STRUCTURAL AND MECHANISTIC CHARACTERIZATION OF CARBON-PHOSPHORUS LYASE MULTI-PROTEIN COMPLEX AND THE PHOSPHOTRIESTERASE FROM *SPHINGOBIUM SP*. STRAIN TCM1

A Dissertation

by

ZHONGJIE REN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Frank M. Raushel
Committee Members,	David P. Barondeau
	Gregory Reinhart
	Paul Straight
Head of Department,	Gregory Reinhart

December 2016

Major Subject: Biochemistry

ABSTRACT

Phosphonates are a class of organophosphorus compounds with a characteristic carbon–phosphorus bond. The carbon–phosphorus lyase complex, encoded by the genes *phnGHIJKLM*, catalyzes the cleavage of the stable carbon–phosphorus bond of organophosphonates to the corresponding hydrocarbon and inorganic phosphate. Recently, complexes of this enzyme containing five subunits (PhnG-H-I-J-K), four subunits (PhnG-H-I-J), and two subunits (PhnG-I) were purified after expression in *Escherichia coli*.

Here we used mass spectrometry, ultracentrifugation, chemical cross-linking, N-terminal protein sequencing and single-particle cryo-electron microscopy experiments, revealing subunits interactions and construct structural models for the $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$ C–P lyase complexes from *E. coli*. The proposed stoichiometry, subunit interaction and structural model of $PhnG_2H_2I_2J_2$ complex agreed quite well with the crystal structure that was published later. $PhnG_2H_2I_2J_2$ is a two-fold symmetric hetero-octamer. Its two PhnJ subunits provide two identical binding sites for PhnK. Only one PhnK binds to $PhnG_2H_2I_2J_2$ due to steric hindrance. Moreover, binding of PhnK exposes the active site residue, Gly32 of PhnJ, located near the interface between PhnJ and PhnH. This structural information provides a basis for further deciphering of the reaction mechanism of the C-P lyase**.**

A group of organophosphorus compounds, the phosphotriester plasticizers and flame retardants, has recently become widely used. This group of compounds lacks an easily hydrolyzable bond, rendering them inert to typical phosphotriesterases. A phosphotriesterase from *Sphingobium sp.* strain TCM1 (*Sb*-PTE) has recently been reported to catalyze the hydrolysis of organophosphorus flame retardants. This enzyme was expressed in *Escherichia coli*, and the activity with a wide variety of organophosphorus substrates has been characterized. *Sb*-PTE exhibits catalytic activity against flame retardants, plasticizers, and industrial solvents.

In this study, we use mass spectrometry and peptide mapping experiments, revealing the active site residues and reaction mechanism for *Sb*-PTE. The proposed active site residues of *Sb*-PTE complex agreed quite well with the crystal structure that was published later. The enzyme catalyzes the hydrolysis of substrates by activation of a nucleophilic water molecule for direct attack at the phosphorus center.

ACKNOWLEDGEMENTS

I am greatly grateful to my principal investigator, Dr. Frank M. Raushel, for his guidance, support and constant encouragement over the last 5 years at Texas A&M University. I am thankful to Dr. David Russell, Dr. William Russell and Dr. Larry Dangott for all their efforts and help with mass spectrometry, Dr. Junjie Zhang and Kailu Yang for their excellent work and collaboration on single particle cyro-electron microscopy, Dr. David P. Barondeau and his lab-members for the use of the glovebox in his laboratory, Dr. Gregory Reinhart and Dr. Paul Straight for serving as a member on my thesis committee. I also thank Dr. Siddhesh Kamat for serving as my mentor in the Raushel lab, and, Dr. Daofeng Xiang, Dr. Andrew Bigley and Dr. Mark Mabanglo for great collaboration and for valuable advice throughout my graduate studies.

I am grateful to my family and friends for showing confidence and faith in me. I also express my deepest gratitude to all members of the Raushel lab, past and present, for all the help and friendship. Finally, thanks to my mother and father for their encouragement and to my wife for her patience and love.

CONTRIBUTORS AND FUNDING SOURCES

CONTRIBUTORS

This work was supervised by a dissertation committee consisting of Professors Frank M. Raushel, Paul Straight and Gregory Reinhart of the Department of Biochemistry and Biophysics and Professor David P. Barondeau of the Department of Chemistry.

All data from cryo-electron microscopy experiments were collected, processed and interpreted by Kailu Yang in Dr. Junjie Zhang'lab. And all data from H/D mass spectrometry experiments were collected, processed and interpreted by Soumya Ranganathan. The data of native mass spectrometry was collected by Dr. William Russell in Dr. David Russell's lab.

All other work conducted for the dissertation was completed by the student independently.

FUNDING SOURCES

This work was made possible in part by NIH under Grant Number GM103917.

TABLE OF CONTENTS

Page

CHAPTERS

LIST OF FIGURES

Page

LIST OF TABLES

LIST OF SCHEMES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW*

1.1 Bacterial Phosphonate Metabolism

1.1.1 Biological importance and industrial applications of phosphonates

Phosphorus, an essential element for life, is a limiting nutrient in various ecosystems. The most abundant phosphorus sources are orthophosphate and phosphate esters, which contain a C-O-P bond. The C-O-P bond is part of DNA and RNA backbones, which are essential for the storage, transcription and expression of genetic information. C-O-P bond is also key component of ATP which is most common cellular fuel. However, organophosphonates may have been the predominant phosphorus source on early earth when oxygen was not a major component of air (1). In contrast to the

---

* Part of this chapter is reprinted with permission from "Subunit Interactions within the Carbon–Phosphorus Lyase Complex from Escherichia coli" by Zhongjie Ren, Soumya Ranganathan, Nathanael F. Zinnel, William K. Russell, David H. Russell, and Frank M. Raushel, *Biochemistry* **2015** *54* (21), 3400-3411, Copyright 2016 by American Chemical Society.

Part of this chapter is reprinted with permission from "Interrogation of the Substrate Profile and Catalytic Properties of the Phosphotriesterase from *Sphingobium sp*. Strain TCM1: An Enzyme Capable of Hydrolyzing Organophosphate Flame Retardants and Plasticizers" by Dao Feng Xiang, Andrew N. Bigley, Zhongjie Ren, Haoran Xue, Kenneth G. Hull, Daniel Romo, and Frank M. Raushel, *Biochemistry* **2015** *54* (51), 7539-7549, Copyright 2016 by American Chemical Society.

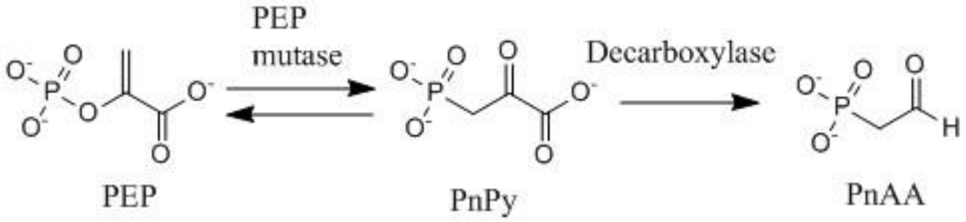
orthophosphate and phosphate esters which contain a C-O-P bond, the C-P bond in organophosphonates is highly resistant to chemical hydrolysis, thermal decomposition and photolysis (2).

Compounds containing a C-P bond constitute as much as 20% to 30% of the available phosphorus in nature (3). The highly stable C-P bond in organophosphonates, such as lipids, exopolysaccharides and glycoproteins, stabilize the cell membrane structure and resist phosphatase hydrolysis. Phosphonates are the main cellular phosphorus-containing compounds in several organisms. For example, 50% of the phosphorus in the sea anemone Tealia could be found in a variety of phosphonates such as phosphonoglycoproteins, phosphonoglycans, and phosphonolipids (4), whereas 95% of phosphorus in egg cells of the freshwater snail Helisoma exists in the form of 2-aminoethylphosphonate modified phosphonoglycans (5). In addition, the organophosphonates are also widely used in the pharmaceutical and agricultural industry. Fosfomycin, dehydrophos, and plumbemycin are broad spectrum antibiotics. Glyphosate, phosphinothricin tripeptide (bialaphos) and phosphonothrixin are common herbicides. Alendronate is an effective drug for the treatment of osteoporosis (3). The structural similarity of bioactive organophosphonates to phosphate esters and carboxylic makes them analogous inhibitors with high affinity which can compete with phosphate esters and carboxylic for binding to the active sites of corresponding enzymes (6). In addition, several phosphonates antibiotics such as fosfomycin can cause irreversible inhibition through covalent modification of the active site of enzymes (7). Phosphate esters and carboxylic group are involved in most of the biological process in cells, so potential enzyme targets for phosphonate inhibitors are widespread in metabolic and regulatory pathways such as glycolysis, protein phosphorylation and carboxylation (8, 9).

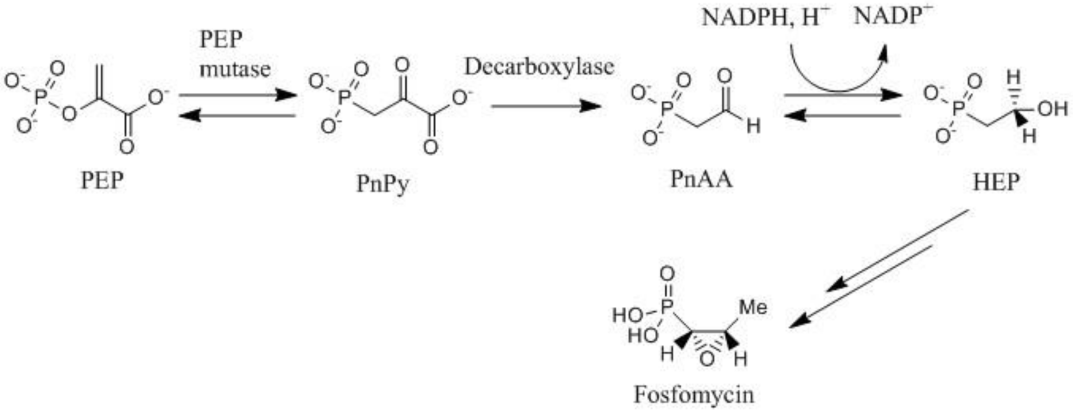
Organophosphonates were initially studied in 1944 by Pikl who synthesized aminomethylphosphonic acid, and then Kosolapoff and Chavane successfully synthesized a series of amino and amino-substituted phosphonates (10). Chavane predicted the natural existence of phophonates based on the inherent stability of the carbon phosphorus bond (11). The occurrence of organophosphonates in biological systems was not discovered until 1959. 2-aminoethylphosphonic acid (2-AEP) was the first natural organophosphonate that was identified, which was isolated from an ether–ethanol-extracted hydrolysate of rumen protozoa (12). Today, natural and man-made phosphonates are found to be widespread in the environment, however, phosphonates were considered of only limited significance for a very long time.

1.1.2 Bacterial biosynthetic pathways of phosphonates

Recent studies suggest that organophosphonates play a major role in the marine phosphorus cycle (13). In the total oceanic dissolved organic phosphorus, the share of organophosphonates can be up to 30 percent. The biosynthetic pathways for organophosphonates are mostly conserved in their initial steps, considering the great diversity of natural phosphonates. In the organophosphonate biosynthesis pathway, the initial carbon phosphorus bond formation occurs through the intramolecular rearrangement of phosphoenolpyruvate (PEP) to phosphonopyruvate (PnPy). This reaction is catalyzed by the enzyme phosphoenolpyruvate phosphomutase (PEP mutase) (Scheme 1) (14). The thermodynamics of the reaction indicates that the product, phosphonopyruvate, is 500 fold less favorable than the substrate, phosphoenolpyruvate. To provide the thermodynamics driving force to push the reaction towards phosphonopyruvate, the initial reaction is usually followed by a coupled enzymatic reaction, catalyzed by phosphonopyruvate decarboxylase (Ppd), which converts phosphonopyruvate to phosphonoacetaldehyde (PnAA) (15). In other cases, the condensation of phosphonopyruvate and acetyl-CoA serves as the coupled enzymatic reaction, and results in the formation of 2-phosphonomethylmalate. This exergonic reaction is catalyzed by a homolog of homocitrate synthase, FrbC (16).



Scheme 1. Biosynthetic pathway for phosphonoacetaldehyde (PnAA) (3)



Scheme 2. Proposed biosynthetic pathway for fosfomycin (3)

Phosphonoacetaldehyde and 2-phosphonomethylmalate then go through a variety of further biochemical reactions; for example, in the biosynthesis of fosfomycin, phosphonoacetaldehyde (PnAA) undergoes further methylation to form 2-hydroxypropyl phosphonic (HEP) by the enzyme phosphonoacetaldehyde reductase (Scheme 2) (16).

1.1.3 Regulation of bacterial phosphonate uptake and metabolism

The uptake and catabolism of phosphonates is subject to the Pho regulon, which links the environmental phosphate sensor to gene expression, providing a sophisticated and adaptive response to phosphate availability in the environment. The Pho regulon, negatively controlled by the phosphate level, includes eight different transcriptional units (17). In the Pho regulon, the PstSCAB protein complex, an ATP-binding cassette (ABC) transporter, acts as the detector of environmental phosphate (18). The organization of the PstSCAB protein complex is illustrated in Figure 1. The PstSCAB protein complex is formed by four different subunits including two transmembrane proteins (PstC and PstA), a cytoplasmic ATP binding protein (PstB) and a periplasmic phosphate binding protein (PstS). If the environmental phosphate is insufficient, then PstS decouples from PstA and PstC, leading to a conformational change in the PstSCAB complex. The conformational change is transmitted through the subunit PhoU to the membrane associated histidine kinase PhoR. The detailed mechanism of transition is unknown. PhoR can be switched into two forms, repressor and activator, which act as phosphatase and kinase of PhoB, respectively (18).

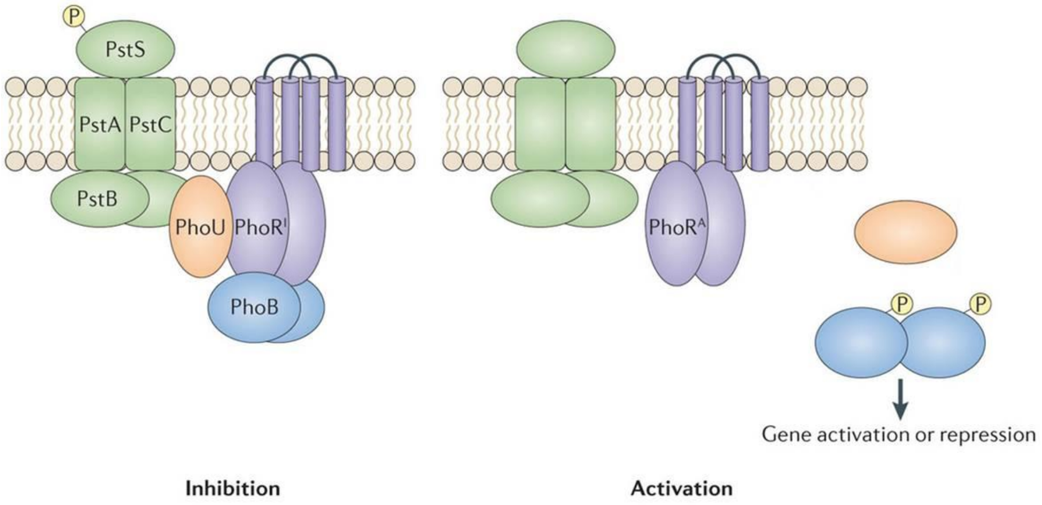


Figure 1: Model of phosphate regulon signal transduction (19)

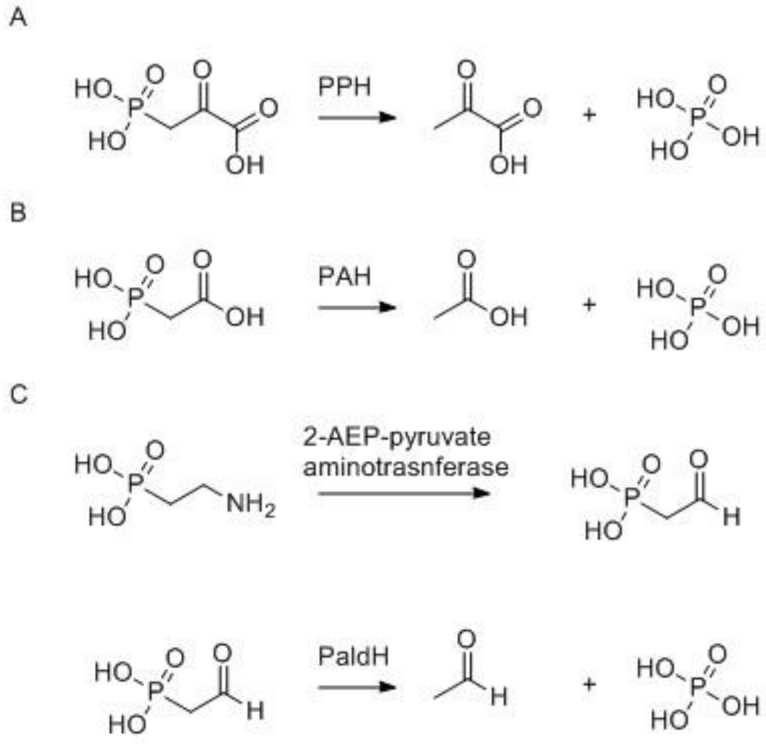
Under phosphate deficiency, PhoR serves as an activator by phosphorylating PhoB. Upon phosphorylation, PhoB binds to the Pho box, which is a conserved DNA sequence, located 10 bp upstream of the Pho regulon. The feature of the Pho box sequence is tandem repeats of a conserved 6-7 base pair motif. The motif is separated by short a sequence which varies in different species. Upon binding between PhoB and the Pho box DNA sequence, PhoB facilitates RNA polymerase binding onto the promoters in the Pho regulon to activate transcription of a series of genes that promote the uptake and catabolism of both varied phosphorus source under conditions of phosphate limitation (18).

1.1.4 Bacterial degradation pathways of phosphonates

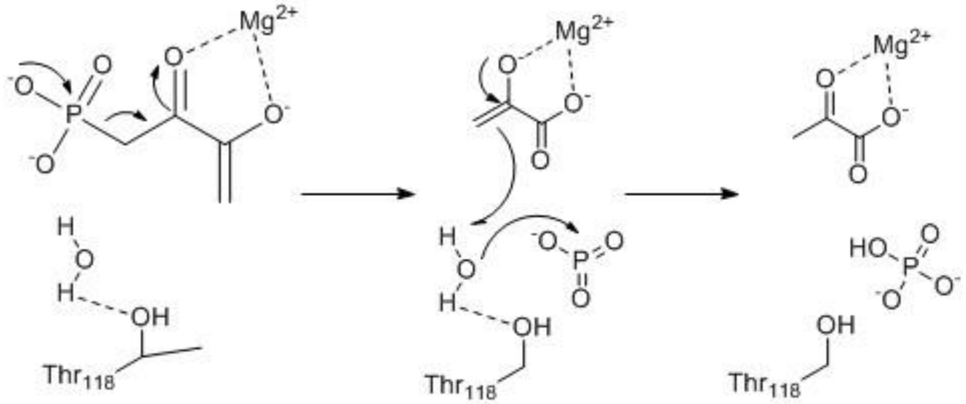
The bacterial catabolism of phosphonates as an alternative phosphorus source has been widely studied. Currently, three groups of distinct organophosphonate degradation enzymes or enzymatic pathways have been characterized, including the C-P lyase, phosphonate hydrolases, and an oxidative pathway (20).

The first group of organophosphonate degradation enzymes is phosphonate hydrolase, including phosphonoacetaldehyde hydrolase (Pald), phosphonoacetate hydrolase (PAA) and phosphonopyruvate hydrolase (PPH). Refereed as phosphonatase, This group of enzymes was firstly studied and has been referred as phosphonatase. Phosphonate hydrolases can specifically cleave the C–P bond of substrates containing an electron withdrawing β-carbonyl group that facilitates bond delocalization. (Scheme 3) (21).

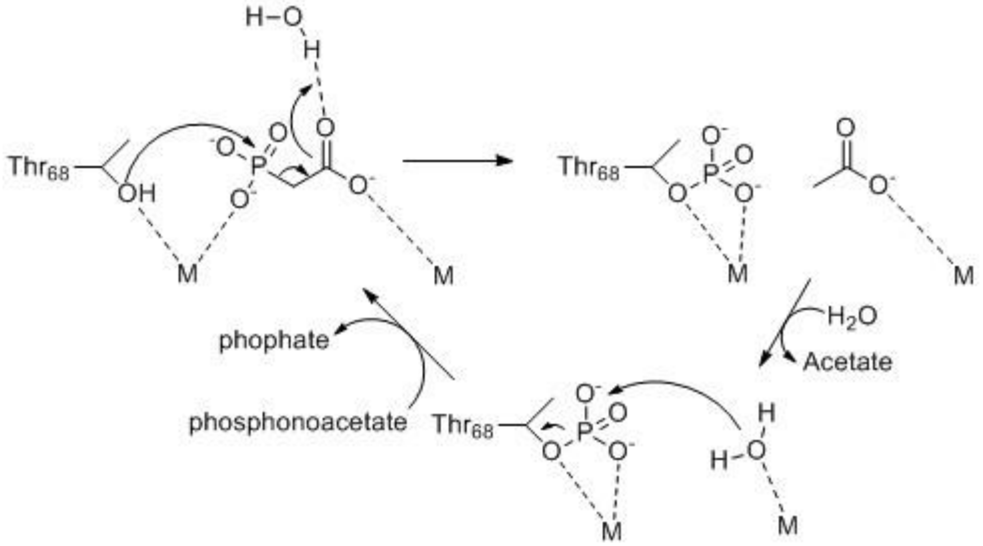
Enzymes in the phosphonate hydrolases group have evolved from varied protein superfamilies. The phosphonate hydrolysis activity was firstly reported by Ternan NG and coworkers. They identified phosphonopyruvate hydrolase (PPH) from cell free extracts of *Burkholderia cepacia*, which is capable of utilizing phosphonoalanine as a phosphorus source (22). In the proposed reaction mechanism catalyzed by PPH, phosphonopyruvate is hydrolyzed into phosphate and pyruvate (Scheme 3a). As a member of the PEP mutase/isocitrate lyase superfamily, PPH possesses a (β/α) 8-barrel structural fold. Monomers of PPH can form homo-tetramers with the eighth helix inserted between two adjacent dimers. The crystal structure of PPH has been solved with magnesium ion and oxalate binding within its active site (PDB id: 2DUA). The Mg^{2+} in the active site attracts the phosphonopyruvate substrate to facilitate further hydrolysis. The details of the proposed reaction mechanism are illustrated in Scheme 4. (23)



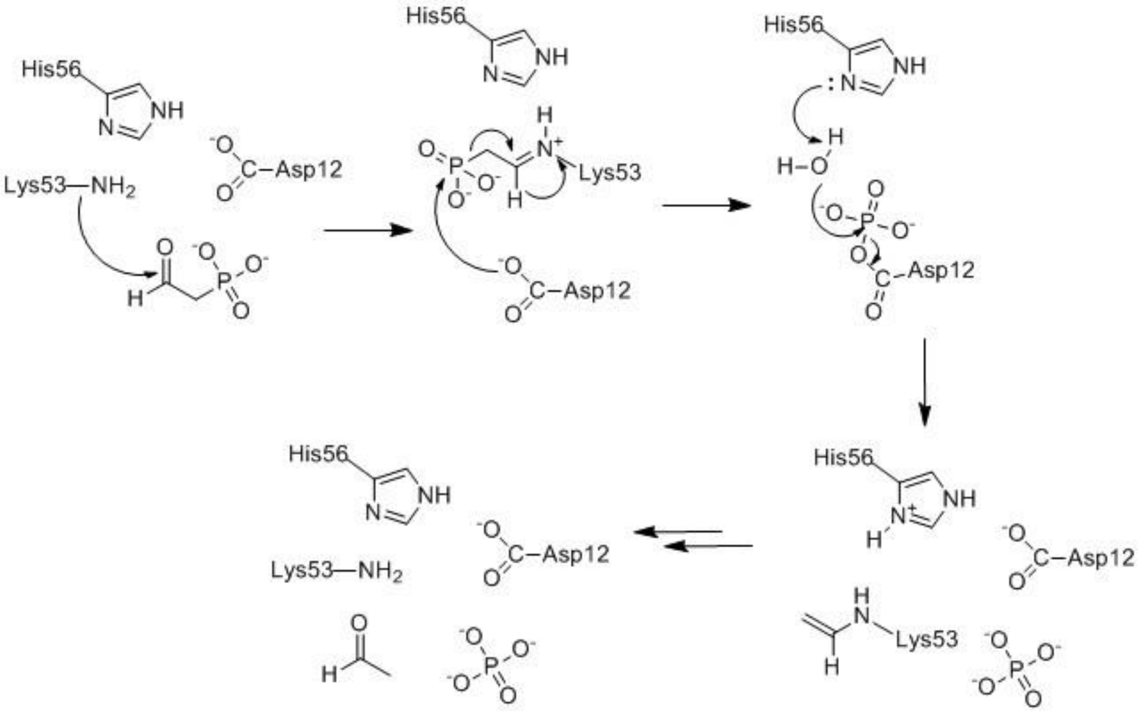
Scheme 3. Typical reactions catalyzed by enzymes of phosphonate hydrolases group (24)



Scheme 4. Mechanism of the reaction catalyzed by phosphonopyruvate hydrolase (PPH) from *Variovorax sp. Pal2*. (25)



Scheme 5. Mechanism of the reaction catalyzed by phosphonoacetate hydrolase (PAH) (26)



Scheme 6. Mechanism of the reaction catalyzed by phosphonoacetaldehyde hydrolase (PaldH) (27)

Another member of the phosphonate hydrolase group, phosphonoacetate hydrolase (PAH) was firstly reported in *Pseudomonas fluorescens*, from bacteria isolated in the sludge of a laundry waste treatment plant. The substrate, phosphonoacetate, is converted to phosphate and acetate under the catalysis of PAH (Scheme 3b). Phosphonoacetate is an intermediate of 2-aminoethylphosphonate (2-AEP) degradation pathway (28). As a member of the alkaline phosphatase superfamily, PAH contains a binuclear metal center and a highly conserved threonine or serine residue at the active site. In the proposed reaction mechanism, the conserved threonine or serine is recruited to attack the substrate to form a covalent intermediate bound to a phosphoryl group (28).

Two different PAH enzymes have been structurally and functionally characterized from *Sinorhizobium meliloti 102*1 and *P. fluorescens 23F*. Both of these enzymes can be activated by Zn^{2+}, but Zn^{2+} can be replaced by Mn^{2+} or Fe^{2+} in the active site for PAH from *S. meliloti*. The structure of these enzymes has been determined for the apo-enzyme andr complexed with substrate/inhibitor (PDB id : 1EI6, 3SZY, 3T02, 3SZZ, 3T00, 3T01) (29). On the basis of these crystal structures, PAH possesses an unprecedented capping domain to shield the active site from solvent during the reaction. A detailed reaction mechanism for the PAH is illustrated in Scheme 5 (30).

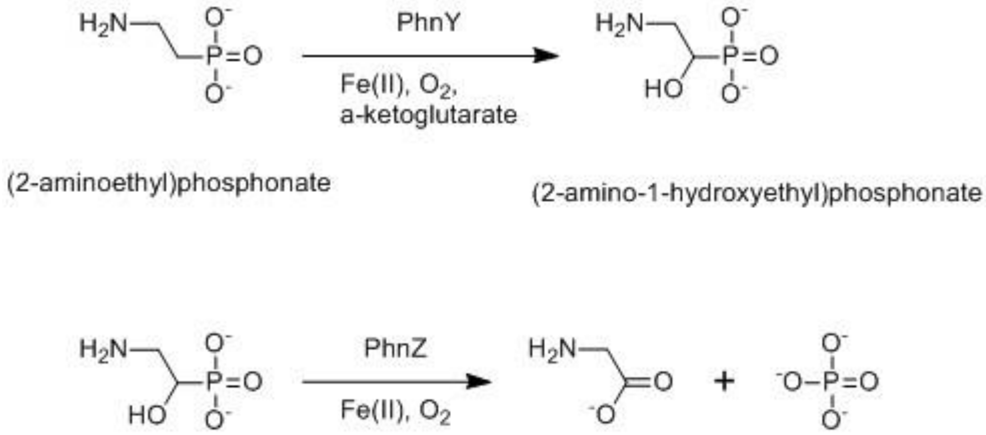
Phosphonoacetaldehyde hydrolase (PaldH) is also a member of the phosphonate hydrolase group. The substrate, phosphonoacetaldehyde (Pald), is synthesized by the reaction catalyzed by 2-AEP-pyruvate transaminase, which coverts 2-AEP and pyruvate to L-alanine and the phosphonoacetaldehyde. The phosphonoacetaldehyde is further hydrolyzed by PaldH to form phosphate and acetaldehyde as products (Scheme 4c) (20).

As a member of the haloalkanoic acid dehydrogenase (HAD) superfamily, Mg^{2+} in its active site will facilitate the covalent modification of an active site aspartate residue by the substrate(31). There are several available structures for PaldH from *Bacillus cereus*: PaldH complexed with Mg^{2+} (PDB id: 1RQN), vinyl sulfonate (PDB id: 1RQL), and a trapped covalently modified intermediate (PDB id: 2IOF) (32-34). A triad of residues (lysine, histidine and aspartic acid) in the active site has been identified, based on structural and functional studies. The lysine residue attacks the carbon atom of the carbonyl group of the substrate. A water molecule is activated by the histidine residue to attack the substrate and cleaves the carbon phosphorus bond. Then the phosphoryl group is transferred to the aspartate residue. Details of the proposed reaction mechanism are summarized in Scheme 6 (20).

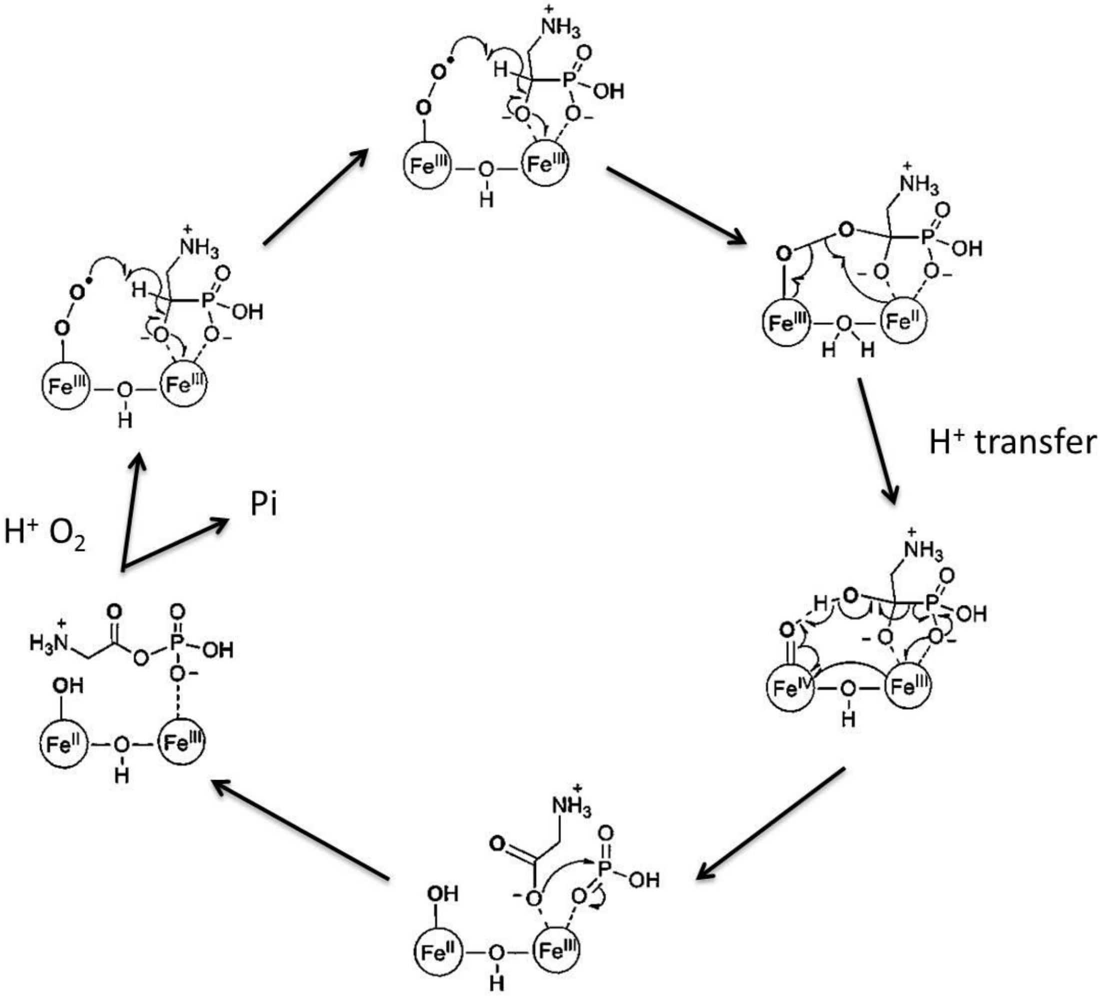
Recently a new class of organophosphonate degradation pathway was found. The gene product of *phnY* (2-oxoglutarate dioxygenase) and *phnZ* (a novel Fe^{2+} dependent phosphonate oxygenase) catalyze an oxidative cleavage of 2-aminoethylphosphonate (Scheme 7) (35). Sequence homology studies indicates that PhnY belongs to the non-heme Fe (II)/a-ketoglutarate-dependent dioxygenase family and PhnZ is a member of the metal-ion dependent phosphohydrolase superfamily (36). The oxidative cleavage of 2-AEP is catalyzed by PhnY and PhnZ sequentially. In the first step, 2-AEP is converted into 2-amino-1-hydroxyl ethylphosphonic acid (2-α-1-HEP) under the catalysis of PhnY. PhnY has very high specificity for the substrate 2-AEP so it cannot catalyze the hydrolysis of other phosphonates (37). The proposed mechanism of the reaction is shown in Scheme 8.

In the next step, PhnZ recruits molecular oxygen to oxidatively hydrolyze the carbon phosphorus bond of 2-α-1-HEP, yielding phosphate and glycine as products. A binuclear iron center in the active site plays an essential role in the oxidation process of the substrate through the formation of a Fe (III)-superoxo species. A detailed reaction mechanism for the hydrolysis of the C–P bond of 2-α-1-HEP is illustrated in Scheme 8 (35).

New biological pathways for phosphonate metabolism were also explored and studied. Fosfomycin is a widely used organophosphonate antibiotic. It is a secondary metabolite of some bacterial species such as *Streptomyces wedmorensis, S. fradiae, S. viridochromogenes* and *Pseudomonas syringae* (19). It contains a C-P bond and an oxirane ring with its structure. Previous studies indicated that bacteria *Mesorhizobium huakuii* was able to utilize fosfomycin as a sole carbon and phosphorus source (38). The fosfomycin degradation pathway is initiated with the hydrolysis of oxirane ring to yield the (*1R*, *2R*)-1,2-dihydroxypropylphosphonic acid (39). Then (*1R*, *2R*)-1, 2-dihydroxypropylphosphonic acid is oxidized to form (*R*)-1-hydroxy-2-oxopropylphosphonic acid. However, other reactions involved in fosfomycin degradation pathway are still unclear (19).



Scheme 7. Oxidative cleave of 2-aminoethylphosphonate catalyzed by PhnY and PhnZ (35)



Scheme 8. Proposed Reactions and Mechanisms for PhnY and PhnZ (35)

1.1.5 Discovery of bacterial C-P lyase pathway

The major route for the metabolism of phosphonates is the C–P lyase pathway. The C-P lyase activity is very promiscuous. It can hydrolyze a broad spectrum of phosphonates containing either activated or unactivated carbon phosphorus bonds. The first evidence for the cleavage of unactivated C-P bonds was reported by Zeleznick *et al* in 1963. They used methyphosphonate or ethylphosphonate as the sole phosphorus source for sustainable *Escherichia coli* growth (40). The Frost laboratory revealed that ribosyl intermediates were invloled in the pathway in 1980s, despite unsuccessful attempts to purify the proteins in C-P lyase pathway (2).

Although the activity of C-P lyase was identified in a variety of bacteria species, the enzymatic activity was never found in the cell free extracts (2). This obstructed the effort of clarifying the chemical reactions of the C-P lyase pathway. Several hypotheses had been proposed for the mechanism of enzymatic cleavage of the C-P bond (17, 40, 41). The first hypothesis proposed the hydrolysis of C-P bond via the oxidation of the α-carbon of alkyphosphonate forming α-hydroperoxy, α-hydroxy, α-keto, or phosphor-monoesters (40). However, there was no proposed intermediate identified in the followed studies (36-38). The most distinctive feature of degradation of methylphosphonate is the ratio between the released molecular methane and the formed cellular phosphate is 1:1. The ratio could be also apply to other alkyphosphonates forming corresponding alkanes such as ethane, propane, butane, pentane, and hexane. In addition, further studies of products of C-P lyase pathway also revealed the presence of corresponding alkenes such as ethene, propene and butane (42).

The second hypothesis proposed a redox dependent radical based dephosphorylation mechanism (41). In the first step, a phosphonyl radical is generated and then cleaved to form an alkyl radical and phosphate. Then the alkyl radical abstracts a hydrogen to yield the corresponding alkane. The presence of alkenes among degradation products suggested that the radical dependent mechanism was possible. The radical based mechanism of the cleavage of C-P bond was also supported by later studies (43, 44). However, alternative hypothesis proposed that the cleavage of C-P bond involved ATP or other nucleotides (44). The identification of intermediate, α-1-ethylphosphono-ribose, was the evidence that supported this hypothesis. It suggested that nucleotides may react with the phosphoryl group of alkylphosphonates to form a ribose-phosphonate intermediate. Despite these great efforts, enzymatic C-P bond cleavage *in vitro* has not been achieved until 2011(45).

1.1.6 Gene organization and functions of enzymes in C-P lyase pathway

The *phn* gene operon that is responsible for C-P lyase pathway was identified by Wash CT *et al* in 1990. Up to 17 genes encoding PhnA to PhnQ was determined to be required for bacteria to utilize unactivated phosphonates as a sole phosphorus source by deletion analysis (46). The identified genes were cloned, mutated and transformed into various plasmids by B. L. Wanner *et al* (35). A collection of plasmids that contained genes from *phn operon* was also constructed by B. L. Wanner *et al* (41). Plasmid libraries containing insertions of different length of DNA fragments from *phn operon* were transformed into *E. coli* host that contained a deletion of the *phn* locus, and these *E.*

coli strains were screened for ones that could utilize phosphonate as a sole P source. A 10.4 kB of DNA that is essential for complementation of *a phn* deletion mutant was identified. This DNA fragment contains fourteen genes (*phnCDEFGHIJKLMNOP*). Further studies suggested that fourteen genes were transcribed as a single operon that is up-regulated by the PhoR/PhoB two-component signaling system (41).

Three genes (*phnCDE)* encode an ATP-binding cassette (ABC) transport system for phosphonate. Previous studies showed that *phnCDE* gene products provided active transport of 2-aminoethyl [32P] phosphonic acid into *E. coli* cells (41). The PhnC protein is an ATP binding component of the ABC transporter system, PhnD is a periplasmic phosphonate binding protein, and PhnE is a trans-membrane protein (41). The PhnD protein also contains a signal sequence for transmembrane export to the periplasm. Affinity analysis has shown that PhnD is able to bind phosphonates with a K_d as low as 5 nM (47). DNA sequence analysis of the *E. coli* K-12 *phn gene operon* revealed that an 8 base pairs insertion mutation in the *phnE* gene leads to the cryptic phenotype of *E. coli* K-12 strain (48).

Based on sequence similarity studies, PhnF is a transcriptional repressor protein of the C-P lyase pathway. Seven genes (*phnGHIJKLM*) have been proposed to encode the essential cassette for carbon-phosphorus (C-P) lyase activity; and three genes (*phnNOP*) have been assigned to encode enzymes with auxiliary functions (19).
The *phnO* gene was characterized to produce an aminoalkyl phosphonate-*N*-acetyltransferase (49). PhnO catalyzed the reactions through a sequential kinetic mechanism with acetyl-CoA binding first and then followed by aminoalkylphosphonate.

The thioester of acetyl-CoA was attacked by the amine of aminoalkylphosphonate forming a tetrahedral intermediate. The intermediate was hydrolyzed to yield the product, *N*-acetylated aminoalkylphosphonate. As a member of the Gcn5-related N-acetyltransferases (GNAT) superfamily, PhnO unprecedentedly requires a divalent metal ion for catalysis. The *phnO* gene is essential for hydrolysis of 1-aminoalkylphosphonates such as aminomethylphosphonic acid and *S*-1-aminoethylphosphonate (49).

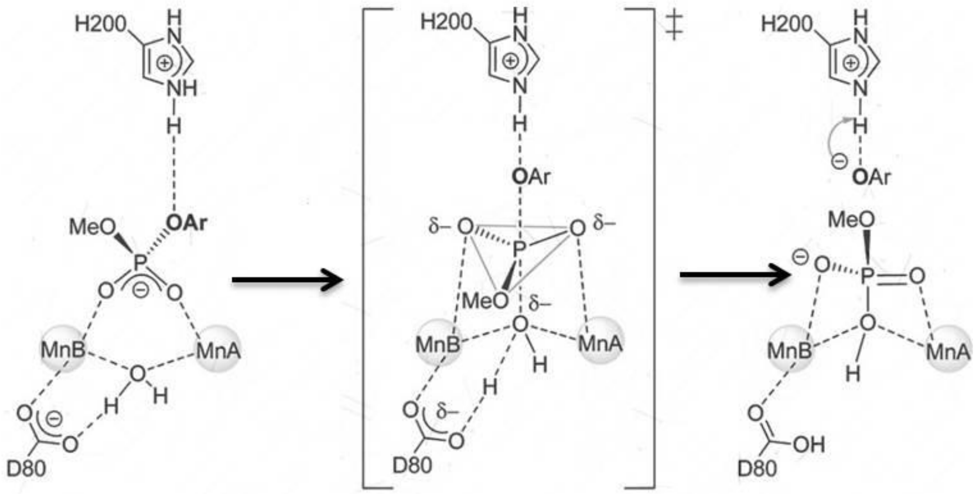
The product of *phnN gene* was identified as a ribosyl 1, 5-bisphosphokinase which can phosphorylate ribose 1, 5-bisphosphate to yield 5-phosphoribosyl α-D-1-diphosphate (PRPP) (50). Genetic studies have shown that *phnN* is not necessary for phosphonate utilization in *E. coli*, which may be attributed to the presence of PhnN redundant activity in the cells (41). The fact that the substrate of PhnN, ribosyl 1, 5-bisphosphate, is also found in cells without the presence of phosphonate in the surroundings suggests the existence of PhnN redundant activity in the cells (51).

As a member of the metallo-β-lactamase superfamily, PhnP is a phosphodiesterase that can recycle a "dead-end" intermediate, 5-phospho-α-ribosy-l , 2-cyclic phosphate, which is formed upon the cleavage of the C-P bond (52). The structure of PhnP has been determined with inhibitor orthovanadate (PDB id: 3P2U). On the basis of the structural model and a Brønsted analysis of PhnP reactions, a mechanism for PhnP is proposed (Scheme 9). In this mechanism, an attacking water molecule is activated by two manganese ions in the active site. The water molecule is then further activated by deprotonation by D80 (52).

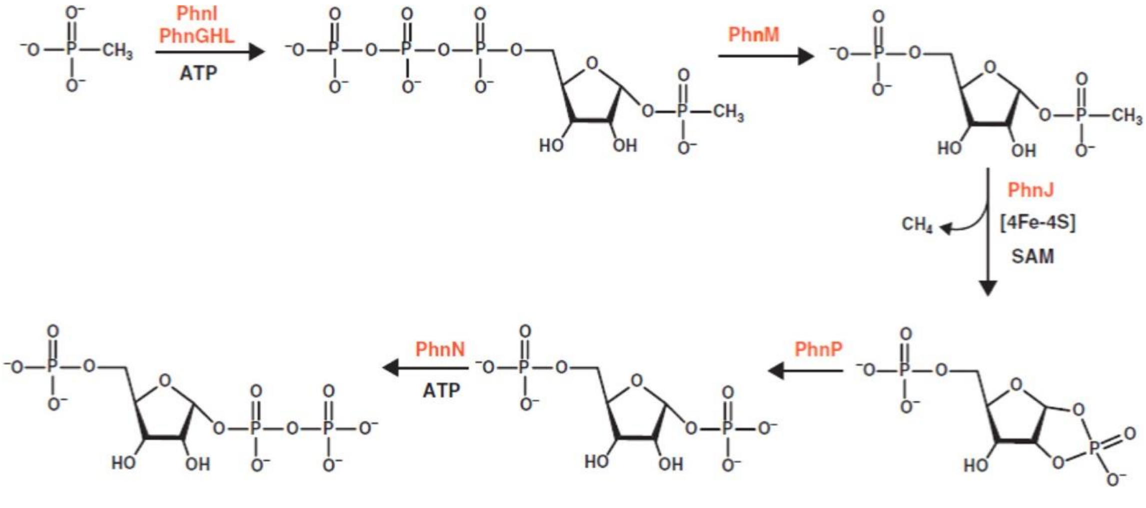
1.1.7 Essential reactions of cleavage of C-P bond

Using methylphosphonate as a model substrate, the essential biochemical reactions for cleavage of C-P bonds were reconstituted *in vitro* (Scheme 10). PhnI, a nucleosidase, catalyzes the displacement of adenine from ATP by methylphosphonate (MePn) to form ribose-1-methylphosphonate-5-triphosphate (RPnTP), in the presence of PhnG, PhnH, and PhnL. In the second step, RPnTP is hydrolyzed by PhnM, a pyrophosphatase, to generate ribose-1-methylphosphonate-5-phosphate (PRPn), and in the last step PhnJ, a SAM radical enzyme, catalyzes the cleavage of C-P bond in PRPn to form molecular methane and ribose-1, 2-cyclic phosphate-5-phosphate (PRcP) (45).

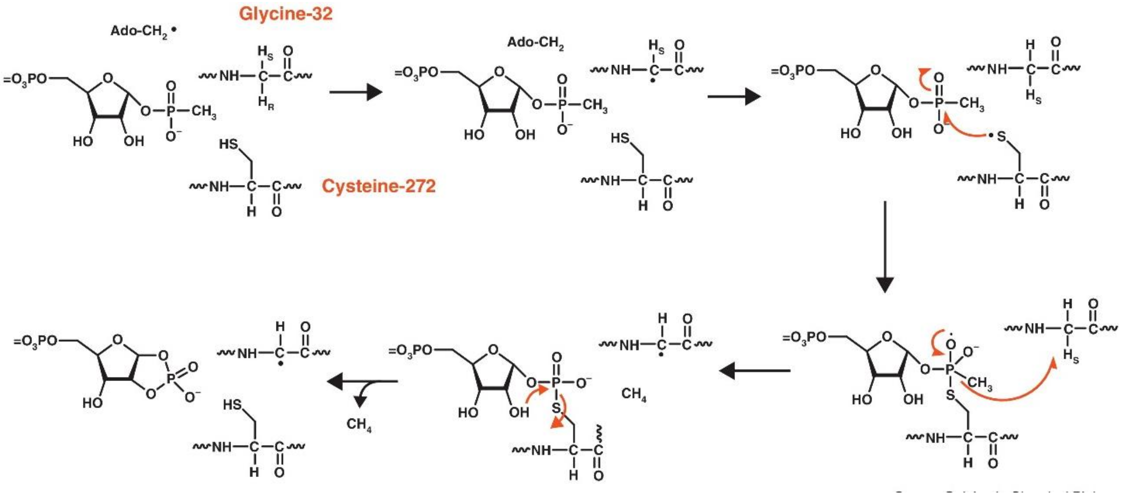
Anaerobic conditions are required for the reaction of the cleavage of C-P bond. Spectroscopy revealed the existence of a 4Fe-4S cluster within PhnJ. In addition, the requirement of the *S*-adenosylmethionine (SAM) for the reaction suggests that the C-P bond is cleaved via a typical radical SAM mechanism. Three cysteine residues (Cys241, Cys244, and Cys266) take part in the formation of the Fe_4S_4 cluster, whereas Cys272 is supposed to participate the reaction as a thiyl radical. The mechanism of the cleavage of carbon phosphorus bond has been proposed (53). A reduced Fe-S cluster cleaves SAM to generate a 5'-deoxyadenosyl radical intermediate. Then a hydrogen atom on Gly-32 will be abstracted by the radical intermediate to form a glycyl radical. The radical will then transfer to Cys-272 to produce a thiyl radical. To cleave the C-P bond, the thiyl radical attacks the methylphosphonate moiety of PRPn to form a transient radical intermediate. The C-P bond will be hydrolyzed by the collapse of the transient radical intermediate. The final products are PRcP and methane. The proposed reaction mechanism is presented in Scheme 11 (53).



Scheme 9. Proposed mechanism for the PhnP reaction (52)



Scheme 10. Reactions *of* C-P lyse pathway in *E. coli* (54)



Scheme 11. Proposed mechanism for the reaction catalyzed by PhnJ (20)

As the product of PhnJ catalyzed reaction, PRcP will be further hydrolyzed by PhnP to ribose-1, 5-bisphosphate (RbP). In the next step, RbP is converted to 5-phosphoribosyl-1-pyrophosphate (PRPP) via the catalysis of PhnN (50). The reactions of C-P lyase are illustrated in Scheme 10.

1.1.8 Bacterial species variation of *phn operon*

In different bacteria species other than *E .coli*, the organization of *phn gene operon* could also be different (55). Studies of *phn operon* in different bacteria species reveals two possible organizations of the genes encoding the phosphonate transport proteins, either *phnCDE*, as in *E. coli*, *Burkholderia pseudomallei*, and the *Pseudomonas stutzeri phn* operon, or *phnCDEE*, with two distinct transmembrane proteins, as in *Sinorhizobium meliloti*, *O. anthropi*, *Agrobacterium radiobacter*, *Mesorhizobium loti*, *Bradyrhizobium japonicum*, *Nostoc* sp. strain PCC7120, and *Trichodesmium erythraeum* (55). In addition, a new *operon htx* that is responsible for uptake and oxidation of hypophosphite and phosphonate was also identified in *P. stutzeri htx* (56). The actual location of genes *phnCDE*(*E*) is not always upstream of the remaining genes in *phn operon* as *E. coli*. In *O. anthropi*, *M. loti*, and *B. pseudomallei*, the position of *phnCDE(E)* genes are downstream of other genes in *phn operon (57)*. In *S. meliloti*, the *phnCDE(E)* genes are located far away from the remaining genes in *phn operon*. In most species, *phnCDE(E) genes* are transcribed in the same direction as the remaining genes in C-P lyase pathway, however, in *Nostoc sp. PCC7120*, the *phnCDE* genesare are transcribed in a divergent direction (55).

In some species, *phnGHIJKLM* genes are not located as a contiguous gene cluster. The location of the *phnM* gene is away from the *phnGHIJKL* genes in *S. meliloti*. There are two copies of *phnM* in *O. anthropi* and *A. radiobacter*. The location of *phnO* gene is downstream of and overlapping the *phnL* gene in *S. meliloti*, *O. anthropi*, *A. radiobacter*, and *M. loti*. *PhnO* gene is located very far away from the rest genes of *phn operon* in *Nostoc* sp. PCC7120 (5). In many species, there is no *phnP* gene in their *phn operon*, but in all strains where *phnP* is absent, *duf1045* gene is a substitute in *phn operon (55)*. This gene encodes an enzyme PhnPP which is a 5-phosphoribosyl 1, 2-cyclic phosphate dihydrolase. PhnPP was isolated from the Gram-positive organisms *Eggerthella lenta* and *Clostridium difficile*. It hydrolyzes 5-phospho-ribose-1, 2-cyclic phosphate at α-carbon of the ribose moiety to yield an intermediate, ribose-2, 5-bisphosphate. And then ribose-2, 5-bisphosphate is hydrolyzed to ribose-5-phosphate and inorganic phosphate (58).

Despite the widely spread *phn operon* in *Eubacteria*, C-P lyase activity was also identified in a few *archaeal* species for phosphonate catabolism. For example, the chromosomes of *Halorhabdus tiamatea* SARL4B (GenBank accession no. HF571520.1) (59), *Haloquadratum walsbyi* C23 (accession no. FR746099.1) (60), and *Natronomonas moolapensis* 8.8.11 (accession no. HF582854.1) (60) all contain *phnGHIJKLM* genes. Located upstream of the *phnG* gene, but transcribed in a divergent direction, the three species contain three genes *phnDCE* (*H. tiamatea*) or four genes *phnDCEO* (*H. walsbyi* and *N. moolapensis*). There were no phnN or phnP gene *or* homolog identified in any archaeal organisms. The physiological role of the *phn operon* is still unclear (55).

1.1.9 Multi-protein complex formed by enzymes of C-P lyase pathway

PhnG, PhnH, and PhnM have been expressed and purified via standard methods. However, PhnI, PhnJ, PhnK, and PhnL have only been expressed and purified as glutathione S-transferase (GST) conjugates(45). The instability of these proteins in the absence of solubility tags has significantly hampered the structural elucidation of those enzymes responsible for phosphonate metabolism. Among those enzymes, PhnH is the only protein that has been structurally characterized as a homodimer. Although it displays a fold related to the pyridoxal-5'-phosphate-dependent transferases, the role of PhnH is still unclear (61).

Zechel, Hove-Jensen, and colleagues have recently demonstrated that multi-subunit fragments of the C–P lyase complex can be expressed from plasmids in *E.coli* and subsequently purified to apparent homogeneity. Partial complexes containing PhnG and PhnI (PhnG-I); PhnG, PhnH, PhnI, and PhnJ (PhnG–H-I-J); and PhnG, PhnH, PhnI, PhnJ, and PhnK (PhnG-H-I-J-K) were expressed and purified in high yield with high solubility and stability without the need for solubility tags (62).

1.2 Bacterial Phosphotriester Metabolism

1.2.1 Environmental concerns raised by the application of phosphotriesters

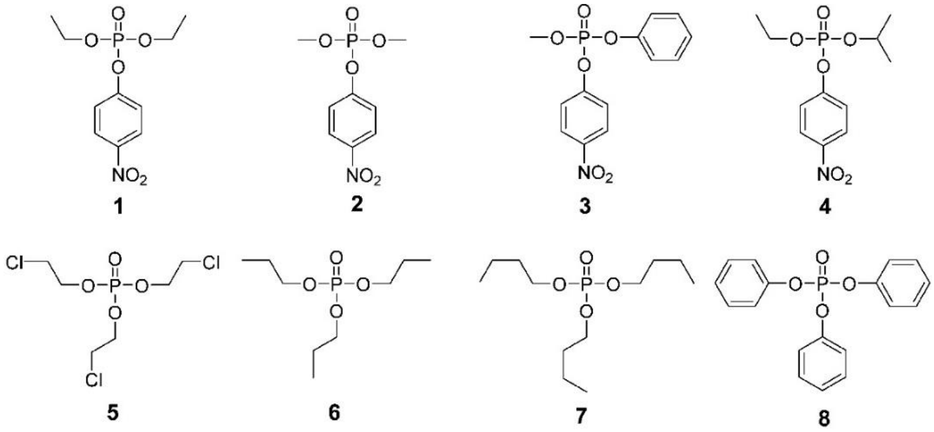
Phosphotriesters are a group of organophosphates that were initially developed during World War II. This group of compounds was first utilized as insecticides for crop protection (63). The structure of this group of compounds is a phosphate center with three ester linked functional groups. The chemical structures of common synthetic phosphotriesters are illustrated in Scheme 1. The most well-known phosphotriesters are the neurotoxic nerve agents and insecticides such as paraoxon (compound 1 in Scheme 12). These compounds can irreversibly inhibit the neural enzyme acetylcholine esterase by the enzymatic cleavage of a labile ester bond (64). In recent years, a related group of organophosphotriesters has gained popularity for use as plasticizers and flame retardants in lubricants, durable plastics, and foams (65, 66). Compounds such as tris(2-chloroethyl) phosphate (compound 5 in Scheme 12) (65) and triphenyl phosphate (66), unlike the neurotoxic organophosphates, do not contain a labile ester bond but rather make use of the stable nature of these compounds to inhibit combustion and improve the performance of other organic compounds.

While not neurotoxins, many of the flame retardants contain halogenated alkyl or aromatic esters and are known, or suspected, chemical carcinogens (66). For example, tris (2-chloroethyl) phosphate could have adverse effects on the nervous system, digestive system, circulatory system and reproductive system of male rats and mice (67). These compounds are currently being produced at hundreds of metric tons per year (2, 3). The contamination of the environment by organophosphorus flame retardants has become widespread, and these compounds are now considered to be major emerging pollutants (3).

1.2.2 Bacterial degradation of phosphotriesters

In various species of bacteria, natural and synthetic organophosphotriesters are hydrolyzed by phosphotriesterase (PTE) (68-71). To utilize the man-made organophosphotriesters as phosphorus source in the environment, bacteria evolved the ability to hydrolyze this type of compounds. The structural similarity of the phosphoryl centers between synthetic and natural substrates may play an essential role for the development of the promiscuous activity towards synthetic organophosphortriesters (72).

Structural models of PTEs from various species reveal different protein folding patterns of PTEs including TIM-barrel, β-lactamase, "pita-bread", and β-propeller folds (73-77). Despite the varied protein structural folds, PTEs are known to possess divalent metal center(s) to bind and activate substrates for hydrolysis (71, 74, 78, 79). PTEs with different folding patterns hydrolyze the organophosphotriesters via different chemical mechanisms. The enzymes with β-propeller folding are proposed to go through a covalent reaction intermediate, while the TIM-barrel and the "pita-bread" enzymes usually hydrolyze the substrate by direct attack of water/a hydroxide (80-82). The chemical mechanism of phosphotriesterase with a β-lactamase fold is not unknown (83).

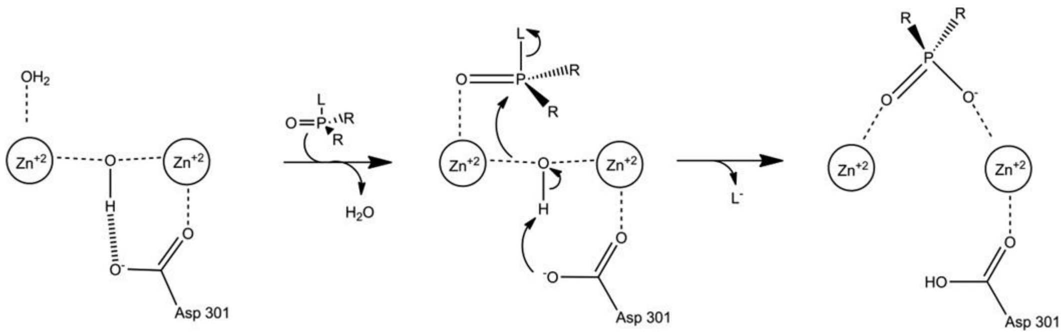


Scheme 12. Common synthetic organophosphotriesters

1.2.3 Phosphotriesterase with TIM barrel fold

As a member of the amidohydrolase superfamily, the structure and chemical mechanism of phosphotriesterases (PTE) with a TIM barrel fold are well studied. This group of enzymes was first identified from the bacterium *Pseudomonas diminuta* (*Pd*-PTE) and *Flavobacterium* sp. in phosphotriester contaminated soil (84, 85). Other enzymes of this group were identified in different soil bacterial species such as *Agrobacterium radiobacter* (OpdA) and *Chryseobacterium balustinum* (86, 87). The identified enzymes of this group share more than 90% sequence identity (88). Moreover, these enzymes also have very similar substrate profiles and kinetic parameters (88). The X-ray structures of these enzymes revealed a distorted $(\beta/\alpha)_8$ or TIM-barrel protein fold (88-91) (Figure 2A). The fold of these proteins contains a core barrel of eight α-helices and eight parallel β-strands. The α-helices and β-strands in the barrel are connected to one another by flexible loops. A binuclear metal center is ligated to histidine and aspartic acid residues in the active site that is located at the C-terminal end of the TIM-barrel (73). The α-metal ion that is ligated two histidine and an aspartate is buried under the surface of the protein while the β-metal that is solvent accessible is ligated by two histidine residues (92, 93) (Figure 2B). The active site is quite promiscuous for metal ions. Although the native enzyme contains Zn^{2+} in its active site, Cd^{2+}, Mn^{2+}, or Ni^{2+} can also result in active enzymes (78).

The activity of this group of enzymes is very promiscuous. It can hydrolyze a broad spectrum of phosphotriesters (83). Containing an electron withdrawing phenolic leaving group, paraoxon is one of the best substrates (83). The catalytic efficiency with paraoxon can be close to the limits of diffusion (94). In addition, PTE is also able to hydrolyze phosphotriesters containing thiol linkages and halide bonds (83). The broad substrate spectrum of this group of enzymes is proposed to be attributed to the promiscuity of the substrate binding site. Three hydrophobic binding pockets on the surface of the enzyme can fit three ester groups of the phosphotriesters (95). In the proposed catalytic mechanism of PTE with a TIM barrel fold, an activated water molecule directly attacks the phosphorus center of the substrate (Scheme 13) (89). Mutagenesis and kinetic studies of PTE have indicated that that the residues in the binding pockets determines the substrate specificity for PTE (96). Based on the available structural information, catalytic efficiency for nerve agents and insecticides has been optimized by directed evolution (83).



Scheme 13. Proposed chemical mechanism of PTE with the TIM-barrel folding (97).

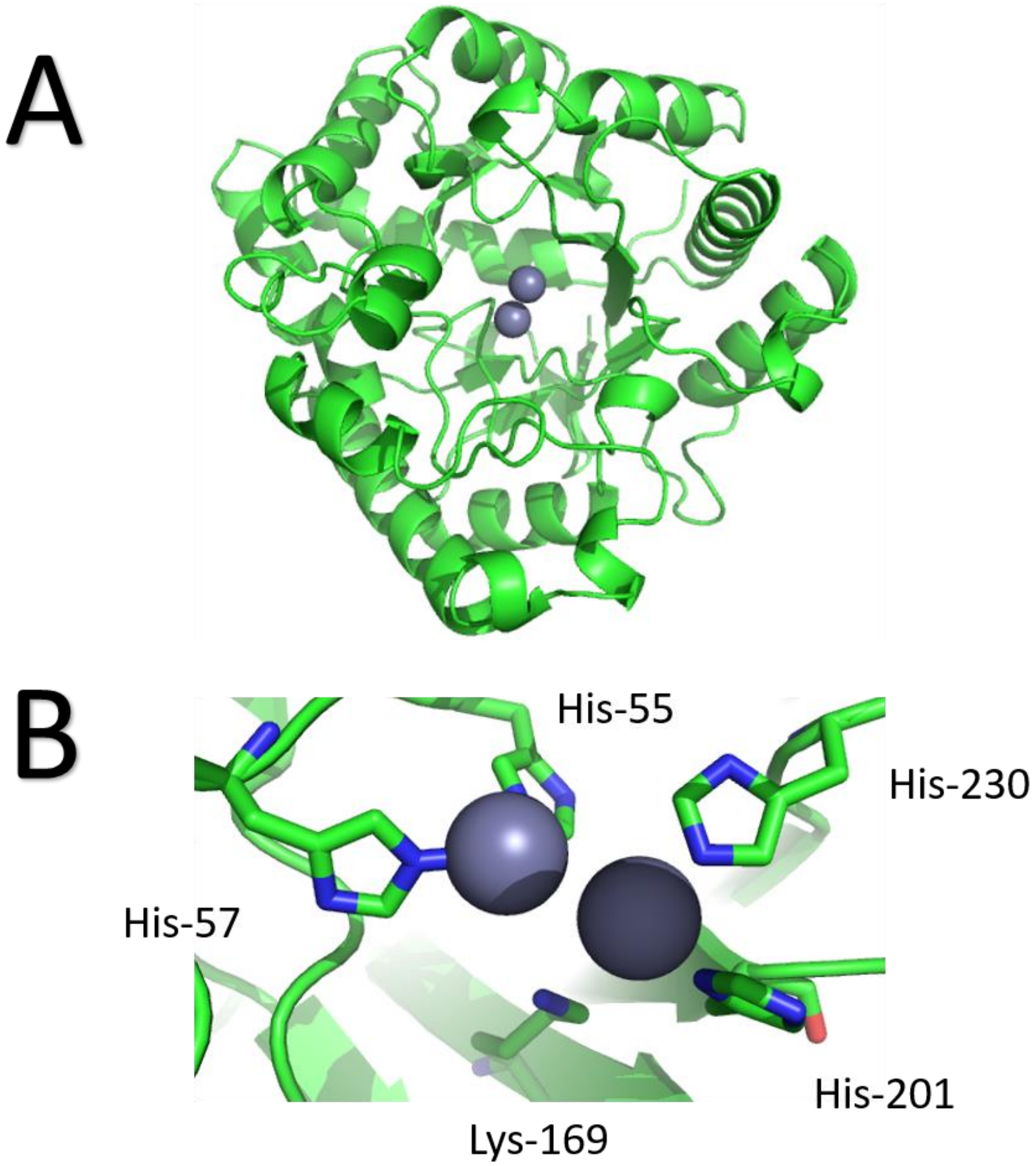
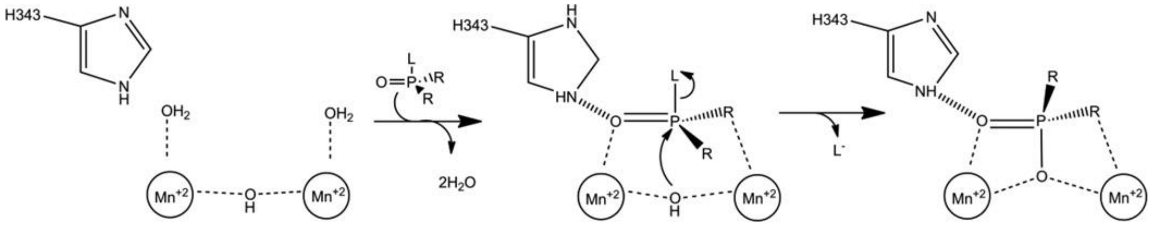


Figure 2: Crystal structure of PTE with TIM-Barrel folding. (PDB id: 1DPM). (A) TIM-barrel fold of PTE. Metals in active site are shown as spheres. (B) Metal binding site of PTE.

1.2.4 Phosphotriesterase with “pita bread” fold

The best characterized enzyme of the group of phosphotriesterase with a “pita bread” fold is organophosphate acid anhydrase (OPAA) from *Alteromonas sp.* (71). The overall architecture of this enzyme is a tetramer that can further be divided into two dimers (98). A binuclear Mn^{2+} site is located in the central cleft of the "pita bread" fold at the C-terminal domain (Figure 3A). The two Mn^{2+} ions are bridged by two carboxylate residues (98). Based on the X-ray structure of OPAA, these two metals are proposed to be bridged by an additional hydroxide (98) (Figure 3B). The proposed chemical mechanism of OPAA is illustrated in Scheme 14. The direct attack at the phosphorus center of the substrate by the bridging hydroxide is observed by O^{18} labeling mass spectroscopy and NMR analysis (77).



Scheme 14. Proposed chemical mechanism of PTE with the Pita-Bread fold (97).

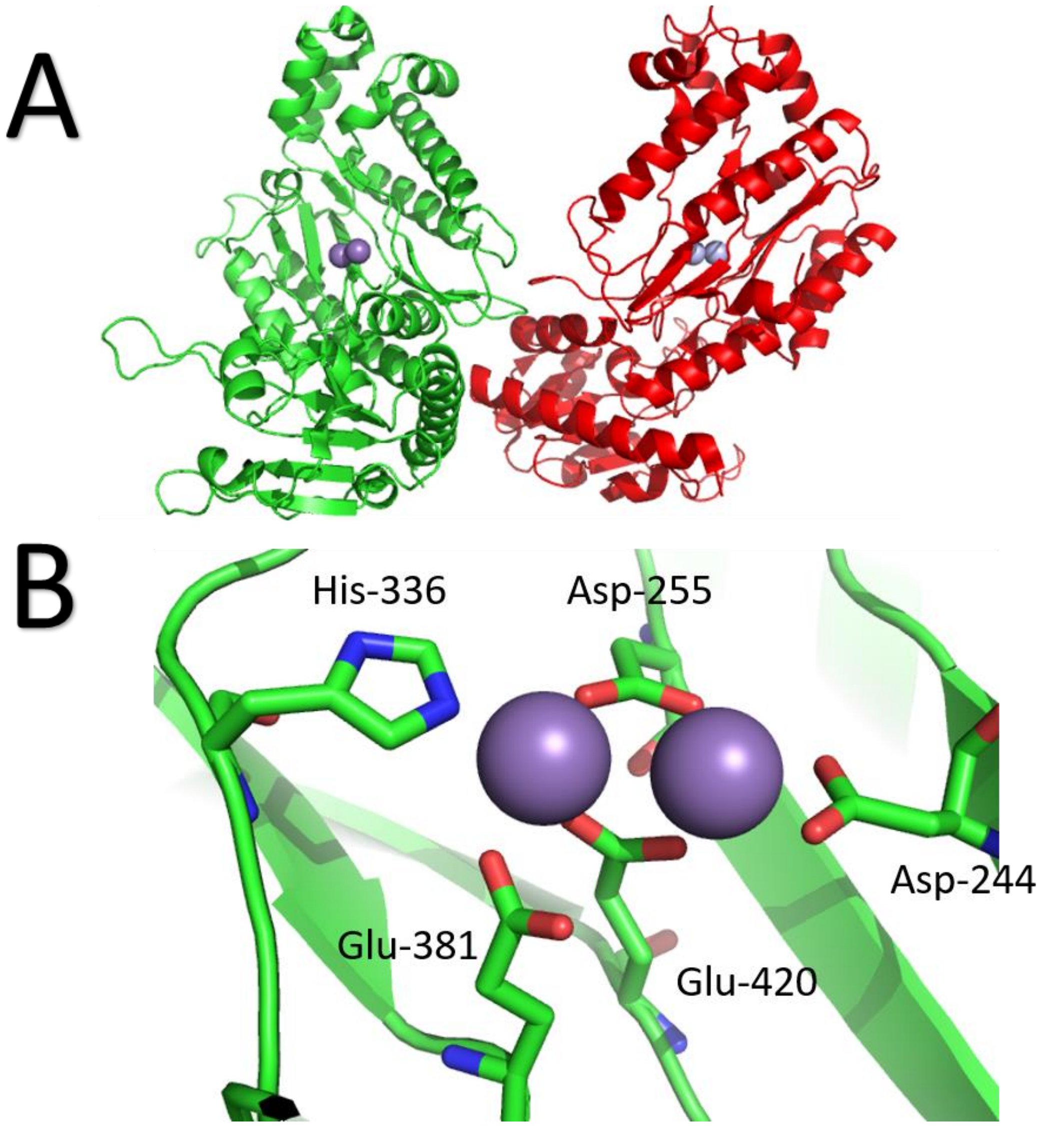
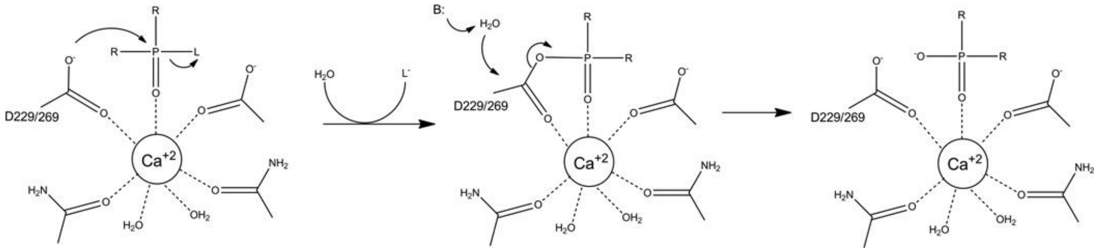


Figure 3: Crystal structure of OPAA, a representative of PTE with a “pita-bread” fold (PDB id: 3l7G). (A) Dimeric structure of OPAA. One momomer is shown with in green and the other monomer is shown in red. Mn^{2+} is shown as a sphere. (B) Expanded view of the OPAA metal center.

1.2.5 Phosphotriesterase with β-propeller fold

Paraoxonase (PON1) and diisopropylfluorophosphatase (DFPase) are PTEs with a β-propeller fold. These two enzymes were identified in mammalian serum and sea squid, respectively (99, 100). Based on the crystal structures, DFPase and PON1 are Ca^{2+} containing enzymes with a six-bladed β-propeller fold (75, 101). The X-ray structure of DFPase is illustrated in Figure 4. Each propeller typically has four pleated β-strands. Two Ca^{2+} are located in the water filled cleft at the center of the β-propeller. One Ca^{2+} is considered to stabilize the β-propeller fold and doesn't participate the catalytic process (79). This Ca^{2+} is ligated to aspartic acid or histidine residues as well as three water molecules. (Figure 4B) (79). It is proposed to maintain the water network in the center of the six-bladed β-propeller fold (79). The other Ca^{2+} is more solvent accessible and takes part in the catalytic process (79). The proposed mechanism for phosphotriester cleavage by PON1 and DFPase is illustrated in Scheme 15. In the proposed reaction mechanism, the reaction proceeds via a covalent enzyme bound intermediate (76, 80). In the crystal structure of DFPase with a substrate analog bound, a conserved aspartate residue is aligned for backside attack to the phosphorus center of the substrate. At the same time, a conserved histidine residue is located in an orientation towards the leaving group of the substrate (80). This histidine residue is proposed to serve as a base for the deprotonation of water molecule while the carboxyl group of the aspartic acid was proposed to attack the substrate yielding a phosphoacyl intermediate. The formation of a phosphoacyl intermediate was confirmed by ^{18}O labeling mass spectrometry under single and multiple turnover conditions (80).



Scheme 15. Proposed catalytic mechanism for phosphotriesterase reaction catalyzed by PTE with β-propeller folding (97).

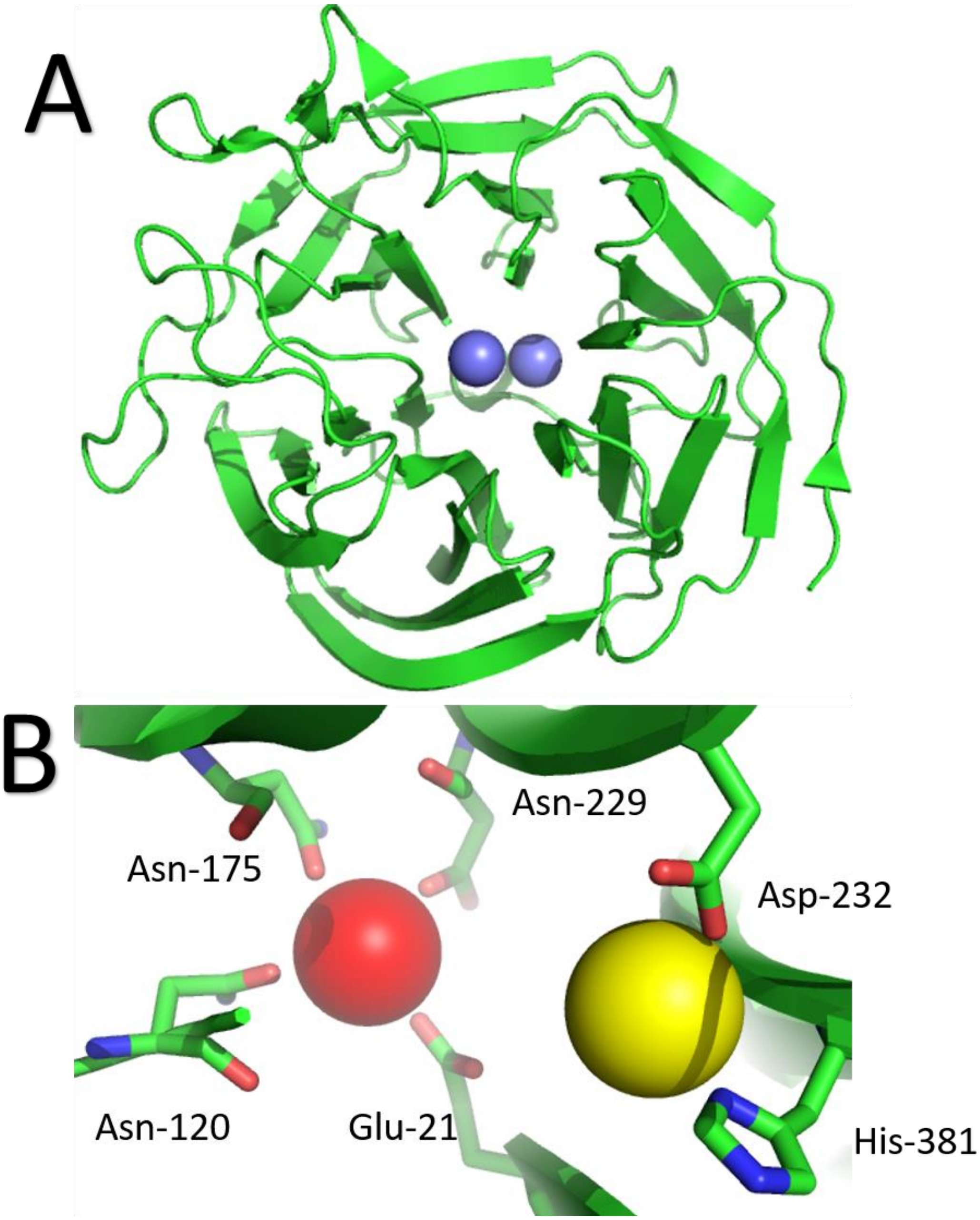


Figure 4: Crystal structure of DFPase, PTE with β-propeller folding (PDB id: 2GVV). Calcium ions are shown as spheres. Catalytic calcium is shown in red while structural calcium is shown in yellow. (A) Top down view of DFPase showing the 6 bladed β-propeller structure. (B) An extended view of DFPase metal center ligation.

1.2.6 Discovery of novel phosphotriesterases

Due to the lack of a sufficiently labile leaving group, currently identified phosphotriesterases have little or no activity against organophosphorus flame retardants and plasticizers (88, 102-104). Two novel phosphotriesterase-like enzymes from *Sphingomonas sp.* strain TDK1 (Sm-PTE) and *Sphingobium sp.* strain TCM1 (*Sb*-PTE) have recently been reported to hydrolyze organophosphates that are not substrates for other phosphotriesterases, such as *Pd*-PTE or OpdA (103). When this study was initiated, very little was known about these enzymes aside from their amino acid sequence and their dependence on the addition of divalent metal ions for catalytic activity. In addition, there was no structural or mechanistic information available for either *Sm*-PTE or *Sb*-PTE, or any close homologue of known function.

CHAPTER II

STRUCTURAL CHARACTERIZATION OF CARBON PHOSPHORUS MULTI-PROTEIN COMPLEX[*]

2.1 Introduction

Phosphorus, an essential element for life, is a limiting nutrient in various ecosystems. The most abundant phosphorus sources are orthophosphate and phosphate esters, which contain a C-O-P bond. However, organophosphonates may have been the predominant phosphorus source on early earth when oxygen was not a major component of air (1). In contrast to the orthophosphate and phosphate esters which contain a C-O-P bond, the C-P bond in organophosphonates is highly resistant to chemical hydrolysis, thermal decomposition and photolysis (2). Compounds containing a C-P bond constitute as much as 20% to 30% of the available phosphorus in nature (3). Phosphonates are the main cellular phosphorus-containing compounds in several organisms. In addition, the organophosphonates are also widely used in the pharmaceutical and agricultural

---

[*] Part of this chapter is reprinted with permission from “Subunit Interactions within the Carbon–Phosphorus Lyase Complex from Escherichia coli” by Zhongjie Ren, Soumya Ranganathan, Nathanael F. Zinnel, William K. Russell, David H. Russell, and Frank M. Raushel, *Biochemistry* **2015** *54* (21), 3400-3411, Copyright 2016 by American Chemical Society.

Part of this chapter is reprinted with permission from “Structures of the Carbon-Phosphorus Lyase Complex Reveal the Binding Mode of the NBD-like PhnK” by Kailu Yang, Zhongjie Ren, Frank M. Raushel, Junjie Zhang, *Structure* **2015** *24*(1), 37-42, Copyright 2016 by Elsevier Inc.

industry. Today, natural and man-made phosphonates are found to be widespread in the environment.

The bacterial catabolism of phosphonates as an alternative phosphorus source has been widely studied. Currently, three groups of distinct organophosphonate degradation enzymes or enzymatic pathways have been characterized, including the C-P lyase, phosphonate hydrolases, and an oxidative pathway (20). The major route for the metabolism of phosphonates is the C–P lyase pathway. The C-P lyase activity is very promiscuous. It can hydrolyze a broad spectrum of phosphonates containing either activated or unactivated carbon phosphorus bonds. Although the activity of C-P lyase was identified in a variety of bacteria species, the enzymatic activity was never found in the cell free extracts (2). This obstructed the effort of clarifying the chemical reactions of the C-P lyase pathway. Despite these great efforts, enzymatic C-P bond cleavage *in vitro* has not been achieved until 2011(45).

The C-P lyase pathway is encoded by a 14-cistron *phnCDEFGHIJKLMNOP* operon. Further studies suggested that fourteen genes were transcribed as a single operon that is up-regulated by the PhoR/PhoB two-component signaling system (41).Seven genes (*phnGHIJKLM*) have been proposed to encode the essential cassette for carbon-phosphorus (C-P) lyase activity (19). Using methylphosphonate as a model substrate, the essential biochemical reactions for cleavage of C-P bonds were reconstituted *in vitro*. PhnI catalyzes the displacement of adenine from ATP by methylphosphonate (MePn) to form ribose-1-methylphosphonate-5-triphosphate (RPnTP), in the presence of PhnG, PhnH, and PhnL. In the second step, RPnTP is hydrolyzed by PhnM, a pyrophosphatase, to generate ribose-1-methylphosphonate-5-phosphate (PRPn), and in the last step PhnJ, a SAM radical enzyme, catalyzes the cleavage of C-P bond in PRPn to form molecular methane and ribose-1, 2-cyclic phosphate-5-phosphate (PRcP) (45).

PhnG, PhnH, and PhnM have been expressed and purified via standard methods. However, PhnI, PhnJ, PhnK, and PhnL have only been expressed and purified as glutathione S-transferase (GST) conjugates(45). The instability of these proteins in the absence of solubility tags has significantly hampered the structural elucidation of those enzymes responsible for phosphonate metabolism. Among those enzymes, PhnH is the only protein that has been structurally characterized as a homodimer. Although it displays a fold related to the pyridoxal-5’-phosphate-dependent transferases, the role of PhnH is still unclear (61). Zechel, Hove-Jensen, and colleagues have recently demonstrated that multi-subunit fragments of the C–P lyase complex can be expressed from plasmids in *E.coli* and subsequently purified to apparent homogeneity. Partial complexes containing PhnG and PhnI (PhnG-I); PhnG, PhnH, PhnI, and PhnJ (PhnG–H-I-J); and PhnG, PhnH, PhnI, PhnJ, and PhnK (PhnG-H-I-J-K) were expressed and purified in high yield with high solubility and stability without the need for solubility tags (62).

2.2 Materials and Methods

2.2.1 Materials

LB broth was purchased from Research Products International Corporation. The HisTrap and gel filtration columns were obtained from GE Healthcare. Mini-PROTEAN precast gels were purchased from Bio-Rad. Bis-sulfosuccinimidyl suberate (BS3), immobilized pepsin resin, and trypsin were purchased from Thermo Scientific. The C18 reverse phase column was purchased from Waters Corporation. D_2O was obtained from Cambridge Isotope Laboratories. All other buffers, purification reagents, and chemicals used in this project were purchased from Sigma/Aldrich, unless otherwise noted.

2.2.2 Cloning and expression of C–P lyase complexes

DNA containing the gene for PhnG through PhnK (his-PhnG-H-I-J-K and PhnG-H-I-J-K-his) from *E. coli* was amplified from the strain BW5328/pGY3 (CGSC, Yale University). For the his-PhnG-H-I-J-K construct, 5’-AGCTCGTCGAT GCGACA TATGCACGCAGATACCGCGACCCGCC-3’ and 5‘-TCAAGTGG ATCCT CAATTCTGCAAAACCGATGACACCAGCAGCTG-3’ were utilized as the forward and reverse primers, respectively. For the PhnG-H-I-J-K-his construct, 5’-AGCTCGTCGAT GCGACATATGCACGCA GATACCGCGACCCG CC-3’ and 5’-TCAAGTCTCGAGATTCTGCAAAACCGATGACACCAG CAGCTG-3’ were utilized as the forward and reverse primers, respectively. The polymerase chain reaction (PCR) was performed using Platinum Pfx DNA polymerase (Life Technology) and the following reaction conditions: 5 min at 95 °C, followed by 30 cycles of 30 s at 95 °C, 30 s at 66 °C, and 5 min at 72 °C. The his-PhnG-H-I-J-K PCR fragment was subsequently digested by the restriction enzymes Nde1 and BamH1 (New England Biolabs).

The digested DNA fragment was ligated to a pET-28b vector for expression with an N-terminus 6x-His-tag for purification. The PhnG-H-I-J-K-his PCR fragment was subsequently digested by the restriction enzymes Nde1 and Xho1 (New England Biolabs). The digested DNA fragment was ligated to a pET-24b vector for expression with a C-terminus 6x-His-tag for purification. The DNA for the expression of his-PhnG, his-PhnH, his-PhnG-I, and PhnG-I-his (containing only PhnG and PhnI) and PhnG-H-I-J-his (containing PhnG, PhnH, PhnI, and PhnJ) was cloned as previously described (45, 62).

2.2.3 Protein purification

The plasmids for the expression of the various complexes of C–P lyase were transformed into Rosetta2 (DE3) pLysS cells (Novagen) by electroporation and plated on LB agar. A single colony was used to inoculate 10 mL of LB medium containing 50 μg/mL kanamycin and allowed to grow overnight at 37 °C (250 rpm agitation). The 10 mL culture was subsequently used to inoculate 1 L of LB medium and incubated (250 rpm agitation) at 37 °C for 2–3 h until the OD600 reached ∼0.6. The temperature was then reduced to 18 °C and the culture induced with 0.5 mM isopropyl-β-thiogalactoside

(IPTG). The cells were harvested by centrifugation after 16–18 h of incubation and then stored at −80 °C. The frozen cells were thawed and then resuspended (4:1, v/w) in buffer A (50 mM HEPES, pH 8.5, 150 mM NaCl, and 20 mM imidazole) containing 0.10 mg/mL PMSF and a protease inhibitor cocktail. The cells were disrupted by sonication, the insoluble debris was removed by centrifugation at 4 °C, and the supernatant solution applied to a 5 mL HisTrap (GE Healthcare) column. The column was pre-equilibrated with 10 column volumes of buffer A, and the proteins were eluted with a gradient of buffer B (500 mM imidazole in buffer A). The fractions were pooled and applied to a High Load 26/60 Superdex 200 prep grade gel filtration column (GE Healthcare), which was previously equilibrated with buffer C (buffer A without imidazole). The fractions were pooled and analyzed by SDS-PAGE. Typical yields for the PhnG-H-I-J-K C–P lyase complexes were ∼30 mg/L of culture. Typical yields for the his-PhnG, his-PhnH, his-PhnG-I, PhnG-I-his, and PhnG-H-I-J-his fragments were ∼18 and ∼25 mg per liter of culture, respectively. The N-terminus His-tag was cleaved by Thrombin CleanCleave Kit (Sigma-Aldrich) following the manufacturer’s protocol.

2.2.4 N-terminus protein sequencing

The amino acid sequence and stoichiometry of the subunits of the two protein complexes were determined by amino terminus protein sequencing using a Procise 494A protein sequencer (Applied Biosystems). Seven cycles of automated Edman degradation of the PhnG-I-his complex was performed. The subunit stoichiometry was calculated by comparing the relative protein concentration of each subunit after each cycle of Edman degradation.

2.2.5 Analytical ultracentrifugation

The molecular mass and oligomeric state of three C–P lyase complexes (his-PhnG-I, his-PhnG-H-I-J, and his-PhnG-H-I-J-K) were characterized by sedimentation velocity using an Optima XL-A analytical ultracentrifuge (Beckman Instruments). All experiments and analyses were conducted at the University of Texas Health Science Center, San Antonio, TX. Samples of the C–P lyase complexes contained 10 mM Tris (pH 8.5), 150 mM NaCl, and 1.0 mM DTT. The protein concentrations ranged from 0.25 to 0.75 mg/mL, and the samples were centrifuged at 30000 rpm at 20 °C. The density and relative viscosity of the buffers were calculated as 1.004700 g/mL and 1.01418 mPa·s, respectively (14). On the basis of the sequence of the C–P lyase complexes, the partial specific volume (vbar) of the protein complexes were calculated as 0.7272 mL g^{-1}. The absorbance of the samples was measured at 231 nm with a 50 s delay.

2.2.6 ESI-MS measurements

ESI MS measurements were acquired with a ThermoScientific Exactive Plus EMR. In general, protein complexes were buffered with 50 mM ammonium acetate and loaded into an in-house pulled tip containing a platinum wire. All mass spectra were calibrated externally using a 1.0 µg/µL solution of cesium iodide. The protein complexes were buffered exchanged into 50 mM ammonium acetate (pH 6.8) prior to analysis using Bio-Rad Micro Bio-Spin columns. A potential of 1500–1800 V was applied to the spray tip and m/z was scanned from 900 to 16000. Important instrument parameters used were: FT resolution was set to 17500, S-Lens RF = 200, S-Lens = 42 V, skimmer = 15 V, ion injection time was limited to 200 ms, AGC target set to 1000000, and 40 microscans were acquired. Typical acquisition times were 4 min. Source CID and HCD voltage were varied to either limit fragmentation or to promote it.

The protein complex composition and stoichiometry were determined via mass differences observed in the intact MS analysis of the protein complex. Obtained data were deconvoluted with Thermo Scientific Protein Deconvolution software. Typical settings for the deconvolution were an m/z range of 2000–13000, output mass range 20000–300000, 75 ppm mass tolerance, charge state range of 4–50, four iterations, minimum adjacent charges 4–7, and a resolution of 12374 at 400m/z. Because of sample heterogeneity and the nature of the native spray analysis, it is estimated that all deconvoluted masses have an error of 400 ppm.

2.2.7 Chemical cross-linking

The chemical cross-linking agent BS3 was added to a solution of 20 µM his-PhnG-H-I-J-K (50 mM HEPES, pH 8.5 and 150 mM NaCl) to a final concentration of 2.5 mM and then incubated on ice for 2 h. The reaction was quenched by Laemmli sample buffer (2% SDS, 10% glycerol, 120 mM Tris-HCl, pH 6.8) and subsequently analyzed by SDS-PAGE. The gel was stained with 3 mg/mL Coomassie brilliant blue R250 and then destained with a solution containing 30% methanol and 10% acetic acid. The sample was proteolyzed via the addition of trypsin using a previously described protocol (105). The final samples were analyzed by LC-MS (Dionex nanoRSLC and an Orbitrap Fusion Tribrid mass spectrometer from Thermo Scientific). Briefly, 1.0 µL of the digested protein sample was loaded onto a 75 µm × 15 cm Acclaim Pepmap C18 column and eluted over 30 min with a total run time of 1 h. The eluted peptides were introduced into the mass spectrometer and analyzed using the Orbitrap (OT) mass spectrometer in positive ion mode with the following settings: resolution 120 K, scan range 400–1500 m/z, maximum injection time 50 ms, AGC target 2 × 104, and S-lens RF level = 60. Charge states from 2 to 7 were accepted, and targeting ions were excluded for 60 s. MS/MS data were obtained using the HCD cell and fragment masses determined using the OT mass spectrometer at a resolution of 30 K. The generated data were analyzed using the cross-linking identification software StavroX (version 3.4.5) (106).

For the analysis of peptide masses from cross-linking experiments, a maximum mass deviation of 7 ppm and a minimal signal-to-noise ratio (S/N) of three were accepted. The mass range was set as 200–5000 Da. Decoy searches were also conducted with the same parameters using the reverse protein sequences to determine the false positive rate. The results were evaluated by the StavroX software (106). For scores larger than 76, the false-positive rate was less than 5%. Candidate peptides with scores higher than 76 were selected and then manually checked before being assigned as positives. Additionally, carbamidomethylation of cysteine was taken into account as a fixed modification and oxidation of methionine was set as a varied modification. Up to three missed-cleavage sites were utilized during data processing.

2.2.8 Hydrogen/deuterium exchange

The his-PhnG, his-PhnH, and his-PhnG-I protein complexes were concentrated to 2.5 mg/mL and the his-PhnG-H-I-J-K complex was concentrated to 4.5 mg/mL. Deuterated buffers were prepared by removing water from the aqueous buffer containing 50 mM HEPES (pH 8.5), 200 mM NaCl, and 2.0 mM TCEP using a centrifugal vacuum concentrator and then adding an equivalent volume of 99.9% D_2O (Cambridge Isotope Laboratories, Inc.). A total of 4.0 µL of the concentrated protein sample was diluted into 36 µL of aqueous buffer for the undeuterated sample. The deuterated samples were obtained by diluting 4 µL of protein into 36 µL of the deuterated buffer (90% deuteration) for various amounts of time (4 and 10 s, 1.0, 2.0, 5.0, 15, and 30 min, and 1.0 h) at 0 °C. The deuterium exchange reaction was quenched with 60 µL of ice-cold buffer containing 3.0 M urea and 50 mM TCEP at pH 2.1 to bring the final pH of the resultant mixture to pH 2.3–2.4 (17, 18). A total of 50 µL of the quenched sample was added to 50 µL of immobilized pepsin resin (Thermo Scientific, USA). The pepsin beads were preactivated by washing with 1% (v/v) trifluoroacetic acid in water, pH 2.0 (107, 108).

The digestion was carried out at ~6 °C for 5 min with frequent vortexing. The solution was then transferred to a 0.22 µm Ultrafree-MC GV centrifugal filter (EMD

Millipore), and the pepsin beads were removed by centrifugation at 8000 rpm for 30 s. The supernatant solution was divided into aliquots and immediately flash-frozen in liquid nitrogen and stored at -80 °C.

The subsequent analysis by mass spectrometry was performed using a Thermo Scientific Q Exactive instrument within two to three days of sample preparation. The samples were thawed on ice prior to injection to minimize back-exchange. The peptic digests of his-PhnG, his-PhnH, his-PhnG-I, and his-PhnG-H-I-J-K were applied to a 2.1 mm × 100 mm Symmetry C18 reverse-phased column (Waters Corporation, Milford, MA) using 0.1% (v/v) formic acid in water at a flow rate of 200 μL/min. The peptides were eluted using a gradient of 8–55% acetonitrile over the course of 8 min. The solvents and column were placed in an ice-bath throughout the experiment to reduce the back-exchange of deuterons during analysis. A blank solution using acetonitrile was utilized between each sample to avoid peak carry-over between runs. The mass spectrometer was calibrated using an LTQ ESI positive ion calibration solution from Thermo Scientific, USA. The peptides were identified using the software Peptide Mapping (Thermo Scientific) with a mass tolerance of 10 ppm for the theoretical mass-to-charge (m/z) ratios. Oxidation of methionine was set as a possible modification. The peptides were further confirmed by LC-MS/MS analysis on an Orbitrap Fusion Tribrid mass spectrometer from Thermo Scientific as described earlier. The pepsin digestion provided sequence coverages of 88%, 83%, and 86% for the PhnG, PhnH, and PhnG-I complexes, respectively. The PhnG-H-I-J-K complex had a sequence coverage of 89% for PhnG and PhnI and 83% for PhnH.

The deuterium content and centroid values of each peptide at various time points were calculated using the software HDExaminer (Sierra Analytics, Modesto, CA). Each identified peptide was manually examined and validated. It was assumed that the maximum number of exchangeable amides in a peptide is two less than the total number of amino acids in the peptide minus any proline residues (107-109). The shifts in the centroid values for the deuterated samples were determined using the undeuterated sample as a control. The various charge states of an individual peptide were used to calculate the deuterium level, and the values between them did not show much change. For regions of the proteins containing multiple overlapping peptides, a maximum of three overlaps encompassing the wider region were used in the analysis. The peptide exchange plots were generated using MS-Excel. The deuterated samples were not corrected for back-exchange, and all of the reactions were carried out under identical conditions.

2.2.9 Homology modeling and secondary structure predictions

The homology model for PhnK was created using the Intensive Mode of the web-based structure prediction software Phyre2 (110) with the nucleotide-binding domain of a dipeptide ABC transporter (PDB id: 4FWI), which has a 32% sequence identity with PhnK. For PhnG, PhnI, and PhnJ, the homology modeling was conducted using Swiss Model (111) and the secondary structure prediction was carried out using SABLE protein structure prediction server (111). All protein structure figures were generated using Pymol (http://www.pymol.org).

2.2.10 Electron microscopy

$\text{PhnG}_2\text{H}_2\text{I}_2\text{J}_2\text{K}$ was diluted to 0.1 mg/mL with dilution buffer (50 mM HEPES, pH=8.5, 150 mM NaCl, 2 mM TCEP). 3 μL of the sample was applied to a C-Flat 1.2/1.3 holey carbon grid at 16 °C with 100% relative humidity and vitrified using a Vitrobot Mark III (FEI). The thin-ice areas that were expected to show clearly visible and mono-dispersed particles were imaged under an FEI Tecnai F20 cryo-electron microscope with a field emission gun (FEI) operated at 200 kV. Data were recorded on a Gatan K2 Summit electron-counting direct detection camera (Gatan) in electron-counting mode (112). A nominal magnification of 25,000× was used, yielding a pixel size of 1.5 Å. The beam intensity was adjusted to a dose rate of ∼10 electrons per pixel per second on the camera. A 25-frame or 50-frame movie stack was recorded for each picture, with 0.2 s per frame for a total exposure time of 5 s or 10 s, respectively.

2.2.11 Image processing

Up to 967 collected image stacks were iteratively aligned and filtered based on the electron dose using Unblur (113). A total number of 694 sum images were selected according to the visibility of particles and power spectra. For particle picking, the sum images were first low-pass filtered to 15 Å to visualize particles unambiguously. The Erase tool of e2boxer.py in EMAN2 (114) was used to erase bad areas that had junk or no particles. Finally, the Swarm tool of e2boxer.py was used to semi-automatically pick 492,889 particles using a box size of 120 pixels, during which the threshold for picking was interactively changed so that false-positive picking was minimized. The coordinates were imported into Relion (115) to extract the particles from the original sum images with a downscaling factor of 1, 2, or 4. Reference-free 2D classification was performed in Relion to remove bad particles that cannot average well with each other, which generated a dataset of 294,203 particles. We selected 3,956 particles with good contrast to generate an initial model using the PRIME routine in the SIMPLE package (116), which was low-pass filtered to 60 Å and used for the first round of unsupervised 3D classification in Relion.

The 3D classification and refinement was performed hierarchically. Briefly, in the first two rounds of 3D classification, particles were separated into three classes based on quality and the presence or absence of PhnK. Up to 50,061 good particles from one class with PhnK were selected and further refined with C1 symmetry to produce the density map of $PhnG_2H_2I_2J_2K$. A total number of 86,391 particles without PhnK were subjected to an additional round of 3D classification to separate them into three classes, which yielded 23,861 good particles in one class. These 23,861 particles were further refined with C2 symmetry to produce the density map of $PhnG_2H_2I_2J_2$. The particle images were downscaled by four times in the first two rounds of the 3D classification and downscaled by two times in the later 3D classification and refinement. Refinement against the particles without downscaling, did not improve the final resolution of the maps.

Blocres of the Bsoft package (117) was used to estimate the local resolution of the cryo-EM maps. Our resolution is limited to 7.8 Å, which may be due to the flexible nature of the specimen.

2.2.12 Molecular modeling and map segmentation

Fitting of models into the density map was done in UCSF Chimera (118). The correct handedness of the density maps of Phn$G_2H_2I_2J_2$ and Phn$G_2H_2I_2J_2$K were initially determined by docking the crystal structure of PhnH (PDB id : 2FSU) (61) into the original maps and the maps with the flipped handedness. The crystal structure of PhnH fitted into the flipped maps with all the secondary structures matching, confirming that the original maps should have their handedness flipped. A homology model of PhnK was built using SWISS Model (119) based on an nucleotide binding domain (PDB id : 4U00) (120). The location of PhnK was determined by the difference map between Phn$G_2H_2I_2J_2$K and Phn$G_2H_2I_2J_2$. Moreover, the homology model of PhnK fitted the difference map with all of the secondary structures in place, confirming the correct location of PhnK in Phn$G_2H_2I_2J_2$K. Segmentation of the Phn$G_2H_2I_2J_2$ core was done using Segger (121) in UCSF Chimera based on the crystal structure of Phn$G_2H_2I_2J_2$ (PDB id: 4XB6). The electrostatic surface of the models was calculated using the APBS module (119)in UCSF Chimera.

2.3 Results

2.3.1 Characterization of PhnG-I

Two different contructions were prepared for the expression of a complex that contained only PhnG and PhnI. The first of these had a polyhistidine purification tag at the N-terminus of PhnG, and the second had the polyhistidine tag at the C-terminus of PhnI. The calculated molecular masses for the subunits within these two complexes are as follows: PhnG (16,524 Da), his-PhnG (18,688 Da), PhnI (38,852 Da), and PhnI-his (39,917 Da). Both of these complexes migrated as a single peak upon elution from a calibrated Superdex 200 10/300 GL gel filtration column (GE Heathcare) (data not shown), and the approximate molecular mass was determined to be 140 ± 26 kDa.

The molecular mass for the his-PhnG-I complex was also determined using analytical ultracentrifugation and ESI mass spectrometry. For the his-PhnG-I complex, the sedimentation coefficient (S_w=20) of 5.89 S corresponds to an apparent molecular mass of 115.8 ± 1.7 kDa at a protein concentration of 0.25 mg/mL. The sedimentation velocity experiment was also conducted at a protein concentration of 0.75 mg/mL and a van Holde–Weischet analysis (122) indicated that the oligomeric state of his-PhnG-I did not change at the higher protein concentration. Under native ESI conditions, the his-PhnG-I complex exhibited charge states in the gas phase with a range of +21 to +25 and a mass of 115, 300 Da (Figure 5). CID fragmentation of the his-PhnG-I complex resulted in the loss of a PhnG monomer and a mass corresponding to PhnGI2 (96,600

Da), as well as the appearance of a mass at 56,000 Da that corresponds to PhnGI, indicating that PhnG and PhnI form a heterodimer (Figure 6).

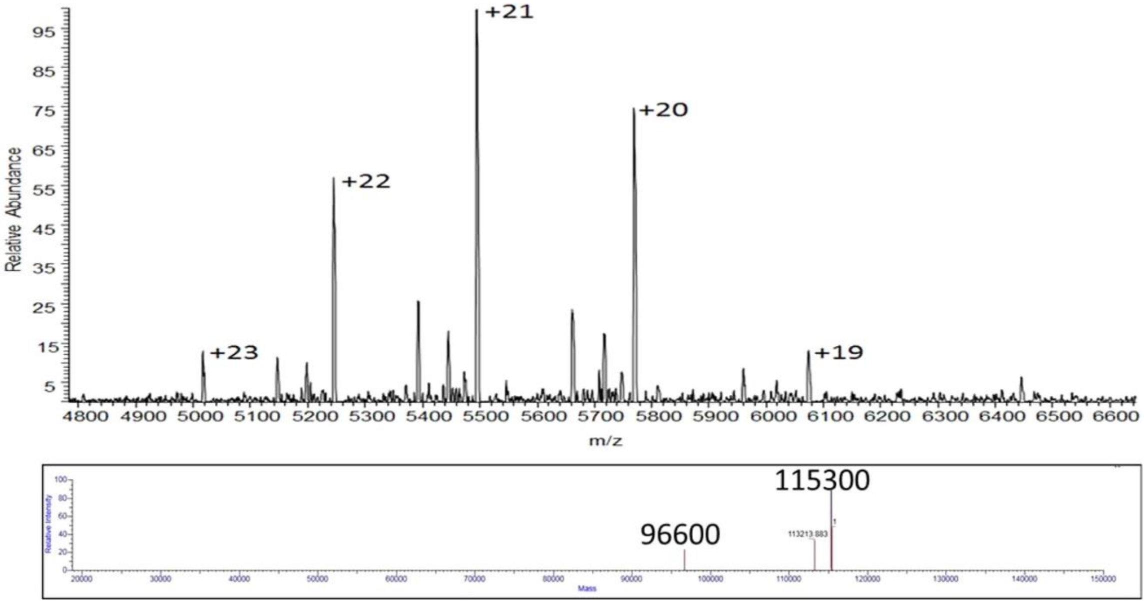


Figure 5: Native ESI of the $PhnG_2I_2$ complex. The upper figure is the raw data obtained using 75 V of energy in the HCD cell and 50 V of energy in the source region. The bottom figure is the deconvoluted data using the isotope clusters from the upper mass spectrum of the complex, which represents the +19 to +23 charge states of the complex.

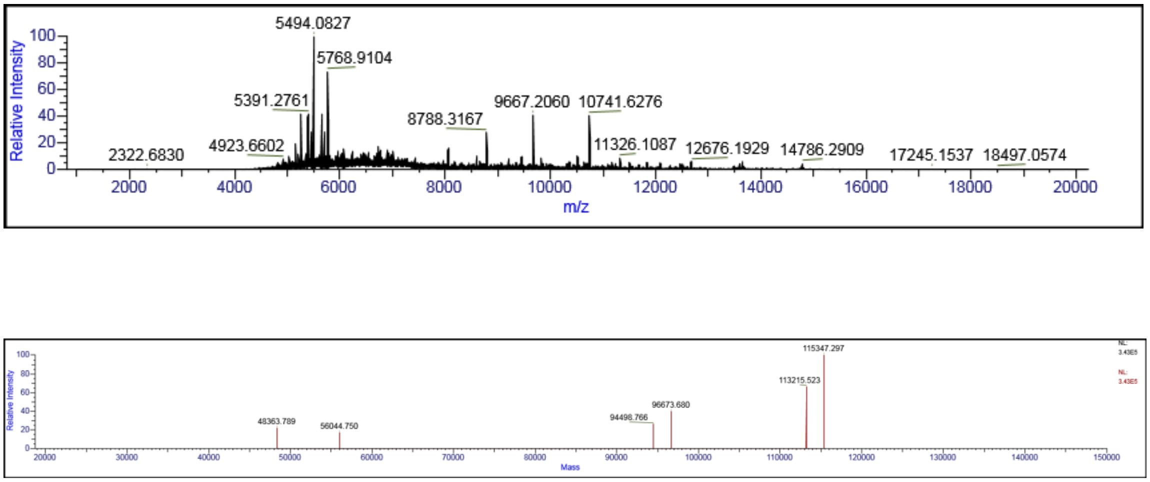


Figure 6: Fragmentation of the PhnG$_2$I$_2$ complex. Upper figure is the raw data obtain using 75 V of energy in the HCD cell and 75V of energy in the source region. The bottom figure is the deconvoluted data.

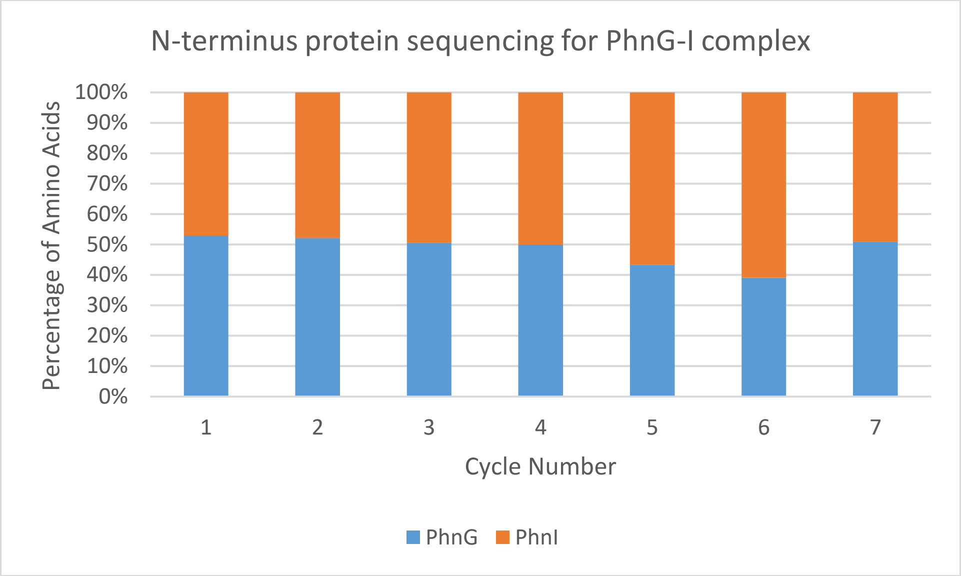


Figure 7: The ratio between the amino acid of PhnG and PhnI in seven N-terminus protein sequencing cycles.

Table 1: The ratio between the amino acid of PhnG and PhnI in seven N-terminus protein sequencing cycles.

Cycle number	1	2	3	4	5	6	7
PhnG amino acids	M	H	A	D	T	A	T
PhnI amino acids	M	Y	V	A	V	K	G
Ratio	-	1.09	1.02	1.00	0.87	0.84	1.04
Average ratio	0.98						

Stoichiometry of the PhnG-I-his complex was addressed by N-terminus amino acid sequence analysis. After seven rounds of N-terminus amino acid sequencing, the ratio of the amino acids specific for PhnG to that for PhnI averaged 0.9 ± 0.1 (Figure 7 and table 1). The data from the first round of sequencing was dismissed due to the overlap methionine of these two proteins. Therefore, the apparent stoichiometry of the two subunits contained within the PhnG-I-his complex is ∼1:1. Given the mass of the his-PhnG-I complex determined from analytical ultracentrifugation (115,800 Da) and ESI mass spectrometry (115,770 Da), it is concluded that the stoichiometry of the complex containing only PhnG and PhnI is $PhnG_2I_2$.

2.3.2 SDS electrophoresis of $PhnG_2I_2$

The subunit composition of the $PhnG_2I_2$ complex was confirmed by SDS-PAGE, and the identity of each band was verified by in-gel digestion with trypsin and mass spectroscopy (Figure 8). The intensity of the Coomassie Blue staining for each of the two separated subunits of $PhnG_2I_2$ was measured using a Gel Doc EZ System densitometer (BioRad). For the $PhnG_2I_2$-his complex, the relative staining intensity for subunits PhnG and PhnI-his was 1.00:1.22 (56.3:68.6) and 1.00:1.07 (56.4:60.3) for his-$PhnG_2I_2$ (Table 2). When the his-tag was removed by proteolysis with thrombin, the relative staining intensity for these two subunits was similar (1.00:1.17). Because the molecular mass ratios for the two subunits are 1.00:2.42 (for PhnG and PhnI-his) and 1.00:2.09 (for his-PhnG and PhnI), PhnG stains approximately twice as dense per amino acid as does PhnI under these conditions.

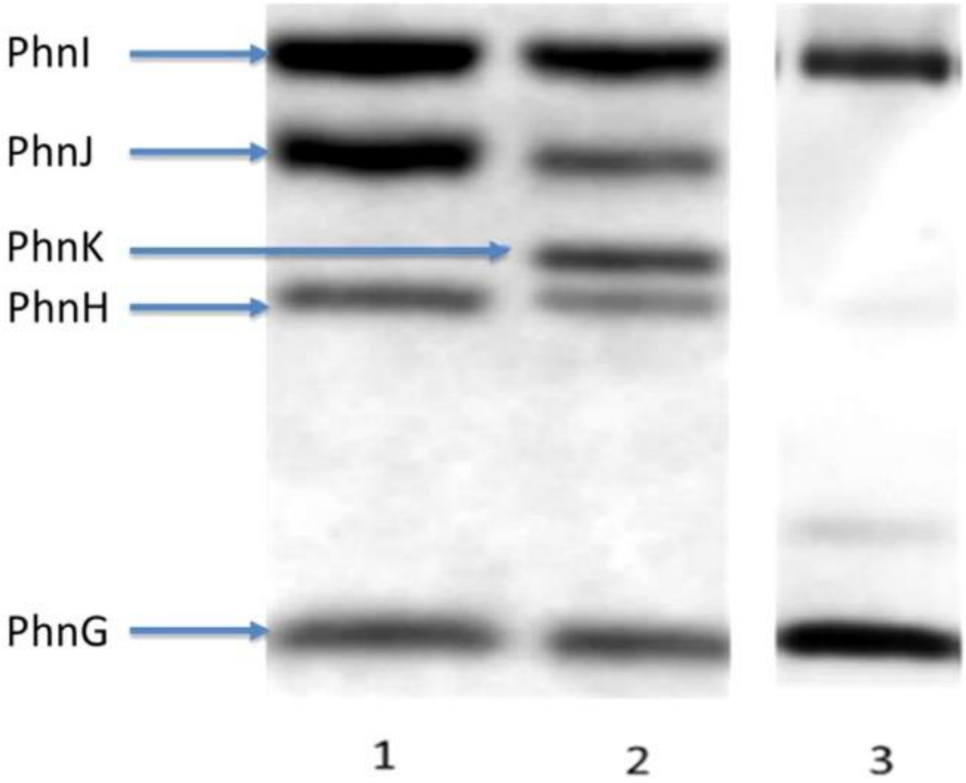


Figure 8. SDS-PAGE separation of the subunits contained within the C–P lyase complexes from *E. coli*. Lane 1: his-PhnG$_2$I$_2$. Lane 2: his- PhnG$_2$H$_2$I$_2$J$_2$. Lane 3: his-PhnG$_2$H$_2$I$_2$J$_2$K. All protein complexes were diluted to 2 mg/mL in sample buffer (65.8 mM Tris-HCl, pH 6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue) and loaded onto a Bio-Rad mini-protean precast gel. The gel was stained by Coomassie Brilliant Blue and the relative intensity of each band quantified using a Gel Doc EZ System densitometer (BioRad).

2.3.3 Characterization of PhnG-H-I-J

The complex that contains subunits PhnG, PhnH, PhnI, and PhnJ was cloned with a polyhistidine tag at the C-terminus of PhnJ. This complex migrated as a single homogeneous peak after elution from a gel filtration column (data not shown). The approximate molecular mass for this complex was determined to be 243 ± 38 kDa using a calibrated Superdex 200 10/300 GL gel filtration column (GE Heathcare). Two species were identified by ultracentrifugation at a protein concentration of 0.25 mg/mL. The major component (79.9%) exhibited a S_w20 value of 9.43 S and an estimated molecular mass of 218.8 ± 7.4 kDa. The minor component (19.1%) had an S_w20 value of 7.90 S with an estimated molecular mass of 191.0 ± 9.1 kDa. The sedimentation velocity experiments were also conducted at a higher protein concentration (0.75 mg/mL), and these experiments indicated that the PhnG-H-I-J-his complex aggregated at the higher protein concentration during the ultracentrifugation process.

The subunit composition of the PhnG-H-I-J-his complex was confirmed by SDS-PAGE, and the identity of each band was verified by in-gel digestion with trypsin and mass spectroscopy (Figure 7). The intensity of the Coomassie Blue staining for each of the separated subunits was measured, and the relative staining intensities for subunits PhnG, PhnH, PhnI, and PhnJ-his are listed in Table 2. From the relative staining intensities for PhnG and PhnI in this complex (1.19–1.28), we can conclude that these two subunits are in a ratio of 1:1 because this is essentially the same ratio as found in the $PhnG_2I_2$ complex (1.17).

Table 2. Relative staining intensities in PhnG-I, PhnG-H-I-J and PhnG-H-I-J-K

	PhnG	**PhnI**
PhnG- I	56.0	65.5
PhnG-I-*his*	56.3	68.6
Phn his-G-I	56.4	60.3
Average	56.2	64.8
Ratio	1.00	1.15

	PhnG	**PhnH**	**PhnI**	**PhnJ**	**G:I**
PhnG-H-I-J	27.6	11.2	35.4	25.9	1:1.28
PhnG-H-I-J-*his*	26.0	13.5	31.0	29.5	1:1.19
PhnG-H-I-J-K	28.4	11.8	35.8	27.0	1:1.26
PhnG-H-I-J-K-*his*	29.7	11.3	38.0	27.9	1:1.28
his-PhnG-H-I-J-K	28.5	9.3	36.5	24.8	1:1.28

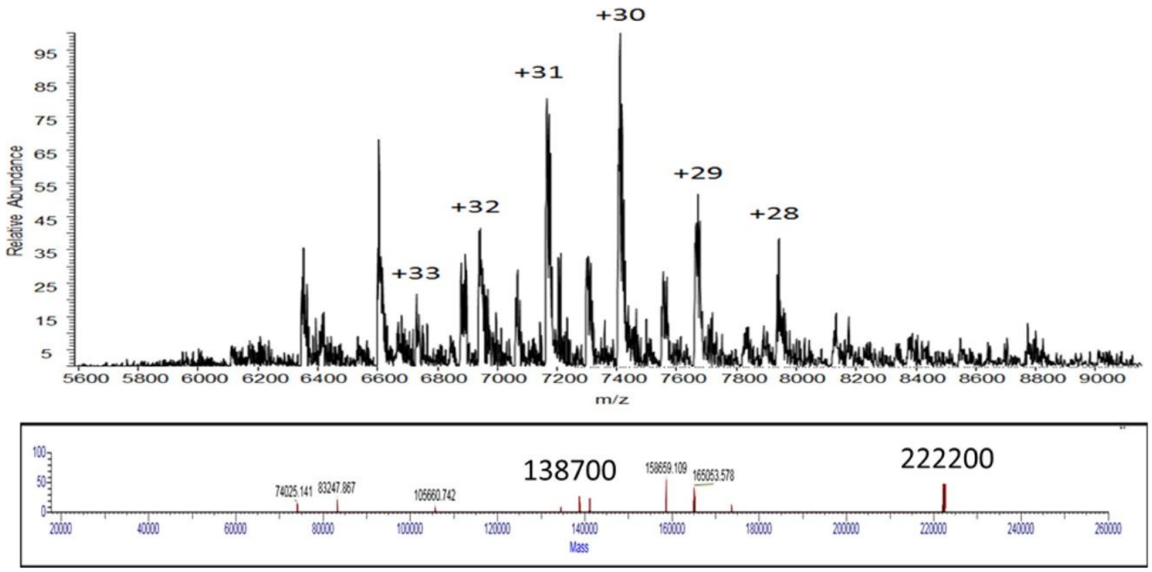


Figure 9: Native ESI of the PhnG$_2$H$_2$I$_2$J$_2$ complex. The upper figure is the raw data obtained using 10 V of energy in the HCD cell and 10 V of energy in the source region. The bottom figure is the deconvoluted data utilizing the isotope clusters from the upper mass spectrum of the complex, which represents the +28 to +33 charge states of the complex.

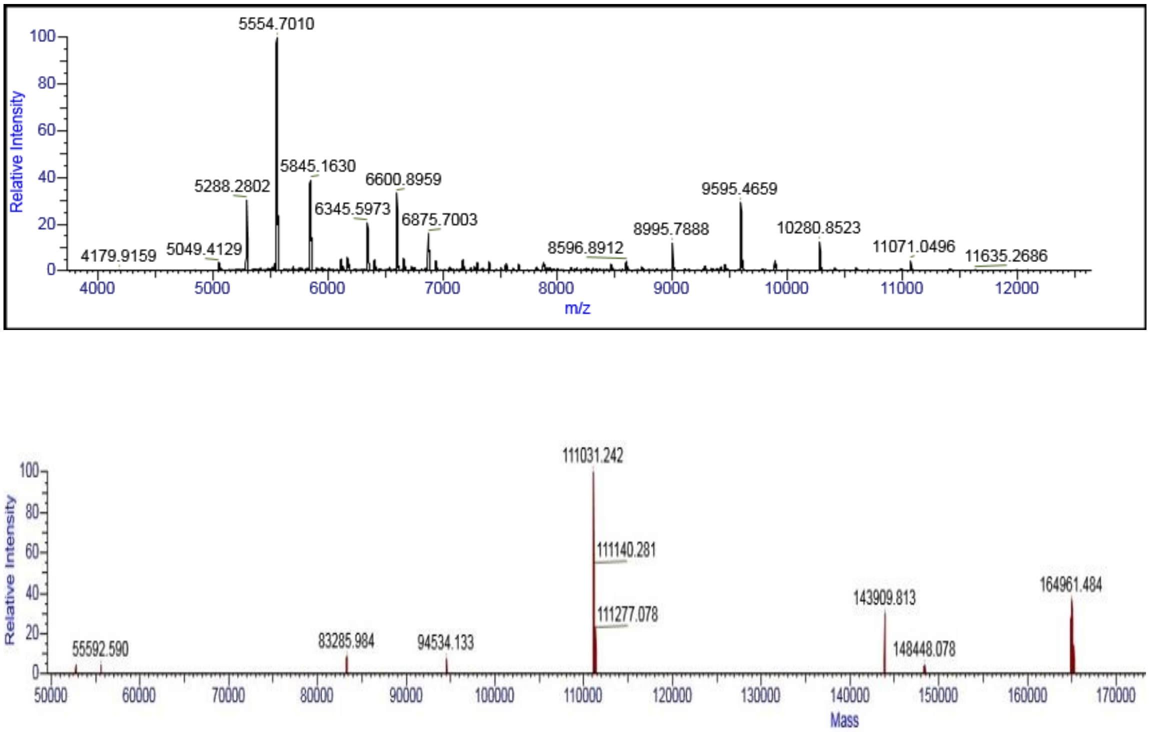


Figure 10: Fragmentation of the PhnG$_2$H$_2$I$_2$J$_2$ complex. Upper figure is the raw data obtain using 75V of energy in the HCD cell and 75V of energy in the source region. The bottom figure is the deconvoluted data.

Under native ESI conditions, the parent mass of the PhnG-H-I-J-his complex was measured to be approximately 222,200 Da (Figure 9). In an effort to elucidate the subunit composition of this complex, the collisional energy was raised as high as 200 V and the resulting deconvoluted fragment masses of 197,900, 166,300, 144,000, 111,100, 94,500, and 55,600 were obtained (Figure 10). These fragments correspond to the sequential loss of PhnH (21.0 kDa), PhnJ-his (32.8 kDa), PhnH (21.0 kDa), PhnJ-his (31.7 kDa), PhnG (16.5 kDa), and PhnI (38.9 kDa). PhnGI is the smallest fragment that was observed. These results are consistent with the formation of PhnHJ and PhnGI heterodimers in the complex and that the overall composition for this complex is $PhnG_2H_2I_2J_2$.

2.3.4 Characterization of PhnG-H-I-J-K

The complex composed of PhnG through PhnK was cloned and subsequently expressed with a polyhistidine tag at either the N-terminus of PhnG or the C-terminus of PhnK. Both of these complexes migrated through a calibrated gel filtration column as a single peak (data not shown) with an estimated molecular mass of 285 ± 46 kDa. The subunit compositions of the two PhnG-H-I-J-K complexes were confirmed by SDS-PAGE, and the identity of each band was verified by in-gel digestion with trypsin and mass spectroscopy (Figure 7). To estimate the stoichiometry of the subunits contained within these two complexes, the relative Coomassie Blue staining intensities for PhnG, PhnH, PhnI, and PhnJ were compared to the relative staining intensities for these same subunits in the complex that contains only four subunits (Table 1). Within experimental error, the relative staining intensities for PhnG, PhnH, PhnI, and PhnJ contained within the $PhnG_2H_2I_2J_2$ complexes matched those contained within either of the complexes that contained all five subunits. Therefore, the $PhnG_2H_2I_2J_2$ core is retained within the two complexes that now contain PhnK.

The molecular mass of the PhnG-H-I-J-K complex was estimated using analytical ultracentrifugation. At a protein concentration of 0.25 mg/mL, the Sw20 value was determined to be 8.94 S for the PhnG-H-I-J-K-his complex with a calculated molecular mass of 236.9 ± 8.3 kDa. The sedimentation velocity experiments were also conducted at a protein concentration of 0.75 mg/mL and a van Holde–Weischet analysis indicated that the oligomeric state of PhnG-H-I-J-K-his was constant at the two protein concentrations. The mass of the PhnG-H-I-J-K-his complex was also determined using ESI mass spectrometry. Under native ESI conditions, the PhnG-H-I-J-K-his complex enters the gas phase with charge states of +30 to +36, corresponding to a mass of 249,700 (Figure 11), a value that is consistent with that determined by analytical ultracentrifugation. Using 75 V of source fragmentation energy and 45 V in the HCD cell, the mass at 249,200 most likely represents the $PhnG_2H_2I_2J_2K$ -his complex (calculated mass = 245,144) and that at 221,300 represents the $PhnG_2H_2I_2J_2$ complex (calculated mass = 216,234) and the loss of PhnK-his (calculated mass = 28,910). The mass at 196,234 represents the loss of PhnHJ (calculated mass = 52,741) to give $PhnG_2HI_2JK$ (calculated mass = 192,403). The mass at 168,598 represents the $PhnG_2HI_2J$ complex with the loss of PhnK (Figure 12).

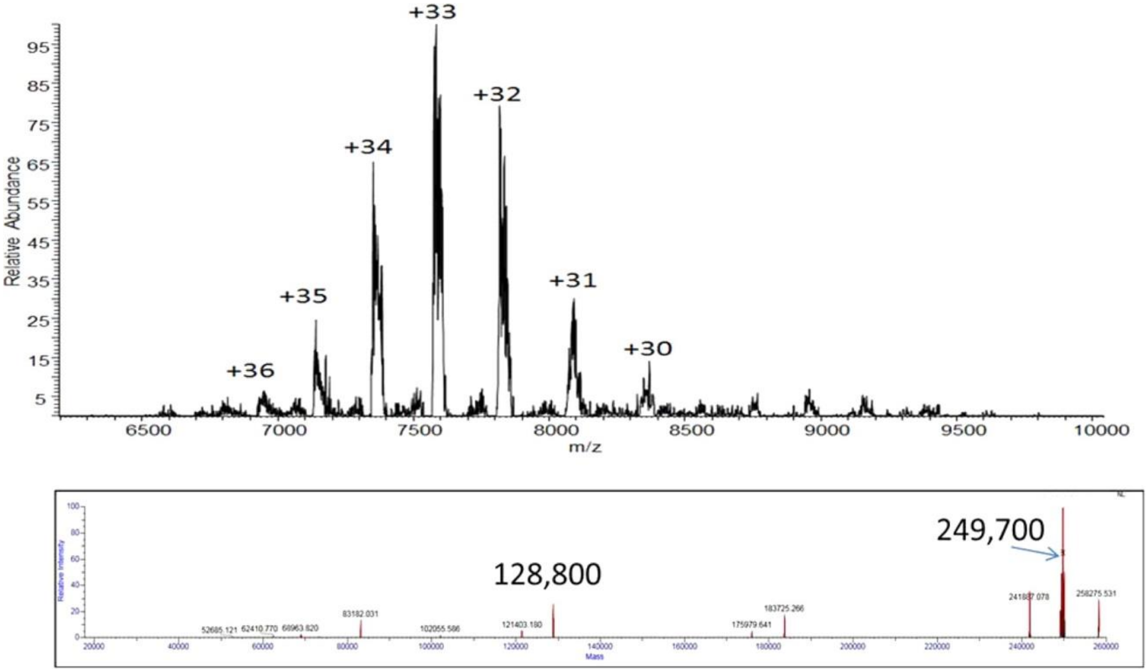


Figure 11: Native ESI of the $PhnG_2H_2I_2J_2K$ complex. The upper figure is the raw data obtained using 13 V of collision energy in the HCD cell and 10 V of energy in the source region. The bottom figure is the deconvoluted data utilizing the isotope clusters from the upper mass spectrum of the complex, which represents the +30 to +36 charge states of the complex.

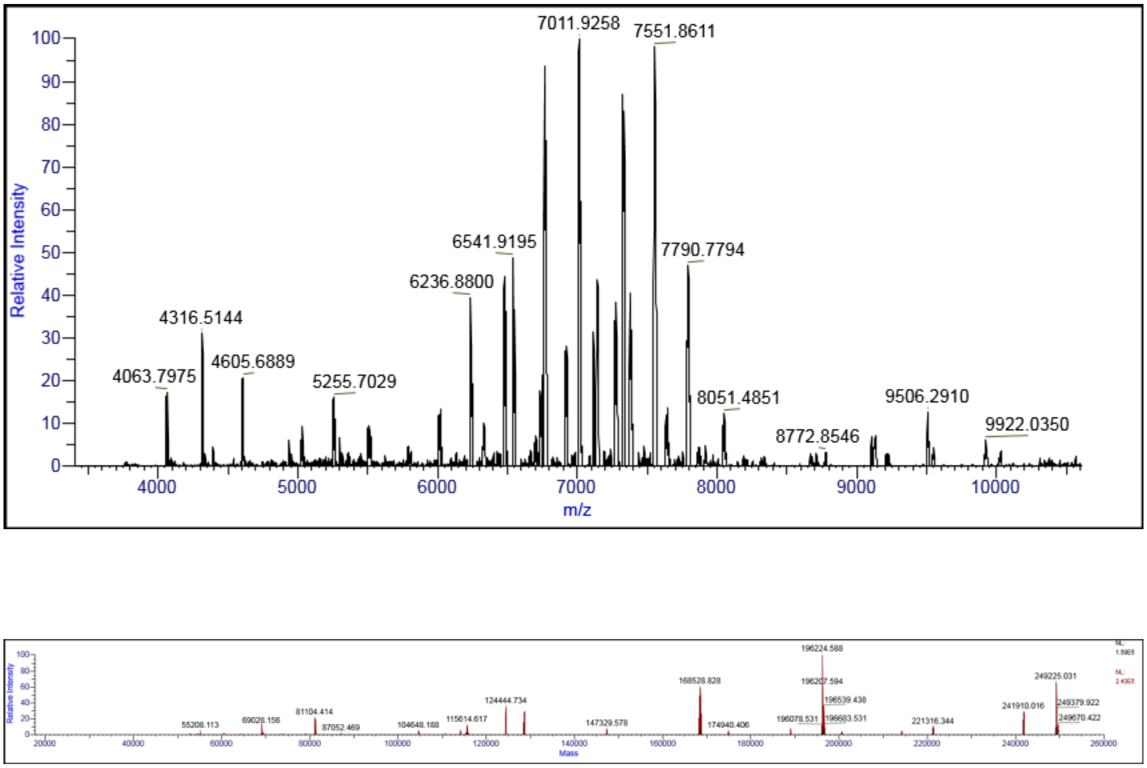


Figure 12: Fragmentation of the PhnG$_2$H$_2$I$_2$J$_2$K complex. Upper figure is the raw data obtain using 75V of energy in the HCD cell and 45V of energy in the source region. The bottom figure is the deconvoluted data.

2.3.5 Structural analysis by chemical cross-linking

The contact regions for interaction of the subunits contained within the $PhnG_2H_2I_2J_2K$ complex were probed by chemical cross-linking. Cross-linking was initiated with the homobifunctional reagent bis-sulfosuccinimidyl suberate (BS3) that can form a linkage of up to 11.4 Å between the ε-amino groups of lysine residues or the free amino group of the N-terminus (123). The products of the cross-linking reactions with $PhnG_2H_2I_2J_2K$ -his were separated by SDS-PAGE, and the results are presented in Figure 13. Four prominent high molecular mass cross-linked protein complexes of approximately 47 (band 1), 70 (band 2), 87 (band 3), and 110 kDa (band 4) were formed after incubation with BS3. Two bands of higher molecular mass at approximately 125 and 165 kDa were also formed but are less distinct.

The cross-linked proteins were excised from the gel and subjected to in-gel digestion with trypsin and characterized by LC-ESI/MS. On the basis of the mass spectrometry data, the 47 kDa band contains one copy each of PhnG (17 kDa) and PhnJ (31 kDa), the 70 kDa band contains one copy each of PhnI (39 kDa) and PhnJ, and the 87 kDa band contains one copy each of PhnG, PhnI, and PhnJ. The bands at 110, 125, and 165 kDa contain mixtures of PhnG, PhnI, and PhnJ. No cross-linked proteins were identified to contain PhnH or PhnK.

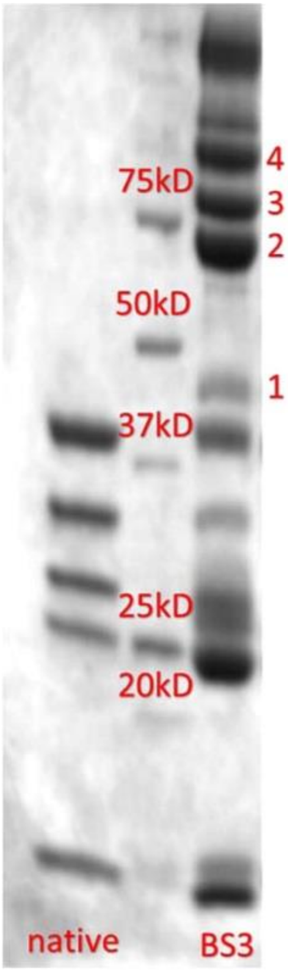


Figure 13: SDS-PAGE of the BS3-initiated cross-linking of those proteins contained within the various carbon–phosphorus lyase complexes. The left lane represents the separation of the subunits contained within the PhnG$_2$H$_2$I$_2$J$_2$K complex prior to the addition of BS3. The middle lane is a standard molecular mass protein ladder (Bio-Rad). The rightmost lane represents the separation of cross-linked proteins after the addition of BS3 to the PhnG$_2$H$_2$I$_2$J$_2$K complex. Four prominent cross-linked bands were labeled as 1, 2, 3, and 4 from the bottom to the top.

Table 3: Identification of Specific Crosslinking Site of PhnG-H-I-J-K C-P Lyase

	M+H+	Dev. (ppm)	Sequence 1	Protein 1	Sequence 2	Protein 2	Site P1	Site P2
Band1	3010.457	-2.65	[DKQHAER]	PhnG	{ANLSGYNFAYLDEQTKR]	PhnJ	K2	{0
Band2	2501.295	2.34	[NFFKR]	PhnJ	[GGEKAIDAAHALQESR]	PhnI	K4	K4
	3214.425	-1.95	{MYVAVKGGEK]	PhnI	[QQSEAKNQ}	PhnJ	{0	K17
	2857.428	-3.37	{mYVAVK]	PhnI	{ANLSGYNFAYLDEQTKR]	PhnJ	K6	{0
	1930.028	1.53	[NFFKR]	PhnJ	{MYVAVKGGEK]	PhnI	K4	K6
Band3	2501.291	0.65	[NFFKR]	PhnJ	[GGEKAIDAAHALQESR]	PhnI	K4	K4
	3010.47	1.68	[DKQHAER]	PhnG	{ANLSGYNFAYLDEQTKR]	PhnJ	K2	{0
	3137.47	-3.92	{MMHADTATR]	PhnG	[MDMmPALQLFGAGREKR]	PhnJ	{0	K16
	2167.068	-0.97	[QQSEAKNQ}	PhnJ	{mYVAVKGGEK]	PhnI	K6	{0
Band4	3010.47	1.68	[DKQHAER]	PhnG	{ANLSGYNFAYLDEQTKR]	PhnJ	K2	{0
	7691.662	-6.61	{mMHADTATRQHWMSVLAHSQPAELAAR]	PhnG	[EMPMPYGWGTGGIQLTASVIGESDVLKVIDQGADDTTNAVSIR]	PhnJ	{0	K27
	2501.280	0.65	[NFFKR]	PhnJ	[GGEKAIDAAHALQESR]	PhnI	K4	K4
	1930.021	-2.17	[NFFKR]	PhnJ	{MYVAVKGGEK]	PhnI	K4	K6

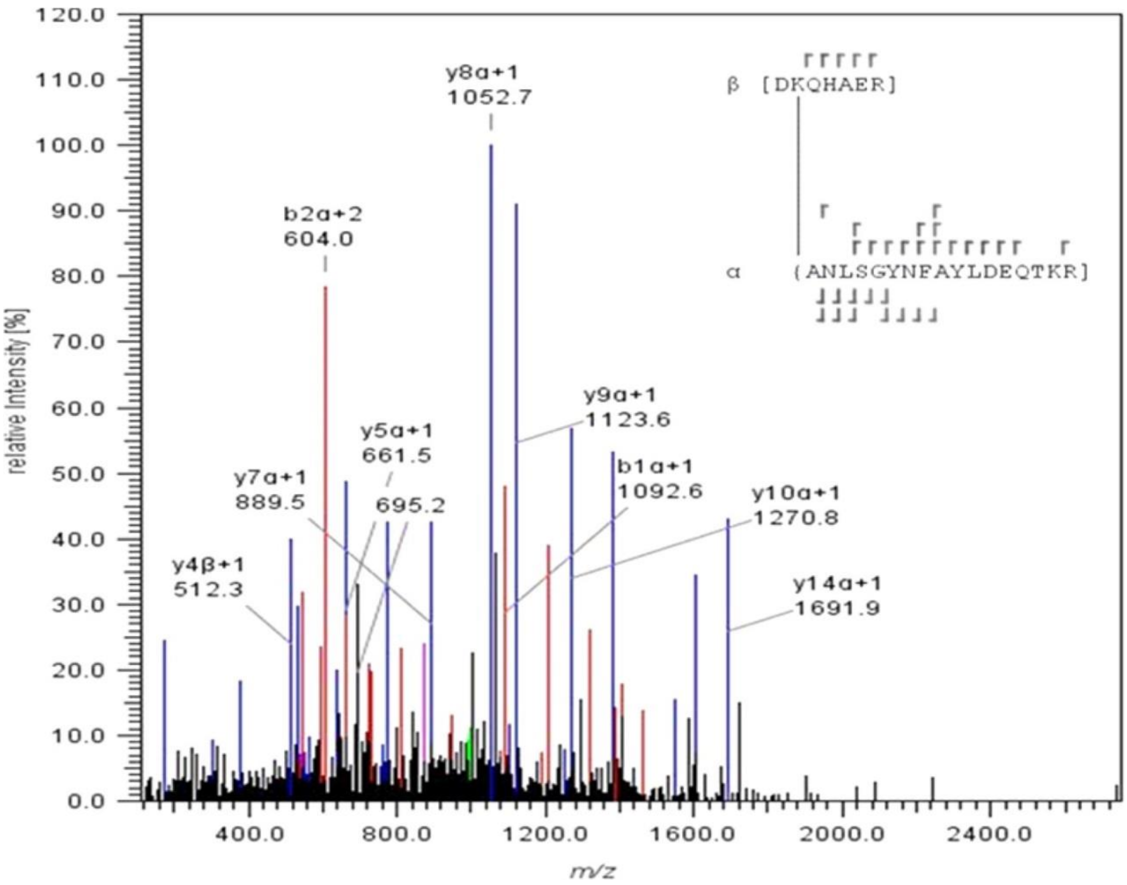


Figure 14: ESI-MS/MS analysis of the cross-linked peptides contained within the $PhnG_2H_2I_2J_2K$ complex. The peptide ^{88}DKHAER^{94} from PhnG and the peptide ^{2}ANLSGNYFAYLDEQTKR^{18} were cross-linked through Lys-89 from DKHAER and the N-terminus of ANLSGNYFAYLDEQTKR.

The cross-linked proteins were excised, proteolyzed with trypsin, and then analyzed by LC-ESI-MS/MS to identify the specific cross-linking sites. The purified peptides were subjected to high-energy collision dissociation (HCD). A representative example of an MS/MS spectrum of a triply charged ion at m/z 1004.16 (corresponding to an M + H+ = 3010.46) is provided in Figure 14. The daughter ion spectrum exhibited most of the y-ion series for two peptides from PhnG (β-peptide, residues 89–95) and PhnJ (α-peptide, residues 2–18). The y16β and y5α ions indicate that Lys-89 of PhnG and the N-terminus of PhnJ are covalently linked to one another. A similar analysis was conducted to identify all of the other cross-linked peptides, and the identity of the cross-linked peptide pairs are presented in Table 3.

In addition to the intermolecular cross-links between different protein subunits, there were several intramolecular cross-links contained within PhnI and PhnJ. In all of the cross-linking experiments, only peptides from PhnG, PhnI, or PhnJ were identified. There was no trace of PhnH or PhnK in the cross-linked proteins. However, further analysis of the ESI-MS data identified several internally cross-linked peptides and individual lysine residues within PhnH and PhnK that were labeled with the cross-linking reagent. For example, Lys-38 from PhnH was derivatized with BS3. In PhnK, the N-terminus, Lys-127, and Lys-220 were labeled with BS3 and an internal cross-link was identified between Lys-19 and Lys-49.

2.3.6 Structural analysis by hydrogen/deuterium exchange

Hydrogen/deuterium exchange experiments were carried out on PhnG, $PhnG_2I_2$, and $PhnG_2H_2I_2J_2K$ from the C–P lyase complex in an attempt to determine those regions of PhnG that are most affected by the association with PhnI. It was assumed that those portions of PhnG that interact directly with PhnI will exhibit a change in the rate of H/D exchange in either $PhnG_2I_2$ or $PhnG_2H_2I_2J_2K$ relative to the H/D exchange profile for PhnG alone. Reduced rates of H/D exchange may also be observed for those portions of PhnG that become more structured in either of these two complexes than in the isolated subunit alone. A difference of 10% or more in the exchange rate between any two protein complexes is considered significant. The overall coverage of peptides from PhnG that were detected by mass spectrometry in all three complexes was approximately 87%.

The time course for the exchange of deuterium from solvent incorporated into the peptide ^{68}TRAAVRLTDGTLGYS^{82} from PhnG is presented in Figure 15. The time courses for H/D exchange within PhnG alone indicate that many regions of the protein undergo H/D exchange very rapidly. This is especially true for those regions of PhnG that extend from residues 28–79 and from residues 116–124 (Figure 16). All of the peptides identified for PhnG, $PhnG_2I_2$, and $PhnG_2H_2I_2J_2K$ exhibit EX2 kinetics for deuterium uptake (124-126). Bimodal isotopic distribution spectra, which indicate the presence of local or global unfolding through EX1 kinetics, were not observed for any of the peptides of PhnG or PhnI (58, 59).

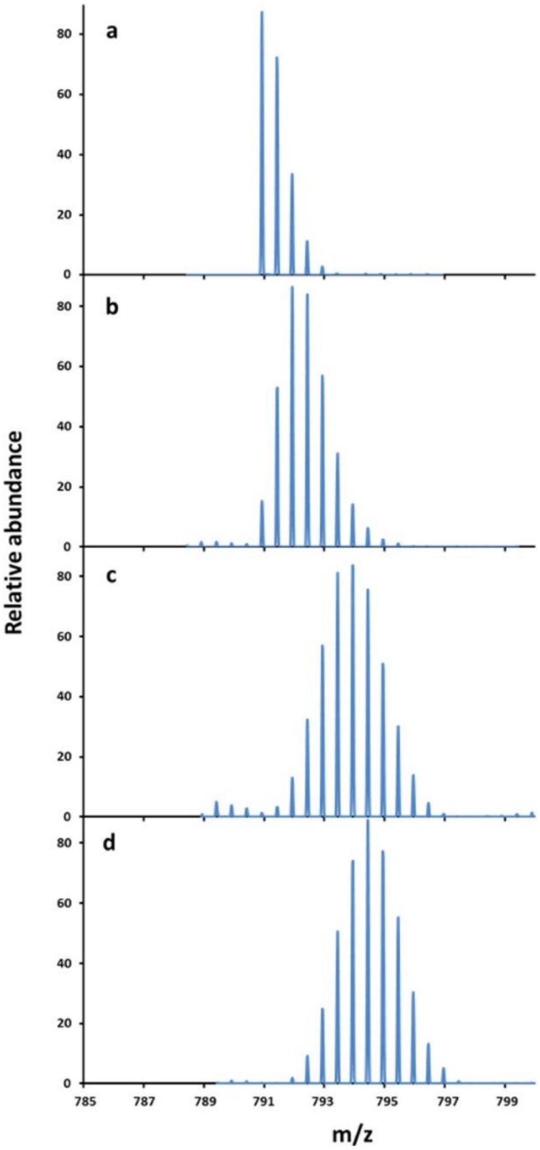


Figure 15: Time course for the uptake of deuterium by the peptide ^{68}TRAAVRLTDGTLGYS^{82} from PhnG as measured by ESI-MS after 0, 4 s, 5 min, and 1 h.

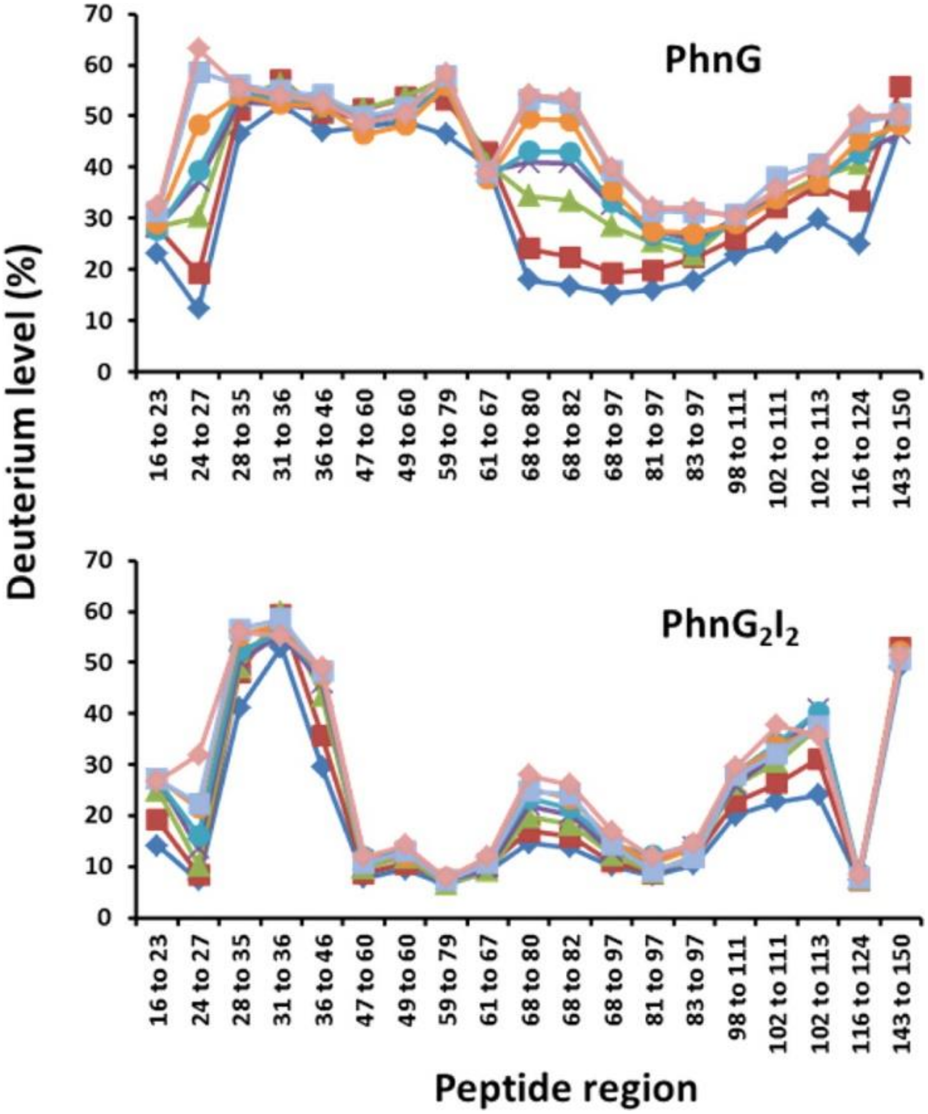


Figure 16: Differential deuterium exchange for the peptides contained within PhnG alone and the same peptide contained within $PhnG_2I_2$. The incubation times are as follows: 4 s (◆), 10 s (■), 1 min (▲), 2 min (□), 5 min (●), 15 min (●), 30 min (■), and 60 min (◆).

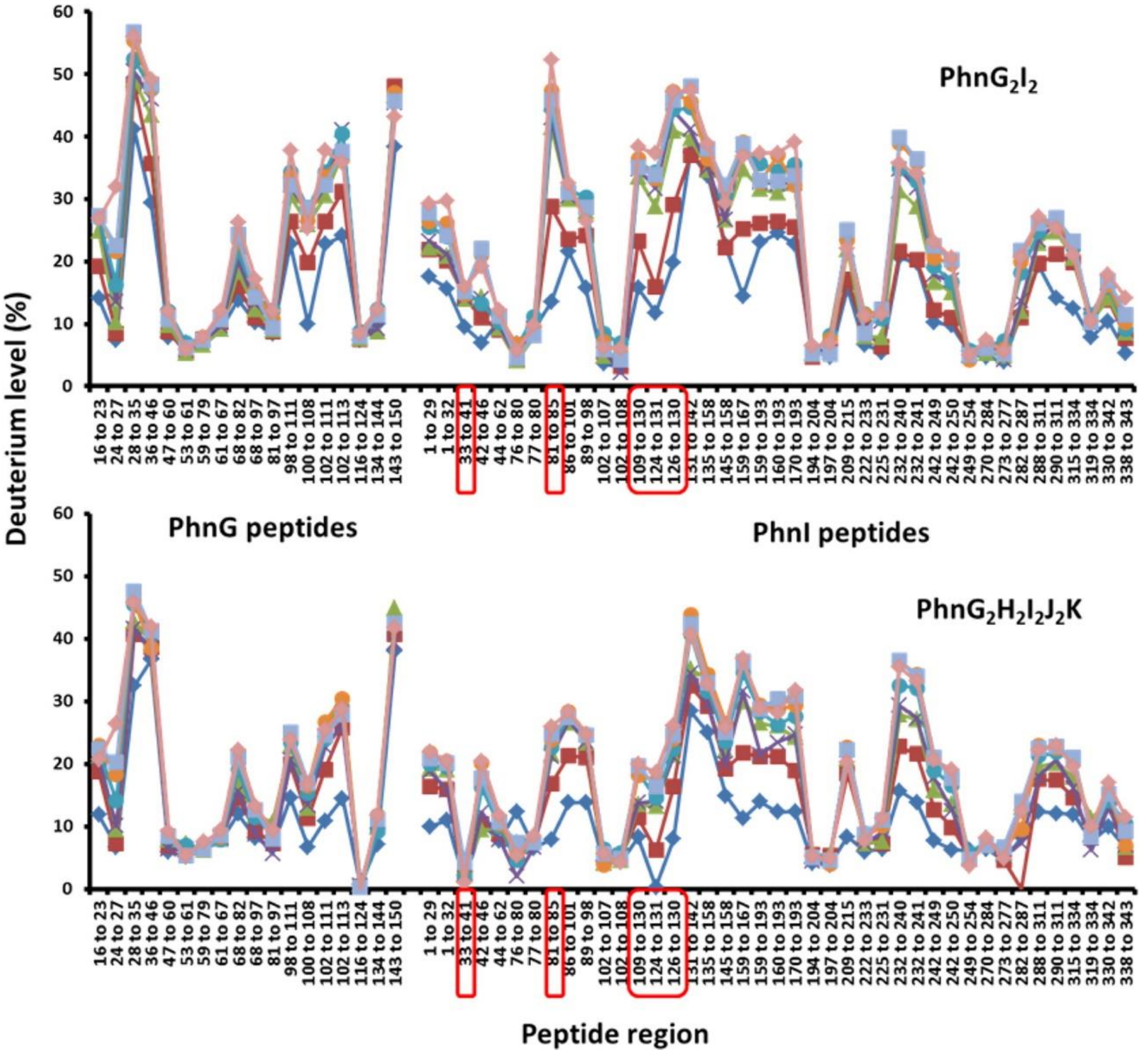


Figure 17: Differential deuterium exchange for the peptides contained within PhnG and PhnI for the complexes $\mathrm{PhnG_2I_2}$ and $\mathrm{PhnG_2H_2I_2J_2K}$. The peptides differing in exchange pattern between the two complexes are highlighted in red. The incubation times are as follows: 4 s (◆), 10 s (■), 1 min (▲), 2 min (□), 5 min (●), 15 min (●), 30 min (■), and 60 min (◆).

There are significant differences in the rates of deuterium exchange for those peptides from PhnG when this protein is contained within the PhnG$_2$I$_2$ or PhnG$_2$H$_2$I$_2$J$_2$ complexes. These differences are graphically depicted in the representative times courses for comparable peptides in Figure 16. For the peptides that encompass residues 16–23, 28–36, and 98–111, there are essentially no differences in the time courses for the uptake of deuterium by PhnG alone and PhnG contained within the PhnG$_2$I$_2$ complex. These results are in stark contrast to the significant differences in the exchange profiles for peptides that encompass residues 47–79 and 68–97 (Figure 16).

The H/D exchange profiles for the PhnG and PhnI subunits contained within the PhnG$_2$I$_2$ and PhnG$_2$H$_2$I$_2$J$_2$K complexes are quite similar to one another (Figure 14). For PhnG, there are virtually no differences in the exchange rates for portions of the protein when this complex is embedded in either PhnG$_2$I$_2$ or PhnG$_2$H$_2$I$_2$J$_2$K. For PhnI, the only regions exhibiting differential deuterium exchange profiles between the two complexes are the peptides contained within PhnI that encompass residues from 33–41, 81–85, and 109–130 (Figure 17).

These peptides show slightly more protection in the PhnG$_2$H$_2$I$_2$J$_2$K complex than in the PhnG$_2$I$_2$ complex. H/D exchange analysis of PhnH alone was carried out and compared with the peptides from PhnH obtained from the PhnG$_2$H$_2$I$_2$J$_2$K complex to help identify those regions of PhnH that undergo structural variations upon complex formation. A total of 23 common peptides, spanning 83% of PhnH, were identified. From the deuterium uptake plots for the PhnH peptides in the individual subunit and the PhnG$_2$H$_2$I$_2$J$_2$K complex, many regions along the entire length of PhnH are more protected in the PhnG$_2$H$_2$I$_2$J$_2$K complex (Figure 18). The N-terminus region of PhnH, which comprises the dimerization interface in the crystal structure of PhnH, is represented by the peptides encompassing residues 6–18, 7–18, 19–24, and 19–31. This region exhibits significant reduction in deuterium exchange in the PhnG$_2$H$_2$I$_2$J$_2$K complex as compared to the individual subunit of PhnH. A major difference in deuterium exchange profile is also observed in the region connecting α-helix D with a disordered loop and stretches up to β-strand β5 for the peptides spanning residues 110–123 (peptides 107–122 and 110–123). A high degree of differential exchange is also observed in the region extending from residues 136–146 as illustrated by the peptides ^{136}LRLTGAG^{142}, ^{136}LRLTGAGIAE^{145}, and ^{136}LRLTGAGIAEE^{146} (Figure 18).

This region extends from β-strand β6 and connects through a disordered loop region to β-strand β7 (Figure 19). Finally, the C-terminus region of PhnH, illustrated by the peptides encompassing residues 157–173, 159–173, and 183–194 are more protected in the complex compared to the PhnH alone (Figure 18).

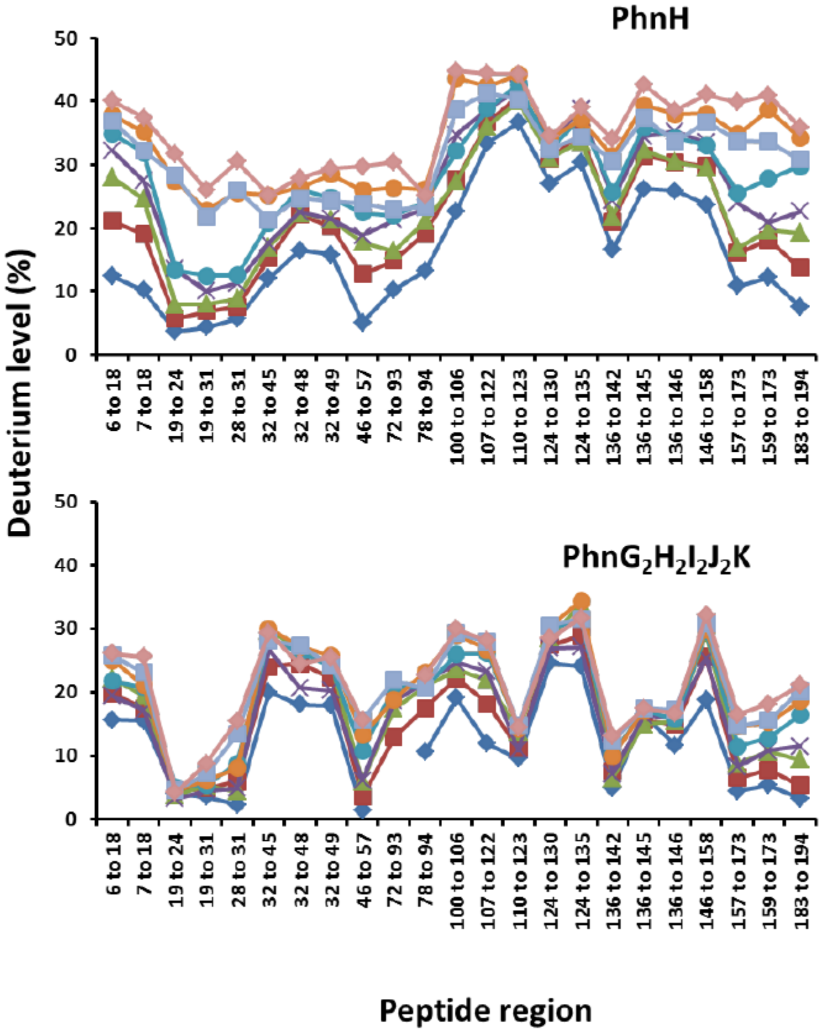


Figure 18: Differential deuterium exchange for the peptides contained within PhnH alone and the same peptides contained within the $\text{PhnG}_2\text{H}_2\text{I}_2\text{J}_2\text{K}$ complex. The incubation times are as follows: 4 seconds (blue square); 10 seconds (red square); 1 minute (green triangle); 2 minutes (purple cross); 5 minutes (blue circle); 15 minutes (light blue square); 30 minutes (orange circle); 60 minutes (pink square).

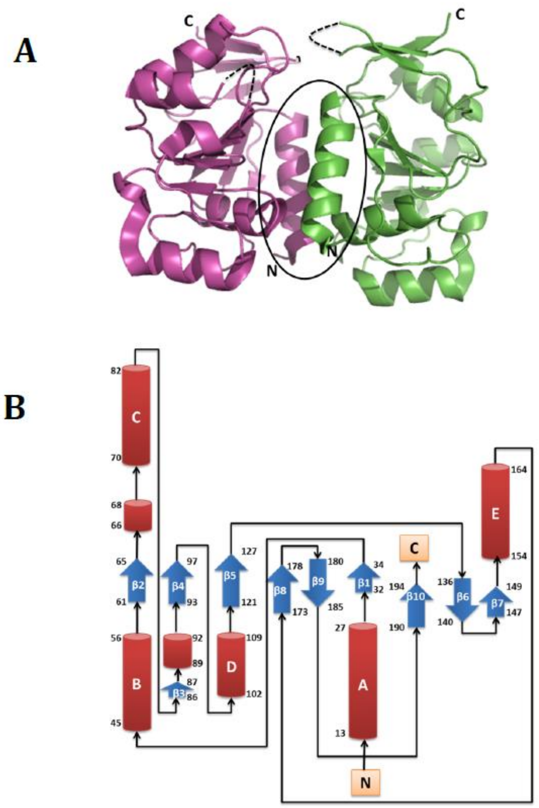


Figure 19: Crystal structure of PhnH. (A) Cartoon representation of PhnH dimer. The N and C terminus ends are labeled for each monomer. The dimer interface region undergoing differential exchange is circled. The dashed lines connecting two regions of the proteins indicate disordered regions not modeled in the crystal structure. These disordered regions undergo differential exchange between individual subunit of PhnH and in the PhnG$_2$H$_2$I$_2$J$_2$K complex. (B) Topology diagram of PhnH. The N and C terminus regions are labeled; structural regions are labeled and the numbers indicate the residue range spanning each region. Red cylinders represent α helices and blue arrows represent β---strands. Coordinates taken from PDB id: 2FSU.

2.3.7 Co-existence of $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$

The purified complexes were vitrified and imaged (Experimental Procedures). The collected movie stacks, from an electron-counting direct detection camera (112), showed clear particles and strong power spectra after aligning the frames within each movie (Figures 19B and 19C). A total of 492,889 particles were semi-automatically picked, and subsequently classified with reference-free two-dimensional (2D) classification and unsupervised three-dimensional (3D) classification in Relion. Two populations of particles were identified and reconstructed into two density maps, which correspond to the $PhnG_2H_2I_2J_2$ (core complex) and $PhnG_2H_2I_2J_2K$ (core complex bound with PhnK), respectively. These two maps were similar except for an extra lobe of density in one of them (Figure 19D). Differences in the extra density were also observed in the reference-free 2D class averages calculated from all particles (Figure 19E). This extra density was later confirmed as the PhnK subunit of the $PhnG_2H_2I_2J_2K$ complex. Further refinements against the separated particles yielded the cryo-EM density maps of $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$, both at 7.8 Å resolution, based on the gold-standard Fourier shell correlation. Local resolution was estimated, showing most regions of the maps were at a resolution range between 7 and 8 Å (Figure 19D). At this resolution, the protein secondary structures were clearly visible.

The existence of a complex missing PhnK, in the purified sample expressed from a plasmid containing the phnGHIJK genes, is consistent with previous native gel electrophoresis and mass spectrometry experiments, suggesting the loose binding of PhnK to the core complex $PhnG_2H_2I_2J_2$. The fact that $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$

co-exist in the same sample apparently hindered previous trials for the crystallization of $PhnG_2H_2I_2J_2K$ (127).

2.3.8 Architecture of $PhnG_2H_2I_2J_2$ with or without PhnK bound

The 3D density maps of the $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$ complexes measure 110 Å × 70 Å × 100 Å and 110 Å × 110 Å × 100 Å, respectively (Figure 20). The common part of these two complexes is the $PhnG_2H_2I_2J_2$ core. The core retains the same architecture after PhnK binding. In both density maps, the two copies of PhnG, PhnH, PhnI, and PhnJ are arranged in a head-to-tail fashion in a sequence of H-J-I-G as labeled in Figure 19B. While the cryo-EM maps were independently determined and not biased by the recently determined crystal structure of $PhnG_2H_2I_2J_2$ (Experimental Procedures), the crystal structure fits well into the cryo-EM density map of the core complex $PhnG_2H_2I_2J_2$ with matching secondary structures (Figure 20C).

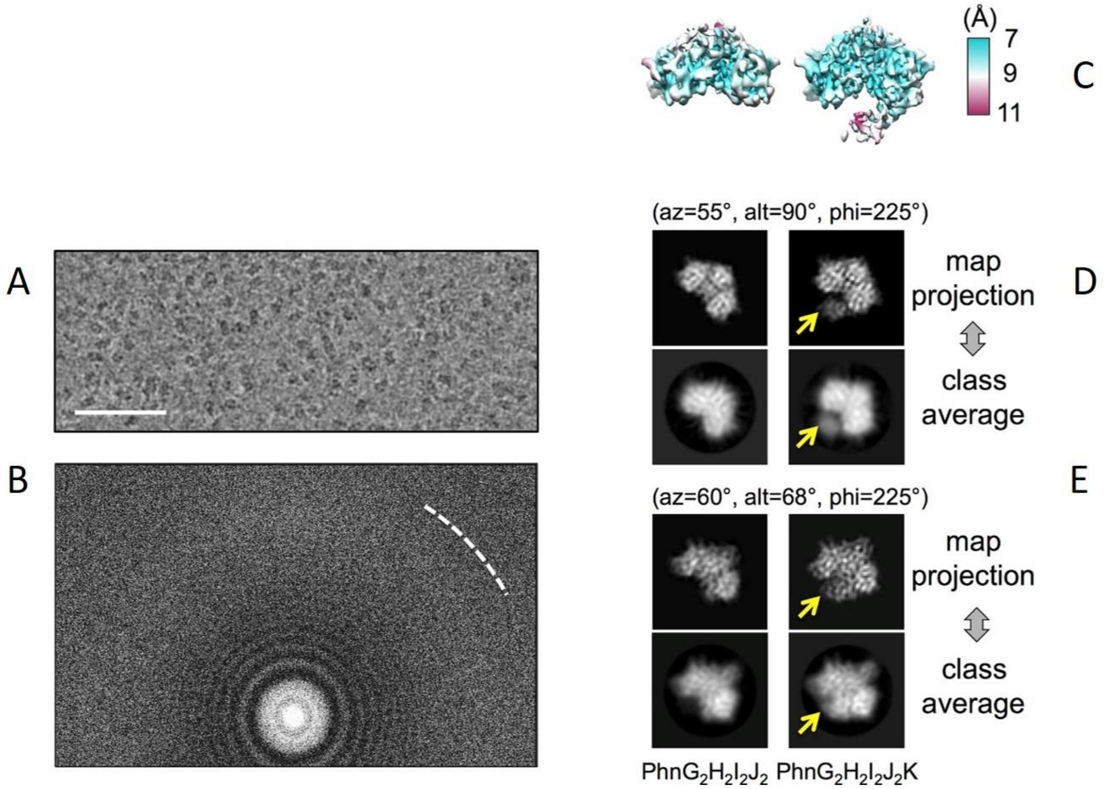


Figure 20: Cryo-EM analysis reveals the co-existence of $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$ (A) The series of the enzymatic steps of C-P lyase. (B) A region of a representative micrograph showing the particles of $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$ with a defocus of 1.2 μm. The scale bar denotes 500 Å. (C) Power spectrum of the same micrograph in (B). The white dashed curve shows the Nyquist at 3 Å. (D) 3D density maps of $PhnG_2H_2I_2J_2$ (left) and $PhnG_2H_2I_2J_2K$ (right). The maps were colored based on the local resolutions. (E) Projections of the two maps compared with the reference-free class averages in two representative orientations. The Euler angles were based on the EMAN2 convention. Yellow arrows indicate the location of PhnK.

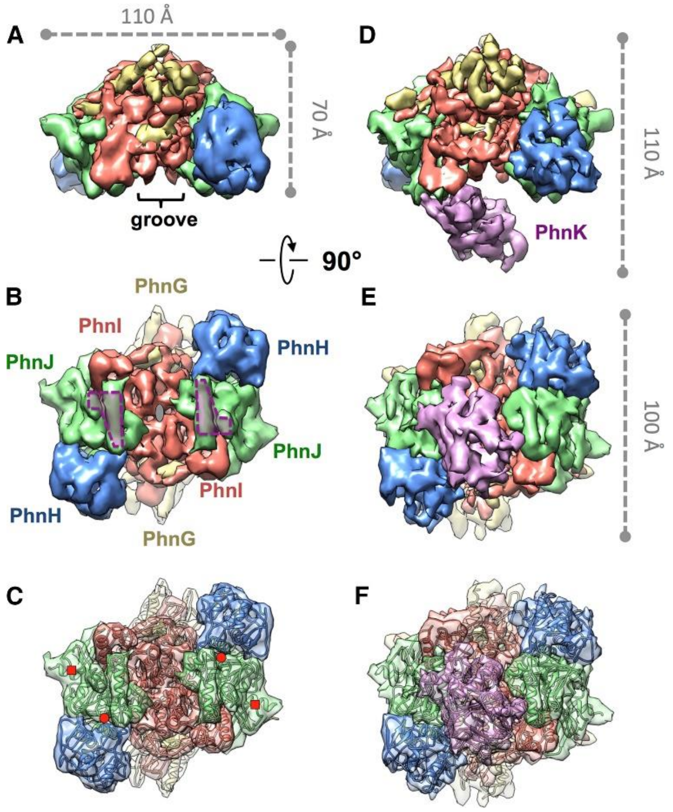


Figure 21: Architecture of PhnG$_2$H$_2$I$_2$J$_2$ without or with PhnK bound (A and B) front and bottom view of PhnG$_2$H$_2$I$_2$J$_2$. In (A), the groove on the core complex PhnG$_2$H$_2$I$_2$J$_2$ is labeled. In (B), the gray ellipse denotes the two-fold symmetry axis, and the two purple polygons denote the two identical binding sites for PhnK. (C) The crystal structure of PhnG$_2$H$_2$I$_2$J$_2$ fitted into the cryo-EM map of PhnG$_2$H$_2$I$_2$J$_2$ in the same orientation as (B). The red circles and squares denote the positions of Gly32 and Cys272 of PhnJ, respectively. (D and E) Front and bottom view of PhnG$_2$H$_2$I$_2$J$_2$K in the same orientations as (A) and (B), respectively. (F) The crystal structure of PhnG$_2$H$_2$I$_2$J$_2$ and the homology model of PhnK fitted into the cryo-EM map of PhnG$_2$H$_2$I$_2$J$_2$K in the same orientation as (E).

2.3.9 Two identical binding sites for PhnK on $PhnG_2H_2I_2J_2$

The binding site for PhnK is located on α helix 6 (residue 147–158) and a nearby loop (residue 227–230) of PhnJ. We term this loop the “chock loop” as it may stabilize the binding of PhnK like a chock. Two PhnJ subunits in the core complex provide two identical binding sites for PhnK (purple polygons in Figure 20C). These two binding sites are approximately 50 Å away from each other. Noticeably, between these two binding sites lies a negatively charged groove, which was previously proposed as the PhnK-binding region based on a low-resolution density map of $PhnG_2H_2I_2J_2K$ from negative stain EM and protein sequence conservation (127). However, the entire groove is exposed to solvent and not in contact with PhnK in our cryo-EM map of $PhnG_2H_2I_2J_2K$ (Figure 21).

Even though the core complex provides two identical binding sites for PhnK, only one copy of PhnK was found to bind (Figures 20C–19E and 21). PhnK is homologous to the conserved NBD of proteins in the ABC transporter family. Like NBDs, PhnK consists of a helical domain and a RecA-like domain. It has all of the sequence motifs (Walker A, Walker B, ABC signature, A-loop, D loop, Q-loop, and switch H-loop) of the NBD (Figure 22A) (128, 129). A sequence alignment of PhnK with other structurally characterized NBDs is presented in Figure 24. NBDs are known to interact with a transmembrane domain or protein (TMD) through its helical domain (130). The L-loop of the TMD inserts into a cavity around the helical domain of the NBD, which forms the major interaction between NBD and TMD (131). In contrast, PhnK, an NBD-like protein, interacts with the cytoplasmic protein complex

$PhnG_2H_2I_2J_2$. This interaction is via a helix-turn-helix motif (residues 100–120 containing the α helices 3 and 4) within the helical domain of PhnK to bind to α helix 6 and chock loop of PhnJ (Figures 22B and 23A).

When PhnK binds to PhnJ, the cavity of PhnK is completely exposed to solvent. The chock loop of PhnJ may instead provide additional support to block the movement of PhnK. The binding interface between PhnK and PhnJ shows complementary electrostatic surface potentials. At the interface, PhnK exhibits a net positive charge while PhnJ exhibits a negative charge (Figure 23B). Attempts to dock a second copy of PhnK, in the same conformation as the first one, to the empty binding site in $PhnG_2H_2I_2J_2K$, resulted in considerable collision between their RecA-like domains, which prevents the simultaneous binding of two PhnK subunits.

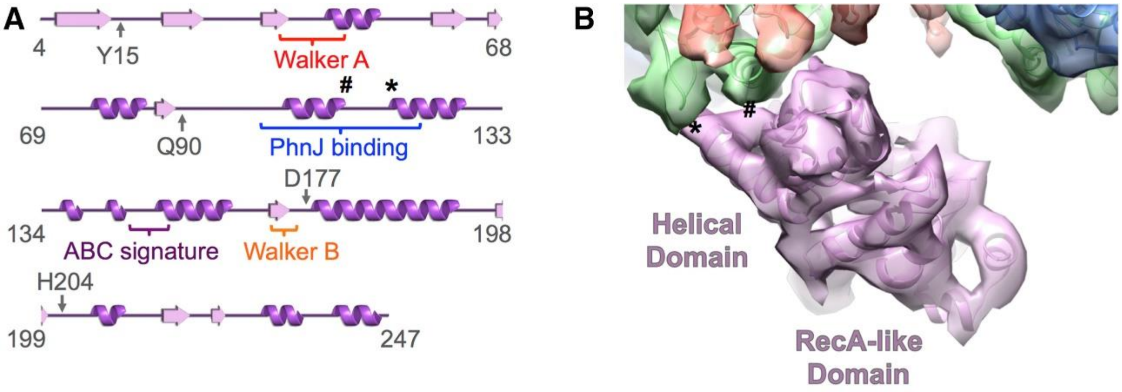


Figure 22: PhnK is an NBD-like protein and interacts with PhnJ (A) Secondary structure of PhnK with the NBD motifs labeled. The symbols ∗ and # label the sites for the interactions between PhnK and PhnJ. (B) One PhnK binds to the core complex with the models fitted in the map.

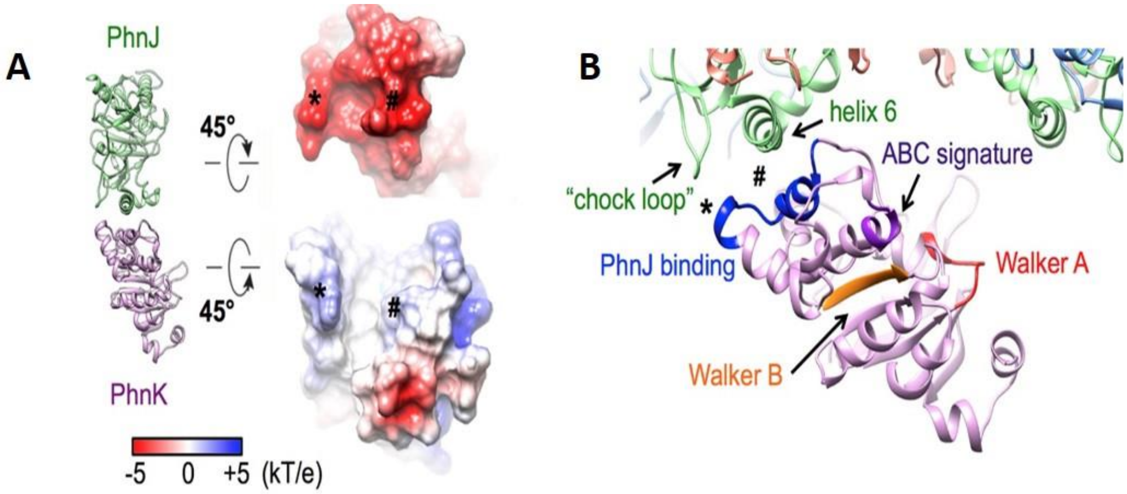


Figure 23: (A) Electrostatic surface potentials at the interface of PhnJ and PhnK (B) One PhnK binds to the core complex with the NBD motifs color labeled.

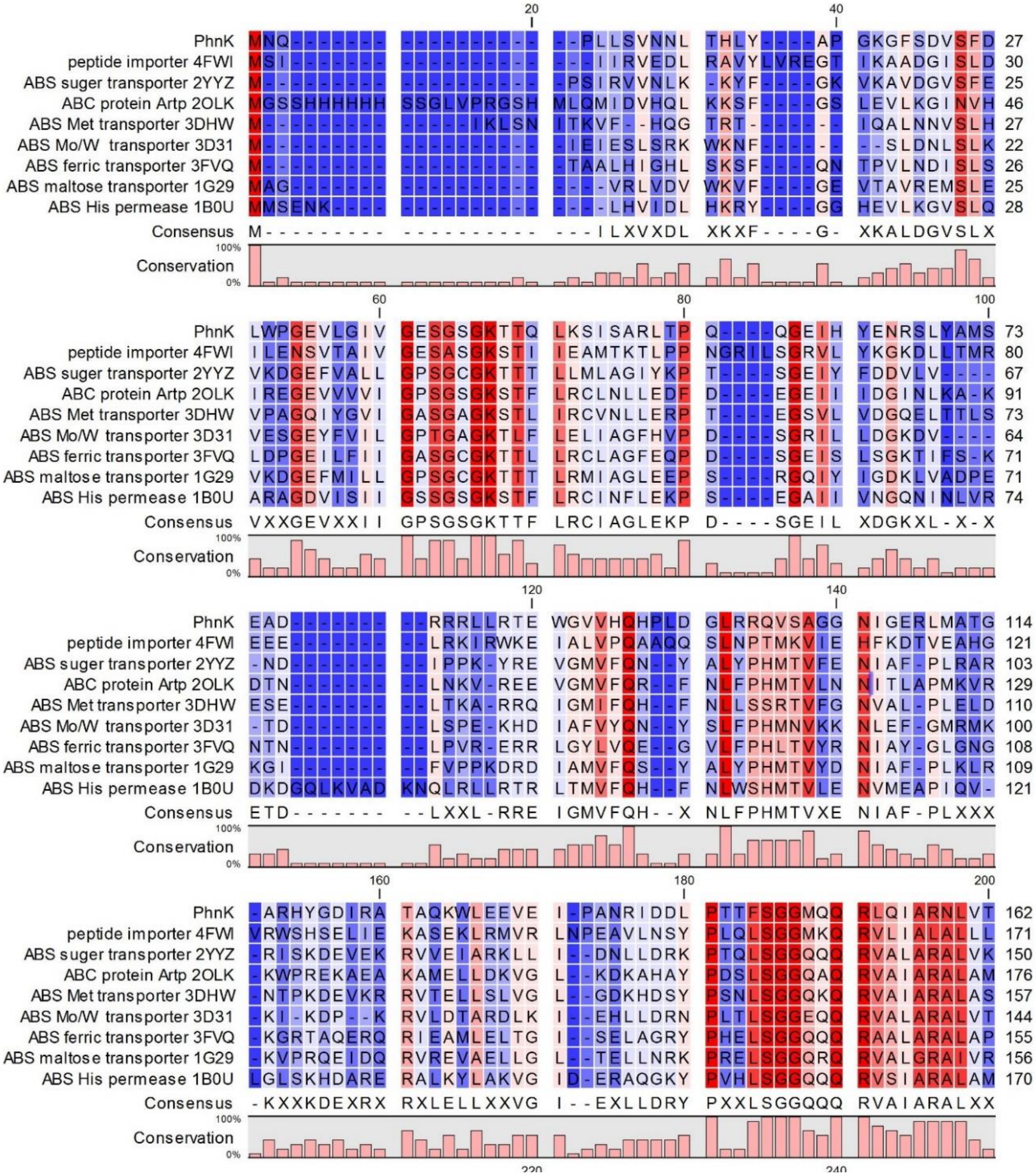


Figure 24: Sequence alignment of PhnK, PhnL with NBDs in the ABC transporter family using CLC sequence viewer (CLC bio, QIAGEN). Binding of PhnK Exposes the Active Site Residue Gly32

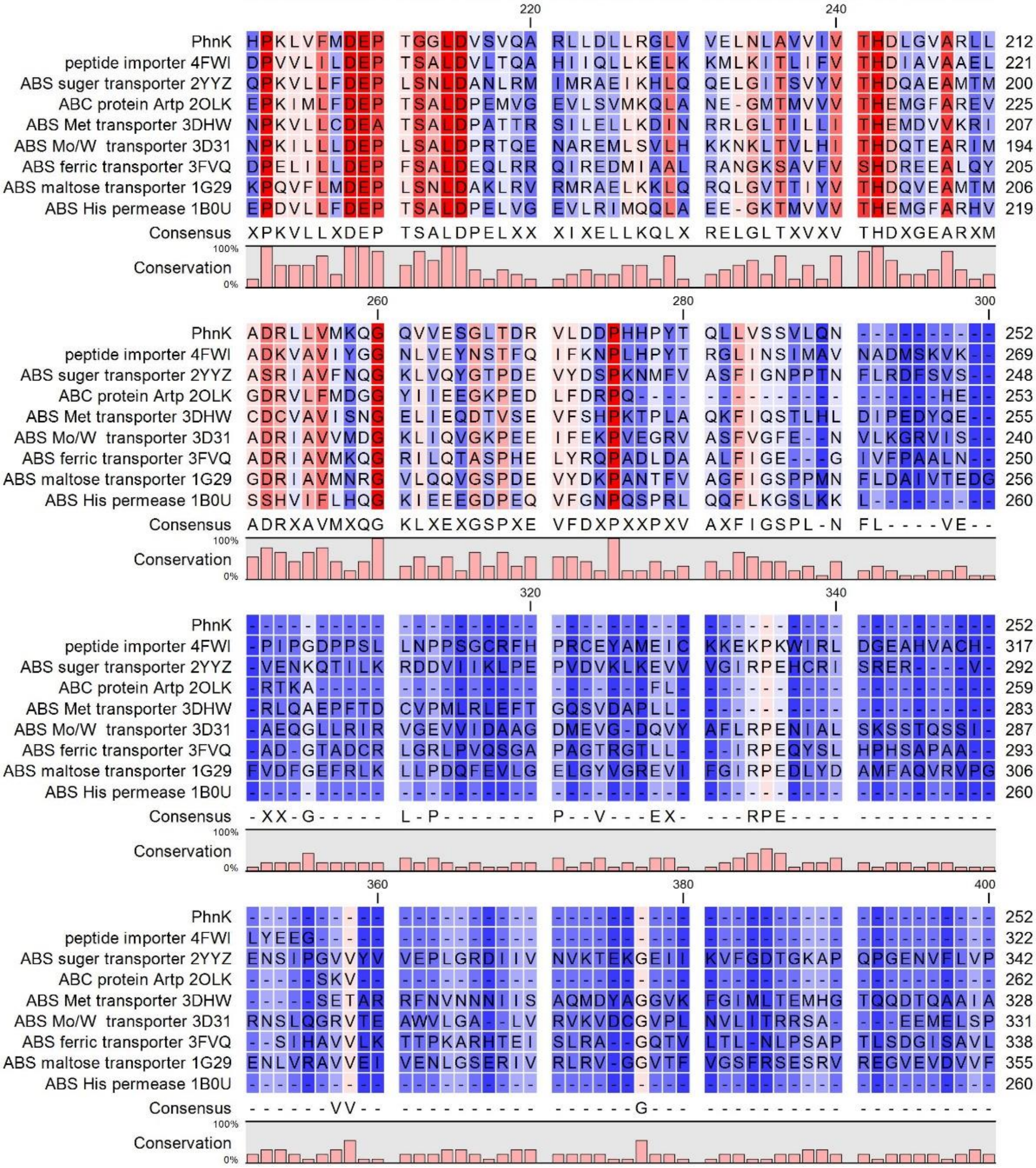


Figure 24 (Continued)

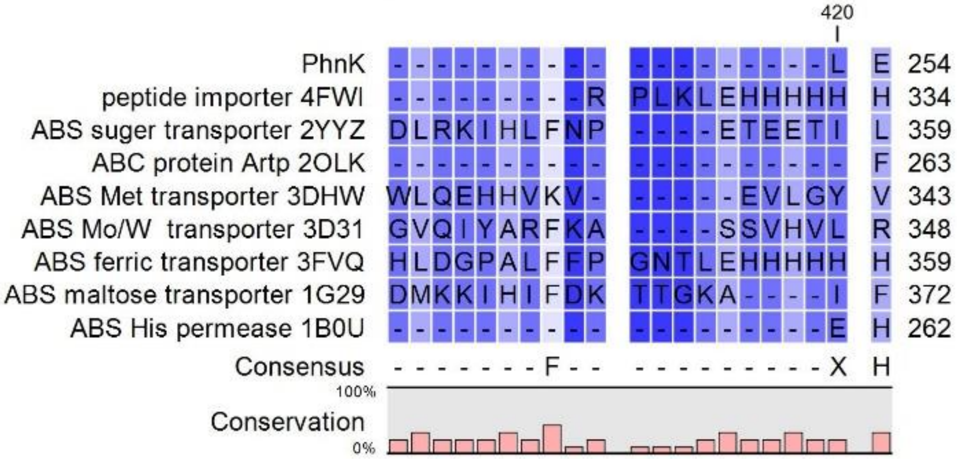


Figure 24 (Continued)

The universally conserved Gly32 of PhnJ was proposed to generate a stable glycyl radical enzyme that supports multiple turnovers without further SAM consumption (53). In the core complex, the two Gly32, each from one subunit of PhnJ, are symmetrically located on the two sides of the complex. Both of these glycine residues are buried in PhnJ near the interface between PhnJ and PhnH. In the map of $PhnG_2H_2I_2J_2K$, the density burying Gly32 of PhnJ shows striking difference between the two copies of PhnJ (Figure 33).

On the side where there is no PhnK bound (apo side), the density burying Gly32 is stronger than the corresponding density on the side where there is a PhnK bound (bound side), indicating the interaction between PhnJ and PhnH in the apo side is tighter than in the bound side. On the apo side, the residue Gly32 is hidden, as in our cryo-EM density map of $PhnG_2H_2I_2J_2$ (Figure 33A). On the bound side, the α helices 5 and 6 of PhnJ move away from PhnH, disrupting the molecular interaction between PhnJ and PhnH to expose the active site residue Gly32 (Figure 33B). On the apo side, the density of α helices 5 and 6 of PhnJ matches well with the crystal structure, while on the bound side, the density of α helices 5 and 6 of PhnJ is shifted relative to the crystal structure. This shift indicates there is a structural rearrangement in this region to reveal the active site underneath.

2.4 Discussion

2.4.1 Stoichiometry of the C–P Lyase complexes

Three different protein complexes from the carbon–phosphorus lyase enzyme were expressed, purified, and characterized by gel electrophoresis, analytical ultracentrifugation, N-terminus protein sequencing, and high-resolution mass spectrometry. For the complex that contains only PhnG and PhnI, the results are consistent with an equal mixture of the two subunits and an overall stoichiometry of $PhnG_2I_2$. In the complex that contains these two subunits, in addition to PhnH and PhnJ, the results are consistent with a stoichiometry of $PhnG_2H_2I_2J_2$. For the complex that contains five different subunits, the complex contains two copies of PhnG, PhnH, PhnI, and PhnJ, plus a single copy of PhnK ($PhnG_2H_2I_2J_2K$).

These results suggest that PhnG and PhnI associate to form a central heterodimeric core ($PhnG_2I_2$) that is required prior to the binding of two additional copies each of PhnH and PhnJ (PhnHJ) to facilitate the assembly of $PhnG_2H_2I_2J_2$. The largest complex is subsequently formed by the binding of a single copy of PhnK. The proposed assembly pathway for the formation of the $PhnG_2H_2I_2J_2$ and the $PhnG_2H_2I_2J_2K$ complexes is further supported by the mass spectrometry results where individual subunits are progressively dissociated from the central core complex. With the $PhnG_2H_2I_2J_2$ complex, we observed the sequential dissociation of PhnH, PhnJ, PhnH, PhnJ, PhnG, and PhnI.

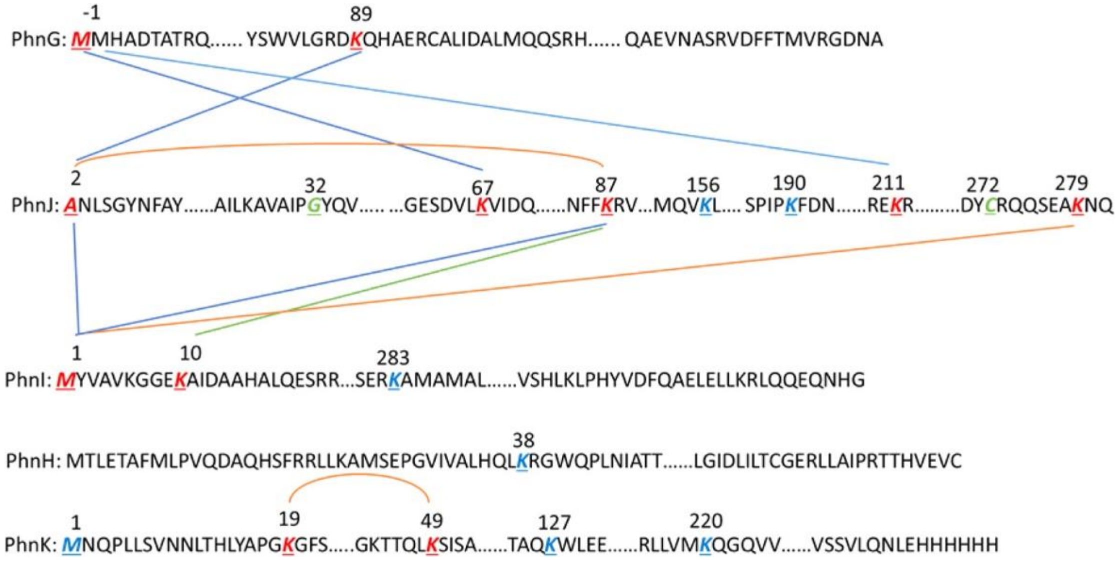


Figure 25: Cross-linking map of the PhnG$_2$H$_2$I$_2$J$_2$K complex. All of the cross-linked residues are highlighted in red, while all of the residues that participate in a monolinkage are highlighted in blue. Two residues that were reported to be key residues within the active site of PhnJ (Gly-32 and Cys-272) are highlighted in green. The cross-linking bonds are depicted as straight lines between different cross-linking sites. The internal cross-linked bonds are presented as a curve connecting two lysine residues.

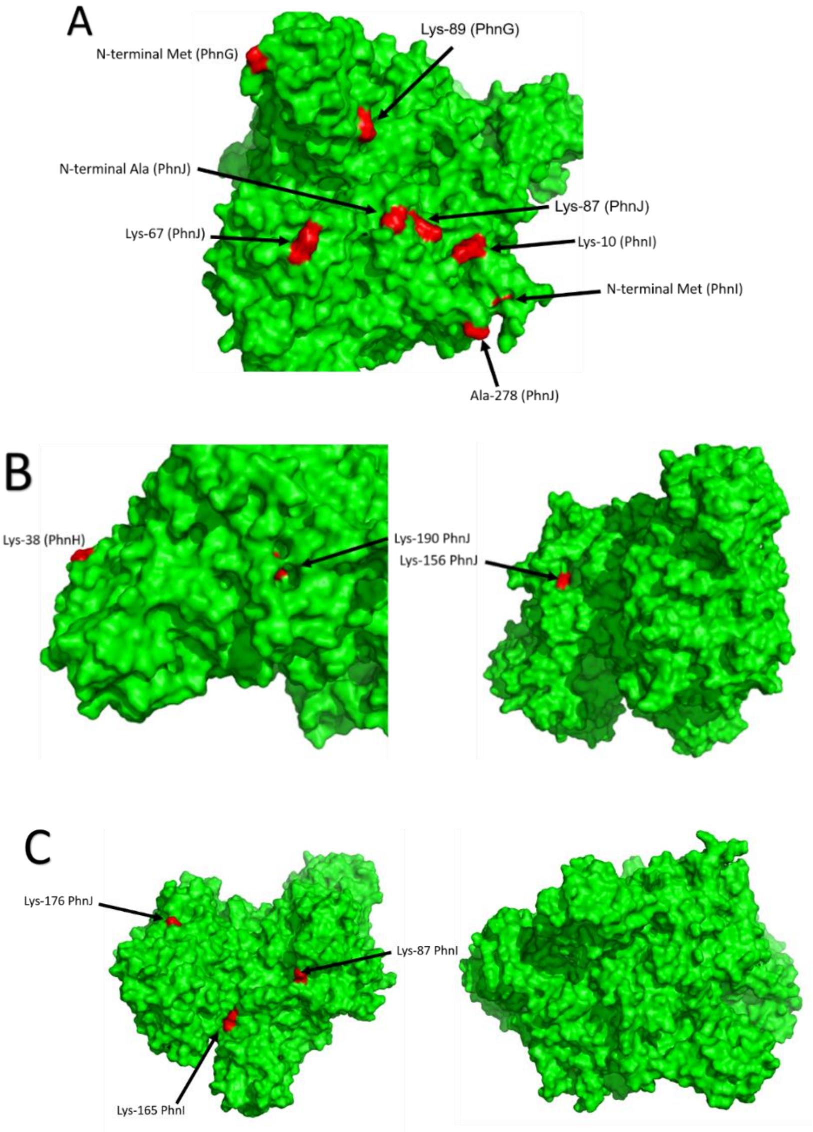


Figure 26: Solvent exposed lysine residues and the N-termini of the Phn$G_2H_2I_2J_2$ complex. (A) Solvent exposed lysine residues and N-termini that participate in the formation of crosslinked bonds. (B) Solvent exposed lysine residues and subunit N-terminus that are labeled by a mono-link. (C) Solvent exposed lysine residues and subunit N-terminus that are not modified by mono-link or cross-link.

2.4.2 Chemical cross-linking

Chemical cross-linking experiments were conducted to identify those subunits that are physically able to contact one another, and these results are graphically illustrated in Figure 25. There are nine identified cross-linked bonds including the inter-subunit and intra-subunit cross-links in the $PhnG_2H_2I_2J_2K$ complex. In addition, seven residues are found to be modified by a mono-link. However, there are 16 additional lysine residues that are not modified by the cross-linking reagent. According to the X-ray structure of $PhnG_2H_2I_2J_2$, most of the residues that are connected by a cross-linked bond or singly modified are located on the surface of the complex while the majority of the unlabeled lysine residues are buried inside the complex (Figure 26). These results indicated that the N-termini or lysine residues with high solvent accessibility are reacted more readily by the cross-linking reagent.

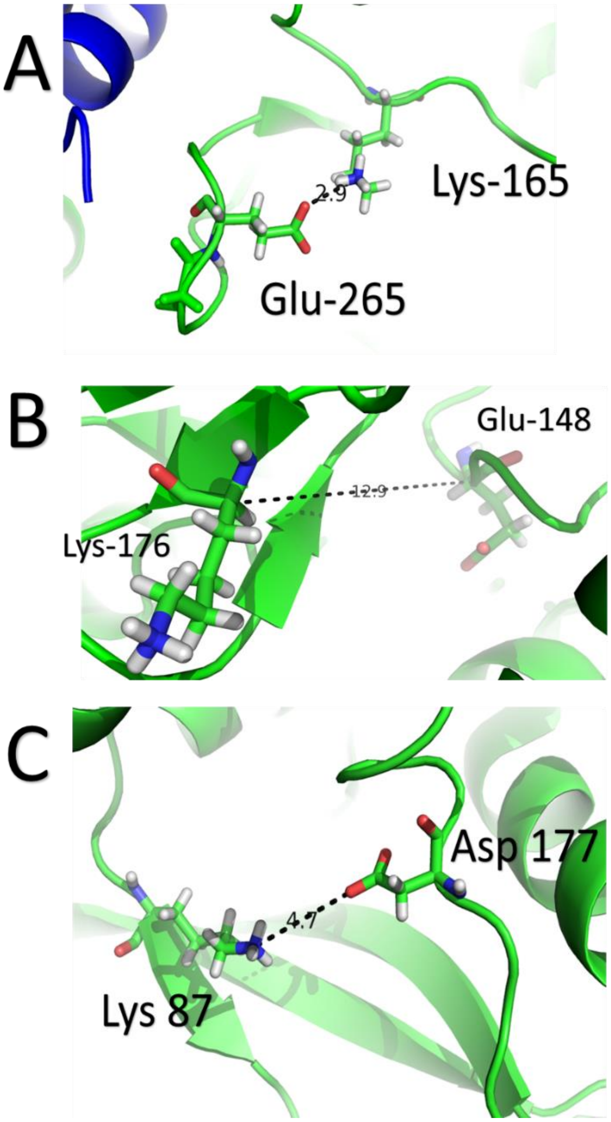


Figure 27: The formation of salt bridge between the unlabeled lysine residues that are the surface and the adjacent aspartic or glutamic acid residues. (A) Lys-165 and Glu-265 are from subunit PhnI. PhnI is colored green and PhnJ is colored blue. (B) Lys-176 and Glu-148 are from subunit PhnJ. PhnJ is colored green. (C) Lys-87 and Asp-177 are from subunit PhnI. PhnI is colored green.

2.4.2.1 Unlabeled lysine residues of $PhnG_2H_2I_2J_2$

In $PhnG_2H_2I_2J_2$ complex, there are 14 lysine residues that are not modified. Based on the X-ray structure, most of the unmodified lysine residues are buried inside the complex. However, three lysine residues (Lys-176 of PhnJ, Lys-165 of PhnI and Lys-87 of PhnI) are found on the surface of the complex (Figure 26C). These three residues can form a salt bridge with an adjacent aspartic acid or glutamic acid residue (Figure 27), which may prevent these lysine residues from being labeled by the cross-linking reagent. The side chain orientation of Lys-176 and Glu-148 of PhnJ cannot be aligned as a salt bridge in the crystal structure of $PhnG_2H_2I_2J_2$ (Figure 27B). However, the distance between α-carbons of these two residues is 12 Å, which is shorter than the maximal length for the formation of a salt bridge, 15 Å. The maximal length is calculated by the sum of the maximal length of a salt bridge, 4 Å, the side chain length of a lysine residue, 6 Å and the side chain length of a glutamic residue, 5 Å. Potentially, a salt bridge could be formed between these two residues. In addition, Lys-165 of PhnI is on a flexible loop (average B-factor 41 $Å^2$) that extends into the solvent, this residue may have been involved in an inter-complex crosslinking which was removed from the sample by SDS-PAGE before mass spectrometry analysis.

2.4.2.2 Cross-linked or mono-linked modifications of subunit PhnG

The N-terminal methionine and Lys-89 of PhnG are shown to be cross-linked with lysine residues of subunit PhnJ. There are no other mono-linked labeled or unlabeled lysine residues on PhnG. The N-terminus of PhnG is not located at the interface between PhnG and PhnJ, despite the identification of cross-linked bonds between the methionine at the N-terminus of PhnG and the two lysines (Lys-211 and Lys-67) of PhnJ. However, these two lysine residues are located at the interface between PhnG and PhnJ. Lys-67 is exposed to the solvent while Lys-211 is buried by the interacting subunit PhnG. The distance between the N-terminus of PhnG and these two lysine residues from PhnJ is as long as 50 Å, a distance that is much longer than the maximal length of a cross-linked bond between these residues. The N-termini of PhnG is at an unstructured sequence with high mobility. The average B-factor of this region is 49 $Å^2$ while the average B-factor of the whole complex is around 29 $Å^2$. The formation of these cross-linked bonds must therefore be attributed to conformational changes within the complex.

Lys-89 of PhnG and the N-terminus of PhnJ were shown to be connected by a cross-linked bond. In the X-ray structure, the distance between Lys-89 of PhnG and the N-terminal alanine of PhnJ is 29 Å, obviously longer than the maximal length of the cross-linked bond, 20Å (Figure 28). The maximal length of a lysine to N-terminus crosslinked bond is calculated by taking the sum of the arm length of the cross-linking reagent (11.5 Å), the side chain length of a lysine residue (6.5Å) and the length of carbon nitrogen bond of the N-terminal alanine (1.5 Å). However, the N-terminus of PhnJ is at the end of a flexible loop. The average B-factor value of this loop is as high as 45 $Å^2$, compared with an average B-factor value 29 $Å^2$ for the whole protein complex. It is possible that the loop is subject to a conformational change, which could reduce the distance between Lys-89 of PhnG and the N-terminus of PhnJ.

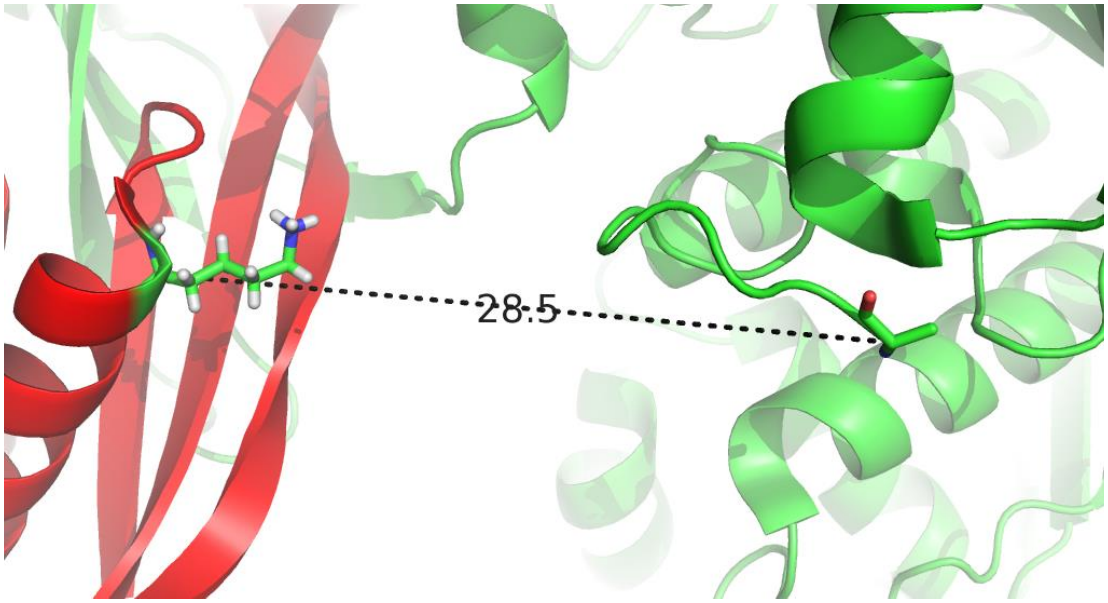


Figure 28: The distance between Lys-89 of PhnG and N-terminus of PhnJ in the $PhnG_2H_2I_2J_2$ complex. A cross-linked bond was identified connecting Lys-89 of PhnG to N-terminus of PhnJ by mass spectrometry analysis. Subunit PhnJ is colored green and subunit PhnG is colored red. The distance between the α-carbons of these two residues was measured as 28.5 Å.

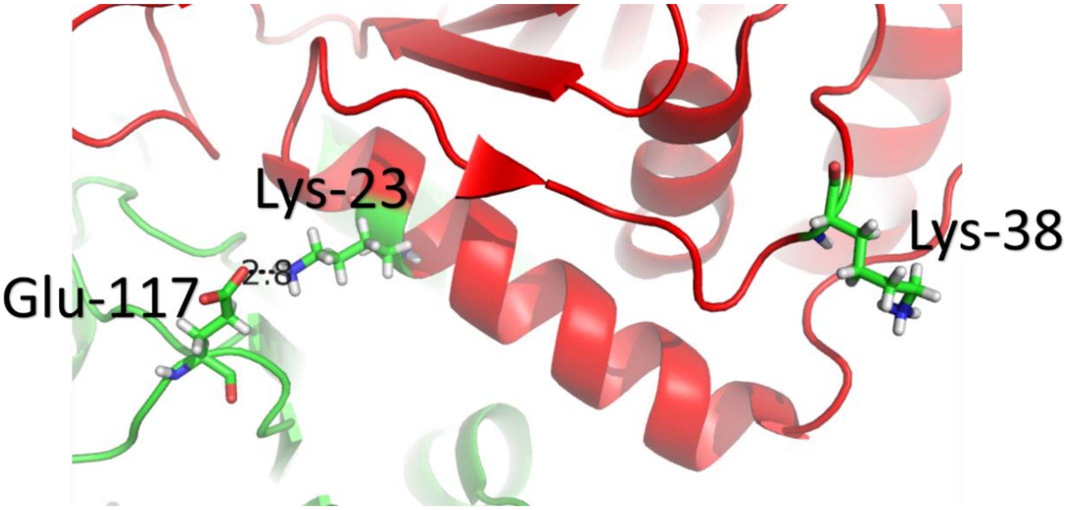


Figure 29: Lys-23 and Lys-38 of PhnH in the $PhnG_2H_2I_2J_2$ complex. A potential salt bridge is observed between Lys-23 of PhnH and Glu-117 of PhnJ. The distance between the ε-amino group and carboxylate group is 2.8 Å. PhnH is colored red and PhnJ is colored green.

2.4.2.3 <u>Cross-linked or mono-linked modifications of subunit PhnH</u>

There are two lysine residues (Ly-23 and Lys-38) on subunit PhnH, however, no cross-linked peptides were identified for these two sites. A mono-linked modification was found on Lys-38. According to the crystal structure, Lys-38 is solvent exposed and located away from the interface between PhnH and PhnJ (Figure 29). The closest lysine residue on PhnJ is 33 Å away from Lys-38, which is longer than the maximal cross-linked length between two lysine residues of 25 Å. The maximal length of the lysine to lysine crosslinked bond is calculated by the sum of the arm length of the cross-linking reagent (11.5 Å) and side chain length of two lysine residues (13.0 Å). Moreover, there are two α-helices between these two residues, blocking the formation of a cross-linked bond. Located at the interaction region between PhnH and PhnJ, Lys-23 is able to form a salt bridge with Glu-177 of PhnJ, which could hinder a cross-link or mono-link modification to this lysine residue.

2.4.2.4 <u>Cross-linked or mono-linked modifications of subunit PhnI and PhnJ</u>

There are three mono-linked labeled lysine residues of PhnI and PhnJ, identified by cross-linking mass spectrometry analysis. In the X-ray structure, Lys-156 and Lys-190 of PhnJ are located at the solvent exposed surface, far away from the protein-protein interaction region. Lys-283 of PhnI is buried inside the complex but it can be solvent accessible through a short hydrophilic channel.

Four cross-linked bonds were located at a solvent exposed region of the interface between PhnI and PhnJ (Figure 30). Four residues (Lys-10 of PhnI, Lys-87of PhnJ, N-terminus of PhnJ and N-terminus of PhnI) were shown to be connected with each other by the cross-linking reagent. In the X-ray structure, the distance between Lys-87 and N-terminal alanine from PhnI is 9 Å, which is shorter than the maximal length of the cross-linked bond, 20 Å. The distance between Lys-87 from PhnJ and Lys-10 from PhnI is 18 Å, which is also shorter than the maximal length of the cross-linked bond, 25 Å. The distance between other crosslinked sites is longer than the maximal length of a cross-linked bond (25 Å for lysine to lysine cross-link, 20 Å for lysine to N-terminus cross-link and 14.5 Å for N-terminus to N-terminus cross-link), which could be attributed to the flexibility of the N-terminus of PhnJ and PhnI. The N-terminus to N-terminus cross-link is calculated by taking the sum of two carbon nitrogen bond (3.0 Å) and the length of the crosslinked bond (11.5 Å). The average B-factors for these two N-terminus loops of PhnI and PhnJ are 45 $Å^2$ and 49 $Å^2$, respectively. Lys-279 of PhnJ is not included in the X-ray structural model so the precise distance between this residue and N-terminal methionine of PhnI could not be measured. However, Ala-278, the residue located before Lys-279, is included in the crystal structure. The distance between the α-carbon of Ala-278 of PhnJ and N-terminus of PhnI is 16 Å, which is shorter than the maximal length of a cross-linked bond, taking the length of side chains and amide bond into consideration (Figure 31).

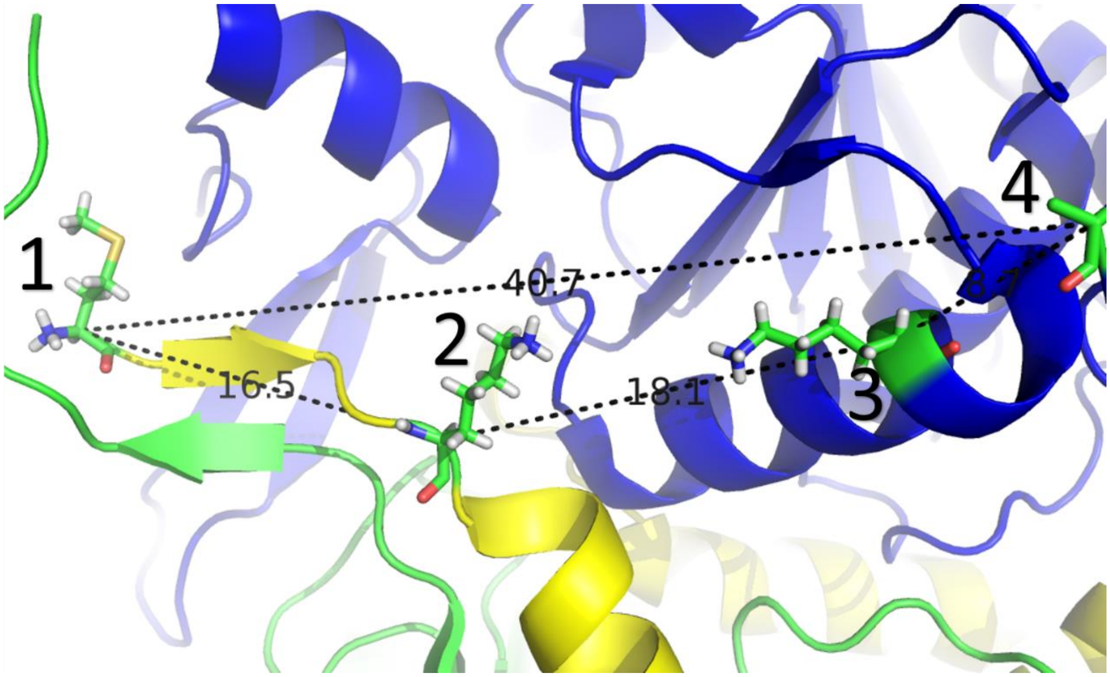


Figure 30: The distance between lysine and N-terminus residues at the interface between PhnI and PhnJ in the PhnG$_2$H$_2$I$_2$J$_2$ complex. PhnI is colored yellow and PhnJ is colored blue. Other subunits are colored green. The N-terminal methionine and Lys-10 of PhnI are labeled as 1 and 2, respectively. Lys-87 and the N-terminal alanine of PhnJ are labeled as 3 and 4, respectively. The distance between the α-carbons of two residues where a cross-linked bond was as follows, residue 1 to residue 2: 16.5 Å; residue 2 to residue 3, 18.1 Å; residue 3 to residue 4, 8.7 Å and residue 4 to residue 1, 40.7 Å.

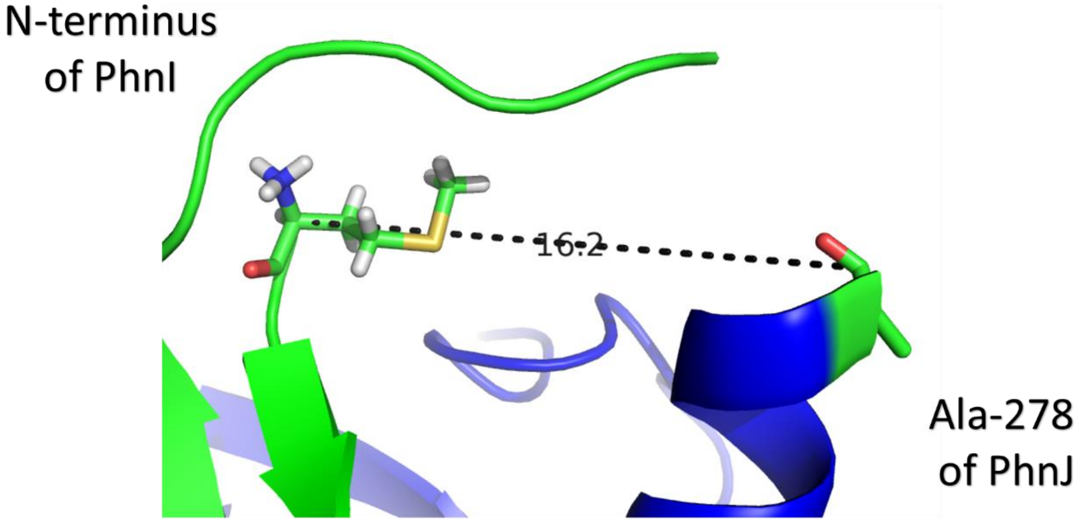


Figure 31: The distance between Ala-278 of PhnG and the N-terminus of PhnI in the $PhnG_2H_2I_2J_2$ complex. The distance between the α-carbons of these two residues was measured as 16.2 Å. A cross-linked bond was identified connecting Lys-279 of PhnG to N-terminus of PhnI by cross-linking mass spectrometry analysis. Subunit PhnJ is colored blue and subunit PhnI is colored green. The distance between the α-carbons of these two residues was measured as 16.2 Å.

2.4.2.5 Cross-link or mono-link modification of subunit PhnK

We identified an internal cross-link between two lysine residues and the labeling of two single lysine residues (Lys-127 and Lys-220) within PhnK. This result indicates that the region containing these two lysine residues (Lys-19 and Lys-49) must be close to each other in the complex. Although the lysine residues in the homology model of PhnK appear to be on the protein surface, no peptides of PhnK were cross-linked to any other subunit. Based on the cryo-EM structural model of PhnG$_2$H$_2$I$_2$J$_2$K, these residues are not involved in the interaction between PhnK and the other subunits. Due to the resolution limitation, an accurate distance between the intra-linked two lysine residues cannot be determined. However, the homology model of PhnK indicates that residues Lys-19 and Lys-49, which are internally cross-linked, are separated by a distance of about 20 Å, which is shorter than the maximal length (25 Å) of a cross-linked bond between two lysine residues.

2.4.3 Hydrogen/deuterium exchange

Although cross-linking mass spectrometry is a powerful tool for identifying the subunit interactions in protein complexes, the requirement for two adjacent lysine residues greatly limits its application. To further reveal more subunit interactions, the assembly of the multi-protein complexes was addressed by measuring the rates of exchange of deuterium into the amide backbone from solvent. Two proteins alone (PhnG and PhnH) and two protein complexes (PhnG$_2$I$_2$ and PhnG$_2$H$_2$I$_2$J$_2$K) were subjected to H/D exchange mass spectrometry analysis. Protein PhnI, PhnJ and PhnK are not used in this H/D exchange analysis because they are insoluble without a solubility tag.

2.4.3.1 Subunit interaction of PhnG within the $PhnG_2I_2$ and $PhnG_2H_2I_2J_2K$ complexes

For PhnG alone, about 40% of the protein backbone undergoes deuterium exchange in a very short time (<10 s). This result demonstrates that significant portions of PhnG alone are highly dynamic and likely to be largely unstructured (Figure 32). The highly dynamic region of PhnG has apparently hindered previous crystallization efforts using PhnG alone. The region with the most rapid H/D exchange rate (residues 28–67) is considered to be mostly unstructured or β-strands. In the X-ray structure of $PhnG_2H_2I_2J_2$, residues 28-36 of PhnG are on a flexible loop (average B-factor 71 Å^2) while residue 37-70 of PhnG are on β-strands (average B-factor 30 Å^2).

Upon forming the $PhnG_2I_2$ complex, some highly dynamic regions of PhnG become more structured upon interaction with PhnI (Figure 32). This is particularly true for the C-terminus region of PhnG and the portion of the protein that extends from Val-47 through Leu-67. Therefore, these two regions were proposed to be located at the interface between PhnG and PhnI in the $PhnG_2I_2$ complex. The X-ray structure of $PhnG_2H_2I_2J_2$ complex revealed more details of the subunit interactions within these two regions of PhnG. The protein sequence containing residues 47-67 forms a hydrophobic hairpin-like β-strand from PhnG that inserts into PhnI (Figure 33A). The C-terminal region of PhnG, which forms an unstructured loop, interacts with the N-terminal loop of PhnI (Figure 33B). However, further differences are not observed for PhnG within the

$PhnG_2H_2I_2J_2K$ complex. Considering the identification of a cross-linked bond between Lys-89 of PhnG and the N-terminus of PhnJ, there could be a small interface between PhnG and PhnJ. In the crystal structure, the C-terminal region of PhnG interacts with PhnI and PhnJ simultaneously, which may explain the failure to identify a change in the H/D exchange level of PhnG upon forming a larger protein complex. There could also be potential interactions between the N-terminal loop of PhnJ and Lys-89 with adjacent residues of PhnG if the N-terminal loop is subject to a conformational change to reduce the distance to PhnG (Figure 28). PhnH is located about 30 Å away from PhnG so there are no interactions observed between these two subunits.

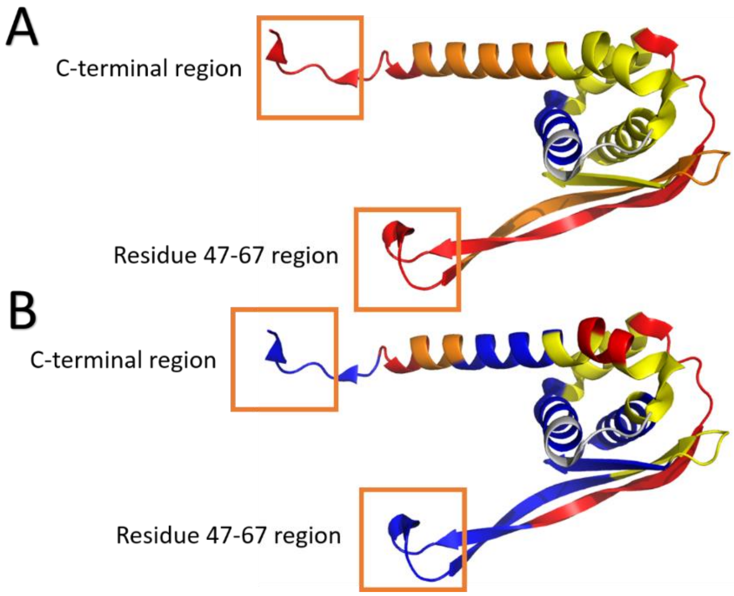


Figure 32: The differential deuterium exchange analysis of PhnG alone and PhnG within the $PhnG_2I_2$ or $PhnG_2H_2I_2J_2K$ complexes. The samples were incubated with D_2O for 30 minutes. Different H/D exchange levels are labeled in different colors. The regions with H/D exchange levels higher than 45% are colored red. The regions with H/D exchange levels from 30% to 45% are colored orange. The regions with H/D exchange levels from 15% to 30% are colored yellow. The regions with H/D exchange levels lower than 15% are colored blue. The regions without H/D exchange data are colored grey. The region containing residues 47-67 and the C-terminal region are enclosed with a square. (A) PhnG alone. (B) PhnG within $PhnG_2I_2$ or $PhnG_2H_2I_2J_2K$.

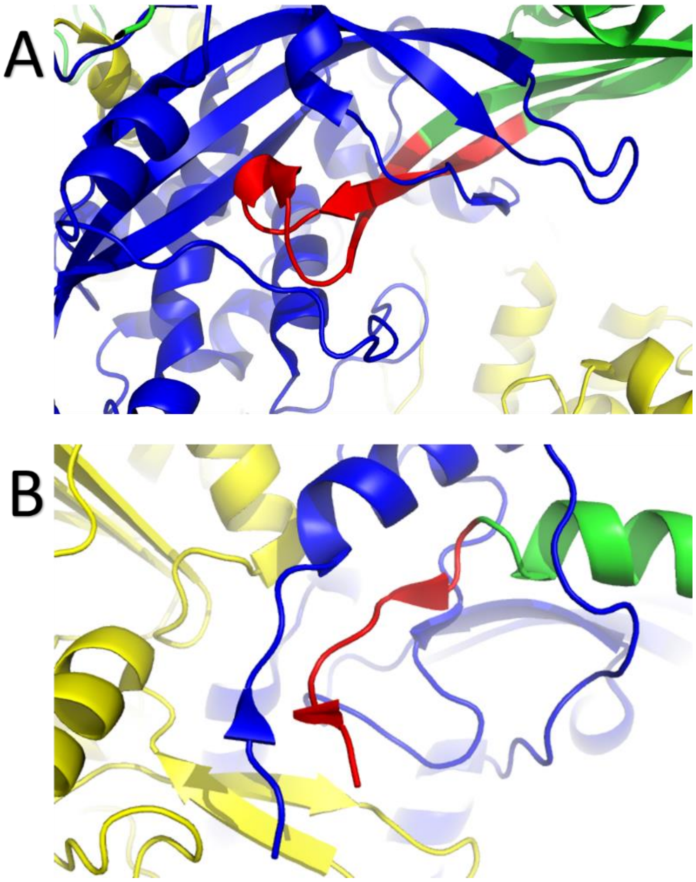


Figure 33: The interaction of PhnG with PhnI in the $\text{PhnG}_2\text{H}_2\text{I}_2\text{J}_2$ complex. PhnG is colored green, PhnI is colored blue and other subunits of the $\text{PhnG}_2\text{H}_2\text{I}_2\text{J}_2$ complex are colored yellow. (A) Residues 45-64 of PhnG interact with PhnI. Residues 45-64 of PhnG are colored red. (B) C-terminal region of PhnG interacts with PhnI. C-terminal region is colored red.

2.4.3.2 Subunit interactions of PhnI within the $PhnG_2I_2$ and $PhnG_2H_2I_2J_2K$ complex

Significant changes in the H/D exchange rates are observed for PhnI when the $PhnG_2I_2$ complex is compared with the $PhnG_2H_2I_2J_2K$ complex. In the X-ray structure, PhnJ is attached to the complex mainly through interactions with PhnI. As a result, there is a dramatic change in the H/D exchange rates for PhnI upon the formation of the $PhnG_2H_2I_2J_2K$ complex relative to the $PhnG_2I_2$ complex. Similar structural transitions are often found in multi-protein complexes and in intrinsically disordered proteins that undergo synergistic folding in the presence of their binding partners (132, 133). The interface between PhnI and PhnJ contains residues 81-86, 109-130 and the N-terminal region of PhnI. As previously discussed, the N-terminal region of PhnI simultaneously interacts with both PhnG and PhnI in the $PhnG_2H_2I_2J_2K$ complex. The H/D exchange rate for this region was already very low in the $PhnG_2I_2$ complex. It was not unexpected that a further change in the H/D exchange rate was not observed in $PhnG_2H_2I_2J_2K$ complex. There is a significant decrease in the H/D exchange rate observed for two other regions (residues 81-86 and residues 109-130) in the $PhnG_2H_2I_2J_2K$ complex, as compared to the $PhnG_2I_2$ complex. The interactions between these two regions and PhnJ are illustrated in Figure 34. Interestingly, region 109-130 is one of the most dynamic regions in the $PhnG_2I_2$ complex. In the crystal structure, this region (residues 109 to 130) is shown as a long unstructured loop (Figure 34B).

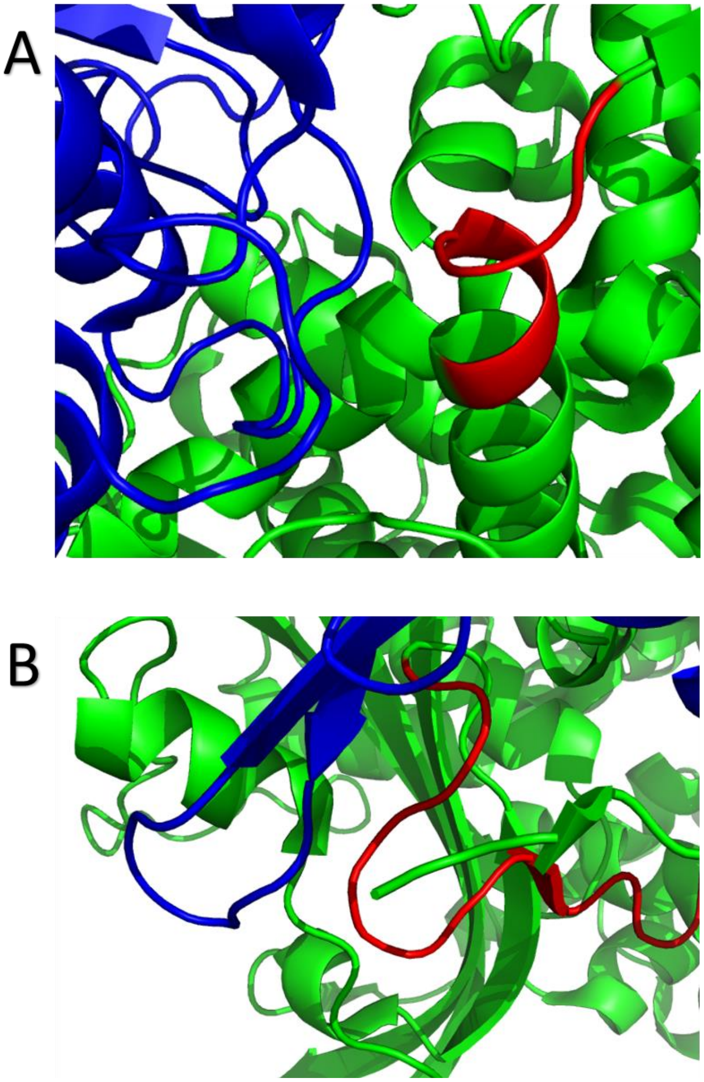


Figure 34: The regions containing residues 81-86 and 109-130 of PhnG interact with PhnJ. (A) The region containing residues 81-86 is colored red. Other residues of PhnI are colored green. PhnJ is colored blue. (B) The region containing residues 109-130 is colored red. Other residues of PhnI are colored green. PhnJ is colored blue

2.4.3.3 *Subunit interaction of PhnH within* $PhnG_2H_2I_2J_2K$ *complex*

To further characterize the interaction between PhnH and the other subunits of $PhnG_2H_2I_2J_2K$ complex, the differential deuterium exchange analysis of PhnH alone and PhnH within $PhnG_2H_2I_2J_2K$ was conducted. The peptides (^{19}RRLLKA^{24} and ^{19}RRLLKAMSEPGVI^{31}) exchange at a much slower rate in the $PhnG_2H_2I_2J_2K$ complex than in PhnH alone. The regions containing residues from Ser-110 to Ala-121, from His-166 to Asp-173 and the C-terminal region from Pro-185 to Cys-194 become more protected in the complex. PhnH alone forms a dimer in both solution and the crystal structure (61). The dimer is assembled through the interactions between two N-terminal α-helices. In the crystal structure of the $PhnG_2H_2I_2J_2$ complex, PhnH is attached to the complex through the interaction with PhnJ. The interface between PhnH and PhnJ closely resembles that observed in the X-ray structure of the PhnH homodimer. The dramatic decrease of the H/D exchange rate of all these region of PhnH, upon forming a complex, indicates that the interface between PhnH and PhnJ is larger and more compact in the complex than that between PhnH in the homodimer.

The crystal structure of the $PhnG_2H_2I_2J_2$ complex is consistent with this conclusion. All regions with significant H/D exchange differences between PhnH alone and PhnH within the complex are located at the interface between PhnH and PhnJ. The region containing residues from Arg-19 to Ile-31 is located on an N-terminal α-helix of PhnH that interacts with another α-helix of PhnJ. When compared with the N-terminal α-helices pair of the PhnH homodimer, the distance between these two α-helices is shorter in the complex (Figure 35), which indicates a more compact interaction between

PhnH and PhnJ. The PhnH regions from Ser-110 to Ala-121 and from His-166 to Asp-173 are located on two separate unstructured loops that interact with a short α-helix of PhnJ (Figure 36). However, similar interactions of these two regions are not observed in the PhnH homodimer. Part of the C-terminal region of PhnH interacts with the end of the N-terminal α-helix of PhnJ or the other PhnH monomer in the $PhnG_2H_2I_2J_2$ complex or PhnH homodimer, respectively. The distance between the C-terminal loop of PhnH and the N-terminal α-helix of the interacting subunit is around 3 Å shorter in the $PhnG_2H_2I_2J_2$ complex than that in PhnH homodimer (Figure 37). These interactions suggest that the interface between PhnH and PhnJ is larger and more compact than that in PhnH homodimer.

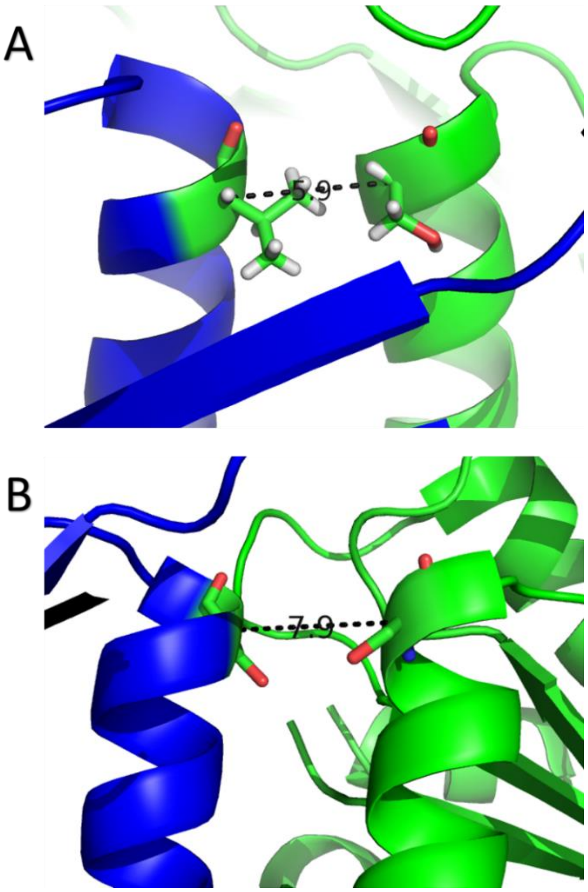


Figure 35: The interaction between two N-terminal α-helices at the interface in PhnG$_2$H$_2$I$_2$J$_2$ complex and PhnH homodimer. (A) The interface between PhnH and PhnJ in PhnG$_2$H$_2$I$_2$J$_2$ complex. PhnH and PhnJ are colored green and blue, respectively. The distance, 5.9 Å, is measured between two α-carbons of Ser-26 of PhnH and Val-28 of PhnJ. (B) The interface of the PhnH homodimer. One PhnH monomer is colored green and the other one is colored blue. The distance, 7.9 Å, is measured between two α-carbons of Ser-26 of both monomers.

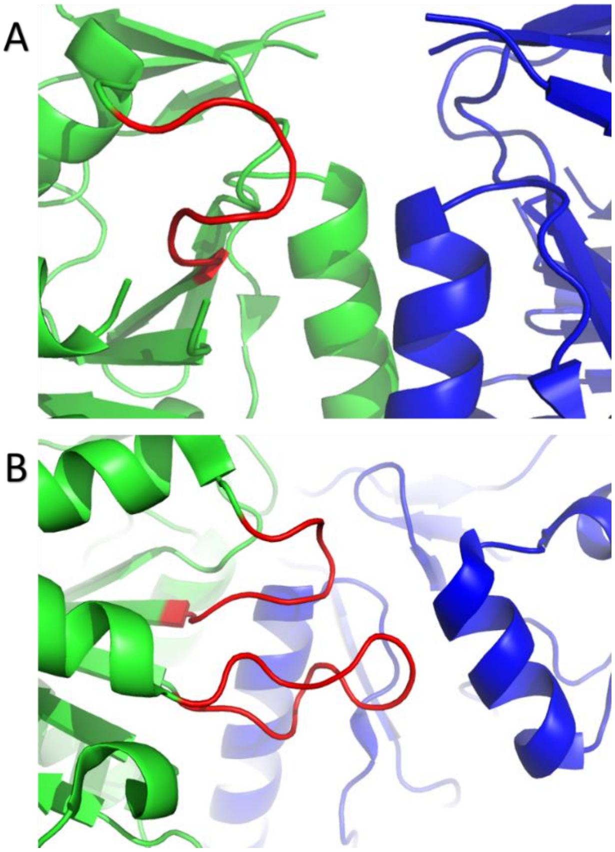


Figure 36: The regions containing residues 110-121 and residues 166-173 of PhnH interact with the other PhnH monomer or PhnJ in the PhnH homodimer and the $PhnG_2H_2I_2J_2$ complex, respectively. (A) PhnH homodimer. The region containing residues 166-173 is colored red. The region containing residues 110-121 is not included in the crystal structure of the PhnH homodimer. Other residues of this monomer are colored green. The other monomer in PhnH homodimer is colored blue. (B) $PhnG_2H_2I_2J_2$ complex. PhnH and PhnJ are colored green and blue, respectively. The regions containing residues 110-121 and residues 166-173 of PhnH are colored red.

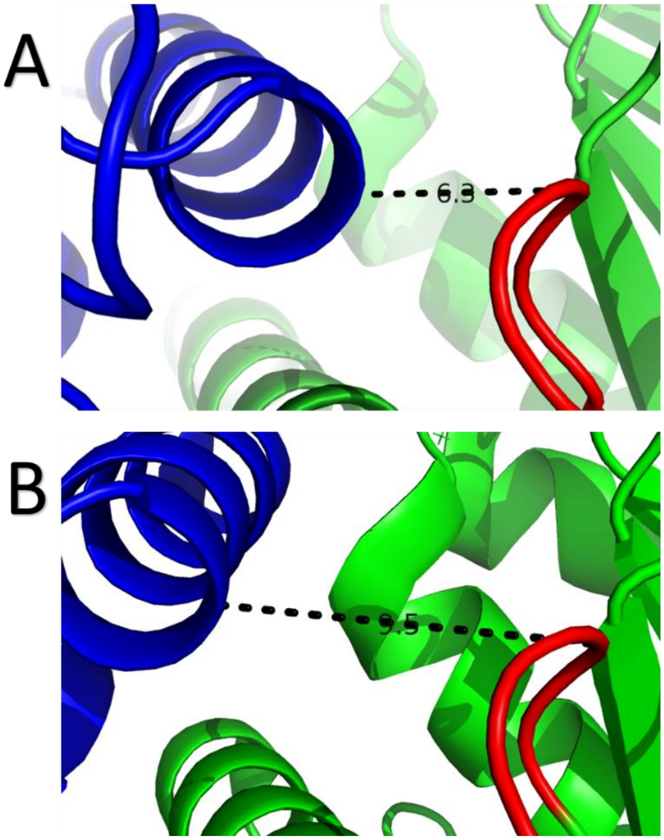
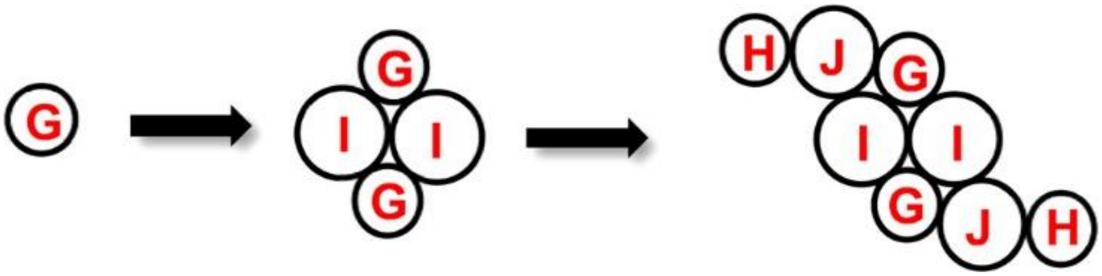


Figure 37: The C-terminal regions of PhnH interact with the other PhnH monomer or PhnJ in the PhnH the homodimer and the $PhnG_2H_2I_2J_2$ complex. (A) $PhnG_2H_2I_2J_2$ complex. PhnH and PhnJ are colored green and blue, respectively. The C-terminal region of PhnH is colored red. The distance, 6.3 Å, is measured between the α-carbons of Arg-187 of PhnH and Ala-29 of PhnJ. (B) PhnH homodimer. C-terminal region of PhnH is colored red. Other residues of this monomer are colored green. The other monomer in PhnH homodimer is colored blue. The distance, 9.5 Å, is measured between the α-carbons of Arg-187 of one PhnH monomer and the α-carbons of Ser-26 of the other PhnH monomer.

2.4.4 Low resolution structural model

The results obtained from the mass spectrometry and chemical cross-linking experiments were used to propose a low-resolution model for the assembly of the C–P lyase complex (Scheme 11). In the crystal structure of $PhnG_2H_2I_2J_2$, the C–P lyase core complex resembles the letter 'H' with rounded arms that are twisted approximately 45° in and out of the plane with respect to each other (127). The arms are composed on opposing sides by the two PhnG molecules and on the other sides by tight complexes between PhnJ and PhnH, which confirmed the interactions between PhnJ and PhnG (Figure 38) (127). At the center of the molecule, a compact PhnI homodimer forms a disc-like structure that serves as a central hub for attachment of the other subunits, which agrees with the $PhnG_2I_2$ core in the low resolution structural model. At the C- and N-termini there are helical extensions of the core domain of PhnI, approximately 35 residues that grasp PhnJ and tether it to the complex via extensive interactions, which is consistent with the identified interactions between PhnI and PhnJ. In turn, PhnJ attaches PhnH to the complex through packing of conserved α-helices in both proteins, which matches the fragment of PhnHJ in MS/MS spectrum.

The closest sequence homologues for PhnG, PhnI, and PhnJ were found to be the putative Trp repressor protein (PDB id: 3KOR), carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (PDB id: 2D4E), and PhnH (PDB id: 2FSU), respectively. However, the sequence identities of these proteins to their closest identified homologues are less than 20%, and so a reliable homology model could not be built for any of these proteins. It was quite interesting that the closest structural model for PhnJ is PhnH. In the X-ray structure of Phn$G_2H_2I_2J_2$, the core folds of PhnJ and PhnH are nearly identical. Moreover, the interface between PhnH and PhnJ closely resemble that observed in the previously published X-ray structure of the PhnH homodimer (Figure 39).



Scheme 16: Low resolution assembly model for C-P lyase complex

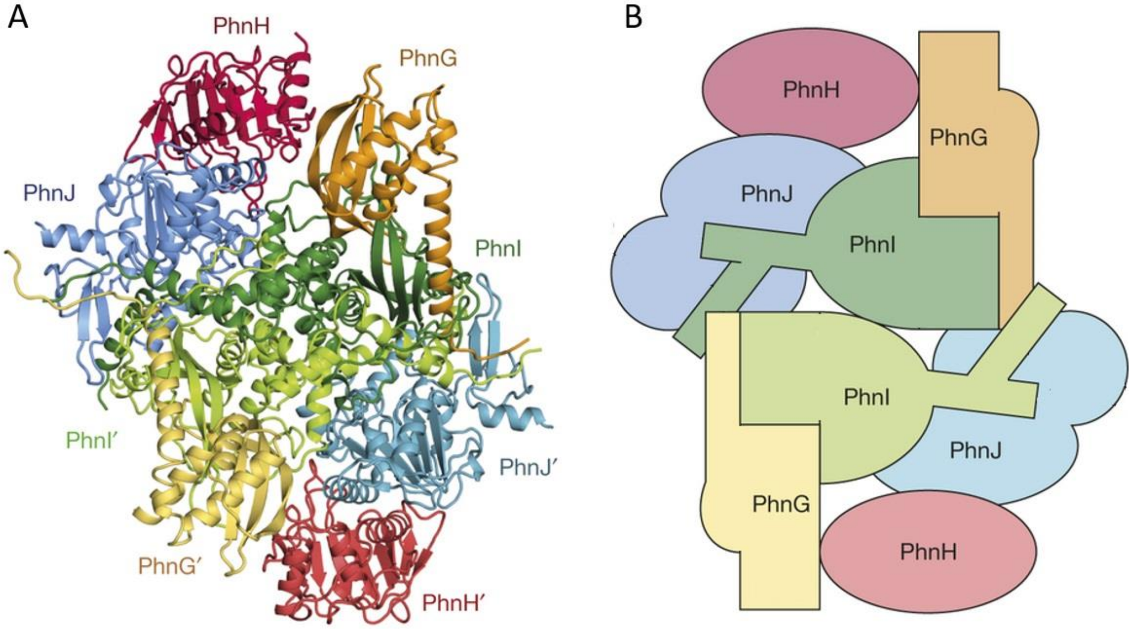


Figure 38: (A) Overall structure of the 240 kDa C–P lyase core complex. (B) Schematic architecture of the complex with structural domains indicated (127).

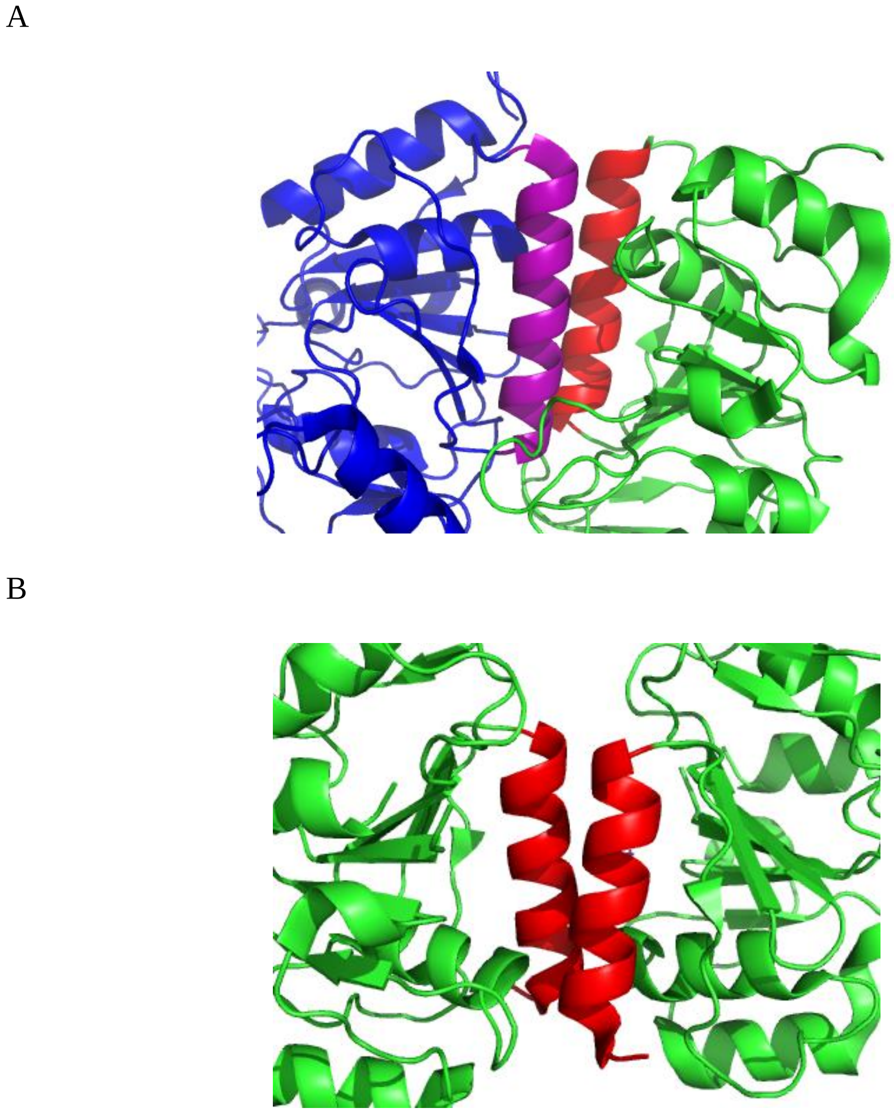


Figure 39: The interface between PhnH and PhnJ closely resembles the PhnH homodimer. (A) The interface between PhnH and PhnJ in the $PhnG_2H_2I_2J_2$ complex. The interacting α-helices of PhnH and PhnJ are colored red and purple, respectively. PhnH is colored green and PhnJ is colored blue. (B) The interface of the PhnH homodimer. The interacting α-helices of both PhnH monomer are colored red.

2.4.5 Cryo-electron microscopy structural model

We have demonstrated a unique binding mode of the NBD-like PhnK to the C-P lyase core complex, $PhnG_2H_2I_2J_2$. Even though there exist two identical binding sites for PhnK within the core complex, only one PhnK can bind at a time, leading to a more exposed proposed active site residue Gly32 from PhnJ on the PhnK bound side (Figure 40). The proposed glycyl radical formed at Gly32 was postulated to abstract the hydrogen of Cys272 to form a thiyl radical during PhnJ catalysis (53). Such rearrangement around the active site may facilitate the delivery of the substrate or release of the reaction product.

However, Seweryn *et al* made the discovery that Gly32 of PhnJ is more than 30 Å away from Cys272 in the crystal structure of the $PhnG_2H_2I_2J_2$ complex. In our cryo-EM structure of $PhnG_2H_2I_2J_2$ with PhnK bound, the overall structure of PhnJ remains similar to the crystal structure with Gly32 still ~30 Å away from Cys272. If the proposed reaction mechanism for C-P lyase, which involves the direct interaction between Cys272 and Gly32, is correct, then substantial structural rearrangements must occur transiently and/or after the binding of substrates or other protein subunits to this complex.

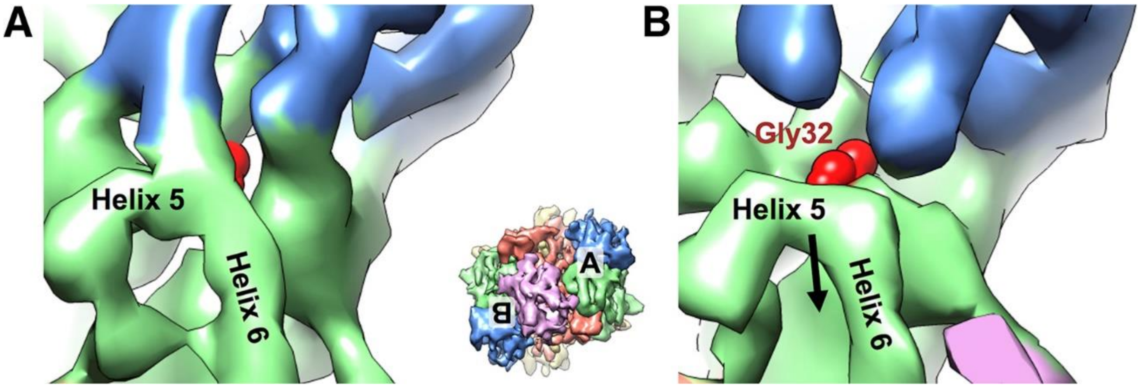


Figure 40: Binding of PhnK to the $PhnG_2H_2I_2J_2$ complex causes Gly32 of PhnJ to become more exposed (A) On the side of the complex without PhnK bound (apo side), helix 5 of PhnJ interacts with PhnH. (B) On the side with PhnK bound (bound side), the densities for helix 5 and helix 6 of PhnJ move away from PhnH to more fully expose Gly32 from PhnJ. The density map of $PhnG_2H_2I_2J_2K$ is rendered at 4.5 σ above the mean and colored coded as follows, yellow: PhnG, blue: PhnH, green: PhnJ, orange: PhnI, lavender: PhnK. The red sphere in the model denotes Gly32 from PhnJ. Inset shows the map of the $PhnG_2H_2I_2J_2K$ complex in the same orientation as Figure 18E with the locations of (A) and (B) labeled**.**

PhnK has all of the structural motifs to bind ATP. These interactions may help recruit ATP as the substrate for the C-P lyase reaction. While the two-fold symmetrical C-P lyase core complex $PhnG_2H_2I_2J_2$ potentially provides two equivalent sets of active sites for two series of reactions, the loose and monomeric binding of PhnK suggests that this subunit may alternate between the two binding sites on the core complex to deliver ATP and facilitate the series of reactions in a staggered time frame. Further biochemical and structural studies are needed to test these hypotheses. The structural information provided here may provide a better understanding of the enzymatic reaction mechanism for phosphonate utilization, especially the function of the NBD-like PhnK. These results clearly demonstrate the power of cryo-EM for the structural analysis of small (~200 kDa) asymmetrical multi-subunit protein complexes with compositional heterogeneity.

In summary, we have used high-resolution mass spectrometry, chemical cross-linking, hydrogen/deuterium exchange, analytical ultracentrifugation, N-terminal protein sequencing and cryo-electron microscopy to characterize the stoichiometry, revealed subunits interactions and construct structural models for the $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$ C–P lyase complexes from *E. coli.* The interactions between the monomeric PhnK and the dimeric $PhnG_2H_2I_2J_2$ complex was revealed. The proposed stoichiometry, subunit interaction and low resolution structural model of $PhnG_2H_2I_2J_2$ complex agreed quite well with the crystal structure that was published later.

2.4.6 Future perspectives

Based on previous genetic and biochemical studies, seven proteins are essential for the cleavage of the C-P bond. Five out of these seven proteins are able to form multi-subunit complexes, $\mathrm{PhnG_2I_2}$, $\mathrm{PhnG_2H_2I_2J_2}$ and $\mathrm{PhnG_2H_2I_2J_2K}$. The determination of the high resolution structural model of $\mathrm{PhnG_2H_2I_2J_2K}$ and $\mathrm{PhnG_2H_2I_2J_2}$ protein complex provides a basis for the greater understanding of the physiological functions of these complexes. However, as essential proteins in the proposed pathway for the hydrolysis of C-P bonds, PhnL and PhnM are not components in the structurally determined C-P lyase complexes. The marginal solubility of these two proteins *in vitro* implies that other proteins including PhnL and PhnM may form a larger protein complex to cleave the C-P bond. It is known that PhnL can work with the subunits PhnG, PhnH, PhnI to catalyze the initial reaction within the C-P lyase pathway. However, the specific role of each protein is unknown. It is possible that each protein has an active site performing a distinct step of the initial reaction. An alternative explanation is that only some of the subunits have catalytic functions while others are responsible for scaffold and regulatory roles. The elucidation of the structure of the C-P lyase protein complex with substrates, inhibitors and cofactors bound to the active sites will be helpful to shed light on these questions.

Despite attempts to reconstitute the C-P lyase complexes catalyzed reactions *in vitro*, the purified C-P lyase complexes are still catalytically incompetent. It may require additional subunits or factors to activate these complexes. Further genetic and biochemical studies such as protein-protein interaction profiling and mass spectrometry based metabolomics will provide powerful tools to find the critical missing part of the C-P lyase protein complex. PhnJ, a core subunit of $PhnG_2H_2I_2J_2$ or $PhnG_2H_2I_2J_2K$ complex, is apparently responsible for cleavage of the C-P bond. Based on the proposed reaction mechanism of PhnJ, a reduced Fe-S cluster cleaves *S*-Adenosyl methionine to generate a 5’-deoxyadenosyl radical intermediate. A hydrogen atom on Gly-32 is then abstracted by the radical intermediate to form a glycyl radical. The radical then transfers from Gly-32 to Cys-272 generating a thiyl radical capable of cleaving the C-P bond of PRPn (45). The participation of Gly32 in the radical transferring process was confirmed by deuterium labeled mass spectrometry. However, Gly32 is 30 Å away from Cys-252 and the proposed Fe-S cluster, which implies that PhnJ must go through a dramatic conformational change during the catalytic cycle or there are other unidentified steps in the radical transferring process. Further studies including fluorescence resonance energy transfer (FRET) analysis and the elucidation of the crystal structure of the complex bound with Fe-S clusters will be helpful to answer these issues.

CHAPTER III

MECHANISTIC CHARACTERIZATION OF THE PHOSPHOTRIESTERASE FROM *SPHINGOBIUM SP*. STRAIN TCM1[*]

3.1 Introduction

Phosphotriesters are a group of organophosphates that were initially developed during World War II. This group of compounds was first utilized as insecticides for crop protection (63). The structure of this group of compounds is a phosphate center with three ester linked functional groups. In recent years, a related group of organophosphotriesters has gained popularity for use as plasticizers and flame retardants in lubricants, durable plastics, and foams (65, 66). Compounds such as tris(2-chloroethyl) phosphate (compound 5 in Scheme 12) (65) and triphenyl phosphate (66) do not contain a labile ester bond but rather make use of the stable nature of these compounds to inhibit combustion and improve the performance of other organic

---

[*] Part of this chapter is reprinted with permission from "Interrogation of the Substrate Profile and Catalytic Properties of the Phosphotriesterase from *Sphingobium sp*. Strain TCM1: An Enzyme Capable of Hydrolyzing Organophosphate Flame Retardants and Plasticizers" by Dao Feng Xiang, Andrew N. Bigley, Zhongjie Ren, Haoran Xue, Kenneth G. Hull, Daniel Romo, and Frank M. Raushel, *Biochemistry* **2015** *54* (51), 7539-7549, Copyright 2016 by American Chemical Society.

Part of this chapter is reprinted with permission from "Chemical Mechanism of the Phosphotriesterase from *Sphingobium sp*. Strain TCM1, an Enzyme Capable of Hydrolyzing Organophosphate Flame Retardants" by Andrew N. Bigley, Dao Feng Xiang, Zhongjie Ren, Haoran Xue, Kenneth G. Hull, Daniel Romo, and Frank M. Raushel, *Journal of the American Chemical Society* **2016** *138* (9), 2921-2924, Copyright 2016 by American Chemical Society.

compounds. While not neurotoxins, many of the flame retardants contain halogenated alkyl or aromatic esters and are known, or suspected, chemical carcinogens (66). The contamination of the environment by organophosphorus flame retardants has become widespread, and these compounds are now considered to be major emerging pollutants (3).

In various species of bacteria, natural and synthetic organophosphotriesters are hydrolyzed by phosphotriesterase (PTE) (68-71). To utilize the man-made organophosphotriesters as phosphorus source in the environment, bacteria evolved the ability to hydrolyze this type of compounds. The structural similarity of the phosphoryl centers between synthetic and natural substrates may play an essential role for the development of the promiscuous activity towards synthetic organophosphortriesters (72). Due to the lack of a sufficiently labile leaving group, currently identified phosphotriesterases have little or no activity against organophosphorus flame retardants and plasticizers (88, 102-104). Two novel phosphotriesterase-like enzymes from *Sphingomonas sp.* strain TDK1 (Sm-PTE) and *Sphingobium sp.* strain TCM1 (*Sb*-PTE) have recently been reported to hydrolyze organophosphates that are not substrates for other phosphotriesterases, such as *Pd*-PTE or OpdA (103). When this study was initiated, very little was known about these enzymes aside from their amino acid sequence and their dependence on the addition of divalent metal ions for catalytic activity. In addition, there was no structural or mechanistic information available for either *Sm*-PTE or *Sb*-PTE, or any close homologue of known function.

Due to the lack of a sufficiently labile leaving group, currently identified phosphotriesterases have little or no activity against organophosphorus flame retardants and plasticizers (88, 102-104). Two novel phosphotriesterase-like enzymes from *Sphingomonas sp.* strain TDK1 (Sm-PTE) and *Sphingobium sp.* strain TCM1 (*Sb*-PTE) have recently been reported to hydrolyze organophosphates that are not substrates for other phosphotriesterases, such as *Pd*-PTE or OpdA (103). When this study was initiated, very little was known about these enzymes aside from their amino acid sequence and their dependence on the addition of divalent metal ions for catalytic activity. In addition, there was no structural or mechanistic information available for either *Sm*-PTE or *Sb*-PTE, or any close homologue of known function.

Structural models of PTEs from various species reveal different protein folding patterns of PTEs including TIM-barrel, β-lactamase, "pita-bread", and β-propeller folds (73-77). Despite the varied protein structural folds, PTEs are known to possess divalent metal center(s) to bind and activate substrates for hydrolysis (71, 74, 78, 79). PTEs with different folding patterns hydrolyze the organophosphotriesters via different chemical mechanisms.

Due to the lack of a sufficiently labile leaving group, currently identified phosphotriesterases have little or no activity against organophosphorus flame retardants and plasticizers (88, 102-104). Two novel phosphotriesterase-like enzymes from *Sphingomonas sp.* strain TDK1 (Sm-PTE) and *Sphingobium sp.* strain TCM1 (*Sb*-PTE) have recently been reported to hydrolyze organophosphates that are not substrates for other phosphotriesterases, such as *Pd*-PTE or OpdA (103). When this study was

initiated, very little was known about these enzymes aside from their amino acid sequence and their dependence on the addition of divalent metal ions for catalytic activity. In addition, there was no structural or mechanistic information available for either *Sm*-PTE or *Sb*-PTE, or any close homologue of known function.

3.2 Materials and Methods

3.2.1 Materials

Unless otherwise noted, the bacterial growth medium was obtained from Research Products International and general chemicals were purchased from Sigma/Aldrich. Paraoxon is synthesized as previously reported (96). The C18 reverse phase column was purchased from Waters Corporation. O^{18} labeled water was obtained from Cambridge Isotope Laboratories. Protein Sb-PTE was expressed and purified by Dr. Dao Feng Xiang.

3.2.2 Inactivation of *Sb*-PTE by 1-butynyl dibutylphosphate

All substrate profile, kinetic parameters and optimal metal determination of *Sb*-PTE were determined by Dr. Dao Feng Xiang. To assess the interaction of Sb-PTE with 1-butynyl dibutylphosphate, the enzyme (32 nM) was mixed with this compound in various molar ratios ranging from 0 to 37.5 in 50 mM CHES buffer (pH 9.0). The residual catalytic activity of Sb-PTE was assayed using 0.40 mM paraoxon as the substrate. Samples of the fully inhibited enzyme were subjected to MS analysis to determine the number of modifications on the intact protein in the presence of excess inhibitor. For reactivation studies, 500 µL samples of the inhibited enzyme were prepared by mixing 1.3 µM enzyme with 30 µM 1-butynyl dibutylphosphate in 50 mM HEPES (pH 7.0). The enzyme was fully inhibited within 10 minutes. Reactivation of the inhibited enzyme was tested by diluting samples 1:2 into a reactivation mixture containing 50 mM CHES (pH 9.0) and hydroxylamine (0–20 mM). The reactivation of catalytic activity was followed over time by measuring the enzymatic activity using 0.60 mM paraoxon as the substrate. Reactivation of Sb-PTE was also examined, in the absence of added hydroxylamine, by variation of the pH from 7 to 10.

3.2.3 MS/MS analysis of inactivated *Sb*-PTE

Samples of *Sb*-PTE, which had been inactivated by 1-butynyl dibutylphosphate, and a control sample were loaded on SDS–PAGE gels at a concentration of 1.5 mg/mL and visualized by Coomassie Blue staining. The samples were proteolyzed via the addition of trypsin using a previously described protocol (134). The samples were analyzed by LC–MS (Dionex nanoRSLC and an Orbitrap Fusion Tribrid Mass Spectrometer from Thermo Scientific). Briefly, 1.0 µL of the digested protein sample was loaded onto a 75 µm × 15 cm Acclaim Pepmap C18 column and eluted over 30 minutes with a total run time of 1 hour. The eluted peptides were introduced into the mass spectrometer and analyzed using the Orbitrap (OT) mass spectrometer in positive ion mode with the following settings: resolution, 120 K; scan range, m/z 200–1800;

maximal injection time, 50 ms; AGC target, 2 × 104; S-lens RF level, 60. Charge states ranging from 2 to 7 were accepted, and targeting ions were excluded for 60 s. MS/MS data were obtained using the HCD cell, and fragment masses were determined using the OT mass spectrometer at a resolution of 30 K. The data were analyzed using PEAKS Studio Software (Thermo Scientific). For the analysis of peptide MS and MS/MS from the inhibitor modification experiments, maximal mass deviations of 10 ppm for MS and 0.5 D for MS/MS were accepted. Additionally, carbamidomethylation of cysteine was taken into account as a fixed modification, and oxidation of methionine was set as a varied modification. Up to three missed cleavage sites were utilized during data processing.

3.2.4 ^{18}O-labeling of *Sb*-PTE protein

50 µL samples of 220 µM *Sb*-PTE were dried *in-vacuo* at room temperature. Similarly, samples of 20 µL 0.5 M CHES (pH 9) and, separately, 66 µL of 8.4 mM paraoxon were dried to completion. Samples were then reconstituted and combined in a total of 120 µL of ^{18}O labeled or unlabeled water as a control. The samples were incubated at room temperature until the hydrolysis of the paraoxon was complete (~2 hr). The protein was run on SDS-PAGE gel and the appropriate bands were digested in-gel with trypsin prior to MS analysis as previously described (54).

3.2.5 ^{18}O-labeling of reaction products by *Sb*-PTE

^{18}O incorporation into the reaction products of *Sb*-PTE were tested under single and multi-turnover conditions. 100 µL reactions were in conducted in 60% ^{18}O labeled water with 20 mM NH_4HCO_3 (pH 9.0). For single turnover conditions 50 µM paraoxon was used with 150 µM *Sb*-PTE. For multiple turnover conditions, 500 µM paraoxon was used with 50 µM *Sb*-PTE. Reactions were run overnight and the protein filtered using a Viaspin 500 centrifugal filtration device (GE Healthcare). Reaction products were analyzed by mass spectrometry using a Thermo Scientific Q Exactive mass spectrometer. The samples were applied to the mass spectrometer by direct injection at a flow rate of 200 µL/min. A blank solution of ammonium bicarbonate was utilized between each sample to avoid peak carry-over between runs. The mass spectrometer was calibrated using a Pierce™ Negative Ion Calibration Solution from Thermo Scientific, USA. The results were analyzed using software Xcalibur (Thermo Scientific).

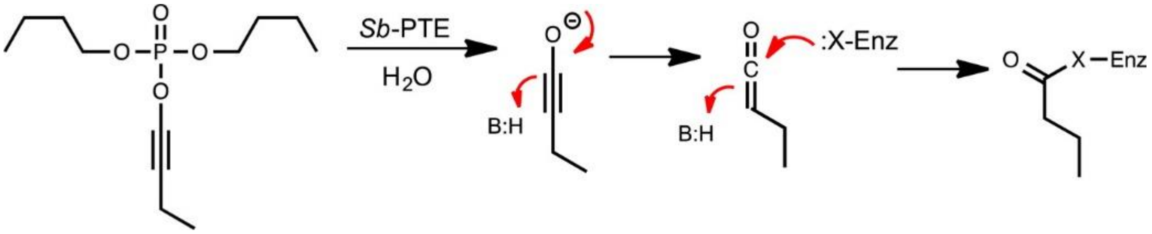
3.3 Results

3.3.1 Inactivation of *Sb*-phosphotriesterase

1-butynyl dibutylphosphate was synthesized and tested as a mechanism-based inactivator of *Sb*-PTE. The enzymatic hydrolysis of this compound is expected to generate a highly reactive ketene intermediate, which can potentially react with available nucleophiles in the active site as illustrated in Scheme 17. The alkynyl phosphate ester was found to be a highly effective inhibitor of *Sb*-PTE with nearly complete inhibition achieved with near stoichiometric amounts of inhibitor (Figure 41). The mass spectrum of the modified protein exhibited two major additional peaks that are shifted by ∼71 and ∼142 Da, relative to the unlabeled protein, and two minor peaks. This result is consistent with the addition of up to four ketene groups to the protein (Figure 42B).

Reactivation of the enzyme by the addition of hydroxylamine and changes in pH were attempted. Incubation of the inhibited enzyme with 5.0 mM hydroxylamine was found to restore approximately 60% of the enzymatic activity in 3 h, and >90% of the original catalytic activity was restored after 24 h. A similar reactivation profile occurs after incubation of the inhibited enzyme at pH 10 (data not shown). Analysis of the MS/MS spectra of the inhibited protein identified residues Tyr-219, Ser-474, Tyr-456, and Ser-458 as being modified by the mechanism-based inactivator (Figure 43).

The addition of the substrate reduced the rate of inactivation of *Sb*-PTE by the suicide substrate. The ability of the substrate to protect the enzymatic activity was tested by changing the order of addition of the substrate and inhibitor in reaction mixtures containing 50 mM CHES (pH 9.0) at 30 °C with 2.0 mM paraoxon as the substrate. The addition of paraoxon to a premixture of *Sb*-PTE (10 nM) and 1-butynyl dibutylphosphate (20 nM) resulted in a reaction rate that was 13% of the uninhibited control reaction rate. The addition of 1-butynyl dibutylphosphate (20 nM) to a premixture of *Sb*-PTE (10 nM) and the substrate paraoxon (2.0 mM) gave a reaction rate that was 62% of the uninhibited control reaction rate. These data demonstrated that the substrate paraoxon can protect *Sb*-PTE against inactivation by the suicide substrate 1-butynyl dibutylphosphate.



Scheme 17: Covalent modification of TCM1 by 1-butynyl dibutylphosphate

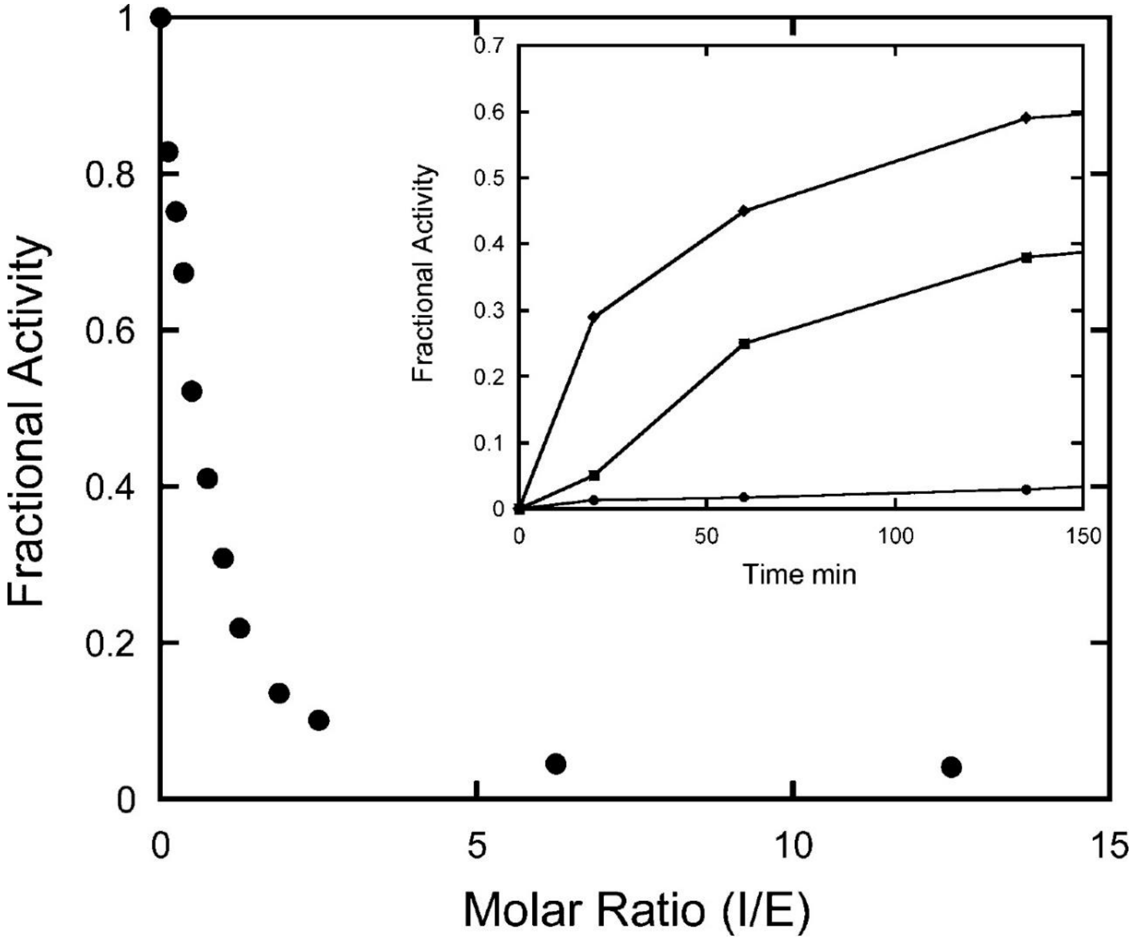


Figure 41: Inactivation of *Sb*-PTE by 1-butynyl dibutylphosphate. Various amounts of the mechanism-based inactivator were added to 32 nM *Sb*-PTE in CHES buffer (pH 9.0). Aliquots were removed after 10 minutes and tested for catalytic activity using 0.40 mM paraoxon at pH 9.0. The inset shows the reactivation of enzyme activity after the addition of hydroxylamine [(●) 0, (■) 1.0, and (◆) 5.0 mM] at pH 7.0. The activity of the rescued enzyme was assayed with 0.6 mM paraoxon in 50 mM CHES (pH 9.0).

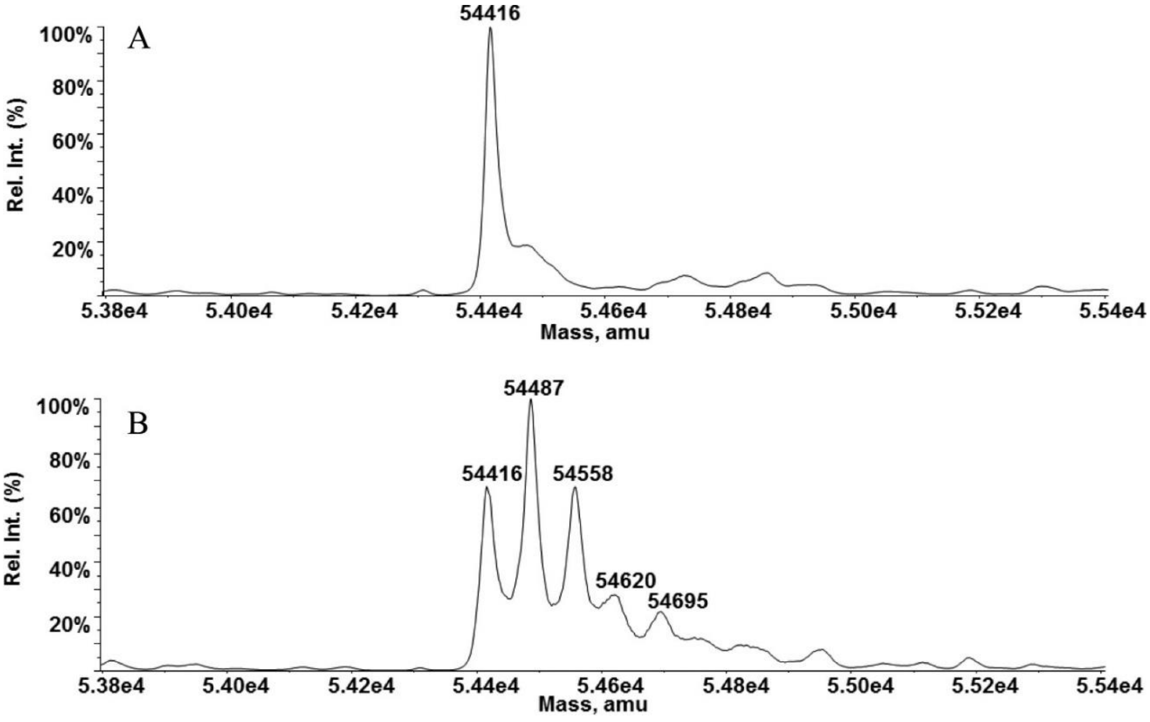


Figure 42: MS of (A) intact *Sb*-PTE and (B) 1-butynyl dibutylphosphate-inactivated *Sb*-PTE. Masses of major species are listed with the labeled protein showing addition of up to four butyl groups.

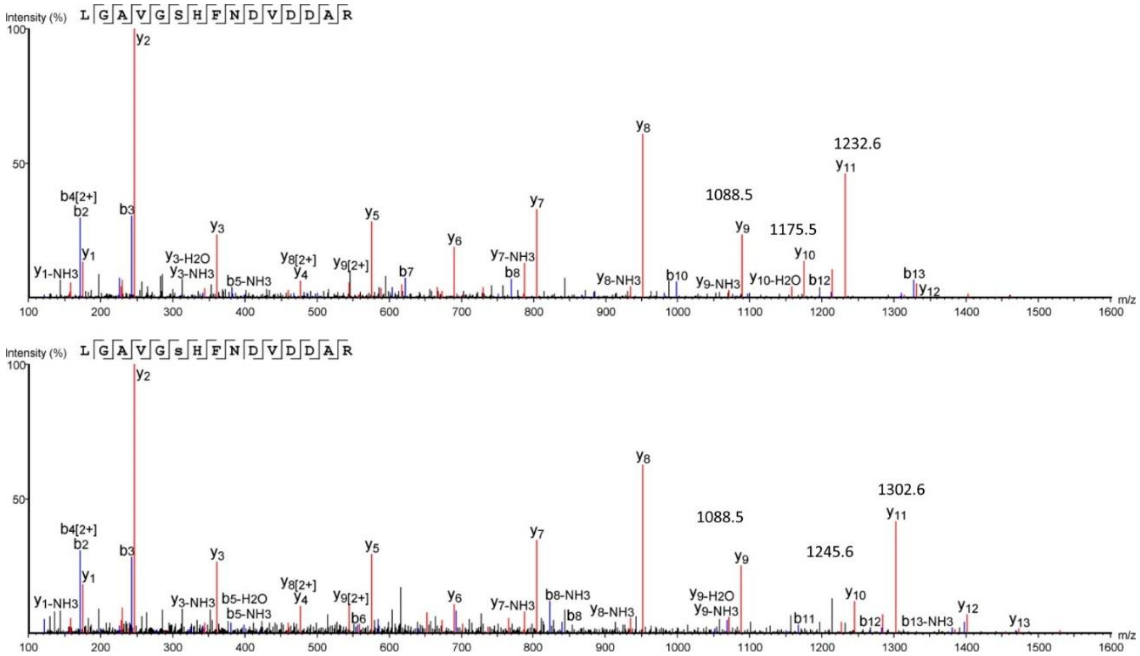


Figure 43: MS/MS of the peptide from residue Leu-469 to Arg-483 of *Sb*-PTE. The top panel shows the spectrum for the unmodified protein, while the bottom panel shows the protein after inactivation by 1-butynyl dibutylphosphate. Modification is identified at Ser-474 (peak y_{10}). Masses for peaks y_9–y_{11} are provided in the plot.

3.3.2 ^{18}O-labeling of *Sb*-PTE protein and its reaction products by *Sb*-PTE

Peptide coverage for the control reaction was 85% while coverage for the ^{18}O experiment was 87% (Table 4). In both cases the peptide containing the proposed active site aspartate (Asp-478) was observed. No peptide was identified that had an M+2 or M+4 peak that differed significantly from that of the unlabeled control experiment. The only peptides containing either glutamate or aspartate that were not identified by mass spectrometry were ^{181}DPR^{183},^{373}SQPKMRLVGEAK^{384}, and ^{499}IVDLR^{503}.

When the hydrolysis reaction was conducted in unlabeled water the *p*-nitrophenolate was isolated with an *m*/*z* of 138 and diethyl phosphate at an *m*/*z* of 153 (Figure 44a). However, when the hydrolysis reaction was conducted under multiple turnover conditions in ∼60% ^{18}O-labeled water, the *p*-nitrophenolate was produced with an *m*/*z* of 138, whereas the diethyl phosphate was isolated as a mixture of species with *m*/*z* values of 153 and 155 (Figure 44b). When we utilized a 3-fold molar excess of *Sb*-PTE to hydrolyze paraoxon in ∼60% ^{18}O-labeled water, the isotopic content of the recovered diethyl phosphate was essentially the same as the solvent (Figure 44c).

Table 4: Predicted and observed masses for peptides found from *Sb*-PTE.

Peptide	**Predicted Mass**	**Observed Mass ^{16}O**	**Observed Mass ^{18}O**
88-QVVIGPGDRPETGLQGQTTIEDVVSGRS-114	2808.44	2808.44	2808.44
125-RLVGRTDIWNRG-134	1229.68	1229.67	1229.67
129-RTDIWNRG-134	804.39	804.39	804.40
135-RGGNLQLSWVDQC(Carbamidomethyl)AYVSTFKQ-153	2173.02	2173.04	2173.03
154-KQAGPITANSRS-163	1014.52	1014.53	NO
164-RSALFLRE-169	706.42	706.42	706.42
170-REPAGVAVIDVRD-180	1125.62	1125.62	1125.62
184-RAPKPVRL-189	667.42	667.42	667.42
190-RLLRD-192	401.28	NO	401.28
193-RDRGSIDAVETMHAIAAPGRK-211	1966.97	1966.97	1966.98

Table 4 (Continued)

Peptide	**Predicted Mass**	**Observed Mass ^{16}O**	**Observed Mass ^{18}O**
195-RGSIDAVETMHAIAAPGRKV-212	1823.95	1823.94	NO
227-RGEEDAAWLSIYDASNC(Carbamidomethyl)LNPKL-246	2253.00	2253.01	2253.01
247-KLQSEFKW-252	751.39	751.39	751.39
268-RRVYGTEVVPGLGSGKG-282	1518.81	1518.82	NO
269-RVYGTEVVPGLGSGKG-282	1362.72	1362.72	1362.72
283-KGGLHVLDISDMKR-294	1284.65	1284.66	1284.66
298-RYLGRF-301	508.28	508.28	508.28
302-RFGVTRPNGLTAGFTPHEVSISHDERR-326	2724.34	2724.34	2724.34

Table 4 (Continued)

Peptide	**Predicted Mass**	**Observed Mass ^{16}O**	**Observed Mass ^{18}O**
385-KQGGFHSVVPASINGVPHLVGAAELGAC(Carbamidomethyl)PGTWPRI-417	3338.66	3338.68	3338.68
418-RIINIADEKN-425	915.51	915.51	915.51
429-KIVGEFKL-434	692.39	692.39	692.39
435-KLQMNIKE-440	746.42	746.42	746.42
441-KENC(Carbamidomethyl)DAIRF-447	877.38	877.38	877.38
448-RFTPRK-451	520.28	520.28	520.28
452-RKEDPYASFIPIPDITARL-468	1932.99	1933.00	1933.00
469-RLGAVGSHFNDVDDARN-483	1572.74	1572.74	1572.74
487-RLGLFPFFAGGVRI-498	1280.71	1280.71	1280.71

Table 4 (Continued)			
Peptide	**Predicted Mass**	**Observed Mass ^{16}O**	**Observed Mass ^{18}O**
504-RDPTKPVEVGYYKPGANPDTPLSGNGLNWTGLNDQVTDGC(Carbamidomethyl)MSHVRY-547	4757.21	4757.23	4757.23
548-RYVPESGHIWFAC(Carbamidomethyl)VTTGFHVVELNPDLRA-574	3143.53	3143.54	3143.54
577-RLGFPTVKL-583	761.45	761.45	761.45
584-KLEHHHHHH-591	1083.50	1083.49	1083.49
M+2 Peptides observed			
441-KENC(Carbamidomethyl)DAIRFTPR-451	1380.65	1380.65	NO
453-KEDPYASFIPIPDITARLGAVGSHFNDVDDAR-483	3360.63	3360.63	3360.63

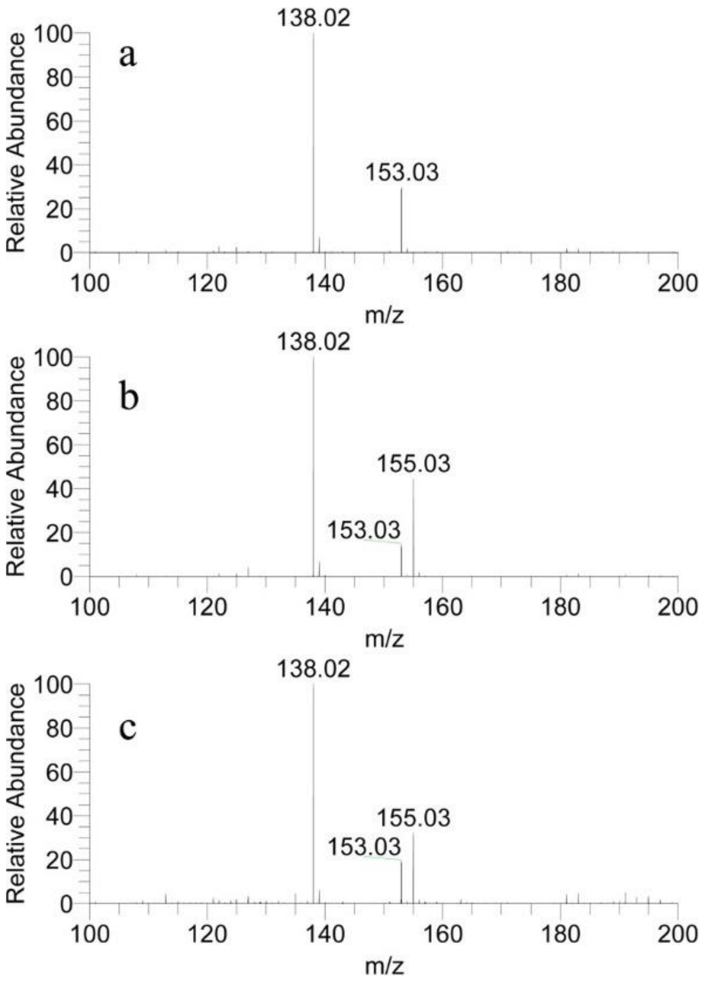


Figure 44: Mass spectra of *p*-nitrophenol and diethlyphosphate from the *Sb*-PTE catalyzed hydrolysis of paraoxon. (a) Control reaction conducted in unlabeled water. (b) Multiple turnover conditions where 500 μM paraoxon was hydrolyzed with 50 μM *Sb*-PTE in ~60% ^{18}O-labeled water. (c) Single turnover conditions in ~60% ^{18}O-labeled water where 50 μM paraoxon was hydrolyzed with 150 μM *Sb*-PTE. The peaks identified are unlabeled *p*-nitrophenolate (*m*/*z* = 138.02), unlabeled diethylphosphate (*m*/*z* = 153.03), and ^{18}O-labeled diethylphosphate (*m*/*z* = 155.03).

3.4 Discussion

3.4.1 X-ray crystal structure of *Sb*-PTE

The X-ray crystal structure of *Sb*-PTE has been recently reported, revealing a seven-bladed β-propeller protein fold (Figure 45A) (135). This enzyme appears as a homodimer in the crystal structure (Figure 46A). A binuclear manganese center is located within the active site of *Sb*-PTE (135). These two metal ions are bridged by Glu-201 and a water molecule. The α manganese ion is further liganded to the protein by the side chain groups of His-389, His-475, and Glu-407 and the β manganese ion is coordinated to His-317 and His-258 (Figure 45B) (135).

3.4.2 Covalent modification of active site residues

The inhibitor 1-butynyl dibutylphosphate proved to be an incredibly potent inhibitor of *Sb*-PTE. Full inhibition was observed with approximately 2 equiv of inhibitor. By contrast, *Pd*-PTE was reported to hydrolyze ∼500 molecules of a similar mechanism-based inhibitor before being fully inactivated (136). The ability to reactivate *Sb*-PTE with hydroxylamine following inactivation with 1-butynyl dibutylphosphate suggests the loss of catalytic activity is most likely due to the modification of one or more histidine residues (136). However, MS/MS analysis identified two serine residues (Ser-458 and Ser-474) and two tyrosine residues (Tyr-219 and Tyr-456) as being labeled. Mutation of Tyr-219, Ser-458, and Tyr-456 residues did not result in substantial loss of catalytic activity or metal binding (137). Mutation of Ser-474, which appears to be adjacent to His-475 (the proposed active site base), did severely diminish the activity of *Sb*-PTE. In the X-ray structure of *Sb*-PTE, two identified residues (Ser-474 and Tyr-219) are located adjacent to the active site. The distance from the α-Mn^{2+} ion at the active site to these two residues is shorter than 12 Å (Figure 47A). The reactive product from the hydrolysis of 1-butynyl dibutylphosphate could migrate within the active site to modify surrounding nucleophilic residues. Two other residues (Tyr-456 and Ser-458) are located in a short hydrophobic loop spanning residues 447-463, far away from the active site. However, in the crystal structure of the *Sb*-PTE dimer, this loop is extended into the active site of the other subunit, precluding phosphotriester binding to the active site (Figure 46B). The distance from Tyr-456 and Ser-458 to the α-Mn^{2+} ion at the active site of the other subunit is 5.8 Å and 16.7 Å, respectively (Figure 47B). In addition, the loop containing these two residues is flexible. The average B-factor of this loop is 50 $Å^2$ while the average B-factor of the whole protein is 39 $Å^2$. This loop could be subject to a conformational change to reduce the distance between Ser-458 and the active site to facilitate the modification. The observed modification of Tyr-456 and Ser-458 suggests that *Sb*-PTE may catalyze the reaction in solution as a dimer. In the catalytic process, the loop containing these two residues could undergo a conformational changes to control the uptake of substrates and the release of products.

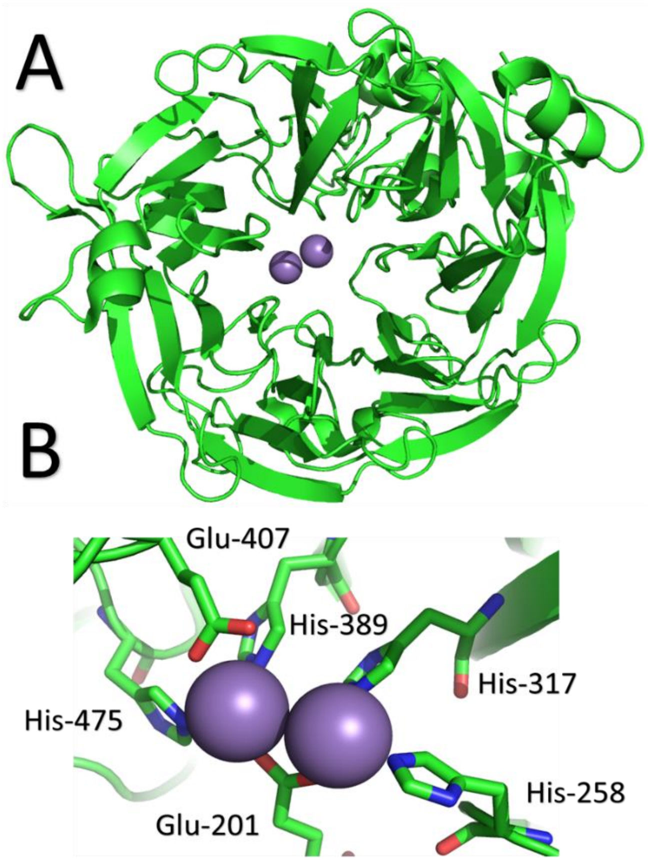


Figure 45: Crystal structure of *Sb*-PTE (PDB id: 5HRM). Manganese ions at the active site are shown in purple. (A) Top down view of *Sb*-PTE showing the seven bladed β-propeller structure. (B) Metal center ligation of *Sb*-PTE.

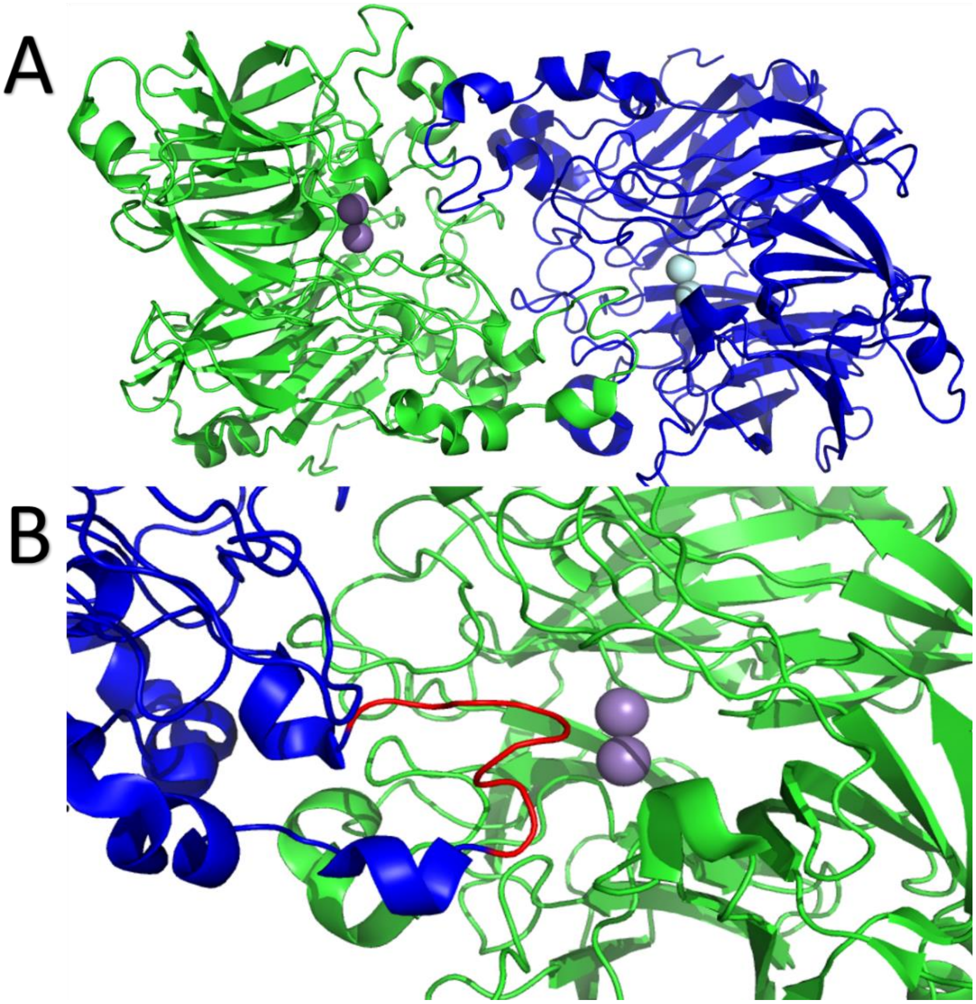


Figure 46: The crystal structure of the *Sb*-PTE dimer. (A) Two subunits of *Sb*-PTE dimer are colored green and blue. The binuclear manganese ions are colored purple in the center of each monomer. (B) The hydrophobic loop (residues 447-463) extending into the active site of the other subunit is colored red.

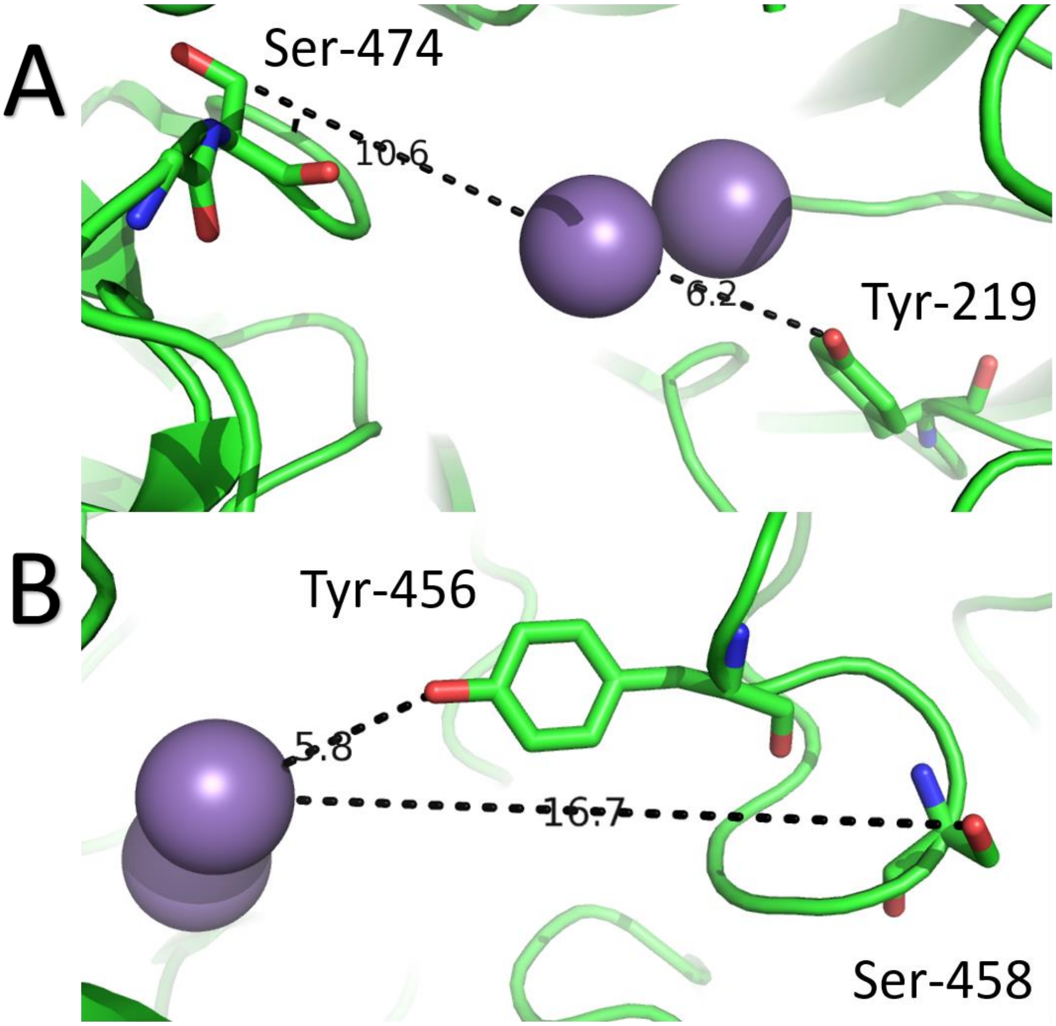


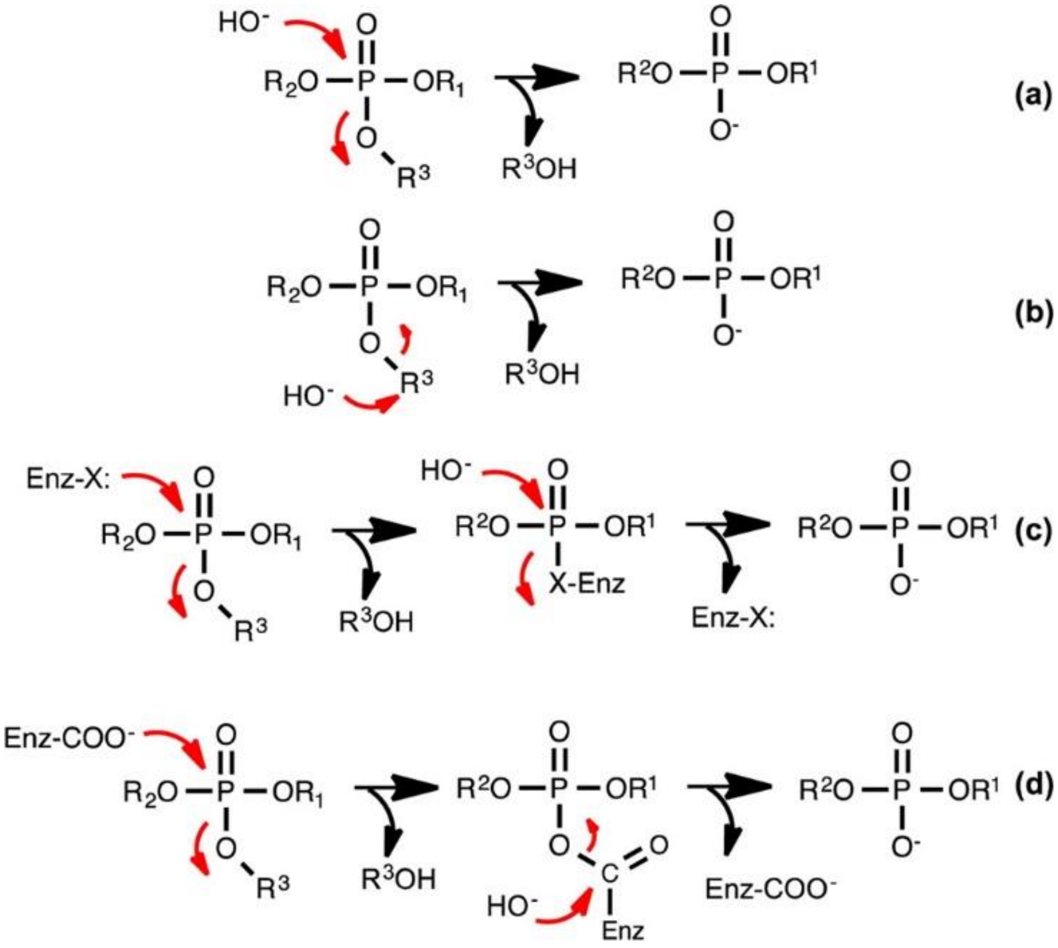
Figure 47: The distance between the active site and four modified residues in the crystal structure of *Sb*-PTE. (A) The distance between Tyr-219 and the α-Mn^{2+} ion is 6.2 Å; the distance between Ser-474 and the α-Mn^{2+} ion is 10.6 Å. (B) The distance between Tyr-456 and the α-Mn^{2+} ion is 5.8 Å; the distance between Ser-458 and the α-Mn^{2+} ion is 16.7 Å.

3.4.3 Proposed chemical mechanism

As a phosphotriesterase with a β-propeller fold, the structure of *Sb*-PTE shows significant differences with other members in this group of enzymes such as DFPase and PON1. Compared with these two enzymes that have a six-bladed β-propeller fold, *Sb*-PTE adopts a seven-bladed β-propeller fold. Instead of a single calcium at the active site, *Sb*-PTE contains a binuclear metal center located in its active site. It is quite surprising that the structure of the binuclear metal center in *Sb*-PTE is quite similar with that in *Pd*-PTE, which adopts a typical TIM barrel fold. The chemical mechanism of the phosphotriester hydrolysis is quite different between phosphotriesterases with β-propeller fold and TIM barrel fold. The enzymes with a β-propeller fold are proposed to go through a covalent reaction intermediate, while the TIM-barrel enzymes usually hydrolyze the substrate by direct attack of water (83).

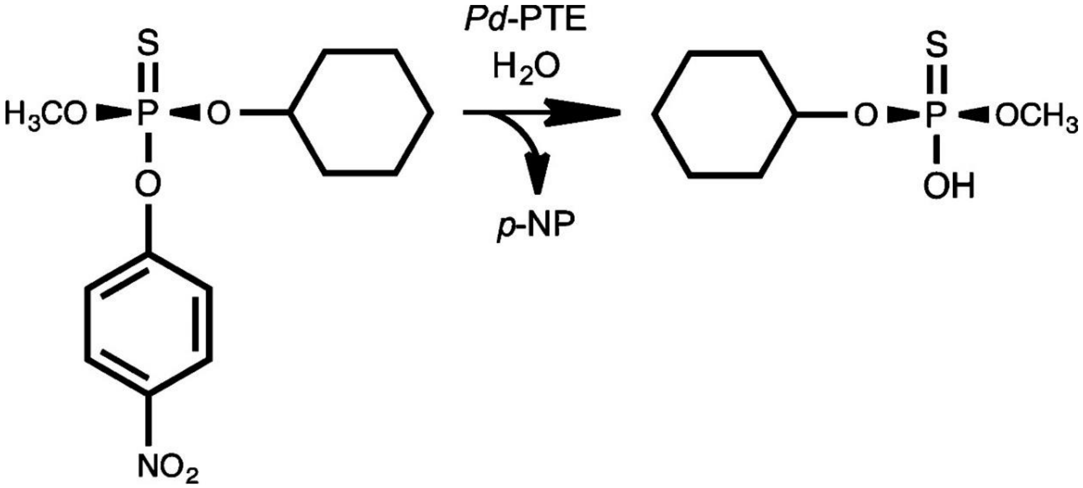
There are a variety of reaction mechanisms that are possible for *Sb*-PTE catalyzed hydrolysis of organophosphatetriesters. The most straightforward mechanism utilizes the direct nucleophilic attack of an activated water/hydroxide on the phosphorus center (Scheme 18a). For substrates with highly activated leaving groups, such as *p*-nitrophenol, it is also possible for the initial nucleophilic attack to occur on the aromatic ring of the leaving group with subsequent C–O bond cleavage through a nucleophilic aromatic substitution mechanism, rather than breakage of the P–O bond (Scheme 18b). Alternatively, the reaction may be initiated by an enzyme-based nucleophile that results in the formation of a covalent intermediate that is subsequently hydrolyzed by an activated water/hydroxide (Scheme 18c). If the enzyme-based nucleophile is the side chain carboxylate from either glutamate or aspartate, then hydrolysis of the covalent intermediate can occur either by nucleophilic attack at phosphorus or attack on the carboxylate carbon proceeding through a typical tetrahedral intermediate (Scheme 18d).

To determine whether the initial nucleophilic attack is directed at the leaving group with C–O bond cleavage (Scheme 18b), *Sb*-PTE was used to catalyze the hydrolysis of paraoxon (diethyl *p*-nitrophenyl phosphate) in ^{18}O-labeled water under conditions of multiple turnovers. If C1 of the *p*-nitrophenyl group is attacked by water/hydroxide then the isolated *p*-nitrophenolate will have a nominal molecular mass of 140, rather than 138. Alternatively, if nucleophilic attack occurs at the phosphorus center, then the diethyl phosphate product will have a nominal mass of 155, rather than 153. As expected, when the hydrolysis reaction was conducted in unlabeled water the *p*-nitrophenolate was isolated with an *m*/*z* of 138 and diethyl phosphate at an *m*/*z* of 153 (Figure 7a). However, when the hydrolysis reaction was conducted under multiple turnover conditions in ∼60% ^{18}O-labeled water, the *p*-nitrophenolate was produced with an *m*/*z* of 138, whereas the diethyl phosphate was isolated as a mixture of species with *m*/*z* values of 153 and 155 (Figure 7b). These results eliminate the mechanism depicted in Scheme 5b. Therefore, the reaction catalyzed by *Sb*-PTE occurs via the direct attack of a nucleophile at the phosphorus center of the substrate.



Scheme 18: Potential chemical mechanisms for *Sb*-PTE

The mechanism depicted in Scheme 5d can be interrogated in two ways. If the reaction is conducted in ^{18}O-labeled water under single turnover conditions, then the recovered diethyl phosphate will not contain an ^{18}O-label. When we utilized a 3-fold molar excess of *Sb*-PTE to hydrolyze paraoxon in ∼60% ^{18}O-labeled water, the isotopic content of the recovered diethyl phosphate was essentially the same as the solvent (Figure 7c). The utilization of a side chain nucleophile from either glutamate or aspartate was also probed by measurement of ^{18}O-incorporation into the protein when hydrolysis of paraoxon was conducted in >95% ^{18}O-labeled water. The net peptide coverage was 87% and no peptide was identified that had an M^{+2} or M^{+4} peak that differed significantly from that of the unlabeled control experiment. These results are inconsistent with the mechanism depicted in Scheme 5d. Stereochemical analysis of the products from the hydrolysis of (R_P)-*O*-methyl *O*-cyclohexyl *O*-(*p*-nitrophenyl), a chiral thiophosphate, showed an inversion of stereochemistry at the phosphorus center indicating a direct attack by the activated water/hydroxide rather than forming a covalent intermediate (Scheme 19). The results from O^{18} labeling mass spectrometry and NMR based stereochemical analysis are consistent with the mechanism depicted in Scheme 18a.



Scheme 19: Hydrolysis of (R_P)-*O*-methyl *O*-cyclohexyl *O*-(*p*-nitrophenyl) thiophosphate (1) to (S_P)-*O*-methyl *O*-cyclohexyl thiophosphate (2) by *Pd*-PTE

3.4.4 Potential applications

There is mounting concern about the environmental impact of the use of organophosphorus flame retardants in plastics and foam materials as well as concerns about organophosphorus solvents used industrially. Many of these compounds are known or suspected carcinogens and are being produced at rates of millions of tons per year. Aviation Syndrome has been directly linked to the use of tricresol phosphate (**9**) in jet fuel and lubricants (138). Unlike insecticides, these compounds are intended to function under the harsh conditions of combustion and consequently are very stable. Typical phosphotriester enzymes have no activity against these compounds because of the lack of a sufficiently labile leaving group. *Sb*-PTE may prove to be useful for the environmental remediation of these compounds because of its unique ability to hydrolyze phosphotriesters without an activated leaving group. However, the structural and biochemical basis for the ability of *Sb*-PTE to hydrolyze the substrate with unactivated leaving group is still a mystery. Deeper understanding on this problem will help to optimize the catalytic efficiency of *Sb*-PTE for more stable fire retardants like tributyl phosphotriesters by direct evolution.

CHAPTER IV

SUMMARY AND CONCLUSION

4.1 Structural Characterization of C-P Lyase Multi-Protein Complex

In this study, structural models of C-P lyase multi-protein complex, PhnG-I, PhnG-H-I-J and PhnG-H-I-J-K were constructed by the integration of various chemical and physical methods including mass spectrometry, protein N-terminal sequencing, analytical ultra-centrifugation and cryo-electron microscopy. Using mass spectrometry and protein N-terminal sequencing, the stoichiometry of protein complexes PhnG-I, PhnG-H-I-J and PhnG-H-I-J-K are determined as $PhnG_2I_2$, $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$, respectively. Analytical ultra-centrifugation results indicated that there are no further dimerization or oligomerization in solution for these complexes. These structural models also revealed the interactions between different subunits of these three protein complexes. Using cryo-electron microscopy, we have solved high resolution (7.8 Å) structural models of the $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$ complexes.

In the structural models of $PhnG_2H_2I_2J_2$ and $PhnG_2H_2I_2J_2K$, $PhnG_2I_2$ works as a central hub where two copies of PhnJ are attached to the PhnI dimer. PhnH is further attached those two copies of PhnJ. The interactions between these subunits were identified and used to create an interaction map for the complex. This map allowed for identification of the amino acids involved in the interface of different subunits. In the structural model of $PhnG_2H_2I_2J_2K$ complex, one copy of PhnK associates with the complex by interacting with the PhnJ subunit to form a five subunit complex. This interaction is shown to be a helix-turn-helix motif within the helical domain of PhnK, which binds to α helix 6 and chock loop of PhnJ. There are two copies of PhnJ associated with the complex in a symmetrical fashion, which results in two identical PhnK binding sites on $PhnG_2H_2I_2J_2$ complex. However, only one copy of PhnK was observed to be associated with the complex by cyro-electron microscopy. Further docking experiments revealed that the binding of the second copy of PhnK is hindered by considerable spacial collision with the other copy of PhnK, which explains that only one copy of PhnK is able to bind to the complex despite two identical binding sites. Furthermore, the binding of PhnK yields to a more exposed Gly-32 residue from. Gly-32 is proposed to be a critical residue in the radical transfer chain under during the hydrolysis of C-P bond. Such rearrangement around the active site may facilitate the delivery of the substrate or release of the reaction product.

Recently, a crystal structure of $PhnG_2H_2I_2J_2$ complex was published. It revealed an intertwined network of subunits PhnG, PhnH, PhnI, and PhnJ with self-homologies. However, the binding of PhnK to the core complex remains unclear. While our interaction map and cryo-electron microscopy models were independently determined and not biased by the recently determined crystal structure of $PhnG_2H_2I_2J_2$, the crystal structure of $PhnG_2H_2I_2J_2$ is consistent with our results. All identified interactions within the complex were observed in the crystal structure. In addition, the crystal structure fits well into the cryo-EM density map of the core complex $PhnG_2H_2I_2J_2$ with matching secondary structures. Compared with the crystal structure, our models provide a more detailed view of the association of PhnK. This structural information provides a basis for further deciphering of the reaction mechanism of the C-P lyase.

4.2 Mechanistic Characterization of *Sb*-PTE

Novel phosphotriesterase-like enzyme from Sphingobium sp. strain TCM1 (*Sb*-PTE) has recently been reported to hydrolyze organophosphates that are not substrates for other phosphotriesterases, such as *Pd*-PTE or OpdA. *Sb*-PTE may prove to be useful for the environmental remediation of these compounds because of its unique ability to hydrolyze phosphotriesters without an activated leaving group. When this study was initiated, very little was known about these enzymes aside from their amino acid sequence and their dependence on the addition of divalent metal ions for catalytic activity. *Sb*-PTE was expressed and purified in our lab by Dr. Daofeng Xiang. She also determined the substrate profile, kinetic parameters and optimal metal of Sb-PTE enzyme.

The inhibitor 1-butynyl dibutylphosphate proved to be an incredibly potent inhibitor of *Sb*-PTE. Full inhibition was observed with approximately 2 equiv of inhibitor. The loss of catalytic activity is most likely due to the modification of one or more residues in the active site. MS/MS analysis identified two serine residues (Ser-458 and Ser-474) and two tyrosine residues (Tyr-219 and Tyr-456) as being labeled. In the X-ray structure of *Sb*-PTE that was published later, two identified residues (Ser-474 and

Tyr-219) are located adjacent to the active site. Two other residues (Tyr-456 and Ser-458) are located in a short hydrophobic loop that is extended into the active site of the other subunit, precluding phosphotriester binding to the active site. The observed modification of theses residues revealed the location of the active site and suggested that *Sb*-PTE may function as a dimer in solution. We also characterized the chemical mechanism of *Sb*-PTE by O^{18} labeling mass spectrometry and NMR based stereochemical analysis. After ruling out the possibility of a covalent reaction intermediate mechanism and mechanisms containing nucleophilic attack on the leaving group, *Sb*-PTE was determined to catalyze the hydrolysis of phosphotriesters by the direct nucleophilic attack of an activated water/hydroxide on the phosphorus center. This mechanism is quite similar with *Pd*-PTE despite the dramatic difference in protein fold.

Currently, the crystal structure and the chemical mechanism of *Sb*-PTE are available. However, the structural and biochemical basis for the ability of *Sb*-PTE to hydrolyze the substrate with unactivated leaving group is still a mystery. Deeper understanding of this problem will allow for the optimization of the catalytic efficiency of *Sb*-PTE for degradation of more stable fire retardants like tributyl phosphotriesters by direct evolution.

REFERENCES

1. De Graaf, R.M. and A.W. Schwartz, *Thermal synthesis of nucleoside h-phosphonates under mild conditions.* Orig Life Evol Biosph, 2005. **35**(1): p. 1-10.
2. Zhang, Q. and W.A. van der Donk, *Answers to the carbon–phosphorus lyase conundrum.* ChemBioChem, 2012. **13**(5): p. 627-629.
3. Metcalf, W.W. and W.A. van der Donk, *Biosynthesis of phosphonic and phosphinic acid natural products.* Annual review of biochemistry, 2009. **78**: p. 65-94.
4. Miceli, M.V., T.O. Henderson, and T.C. Myers, *2-aminoethylphosphonic acid metabolism during embryonic development of the planorbid snail helisoma.* Science, 1980. **209**(4462): p. 1245-7.
5. Quin, L.D., *The prsence of compounds with a carbon-phosphorus bond in some marine invertebrates**. Biochemistry, 1965. **4**(2): p. 324-330.
6. Wanke, C. and N. Amrhein, *Evidence that the reaction of the udp-n-acetylglucosamine 1-carboxyvinyltransferase proceeds through the o-phosphothioketal of pyruvic acid bound to cys115 of the enzyme.* Eur J Biochem, 1993. **218**(3): p. 861-70.
7. Beharry, Z. and T. Palzkill, *Functional analysis of active site residues of the fosfomycin resistance enzyme fosa from pseudomonas aeruginosa.* J Biol Chem, 2005. **280**(18): p. 17786-91.

8. Itoh, H., et al., *Mechanically driven atp synthesis by f1-atpase.* Nature, 2004. **427**(6973): p. 465-468.

9. Bruender, N.A. and V. Bandarian, *The radical s-adenosyl-l-methionine enzyme mftc catalyzes an oxidative decarboxylation of the c-terminus of the mfta peptide.* Biochemistry, 2016. **55**(20): p. 2813-2816.

10. Naydenova, E.D., P.T. Todorov, and K.D. Troev, *Recent synthesis of aminophosphonic acids as potential biological importance.* Amino Acids, 2010. **38**(1): p. 23-30.

11. Hoagland, R.E., *Naturally occurring carbon—phosphorus compounds as herbicides*, in *Biologically active natural products*. 1988, American Chemical Society. p. 182-210.

12. Whitelaw, F.G., et al., *2-aminoethylphosphonic acid concentrations in some rumen ciliate protozoa.* Applied and Environmental Microbiology, 1983. **46**(4): p. 951-953.

13. Quinn, J.P., et al., *New ways to break an old bond: The bacterial carbon–phosphorus hydrolases and their role in biogeochemical phosphorus cycling.* Environmental Microbiology, 2007. **9**(10): p. 2392-2400.

14. Yu, X., et al., *Diversity and abundance of phosphonate biosynthetic genes in nature.* Proc Natl Acad Sci U S A, 2013. **110**(51): p. 20759-64.

15. Villarreal-Chiu, J.F., J.P. Quinn, and J.W. McGrath, *The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment.* Frontiers in Microbiology, 2012. **3**: p. 19.

16. Kim, S.Y., et al., *Different biosynthetic pathways to fosfomycin in pseudomonas syringae and streptomyces species.* Antimicrobial Agents and Chemotherapy, 2012. **56**(8): p. 4175-4183.

17. Hsieh, Y.J. and B.L. Wanner, *Global regulation by the seven-component pi signaling system.* Curr Opin Microbiol, 2010. **13**(2): p. 198-203.

18. Yuan, Z.C., R. Zaheer, and T.M. Finan, *Regulation and properties of pstscab, a high-affinity, high-velocity phosphate transport system of sinorhizobium meliloti.* J Bacteriol, 2006. **188**(3): p. 1089-102.

19. McGrath, J.W., J.P. Chin, and J.P. Quinn, *Organophosphonates revealed: New insights into the microbial metabolism of ancient molecules.* Nature Reviews: Microbiology, 2013. **11**(6): p. 412-419.

20. Kamat, S.S. and F.M. Raushel, *The enzymatic conversion of phosphonates to phosphate by bacteria.* Current Opinion in Chemical Biology, 2013. **17**(4): p. 589-596.

21. Kulakova, A.N., et al., *The purification and characterization of phosphonopyruvate hydrolase, a novel carbon-phosphorus bond cleavage enzyme from variovorax sp pal2.* J Biol Chem, 2003. **278**(26): p. 23426-31.

22. Ternan, N.G. and J.P. Quinn, *In vitrocleavage of the carbon–phosphorus bond of phosphonopyruvate by cell extracts of an environmentalburkholderia cepaciaisolate.* Biochemical and Biophysical Research Communications, 1998. **248**(2): p. 378-381.

23. Chen, C.C.H., et al., *Structure and kinetics of phosphonopyruvate hydrolase from voriovorax sp. Pal2: New insight into the divergence of catalysis within the pep mutase/isocitrate lyase superfamily.* Biochemistry, 2006. **45**(38): p. 11491-11504.

24. Berman, H.A., et al., *Fluorescent phosphonate labels for serine hydrolases. Kinetic and spectroscopic properties of (7-nitrobenz-2-oxa-1,3-diazole)aminoalkyl methylphosphonofluoridates and their conjugates with acetylcholinesterase molecular forms.* J Biol Chem, 1985. **260**(6): p. 3462-8.

25. Kulakova, A.N., et al., *The purification and characterization of phosphonopyruvate hydrolase, a novel carbon-phosphorus bond cleavage enzyme from variovorax sp. Pal2.* Journal of Biological Chemistry, 2003. **278**(26): p. 23426-23431.

26. Kulakova, A.N., et al., *Structural and functional analysis of the phosphonoacetate hydrolase (phna) gene region in pseudomonas fluorescens 23f.* J Bacteriol, 2001. **183**(11): p. 3268-75.

27. Klimek-Ochab, M., A. Mucha, and E. Zymanczyk-Duda, *2-aminoethylphosphonate utilization by the cold-adapted geomyces pannorum p11 strain.* Curr Microbiol, 2014. **68**(3): p. 330-5.

28. Agarwal, V., et al., *Structural and mechanistic insights into c-p bond hydrolysis by phosphonoacetate hydrolase.* Chem Biol, 2011. **18**(10): p. 1230-40.

29. Kim, A., et al., *Divergence of chemical function in the alkaline phosphatase superfamily: Structure and mechanism of the p−c bond cleaving enzyme phosphonoacetate hydrolase.* Biochemistry, 2011. **50**(17): p. 3481-3494.

30. Agarwal, V., et al., *Structural and mechanistic insights into c-p bond hydrolysis by phosphonoacetate hydrolase.* Chemistry & Biology, 2011. **18**(10): p. 1230-1240.

31. Baker, A.S., et al., *Insights into the mechanism of catalysis by the p−c bond-cleaving enzyme phosphonoacetaldehyde hydrolase derived from gene sequence analysis and mutagenesis.* Biochemistry, 1998. **37**(26): p. 9305-9315.

32. Morais, M.C., et al., *The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: Insight into catalysis of phosphorus bond cleavage and catalytic diversification within the had enzyme superfamily.* Biochemistry, 2000. **39**(34): p. 10385-10396.

33. Morais, M.C., et al., *X-ray crystallographic and site-directed mutagenesis analysis of the mechanism of schiff-base formation in phosphonoacetaldehyde hydrolase catalysis.* J Biol Chem, 2004. **279**(10): p. 9353-61.

34. Lahiri, S.D., et al., *Diversification of function in the haloacid dehalogenase enzyme superfamily: The role of the cap domain in hydrolytic phosphoruscarbon bond cleavage.* Bioorganic Chemistry, 2006. **34**(6): p. 394-409.

35. McSorley, F.R., et al., *Phny and phnz comprise a new oxidative pathway for enzymatic cleavage of a carbon–phosphorus bond.* Journal of the American Chemical Society, 2012. **134**(20): p. 8364-8367.

36. Price, J.C., et al., *Kinetic dissection of the catalytic mechanism of taurine:A-ketoglutarate dioxygenase (taud) from escherichia coli.* Biochemistry, 2005. **44**(22): p. 8138-8147.

37. Brown, P.M., et al., *Crystal structure of a substrate complex of myo-inositol oxygenase, a di-iron oxygenase with a key role in inositol metabolism.* Proceedings of the National Academy of Sciences, 2006. **103**(41): p. 15032-15037.

38. McGrath, J.W., F. Hammerschmidt, and J.P. Quinn, *Biodegradation of phosphonomycin by rhizobium huakuii pmy1.* Appl Environ Microbiol, 1998. **64**(1): p. 356-8.

39. McGrath, J.W., et al., *Studies on the biodegradation of fosfomycin: Growth of rhizobium huakuii pmy1 on possible intermediates synthesised chemically.* Org Biomol Chem, 2009. **7**(9): p. 1944-53.

40. Makino, K., et al., *Molecular analysis of the cryptic and functional phn operons for phosphonate use in escherichia coli k-12.* J Bacteriol, 1991. **173**(8): p. 2665-72.

41. Metcalf, W.W. and B.L. Wanner, *Mutational analysis of an escherichia coli fourteen-gene operon for phosphonate degradation, using tnphoa' elements.* Journal of Bacteriology, 1993. **175**(11): p. 3430-3442.

42. Pipke, R. and N. Amrhein, *Carbon-phosphorus lyase activity in permeabilized cells of arthrobacter sp. Glp-1.* FEBS Letters, 1988. **236**(1): p. 135-138.

43. Shames, S.L., et al., *Fragmentative and stereochemical isomerization probes for homolytic carbon to phosphorus bond scission catalyzed by bacterial carbon-phosphorus lyase.* Bioorganic Chemistry, 1987. **15**(4): p. 366-373.

44. Okamoto, Y., et al., *Photochemical c-p bond cleavage of some (substituted benzyl)phosphonic acid derivatives.* NIPPON KAGAKU KAISHI, 1987. **1987**(7): p. 1255-1261.

45. Kamat, S.S., H.J. Williams, and F.M. Raushel, *Intermediates in the transformation of phosphonates to phosphate by bacteria.* Nature, 2011. **480**(7378): p. 570-573.

46. Chen, C.M., et al., *Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the phn (psid) genes involved in alkylphosphonate uptake and c-p lyase activity in escherichia coli b.* J Biol Chem, 1990. **265**(8): p. 4461-71.

47. Rizk, S.S., M.J. Cuneo, and H.W. Hellinga, *Identification of cognate ligands for the escherichia coli phnd protein product and engineering of a reagentless fluorescent biosensor for phosphonates.* Protein Science : A Publication of the Protein Society, 2006. **15**(7): p. 1745-1751.

48. Makino, K., et al., *Molecular analysis of the cryptic and functional phn operons for phosphonate use in escherichia coli k-12.* Journal of Bacteriology, 1991. **173**(8): p. 2665-2672.

49. Errey, J.C. and J.S. Blanchard, *Functional annotation and kinetic characterization of phno from salmonella enterica.* Biochemistry, 2006. **45**(9): p. 3033-9.

50. Hove-Jensen, B., et al., *Escherichia coli phnn, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): Dual role in phosphonate degradation and nad biosynthesis pathways.* J Bacteriol, 2003. **185**(9): p. 2793-801.

51. Hove-Jensen, B., et al., *Escherichia coli phnn, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): Dual role in phosphonate degradation and nad biosynthesis pathways.* Journal of Bacteriology, 2003. **185**(9): p. 2793-2801.

52. He, S.M., et al., *Structure and mechanism of phnp, a phosphodiesterase of the carbon-phosphorus lyase pathway.* Biochemistry, 2011. **50**(40): p. 8603-15.

53. Kamat, S.S., et al., *The catalytic mechanism for aerobic formation of methane by bacteria.* Nature, 2013. **497**(7447): p. 132-136.

54. Ren, Z., et al., *Subunit interactions within the carbon–phosphorus lyase complex from escherichia coli.* Biochemistry, 2015. **54**(21): p. 3400-3411.

55. Hove-Jensen, B., D.L. Zechel, and B. Jochimsen, *Utilization of glyphosate as phosphate source: Biochemistry and genetics of bacterial carbon-phosphorus lyase.* Microbiol Mol Biol Rev, 2014. **78**(1): p. 176-97.

56. White, A.K. and W.W. Metcalf, *Two c-p lyase operons in pseudomonas stutzeri and their roles in the oxidation of phosphonates, phosphite, and hypophosphite.* J Bacteriol, 2004. **186**(14): p. 4730-9.

57. Villarreal-Chiu, J.F., J.P. Quinn, and J.W. McGrath, *The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment.* Front Microbiol, 2012. **3**: p. 19.

58. Ghodge, S.V., et al., *Discovery of a cyclic phosphodiesterase that catalyzes the sequential hydrolysis of both ester bonds to phosphorus.* Journal of the American Chemical Society, 2013. **135**(44): p. 16360-16363.

59. Dyall-Smith, M.L., et al., *Haloquadratum walsbyi: Limited diversity in a global pond.* PLoS One, 2011. **6**(6): p. e20968.

60. Antunes, A., et al., *Genome sequence of halorhabdus tiamatea, the first archaeon isolated from a deep-sea anoxic brine lake.* J Bacteriol, 2011. **193**(17): p. 4553-4.

61. Adams, M.A., et al., *Crystal structure of phnh: An essential component of carbon-phosphorus lyase in escherichia coli.* Journal of Bacteriology, 2008. **190**(3): p. 1072-1083.

62. Jochimsen, B., et al., *Five phosphonate operon gene products as components of a multi-subunit complex of the carbon-phosphorus lyase pathway.* Proceedings of the National Academy of Sciences, 2011. **108**(28): p. 11393-11398.

63. Musilek, K., et al., *Design, evaluation and structure-activity relationship studies of the ache reactivators against organophosphorus pesticides.* Med Res Rev, 2011. **31**(4): p. 548-75.

64. Maxwell, D.M., et al., *Acetylcholinesterase inhibition: Does it explain the toxicity of organophosphorus compounds?* Arch Toxicol, 2006. **80**(11): p. 756-60.

65. Reemtsma, T., et al., *Organophosphorus flame retardants and plasticizers in water and air i. Occurrence and fate.* TrAC Trends in Analytical Chemistry, 2008. **27**(9): p. 727-737.

66. van der Veen, I. and J. de Boer, *Phosphorus flame retardants: Properties, production, environmental occurrence, toxicity and analysis.* Chemosphere, 2012. **88**(10): p. 1119-53.

67. Abe, K., et al., *Haloalkylphosphorus hydrolases purified from sphingomonas sp. Strain tdk1 and sphingobium sp. Strain tcm1.* Appl Environ Microbiol, 2014. **80**(18): p. 5866-73.

68. Harper, L.L., et al., *Dissimilar plasmids isolated from pseudomonas diminuta mg and a flavobacterium sp. (atcc 27551) contain identical opd genes.* Appl Environ Microbiol, 1988. **54**(10): p. 2586-9.

69. Sun, L., et al., *Crystallization and preliminary x-ray studies of methyl parathion hydrolase from pseudomonas sp. Wbc-3.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 5): p. 954-6.

70. Zhongli, C., L. Shunpeng, and F. Guoping, *Isolation of methyl parathion-degrading strain m6 and cloning of the methyl parathion hydrolase gene.* Appl Environ Microbiol, 2001. **67**(10): p. 4922-5.

71. DeFrank, J.J. and T.C. Cheng, *Purification and properties of an organophosphorus acid anhydrase from a halophilic bacterial isolate.* J Bacteriol, 1991. **173**(6): p. 1938-43.

72. Elias, M., et al., *Structural basis for natural lactonase and promiscuous phosphotriesterase activities.* J Mol Biol, 2008. **379**(5): p. 1017-28.

73. Benning, M.M., et al., *Three-dimensional structure of the binuclear metal center of phosphotriesterase.* Biochemistry, 1995. **34**(25): p. 7973-8.

74. Dong, Y.J., et al., *Crystal structure of methyl parathion hydrolase from pseudomonas sp. Wbc-3.* J Mol Biol, 2005. **353**(3): p. 655-63.

75. Koepke, J., et al., *Statistical analysis of crystallographic data obtained from squid ganglion dfpase at 0.85 a resolution.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 10): p. 1744-54.

76. Blum, M.M., et al., *Rapid determination of hydrogen positions and protonation states of diisopropyl fluorophosphatase by joint neutron and x-ray diffraction refinement.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 713-8.

77. Vyas, N.K., et al., *Structural insights into the dual activities of the nerve agent degrading organophosphate anhydrolase/prolidase.* Biochemistry, 2010. **49**(3): p. 547-59.

78. Omburo, G.A., et al., *Characterization of the zinc binding site of bacterial phosphotriesterase.* J Biol Chem, 1992. **267**(19): p. 13278-83.

79. Hartleib, J., et al., *Role of calcium ions in the structure and function of the di-isopropylfluorophosphatase from loligo vulgaris.* Biochem J, 2001. **353**(Pt 3): p. 579-89.

80. Blum, M.M., et al., *Binding of a designed substrate analogue to diisopropyl fluorophosphatase: Implications for the phosphotriesterase mechanism.* J Am Chem Soc, 2006. **128**(39): p. 12750-7.

81. Aubert, S.D., Y. Li, and F.M. Raushel, *Mechanism for the hydrolysis of organophosphates by the bacterial phosphotriesterase.* Biochemistry, 2004. **43**(19): p. 5707-15.

82. Chen, J.C., et al., *Neutron structure and mechanistic studies of diisopropyl fluorophosphatase (dfpase).* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 11): p. 1131-8.

83. Bigley, A.N. and F.M. Raushel, *Catalytic mechanisms for phosphotriesterases.* Biochim Biophys Acta, 2013. **1834**(1): p. 443-53.

84. Serdar, C.M., et al., *Plasmid involvement in parathion hydrolysis by pseudomonas diminuta.* Applied and Environmental Microbiology, 1982. **44**(1): p. 246-249.

85. Mulbry, W.W., et al., *Identification of a plasmid-borne parathion hydrolase gene from flavobacterium sp. By southern hybridization with opd from pseudomonas diminuta.* Applied and Environmental Microbiology, 1986. **51**(5): p. 926-930.

86. Horne, I., et al., *Identification of an opd (organophosphate degradation) gene in an agrobacterium isolate.* Appl Environ Microbiol, 2002. **68**(7): p. 3371-6.

87. Somara, S. and D. Siddavattam, *Plasmid mediated organophosphate pesticide degradation by flavobacterium balustinum.* Biochem Mol Biol Int, 1995. **36**(3): p. 627-31.

88. Yang, H., et al., *Evolution of an organophosphate-degrading enzyme: A comparison of natural and directed evolution.* Protein Eng, 2003. **16**(2): p. 135-45.

89. Benning, M.M., et al., *Three-dimensional structure of the binuclear metal center of phosphotriesterase.* Biochemistry, 1995. **34**(25): p. 7973-7978.

90. Kim, J., et al., *Structure of diethyl phosphate bound to the binuclear metal center of phosphotriesterase.* Biochemistry, 2008. **47**(36): p. 9497-9504.

91. Jackson, C., et al., *The structure of an enzyme–product complex reveals the critical role of a terminal hydroxide nucleophile in the bacterial phosphotriesterase mechanism.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2005. **1752**(1): p. 56-64.

92. Benning, M.M., et al., *High resolution x-ray structures of different metal-substituted forms of phosphotriesterase from pseudomonas diminuta.* Biochemistry, 2001. **40**(9): p. 2712-22.

93. Omburo, G.A., L.S. Mullins, and F.M. Raushel, *Structural characterization of the divalent cation sites of bacterial phosphotriesterase by 113cd nmr spectroscopy.* Biochemistry, 1993. **32**(35): p. 9148-55.

94. Hong, S.B. and F.M. Raushel, *Metal-substrate interactions facilitate the catalytic activity of the bacterial phosphotriesterase.* Biochemistry, 1996. **35**(33): p. 10904-12.

95. Vanhooke, J.L., et al., *Three-dimensional structure of the zinc-containing phosphotriesterase with the bound substrate analog diethyl 4-methylbenzylphosphonate.* Biochemistry, 1996. **35**(19): p. 6020-5.

96. Chen-Goodspeed, M., et al., *Structural determinants of the substrate and stereochemical specificity of phosphotriesterase.* Biochemistry, 2001. **40**(5): p. 1325-31.

97. Bigley, A.N. and F.M. Raushel, *Catalytic mechanisms for phosphotriesterases.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013. **1834**(1): p. 443-453.

98. Wilce, M.C., et al., *Structure and mechanism of a proline-specific aminopeptidase from escherichia coli.* Proc Natl Acad Sci U S A, 1998. **95**(7): p. 3472-7.

99. Hoskin, F.C. and R.J. Long, *Purification of a dfp-hydrolyzing enzyme from squid head ganglion.* Arch Biochem Biophys, 1972. **150**(2): p. 548-55.

100. Tiegang, L., et al., *Polyethylene glycosylation prolongs the stability of recombinant human paraoxonase-1.* Toxicol Lett, 2012. **210**(3): p. 366-71.

101. Harel, M., et al., *Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes.* Nat Struct Mol Biol, 2004. **11**(5): p. 412-9.

102. Dawson, R.M., et al., *Degradation of nerve agents by an organophosphate-degrading agent (opda).* J Hazard Mater, 2008. **157**(2-3): p. 308-14.

103. Cherny, I., et al., *Engineering v-type nerve agents detoxifying enzymes using computationally focused libraries.* ACS Chemical Biology, 2013. **8**(11): p. 2394-2403.

104. Dumas, D.P., et al., *Purification and properties of the phosphotriesterase from pseudomonas diminuta.* J Biol Chem, 1989. **264**(33): p. 19659-65.

105. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat. Protocols, 2007. **1**(6): p. 2856-2860.

106. Gotze, M., et al., *Stavrox--a software for analyzing crosslinked products in protein interaction studies.* J Am Soc Mass Spectrom, 2012. **23**(1): p. 76-87.

107. Bai, Y., et al., *Primary structure effects on peptide group hydrogen exchange.* Proteins, 1993. **17**(1): p. 75-86.

108. Englander, S.W. and N.R. Kallenbach, *Hydrogen exchange and structural dynamics of proteins and nucleic acids.* Q Rev Biophys, 1983. **16**(4): p. 521-655.

109. Walters, B.T., et al., *Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment.* J Am Soc Mass Spectrom, 2012. **23**(12): p. 2132-9.

110. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the web: A case study using the phyre server.* Nat Protoc, 2009. **4**(3): p. 363-71.

111. Arnold, K., et al., *The swiss-model workspace: A web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

112. Li, X., et al., *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em.* Nat Meth, 2013. **10**(6): p. 584-590.

113. Grant, T. and N. Grigorieff, *Measuring the optimal exposure for single particle cryo-em using a 2.6 å reconstruction of rotavirus vp6.* eLife, 2015. **4**: p. e06980.

114. Tang, G., et al., *Eman2: An extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38-46.

115. Scheres, S.H.W., *Relion: Implementation of a bayesian approach to cryo-em structure determination.* Journal of Structural Biology, 2012. **180**(3): p. 519-530.

116. Elmlund, H., D. Elmlund, and S. Bengio, *Prime: Probabilistic initial 3d model generation for single-particle cryo-electron microscopy.* Structure, 2013. **21**(8): p. 1299-1306.

117. Heymann, J.B. and D.M. Belnap, *Bsoft: Image processing and molecular modeling for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 3-18.

118. Pettersen, E.F., et al., *Ucsf chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

119. Guex, N., M.C. Peitsch, and T. Schwede, *Automated comparative protein structure modeling with swiss-model and swiss-pdbviewer: A historical perspective.* Electrophoresis, 2009. **30 Suppl 1**: p. S162-73.

120. Devi, S.K., et al., *Structural basis for the hydrolysis of atp by a nucleotide binding subunit of an amino acid abc transporter from thermus thermophilus.* J Struct Biol, 2015. **190**(3): p. 367-72.

121. Pintilie, G.D., et al., *Quantitative analysis of cryo-em density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions.* J Struct Biol, 2010. **170**(3): p. 427-38.

122. Schuck, P., et al., *Size-distribution analysis of proteins by analytical ultracentrifugation: Strategies and application to model systems.* Biophysical Journal, 2002. **82**(2): p. 1096-1111.

123. Sinz, A., *Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes.* J Mass Spectrom, 2003. **38**(12): p. 1225-37.

124. Weis, D.D., et al., *Identification and characterization of ex1 kinetics in h/d exchange mass spectrometry by peak width analysis.* J Am Soc Mass Spectrom, 2006. **17**(11): p. 1498-509.

125. Wales, T.E. and J.R. Engen, *Hydrogen exchange mass spectrometry for the analysis of protein dynamics.* Mass Spectrom Rev, 2006. **25**(1): p. 158-70.

126. Liyanage, R., et al., *Comparison of two esi ms based h/d exchange methods for extracting protein folding energies.* Int J Mass Spectrom, 2009. **287**(1-3): p. 96-104.

127. Seweryn, P., et al., *Structural insights into the bacterial carbon-phosphorus lyase machinery.* Nature, 2015. **525**(7567): p. 68-72.

128. Ambudkar, S.V., et al., *The a-loop, a novel conserved aromatic acid subdomain upstream of the walker a motif in abc transporters, is critical for atp binding.* FEBS Lett, 2006. **580**(4): p. 1049-55.

129. Maqbool, A., et al., *The substrate-binding protein in bacterial abc transporters: Dissecting roles in the evolution of substrate specificity.* Biochem Soc Trans, 2015. **43**(5): p. 1011-7.

130. Schmitt, L., et al., *Crystal structure of the nucleotide-binding domain of the abc-transporter haemolysin b: Identification of a variable region within abc helical domains.* J Mol Biol, 2003. **330**(2): p. 333-42.

131. Locher, K.P., A.T. Lee, and D.C. Rees, *The e. Coli btucd structure: A framework for abc transporter architecture and mechanism.* Science, 2002. **296**(5570): p. 1091-8.

132. Demarest, S.J., et al., *Mutual synergistic folding in recruitment of cbp/p300 by p160 nuclear receptor coactivators.* Nature, 2002. **415**(6871): p. 549-53.

133. Bhattacherjee, A. and S. Wallin, *Coupled folding-binding in a hydrophobic/polar protein model: Impact of synergistic folding and disordered flanks.* Biophys J, 2012. **102**(3): p. 569-78.

134. Shevchenko, A., et al., *In-gel digestion for mass spectrometric characterization of proteins and proteomes.* Nat Protoc, 2006. **1**(6): p. 2856-60.

135. Mabanglo, M.F., et al., *Structure of a novel phosphotriesterase from sphingobium sp. Tcm1: A familiar binuclear metal center embedded in a seven-bladed β-propeller protein fold.* Biochemistry, 2016. **55**(28): p. 3963-3974.

136. Banzon, J.A., et al., *Mechanism-based inactivation of phosphotriesterase by reaction of a critical histidine with a ketene intermediate.* Biochemistry, 1995. **34**(3): p. 743-9.

137. Xiang, D.F., et al., *Interrogation of the substrate profile and catalytic properties of the phosphotriesterase from sphingobium sp. Strain tcm1: An enzyme capable of hydrolyzing organophosphate flame retardants and plasticizers.* Biochemistry, 2015. **54**(51): p. 7539-7549.

138. Mackerer, C.R., et al., *Comparison of neurotoxic effects and potential risks from oral administration or ingestion of tricresyl phosphate and jet engine oil containing tricresyl phosphate.* J Toxicol Environ Health A, 1999. **57**(5): p. 293-328.