A Dynamics-enabled Structural Characterization of a Set of Poly(aspartic acid) Hydrolases from *Sphingomonas* sp. KT-1

By

Chad A. BRAMBLEY

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Molecular Biosciences

MIDDLE TENNESSEE STATE UNIVERSITY
August, 2022

Dissertation Advisory Committee:

Dr. Justin Miller, Research Advisor

Dr. Kevin Bicker, Chair

Dr. Anatoliy Volkov

Dr. David Nelson

Dr. Scott Handy

“I am pleased to see that we have differences. May we together become greater than the sum of both of us.”

-Surak

“Dedicated in loving memory of Nevada Hurst. So much of who I am is because of you.”

“To my beautiful wife, Brooklyn. For your patience, your support, and your never yielding unconditional acceptance and love. ”

“To our family.”

ACKNOWLEDGMENTS

This work is made possible through the help and support of others. I would first like to thank all contributors of funding and facilities, including the Molecular Biosciences Doctoral Program (MOBI) and the Department of Chemistry at Middle Tennessee State University (MTSU).

I am honored to thank my advisor, Dr. Justin Miller, for his invaluable mentorship and instruction. He has supported me in ways that far exceed the science. Special thanks as well to my brother in science, Justin Marsee, in whom I can always find a valued collaborator and friend. Additional thanks are extended to all others in the Miller lab, past and present. Each of you have helped guide this project through your feedback, criticisms, support, and enthusiasm.

I am also thankful for assistance from our various collaborators, including the labs of Dr. Jamie Wallen and Dr. Mitch Weiland of Western Carolina and Georgia Southern Universities, respectively. This list further includes a special acknowledgement of Dr. Anatoliy Volkov, also of MTSU, whose generous offering of high performance computing resources has been of immeasurable value to this work. Dr. Volkov additionally serves on my dissertation committee, alongside Drs. Justin Miller, Kevin Bicker, David Nelson, and Scott Handy.

Finally, I express my sincerest gratitude to all of my friends and family who have supported me with unwavering love and guidance. The list of names I wish to credit is too lengthy to include in its entirety, so I shall simply say:

I love you all.

ABSTRACT

Polymers of aspartic acid (PAA) belong to a class of water-soluble polycarboxylates (WSP) with widespread usage in a variety of commercial applications. Given increasing evidence that such WSPs may pose environmental risks via leeching and accumulation into ground and surface water, the need to develop an effective means of degradation and bioremediation is clear. Biodegradation of PAA has most recently been made possible by the discovery of a collection of PAA hydrolases isolated from river water bacteria, *Sphingomonas* sp. KT-1 and *Pedobacter* sp. KP-2. PAA-hydrolase-1 (PahZ1) was identified in both species, while a subsequent PAA-hydrolase-2 (PahZ2) was found only in *Sphingomonas* sp. KT-1. The two enzymes have previously been shown to cooperatively carry out the hydrolysis of PAA as part of a multi-step catalytic mechanism. $\mathrm{PahZ1}_{\mathrm{KT\text{-}1}}$ targets β-amide linkages to yield oligo(aspartic acid)(OAA), which is in turn further degraded by $\mathrm{PahZ2}_{\mathrm{KT\text{-}1}}$. Here, we present a collection of dynamics focused structural characterizations of both enzymes. We additionally report a machine-learning strategy to model the structure and motions of a putative chimeric enzyme engineered to include both $\mathrm{PahZ1}_{\mathrm{KT\text{-}1}}$ and $\mathrm{PahZ2}_{\mathrm{KT\text{-}1}}$ functional domains.

TABLE OF CONTENTS

Chapter 2: Sphingomonas sp. KT-1 PahZ1 **26**

Chapter 3: Sphingomonas sp. KT-1 PahZ2 **71**

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PAA	**P**oly **a**aspartic **a**cid
tPAA	**T**hermally synthesized **P**oly **a**aspartic **a**cid
dPAA	**D**igested **P**oly **a**aspartic **a**cid
OAA	**O**ligo**a**aspartic **a**cid
WSP	**W**ater **s**oluble **p**olymers
PahZ1	**P**oly **a**spartic acid **h**ydrolase **1**
PahZ2	**P**oly **a**spartic acid **h**ydrolase **2**
PE	**P**oly**e**thylene
PVC	**P**oly**v**inyl **c**hloride
PET	**P**oly**e**thylene **t**erephthalate
SP	**S**odium **p**olyacrylate
SAP	**S**super **a**bsorbent polymer
γ-PGA	**P**oly-γ-**g**lutamic **a**cid
PSI	**P**oly **s**uccin**i**mide
GPC	**G**el **P**ermeation **c**hromatography
DFP	**D**iisopropyl **f**luoro**p**hosphate
PMSF	**P**henyl**m**ethane**s**ulfonyl **f**luoride
NMR	**N**uclear **m**agnetic **r**resonance
WGS	**W**hole **g**enome **s**hotgun sequencing

EBI	**E**uropean **B**ioinformatics **I**nstitute
EDTA	**E**thylene**d**iamine**t**etraacetic **a**cid
CPG2	**C**arboxy**p**eptidase **G2**
MD	**M**olecular **d**ynamics
SAXS	**S**mall **a**ngle **X**-ray **s**cattering
MALS	**M**ulti **a**ngle **l**ight **s**cattering
AEBSF	4-2-**a**mino**e**thyl**b**enzene**s**ulfonyl **f**luoride
SEC	**S**ize **e**xclusion **c**hromatography
HPLC	**H**igh **p**erformance **l**iquid **c**hromatography
PDB	**P**rotein **D**ata **B**ank
NVT	**N**umber of particles, **v**olume, and **t**emperature
NPT	**N**umber of particles, **p**ressure, and **t**emperature
COM	**C**enter **o**f **m**ass
SMD	**S**teered **m**olecular **d**ynamics
PMF	**P**otential of **m**ean **f**orce
PCA	**P**rincipal **c**omponent **a**nalysis
RMSD	**R**oot **m**ean **s**quare **d**eviation
RMSF	**R**oot **m**ean **s**quare **f**luctuation
FEL	**F**ree **e**nergy **l**andscape
LB	**L**ysogeny **b**roth
NTA	**N**i**t**riloacetic **a**cid

Chapter 1

Introduction

1.1 A brief history of synthetic polymers

Man-made synthetic polymers, including polyethylene (PE), poly(vinyl chloride)(PVC), and poly(ethylene terephthalate)(PET) have enjoyed consistent utility in modern industry for greater than a half-century.[1] Packaging, building materials, hygiene products, and medical devices are just a few example applications for these these ubiquitous plastics. Their decades-long popularity has unfortunately left the current generation with addressing the environmental impact of the accumulation of plastic waste products. For scale, it is estimated that only 30% of the total plastic ever produced ($\sim$7800 Mt between 1950 and 2015) is currently in use,[1] leaving the rest to slowly degrade in landfills and oceans. Such shortcomings in waste management are not without consequences. For example, the presence of plastic debris in the earth's oceans has been shown to threaten the integrity of marine ecosystems.[2] Further, it may pose threats to human health via the leeching of toxic nanoplastics and subsequent uptake into the food chain.[3,4]

Attempts to alleviate these concerns have led to increased utilization of a class of water-soluble poly(carboxylate) polymers (WSPs). Common WSPs, including poly(acrylate), polyacrylamide, and poly(vinyl alcohol), have seen application in the manufacturing of detergents, adhesives, food packaging, and water treatment materials.[5] A prevailing notion leading to their adoption is that many WSPs offer a greener alternative to other synthetic polymers, with little impact on the environment. However, the belief that these products are eco-friendly is an assumption based on a lack of noticeable accumulation.[5] In reality, many WSPs tend to degrade only slowly, if at all. The environmental persistence of poly(carboxylates) and other WSPs has been well established, with potentially damaging ecological consequences.[6–8] When such polymers do demonstrate measurable biodegradation, the resulting products are not necessarily benign. For example, fears have been expressed that polyacrylamide may slowly leech monomers of acrylamide, a compound with potent neurotoxic effects, into ground or surface water.[6,9,10]

Another example of a polymeric threat to the environment can be found in sodium poly(acrylate) (SP), better known to industry as a superabsorbant polymer (SAP). SP consists of crosslinked networks of hydrophilic and anionic acrylate monomers capable of trapping large amounts of water.[11] This property makes them an ideal candidate for a variety of household and medical applications, including diapers, feminine hygiene products, absorbant pads, drug-delivery methods, and soil treatment.[12] SP is therefore widely used, owed to both its utility and a perceived environmental friendliness. While it is known to be non-toxic, it is also non-biodegradable, again raising the issue of environmental persistence. It presents a particular concern for the management of waste water when SP is deployed as a dispersant or antiscaling agent. Its high solubility in water means a recovery mechanism dependent on chemical precipitation, yielding masses of insoluble material that contribute to the global carbon footprint by accumulating in landfills.[13] Such environmental consequences illustrate a need to develop greener alternatives.

1.1.1 Understanding the versatility of poly(amino acids)

Polymers of naturally-occurring amino acids may offer several biodegradable and biocompatible candidates for human utilization including, but certainly not limited to, environmental and biomedical applications. Of relevance to the latter, poly-gamma-glutamate (γ-PGA) has been shown to support a healthier composition of gut microbiota in murine models, leading to its proposed use as a promising prebiotic.[14] It has also been employed as a drug-delivery agent of anticancerous chemotherapeutic compounds. Conjugation of these drugs with γ-PGA effectively alters the pharmacokinetics and biodistribution of the drug and thus lowers biological toxicity.[15,16] Also pursued for medical importance are various enzymes capable of hydrolyzing γ-PGA. The pathogenicity of *Bacillus anthracis*, for example, is largely attributed to the antiphagocytic properties of its γ-D-PGA composed capsule. Treatment with CapD, a γ-PGA depolymerase, has been shown to degrade the capsule and thus enable recognition by macrophages.[17] A more promising replacement for traditional SAPs may be found instead in poly(aspartic acid)(PAA). When coupled with a group of recently isolated PAA hydrolases, these polymers may present a robust biodegradable solution to the previously detailed ecological consequences of existing products.[6–8] In this review, we present an introduction to thermally synthesized PAA, as well as to the hydrolases that facilitate its enzyme-catalyzed biodegradation.

1.2 Poly(aspartic acid) as a biodegradable alternative to conventional superabsorbents

As previously mentioned, it has been proposed that polymers of naturally occurring amino acids may offer the best environmental bang for our industrial buck. Fully biodegradable and biocompatible, such polymers may find utilization in biomedical applications, cosmetics, soil treatment, and absorbent materials.[18] Poly(aspartic acid)(PAA), in particular, has been echoed in the field as not only an alternative, but as an *improvement* over existing products.[18] For instance, PAA is easily synthesized, biodegradable, has a lower molecular weight, and boasts greater absorbency than the leading non-biodegradable SAPs.[18,19] Additionally, PAA can be utilized in diverse ways. For example, in addition to acting as an antiscaling agent and superabsorbent, PAA can also reduce the need for excess plant fertilizer by forming a thin protective layer around roots, increasing nutrient retention and uptake.[18,20] In order to better illustrate how PAA might serve as a practical alternative to other commercial polymers, it is first necessary to explore the biochemical details of its manufacture and biodegradation.

1.2.1 Thermal polymerization of PAA

Procedures detailing the synthesis of PAA from naturally occurring aspartic acid monomers has been summarized by Thombre and coworkers.[18] Briefly, the process involves the thermal poly-condensation of either L-aspartic acid or maleamic acid, resulting in closed rings of a poly(succinimide)(PSI) intermediate. These rings are then opened via alkaline hydrolysis,[21] yielding thermally synthesized polymers of aspartic acid (tPAA). As already stated, the final product of this method tends to exhibit structural inconsistencies, including irregular end groups, the presence of both α and β-amide linkages, as well as extensive branching and cross-linking.[22] tPAA may also present with racemic mixtures of both L and D aspartate isomers, regardless of which was used as an initial reactant.[18,21]

Figure 1.1C and Figure 1.1D present a more detailed illustration of tPAA synthesis. Both α and β-amide linked subunits are shown as the product of the alkaline hydrolysis of polysuccinimide rings. Structural examples of other water-soluble poly(carboxylates), as well as more traditional plastics, are included for reference in Figure 1.1A and Figure 1.1B. As with any chemical synthesis, optimization of tPAA polymerization may yield products with increased functionality and performance. To that end, structure-controlled methods of polymerization have been investigated, including the ring-opening of α-amino acid N-carboxylic anhydrides.[22,23] Such a protocol is considerably more challenging, however, requiring high monomer purity and the use of highly toxic phosphene

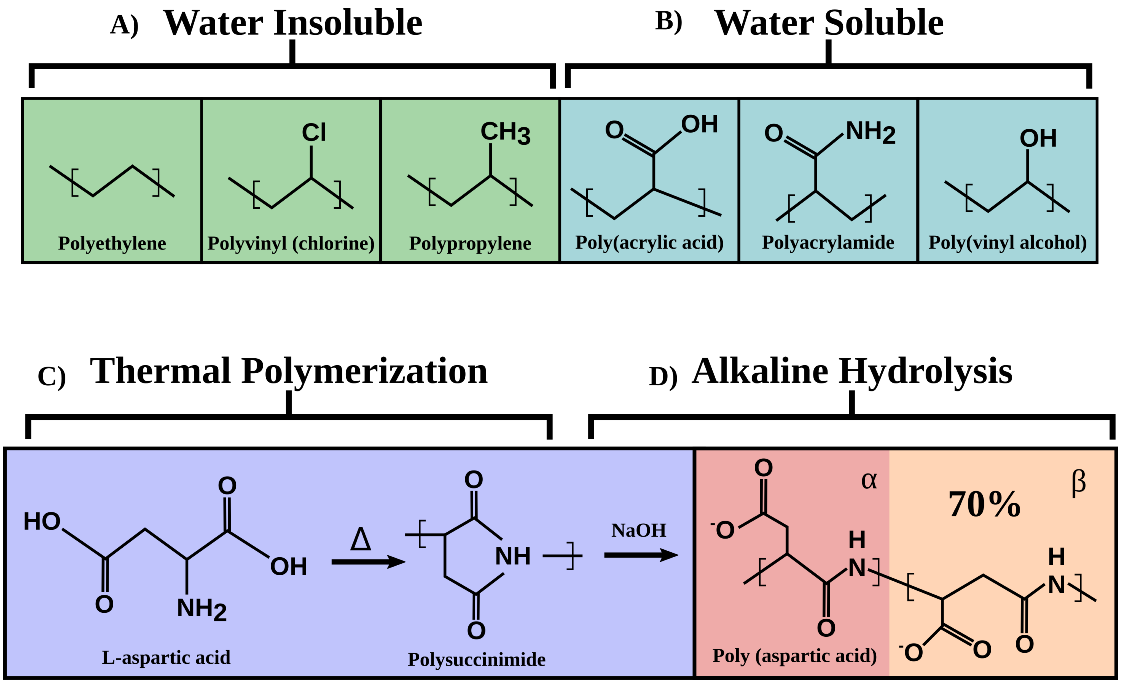


Figure 1.1: Line drawing representations of various polymerized substrates. Commonly used water-insoluble (A) and water-soluble (B) polymers are shown highlighted in green and blue, respectively. Thermal polymerization of L-aspartic acid, yielding rings of polysuccinimide is illustrated in panel C. The subsequent alkaline hydrolysis of of polysuccinimide is shown in panel D, yielding both α and β linked poly (aspartic acid). Repeating units are indicated by square brackets.

gas. Enzymatic synthesis has instead been offered as a safer structure-controlled alternative.[22,24] Comprehensive discussion of tPAA synthetic mechanisms is beyond the scope of this review, as the primary environmental focus rests with biodegradability of tPAA substrates.

1.2.2 Microbial degradation of tPAA by *Sphingamonas* sp. KT-1 and *Pedobacter* sp. KP-2

In order to understand how and why tPAA substrates may be considered environmentally friendly, it is first important to make note of existing challenges to biodegradation. For example, biodegradability of polymers is known to be limited by substrate availability to the local microbial populations.[18] It has therefore been suggested that the tendency of tPAA to be removed from solution via adsorption, for example onto the roots of plants, may be problematic for its availability to microbial hydrolases.[18,20,25,26] Further, the previously described structural distinctions from natural proteins are another obstacle for tPAA biodegradation. Recall that tPAA structures are highly unusual, exhibiting irregular end-groups, branching, and around 70% β-linkages,[22] all of which are features that may impact the rate of microbial breakdown.[24–26]

To elucidate these possible roadblocks to biodegradation, Tabata and coworkers published their findings detailing two strains of bacteria from samples of river water.[27,28] A summary of their results is presented as follows. Both *Sphingamonas* sp. KT-1 and *Pedobacter* sp. KP-2 were identified as tPAA-degrading organisms.[28] In the case of *Sphingamonas* sp. KT-1, biodegradation of tPAA was observed following incubation with both whole cell isolates and cellular extracts, with the latter being capable of degrading high-molecular-weight tPAA up to 150 kDa.[27] Product weights were consistent with monomers of aspartic acid. In contrast, whole cell isolates were only able to digest low-molecular-weight tPAA (<5.0 kDa). These results suggest that, while *Sphingamonas* sp. KT-1 contains enzymes capable of degrading tPAA, complete degradation by living cells seemed to be limited to low-molecular-weight tPAA. This finding was attributed to the inability of heavier polymers to enter the cell's cytoplasm.[27,28]

Conversely, similar experiments carried out in the presence of *Pedobacter* sp. KP-2 found the whole cell isolate was capable of degrading tPAA with measured weights between 50 and 150 kDa.[28] While GPC assays measured diminishing tPAA over 20 days, the team also reportedb a concomitant accumulation of low-molecular-weight products between 2.5 and 5.0 kDa. Thus, access to intracellular enzymes was not restricted to low-molecular-weight tPAA as was observed for *Spingomonas* sp. KT-1. This finding led researchers to suggest a model by which outer-membrane bound PAA-hydrolyzing enzymes were able to degrade high-molecular-weight products, reducing their weight and enabling their transport into the cell. When tPAA was incubated in a mixed culture of both *Pedobacter* sp. KP-2 and *Sphingomonas* sp. KT-1, however, GPC measurements reveal complete degradation of tPAA after only 12 days without the same accumulation of low-molecular-weight products. Although it should be noted that when replicated on samples exhibiting more extensive branching, minimal quantities of low-molecular-weight tPAA was still present after 12 days.[28] Taken together, these data suggest that the microbial biodegradation of tPAA by *Sphingomonas* sp. KT-1 and *Pedobacter* sp. KP-2 is a structure-dependent process, wherein the molecular-weight and amount of subunit branching restricts access to the cell's cytoplasm.

1.2.3 Biodegradation of tPAA by PAA hydrolases purified from *Sphingomonas* sp. KT-1

As recounted above, direct biodegradation of tPAA by microbial culture requires strict structural control during synthesis to enable access of the substrate to cytoplasmic hydrolases. If tPAA is to be utilized as an eco-friendly alternative to existing SAPs in commercial settings, it will be necessary to overcome these structure-restricted hurdles. To that end, more data was needed to understand the biochemical nature of the relevant hydrolases, their preferred tPAA architectures, and structural characterization of their catalyzed products.

Biochemical description of PAA hydrolase-1 from *Sphingomonas* sp. KT-1

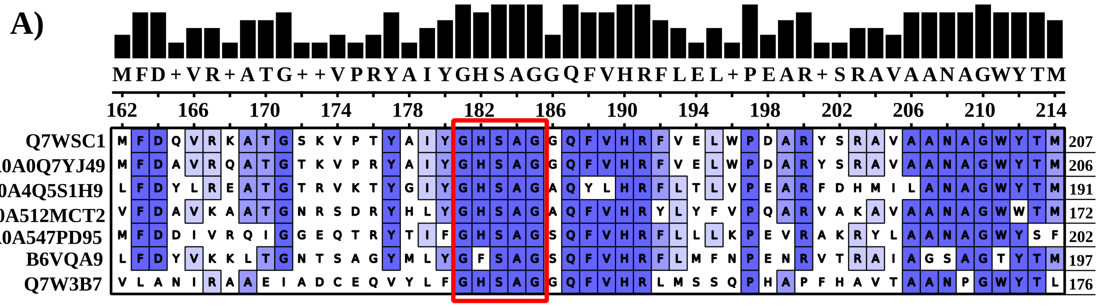
Early efforts to isolate the enzyme(s) responsible for tPAA degradation led to the identification of PAA-hydrolase-1 (PahZ1$_{\text{KT-1}}$), a 30 kDa enzyme from *Sphingomonas* sp. KT-1.[29] A summary of the resulting biochemical characterization can be found in Table 1.1 in comparison with other tPAA-hydrolyzing enzymes to be discussed later. Briefly, PahZ1$_{\text{KT-1}}$ proteolytic activity was found to be inhibited by the presence of diisopropyl fluorophosphates (DFP) or phenylmethane sulfonyl fluoride (PMSF), whereas other protease inhibitors had little or no effect.[29,30] This finding led the group to conclude that PahZ1$_{\text{KT-1}}$ is a serine-type hydrolase, as DFP and PMSF are known inhibitors of serine-center active sites.[29,31,32] This was later confirmed by genetic analysis, wherein a conserved lipase box motif (Gly-Xaa-Ser176-Xaa-Gly) was identified with the active site serine

176 (S176) at its center.[30] Specific activity measurements carried out on Serine-to-alanine (S176A) mutants found no accumulation of tPAA-hydrolyzed products, revealing S176 to be a catalytically important residue for tPAA hyrolysis.[30]

To determine whether PahZ1$_{\text{KT-1}}$ is responsible for the recovery of aspartic acid monomers, samples of tPAA were incubated in the presence of enzyme for 24 hours.[29] Analysis via GPC revealed no such monomers, and instead identified products with molecular weights consistent with oligomers of aspartic acid (OAA). To be sure, no further degradation was observed following additional incubation with fresh enzyme.[29] Structural characterization of both reactant tPAA and the resultant OAA via NMR spectroscopy revealed the catalytic preferences of PahZ1$_{\text{KT-1}}$ activity.[29] Prior to degradation, peaks were observed corresponding to each of the four diad sequences that comprise the tPAA network, including α-α, β-β, as well as the peak sum of α-β and β-α.[29] Following 24 hours of incubation with purified enzyme, only the β-β peak was found to have been reduced. These results suggest that PahZ1$_{\text{KT-1}}$ displays a preference for the cleavage of β-β diad sequences via endo-mode hydrolysis.[29]

Investigating conservation of the PahZ1$_{\text{KT-1}}$ primary sequence

At the time of publication,[29] no proteins showing significant sequence homology with PahZ1$_{\text{KT-1}}$ had been discovered. To investigate whether this had changed, we conducted a BLASTP search against PahZ1$_{\text{KT-1}}$. The output identified ASD67_20930, an uncharacterized protein from *Sphingopyxis* sp. Root1497, as sharing 88% of its primary sequence identity with PahZ1$_{\text{KT-1}}$. The subsequent results of a multiple sequence alignment are shown in Figure 1.2A. Each gene product examined feature the same highly conserved serine-centered lipase-box motif, highlighted here in red. Alignment entries are ordered by percent similarity to PahZ1$_{\text{KT-1}}$, the values of which are shown in Figure 1.2B. The sequences of ASD67_20930, along with other uncharacterized proteins listed here (Uniprot IDs A0A*), were derived from whole genome shotgun (WGS) sequecing courtesy of the European Bioinformatics Institute (EMBL-EBI).[33,34]



B)

Uniprot ID	Gene	Species	Similarity to PahZ1$_{\text{KT-1}}$
Q7WSC1	PahZ	*Sphingomonas* sp. KT-1	100 %
A0A0Q7YJ49	ASD67_20930	*Sphingopyxis* sp. Root 1497	88%
A0A4Q5S1H9	EOP62_09125	*Sphingomonadales bacterium*	55%
A0A512MCT2	BGE01nite_38030	*Brevifollis gellanilyticus*	40%
A0A547PD95	FGU71_09770	*Erythrobacter insulae*	40%
B6VQA9	PahZ	*Pedobacter* sp. KP-2	35%
Q7W3B7	BPP4128	*Bordatella parapertussis*	29%

Figure 1.2: Primary sequence analysis of PahZ1$_{\text{KT-1}}$, PahZ1$_{\text{KP-2}}$, LpqC, and uncharacterized gene products. (A) Multiple sequence alignment with PahZ1$_{\text{KT-1}}$ and related proteins, labelled according to Uniprot ID. Consensus sequence and histograms are shown at the top of the figure in black. Color scheme indicates relative conservation, whereas darker violet implies a highly conserved position. Red box indicates the position of the highly conserved lipase-box motif. (B) Table associating each Uniprot ID with gene name and corresponding species. Percent similarity to PahZ1$_{\text{KT-1}}$ is listed in descending order for each gene product.

Perhaps most surprising are the final two entries in the alignment, each reflecting only minimal sequence homology with PahZ1$_{\text{KT-1}}$. The PahZ gene from *Pedobacter* sp. KP-2 yields a protein of only 35% similarity, whereas BPP4128 from *Bordatella parapertussis* achieves just 29%. The latter represents LpqC, a putative poly(3-hydroxybutyrate) depolymerase. Given that LpqC has already been shown to exhibit significant *structural* homology with PahZ1$_{\text{KT-1}}$ (see Chapter2),[35] we considered the same may be true of the other constituents of this alignment. To address this question, we performed an AI-guided structure prediction of the PahZ hydrolase from *Pedobacter* sp. KP-2, along with each of the uncharacterized proteins shown in Figure 1.2 and subsequently report the results later in Chapter 4.

Biochemical description of PAA hydrolase-2 from *Sphingomonas* sp. KT-1

Tabata and coworkers had previously found that tPAA degradation by *Sphingomonas* sp. KT-1 whole-cell extract would yield the appearance of aspartic acid monomers.[27] The lack of such monomers recovered from PahZ1$_{\text{KT-1}}$-catalyzed degradation of tPAA, led the team to conclude the existence of a second unidentified enzyme that must play a role in the further degradation of OAA. This hypothesis would lead to the purification by Hiraishi et al. of PAA hydrolase-2 from *Sphingomonas* sp. KT-1 (PahZ2$_{\text{KT-1}}$) from microbial cellular extracts.[36] The study reports an N-terminal sequence of PahZ2$_{\text{KT-1}}$ that is consistent with the M20/M25/M40 family of peptidases. Preliminary hydrolysis assays on α-di-PAA revealed optimum activity at a pH of 7.0 and 55°C. The latter measurement was noted to be inconsistent with the ideal growth temperature for *Sphingomonas* sp. KT-1 cultures (28°C).[36] PahZ2$_{\text{KT-1}}$ activity on α-di-PAA was further found to be inhibited by EDTA, suggesting the enzyme may require metal ion cofactors for catalysis.[36] Additional biochemical parameters[22,28,29,36] for PahZ2$_{\text{KT-1}}$ can be found alongside PahZ1$_{\text{KT-1}}$ and PahZ1$_{\text{KP-2}}$ in Table 1.1.

Table 1.1: Biochemical details of the PAA hydrolases from *Spingomonas* sp. KT-1 and *Pedobacter* sp. KP-2

Property	**PahZ1$_{\text{KT-1}}$**	**PahZ2$_{\text{KT-1}}$**	**PahZ1$_{\text{KP-2}}$**
Subunit Mol.Wt. (kDa)	30	42	30
Localization	Cytosol	Cytosol	Periplasm
Isoelectric Point	8.9	9.6	8.57
Inhibitors	PMSF, DFP	EDTA, PMSF, DFP	PMSF, DFP
Degradation Type	Endo	Exo	Endo
Opt. Temperature (°C)	40	55	40
Opt. pH	10.0	7.0	7.5
Catalytic Mechanism	Serine hydrolase	Zinc metallopeptidase	Serine hydrolase

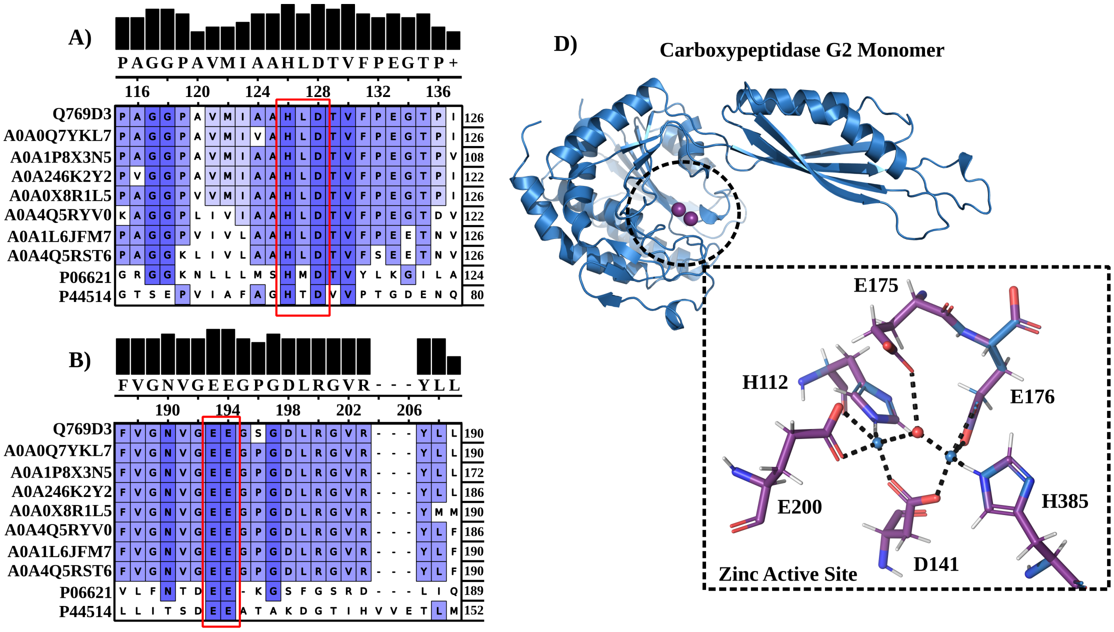
To test the hypothesis that PahZ2$_{\text{KT-1}}$ is required to break down OAA, as well as to further characterize its ability to hydrolyze PAA, the group performed activity assays in the presence of tPAA, OAA, and a structure-controlled polymer of α-L-aspartic acid (α-L-PAA). Subsequent product analysis following 24 hours of incubation revealed that both α-L-PAA and OAA was degraded by PahZ2$_{\text{KT-1}}$ with a concominant appearance of monomeric aspartic acid. In contrast, tPAA samples were mostly unchanged following the same 24 hour incubation.[36] The ability of PahZ2$_{\text{KT-1}}$ to fully degrade α-L-PAA but not tPAA was attributed to the presence of irregular end-groups of the tPAA structure, a finding consistent with substrate recognition of an exo-mode hydrolase.[36] Further, the structural compatibility with tPAA-derived OAA suggests the PahZ1$_{\text{KT-1}}$-catalyzed degradation of tPAA results in a product with end-groups amenable to PahZ2$_{\text{KT-1}}$ activity.[36] To further elucidate substrate recognition patterns, researchers assayed PahZ2$_{\text{KT-1}}$-catalyzed hydrolysis of both α-penta-L-aspartic acid and β-penta-L-aspartic acid.[36] In both cases, penta-L-aspartic acid was observed to decrease over time and was matched by an increasing appearance of tetra-L-aspartic acid.[36] Measurable quantities of trimers, however, were not observed until at least 60 minutes of reaction time.[36] These data led the group to conclude that PahZ2$_{\text{KT-1}}$ does in fact act upon OAA via exo-mode hydrolysis.[36] Further, it does so without significant preference for α- or β-amide linkages.[36]

These results, considered in context with the previously described activity of PahZ1$_{\text{KT-1}}$, suggest a model by which these two hydrolases may work together to catalyze the conversion of tPAA into monomers of aspartic acid. To summarize, tPAA β-amide linkages are clipped by PahZ1$_{\text{KT-1}}$. The yielded OAA may then be recognized by PahZ2$_{\text{KT-1}}$. The end of this pathway returns monomers of aspartic acid, which can subsequently be used as raw materials for polymerizing more tPAA. In this way, PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ could be deployed in tandem as needed to complete a cradle-to-cradle manufacturing process.

Exploration of PahZ2$_{\text{KT-1}}$ primary sequence conservation

As mentioned in the previous section, the PahZ2$_{\text{KT-1}}$ N-terminus had already been shown to present sequence homology with the M20/M25/M40 family of peptidases. Hiraishi et. al later cloned the complete PahZ2 gene, revealing no significant homology with PahZ1$_{\text{KT-1}}$.[36] Instead they discovered conservation with carboxypeptidase G2 (CPG2) from *Pseudomonas* sp. strain RS-16 (27.5% identity).[36] CPG2 is a known zinc-dependent exopeptidase, often used to reduce methotrexate toxicity during cancer treatment by hydrolyzing the drug to an inactive isoform.[37,38] This finding is consistent with the existing description of PahZ2$_{\text{KT-1}}$ as a metal-dependent exopeptidase. In the absence of any PahZ2$_{\text{KT-1}}$ structural determination, the comparison with CPG2 was also able to yield insights into a possible PahZ2$_{\text{KT-1}}$ structure and mechanism. Briefly, CPG2 structural data reveals a homodimeric complex wherein a zinc-bound catalytic domain is connected via a flexible hinge region to a dimerization domain.[38] The catalytic domain is characterized by a dual-zinc coordinating active site built upon conserved His-Xaa-Asp and Glu-Glu motifs, both of which are shared by PahZ2$_{\text{KT-1}}$.

Figure 1.3 reports the results of a multiple sequence alignment of PahZ2$_{\text{KT-1}}$ and CPG2, as well as a succinyl-diaminopimelate desuccinylase (dapE) from *Haemophilus influenzae*, a CPG2 structural homologue.[38–40] The remaining proteins shown represent uncharacterized peptidases uncovered by a BLASTP[33] search. Figure 1.3A-B indicates the positions of the conserved zinc-coordinating residues, shown here with red outlines. The gene and associated gene product



C)

Uniprot ID	Gene	Protein Classification	Organism
Q769D3	pahZ2	Poly(aspartic acid)-hydrolase	*Sphingomonas* sp. KT-1
A0A0Q7YKL7	ASD67_20940	Peptidase M20	*Sphingopyxis* sp. Root1497
A0A1P8X3N5	BWQ93_12675	Peptidase M20	*Sphingopyxis* sp. QXT-31
A0A246K2Y2	CDQ92_06400	Peptidase M20	*Sphingopyxis bauzanensis*
A0A0X8R1L5	pahZ2	Poly(aspartic acid)-hydrolase	*Sphingopyxis granuli*
A0A4Q5RYV0	EOP62_09115	M20/M25/M40 metallo-hydrolase	*Sphingomonadales bacterium*
A0A1L6JFM7	BRX40_21725	M20/M25/M40 metallo-hydrolase	*Sphingomonas koreensis*
A0A4Q5RST6	EOP62_16375	M20/M25/M40 metallo-hydrolase	*Sphingomonadales bacterium*
P06621	cpg2	Carboxypeptidase G2	*Pseudomonas sp. (strain RS-16)*
P44514	dapE	Succinyl-diaminopimelate desuccinylase	*Haemophilus influenzae*

Figure 1.3: Primary sequence conservation study of PahZ2$_{\text{KT-1}}$ with highest percent identity gene products obtained via Uniprot. (A-B) Multiple sequence alignment with gene products labeled by Uniprot ID and ranked by percent identity with PahZ2$_{\text{KT-1}}$. Color scheme indicates degree of position conservation whereas dark shades of violet signals highly conserved residues. The red boxes are used to highlight conserved metal ion coordinating motifs. Consensus sequence is shown at the top edge of each alignment view in black text. (C) Informational table indicating the gene, protein type, and species of origin associated with each Uniprot ID. (D) Cartoon representation of the structure of a carboxypeptidase G2 monomer (PDB: 1CG2). Zinc ions shown as violet spheres. The dashed box presents enlarged view of the zinc active site. Zinc atoms and coordinating residues are shown in blue and violet, respectively. Bridging water molecule is included and shown in red.

is shown along with the species of origin in Figure 1.3C. Figure 1.3D depicts a cartoon representation of a CPG2 monomeric subunit. The two bound zinc ions are shown here as violet spheres within the binding pocket of the labelled catalytic domain. Both cofactors, labelled as ZnI and ZnII, are tetrahedrally coordinated by 3-4 residues and a bridging water molecule. This arrangement, as shown in the enlarged sub-panel of Figure 1.3D, is consistent with the M28 family of zinc metallopeptidases.[41] The sequence homology with CPG2 taken together with empirical observations of $\text{PahZ2}_{\text{KT-1}}$ biochemical behavior presents the case for its inclusion as a member of the M28 family of zinc peptidases. However, more data is still needed to fully understand the structural nature of $\text{PahZ2}_{\text{KT-1}}$ cleavage of OAA and its role in the tPAA hydrolysis pathway. For this reason we present, in Chapter 3, the structure of $\text{PahZ2}_{\text{KT-1}}$ and further characterize its catalytic mechanism.

1.3 Engineering the future of tPAA synthesis and reclamation

The effectiveness of tPAA in regards to its application as a ecofriendly alternative to existing polycarboxylates rests on the efficacy of the enzymes involved in its biodegradation. We have alluded to the promise of a cradle-to-cradle synthetic process wherein tPAA is cleaved by $\text{PahZ1}_{\text{KT-1}}$ into shorter chains of OAA, the target substrate for $\text{PahZ2}_{\text{KT-1}}$. Therefore, the end product of the proposed catalytic workflow is monomeric subunits of aspartic acid that could in theory support additional tPAA synthesis. In this way, these hydrolases represent a mechanism for developing an immeasurably sustainable polymer with little to no threat of environmental accumulation. As already described, however, there are several challenges that must first be overcome if such a reality is to be realized. For example, PahZ1$_{\text{KT-1}}$ activity is limited by its preference for cleaving β-linked units of tPAA. Similarly, and as further described later in Chapter 3, PahZ2$_{\text{KT-1}}$ activity falls short of achieving the complete conversion of OAA into aspartic acid. Such restricted selectivity and incomplete degradation illustrates a major shortcoming of using both enzymes in their current form. Less polymer would be ultimately recoverable and could lead to the unwanted leeching of uncleaved polymer back into the environment, further compounding the very phenomenon this work is attempting to curb.

To that end, the overarching goal of the following research is to explore possible strategies toward developing a novel tPAA hydrolase combining the respective functions of PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ while limiting the system of their catalytic inefficiencies. We emphasize here that the following studies concerning PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-2}}$ in Chapter2 and Chapter3, respectively, serve to understand the structural and dynamic details driving their respective functions, as well as to elucidate a structural basis underpinning their described limitations. It is only with such knowledge that they may be optimally made of use through future protein engineering efforts to improve activity or introduce additional functionality. For example, we hope that this work will lead to synthetic PahZ1-like complexes capable of non-selectively cleaving both α- and β-linked units of tPAA to enable a more complete conversion into OAA. Similar constructs may be further improved upon such that additional polymers, including other polycarboxylates, could become potential targets for degradation. Another limitation that we hope to address through protein engineering is the intrinsic inefficiency of requiring multiple enzymes to facilitate the complete degradation and recovery of tPAA. We propose later in Chapter 4 that a chimeric enzyme, engineered to include both $\mathrm{PahZ1}_{\mathrm{KT\text{-}1}}$ and $\mathrm{PahZ2}_{\mathrm{KT\text{-}1}}$ functional domains, may offer a promising alternative wherein tPAA could be directly converted to monomeric aspartic acid without the need for additional enzymes or any intermediate steps. We will combine sophisticated machine learning techniques with molecular dynamics and other advanced modelling approaches to investigate putative structures of two candidate designs. Taken together, the succeeding chapters present a dynamics-driven structural analysis yielding molecular insights that are expected to enable the development of novel enzymes for the degradation and reclamation of synthetic tPAA and other polymers. At the very least, it is hoped that these works will inspire new approaches and more ecofriendly sensibilities when it comes to polymer production and distribution.

References

(1) Geyer, R.; Jambeck, J.; Law, K. Production, use, and fate of all plastics ever made. *Science Advances* **2017**, *3*, 1700782.

(2) Taylor M. L., G. C. R. L. F.; Woodall, L. C. Plastic microfibre ingestion by deep-sea organisms. *Sci Rep* **2016**, *6*, 33997.

(3) Jahnke, A.; Arp, H.; Escher, B.; Gewert, B.; Gorokhova, E.; Kühnel, D.; Ogonowski, M.; Potthoff, A.; Rummel, C.; Schmitt-Jansen, M. Reducing Uncertainty and Confronting Ignorance about the Possible Impacts of Weathering Plastic in the Marine Environment. *Environmental Science & Technology Letters* **2017**, *4*, 85–90.

(4) Waring, R.; Harris, R.; Mitchell, S. Plastic contamination of the food chain: A threat to human health? *Maturitas* **2018**, *115*, 64–68.

(5) Swift, G. In *Degradable Polymers: Principles and Applications*, Scott, G., Ed.; Dordrecht: Springer Netherlands, 2002, 379–412.

(6) Arp, H.; Knutsen, H. Could We Spare a Moment of the Spotlight for Persistent, Water-Soluble Polymers? *Environmental Science & Technology* **2020**, *54*, 3–5.

(7) Hennecke, D.; Bauer, A.; Herrchen, M.; Wischerhoff, E.; Gores, F. Cationic polyacrylamide copolymers (PAMs): environmental half life determination in sludge-treated soil. *Environ Sci Eur* **2018**, *30*, 16.

(8) Jop, K.; Guiney, P.; Christensen, K.; Silberhorn, E. Environmental fate assessment of two synthetic polycarboxylate polymers. *Ecotoxicol Environ Saf* **1997**, *37*, 229–37.

(9) Xiong, B.; Loss, R.; Shields, D.; Pawlik, T.; Hochreiter, R.; Zydney, A.; Kumar, M. Polyacrylamide degradation and its implications in environmental systems. *npj Clean Water* **2018**, *1*, DOI: `10.1038/s41545-018-0016-8`.

(10) Buczek, S.; Cope, W.; McLaughlin, R.; Kwak, T. Acute toxicity of polyacrylamide flocculants to early life stages of freshwater mussels. *Environ Toxicol Chem* **2017**, *36*, 2715–2721.

(11) Zhao, C.; Zhang, M.; Liu, Z.; Guo, Y.; Zhang, Q. Salt-Tolerant Superabsorbent Polymer with High Capacity of Water-Nutrient Retention Derived from Sulfamic Acid-Modified Starch. *ACS Omega* **2019**, *4*, 5923–5930.

(12) Sharma, S.; Dua, A.; Malik, A. Polyaspartic acid based superabsorbent polymers. *European Polymer Journal* **2014**, *59*, 363–376.

(13) Kumar, A. Polyaspartic Acid - A Versatile Green Chemical. **2012**, 6.

(14) Jin, H.-E.; Choi, J.-C.; Lim, Y. T.; Sung, M.-H. Prebiotic effects of poly-gamma-glutamate on bacterial flora in murine gut. *Journal of microbiology and biotechnology* **2017**, *27*, 412–415.

(15) Zhang, S.-Q., *Poly (L-glutamic acid)-paclitaxel conjugates for cancer treatment*; IntechOpen: 2012.

(16) Kim, C.-K.; Lim, S.-J. Recent progress in drug delivery systems for anticancer agents. *Archives of pharmacal research* **2002**, *25*, 229–239.

(17) Scorpio, A.; Chabot, D. J.; Day, W. A.; O'brien, D. K.; Vietri, N. J.; Itoh, Y.; Mohamadzadeh, M.; Friedlander, A. M. Poly-γ-glutamate capsule-degrading enzyme treatment enhances phagocytosis and killing of encapsulated Bacillus anthracis. *Antimicrobial agents and chemotherapy* **2007**, *51*, 215–222.

(18) Thombre, S.; Sarwade, B. Synthesis and Biodegradability of Polyaspartic Acid: A Critical Review. *Journal of Macromolecular Science, Part A* **2005**, *42*, 1299–1315.

(19) Katritzky, A. R.; Yao, J.; Qi, M.; Qiu, G.; Bao, W.; Yang, B.; Denisko, O.; Davis, S.; Zhang, J. Preparation and physical properties of N-defunctionalized derivatives of poly(aspartic acid). *Journal of Applied Polymer Science* **2001**, *81*, 85–90.

(20) Reisch, M. Butting heads in polyaspartic acid. *Chemical & Engineering News* **2002**, *80*, 23–24.

(21) Adelnia, H.; Blakey, I.; Little, P. J.; Ta, H. T. Hydrogels Based on Poly(aspartic acid): Synthesis and Applications. *Frontiers in Chemistry* **2019**, *7*, 755.

(22) Hiraishi, T.; Maeda, M. Poly(aspartate) hydrolases: biochemical properties and applications. *Appl Microbiol Biotechnol* **2011**, *91*, 895–903.

(23) Kricheldorf, H. R. Polypeptides and 100 Years of Chemistry of α-Amino Acid N-Carboxyanhydrides. *Angewandte Chemie International Edition* **2006**, *45*, 5752–5784.

(24) Hiraishi, T.; Taguchi, S. Enzyme-catalyzed Synthesis and Degradation of Biopolymers. *Mini-Reviews in Organic Chemistry* **2009**, *6*, 44–54.

(25) Tang, Y.; Wheeler, A. Environmental Factors That Influence Biodegradation of Thermal Poly(aspartate). *ACS Symposium Series* **2001**, *786*, 157–171.

(26) Alford, D; Wheeler, A.; Pettigrew, C. Biodegradation of thermally synthesized polyaspartate. *CA J Environ Polym Degrad* **1994**, *2*, 225.

(27) Tabata, K.; Kasuya, K.; Abe, H.; Masuda, K.; Doi, Y. Poly(aspartic acid) degradation by a Sphingomonas sp. isolated from freshwater. *Appl Environ Microbiol* **1999**, *65*, 4268–4270.

(28) Tabata, K.; Abe, H.; Doi, Y. Microbial Degradation of Poly(aspartic acid) by Two Isolated Strains of Pedobacter sp. and Sphingomonas sp. *Biomacromolecules* **2000**, *1*, 157–161.

(29) Tabata, K.; Kajiyama, M.; Hiraishi, T.; Abe, H.; Yamato, I.; Doi, Y. Purification and characterization of poly(aspartic acid) hydrolase from Sphingomonas sp. *KT-1. Biomacromolecules* **2001**, *2*, 1155–60.

(30) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Yamato, I.; Doi, Y. Genetic Analysis and Characterization of Poly(aspartic acid) Hydrolase-1 from Sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 80–86.

(31) Kita, K; Ishimaru, K; Teraoka, M; Yanase, H; Kato, N Properties of poly(3-hydroxybutyrate) depolymerase from a marine bacterium, Alcaligenes faecalis AE122. *Applied and environmental microbiology* **1995**, *61*, 1727—1730.

(32) Nakamura, K.; Tomita, T.; Abe, N.; Kamio, Y. Purification and Characterization of an Extracellular Poly(Lactic Acid) Depolymerase from a Soil Isolate, Amycolatopsis sp. Strain K104-1. *Applied and Environmental Microbiology* **2001**, *67*, 345–353.

(33) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **1990**, *215*, 403–410.

(34) Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **2020**, *49*, D480–D489.

(35) Brambley, C.; Bolay, A.; Salvo, H.; Jansch, A.; Yared, T.; Miller, J.; Wallen, J.; Weiland, M. Structural Characterization of Sphingomonas sp. KT-1 PahZ1-Catalyzed Biodegradation of Thermally Synthesized Poly(aspartic acid. *ACS Sustainable Chemistry & Engineering* **2020**, *8*, 10702–10713.

(36) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Abe, H.; Yamato, I.; Doi, Y. Biochemical and molecular characterization of poly(aspartic acid) hydrolase-2 from sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 1285–92.

(37) In *Meyler's Side Effects of Drugs (Sixteenth Edition)*, Aronson, J., Ed., Sixteenth Edition; Elsevier: Oxford, 2016, pp 115–116.

(38) Rowsell, S.; Pauptit, R.; Tucker, A.; Melton, R.; Blow, D.; Brick, P. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure* **1997**, *5*, 337–47.

(39) Nocek, B.; Reidl, C.; Starus, A.; Heath, T.; Bienvenue, D.; Osipiuk, J.; Jedrzejczak, R.; Joachimiak, A.; Becker, D.; Holz, R. Structural Evidence of a Major

Conformational Change Triggered by Substrate Binding in DapE Enzymes: Impact on the Catalytic Mechanism. *Biochemistry* **2018**, *57*, 574–584.

(40) Nocek, B.; Gillner, D.; Fan, Y.; Holz, R.; Joachimiak, A. Structural basis for catalysis by the mono- and dimetalated forms of the dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase. *J Mol Biol* **2010**, *397*, 617–26.

(41) Rawlings, N.; Barrett, A.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **2012**, *40*, 343–350.

Chapter 2

Structural characterization of *Sphingomonas* sp. KT-1 PahZ1 catalyzed biodegradation of thermally synthesized poly(aspartic acid)

Chad A. Brambley, Austin L. Bolay, Henry Salvo, Amanda L. Jansch, Tarah J. Yared, Justin M. Miller, Jamie R. Wallen, and Mitch H. Weiland

Reprinted (adapted) with permission from ACS Sustainable Chem. Eng. 2020, 8, 29, 10702–10713. Copyright 2020 American Chemical Society.

2.1 Abstract

Poly(aspartic acid) (PAA) is a green alternative to non-biodegradable poly(carboxylates) and has applications in both industrial and biomedical settings. PAA is synthesized by heating monomeric aspartic acid to yield a polysuccinamide that can be ring opened to yield thermal PAA composed of 30% α-amide and 70% β-amide linkages. Here we report the first X-ray crystal structure of a poly(aspartic acid) hydrolase from the bacteria Sphingomonas sp. KT-1 ($\text{PahZ1}_{\text{KT-1}}$) which functions to degrade synthetic PAA to oligo(aspartic acid) by selective cleavage of β -amide linkages. The structure was solved to 2.45 Å and

shows a dimeric assembly where each monomer maintains an α/β hydrolase fold with a prominent, positively lined trough responsible for binding the anionic polymeric substrate. The putative catalytic sites of each monomer lie at the surface of the enzyme on opposite faces. The dimeric interface, as supported by SAXS/MALS data, is primarily hydrophobic and is further stabilized by flanking hydrogen bonds. Molecular dynamics (MD) simulations support the previously determined specific cleavage of only the β-amide linkage through a conformational change that aligns the substrate with the active site Ser. This data provides a scaffold for further understanding the mechanism of PAA hydrolysis and opens the opportunity for using protein engineering to catalyze the biodegradation of other xenobiotics.

2.2 Introduction

Polycarboxylates, such as poly(acrylate) and its derivatives, are versatile water-soluble polymers that have found commercial applications in detergents, as antiscaling agents, and as superabsorbent materials.[1,2] These synthetic polymers are produced from fossil feedstocks, are relatively environmentally stable, and undergo minimal biodegradation.[3,4] However, concerns have been raised about the potential for these polymers leaching into ground or surface water.[5,6] While they do not currently present environmental concerns, there is a need to develop biodegradable alternatives. Poly(aspartic acid) (PAA) is a water soluble, biodegradable, eco-friendly polymer that has found success as a viable replacement for polycarboxylates and due to its naturally occurring components, it has found numerous applications in the medical field.[1,7,8]

PAA is thermally synthesized (tPAA) by incubating L-Asp to form a polysuccininamide that is subsequently ring-opened using sodium hydroxide. This process results in atactic tPAA having an equal distribution of racemized aspartates, irregular end groups, and branched units containing α- and β-amide linkages randomly distributed, with the β-amide linkage occurring 70% of the time, (Figure 2.1).[9] A river water bacterial strain, Sphingomonas sp. KT-1 was identified as being capable of degrading tPAA to monomeric aspartic acid.[10] Subsequent genetic and biochemical studies elucidated two different enzymes, poly(aspartic acid) hydrolase-1 (PahZ1$_{\text{KT-1}}$) and poly(aspartic acid) hydrolase-2 (PahZ1$_{\text{KT-1}}$), responsible for selective cleavage of the β-amide or both the α- and β-amides, respectively.[10–15] Additionally, another PahZ1 enzyme was identified in a different bacterial strain, Pedobacter sp. KP-2, but a corresponding PahZ2 was not reported in this strain.[10,14]

Complete degradation of tPAA is accomplished by the initial addition of PahZ1 to create oligo(aspartic acid) (OAA), which becomes the substrate for PahZ2 to create monomeric aspartic acid.[11] The present study focuses on PahZ1 from Sphingomonas sp. KT-1 (PahZ1$_{\text{KT-1}}$), a hydrolase that follows an endo-mode of hydrolysis to create OAA through specific cleavage of only the β-amide linkages.[12,13] PahZ1$_{\text{KT-1}}$ contains a lipase box consensus sequence (G-X-S-X-G/A)

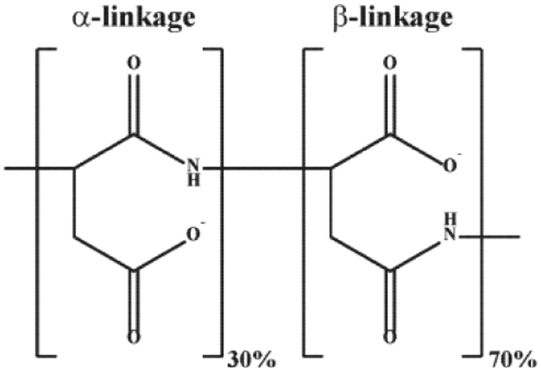


Figure 2.1: Representative poly(aspartic acid) substrate structure with α- vs β-linkages highlighted. Percentage frequencies for each linkage in a thermally synthesized PAA substrate are provided.

which corresponds to the active site serine of other serine proteases and is susceptible to inhibition by diisopropyl fluorophosphate (DFP) and phenylmethylsulfonyl fluoride (PMSF).[12,15] Mutation of this catalytic serine to an alanine impaired function as a PAA hydrolase. Sequence alignments with PahZ1KT-1 shows homology with poly(3-hydroxybutyric acid) (PHB) depolymerases which contain a catalytic triad of Ser/His/Asp residues.[12] Therefore, based on the literature PahZ1$_{\text{KT-1}}$ likely functions as a serine protease.

Herein, we report the first X-ray crystal structure of a poly(aspartic acid) hydrolase, PahZ1$_{\text{KT-1}}$, along with biophysical data to further characterize the enzyme. PahZ1$_{\text{KT-1}}$ is a dimeric protein that contains an α/β hydrolase fold typical of other serine proteases. An electrostatic map of the structure shows a prominent, positively charged trough that is likely responsible for directing substrate binding of the negatively charged tPAA. The dimeric structure is stabilized primarily through hydrophobic interactions that are flanked by hydrogen bonds.

Molecular dynamics (MD) methods have revealed a conformational change that provides a structural model to explain the reported cleavage specificity toward the β-amide linkage substrate. Using the PahZ1$_{\text{KT-1}}$ with a substrate harboring the β-amide linkage docked, MD simulations result in the observation of the substrate being properly positioned for reaction with the conserved catalytic serine residue. Collectively, the structure/function relationships presented here may aid in further understanding the catalytic mechanism of PahZ1 and serve as a guide for future protein engineering to catalyze the biodegradation of additional polymers that may accumulate in the environment.

2.3 Experimental Section

2.3.1 PahZ1$_{\text{KT-1}}$ Construct Design.

The PahZ1$_{\text{KT-1}}$ gene codes for 314 amino acids; however, the first 35 amino acids correspond to a signal sequence and residues 36-40 are cleaved by a protease during expression in BL21.[12] Therefore, the nucleotide sequence corresponding to amino acid residues 41-314 of PahZ1$_{\text{KT-1}}$[16] was cloned in the pET15b vector using the 5′ NdeI and 3′ XhoI restriction sites. DNA synthesis and cloning was performed by GenScript (Piscataway, NJ). Cloning into this vector added the following N-terminal residues: MGSSHHHHHHSSGLVPRGSHMSKG. This gave the construct a N-terminal 6x His tag for Ni-NTA purification. The start of the mature protein is underlined above. Site-directed mutagenesis to create the mutant constructs (S157A, R246K, R246E, R246A, and D225A) was performed by GenScript using the PahZ1$_{\text{KT-1}}$ template described above.

2.3.2 PahZ1$_{\text{KT-1}}$ and PahZ1$_{\text{KT-1}}$ Mutant Expression and Purification.

All constructs were transformed into BL21(DE3) *Escherichia coli* cells (New England Biolabs, Ipswich, MA), and expression conditions were the same for the wild-type and the mutants. A single colony was used to inoculate 50 mL of LB media containing 50 µg/mL final carbenicillin. Cells were incubated overnight at 37 °C, shaking at 225 rpm. The next day, 12.5 mL of this growth was used to inoculate 500 mL of LB-carb media in a 2 L Erlenmeyer flask, and the cells were grown to an O.D.$_{600}$ = 1.0 by keeping at the same temperature and shaking. Protein expression was induced with 0.1 mM isopropyl β-d-1-thiogalactopyranoside (GoldBio, St. Louis, MO) final, and the flask was incubated at 20 °C for 16 h with shaking at 225 rpm. The cells were harvested by centrifugation at 10,000g for 10 min at 4 °C. Cell pellets were stored at -20 °C.

Pelleted cells were resuspended in 50 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 1 mM 4-(2-aminoethy)benzenesulfonyl fluoride hydrochloride (AEBSF; EMD Millipore, Billerica, MA). This step was also performed in the absence of AEBSF and activity assays indicated that the inhibitor did not affect enzymatic activity, data not shown. The cells were lysed by sonication over 45 min at 23 °C. Lysozyme (Thermo Scientific, Rockford, IL) at 10 µg/mL final and 125 units/mL of turbonuclease (Accelagen, San Diego, CA) were added and allowed to rock for 45 min to enhance lysis and remove genomic DNA, respectively. The cell lysate was clarified by centrifugation at 8000g at 4 °C and subsequently loaded on a HisPur Ni-NTA 5 mL cartridge (Thermo Scientific) equilibrated in 50 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole. Protein was eluted from the column with 50 mM Tris, pH 8.0, 300 mM NaCl, and 500 mM imidazole using a gradient from 0 to 100% over 60 min. The second half of the eluted peak corresponding to PahZ1$_{\text{KT-1}}$ was collected. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis analysis showed the protein to be >98% pure after this single column. Protein was then dialyzed against 50 mM potassium phosphate, pH 7.4 for use in gel-permeation chromatography (GPC) assays or 20 mM Tris, pH 8.0 for crystallization and then stored on ice.

2.3.3 GPC Activity Assays

The tPAA substrate was synthesized according to Andrzejak et al.[17] and solubilized to 100 mg/mL in 1 mL of 50 mM potassium phosphate, pH 7.4, and then 3 mg of thyroglobulin (MilliporeSigma, Burlington, MA) was added to create this stock solution. Activity assays were set up as 100 µL reactions containing PahZ1$_{\text{KT-1}}$ and 24.5 µL (2.45 mg) of PAA/thyroglobulin stock solution. Samples were incubated at 37 °C for 20 and 48 h, followed by incubation at 70 °C for 10 min to stop the reaction. Incubation of PahZ1$_{\text{KT-1}}$ at 70 °C is sufficient to cause precipitation of the enzyme. Samples were clarified by centrifugation and stored at -80 °C until they were loaded on the GPC. Activity of PahZ1$_{\text{KT-1}}$ was determined using a Shimadzu GPC with a Yarra 3 µm SEC-200, 300 x 7.8 mm

(Phenomenex, Torrance, CA) column equilibrated in 50 mM potassium phosphate, pH 7.4 and a refractive index detector. Each injection was 10 µL, and the thyroglobulin served as an internal loading control. PahZ1$_{\text{KT-1}}$ degrades PAA, a common OAA product peak is present at a retention time of 10.9 min. Percent activity was calculated using a ratio of OAA peak height to thyroglobulin peak height at 5.1 min. Relative activity of each mutant was compared relative to the wild-type enzyme which was set at 100% activity. Polymer molecular weights were determined by creating a calibration curve using polyethylene glycol EasiVial GPC/SEC Calibration Standards (Agilent, Santa Clara, CA) and using the LabSolutions GPC software (Shimadzu).

2.3.4 X-ray Structure Determination and Refinement.

PahZ1$_{\text{KT-1}}$ crystallization screening was initially performed at Hauptman-Woodward Institute High-Throughput Crystallization Screening Center[18] with PahZ1$_{\text{KT-1}}$ in 20 mM Tris, pH 8.0 at 8.4 mg/mL, with crystals observed in a variety of conditions. Single PahZ1$_{\text{KT-1}}$ crystals were obtained by hanging-drop vapor diffusion at 22 °C by mixing 2 µL of the protein (8 mg/mL PahZ1$_{\text{KT-1}}$) with a 2 µL reservoir placed over a 500 µL reservoir containing 15% polyethyleneglycol 3350, 2% 1,4-dioxane, and 0.1 M Tris pH 8.0. Crystals appeared within a few days and grew to full size in approximately 1 to 2 weeks. For preparation for data collection, crystals were transferred to a solution containing a 1:1 mixture of mineral oil/paratone-N (50% final concentration for both solutions), then flash-cooled in a gaseous nitrogen stream at 100 K. X-ray data were collected at the Wake Forest University Center for Structural Biology using a Rigaku/MSC MicroMax-007 rotating anode generator and a Pilatus3R pixel array detector. Data were processed using HKL3000.[19]

The structure of PahZ1$_{\text{KT-1}}$ was determined by molecular replacement using the coordinates of the putative poly(3-hydroxybutyrate) depolymerase from Bordetella parapertussis (PDBID 3D0K) as a starting model. Molecular replacement was performed using the program Phaser[20] in the Phenix software suite.[21] Phaser correctly located the position of four PahZ1$_{\text{KT-1}}$ molecules in the asymmetric unit. The program Autobuild[22] was able to successfully build residues 34-294 of chain A, 30-294 of chain B, 34-158 and 203-295 of chain C, and 33-295 of chain D. Autobuild also located the positions of 393 water molecules. Missing residues were manually built into density-modified electron density maps generated in Phenix. Manual structure building was followed by xyz coordinate, real space, and individual B-factor refinement using the program phenix.refine. Non-crystallographic and secondary structure restraints were included during refinement. A summary of data collection and refinement statistics are provided in Table 3.1. X-ray structure figures, as well as electrostatic surface potential maps, were generated using the program PyMol (the PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC.).

Table 2.1: Data Collection and Refinement Statistics for the $\mathrm{PahZ1_{KT\text{-}1}}$ Crystal Structure[a]

PahZ1KT-1	
Data Collection	
space group	$P2_12_12_1$
Cell Dimensions	
a, b, c (Å)	81.75, 87.00, 171.10
α,β,γ (deg)	90, 90, 90
resolution (Å)	50-2.45 (2.49-2.45)
unique reflections	45,675 (2278)
CC1/2	1.00 (0.731)
R_{merge} (%)	16.0 (75.6)
I/σ	13.5 (2.6)
completeness (%)	97.9 (97.6)
redundancy	5.0 (5.2)
Refinement	
resolution (Å)	47.69-2.45 (2.50-2.45)
no. reflections	44,451
R_{work}/R_{free} (%)	23.58/29.03 (26.83/32.66)
No. Atoms	
protein	8408
water	425
B-Factors	
protein	34.69
water	33.96
Stereochemical Ideality	
bond lengths (Å)	0.006
bond angles (deg)	0.922
ϕ, ψ most favored (%)	97.82
ϕ, ψ allowed (%)	1.99
ϕ, ψ outliers (%)	0.19

[a] Values in parentheses are for the highest resolution shell. Coordinates have been deposited in the Protein Data Bank as entry 6VE6.

2.3.5 Small-Angle X-ray Scattering and Multi-Angle Light Scattering Data Acquisition in-Line with Size-Exclusion Chromatography

A 60 µL sample containing 8 mg/mL PahZ1$_{\text{KT-1}}$ was prepared in 20 mM Tris pH 7.0, 100 mM NaCl, and 1 mM DTT. Small-angle X-ray scattering coupled with multi-angle light scattering in-line with size-exclusion chromatography (SEC-SAXS-MALS) data was collected at the ALS beamline 12.3.1 LBNL Berkeley, California.[23] The X-ray wavelength was set at λ = 1.127 Å, and the sample-to-detector distance was 2100 mm resulting in scattering vectors, q, ranging from 0.01 to 0.4 Å^{-1}. The scattering vector is defined as $q = 4\pi \sin \theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20 °C and data was collected and processed as previously described.[24] A SAXS flow cell was directly coupled with an online Agilent 1260 Infinity HPLC system using a Shodex KW803 column. The column was equilibrated with a running buffer (20 mM Tris pH 7.0, 100 mM NaCl, 1 mM DTT) with a flow rate of 0.6 mL/min. Data were processed using the program SCÅTTER.

2.3.6 SEC to Assess PahZ1$_{\text{KT-1}}$ Dimer Stability.

To determine if the dimeric interface of the wild-type PahZ1$_{\text{KT-1}}$ was susceptible to increasing NaCl concentration, 1 mL of 5 mg/mL samples were loaded onto a HiPrep Sephacryl S-200 HR SEC column (GE Healthcare, Chicago, IL) attached to an ÄKTA Purifier (GE Healthcare) that was equilibrated in 50 mM Tris, pH 7.4

and NaCl concentrations of 0, 150, and 500 mM. Protein was detected by UV at λ 280 to determine the elution volumes.

2.3.7 Molecular Dynamics

Docking. Protein data bank (PDB) files for both α- and β-linked tripeptide polyaspartic acid substrates (α- and β-PAA, respectively) were designed in Avogadro molecular modeling software.[25] The β-PAA substrate has been modeled as a representative fragment of a β-linkage contained within a tPAA substrate. Given the irregular nature of tPAA with regards to linkage (α vs β) and branches, the β-PAA substrate was designed with two β-linkages and two α-linkages connected by an internal branch occurring at the middle C α carbon position (Figure A.2). To optimally recapitulate the hydrogen bonding capability of tPAA, all amide functional groups have been N-methylated. Additionally, the N-terminus has been replaced with a methyl group to remove positive charge that may destabilize tripeptide binding to the positively charged PahZ1$_{\text{KT-1}}$ binding surface. In contrast, the α-PAA tripeptide substrate contains only α-linkages occurring through the α-carboxylate functional group. The ligand models were subsequently prepared alongside a PDB of a PahZ1$_{\text{KT-1}}$ dimer in UCSF Chimera to generate the input files for docking via Autodock Vina. The search box was placed such that the coordinates included the proposed active-site residues. Docking was then performed for both ligands, each generating nine modes of coordinates. PDB files of resulting structures exhibiting the highest scores were exported for further analysis.

Molecular Dynamics. All MD simulations were carried out with the Gromacs software package.[26–29] Familiarity with Gromacs software and workflow is assumed. Specific Gromacs modules are italicized for clarity. For all conditions described, a PDB file of the PahZ1$_{\text{KT-1}}$ dimer structure was used as input for the Gromacs module *pdb2gmx*, which yields the initial system topology. For PAA substrate-bound simulations, included topologies for all ligands were generated via the Swiss-param server courtesy of the Swiss Institute of Bioinformatics. The Charmm-36 all-atom force field[30,31] was chosen as well as the TIP3P[32,33] water model. The resulting Gromacs compatible structures were then placed into a rhombic dodecahedron water box and subsequently solvated with simple point charge water with neutralizing ions added. Following a short energy minimization, 100 ps of constant volume *NVT* [number of particles (N), volume (V), and temperature (T)] equilibration was performed, using a velocity rescaling thermostat to insure the system temperature stabilized at 310 K. An additional 100 ps of constant pressure *NPT* [number of particles (N), pressure (P), and temperature (T)] equilibration MD was then performed using Berendsen pressure coupling[34] to bring the system to 1.0 bar. Position restraints were then removed and 200 ns of unrestrained production MD was performed for each condition. During production MD, the Nosé-Hoover thermostat[35,36] and Parinello-Rahman barostat[37,38] were used to maintain system temperature and pressure, respectively.

Center-of-Mass Pulling and Umbrella Sampling. The system was prepared as reported in the previous section with notable variations. After passing *pdb2gmx* to generate the initial topology and the Gromacs structure, chain B of the PahZ1$_{\text{KT-1}}$

dimer was position restrained by inserting the *pdb2gmx*-generated position restraint include topology. Next, the structure was inserted via editconf into a water box measuring 8.00 x 25.00 x 8.00 nm to insure adequate space for pulling. Energy minimization and equilibration steps were carried out as already described. Pull code was used that designated chain B to be pulled along the Y direction with a force constant of 1000 kJ mol^{-1} nm $^{-}$2 for 500 ps. Extraction of center-of-mass (COM) distances was then carried out on the resultant pulling trajectory by passing the *-sep* argument to *trjconv*. A total of 33 COM distance windows with 0.1 nm spacing were then selected and used to generate binary run files for each configuration by passing the Gromacs structure files as input to *grompp*. Finally mdrun was invoked for each of the 33 instances and the simulations carried out for 10 ns. The resultant pull force data files for each COM window was passed along with the associated binary run files to wham,[39] which utilizes the weighted histogram analysis method to generate a plot of potential mean force (PMF).

2.4 Results

2.4.1 Crystal Structure of PahZ1$_{\text{KT-1}}$

Results from the Hauptman-Woodward Institute screening revealed that the PahZ1$_{\text{KT-1}}$ protein crystallized in a variety of conditions, and we pursued a condition that gave large, single crystals. PahZ1$_{\text{KT-1}}$ crystals diffracted to 2.45 Å resolution, and the structure was determined using molecular replacement,

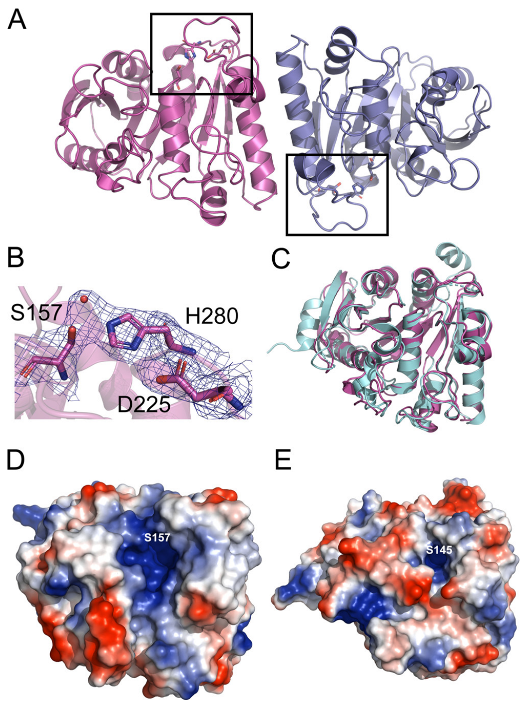


Figure 2.2: Crystal structure of PahZ1$_{\text{KT-1}}$. (A) Dimeric assembly of PahZ1$_{\text{KT-1}}$ is shown, with monomers colored magenta and slate blue. The active sites of PahZ1$_{\text{KT-1}}$ (boxed region of each monomer) lie on the surface on opposite faces of the enzyme. (B) Zoomed-in view of the magenta active site that is boxed in panel (A). Residues that make up the catalytic triad (S157, D225, and H280) are shown, with $2F_o$ - F_c electron density (in blue) contoured at 1 sigma. A water molecule (red sphere) was modeled in the active site of the enzyme, positioned between the catalytic S157 and H280 residues. (C) Superposition of a PahZ1$_{\text{KT-1}}$ monomer (magenta) with the putative poly(3-hydroxybutyrate) depolymerase from B. parapertussis (cyan). The two structures superimpose with an rmsd of 1.7 Å. (D) Electrostatic surface potential map of PahZ1$_{\text{KT-1}}$ reveals a long positively charged cleft that originates at the enzyme active site. The location of the catalytic serine (S157) is labeled for clarity. (E) Electrostatic surface potential map of the *B. parapertussis* poly(3-hydroxybutyrate) depolymerase shows that while the active site is also positively charged, the enzyme lacks both the open cleft and additional positive charge that surrounds the PahZ1$_{\text{KT-1}}$ active site shown in panel (D). The location of the catalytic serine (S145) is labeled for clarity.

Table 1. The structure contains four PahZ1$_{\text{KT-1}}$ monomers in the asymmetric unit, and Figure 2.2 shows the structure that we believe is the biological PahZ1$_{\text{KT-1}}$ This dimeric assembly is supported by SAXS solution studies that are discussed below. The two monomers of the dimer are related by a two-fold rotational symmetry at the dimer interface, and an α helix spanning residues 244-265 from each monomer subunit make up the majority of the dimer interface. PahZ1$_{\text{KT-1}}$ adopts a canonical α/β hydrolase fold (Figure 2.2A) with the putative catalytic triad, based upon sequence alignments with PHB family members,[12] located on the surface of the enzyme and consisting of residues S157, D225, and H280 (Figure 2.2B). Given the two-fold rotational symmetry, the catalytic triad of each monomer lies on opposite faces of the dimeric assembly (black boxes in Figure 2.2A).

A structural comparison of PahZ1$_{\text{KT-1}}$ using the DALI server[40] reveals that the closest structural homolog is a putative PHB depolymerase from *B. parapertussis* (PDBID 3D0K, Z-score = 36.3), which was the starting model used during molecular replacement. PahZ1$_{\text{KT-1}}$ and the *B. parapertussis* PHB depolymerase share 33% primary sequence identity and superimpose well (root mean square deviation = 1.7 Å, Figure 2.2C), and the PHB depolymerase conserves the Ser/Asp/His catalytic triad residues (discussed below). Other top hits identified from the DALI server consist of various hydrolase enzymes that range from 10 to 16% primary sequence identity to PahZ1$_{\text{KT-1}}$ and Z-scores from 16.0 to 18.6. While the structural superposition shown in Figure 2.2C shows a similar fold between

PahZ1$_{\text{KT-1}}$ and the *B. parapertussis* PHB depolymerase, electrostatic surface potential maps for the two enzymes show a marked difference at the enzyme active sites. The electrostatic surface potential map for PahZ1$_{\text{KT-1}}$ shows a large, open, positive trough that leads toward the enzyme active site (Figure 2.2D), which is expected for substrate binding given the significant negative charge imparted by the PAA substrate. While the PHB depolymerase active site is also positively charged (Figure 2.2E), the positive charge near the active site is much less than observed with $_{\text{KT-1}}$. Additionally, while the PahZ1$_{\text{KT-1}}$ active-site binding pocket is open and appears to be capable of binding large polymers, the PHB depolymerase active-site binding pocket is capped by a region of negative charge. These features highlight the differing active-site architectures that would allow these enzymes to recognize different substrates.

PahZ1$_{\text{KT-1}}$ sequence analysis revealed that it shared similarity with the catalytic domain of the *Paucimonas lemoignei* PhaZ7 PHB depolymerase.[12] A comparison of both the PahZ1$_{\text{KT-1}}$ and *B. parapertussis* PHB structures to the structurally well- characterized PhaZ7 PHB depolymerase from *P. lemoignei* reveals interesting structural observations. A superposition of the three structures show that the catalytic triad Ser/His/Asp of the enzymes superimpose well (Figure A.1). However, there are two key differences in the structures that show the *P. lemoignei* enzyme is distinct from the other two. First, a structure of the *P. lemoignei* PHB depolymerase bound to PHB substrate shows that the enzyme binds PHB in a cleft formed by a series of aromatic residues (Y105, Y176, Y189, Y190, and

W207).[41] While the PHB binding cleft is located away from the active-site catalytic triad, the authors propose that a flexible loop made of residues 281-294 could deliver the PHB substrate to the active site for catalysis. Figure A.1 shows that both the PahZ1$_{\text{KT-1}}$ and *B. parapertussis* PHB structures lack this PHB binding cleft. The equivalent of the *P. lemoignei* Y105 residue is L108 and E96 in the PahZ1$_{\text{KT-1}}$ and *B. parapertussis* PHB structures, respectively. The additional four *P. lemoignei* aromatic residues are missing altogether in both the PahZ1$_{\text{KT-1}}$ and the *B. parapertussis* structures (Figure A.1). The second major observation is that both the PahZ1$_{\text{KT-1}}$ and *B. parapertussis* enzymes conserve an arginine residue in the active site (R246 and R234, respectively), and our MD and mutagenesis studies support an essential role for R246 in the PahZ1$_{\text{KT-1}}$ catalytic cycle (discussed below). In the *P. lemoignei* PHB depolymerase structure, an aspartic acid (D256) is at the equivalent position of PahZ1$_{\text{KT-1}}$ R246 (Figure A.1). These observations highlight key structural differences between PahZ1$_{\text{KT-1}}$ and the *P. lemoignei* PHB depolymerase that allow for the recognition of different substrates, and it also appears that the enzymes utilize a different catalytic mechanism during turnover. The structural comparisons also bring up the question of whether the *B. parapertussis* enzyme is truly a PHB depolymerase. The *B. parapertussis* structure was determined as part of a structural genomics project, and to our knowledge the enzyme has not been biochemically characterized. The top hits of a DALI search[40] for the *B. parapertussis* structure reveal a general series of esterases and hydrolases, with no clear suggestion of a natural substrate. The PhaZ7 PHB depolymerase is the 123rd hit in the DALI search, with a Z-score of 11.9, a root mean square deviation = 3.6 Å, and 13% identity. Of the 169 hits provided in the DALI search output, the *P. lemoignei* enzyme is the only PHB depolymerase observed. Given the lack of the PHB binding cleft, as well as the conserved R234 in the active site, we hypothesize that the *B. parapertussis* enzyme recognizes a substrate other than PHB and utilizes a catalytic mechanism similar to PahZ1$_{\text{KT-1}}$. Although they may use a similar catalytic mechanism, the differences in active-site architectures shown in Figure 2.2D,E provide strong evidence that PahZ1$_{\text{KT-1}}$ and the *B. parapertussis* recognize different substrates.

2.4.2 SEC-SAXS-MALS Analysis of PahZ1$_{\text{KT-1}}$

To confirm that our structure shown in Figure 2.2A is correct and that PahZ1$_{\text{KT-1}}$ is indeed a dimer in solution, we analyzed the PahZ1$_{\text{KT-1}}$ enzyme using SEC-SAXS-MALS. The monomeric mass of PahZ1$_{\text{KT-1}}$ is predicted to be 32.7 kDa, and our MALS data, which shows a PahZ1$_{\text{KT-1}}$ mass of 60.9 kDa in solution, is consistent with a PahZ1$_{\text{KT-1}}$ dimer. Both Guinier and Kratky plots from the SAXS data show that PahZ1$_{\text{KT-1}}$ is well-behaved and folded in solution, with a radius of gyration (R_G) of 27.03 Å and a volume of 1.2 x 10^5 nm^3 . A pair distribution function *(P(r))* calculated in SCÅTTER shows that PahZ1$_{\text{KT-1}}$ has a maximal dimension (D_{max}) of $\sim$89 Å (Figure 2.3, inset). This agrees well with the PahZ1$_{\text{KT-1}}$ crystal structure, where measurements made along the longest dimension of the dimer (from residue L32 of the magenta monomer to P30 of the blue monomer in Figure 2.2A) give a distance of $\sim$86 Å. We compared the PahZ1$_{\text{KT-1}}$ SAXS scattering profile with the PahZ1$_{\text{KT-1}}$ dimeric structure presented in Figure 2.2 using FoXS,[42,43] and

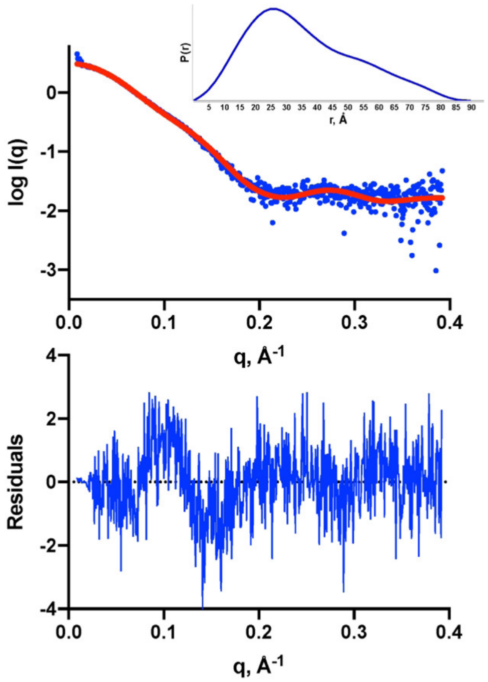
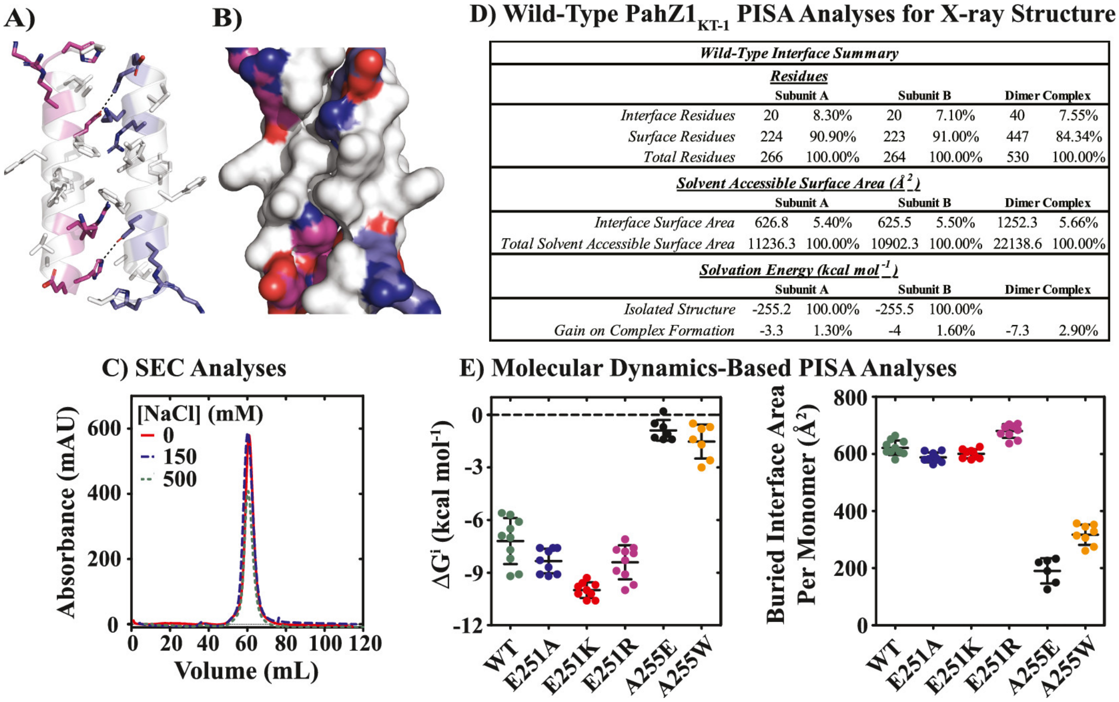


Figure 2.3: SAXS analysis of PahZ1$_{\text{KT-1}}$. Experimental (blue) and theoretical (red) SAXS curves are shown along with fit-residuals. The theoretical curve, which was calculated from the dimer presented in Figure 2.2, matches the experimental curve well with χ^2 = 1.23. Inset: $P(r)$ function calculated from the experimental data. The distance r where $P(r)$ approaches zero intensity is the maximal dimension of the macromolecule (D_{max}). For PahZ1$_{\text{KT-1}}$, the D_{max} value of ∼89 Å agrees well with the PahZ1$_{\text{KT-1}}$ crystal structure.

the results show that the scattering profile agrees well with the crystal structure (Figure 2.3, χ^2 = 1.23). Overall, the results show that our structure presented in Figure 2.2 provides an optimal fit to the SAXS scattering data and provides further support for a PahZ1$_{\text{KT-1}}$ dimer in solution.



D) Wild-Type PahZ1$_{\text{KT-1}}$ PISA Analyses for X-ray Structure

Wild-Type Interface Summary						
Residues						
	Subunit A		**Subunit B**		**Dimer Complex**	
Interface Residues	20	8.30%	20	7.10%	40	7.55%
Surface Residues	224	90.90%	223	91.00%	447	84.34%
Total Residues	266	100.00%	264	100.00%	530	100.00%
Solvent Accessible Surface Area ($\AA^2$)						
	Subunit A		**Subunit B**		**Dimer Complex**	
Interface Surface Area	626.8	5.40%	625.5	5.50%	1252.3	5.66%
Total Solvent Accessible Surface Area	11236.3	100.00%	10902.3	100.00%	22138.6	100.00%
Solvation Energy (kcal mol^{-1})						
	Subunit A		**Subunit B**		**Dimer Complex**	
Isolated Structure	-255.2	100.00%	-255.5	100.00%		
Gain on Complex Formation	-3.3	1.30%	-4	1.60%	-7.3	2.90%

Figure 2.4: PahZ1$_{\text{KT-1}}$ dimer interface is primarily nonpolar. (A) Dimeric interface of PahZ1$_{\text{KT-1}}$ is shown with polar contributions from each monomer colored magenta and slate blue. Nonpolar dimer interface features are colored white. PahZ1$_{\text{KT-1}}$ dimer interface is represented in cartoon (A) and surface (B) representations. (C) SEC experiments performed in the presence of 0, 150, and 500 mM NaCl are represented as solid red, broken blue, and dashed green lines, respectively. Elution volumes are observed to be independent of [NaCl] and are each consistent with the molecular weight of a PahZ1$_{\text{KT-1}}$ dimer. (D) Summary of Proteins, Interfaces, Structures, and Assemblies (PISA) analyses for the wild-type PahZ1$_{\text{KT-1}}$ X-ray structure. (E) Extended PISA analyses for PahZ1$_{\text{KT-1}}$. PahZ1$_{\text{KT-1}}$ variants were generated in silico and equilibrated using MD techniques. Extracted frames from each molecular dynamics trajectory were subjected to PISA analysis. The resulting estimates of gain in solvation free energy for complex formation, ΔG^i, and buried interface area per monomer are plotted separately as colored spheres labeled by mutation. All replicate values are plotted with mean and standard deviations indicated by bars.

2.4.3 PahZ1$_{\text{KT-1}}$ Dimer Interface is Primarily Non-polar.

Analysis of the PahZ1$_{\text{KT-1}}$ structure suggests that complex formation occurs via interaction between two copies of an α-helix represented by residues R246-L264 on both subunits, wherein each helix is positioned antiparallel relative to the other. Figure 2.4A,B highlights specific features of the dimer interface wherein nonpolar features and electrostatic interactions are colored white and magenta-blue, respectively. Two electrostatic interactions are observed at the terminal ends of the helix-helix structure between H262 and E251, with an interaction distance equal to 3 Å. Because of the antiparallel orientation of the helix-helix interface, these residues interact identically at each terminal end. Given that the majority of residues observed at the dimer interface are nonpolar, we applied SEC methods in the presence of 0, 150, and 500 mM NaCl to determine the impact of elevated salt concentrations on PahZ1$_{\text{KT-1}}$ complex molecular weight. Figure 2.4C highlights that the apparent molecular weight, 65 kDa, is independent of [NaCl], which supports nonpolar interactions as the primary driver of dimer formation.

Additional analysis of the PahZ1$_{\text{KT-1}}$ dimer interface was performed using the Protein Data Bank in Europe Proteins, Interfaces, Structures, and Assemblies (PDBePISA)[44] interactive tool. The results of PDBePISA analysis of the PahZ1$_{\text{KT-1}}$ structure reported here are summarized in Figure 2.4D, where data is provided to describe each individual subunit as well as the overall dimer complex. First, 7.6% of the 530 residues represented by both chains occur at the subunit interface. Figure 2.4D highlights that subunits A and B contribute 626.8 Å 2 (5.4% of subunit A surface area) and 625.5 Å 2 (5.5% of subunit B surface area), respectively, thereby yielding an overall interface area equal to 1252.3 Å. The average buried surface area formed on interfacing monomers can be readily calculated as the average of the individual interface areas for each subunit, where this value is equal to 626.2 Å 2 . PDBePISA also calculates solvation free-energy gain associated with complex formation by calculation of the difference between solvation energies for the isolated chains and the complex structure. Based on this, the overall solvation free-energy gain associated with complex formation for wild-type PahZ1$_{\text{KT-1}}$ is equal to -7.3 kcal mol -1 , where negative values for ΔG^i correspond to hydrophobic interfaces. Salt bridges observed reflect the symmetric nature of the PahZ1$_{\text{KT-1}}$ interface and include H262:E251 and R258:E251. Detailed analysis of the dimer interface identifies residues M240 to H245, L247 to A248, E251 to F252, A255 to Y256, R258 to K259, and H262 to E263 as interfacing residues. Based on the calculated ΔG^i and the observation that 35% of the interface residues are nonpolar, analysis using the PDBePISA tool suggests that the dimer interface is stabilized by nonpolar effects.

In order to better understand the contribution of polar and nonpolar residues to dimer stability, we generated a series of in silico PahZ1$_{\text{KT-1}}$ interface mutants for calculation of solvation free-energy gain, ΔG^i , by PDBePISA analysis. In brief, we generated five PahZ1$_{\text{KT-1}}$ mutants in the PyMol software package,[45] followed by energy minimization and equilibration using MD methods in the Gromacs[28,29] environment. All MD-based equilibrations were allowed to proceed to at least 50 nanoseconds beyond that required for the system to reach equilibration. Using the equilibrated system, 5-10 frames were extracted for each mutant and subjected to analysis using the PDBePISA tool, thereby yielding estimates for ΔG^i and interface area. Our rationale for this workflow was two-fold: (1) extended MD trajectories allow us to sample multiple conformations to allow for a dynamic observation of ΔG^i and (2) energy minimization and equilibration allows for conformational rearrangements necessary to compensate for interface mutations not found in the wild-type crystal structure. Figure 2.4E highlights the results of this analysis where wild-type PahZ1$_{\text{KT-1}}$ is observed with ΔG^i = -7.2 ± 0.4 kcal mol^{-1} and buried interface area per monomer = 622 ± 8 Å^2. All E251 mutations (E251A, E251K, and E251R) are qualitatively observed to display solvation free-energy gains and interface areas similar to wild-type enzyme. However, application of unpaired, one- or two-tailed t testing indicates with 95% confidence that all E251 mutants display marginally enhanced complex stability, which may be a consequence of conformational rearrangements not accessible with E251 intact. In contrast, A255E and A255W mutations each significantly disrupt complex stability as assessed by decreases in solvation free-energy gains equal to $\Delta\ \Delta G^i$ = 6.3 and 5.7 kcal mol^{-1} for each mutant, respectively (Figure 2.4E). All A255 mutations are observed to yield significant decreases in buried interface area per monomer as well, which is consistent with disruption to the nonpolar surface necessary for dimer stabilization. Taken together, these data strongly implicate nonpolar interface residues such as A255 in stabilization of the PahZ1$_{\text{KT-1}}$ dimer.

To increase confidence in the application of the PDBePISA tool toward conclusions regarding complex stability, we employed an alternative method to assess the free energy of PahZ1$_{\text{KT-1}}$ complex formation. We employed a steered MD (SMD) technique known as COM pulling to separate subunits toward the goal of calculating the free energy necessary for dimer dissociation. Figure 2.5A visualizes the start and end states or 53 ± 3 kJ mol^{-1} . We note the high degree of similarity between free-energy changes determined from either changes in solvation energy or PMF. However, we note that one should not expect the two estimates of free-energy change to be identical given differences in the methods of calculation. The data presented in Figures 4 and 5 collectively support a model wherein dimerization is driven largely by the hydrophobic effect, thereby predicting an energetically favorable association free energy that is controlled by entropic factors.

2.4.4 R246 Side-Chain Orientation Directs Orientation of Bound Ligand.

Previous studies examining the substrate specificity of PahZ1$_{\text{KT-1}}$-catalyzed PAA hydrolysis have demonstrated a preference toward cleavage of β-amide linkages.[13,15] To better understand how PahZ1$_{\text{KT-1}}$ discriminates between β-β and α-α linkages, we performed additional MD simulations with tripeptide substrates allowed to dock without positional constraints. In each docking attempt, we observe the tripeptide to associate with a cleft of positive charge adjacent to the catalytic triad (Figure 2.2D). To insure that each docked PahZ1$_{\text{KT-1}}$:tripeptide complex was stable, we performed an additional production MD simulation for 200 ns to allow for equilibration prior to structural analysis.

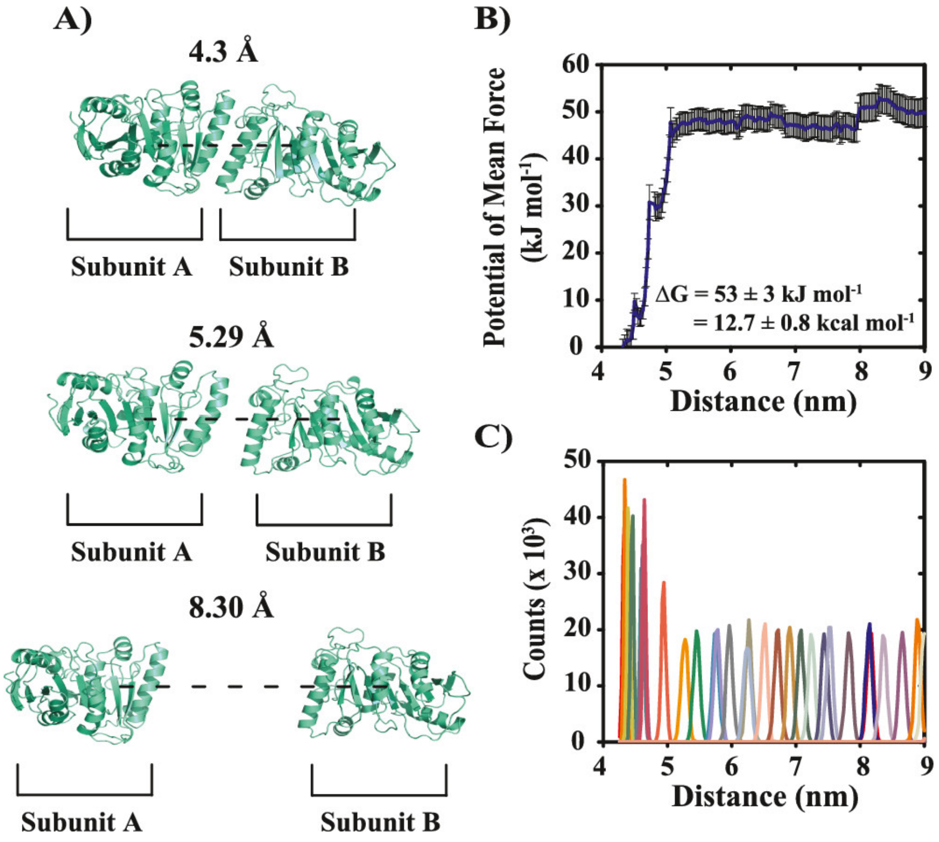


Figure 2.5: SMD techniques were utilized alongside umbrella sampling methods to observe PahZ1$_{\text{KT-1}}$ dimer dissociation. (A) Frames observed along the SMD pulling trajectory are shown. Subunit A is position restrained, while subunit B is pulled along a defined path by application of a static force vector. (B) Application of umbrella sampling techniques to frames along the SMD trajectory allow for calculation of PMF. The difference between maximum and minimum PMF values allows for estimation of $\Delta G = 53 \pm 3$ kJ mol^{-1} or 12.7 ± 0.8 kcal mol^{-1} . Error bars represent the results of bootstrapping analysis with 1000 replicates to insure statistical significance. (C) Weighted histogram analysis methods were applied to insure sufficient sampling along the SMD trajectory for performance of free-energy calculations.

Conditions examined include PahZ1$_{\text{KT-1}}$ in water with and without α- and β-PAA substrates bound to the proposed active site. All simulations were carried out for

200 ns of total production MD for all conditions with equilibrium convergence confirmed by root mean square deviation block averaging. Experiments were also performed in which polyalanine, poly(3-hydroxybutyrate), and polymalate were separately docked to PahZ1$_{\text{KT-1}}$, but none formed interactions that stably positioned these alternative substrates in a catalytically relevant orientation (Figure A.3). Figure 2.6A visually highlights our observation of reproducible interactions of α- PAA with arginine residues R246 and R74, which were not frequently replicated in β-PAA bound conditions. Together, these arginine residues represent part of a larger positively charged cleft on the surface of PahZ1$_{\text{KT-1}}$ to which negatively charged PAA substrates may bind.

To predict whether the identified arginine residues might contribute to PahZ1$_{\text{KT-1}}$ substrate specificity, we performed distance and hydrogen bonding analyses of the MD data. As presented above, vacuum electrostatic surface representations of the active-site surface with substrates docked are illustrated (Figure 2.6A). The positions of R246 and R74 relative to the substrate indicate conformational deviations between α-PAA and β-PAA bound structures. Consistent with this, root-mean-square fluctuation (RMSF) analysis of apo PahZ1$_{\text{KT-1}}$ compared with α-PAA and β-PAA bound structures suggests substrate-dependent dynamics for loop features near the proteolytic active site (Figure A.4). In particular, Figure 2.6B highlights hydrogen bonding activity of R246 at 150 ns for trajectories collected with either α-PAA and β-PAA substrates docked. In α-PAA bound conditions, R246 closely associates with and forms multiple hydrogen

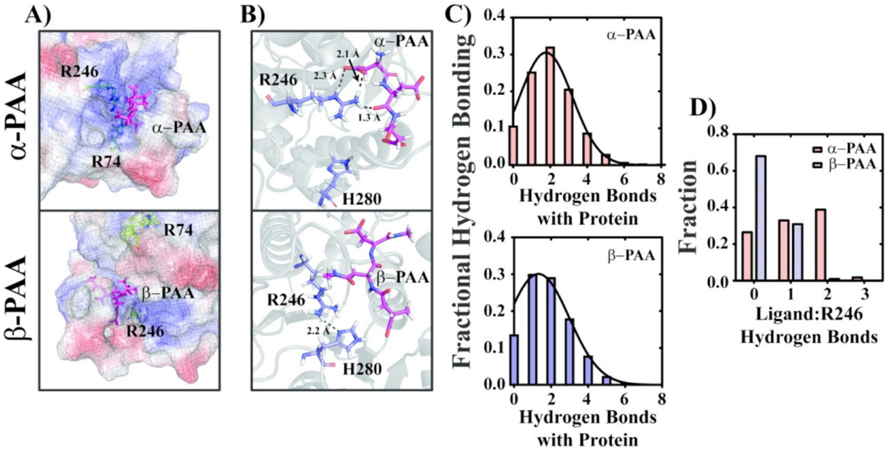
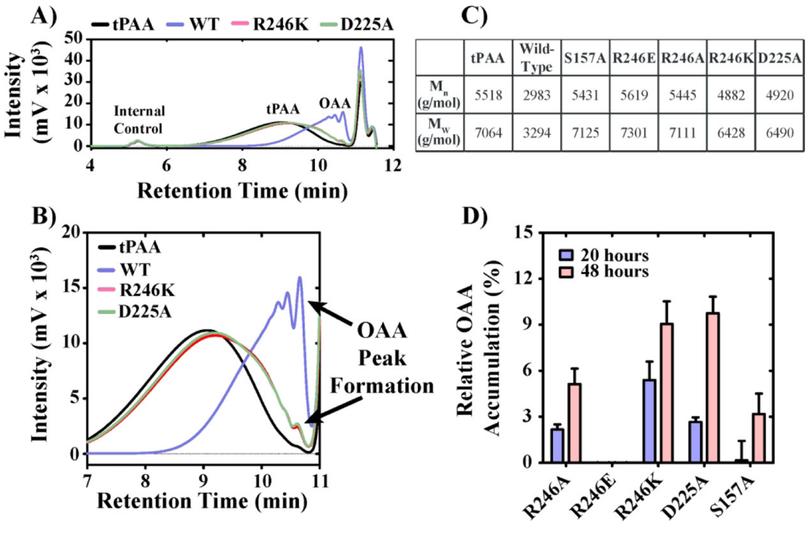


Figure 2.6: PAA ligands dock to a positively charged PahZ1$_{\text{KT-1}}$ surface. (A) Positions of α- and β-PAA are shown in relation to arginines R246 and R74 following 150 ns of unrestrained molecular dynamics. (B) R246 exhibits notable capacity to hydrogen bond with α-PAA as shown by distance measurements between hydrogen bonding partners of less than 2.5 Å. In the presence of β-PAA, R246 conversely displays increased side-chain dynamics involving interactions with D225, H280, and β-PAA. (C) Analysis of overall hydrogen bonding frequency between PahZ1$_{\text{KT-1}}$ and either α- or β-PAA substrate. Mean hydrogen bonds estimated by population frequency analysis are 2.0 ± 1.3 and 1.8 ± 1.2 for α- and β-PAA substrates, respectively. (D) Analysis of hydrogen bonding frequency between R246 and the PAA substrate reveal mean values equal to 1.2 ± 0.8 and 0.3 ± 0.5 hydrogen bonds formed between R246 and α- and β-PAA, respectively.

bonds with the substrate, as indicated by interatomic distances <2.5 Å between the hydrogen bonding partners. For conditions with bound β- PAA, R246 is observed to adopt multiple side-chain conformations that generally involve fewer hydrogen bonds with a bound ligand relative to conditions including α-PAA (Figure 2.6B). It is important to note here that the formation of additional interactions for α-PAA versus β-PAA with PahZ1$_{\text{KT-1}}$ does not imply greater affinity for a particular substrate because these are dynamic interactions overall. The R246 side-chain is also observed to form dynamic hydrogen bonding interactions with β-PAA, H280, and D225 during replicate 200 ns trajectories. On average, α- and β-PAA substrates are observed to form 2.0 $\pm$ 1.3 and 1.8 $\pm$ 1.2 hydrogen bonding interactions with PahZ1$_{\text{KT-1}}$, respectively, as illustrated by histograms presented in Figure 2.6C. However, R246 appears to be the dominant driver of ligand specificity given the observation of 1.2 $\pm$ 0.8 and 0.3 $\pm$ 0.5 hydrogen bonds formed between R246 and α- and β-PAA, respectively, over the course of each simulation. Analyses of the frequency distributions for hydrogen bonding between R246 and PAA ligand indicate no hydrogen bonding with the β-PAA substrate for $\sim$70% of the simulation time, whereas dynamic hydrogen bonding is observed between R246 and α-PAA with nearly equivalent occupancy in arrangements involving, 0, 1, and 2 hydrogen bonds (Figure 2.6D). These data strongly suggest a role for R246 in the specific positioning of α-PAA and β-PAA substrates within the catalytic site.

2.4.5 Wild-Type PahZ1$_{\text{KT-1}}$ Activity Requires the R246 Guanidinium Functional Group.

To investigate the role of R246 in PahZ1$_{\text{KT-1}}$-catalyzed tPAA hydrolysis, a GPC assay was employed to assess tPAA degradation for the wild-type enzyme alongside selected active-site variants. tPAA alone shows a polymer peak at $\sim$9 min (M_n = 5518 g/mol, M_w = 7064 g/mol) and when incubated with PahZ1$_{\text{KT-1}}$, the polymer begins to degrade as seen with a shift to longer retention times and, ultimately, the formation of a common OAA peak (M_n = 1494 g/mol, M_w = 1517 g/mol) with a retention time of ~11 min (Figure 2.7A,B).



	tPAA	Wild-Type	S157A	R246E	R246A	R246K	D225A
M_n (g/mol)	5518	2983	5431	5619	5445	4882	4920
M_w (g/mol)	7064	3294	7125	7301	7111	6428	6490

Figure 2.7: $PahZ1_{KT\text{-}1}$ GPC activity assay comparison. (A) Overlay of four chromatograms resulting from incubations with tPAA alone (black line), wild-type $PahZ1_{KT\text{-}1}$ and tPAA (blue line), R246K mutant with tPAA (red line), and D225A mutant with tPAA (green line). The internal loading control is thyroglobulin. Peaks corresponding to tPAA and OAA are labeled. (B) Zoomed in look at the common OAA peak shown in panel (A). (C) Summary of the tPAA polymer molecular weights determined from chromatogram analyses for conditions incubating tPAA alone and tPAA with wild-type $PahZ1_{KT\text{-}1}$ or with $PahZ1_{KT\text{-}1}$ mutant constructs. (D) OAA accumulation calculated relative to wild-type activity for time points collected after incubation for 20 (blue bar) or 48 (pink bar) hours. Error bars represent the standard deviation were calculated from duplicate measurements of the 20 h and triplicate measurements of the 48 h samples.

The activity of active-site variants of $PahZ1_{KT\text{-}1}$ constructs (R246A, R246E, R246K, D225A, and S157A) were compared against wild-type enzyme after 20 and 48 h incubations using OAA peak intensity as a signal. While wild-type $PahZ1_{KT\text{-}1}$ shows rapid formation of the OAA peak that ultimately stops accumulating after approximately 1 h (Figure A.5), this activity was significantly diminished when mutations were introduced and substantially longer incubation times were needed to show OAA peak presence, Figure 2.7B. Activity is also evident by the decreased tPAA molecular weights with the R246K and D225A mutants, Figure 2.7C. Figure 2.7D highlights the effect of these mutations on enzyme-catalyzed OAA production as a percentage of wild-type PahZ1$_{\text{KT-1}}$ hydrolysis activity. As expected, a S157A mutant exhibits significantly ablated hydrolysis activity relative to the wild-type enzyme. Incubation of this S157A variant for 48 h yields the observation of low-level hydrolase activity. Complete loss of catalytic function is not expected based on a similar mutation, S195A, in trypsin, which also exhibits low-level proteolytic functionality.[46] A PahZ1$_{\text{KT-1}}$ D225A mutation exhibits diminished activity relative to the wild-type enzyme, but still three-fold greater activity at the 48 h time point relative to that in the S157A construct. In contrast, mutation of R246 to alanine, glutamate, or lysine yielded three different outcomes, where R246E yielded complete loss of PahZ1$_{\text{KT-1}}$ hydrolase activity (Figure 2.7D) likely due to electrostatic repulsion of the carboxylate with the polyanionic substrate. R246A and R246K mutants retained function comparable to S157A and D225A, respectively. These data strongly suggest that R246 is involved in the PahZ1$_{\text{KT-1}}$ hydrolase mechanism.

Additional MD simulations were performed in order to develop an atomistic understanding of the impact of R246 mutations on active-site arrangement. We first examined the arrangement of the wild-type PahZ1$_{\text{KT-1}}$ catalytic triad observed when α- or β-PAA substrate is bound. Figure 2.8A demonstrates the occurrence of a catalytically relevant triad formed by D225, H280, and S157, which positions S157 directly adjacent to the scissile peptide bond of the bound β- PAA substrate. As noted above, the dynamic nature of the R246 side-chain allows for the guanidinium functional group to simultaneously interact with the β-PAA substrate and the D225 carboxylate group. Thus, R246 likely fulfills roles in stabilizing catalytically relevant orientations for both the β-PAA substrate and D225. Using this active conformation as a benchmark, we performed an analysis of the average β-PAA scissile bond distance to S157, which is estimated as 5.39 ± 0.03 Å. Figure 2.8B reveals a distorted catalytic triad that results when the α-PAA substrate is docked such that increased R246:α-PAA interactions lead to repositioning of D225/H280 away from S157 and a S157:α-PAA scissile bond distance is equal to 8.1 ± 0.1 Å.

Comparison of wild-type simulations with runs performed with R246K, D225A, and R246A PahZ1$_{\text{KT-1}}$ variants provide atomistic insight into our observed functional defects. Figure 2.8C-E highlight the resulting data in the same format used in Figure 2.8A,B but instead with β-PAA exclusively bound. Examination of the R246K and D225A catalytic triad arrangements suggests that these residue positions are closely linked. Figure 2.8C reveals that mutation of R246 to lysine (R246K) decreases hydrogen bonding potential by conversion of the guanidinium functional group to a primary amine, thereby disallowing simultaneous D225/β-PAA interactions observed in Figure 2.8A. Consequently, D225 interactions with H280 do not occur as efficiently as in the wild-type enzyme. For either R246K (Figure 2.8C) or D225A (Figure 2.8D) mutations, H280 and S157 are not aligned properly, thereby predicting impaired catalytic function. Based

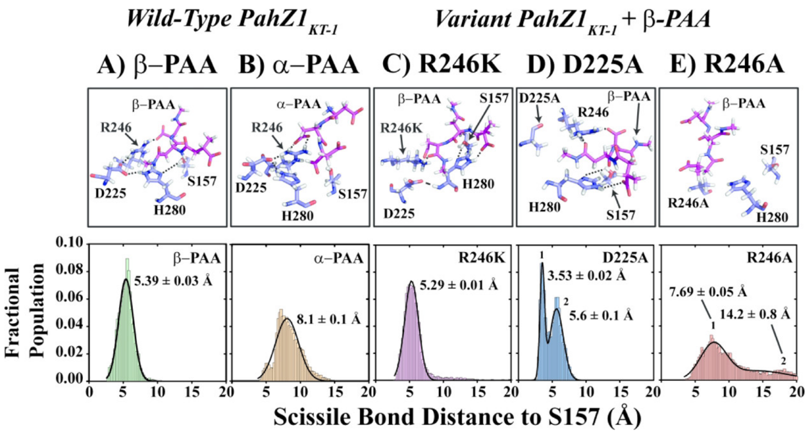


Figure 2.8: PahZ1$_{\text{KT-1}}$ mutations impact catalytic triad arrangements. Atomistic molecular dynamics simulations were performed with PahZ1$_{\text{KT-1}}$ bound with β-PAA (A) or α-PAA substrate (B). Identical 200 nanosecond simulations were performed with variant PahZ1$_{\text{KT-1}}$ constructs, R246K (C), D225A (D), and R246A (E), bound to β-PAA. For each simulation, structural representations of the catalytic triad are provided that highlight the position of bound substrate relative to the residues D225, H280, S157, and R246. All residues are labeled accordingly. Each condition provides an analysis of the fractional population frequency describing the distance of the scissile bond to the catalytic S157 residue. Mean distance values are provided in each case.

on these results, our observation of comparable tPAA hydrolase activities for R246K and D225A (Figure 2.7C) PahZ1$_{\text{KT-1}}$ variants suggests impairments to H280 positioning, and likely also H280 polarization, as the underlying cause for the loss of function. However, it is important to take note that both mutants still retain the ability to position the β-PAA substrate in close proximity to S157. As such, we expect that R246K and D225A hydrolase activities are probabilistic and depend on random sampling of H280 side-chain conformations leading to productive catalysis. This idea is supported by observations for R246A mutants of significantly altered positioning of both the β-PAA substrate and D225 within the PahZ1$_{\text{KT-1}}$ active site (Figure 2.8E). The PAA:S157 scissile bond distances are comparable between the R246A:β-PAA (Figure 2.8E) and wild-type:α-PAA (Figure 2.8B) simulations and are consistent with the decreased hydrolase activity for R246A reported above (Figure 2.7C). Taken together, these data strongly implicate the R246 guanidinium functional group in substrate binding/positioning as well as stabilization of the functional catalytic triad.

2.5 Discussion

In the modern era, polymers have become ubiquitous in our daily lives. With the production of these xenobiotics, the dire effect on the environment needs to be considered. Biodegradation is a realistic approach toward combating polymer accumulation and knowing protein structures allows for subsequent protein engineering to enhance activity or even change substrate selectivity to breakdown these environmental contaminants. Recently, two enzymes were discovered that are capable of degrading highly crystalline poly(ethylene terephthalate)[47,48] which may be of utility to combat issues like the rapidly accumulating Great Pacific Garbage Patch.[49] While water-soluble polymers, like non-biodegradable poly(carboxylates), may not garner the same attention as plastics because consumers are unable to see them, long-term accumulation of these polymers may pose issues for the fauna and flora in the marine environment.[50] PAA is a green alternative that may pose less of an environmental risk. To begin understanding the mechanism of PAA biodegradation, we determined the structure of

$\mathrm{PahZ1_{KT\text{-}1}}$.

Active $\mathrm{PahZ1_{KT\text{-}1}}$ was shown to be composed of a dimeric assembly in the crystal structure (Figure 2.2) which was further supported through solution SAXS/MALS data (Figure 2.3). The dimer interface is lined with hydrophobic residues and is further stabilized through hydrogen bonding at the termini of the helical interface (Figure 2.4). Through MD, this interface was calculated to be relatively weak with a $\Delta G = 53 \pm 3$ kJ $\mathrm{mol}^{-}1$, which may lend some insight into the protein being unstable at temperatures above 40 °C.[7] The structure shows a large, open, positive trough (Figure 2.2D) which likely directs the negatively charged polymer substrate binding to the catalytic sites that reside on opposite faces of the structure. Consistent with this prediction, computational docking experiments failed to observe stable structures formed between $\mathrm{PahZ1_{KT\text{-}1}}$ and polymer substrates with uncharged backbones including polyalanine, poly(3-hydroxybutyrate), and polymalate (Figure A.3). Potentially, this binding trough architecture may allow for a single, longer polymer substrate to bind each monomer to facilitate more efficient degradation.

Collectively, $\mathrm{PahZ1_{KT\text{-}1}}$ and $\mathrm{PahZ2_{KT\text{-}1}}$ are both needed to fully digest tPAA to aspartic acid. While $\mathrm{PahZ1_{KT\text{-}1}}$ only cleaves the β-amide linkages and $\mathrm{PahZ2_{KT\text{-}1}}$ cleaves both the α- and β-amide linkages, $\mathrm{PahZ1_{KT\text{-}1}}$ needs to act upon tPAA first. A rationale for this ordered activity is that it is likely due to irregular end formation during the thermal synthesis of PAA.[13] While $\mathrm{PahZ2_{KT\text{-}1}}$ acts on either PAA amide linkage, it is an exo-type hydrolase. Therefore, $\mathrm{PahZ1_{KT\text{-}1}}$, as an endo-type hydrolase, cleaves the interior of the polymer chain yielding homogenous ends for PahZ2$_{\text{KT-1}}$ activity. This mode of hydrolysis and substrate specificity is supported by our MD studies. When a β-linked tripeptide is docked into the enzyme active site, the penultimate β-linkage is properly positioned for nucleophilic attack by the active-site serine residue (Figures 6B and 8A). Additionally, docking suggests that the substrate specificity is dictated by R246, which preferentially forms 1-2 hydrogen bonds with the α-linked peptide and moves the substrate away from the catalytic serine; however, R246 undergoes significant movement when the β-linked peptide is present. In the latter, only $\sim$1 hydrogen bond exists and the directionality of the R246 indicates that bonding moves away from the substrate toward stabilization of catalytic triad residues D225 and H280.

While all evidence supports PahZ1$_{\text{KT-1}}$ functioning as a serine protease, the potential hydrogen bonding between R246 and D225 is interesting from a mechanistic perspective. The apo crystal structure presented in Figure 2.2shows the ideal positioning of the canonical Ser/His/Asp catalytic triad. If R246 were to hydrogen bond with the D225 side-chain carbonyl oxygen, then a D225-H280 interaction could improve the likelihood of successful proton abstraction from S157 in the first steps of the classical serine protease mechanism. Interestingly, this arginine residue is also conserved in the putative *B. parapertussis* poly(3-hydroxybutyrate) depolymerase (Figure A.1). Of the top 20 hits identified in our DALI search with the PahZ1$_{\text{KT-1}}$ structure, the putative *B. parapertussis* poly(3-hydroxybutyrate) depolymerase is the only other structure that has an arginine at this position. In the other structures, the equivalent position is usually occupied by a hydrophobic residue. Our data suggests that R246 does play an essential role in PahZ1$_{\text{KT-1}}$ substrate binding and/or catalysis, and likely the *B. parapertussis* enzyme functions similarly. Our structural comparison of the *B. parapertussis* enzyme to the well-characterized *P. lemoignei* PHB depolymerase suggests that the *B. parapertussis* enzyme likely recognizes a substrate other than PHB (Figure A.1). Additionally, although both PahZ1$_{\text{KT-1}}$ and the *B. parapertussis* enzyme have very similar folds (Figure 2.2C) and conserve the active-site arginine residue (Figure A.1), our analysis of their active-site architectures (Figure 2.2D,E) suggest that they bind different substrates.

2.6 Conclusions

The dimeric structure of PahZ1$_{\text{KT-1}}$ is stabilized primarily through hydrophobic interactions and each subunit displays a positively lined trough that binds the polyanionic PAA substrate and directs it toward the catalytic active site. MD provides a rationale for the previous reports of PahZ1$_{\text{KT-1}}$ specific cleavage of only the PAA β-linkage through the movement of R246. Furthermore, R246 appears to directly interact with the catalytic D225 and may be responsible for directing binding or stabilizing the functional catalytic triad orientation. Importantly, knowing the structure of PahZ1$_{\text{KT-1}}$ provides the framework to begin protein engineering studies to evaluate this enzyme's ability to serve as a biodegradation agent toward other water-soluble polymers.

References

(1) Thombre, S.; Sarwade, B. Synthesis and Biodegradability of Polyaspartic Acid: A Critical Review. *Journal of Macromolecular Science, Part A* **2005**, *42*, 1299–1315.

(2) Paik, Y.; Simon, E.; Swift, G. In *Hydrophilic Polymers, American Chemical Society*, 1996; Vol. 248, 79–98.

(3) Bai, M.; Wilske, B.; Buegger, F.; Esperschütz, J.; Bach, M.; Frede, H.-G.; Breuer, L. Relevance of nonfunctional linear polyacrylic acid for the biodegradation of superabsorbent polymer in soils. *Environmental Science & Pollution Research* **2015**, *22*, 5444–5452.

(4) Wilske, B.; Bai, M.; Lindenstruth, B.; Bach, M.; Rezaie, Z.; Frede, H.-G.; Breuer, L. Biodegradability of a polyacrylate superabsorbent in agricultural soil. *Environmental Science & Pollution Research* **2014**, *21*, 9453–9460.

(5) Amass, W.; Amass, A.; Tighe, B. A review of biodegradable polymers: uses, current developments in the synthesis and characterization of biodegradable polyesters, blends of biodegradable polymers and recent advances in biodegradation studies. *Polymer International* **1998**, *47*, 89–144.

(6) Freeman, M.; Paik, Y.; Swift, G.; Wilczynski, R.; Wolk, S.; Yocom, K. Biodegradability of Polycarboxylates: Structure—Activity Studies. In Hydrogels and Biodegradable Polymers for Bioapplications. *American Chemical Society* **1996**, *627*, 118–136.

(7) Hiraishi, T. Poly(aspartic acid) (PAA) hydrolases and PAA biodegradation: current knowledge and impact on applications. *Appl Microbiol Biotechnol* **2016**, *100*, 1623–30.

(8) Piatkowski, M.; Radwan-Pragłowska, J.; Raclavsky, K. Application of Poly (aspartic acid) and its Derivatives in Medicine and Pharmacy. *Asian Journal of Applied Sciences* **2015**, *03*, 718.

(9) Hiraishi, T.; Maeda, M. Poly(aspartate) hydrolases: biochemical properties and applications. *Appl Microbiol Biotechnol* **2011**, *91*, 895–903.

(10) Tabata, K.; Abe, H.; Doi, Y. Microbial Degradation of Poly(aspartic acid) by Two Isolated Strains of Pedobacter sp. and Sphingomonas sp. *Biomacromolecules* **2000**, *1*, 157–161.

(11) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Abe, H.; Yamato, I.; Doi, Y. Biochemical and molecular characterization of poly(aspartic acid) hydrolase-2 from sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 1285–92.

(12) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Yamato, I.; Doi, Y. Genetic Analysis and Characterization of Poly(aspartic acid) Hydrolase-1 from Sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 80–86.

(13) Hiraishi, T.; Kajiyama, M.; Yamato, I.; Doi, Y. Enzymatic hydrolysis of alpha- and beta-oligo(L-aspartic acid)s by poly(aspartic acid) hydrolases-1 and 2 from Sphingomonas sp. *KT-1. Macromol Biosci* **2004**, *4*, 330–9.

(14) Hiraishi, T.; Masuda, E.; Kanayama, N.; Nagata, M.; Doi, Y.; Abe, H.; Maeda, M. Cloning of poly(aspartic acid) (PAA) hydrolase-1 gene from Pedobacter

sp. KP-2 and hydrolysis of thermally synthesized PAA by its gene product. *Macromol Biosci* **2009**, *9*, 10–9.

(15) Tabata, K.; Kajiyama, M.; Hiraishi, T.; Abe, H.; Yamato, I.; Doi, Y. Purification and characterization of poly(aspartic acid) hydrolase from Sphingomonas sp. *KT-1. Biomacromolecules* **2001**, *2*, 1155–60.

(16) *European Nucleotide Archive.* https://www.ebi.ac.uk/ena/data/view/BAC78155 (accessed January 18, 2018).

(17) Andrzejak, S.; Moyer, S.; Quillian, B.; Shank, N. Revisiting a Green Polymerization of Aspartic Acid for a Second-Semester Introductory Organic Chemistry Laboratory: Improvements. *Learning Objectives, and Post-Laboratory Assignment Chemical Educator* **2018**, *23*, 221–226.

(18) Luft, J.; Collins, R.; Fehrman, N.; Lauricella, A.; Veatch, C.; DeTitta, G. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J Struct Biol* **2003**, *142*, 170–9.

(19) Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 859–66.

(20) McCoy, A.; Grosse-Kunstleve, R.; Adams, P.; Winn, M.; Storoni, L.; Read, R. Phaser crystallographic software. *J Appl Crystallogr* **2007**, *40*, 658–674.

(21) Adams, P. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–21.

(22) Terwilliger, T.; Grosse-Kunstleve, R.; Afonine, P.; Moriarty, N.; Zwart, P.; Hung, L.; Read, R.; Adams, P. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **2008**, *64*, 61–9.

(23) Classen, S.; Hura, G.; Holton, J.; Rambo, R.; Rodic, I.; McGuire, P.; Dyer, K.; Hammel, M.; Meigs, G.; Frankel, K.; Tainer, J. Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* **2013**, *46*, 1–13.

(24) Foster, B.; Rosenberg, D.; Salvo, H.; Stephens, K.; Bintz, B.; Hammel, M.; Ellenberger, T.; Gainey, M.; Wallen, J. Combined Solution and Crystal Methods Reveal the Electrostatic Tethers That Provide a Flexible Platform for Replication Activities in the Bacteriophage T7 Replisome. *Biochemistry* **2019**, *58*, 4466–4479.

(25) Hanwell, M.; Curtis, D.; Lonie, D.; Vandermeersch, T.; Zurek, E.; Hutchison, G. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **2012**, *4*, 1–17.

(26) Hess, B.; Kutzner, C.; Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, *4*, 435–47.

(27) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M.; Smith, J.; Kasson, P.; Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–54.

(28) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, *26*, 1701–18.

(29) Abraham, M.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(30) Best, R.; Zhu, X.; Shim, J.; Lopes, P.; Mittal, J.; Feig, M.; Mackerell Jr, A. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **2012**, *8*, 3257–3273.

(31) Huang, J.; MacKerell Jr, A. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **2013**, *34*, 2135–45.

(32) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(33) MacKerell, A. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998**, *102*, 3586–616.

(34) Berendsen, H.; Postma, J.; Gunsteren, W.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(35) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(36) Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31*, 1695–1697.

(37) Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(38) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Stud. *Physical Review Letters* **1980**, *45*, 1196–1199.

(39) Grossfield, A. WHAM: the weighted histogram analysis method, en.

(40) Holm, L.; Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **1995**, *20*, 478–80.

(41) Jendrossek, D.; Hermawan, S.; Subedi, B.; Papageorgiou, A. Biochemical analysis and structure determination of Paucimonas lemoignei poly(3-hydroxybutyrate) (PHB) depolymerase PhaZ7 muteins reveal the PHB binding site and details of substrate–enzyme interactions. *Molecular Microbiology* **2013**, *90*, 649–664.

(42) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J.; Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* **2013**, *105*, 962–74.

(43) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* **2010**, *38*, Web Server issue), W540-4.

(44) Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007**, *372*, 774–97.

(45) The PyMOL Molecular Graphics System, Version 1.8, de, Schrödinger, LLC.

(46) Corey, D.; Craik, C. An Investigation into the Minimum Requirements for Peptide Hydrolysis by Mutation of the Catalytic Triad of Trypsin. *J. Am. Chem. Soc* **1992**, *114*, 1784–1790.

(47) Palm, G.; Reisky, L.; Böttcher, D.; Müller, H.; Michels, E.; Walczak, M.; Berndt, L.; Weiss, M.; Bornscheuer, U.; Weber, G. Structure of the plastic-degrading Ideonella sakaiensis MHETase bound to a substrate. *Nature Communications* **2019**, *10*, 1717.

(48) Austin, H. et al. Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proceedings of the National Academy of Sciences* **2018**, *115*, 4350– 4357.

(49) Lebreton, L. et al. Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. *Scientific Reports* **2018**, *8*, 4666.

(50) Gross, R.; Kalra, B. Biodegradable Polymers for the Environment. *Science* **2002**, *297*, 803–807.

Chapter 3

Sphingomonas sp. KT-1 PahZ2 Structure Reveals a Role for Conformational Dynamics in Peptide Bond Hydrolysis

Chad A Brambley, Tarah J Yared, Marriah Gonzalez, Amanda L Jansch, Jamie R Wallen, Mitch H Weiland, and Justin M Miller

Reprinted (adapted) with permission from J Phys Chem B. 2021 Jun 10;125(22):5722-5739.
Copyright 2021 American Chemical Society.

3.1 Abstract

Poly(aspartic acid) (PAA) is a common water-soluble polycarboxylate used in a broad range of applications. PAA biodegradation and environmental assimilation were first identified in river water bacterial strains, Sphingomonas sp. KT-1 and Pedobacter sp. KP-2. Within Sphingomonas sp. KT-1, PahZ1$_{\text{KT-1}}$ cleaves β-amide linkages to oligo(aspartic acid) and then is degraded by PahZ2$_{\text{KT-1}}$. Recently, we reported the first structure for PahZ1$_{\text{KT-1}}$. Here, we report novel structures for PahZ2$_{\text{KT-1}}$ bound to either Gd^{3+}/Sm^{3+} or Zn^{2+} cations in a dimeric state consistent with M28 metallopeptidase family members. PahZ2$_{\text{KT-1}}$ monomers include a dimerization domain and a catalytic domain with dual Zn^{2+} cations.

MD methods predict the putative substrate binding site to span across the dimerization and catalytic domains, where NaCl promotes the transition from an open conformation to a closed conformation that positions the substrate adjacent to catalytic zinc ions. Structural knowledge of PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ will allow for protein engineering endeavors to develop novel biodegradation reagents.

3.2 Introduction

Polymers are ubiquitous in our daily lives and are easily recognized in the form of packaging materials and single-use containers. The general public is acknowledging that these versatile synthetic products do not naturally decompose and are readily accumulating in landfills and the environment, as evidenced by the ever expanding Great Pacific Garbage Patch.[1,2] These tangible items have rightfully garnered significant interest about their environmental impacts,[3–6] shortcomings associated with recycling,[7] and potential benefits through biodegradation of persistent plastics like poly(ethylene terephthalate).[8–12] However, perhaps equally as important is to investigate those polymers which we cannot see and are being directly or indirectly released into the environment: water-soluble polymers (WSPs).[13,14] Just because they are not seen does not warrant labeling all WSPs as being environmentally friendly or biodegradable.[15]

WSPs are found in a myriad of industrial applications ranging from washing agents to wastewater treatment to pharmaceuticals.[13] The broad use of WSPs has increased as market demand has shifted away from water-insoluble polymers that are visibly damaging to ecological systems. This transition is motivated by the assumption that WSPs are environmentally friendly, biodegradable, and generally non-toxic. However, it is well-known that water-soluble polymers such as polyacrylamides and polycarboxylates exhibit slow biodegradation contributing to significant environmental persistence.[16,17] The negative effects of specific biodegradation products are clear in some cases; polyacrylamide breakdown yields the potent neurotoxin acrylamide. Acrylamide monomer concentrations in wastewater derived from oil-and-gas industrial applications have been reported to exceed the threshold for ecotoxic effects at 10-1000 mg/L.[14,18,19] In other cases, increased WSP use may lead to increased environmental emissions leading to unintended and unexpected ecological consequences not related to toxicity effects.[14] Ecological effects may instead be related to WSP behavior as flocculants, detergents, or water/soil conditioners in environments extending beyond those intended.

Polycarboxylates like poly(acrylate) and its derivatives are WSPs that have found utility as superabsorbent materials in diapers and feminine hygiene products in addition to being used as detergents, as antiscaling agents, and as a fertilizer synergist.[20–22] Reports have found these synthetic polymers to be environmentally stable and undergo minimal biodegradation.[23,24] As such, there is concern about their environmental accumulation through leaching into ground or surface water, which may also result in metal contamination due to the ability of polycarboxylates to chelate metal ions.[25,26] While poly(acrylate) is generally considered safe, unless ingested in very high quantities,[27] there is a need to shift toward greener alternatives to mitigate environmental accumulation. Poly(aspartic acid) (PAA) is an eco-friendly alternative to polycarboxylates like poly(acrylate) that has been shown to be biodegradable.

PAA is easily synthesized using naturally occurring aspartic acid. Thermal synthesis of PAA (tPAA) is accomplished by heating aspartic acid to form a poly(succinimide) that is ring-opened by the addition of sodium hydroxide to yield a tPAA with α- and β-amide linkages. The process results in formation of an atactic polymer of racemized aspartates, irregular end groups, and preferential branching ($\sim$70%) through the β-linkage.[28] Biodegradation and assimilation of tPAA were first identified in two river water bacterial strains, *Sphingomonas* sp. KT-1 and *Pedobacter* sp. KP-2.[29,30] Subsequent studies of these bacteria identified, isolated, and initially characterized three different proteins capable of tPAA degradation: PahZ1$_{\text{KT-1}}$, PahZ2$_{\text{KT-1}}$, and PahZ1$_{\text{KP-2}}$. Within *Sphingomonas* sp. KT-1, tPAA degradation is an intracellular process whereby PahZ1 cleaves the β-amide linkages of low-molecular-weight tPAA (<5000 Da) leaving oligo(aspartic acid) to be cleaved at both the α- and β-amide linkages by PahZ2 to yield monomeric aspartic acid.[29,31–34] PahZ1$_{\text{KP-2}}$was shown to extracellularly cleave both low-and high-molecular-weight tPAA; however, a corresponding PahZ2 has not been identified in *Pedobacter* sp. KP-2.[29,35]

Recently, we determined the first structure of a poly(aspartic acid) hydrolase, PahZ1$_{\text{KT-1}}$. This serine protease was shown to have a dimeric assembly and displays a positively lined trough of residues that directs the polyanionic tPAA substrate toward the catalytic active site allowing for endolytic cleavage.[36] Here, we further build upon our mechanistic understanding of tPAA degradation with the crystal structure of PahZ2$_{\text{KT-1}}$ which was solved to 1.85 Å. PahZ2$_{\text{KT-1}}$ is also a dimer but is structurally related to M28 metallopeptidases, a family of dinuclear Zn(II)- dependent exopeptidases.[37] PahZ2$_{\text{KT-1}}$ subunits are similar to two other well-characterized M28 family members, carboxypeptidase G2 and DapE, containing both a dimerization and a catalytic domain that harbors two Zn binding sites.[38–40] On the basis of activity assays and molecular dynamics (MD) simulations, we show the OAA substrate binds at the interface between these two domains, through electrostatic interactions, and the presence of NaCl enhances the activity of PahZ2$_{\text{KT-1}}$. The latter is likely through a conformational change that shifts the protein from an open-complex to a closed-complex allowing for enhanced catalytic activity.

Having the structures of PahZ1$_{\text{KT-1}}$ and now PahZ2$_{\text{KT-1}}$ allows for an atomistic understanding of how PAA can be recycled back to monomeric aspartic acid. While the activity of PahZ2$_{\text{KT-1}}$ does not yield complete degradation of OAA to Asp, knowing its structure provides the framework for future protein engineering endeavors to enhance the catalytic activity of PahZ2$_{\text{KT-1}}$, which is exceptionally important from a green chemistry perspective of using PAA, or potentially other WSPs, in a complete cradle-to-cradle process.

3.3 Experimental Section

3.3.1 Materials

All solutions used to generate data reported here were prepared with double-distilled water produced from a Purelab Ultra Genetic System (Siemens Water Technology). The DNA coding for the mature PahZ2$_{\text{KT-1}}$ sequence[31] was cloned into the pET15b vector using the 5′ NdeI and 3′ XhoI restriction sites. DNA synthesis, cloning, and site-directed mutagenesis were performed by GenScript (Piscataway, NJ). PahZ2$_{\text{KT-1}}$ E155A, E156A, D121A, D122A/D184A, and E156A/H374A mutant constructs were created using the PahZ2$_{\text{KT-1}}$ plasmid described above. Preparation of Digested Poly(aspartic acid) (dPAA). It has been shown that PahZ2$_{\text{KT-1}}$ does not show enzymatic activity against tPAA due to the presence of irregular end groups.[31] To produce the substrate for PahZ2$_{\text{KT-1}}$, thermally synthesized poly(aspartic acid), tPAA,[41] was solubilized to 100 mg/mL in 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.0. Subsequently, the tPAA was digested overnight at 37 °C with 0.5 mg/mL PahZ1$_{\text{KT-1}}$.[36] The sample was then incubated at 70 °C to precipitate the PahZ1$_{\text{KT-1}}$, and the supernatant containing the digested tPAA (dPAA) served as the substrate for PahZ2$_{\text{KT-1}}$ in all gel permeation chromatography (GPC) assays.

3.3.2 Methods

Protein Expression and Purification. Plasmid encoding for PahZ2$_{\text{KT-1}}$ or PahZ2$_{\text{KT-1}}$ mutant constructs was transformed into BL21(DE3) *Escherichia coli* cells (New

England Biolabs, Ipswich, MA), and all followed the same expression and purification protocol as described. A single colony was used to inoculate 50 mL of lysogeny broth (LB) media containing 50 µg/mL final carbenicillin. Cells were incubated overnight at 37 °C, with shaking at 225 rpm. After incubation, 10 mL of this growth was used to inoculate 500 mL of LB-media in a 2 L Erlenmeyer flask where cells were grown to an optical density at 600 nm (OD 600) equal to 1.0 at the same temperature and shaking speed. Expression of PahZ2$_{\text{KT-1}}$ was induced with a final concentration of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, GoldBio, St. Louis, MO), and the flask was incubated at 20 °C for 4 h with shaking at 250 rpm. The cells were harvested by centrifugation at 10,000g for 10 min at room temperature. Cell pellets were stored at -20 °C prior to further use.

Pelleted cells were resuspended in 50 mM tris (2-amino-2-(hydroxymethyl)-1,3-propanediol), 300 mM NaCl, and 20 mM imidazole and lysed by sonication for 45 min on ice. Turbonuclease (50 µL, Accelagen, San Diego, CA) was added and allowed to rock for 45 min at room temperature to increase lysis and remove genomic DNA. The cell lysate was clarified by centrifugation at 10,000g for 20 min at 4 °C and subsequently loaded on a HisPur Ni-nitriloacetic acid (NTA) 5 mL cartridge (Thermo Scientific, Waltham, MA) equilibrated in 50 mM Tris, 300 mM NaCl, and 20 mM imidazole. After loading, PahZ2$_{\text{KT-1}}$ was eluted from the column with 50 mM Tris, 300 mM NaCl, and 500 mM imidazole using a gradient of 0-100% over 40 min. The eluted peak corresponding to PahZ2$_{\text{KT-1}}$ was collected and analyzed via SDS-PAGE which showed that the protein was >98% pure after this single purification. Protein was dialyzed against 50 mM

HEPES, pH 7.40, or 20 mM Tris, pH 8.0, and stored at 4 °C.

X-ray Structure Determination and Refinement. PahZ2$_{\text{KT-1}}$ crystallization screening was initially performed at the Hauptman-Woodward Institute High-Throughput Crystallization Screening Center,[42] with crystals observed in a variety of conditions. Single PahZ2$_{\text{KT-1}}$ crystals were obtained by hanging-drop vapor diffusion at 22 °C by mixing 2 µL of protein (13-15 mg/mL PahZ2$_{\text{KT-1}}$ in 20 mM Tris pH 8.0) with 2 µL of reservoir solution placed over a 500 µL reservoir solution containing 15% PEG 3350, 0.1 M HEPES pH 6.5, and the silver bullets D4 (Hampton) cocktail (0.005 M gadolinium(III) chloride hexahydrate, 0.005 M samarium(III) chloride hexahydrate, 0.05 M benzamidine hydrochloride, 0.25% w/v salicin, and 0.02 M HEPES pH 6.8). Crystals grew to full size in approximately 2-3 weeks. For preparation for data collection, crystals were quickly transferred through a solution containing 15% PEG 3350, 0.1 M HEPES pH 6.5, silver bullets D4, and 30% w/v glycerol, and then flash-cooled by plunging into liquid nitrogen. X-ray data were collected at a wavelength of 1.11608 Å at beamline 8.3.1 at the Advanced Light Source on a Dectris Pilatus3 S 6 M detector. Data were processed using XDS.[43]

A strong anomalous signal was detected during data processing, likely due to the Gd^{3+} and/or Sm^{3+} present in the crystallization condition. The structure of PahZ2$_{\text{KT-1}}$ was determined by single-wavelength anomalous diffraction (SAD) phasing using the program AutoSol[44] in the Phenix software suite.[45] AutoSol successfully solved the structure with a figure of merit of 0.412 and an R_{work} and R_{free} of 23.5% and 24.3%, respectively. The heavy atom sites and density modified maps from AutoSol were then fed into Autobuild,[46] which was able to successfully build residues 7-404 of both protein monomers in the asymmetric unit, with an R_{work} and R_{free} of 21.0% and 21.8%, respectively. Manual structure building was followed by xyz coordinate, real space, TLS, and individual B-factor refinement using the program phenix.refine. Noncrystallographic restraints were included during refinement.

For the Zn^{2+}-bound $PahZ2_{KT\text{-}1}$ structure, crystals grown in 15% PEG 3350, 0.1 M HEPES pH 6.5, and silver bullets D4 were soaked in a solution containing 15% PEG 3350, 0.1 M HEPES pH 6.5, silver bullets D4, and 25 mM $ZnCl_2$ for 4 h at room temperature. No back-soaking was performed in an effort to prevent the active-site Zn^{2+} metals from being replaced by the Gd^{3+}/ Sm^{3+} present in the crystallization solution, which led to 28 total zinc atoms present in the final structure. After soaking, the crystals were quickly transferred through a solution containing 15% PEG 3350, 0.1 M HEPES pH 6.5, silver bullets D4, 25 mM $ZnCl_2$, and 30% w/v glycerol, then flash-cooled by plunging into liquid nitrogen. X-ray data were collected at a wavelength of 1.28273 Å at beamline 8.3.1 at the Advanced Light Source on a Dectris Pilatus3 S 6 M detector. Data were processed using XDS.[43] The structure of $PahZ2_{KT\text{-}1}$ bound to zinc was determined by SAD phasing to 2.50 Å using the program AutoSol[44] in the Phenix software suite.[45] AutoSol successfully solved the structure with a figure of merit of 0.322 and an R_{work} and R_{free} of 27.1% and 29.5%, respectively. The heavy atom sites and density modified maps from AutoSol were then fed into Autobuild,[46] which was able to successfully build residues 7-102 and 122-405 of chain A, and residues 12-404 of chain B, with an R_{work} and R_{free} of 20.6% and 22.9%, respectively. Manual structure building was followed by xyz coordinate, real space, TLS, and individual B-factor refinement using the program phenix.refine. Noncrystallographic restraints were included during refinement. A summary of data collection and refinement statistics for both structures is provided in Table 3.1. X-ray structure figures, as well as electrostatic surface potential maps, were generated using the program PyMol.[47]

PahZ2$_{KT\text{-}1}$ Activity Assays in the Presence of Metals. PahZ2$_{KT\text{-}1}$ was exhaustively dialyzed against 50 mM HEPES, 1 mM EDTA, and pH 7.0 and concentrated to 1.27 mg/mL as determined from absorbance measurements at 280 nm. All metals were the chloride salt except for magnesium which had sulfate as the counterion. Metal solutions were made in 50 mM HEPES, pH 7.0, at a concentration of 50 mM (except Zn^{2+}), added to PahZ2$_{KT\text{-}1}$ to a 2.5 mM final concentration, and incubated on ice overnight. Since zinc is poorly soluble at physiological pH, PahZ2$_{KT\text{-}1}$ was dialyzed against 50 mM HEPES, pH 7.0, 2 mM $ZnCl_2$, at 4 °C. Activity assays were set up as 100 µL reactions containing 80 µg PahZ2$_{KT\text{-}1}$ and 24.5 µL of dPAA spiked with 3 mg/mL thyroglobulin (MilliporeSigma, Burlington, MA). Samples were incubated at 37 °C overnight, and the reactions were stopped by heating at 80 °C for 10 min to precipitate the PahZ2$_{KT\text{-}1}$. Samples were stored at -80 °C until they were loaded on the GPC. PahZ2$_{KT\text{-}1}$ activity was assayed using a Shimadzu GPC with a 3 µm SEC-2000, 300 mm x 7.8 mm (Phenomenex, Torrance, CA) column equilibrated in 50 mM HEPES, pH 7.0. Injections were 10 µL, and the thyroglobulin served as an internal loading control. As PahZ2$_{KT\text{-}1}$

Table 3.1: Data Collection and Refinement Statistics for the PahZ2$_{\text{KT-1}}$ Crystal Structures[a]

	PahZ2$_{\text{KT-1}}$ Zn^{2+}	**PahZ2$_{\text{KT-1}}$ Gd^{3+}/ Sm^{3+}**
Data Collection		
space group	$P2_12_12_1$	$P2_12_12_1$
Cell Dimensions		
a, b, c (Å)	48.40, 146.17, 197.24	48.67, 147.87, 197.14
α,β,γ (deg)	90, 90, 90	90, 90, 90
resolution (Å)	117.4-2.50 (2.59-2.50)	98.57-1.85 (1.92-1.85)
total reflections	98 445 (9,103)	244 755 (24 036)
unique reflections	49 265 (4588)	122 397 (12 022)
CC1/2	0.997 (0.591)	0.999 (0.886)
CC	0.999 (0.862)	1 (0.969)
R_{merge} (%)	4.60 (50.8)	3.16 (27.6)
I/σ	16.43 (1.40)	23.23 (2.53)
completeness (%)	99.34 (94.64)	94.50 (82.82)
redundancy	2.0 (2.0)	2.0 (2.0)
Refinement		
no. reflns	49 257 (4587)	115 801 (9982)
R_{work}/R_{free} (%)	18.21/21.00 (30.68/36.92)	14.51/16.16 (25.33/27.34)
No. Atoms		
protein	5926	5955
water	202	879
metals	28	3
B-Factors		
protein	67.34	36.50
water	63.34	48.23
metals	113.17	34.01
Stereochemical Ideality		
bond lengths (Å)	0.002	0.013
bond angles (deg)	0.54	1.21
ϕ, ψ most favored (%)	97.99	99.37
ϕ, ψ allowed (%)	2.01	0.63
ϕ, ψ outliers (%)	0.00	0.00

[a] Values in parentheses are for the highest-resolution shell. Coordinates have been deposited in the Protein Data Bank as entry 7LJH and 7LJI.

degrades dPAA, it forms aspartate which can be evaluated with changes in intensity in the peak with a retention time of 11.18 min. Percent activity was calculated using a ratio of this peak height to thyroglobulin peak at 4.95 min. Alternatively and equivalently, the decrease in dPAA oligomer peak height at 7.0 min can be used to assess enzyme activity by again using the ratio of the peak height to thyroglobulin. Figure B.1 shows the inverse correlation between product formation at time ∼11.2 min corresponding to polymer degradation at time ∼7.0 min. Since the presence of Zn^{2+} showed the greatest activity, all other metals were compared relative to Zn^{2+} .

PahZ2$_{KT\text{-}1}$ Salt Activity Assay. Digested PAA was dialyzed against 50 mM HEPES pH 7.0 without NaCl, or with 150 mM NaCl or 500 mM NaCl. Dialyzed dPAA samples were lyophilized and resuspended to a final concentration of 100 mg/mL. PahZ2$_{KT\text{-}1}$ was prepared by an initial dialysis against 50 mM HEPES with 2 mM $ZnCl_2$ at pH 7.0 and subsequent dialysis against 50 mM HEPES pH 7.0 with or without NaCl. Assay samples were made with 24.5 mg of dPAA and 0.1 mg/ mL PahZ2$_{KT\text{-}1}$ in a final reaction volume of 100 µL and incubated at 37 °C. Aliquots of 20 µL were taken at the following time points: 5 min, 30 min, 1 h, 2 h, 3 h, 6 h, 12 h, and 24 h. Reactions were stopped by incubating at 80 °C for 10 min and stored at -80 °C. Prior to running on the GPC, samples were spiked with 10 µL of 2.2 mg/mL thyroglobulin. A new GPC 3 µm SEC-2000, 300 mm x 7.8 mm column from the same manufacturer but from different lots was used for the salt assays; however, the same corresponding peak as the metal assays was analyzed for the 0 mM salt assays. The activity of PahZ2$_{KT\text{-}1}$ was determined by analyzing the peak formed from the formation of aspartate with a retention time of 11.02 min, on this new column, relative to the thyroglobin for the 0 mM NaCl assays. However, the presence of NaCl shifted the chromatogram, and the aspartate peaks for 150 and 500 mM NaCl were at 10.75 and 10.80 min, respectively. Figure B.2 shows chromatogram deviations based on the presence of salt and includes monomeric Asp along with Asp + dPAA as controls to show that PahZ2$_{\text{KT-1}}$ cleaves dPAA to Asp.

PahZ2$_{KT\text{-}1}$ Catalytic Mutant Activity Assays. The procedure mimicked the 0 mM NaCl activity assays described above with the exception of the only time point being assayed at 24 h. Given the abolishment of activity with the catalytic mutants, assays were assessed using the degradation of the dPAA peak at a time of 7.0 min because there was more noise in the measurements at time $\sim$11 min where the salt peak and product (monomeric aspartic acid) peak overlap.

Molecular Dynamics Simulations. Docking and Ligand Preparation. Molecular docking of poly(aspartic acid) and other ligands to the PahZ2$_{\text{KT-1}}$ active site was performed in several steps. First, Protein Data Bank (PDB) models of PahZ2$_{\text{KT-1}}$ were generated using the build function of the Pymol molecular visualization program 30 or downloaded from the Protein Data Bank.[48] All subsequent preparation prior to docking was completed using UCSF Chimera.[49] Structure files were converted to the requisite PDBQT file type, and polar hydrogens were added where applicable. Search coordinates were constructed corresponding to the known active site, and docking was carried out without constraints in Autodock Vina.[50] The docked ligand coordinates were then exported, and the

SwissParam small molecule force field generation tool[51] was used to build parameter sets for later use in molecular simulations.

Production Molecular Dynamics. All models prepared for in silico studies were structurally refined with 200 ns of production molecular dynamics (MD) via the Gromacs molecular simulation package.[52,53] Briefly, all models were solvated in triclinic boxes with simple point charge water and neutralizing ions. Where applicable, 500 mM NaCl was also added. Ligand-bound simulations feature parameter sets generated by the SwissParam small molecule force field generation tool (see Docking and Ligand Preparation).[51] Solvated systems were minimized using a steepest descent energy minimization protocol. Except where otherwise noted, all minimized structures were thermally equilibrated for up to 5000 ps of position restrained NVT and NPT ensembles with a Berendsen thermostat.[53,54] Unrestrained production MD was then carried out for 200 ns with Nose-Hoover[55,56] and Parinello-Rahman[57,58] temperature and pressure coupling, respectively, using the Charmm-36 all-atom force field.[59,60]

Free Energy Landscapes. Free energy landscapes (FEL) were generated using the sham analysis package as included in the Gromacs suite of molecular simulation tools. To prepare for sham input, all models were equilibrated by 200 ns of production MD as described above. Simulations performed with high salt conditions were solvated in simple point charge water and supplemented with 0.5 M NaCl. Following equilibration, root-mean-square deviation (RMSD) and radius of gyration (RoG) calculations were performed for each system. Gromacs-generated average structures were extracted and used as reference files for these calculations. A text file containing RMSD and RoG data per frame was constructed and passed to sham. The resulting FEL data was then visualized by the Gnuplot suite of plotting tools.[61]

Principal Component Analysis. Following 200 ns of unrestrained production MD (see Production Molecular Dynamics section above), all protein coordinates were extracted as DCD format trajectories. These files were then read into Bio3d, a powerful R package for molecular simulation and structure analysis.[62,63] C-α atoms were selected and used for subsequent superposition of the trajectory frames in preparation of principal component calculations. Plotting tools provided by R Studio[64] and Bio3d[62,63] were used for all component visualization.

Center-of-Mass (COM) Pulling. All PahZ2$_{\text{KT-1}}$ models with bound Zn^{2+} atoms were prepared and simulated as described above. A 200 ns frame representing conformational con- vergence was taken from the resulting trajectory. The catalytic domain featuring the Zn^{2+} binding site was then extracted and used as input for the subsequent COM pulling simulations using Gromacs.[52,53] The catalytic domain was first solvated in a TIP3P water box[65,66] measuring 13 nm along the reaction coordinate (ξ). Neutralizing ions were added, and the system was brought to an energetic minimum by steepest descent energy minimization. Convergence of thermodynamic parameters was accomplished following 5000 ps of NPT ensemble dynamics in the presence of position restraints. Restraints were removed, and each Zn^{2+} atom was then pulled along ξ in a stepwise procedure. Unrestrained NPT dynamics were used for 1000 ps after pulling the first Zn^{2+} atom to allow for relaxation of active-site residues before extracting the remaining Zn^{2+} atom. All pulls used for umbrella sampling were performed at a rate of 0.01 nm/ps and a force constant of 500 kJ mol^{-1} nm^2. Additional pulls for each Zn^{2+} were also performed across a wide range of rates of constants between 0.001 and 0.01 nm ps^{-1} and between 500 and 1000 kJ mol^{-1} nm^{-2}, totaling five complete COM pulling simulations for each bound Zn^{2+} atom.

Umbrella Sampling and Potential of Mean Force (PMF) Calculations. Configurations with varying COM distances between 0.02 and 0.5 nm along ξ were extracted and prepared for umbrella sampling simulations. Additional configurations were used as necessary to ensure sufficient sampling, resulting in 32 and 34 windows for the first and second Zn^{2+} steered molecular dynamics dissociations, respectively. Each chosen configuration along ξ was then brought to thermodynamic convergence with 100 ps of restrained NPT ensemble equilibration. Restraints were removed, and each configuration was simulated for 10 ns with an applied umbrella potential at a force constant of 500 kJ mol^{-1} nm^2 . The resulting force data was then used as input for subsequent potential of mean force calculations by the WHAM Gromacs package.[67] Adequate sampling was confirmed by histogram analysis, and error estimation was obtained by bootstrap methods (n = 200).

Dynamic Light Scattering (DLS) to Assess PahZ2$_{KT\text{-}1}$ Dimerization in Solution. To determine if differing concentrations of NaCl would disrupt the dimeric structure of PahZ2$_{\text{KT-1}}$, purified PahZ2$_{\text{KT-1}}$ was dialyzed against 50 mM HEPES, pH 7.0, and supplemented with NaCl to final concentrations of 0, 150, and 500 mM. All protein solutions were at 0.5 mg/mL final and centrifuged at 17 000g for 5 min at room temperature to remove any particulate. Samples were analyzed on a Zetasizer Nano ZS instrument (Malvern Panalytical, Westborough, MA) with a size detection range of 0.3 nm to 10 µm where 500 µL of sample was added to a disposable cell.

3.4 Results and Discussion

3.4.1 Overview of the PahZ2$_{\text{KT-1}}$ Crystal Structure.

Screening for crystallization conditions was performed at the Hauptman-Woodward Institute High-Throughput Crystallization Screening Center, which revealed PahZ2$_{\text{KT-1}}$ crystals in a variety of conditions. Single crystals grown in the presence of Gd^{3+} / Sm^{3+} diffracted to 1.85 Å, whereas crystals soaked in 25 mM Zn^{2+} diffracted to 2.50 Å (Table 3.1). Metal-bound structures were determined independently by single-wavelength anomalous diffraction (SAD) phasing methods. For the PahZ2$_{\text{KT-1}}$ structure in the presence of Gd^{3+} / Sm^{3+}, refinement was carried out with Gd^{3+} placed in the structure. However, both metals have similar anomalous scattering coefficients at the wavelength used for data collection, so differentiation between the bound metals is ambiguous. It is likely that a mixture of both Gd^{3+} and Sm^{3+} occupy the metal coordination sites in the structure, and activity assays described below show comparable PahZ2$_{\text{KT-1}}$ activity in the presence of these metals. Therefore, we conclude that the structure presented accurately represents the structure with either Gd^{3+} or Sm^{3+} bound.

The structures each contain two PahZ2$_{\text{KT-1}}$ monomers in the asymmetric unit, and Figure 3.1A shows the believed biological PahZ2$_{\text{KT-1}}$ dimer. Support for this dimer assembly is drawn from structurally related enzymes, size-exclusion chromatography experiments, and dynamic light scattering experiments discussed below. Individual PahZ2$_{\text{KT-1}}$ subunits are related by a 2-fold rotational symmetry at the dimer interface and reveal two distinct domains: (1) dimerization domain and (2) catalytic domain. Figure 3.1B illustrates that the two domains are connected by a hinge region, with the dimerization domain occurring as an insertion spanning V196-A318. The PahZ2$_{\text{KT-1}}$ catalytic domain is represented by residues M1-G195 and G319-Q405. Structures determined with bound Gd^{3+} / Sm^{3+} or Zn^{2+} superimpose well, with root-mean-square deviation equal to 0.652 Å for the dimer (767 residues) and 0.558 Å for the monomer (400 residues) superpositions (Figure 3.1A,B) and do not reveal significant differences in side-chain positioning. Additional discussion of each PahZ2$_{\text{KT-1}}$ structural domain is provided below.

A structural comparison using the DALI server[68,69] reveals multiple closely related structural homologues. Structural homologues that have been experimentally characterized are observed to commonly be dimeric with Zn^{2+} bound to the catalytic domain. Examples include *Pseudomonas* sp. RS-16 carboxypeptidase G2 40 (PDB ID 1CG2, Z-score = 35.6) and *Haemophilus influenza* N-succinyl-l,l-diaminopimelic acid desuccinylase (DapE)[38,39] (PDB ID 5VO3, Z-score = 31.8), which share 21% and 18% sequence identity, respectively, with PahZ2$_{\text{KT-1}}$. Structural superposition of PahZ2$_{\text{KT-1}}$, carboxypeptidase G2, and DapE reveals a

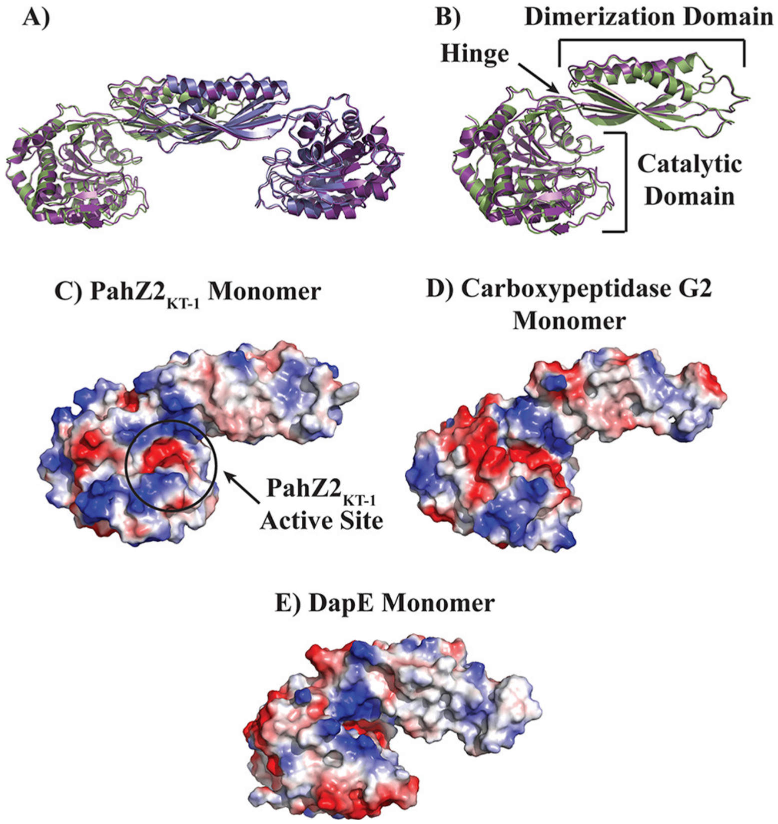


Figure 3.1: Overview of PahZ2$_{KT-1}$ structure. (A) Superposition of the PahZ2$_{KT-1}$ dimers bound with Gd^{3+} / Sm^{3+} or Zn^{2+} in the active sites. Both monomers of the Gd^{3+} / Sm^{3+} dimer are shown in purple, while the two Zn-bound dimers are colored blue and green. The dimers superimpose with an RMSD of 0.625 Å over 767 residues. (B) Superposition of the PahZ2$_{KT-1}$ monomers bound with Gd^{3+} / Sm^{3+} (purple) or zinc (green) in the active sites. The two structures superimpose with an RMSD of 0.558 Å over 400 residues. Labels are included to highlight the location of the dimerization and catalytic domains as well as the connector hinge region. PahZ2$_{KT-1}$ structures are represented in cartoon view. (C-E) Calculated electrostatic surface potentials for PahZ2$_{KT-1}$ (C), carboxypeptidase G2 (D), and DapE (E) monomer structures. All structural representations were generated using the Pymol software package and Protein Data Bank accession files 1CG2 and 5VO3 for carboxypeptidase G2 and DapE, respectively.

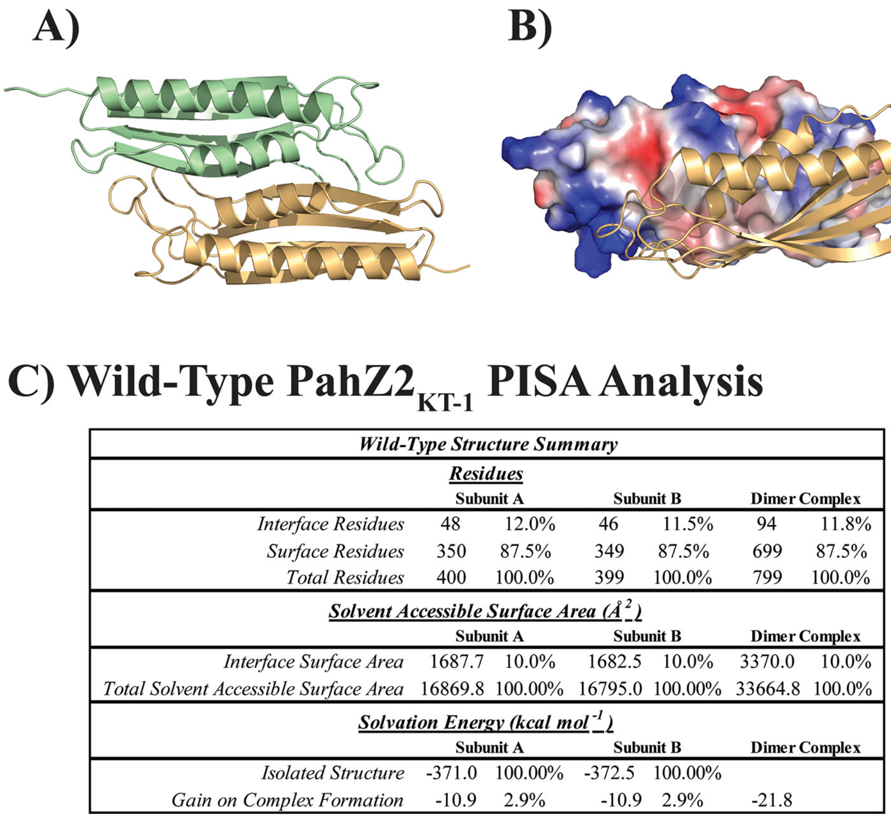
shared fold, but electrostatic surface potential maps demonstrate a marked difference in surface charge. Figure 3.1C-E illustrates these differences by showing calculated electrostatic surface potential maps for the PahZ2$_{\text{KT-1}}$ (Figure 3.1C), carboxypeptidase G2 (Figure 3.1D), and DapE (Figure 3.1E) monomer structures. Each structure harbors nonpolar content at the core of the dimer interface (white, Figure 3.1C-E), consistent with the formation of a protein/protein interface. In contrast, the PahZ2$_{\text{KT-1}}$ catalytic domain surface is enriched in positive charge (blue, Figure 3.1C), which contrasts the solvent accessible surfaces for carboxypeptidase G2 and DapE. However, the active site is conserved in each enzyme, thereby leading to similar charge distribution in this region. Calculations of the sum of formal charges on DapE, carboxypeptidase G2, and PahZ2$_{\text{KT-1}}$ predict net charges equal to -18, -14, and +2, respectively. Differences in electrostatic surface potential for PahZ2$_{\text{KT-1}}$ solvent accessible surfaces relative to other enzymes may reflect an evolved ability to interact with negatively charged oligo(aspartic acid) substrates.

The overall structures of PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ are notably different in domain architecture and classification (Figure B.3A,B). While both enzymes exist in solution as a homodimer, PahZ1$_{\text{KT-1}}$ monomers consist of only a single domain, while PahZ2$_{\text{KT-1}}$ monomers each have distinct dimerization and catalytic domains as discussed here. Visualization of the electrostatic surface potential for each enzyme also highlights how each has evolved the ability to distinguish between poly(aspartic acid) and oligo(aspartic acid). Figure B.3C highlights the electrostatic surface for PahZ1$_{\text{KT-1}}$, where a large trough exists that is lined with basic residues. In our previous report of the PahZ1$_{\text{KT-1}}$ structure,[36] we utilized in silico molecular dynamics methods to examine how the enzyme utilizes this cationic surface to distinguish between α- and β-linkages in its substrates consistent with previous reports of a β-linkage specificity.[34] In contrast, no such cationic trough is observed for PahZ2$_{\text{KT-1}}$ (Figure B.3D), which is consistent with previous kinetic data examining the length-dependent kinetic properties for PahZ2$_{\text{KT-1}}$ catalyzed oligo(aspartic acid) degradation. Hiraishi and co-workers reported that PahZ2$_{\text{KT-1}}$ exhibits a V_{max} and K_{m} that are dependent and independent, respectively, on substrate length such that optimum activity is observed for a pentameric substrate.[33] Thus, PahZ2$_{\text{KT-1}}$ does not possess a binding surface that would accommodate tPAA similar to that found in PahZ1$_{\text{KT-1}}$. Last, PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ each belong to unique protease families. Figure B.3E,F highlights the active-site architectures observed in PahZ1$_{\text{KT-1}}$ (Figure B.3E) and PahZ2$_{\text{KT-1}}$ (Figure B.3F), which are consistent with membership in serine protease and metalloprotease families, respectively.

3.4.2 PahZ2$_{\text{KT-1}}$ Dimerization Domain.

Examination of the PahZ2$_{\text{KT-1}}$ dimer interface reveals symmetric interactions that occur between subunits involving two separate contact points for both Gd^{3+}/ Sm^{3+} - and Zn^{2+}-bound structures. Figure 3.2A highlights the two-tiered arrangement wherein dimerization is mediated through interactions that occur between α-helical and β-sheet layers of associated subunits. Intersubunit interactions involve a mixture of nonpolar and polar residues that occur in both tiers. Figure 3.2B provides insight into how polar and nonpolar residues are distributed about the interface. The symmetric interface is depicted such that one dimerization domain is shown as an electrostatic surface and the other is shown in cartoon representation. From this, it is apparent that nonpolar residues populate the core of the dimer interface with increased positive and negative charge existing at the periphery of the interface. Furthermore, protein-protein contact occurs primarily through a single α-helix and β-strand on each subunit consisting of residues M222-K234 and S248-T254, respectively. The interaction of each secondary structural element is symmetric such that the α-helix or β-strand is positioned antiparallel to the same feature on the adjacent subunit.

To identify interactions that stabilize the PahZ2$_{\text{KT-1}}$ dimer, we subjected the interface to analysis using the Protein Data Bank in Europe Proteins, Interfaces, Structures, and Assemblies (PDBePISA)[70] interactive tool. The summarized PDBePISA output for the higher-resolution interface derived from the Gd^{3+}/ Sm^{3+}-bound structure is presented in Figure 3.2C with data tabulated for each individual subunit as well as the overall dimer complex. Of the 799 residues represented by both PahZ2$_{\text{KT-1}}$ chains, 11.8% are found at the dimer interface. Figure 3.2C further highlights the contribution of subunits A and B to the interface surface as 1687.7 Å^2 (10.0% of subunit A surface area) and 1682.5 Å^2 (10.0% of subunit B surface area), respectively. The overall dimer interface area is equal to the mean of individual subunit surface areas, 1685.1 Å^2. On the basis of the difference in solvation energies for the isolated chains versus the dimer structure, PDBePISA also provides an estimate of solvation free energy gain associated



C) Wild-Type PahZ2$_{KT-1}$ PISA Analysis

Wild-Type Structure Summary						
Residues						
	Subunit A		**Subunit B**		**Dimer Complex**	
Interface Residues	48	12.0%	46	11.5%	94	11.8%
Surface Residues	350	87.5%	349	87.5%	699	87.5%
Total Residues	400	100.0%	399	100.0%	799	100.0%
Solvent Accessible Surface Area (Å^2)						
	Subunit A		**Subunit B**		**Dimer Complex**	
Interface Surface Area	1687.7	10.0%	1682.5	10.0%	3370.0	10.0%
Total Solvent Accessible Surface Area	16869.8	100.00%	16795.0	100.00%	33664.8	100.0%
Solvation Energy (kcal mol^{-1})						
	Subunit A		**Subunit B**		**Dimer Complex**	
Isolated Structure	-371.0	100.00%	-372.5	100.00%		
Gain on Complex Formation	-10.9	2.9%	-10.9	2.9%	-21.8	

Figure 3.2: Dimerization domain is a nonpolar interface. (A) The dimer interface of PahZ2$_{KT-1}$ is shown in cartoon view to highlight interactions occurring between secondary structures positioned antiparallel to one another between a single α-helix and β-strand on each subunit, represented by residues M222-K234 and S248-T254, respectively. (B) An electrostatic surface reveals the core of the dimer interface to be largely nonpolar. (C) Summary of proteins, interfaces, structures, and assemblies (PISA) analyses for the wild-type PahZ2$_{KT-1}$ X-ray structure.

with complex formation, ΔG^i . The solvation free energy gain is estimated as ΔG^i = -21.8 kcal mol^{-1} (-91.2 kJ mol^{-1}), where the strongly negative value here indicates the presence of a hydrophobic interface. By comparison, estimates of overall dimer interface area and solvation free energy gain have been reported for the PahZ1$_{KT-1}$ dimer as 626.2 Å^2 and -7.3 kcal mol^{-1}, respectively, consistent with reports of increased thermal stability for PahZ2$_{\text{KT-1}}$ versus PahZ1$_{\text{KT-1}}$.[36,71]

Examination of the interface amino acid composition suggests a mixture of nonpolar and polar residues that collectively promote dimerization. All interactions are observed to be symmetric due to the antiparallel orientation of dimerization domains relative to one another. Multiple salt bridges are present at the dimer interface that include 19 total interactions occurring as side-chain/side-chain, side-chain/main-chain, and main-chain/main-chain arrangements. Of these, 4 hydrogen bonding interactions occur between side-chain functional groups R200-N257, N257-R270, N220-Y233, and N268-S212 with distances equal to 2.68, 3.88, 2.7, and 3.20 Å, respectively. Additional hydrogen bonding interactions involving main-chain atoms include K242-G214, R299-Y233, N220-A247, G252-T250, Y213-T244, Y233-R299, S248-T254, and V256-R270. Of these interactions, two hydrogen bonding interactions, N220-Y233 and N257-R200, are notable for short distances equal to 2.6 and 2.4 Å, respectively, that may suggest an increased bond strength. Given the symmetric dimer interface, two N220-Y233 interactions are observed to stabilize the terminal ends of the helix-helix interaction spanning residues M222-K234 for each helix and two N257-R200 interactions that stabilize the β-tier of the dimerization domain. A detailed MD inspection of in silico-generated interface residue mutations that either stabilize or destabilize PahZ2$_{\text{KT-1}}$ dimer formation is presented in the Supporting Information, which generally confirm the role of nonpolar residues in stabilizing the dimer (Figures S4 and S5).

3.4.3 PahZ2$_{\text{KT-1}}$ Catalytic Site Analysis.

Overall Catalytic Site Description. The PahZ2$_{\text{KT-1}}$ structure reveals an active site consistent with classification as an M28 family member.[37] The M28 family of metallopeptidases contains both aminopeptidases and carboxypeptidases that each harbor two zinc ions at the active site. From available crystal structures for M28 family members, each zinc ion is expected to coordinate in a tetrahedral molecular geometry via 3-4 amino acids and additional water molecules. Figure 3.3A,B highlights the metal- coordinating residues in the PahZ2$_{\text{KT-1}}$ active site when Gd^{3+} / Sm^{3+} (Figure 3.3A) versus Zn^{2+} (Figure 3.3B) is bound, where anomalous difference electron density allows for unambiguous metal placement in each case. The overlay of active-site residues with bound Gd^{3+} / Sm^{3+} and Zn^{2+} shown in Figure 3.3C illustrates no significant difference in positioning of nearby side-chain functional groups but instead shows a difference in metal placement. The PahZ2$_{\text{KT-1}}$:Gd^{3+} / Sm^{3+} complex reveals the metal to adopt a coordination number of seven involving interactions with side-chain functional groups from E155, D184, H94, D121, E156, and two water molecules (Figure 3.3A). Inspection of the zinc-bound PahZ2$_{\text{KT-1}}$ active site reveals two zinc ions that occupy unique binding sites, which is consistent with other M28 family members (Figure 3.3B).[38–40] Within the M28 family,[37] the two Zn^{2+} binding sites are generally annotated as Zn_{I} and Zn_{II} as shown in Figure 3.3B. At the Zn_{I} coordination site, H94, D184, and D121 function as zinc ligands, whereas ligands found at the Zn_{II} site include H374, E156, and D121.

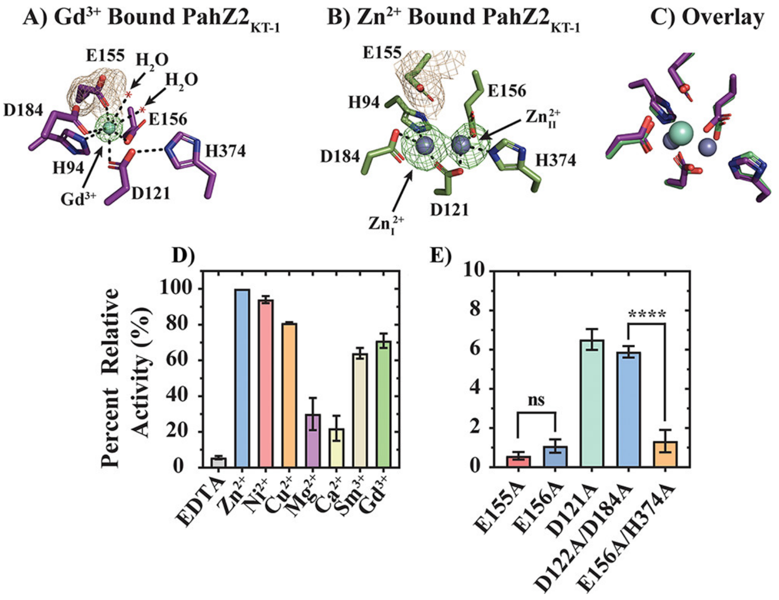


Figure 3.3: PahZ2$_{\text{KT-1}}$ catalytic site description. (A) Active site with Gd^{3+} / Sm^{3+} bound. Anomalous difference density, contoured at 3σ, is shown in green surrounding the Gd^{3+}. Additionally, $2F_o$ - F_c electron density, contoured at 1σ, in shown in brown for residue E155 to show the overall quality of the data. (B) Active site with Zn^{2+} bound. Anomalous difference density, contoured at 5σ, is shown in green surrounding the Zn^{2+} . $2F_o$ - F_c electron density, contoured at 1σ, is shown in brown for residue E155 to show the overall quality of the data. (C) Overlay of Gd^{3+} / Sm^{3+} - and Zn^{2+} -bound active-site features. (D) Relative activity comparing maximal PahZ2$_{\text{KT-1}}$ catalyzed dPAA degradation activity observed in the presence of ethylenediamine tetraacetic acid (EDTA) and Zn^{2+}, Ni^{2+}, Cu^{2+}, Mg^{2+}, Ca^{2+}, Sm^{3+}, and Gd^{3+} metal ions. (E) Relative activity comparing maximal PahZ2$_{\text{KT-1}}$ catalyzed dPAA degradation activity observed for reactions performed with wild-type PahZ2$_{\text{KT-1}}$ alongside Zn_{I} and Zn_{II} site mutant constructs. All bar graph plots are represented with mean and standard deviation derived from triplicate measurements. All structural representations were generated using the Pymol software package with active-site residues presented in stick view. The Gd^{3+} / Sm^{3+} and Zn^{2+} metal ions are represented as teal or gray spheres, respectively. Individual Zn_{I} and Zn_{II} binding sites are indicated by arrows.

PahZ2$_{\text{KT-1}}$ Catalysis is Supported by Multiple Metals. The tetrahedral arrangement of zinc atoms associated with M28 family active sites suggests that Zn^{2+}

may not be the only metal capable of supporting PahZ2$_{\text{KT-1}}$ catalytic function. In fact, Co^{2+}, Zn^{2+}, Cd^{2+}, Mn^{2+}, Ni^{2+}, Cu^{2+}, and Mg^{2+} metal ions are known to support DapE catalytic function.[72] To assess whether PahZ2$_{\text{KT-1}}$ activity can be supported by alternative metals, we performed activity assays reporting on PahZ2$_{\text{KT-1}}$ catalyzed poly(aspartic acid) degradation as described in the Materials and Methods sections. Activity assays were performed in the presence of ethylenediamine tetraacetic acid (EDTA), Zn^{2+}, Ni^{2+}, Cu^{2+}, Mg^{2+}, Ca^{2+}, Sm^{3+}, and Gd^{3+}. As shown in Figure 3.3D, incubation of PahZ2$_{\text{KT-1}}$ in the presence of EDTA significantly inhibits catalytic function such that triplicate measurements yield a mean relative activity equal to 5.61 ± 0.9%. By comparison, Hiraishi and co-workers reported a relative activity equal to 16% in the presence of 1 mM EDTA, which suggested metal-dependent PahZ2$_{\text{KT-1}}$ activity.[31,71] We report here maximum PahZ2$_{\text{KT-1}}$ catalytic activity when reactions are performed in the presence of 2 mM $ZnCl_2$. However, we observe relative activities equal to 93.52 ± 2.18% and 80.99 ± 0.52% when reactions are performed in the presence of Ni^{2+} or Cu^{2+}, respectively. Interestingly, Ni^{2+} commonly adopts a tetrahedral coordination geometry, whereas Cu^{2+} is capable of adopting either square planar or octahedral arrangements. In contrast, divalent cations Mg^{2+} and Ca^{2+} adopt octahedral coordination geometries, which drives an apparent decrease in relative activity to 29.38 ± 9.02% and 21.94 ± 7.20%, respectively. Taken together, these data suggest that the PahZ2$_{\text{KT-1}}$ catalytic output is strongly dependent upon the preferred coordination geometry of a bound metal ion.

Our initial PahZ2$_{\text{KT-1}}$ structures obtained in the presence of samarium or gadolinium salts revealed electron density consistent with a single bound metal ion. To determine whether single metal binding may support PahZ2$_{\text{KT-1}}$ function, additional activity assays were performed in the presence of Sm^{3+} or Gd^{3+}, which each yield relative activities equal to 64.38 ± 2.96% and 70.82 ± 4.37%, respectively. However, Sm^{3+} and Gd^{3+} do not form tetrahedral complexes; instead, they adopt complexes with variable coordination numbers ranging from 6 to 8. Reference to the Gd^{3+} / Sm^{3+}-bound PahZ2$_{\text{KT-1}}$ structure (Figure 3.3A,C) reveals the metal binds directly to the Zn$_{\text{I}}$ site, but such that the Zn$_{\text{II}}$ Site E156 residue also coordinates. As such, we expect that metal binding to the Zn$_{\text{II}}$ site of Gd^{3+} / Sm^{3+}-bound PahZ2$_{\text{KT-1}}$ would be blocked, thereby retaining only activity associated with the Zn$_{\text{I}}$ site.

Bioinformatic Analysis. Figure 3.3B clearly shows that D121 forms interactions at both the Zn$_{\text{I}}$ and Zn$_{\text{II}}$ sites, which is further consistent with classification as an M28 metallopeptidase family member. M28 metallopeptidase family members are classified on the basis of the stated zinc coordination arrangement discussed above, which positions zinc ligands based on the His-X$_{\text{aa}}$-Asp and Glu-Glu motifs (X$_{\text{aa}}$ represents any amino acid). A multiple sequence alignment was performed to compare PahZ2$_{\text{KT-1}}$ primary sequence features to structurally similar proteins identified from the DALI Protein Structure Comparison Server.[68,69] Inspection of the PahZ2$_{\text{KT-1}}$ primary sequence reveals the sequences HLD (H94-D96) and HSLD (H374-D377) at the Zn$_{\text{I}}$ and Zn$_{\text{II}}$ sites, respectively. Figure B.6A illustrates a high degree of sequence conservation at the Zn$_{\text{I}}$ site such that a conserved His-X$_{\text{aa}}$-Asp (HXD, where X represents a variable amino acid position) motif is readily identified. A similar degree of conservation is observed at the $\mathrm{Zn_{II}}$ site that yields a His-$\mathrm{X_{aa}}$-$\mathrm{X_{aa}}$-Asp (HXXD) motif (Figure B.6C).

Figure B.6B demonstrates invariable conservation of the Glu-Glu motif across all structural homologues examined here. On the basis of structural alignments with related proteins such as carboxypeptidase G2 and DapE,[38–40] the conserved Glu-Glu motif positions an acidic functional group to likely participate in water activation essential to formation of the catalytic nucleophile. Once substrate is bound, DapE has been suggested to utilize E134 (Figure B.7) in the deprotonation of a metal-bound water molecule that will undergo nucleophilic attack on the bound peptide substrate.[39,72] Steady-state kinetic studies with DapE performed as a function of pH reveal the maximum reaction velocity to be independent of pH when pH ≥ 7, but to decrease with decreasing pH when pH < 7 with an apparent pKa equal to 6.6 ± 0.2.[73] Given similar observations for DapE,[73] carbonic anhydrase,[74,75] and prolidase,[76] this pKa suggests the pH-sensitive functional group to be a bound water molecule, where metal binding is known to depress the acid dissociation constant for water. Moreover, the positioning of E134 adjacent to the metal-bound water as well as kinetic studies with E134 mutants strongly implicate this glutamate as a general acid/base residue.[38,39,77] The observed structural/sequence conservation observed between $\mathrm{PahZ2_{KT\text{-}1}}$ and DapE strongly implicates $\mathrm{PahZ2_{KT\text{-}1}}$ E155 as a general acid/base in the water activation mechanism presented in Figure B.6.

Catalytic Site Mutagenesis Reveals Zn^{2+} *Binding Site Asymmetry.* To further examine the catalytic importance of the acidic resides discussed above, recombinant

PahZ2$_{\text{KT-1}}$ mutant constructs were prepared. Mutations were made as follows: E155A, E156A, D121A, D122A/D184A, and E156A/H374A. Double-mutations were designed to target acid/metal chelating residues at either the Zn$_{\text{I}}$ or Zn$_{\text{II}}$ sites to disrupt metal binding at either site. Activity assays were performed as described in the Materials and Methods sections in triplicate by incubating PahZ2$_{\text{KT-1}}$ with dPAA for 24 h followed by GPC-based polymer degradation analysis. The resulting data were reported as percent activities relative to wild-type activity. Figure 3.3E presents relative activities for each mutant construct relative to wild-type PahZ2$_{\text{KT-1}}$. Percent relative activities for E155A, E156A, D121A, D122A/D184A, and E156A/H374A were estimated as 0.6 ± 0.2%, 1.1 ± 0.3%, 6.5 ± 0.5%, 5.9 ± 0.3%, and 1.3 ± 0.6%, respectively. The application of ANOVA testing provided support for the conclusion of statistically unique means ($P < 0.0001$). Further application of pairwise Student's t tests were performed to determine whether mean activities are statistically unique for each comparison: E155A and E156A (not significant, $P = 0.0929$), E156A and E156A/ H374A (not significant, $P = 0.5483$), E156A/H374A (Zn$_{\text{II}}$ site) and D122A/D184A (Zn$_{\text{I}}$ site) (significant, $P = 0.0003$), D121A and E155A (significant, $P < 0.0001$), and D121A and E156A (significant, $P = 0.0001$). While all active-site mutations nearly abolish catalytic activity, we highlight that mutations introduced to the Zn$_{\text{II}}$ site have a more negative, statistically significant, effect on relative activity compared to similar mutations introduced at the Zn$_{\text{I}}$ site.

Differences in mean relative activities for each PahZ2$_{\text{KT-1}}$ active-site mutant present multiple mechanistic possibilities that center on nucleophilic water activation. First, we note that all mutations examined here yield a significant loss of relative activity, which strongly supports labeling these residues as catalytic. The noted differences in relative activities for $\mathrm{Zn_{I}}$ and $\mathrm{Zn_{II}}$ site mutant constructs further suggests that each site is nonidentical in catalytic contribution. E156A and E156A/H374A mutants render more significant catalytic impairment relative to the D122A/D184A construct. On the basis of the $\mathrm{PahZ2_{KT\text{-}1}}$ structure, we expect that each double-mutation would decrease the metal binding affinity at the $\mathrm{Zn_{I}}$ and $\mathrm{Zn_{II}}$ sites. Thus, the relative activities presented in Figure 3.3E provide the first clue as to whether one site is more critical to water activation or peptide substrate binding/orientation. It is not possible from these data alone to associate water activation versus peptide substrate binding with a specific site.

Identification of $\mathrm{PahZ2_{KT\text{-}1}}$ E155 as a General Acid/Base Catalytic Residue. Comparison with literature reports for DapE mutants implicate E155 of $\mathrm{PahZ2_{KT\text{-}1}}$ as the only active-site residue functioning as a general acid/base in the water activation mechanism. Consistent with this, an E155A mutation yields 0.6 ± 0.2% relative activity for data points collected after a 24 h incubation. Given that water is not resolved at the active site in the zinc-bound $\mathrm{PahZ2_{KT\text{-}1}}$ active-site structure reported here, we next sought to evaluate the possibility of E155 functioning as a catalytic acid/base by computational methods where water molecules could be readily visualized. Molecular dynamics (MD) simulations were performed as described in the Materials and Methods sections at 4 and 25 °C with the goal of understanding how individual water molecules interact with bound zinc ions and E155. Multiple temperatures have been examined here to account for highly labile interactions involving zinc. Figure 3.4A provides a snapshot of the PahZ2$_{\text{KT-1}}$ metal-binding site for the system equilibrated at 4 °C. MD trajectories reveal two distinct PahZ2$_{\text{KT-1}}$ active-site metal coordination geometries. As shown in Figure 3.4A, the Zn$_{\text{I}}$ site harbors a tetrahedral zinc atom, Zn$_{\text{I}}$, that engages in interactions with E155, D184, and D121 as observed in the X-ray structure. However, these data reveal a rotamer shift for H94 relative to the crystal structure that allows for D122 to complete the zinc coordination sphere. This is consistent with isomerization free energies reported by Dudev and Lim, where the transition from a Zn^{2+} octahedral to tetrahedral geometry occurred with ΔG = -8.4 or -9.6 kcal mol^{-1} when one imidazole or carboxylate ligand was coordinated, respectively.[78] Thus, the replacement of H94 with D122 at the Zn$_{\text{I}}$ site likely represents an arrangement with a marginal stability increase.

The Zn$_{\text{II}}$ site differs from the Zn$_{\text{I}}$ site through the coordination of three ligands from the protein and additional interactions with water. Figure 3.4A highlights the Zn$_{\text{II}}$ site coordination geometry wherein Zn^{2+} interacts with D121, E156, and H374 alongside a bridging water molecule. We note that this water molecule bridges E155 and the Zn$_{\text{II}}$ ion, which is consistent with the proposed DapE mechanism.[39] These data suggest that the Zn$_{\text{II}}$ ion functions to polarize the water molecule, while E155 is positioned to abstract a proton leading to hydroxide formation. Thus, the residual activity reported for E155A in Figure 3.3E may be the consequence of another residue such as D96 functioning as a general acid/base in the absence of E155. We reason then that the role of the Zn$_{\text{I}}$ metal

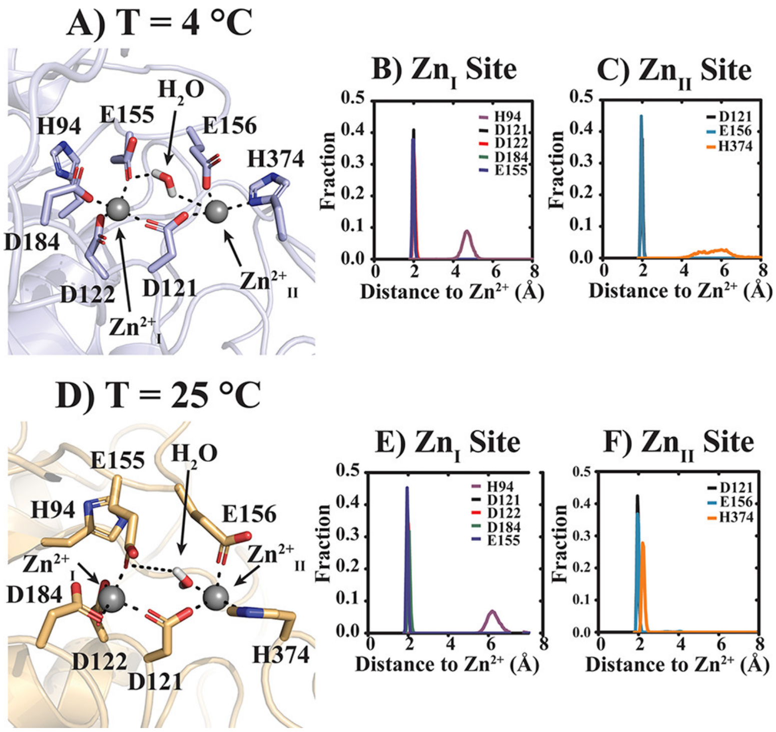


Figure 3.4: Molecular dynamics simulations performed for 200 ns at different temperatures capture a bridging water molecule. (A) MD simulations were performed at 4 °C that reveal a stably coordinated water molecule that interacts with the $\mathrm{Zn}_I^{2+} I$ metal ion and E155. (B, C) Distribution analyses indicate that, with the exception of H94, all active-site residues form stable interactions with nearby Zn_I^{2+} (B) or $\mathrm{Zn}_I^{2+} I$ (C) metal ions based on the distance between side-chain and metal. (D) MD simulations performed at 25 °C reveal water coordination as described in panel A. (E, F) Distribution analyses for active-site side-chain/metal ion distances observed at 25 °C for Zn_I^{2+} (E) or $\mathrm{Zn}_I^{2+} I$ (F) sites. All structural representations were generated using the Pymol software package with active-site residues and water molecules presented in stick view. The Zn^{2+} metal ions are represented as gray spheres.

ion is likely to facilitate both E155 and oligo(aspartic acid) substrate positioning. Support for this model can be found in the observation that all Zn_{I} and $\mathrm{Zn}_{\mathrm{II}}$ site residues are nearly static in position over the course of the low-temperature simulation (Figure 3.4B,C).

Additional simulations performed at 25 °C demonstrate the presence of a bridging water molecule between the $\mathrm{Zn}_{\mathrm{II}}$ metal ion and E155 at a temperature expected to increase water exchange rates. As shown in Figure 3.4D, a bridging water molecule is observed as described, though it is not positioned correctly for nucleophile generation. Improper positioning of this water molecule is likely the consequence of frame sampling. We expect that finer sampling would reveal the water molecule to adopt a catalytic orientation at some point during the simulation. Analysis of the distance between Zn_{I} and $\mathrm{Zn}_{\mathrm{II}}$ site residues and the bound metal ion at 25 °C reveals stable interactions that do not vary significantly over time (Figure B.8). Figure 3.4E demonstrates mean distances between the side-chain and Zn^{2+} for D121, D122, E155, D184, and H94 equal to 1.97 ± 0.04, 2.03 ± 0.06, 1.96 ± 0.04, 2.04 ± 0.06, and 6.3 ± 0.3 Å, respectively. Mean distances between $\mathrm{Zn}_{\mathrm{II}}$ and site 2 residues for D121, H374, and E156 are equal to 1.97 ± 0.04, 2.23 ± 0.07, and 1.98 ± 0.05 Å, respectively (Figure 3.4F). Inspection of a plot of side-chain-Zn^{2+} distance versus time for E156 reveals side-chain oscillations within the first $\sim$10 ns that may reflect artifactual behavior as the system establishes equilibrium. The MD trajectory shown in Figure B.8 is observed to arrange the D156 side-chain at a mean distance equal to 2.06 ± 0.07 Å for the $\mathrm{Zn}_{\mathrm{II}}$ metal ion for the first 3.87 ns followed by an increase to 3.7 ± 0.3 for the next 5.64 ns, decrease to 2.1 ± 0.1 Å for 0.49 ns, increase to 3.6 ± 0.3 Å for 0.39 ns, and, finally, decrease to 1.98 ± 0.05 Å for the remaining 190 ns. This increased side-chain flexibility for E156 is reflected in a statistically significant increase in root-mean-square fluctuation (RMSF) for E156 relative to E155 (Figure B.9, P = 0.0316). Taken together, these data suggest that all Zn_I and Zn_{II} site residues appear to be nearly static in position alongside the bound zinc ions. On this basis, the rate of water activation is likely limited by water exchange.

3.4.4 Asymmetric Zn^{2+} Binding at Sites I and II.

The Zn_I and Zn_{II} sites each contribute unique ligands that function separately and uniquely to coordinate Zn^{2+} ions. From Figure 3.4A, site I utilizes a tetrahedral coordination geometry exclusively involving carboxylate functional groups, whereas site II contributes the side-chains of D121, E156, and H374 alongside water to function as Zn ligands. On this basis, we next asked if $PahZ2_{KT\text{-}1}$ binds each Zn^{2+} with unique affinity. To address this question, we utilized a steered molecular dynamics (SMD) technique known as center-of-mass (COM) pulling to stepwise dissociate each Zn^{2+} ion toward the estimation of dissociation free energy values. Figure 3.5A provides an illustrative look at the method itself. A Zn^{2+} - bound $PahZ2_{KT\text{-}1}$ model for the isolated catalytic domain that has previously been subjected to 200 ns of production molecular dynamics simulation is subjected to an applied force centered on the Zn^{2+} bound at the Zn_{II} site. Through a continued application of force, the metal ion dissociates and moves further away from the protein with increasing time. The same process can then be repeated for the metal bound at the Zn_I site (Figure 3.5B). For experiments removing Zn^{2+} from sites I and II, the centers-of-mass between Zn^{2+} and $PahZ2_{KT\text{-}1}$ are $\sim$1 Å prior to pull force application but reach 6 Å at the end of the 500 ps simulation.

In order to use SMD trajectories for thermodynamic calculations, the bias introduced through the application of pull force must be removed. To this end, umbrella sampling techniques were applied to the extraction of frames along the SMD trajectory for re-equilibration. Configuration of umbrella sampling simulations involved sampling 27 and 34 intersubunit COM distances for the Zn_{II} and Zn_{I} sites, respectively. Each frame was subjected to a 10 ns simulation to allow for equilibrium establishment and then used for PMF calculations summarized in Figure 3.5C,D. Each SMD trajectory was judged by weighted histogram analysis (Figure 3.5E,F) as sufficiently sampled for use in PMF calculations. Each plot of PMF versus distance shown in Figure 3.5C,D is observed to exhibit a sharp increase in energy with increasing COM distance until a distance of $\sim$3 nm is reached and forms an apparent plateau. Calculation of the difference between the PMF maximum and minimum provides an estimate of the apparent free energy change, ΔG, describing the dissociation of Zn^{2+} from each binding site. Figure 3.5C,D reveals estimates for the dissociation free energies at Zn_{II} and Zn_{I} as ΔG ZnII = 42 $\pm$ 4 kJ mol^{-1} and ΔG ZnI = 62 $\pm$ 2 kJ mol^{-1} , which can be converted to dissociation equilibrium constants equal to 49 nM and 16 pM, respectively. These data indicate that binding of Zn^{2+} to each $PahZ2_{KT\text{-}1}$ metal binding site occurs with unique affinity.

We can categorize Zn_{I} and Zn_{II} sites as high- and intermediate-affinity binding sites. By comparison, the suppressor of fused (SUFU), metal responsive element-binding transcription factor (MTF1), carbonic anhydrase II, and rhodopsin each bind zinc with an affinity equal to 0.5 nM,[79] 30 pM,[80] 0.8 pM,[81] and 0.1 µM,[82]

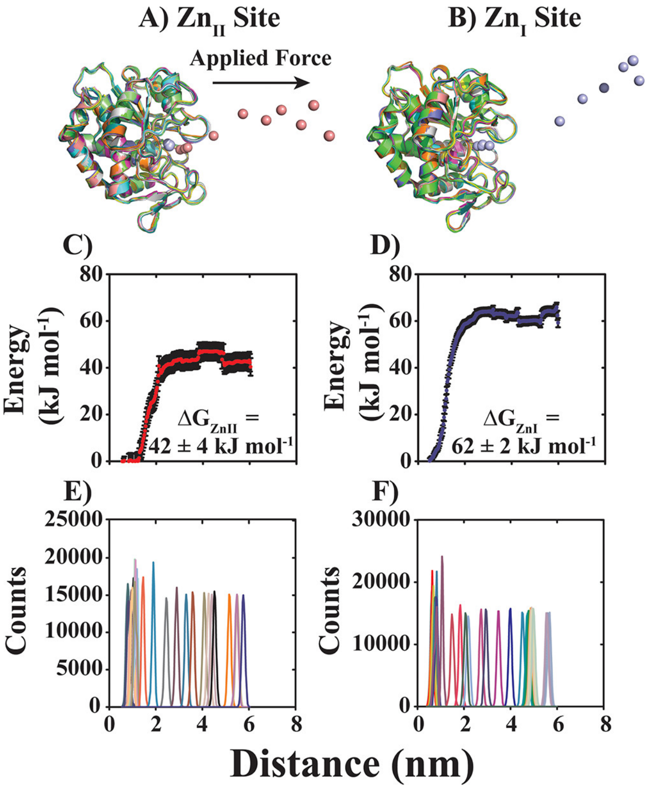


Figure 3.5: In silico calculation of PahZ2$_{KT\text{-}1}$ affinity for zinc metal binding. Steered molecular dynamics techniques were employed alongside umbrella sampling methods to simulate stepwise metal dissociation events at the Zn$_{II}$ and Zn$_{I}$ sites. For each dissociation event, frames observed along the SMD pulling trajectory are shown for the (A) Zn$_{II}$ and (B) Zn$_{I}$ sites. The PahZ2$_{KT\text{-}1}$ catalytic domain is position restrained, while each zinc metal ion is sequentially pulled along a defined path by application of a static force vector. The resulting SMD trajectories were then subjected to umbrella sampling techniques in order to calculate the potential of mean force (PMF). The amplitude of a resulting plot of PMF versus distance between metal and PahZ2$_{KT\text{-}1}$ centers-of-mass yields an estimate of the dissociation free energy. The zinc dissociation free energies from the Zn$_{II}$ (C) and Zn$_{I}$ (D) sites is estimated as 42 ± 4 and 62 ± 2 kJ mol^{-1}, respectively. Adequate sampling was confirmed by weighted histogram analysis (E, F), and error estimation was obtained by bootstrap methods (n = 200).

respectively. High-affinity zinc binding allows for the cell to keep the concentration of free zinc low, which is necessary to avoid accidental binding of other metals such as iron.[83] In fact, the concentration of free zinc in Escherichia coli has been estimated to be on the order of femtomolar,[84] which allows for activity modulation based on the affinities discussed here. From this, we can predict the dominant PahZ2$_{\text{KT-1}}$ ligation state under physiologically relevant conditions. If we assume a free [Zn^{2+}] = 500 fM, the distribution of free PahZ2$_{\text{KT-1}}$ (E), singly bound (EZn) species, and doubly bound (EZn$_2$) can be calculated as 96.9%, 3.0%, and $\sim$0%, respectively, based on a two-site, independent binding model. Figure B.10 demonstrates that a further increase to 16 pM promotes a distribution of 49.9% E, 49.9% EZn, and 0.02% EZn$_2$. On the basis, the occurrence of asymmetric PahZ2$_{\text{KT-1}}$ metal binding sites leading to unique Zn^{2+} binding affinities provides a path to the regulation of enzyme activity. The DapE structural homologue displays 60% relative activity when only a single Zn^{2+} ion is bound, but binding of two zinc ions is required for full enzymatic activity.[39,73] This observation for DapE provides the basis for a model wherein PahZ2$_{\text{KT-1}}$ activity is directly dependent upon free zinc available from the local environment. That is to say, PahZ2$_{\text{KT-1}}$ likely has a single high-affinity zinc atom always bound at the Zn$_{\text{I}}$ site that would support activity, though full activity would only result when both Zn$_{\text{I}}$ and Zn$_{\text{II}}$ sites are metal-bound. Thus, the metal binding at the Zn$_{\text{II}}$ site may serve to regulate PahZ2$_{\text{KT-1}}$ activity with varied [Zn^{2+}]. PahZ2$_{\text{KT-1}}$ may have evolved to have a tunable enzymatic activity that allows for increased function when Sphingomonas sp. KT-1 encounters locally high [Zn^{2+}] associated with metal-chelated poly(aspartic acid) in river water.

3.4.5 NaCl Promotes PahZ2$_{\text{KT-1}}$ Activity.

To date, all studies reporting on the activity of PahZ2$_{\text{KT-1}}$ catalyzed hydrolysis of oligo(aspartic acid) have utilized buffer conditions lacking supplemented salt.[31,32] This strategy ignores the fact that PahZ2$_{\text{KT-1}}$ function under physiologic conditions would necessarily include the presence of salt ions that may impact how PahZ2$_{\text{KT-1}}$ interacts with its anionic substrate. To examine the effect of salt on PahZ2$_{\text{KT-1}}$ activity, we performed additional PahZ2$_{\text{KT-1}}$ activity assays wherein enzyme was incubated with poly(aspartic acid) substrate in the presence of [NaCl] = 0-0.5 M. For each condition, time points were quenched at specific time intervals and subjected to analysis by gel permeation chromatography (GPC). The resulting chromatograms can then be used to quantify product formation as a function of time. Figure 3.6A illustrates the resulting plot of relative product formation as a function of time. In the absence of NaCl, a $\sim$360 min lag phase is observed prior to product formation. By determining the slope of the line describing product formation from 360 to 1500 min, we can place an estimate on the reaction initial velocity. In contrast, conditions including 0.15 and 0.5 M NaCl yield the observation of product formation within 5-10 min of incubation (Figure 3.6B,C). Figure 3.6D presents a plot of apparent initial velocity as a function of [NaCl]. The solid line represents a linear least-squares fit to a straight line with slope equal to (377 $\pm$ 6) x 10^{-6} relative product min^{-1} mM^{-1} NaCl. Moreover, the observation of a linear and positive increase in initial velocity with increasing [NaCl] without a significant decrease in apparent time course amplitude (Figure 3.6B,C) suggests that salt stimulates PahZ2$_{\text{KT-1}}$ catalytic function without significant impairment to substrate binding leading to product formation.

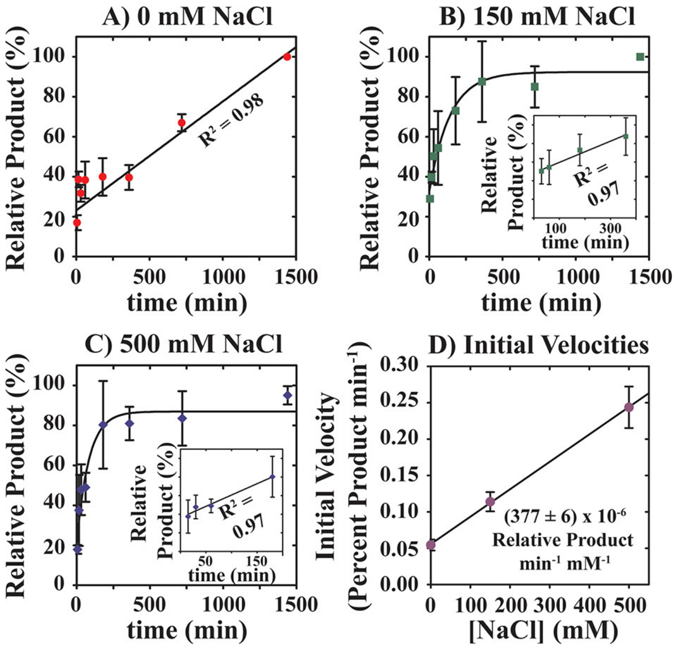


Figure 3.6: Initial velocity analysis reveals PahZ2$_{\text{KT-1}}$ catalyzed dPAA degradation is stimulated by [NaCl]. (A-C) Plots of relative product formation as a function of incubation time in the presence of 0 (A), 150 (B), and 500 (C) mM NaCl were subjected to linear least-squares analysis to estimate the slope describing product formation as a function of time for the linear phase observed in each plot. (D) A plot of initial velocity as a function of [NaCl] reveals a linear dependence with slope equal to (377 ± 6) x 10 6 relative product min^{-1} mM^{-1} NaCl. All data points represent the mean of triplicate measurements with associated standard deviation presented as vertical error bars. All insets depict linear data utilized for initial velocity analysis alongside indicated correlation coefficient.

3.4.6 Poly(aspartic acid) Binding Site Prediction.

Inspection of the PahZ2$_{\text{KT-1}}$ crystal structure does not sufficiently address two critical components of mechanistic importance: (1) poly(aspartic acid) binding site identification and (2) the role of NaCl in promoting PahZ2$_{\text{KT-1}}$ activity. To first predict where substrate binding may occur, we turned to the cocrystal structure of DapE bound to products succinic acid and 2,6-diaminopimelic acid, which reveals protein/ligand interactions that occur in both the catalytic and dimerization domains.[38] The interaction of DapE with 2,6-diaminopimelic acid is supported by R258 and S181, which are each contributed from the dimerization domain. By comparison, DapE R258 is conserved in both sequence and structure in PahZ2$_{\text{KT-1}}$ as R270. Similarly, DapE S181 is conserved in PahZ2$_{\text{KT-1}}$ sequence and structure as S198. In contrast, similar conserved elements were not located in the PahZ2$_{\text{KT-1}}$ catalytic domain.

To identify the putative poly(aspartic acid) binding site on PahZ2$_{\text{KT-1}}$, an α-linked oligo(aspartic acid) pentapeptide substrate, PAA$_5$, was docked to PahZ2$_{\text{KT-1}}$ without constraints as described in the Materials and Methods sections. Following PAA$_5$ docking, the protein:substrate complex was brought to equilibrium over 200 ns in the absence and presence of 0.5 M NaCl in order to assess the impact of solution ionic strength on stabilizing electrostatic interactions between protein and ligand. Inspection of the equilibrated PahZ2$_{\text{KT-1}}$:PAA$_5$ complex reveals that substrate binds at the hinge between dimerization and catalytic domains such that contacts are made across both domains independent of incubation conditions (Figure 3.7A,B). The overall positioning of PAA$_5$ in the PahZ2$_{\text{KT-1}}$ binding site is similar to contacts made between the structural homologues DapE and carboxypeptidase G2 with their substrates where intermolecular interactions occur between substrate and both dimerization/catalytic domains.[38,85,86]

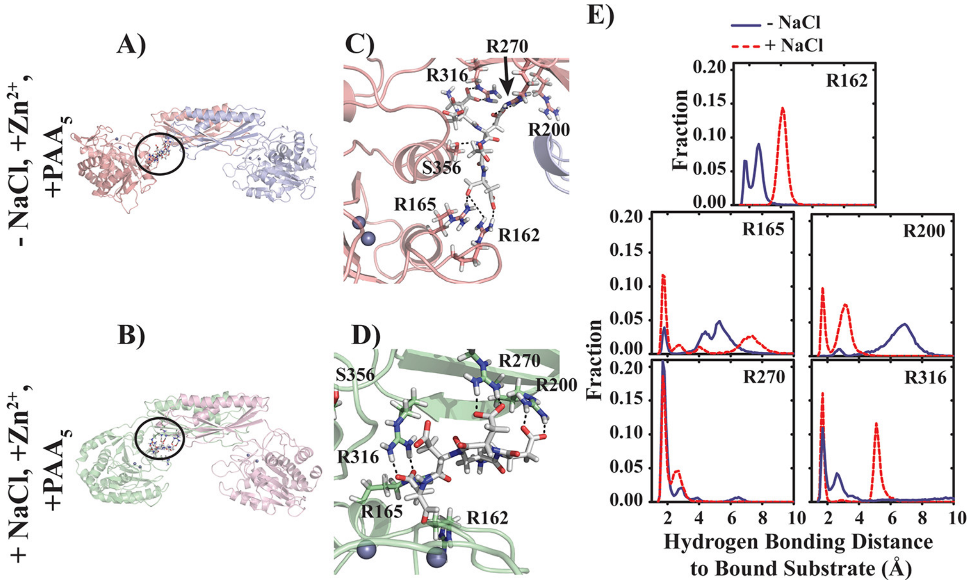


Figure 3.7: In silico identification of ligand binding site. PAA$_5$ ligand docks to a PahZ2$_{\text{KT-1}}$ surface spanning the dimerization and catalytic domains with NaCl absent and Zn^{2+} present (A) as well as with both NaCl and Zn^{2+} present (B). (C, D) The PAA$_5$ position is shown for each simulation condition and reveals hydrogen bonding interactions with R162, R165, R200, R270, and R316 that exhibit NaCl-specific behaviors. (E) Analysis of overall hydrogen bonding frequency between PahZ2$_{\text{KT-1}}$ arginine residues and PAA$_5$ in the absence (solid blue line) or presence (dashed red line) of NaCl. Mean hydrogen bonding frequency under each condition is described in the text.

We note condition-dependent differences that exist for PahZ2$_{\text{KT-1}}$ with respect to the orientation of PAA$_5$ in the binding site not previously reported for structural homologues. Figure 3.7A reveals that incubation of Zn^{2+}/ PAA$_5$-bound

PahZ2$_{\text{KT-1}}$ in the absence of salt promotes an open conformation that does not position the catalytic zinc ions adjacent to bound polypeptide substrate. In contrast, the incubation of Zn^{2+} / PAA$_5$-bound PahZ2$_{\text{KT-1}}$ with 0.5 M NaCl promotes a closed conformation that brings catalytic zinc atoms into proximity of the bound PAA$_5$ substrate (Figure 3.7B). Inspection of the local electrostatic environment for each PahZ2$_{\text{KT-1}}$ conformation indicates that each case involves substrate binding to surfaces enriched with cationic residues (Figure 3.7C,D). However, in all cases, we note that the C-terminal end of the substrate is positioned adjacent to catalytic zinc atoms, which is consistent with the existing literature reports preference for cleavage of C-terminal aspartic acid units.[28,31]

PAA$_5$ binding to open versus closed PahZ2$_{\text{KT-1}}$ involves two unique substrate binding poses. Figure 3.7C demonstrates that PAA$_5$ binding to the open conformation involves interactions in the dimerization domain, the hinge region, and the catalytic domain. Hydrogen bonding interactions are observed between PAA$_5$ and R270 (dimerization domain), R316 (dimerization domain), S356 (hinge region), R162 (catalytic domain), and R165 (catalytic domain). In contrast, incubation conditions that include 0.5 M NaCl and Zn^{2+} / PAA$_5$-bound PahZ2$_{\text{KT-1}}$ promote a closed conformation that brings catalytic zinc atoms into the proximity of bound substrate. However, the closed PahZ2$_{\text{KT-1}}$ conformation promotes a unique substrate binding pose that eliminates interactions in the hinge region. Instead, interactions between PAA$_5$ with enzyme involve dimerization domain residues R270, R200, and R316 as well as catalytic domain residues R165 and R162 (Figure 3.7D).

Calculation of hydrogen bonding distance distributions from MD trajectories for interactions between PAA_5 and R162, R165, R200, R70, and R316 reveal NaCl-dependent behaviors. Figure 3.7E reveals three general groupings for these residues. We observe that the distribution of R162:PAA_5 hydrogen bonding interactions is bimodal in the absence of NaCl with a median distance equal to 2.58 Å but transitions to a unimodal distribution with median distance of 4.11 Å when salt is present. From this, R162:PAA_5 interactions apparently weaken in the presence of NaCl. In contrast, R165 and R200 interactions with PAA_5 undergo stabilization in the presence of NaCl. The median R165 interaction distances in the absence or presence of 0.5 M NaCl are 5.10 and 3.79 Å, respectively. However, the R165 interaction distance distribution is bimodal in the presence of 0.5 M NaCl with a major population adopting distances equal to 1.72 Å alongside a noninteracting minor population with a mean distance equal to 7.21 Å. Similarly, R200 interaction distances shorten in the presence of NaCl such that median distances equal 6.60 and 2.93 Å in the absence and presence of NaCl, respectively. Thus, R165 and R200 appear to form stronger interactions with PAA_5 in the presence of NaCl, where evidence for this conclusion is drawn from the apparent shortened hydrogen bonding distances observed in the presence of NaCl relative to conditions without NaCl. The last general grouping includes only R270, which is observed to interact with PAA_5 without impact from NaCl. Observed median interaction distances for R270 in the absence or presence of NaCl are 1.84 and 1.90 Å, respectively.

An exception to this overall grouping strategy is R316, which transitions from a median hydrogen bonding distance equal to 2.68 Å in the absence of NaCl to a highly bimodal distribution with a median distance equal to 4.98 Å in the presence of NaCl. R316 cannot be strictly classified alongside R162 as undergoing hydrogen bond weakening in the presence of NaCl. The R316:PAA$_5$ hydrogen bonding distance distribution calculated in the presence of NaCl includes two populations, where one increases the frequency of short hydrogen bonds and the other effectively does not interact. Taken together, these data provide insight into how PahZ2$_{\text{KT-1}}$ conformation switching between open and closed states may directly translate into catalytic activity.

3.4.7 NaCl Influences Conformational Dynamics.

The observation of an apparent NaCl-dependent shift in PahZ2$_{\text{KT-1}}$ between open and closed conformations provides a structural rationale for the [NaCl] promoted oligo(aspartic acid) hydrolysis reported in Figure 3.6. Given that a closed conformation brings catalytic zinc atoms into the proximity of bound substrate and that NaCl may promote such a transition, it is reasonable to infer that salt may consequently stimulate catalytic function by shifting the equilibrium between open and closed states. Additional molecular dynamics experiments were performed in order to better understand the impact of NaCl on the PahZ2$_{\text{KT-1}}$ conformational landscape. MD trajectories were prepared as described in the Materials and Methods sections in the presence of 0 and 0.5 M NaCl for 200 ns each in triplicate with bound PAA$_5$ ligand. A plot of free energy as a function of radius of gyration and root-mean-square deviation (RMSD) provides a topological perspective that relates protein conformation to free energy. Figure 3.8A presents such a conformational landscape plot for PahZ2$_{\text{KT-1}}$ incubated in the absence of supplemental NaCl and illustrates a restricted set of accessible low-energy conformations that are described by radii of gyration between ~3.6 and 3.7 nm and RMSD values equal to 0.2-0.3 Å. In contrast, MD simulations including 0.5 M NaCl yield a broadened accessible conformational space that spans radii of gyration between ~3.7 and 3.9 nm and RMSD values equal to 0.2-0.5 Å (Figure 3.8B). Therefore, these data allow for the conclusion that NaCl promotes an increase in accessible conformational switching between similar low-energy states.

Complementary dynamic light scattering (DLS) experiments were performed as a function of [NaCl] to assess dimer integrity over the discussed experimental conditions. DLS experiments were performed in the presence of [NaCl] = 0-0.5 M. Figure 3.8C presents a plot of signal intensity as a function of particle diameter. In the absence of salt, the mean particle diameter is estimated as 8.8 ± 0.7 nm. By comparison, the PahZ2$_{\text{KT-1}}$ crystal structure predicts a diameter of ~12 nm, which is consistent with the DLS data shown in Figure 3.8C. In the presence of 0.15 or 0.5 M NaCl, particle diameters are estimated from DLS experiments as 8.9 ± 0.9 and 10 ± 2 nm, respectively. A comparison of mean particle diameter estimates for each condition by the Student's t test indicates no significant differences in means across all conditions examined here. We note an increase in peak width (polydispersity) for conditions that include 0 versus 0.5 M NaCl, which can be quantitatively described from the apparent increase in

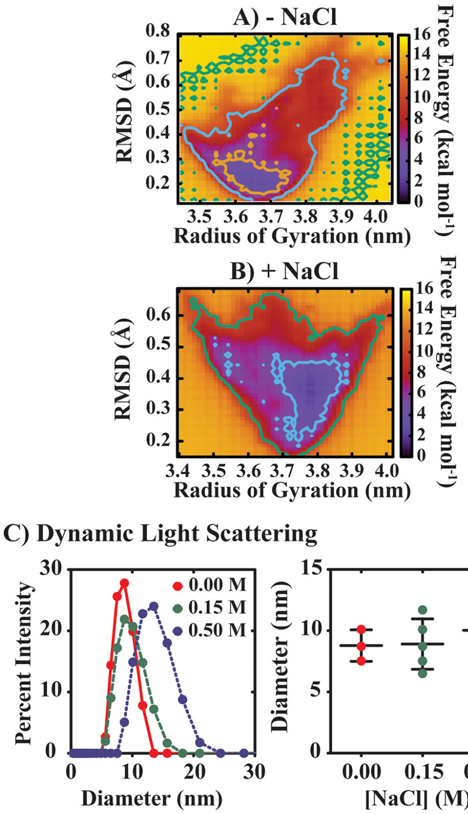


Figure 3.8: Impact of NaCl on PahZ2$_{\text{KT-1}}$ dimer structure and conformational dynamics. (A, B) Molecular dynamics simulations were performed wherein the PahZ2$_{\text{KT-1}}$ dimer was incubated in the (A) absence or (B) presence of 0.5 M NaCl, which indicates that NaCl promotes expanded conformational access. Three-dimensional plots of RMSD, radius of gyration, and free energy (x, y, z) provide an estimate of accessible PahZ2$_{\text{KT-1}}$ conformations for each incubation condition. Calculated free energy values are colored on the indicated scale from low to high from dark purple to yellow, respectively. All simulations were performed as described in Materials and Methods sections. (C) Dynamic light scattering experiments indicate that incubation of PahZ2$_{\text{KT-1}}$ with varied [NaCl] does not perturb dimer integrity. From Z-average distributions that plot fractional scattering intensity as a function of particle size, the mean particle sizes observed in the presence of 0.00, 0.15, and 0.50 M NaCl are estimated as 8.78 ± 0.74, 8.91 ± 0.92, and 10.0 ± 2.0 nm, respectively. Each sphere represents an individual particle size measurement with horizontal lines highlighting mean and standard deviation values.

standard deviation for the high-salt condition. However, application of pairwise F-tests indicates that all combinations of variances are statistically identical.

Size-exclusion chromatography experiments were performed to complement DLS data which indicate that PahZ2$_{\text{KT-1}}$ incubation with NaCl retains the dimeric complex intact. We performed size-exclusion chromatography (SEC) experiments in the presence of 0 or 0.5 M NaCl. Samples were prepared for SEC analysis by incubation of PahZ2$_{\text{KT-1}}$ (0.75 mg/mL) overnight at 25 C° in the presence or absence of 0.5 M NaCl. Figure 3.9A presents the resulting chromatograms for each condition where conditions with 0 or 0.5 M NaCl are colored as solid blue or broken teal lines, respectively. Elution volumes corresponding to PahZ2$_{\text{KT-1}}$ monomers, dimers, and trimers have been indicated as A_1 , A_2 , and A_3, respectively. Equilibration conditions including 0 or 0.5 M NaCl yield elution profiles that are unique with respect to predicted molecular weight. Figure 3.9A,B highlights these observations wherein the PahZ2$_{\text{KT-1}}$ molecular weight is estimated as 86.6 ± 0.8 and 106.6 ± 0.9 kDa in the presence of 0 or 0.5 M NaCl, respectively. The introduction of NaCl causes an apparent shift in PahZ2$_{\text{KT-1}}$ complex stoichiometry from 1.94 ± 0.02 to 2.4 ± 0.02 when 0 or 0.5 M NaCl is present, respectively. Thus, the PahZ2$_{\text{KT-1}}$ complex assembles into an apparent dimer in the presence and absence of NaCl. The apparent shift in complex molecular weight may reflect a [NaCl]-dependent conformational change that alters migration through the size-exclusion column. Taken together, these data are consistent with the formation of a stable PahZ2$_{\text{KT-1}}$ dimer independent of NaCl, though with condition-dependent changes in dimer conformational landscape.

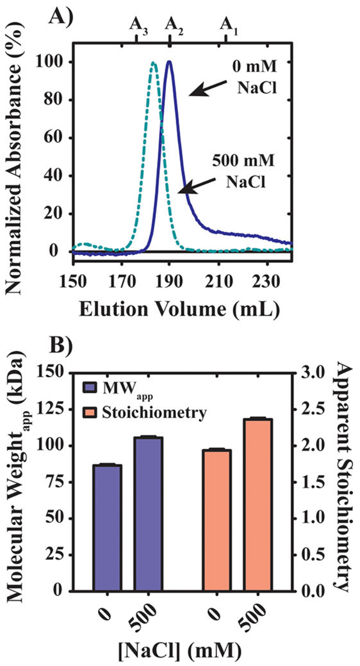


Figure 3.9: NaCl promotes altered behavior by size-exclusion chromatography. (A) Samples were prepared for SEC analysis by incubation of PahZ2$_{\text{KT-1}}$ overnight at 25 C° in the presence or absence of NaCl. Chromatograms resulting from incubation conditions including 0 or 500 mM NaCl are represented as blue solid or broken teal lines, respectively. Elution volumes that correspond to predicted PahZ2$_{\text{KT-1}}$ monomeric, dimeric, or trimeric species are labeled as A_1, A_2, and A_3, respectively. (B) Double-Y plot of apparent molecular weight and apparent stoichiometry determined in the presence of 0 and 500 mM NaCl. All experiments were performed in triplicate. Error bars represent the corresponding standard deviation.

3.4.8 Principal Component Analyses Reveal NaCl-Dependent Domain Movements.

From our activity data reported in Figure 3.6, it is apparent that NaCl stimulates PahZ2$_{\text{KT-1}}$ catalyzed oligo(aspartic acid) degradation. Combined molecular dynamics (Figures 7A,B and 8A,B) and size-exclusion (Figure 3.9A) data collectively suggest that PahZ2$_{\text{KT-1}}$ incubation with NaCl also leads to an altered conformational landscape involving open and closed states. On the basis of these observations, we asked the following: How do incubation conditions impact PahZ2$_{\text{KT-1}}$ conformational motions leading to the stabilization of a closed conformation that positions active-site zinc ions adjacent to bound peptide? To address this question, additional 200 ns MD simulations were performed in varied combinations of NaCl, Zn^{2+}, and PAA_5 as (1) -NaCl, -Zn^{2+}, -PAA_5; (2) +NaCl, -Zn^{2+}, -PAA_5; (3) -NaCl, +Zn^{2+}, -PAA_5; (4) -NaCl, +Zn^{2+}, +PAA_5; (5) +NaCl, +Zn^{2+}, -PAA_5; and (6) +NaCl, +Zn^{2+}, +PAA_5. For each experiment including bound Zn^{2+}, the cation was placed in the active site using coordinates from the X-ray structure without modification. Each production quality MD trajectory was then subjected to principal component analysis (PCA) to identify protein motions that contribute to overall conformational dynamics. Principal component analysis has found widespread application in the analysis of MD trajectories since it offers a straightforward path to the reduction of multidimensional data sets into simplified data representations with lower dimensionality while still capturing most of the information contained initially.[87–91] Thus, we can use PCA to identify the principal components contributing to overall system variance leading to overall groups of correlated atom motions.

Figure 3.10 presents the results of PCA for each condition examined by molecular dynamics alongside the major structural motion, principal component 1 (PC1). For conditions lacking NaCl, Zn^{2+}, or PAA_5, three principal components drive ∼80% of system variance, where PC1 and PC2 each contribute 46.5% and 21.6%, respectively, as indicated in Figure 3.10A. Inspection of extracted structural models that describe PC1 and PC2 reveals subtle motions related to the catalytic domain. Calculation of the contribution per residue to PC1 and PC2 reveals a minimal contribution from the dimerization domain but a greater contribution from the catalytic domain in both chains A and B (Figure B.11). Structural movements captured in PC1 can best be described as a rotational motion parallel to the plane formed by the dimerization domain. In contrast, the structural movements associated with PC2 occur as an open-close motion perpendicular to the plane of the dimerization domain. However, the range of motion for PC2 combined with its contribution to overall system variance suggests that while the closed PahZ2$_{\text{KT-1}}$ conformation is sampled under these conditions, it is not significantly stabilized.

In contrast to conditions lacking salt, PCA resulting from MD trajectories collected in the presence of 0.5 M NaCl without Zn^{2+} or PAA_5 suggests condition-dependent conformational dynamics. Figure 3.10B illustrates similar principal component contributions relative to conditions lacking salt such that PC1 and PC2 each contribute 43.5% and 29.7%, respectively, to overall system variance. However, inspection of extracted motions for PC1 and PC2 reveals significant differences between 0 and 0.5 M NaCl conditions. In the presence of NaCl, PC1 is observed to describe a significant open-close event involving movement of the catalytic domain perpendicular to the plane formed by the dimerization domain. While similar to PC2 in the absence of salt, the range of movement

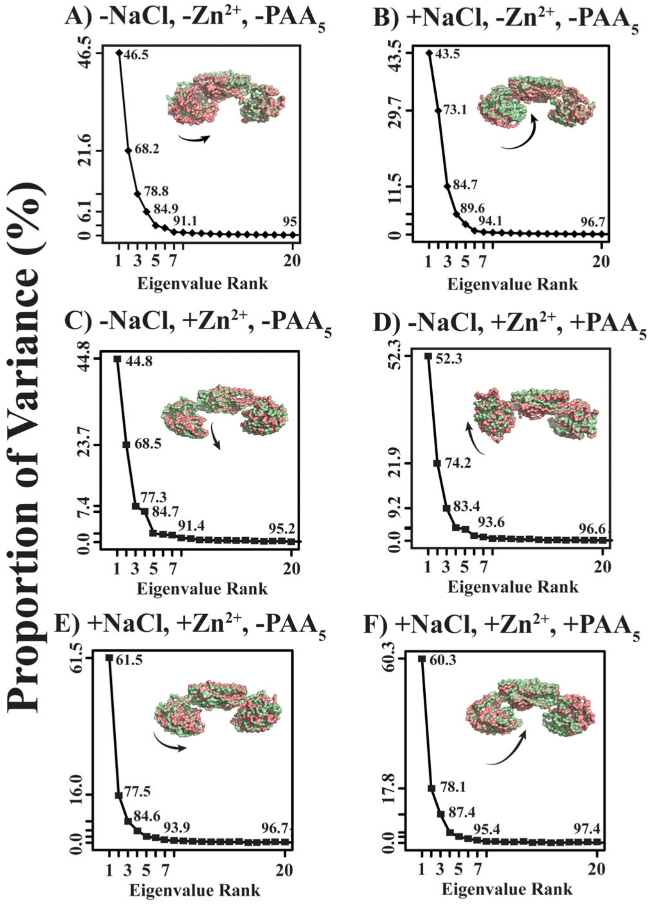


Figure 3.10: Principal component analysis (PCA) indicates that PahZ2$_{\text{KT-1}}$ conformational dynamics are condition-dependent. Scree plots illustrate the contribution of individual principal components to overall system variance as a percentage for (A) apoprotein and conditions that include (B) +NaCl, -Zn^{2+}, -PAA5; (C) -NaCl, +Zn^{2+}, -PAA5; (D) -NaCl, +Zn^{2+}, +PAA5; (E) +NaCl, +Zn^{2+}, -PAA5; and (F) +NaCl, +Zn^{2+}, +PAA5. Y-Axis labeling indicates the contribution of major principal components to system variance. Internal plot labeling indicates total variance captured with each additional principal component. Each condition includes a structural representation of the motion represented in principal component 1 (PC1). All structural representations are shown as surface views with the direction of motion indicated by an arrow.

is greater for PC1 when 0.5 M NaCl is present such that PC2 (0 M NaCl) and PC1 (0.5 M NaCl) involve ∼6 and 11 Å motions, respectively. Furthermore, PC1 observed with salt present contributes more than twice as much to overall system variance relative to PC2 (0 M NaCl), thereby predicting a significant open-close motion for PahZ2$_{\text{KT-1}}$ when salt is present. However, it is worth noting that the conformational landscapes presented in Figure 3.8A,B for conditions with 0 versus 0.5 M NaCl revealed an increased overall RMSD when salt is present, which suggests that this open-close event captured by PC1 in the presence of salt likely does not lead to a stable closed conformation. Instead, PahZ2$_{\text{KT-1}}$ likely experiences frequent transitions between open and closed conformational states when in the presence of NaCl without Zn^{2+} or PAA_5.

Additional PCA analysis of MD trajectories performed in the presence of different combinations of NaCl, Zn^{2+}, and PAA_5, but with not all present, do not suggest the formation of a stable PahZ2$_{\text{KT-1}}$ closed conformation. Figure 3.10C,D highlight that conditions including Zn^{2+} without or with bound PAA_5 do not promote catalytic domain motion leading to a closed conformation. Supplemental free energy landscape calculations support this notion through increased RMSD and radii of gyration estimates (Figure B.12). For each of these conditions, PC1 is observed to correspond to catalytic domain movements yielding an open or semiopen conformation with a high degree of active-site solvent accessibility. Similarly, simulation conditions including 0.5 M NaCl for Zn^{2+} -bound PahZ2$_{\text{KT-1}}$ reveal little preference for motions supporting a stable closed conformation (Figure 3.10E). Examination of extracted structural motions for

PC1 reveals these motions to represent a rocking motion wherein the catalytic domain rotates perpendicular to the dimerization domain such that the protein moves between open and semiopen states.

In contrast, Figure 3.10F demonstrates that incubation of Zn^{2+} / PAA_5-bound $PahZ2_{KT\text{-}1}$ with 0.5 M NaCl promotes a closed conformation that is unique compared to that observed under any other conditions examined here. Extracted structural motions reveal a stable closed conformation wherein the catalytic and dimerization domains frequently contact one another via rocking motions that never lead to a fully open conformation. In fact, these motions attributed to PC1 under conditions with 0.5 M NaCl and bound Zn^{2+} / PAA_5 contribute 60.3% to total system variance. A supplemental calculation of conformational landscapes indicates that the incubation of Zn^{2+} / PAA_5-bound $PahZ2_{KT\text{-}1}$ with 0.5 M NaCl promotes a restriction in accessible low-energy states with decreased RMSD and radii of gyration relative to other incubation conditions (Figure B.12). These data support a model wherein $PahZ2_{KT\text{-}1}$ binding of Zn^{2+} and PAA_5 substrate only adopts a closed conformation in the presence of NaCl, where the closed conformation likely stabilizes the position of all relevant active-site elements for catalysis.

3.5 Conclusion

$PahZ2_{KT\text{-}1}$ adopts a stable dimeric structure supported by nonpolar interactions occurring at the subunit interface. The X-ray structure reveals that each subunit contains two distinct domains, a dimerization and a catalytic domain. The catalytic domain is characterized as an M28 metallopeptidase family member that displays optimal enzymatic activity upon binding of dual zinc metal ions, where zinc binding occurs at Zn_I and Zn_{II} sites with dissociation equilibrium constants equal to 16 pM and 49 nM, respectively. Oligo(aspartic acid) substrate binding involves electrostatic interactions that occur across both $PahZ2_{KT\text{-}1}$ domains. The data reported here indicate that optimal catalysis occurs only in the presence of NaCl, Zn^{2+}, and peptide substrate due to stabilization of a closed $PahZ2_{KT\text{-}1}$ conformational state that precisely positions catalytic zinc metal ions adjacent to bound peptide. The linear dependence in $PahZ2_{KT\text{-}1}$ catalyzed oligo(aspartic acid) degradation as a function of [NaCl] suggests that salt promotes a shift in the equilibrium between the open and closed conformation toward the stabilization of a closed state enzyme of catalytic relevance. Importantly, knowledge of the $PahZ2_{KT\text{-}1}$ structure alongside the PahZ1 KT-1 structure recently reported provides a structural picture of the enzymatic participants responsible for conversion of poly(aspartic acid) into free aspartic acid. Therefore, the current work provides a complete structural and mechanistic framework to build from in protein engineering studies that bring $PahZ1_{KT\text{-}1}$ and $PahZ2_{KT\text{-}1}$ together toward the goal of developing novel biodegradation agents toward poly(aspartic acid) and other water-soluble polymers.

References

(1) Geyer, R.; Jambeck, J.; Law, K. Production, use, and fate of all plastics ever made. *Science Advances* **2017**, *3*, 1700782.

(2) Lebreton, L. et al. Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. *Scientific Reports* **2018**, *8*, 4666.

(3) Jahnke, A.; Arp, H.; Escher, B.; Gewert, B.; Gorokhova, E.; Kühnel, D.; Ogonowski, M.; Potthoff, A.; Rummel, C.; Schmitt-Jansen, M. Reducing Uncertainty and Confronting Ignorance about the Possible Impacts of Weathering Plastic in the Marine Environment. *Environmental Science & Technology Letters* **2017**, *4*, 85–90.

(4) Worm, B.; Lotze, H.; Jubinville, I.; Wilcox, C.; Jambeck, J. Plastic as a Persistent Marine Pollutant. *Annual Review of Environment and Resources* **2017**, *42*, 1–26.

(5) Jambeck, J.; Geyer, R.; Wilcox, C.; Siegler, T.; Perryman, M.; Andrady, A.; Narayan, R.; Law, K. Plastic waste inputs from land into the ocean. *Science* **2015**, *347*, 768–771.

(6) Waring, R.; Harris, R.; Mitchell, S. Plastic contamination of the food chain: A threat to human health? *Maturitas* **2018**, *115*, 64–68.

(7) Billiet, S.; Trenor, S. 100th Anniversary of Macromolecular Science Viewpoint: Needs for Plastics Packaging Circularity. *ACS Macro Letters* **2020**, *9*, 1376–1390.

(8) Han, X.; Liu, W.; Huang, J.-W.; Ma, J.; Zheng, Y.; Ko, T.-P.; Xu, L.; Cheng, Y.-S.; Chen, C.-C.; Guo, R.-T. Structural insight into catalytic mechanism of PET hydrolase. *Nature Communications* **2017**, *8*, 2106.

(9) Hiraga, K.; Taniguchi, I.; Yoshida, S.; Kimura, Y.; Oda, K. Biodegradation of waste PET. *EMBO reports* **2019**, *20*, 49365.

(10) Joo, S.; Cho, I.; Seo, H.; Son, H.; Sagong, H.-Y.; Shin, T.; Choi, S.; Lee, S.; Kim, K.-J. Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nature Communications* **2018**, *9*, 382.

(11) Palm, G.; Reisky, L.; Böttcher, D.; Müller, H.; Michels, E.; Walczak, M.; Berndt, L.; Weiss, M.; Bornscheuer, U.; Weber, G. Structure of the plastic-degrading Ideonella sakaiensis MHETase bound to a substrate. *Nature Communications* **2019**, *10*, 1717.

(12) Yoshida, S.; Hiraga, K.; Takehana, T.; Taniguchi, I.; Yamaji, H.; Maeda, Y.; Toyohara, K.; Miyamoto, K.; Kimura, Y.; Oda, K. A bacterium that degrades and assimilates poly(ethylene terephthalate. *Science* **2016**, *351*, 1196–1199.

(13) Huppertsberg, S.; Zahn, D.; Pauelsen, F.; Reemtsma, T.; Knepper, T. Making waves: Water-soluble polymers in the aquatic environment: An overlooked class of synthetic polymers? *Water Research* **2020**, *181*, 115931.

(14) Arp, H.; Knutsen, H. Could We Spare a Moment of the Spotlight for Persistent, Water-Soluble Polymers? *Environmental Science & Technology* **2020**, *54*, 3–5.

(15) Swift, G. In *Degradable Polymers: Principles and Applications*, Scott, G., Ed.; Dordrecht: Springer Netherlands, 2002, 379–412.

(16) Hennecke, D.; Bauer, A.; Herrchen, M.; Wischerhoff, E.; Gores, F. Cationic polyacrylamide copolymers (PAMs): environmental half life determination in sludge-treated soil. *Environ Sci Eur* **2018**, *30*, 16.

(17) Jop, K.; Guiney, P.; Christensen, K.; Silberhorn, E. Environmental fate assessment of two synthetic polycarboxylate polymers. *Ecotoxicol Environ Saf* **1997**, *37*, 229–37.

(18) Xiong, B.; Loss, R.; Shields, D.; Pawlik, T.; Hochreiter, R.; Zydney, A.; Kumar, M. Polyacrylamide degradation and its implications in environmental systems. *npj Clean Water* **2018**, *1*, DOI: `10.1038/s41545-018-0016-8`.

(19) Buczek, S.; Cope, W.; McLaughlin, R.; Kwak, T. Acute toxicity of polyacrylamide flocculants to early life stages of freshwater mussels. *Environ Toxicol Chem* **2017**, *36*, 2715–2721.

(20) Thombre, S.; Sarwade, B. Synthesis and Biodegradability of Polyaspartic Acid: A Critical Review. *Journal of Macromolecular Science, Part A* **2005**, *42*, 1299–1315.

(21) Paik, Y.; Simon, E.; Swift, G. In *Hydrophilic Polymers, American Chemical Society*, 1996; Vol. 248, 79–98.

(22) Hu, M.-m; Dou, Q.-h; Cui, X.-m; Lou, Y.-h; Zhuge, Y.-p Polyaspartic acid mediates the absorption and translocation of mineral elements in tomato

seedlings under combined copper and cadmium stress. *Journal of Integrative Agriculture* **2019**, *18*, 1130–1137.

(23) Bai, M.; Wilske, B.; Buegger, F.; Esperschütz, J.; Bach, M.; Frede, H.-G.; Breuer, L. Relevance of nonfunctional linear polyacrylic acid for the biodegradation of superabsorbent polymer in soils. *Environmental Science & Pollution Research* **2015**, *22*, 5444–5452.

(24) Wilske, B.; Bai, M.; Lindenstruth, B.; Bach, M.; Rezaie, Z.; Frede, H.-G.; Breuer, L. Biodegradability of a polyacrylate superabsorbent in agricultural soil. *Environmental Science & Pollution Research* **2014**, *21*, 9453–9460.

(25) Amass, W.; Amass, A.; Tighe, B. A review of biodegradable polymers: uses, current developments in the synthesis and characterization of biodegradable polyesters, blends of biodegradable polymers and recent advances in biodegradation studies. *Polymer International* **1998**, *47*, 89–144.

(26) Freeman, M.; Paik, Y.; Swift, G.; Wilczynski, R.; Wolk, S.; Yocom, K. Biodegradability of Polycarboxylates: Structure—Activity Studies. In Hydrogels and Biodegradable Polymers for Bioapplications. *American Chemical Society* **1996**, *627*, 118–136.

(27) Dorman, D.; Foster, M.; Olesnevich, B.; Bolon, B.; Castel, A.; Sokolsky-Papkov, M.; Mariani, C. Toxicity associated with ingestion of a polyacrylic acid hydrogel dog pad. *J Vet Diagn Invest* **2018**, *30*, 708–714.

(28) Hiraishi, T.; Maeda, M. Poly(aspartate) hydrolases: biochemical properties and applications. *Appl Microbiol Biotechnol* **2011**, *91*, 895–903.

(29) Tabata, K.; Abe, H.; Doi, Y. Microbial Degradation of Poly(aspartic acid) by Two Isolated Strains of Pedobacter sp. and Sphingomonas sp. *Biomacromolecules* **2000**, *1*, 157–161.

(30) Tabata, K.; Kasuya, K.; Abe, H.; Masuda, K.; Doi, Y. Poly(aspartic acid) degradation by a Sphingomonas sp. isolated from freshwater. *Appl Environ Microbiol* **1999**, *65*, 4268–4270.

(31) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Abe, H.; Yamato, I.; Doi, Y. Biochemical and molecular characterization of poly(aspartic acid) hydrolase-2 from sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 1285–92.

(32) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Yamato, I.; Doi, Y. Genetic Analysis and Characterization of Poly(aspartic acid) Hydrolase-1 from Sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 80–86.

(33) Hiraishi, T.; Kajiyama, M.; Yamato, I.; Doi, Y. Enzymatic hydrolysis of alpha- and beta-oligo(L-aspartic acid)s by poly(aspartic acid) hydrolases-1 and 2 from Sphingomonas sp. *KT-1. Macromol Biosci* **2004**, *4*, 330–9.

(34) Tabata, K.; Kajiyama, M.; Hiraishi, T.; Abe, H.; Yamato, I.; Doi, Y. Purification and characterization of poly(aspartic acid) hydrolase from Sphingomonas sp. *KT-1. Biomacromolecules* **2001**, *2*, 1155–60.

(35) Hiraishi, T.; Masuda, E.; Kanayama, N.; Nagata, M.; Doi, Y.; Abe, H.; Maeda, M. Cloning of poly(aspartic acid) (PAA) hydrolase-1 gene from Pedobacter sp. KP-2 and hydrolysis of thermally synthesized PAA by its gene product. *Macromol Biosci* **2009**, *9*, 10–9.

(36) Brambley, C.; Bolay, A.; Salvo, H.; Jansch, A.; Yared, T.; Miller, J.; Wallen, J.; Weiland, M. Structural Characterization of Sphingomonas sp. KT-1 PahZ1-Catalyzed Biodegradation of Thermally Synthesized Poly(aspartic acid. *ACS Sustainable Chemistry & Engineering* **2020**, *8*, 10702–10713.

(37) Rawlings, N.; Barrett, A.; Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **2012**, *40*, 343–350.

(38) Nocek, B.; Reidl, C.; Starus, A.; Heath, T.; Bienvenue, D.; Osipiuk, J.; Jedrzejczak, R.; Joachimiak, A.; Becker, D.; Holz, R. Structural Evidence of a Major Conformational Change Triggered by Substrate Binding in DapE Enzymes: Impact on the Catalytic Mechanism. *Biochemistry* **2018**, *57*, 574–584.

(39) Nocek, B.; Gillner, D.; Fan, Y.; Holz, R.; Joachimiak, A. Structural basis for catalysis by the mono- and dimetalated forms of the dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase. *J Mol Biol* **2010**, *397*, 617–26.

(40) Rowsell, S.; Pauptit, R.; Tucker, A.; Melton, R.; Blow, D.; Brick, P. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure* **1997**, *5*, 337–47.

(41) Andrzejak, S.; Moyer, S.; Quillian, B.; Shank, N. Revisiting a Green Polymerization of Aspartic Acid for a Second-Semester Introductory Organic Chemistry Laboratory: Improvements. *Learning Objectives, and Post-Laboratory Assignment Chemical Educator* **2018**, *23*, 221–226.

(42) Luft, J.; Collins, R.; Fehrman, N.; Lauricella, A.; Veatch, C.; DeTitta, G. A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J Struct Biol* **2003**, *142*, 170–9.

(43) Kabsch, W. XDS. Acta Crystallogr., Sect. D: Biol. Crystallogr. **2010**, *66*, 125–132.

(44) Terwilliger, T.; Adams, P.; Read, R.; Mccoy, A.; Moriarty, N.; Grosse-Kunstleve, R.; Afonine, P.; Zwart, P.; Hung, L. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D* **2009**, *65*, 582–601.

(45) Adams, P. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–21.

(46) Terwilliger, T.; Grosse-Kunstleve, R.; Afonine, P.; Moriarty, N.; Zwart, P.; Hung, L.; Read, R.; Adams, P. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **2008**, *64*, 61–9.

(47) Schrödinger, LLC The PyMOL Molecular Graphics System, Version 1.8, 2015.

(48) Berman, H.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I.; Bourne, P. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235–42.

(49) Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **2004**, *25*, 1605–12.

(50) Trott, O.; Olson, A. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010**, *31*, 455–61.

(51) Zoete, V.; Cuendet, M.; Grosdidier, A.; Michielin, O. SwissParam: a fast force field generation tool for small organic molecules. *J Comput Chem* **2011**, *32*, 2359–68.

(52) Abraham, M.; Murtola, T.; Schulz, R.; Pall, S.; Smith, J.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(53) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.; Berendsen, H. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, *26*, 1701–18.

(54) Berendsen, H.; Postma, J.; Gunsteren, W.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(55) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(56) Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A* **1985**, *31*, 1695–1697.

(57) Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(58) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Stud. *Physical Review Letters* **1980**, *45*, 1196–1199.

(59) Best, R.; Zhu, X.; Shim, J.; Lopes, P.; Mittal, J.; Feig, M.; Mackerell Jr, A. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **2012**, *8*, 3257–3273.

(60) Huang, J.; MacKerell Jr, A. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **2013**, *34*, 2135–45.

(61) Williams, T.; Kelley, C. Gnuplot 4.5: an interactive plotting program, pt, 2011.

(62) Grant, B.; Rodrigues, A.; ElSawy, K.; McCammon, J.; Caves, L. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695–6.

(63) Grant, B.; Skjaerven, L.; Yao, X. The Bio3D packages for structural bioinformatics. *Protein Sci* **2021**, *30*, 20–30.

(64) Team, R. RStudio: Integrated Development for R, RStudio, en, Boston, MA, 2020.

(65) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(66) MacKerell, A. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **1998**, *102*, 3586–616.

(67) Grossfield, A. WHAM: The Weighted Histogram Analysis Method, en, 2013.

(68) Holm, L. DALI and the persistence of protein shape. *Protein Sci* **2020**, *29*, 128–140.

(69) Holm, L. Using Dali for Protein Structure Comparison. *Methods Mol Biol* **2020**, *2112*, 29–42.

(70) Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007**, *372*, 774–97.

(71) Hiraishi, T.; Maeda, M. Poly(aspartate) hydrolases: biochemical properties and applications. *Appl. Microbiol. Biotechnol* **2011**, *91*, 895–903.

(72) Bienvenue, D.; Gilner, D.; Davis, R.; Bennett, B.; Holz, R. Substrate specificity, metal binding properties, and spectroscopic characterization of the DapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from Haemophilus influenzae. *Biochemistry* **2003**, *42*, 10756–63.

(73) Born, T.; Zheng, R.; Blanchard, J. Hydrolysis of N-succinyl-L,L-diaminopimelic acid by the Haemophilus influenzae dapE-encoded desuccinylase: metal activation, solvent isotope effects, and kinetic mechanism. *Biochemistry* **1998**, *37*, 10478–87.

(74) Jewell, D.; Tu, C.; Paranawithana, S.; Tanhauser, S.; LoGrasso, P.; Laipis, P.; Silverman, D. Enhancement of the catalytic properties of human carbonic anhydrase III by site-directed mutagenesis. *Biochemistry*, *1991, (30*, 1484–1490.

(75) Kogut, K.; Rowlett, R. A comparison of the mechanisms of CO2 hydration by native and Co2+-substituted carbonic anhydrase II. *J. Biol. Chem* **1987**, *262*, 16417–16424.

(76) Mock, W.; Liu, Y. Hydrolysis of picolinylprolines by prolidase. A general mechanism for the dual-metal ion containing aminopeptidases. *J. Biol. Chem* **1995**, *270*, 18437–18446.

(77) Davis, R.; Bienvenue, D.; Swierczek, S.; Gilner, D.; Rajagopal, L.; Bennett, B.; Holz, R. Kinetic and spectroscopic characterization of the E134A- and E134D-altered dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from Haemophilus influenzae. *J. Biol. Inorg. Chem*, *2206, (11*, 206–216.

(78) Dudev, T.; Lim, C. Tetrahedral vs Octahedral Zinc Complexes with Ligands of Biological Interest: A DFT/CDM Study. *J. Am. Chem. Soc* **2000**, *122*, 11146–11153.

(79) Jabrani, A.; Makamte, S.; Moreau, E.; Gharbi, Y.; Plessis, A.; Bruzzone, L.; Sanial, M.; Biou, V. Biophysical characterisation of the novel zinc binding property in Suppressor of Fused. *Sci Rep* **2017**, *7*, 11139.

(80) Guerrerio, A.; Berg, J. Metal ion affinities of the zinc finger domains of the metal responsive element-binding transcription factor-1 (MTF1. *Biochemistry* **2004**, *43*, 5437–44.

(81) Lesburg, C.; Huang, C.; Christianson, D.; Fierke, C. Histidine –> carboxamide ligand substitutions in the zinc binding site of carbonic anhydrase II alter metal coordination geometry but retain catalytic activity. *Biochemistry* **1997**, *36*, 15780–91.

(82) Stojanovic, A.; Stitham, J.; Hwa, J. Critical role of transmembrane segment zinc binding in the structure and function of rhodopsin. *J Biol Chem* **2004**, *34*, 279.

(83) Krezel, A.; Maret, W. The biological inorganic chemistry of zinc ions. *Arch Biochem Biophys* **2016**, *611*, 3–19.

(84) Outten, C.; O'Halloran, T. Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* **2001**, *292*, 2488–92.

(85) Jeyaharan, D.; Brackstone, C.; Schouten, J.; Davis, P.; Dixon, A. Characterisation of the Carboxypeptidase G2 Catalytic Site and Design of New Inhibitors for Cancer Therapy. *Chembiochem* **2018**, *19*, 1959–1968.

(86) Turra, K.; Pasqualoto, K.; Ferreira, E.; Rando, D. Molecular modeling approach to predict a binding mode for the complex methotrexate-carboxypeptidase G2. *J Mol Model* **2012**, *18*, 1867–75.

(87) Amadei, A.; Linssen, A.; Berendsen, H. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–25.

(88) Das, P.; Moll, M.; Stamati, H.; Kavraki, L.; Clementi, C. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proc Natl Acad Sci U S A* **2006**, *103*, 9885–90.

(89) Lange, O.; Grubmuller, H. Generalized correlation for biomolecular dynamics. *Proteins* **2006**, *62*, 1053–61.

(90) Hub, J.; Groot, B. Detection of functional modes in protein dynamics. *PLoS Comput Biol* **2009**, *5*, 1000480.

(91) Perez-Hernandez, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noe, F. Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* **2013**, *139*, 015102.

Chapter 4

Modelling a rationally designed chimeric poly(aspartic acid) hydrolase

Chad A. Brambley and Justin M. Miller

4.1 Abstract

Polymers of naturally-occurring aspartic acid are of interest for a variety of commercial and medical applications. Both eco-friendly and bio-compatible, poly(aspartic acid)(PAA) offers an attractive alternative to existing polycarboxylates that are believed to contribute to emerging ecological threats. A key requisite for the environmentally benign application of PAA is its enzymatic degradation. To that end, we have previously characterized a set of PAA hydrolases isolated from *Sphingomonas* sp. KT-1, $\text{PahZ1}_{\text{KT-1}}$ and $\text{PahZ2}_{\text{KT-1}}$. Here, we couple deep-learning structure prediction with other in silico modelling methods to elucidate the structure and dynamics of a chimeric poly(aspartic acid) hydrolase comprised of both $\text{PahZ1}_{\text{KT-1}}$ and $\text{PahZ2}_{\text{KT-1}}$ domains. Our results predict a dimeric assembly for the aptly named super-poly(aspartic acid) hydrolase 2:1 (suPAH2:1). Monomers are predicted to associate along a non-polar interface,

with complex formation driven largely by interactions between anti-parallel disordered regions. Taken together, these data present a modelling workflow by which the structure and dynamic details of a novel enzyme may be elucidated prior to in vitro structural determination.

4.2 Introduction

Water-soluble polymers (WSPs) are a popular class of materials used for a wide range of industrial applications, including food packaging, water treatment, and biomedicine.[1] The increasing adoption of WSPs, including both polycarboxylates and polyacrylamides, is largely predicated on a perceived set of eco-friendly attributes. For example, the water-soluble nature of these compounds means a noticeable accumulation of leeched products is rare or nonexistent. This has contributed to a prevailing assumption that they are are environmentally inert and biodegradable. The lack of an ecological impact has been challenged by evidence indicating polycarboxylates, popular for their superabsorbant properties, are environmentally stable. In fact, biodegradation occurs only rarely and at slow rates, leading to suggestions that the continued widespread use of superabsorbant polymers (SAPs) may represent an emerging threat of environmental persistence.[2–5] Additionally, while many WSPs are non-toxic, concerns have been raised over the possible accumulation of dangerous by-products, such as polyacrylamide yielding neurotoxic acrylamide monomers.[6–8] To combat these potential harms, polymers of naturally-occurring aspartic acid (PAA) may provide a commercially viable and fully biodegradable alternative.

The synthetic process for PAA involves the thermal poly-condensation of aspartic acid monomers (ASP), yielding rings of poly(succinimide) that can be opened by alkaline hydrolysis.[9,10] The result is a thermally synthesized PAA (tPAA) comprised of a mixture of α- and β-linked monomers, extensive branching, and irregular end groups. The unique structures of tPAA enable its diverse range of applications. For example, the presence of cross-linking facilitates an affinity for water molecules, allowing for use as an SAP. Pairing this with tPAA's polyanionic character further provides a means for chelation of metal ions. Several studies thus support its application toward antiscaling, agricultural soil recovery and treatment, and as a biocompatible scaffolding agent for bone grafts.[11–13] The utility of tPAA for both environmental and biomedical applications is thus well established. The true eco-friendly aptitude, however, is dependent on effective strategies for breaking down the material after its deployment.

Biodegradation of tPAA was first observed in *Sphingamonas* sp. KT-1 and *Pedobacter* sp. KT-2, two bacterial strains isolated from river water.[14,15] The enzymes responsible were later determined to be a collection of PAA hydrolases: $\text{PahZ1}_{\text{KT-1}}$, $\text{PahZ2}_{\text{KT-1}}$, and $\text{PahZ1}_{\text{KP-2}}$. In the case of *Spingamonas* sp. KT-1, $\text{PahZ1}_{\text{KT-1}}$ acts first to hydrolyze tPAA β-linkages to yield shorter chains of oligo(aspartic acid) OAA. $\text{PahZ2}_{\text{KT-1}}$ subsequently cleaves the OAA into yet smaller units of ASP, although a complete conversion into monomeric ASP has not so far been described. We have previously reported structural and dynamic characterizations of both $\text{PahZ1}_{\text{KT-1}}$ and $\text{PahZ2}_{\text{KT-1}}$ in order to gain a stronger understanding of the role of domain motions and conformational changes to substrate binding and specificity.[16,17]

Our work demonstrates PahZ1$_{\text{KT-1}}$ to function as a homodimer with anti-parallel serine triad active sites. Structural data indicate the presence of a solvent-accessible binding ridge characterized by positively charged flexible side-chains amenable to interactions with polyanionic tPAA. Molecular dynamics (MD) and mutational activity assays further identified key residues comprising the ridge to be necessary for catalysis, including a flexible arginine that may aid in the positioning of active site residues and/or substrate binding. PahZ2$_{\text{KT-1}}$ works through a notably different mechanism, belonging to the M28 family of Zn^{2+}-binding metallopeptidases rather than the serine proteases. PahZ2$_{\text{KT-1}}$ is comprised of multidomain monomeric subunits that form dimeric complexes via a conserved dimerization domain. OAA associates with a Zn^{2+}-bound catalytic domain along a flexible hinge-like region.

Taken together, the diverse mechanisms required for the efficient biodegradation of tPAA serve here to illustrate a major roadblock in the development of alternative eco-friendly polymers. A multistep degradation pathway requiring the production and purification of dual enzymes is inherently less efficient, more cost prohibitive, and more time consuming relative to a single enzyme solution. Thus we propose the construction of a rationally designed chimeric complex, featuring a fusion of both PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ functional domains. The aptly named super-PahZ (suPAH) construct could theoretically facilitate the complete biodegradation of tPAA to OAA, and subsequently of OAA to smaller units of ASP. It is hoped that additional future engineering efforts will allow for a more complete conversion of OAA to monomeric ASP. These products could then be recovered and recycled to manufacture more tPAA in a so-called cradle-to-cradle synthetic mechanism. Here, we present a prerequisite modelling strategy, guided by deep learning and molecular dynamics, to both probe the feasibility of our chimeric design, as well as to improve our understanding of how such a chimera might fold and function in solution.

4.3 Results and discussion

4.3.1 AI assisted modelling of putative PahZ1$_{\text{KT-1}}$ homologues reveals modularity of PahZ-like domains

To better understand how a suPAH fusion protein might behave, we sought first to develop a broader awareness of possible variations in domain architecture, as well as to identify potential PahZ-like domains being utilized elsewhere in nature. To begin, we performed a Blast[18] search of the PahZ1$_{\text{KT-1}}$ primary sequence to identify a series of putative homologues sharing greater than 40% sequence identity. Those demonstrating greater than 70% identity, along with the sequence of PahZ1$_{\text{KT-1}}$ itself as controlled reference, were subsequently used for theoretical structure prediction using Robetta, a RoseTTAFold web server.[19,20] The results of these predictions are shown in Figure 4.1 and are labelled according to Uniprot accession number.[21] Uniprot ID Q7WSC1 corresponds to the Robetta predicted PahZ1$_{\text{KT-1}}$ control sequence. Overall confidence is reported for each prediction on a scale of 0.0 - 1.0 where 1.0 represents the highest possible score. Prediction error estimates per residue are also provided in Figure 4.1C for each model. Note that large peaks indicate decreased confidence in the predicted fold, however, nearly all error exceeding 2 Å corresponds to unstructured motifs. Root mean square deviation (RMSD) calculations are reported here as a result of a pairwise alignment of each homologue against the PahZ1$_{\text{KT-1}}$ crystal structure, shown in Figure 4.1A-B in blue and violet, respectively. All predicted structures demonstrate considerable conservation of secondary and tertiary structures relative to PahZ1, as evidenced by RMSD values of less than 1.0 Å. However, catalytically-relevant deviations can be observed along the PAA binding pocket, as can be seen in the provided electrostatic surface maps. For clarity, positive and negatively charged regions are highlighted in blue and red, respectively. As previously characterized for PahZ1$_{\text{KT-1}}$,[17] this pocket is comprised of a cleft of positive charge adjacent to a catalytic serine triad. Analogous clefts are observed among the putative structures, with each one demonstrating unique arrangements and surface coverage. Considering each homologous sequence belongs to a different bacterial genus, variant binding site architectures are likely owed to differences in substrate specificity. Future experiments involving recombinant enzyme activity assays and structural determination are expected to yield further insights.

The largest structural deviation from PahZ1$_{\text{KT-1}}$ reported here is shown in Figure 4.1B. Uniprot accession number A0A512MCT2 (MCT2) reveals a predicted

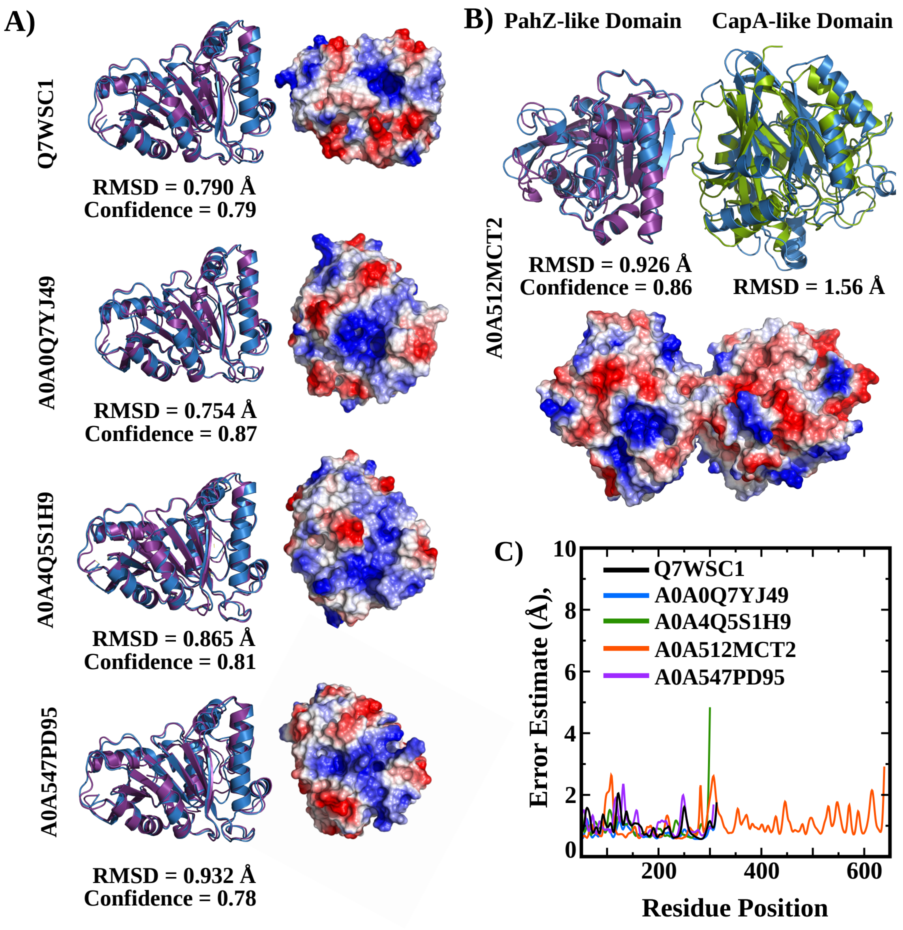


Figure 4.1: Predicted structures of putative PahZ1$_{\mathrm{KT\text{-}1}}$ homologues. **A-B:** Each model labelled by Uniprot accession, includes pairwise alignments with PahZ1$_{\mathrm{KT\text{-}1}}$ and electrostatic surface maps. Predicted structures and PahZ1$_{\mathrm{KT\text{-}1}}$ are shown in blue and violet, respectively. RMSD and confidence scores are shown for each structure. **C:** Per-residue error estimates in angstroms for each predicted model.

multidomain complex, wherein a PahZ-like domain has been fused with a sequence sharing considerable homology with a capsular poly-γ-glutamic acid (PGA) synthetase (capA) from *Bacillus subtilis*. CapA is believed to contribute to the synthesis of *Bacillus sp.* capsules via transport of PGA.[22–26] A search of the protein data bank for CapA yielded no structures comparable to our predicted result, however, a deposition into the Alphafold structure prediction database[27,28] identified one such model with high confidence (pLDDT $>$ 90). The result of a pairwise alignment of MCT2 against PahZ1$_{\text{KT-1}}$ and the Alphafold predicted CapA (AfCapA) is reported in Figure 4.1B. The multidomain MCT2 is shown in blue, while PahZ2$_{\text{KT-1}}$ and AfCapA are shown in violet and green, respectively. Both models exhibit significant structural homology with MCT2, as evidenced here by RMSD. These predictions lead us to consider an evolutionary fusion of domains. Additional Blast[18] searches for MCT2, PahZ1$_{\text{KT-1}}$, and AfCapA, with subsequent structure predictions of candidate sequences, yielded no other analogous fusion proteins. Nonetheless, the recombination of protein domains is a well-described and efficient evolutionary strategy, providing cells with modular scaffolds on which to construct new complexes with novel functionality.[29] These results additionally provide a theoretical framework from which to begin designing a chimeric PAA hydrolase.

4.3.2 Designing a modelling workflow for multidomain enzyme engineering

The rational reconstruction of enzyme domains requires a strong understanding of known structural features. To best ensure our efforts would yield a functional enzyme, we referred to our previous structural characterizations[16,17] to guide the computationally-led design of suPAH. Briefly, both PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$

exist in solution as homodimeric complexes. In the case of the former, subunits associate along a largely nonpolar interface in an antiparallel configuration as shown in Figure 4.2A. Each monomer features solvent-accessible and positively charged side-chains thought to be instrumental in ligand binding by positioning PAA adjacent to a serine triad active site.[17] Similarly, PahZ2$_{\text{KT-1}}$ subunits associate along antiparallel α helices and β strands with a primarily nonpolar character. The biological assembly is shown in Figure 4.2B and colorized according to chain ID. In contrast to PahZ1$_{\text{KT-1}}$, PahZ2$_{\text{KT-1}}$ subunits are of a multidomain construction. The aforementioned interface is facilitated by a structurally conserved dimerization domain, which in turn is linked via a flexible hindge to a Zn^{2+}-binding catalytic domain.[16] Here, we use these and other structural data to inform our subsequent architectural rearrangements.

Constructing the blueprints: Primary sequence determination was carried out with the goal of yielding a catalytically optimal domain arrangement. To that end, we chose to emphasize PahZ2$_{\text{KT-1}}$ as the base catalytic unit in our design. As evidenced by the presented tertiary structures in Figure 4.2A-B, PahZ2$_{\text{KT-1}}$ exhibits a more complex multi-domain architecture relative to PahZ1$_{\text{KT-1}}$. Our data also indicate that PahZ2$_{\text{KT-1}}$ catalytic mechanisms incorporate dynamic cross-chain interactions with polymer. As such, it was determined that the dimeric character of the PahZ2$_{\text{KT-1}}$ component remain intact for the suPAH complex. The strategy then was to introduce PahZ1$_{\text{KT-1}}$ monomers, engineered with a dimer-disrupting interface mutation (PahZ1$_{\text{KT-1}}$-IM), into either terminus of the PahZ2$_{\text{KT-1}}$ primary sequence. This led to two candidate designs, shown in

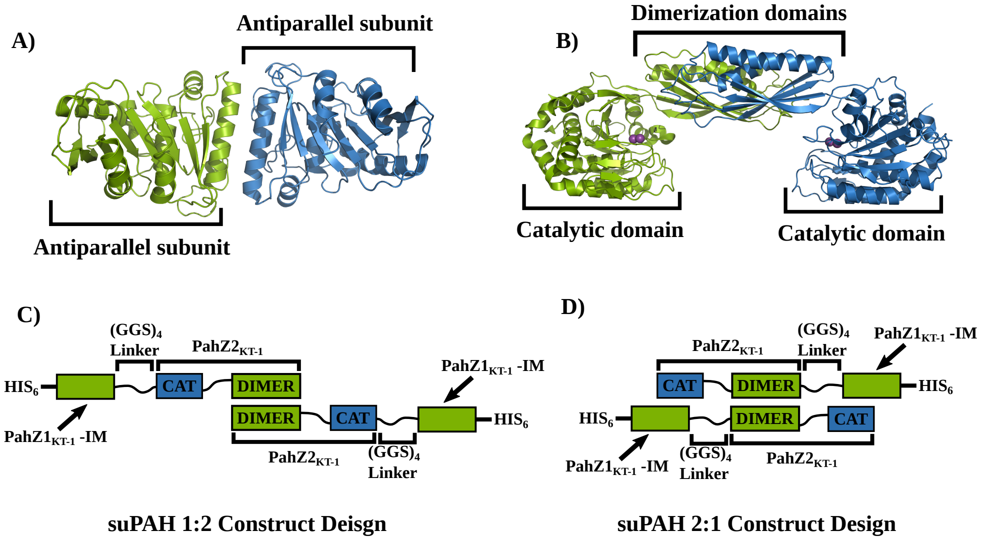


Figure 4.2: Candidate suPAH domain arrangements. **A-B:** Cartoon representations of PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ dimeric assemblies colorized according to protein chain ID. Zn ions bound to PahZ2 catalytic domains shown as violet spheres. **C-D:** Domain schematics for suPAH construction wherein PahZ1$_{\text{KT-1}}$ interface mutant (IM) monomers are inserted via flexible (GGS)$_4$ linker at PahZ2$_{\text{KT-1}}$ (**C**) N-termini (suPAH 1:2) and **D** C-termini (suPAH 2:1).

Figure 4.2C-B, wherein the IM sequence was inserted at either the N-terminus (suPAH1:2) or the C-terminus (suPAH2:1). In both cases, the attachment was coordinated via a flexible (GGS)$_4$ linker motif. N-terminal HIS$_6$ tags are included in both designs to facilitate the eventual purification of the recombinant enzyme.

Predicting suPAH structures from primary sequence: The resulting primary sequences sans HIS$_6$ tag were used for further AI-assisted structure prediction to probe feasibility of each design. Briefly, fasta files were generated for each sequence and uploaded to a Colab-fold runtime,[30] an open-access GPU platform for AI-assisted protein folding using AlphaFold2[31] or RosettaFold.[19] The top ranking predicted models for each construct can be viewed in Figure 4.3. Monomers of suPAH1:2 and 2:1 are shown in panels A and C, respectively. Model confidence is reported via predicted local difference distance test (pLDDT), a superposition-free structural comparison.[32] Unlike Robetta, ColabFold reports confidence on a scale between 1-100, whereas 100 represents complete confidence. Both constructs thus were predicted with high confidence, defined here as having a mean pLDDT > 90, although suPAH2:1 narrowly achieves a stronger result of 95.3. It should be noted that the individual suPAH domains are essentially identical to experimental observation. Figure C.1 reports the RMSD results of pairwise alignments of each suPAH domain to the relevant crystal structure. Specifically, the ColabFold prediction yielded domains that align with RMSD values of 0.397 Å, 0.435 Å, and 0.231 Å to $\text{PahZ1}_{\text{KT-1}}$, and to the $\text{PahZ2}_{\text{KT-1}}$ catalytic and dimerization domains, respectively. Confidence in the ColabFold predicted models are further elucidated by per-residue pLDDT presented for suPAH1:2 and 2:1 in panels B and D, respectively. Data are colorized to indicate pLDDT scores for all predicted models ordered by global confidence. In both instances, pLDDT scores are only observed to decrease for disordered secondary structures. This is most apparent for residues corresponding to the attached $(\text{GGS})_4$ linker, a motif designed to maintain flexibility around domains. As such, the observed drop in pLDDT for these residues to < 20 is an expected result, as regions of extremely low confidence are known to correlate with protein disorder.[31]

To address the question of whether the predicted domain arrangements are likely to work together in the sequential conversion of PAA to diAA or

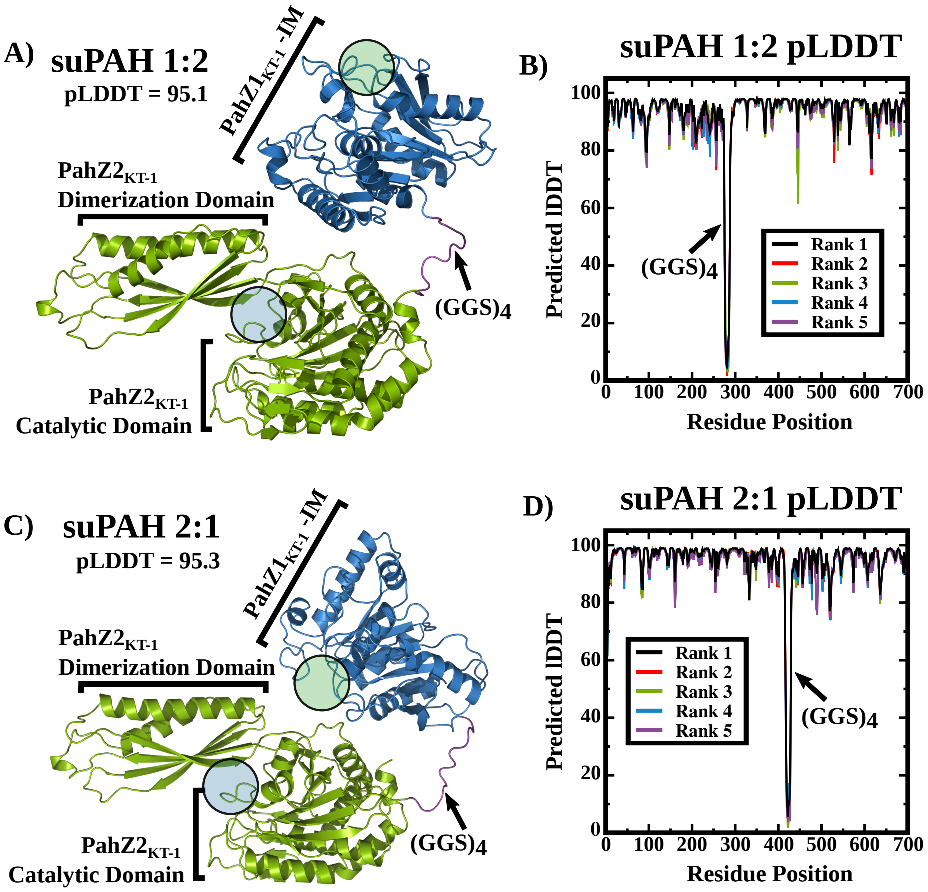


Figure 4.3: AI-generated suPAH structures. **A:** Cartoon depiction of suPAH1:2 predicted domain arrangement. PahZ1$_{\text{KT-1}}$-IM and PahZ2$_{\text{KT-1}}$ domains are represented here in blue and green, respectively. (GGS)$_4$ linker shown in violet. **B:** Per-residue pLDDT for suPAH1:2 models. **C:** Cartoon of suPAH2:1 AI prediction colorized using the same scheme as described in previous panels. **D:** Per-residue pLDDT for suPAH2:1 models. All ranks shown in both panels are observed to exhibit high confidence (pLDDT $>$ 90)

smaller units, we considered the proximity of PAA and OAA binding sites for the PahZ1$_{\text{KT-1}}$-IM and PahZ2$_{\text{KT-1}}$ domains, respectively. These regions are indicated in Figure 4.3A and 3C by transparent overlays. In the case of suPAH1:2, the PAA binding site is fully solvent-exposed and oriented away from the PahZ2$_{\text{KT-1}}$ domains. This arrangement would predict the free diffusion of OAA products into solution, rather than toward the OAA binding site. A more desirable outcome would be the channeling of OAA directly into the PahZ2$_{\text{KT-1}}$ catalytic domain, a process more easily enabled by the binding site proximity afforded by the reported suPAH2:1 construct. Substrate channeling has been shown to be beneficial for the engineered fusion of catalytic domains by directing multi-step intermediates like OAA to subsequent reaction centers. In this way, the design may avoid various time-consuming steps, including those associated with intermediate equilibration with bulk solvent and substrate pre-concentration.[33] The same orientation of binding sites is conserved amongst all ColabFold-generated suPAH2:1 models, each of which are shown in Figure C.2. RMSD calculations are also shown, resulting from pairwise alignments of each rank against the top ranking model's PahZ2$_{\text{KT-1}}$ domain. The choice of alignment selection was made to illustrate the apparent flexibility of the PahZ1$_{\text{KT-1}}$-IM domain arising from its attachment via (GGS)$_4$ linker. This observation of inter-domain substrate binding site proximity, taken together with a relative increase in global confidence, leads us to determine suPAH2:1 to be a design with greater potential of providing the desired catalytic outcome.

Molecular dynamics model refinement: To better assess whether the ColabFold suPAH2:1 structure presented previously represents an energetically favorable conformation, as well as to further refine the model, we performed a series of molecular dynamics (MD) simulations. Briefly, we prepared the system for simulation in Gromacs[34–37] with parameter sets derived from the Charmm36 all-atom forcefield.[38] Production MD was carried out for 200 ns of simulation time, with conformational convergence confirmed by RMSD block averaging. The overall conformational landscape was investigated by principal component analysis (PCA) in order to elucidate the essential dynamics. The revealed motions, along with a description of the final refined model is shown here in Figure 4.4.

With regards to the latter, suPAH2:1 was found to largely maintain a domain arrangement analogous to those detailed previously in Figure 4.3C. Specifically, the proximity of the $\mathrm{PahZ1_{KT\text{-}1}}$ and $\mathrm{PahZ2_{KT\text{-}1}}$ active sites are observed to remain consistently adjacent throughout the simulation, as evidenced by the final 200 ns structure represented in Figure 4.4A. In order to address any structural deviations, as well as to further explore the utility of substrate channeling, we performed a series of pairwise alignments with the raw ColabFold predicted output. These results, shown in Figure 4.4B, indicate a rotation of the $\mathrm{PahZ1_{KT\text{-}1}}$-IM domain relative to the $\mathrm{PahZ2_{KT\text{-}1}}$ catalytic domain. Alignment of residues comprising the dimerization domains confirm no disturbance of the conserved motif itself is observed. Interestingly, this domain rotation seems to more closely orient the two active sites, further supporting the possibility to channel substrate from one catalytic center to the next. Further studies will be needed, however, to more completely ascertain such functionality, including comprehensive in vitro characterization of recombinant enzyme activity.

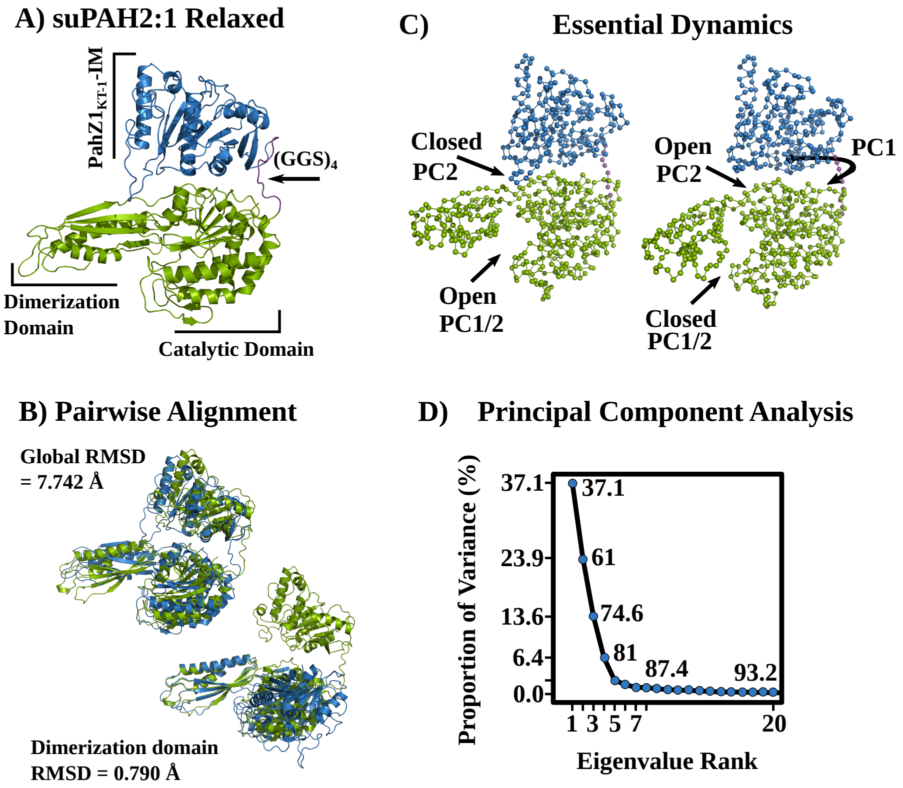


Figure 4.4: Molecular dynamics suPAH2:1 refinement. **A:** Cartoon representation of suPAH2:1 monomers following 200 ns of simulation. Individual domains are colorized and labeled such that the PahZ1$_{\text{KT-1}}$-IM and PahZ2$_{\text{KT-1}}$ domains are blue and green, respectively. **B:** The result of pairwise alignments of the refined structure against the raw Collabfold output, shown here in green. The top and bottom panels present alignments to all atoms and the dimerization domains, respectively. **C-D:** A summary of the protein's essential dynamics as revealed by principal component analysis.

The next question we sought to address was whether our modelled complex is capable of demonstrating catalytically relevant domain motions. For example, we have previously characterized PahZ2$_{\text{KT-1}}$ conformational changes wherein the Zn^{2+}-binding catalytic domain is observed to separate from the dimerization domain along a hinge region, facilitating a mechanism for substrate binding and release.[16] If our suPAH2:1 design is to be similarly active, then these domain motions must be conserved. To investigate, we subjected MD data sets to PCA, a dimensional reduction strategy often used to elucidate domain motions that are statistically likely to be involved in enzyme catalysis.

Our results reveal two principal components corresponding to motions that together account for 61% of the total system variance. A summary of the PCA is shown in Figure 4.4C-D. The $\text{PahZ1}_{\text{KT-1}}$-IM domain is observed to periodically lift from the $\text{PahZ2}_{\text{KT-1}}$ complex and subsequently re-seat, with the $(\text{GGS})_4$ linker sequence serving as a hinge region. Conversely, the $\text{PahZ2}_{\text{KT-1}}$ catalytic domain undergoes a similar opening event in agreement with results described in our earlier characterization.[16] It is important to highlight here that these motions appear to be inversely correlated, such that the $\text{PahZ1}_{\text{KT-1}}$-IM domain is opened as the $\text{PahZ2}_{\text{KT-1}}$ catalytic domain is closed and vice versa. To summarize, these components suggest a model by which suPAH2:1 may be able to sequentially catalyze the conversion of PAA and OAA at both catalytic centers. First, we expect PAA binding at the $\text{PahZ1}_{\text{KT-1}}$-IM active site to stabilize the open conformation for that domain. The release of OAA may then trigger the opening of the $\text{PahZ2}_{\text{KT-1}}$ domain and thus its binding and subsequent cleavage.

Predicting the suPAH2:1 oligomeric complex: Thus far in our modelling approach, we have focused on describing and refining monomeric subunits of suPAH2:1. In the following sections, we will attempt to elucidate a possible multimeric assembly. Given the inclusion of a PahZ1$_{\text{KT-1}}$ domain engineered to negate oligomerization along its dimeric interface, we expect suPAH2:1 monomers to instead associate along the conserved PahZ2$_{\text{KT-1}}$ dimerization domains. To model this, we subjected equilibrated suPAH2:1 monomers to protein-protein docking procedures via ClusPro 2.0, a powerful approach pairing rigid body docking with fast-Fourier transform (FFT) for evaluation of energy expressions.[39–42] This method allows for accurate protein-protein docking without any prior information regarding the complex. The results lend support for our prediction, with the most populated clusters revealing associations along the conserved dimerization domain.

Representative models of the 9 most populated clusters are shown in Figure 4.5A and are colored according to secondary structure. A comprehensive list of the top 9 clusters along with the associated membership numbers and interaction energies are shown in table S1. Clusters 0 and 1 account for 144 and 138 members respectively, and are shown in Figure 4.5B to demonstrate high structural conservation with the PahZ2$_{\text{KT-1}}$ crystal. This finding is evidenced by pairwise alignments to the dimerization domain yielding an RMSD of 0.707 and 0.784 Å respectively for model 00 and 01. A closer analysis of the predicted interface highlights the elements of secondary structure observed to define the association. As is the case with PahZ2$_{\text{KT-1}}$, Figure 4.5C reveals the interface of model 00 is comprised of interactions along both α-helices and β-strands in an anti-parallel arrangement. To further understand the physical characteristics that

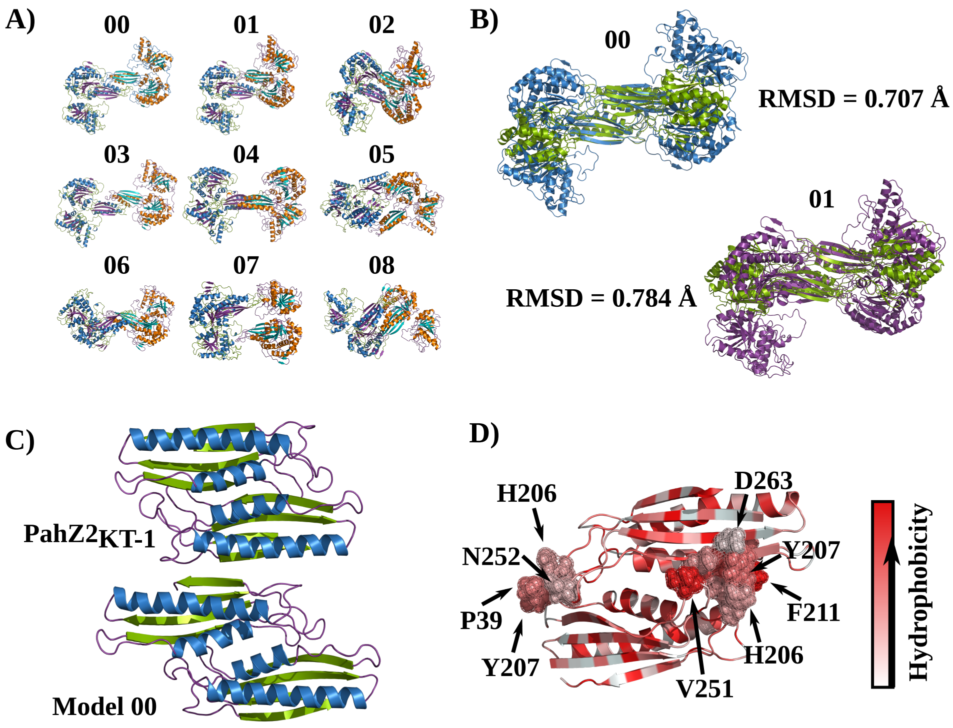


Figure 4.5: Evaluation of suPAH2:1 dimeric interface. **A:** Results of docking suPAH2:1 monomers into a dimeric assembly. The 9 structures shown are colorized according to secondary structure and ranked by cluster membership such that model 00 is the most populated arrangement. **B:** The results of pairwise alignments of models 00 and 01 (shown here in blue and violet, respectively) to the dimzerization domain of PahZ2$_{\text{KT-1}}$, shown here in green. **C:** Close-up perspective of the dimzerization interface of model 00 as compared to that of PahZ2$_{\text{KT-1}}$. Colors used to differentiate elements of secondary structure. **D:** Hotspot prediction results. Designated hotspot residues are labeled and shown as spheres with a mesh surface representation. A normalized scale of hydrophobicity is used to color all residues along a gradient, such that polar and nonpolar character is reflected by white and red, respectively.

might drive dimerization, we briefly minimized the suPAH2:1 model dimer 00 in Gromacs and subsequently performed a search for protein-protein interaction (PPI) hotspots along the interface. PPI hotspots are amino acid residues found to contribute to the energetic stability of a protein-protein interface, and are becoming a popular target for computational drug discovery.[43–45] For our purposes here, the identification of PPI hotspot residues was expected to yield insights into the stabilizing factors of the dimeric assembly. The results, shown in Figure 4.5D, suggest a strong influence of the disordered loops on either end of the dimerization domain to complex formation. This observation is further supported by the findings of others, as so-called "hot loops" are known to mediate a majority of PPIs.[46] As with PahZ2$_{\text{KT-1}}$, we additionally report the suPAH2:1 interface to be occupied with strong hydrophobic character. This is reflected in Figure 4.5D, wherein increasing hydrophobicity is represented by a gradient from white to red, such that the deepest red indicates the most hydrophobic quality. The coloration used here is based on a normalized hydrophobicity scale determined by Eisenberg and co-workers.[47] Taken together, these data appear to suggest a viable model for suPAH2:1 dimer complex formation. Further, our modelling-derived predictions are in good agreement with experimental observations of similar peptidases, including of course PahZ2$_{\text{KT-1}}$.

4.4 Methods

4.4.1 Deep learning-based structure prediction

PahZ1$_{\text{KT-1}}$ homologue predictions: Protein sequences of interest were identified with the basic local alignment search tool (Blast)[18] courtesy of the Universal Protein Resource (Uniprot).[21] Sequences of homologues expressing greater than 70% identity relative to PahZ1$_{\text{KT-1}}$ were downloaded and saved as FASTA formatted files. Each FASTA sequence was subsequently uploaded to the Robetta web server for deep learning-based structural determination via RosettaFold.[19] The resulting models were ranked according to mean pLDDT scores on a 0-1 scale, with those demonstrating the highest confidence reported here. All rendered images and electrostatic surface maps were generated in Pymol.[48] Pymol was also used for pairwise alignment of homologues via the align function to the PahZ1$_{\text{KT-1}}$ crystal structure and subsequent RMSD determination.

suPAH construct predictions: Primary sequences for suPAH1:2 and 2:1 were constructed as FASTA formatted files in accordance with the designed schemes presented in Figure 4.2C-D. They then served as input data for structural determination via the Alphafold2[31] functionality of ColabFold.[30] Template search and multiple sequence alignments were carried out with HHsuite[49] and MMseqs2,[50–53] respectively. All predicted models were subsequently energy minimized in Amber[54] and ranked according to mean pLDDT on a 0-100 scale. The highest ranks for each construct are reported here and used for subsequent modelling. All representations, rendering, and pairwise alignments were carried out in Pymol[48] as previously described.

4.4.2 Model refinement and dimer assemblies

suPAh2:1 monomer simulations: The highest ranked suPAH2:1 prediction was subject to structural relaxation via brief unbiased all-atom molecular dynamics (MD) in Gromacs.[34–37] Gromacs topologies were first generated with Charmm36[38] parameter sets. The model was then solvated in a dodecahedron with TIP3P water and with a minimum periodic distance of 1.0 nm. Neutralizing ions were inserted and the system was minimized by steepest descent until convergence of forces could be observed. Thermodynamic equilibration was carried out in the presence of heavy atom restraints in accordance with the isothermal-isobaric ensemble (NPT). The v-rescale thermostat and Berendsen barostat[55] were used to bring the system to 298.15 K and 1.0 bar, respectively. Restraints were then removed and production steps were initiated with temperature and pressure now maintained with Nose-Hoover[56,57] and Parinello-Rahman[58] temperature and pressure coupling. Long range electrostatics were handled via Particle Mesh Ewald (PME) summation with verlet cutoff scheme.[59] Production MD was allowed to proceed for 200 ns prior to extraction of the relaxed suPAH2:1 monomeric structure.

Principal component analysis: The refinement 200 ns trajectory described above was loaded into visual molecular dynamics (VMD)[60] for confirmation of trajectory integrity and subsequently converted to the DCD file format. These files were necessary for compatibility with Bio3d,[61] a versatile R[62] package for analyzing both structural and MD data sets, including principal component analysis (PCA). Protein coordinates throughout the trajectory were prepared for PCA by superposition to selected C_α atoms. The resulting ranks were then plotted as a function of their contribution to the conformational variance. All structures illustrating simulation results were prepared and rendered with Pymol.[48]

Evaluation of the dimeric interface: The MD-refined suPAH2:1 monomer was extracted from the simulation trajectory described in the previous section and used as input for the ClusPro2 protein-protein docking server.[39–42] Briefly, ClusPro2 utilizes a rigid-body docking approach wherein fast Fourier transform (FFT) is used for energy evaluation, allowing for the sampling of billions of docked conformations. The docking component of the algorithm, named PIPER, places one subunit at the origin of a coordinate grid, while the other is placed at various translational and rotational positions along the grid. Interaction energies are then evaluated according the equation,[39–42]

$$E = {}_{W1}E_{rep} +_{W2} E_{attr} +_{W3} E_{elec} +_{W4} E_{DARS} \tag{4.1}$$

where E_{rep} and E_{attr} account for van der Waals repulsion and attraction energies. E_{elec} defines electrostatic energies and E_{DARS} represents the free energy change associated with desolvation at the interface. Each w coefficient represents the relative weights of each energy term. Finally, the 1000 docked configurations with the lowest energy are extracted and clustered with a distance cutoff of 9 Å. In this work, the top 9 models were saved for subsequent analysis and are presented in Figure 4.5A. PPI hotspot residues were subsequently identified by the KFC2 hotspot prediction server[63,64] courtesy of Oak Ridge National Laboratory. The algorithm analyzes each residue comprising a given protein-protein interface and predicts hotspot classification trained on various chemical and physical features. Residues determined to be interfacing hotspots were then visualized in Pymol[48] in the context of the associated dimer and colorized according to the normalized hydrophobicity scale.[47]

4.5 Conclusion

Deep-learning affords time and cost-effective new strategies to resolve structural details of novel protein sequences, making it an indispensable tool for protein engineers. Here, we have utilized it to aid in the rational design of a chimeric PAA-hydrolase. PahZ1$_{\text{KT-1}}$ homologous structures predicted with high confidence by RosettaFold initially reveal an evolutionary basis for the inherent modularity of the PahZ1$_{\text{KT-1}}$ monomeric subunit. This finding thus provided the groundwork for designing a new primary sequence by which the PahZ1$_{\text{KT-1}}$ domain is joined to either the N-terminus (suPAH1:2) or C-terminus (suPAH2:1) of PahZ2$_{\text{KT-1}}$. Prerequisite structure predictions using ColabFold revealed the latter arrangement appears more amenable to the channeling of bound PAA from one catalytic center to the next and thus a more efficient design.

The conformational stability of the predicted structure was assessed by MD simulations, allowing for further refinement of the model as well as a comprehensive dynamics analysis via PCA. The results reveal the most variant motions correspond to inversely correlated and open/close events of both the PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ catalytic domains. The observed motions are expected, and in the case of the PahZ2$_{\text{KT-1}}$ catalytic domain are experimentally known to be a requisite for catalytic function. Subsequent protein-protein docking studies suggest suPAH2:1 monomers to associate into dimers in a manner consistent with the PahZ2$_{\text{KT-1}}$ crystal structure. Hotspot prediction further determines complex formation is largely dependent on the presence of disordered loops adjacent to the dimerization domains of each subunit. Overall, the data presented here serve as the first steps to engineer a novel enzyme for the timely and affordable biodegradation of PAA. The extensive use of in silico modelling strategies are known to decrease both the time and cost of such engineering efforts, and have thus become an inseparable part of the protein design pipeline.

References

(1) Huppertsberg, S.; Zahn, D.; Pauelsen, F.; Reemtsma, T.; Knepper, T. Making waves: Water-soluble polymers in the aquatic environment: An overlooked class of synthetic polymers? *Water Research* **2020**, *181*, 115931.

(2) Hennecke, D.; Bauer, A.; Herrchen, M.; Wischerhoff, E.; Gores, F. Cationic polyacrylamide copolymers (PAMs): environmental half life determination in sludge-treated soil. *Environ Sci Eur* **2018**, *30*, 16.

(3) Jop, K.; Guiney, P.; Christensen, K.; Silberhorn, E. Environmental fate assessment of two synthetic polycarboxylate polymers. *Ecotoxicol Environ Saf* **1997**, *37*, 229–37.

(4) Bai, M.; Wilske, B.; Buegger, F.; Esperschütz, J.; Bach, M.; Frede, H.-G.; Breuer, L. Relevance of nonfunctional linear polyacrylic acid for the biodegradation of superabsorbent polymer in soils. *Environmental Science & Pollution Research* **2015**, *22*, 5444–5452.

(5) Wilske, B.; Bai, M.; Lindenstruth, B.; Bach, M.; Rezaie, Z.; Frede, H.-G.; Breuer, L. Biodegradability of a polyacrylate superabsorbent in agricultural soil. *Environmental Science & Pollution Research* **2014**, *21*, 9453–9460.

(6) Arp, H.; Knutsen, H. Could We Spare a Moment of the Spotlight for Persistent, Water-Soluble Polymers? *Environmental Science & Technology* **2020**, *54*, 3–5.

(7) Buczek, S.; Cope, W.; McLaughlin, R.; Kwak, T. Acute toxicity of polyacrylamide flocculants to early life stages of freshwater mussels. *Environ Toxicol Chem* **2017**, *36*, 2715–2721.

(8) Xiong, B.; Loss, R.; Shields, D.; Pawlik, T.; Hochreiter, R.; Zydney, A.; Kumar, M. Polyacrylamide degradation and its implications in environmental systems. *npj Clean Water* **2018**, *1*, DOI: `10.1038/s41545-018-0016-8`.

(9) Thombre, S.; Sarwade, B. Synthesis and Biodegradability of Polyaspartic Acid: A Critical Review. *Journal of Macromolecular Science, Part A* **2005**, *42*, 1299–1315.

(10) Adelnia, H.; Blakey, I.; Little, P. J.; Ta, H. T. Hydrogels Based on Poly(aspartic acid): Synthesis and Applications. *Frontiers in Chemistry* **2019**, *7*, 755.

(11) Pramanik, B. K.; Gao, Y.; Fan, L.; Roddick, F. A.; Liu, Z. Antiscaling effect of polyaspartic acid and its derivative for RO membranes used for saline wastewater and brackish water desalination. *Desalination* **2017**, *404*, 224–229.

(12) Mei HU, M.; hui DOU, Q.; min CUI, X.; hong LOU, Y.; ping ZHUGE, Y. Polyaspartic acid mediates the absorption and translocation of mineral elements in tomato seedlings under combined copper and cadmium stress. *Journal of Integrative Agriculture* **2019**, *18*, 1130–1137.

(13) Juriga, D.; Nagy, K.; Jedlovszky-Hajdú, A.; Perczel-Kovách, K.; Chen, Y. M.; Varga, G.; Zrínyi, M. Biodegradation and Osteosarcoma Cell Cultivation on Poly(aspartic acid) Based Hydrogels. *ACS Applied Materials & Interfaces* **2016**, *8*, PMID: 27541725, 23463–23476.

(14) Tabata, K.; Abe, H.; Doi, Y. Microbial Degradation of Poly(aspartic acid) by Two Isolated Strains of Pedobacter sp. and Sphingomonas sp. *Biomacromolecules* **2000**, *1*, 157–161.

(15) Tabata, K.; Kasuya, K.; Abe, H.; Masuda, K.; Doi, Y. Poly(aspartic acid) degradation by a Sphingomonas sp. isolated from freshwater. *Appl Environ Microbiol* **1999**, *65*, 4268–4270.

(16) Brambley, C. A.; Yared, T. J.; Gonzalez, M.; Jansch, A. L.; Wallen, J. R.; Weiland, M. H.; Miller, J. M. Sphingomonas sp. KT-1 PahZ2 Structure Reveals a Role for Conformational Dynamics in Peptide Bond Hydrolysis. *The Journal of Physical Chemistry B* **2021**, *125*, PMID: 34060838, 5722–5739.

(17) Brambley, C. A.; Bolay, A. L.; Salvo, H.; Jansch, A. L.; Yared, T. J.; Miller, J. M.; Wallen, J. R.; Weiland, M. H. Structural Characterization of Sphingomonas sp. KT-1 PahZ1-Catalyzed Biodegradation of Thermally Synthesized Poly (aspartic acid). *ACS Sustainable Chemistry & Engineering* **2020**, *8*, 10702–10713.

(18) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **1990**, *215*, 403–410.

(19) Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **2021**, *373*, 871–876.

(20) Hiranuma, N.; Park, H.; Baek, M.; Anishchenko, I.; Dauparas, J.; Baker, D. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature communications* **2021**, *12*, 1340.

(21) Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **2020**, *49*, D480–D489.

(22) Urushibata, Y.; Tokuyama, S.; Tahara, Y. Characterization of the Bacillus subtilis ywsC gene, involved in gamma-polyglutamic acid production. *Journal of bacteriology* **2002**, *184*, 337—343.

(23) Hsueh, Y.-H.; Huang, K.-Y.; Kunene, S. C.; Lee, T.-Y. Poly-γ-glutamic Acid Synthesis, Gene Regulation, Phylogenetic Relationships, and Role in Fermentation. *International journal of molecular sciences* **2017**, *18*, DOI: `10.3390/ijms18122644`.

(24) Sharma, S.; Bhatnagar, R.; Gaur, D. Bacillus anthracis Poly-γ-D-Glutamate Capsule Inhibits Opsonic Phagocytosis by Impeding Complement Activation. *Frontiers in immunology* **2020**, *11*, 462.

(25) Chateau, A.; Van der Verren, S. E.; Remaut, H.; Fioravanti, A. The Bacillus anthracis Cell Envelope: Composition, Physiological Role, and Clinical Relevance. *Microorganisms* **2020**, *8*, DOI: `10.3390/microorganisms8121864`.

(26) Zorigt, T.; Furuta, Y.; Simbotwe, M.; Ochi, A.; Tsujinouchi, M.; Shawa, M.; Shimizu, T.; Isoda, N.; Enkhtuya, J.; Higashi, H. Development of ELISA based on Bacillus anthracis capsule biosynthesis protein CapA for naturally acquired antibodies against anthrax. *PloS one* **2021**, *16*, e0258317.

(27) Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **2021**, *50*, D439–D444.

(28) Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590—596.

(29) Dohmen, E.; Klasberg, S.; Bornberg-Bauer, E.; Perrey, S.; Kemena, C. The modular nature of protein evolution: domain rearrangement rates across eukaryotic life. *BMC evolutionary biology* **2020**, *20*, 1–13.

(30) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold - Making protein folding accessible to all. *bioRxiv* **2022**, DOI: `10.1101/2021.08.15.456425`.

(31) Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590—596.

(32) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **2013**, *29*, 2722–2728.

(33) Kummer, M. J.; Lee, Y. S.; Yuan, M.; Alkotaini, B.; Zhao, J.; Blumenthal, E.; Minteer, S. D. Substrate Channeling by a Rationally Designed Fusion Protein in a Biocatalytic Cascade. *JACS Au* **2021**, *1*, PMID: 34467357, 1187–1197.

(34) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, PMID: 26620784, 435–447.

(35) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In *Solving Software Challenges for Exascale*, ed. by Markidis, S.; Laure, E., Springer International Publishing: Cham, 2015, pp 3–27.

(36) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl,

E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(37) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(38) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **2017**, *14*, 71—73.

(39) Desta, I. T.; Porter, K. A.; Xia, B.; Kozakov, D.; Vajda, S. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure*, *28*, DOI: `10.1016/j.str.2020.06.006`.

(40) Vajda, S.; Yueh, C.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Kozakov, D. New additions to the ClusPro server motivated by CAPRI. *Proteins* **2017**, *85*, 435—444.

(41) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro web server for protein–protein docking. *Nature protocols* **2017**, *12*, 255–278.

(42) Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Vajda, S. How good is automated protein docking? *Proteins* **2013**, *81*, 2159—2166.

(43) Ghanakota, P.; van Vlijmen, H.; Sherman, W.; Beuming, T. Large-Scale Validation of Mixed-Solvent Simulations to Assess Hotspots at Protein–Protein Interaction Interfaces. *Journal of Chemical Information and Modeling* **2018**, *58*, PMID: 29617116, 784–793.

(44) Cukuroglu, E.; Engin, H. B.; Gursoy, A.; Keskin, O. Hot spots in protein–protein interfaces: Towards drug discovery. *Progress in Biophysics and Molecular Biology* **2014**, *116*, 165–173.

(45) Bakail, M.; Ochsenbein, F. Targeting protein–protein interactions, a wide open field for drug design. *Comptes Rendus Chimie* **2016**, *19*, Emerging Chemistry in France, 19–27.

(46) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nature chemical biology* **2014**, *10*, 716—722.

(47) Eisenberg, D.; Schwarz, E.; Komaromy, M.; Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology* **1984**, *179*, 125–142.

(48) Schrödinger, LLC The PyMOL Molecular Graphics System, Version 1.8, 2015.

(49) Steinegger, M.; Meier, M.; Mirdita, M.; Vöhringer, H.; Haunsberger, S.; Söding, J. HH-suite3 for fast remote homology detection and deep protein annotation, 2019.

(50) Steinegger, M.; Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* **2017**, *35*, 1026—1028.

(51) Steinegger, M.; Söding, J. Clustering huge protein sequence sets in linear time. *Nature communications* **2018**, *9*, 2542.

(52) Mirdita, M.; Steinegger, M.; Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **2019**, *35*, 2856–2858.

(53) Mirdita, M; Steinegger, M; Breitwieser, F; Söding, J; Levy Karin, E Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* **2021**, *37*, 3029–3031.

(54) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **2013**, *3*, 198–210.

(55) Berendsen, H.; Postma, J.; van Gunsteren, W.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.

(56) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(57) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(58) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.

(59) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(60) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(61) Grant, B. J.; Rodrigues, A. P. C.; Elsawy, K. M.; Mccammon, J. A.; S, L.; Caves, D. Bio3d: An R Package.

(62) R Development Core Team R: A Language and Environment for Statistical Computing, ISBN 3-900051-07-0; R Foundation for Statistical Computing, Vienna, Austria, 2011.

(63) Darnell, S. J.; Page, D.; Mitchell, J. C. An automated decision-tree approach to predicting protein interaction hot spots. *Proteins: Structure, Function, and Bioinformatics* **2007**, *68*, 813–823.

(64) Zhu, X.; Mitchell, J. C. KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79*, 2671–2683.

Chapter 5

Conclusion

5.1 Introduction

The following chapter serves to conclude the current investigation of two poly(aspartic acid) hydrolases (PahZ1 and PahZ2) from *Spingomonas* sp. KT-1. We will summarize key findings relative to our research aims along with our primary contributions to the field at large. This includes some discussion regarding how our structural characterizations and molecular modelling may impact the widespread commercial use of polycarboxylates and other polymers in favor of more ecologically inert poly(aspartic acid) (PAA). It is hoped that by improving the structural and dynamic understanding of these hydrolases, this work will contribute to the development of a cradle-to-cradle system of biodegradation wherein recoverable monomeric units of aspartic acid may be generated from thermally synthesized PAA (tPAA). Finally, we will address here the major limitations of the present study as well as opportunities for future work.

5.2 Summary of findings

5.2.1 PahZ1 from *Sphingomonas* sp. 1

PahZ1$_{\text{KT-1}}$ is a dimeric serine protease and poly(aspartic acid)(PAA) hydrolase isolated from a strain of river water bacteria, *Sphingomonas* sp. KT-1. PahZ1$_{\text{KT-1}}$ was found to participate in the conversion of tPAA to shorter chains of oligo (aspartic acid) (OAA), a prerequisite step in the complete biodegradation of tPAA into recoverable units of aspartic acid monomers.[1–3] The step is imperfect, however, as PahZ1$_{\text{KT-1}}$ was found to exhibit a selective preference for β-amide linkages over α linked chains.[3] This finding is problematic for the use of PahZ1$_{\text{KT-1}}$ in any catalytic pipeline designed around the complete biodegradation of tPAA. To understand more about the structural basis for this selectivity, as well as to begin probing for ways to engineer a solution, we resolved the x-ray crystal structure of PahZ1$_{\text{KT-1}}$.[4]

Structural characterization of PahZ1$_{\text{KT-1}}$: The structure of PahZ1$_{\text{KT-1}}$ was solved to a resolution of 2.45 Å and reveals four monomers in solution. We determined the biological assembly to consist of homodimers, a finding later supported by SAXS analysis (Figure 2.3). Monomers are observed to associate along interface residues comprising helical secondary structures, with catalytic triad active sites oriented on opposing surfaces of the dimer in an anti-parallel arrangement. Structural data indicate the constituents of the triad are residues S157, D184, and H280. Further insights regarding the association of tPAA to these triad residues was gained by the identification of putative binding pockets characterized by clefts of concentrated positive charge, a finding enabled by maps of electrostatic surface potential. Such an observation led to the hypothesis that PahZ1$_{\text{KT-1}}$ association with the polyanionic tPAA is governed largely by electrostatic interactions, a question that would later be assessed both by molecular modelling and mutational activity assays (see below).

Association of monomers into dimers, however, was found to be driven primarily by non-polar forces, as reflected in Figure 2.4. This was first evidenced directly by the structural data indicating the helical interface was largely populated by non-polar residues. An NaCl-dependent size-exclusion-chromatography (SEC) analysis provided further support, revealing no change in the apparent molecular weight of the complex as a function of increasing NaCl concentration. We then began to determine the contribution of different interface residues to the solvation free energy of complex formation (ΔG^i) by pairing molecular dynamics (MD) with a Protein Data Bank in Europe proteins, interfaces, structures, and assemblies (PDBePISA) analysis. Briefly, a series of mutations were introduced at positions 251 and 255, including E251A, E251K, E251R, A255E, and A255W. Each mutant construct was subjected to 200 ns of simulation time, from which ten frames were extracted and used for subsequent PDBePISA analysis. The results, shown in Figure 2.4E, reveal no change in ΔG^i for all E251 mutations relative to wild type. In contrast, A255E and A255W mutants are observed to destabilize the dimeric complex, as reflected by $\Delta\Delta G^i$ = 6.3 and 5.7 kcal mol^{-1}, respectively. All things considered, these data suggest the PahZ1$_{\text{KT-1}}$ dimeric assembly is largely stabilized by the presence of non-polar interface residues.

Assessing PahZ1$_{\text{KT-1}}$ substrate selectivity: As previously described, our initial structural characterization of PahZ1$_{\text{KT-1}}$ provided clues as to how substrate association is carried out. The active site triad-adjacent pocket of positive charge thus became the target of a subsequent tPAA docking and simulation strategy. The rationale for these simulations was two-fold: (1) elucidate modes of tPAA binding and (2) attempt to identify explanations for the β selectivity exhibited by PahZ1$_{\text{KT-1}}$. The first of these aims was addressed by the aforementioned docking of both α- and β-linked tPAA ligands to search coordinates corresponding to the positively charged binding pocket. Binding poses yielding the highest docking scores are shown in Figure 2.6A. To establish whether the obtained poses represent stable interactions, we subjected protein-ligand complexes to 200 ns of simulation time in Gromacs.[5–8] As can be seen in representative frames in Figure 2.6C, the trajectories revealed interactions between substrate and residue R246 across all simulation conditions. The subsequent quantification of these interactions via hydrogen bonding frequency analysis reports values of 1.2 ± 0.8 and 0.3 ± 0.5 hydrogen bonds between R246 and α- and β-PAA, respectively. These results led us to hypothesize that R246 plays a role in substrate recognition and presentation to triad residues.

To test this hypothesis, we expressed and purified R246 and other mutant constructs for use in gel permeation chromatography (GPC) activity assays reporting on OAA formation. A summary of R246A, R246K, R246E, D225A, and S157A activities are reported as percentages of wild type activity in Figure 2.7D. All mutants tested revealed diminished hydrolase activity to less than 12% after

48 hours of incubation. This was expected in the case of S157A, however, the reduced activity for the D225, a noted residue involved in the catalytic triad, was perhaps not as diminished as expected. Further, the reported effect of R246 mutants, including the complete ablation of PahZ1$_{\text{KT-1}}$ activity by R246E, strongly suggests a role for the residue in tPAA binding and/or active site orientation. To infer any mechanistic atomic-level insights related to these results, we performed additional MD simulations on R246A, R246K, and D225A. The results, reported in Figure 2.8, can best be summarized by describing each mutant as inducing various rearrangements of the catalytic triad. Particularly, the residual activity of D225A may be explained by the presence of R246 itself, which is observed to interact with β-PAA such that it may be positioned as close as 3.53 ± 0.02 Å from S157. For all other R246 mutants, β-PAA was observed to be effectively excluded from active site access. Taken together, these results suggest a model by which R246 is instrumental to PahZ1$_{\text{KT-1}}$ binding of tPAA, as well as to tPAA positioning within the active site.

5.2.2 PahZ2 from *Sphingomonas* sp. 1

Though both isolated from *Sphingomonas* sp. KT-1, PahZ2$_{\text{KT-1}}$ is unlike PahZ1$_{\text{KT-1}}$ in both structure and catalytic mechanism. While the latter is largely a canonical serine protease, PahZ2$_{\text{KT-1}}$ functions as an M28 family zinc-dependent exoopeptidase in order to cleave OAA into smaller units of aspartic acid chains. It is in this way that the two enzymes work together to accomplish the previously described cradle-to-cradle pathway of tPAA biodegradation. Further distinguishing it from PahZ1$_{\text{KT-1}}$, PahZ2$_{\text{KT-1}}$ demonstrates no selective preference of either α- or β-amide linkages.[2,3,9,10] However, its activity leaves much room for improvement in that it fails to achieve the complete conversion of OAA into monomeric units of aspartic acid. To that end, we set out to characterize the structure and essential dynamics of PahZ2$_{\text{KT-1}}$ to better enable the possibility for future engineering efforts toward a more effective enzyme.

PahZ2$_{\text{KT-1}}$ structural characterization: PahZ2$_{\text{KT-1}}$ crystals were grown both in the presence of Zn^{2+} and Gd^{3+}/Sm^{3+} and diffracted to 2.50 Å and 1.85 Å, respectively. The results revealed two monomers in the asymmetric unit, however, the dimeric complex shown in Figure 3.1A is the predicted biological assembly. This determination was based on structural conservation with analogous metallopeptidases, including carboxypeptidase G2 and DapE (Figure 3.1D-E). Each monomeric subunit is comprised of a metal binding catalytic domain attached by a flexible hinge region to a conserved dimerization domain, as illustrated in Figure 3.1B. This domain forms protein-protein interactions along an α helix and single β strand, each associating with the partner chain in an anti-parallel orientation as illustrated in Figure 3.2A. Complex formation appears to be governed by interactions between a mixture of polar and non-polar interface residues. To further elucidate, and similar to our approach with PahZ1$_{\text{KT-1}}$, we sought to characterize the nature of complex formation via a PDBePISA analysis. The findings support an interface driven by hydrophobic interactions, as evidence by a strongly negative ΔG^{i} of -21.8 kcal mol^{-1} (Figure 3.2C). This conclusion was further supported by supplemental MD simulations whereby interface mutant constructs were equilibrated for 100 ns prior to additional PDBePISA calculations. The results indicate mutations made at hydrophobic residue positions were most likely to destabilize the complex (Figure 3.S4).

Zn^{2+} binding active site analysis: Our structural data further confirmed $PahZ2_{KT\text{-}1}$ classification as an M28 metallopeptidase by the inclusion of two Zn^{2+} ions, designated Zn_I and Zn_{II}, bound to unique sites of each catalytic domain. Each zinc is observed to demonstrate a tetrahedral geometry, wherein E155, D184, H94, D121, E156, and two waters operate as metal-coordinating residues. A summary of the active site arrangement was shown in Figure 3.3. Other metals were subsequently found to support $PahZ2_{KT\text{-}1}$ function, with those preferring to adopt tetrahedral coordination geometries demonstrating the highest percent activity relative to Zn^{2+} bound enzyme. These data strongly suggest the rearrangement of metal coordinating residues may influence catalytic activity by $PahZ2_{KT\text{-}1}$.

This was not unexpected given the high sequence conservation observed for active site amino acids for M28 family members. This includes the presence of the Glu-Glu motif that was identified among all $PahZ2_{KT\text{-}1}$ structural homologues investigated. The Glu-Glu motif in $PahZ2_{KT\text{-}1}$ is thus believed to indicate a similar hydrolysis mechanism as has been suggested for DapE.[11,12] That is, an acidic side chain (E155 in the case of $PahZ2_{KT\text{-}1}$) is positioned such that it may activate a nearby water to serve in the nucleophilic attack on substrate. This proposed mechanism was illustrated in Figure S7D. The subsequent analysis of $PahZ2_{KT\text{-}1}$ activities as a function of active site mutation revealed that all metal-coordinating residues are catalytic. However, mutations of Zn_{II} coordinating residues resulted in more significant reductions of activity relative to wild type enzyme as shown in Figure 3.3E. Taken together, these data allow for the determination that the Zn_{I} and Zn_{II} sites offer unique contributions to the $\mathrm{PahZ2}_{\text{KT-1}}$ function, although more work is yet needed to further elucidate the molecular details not resolved by these experiments.

Active site dynamics: Structural comparison of $\mathrm{PahZ2}_{\text{KT-1}}$ with DapE suggests E155 serves as a general acid/base in a catalytic mechanism requiring the activation of a water molecule. Considering no active site water molecules were resolved in the Zn^{2+} $\mathrm{PahZ2}_{\text{KT-1}}$ structure, we performed a series of MD simulations on the solvated construct to further investigate. What we found was the occasional rearrangement of active site residues into two primary coordination geometries. One of these was consistent with structural data wherein Zn_{I} is observed to interact with E155, H94, D184, and D121. For the other, H94 undergoes a rotamer shift such that D122 is able to take its place in metal-coordination. We suggest this represents an increase in complex stability, as replacement of an imidazole group with a carboxylate in a Zn^{2+} tetrahedral coordination sphere has been shown to demonstrate a more negative change in the free energy of isomerization from ΔG = -8.4 to -9.6 kcal mol^{-1} relative to octahedrally coordinated zinc atoms.[13]

Pertinent to the purpose of performing the MD, our simulations captured a water molecule interaction that was observed to bridge E155 with the Zn_{II} site, as seen in Figure 3.4A. This finding is consistent with the proposed DapE mechanism of water activation. These data led to a working model wherein the purpose of Zn_{II} is to polarize a water molecule, allowing for E155 to facilitate proton removal. It is likely that in this scenario, Zn_{I} is tasked with positioning both E155 and substrate to a configuration optimal for hydrolysis. The MD further supports this notion by the observation that metal-coordinating residues are largely immobile during the 200 ns trajectories, as evidenced by their distance to the nearest Zn^{2+} ion in Figure S8.

To further elucidate the respective roles of each zinc site, we subsequently performed umbrella sampling of steered simulations wherein each zinc ion was removed from the active site by a biasing harmonic potential. The resulting trajectories were used for the calculation of potential of mean force (PMF) for each process, yielding changes in free energies for Zn_{II} and Zn_{I} equal to 42 $\pm$ 4 kJ mol^{-1} and 62 $\pm$ 2 kJ mol^{-1} respectively. Thus we conclude that each zinc ion binds to PahZ2$_{\text{KT-1}}$ with unique affinities. The observation of high- and intermediate-affinity zinc binding is not without precedent, as other such relationships have been previously described[14–17] wherein asymmetric binding allows for regulation of enzyme activity within the cell. For example, we propose that PahZ2$_{\text{KT-1}}$ may demonstrate increased catalytic output when *Sphingomonas* sp. KT-1 is exposed to locally high concentrations of zinc-chelated PAA in river water.

PahZ2$_{\text{KT-1}}$ NaCl-dependent activity and dynamics: To address a gap in the experimental conditions in which PahZ2$_{\text{KT-1}}$ activity has traditionally been investigated, we conducted GPC activity assays in a range of NaCl concentrations.

Briefly, product formation was tracked as a function of time for 0, 150, and 500 mM NaCl. These data allow for the determination of initial velocities, as shown in Figure 3.6, revealing an increase in percent product per minute with increasing NaCl concentration. Thus, PahZ2$_{\text{KT-1}}$ demonstrates a linear stimulation of catalytic function with the inclusion of NaCl. To develop a structural understanding of the underlying mechanism, as well as to predict putative OAA binding sites, we again turned to an MD based modelling approach. We began by docking an α-penta-PAA (PAA$_5$) substrate to the surface comprising the interface between the PahZ2$_{\text{KT-1}}$ catalytic and dimerization domain. We then simulated the construct in both the presence and absence of 500 mM NaCl for no less than 200 ns. Consistent with structural observations taken from the homologous DapE system, we found substrate to be stably bound along the hinge region, with the positioning driven largely by arginine residues belonging to both the catalytic and dimerization domains as seen in Figure 3.7.

These simulations further provided insights regarding the nature of NaCl-stimulated activity. Specifically, trajectories calculated in the absence of NaCl revealed a primary conformation wherein the catalytic domain remains open along the hinge. These conditions appear to disallow the substrate from being optimally positioned near the catalytic zinc atoms. The inclusion of 500 mM NaCl conversely promotes a conformation that remains largely closed, a state which is observed to yield a substrate binding pose that is closer to the relevant catalytic residues. In all conditions explored, it is the C-terminal carboxylate group of PAA$_5$ that is oriented toward the active site. This is in good agreement with existing descriptions of PahZ2$_{\text{KT-1}}$ as as exopeptidase.[9,18] Additional MD was used to elucidate the nature of these conformational differences by the extraction of free energy landscapes shown in Figure 3.8. The results revealed a wider accessibility to the lowest energy states when NaCl was included in solution. Additional dynamic light scattering (DLS) and size-exclusion chromatography (SEC) (Figure 3.8 and Figure 3.9 respectively) were used to confirm that the PahZ2$_{\text{KT-1}}$ dimeric construct remains in tact at all NaCl concentrations investigated.

Finally, the results of principal component analyses (PCA) (Figure 3.10) for a variety of enzyme conditions aided in the interpretation of these data. Briefly, PCA allows for one to extract statistically significant dynamics of a system from the assemblage of motions typical of a long MD trajectory. The results reveal principal components corresponding to catalytic domain open/close events with increased ranges of motion in the presence of NaCl in solution. However, as Figure 3.10F illustrates, it is only when Zn^{2+} and PAA_5 are also included that we observe the formation of a stable closed conformation for PC1 in the presence of 500 mM NaCl. Additional energy landscapes shown in Figure S12 suggest this may be the result of a restriction of the accessible conformational space such that the optimal catalytic configuration is favored. Overall, the data presented here express a model wherein PahZ2$_{\text{KT-1}}$ activity is extensively tunable by altering various environmental factors, including salt and metal ion availability.

5.2.3 Engineering a chimeric PahZ hydrolase:

A recurring theme of the present work has been the overall inefficiency in the catalytic mechanisms of PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$. Particularly, such complaints are usually targeting poor substrate selectivity and incomplete OAA degradation for PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$, respectively. However, we have also highlighted the inherent shortcomings of deploying a dual-enzyme system involving varied catalytic mechanisms in a commercial setting to degrade PAA or other polymers. It is a more costly and less efficient solution than a single-enzyme delivery. Thus, we have proposed to engineer a chimeric enzyme, affectionately named "super PahZ" (suPAH), wherein PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ monomers are fused into a single chain. Our first steps into the process, as described here, present a deep-learning and molecular dynamics-based modelling strategy to preemptively evaluate the feasibility of two constructed sequences.

Designing and modelling suPAH constructs: The rationale during the design stages was to directly connect PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ monomers in primary sequence in order to create a multi-domain complex theoretically capable of degrading tPAA substrates into OAA, and subsequently cleaving the OAA into monomeric units of aspartic acid. It is in this way that we propose to utilize a cradle-to-cradle synthetic process by which the end product of tPAA degradation can be used to synthesize additional starting material. As illustrated in Figure 4.2, we opted to first build suPAH fusions by simply attaching a PahZ1$_{\text{KT-1}}$ monomer via a flexible (GGS)$_4$ linker motif to either termini of PahZ2$_{\text{KT-1}}$, yielding two putative sequences. The resulting constructs were named dependent upon the placement of PahZ1$_{\text{KT-1}}$ relative to PahZ2$_{\text{KT-1}}$. That is, suPAH1:2 and suPAH2:1 when PahZ1$_{\text{KT-1}}$ is fused to the N-terminus and C-terminus of PahZ2$_{\text{KT-1}}$, respectively.

We next sought to further explore the feasibility of each design by performing a deep-learning based protein structure prediction of each primary sequence. To do this, we uploaded FASTA files for suPAH1:2 and 2:1 to the ColabFold collaborative notebook in order to obtain high confidence models via its AlphaFold2 functionality.[19,20] The results, presented in Figure 4.3, reveal largely analogous structures with comparable confidence levels. Conformationally, the suPAH2:1 domain arrangement suggests it may allow for the most efficient catalytic activity. This is evidenced by the positioning of the PahZ1$_{\text{KT-1}}$ active site relative to that of PahZ2$_{\text{KT-1}}$. This is highlighted by the transparent overlays in Figure 4.3A and C. The closer proximity of active sites is believed to offer the greatest likliehood of substrate channeling, wherein an intermediate product released from one active site can migrate to the next catalytic center in a multi-step mechanism without returning to bulk solvent.[21] We thus chose to focus on the suPAH2:1 construct for the duration of the current study. The high confidence for both models is quantified in the provided pLDDT per residue plots, wherein a global pLDDT is determined for suPAH1:2 and 2:1 of 95.1 and 95.3 respectively. Regions of localized low confidence are attributed to protein disorder. This is most notably associated here with the flexible $(GGS)_4$ linker connecting the PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ domains.

Analyzing suPAH2:1 essential dynamics and dimer assessment: Following structure prediction, we subjected the suPAH2:1 model to 200 ns of production MD in order to further refine the structure as well as to investigate catalytically relevant domain motions. The results reveal a stable association of the PahZ1$_{\text{KT-1}}$ domain with the dimerization domain along the end of a solvent accessible β strand, as shown in Figure 4.4A. This conformation places each catalytic center within even closer proximity, further supporting the notion of substrate channeling between active sites. The subsequent pairwise alignment with the pre-equilibrated model reveals no disruption in the arrangement of conserved dimerization residues, as evidenced by an RMSD = 0.790 Å in Figure 4.4B. The resulting 200 ns trajectories also allowed for the extraction of variant motions via the calculation of principal components. PC1 and PC2 are highlighted in Figure 4.4C, which reveal open and close events for the PahZ2$_{\text{KT-1}}$ catalytic and PahZ1$_{\text{KT-1}}$ domains respectively. The former is noted to be a required motion for successful catalysis by PahZ2$_{\text{KT-1}}$, as reported in our previous work.[22]

The final step of the current study was the prediction of the suPAH2:1 oligomeric assembly. Given the conservation of the included dimerization domain, along with structural data from PahZ2$_{\text{KT-1}}$ and other homologues reported previously,[22] we based all subsequent modelling around a dimeric assembly of suPAH2:1 monomers. To that end, we employed a clustering based protein-protein docking strategy in order to predict energetically favored dimeric arrangements. The most highly populated clusters revealed conformations in good agreement with PahZ2$_{\text{KT-1}}$ dimers, as evidenced by RMSD and structural comparisons shown in Figure 4.5B-C. Similar to PahZ2$_{\text{KT-1}}$, the dimeric interface was also found to be comprised of a mixture of polar and non-polar residues with hydrophobic interactions likely contributing to complex formation. This finding is supported by the results of a subsequent hotspot analysis, which predict residues of the disordered loops on either side of the dimerization domain to drive dimer association. Taken together, the presented suPAH2:1 model presents a structure with high confidence that is predicted to demonstrate conformational dynamics that are favorable for successful multi-step catalysis. The further modelling of the dimeric complex demonstrates interface features that are consistent with experimental observations taken from PahZ2$_{\text{KT-1}}$.

5.3 Discussion

Polymers of aspartic acid (PAA) are a promising avenue for materials research, offering superabsorbent properties and biosafe applications ranging from agricultural soil treatment to product packaging to biomedical implants and drug delivery. Unlike existing polymers, including the polycarboxylates, thermally synthesized PAA (tPAA)-based materials are thought to be much more ecofriendly thanks to their ability to be degraded by the previously characterized hydrolases. Throughout this work, we have attempted to emphasize both the strengths and the various catalytic shortcomings of PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ activity. For example, the former has been characterized to demonstrate a preference for cleaving β-amide linkages within the tPAA structure, while α-linkages remain largely in tact. Such substrate selectivity may significantly restrict the utility of

$\text{PahZ1}_{\text{KT-1}}$ in commercial applications and bioremediation. We therefore hope that our structural work elucidating possible mechanisms for substrate selection will lead to future strategies to engineer a more promiscuous enzyme. An ideal candidate would be capable of acting not only on α- and β-linked PAA, but other polycarboxylates currently threatening the ecosystem with environmental persistence.

In contrast, $\text{PahZ2}_{\text{KT-1}}$ exhibits no such β-linked preferences. However, it was previously noted that it struggles to achieve complete degradation of OAA into aspartic acid monomers. Such inefficiency is a hindrance to the proposed cradle-to-cradle synthetic process of tPAA, wherein the product of $\text{PahZ2}_{\text{KT-1}}$ catalysis can be reclaimed and used as source material for manufacturing additional polymer. Similar to our rationale with $\text{PahZ1}_{\text{KT-1}}$, our work to characterize the structure and dynamics of $\text{PahZ2}_{\text{KT-1}}$ is hoped to enable the future engineering of a more catalytically efficient tPAA hydrolase. For example, we note in the current work that there may be a link between catalytic efficiency of $\text{PahZ2}_{\text{KT-1}}$ and conformational switching between open and closed states. This was further found to be an NaCl-dependent process, such that high salt conditions were observed to stimulate the transition from an open to closed conformation resulting in the closer proximity of bound substrate to the catalytic zinc atoms. Positioning of substrate is guided by a collection of arginine residues belonging to both the dimerization and catalytic domains. We note that many of these interactions are similarly NaCl-dependent. For example, R162 is observed to form stable hydrogen bonding interactions with bound PAA_5 that is weakened by the inclusion of salt in simulation conditions. In contrast, R165 and R200 interactions appear to be stabilized by NaCl. R270 uniquely seems to interact with PAA_5 substrates regardless of salt concentration. Overall, our findings suggest the catalytic output of $PahZ2_{KT-1}$ may be environmentally tunable by such factors as NaCl concentration. Further work will be needed, however, to more completely elucidate the role of various cationic residues in substrate binding. A stronger understanding here is expected to enable additional engineering efforts to build a more efficient enzymatic process wherein $PahZ2_{KT-1}$ domains are used to generate a greater yield of aspartic acid monomers. Designing such additional functionality for both enzymes would effectually render a feasible approach to the sustainable manufacturing and reclamation of tPAA.

We have additionally described the inherent inefficiency of a dual-enzyme strategy to tPAA remediation. The expression, purification, and mass production of two enzymes is far less cost-effective and less scalable than a single enzyme solution. To that end, we have proposed the development of a chimeric hydrolase consisting of both $PahZ1_{KT-1}$ and $PahZ2_{KT-1}$ functional domains in chapter 4. We explored two construct designs, wherein a $PahZ1_{KT-1}$ domain is attached to either terminus of a $PahZ2_{KT-1}$ monomer via a flexible $(GGS)_4$ linker (suPAH1:2 and 2:1). The rationale for this work is that the construction of such a chimera would enable the conversion of tPAA directly to aspartic acid monomers without the need for additional intermediate steps. We heavily employed state of the art modelling techniques, including deep-learning protein structure prediction and molecular dynamics simulations, to explore the feasibility of our designs.

The results suggest a possible route for substrate channeling such that tPAA is able to be converted into OAA at the PahZ1$_{\text{KT-1}}$ active site and subsequently orient toward the PahZ2$_{\text{KT-1}}$ catalytic zinc atoms without first returning to bulk solvent. We emphasize the modelling presented here is largely preliminary, and more work will be needed to confirm whether a given sequence design is truly catalytic.

A major limitation of our approach is the lack of crystal structure data of recombinant suPAH1:2 or 2:1 enzymes. To mitigate these shortcomings, we have utilized both PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ structures, as well as other structural homologues, as reference material for control comparisons with all in silico modelling. These comparisons, along with impressively favorable pLDDT scores presented for each model in chapter 4, allow us to place high confidence in the reported suPAH structures. Regardless, further characterization is necessary to more fully elucidate the feasibility of each construct. We ultimately seek as well to engineer additional functionality and enable a truly novel system of tPAA synthesis, biodegradation, and reclamation. In doing so, we hope to further the development of tPAA as a sustainable and ecofriendly biopolymer for a wide range of environmental, agricultural, commercial, and biomedical applications.

References

(1) Tabata, K.; Kasuya, K.; Abe, H.; Masuda, K.; Doi, Y. Poly(aspartic acid) degradation by a Sphingomonas sp. isolated from freshwater. *Appl Environ Microbiol* **1999**, *65*, 4268–4270.

(2) Tabata, K.; Abe, H.; Doi, Y. Microbial Degradation of Poly(aspartic acid) by Two Isolated Strains of Pedobacter sp. and Sphingomonas sp. *Biomacromolecules* **2000**, *1*, 157–161.

(3) Tabata, K.; Kajiyama, M.; Hiraishi, T.; Abe, H.; Yamato, I.; Doi, Y. Purification and characterization of poly(aspartic acid) hydrolase from Sphingomonas sp. *KT-1. Biomacromolecules* **2001**, *2*, 1155–60.

(4) Brambley, C.; Bolay, A.; Salvo, H.; Jansch, A.; Yared, T.; Miller, J.; Wallen, J.; Weiland, M. Structural Characterization of Sphingomonas sp. KT-1 PahZ1-Catalyzed Biodegradation of Thermally Synthesized Poly(aspartic acid. *ACS Sustainable Chemistry & Engineering* **2020**, *8*, 10702–10713.

(5) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, PMID: 26620784, 435–447.

(6) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In *Solving Software Challenges for Exascale*, ed. by Markidis, S.; Laure, E., Springer International Publishing: Cham, 2015, pp 3–27.

(7) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(8) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(9) Hiraishi, T.; Kajiyama, M.; Tabata, K.; Abe, H.; Yamato, I.; Doi, Y. Biochemical and molecular characterization of poly(aspartic acid) hydrolase-2 from sphingomonas sp. *KT-1. Biomacromolecules* **2003**, *4*, 1285–92.

(10) Hiraishi, T.; Kajiyama, M.; Yamato, I.; Doi, Y. Enzymatic hydrolysis of alpha- and beta-oligo(L-aspartic acid)s by poly(aspartic acid) hydrolases-1 and 2 from Sphingomonas sp. *KT-1. Macromol Biosci* **2004**, *4*, 330–9.

(11) Nocek, B.; Gillner, D.; Fan, Y.; Holz, R.; Joachimiak, A. Structural basis for catalysis by the mono- and dimetalated forms of the dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase. *J Mol Biol* **2010**, *397*, 617–26.

(12) Bienvenue, D.; Gilner, D.; Davis, R.; Bennett, B.; Holz, R. Substrate specificity, metal binding properties, and spectroscopic characterization of the DapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from Haemophilus influenzae. *Biochemistry* **2003**, *42*, 10756–63.

(13) Dudev, T.; Lim, C. Tetrahedral vs Octahedral Zinc Complexes with Ligands of Biological Interest: A DFT/CDM Study. *J. Am. Chem. Soc* **2000**, *122*, 11146–11153.

(14) Jabrani, A.; Makamte, S.; Moreau, E.; Gharbi, Y.; Plessis, A.; Bruzzone, L.; Sanial, M.; Biou, V. Biophysical characterisation of the novel zinc binding property in Suppressor of Fused. *Sci Rep* **2017**, *7*, 11139.

(15) Guerrerio, A.; Berg, J. Metal ion affinities of the zinc finger domains of the metal responsive element-binding transcription factor-1 (MTF1. *Biochemistry* **2004**, *43*, 5437–44.

(16) Lesburg, C.; Huang, C.; Christianson, D.; Fierke, C. Histidine –> carboxamide ligand substitutions in the zinc binding site of carbonic anhydrase II alter metal coordination geometry but retain catalytic activity. *Biochemistry* **1997**, *36*, 15780–91.

(17) Stojanovic, A.; Stitham, J.; Hwa, J. Critical role of transmembrane segment zinc binding in the structure and function of rhodopsin. *J Biol Chem* **2004**, *34*, 279.

(18) Hiraishi, T.; Maeda, M. Poly(aspartate) hydrolases: biochemical properties and applications. *Appl Microbiol Biotechnol* **2011**, *91*, 895–903.

(19) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold - Making protein folding accessible to all. *bioRxiv* **2022**, DOI: `10.1101/2021.08.15.456425`.

(20) Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **2021**, *596*, 590—596.

(21) Kummer, M. J.; Lee, Y. S.; Yuan, M.; Alkotaini, B.; Zhao, J.; Blumenthal, E.; Minteer, S. D. Substrate Channeling by a Rationally Designed Fusion Protein in a Biocatalytic Cascade. *JACS Au* **2021**, *1*, PMID: 34467357, 1187–1197.

(22) Brambley, C. A.; Yared, T. J.; Gonzalez, M.; Jansch, A. L.; Wallen, J. R.; Weiland, M. H.; Miller, J. M. Sphingomonas sp. KT-1 PahZ2 Structure Reveals a Role for Conformational Dynamics in Peptide Bond Hydrolysis. *The Journal of Physical Chemistry B* **2021**, *125*, PMID: 34060838, 5722–5739.

APPENDICES

Appendix A

PahZ1$_{KT\text{-}1}$ Supporting Information

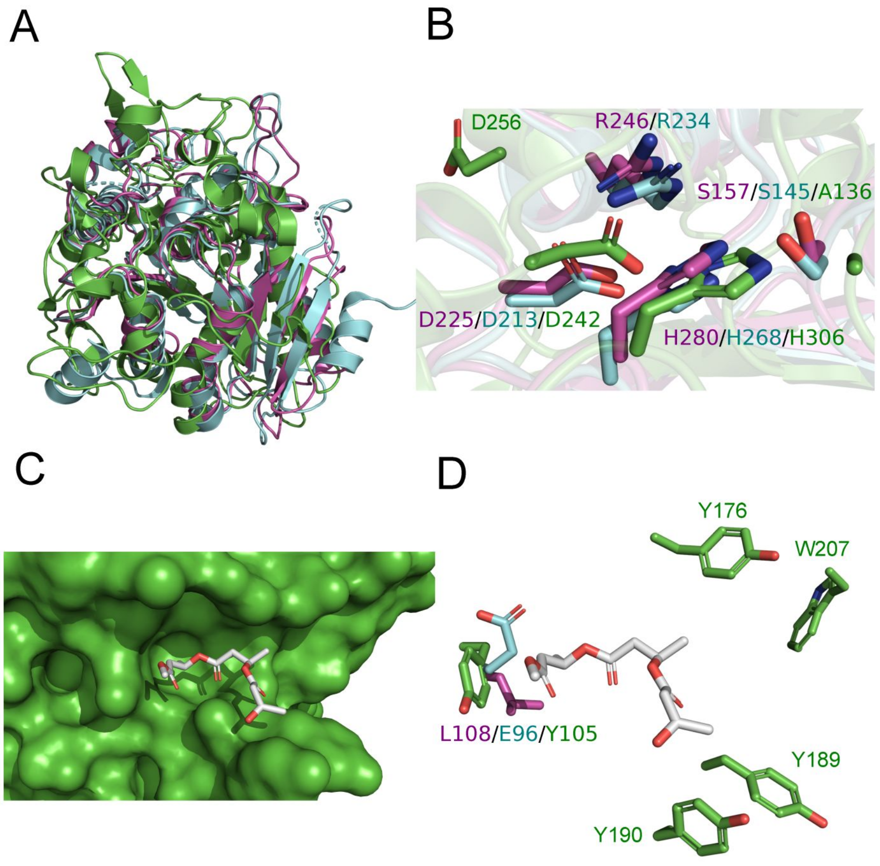


Figure A.1: Structural Comparison of PahZ1$_{KT\text{-}1}$, *P. lemoignei* PHB depolymerase (PDBID 4BTV), and *B. parapertussis* PHB depolymerase (PDBID 3D0K). A) Structural superposition of all three structures, with PahZ1$_{KT\text{-}1}$ colored magenta, *P. lemoignei* PHB depolymerase colored green, and *B. parapertussis* PHB depolymerase colored cyan. The *P. lemoignei* and *B. parapertussis* structures superimpose on PahZ1$_{KT\text{-}1}$ with Root Mean Square Deviations of 3.04 Å and 1.57 Å, respectively. B) The superposition aligns the active site residues of all three structures well, with the catalytic triad Ser/His/Asp of the enzymes structurally conserved. In the *P. lemoignei* PHB depolymerase structure used in this analysis, an S136A catalytically impaired mutant was used in order to obtain the structure of the enzyme bound to the PHB substrate. A key difference in the active site lies at the R246 position of PahZ1$_{KT\text{-}1}$. While the equivalent residue in *B. parapertussis* is also an arginine (R234), an aspartic acid (D256) is present in the *P. lemoignei* structure. C) Surface representation of the *P. lemoignei* PHB depolymerase structure with PHB substrate bound. The PHB substrate is shown in white. D) In the *P. lemoignei* structure a series of aromatic residues make up the binding pocket that recognizes the PHB substrate. Both PahZ1$_{KT\text{-}1}$ and the *B. parapertussis* structure lack this binding pocket. L108 in PahZ1$_{KT\text{-}1}$ and E96 in the *B. parapertussis* structure are found at the equivalent position of *P. lemoignei* residue Y105. The remaining aromatic residues are unique to the *P. lemoignei* PHB depolymerase, with no structural equivalent found in the other two enzymes.

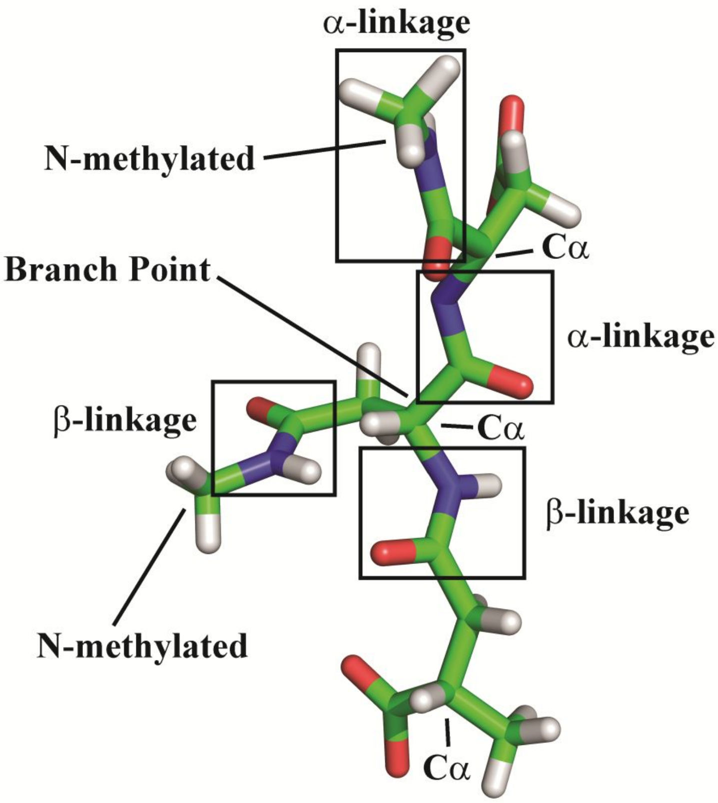
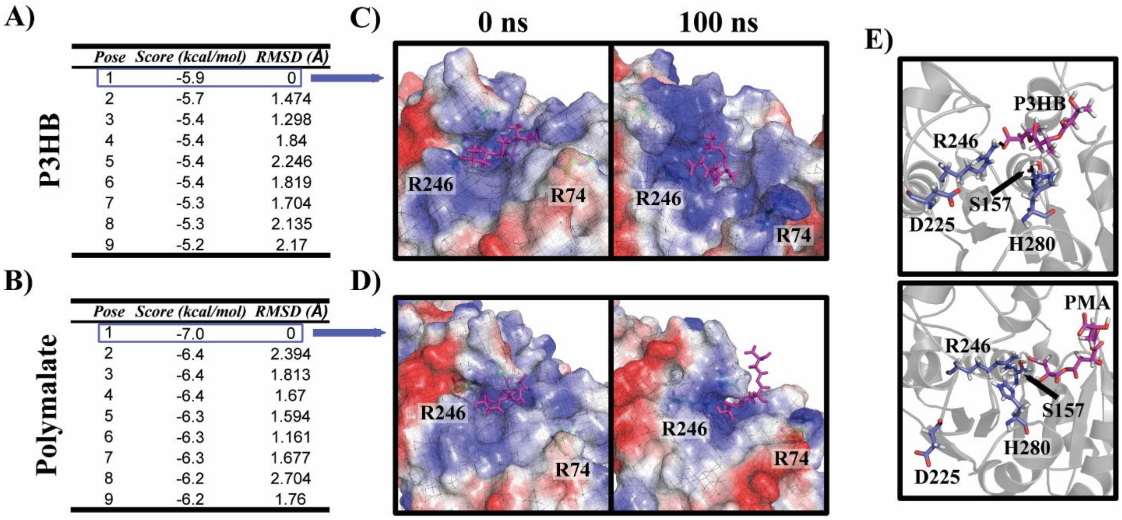


Figure A.2: Schematic representation of in silico-generated β-PAA substrate. β-PAA substrate has been modeled as a representative fragment of tPAA containing α- and β-linkages as well as an internal branch site. All amide functional groups have been N-methylated to mimic tPAA hydrogen bonding potential. All features are indicated by labeling.



Pose	*Score (kcal/mol)*	*RMSD (Å)*
1	-5.9	0
2	-5.7	1.474
3	-5.4	1.298
4	-5.4	1.84
5	-5.4	2.246
6	-5.4	1.819
7	-5.3	1.704
8	-5.3	2.135
9	-5.2	2.17

Pose	*Score (kcal/mol)*	*RMSD (Å)*
1	-7.0	0
2	-6.4	2.394
3	-6.4	1.813
4	-6.4	1.67
5	-6.3	1.594
6	-6.3	1.161
7	-6.3	1.677
8	-6.2	2.704
9	-6.2	1.76

Figure A.3: Docking results for poly(3-hydroxybutyrate) and polymalate to PahZ1$_{KT\text{-}1}$. (A-B) Table of docking scores for the top nine hits for poly(3-hydroxybutyrate) (A) and polymalate (B). (C-D) The positions of each polymer are shown in relation to arginines R246 and R74 at the t = 0 and 100 nanoseconds of unrestrained molecular dynamics for poly(3-hydroxybutyrate) (C) and polymalate (D). PahZ1$_{KT\text{-}1}$ vacuum electrostatics are represented in surface view and colored by surface charge with blue and red indicating positive and negative charge, respectively. E) Atomic positions of equilibrated PahZ1$_{KT\text{-}1}$ complex with poly(3-hydroxybutyrate) and polymalate are represented in stick. Polymer substrate and catalytic triad elements are colored in magenta stick and blue stick, respectively.

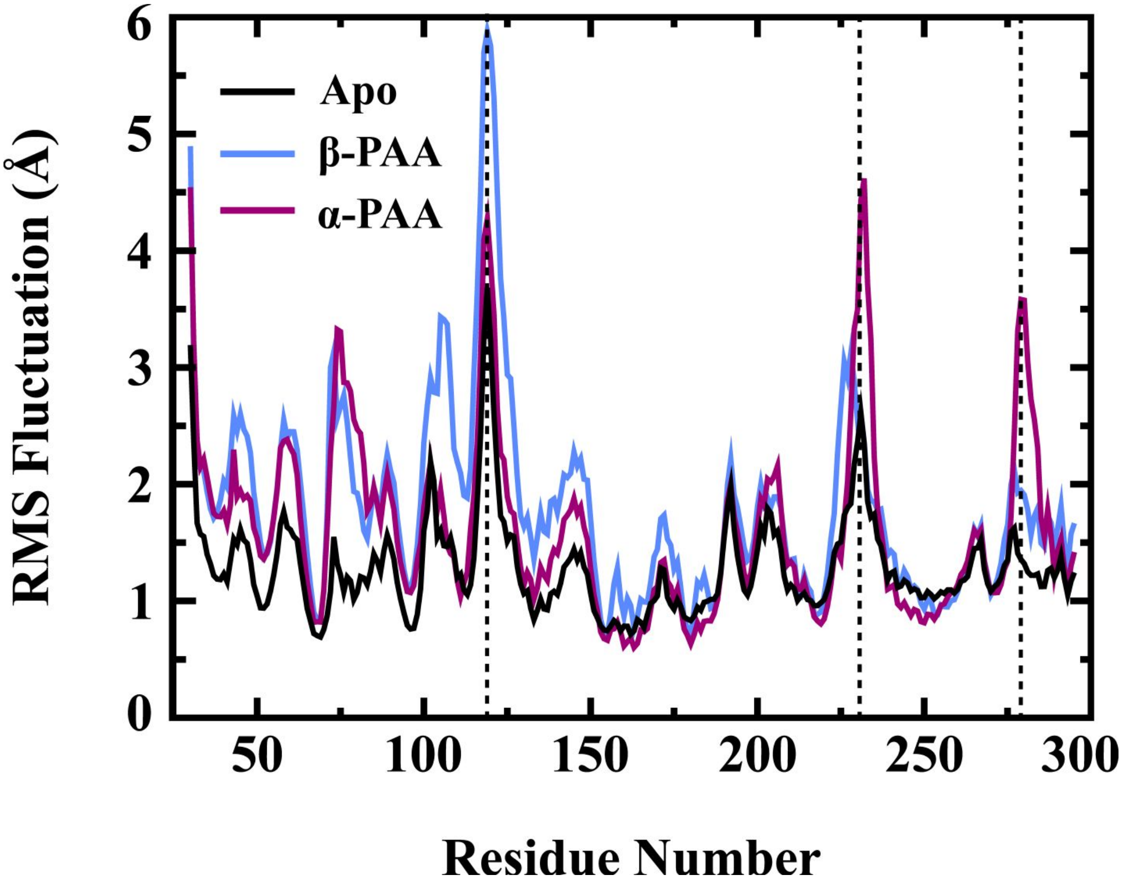


Figure A.4: Analysis of root mean square fluctuation in angstroms plotted as a function of residue number for the substrate bound chain of PahZ1$_{\text{KT-1}}$. Apoenzyme, α-PAA, and β-PAA conditions are shown in black, magenta, and blue respectively. Dashed lines indicate disordered and highly flexible loop regions

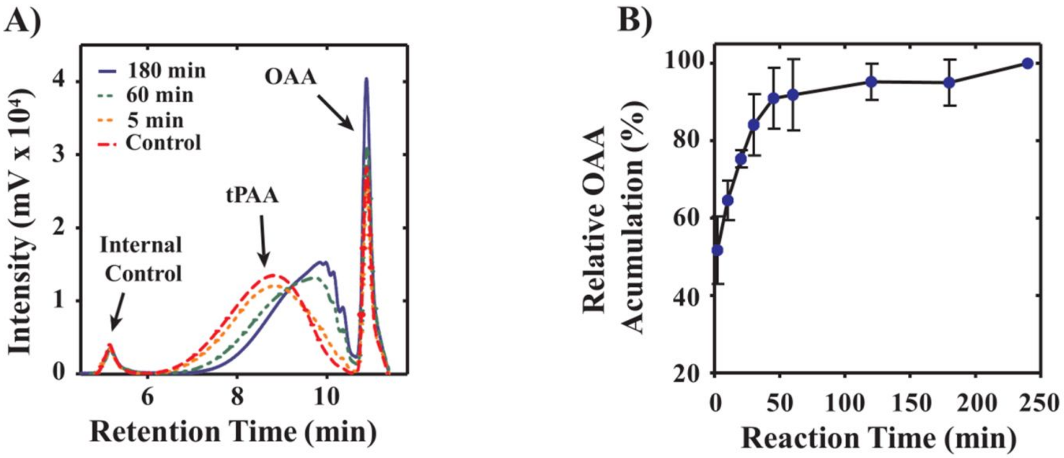


Figure A.5: PahZ1$_{KT\text{-}1}$ wild-type GPC activity assay. A) Overlay of three assays run on the GPC showing increasing degradation of the tPAA and formation of an OAA peak as the enzyme is allowed to incubate with substrate. The internal loading control is thyroglobulin. B) Determination of percent PahZ1$_{KT\text{-}1}$ activity at multiple time points using GPC data, each time-point has a n = 3. Shown is the plateauing of OAA peak formation after approximately one hour of incubation.

Appendix B

PahZ2$_{KT-1}$ Supporting Information

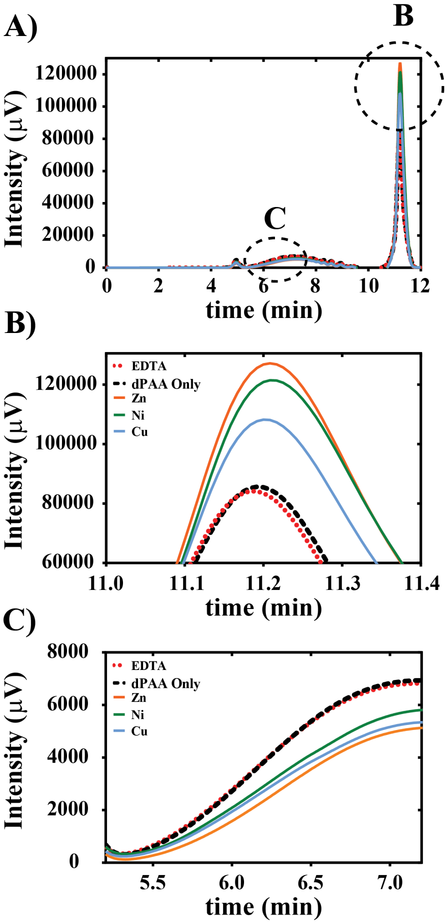


Figure B.1: GPC Chromatograms showing PahZ2$_{\text{KT-1}}$ activity based on changing the metal cofactor and descriptions of chromatogram analysis. (A) Overlay of chromatograms showing the substrate dPAA alone, PahZ2$_{\text{KT-1}}$ incubated with EDTA, or different metals. Select metals were chosen to simplify the figure but all followed the same trends seen. Circled and labeled as either "A" or "B" are the regions zoomed in and shown in panels B and C. (B) While the product formation of aspartic acid can be seen at $\approx$11.2 minutes, as evidenced by an increase in intensity with the Zn^{2+} showing maximal product formation, it also overlaps with the large refractive index peak appearing at the end of the chromatogram. (C) To ensure the change in intensity was due to enzyme activity, chromatographic analysis was also performed at a time of 7.0 minutes by looking at the loss of the dPAA oligomeric peak. Here decreases in intensity directly correspond to enzyme activity. A comparison of enzyme activity via product formation at time 11.2 min or polymer degradation at time 7.0 resulted in nearly identical values (data not shown) indicating either chromatographic approach is acceptable to assess PahZ2$_{\text{KT-1}}$ activity.

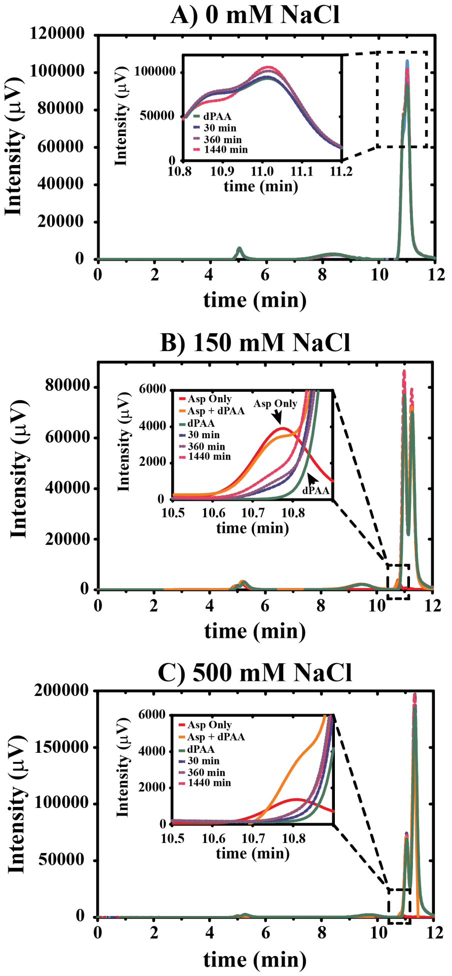


Figure B.2: GPC Chromatograms showing PahZ2$_{KT\text{-}1}$ activity based on changing salt concentrations. (A) Time-course activity assay of PahZ2$_{KT\text{-}1}$ in the absence of NaCl. The inset shows representative data and increased product formation with time. (B) and (C) The presence of NaCl changed the shapes of chromatograms and the large refractive index peak used to measure activity for (A) and moved down field. Therefore to identify where the aspartic acid product was eluting, controls of monomeric aspartic acid and dPAA + monomeric aspartic acid were run. (B) Representative chromatograms of PahZ2$_{KT\text{-}1}$ activity in 150 mM NaCl. The inset shows the aspartic acid alone elutes with a peak intensity at 10.75 minutes and the dPAA + aspartic acid shows a peak at the same location. Therefore, activity in 150 mM NaCl was measured using this retention time. (C) PahZ2$_{KT\text{-}1}$ activity in the presence of 500 mM NaCl. Monomeric aspartic acid further shifted down field and therefore activity was measured at 10.8 minutes. In all chromatograms, the peak shown at ≈5.0 minutes corresponds to thyroglobulin which served as an internal loading control. This data is intended to show PahZ2$_{KT\text{-}1}$ forms monomeric aspartic acid; however, the yield is relatively low. The data was also analyzed by the degradation of dPAA, as described in Figure S1, and that analysis correlated directly with what is shown by the formation of aspartic acid here.

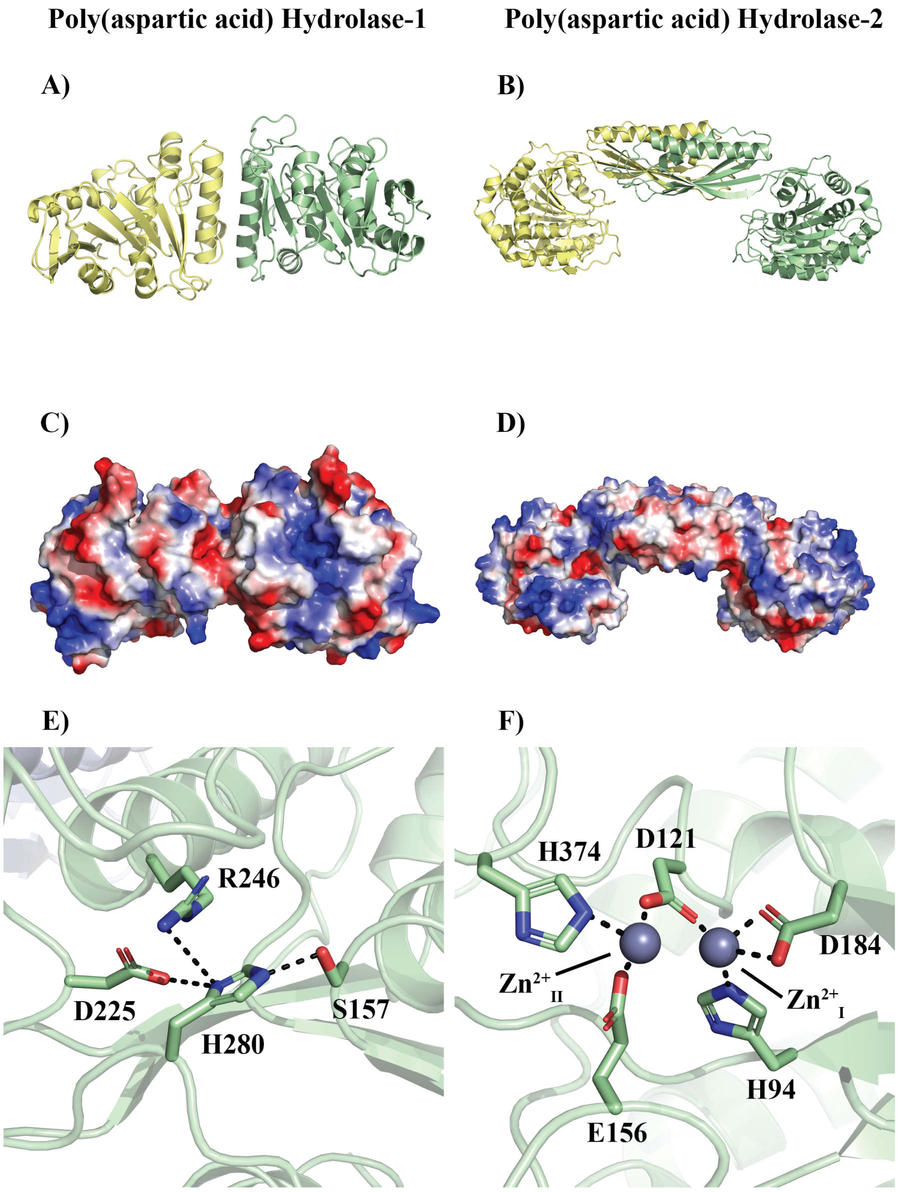
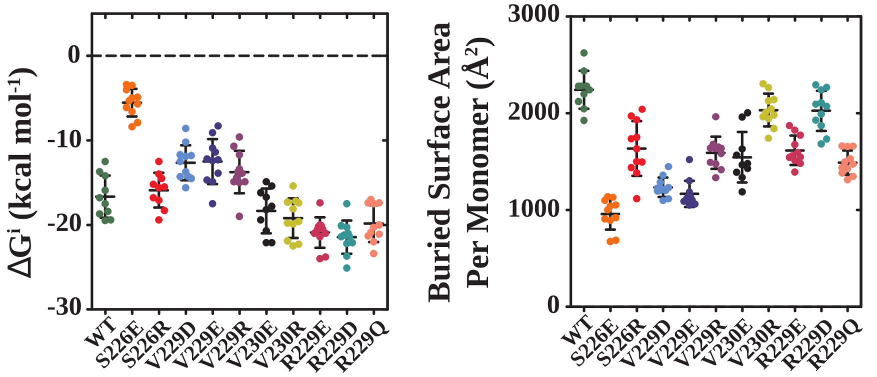


Figure B.3: Structural comparison between Sphingomonas sp. KT-1 PahZ1 and PahZ2. A-B) Cartoon representation of PahZ1$_{KT\text{-}1}$ (A) and PahZ2$_{KT\text{-}1}$ (B) structures. Individual subunits are colored yellow and green. C-D) Electrostatic potential surface representations for PahZ1$_{KT\text{-}1}$ (C) and PahZ2$_{KT\text{-}1}$ (D). Comparison of the active sites for PahZ1$_{KT\text{-}1}$ (E) and PahZ2$_{KT\text{-}1}$ (F) identify each enzyme as belonging to the serine protease and M28 metallopeptidase families, respectively. All structural representations were prepared using the Pymol software package and Protein Data Bank accession codes 6VE6 and 7LIH. Backbone representations are shown in cartoon and side-chains are shown in stick view with coloring by element.

A) Molecular Dynamics-Based PISA Analysis (Single Mutant Constructs)



B) Molecular Dynamics-Based PISA Analysis (Multiple Mutant Constructs)

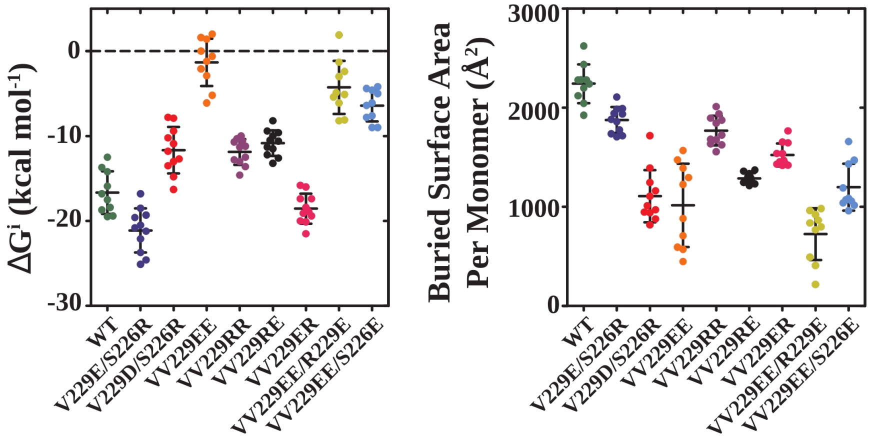
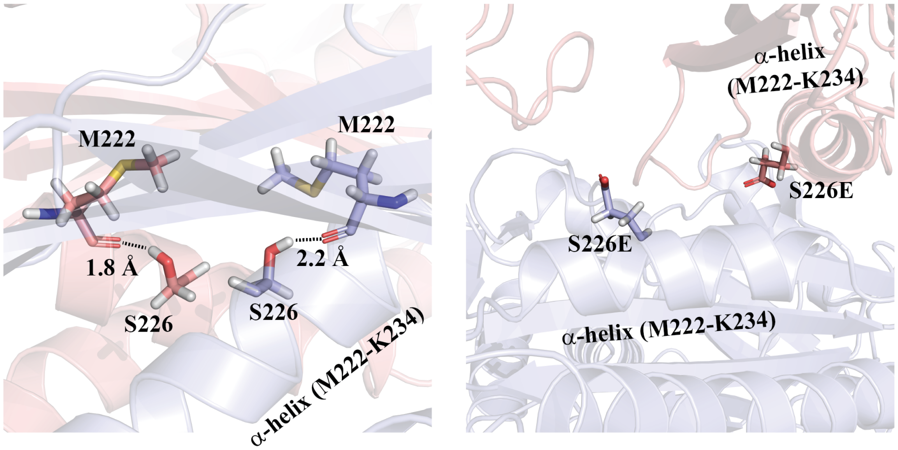
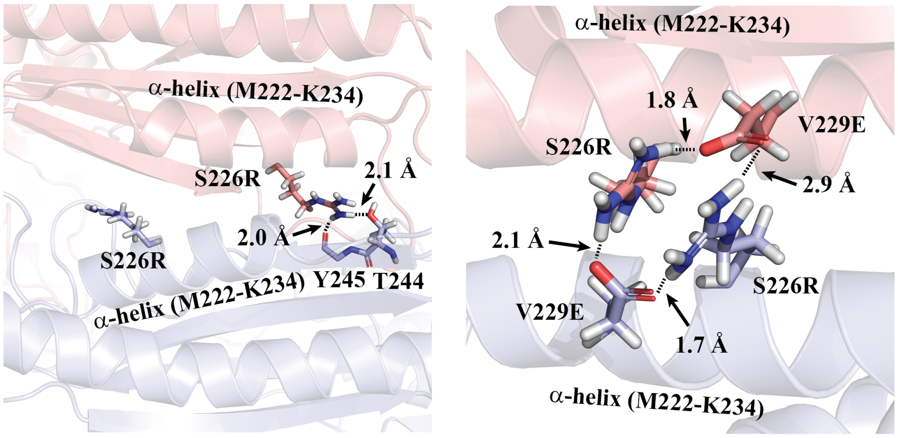


Figure B.4: Summary of Proteins, Interfaces, Structures, and Assemblies (PISA) analyses for the wild-type and mutant PahZ1$_{KT\text{-}1}$ structures. PahZ1$_{KT\text{-}1}$ variants were generated in silico and equilibrated using MD techniques. Extracted frames from each molecular dynamics trajectory were subjected to PISA analysis. The resulting estimates of gain in solvation free energy for complex formation, ΔG^{i}, and buried interface area per monomer are plotted separately as colored spheres labeled by mutation. All replicate values are plotted with mean and standard deviations indicated by bars. The results of single- versus multi-point mutant constructs are presented in Panels A and B, respectively.



A) Wild-Type PahZ2$_{KT\text{-}1}$

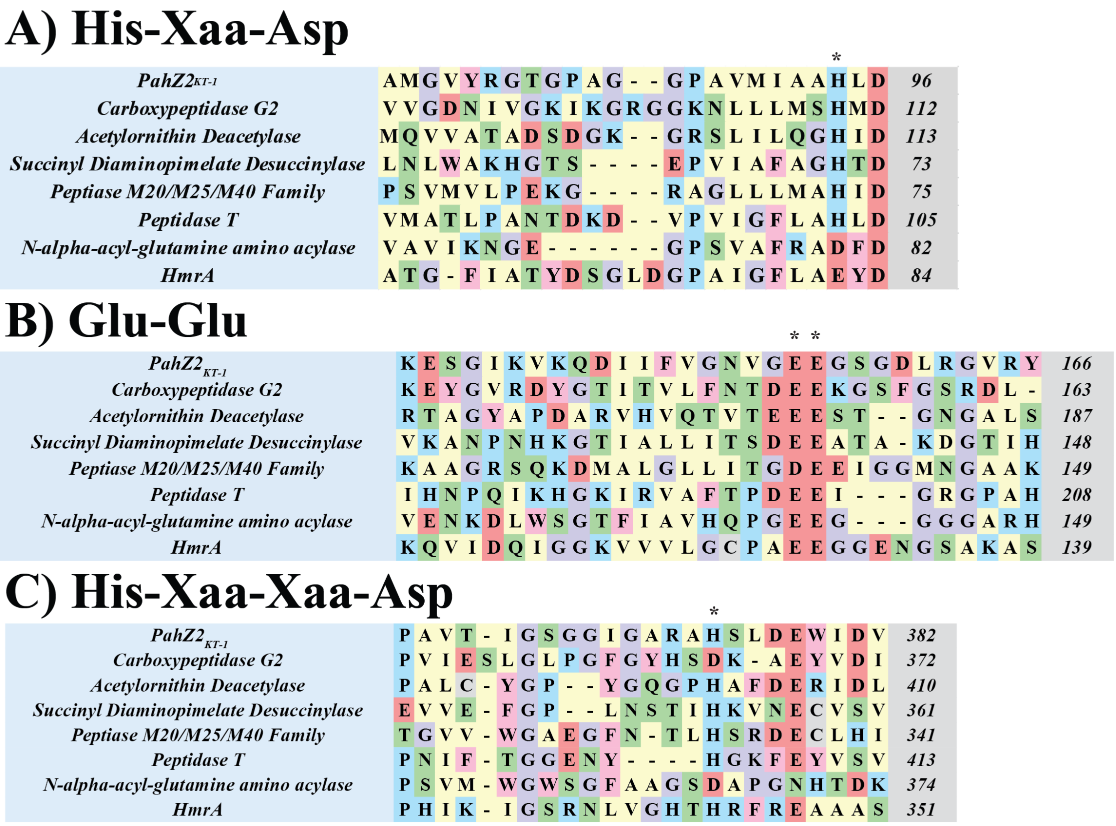
B) S226E PahZ2$_{KT\text{-}1}$



C) S226R PahZ2$_{KT\text{-}1}$

D) S226R/V229E PahZ2$_{KT\text{-}1}$

Figure B.5: Structural rationalization for trends in ΔG^i and buried interface area per monomer for in silico PahZ2$_{KT\text{-}1}$ mutant constructs. All structural representations were prepared using the Pymol software package. Backbone representations are shown in cartoon and side-chains are shown in stick view with coloring by element.



A) His-Xaa-Asp

```
                                                              *
PahZ2KT-1                                AMGVYRGTGPAG--GPAVMIAAHLD   96
Carboxypeptidase G2                      VVGDNIVGKIKGRGGKNLLLMSHMD  112
Acetylornithin Deacetylase               MQVVATADSDGK--GRSLILQGHID  113
Succinyl Diaminopimelate Desuccinylase   LNLWAKHGTS----EPVIAFAGHTD   73
Peptiase M20/M25/M40 Family              PSVMVLPEKG----RAGLLLMAHID   75
Peptidase T                              VMATLPANTDKD--VPVIGFLAHLD  105
N-alpha-acyl-glutamine amino acylase     VAVIKNGE------GPSVAFRADFD   82
HmrA                                     ATG-FIATYDSGLDGPAIGFLAEYD   84
```

B) Glu-Glu

```
                                                           **
PahZ2KT-1                                KESGIKVKQDIIFVGNVGEEGSGDLRGVRY  166
Carboxypeptidase G2                      KEYGVRDYGTITVLFNTDEEKGSFGSRDL-  163
Acetylornithin Deacetylase               RTAGYAPDARVHVQTVTEEEST--GNGALS  187
Succinyl Diaminopimelate Desuccinylase   VKANPNHKGTIALLITSDEEATA-KDGTIH  148
Peptiase M20/M25/M40 Family              KAAGRSQKDMALGLLITGDEEIGGMNGAAK  149
Peptidase T                              IHNPQIKHGKIRVAFTPDEEI---GRGPAH  208
N-alpha-acyl-glutamine amino acylase     VENKDLWSGTFIAVHQPGEEG---GGGARH  149
HmrA                                     KQVIDQIGGKVVVLGCPAEEGGENGSAKAS  139
```

C) His-Xaa-Xaa-Asp

```
                                                        *
PahZ2KT-1                                PAVT-IGSGGIGARAHSLDEWIDV  382
Carboxypeptidase G2                      PVIESLGLPGFGYHSDK-AEYVDI  372
Acetylornithin Deacetylase               PALC-YGP--YGQGPHAFDERIDL  410
Succinyl Diaminopimelate Desuccinylase   EVVE-FGP--LNSTIHKVNECVSV  361
Peptiase M20/M25/M40 Family              TGVV-WGAEGFN-TLHSRDECLHI  341
Peptidase T                              PNIF-TGGENY----HGKFEYVSV  413
N-alpha-acyl-glutamine amino acylase     PSVM-WGWSGFAAGSDAPGNHTDK  374
HmrA                                     PHIK-IGSRNLVGHTHRFREAAAS  351
```

Figure B.6: Multiple sequence alignments highlights the His-Xaa-Asp (A), Glu-Glu (B), and His-Xaa-Xaa-Asp (C) motifs located at the ZnI and ZnII sites for top DALI search results.

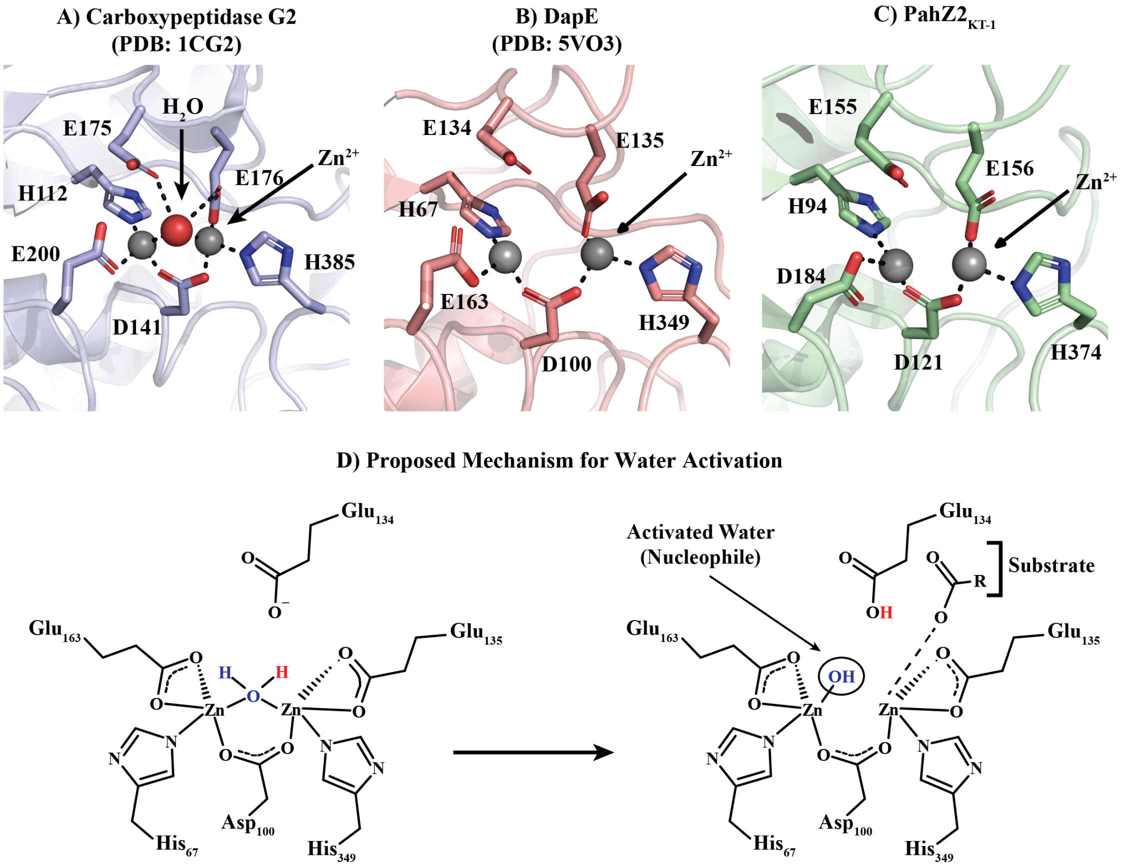


Figure B.7: Active site architecture for A) carboxypeptidase G2, B) DapE, and C) PahZ2$_{KT-1}$. All structural representations were prepared using the Pymol software package. Backbone representations are shown in cartoon and side-chains are shown in stick view with coloring by element. Zinc metal ions and water molecules are shown are gray or red spheres, respectively. All active site features are labeled. D) The proposed DapE chemical mechanism for general base catalysis involving nucleophile generation from bulk water.

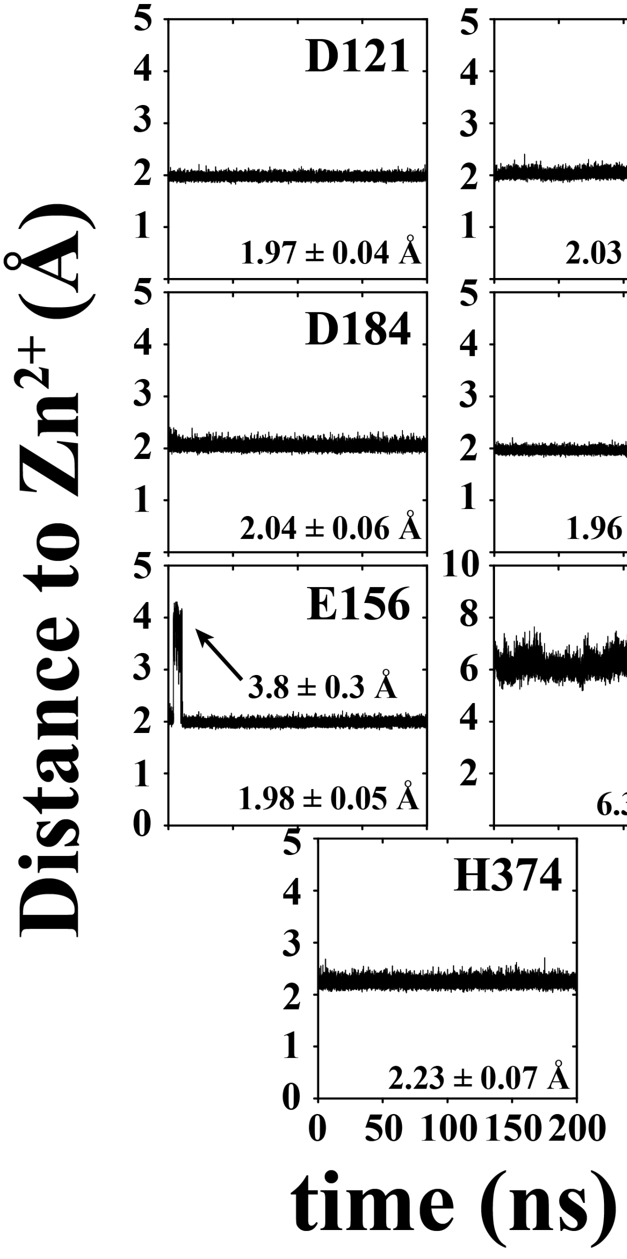


Figure B.8: Distance to nearest zinc metal ion for all wild-type PahZ2$_{KT-1}$ active site residues as a function of time. The side-chain to Zn^{2+} distance has been measured as a function of molecular dynamics simulation time. The mean distance and standard deviation calculated over the entire MD trajectory is provided for each plot.

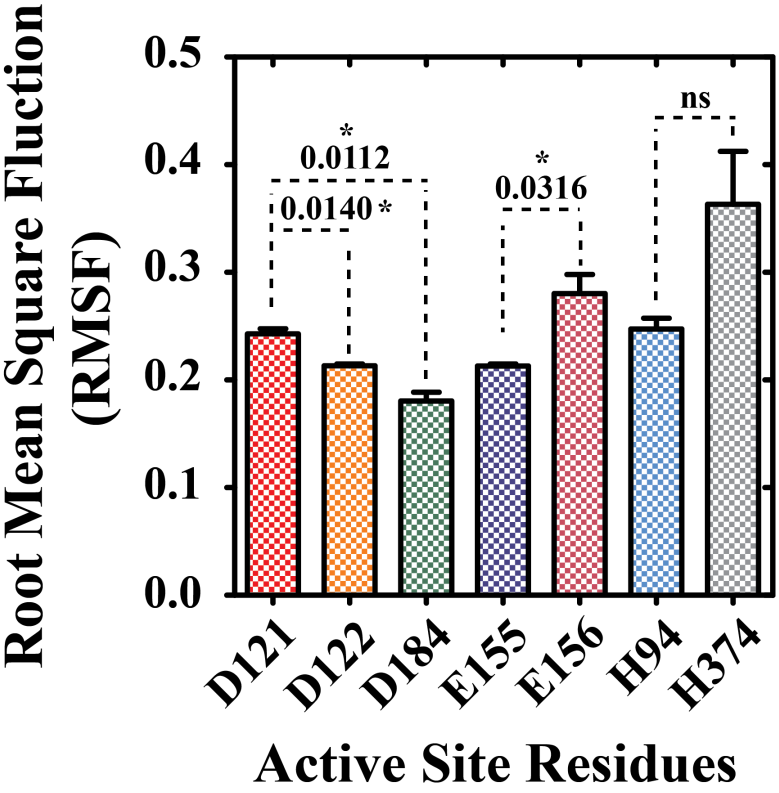


Figure B.9: Root Mean Square Fluctuations calculated for each wild-type PahZ2$_{KT\text{-}1}$ active site residue using 200 ns molecular dynamics trajectory data. Application of pairwise T-testing was utilized to determine whether RMSF values were statistically unique for specific comparisons, which are indicated by dashed lines. Numerical labels indicate P-value. NS label indicates "Not Significant." RMSF standard deviations are indicated by error bars.

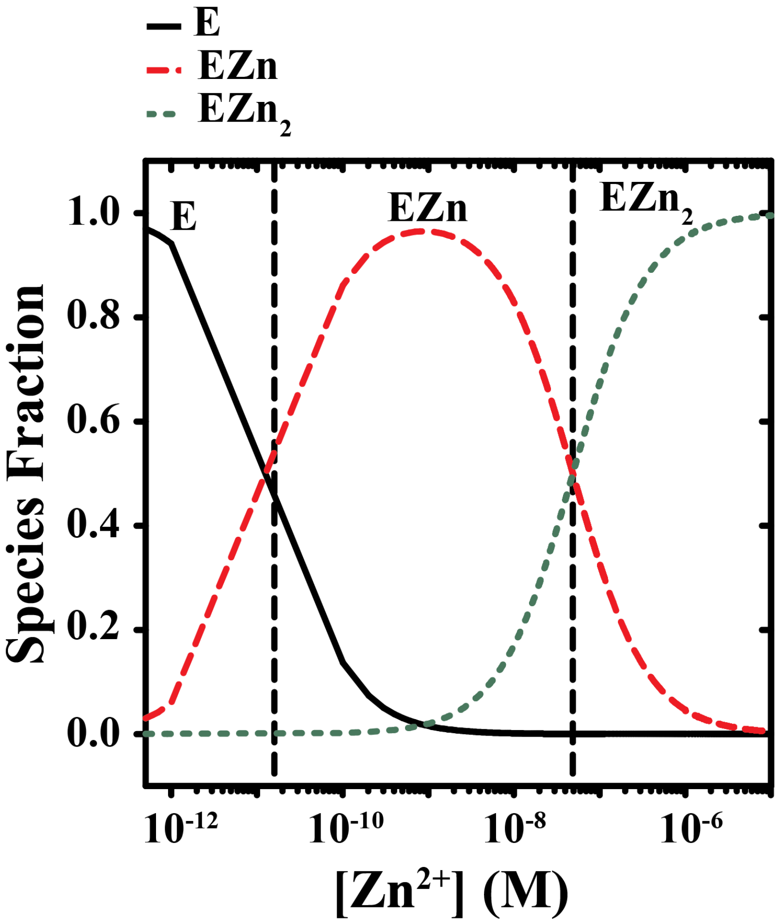


Figure B.10: Calculated species distribution for E, EZn, EZn_2 species as a function of $[Zn^{2+}]$. Calculations were performed using a two-site, independent binding model with dissociation constants $K_{D,1}$ and $K_{D,2}$ set equal to 16 pM and 48.7 nM, respectively. The fraction E, EZn, and EZn_2 species are indicated using black solid, red dashed, and green dotted lines, respectively. Vertical black dashed lines mark the numerical values on the x-axis for $K_{D,1}$ and $K_{D,2}$.

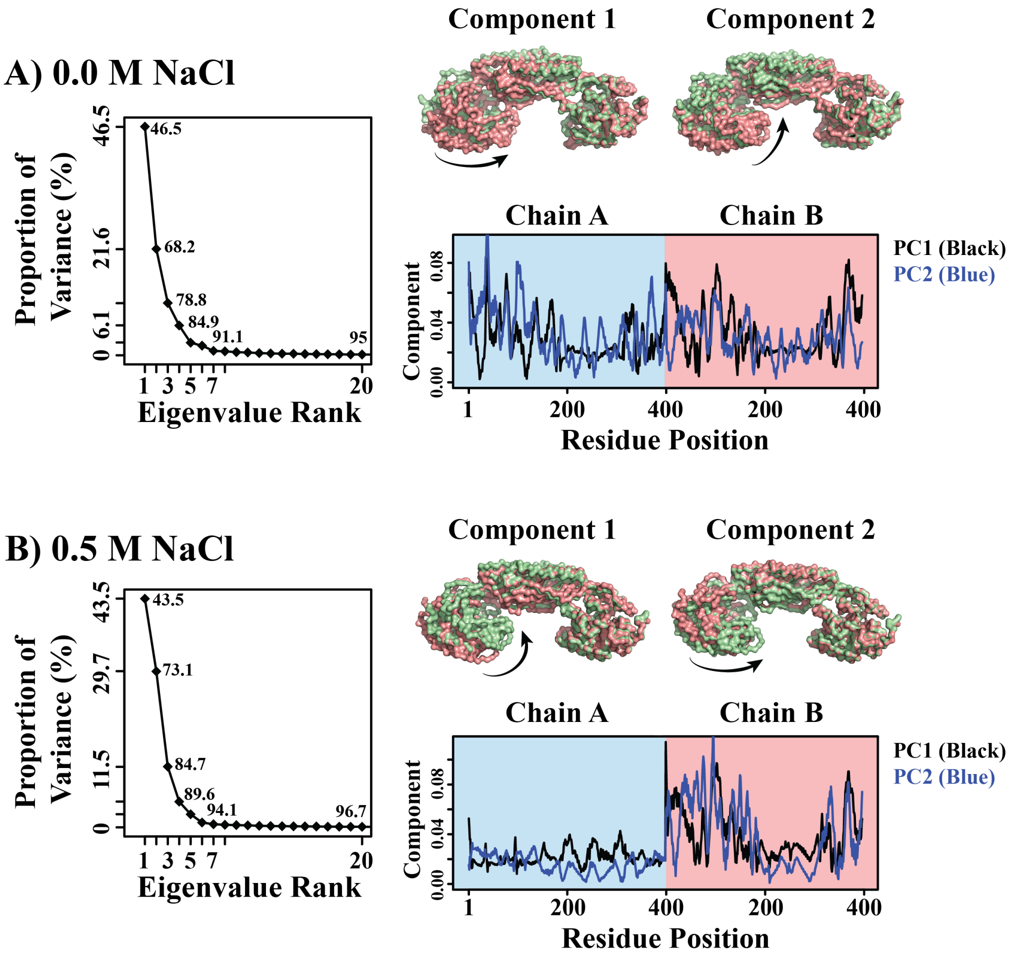


Figure B.11: Principal Component Analyses for 200 ns Molecular Dynamics trajectories simulated in the presence of 0 (A) or 0.5 (B) M NaCl. Scree plots highlight the number of Principal components accounting for at least 95% of system variance. Principal Components 1 and 2 (PC1 and PC2) have each been projected onto the PahZ2$_{KT\text{-}1}$ structure to visualize isolated motions corresponding to each component. PahZ2$_{KT\text{-}1}$ motions are represented in surface views with superimposed structures for the beginning and end frame for each principal component. The fractional contribution of each PahZ2$_{KT\text{-}1}$ residue to PC1 and PC2 have been calculated are plotted as fractional contribution versus residue position.

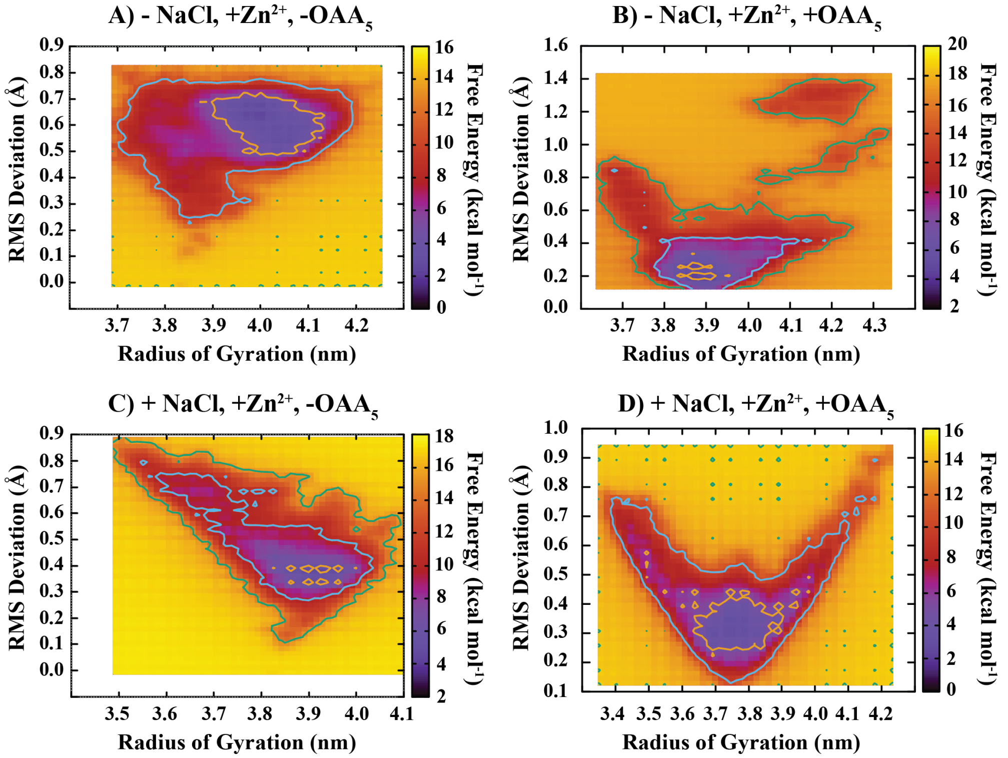


Figure B.12: Molecular dynamics simulations reveal the effect of solution conditions on the PahZ2$_{KT\text{-}1}$ free energy landscape. Three-dimensional plots of RMSD, radius of gyration, and free energy (x, y, z) provide an estimate of accessible PahZ2$_{KT\text{-}1}$ conformations for conditions including A) -NaCl, Zn^{2+}, -OAA$_5$, B) -NaCl, +Zn^{2+}, +OAA$_5$, C) +NaCl, +Zn^{2+}, and D) +NaCl, +Zn^{2+}, +OAA$_5$. Calculated free energy values are colored on the indicated scale from low to high from dark purple to yellow, respectively. All simulations were performed as described in Materials and Methods.

B.1 Supplemental Methods

Production Molecular Dynamics

All models prepared for in silico studies were structurally equilibrated with 200 ns of production MD via the Gromacs molecular simulation package.[1,2] Briefly, all models were solvated in triclinic boxes with simple point charge water and neutralizing ions. Solvated systems were minimized using a steepest descent energy minimization protocol. Except where otherwise noted, all minimized structures were thermally equilibrated for up to 5,000 ps of position restrained NVT and NPT ensembles with a Berendsen thermostat. Unrestrained production MD was then carried out for 200 ns with Nose-Hoover[3,4] and Parinello-Rahman[5,6] temperature and pressure coupling respectively. Frames were then extracted from the equilibrated wild-type and mutant PahZ2$_{\text{KT-1}}$ systems and subjected to PDBePISA analysis as previously described.[7]

B.2 Supplemental Results

In silico Examination of PahZ2$_{KT\text{-}1}$ Dimer Interface

Critical examination of the amino acids found at the dimer interface reveals a significant nonpolar population. PDBePISA analysis indicates the presence of multiple nonpolar residues in the α-helix spanning residues M222-K234, which has the sequence MVAMSQTVVDFYK. In order to better understand the contribution of these residues to dimer stability, we generated a series of in silico PahZ2$_{\text{KT-1}}$ interface mutants for calculation of solvation free energy gain, ΔG^i, and buried interface surface area per monomer by PDBePISA analysis as described.[7] We generated nineteen PahZ2$_{\text{KT-1}}$ mutants using the Pymol software package,8 followed by energy minimization, solvation, and equilibration using molecular dynamics methods available within the GROMACS environment. Each PahZ2$_{\text{KT-1}}$ mutant model was brought to equilibrium through a minimum of 200 ns of production molecular dynamics simulation. From each equilibrated system, 10 frames were extracted and subjected to analysis using the PDBePISA tool for calculation of ΔG^i and interface area. We have previously applied this workflow to the study of PahZ1$_{\text{KT-1}}$ and noted significant advantages.[7] First, the introduction of a mutation as described, followed by system equilibration, allows conformational rearrangements to occur that account for newly incorporated interface residues that were not present in the wild-type crystal structure. Second, molecular dynamics-based analysis as described provides the ability to capture a dynamic view of ΔG^i and interface area. Lastly, in our previous report,[7] we applied rigorous Center-of-Mass steered molecular dynamics/Umbrella Sampling techniques to conclude that calculation of solvation free energy gains using the PDBePISA tool yields comparable conclusions to more rigorous calculation methods.

Figure B.3A-B highlight the resulting solvation free energy gains and buried surface area per monomer (interface area) for each in silico-generated PahZ2$_{\text{KT-1}}$ mutant. For wild-type PahZ2$_{\text{KT-1}}$, Figure B.3A reveals ΔG^i = -16.7 ± 2.5 kcal mol^{-1} and interface area = 2242 ± 195 Å^2, which deviate by ≈ 25 % from ΔG^i and interface area estimated from the crystal structure. Such deviations between the x-ray structure and MD-generated structures may reflect differences in system energy with the crystal structure likely the lowest energy thermodynamic state. Given the intimate interactions formed between α-helices spanning residues M222-K234, and its nonpolar residue composition, we next generated PahZ2$_{\text{KT-1}}$ constructs containing single interface mutations to determine whether the introduction of charged residues would lead to dimer destabilization. Figure B.3A illustrates the effect of S226 mutations on ΔG^i and interface area where a S226E mutation destabilizes the PahZ2$_{\text{KT-1}}$ dimer with ΔG^i = -6 ± 2 kcal mol^{-1} and interface area = 958 ± 161 Å^2. Interestingly, S226 mutation to an arginine (S226R) yields ΔG^i = -15 ± 3 kcal mol^{-1} and interface area = 1636 ± 282 Å^2. Statistical analyses of ΔG^i indicate significant differences between wild-type PahZ2$_{\text{KT-1}}$ and S226E (P value < 0.0001) by t-testing, but no significant difference between mean values for wild-type PahZ2$_{\text{KT-1}}$ and S226R. However, the same comparisons for interface area reveal significant differences for both S226E and S226R relative to wild-type enzyme with mean differences equal to 1284 ± 78 Å^2 and 607 ± 107 Å^2, respectively. Moreover, S226E and S226R display differences when compared such that S226R displays increased stability with $\Delta\Delta G^i$ = -10.4 ± 0.8 kcal mol^{-1} and ΔIA = 677 ± 98 Å^2. Inspection of MD-equilibrated structures for wild-type, S226E, and S226R PahZ2$_{\text{KT-1}}$ models readily explain differences in dimer stability.

Figure B.4A reveals that S226 side-chain polarity is neutralized in the wild-type structure via a hydrogen bonding interaction formed between the S226 alcohol and the main-chain carbonyl oxygen of nearby M222, thereby promoting monomer stabilization and neutralization of polarity at a largely nonpolar dimer interface. However, mutation of S226 to glutamate (S226E) significantly disrupts the dimer interface such that interaction between α-helices is impaired as shown in Figure B.4B. In contrast, the introduction of an arginine in place of S226 yields an asymmetric effect such that the Chain A arginine forms new hydrogen bonding interactions, while the Chain B S226R position introduces charge without forming new hydrogen bonding interactions similar to the S226E construct (Figure B.4C). Thus, S226R is capable of dimer stabilization through the introduction of new hydrogen bonding interactions, though at the expense of interface area. Additional mutations that introduce charged functional groups at positions 229 and 230, V229D, V229E, V229R, V230D, V230E, and V230R, offer mild effects on ΔG^i, while strongly disrupting the dimer interface.

Combinatorial in silico PahZ2$_{\text{KT-1}}$ constructs reveal the potential to effect greater stabilization or destabilization on the dimer complex. Figure B.4D shows that, for example, a PahZ2$_{\text{KT-1}}$ V229E/S226R double-mutant is more stable than wild-type enzyme with $\Delta\Delta G^i$ = -4.4 ± 1.1 kcal mol^{-1} and ΔIA = -367 ± 72 Å^2. For S226R/V229E PahZ2$_{\text{KT-1}}$, the marginal decrease in interface area is associated with increased steric bulk at the interface between α-helices but is accompanied by the introduction four new hydrogen bonding interactions between S226R and V229R (Figure B.4D). Figure B.4D highlights that such stabilization is not possible when V229E is transitioned to V229D due to issues related to side-chain-length. However, the introduction of individually bulky side-chains does appear to promote dimer destabilization. PahZ2$_{\text{KT-1}}$ mutant constructs that alter the residue identities of both V229 and V230 have variable effects such that VV229EE nearly disrupts the dimer completely, while VV229RR and VV229RE impact complex stability like V229D/E/R mutations alone. No single- or double-mutation positively stabilized the dimer interface as evaluated by both interface area and ΔG^i, though a S226R/V229E double-mutation may represent the first step towards this goal.

References

(1) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.

(2) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(3) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **1984**, *81*, 511–519.

(4) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(5) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Stud. *Physical Review Letters* **1980**, *45*, 1196–1199.

(6) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **1981**, *52*, 7182–7190.

(7) Brambley, C.; Bolay, A.; Salvo, H.; Jansch, A.; Yared, T.; Miller, J.; Wallen, J.; Weiland, M. Structural Characterization of Sphingomonas sp. KT-1 PahZ1-Catalyzed Biodegradation of Thermally Synthesized Poly(aspartic acid. *ACS Sustainable Chemistry & Engineering* **2020**, *8*, 10702–10713.

Appendix C

suPAH Supporting Information

Table C.1: Weighted scores derived from ClusPro2 protein-protein docking of suPAH2:1 monomers

Cluster	**Members**	**Representative**	**Weighted Score**
0	144	Center	-966.0
0	144	Lowest Energy	-1075.5
1	138	Center	-980.8
1	138	Lowest Energy	-1076.5
2	79	Center	-633.8
2	79	Lowest Energy	-855.2
3	66	Center	-646.4
3	66	Lowest Energy	-785.3
4	64	Center	-704.5
4	64	Lowest Energy	-789.4
5	60	Center	-805.8
5	60	Lowest Energy	-817.0
6	55	Center	-609.0
6	55	Lowest Energy	-833.2
7	55	Center	-642.9
7	55	Lowest Energy	-842.6
8	48	Center	-658.5
8	48	Lowest Energy	-958.4

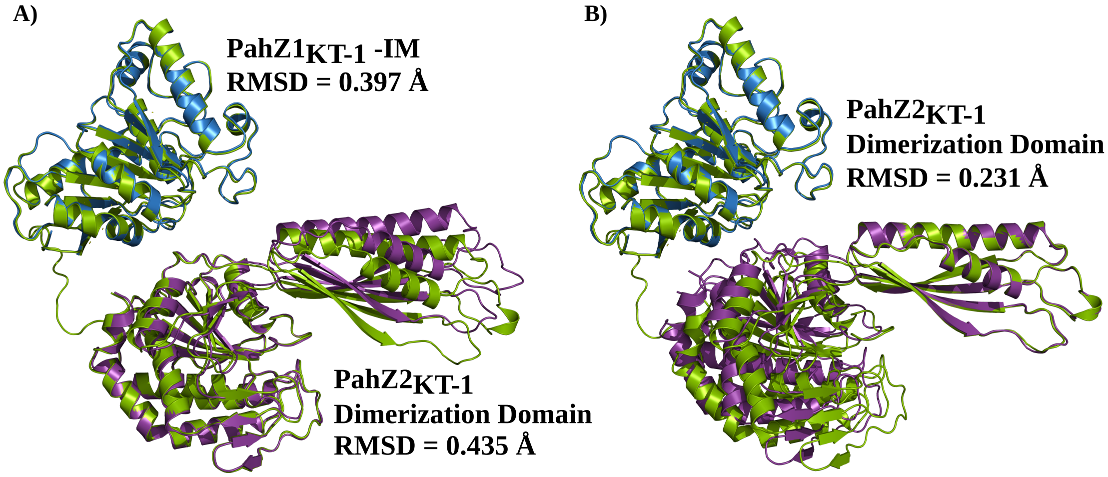
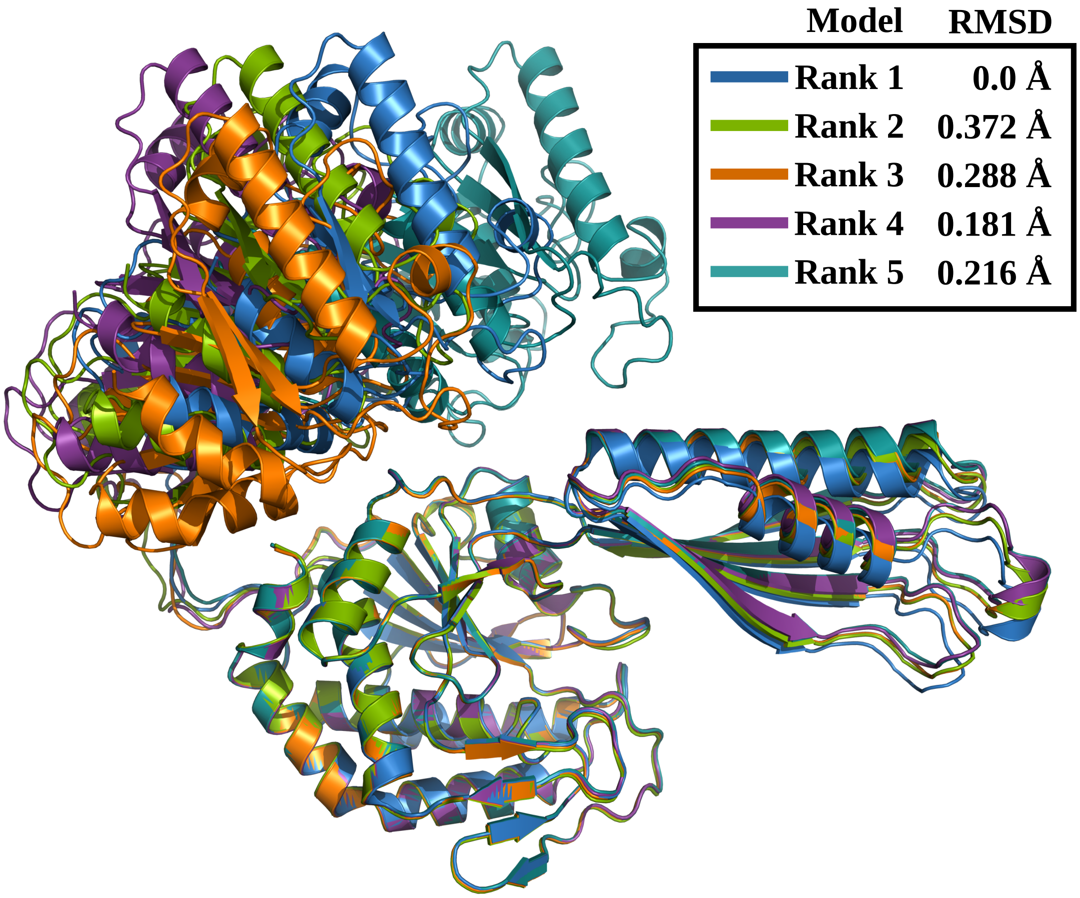


Figure C.1: Pairwise alignment of suPAH2:1 against experimental PahZ structures. All domains are colored such that PahZ1$_{\text{KT-1}}$ and PahZ2$_{\text{KT-1}}$ are blue and violet, respectively. suPAH2:1 is shown in green. **A:** The PahZ1$_{\text{KT-1}}$-IM domain of suPAH2:1 fits the PahZ1$_{\text{KT-1}}$ experimental structure with an RMSD = 0.397 Å, while the PahZ2$_{\text{KT-1}}$ catalytic domain achieves 0.435 Å. **B:** The same strategy presented above, illustrating here the alignment of suPAH2:1 against the PahZ2$_{\text{KT-1}}$ dimerization domain. An RMSD value of 0.231 Å is revealed.



Model	RMSD
Rank 1	0.0 Å
Rank 2	0.372 Å
Rank 3	0.288 Å
Rank 4	0.181 Å
Rank 5	0.216 Å

Figure C.2: Overview of ColabFold generated suPAH2:1 models ranked by global pLDDT confidence score. RMSD values calculated against rank 1 for each subsequent rank.

Appendix D

Statement of Candidate Contributions

The bulk of the findings presented previously represent a degree of scientific collaboration. For clarity, a summary of Chad A. Brambley's (C.A.B) contributions to these works is included by chapter as follows:

D.1 Chapter 1:

Multiple sequence alignments presented in Figure 1.2 and Figure 1.3 were assembled by C.A.B from publicly available data sets obtained via BLAST search results.[1]

D.2 Chapter 2:

C.A.B was the primary contributor of all in silico based experiments and protocols. This includes the generation of molecular dynamics (MD) equilibrated PahZ1$_{\text{KT-1}}$ interface mutants for subsequent PDBePISA[2] analysis shown in Figure 2.4E. All steered molecular dynamics (SMD) and computational determination of free energy of complex formation was performed by C.A.B in Gromacs[3–7] using the weighted histogram analysis method (WHAM).[8] These data were presented in Figure 2.5. The modelling and docking of all tPAA structures shown in Figure 2.6 was accomplished with Avogadro and Autodock Vina, respectively, with subsequent hydrogen bonding data collected in VMD.[9] All subsequent MD experiments carried out on PahZ1$_{\text{KT-1}}$ mutant constructs as shown in Figure 2.8 were performed by C.A.B.

D.3 Chapter 3:

All MD and computational measurements shown in Figure 3.4 was performed by C.A.B using software packages described above. This extends to all free energy calculations described in Figure 3.5. Modelling and docking of PAA$_5$ peptides to PahZ2$_{\text{KT-1}}$ was performed in Autodock Vina by C.A.B, including all subsequent MD calculations and analysis as shown in Figure 3.7. All NaCl-dependent free energy landscapes and principal component analysis shown in Figure 3.8 and Figure 3.10 were calculated and plotted by C.A.B from equilibrium MD trajectories.

D.4 Chapter 4:

C.A.B was the sole contributor of all modelling strategies and data set collections presented in chapter 4.

References

(1) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **1990**, *215*, 403–410.

(2) Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **2007**, *372*, 774–97.

(3) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, PMID: 26620784, 435–447.

(4) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. In *Solving Software Challenges for Exascale*, ed. by Markidis, S.; Laure, E., Springer International Publishing: Cham, 2015, pp 3–27.

(5) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(6) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(7) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.

(8) Grossfield, A. WHAM: the weighted histogram analysis method, en.

(9) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.