Research Article

Crystal structure of *Ralstonia eutropha* polyhydroxyalkanoate synthase C-terminal domain and reaction mechanisms

Jieun Kim\(^1,3\)
Yeo-Jin Kim\(^1,3\)
So Young Choi\(^2,3\)
Sang Yup Lee\(^2\)
Kyung-Jin Kim\(^1\)

\(^1\)School of Life Sciences, KNU Creative BioResearch Group, Kyungpook National University, Daegu, Republic of Korea
\(^2\)Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering (BK21 Plus Program), BioProcess Engineering Research Center, Center for Systems and Synthetic Biotechnology, and Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
\(^3\)Co-first author

**Correspondence:** Prof. Kyung-Jin Kim, School of Life Sciences, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 41566, Republic of Korea.

**E-mail:** kkim@knu.ac.kr

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/biot.201600648.

Submitted: 26-Oct-2016
Revised: 31-Oct-2016
Accepted: 02-Nov-2016

This article is protected by copyright. All rights reserved.
Additional Correspondence: Prof. Sang Yup Lee, Department of Chemical and Biomolecular Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea.

E-mail: leesy@kaist.ac.kr

Keywords: Crystal structure; Enzyme mechanism; PHA synthase; Polyhydroxyalkanoates; Ralstonia eutropha

Abbreviations: CD, circular dichroism; D-loop, dimerization loop; EC-region, extended C-terminal region; GC, gas chromatography; HTH, helix-turn-helix; PHA, polyhydroxyalkanoate; PHB, poly(3-hydroxybutyrate); PISA, Protein Interfaces, Surfaces and Assemblies; PS-region, protruding structure region; SEC, size-exclusion chromatography; 3HV, 3-hydroxyvaleryl-CoA; PHBHV, poly(3-hydroxybutyrate-co-3-hydroxyvalerate); 3HH, 3-hydroxyhexanoyl-CoA; PHBHH, poly(3-hydroxybutyrate-co-3-hydroxyhexanoate)
Abstract

Polyhydroxyalkanoates (PHAs) are natural polyesters synthesized by numerous microorganisms as energy and reducing power storage materials, and have attracted much attention as substitutes for petroleum-based plastics. Here, we report the first crystal structure of *Ralstonia eutropha* PHA synthase at 1.8 Å resolution and structure-based mechanisms for PHA polymerization. *RePhaC1* contains two distinct domains, the N-terminal (*RePhaC1_ND*) and C-terminal domains (*RePhaC1_CD*), and exists as a dimer. *RePhaC1_CD* catalyzes polymerization via non-processive ping-pong mechanism using a Cys-His-Asp catalytic triad. Molecular docking simulation of 3-hydroxybutyryl-CoA to the active site of *RePhaC1_CD* reveals residues involved in the formation of 3-hydroxybutyryl-CoA binding pocket and substrate binding tunnel. Comparative analysis with other polymerases elucidates how different classes of PHA synthases show different substrate specificities. Furthermore, we attempted structure-based protein engineering and developed a *RePhaC1* mutant with enhanced PHA synthase activity.
1 Introduction

Polyhydroxyalkanoates (PHAs) are natural polyesters consisting of various hydroxycarboxylic acids that are synthesized intracellularly as distinct granules by numerous bacteria, including species of the genera *Ralstonia*, *Bacillus*, and *Pseudomonas* [1-4]. PHAs are accumulated as an energy storage material when cell growth is hampered by nutrient limitation in the presence of excess carbon [1-4]. PHAs exhibit thermoplastic and/or elastomeric properties depending on the monomer composition, as well as biodegradability and biocompatibility. Thus, PHAs have attracted much attention as substitutes for petroleum-based plastics and elastomers in packaging, agricultural, medical, and other industrial applications [1-4].

In PHA biosynthesis, PHA synthase (PhaC) is the key enzyme responsible for polymerizing various coenzyme A thioesters of hydroxyalkanoates [1-4]. One of the prominent characteristics of PHA synthase is its broad substrate specificity; to date, more than 150 coenzyme A thioesters of hydroxyalkanoates, including 3-, 4-, 5-, and 6-hydroxycarboxylate, have been reported as substrates of PHA synthases. PHA synthases are divided into four classes according to the carbon chain length of the preferred substrate and the enzyme subunit type.

PHA synthase plays important roles in determining the general characteristics of PHAs produced, including molecular weight, polydispersity, monomer composition, and productivity. It is astounding to see that PHAs from 200 kDa up to 3,000 kDa (and in extreme cases, up to 10,000 kDa) can be synthesized by different PHA synthases [1-5]. In order to understand the fascinating mechanism of polymerization by PHA synthase, there has been great interest in determining the crystal structure of PHA synthase over the last 30 years, unfortunately without success [6-9]. Thus, the characteristics and molecular mechanisms of PHA synthase based on real structural information have been under a dark veil.

Here, we report for the crystal structure of the C-terminal domain of *Ralstonia eutropha* (*Cupriavidus necator*) PhaC at 1.8 Å resolution. While we determined the crystal structure and deposited it to Protein Data Bank under accession code 5HZ2 (deposition date of February 2, 2016), another group independently determined the crystal structure of C-terminal domain containing residues 201-368, 378-589 (residues 369-377 are disordered; structure deposited under accession
Based on the crystal structure, we performed mutagenesis studies for several amino acid residues participating in substrate binding and dimerization and comparative analysis with other class PhaCs. Our structural and biochemical studies provide a structural basis for the mechanisms of PHA polymerization. This work also presents a framework for rational engineering of PHA synthase toward more efficient production of bioplastics from renewable resources and production of polymers composed of desired monomers.

2 Materials and methods

2.1 Purification of RePhaC1

For the expression of full-length RePhaC1 (RePhaC1F), the RePhaC1-coding gene was amplified from *R. eutropha* chromosome H16 by polymerase chain reaction (PCR). The PCR product was subcloned into pET30a (Invitrogen) with a C-terminal 6-His-tag. The resulting expression vector pET30a:RephaC1 was transformed into *E. coli* strain B834, and the cells were grown overnight at 37 °C in Luria-Bertani (LB) medium containing 100 μM kanamycin. At an OD$_{600}$ of 0.7, RePhaC1 protein expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside. After 20 h at 18 °C, the cells were harvested by centrifugation at 4000 x g for 15 min at 4 °C. The cell pellet was resuspended in buffer A (40 mM Tris-HCl, pH 8.0, and 5 mM β-mercaptoethanol [BME]) and disrupted by ultrasonication. The cell debris was removed by centrifugation at 13,500 x g for 25 min, and the lysate was applied to a Ni-NTA agarose column (QIAGEN). After washing with buffer A containing 20 mM imidazole, the bound proteins were eluted with 300 mM imidazole in buffer A. Finally, the trace amounts of contaminants were removed by size-exclusion chromatography (SEC) using a Superdex 200 prep grade (320 mL, GE Healthcare) equilibrated with buffer A. The protein eluted at a molecular weight of approximately 120 kDa, indicating that RePhaC1 protein, with molecular weight of 65.5 kDa, formed a dimer. All purification experiments were performed at 4 °C. The purified protein was concentrated to 30 mg/mL in 40 mM Tris-HCl (pH 8.0) and 5 mM BME. The N-terminal (RePhaC1$_{ND}$) and C-terminal (RePhaC1$_{CD}$) domains of RePhaC1 were prepared using procedures similar to that for RePhaC1$_{F}$. Mutations in RePhaC1 were introduced using the QuikChange kit (Stratagene), and
confirmed by sequencing. The mutant proteins were prepared using similar procedures as for the wild-type protein.

2.2 Crystallization of RePhaC1

Crystallization of the purified RePhaC1 protein was initially performed with commercially available sparse-matrix screens, including Wizard I and II from Emerald Bio, Index from Hampton Research, and Structure Screen 1 and 2 from Molecular Dimensions, using the hanging-drop vapour-diffusion method at 20 °C. Each experiment consisted of mixing 1.0 μL of protein solution (30 mg/mL in 40 mM Tris-HCl, pH 8.0, and 5 mM BME) with 1.0 μL of reservoir solution and then equilibrating the mixture against 0.5 mL of the reservoir solution. RePhaC1 crystals were observed from the following crystallization screening conditions: 1.26 M ammonium sulphate and 0.1 M sodium 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid (HEPES), pH 7.5. After several rounds of crystal improvement, suitable crystals for diffraction experiments were obtained using 1.2 M ammonium sulphate and 0.1 M sodium HEPES pH 7.0 and reached their maximum dimensions of 0.05 x 0.05 x 0.05 mm in 8 weeks.

2.3 X-ray diffraction analysis and structure determination of RePhaC1

The native data set was collected at a resolution of 1.8 Å at the 7A beamline (BL) of the Pohang Accelerator Laboratory (PAL, Pohang, Republic of Korea) using a Quantum 270 CCD detector (ADSC, USA) with a wavelength of 0.9798 Å (Table 1). All data were indexed, integrated and scaled together using the HKL-2000 software package [11]. For structure determination, single-wavelength anomalous dispersion was used with a mercury atom as the anomalous scatterer. Briefly, 2 μL of 12 mM mercury(II) chloride was added to a 2-μL hanging-drop containing crystals. After incubation for 2 h at 20 °C, the crystals were transferred to 2 μL of the original reservoir solution and incubated for an additional 30 min. The mercury-derivative crystal data set was collected at a resolution of 2.0 Å at the 7A BL at the PAL (Pohang, Republic of Korea) using a Quantum 270 CCD detector (ADSC, USA) with a wavelength of 1.0072 Å (Table 1). The peak wavelength for the Hg LIII edge was determined by X-ray
absorption fine structure (XAFS) measurements before data collection. The diffraction data were processed with HKL-2000 [11] and merged and scaled against the native data. Phases were calculated by the single isomorphous replacement with anomalous signal (SIRAS) method including a heavy atom search with SOLVE [12]. The electron density was improved by density modification using RESOLVE [13]. After crystallographic model building and refinement of the Hg-derivative structure, we performed molecular replacement using the Hg-derivative structure as a search model for the refinement of the native structure. Further model building was performed manually using the program WinCoot [14], and refinement was conducted with REFMAC5 [15]. The crystals of RePhaC1 belonged to space group I222, with the following unit-cell parameters: a = 73.246, b = 87.856, c = 136.48 Å, α = β = 90°, and γ = 120°. With one molecule per asymmetric unit, the crystal volume per unit protein weight (VM) was 1.80 Å³/Da, which corresponds to a solvent content of approximately 50.73% [16]. The data statistics are summarized in Table 1. The refinement of the native structure to 1.8-Å resolution was deposited in the Protein Data Bank under PDB code 5HZ2.

2.4 Molecular docking simulations

Molecular docking simulations of CoA and (3-HB)₃ trimer to RePhaC1₇ structures were carried out using Auto dock Vina [17]. The ligand of RePhaC1₇ was prepared with Chemskech [18], and nonpolar H atoms were merged onto both the ligands and the targets using AutoDockTools prior to performing these docking simulations. The grid box was designated with x: 43.146, y: 7.021, and z: 51.932, and the size of the grid was 32.0 × 34.0 × 26.0 Å. The final conformations produced in this simulation were checked using PyMOL software.

2.5 Circular Dichroism (CD)

Far-UV (190–360 nm) CD experiments were carried out in a Jasco J-815 CD Spectrometer (JASCO Corporation, Japan). Scans were recorded at 25 °C between 190 and 360 nm as an average of three scans with 0.5 nm step size, 1.5 s dwell time in a 10 μm pathlength demountable Suprasil quartz cell (Hellma Ltd, UK) and smoothed to obtain the final data of RePhaC1₇ (2 mg/mL) with or without 3-
HB-CoA (1:30 ratio). Spectra were collected at 1.0-nm intervals with a bandwidth of 1 nm in a buffer containing 10 mm potassium sodium phosphate pH 7.0 in a 1-cm quartz cuvette.

2.6 Fluorescence Spectroscopy

Intrinsic tryptophan fluorescence emission spectra of RePhaC1F were recorded on a Hitachi F-7000 fluorescence spectrofluorometer (Edinburgh Instruments) with 2 mg/mL protein in 10 mM potassium sodium phosphate pH 7.0. The spectra were recorded from 300 to 500 nm at an excitation wavelength of 295 nm in the correct spectrum mode of the instrument using excitation and emission band passes of 5 nm each. Structure change with or without CoA analogue (1:30 ratio), acetyl-CoA were monitored at 25 °C.

2.7 PHB polymerization activity assay

The in vitro PHB polymerization activities of PHA synthases and their variants were measured by monitoring the decrease in absorbance at 236 nm due to the cleavage of the thioester bond in 3-HB-CoA [19]. The reaction mixture comprised 800-nM RePhaC1 enzyme, 50 mM sodium phosphate buffer (pH 7.0), and 2 mM DTT, with or without 0.02% (w/v) Triton X-100. The reaction was initiated by adding 30 μM DL-β-hydroxybutyryl-CoA (Sigma) to the reaction mixture. The molar extinction coefficient (ε = 23604500 /M/cm at 236 nm) was used to calculate the activity.

2.8 In vivo PHB biosynthesis in E. coli

For in vivo PHB biosynthesis, E. coli K-12 strain W3110 was used. To construct the PHB biosynthetic pathway in E. coli, the RephaC1 and RephaAB genes were expressed by the pCnCAB plasmid [20]. For the construction of plasmids expressing variants of PHA synthase, the RephaC1 gene was replaced with RephaC1CD (pCnCAB plasmid) or RephaC1 variants harbouring point mutations for site-directed mutagenesis studies. For in vivo polymer production, seed cultures were prepared in 25-mL test tubes containing 10 mL of LB medium at 30 °C overnight in a rotary shaker at 250 rpm. A total of 1 mL of the seed culture was inoculated into a 250-mL flask containing 100 mL of LB medium
supplemented with 20 g/L of glucose. Ampicillin and kanamycin (each at 50 μg/mL) were added to the medium depending on the resistance markers of the employed plasmids. The content and monomer compositions of the synthesized polymers in E. coli cells were analysed by gas chromatography (GC) [21, 22]. The cultured cells were washed twice with distilled deionized water, and the cells were lyophilized. The lyophilized cell pellet was subjected to methanolysis with benzoic acid as an internal standard in acidified methanol with 15% (v/v) sulphuric acid. The resulting methyl ester of 3-hydroxybutyrate was measured by GC. The GC analysis was performed by injecting 1 μL of sample into a Agilent 6890N GC system (Agilent Technologies, Palo Alto, CA, USA) equipped with an Agilent 7683 automatic injector, a flame ionization detector, and a fused silica capillary column (ATTM-Wax, 30 m, ID 0.53 mm, film thickness 1.20 μm, Alltech, Deerfield, IL) using nitrogen as a carrier gas. The GC oven temperature was initially maintained at 80 °C for 5 min and ramped to 230 °C at 7.5 °C/min. The temperature was increased with a gradient of 10 °C/min until reaching 260 °C and held for 5 min. The injector and detector were maintained at 250 °C and 300 °C, respectively.

3 Results and Discussion

3.1 Structure determination and domain characterization of RePhaC1

To elucidate the mechanism of PHA polymerization catalyzed by PHA synthase, it is important to determine its crystal structure. To date, however, crystallization of PHA synthase has not been successful even after numerous attempts by several research groups. After repeated attempts, we were able to determine the crystal structure of PhaC1 from R. eutropha (RePhaC1) at 1.8 Å resolution via single-wavelength anomalous dispersion using a mercury metal ion (Figs. 1A, 1B and Table 1). It was interesting to observe that the N-terminal fragment (Met1-Asp199) was missing from the electron density map. Also, SDS-PAGE analysis of the crystallized RePhaC1 showed a fragment with a molecular weight of ~45 kDa, smaller than the full expected size of 65 kDa (Fig. 1C). These results indicate that RePhaC1 underwent proteolytic cleavage during the two-month crystallization process and only the C-terminal portion of the protein was crystallized; this seems to be the key for successful crystallization of PHA synthase. In vitro proteolytic analysis of purified RePhaC1 also showed that
RePhaC1 underwent proteolysis into two fragments with molecular weights of ~45 and ~20 kDa after incubation at room temperature for several days (Fig. 1D). N-terminal sequencing of the cleaved fragments revealed that proteolysis occurs at Arg192. Based on these results, we suggest that RePhaC1 is composed of two distinct domains, the N-terminal domain (RePhaC1_ND; Met1–Arg192) and the C-terminal domain (RePhaC1_CD; Gly193–Ala589; Figs 1A, 1B, 1C and 1D).

3.2 Monomeric structure of RePhaC1

The overall structure of the RePhaC1_CD monomer can be divided into two subdomains, a core subdomain and a dimerization subdomain (Fig. 1). The core subdomain (Gly193–Phe352 and Leu450–Ala589) exhibits a conventional α/β hydrolase fold. Interestingly, compared with a typical α/β hydrolase fold in which an eight-stranded central β-sheet is surrounded by several α-helices, the core subdomain contains an extended C-terminal region (EC-region; Arg521–Ala589) that is composed of two α-helices (α8 and α9) and four short stretched β-strands (β12–β15). The EC-region wraps the α/β hydrolase fold on the opposite side of the dimerization subdomain and causes the core subdomain to form a more extended central β-sheet (11-stranded) than a typical α/β hydrolase fold (Fig. 1E). Among the secondary structure elements comprising the EC-region, α9 participates in formation of the α/β hydrolase fold and β12-α8-β13 forms a protruding structure region (PS-region) (Figs. 1A and 1E).

Unlike the class I (including RePhaC1) and class II PHA synthases (including PhaC from Pseudomonas sp.), the class III and IV PHA synthases lack the EC-region (Fig. 1A). In the dimerization subdomain (Ala353–Glu449), a two-stranded β-sheet is located at the center, and a helix-turn-helix (HTH) motif and a dimerization loop (D-loop) are located on either side of the β-sheet (Fig. 1E).

3.3 Dimeric structure of RePhaC1

It was previously reported that RePhaC1 undergoes conversion from an inactive monomeric state to an active dimeric state, and the dimeric form is induced by addition of substrates [9]. The crystallographic analysis showed that one RePhaC1_CD molecule exists in the asymmetric unit, and the dimeric structure of RePhaC1_CD can be applied under I222 crystallographic symmetry operation (Fig. 2A). In dimeric RePhaC1_CD, the HTH motif of one RePhaC1_CD subunit is inserted into the D-loop of the
other subunit (Figs. 2A, 2B and 2C). The HTH motif and the D-loop consist of mostly hydrophobic residues and the residues mediate hydrophobic interaction between two monomers (Figs. 2B and 2C). The Protein Interfaces, Surfaces and Assemblies (PISA) analysis [23] also generated the dimeric interface, in which 75 amino acids of one monomer and 74 amino acids of the other monomer are involved. The solvent-accessible surface areas buried in dimerization interface were 2,882.2 and 2,891.0 Å², respectively, corresponding to ~16% of the total surface area of each monomer. To confirm the dimerization mode of the protein, we replaced several residues located at the dimerization interface by bulky hydrophilic residues such as arginine and glutamate and generated the following mutants: \( \text{RePhaC1}^{I357E} \), \( \text{RePhaC1}^{V360E} \), \( \text{RePhaC1}^{E364R} \), \( \text{RePhaC1}^{D401R} \), \( \text{RePhaC1}^{V403E} \), \( \text{RePhaC1}^{V407E} \), \( \text{RePhaC1}^{V408E} \), and \( \text{RePhaC1}^{L412R} \). When size-exclusion chromatography (SEC) analysis was performed as reported previously [9], the wild-type \( \text{RePhaC1} \) eluted as a mixture of a monomeric and dimeric forms. However, the dimeric form was dramatically decreased in all mutants compared the wild-type protein (Figs. 2D and 2E). Moreover, the mutants also exhibited reduced or complete loss of PHA synthase activity (Fig. 2E).

### 3.4 Active site of \( \text{RePhaC1} \)

Several models for the polymerization mechanism of PHA synthase have been proposed based on computational structure prediction and related biochemical studies [6-9]. Although the catalytic triad was correctly predicted without the crystal structure, other aspects of the reaction mechanism need to be revised based on the \( \text{RePhaC1}_{\text{CD}} \) structure we determined. At the core subdomain of each \( \text{RePhaC1}_{\text{CD}} \) monomer, one catalytic site was found. The two catalytic sites of the \( \text{RePhaC1}_{\text{CD}} \) dimer are positioned 33.4 Å apart (Fig. 3A). Also, two substrate-binding tunnels of approximately 13 Å in length were found in the vicinity of the two catalytic sites (Figs. 3A and 3B), suggesting that the polymerization reaction occurs independently at each site.

To elucidate the substrate-binding mode, we tried to determine the structure of \( \text{RePhaC1} \) in complex with either of two substrates, 3-HB-CoA or poly(3-hydroxybutyrate), PHB. However, neither co-crystallization of \( \text{RePhaC1}_{\text{CD}} \) with substrate or soaking of 3-HB-CoA into the \( \text{RePhaC1} \) crystal was successful, again proving the notoriously difficult nature of PhaC crystallization. Thus, as an
alternative, molecular docking simulations of 3-HB-CoA and 3-HB trimer (3-HB)_3 were performed. To validate the molecular docking simulation, far-UV circular dichroism and fluorescence emission spectroscopic analyses of RePhaC1 were performed. Since no noticeable structural changes were observed by the addition of 3-HB-CoA, the molecular docking simulation results represent natural binding mode of the 3-HB-CoA substrate (Figs. 3C and 3D). Both substrates fit into the substrate-binding tunnel (Figs. 3E and 3F). These observations indicate that the two substrates, 3-HB-CoA and the growing PHB polymer, share the same substrate-binding tunnel. Thus, we propose that the PHB polymerization reaction of RePhaC1 is non-processive and follows a ping-pong mechanism. The first substrate 3-HB-CoA binds to the substrate-binding tunnel, and free CoA leaves the tunnel. Then, the second substrate, the PHB\_n polymer, enters the substrate-binding tunnel, and the PHB\_n+1 polymer leaves the tunnel. Molecular docking simulations also allowed us to identify the residues involved in stabilizing the substrates. The β-mercaptoethylamine/pantothenate (β-MP) moiety of CoA is stabilized in the substrate-binding tunnel through interactions with residues including I247, L397, R398, V360, T393, F396, Y440, Y445 and I482 (Figs. 3E, 3F and 3G). The nucleotide moiety is positioned at the shallow pocket that is formed on the outside of the tunnel and stabilized by residues such as D359, D354, H481, W485, and T486 (Figs. 3E, 3F and 3G). The (3-HB)_3 is also positioned in the substrate-binding tunnel in a manner similar to that of the β-MP moiety of CoA and stabilized by residues of Ile247, Val360, Thr393, Phe396, L397, Arg398, Tyr445, His481, and Ile482 (Figs. 3E, 3F and 3H). Some of these residues involved in substrate stabilization were confirmed by site directed mutagenesis experiments (Fig. 3I). Interestingly, the chemical property of the (3-HB)_3 is similar to that of the β-MP moiety of CoA (Fig. 3E). In addition, the lengths of these two chemicals are almost identical; 12.42 Å for (3-HB)_3 and 12.37 Å for the β-MP moiety (Fig. 3E). Also, this length is quite similar to that (13 Å) of the substrate-binding tunnel (Fig. 3E). These observations explain how RePhaC1 utilizes a single substrate-binding tunnel for both 3HB-CoA and the growing PHB polymer. Several residues, including I247, V360, T393, F396, R398, Y445, H481, and I482 are involved in the stabilization of both substrates (Figs. 3G and 3H).
As mentioned above, RePhaC1 exists as a mixture of monomeric and dimeric forms (Fig. 2D), and the dimeric form is induced by addition of substrates [9]. In the dimeric form of RePhaC1CD, the tunnel entrances of the two subunits are located at the dimerization interface, which consequently makes 3-HB-CoA be stabilized by the neighboring subunit when it binds to one subunit (Fig. 3K). More specifically, the pyrophosphate moiety of CoA intensively contacts with G356, D359, V360, and R398 provided by the neighboring subunit; main-chain of G356 and side-chain of D359 form hydrogen bonds with the pyrophosphate moiety (Fig. 3K). These observations suggest that 3-HB-CoA plays a role as a mediator of dimerization by enabling the enzyme to form a dimer more tightly than its apo-form. The lag phase in the initial PHA polymerization has been suggested to be caused by the low affinity of the enzyme for the nucleotide moiety. Also, the lag phase could be decreased when the synthesized (3-HB)₃ was added to the reaction mixture [24, 25]. These previous observations can be explained by our findings above showing that the growing polymer can also bind to the same substrate binding pocket, which can consequently induce dimerization of enzyme (Figs. 3E and 3F).

3.5 Catalytic mechanisms of RePhaC1

At the active site of the RePhaC1CD monomer, the conserved Cys319, Asp480, and His508 residues form a Cys-His-Asp catalytic triad and function as a covalent nucleophile, a general base, and an electron donor, respectively (Fig. 3B). Mutation of these residues to alanine resulted in a complete loss of enzyme activity both in in vitro assay and in vivo granule formation experiments (Fig. 3J). Based on the structure determined, the following catalytic mechanism of PHB polymerization by PHA synthase can be proposed. The polymerization reaction can be divided into two steps: covalent bonding of 3-HB to the active site and PHB chain elongation (Fig. 4A). In the first step, the 3-HB-CoA substrate enters the active site and the first charge relay occurs in the Cys-His-Asp catalytic triad. The deprotonated thiol-group of Cys319 serves as the first nucleophile and attacks the carbonyl carbon of 3-HB-CoA substrate. The carbonyl oxygen is then forced to accept an electron, and the first tetrahedral intermediate is formed. Collapse of this intermediate leads to covalent bonding of the 3-HB moiety to Cys319, transfer of an electron back to Asp480, and release of the protonated CoA molecule. In the second step, the growing PHBₙ polymer substrate enters the active site, and the second charge relay
occurs between Asp480, His508, and the terminal hydroxyl-group of the growing PHB$_n$ polymer. The deprotonated hydroxyl group serves as a second nucleophile and attacks the carbonyl carbon of the 3-HB moiety that is covalently bonded to Cys319 after the first step. The carbonyl oxygen is then forced to accept an electron, and the second tetrahedral intermediate is formed. Collapse of this intermediate results in the formation of the PHB$_{n+1}$ polymer, and the catalytic residues are restored to their original electro-states. PHB$_{n+1}$ polymer leaves the tunnel making the tunnel available for 3-HB-CoA to enter again (Fig. 4).

The first substrate 3-HB-CoA coming into the empty tunnel can always form a covalent bond with Cys319 during the entire polymerization process as described above. However, the binding affinity of the second substrate (PHB$_n$) coming into the tunnel, which is already occupied by covalently bound 3-HB, might be weaker than that of the first substrate during the elongation step of the initial phase of polymerization (see below for the later phase of polymerization). This is because the terminal hydroxyl group, instead of the carbonyl group, of the second substrate is positioned in the vicinity of the catalytic site to function as a second nucleophile (Fig. 4A). Positioning of the terminal hydroxyl group might push the second substrate out of the tunnel, which in turn prevents proper binding of the nucleotide moiety of CoA to its binding pocket and reduces the binding affinity of the second substrates, 3-HB-CoA, (3-HB)$_2$-CoA, and so on, to their binding sites. This phenomenon might become more serious as the polymerization reaction proceeds, because the nucleotide moiety of CoA would completely leave its binding pocket (Fig. 4B). It was previously reported that RePhaC1 shows a lag phase at the start of polymerization [26, 27]. This lag phase can be explained by the low affinity for the nucleotide moiety of the second substrate, 3-HB-CoA or (3-HB)$_r$-CoA. Additional binding forces between the enzyme and the second substrate seem to be required to reduce this lag phase.

### 3.6 Substrate specificities of PHA synthases

RePhaC1 is a class I PHA synthase, and class I, III, and IV PHA synthases show preference towards the formation of short-chain-length PHAs [28]. In contrast, class II PHA synthases, such as PhaC1 and PhaC2 from *Pseudomonas aeruginosa*, utilize medium/long-chain-length 3-hydroxyacyl-CoAs as their main substrates[29, 30]. The RePhaC1$_{CD}$ structure determined here provides insights
into why different classes of PHA synthases exhibit different substrate specificities. Considering the first step of the polymerization reaction, in which the 3-HB moiety forms a covalent bond with Cys319 (Fig. 4A), a pocket that stabilizes the 3-HB moiety is required at the internal edge of the substrate-binding tunnel. Indeed, a small hydrophobic pocket was found in the vicinity of the catalytic triad of RePhaC1 that seems to serve as a 3-HB binding pocket (Fig. 5A). This pocket comprises hydrophobic residues, such as Pro245, Ile252, Leu253, Phe318, Thr393, and Trp425 (Fig. 5A). Interestingly, these residues are highly conserved among class I, III, and IV PHA synthases (Fig. 5C), suggesting that other PHA synthases from these classes form a 3-HB binding pocket similar to that of RePhaC1 and consequently have similar substrate specificities. In contrast, different residues are found at these positions in class II PHA synthases, and these residues are highly conserved only within this class (Fig. 5C). Among these residues, one striking difference was found at position 318 of RePhaC1; predominantly phenylalanine or tyrosine in class I, III, and IV PHA synthases, but alanine, relatively small amino acid in class II PHA synthases. These observations indicate that class II PHA synthases might have a much larger pocket than those of other classes to accommodate a large 3-hydroxyacyl moiety (Fig. 5D). The substrate-binding tunnel through which substrates pass during catalysis also exhibits different conformation depending on the class. For RePhaC1, the residues such as Thr393, Phe396, Leu397, Tyr445, Ile482, and Val483 constitute the substrate-binding tunnel (Fig. 5B), and these residues are highly conserved in PHA synthases of classes I, III, and IV (Fig. 5C). On the other hand, Val, Trp, Met, Phe, Ile, and Thr are positioned in the substrate-binding tunnel of class II PHA synthases, and these residues are highly conserved only within this class (Fig. 5C). Based on these findings, it is concluded that the conformation of the active site tunnels of class I, III, and IV PHA synthases are quite similar and suitable for short-chain-length 3-hydroxyacyl-CoAs as substrates, while the unique conformation of class II PHA synthases is suitable for longer-chain-length substrates; 3-hydroxyhexanoyl-CoA is showcased as an example (Fig. 5D).

To further validate the substrate specificity and also propose enzyme engineering strategies based on the determined structure, we performed rational protein engineering on RePhaC1. Those residues involved in the formation of 3-HB binding pocket and substrate binding tunnel were selected and targeted for mutations (Figs. 3G, 3H, 5A, 5B, and 5C); the mutations introduced in the engineered
RePhaC1 include V320L, F396M, F396L, V483A, P245A, F318Y, and T393S. Then, the engineered PHA synthases were expressed *E. coli* and purified for measuring activities in comparison with the wild-type enzyme. All of these engineered enzymes were expressed well in soluble form, but showed varying relative activities (Fig. 5E). Most of them showed decreased activities compared with that of the wild-type enzyme, but one engineered RePhaC1 having P245A mutation showed 25 % higher activity. This residue seems to affect hydrophobic property near the terminal methyl group of the substrate. Also, this mutation made the 3-HB binding pocket become further extended (Fig. 5F). It is interesting to note that most class I PHA synthase has alanine at the position of P245 in RePhaC1 (Fig. 5C). These results showcase that rational engineering of PHA synthase based on the determined structure can enhance the enzyme activity.

4 Concluding remarks

In this study, we report the crystal structure of the C-terminal domain of PHA synthase from *R. eutropha* (*RePhaC1CD*) at 1.8 Å resolution and describe the molecular mechanisms of PHA biosynthesis. We also report structure-based reasons for the differences in substrate specificities among the PHA synthases in different classes. The results reported here will be useful for further studies on understanding the mechanisms of PHA synthases belonging to other classes. Due to the climate change and other environmental problems, we need to reduce our dependence on fossil oil and move towards bio-based production of polymers from renewable non-food biomass. This study revealing the first structure of PHA synthase will facilitate development of strategies for rational engineering of PHA synthases to produce tailor-made PHAs having various monomers and diverse material properties, and also to enhance PHA production. Also, as can be seen from our accompanying paper [31], we propose the structure and function of the whole PHA synthase based on our studies on the N-terminal domain of PHA synthase and associated proteins such as PhaM.
Acknowledgement

This work was supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries from the Ministry of Science, ICT and Future Planning (MSIP) through the National Research Foundation (NRF) of Korea (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557). The atomic coordinates and structure factors of RePhaCD crystal structure have been deposited in the Protein Data Bank with accession number of 5HZ2 on February 2, 2016.

Conflict of interest

The authors declare no financial or commercial conflict of interest.
5 References


[26] Zhang, S., Yasuo, T., Lenz, R.W., Goodwin, S. Kinetic and mechanistic characterization of the polyhydroxybutyrate synthase from *Ralstonia eutropha*. *Biomacromolecules*, 2000, **1**, 244-251.


Table 1. Data collection and structure refinement of RePhaC1<sub>CD</sub>

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Derivative (HgCl&lt;sub&gt;2&lt;/sub&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Space group</strong></td>
<td>I222</td>
<td>I222</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>73.3, 87.9, 136.5</td>
<td>73.0, 87.6, 136.4</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.0, 90.0, 90.0</td>
<td>90.0, 90.0, 90.0</td>
</tr>
<tr>
<td><strong>Wavelength (Å)</strong></td>
<td>0.9798</td>
<td>1.0072</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>50.00-1.80 (1.83-1.80)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>50.00-2.01 (2.04-2.01)</td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;sym&lt;/sub&gt; or R&lt;sub&gt;merge&lt;/sub&gt;</strong></td>
<td>5.4 (26.1)</td>
<td>7.1 (17.1)</td>
</tr>
<tr>
<td><strong>I/σ</strong></td>
<td>52.2 (3.2)</td>
<td>76.1 (18.7)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>98.1 (93.5)</td>
<td>99.4 (92.0)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>9.8 (5.1)</td>
<td>14.4 (11.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Native</th>
<th>Derivative (HgCl&lt;sub&gt;2&lt;/sub&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>50.00-1.80</td>
<td></td>
</tr>
<tr>
<td><strong>No. reflections</strong></td>
<td>38300</td>
<td></td>
</tr>
<tr>
<td><strong>R&lt;sub&gt;work&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt;</strong></td>
<td>15.1 / 19.0</td>
<td></td>
</tr>
<tr>
<td><strong>No. atoms</strong></td>
<td>3343</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>3014</td>
<td></td>
</tr>
<tr>
<td>Ligand(GOL)/ion(SO4)</td>
<td>12 / 28</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>289</td>
<td></td>
</tr>
<tr>
<td><strong>B-factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>29.4</td>
<td></td>
</tr>
<tr>
<td>Ligand(GOL)/ion(SO4)</td>
<td>32.9 / 61.6</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>37.9</td>
<td></td>
</tr>
<tr>
<td><strong>R.m.s deviations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.952</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Values in parentheses are for highest-resolution shell.
Figure 1. Monomeric structure of RePhaC1

(A) Amino acid sequence alignment of representative PHA synthases from Classes I, II, III, and IV.

Secondary structure elements from the RePhaC1\text{CD} crystal structure are labeled. The N- and C-domains, D-loop, HTH motif, EC-region, and PS-region are shown. Residues involved in enzyme catalysis and CoA binding as well as the 3-HB binding pocket and substrate-binding tunnel are shown as red, cyan,
blue, and violet triangles, respectively. Re, Pa, Bc, and Bm are PHA synthases from *Ralstonia eutropha*, *Pseudomonas aeruginosa*, *Bacillus cereus*, and *Bacillus megaterium*, respectively.

(B) Schematic diagram of the domains of *Re*PhaC1\(_F\).

(C) SDS-PAGE of the crystallized *Re*PhaC1\(_{CD}\) protein. Lanes 1 and 2 are the size marker and crystallized *Re*PhaC1\(_{CD}\) protein, respectively.

(D) *In vitro* proteolysis of *Re*PhaC1\(_F\). Lanes 1–3 are *Re*PhaC1\(_F\) before proteolysis, the size marker, and *Re*PhaC1\(_F\) after proteolysis, respectively.

(E) Monomeric structure of *Re*PhaC1\(_{CD}\). The structure of *Re*PhaC1\(_{CD}\) is shown as a cartoon diagram. The \(\alpha/\beta\) hydrolase fold, dimerization subdomain, and EC-region are shown in cyan, orange, and magenta, respectively. The protruding structure region (PS-region) is distinguished with green-colored dotted-box. The figure on the right is 90 degrees rotated compared to the figure on the left in a horizontal direction.
Figure 2. Dimeric structure of RePhaC1

(A) Dimeric structure of RePhaC1\textsubscript{CD}. A dimer of RePhaC1\textsubscript{CD} is shown as a cartoon diagram. Core subdomains and EC-regions of two monomers are shown with gray and magenta, respectively. The dimerization domains of two monomers are distinguished with cyan and orange, respectively. The right side figure is 90 degree rotation horizontally from the left side figure.

(B and C) Dimerization interface of RePhaC1\textsubscript{CD}. Interaction modes of the HTH motif (B) and the D-loop (C) of one monomer to the other monomer. One monomer is presented with a cartoon diagram and the other is with an electrostatic surface model. The HTH motif and the D-loop of one monomer are distinguished with colors of an orange and a cyan, respectively. Residues on the HTH-motif and the D-loop are shown as a line model and those applied to site-directed mutagenesis are shown as a stick model.
(D and E) SEC analysis (D) and PHA synthase activity assay (E) of RePhaC1 mutants at the dimerization interface. Standard samples of 1, 2, and 3 represent Aldolase (158 kDa), Ovalbumin (44 kDa) and Ribonuclease A (13.7 kDa), respectively. The ratios of dimer/monomer eluted from the SEC are calculated based on the areas of the eluted peaks.
Figure 3. Active site of RePhaC1<sub>CD</sub>

(A) A surface cross-section view of the RePhaC1<sub>CD</sub> dimer is presented to show the substrate-binding tunnels. The sulfur atoms of the two catalytic cysteine residues are shown as sphere models. The substrate-binding tunnel is labeled. The distance between two catalytic sites and the length of the substrate binding tunnel were labeled.
(B) Catalytic site of \( \text{RePhaC1}_{\text{CD}} \). The catalytic triad is shown as a stick model. The bound 3-HB-CoA molecule is also shown as a stick model with magenta color.

(C and D) Far-UV circular dichroism (C) and fluorescence emission (D) spectroscopic analysis of \( \text{RePhaC1}_F \) with and without acetyl-CoA.

(E and F) Substrate-binding tunnel of \( \text{RePhaC1}_{\text{CD}} \). The side view (E) and the top view (F) of the substrate binding tunnel are shown as a surface model. The tunnel length is indicated with an arrow and labeled. The bound 3-HB-CoA (magenta) and (3-HB)_3 trimer (green) are shown as stick models. The catalytic triad is shown as stick models and labeled.

(G and H) Binding mode of 3-HB-CoA (G) and the (3-HB)_3 trimer (H). Hydrogen bonds are shown as green-colored dotted-lines and the distances were labeled. Residues involved in the stabilization of both 3-HB-CoA and (3-HB)_3 are distinguished with a star mark.

(I and J) \textit{In vitro} and \textit{in vivo} PHB polymerization activities of the \( \text{RePhaC1}_F \) mutants. The relative activities of the mutants with changes in the residues involved in CoA and (3-HB)_3 trimer binding (I) and enzyme catalysis (J) are shown relative to that of the wild-type enzyme.

(K) Stabilization of pyrophosphate moiety of 3-HB-CoA. The \( \text{RePhaC1}_{\text{CD}} \) structure is shown as a surface and cartoon models. Two monomers are distinguished with colors of gray and red, respectively. The hydrogen bonds formed between the pyrophosphate moiety of 3-HB-CoA in one monomer and the neighboring monomer are shown as yellow-colored dotted lines.
Figure 4. Catalytic mechanism of RePhaC1

(A) Catalytic mechanism of RePhaC1. The polymerization reaction can be divided into two steps, covalent bonding of 3-HB and PHB chain elongation.

(B) Initial PHB polymerization. The first and second substrates and production during the initial PHB polymerization phase are shown.
Figure 5. Substrate specificities of PHA synthases

(A) 3-HB binding pocket. The binding pocket for the 3-HB moiety covalently bonded to C319 after the first enzyme reaction is shown as a surface cross-section model. The catalytic residues are shown as cyan-colored sticks, and the residues constituting the pocket are shown as orange-colored sticks. The bound 3-HB-CoA is shown as a stick model (magenta).

(B) Substrate-binding tunnel. The substrate-binding tunnel is shown as in (A). The bound (3-HB)$_3$ trimer is shown as a stick model (green).

(C) Residues constituting the 3-HB binding pocket and the substrate-binding tunnel in class I, II, III and IV PHA synthases are compared.

(D) Schematic diagram comparing the 3-hydroxyacyl binding pockets. 3HV, PHBHV, 3HH, PHBHH are abbreviations for 3-hydroxyvaleryl-CoA, poly(3-hydroxybutyrate-co-3-hydroxyvalerate), 3-hydroxyhexanoyl-CoA, and poly(3-hydroxybutyrate-co-3-hydroxyhexanoate), respectively.

(E) Enhanced PHA synthase activity. Top figure shows SDS-PAGE of the purified $Re$PhaC$_1_F$ mutants, and bottom figure shows relative activity of the mutants.

(F) Enlarged 3-HB-CoA binding pocket by mutation of P245 (left) to alanine (right). The 3-HB-CoA binding pocket is shown with cross-section of the surface model of $Re$PhaC$_1_{CD}$ structure.