

STRUCTURAL STUDIES OF ACID PHOSPHATASES FROM PATHOGENIC
BACTERIA

A Dissertation
presented to
the faculty of the Graduate School
University of Missouri-Columbia

In Partial Fulfillment
Of the Requirements for the Degree

Doctor of Philosophy

by
RICHARD LEVI FELTS
Dr. John J. Tanner, Dissertation Supervisor
MAY 2007

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

STRUCTURAL STUDIES OF ACID PHOSPHATASES FROM PATHOGENIC
BACTERIA

Presented by Richard Levi Felts

A candidate for the degree of Doctor of Philosophy

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

Dr. John J. Tanner

Dr. Thomas J. Reilly

Dr. Carol Deakyne

Dr. Lesa J. Beamer

Dr. Kent S. Gates

To my wife and best friend, Kim.

ACKNOWLEDGEMENTS

This dissertation is a reflection of the work that I have accomplished while at the University of Missouri, but I was not without guidance and assistance, throughout the graduate process there were many people that worked tirelessly to help me succeed. Here I acknowledge those who have contributed to my successful completion of this degree.

First off I would like to extend my great appreciation and admiration to my mentor and friend Dr. Jack Tanner. The art of crystallography is very difficult to teach, and even more difficult to understand but Jack has a way of making even some of the most challenging of concepts seem rather trivial. When asked for advice on anything from general science ideas to prospective employment he was always willing to listen and give advice. His attention to detail would make it sometimes difficult to get a quick answer, and many times a five minute meeting would turn into two-hours of brainstorming. Jack is one of the few people that I have been associated with that enjoyed working late into the night. It was great when I would work late and hit a roadblock and he would be around for help. I often wondered when he found time to sleep, it seemed that if he wasn't at work, or at the beamline, he was trying to spend time with his family. I will miss all of the great time that we had together outside of the lab, from the beamline trips in Berkeley, to conferences in Massachusetts, where he was assaulted on an airplane, I greatly enjoyed learning from you.

Equally important to my success as a scientist is Dr. Tom Reilly. When I entered graduate school I was lacking basic molecular biology skills, but all of

that changed when I was introduced to Tom. Tom is one of the most well rounded scientists that I have had the pleasure of knowing. Many of the first conversation that we had often involved a scientific history lesson, ending with "I can't believe that you don't know who (insert famous scientist here) is." I would like to think that I have gotten better, but who knows? Tom took the time to teach me many molecular biology skills that have made me a better scientist. I have had the great fortune of returning the favor by teaching Tom, everything that I know about crystallography. He has a great love of science and I will miss many of our lunches where we would solve the world's problems on the back of a napkin, and then somehow misplace it on the way back to the lab. I will miss singing Journey songs, playing musical trivia, and trying to fit the word homunculus into a presentation all while working in a 3 x 5 foot lab space. I hope that you realize that prisoners get more lab space than you.

As a group I would like to thank the remaining members of my committee, Dr. Carol Deakyne, Dr. Lesa Beamer, and Dr. Kent Gates. I tried not to bother you with my scientific problems, but rather enjoyed the many conversations that we had about unimportant issues. I realize that you have higher priorities than me but I greatly appreciate all that you have done in helping me succeed as a scientist.

Members of the Tanner lab who I would like to thank are Dr. Jon Schuermann, Dr. Jermaine Lemont Jenkins, Dr. Tommi White, Dr. John Larson, Dale Karr, Zhonghui Ou, Joshua Pearson, Dhiraj Srivastava, Emily Aurturo, Min Zhang, Li Ma, and Harkewal Singh. It has been my pleasure to work and learn with you. I think that we had great lab chemistry, which made all of us better

scientists; each of you had a different impact on my research and my life and for that I am forever grateful.

Finally, I would have never chosen chemistry as my career without first meeting Dr. Lawrence Pilgram, my high school chemistry teacher. Dr. Pilgram had a love for teaching chemistry that I have yet to see duplicated. He understood that chemistry was a difficult subject, so he had to make it fun and exciting in order to keep teenagers with raging hormones interested in what was on the chalkboard and not who was sitting around them. From the first lecture I was hooked on chemistry, it fascinated me as much then as it does today. Without a doubt I owe Dr. Pilgram a great deal of thanks for introducing me to my hobby, my passion, and my love of science. Thank you.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	x
LIST OF TABLES	xiii
ABSTRACT	xiv
ABBREVIATIONS	xv
Chapter	
1. INTRODUCTION	
1.1 INTRODUCTION	2
1.2 FRANCISELLA TULARENSIS	11
1.3 VIRULENCE MECHANISM OF F. TULARENSIS	14
1.4 REFERENCES	22
2. CRYSTALLIZATION OF ACPA	
2.1 METHODS AND RESULTS	31
2.1.1 Cloning, expression and purification	31
2.1.2 Crystallization, Data collection and processing	32
2.2 REFERENCES	39
3. THE STRUCTURE OF ACPA	
3.1 EXPERIMENTAL PROCEDURES	41
3.1.1 Crystallization and X-ray diffraction and data collection	41
3.1.2 Phasing and refinement calculations	42
3.1.3 Site directed mutagenesis	46
3.1.4 Determination of metal ion in AcpA	46

3.2	RESULTS	47
3.2.1	Overall structure of AcpA	47
3.2.2	Active site architecture and implications for catalytic mechanism	56
3.2.3	Comparison to other protein structures	67
3.2.4	Conservation of active site residues in the AcpA/PlcH superfamily	73
3.3	DISCUSSION	79
3.3.1	The AcpA family of phosphataees	79
3.3.2	Catalytic mechanism of AcpA-like PLC's	80
3.3.3	Role of AcpA in virulence of <i>F. tularensis</i>	81
3.4	REFERENCES	83
4.	CRYSTALLIZATION OF HISTIDINE ACID PHOSPHATASE	
4.1	INTRODUCTION	89
4.2	METHODS AND RESULTS	91
5.2.1	Cloning, expression and protein purification	91
5.2.2	Crystallization	92
5.2.3	Data collection and processing	95
4.3	REFERENCES	98
5.	STRUCTURE OF HISTIDINE ACID PHOSPHATASE	
5.1	INTRODUCTION	101
5.2	MATERIALS AND METHODS	103
5.2.1	Enzyme assay	103
5.2.2	Expression of HAP	106

5.2.3	Purification of HAP	107
5.2.4	Crystallization of HAP	108
5.2.5	Data collection	110
5.2.6	Structure determination and refinement	112
5.3	RESULTS AND DISCUSSION	115
5.3.1	Overall structure of HAP	115
5.3.2	Structure of HAP/vanadate complex	121
5.3.3	Structure of HAP/tartrate complex	122
5.3.4	Structure of HAP/acetate complex	123
5.3.5	Development of hydrophobic clamp hypothesis	125
5.3.6	Structure of the HAP/tartranilic acid	128
5.3.7	Determination of kinetic parameters	132
5.3.8	Comparison of HAP to HPAP	132
5.3.9	Other histidine acid phosphatase structures	133
5.4	REFERENCES	141
6.	CRYSTALLIZATION AND STRUCTURE DETERMINATION OF BACILLUS ANTHRACIS CLASS C ACID PHOSPHATASE	
6.1	INTRODUCTION	146
6.2	METHODS AND RESULTS	148
6.2.1	Cloning expression and purification	148
6.2.2	Crystallization	152
6.2.3	Data collection and processing	154
6.2.4	Structure determination	158
6.3	DISCUSSION	165

6.4	REFERENCES	170
7.	STRUCTURE OF RP4	
7.1	INTRODUCTION	175
7.2	MATERIALS AND METHODS	177
7.2.1	Crystallization and heavy atom derivative preparation	177
7.2.2	X-ray diffraction data collection	178
7.2.3	Phasing and refinement calculations	179
7.3	RESULTS	182
7.3.1	P4:a new member of the HAD superfamily	182
7.3.2	Structure of the active side	189
7.3.3	Dimeric structure of rP4	194
7.3.4	Structural context of rP4 mutations designed for vaccine studies	199
7.3.5	Structural context of the proposed KVAFDH heme binding motif	200
7.4	DISCUSSION	203
7.4.1	Structure of rP4 and relationship to other HAD enzymes	203
7.4.2	Implications for vaccine design	204
7.4.3	Role of e(P4) in heme acquisition by <i>H. influenzae</i>	206
7.5	REFERENCES	212

APPENDIX

A1. EXPRESSION, PURIFICATION, CRYSTALLIZATION AND STRUCTURE DETERMINATION OF FRANCISELLA TULARENSIS CLASS C NON SPECIFIC ACID PHOSPHATASE

A1.1 EXPERIMENTAL 219

A1.1.1 Expression and purification 219

A1.1.2 Crystallization 221

A1.1.3 Structure determination 223

A1.2 REFERENCES 225

LIST OF PUBLICATIONS 226

VITA 227

LIST OF FIGURES

Figure	Page
1.1 Phosphatase mechanistic pathways	4
1.2 Phosphatase reactions	5
1.3 A classification scheme for phosphatases	8
1.4 Reported cases of tularemia in the US 1990-2000	13
1.5 The <i>Francisella tularensis</i> pathogenicity island	16
2.1 Three crystal forms of recombinant <i>F. tularensis</i> AcpA	38
3.1 Anomalous difference Patterson Map	43
3.2 Stereographic ribbon drawing of AcpA chain A	50
3.3 Secondary structure topology diagram of AcpA	52
3.4 Dimeric structure of AcpA	53
3.5 Inductively coupled plasma-emission spectroscopy	59
3.6 <i>p</i> NPP modeled in the AcpA trough	60
3.7 Active site of AcpA	61
3.8 Summaries of the proposed catalytic mechanisms of AcpA and arylsulfatase A (ASA)	65
3.9 Comparison of AcpA to arylsulfatase A (ASA) and alkaline phosphatase (AlkP)	69
3.10 Amino acid sequence alignment of AcpA/PlcH superfamily members	71
3.11A Ramachandran plot of AcpA chain A	75

LIST OF FIGURES

Figure	Page
3.11B Ramachandran plot of AcpA chain B.	76
3.12 Decavanadate and its interactions with AcpA	78
4.1 Sequence alignment between <i>F. tularensis</i> HAP and <i>L. pneumophila</i> MAP	90
4.2 Crystals of <i>F. tularensis</i> HAP	94
5.1 Kinetic analysis of HAP	104
5.2 Overall structure of HAP	117
5.3 HAP dimer	119
5.4 Inhibitors of HAP	120
5.5 Inhibitor interactions within the active site	126
5.6 Hydrophobic clamp interactions	129
5.7 Active site overlap of HAP structures	130
5.8 Overlap of HAP with HPAP	134
5.9 Sequence alignments between HAP and other histidine acid phosphatases with known structures	135
5.10 A-D Ramachandran plots of HAP structures	138
6.1 SDS-PAGE analysis of fractions from Ni-chelate chromatography	151
6.2 Crystals of <i>B. anthracis</i> class C acid phosphatase	153
6.3 Tungstate derivative anomalous difference Patterson from <i>Bacillus anthracis</i> Class C acid phosphatase	160

LIST OF FIGURES

Figure	Page
6.4 Ramachandran plot of <i>Bacillus anthracis</i> class C NSAP	162
6.5 Overlap of class C NSAP's	167
6.6 Class C NSAP active site overlap	168
6.7 Class C NSAP sequence alignment	169
7.1 Overall structure of rP4	185
7.2 Comparison of rP4 with 4 other HAD superfamily enzymes	187
7.3 Active site of rP4	192
7.4 Dimeric structure of rP4	196
7.5 Stabilization of the active site through intersubunit interactions	198
7.6 The KVAFDH heme binding motif	202
7.7 Structural context of rP4 mutations	208
7.8 Ramachandran plot of rP4	210
7.8 Anomalous difference Patterson map	211
A1.1 Ramachandran plot of <i>Francisella tularensis</i> class C NSAP	224

LIST OF TABLES

Table	Page
2.1 Summary of three crystal forms of recombinant AcpA	36
2.2 Data-collection and processing statistics (crystal form III)	37
3.1 Data collection and refinement statistics	45
3.2 Intersubunit hydrogen bonds	55
4.1 Data collection and processing statistics for <i>F. tularensis</i> HAP	97
5.1 Data collection, phasing, and refinement statistics	114
6.1 Data collection and processing statistics for <i>B. anthracis</i> class C.	157
6.2 Data collection and refinement statistics	164
7.1 Data collection and processing statistics for rP4	181
A1.1 Data collection and refinement statistics for <i>Francisella tularensis</i> class C NSAP	222

STRUCTURAL STUDIES OF ACID PHOSPHATASES FROM PATHOGENIC BACTERIA

Richard Levi Felts

Dr. John J. Tanner, Dissertation Supervisor

ABSTRACT

Acid phosphatases are important, ubiquitous, and diverse group of enzymes that catalyze the transfer of phosphoryl from phosphomonoester to water forming inorganic phosphate and alcohol. These enzymes play critical roles in numerous processes and pathways including virulence. It has been suggested that the pathogenic bacteria *Francisella tularensis* utilizes an acid phosphatase to aid in the escape from the phagosome. As part of this research, the three dimensional X-ray crystal structures of three phosphatases from *F. tularensis* were determined. The three structures from *F. tularensis* were AcpA, HAP, and class C. Parallel to those structural studies homologous structures of class C acid phosphatases were determined from *Haemophilus influenzae* and *Bacillus anthracis*, those structures are also reported here. In all five unique acid phosphatase structures were determined in this research project that have led to a better understanding of the novelty of catalytic mechanisms, a clearer understanding of substrate binding, and a new foundation for structural investigation.

ABBREVIATIONS

Å	Angstrom
ACP	Acid Phosphatase
AlkP	Alkaline Phosphatase
ALS	Advanced Light Source
AMP	Adenosine Monophosphate
APS	Advanced Photon Source
ATP	Adenosine Triphosphate
Bis-Tris	3-bis(tris(hydroxymethyl)methylamino)methane
CDC	Centers for Disease Control and Prevention
CPK	Corey, Pauling, Koltin Color Scheme
EDTA	Ethylenediaminetetraacetic acid
FPI	<i>Francisella tularensis</i> 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
IPTG	isopropylthiogalactoside
K_{av}	Partition Coefficient
K_m	Michaelis Constant
LB	Luria-Bertani
LVS	Live Vaccine Strain
MALDI-TOF	Matrix Assisted Laser Desorption Time-of-Flight Mass

Spectrometry

MAP	Major Acid Phosphatase
MES	2-Morpholinoethanesulfonic acid
NADP ⁺	Nicotinamideadenine dinucleotide phosphate
NADPH	Reduced Nicotinamideadenine dinucleotide phosphate
NCS	Non-crystallographic Symmetry
NSAP	Bacterial Non-specific Acid Phosphatase
OD	Optical Density
PEG	Polyethylene Glycol
PEGMME	Polyethylene Glycol Mono-methyl ether
PCR	Polymerase Chain Reaction
PLC	Phospholipase C
PlcH	Hemolytic Phospholipase C
<i>p</i> NP	<i>p</i> -nitrophenol
<i>p</i> NPP	<i>p</i> -nitrophenyl phosphate
<i>p</i> NPPC	<i>p</i> -nitrophenyl phosphorylcholine
PTP	Protein Tyrosine Phosphatase
PVDF	Polyvinylidene fluoride
SAD	Singlewavelength Anomalous Diffraction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Tris	Tris(hydroxymethyl) methylamine
V_{max}	Maximal Velocity

1.

INTRODUCTION

1.1 Introduction

Phosphatases are an important, ubiquitous, and diverse group of enzymes that catalyze the transfer of phosphoryl from phosphomonoesters to water forming inorganic phosphate and alcohol. In humans, as well as other eukaryotes, phosphatases play critical roles in numerous processes and pathways, including signal transduction, cell cycle control, cell transformation, glycogen metabolism, muscle contraction, protein synthesis, T-cell and B-cell activation, insulin activation, transformation of cancer and cell adhesion¹. There is such a dependence on the dephosphorylation of important molecules within higher level organisms it is unsurprising that pathogenic bacteria have evolved phosphatases and kinases that alter host phosphorylation levels, thereby disrupting host signaling pathways and facilitating virulence. Because of their central importance in eukaryotic and prokaryotic biology, much attention has been given to understanding the structures, functions, and catalytic mechanisms of phosphatases.

It is generally accepted that the catalytic mechanism involves two consecutive in-line nucleophilic attacks at the phosphorus atom. In this reaction the first attack is by a protein-associated nucleophile on the incoming phosphomonoester, which produces a phosphoenzyme intermediate, and the second attack is by solvent water on the phosphoenzyme intermediate²⁻⁶. As a result of the two nucleophilic attacks there are two inversions of the phosphate moiety and retention of configuration at the phosphorus. Figure 1.1 shows a general mechanism. An exception is purple acid phosphatase, which has a bi-

nuclear metal center that activates a water molecule for nucleophilic attack at the substrate P atom.

In addition to providing an active site that is complementary in shape and electrostatic interaction to the transition state, phosphatases typically provide a general acid to protonate the leaving group during formation of the phosphoenzyme intermediate, and a general base to activate the water nucleophile in the final step. In some phosphatases these two roles are shared by a single amino acid. For example human protein tyrosine phosphatase 1B (PTP1B) has Asp181 that has been determined to be both the general acid and general base⁷. Overall phosphatase reaction can proceed via one of three reaction pathways: dissociative, associative, or concerted. Figure 1.2 highlights these three general mechanistic pathways. In the dissociative mechanism a transient PO_3^{-1} ion, metaphosphate, is formed prior to attack of the nucleophile. This reaction is also referred to as an $\text{S}_{\text{N}}1$ -type two-step reaction. The associative mechanism is also a two-step reaction although it proceeds through a pentacoordinate intermediate called phosphorane. The concerted mechanism has bond formation of the nucleophile and bond cleavage of the leaving group taking place in the transition state. Both associative and dissociative mechanistic characteristics are possible in the transition state depending on the relationship between bond formation and cleavage⁸. The available biochemical data for many phosphatases indicate that the reaction proceeds via a dissociative-like transition state. These phosphatases include alkaline phosphatase (AlkP)⁹, protein tyrosine phosphatase (PTP)^{3, 10}, and Cdc25A phosphatase¹¹.

Figure 1.1

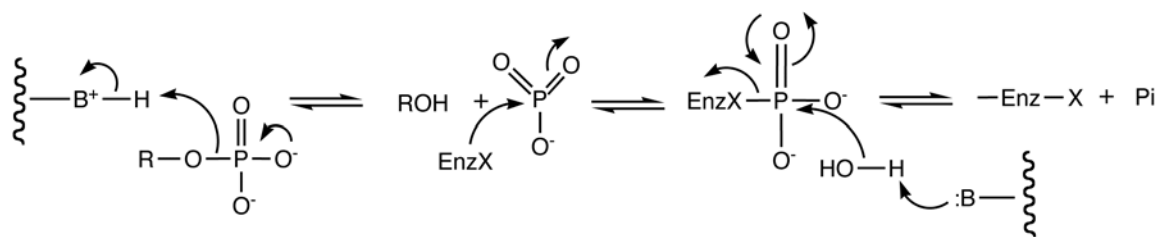


Figure 1.1. General mechanism for phosphomonoester hydrolysis.

In this mechanism a general acid donates a proton forming the alcohol and the meta phosphate intermediate. Nucleophilic attack at the P atom then takes place forming the phosphoenzyme intermediate. The general base abstracts a proton from water forming the hydroxyl nucleophile that attacks the P atom forming inorganic phosphate and retuning the enzyme to its native form.

Figure 1.2

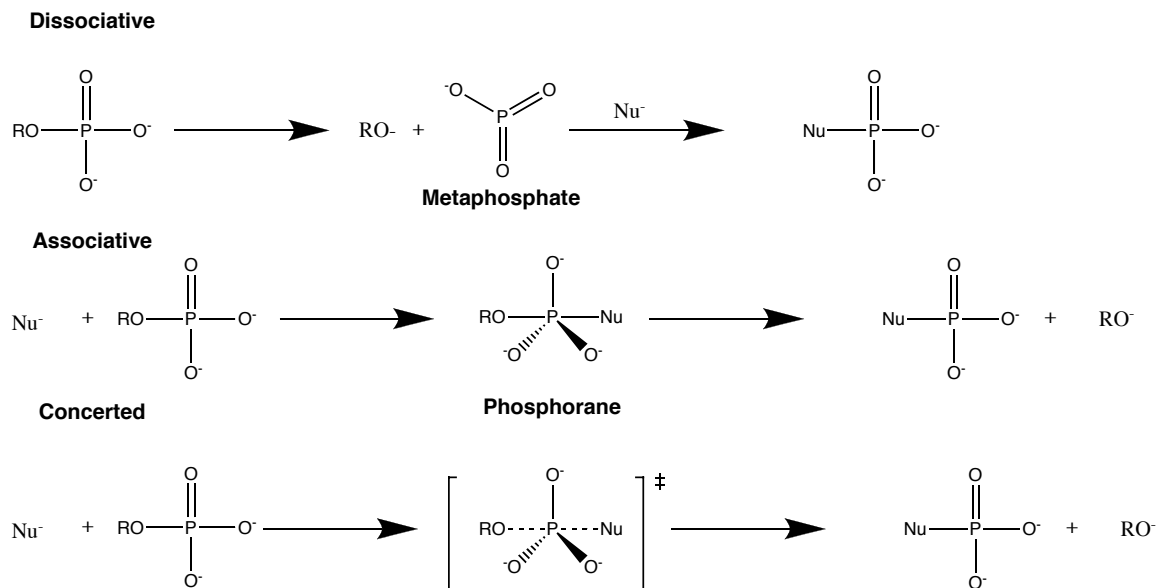


Figure 1.2. Phosphatase reactions.

The dissociative mechanism is shown proceeding through a metaphosphate intermediate, while the associative mechanism proceeds via the phosphorane intermediate. The concerted mechanism lies between the two limiting cases of dissociative and associative.

Although the enzymatic hydrolysis of phosphate monoesters looks to be a well-defined and simple chemical reaction, phosphatases are, in fact, a very diverse group of enzymes. The classification scheme of Taylor and Widlansky for phosphatases is based on substrate preference and leads to 3 general categories¹² Figure 1.3. The first category is the nonspecific phosphatases. These enzymes have the ability to catalyze a variety of phosphorylated substrates. Members of this class are further divided based on the pH at which hydrolysis occurs and have been aptly classified the alkaline and acid phosphatases (ACPs). The second category is the phosphoprotein-specific phosphatases, which cleave phosphorylated proteins or peptides as substrates. Members of this class include dual specificity phosphatases, phosphotyrosine phosphatases, phosphoserine phosphatases, and phospho-Ser/Thr phosphatases. The third category of phosphatase is the small molecule specific phosphatases, and these include enzymes that hydrolyze either one or many structurally similar substrates. This class includes phosphatases that hydrolyze substrates such as glucose-6-phosphate, inositol monophosphate, and fructose 1,6-bisphosphate just to name a few. The work that will be presented in the following chapters are ACP related; therefore, a more in-depth introduction to the classification scheme of these enzymes will be presented.

ACPs can be divided based on a variety of characteristics, such as molecular weight, substrate specificity, optimal pH of hydrolysis, and nucleophile. My proposed classification scheme has been divided into five categories: high molecular weight ACPs, low molecular weight ACPs, purple

ACPs, bacterial non-specific acid phosphatases (NSAP), and AcpA-like ACPs (Figure 1.3).

```

graph TD
    NS[Non-specific] --> AS[Alkaline Serine]
    NS --> AC[Acid]
    AS --> ACPA[AcpA-Like Serine]
    AS --> HMW[High MW Histidine]
    AC --> P2M[Purple 2-metal-ion]
    AC --> BNSAP[Bacterial NSAP]
    AC --> LMC[Low MW Cysteine]
    AC --> VII1[VII-1 like]
    VII1 --> SOL[Soluble]
    VII1 --> TM[Transmembrane]
    SOL --> CA[Class A Histidine]
    SOL --> CB[Class B Aspartate]
    SOL --> CC[Class C Aspartate]
    TM --> 2A[2A]
    TM --> 2B[2B]
    TM --> 2C[2C]
    VII1 --> DSC[Dual Specificity Cysteine]
    VII1 --> PTC[Phosphotyrosine Cysteine]
    VII1 --> PST[Phospho-Ser/Thr 2-metal-ion]
    DSC --> T1[Type 1]
    DSC --> T2[Type 2]
    PTC --> T1
    PST --> T1
    T1 --> SMS[Small Molecule Specific]
    T1 --> TS[Typical Substrates]
    SMS --> G6P[Glucose-6-phosphate]
    SMS --> F16BP[Fructose-1,6-bisphosphate]
    SMS --> IMP[Inositol monophosphate]
    SMS --> IP[Inositol polyphosphate]
    SMS --> PA[Phosphatidic acid]
    TS --> G6P
    TS --> F16BP
    TS --> IMP
    TS --> IP
    TS --> PA
  
```

The initial classification is based on substrate specificity. The secondary classifications are based on mechanism. The low molecular weight phosphatases have been found to have properties of both specific and non-specific phosphatases. When applicable the nucleophile is noted below each class. The black boxes indicate the phosphatases that have been studied as part of this thesis.

High molecular weight ACPs have been isolated and characterized from a variety of plant, bacterial, and mammalian sources. These phosphatases have broad substrate specificity and can hydrolyze alkyl-phosphates, di- and triphosphates, as well as other compounds with phosphorylated groups^{5, 13-15}. The signature sequence for this class of phosphatase is RHGXRXP that is generally located near the N-terminus. The nucleophile in this ACP class is the histidine of the signature sequence; this group of ACPs has been redefined as histidine ACPs¹⁶. The most well studied member of this class of phosphatase is the human prostatic ACP¹⁶⁻¹⁹.

Members of the low molecular weight ACP class are known to be cytoplasmic phosphotyrosine phosphatases also known as low M_r PTPases. These enzymes usually have a molecular weight less than 20 kDa and sometimes display both phosphotyrosine and phosphoserine/phosphothreonine phosphatase activity. This unique feature allows these enzymes to be considered dual specificity PTPases²⁰. These enzymes have the signature sequence motif CXXXXXRS/T, with the cysteine residue as the nucleophile. The low molecular weight ACPs are inhibited by orthovanadate, but not inhibited by tartrate.

Purple ACPs contain a dinuclear metal center, either Fe(III)-Fe(II) as in the active form of mammalian enzymes, or Fe(III)-Zn(II) as found in plant enzymes. The distinctive purple color of these enzymes result from the tyrosinate to Fe(III) charge transfer transition. Kinetic and structural data suggests a unique mechanism in which the phosphate moiety of the substrate binds directly to M(II) and hydrolysis is achieved by nucleophilic attack of a hydroxide ion

bound to the Fe(III)²¹⁻²⁴. In this reaction a phosphoenzyme intermediate is not formed as discussed above resulting in inversion of configuration at the P atom²⁵.

NSAPs are secreted enzymes that serve as soluble periplasmic proteins or as outer membrane-bound lipoproteins. Rossolini *et al* have categorized NSAPs in classes A, B, and C based on short conserved sequence motifs²⁶. The class A enzymes (25-27 kDa) contain the short signature sequence GSYP**SGH**, including the histidine nucleophile, and lack a metal center²⁷. They are not inhibited by inorganic phosphate, fluoride, EDTA, and tartrate. The X-ray crystal structure of a class A NSAP from *Escherichia blattae* has been solved and has been used as the model for this class²⁸.

Class B NSAPS include a group of secreted bacterial phosphohydrolases that are approximately 25 kDa in size. Class B NSAP share very little, if any, sequence homology to class A NSAPs. The class B NSAP signature sequence is divided into two components, an N-terminal motif of **FD**IDDTVLFSSP and a C-terminal motif of YGD-[A/S]-DXD-[I/V]^{26, 29}. These ACPs have an aspartate nucleophile located in the N-terminal region of the signature sequence (shown in bold), and are members of the haloacid dehalogenase (HAD) superfamily. The X-ray crystal structure of AphA from *Escherichia coli* has been solved and is the best characterized representative of the class B family of NSAPs³⁰.

The class C NSAPs are membrane anchored lipoproteins with an average mass of 28 kDa and are also resistant to inorganic phosphate and tartrate. These enzymes are distinguished by the bipartite sequence motif [IV]-[VAL]-D-[IL]-DET-[VM]-LX-[NT]-X(2)-Y in the N-terminus and [IV]-[LM]-X(2)-GD-[NT]-LXDF in the C-terminus²⁶. Class C NSAPs are similar to class B NSAPs in that

they have an aspartate nucleophile located in the N-terminal signature sequence motif and are members of the DDDD and HAD superfamilies. Genes encoding class C NSAPs have been identified in *Chryseobacterium meningosepticum* (*OlpA*)³¹, *Streptococcus equisimilis* (*LppC*)³², *Haemophilus influenzae* (*hel*)³³, *Helicobacter pylori* (HP1285)³⁴, *Staphylococcus aureus*³⁵, *Francisella tularensis* (unpublished data), *Pasteurella multocida*³⁶, *Bacillus anthracis*³⁷, and *Mycoplasma bovis* (unpublished data). Class C enzymes from the first four sources have been purified and characterized, but three-dimensional structures had not been reported prior to work described in this thesis.

There are some ACPs that do not fit into any of the previously described phosphatase categories. AcpA from *F. tularensis* is one such example. AcpA is structurally similar to human alkaline phosphatase (AlkP), has a serine nucleophile, and is a mononuclear metallo-enzyme. Thus, AcpA is the prototype member of the AcpA-like ACPs³⁸.

1.2 *Francisella tularensis*

Francisella tularensis is a facultative intracellular pathogen and etiological agent of tularemia, a serious and potentially life-threatening illness. The organism grows readily in broth culture, and can be isolated from numerous mammalian hosts and arthropod vectors³⁹. At least one strain of *F. tularensis* has been isolated on every continent except Antarctica. However the majority of incidences take place in the northern hemisphere⁴⁰. In the United States for instance, the Midwest has the highest incidence of tularemia particularly

Arkansas and Missouri. Figure 1.4 shows reported outbreaks of tularemia between 1990 and 2000. This pathogen is highly infectious requiring less than ten organisms to establish infection. The types of infection that occur are ulceroglandular, pneumonic, typhoidal, ocularglandular, and glandular with the first ulceroglandular being the most common. Infection/colonization can take place via various routes including ingestion, inoculation, contamination, or inhalation. Inhalation tularemia is by far the most dangerous form of tularemia, and has a case fatality rate of up to 30% if left untreated³⁹. In the light of events of recent years *F. tularensis* has become a great concern in terms of bioterrorism.

F. tularensis is considered to be a potent bioterrorism threat because of its pathogenicity, presence in a multitude of host species, and ease of isolation, growth, and dissemination. The World Health Organization estimates that an aerosol dispersal of 50 kg of virulent *F. tularensis* over a metropolitan area of 5 million inhabitants would result in 250,000 incapacitating casualties, including 19,000 deaths⁴¹. It is estimated that a third of the population would flee the city and 1 million people would require preventative antibiotic treatment for a minimum of ten days. There could be subsequent outbreaks of the disease in humans from enzootic reservoirs established in the initial attack³⁹. The Center for Disease Control and Prevention (CDC) has estimated that the total base cost to society of an *F. tularensis* attack could reach nearly 6 billion dollars for every 100,000 persons exposed⁴². For this reason the CDC has designated *F. tularensis* to be a Category A Priority Pathogen, a class that includes *B. anthracis* (anthrax), *Yersenia Pestis* (plague), *Clostridium Botulinum* (botulism), *Variola major* (smallpox), as well as viruses that cause hemorrhagic fevers.

Figure 1.4.

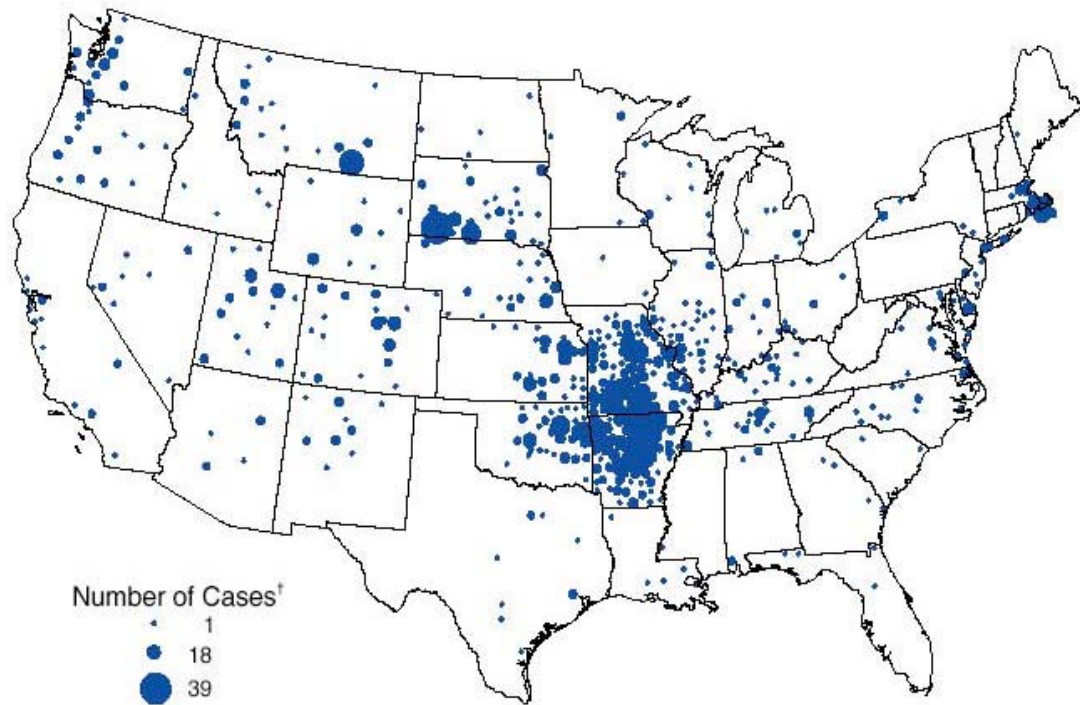


Figure 1.4. Reported cases of tularemia in the US 1990-2000.

Arkansas and Missouri have consistently recorded the highest number of tularemia cases in the last decade of the 20th century. This is partially due to the climate of the Ozark Plateau, which yields excellent conditions for a *F. tularensis* reservoir. This figure was taken from the United States Center for Disease Control and Prevention's website. www.cdc.gov.

While currently available antimicrobials are effective in treating tularemia, the organism can be engineered to carry numerous resistance genes whose presence would impede treatment if such a strain were employed in attack. An attenuated live vaccine strain (LVS) is available, but its success in preventing human disease is variable and little is known about the vaccine's efficacy against the aerosolized agent⁴³. The United States Food and Drug Administration has rejected applications to license the LVS vaccine due to the possibility that it could revert to full or partial virulence³⁹. Current antimicrobial treatments include the use of doxycycline or ciprofloxacin which are taken orally and streptomycin or gentamicin which require intravenous injection⁴⁴. These treatments are curative if administered promptly after infection.

1.3 Virulence mechanism of *F. tularensis*

Very little is known about which proteins are involved in virulence of highly virulent strains of *F. tularensis*. Several candidate virulence proteins have been proposed based on proteomics, molecular genetics, and cell biological studies. The ability to replicate within the host is the trademark of *F. tularensis*^{45, 46}, and this ability is essential to induce macrophage apoptosis⁴⁷. Much of the research to this point to identify virulence proteins has focused on intracellular survival mechanisms. Research from the past few years has demonstrated that the *F. tularensis* pathogenicity island (FPI) contains many of the virulence proteins associated with infection, including IgIA-D and PdpA-D^{48, 49}. The FPI proteins IgIA, IgIB, which are encoded by the *iglABCD* operon, have some

homology to IcmF-associated homologous proteins (IAHPs) found in *Vibrio cholerae*, *Salmonella enterica*, *Rhizobium leguminosarum*, and other bacteria⁵⁰⁻⁵². The genes that encode these homologous proteins are clustered together and are thought to be components of a proposed type VI secretion system in *V. cholerae*⁵³. The MglA and MglB proteins are encoded by an operon located outside the FPI and share homology with SspA and SspB of *E. coli* respectively. The genes *iglA*, *iglC*, *pdpA* and *pdpD* have been shown to be regulated by the MglA protein^{46, 54, 55}. Many of these proteins have limited sequence homology to proteins in Genbank and therefore biological function is unknown. However much work has been done with IglC and MglA in terms of understanding their functions.

Figure 1.5

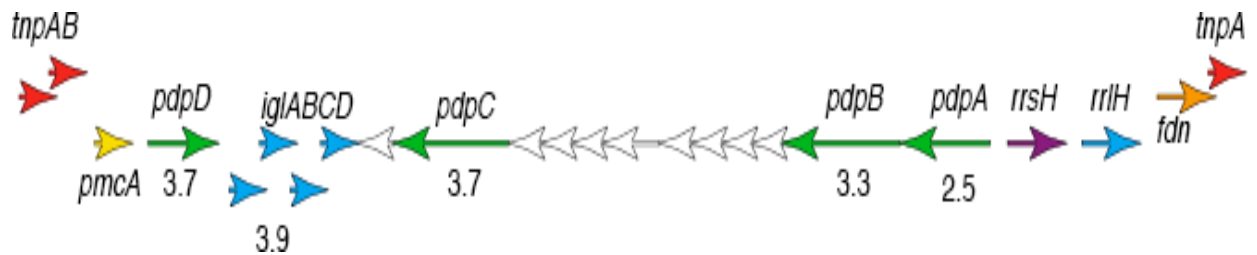


Figure 1.5. The *Francisella tularensis* Pathogenicity Island

The FPI is 30 kb in size and has 17 ORFs. The *iglABCD* (blue) genes seem to be organized in an operon. The deduced IgIC and IgID proteins have no significant homologs whereas IgIA and IgIB are ~30% identical to hypothetical proteins of plant pathogens and endosymbionts. None of the deduced *pdpABC* (green) genes or the eight short ORFs (white; 800–1431 bp) between *pdpB* and *pdpC* has significant homologs. The gene encoding the putative PmcA protein (yellow) is 44% identical to conserved domains found in putative chaperones, such as VdcA of *Bacillus anthracis*. The GC content of the *Francisella tularensis* genome is 33.2% but the GC content of different regions of the FPI varies, ranging from 26.6% (*pdpC* through *pdpA*) to 31.0% (*pdpD* through *IgID*). Upstream of *pdpA* is a 5-kb rRNA-encoding region (purple; light blue) with a GC content of 51%. The *tnpAB* and *tnpA* genes (red) on both ends of the FPI are presumed to encode transposases. There is one copy of FPI in the *F. tularensis* subspecies *novicida* genome but there are two copies in *F. tularensis* subspecies *tularensis* and *F. tularensis* subspecies *holarctica*-derived LVS. *pdpD* is deleted from both copies of the FPI in the LVS and other *F. tularensis* subspecies *holarctica*-derived

clinical isolates. Numbers beneath the arrows indicate the gene size in kb.

This figure and caption was taken from reference 49⁴⁹. Used with permission.

IgIC is a 23kDa cytosolic protein that has been implicated in *F. tularensis* virulence. When compared to the non-virulent LVS IgIC has been shown to be preferentially expressed in the more virulent strains⁵⁶. IgIC is also over-expressed in *F. tularensis* grown in macrophages compared to bacteria grown extracellularly, which suggests that IgIC could be important for survival within the macrophage⁵⁵. Recent studies by Lai *et al* and Lauriano *et al* have shown that virulent strains lacking the IgIC gene are unable to grow and survive within the macrophage^{47, 54}. When the sequence of IgIC is compared to other proteins in Genbank no suitable homologs are returned, therefore little if any functional comparisons can be accurately made. Speculation of IgIC's function is that it has a role in down-regulation of Toll-like Receptor-4-dependant signaling in infected macrophages⁵⁷, but other important details of the interaction of IgIC with the immune system are unavailable.

Macrophage growth locus A (MglA) controls transcription of FPI genes implicated in virulence including IgIC⁴⁸. A *F. tularensis* mutant deficient in MglA expression showed decreased levels of *igIC* transcripts, and this mutant was unable to survive and replicate within macrophages⁵⁴. Such a result suggests that the genes controlled by MglA are critically important for the ability of the pathogen to survive within macrophages and therefore its overall virulence. When the sequence of MglA is compared to all sequences in Genbank the top result is stringent starvation protein A (SspA) from *E. coli* with a sequence identity of 25%. In *E. coli*, SspA interacts with the RNA polymerase holoenzyme and has a role in the expression of at least 11 proteins⁵⁸. SspA homologues have been shown to be involved in virulence for other bacterial pathogens that include

*Neisseria gonorrhoea*⁵⁹, *Yersenia enterocolitica*⁶⁰, and *Providencia stuartii*⁶¹. The crystal structure of SspA from *Y. pestis* has been solved to 2.0 Å resolution⁶²,⁶³. This structure has a fold that is characteristic of glutathione *S*-transferase (GST) proteins, however SspA does not have GST activity. The SspA protein of *Y. pestis* shares high sequence identity with that of *E. coli*, therefore, the authors modeled SspA of *E. coli* from the *Y. pestis* structure. *E. coli* SspA is known to be important for transcriptional activation of the phage P1 late promoter and stationary phase-induced acid tolerance. This structure and functional analysis could help in the elucidation of the functional role of MglA.

Although not encoded by the FPI, AcpA has a role in virulence. Acid phosphatase A (AcpA) is the most well characterized protein from *F. tularensis*. AcpA is a highly expressed 57 kDa polyspecific periplasmic ACP⁶⁴ that hydrolyzes a variety of substrates, including *p*-nitrophenylphosphate (*p*NPP), *p*-nitrophenylphosphorylcholine (*p*NPPC), peptides containing phosphotyrosine, inositol phosphates, AMP, ATP, fructose 1,6-bisphosphate, glucose and fructose 6-phosphates, NADP⁺, and ribose 5-phosphate⁶⁴,⁶⁵. The enzyme is inhibited by the metal oxyanions orthovanadate, molybdate, and tungstate. Based on amino acid sequence analysis, AcpA is distinct from histidine ACPs¹⁷, purple ACPs¹, as well as class A, B and C NSAPs²⁹.

Purified AcpA inhibits the respiratory burst of stimulated neutrophils, which suggests that AcpA helps the pathogen elude the host oxidative defense system during the initial stages of macrophage infection⁶⁴. Furthermore, proteomics studies have shown that AcpA is expressed at a higher level in virulent *F. tularensis* strains compared to the non-virulent vaccine strain⁵⁶. Most

recently, it has been shown that a mutant strain of *F. tularensis* subspecies *novicida* lacking a functional *acpA* gene is less virulent in mice than the wild-type strain due to a defect in phagosomal escape⁶⁶. Thus, AcpA appears to be important for survival of the microbe at two critical junctures of infection: colonization and intracellular survival.

Amino acid sequence alignments show that AcpA belongs to a superfamily of bacterial enzymes that includes ACPs and phospholipases C (PLCs) from a variety of microbial pathogens including *Pseudomonas aeruginosa*, *Mycobacterium tuberculosis*, *Bordetella pertussis* and several *Burkholderia* species⁶⁷. AcpA is the only characterized enzyme from the ACP branch of the superfamily. PLCs from this superfamily are important virulence factors in *P. aeruginosa*⁶⁸ and *M. tuberculosis*⁶⁹ infections, with the hemolytic PLC from *P. aeruginosa* (PlcH) being the best characterized example from the PLC branch of the superfamily⁷⁰. PlcH is particularly interesting because it is a multifunctional enzyme that displays sphingomyelin synthase activity in addition to PLC activity⁷¹. PLCs of the AcpA/PlcH superfamily share no sequence homology with the well-studied zinc metallophospholipases *Clostridium perfringens* alpha-toxin and *Bacillus cereus* phosphatidylcholine-preferring PLC⁷², which suggests that PLCs of the AcpA/PlcH superfamily have a novel, and as yet uncharacterized, catalytic mechanism.

As part of this research four other phosphatases from *F. tularensis* have been identified: histidine ACP (HAP), class C NSAP, class B NSAP, and protein tyrosine phosphatase. A sixth phosphatases, lipid A 4'-phosphatase, has also been identified by Wang *et al*⁷³. We set out to determine the structures of AcpA,

HAP and the class C NSAP. Parallel structure determination experiments were done using class C NSAPs from *F. tularensis*, *B. anthracis* and *H. influenzae* in order to develop a basis for understanding the class C NSAP family. In the following chapters a detailed description of the crystallization and structure determination of AcpA from *F. tularensis* will be presented, as well as the structures of *F. tularensis* HAP, and two class C acid phosphatases from *H. influenzae* and *B. anthracis*.

1.4 References

1. 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. Coleman, J. E., Structure and mechanism of alkaline phosphatase. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, 21, 441-83.
3. Zhang, Z. Y., Chemical and mechanistic approaches to the study of protein tyrosine phosphatases. *Acc. Chem. Res.* **2003**, 36, (6), 385-92.
4. Buchwald, S. L.; Saini, M. S.; Knowles, J. R.; Van Etten, R. L., Stereochemical course of phospho group transfer by human prostatic acid phosphatase. *J. Biol. Chem.* **1984**, 259, (4), 2208-13.
5. Saini, M. S.; Buchwald, S. L.; Van Etten, R. L.; Knowles, J. R., Stereochemistry of phospho transfer catalyzed by bovine liver acid phosphatase. *J. Biol. Chem.* **1981**, 256, (20), 10453-5.
6. Vincent, J. B.; Crowder, M. W.; Averill, B. A., Evidence for a phosphoryl-enzyme intermediate in phosphate ester hydrolysis by purple acid phosphatase from bovine spleen. *J. Biol. Chem.* **1991**, 266, (27), 17737-40.
7. Denu, J. M.; Lohse, D. L.; Vijayalakshmi, J.; Saper, M. A.; Dixon, J. E., Visualization of intermediate and transition-state structures in protein-tyrosine phosphatase catalysis. *Proc Natl Acad Sci U S A* **1996**, 93, (6), 2493-8.
8. Hengge, A. C., *Phosphoryl Transfer in Solution*. Academic Press: San Diego, CA, 1998; Vol. 1, p 517-542.
9. O'Brien, P. J.; Herschlag, D., Alkaline phosphatase revisited: hydrolysis of alkyl phosphates. *Biochemistry* **2002**, 41, (9), 3207-25.
10. Fauman, E. B.; Yuvaniyama, C.; Schubert, H. L.; Stuckey, J. A.; Saper, M. A., The X-ray crystal structures of *Yersinia* tyrosine phosphatase with bound tungstate and nitrate. Mechanistic implications. *J Biol Chem* **1996**, 271, (31), 18780-8.

11. McCain, D. F.; Catrina, I. E.; Hengge, A. C.; Zhang, Z. Y., The catalytic mechanism of Cdc25A phosphatase. *J. Biol. Chem.* **2002**, 277, (13), 11190-200.
12. Taylor, W. P.; Widlanski, T. S., Charged with meaning: the structure and mechanism of phosphoprotein phosphatases. *Chem Biol* **1995**, 2, (11), 713-8.
13. Kilsheimer, G. S.; Axelrod, B., Inhibition of prostatic acid phosphatase by alpha-hydroxycarboxylic acids. *J Biol Chem* **1957**, 227, (2), 879-90.
14. Ostanin, K.; Harms, E. H.; Stevis, P. E.; Kuciel, R.; Zhou, M. M.; Van Etten, R. L., Overexpression, site-directed mutagenesis, and mechanism of *Escherichia coli* acid phosphatase. *J Biol Chem* **1992**, 267, (32), 22830-6.
15. 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. Van Etten, R. L., Human prostatic acid phosphatase: a histidine phosphatase. *Ann N Y Acad Sci* **1982**, 390, 27-51.
17. 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.
18. Van Etten, R. L.; McTigue, J. J., pH dependence and solvent isotope effects in the hydrolysis of phosphomonoesters by human prostatic acid phosphatase. *Biochim Biophys Acta* **1977**, 484, (2), 386-97.
19. 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.
20. Ramponi, G.; Stefani, M., Structural, catalytic, and functional properties of low M(r), phosphotyrosine protein phosphatases. Evidence of a long evolutionary history. *Int J Biochem Cell Biol* **1997**, 29, (2), 279-92.

21. Uppenberg, J.; Lindqvist, F.; Svensson, C.; Ek-Rylander, B.; Andersson, G., Crystal structure of a mammalian purple acid phosphatase. *J. Mol. Biol.* **1999**, 290, (1), 201-11.
22. Strater, N.; Klabunde, T.; Tucker, P.; Witzel, H.; Krebs, B., Crystal structure of a purple acid phosphatase containing a dinuclear Fe(III)-Zn(II) active site. *Science* **1995**, 268, (5216), 1489-92.
23. Klabunde, T.; Strater, N.; Frohlich, R.; Witzel, H.; Krebs, B., Mechanism of Fe(III)-Zn(II) purple acid phosphatase based on crystal structures. *J. Mol. Biol.* **1996**, 259, (4), 737-48.
24. Pinkse, M. W.; Merkx, M.; Averill, B. A., Fluoride inhibition of bovine spleen purple acid phosphatase: characterization of a ternary enzyme-phosphate-fluoride complex as a model for the active enzyme-substrate-hydroxide complex. *Biochemistry* **1999**, 38, (31), 9926-36.
25. Mueller, E. G.; Crowder, M. W.; Averil, B. A.; Knowles, J. R., Purple Acid Phosphatase: A Diiron Enzyme that Catalyzes a Direct Phospho Group Transfer to Water. *J. Am. Chem. Soc.* **1993**, 115, 2914-2915.
26. 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.
27. Stuke, J.; Carman, G. M., Identification of a novel phosphatase sequence motif. *Protein Sci* **1997**, 6, (2), 469-72.
28. Ishikawa, K.; Mihara, Y.; Gondoh, K.; Suzuki, E.; Asano, Y., X-ray structures of a novel acid phosphatase from *Escherichia blattae* and its complex with the transition-state analog molybdate. *Embo J* **2000**, 19, (11), 2412-23.
29. 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.
30. Calderone, V.; Forleo, C.; Benvenuti, M.; Cristina Thaller, M.; Maria Rossolini, G.; Mangani, S., The first structure of a bacterial class B Acid

phosphatase reveals further structural heterogeneity among phosphatases of the haloacid dehalogenase fold. *J. Mol. Biol.* **2004**, 335, (3), 761-73.

31. 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.
32. Malke, H., Cytoplasmic membrane lipoprotein LppC of *Streptococcus equisimilis* functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998**, 64, (7), 2439-42.
33. Green, B. A.; Farley, J. E.; Quinn-Dey, T.; Deich, R. A.; Zlotnick, G. W., The e (P4) outer membrane protein of *Haemophilus influenzae*: biologic activity of anti-e serum and cloning and sequencing of the structural gene. *Infect. Immun.* **1991**, 59, (9), 3191-3198.
34. Godlewska, R.; Bujnicki, J. M.; Ostrowski, J.; Jagusztyn-Krynica, E. K., The hppA gene of *Helicobacter pylori* encodes the class C acid phosphatase precursor. *FEBS Lett.* **2002**, 525, (1-3), 39-42.
35. 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.
36. 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. U S A* **2001**, 98, (6), 3460-5.
37. Read, T. D.; Peterson, S. N.; Tourasse, N.; Baillie, L. W.; Paulsen, I. T.; Nelson, K. E.; Tettelin, H.; Fouts, D. E.; Eisen, J. A.; Gill, S. R.; Holtzapple, E. K.; Okstad, O. A.; Helgason, E.; Rillstone, J.; Wu, M.; Kolonay, J. F.; Beanan, M. J.; Dodson, R. J.; Brinkac, L. M.; Gwinn, M.; DeBoy, R. T.; Madpu, R.; Daugherty, S. C.; Durkin, A. S.; Haft, D. H.; Nelson, W. C.; Peterson, J. D.; Pop, M.; Khouri, H. M.; Radune, D.; Benton, J. L.; Mahamoud, Y.; Jiang, L.; Hance, I. R.; Weidman, J. F.; Berry, K. J.; Plaut, R. D.; Wolf, A. M.; Watkins, K. L.; Nierman, W. C.; Hazen, A.; Cline, R.; Redmond, C.; Thwaite, J. E.; White, O.; Salzberg, S. L.; Thomason, B.; Friedlander, A. M.; Koehler, T. M.; Hanna, P. C.; Kolsto, A. B.; Fraser, C. M., The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **2003**, 423, (6935), 81-6.

38. 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.
39. Oyston, P. C.; Sjostedt, A.; Titball, R. W., Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nat. Rev. Microbiol.* **2004**, 2, (12), 967-78.
40. Titball, R. W.; Johansson, A.; Forsman, M., Will the enigma of *Francisella tularensis* virulence soon be solved? *Trends Microbiol.* **2003**, 11, (3), 118-23.
41. *Health aspects of chemical and biological weapons*; World Health Organization: 1970; pp 105-107.
42. Kaufmann, A. F.; Meltzer, M. I.; Schmid, G. P., The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? *Emerg. Infect. Dis.* **1997**, 3, (2), 83-94.
43. Elkins, K. L.; Leiby, D. A.; Winegar, R. K.; Nacy, C. A.; Fortier, A. H., Rapid generation of specific protective immunity to *Francisella tularensis*. *Infect. Immun.* **1992**, 60, (11), 4571-7.
44. Ellis, J.; Oyston, P. C.; Green, M.; Titball, R. W., Tularemia. *Clin. Microbiol. Rev.* **2002**, 15, (4), 631-46.
45. Fortier, A. H.; Green, S. J.; Polsinelli, T.; Jones, T. R.; Crawford, R. M.; Leiby, D. A.; Elkins, K. L.; Meltzer, M. S.; Nacy, C. A., Life and death of an intracellular pathogen: *Francisella tularensis* and the macrophage. *Immunol. Ser.* **1994**, 60, 349-61.
46. Sjostedt, A., Virulence determinants and protective antigens of *Francisella tularensis*. *Curr. Opin. Microbiol.* **2003**, 6, (1), 66-71.
47. Lai, X. H.; Golovliov, I.; Sjostedt, A., Expression of IgIC is necessary for intracellular growth and induction of apoptosis in murine macrophages by *Francisella tularensis*. *Microb. Pathog.* **2004**, 37, (5), 225-30.
48. Nano, F. E.; Zhang, N.; Cowley, S. C.; Klose, K. E.; Cheung, K. K.; Roberts, M. J.; Ludu, J. S.; Letendre, G. W.; Meierovics, A. I.; Stephens, G.; Elkins, K. L., A

Francisella tularensis pathogenicity island required for intramacrophage growth. *J. Bacteriol.* **2004**, 186, (19), 6430-6.

49. Santic, M.; Molmeret, M.; Klose, K. E.; Abu Kwaik, Y., *Francisella tularensis* travels a novel, twisted road within macrophages. *Trends Microbiol.* **2006**, 14, (1), 37-44.

50. Sexton, J. A.; Miller, J. L.; Yoneda, A.; Kehl-Fie, T. E.; Vogel, J. P., *Legionella pneumophila* DotU and IcmF are required for stability of the Dot/Icm complex. *Infect Immun* **2004**, 72, (10), 5983-92.

51. Das, S.; Chaudhuri, K., Identification of a unique IAHP (IcmF associated homologous proteins) cluster in *Vibrio cholerae* and other proteobacteria through in silico analysis. *In Silico Biol* **2003**, 3, (3), 287-300.

52. Folkesson, A.; Lofdahl, S.; Normark, S., The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. *Res Microbiol* **2002**, 153, (8), 537-45.

53. Pukatzki, S.; Ma, A. T.; Sturtevant, D.; Krastins, B.; Sarracino, D.; Nelson, W. C.; Heidelberg, J. F.; Mekalanos, J. J., Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl. Acad. Sci. USA* **2006**, 103, (5), 1528-33.

54. Lauriano, C. M.; Barker, J. R.; Yoon, S. S.; Nano, F. E.; Arulanandam, B. P.; Hassett, D. J.; Klose, K. E., MglA regulates transcription of virulence factors necessary for *Francisella tularensis* intraamoebae and intramacrophage survival. *Proc. Natl. Acad. Sci. USA* **2004**, 101, (12), 4246-9.

55. Golovliov, I.; Ericsson, M.; Sandstrom, G.; Tarnvik, A.; Sjostedt, A., Identification of proteins of *Francisella tularensis* induced during growth in macrophages and cloning of the gene encoding a prominently induced 23-kilodalton protein. *Infect. Immun.* **1997**, 65, (6), 2183-9.

56. Hernychova, L.; Stulik, J.; Halada, P.; Macela, A.; Kroca, M.; Johansson, T.; Malina, M., Construction of a *Francisella tularensis* two-dimensional electrophoresis protein database. *Proteomics* **2001**, 1, (4), 508-15.

57. Telepnev, M.; Golovliov, I.; Grundstrom, T.; Tarnvik, A.; Sjostedt, A., *Francisella tularensis* inhibits Toll-like receptor-mediated activation of intracellular signalling and secretion of TNF-alpha and IL-1 from murine macrophages. *Cell Microbiol.* **2003**, 5, (1), 41-51.
58. Hansen, A. M.; Lehnherr, H.; Wang, X.; Mobley, V.; Jin, D. J., *Escherichia coli* SspA is a transcription activator for bacteriophage P1 late genes. *Mol Microbiol* **2003**, 48, (6), 1621-31.
59. De Reuse, H.; Taha, M. K., RegF, an SspA homologue, regulates the expression of the *Neisseria gonorrhoeae* pilE gene. *Res. Microbiol.* **1997**, 148, (4), 289-303.
60. Badger, J. L.; Miller, V. L., Expression of invasin and motility are coordinately regulated in *Yersinia enterocolitica*. *J. Bacteriol.* **1998**, 180, (4), 793-800.
61. Ding, X.; Baca-DeLancey, R. R.; Rather, P. N., Role of SspA in the density-dependent expression of the transcriptional activator AarP in *Providencia stuartii*. *FEMS Microbiol. Lett.* **2001**, 196, (1), 25-9.
62. Andrykovitch, M.; Routzahn, K. M.; Li, M.; Gu, Y.; Waugh, D. S.; Ji, X., Characterization of four orthologs of stringent starvation protein A. *Acta Crystallogr.* **2003**, D59, (Pt 5), 881-6.
63. Hansen, A. M.; Gu, Y.; Li, M.; Andrykovitch, M.; Waugh, D. S.; Jin, D. J.; Ji, X., Structural basis for the function of stringent starvation protein a as a transcription factor. *J. Biol. Chem.* **2005**, 280, (17), 17380-91.
64. 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.
65. 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.

66. 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.
67. Stonehouse, M. J.; Cota-Gomez, A.; Parker, S. K.; Martin, W. E.; Hankin, J. A.; Murphy, R. C.; Chen, W.; Lim, K. B.; Hackett, M.; Vasil, A. I.; Vasil, M. L., A novel class of microbial phosphocholine-specific phospholipases C. *Mol. Microbiol.* **2002**, 46, (3), 661-76.
68. Ostroff, R. M.; Wretling, B.; Vasil, M. L., Mutations in the hemolytic-phospholipase C operon result in decreased virulence of *Pseudomonas aeruginosa* PAO1 grown under phosphate-limiting conditions. *Infect. Immun.* **1989**, 57, (5), 1369-73.
69. Raynaud, C.; Guilhot, C.; Rauzier, J.; Bordat, Y.; Pelicic, V.; Manganeli, R.; Smith, I.; Gicquel, B.; Jackson, M., Phospholipases C are involved in the virulence of *Mycobacterium tuberculosis*. *Mol. Microbiol.* **2002**, 45, (1), 203-17.
70. Ostroff, R. M.; Vasil, A. I.; Vasil, M. L., Molecular comparison of a nonhemolytic and a hemolytic phospholipase C from *Pseudomonas aeruginosa*. *J. Bacteriol.* **1990**, 172, (10), 5915-23.
71. Luberto, C.; Stonehouse, M. J.; Collins, E. A.; Marchesini, N.; El-Bawab, S.; Vasil, A. I.; Vasil, M. L.; Hannun, Y. A., Purification, characterization, and identification of a sphingomyelin synthase from *Pseudomonas aeruginosa*. PlcH is a multifunctional enzyme. *J. Biol. Chem.* **2003**, 278, (35), 32733-43.
72. Titball, R. W., Bacterial phospholipases C. *Microbiol. Rev.* **1993**, 57, (2), 347-66.
73. Wang, X.; Ribeiro, A. A.; Guan, Z.; Abraham, S. N.; Raetz, C. R., Attenuated virulence of a *Francisella* mutant lacking the lipid A 4'-phosphatase. *Proc Natl Acad Sci U S A* **2007**, 104, (10), 4136-41.

2.

CRYSTALLIZATION OF ACPA FROM FRANCISELLA TULARENSIS

Portions of this chapter have been adapted from:

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, 107-110.

2.1 Methods and Results

2.1.1 Cloning, expression and purification

Recombinant AcpA (rAcpA) was expressed and purified using methods described elsewhere¹. The purified enzyme was dialyzed against 50 mM sodium acetate, 150 mM NaCl, pH 6.0, and concentrated to 2 - 20 mg/mL. The protein concentration was measured with the Coomassie Plus assay (Pierce). Purified rAcpA was judged to be dimeric in solution and highly monodisperse based on gel filtration chromatography, dynamic light scattering and analytical ultracentrifugation¹. Gel filtration chromatography of rAcpA under nondenaturing conditions was performed on a Superdex 200 gel filtration chromatography resin with 50 mM sodium acetate, pH 6.0, containing 150 mM NaCl and 0.05% (wt/vol) azide as the mobile phase, with a flow rate of 1 mL/min. The partition coefficient of the phosphatase was compared to those of the following gel filtration standards: alcohol dehydrogenase (150,000 Da, K_{av} = 0.310), bovine serum albumin (BSA) (66,000 Da, K_{av} = 0.394), carbonic anhydrase (29,000 Da, K_{av} = 0.487), and cytochrome c (12,400 Da, K_{av} = 0.569). Dynamic light scattering analysis was performed using a Protein Solutions DynaPro 99 Molecular Sizing Instrument. An average of 10–20 measurements were collected for each sample at 20 °C using a laser wavelength of 836.3 nm and a scattering angle of 90°. The measurements were analyzed using Dynamics Version 5.26.38 software.

Results of gel filtration and dynamic light scattering analysis are consistent with the observation of an apparent rAcpA dimer using ultracentrifugation. Gel filtration chromatography of rAcpA through Superdex 200 chromatography resin yielded a partition coefficient of 0.352 (data not shown). This value was compared to the regression line generated from standards, and corresponded to an apparent molecular mass of 100 kDa. Dynamic light scattering experiments (data not shown) performed at a concentration of 1 mg/mL indicated that the rAcpA solution was highly monodisperse with a mean hydrodynamic radius (R_h) of 4.446 nm which corresponds to a molecular mass of 108 kDa. As the monomeric molecular mass was 56 kDa, determined by mass spectrometry. Results from dynamic light scattering indicate that the prevalent form of rAcpA present in solution is a dimer. Collectively these data suggest that the recombinant enzyme is a apparent homodimer under the conditions tested.

2.1.2 Crystallization, data collection and processing

All crystallization experiments were performed at 295 K using the sitting drop method of vapor diffusion with drops formed by mixing equal volumes of the reservoir and protein solution. The Hampton Index Screen was used to identify initial crystallization conditions. Several conditions in the screen yielded crystals of various size and quantity. Polyethylene glycol (PEG) in the molecular weight range 1500 - 5000 was common to all of the positive conditions.

Optimization studies using PEGs with different molecular weights, variation of the protein concentration, and inclusion of an inhibitor in the crystallization setup resulted in the growth of three crystal forms of rAcpA (Table 3.1). Crystal form I was grown with a 2 - 4 mg/mL protein stock solution and reservoir consisting of 5.0% Tacsimate, 0.1 M HEPES pH 7.0, 10% (w/v) PEG monomethyl ether (PEGMME) 5000. These crystals displayed a coffin-shaped plate-like morphology (Figure 3.1a) and diffracted to 3.2 Å resolution when analyzed on a Cu rotating anode X-ray diffraction system equipped with an Raxis-IV image plate detector and Osmic confocal optics. Autoindexing calculations performed with HKL² indicated a primitive orthorhombic lattice with unit cell dimensions $a = 85 \text{ Å}$, $b = 296 \text{ Å}$, $c = 262 \text{ Å}$. Analysis of the solvent content using the method of Matthews³ suggested the presence of 10 - 13 rAcpA molecules per asymmetric unit and a solvent content of 44 – 57%. Clearly, form I is not optimal for high-resolution structure determination.

Crystal form II was obtained by incubating the enzyme with the competitive inhibitor sodium orthovanadate (5 mM, Na_3VO_4) prior to crystallization. Monomeric vanadate is a structural and electronic analogue of phosphate and thus it is a potent inhibitor of many phosphatases⁴, including AcpA^{1, 5}. A protein stock solution of 2 – 4 mg/mL was used and the optimal reservoir condition was 16 – 18% (w/v) PEG 1500. These crystals grew within 3 days after setup and took the form of long, rectangular solids with pointed ends (Figure 3.1b). Diffraction to 2.2 Å resolution was observed, and autoindexing suggested a primitive orthorhombic lattice with unit cell dimensions $a = 110 \text{ Å}$, b

= 126 Å, c = 144 Å. The asymmetric unit of crystal form II is expected to contain 3 – 4 rAcpA molecules and 43 - 57% solvent.

As with form II, crystal form III grew in the presence of 5 mM Na₃VO₄ and PEG 1500; however, the protein concentration was approximately 5 – 10 times higher (17 - 20 mg/mL) than that used to crystallize forms I and II (Table 3.1). The optimal reservoir condition was 19.5% (w/v) PEG 1500. Tetrahedron-shaped crystals typically grew to a maximum dimension of 0.4 mm within 36 hours (Figure 3.1c). In preparation for low temperature data collection, the crystals were cryoprotected by exchanging the mother liquor with a solution containing 25% (w/v) PEG 1500 and 25% (v/v) PEG 200. After a few minutes the crystals were picked up with Hampton mounting loops and plunged into liquid nitrogen.

Form III crystals have space group C222₁ with unit cell dimensions a = 112 Å, b = 144 Å, c = 124 Å. The asymmetric unit is predicted to contain only two rAcpA molecules, with a Matthews coefficient of 2.2 Å³/Da and a solvent content of 43%. These crystals diffracted well beyond 2 Å when analyzed using the rotating anode system.

Form III has been chosen for structure determination because of its exceptional diffraction characteristics, low mosaicity, and amenable unit cell dimensions. A 1.75 Å resolution native data set was collected at beamline 8.3.1 of the Advanced Light Source, and the data were processed with HKL2000². The data set consisted of 100 frames with an oscillation angle of 1°, exposure time of 1.2 s per frame, and detector distance of 100 mm. See Table 3.2 for data processing statistics. AcpA is sufficiently novel that suitable molecular replacement search models are not available from the PDB. Structure

determination by the method of multiple isomorphous replacement is currently in progress.

Table 2.1

Summary of three crystal forms of recombinant AcpA

	Form I	Form II	Form III
Precipitating agent	PEGMME 5000	PEG 1500	PEG 1500
Protein concentration (mg/mL)	2 - 4	2 - 4	17 - 20
Inhibitor	None	Na ₃ VO ₄	Na ₃ VO ₄
Bravais lattice	Pmmm	Pmmm	Cmmm
Unit-cell dimensions (Å)	a = 85 b = 296 c = 262	a = 110 b = 126 c = 144	a = 112 b = 144 c = 124
Diffraction limit (Å)	3.2	2.2	1.75
Molecules in asymmetric unit	10 - 13	3 - 4	2
V _M (Å ³ /Da)	2.2 – 2.9	2.2 – 2.9	2.2
Solvent content (%)	44 - 57	43 - 57	43

Table 2.2

Data-collection and processing statistics (crystal form III)

Beamline	ALS 8.3.1
Wavelength (Å)	1.127129
Space group	C2221
Unit-cell dimensions (Å)	a = 112.1, b = 144.4, c = 123.9
Diffraction resolution (Å)	50 - 1.75 (1.81 - 1.75)
No. of observations	397516
No. of unique reflections	98137
Redundancy	4.1 (4.0)
Refined mosaicity (°)	0.3
Completeness (%)	97.2 (94.9)
Average I/ σ (I)	26.6 (3.8)
R _{merge}	0.047 (0.357)

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

Figure 2.1

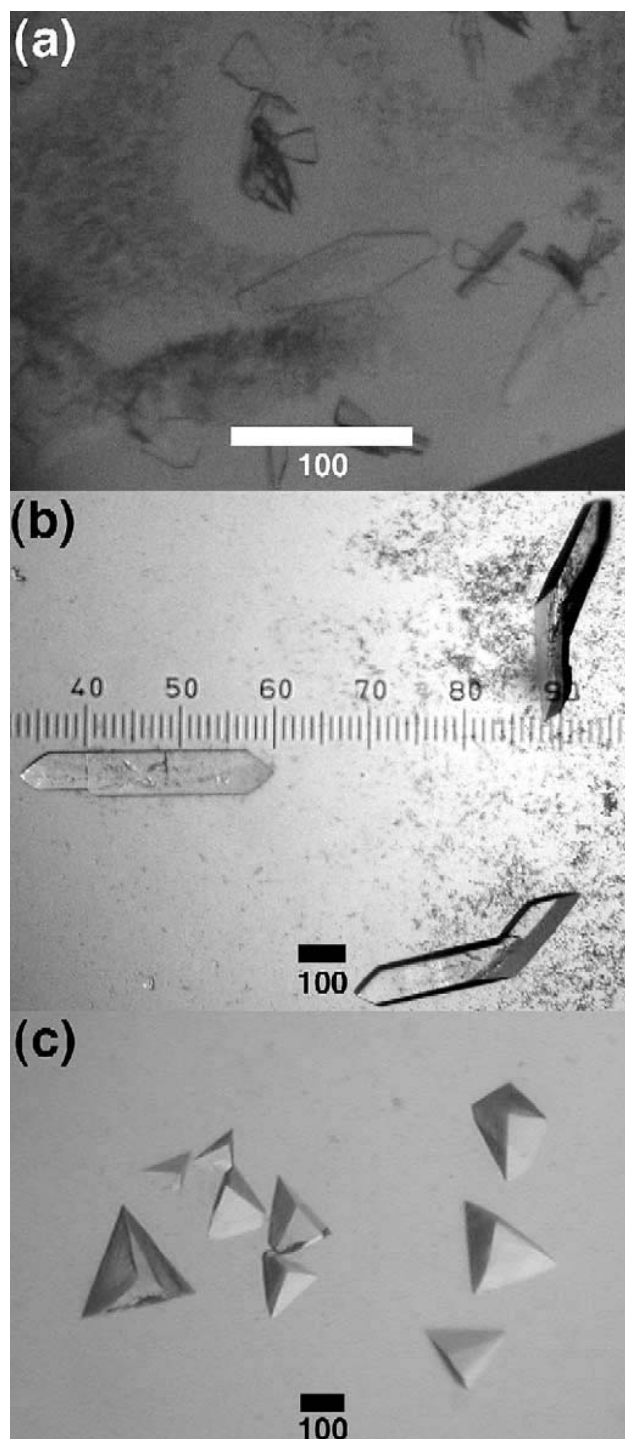


Figure 2.1 Three crystal forms of recombinant *F. tularensis* AcpA.

Forms I, II, and III are shown in panels (a), (b), and (c), respectively. The bar in each panel represents approximately 0.1 mm.

2.2 References

1. Reilly, T. J.; Felts, R. L.; Henzl, M. T.; Calcutt, M. J.; Tanner, J. J., Characterization of recombinant *Francisella tularensis* acid phosphatase. *Biochem. Biophys. Acta* **2005**, to be submitted.
2. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, 276, 307-326.
3. Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968**, 33, 491-497.
4. Crans, D. C.; Smee, J. J.; Gaidamauskas, E.; Yang, L., The chemistry and biochemistry of vanadium and the biological activities exerted by vanadium compounds. *Chem. Rev.* **2004**, 104, (2), 849-902.
5. 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.

3.

THE STRUCTURE OF ACPA

Portions of this chapter have been adapted from:

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.

3.1 Experimental Procedures

3.1.1 Crystallization and X-ray diffraction data collection

Expression of recombinant *F. tularensis* AcpA in *E. coli*, protein purification and the growth of three different crystal forms were described previously¹. Structure determination utilized crystal form III, which was obtained by incubating the enzyme with the competitive inhibitor sodium orthovanadate (Na_3VO_4 , 5 mM) prior to crystallization and using PEG 1500 as the precipitating agent¹. These crystals have space group $\text{C}222_1$ with unit cell dimensions $a = 112 \text{ \AA}$, $b = 144 \text{ \AA}$, $c = 124 \text{ \AA}$, two molecules per asymmetric unit and 43% solvent content.

The derivative used for phasing was produced by soaking an AcpA/orthovanadate crystal in 40 mM $\text{Sm}(\text{C}_2\text{H}_3\text{O}_2)_3$ for 10 minutes. Diffraction data extending to 2.4 \AA resolution were collected from the Sm derivative at Advanced Photon Source (APS) beamline 19-ID using $\lambda = 1.6531 \text{ \AA}$, which corresponds to an energy between the L-I and L-II absorption edges of Sm. Data processing was done with HKL2000 (Table 3.1). Anomalous difference Patterson maps showed several strong features on the $u = 0$ Harker section (Figure 3.1).

The data set used for phase extension and refinement calculations at 1.75 \AA resolution was collected from an AcpA/orthovanadate crystal at Advanced Light Source (ALS) beamline 8.3.1. A second data set, which was used for anomalous difference Fourier analysis of the active site metal center, was collected from another AcpA/orthovanadate crystal at beamline 8.3.1. This data

set was collected at low energy ($\lambda = 1.74 \text{ \AA}$) to enhance the anomalous signal of the metal ion. Both data sets were processed with HKL2000 ². See Table 3.1 for a summary of data processing statistics.

3.1.2 Phasing and refinement calculations

The structure was solved using single wavelength anomalous diffraction (SAD) phasing. Prior to phase calculations an anomalous difference Patterson map was created to determine the presence of the heavy atoms, Figure 3.1. Shake n Bake (SnB) ³ was used to identify a 10-atom anomalous constellation for the Sm derivative, which was input to SHARP ⁴ for SAD phase calculations and solvent flattening. The resulting SHARP phases had figure of merit of 0.84 for reflections to 2.4 \AA resolution. An electron density map calculated from the SHARP phases clearly showed features resembling protein secondary structural elements. A partial backbone tracing consisting of a few α -helices and β -strands was obtained with the automated model building program MAID ⁵. The MAID tracing was used to determine the noncrystallographic symmetry (NCS) transformation relating the two protein molecules in the asymmetric unit. In preparation for NCS averaging, the programs MAMA ⁶ and CNS ⁷ were used to create a mask that covered one of the molecules in the asymmetric unit. The SHARP phases were then improved and extended to 1.75 \AA resolution with 2-fold NCS averaging and solvent flipping in CNS. The 1.75 \AA resolution density modified phases were input to ARP/wARP ⁸ for automated electron density map interpretation. The best model from ARP/wARP included the backbone for 97%

of the expected residues in the asymmetric unit and 83% of the expected side-chains. The model was improved with several rounds of model building in COOT⁹ followed by refinement with REFMAC5¹⁰. Refinement statistics are listed in Table 4.1.

The asymmetric unit includes 953 amino acid residues belonging to two AcpA molecules (chains labeled A, B). The following sections of the polypeptide chains are disordered: A1 – A4, A15 – A18, A490 – A498, B1 – B5, B12 – B18, B129 – B137, B490 – B494. The root mean square difference between chains A and B is 0.27 Å for C α atoms and 0.55 Å for all atoms, which indicates that the two chains have nearly identical conformations. Each AcpA molecule contains one orthovanadate ion (HVO₄²⁻) and one metal ion bound in the active site. The metal ion was modeled as Ca²⁺ for purposes of crystallographic refinement, but appears with atom name X1 and residue name UNK in the coordinate file deposited in the Protein Data Bank (PDB¹¹) to indicate that the identity of the metal is unknown at this time. The solvent structure includes 550 ordered water molecules and 4 bound PEG fragments. There is also a decavanadate ion (V₁₀O₂₈⁶⁻) bound in a crystal contact region, where it interacts with the C-terminal histidine affinity tag of one of the AcpA molecules (Figure 3.12). Coordinates and structure factor amplitudes have been deposited in the PDB under accession code 2D1G.

Figure 3.1

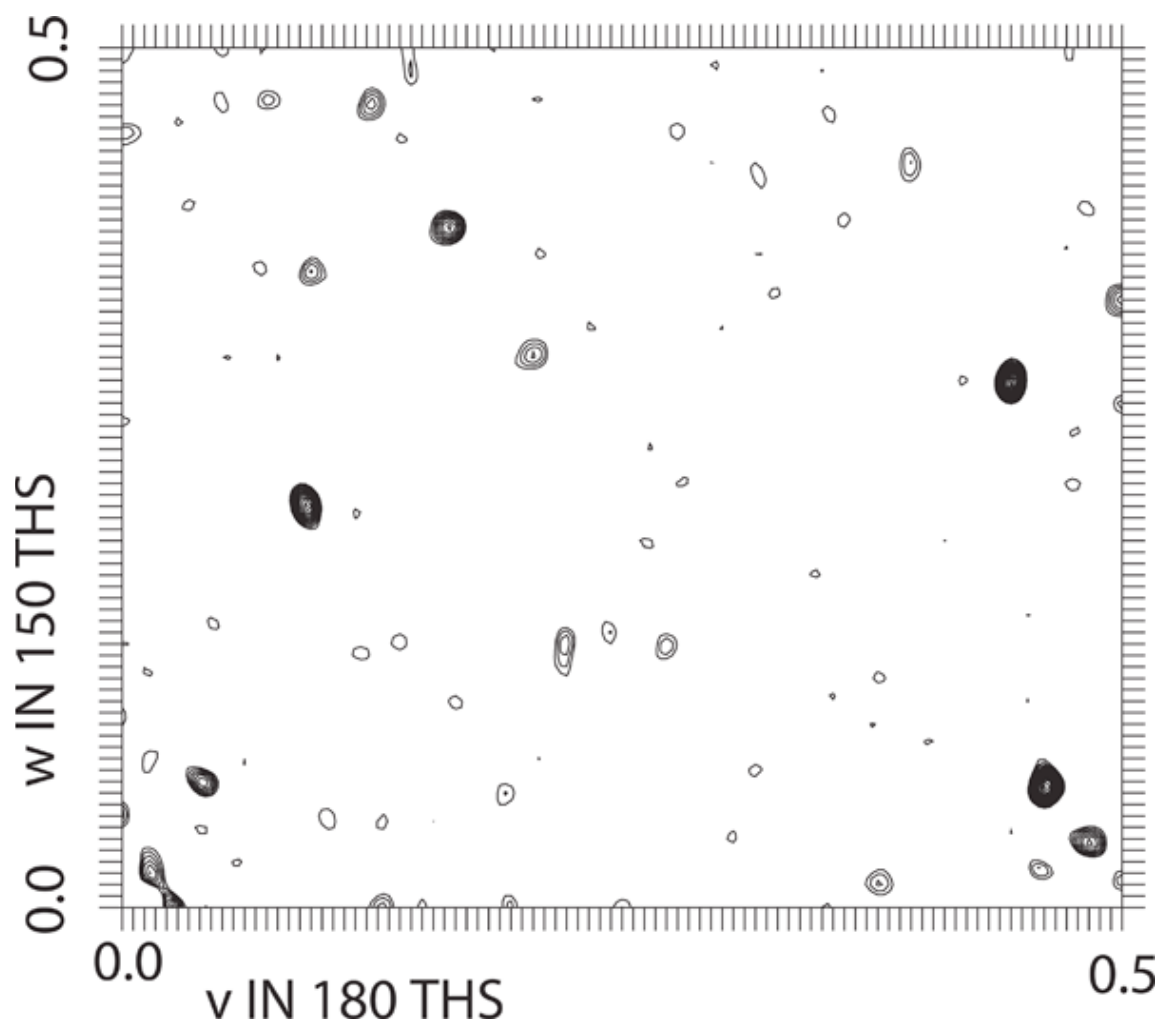


Figure 3.1 Anomalous Difference Patterson Map

This is the $u=0$ Harker section of the anomalous difference Patterson map using the Sm derivative data. The large peaks are indications that Sm atoms are bound to the protein.

Table 3.1

Data collection and refinement statistics ^a			
	Orthovanadate (high energy)	Orthovanadate (low energy)	Sm Derivative
Wavelength (Å)	1.1271	1.740	1.6531
Space group	C2221	C2221	C2221
Unit cell dimensions (Å)	a = 112.1, b = 144.4, c = 123.9	a = 111.2, b = 142.6, c = 125.6	a = 112.2, b = 144.2, c = 123.9
Diffraction resolution (Å)	50 - 1.75 (1.81 - 1.75)	50 - 2.20 (2.28 - 2.20)	50 - 2.40 (2.48 - 2.40)
No. of observations	397516	295644	572236
No. of unique reflections	98137	51195	39618
Redundancy	4.1 (4.0)	5.8 (4.9)	14.4 (13.4)
Completeness (%)	97.2 (94.9)	99.9 (99.9)	99.9 (99.9)
Average I/σ(I)	26.6 (2.2)	16.8 (3.4)	38.6 (23.8)
$R_{sym}(I)$	0.047 (0.357)	0.098 (0.443)	0.065 (0.118)
No. of non-hydrogen atoms	8182		
No. of residues in chain A	481		
No. of residues in chain B	472		
No. of water molecules	550		
R_{cryst}	0.198 (0.244)		
R_{free}^b	0.231 (0.291)		
RMSD ^c			
Bond lengths (Å)	0.012		
Bond angles (deg.)	1.7		
Ramachandran plot ^d			
Favored (%)	96.4		
Allowed (%)	3.1		
Average B-factors (Å ²)			
Protein	25		
Orthovanadate	27		
Active site metal ion	16		
Water	28		
PDB accession code	2D1G		

^aValues for the outer resolution shell of data are given in parenthesis.

^b5% random test set.

^cCompared to the Engh and Huber force field ¹².

^dThe Ramachandran plot was generated with RAMPAGE ¹³.

3.1.3 Site-directed mutagenesis and activity assays

AcpA mutant Ser175Ala was generated using the QuikChange mutagenesis kit (Stratagene) and the mutation was verified by DNA sequencing. The mutant enzyme was expressed and purified using methods employed for recombinant AcpA¹. SDS-PAGE analysis showed that Ser175Ala had the expected molecular weight and western blots using rabbit anti-AcpA polyclonal and anti-His-tag antibodies were positive. Enzymatic activities of AcpA and Ser175Ala were measured using a discontinuous colorimetric assay with *p*-nitrophenyl phosphate (*p*NPP) and *p*-nitrophenylphosphoryl choline (*p*NPPC) as substrates ¹⁴.

3.1.4 Determination of metal ion in AcpA

Elucidation of the metal ion in AcpA was done using Inductively Coupled Plasma (ICP) spectroscopy in conjunction with the colorimetric Ferrozine assay. Samples used for ICP were sent to the University of Georgia spectroscopy facility for analysis. The Ferrozine assay was done using a method described previously by Dawson and Lyle¹⁵. Ferrozine was purchased from Sigma (St. Louis, MO) and needed no further purification prior to its use. Prior to sample measurement a standard curve was made using FeCl₂. The experimentally determined extinction coefficient was 28,031 L mol⁻¹ cm⁻¹ which was close to what was reported in the literature, 27, 900 L mol⁻¹ cm⁻¹. To determine if Fe was present in AcpA, 100 µL of AcpA was mixed with 2 mL of reducing agent (100 mg/mL hydroxylamine), then 2 mL of 0.75 mg/mL Ferrozine was added. 5 mL of buffer, 0.5 M sodium

acetate/0.1 M sodium tartrate pH 6.8 was added and allowed to develop for 10 minutes. Once development was complete the mixture was diluted to 25 mL and the absorbance was measured at 562 nm. The concentration of Fe was then estimated using the standard curve.

3.2 Results

3.2.1 Overall structure of AcpA

The structure of AcpA comprises three domains and has approximate dimensions of 60 Å x 48 Å x 66 Å. The core domain is a highly twisted, 8-stranded β -sheet flanked by three α -helices on either side (Figure 3.2A). The strand order of the β -sheet is 12, 2, 11, 10, 1, 9, 3, 8, with all but strand 11 in parallel (Figure 3.3). There are two smaller domains located above the carboxyl terminal edge of the 8-stranded β -sheet. One of these small domains consists of residues 47-147 and features four short α -helices (labeled A-D) connected by rather long loops (Figure 3.2, blue domain). This domain has a disulfide bond linking Cys102 and Cys138 (Figures 3.2A, 3.3). As discussed below, this domain forms part of the dimer interface. The other small domain (residues 258-283) consists of a pair of 2-stranded anti-parallel β -sheets (β 4 – β 7), resembling a flap (Figure 3.2, orange domain). There is a disulfide bond that links Cys269 of this domain to Cys216 of the β -sheet core domain.

Purified AcpA forms an apparent dimer according to analytical ultracentrifugation and gel filtration chromatography data ¹⁴. The two proteins

chosen for the asymmetric unit (Figure 3.4A) form the largest intermolecular surface between any two proteins in the crystal lattice, based on analysis with PISA ¹⁶. This interface buries 2398 Å² of surface area while the next largest interface buries only 932 Å² of surface area. Furthermore, this interface had the highest possible PISA complexation significance score (CSS = 1.0), compared to CSS = 0 for all other possible interfaces. It is concluded that the pair of protein molecules in the asymmetric unit represents the AcpA dimer in solution.

The small helical domain (residues 47-147) and the β 12 face of the β -sheet core domain form the dimer interface (Figure 3.4A). Secondary structural elements involved in dimerization include α B and its adjacent loops (residues 73-87), the loop following α C (residues 116-119), a 10-residue section of the loop connecting β 10 and β 11 (394-404), β 12 and its adjacent loops (residues 425 - 433), and residues in a loop near the C-terminus (residues 459-466). Together, these residues form a flat surface (Figure 3.4B) that spans 40 Å in one direction and 30 Å in the other.

The dimer interface is highly hydrophilic and hydrogen bonding appears to play a major role in dimer stability. There are 14 direct intersubunit hydrogen bonds (Table 4.4), but no ion pairs. Hydrogen bonding side chains in the interface include Asn74, Thr79, Gln81, Asn116, Gln401, Asp404 and Tyr428. Note that the intersubunit hydrogen bonds display 2-fold symmetry (Table 4.2). In addition, there are 16 interfacial water molecules that mediate 20 intersubunit hydrogen bonds (Figure 3.4C). As with the intersubunit hydrogen bonds, the 16 bridging water molecules obey the 2-fold symmetry of the dimer (Figure 3.4C). Although hydrogen bonding is prominent in the interface, a few nonpolar

residues contribute significant surface area to the interface. For example, Leu82 packs against Leu119, while Leu433 from one subunit packs against Leu433 of the opposite subunit at the centroid of the dimer.

Prior to structure deposition into the PDB Ramachandran plots of both chains A and B were calculated in order to determine if the dihedral angles of the amino acids were in the allowable conformations. These plots can be found in Figures 3.10 A and B and the reported statistics can be found in Table 4.1.

Figure 3.2 Overall structure of AcpA

A) Stereographic ribbon drawing of AcpA chain A.

The vanadate inhibitor (magenta/red) and bound metal ion (yellow) are drawn in CPK mode. Strands and helices of the core β -sheet domain are colored green and red, respectively. The two smaller domains are colored blue (residues 47-147) and orange (flap domain, residues 258-283). Selected strands and helices are labeled as in the topology diagram (see Figure 3.4). Disulfide-bonded Cys side chains are drawn as grey CPK spheres. The thin dashed line indicates disordered residues 15 – 18.

B) Surface topography of the active site entrance. The orthovanadate inhibitor (green/red) and the four water molecules in the trough (red) are drawn as CPK spheres. The β -sheet core domain is colored white. The two smaller domains are colored blue (residues 47-147) and orange (residues 258-283). Residues lining the active site trough are indicated in the schematic diagram on the right. This figure, and others, were created with PyMol ¹⁷.

Figure 3.3

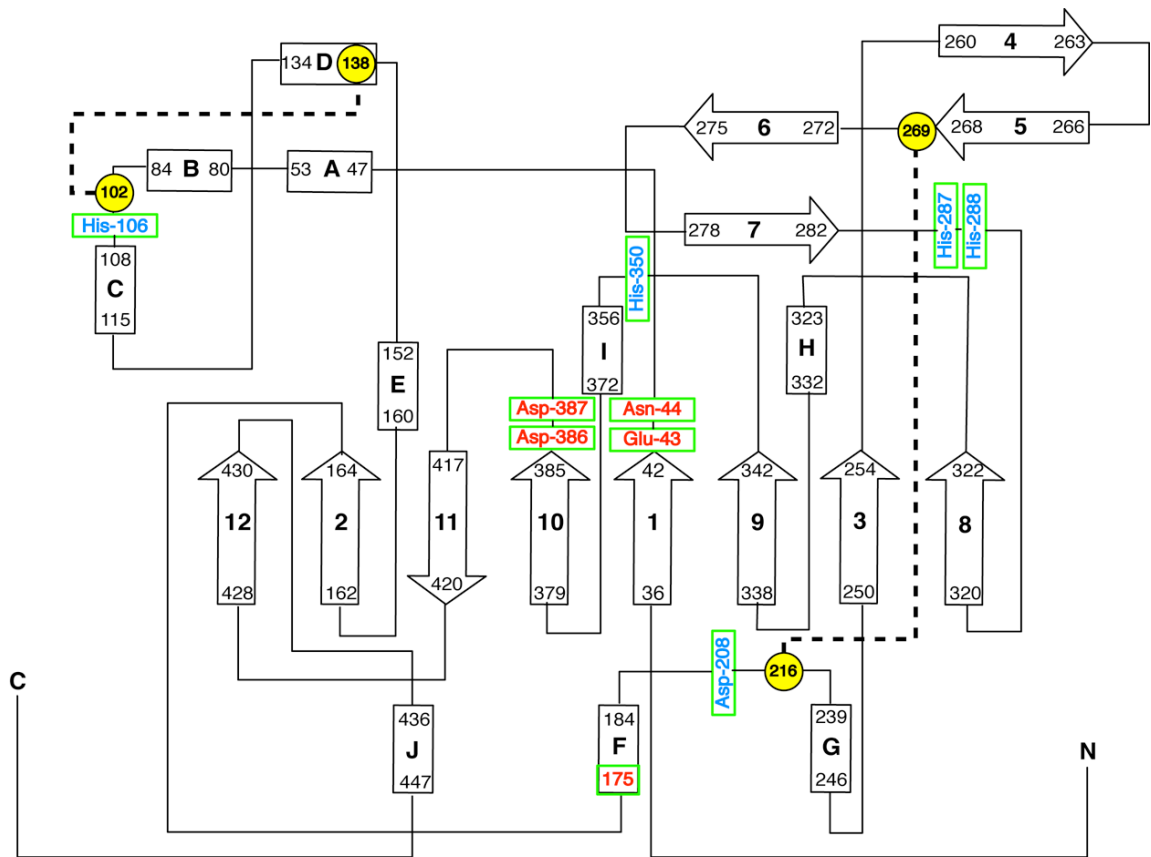


Figure 3.3 Secondary structure topology diagram of AcpA.

α -helices are shown as rectangles labeled A-J, and β -strands are shown as arrows numbered 1-12. Cys residues are represented in yellow with disulfides bridges shown as dashed lines. The green boxes denote active site residues, with red numbering for residues coordinating to the bound metal and blue numbering for residues interacting with the vanadate inhibitor.

Figure 3.4

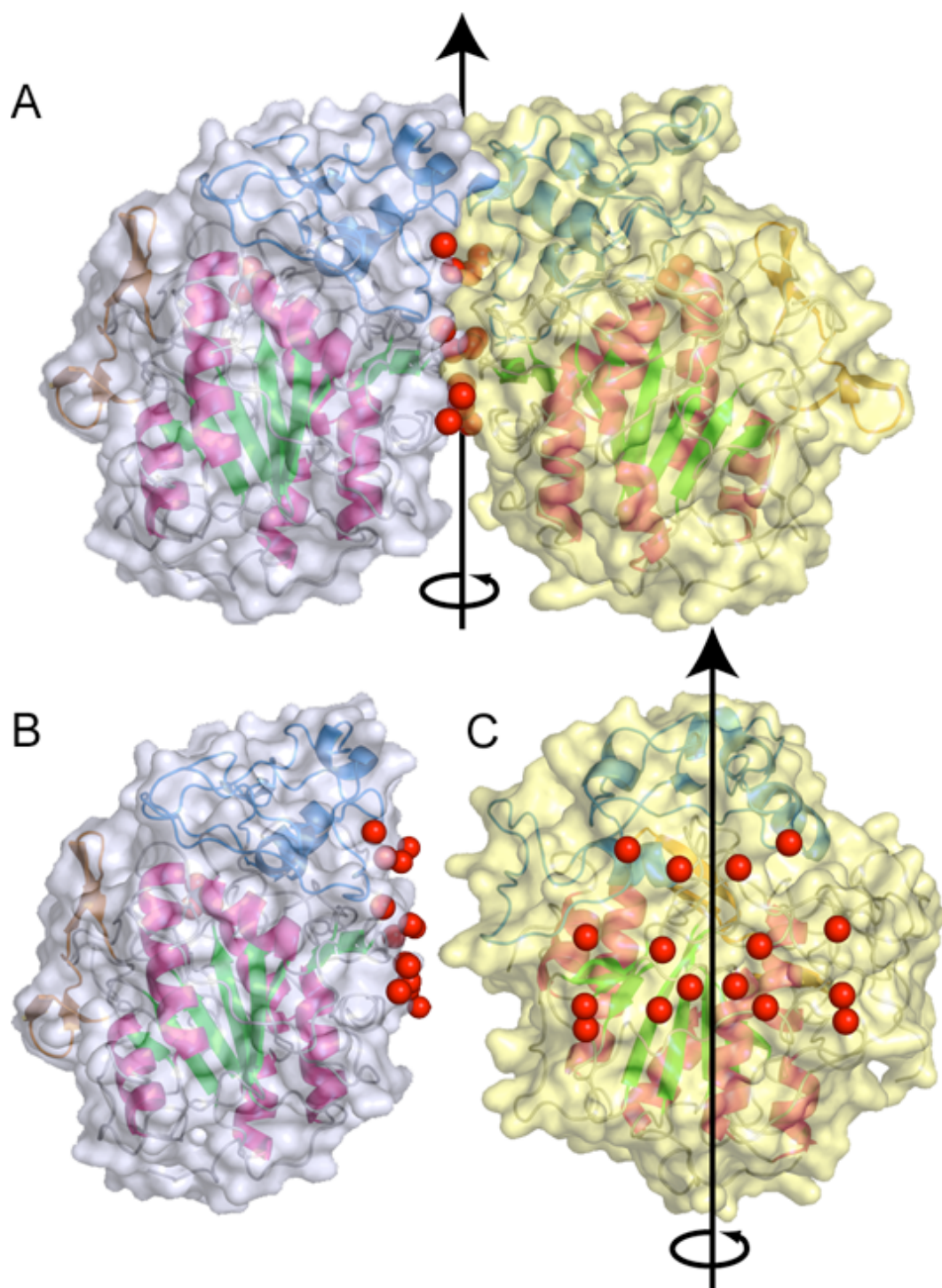


Figure 3.4. Dimeric structure of AcpA.

A) The AcpA dimer is drawn in ribbon representation and covered with a semi-transparent molecular surface. The surfaces are colored blue for one subunit and yellow for the other subunit. The orientation of the yellow subunit is nearly identical to that of Figure 3.2A. For each subunit, the coloring scheme of the ribbon is the same as that used in Figure 3.2. The red spheres represent interfacial water molecules. The arrow denotes the non-crystallographic molecular 2-fold symmetry axis. **B)** This panel is the same as panel A except that the yellow subunit has been removed to show the flatness of the dimer interface. **C)** This panel is the same as panel A except that the blue subunit has been removed and the yellow subunit has been rotated 90° so that the interfacial surface points toward the viewer. Note the 2-fold symmetry in the constellation of interfacial water molecules.

Table 3.2

Intersubunit hydrogen bonds

Hydrogen Bond Partners		Distance (Å)
Chain A	Chain B	
Asn 74 (ND2)	Asp 404 (OD1)	3.1
Asn 74 (N)	Asp 404 (OD2)	3.0
Asp 404 (OD1)	Asn 74 (ND2)	2.8
Asp 404 (OD2)	Asn 74 (N)	3.1
Thr 79 (OG1)	Gly 117 (O)	2.6
Gly 117 (O)	Thr 79 (OG1)	2.7
Gln 81 (NE2)	Asn 116 (OD1)	3.3
Asn 116 (OD1)	Gln 81 (NE2)	3.1
Gln 401 (NE2)	Tyr 428 (OH)	3.2
Gln 401 (NE2)	Val 429 (O)	3.0
Gln 401 (OE1)	His 431 (N)	2.8
Tyr 428 (OH)	Gln 401 (NE2)	3.2
Val 429 (O)	Gln 401 (NE2)	3.1
His 431 (N)	Gln 401 (OE1)	2.9

3.2.2 Active site architecture and implications for catalytic mechanism

The location of the active site was clearly indicated by a strong electron density feature corresponding to the bound orthovanadate inhibitor (Figure 3.7A). The active site is located above the carboxyl-terminal edge of the 8-stranded β -sheet near $\beta 1$ and αF (Figure 3.2A). The inhibitor binds in one end of a 12 Å long trough, which is located in a broad, shallow depression formed by residues from all three domains (Figure 3.2B). Note that four water molecules are bound in the trough (Figure 3.2B). The shape of the trough suggests that it may be involved in binding the leaving group of the substrate. This idea was tested by modeling *p*NPP in the active site. We found that the nitrophenyl group of *p*NPP fits edgewise into the trough (water removed) without causing steric clash (Figure 3.6).

Surprisingly, there is a metal ion bound in the active site, based on the observation of a very strong electron density feature that could not be assigned to the protein, inhibitor or solvent (Figure 3.7A). Four lines of evidence suggest that this feature represents a metal ion. First, it is surrounded by an octahedral array of six oxygen ligands (Figures 3.7A, 3.7B): Glu43, Asn44, Ser175, Asp386, Asp387, and the vanadate inhibitor. Three of the six coordinating ligands are carboxyl groups, which is suggestive of a bound metal ion with charge of at least +2. Second, the proposed metal site corresponded to the highest peak in an anomalous difference Fourier map calculated from diffraction data collected at low energy ($\lambda = 1.74$ Å). The anomalous difference density feature was quite

prominent when viewed at the 3.5 σ contour level (Figure 3.7A, red cage) and remained visible even at 10 σ . Third, the proposed metal ion site was the second strongest binding site of the Sm derivative used for SAD phasing. We note that lanthanides readily replace metal ions of protein active/binding sites, including Mg^{2+} , Ca^{2+} and first row transition metal ions¹⁸. Fourth, as described in the next section, structural homologs of AcpA, such as arylsulfatase A (ASA) and alkaline phosphatase (AlkP), have metal ions bound in the active site at locations that are structurally analogous to the proposed AcpA metal site.

Electron density maps were analyzed to gain insights into the elemental identity of the bound metal ion. The anomalous difference Fourier peak corresponding to the metal ion was much stronger than that of the vanadate (Figure 3.7A, red cage), which implies that the metal ion is a stronger anomalous scatterer than the V atom of the inhibitor. This result is consistent with the active site metal ion being a first row transition metal. Also, several simulated annealing refinements were performed against the 1.75 Å data set with different metal ions modeled in the active site. The resulting difference electron density maps (σ_A -weighted mFo-DFc) suggested that the metal ion has at least the number of electrons of Ca^{2+} and is more likely a first row transition metal ion. Therefore, the metal ion was conservatively modeled as Ca^{2+} in the published structure pending further biochemical and analytical studies of the metal content of AcpA.

Using chemical and spectroscopic analysis the identity of the metal ion in AcpA has been investigated. Ferrozine is a Fe(II) specific chelator, that is used for colorimetric detection of Fe in water. The sensitivity of this reaction is 1-25

μg/mL. When AcpA was treated with Ferrozine as described in the methods section, the solution turned purple indicating that Fe was present in the sample. Quantitation of the Fe in AcpA was done using the method that had been previously reported in the literature¹⁵. The initial results indicated that there were 0.7 -1.2 Fe ions per AcpA molecule. This data was interpreted as only one Fe ion per molecule and that fluctuations were measured within the samples. These fluctuations are due to precipitation of AcpA over time in the mixture.

Analysis for metals in AcpA was done using Inductively Coupled Plasma (ICP) spectroscopy. To prepare the samples the protein was dialyzed against either 50 mM sodium acetate buffer for at least 12 hours, keeping a small sample for a pre dialysis analysis. After dialysis the protein, buffer and pre-dialysis samples were sent to the University of Georgia Spectroscopy facility for analysis. For the AcpA sample ICP-ES was used. With emission spectroscopy many elements can be measured simultaneously. The elements chosen for AcpA included many of the first row transition metals as well as calcium and magnesium. The results of this analysis (Figure 3.5) clearly indicate that iron has the highest relative signal compared to all other metals tested. No other metals had significant signal greater than either the water or the pre-dialysis buffer. The signal from Fe however is less than what was calculated for the amount of enzyme used for analysis. A decrease in this signal is most likely due to the lengthy dialysis procedure used for sample preparation. This data in conjunction with the Ferrozine assay data would support the hypothesis that iron is the most likely metal in AcpA.

Figure 3.5

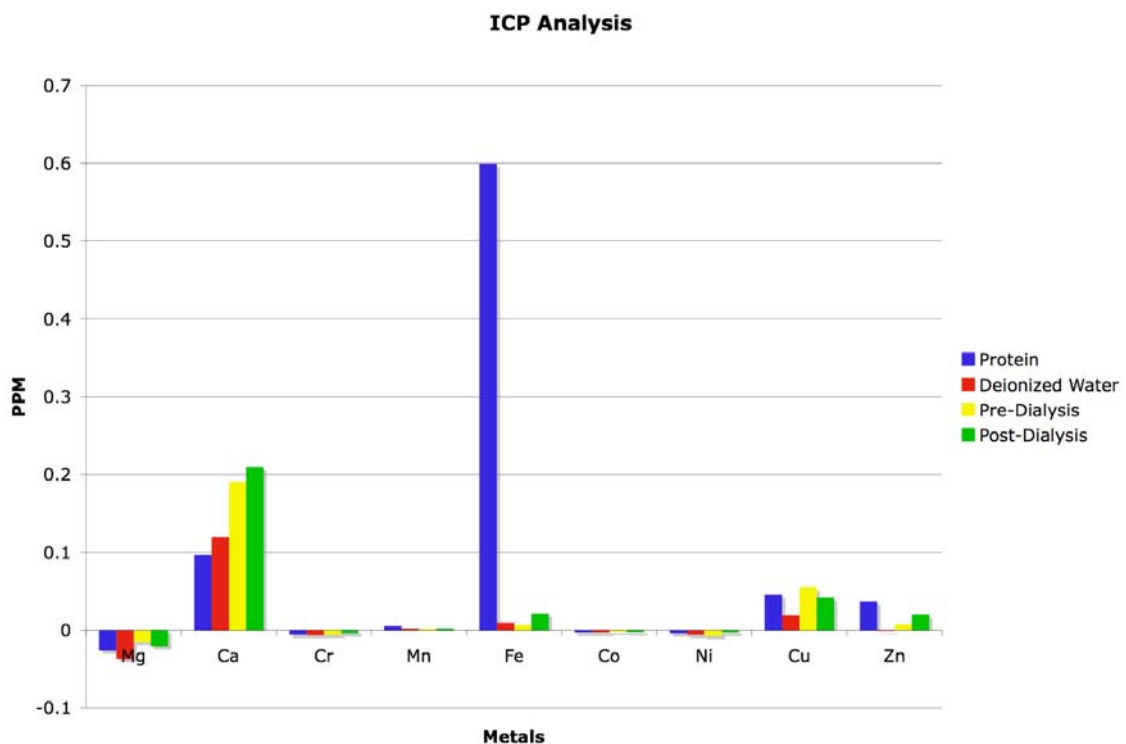


Figure 3.5 Inductively Coupled Plasma-Emission Spectroscopy Results

The protein sample is shown in blue, the deionized water is in red, the pre-dialysis buffer is in yellow, and the post-dialysis buffer sample is in green.

Figure 3.6

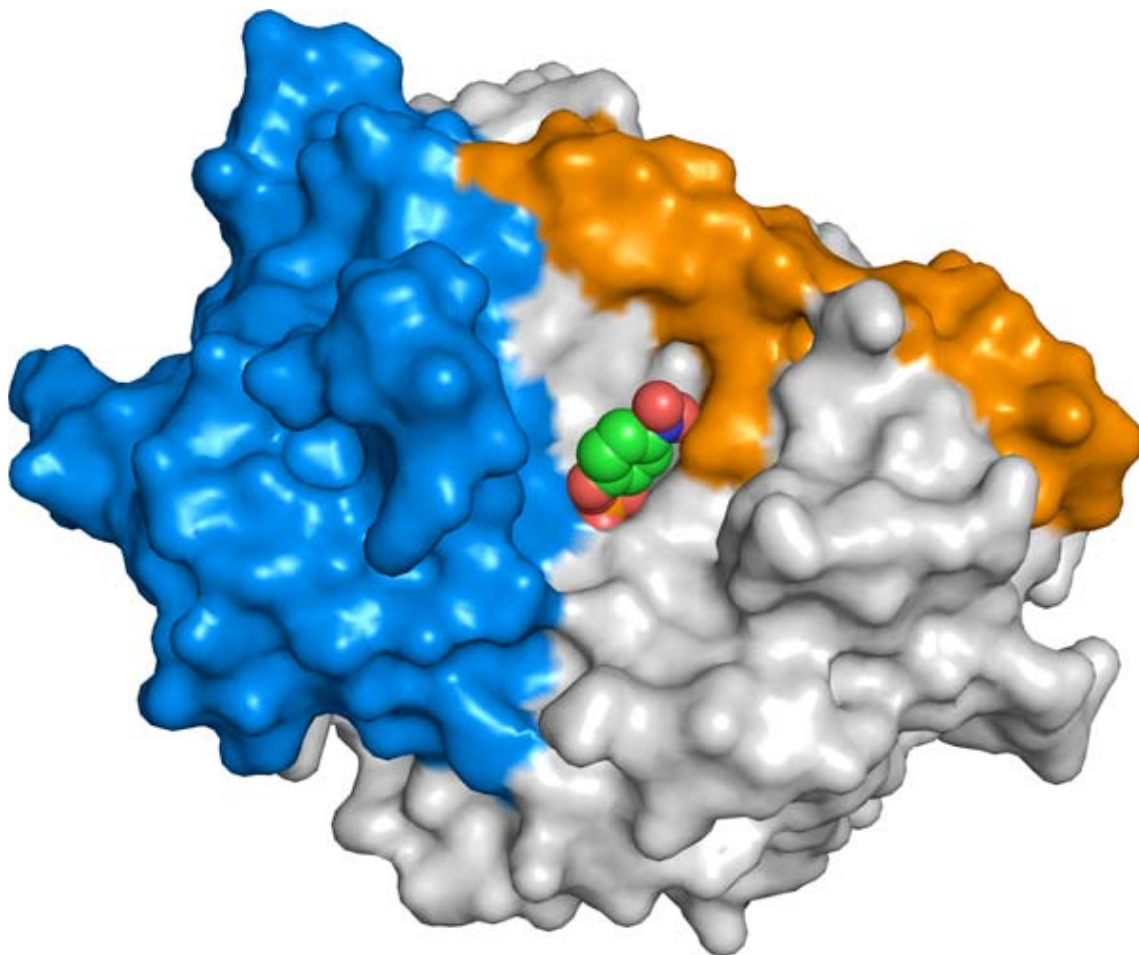


Figure 3.6 *p*NPP modeled in the AcpA trough

The *p*NPP molecule is located in the trough. The phosphate group overlays the enzyme-bound orthovanadate molecule. The R group fits into the trough that was occupied with water molecules in the structure. The phenyl ring is depicted as green spheres.

Figure 3.7

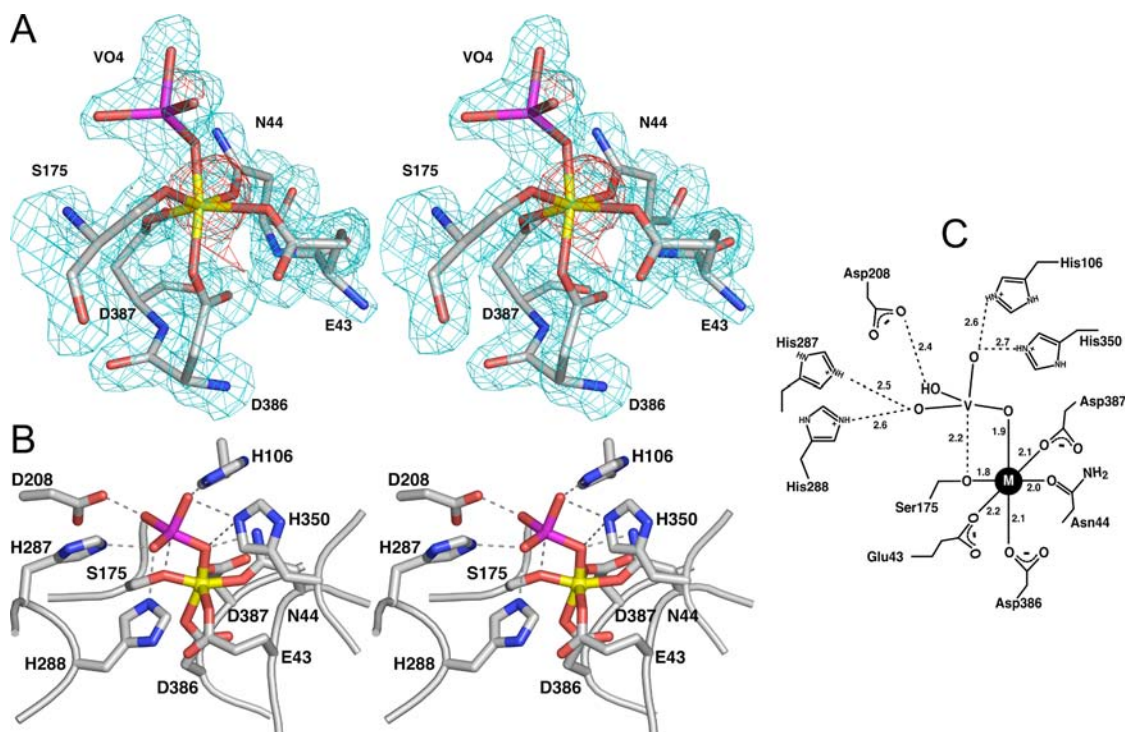


Figure 3.7 Active site of AcpA.

A) Stereographic drawing of the AcpA metal center covered by two electron density maps. The cyan cage represents a simulated annealing σ A-weighted mFo-DFc electron density map (3σ). The metal ion, orthovanadate, and surrounding residues within 3.9 Å were omitted prior to simulated annealing refinement and map calculation. The red cage represents an anomalous difference Fourier map (3.5σ) calculated with phases from the final model and anomalous differences from the low energy orthovanadate data set. The orthovanadate inhibitor is shown in magenta/red and the metal ion is colored yellow. Protein residues appear in white. **B)** Stereographic drawing of the AcpA

active site highlighting protein-inhibitor electrostatic interactions (dashed lines). The orthovanadate inhibitor is shown in magenta/red and the metal ion is colored yellow. Protein residues appear in white. **C)** Schematic diagram of the active site.

The orthovanadate inhibitor exhibits distorted trigonal bipyramidal geometry and is bound by 6 side-chains and the metal ion (Figures. 3.6B, 3.6C). The inhibitor axial oxygen atom interacts with His106 and His350, while the equatorial oxygen atoms bind to Asn44, His287, His288, Asp208, His350 and the metal ion. Asp208 appears to share a proton with the inhibitor.

The location of Ser175 relative to the inhibitor and metal ion suggests that it plays the role of nucleophile that attacks the substrate P atom. The hydroxyl oxygen atom of Ser175 is 1.8 Å from the metal ion and 2.2 Å from the inhibitor V atom (Figure 3.7C). Ser175 appears to be in an ideal location for backside nucleophilic attack at the substrate P atom. Thus, the active site structure strongly suggests that Ser175 is the enzyme nucleophile and the role of the metal ion is to activate Ser175 for nucleophilic attack (Figure 3.8A, step 2). This hypothesis implies formation of a covalent Ser175-phosphoryl intermediate during catalysis (Figure 3.8A, step 2).

We engineered the Ser175Ala mutant to test the importance of this residue for catalysis. The mutant exhibited no detectable activity using either *p*NPP or *p*NPPC as the substrate even at enzyme concentrations greater than 0.01 mM and substrate concentrations up to 20 mM. Thus, Ser175 plays an essential role in catalysis, consistent with our hypothesis that it is the enzyme nucleophile.

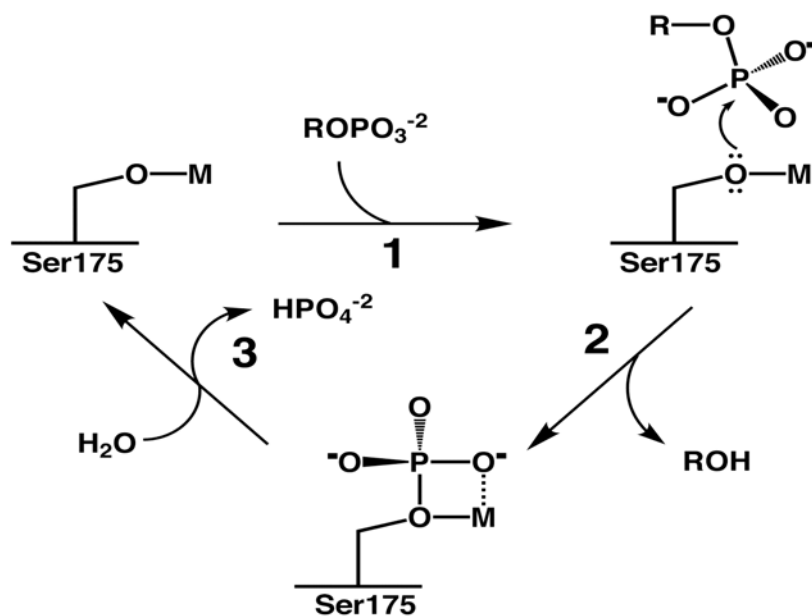
Ser175 is bound to an unknown metal in the structure of AcpA. This unknown metal may serve to deprotonate the serine thereby activating it for nucleophilic attack at the phosphorus atom of the substrate. This alkoxide formation is supported in the literature with the well-studied alkaline phosphatase (AlkP)^{19, 20}. In AlkP the serine nucleophile is deprotonated by Zn(II)

and acts as the initial nucleophile to attack the phosphate yielding a phospho-enzyme intermediate. In the free enzyme form the pK_a of the serine nucleophile is reported to be 5.5²⁰. This change in pK_a represents large shift from the value of free serine. Although the pK_a of the hydroxyl of free serine is not commonly known, we note that the pK_a of methanol is 15.5.

Hydrolysis of the Ser175-phosphoryl intermediate (Figure 3.8A, step 3) presumably requires a general base to activate a water molecule. Residues that bind the inhibitor and that are located on the solvent side of the active site are possible candidates for this role. Asp208 is, perhaps, the most likely candidate because aspartic acid residues serve as the general base in other phosphatases, such as protein tyrosine phosphatase ²¹, and the carboxyl of Asp208 forms a hydrogen bond (2.8 Å) with a molecule (Wat285) in our structure.

Figure 3.8

A



B

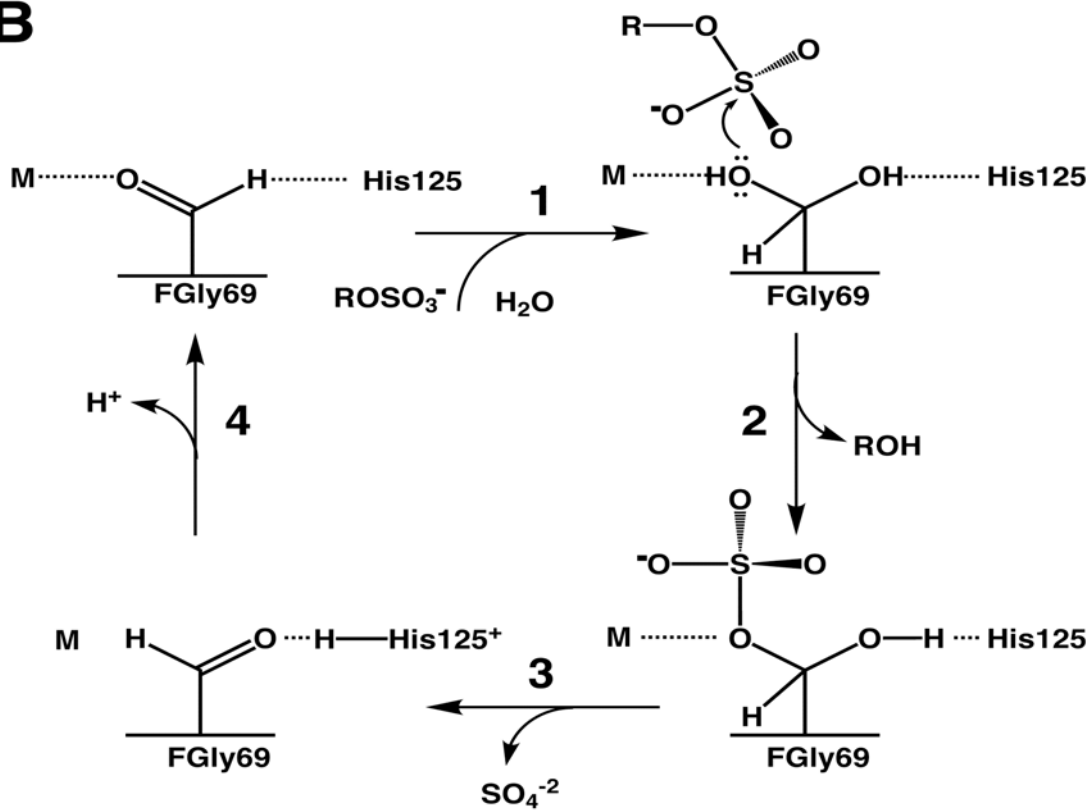


Figure 3.8 Summaries of the proposed catalytic mechanisms of AcpA and arylsulfatase A (ASA).

A) The proposed mechanism of AcpA. The steps shown are (1) substrate binding (see Figure 3.7 for more details), (2) nucleophilic attack at the substrate P atom by Ser175, which is activated by a metal ion, and (3) hydrolysis of the enzyme-phosphoryl intermediate to release the phosphate product and regenerate the active site. **B)** The ASA mechanism proposed previously ^{22, 23}. The steps shown are (1) formation of a gem-diol from the reaction of water with FGly69, (2) nucleophilic attack at the substrate S atom by one of the gem-diol hydroxyl groups and release of ROH, (3) elimination of the product sulfate facilitated by proton transfer to His125 and (4) regeneration of the active site.

3.2.3 Comparison to other protein structures

To understand the relationship of AcpA to other phosphatases, we searched the PDB for structural homologs of AcpA using the program DALI ²⁴. Surprisingly, the closest homolog was not a phosphatase, but was human arylsulfatase A (ASA, PDB code 1AUK, DALI Z-score = 16), followed by phosphoglycerate mutase (PDB code 1EJJ, Z = 15), phosphonoacetate hydrolase (PDB code 1EI6, Z = 10) and *E. coli* alkaline phosphatase (AlkP, PDB code 1B8J, Z = 9). All four enzymes belong to the AlkP superfamily, which has been described in detail ²⁵. AcpA shares a common β -sheet core domain and active site location with AlkP superfamily members. The shared secondary structural elements consist of the middle 6 strands of the central β -sheet along with the 6 flanking α -helices (Figure 3.9A). We note that DALI did not identify a single ACP with structural similarity to AcpA.

Structural homology of AcpA to AlkP enzymes, especially ASA and AlkP, extends to details of the active site. ASA uses a formylglycine (FGly69) as the nucleophile and binds a single metal ion in the active site (Ca^{2+} or Mg^{2+}) ²². The array of metal-binding ligands in ASA is remarkably similar to that of AcpA (Figure 3.9B). In both enzymes, the metal ion has octahedral coordination with three carboxyl groups, an asparagine side chain, the nucleophilic oxygen atom and the inhibitor/phosphoryl. Recognition of the substrate phosphoryl is also similar in the two enzymes. For example, His287 and His350 of AcpA are structurally analogous to His125 and His229, respectively, of ASA (Figure 3.9B). Moreover, imidazole nitrogen atoms of AcpA His106 and His288 overlap nearly

perfectly with the ϵ -amino groups of ASA Lys302 and Lys123 (Figure 3.9B). The only major difference between the two active sites, besides the nucleophile, appears to be D208 of AcpA, which is replaced by Val91 in ASA.

Figure 3.9

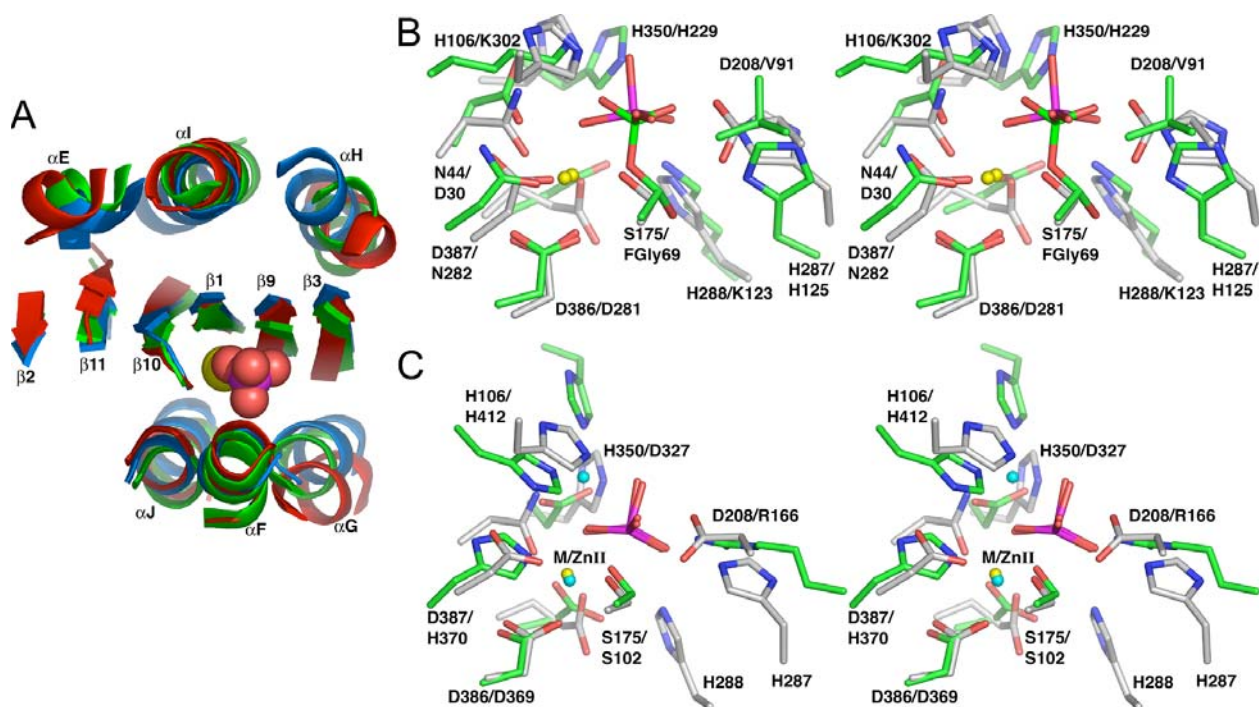


Figure 3.9. Comparison of AcpA to arylsulfatase A (ASA) and alkaline phosphatase (AlkP).

A) Ribbon representation of the conserved cores of AcpA (red), ASA (green, PDB code 1N2K ²³) and AlkP (blue, PDB code 1B8J ²⁶). The orthovanadate inhibitor (magenta/red) and metal ion (yellow) of AcpA are drawn as CPK spheres. Strands and helices of AcpA are labeled as in the topology diagram (Figure 3.4). **B)** Stereographic drawing of the ASA active site (green) superimposed onto the AcpA active site (white). Orthovanadate of AcpA is colored magenta/red. The phosphoryl intermediate of ASA is colored green/red. The metal ions of each structure are colored yellow. Selected residues are labeled as AcpA/ASA. **C)** Stereographic drawing of the AlkP active site (green) superimposed onto the AcpA active site (white). The orthovanadate inhibitors of both structures are colored magenta/red. The AcpA metal ion appears in yellow and the Zn ions of AlkP are colored cyan. Selected residues are labeled as AcpA/AlkP.

Figure 3.10

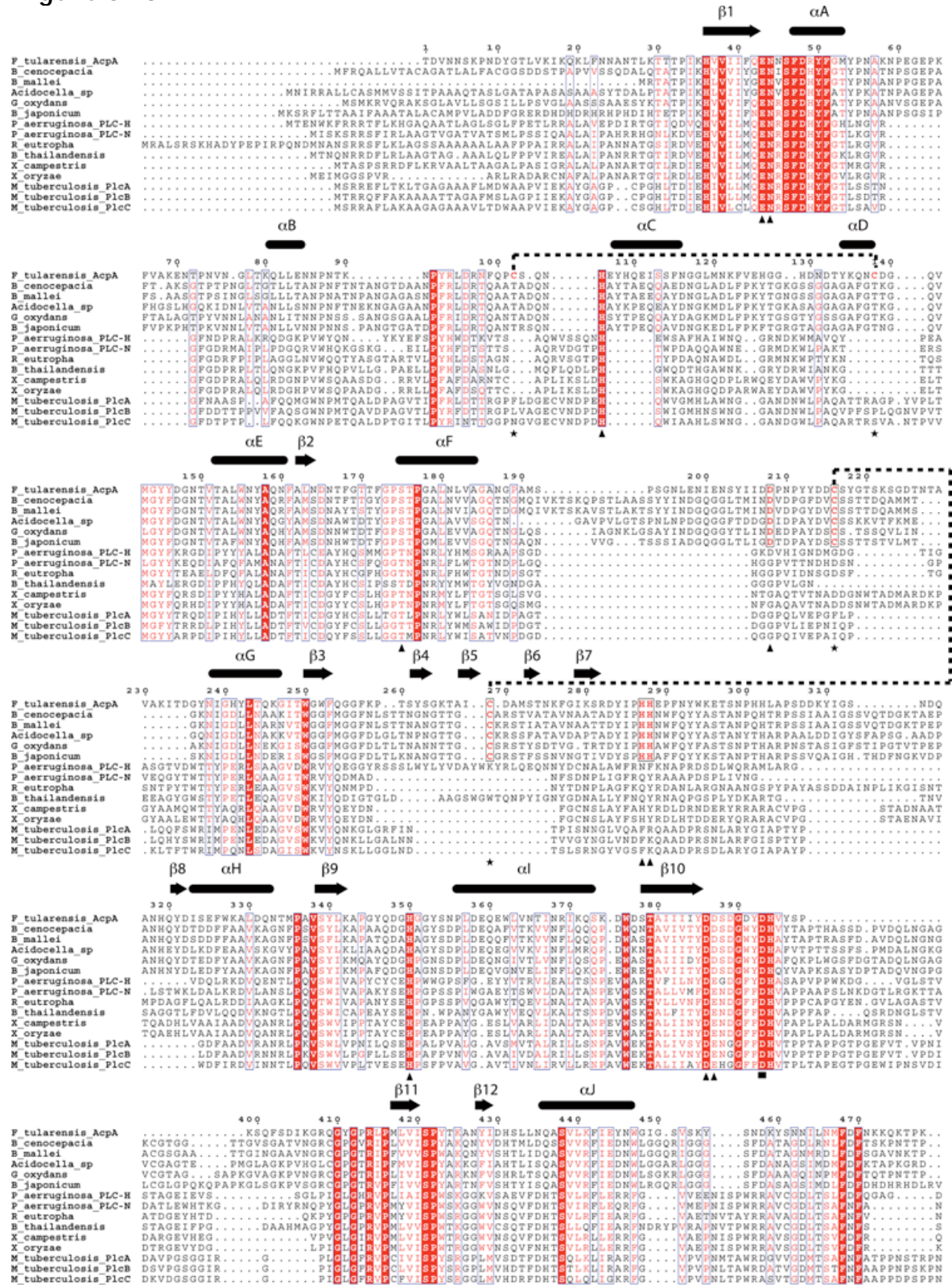


Figure 3.10 Amino acid sequence alignment of AcpA/PlcH superfamily members.

The upper six proteins of the alignment are AcpA and 5 close homologs of AcpA. These proteins share 37 - 84% pairwise amino acid sequence identity. The lower 9 proteins are PlcH and its close homologs. These proteins have 31 - 75% pairwise identity. Numbers above the alignment correspond to AcpA residue numbering. Residues identically conserved in all sequences are denoted by white letters on red background. Other regions of high sequence conservation are indicated by boxed red letters on white background. Triangles below the alignment denote active site residues that contact either the metal ion or vanadate inhibitor in AcpA. Stars below the alignment indicate disulfide-bonded Cys in AcpA. Dotted lines connect disulfide-bonded Cys residues. Filled rectangles below the alignment indicate conserved residues Asp393 and Arg414 (AcpA numbering), which form an ion pair in AcpA. This figure was created with ESPript ²⁷.

Like AcpA, AlkP has a Ser nucleophile ²⁶, but AlkP binds three metal ions: two Zn²⁺ (ZnI, ZnII) and Mg²⁺. The AcpA metal ion and AlkP ZnII occupy analogous locations in their respective structures, although ZnII has tetrahedral coordination and the AcpA metal ion has octahedral coordination (Figure 3.9C). Despite the difference in coordination geometry, Glu43 and Asp386 of AcpA are analogous to Asp51 and Asp369 of AlkP. There are also similarities between the two enzymes in terms of binding the substrate phosphoryl group. For example, AcpA His106 is analogous to AlkP His412, while the side chains of AcpA His287 and AlkP Arg166 occupy similar locations in their respective structures (Figure 3.9C). One notable difference between the two enzymes is that AlkP does not have an acidic residue equivalent to AcpA Asp208.

3.2.4 Conservation of active site residues in the AcpA/PlcH superfamily

Analysis of available sequence databases using BLAST ²⁸ shows that close homologs of AcpA are present in other bacteria including *Burkholderia mallei*, *Corynebacterium jeikeium* and *Bradyrhizobium japonicum*. These proteins have 520-648 residues and share 38-46% global amino acid sequence identity with AcpA. The 10 residues that contact vanadate or the metal ion in AcpA are identically conserved in these proteins: Glu43, Asn44, His106, Ser175, Asp208, His287, His288, His350, Asp386 and Asp387 (Figure 3.10, upper 6 protein sequences). Moreover, residues that form a disulfide bond (Cys216 - Cys269) and an ion pair (Asp393 - Arg414) are also identically conserved (Figure 3.10, upper 6

sequences). The conserved ion pair links residues within the long loop between $\beta 10$ and $\beta 11$. This loop also contains metal-binding residues Asp386 and Asp387 and thus the conserved ion pair may be crucial for stabilizing the metal binding site.

The AcpA structure also provides insights into the active site architectures of enzymes from the PLC branch of the AcpA/PlcH superfamily. PlcH has 730 residues and the N-terminal two-thirds of the enzyme shares 23 % amino acid sequence identity with AcpA. The AcpA structure, however, suggests that the sequence homology between AcpA and PlcH is much stronger within the active site. For example, five of the ten AcpA active site residues (Glu43, Asn44, His106, His350 and Asp386) are identically conserved in PlcH (Figure 3.10). Moreover, PlcH residue Thr178 aligns with AcpA nucleophile Ser175, and PlcH residue Glu358 aligns with AcpA metal-binding residue Asp387 (Figure 3.10). In addition, ion pair residues Asp393 and Arg414 of AcpA are also present in the PlcH sequence (Asp364, Arg401). All nine of these conserved residues are also present in close homologs of PlcH (Figure 3.10, lower 9 protein sequences). Thus, enzymes in the PLC branch of the AcpA/PlcH superfamily likely retain the essential hydroxyl nucleophile and octahedral metal-binding site of AcpA.

Figure 3.11A

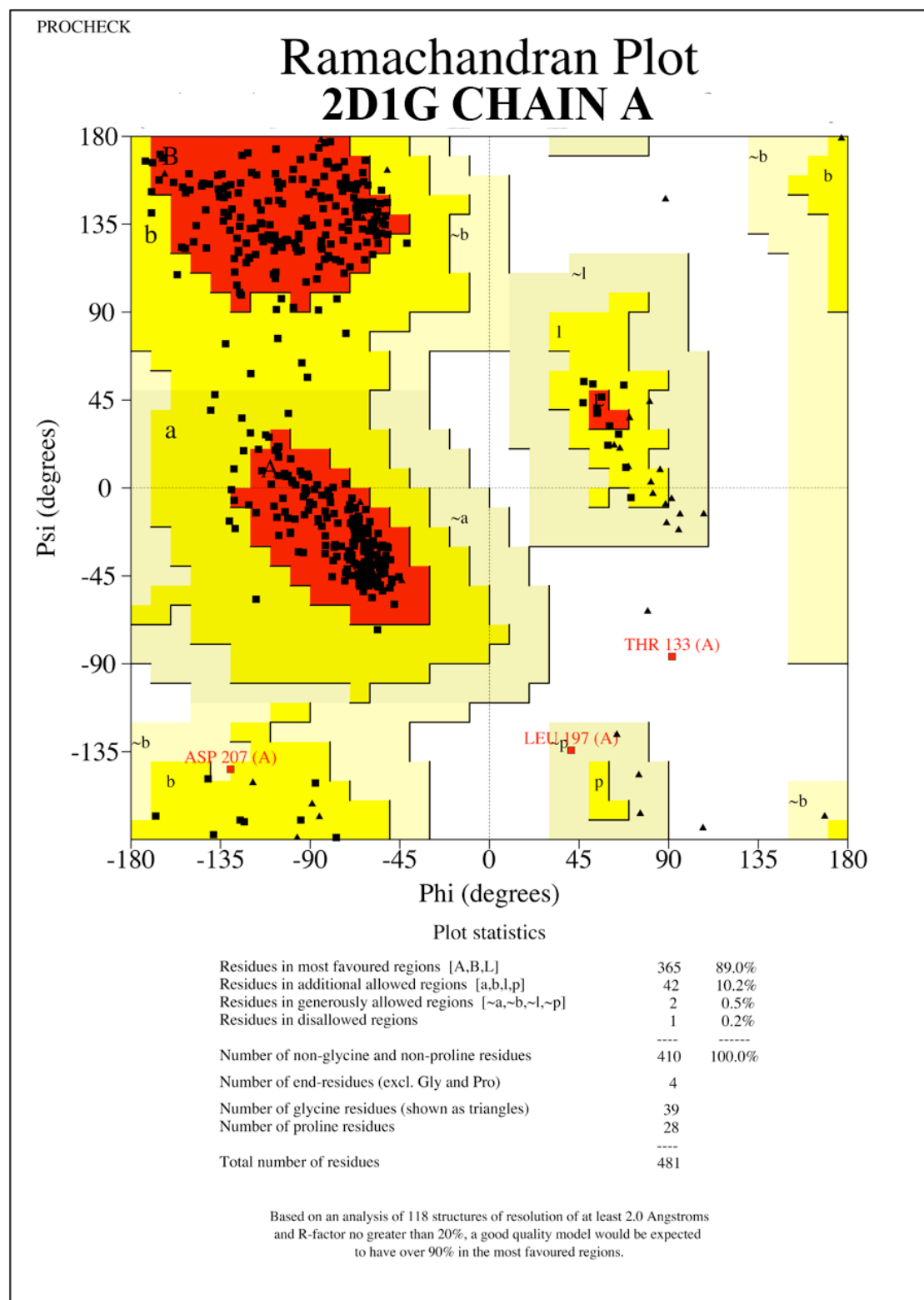


Figure 3.11B

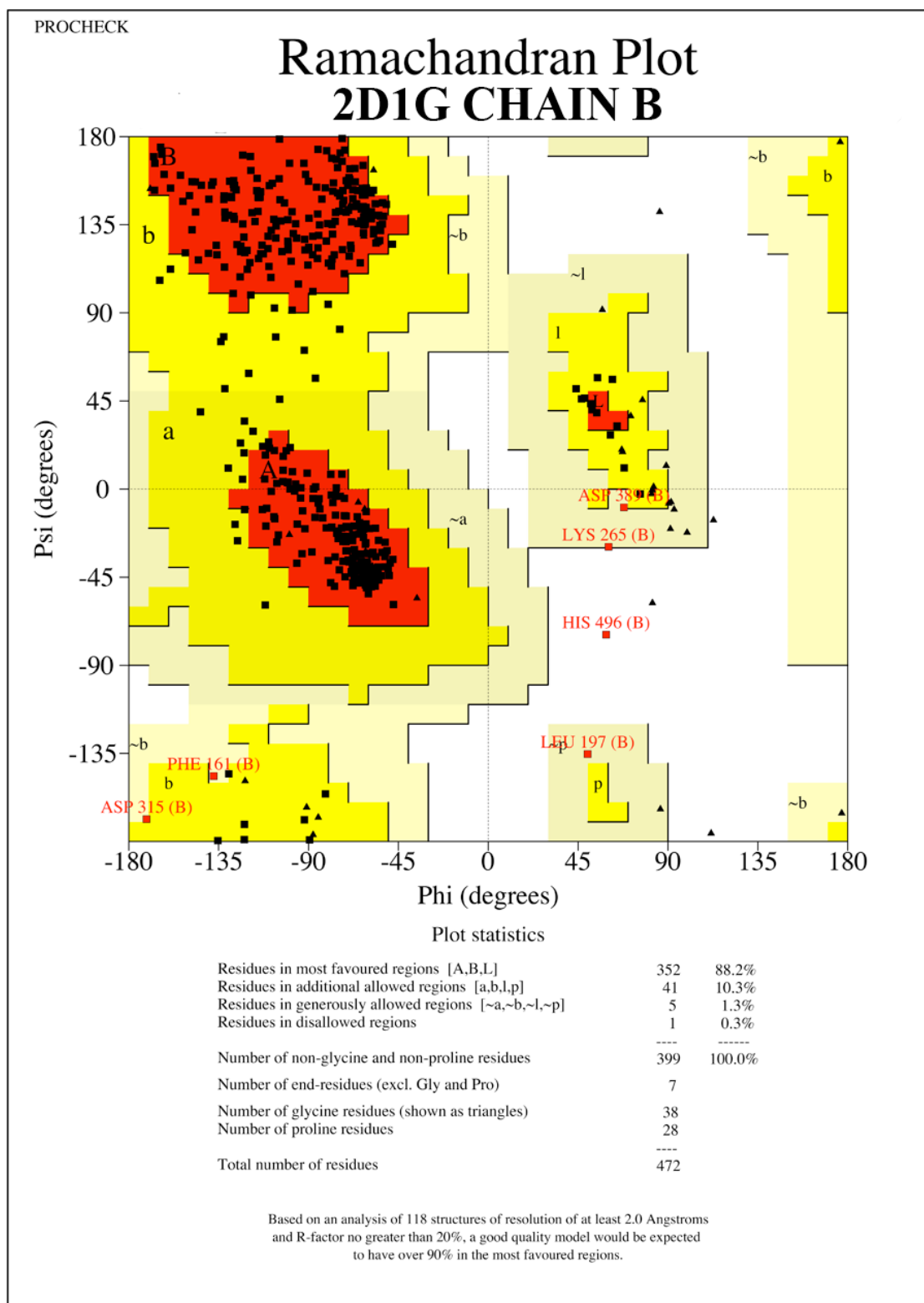


Figure 3.11A&B Ramachandran Plots of AcpA chain A and B

These Ramachandran plots were generated using Procheck²⁹. A represents chain A with 89% of the residues in the allowable region, an 11% in the other regions. B represents chain B and has 88% of the residues in the allowable region and 12% in the other regions. These statistics are less favorable than those calculated using the RAMPAGE¹³ server. Those statistics can be found in Table 4.1.

Figure 3.12

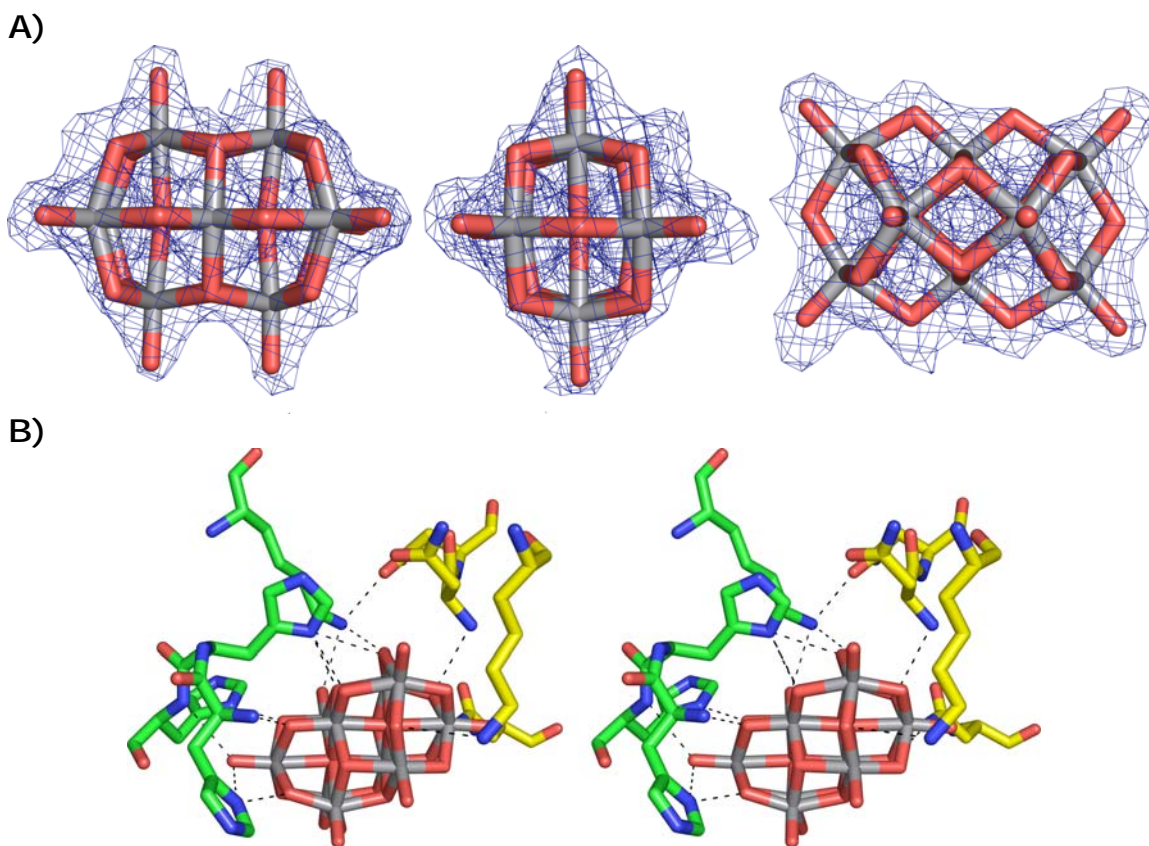


Figure 3.12 Decavanadate and its interactions with AcpA

A) This is the decavanadate molecule from the AcpA structure shown with electron density. The grey sticks are the vanadium atoms and the red sticks are the oxygen atoms. Decavanadate has the molecular formula $V_{10}O_{28}^{6-}$. The density is from a simulated annealing omit map calculated in CNS at 3.5σ . **B)** Stereo drawing of the histidine tag of molecule B interacting with the decavanadate as well as K406, N202, and N196 of molecule A. The decavanadate is located in a crystal contact region and has a significant role in crystallization. This is the first high-resolution decavanadate structure in the Protein Data Bank.

3.3 Discussion

3.3.1 The AcpA family of phosphatases

The structure reported here shows that AcpA from *F. tularensis* is distinct from other ACPs in terms of overall fold and active site architecture and that AcpA shares a common α/β core with enzymes in the AlkP superfamily. A major result from our work is the discovery that AcpA is a Ser-based metallophosphatase. This result was unexpected because AcpA had been predicted to be a Cys-based phosphatase based on a putative catalytic motif in residues 216-224 (CX5KSG). Furthermore, there were no reports in the literature showing that metal ion is required for activity. The AcpA structure thus provides the framework for experiments that will establish new paradigms for the AcpA family of ACPs.

The structural similarity between AcpA and AlkP enzymes sheds new light on the catalytic mechanism of AcpA. For example, the mechanism of AlkP involves two consecutive in-line nucleophilic attacks at the phosphorous, which results in retention of configuration at the P center^{19, 30-32}. ZnII activates Ser102 for the first nucleophilic attack and ZnI activates a water molecule for the second attack. We propose that AcpA follows an AlkP-like mechanism with AcpA Ser175 serving as nucleophile, the AcpA metal ion playing the role of AlkP ZnII and a protein side chain - possibly Asp208 - substituting for AlkP ZnI.

Interestingly, although the active site structures of ASA and AcpA are very similar (Figure 3.9B), it is unlikely that they share a common catalytic

mechanism. The ASA mechanism proceeds through a gem-diol intermediate formed by reaction of water with FGly69 (Figure 3.7B, step 1). One of the hydroxyl groups of the gem-diol is activated for nucleophilic attack at the substrate S atom (Figure 3.7B, step 2). Proton transfer from the other hydroxyl to His125 facilitates release of the product sulfate (Figure 3.7B, step 3). Since the nucleophilic oxygen atom leaves with the product, the reaction occurs with overall inversion of configuration of the sulfate ^{22, 23}. Formation of a gem-diol intermediate from AcpA Ser175 is chemically unfavorable. Thus it appears that AcpA and ASA have quite different catalytic mechanisms despite having nearly identical metal ion binding sites, similar constellations of substrate-binding residues, and a common protein fold.

3.3.2 Catalytic mechanism of AcpA-like PLCs

The AcpA structure also provides new insights into the catalytic mechanism of PLCs of the AcpA/PlcH superfamily. As discussed above, structure-based sequence analysis suggests that these PLCs have a hydroxyl nucleophile (e.g. Thr178 in PlcH) coupled to an octahedral metal center in the active site. Involvement of a threonine nucleophile implies a double-displacement catalytic mechanism for PlcH and related PLCs in which a covalent intermediate is formed between the nucleophilic threonine and the phosphoryl head group of the substrate. This predicted mechanism is radically different from that of zinc metallophospholipases *C. perfringens* alpha-toxin and *B. cereus* phosphatidylcholine-preferring PLC, which utilize a single nucleophilic attack on

the phosphodiester substrate by an activated water molecule without formation of a covalent intermediate ³³. The AcpA structure provides a basis for designing experiments to test this proposed mechanism.

3.3.3 Role of AcpA in virulence of *F. tularensis*

F. tularensis is a facultative intracellular pathogen whose primary target of infection is the macrophage ³⁴. An essential aspect of virulence is the ability of *F. tularensis* to escape phagosomal containment, which leads to over 1000-fold replication of the pathogen in the cytoplasm and eventual apoptosis of the infected macrophage. Several proteins are thought to contribute to intramacrophage growth and survival of *F. tularensis*, including putative transcriptional regulators MglA and MglB ^{35, 36}, phosphatases such as AcpA, and proteins with unknown functions IgIC ^{37, 38} and FTT0918 ³⁹.

A current challenge is to understand the role of AcpA in intramacrophage survival. Our structure-based sequence analysis shows that essential elements of the AcpA active site are shared by PlcH-like PLCs, in particular, the hydroxyl nucleophile and mononuclear metal center. This structural similarity raises the possibility that AcpA exhibits PLC activity and suggests a new hypothesis that AcpA facilitates phagosomal escape by hydrolyzing phospholipids of the phagosomal inner membrane. Interestingly, AcpA efficiently hydrolyzes the phospholipid-like substrate *p*NPPC ¹⁴. But, this is not necessarily an accurate indicator of bona-fide PLC activity because *p*NPPC lacks a hydrocarbon tail. We note that purified AcpA does not exhibit lecithinase activity on egg yolk agar nor

does it lyse red blood cells (data not shown). Thus, additional studies are needed to determine whether AcpA exhibits true PLC activity.

A second hypothesis about the role of AcpA in virulence is that AcpA might affect host signaling pathways by dephosphorylation of host proteins, inositol phosphates or phosphoinositides, the latter being critically important for phagosome formation ⁴⁰ and respiratory burst activation ⁴¹. The wide and relatively flat surface surrounding the active site (Figure 3.1B) is compatible with AcpA docking to a protein substrate.

Finally, it is possible that AcpA functions in a phosphate retrieval system that is activated upon phagosomal containment. AcpA would be an effective phosphate scavenger because of its broad substrate specificity and high abundance. Central to all three hypotheses is the question of whether expression of AcpA is controlled by MglA/B, as has been suggested by Baron and Nano ³⁵. The AcpA structure provides a framework for exploring these hypotheses and for designing AcpA and PlcH inhibitors.

3.4 References

1. 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.
2. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, 276, 307-326.
3. Weeks, C. M.; Miller, R., The design and implementation of SnB version 2.0. *J. Appl. Crystallogr.* **1999**, 32, 120-124.
4. Bricogne, G.; Vonrhein, C.; Flensburg, C.; Schiltz, M.; Paciorek, W., Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr.* **2003**, D59, (Pt 11), 2023-30.
5. Levitt, D. G., A new software routine that automates the fitting of protein X-ray crystallographic electron-density maps. *Acta Crystallogr.* **2001**, D57, (Pt 7), 1013-9.
6. Kleywegt, G. J.; Jones, T. A., Software for handling macromolecular envelopes. *Acta Crystallogr.* **1999**, D55, 941-944.
7. Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **1998**, D54, (Pt 5), 905-921.
8. Morris, R. J.; Perrakis, A.; Lamzin, V. S., ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr.* **2002**, D58, (Pt 6 Pt 2), 968-75.
9. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **2004**, D60, (Pt 12 Pt 1), 2126-32.
10. Winn, M. D.; Murshudov, G. N.; Papiz, M. Z., Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **2003**, 374, 300-21.

11. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000**, 28, 235-242.
12. Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* **1991**, A47, (4), 392-400.
13. 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.
14. 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.
15. Dawson, M. V.; Lyle, S. J., Spectrophotometric determination of iron and cobalt with ferrozine and dithizone. *Talanta* **1990**, 12, 1189-1191.
16. Krissinel, E.; Henrick, K. In *Detection of protein assemblies in crystals*, Computational Life Sciences: First International Symposium, Konstanz, Germany, 2005; Berthold, M. R.; Glen, R.; Diederichs, K.; Kohlbacher, O.; Fischer, I., Eds. Springer: Konstanz, Germany, 2005; pp 163-174.
17. DeLano, W. L. The PyMOL Molecular Graphics System (<http://www.pymol.org>). <http://www.pymol.org>
18. Turro, C.; Fu, P. K.-L.; Bradley, P. M., Lanthanide ions as luminescent probes of proteins and nucleic acids. In *Metal ions in biological systems*, Sigel, A.; Sigel, H., Eds. Marcel Dekker: New York, 2003; Vol. 40, pp 323-353.
19. Coleman, J. E., Structure and mechanism of alkaline phosphatase. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, 21, 441-83.
20. O'Brien, P. J.; Herschlag, D., Alkaline phosphatase revisited: hydrolysis of alkyl phosphates. *Biochemistry* **2002**, 41, (9), 3207-25.
21. Lohse, D. L.; Denu, J. M.; Santoro, N.; Dixon, J. E., Roles of aspartic acid-181 and serine-222 in intermediate formation and hydrolysis of the mammalian protein-tyrosine-phosphatase PTP1. *Biochemistry* **1997**, 36, (15), 4568-75.

22. Lukatela, G.; Krauss, N.; Theis, K.; Selmer, T.; Gieselmann, V.; von Figura, K.; Saenger, W., Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis. *Biochemistry* **1998**, 37, (11), 3654-64.
23. Chruszcz, M.; Laidler, P.; Monkiewicz, M.; Ortlund, E.; Lebioda, L.; Lewinski, K., Crystal structure of a covalent intermediate of endogenous human arylsulfatase A. *J. Inorg. Biochem.* **2003**, 96, (2-3), 386-92.
24. Holm, L.; Sander, C., Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **1993**, 233, 123-138.
25. Galperin, M. Y.; Jedrzejewski, M. J., Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes. *Proteins* **2001**, 45, (4), 318-24.
26. Holtz, K. M.; Stec, B.; Kantrowitz, E. R., A model of the transition state in the alkaline phosphatase reaction. *J. Biol. Chem.* **1999**, 274, (13), 8351-4.
27. Gouet, P.; Courcelle, E.; Stuart, D. I.; Metz, F., ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **1999**, 15, (4), 305-308.
28. 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.
29. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, 26, 283-291.
30. 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.
31. Knowles, J. R., Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **1980**, 49, 877-919.

32. Hendry, P.; Sargeson, A. M., Metal ion promoted reactions of phosphate derivatives. In *Progress in inorganic chemistry: bioinorganic chemistry*, Lippard, S. J., Ed. John Wiley & Sons, Inc.: 1990; Vol. 38, pp 201-258.
33. Martin, S. F.; Hergenrother, P. J., Catalytic cycle of the phosphatidylcholine-preferring phospholipase C from *Bacillus cereus*. Solvent viscosity, deuterium isotope effects, and proton inventory studies. *Biochemistry* **1999**, 38, (14), 4403-8.
34. Oyston, P. C.; Sjostedt, A.; Titball, R. W., Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nat. Rev. Microbiol.* **2004**, 2, (12), 967-78.
35. Baron, G. S.; Nano, F. E., MglA and MglB are required for the intramacrophage growth of *Francisella novicida*. *Mol. Microbiol.* **1998**, 29, (1), 247-59.
36. Lauriano, C. M.; Barker, J. R.; Yoon, S. S.; Nano, F. E.; Arulanandam, B. P.; Hassett, D. J.; Klose, K. E., MglA regulates transcription of virulence factors necessary for *Francisella tularensis* intraamoebae and intramacrophage survival. *Proc. Natl. Acad. Sci. USA* **2004**, 101, (12), 4246-9.
37. Lai, X. H.; Golovliov, I.; Sjostedt, A., Expression of IgIC is necessary for intracellular growth and induction of apoptosis in murine macrophages by *Francisella tularensis*. *Microb. Pathog.* **2004**, 37, (5), 225-30.
38. Santic, M.; Molmeret, M.; Klose, K. E.; Jones, S.; Kwaik, Y. A., The *Francisella tularensis* pathogenicity island protein IgIC and its regulator MglA are essential for modulating phagosome biogenesis and subsequent bacterial escape into the cytoplasm. *Cell. Microbiol.* **2005**, 7, (7), 969-79.
39. Twine, S.; Bystrom, M.; Chen, W.; Forsman, M.; Golovliov, I.; Johansson, A.; Kelly, J.; Lindgren, H.; Svensson, K.; Zingmark, C.; Conlan, W.; Sjostedt, A., A Mutant of *Francisella tularensis* Strain SCHU S4 Lacking the Ability To Express a 58-Kilodalton Protein Is Attenuated for Virulence and Is an Effective Live Vaccine. *Infect. Immun.* **2005**, 73, (12), 8345-52.
40. Simonsen, A.; Wurmser, A. E.; Emr, S. D.; Stenmark, H., The role of phosphoinositides in membrane transport. *Curr. Opin. Cell Biol.* **2001**, 13, (4), 485-92.

41. Baggiolini, M.; Wymann, M. P., Turning on the respiratory burst. *Trends Biochem. Sci.* **1990**, 15, (2), 69-72.

4.

CRYSTALLIZATION OF HISTIDINE ACID PHOSPHATASE

Portions of this chapter have been adapted from:

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. F*. **2006**, F62, 32-35.

4.1 Introduction

As part of our ongoing research on the roles of phosphatases in *Francisella tularensis* intracellular survival and virulence¹⁻⁴, we analyzed the complete but unannotated genome sequence of *F. tularensis* subsp. *holarctica* live vaccine strain (LVS) in search of possible phosphatase genes. Comparison of the open reading frames of the LVS genome with current sequence databases using BLAST⁵ identified a 37.2-kD ortholog of the major acid phosphatase from *Legionella pneumophila* (MAP,⁶). The deduced amino acid sequence of the *F. tularensis* protein is 41% identical (134 out of 330 residues) to that of MAP (Figure 4.1) and exhibited lower, but significant homology to several phosphatases from eukaryotic sources. These enzymes share the conserved RHGXRXP motif (Figure 4.1) that is characteristic of the histidine acid phosphatase (HAP) family⁷.

MAP is thought to be secreted via the type II secretion system and may be involved in intracellular infection⁶. It is not known whether *F. tularensis* HAP (FtHAP) is secreted by an analogous system or whether FtHAP plays a role in virulence or intracellular survival. In parallel with our studies of substrate specificity, kinetics and biological function of this newly discovered enzyme, we have crystallized FtHAP as a first step toward structure determination.

Figure 4.1

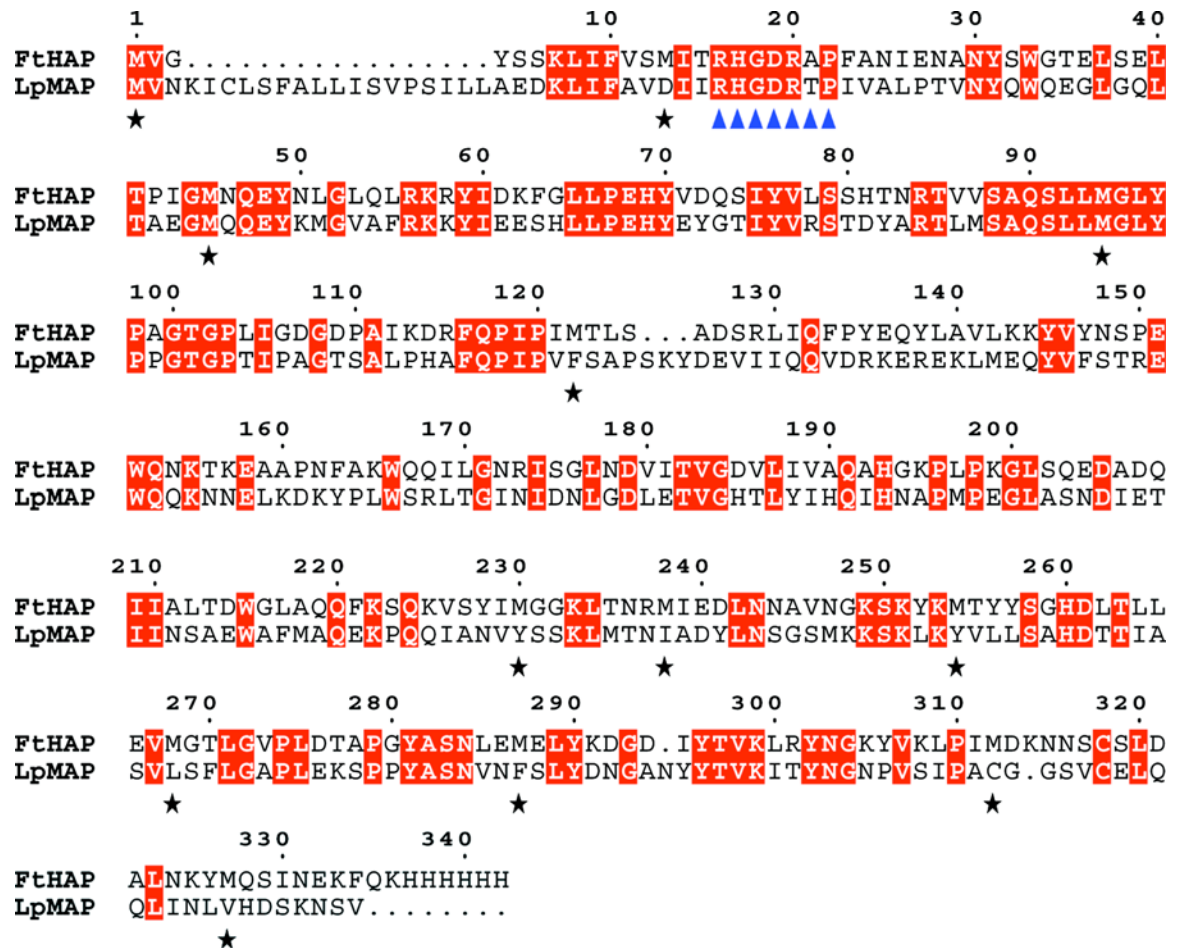


Figure 4.1. Amino acid sequence alignment of *Francisella tularensis* histidine acid phosphatase (FtHAP) and the major acid phosphatase of *Legionella pneumophila* (LpMAP). Identical residues are highlighted in red, the RHGXRXP signature motif of histidine acid phosphatases is denoted by blue triangles, and black stars denote the Met residues of FtHAP.

4.2 Methods and results

4.2.1 Cloning, expression and protein purification

The *hap* gene was amplified by PCR from genomic DNA obtained from the *F. tularensis* subsp. *holarctica* live vaccine strain and cloned into pET20b (Novagen) using *NcoI* and *XhoI* sites. The recombinant protein was expressed in *E. coli* using procedures similar to those described previously for the AcpA phosphatase³. Harvested cells were resuspended in 50 mM sodium acetate pH 6.0 (buffer A) and lysed in a French pressure cell. The resulting mixture exhibited strong phosphatase activity as measured by a discontinuous colorimetric activity assay using *p*-nitrophenylphosphate as the substrate⁸. This assay was also performed after each step in the following purification procedure to identify the enzyme of interest and its relative activity amongst the collected fractions. Sodium chloride was added to the broken cells to a final concentration of 1.0 M and two centrifugation cycles of (1) 27,200 *g* for 20 min and (2) 184,000 *g* for 1.5 hr were performed. The supernatant from the latter centrifugation step was dialyzed against buffer A for 24 hours. The sample was applied to a HiTrap SP HP cation exchange column (Amersham Biosciences) that had been equilibrated with buffer A. A linear NaCl gradient of 0 - 1.0 M over 8 column volumes was used to elute the protein from the column. FtHAP was eluted at 500 mM NaCl. The protein was then dialyzed against 10 mM Na₃PO₄ pH 7.0 (buffer B) for 24 hours. The dialyzed sample was applied to a HiTrap Chelating HP column (Amersham Biosciences) that had been charged with 0.1 M NiSO₄ and

equilibrated with buffer B. A linear imidazole gradient of 0 - 1.0 M was used to elute the protein from the column. FtHAP eluted at 350 mM imidazole. The protein was then dialyzed against buffer A for 24 hours and concentrated to 10 mg / mL. Protein concentration was determined with absorption spectroscopy using an extinction coefficient ($\lambda = 280$ nm) of 48,360 M⁻¹ cm⁻¹ predicted by the ExPASy server⁹. Protein purity was evaluated by SDS-PAGE.

4.2.2 Crystallization

Crystallization trials were performed using the sitting drop method of vapor diffusion at 293 K. Initial screening for crystallization conditions was done using the Index Screen from Hampton Research. Equal volumes of the protein (1.5 mL) and the reservoir (1.5 mL) were mixed and allowed to equilibrate with 1.0 mL of reservoir for 24 - 48 hours. Index Screen reagents 3 - 6, 21, 39, 42, 57, 63, 65, 67 - 69, 71, 72, 75 - 80, 83 - 85, 87, 92 and 94 produced crystals of various size and quality. Condition 63 produced the largest and most well-defined crystals. This condition contained 5% (v / v) Tacsimate, 0.1 M HEPES pH 7.0 and 25% (w / v) PEG 3350. Tacsimate (Hampton Research, HR2-755) is a mixture of organic acids that includes sodium malonate, sodium acetate, tri-ammonium citrate, succinic acid, DL-malic acid and sodium formate (Bob Cudney, personal communication). The initial crystals were diamond shaped with jagged edges, and, in some cases, multiple crystals were fused together (Figure 4.2A). One round of optimization resulted in large, single crystals having sharp edges and dimensions of 0.4 mm x 0.4 mm x 0.7 mm (Figure 4.2B). The final optimized

condition contained 10% (v / v) Tascimate, 0.1 M HEPES pH 7.0 and 19% (w / v) PEG 3350. These crystals would typically appear within 48 hours after setup.

The crystals were cryoprotected in the harvest buffer (10% (v / v) Tascimate, 0.1 M HEPES pH 7.0, 24% PEG 3350) supplemented with 25% PEG 200 as follows. First, 10 – 30 mL of harvest buffer was added to a sitting drop that contained crystals. Next, the liquid in the drop was exchanged for the cryoprotectant in 5 steps over a period of 10 - 20 minutes without moving the crystals. At each step, the concentration of PEG 200 was increased by 5% and the drop was gently mixed by aspiration without disturbing the crystals. Finally, the crystals were picked up with Hampton mounting loops and plunged into liquid nitrogen.

Figure 4.2

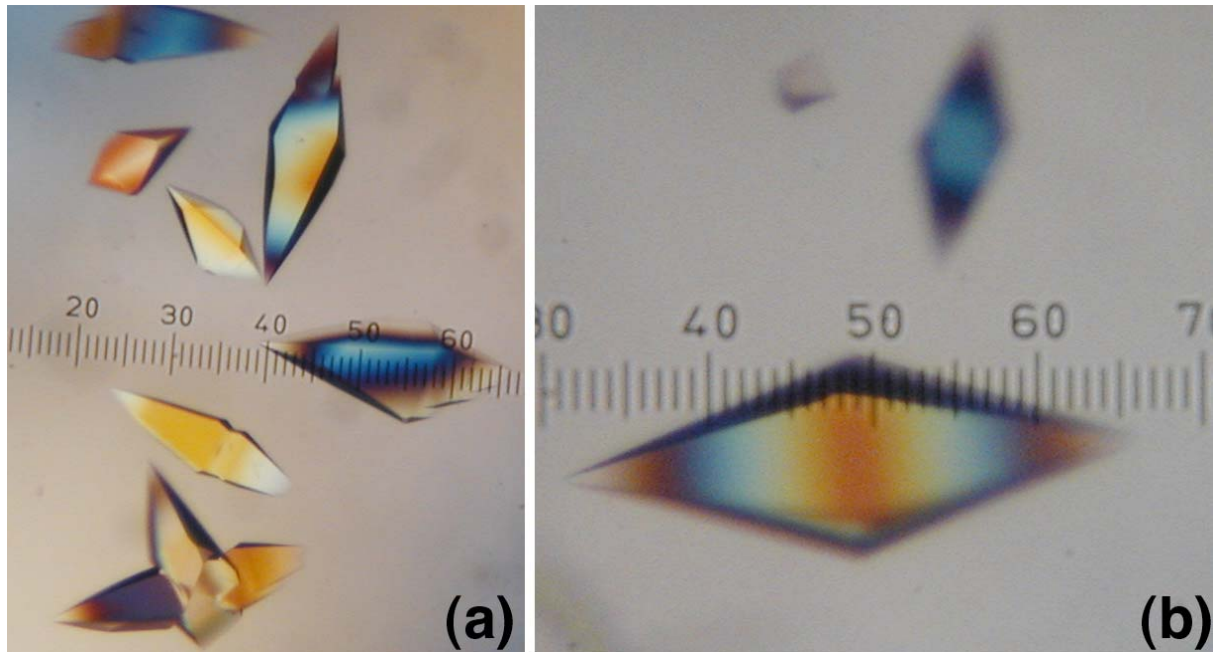


Figure 4.2. Crystals of *F. tularensis* HAP. A) Initial crystals grown from Hampton Research Index Screen reagent 63. These crystals grew to an approximate size of 0.5 mm x 0.2 mm x 0.2 mm within 48 hours. B) Crystals of FtHAP obtained after optimizing the crystallization condition to 10% (v /v) Tacsimate, 0.1 M HEPES pH 7.0 and 19% (w /v) PEG 3350. This crystal has dimensions of 0.4 mm x 0.4 mm x 0.7 mm. The smallest division of the ruler in both panels corresponds to 0.02 mm.

5.2.3 Data Collection and Processing

Initial characterization of X-ray diffraction was done using an R-Axis IV image plate detector coupled to a Rigaku copper rotating anode generator. Auto-indexing of diffraction images using CrystalClear¹⁰ suggested a primitive tetragonal lattice with unit cell dimensions $a = 62.0 \text{ \AA}$, $c = 211.0 \text{ \AA}$. Frozen crystals were transported to Lawrence Berkeley National Laboratory for high-resolution data collection at the Advanced Light Source (ALS).

Diffraction data were collected at ALS beamline 4.2.2 using a NOIR-1 detector with the detector distance and angle set to 150 mm and 17° , respectively. A total of 180° of data was collected using an oscillation angle of 0.5° and exposure time of 8 sec per image. The data were integrated and scaled to 1.75 \AA resolution using d*TREK¹⁰. The refined unit cell dimensions were determined to be $a = 61.96 \text{ \AA}$, $c = 210.78 \text{ \AA}$. Analysis of the data with dtcell¹⁰ suggested space group $P4_12_12$. Matthews calculations suggested that this crystal form has 1 molecule in the asymmetric unit, 53 % solvent and Matthews coefficient of $2.7 \text{ \AA}^3/\text{Da}^{11}$. See Table 4.1 for data processing statistics.

Since the apparent Laue class is $4/mmm$, the possibility of merohedral twinning was considered. A plot of the cumulative intensity distribution for acentric reflections did not display the sigmoidal shape that is characteristic of twinned data¹². The average value of $\langle I^2(h) \rangle / \langle I(h) \rangle^2$ was 2.2 for acentric reflections, which is close to the value of 2.0 that is expected for untwinned data¹³. For reference, a value of 1.5 is expected in the case of perfect hemihedral

twinning¹³. Based on these results, we do not anticipate difficulties due to twinning during structure determination.

The FtHAP sequence was compared to sequences in Protein Data Bank (PDB,¹⁴) to identify a search model for molecular replacement calculations. The closest relative in the PDB is human prostatic acid phosphatase (1CVI,¹⁵), which shares 28% global amino acid sequence identity with FtHAP. Molecular replacement trials were performed with MOLREP¹⁶ using 1CVI as the search model. All possible space groups with Laue symmetry 4/mmm were tested. The top solution had $R > 0.6$ and correlation coefficient < 0.12 , which indicated that molecular replacement is not a suitable phasing method. Structure determination using crystals of selenomethionyl FtHAP is in progress. Twelve Met residues are expected in the asymmetric unit (Figure 4.1).

Table 4.1.Data collection and processing statistics^a

Space Group	P4 ₁ 2 ₁ 2
Unit Cell (Å)	a = 61.96, c = 210.78
Wavelength (Å)	1.12718
Resolution Range (Å)	46.47 - 1.75 (1.81 - 1.75)
Total Reflections	243916
Unique Reflections	41088
Redundancy	5.94 (4.72)
Mosaicity (°)	0.466
Rmerge	0.041 (0.299)
Completeness (%)	96.2 (91.7)
Average I/σ(I)	25.4 (5.2)

^aValues for the outer resolution shell of data are given in parentheses.

4.3 References

1. Baron, G. S.; Reilly, T. J.; Nano, F. E., The respiratory burst-inhibiting acid phosphatase AcpA is not essential for the intramacrophage growth or virulence of *Francisella novicida*. *FEMS Microbiol. Lett.* **1999**, 176, (1), 85-90.
2. 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.
3. Reilly, T. J.; Felts, R. L.; Henzl, M. T.; Calcutt, M. J.; Tanner, J. J., Characterization of recombinant iacid phosphatase. *Prot. Exp. Purif.* **2005**, In press.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., *Protein Identification and Analysis Tools on ExPASy Server*. Humana Press: 2005; p 571 - 607.

10. Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Crystallogr.* **1999**, D55, 1718-1725.
11. Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968**, 33, 491-497.
12. Rees, D. C., The influence of twinning by merohedry on intensity statistics. *Acta Crystallogr.* **1980**, A36, 578-581.
13. Redinbo, M. R.; Yeates, T. O., Structure determination of platocyanin from a specimen with a hemihedral twinning fraction of one-half. *Acta Crystallogr.* **1993**, D49, 375-380.
14. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000**, 28, 235-242.
15. Jakob, C. G.; Lewinski, K.; Kuciel, R.; Ostrowski, W.; Lebioda, L., Crystal structure of human prostatic acid phosphatase. *Prostate* **2000**, 42, (3), 211-8.
16. Vagin, A.; Teplyakov, A., MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **1997**, 30, 1022-1025.

5.

STRUCTURE OF HISTIDINE ACID PHOSPHATASE

5.1 Introduction

Francisella tularensis is a highly infectious intracellular bacterial pathogen and is the causative agent for the disease tularemia. The CDC (Centers for Disease Control and Prevention) has assigned *F. tularensis* to be a Category A pathogen due to its potential for use as a bioterrorism agent¹. The bacteria can be isolated from numerous mammalian hosts and arthropod vectors, readily grown in broth culture, mechanically aerosolized, and it takes less than ten organisms to establish infection^{2, 3}. Pneumonic tularemia is acquired via inhalation of *F. tularensis* and is the most serious infectious concern due to its high fatality rate, >30%, if left untreated². Understanding the biochemistry and biology of *F. tularensis* virulence is necessary for the development of improved vaccines and antimicrobial agents that will provide protection and treatment in the event of a *F. tularensis*-based attack.

Histidine acid phosphatase (HAP) from *F. tularensis* is a 38-kDa polyspecific, periplasmic acid phosphatase that contains the characteristic RHGXRXP sequence motif that is present in most members of the histidine acid phosphatase family⁹. Through sequence analysis it was determined that HAP shares high similarity (41%) to MAP, the major acid phosphatase from *Legionella pneumophila*¹⁰. MAP has been shown to be extracellularly secreted via the type II secretion system and plays a role in the virulence of *L. pneumophila*. Further sequence analysis showed lower yet significant homology to the prostatic acid phosphatases from both human and rat sources.

Human prostatic acid phosphatase (HPAP) is the most well studied member of the histidine acid phosphatase family^{9, 11-19}. HPAP is a major protein secreted by epithelial cells of the prostate and exhibits an ability to hydrolyze a wide range of alkyl and aryl phosphate monoesters¹³. The reaction mechanism for this enzyme has been well characterized including site directed mutagenesis of active site residues and secondary isotope effect studies. Breakdown of the phosphohistidine intermediate via attack of a nucleophilic water molecule to form inorganic phosphate has been shown to be the rate-limiting step of the reaction. This reaction is inhibited by the addition of L-tartrate in sub millimolar amounts. The crystal structure of both the human and rat PAP have been solved previously^{12, 20, 21}.

As part of our ongoing studies to understand the pathogenesis of *F. tularensis* we have initiated structural studies as a prelude to the full biochemical characterization of HAP. Here we report the three dimensional structures of HAP in complex with acetate, orthovanadate, L-tartrate, and (2*R*, 3*R*)-tartranilate.

5.2 Materials and Methods

5.2.1 Enzyme Assay

ACP activity was measured using a discontinuous assay with *p*-nitrophenyl phosphate (*p*NPP) as the substrate. The product *p*-nitrophenolate (*p*NP) was monitored spectrophotometrically at 405 nm using a Shimadzu BioSpec-mini spectrometer. The extinction coefficient for *p*NP used in the calculations was $1.78 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$. Linearity of the activity was measured using a 600 μL assay mixture containing of 0.4 M sodium acetate buffer, pH 5.0, 10 mM *p*NPP and various amounts of enzyme. The mixtures were incubated at 37 °C for 15 minutes and 300 μL of 0.5 M glycine pH 10.0 was added to quench the reaction. Using these conditions the enzymatic activity was linear with the amount of enzyme added. Enzymatic activity was also linear over time for at least 60 minutes (Figure 5.1A). Only the initial rates (slope over the first 15 minutes) were used for calculation of enzyme activity and associated kinetic parameters. One unit of activity is described as the amount of enzyme required to convert 1 nmole of substrate to product per minute. Determination of the Michaelis-Menton constants for *p*NPP was performed using 5.5×10^{-5} mg of HAP and a wide range ($K_m/10$ to $5 K_m$) of substrate concentrations. The K_m and K_i values reported in the results section represent averages of multiple trials performed for each substrate or inhibitor. Data were analyzed using Origin (Microcal).

Figure 5.1

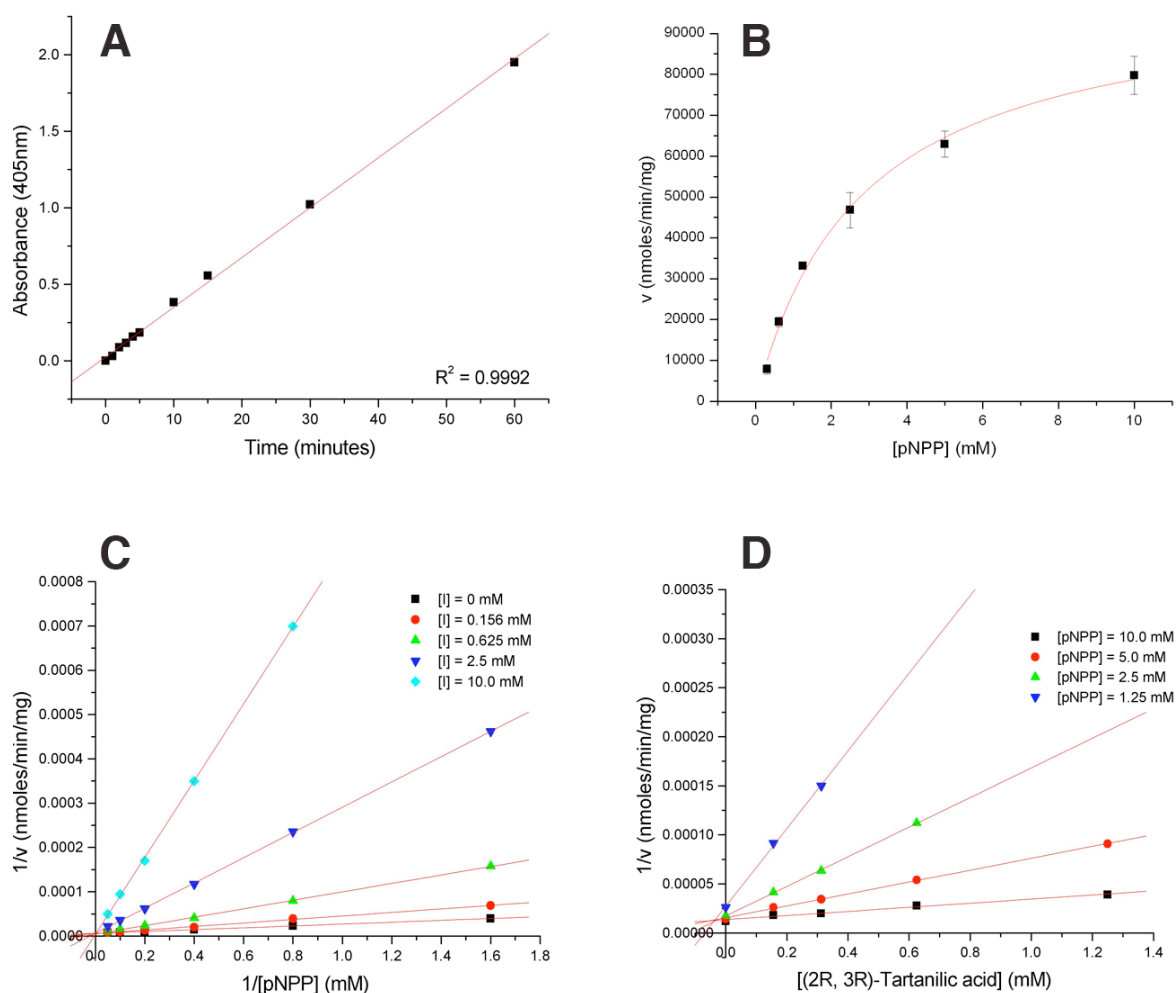


Figure 5.1 Kinetic analysis of HAP.

A) This plot represents the linearity of the phosphatase assay used to determine the kinetic parameters for HAP. **B)** Activity is plotted against the substrate concentration and fit to the Michaelis-Menton equation in order to estimate the K_m and V_{max} values for HAP using pNPP as the substrate. **C)** This Lineweaver-Burk plot indicates that (2R, 3R)-tartranilic acid is a competitive inhibitor of HAP. Note that V_{max} remains constant while the K_m changes with

increasing amounts of substrate. **D)** The Dixon plot was used to calculate the K_i for (2*R*, 3*R*)-tartranilic acid.

5.2.2 Expression of HAP

A single colony of *Escherichia coli* BL21 (DE3) cells containing the plasmid pET20B harboring the *hap* gene was used to inoculate 5 mL LB containing ampicillin (50 µg/mL) and incubated overnight at 37 °C with constant aeration. A 1:1000 dilution of this overnight culture was then used to inoculate 25 mL of fresh media. This sample was incubated as before until the density of the culture reached an OD A_{600} of 0.6. The sample was then chilled on ice for 10 minutes then placed at 4 °C for overnight storage. To remove the secreted β -lactamases, the *E. coli* starter culture was centrifuged at 3,600 *g* for 10 minutes at 4 °C and suspended in fresh LB prior to distribution into four 1.8 L flasks containing 500 mL LB supplemented with ampicillin 50 µg/mL and 0.2% (w/v) filter-sterilized glucose. The culture was grown with constant aeration to an OD A_{600} of 0.4 at which time IPTG was added to a final concentration of 0.4 mM. To determine time of maximum induction as determined by specific activity, 1 mL aliquots were removed at regular intervals from each of the four flasks, pelleted, and suspended in 50 mM acetate, pH 5.0, containing 150 mM NaCl prior to quantitation of enzyme activity and protein concentration. Three hours was found to be optimum time for expression and used as a stopping point in subsequent inductions.

Selenomethionine (Se-Met) incorporated enzyme was overexpressed in *E. coli* BL21 (DE3) cells containing the same plasmid and gene as above using the Overnight Express Autoinduction System 2 (Novagen). Once the expression was

complete the Se-Met protein was purified in the same way as the wild type enzyme with the addition of 5 mM DTT to all of the buffers.

5.2.3 Purification of HAP

The cells containing recombinant HAP were pelleted from the bulk media and resuspended in a minimal amount of buffer containing 50 mM sodium acetate pH 5.0 (buffer A). The resuspended cells were subjected to two cycles of French Press (Aminco) adjusted to 10,000 psi in a rapid fill cell and a flow rate of 20 drops/min. All subsequent purification steps were conducted at 4° C. The French Press treated cell extract was stirred in the presence of 1 M NaCl for 60 minutes to dissociate peripherally associated proteins. Unbroken cells and pelletable debris were removed by centrifugation at 27,200 *g* for 15 minutes. Bacterial membranes were pelleted using ultracentrifugation at 184,000 *g* for 60 minutes.

The supernatant from the ultracentrifugation containing the bulk of phosphatase activity, as measured using a discontinuous colorimetric assay described above, was dialyzed in 2L of 20 mM sodium phosphate, pH 7.4 (buffer B) for 24 hours. The dialyzed supernatant was applied to a HiTrap Chelating HP column (Amersham Biosciences) that had been charged with 0.1 M NiSO₄ and equilibrated with buffer B. The elution buffer containing 20 mM sodium phosphate, pH 7.4 and 1 M imidazole was run as a linear gradient of 0 – 1 M imidazole. The enzyme eluted at 350 mM imidazole. The protein solution was then dialyzed against buffer A 24 hours prior to concentration. The enzyme was

concentrated to 10 mg/mL using Amicon Ultra-15 centrifugal filter devices (Millipore) with a 10,000 mw cutoff. Concentration was determined using absorption spectroscopy with an extinction coefficient ($\lambda = 280$ nm) of 48,360 M⁻¹cm⁻¹ ²². Enzyme purity was evaluated using SDS-PAGE.

5.2.4 Crystallization of HAP

Purified Se-Met enzyme was crystallized using sitting drop vapor diffusion in the same condition as previously reported¹¹: 10% (w/v) Tacsimate, 0.1 M HEPES pH 7.0, 5 mM DTT, and 19% (w/v) PEG 3350. In the crystallization tray 1.5 μ L of purified enzyme was mixed with 1.5 μ L of the reservoir solution that had a total volume of 1 mL. The resulting crystals were morphologically similar to the original crystals in size and shape having unit cell dimensions of $a = 61.93$ Å and $c = 210.91$ Å occupying space group P4₁2₁2. The Se-Met crystals were cryoprotected in the same way as the original crystals. When the structure was solved it was discovered that the competitive inhibitor L-tartrate was bound in the active site, thus this crystal form will hereby be referred to as the tartrate bound form.

Crystals of ligand-free HAP were obtained. As previously mentioned, HAP crystals grew in several conditions of the Index screen. After assessing the size and preliminary diffraction of many of the crystals in the Index screen, conditions 3-6 were chosen to grow ligand-free crystals. After optimization large well diffracting crystals were grown in a range of conditions consisting of 0.05 - 0.2 M bis-tris, pH 5.0, 1.6 - 2.0 M ammonium sulfate and typically appeared

between 3 days and 3 weeks. To cryoprotect these crystals ethylene glycol (0% - 35%) was added stepwise to each of the conditions that produced crystals. From the best crystals a 1.85 Å data set was collected. The unit cell dimensions were $a = 62.4$ Å and $c = 209.8$ Å and the space group was $P4_12_12$. The data collection and processing statistics for this crystal form as well as all of the others can be found in Table 5.1.

The ease of crystal growth using the Index screen allowed for parallel crystallization experiments to be conducted using the competitive inhibitors sodium orthovanadate and (2*R*,3*R*)-tartranilic acid. It is widely known that orthovanadate is a phosphate analog that is a common inhibitor of many phosphatases. Having the structure of HAP with orthovanadate bound will give insight to the catalytic mechanism, as well as binding interaction of the phosphate moiety. For the co-crystallization experiment a final concentration of 10 mM activated orthovanadate was added to purified HAP (10 mg/mL) and set up in the Index screen as previously described for the tartrate-bound structure. The inhibitor solution was activated by raising the pH to 10.0 followed by 2-4 cycles of heating; this procedure prevented the formation of vanadate polymers. After a few days of incubation at 293 K, crystals appeared in many of the same conditions as the previous forms, with condition 30 chosen due to the large size and good diffraction of the initial crystals. The optimized condition was 0.1 M NaCl, 0.1 M Bis-Tris, pH 6.5, and 1.5 M ammonium sulfate. This condition is very similar to that used for the ligand-free HAP crystals except for the addition of NaCl and a slight change in pH. Cryoprotection of these crystals was also done with a stepwise addition of ethylene glycol up to 35%. These crystals exhibited

diffraction to 1.65 Å and had unit cell dimensions of $a = 61.8\text{Å}$ and $c = 209.3\text{Å}$ with space group $P4_12_12$.

Co-crystals of HAP with the competitive inhibitor (*2R, 3R*)-tartranilic acid were obtained. This inhibitor was chosen for its likeness to L-tartrate and for its aniline group. This crystal form was grown in Index #6 and diffracted so well that further optimization was unnecessary. A final concentration of 10 mM (*2R, 3R*)-tartranilic acid was added to purified HAP (10 mg/mL) and set up in the tray as described earlier. The final condition was 0.1 M tris, pH 8.5 and 2.0 M ammonium sulfate. These crystals were similar in morphology to all of the previous crystal forms, exhibiting diffraction to 2.0 Å and having unit cell dimensions of $a = 61.9\text{Å}$ and $c = 210.5\text{Å}$ with space group $P4_12_12$.

5.2.5 Data Collection

Data were collected at the Molecular Biology Consortium (MBC) beamline 4.2.2 at the Advanced Light Source at Lawrence Berkeley National Laboratory. The data were collected on a Noir 1 CCD detector at a variety of wavelengths. For the Se-met crystals the wavelength was set to 0.97915 Å, at this wavelength the maximum anomalous signal for Se could be achieved. The data collection consisted of 360 frames with an oscillation width of 0.5° per image. The crystal to detector distance was 150 mm and the exposure time was 3.5 seconds. All HAP X-ray data were processed with the program d*TREK. The refined unit cell dimensions were $a = 61.93\text{Å}$ and $c = 210.91\text{Å}$. Using the unit cell dimensions, the molecular weight and the space group a Matthews calculation was done to

determine the approximate solvent content (52.8%) and the number of molecules in the asymmetric unit (1)²³. This calculation gave consistent results for each of the four different crystal forms. The Se-Met data were collected to 2.10 Å for the purpose of collecting highly redundant anomalous data. The tartaric acid bound crystals exhibited diffraction to 1.75 Å. The data collection and processing statistics for all 5 data sets can be found in Table 5.1.

The ligand-free HAP crystals were collected at a wavelength of 1.12711 Å with a total of 360 images, 0.5° oscillation angle, detector distance of 160 mm, 2 θ angle of 10°, and exposure length of 10 seconds per frame. HAP crystals of all forms discussed here have a long unit cell dimension of around 210 Å. In order to collect the highest resolution data possible at 4.2.2. the 2 θ angle was increased from 0° to as much as 10°.

The data collection protocol for the co-crystallized HAP with orthovanadate was similar to the previous crystal forms. These crystals were collected at a wavelength of 1.12711 Å. The data collection consisted of 360 images with 0.5° oscillation, 160 mm crystal to detector distance, 10 second exposure, and a 2 θ angle of 10°. These crystals exhibited diffraction to 1.65 Å and had refined unit cell dimensions of $a = 61.78$ Å and $c = 209.29$ Å.

The final crystal form used in this research was the co-crystal of HAP and the competitive inhibitor (2*R*, 3*R*)-tartranilic acid. Data for these crystals were collected at 1.000 Å using the method of fine slicing²⁴. This protocol calls for taking more images, 900 in this case, at very low oscillation angle, 0.2°. The crystal to detector distance was 150 mm, the exposure time was 1 second, and the 2 θ angle was set to zero. The fine slicing method was used because of the long

unit cell dimension and relatively close detector distance. These crystals exhibited diffraction to greater than 1.9 Å, but a high quality data set was truncated at 2.0 Å. The refined unit cell dimensions for this crystal form were $a = 61.90$ Å and $c = 210.49$ Å.

5.2.6 Structure Determination and Refinement

Initial phases were determined using single-wavelength anomalous diffraction with the program SOLVE²⁵. SOLVE was able to locate 11 of the 12 selenium sites in the crystal at a resolution to 2.1 Å with a resulting figure of merit equal to 0.31. RESOLVE²⁵ was then used for density modification and model building. The phases from RESOLVE had a figure of merit equal to 0.65, and the tracing algorithm produced 63% of the 331 residues built and assigned to sequence. The remaining residues were built in manually using the program COOT²⁶. The Se-Met derived model was refined to 2.1 Å using REFMAC5²⁷ with a $R_{cryst} = 0.237$ and an $R_{free} = 0.261$. This model was then phase extended to 1.75 Å using the high-resolution tartaric acid data. After refinement the resulting R_{cryst} and R_{free} are 0.193 and 0.219 respectively. Solvent molecules were added using the water picking option of COOT and verified manually. An L-tartrate ion was manually fit into the corresponding $F_o - F_c$ density located in the active site. Atomic coordinates have been deposited in the Protein Data Bank²⁸ (PDB) with the accession code of 2GLA.

The solution of the three structures that followed was done using similar protocols maintaining the same test set of reflections. The tartrate-bound

structure with the solvent and ligand molecules removed was used as the initial model. Refmac5 was used for refinement of the structure to the newly collected data. For each data set the first step was rigid body refinement of the model to 3.0Å. The resulting structure would then be used for a second round of rigid body refinement to the maximum resolution of the data set. After the second round of rigid body refinement the resulting structure was subjected to restrained refinement with TLS. Model building was then done in COOT. The refinement statistics, and PDB codes for each of these structures can be found in Table 5.1. Ramachandran plots were made to assess main-chain geometry (Figure 5.10A-D).

Table 5.1 Data collection, phasing and refinement statistics ^a

	Se-MET	2GLA	2GLB	2GLC	2P36
Data collection					
Active Site Ligand	L-tartrate	L-tartrate	None	Vanadate	L-tartranilate
Space group	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2
Cell dimensions					
<i>a</i> , <i>c</i> (Å)	61.93, 210.9	61.96, 210.8	62.41, 209.98	61.78, 209.29	61.93, 210.46
Wavelength (Å)	0.979091	1.12711	1.12711	1.20372	1.0000
Resolution (Å)	43.79 - 2.10	46.47 - 1.75	40.17 - 1.85	59.25 - 1.65	46.43 - 1.85
	(2.18 - 2.10)	(1.81 - 1.75)	(1.92 - 1.85)	(1.17 - 1.65)	(1.92 - 1.85)
No. of Observations	338002	243916	232644	291325	416643
No. of Unique reflections	25062	41088	36481	46663	36027
<i>R</i> _{merge}	0.094 (0.338)	0.041 (0.299)	0.072 (0.285)	0.065 (0.476)	0.104 (0.474)
<i>I</i> / σ _{<i>I</i>}	15.2 (7.3)	25.4 (5.2)	14.9 (4.0)	11.9 (2.6)	9.7 (2.1)
Completeness (%)	100.0 (100.0)	96.2 (91.7)	99.5 (99.5)	93.2 (83.7)	99.8 (99.7)
Redundancy	13.49 (14.38)	5.94 (4.72)	6.38 (5.26)	6.24 (5.58)	11.56 (97.15)
Refinement					
Resolution (Å)		46.47 - 1.75	40.17 - 1.85	59.25 - 1.65	46.43-2.00
No. Reflections		38900	34544	44204	28648
<i>R</i> _{work}		0.175 (0.214)	0.191 (0.221)	0.195 (0.326)	0.202 (0.273)
<i>R</i> _{free} ^b		0.196 (0.275)	0.220 (0.295)	0.219 (0.383)	0.233 (0.414)
No. Residues		329	327	327	327
No. Atoms					
Protein		2626	2549	2559	2569
Water		279	230	271	247
Acetate ions		1	6	0	1
HVO ₄ ²⁻ ions		0	0	2	0
SO ₄ ²⁻ ions		0	0	0	1
L-tartrate ions		1	0	0	0
L-tartranilate		0	0	0	1
PEG Fragments		1	0	0	0
<i>B</i> -factors (Å ²)					
Protein		31.6	36.0	37.8	32.7
Water		35.9	40.4	42.1	33.5
Acetate ions		45.1	41.1	NA	51.8
HVO ₄ ²⁻ ions		NA	NA	43.2	NA
SO ₄ ²⁻ ions					58.13
L-tartrate ions		20.0	NA	NA	NA
L-tartranilate		NA	NA	NA	26.9
PEG Fragments		47.6	NA	NA	NA
RMSD ^c					
Bond lengths (Å)		0.010	0.011	0.009	0.014
Bond angles (°)		1.227	1.191	1.151	1.366
Ramachandran					
Favored (%)		98.5 ^d	99.1 ^d	99.1 ^d	92.6 ^e
Allowed (%)		1.5	0.9	0.9	7.0

^aValues for the outer resolution shell are in parenthesis.

^b5% random test set.

^cCompared to Engh and Huber force field²⁹.

^dThe Ramachandran polts for 2GLA-C were generated using RAMPAGE³⁰.

^eThe Ramachandran plots for 2P36 was generated using PROCHECK³¹.

5.3 Results and Discussion

The asymmetric unit includes 326 amino acid residues belonging to one HAP molecule. The following sections of the polypeptide chain are disordered: 1-5, 106-107, and 335-340. The four structures have similar numbers of water molecules. The exact number of waters for each structure can be found in Table 5.1. Other solvent molecules modeled in the electron density include 6 acetate ions in the ligand-free structure, 1 acetate ion and 3 sulfate ions in the (*2R*, *3R*)-tartranilate structure, and one PEG fragment in the tartrate bound structure. Inhibitors built into density for the appropriate structures include orthovanadate ions (2), L-tartrate (1), and (*2R*, *3R*)-tartranilate (1). The four structures have an RMSD range of 0.17 to 0.30 Å, indicating that they have nearly identical conformations. The important differences between the four structures are located within the active site.

5.3.1 Overall structure of HAP

The structure of HAP comprises two distinct domains and has approximate overall dimensions of 60 x 32 x 30Å. The core domain is a twisted, 5-stranded β -sheet flanked by two α -helices on one side and three on the other (Figure 5.2A). The strand order of the β -sheet is 5, 4, 1, 3, 2 with all but strand 4 in parallel (Figure 5.2A). The second domain is made entirely of α -helices and is located directly above the core domain. This domain comprises residues 23-39 and 134-232 and features six α -helices labeled C-G, I in Figure 5.2A. When the

solvent is removed and the structure is inspected as a surface model, there is a solvent channel running directly between the two domains near the active site (Figure 5.2B). There are 15 waters molecules bound in the channel.

According to gel-filtration experiments (data not shown) purified HAP forms an apparent dimer. When the crystallographic symmetry operator $Y, X, -Z+1$ is applied to the HAP structure a dimer is formed. Based on analysis, using the program PISA³², this dimer has the largest buried surface area of any two proteins in the crystal lattice. According to PISA this area is 1,560 Å² while the largest interface only buries 415 Å². This interface also had the highest complexation significance score ($CSS=1$), compared to $CSS = 0$ for all other possible interfaces. It is therefore concluded that this dimer represents HAP in solution.

Figure 5.2

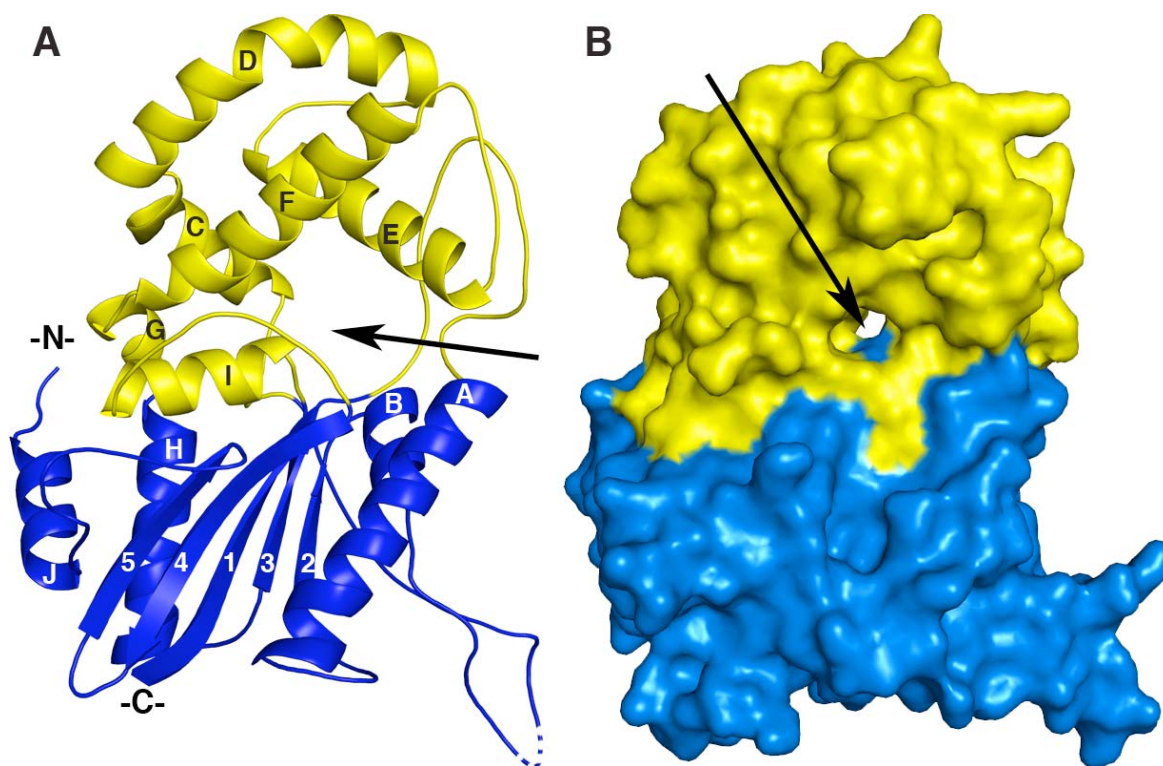


Figure 5.2 Overall structure of HAP.

A) A ribbon drawing of a HAP subunit. The upper helical domain, consisting of α helices C, D, E, F, G, and I are colored in yellow. The lower core domain, consisting of a mixed β sheet flanked by α helices is colored in blue and labeled accordingly. The active site is located in a cleft between the two domains. A chain break from residue 106-108 is indicated by dashes. **B)** A surface rendering of HAP. The solvent has been removed revealing a hole that leads from the active site. In this orientation the entrance to the active site would be directly across from this hole between the two colored domains. The black arrows indicate the position of the active site.

The dimer interface is formed from α -helix B (residues 83-98) in the core domain as well as the large loop region consisting of residues 99-125 which connects α B and α C (Figure 5.3). Dimer interactions are, besides a number of hydrophobic contacts, hydrogen bonds between side chain atoms of Gln90 and its symmetry mate and between Lys113 and the carbonyl of Ala99. The side chains of residue Phe116 are stacked on top of each other across the two-fold axis relating the two subunits. This type of interaction is consistent with what has been seen with the rat and human prostatic acid phosphatases^{17, 20}. The $2F_o - F_c$ density is rather weak for the aromatic ring of Phe116 indicating that there is some disruption within the crystal in this area. However, when inspecting the difference density there is a large amount of positive density that corresponds to an aromatic ring. When the side chain is built in there is some residual negative difference density and incomplete $2F_o - F_c$ density. The phenylalanine side chain looks to be clashing with another residue in the dimer interface, Met94.

Met94 looks to be important for dimerization. In all of the HAP structures discussed here Met94 appears to adopt two conformations. The anomalous difference Fourier map calculated from the Se-Met data showed positive density indicating two positions for the Se atom. The methyl group of the methionine of conformation B clashes with the aromatic ring of Phe116 when both side chains are present. To alleviate any unfavorable interaction within the structure the side chain of Phe116 was deleted. This interaction will require further investigation.

Figure 5.3

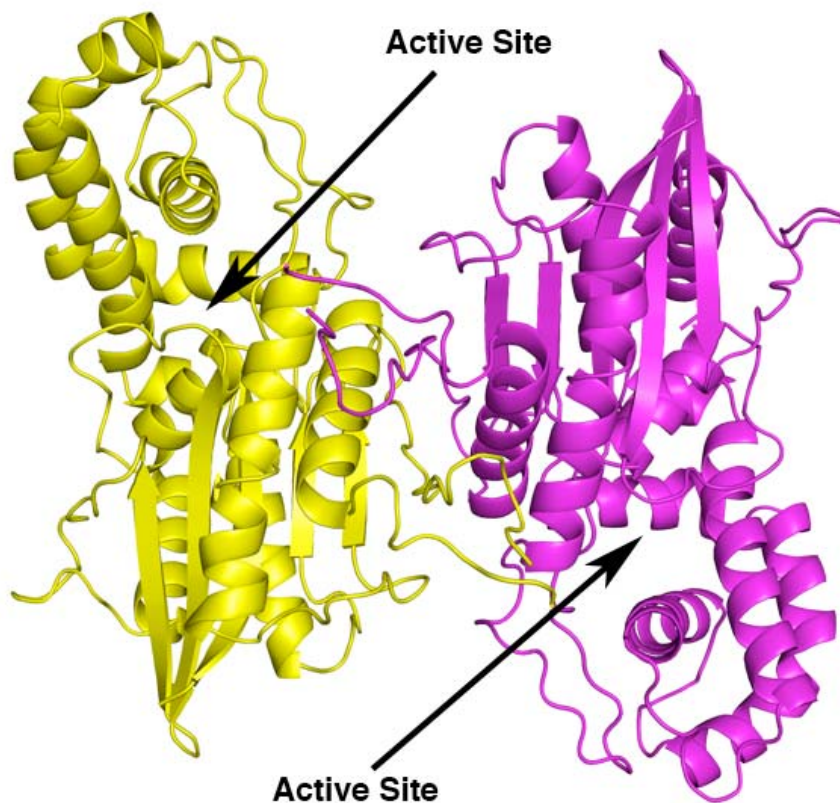


Figure 5.3 HAP dimer.

A ribbon drawing of the HAP dimer. This dimer was generated by applying the symmetry operator $Y, X, -Z+1$ to the one molecule in the asymmetric unit. The active sites are labeled for each subunit.

Figure 5.4

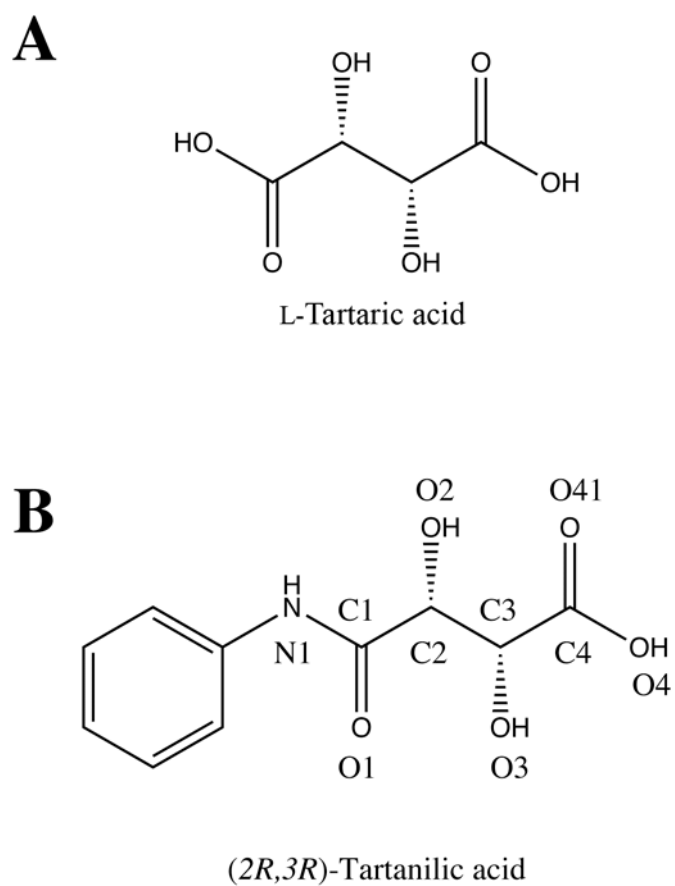


Figure 5.4 Inhibitors of HAP.

A) Drawing of L-tartaric acid indicating its overall shape and stereochemistry. **B)** Drawing of the competitive inhibitor (2*R*,3*R*)-tartranilic acid.

5.3.2. Structure of HAP/vanadate complex

The active site of the vanadate bound HAP structure is comprised of seven residues and six water molecules (Figure 5.5A). Of the seven amino acid residues only histidine 17 is in the position for nucleophilic attack. The uncharged nitrogen of the imidazole ring of histidine is approximately 3.8 Å from the vanadium atom of the bound orthovanadate. Of the six water molecules in the active site four coordinate to the bound vanadate. Two that are of most interest are also bound to aspartate 261. This residue is in the correct location to play the role of general acid/general base. When compared to the mammalian prostatic acid phosphatases this residue is in the same location as their general acid/base that is also an aspartate.

The other residues in the active site that interact with the bound vanadate include arginines 16, 20, and 84, glutamine 132, histidine 260, and aspartate 261. As with the aspartate 261 that was discussed previously, Gln132 interacts with the vanadate through water mediated hydrogen bond.

When inspecting the electron density in and around the active site there was a large feature that could not be attributed to the polypeptide or solvent near Arg84. This density was first modeled as water. However the difference density after refinement indicated that it was not a large enough molecule to occupy that location. The electron density was then modeled as vanadate and refined. There was no residual difference density after refinement indicating that vanadate was the correct molecule for that density.

The guanidinium group of arginine 84 shares two hydrogen bonds with two oxygen atoms of the second vanadate ion. These interactions stabilize the vanadate near the entrance of the active site.

5.3.3. Structure of HAP/tartrate complex

When the active site was located there was unidentified electron density that could not be built in as part of the protein or as solvent. To determine what molecule would fit into this electron density we investigated the components in the crystallization reagent Tacsimate. This reagent is a mixture of organic acids including, malate, citrate, succinate, malonate, formate, acetate, and tartrate. When trying to fit each one of these molecules into the electron density only one fit. The molecule that fit best into the electron density was L-tartrate. This dicarboxylic acid is a well-known inhibitor of histidine acid phosphatase. The stereochemistry is such that only the L isomer of tartrate would fit the electron density. L-tartrate has 13 interactions with residues in the active site (Figure 5.5B). The nucleophile, His17 has three interactions with the inhibitor. These include a hydrogen bond to the hydroxyl group in position O2, and non-bonded interactions with the carboxylate in positions O4 and O41. The amine of Gln132 and the guanidinium of Arg84 hydrogen bond to the O1 oxygen of the inhibitor. A hydrogen bond between the NH of the imidazole group of His260 is made to the hydroxyl at position O2. Asp262 has two interactions with the inhibitor. The first is a hydrogen bond between the backbone amine and oxygen O4. The second interaction is a hydrogen bond between its carboxylate and the inhibitors

hydroxyl in position O3. The guanidinium of Arg16 forms two hydrogen bonds with the O4 oxygen of the inhibitor. The guanidinium of Arg20 forms two hydrogen bonds with the O41 oxygen of the inhibitor.

5.3.4. Structure of HAP/acetate complex

In the ligand free HAP structure there are four water molecules and two acetate ions near the entrance of the active site (Figure 5.5C). The acetate ions were modeled into Y-shaped electron density that would not refine correctly with water. When water was built in and refined the bond lengths between the two water molecules was less than that of a hydrogen bond and there was residual difference density indicating that there needed to be more atoms than just the two water molecules.

When refinement with acetate ions was done there was no residual difference density visible. The carboxylate group of one of the acetate ions accepts hydrogen bonds from the guanidinium of Arg20. The second acetate ion forms hydrogen bonds with guanidinium of Arg84. The guanidinium of Arg84 also forms a hydrogen bond with the hydroxyl of Ser38.

When inspecting the three HAP structures it was observed that upon binding of tartrate or vanadate Arg20 has a 1.77 Å shift into the active site relative to the HAP/acetate structure. This shift allows the guanidinium of the arginine to form hydrogen bonds with the vanadate ion or to the O41 oxygen atom of the L-tartrate ion. Other interactions that this residue has in the acetate complex

include an ion pair with Asp185 and a hydrogen bond with the hydroxyl of tyrosine 281.

Another residue that has different conformations between the three structures is Phe23. In the HAP/acetate structure the side chain is rotated out of the active site. In the vanadate structure the Phe23 side chain is disordered.

In the tartrate bound structure, the side chain of Phe23 is well ordered, but rotated 180° into the active site. This stabilization of Phe23 is most likely due to an interaction with a HEPES molecule. When building the tartrate bound structure some unidentified density appeared in a cleft between Phe23 and Tyr135. This density did not represent the protein and was too large for water. In the initial crystallization condition for the tartrate bound structure one of the components was HEPES. A HEPES molecule was built into the density but it had very high B-factors and was not kept for the final structure.

In concert with the movement of Phe23, Ser38 was also observed to move when a ligand was bound in the active site. In the ligand free and the vanadate bound structures the hydroxyl group is rotated in towards the active site. In contrast when L-tartrate is bound the hydroxyl group is rotated away from the active site. In the ligand free structure this serine forms a hydrogen bond with an acetate ion that is also interacting with Arg84 as described above.

In the vanadate bound HAP structure Gln132 was built in multiple conformations. In one conformation the amino group is hydrogen bonding to a water molecule that is hydrogen bonding to the vanadate ion. In the second conformation the carboxylate forms a hydrogen bond with the amino group of

His260. Both of these conformations are slightly different from the ligand free and the tartrate bound structures.

5.3.5. Development of hydrophobic clamp hypothesis

Inspection of the three HAP structure active sites revealed that two residues, Phe23 and Tyr135 might have roles in substrate binding. In the tartrate bound structure there was density representing a HEPES molecule bound between these two residues possibly stabilizing Phe23. This side chain is disordered when the inhibitor vanadate is bound, and it is rotated away from the active site in the acetate bound structure. When initial substrate specificity studies were done on HAP those having the highest degree of hydrolysis did in fact have aromatic moieties that could interact with these two hydrophobic residues (Data not shown).

Mutation of these residues both individually and together in a double mutation should decrease activity by compromising substrate recognition. If kinetic experiments were done using a molecule that had both a L-tartrate moiety as well as an aromatic group that would interact with the hydrophobic clamp, a lower inhibition constant should be observed when compared to that of L-tartrate. We tested this prediction using (2*R*, 3*R*)-tartranilic acid (Figure 5.4B)

Figure 5.5

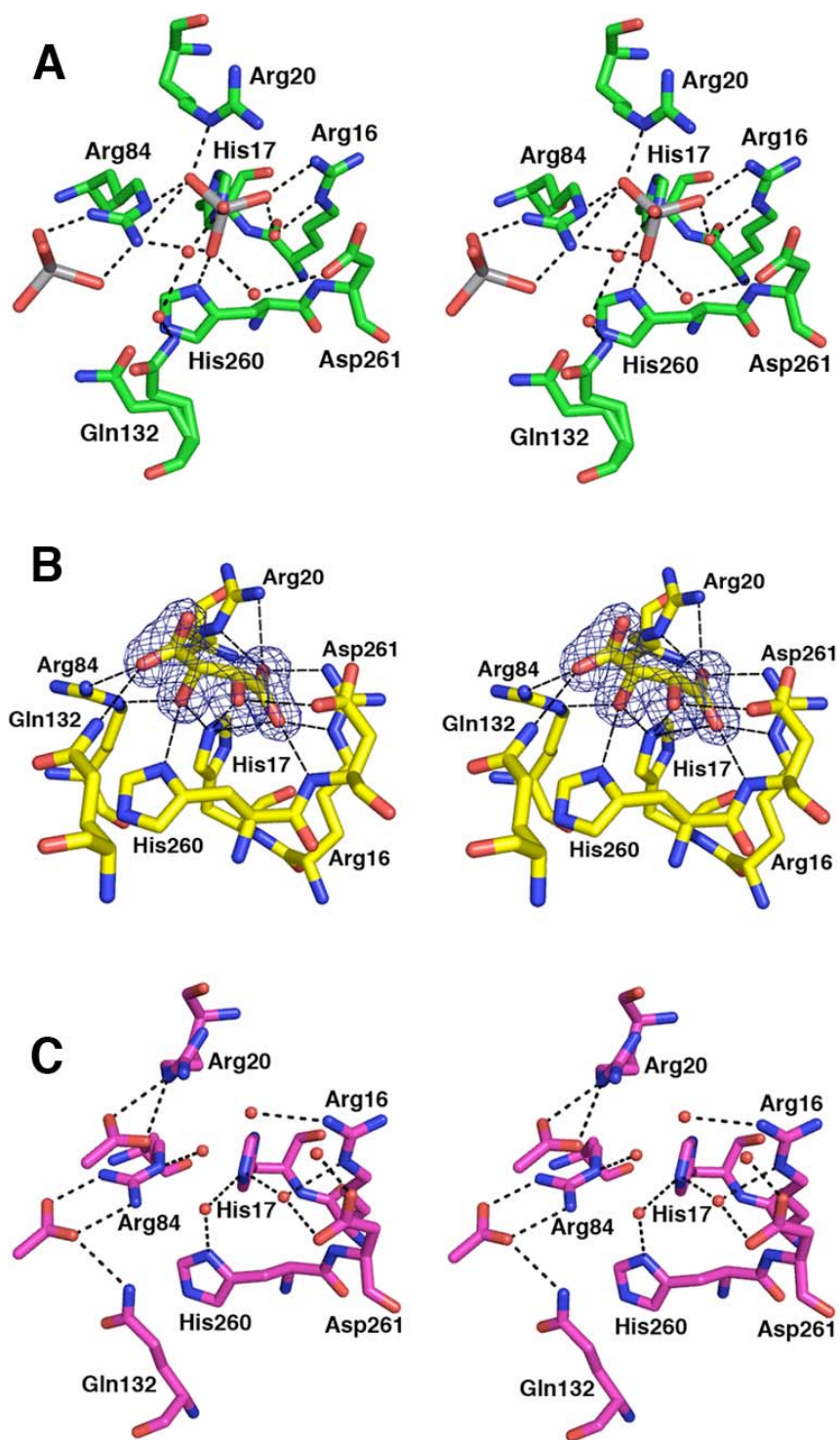


Figure 5.5 Inhibitor interactions within the active site.

A) Stereo drawing of vanadate bound in the HAP active site. **B)** The 1.75Å structure of L-tartrate bound in the HAP active site drawn in stereo. The electron density map around the inhibitor was calculated directly from the RESOLVE phases. The hydrogen bonds are represented as black dashed lines. **C)** Stereo drawing of the HAP active site with acetate bound.

5.3.6. Structure of the HAP/tartranillic acid

When the structure of HAP with (2*R*,3*R*)-tartranalate bound it was determined the molecule was bound just as predicted (Figure 5.6). The L-tartrate end of the molecule was in the identical position to that of the tartrate bound structure. The same interactions between the active site residues were observed for this structure. The aniline end of the molecule fits between the two hydrophobic residues Phe23 and Tyr135. Phe23 is ordered in this structure and overlaps with the tartrate bound structure. Tyr135 is also ordered in this structure however there is a slight movement of 0.5 Å towards the aromatic ring of the ligand (Figure 5.7A).

Figure 5.6

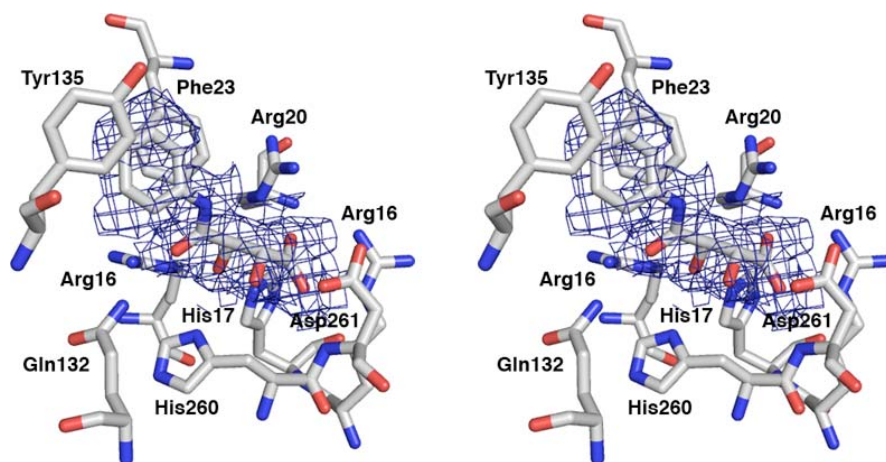


Figure 5.6 Hydrophobic clamp interactions.

The 2.0Å structure of (2*R*, 3*R*)-tartranilic acid bound in the active site of HAP drawn in stereo. The $2F_o - F_c$ electron density was calculated using phases prior to including the inhibitor in the model.

Figure 5.7

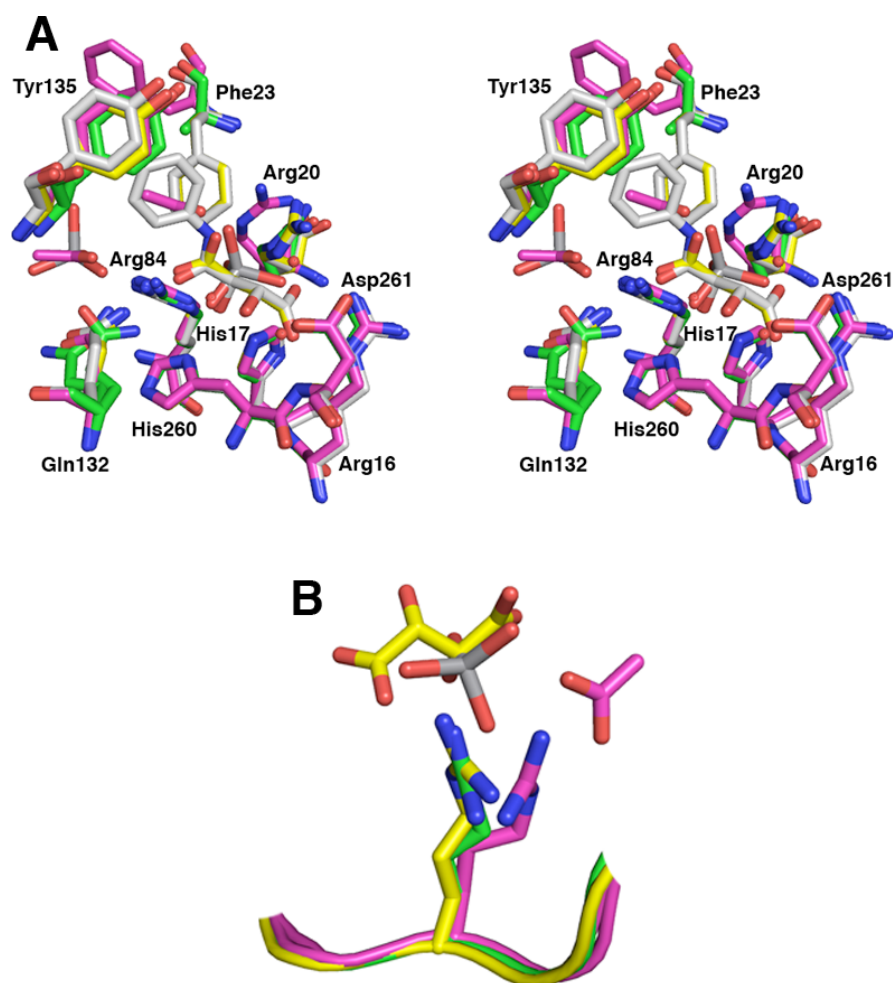


Figure 5.7 Active site overlap of HAP structures.

A) The four HAP structures are shown in stereo view. Coloring is consistent with figures 5.5 and 5.6. Major differences include disorder in Phe23, multiple conformations of Gln132 for the vanadate bound structure, and a rotation of Phe23 in the acetate/HAP structure. **B)** Arg20 of the vanadate and tartrate bound structures has a 1.77Å shift from that of that of acetate/HAP structure. This shift is due to binding of the inhibitor. The color schemes are the same as in panel A.

5.3.7 Determination of kinetic parameters

The K_m of HAP for *p*NPP is estimated to be 2.8 ± 0.3 mM at pH 5.0 in acetate buffer. The V_{max} value for this substrate is $1.0 \times 10^5 \pm 4.0 \times 10^3$ nmoles of *p*-nitrophenolate per minute per mg of enzyme. Inhibition constants were estimated for orthovanadate, L-tartaric acid, and (2*R*, 3*R*)-Tartanilic acid. Of the compounds tested orthovanadate had the greatest inhibitory effect with $K_i = 2.2$ μ M. This was not an unexpected result due the similarity of vanadate to inorganic phosphate, a product of the hydrolysis of phosphomonoesters. The K_i values for the tartrate compounds were 172 μ M for L-tartaric acid and 37.5 μ M for (2*R*, 3*R*)-Tartranilic acid. The increase in affinity of tartranilic acid over tartaric acid is consistent with the observation that the aniline moiety packs between Tyr135 and Phe23. Michaelis-Menton plots for *p*NPP as well as Lineweaver-Burk, and a Dixon plot for the (2*R*, 3*R*)-Tartranilic acid can be found in Figure 5.1B-D.

5.3.8 Comparison to HPAP

A superposition of HAP from *F. tularensis* with the human prostatic acid phosphatase (HPAP)¹², based on the C $_{\alpha}$ positions of the two domains, reveals a high structural similarity between these two enzymes (Figure 5.8). Since both of these enzymes are histidine acid phosphatases it is not surprising that the active sites were very similar, including the conformations of active site side chains. The most striking difference between HAP and HPAP is the position of helix C.

In the HPAP structure the location of this helix is approximately 12.5Å further away from the active site than in HAP. This difference results in a more open active site for HPAP possibly allowing for different substrate binding interactions. A full biochemical characterization of HAP is currently underway and initial results indicate that HAP is more specific for 2' and 3' AMP than it is for *p*NPP. This is not the case for HPAP, where *p*NPP is clearly the best substrate¹³. Located at the N-terminal end of helix C in HAP is amino acid Tyr135. This residue is thought to play an important role in substrate specificity, recognition, and orientation. When HAP was co-crystallized with the competitive inhibitor (2*R*, 3*R*)-tartranilic acid it was clear that this residue was involved in stabilization via non-polar interactions. In HAP we propose that Tyr135 and Phe23 form a clamp that interacts with aromatic rings of the substrate, possibly contributing to substrate preference. Currently site directed mutagenesis on this residue as well as Phe23 are underway to determine if activity is affected.

5.3.9 Other Histidine Acid Phosphatase Structures

HPAP is by far the most well studied member of the histidine acid phosphatase but it is not the only member with its crystal structure solved. Figure 5.9 is a sequence alignment of enzymes in the histidine acid phosphatase class with their structures solved. These include, prostatic acid phosphatases, glucose-1-phosphatases, and phytases. The X-ray crystal structures of HAP reported here represent the highest resolution examples of any histidine acid phosphatase member.

Figure 5.8



Figure 5.8 Overlap of HAP with HPAP

Human prostatic acid phosphatase (HPAP) in yellow, superimposed onto the structure of HAP (blue). The nucleophile, histidine, is shown in magenta. These two structures have an RMSD of 1.9Å with the largest deviation in the two structures belonging to a shift in helix αC , HAP nomenclature. This helix is colored in red for HAP and wheat for HPAP. In HAP this helix is shifted 12.5Å from the equivalent position in the HPAP structure. This shift effectively narrows the entrance to the HAP active site when compared to that of HPAP.

Figure 5.9

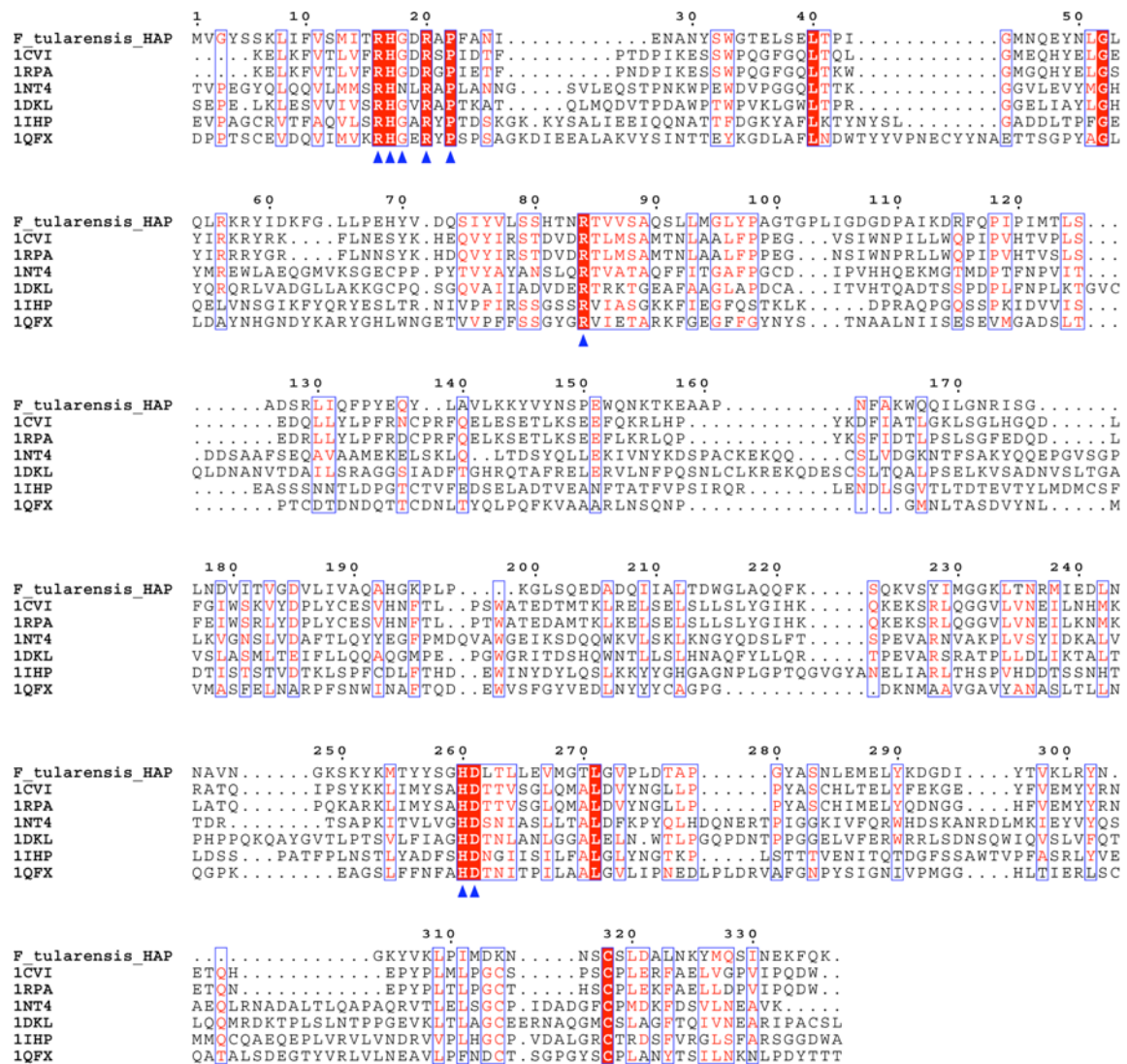


Figure 5.9 Sequence alignments between HAP and other histidine acid phosphatases with known structures.

The PDB accession codes represented in this sequence alignment correspond to known histidine acid phosphatase structures. 1CVI, human prostatic acid phosphatase; 1RPA, rat prostatic acid phosphatase; 1NT4, *E. coli* glucose-1-phosphatase; 1DKL, *E. coli* phytase; 1IHP, *Aspergillus ficuum* phytase; and 1QFX, *A. niger*. The active site residues are indicated by black triangles. These structures share limited sequence similarity, beyond the signature motif, with HAP.

Figure 5.10A

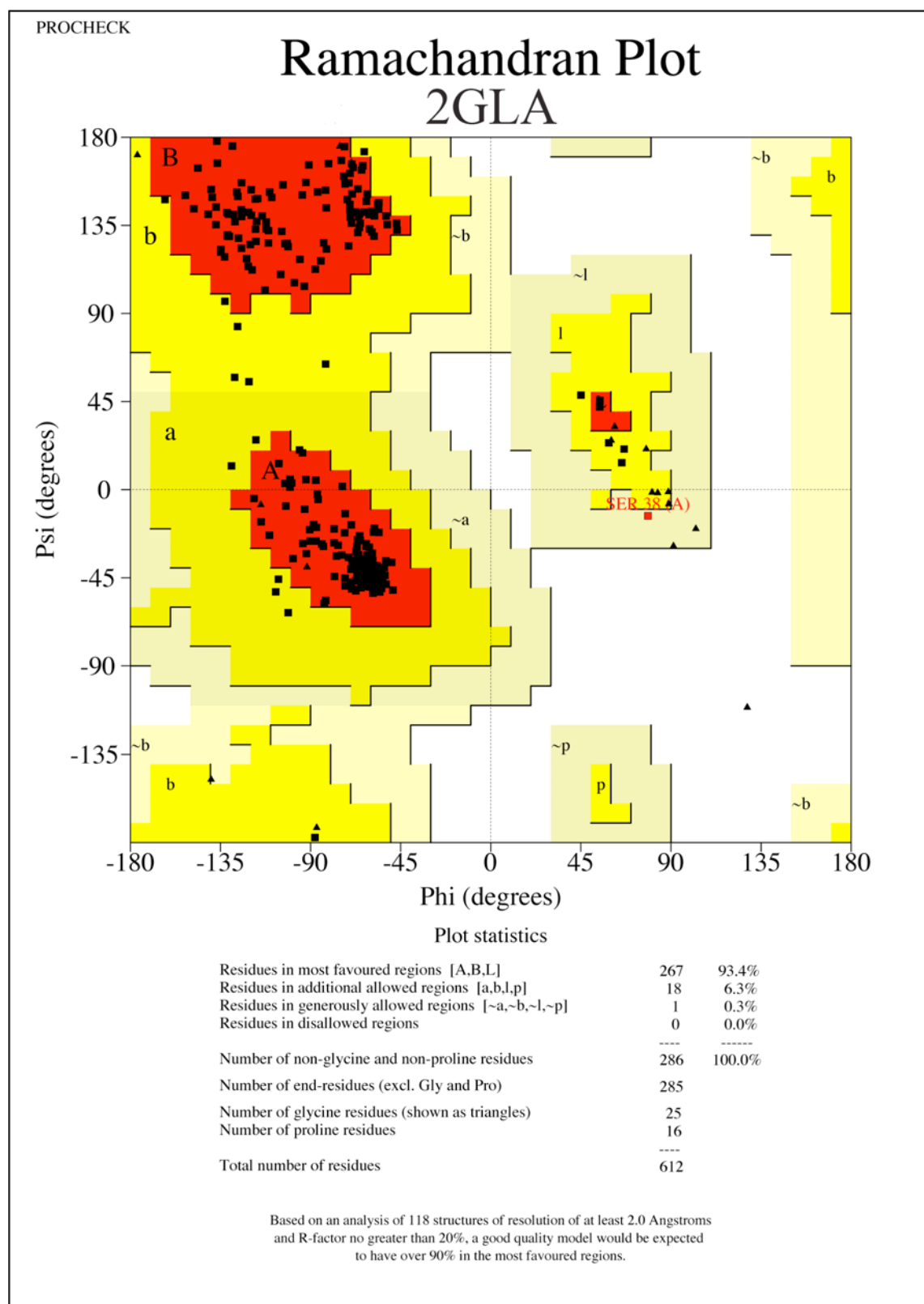


Figure 5.10B

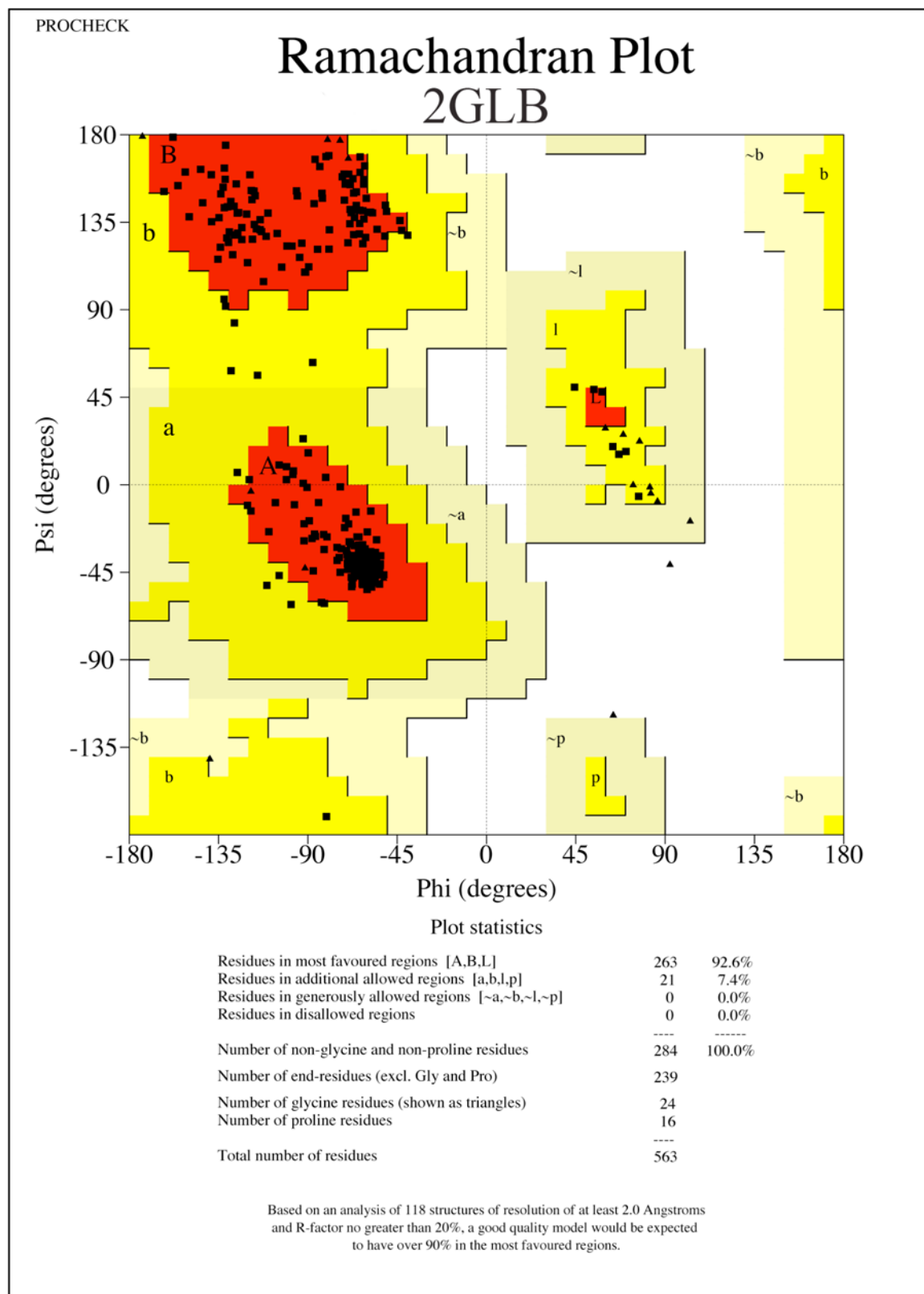


Figure 5.10C

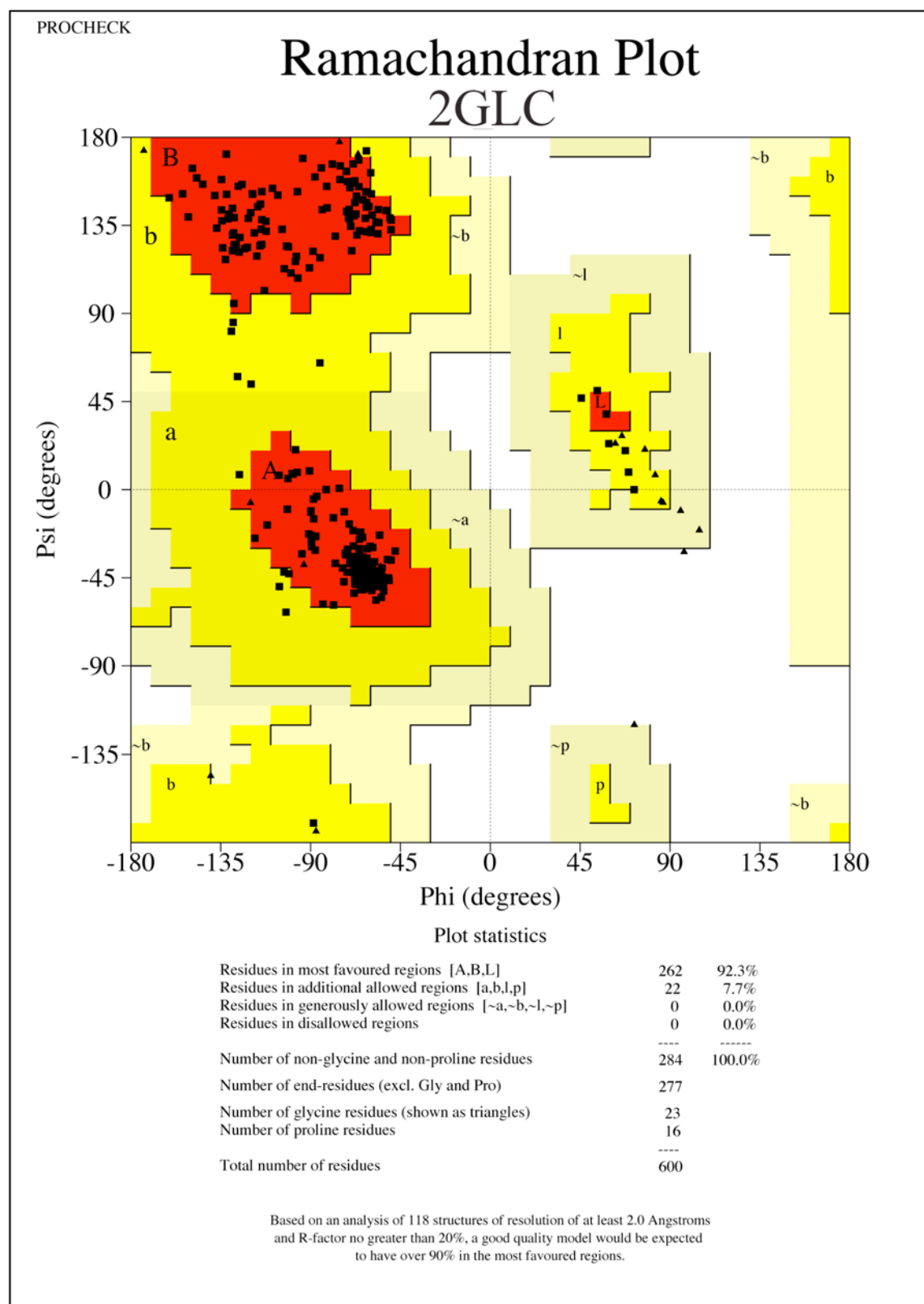
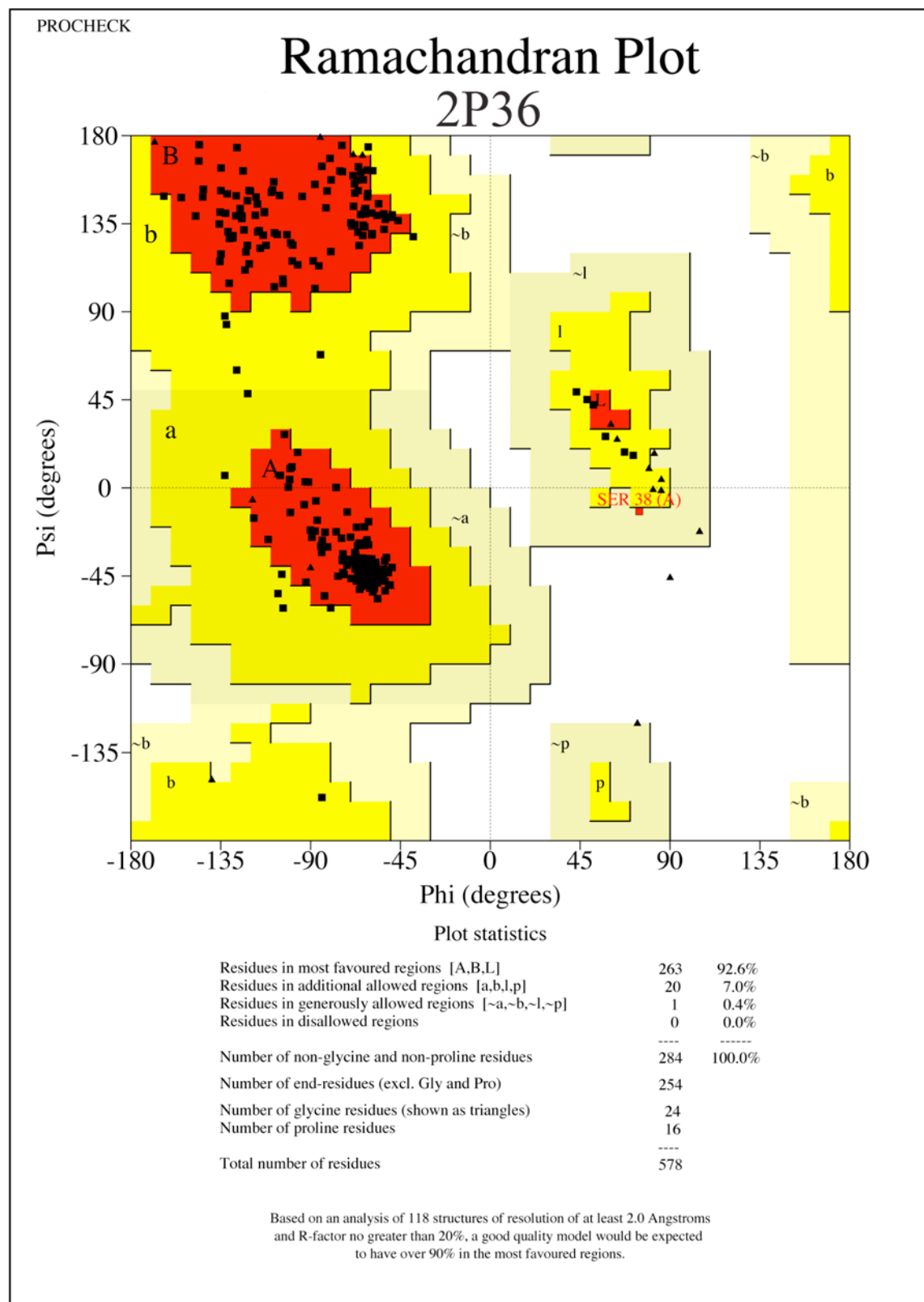


Figure 5.10D



5.4 References

1. Kaufmann, A. F.; Meltzer, M. I.; Schmid, G. P., The economic impact of a bioterrorist attack: are prevention and postattack intervention programs justifiable? *Emerg. Infect. Dis.* **1997**, 3, (2), 83-94.
2. Ellis, J.; Oyston, P. C.; Green, M.; Titball, R. W., Tularemia. *Clin. Microbiol. Rev.* **2002**, 15, (4), 631-46.
3. Oyston, P. C.; Sjøstedt, A.; Titball, R. W., Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nat. Rev. Microbiol.* **2004**, 2, (12), 967-78.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. Van Etten, R. L., Human prostatic acid phosphatase: a histidine phosphatase. *Ann N Y Acad Sci* **1982**, 390, 27-51.
10. 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.

11. Felts, R. L.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Crystallization of a newly discovered histidine acid phosphatase from *Francisella tularensis*. *Acta Crystallogr.* **2006**, F62, (Pt 1), 32-35.
12. Jakob, C. G.; Lewinski, K.; Kuciel, R.; Ostrowski, W.; Lebioda, L., Crystal structure of human prostatic acid phosphatase. *Prostate* **2000**, 42, (3), 211-8.
13. Kilsheimer, G. S.; Axelrod, B., Inhibition of prostatic acid phosphatase by alpha-hydroxycarboxylic acids. *J Biol Chem* **1957**, 227, (2), 879-90.
14. Myers, J. K.; Cohen, J. D.; Widlanski, T. S., Substituent Effects on the Mechanism-Based Inactivation of Prostatic Acid Phosphatase. *Journal of the American Chemical Society* **1995**, 117, (45), 11049-54.
15. Myers, J. K.; Widlanski, T. S., Mechanism-based inactivation of prostatic acid phosphatase. *Science* **1993**, 262, (5138), 1451-3.
16. 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.
17. 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.
18. 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.
19. Van Etten, R. L.; McTigue, J. J., pH dependence and solvent isotope effects in the hydrolysis of phosphomonoesters by human prostatic acid phosphatase. *Biochim Biophys Acta* **1977**, 484, (2), 386-97.
20. 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.

21. 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.
22. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A., *Protein Identification and Analysis Tools on ExPASy Server*. Humana Press: 2005; p 571 - 607.
23. Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968**, 33, 491-497.
24. Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Crystallogr.* **1999**, D55, 1718-1725.
25. Terwilliger, T. C., SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **2003**, 374, 22-37.
26. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **2004**, D60, (Pt 12 Pt 1), 2126-32.
27. 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.
28. Abola, E. E.; Bernstein, F. C.; Bryant, S. H.; Koetzle, T. F.; Weng, J., Protein Data Bank. In *Crystallographic databases - Information content, software systems, scientific applications*, Allen, F. H.; Bergerhoff, G.; Sievers, R., Eds. Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester, 1987; pp 107-132.
29. Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* **1991**, A47, (4), 392-400.
30. 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.

31. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, 26, 283-291.
32. Krissinel, E.; Henick, K., Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). *Comp. Life* **2005**, 163-174.
33. Dixon, M., The determination of enzyme inhibitor constants. *Biochem. J.* **1953**, 55, (1), 170-1.

6.

STRUCTURE OF BACILLUS ANTHRACIS CLASS C ACID PHOSPHATASE

Portions of this chapter have been adapted from:

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. F.*, **2006**, F62, 705-708.

6.1. Introduction

Phosphatases are ubiquitous enzymes that catalyze phosphoryl transfer from phosphomonoesters to water¹. The subgroup of phosphatases known as nonspecific acid phosphatases (NSAPs) refers to bacterial polyspecific phosphatases that are secreted to the periplasmic space or cell surface and that exhibit optimum activity at acidic pH. Rossolini and co-workers first recognized NSAPs as a distinct group of phosphatases and further categorized them into three classes denoted A, B, and C based on subcellular localization and conserved amino acid sequence motifs².

Class C NSAPs are ~30 kDa acid phosphatases that are anchored to the outer membrane surface by an N-terminal lipid modification³. The signature sequence for class C NSAPs is a bipartite motif consisting of (I/V)-(V/A/L)-D-(I/L)-D-E-T-(V/M)-L-X-(N/T)-XX-Y near the N-terminus and (I/V)-(L/M)-XX-G-D-(N/T)-L-X-D-F near the C-terminus, separated by a linker region of 180-220 residues⁴. Class C NSAPs that have been characterized to date include enzymes from *Haemophilus influenzae* (e(P4))⁵, *Streptococcus equisimilis* (LlpC)⁶, *Staphylococcus aureus*⁷, *Helicobacter pylori*⁸, and *Chryseobacterium meningosepticum* (OlpA)⁸. Enzymatically inactive mutants of e(P4) were recently shown to be potential vaccine candidates for *H. influenzae*⁹. Thus, class C enzymes may be useful vaccine reagents for other bacteria. Crystal structures for class C NSAPs have not been reported.

Class C NSAPs are distantly related to class B NSAPs, which also contain two pairs of absolutely conserved aspartic acid residues (DDDD motif) separated

by a linker region⁴. The structure of AphA from *E. coli* shows that class B enzymes have the haloacid dehalogenase (HAD) fold and that the four conserved aspartic acid residues bind a divalent metal ion in the active site¹⁰. The first Asp of the DDDD motif is proposed to be the nucleophile that attacks the P atom of the substrate¹¹. Although class B and C enzymes have similar bipartite DDDD motifs, they share only 14 -22 % global sequence identity⁴. In the following chapter, we will show that class C NSAPs have the HAD fold and active site structure of class B NSAPs.

During our analysis of putative surface proteins in Category A biothreat pathogens, a homolog of the class C NSAP of *H. pylori* was found in the annotated genome sequence of *Bacillus anthracis* strain *Sterne*. *B. anthracis* is a Gram-positive, aerobic, spore-forming bacterium that is the causative agent of anthrax. A gene encoding an identical protein is present in *B. anthracis* strains *Ames* and *A2012*^{12, 13}. Closely related proteins (>97% amino acid sequence identity) are encoded in the genomes of multiple *B. cereus* isolates as well as in strains of *B. thuringiensis*, but no homolog was identified in the complete genome sequences of *B. subtilis*, *B. halodurans* or *B. clausii*. The identified *B. anthracis* protein (RefSeq accession number NP_846955) shares 32 - 44% global amino acid sequence identity with known class C NSAPs from *H. influenzae*, *S. equisimilis*, *S. aureus*, *H. pylori*, and *C. meningosepticum*. The defining class C signature motif for this protein is VLDLDETVDNSPH in the N-terminal half of the protein and VLFFGDNLSDF in the C-terminal half.

As part of our ongoing research on the roles of acid phosphatases in bacterial pathogenesis¹⁴⁻¹⁸, we have initiated structural investigations of the class

C acid phosphatase from *B. anthracis*. We report here the expression, purification, crystallization and structure determination of this class C acid phosphatase.

6.2. Methods and results

6.2.1 Cloning, expression and purification

The gene encoding the class C NSAP from *B. anthracis* delta *Sterne* (strain *Sterne* lacking plasmid pXO1) was amplified by PCR under standard cycling conditions using primers containing *Nco*I and *Xho*I sites. Following initial cloning into pZero Blunt (Invitrogen), the gene was inserted into the expression vector pET20b (Novagen) using *Escherichia coli* DH10B as cloning host. The resulting recombinant protein consists of 251 residues with an approximate molecular weight of 28 kDa, a hexa-histidine tag at the C-terminus, and a pelB leader sequence for targeting to the *E. coli* periplasm. The pelB leader sequence replaces the signal peptide corresponding to residues 1-24 of the native protein. Following confirmation of the insert sequence, the recombinant plasmid, designated pBanxExp, was introduced into the expression host *E. coli* BL21 (DE3) by standard transformation.

A single colony of *E. coli* BL21 (DE3) containing the plasmid pBanxExp was used to inoculate 5 mL LB containing ampicillin (50 mg/mL) and incubated overnight at 37 °C with constant aeration. A 1:1,000 dilution of this overnight culture was then used to inoculate 25 mL of fresh media. The sample was

incubated as before until the density of the culture reached an OD A600 of 0.6. The sample was then chilled on ice for 10 minutes and placed at 4 °C for overnight storage. To remove secreted β -lactamases, the *E. coli* starter culture was centrifuged at 3,660 *g* for 10 minutes at 4 °C and suspended in fresh LB prior to equal distribution into four 1.8 L flasks containing 500 mL LB supplemented with ampicillin (50 μ g/mL) and 0.2% (w/v) filter-sterilized glucose. The culture was grown with constant aeration at 37 °C to an OD A600 of 0.4 at which time IPTG was added to a final concentration of 0.4 mM. The cells were harvested 3 hours after induction by centrifuging at 5,000 *g* for 10 minutes. The resulting pellet was suspended in 25 mL 50 mM sodium acetate pH 6.0. All subsequent purification steps were carried out at 4 °C unless otherwise noted. The cells were disrupted by two passes through a French pressure cell adjusted to 10,000 psi in a 40k rapid fill cell and a flow rate of 20 drops/minute. The resulting extract was then stirred in the presence of 1 M NaCl. Unbroken cells and pelletable debris were removed by centrifugation at 31,000 *g* for 15 minutes. Bacterial membranes were pelleted by ultracentrifugation at 192,000 *g* for 1 hour. The ultracentrifugation supernatant, which contained the majority of acid phosphatase activity, was dialyzed against 100 volumes of 20 mM sodium phosphate pH 7.4 containing 1.0 M NaCl (buffer A) for 24 hours.

The dialyzed sample was applied (5 mL/minute) to a HiTrap Chelating HP chromatography column (Amersham Biosciences, 5 mL) pre-equilibrated with buffer A and charged with 100 mM NiCl₂. The applied sample was washed with several column volumes of buffer A, followed by a second wash with buffer A supplemented with 100 mM imidazole. The latter wash liberated large amounts

of contaminating proteins as shown in Figure 6.1 (chromatogram peak B). After the washing steps, a linear gradient of 100 mM - 500 mM imidazole in buffer A (over 8 column volumes) was performed. The enzyme of interest eluted at approximately 200 mM imidazole and was judged by SDS-PAGE to be the correct molecular weight and sufficiently pure for crystallization trials (Figure 6.1, chromatogram peak C). The purified phosphatase was dialyzed against 50 mM sodium acetate pH 6.0 for 24 hours and concentrated to 10 mg/mL using Amicon Ultra 15 centrifugal filter devices (10-kDa mw cutoff). Purified *B. anthracis* class C NSAP exhibited acid phosphatase activity using a variety of substrates including nucleotides, *p*-nitrophenyl phosphate, and 4-methylumbelliferyl-phosphate.

Figure 6.1

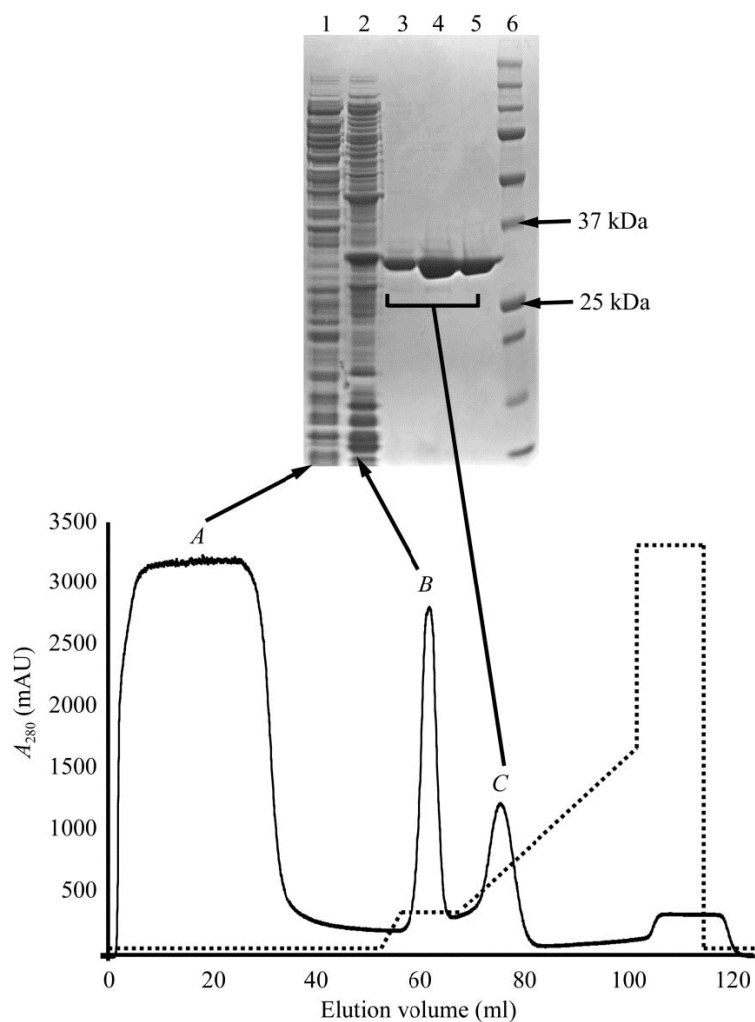


Figure 6.1. SDS-PAGE analysis of fractions from Ni-chelate chromatography.

Lane 1 is the flow-through indicated as peak A on the chromatogram. Lane 2 represents a sample from peak B, which resulted from a washing step at 100 mM imidazole. Lanes 3, 4, and 5 correspond to the first, middle, and last fractions from peak C. Peak C eluted at 100 - 200 mM and contained the desired enzyme. Lane 6 shows molecular mass standards.

6.2.2 Crystallization

Crystals were grown at 293 K using Cryschem 24-well sitting drop plates (Hampton Research). Initial screening was performed using the Index Screen from Hampton Research. Equal volumes of the protein at a concentration of 10 mg/mL (1.5 mL) and the reservoir (1.5 mL) were mixed and equilibrated with 1.0 mL of reservoir solution. Large, well diffracting crystals were obtained directly from Index reagent 39 (0.1 M HEPES pH 7.0, 30% (w/v) Jeffamine ED – 2001 pH 7.0). These crystals typically appeared within two weeks after setup as rectangular blocks with dimensions of 0.4 x 0.2 x 0.2 mm (Figure 6.2). Prior to cryogenic data collection the crystals were soaked in the harvest buffer (0.1 M HEPES pH 7.0, 35% Jeffamine ED – 2001 pH 7.0) supplemented with 25% (w/v) PEG 200.

Figure 6.2

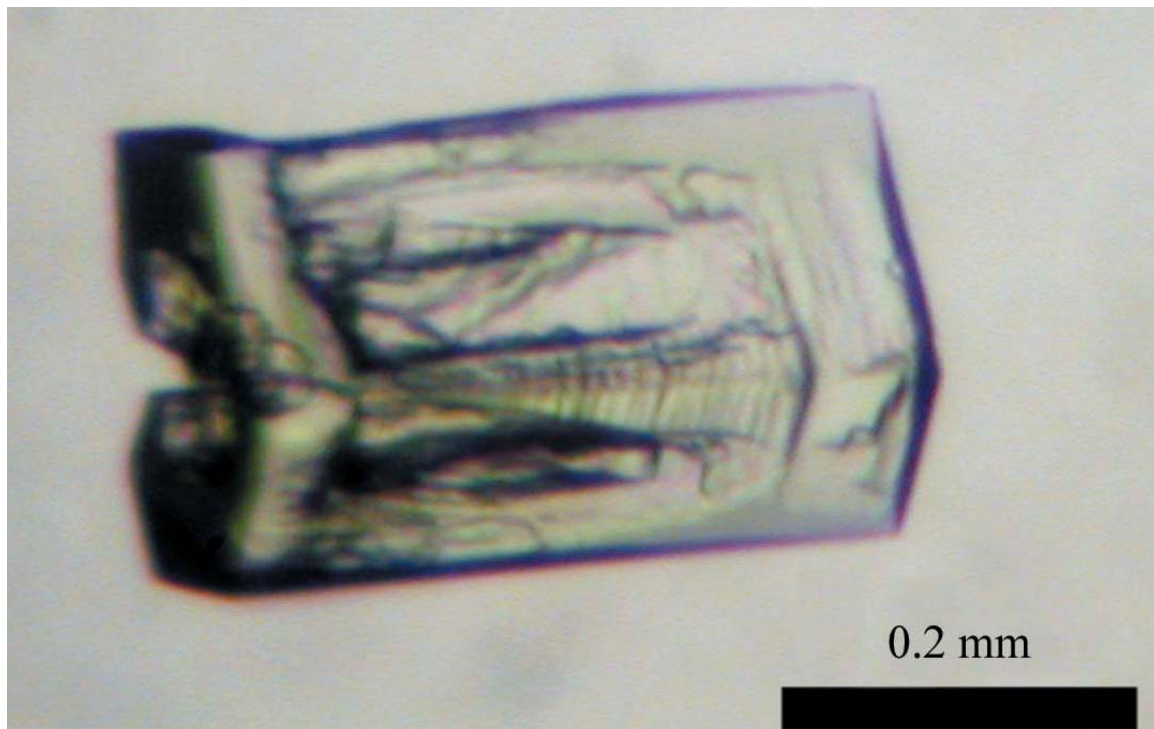


Figure 6.2. Crystals of *B. anthracis* class C acid phosphatase.

A crystal of *B. anthracis* class C NSAP grown by vapor diffusion over a reservoir containing 0.1 M HEPES pH 7.0, 30% (w/v) Jeffamine ED – 2001 pH 7.0.

6.2.3 Data collection and processing

Two diffraction data sets, corresponding to before and after flash annealing, were collected from the same crystal at ALS beamline 4.2.2 using a NOIR-1 CCD detector. For both data sets, the detector distance and angle were set to 125 mm and 0° , respectively, and the oscillation width was 0.5° per frame. The two data sets differed in exposure time per frame and total oscillation width. The first data set consisted of 110° of data collected with an exposure time of 7 seconds per frame, whereas the second data set consisted of 180° of data with exposure time of 10 seconds per frame.

Processing of the first data set in d*TREK¹⁹ suggested space group $P2_12_12_1$ with unit cell lengths of $a = 53.3 \text{ \AA}$, $b = 90.1 \text{ \AA}$, and $c = 104.1 \text{ \AA}$. Analysis of crystal packing using the method of Matthews indicated a probable solvent content of 38 % with 2 protein molecules in the asymmetric unit and $V_M = 2.8 \text{ \AA}^3/\text{Da}^{20}$. This data set exhibited acceptable statistics to 1.8 \AA resolution and the refined mosaicity was 1.6° (Table 6.1, Before annealing).

Although the 1.8 \AA data set was quite satisfactory, we nonetheless decided to test the response of the crystal to flash annealing²¹. With the crystal remaining in data collection position, the cold nitrogen stream was blocked for approximately 4 seconds using a plastic card. Visualization of the crystal through the beamline video system verified that the crystal had indeed thawed during the brief blockage of the stream, and that the crystal was re-frozen upon removing the card. This procedure was repeated three times.

Test diffraction images collected from the annealed crystal suggested an improvement in resolution and so another complete data set was recorded. This second data set could be processed to 1.6 Å resolution and the mosaicity improved to 0.85° (Table 6.1, After annealing). Thus, flash annealing appeared to cause a dramatic improvement in diffraction quality.

To more accurately assess the improvement in crystal quality caused by annealing, the data set collected after annealing was truncated at 1.8 Å resolution and only the first 95° of data were used so that the resulting redundancy was comparable to that of pre-annealing data set. Statistics from this calculation are shown in Table 6.1 (last column). Comparison of the two 1.8 Å data sets shows that the data set collected before annealing had overall average $I/\sigma(I)$ of 8.9, compared to $I/\sigma(I) = 20$ after annealing (Table 6.1). The average $I/\sigma(I)$ in the high-resolution bin improved from 2.7 before annealing to 9.2 after annealing. Likewise, the overall R_{merge} improved from 0.066 (0.372) before annealing to 0.038 (0.109) after annealing. We note that the factor of 2-3 increase in $I/\sigma(I)$ exceeds the increase expected from counting statistics due to the longer exposure time used for the second data set²². We therefore conclude that the observed improvement in data quality was primarily due to increased lattice order induced by flash annealing.

The amino acid sequence of *B. anthracis* class C NSAP was used in a BLAST²³ search of the Protein Data Bank²⁴ to determine if suitable homologs were available for use as search models in molecular replacement studies. The closest homolog was AphA from *E. coli*, which shares only 17% global sequence

identity with *B. anthracis* class C NSAP. Thus, structure determination is being pursued with experimental phasing techniques.

Table 6.1**Data collection and processing statistics^a**

	Before annealing	After annealing	After annealing (reprocessed to 1.8 Å)
Beamline	ALS 4.2.2	ALS 4.2.2	ALS 4.2.2
Wavelength (Å)	1.2037	1.2037	1.2037
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit-cell dimensions (Å)	$a = 53.35, b = 90.09,$ $c = 104.15$	$a = 53.36, b = 90.15,$ $c = 104.23$	$a = 53.36, b = 90.15,$ $c = 104.23$
Diffraction resolution (Å)	45.91 – 1.80 (1.86 – 1.80)	45.12 – 1.60 (1.66 – 1.60)	45.12 – 1.60 (1.86 – 1.80)
Exposure time per frame (s)	7	10	10
Total oscillation range (°)	110	180	95
No. of observations	170571	451942	172003
No. of unique reflection	45878	66821	46860
Redundancy	3.7 (3.7)	6.7 (6.4)	3.76 (3.5)
Refined mosaicity (°)	1.6	0.8	0.8
Completeness (%)	96.9 (95.4)	99.5 (98.5)	98.8 (97.2)
Average $\langle I/\sigma(I) \rangle$	8.9 (2.7)	11.1 (3.0)	20 (9.2)
R_{merge}^b	0.066 (0.372)	0.062 (0.443)	0.038 (0.109)

^a Values for the outer resolution shell of data are given in parentheses

^b $R_{merge} = \sum_{hkl} \sum_j |I_{hkl}(j) - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_j I_{hkl}(j)$, where $\langle I_{hkl} \rangle$ is the mean intensity and $I_{hkl}(j)$ are individual intensity measurements of the reflection (hkl).

6.2.4 Structure determination

To solve the crystal structure of *B. anthracis* class C acid phosphatase, native crystals were soaked for one hour in a 40 mM sodium tungstate (Na_2WO_4) solution in order to incorporate the heavy atom into the crystal. These derivative crystals were screened on the home X-ray source in order to determine if they were suitable for beamline data collection. Those crystals that met the standards were stored for data collection at the ALS beamline 4.2.2. Data from the tungstate derivatives were collected at a wavelength of 1.20372 Å in order to take advantage of the tungsten anomalous signal. At this wavelength tungsten has an approximate f'' of 10 electrons. The data collection statistics can be found in Table 6.1.

An anomalous difference Patterson map was calculated in order to determine if there were heavy atoms present in the structure. Harker sections of this map are shown in Figure 6.3. The derivative data were processed using d*TREK and the reflection file was then input in to the SOLVE/RESOLVE programs²⁵. SOLVE identified two heavy atom sites and the resulting phases had figure of merit 0.27 to 2.5 Å resolution. Density modification by solvent flattening was performed with resolve at 2.5 Å resolution. The initial trace from RESOLVE consisted of 120 residues including 20 with sequence assigned. This result was interpreted to be only one molecule in the asymmetric unit. A new heavy atom input file was made to incorporate not only the tungsten atoms but also three atoms that we predicted to be related by non-crystallographic symmetry (NCS). This file was then input back into RESOLVE, which was then

run using non-crystallographic symmetry averaging. The result of this run was 2 molecules in the asymmetric unit with 324 of the 504 residues traced. Of those traced 201 were assigned to sequence. The figure of merit from RESOLVE was 0.70 to 2.0 Å resolution. The resulting file containing the phases to 2.0Å was then merged with the high resolution native data using CAD in the CCP4 suite²⁶. DM was used for density modification and phase extension to the full 1.6Å resolution of the native data. The resulting phases were then input into ARP/wARP²⁷. The resulting model consisted of 369 residues assigned to sequence and a connectivity of 92%. The remaining residues were built in manually using COOT²⁸. Between model building cycles refinement was done using REFMAC5²⁹. Ramachandran plots were made to ensure that the amino acids in the structure obeyed the correct geometries (Figure 6.4A & B). The refinement statistics can be found in Table 6.2.

Figure 6.3

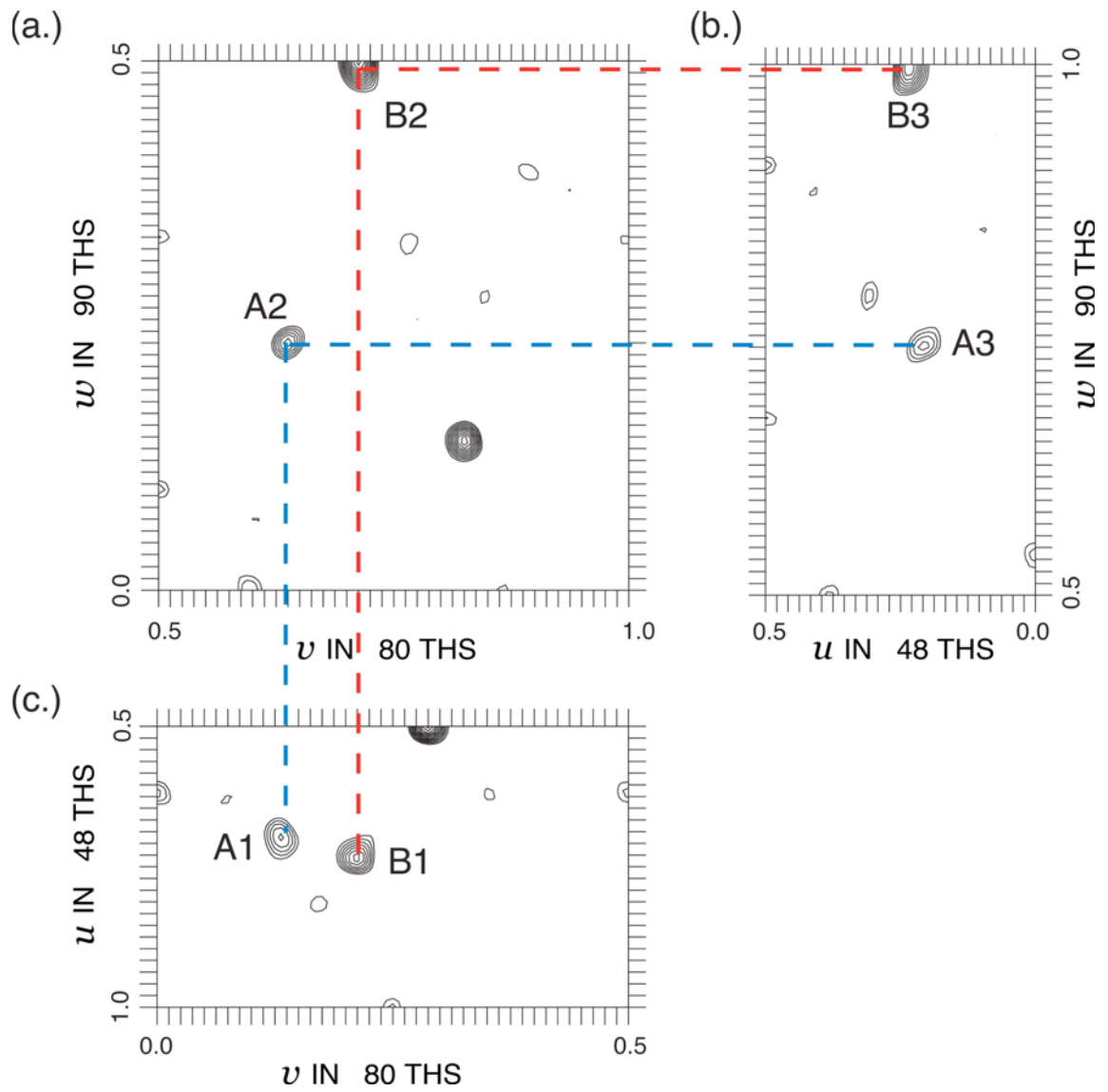


Figure 6.3. Tungstate derivative anomalous difference Patterson from *Bacillus anthracis* class C acid phosphatase. (A-C) Harker sections of the anomalous difference Patterson map $u = 0.5$, $v = 0.5$, and $w = 0.5$ corresponding to the Harker planes $1/2$, $1/2 \pm 2y$, $\pm z$; $\pm 2x$, $1/2$, $2z \pm 1/2$; $1/2 \pm 2x$, $\pm 2y$, $1/2$ at 3.5 Å resolution. The Patterson coordinates for the A2 peak are $(u,v,w) = (1/2, 51/80, 21/90) = (1/2, 1/2 \pm 2y, \pm z)$ converting to real space coordinates $(x,y,z) = (x, -0.0687, -0.1167)$. For peak A3 $(u,v,w) = (10/48, 1/2, 66/90) = (\pm 2x, 1/2, 2z \pm 1/2)$ converting to real space coordinates $(x,y,z) = (-0.1042, y, -0.1167)$. For peak A1 $(u,v,w) = (34/48, 10.5/80, 1/2) = (-0.1042, -0.0687, z)$. The same calculations were done for peaks B1, B2, and B3 with the real Patterson coordinates equal to $(u,v,w) = (35/48, 17/80, 1/2)$; $(u,v,w) = (1/2, 57/80, 45/90)$; and $(u,v,w) = (11/48, 1/2, 90/90)$ respectively. These Patterson peaks for each Harker section convert to real space coordinates $(x,y,z) = (0.1146, -0.1063, z)$; $(x,y,z) = (x, -0.1063, -0.25)$; and $(x,y,z) = (0.1146, y, -0.25)$. The position of the heavy atoms in real space are $(x,y,z) = (-0.1042, -0.0687, -0.1167)$ corresponding to peak A and $(x,y,z) = (0.1146, -0.1063, -0.25)$ corresponding to peak B.

Figure 6.4A

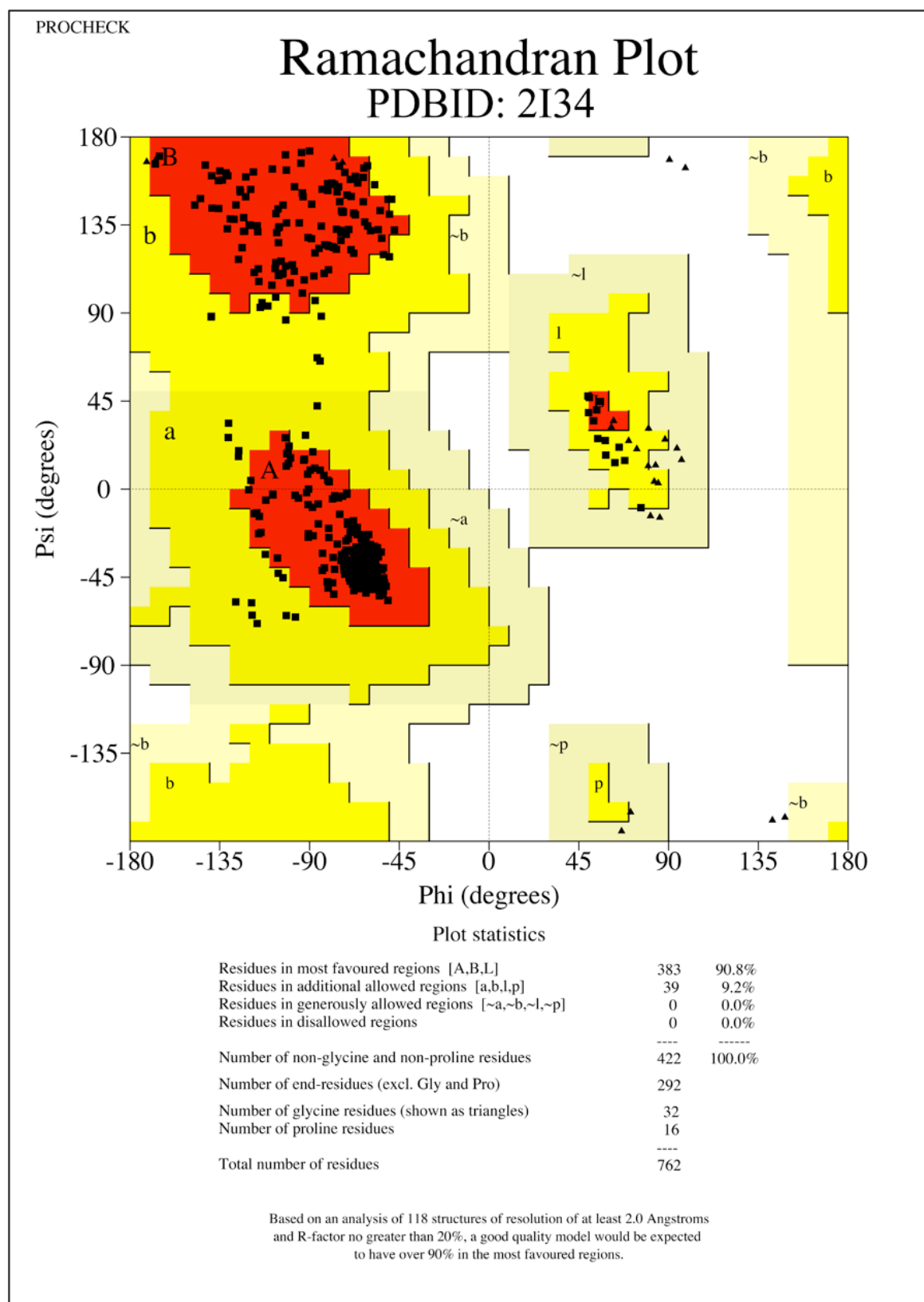


Figure 6.4B

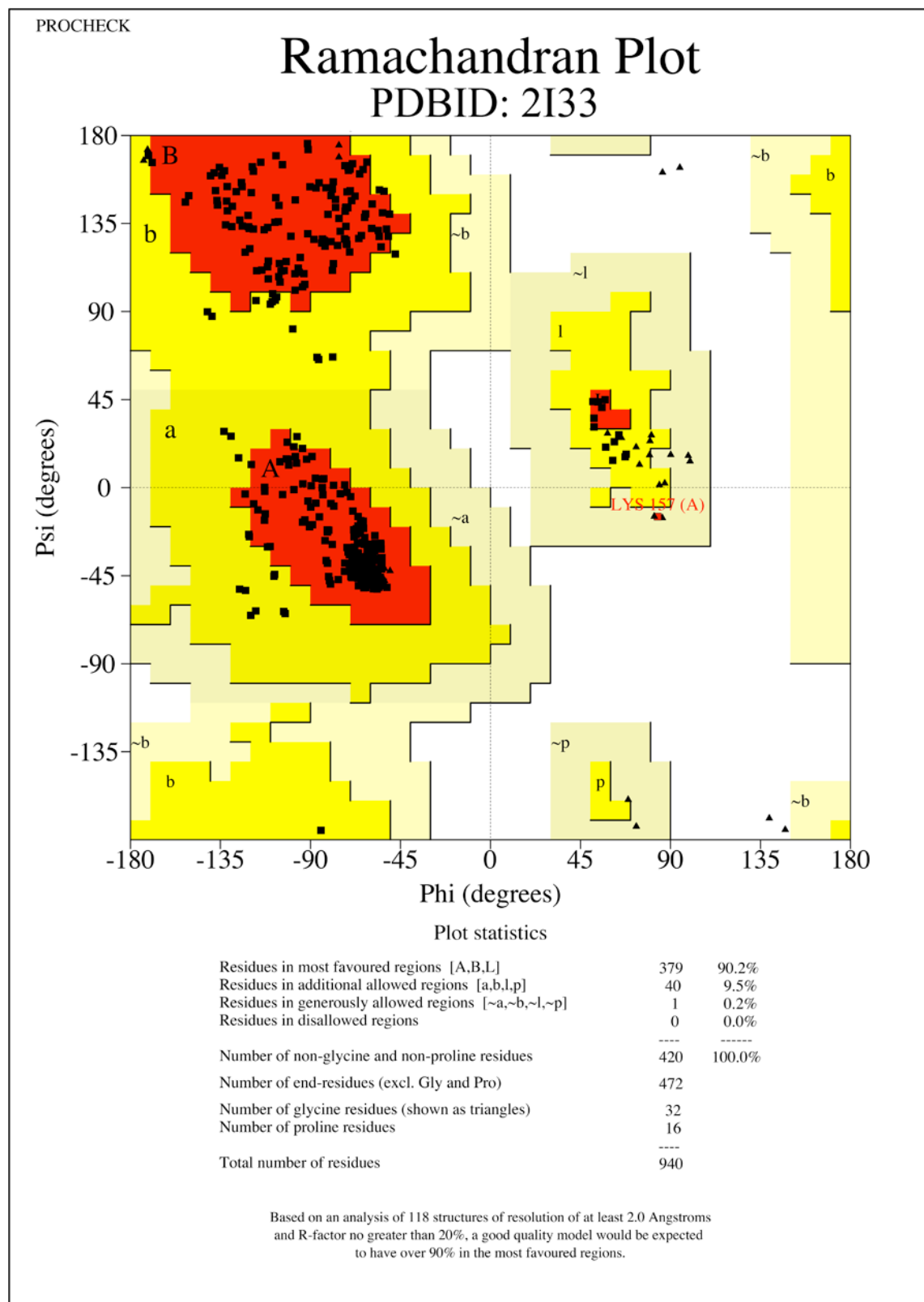


Table 6.2

Data collection and refinement statistics ^a		
	Native	Tungstate Derivative
Wavelength (Å)	1.000	1.005
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit cell dimensions (Å)	a = 53.21, b = 89.92, c = 104.27	a = 53.00, b = 89.94, c = 104.12
Diffraction resolution (Å)	42.2-1.57 (1.61-1.57)	47.23-2.00 (2.05-2.00)
No. of observations	451942	142819
No. of unique reflections	66821	34308
Redundancy	6.7 (6.4)	4.2 (4.0)
Completeness (%)	99.5 (98.5)	99.4 (92.7)
Average I/σ	11.1 (3.0)	8.4 (3.5)
R _{sym} (I)	0.062 (0.443)	0.110 (0.391)
No. of non-hydrogen atoms	4257	4071
No. of residues in chain A	236	236
No. of residues in chain B	236	236
No. of water molecules	574	279
R _{cryst}	0.173 (0.319)	0.181 (0.201)
R _{free} ^b	0.207 (0.356)	0.224 (0.247)
RMSD ^c		
Bond lengths (Å)	0.011	0.010
Bond angles (deg.)	1.318	1.179
Ramachandran plot ^d		
Favored (%)	90.2	90.8
Allowed (%)	9.5	9.2
Average B-factors (Å ²)		
Protein	29.38	18.86
Orthovanadate	NA	16.27
Active site metal ion	14.12	14.60
Water	34.45	17.89
PDB accession code	2I33	2I34

^aValues for the outer resolution shell of data are given in parenthesis.

^b5 % random test set.

^cCompared to the Engh and Huber force field ³⁰.

^dThe Ramachandran plot was generated with Procheck³¹.

6.4 Discussion

The overall structures of class c NSAPs from *B. anthracis* and *F. tularensis* are very similar to that of rP4 (Figure 6.5). There is a central core domain composed of a central β -sheet with α helices flanking either side. There is an upper domain made up of α helices. The structures have an RMSD range of 0.7 Å to 1.2 Å. This is an indication that there are very small differences in structure, but globally they are similar. When a sequence alignment was done only the active sites were highly conserved. In Figure 6.7 the sequence alignment shows which residues are conserved among the three enzymes. The pairwise sequence identity among the three enzymes is only 21-34%. It was thus surprising that the *F. tularensis* structure could be solved using molecular replacement with the *B. anthracis* structure as the starting model. There is a section of the *F. tularensis* sequence that does not correlate to the others. This section is an extra loop region between α G and β 4. The role of this loop is unknown. The *F. tularensis* enzyme is approximately 35 residues shorter than the other two enzymes and therefore lacks the c-terminal loop and helix secondary structures that the others share.

The active sites of both *B. anthracis* and *F. tularensis* overlap perfectly with the active site of the *H. influenzae* class C NSAP (Figure 6.6). This also was not unexpected as they are all members of both the DDDD and HAD superfamilies. There is minimal biochemical characterization of the *F. tularensis* and *B. anthracis* enzymes but judging from the in-depth biochemical characterizations completed for *H. influenzae*^{5, 32, 33}, *H. pylori*³, and *C.*

*meningosepticum*⁸ class C NSAPs it would not be surprising to find that these two enzymes have similar properties, including substrate specificity and inhibition.

The structures of the class C NSAPs reported here should provide a framework for biochemical characterization, including site directed mutagenesis and mechanistic studies, in order to determine a biological significance for these enzymes. In chapter 7 the vaccine possibilities for the *H. influenzae* class C NSAP are discussed, it will be interesting to know if any of the other class C NSAP share this biological property.

Figure 6.5

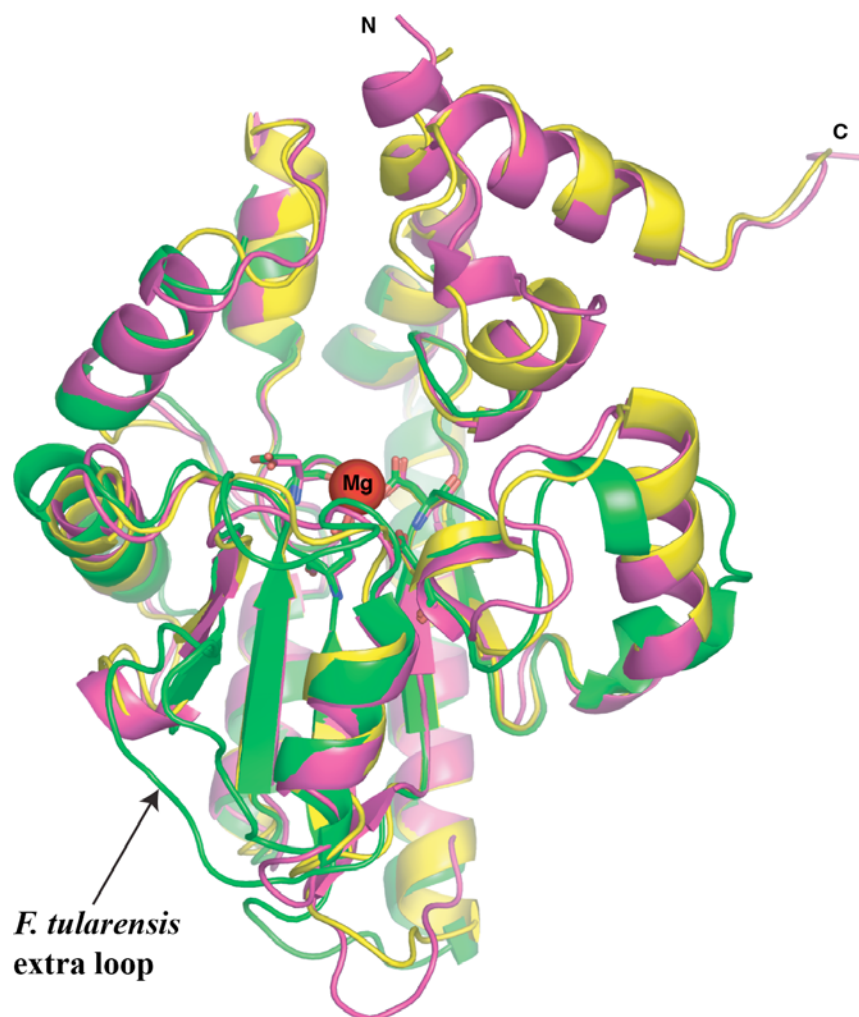


Figure 6.5 Overlap of class C NSAPs

Enzymes from *B. anthracis*, *F. tularensis*, and *H. influenzae* are colored yellow, green and magenta respectively. The N and C termini have been labeled accordingly. The extra loop region of the *F. tularensis* structure has also been labeled. The role of this extra region is unknown at this time.

Figure 6.6

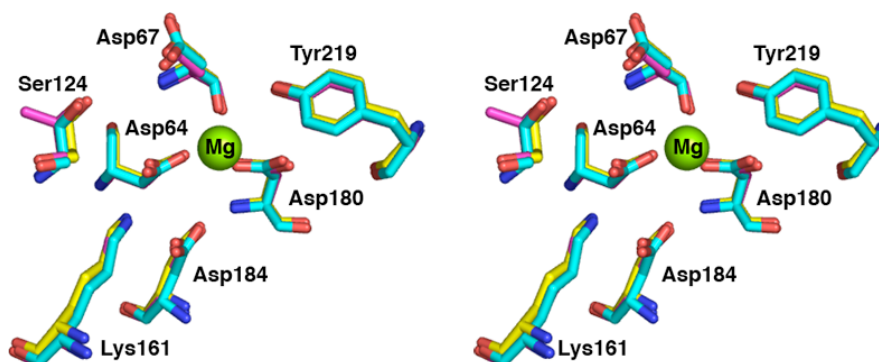


Figure 6.6 Class C NSAP active site overlap.

Stereo drawing of the three active sites overlaid onto one another. The yellow residues and numbering correspond to the *B. anthracis* structure, the teal *F. tularensis*, and the magenta *H. influenzae*. The only difference in the three active sites is the substitution of a threonine for a serine in the *H. influenzae* structure.

Figure 6.7

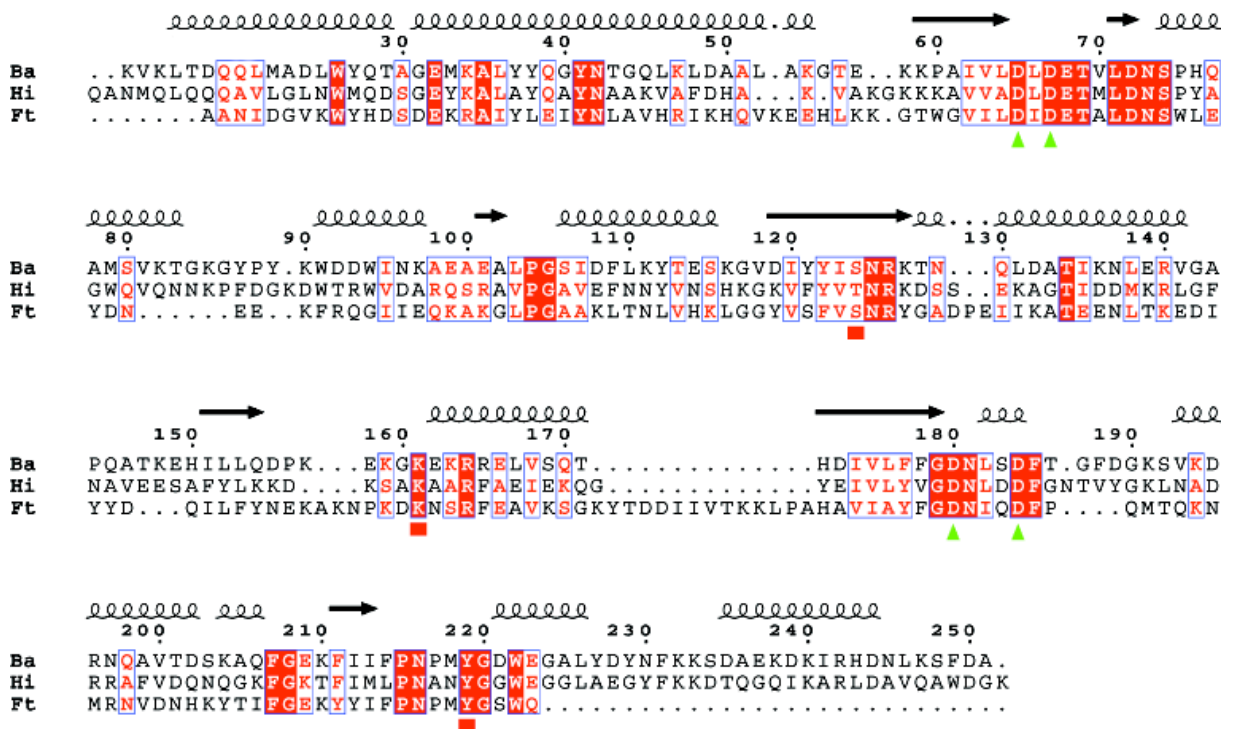


Figure 6.7 Class C NSAP sequence alignment.

This is a sequence alignment between the three class C NSAP with known structures. The secondary structure elements are derived from the *B. anthracis* structure. The green triangles indicate the position of the four aspartic acid residues of the DDDD superfamily, as well as the active site residues. The red squares denote other conserved active site residues. The pairwise sequence identity is 34% for *B. anthracis* and *H. influenzae*, 24% for *B. anthracis* and *F. tularensis*, and 21% for *H. influenzae* and *F. tularensis*.

6.5 References

1. Knowles, J. R., Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **1980**, 49, 877-919.
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. 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.
4. 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.
5. 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.
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. 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. 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.

10. Calderone, V.; Forleo, C.; Benvenuti, M.; Cristina Thaller, M.; Maria Rossolini, G.; Mangani, S., The first structure of a bacterial class B Acid phosphatase reveals further structural heterogeneity among phosphatases of the haloacid dehalogenase fold. *J. Mol. Biol.* **2004**, 335, (3), 761-73.
11. 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.
12. Read, T. D.; Peterson, S. N.; Tourasse, N.; Baillie, L. W.; Paulsen, I. T.; Nelson, K. E.; Tettelin, H.; Fouts, D. E.; Eisen, J. A.; Gill, S. R.; Holtzapple, E. K.; Okstad, O. A.; Helgason, E.; Rilstone, J.; Wu, M.; Kolonay, J. F.; Beanan, M. J.; Dodson, R. J.; Brinkac, L. M.; Gwinn, M.; DeBoy, R. T.; Madpu, R.; Daugherty, S. C.; Durkin, A. S.; Haft, D. H.; Nelson, W. C.; Peterson, J. D.; Pop, M.; Khouri, H. M.; Radune, D.; Benton, J. L.; Mahamoud, Y.; Jiang, L.; Hance, I. R.; Weidman, J. F.; Berry, K. J.; Plaut, R. D.; Wolf, A. M.; Watkins, K. L.; Nierman, W. C.; Hazen, A.; Cline, R.; Redmond, C.; Thwaite, J. E.; White, O.; Salzberg, S. L.; Thomason, B.; Friedlander, A. M.; Koehler, T. M.; Hanna, P. C.; Kolsto, A. B.; Fraser, C. M., The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **2003**, 423, (6935), 81-6.
13. Read, T. D.; Salzberg, S. L.; Pop, M.; Shumway, M.; Umayam, L.; Jiang, L.; Holtzapple, E.; Busch, J. D.; Smith, K. L.; Schupp, J. M.; Solomon, D.; Keim, P.; Fraser, C. M., Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **2002**, 296, (5575), 2028-33.
14. Felts, R. L.; Reilly, T. J.; Calcutt, M. J.; Tanner, J. J., Crystallization of a newly discovered histidine acid phosphatase from *Francisella tularensis*. *Acta Crystallogr.* **2006**, F62, (Pt 1), 32-35.
15. 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.
16. 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.
17. Ou, Z.; Felts, R. L.; Reilly, T. J.; Nix, J. C.; Tanner, J. J., Crystallization of recombinant *Haemophilus influenzae* e (P4) acid phosphatase. *Acta Crystallogr.* **2006**, F62, (Pt 5), 464-6.

18. 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.
19. Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Crystallogr.* **1999**, D55, 1718-1725.
20. Matthews, B. W., Solvent content of protein crystals. *J. Mol. Biol.* **1968**, 33, 491-497.
21. Yeh, J. I.; Hol, W. G., A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Crystallogr D Biol Crystallogr* **1998**, 54, (Pt 3), 479-80.
22. Dauter, Z.; Wilson, K. S., Principles of monochromatic data collection. In *International Tables for Crystallography*, Rossmann, M. G.; Arnold, E., Eds. Kluwer Academic Publishers: Dordrecht, 2001; Vol. F, pp 177-196.
23. 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.
24. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucl. Acids Res.* **2000**, 28, 235-242.
25. Terwilliger, T. C., SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **2003**, 374, 22-37.
26. Project, C. C., The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* **1994**, D50, 760-763.
27. Morris, R. J.; Perrakis, A.; Lamzin, V. S., ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr.* **2002**, D58, (Pt 6 Pt 2), 968-75.
28. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **2004**, D60, (Pt 12 Pt 1), 2126-32.

29. 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.
30. Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* **1991**, **A47**, (4), 392-400.
31. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, **26**, 283-291.
32. 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.
33. 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.

STRUCTURE OF P4 FROM HAEMOPHILUS INFLUENZAE

7.1 Introduction

Lipoprotein *e*(P4) is one of 6 highly conserved outer membrane proteins of *Haemophilus influenzae*, a Gram-negative facultatively anaerobic coccobacillus. The organism is a common commensal inhabitant of the human nasopharynx and the etiologic agent of local and invasive infections in both pediatric and adult populations. *H. influenzae* strains cause a variety of infections including otitis media, sinusitis, bronchitis, conjunctivitis and pneumonia. In particular, *H. influenzae* infections of the lower respiratory tract are a major cause of mortality in both infants and children in developing countries¹.

An obvious prelude to clinical infection is colonization of the host. Colonization and survival of the microbe on the mucosal surface is dependent on the activity of proteins located at the periphery of the bacterium, juxtaposed to the host's mucosal surface. These outer membrane proteins have been under intensive investigation as vaccine candidates and as targets for development of chemotherapeutic agents. Many of the 36 different protein species found in the *H. influenzae* envelope have been characterized with regard to immunogenicity, antigenicity and protective attributes².

Lipoprotein *e*(P4), the subject of this study, is a highly conserved 28 kDa acid phosphatase found in all strains of *H. influenzae*^{1, 3}. The immature protein has a lipoprotein signal peptide at the N-terminus, which results in anchoring of the mature protein in the outer membrane surface by a lipidated N-terminal Cys residue. P4 is currently being investigated for use in a vaccine against nontypeable *H. influenzae*⁴⁻⁶. Antibodies to both purified recombinant P4 and

P4 mutants are highly immunogenic and exhibit bacteriocidal activity against nontypeable *H. influenzae* strains⁶. Moreover, in a nasal challenge mouse model, intranasal immunization with formulations containing recombinant P4 and other proteins resulted in greater than 100x reduction of nasal colonization of nontypeable *H. influenzae* strain SR7332⁵.

Wild type *e*(P4) and the recombinant enzyme (denoted rP4) exhibit potent phosphomonoesterase activity with aryl-phosphate substrates including tyrosine phosphate, phenyl phosphate, *p*-nitrophenyl phosphate, 4 methylumbelliferyl phosphate and NMN^{3, 7}. Dephosphorylation of NMN by P4 is biologically relevant because *H. influenzae* lacks almost all the enzymes necessary for biosynthesis of NAD⁺ and therefore requires NAD⁺, NMN, and nicotinamide riboside as necessary growth factors. The role of P4 in NAD⁺ uptake appears to be the dephosphorylation of NMN to nicotinamide riboside, which is subsequently taken up by the organism⁷. Additionally, P4 is essential for uptake of hemin⁸ under aerobic conditions, although the molecular mechanism of this function remains obscure.

Amino acid sequence analysis has provided insights into the relationship of P4 to other phosphoryl transfer enzymes. P4 belongs to the DDDD superfamily of phosphohydrolases, which is distinguished by a bipartite sequence motif consisting of two pairs of essential aspartic acid residues separated by a linker region of 100 - 200 residues⁹. The DDDD motif of *e*(P4) corresponds to Asp64, Asp66, Asp181 and Asp185. The DDDD superfamily comprises diverse phosphatases from bacteria, archaea and eukaryotes, including nonspecific acid

phosphatases, phosphoglycolate phosphatases, histidinol phosphatases, phosphoserine phosphatases and threose-6-phosphatases⁹.

Within the DDDD superfamily, P4 is the prototype of a group of secreted bacterial enzymes known as class C nonspecific acid phosphatases (NSAPs). Class C NSAPs are distinguished by the bipartite sequence motif [IV]-[VAL]-D-[IL]-D-E-T-[VM]-L-X-[NT]-X(2)-Y in the N-terminus and [IV]-[LM]-X(2)-G-D-[NT]-L-X-D-F in the C-terminus⁹. Genes encoding class C NSAPs have been identified in *Chryseobacterium meningosepticum* (*OlpA*)¹⁰, *Streptococcus equisimilis* (*LppC*)¹¹, *H. influenzae* (*hel*)¹², *Helicobacter pylori* (HP1285)¹³, *Staphylococcus aureus*¹⁴, *Francisella tularensis* (unpublished data), *Pasteurella multocida*¹⁵ and *Bacillus anthracis*¹⁶. Class C enzymes from the first four sources have been purified and characterized, but three-dimensional structures have not been reported. To understand the structural basis of phosphatase activity and hemin binding by P4, and to aid in vaccine development, we have solved the crystal structure of rP4.

7.2 Materials and Methods

7.2.1 Crystallization and Heavy Atom Derivative Preparation

(Research described in this section was performed by Zhonghui Ou and Dr. Jack Tanner)

Expression, purification and crystallization of rP4 have been previously described¹⁷. Briefly, crystals were grown in sitting drops using a reservoir

solution of 0.2 M MgCl_2 , 0.1 M Tris-HCl pH 8.1 - 8.5 and 28 - 36% (w/v) PEG 4000. The space group is $P4_22_12$ with unit cell dimensions of $a = 65.6 \text{ \AA}$, $c = 101.4 \text{ \AA}$ and one rP4 chain per asymmetric unit. The structure was solved using single isomorphous replacement with anomalous scattering (SIRAS) phasing based on a tungstate (Na_2WO_4) derivative. The derivative used for phasing was obtained as follows. A native crystal was transferred into cryobuffer consisting of 0.2M MgCl_2 , 0.1M Tris-HCl pH 8.5, 40% (w/v) PEG 4000 and 15% (w/v) PEG 200. The crystal was then transferred into 20 mL of the cryobuffer supplemented with 50 mM Na_2WO_4 . After soaking in the tungstate solution for 40 min, the crystal was picked up with a Hampton mounting loop and plunged into liquid nitrogen.

7.2.2 X-ray Diffraction Data Collection

Diffraction data were collected at Advanced Light Source beamline 4.2.2. Acquisition of a 1.7 \AA resolution native data set was described previously¹⁷. A data set for the tungstate derivative was collected with $\lambda = 1.0052 \text{ \AA}$ ($E = 12,335 \text{ eV}$), which corresponds to an energy slightly higher than the L_1 absorption edge of W ($E = 12,100 \text{ eV}$). We note that the theoretical value of f'' is approximately 12 electrons at $\lambda = 1.0052 \text{ \AA}$. The tungstate data set consisted of 90° of data collected with an oscillation angle of 1° , exposure time of 5 s/frame, detector distance of 140 mm, and detector angle of zero. The tungstate data set was processed to 2.0 \AA resolution using d*TREK¹⁸. Data collection statistics for the native and derivative data sets are listed in Table 1.

An anomalous difference Patterson map (15 - 3.5 Å resolution) was calculated to assess the suitability of the tungstate data set for phasing. The $w = 0.5$ Harker section displayed a single strong peak (15σ at $(u, v) = (0.15, 0.33)$, (Figure 7.9) suggesting the presence of one high-occupancy tungstate site per asymmetric unit.

7.2.3 Phasing and Refinement Calculations

SIRAS phasing calculations were performed with SOLVE¹⁹ using the tungstate and native data sets. One heavy atom site was identified by SOLVE and resulted in a figure of merit of 0.27 for data to 2.0 Å resolution. Density modification and phase extension to 1.7 Å resolution were performed with RESOLVE¹⁹. The figure of merit from RESOLVE was 0.60 for reflections to 1.7 Å resolution. The density-modified phases were input to ARP/wARP²⁰ for automated backbone tracing. GuiSIDE²⁰ was used to add side chains to the ARP/wARP backbone model. The resulting model included 211 residues with side chains out of the expected 254 residues in the asymmetric unit.

The model from ARP/wARP/GuiSIDE was improved through several rounds of model building in COOT²¹ followed by refinement to 1.7 Å resolution with REFMAC5²². The final model includes residues 11 - 254, one Mg²⁺ ion bound in the active site, 185 water molecules and one PEG fragment. See Table 7.1 for refinement statistics.

The 1.7 Å resolution native structure, with selected active site residues, Mg²⁺ and solvent omitted, served as the starting point for refinement of the

tungstate complex at 2.0 Å resolution. The final model of the tungstate-bound enzyme includes residues 11 - 254, one Mg²⁺ ion, one tungstate ion, 148 water molecules and one PEG fragment. Refinement statistics are listed in Table 7.1.

Table 7.1. Data collection and refinement statistics^a

	Native	WO ₄ ²⁻ Derivative
Wavelength (Å)	1.0359	1.0052
Space Group	P4 ₂ 2 ₁ 2	P4 ₂ 2 ₁ 2
Unit Cell Dimensions	<i>a</i> = 65.6, <i>b</i> = 101.4	<i>a</i> = 65.8, <i>c</i> = 101.9
Number of Proteins/ASU	1	1
Resolution (Å)	42-1.70 (1.79-1.70)	47 - 2.00 (2.07 - 2.00)
Observations	158442	111384
Unique Reflections	25076	15761
Average Redundancy	6.3 (3.6)	7.1 (7.1)
Completeness (%)	99.8 (98.3)	99.7 (99.9)
Mean <i>I</i> / σ (<i>I</i>)	17.5 (2.6)	9.1 (3.8)
<i>R</i> _{merge}	0.071 (0.410)	0.117 (0.416)
Protein Atoms	1942	1934
Water Molecules	183	145
PEG Fragments	1	1
Mg ²⁺ ions	1	1
WO ₄ ²⁻ ions	0	1
<i>R</i> _{cryst}	0.186 (0.262)	0.182 (0.220)
<i>R</i> _{free} ^b	0.222 (0.329)	0.225 (0.282)
RMSD ^c		
Bond Lengths (Å)	0.010	0.013
Bond Angles (°)	1.137	1.313
Ramachandran Plot ^d		
Favored (%)	97.3	97.3
Allowed (%)	3.7	3.7
Average B-factors (Å ²)		
Protein	23.6	19.1
Mg ²⁺ ion	15.0	11.1
PEG Fragment	37.6	27.6
Water Molecules	27.2	22.1
WO ₄ ²⁻ ion		9.56
PDB accession code	2HLK	2HLL

^aValues for the outer resolution shell are in parenthesis.^b5 % random test set.^cCompared to Engh and Huber force field²³.^dThe Ramachandran plot was generated with RAMPAGE²⁴.

7.3 Results

7.3.1 P4: a new member of the HAD superfamily

The 1.7 Å resolution structure of rP4 was solved using single isomorphous replacement with anomalous scattering (SIRAS) phasing based on a high-occupancy single-site tungstate derivative. The structure of the tungstate-bound enzyme was subsequently solved at 2.0 Å resolution using the native structure as the starting model.

The rP4 structure comprises two structural domains: an α/β core domain and an α domain (Figures 7.1A, 7.1B). The active site is located in a cleft between the two domains, as denoted by the tungstate and Mg^{2+} ions in Figure 1A. The α/β core domain (residues 31-70, 106-216) consists of a central five-stranded parallel β -sheet flanked by three α -helices (αB , αE , αF) on one side and two α -helices (αG , αH) on the other side. The β -sheet has strand order of 3,2,1,4,5 (Figure 7.1C). Note that the topology of the α/β core domain resembles that of the Rossmann dinucleotide-binding fold.

The α domain is located above the carboxyl-terminal edge of the strands of the core β -sheet and consists of three components. The first component is a long α -helix at the N-terminus of the polypeptide chain (αA , residues 11-30). This helix connects to another long helix (αB), which is part of the core domain, and together αA and αB form a long helical elbow that spans the entire length of rP4 (Figure 7.1B). The second component of the α domain is a helix-loop-helix

substructure (α C and α D, residues 71 - 105) inserted between β 1 and α E of the core domain. The third component consists of 38 residues at the C-terminus (residues 217-254) that form two helices (α I, α J) followed by an extension. As discussed below, the α domain is involved in dimerization.

Comparison of the rP4 structure against the Protein Data Bank (PDB)²⁵ revealed that rP4 is a member of the haloacid dehalogenase (HAD) enzyme superfamily. The closest structural homologs, based on a search of the PDB with the DALI²⁶, include the class B NSAP AphA (DALI Z-score = 9.8, PDB code 2B8J²⁷), bacteriophage T4 polynucleotide kinase (Z = 9.7, PDB code 1LTQ²⁸), mitochondrial 5'(3')-deoxyribonucleotidase (Z = 8.1, PDB code 1Z4Q²⁹), magnesium-dependent phosphatase-1 from *Mus musculus* (Z = 7.6, PDB accession code 1U7P³⁰) and *Bacillus cereus* phosphonoacetaldehyde hydrolase (Z = 7.6, PDB accession code 1FEZ³¹). These enzymes are members of the HAD superfamily, and they share with rP4 a common α/β core domain (Figure 7.2A) and active site (Figure 7.2B), despite negligible overall amino acid sequence identity to rP4 (< 19% for aligned residues).

HAD superfamily enzymes are organized into three structural subclasses based on the presence (subclasses I, II) or absence (subclass III) of a cap domain located on the C-terminal edge of the conserved β -sheet³². Subclass I enzymes have a small α helical bundle cap domain, whereas subclass II enzymes have a larger β -sandwich cap domain³³. Subclass III encompasses HAD enzymes lacking subclass I or II cap domains. In rP4, the α domain substitutes for the cap domain. Since the α domain resembles neither the helical bundle of subclass I

enzymes nor the β -sandwich of subclass II enzymes, rP4 is best described as a class III HAD enzyme. We note that the close homologs of rP4 identified by DALI are all members of HAD subclass III.

Figure 7.1

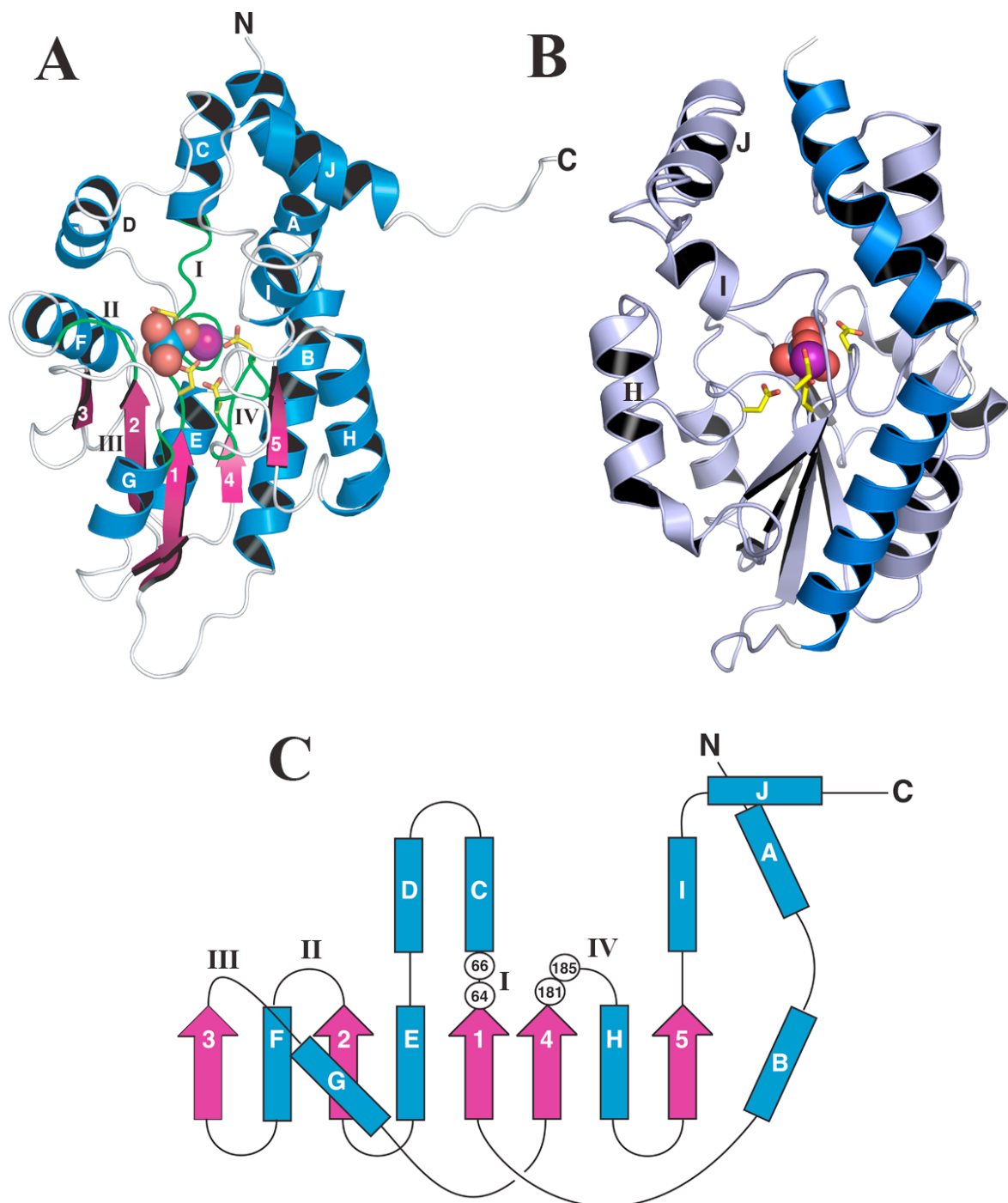


Figure 7.1. Overall structure of rP4

A) Ribbon drawing of the rP4 subunit. The tungstate inhibitor (blue/red) and the bound Mg^{2+} ion (violet) are drawn as spheres. Strands are colored in pink and labeled 1-5. Helices are colored blue and labeled A-J. Active site loops are colored green and labeled I-IV. The N and C terminus of the enzyme are labeled appropriately. **B)** Another view of the rP4 subunit highlighting the long helical elbow formed by αA and αB . **C)** Secondary structure topology diagram of rP4. Blue rectangles denote α -helices. Pink arrows denote β -strands. The four conserved aspartic acid residues of the DDDD superfamily are denoted by white circles.

Figure 7.2

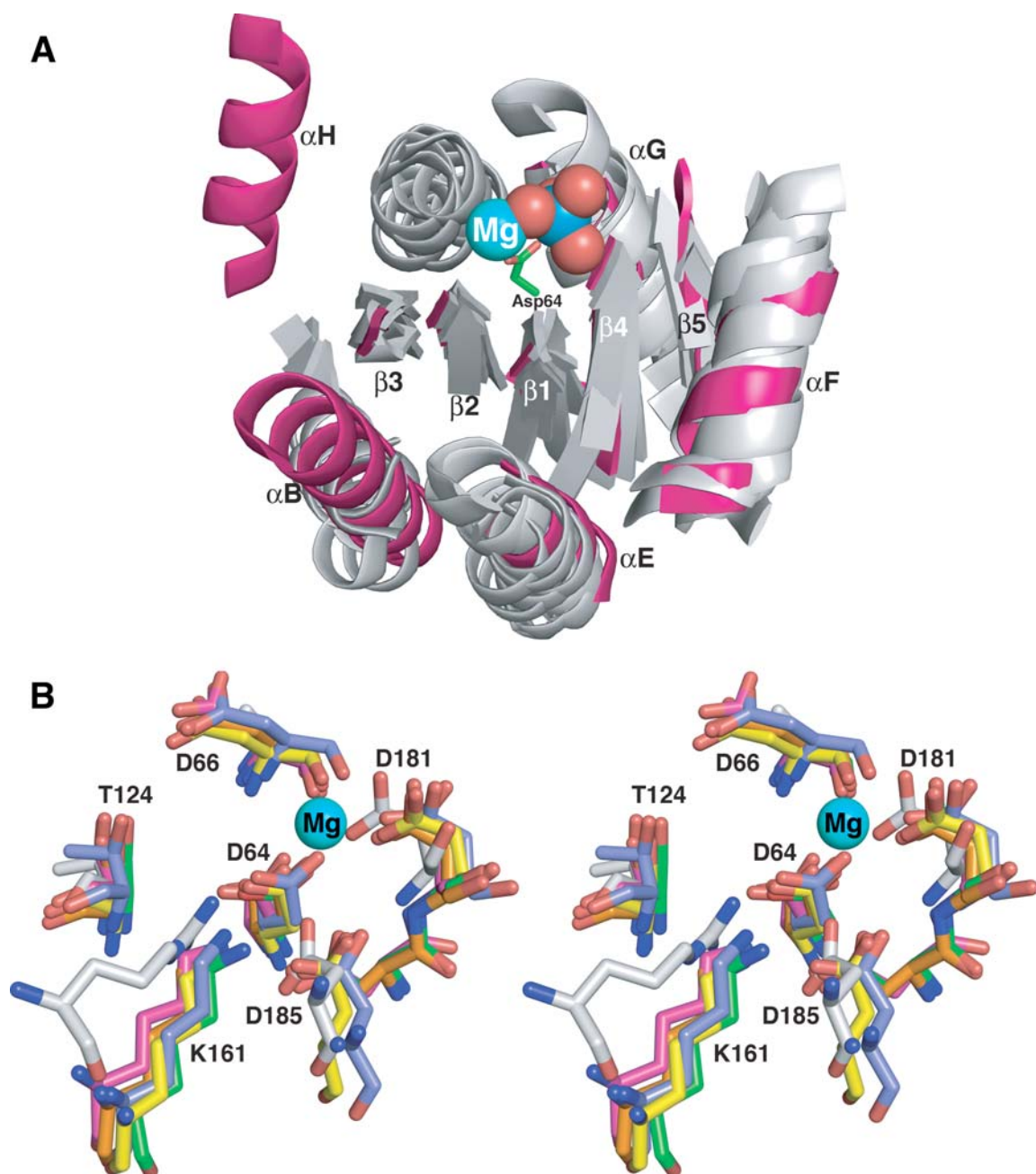


Figure 7.2. Comparison of rP4 with 4 other HAD superfamily enzymes.

The structures used in this figure are 1LTQ²⁸, 2B8J²⁷, 1U7P³⁰, and 1Z4Q²⁹.

A) Stereo view of the α/β core domains, with rP4 shown in pink and the other HAD superfamily enzymes in white. The Mg^{2+} ion (blue) and the tungstate (blue/red) of rP4 are drawn as spheres. The proposed nucleophile, Asp64, is shown in stick form and is located at the C terminus of strand $\beta 1$. The strands and helices are labeled in accordance with the topology diagram in Figure 7.1. **B)** Stereo view of the active sites, with rP4 shown in yellow and the enzymes shown in various colors. The Mg^{2+} ion (blue) and residue labels correspond to rP4.

7.3.2 Structure of the active site

The active site of rP4 is located in a cleft between the α/β core domain and the α domain. The cleft has approximate dimensions of 29 Å x 25 Å x 22 Å and is accessible to solvent. The tungstate ion binds in the lower part of the cleft, atop the carboxyl terminal edge of the β -sheet. There is a large space in the upper part of the cleft that is possibly the binding site for the non-phosphoryl groups of substrates (Figure 7.1A).

Following the nomenclature used for HAD superfamily enzymes, the active site consists primarily of four catalytic loops, labeled I - IV (Figure 7.1A). The catalytic loops of rP4 connect secondary structure elements β 1 and α C (loop I, residues 63-73), β 2 and α F (loop II, residues 124-132), β 3 and α G (loop III, residues 155-161), and β 4 and α H (loop IV, residues 180-186).

Electron density maps clearly indicated the presence of a metal ion in the active site (Figure 7.3A), which is consistent with the fact that most HAD superfamily members are metalloenzymes³². Structure refinement with Mg^{2+} modeled in the active site produced F_o-F_c electron density maps having negligible residual density at the metal site and a reasonable B-value for the metal ion (16 Å²). It is concluded that Mg^{2+} is the most likely candidate for the metal ion in rP4 expressed in *E. coli*.

The Mg^{2+} ion has octahedral geometry and the conserved DDDD motif plays a prominent role in metal binding (Figure 7.3). The protein provides three of the six ligands for the metal ion, which are side chains of Asp64 and Asp181, and the main chain carbonyl of Asp66. All three residues are part of the DDDD motif.

Water molecules are the other three ligands in the inhibitor-free conformation. These water molecules bridge the metal ion with Asp64, Asp181, Asp185 and Tyr 221.

The tungstate-bound structure provides insights into the roles of active site residues because tungstate mimics the phosphate moiety of phosphomonoester substrates. The tungstate ion is bound by the Mg^{2+} ion and catalytic loops I-III. The oxygen atoms of the tungstate interact with the side chains of Asp66, Thr124 and Lys161 (Figures 7.3B, 7.3C), as well as the backbone amine groups of Leu65, Asp66 and Asn125. The binding of tungstate displaces four water molecules but causes no major conformational changes in the protein. The four displaced water molecules occupy the same locations in the native structure as the tungstate oxygen atoms in the inhibitor-bound structure.

The rP4/tungstate structure suggests that Asp64 is the nucleophile that attacks the P atom of the substrate. A carboxyl oxygen atom of Asp64 is only 2.4 Å from the W atom of the inhibitor, which is the closest contact observed between the protein and the W atom (Figures 7.4B, 7.4C). Moreover, the other carboxyl oxygen atom of Asp64 is bound to the Mg^{2+} ion, which would depress the pK_A and increase the nucleophilicity of Asp64.

The constellation of residues in the active site of rP4 is similar to those of close homologs from the HAD superfamily, as shown in Figure 7.2B. In particular, there is high structural similarity among these enzymes in the location of the metal ion and the four Asp residues of the DDDD motif. Moreover, two residues that interact with the substrate phosphoryl group, corresponding to

Thr124 and Lys161 of rP4, are also structurally conserved by this group of enzymes.

Figure 7.3

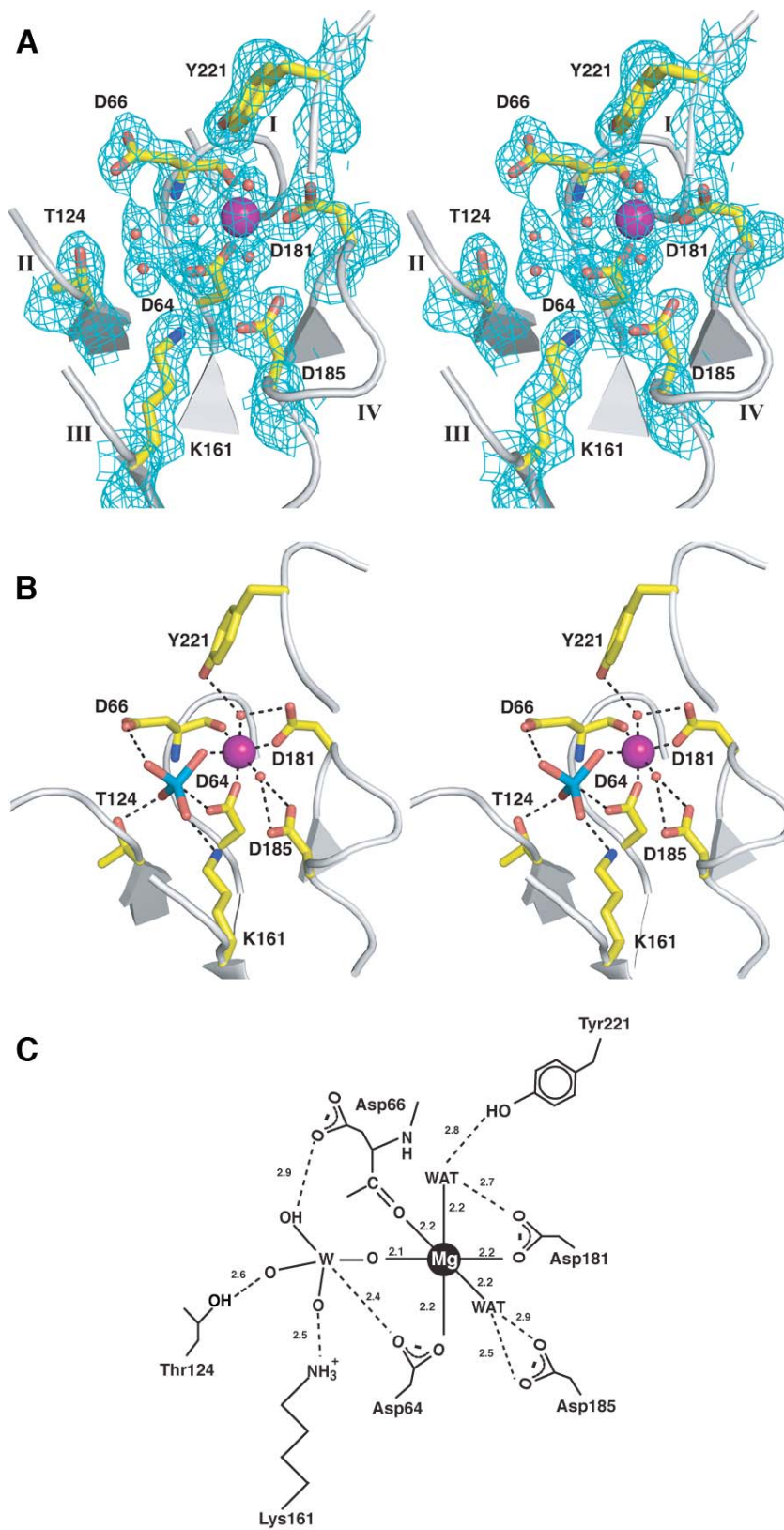


Figure 7.3. Active site of rP4

A) Stereographic drawing of the rP4 active site with the Mg^{2+} ion drawn in violet. Electron density is from simulated annealing F_o-F_c map. Residues of the active site were omitted prior to simulated annealing refinement in CNS. The map is contoured at 1.0σ . **B)** Stereographic drawing of the rP4 active site highlighting the protein-inhibitor electrostatic interactions (dashed lines). The tungstate ion is shown in blue/red and the Mg^{2+} ion is in violet. **C)** Schematic diagram of the active site. Dotted lines indicate electrostatic interactions.

7.3.3 Dimeric Structure of rP4

Purified rP4 forms an apparent homodimer in solution based on gel filtration chromatography³⁶. The asymmetric unit contains one rP4 chain, so interfaces generated by crystallographic symmetry were analyzed with the PISA server³⁷ to gain insights into the dimeric structure of rP4. The largest interface in the crystal lattice is generated by 2-fold rotation around the crystallographic c-axis. This interface buries 3,053 Å² of surface area, while the other interfaces each bury less than 350 Å² of surface area. There are 44 hydrogen bonds and 4 ion pairs in this interface, implying substantial electrostatic complementarity. Based on these data, we propose that the assembly generated by 2-fold rotation around the c-axis represents the dimer formed in solution (Figure 7.4A).

The primary dimer interface is formed by α A, α B, α H, β 5, α I and α J. These elements form a large groove containing two pronounced holes separated by a crosspiece (Figure 7.4B). The crosspiece is formed by α H and its preceding loop. In the dimer, the α A- α B elbow of one subunit packs into the groove of the other subunit (Figure 7.4C), which allows side chains of α A and α B to poke into the holes of the groove and interact with the active site of the other subunit. The side chains of Glu32 and Leu22 seem to be particularly important in this regard, as described below.

In addition to this large interface, there is a smaller intersubunit interface formed by interaction of the C-terminal extension (residues 248-254) of one subunit with the helix-loop-helix insert of the other subunit (Figure 7.4A). Here,

Trp251 of the extension packs against α C and α D, making nonpolar contacts with the side chains of Tyr75, Trp78, Arg93 and Trp94.

Dimerization appears to be important for catalysis because side chains from α A and α B of one subunit poke through the holes in the dimer interface into the active site of the opposite subunit. Most notably, Glu32 forms direct and indirect electrostatic interactions with catalytic loop IV of the opposite subunit (Figure 7.5A). The direct interaction is a hydrogen bond between the carboxyl of Glu32 and the amine of Leu183 of loop IV. In the indirect interaction, Glu32 forms an intersubunit ion pair with Arg198, which in turn, interacts with Asn182 and Asp184 of loop IV. We note that Glu32 and Arg198 are highly conserved among class C NSAPs. These interactions most likely help define and stabilize the conformation of catalytic loop IV, which contributes two of the metal-binding ligands, Asp181 and Asp185.

The second place where dimerization appears to be important for catalysis is the upper part of the active site cleft in the region bounded by the helix-loop-helix insert on one side and by α A, α I and α J on the other side. In this region, Leu22 of one subunit pokes into the active site of the other subunit and fills a hole surrounded by α A, α I and α J. As discussed above, it is possible that the upper part of the active site is involved in binding substrates and so this intersubunit contact may be important for substrate recognition.

Figure 7.4

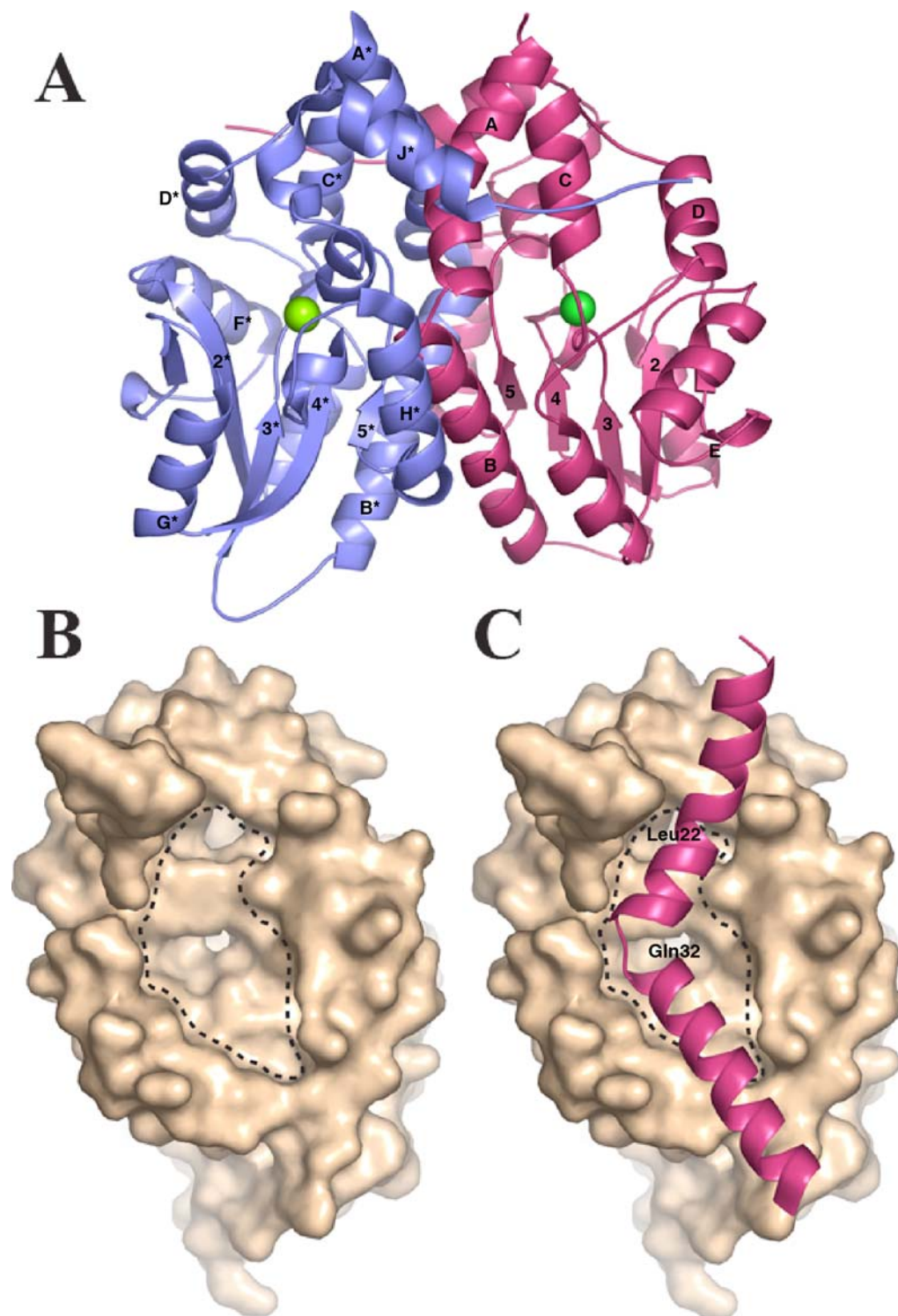


Figure 7.4. Dimeric structure of rP4

A) Ribbon drawing of the rP4 dimer. The two subunits are colored blue and pink. **B)** The large groove that forms intersubunit contacts with the helical elbow of the opposite subunit. Note the two large holes in the groove, which lead to the active site. **C)** Depiction of the helical elbow of one subunit fitting into the groove of the opposite subunit.

Figure 7.5

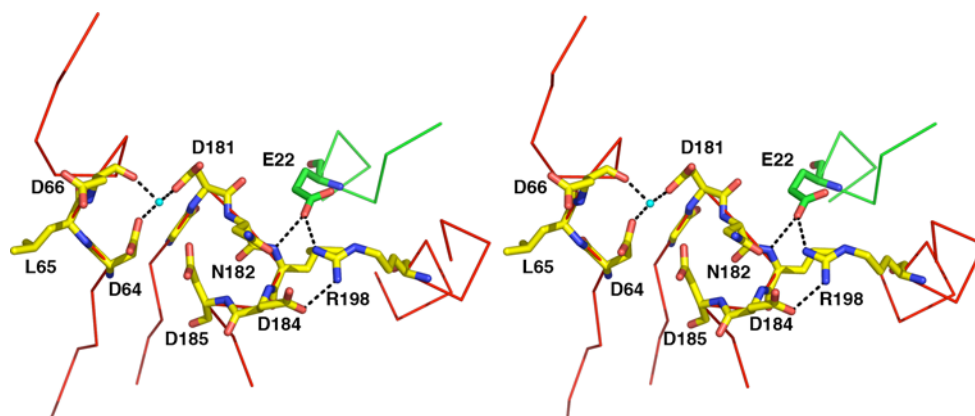


Figure 7.5. Stabilization of the active site through intersubunit interactions.

Intersubunit interactions formed between Glu32 and catalytic loop IV. Active site residues of one subunit are colored yellow. Residues of the other subunit are colored green. The Mg^{2+} ion is colored cyan.

7.3.4 Structural context of rP4 mutations designed for vaccine studies

Recombinant P4 is a promising vaccine candidate against nontypeable *H. influenzae* infections and the enzyme has been extensively mutated in an effort to engineer catalytically inactive proteins for use in a vaccine. Green and coworkers recently measured the immune responses of mice vaccinated individually with 12 different rP4 mutants³. Three mutants, Asn218Gln, Gln39Glu, Phe48Cys, were notable because they displayed the best combination of reduced phosphatase activity and favorable immune response.

Interestingly, the structure shows that none of these three residues participate directly in catalysis or substrate binding. Asn218 is located in the loop that connects $\beta 5$ to αI . It forms a hydrogen bond with Asp181, which is one of Mg^{2+} ligands (Figure 7.3B). Thus, the role of Asn218 may be to help orient Asp181 for interaction with the metal ion. Mutation of Asn218 to Glu will cause electrostatic repulsion with Asp181, which would likely alter the conformation of Asp181 and adversely affect metal binding.

Gln39 is located in the middle of αB and forms hydrogen bonds with $\beta 5$ of the opposite subunit (Figure 7.7B). The side chain carbonyl and amine of Gln39 form intersubunit hydrogen bonds with the backbone amine and carbonyl, respectively, of Met215. Mutation of Gln39 to Glu would disrupt this intersubunit hydrogen bonding network. In addition, because αB is close to and nearly parallel with the 2-fold axis of the dimer, Gln39 side chains of opposite subunits are only 5 Å apart. Thus, mutation of Gln39 to Glu buries two carboxyl

groups close together in the dimer interface. It is concluded that the diminished activity of Gln39Glu results from disruption of the dimer interface.

Lastly, Phe48 is located at the C-terminus of α B and its side chain makes nonpolar contacts with Val61, Val113 and Lys59 (Figure 7.7C). The side chain of Lys59 reaches over the phenyl ring of Phe48 to form a hydrogen bond with the carbonyl of Phe48 thus capping α B. Mutation of Phe48 to Cys would certainly affect the tight packing in this region. Since α B is involved in dimerization, this mutation, like Gln39Glu, may diminish activity by disrupting the dimer interface.

7.3.5 Structural context of the proposed KVAFDH heme binding motif

P4 is essential for aerobic growth of *H. influenzae* by mediating uptake of heme from the environment^{38, 39}. The heme binding domain was originally proposed by Reidl and Mekalanos³⁸ as the short sequence motif KVAFDH in residues 45 - 50 due to its similarity to the heme binding motifs found in hemoglobin, cytochrome *c*₃, HbpA and HAP-1.

Residues 45 - 50 are located on the C-terminus of α B (Figure 7.6). Because this helix is near the 2-fold axis of the dimer, the α B helices of the two subunits pack against each other in the dimer interface. As a result, the two KVAFDH motifs of the dimer form a V-shaped intersubunit cleft with the side chains of Val46 and His50 pointing into the cleft. The two His50 residues of the motif face each other with imidazole N atoms separated by 5.3 Å. It is tempting to speculate that a molecule of heme could bind in this cleft edgewise and engage in bis-

histidinyll coordination to His50 in a manner similar to that found in cytochrome *C₃*.

Figure 7.6

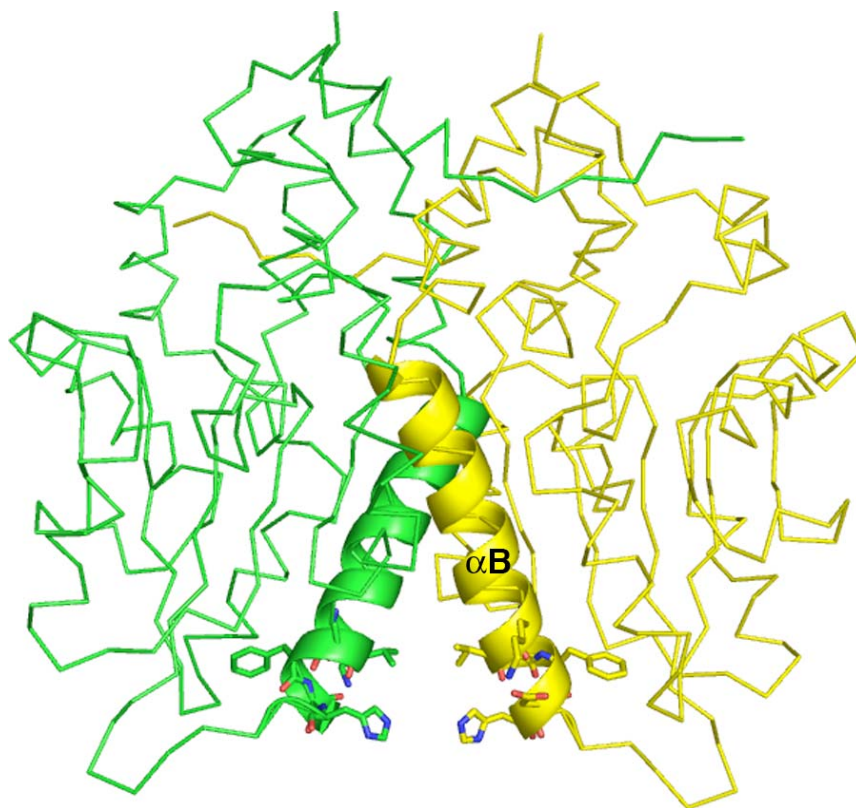


Figure 7.6 The KVAFDH hemin binding motif.

The KVAFDH sequence is located at the C-terminus of helix αB . The residues are drawn as sticks and colored according to their respective subunit.

7.4 Discussion

7.4.1 Structure of rP4 and relationships to other HAD enzymes

The structure of rP4 revealed a two-domain architecture consisting of the quintessential HAD superfamily α/β core domain and a unique α domain that is involved in dimerization. The rP4 active site features the 4-loop structure and metal binding site that are characteristic of the HAD superfamily. The structure of rP4 complexed with tungstate suggested that Asp64 is the nucleophile that attacks the P atom of the substrate, which is expected for a HAD superfamily member. Thus, P4 is a fairly typical example of a HAD superfamily enzyme in terms of the α/β core domain structure and active site composition.

On the other hand, the α domain of P4 appears to be unique. This domain mediates a dimeric structure that distinguishes P4 from other HAD and DDDD enzymes. In particular, the protein-protein interaction surface observed in the rP4 dimer is not seen in AphA, the closest structural homolog of rP4 in the PDB and prototypical class B NSAP. Class B NSAPs form homotetramers in which the subunits interact via intermolecular β -sheet formation involving an N-terminal extension⁴⁰. It thus appears that the mode of self-association is the major structural difference between class B and C NSAPs.

We have recently solved structures of class C enzymes from *Francisella tularensis* and *Bacillus anthracis* (unpublished results). These enzymes exhibit the same subunit fold and dimeric structure as rP4. Detailed comparison of our three class C NSAP structures with those of class B enzymes should provide new

insights into the structural and evolutionary relationships between class B and C NSAPs.

7.4.2 Implications for vaccine design

Considerable effort is currently focused on designing mutants of rP4 for use in a vaccine against nontypeable *H. influenzae* infections³⁻⁵. An important design constraint is that the protein should lack catalytic activity because vaccination with an active phosphatase could affect host metabolism and signalling pathways. The four aspartate residues of the DDDD motif are obvious candidates for mutagenesis since these residues directly participate in catalysis and metal binding. Interestingly, Green and coworkers showed that such Asp mutants, while devoid of activity, do not elicit the best immune responses compared to the mutants Asn218Gln, Gln39Glu and Phe48Cys³.

We used the rP4 structures to understand why Asn218Gln, Gln39Glu and Phe48Cys have diminished activity. This analysis showed that these mutations likely decrease activity through indirect mechanisms, since none of these residues directly bind the active site metal ion or tungstate inhibitor. Asn218 participates indirectly in metal binding via a hydrogen bond interaction with a metal binding ligand, Asp181. On the other hand, we reasoned that the mutations Gln39Glu and Phe48Cys adversely affect activity by disrupting the dimer interface.

These results suggest a structure-based vaccine design strategy for rP4 in which the residues targeted for mutagenesis are located (1) on the periphery of the active site or (2) in the dimer interface. In the former category, good

candidates are Thr68, Asn72 and Arg126. Thr68 forms hydrogen bonds with two residues of the DDDD motif, Asp64 and Asp181, while Asn72 and Arg126 interact with the DDDD motif residue Asp66. The dimer interface is large and offers many possibilities for mutagenesis. Leu22 and Glu32 are two intriguing candidates since they stabilize the active site via intersubunit contacts. Also, Trp26 side chains from the two subunits interact with each other. Mutation to Ala would likely result in a destabilizing cavity buried in the dimer interface. Finally, a truncation mutation that removes the C-terminal extension may have a significant effect on dimerization and hence activity.

7.4.3 Role of e (P4) in hemin acquisition by *H. influenzae*

H. influenzae requires exogenous hemin for growth under aerobic conditions. Lipoprotein e(P4) plays an essential role in hemin acquisition by the organism, based on a variety of bacterial growth studies. Reidl and Mekalanos³⁸ showed that mutant strains of *H. influenzae* lacking a functional e(P4) gene (*hel*) showed almost no growth under aerobic conditions in the presence of hemin. Complementation of the mutant strain with a plasmid carrying the *hel* gene rescued this growth defect. Studies using *hemA* mutants of *E. coli* also point to a role for e(P4) in hemin acquisition. The *hemA* gene encodes the porphyrin biosynthetic enzyme Glu-tRNA-reductase, which catalyzes the synthesis of 5-amino-levulinic acid from Glu-tRNA. *E. coli hemA* mutants fail to grow aerobically in hemin-enriched media, indicating that they are unable to acquire hemin from the environment to overcome the defect in porphyrin biosynthesis. At least two studies have shown that *hemA* mutants transformed with a plasmid harboring the *hel* gene grow aerobically in hemin-enriched media^{38, 39}. These studies strongly suggest that P4 is essential for hemin acquisition, but the molecular details of this function are unknown. For example, it is not known whether e(P4) binds directly to hemin or whether it functions in a more indirect way, for example, by interacting with another factor or increasing the permeability of the outer membrane.

The binding site for hemin is unknown. Reidl and Mekalanos proposed that residues 45 - 50 form the hemin binding site³⁸. The rP4 structure shows that

these residues appear in the dimer interface and form a cleft that is suggestive of bis-histidinyI coordination of hemin.

An alternative hypothesis is that hemin binds in the active site. Data supporting this hypothesis comes from Reilly *et al.* who performed growth studies of *hemA* mutants transformed with plasmids containing mutant rP4 genes³⁹. They showed that mutation of Asp64 or Asp66 to Ala compromises the ability of rP4 to rescue the growth defect of an *E. coli hemA* mutant in hemin-enriched media. These two aspartate residues interact with the phosphoryl moiety of the substrate based on our structure of rP4 complexed with tungstate, which supports the idea that hemin binds in the active site, possibly mimicking the binding of substrate. The rP4 structure will provide a foundation for elucidating the details of P4-hemin interaction and for understanding the biological role of P4 in hemin acquisition.

Figure 7.7

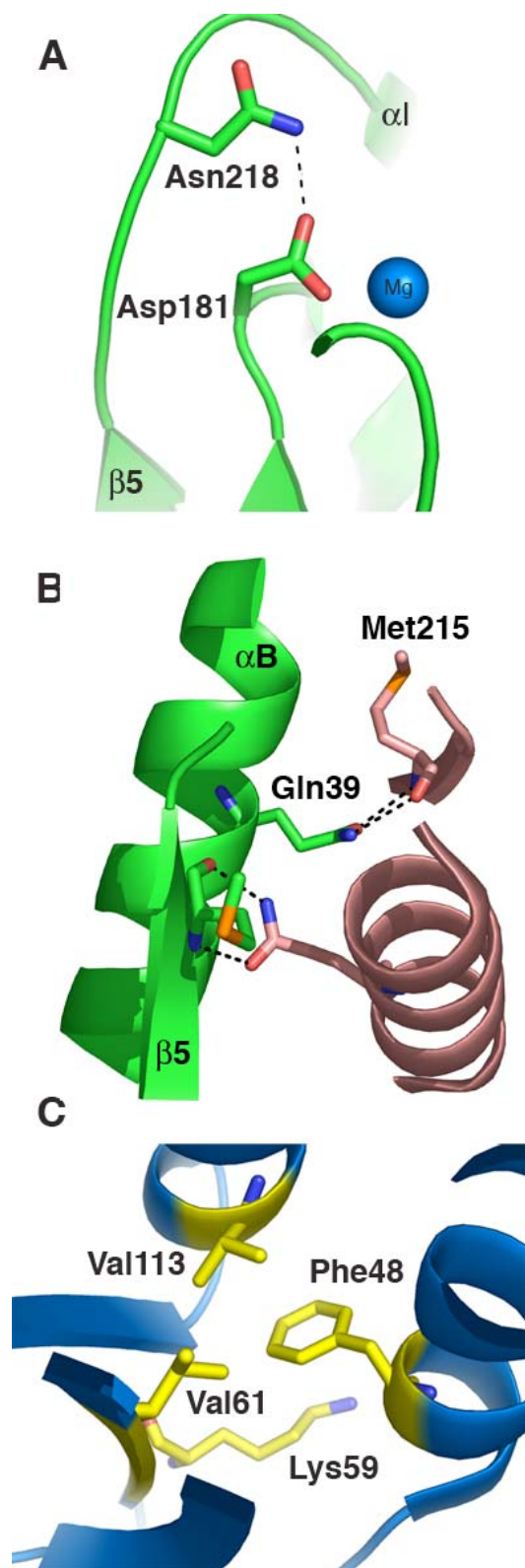


Figure 7.7. Structural context of rP4 mutations (a) Asn218Gln, (b) Gln39Glu and (c) Phe48Cys.

Figure 7.8

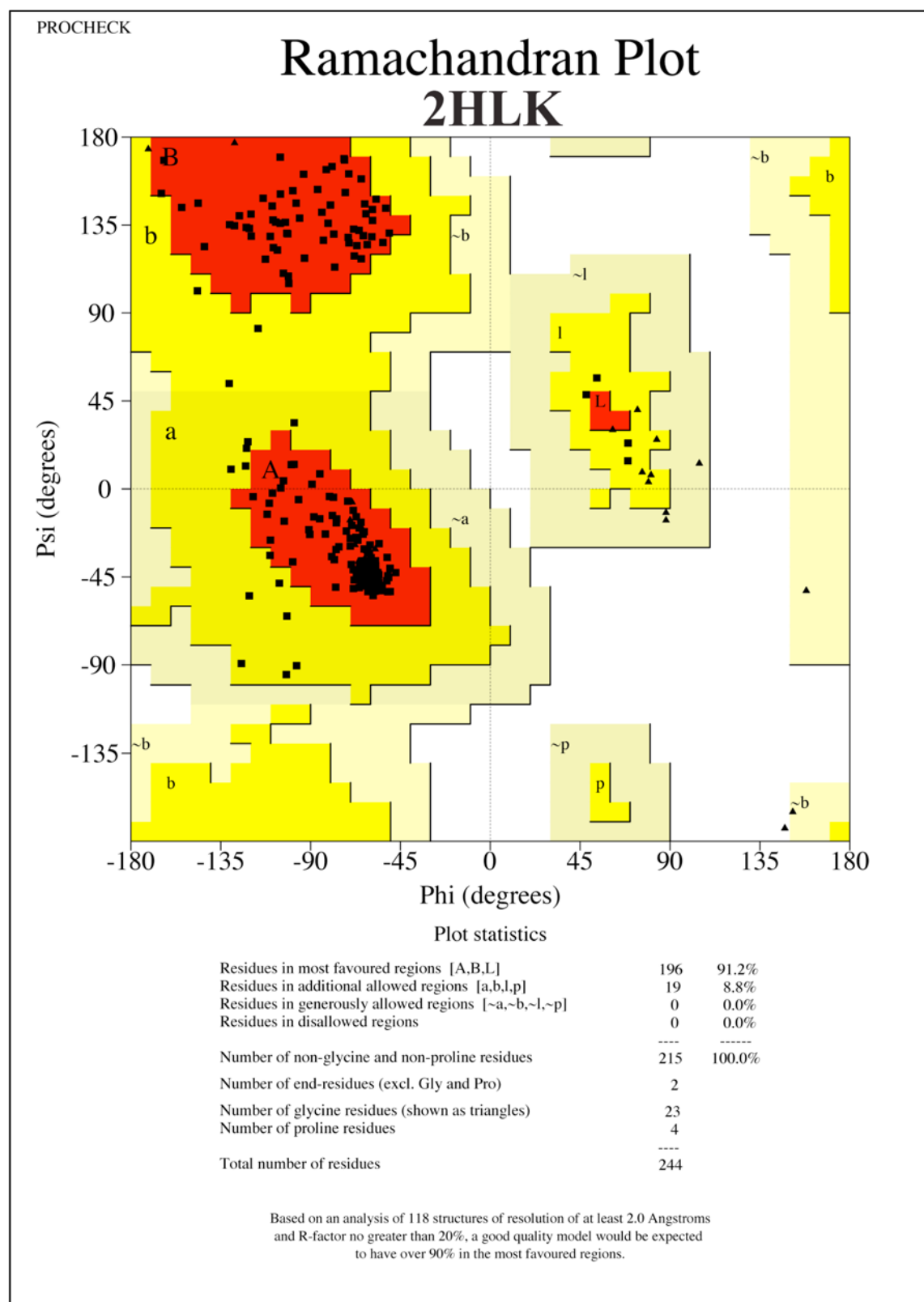


Figure 7.9

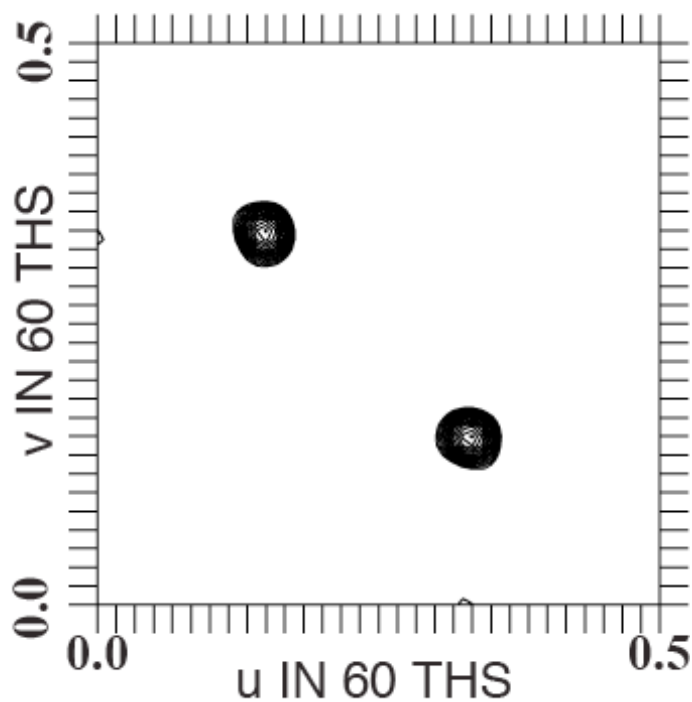


Figure 7.9 Anomalous difference Patterson Map

The $w=0.5$ Harker section of the anomalous difference Patterson map for the tungstate derivative.

7.5 References

1. Foxwell, A. R.; Kyd, J. M.; Cripps, A. W., Nontypeable *Haemophilus influenzae*: pathogenesis and prevention. *Microbiol. Mol. Biol. Rev.* **1998**, 62, (2), 294-308.
2. Loeb, M. R.; Smith, D. H., Outer membrane protein composition in disease isolates of *Haemophilus influenzae*: pathogenic and epidemiological implications. *Infect. Immun.* **1980**, 30, (3), 709-717.
3. 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.
4. 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.
5. 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.
6. Oropeza, V. C.; Page, M. E.; Van Bockstaele, E. J., Systemic administration of WIN 55,212-2 increases norepinephrine release in the rat frontal cortex. *Brain Res* **2005**, 1046, (1-2), 45-54.
7. 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), 3974-81.
8. Reidl, J.; Schlor, S.; Kraiss, A.; Schmidt-Brauns, J.; Kemmer, G.; Soleva, E., NADP and NAD utilization in *Haemophilus influenzae*. *Mol. Microbiol.* **2000**, 35, (6), 1573-1581.

9. 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.
10. 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.
11. Malke, H., Cytoplasmic membrane lipoprotein LppC of *Streptococcus equisimilis* functions as an acid phosphatase. *Appl. Environ. Microbiol.* **1998**, 64, (7), 2439-42.
12. Green, B. A.; Farley, J. E.; Quinn-Dey, T.; Deich, R. A.; Zlotnick, G. W., The e (P4) outer membrane protein of *Haemophilus influenzae*: biologic activity of anti-e serum and cloning and sequencing of the structural gene. *Infect. Immun.* **1991**, 59, (9), 3191-3198.
13. Godlewska, R.; Bujnicki, J. M.; Ostrowski, J.; Jagusztyn-Krynicka, E. K., The hppA gene of *Helicobacter pylori* encodes the class C acid phosphatase precursor. *FEBS Lett.* **2002**, 525, (1-3), 39-42.
14. 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.
15. 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. U S A* **2001**, 98, (6), 3460-5.
16. Read, T. D.; Peterson, S. N.; Tourasse, N.; Baillie, L. W.; Paulsen, I. T.; Nelson, K. E.; Tettelin, H.; Fouts, D. E.; Eisen, J. A.; Gill, S. R.; Holtzapple, E. K.; Okstad, O. A.; Helgason, E.; Rillstone, J.; Wu, M.; Kolonay, J. F.; Beanan, M. J.; Dodson, R. J.; Brinkac, L. M.; Gwinn, M.; DeBoy, R. T.; Madpu, R.; Daugherty, S. C.; Durkin, A. S.; Haft, D. H.; Nelson, W. C.; Peterson, J. D.; Pop, M.; Khouri, H. M.; Radune, D.; Benton, J. L.; Mahamoud, Y.; Jiang, L.; Hance, I. R.; Weidman, J. F.; Berry, K. J.; Plaut, R. D.; Wolf, A. M.; Watkins, K. L.; Nierman, W. C.; Hazen, A.; Cline, R.; Redmond, C.; Thwaite, J. E.; White, O.; Salzberg, S. L.; Thomason, B.; Friedlander, A. M.; Koehler, T. M.; Hanna, P. C.; Kolsto, A. B.; Fraser, C. M., The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **2003**, 423, (6935), 81-6.

17. Ou, Z.; Felts, R. L.; Reilly, T. J.; Nix, J. C.; Tanner, J. J., Crystallization of recombinant *Haemophilus influenzae* e (P4) acid phosphatase. *Acta Crystallogr.* **2006**, F62, (Pt 5), 464-6.
18. Pflugrath, J. W., The finer things in X-ray diffraction data collection. *Acta Crystallogr.* **1999**, D55, 1718-1725.
19. Terwilliger, T. C., SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **2003**, 374, 22-37.
20. Morris, R. J.; Perrakis, A.; Lamzin, V. S., ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr.* **2002**, D58, (Pt 6 Pt 2), 968-75.
21. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **2004**, D60, (Pt 12 Pt 1), 2126-32.
22. 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.
23. Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* **1991**, A47, (4), 392-400.
24. 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.
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. *Nucl. Acids Res.* **2000**, 28, 235-242.
26. Dietmann, S.; Park, J.; Notredame, C.; Heger, A.; Lappe, M.; Holm, L., A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3. *Nucleic Acids Res.* **2001**, 29, (1), 55-7.
27. 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.

28. Galburt, E. A.; Pelletier, J.; Wilson, G.; Stoddard, B. L., Structure of a tRNA repair enzyme and molecular biology workhorse: T4 polynucleotide kinase. *Structure* **2002**, 10, (9), 1249-60.
29. 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.
30. Peisach, E.; Selengut, J. D.; Dunaway-Mariano, D.; Allen, K. N., X-ray crystal structure of the hypothetical phosphotyrosine phosphatase MDP-1 of the haloacid dehalogenase superfamily. *Biochemistry* **2004**, 43, (40), 12770-9.
31. Morais, M. C.; Zhang, W.; Baker, A. S.; Zhang, G.; Dunaway-Mariano, D.; Allen, K. N., The crystal structure of *Bacillus cereus* phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily. *Biochemistry* **2000**, 39, (34), 10385-96.
32. Allen, K. N.; Dunaway-Mariano, D., Phosphoryl group transfer: evolution of a catalytic scaffold. *Trends Biochem. Sci.* **2004**, 29, (9), 495-503.
33. Shin, D. H.; Roberts, A.; Jancarik, J.; Yokota, H.; Kim, R.; Wemmer, D. E.; Kim, S. H., Crystal structure of a phosphatase with a unique substrate binding domain from *Thermotoga maritima*. *Protein Sci.* **2003**, 12, (7), 1464-72.
34. Coleman, J. E., Structure and mechanism of alkaline phosphatase. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, 21, 441-83.
35. Knowles, J. R., Enzyme-catalyzed phosphoryl transfer reactions. *Annu. Rev. Biochem.* **1980**, 49, 877-919.
36. 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.
37. Krissinel, E.; Henick, K., Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). *Comp. Life* **2005**, 163-174.

38. Reidl, J.; Mekalanos, J. J., Lipoprotein e(P4) is essential for heme uptake by *Haemophilus influenzae*. *J. Exp. Med.* **1996**, 183, (2), 621-629.
39. 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.
40. Calderone, V.; Forleo, C.; Benvenuti, M.; Cristina Thaller, M.; Maria Rossolini, G.; Mangani, S., The first structure of a bacterial class B Acid phosphatase reveals further structural heterogeneity among phosphatases of the haloacid dehalogenase fold. *J. Mol. Biol.* **2004**, 335, (3), 761-73.

APPENDIX

A1.

EXPRESSION, PURIFICATION, CRYSTALLIZATION, AND STRUCTURE
DETERMINATION OF FRANCISELLA TULARENSIS CLASS C NON-SPECIFIC
ACID PHOSPHATASE

A1.1 Experimental

A1.1.1 Expression and purification

A single colony of *Escherichia coli* BL21(DE3) cells containing the plasmid encoding the *F. tularensis* class C non-specific acid phosphatase (NSAP) was used to inoculate 5 mL LB containing ampicillin at a concentration of 50 µg/mL and incubated overnight at 37 °C with constant aeration. A 1:1000 dilution of this overnight culture was used to inoculate 25 mL of fresh LB. The sample was incubated as before until the optical density reached 0.6. The sample was then chilled on ice for 10 minutes and placed at 4 °C for overnight storage. Prior to expression scale up the 25 mL starter culture was spun at 3,660 *g* for 10 minutes at 4 °C. The pellet was then re-suspended in fresh LB. The sample was then distributed into 1.8 L of LB containing 50 µg/mL ampicillin and 0.2% glucose. The culture was grown at 37 °C with constant aeration until the optical density reached 0.4. At this time IPTG was added to a final concentration of 0.4 mM. The cells were harvested when the optical density was between 1.0 and 1.5. This normally took between three and four hours. The cells were pelleted by centrifugation at 5,000 *g* for 10 minutes. The pellet was then resuspended in 25 mL of 50 mM sodium acetate pH6.0.

After resuspension of the cells into 50 mM sodium acetate they were broken using a French Pressure Cell adjusted to 10,000 psi and a flow rate of 20 drops/minute. The extracts from French Press were stirred in the presence of 1 M NaCl for 30 minutes. Unbroken cells and pelletable debris were removed by

centrifugation at 30,000 *g* for 15 minutes. Bacterial membranes were pelleted by ultracentrifugation at 192,000 *g* for 1 hour. The supernatant of the ultracentrifugation step contained the majority of acid phosphatase activity as measured using the discontinuous assay described in Chapter 6. The supernatant was then dialyzed against 100 volumes of 50 mM Tris, pH 8.4 for 24 hours.

The dialyzed sample was applied (5 mL/minute) to a Q sepharose anion exchange column (Amersham biosciences, 5 mL). The sample was eluted from the column using a linear gradient from 0.0 M to 1.0 M NaCl. The protein would typically elute at 60-100 mM NaCl. The fractions containing the bulk of acid phosphatase activity were then pooled and dialyzed against 20 mM PO_4^{3-} , 1.0 M NaCl, pH 7.0 overnight. The dialyzed sample was then applied (5 mL/minute) to a HiTrap Chelating HP chromatography column pre-equilibrated with 20 mM PO_4^{3-} , 1.0 M NaCl, pH 7.0 and charged with 100 mM NiCl_2 . The applied sample was washed with several column volumes of the buffer. The purified protein was eluted from the column using a linear gradient from 0.0 M to 1.0 M imidazole plus 20 mM PO_4^{3-} , 1.0 M NaCl, pH 7.0. The protein normally eluted in the range 350-400 mM of imidazole. The fractions containing the phosphatase activity were pooled and dialyzed against 50 mM sodium acetate pH 6.0, 100 mM NaCl overnight. The dialyzed protein was concentrated to 10 mg/mL for use in crystallization studies.

A1.1.2 Crystallization

Sitting drop crystallization trays and the Index solutions from Hampton Research were used to identify initial crystallization conditions. A 1.0 mL reservoir was used with a 3.0 μ L crystallization drop containing 1.5 μ L of purified protein and 1.5 μ L of the reservoir solution. After 2 weeks crystals appeared in conditions 42-44, 70, 71, and 78-80. These crystallization conditions were optimized and the resulting crystals were surveyed for diffraction using a Rigaku Raxis-IX detector coupled to a Cu rotating anode with Osmic confocal optics. The best crystals grew in 0.2 M ammonium acetate, 0.1 M Bis-Tris pH 6.5, 27% PEG 3350. These crystals exhibited diffraction to 2.5Å and were frozen for collection at a synchrotron radiation source.

Diffraction data were collected at the Molecular Biology Consortium Beamline 4.2.2 at the Advanced Light Source of Lawrence Berkeley National Laboratory. The crystals diffracted to 2.0Å, indexed in space group C222₁ and had unit cell dimensions of $a=59.07$, $b=124.03$ and $c=61.81$. The resulting data collection statistics can be found in Table A1.1.

Table A1.1

Data collection and refinement statistics ^a	
	Native
Wavelength (Å)	1.28199
Space group	C222 ₁
Unit cell dimensions (Å)	a = 59.07, b = 124.03, c = 104.12
Diffraction resolution (Å)	43.78-2.00 (2.07-2.00)
No. of observations	107363
No. of unique reflections	15722
Redundancy	6.8 (6.8)
Completeness (%)	99.8 (100)
Average I/σ	10.5 (3.1)
R _{sym} (I)	0.099 (0.410)
No. of non-hydrogen atoms	3706
No. of residues in chain A	209
No. of water molecules	113
R _{cryst}	0.214 (0.283)
R _{free} ^b	0.254 (0.348)
RMSD ^c	
Bond lengths (Å)	0.015
Bond angles (deg.)	1.505
Ramachandran plot ^d	
Favored (%)	92.0
Allowed (%)	6.9
Average B-factors (Å ²)	
Protein	39.12
PEG Fragment	67.08
Active site metal ion	26.01
Water	38.53
PDB accession code	NA

^aValues for the outer resolution shell of data are given in parenthesis.

^b5 % random test set.

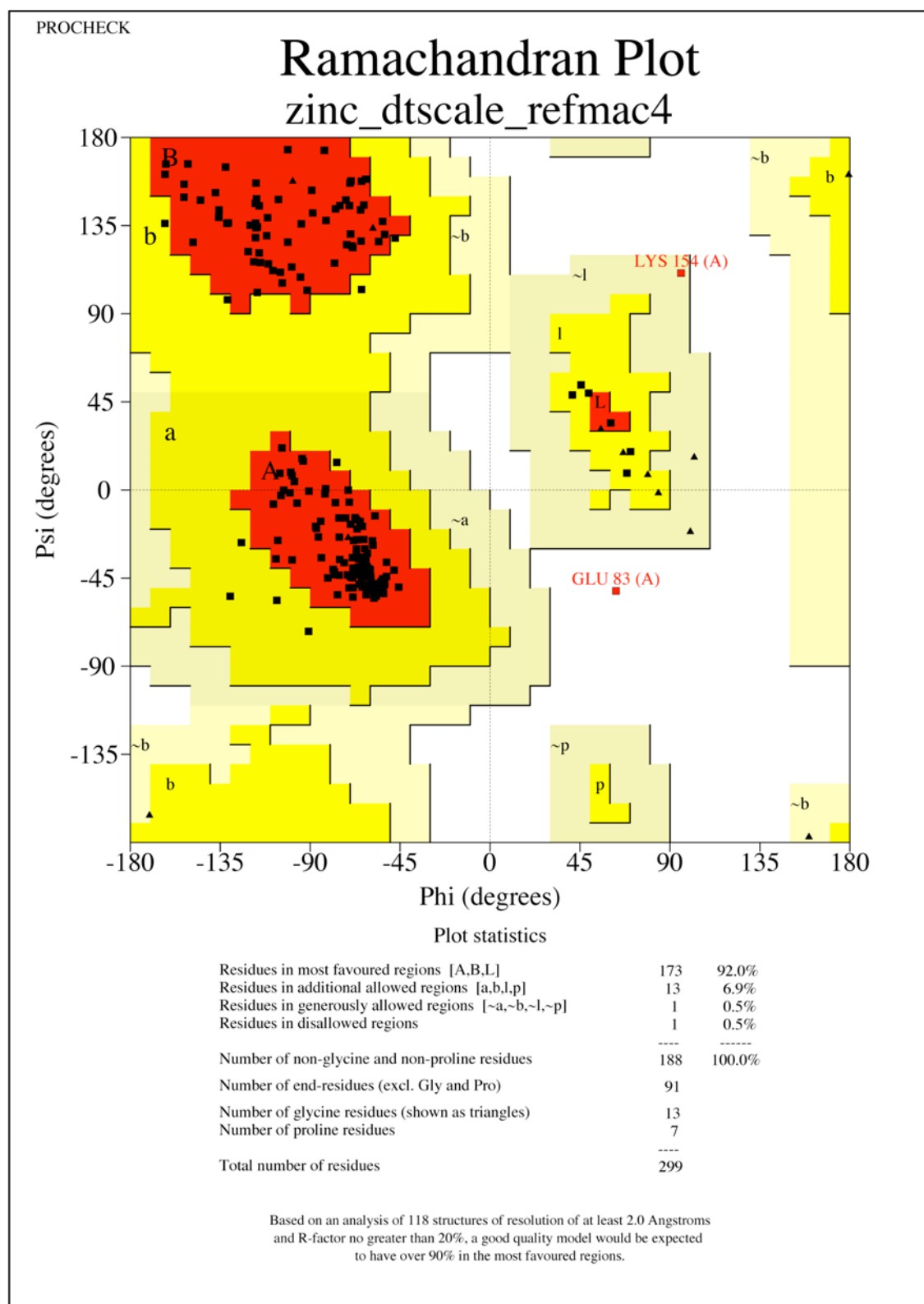
^cCompared to the Engh and Huber force field ¹.

^dThe Ramachandran plot was generated with Procheck².

A1.1.3 Structure determination

The structure of *F. tularensis* class C NSAP was solved using molecular replacement with the program MOLREP³. The starting model was the entire class C NSAP from *Bacillus anthracis* with all solvent and ligands removed. These two enzymes only share 24% sequence identity but molecular replacement was tried anyway. The results from MOLREP indicated that an initial solution had been found. The correlation coefficient was 0.263 and the R-factor was 0.588. The side chains were then truncated to alanine residues for refinement using REFMAC⁴. After refinement the missing residues as well as solvent molecules were built using COOT⁵ with refinement after each COOT session. The final structure had $R_{cryst}=0.214$ and $R_{free}=0.254$ with 209 of the 227 possible residues built. The structure also includes one magnesium ion, one PEG fragment, and 113 water molecules. A Ramachandran plot was calculated for this structure to assess backbone stereochemistry (Figure A1.1).

Figure A1.1



A1.2 References

1. Engh, R. A.; Huber, R., Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallogr.* **1991**, A47, (4), 392-400.
2. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, 26, 283-291.
3. Vagin, A.; Teplyakov, A., MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **1997**, 30, 1022-1025.
4. 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.
5. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **2004**, D60, (Pt 12 Pt 1), 2126-32.

LIST OF PUBLICATIONS

1. 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.
2. 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.
3. 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. F.* **2006**, F62, 32-35.
4. Ou, Z.; Felts, R. L.; Reilly, T. J.; Nix, J. C.; Tanner, J. J., Crystallization of recombinant *Haemophilus influenzae* e (P4) acid phosphatase. *Acta Crystallogr.* **2006**, F62, (Pt 5), 464-6.
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. F.*, **2006**, F62, 705-708.
6. 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.

VITA

On March 28, 1979 two significant events took place that would change the world forever. The first was the partial meltdown of the nuclear reactor at Three Mile Island, the worst nuclear disaster in the history of the United States; the second was the birth of Richard Levi Felts in the small town of St. Joseph, Missouri to H. Rick and Josie A. Felts. Levi was educated in the St. Joseph school district and graduated from Lafayette High School in 1997. In high school he excelled in science, JROTC, and industrial technologies choosing science as his career path. Levi decided to stay in St. Joseph and attend Missouri Western State College now University. At MWSC Levi quickly declared chemistry as his major and graduated in 2003 with a B.S. in chemistry with minors in math and physics. While attending MWSC he was a member of Phi Delta Theta fraternity where he held various positions including vice-president, he was also elected to the student government as a senator for eight consecutive semesters. Levi's passion however was chemistry, in his final two summers as an undergraduate he completed research at MWSC where he developed a physical chemistry lab for students in biological sciences, and at the University of Missouri, where he was awarded a Stevens research fellowship, there he synthesized ligands for Cu(I) coordination chemistry. After graduation Levi accepted an invitation to attend the University

of Missouri to earn his Ph.D. in chemistry. There he taught general chemistry lab in his first year, and quickly decided that he preferred laboratory research to teaching. He joined Dr. Jack Tanner's research group in his second semester at MU. Within one year he had published his first manuscript on the crystallization of a unique enzyme, AcpA, from *Francisella tularensis*. Over the next three years Levi, with the help of many, would go on to publish six more articles before graduating. Levi was privileged to attend numerous trips to conferences and synchrotrons throughout his graduate career. In November 2006 at the 5th International Conference on Tularemia he was awarded the prize for Best Poster Presentation on his work with AcpA, in February 2007 he won the University of Missouri Department of Chemistry Breckenridge/Lyons Award for Outstanding Graduate Research.

In March 2007, Levi married Kimberly Joy Elling who had graduated from The Ohio State University in 2000 and was an employee at John Deere Company. After graduating from MU in May 2007, Levi accepted a postdoctoral position at the National Cancer Institute at the National Institutes of Health in Bethesda, MD with Dr. Sriram Subramaniam.