Chimera for *in-silico* mutation, PyMOL version 1.7.4.5 and Visual Molecular Dynamics (VMD) for visualization and image rendering, and GRACE version 5.1.2.3 for graph plotting (Turner and Stambulchik, 1998). Additional details on InterProSurf and Protein Interaction Calculator are described as follows:

InterProSurf (http://curie.utmb.edu/prosurf.html) (Negi et al., 2007).

Wild-type PhaC$_{Cs}$ dimer protein structure was uploaded in the server to identify the interface residues.

Protein Interaction Calculator (http://pic.mbu.iisc.ernet.in/) (Tina et al., 2007).

Average wild-type and mutant PhaC$_{Cs}$ dimer structures obtained from the final 10 ns simulation was uploaded to the server to analyse the residues involved in hydrophobic interactions.

CAVER (https://caver.cz/) (Chovancova et al., 2012)

Average wild-type and mutant PhaC$_{Cs}$ monomer and dimer structures obtained from final 10 ns simulation was analysed using the plugin installed in Pymol to identify the substrate channel in PhaC$_{Cs}$ structure.

g-mmpbsa (Kumari et al., 2014).

The energy components of each component were calculated for 100 snapshots extracted every 0.25 ps from the production trajectory from 40 ns to 50 ns in trajectory file format (.xtc). Other parameters used by g_mmpbsa for the energy calculation were set as recommended in the developer's publication (Kumari et al., 2014)

3.3 Visualization of protein models

The visualization of the 3D structure of the PhaC$_{Cs}$ monomer and homodimer models was performed using Pymol software.

3.4 *In-silico* mutation of the PhaC$_{Cs}$ monomer and dimer models

Single and double mutation of Arginine residue to Alanine at residue position 409 and 490 of monomer and dimer protein of PhaC$_{Cs}$ were performed using Rotamer

wizard available in UCSF Chimera tool (Pettersen et al., 2004). The side chains rotamers of mutated residues were selected based on the dynameomics rotamers library (Scouras and Daggett, 2011). For PhaC$_{Cs}$ dimer structure, the mutation was performed on both the monomer chains. The resultant mutant structures were labeled as R409A, R490A and R409A-R490A mutants for both monomer and dimer structures based on the mutated residue position in each structure. Hence, this resulted in a total of eight protein structures including the wild-type structure of PhaC$_{Cs}$ monomer and dimer to be used as starting models for molecular dynamics simulation. Furthermore,

the numbering of residues in Chain B of all dimer structures has been assigned to be in continuous after the last residue of Chain A in order to be able to create two separate topology files that can be read during simulation. Hence, the residues in Chain A are numbered as residue 1 to residue 567 whereas the residues in Chain B are numbered as residue 568 to residue 1134 without any changes in amino acids.

3.5 Molecular dynamics simulations of wild-type and mutants of PhaC$_{Cs}$ monomer and dimer protein model in explicit water conditions.

In this study, molecular dynamics simulation was performed for a time period of 50 ns in explicit water condition with a periodic condition applied in all dimension for both the monomer and dimer structure of PhaC$_{Cs}$ that included the wild-type and mutants. All the simulation were run separately for each protein models using standalone open source GROMACS 4.6.5 software package (Pronk et al., 2013).The simulations were performed by adhering to the standard protocols of basic molecular dynamics simulation that comprise of energy minimization, two phase equilibration, and finally production stage of simulation of protein systems (Hess et al., 2013). The workflow used in this study was summarized in Figure 3.1.