STRUCTURAL BASIS OF SUBSTRATE RECOGNITION AND INACTIVATION
IN PHOSPHATASES

_____

A Dissertation presented to the faculty of the Graduate School

at the University of Missouri-Columbia

_____

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

_____

by

Harkewal Singh

Dr. John J. Tanner, Dissertation Supervisor

December, 2011

The undersigned, appointed by the Dean of The Graduate School, have examined the dissertation entitled:

STRUCTURAL BASIS OF SUBSTRATE RECOGNITION AND INACTIVATION IN PHOSPHATASES

Presented by Harkewal Singh

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of acceptance.

_____

Professor John J. Tanner

_____

Professor Thomas J. Reilly

_____

Professor Lesa J. Beamer

_____

Professor Kent S. Gates

Acknowledgements

While I write this dissertation, I see that my scientific journey is reaching the first major milestone. However, this milestone is not just my sole effort. There have been several key people, whom I would like to acknowledge with all the humility.

This milestone of my life would not have been possible without the support of my thesis advisor Prof. John J. Tanner. He has tirelessly spent last four and a half years with me in this journey. In the spring of 2007, I joined the Tanner lab without any prior knowledge of biochemistry or X-ray crystallography. Prof. Tanner took his time out to teach me the essence of X-ray crystallography. He has always been around when I needed the help. If I was working late nights then he was too.

Prof. Tanner has answered my scientific questions on most of the occasions but also has taught me the way to find answers by guiding me to right directions. He always supported the ideas I had and amended them if needed. He strives for making his students independent thinkers and I think, he has succeeded very well in this endeavor. He taught me how to collect; process and most importantly interpret data. All this data collection lasted late nights and he was always right next-door to help, if I came across any hurdles. I know this for sure that my success in this lab is due to my advisor. I am a big fan of Dr. Tanner's writing style. He is one of the few people who know how to write well. At least in my little exposure to the world of writing scientific papers, he stands tall.

He not only motivated me by his scientific thinking or his writing style but also motivated me outside the lab. I used to hear from people that sometimes students are a reflection of their advisors and I hope that's true. I would like to assert here that I owe Prof. Tanner for his teachings. I hope our paths cross in future and I again get the chance to work with him. Prof. Tanner is a courageous advisor and I am fortunate that I did my Ph.D. under his guidance. I enjoyed the beer and sushi with him, those were some moments that I will always cherish throughout my life. I cannot thank him enough for his support and insight in deciding my postdoctoral advisor. Finally, I would also like to extend my gratitude to Prof. Tanner's family. I am sure their sacrifices are in part our successes.

This Ph.D. will not be even near completion if I had not met Prof. Thomas J. Reilly. I was introduced to Prof. Reilly, as one of Prof. Tanner's collaborator in 2007. But in the beginning of summer 2007, I started going to Reilly lab to learn the basics of molecular biology. Whatever little molecular biology techniques that I learnt are due to the intense work of Prof. Reilly. He taught me for the next few months and kept guiding me till date. Prof. Reilly's teaching style is immaculate, he taught with passion and style. I cannot forget the first day when I walked into his small laboratory. Without any greetings he asked, "who is Rodney Porter?" and as expected from a novice in the field, I said " I don't know". The next sentence from Dr. Reilly was " I can't believe you don't know him". As any self-respecting student, I went straight to the library and read about him so that I could claim my respect back. The very next day, before I could talk about Dr.

Porter, he asked me " Who was Barbara McClintock?" and again I said, " I don't know". This kept going on for a long time. Although at first what seemed like a challenge to my knowledge, later proved to be very helpful. It was a master plan by Dr. Reilly, to teach me the historical aspects of science and appreciate the depth behind each experiment I went on to do during my graduate career. He introduced me to overlap PCR, protein purification, enzyme kinetics and several other techniques, which I am sure I will use for the rest of my career. I cannot thank Prof. Reilly enough, for his support and help in my career. In my humble opinion, he is underrated and a walking encyclopedia of knowledge. I will miss his lab and yes I know who "Shutruk Nahhunte" was!

I would also like to thank my committee members, Prof. Kent S. Gates and Prof. Leas J. Beamer. We recently collaborated with Prof. Gates's group to work on PTP1B project and I hope that this collaboration lasts long. I am thankful to him for his insightful discussions on the PTP1B project and his confidence in me.

Prof. Beamer had been a great help in preparing me for my current postdoctoral position. Incidentally, she and I will now have the same postdoctoral mentor. I hope that I also do well in my new lab!

I would like to thank my parents Mr. Kulwant Singh and Mrs. Jaswinder Kaur for their love, support and confidence. My father always wanted me to be highly educated and I would proudly say that I have probably earned which one could call a terminal degree. My mother's teachings of honesty are inseparable part of my character.

I would also like to thank Thomas J. Malinski, Linda Malinski, Luke

Malinski and Matt Malinski for their unconditional support. They are an extended family to me in the USA. I can never forget their love and care. Thomas had the patience to teach me microbiology when I was taking a graduate level microbiology course. I would have probably flunked that exam if Thomas was not teaching me secretion systems in bacteria. I enjoyed our discussions on a variety of issues ranging from science to politics and religion. Thomas is one of the most humble men I met and I admire him and his lovely family that of course includes "toby". I hope that he finishes his dissertation soon with flying colors.

I would also like to thank Nisha Kaur. She had been a great source of motivation that made me a better person. I wish her all the success and happiness in her life.

I would also like to acknowledge Dr. Soma S. Marla. Dr. Marla, helped me tremendously in getting me to a graduate school. I sincerely thank him for his help.

Finally, I would like to extend my gratitude to the former members of Tanner lab who are Dr. R. L. Felts, Dr. T. A. White, Dr. J.P. Schuermann. Dr. J. Larson and Dr. Dhiraj Srivastava for good times in the lab. Current members of Tanner lab -Travis Pemberton and Min Luo and structural biology core personnel – Dale Karr and Dr. Ritcha Chaudhary. I would like to extend my special thanks to Richa Dhatwalia for being such a wonderful person. Richa has helped me immensely in preparing this thesis. In my opinion, she has all the qualities that need to be in a good crystallographer and I am confident that with her knowledge and effort that is on the horizon.

Harkewal Singh

TABLE OF CONTENTS

List of Figures

xiv

List of Tables

Abbreviations

Å	Angstrom
ACP	Acid phosphatase
AlkP	Alkaline phosphatase
ALS	Advanced Light Source
AMP	Adenosine monophosphate
AMPS	Adenosine 5´-phosphorothioate
APS	Advanced Photon Source
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BHI	Brain heart infusion broth
Bis-Tris	3-bis(tris(hydroxymethyl)methylamino)methane
BLAST	Basic Local Alignment Search Tool
CBAP	Class B acid phosphatases
CCAP	Class C acid phosphatase
CDC	Centers for Disease Control and Prevention
CPK	Corey, Pauling, Kultin color scheme
D261A	Mutant of FtHAP in which Asp261 is replaced by Ala

D64N   Mutant of rP4 in which Asp64 is replaced by Asn

D66N   Mutant of rP4 in which Asp66 is replaced by Asn

EDTA   Ethylenediaminetetraacetic acid

F23A   Mutant of FtHAP in which Phe23 is replaced by Ala

FtHAP   *Francisella tularensis* histidine acid phosphatase

FtCCAP   *Francisella tularensis* class C acid phosphatase

FPI   *Francisella tularensis* pathogenicity island

GST   Glutathione *S*-transferase

HAD   Haloacid dehalogenase

HAP   Histidine acid phosphatase

HEPES   4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

hPAP   Human prostatic acid phosphatase

HppA   Class C acid phosphatase from *Helicobacter pylori*

IPTG   Isopropylthiolgalactoside

K*av*   Partition coefficient

*K*m   Michaelis constant

LB   Luria-Bertani

LppC   Class C acid phosphatase from *Streptococcus equisimilis*

LVS   Live vaccine strain

MALDI-TOF	Matrix assisted laser desorption time-of- flight Mass spectrometry
MAP	Major acid phosphatase
MbCCAP	*Mycoplasma bovis* class C acid phosphatase
mdN	human mitochondrial deoxyribonucleotidase
MES	2-Morpholinoethanesulfonic acid
MOLREP	Molecular replacement
NADP+	Nicotinamideadenine dinucleotide phosphate
NADPH	Reduced nicotinamideadenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NCS	Non-crystallographic symmetry
NMN	Nicotinamide mononucleotide
NR	Nicotinamide riboside
NSAP	Bacterial Non-specific acid phosphatase
OD	Optical density
OlpA	Class C acid phosphatase from *Chryseobacterium meningoseptica*
PEG	Polyethylene glycol
PEGMME	Polyethylene glycol mono-methyl ether

PCR	Polymerase chain reaction
P_i	Inorganic phosphate
PISA	Protein Interactions, Surfaces and Assemblies
PLC	Phospholipase C
PlcH	Hemolytic phospholipase C
*p*NP	*p*-nitrophenol
*p*NPP	*p*-nitrophenyl phosphate
*p*NPPC	*p*-nitrophenyl phosphorylcholine
PDB	Protein data bank
PTP	Protein tyrosine phosphatase
PVDF	Polyvinylidene fluoride
rPAP	Rat prostatic acid phosphatase
rP4	Recombinant *Haemophilus influenzae e* (P4) class C acid phosphatase
SAD	Single wavelength anomalous diffraction
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
TEVP	Tobacco etch virus protease
TLS	Translation libration screw-motion
Tris	Tris(hydroxymethyl) methylamine

Vmax   Maximal velocity

Y135A   Mutant of FtHAP in which Tyr135 is replaced by Ala

STRUCTURAL BASIS OF SUBSTRATE RECOGNITION AND INACTIVATION IN PHOSPHATASES

Harkewal Singh

Prof. John J. Tanner, Dissertation Supervisor

ABSTRACT

Phosphatases are ubiquitous enzymes that catalyze the transfer of phosphoryl group to water. In addition to being one of the most important and fundamental reactions in biological systems, phosphatases are critical to cellular processes such as metabolism, energy balance, nucleotide pool regulation, signal transduction pathways, and sequestration of inorganic phosphate. In bacteria, phosphatases are not only implicated in virulence but also are important vaccine candidates. In humans, phosphatases play crucial roles in disease and metabolism. For instance, human prostatic acid phosphatase has been shown to suppress pain by generating adenosine from extracellular adenosine 5' monophosphate whereas protein tyrosine phosphatase 1B is a drug target against type II diabetes. The larger context of the research described in this dissertation addresses the structural basis of substrate preferences and recognition by *Francisella tularensis* histidine acid phosphatase, class C prototype acid phosphatases from *Haemophilus influenzae*, *Pasteurella multocida* acid phosphatase, *Mycoplasma bovis* acid phosphatase and structural basis of H_2O_2/$KHCO_3$ mediated inactivation of PTP1B. Finally, this thesis addresses the structural context of oligomerization in class C acid phosphatases using small angle x-ray scattering and analytical ultracentrifugation methods.

1. Introduction

Phosphatases are the enzymes that catalyze the hydrolysis of phosphomonoesters by transferring the phosphoryl group to water[1]. This enzyme-catalyzed process is not only ubiquitous to metabolism but also central to the energy balance of all. Events like maintenance of transmembrane chemical potential, synthesis of metabolites of higher energy, regulation of nucleotides, signal transduction pathways, and sequestration of inorganic phosphate in the cell all involve the displacement of phosphorous atom from a phosphomonoester or anhydride.

In bacteria, there are several phosphatases that can deliver the product of phosphomonoester hydrolysis. These enzymes are crucial to the physiology of the bacterial cell. For example, molecules such as nucleotides, sugar phosphates, phytic acid cannot be transported across the cytoplasmic membrane by itself. However, phosphatases play an important role in essentially scavenging inorganic phosphate and other byproducts from these molecules that can be transported across the membrane in the form of essential nutrients to the cell.

Recent studies have also suggested the role of phosphatases in the virulence of *Francisella tularensis,* a category A bioterrorism agent. Studies from *Mohapatra et al.* suggest that acid phosphatases are essential for intramacrophage survival and virulence of *F. tularensis*[2, 3]. The authors suggest that four acid phosphatase genes, identified in highly virulent type A strains (*acpA*, *acpB*, *acpC*, and *hap)* are implicated in the virulence of *F. tularensis*.

Real-time PCR analysis of gene expression in macrophages infected with *F. tularensis* subsp. *novicida* showed that the *acpA* and *hap* genes are induced 219- and 10-fold, respectively, upon infection, indicating that the intracellular environment activates these genes[3]. Furthermore, a triple mutant strain of *F. novicida* lacking functional *acpA*, *acpC*, and *hap* genes is defective for survival inside macrophages and exhibited a marked defect in time to death in a mouse model of tularemia[3]. Moreover, mice infected with a quadruple mutant lacking all four acid phosphatase genes showed 100% survival at nine weeks after infection, and this mutant showed promising protective capacity as a single-dose live vaccine.

Another phosphatase, *e(*P4)[4, 5] from a common human pathogen, *Haemophilus influenzae*, is a major component of the outer membrane and is highly conserved among *H. influenzae* strains. Studies have shown that recombinant P4 (rP4) and rP4 mutant enzymes are highly immunogenic, anti-rP4 antibodies exhibit bactericidal activity, and immunization of mice with rP4 reduces nasal colonization of nontypeable *H. influenzae* strains[6, 7].

In humans, phosphatases have been linked to several diseases. Historically, human prostatic acid phosphatases (hPAP), has been implicated in prostate cancer[8] but recent studies[9, 10] have shown that hPAP suppresses pain by dephosphorylating extracellular adenosine monophosphate (5′ AMP). Another eukaryotic phosphatase PTP1B is an important drug target against diabetes mellitus[11] and is also involved in signaling due to its inactivation by intracellular H_2O_2[12].

Finally, 5′-nucleotidases represent a class of specific mammalian phosphatases that dephosphrylate non-cyclic nucleoside monophosphates to nucleosides and inorganic phosphate[13] Human mitochondrial deoxyribonucleotidase (mdN[14]), a 5′-nucleotidase, catalyzes the hydrolysis of thymidine and deoxy-uridine monophosphates and regulates the deoxythymidine triphosphate (dTTP) pool in the mitochondria that is involved in the *in vivo* synthesis of DNA[14] Furthermore, these mammalian 5′-nucleotidases regulate the nucleotide pool by reversing the phosphorylation process of nucleosides by nucleoside kinases[13].

In conjunction with cytosolic and mitochondrial thymidine kinases, cytosolic deoxyribonucleotidase (cdN) and mdN are known to control the phosphorylation of thymidine and deoxyuridine[15-18]. These enzymes have also been implicated in decreasing the efficacy of antiviral and antiblastic nucleoside analogs. These drugs must be activated by phosphorylation to exert their therapeutic effect, and intracellular 5′-nucleotidases can reverse the activation step and thus decrease efficacy. For this reason, there is an interest in designing inhibitors of 5′-nucleotidases for co-therapy.

Due to such diverse functions that range from sequestration of inorganic phosphate in bacteria and pain suppressive abilities in mice, phosphatases are clearly a family of enzymes with varied biochemical and metabolic functions via a common enzymatic path – the hydrolysis of phosphomonoesters. Substrates of phosphatases vary from small molecules (e.g. *p*NPP, glucose-6-phosphate), nucleoside monophosphates and finally to large molecules such as

phosphorylated proteins. Therefore, it is useful to have a classification of these enzymes based upon their substrate choices and ultimately structural motifs.

1.1 Classification of phosphatases

1.1.1 Nonspecific phosphatases

The primary basis of this classification is based upon the broad substrate choices and pH optimum exhibited by these enzymes. The enzymes classified under this category hydrolyze a variety of phosphomonoesters. Bacterial nonspecific acid phosphatases (NSAPs) and enzymes belonging to histidine phosphatase superfamily are the major constituents of this category. It is noteworthy that there are no sequence similarities between NSAPs and histidine phosphatase superfamily members. Both families have distinct structural motifs and reaction centers for the catalysis of various phosphomonoesters.

1.1.2 Specific phosphatases

Specific phosphatases can be further categorized into different subclasses depending upon their substrate choices. For example, *phosphoprotein specific phosphatases* that hydrolyze phosphoprotein or phosphopeptide substrates represent one subclass. Some major examples belonging to this class include dual specificity phosphatases, phospho Ser/Thr two metal ion phosphatases and

protein tyrosine phosphatases. One such example from this sub-class, protein tyrosine phosphatases1B (PTP1B), will be discussed in detail in chapter 7.

A second convenient subclass is *small molecule specific phosphatases* that prefer substrates such as glucose-6-phosphate, fructose-1,6 bisphosphate, inositol mono and polyphosphate and phosphatidic acid.

1.2 Nonspecific acid phosphatases (NSAPs)

Thaller et al.[20] originally used the term NSAP to describe bacterial enzymes that were able to show optimal activity at neutral to acidic pH and also promiscuous with respect to the substrate choices.

Bacterial NSAPs are secreted phosphohydrolases produced as membrane bound lipoproteins or soluble periplasmic proteins[21] They exhibit broad substrate specificity and show optimal activity at acidic to neutral pH. Based upon their subcellular localization, amino acid sequence and biophysical characteristics, they are further classified into class A, class B and class C NSAPs. Some salient features and characteristic of each class are described below.

1.2.1 Class A nonspecific acid phosphatases

Class A NSAPs (CAAPs) are secreted phosphohydrolases composed of a polypeptide component of 25 - 27 kDa and possess a   conserved sequence motif. They are homodimeric or homotetrameric[21] in solution and their activity is

unaffected by EDTA and divalent cations. The enzymes belonging to this class catalyze the hydrolysis of a variety of substrates such as 3′ and 5′ nucleoside monophosphates, synthetic aryl phosphomonoesters such as p-nitro phenyl phosphate (pNPP) and β– glycerophosphate. In addition to the phosphomonoesters, these enzymes also hydrolyze nucleoside diphosphates, nucleoside triphosphates, hexose and pentose phosphates. The signature motif of CAPPs is **K-X-R-P-X-P-S-G-H-X-S-R-X-H-X-D.** Some examples of class A NSAPs include PhoC of *M. morganii* and PhoN of *P. stuaritii* and *S. enterica ser. typhimurium*.

1.2.2 Class B nonspecific acid phosphatases

Class B NSAPs (CBAPs) are also secreted phosphohydrolases composed of 25 kDa subunits. Unlike class A NSAPs, EDTA inhibits these enzymes. Members of this class belong to the DDDD molecular superfamily and have a bipartite signature motif **F-D-I-D-D-T-V-L-F-S-S-P** located in the N-terminal part of the chain, and **Y-G-D- [AS]-D-X-D- [IV]** located near the C-terminus. CBAPs are homotetramers in solution[21]. Examples of CBAPs include NapA-Mm enzyme from *M. morganii*, NapA from *H. influenzae*, AphA from *E. coli* and AphA from *S. enterica. E. Coli* AphA exhibits strong preference for 5′ and 3′ mononucleotides[22] Structural work for the *E.coli* AphA has shown that the CBAPs have four aspartates in the active site, which are essential for catalysis. Additionally, these

enzymes have structural features that are consistent with haloacid dehalogenase family members[22].

1.2.3 Class C nonspecific acid phosphatases

Class C nonspecific acid phosphatases (CCAPs) are bacterial outer-membrane lipoproteins and form homodimers[23] in solution as compared to homotetramers in CBAPs. Enzymes belonging to this class are inhibited by EDTA, molybdate and orthovanadate[21]. This class includes enzymes having 25 - 30 kDa polypeptide subunit and requires Mg^2 for catalysis. CCAPs are characterized by a similar DDDD motif as CBAPs, which is essential for catalytic activity. These residues are located in a bipartite signature motif of [IV]-[VAL]-**D**-[IL]-**D**-E-T-[VM]-L-X-[NT]-X-X-Y near the N- terminus and [IV]-[LM]-X-X-G-**D**-[NT]-L-X-**D**-F near the C-terminus[20]. Note that the four aspartic acid residues are highlighted. CCAPs from *H. influenzae* (rP4), *Fransicella tularensis*, *Bacillus anthracis*, *C. meningosepticum*, *H. pylori*, *S. equisimilis, Pasteruella multocida* (PmCCAP) and *C. perfringens* have been identified and characterized to various degrees. CCAPs have been found to hydrolyze a range of substrates including 5' and 3' nucleoside monophosphate substrates and aryl phosphates like 4-MUP and *p*NPP. For example, studies on one CCAP from *Chryseobacterium meningosepticum* (OlpA) showed that it has 10 to 50 fold preference for 5' mononucleotides over 3' mononucleotides[24]. Our structural work on rP4[25, 26] and PmCCAP[23] shows that the catalytic scaffold of these enzymes have four invariant

Asp (D) residues clustered around a Mg^{2+} ion. The enzymes have an α/β core domain and also a shorter α domain. The core domain features five β strands flanked by α helices that is a reminiscent of the haloacid dehalogenase superfamily of enzymes. Furthermore, our biochemical data[23, 26] reveals that these enzymes are able to hydrolyze 5', 3', 2' nucleoside monophosphates and synthetic substrates such as pNPP, with slight preferences for 5' nucleoside monophosphate over 3' nucleoside monophosphate.

1.3 Histidine phosphatase superfamily

The histidine phosphatase superfamily includes enzymes that are functionally and biochemically diverse. Recent reviews suggest that enzymes belonging to the histidine phosphatase superfamily are dominated by phosphatases but the superfamily is also known to have diverse members such as co-factor dependent phospho glycerate mutase (dPGM)[27]. For simplicity, the histidine phosphatase superfamily can be divided into two major branches (Fig. 1). Branch one is the larger subfamily mainly consisted of dPGM, fructose-2,6-bisphosphatase, PhoE, SixA, TIGAR [TP53 (tumor protein 53)-induced glycolysis and apoptosis regulator], and Sts-1. Whereas, branch two represents few enzymes consisting of acid phosphatases and phytases. Note that these two sub-branches of histidine phosphatase superfamily share very little sequence similarity. Histidine acid phosphatase or HAP belongs to the second smaller class of histidine phosphatases.
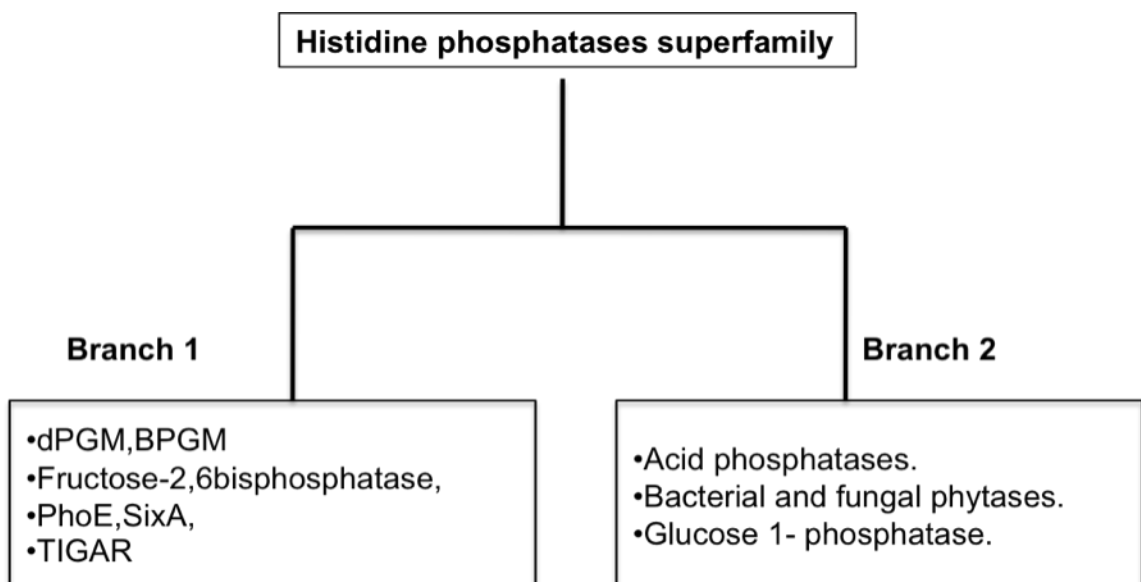
Fig. 1.1 Classification scheme of the histidine phosphatase superfamily.

One of the major focuses of this dissertation is the structural and biochemical characterization of histidine acid phosphatase from *Francisella tularensis* (FtHAP), therefore, the next section will be mainly focused on defining the biochemical, structural and biological features of HAPs.

1.3.1 Histidine acid phosphatases

Histidine acid phosphatases (HAPs) refer to enzymes that utilize an active site histidine residue for catalysis. As the name suggests, these enzymes also catalyze the transfer of phosphoryl group to water at acidic pH. Members of this family share a conserved catalytic core exhibiting a common sequence motif as R**H**GXRXP. The second residue in the motif is the catalytic histidine that is

phosphorylated during the enzymatic reaction. Interestingly the arginines in the motif are located in/near the active site and are presumed to enhance the binding of negatively charged phosphoryl moiety of the substrate via a set of electrostatic interactions[28, 29]. Note that, unlike HAD superfamily members CCAPs, enzymes belonging to this superfamily do not require the metal ion for activity.

Structural and biochemical work done by our and other labs suggest the presence of another histidine in the active site. Interestingly, this histidine is not the part of the conserved sequence motif. Adjacent to this second histidine, HAPs also have an aspartate residue that is believed to behave as general acid/base during the phosphomonoester hydrolysis[30, 31]. Note, that a similar aspartate (Asp 181) is also present in PTP1B that has been shown to behave as general acid/base. It is worthwhile mentioning that these two enzymes share no sequence similarity.

HAPs are present in both bacteria and mammals. These enzymes perform a variety of functions based upon their location and source organism. For example, in pathogenic bacteria, *F. tularensis*, *hap* gene product is implicated in the survival of the bacteria[3].

A recent study shows that human hisitidine acid phosphatase a.k.a. hPAP, dephosphorylates the extracellular adenosine 5' monophosphate generating adenosine. The adenosine thus generated is taken up by A_1R receptors (Fig. 2) in the dorsal spinal cord and relieves pain[9]. Researchers associated with the study, suggest that human PAP thus used is eight times more potent than the commonly used analgesic morphine and that the effect of hPAP lasted three

days post injection[9]. Further, their pain suppression assay suggested that hPAP has anti-allodynic and anti-nociceptive properties[9].

In another study[10] by the same group, these authors suggest that a single intraspinal injection of recombinant mouse prostatic acid phosphatase (mPAP) had similar antihyperalgesic and antiallodynic effects as tested in their inflammatory pain model. These antinociceptive effects were transiently blocked by the A_1R antagonist 8-cyclopentyl-1,3-dipropylxanthine (CPX), suggesting mPAP dephosphorylates nucleotides to adenosine to mediate antinociception just like human and bovine PAP[10]. Thus, different adenosine receptors are potential targets for treating chronic pain in humans either by using agonists or engineered ectonucleotidases such as hPAP.
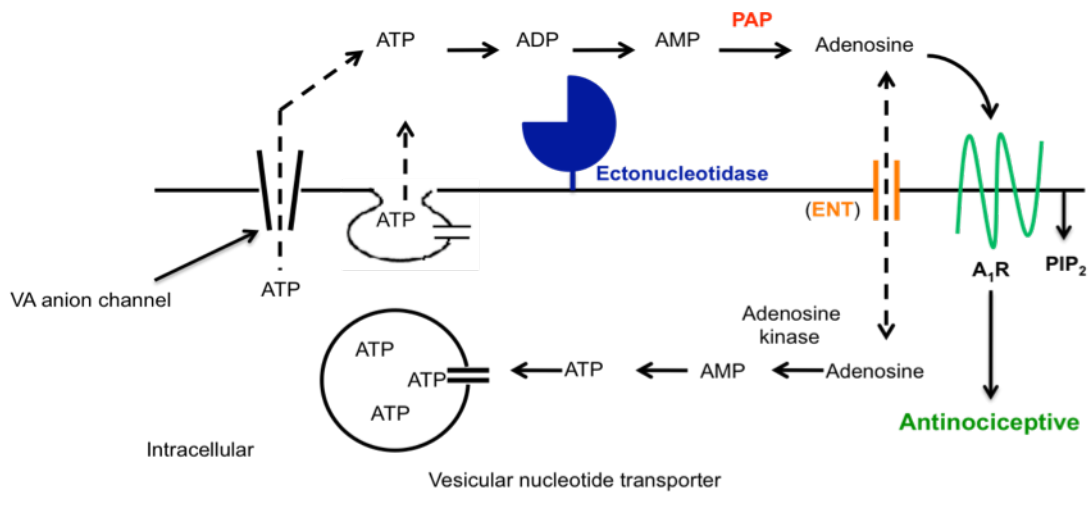
Fig. 1.2 –Schematic diagram of pain suppression via A_1R activation[32].

1.4 References

1.   Vincent, J. B.; Crowder, M. W.; Averill, B. A., Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions. *Trends Biochem Sci* **1992,** *17* (3), 105-10.

2.   Mohapatra, N. P.; Balagopal, A.; Soni, S.; Schlesinger, L. S.; Gunn, J. S., AcpA is a Francisella acid phosphatase that affects intramacrophage survival and virulence. *Infect Immun* **2007,** *75* (1), 390-6.

3.   Mohapatra, N. P.; Soni, S.; Reilly, T. J.; Liu, J.; Klose, K. E.; Gunn, J. S., Combined deletion of four Francisella novicida acid phosphatases attenuates virulence and macrophage vacuolar escape. *Infect Immun* **2008,** *76* (8), 3690-9.

4.   Reilly, T. J.; Chance, D. L.; Smith, A. L., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J Bacteriol* **1999,** *181* (21), 6797-805.

5.   Reilly, T. J.; Smith, A. L., Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr Purif* **1999,** *17* (3), 401-9.

6.   Green, B. A.; Baranyi, E.; Reilly, T. J.; Smith, A. L.; Zlotnick, G. W., Certain site-directed, nonenzymatically active mutants of the Haemophilus influenzae P4 lipoprotein are able to elicit bactericidal antibodies. *Infect Immun* **2005,** *73* (7), 4454-7.

7.   Hotomi, M.; Ikeda, Y.; Suzumoto, M.; Yamauchi, K.; Green, B. A.; Zlotnick, G.; Billal, D. S.; Shimada, J.; Fujihara, K.; Yamanaka, N., A recombinant P4 protein of Haemophilus influenzae induces specific immune responses biologically active against nasopharyngeal colonization in mice after intranasal immunization. *Vaccine* **2005,** *23* (10), 1294-300.

8.   Gutman, A. B.; Gutman, E. B., An " Acid " Phosphatase Occurring in the Serum of Patients with Metastasizing Carcinoma of the Prostate Gland. *J Clin Invest* **1938,** *17* (4), 473-8.

9.   Zylka, M. J.; Sowa, N. A.; Taylor-Blake, B.; Twomey, M. A.; Herrala, A.; Voikar, V.; Vihko, P., Prostatic acid phosphatase is an ectonucleotidase and suppresses pain by generating adenosine. *Neuron* **2008,** *60* (1), 111-22.

10.   Sowa, N. A.; Street, S. E.; Vihko, P.; Zylka, M. J., Prostatic acid phosphatase reduces thermal sensitivity and chronic pain sensitization by depleting phosphatidylinositol 4,5-bisphosphate. *J Neurosci* **2010,** *30* (31), 10282-93.

11. Patti, M. E.; Kahn, C. R., The insulin receptor--a critical link in glucose homeostasis and insulin action. *J Basic Clin Physiol Pharmacol* **1998,** *9* (2-4), 89-109.

12. Rhee, S. G., Cell signaling. H2O2, a necessary evil for cell signaling. *Science* **2006,** *312* (5782), 1882-3.

13. Bianchi, V.; Spychala, J., Mammalian 5'-nucleotidases. *J Biol Chem* **2003,** *278* (47), 46195-8.

14. Wallden, K.; Ruzzenente, B.; Rinaldo-Matthis, A.; Bianchi, V.; Nordlund, P., Structural basis for substrate specificity of the human mitochondrial deoxyribonucleotidase. *Structure* **2005,** *13* (7), 1081-8.

15. Rampazzo, C.; Ferraro, P.; Pontarin, G.; Fabris, S.; Reichard, P.; Bianchi, V., Mitochondrial deoxyribonucleotides, pool sizes, synthesis, and regulation. *J Biol Chem* **2004,** *279* (17), 17019-26.

16. Bradshaw, H. D., Jr.; Deininger, P. L., Human thymidine kinase gene: molecular cloning and nucleotide sequence of a cDNA expressible in mammalian cells. *Mol Cell Biol* **1984,** *4* (11), 2316-20.

17. Gazziola, C.; Ferraro, P.; Moras, M.; Reichard, P.; Bianchi, V., Cytosolic high K(m) 5'-nucleotidase and 5'(3')-deoxyribonucleotidase in substrate cycles involved in nucleotide metabolism. *J Biol Chem* **2001,** *276* (9), 6185-90.

18. Johansson, M.; Karlsson, A., Cloning of the cDNA and chromosome localization of the gene for human thymidine kinase 2. *J Biol Chem* **1997,** *272* (13), 8454-8.

19. Taylor, W. P.; Widlanski, T. S., Charged with meaning: the structure and mechanism of phosphoprotein phosphatases. *Chem Biol* **1995,** *2* (11), 713-8.

20. Thaller, M. C.; Schippa, S.; Rossolini, G. M., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci* **1998,** *7* (7), 1647-52.

21. Rossolini, G. M.; Schippa, S.; Riccio, M. L.; Berlutti, F.; Macaskie, L. E.; Thaller, M. C., Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. *Cell Mol Life Sci* **1998,** *54* (8), 833-50.

22. Calderone, V.; Forleo, C.; Benvenuti, M.; Thaller, M. C.; Rossolini, G. M.; Mangani, S., A structure-based proposal for the catalytic mechanism of the bacterial acid phosphatase AphA belonging to the DDDD superfamily of phosphohydrolases. *J Mol Biol* **2006,** *355* (4), 708-21.

23.   Singh, H.; Malinski, T. J.; Reilly, T. J.; Henzl, M. T.; Tanner, J. J., Crystal structure and immunogenicity of the class C acid phosphatase from Pasteurella multocida. *Arch Biochem Biophys* **2011,** *509* (1), 76-81.

24.   Passariello, C.; Schippa, S.; Iori, P.; Berlutti, F.; Thaller, M. C.; Rossolini, G. M., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5'-nucleotides. *Biochim Biophys Acta* **2003,** *1648* (1-2), 203-9.

25.   Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of recombinant Haemophilus influenzae e (P4) acid phosphatase reveals a new member of the haloacid dehalogenase superfamily. *Biochemistry* **2007,** *46* (39), 11110-9.

26.   Singh, H.; Schuermann, J. P.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Recognition of nucleoside monophosphate substrates by Haemophilus influenzae class C acid phosphatase. *J Mol Biol* **2010,** *404* (4), 639-49.

27.   Rigden, D. J., The histidine phosphatase superfamily: structure and function. *Biochem J* **2008,** *409* (2), 333-48.

28.   Schneider, G.; Lindqvist, Y.; Vihko, P., Three-dimensional structure of rat acid phosphatase. *EMBO J* **1993,** *12* (7), 2609-15.

29.   Saini, M. S.; van Etten, R. L., Dimeric nature and amino acid compositions of homogeneous canine prostatic human liver and rat liver acid phosphatase isoenzymes. Specificity and pH-dependence of the canine enzyme. *Biochim Biophys Acta* **1978,** *526* (2), 468-78.

30.   Porvari, K. S.; Herrala, A. M.; Kurkela, R. M.; Taavitsainen, P. A.; Lindqvist, Y.; Schneider, G.; Vihko, P. T., Site-directed mutagenesis of prostatic acid phosphatase. Catalytically important aspartic acid 258, substrate specificity, and oligomerization. *J Biol Chem* **1994,** *269* (36), 22642-6.

31.   Saini, M. S.; Van Etten, R. L., An essential carboxylic acid group in human prostate acid phosphatase. *Biochim Biophys Acta* **1979,** *568* (2), 370-6.

32.   Zylka, M. J., Pain-relieving prospects for adenosine receptors and ectonucleotidases. *Trends Mol Med* **2011,** *17* (4), 188-96.

Chapter 2.

Crystal Structures of the Histidine Acid Phosphatase from Francisella tularensis Provide Insight into Substrate Recognition

Harkewal Singh, Richard L. Felts, Jonathan P. Schuermann, Thomas J. Reilly,

and John J. Tanner

Author contribution – H.S. expressed purified, crystallized wild type FtHAP and D261A mutant, determined crystal structures of FtHAP-Pi, D261A-3'AMP complex, refined native and inhibitor bound structures.

Abstract

Histidine acid phosphatases catalyze the transfer of a phosphoryl group from phosphomonoesters to water at acidic pH using an active site histidine. The histidine acid phosphatase from the category A pathogen *Francisella tularensis* (FtHAP) has been implicated in intramacrophage survival and virulence, motivating interest in understanding the structure and mechanism of this enzyme. Here we report a structure-based study of ligand recognition by FtHAP. The 1.70 Å resolution structure of FtHAP complexed with the competitive inhibitor L(+)-tartrate was solved using single-wavelength anomalous phasing. Structures of the ligand-free enzyme and the complex with inorganic phosphate were determined at resolutions of 1.85 Å and 1.70 Å, respectively. To gain insight into substrate recognition, the structure of the Asp261Ala mutant enzyme complexed with the substrate 3'-AMP was determined at 1.50 Å resolution. FtHAP exhibits a two-domain fold similar to that of human prostatic acid phosphatase, consisting of an α/β core domain and a smaller domain that caps the core domain. The structures show that the core domain supplies the phosphoryl binding site, catalytic histidine (His17), and an aspartic acid residue (Asp261) that protonates the leaving group, while the cap domain contributes residues that enforce substrate preference. FtHAP and human prostatic acid phosphatase differ in the orientation of the crucial first helix of the cap domain, implying differences in the substrate preferences of the two enzymes. 3'-AMP binds in one end of a 15-Å long tunnel, with the adenine clamped between Phe23 and Tyr135, and the

ribose 2'-hydroxyl interacting with Gln132. The importance of the clamp is confirmed with site-directed mutagenesis; mutation of Phe23 and Tyr135 individually to Ala increases K_m by factors of seven and ten, respectively. The structural data are consistent with a role for FtHAP in scavenging phosphate from small molecules present in host macrophage cells.

2.1 Introduction

Francisella tularensis is a highly infectious facultative intracellular bacterial pathogen and the causative agent of the potentially fatal disease tularemia.[1] Due to its high infectivity and lethality, and the potential for use as a bioterrorism agent, the Centers for Disease Control and Prevention consider *F. tularensis* to be a category A agent. Currently, there is no licensed vaccine available, and new antimicrobial agents are needed in anticipation of strains engineered to have enhanced virulence.

Acid phosphatases have been shown to be essential for intramacrophage survival and virulence of *F. tularensis*.[2] Four acid phosphatase genes have been identified in highly virulent type A strains: *acpA*, *acpB*, *acpC*, and *hap* (encodes the enzyme studied here). AcpA is the prototype of a unique superfamily of acid phosphatases and phospholipases C.[3] AcpB and AcpC[4] belong to the DDDD superfamily of phosphohydrolases.[5] The *hap* gene product is unrelated to the other three phosphatases. Real-time PCR analysis of gene expression in macrophages infected with *F. tularensis* subsp. *novicida* showed that the *acpA* and *hap* genes are induced 219- and 10-fold, respectively, upon infection, indicating that the intracellular environment activates these genes.[2b] Furthermore, a triple mutant strain of *F. novicida* lacking functional *acpA*, *acpC*, and *hap* genes is defective for survival inside macrophages and exhibited a marked defect in time to death in a mouse model of tularemia.[2b] Moreover, mice infected with a quadruple mutant lacking all four acid phosphatase genes showed

100% survival at nine weeks after infection, and this mutant showed promising protective capacity as a single-dose live vaccine.[2b]

The demonstration that acid phosphatase activity is essential for intracellular survival and virulence of *F. tularensis* has generated new interest in understanding the structures and molecular functions of the four acid phosphatases. The enzymes encoded by *acpA*,[6] *acpC*,[4] and *hap*[7] have been crystallized, and the structure of AcpA has been determined.[3] Also, kinetic and biophysical studies of AcpA have been reported.[8]

Herein we present a structure-based study of the *hap* gene product (FtHAP) focused on substrate recognition. FtHAP is a 37 kDa periplasmic acid phosphatase that contains the characteristic histidine acid phosphatase (HAP) sequence motif of RHGXRXP, which was first identified in human prostatic acid phosphatase (hPAP) by Van Etten and coworkers.[9] Crystal structures of hPAP[10] and the rat homolog[11] (rPAP) are known, but heretofore the structure of a bacterial HAP has not been reported. In the following, we present four high-resolution crystal structures of FtHAP (Table 2.1), including structures of the enzyme complexed with the product inorganic phosphate (P_i) and the competitive inhibitor L(+)-tartrate, as well as the first structure of a HAP/substrate complex. The data provide insight into the structural determinants of substrate recognition by FtHAP and other HAPs.

2.2 Materials and Methods

2.2.1 Crystallization of native, Se-Met, and mutant FtHAP

Native FtHAP was expressed, purified, and crystallized as described previously.[7] Briefly, tetragonal crystals having unit cell dimensions of a = 62 Å, c = 211 Å were grown in sitting drops using a reservoir solution of 10% (w/v) Tacsimate, 0.1 M HEPES pH 7.0, and 19% (w/v) PEG 3350.[7] The active site of the crystallized enzyme contains a bound L(+)-tartrate ion (*vide infra*), which is a known competitive inhibitor of HAPs.[15a, 22] The source of the inhibitor is Tacsimate, which contains 0.16 M ammonium tartrate.

The Se-Met derivative was created with the Overnight Express Autoinduction System 2 (Novagen). The purification scheme was similar to the one used for native FtHAP. The Se-Met enzyme was crystallized in the Tacsimate/PEG condition with 5 mM dithiothreitol added to the reservoir solution. The resulting crystals are isomorphic to the native ones. Crystals of Se-Met FtHAP were cryoprotected using 25 % PEG 200 as described for native FtHAP.[7]

Crystals of the inhibitor-free enzyme were grown using an ammonium sulfate-based recipe that had been identified previously during crystallization screening trials.[7] These crystals were grown in sitting drops at room temperature using a protein stock solution of 10 mg/mL enzyme in 50 mM sodium acetate pH 6.0, and reservoir solutions consisting of 0.05 - 0.2 M Bis-Tris buffer pH 5.0, and 1.6 - 2.0 M ammonium sulfate. The crystals are isomorphic to those grown in Tacsimate and PEG.

Crystals used to obtain the complex with P_i were grown using a variation of the PEG/Tacsimate recipe in which Tacsimate was omitted. Large tetragonal crystals were grown in sitting drops at room temperature using reservoir solutions of 0.05- 0.20 M Bis-Tris pH 5.0 - 6.5 and 17 - 25 % PEG 3350. The P_i complex was formed by soaking a crystal for 30 minutes in 21 % (w/v) PEG 3350, 0.1 M Bis-Tris pH 6.0, 25% (v/v) PEG 200, and 20 mM P_i (from a stock solution of 100 mM KH_2PO_4 adjusted to pH 6.0). The crystal was then picked up with a Hampton loop and plunged into liquid nitrogen.

Crystals of a substrate-trapping mutant enzyme in which Asp261 is replaced by Ala (D261A) were grown in sitting drops at room temperature using a variation of the ammonium sulfate-based recipe described above. In this case, the protein stock solution consisted of 15 mg/mL D261A in a buffer of 50 mM sodium acetate at pH 6.0, and the reservoir contained 0.1 M Bis-Tris buffer at pH 6.5 and 2.0 M ammonium sulfate. The complex with the substrate 3'-AMP was formed by soaking a D261A crystal for 45 minutes in 20 mM 3'-AMP, 2 M ammonium sulfate, 0.1 M Bis-Tris pH 6.5, and 20 % (v/v) glycerol prior to freeze-trapping in liquid nitrogen.

2.2.2 X-ray diffraction data collection

Crystals of Se-Met FtHAP were analyzed at beamline 4.2.2 of the Advanced Light Source using a NOIR-1 detector, and the data were processed with D*TREK.[23] The data set that was used for single-wavelength anomalous

diffraction (SAD) phasing was collected at the energy corresponding to the peak of f'' (λ = 0.97915 Å). This data set consisted of 360 frames with an oscillation width of 0.5° per image, detector distance of 150 mm, and exposure time of 3.5 s/image. The data were processed to 2.10 Å resolution (Table 2.1).

Data from crystals of the native enzyme were also collected at beamline 4.2.2 and processed with D*TREK (Table 2.1). The 1.70 Å resolution data set for the L(+)-tartrate complex was collected using a detector distance of 150 mm and detector offset angle of 17°. A total of 180° of data were collected using an oscillation angle of 0.5° and an exposure time of 8 s/image. The 1.70 Å resolution data set for the P_i complex consisted of 180° of data collected with the detector distance of 150 mm, offset angle of 10°, oscillation angle of 0.5°, and exposure time of 3 s/image. The data set used to refine the structure of the inhibitor-free enzyme consisted of 360 images acquired with an oscillation width of 0.5° per image, detector distance of 160 mm, detector offset angle of 10°, and exposure time of 3.5 s/image.

A 1.50 Å resolution data set was collected from a crystal of the D261A/3'-AMP complex at NE-CAT beamline 24-ID-C at the Advanced Photon Source using a Quantum 315 detector. The data set consisted of 180 images collected with detector distance of 170 mm and an oscillation width of 1°. The data were processed with HKL2000.[24]

Table 2.1. Data collection and refinement statistics[a]

enzyme	Se-Met	native	native	native	D261A
active site ligand	L(+)-tartrate	phosphate	L(+)-tartrate	none	3'-AMP
space group	$P4_12_12$	$P4_1$	$P4_1$	$P4_1$	$P4_1$
unit cell lengths (Å)	a = 61.9,	a = 61.9	a = 62.0	a = 62.4	a = 61.6
	c = 210.9	c = 211.6	c = 210.9	c = 209.9	c = 211.4
wavelength	0.979091	1.00000	1.12711	1.12711	0.97918
resolution (Å)	43.79 - 2.10	34.95 - 1.70	46.51 - 1.70	40.17 - 1.85	33.63 – 1.50
	(2.18 - 2.10)	(1.76 - 1.70)	(1.76 - 1.70)	(1.92 - 1.85)	(1.55 – 1.50)
no. of observations	338002	351571	262561	231368	938410
no. of unique reflections	25062	86186	79753	67426	125320
$R_{merge}(I)$	0.094 (0.338)	0.095(0.249)	0.067 (0.384)	0.077 (0.290)	0.074 (0.539)
average I/σ	15.2 (7.3)	8.4 (2.9)	12.7 (2.2)	10.0 (2.2)	32.3 (2.8)
completeness (%)	100.0 (100.0)	99.0 (95.1)	91.7 (84.6)	99.1(99.7)	99.9 (100.0)
redundancy	13.49 (14.38)	4.08 (3.12)	3.29 (2.66)	3.43 (2 .76)	7.5 (7.4)
R_{cryst}		0.189	0.184	0.203	0.172
R_{free}[b]		0.215	0.207	0.227	0.187
no. of protein residues		658	657	656	672
no. of protein atoms		5099	5129	5096	5295
no. of water molecules		516	413	283	580
average B-factor (Å^2)					
protein		27.6	28.2	32.9	19.9
water		34.6	33.6	36.0	28.3
active site ligand		25.1	20.5		18.4
rmsd[c]					
bond lengths (Å)		0.004	0.007	0.006	0.005
bond angles (deg)		0.91	1.03	0.89	0.98
Ramachandran plot[d]					
favored (%)		99.24	99.23	99.38	99.25
allowed (%)		0.76	0.77	0.62	0.75
outliers (%)		0	0	0	0
PDB accession code		3IT0	3IT1	3IT2	3IT3

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]5% test set. A common set of test reflections was used for refinement of all structures.
[c]Compared to the parameters of Engh and Huber.[33]
[d]The Ramachandran plot was generated with RAMPAGE.[34]

2.2.3 Phasing and refinement

The structure was solved using SAD phasing based on data collected from a crystal of Se-Met FtHAP complexed with L(+)-tartrate. SAD phasing calculations were performed with SOLVE[25] using $P4_12_12$ as the space group. Eleven of the expected twelve selenium sites (i.e., one FtHAP molecule per asymmetric unit) were identified, which resulted in a figure of merit of 0.31 for reflections to 2.1 Å resolution. RESOLVE[25] was then used for density modification and automated model building. The phase set from RESOLVE had a figure of merit of 0.65. The model from automated building included 63 % of the expected residues in the asymmetric unit.

REFMAC5[26] and PHENIX[27] were used for refinement. The B-factor model used during refinement consisted of an isotropic B-factor for each non-hydrogen atom plus one TLS group per chain. COOT[28] was used for model building. The model from RESOLVE was extended through iterative rounds of model building and refinement against the SAD data set to 2.1 Å resolution. The resulting model had an R_{cryst} value of 0.24 and R_{free} of 0.26. This model served as the starting point for further model building and refinement against the 1.70 Å resolution native data set for the L(+)-tartrate complex. As the model neared completion, it became apparent that there was a steric clash between two molecules related by the symmetry operation (y, x, -z) and involving Met94 and Phe116. This conflict was interpreted as being a violation of $P4_12_12$ symmetry and could be resolved by lowering the symmetry to $P4_1$ and assuming two molecules in the asymmetric unit rather than one. Therefore, subsequent refinement calculations were

performed with $P4_1$ as the space group. The final model of the L(+)-tartrate complex was used to initiate refinements of the other structures. A common set of test reflections (5%) was used for all the refinements. See Table 2.1 for refinement statistics.

2.2.4 Site-directed Mutagenesis

Three site-directed mutants of FtHAP were created in which Ala replaced Phe23 (F23A), Tyr135 (Y135A), or Asp261 (D261A). These alterations to the FtHAP gene were made using the overlap extension PCR method as originally described by Horton *et al*.[29] PCR cycling parameters for generation of the overlap amplicons were as follows: initial denaturation at 94° C for 60 s followed by 30 cycles of 94° C denaturation (30 s), annealing at 55° C (60 s), and elongation at 72° C (120 s). The blunt-ended final PCR product was cloned into pZERO Blunt, transformed into *E. coli* DH5α and cultured on LB containing 50 µg/mL kanamycin. The blunt insert was ligated into pET20b using *Nco*I and *Xho*I restriction sites. The mutant enzymes were expressed and purified using protocols similar to those described above for the native enzyme.

2.2.5 Kinetic Characterization

Initial kinetic screening trials identified 3'-AMP and *p*-nitrophenyl phosphate (*p*NPP) as two of the best *in vitro* substrates for FtHAP (data not shown),

25

therefore, these molecules were used for characterization of steady-state kinetics. Enzymatic activity with 3'-AMP as the substrate was assessed using a discontinuous assay that measures the production of inorganic phosphate.[30] The assays were performed at 37 °C in a buffer of 0.2 M sodium acetate at pH 6.0. We note that FtHAP exhibits negligible activity at pH values greater than 7.0 with pNPP as the substrate, suggesting that the enzyme can be considered to be an acid phosphatase (data not shown). For each substrate concentration, the reaction was stopped by the addition of the malachite green reagent after reaction times of 15 s, 30 s, 45 s and 60 s, and the inorganic phosphate concentrations were then determined spectrophotometrically. The initial rate was estimated by fitting data from the four time points to a line. The kinetic constants K_m and V_{max} were estimated by fitting the initial rate data to the Michaelis-Menten equation using Origin 8 software. Inhibition by L(+)-tartrate was studied using pNPP as the substrate at 37 °C in a buffer of 0.2 M sodium acetate at pH 6.0. In this case, the production of p-nitrophenolate was monitored at 405 nm as described previously.[8b] The inhibition constant (K_i) was estimated using simultaneous nonlinear regression as described by Kakkar.[31]

2.3 Results

2.3.1 Structure of the FtHAP protomer

The FtHAP protomer comprises two domains and has approximate dimensions

of 60 Å by 32 Å by 30 Å. The core domain is a twisted, 7-stranded β-sheet flanked by two α-helices on one side and three on the other side (Fig. 2.1a). The strand order of the β-sheet is 3, 2, 4, 1, 5, 6, and 7, with all but strands 5 and 7 in parallel. This domain is similar to the α/β core domains of enzymes in the histidine phosphatase superfamily[12] and thus identifies FtHAP a member of this group. The second domain (Fig. 2.1b, pink) caps the core domain and consists of two sections of the polypeptide chain that are inserted between secondary structural elements of the core domain. The first insertion follows α1 and consists of a loop formed by residues 21 - 38. The second insertion follows α3 and includes residues 134 - 222, which form a bundle of α-helices.

The active site is located in the junction of the two domains, at one end of a solvent-filled tunnel that runs through the protein, as indicated by the bound P_i anion in Figs. 2a and 2b. The core domain contributes the presumptive catalytic histidine residue (His17) and other residues that form the phosphoryl-binding pocket, which is located in the floor of the tunnel. The cap domain forms the walls and ceiling of the tunnel. As shown below, residues of the cap domain are important for binding the non-phosphoryl groups of the substrate 3'-AMP, implying that this domain plays a role in enforcing substrate preference.
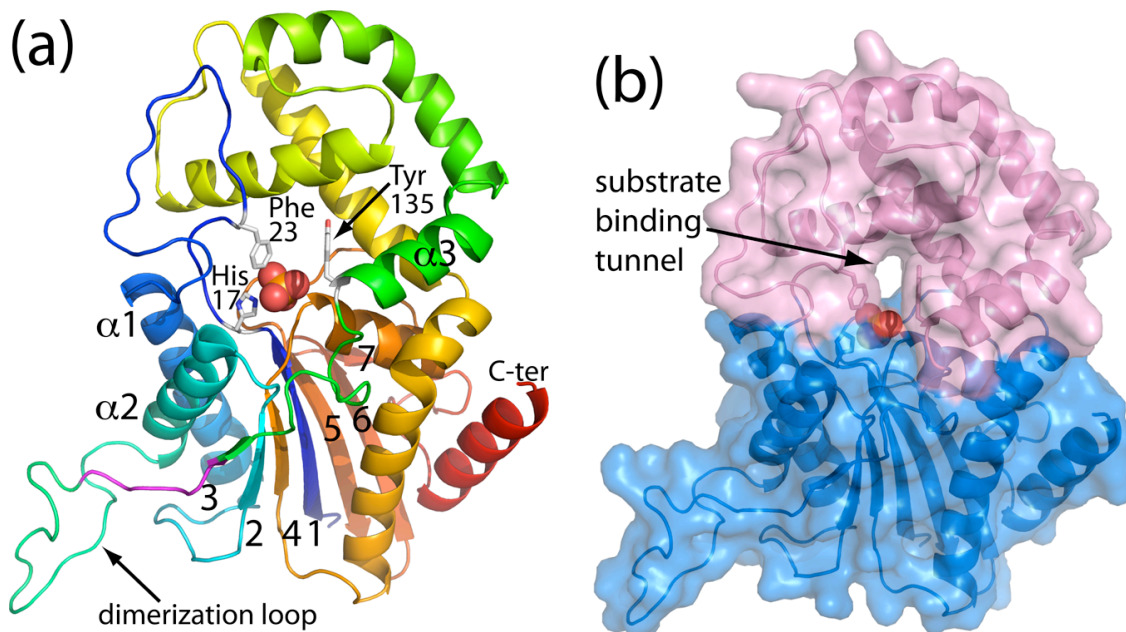
Fig. 2.1. Overall fold of FtHAP. (a) Ribbon drawing of the FtHAP protomer. Residues are colored according to a rainbow scheme, with blue at the N-terminus and red at the C-terminus. The P_i ligand is shown in spheres. The side chains of His17, Phe23, and Tyr135 are drawn as sticks in white. Strands of the β-sheet are numbered. The first three helices of the polypeptide chain are labeled α1, α2, and α3. The conserved portion of the dimerization loop (residues 116 - 121) is colored magenta. (b) Ribbon and surface representations of the protomer, highlighting domain structure. The orientation is the same as in panel (a). The core domain is colored blue, and the cap domain is colored pink. The P_i ligand is shown in spheres, and the side chains of His17, Phe23, and Tyr135 are drawn as sticks. This figure and others were created with PyMOL.[35]

hPAP[10] is the closest structural homolog of FtHAP, based on a query of the Protein Data Bank with SSM.[13] The root mean square deviation between the two enzymes is 1.7 Å over 273 residues (out of 329), despite sharing only 27 %

amino acid sequence identity. The superposition of FtHAP and hPAP shown in Fig. 2a confirms that the two enzymes share a common fold. There is, however, a major difference in the orientation of the first helix of the cap domain (α3, colored red in Fig. 2.2). The corresponding helix of hPAP (colored yellow in Fig. 2.2) is shifted 12 Å away from the active site compared to FtHAP. This difference appears to be due to a disulfide bond in hPAP that links residue 129 at the N-terminus of this helix with residue 340 near the C-terminus of the polypeptide chain (Fig. 2.2b). These Cys residues are absent in FtHAP, which evidently allows α3 to sit closer to the phosphoryl binding site. Consequently, the active site entrance is narrower in FtHAP than in hPAP.

2.3.2 Structure of the dimer

FtHAP forms a dimer in solution based on gel filtration and sedimentation equilibrium studies (data not shown), and the two-molecule assembly in the asymmetric unit (Fig. 2.3) likely represents this dimer. This assembly has an interface area of 1600 Å^2 and a complexation significance score of 1.0, based on analysis with PISA.[14] The next largest interface in the lattice buries only 400 Å^2 of surface area, and the complexation significance score is 0.0. Furthermore, an identical dimer is present in the crystal lattices of the close structural homologs, hPAP and rPAP.
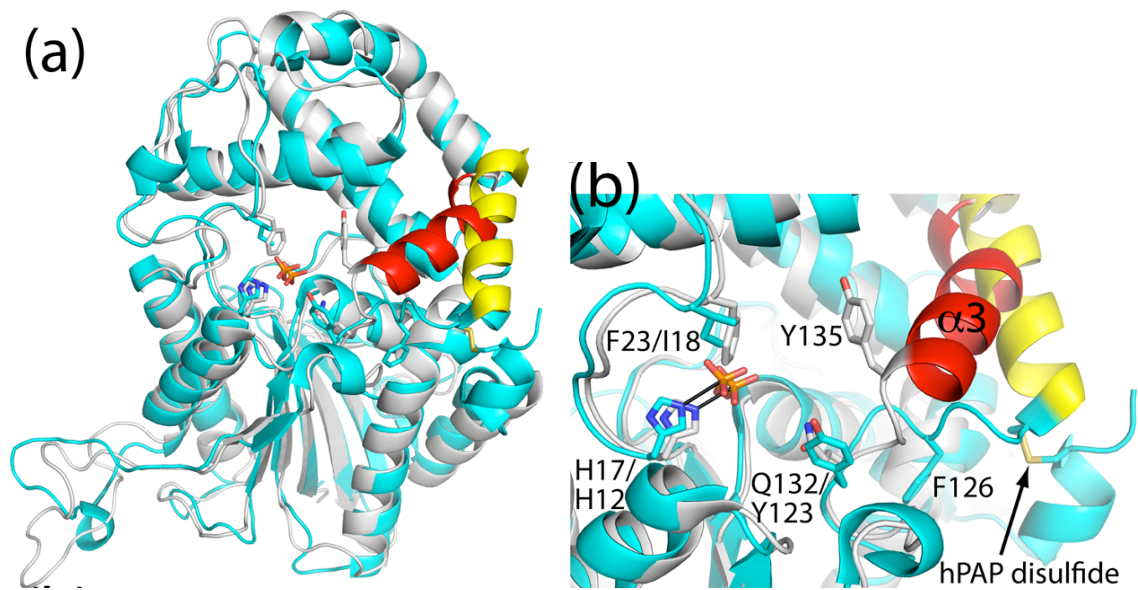
Fig. 2.2. Comparison of FtHAP and hPAP. (a) Ribbon drawings of FtHAP (white) and hPAP (cyan, PDB code 1ND6) complexed with P_i. Helix α3 of FtHAP is colored red. The corresponding helix of hPAP is colored yellow. (b) Close-up view of the active sites using the same coloring scheme as in panel (a). Where appropriate, residue numbers are listed as FtHAP/hPAP.
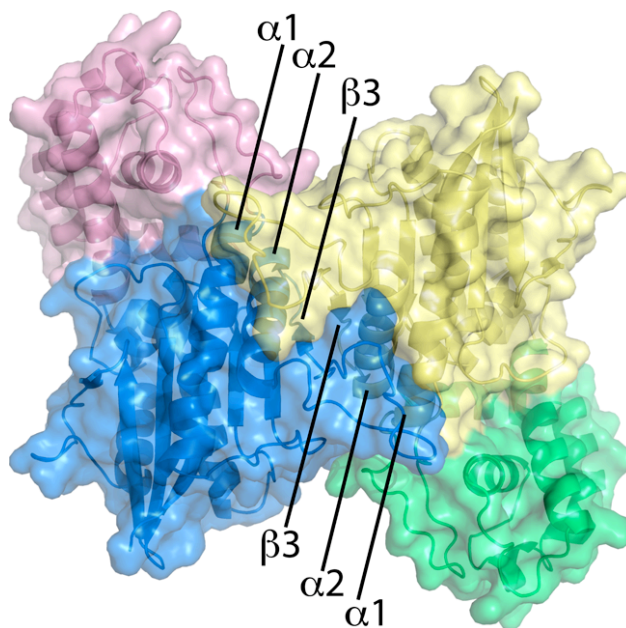
Fig. 2.3. Dimer structure of FtHAP. The core domains of the two protomers are colored blue and yellow. The cap domains are colored pink and green. The blue/pink protomer is related to the one in Fig. 2b by an approximate rotation of 180° around the vertical axis. Secondary structural elements that contribute to the dimer interface are indicated.

The FtHAP dimer comprises two interlocking protomers (Fig. 2.3). A prominent feature of the dimer is the long, protruding loop that connects α2 and β3 of the core domain (residues 98-121, Fig. 2.1a). This loop reaches out and embraces the core domain of the other protomer, thus forming an interlocking dimer. Additional secondary structural elements buried in the dimer interface include α1, α2, and β3 (Fig. 2.3).

2.3.4 Structure of FtHAP complexed with inorganic phosphate

The structure of FtHAP complexed with the product P_i was determined at 1.70 Å resolution (Table 2.1). This complex provides information about the identities of residues that directly participate in catalysis as well as those stabilizing the phosphoryl group of phosphomonoester substrates. The ion interacts with Arg16, His17, Arg20, Arg84, His260, Asp261, and one water molecule (Fig. 2.4a). Note that Arg16, His17, and Arg20 are part of the conserved RHGXRXP signature motif.

The oxygen atom of P_i denoted by O1 in Fig. 2.4a likely represents the O atom of the scissile P-O bond of the substrate, or equivalently, an O atom derived from the water molecule that attacks the phosphoenzyme intermediate. While the other oxygen atoms of P_i are buried in the active site, the P-O1 bond is directed toward the protein surface. Furthermore, O1 is the only oxygen atom of P_i that interacts with Asp261, a conserved residue proposed by Van Etten's group to protonate the leaving group.[15]

The FtHAP/P_i structure is consistent with His17 being the enzyme nucleophile. His17 forms the closest contact between the protein and the P atom (3.0 Å). Furthermore, the angle formed by the ε2 nitrogen atom of His17, the P atom, and O1 is approximately 180°, which is consistent with backside, in-line nucleophilic attack by His17. We note that the aforementioned interactions are essentially identical to those found in the structure of hPAP complexed with P_i (PDB code 1ND6[10c]).

2.3.5 Structure of FtHAP complexed with L(+)-tartrate

Inhibition by L(+)-tartrate is a common characteristic of HAPs, and as expected, L(+)-tartrate is a competitive inhibitor of FtHAP. The inhibition constant (K_i) is 0.2 mM, which is similar to the value of 0.15 mM reported for hPAP.[15a] The 1.70 Å resolution structure of FtHAP complexed with L(+)-tartrate is the first high resolution structure of any HAP complexed with this inhibitor. A moderate resolution (2.9 Å) structure of hPAP complexed with the L(+)-tartrate analog *N*-propyl-L-tartramate has been reported (PDB code 2HPA[10b]).

The bound L(+)-tartrate anion interacts with eight residues, and every functional group of the inhibitor is engaged in at least one electrostatic interaction with the protein (Fig. 2.4b). One of the carboxylate groups (labeled 1 in Fig. 2.4b) is buried deep in the P_i site, where it interacts with the side chains of Arg16 and Arg20, as well as the backbone N-H group of Asp261. The adjacent hydroxyl group of the inhibitor occupies the P_i O1 binding site and forms a hydrogen bond with Asp261. The other hydroxyl group of the inhibitor forms hydrogen bonds with His17, His260, and Arg84. Presumably, His17 - the catalytic residue - is a hydrogen bond acceptor, while His260 and Arg84 are donors. Finally, the carboxylate group of the inhibitor distal from the P_i site (labeled 4 in Fig. 2.4b) interacts with Arg84 and Gln132. All of these enzyme-inhibitor interactions, except for the hydrogen bond with Gln132, are also observed in the hPAP/*N*-propyl-L-tartramate complex.
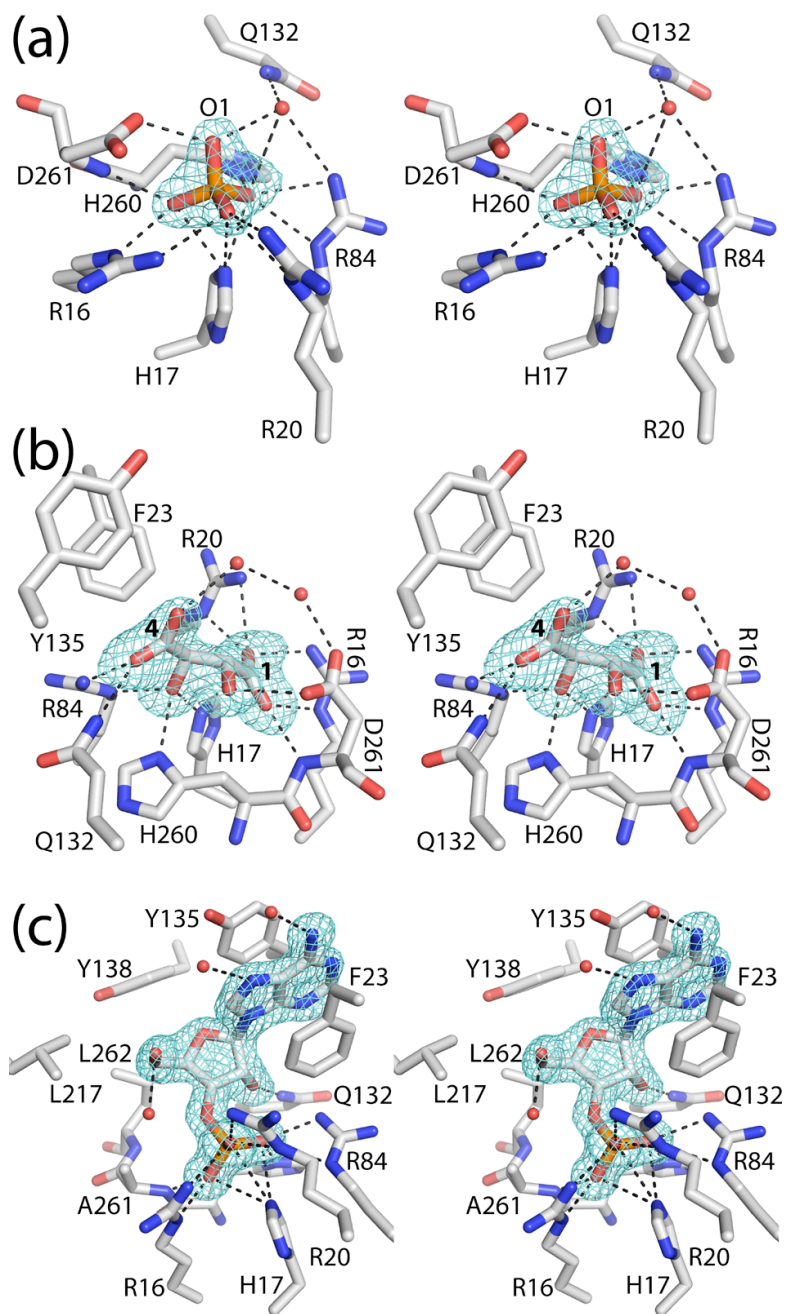
Fig. 2.4. Stereographic views (relaxed) of the actives sites of (a) FtHAP complexed with P$_i$, (b) FtHAP complexed with L(+)-tartrate, and (c) FtHAP mutant D261A complexed with 3'-AMP. In each panel, the cage represents a simulated annealing σ$_A$-weighted F_o - F_c map contoured at 3 σ. Prior to calculation of the map, the active site ligand and surrounding side chains were omitted, and simulated annealing refinement was performed with PHENIX.

2.3.6 Structure of D261A complexed with 3'-AMP

How HAPs recognize the non-phosphoryl groups of substrates is a major outstanding question. According to our initial analysis of substrate preferences, FtHAP dephosphorylates a number of different phosphomonoesters, with 3'-AMP being one of the best substrates tested (data not shown). Therefore, several site-directed mutants of FtHAP were created for use in structure determination of an enzyme/3'-AMP complex. D261A proved to be the most useful of these mutant enzymes for trapping 3'-AMP in the active site. The choice of D261A was motivated by analogous work on the protein tyrosine phosphatase PTP1B (unrelated in sequence and structure to HAPs), which showed that PTP1B mutant D181A is an effective substrate trap.[16] Asp181 of PTP1B is the general acid that protonates the substrate leaving group and Asp261 is predicted to play this role in FtHAP based on studies of homologous HAP enzymes.[15]

The structure of D261A complexed with the *in vitro* substrate 3'-AMP (K_m = 0.3 mM, k_{cat} = 120 s^{-1}) was solved at 1.50 Å resolution (Table 2.1). The electron density maps clearly indicated the conformation of the substrate (Fig. 2.4c). 3'-AMP binds in one end of a tunnel of length 15 Å that is connected to the bulk medium at both ends (Fig. 2.5). One edge of the adenine base is exposed to the bulk medium, while the rest of the substrate is inside the tunnel (Fig. 2.5).

The environment of the 3'-AMP phosphoryl group is identical to that described for P$_i$ above, except for the interaction with Asp261, which is obviously absent in D261A. Modeling of Asp261 based on the FtHAP/P$_i$ complex suggests that this residue could potentially interact with the 3' oxygen atom of 3'-AMP at a distance

of 2.5 Å. Such an interaction is consistent with Asp261 functioning as the acid that protonates the leaving group.

The enzyme also interacts with the ribose moiety of the 3'-AMP (Fig. 2.4c). The 2'-hydroxyl interacts with Gln132 via a direct hydrogen bond and a water-mediated one. The 5'-hydroxyl does not directly interact with the enzyme, but it does form a hydrogen bond with a water molecule that is part of a network of hydrogen-bonded water molecules in the tunnel (Fig. 2.5).

The base of 3'-AMP stacks in parallel between the side chains of Phe23 and Tyr135, forming a sandwich (Figs. 2.4c, 2.5). The adenine is within van der Waals contact with the aromatic rings of these residues, indicating that the stacking is intimate. Note that the Watson-Crick hydrogen bonding groups of the base are directed out of the tunnel toward the bulk medium. As a result, the base does not form hydrogen bonds with the enzyme. The adenine ring does interact with two water molecules via hydrogen bonding (Fig. 2.4c).
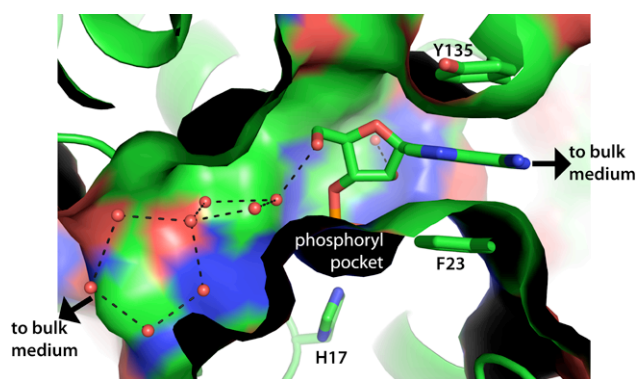
Fig. 2.5. Cut-away view of the substrate-binding tunnel with bound 3'-AMP.

Finally, although all three groups of 3'-AMP - phosphoryl, ribose, and adenine - interact with the enzyme, the fit between this substrate and enzyme is not perfect. As shown in Fig. 2.5, the substrate does not completely fill the substrate-binding tunnel, implying suboptimal shape complementarity with 3'-AMP. In particular, the pocket has substantial available space around the ribose 5'-hydroxyl, which is directed toward the solvent-filled end of the tunnel.

2.3.7 Analysis of conformational changes induced by ligand binding

The structure of inhibitor-free FtHAP was determined to gain insight into the conformational changes that accompany the binding of inhibitors and substrates. A superposition of the four FtHAP structures is shown in Fig. 2.6. The actives sites are nearly identical, except for the conformations of Arg20, Phe23, and Tyr135.

The unique conformation of Arg20 in the ligand-free enzyme appears to be due to the presence of an acetate ion near this residue. (We note that sodium acetate was included in the enzyme storage buffer.) In the ligand-free enzyme, the bound acetate ion occupies the location corresponding to the phenyl group of Phe23 in the other structures. Electron density for the side chain of Phe23 was very weak in the ligand-free enzyme, indicating that this side chain is disordered. Note that Phe23 has been truncated after the C^b atom in the ligand-free enzyme structure.

The structural comparison also reveals perceptible differences involving the substrate-clamping residues Phe23 and Tyr135. These side chains are closest together in the 3'-AMP complex (Fig. 2.6, magenta). Furthermore, the electron density for these side chains is very strong in the 3'-AMP complex, but comparatively weaker in the other structures. These observations suggest that the binding of 3'-AMP induces ordering of the side chains of Phe23 and Tyr135 and closure of the clamp.
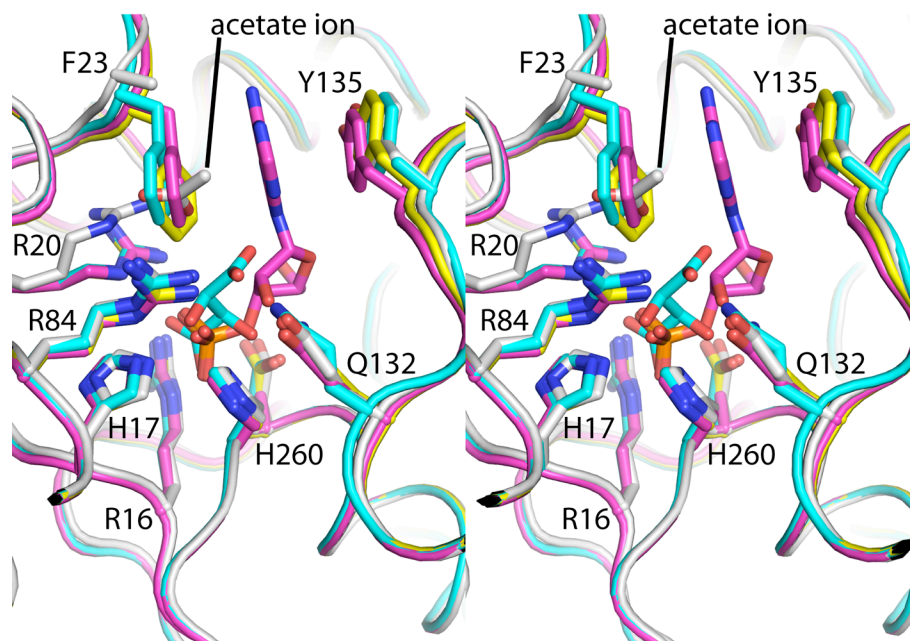
Fig. 2.6. Superposition (relaxed stereographic view) of the active sites of the four FtHAP structures. The complexes are colored as follows: inhibitor-free FtHAP, white; P_i complex, yellow; L(+)-tartrate complex, cyan; 3'-AMP complex, magenta.

2.3.8 Kinetic characterization of F23A and Y135A

The importance of the substrate-clamping residues was studied by creating the mutants F23A and Y135A. Steady-state kinetics measurements were performed using 3'-AMP as the substrate, and the resulting kinetic parameters are listed in Table 2.2.

Table 2.2. Kinetic parameters for native FtHAP, F23A, and Y135A using 3'-AMP as the substrate

	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mM^{-1})
Native	0.315 ± 0.001	119.9 ± 0.1	381
F23A	2.195 ± 0.002	37.5 ± 0.4	17
Y135A	3.0 ± 0.5	$93. \pm 31.$	31

The kinetic data suggest that Phe23 and Tyr135 play a significant role in binding 3'-AMP. The kinetic parameters for the native enzyme are 0.3 mM and 120 s^{-1} for K_m and k_{cat}, respectively, which yields a catalytic efficiency of 381 s^{-1}mM^{-1}. The efficiencies of F23A and Y135A are twenty-fold and ten-fold lower, respectively. These decreases in efficiency reflect primarily changes in K_m, which increases from 0.3 mM for the native enzyme to 2.2 mM for F23A and 3.0 mM for Y135A. These results suggest that the stacking interactions implied by the crystal structure contribute substantially to the binding energy of 3'-AMP.

2.4 Discussion

FtHAP belongs to a large class of enzymes known as the histidine phosphatase superfamily.[12] This superfamily comprises phosphoryl transfer enzymes that share a common catalytic core featuring an active site histidine that is phosphorylated during the catalytic cycle. There are two main branches in the superfamily, which are distinguished by sequence and structural characteristics. Branch 1 comprises several functionally diverse enzymes, including cofactor-dependent phosphoglycerate mutases and a variety of phosphatases. Branch 2 is less diverse, and consists mainly of HAPs, such as FtHAP and hPAP, as well as phytases.

Here we described the first structure of a bacterial member of the HAP subbranch of the histidine phosphatase superfamily. FtHAP has the same fold and dimeric structure as hPAP. The core catalytic domains of FtHAP and hPAP superimpose almost perfectly, which is consistent with the fact that the core domain is common to all members of the superfamily. It is thus likely that the fold displayed by FtHAP and hPAP is conserved throughout the HAP subbranch of the histidine phosphatase superfamily. Multiple sequence alignments suggest that the dimerization loop is present in other HAPs. Furthermore, the alignments reveal that seven residues near the end of the dimerization loop, corresponding to residues 116 - 121 of FtHAP, are highly conserved (colored magenta in Figure 2.1a). This section of the loop has the sequence FQPIPI in FtHAP. The corresponding sequence in hPAP and many other predicted HAPs is WQPIPV.

These observations suggest that the dimeric structure displayed by FtHAP and hPAP is also conserved throughout the HAP subbranch. Where FtHAP and hPAP differ is in the cap domain (pink in Fig. 2.1b), which we have shown is important for substrate recognition. In particular, the crucial first helix of the cap domain (α3) is shifted due to a disulfide bond that is present in hPAP but absent in FtHAP (Fig. 2.2b). This key structural difference in the active site implies that the two enzymes have different substrate preferences.

Although some of the physiological substrates for hPAP have been identified, nothing is known about the *in vivo* substrates for FtHAP. Previous work has suggested that the cellular form of hPAP functions as a protein phosphatase, with potential substrates including c-ErbB-2[17] and the epidermal growth factor receptor.[18] Also, the transmembrane isoform of hPAP has been recently shown to suppress pain by dephosphorylating extracellular 5'-AMP to adenosine.[19]

The structural data reported here provides clues about potential *in vivo* substrates and possible roles for FtHAP in intracellular survival. The active site entrance of FtHAP is narrower than that of hPAP due to the difference in the orientation of α3 (Fig. 2.2b). The width of the active site entrance could help dictate substrate specificity, with wider entrances allowing broader specificity, as has been suggested for PhoE, another member of the histidine phosphatase superfamily.[20] The wider entrance of hPAP is consistent with phosphoprotein substrates, but evidently does not preclude 5'-AMP as a substrate. Conversely, the narrow entrance of FtHAP is predicted to preclude phosphoproteins as potential substrates, and instead suggests small molecule phosphomonoesters

as potential substrates. In fact, docking of a phosphotyrosine pentapeptide into the FtHAP and hPAP structures suggests that residues flanking the phosphotyrosine would clash with α3 of FtHAP but not with the corresponding helix of hPAP (data not shown). Thus, our data argue against a role for FtHAP in modulating host cell signaling pathways by dephosphorylating host proteins.

The D261A/3'-AMP structure is the first structure of a substrate complex for any HAP, and thus a major result of this work is to identify some of the structural determinants of substrate recognition by HAPs.   In particular, the structure provides the first clues about residues that bind the non-phosphoryl groups of substrates. The structure shows that residues of the cap domain interact with the ribose and adenine, which suggests that the role of this domain is to help enforce substrate preference.   The cap domain provides two main contact points with 3'-AMP. Primary among these is the first helix of the domain, which contains the key residues Gln132 and Tyr135.   The former residue forms a hydrogen bond with the 2'-hydroxyl of 3'-AMP, while the latter forms one side of an aromatic sandwich that clamps the adenine base.   The second locus of interaction is the large loop insertion that follows β1. This loop contributes Phe23, which forms the other side of the clamp. We thus propose that α3 and the large loop insertion are "hot-spots" for substrate recognition.

It is interesting that Phe23, Gln132, and Tyr135 are not widely conserved among HAPs.   Isoleucine and valine are frequently found at the residue corresponding to Phe23, with Ile appearing in hPAP, rPAP, and the closest sequence homolog to FtHAP in GenBank, *Legionella pneumophila* major acid

phosphatase[21] (41 % identity). The residues corresponding to Gln132 and Tyr135 display wide variation among HAPs. The *Legionella* enzyme appears to be the only other HAP having a glutamine aligning with FtHAP Gln132. Arginine substitutes for Tyr135 in the *Legionella* enzyme. The equivalent residues in hPAP and rPAP are Tyr123 and Phe126. Tyr123 of hPAP aligns structurally with FtHAP Gln132 (Fig. 2.2b); the hydroxyl group of Tyr123 could potentially play a role in substrate binding analogous to the carboxamide of FtHAP Gln132. On the other hand, Phe126 of the human enzyme is located 13 Å (C^ζ-C^ζ distance) from Tyr135 of FtHAP because of the shift in α3 discussed above (Fig. 2.2b). Thus, significant conformational change would be required to bring this residue of hPAP into the active site to bind substrates in the manner demonstrated by the D261A/3'-AMP structure. The lack of conservation of these key residues suggests that the substrate preference of FtHAP might be unique among HAPs.

Information about substrate recognition hot spots should aid the discovery of physiological substrates of FtHAP, which is important for understanding the enzyme's biological role. For example, gene deletion studies have implicated phosphatases, including FtHAP, in intramacrophage survival and virulence of *F. tularensis*, but a more detailed, molecular-level understanding of the biological roles of FtHAP, AcpA, AcpB, and AcpC is lacking. Our data provide some general ideas that might be useful in future studies intended to identify the biological function of FtHAP. For example, the D261A/3'-AMP structure shows that 3'-AMP perfectly spans the distance between the P_i binding site and the hydrophobic clamp. Thus, it appears that the active site is suited for binding

nucleoside 3'-monophosphates. Since the enzyme does not form hydrogen bonds with the adenine base, FtHAP is predicted to lack a significant base preference for nucleoside 3'-monophosphates. Because 3'-AMP does not completely fill the substrate-binding tunnel, it is unlikely that the substrate preference of FtHAP is tightly restricted to nucleoside 3'-monophosphates. Indeed, initial *in vitro* kinetic screening trials also identified 2'- and 5'-nucleoside monophosphates as substrates for FtHAP (data not shown). How the enzyme recognizes these substrates remains to be determined. Taken together, the structural data presented here point to a possible role for FtHAP in scavenging phosphate from small molecules present in host macrophage cells, perhaps with a preference for acquiring phosphate from nucleoside monophosphates.

Accession codes

Atomic coordinates and structure factor amplitudes have been deposited in the PDB[32] with accession codes 3IT0, 3IT1, 3IT2, and 3IT3.

2.5 References

1.   Oyston, P. C.; Sjostedt, A.; Titball, R. W., Tularaemia: bioterrorism defence renews interest in Francisella tularensis. *Nat. Rev. Microbiol.* **2004,** *2* (12), 967-78.

2.   (a) Mohapatra, N. P.; Balagopal, A.; Soni, S.; Schlesinger, L. S.; Gunn, J. S., AcpA is a Francisella acid phosphatase that affects intramacrophage survival and virulence. *Infect. Immun.* **2007,** *75* (1), 390-6; (b) Mohapatra, N. P.; Soni, S.; Reilly, T. J.; Liu, J.; Klose, K. E.; Gunn, J. S., Combined deletion of four Francisella novicida acid phosphatases attenuates virulence and macrophage vacuolar escape. *Infect. Immun.* **2008,** *76* (8), 3690-9.

3.   Felts, R. L.; Reilly, T. J.; Tanner, J. J., Structure of Francisella tularensis AcpA: prototype of a unique superfamily of acid phosphatases and phospholipases C. *J. Biol. Chem.* **2006,** *281* (40), 30289-30298.

4.   Singh, H.; Felts, R. L.; Ma, L.; Malinski, T. J.; Calcutt, M. J.; Reilly, T. J.; Tanner, J. J., Expression, purification and crystallization of class C acid phosphatases from Francisella tularensis and Pasteurella multocida. *Acta Crystallogr.* **2009,** *F65* (Pt 3), 226-31.

5.   Rossolini, G. M.; Schippa, S.; Riccio, M. L.; Berlutti, F.; Macaskie, L. E.; Thaller, M. C., Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. *Cell. Mol. Life Sci.* **1998,** *54* (8), 833-850.

6.   Felts, R. L.; Reilly, T. J.; Tanner, J. J., Crystallization of AcpA, a respiratory burst-inhibiting acid phosphatase from Francisella tularensis. *Biochim. Biophys. Acta* **2005,** *1752* (1), 107-10.

7.   Felts, R. L.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Crystallization of a newly discovered histidine acid phosphatase from Francisella tularensis. *Acta Cryst.* **2006,** *F62* (Pt 1), 32-35.

8.   (a) Reilly, T. J.; Baron, G. S.; Nano, F. E.; Kuhlenschmidt, M. S., Characterization and sequencing of a respiratory burst-inhibiting acid phosphatase from Francisella tularensis. *J. Biol. Chem.* **1996,** *271* (18), 10973-10983; (b) Reilly, T. J.; Felts, R. L.; Henzl, M. T.; Calcutt, M. J.; Tanner, J. J., Characterization of recombinant Francisella tularensis acid phosphatase A. *Protein Expr. Purif.* **2006,** *45* (1), 132-141.

9.   Van Etten, R. L.; Davidson, R.; Stevis, P. E.; MacArthur, H.; Moore, D. L., Covalent structure, disulfide bonding, and identification of reactive surface and active site residues of human prostatic acid phosphatase. *J. Biol. Chem.* **1991,** *266* (4), 2313-9.

10.   (a) Jakob, C. G.; Lewinski, K.; Kuciel, R.; Ostrowski, W.; Lebioda, L., Crystal structure of human prostatic acid phosphatase. *Prostate* **2000,** *42* (3), 211-8; (b) LaCount, M. W.; Handy, G.; Lebioda, L., Structural origins of L(+)-tartrate inhibition of human prostatic acid phosphatase. *J. Biol. Chem.* **1998,** *273* (46), 30406-9; (c) Ortlund, E.; LaCount, M. W.; Lebioda, L., Crystal structures of human prostatic acid phosphatase in complex with a phosphate ion and alpha-benzylaminobenzylphosphonic acid update the mechanistic picture and offer new insights into inhibitor design. *Biochemistry* **2003,** *42* (2), 383-9.

11.   (a) Lindqvist, Y.; Schneider, G.; Vihko, P., Three-dimensional structure of rat acid phosphatase in complex with L(+)-tartrate. *J. Biol. Chem.* **1993,** *268* (28), 20744-6; (b) Lindqvist, Y.; Schneider, G.; Vihko, P., Crystal structures of rat acid phosphatase complexed with the transition-state analogs vanadate and molybdate. Implications for the reaction mechanism. *Eur. J. Biochem.* **1994,** *221* (1), 139-42; (c) Schneider, G.; Lindqvist, Y.; Vihko, P., Three-dimensional structure of rat acid phosphatase. *Embo J.* **1993,** *12* (7), 2609-15.

12.   Rigden, D. J., The histidine phosphatase superfamily: structure and function. *Biochem. J.* **2008,** *409* (2), 333-48.

13.   Krissinel, E.; Henrick, K. In *Protein structure comparison in 3D based on secondary structure matching (SSM) followed by Ca alignment, scored by a new structural similarity function*, 5th International Conference on Molecular Structural Biology, Vienna, Kungl, A. J.; Kungl, P. J., Eds. Vienna, 2003; p 88.

14.   Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-97.

15.   (a) Ostanin, K.; Saeed, A.; Van Etten, R. L., Heterologous expression of human prostatic acid phosphatase and site-directed mutagenesis of the enzyme active site. *J. Biol. Chem.* **1994,** *269* (12), 8971-8; (b) Ostanin, K.; Van Etten, R. L., Asp304 of Escherichia coli acid phosphatase is involved in leaving group protonation. *J. Biol. Chem.* **1993,** *268* (28), 20778-84.

16.   Flint, A. J.; Tiganis, T.; Barford, D.; Tonks, N. K., Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. U S A* **1997,** *94* (5), 1680-5.

17.   Meng, T. C.; Lin, M. F., Tyrosine phosphorylation of c-ErbB-2 is regulated by the cellular form of prostatic acid phosphatase in human prostate cancer cells. *J. Biol. Chem.* **1998,** *273* (34), 22096-104.

18.   Lin, M. F.; Clinton, G. M., The epidermal growth factor receptor from prostate cells is dephosphorylated by a prostate-specific phosphotyrosyl phosphatase. *Mol. Cell. Biol.* **1988,** *8* (12), 5477-85.

19.   (a) Sowa, N. A.; Vadakkan, K. I.; Zylka, M. J., Recombinant mouse PAP Has pH-dependent ectonucleotidase activity and acts through A(1)-adenosine receptors to mediate antinociception. *PLoS ONE* **2009,** *4* (1), e4248; (b) Zylka, M. J.; Sowa, N. A.; Taylor-Blake, B.; Twomey, M. A.; Herrala, A.; Voikar, V.; Vihko, P., Prostatic acid phosphatase is an ectonucleotidase and suppresses pain by generating adenosine. *Neuron* **2008,** *60* (1), 111-22.

20.   Rigden, D. J.; Mello, L. V.; Setlow, P.; Jedrzejas, M. J., Structure and mechanism of action of a cofactor-dependent phosphoglycerate mutase homolog from Bacillus stearothermophilus with broad specificity phosphatase activity. *J. Mol. Biol.* **2002,** *315* (5), 1129-43.

21.   Aragon, V.; Kurtz, S.; Cianciotto, N. P., Legionella pneumophila major acid phosphatase and its role in intracellular infection. *Infect. Immun.* **2001,** *69* (1), 177-85.

22.   Van Etten, R. L.; Saini, M. S., Selective purification of tartrate-inhibitable acid phosphatases: rapid and efficient purification (to homogeneity) of human and canine prostatic acid phosphatases. *Clin. Chem.* **1978,** *24* (9), 1525-30.

23.   Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Cryst.* **1999,** *D55* (Pt 10), 1718-25.

24.   Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307-326.

25.   Terwilliger, T. C., SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **2003,** *374*, 22-37.

26.   Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **1997,** *F53* (Pt 3), 240-55.

27.   Adams, P. D.; Gopal, K.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Pai, R. K.; Read, R. J.; Romo, T. D.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C., Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron Rad.* **2004,** *11* (Pt 1), 53-5.

28.   Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004,** *D60* (Pt 12 Pt 1), 2126-32.

29.   (a) Horton, R. M.; Cai, Z. L.; Ho, S. N.; Pease, L. R., Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* **1990,** *8* (5), 528-35; (b) Horton, R. M.; Hunt, H. D.; Ho, S. N.; Pullen, J. K.; Pease, L. R., Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **1989,** *77* (1), 61-8.

30.   (a) Lanzetta, P. A.; Alvarez, L. J.; Reinach, P. S.; Candia, O. A., An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* **1979,** *100* (1), 95-97; (b) Carter, S. G.; Karl, D. W., Inorganic phosphate assay with malachite green: an improvement and evaluation. *J. Biochem. Biophys. Methods* **1982,** *7* (1), 7-13.

31.   Kakkar, T.; Boxenbaum, H.; Mayersohn, M., Estimation of Ki in a competitive enzyme-inhibition model: comparisons among three methods of data analysis. *Drug Metab. Dispos.* **1999,** *27* (6), 756-62.

32.   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-242.

33.   Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **1991,** *A47* (4), 392-400.

34.   Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.

35.   DeLano, W. L., *The PyMOL User's Manual*. DeLano Scientific: Palo Alto, CA, USA, 2002.

Chapter 3.

Recognition of Nucleoside Monophosphate Substrates by Haemophilus influenzae Class C Acid Phosphatase

Harkewal Singh, Jonathan P. Schuermann, Thomas J. Reilly, Michael J. Calcutt,

and John J. Tanner

Author contribution- H.S. cloned, expressed, purified and crystallized wt and D66N mutant. He collected data, and determined crystal structures of all complexes. He also perfomed the enzyme kinetics experiments.

Abstract

The *e* (P4) phosphatase from *Haemophilus influenzae* functions in a vestigial NAD$^+$ utilization pathway by dephosphorylating NMN to nicotinamide riboside. P4 is also the prototype of class C acid phosphatases, which are nonspecific 5′-, 3′-nucleotidases localized to the bacterial outer membrane. To understand substrate recognition by P4 and other class C phosphatases, we have determined the crystal structures of a substrate-trapping mutant P4 enzyme complexed with NMN, 5′-AMP, 3′-AMP, and 2′-AMP. The structures reveal an anchor-shaped substrate-binding cavity comprising a conserved hydrophobic box that clamps the nucleotide base, a buried phosphoryl binding site, and three solvent-filled pockets that contact the ribose and hydrogen-bonding edge of the base. The span between the hydrophobic box and phosphoryl site is optimal for recognizing nucleoside monophosphates, which explains the general preference for this class of substrate. The base makes no hydrogen bonds with the enzyme, which is consistent with observed lack of base specificity. Two solvent-filled pockets flanking the ribose are key to the dual recognition of 5′- and 3′-nucleotides. These pockets minimize the enzyme's direct interactions with the ribose and provide sufficient space to accommodate 5′ substrates in an *anti* conformation and 3′ substrates in a *syn* conformation. Finally, the structures suggest that class B and C acid phosphatases share a common strategy for nucleotide recognition.

3.1 Introduction

The lipoprotein *e* (P4)[1] is a 28-kDa outer membrane acid phosphatase from *Haemophilus influenzae*, a common commensal inhabitant of the human nasopharynx and the etiologic agent of local and invasive infections in humans, particularly children.[2] The enzyme is a major component of the outer membrane and is highly conserved among *H. influenzae* strains. The high conservation and outer membrane location of *e* (P4) has motivated investigations of the enzyme as a potential vaccine component. Studies have shown that recombinant P4 (rP4) and rP4 mutant enzymes are highly immunogenic, anti-rP4 antibodies exhibit bactericidal activity, and immunization of mice with rP4 reduces nasal colonization of nontypeable *H. influenzae* strains.[3]

The main biological role of *e* (P4) is to catalyze the conversion of nicotinamide mononucleotide (NMN) to nicotinamide riboside (NR) as part of vestigial NAD^+ utilization pathway.[4]   *H. influenzae* lacks the full repertoire of enzymes needed for the *de novo* biosynthesis of NAD^+, therefore, it must obtain this essential cofactor from the host. The NAD^+ utilization pathway includes an uptake system that imports NAD^+, NMN, and NR into the periplasm.   Within the periplasm, the NAD^+ nucleotidase NadN catalyzes the hydrolysis of NAD^+ to generate NMN and AMP. NMN produced by NadN, or imported by the uptake system, is dephosphorylated to NR by *e* (P4). NadN also has NMN 5′-nucleotidase activity, but *e* (P4) has higher efficiency for NMN and is thus thought to be the major catalyst for the production of NR for the pathway.[4a] NR is

51

then transported across the inner membrane into the cytosol by the NR-specific permease PnuC, where it is converted into NAD^+ by the bifunctional NR kinase/NMN adenylyltransferase NadR[5]. Although the biological function of *e* (P4) in NAD^+ utilization is well established, the structural basis for the recognition of NMN by P4 is unknown.

The larger context for the research described here is that *e* (P4) is the prototype of class C acid phosphatases (CCAPs). First recognized as a family of related bacterial enzymes by Thaller *et al.* in 1998,[6] CCAPs belong to the DDDD superfamily of phosphohydrolases and are defined at the primary structure level by the conserved bipartite sequence motif of [IV]-[VAL]-**D**-[IL]-**D**-E-T-[VM]-L-X-[NT]-X(2)-Y and [IV]-[LM]-X(2)-G-**D**-[NT]-L-X-**D**-F (Asp residues of the DDDD motif in bold).

In addition to *e* (P4),[1, 7] several CCAPs have been characterized to various degrees, including those from *Elizabethkingia meningosepticum* (OlpA[8]), *Streptococcus equisimilis* (LppC[9]), *Helicobacter pylori* (HppA[10]), and *Clostridium perfringens*.[11] CCAPs are dimeric enzymes that exhibit phosphomonoesterase activity for commonly used aryl phosphate substrates such as *p*-nitrophenyl phosphate (*p*NPP) and 4-methylumbelliferyl phosphate. Among the physiologically relevant molecules that have been tested, the highest catalytic efficiencies have been achieved with nucleoside 5′-monophosphates. CCAPs do not exhibit a strong base preference among this class of substrate. OlpA, LppC,[12] rP4 (*vide infra*) and the *C. perfringens* enzyme[11a] also exhibit activity with nucleoside 3′-monophosphates, but with lower efficiency than nucleoside 5′-

monophosphate substrates. Thus, the available *in vitro* data suggest that CCAPs function primarily as nonspecific 5′-, 3′-nucleotidases.

The structure of one CCAP - rP4 - has been determined.[7c] rP4 has a two-domain fold consisting of a core α/β domain (Fig. 3.1, blue) and an α-helical cap domain (Fig. 3.1, pink). The four Asp residues of the DDDD motif are clustered around a Mg^{2+} ion at the base of active site (Fig. 3.1, yellow sphere). The core domain fold indicates that rP4 belongs to the haloacid dehalogenase (HAD) structural superfamily.[13]
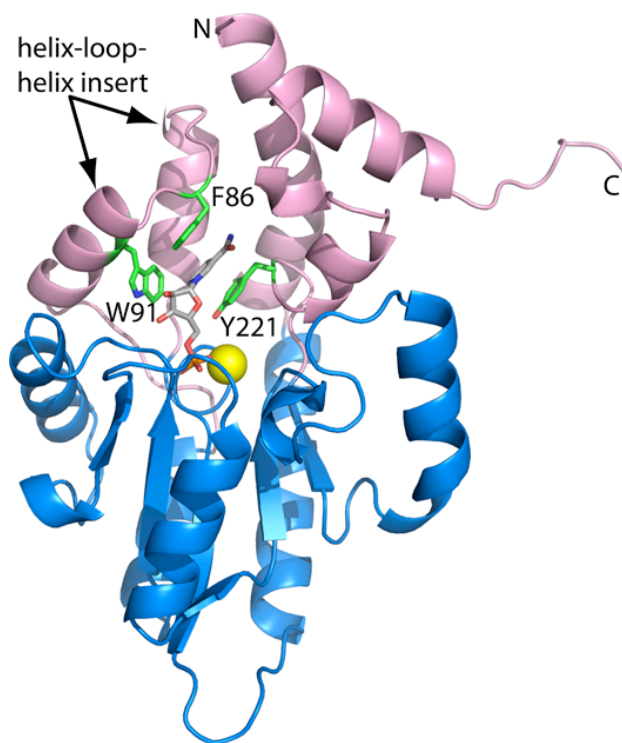
Fig. 3.1. Ribbon representation of D66N complexed with NMN. The core and cap domains are colored blue and pink, respectively. NMN is colored gray. The yellow sphere represents Mg^{2+}. Residues of the aromatic box are colored green. This figure and others were created with PyMOL.[28]

The structure of rP4 complexed with the inhibitor tungstate provided insight into the identities of the nucleophile that attacks the substrate phosphoryl group (Asp64), the residue that protonates the leaving group (Asp66), and side chains that stabilize the substrate phosphoryl group (Lys161, Thr124). However, the residues that interact with the non-phosphoryl groups of substrates have not been identified. Thus, the structural elements that enforce the preference for nucleoside monophosphates are unknown. Furthermore, the question of how CCAPs achieve the dual recognition of nucleoside 5′- and 3′-monophosphates remains unanswered.

Within this context we initiated a structure-based study of substrate recognition having the goals of understanding how rP4 binds its known biological substrate, NMN, and more generally, elucidating the structural features of CCAPs that are responsible for recognizing nucleoside monophosphate substrates. To this end, we have determined high resolution crystal structures of a substrate-trapping mutant of rP4 complexed with NMN, 5′-AMP, 3′-AMP, and 2′-AMP, as well as a structure of rP4 complexed with the product inorganic phosphate (P_i) (Table 3.1). The structures and the accompanying kinetic data provide insight into the basis of the nucleotidase activity of P4 and other CCAPs.

3.2 Materials and Methods

3.2.1 Subcloning and mutagenesis

Previous structural studies of rP4 used a recombinant enzyme lacking a polyhistidine affinity tag, but the purification of that enzyme was inefficient,[7b] and crystallization was not highly reproducible. Therefore, for this work, an rP4 construct encoding the enzyme fused to a C-terminal hexahistidine tag was created to aid purification. The *hel* gene was subcloned into pET20b using *Nco*1 and *Xho*1 sites such that the N-terminal signal sequence was replaced with the *pelB* leader sequence from *Erwinia chrysanthemi* and the N-terminal Cys of the mature protein was replaced with Met. As a result, the rP4 protein used here contains a C-terminal hexahistidine tag, is free of lipid modification, and targeted to the *Escherichia coli* periplasm.

A site-directed mutant of rP4 in which the residue that protonates the leaving group, Asp66, is changed to Asn (D66N) was created for the purpose of determining crystal structures of enzyme/substrate complexes. We note that an analogous strategy has been used to trap substrate complexes of other phosphatases.[17] The mutation was introduced into the aforementioned plasmid using the QuickChange kit (Stratagene) and confirmed by DNA sequencing.

3.2.2 Expression and purification of rP4

The rP4-pET20b plasmid was transformed into *E. coli* BL21AI cells and plated on LB medium containing ampicillin (50 µg/mL). A single colony of the transformant was picked and used to inoculate 1 L of culture. The protein was expressed using auto-induction[18] at 37 °C with constant shaking at 300 rpm. The cells were harvested by centrifugation at 3500 rpm for 30 min at 4°C and resuspended in 20 mM phosphate, 20 mM imidazole, and 500 mM NaCl at pH 7.0. The cell pellet was quick frozen in liquid nitrogen and stored at -80°C.

Frozen cells were thawed at 4 °C and ruptured using a French press at 1000 psi. Unbroken cells and cellular debris were removed by centrifugation for 60 min. at 17500 rpm and 4 °C. The supernatant was collected and subjected to a second centrifugation step (30 min, 17500 rpm, 4 °C). The resulting supernatant was used for further purification by immobilized metal ion affinity chromatography (Ni^{2+}-charged HiTRAP, GE Healthcare) followed by cation exchange chromatography (HiTRAP SP, GE Healthcare). The purified enzyme was dialyzed into 50 mM sodium acetate, 50 mM NaCl, 2.5 mM $MgCl_2$ at pH 6.0. The sample was concentrated using a centrifugal ultrafiltration device (10 kDa cutoff) to 2 - 5 mg/mL (based on the bicinchoninic acid assay, Pierce).

3.2.3 Expression and purification of D66N

The D66N mutant plasmid was transformed into *E. coli* BL21(DE3) cells and

plated on LB medium containing ampicillin (50 µg/mL). A single colony of the transformant was picked and used to inoculate a 10 mL starter culture containing ampicillin (50 µg/mL). After overnight growth at 37 °C with 250 rpm shaking, the starter culture was used to inoculate 1 L of LB medium supplemented with ampicillin (50 µg/mL). The culture was then grown at 37 °C with 250 rpm shaking until the optical density at 600 nm reached 0.6. Protein expression was induced by adding isopropyl ß-D-thiogalactoside (0.5 mM), and the culture was incubated for 8 hours at 25 °C with 200 rpm shaking. The expressed protein was purified as described above for rP4. After purification, the sample was dialyzed overnight into 50 mM sodium acetate, 50 mM NaCl at pH 6.0 and concentrated to 10 mg/ml.

3.2.4 Crystallization and preparation of enzyme-ligand complexes

Crystallization trials of rP4 and D66N were performed at 20 °C using the sitting drop method with drops formed of 2 mL each of the enzyme and reservoir solutions. Initial crystallization conditions were identified using commercially available screens (Hampton Research). Promising results were obtained with reservoirs containing ammonium citrate and PEG 3350. After optimization, diffraction quality crystals having hexagonal external morphology were grown over reservoirs containing 0.05 - 0.2 M ammonium citrate, 0.05-0.15 mM MgCl$_2$, and 18 - 28 % (w/v) PEG 3350 in the pH range of 6.8 - 7.2. The best rP4 crystals typically grew in 18 - 23 % (w/v) PEG 3350, whereas a higher concentration of 23

- 28 % (w/v) PEG 3350 was used for crystallization of D66N. Typical protein concentrations used for optimal crystal growth were 1 - 3 mg/mL for rP4 and 8 - 10 mg/mL for D66N.

Crystals of the D66N/substrate complexes were obtained by soaking as follows. Stock solutions of the substrates were prepared in water, and the pH was adjusted to 6.0. D66N crystals were cryoprotected in 28 - 30 % PEG 3350, 0.1 M ammonium citrate buffer pH 7.0, and 20% PEG 200. The cryoprotected crystals were transferred to a solution of the cryobuffer supplemented with 5 - 20 mM of a substrate and 100 - 200 mM $MgCl_2$. The soaking time was in the range of 5 - 45 minutes.

Crystals of rP4 complexed with P_i were also obtained by soaking. A stock solution of potassium phosphate (100 mM KH_2PO_4 at pH 6.0) was first prepared. Next, crystals of rP4 were cryoprotected at room temperature in 23-28 % (w/v) PEG 3350, 0.1 M ammonium citrate buffer pH 7.0, and 20% PEG 200. The cryoprotected crystals were transferred to a solution of the cryobuffer supplemented with 25 mM P_i and 200 mM $MgCl_2$. After 30 minutes the crystals were picked up with Hampton loops and plunged into liquid nitrogen.

3.2.5 Structure determination

X-ray diffraction data sets were collected at the Advanced Light Source beamline 4.2.2 and Advanced Photon Source beamline 24-ID-C (Table 3.1). The data sets were processed with HKL2000.[19]

Table 3.1. Data collection and refinement statistics

	NMN	5′-AMP	3′-AMP	2′-AMP	phosphate
Beamline	24-ID-C	4.2.2	4.2.2	4.2.2	24-ID-C
Wavelength	0.9792	1.00	1.00	1.00	0.9792
Data collection resolution (Å)	50 – 1.35 (1.40 – 1.35)	50 - 1.55 (1.61-1.55)	50 - 1.85 (1.92 – 1.85)	50 - 1.90 (1.93 – 1.90)	50 - 1.40 (1.45 – 1.40)
No. of observations	622380	944628	536096	499757	340873
No. of unique reflections	66003	44418	26166	24575	57486
$R_{merge}(I)$	0.074 (0.603)	0.088 (0.427)	0.114 (0.439)	0.120 (0.534)	0.064 (0.557)
Average I/σ	39.6 (4.3)	47.0 (5.9)	27.6 (4.35)	30.2 (3.47)	25.4 (2.8)
Completenes(%)	99.7 (100)	100 (100)	99.9 (99.4)	99.9 (98.3)	95.9 (98.3)
Redundancy	9.4 (9.4)	21.3 (19.0)	20.5 (14.3)	20.3 (13.0)	5.9 (5.4)
Refinement resolution (Å)	50 – 1.35 (1.37 –1.35)	50 - 1.55 (1.58-1.55)	50 - 1.85 (1.92 – 1.85)	50 - 1.90 (1.98 – 1.90)	50 - 1.40 (1.42 – 1.40)
R_{cryst}	0.135(0.211)	0.141(0.159)	0.164 (0.177)	0.178 (0.196)	0.139 (0.207)
R_{free}[a]	0.154(0.245)	0.171(0.219)	0.189 (0.202)	0.217 (0.245)	0.168 (0.277)
No. of protein residues	247	247	247	247	247
No. of protein atoms	1933	1938	1931	1912	1929
No. of water molecules	274	280	231	173	256
Average B-factor ($Å^2$)					
Protein	14.1	15.8	16.8	26.8	14.2
Water	25.0	27.5	24.4	30.4	24.9
Ligand	12.2	15.7	22.1	23.3	11.7
Mg^{+2}	7.7	9.7	11.8	15.8	7.6
rmsd[b]					
Bonds (Å)	0.006	0.005	0.006	0.006	0.005
Angles (deg)	1.17	1.04	1.00	1.02	1.03
Ramachandran plot[c]					
Favored (%)	98.0	98.0	98.0	97.5	98.0
Allowed (%)	2.0	2.0	2.0	2.5	2.0
Coordinate error (Å)[d]	0.13	0.17	0.18	0.21	0.14

Values for the outer resolution shell of data are given in parenthesis.
[a]A common set of test reflections (5 %) was used for refinement of all structures.
[b]Compared to the parameters of Engh and Huber.[26]
[c]The Ramachandran plot was generated with RAMPAGE.[27]
[d]Maximum likelihood-based coordinate error estimate

The crystals have space group $P6_522$ with unit cell lengths of $a = 98$ Å, $c = 107$ Å, one molecule in the asymmetric unit, 54 % solvent, and V_m of 2.65 Å^3/Da.[20] We note that this form is different from the tetragonal one used in our earlier work.[7c] Initial phases were estimated using molecular replacement as implemented in PHASER[21] with the search model derived from a previously determined rP4 structure (PDB code 3ET4[7c]). COOT[22] was used for model building, and PHENIX[23] was used for refinement. A common set of test reflections (5 %) was used for the refinement calculations. For each structure, the B-factor model using during the initial rounds of refinement consisted of an isotropic B-factor for each non-hydrogen atom and TLS refinement with one TLS group corresponding to the protein chain. Anisotropic B-factors were used during the final few rounds of refinement of the NMN and P_i complexes. The introduction of anisotropic B-factors decreased R_{free} by 0.08 for the NMN complex and 0.06 for the P_i complex.

3.2.6 Kinetic characterization

Steady-state enzymatic activity was assessed at 25 °C using a discontinuous assay that measures the production of inorganic phosphate.[24] The assay buffer consisted of 100 mM sodium acetate, 100 mM NaCl, 1 mM MgCl$_2$ at pH 5.5. For each substrate concentration, the reaction was stopped using the malachite green reagent after reaction times of 15 s, 75 s, 135 s and 195 s, and the citrate color development reagent (34 % (w/v) sodium citrate) was added 60 s after

stopping each reaction. After 30 minutes, the inorganic phosphate concentrations were determined spectrophotometrically at 625 nm by reference to a standard curve constructed from solutions of known P_i concentration. The initial rate was estimated by fitting data from the four time points to a line. Apparent values of K_m and V_{max} were estimated by fitting the initial rate data to the Michaelis-Menten equation using Origin 8 software.

3.3 Results

3.3.1 Structures D66N complexed NMN and 5′-AMP

Crystal structures of rP4 complexed with NMN and 5′-AMP were determined to understand how the enzyme recognizes nucleoside 5′-monophosphate substrates. A mutant of rP4 in which the residue that protonates the leaving group, Asp66, has been changed to Asn (D66N) was used for structure determination of enzyme-substrate complexes. Structures of D66N complexed with NMN (Fig. 3.1) and 5′-AMP were determined at high resolution limits of 1.35 Å and 1.55 Å, respectively (Table 3.1). The electron density maps at these resolutions allowed unambiguous determination of the conformations of the bound substrates, enumeration of enzyme-substrate interactions, and identification of water molecules involved in substrate binding (Fig. 3.2a, 3.2b).
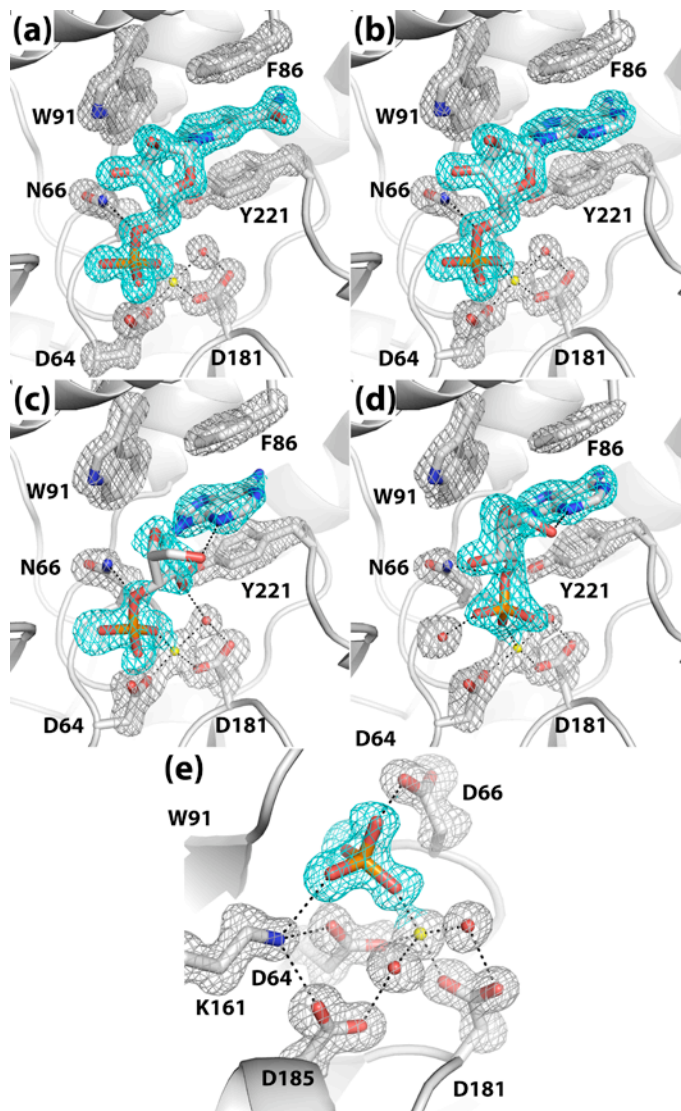
Fig. 3.2. Conformations of ligands bound to rP4: (a) NMN, (b) 5′-AMP, (c) 3′-AMP, (d) 2′-AMP, and (e) P_i. In each panel, the cage (cyan for the ligand, silver for protein side chains) represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. Prior to map calculation, the ligand and surrounding residues and water molecules were removed, and simulated annealing refinement was performed using PHENIX.

NMN and 5′-AMP bind in an anchor-shaped cavity located in the junction of the α/β core domain and the cap domain, as shown for 5′-AMP in Fig. 3.3a. NMN and 5′-AMP adopt identical conformations when bound to the enzyme (Fig. 3.4). Using the nomenclature described by Sanger,[14] the ribose in each case displays an unsymmetrical twist with major $C_{3'}$-*endo* and minor $C_{2'}$-*exo* pucker ($^3T_2$). The bases adopt *anti* orientations about the glycosyl bond.

The substrate phosphoryl group occupies the well-known phosphoryl binding site of HAD superfamily enzymes (Fig. 3.4). In both structures, the phosphoryl group interacts with the active site Mg^{2+}, conserved residues Thr124 and Lys161, and the backbone N-H group of Asn66. These interactions are identical to those observed in the structure of rP4 complexed with P_i (Fig. 3.2e). Furthermore, the substrate phosphoryl sits above the nucleophilic O atom of Asp64 in an orientation suggestive of backside nucleophilic attack. In particular, the nucleophilic O atom is poised 3.0 Å from the P atom, and the angle formed by the nucleophile, P, and $O_{5'}$ is 179°. Finally, the O atom of the scissile bond forms a hydrogen bond with Asn66 (2.8 Å). This interaction is consistent with Asp66 functioning as the acid that protonates the leaving group.

The ribose moieties of NMN and 5′-AMP form identical interactions with rP4 (Fig. 3.4). In each structure, the ribose is oriented such that the $C_{2'}$-$C_{3'}$ locus contacts the pyrrole ring of Trp91, while the opposite edge of the ring (i.e., $O_{4'}$) points towards a large solvent-filled pocket on the right hand side of the active site (pocket 1 in Fig. 3.3A). The hydroxyl groups of the ribose occupy a second, smaller pocket on the left side (pocket 2 in Fig. 3.3A). Within pocket 2, the 2′-

63

hydroxyl group forms a water-mediated hydrogen bond with Glu131 (Fig. 3.4).

Finally, the bases of NMN and 5′-AMP bind in an aromatic box formed by Phe86, Trp91, and Tyr221 (Fig. 3.4). The first two residues are part of the helix-loop-helix substructure of the cap domain, while Tyr221 is located on the loop that follows the last strand of the core domain (Fig. 3.1). Phe86 and Tyr221 form the sides of the box, while Trp91 forms the floor. In both complexes, the base stacks in parallel between Phe86 and Tyr221 forming an aromatic sandwich. Trp91 contacts the C_5 - C_6 locus of the NMN nicotinamide and the pyrrole ring of adenine. The hydrogen bonding groups of the bases are directed toward a solvent-filled cavity at the top of the active site (pocket 3 in Fig. 3.3A). As a result, neither base forms direct hydrogen bonds with the enzyme, but there are water-mediated hydrogen bonds with Gln79, Asn220, and Glu225.
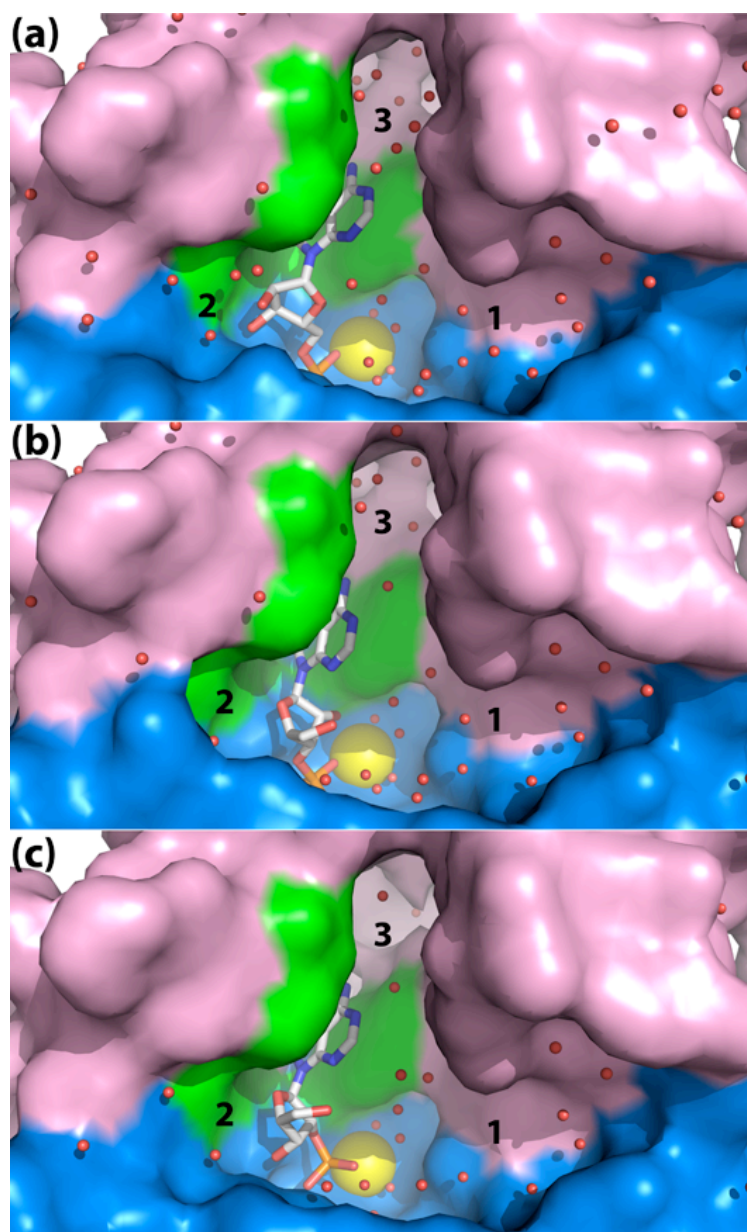
Fig. 3.3. Close-up view of the rP4 active site emphasizing the shape and solvent content. The panels correspond to D66N complexed with (a) 5'-AMP, (b) 3'-AMP, and (c) 2'-AMP. Three solvent-filled pockets are labeled 1, 2, and 3. The core and cap domains are colored blue and pink, respectively, and residues of the aromatic box are colored green. The yellow sphere represents Mg^{2+}.

65
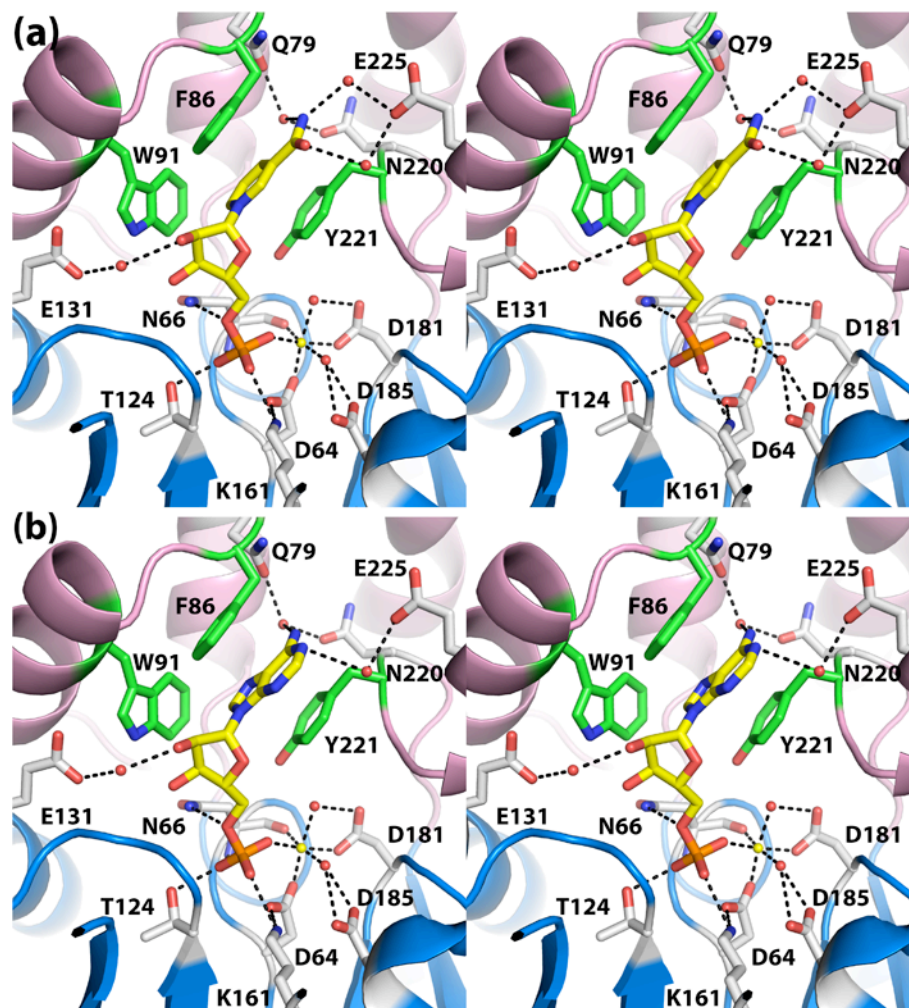
Fig. 3.4. Recognition of the nucleoside 5′-monophosphate substrates (a) NMN and (b) 5′-AMP (stereographic views). In both panels, the substrate is represented in yellow sticks, and Mg^{2+} is depicted as a yellow sphere. Secondary structural elements of the core and cap domains are colored blue and pink, respectively, and residues of the aromatic box are colored green.

3.3.2 Structures D66N complexed with 3′-AMP and 2′-AMP

Structures of D66N complexed with 3′-AMP and 2′-AMP were determined to

understand how rP4 and other CCAPs accommodate nucleoside monophosphate substrates differing in the position of the phosphoryl group on the ribose. Both electron density maps exhibited a strong feature representing the bound substrates. The quality of the 2′-AMP map (Fig. 3.2d) rivaled those of NMN and 5′AMP. Although the 3′-AMP map exhibited lower quality, the locations of the adenine ring, phosphoryl group, 2′ hydroxyl, and $O_{4'}$ were unambiguous (Fig. 3.2c). Density for the $C_{5'}$-$O_{5'}$ bond of the ribose was weaker and suggested the same conformation as in 2′-AMP. The weaker density for 3′-AMP also suggested that the occupancy of the ligand is less than 1.0; refinement of the occupancy resulted in a value of 0.81.

The conformations of 2′- and 3′-AMP differ substantially from that of 5′-AMP (Fig. 3.5). In both 2′-AMP and 3′-AMP, the ribose adopts the $C_{2'}$-*endo* pucker (^2E), and the base is in a *syn* orientation. We note that the $C_{2'}$-*endo* pucker is favored for *syn* nucleosides.[14] The *syn* conformation is stabilized by an intramolecular hydrogen bond between N_3 and $O_{5'}$ (2.9 Å in 3′-AMP, 2.8 Å in 2′-AMP). We note that this type of interaction is commonly found in *syn* nucleotides.[14] As observed with 5′-AMP, the bases of 2′-AMP and 3′-AMP bind in the aromatic box with the N_6-N_7 edge directed toward pocket 3 (Fig. 3.3b, 3.3c). However, the ribose rings of 2′-AMP and 3′-AMP are oriented with the $O_{4'}$ pointing towards the left into pocket 2, which is the reverse of the orientation of the bound 5′ substrates (compare Figs. 3.3b and c to Fig. 3.3a).

Finally, there is a difference in the position of the phosphoryl group of 2′-AMP compared to the other substrates (Fig. 3.5b). The phosphoryl of 2′-AMP is

shifted by 1.8 Å from the expected site such that the P atom is 4.3 Å from the nucleophilic O atom of Asp64, which is obviously not optimal for catalysis. As a result of the shift, the phosphoryl group does not form the expected interactions with the conserved residues Thr124 and Lys161.
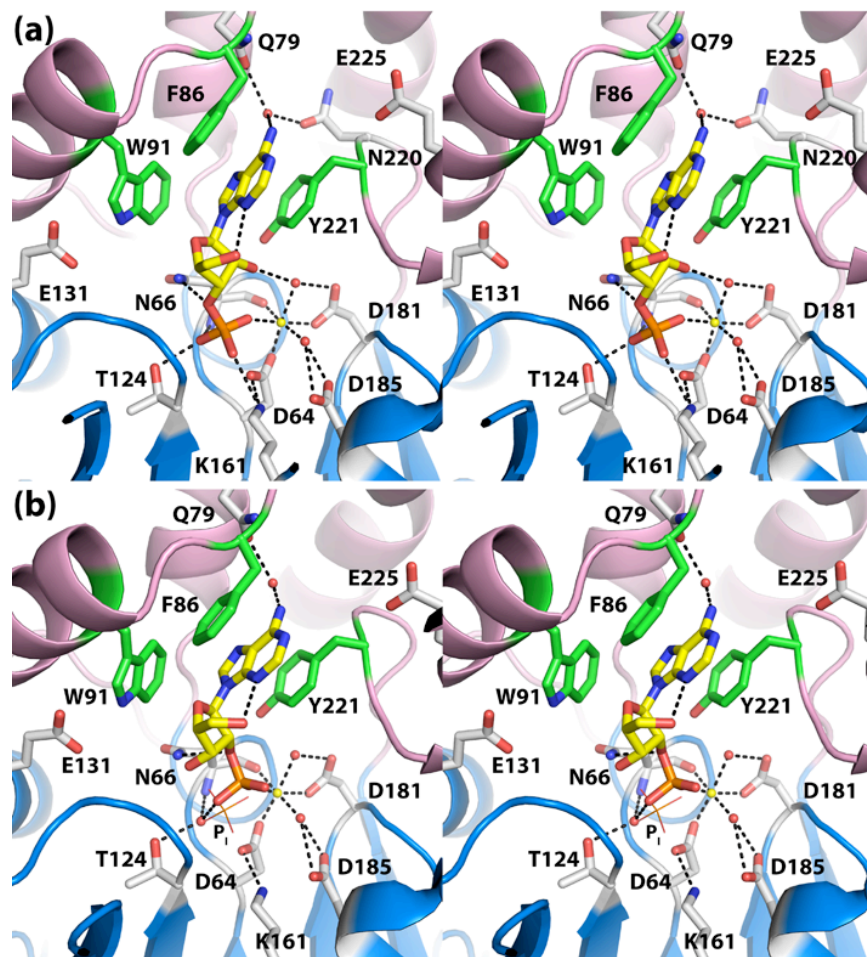
Fig. 3.5. The active sites of D66N complexed with (a) 3′-AMP (b) 2′-AMP (stereographic views). In both panels, the substrate is represented in yellow sticks, and Mg^{2+} is depicted as a yellow sphere. Secondary structural elements of the core and cap domains are colored blue and pink, respectively, and residues of the aromatic box are colored green. In panel (b), the location of normal phosphoryl binding site is indicated by the phosphate ion shown in lines and labeled P_i.

3.3.3 Kinetic characterization of rP4

The catalytic efficiencies (k_{cat}/K_m) of rP4 for certain nucleoside monophosphate substrates were estimated using steady-state kinetic assays. NMN, 5′-AMP, 3′-AMP, and 2′-AMP were used as substrates; the results are summarized in Table 3.2. Among the substrates tested, the enzyme has the highest efficiency for NMN, followed by 5′-AMP, 3′-AMP, and 2′-AMP. The efficiency for NMN is only 2.5 times that for 5′-AMP, suggesting that rP4 does not exhibit a strong preference for NMN over other nucleoside 5′-monophosphatases. Comparing the data for the adenosine monophosphate substrates, rP4 exhibits only a 2-fold preference for 5′-AMP over 3′-AMP. 2′-AMP is the poorest substrate tested, having a 20-fold lower efficiency than 5′-AMP. Thus, like some other CCAPs, it is reasonable to classify rP4 as a dual 5′-, 3′-nucleotidase.

Table 3.2. Kinetic parameters for rP4.

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (%)*
NMN	0.7 ± 0.2	0.52 ± 0.05	100
5′-AMP	0.23 ± 0.04	0.070 ± 0.004	40
3′-AMP	0.8 ± 0.2	0.10 ± 0.01	20
2′-AMP	6 ± 2	0.11 ± 0.01	2

*relative to NMN

3.4 Discussion

3.4.1 Structural basis of substrate recognition

The main result of our work is to provide structure-based insight into substrate preference and promiscuity of rP4 and CCAPs closely related to rP4. CCAPs are somewhat selective in the sense that the *in vivo* substrates are thought to be nucleoside monophosphates. For example, genetic and molecular studies suggest that NMN is a biological substrate for *e* (P4)[4b]. Although the *in vivo* substrates of other CCAPs have not been similarly identified, the available *in vitro* kinetic data suggest nucleoside monophosphates as biologically relevant substrates. On the other hand, CCAPs are generally promiscuous with regard to the identity of the base and whether the phosphoryl is attached at the 5′ or 3′ position of the sugar. For example, we have shown here that the catalytic efficiency of rP4 for 5′-AMP is approximately half that of NMN and only twice that of 3′-AMP. Similar results have been reported for other CCAPs.[8, 10-11]

The basis for the preference for nucleoside monophosphates is evident from the D66N complexes. The aromatic box is well suited for binding the aromatic ring systems of nucleotides. In particular the box provides two aromatic residues that stack in parallel with the base forming a sandwich. Furthermore, the span between the aromatic box and phosphoryl binding pocket is optimal for nucleoside monophosphates.

The structures also provide insight into substrate promiscuity with respect to the base. The base is aligned in the aromatic box such that the hydrogen

70

bonding groups point into a solvent-filled pocked. As a result, there are no direct hydrogen bonds with the enzyme implying low base selectivity. Indeed we have shown here that the catalytic efficiency of rP4 for 5′-AMP rivals that with NMN. These results imply that P4 is not tuned to exclusively recognize NMN, and that the enzyme may have other biological functions beyond NAD^+ utilization such as acquiring inorganic phosphate from nucleoside monophosphates found in the bacterium's environment.

The D66N structures also shed light on how rP4 achieves the dual recognition of 5′- and 3′-nucleoside monophosphates, a characteristic of some CCAPs. The 5′ substrates bind with the base in an *anti* conformation, whereas 3′-AMP adopts a *syn* conformation. The somewhat higher catalytic efficiency of 5′ substrates likely reflects the lower conformational energy of the *anti* conformation. The two solvent-filled pockets flanking the ribose appear to be important for the enzyme's ability to bind both types of substrates. These pockets minimize the enzyme's direct interactions with the ribose and provide sufficient space to accommodate the different ribose orientations. Indeed, the ribose occupies the widest part of the active site (Fig. 3.3). Thus, the open active site of rP4 appears to underlie both the weak base specificity and the dual recognition of 5′- and 3′-nucleoside monophosphates.

The structure of D66N complexed with 2′-AMP further demonstrates the ability of the active site to bind different nucleoside monophosphates. The structure is unusual in that the phosphoryl is shifted away from the catalytic Asp into a solvent-filled pocket and is thus not aligned optimally for catalysis. This

conformation perhaps represents a non-productive complex. As discussed by Cornish-Bowden, non-productive binding of the substrate is a form of competitive inhibition, and the measured values of V_{max} and K_m are lower than expected by an unknown and typically immeasurable amount.[15] Nevertheless, V_{max}/K_m does provide a correct measure of the catalytic properties of the enzyme in such cases.[15] We found that the apparent catalytic efficiency of rP4 for 2′-AMP is substantially lower (10 - 50 times) than those for the other substrates tested, indicating that it is a poor substrate. Regardless of whether the observed conformation represents non-productive substrate binding, the rP4/2′-AMP structure provides another illustration of how the open active site plays a role in binding different nucleoside monophosphate ligands.

3.4.2 Connections with other CCAPs and class B acid phosphatases

Sequence conservation suggests that the rP4 complexes reported here are representative of other CCAPs. The residues of the phosphate binding pocket (Asp64 Asp66, Lys161, and Thr124) are highly conserved in the HAD superfamily, and as expected, these residues are highly conserved among CCAPs. In fact, Asp64, Asp66, and Lys161 are invariant among the CCAPs, and Thr124 appears as Ser, a conservative substitution, in some CCAPs. Residues of the aromatic box are also highly conserved. Tyr221 appears to be universally conserved among CCAPs. An aromatic ring in the form of Phe, Tyr, Trp, or His is always present at the residue corresponding to rP4 Phe86. Thus, all CCAPs

appear to have residues capable of forming an aromatic sandwich with the nucleotide base. Finally, the floor of the box, Trp91, is almost invariant; substitution with Phe, a conservative change, is observed in some CCAP sequences. Thus, we suggest that the structures reported here provide a model for understanding substrate recognition in other CCAPs.

The rP4 structures also reveal a new relationship between CCAPs and the related phosphatases known as class B acid phosphatases (CBAPs). The D66N/5′-AMP complex is reminiscent of the structure of the CBAP AphA complexed with the 5′-AMP analog 9-[(R)-2-(phosphono-methoxy)ethyl]adenine (PMEA, PDB code 2G1A[16]). CBAPs also belong to the DDDD superfamily, and AphA is regarded as the prototype of the family. CBAPs have a bipartite sequence motif that is similar to that of CCAPs, and like CCAPs, class B enzymes also show a preference for nucleoside 5′- and 3′-monophosphate substrates. Although rP4 and AphA share a common HAD superfamily core domain, their sequences have negligible similarity outside of the bipartite sequence motif (13 % overall identity), and their cap domains have different folds. The latter difference is related to the different quaternary structures of the two enzymes (dimer for CCAPs, tetramer for CBAPs).
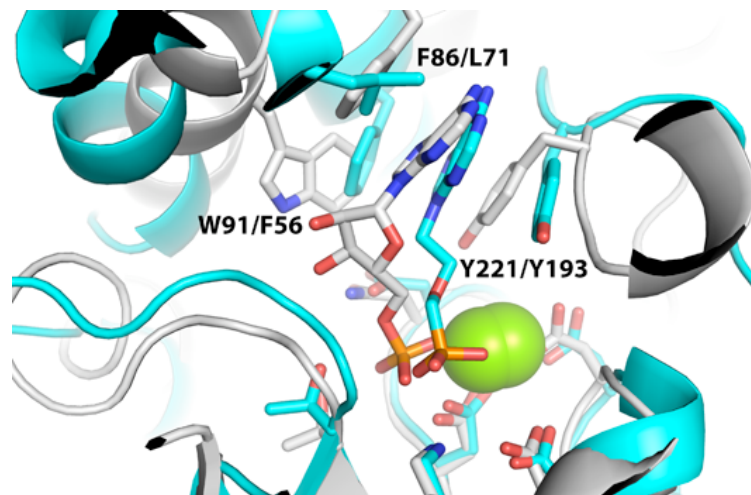
Fig. 3.6. Comparison of the active sites of D66N/5′-AMP (white) and AphA/PMEA (cyan). Residue numbers for the aromatic boxes are listed as rP4/AphA. The yellow spheres represent Mg^{2+}.

Despite the different cap domain structures, AphA and P4 exhibit commonalities in substrate recognition. In particular, the adenine ring of PMEA packs into a hydrophobic pocket that is similar to the aromatic box of rP4 (Fig. 3.6). The hydrophobic pocket of AphA consists of Leu71, Tyr193, and Phe56, which are analogous to the rP4 aromatic box residues Phe86, Tyr221, and Trp91, respectively. Note also that the adenine rings have nearly identical orientations in the two structures. We thus suggest that class B and C acid phosphatases share a common strategy for nucleotide recognition.

Accession codes

Atomic coordinates and structure factor amplitudes have been deposited in the PDB[25] with accession codes 3OCU (NMN), 3OCV (5′-AMP), 3OCW (3′-AMP), 3OCX (2′-AMP), and 3OCY (P_i).

3.5 References

1.	(a) Reilly, T. J.; Chance, D. L.; Smith, A. L., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J. Bacteriol.* **1999,** *181* (21), 6797-6805; (b) Reilly, T. J.; Smith, A. L., Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr. Purif.* **1999,** *17* (3), 401-409.

2.	(a) Foxwell, A. R.; Kyd, J. M.; Cripps, A. W., Nontypeable Haemophilus influenzae: pathogenesis and prevention. *Microbiol. Mol. Biol. Rev.* **1998,** *62* (2), 294-308; (b) Murphy, T. F.; Faden, H.; Bakaletz, L. O.; Kyd, J. M.; Forsgren, A.; Campos, J.; Virji, M.; Pelton, S. I., Nontypeable Haemophilus influenzae as a pathogen in children. *Pediatr. Infect. Dis. J.* **2009,** *28* (1), 43-8.

3.	(a) Green, B. A.; Baranyi, E.; Reilly, T. J.; Smith, A. L.; Zlotnick, G. W., Certain site-directed, nonenzymatically active mutants of the Haemophilus influenzae P4 lipoprotein are able to elicit bactericidal antibodies. *Infect. Immun.* **2005,** *73* (7), 4454-4457; (b) Mason, K. W.; Zhu, D.; Scheuer, C. A.; McMichael, J. C.; Zlotnick, G. W.; Green, B. A., Reduction of nasal colonization of nontypeable Haemophilus influenzae following intranasal immunization with rLP4/rLP6/UspA2 proteins combined with aqueous formulation of RC529. *Vaccine* **2004,** *22* (25-26), 3449-56; (c) Hotomi, M.; Ikeda, Y.; Suzumoto, M.; Yamauchi, K.; Green, B. A.; Zlotnick, G.; Billal, D. S.; Shimada, J.; Fujihara, K.; Yamanaka, N., A recombinant P4 protein of Haemophilus influenzae induces specific immune responses biologically active against nasopharyngeal colonization in mice after intranasal immunization. *Vaccine* **2005,** *23* (10), 1294-300.

4.	(a) Kemmer, G.; Reilly, T. J.; Schmidt-Brauns, J.; Zlotnik, G. W.; Green, B. A.; Fiske, M. J.; Herbert, M.; Kraiss, A.; Schlor, S.; Smith, A.; Reidl, J., NadN and e (P4) are essential for utilization of NAD and nicotinamide mononucleotide but not nicotinamide riboside in Haemophilus influenzae. *J. Bacteriol.* **2001,** *183* (13), 3974-3981; (b) Gerlach, G.; Reidl, J., NAD+ utilization in Pasteurellaceae: simplification of a complex pathway. *J. Bacteriol.* **2006,** *188* (19), 6719-27.

5.	Singh, S. K.; Kurnasov, O. V.; Chen, B.; Robinson, H.; Grishin, N. V.; Osterman, A. L.; Zhang, H., Crystal structure of Haemophilus influenzae NadR protein. A bifunctional enzyme endowed with NMN adenyltransferase and ribosylnicotinimide kinase activities. *J. Biol. Chem.* **2002,** *277* (36), 33291-9.

6.	Thaller, M. C.; Schippa, S.; Rossolini, G. M., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci.* **1998,** *7* (7), 1647-52.

7.   (a) Reilly, T. J.; Green, B. A.; Zlotnick, G. W.; Smith, A. L., Contribution of the DDDD motif of H. influenzae e (P4) to phosphomonoesterase activity and heme transport. *FEBS Lett.* **2001,** *494* (1-2), 19-23; (b) Ou, Z.; Felts, R. L.; Reilly, T. J.; Nix, J. C.; Tanner, J. J., Crystallization of recombinant Haemophilus influenzae e (P4) acid phosphatase. *Acta Cryst.* **2006,** *F62* (Pt 5), 464-6; (c) Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of Recombinant Haemophilus Influenzae e (P4) Acid Phosphatase Reveals a New Member of the Haloacid Dehalogenase Superfamily. *Biochemistry* **2007,** *46* (39), 11110-9.

8.   Passariello, C.; Schippa, S.; Iori, P.; Berlutti, F.; Thaller, M. C.; Rossolini, G. M., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5'-nucleotides. *Biochim. Biophys. Acta* **2003,** *1648* (1-2), 203-9.

9.   Malke, H., Cytoplasmic membrane lipoprotein LppC of Streptococcus equisimilis functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998,** *64* (7), 2439-42.

10.   Reilly, T. J.; Calcutt, M. J., The class C acid phosphatase of Helicobacter pylori is a 5' nucleotidase. *Protein Expr. Purif.* **2004,** *33* (1), 48-56.

11.   (a) Reilly, T. J.; Chance, D. L.; Calcutt, M. J.; Tanner, J. J.; Felts, R. L.; Waller, S. C.; Henzl, M. T.; Mawhinney, T. P.; Ganjam, I. K.; Fales, W. H., Characterization of a unique class C acid phosphatase from Clostridium perfringens. *Appl. Environ. Microbiol.* **2009,** *75* (11), 3745-54; (b) Wang, R.; Ohtani, K.; Wang, Y.; Yuan, Y.; Hassan, S.; Shimizu, T., Genetic and biochemical analysis of a class C non-specific acid phosphatase (NSAP) of Clostridium perfringens. *Microbiology* **2010,** *156* (Pt 1), 167-73.

12.   Malke, H.; Steiner, K. In *Functional properties of the cytoplasmic membrane lipoprotein acid phosphatase of Streptococcus dysgalactiae subsp. equisimilis*, XIV Lancefield International Symposium on Streptococci and Streptococcal Diseases, Aukland New Zealand, 10-11-1999; Martin, D. R.; Tagg, J. R., Eds. Aukland New Zealand, 1999; pp 879–883.

13.   Allen, K. N.; Dunaway-Mariano, D., Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **2004,** *29* (9), 495-503.

14.   Saenger, W., *Principles of nucleic acid structure*. Springer-Verlag: New York, 1984.

15.   Cornish-Bowden, A., *Fundamentals of enzyme kinetics*. Butterworths: London, 1979.

16.   Leone, R.; Cappelletti, E.; Benvenuti, M.; Lentini, G.; Thaller, M. C.; Mangani, S., Structural insights into the catalytic mechanism of the bacterial class B phosphatase AphA belonging to the DDDD superfamily of phosphohydrolases. *J. Mol. Biol.* **2008,** *384* (2), 478-88.

17.   (a) Flint, A. J.; Tiganis, T.; Barford, D.; Tonks, N. K., Development of "substrate-trapping" mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. U S A* **1997,** *94* (5), 1680-5; (b) Singh, H.; Felts, R. L.; Schuermann, J. P.; Reilly, T. J.; Tanner, J. J., Crystal Structures of the histidine acid phosphatase from Francisella tularensis provide insight into substrate recognition. *J. Mol. Biol.* **2009,** *394* (5), 893-904.

18.   Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005,** *41* (1), 207-34.

19.   Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276*, 307-326.

20.   (a) Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968,** *33*, 491-497; (b) Kantardjieff, K. A.; Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **2003,** *12* (9), 1865-71.

21.   McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.

22.   Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004,** *D60* (Pt 12 Pt 1), 2126-32.

23.   Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D* **2010,** *66* (Pt 2), 213-21.

24.   (a) Lanzetta, P. A.; Alvarez, L. J.; Reinach, P. S.; Candia, O. A., An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* **1979,** *100* (1), 95-97; (b) Carter, S. G.; Karl, D. W., Inorganic phosphate assay with malachite green: an improvement and evaluation. *J. Biochem. Biophys. Methods* **1982,** *7* (1), 7-13.

25.   Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-242.

26.   Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **1991,** *A47* (4), 392-400.

27.   Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.

28.   DeLano, W. L., *The PyMOL User's Manual*. DeLano Scientific: Palo Alto, CA, USA, 2002.

Chapter 4.

Expression, purification and crystallization of class C acid phosphatases from *Francisella tularensis* and *Pasteurella multocida*

Harkewal Singh, Richard L. Felts, Li Ma, Thomas J.Malinski, Michael J.Calcutt, Thomas J.Reilly and John J.Tanner

Author contribution – H.S. expressed, purified, crystallized PmCCAP and also collected data.

Abstract

Class C nonspecific acid phosphatases are bacterial enzymes that are secreted across the cytoplasmic membrane and hydrolyze a variety of phosphomonoesters at acidic pH. These enzymes are of interest for the development of improved vaccines and clinical diagnostic methods. In one case, the category A pathogen *Francisella tularensis*, the class C phosphatase plays a role in bacterial fitness. Here, we report the cloning, expression, purification and crystallization methods for the class C acid phosphatases from *F. tularensis* and *Pasteurella multocida*. Crystals of the *F. tularensis* enzyme diffract to 2.0 Å resolution and have space group $C222_1$ with one enzyme molecule in the asymmetric unit. Crystals of the *P. multocida* enzyme diffract to 1.85 Å resolution and have space group $C2$ with three molecules in the asymmetric unit. Diffraction patterns from crystals of the *P. multocida* enzyme exhibit multiple interpenetrating reciprocal space lattices, indicating epitaxial twinning. Despite this aberrance, autoindexing was robust and the data could be satisfactorily processed to 1.85 Å resolution using MOSFLM and SCALA.

4.1 Introduction

Acid phosphatases (EC 3.1.3.2) are ubiquitous and catalyze, at acidic pH, the transfer of phosphoryl groups from phosphomonoesters to water [1]. They play essential roles in the generation, acquisition and mobilization of inorganic phosphate, as well as critical roles in phosphoryl relay systems involved in signal transduction pathways in both prokaryotes and eukaryotes.

Three classes of bacterial nonspecific acid phosphatases (NSAPs) have been identified (denoted A, B and C) based on subcellular localization and conserved amino acid sequence motifs [2]. The enzymes studied here belong to class C, which is characterized by the presence of four invariant aspartate residues known as the DDDD motif (in bold) contained within the bipartite signature motif of [IV]-[VAL]-**D**-[IL]-**D**-E-T-[VM]-L-X-[NT]-X-X-Y near the N-terminus and [IV]-[LM]-X-X-G-**D**-[NT]-L-X-**D**-F near the C-terminus [3]. Several class C enzymes have been purified and/or characterized, including those from *Haemophilus influenzae* (known as *e* (P4)) [4], *Bacillus anthracis* [5], *Streptococcus equisimilis* [6], *Staphylococcus aureus* [7], *Helicobacter pylori* [8], *Chryseobacterium meningosepticum* [9] and *Clostridium perfringens* [10]. Shared attributes include polypeptide size of 25 - 30 kDa, requirement of metal cation for catalytic activity and rather broad substrate specificity. Most have an N-terminal lipidated Cys that anchors them to the outer membrane of the bacterium. The *S. aureus* and *C. perfringens* enzymes appear to be exceptions in this regard.

Crystal structures of class C NSAPs from *H. influenzae* [11] and *B. anthracis* (PDB code 2I33) have been determined. The structures show that class C

NSAPs belong to the haloacid dehalogenase structural superfamily and that the conserved Asp residues of the DDDD motif participate in binding an active site Mg^{2+} ion [11a]. Furthermore, the structure of the *H. influenzae* enzyme complexed with a tungstate ion inhibitor suggests that the first Asp of the motif is the nucleophile that attacks the substrate P atom [11a].

There is emerging interest in class C NSAPs. Because of their localization to the bacterial outer membrane, they are potentially attractive candidates for vaccine development. In fact, significant progress has been made toward creating a vaccine against nontypeable *H. influenzae* using catalytically inactive mutants of recombinant *e* (P4) [12]. Moreover, the class C NSAP from the category A bioterrorism pathogen *Francisella tularensis* has been implicated in phagosomal escape and virulence [13]. Finally, detection of acid phosphatase activity is a useful diagnostic tool for clinical identification of *C. perfringens* [14]. Presumably, these methods detect the *Clostridium* class C NSAP. By extension, it may be possible to develop analogous tools for identifying other bacteria based on immunological and biochemical detection of an organism's unique class C NSAP. Elucidation of crystal structures of class C NSAPs will aid this effort.

To facilitate further investigations of this enzyme family, we have developed recombinant expression systems for the class C enzymes from *F. tularensis* (FtAcpC) and *Pasteurella multocida* (PmAcpC), the latter being a pathogen of significant agricultural importance [15]. Methods for expressing, purifying and crystallizing these enzymes are described herein, along with preliminary analyses of X-ray diffraction data.

4.2 Methods and results

4.2.1 Cloning of the FtAcpC gene

The *FtAcpC* gene was identified previously as one of five phosphatase genes present in the genome sequence of *F. tularensis* subsp. *novicida* [13]. The gene encodes a 247-residue protein that has 25 % identity over 225 amino acids with *H. influenzae e* (P4), the archetype of class C NSAPs. Moreover, the sequence contains a bipartite motif that is characteristic of class C NSAPs. The motif for FtAcpC is IL**DID**ETALDNS at residues 73 - 84 followed by IAYFG**D**NIQ**D**F at residues 202 - 212. Secretion across the cytoplasmic membrane is a defining characteristic of class C NSAPs, so the protein sequence was analyzed for potential signal peptides. Analysis with the SignalP 3.0 [16], Signal-BLAST [17] and PSORT 6.4 [18] servers suggests that the N-terminus contains a signal peptide of length 20 - 23 residues, implying that the protein is expressed as a precursor polypeptide that is exported from the cytoplasm and cleaved to yield the mature enzyme. There is a cysteine at position 24, which could potentially be lipidated, as is predicted for most other class C NSAPs, but this remains to be determined. The sequence similarity to *e* (P4), bipartite DDDD motif and predicted export from the cytoplasm are consistent with classification of the enzyme into the class C NSAP family.

The *FtAcpC* gene (NCBI RefSeq number YP_169641) was cloned from genomic DNA of *F. tularensis* strain SCHU S4 [19]. The gene was amplified by PCR and inserted into pET20b using *Nco*I and *Xho*I restriction sites. To avoid

potential problems with membrane association, the N-terminal 24 residues were replaced by the *pelB* leader sequence followed by Met-Gly. Thus, the mature recombinant enzyme expressed in *E. coli* is exported to the periplasmic space and is predicted to have N-terminal residues MGNSVNI and a C-terminal $(His)_6$ tag.

4.2.2 Identification and cloning of the PmAcpC gene

Query of the complete genome sequence of *P. multocida* strain Pm70 [20] with class C NSAP genes identified an open reading frame, designated PM1064, which had the potential to be a class C NSAP. PM1064 encodes 272 residues having 59 % global amino acid sequence identity with *e* (P4). PmAcpC contains a classic 19-residue lipoprotein signal peptide in the N-terminus [21]. The predicted signal peptide contains positively charged residues (Lys) at positions - 18 and -15, followed by a hydrophobic region, and ending in the lipoprotein box LLAA-C. Thus, the mature form of PmAcpC is predicted to be a 253-residue lipoprotein, with Cys1 modified by lipidation. Furthermore, the predicted protein contains a class C bipartite sequence motif of VV**DLD**ETMIDNS at residues 61 - 72 of the mature protein and VLFVG**D**NLN**D**F at residues 175-185. The high sequence identity with *e* (P4), bipartite DDDD motif and lipoprotein prediction are consistent with assignment of this protein as a class C NSAP.

The *PmAcpC* gene (GenBank accession number FJ609981) was cloned from genomic DNA from a clinical isolate. The gene was amplified by PCR and

cloned as an *Nco*I-*Xho*I fragment into pET20b. To produce a soluble protein in *E. coli*, the N-terminal 20 residues were replaced by the *pelB* leader sequence followed by Met-Val. The mature recombinant enzyme is thus exported to the periplasmic space and is predicted to have N-terminal residues MVSNQQA and a C-terminal (His)$_6$ tag.

4.2.3 Expression and purification of FtAcpC

Recombinant FtAcpC was expressed in *E. coli* BL21 (DE3). A starter culture was grown overnight at 37 °C in 5 mL LB supplemented with 50 µg/mL ampicillin and 0.2% (w/v) glucose. The starter culture was diluted 1:500 into 35 mL LB containing 50 µg/mL ampicillin and 0.2% (w/v) glucose. The culture was incubated at 37 °C with constant aeration supplied by orbital shaking at 250 rev min^{-1} until OD$_{600}$ reached 0.6. The cells were then centrifuged at 4 °C (3600g for 10 minutes) and the resulting pellet was re-suspended in 4 mL of LB and diluted 1000-fold into 1 L LB containing 50 µg/mL ampicillin and 0.2% (w/v) glucose. The culture was incubated at 37 °C with constant aeration (250 rev min^{-1}) and induced with 0.4 mM IPTG after OD$_{600}$ reached 0.4. The incubation was continued at the lower temperature of 25 °C and decreased shaking rate of 220 rev min^{-1} until OD$_{600}$ reached 1.0. The cells were harvested by centrifugation (3600g, 4 °C), suspended in 50 mM Tris-HCl pH 8.4 and frozen.

The frozen cell suspension was thawed and the cells were disrupted in a Thermo 40K cell French Press adjusted to 10,000 psi for two cycles at a low flow

rate to maintain the cell pressure. Unbroken cells and cellular debris were removed by centrifugation at 31,000g for 20 min at 4 °C. Remaining bacterial membranes were pelleted by centrifugation at 183,960g for 1 hr at 4 °C.

The supernatant from ultracentrifugation was filtered through a 0.2 μm Millipore filter and applied to a Q Sepharose anion exchange column with an ÄKTA FPLC chromatography system. The sample was loaded on the column at 2 mL/min using a buffer of 50 mM Tris-HCl at pH 8.4. A linear NaCl gradient was then applied, eluting FtAcpC in the range 60 - 100 mM NaCl. Fractions containing FtAcpC were identified using SDS-PAGE and phosphatase activity assays, the latter consisting of a discontinuous colorimetric assay with *p*-nitrophenylphosphate as the substrate [22]. The pooled fractions were dialyzed overnight at 4 °C against 20 mM phosphate buffer at pH 7.0 containing 0.5 M NaCl in preparation for immobilized metal ion affinity chromatography. The dialyzed sample was loaded onto a Ni(II)-charged affinity column and eluted with 175 mM - 200 mM imidazole. Fractions having the highest purity level, based on the SDS-PAGE, were pooled and dialyzed overnight into 50 mM sodium acetate buffer at pH 6.0. Finally, the sample was concentrated using a centrifugal concentrating device to 10 mg/mL. Protein concentration was assessed using the BCA method (Pierce kit).

4.2.4 Expression and purification of PmAcpC

Recombinant PmAcpC was expressed using the auto-induction method [23]. A single colony of *E. coli* BL21 (DE3) containing the engineered plasmid was used to inoculate 5 mL of LB containing ampicillin (50 µg/mL) and incubated overnight at 37 °C. The culture was pelleted at 3,660g and suspended in fresh LB prior to inoculation of four 2.8 L flasks, each containing 500 mL of TY broth supplemented with the ZYM-5052 media components and ampicillin (50 µg/mL). The culture was grown with constant aeration at 37 °C overnight.

The enzyme was purified using standard non-denaturing methods. All procedures were conducted at 4 °C unless otherwise noted. Cultures from auto-induction were centrifuged at 3,660g for 20 min and the pellet was suspended in 40 mL of pH 7.0 buffer containing 20 mM sodium phosphate and 0.5 M NaCl. The suspension was stored at -20 °C overnight and thawed the next day. A protease inhibitor tablet (SigmaFAST) and DNase were added to the thawed suspension and mixed for approximately 30 minutes. The cells were disrupted and subjected to low-speed centrifugation and ultracentrifugation steps as described above for FtAcpC. The ultracentrifugation supernatant was retained and refrigerated.

The supernatant from ultracentrifugation was loaded at 2 mL/min onto a 5 mL Ni(II)-charged HiTrap metal-chelate chromatography resin using an ÄKTAprime system. The resin was washed with 20 mM sodium phosphate, 0.5 M NaCl, 50 mM imidazole pH 7.0 until A_{280} returned to baseline. A 200 mL linear imidazole gradient (0.05–2.0 M imidazole in 20 mM sodium phosphate pH 7.0, 0.5 M NaCl)

was applied and the eluate collected in 2 mL fractions. PmAcpC eluted at approximately 200 mM imidazole. Fractions containing high phosphatase activity (*p*-nitrophenylphosphate substrate) were pooled and dialyzed overnight against 50 mM sodium acetate pH 6.0.

The dialyzed sample was loaded at a flow rate of 2.0 mL/min onto a 5 mL SP-Sepharose cation exchange column that had been equilibrated with 50 mM sodium acetate pH 6.0. The column was washed with 50 mM sodium acetate pH 6.0 until A_{280} returned to baseline. The phosphatase was eluted from the column with a 200 mL linear NaCl gradient (0.0 – 1.0 M) and collected in 2 mL fractions. PmAcpC eluted in the range 0.44 M - 0.53 M NaCl. Enzymatically active fractions that were judged to be highly pure by SDS-PAGE were pooled and dialyzed overnight against 20 mM sodium phosphate pH 7.0. The sample was then concentrated to 10 mg/mL (BCA assay) with a 10,000 MWCO Centricon filtering device.

4.2.5 Crystallization of FtAcpC and preliminary analysis of X-ray diffraction data

Crystallization trials were performed at 293 K using the sitting drop method of vapor diffusion with Cryschem plates and reservoir volumes of 1 mL. Drops were formed by mixing 2 µl of the protein stock solution and 2 µl of the reservoir solution. Crystal screening trials using commercially available reagent kits revealed promising crystallization conditions consisting of 25 - 30 % (w/v) PEG

3350, 0.1 M Bis-Tris pH 5.5 - 6.5 and 0.1 - 0.2 M of either NaCl or ammonium acetate. The latter salt proved to be the preferred additive, and eventually, the best crystals were grown using reservoir solutions of 24 - 32 % (w/v) PEG 3350, 0.1 M Bis-Tris pH 6.5 and 0.1 M ammonium acetate. The optimized crystals appeared as rectangular blocks with maximum dimension of approximately 0.2 mm (Fig. 4.1). In preparation for low temperature data collection, the crystals were soaked in 30 % (w/v) PEG 3350, 0.1 M Bis-Tris pH 6.5, 0.1 M ammonium acetate and 20 % (v/v) PEG 200. The cryoprotected crystals were picked up with Hampton loops and plunged into liquid nitrogen.
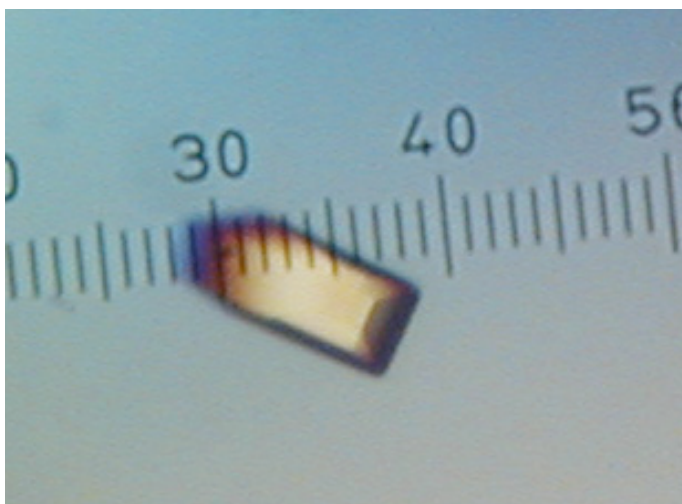
Fig. 4.1 A crystal of FtAcpC. The smallest dimension of the ruler corresponds to 20 μm.

FtAcpC crystals were analyzed at Advanced Light Source beamline 4.2.2 using a NOIR-1 CCD detector. Autoindexing calculations with D*TREK [24] and MOSFLM [25] indicated a C centered orthorhombic lattice with unit-cell lengths of *a*

= 59.2 Å, *b* = 124.2 Å and *c* = 62.2 Å.   Using the method of Matthews [26], the

asymmetric unit is predicted to contain one FtAcpC molecule and 46 % solvent.

A data set consisting of 115 frames was collected with detector distance of 150

mm, oscillation width of 1.0°/frame and exposure time of 5 s/frame. The data set

was integrated with MOSFLM through the iMosflm graphical interface and scaled

to 2.0 Å resolution with SCALA [27] using the CCP4i interface [28].   Data processing

statistics are listed in Table 4.1.

Table 4.1 Data-processing statistics for FtAcpC

Space group	$C222_1$
Wavelength (Å)	0.97909
Unit-cell parameters (Å, °)	*a* = 59.2, *b* = 124.2, *c* = 62.2
Protein molecules in asymmetric unit	1
V_m (Å^3/Da)	2.3
Solvent content (%)	46
Resolution (Å)	27.8 - 2.00 (2.11 - 2.00)
Total observations	70800
Unique reflections	15842
Redundancy	4.5 (4.0)
Completeness (%)	99.8 (99.1)
Mean $I/\sigma(I)$	16.3 (3.8)
R_{merge}†	0.073 (0.470)
R_{merge}† in low-resolution bin	0.034

$^†R_{merge} = S_{hkl}S_i \mid I_i(hkl) - \langle I(hkl) \rangle \mid / S_{hkl}S_i \, I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations of reflection *hkl*.

Values in parentheses are for the outer resolution shell of data

4.2.6 Crystallization of PmAcpC and preliminary analysis of X-ray diffraction data

Crystallization trials were performed at 293 K using the sitting drop method (Cryschem plates, 1 mL reservoir volume) with drops formed by mixing 5 μl of the protein stock solution and 5μl of the reservoir solution. Initial crystal screening trials revealed several promising conditions having PEG as the precipitating agent. The crystals invariably appeared as clusters of needles, and although extensive optimization trials reduced the degree of fusing, it was not possible to grow truly single crystals. For example, the best crystals, which were grown using a reservoir solution of 20 % (w/v) PEG 3350, 0.2 M ammonium citrate dibasic and 10% (v/v) n-propanol, still exhibited significant clustering and needle-like morphology (Fig. 4.2). In preparation for low temperature data collection, the crystals were soaked in 25 % (w/v) PEG 3350, 0.2 M ammonium citrate dibasic and 25 % (v/v) PEG 200. The clusters were teased apart with Hampton loops in an attempt to isolate single crystals for data collection. The cryoprotected crystals were then plunged into liquid nitrogen.

X-ray diffraction data were collected at Advanced Light Source beamline 4.2.2 using a NOIR-1 CCD detector. The diffraction images clearly exhibited multiple interpenetrating reciprocal space lattices (Fig. 4.3*a*), particularly at low resolution (Fig. 4.3*b*). The interpenetrating reciprocal lattices are symptomatic of epitaxial twinning and are due to multiple real space lattices in the sample that are not perfectly aligned [29]. The crystals diffracted to surprisingly high resolution despite being rather thin.
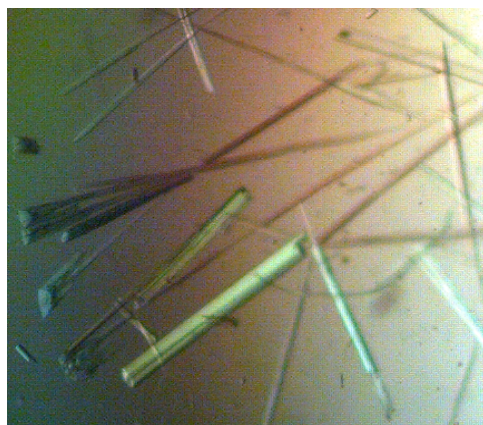
Fig. 4.2 Crystals of PmAcpC. Note that the crystals grow as clusters of needles. The thickest clusters are approximately 20 - 40 μm thick.

As shown in Fig. 4.3*c*, reflections extending beyond 1.84 Å resolution were observed. Despite the epitaxial twinning, autoindexing using either D*TREK [24] or MOSFLM was robust and consistently indicated C2 as the space group with unit-cell parameters of *a* = 80.0 Å, *b* = 106.1 Å, *c* = 89.7 Å and β = 93.1°. The assumption of three PmAcpC molecules in the asymmetric unit implies 45 % solvent and V_m = 2.2 Å^3/Da [26]. A data set consisting of 360 frames with oscillation width of 0.5°/frame and detector distance of 140 mm was collected. Integration was performed with MOSFLM through the iMosflm interface and scaling calculations were performed with SCALA via CCP4i. The data could be processed satisfactorily to a resolution of 1.85 Å resolution, as shown in Table 4.2. Note, however, that R_{merge} is 0.065 in the low-resolution bin, which is rather high. Presumably, this suboptimal result is due to the presence of multiple reciprocal space lattices, which are particularly evident at low resolution (Fig.

4.3*b*). For comparison, R_{merge} is 0.034 in the low-resolution bin for the FtAcpC

data (Table 4.1).
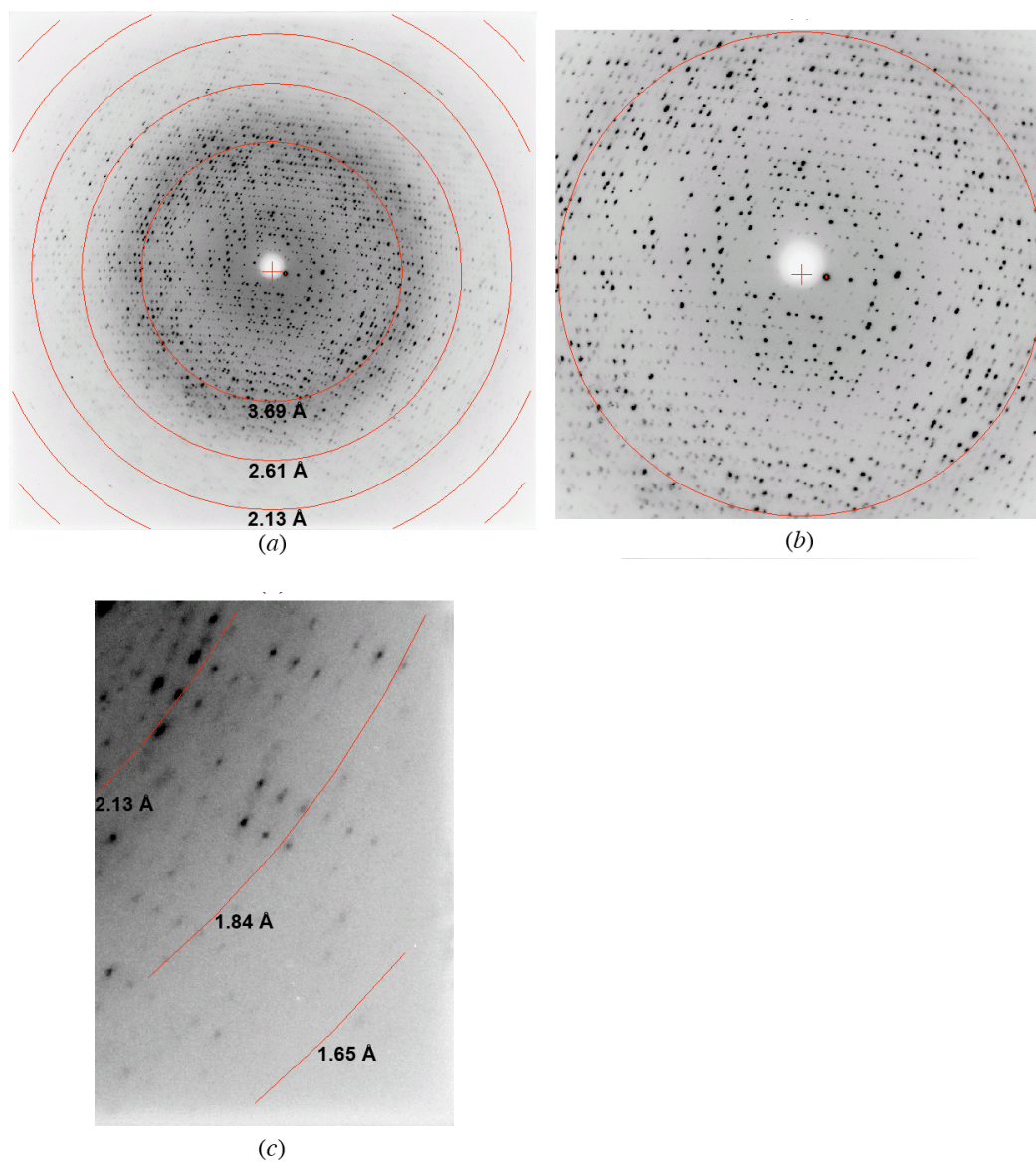
Fig. 4.3 Diffraction image collected from a PmAcpC crystal. Panel (*a*) shows the entire resolution range. (*b*) Enlarged view of the low resolution region. The circle corresponds to 3.69 Å resolution. (*c*) Enlarged view of the high resolution data.

Table 4.2   Data-processing statistics for PmAcpC

Space group	*C*2
Wavelength (Å)	1.00000
Unit-cell parameters (Å, °)	*a* = 80.0, *b* = 106.1, *c* = 89.7, *â* = 93.1°
Protein molecules in asymmetric unit	3
V_m (Å^3/Da)	2.2
Solvent content (%)	45
Resolution (Å)	27.5 - 1.85 (1.95 - 1.85)
Total observations	215684
Unique reflections	62294
Redundancy	3.5 (2.5)
Completeness (%)	97.8 (86.7)
Mean *I/σ(I)*	10.4 (3.1)
R_{merge}†	0.094 (0.272)
R_{merge}† in low-resolution bin	0.065

†R_{merge} = $S_{hkl}S_i$ | $I_i(hkl)$ - 〈 $I(hkl)$ 〉 | / $S_{hkl}S_i$ $I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and 〈 $I(hkl)$ 〉 is the weighted average intensity for all observations of reflection *hkl*.

Values in parentheses are for the outer resolution shell of data.

4.3 References

1.   Vincent, J. B.; Crowder, M. W.; Averill, B. A., Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions. *Trends Biochem. Sci.* **1992,** *17* (3), 105-10.

2.   Rossolini, G. M.; Schippa, S.; Riccio, M. L.; Berlutti, F.; Macaskie, L. E.; Thaller, M. C., Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. *Cell. Mol. Life Sci.* **1998,** *54* (8), 833-850.

3.   Thaller, M. C.; Schippa, S.; Rossolini, G. M., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci.* **1998,** *7* (7), 1647-52.

4.   (a) Reilly, T. J.; Chance, D. L.; Smith, A. L., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J. Bacteriol.* **1999,** *181* (21), 6797-6805; (b) Reilly, T. J.; Smith, A. L., Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr. Purif.* **1999,** *17* (3), 401-409.

5.   Felts, R. L.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Cloning, purification and crystallization of Bacillus anthracis class C acid phosphatase. *Acta Cryst.* **2006,** *F62* (Pt 7), 705-8.

6.   Malke, H., Cytoplasmic membrane lipoprotein LppC of Streptococcus equisimilis functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998,** *64* (7), 2439-42.

7.   du Plessis, E. M.; Theron, J.; Joubert, L.; Lotter, T.; Watson, T. G., Characterization of a phosphatase secreted by Staphylococcus aureus strain 154, a new member of the bacterial class C family of nonspecific acid phosphatases. *Syst. Appl. Microbiol.* **2002,** *25* (1), 21-30.

8.   Reilly, T. J.; Calcutt, M. J., The class C acid phosphatase of Helicobacter pylori is a 5' nucleotidase. *Protein Expr. Purif.* **2004,** *33* (1), 48-56.

9.   Passariello, C.; Schippa, S.; Iori, P.; Berlutti, F.; Thaller, M. C.; Rossolini, G. M., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5'-nucleotides. *Biochim. Biophys. Acta* **2003,** *1648* (1-2), 203-9.

10.   Reilly, T. J.; Chance, D. L.; Calcutt, M. J.; Tanner, J. J.; Felts, R. L.; Waller, S. C.; Henzl, M. T.; Mawhinney, T. P.; Ganjam, I. K.; Fales, W. H.,

Characterization of a unique class C acid phosphatase of Clostridium perfringens. *Appl. Environ. Microbiol.* **2008**, Submitted.

11. (a) Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of Recombinant Haemophilus Influenzae e (P4) Acid Phosphatase Reveals a New Member of the Haloacid Dehalogenase Superfamily. *Biochemistry* **2007,** *46* (39), 11110-9; (b) Ou, Z.; Felts, R. L.; Reilly, T. J.; Nix, J. C.; Tanner, J. J., Crystallization of recombinant Haemophilus influenzae e (P4) acid phosphatase. *Acta Cryst.* **2006,** *F62* (Pt 5), 464-6.

12. (a) Hotomi, M.; Ikeda, Y.; Suzumoto, M.; Yamauchi, K.; Green, B. A.; Zlotnick, G.; Billal, D. S.; Shimada, J.; Fujihara, K.; Yamanaka, N., A recombinant P4 protein of Haemophilus influenzae induces specific immune responses biologically active against nasopharyngeal colonization in mice after intranasal immunization. *Vaccine* **2005,** *23* (10), 1294-300; (b) Mason, K. W.; Zhu, D.; Scheuer, C. A.; McMichael, J. C.; Zlotnick, G. W.; Green, B. A., Reduction of nasal colonization of nontypeable Haemophilus influenzae following intranasal immunization with rLP4/rLP6/UspA2 proteins combined with aqueous formulation of RC529. *Vaccine* **2004,** *22* (25-26), 3449-56; (c) Green, B. A.; Baranyi, E.; Reilly, T. J.; Smith, A. L.; Zlotnick, G. W., Certain site-directed, nonenzymatically active mutants of the Haemophilus influenzae P4 lipoprotein are able to elicit bactericidal antibodies. *Infect. Immun.* **2005,** *73* (7), 4454-4457.

13. Mohapatra, N. P.; Soni, S.; Reilly, T. J.; Liu, J.; Klose, K. E.; Gunn, J. S., Combined deletion of four Francisella novicida acid phosphatases attenuates virulence and macrophage vacuolar escape. *Infect. Immun.* **2008,** *76* (8), 3690-9.

14. (a) Adcock, P. W.; Saint, C. P., Rapid confirmation of Clostridium perfringens by using chromogenic and fluorogenic substrates. *Appl. Environ. Microbiol.* **2001,** *67* (9), 4382-4; (b) Eisgruber, H.; Geppert, P.; Sperner, B.; Stolle, A., Evaluation of different methods for the detection of Clostridium perfringens phosphatases. *Int. J. Food Microbiol.* **2003,** *82* (1), 81-6; (c) Sartory, D. P.; Waldock, R.; Davies, C. E.; Field, A. M., Evaluation of acid phosphatase as a confirmation test for Clostridium perfringens isolated from water. *Lett. Appl. Microbiol.* **2006,** *42* (4), 418-24.

15. Harper, M.; Boyce, J. D.; Adler, B., Pasteurella multocida pathogenesis: 125 years after Pasteur. *FEMS Microbiol. Lett.* **2006,** *265* (1), 1-10.

16. Bendtsen, J. D.; Nielsen, H.; von Heijne, G.; Brunak, S., Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004,** *340* (4), 783-95.

17. Frank, K.; Sippl, M. J., High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics* **2008,** *24* (19), 2172-6.

18. Nakai, K.; Kanehisa, M., Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* **1991,** *11* (2), 95-110.

19.   Larsson, P.; Oyston, P. C.; Chain, P.; Chu, M. C.; Duffield, M.; Fuxelius, H. H.; Garcia, E.; Halltorp, G.; Johansson, D.; Isherwood, K. E.; Karp, P. D.; Larsson, E.; Liu, Y.; Michell, S.; Prior, J.; Prior, R.; Malfatti, S.; Sjostedt, A.; Svensson, K.; Thompson, N.; Vergez, L.; Wagg, J. K.; Wren, B. W.; Lindler, L. E.; Andersson, S. G.; Forsman, M.; Titball, R. W., The complete genome sequence of Francisella tularensis, the causative agent of tularemia. *Nat. Genet.* **2005,** *37* (2), 153-9.

20.   May, B. J.; Zhang, Q.; Li, L. L.; Paustian, M. L.; Whittam, T. S.; Kapur, V., Complete genomic sequence of Pasteurella multocida, Pm70. *Proc. Natl. Acad. Sci. USA* **2001,** *98* (6), 3460-5.

21.   Hayashi, S.; Wu, H. C., Lipoproteins in bacteria. *J. Bioenerg. Biomembr.* **1990,** *22* (3), 451-71.

22.   Reilly, T. J.; Felts, R. L.; Henzl, M. T.; Calcutt, M. J.; Tanner, J. J., Characterization of recombinant Francisella tularensis acid phosphatase A. *Protein Expr. Purif.* **2006,** *45* (1), 132-141.

23.   Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005,** *41* (1), 207-34.

24.   Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Cryst.* **1999,** *D55* (Pt 10), 1718-25.
25.   (a) Leslie, A. G., The integration of macromolecular diffraction data. *Acta Cryst.* **2006,** *D62* (Pt 1), 48-57; (b) Leslie, A. G. W., *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography No. 26. Recent changes to the MOSFLM package for processing film and image plate data*. 1992; Vol. 26.

26.   (a) Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968,** *33*, 491-497; (b) Kantardjieff, K. A.; Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **2003,** *12* (9), 1865-71.

27.   Evans, P., Scaling and assessment of data quality. *Acta Cryst.* **2006,** *D62* (Pt 1), 72-82.

28.   Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson, E., A graphical user interface to the CCP4 program suite. *Acta Cryst.* **2003,** *D59* (Pt 7), 1131-7.

29.   (a) Yeates, T. O.; Fam, B. C., Protein crystals and their evil twins. *Structure Fold. Des.* **1999,** *7* (2), R25-9; (b) Chandra, N.; Acharya, K. R.; Moody, P. C., Analysis and characterization of data from twinned crystals. *Acta Cryst.* **1999,** *D55*, 1750-8.

Chapter 5.

Crystal structure and immunogenicity of the class C acid phosphatase from Pasteurella multocida

Harkewal Singh, Thomas J. Malinski, Thomas J. Reilly, Michael T. Henzl,

and John J. Tanner

Author contribution – H.S. expressed and purified PmCCAP and P4. He also determined the crystal structure and peformed enzyme kinetics experiments.

Abstract

Pasteurella multocida is a pathogen of veterinary and medical importance. Here, we report the 1.85 Å resolution crystal structure of the class C acid phosphatase from this organism (denoted rPmCCAP). The structure shows that rPmCCAP exhibits the same haloacid dehalogenase fold and dimeric assembly as the class C enzyme from *Haemophilus influenzae*. Formation of the dimer in solution is demonstrated using analytical ultracentrifugation. The active site is devoid of a magnesium ion due to the presence of citrate in the crystallization buffer. Absence of the metal ion minimally perturbs the active site structure, which suggests that the main role of the ion is to balance the negative charge of the substrate rather than stabilize the active site structure. The crystal lattice displays unusual crystal packing involving the C-terminal polyhistidine tag mimicking the substrate. Steady-state kinetic constants are determined for the substrates NMN, 5′-AMP, 3′-AMP, 2′-AMP, and *p*-nitrophenyl phosphate. The highest catalytic efficiency is observed with NMN. The production of polyclonal anti-rPmCCAP antibodies is demonstrated, and these antibodies are shown to cross-react with the *H. influenzae* class C phosphatase. The antibodies are used to detect PmCCAP in clinical *P. multocida* and *Mannheimia haemolytica* strains cultured from infected animals.

5.1 Introduction

Class C acid phosphatases (CCAPs) are outer membrane acid phosphatases that contain four essential Asp residues imbedded in the bipartite sequence motif of [IV]-[VAL]-**D**-[IL]-**D**-E-T-[VM]-L-X-[NT]-X(2)-Y in the N-terminal half of the polypeptide chain and [IV]-[LM]-X(2)-G-**D**-[NT]-L-X-**D**-F in the C-terminal half [1]. Because of their localization to the bacterial outer membrane, CCAPs are potential candidates for vaccine development, and some progress has been made toward creating a vaccine against nontypeable *Haemophilus influenzae* using catalytically inactive mutants of the *H. influenzae* CCAP [2]. Biochemical studies have shown that CCAPs typically hydrolyze a variety of aryl phosphatases as well as nucleoside 5′- and 3′- monophosphates [3]. Crystallographic studies of the recombinant CCAP from *H. influenzae*, a.k.a. rP4, have shown that CCAPs belong to the haloacid dehalogenase (HAD) structural superfamily and revealed the basis of selectivity for nucleoside monophosphate substrates [3e, 3k]. The structural and biochemical data also showed that the first of the four conserved Asp residues is the nucleophile that attacks the P atom of the substrate, the second one functions as the acid that protonates the leaving group, and all four Asp residues contribute to the binding of a Mg^{2+} ion.

Previously, it was shown that the *Pasteurella multocida* genome encodes a protein that shares 58 % global amino acid sequence identity with rP4 suggesting that the organism produces a CCAP [4]. This finding is potentially significant, because *P. multocida* is a pathogen of medical and veterinary importance [5]. The

organism is a ubiquitous Gram-negative, facultative anaerobic bacterium and is often a constituent of the normal respiratory microbiota, but is also capable of causing pneumonia, osteomyelitis, and endocarditis and meningitis infections. In humans, *P. multocida* infections are common from dog and cat bites. *P. multocida* and other *Pasteurellaceae*, including *Mannheimia haemolytica* (formerly *Pasteurella haemolytica*), are the main causative agents of bovine respiratory disease, which causes annual losses in the U.S. animal agriculture industry of $750 million to $1 billion [6]. Although there has been some progress in understanding the factors involved in the pathogenesis of *P. multocida,* including the polysaccharide capsule, lipopolysaccharide, outer membrane proteins, fimbriae and adhesions, there remains a lack of knowledge about the host-pathogen relationship [5].

Given the value of outer membrane proteins as potential targets for vaccine development, we previously cloned and expressed a recombinant form of the predicted CCAP from *P. multocida* [4]. Here report the crystal structure of this protein (denoted rPmCCAP), show that the enzyme forms a dimer in solution and exhibits nucleotidase activity, and demonstrate the production of anti-rPmCCAP polyclonal antibodies.

5.2 Material and methods

5.2.1 Crystal structure determination

Methods for the expression, purification, and crystallization of rPmCCAP were reported previously [4]. Briefly, the crystals were grown in sitting drops at room temperature over a reservoir containing 20% (w/v) PEG 3350, 0.2 M ammonium citrate, and 10% (v/v) n-propanol. The space group is *C*2 with unit cell parameters of *a* = 80.0, *b* = 106.1, *c* = 89.7, and ß = 93.1°. The 1.85 Å resolution X-ray diffraction data set used for structure determination was collected from an epitaxially-twinned crystal at beamline 4.2.2 of the Advanced Light Source as described previously [4].

The structure was solved using molecular replacement phasing as implemented in MOLREP [7]. The search model was a monomer extracted from the coordinates of rP4 (PDB code 3ET4), which shares 58 % amino acid sequence identity with rPmCCAP. A clear solution having three molecules in the asymmetric unit was obtained. Based on the method of Matthews [8], the solvent content is estimated to be 44 % with a V_m value of 2.2 $Å^3$/Da. The model from molecular replacement was refined with simulated annealing in PHENIX [9]. The refined model was used to calculate phases, which were input to ARP/wARP [10] for automated building and sequence assignment. The model from ARP/wARP was used as the starting point for several rounds of manual model building in COOT [11] followed by refinement in PHENIX. Refinement statistics for the final model are listed in Table 5.1. The pairwise root mean square deviations for Cα

atoms of the three chains are 0.21 - 0.33 Å, indicating that the overall conformations are identical within experimental error.

5.2.2 Enzyme activity assays

Steady-state kinetic constants were determined using a discontinuous assay that measures the production of inorganic phosphate [12]. The assays were performed at 37 °C in 200 mM sodium acetate and 1 mM $MgCl_2$ at pH 5.5 (total assay volume of 0.5 mL). The pH of 5.5 was chosen based on measurements of the activity as a function of pH using NMN as substrate, which showed a plateau of maximal activity at pH 5.0 - 6.0 (data not shown). The enzyme concentration was 0.80 µM when using NMN as the substrate, 1.2 µM for *p*-nitrophenyl phosphate (pNPP), 4.0 µM for 5´-AMP and 3´-AMP, and 6.0 µM for and 2´-AMP. For each substrate concentration, the reaction was stopped by adding 1 mL of a quenching/detection solution (0.045 % (w/v) malachite green and 4.2% (w/v) ammonium molybdate in 4N HCl) after reaction times of 15 s, 75 s, 135 s and 195 s, and the citrate color development reagent (34 % (w/v) sodium citrate) was added 70 s after stopping each reaction. After 30 minutes, the inorganic phosphate concentrations were determined spectrophotometrically at 625 nm by reference to a standard curve constructed from solutions of known phosphate concentration. The initial rate was estimated by fitting data from the four time points to a line. Apparent values of K_m and V_{max} were estimated by fitting the initial rate data to the Michaelis-Menten equation using Origin 8.5 software. For

each substrate tested, a preliminary data set was collected in order to estimate K_m, and then a second data set was collected using a substrate range that spanned from less than $0.3K_m$ to at least $3K_m$. The reported kinetic constants were determined from the second data set.

5.2.3 Analytical ultracentrifugation

Equilibrium sedimentation data were acquired at 20 °C using a Beckman XL-I Optima analytical ultracentrifuge equipped with an An50Ti rotor. Prior to centrifugation, the protein sample was dialyzed into a reference buffer of 50 mM acetate buffer at pH 6.0. Absorbance data (280 nm) were collected at three protein concentrations and three rotor speeds (11,000, 15,000, and 22,000 rpm). The 9 sets of data were fit globally using Origin 8 to an ideal single-species model (equation 9 of [13]). A solvent density of 1.00 g/cm^3 was employed for the calculations, and the partial specific volume was assumed to be 0.73 cm^3/g.

Table 5.1. Data collection and refinement statistics[a]

Space group	*C*2
Wavelength (Å)	1.00000
Data collection resolution (Å)	27 - 1.85 (1.95 - 1.85)
Unit-cell parameters (Å, °)	*a* = 80.0, *b* = 106.1, *c* = 89.7, ß = 93.1°
No. of observations	215684
No. of unique reflections	62294
$R_{merge}(I)$	0.094 (0.272)
Average I/σ	10.4 (3.1)
Completeness (%)	97.8 (86.7)
Redundancy	3.5 (2.5)
Refinement resolution (Å)	27 - 1.85 (1.88 - 1.85)
R_{cryst}	0.211 (0.302)
R_{free}[b]	0.267 (0.385)
No. of protein residues	754
No. of protein atoms	5814
No. of water molecules	583
Average B-factor (Å^2)	
Protein	23.2
Water	27.1
rmsd[c]	
Bonds (Å)	0.006
Angles (°)	0.93
Ramachandran plot[d]	
Favored (%)	97.7
Allowed (%)	2.3
Coordinate error (Å)[e]	0.29
PDB code	3PCT

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]Random 5 % test set.
[c]Compared to the parameters of Engh and Huber [18].
[d]The Ramachandran plot was generated with RAMPAGE [19].
[e]Maximum likelihood-based coordinate error estimate.

5.2.4 Immunological methods

Rabbit anti-rPmCCAP antiserum was obtained using standard immunological methods. Laboratory animals were utilized with approval of the University of Missouri Animal Care and Use Committee. Purified rPmCCAP (~250 µg) was dialyzed against 0.9 % NaCl, filter-sterilized, and emulsified in Freund's complete adjuvant. The immunogen was then injected subcutaneously at multiple sites into a New Zealand white rabbit. Approximately 24 days after primary immunization, the response was boosted by single subcutaneous injection with similarly treated purified enzyme. Serum was collected by ear puncture 14 days following secondary immunization.

For Western blot detection of purified rPmCCAP and rP4, the protein was subjected to SDS-PAGE and then transferred to a Nitrocellulose (Bio Trace™ Pall Life Science) membrane in G-Biosciences 1X Tris-Glycine transfer buffer at 100 mA for 4 hr in an Owl Bandit Transblot apparatus. Rabbit anti-rPmCCAP polyclonal antibodies were used as primary antibody and detected with goat anti-rabbit or anti-mouse IgG (H+L) conjugated to alkaline phosphatase. Purified rP4 was obtained as previously described [3d].

5.2.5 Detection of CCAP in clinical Pasteurella strains

Two *P. multocida* strains and one *M. haemolytica* strain were cultured from samples submitted to the Veterinary Medical Diagnostic Laboratory at the

University Of Missouri College Of Veterinary Medicine. The two *P. multocida* samples originated from avian and porcine sources. The *M. haemolytica* strain originated from a bovine source. The strains were plated onto blood agar plates and incubated at 37 °C. After incubation, a colony was picked, resuspended in 10 mL of BHI broth and incubated overnight at 37°C under aerated conditions. The cells were harvested, disrupted using freeze/thaw cycles, and centrifuged. The supernatant was discarded, and the pellet was suspended in buffer and analyzed with Western blotting using rabbit anti-rPmCCAP IgG.

5.3 Results and discussion

5.3.1 Overall fold and active site structure

The fold of rPmCCAP comprises an α/β core domain situated below an α-helical cap domain and is almost identical to that of rP4 (Fig. 5.1A). The two enzymes superimpose with a root mean square deviation of 0.66 Å for Cα atoms.

The structural similarity of rPmCCAP to rP4 extends to the details of the active site, except for the Mg^{2+} ion, which is replaced by a water molecule in the rPmCCAP structure. Structures of rP4 have shown that the lower part of the active site contains the substrate phosphoryl binding site, which includes an essential Mg^{2+} ion surrounded by the four Asp residues of the DDDD motif [3e, 3k]. The roles of these Asp residues in rPmCCAP can be inferred from structures of

108

rP4 complexed with substrates [3k] and more generally, studies of HAD superfamily enzymes [14].

The first Asp (Asp63 in rPmCCAP, Fig. 5.2A) is the nucleophile that attacks the substrate phosphoryl group. The second Asp, aka Asp+2, is positioned two residues upstream from the nucleophile (Asp65, Fig. 5.2A), and functions as a general acid-base that protonates the substrate leaving group and deprotonates the water molecule that attacks the phosphoenzyme intermediate. The other two Asp residues of the DDDD motif (Asp180, Asp184 in rPmCCAP) are expected to bind the Mg^{2+} ion. Electron density maps indeed displayed a peak at the location expected for the metal ion (Fig. 5.2A, feature denoted "wa"). Consistent with the known metal ion site of rP4, the electron density peak is surrounded by the carboxylate groups of Asp63 and Asp180 as well as the carbonyl of Asp65. However, when a Mg^{2+} ion was modelled at this location, the distances between the metal ion and these coordinating oxygen ligands refined to values in the range of 2.7 - 2.9 Å, which are significantly longer than the Mg^{2+}-O distance of 2.2 Å observed in the rP4 structure.

Fig. 5.1. Structure of rPmCCAP. (A) Ribbon drawing of the protomer highlighting domain structure and similarity to rP4. rPmCCAP and rP4 (PDB code 3ET4) are shown in silver and green, respectively. Side chains of the four conserved Asp residues (DDDD motif) of rPmCCAP are shown in yellow. (B) Ribbon drawing of a dimer of rPmCCAP. The two chains are colored blue and pink. Asp residues of the DDDD motif are colored yellow. The dashed curve represents a disordered loop (residues 52-56) of chain A. The two insets show close-up views of intersubunit interactions.
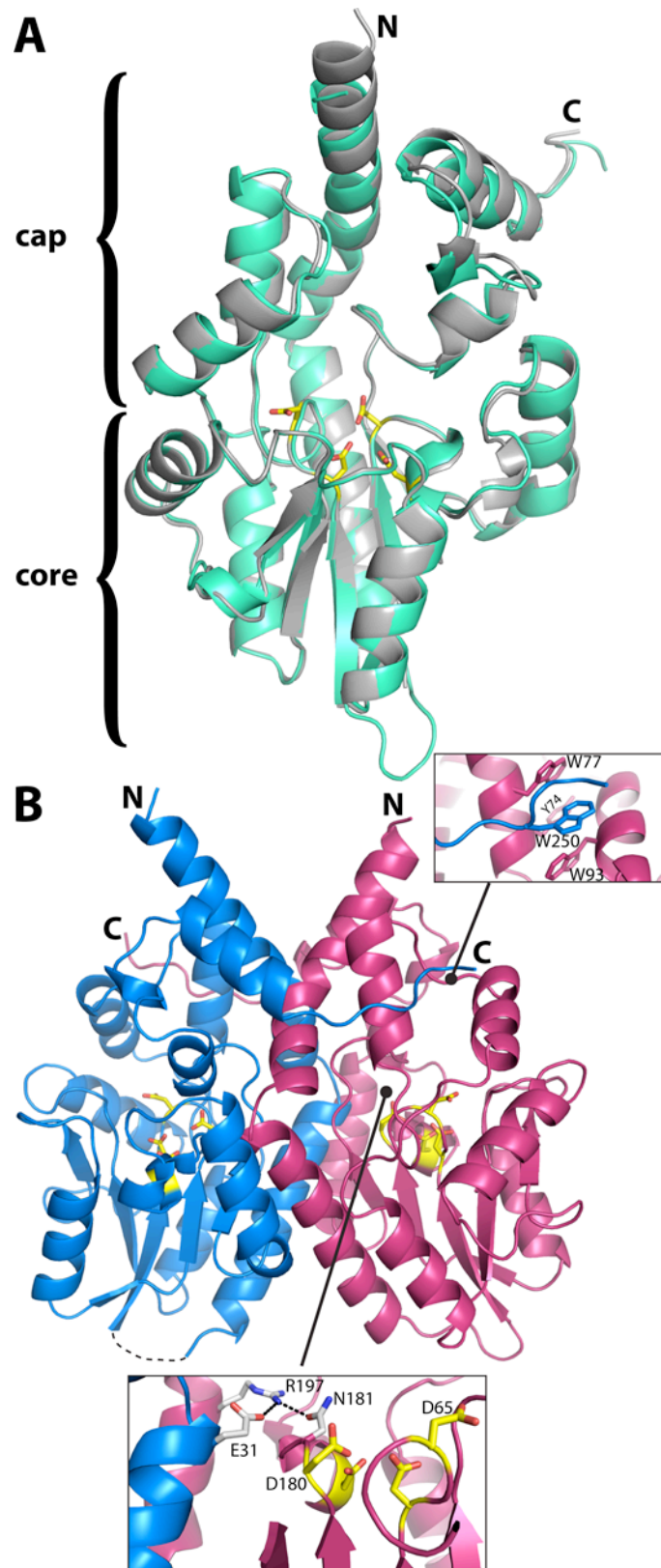
Furthermore, the electron density map did not have features representing the three water molecules that complete the octahedral coordination sphere of the metal ion in rP4. It was therefore concluded that the electron density feature in rPmCCAP represents a water molecule (wa in Fig. 5.2A) rather than a metal ion. This assignment is consistent with the fact that the crystallization and cryoprotection solutions contained 0.2 M citrate [4], a Mg^{2+} chelator.

Aside from the absence of a metal ion and its three coordinating water molecules, the active site of metal-free rPmCCAP is remarkably similar to that of Mg^{2+}-bound rP4 (Fig. 5.2B). The only perceptible difference in the protein conformation involves Asp63, which is the enzyme nucleophile. This side chain is rotated around χ_2 by 45° compared to Asp64 of rP4. The other three Asp residues display essentially the same conformations as in rP4 despite the absence of the metal ion. The structural integrity of the metal-free active site suggests that the metal ion plays a minimal role in organizing the active site residues for catalysis, and that its major role is to neutralize the negative charge of the substrate.
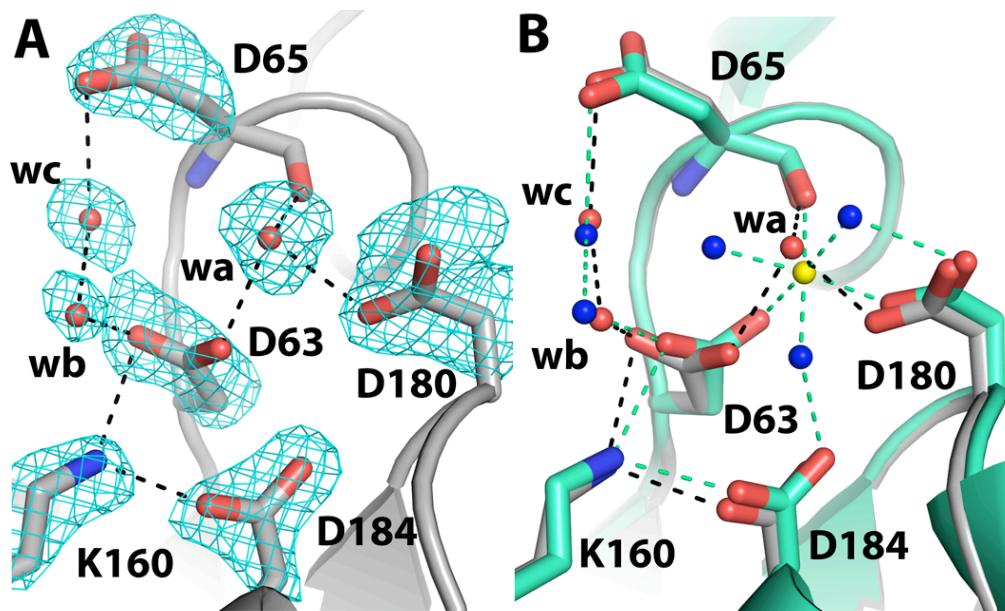
Fig. 5.2. Close-up views of the phosphoryl binding site of rPmCCAP. (A) Electron density near the phosphoryl binding site. The cage represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. Prior to map calculation, the side chains and water molecules shown were removed, and simulated annealing refinement was performed using PHENIX. The water molecule denoted wa occupies the expected Mg^{2+} site. (B) Comparison of rPmCCAP (silver) and rP4 (green, PDB code 3ET4). Water molecules of rPmCCAP are colored red, while those of rP4 are colored blue. The Mg^{2+} ion of rP4 is shown in yellow. Black and green dashed lines denote electrostatic interactions in rPmCCAP and rP4, respectively.

5.3.2 Oligomeric state

The oligomeric state of rPmCCAP in solution was determined with equilibrium analytical ultracentrifugation. Data from nine sedimentation experiments corresponding to three rotor speeds and three loading concentrations were fit globally to a single species model (Fig. 5.3).

112
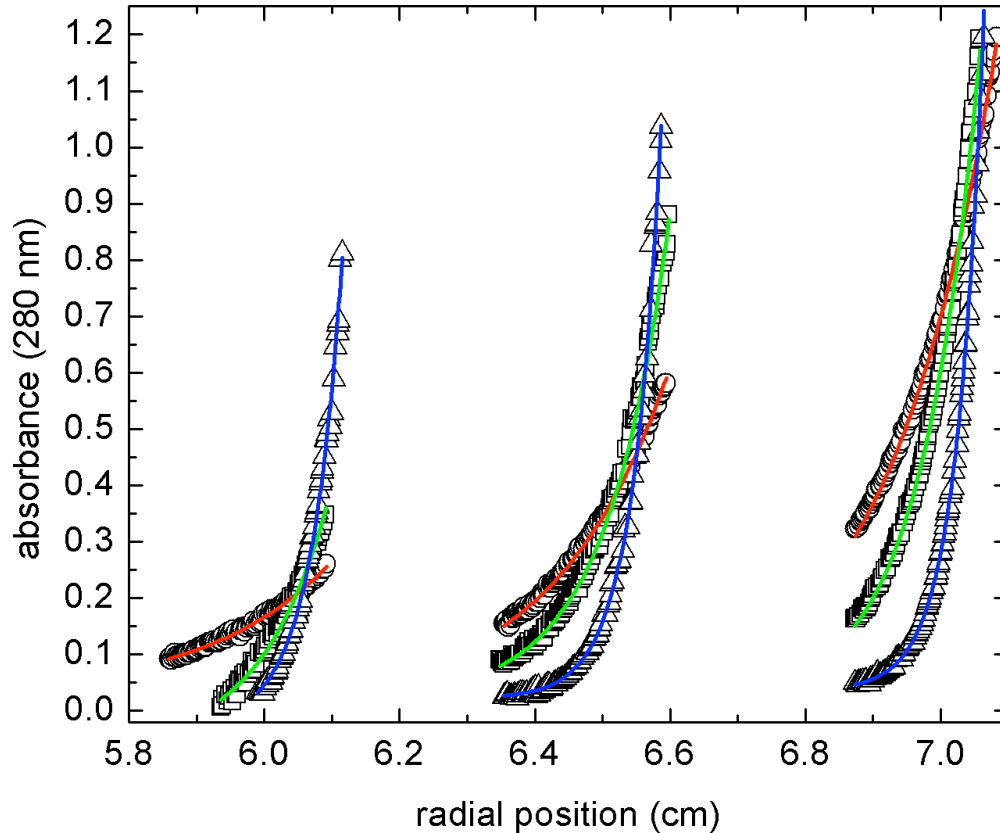
Fig. 5.3. Global fit of sedimentation equilibrium data to an ideal single-species model. Data were acquired at 11,000 (circles), 15,000 (squares), and 22,000 (triangles) rpm at three different loading concentrations. The 9 data sets were subjected to non-linear least squares minimization as described in the text. The solid curves indicate the best fit to the data (red, 11,000 rpm; green, 15,000 rpm; blue, 22,000 rpm).

The best-fit value of the molecular weight is 60.1 ± 0.2 kDa, which is within 3 % of the theoretical molecular weight of a dimer (58.1 kDa). This result is consistent with sedimentation and size exclusion chromatography data showing

113

that rP4 and the CCAPs from *Helicobacter pylori* and *Clostridium perfringens* are dimeric in solution [3b, 3h, i].

The dimeric assembly shown in Fig. 5.1B is most likely the one formed in solution. The asymmetric unit contains one such dimer, which is formed by chains A and B (AB dimer, Fig. 5.4A). There is a third molecule in the asymmetric unit (chain C), and rotation of it around the crystallographic 2-fold axis (-x, y, -z) followed by translation of (1, 0, 0) generates the CC′ dimer shown in Fig. 5.4A. Analysis of protein-protein interfaces using the PISA server [15] shows that this dimer has a buried interface area of 2900 $Å^2$, which is the largest interface in the lattice. The estimated free energy of dissociation from PISA (42 kcal/mol) suggests that the dimer is stable in solution. Furthermore, an identical dimer is found in both the tetragonal and hexagonal rP4 crystal lattices [3e, 3k], and several intersubunit interactions are observed in both enzymes. These conserved interactions include an intersubunit Glu-Arg-Asn electrostatic network that involves an active site loop (Fig. 5.1B, lower inset) and a hydrophobic interaction in which a tryptophan near the C-terminus of one protomer (Trp250 in rPmCCAP) packs against three aromatic residues (Tyr74, Trp77, Trp93) of the cap domain of the other protomer (Fig. 5.1B, upper inset). Taken together, these results suggest that the dimer of Fig. 5.1B represents the solution form of rPmCCAP and confirms this assembly as the dimer of rP4.
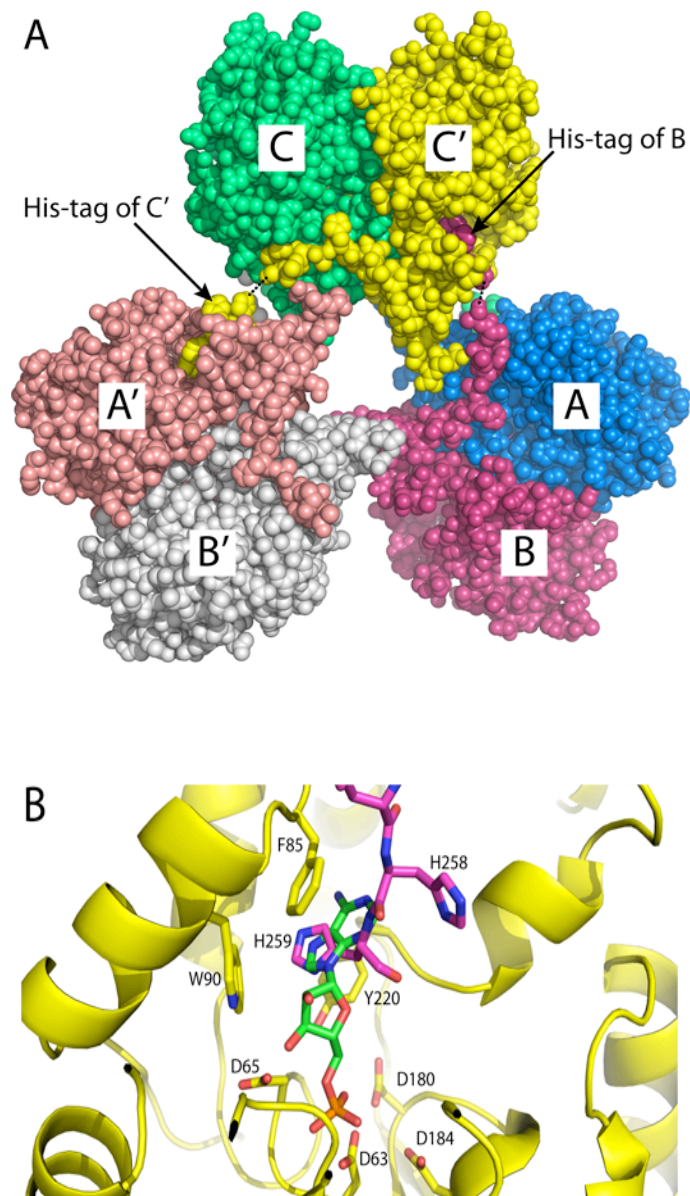
Fig. 5.4. Involvement of the C-terminal polyhistidine tag in crystal packing. (A) Section of the crystal lattice showing three dimers interacting via His-tags to form a hexameric assembly. The asymmetric unit consists of chains A, B, and C. Crystallographic symmetry was used to generate chains A′, B′, and C′. The dotted lines indicate disordered residues of chains B and C. The view is looking along the *a-c* diagonal with the *b*-axis vertical. (B) Close-up view of the binding of the tag of chain B (magenta) in the active site of chain C (yellow). 5′-AMP from the structure of a substrate-trapping mutant of rP4 complexed with 5′-AMP is shown in green (PDB code 3OCV).

5.3.3 Stabilization of the crystal lattice by the polyhistidine affinity tag

The crystal lattice exhibits an unusual case of the polyhistidine tag participating in crystal contacts (Fig. 5.4A). We note that the rPmCCAP gene was created so that the expressed protein has a non-cleavable hexahistidine tag following Lys253. Electron density was evident for the last five histidine residues of chain C (residues 255-259) but absent for Lys253 and His254. The connectivity of the C-terminus of chain B was better, with only His255 omitted from the final model. Electron density for the tag of chain A was not evident. As shown in Fig. 5.4A, the His-tag of chain B interacts with chain C (or equivalently, C´), while the tag of chain C (or C´) interacts with chain A (or A´), resulting in a hexameric assembly. The hexamer is an artifact of crystallization, since the sedimentation data showed that His-tagged rPmCCAP forms a dimer in solution.

The terminal residue His259 appears to mimic the base of nucleoside monophosphate substrates (Fig. 5.4B). Previous structures of a substrate-trapping mutant of rP4 complexed with nucleoside monophosphates showed that the base of the substrate binds in an aromatic box formed by Phe86, Trp91, and Tyr221 [3k]. The analogous residues of rPmCCAP are Phe85, Trp90, and Tyr220, and they form a substructure identical to that of rP4. The imidazole ring of His259 binds in the aromatic box and is sandwiched between Phe85 and Tyr220. This arrangement is reminiscent of the adenine ring of 5′-AMP binding in the aromatic box of rP4. Thus, substrate mimicry appears to contribute to the stabilization of the polyhistidine affinity tag, and hence, crystal formation.

We note that ordered polyhistidine tags in crystal structures are common. A recent query of the PDB using PDBeMotif [16] revealed 15 structures with at least 5 histidine residues built at the N-terminus and 69 structures with at least 5 histidine residues built at the C-terminus.

5.3.4 Kinetic analysis

Preliminary activity assays showed that the enzyme exhibited activity with nucleoside monophosphate substrates as well as the synthetic aryl phosphate substrates pNPP and 4-methylumbelliferyl phosphate. The steady-state kinetic constants were measured at 37°C and pH 5.5 for the substrates NMN, 5′-AMP, 3′-AMP, 2′-AMP, and pNPP (Table 5.2). Among the nucleoside monophosphate substrates tested, the enzyme exhibits the highest catalytic efficiency (k_{cat}/K_m) for NMN, followed by 5′-AMP, 3′-AMP, and 2′-AMP. This rank ordering is the same as that of rP4 [3k]. The lowest K_m was obtained with 5′-AMP (0.08 mM), and the highest was obtained with 2′-AMP (1.4 mM). This trend was also observed for rP4. The observation that rPmCCAP catalyzes the hydrolysis of nucleoside 5′- and 3′-monophosphates is consistent with other kinetic analyses suggesting that CCAPs are nonspecific nucleotidases [3a-j].

Table 5.2 Kinetic parameters for rPmCCAP.

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}M^{-1})
NMN	0.29 ± 0.06	4.4 ± 0.3	15000 ± 3300
pNPP	0.5 ± 0.1	6.2 ± 0.5	12000 ± 2700
5'-AMP	0.08 ± 0.01	0.18 ± 0.01	2300 ± 310
3'-AMP	0.8 ± 0.2	0.19 ± 0.01	240 ± 61
2'-AMP	1.4 ± 0.3	0.13 ± 0.01	90 ± 21

5.3.5 Detection of proteins using anti-rPmCCAP IgG

Since antibodies against outer membrane proteins are potentially useful in vaccine development, we raised polyclonal anti-rPmCCAP antibodies in a rabbit. Western blotting confirms that the anti-rPmCCAP antibodies react with rPmCCAP (Fig. 5.5A, lane 1). The high structural similarity between rPmCCAP and rP4 prompted us to test whether anti-rPmCCAP antibodies cross-react with rP4. As shown in Fig. 5.5A (lane 3), anti-rPmCCAP antibodies also recognize rP4.

The rabbit anti-rPmCCAP IgG was used to detect PmCCAP in clinical strains. The strains of *P. multocida* tested were cultured from avian and porcine sources. The cells were disrupted, and the pellet obtained by centrifugation, which presumably includes membrane-anchored proteins, was assayed using Western

blotting. As shown in Fig. 5.5B, PmCCAP is detected in both clinical strains (lanes 2 and 3). These results are consistent with the prediction that PmCCAP is lipidated and exported from the cytoplasm [4]. Further work will be needed to determine whether the enzyme is associated with the inner or outer membrane, but the high sequence identity to P4 implies that it is an outer membrane protein.
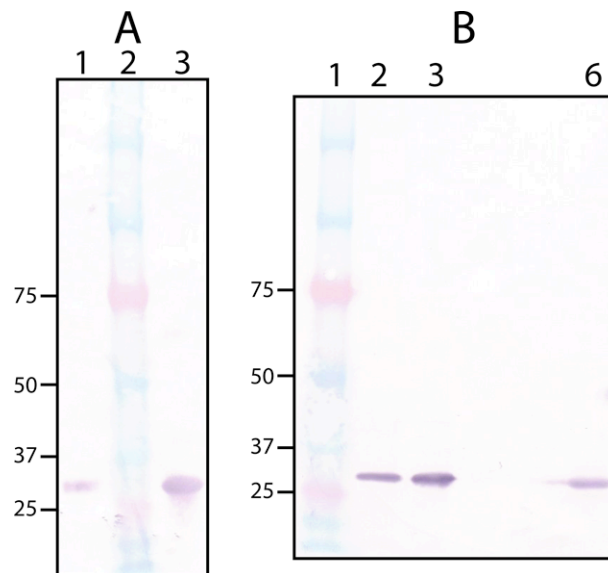
Fig. 5.5. Western blot analysis using anti-rPmCCAP IgG. (A) Detection of purified rPmCCAP (lane 1) and rP4 (lane 3) using a 1:80,000 antiserum dilution. Lane 2 contains molecular weight standards. (B) Analysis of clinical *P. multocida* and *M. haemolytica* clinical strains. Lane 1 contains molecular weight standards. Lanes 2 and 3 contain cell pellet fractions of *P. multocida* clinical strains cultured from avian and porcine sources, respectively. Lane 6 contains the cell pellet fraction from a *M. haemolytica* clinical strain (bovine source). The antiserum was diluted 1:160,000 in each case.

Anti-rPmCCAP was also used to detect the CCAP in *M. haemolytica* cultured from an infected bovine source. Formerly known as *Pasteurella haemolytica*, *M.*

119

haemolytica is closely related to *P. multocida*. Comparison of the PmCCAP amino acid sequence with the non-redundant sequence database using BLAST [17] revealed a predicted *M. haemolytica* CCAP having 65 % global sequence identity with PmCCAP (RefSeq accession number ZP_05990345). As shown in Fig. 5.5B (lane 6), Western blot analysis of the *M. haemolytica* clinical strain using anti-rPmCCAP was positive, consistent with the prediction from sequence analysis.

Antibodies against rPmCCAP may have clinical utility. The use of rP4 as a component of a recombinant vaccine against nontypeable *H. influenzae* [2] suggests the potential for developing a vaccine to protect animals against *P. multocida* infection. Another possibility is to develop a diagnostic test for *P. multocida* and closely related pathogens using anti-rPmCCAP. The structure reported here should aid these efforts by providing a three-dimensional structural basis for the design of catalytically inactive mutant enzymes having enhanced immunogenicity.

5.4 References

1.   Thaller, M. C.; Schippa, S.; Rossolini, G. M., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci.* **1998,** *7* (7), 1647-52.

2.   (a) Hotomi, M.; Ikeda, Y.; Suzumoto, M.; Yamauchi, K.; Green, B. A.; Zlotnick, G.; Billal, D. S.; Shimada, J.; Fujihara, K.; Yamanaka, N., A recombinant P4 protein of Haemophilus influenzae induces specific immune responses biologically active against nasopharyngeal colonization in mice after intranasal immunization. *Vaccine* **2005,** *23* (10), 1294-300; (b) Mason, K. W.; Zhu, D.; Scheuer, C. A.; McMichael, J. C.; Zlotnick, G. W.; Green, B. A., Reduction of nasal colonization of nontypeable Haemophilus influenzae following intranasal immunization with rLP4/rLP6/UspA2 proteins combined with aqueous formulation of RC529. *Vaccine* **2004,** *22* (25-26), 3449-56; (c) Green, B. A.; Baranyi, E.; Reilly, T. J.; Smith, A. L.; Zlotnick, G. W., Certain site-directed, nonenzymatically active mutants of the Haemophilus influenzae P4 lipoprotein are able to elicit bactericidal antibodies. *Infect. Immun.* **2005,** *73* (7), 4454-4457.

3.   (a) Reilly, T. J.; Chance, D. L.; Smith, A. L., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J. Bacteriol.* **1999,** *181* (21), 6797-6805; (b) Reilly, T. J.; Smith, A. L., Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr. Purif.* **1999,** *17* (3), 401-409; (c) Reilly, T. J.; Green, B. A.; Zlotnick, G. W.; Smith, A. L., Contribution of the DDDD motif of H. influenzae e (P4) to phosphomonoesterase activity and heme transport. *FEBS Lett.* **2001,** *494* (1-2), 19-23; (d) Ou, Z.; Felts, R. L.; Reilly, T. J.; Nix, J. C.; Tanner, J. J., Crystallization of recombinant Haemophilus influenzae e (P4) acid phosphatase. *Acta Cryst.* **2006,** *F62* (Pt 5), 464-6; (e) Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of Recombinant Haemophilus Influenzae e (P4) Acid Phosphatase Reveals a New Member of the Haloacid Dehalogenase Superfamily. *Biochemistry* **2007,** *46* (39), 11110-9; (f) Passariello, C.; Schippa, S.; Iori, P.; Berlutti, F.; Thaller, M. C.; Rossolini, G. M., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5'-nucleotides. *Biochim. Biophys. Acta* **2003,** *1648* (1-2), 203-9; (g) Malke, H., Cytoplasmic membrane lipoprotein LppC of Streptococcus equisimilis functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998,** *64* (7), 2439-42; (h) Reilly, T. J.; Calcutt, M. J., The class C acid phosphatase of Helicobacter pylori is a 5' nucleotidase. *Protein Expr. Purif.* **2004,** *33* (1), 48-56; (i) Reilly, T. J.; Chance, D. L.; Calcutt, M. J.; Tanner, J. J.; Felts, R. L.; Waller, S. C.; Henzl, M. T.; Mawhinney, T. P.; Ganjam, I. K.; Fales, W. H., Characterization of a unique class C acid phosphatase from Clostridium perfringens. *Appl. Environ. Microbiol.* **2009,** *75* (11), 3745-54; (j) Wang, R.; Ohtani, K.; Wang, Y.; Yuan, Y.; Hassan, S.; Shimizu, T., Genetic and

biochemical analysis of a class C non-specific acid phosphatase (NSAP) of Clostridium perfringens. *Microbiology* **2010,** *156* (Pt 1), 167-73; (k) Singh, H.; Schuermann, J. P.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Recognition of Nucleoside Monophosphate Substrates by Haemophilus influenzae Class C Acid Phosphatase. *J. Mol. Biol.* **2010,** *404* (4), 639-649.

4.   Singh, H.; Felts, R. L.; Ma, L.; Malinski, T. J.; Calcutt, M. J.; Reilly, T. J.; Tanner, J. J., Expression, purification and crystallization of class C acid phosphatases from Francisella tularensis and Pasteurella multocida. *Acta Crystallogr.* **2009,** *F65* (Pt 3), 226-31.

5.   Harper, M.; Boyce, J. D.; Adler, B., Pasteurella multocida pathogenesis: 125 years after Pasteur. *FEMS Microbiol. Lett.* **2006,** *265* (1), 1-10.

6.   Rice, J. A.; Carrasco-Medina, L.; Hodgins, D. C.; Shewen, P. E., Mannheimia haemolytica and bovine respiratory disease. *Anim. Health Res. Rev.* **2007,** *8* (2), 117-28.

7.   Vagin, A.; Teplyakov, A., MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **1997,** *30*, 1022-1025.

8.   (a) Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968,** *33*, 491-497; (b) Kantardjieff, K. A.; Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **2003,** *12* (9), 1865-71.

9.   Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D* **2010,** *66* (Pt 2), 213-21.

10.   Morris, R. J.; Perrakis, A.; Lamzin, V. S., ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **2003,** *374*, 229-44.

11.   Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004,** *D60* (Pt 12 Pt 1), 2126-32.

12.   Lanzetta, P. A.; Alvarez, L. J.; Reinach, P. S.; Candia, O. A., An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* **1979,** *100* (1), 95-97.

13.   Lebowitz, J.; Lewis, M. S.; Schuck, P., Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **2002,** *11* (9), 2067-79.

14.   Allen, K. N.; Dunaway-Mariano, D., Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **2004,** *29* (9), 495-503.

15.   Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-97.

16.   Golovin, A.; Henrick, K., MSDmotif: exploring protein sites and motifs. *BMC Bioinformatics* **2008,** *9*, 312.

17.   Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* **1990,** *215* (3), 403-10.

18.   Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **1991,** *A47* (4), 392-400.

19.   Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.

Chapter 6.

Structural Basis of the Inhibition of Class C Acid Phosphatases by Adenosine 5′-phosphorothioate

Harkewal Singh, Thomas J. Reilly, and John J. Tanner

Author contribution – H.S. cloned, expressed, purified, crystallized P4 and D64N. He also determined crystal structures described and performed enzyme kinetics experiments.

Abstract

The inhibition of phosphatases by adenosine 5′-phosphorothioate (AMPS) was first reported in the late 1960s, however, the structural basis for the inhibition has remained unknown. Here, it is shown that AMPS is a submicromolar inhibitor of class C acid phosphatases, a group of bacterial outer membrane enzymes belonging to the haloacid dehalogenase structural superfamily. The 1.35 Å resolution structure of the inhibited *Haemophilus influenzae* enzyme (aka rP4) is determined, which is the first structure of any enzyme complexed with AMPS. The conformation of the inhibitor is identical to that of the substrate 5′-AMP, except for the thiophosphoryl, which is rotated out of the phosphoryl binding site so that it binds 1.2 Å from the expected site. The rotation appears to be due to the larger size of S compared to O. This conformation is catalytically nonproductive, because the P atom is not positioned optimally for nucleophilic attack by Asp64, and the O atom of the scissile O-P bond is too far from the Asp residue (Asp66) that protonates the leaving group. The structure of 5′-AMP complexed with the Asp64Asn mutant enzyme was also determined at 1.35 Å resolution. The structure shows that mutation of this essential residue induces the substrate to adopt the same nonproductive mode of binding that is observed in the AMPS complex. The structures not only provide the basis for the inhibition of rP4 by AMPS, but also highlight the precise steric and electrostatic requirements of phosphoryl recognition.

6.1 Introduction

Class C acid phosphatases (CCAPs) constitute a large group of bacterial enzymes related by several shared characteristics. The conserved features that are evident from the amino acid sequence include a 25-residue bipartite motif featuring four essential Asp residues [1], subunit molecular weight of approximately 30 kDa, and subcellular localization to the bacterial outer membrane via a lipidated N-terminal Cys residue. Structural and biophysical studies have further shown that CCAPs exhibit the haloacid dehalogenase (HAD) fold and are dimeric in solution [2]. Kinetic studies of several CCAPs indicate a substrate preference for nucleoside monophosphates, which has led to the classification of CCAPs as broad specificity 5′-,3′-nucleotidases [2b, 3].

Although the defining characteristics of the family are established, the biological functions of CCAPs are less certain. The *Haemophilus influenzae* enzyme (aka P4), is the best studied member of the CCAP family. Kemmer *et al*. showed that one function of P4 is to dephosphorylate nicotinamide mononucleotide to nicotinamide riboside (NR) as part of a vestigial NAD$^+$ utilization pathway [4]. Whether other CCAPs also function in NAD$^+$ utilization is unknown.

High resolution crystal structures of recombinant P4 (rP4) complexed with substrates have provided insight into the nucleotidase activity of CCAPs [5]. Structures of a substrate-trapping mutant rP4 complexed with NMN, 5′-AMP, 3′-AMP, and 2′-AMP were determined. The structures revealed that the base binds

in an aromatic box such that the hydrogen bonding groups do not contact the protein. This result is consistent with the general lack of base specificity of rP4 and other CCAPs. The region around the ribose is relatively open, which allows recognition of both 5′- and 3-′ substrates. The phosphoryl is tightly surrounded by numerous charged and polar groups. This congested environment is necessary for positioning the phosphoryl group for nucleophilic attack by Asp64 and protonation of the leaving group by Asp66. Importantly, the span between the phosphoryl pocket and the aromatic box appears to be optimal for binding nucleoside monophosphates, which is consistent with the classification of CCAPs as 5′-,3′-nucleotidases. Given the high sequence identity of the active sites of CCAPs, these features of substrate recognition are expected to be valid for the entire family.

The observation that the rP4 active site is well suited for binding nucleoside 5′-monophosphate substrates motivated us to investigate the inhibition of CCAPs by the nucleotide analog adenosine 5′-phosphorothioate (AMPS). AMPS differs from 5′-AMP by just the substitution of an S atom for one of the O atoms of the phosphoryl and is thus isostructural with 5′-AMP (Scheme 6.1). AMPS has been shown to be a competitive inhibitor or very poor substrate of 5′-nucleotidases [6], but the structural basis of the inhibition is not known. We show here that AMPS is also an inhibitor of CCAPs and furthermore provide a structural explanation for the observed inhibition.

Scheme 1.
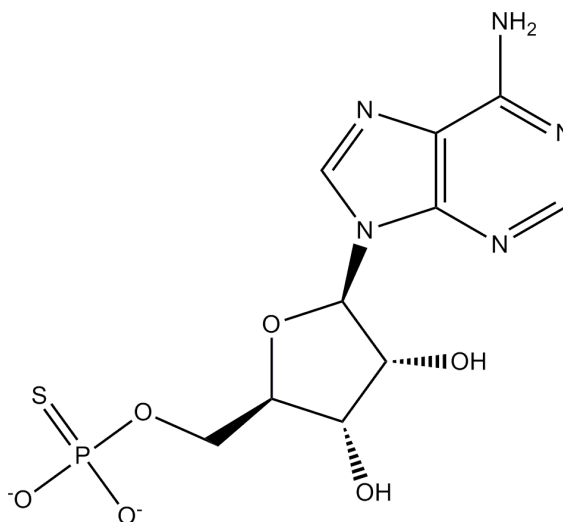
6.2 Materials and Methods

6.2.2 Kinetics of inhibition

His-tagged rP4 and *Francisella tularensis* CCAP (FtCCAP) were expressed and purified as described previously [5, 7]. Steady-state enzymatic activity was assessed at 25 °C with *p*-nitrophenyl phosphate (*p*NPP) as the substrate and using a discontinuous assay that measures the production of *p*-nitrophenolate [8]. The assay buffer consisted of 0.2 M sodium acetate, and 1 mM MgCl$_2$ at pH 5.5. The substrate concentration was varied in the range of 0.1 - 5 mM at fixed AMPS concentrations of 0.25 - 1.0 µM. The reaction was stopped using 1.0 M glycine at pH 10 after reaction times of 15 s, 75 s, 135 s and 195 s. The *p*-nitrophenolate concentration was then determined spectrophotometrically at 405 nm by reference to a standard curve constructed from solutions of known *p*-

nitrophenolate concentration. The initial rate was estimated by fitting data from the four time points to a line. The kinetic parameters for pNPP (V_{max}, K_m) and K_i for AMPS were estimated by the global fitting using Origin 8.5 software to the competitive inhibition model,

$$v = V_{max} \,[\text{S}] \,/\, \{K_m \,(1 + [\text{I}]/K_i) + [\text{S}] \}. \tag{1}$$

6.2.3 Preparation of the D64N mutant

A site-directed mutant of rP4 in which Asp64, the nucleophile that attacks the substrate phosphoryl group, is changed to Asn (D64N) was created using overlap extension PCR as follows. The pET20b plasmid containing the rP4 coding sequence was used to create the D64N mutation as follows. Oligonucleotides for D64N forward and D64N reverse primers were designed for initial PCR reactions. The PCR products were confirmed by 1% agarose gel electrophoresis and quantified using a Nanodrop spectrophotometer. Fifteen cycles of overlap PCR was performed as follows. The samples were heated to 94 °C to melt the DNA, and then primers were annealed at 25 °C followed by elongation at 72 °C for 1 minute per cycle. The T7 reverse and forward primers, and 0.5ul *Pyrococcus furiosus* DNA polymerase were added to the reaction mixture, and PCR was performed for 25 cycles by heating at 94 °C, annealing at 50 °C, and then holding at 72 °C for 1 minute per cycle. The presence of insert was confirmed with gel electrophoresis, and the specific mutation was verified by DNA sequencing. The PCR product was gel purified and then ligated into the

129

pZERO vector. *E. coli* TOP10 competent cells were transformed and plated on LB supplemented with kanamycin. Four colonies were picked and added to 5 ml LB (ampicillin) cultures for growth at 37 °C with shaking at 300 rpm. The plasmids were extracted, digested with *Nco1* and *Xho1,* and the resulting insert was ligated into pET20b at 16 °C overnight. The resulting pET20b plasmid containing the insert (pET20b16 D64N-rP4) was transformed into *E. coli*. DH5α. The transformants were plated onto LB (ampicillin) plates and incubated overnight at 37 °C. Four colonies were picked and grown in 5 ml LB (ampicillin) cultures. The plasmids were extracted and transformed into BL21(AI) for large scale protein expression using Studier's auto-induction method. D64N was purified as described previously for rP4 [5].

6.2.4 Preparation of rP4/AMPS and D64N/5′-AMP crystals

Hexagonal crystals of rP4 and D64N were grown as described previously for rP4. The crystals were grown in sitting drops at room temperature over reservoirs containing 0.05 - 0.2 M ammonium citrate, 0.05 - 0.15 mM $MgCl_2$, and 18 - 23 % (w/v) PEG 3350 in the pH range of 6.8 - 7.2.

Crystals of enzyme/ligand complexes were obtained by soaking. Crystals of rP4 were first cryoprotected at room temperature in 23 - 28 % (w/v) PEG 3350, 0.1 M ammonium citrate buffer pH 7.0, and 20% PEG 200. The cryoprotected crystals were transferred to a solution of the cryobuffer supplemented with 10 - 15 mM AMPS and 200 mM $MgCl_2$. After 15- 35 minutes the crystals were picked

up with Hampton loops and plunged into liquid nitrogen. Crystals of D64N complexed with 5′-AMP were prepared similarly, except that the substrate concentration was 10 - 23 mM, the $MgCl_2$ concentration was 100 mM, and the soaking times were 30 - 75 minutes.

6.2.5 Structure determination

X-ray diffraction data were collected at the Advanced Photon Source beamline 24-ID-C and processed with HKL2000 [9] (Table 6.1). The crystals have space group $P6_522$ with unit cell lengths of a = 98 Å and c = 107 Å, and there is one molecule in the asymmetrical unit with 54% solvent content and V_m of 2.6 $Å^3$/Da [10]. We note that this crystal form is the same one used to determine structures of the D66N mutant of rP4 complexed with substrates [5]. PHENIX [11] and COOT [12] were used for structure refinement and model building, respectively. Refinement was initiated from a model derived from the coordinates of a 1.35 Å resolution structure of rP4 (PDB code 3OCU). The test set of reflections (5 %) was based on the one used for previous refinements of rP4 structures [5]. The B-factor model used during the initial rounds of refinement consisted of an isotropic B-factor for each non-hydrogen atom and TLS refinement (one TLS group for the entire protein chain). During the final few rounds of refinement, anisotropic B-factors were used, which decreased R_{free} by 0.09 for the AMPS structure and 0.0? for the D64N structure.

131

Table 6.1. Data collection and refinement statistics.

Complex	rP4/AMPS	D64N/5′-AMP
Wavelength	0.9792	
Data collection resolution (Å)	50 - 1.35 (1.40 – 1.35)	50 - 1.35 (1.40 – 1.35)
No. of observations	547582	614437
No. of unique reflections	67160	66328
$R_{merge}(I)$	0.072 (0.607)	0.063 (0.537)
Average I/σ	31.8 (3.2)	39.2 (3.4)
Completeness (%)	100 (99.9)	99.8 (99.8)
Redundancy	8.2 (7.9)	9.3 (9.1)
Refinement resolution (Å)	50 - 1.35 (1.37 – 1.35)	50 - 1.35 (1.40 – 1.35)
R_{cryst}	0.138 (0.224)	0.142 (0.214)
R_{free}[a]	0.159 (0.252)	0.159 (0.257)
No. of protein residues	247	247
No. of protein atoms	1938	1928
No. of water molecules	305	251
Average B-factor ($Å^2$)		
Protein	13.2	15.1
Water	24.6	25.1
Ligand	13.3	13.9
Mg^{+2}	7.4	8.6
rmsd[b]		
Bonds (Å)	0.006	0.006
Angles (deg)	1.08	1.13
Ramachandran plot[c]		
Favored (%)	98.0	97.6
Allowed (%)	2.0	2.4
Outliers (%)	0.0	0.0
Coordinate error (Å)[d]	0.13	0.14
PDB code	3OCZ	

Values for the outer resolution shell of data are given in parenthesis.
  [a]A common set of test reflections (5 %) was used for refinement of all structures.
  [b]Compared to the parameters of Engh and Huber.[16]
  [c]The Ramachandran plot was generated with RAMPAGE.[17]
  [d]Maximum likelihood-based coordinate error estimate.

6.3 Results

6.3.1 Inhibition kinetics

AMPS was tested as an inhibitor of CCAPs using rP4 (UniProtKB C9MJZ6) and the CCAP from *Francisella tularensis* (FtCCAP, UniProtKB Q5NH51). The two enzymes share only 20 % global amino acid sequence identity and thus represent two distant branches of the CCAP family. Steady-state kinetic data could be fit satisfactorily to the competitive inhibition model (eq 1), which yielded and K_i values of 0.03 µM for rP4 and 0.3 µM for FtCCAP. Thus, AMPS is a submicromolar inhibitor of these enzymes.

6.3.2 Structure of rP4 inhibited by AMPS

The crystal structure of rP4 complexed with AMPS was determined at 1.35 Å resolution (Table 6.1). To our knowledge, this is the first structure of an enzyme complexed with this ligand.

Table 6.2. Inhibition of CCAPs by AMPS using *p*NPP as the substrate.

CCAP	K_m of (mM)	k_{cat} (s^{-1})	K_i (µM)
rP4	0.12 ± 0.02	3.3 ± 0.1	0.0346 ± 0.006
FtCCAP	0.56 ± 0.04	0.174 ± 0.004	0.30 ± 0.03

AMPS binds in the previously identified substrate binding site, which is located at the junction between the core and cap domains (Fig. 6.1). The adenine base interacts exclusively with the cap domain, while the thiophosphoryl interacts with the core domain.
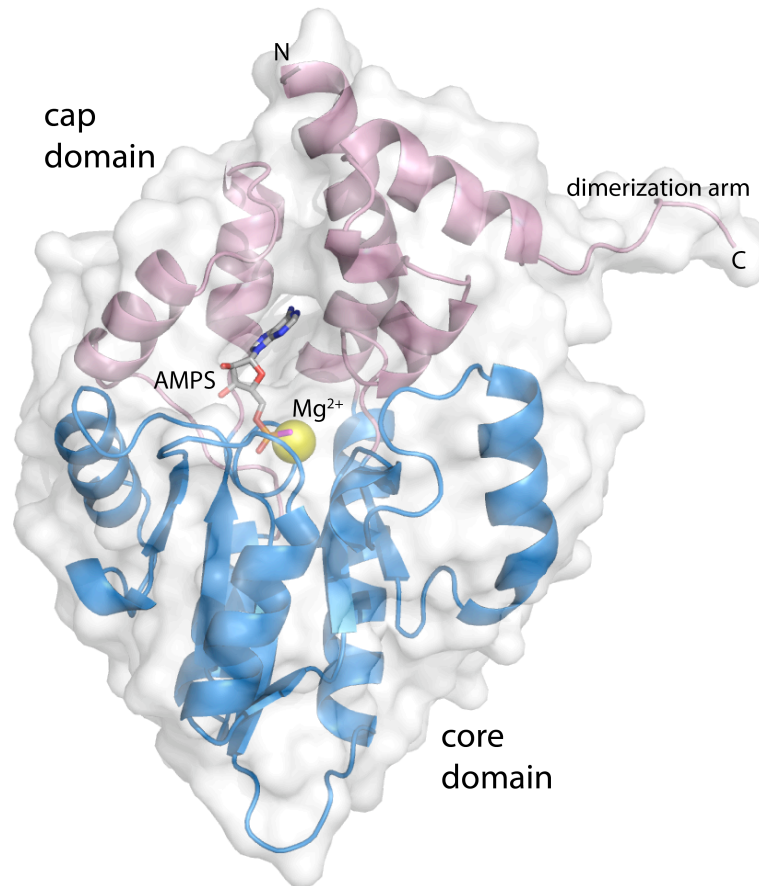
Fig. 6.1. Ribbon representation of rP4 complexed with AMPS. The core and cap domains are colored blue and pink, respectively. AMPS is colored gray; the S atom of AMPS is colored magenta. The yellow sphere represents Mg^{2+}. This figure and others were created with PyMOL.[18]

The electron density at 1.35 Å resolution allowed for unequivocal determination of the inhibitor conformation, as well as the conformations of surrounding side chains and positions of active site water molecules (Fig. 6.2). The conformation of AMPS is identical to that of 5′-AMP bound to the D66N substrate trapping mutant rP4 solved previously [5], except that the thiophosphoryl is shifted by 1.2 Å from the phosphoryl of 5′-AMP (Fig. 6.3). In both structures, the adenine base binds in an aromatic box formed by residues Phe86, Trp91 and Tyr221, and the ribose interacts with Glu131 via a water-mediated hydrogen bonds. The shift of the thiophosphoryl reflects an apparent rotation of about 20° around the C4′-C5′ bond.
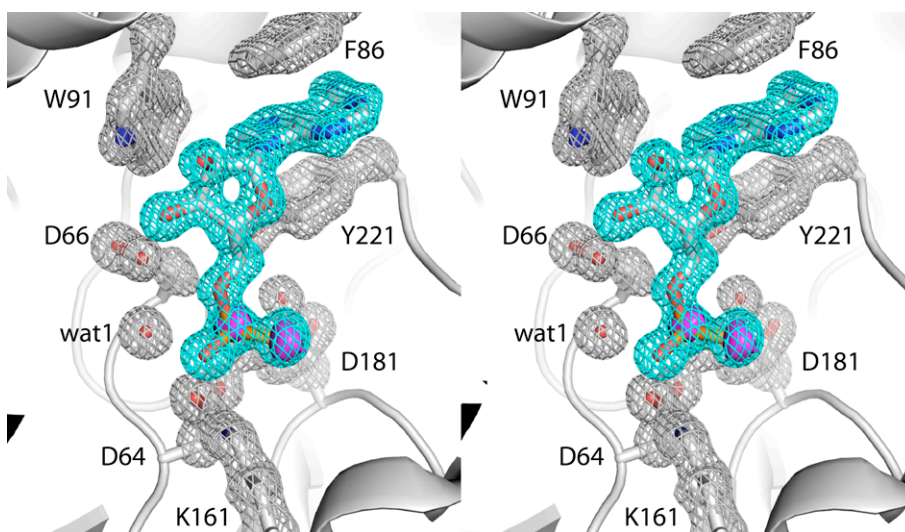
Fig. 6.2. Electron density maps showing the presence of AMPS in the active site (relaxed stereographic view). The cage (cyan for the ligand, silver for protein side chains) represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. Prior to map calculation, the ligand and surrounding residues and water molecules were removed, and simulated annealing refinement was performed using PHENIX. The magenta surface represents an

anomalous difference Fourier map contoured at 3.0 σ. Note that this map shows strong peaks corresponding to the P and S atoms of the thiophosphoryl.

The 20° rotation around the C4′-C5′ bond of AMPS is significant, because it moves the thiophosphoryl away from two essential catalytic residues, Asp64 and Asp66. The former residue functions as the nucleophile that attacks the substrate P atom. The latter residue, which is also referred to as the Asp+2 residue of HAD superfamily enzymes, functions as an acid that protonates the substrate leaving group. We note that this residue is mutated to Asn in the substrate-trapping enzyme D66N. The distance between the nucleophile and the P atom of AMPS is 3.3 Å, and the angle formed by the nucleophile, P, and O5′ is 137°. The corresponding values for 5′-AMP are 3.0 Å and 173°, which are more consistent with backside nucleophilic attack. The rotation also displaces the O atom of the scissile O-P bond out of reach from the Asp+2 residue. The O5′-Asp66 distances are 3.4 Å and 2.8 Å for AMPS and 5′-AMP, respectively. These subtle changes in the positioning of thiophosphoryl group lock the inhibitor in a nonproductive binding mode.

The identity of the S atom of AMPS was deduced from anomalous difference Fourier analysis in order to better understand the basis for the displaced thiophosphoryl group. Although the anomalous signal for sulfur is low at the wavelength of the data collection (f'' is 0.18 e$^-$ for P and 0.23 e$^-$ for S), the anomalous difference Fourier map was unambiguous. Peaks 1 and 5 of the map correspond to the P and S atom of AMPS, respectively (magenta surface in Fig. 6.2). We note that peaks 2 – 4 and 6 correspond to Met residues. The P-S bond points 180° away from the back wall of the active site (Fig. 6.4). Consequently,

the S atom sits in an open solvent-filled region of the active site and makes no direct contacts with the enzyme (Fig. 6.4).

With the S atom identified, the interactions involving the O atoms of the thiophosphoryl are evident (Fig. 6.3). One of the O atoms interacts with the Mg^{2+} ion, while the other one interacts with Lys161. Unlike the phosphoryl of 5′-AMP, the thiophosphoryl does not interact directly with conserved Thr124 or the backbone of Asp66. Instead, a water molecule mediates interaction these groups (denoted wat1 in Fig. 6.3). Binding of AMPS does not disrupt the octahedral coordination geometry of the Mg^{2+} ion. In ligand-free rP4 (PDB code 3ET4), the metal ion binds to Asp64, Asp81, the backbone carbonyl of Asp66, and three water molecules. One of the coordinating water molecules dissociates upon substrate binding, and an O atom of the phosphoryl takes its place. An analogous exchange occurs during the binding of AMPS.
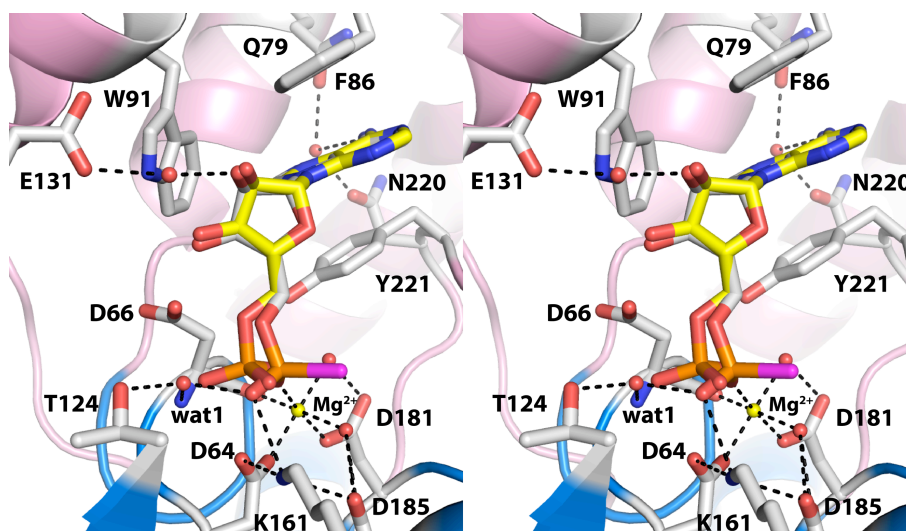
Fig. 6.3. Active site of rP4 inhibited by AMPS (relaxed stereographic view). AMPS is colored white; the S atom of AMPS is colored magenta. For reference, 5′-AMP from the D66N/5′-AMP structure (PDB code 3OCV) is also shown (yellow). Secondary structural elements of the core and cap domains are colored blue and pink, respectively, and residues of the aromatic box are colored green. The yellow sphere represents Mg^{2+}. The dashed lines denote electrostatic interactions involving AMPS.
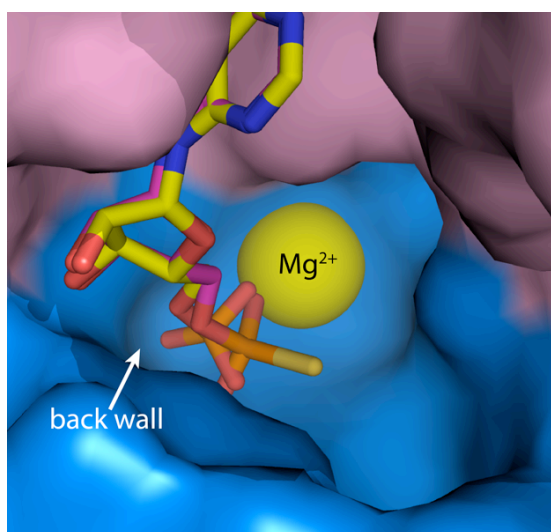
Fig. 6.4. Surface representation of the active site highlighting the contrasting conformations of 5′-AMP (yellow) and AMPS (magenta). Note that the S atom of AMPS point away from the back wall of the active site and toward an open cavity that leads to the bulk medium.

6.3.3 Mutation of the Essential Asp Nucleophile Induces the Nonproductive Conformation in the Substrate

The conformation of AMPS is reminiscent of substrates bound to human mitochondrial deoxyribonucleotidase (mdN) [13]. This enzyme also belongs to the

HAD structural superfamily, but the cap domain fold and quaternary structure are substantially different from those of rP4. In order to trap nucleoside monophosphate substrates bound to mdN, the essential nucleophile (Asp41) was mutated to Asn. The substrates bound with the phosphoryl in one or both of two alternative conformations, which were dubbed Mode A and Mode B. Mode A is the active conformation and is identical to the phosphoryl position of 5′-AMP bound to D66N. Mode B is a nonproductive conformation and resembles AMPS. Presumably, the adoption of the Mode B conformation results from the alteration of the electrostatic environment caused by the mutation of the negatively charged Asp41 to Asn.

Mutation of the Asp64 nucleophile in rP4 also induces the substrate to adopt the nonproductive B mode conformation. The structure of D64N complexed with 5′-AMP was determined at 1.35 Å resolution (Table 6.1). The electron density map definitively indicates just a single conformation, which is very similar to that of AMPS (Fig. 6.5). The only substantial difference between the two structures involves Lys161. In the AMPS complex, Lys161 interacts with Asp64, Asp185, and a thiophosphoryl O atom (Fig. 6.6, black dashes). In the D64N/5′-AMP complex, however, Lys161 has rotated 90° around the $C\delta$-$C\epsilon$ bond to avoid an electrostatic clash with the side chain amine group of Asn64. This new conformation of Lys161 maintains the ion pair with Asp185 and also allows a new interaction with a different phosphoryl oxygen atom (the O atom corresponding to the S of the thiophosphoryl, see Fig. 6.6, green dashes).
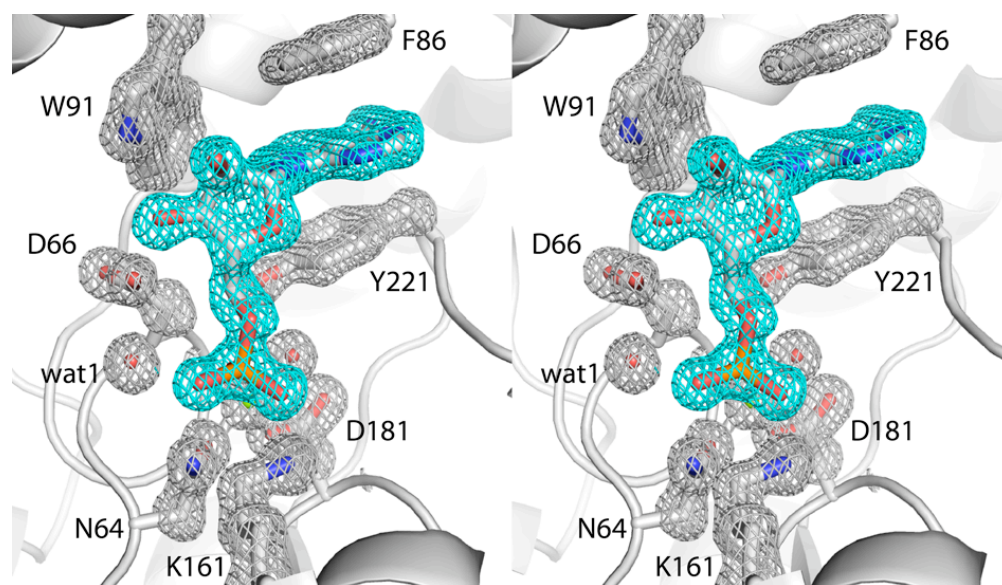
Fig. 6.5. Electron density map showing the presence of 5′-AMP in the active site of D64N (relaxed stereographic view). The cage (cyan for the ligand, silver for protein side chains) represents a simulated annealing σ_A-weighted F_o - F_c omit map contoured at 3.0 σ. Prior to map calculation, the ligand and surrounding residues and water molecules were removed, and simulated annealing refinement was performed using PHENIX.
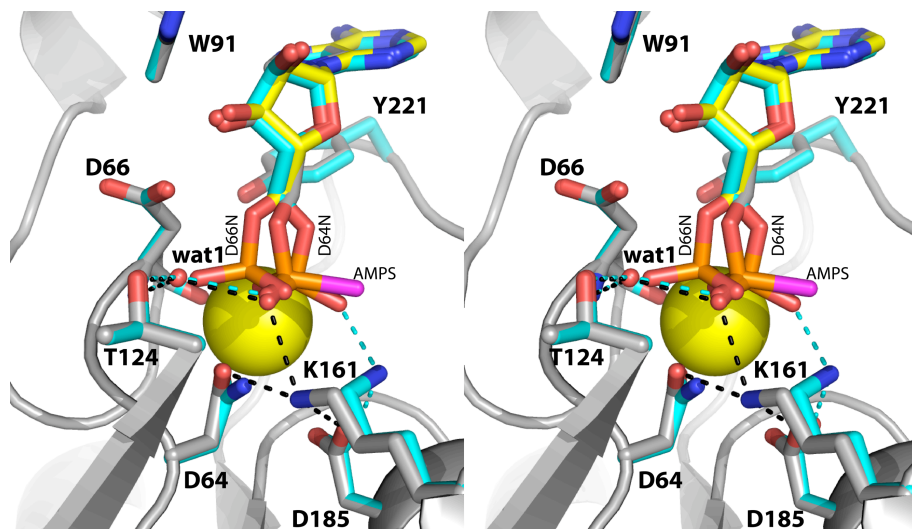
Fig. 6.6. Comparison of rP4/AMPS (gray with magenta S atom) and D64N/5′-AMP (cyan) (relaxed stereographic view). The black and cyan dashes denote electrostatic interactions in rP4/AMPS and D64N/5′-AMP, respectively. For reference, the 5′-AMP bound to D66N is shown in yellow. The yellow sphere represents Mg^{2+}.

6.4 Discussion

Murray and Atkinson reported in 1968 that AMPS is a competitive inhibitor (K_i value of 20 μM) of the type II 5′-nucleotidase from *Crotalus adamanteus* venom [6a]. One plausible explanation for the inhibition is that the lower electronegativity of S compared to O causes the P atom of AMPS to be less electropositive than that of 5′-AMP and therefore less susceptible to nucleophilic attack by the enzyme. In this scenario, one might expect the inhibitor to adopt the same conformation as the substrate.

The rP4/AMPS structure suggests a different mechanism of inhibition, which

is based on steric considerations. The larger size of S compared to O prevents the thiophosphoryl from occupying the catalytically active conformation. The phosphoryl binding site appears to be too small, and apparently not flexible enough, to accommodate the larger S atom. This steric conundrum is solved by rotating the thiophosphoryl away from the back wall of the active site, which allows the S atom to protrude into an open solvent filled cavity. As a consequence of the rotation, the P atom is not positioned optimally for nucleophilic attack by Asp64, and the O atom of the scissile O-P bond is too far from the Asp residue (Asp66) that protonates the leaving group. Thus, the AMPS is an inhibitor of rP4.

This nonproductive mode of binding is also adopted by substrates bound the mutant enzyme D64N. In this case, electrostatic considerations underlie the rotation of the phosphoryl. Mutation of Asp to Asn decreases the negative charge of the active site and places an obligate hydrogen bond donor group next to Lys161. It appears that Lys161 rotates to avoid the electrostatic repulsion, which induces rotation of the phosphoryl into the Mode B position. We note that a similar rotation was not observed for the mutant D66N. Asp66 functions as the acid that protonates the leaving group, and is therefore protonated at the beginning of the catalytic cycle. Thus, mutation of Asp66 to Asn does not change the overall charge of the active site. For this reason, it is a superior to D64N as a substrate-trapping reagent. Taken together, the two rP4 structures reported here highlight the precise steric and electrostatic requirements of phosphoryl recognition by rP4.

We note that a similar rotation of the phosphoryl was not observed for D66N. Asp66 functions as the acid that protonates the leaving group, and is therefore protonated at the beginning of the catalytic cycle. Thus, mutation of Asp66 to Asn does not change the overall charge of the active site. For this reason, D66N is superior to D64N as a substrate-trapping enzyme.

The D64N structure illustrates the potential pitfalls of the common practice of using active site mutants to understand enzyme structure and mechanism. Most biochemists consider the mutation of Asp to Asn to be a conservative change. Nevertheless, this mutation at residue 64 of rP4 induced the substrate to adopt a conformation that does not reflect the conformation of the substrate bound to the native enzyme. Interestingly, this artifact was avoided in rP4 by mutating Asp66 rather than Asp64. These results remind us that seemingly innocuous mutations can induce unexpected structural artifacts depending on the three-dimensional context of the mutated residue.

The data reported here could potentially aid inhibitor design efforts. For example, it has been suggested that targeting the NAD^+ utilization pathway may be a possible route to the development of narrow spectrum antimicrobial agents against *H. influenzae* [14]. The NAD^+ utilization pathway of *H. influenzae* includes an uptake system that imports NAD^+, NMN, and nicotinamide riboside (NR) into the periplasm. The periplasmic NAD^+ nucleotidase NadN catalyzes the hydrolysis of NAD^+ to generate NMN and AMP. NMN is dephosphorylated in the periplasm by P4 to NR, which is transported across the inner membrane into the cytosol by the NR-specific permease PnuC. Finally, the bifunctional NR kinase/NMN

adenylyltransferase NadR converts NR into NAD$^+$. All of these enzymes and transporters are potential targets for inhibitor design. The structural and kinetic data reported here suggest that nucleotide analogs have the potential to be potent inhibitors of P4.

The similarities between the nonproductive nucleotide conformations reported here for rP4 and observed previously for mdN may provide new insight into designing inhibitors of human 5′-nucleotidases. 5-nucleotidases regulate the nucleotide pool by reversing the phosphorylation process of nucleosides by nucleoside kinases [15]. They have been implicated in decreasing the efficacy of antiviral and antiblastic nucleoside analogs. These drugs must be activated by phosphorylation to exert their therapeutic effect, and intracellular 5-nucleotidases can reverse the activation step and thus decrease efficacy. For this reason, designing inhibitors of 5′-nucleotidases for co-therapy is of interest. P4 and mdN share a similar core domain that includes the essential DDDD motif and a hydrophobic region for binding the nucleotide base. These similarities raise the possibility that AMPS could be an inhibitor of mdN.

6.5 References

1.   Thaller, M. C.; Schippa, S.; Rossolini, G. M., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci.* **1998,** *7* (7), 1647-52.

2.   (a) Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of Recombinant Haemophilus Influenzae e (P4) Acid Phosphatase Reveals a New Member of the Haloacid Dehalogenase Superfamily. *Biochemistry* **2007,** *46* (39), 11110-9; (b) Reilly, T. J.; Chance, D. L.; Calcutt, M. J.; Tanner, J. J.; Felts, R. L.; Waller, S. C.; Henzl, M. T.; Mawhinney, T. P.; Ganjam, I. K.; Fales, W. H., Characterization of a unique class C acid phosphatase from Clostridium perfringens. *Appl. Environ. Microbiol.* **2009,** *75* (11), 3745-54; (c) Singh, H.; Malinski, T. J.; Reilly, T. J.; Henzl, M. T.; Tanner, J. J., Crystal structure and immunogenicity of the class C acid phosphatase from Pasteurella multocida. *Arch. Biochem. Biophys.* **2011,** *509* (1), 76-81.

3.   (a) Reilly, T. J.; Chance, D. L.; Smith, A. L., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J. Bacteriol.* **1999,** *181* (21), 6797-6805; (b) Reilly, T. J.; Smith, A. L., Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr. Purif.* **1999,** *17* (3), 401-409; (c) Malke, H., Cytoplasmic membrane lipoprotein LppC of Streptococcus equisimilis functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998,** *64* (7), 2439-42; (d) Passariello, C.; Schippa, S.; Iori, P.; Berlutti, F.; Thaller, M. C.; Rossolini, G. M., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5'-nucleotides. *Biochim. Biophys. Acta* **2003,** *1648* (1-2), 203-9; (e) Reilly, T. J.; Calcutt, M. J., The class C acid phosphatase of Helicobacter pylori is a 5' nucleotidase. *Protein Expr. Purif.* **2004,** *33* (1), 48-56; (f) Wang, R.; Ohtani, K.; Wang, Y.; Yuan, Y.; Hassan, S.; Shimizu, T., Genetic and biochemical analysis of a class C non-specific acid phosphatase (NSAP) of Clostridium perfringens. *Microbiology* **2010,** *156* (Pt 1), 167-73.

4.   Kemmer, G.; Reilly, T. J.; Schmidt-Brauns, J.; Zlotnik, G. W.; Green, B. A.; Fiske, M. J.; Herbert, M.; Kraiss, A.; Schlor, S.; Smith, A.; Reidl, J., NadN and e (P4) are essential for utilization of NAD and nicotinamide mononucleotide but not nicotinamide riboside in Haemophilus influenzae. *J. Bacteriol.* **2001,** *183* (13), 3974-3981.

5.   Singh, H.; Schuermann, J. P.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Recognition of Nucleoside Monophosphate Substrates by Haemophilus influenzae Class C Acid Phosphatase. *J. Mol. Biol.* **2010,** *404* (4), 639-649.

6.　　(a) Murray, A. W.; Atkinson, M. R., Adenosine 5'-phosphorothioate. A nucleotide analog that is a substrate, competitive inhibitor, or regulator of some enzymes that interact with adenosine 5'-phosphate. *Biochemistry* **1968,** *7* (11), 4023-9; (b) Skladanowski, A. C.; Hoffmann, C.; Krass, J.; Jastorff, B.; Makarewicz, W., Structure-activity relationship of cytoplasmic 5'-nucleotidase substrate sites. *Biochem. J.* **1996,** *314 (Pt 3),* 1001-7.

7.　　Singh, H.; Felts, R. L.; Ma, L.; Malinski, T. J.; Calcutt, M. J.; Reilly, T. J.; Tanner, J. J., Expression, purification and crystallization of class C acid phosphatases from Francisella tularensis and Pasteurella multocida. *Acta Crystallogr.* **2009,** *F65* (Pt 3), 226-31.

8.　　(a) Lanzetta, P. A.; Alvarez, L. J.; Reinach, P. S.; Candia, O. A., An improved assay for nanomole amounts of inorganic phosphate. *Anal. Biochem.* **1979,** *100* (1), 95-97; (b) Carter, S. G.; Karl, D. W., Inorganic phosphate assay with malachite green: an improvement and evaluation. *J. Biochem. Biophys. Methods* **1982,** *7* (1), 7-13.

9.　　Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997,** *276,* 307-326.

10.　　(a) Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968,** *33,* 491-497; (b) Kantardjieff, K. A.; Rupp, B., Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **2003,** *12* (9), 1865-71.

11.　　Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D* **2010,** *66* (Pt 2), 213-21.

12.　　Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Cryst.* **2004,** *D60* (Pt 12 Pt 1), 2126-32.

13.　　Wallden, K.; Ruzzenente, B.; Rinaldo-Matthis, A.; Bianchi, V.; Nordlund, P., Structural basis for substrate specificity of the human mitochondrial deoxyribonucleotidase. *Structure* **2005,** *13* (7), 1081-8.

14.　　Gerlach, G.; Reidl, J., NAD+ utilization in Pasteurellaceae: simplification of a complex pathway. *J. Bacteriol.* **2006,** *188* (19), 6719-27.

15.　　Bianchi, V.; Spychala, J., Mammalian 5'-nucleotidases. *J. Biol. Chem.* **2003,** *278* (47), 46195-8.

16.   Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst.* **1991,** *A47* (4), 392-400.

17.   Lovell, S. C.; Davis, I. W.; Arendall, W. B., 3rd; de Bakker, P. I.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C., Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **2003,** *50* (3), 437-50.
18.   DeLano, W. L., *The PyMOL User's Manual*. DeLano Scientific: Palo Alto, CA, USA, 2002.

Chapter 7.

Cloning, expression, purification and structure determination of Protein tyrosine phopshatases 1B (PTP1B)

Author contribution – H.S. cloned, expressed, purified and crystallized PTP1B. He also collected the data and solved crystal structures. All the enzyme kinetics experiments were performed by Prof. Kent S. Gates's lab.

Abstract

Protein tyrosine phosphatases (PTPs) are important drug target against diabetes mellitus and are also inactivated by intracellular hydrogen peroxide (H_2O_2). Although, H_2O_2 mediated inactivation in PTP1B, a prototypical member of the PTP superfamily member, is well known but the process is rather slow. The research described in this chapter addresses the novel use of bicarbonate-based buffer for H_2O_2 mediated PTP1B inactivation rate enhancement. The methods used for PTP1B cloning, expression, purification, crystallization and structure determination are described here. The results of biochemical experiments conducted by Prof. Gates's laboratory and 1.7 Å resolution $KHCO_3$-H_2O_2 inactivated PTP1B crystal structure suggest that $KHCO_3$ assisted and H_2O_2 mediated PTP1B inactivation process involves oxidation of the catalytic cysteine residue in PTP1B.

7.1 Introduction

Cells utilize phoshorylation of specific protein residues such as tyrosine to regulate a variety of intracellular events. Protein kinases are those enzymes that serve the purpose of phosphorylating tyrosine residues while protein phosphatases can revert this process by dephosphorylating phospho-tyrosine residues[1].

Phosphotyrosine based signaling can be imagined as a molecular switch, implicated in several biological events such as metabolism, proliferation, differentiation, and cell cycle regulation. For instance, the insulin binding to the receptors triggers a cascade of intracellular signal transduction events that can influence the glucose uptake from the blood. These signal transduction events are often accompanied by the reversible phohsphorylation of the protein tyrosine residues [2-4].

The mechanism central to this molecular switch involves the phosphorylation of tyrosine residues by protein kinases (PTKs) and then de-phosphorylation of the phosphorylated tyrosine residue by protein tyrosine phosphatases (PTPs)[1]. The phosphorylation status of these tyrosine residues is well regulated by the opposing reactions carried out by PTKs and PTPs respectively. In some cases, the dephosphorylation of these tyrosine residues by PTPs acts as the turn "off" switch for the cellular response [5].

PTPs also play important role in the insulin cascade system and glucose homeostasis. Insulin receptor (IR) belongs to receptor tyrosine kinases family

that are tetrameric allosteric enzymes and are composed of two α and two β subunits[6]. In these enzymes the α subunits inhibit the kinase activity of the β subunit. Due to the binding of insulin to the α subunit, the kinase activity of the β subunit is suppressed that results in the transphosphorylation of the β subunit leading to conformational change and further augmentation of kinase activity [7]. Specifically, the allosteric induced change in the structure of IR leads to the autophosphorylation of β subunit Tyr1146, Tyr1162 and Tyr1163 residues[7]. This turns the IR into kinase that can subsequently phosphorylate various insulin receptor substrates. These phosphorylated IR substrates are the "hotspots" for the dephosphorylation by phosphatases. In particular, protein tyrosine phosphatase 1B (PTP1B) dephosphorylates the phosphorylated β subunit of IR, therefore, potentially acting as negative regulator for the insulin signal transduction, hence a drug target against diabetes mellitus.

Initially thought as the "house keeping" genes, PTPs have now been appreciated as a diverse family of enzymes that is represented by about 100 genes in humans[8]. The combined actions of PTPs and PTKs regulate a multitude of cellular and biochemical functions in the cell[9]. The prototypic member of PTP family, PTP1B was originally isolated from human placenta as a 37 kDa soluble protein [10, 11] but further analysis revealed that it is indeed 50 KDa protein encoded by 435 amino acids and that the 37 KDa domain corresponds to N terminal 1-321 amino acid residues[12]. In fact work from other groups has suggested the use of a smaller version of PTP1B catalytic domain consisting 1-298 amino acid residues [13, 14]. Briefly PTP1B is a cysteine dependent enzyme

exhibiting the characteristic PTP signature motif **[I/V] HCXXGXXR [S/T]** carrying an invariant Cys residue that functions as nucleophile during the catalysis[15]. PTP1B active site provides a unique environment that lowers the pKa of the active site Cys that enhances the nucleophilicity of Cys and renders it susceptible for oxidation by peroxides such as H_2O_2 [16]. The oxidation of the active site Cys in PTP1B abolishes its activity and therefore inhibition of PTP activity. This represents as a control mechanism of tyrosine –phosphorylation based signaling using the inhibition of PTPs.

H_2O_2 is a signaling agent and is known to mediate several physiological processes such as cell proliferation, differentiation and migration[17]. It also affects several signaling pathways and modifies the activity of key signaling proteins by selectively oxidizing the cysteine residues in the proteins[18]. In that context, PTPs are major targets for H_2O_2 mediated reversible oxidation. The oxidation of the active site Cys can yield either of sulphenic acid (SOH), sulphinic acid (SO_2H) or sulphonic acid (SO_3H), depending upon the extend of oxidation[19]. For this oxidation to be a control mechanism it is absolutely necessary that the active site Cys be not oxidized beyond the sulphenic acid form since the other two forms of oxidation might render the enzyme irreversibly modified[19]. Earlier work [15, 16] has shown that H_2O_2 mediated inactivation of PTP1B proceeds via five membered cyclic sulfenyl amide form of Cys215. This oxidation form of Cys215 is known to induce large conformational changes in the tertiary structure of PTP1B and is also mechanistically reversible upon thiol addition.

During signaling events peroxide mediated inactivation of PTPs is known to occur in the range of 5-15 min [20-22]. In vitro experimental evidences indicate that the loss of recombinant PTPs activity by H_2O_2 is rather poor i.e. the rate constants 10-40 M^{-1} s^{-1} [23]. This observation suggest that the loss of enzymatic activity of PTPs during signaling processes should be slower owing to the low amounts of H_2O_2 present during signaling events (0.1-1 μM) [24, 25]. This raises a possibility that H_2O_2 might be converting to more reactive oxidizing species that could be responsible for fast inactivation of PTPs intracellularly [26, 27].

In this context of the above kinetic discrepancy, Prof. Gates group set out to explore the use of bicarbonate/CO_2 buffer system in H_2O_2-mediated signal transduction. The larger context of the work described here is to elucidate the effect of bicarbonate/CO_2 buffer and H_2O_2 in PTP1B inactivation. Due to the collaborative nature of this work between Prof. Gates and Prof. Tanner's laboratory, the major focus of this chapter will be to describe the work done by me in Prof. Tanner's laboratory. Therefore, this chapter will focus on cloning, expression, purification, crystallization and structure determination of various PTP1B structures. The main thrust of the work described here will provide details about the methods used to sub-clone and express PTP1B for efficient and reproducible purification, crystallization and finally crystal structure of PTP1B inactivated by $KHCO_3$ and H_2O_2. Additionally, I have also determined crystal structures of PTP1B inactivated by dietary supplement allyl isothiocyanate (AITC) and cyclic TPZ inactivator, which will be described in the appendix of this thesis.

7.2 Materials and Methods

7.2.1 Cloning of PTP1B 1-321 and 1-298 constructs

For the crystallographic studies, PTP1B catalytic domain constructs having cleavable polyhistidine tags were engineered. The coding sequence for residues 1-321 was subcloned into pKA8H vector using *Nde*I and *Bam*HI sites such that the *N*-terminal polyhistidine tag was cleavable using tobacco etch virus protease (TEVP). Briefly, the plasmid containing PTP1B coding sequence was amplified using PCR. The PCR product was gel purified using a 1% DNA agarose gel. The purified PCR product was ligated into pZERO vector at 16 °C. The resulting ligation product was transformed into DH5α and plated onto LB agar plates supplemented with 40 µg/mL kanamycin and the plate was incubated at 37 °C overnight. Following incubation, four single colonies were picked and grown in LB media supplemented with 40 µg/mL kanamycin. These cultures were incubated at 37 °C and 250 rpm overnight in an incubator-shaker and further used for plasmid preparation. The isolated plasmids were excised with *Nde*I and *Bam*HI and gel purified as described above. These inserts were ligated into pKA8H vector (cut with *Nde*I/*Bam*HI and purified) followed by transformation into DH5α. The transformants were plated onto LB agar plates supplemented with 50µg/mL ampicillin. Four single colonies were picked again for plasmid preparation and the clone was verified by DNA sequencing.

A shorter version of PTP1B including residues 1 - 298 was created by inserting a stop codon into the aforementioned plasmid. The Quick Change kit

(Stratagene) was used for this purpose, and the clone was confirmed by DNA sequencing.

7.2.2 Expression and Purification of PTP1B (1-298 domain)

The PTP1B (1-298) plasmid was transformed into *E. coli* BL21AI cells and plated on LB Agar containing ampicillin (50 µg/mL). The plate was incubated at 37 ºC overnight and a single colony of the transformants was picked to inoculate 10 mL starter culture made of 1% tryptone and 0.5% yeast extract. This was incubated at 37 ºC with constant shaking at 250 rpm, overnight. The starter culture was used to inoculate 1 L of auto-induction media [28] and the cells were allowed to shake constantly for two hours at 37 ºC and 250 rpm. After two hours of cell growth, the temperature was reduced to 25 ºC, and 0.2 % arabinose was added to the media. Cells were harvested after 20-21 hours by centrifugation at 4 ºC and 3500 rpm and resuspended in Buffer A (20 mM Tris, 150 mM NaCl, 10% glycerol pH 7.5). The cell pellet was quick-frozen into liquid nitrogen for later use.

Frozen cells were thawed at 4 ºC in the presence of the following protease inhibitors: 10 µM leupeptin, 1 µM pepstatin A, 1 mM PMSF. Cells were stirred for 15-20 minutes at 4 ºC followed by disruption using sonication. Unbroken cells and debris were removed by centrifugation for 60 min at 17,000 rpm. The supernatant was collected and subjected to a second centrifugation step for 30 min at 17,000 rpm. The resulting supernatant was used for further purification by immobilized metal-ion affinity chromatography (Ni^{2+}-charged HiTRAP; GE

Healthcare). The fractions were eluted using buffer B (Buffer A supplemented with 1 M imidazole). Fractions containing PTP1B were pooled and mixed with TEVP (1 mg of TEVP per 40 mg of PTP1B) and 1 mM THP. The sample was incubated for 8 h at 20 ºC and then dialyzed against buffer A. The dialyzed protein was again loaded onto the Ni^{2+} charged column using buffer A. Tag-free PTP1B was collected in both the flow-through and by elution in 3% buffer B. The purified protein was dialyzed into 10 mM Tris, 25 mM NaCl, 1 mM EDTA, 1 mM THP pH 7.5. Finally, the protein was distributed into thin-walled PCR tubes, quick-frozen in liquid nitrogen, and stored at -80 ºC.

7.2.3 Crystallization of PTP1B (1-298)

All crystallization trials were performed at 4ºC using sitting drop vapor diffusion method. Commercially available screens were used for initial crystallization condition. Crystals appeared within a week into several conditions mostly having MPD and PEG 3000 as the precipitant. Briefly, for this work, the diffraction quality crystals were grown using 11-18 % PEG3000, 0.1M HEPES pH 7.0-8.0, 0.2 M magnesium acetate and 2mM TCEP. Note that this condition has been reported earlier [29]. Typical protein concentration used for optimal crystal growth was 10 mg/mL.

7.2.4 Soaking with H_2O_2 and $KHCO_3$ and cryoprotection

Soaking of the native PTP1B crystals was done at room temperature. Stock solution for $KHCO_3$ was prepared in water and the crystals were cryoprotected using 20% PEG3000, 0.1M HEPES pH 7.0, 0.2 M magnesium acetate and 20% PEG 200. The final cryoprotected crystals were transferred to a cryobuffer supplemented with 25mM $KHCO_3$ and 50µM H_2O_2. The soak time was varied from 20-45 minutes. Crystals were picked up Hampton loop and plunged into liquid N_2 for low temperature data collection.

7.2.5 Data collection, structure determination and refinement

Low temperature data was collected at ALS 4.2.2. Data was processed using d*trek and structure was determined using molecular replacement. Briefly the crystals diffracted to 1.70 Å and belonged to $P3_12_1$ space group with unit cell dimensions of a=b= 88.41 Å and c=104.28 Å. (Table 1). Search model was derived from a native PTP1B structure (PDB code 2f71 [30]) with following residues omitted: 46-49, 180-188, and 215-221. Note that these residues are known to be mobile depending upon the oxidation state of the active site [15, 16]. Therefore, to ascertain the oxidation status of our structure and pinpoint the conformation of these residues, we omitted these residues in the starting model and determined the structure using molecular replacement. Briefly, Auto MR was used from PHENIX[31] for molecular replacement. A clear solution, having one molecule in

the asymmetric unit was obtained. A common set of test reflections (5%) was used for refinement calculations. For this structure, the B-factor model used during the initial rounds of refinement consisted of an isotropic B-factor for each nonhydrogen atom and TLS refinement, with one TLS group corresponding to the protein chain. The model was improved with iterative rounds of model building in Coot [32] and refinement in PHENIX [33].

7.3 Results and Discussion

7.3.1 Effect of Bicarbonate on PTP1B inactivation by H_2O_2

To explore the effect of bicarbonate on H_2O_2 mediated signaling, Dr. Gates's group measured the inactivation of PTP1B catalytic domain (1-322) in the presence of bicarbonate ($KHCO_3$). The extracellular and intracellular concentration of bicarbonate is 14.4 mM and 25 mM respectively [34]. Therefore, to examine the effects of $KHCO_3$ in context of biological system, they employed the concentration of $KHCO_3$ in the range of 14.4 – 25 mM in their experiments.

Their experiments suggest that the presence of potassium bicarbonate markedly increased the rate of time-dependent enzyme inactivation by H_2O_2. For example, Dr. Gates group showed that the potassium bicarbonate increased the apparent second-order rate constants for inactivation of PTP1B by H_2O_2 to 202 ± 4 M^{-1} s^{-1} and 330 ± 11 M^{-1} s^{-1} at concentrations of 25 and 50 mM, respectively. At physiological temperature (37 °C), the rate of inactivation by the H_2O_2-$KHCO_3$ system increases further to 396 ± 10 M^{-1} s^{-1}, when used in the presence of 25mM

KHCO$_3$. It is worthwhile to note that potassium bicarbonate alone did not cause time-dependent inactivation of PTP1B at 24 °C. The time-dependent nature of the inactivation observed by them is consistent with a process involving covalent chemical modification of the enzyme.

Time-dependent inactivation of PTP1B by the H$_2$O$_2$-KHCO$_3$ system was slowed by competitive inhibitors. For example, the competitive inhibitor phosphate (50 mM, K$_i$ = 20 mM) slowed inactivation by a factor of 1.7 ± 0.1. Activity did not return to the inactivated enzyme following gel filtration or dialysis to remove H$_2$O$_2$ and potassium bicarbonate. Together the results suggest that inactivation of PTP1B by H$_2$O$_2$-KHCO$_3$ involves covalent modification of an active site residue. Catalytic activity was recovered upon treatment of the inactivated enzyme with thiols such as dithiothreitol (DTT, 5-50 mM). Note that similar observation has been made by others upon PTP1B treatment with H$_2$O$_2$ alone [15, 16].

The exact mechanism by which bicarbonate accelerates H$_2$O$_2$-mediated inactivation of PTP1B remains uncertain. However, to further understand the basis of this reversible inactivation, we determined the crystal structure of PTP1B inactivated by KHCO$_3$ and H$_2$O$_2$ at 1.7 Å resolution as described below.

7.3.2 Structural analysis of PTP1B inactivated by KHCO$_3$ and H$_2$O$_2$

Several data sets were collected and the electron density maps were inspected. Overall structure of the oxidized PTP1B is exactly similar to the one

previous reported by Salmeen et al. [15]. The two structures superpose exactly on each other (Fig. 7.1). The maps clearly indicated the presence of both the native Cys215 conformation and the cyclic sulfenyl amide conformation, with the occupancies of the two forms dependent on the concentrations of $KHCO_3$ and H_2O_2 used for soaking. The 1.7 Å resolution structure reported here (Table 7.1) was obtained from a crystal soaked in 25 mM $KHCO_3$ and 50 µM H_2O_2, which resulted in nearly 100% occupancy of the sulfenyl amide (Fig. 7.2). The conformations of the P-loop and other flexible active site loops are identical to those described previously for crystals soaked in H_2O_2 [15, 16]. For example, electron density for the P-loop is shown in Fig. 7.2. The crystal structure presented here is consisted with the biochemical data pointing towards the covalent modification of Cys215 in the active site.

The kinetics data provided here along with structural information, shows that the presence of biological buffer system, bicarbonate/CO_2, enhances the H_2O_2 mediated PTP1B inactivation. Although, sufficient experiments are warranted to elucidate the exact mechanism employed by PTP1B for the higher inactivation rates observed nonetheless our structural data convincingly points towards the presence of five-membered sulfenyl amide ring implicated in the process.
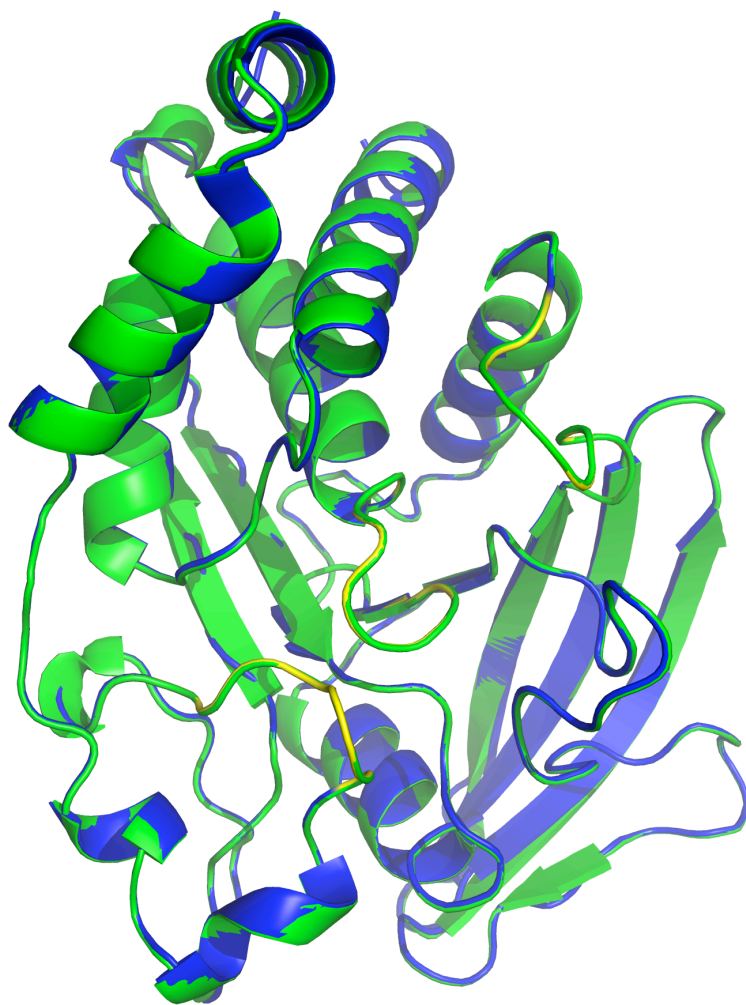
Fig. 7.1 Superposition of 1.7 Å PTP1B structure (blue) oxidized by 25 mM $KHCO_3$ and 50 μM H_2O_2 with PTP1B structure (green) obtained after soaking with H_2O_2 (PDB ID : 1OEM).
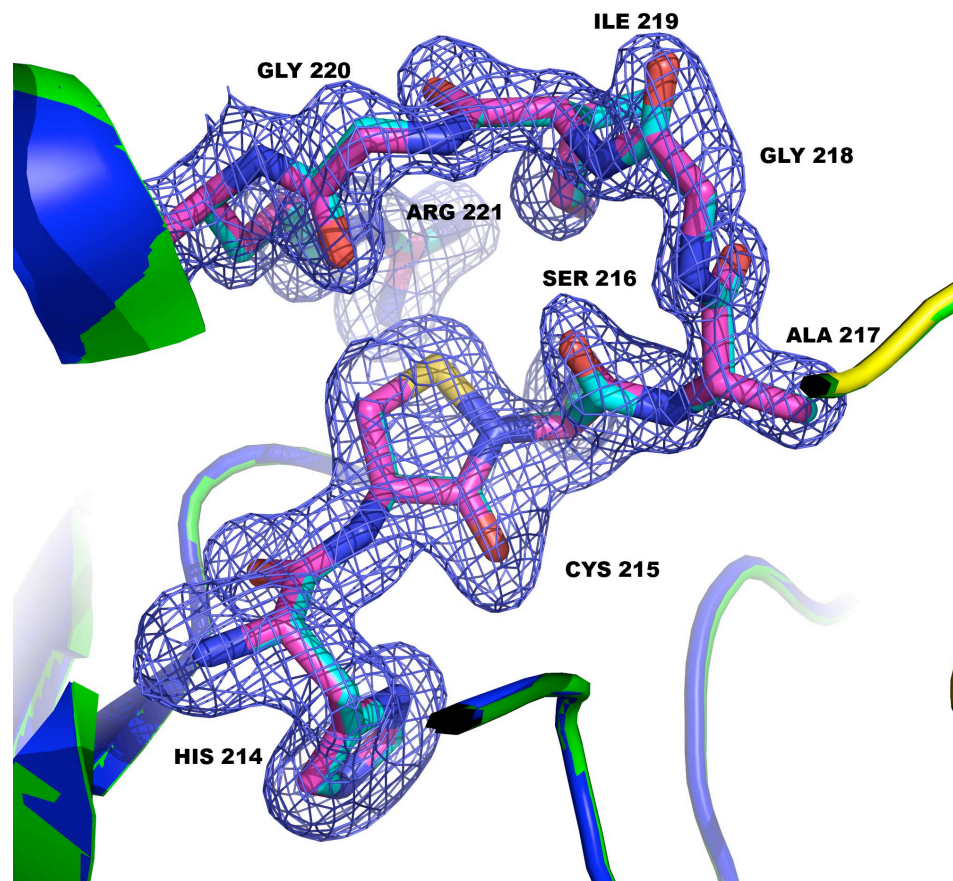
Fig. 7.2 Electron density maps showing the conformation of P loop. Note that the residues 214-221 corresponding to the loop were omitted before the calculation of maps. The blue cage represents the simulated annealing σ_A-weighted F_o-F_c omit map contoured at 3.0σ covering five-membered sulfenyl amide conformation of Cys215 and flanking residues.

Table 7.1 Data collection and refinement statistics.

Wavelength (Å)	1. 0000
Data collection resolution (Å)	44.21 - 1.70 (1.76 - 1.70)
No. of Observations	571812
No. of unique reflections	52290
$R_{merge}(I)$	0.063 (0.552)
Average I/σ	12.6 (2.8)
Completeness (%)	100.0 (100.0)
Redundancy	10.94 (11.00)
Refinement resolution (Å)	44.21 - 1.70 (1.76 - 1.70)
R_{cryst}	0.197 (0.290)
R_{free}	0.210 (0.303)
No. of protein residues	282
No. of protein atoms	2265
No. of water molecules	170
Average B-factor ($Å^2$)	
Protein	32.1
Water	38.0
Mg^{+2}	35.0
rmsd	
Bonds (Å)	0.006
Angles (deg)	1.051
Ramachandran plot	
Favored (%)	98.21
Allowed (%)	1.79
Outliers (%)	0.00
Coordinate error (Å)	0.21
PDB code	3SME

7.4 References

1.   Pawson, W. A. L. a. T., Phosphotyrosine Signaling: Evolving a New Cellular Communication System. *Cell* **2010,** *142* (5), 661-667.

2.   Saltiel, A. R. K. C. R., Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **2001,** *414* (6865), 799-806.

3.   Cohen, S., Epidermal growth factor (EGF): historical perspectives. *Horm. Proteins Pept.* **1985,** *12*, 299-304.

4.   Ahn, N., Introduction: Protein phosphorylation and signaling. Chem. Rev. 2001,; 101 (8), Introduction: Protein phosphorylation and signaling. *Chem. Rev.* **2001,** *101* (8).

5.   Johnson, L. N., Lewis, R. J., Structural Basis for Control by Phosphorylation.
. *Chem. Rev.* **2001,** *101* (8), 2209-2242.

6.   Alan R. Saltiel, C. R. K., Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **2001,** *414* (13), 799-806.

7.   Patti, M. E. K., C. R. , The insulin receptor—a critical link in glucose homeostasis and insulin action. *J. Basic Clin. Physiol. Pharmacol.* **1998,** *9*, 89-109.

8.   Tonks, N. K., PTP1B: From the sidelines to the front lines. *FEBS Letters* **2003,** *546*, 140-148.

9.   JANNIK N. ANDERSEN, O. H. M., GU¨NTHER H. PETERS, PAUL G. DRAKE, LARS F. IVERSEN, O. H. O., PETER G. JANSEN, HENRIK S. ANDERSEN, NICHOLAS K. TONKS, A. N. P. H. M., Structural and Evolutionary Relationships among Protein Tyrosine Phosphatase Domains. *MOLECULAR AND CELLULAR BIOLOGY* **2001,** *21*, 7117-7136.

10.   Tonks NK, D. C., Fischer EH., Characterization of the major protein-tyrosine-phosphatases of human placenta. *The Journal of Biological Chemistry* **1988,** *263* (14), 6731-7.

11.   Tonks NK, D. C., Fischer EH., Purification of the major protein-tyrosine-phosphatases of human placenta. *The Journal of Biological Chemistry* **1998,** *263* (14), 6722-30.

12.   Chernoff J, S. A., Jost CA, Erikson RL, Neel BG., Cloning of a cDNA for a major human protein-tyrosine-phosphatase. *Proc Natl Acad Sci U S A* **1990,** *87* (7), 2735-9.

13.   David Barford, J. C. K., Andrew J. Flint, Nicholas K. Tonks, Purification and Crystallization of the Catalytic Domain of Human Protein Tyrosine Phosphatase 1B Expressed in Escherichia coli. *Journal of molecular biology* **1994,** *239* (5), 726-730.

14.   Paul J. Ala, L. G., Milton C. Hillman, Mary Becker-Pasha, MinWei, Brian G. Reid, Ronald Klabe,Eddy W. Yue, Brian Wayland, Brent Douty, Padmaja Polam, Zelda Wasserman, Michael Bower,Andrew P. Combs, Timothy C. Burn, Gregory F. Hollis, and Richard Wynn, Structural Basis for Inhibition of Protein-tyrosine Phosphatase 1B by Isothiazolidinone Heterocyclic Phosphonate Mimetics. *The Journal of Biological Chemistry* **2006,** *281* (43), 32784-32795.

15.   Salmeen A, A. J., Myers MP, Meng TC, Hinks JA, Tonks NK, Barford D., Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate.*Nature* **2003,** *423* (12), 769-773.

16.   Rob L. M. van Montfort, M. C., Dominic Tisi, Robin Carr; Jhoti, H., Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B. *Nature* **2003,** *423*, 773-777.

17.   Sundaresan M, Y. Z., Ferrans VJ, Irani K, Finkel T., Requirement for generation of H2O2 for platelet-derived growth factor signal transduction. *Science* **1995,** *270* (5234), 296-9.

18.   Rhee, S. G., H2O2, a Necessary Evil for Cell Signaling. *Science* **2006,** *312*, 1882-1883.

19.   TONKS, N. K., Redox Redux: Revisiting PTPs and the Control of Cell Signaling. *Cell* **2005,** *121*, 667-670.

20.   Lee, S. R., Kwon, K.S., Kim, S.R. & Rhee, S.G. , Reversible inactivation of protein-tyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor. *The Journal of Biological Chemistry* **1998,** *273*, 15366-72

21.   Mahedev, K., Zilbering, A., Zhu, L. & Goldstein, B.J., Insulin-stimulated hydrogen peroxide reversibly inhibits protein-tyrosine phosphatase 1B in vivo and enhances the early insulin action cascade. *The Journal of Biological Chemistry* **2001,** *276*, 21938-21942

22.   Meng, T.-C., Buckley, D.A., Galic, S., Tiganis, T. & Tonks, N.K. , Regulation of insulin signaling through reversible oxidation of the protein tyrosine

phosphatases TC45 and PTP1B. *The Journal of Biological Chemistry* **2004,** *279*, 37716-37725.

23.   Denu, J. M. T., K. G. , Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry* **1998,** *37*, 5633–5642

24.   Stone, J. R., Hydrogen peroxide: a signaling messenger. . *Antioxidants Redox Signaling* **2006,** *8*, 243-270

25.   Winterbourn, C. C., Reconciling the chemistry and biology of reactive oxygen species. *Nature Chemical Biology* **2008,** *4*, 278-286.

26.   Bhattacharya, S., LaButti, J.N., Seiner, D.R. & Gates, K.S. , Oxidative inactivation of PTP1B by organic peroxides. . *Bioorganic Med. Chem. Lett.* **2008,** *18*, 5856-5859.

27.   LaButti, J. N., Chowdhury, G., Reilly, T.J. & Gates, K.S., Redox regulation of protein tyrosine phosphatase 1B by peroxymonophosphate. *J. Am. Chem. Soc.* **2007,** *129*, 5320-5321.

28.   Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **2005,** *41* (1), 207-34.

29.   Paul J. Ala1, L. G., Milton C. Hillman, Mary Becker-Pasha, MinWei, Brian G. Reid, Ronald Klabe,; Eddy W. Yue, B. W., Brent Douty, Padmaja Polam, Zelda Wasserman, Michael Bower,; Andrew P. Combs, T. C. B., Gregory F. Hollis, and Richard Wynn, Structural Basis for Inhibition of Protein-tyrosine Phosphatase 1B by Isothiazolidinone Heterocyclic Phosphonate Mimetics. *The Journal of Biological Chemistry* **2006,** *281*, 32784-32795.

30.   Sean R. Klopfenstein, a., , Artem G. Evdokimova, Anny-Odile Colson†, a, Neil T. Fairweathera, Jeffrey J. Neumana, Matthew B. Maiera, Jeffrey L. Graya, Gina S. Gerwea, George E. Stakea, Brian W. Howarda, Julie A. Farmera, Matthew E. Pokrossa, Thomas R. Downsa, Bhavani Kasibhatlaa and Kevin G. Petersa, 1,2,3,4-Tetrahydroisoquinolinyl sulfamic acids as phosphatase PTP1B inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2006,** *16* (6), 1574-1578.

31.   Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D., Automated structure solution with the PHENIX suite. *Methods Mol Biol* **2008,** *426*, 419-35.

32.   Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

33.	Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** *58* (Pt 11), 1948-54.

34.	Medinas, D. B., Cerchiaro, G., Trinidade, D.F. & Augusto, O. , The carbonate radical and related oxidants derived from bicarbonate buffer. *IUBMB Life* **2007,** *59*, 255-262.

Chapter 8.

Expression, purification, and crystallization of an atypical class C acid phosphatase from *Mycoplasma bovis*

Harkewal Singh, Thomas J. Reilly, Michael J.Calcutt and John J. Tanner

Abstract

Class C acid phosphatases (CCAPs) are 25 - 30 kDa bacterial surface proteins that are thought to function as broad specificity 5′-,3′-nucleotidases. Analysis of the newly published complete genome sequence of *Mycoplasma bovis* PG45 revealed a putative CCAP having a molecular weight of 49.9 kDa. The expression, purification, and crystallization of this new family member are described here. Standard purification procedures involving immobilized metal ion affinity chromatography and ion exchange chromatography yield highly pure and crystallizable protein. The crystals are grown in sitting drops at room temperature in presence of PEG 3350 and HEPES buffer at pH 7.5 and diffract to 2.3 Å resolution. Analysis of diffraction data suggests a primitive monoclinic space group with unit-cell parameters of a = 78 Å, b = 101 Å, c = 180 Å, and β = 92°. The asymmetric unit is predicted to contain six molecules, which are likely arranged as three dimers.

8.1 Introduction

Phosphatases are ubiquitous enzymes that catalyze the transfer of the phosphoryl group of a phosphomonoester to water [1]. They play many important roles in biology, including the regulation of complex signal transduction pathways, nucleotide metabolism, bone metabolism, antinociception, NAD$^+$ utilization, and the generation, acquisition and mobilization of inorganic phosphate.

Acid phosphatases, as the name implies, are optimally active at acidic pH. Multiple classes of acid phosphatases have been identified, including histidine acid phosphatases [2], purple acid phosphatases [3], alkaline phosphatase-like acid phosphatases [4], and bacterial nonspecific acid phosphatases [5]. The latter category includes class C acid phosphatases (CCAPs), the subject of this note.

First recognized as a group of related enzymes by Thaller *et al*. in 1998 [5b], CCAPs are membrane-anchored proteins that are distinguished by several conserved properties. At the primary structure level, CCAPs have a bipartite signature motif of [IV]-[VAL]-**D**-[IL]-**D**-E-T-[VM]-L-X-[NT]-X-X-Y near the N-terminus and [IV]-[LM]-X-X-G-**D**-[NT]-L-X-**D**-F near the C-terminus. The four essential Asp residues imbedded in the motif (in bold) identify CCAPs as DDDD superfamily phosphohydrolases. Other shared attributes include a polypeptide size of 25 - 30 kDa, dimer oligomeric state in solution, and the requirement of a metal cation for catalytic activity. Several class C enzymes have been characterized, including those from *Haemophilus influenzae* [6], *Bacillus anthracis*

[7], *Streptococcus equisimilis* [8], *Staphylococcus aureus* [9], *Helicobacter pylori* [10], *Elizabethkingia meningoseptica* [11] and *Clostridium perfringens* [12]

Crystal structures have been reported for the CCAPs from *Haemophilus influenzae* (aka P4), [13]) and *Pasteurella multocida* [14]. The structures showed that CCAPs belong to the haloacid dehalogenase structural superfamily, revealed the roles of the four conserved Asp residues and active site Mg^{2+} ion, and provided an understanding of substrate recognition.

Here, we report the crystallization of the CCAP from *Mycoplasma bovis*. This new family member is of interest because of its long polypeptide chain length, which is almost twice that of other CCAPs.

8.2   Methods and results

8.2.1 Identification of the MbCCAP gene and subcloning

Query of the complete genome sequence of *Mycoplasma bovis* PG45 [15] with CCAP genes identified an ORF that is predicted to encode a CCAP (MbCCAP, NCBI RefSeq number YP_004056356.1). The ORF encodes 463 residues. Multiple sequence alignments calculated using ClustalW2 show that MbCCAP has 18 - 25 % pairwise amino acid sequence identity with known CCAPs (Fig. 8.1). Furthermore, the alignments indicate the presence of the bipartite sequence common to CCAPs (green bars in Fig. 8.1). The MbCCAP polypeptide is considerably longer those of other CCAPs (Fig. 8.1). The extra residues appear as six insertions scattered throughout the polypeptide chain.

CCAPs are secreted and typically anchored to the outermost membrane by a non-cleaved signal sequence or an N-terminal lipidated Cys residue; therefore, the MbCCAP sequence was analyzed for signal peptides. Analysis with the SignalP 3.0 [16], PSORT 6.4 [17], PrediSi [18], and PSORT Signal-BLAST [19] servers suggests that the encoded protein contains a signal peptide with cleavage site between Thr26 and Cys27 (black triangle in Fig. 8.1). The signal peptide has Lys/Arg residues at positions -19, -22, and -23 relative to Cys27, followed by a hydrophobic region, and ending in the non-canonical lipoprotein box IAAT-C. These results suggest that the protein is expressed as a precursor polypeptide that is exported from the cytoplasm and cleaved to yield a 437-residue, 49.9 kDa mature enzyme. Furthermore, it is predicted that Cys27 is lipidated. It is noted that Cys27 aligns with cysteine residues of other CCAPs, including the Cys of P4 that is known to be lipidated [6b] (Fig. 8.1). The sequence similarity to other CCAPs, bipartite DDDD motif, predicted subcellular localization, and possible lipidation are consistent with classification of the enzyme into the CCAP family.

A synthetic gene encoding MbCCAP was obtained from Bio Basics Inc. (Markham, Ontario, Canada).. The gene was subcloned as an *NcoI–XhoI* fragment into pET20b such that the predicted signal peptide is replaced by the *pelB* leader sequence, and Cys27 is replaced by Met. The mature protein is exported to the *E. coli* periplasm and has a C-terminal His tag of Leu-Glu-His$_6$.

```
Hi     1   ......MKTTLKMTALAALSAFVLAGCG.......SHQMKSEE..................
Se     1   ....MKTKQ.VASVISLALSLFLVTGCAQLDHKANVNSKETVK..................
Ba     1   ...MKMKRGITTLLSVAVLSTSLVACS......GTTEKTVAK...................
Mb     1   MTNRKFFKGMLIGSAPVILLPTIAATCENTATKSDQKDNKIYENTKLNTVLSSLKSTWET
                                       ▲
                              α1                        α2
3ET4                       llllll  l............... lllllll  lllllll

Hi    31   ................HANMQLQQQAV.................LGLNWMQDSGEYKALAYQ
Se    39   ................QTKVTYSDEQLRSNE........NTMSVLWYQRAAEAKALYLQ
Ba    34   ................EEKVKLTDQQL..............MADLWYQTAGEMKALYYQ
Mb    61   LKDNEKHDAIKTIAKAMKSDVKLSDDEVHYLISELNKDVGTFGGVVNYISSVESLIGKEQ

           α3                              β1          β2      α4
3ET4    lllllllllll                     ----->      ----->  lllll

Hi    60   AYNAAKVAFDHAKVAK.....................GKKK.AVVADLDETMLDNSPYAG
Se    74   GYQLATDRLKNQLGQA.....................TDKPYSIVLDIDETVLDNSPYQA
Ba    63   GYNTGQLKLDAALAKG.....................TEKKPAIVLDLDETVLDNSPHQA
Mb   121   SYKFAKLAFDKLKESNKDKIDFGKKFMDNTSKIEADSNKSLPVVFMDIDETVLQN.DLTE

                         α5         β3        α6            β4
3ET4    lllll    lllllllll    ----->   lllllllll    ----------->    lllll

Hi    98   WQVQNNKPFDGKDWTRWVDARQSRAVPGAVEFNNYVNSHNGKVFYVTNRKD.STEKSGTI
Se   113   KNILEGTSFTPESWDVWVQKKEAKPVAGAKEFLQFADQNGVQIYYISDRA..VSQVDATM
Ba   102   MSVKTGKGY.PYKWDDWINKAEAEALPGSIDFLKYTESKGVDIYYISNRK..TNQLDATI
Mb   180   ASAMINGGYSGATKEKNDLKANRFAIPGAVEFINYVHENGGLVFYNSDMNQSTAVRDAVK

           α7          η1  β5                      α8
3ET4    llllll      lll  ----->                lllllllll

Hi   157   DDMKRLGFNGVEESAFYLK.........KDKSAKAARFAEIEKQ................
Se   171   ENLQKEGIPVQGRDHLLFL.........EEGVKSKEARRQKVKE................
Ba   159   KNLERVGAPQATKEHILLQ.........DPKEKGKEKRRELVSQ................
Mb   240   ENLKKVGVKFVQDFQFWMRGSMPYLAENEAEISDEKTKDLSDDSLKKLANKLKFTTEYRE

                                             β6          η2
3ET4                                    ----->        lll

Hi   192   ..............................................GYEIVLYVGDNLDDFGN
Se   206   ..............................................TTNLIMLFGDNLVDFAD
Ba   194   ..............................................THDIVLFFGDNLSDFTG
Mb   300   TPWITWTNSSAAYRLGKKVYKTDRMEALDKNNEGFKLMGDKKVPLKTIMKIGDNYNDFFD

                               α9                        η3      β7
3ET4    l...........................  lllllllll lll     ----->

Hi   209   TVYGKLN...............................ADRRAFVDQNQGKFGKTFIMLP
Se   223   .FSKKSE...............................EDRTALLSELQEEFGRQFIIFP
Ba   211   .FDGKSV...............................KDRNQAVTDSKAQFGEKFIIFP
Mb   360   RISKGKTNKERVDSYRSKELEINKYFDINGADVKHTVKKDNKYTLEKVPNPWKQVYVLVP

           α10       η4       α11
3ET4    llll      lll   lllllllllll

Hi   238   .NANYGGWEGGLAEGYFKKDTQGQIKARLDAVQAWDGK......
Se   251   .NPMYGSWESAVYKGD.KLDASHQLKERRKALESFEK.......
Ba   239   .NPMYGDWEGALYDYNFKKSDAEKDKIRHDNLKSFDAK......
Mb   420   GNAEYGGWLDKLGYGNVYNLYKEILNILKNTKYQEGPKPENNVI
```

Fig. 8.1 Sequence alignment of the CCAPs from *Haemophilus influenzae*, *Streptococcus equisimilis*, *Bacillus anthracis*, and *Mycoplasma bovis*. The black triangle denotes the predicted cleavage site. The green bars denote the CCAP signature sequence motif. The secondary structure elements were obtained from the *Haemophilus influenzae* structure (PDB code 3ET4). This figure was prepared with CLUSTALW2 and ESPript [23].

8.2.2 Expression and purification

The aforementioned plasmid was transformed into *E. coli* BL21(AI). The resulting transformants were plated on MDG_{amp} plates and incubated overnight at 37 °C. A single colony was picked and used to inoculate 1 L of culture. The protein was expressed via autoinduction [20] at 310 K with constant shaking at 300 rev min^{-1}. The cells were harvested by centrifugation at 2681g for 30 min at 277 K and resuspended in 20 mM phosphate, 20 mM imidazole, and 500 mM NaCl at pH 7.0. The cell pellet was flash frozen in liquid nitrogen and stored at 193 K.

Frozen cells were thawed at 277 K and ruptured using a French press at 1000 psi. Cell debris and unbroken cells were removed by centrifugation at 26891g for 30 minutes, followed by ultracentrifugation at 183960g for 1 hr. The resulting supernatant was loaded onto an immobilized metal-ion affinity chromatography column (Ni^{2+}-charged HiTrap, GE Healthcare) that had been equilibrated in 20 mM phosphate, 20 mM imidazole, and 500 mM NaCl at pH 7.0. The column was eluted with the equilibration buffer supplemented with 300 mM imidazole. Fractions exhibiting acid phosphatase activity with *p*-nitrophenylphosphate as the substrate were pooled and dialyzed overnight at 277 into 50 mM sodium acetate, 2.5 mM $MgCl_2$, and 5% (v/v) glycerol at pH 5.7 (buffer A). The resulting sample was loaded onto a cation exchange column (HiTrap SP, GE Healthcare) that had been equilibrated in buffer A. MbCCAP was eluted with linear gradient (0 – 1 M NaCl over 50 column volumes). The fractions were analyzed using SDS-PAGE, and the purest fractions were pooled and dialyzed into 50 mM sodium acetate, 50 mM NaCl, 2.5 mM $MgCl_2$, and 5 % (v/v) glycerol at pH 5.7. The sample was

concentrated to 5 - 6 mg mL^{-1} using a centrifugal concentrating device having a 30-kDa cutoff membrane. The protein concentration was estimated using the BCA method (Pierce kit).

8.2.3 Crystallization and preliminary analysis of X-ray diffraction data

Crystallization trials were performed at 293 K using the sitting drop method of vapor diffusion with Cryschem plates and reservoir volumes of 1 mL. Drops were formed by mixing 2 μl of the protein stock solution and 2 μl of the reservoir solution. Crystal screening trials using commercially available reagent kits were used to identify promising crystallization conditions. Optimization of the leads resulted in the growth of diffraction quality crystals using reservoir solutions containing 0.1 M - 0.2 M HEPES pH 7.5 and 22 - 27 % (w/v) PEG3350. The optimized crystals appeared as rods with a maximum dimensions of approximately 0.6 mm long and 0.08 mm thick (Fig. 2). In preparation for low temperature data collection, the crystals were soaked in 28 - 30 % PEG3350, 25% PEG 200, 0.1 M HEPES pH 7.5. The cryoprotected crystals were picked up with Hampton loops and plunged into liquid nitrogen.

The crystals were analyzed at Advanced Light Source beamline 4.2.2 using a NOIR-1 CCD detector. Autoindexing calculations with D*TREK [21] suggested a primitive monoclinic lattice with unit-cell parameters of a = 78 Å, b = 101 Å, c = 180 Å, and β = 92°. Using the method of Matthews [22], the asymmetric unit is predicted to contain six MbCCAP molecules and 48 % solvent (V_M = 2.37 Å^3 Da$^-$

[1]). Since CCAPs are known to form dimers in solution, the six molecules in the asymmetric unit are likely arranged as three dimers.
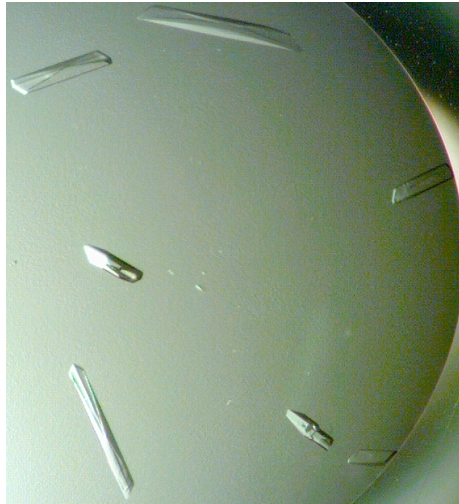
Fig. 8.2 Crystals of MbCCAP grown in a sitting drop.

A data set consisting of 719 frames was collected with detector distance of 160 mm, oscillation width of 0.25°/frame and exposure time of 5 s/frame. The data set was integrated and scaled to 2.3 Å resolution with D*TREK. Data processing statistics are listed in Table 8.1. Scaling of the data confirmed 2/*m* as the likely Laue class, as indicated by acceptable R_{merge} values of 0.033 in the lowest resolution bin, 0.416 in the highest bin, and 0.073 overall. The structure of MbCCAP will be solved with experimental phasing owing to a lack of suitable search model for molecular replacement calculations.

Table 8.1 Data- processing statistics

Space group	*P*2 or *P*2$_1$
Wavelength (Å)	1.0000
Unit-cell parameters (Å, °)	*a* = 78.0, *b* = 101.1, *c* = 180.5, b = 91.9
Protein molecules in asymmetric unit	6
V_m (Å^3/Da)	2.37
Solvent content (%)	48
Resolution (Å)	45.10 - 2.30 (2.38 - 2.30)
Total observations	442544
Unique reflections	124090
Redundancy	3.57 (3.39)
Completeness (%)	99.6 (99.0)
Mean *I/s(I)*	8.1 (2.1)
R_{merge}[†]	0.073 (0.416)
R_{merge}[†] in low-resolution bin	0.033

[†]R_{merge} = S$_{hkl}$S$_i$ | I_i(*hkl*) - $\langle$ *I*(*hkl*) $\rangle$ | / S$_{hkl}$S$_i$ I_i(*hkl*), where I_i(*hkl*) is the *i*th observation of reflection *hkl* and $\langle$ *I*(*hkl*) $\rangle$ is the weighted average intensity for all observations of reflection *hkl*.
Values in parentheses are for the outer resolution shell of data.

8.3 References

1.	(a) Vincent, J. B.; Crowder, M. W.; Averill, B. A., Hydrolysis of phosphate monoesters: a biological problem with multiple chemical solutions. *Trends Biochem. Sci.* **1992,** *17* (3), 105-10; (b) Vincent, J. B.; Crowder, M. W., *Phosphatases in cell metabolism and signal transduction: structure, function, and mechanism of action*. R.G. Landes Company: Austin, 1995.

2.	Rigden, D. J., The histidine phosphatase superfamily: structure and function. *Biochem. J.* **2008,** *409* (2), 333-48.

3.	Olczak, M.; Morawiecka, B.; Watorek, W., Plant purple acid phosphatases - genes, structures and biological function. *Acta Biochim. Pol.* **2003,** *50* (4), 1245-56.

4.	Felts, R. L.; Reilly, T. J.; Tanner, J. J., Structure of Francisella tularensis AcpA: prototype of a unique superfamily of acid phosphatases and phospholipases C. *J. Biol. Chem.* **2006,** *281* (40), 30289-30298.

5.	(a) Rossolini, G. M.; Schippa, S.; Riccio, M. L.; Berlutti, F.; Macaskie, L. E.; Thaller, M. C., Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. *Cell. Mol. Life Sci.* **1998,** *54* (8), 833-850; (b) Thaller, M. C.; Schippa, S.; Rossolini, G. M., Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. *Protein Sci.* **1998,** *7* (7), 1647-52.

6.	(a) Reilly, T. J.; Chance, D. L.; Smith, A. L., Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J. Bacteriol.* **1999,** *181* (21), 6797-6805; (b) Reilly, T. J.; Smith, A. L., Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr. Purif.* **1999,** *17* (3), 401-409.

7.	Felts, R. L.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Cloning, purification and crystallization of Bacillus anthracis class C acid phosphatase. *Acta Cryst.* **2006,** *F62* (Pt 7), 705-8.

8.	Malke, H., Cytoplasmic membrane lipoprotein LppC of Streptococcus equisimilis functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998,** *64* (7), 2439-42.

9.	du Plessis, E. M.; Theron, J.; Joubert, L.; Lotter, T.; Watson, T. G., Characterization of a phosphatase secreted by Staphylococcus aureus strain 154, a new member of the bacterial class C family of nonspecific acid phosphatases. *Syst. Appl. Microbiol.* **2002,** *25* (1), 21-30.

10.   Reilly, T. J.; Calcutt, M. J., The class C acid phosphatase of Helicobacter pylori is a 5' nucleotidase. *Protein Expr. Purif.* **2004,** *33* (1), 48-56.

11.   Passariello, C.; Schippa, S.; Iori, P.; Berlutti, F.; Thaller, M. C.; Rossolini, G. M., The molecular class C acid phosphatase of Chryseobacterium meningosepticum (OlpA) is a broad-spectrum nucleotidase with preferential activity on 5'-nucleotides. *Biochim. Biophys. Acta* **2003,** *1648* (1-2), 203-9.

12.   (a) Reilly, T. J.; Chance, D. L.; Calcutt, M. J.; Tanner, J. J.; Felts, R. L.; Waller, S. C.; Henzl, M. T.; Mawhinney, T. P.; Ganjam, I. K.; Fales, W. H., Characterization of a unique class C acid phosphatase from Clostridium perfringens. *Appl. Environ. Microbiol.* **2009,** *75* (11), 3745-54; (b) Wang, R.; Ohtani, K.; Wang, Y.; Yuan, Y.; Hassan, S.; Shimizu, T., Genetic and biochemical analysis of a class C non-specific acid phosphatase (NSAP) of Clostridium perfringens. *Microbiology* **2010,** *156* (Pt 1), 167-73.

13.   (a) Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of Recombinant Haemophilus Influenzae e (P4) Acid Phosphatase Reveals a New Member of the Haloacid Dehalogenase Superfamily. *Biochemistry* **2007,** *46* (39), 11110-9; (b) Singh, H.; Schuermann, J. P.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Recognition of Nucleoside Monophosphate Substrates by Haemophilus influenzae Class C Acid Phosphatase. *J. Mol. Biol.* **2010,** *404* (4), 639-649.

14.   Singh, H.; Malinski, T. J.; Reilly, T. J.; Henzl, M. T.; Tanner, J. J., Crystal structure and immunogenicity of the class C acid phosphatase from Pasteurella multocida. *Arch. Biochem. Biophys.* **2011,** *509* (1), 76-81.

15.   Wise, K. S.; Calcutt, M. J.; Foecking, M. F.; Roske, K.; Madupu, R.; Methe, B. A., Complete genome sequence of Mycoplasma bovis type strain PG45 (ATCC 25523). *Infect. Immun.* **2011,** *79* (2), 982-3.

16.   Bendtsen, J. D.; Nielsen, H.; von Heijne, G.; Brunak, S., Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004,** *340* (4), 783-95.

17.   Nakai, K.; Kanehisa, M., Expert system for predicting protein localization sites in gram-negative bacteria. *Proteins* **1991,** *11* (2), 95-110.

18.   Hiller, K.; Grote, A.; Scheer, M.; Munch, R.; Jahn, D., PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.* **2004,** *32* (Web Server issue), W375-9.

19.   Frank, K.; Sippl, M. J., High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics* **2008,** *24* (19), 2172-6.

20.   Studier, F. W., Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005,** *41* (1), 207-34.

21.   Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Cryst.* **1999,** *D55* (Pt 10), 1718-25.

22.   Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968,** *33*, 491-497.

23.   Gouet, P.; Courcelle, E.; Stuart, D. I.; Metoz, F., ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **1999,** *15* (4), 305-308.

Chapter 9.

Crystal structure and small angle X-ray scattering analysis of *Mycoplasma bovis* acid phosphatase

Author contribution – H.S. cloned the MbCCAP for SeMet. He expressed, purified and crystallized MbCCAP, P4, FtCCAP, BaCCAP and determined structure. H.S. also peformed SAXS experiments.

9.1 Background

To understand the structural diversity of CCAPs, we decided to determine the three dimensional structures of several family members. Since the structures of Bacilllus anthracis class C acid phosphatase (BaCCAP) and FtCCAP were determined earlier, this chapter will not describe those structures in detail. However, during my PhD tenure at Tanner lab, I was able to obtain new crystal forms of FtCCAP and BaCCAP that diffracted to 1.65Å and 1.30Å respectively (Table 9.2).

The major focus of this chapter is to highlight the structural diversity among CCAPs and also to shed light on the oligomerization of CCAPs. Hence this part of the thesis will focus on these aspects.

9.2 Structure determination of MbCCAP

There is no suitable homolog of MbCCAP is in the protein data bank, therefore molecular replacement is ruled out as a possible route to determine the crystal structure. An, obvious strategy is to use heavy atom methods to obtain the phasing information. We attempted to obtain a heavy atom derivative by soaking the crystals of native MbCCAP with various heavy atoms, but the

resulting maps obtained were not interpretable. Therefore, we decided to make a Se-Met[1] derivative of MbCCAP. Note that the sequence analysis of MbCCAP reveals that there are 12 methionine residues in overall 463 amino acids, which should be acceptable for selenomethionine labeling of the enzyme.

9.3 Materials and Methods

9.3.1 Sub-cloning of MbCCAP into pET26b(+) vector

Originally the MbCCAP gene was cloned using NcoI and XhoI restriction sites in pET20b(+) vector. During expression trials we realized that we could only obtain the transformants of MbCCAP if the expression strain was BL21AI. We were not able to obtain any colonies when we tried to transform the MbCCAP plasmid into BL21DE3 or even BL21(DE3)pLysS. The one plausible explanation for this is that pET20b(+) vector does not have stringent control over the protein expression or in other words the vector is "leaky", and MbCCAP being a toxic protein the optimal strain for the expression should have stringent control over protein expression. In fact, BL21AI strain is under the control of pBAD promotor[2] which allows stringent control.

However, for Se-Met labeleling, the protocol calls for the use of minimal media and expression either in BL21DE3 or BL21DE3plysS. Our attempts to transform the MbCCAP-pET20b(+) plasmid into the above strains were unsuccessful and also we were not able to overproduce MbCCAP Se-Met protein using pET20b(+) -BL21AI strain therefore we decided to subclone this gene into

a vector that has stringent control over protein expression. Obviously, there are several choices available for such vectors. However, pET20b(+) is ampicillin resistant so we wanted to use a vector, which is not ampicillin resistant and allowed us to use the NcoI and XhoI cloning sites such that the original gene could be directly transferred to the new vector. Hence, we decided to use pET26b(+) vector for this purpose. This vector has NcoI and XhoI sites available and also has LacI repressor, therefore, optimal for toxic protein expression. Additionally, this vector is kanamycin resistant so we could directly excise the plasmid from pET20b(+) and ligate it into pET26b(+) vector.

9.3.2 Protocol

Plasmid containing the MbCCAP gene was extracted from DH5α cells and excised using NcoI and XhoI restriction enzymes. The resulting construct was gel purified using 1% DNA agarose gel. Similarly, pET26b(+) vector was digested using NcoI and XhoI and gel purified as above. Ligation of MbCCAP excised gene into pET26b(+) was performed at 16 °C for overnight. The ligation product was transformed into DH5α and the transformants were plated onto LB agar supplemented with 40 µg/mL kanamycin. The following day, four single colonies were picked and a 10 mL LB broth culture was grown at 37 °C and 300 rev min^{-1}. Note that the culture was also supplemented with 40 µg/mL kanamycin. Four plasmids were purified and again excised with NcoI and XhoI restriction enzymes. A 0.8% DNA agarose gel was run and band corresponding to ~ 1500

bp was observed, confirming the presence of the right size gene product. Further, these plasmid samples were submitted to DNA core for sequencing and the sub-cloning was confirmed by DNA sequencing.

9.3.3 Expression of SeMet labeled MbCCAP

The MbCCAP-pET26b(+) construct was finally transformed into BL21(DE3) cells and plated onto LB agar supplemented with 40 µg/mL kanamycin. The selenomethionyl (Se-Met) labeled MbCCAP was expressed using metabolic inhibition method as described o by Sylvie Doublie[1]. The Se-Met labeled enzyme was purified using the protocol described for native MbCCAP except that all the buffers after Ni^{2+} chromatography contained 1mM THP as reducing agent. The Se-Met incorporation in the protein was verified by MALDI TOF MS.

9.3.4 Crystallization and data collection of Se-Met labeled MbCCAP

Several crystallization conditions were tested for the growth of Se-Met labeled MbCCAP, but most promising and well diffracting crystals were obtained in 23-28 % PEG3350, 0.1-0.2M HEPES at pH 7.5. Additionally, additive screening was performed in quest to achieve high-resolution diffraction but the crystals diffracted to almost similar resolution in the absence or presence of

additive screening. Briefly, the data from Se-Met derived crystals were collected at the NE-CAT beamline 24IDC of the Advanced Photon Source using a Quantum 315 detector and processed with HKL2000[3]. The data set that was used for single-wavelength anomalous diffraction (SAD) phasing was collected at the energy corresponding to the experimentally determined peak of f'' (wavelength = 0.97922 Å). This data set consisted of 180 frames collected with an oscillation width of 1.0° per image, detector distance of 350 mm, and exposure time of 1.0 s/image. The data were processed to 2.45 Å (Table 9.1).

9.3.5 Phasing and refinement

As mentioned earlier, molecular replacement calculations using search models derived from the rP4 and BaCCAP structure were unsuccessful, therefore the structure of MbCCAP was solved using SAD phasing. The phasing potential of each data set was analyzed with the HKL2MAP[4] interface to the SHELXC/D/E programs[4]. Promising data sets were input to PHENIX AutoSol for SAD phasing, density modification, and automated building calculation[5]. 64 of the expected 66 selenium sites (i.e., six MbCCAP molecules per asymmetric unit) were identified, which resulted in a figure of merit of 0.39 for reflections to 2.45 Å resolution. Density modification, which included 6-fold noncrystallographic symmetry averaging, increased the figure of merit to 0.69. PHENIX AutoBuild[5] was used for phase extension to 2.0 Å resolution and additional model building. The resulting model included 2206 protein residues and had an *R-factor* of 0.24,

186

R-free of 0.26, and map-model correlation coefficient of 0.79. The model from automated building was used as the starting point for several rounds of manual building in COOT[6] and simulated annealing refinement against the 2.0 Å resolution native data set in PHENIX[5]. The B-factor model used during refinement consisted of an isotropic B-factor for each non-hydrogen atom plus one TLS group per chain.

9.3.6 Small-Angle X-ray Scattering

Small angle x-ray scattering (SAXS) experiments were performed at beamline 12.3.1 of the Advanced Light Source. Prior to analysis, samples of MbCCAPb (6.0 mg/mL) and P4 (4.2 mg/mL) were dialyzed against 50 mM acetate, 50 mM NaCl, 2.5 mM $MgCl_2$ and 5% glycerol at pH 5.7.

Scattering intensities (I) were measured at different protein concentration as a function of the scattering vector $q = (4\pi \sin \theta) / \lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident beam. Exposure times of 0.5, 1.0, and 5.0 sec were used. The scattering curves collected from the protein sample were corrected for background scattering using intensity data collected from the dialysis buffer for each protein. The composite scattering curve used for shape reconstructions was generated with PRIMUS[7]. Scattering curves were subjected to indirect Fourier transform using GNOM[8] to yield the pair distribution function (P(r)), from which the radius of gyration (R_g) and the maximum particle dimension (D_{max}) were estimated. Shape reconstructions were performed using GASBOR[9]

with a D_{max} value of 77Å for P4 and 94Å for MbCCAP. Twenty independent structure calculations were performed, and the models were averaged and filtered using DAMAVER[10]. Theoretical scattering curves were calculated from atomic coordinates using FosX web server[11]. R_g from atomic coordinates was calculated using MOLEMAN[12].

Table 9.1 Data collection and refinement statistics

enzyme	Se-Met-MbCCAP	Native MbCCAP
active site ligand	None	Phosphate
space group	$P2_1$	$P2_1$
unit cell lengths (Å)	a = 78.3,b=101.1, c =180.8 β=92.2	a = 78.2,b=101.7, c =180.7 β=92.0
wavelength	0.97922	0.97922
resolution (Å)	50.00 – 2.45 (2.54 – 2.45)	50.00 – 2.00 (2.07 – 2.00)
no. of observations	369321	714573
no. of unique reflections	103559	190624
$R_{merge}(I)$	0.106 (0.434)	0.059 (0.338)
average I/σ	18.22 (2.62)	30 (4.5)
completeness (%)	99.0 (96.8)	99.8 (99.9)
redundancy	3.6 (3.3)	3.7 (3.7)
R_{cryst}		0.201
R_{free}[b]		0.226
no. of protein residues		2454
no. of protein atoms		18525
no. of water molecules		701
average B-factor (Å^2)		
protein		34.36
water		30.94
active site ligand		29.43
rmsd[c]		
bond lengths (Å)		0.008
bond angles (deg)		1.077
Ramachandran plot[d]		
favored (%)		97.0
allowed (%)		2.70
outliers (%)		0.2
PDB accession code		NA

Table 9.2 Data collection and refinement statistics for FtCCAP and BaCCAP

| Enzyme	FtCCAP	BaCCAP
Space group	$P2_12_12_1$	$P2_1$
Wavelength (Å)	0.9722	0.9722
Molecules in the asymmetric unit	6	4
Unit-cell parameters (Å, °)	a = 45.5, b = 126.0, c = 318.5	a = 65.0, b = 101.0, c = 75.84, β = 105.6
No. of observations	1449876	625677
No. of unique reflections	219089	229434
Rmerge(I)	0.069(0.381)	0.044 (0.428)
Average I/σ	35.3 (4.6)	21.0 (2.1)
Completeness (%)	99.9 (99.6)	99.7 (99.9)
Redundancy	6.6(6.4)	2.7(2.7)
Resolution (Å)	50.00-1.65 (1.71 - 1.65)	50.00-1.30 (1.35 - 1.30)

9.4 Results and Discussion

### 9.4.1 Structure of MbCCAP: A new member of the CCAP family

The 2.0Å resolution structure of MbCCAP was solved by single wavelength anomalous dispersion (SAD). The structure consists of two domains: an α/β core domain and an α domain (cap domain). The core domain is a reminiscent of the HAD superfamily members and supplies residues that participate in the catalysis. The active site is located at the junction of core and cap domain. The core domain consists of the five beta strand named in 3,2,1,4,5 order that are flanked by α helices on both sides. Note that the similar nomenclature was also used for the prototypical CCAP structure of P4[13, 14]. Briefly, the alpha helices that flank the beta strands are in the order as observed

for P4 i.e. αB, αE, αF on one side and αG and αH on the other side.

For convenience, we adopted the same nomenclature as was used for P4[14]. The mainly alpha helical domain is located above the C terminus edge of the strands of the beta sheets. MbCCAP structure starts with a span of three large alpha helices that are not seen in any other class C structures. Since we adopted a similar nomenclature as for P4, therefore these three alpha helices are labeled as αA-3 (resi 18-33), αA-2 (resi 36-55) and αA-1 (resi 59-69). This represents the insert1 (colored green in figure) in the MbCCAP structure as compared to the P4 structure. αA spans residues 73-84 and connects to αB of the core domain that spans residues 86-110. αA and αB , together they travel the majority of the length of the MbCCAP structure.

A

B
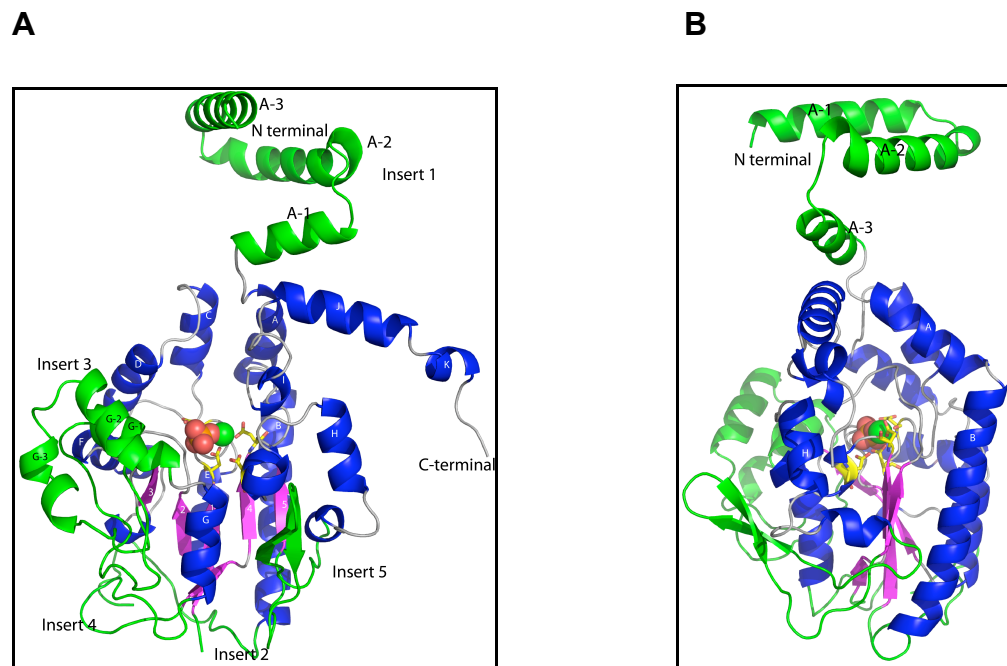
C
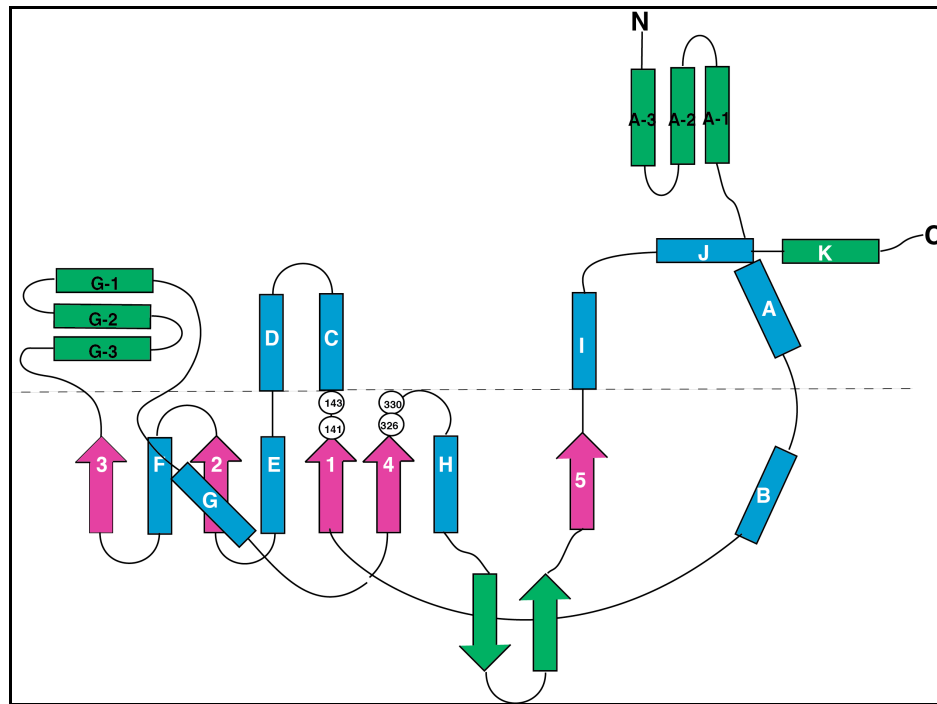
Fig. 9.1 Overall structure of MbCCAP. (a) Ribbon diagram of the MbCCAP subunit. The phosphate ion and Mg metal (green) are drawn as spheres. The side chains of the DDDD motif are shown in yellow. Strands are colored pink while helices are colored blue. All the inserts are shown in green. The N -and C- termini are labeled. (b) Another view of the MbCCAP protomer showing the long αB spanning through the protomer. αA and αH are highlighted as well. Note that view (b) is 90° rotated compared to the first view. (c) Secondary structure topology diagram of MbCCAP. The coloring scheme is similar to the one described above. All inserts (compared to P4) are colored green.

αB helix connects to the first β strand labeled as β1 via another insert that spans residues 111-134. This insert represents the second insert (labeled as

insert 2) as compared to the P4 structure is the form of an extended loop. Note that in P4 the αB and β1 connects via a short loop consisting of only seven residues. Therefore, in MbCCAP the insert is almost longer by 16 residues.

Briefly, β1 is composed of residues 135-140 and connects to another alpha helix αC (residues 149-157) via a loop that extends to next helix αD (residues 164-157). Together the αC and αD are inserted β1 between and αE (residues 182-192) of the core domain like P4. αE connects to β2 of the structure which then connects to αF that leads to another secondary structure element of MbCCAP labeled as β3.

From here, the MbCCAP has helix-loop-helix insert that connects the β3 and αG of the substructure. The insert is 59 residues long and spans from residues 234-293 in the MbCCAP structure. This insert is mainly comprised of alpha helical structure connected by loops varying in size. For instance, the smallest loop is made of four residues while the longest is 17 residues. The α helices belonging to this insert start after αF and going into αG therefore they are named as αG-3, αG-2 and αG-1 respectively, where αG-1 is the helix that is closest to αG (Fig. 9.1). Note, that this insert is labeled as insert 3 in the figure 9.1.

The αG (residues 294-303) connects to β4 of the structure via a loop made of 18 residues. Note that in P4 structure similar connection is observed but the loop connecting these two secondary structural elements is only five residues long, therefore in MbCCAP we have another insert that is 13 residues longer than P4. This insert is labeled as insert 4 in the figure. β4 of the structure is

connected to αH of the structure which is also similar to the one observed in P4.

In P4, the αH connects to the last beta strand that is β5 via a very short loop consisted of four residues, however that does not hold true for MbCCAP structure. Interestingly, the connection between β5 and αH involves 26 residues. Infact, upon closer inspection, it becomes evident that this connection has more than just 26 residues. This insert (labeled as insert 5) is comprised of two anti-parallel beta strands that were not observed before in any of the CCAP structures determined so far. This insert is also novel to MbCCAP in the sense that it was absent from all other CCAP structures.

Finally, the structural features extend into three more alpha helices labeled as αI, αJ and αK respectively and ends in with additional four residues extension.

9.4.2 Architecture of the active site

The active site of MbCCAP is located between the conjunction of core and α domain. Note the MbCCAP native structure is obtained with phosphate as ligand. The phosphate bound structure gives us a glimpse of the unperturbed active site. The four active site aspartates labeled as D141,D143,D326 and D330 (Fig. 9.2) represent the DDDD motif essential for catalysis which is the hallmark of CCAPs family. The D141 is presumably the nucleophile in MbCCAP that corresponds to D64 in P4. Briefly, the carbonyl oxygen of Asp141 is only 2.8 Å away from the P atom of the bound phosphate further confirming the role of this

residue as the nucleophile. The O1 atom of the phosphate is 2.5 Å away from the second Asp (D143) and likely represents the scissile oxygen of the phosphomonoester. Note that the residue D143, is the common Asp+2 present in the HAD superfamily phosphatases. This Asp+2 is known to behave as general acid/base that protonates the leaving group.

Four asp residues are clustered around the Mg^{2+} ion. Asp141, Asp143, Asp326, two water molecules (denoted as red sphere in the figure) and one O atom of phosphate act as six ligands to the octahedrally coordinated Mg^{2+} ion in the active site. Note that HAD superfamily members have an absolute requirement of Mg^{2+} ion for activity[15]. As indicated in the methods section, that during the purification and dialysis steps buffers were supplemented with $MgCl_2$ therefore Mg^{2+} seems very reasonable candidate for the metal. This is further supported by octahedral geometry and low B factors observed for Mg^{2+} during the refinement process (Table 9.1). Other important residues that participate significantly in the active site interactions are Lys294 and Ser200. In P4 the analogous residues are Lys161 and Thr124. To summarize, the overall active site of MbCCAP is similar to that of prototypical member P4.
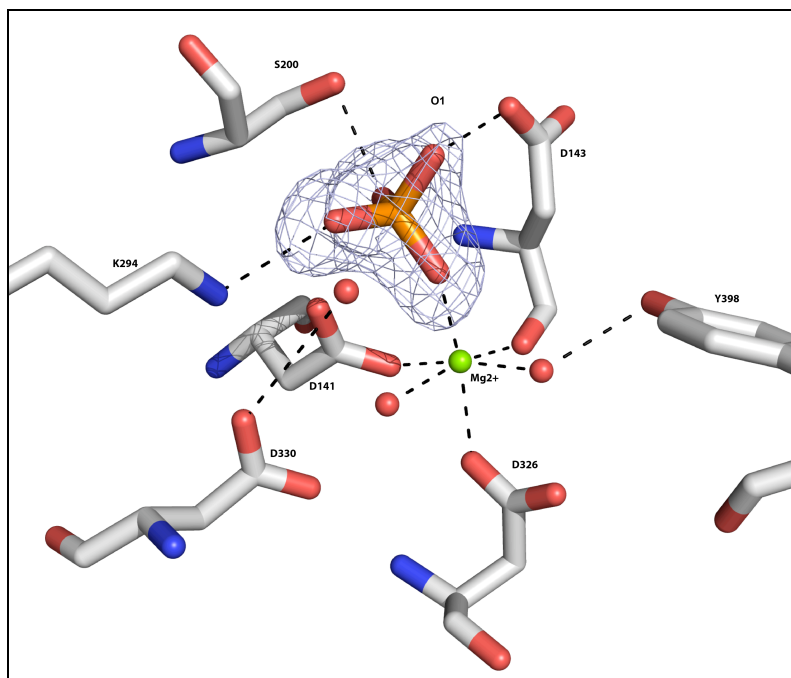
Fig. 9.2. The active site of MbCCAP in complex with inorganic phosphate. All the residues are labeled and shown as sticks. Mg^{2+} is represented by green sphere. The electron density map around phosphate is a σ_A weighted Fo-Fc omit map, contoured at 3σ. Note that prior to the calculation of the map, phosphate was omitted from the model.

9.4.3 Analysis of substrate binding in MbCCAP and P4

Superposition of substrate bound P4 and phosphate bound MbCCAP structure reveals that MbCCAP and P4 might utilize similar substrate binding strategies. For instance, the aromatic clamp in MbCCAP consists of Y162, K167 and Y398, whereas, in P4 these residues were F86, W91 and Y221 respectively. The only interesting difference between the two aromatic clamps is the presence of lysine in MbCCAP instead of tryptophan in P4 (Fig. 9.3). Based upon the two

structures, the base of the substrate might be clamped between the Y162 and Y398. Upon further inspection we notice that the major difference near the base-binding pocket of MbCCAP and P4 is that the presence of Glu153 in MbCCAP. This residue is predicted to be within the hydrogen bonding distance of the base of substrate. It is worthwhile mentioning that P4 does not have any residue that can make any direct contact to the base of the substrate. The corresponding residue in P4 is alanine. The one plausible implication of having a residue closer to the substrate-binding site is that it might help enforce substrate preferences in MbCCAP.
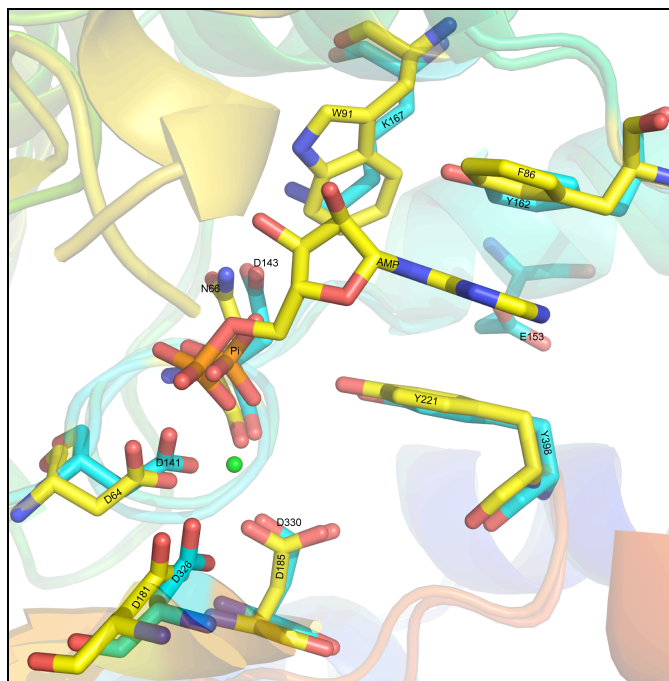

Fig. 9.3 Superposition of rP4-substrate complex (PDB code 3OCY) and MbCCAP acitve site showing the plausible substrate binding strategy in MbCCAP. The residues belonging to rP4 are shown in yellow while MbCCAP residues are shown in cyan. Each residue is numbered and labeled as represented in the crystal structure.

9.4.4 Dimeric structure of MbCCAP

The crystal structure of MbCCAP shows six molecules in the asymmetric unit that are arranged as three dimers. According to the crystal structure and sequence alignment, it is evident that this enzyme has several extra inserts. One of the first insert in the MbCCAP is arranged as three α helices, which are intertwined such that αA-1 and αA-3 of one subunit interacts with the similar α helices of another subunit to contribute in dimerization. This potential dimer interface area is unique to MbCCAP. Other potential dimer interfaces in the MbCCAP are represented by αA, αB, αH, αI, αJ and β5 that are also present in the prototype structure P4. Furthermore, another CCAP from *Francisella tularensis* has all the elements but lacks the N terminal insert and C terminal extension αJ (Fig. 9.4).

A
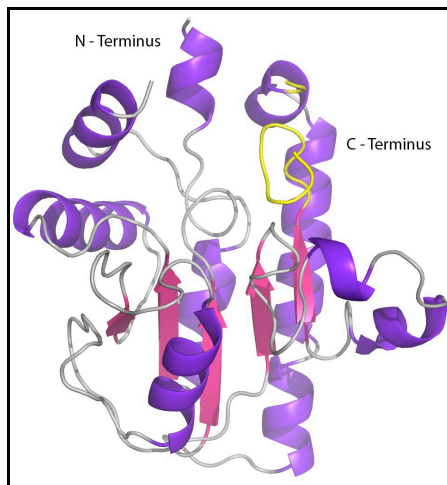
B
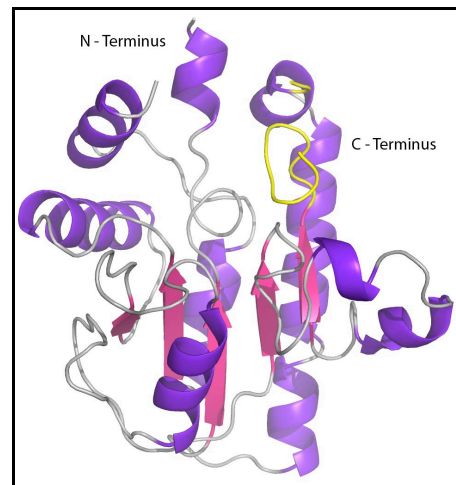
Fig. 9.4 (a) Figure showing the protomer of FtCCAP. Note that the C –termini is considerably shorter in this enzyme. (b) Figure showing rP4 protomer. N- and C- termini are labeled. The secondary structure coloring scheme is identical in both the structures and is similar to the one shown for MbCCAP.

Analytical ultracentrifugation studies (data not presented) have shown that FtCCAP also exists as dimer in solution. That raises a question about the biological significance of insert 1 of MbCCAP. Note that FtCCAP, P4, PmCCAP[16] and BaCCAP also lack this N terminal extension.

Also the dimer interface area calculations from PISA[17] revealed that MbCCAP has significantly higher dimer interface area (4865 $Å^2$) as compared to P4 (3054 $Å^2$). Therefore to understand the oligomeric state of MbCCAP and P4, we performed small angle X-ray scattering (SAXS) studies. The scattering curves calculated from the dimer and monomer of P4 and MbCCAP respectively are shown in figure. The scattering profile calculated from the dimer and monomers of both the enzymes is compared with the experimentally determined scattering curve. These results show that P4 and MbCCAP both exist as dimer in solution. Further we performed shape reconstruction calculations using GASBOR assuming a two-fold symmetry. The normalized spatial discrepancy (NSD) for the set of 20 GASBOR models is 1.08 ± 0.06 (MbCCAP) and for P4 0.930 ± 0.020 (Table 9.3), which is an indicator of good quality shape reconstruction. The model from gasbor was used for volumetric map calculations. The P4 and MbCCAP dimer were then superposed (Fig. 9.5) on this volumetric map to visually inspect the agreement between the crystal structure and shape of the

protein in solution.
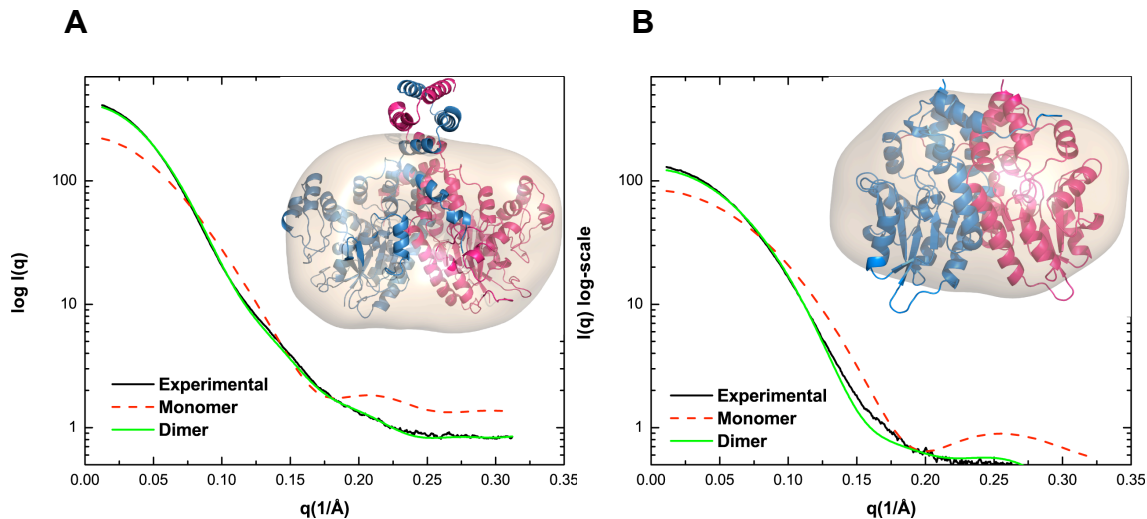
Fig. 9.5 SAXS analysis of MbCCAP(a) and P4(b). The thick solid black line represents the experimental scattering curve. The green solid line represents the scattering profile calculated for respective dimer. The dashed red line represents the scattering profile calculated for respective monomer. The inset shows the GASBOR shape reconstruction assuming two-fold symmetry. The surface represents the SAXS reconstruction (averaged and filtered). The structure of MbCCAP dimer and P4 dimer is superposed onto the surface in (a) and (b) respectively.

Table 9.3 Summary of SAXS analyses

R_g (Å) for MbCCA	30.8 ± 0.01
R_g (Å) for P4	24.5 ± 0.01
D_{max} (Å) for MbCCAP	94
D_{max} (Å) for P4	77
NSD of GASBOR reconstruction for MbCCAP	1.08 ± 0.06
NSD of GASBOR reconstruction for P4	0.93 ± 0.02
χ from FoxS	
MbCCAP Monomer	73.3
MbCCAP Dimer	7.9
P4 Monomer	42.5
P4 Dimer	6.5

This analysis reveals that overall P4 and MbCCAP dimer has good agreement between the dimer as represented by crystal structure and shape

reconstruction in solution. Upon close inspection, it becomes clear that the three N terminal α helical insert that is unique to MbCCAP are not in good agreement with the volumetric map calculated. This could be due the reason that these α helices are mobile in solution. Overall, our results indicate that MbCCAP and P4 exist as dimeric assembly in solution as indicated by the crystal structures. Furthermore, the presence of N terminal α helical insert that is unique to MbCCAP is not a required feature for the entire CCAPs. The result from AUC that FtCCAP is also a dimer excludes the requirement of C terminal αJ for dimerization. Finally, P4, PmCCAP, and FtCCAP exist as dimers in solution and none have the N terminal α helical insert, pointing out to a possibility that perhaps over the course of the evolution of these enzymes, the only essential part of dimerization was preserved as exhibited by most of the other members of this family. Therefore, neither N – terminal insert nor the C-terminal insert extenions are essential for dimerization.

9.5 References

1.   Doublie, S., Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* **1997,** *276*, 523-30.

2.   Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J., Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **1995,** *177* (14), 4121-30.

3.   Minor, Z. O. a. W., Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology* **1997,** *276:Macromolecular Crystallography* (A), 307-326.

4.   Schneider, T. P. T. R., HKL2MAP: a graphical user interface for phasing with SHELX programs. *J. Appl. Cryst.* **2004,** *37*, 843-844.

5.   Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **2002,** *58* (Pt 11), 1948-54.

6.   Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2126-32.

7.   P. V. Konarev, V. V. V., A. V. Sokolova, M. H. J. Koch and D. I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **2003,** *36*, 1277-1282.

8.   Svergun, D. I., Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **1992,** *25*, 495-503.

9.   Svergun, D. I.; Petoukhov, M. V.; Koch, M. H., Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **2001,** *80* (6), 2946-53.

10.   Svergun, V. V. V. a. D. I., Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* **2003,** (36), 860-864.

11.   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.

12.   Kleywegt, G. J., Validation of protein models from Calpha coordinates alone. *J Mol Biol* **1997,** *273* (2), 371-6.

13.   Singh, H.; Schuermann, J. P.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Recognition of nucleoside monophosphate substrates by Haemophilus influenzae class C acid phosphatase. *J Mol Biol* **2010,** *404* (4), 639-49.

14.   Felts, R. L.; Ou, Z.; Reilly, T. J.; Tanner, J. J., Structure of recombinant Haemophilus influenzae e (P4) acid phosphatase reveals a new member of the haloacid dehalogenase superfamily. *Biochemistry* **2007,** *46* (39), 11110-9.

15.   Lu, Z.; Dunaway-Mariano, D.; Allen, K. N., The catalytic scaffold of the haloalkanoic acid dehalogenase enzyme superfamily acts as a mold for the trigonal bipyramidal transition state. *Proc Natl Acad Sci U S A* **2008,** *105* (15), 5687-92.

16.   Singh, H.; Malinski, T. J.; Reilly, T. J.; Henzl, M. T.; Tanner, J. J., Crystal structure and immunogenicity of the class C acid phosphatase from Pasteurella multocida. *Arch Biochem Biophys* **2011,** *509* (1), 76-81.

17.   Henrick, E. K. a. K., Detection of Protein Assemblies in Crystals. *Lecture Notes in Computer Science* **2005,** *3695*, 163-174.

VITA

Harkewal Singh was born in India. He did his M.Sc. in chemistry from Kumaun University at Nainital (India), and Post Graduate Diploma in Bioinformatics and Applied Biotechnology from Institute of Bioinformatics and Applied Bioctechnology (IBAB), Bangalore (India). He worked as an Application Specialist with Millipore India from November 2003-October 2004. He then joined Jaypee University of Information Technology (H.P.) as lecturer in bioinformatics from November 2004- July 2006. He joined chemistry department in the Fall-06 and subsequently joined Prof. John J. Tanner's group in January 2007. Starting Fall-11 he will be working as Postdoctoral Research Associate at University of California, Los Angeles in Prof. David Eisenberg's research group.