2011

The catalytic mechanism of archaeal inorganic pyrophosphatase revealed by x-ray and neutron crystallography

Ronny C. Hughes

THE CATALYTIC MECHANISM OF ARCHAEAL INORGANIC PYROPHOSPHATASE REVEALED BY X-RAY AND NEUTRON CRYSTALLOGRAPHY

by

RONNY C. HUGHES

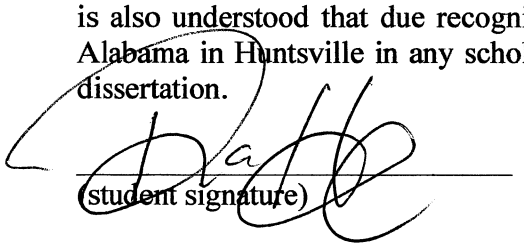
A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
The Biotechnology Science & Engineering Program
to
The School of Graduate Studies
of
The University of Alabama in Huntsville

HUNTSVILLE, ALABAMA

2011

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or, in his/her absence, by the Director of the Program or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this dissertation.



(student signature) 11/1/11 (date)

DISSERTATION APPROVAL FORM

Submitted by Ronny C. Hughes in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Biotechnology Science & Engineering Program and accepted on behalf of the Faculty of the School of Graduate Studies by the dissertation committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Biotechnology Science & Engineering Program.

Dr. Joseph D. Ng (Date) Committee Chair

Dr. Edward J. Meehan

Dr. Robert McFeeters

Dr. Pamela Twigg 10/31/11

Dr. Leighton Coates

Dr. Joseph D. Ng Program Director

Dr. John J. Fix 11/1/11 College Dean

Dr. Rhonda K. Gaede 12/5/11 Graduate Dean

ABSTRACT
The School of Graduate Studies
The University of Alabama in Huntsville

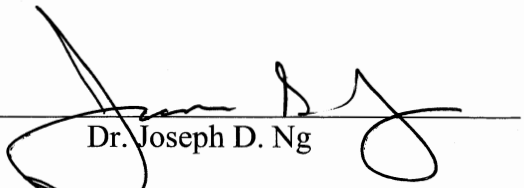
Degree Doctor of Philosophy Program Biotechnology Science & Engineering

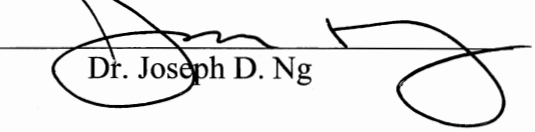
Name of Candidate Ronny C. Hughes

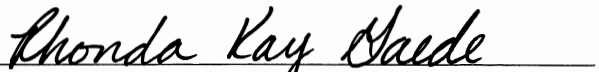
Title The Catalytic Mechanism of Archaeal Inorganic Pyrophosphatase revealed by X-ray and Neutron Crystallography

A structure-based model for the catalytic cycle of archaeal inorganic pyrophosphatase (IPPase) from the sulfur-reducing hyperthermophilic archaeon *Thermococcus thioreducens* has been proposed providing new insight into a mechanism for proton transfer-mediated, metal-assisted enzyme catalysis. Eight high-resolution X-ray crystallographic structures of the enzyme have been determined in complex with the substrate, product, and in the presence of different metal cofactors and reactive analogues. As a result, structural snapshots of the catalytic site revealed important catalytic events that could not have been obtained otherwise. Moreover, a method for growing large volume protein crystals (>1 mm^3) suitable for neutron diffraction studies has been developed through an advancement of the capillary counter-diffusion crystallization technique in restricted geometry. Large volume crystals of IPPase were obtained and used for neutron crystallographic structure determination of the enzyme solely bound to Ca^{2+} at 2.50Å resolution. The IPPase neutron crystallographic structure provided further structural information at the level of water orientation and the definitive location of hydrogen atoms in the catalytic site. The archaeal enzyme was also observed to exhibit active site rearrangements and activities closely resembling those described in

other investigations conducted on bacterial and eukaryotic IPPases. In this investigation, it was hypothesized that Asp68 in *T. thioreducens* IPPase (equivalent to Asp67 and Asp117 in *E. coli* and *S. cerevisiae* respectively) was the most likely principal proton acceptor for the deprotonation of a water molecule that bridges two metals, resulting in the formation of an attacking hydroxide nucleophile toward the hydrolysis of the pyrophosphate (PP_i) substrate. This hypothesis was tested by measuring the enzymatic activity of an Asp68Asn mutant construct of *T. thioreducens* IPPase and comparing it with the activity of the unaltered recombinant enzyme. Consequently, it was determined that the mutant enzyme retained significant activity (~10-15%) towards PP_i hydrolysis. This finding led to the proposal of an alternative pathway for nucleophile activation in *T. thioreducens* IPPase that is independent of Asp68. In the alternate pathway, nucleophile activation is accomplished by transferring the proton along a chain of water molecules to the bulk solvent. Thus, the X-ray and neutron crystallographic models coupled to supporting biochemical data allowed, for the first time, a structure-based mechanism for IPPase catalysis from an archaeal system to be proposed.

Abstract Approval: Committee Chair 
Dr. Joseph D. Ng

Department Chair 
Dr. Joseph D. Ng

Graduate Dean 
Dr. Rhonda K. Gaede

ACKNOWLEDGMENTS

I would like to foremost express my appreciation to my committee members, Drs. Pamela Twigg, Robert McFeeters, Edward Meehan, Leighton Coates, and Dr. Joseph Ng for challenging me to work to the best of my ability. My gratitude is also extended to my fellow lab members Dr. Miranda Byrne-Steele, Dr. Damien Marsic, Dr. Randall Wilson, and Amicia Elliot for their encouragement and friendship. I would like to offer a special thanks to my primary advisor Dr. Joseph Ng for challenging me to do my best and providing me with opportunities for academic advancement. I have considered it a privileged to work in a laboratory where personal friendship is held at high value, while maintaining a challenging and demanding work environment that encourages learning and professionalism. I also offer my sincere thanks to my friends Sheila Fore-Williams and Sebrina England who were staff members during the time I spent at UA Huntsville. I sincerely appreciate Dr. Ernest Curto for his technical assistance and his friendship. I would like to express my gratitude towards the late Dr. Maria Davis and Dr. Gopi Podilla for the standard of science they established at UA Huntsville. As members of my Ph.D. committee, they encouraged me to take advantage of available opportunities and not take them for granted. It is because of the example that they provided that I strive to work hard to contribute towards the advancement of science.

I appreciate the financial support provided by the National Science Foundation EPsCOR program, and the UA Huntsville, Department of Biological Sciences.

Most of all, I would like to thank my parents and my sister for their love, encouragement and support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

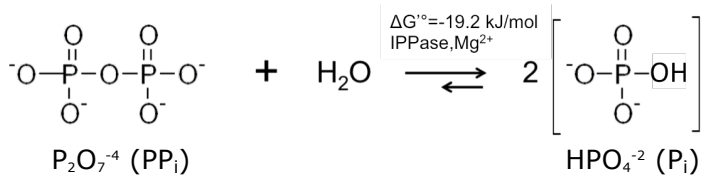
Table **Page**

CHAPTER 1

INTRODUCTION

1.1 Inorganic pyrophosphatase

Soluble inorganic pyrophosphatase (IPPase, EC 3.6.1.1) catalyzes the hydrolysis of inorganic pyrophosphate (PP_i) to form orthophosphate (P_i).



The action of this enzyme shifts the overall equilibrium in favor of synthesis during a number of ATP-dependent cellular processes, such as in the polymerization of nucleic acids and in the production of coenzymes, proteins, and sulfate assimilation pathways. IPPases have been studied from a range of sources at the biochemical and structural level. Structural features of the IPPase catalytic site in *E. coli*[1; 2] and *S. cerevisiae*[3; 4] have been revealed by X-ray crystallography, and based on these studies, along with biochemical investigations, models for the enzyme's catalytic mechanism have been proposed. However, no catalytic mechanism derived from an archeael IPPase has been put forth to-date.

Cytosolic IPPases have been divided into two subfamilies, family I and II, based on their binding preferences to Mg^{2+} and Mn^{2+} cofactors, respectively, their sequence conservation, and their activating efficiency. Family I IPPases are classified as single-

domain OB (Oligonucleotide Binding)-fold proteins, whereas family II enzymes have two domains. Bacterial family I IPPases exist predominantly as homohexamers with a subunit molecular mass of approximately 21 kDa.[2; 5-7] Figure 1.1 provides a comparison of IPPase active centers of family I and family II enzymes.

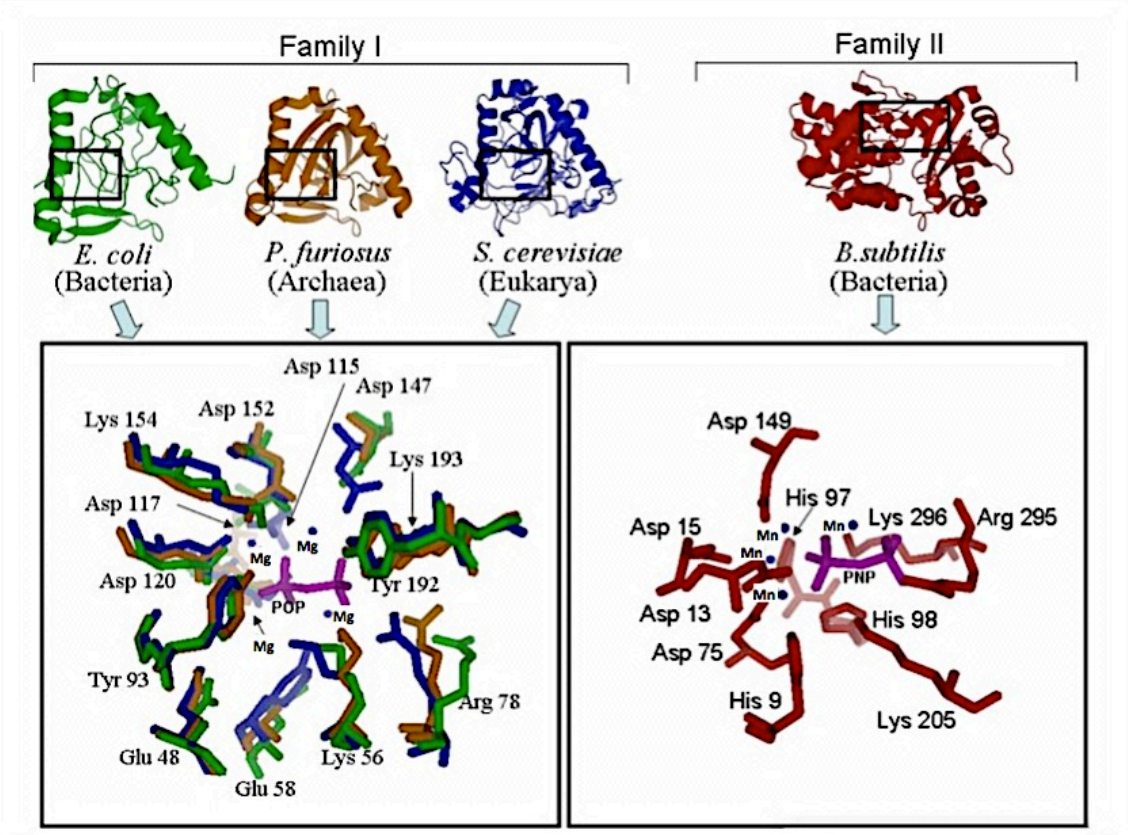


Figure 1.1. A comparison of the active site of family I and family II IPPase. In the case of family I IPPase a complete active center is contained entirely within the single domain of the monomeric enzyme subunit. In contrast, family II enzymes have two domains and the active center is located at the interface of the domains. In family I enzymes, magnesium is the preferred metal activator and metal binding is dependent on the amino acid side chains of four aspartate residues and a glutamate residue. In the case of family II IPPase, manganese has been shown to be the best activator and metal coordination is aspartate and histidine dependent. The illustration was provided by Joseph D. Ng at The University of Alabama in Huntsville.

Eukaryotic family I IPPases exist predominantly as homodimers consisting of two 35 kDa monomers. IPPases have a highly conserved active site structure comprised of at least 13 functionally important residues and three to four binding sites for divalent metal ions.[4; 8; 9] In yeast (*S. cerevisiae*), the catalytic action of IPPase is thought

to be analogous to the alkaline phosphatases, exonucleases, and polymerases involving a "two-metal ion" mechanism with the addition of a third metal ion that coordinates water molecules. Structure-based models for the catalytic cycle of *S. cerevisiae* and *E. coli* IPPases have been described.[2; 3] The ribbon diagrams of the catalytic domain of IPPase from the three domains of life are compared in Figure 1.2.

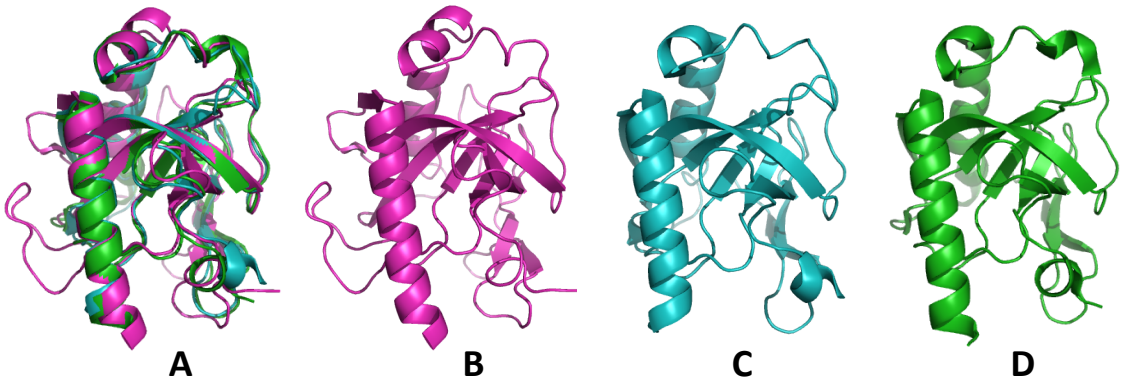


Figure 1.2. The conserved fold of family I IPPases. Ribbon diagrams of the IPPase monomeric unit from the three domains of life are shown. A. *S. cerevisiae* (magenta), *E. coli* (light blue), and *P. furiosus* (green) enzyme structures are overlaid with each other showing their topological similarity. To the right of the overlaid structures are the same structural ribbon diagrams shown separately to distinctly highlight the conserved beta-barrel and helical architecture of the enzymes in *S. cerevisiae* (B), *E. coli* (C), and *P. furiosus* (D).

A comparative illustration of the primary amino acid sequence and oligomeric structure of *S. cerevisiae* and *E. coli* enzyme IPPases is provided in Figure 1.3 and their secondary monomeric structures and topologies are illustrated in Figure 1.4.

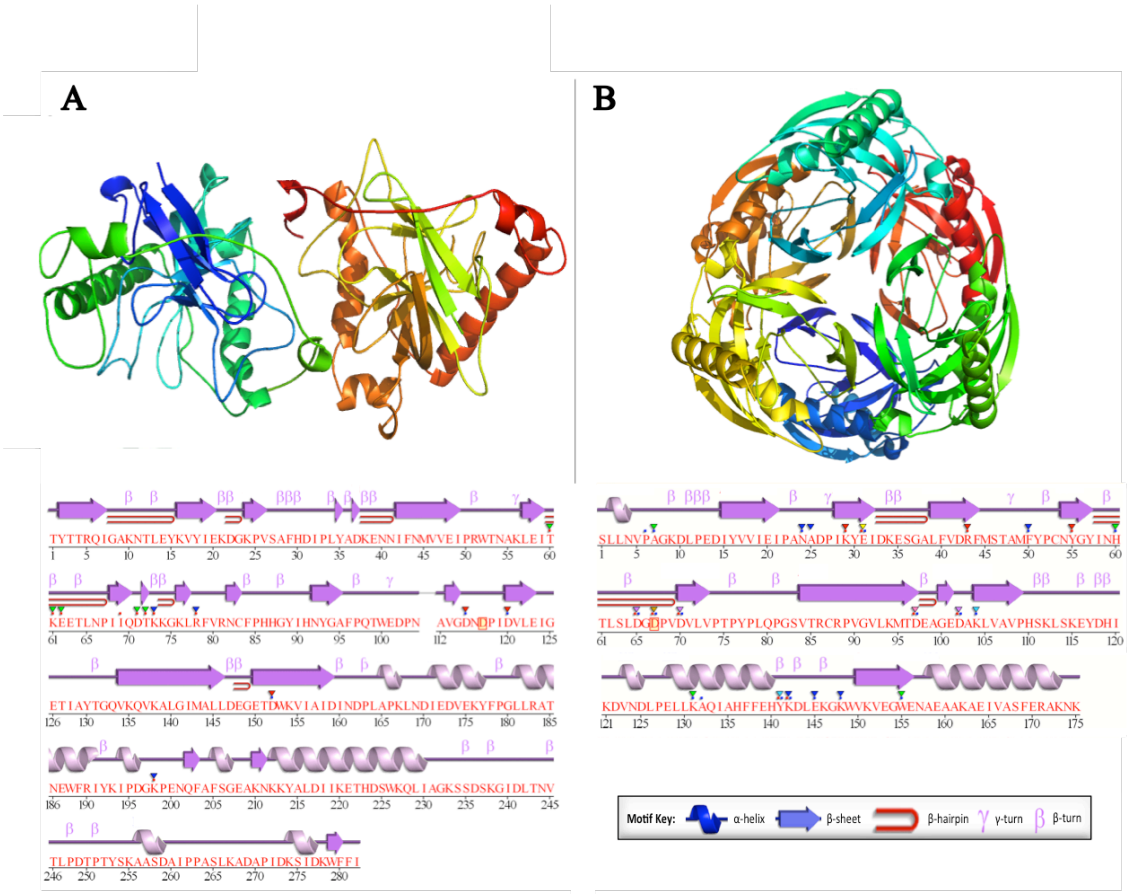


Figure 1.3. The oligomeric structure, secondary structure elements, and the sequence of amino acids found in family I IPPases. A. The three-dimensional structure of the *S. cerevisiae* dimer is shown above its primary sequence of amino acids. (PDB ID: 1E6A) B. The three-dimensional structure of the *E. coli* hexamer is shown above its sequence of amino acids (PDB ID: 2AU9). The primary sequences are labeled to highlight the conserved secondary structure elements and functionally important residues according to the provided key. The illustrations in this figure were created using Pymol and PDBsum generate tool resource developed by the European Molecular Institute division of the European Molecular Biology Laboratory (EMBL).

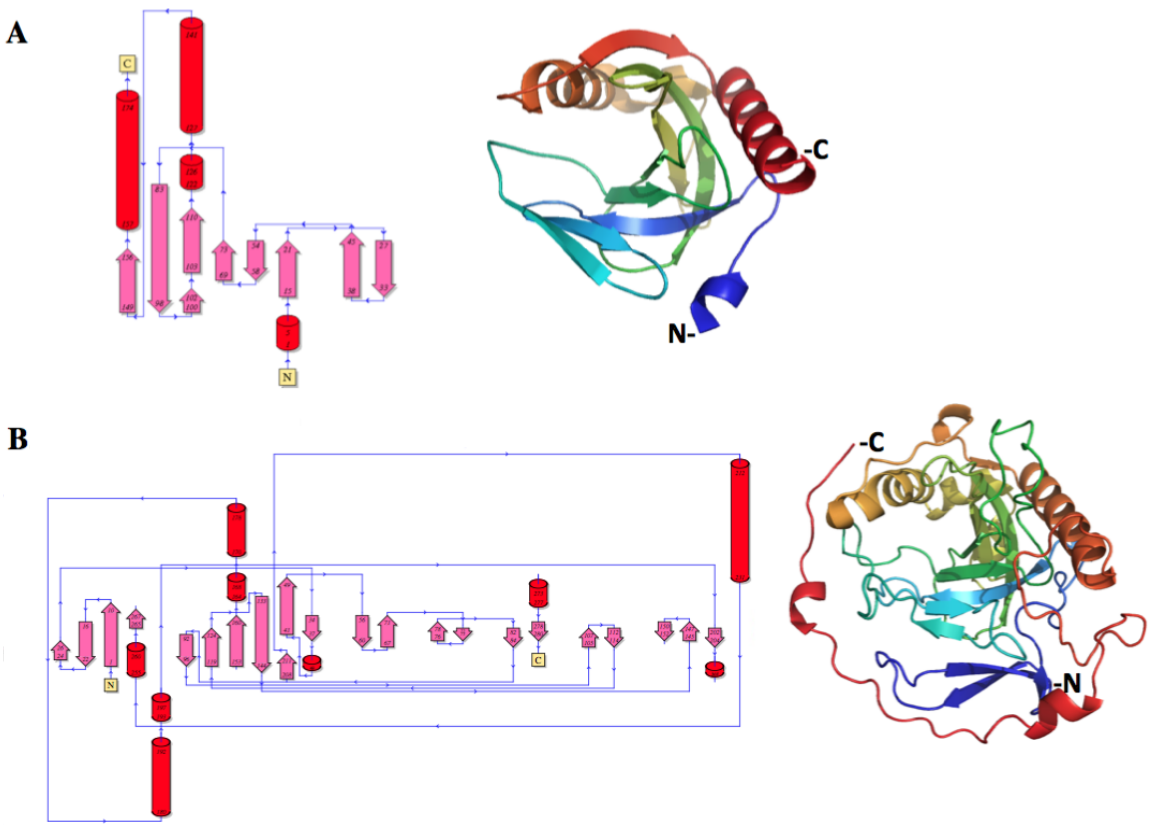


Figure 1.4. The 2-D molecular topology and 3-D monomeric structures of eukaryotic and prokaryotic family I IPPase enzymes. The diagrams provide a view of the secondary structure topologies and quaternary structure of the IPPase monomer. A comparison of the general topology of the enzymes native fold illustrates the high level of conservation present in the active site cleft of the enzyme across domains in eurkaryotic and bacterial derived enzymes. A. The 2-D topology of the *S. cerevisiae* monomer is shown next to a ribbon diagram illustration of its quaternary structure (PDB ID: 1E6A, Chain-A). B. The 2-D secondary structure of the *E. coli* monomer is shown next to its ribbon diagram. (PDB ID: 2AU9, Chain A). The ribbon diagrams in this figure were made using Pymol.

In the studied model systems, up to four essential Mg^{2+} ions bind and precisely orient the PP_i substrate, and are coordinated to active site water molecules to fill their octahedral coordination geometry. Four aspartate residues and one glutamate residue form the divalent metal binding sites. Three lysine residues, two tyrosine residues, and an arginine residue are strictly conserved and function to coordinate the substrate. One histidine is strictly conserved among prokaryotic IPPases, and it forms an important intermolecular contact at the trimer-trimer interface in these enzymes. Also among the conserved residues found in prokaryotic IPPases are hydrophobic amino acid residues

located at the dimer interface and hydrophilic residues that form the ionic pairs to stabilize the monomeric contacts of the trimer. Several glycine and two proline residues are also highly conserved. Charged amino acid residues are loosely conserved near the carboxyl termini of thermally stable IPPases and help to stabilize the structure near the terminus through ionic interactions. Among the prokaryotic sources, the *E. coli* enzyme has been the most studied. The enzyme has been shown to exist predominantly as a homohexamer under physiological conditions, consisting of two trimers related by a two-fold axis. Each of the two trimers has three 21 kDa monomers and a separate active center that binds the substrate and metal ions. While it is not fully understood, the homohexamer form of the enzyme has been shown to act in a coordinated manner to cleave the substrate.[10] In *E. coli,* both the hexameric and trimeric forms of the enzyme have been shown to be catalytically active.[10] In these studies the trimeric form of the enzyme could be obtained in mildly acid solutions or at neutral pH by site-directed mutagenesis of critical intermolecular contacts.[11; 12] In contrast, eukaryotic IPPases function as 60 kDa homodimers.[13]

Despite differences in their oligomeric structure, both prokaryotic and eukaryotic enzymes depend on the binding of three or four divalent metal ions, with two metal ions binding as cofactors to order the active site in preparation for substrate binding. The third metal binds simultaneously with the substrate, and the fourth metal appears in the presence of phosphate reaction products.[13; 14] In studies on *E. coli* and *S. cerevisiae,* Mg^{2+} was determined to be the most efficient activator, although other divalent metal cations, such as Co^{2+} and Mn^{2+}, can also be activating.[15] However, this is not true for all IPPases. Recently, an IPPase was identified from the hyperthermophilic archeaon

Thermococcus onnurineus for which Ni^{2+} was shown to be the most efficient activator of PP_i hydrolysis.[16] In *E coli* and *S. cerevisiae,* Ca^{2+} has been shown to be an inhibitor of IPPase. The kinetic data on *S. cerevisiae* IPPase determined that inhibition resulted from the competitive binding of Ca^{2+} at the M2 site.[17; 18] In general, several hypotheses have been advanced to explain differential metal activation. Among the factors that could explain the differential metal activation in IPPase include changes in the hydration state of the metal upon metal-substrate or metal-protein coordination, changes in metal-water coordination geometry, effects related to the atomic radius of the metal, and differences in metal hydration energies. Calcium inhibition of IPPase is believed to serve an important biological function by indirectly controlling the rate of many growth dependent reactions coupled to the breakdown of PP_i. This proposed regulatory role is based on biochemical investigations which revealed that inhibition of IPPase only occurs near the highest Ca^{2+} levels observed under physiological conditions, but not at normal levels. In addition to differential activation of IPPase towards PP_i hydrolysis, the identity of the metal has also been shown to control the substrate specificity of IPPase. In a biochemical study on *S. cerevisiae* IPPase it was revealed that Mn^{2+} and Zn^{2+} were activators of both ATP and PP_i hydrolyisis. However, in the presence of Mg^{2+}, the enzyme was only activated towards the hydrolysis of linear substrates (i.e., PP_i, PPP_i).[19] Metal cofactor control of substrate specificity and catalytic activity in IPPase may also have significant evolutionary significance. Although not observed for family I enzymes, family II IPPases have been characterized to preferentially bind Mn^{2+} over Mg^{2+} in only a subset of metal binding sites, and to exhibit the highest activity in the presence of both metals.

Structure-guided catalytic investigations of the hydrolysis of PP_i have been well studied using the model systems of *S. cerevisiae* and *E. coli*, and these investigations couple the usage of high-resolution crystallographic structures with biochemical analyses.[1; 3] Based on these investigations, it was proposed that the reaction nucleophile is generated by a "two-metal ion" mechanism. In this mechanism, the hydrogen bonds and extensive Lewis acid coordination provide charge shielding of the electrophile, lowering the leaving group's pK_a. A complete structural description of the catalytic cycle for the yeast enzyme has been described by determining the structure of the wild-type enzyme and seven active site variants in the presence of Mg^{2+} and phosphate.[3] In this study, the investigators were able to create a "closed" wild-type and an "open" conformation variant active site, revealing the dissociation of phosphate products, which led to the proposal of a “two water-two metal” mechanism for nucleophile generation. In the resulting structures from the yeast IPPase studies, the active site conformations representing the various catalytic cycle intermediates can be observed. This enabled a complete reaction scheme to be proposed for the enzyme as described by Oksanen *et al.*[3]

The reaction mechanism of the *E. coli* enzyme has been proposed to proceed by a more simplified scheme,[2] and more recent details have been put forth by a number of kinetic,[13; 20-22] structural[2; 23] and computational studies.[24] Four essential Mg^{2+} ions are required to coordinate the PP_i substrate. The metal coordination to the substrate and active site water displays octahedral geometry among the four Mg^{2+} ions, as described for the yeast enzyme. In addition to metal coordination, the substrate is bound to Lys29, Lys142, and Arg43 through electrostatic interactions, with Asp67 being the proposed general base to deprotonate the bridging water molecule. In this manner, the water

displacements and general reaction schemes are similar to that of the yeast enzyme. The major distinction is in the process proposed for nucleophile generation, in that no shortening of a "two metal-water bridge" was revealed from studies on the *E. coli* enzyme.[1]

Despite the extensive experiments performed on *S. cerevisiae* and *E. coli*, there are still questions regarding several key aspects of IPPase catalysis. These questions are about the chemical properties of the nucleophile and the precise mode of its activation for substrate hydrolysis, the biological role of oligomerization among hexameric IPPases, the role of active site water in terms of hydrogen relay, and the essential displacements and interactions of protein side chains with solvent molecules, substrates, and metal ions, which lead to efficient catalysis. Furthermore, while thermally stable IPPases from the hyperthermophilic archaea *P. furiosus* and *P. horikoshii* have been studied both biochemically and structurally,[25] the archaeal IPPases have been studied to a far less extent and currently no model been proposed for archaeal IPPase catalysis.

1.2 The role of hydrogen in enzyme catalysis

Hydrogen bonds play a major role in all forms of enzymatic catalysis. However, the full capacity that hydrogen bonds have towards catalysis remains to be determined.[26] Following the first description of the hydrogen charge relay system in the X-ray crystallographic structure of the enzyme alpha-chymotrypsin,[27] the catalytic importance of hydrogen atoms in enzyme catalysis has become increasingly more apparent. This has led to a number of hypothesized mechanisms for hydrogen-mediated catalysis. Among the earlier proposed mechanisms included a description of oxyanion holes, and low barrier hydrogen bonds (LBHB) that have been described to exist in the active sites of

enzymes. Subtilisin, chymotrypsin,[27] and more recently elastase,[28] all members of the serine protease family of enzymes, have provided suitable model systems from which to test hypothesizes regarding the roles of hydrogen atoms in enzyme catalysis. In this family of enzymes, an aspartate-histidine-serine (or in some cases glutamate-histidine-serine) catalytic triad forms a charge-relay network to shuttle a proton from the catalytic serine to histidine, thereby activating the serine towards nucleophilic attack on the carbonyl carbon of the protein/poly-peptide substrate. The hypothesis presumes that as the hydrogen bond length shortens between the carboxylate oxygen of the aspartate side chain and histidine nitrogen to approach the sum of the van der Waals distances of the bonded atoms (~2.5Å), the difference between the proton acceptor and donor pK_a's values are minimized. Therefore, in theory the LBHB bond would be much stronger (~5-10x) than a traditional polar hydrogen bond.[29] In addition to the LBHB, an oxyanion hole was hypothesized to be formed by the backbone amide hydrogen atoms of two active site glycine residues. In the proposed mechanism, the filling of the oxyanion hole was necessary to stabilize the peripheral charge on the substrate's carbonyl oxygen in order to allow the formation of a covalent serine-peptide acyl intermediate. Subsequently, the oxyanion hole is emptied and the peptide bond is cleaved to produce the reaction product.[30] Only recently neutron diffraction studies of the structure of elastase[28] and photoactive yellow protein[31] have provided the means to test the validity of LBHB and oxyanion hole proposals.

Model enzyme systems for studying water-facilitated hydrogen relay chains have also been pursued. The enzyme carbonic anhydrase C has proven to be a good model system to study water-mediated proton relay. Carbonic anhydrases operate via a two-step

mechanism to catalyze the reaction between a zinc-bound hydroxide ion and carbon dioxide to form bicarbonate. In carbonic anhydrase, early investigations of intermolecular proton transfer were based on hydrogen-deuterium studies.[32] Many years passed before mutational studies were able to provide insight and test hypotheses regarding the mechanism leading to the enzymes high catalytic efficiency.[33] To date, extensive experiments including neutron diffraction and NMR structural studies have been conducted to better understand the structural and biochemical basis of proton transfer in this enzyme. Specifically, the model systems established for the enzyme have proven to be good models for studying water proton transfer.[34] This is because a critical aspect of the enzyme's mechanism of catalysis has been proposed to depend on a three water relay system in which the pK_a of the water molecules in the relay chain are lowered by their coordination to metal. In the model, the water relay system acts as proton transfer bridge leading to an active site histidine residue. In 2010 a neutron crystallographic structure of human carbonic anhydrase II provided the means to directly visualize the proton positions within the enzyme's active center.[35] Studies on proton transfer in carbonic anhydrase have recently been reviewed.[36] The insight provided has led to an increasing number of testable hypotheses for the role of hydrogen bonded water charge relay systems in enzyme catalysis.

In recent years, effort has been extended to better understand long-range hydrogen relay systems and their role in enzyme catalysis and protein signaling. Among the most notable studies are those conducted for enzymes dependent on small molecule cofactors such as the multi-enzyme complex of pyruvate dehydrogenase.[37] Pyruvate dehydrogenase is a highly symmetric multi-enzyme complex that has a molecular mass of

up to 10 MDa, depending on the source organism.[38] In these studies a "proton wire" was observed and proposed to allow for long-range communication through a proton shuttle mechanism.[39] As a result it has been proposed that proton relay networks and clusters within enzymes could allow co-factors to function as effective general acids and bases in catalysis.[40] As a result of structural and biochemical studies targeting the thiamine diphosphate-dependent E1 enzyme component of the pyruvate dehydrogenase complex, a proton relay chain was suggested to enable active sites coordination in the oligomeric enzyme complex as an explanation for the biological significance of oligomerization in the enzyme.[39; 40]

Despite these extensive efforts there remains an opportunity for the development of new model systems for studying the catalytic role of hydrogen. Experimental methods for validation are also of great necessity, considering that the mechanism of proton transfer catalysis of even the most well studied systems are at best poorly understood. Recently, model systems have been established to provide much needed structural information regarding the protonation state of critical residues in several enzymes with catalytic mechanisms that rely on proton transfer. These model systems include alcohol dehydrogenase,[41] diisopropyl fluorophosphatase,[42] bovine trypsin,[43] and β-lactamase.[44] Nevertheless, additional model systems are needed in order to better understand the roles that hydrogen relays and hydrogen bonding serve in enzyme catalysis.

1.3 Purpose of this study

A proposed model for the catalytic cycle has been established for the enzyme from *S. cerevisiae* and *E. coli*. Nevertheless, significant work is still needed to validate the proposed catalytic mechanisms of these enzymes. Many aspects related to the process of nucleophile generation and the precise atomic details of substrate hydrolysis have yet to be described. It still remains unclear whether active-site water molecules can act in a potential hydrogen relay system to generate the nucleophile. If so, does the water relay system depend on aspartate for efficient proton transfer? The mechanism of proton transfer, including the preferred path of the proton from the site of nucleophile generation, needs to be further investigated. While enzymes from *P. furiosus* and *P. horikoshii* have been studied, no catalytic model system has been established for an archaeal IPPase. This study focuses on deciphering the precise mechanism of the family I IPPase from the hyperthermophilic archaeon *T. thioreducens* (IPPase) and is aimed at providing an atomic-level description of the catalytic mechanism of the enzyme. An equally significant goal of this work is to provide a model system for studying the mechanism of proton transfer in IPPase as a means to discriminate between conflicting models for nucleophile activation.

The most powerful method to obtain hydrogen positions is neutron crystallography. The enzyme will be a target for neutron structure determination as a means to identify the hydrogen positions in critical amino acid residues and water molecules within the active site. While neutron diffraction studies have been attempted for *S. cerevisiae* IPPase, no significant results were produced due to instability of the crystals and experimental complications caused by the neutron scattering properties of

manganese ions in the active site. *T. thioreducens* IPPase's thermal stability and high propensity to crystallize could make the enzyme a suitable candidate for neutron structural studies. This study is based on the hypothesis that archaeal IPPase from *T. thioreducens* catalyzes the hydrolysis of PP_i through a metal-assisted mechanism to activate a water molecule (i.e., hydroxide ion) for a backside nucleophilic attack on the innermost phosphate group of the substrate. An extensive set of high-resolution X-ray crystallographic structures for IPPase is also desired to provide a complete description of the active site rearrangements associated with catalytic events. Thereby, the X-ray crystallographic structures coupled to the neutron crystallographic structure of IPPase will be used to guide subsequent biochemical studies, aimed at significantly increasing our understanding of the catalytic mechanism of the archaeal enzyme. The activity of the IPPase wild-type enzyme and active site mutants will be examined and used to establish a model system in archaea. Physical parameters of IPPase activation will be determined as a function of metal cofactors, variations in pH, and temperature. The conditions for optimal activity will be described, and IPPase mutants will be assayed for their capacity towards efficient PP_i hydrolysis.

CHAPTER 2

MATERIALS AND METHODS

2.1 Recombinant cloning, expression, and purification

2.1.1 Gene-to-structure pipeline

The inorganic pyroposphatase was one of the proteins targeted in a manual gene-to-structure pipeline constructed to clone, recombinantly express, and crystallize protein targets of interest in a high-throughput process without any automation. The protein coding DNA sequences were identified in the sequenced genome of *T. thioreducens*[45] and were targeted for Polymerase Chain Reaction (PCR)[46; 47] amplification in a 96-well configuration. The amplified products were subcloned into expression vectors using *in vivo* homologous recombination. Recombinant protein expression trials were performed in *E. coli,* and clones that displayed over-expression were used for large-scale protein production. In the final stage of the pipeline, purified protein were screened for crystallization using high-throughput methods with commercially available screening reagents. The goal of this pipeline approach was to provide a means to go from gene to structure in a time- and cost-effective manner without the necessity for automation. Among the initial targets selected for structure-function studies was the inorganic pyrophosphatase, which is the focus of this study. Figure 2.1 illustrates the steps associated with this pipeline from targeting to structure determination.

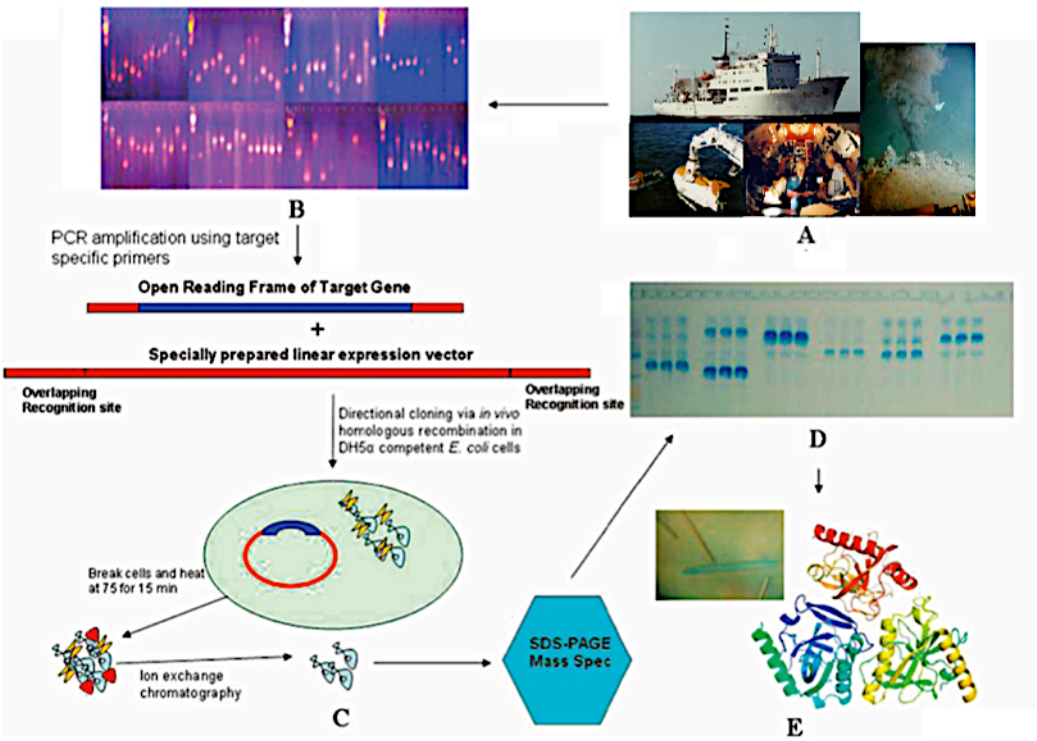


Figure 2.1. Representative steps of the gene-to-structure pipeline targeting the genome of the hyperthermophilic archaeon *T. thioreducens.* A. The genome of the cultured hypothermophilic archaeon *T. thioreducens* was sequenced. The archaeon isolated from a deep-sea thermal vent near the mid-Atlantic ridge. B. Cloning is performed by PCR amplification using selected oligonucleotides against targeted coding regions. PCR amplification is usually very specific, and successful reactions are observed to occur over 95% of the time as shown when reaction samples are analyzed by agarose gel electrophoresis. The samples shown are run on 1% agarose, stained with ethidum bromide and visualized under UV. C. The PCR products are subcloned directly into an expression vector via homologous annealing and transformed into *E. coli* cells without ligation or extensive restriction digest. The resulting plasmids are propagated and transformed into an expression host for small-scale expression trials. D. Different clones are overexpressed and the fractions are heat selected such that the remaining soluble protein fractions are analyzed on SDS-PAGE analysis. Overexpression is clearly seen as the heaviest band fragment and usually comprises of more than 50% of the protein at this step. Approximately one-third of the cloned fragments demonstrate strong expression, and they undergo large-scale expression where they are purified to more than 90% homogeneity ion exchange or size exclusion chromatography. E. The purified protein fractions are screened for crystallization. Protein crystals that diffract to atomic resolutions are use to collect a complete crystallographic data set for structure determination.

IPPase was among the initial targets that were carried all the way through the pipeline from cloning, expression, and purification, to X-ray crystallographic structure determination. Representative structures were determined in the course of this study to show the effectiveness of the gene-to-structure pipeline (Table 2.1, Figure 2.2). Since IPPase is the focus of this thesis, other protein structures will not be discussed.

Table 2.1. X-ray crystallographic structure determined initial phase of the pipeline.

Protein	MW(kDa)	Space Group	Unit Cell (Dimensions in Å, Angles in °)	Ribbon Diagram
Inorganic pyrophosphatase	20.7	C2	a=106.03,b=95.51,c=113.9 α=γ=90,β=98.12	
N-aminopeptidase	50.4	$P6_3$	a=b=127.273,c=195.92 α=β=90,γ=120	
Alcohol dehydrogenase	43.8	$P4_32_12$	a=b=69.400,c=166.59 α=β=γ=90	
Lactate dehydrogenase	38.8	$P2_1$	a=75.783,b=108.096,c=107.8 α=γ=90,β=9369	
Iron-uptake protein-FeoA	8.8	C2	a=93.78,b= 68.38,c=68.96 α=γ=90,β=132.1	
Piwi/Argonaute/Zwille	15.7	$P2_12_12_1$	a=57.39,b=60.89,c= 66.44 α=β=γ=90	
Rnase PH	24.3	$C222_1$	a=83.14,b=123.72,c=62.65 α=β=γ=90	

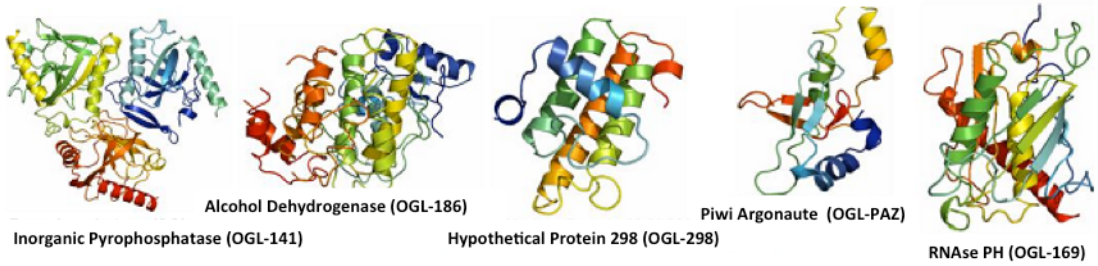


Figure 2.2. Examples of X-ray crystallographic structures produced by the gene-to-structure pipeline. The ribbon diagram representations of high-resolution X-ray crystallographic structures solved in the course of this study are shown. The ribbon diagram shown far left represents the structure of IPPase, the focus of this study.

2.1.2 PCR amplification of the IPPase DNA coding region

The coding region for IPPase was selectively amplified by the Polymerase Chain Reaction (PCR)[47] from the purified genomic DNA of *T. thioreducens* using the PCR primer pairs: 5'-tttgtttaactttaagaaggagatatacatATGAACCCGTTCCACGAGCT-3' (Operon) and 5'- tcctttcgggctttgttagcagccggatccTCACTCCTCCTTGCCGAACT-3. The primers were designed from the exact alignment of the first and last 20 nucleotides of the IPPase open reading frame (uppercase letters). The oligonucleotide primers were also preceded by linker sequences corresponding to 30 nucleotides of the expression plasmid vector pET3a (Novagen), which contain the NdeI/BamHI insertion sites (lowercase letters). The amplification reaction was performed in a total volume of 50 μl containing 20 pmole of each primer, 20 ng of genomic template, 2 units of Rainbow DNA polymerase (iXpressGenes), 200 mM of each dNTP, 1 mM $MgCl_2$, and standard PCR buffer[20] in sterile PCR grade water. PCR was carried out using a Perkin Elmer GeneAmp PCR System 9600 thermal cycler for 30 seconds at 95˚C initial denaturation, followed by 30 cycles of 30 seconds at 95˚C, 30 seconds at 55˚C and 1 minute at 74˚C. A final extension step was performed for 2 minutes at 74˚C.

2.1.3 Plasmid preparation

A linear pET3a expression vector (Novagen) was prepared as described by Marsic *et al.*[48] One μg of pET3a vector was digested in sequential reactions with NdeI/BamHI restriction enzyme endonucleases and subsequently purified with a commercially supplied spin column (Genlantis) according to the manufacturer's recommendations. The linear, purified fragment was eluted from the inverted spin column in 50 μL of PCR grade water. In order to minimize the background of the

cloning procedure, Klenow DNA polymerase (New England Biolabs) was used to produce a blunt end on the purified linear plasmid vector. The reaction was performed in the supplied reaction buffer by adding 2 units of enzyme per 0.5 µg of linear plasmid in the presence of a mixture of dNTP. The reaction was allowed to stand for 30 minutes at room temperature before being loaded onto a PCR cleanup spin column, which purified the linear fragment and exchanged the reaction buffer. The eluted, linear DNA was then diluted to 20 ng/µL and stored at -20˚C.

2.1.4 Production of competent cells

E. coli strain DH5α was rendered competent using the rubidium chloride method[49]. A commercial stock of DH5α was used to grow a 10 ml inoculate in Luria-Bertani (LB) medium, and it was incubated at 37˚C in a shaker incubator. The overnight inoculate was used as a starter culture for growth in 250 ml "SOB" media. The resulting culture was placed in a refrigerated shaker incubator at 18˚C until O.D. 600 nm absorbance was approximately 0.6. The cell suspension was then placed on ice for 10 minutes, followed by centrifugation at 5000 rpm in a JA-25.50 rotor for 10 minutes at 4˚C. The cells were then gently resuspended in 80 ml of ice-cold transformation buffer and placed on ice for 10 minutes. The cells were pelleted a second time in the manner previously described. Next, the second pellet was gently resuspended in 20 ml of ice-cold transformation buffer (10 mM Pipes pH 6.7, 15 mM $CaCl_2$, 250 mM KCl, and 50 mM $RbCl_2$). The suspended cells were supplemented with DMSO to a final concentration of 7% and placed on ice for a final 10-minute period. Then, they were divided into 100 ml aliquots in 1.5 ml pre-chilled microcentrifuge tubes and placed at 80˚C.

2.1.5 Recombinant cloning

The PCR primers were designed to include additional 3' and 5' vector specific sequences to facilitate ligation-independent cloning in the specially prepared pET3a expression vector. Approximately 40 ng of the resulting amplified product was mixed with an equal amount of the linearized expression vector, and the resulting mixture was transformed into chemically competent DH5α *E. coli*, as described by Marsic *et al.*[50] This method of cloning was possible because the 5' and 3' extreme DNA ends flanking the IPPase open reading frame were designed to be homologous to the end sequences of a prepared linear expression vector. When the DNA fragment was mixed with the prepared linear expression vector and transformed into competent *E. coli* cells (e.g., DH5α), an intrinsic recombinase activity joined the two DNA fragments. This resulted in a circular expression plasmid containing the coding region of the targeted protein. Thus, cloning can be performed without the use of overnight ligation or restriction enzymes. To accomplish this, 40 ng of the prepared linear vector was mixed with 100 ng of the amplified DNA sequence in a sterile microcentrifuge tube on ice. After twenty minutes, the mixture was used for transformation into DH5α competent *E. coli* cells. This was accomplished by adding the entire mixture (~5 µL) to 50 µl competent cells thawed on ice. After 30 minutes of incubation, the cells were heat-shocked at 42˚C for 1 minute. Thereafter, the transformed cells were diluted with 400 µl Luria-Bertani (LB) medium and incubated for 1 hour at 37˚C before being spread on LB-agar plates containing 100 mg/L carbenicillin. Colonies were grown overnight in selective LB miller broth to propagate the recombinant plasmid. The plasmids were then extracted by standard molecular biology protocols[51] using a Qiagen plasmid mini-prep kit according to the

manufacturer's instructions. Finally, the purified plasmids were sequenced by MWG DNA sequencing service to identify the clones that contained the correct IPPase coding sequence.

2.1.6 Transformation of expression host

The sequence-confirmed plasmid construct was transformed into *E. coli* Rosetta2^{TM} cells (Novagen) for expression by adding approximately 20 ng of plasmid DNA to 50 µl of competent cells. After 30 minutes of incubation on ice, the cells were heat-shocked at 42˚C for 1 minute. The transformed cells were then diluted with 400 µl Luria-Bertani (LB) medium, and incubated for 1 hour at 37˚C before being spread on LB-agar plates containing 100 mg/L carbenicillin and 35 mg/L chloramphenicol.

2.1.7 Production of glycerol stocks

Glycerol stocks were stored from first generation growth cultures that had been inoculated with a single colony of the transformed expression host. The cells were taken from the growing culture during the exponential growth phase just prior to IPTG induction. Sterile glycerol was added to 1 ml aliquots of cell suspension to reach a final concentration of 12%. Gentle agitations of the sample mixtures were performed to allow the glycerol to evenly distribute within the cell suspension. The solutions were prepared in sterilized 1.5 ml microcentrifuge tubes and stored at -80˚C. These stocks were used as starter inoculates for overnight cultures for all subsequent large-scale protein expression experiments.

2.1.8 Recombinant expression

Cultures for recombinant expression were grown in 2 L of Luria-Bertani (LB) broth containing 100 μg/ml carbenicillin and 35 μg/ml chloramphenicol. Following inoculation of the 2 L culture with a 10 ml overnight starter culture, the cells were allowed to propagate at 37˚C until the point of induction corresponded to an optical density of 0.8 $O.D._{600nm}$. At this time, protein expression was induced by the addition of isopropyl thio-ß-D-galactoside (IPTG) to reach a final concentration of 1 mM, and it was then transferred to a refrigerated incubator to shake overnight at 250 rpm at 18˚C. After 18 hours, cells were harvested by centrifugation at 8,000 x g in a JLA 9.1000 rotor in a Beckman Avanti J-25 centrifuge.

2.1.9 Protein purification

Protein purification was accomplished by resuspending the pellet (~10 g wet weight) in 100 ml of buffer A (50 mM Tris pH 7.5, 50 mM NaCl and 1 mM EDTA). Cells were disrupted by 8 cycles of sonication using a Branson sonifier 250 cell disruptor (VWR Scientific, USA). The resulting lysate was centrifuged for 20 minutes at 10,000 x g in a JA-25.50 rotor, and the supernatant was subjected to heat treatment for 30 minutes in an Erlenmeyer flask submersed in a 75ºC water bath. The denatured protein was removed by centrifugation at 10,000 x g for 20 minutes in the same rotor, and the supernatant was loaded onto a 5 ml HiTrap Q sepharose XL cation exchange column (GE Heathcare, USA) that had been pre-equilibrated with buffer. The loading buffer used for anion exchange contained 0.05 M Tris pH 7.5, 0.05 M NaCl, and 1 mM EDTA and the elution buffer was 0.05 M Tris pH 7.5, 1 M NaCl and 1 mM EDTA. The protein was eluted using a flow rate of 2.5 ml/min into 2.5 ml fractions as function of a gradient from

0.05 – 1.0 M NaCl using an Äkta Explorer FPLC system (Amersham-Pharmacia). The collected fractions corresponding to the major elution peak were pooled and concentrated to a volume of 2 ml using a centricon-10 ultrafiltration device (Millipore) and loaded on to a S-200 sephadex size-exclusion column. The column was pre-equilibrated with an elution buffer that contained 50 mM Tris pH 7.5 and 50 mM NaCl buffer. No EDTA was added to the final elution buffer. The fractions of the principal elution peak from the size-exclusion column were eluted by approximately 60 ml of buffer, and were collected and analyzed by SDS-PAGE. To assess the level of nucleic acid contaminants present in the final purified sample the relative absorbencies measured at 280 nm and 260 nm were measured using a UV spectrophotometer.

2.2 Characterization of enzymatic activity

The biochemical characterization of IPPase was performed using a colorimetric molybdate assay.[52] Metal dependent IPPase activity was measured to determine the best activating metal. The degree of metal activation was determined by P_i liberation at 75˚C in a pH 7.5, 50 mM Tris buffer, containing 5 mM of the metal being evaluated. Rate experiments were also conducted at various initial substrate concentrations. Mg^{2+}, Mn^{2+}, Ca^{2+}, Co^{2+}, Cd^{2+}, Zn^{2+} were tested. A sodium phosphate standard was used to correlate absorbance at 660nm with $[P_i]$. Units for activity were calculated in µmol of PP_i/min. The experiments were repeated at 60˚C and 80˚C. The three metals with highest activation were tested at 37, 60, 70, and 80˚C at pH 6.5 and 8.5. Once the best activating metal had been determined, IPPase was assayed for activity in the presence of this metal over a broad range of pH and temperature in three alternate buffers corresponding to the

ranges determined. Each reaction was initiated by adding the enzyme to 1 ml aliquots in a 96-well block. Fifty µl were taken at 1-minute intervals and placed on ice in a 96-well 100 µl assay plate for the duration of the experiment. A solution of 1% ammonium heptamolybdate (w/v), 0.83 M sulfuric acid, 8% iron (II) chloride (w/v), and 2% SDS were used to stop the reaction and induce the color change as in the method described by van Alebeek *et al.*[52] The absorbances of these samples were measured after thirty minutes.

2.3 Crystallization screening for X-ray diffraction

2.3.1 General screening strategy

Crystallization screening trials were performed using freshly prepared protein, which was judged to be substantially pure from both protein and nucleic acid contaminants by SDS-PAGE and by the relative absorbance (~1.92:1) at 280 nm and 260 nm, respectively. The initial crystallization trials were performed with 15 mg/ml and 7.5 mg/ml solution using the sitting drop vapor diffusion method in a 96-well intelliplate (Art Robbins) with commercial sparse-matrix screens[53] from Hampton Research. Hampton Screen High-Throughput (HSHT) reagents were used as the reservoir solutions in the initial trial. Protein and precipitant ratios of 1:1 and 2:1 were placed in the top and bottom sample wells, respectively. The plate was allowed to equilibrate undisturbed for one week at 20˚C in a low temperature incubator.

2.3.2 Crystallization of the metal-free enzyme

Protein crystallization trials were performed with purified IPPase prepared at concentrations of 7 mg/ml and 15 mg/ml. Proteins were equilibrated against the HSHT

reagents (Hampton Research) using 96-well intelliplates (Art Robbins). Protein and precipitant ratios of 1:1 and 2:1 were placed in the top and bottom sample wells, respectively, with each reservoir containing 100 µl of the undiluted precipitating reagent. The plate was allowed to equilibrate undisturbed for one week at 20˚C in a low temperature incubator. After 5 days of equilibration, initial crystals were observed in conditions containing 2 M ammonium sulfate and 5% v/v 2-propanol. The crystallization conditions were further optimized in a 96-well sitting-drop vapor diffusion plate (Corning) by varying the concentration of ammonium sulfate and 2-propanol over a broad range of protein concentration and pH. Optimal crystals were finally obtained within a 5 µl droplet, which contained 15 mg/ml protein when equilibrated against 1 M ammonium sulfate and 10% 2-propanol for 1 month at 20˚C.

2.3.3 Crystallization of divalent metal complexes

This screening approach was repeated with an array of purified enzyme solutions, each containing a different divalent metal at a final concentration of 10 mM. The initial protein solution for the divalent metal crystallization trials was produced in precisely the same manner as was used for the initial screening trials, with the exception that no EDTA was used during the final purification procedure. In total, seven 0.5 ml aliquots of 15 mg/ml IPPase were prepared in 1.5 ml Eppendorf microcentrifuge tubes. Using a 100 mM stock, six of the tubes were supplemented to a final concentration of 10 mM divalent metal. The metal solutions that were used were as follows: magnesium chloride, manganese chloride, calcium chloride, cobaltous chloride, ferrous sulfate, and zinc sulfate. Only one divalent metal was added to each tube. The remaining aliquot was used “as is” to serve as a control. Crystal trials were performed with a 15 mg/ml protein

solution using the sitting-drop vapor diffusion method. Three, 96-well intelliplates that were prefilled with HSHT screening reagents were used for this trial. Two microliters of each protein-metal solution were mixed with an equal volume of reservoir solution and allowed to equilibrate against the reservoir solution undisturbed for one week, as described for the initial screen. Using the control plate as a guide to differentiate between conditions that were only occurring in the presence of a particular metal ion, additional non-redundant solutions were determined to yield single crystals suitable for X-ray data collection. Of the metals tested, only magnesium and manganese yielded new conditions with well diffracting crystals.

For IPPase supplemented with magnesium chloride, conditions 0.1 M sodium chloride, 0.1 M BICINE pH 9.0, 20% v/v polyethylene glycol monomethyl ether 550; 0.1 M HEPES pH 7.5, 70% v/v (+/-)-2-methyl-2,4-pentanediol; and 0.2 M sodium chloride, 0.1 M sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-methyl-2,4-pentanediol yielded crystals suitable to obtain a reasonable quality data set. In the case of manganese chloride supplemented protein, only condition 0.1 M HEPES pH 7.5, 10% w/v polyethylene glycol 8,000, 8% v/v ethylene glycol yielded crystals suitable to obtain a reasonable quality data set. For calcium chloride supplemented protein, HSHT condition G11 and E10 yielded small single crystals; however, the diffraction quality of the crystals was not suitable for data collection.

2.3.4 Crystallization of substrate and product complexes

The protein preparation was performed in precisely the same manner as the original screen using 10 mg/ml IPPase with 5 mM sodium pyrophosphate added to the concentrated protein solution just prior to screening. A 1:1 and a 2:1 ratio of protein to

precipitant was used in the top and bottom well of each trial. The plate was allowed to equilibrate undisturbed for one week at 20˚C in a low temperature incubator.

Crystals were used to determine the X-ray crystallographic IPPase structure PDB ID 3Q46 were obtained in the following conditions: 0.1 M magnesium chloride hexahydrate, 0.1 M sodium HEPES pH 7.5, and 15% 2-propanol. The crystals were collected after 3 months of equilibration at 20˚C. Crystals used to determine structure 3Q9M were obtained under conditions of 0.02 M calcium chloride dihydrate, 0.1 M sodium acetate trihydrate pH 4.6, and 20% 2-methyl-2,4-pentanediol. Crystals were obtained within 2-7 days of undisturbed equilibration at 20ºC. Finally, crystals used for the determination of structure 3Q4W were grown under 0.2 M calcium chloride dihydrate, 0.1 M sodium acetate trihydrate pH 4.6, and 30% v/v 2-propanol within one month of incubation at 20˚C. The vapor diffusion crystallization conditions used to obtain the X-ray crystallographic structures in this dissertation are provided in Table 2.2.

Table 2.2. The conditions for IPPase crystallization determined by vapor diffusion.

Crystallized component	Crystal Lattice, molecules per asymmetric unit	Protein solution	Precipitate solution	Ratio of solutions (Equilibration time, Temperature)
Metal-free IPPase bound to sulfate	Triclinic, six	15 mg/mL IPPase 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.001M EDTA	1.0M ammonium sulfate, 10% v/v 2-propanol	1:1 (6 weeks, 293K)
Magnesium-coordinated IPPase bound to sulfate	Trigonal rhombohedral, three	15 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.01M magnesium chloride	0.1M sodium chloride, 0.1M BICINE pH 9.0, 20% v/v PEG-MME 550	1:1 (3 months, 293K)
Manganese-coordinated IPPase	Trigonal rhombohedral, two	15 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.005M manganese chloride	0.1M HEPES pH 7.5, 10% w/v PEG 8,000, 8% v/v ethylene glycol	1:1 (4 weeks, 293K)
Magnesium-activated product complex	Trigonal rhombohedral, one	10 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.005M sodium pyrophosphate	0.1M magnesium chloride, 0.1M sodium HEPES pH 7.5, 15% v/v 2-propanol	1:1 (2 months, 293K)
Calcium-inhibited substrate complex I	Trigonal, rhombohedral, one	10 mg/mL IPPase. 0.05M TRIS pH 7.5, 0.05 M sodium chloride, 0.005M sodium pyrophosphate	0.2M calcium chloride, 0.1M sodium acetate pH 4.6, 30% v/v 2-propanol	2:1 (2 weeks, 293K)
Calcium-inhibited substrate complex II	Trigonal, hexagonal, three	10 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.005M sodium pyrophosphate	0.02M calcium chloride, 0.1M sodium acetate pH 4.6, 20% v/v MPD	(2:1 or 3:1) (7 days, 293K)

2.4 X-ray structure determination and refinement

2.4.1 Collection and processing of diffraction data

The major steps associated with X-ray crystallographic structure determination are shown in Figure 2.3. IPPase crystals that were grown in the presence of 2-methyl-2,4- pentanediol were mounted on an appropriate size nylon loop (Hampton Research) and were subjected to flash freezing and storage in liquid nitrogen prior to X-ray data collection. In cases where the mother liquor was not a sufficient cryo-protectant the crystals were briefly soaked in an equivalent solution that was supplemented with sterile glycerol.

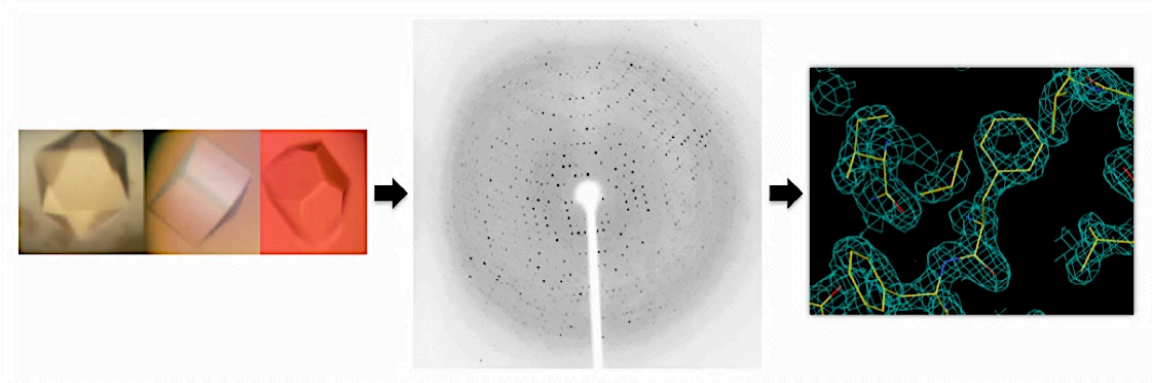


Figure 2.3. The steps associated with X-ray crystallographic structure determination. A critical step in X-ray crystallography is obtaining the crystals. In order to do this, targeted proteins are cloned into an *E. coli* expression plasmid, over-expressed, purified, and set-up for crystallization. This figure illustrates the steps after crystallization. Crystals are diffracted, and data is collected at the synchrotron at Argonne National Lab in Chicago, IL. The intensity and relative spatial orientation of the reflections together with phase information allows one to fit all of the carbon, oxygen, and nitrogen atoms into the cage of electron density.

The crystals were mounted on a nylon loop, flash cooled, and stored as previously mentioned. X-ray data sets were collected at SER-CAT beam line 22 (Argonne National Lab, Chicago, IL) using a MAR CCD (charge coupled device) to record the diffraction images. The resulting images were indexed and integrated using DENZO and they were scaled with SCALEPACK program packages that were included in the HKL2000 software suite.[54] The resolution of the scaled data was limited to Bragg reflections with I/σ's above 2. The scaled diffraction data was further processed using TRUNCATE[55] and FREERFLAGS[56] to convert the scaled intensities to structure factors and to select the set of free reflections. The number of expected molecules in the asymmetric unit was determined by calculating the Matthew's coefficient. IPPase structures were phased by molecular replacement using either the program MOLREP,[57] which was included in the CCP4[58] program suite, or PHASER,[59] associated with PHENIX.[60] The *Pyrococcus furiosus* IPPase structure PDB ID: 1TWL was used as the initial search model.

2.4.2 Structure determination

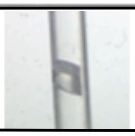
The resolution of the reflection data was determined based on an overall completeness of ~90% and an outer shell completeness of ~50%. All final structures were refined using *phenix.refine*[60] in order to refine the occupancies of bound ligands and all atoms modelled in alternate conformations. In the case of structure 3Q46, initial rounds of refinement were performed using REFMAC5.[61] The COOT[62] program package was used to make adjustments to the structural models by manually fitting the calculated F_o-F_c and $2F_o$-F_c electron density maps. In general, iterative rounds consisted of three cycles of restrained refinement, one cycle of solvent update, and one cycle of manual modeling within COOT. All adjustments were continued until the protein model was substantially complete, all metal ions and ligands were modelled, and a reasonable attempt had been made to account for all peaks above 5σ in the F_o-F_c density map. The restraints for metal ion and ligands coordination geometries were determined using *Readyset!*, a subprogram accessible through the *phenix.refine* graphical user interface. For structures in which the total number of refined atoms did not exceed 1/8 of the total number of unique reflections, the B-factors were refined anisotropically. Prior to validation, an extended round (10-15 cycles) of xyz, occupancy, and B-factor refinement was carried out in *phenix.refine*. During this round, the refinement target weights were optimized in the automatic mode and asparagine and glutamine B-factors were used to suggest the correct side chain orientation in cases of ambiguity. Final structural validations were performed using MOLPROBITY.[63] After each model was validated, a final round, which consists of an extended number of cycles (10-15) of occupancy and B-factor refinement in *phenix.refine*, was performed. A summary of the structure

determination and refinement strategies for IPPase structures are provide in Tables 2.3 and 2.4.

Table 2.3. Strategies for X-ray crystallographic structures determination

	X-ray diffraction Tt-IPPase complexed with SO_4	X-ray diffraction Tt-IPPase complexed with Mn^{2+}, SO_4	X-ray diffraction Tt-IPPase complexed with PP_i, Ca^{2+}	X-ray diffraction Tt-IPPase complexed with PP_i, Ca^{2+}	X-ray diffraction Tt-IPPase complexed with P_i, Mg^{2+}	X-ray diffraction Tt-IPPase complexed with Mg^{2+}, SO_4
Mounted crystal						
Data collection						
Strategy	Single wavelength X-ray	Single wavelength X-ray	Single wavelength X-ray	Single wavelength X-ray	Single wavelength X-ray	Single wavelength X-ray
Crystal Lattice	Triclinic	Primative Rhombohedral	Trigonal Hexagonal	Trigonal Rhombohedral	Primative Rhombohedral	Primative Rhombohedral
Beam line facility	SER-CAT 22BM	SER-CAT 22BM	SER-CAT 22ID	SER-CAT 22BM	SER-CAT 22BM	SER-CAT 22BM
Detector	MAR225 CCD	MAR225 CCD	MAR300 CCD	MAR225 CCD	MAR225 CCD	MAR225 CCD
Cryoprotectant	28% Glycerol	28% Glycerol	15% Glycerol, 15% MPD	20% PEG 4000	28% Glycerol	28% Glycerol
Data sets used	1	2	1	1	2	2
Indexing						
Program	Denzo HKL	Denzo HKL	Denzo HKL	Denzo HKL	Denzo HKL	Denzo HKL
Scaling						
Program	Scalepack	Scalepack	Scalepack	Scalepack	Scalepack	Scalepack
Structure factor calculations						
program	Scalepack2mtz (Truncate)	Scalepack2mtz (Truncate)	Scalepack2mtz (Truncate)	Scalepack2mtz (Truncate)	Scalepack2mtz (Truncate)	Scalepack2mtz (Truncate)
% of free reflections	5%	5%	5%	5%	5%	5%
Phasing						
Program	Phaser	Phaser	Molrep	Molrep	Molrep	Molrep
method	Molecular Replacement	Molecular Replacement	Molecular Replacement	Molecular Replacement	Molecular Replacement	Molecular Replacement
starting model	3I98	3I98	1TWL	1TWL	1TWL	1TWL
Refinement						
program	Phenix.Refine	Phenix.Refine	*Refmac, Phenix.Refine	*Refmac, Phenix.Refine	*Refmac, Phenix.Refine	*Refmac, Phenix.Refine
xyz refinement	all atoms	all atoms	all atoms	all atoms	all atoms	all atoms
ADP parameters	Anisotropic all atoms	Anisotropic all atoms	Anisotropic all atoms	Anisotropic all atoms	Anisotropic all atoms	Anisotropic all atoms
Aniso. Resolution limit	1.1	-	1.1	1.4	1.0	1.0
Occupancy	selected atoms	selected atoms	selected atoms	selected atoms	selected atoms	selected atoms
Ligand restraints	phenix_readyset	phenix_readyset	phenix_readyset	phenix_readyset, reel	phenix_readyset, reel	phenix_readyset, reel
Program used to add H	Reduce, elBOW	Reduce, elBOW	Reduce, elBOW	Reduce, elBOW	Reduce, elBOW	Reduce, elBOW
Model Building						
Program	Coot	Coot	Coot	Coot	Coot	Coot
Validation						
Program	Molprobity	Molprobity	Molprobity	Molprobity	Molprobity	Molprobity

Table 2.4. Strategies for X-ray crystallographic structure determination for capillary grown crystals

Complex	**X-ray diffraction Tt-IPPase complexed with Ca^{2+}**	**X-ray diffraction Tt-IPPase complexed with Ca^{2+}**	**Neutron diffraction Tt-IPPase complexed with Ca^{2+}**
Mounted crystal			
PDB ID	**3I98**	**3R5V**	**3Q3L**
Method	**Counter diffusion**	**Counter diffusion**	**Counter diffusion**
Volume of Crystal	0.90 mm^3	0.90 mm^3	6.00 mm^3
Base reagent	HSHT-A1	HSHT-A1	HSHT-A1
Protein buffer	Hepes pH 7.5	Hepes pH 7.5	Hepes pH 7.8
protein solution supplements	-	-	-
Ratio used protein:reagent	1:1	1:1	1:1
Crystallization pH	6.0	6.0	6.0
Data collection			
Strategy	Single wavelength X-ray	Single wavelength X-ray	Multi-wavelength, neutron
Crystal Lattice	C-centered Monoclinic	C-centered Monoclinic	C-centered Monoclinic
Beam line facility	SER-CAT 22BM	SER-CAT 22BM	Ladi-3
Detector	MAR225 CCD	MAR225 CCD	NSI
Cryoprotectant	-	-	-
Data sets used	1	1	3
Indexing			
Program	Denzo HKL	Denzo HKL	Lauegen
Scaling			
Program	Scalepack	Scalepack	Lauescale
Structure factors			
program	Scalepack2mtz (Truncate)	Scalepack2mtz (Truncate)	Phenix
% of free reflections	5%	5%	5%
Phasing			
Program	Molrep	Molrep	Phenix.Refine
method	Molecular Replacement	Molecular Replacement	Fix Model
starting model	1TWL	1TWL	3I98
Refinement			
program	*Refmac, Phenix.Refine	*Refmac, Phenix.Refine	Phenix.Refine
xyz refinement	all atoms	all atoms	-
ADP parameters	Isotropic all atoms	Isotropic all atoms	Isotropic all atoms
Aniso. Resolution limit	-	-	-
Occupancy	selected atoms	selected atoms	selected atoms, H/D
Ligand restraints	phenix_readyset	phenix_readyset	phenix_readyset, reel
Program used to add H	Reduce, elBOW	Reduce, elBOW	Reduce, elBOW
Model Building			
Program	Coot	Coot	Coot
Validation			
Program	Molprobity	Molprobity	Molprobity

2.5 Protein crystallization for neutron diffraction

2.5.1 Large volume crystal growth in restricted geometry

Purified IPPase at a concentration of 35 mg/ml in 50 mM HEPES pH 7.5, 50 mM NaCl was used for crystallization in quartz capillaries (Vitrocom) with inner diameters (ID) of 0.3 mm-1.0 mm using the counter-diffusion crystallization technique.[25; 38; 39] The protein solution was loaded from one end and allowed to fill to the mid point of the capillary followed by an equal volume of the precipitating reagent (25% 2-methyl-2-4-pentanediol, 0.1 M sodium acetate pH 4.6, and 10 mM calcium chloride) prior to being sealed with wax and enamel. Crystals were observed along the length of the capillary within 1-2 days of equilibration.

In collaboration with Juan M. Garcia-Ruiz, the counter-diffusion crystallization method in restricted geometry was adapted through a series of experiments to obtain large volume crystals for neutron diffraction studies. The initial crystallization conditions for obtaining IPPase crystals were determined from prior vapor-diffusion crystallization trials. Counter-diffusion crystallization trials were then employed in borosilicate capillaries to develop a reliable protocol for producing large crystals for room temperature X-ray diffraction studies. The largest and best diffracting crystals (based on resolution and I/σ of diffraction spots) were obtained in the condition 25% 2-methyl-2-4-pentanediol, 100 mM sodium acetate pH 4.6, and 10 mM calcium chloride. This optimized condition was subsequently used for a series of counter-diffusion crystallization experiments in restricted geometry based on the current methodologies used in the laboratory[64; 65; 66]. Ten-centimeter long quartz capillaries (Vitrocom) were used with ID's of up to 1.0 mm. The precipitating agent and protein solutions were

arranged in juxtaposition to one another inside the capillary tube, as described by Ng *et al*.,[67] except that both solutions were entirely sealed within a single capillary of uniform diameter. Specifically, half the volume of a capillary tube was filled with purified IPPase in a buffer consisting of 0.05 M HEPES pH 7.5, 0.05 M sodium chloride) at a concentration of 35 mg/ml, and it was allowed to diffuse against an equal volume of the same precipitating solution as mentioned above. The tubes were filled by capillary action and sealed with soft wax and enamel. Experiments were conducted using slight variations in the ratios and concentrations of the components for 0.3-0.7 mm I.D. capillaries.

2.5.2 Room temperature X-ray diffraction analysis of capillary-grown crystals

A single IPPase crystal obtained in a 1 mm diameter quartz capillary was used for *in situ* X-ray data analysis at 298K (25˚C). The entire capillary containing the targeted crystal was mounted horizontally and centered directly in front of the X-ray source. The IPPase crystal filled up the diameter of the capillary and was immobilized by its attachment to the capillary wall. X-ray diffraction data was recorded with a MAR225 CCD (charge coupled device) detector at the SERCAT beam line BM22 (Argonne National Lab, Chicago, IL) using 0.92Å radiation. The beam was attenuated to 20% intensity through a narrow slit collimator. Exposure time for each image was 1 second for each 1.0˚ oscillation angle recorded by the detector. Data were processed and reduced within HKL2000.[54] The resulting data was used for structure determination of calcium coordinated IPPase at room temperature (PDB ID: 3R5V and 3I98).

2.5.3 Growth of crystals in larger diameter quartz capillaries

The ability to routinely obtain large crystals using the previously described capillary counter-diffusion crystallization method was found to diminish when using capillaries with inner diameters (I.D.) greater than 0.7 mm, due to the increase in convection during the equilibration process. Increasing the viscosity of the solutions and incorporating a 1% low melting temperature agarose plug between the protein and precipitant could circumvent this problem. The initial test experiments were performed using 1.5 mm-2 mm I.D. capillaries adapted to a 1.5 ml syringe by an equivalent diameter section of tygon tubing. The syringe was required to pull the solutions in place and position the 1% low melting temperature agarose plug within the capillary tube. The proper installation and placement of the agarose plug allows the precipitant to diffuse more slowly into the protein solution and prevented heavy precipitation from forming at the interface of the protein and precipitating solutions. The concentration of protein necessary to routinely obtain large volume crystals in this format was determined to be approximately 60 mg/ml. However, crystals approaching 8 mm^3 in volume have been obtained at concentrations as low as 25 mg/ml. A precipitating solution comprised of 25% 2-methyl-2-4-pentanediol (MPD), 100 mM sodium acetate pH 4.6, 10 mM calcium chloride was most amenable to large crystal growth by this technique.

The general process is similar to the process described for small diameter capillary tubes. However, extra care must be taken to completely seal both ends with soft wax, followed by the application of two layers of fast-drying enamel. Effort was required to prevent or remove any air pockets remaining in the capillaries prior to sealing. In some cases it was necessary to add a small volume (<10 μL) of protein solution or precipitant

solution to the end of the capillary just before sealing with it with wax. The crystallization process took place at 293K (20˚C) in a low-temperature thermal-coupled incubator. The crystallization experiments were performed in 1.0, 1.2, 1.5, 2, 2.5, and 3 mm I.D. capillaries. The largest crystals were obtained within quartz capillaries with diameters between 2-4 mm against the precipitating agent containing 10 mM calcium chloride dihydrate, 0.1 M sodium acetate trihydrate pH 4.6, and 25% 2-methyl-2,4-pentanediol. Crystals as big as 8 mm^3 were obtained after 6-8 weeks equilibration at 293K (20˚C).

2.5.4 Deuterium exchange of capillary-grown crystals

In neutron diffraction, significant background noise results from the incoherent scattering properties of hydrogen atoms. Several methods have been employed to replace exchangeable hydrogen atoms with deuterium. The general process utilized most commonly for neutron crystallographic study is vapor diffusion exchange of crystals mounted in capillaries.[44]

Since conditions have been identified to grow IPPase crystals to fill the diameter of a capillary, the deuterium exchange process was performed by exchanging solvents without physical manipulations of the protein crystal. The hydrogenated mother liquor was carefully removed and replaced with an equivalent solution containing D_2O. The exchange procedure was performed by sequentially opening each end of the capillary and gently aspirating away the equilibrated protein solution. This was accomplished by using a narrow diameter capillary to carefully remove the hydrogenated solution, leaving the largest crystal undisturbed within a small amount of residual mother liquor. The deuterated solution was then layered back into the capillary from both sides, leaving an

empty, air-filled space between the solution and the crystal. After one side had been filled with deuterated solution, the open end was filled with waxed and double sealed with fast drying enamel prior to repeating the same procedure on the remaining end of the capillary tube. Crystals obtained in the presence of an agarose plug were exchanged from one side by placing the deuterated solutions directly against the agarose plug, allowing free liquid diffusion. The optimal conditions for exchange occurred when the protein crystals grew attached to the agarose plug, providing a cushion and added stability during transport exchange procedures.

2.5.5 X-ray analysis of deuterium-exchanged crystals in capillaries

A single IPPase crystal obtained in a 1 mm diameter quartz capillary was subjected to deuterium exchange by vapor diffusion. The exchange was performed by removing the surrounding liquid around the targeted crystal and replacing it with a 78% precipitating reagent formulated in D_2O to mimic the natural mother liquor. The initial exchange was performed with 78% deuterated solution. After one week, the solution was replaced with a 91% deuterated solution formulated with uniformly deuterated MPD (CDN isotopes). The solution was removed and replaced a second time to ensure that a high percentage of deuterium exchange was achieved. The crystal was subsequently used for *in situ* X-ray data analysis at 298K (25˚C) by mounting the entire capillary directly onto the goniometer head using a soft mounting clay. The X-ray diffraction data was recorded with a MAR225 CCD (charge coupled device) detector at the SERCAT beam line BM22 (Argonne National Lab, Chicago, IL) using 0.92Å radiation. The beam was attenuated to 20% intensity through a narrow slit collimator. Exposure time for each image was 1 second for each 1.0˚ oscillation angle recorded by the detector. A total of

180 images was collected. A sharp decline in diffraction quality and resolution refection occurred after 140 diffraction images were collected. As a result only 149 of the 180 images collected were included in the final data set. The highest quality data collected on the deuterium-exchanged and non-exchanged crystals were compared to determine any observable changes in the crystal quality resulted from the exchange manipulation.

2.5.6 Preliminary neutron diffraction of large volume crystals

The initial crystals grown using the modified capillary growth technique were tested to determine their diffraction quality. An IPPase crystal grown in a quartz capillary tube (inner diameter 2.0 mm) was measured to be about 5.0 mm^3 and targeted for preliminary neutron diffraction analysis. Labile hydrogen atoms were exchanged with deuterium by removing the surrounding liquid around the targeted crystal and allowing it to equilibrate against a D_2O mother liquor solution, which contained 78 – 91% deuterated precipitating agent, including deuterated MPD in D_2O (CDN isotopes) in the capillary. The exchange was performed by vapor diffusion against an air space between the crystal and the new deuterated solution for a period of one month with biweekly exchanges of new deuterated solutions. A wavelength-resolved time-of-flight neutron Laue diffraction image of a single IPPase crystal setting was recorded at the Protein Crystallography Station (PCS) on the spallation neutron source operated by Los Alamos Neutron Science Center (LANSCE).[68] The crystal was exposed for a duration of 16 hours. The crystal-to-detector distance was 730 mm, giving 120˚ detector coverage in the horizontal plane with 16˚ of coverage in the vertical plane. The neutrons were generated from a tungsten target in pulses at a frequency of 20 Hz, and their energies

were moderated by a specially designed, partially coupled water moderator. The primary neutron flight path was 28 meters long, and the wavelength range and angular divergence of the incident neutron beam were shaped by a composite T0/T1 chopper, along with a series of collimating inserts to 0.6-6Å and 0.12_{o}. Diffraction data were processed using a modified version of d*TREK.[69]

One of the advantages of a spallation source is that diffraction data is collected with excellent wavelength resolution for all diffracted neutrons. The data is thereby "wavelength resolved," thus the Laue pattern is integrated wavelength by wavelength as if it were the result of several monochromatic experiments. Consequently, there is a massive reduction in the background beneath each reflection, increasing the signal to noise ratio. These initial experiments were used to prove the suitability of the crystal neutron diffraction studies and to assess the effectiveness of the deuterium-exchanged procedure. As a result, additional neutron beam time was awarded based on an accepted proposal to utilize the obtained crystals for neutron structure determination. Freshly grown crystals were tested to assess their ability to diffract neutrons and to confirm the efficiency of the deuterium exchange method a second time using the Laue diffractometer (LADi-III)[70] at the Institute Laue-Langevin (ILL) in Grenoble, France. The primary motivation of this test was to determine the suitability of IPPase crystals for subsequent complete data set collection for neutron structure determination.

2.6 Neutron macromolecular crystallography

2.6.1 Collection and processing of diffraction data

A total of three IPPase crystals grown in a quartz capillary tube (I.D. 2.0 mm) ranging in volume from 5.0-8.0 mm^3 were used for neutron diffraction analysis. Labile hydrogen atoms were exchanged with deuterium by removing the surrounding liquid around the targeted crystal and by allowing it to equilibrate against a D_2O mother liquor solution (10 mM calcium chloride dihydrate, 50 mM sodium acetate pH 4.6, 25 mM HEPES pH 7.8, 50 mM sodium chloride, and 12.5% MPD) in two steps (~78%, ~95% deuterated). In the case of the 95% deuterated solution, deuterated MPD (CDN isotopes) was used, and non-buffer components were made from dried powder dissolved in D_2O. Buffers were prepared as 2 M stocks with H_2O. The exchange was performed by vapor diffusion against an air space between the crystal and the new deuterated solution for a period of one month with biweekly exchanges of new deuterated solutions. Neutron Laue diffraction data were recorded for the deuterium exchanged IPPase crystals at room temperature using the LADi-III diffractometer at ILL in Grenoble, France. The cylindrical area detector of the LADi III diffractometer allowed a large number of Bragg reflections to be recorded simultaneously. A photo and diagram of the LADi-III beamline detector is shown in Figure 2.4. The neutron radiation was filtered with a nickel/titanium multi-layer band-pass. Consequently, neutron wavelength range ($\Delta\lambda/\lambda \sim 25\%$) corresponding to 3.1-4.2Å centered at 3.7Å was used for the data collection. The data were recorded from three different crystals, each collected at two different orientations. In each case, the diffraction image represented 6-12 hour exposure to the neutron beam with each image accounting for a 7˚ wedge of the vertical

rotation axis relative to the axis of the detector. Neutron Laue data were indexed and integrated using LAUEGEN[71] and modified for the cylindrical geometry of the LADi-III detector. The LSCALE[72] program was used to derive the wavelength-normalization curve by using the intensities of symmetry-equivalent reflections, which were measured at different wavelengths. SCALA[73] was used to combine and merge the observed data. A D_{min} of 1.0 was used, and the resolution of the data was limited to 2.5Å in the final merged data set, which resulted from the 38 collected diffraction images.

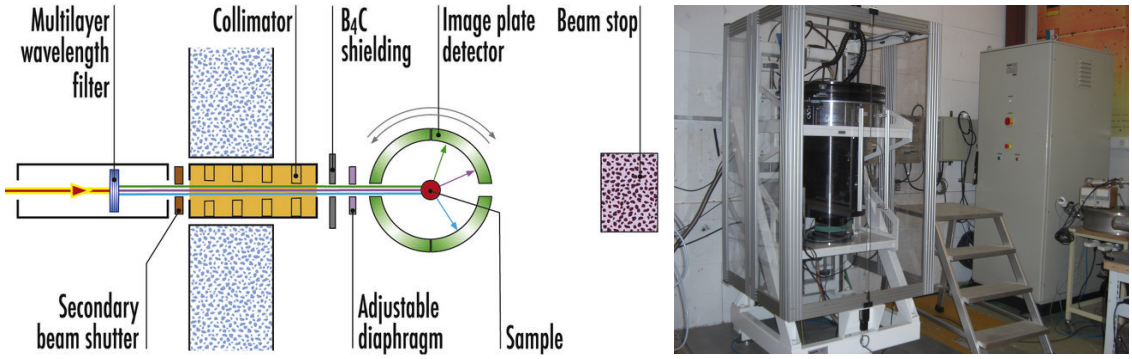


Figure 2.4. Description of the LADi-III cyclindrical detector used to collect the neutron data for IPPase. Shown is a diagram and photograph of the detector of the LADi-III neutron diffractometer and detector at the Institute Laue-Langevin (ILL) in Grenoble, France.

2.6.2 Neutron crystallographic structure determination

The 1.65Å X-ray crystallographic structure, which was deciphered from the IPPase crystal within the C2 monoclinic space group at room temperature (PDB ID: 3R5V), was used as a starting model for the neutron structure determination. The structure coordinates of the IPPase X-ray crystallographic model were modified by removing low occupancy alternate conformations and water molecules. Thereafter, hydrogen and exchangeable deuterium atoms were added to the initial protein model using the *ReadySet!* program in the PHENIX (Version 1.5.2) program suite. The isotropic atomic displacement parameters (ADP) and occupancies of hydrogen and deuterium at

exchangeable positions, as well as residues, which were also modeled in alternate conformations, were refined over the resolution range of 2.5-19.4Å using the *phenix.refine* program.

2.6.3 Structural refinement and validation

Real-space refinement was performed in *Coot* on regions of the structure having low $2F_o$-F_c and F_o-F_c. A subsequent round of ADP parameter and occupancy refinement was performed. Water was automatically added to F_o-F_c to the model using the default settings of phenix.refine "update water". This refinement strategy was repeated for several cycles, each containing 8 rounds of ADP and occupancy refinement coupled to least squares fitting and manual adjustments to the model in *Coot.*[62] The *ReadySet!* Program was then used to add deuteriums to the modeled water molecules. The nuclear density maps were then compared to the maps calculated with only oxygen atoms modeled at these positions. In the case where positive F_o-F_c nuclear density was observed to coincide with the positions of the model deuterium atoms that surround the modeled oxygen atom correlated to positive (>0.5σ) $2F_o$-F_c nuclear density, the modeled D_2O coordinates were retained. Manual adjustment of all D_2O molecules was completed using both $2F_o$-F_c and F_o-F_c nuclear density maps. The final structure was deposited in the Protein Data Bank (PDB) and was assigned the accession code 3Q3L.

2.6.4 Site-directed mutagenesis

The Quik Change site-directed mutagenesis kit (Stratagene) was used to make point mutations leading to selective amino acid changes in IPPase. Primers incorporating the desired point mutation were designed to be complementary to the

sequence confirmed IPPase expression vector. The primers were designed according to manufacturers recommendations. According to primer design guidelines, each of the designed primer pairs were approximately 40 nucleotides in length with as close as possible melting temperatures of ≥78˚C based on calculations of length and %GC content. Each primer contained the desired mutation near the middle of the oligomers with at least 10 bases of perfectly complimentary sequences on each side of the mutation. The sequences of the designed primer pairs for each targeted mutation are shown in Table 2.5. The mutagenesis reaction was performed by mixing each primer pair with the isolated IPPase encoding recombinant plasmid on ice. PCR extension was subsequently performed by using 12 cycles in the presence of *PfuTurbo* DNA polymerase and dNTP. As a result, mutated plasmid constructs were created containing staggered nicks. The resulting product was then treated with *Dpn* I, a methyl specific endonuclease, in order to cleave the template strands of the original plasmid template. The dpnI treatment was performed at room temperature according to the manufacturer's recommendations. The remaining reaction mixture containing the uncleaved mutated copies of pET3a was transformed into XL1-Blue supercompetent cells by the same heat shock procedure previously described for wild-type plasmid transformations. Colonies were selected and grown overnight in selective LB miller broth to propagate the recombinant plasmid. The plasmids were then extracted by standard molecular biology protocols[22] using a Qiagen Plasmid Mini kit according to the manufacturer's instructions. Finally, the purified plasmids were sequenced by MWG DNA sequencing service to identify the clones that contained the correct IPPase coding sequence.

Table 2.5. The primer sequences used for site-directed mutagenesis of IPPase

Target mutation	Primer type	Sequence	Length	Tm (˚C)	Duplex energy @68˚C (kcal/mole)	Energy cost of mismatch(%)
His5Phe	FORWARD PRIMER	**5'-atatgatgaacccgttcttcgagcttgagcccggac-3'**	36	78.83	-49.48	6.40%
CAC➔TTC	REVERSE PRIMER	**5'-gtccgggctcaagctcgaagaacgggttcatcatat-3'**	36	78.83	-50.95	8.50%
Asn2Ala	FORWARD PRIMER	**5'-agaaggagatatacatatgatggatccgttccacgagcttgagc-3'**	44	79.32	-50.59	8.5
AAC➔GAU	REVERSE PRIMER	**5'-gctcaagctcgtggaacggatccatcatatgtatatctccttct-3'**	44	79.32	-52.41	6.7
Lys35Arg	FORWARD PRIMER	**5'-agaaggagatatacatatgatggatccgttccacgagcttgagc-3'**	47	79.07	-56.68	6.6
AAG➔GCG	REVERSE PRIMER	**5'-gctcaagctcgtggaacggatccatcatatgtatatctccttct-3'**	47	79.07	-53.75	4.9
Lys35Ala	FORWARD PRIMER	**5'-caggaacaagtacgagctcgaccgtaagactggtttacttaagctcg-3'**	42	79.21	-50.63	3.8
AAG➔GCG	REVERSE PRIMER	**5'-cgagcttaagtaaaccagtcttacggtcgagctcgtacttgttcctg-3'**	42	79.21	-48.72	0.4
Lys35Met	FORWARD PRIMER	**5'-ggaacaagtacgagctcgacgcgaagactggtttacttaagc-3'**	39	78.45	-47.14	3.6
AAG➔ATG	REVERSE PRIMER	**5'-gcttaagtaaaccagtcttcgcgtcgagctcgtacttgttcc-3**	39	78.45	-42.33	0.8
Asp68Asn	FORWARD PRIMER	**5'-ggtacgacgacggcaacccctttcgacatc-3'**	29	80.22	-46.22	7.6
GAC➔AAC	REVERSE PRIMER	**5'-gatgtcgaaggggttgccgtcgtcgtacc-3'**	29	80.22	-44.88	9.1
Asp71Asn	FORWARD PRIMER	**5'-gacggcgacccctttcaacatcatggtcatca-3'**	31	78.98	-45.99	7.7
GAC➔AAC	REVERSE PRIMER	**5'-tgatgaccatgatgttgaaggggtcgccgtc-3**	31	78.98	-44.88	9.1
Glu32Asp	FORWARD PRIMER	**5'-agcaggaacaagtacgatctcgacaagaagactgg-3'**	35	79.27	-48.23	4.9
GAG➔GAU	REVERSE PRIMER	**5'-ccagtcttcttgtcgagatcgtacttgttcctgct-3'**	35	79.27	-44.32	1.4
Tyr56Phe	FORWARD PRIMER	**5'-cgttcttctatccggtcgacttcggaataatcccg-3'**	34	80.44	-49.69	3.8
TAC➔TTC	REVERSE PRIMER	**5'-cgggattattccgaagtcgaccggatagaagaacg-3'**	34	80.44	-48.53	1.4
Tyr140Phe	FORWARD PRIMER	**5'-cacttcttccagaggttcaaggagctgcaagga-3'**	33	79.14	-49.4	3.9
TAC➔TTC	REVERSE PRIMER	**5'-tccttgcagctccttgaacctctggaagaagtg-3'**	33	79.14	-44.35	1.5
His135Ala	FORWARD PRIMER	**5'-cctcgacgagatcgccgccttcttccagaggtac-3'**	34	79.88	-47.24	9.3
CAC➔GCC	REVERSE PRIMER	**5'-gtacctctggaagaaggcggcgatctcgtcgagg-3'**	34	79.88	-49.92	5.4
Glu142Gln	FORWARD PRIMER	**5'-cttccagaggtacaagcagctgcaaggaaagac-3'**	33	79.14	-42.21	7.8
GAG➔CAG	REVERSE PRIMER	**5'-gtctttccttgcagctgcttgtacctctggaag-3'**	33	79.14	-38.66	9.7
Ser48Ala	FORWARD PRIMER	**5'-cgaccgcgtcctttacgccccgttcttctatccg-3'**	34	79.88	-47.38	7.7
AGC➔GCC	REVERSE PRIMER	**5'-cggatagaagaacggggcgtaaaggacgcggtcg-3'**	34	79.88	-49.51	7.9

2.6.5 Recombinant expression of mutants

Recombinant mutants were expressed and purified in *E. coli* by methods similar to that performed for the wild type enzyme. The mutant cell cultures were grown in selective LB miller broth to an OD of 0.6 prior to being induced by the addition of 1 M isopropyl β-D-1-thiogalactopyranoside to a final concentration of 1 mM. The cultures were grown for 16 hours at a temperature of 25˚C before being harvested by centrifugation. Protein purification was performed on the mutant cell extracts. In brief, the cells were lysed by ultra-sonification, centrifuged at 10,000 rpm for 30 minutes to

remove cellular debris, and the supernatant was then subjected to a heat cut at 75˚C for 30 minutes. The heat-precipitated material was removed by centrifugation of the suspension at 10,000 rpm. The clarified supernatant was then loaded onto a 5 ml Q-sepharose ion exchange column (GE Healthcare) and pre-equilibrated with buffer as previously described for wild type enzyme purification. The protein was eluted from the anion exchange column by a two-step elution procedure using 30 ml syringe. The first step was used to wash the column with 15 ml of 0.10 M NaCl containing buffer solution (50 mM Tris pH 7.5) to remove unbound sample. In the second step 10 ml of 0.35 M NaCl containing buffer solution was passed through the column to elute IPPase from the column. The eluted protein was then concentrated to 2 ml and loaded onto a superdex S200 gel filtration column to further separate contaminants based on molecular weight. The IPPase enzyme was eluted from the gel filtration column in 3 ml fractions by passing approximately 75 ml of buffer through the column. The fractions from gel filtration were pooled and concentrated to 10 mg/ml for crystallization trials.

2.6.6 Activity assays for mutants

The colorimetric molybdate assay was used to measure IPPase activity under the optimal conditions determined from biochemical studies of the wild-type enzyme. The assayed activity was based on the amount of P_i generated at 75˚C in a pH 7.5 50 mM Tris buffer containing 5 mM magnesium chloride. A 0.1 µmol sodium phosphate solution was used as a standard to correlate absorbance at 660 nm with concentration of liberated P_i from the reaction in order to determine the activity of the enzyme as a function of different metals. The activity unit is defined as the µmol of PP_i hydrolyzed per min per mg of protein. The experiments were repeated for each mutant to establish the relative

activity of each compared to a wild-type control. For all experiments, the reaction was initiated in the same manner as previously described by adding the enzyme to 1 ml aliquots in a 96-well block. In 1-minute intervals, 50 µl was taken at 1-minute intervals and placed on ice in a 96-well 100 µl assay plate for the duration of the experiment. A solution containing 1% ammonium heptamolybdate (w/v), 0.83 M sulfuric acid, 8% iron (II) chloride (w/v), and 2% SDS was added to stop the reaction and induce the color change, as in the method described by van Alebeek *et al.*[52] The absorbance of each of samples was measured after 30 minutes at the tested temperature. The best activating metal, optimal temperature, and pH were determined. Three experiments were performed for each test condition and the measured activity was averaged of the three experiments were averaged.

CHAPTER 3

RESULTS

3.1 Primary structure analysis

The primary structure of IPPase shares a high level of amino acid sequence similarity with others derived from archaea and thermophilic bacterial. Compared to other archaeal IPPases of known three-dimensional structures, *T. thioreducens* IPPase shares 88% and 87% sequence identity with those from *P. furiosus* and *P. horikoshii* respectively. Less sequence identity are observed when *T. thioreducens* IPPase is compared with bacterial IPPases, revealing 51% for *S. acidocaldarius* and 41% for *E. coli.* Among the conserved active site residues are the five aspartate (66,68,71,98,103) and one glutamate (32) residues that form divalent metal binding sites. In addition, two lysine residues (141, 30), two tyrosine residues (140,156), and one arginine residue (44) are also conserved and they coordinate the PP_i substrate. A pairwise amino acid sequence alignment of IPPase with those from other IPPases is illustrated in Figure 3.1.

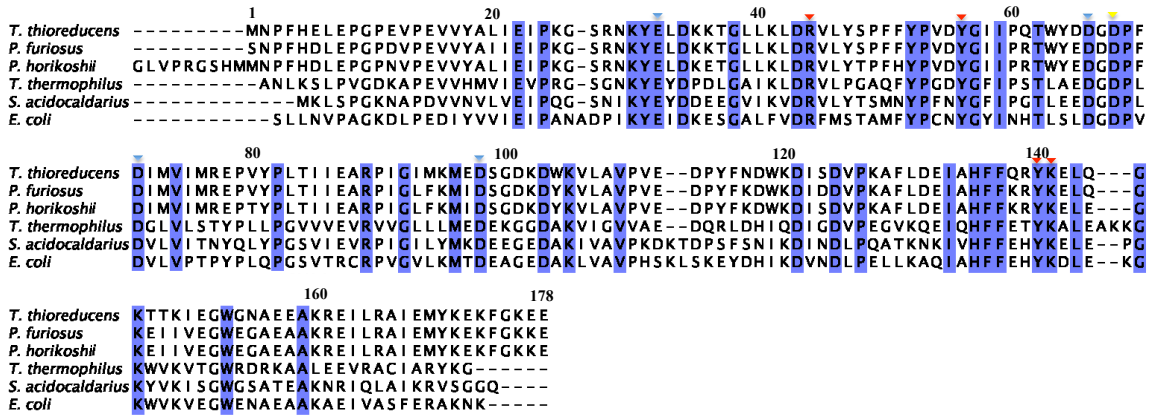


Figure 3.1. Conserved amino acid residues found in prokaryotic Family I. The amino acid sequences of IPPase from *P. furiosus*, *P. horikoshii*, *T. thermophilus*, *S. acidocaldarius,* and *E. coli* enzymes have been aligned using the Clustalw multiple sequence alignment program[74]. The conserved residues have been highlighted blue. Among conserved residues are the four asparate (66,71,98,103), and one glutamate (32) that form the divalent metal binding sites (identified with blue arrows), and the three lysine residues (105, 141, 30), two tyrosine residues (140,156), and one arginine residue (44) that coordinate the substrate (red arrows). The proposed catalytic aspartate residue (68) is also conserved in IPPase (yellow arrow). A single histidine residue (135) is conserved. Hydrophobic amino acids located at the dimer interface and hydrophilic residues forming ionic pairs that stabilize the trimer are conserved. Several glycines and two prolines are also conserved. Thermally stable IPPases have loosely conserved charged residues near the carboxyl terminus.

Other conserved residues found in IPPase include two critical glycine (e.g., 67,100) and two proline residues (69, 82). These residues are believed to be important for maintaining the structural fold and proper balance of rigidity and flexibility in IPPase. Semi-conserved charged amino acids not seen in all IPPases are found near the carboxyl terminus of IPPase. These residues may contribute to the enzyme's thermal stability, as suggested by studies reported for other enzymes derived from hyperthermophilic archaea, in which similar ionic networks have been shown to prevent unfolding of the termini at elevated temperatures.[7]

3.2 Protein expression and purification

Recombinant IPPase can be expressed in *E. coli* using the T7 expression system[75] and purified to homogeneity in four major steps (Figure 3.2). The initial apparent molecular weight of the protein was estimated to be about 20 kDa based on SDS-PAGE

analysis. This value is consistent with the calculated value deciphered from the translated nucleotide open reading frame for IPPase. More than 75% purity can be achieved after a 30 minute heat treatment at ~80˚C. Purified IPPase can ultimately be obtained having more than 90% homogeneity after two chromatography steps. An approximate yield of 90 mg of purified protein can be routinely obtained from a 2 L culture.

IPPase was observed to be resistant to denaturation at temperatures above 85˚C. The enzyme's thermal stability requires the presence of NaCl. A concentration of 5 mM NaCl was deemed to be sufficient to maintain thermal stability during the purification procedure. The enzyme exhibited a high solubility of greater than 100 mg/ml in aqueous solutions over a broad range of pH and buffer conditions.

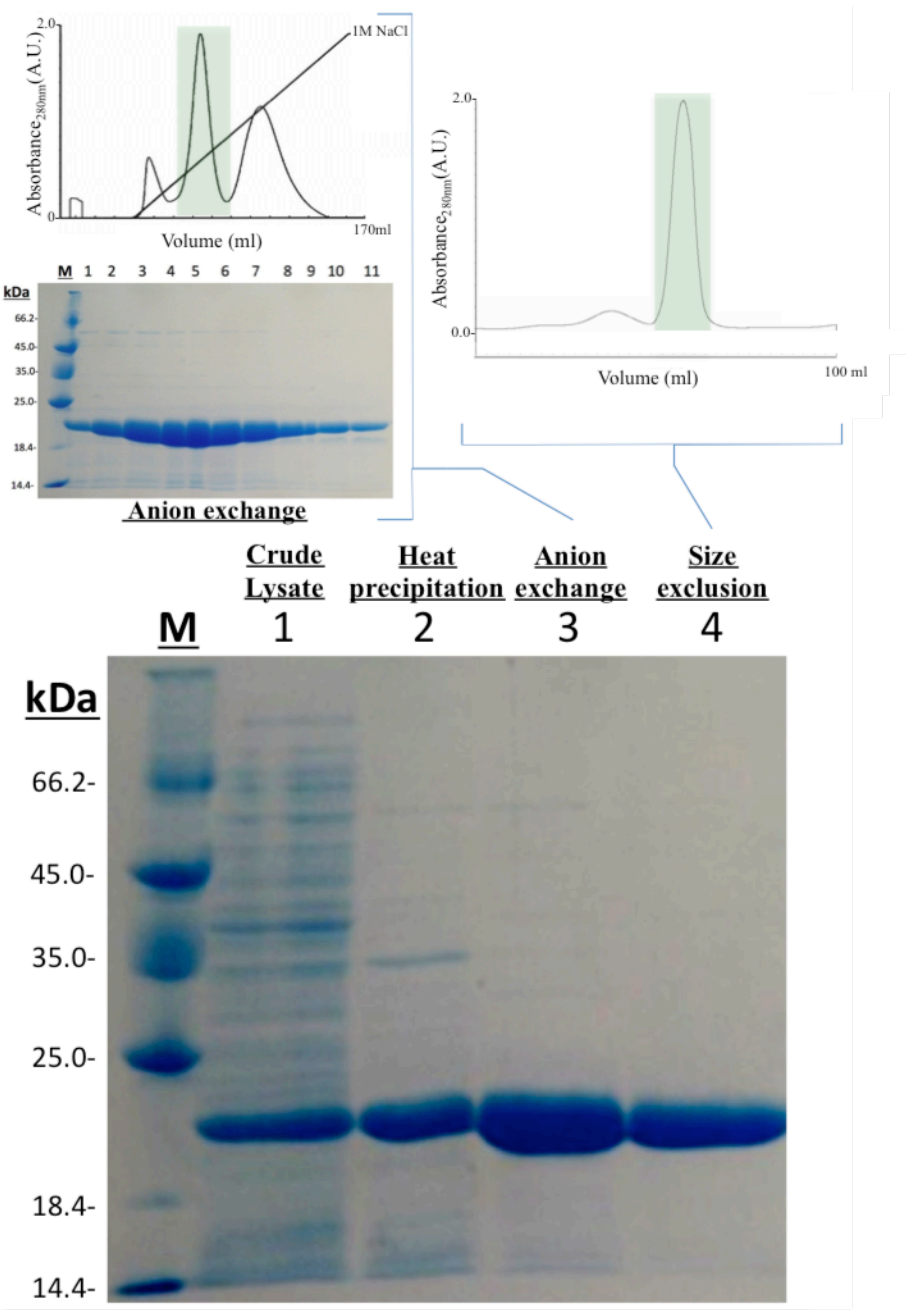


Figure 3.2. IPPase purification steps. The purification of IPPase entails an initial heat selection step, anion exchange chromatography, (shown top-left) and size-exclusion chromatography (shown top-right). The purity of IPPase was evaluated on a 12% SDS-PAGE as a function of the purification steps (lower panel). A molecular weight marker was loaded into Lane M. Lanes 1-4 were loaded with samples from the crude cell lysate, heat selection, anion exchange separation, and size-exclusion chromatography steps respectively.

3.3 IPPase enzymatic activity

The activity of the IPPase enzyme was assayed at 75˚C to compare the enzymatic activity in the presence of different divalent metal cations. As expected, magnesium was the best activator towards PP_i hydrolysis. The activity observed in the presence of the tested divalent metals was quantified as percentage of maximal activity as shown in Figure 3.3. Ni^{2+} and Ca^{2+} displayed the lowest activity of the metals tested. The remaining metals tested all showed significant activity towards substrate hydrolysis with Co^{2+} and Mn^{2+} having approximately 12% and 25% activation efficiency compared to Mg^{2+} respectively.

IPPase activity was also assayed as a function of pH at 75˚C in the presence of Mg^{2+}. The highest activity was measured at pH 9.0 when buffered by Tris. However, the enzyme in Tris buffer at pH 7.5 and 8.5 also displayed high activity towards substrate hydrolyisis with activities of greater than 80% compared to the optimal pH. In these assays the actual pH at 75˚C was experimentally determined to be ~8.7. When HEPES buffer was used in place of Tris at equivalent pH's, the catalytic activity was observed to decrease significantly. The activity measured at pH 5.5 is approximately half and one-third of measured at pH 6.5 and pH 7.0 respectively. The wild-type enzyme was tested in a Tris buffer at pH 7.5 to determine the activity towards substrate hydrolysis as a function of temperature. The highest activity observed using the assay conditions described was measured at 75˚C. However, recent studies additional (not reported in this dissertation) have shown an increase in activity in the presence higher concentrations of sodium chloride (>10 mM).

An Asp68Asn mutant was created to evaluate if the role of Asp68 was critical in catalysis. The enzymatic activity of the Asp68Asn mutant was measured and compared to that of the wild-type enzyme. As a result, in the mutant IPPase as much as 15% of the wild-type activity was retained.

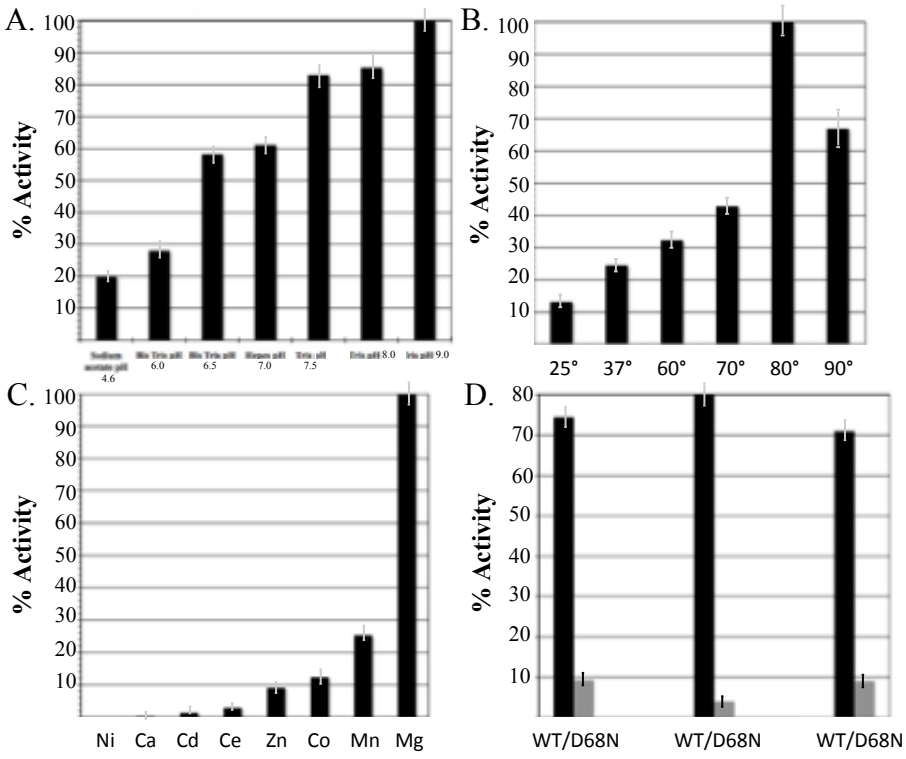


Figure 3.3. Activity measurements on IPPase. IPPase activity was measured under different conditions with varying buffer types, pH, temperature, and metals. A. The influence of different buffer types with varying pH on activity was measured. The activity percentage was experimentally measured in respect to the conditions that gave rise to the highest activity at Tris pH 9.0. B. Similarly, the percent activity was determined as a function of temperature when the highest activity at ~80˚C was used as a reference.
C. Magnesium gave rise to the highest activity out of eight divalent metals examined. D. Three different experiments comparing the activity of the wild-type IPPase to that of the Asp68Asn mutant in a 7.5 pH Tris buffer. The data shown represents the average of three experimental sets with error ranges shown by the error bars.

3.4 Crystallization for X-ray diffraction

3.4.1 Initial crystallization screening

A commercial sparse-matrix screen (Hampton Screen High-Throughput crystallization, HSHT) was employed to find the initial crystallization conditions for IPPase. As a result, several conditions yielded crystals suitable for X-ray diffraction analysis. These conditions include 2.0 M ammonium sulfate, 5% 2-propanol; 0.1 M Tris pH 8.5, 25% v/v tert-Butanol; 0.1 M sodium acetate trihydrate pH 4.6, 2.0 M ammonium sulfate; 1.5 M ammonium sulfate, 0.1 M Tris pH 8.5, 12% glycerol; 0.01 M cobalt(II) chloride hexahydrate, 0.1 M MES monohydrate pH 6.5, 1.8 M ammonium sulfate; 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% v/v (+/-)-2-methyl-2,4-pentanediol; and 0.01 M zinc sulfate heptahydrate, 0.1 M MES monohydrate pH 6.5, 25% v/v polyethylene glycol monomethyl ether 550.

Further crystallization screening experiments led to the identification of additional conditions for IPPase crystallization that proved to be adequate for X-ray diffraction analysis. Concentrated IPPase supplemented with 10 mM magnesium chloride diffraction quality crystals were obtained using the conditons 0.1 M sodium chloride, 0.1 M BICINE pH 9.0, 20% v/v polyethylene glycol monomethyl ether 550; 0.1 M HEPES pH 7.5, 70% v/v (+/-)-2-methyl-2,4-pentanediol; and 0.2 M sodium chloride, 0.1 M sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-methyl-2,4-pentanediol.

Crystals of a number of metal coordinated IPPase enzyme complexes were obtained by repeating the initial screening procedure in the presence of different divalent metals. Of the metals tested, only magnesium and manganese yielded new conditions that produced crystals acceptable for X-ray diffraction. The conditions containing

magnesium chloride include 0.1 M sodium chloride, 0.1 M BICINE pH 9.0, 20% v/v polyethylene glycol monomethyl ether 550; 0.1 M HEPES pH 7.5, 70% v/v (+/-)-2-methyl-2,4-pentanediol; and 0.2 M sodium chloride, 0.1 M sodium acetate trihydrate pH 4.6, 30% v/v (+/-)-2-methyl-2,4-pentanediol. Crystals obtained under these conditions diffracted X-rays to atomic resolution for structure determination.

In the case of IPPase supplemented with manganese chloride, only the condition 0.1 M HEPES pH 7.5, 10% w/v polyethylene glycol 8,000, 8% v/v ethylene glycol yielded crystals that were usable for subsequent X-ray diffraction analysis. Substrate-bound IPPase was crystallized in a similar manner and the resulting crystals were both analyzed at the home laboratory and synchrotron X-ray source. The optimized vapor diffusion crystallization conditions used to determine the structures reported in this dissertation are summarized in Table 3.1.

Table 3.1. Crystallization conditions for IPPase

Crystallized component	Crystal lattice, molecules per asymmetric unit	Protein solution	Precipitate solution	Ratio of solutions (Equilibration time, Temperature)
Metal-free IPPase bound to sulfate	Triclinic, six	15 mg/mL IPPase 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.001M EDTA	1.0M ammonium sulfate, 10% v/v 2-propanol	1:1 (6 weeks, 293K)
Magnesium-coordinated IPPase bound to sulfate	Tringonal rhombohedral, three	15 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.01M magnesium chloride	0.1M sodium chloride, 0.1M BICINE pH 9.0, 20% v/v PEG-MME 550	1:1 (3 months, 293K)
Manganese-coordinated IPPase	Trigonal rhombohedral, two	15 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.005M manganese chloride	0.1M HEPES pH 7.5, 10% w/v PEG 8,000, 8% v/v ethylene glycol	1:1 (4 weeks, 293K)
Magnesium-activated product complex	Trigonal rhombohedral, one	10 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.005M sodium pyrophosphate	0.1M magnesium chloride, 0.1M sodium HEPES pH 7.5, 15% v/v 2-propanol	1:1 (2 months, 293K)
Calcium-inhibited substrate complex I	Trigonal, rhombohedral, one	10 mg/mL IPPase. 0.05M TRIS pH 7.5, 0.05 M sodium chloride, 0.005M sodium pyrophosphate	0.2M calcium chloride, 0.1M sodium acetate pH 4.6, 30% v/v 2-propanol	2:1 (2 weeks, 293K)
Calcium-inhibited substrate complex II	Trigonal, hexagonal, three	10 mg/mL IPPase, 0.05M TRIS pH 7.5, 0.05M sodium chloride, 0.005M sodium pyrophosphate	0.02M calcium chloride, 0.1M sodium acetate pH 4.6, 20% v/v MPD	(2:1 or 3:1) (7 days, 293K)

Most of the IPPase crystals observed had well defined habits, were prismatic, and showed strong birefringence when observed under polarized light. Crystal sizes ranged from 0.05 mm to 0.3 mm in the longest dimension. The protein crystals were resilient to long-term storage time at room temperature and compatible with a wide range of cryogenic soaking conditions. In all cases, IPPase crystals were physically robust as they resisted breakage during physical manipulation. The quality of the crystals assessed at the home and synchrotron X-ray sources are exemplified in Figure 3.4.

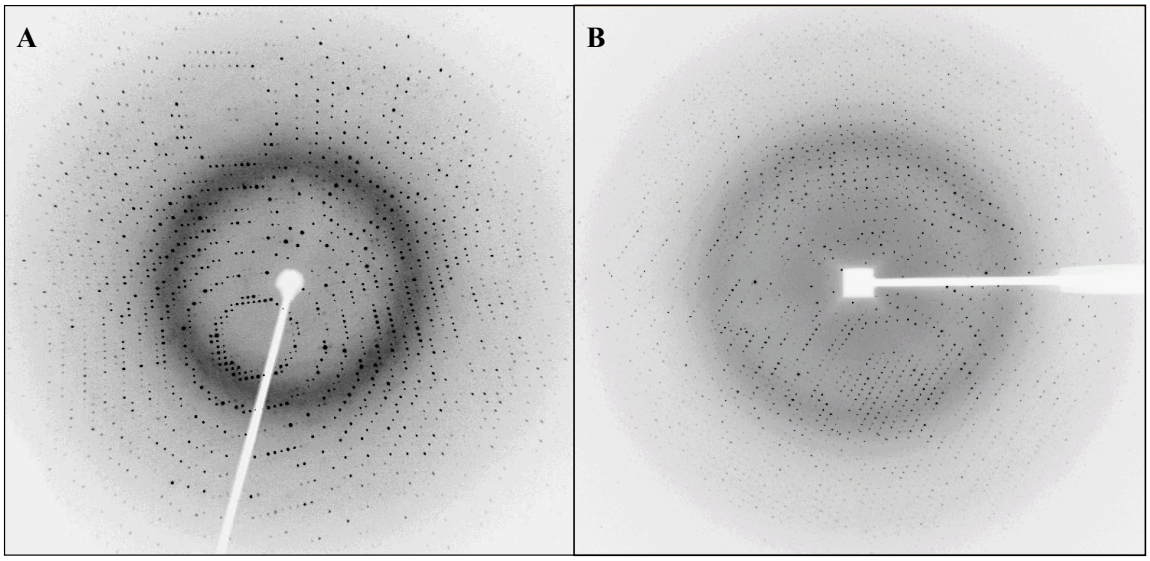


Figure 3.4. Example diffraction images of IPPase crystals. A diffraction image of an optimized IPPase crystal grown in the presence of ammonium sulfate was collected at the home X-ray source (A). The image is a result of a five-minute exposure at 1^{o} oscillation showing diffraction to 2.0Å. resolution. Similarily, an IPPase crystal containing manganese is shown to diffract to 1.9Å using an X-ray synchrotron source (B). In both cases, the images represent the high quality of the IPPase crystals typically obtained based on mosaicity and I/σ measurements.

3.4.2 Crystallization in capillaries

Capillary counter-diffusion crystallization is an effective method to screen and optimize crystal growth conditions in restricted geometry.[66; 76] The crystal growth principles and setup are described in Figure 3.5. Counter-diffusion crystallization in restricted geometry offers an advantage of being able to analyze the crystal grown in the

capillary by X-ray diffraction without any need for physical manipulation.[77] Crystallization setups were prepared using 0.3, 0.5, and 0.7 mm inner diameter (I.D.) glass capillaries. The composition of the precipitant used was derived from initial vapor diffusion crystallization experiments. The final precipitating reagent contained 0.02M calcium chloride dihydrate, 0.1 M sodium acetate trihydrate pH 4.6, and 30% 2-methyl-2,4-pentanediol and was determined to be most amendable to obtain large volume crystals. Using capillary diameters of up to 0.7 mm I.D., crystals could be obtained that fill the entire diameter of the tube (Figure 3.6). However, the crystal size did not exceed 0.2 mm^3 when larger than 0.7 mm diameter capillaries were used. When the IPPase protein solution was exchanged with HEPES buffer under the same conditions, crystal volumes of at least 1 mm^3 can be achieved.

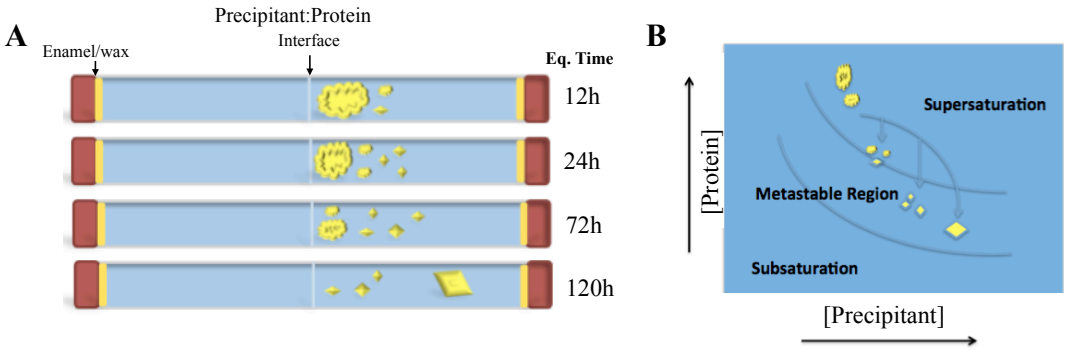


Figure 3.5. Capillary counter-diffusion crystallization. The counter-diffusion configuration is comprised of a precipitating agent diffusing against a concentrated protein volume. As the equilibration process proceeds, a gradient of supersaturation is created along the length of the capillary producing a nucleation front (panel A). The solubility diagram shown in panel B represents the phase transition process of IPPase during counter-diffusion crystallization.

This method was used to grow large volume (>1 mm^3) crystals of IPPase suitable for neutron diffraction studies. The crystallization conditions used to grow crystals filling the inner diameter of 0.3 mm I.D. glass capillaries were used to grow IPPase crystals by counter-diffusion with capillary diameters measuring up to 0.7 mm I.D. The

glass capillaries were made of borosilicate. Since borosilicate has strong neutron absorbance due to the boron, quartz capillaries were used for the larger volume crystal growth. Using quartz capillaries with diameters ranging from 1-2 mm I.D., purified IPPase was carefully aspirated into the capillary followed by the precipitating agent as described in the materials and methods section.

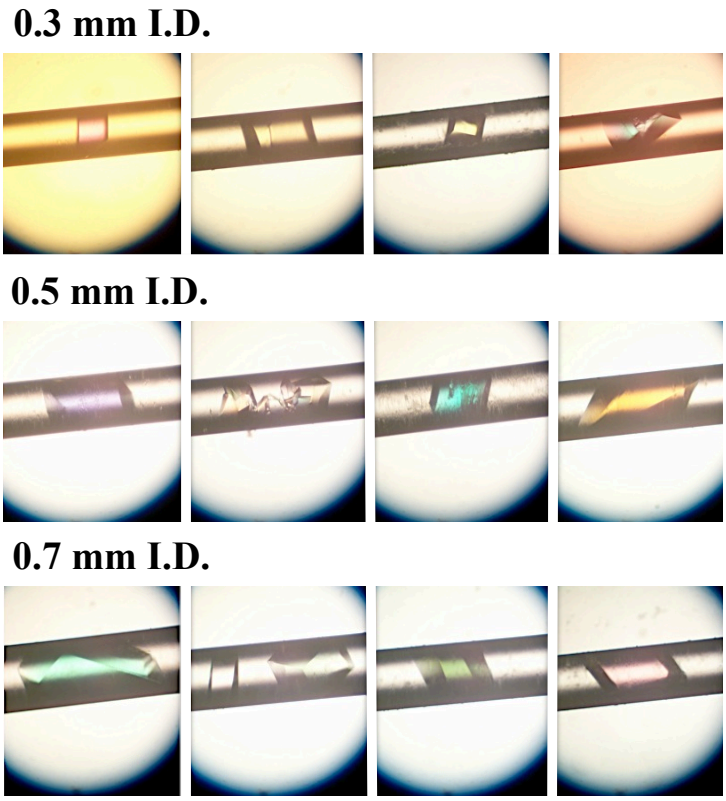


Figure 3.6. IPPase crystals grown by counter-diffusion crystallization. Crystals of IPPase are shown to fill up the inner diameter of a capillary during the counter diffusion crystallization process. Each row shows different crystals of IPPase rountinely observed. Single crystals can be obtained with capillaries having capillary sizes as wide as 0.7 mm I.D.

The counter-diffusion process occurs in a virtually convection free environment where diffusion governs mass transport. When the diameter of the capillary tube exceeds 0.5 mm I.D., diffusion is no longer limiting and convection takes place. However, the diffusion limiting effect can be retained in larger diameter capillaries if the viscosity of

the solution is increased. The increased viscosity associated with the protein solution in the presence of 2-methyl-2-4-pentanediol (MPD) renders a buoyancy and viscous force ratio that can allow diffusion to govern the mass transport process despite a large tube volume. This was evident in 1 mm I.D. (or greater) tubes by the formation of a precipitation front closest to the protein-precipitate interface accompanied by the progression of a high to low number of crystal nucleation along the length of the capillary indicative of a supersaturation gradient.

Single crystals filling the entire diameter of the capillary can be obtained within 10 days. The gradient is unstable as a function of time and slow mixing does occur. However, existing crystals undergo Ostwald ripening[78] where depletion zones around growing crystals are revealed such that the most stable crystals (not necessarily the biggest) grow at the expense of the weaker ones (not necessarily the smallest). Figure 3.7 shows merged photographs taken under a light microscope of a counter diffusion process in a 1 mm I.D. quartz capillary.

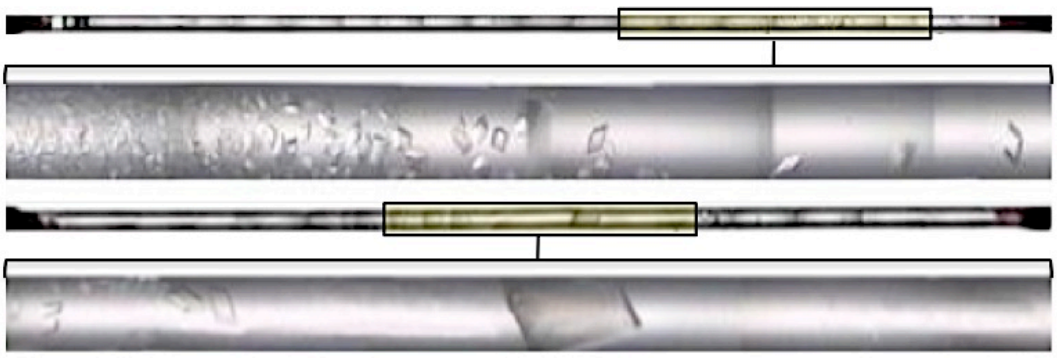


Figure 3.7. IPPase crystals grown in a 1 mm I.D quartz capillary. A prismatic crystal of IPPase is shown to fill up the inner diameter (1 mm) of the capillary. The crystal dimensions were approximately 1.0 x 1.0 x 0.5 mm and it continued to grow at the expense of the adjacent smaller crystal seen in the bottom left. The principal crystal was completely attached to the capillary wall and used directly for *in situ* crystallographic data collection.

In capillaries with diameters exceeding more than 1 mm I.D., an agrose plug was required to prevent the initial rapid mixing of the precipitating and protein solutions.

When agarose plugs are implemented, the progression of high nucleation to single crystal growth is observed as described previously. However, the most stable and ultimately the largest crystals grow out of the agarose plug within 3 weeks (Figure 3.8).

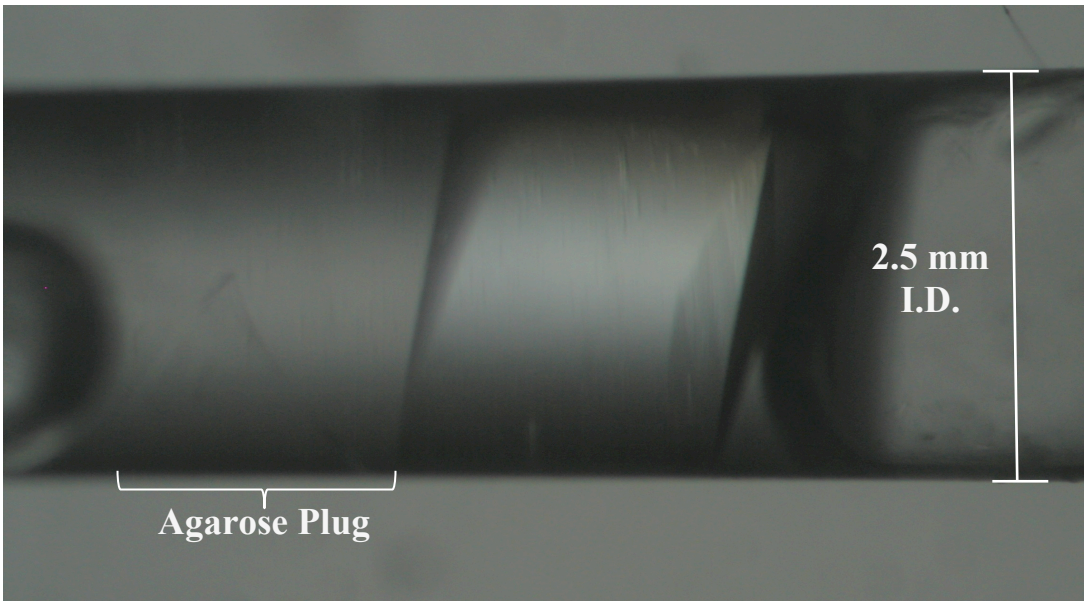


Figure 3.8. IPPase crystals can grow and fill the entire diameter of a 2.5 mm I.D capillary tube. The crystal is shown to be attached to a 1% w/v agarose plug as indicated.

Almost invariably, the largest crystals are attached to the agarose plug. The agarose seemingly serves as an epitaxial surface to nucleate, immobilize, and stabilize the crystal. The agarose matrix also proved to be invaluable for subsequent D_2O exchange as explained in the next section.

3.4.3 Deuterium exchange of capillary-grown crystals

In neutron diffraction significant background noise results from the incoherent scattering properties of hydrogen atoms. In practice, several methods have been employed to replace exchangeable hydrogen atoms with deuterium.[44] The general process utilized most commonly for neutron crystallographic study is vapor diffusion

exchange on crystals mounted in capillaries. Crystals grown in capillaries are generally immobilized against the capillary wall or agarose plug such that the deuterium exchange process can be performed by exchanging solvents without physically manipulating the protein crystal.

A minimum of four exchanges with D_2O with equilibration time of 3 days each was required for adequate deuteration. The efficiency of deuterium exchange was greater than 60% of the labile hydrogen atoms as determined by refined occupancies of exchangeable positions in the neutron crystallographic structure.[79]

3.4.4 Evaluation of deuterium exchanged IPPase crystals

IPPase crystals grown in capillary tubes and deuterium exchanged were subjected to *in situ* X-ray analysis. The crystals were determined to belong to the monoclinic space group C2 and have been observed to tolerate long exposures to synchrotron X-ray radiation performed at room temperature (Figure 3.9).

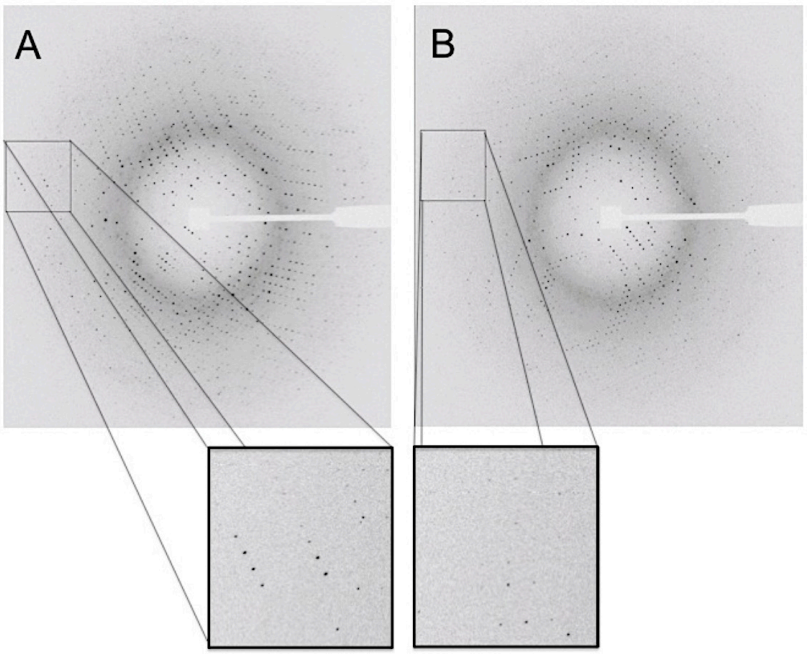


Figure 3.9. X-ray diffraction of deuterium exchanged IPPase crystals at room temperature. IPPase crystals grown in capillaries showed high resilliance against synchrotron radiation damage. One degree oscillations photographs were taken over 150 degrees with an exposure time of one second per frame. Panel shows the initial diffraction image with diffraction spots extending out to 1.8Å. Panel B shows the last image still revealing reflection spots near the maximum resolution.

The complete data collection with synchrotron X-ray radiation was performed on monoclinic crystals of IPPase bound to Ca^{2+} at room temperature. As a result, the X-ray crystallographic structures of Ca^{2+}-bound IPPase was determined at 1.85Å (PDB ID 3I98) and 1.65Å (PDB ID 3Q5V). The purpose of obtaining room temperature X-ray data was to assess the effects of the deuterium exchange manipulations on the diffraction quality of the crystal and to obtain a room-temperature X-ray model that could be used for refinement against the neutron crystallographic data once it had been obtained.

The overall quality of the deuterium-exchanged crystal was excellent with an overall R_{sym} of 7.8%. The highest quality diffraction data obtained from a non-exchanged crystal was only slightly lower.

The deuterium exchanged crystals showed little evidence of disorder and/or radiation damage after a complete data set collection. The observed reflection spots in the diffraction pattern had low mosaic spread over the course of the data collection and were easily indexed and scaled. As a result it was possible to obtain sufficient data for complete structure determination.

3.4.5 Preliminary neutron diffraction analysis

Large volume IPPase crystals were tested for their ability to diffract neutrons at the Protein Crystallography Station (PCS) neutron beamline at the LANSCE facility. Several test exposures were taken from different crystals to assess their diffraction quality. Reflections with I/δ greater than 2.0 were measured in the highest resolution shell (2.1Å). The mosaicity of the reflection spots were very low, indicative of a high quality crystal. The best diffracting crystal was used to capture a complete image from a 120˚ curved (arc length of 1.5 m) helium gas neutron detector with a 12 h exposure. As the result a Laue diffraction pattern showing reflections measuring to 2.1Å was obtained (Figure 3.10)

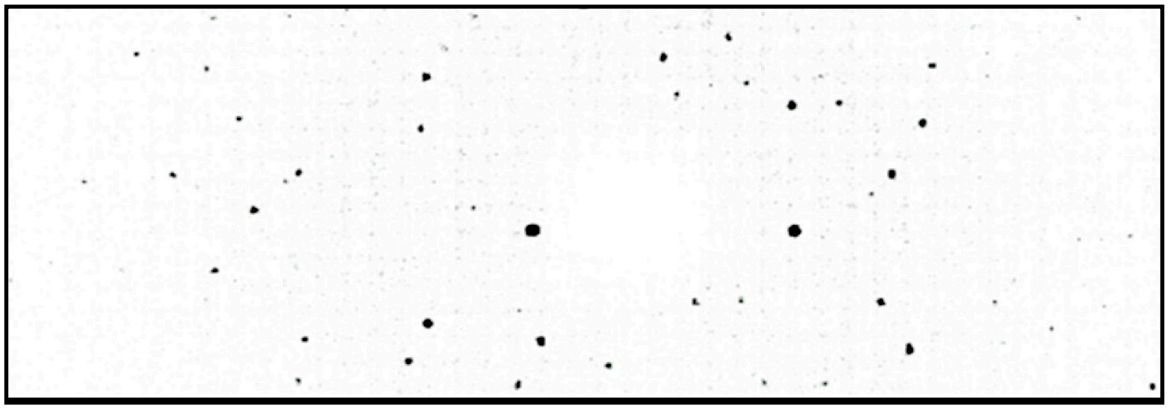


Figure 3.10. Initial neutron Laue diffraction image from an IPPase crystal. A Laue diffraction image recorded 12 hours of exposure. The image was recorded on a 120˚ curved (arc length of 1.5m) helium-free gas neutron detector at the PCS beamline at the LANCE facility in Los Alamos, NM. Reflections were measured to 2.1Å with I/s greater than 2.5

Following the initial diffraction experiments performed at PCS, additional large volume IPPase crystals were then grown and prepared for a full data collection at the LADi III instrument at ILL in Grenoble, France. Figure 3.11 shows a photo taken of a capillary containing an IPPase crystal mounted on the LADi-III instrument. A total of 38 Laue diffraction images were taken at discrete orientations for structure determination. Figure 3.12 shows and diffraction image obtained after a 12 hour exposure of a single IPPase crystal to the neutron beam.

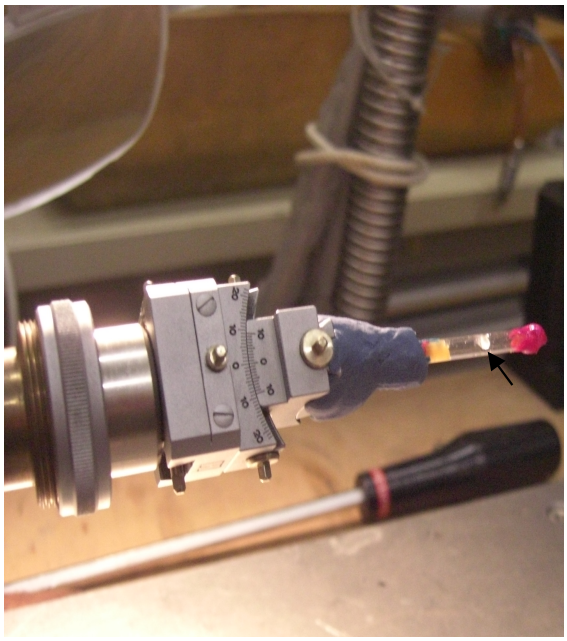


Figure 3.11. Large volume crystals of IPPase prepared for neutron diffraction studies. IPPase crystal grown in a 3 mm I.D. quartz capillary is shown mounted on the LADi-III goniometer head. The volume of the crystal (identified with an arrow) was approximately 8 mm^3

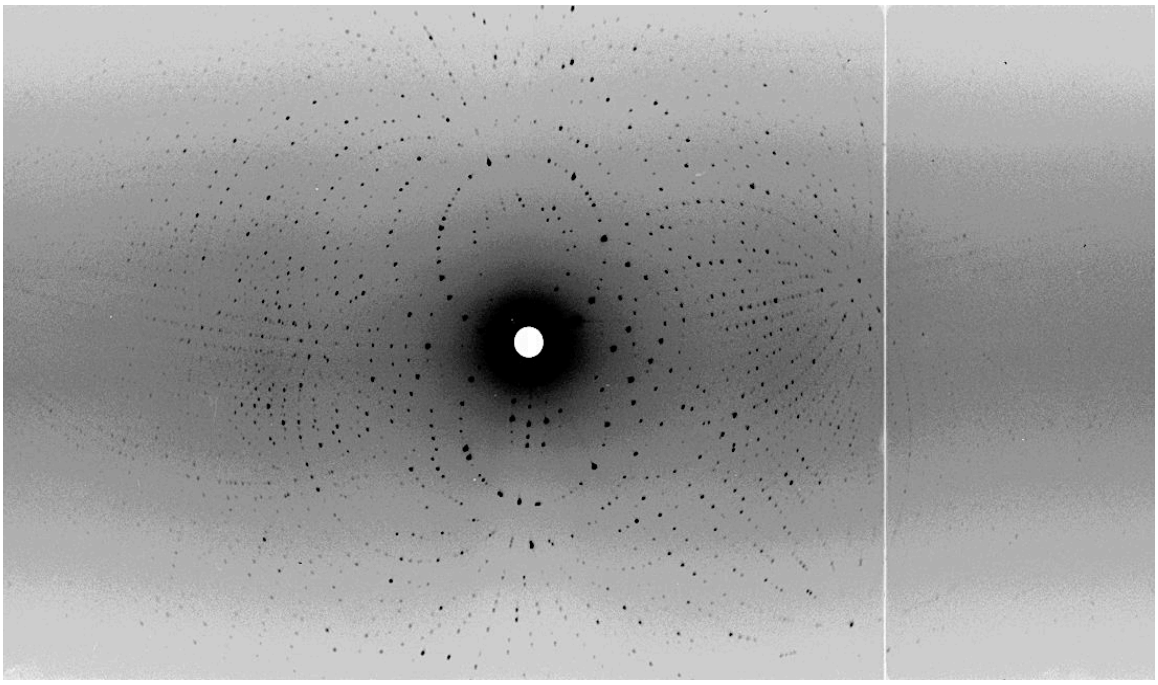


Figure 3.12. Laue diffraction image for IPPase recorded at ILL. An example neutron diffraction image is shown from the data set used to determine the neutron crystallographic structure of IPPase. Reflection spots extends out to 2.1Å with I/δ greater than 2.

3.5 X-ray crystallographic structure determination

Eight high resolution X-ray and one neutron crystallographic structures (PDB ID: 3R6E, 3R5U, 3Q5V, 3Q4W, 3R5V, 3Q9M, 3Q46, 3I98, 3R5V, 3Q3L) were determined. The crystallographic statistics (Tables 3.2-3.5) indicate that the analyses provided acceptable and accurate models.

3.5.1 The structure of IPPase monomer

The monomeric structure has a single-domain OB fold composed of nine ß-strands and two α-helices arranged in a ß1- ß2- ß3- ß4- ß5- ß6- ß7- ß8- α1- ß9- α2 topology (Figure 3.13).

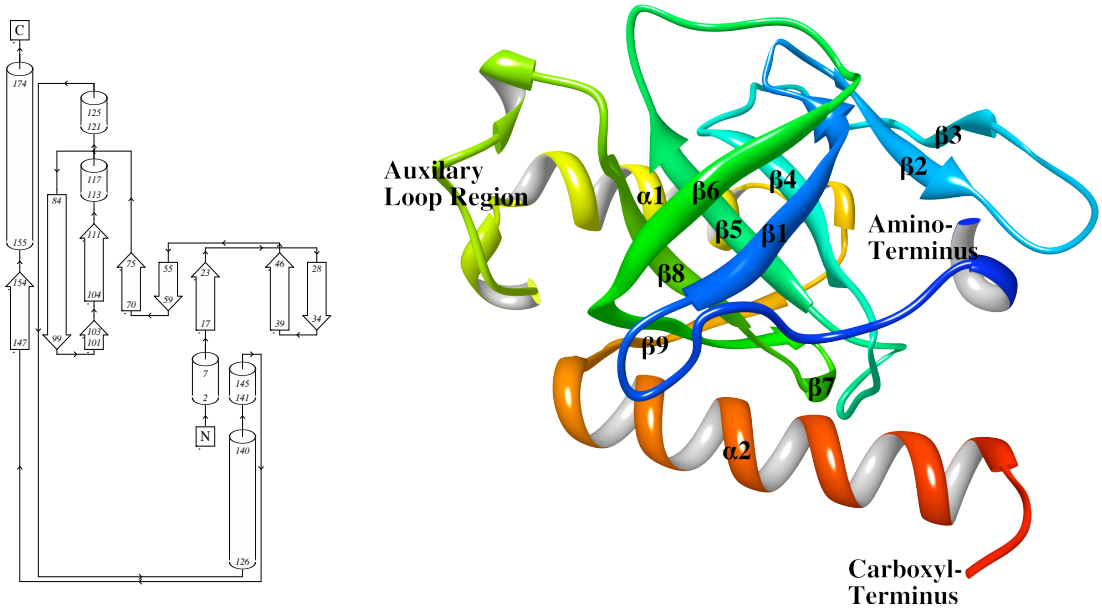


Figure 3.13. The secondary structure topology and quaternary structure of the IPPase monomer. The diagram shown left is a two-dimensional representation of the topology of the IPPase monomer. Alpha-helices are represented by block arrows and ß-sheets are represented by cylinders. Shown right is a ribbon diagram illustration of the IPPase monomer (PDB ID 3I98, Chain-A) labeled to reveal the conserved ß1- ß2- ß3- ß4- ß5- ß6- ß7- ß8- α1- ß9- α2 architecture.

Table 3.2. X-ray data collection and refinement statistics for metal-free, magnesium-coordinated, and manganese-coordinated IPPase crystals

Tt-IPPase Complex	**SO_4^{-2}**	**Mn^{+2}**	**Mg^{+2}, SO_4^{-2}**
PDB ID	**3R6E**	**3R5U**	**3Q5V**
Data collection	**X-ray**	**X-ray**	**X-ray**
Temperature, K	100	100	100
Wavelength, Å	0.97	0.97	0.97
Space group	P1	R32	R32
Unit cell dimensions a, b, c, Å	62.8, 67.3, 74.1	114.5, 114.5.1, 148.4	99.6, 99.6, 99.4
Unit cell dimensions α,β, γ, °	105.9, 106.6, 101.9	90, 90, 120	90, 90, 120
Resolution range, Å	50.00-1.14 (1.18.-1.14)[a]	50.00-1.92 (1.99-1.92)[a]	50-1.08 (1.10-1.08)[a]
No. of observed reflections	1,286,479 (15,401)[a]	160,380 (15,401)[a]	421,279 (17,789)[a]
No. of unique reflections	327,231 (11.531))[a]	28,819 (2,865)[a]	77,505 (6,353)[a]
Redundancy	4.1 (1.2)[a]	5.8 (5.8)[a]	5.4 (2.8)[a]
Completeness	84.6 (29.8)[a]	99.5 (99.9)[a]	88.1% (72.1)[a]
R_{merge} (%)	8.1 (25.7)[a]	8.3 (15.4)[a]	6.0 (25.1)[a]
Wilson plot B	10.4	14.2	9.8
I/σ(I)	13.6 (2.0)[a]	11.2 (6.65)[a]	17.99 (1.99)[a]
Refinement			
Resolution range, Å	27.36-1.18 (1.22-1.18)[a]	29.71-1.92 (1.99-1.92)[a]	32.57-1.08 (1.1-1.08)[a]
No. reflections used	308,458 (14,823)[a]	28,786 (2,818)[a]	76,989 (16,956)[a]
R_{work}, %	12.3	15.0	12.2
R_{free}, %	14.8	19.6	15.5
Solvent % (Mol. Per AU)	45.1 (6)	45.1 (2)	34.9 (1)
No. of atoms (non H or D atoms)	11,586	3,451	-
No. of included H atoms	9,756	-	-
No. of included D atoms	-	-	-
No. of water molecules	1825	419	-
H_2O	-	-	**1,842**
D_2O	-	-	-
DO	-	-	-
O	1864	419	**1,842**
Average B factor, $Å^2$			
Overall	22.9	17.1	16.0
Main-chain atoms	10.4	14.1	13.6
Side chains	25.4	26.7	21.5
Substrate or product	-	-	-
Metal ions	-	4.3	-
Solvent atoms	43.9	34.8	24.2
Rmsd			
Bond length, Å	0.011	0.009	0.010
Bond angles, °	1.421	1.156	1.35
Ramachandran plot (favored/allowed/outliers), %	98.7/1.3/0.0	97.8/2.3/0.0	97.3/1.2/0.5

Table 3.3. X-ray data collection and refinement statistics for substrate and product complexes.

Tt-IPPase Complex	**PP_i, Ca^{+2}**	**PP_i, Ca^{+2}**	**P_i,Mg^{+2}**
PDB ID	**3Q4W**	**3Q9M**	**3Q46**
Data collection		**X-ray**	**X-ray**
Temperature, K	100	100	100
Wavelength, Å	1.07	1.00	0.97
Space group	R32	P3221	R32
Unit cell dimensions a, b, c, Å	100.1, 100.1, 98.7	100.2, 100.2, 97.8	99.3, 99.3, 98.6
Unit cell dimensions α,β, γ, °	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range, Å	50.00-1.44 (1.46-1.44)[a]	50.00-1.34 (1.39-1.34)[a]	50.00-0.95 (1.02-0.95)[a]
No. of observed reflections	216,024 (15,195)[a]	947,788(5,430)[a]	523,484 (10,969)[a]
No. of unique reflections	32,860 (3,233)[a]	116,178 (3,394)[a]	99,827 (4,219)[a]
Redundancy	6.6 (4.7)[a]	8.6((1.6)[a]	6.2 (2.6)[a]
Completeness	95.6 (95.5)[a]	87.9 (50.0)[a]	91.2 (50.7)[a]
R_{merge} (%)	4.7 (36.0)[a]	6.4 (31.4)[a]	6.6 (46.6)[a]
Wilson plot B	37.16 (3.31)[a]	8.8	6.8
I/σ(I)	14.2 (3.6)[a]	13.4 (2.4)[a]	12.9 (2.00)[a]
Refinement			
Resolution range, Å	28.89-1.44 (1.49-1.44)[a]	44.58-1.34 (1.44-1.34)[a]	18.76-0.99 (1.02-0.99)[a]
No. reflections used	32,856(3,080)[a]	110,517 (12,068)[a]	94,537 (5,195)[a]
R_{work}, %	11.9	11.4	11.0
R_{free}, %	14.9	14.6	12.3
Solvent % (Mol. Per AU)	34.1(1)	45.6 (3)	45.17 (1)
No. of atoms (non H or D atoms)	1835	4,091	1,921
No. of included H atoms	-	-	3,546
No. of included D atoms	-	-	-
No. of water molecules	184	627	293
H_2O	-	-	-
D_2O	-	-	-
DO	-	-	-
O	184	627	293
Average B factor, $Å^2$			
Overall	18.6	24.6	14.4
Main-chain atoms	15.9	12.9	11.6
Side chains	23.8	21.3	20.4
Substrate or product	11.1	-	5.5
Metal ions	10.6	7.6	5.9
Solvent atoms	30.3	29.9	21.5
Rmsd			
Bond length, Å	0.01	0.014	0.022
Bond angles, °	1.31	1.67	2.04
Ramachandran plot (favored/allowed/outliers), %	98.4/1.6/0.0	98.6/1.4/0.0	97.6/98.0/0.2

Table 3.4. X-ray and neutron data collection and refinement statistics for calcium-coordinated IPPase

Tt-IPPase Complex	**Ca^{2+}**	**Ca^{2+}**	**Ca^{2+}**
PDB ID	**3I98**	**3R5V**	**3Q3L**
Data collection	**X-ray**	**X-ray**	**Neutron**
Temperature, K	273	273	273
Wavelength, Å	0.92	1.00	4.2-3.1
Space group	C2	C2	C2
Unit cell dimensions a, b, c, Å	106.1, 95.5, 113.7	106.1, 95.5, 113.7	106.1, 95.5, 113.7
Unit cell dimensions α,β, γ, °	90, 98.1, 90	90, 98.1, 90	90.0, 98.1, 90
Resolution range, Å	50.00-1.85 (1.92-1.85)[a]	50.00-1.65 (1.71-1.65)[a]	37.53-2.5 (2.64-2.5)[a]
No. of observed reflections	296,646 (26,292)[a]	223,275 (19,269)[a]	97,346 (7850)[a]
No. of unique reflections	95,841 (9,390)[a]	127,642 (13,790)[a]	28,587 (3343)[a]
Redundancy	3.1 (2.8)[a]	1.7 (1.7)[a]	3.4 (2.3)[a]
Completeness	99.0 (98.4)[a]	90.1 (77.4)[a]	73.7 (58.7)[a]
R_{merge} (%)	7.8 (39.6)[a]	5.5 (39.0)[a]	14.6 (17.7)[a]
Wilson plot B	13.5 (2.6)[a]	16.4	14.2
I/σ(I)	15.2 (7.8)[a]	13.5 (2.0)[a]	7.4 (5.4)[a]
Refinement			
Resolution range, Å	46.23-1.85(1.92-1.85)[a]	46.23-1.65(1.71-1.65)[a]	36.42-2.50 (2.59-2.50)[a]
No. reflections used	90,079 (9,122)[a]	121,472 (12,165)[a]	28,530 (3,324)[a]
R_{work}, %	13.9	12.3	24.1
R_{free}, %	18.7	16.3	27.8
Solvent % (Mol. Per AU)	39.4 (6)	39.4 (6)	39.8 (6)
No. of atoms (non H or D atoms)	6,950	9,258	18,884
No. of included H atoms	-	-	8,364
No. of included D atoms	-	-	1,767
No. of water molecules	654	631	131
H_2O	-	-	-
D_2O	-	-	45
DO	-	-	-
O	654	631	86
Average B factor, $Å^2$			
Overall	23.2	25.8	14.4
Main-chain atoms	20.1	20.1	23.6
Side chains	34.4	34.4	24.8
Substrate or product	-	-	-
Metal ions	7.6	7.6	30.1
Solvent atoms	50.4	46.4	16.6
Rmsd			
Bond length, Å	0.023	0.008	0.007
Bond angles, °	1.87	0.960	1.07
Ramachandran plot (favored/allowed/outliers), %	98.35/1.5/0.2	98.5/1.5/0.0	98.5/1.5/0.0

The superposition of the minor fold-determining, ß-harpin- and ß-turn-, motifs found in IPPase are structurally conserved. Specifically, sixteen ß-turn motifs, three ß-harpin motifs, and one ß-turn motif are found in the monomeric structure of IPPases (Figure 3.14). While the position of the ß-hairpin motifs are structurally conserved with *E. coli* IPPase, sequence indentity of the participating contact residues of the adjacent anti-parallel ß-strands in IPPase are only conserved among closely related hyperthermophilic archaea (i.e., IPPase, *P. furiosus,* and *P. horikoshii*). The loops connecting the forth and fifth ß-strands (ß4-ß5), ß8 and the second α-helix (α2), and α3 to ß9 are truncated in *P. furiosus* and *P. horikoshii* IPPase relative to the length of the equivalent loops in the *E. coli* enzyme.

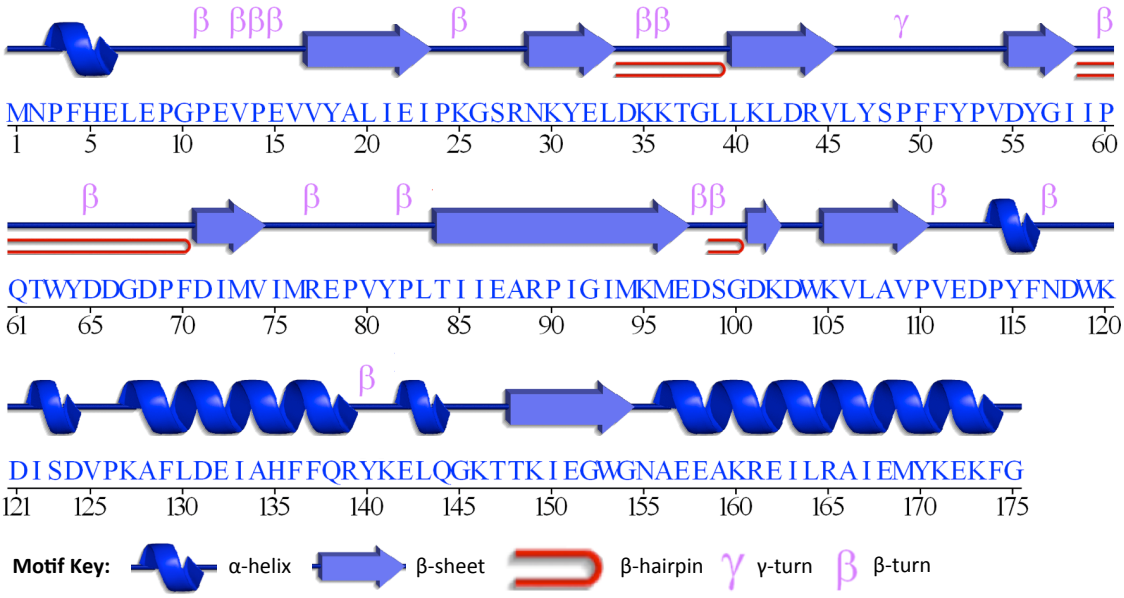


Figure 3.14. The major secondary structure elements and minor structural motifs of IPPase. The sequence of amino acid of IPPase are shown to coorespond with its major and minor secondary structure elements comprising the monomeric structure of IPPase (PDB ID: 3Q9M, Chain-A). The major α-helical and ß-strands elements and the minor ß-hairpins, ß-turns, and γ-turn motifs are represented according to the provided key.

The shorter connecting loops lead to a significant reduction in the total number of ß-turn motifs in the structure of IPPase (16 vs. 21) relative to *E. coli* IPPase. The loop connecting ß1 and ß2 is also truncated relative to the equivalent loop in *E. coli* enzyme. As a result, the ß-turn motif observed prior to ß2 in *E. coli* IPPase is not observed in IPPase.

Despite the low sequence homology (<30%) and different monomeric molecular weights, IPPase shares a high degree of architechtural similarity with the eukaryotic IPPase derived from *S. cereversiae*. The ribbon diagram of the IPPase monomer has been overlayed with the *S. cerevisiae and E. coli* enzyme monomers displayed in Figure 3.15.

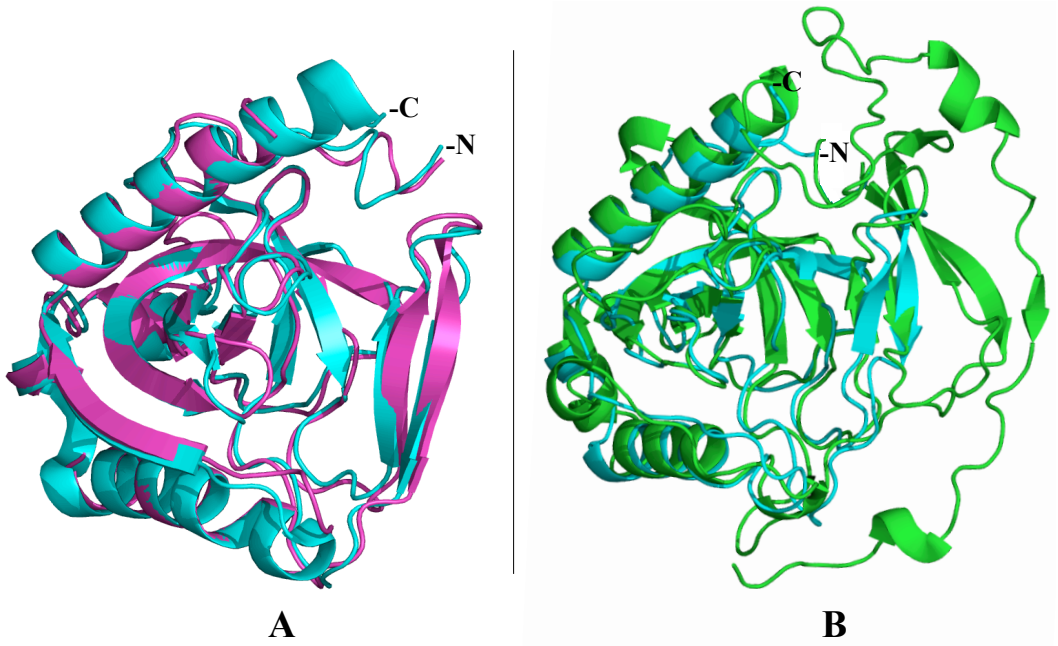


Figure 3.15. Structural alignment of the IPPase monomer with those from prokaryotic and eukaryotic sources. The ribbon diagrams show the structural superposition of *T. thioreducens* (cyan) against *E. coli* (magenta)(A) and *S. cerevisiae* (green)(B). IPPase shows high topological similarities when compared to *E. coli* and *S. cerevisiae* IPPase with a calculated average RMSD value of 0.79Å^2 and 1.66Å^2 respectively for the main chain atoms.

The average polypeptide backbone root-mean-square deviation (RMSD) among the crystallographic models of the IPPase monomer (residues 4 to 174 of the main referenced chain) is 0.38Å. The three-dimensional structure of the IPPase monomeric subunit displayed a high degree of similarity with the family I IPPases from *E. coli* and *S. cerevisiae* as illustrated in Figure 3.12. The average RMSD value for the main chain atoms for the structures shown was 0.79Å^2 and 1.66 Å^2 when compared against *E. coli* and *S. cerevisiae* IPPase respectively.

3.5.2 IPPase oligomerization

The PISA (Protein Interfaces, Surfaces and Assemblies) server was used to predict the oligmeric assembly for IPPase by examining structurally similar interfaces and using the parameters associated with these interfaces to predict the most probable biological assembly. As a result of these calculations, the predicted assembly for the structures of the calcium coordinated IPPase solved at room temperature (3R5V, 3I98) cooresponded with the tightly packed hexameric assymertric unit observed in the crystal structure of the IPPase apo-enzyme (3R6E).

IPPase is composed of six identical 21 kDa monomeric subunits arranged as a symmetrical "dimer of trimers" with each monomer having a self-contained active center (Figure 3.16). A 60˚ rotation about the three-fold axis allows each of the monomeric subunits in neighboring trimers to coordinate with four of the remaining five monomers in the hexamer.

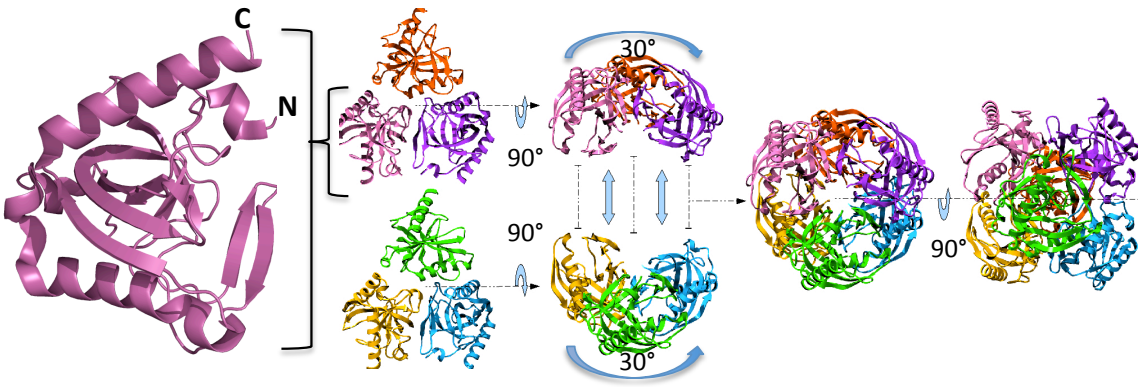


Figure 3.16. IPPase oligomerization. Two trimers form a dimer related by two-fold symmetry making up a homohexamer. There is a 30˚ rotation of each trimer relative to the other upon dimerization, allowing each monomer to contact four other monomers such that the calculate bonding energies associated with interfacial contacts are maximized.

Three small pores, formed by the active sites of the trimers are arranged symmetrically about a large central pore that vertically transverses the complex. A large solvent channel is observed to transverse horizontally along the axis orthogonal to the trimer-trimer interface. The IPPase hexamer is stabilized by a relatively small number of residues that vary in both identity and degree of conservation among prokaryotic IPPases.

The IPPase trimer is primarily stabilized by hydrophobic contacts between the side chains of two valine (45,80), three isoleucine (83,85,86), and two leucine (39,40) amino acid residues. Other contacts observed at the trimer interface in IPPase include, Pro82, Gly38, Tyr81, Thr84, and Met76. The stabilizing contacts at the dimer interface are mostly formed by ionic interactions between the side chains of Glu132, Arg139, Ser48, Phe50, His135, Glu142, and Arg77.

3.5.3 IPPase active site coordination

The X-ray crystallographic structures have provided atomic level snapshots of the IPPase active center at resolutions as high as 0.99Å showing the discrete conformations of important water molecules and protein groups coordinating metal ions, substrate,

product, and sulfate product analogue (Figure 3.17). Four divalent metal binding sites have been identified in the active site of IPPase and are referenced by number from highest to lowest affinity (M1, M2, M3, M4) based on conventional measurements determined by crystallographic and biochemical calculations.[80] In the IPPase structures (PDB ID: 3R5V, 3R5U, and 3I98) obtained in the absence of substrate, product, and product analogue only the M1 site is fully occupied by a metal ion.

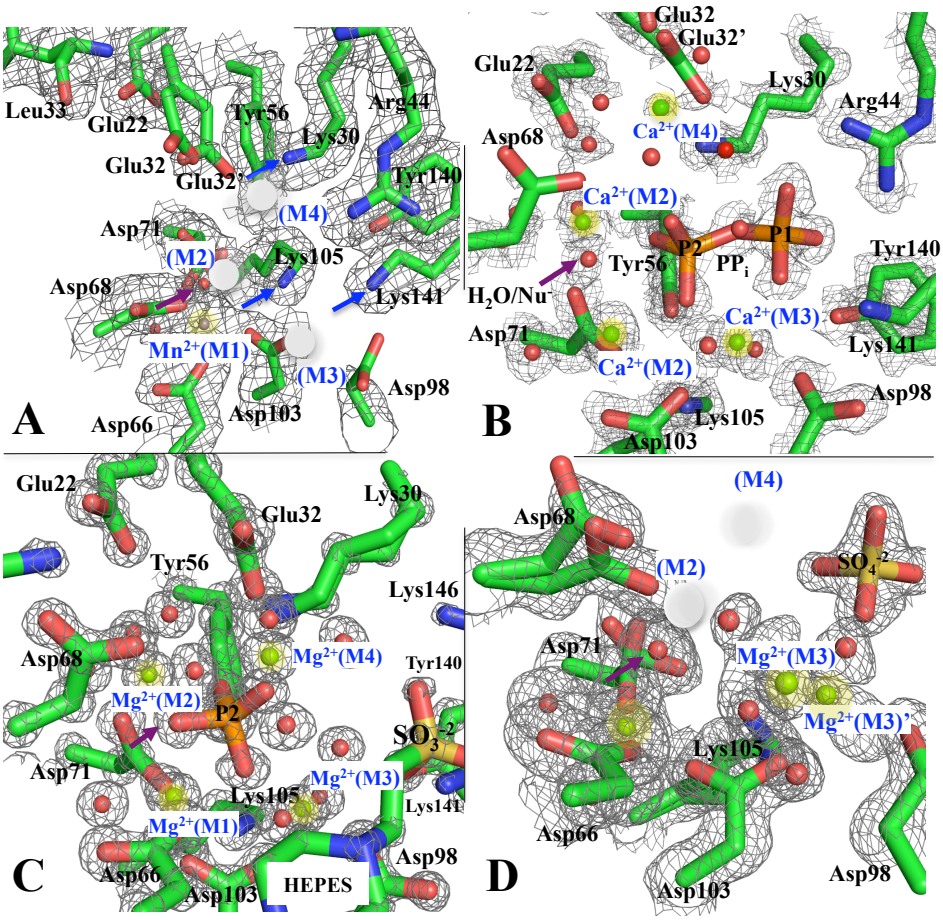


Figure 3.17. Atomic resolution snapshots of IPPase active centers determined by X-ray crystallography. The active site coordination of IPPase determined from crystals obtained in the presence of manganese (A), calcium and pyrophosphate (B), magnesium and phosphate (C), and magnesium and sulfate product analogue (D) are shown. The atomic positions of carbon (green), oxygen (red), nitrogen (blue), water molecules (red sphere) are shown modeled into a mesh of electron density determined by crystallographic calculation. The metal ions have been identified according to the conventional numbering system (M1-M4)[80].

The remaining metal binding sites (M2, M3, M4) are not coordinated to a metal ion. In IPPase structure determined with manganese and calcium (PDB ID: 3R5U, 3R5V, 3I98) the M2 and M3 coordinating aspartate residues (71,98) are interacting with ε-amino group of two lysines residues (105, 146). For example, Lys105 participates in an ionic bond with Asp98 at the M3 site. The active site of the 1.65Å calcium-coordinated IPPase crystallographic structure (PDB ID: 3R5V) is shown in Figure 3.18 and provides a high resolution snapshot of the metal coordinated active site of IPPase in the absence of substrate. The M1 metal is coordinated by Asp66, Asp71, and Asp103 and exhibits reasonable temperature factors refinement calculations when the metal ion is refined with an occupancy value of 1.0.

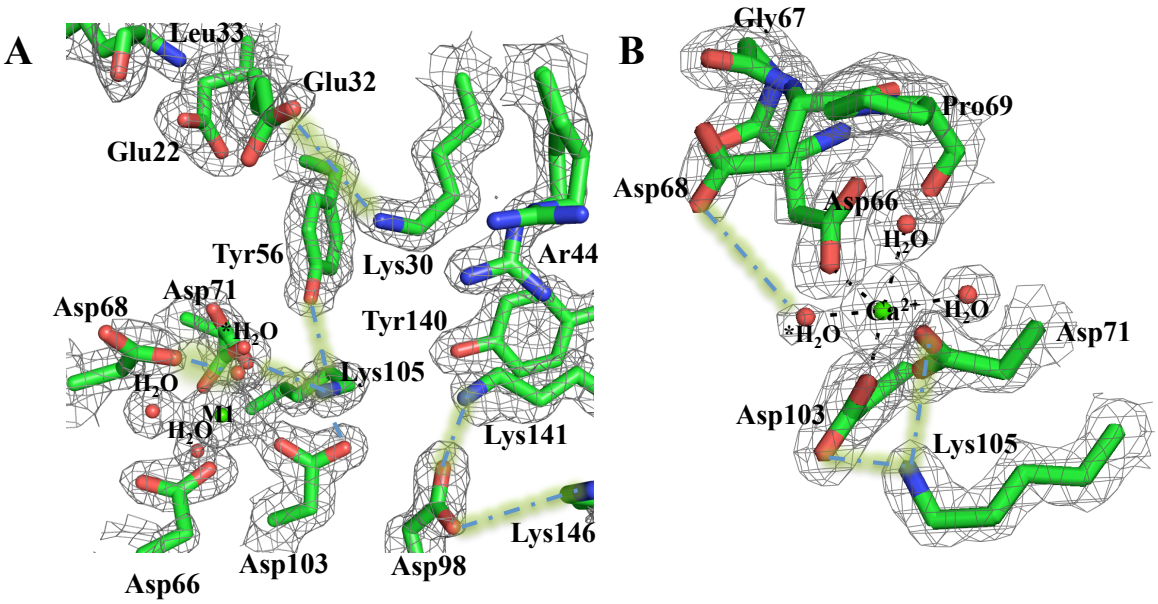


Figure 3.18. M1 coordination in IPPase. The M1 site coordination is shown with a calicum Ca^{2+} ion bound. Three aspartate (66,71,103) residues can be seen to coordinate the M1 metal ion. A. In the absence of substrate or product only the M1 site is filled. However, the active site structure is maintained in absence of substrate or additional metals by ionic interactions between the M2, M3, and M4 coordinating residues and active site lysine residues (shown as yellow highlighted dashed lines). Lys30 is paired with Glu32. Lys141 and Lys146 pair with Asp98. Lys105 coordinates with Asp103, and Asp71. Lys105 can also be seen in a bonding interaction with Tyr56. B. An alternate view is shown of the M1 site with black dashed lines drawn between the M1 Ca^{2+} ion and the waters that fill its coordination shell. The coordination shell of M1 is an imperfect octahedron. The skewing of the octahedron is evident by the angle between the water molecule (Red sphere,H_2O) that is between Pro69 and the Ca^{2+} ion bound at the M1 site.

In the Ca^{2+} inhibited substrate complexes of IPPase (3Q9M, 3Q5W) the substrate is coordinated by three fully occupied metal ions (M1, M2, and M3). A fourth metal M4 can be seen to be partially occupied (~30-50%) in the M4 site. The occupancy and the coordination of the metal appear to coincide with the position and occupancy of Glu32. Active site residues Lys141, Arg44, Lys30, Tyr56, Tyr140 coordinate to the remaining substrate phosphate oxygen atoms. Arg44, Lys141, and Tyr140 coordinate with the outermost (P1) substrate group. Tyr56 coordinates to the inner most (P2) group. The phosphoester oxygen is coordinated to a water molecule and Lys30. In the presence of substrate a single water molecule is shared between the metal coordination sphere of M1 and M2. The analogous water molecule was referenced as the "bridging water" that was proposed from previous studies on bacterial and eukaryotic IPPase to be active towards nucleophilic attack on the P2 phosphate. The proposed proton acceptor in the corresponding studies, Asp67 in *E. coli* and Asp117 in *S. cerevisiae* are structurally conserved with Asp68 in *T. thioreducens* IPPase. The hydrogen bonding distance is observed to be between 3.7-3.2Å between Asp68 and the "bridging water" in the structures determined in the presence of substrate. When comparing the substrate free active site of IPPase the relative position of the water molecule is conserved in absence of M2.

Two confirmations of the substrate are observed and modeled in structural complexes of IPPase (Figure 3.19). In the predominant confirmation the substrate is not optimally positioned for a backside nucleophilic attack by the bridging water molecule. The changes in substrate confirmation coincide with discrete changes in the active site solvent structure as a function changes in the M4 metal and its coordination to Glu32.

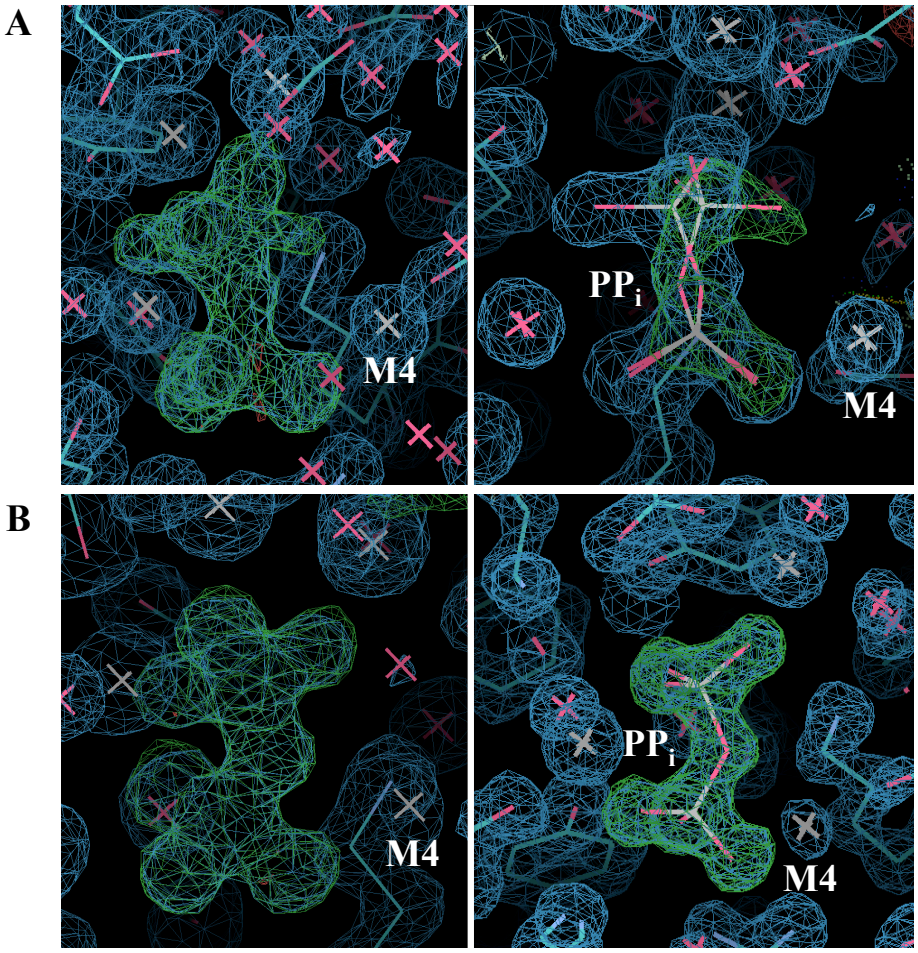


Figure 3.19. Modeling of alternate conformations of PP_i substrate. Shown are $2F_o$-F_c (shown at 1.5δ, blue mesh) and F_o-F_c (shown at 6.0δ, green mesh) electron density maps covering the active sites of chain A (A, upper panels) and (B, lower panels) of the IPPase substrate complex PDB ID: 3Q9M. In two of the three monomers that comprise the crystallographic asymmetric unit, two conformations of the substrate are apparent and can be modeled at ~70/30% occupancies. In chain B only one conformation of the substrate was modeled. The alternate conformation coincides with M4 coordination to the P2 phosphate group of the substrate.

Among the observations that were seen include the shortening of the hydrogen bonds in the water chain that leads from the nucleophile towards Glu22. In the highest occupancy (~75%) (i.e., when the substrate is not in the optimal orientation for a backside attack) two water molecules form a connection between W1 and Glu22 (Figure 3.20). The positions of the water molecules and their coordination is very similar to the

water chain that was proposed in *E. coli* studies to explain the retained activity when the proposed proton acceptor Asp68 was mutated to asparagine.[1]

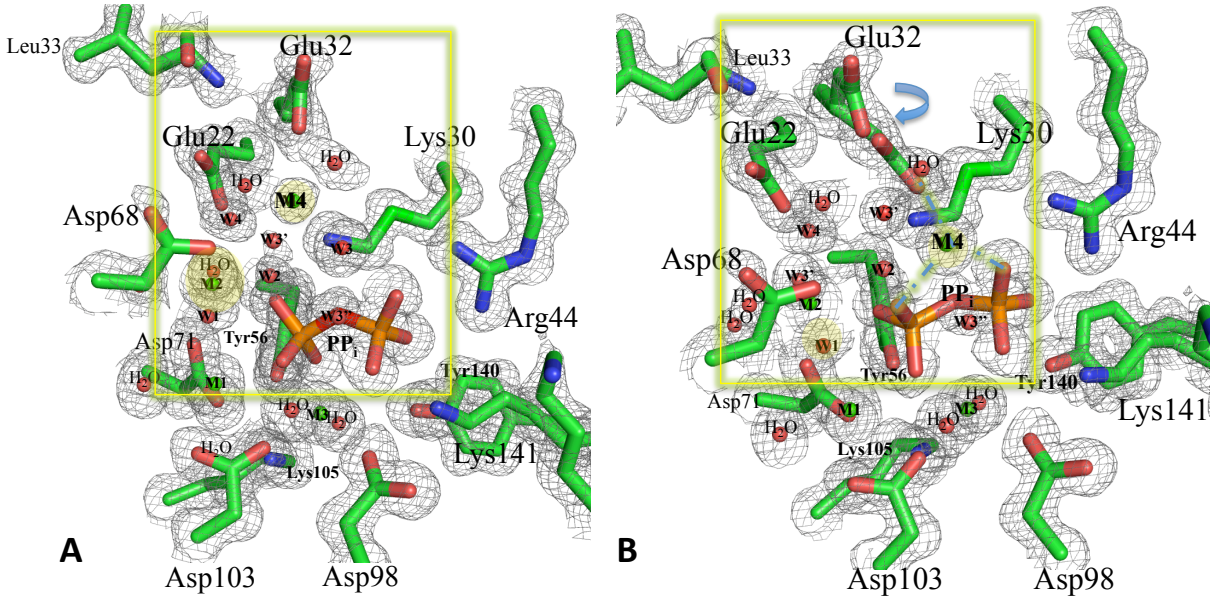


Figure 3.20. Substrate coordination in IPPase. Two confirmations of the substrate are observed in structural complexes of IPPase. In the predominant confirmation of the substrate is not optimally positioned for a backside nucleophilic attack by the bridging water molecule (A). In the IPPase complex at 1.3Å the two substrate confirmations appear to coincide with the occupancy of the bridging water. In the lower occupancy conformation of the substrate (~30% occupied) the substrate can be seen in the correct orientation for a backside attack on the innermost phosphate group of the substrate with two water molecules forming a connection between W1 and Glu22 (B).

3.5.4 Neutron crystallographic structure

A complete neutron crystallographic structure of the non-substrate bound form of IPPase complexed with Ca^{2+} was determined to 2.5Å resolution (PDB ID 3Q3L; Table 3.4). The counter-diffusion crystallization method in restricted geometry provided suitable crystals for a neutron structure determination revealing valuable images within the catalytic site not seen by X-ray crystallographic data. As a result the protonation of the critical active-site residue were determined in the active center (Figure 3.20).

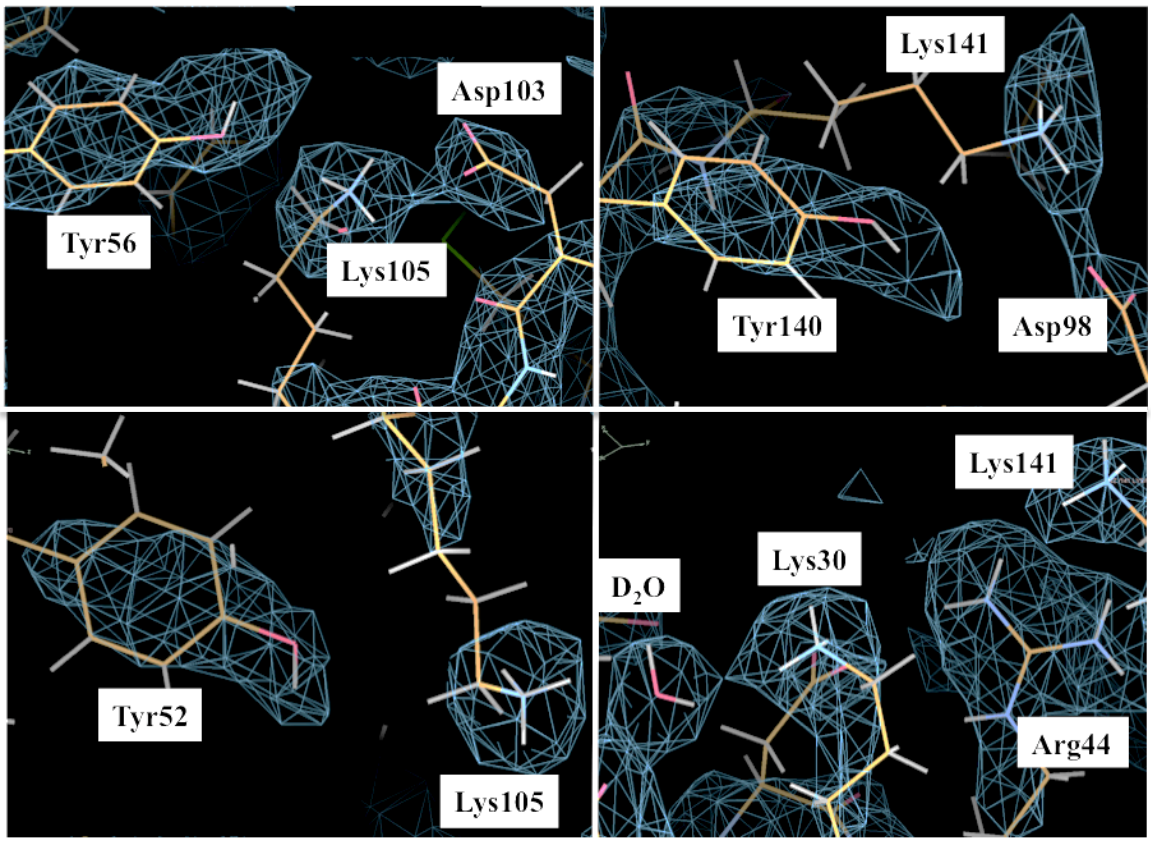


Figure 3.21. The positions of protons in the calcium-coordinated active site of IPPase. The neutron density can be seen to cover the deuterium of exchangeable nitrogen and oxygen atoms within the IPPase active center. Carbon-bound hydrogen atoms do not exchange and result in density cancelation as seen in the figure.

As mentioned in the previous section, the structural location of Asp68 in *T. thioreducens* IPPase is equivalent to Asp67 in the *E. coli* and Asp117 in *S. cerevisiae* active sites. This Asp residue was thought to play a major role in the deprotonation of a water molecule that bridges two metals (M1 and M2) for the nucleophilic attack of the inner most phosphate (P2) of the PP_i substrate.

In effort to explain the residual activity present in the IPPase mutant lead to the consideration of the mechanism of the activation of the water bridging M1 and M2. The neutron structure reveals precisely positioned water molecules through tight hydrogen bond contacts that are conserved in the crystallographic structures of IPPase determined

in this study and will be further discussed in the next chapter. Figure 3.21 shows a side-by-side comparison of a chain of hydrogen bonded water molecules in the neutron and X-ray crystallographic structures of IPPase (PDB ID: 3Q3L , 3R5V). The neutron structure of IPPase has provided the means to unambiguously determine protonation states of the IPPase residues and the orientation of active site water molecules.

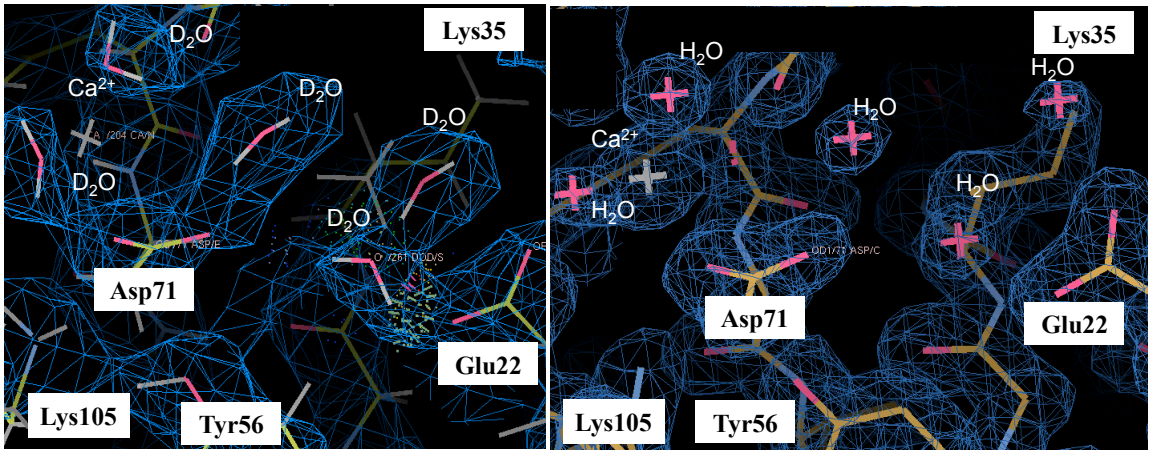


Figure 3.22. Active site water coordination in IPPase. Images are shown of the neutron density (left) and electron density (right) covering the highest affinity metal binding site and the surrounding solvent and protein groups is shown. The blue mesh represents the respective $2F_o$-F_c neutron density (shown at 2δ) and electron density map (shown at 2δ) covering active site residues of calcium coordinated IPPase (PDB ID: 3R5V, 3Q3L). The neutron structure enabled the determination of the orientation of the coordinating water molecules in the active site. In the figure water molecules have been labeled as D_2O in the neutron crystal structure screenshot (shown left) and as H_2O in the X-ray crystal structure screenshot (shown right). The orientation of water molecules identified D_2O can be deciphered even at low resolution of 2.5Å by neutron crystallography. These determinations are not possible with X-ray diffraction even at much higher resolutions.

CHAPTER 4

DISCUSSIONS

4.1 Oligomeric structure and assembly

The overall structural features of IPPase reveal valuable insight into the assembly, symmetry, and mechanism of the enzyme. The predominance of the hexameric form of *T. thioreducens* IPPase in solution was initially predicted based on size-exclusion chromatographic analysis observed during the initial protein purification procedures. The residues essential for stabilizing the hexameric form were identified (Table 4.1; right column) and structurally conserved with those of *E. coli* IPPase. Therefore, it is most likely the predominant of active form of *T. thioreducens* IPPase is hexameric as reported for IPPase in *E. coli.*[10] Oligomerization has been suggested to be important for efficient catalysis, potentially by establishing channels for communication between monomeric active sites of the hexamer.[10] The oligomeric contacts found in the asymmetric units of *T. thioreducens* IPPase X-ray crystallographic structures are shown Appendix B. These oligomeric contacts are similar to those found in other hyperthermophilic archaea but different from those observed in *E. coli* or *S. cerevisiae*. Thereby the oligomeric contacts may have a possible role in thermostability. In the IPPase trimer the N-terminal residues are located in the central pore near the three-fold center and are stabilized by contacts from neighboring monomers. In contrary to monomeric proteins with disordered N-terminal structures, clear defined electron densities can be readily observed in the IPPase crystallographic difference maps. Other

residues that contribute to the trimer contact stability include lys35 and Lys36 near the mobile active site loops.

Table 4.1. Functionally conserved residues in IPPase

Active site		Oligomeric contacts	
Function	**Conserved Residue**	**Function**	**Conserved Residue**
Substrate binding	Arg44	*Dimer contact*	Ser48
	Lys30		Phe50
	Tyr56		Arg77
	Tyr140		Glu132
	Lys141		His135
Metal binding	Glu32	*Trimer contact*	Leu39
	Asp66		Leu40
	Asp71		Phe50
	Asp98		Val45
	Asp103		Val80
Other	Glu22		Pro82
	Tyr52		Ile83
	Thr62		Thr84
	Lys105		Ile85
	Phe136		Ile86
	Phe137		Arg139
	Ala158		Glu142

Initial observations suggest that the IPPase trimer may be necessary for inter-oligomeric communication and thermostability Biochemical investigations conducted on *E. coli* IPPase concluded that the hexamer acts in a coordinated manner to cleave the substrate.[10] This finding was also consistent with the observation that sulfate binding in the active center can result in formation of an asymmetric hexamer in which the three-fold symmetry about the trimer is non-crystallographic. As a result, it was proposed that a transient asymmetric hexamer forms as a result of product formation and is subsequently restored upon substrate binding.[10] Mutational studies on the *E. coli.* enzyme also showed that the hexamer and trimer units are active in *E. coli*, with the hexamer being favored at neutral pH and the trimer being favored in mildly acidic

conditions. However, similar investigations have not been performed on IPPases from other sources.

A close examination of *T. thioreducens* IPPase structures bound to substrate, product, and reaction analogues, provide clues into the molecular location of the communication points in the IPPase oligomer. The average RMSD values obtained for the main chain atoms of the IPPase structural complexes reveal that asymmetry within the trimer appears to be related to the bound state of the enzyme and concentration of the ligand used for the crystallization experiment. Most notably, the average RMSD values among intra- molecular asymmetric units are about the same as those compared among asymmetric units in different structural complexes. For example, the average RMSD values for main chain atoms among the six asymmetric unit monomers of the IPPase structure crystallized in the presence of ammonium sulfate (PDB ID: 3R6E) (RMSD = 0.39Å^2) is nearly equivalent to the average RMSD measured between the A-chains of the different IPPase complexes (RMSD = 0.38Å^2) reported in this dissertation.

Similarly, a comparison of the RMSD of side-chains of the different IPPase complexes can be performed to provide a global view of structural differences observed for important residues as a function of the ligand-bound state of the enzyme. Upon close examination it becomes apparent that the global changes occurring in the hexamer are a result of structural changes that occur in a subset of residues that directly interact with the bound ligand. Changes in coordination to the ligand lead to changes occurring in dimer and trimer interfacial contact residues.

Structural differences between monomers within the hexamer also become evident when examining the degree of deuterium exchange for backbone amide groups in

the neutron crystallographic structure of IPPase (PDB ID: 3I98). Figure 4.1 illustrates how the exchange rate of backbone amide protons varies with the average atomic displacement parameter for residues in the neutron crystallographic trimer. The N-terminal residues including His5, Asn2, and the Lys35-containing loop near the IPPase active center have a high temperature factor and a high exchange rate. As expected, the residues within the hydrophobic pocket adjacent to the active site have high deuterium occupancies for the backbone amide nitrogen and overall temperature factor relative to other regions of the molecule. The metal-coordinating and key active site residues have low temperature factors where as other active site residues are well-exchanged in the metal preordered active site.

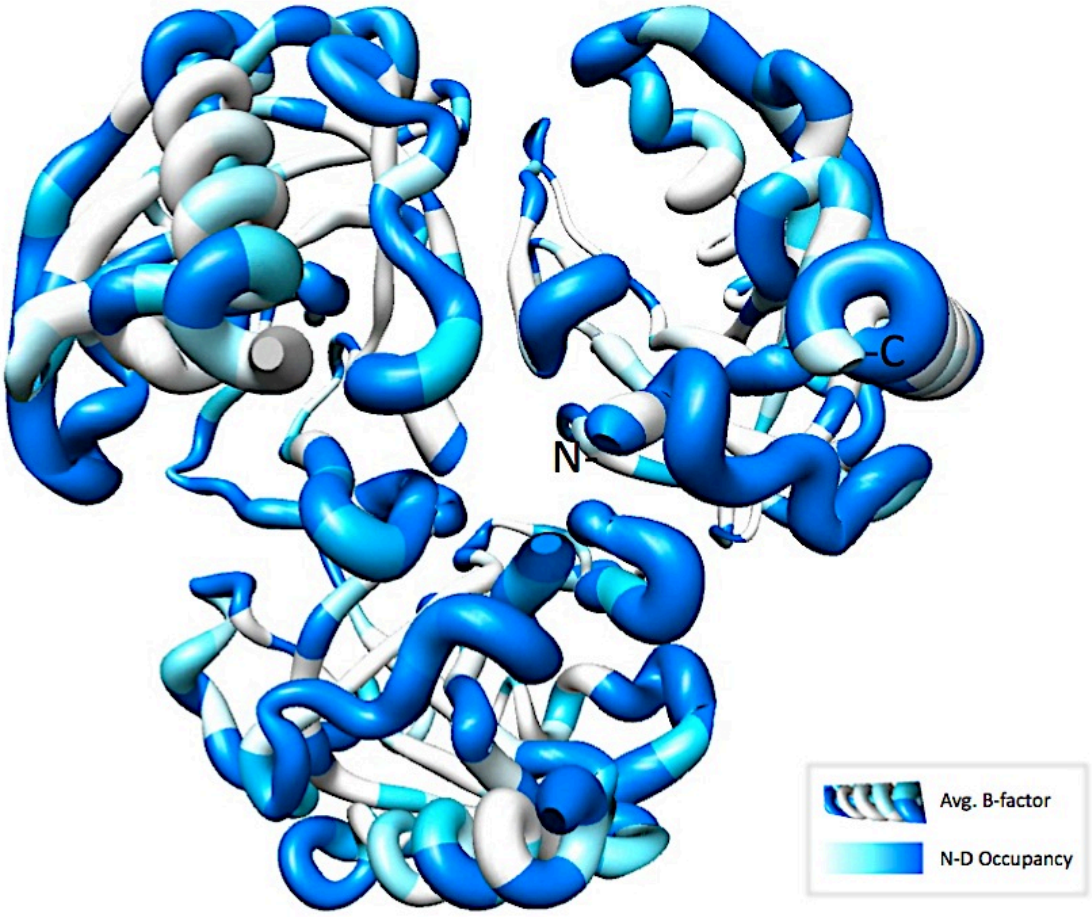


Figure 4.1. Temperature factor versus backbone amide deuterium occupancy for the crystallographic trimer determined by neutron crystallography. Illustrates how the exchange rate on the amide backbone nitrogen varies with the average atomic displacement parameter for residues in the neutron crystallographic trimer. The average temperature factor (Avg. B-factor) is displayed as increasing with the diameter "worm" representation of enzyme's backbone atoms. The amide exchange percentage is represented by coloring with white representing resides that have less than 20% exchange of the amide deuteron in the neutron crystal structure of IPPase. Residues colored in light blue and dark blue represent Residues with amide nitrogen deuterium occupancies of between 20-60% and 61-100% respectively. Equivalent residues in neighboring monomers exhibit different degrees of amide exchange and serve to highlight potential intrinsic differences within the monomers of a trimer. The data discussed here was obtained by using the *Phenix.Refine* program to simultaneously refine the occupancies of hydrogen and deuterium at each exchangeable position within the protein structure of IPPase.

4.2 Ligand-binding effects on symmetry

The structures of *T. thioreducens* IPPase, complexed with Mn^{2+} (PDB ID: 3R5U) Mg^{2+} and phosphate (PDB ID: 3Q46) were determined in the rhombohedral space group R32. In contrast, the structures bound with Ca^{2+} (PDB ID: 3Q5V) and sulfate (PDB ID: 3R6E) were obtained in the lower symmetry space groups C2 and P1, respectively. Translational pseudosymmetry was observed in which the non-crystallographic symmetry (NCS) operators calculated for the lower symmetry space groups were nearly equal to those belonging to R32 lattice symmetry. Therefore, an investigation into the source of the symmetry breakdown was necessary to ensure that the diffraction data were processed correctly. The transformation matrix relating the C2 lattice and R32 lattice are shown below from the *International Tables of Crystallography. Vol. A).*

$$\begin{bmatrix} h \\ k \\ l \end{bmatrix}_{R32} = \begin{bmatrix} ½ & ½ & ½ \\ -½ & ½ & 0 \\ 0 & 0 & ½ \end{bmatrix} \begin{bmatrix} h \\ k \\ l \end{bmatrix}_{C2}$$

$$\begin{bmatrix} h \\ k \\ l \end{bmatrix}_{C2} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} h \\ k \\ l \end{bmatrix}_{R32}$$

To understand the nature of the prevalence of psuedosymmetry in the crystallographic data, it is important to consider the oligomeric structure of IPPase. IPPase exists in solution as a homohexamer, which itself has intrinsic hexameric symmetry. Based on crystallographic occupancy calculations, differences in affinity of individual monomeric active sites for metals, substrate, or product can lead to structural differences between the monomeric subunits of the hexamer. When the differences are significant they can cause a breakdown in the higher crystallographic symmetry. In the

case of IPPase, ligand binding has been observed to induce asymmetry in the trimer of *E. coli*[80] *and is* also observed in the *T. thioreducens* IPPase.

Depending on the degree of symmetry breakdown in the reflection data, two cases are described in which the diffraction data were processed in the lower symmetry subgroups C2 and P1. An analysis of the diffraction data leading to the lower symmetry space groups P1 and C2 structures revealed the presence of translational pseudosymmetry based on Patterson methods and calculations performed using the *Phenix.Xtriage* program. In the case of the Ca^{2+}-bound structure processed in space group C2, a peak was found in the Patterson function (at fractional coordinates: 0.5, 0.025, 0.5) that was 38% of the height of the origin. To eliminate any ambiguity or uncertainty regarding the space group selected in each case it was necessary to understand the source of the breakdown in symmetry by comparing the crystal packing arrangements of the different crystal forms of IPPase.

Crystals of *T. thioreducens* were obtained in 3 crystal forms within R32, or C2 and P1 subgroups each having a different asymmetric unit composition (Table 4.2).

Table 4.2. IPPase crystal forms exhibit translational pseudosymmetry

Crystallized component	*Molecules per A.U.*	*Space group*	*Unit cell parameters (dimensions in Å , angles in °)*	*Crystallization conditions*
Metal-free IPPase bound to sulfate	6	P1	a=62.8, b=67.3,c=74.1 α=105.9, β-106.6, γ=101.9	1.0M ammonium sulfate, 10% v/v 2-propanol
Calcium- coordinated complex	6	C2	a=106.1, b=95.5, c=113.7 α =90, β=98.1, γ=90	0.02M calcium chloride, 0.1M sodium acetate pH 4.6, 20% v/v MPD
Manganese- coordinated IPPase	2	R32	a=114.5, b=114.5.1, c=148 α, β=90, γ=120	0.1M HEPES pH 7.5, 10% w/v PEG 8,000, 8%,v/v ethylene glycol
Magnesium- activated product complex	1	R32	a=99.3, b=99.3, c=98.6 α, β=90, γ=120	0.1M magnesium chloride, 0.1M sodium HEPES pH 7.5, 15% v/v 2-propanol

To compare the structural differences in the monomeric structures of IPPase, asymmetric unit monomers of the structures solved in the lower symmetry space groups were superimposed and compared to determine the average Root-Mean-Square-Deviation (RMSD) using the *Phenix.Superpose* program. As a result, there are significant structural differences among the IPPase monomers in the asymmetric unit of the sulfate bound structure (PDB ID: 3R6E). Average RMSD values of greater than 0.25$Å^2$ were obtained by comparing the backbone atoms of the individual monomers. This was also observed when comparing the two monomers of the Mn^{2+}-coordinated structure (PDB ID: 3R5U) and six monomers of the asymmetric unit of the Ca^{2+} (PDB ID: 3Q5V), in which a 0.504$Å^2$ and a 0.384 average RMSD value was obtained when the backbones atoms were superimposed.

Structural differences were also evident when comparing the inter-molecular contacts of the structures. A summary of the differences in oligomeric contacts between the six monomers comprising the P1 triclinic lattice of the sulfate-bound structure (PDB ID: 3R6E) has been provided in Appendix B. The number of hydrogen bonds and other non- covalent interactions and their bonding pattern differ when comparing the intermolecular contacts residues between asymmetric unit monomers.

Similarly, crystallographic symmetry breakdown was detected in the Mn^{2+}-coordinated structure (PDB ID:3R5U), in which a breakdown of symmetry in the two-fold crystallographic axis was observed. In the case of the Ca^{2+}-coordinated structure (PDB ID:3Q5V), the structural differences among the monomers caused a breakdown in the three-fold crystallographic axis. The structural differences among the monomers in the sulfate complex (PDB ID:3R6E) were also seen to break symmetry in both the two-fold and three-fold crystallographic axis. In each case, the breakdown in symmetry prevented the data from being processed in the higher symmetry R32 space group.

It was concluded that in the absence of saturating metal conditions or in the presence of specific ligands (i.e., sulfate) there is evidence that common pathologies arise in space group assignment involving translational pseudosymmetry. These observations suggest that substrate, or product ligand binding may have substantial effects on the oligomeric structure of IPPase.

4.3 Proposed mechanism of catalysis

The detailed structural analysis of the *T. thioreducens* IPPase substrate/product binding site from eight X-ray and one neutron crystallographic structures allowed the

proposal of a structure-based model for the enzyme's catalytic mechanism. The catalytic domain of IPPase shares topological similarity with IPPases studied from prokaryotic and eurkaryotic sources. A sequence alignment of IPPase with 15 other IPPases from the three domains of life reveals that only 17 amino acids are invariable despite the high degree of structure conservation (Table 4.1; left column). Out of the 17 conserved amino acids, 13 are involved in substrate and metal binding. The remaining conserved residues are found in the active but do not directly interact with the substrate. Instead, these residues help to shape and stabilize the active site fold.

Biochemical experiment and X-ray crystallographic structures of IPPase have shown the archaeal enzyme to exhibit active site rearrangements that are consistent with studies conducted on family I IPPase enzymes from prokaryotic and eukaryotic sources.[2;3] In these investigations, it was proposed that Asp68 in *T. thioreducens* IPPase (equivalent to Asp67 in *E. coli* and Asp117 in *S. cerevisiae*) was the most probable proton acceptor for water activation ($^{-}$OH formation) toward the nucleophilic attack and subsequent hydrolysis of the PP_i substrate.[1; 3] While the position and bonding properties of Asp68 are consistent with studies performed on other IPPase enzymes, the data reported here provide evidence of an alternate pathway for nucleophile activation independent of aspartate. In this alternate pathway, nucleophile activation is accomplished by transferring the proton along a chain of water molecules to the bulk solvent.

The possibility of a water proton relay system was suggested in light of the retained activity in the absence of Asp67 in studies on the *E. coli* enzyme.[2] To explain the activity exhibited by the *E. coli* Asp67Asn mutant enzyme, it was suggested that a

short chain of hydrogen-bonded water molecules could relay the proton to the bulk solvent. However, neither the *E. coli* or *S. cerevisiae* could definitively identify and discriminate between potential alternate pathways for proton transfer in IPPase.[1] Currently, the positions of protons in critical active site residues are not known for the *E. coli* or the *S. cerevisiae* enzymes.

The most powerful method to identify the spatial arrangement of hydrogen atoms is neutron crystallography. While neutron diffraction studies have been attempted for *S. cerevisiae* IPPase, no significant results were produced due to instability of the crystals and experimental complications caused by the neutron scattering properties of manganese ions in the enzyme's active site. Therefore, a model system to study proton transfer in IPPase was desirable.

The IPPase crystals used for neutron crystallography contained the largest unit cell volume of all neutron structures determined to-date. The resolution of the neutron structure (2.5Å) was limited by the long C-axis (~113Å) and the low symmetry monoclinic space group. The true diffraction limit of the best diffracting crystal was measured to be ~2.1Å. However, the exposure time for each diffraction image was limiting such that short diffraction time was necessary to obtain the dataset completeness suitable for structure determination. Thus, only a 2.5Å resolution neutron structure of IPPase could be obtained within the experimental time constraints.

By coupling X-ray and neutron crystallographic structures of IPPase, the molecular sequence of events associated with the enzymatic hydrolysis of PP_i are revealed. First, the high-resolution X-ray crystallographic structures provided molecular snapshots of the events associated with IPPase catalysis (Figure 4.2). It is observed that a

single divalent metal ion, coordinated by three aspartates (66,71,103) preorders the active site, as seen in the representative IPPase structure (PDB ID 3R5V) (Figure 4.2A). Two additional divalent metals are required to coordinate the substrate as represented by the calcium inhibited enzyme substrate complex (PDB ID: 3Q9M) (Figure 4.2B). The proposed nucleophile is generated by activation of the water molecule between the M1 and M2 (shown as red sphere labeled "W1"). Amino acid side chains of Arg44, Lys30, Lys141, Tyr56, and Tyr140 coordinate the substrate. The third metal (M3) coordinates the phosphonyl oxygen of P2 to allow for the formation of a trigonal-bipyramidal transition state as a result of backside nucleophilic attack of the activated water molecule. Residue Asp68 is in position to deprotonate the attacking water. However, activation of the water molecule could be independent of Asp68 and be accomplished by a water relay chain established by M2 coordination and optimized for proton transfer upon M4 coordination to the P2 substrate group. Tyr56 and Tyr140 have been observed to form short hydrogen bonds (~2.5Å) with the substrate's oxygen atoms and may help stabilize the transition state. A fourth metal (M4) coordinates the substrate prior to substrate hydrolysis and is believed to play an important role in the polarization of the P1 phosphate in favor of substrate bond cleavage. The isomerization of the side chain of Glu32 coincides with the observation of a metal ion at the M4 site. As the electrons resonate back into the P2 phosphate group, the bonding electrons leave with the P1 phosphate group. M2 and M3 also likely serve as the electron sink for the P2 electrons and thus help stabilize the trigonal-bipyramidal transition state (Figure 4.2C). The hydrolyzed product structure (PDB ID 3Q46) at 0.99Å provides support for an addition-elimination mechanism of phosphoryl transfer in which the hydrolyzed product inverts as

a function of the back-side nucleophilic attack (Figure 4.2D). The catalyzed reaction produces two molecules of orthophosphate that are shuffled from the active site along a network of positively charged residues (PDB ID 3Q5V) (Figure 4.2E).

Second, both X-ray and neutron crystallographic structures were imperative in showing the positions of water molecules through tight hydrogen-bond contacts that are conserved in the X-ray crystallographic structures of IPPase (Figures 4.3 and 4.4). These protein residues, metal ions, and water molecules coordinate to form a proton relay system that could activate the nucleophile independent of Asp68. In the neutron crystallographic structure of IPPase (PDB ID: 3Q3L), the M1 site is fully occupied by a calcium ion. The neutron scattering properties of Ca^{2+} enabled the density of the coordinating waters to be clearly seen. This is significant because even if suitable crystals of the *S. cerevisiae* enzyme could have been used for neutron structure determination, the scattering properties of the bound Mn^{2+} ions would contribute toward significant density cancelation as discussed previously. In addition, there are four water molecules in line with one another to form the shortest path for proton transfer to activate the nucleophile. These water molecules, as modeled in the neutron density map, could be unambiguously oriented (Figure 3.22).

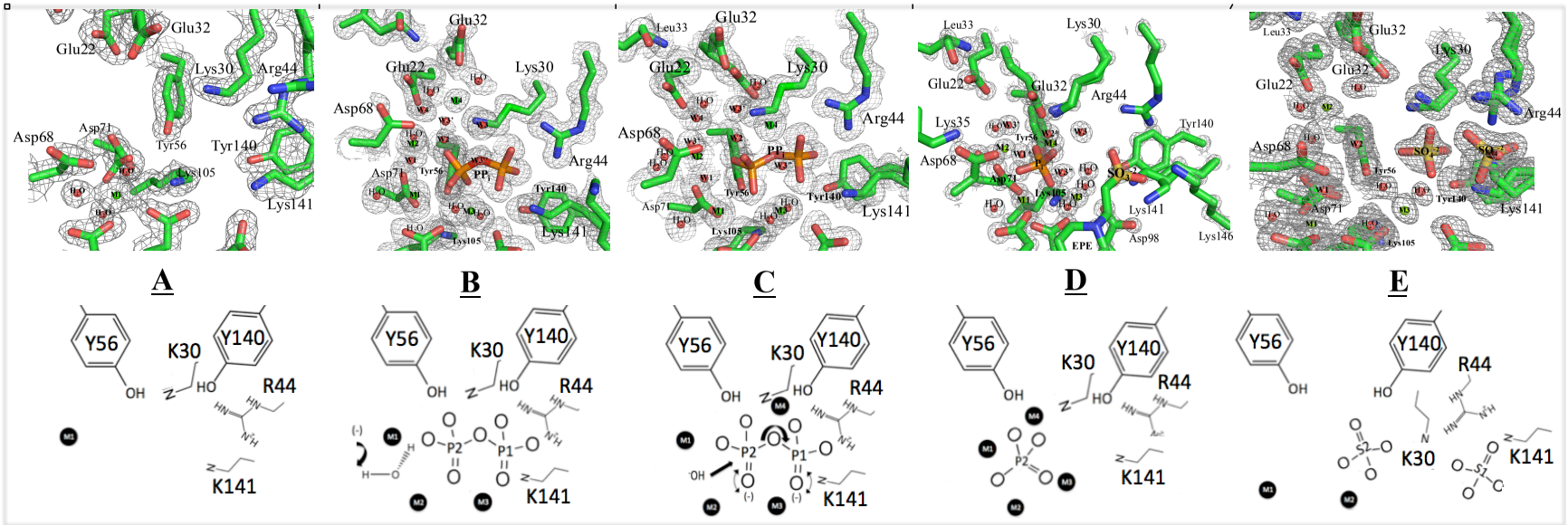


Figure 4.2. X-ray crystallographic snapshots of the events associated with IPPase catalysis. Shown are the active sites of the X-ray crystallographic structures used for the proposed structure-based mechanism of IPPase catalysis. Carbon, oxygen, and nitrogen atoms are colored green, blue, and red respectively. Water molecules and metal ions are shown as red and green spheres, respectively. Panels A-E correspond to the sequence of proposed events associated with the catalytic process as described in text. A. In the absence of substrate, M1 stabilizes Asp66, 71 and 103. B. Two additional metals, M2 and M3 coordinate the substrate upon binding. C. Isomerization of Glu32 and its coordination to M4 position the substrate for nucleophilic attack. D. Substrate is hydrolyzed leaving P2 bound coordinated with the four metals. E. Product analogues (sulfates) are coordinated by positively charged residues and positioned for removal.

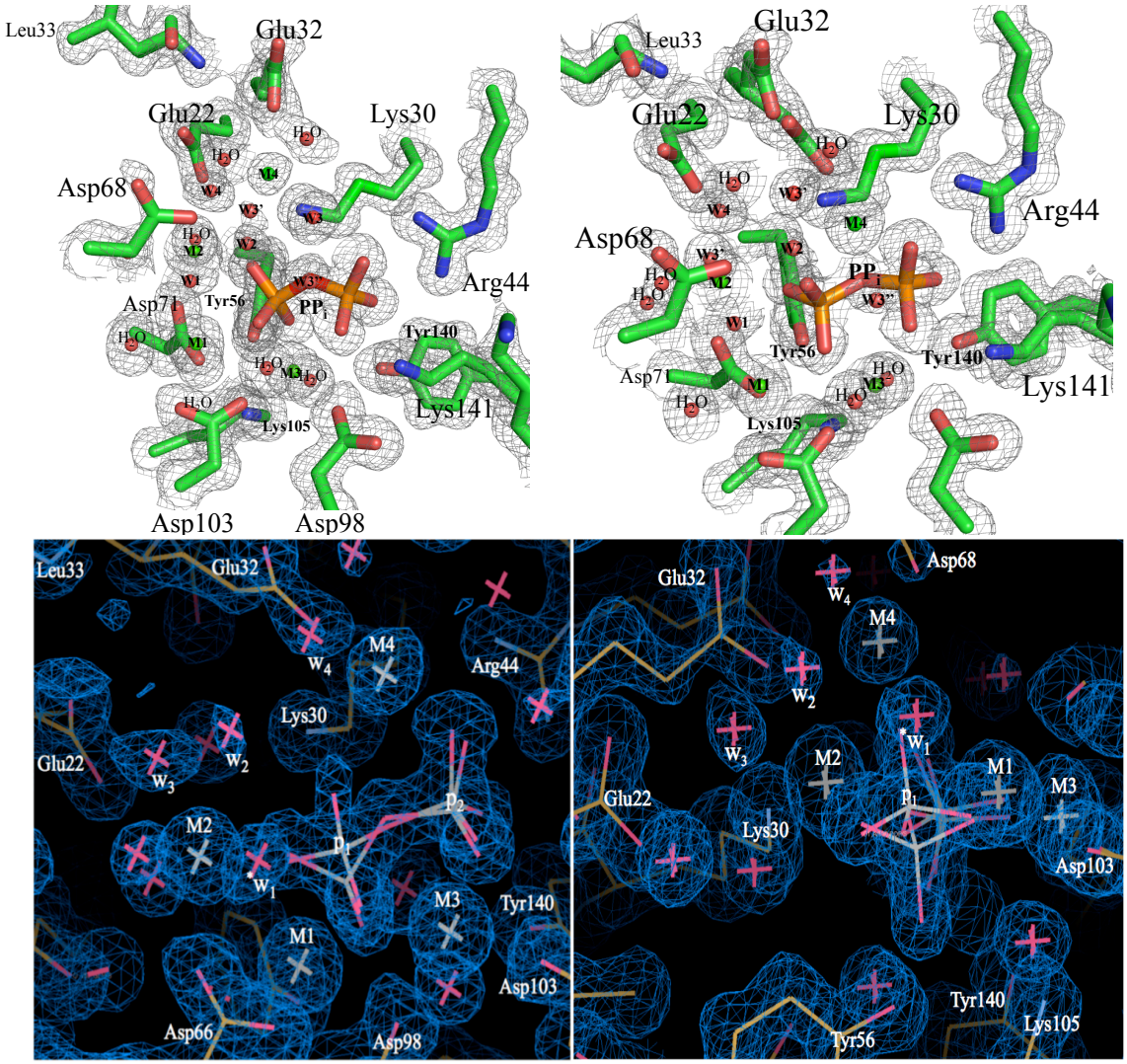


Figure 4.3. Glu32 isomerization as a function of substrate binding leads to changes in metal-water coordination. Shown are 1.5δ $2F_o$-F_c electron density screen captures of the water network that is conserved among the IPPase complexes. The orientations of a significant portion of these water molecules could be unambiguously determined in the neutron structure of IPPase. Seen in the panel of screen shots are different views of the water chain leading from M1, the site of nucleophile generation to the bulk solvent past Glu22.

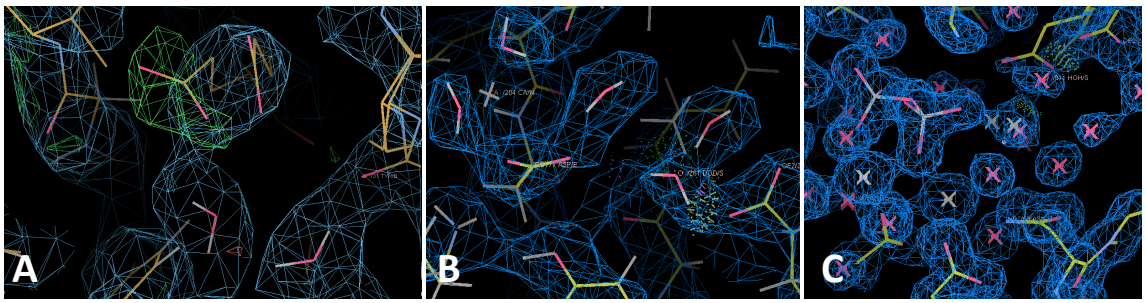


Figure 4.4. Water networks capable of proton transfer revealed by neutron diffraction studies. An alternate pathway for proton transfer during nucleophile activation is shown. Glu22 coordinates to a short chain of hydrogen bonded water molecules oriented to relay the proton from the water molecule positioned for nucleophilic attack between M1 and M2 (Panels A,B). The positions of the water molecules in the network are conserved among the IPPase substrate complexes (Panel C).

A comparison of all the *T. thioreducens* IPPase structures reported here suggests the water network coincides with the filling of the M1 site and may be optimized for proton transfer upon M2 and M4 binding. The water coordination of M2 and M4 undergoes discrete changes upon substrate binding and nucleophile generation and may represent a critical aspect of a metal-assisted process of nucleophile activation (Figure 3.19, 4.3).

The short chain of water molecules leading to Glu22 at the protein surface represent a water-based proton relay system. As seen in the neutron crystallographic structure of IPPase (PDB ID:3Q3L), the water molecules in this chain are oriented properly for proton transfer. As Glu32 coordinates to M4, changes in the metal-water coordination could result in partial proton donation to the substrate and thereby lowering the energy barrier of PP_i substrate hydrolysis. The coordination of Glu22 to the backbone amide nitrogen of Leu33 may be an important component of the activated

proton relay chain. Figure 4.5 provides an illustration of an alternate model for proton transfer along two chains hydrogen bonded water molecules.

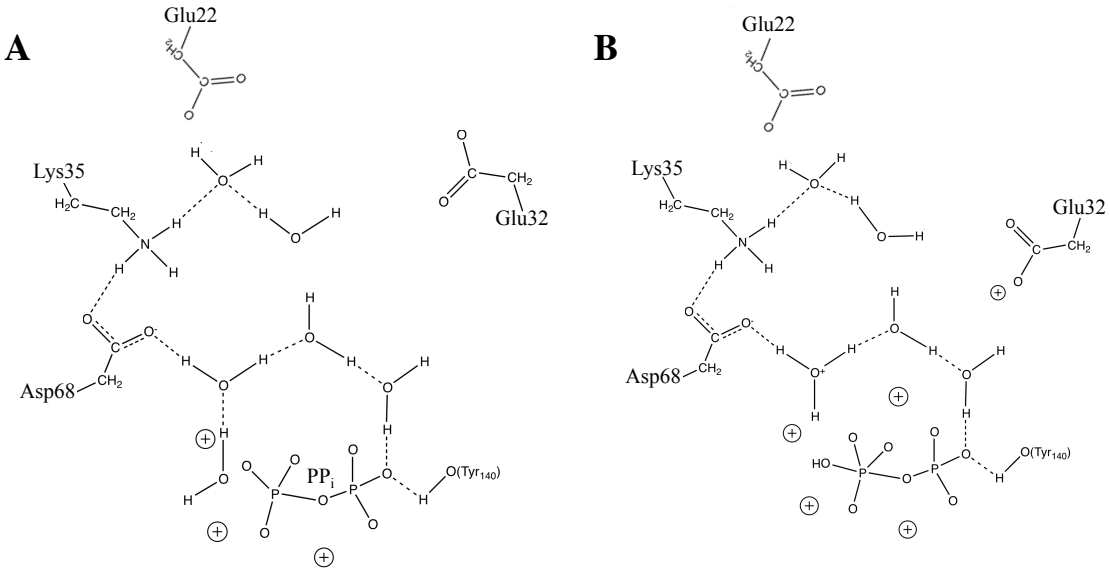


Figure 4.5. Role of Asp68 in the alternative model for proton transfer. In the alternative model for proton transfer the proton is transferred along a short-chain of hydrogen bonded water molecules to the bulk solvent. The isomerization of Glu32 cause discrete changes in active site coordination. M4 coordinates to Glu32 and to the P2 substrate group. The coordination of M4 to the P2 group orients the substrate properly a backside nucleophilic attack. The coordination of Asp68 and Lys35 could help coordinate the events of M4 substrate coordination, Glu32 isomerization, and nucleophile activation. Hydrogen bonding between Asp68 and the nucleophile may also help orient the nucleophile for attack on the P2 phosphorus atom.

4.4 Short hydrogen bonds

Tyr56 and Tyr140 form short hydrogen bonds (~2.5Å**)** with the substrate's oxygen atoms of the P2 and P1 phosphate groups respectively. Both residues could potentially be sharing a proton with each substrate group. Tyr140 and the fourth metal (M4) coordinate with the P1 group of the substrate prior to hydrolysis and could contribute to the polarization of the P1 phosphate in favor of substrate bond cleavage (Figure 4.5).

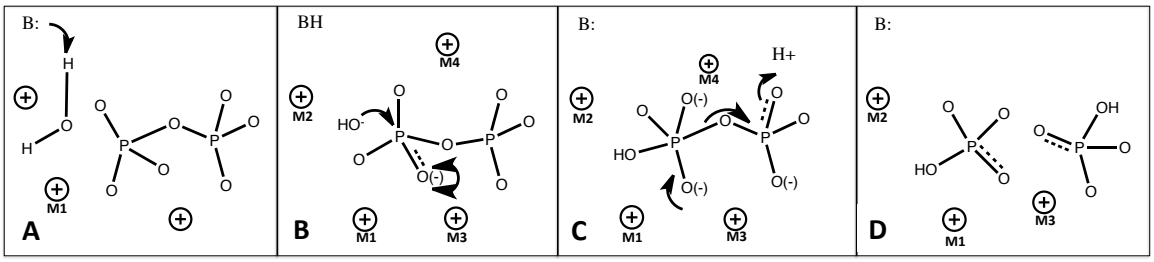


Figure 4.6 Model for IPPase substrate hydrolysis. Shown is a proposed atomic model for the IPPase substrate hydrolysis. In order for the transition intermediate to form there is a charge build up that must be stabilized on the substrate oxygen atoms. Therefore, partial donation of charge from the coordinating residues and solvent protons is important to stabilize the transition state.

As the electrons resonate back into the P2 phosphate group, the bonding electrons leave with the P1 phosphate group. In a similar manner, Tyr56, M1, M2, and M3 could serve as the electron sink for the P2 electrons and thus help stabilize the peripheral charge associated with a trigonal bipyramidal transition state (Figure 4.7). As mentioned, the hydrolyzed product structure (PDB ID 3Q46) at 0.99Å provides support for the concerted mechanism of phosphoryl transfer, in which the hydrolyzed product inverts following a back-side nucleophilic attack. The cleavage of the phosphoester bond produces two molecules of orthophosphate. The products are then shuffled from the active site along a network of positively charged residues to the protein surface. In the proposed model, no major rearrangements of the active site is required for the cycle to be reset. The water molecule that forms a bridge between the product oxygen and Asp68 in the hydrolyzed product structure of IPPase (PDB ID: 3Q46) displaces the product between M1 and M2.

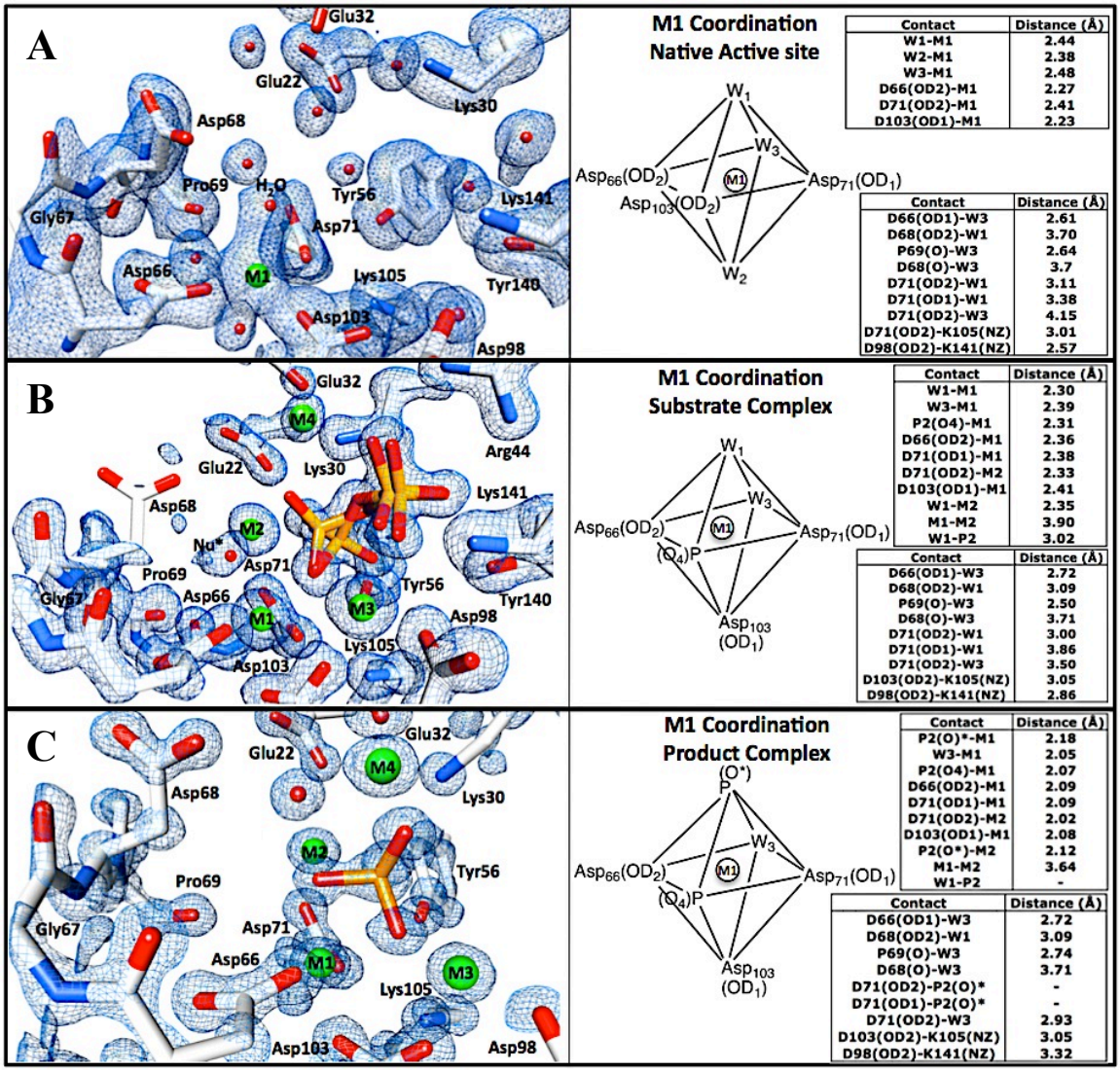


Figure 4.7. A comparison of active site coordination in IPPase X-ray structures. Shown left are the active sites of three IPPase complexes showing 2Fo-Fc electron density maps covering important active site residues. A. 1.65Å calcium-bound IPPase collected at room temperature. B. 1.35Å substrate bound IPPase inhibited by calcium. C. 0.99Å product bound IPPase activated by magnesium. Shown in the right frame are diagrams of the M1 coordination geometries and related contact distances.

The structural evidence reported here supports the hypothesis of this dissertation that the hydrolysis of inorganic pyrophosphate catalyzed by IPPase proceeds via direct phosphoryl transfer to water without the formation of a phosphorylated intermediate. It is believed that an extended network of hydrogen bonding interactions mediate and perhaps dictate the architecture of the catalytic site affecting the degree of catalysis and substrate recognition.

CHAPTER 5

CONCLUSION AND PROSPECTIVE

A family I IPPase enzyme from the isolated and characterized novel archaeon[45] *Thermococcus thioreducens* has been the subject of extensive structure-function studies. The open reading frame coding for the enzyme was amplified from the organism's genomic DNA, cloned into a bacterial experession vector, and over-expressed in *E. coli*. The three-dimensional structure of the overproduced enzyme was determined by X-ray crystallography at ultra-high resolution (>1Å) in complex with its substrate and product, and in the presence of various metals and reactive species analogues. Biochemical experiments were performed to determine the solution conditions for optimal enzymatic activity.

The biochemical data coupled with the X-ray crystallographic structures were used to establish a mechanistic model for IPPase catalysis in the domain Archaea. The proposed mechanism was compared with the established models derived from studies on the bacterial and eukaryotic enzymes of *E. coli* and *S. cerevisiae*. As a result, the archaeal enzyme was shown to undergo conserved structural changes as a function of substrate and product coordination consistent with existing model mechanisms proposed for the bacterial and eukaryotic enzymes. The general mechanism of family I IPPase substrate hydrolysis appears to be tightly conserved in the archaeal domain. In the proposed mechanism a single water molecule is activated and undergoes a nucleophilic attack on the innermost phosphate group of the PP_i substrate. IPPase structures provide support for a mechanism of phosphoryl transfer in which the hydrolyzed product would

undergo an inversion as a function of the back-side nucleophilic attack. This is consistent with a reaction pathway proceeding through a trigonal-bipyramidal transition state.

In existing IPPase models, Asp68 (equivalent to the Asp67 in *E. coli* and Asp117 in *S. cerevisiae*) is the most likely principal acceptor for the deprotonation of the attacking nucleophile $^{-}$OH toward the hydrolysis of pyrophosphate substrate. Enzyme activity studies on an IPPase mutant construct Asp68Asn mutants revealed significant activity in spite of the loss of Asp68's proton-accepting capacity, an observation also made for the enzymes of *E. coli* and *S. cerevisiae.* These observations provided motivation to establish a model system based on the archaeal enzyme that could hypothesize potential alternate pathways for proton transfer capable of nucleophile activation.

Additional opportunities are also presented for future studies utilizing IPPase to examine and determine the role of hydrogen bonding in inter- active site communications within the IPPase oligomer. Cooperatively of monomeric active sites could be facilitated in a manner as for enzymes that depend on thiamine diphosphate cofactors for efficient catalysis.[39] It is possible that coordination in IPPase could also rely on proton relays or a "proton wire" which connects the active sites of the neighboring monomers in the hexamer. Therefore, IPPase represents an excellent model system for further insight in to the functional significance of oligomerization and the structural basis for differential metal activation.

APPENDICES

APPENDIX A

NUCLEOTIDE SEQUENCE AND PRIMARY SEQUENCE OF IPPASE

NUCLEOTIDE SEQUENCE

ATGAACCCGTTCCACGAGCTTGAGCCCGGACCGGAGGTTCCAGAGGTCGTTTACG
CTCTCATAGAGATACCGAAGGGGAGCAGGAACAAGTACGAGCTCGACAAGAAGAC
TGGTTTACTTAAGCTCGACCGCGTCCTTTACAGCCCGTTCTTCTATCCGGTCGAC
TACGGAATAATCCCGCAGACCTGGTACGACGACGGCGACCCCTTCGACATCATGG
TCATCATGCGCGAACCTGTCTACCCGCTCACCATCATCGAGGCCAGGCCGATAGG
CATCATGAAGATGGAGGACTCTGGCGATAAGGACTGGAAGGTTCTCGCCGTTCCG
GTCGAGGACCCCTACTTCAACGACTGGAAGGACATAAGTGATGTTCCCAAGGCTT
TCCTCGACGAGATCGCCCACTTCTTCCAGAGGTACAAGGAGCTGCAAGGAAAGAC
CACCAAGATAGAGGGCTGGGGCAACGCCGAGGAGGCCAAGAGGGAGATACTCAGG
GCCATCGAGATGTACAAGGAGAAGTTCGGCAAGGAGGAGTGA

PRIMARY AMINO ACID SEQUENCE

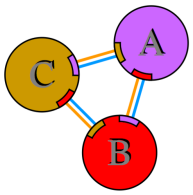
MNPFHELEPGPEVPEVVYALIEIPKGSRNKYELDKKTGLLKLDRVLYSPFFYPVD
YGIIPQTWYDDGDPFDIMVIMREPVYPLTIIEARPIGIMKMEDSGDKDWKVLAVP
VEDPYFNDWKDISDVPKAFLDEIAHFFQRYKELQGKTTKIEGWGNAEEAKREILR
AIEMYKEKFGKEE

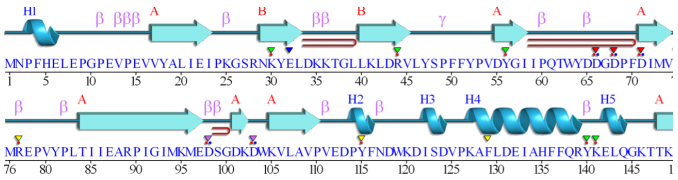
APPENDIX B

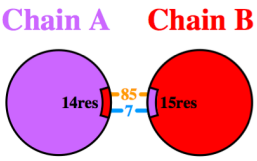
OLIOMERIC CONTACTS OF IPPASE STRUCTURES

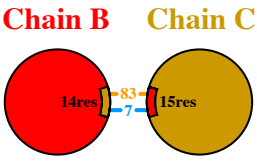
PDB ID 3Q9M

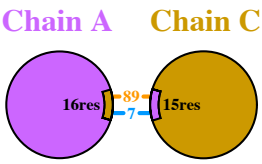
Chains	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
A⌘B	14:15	786:775	-	-	7	85
& B⌘C	14:15	796:784	-	-	7	83
A⌘C	16:15	778:801	-	-	7	89

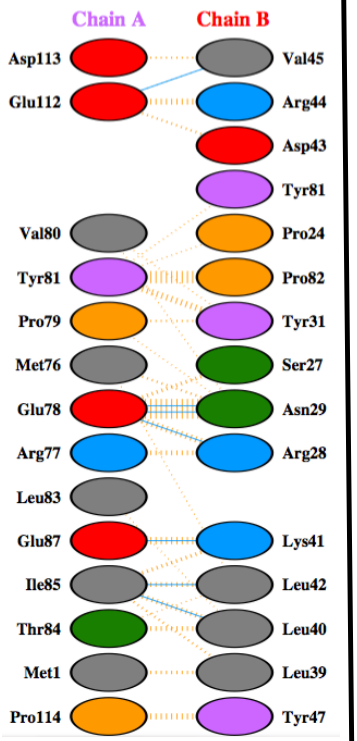


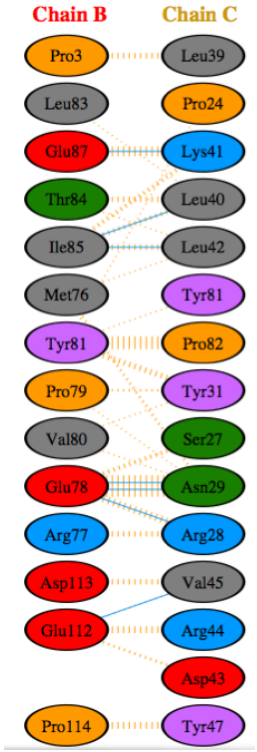


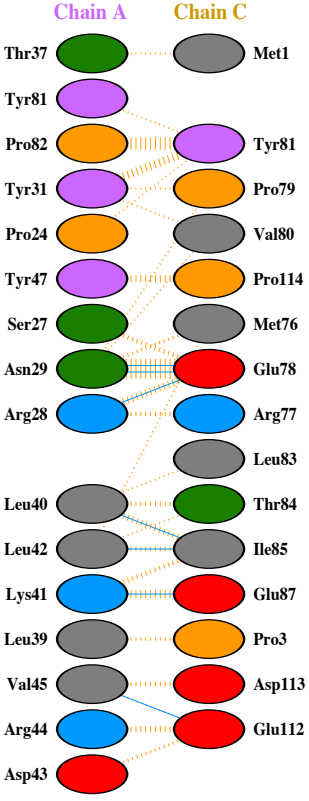






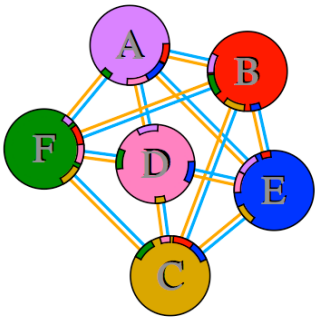




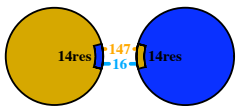


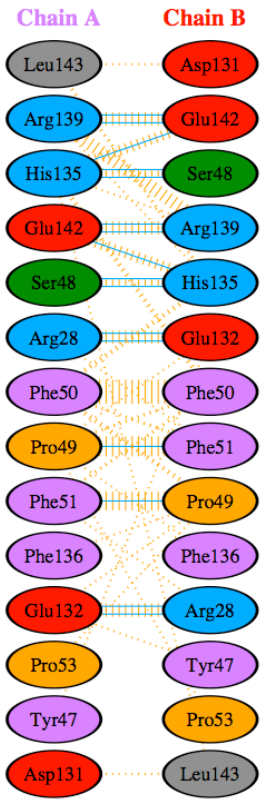
Key: Salt bridges; Disulphide bonds; Hydrogen bonds; Non-bonded contacts

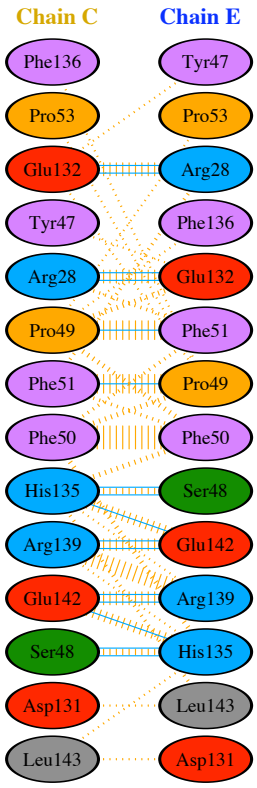
PDB ID 3R6E (PAGE 1 of 4)

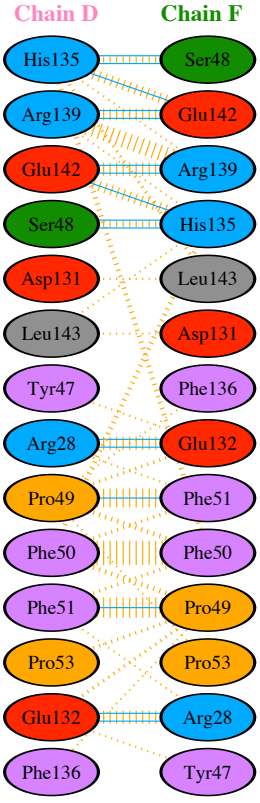


Chains	No. of interface residues	Interface area (A^2)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
A H B	14 : 14	718 : 718	-	-	16	143
& C H E	14 : 14	719 : 718	-	-	16	147
& D H F	14 : 14	721 : 716	-	-	16	148
A H D	16 : 17	780 : 752	-	-	6	97
& B H C	16 : 17	775 : 768	-	-	6	93
& C H F	16 : 16	781 : 773	-	-	6	89
& D H E	16 : 17	770 : 767	-	-	6	96
A H E	17 : 16	775 : 781	-	-	7	98
& B H F	17 : 16	770 : 775	-	-	6	94
A H F	8 : 8	360 : 364	-	-	2	37
& B H E	8 : 8	362 : 363	-	-	2	36
& C H D	8 : 8	366 : 362	-	-	2	36

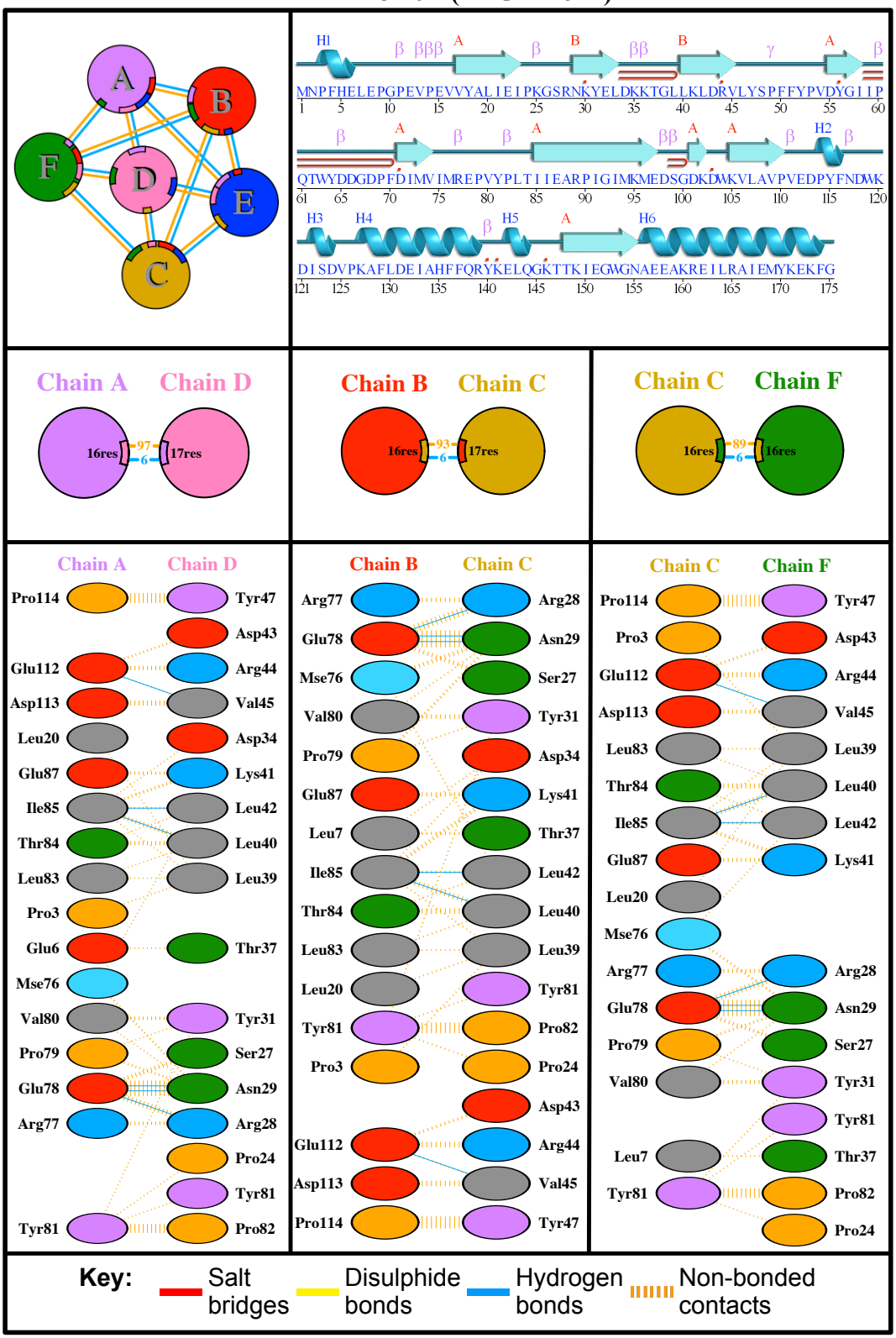


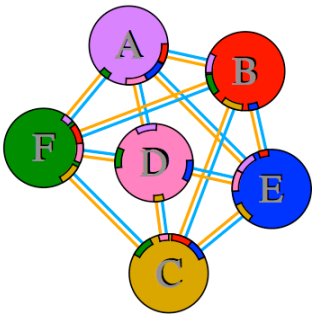




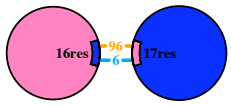


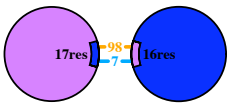
Key: Salt bridges | Disulphide bonds | Hydrogen bonds | Non-bonded contacts

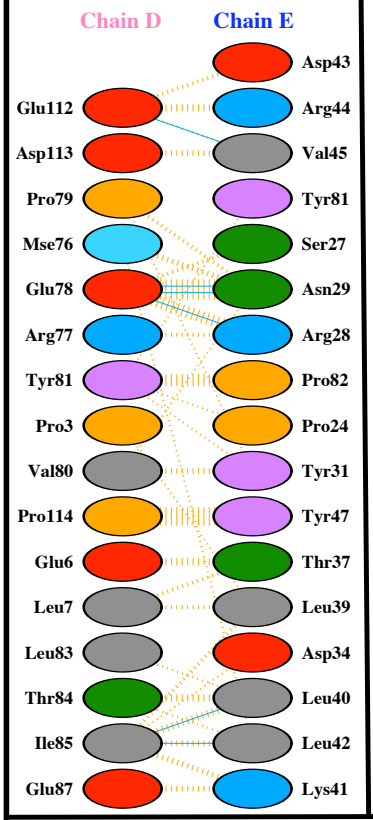

A
B
C
D
E
F
MNPFHELEPGPEVPEVVYALIEIPKGSRNKYELDKKTGLLKLDRVLYSPFFYPVDYGIIP
QTWYDDGDPFDIMVIMREPVYPLTIIEARPIGIMKMEDSGDKDWKVLAVPVEDPYFNDWK
DISDVPKAFLDEIAHFFQRYKELQGKTTKIEGWGNAEEAKREILRAIEMYKEKFG
Chain A
Chain D
16res
17res
Chain B
Chain C
Chain C
Chain F
Pro114
Glu112
Asp113
Leu20
Glu87
Ile85
Thr84
Leu83
Pro3
Glu6
Mse76
Val80
Pro79
Glu78
Arg77
Tyr81
Tyr47
Asp43
Arg44
Val45
Asp34
Lys41
Leu42
Leu40
Leu39
Thr37
Tyr31
Ser27
Asn29
Arg28
Pro24
Pro82
Leu7
Arg77
Glu78
Mse76
Val80
Pro79
Tyr81
Pro3
Glu112
Asp113
Pro114
Arg28
Asn29
Ser27
Tyr31
Asp34
Lys41
Thr37
Leu42
Leu40
Leu39
Tyr81
Pro82
Pro24
Asp43
Arg44
Val45
Tyr47
Pro114
Pro3
Glu112
Asp113
Leu83
Thr84
Ile85
Glu87
Leu20
Mse76
Arg77
Glu78
Pro79
Val80
Leu7
Tyr81
Tyr47
Asp43
Arg44
Val45
Leu39
Leu40
Leu42
Lys41
Arg28
Asn29
Ser27
Tyr31
Tyr81
Thr37
Pro82
Pro24
Key:
Salt bridges
Disulphide bonds
Hydrogen bonds
Non-bonded contacts



Chains	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
A & B	14 : 14	718:718	-	-	16	143
C & E	14 : 14	719:718	-	-	16	147
D & F	14 : 14	721:716	-	-	16	148
A & D	16 : 17	780:752	-	-	6	97
B & C	16 : 17	775:768	-	-	6	93
C & F	16 : 16	781:773	-	-	6	89
D & E	16 : 17	770:767	-	-	6	96
A & E	17 : 16	775:781	-	-	7	98
B & F	17 : 16	770:775	-	-	6	94
A & F	8 : 8	360:364	-	-	2	37
B & E	8 : 8	362:363	-	-	2	36
C & D	8 : 8	366:362	-	-	2	36

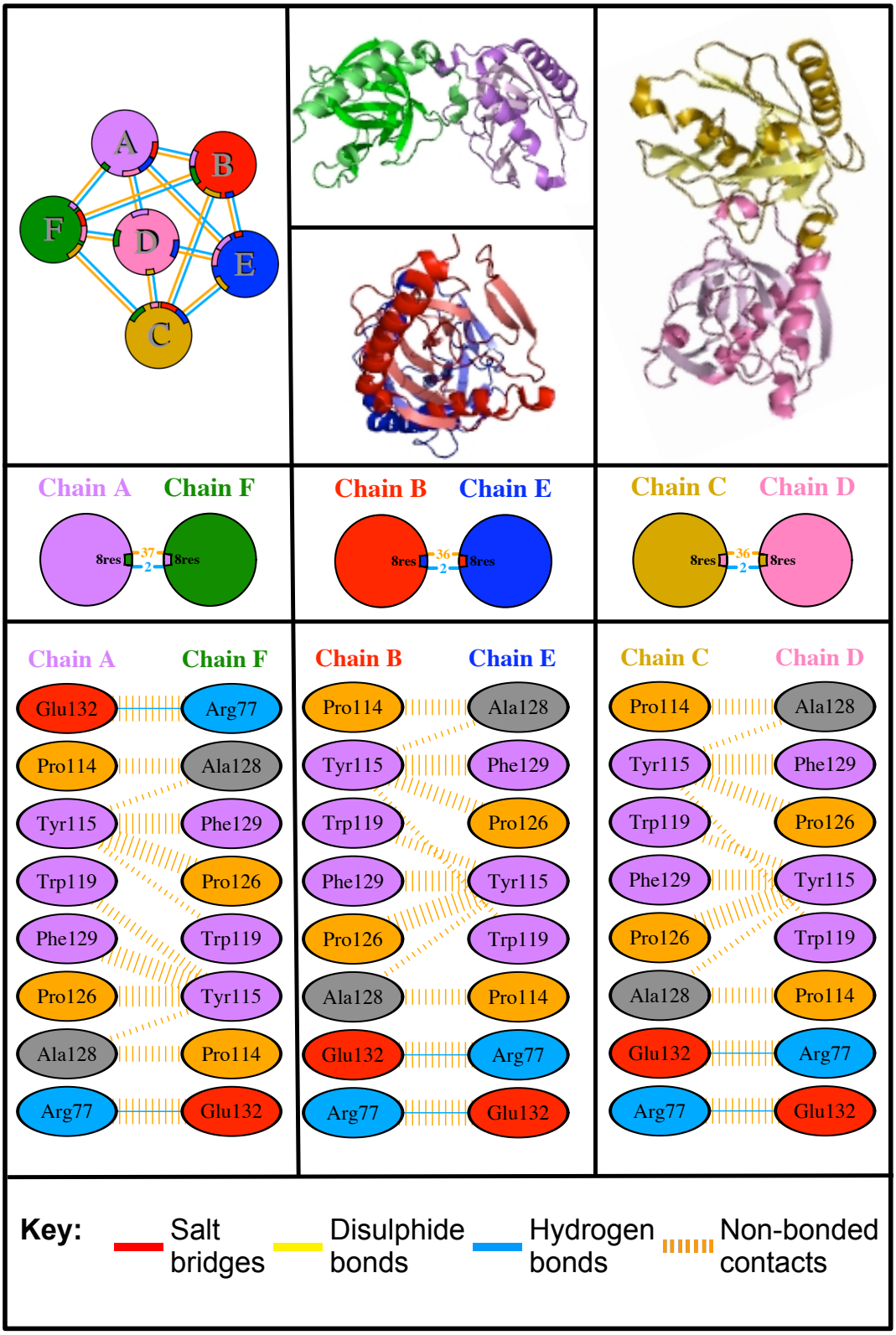






Key: Salt bridges — Disulphide bonds — Hydrogen bonds — Non-bonded contacts

PDB ID 3R6E (PAGE 4 of 4)



REFERENCES

[1]Samygina, V. R., Moiseev, V. M., Rodina, E. V., Vorobyeva, N. N., Popov, A. N., Kurilova, S. A., Nazarova, T. I., Avaeva, S. M. & Bartunik, H. D. (2007). Reversible inhibition of *Escherichia coli* inorganic pyrophosphatase by fluoride: Trapped catalytic intermediates in cryo-crystallographic studies. *J Mol Biol* **366**, 1305-17.

[2]Harutyunyan, E. H., Oganessyan, V. Y., Oganessyan, N. N., Avaeva, S. M., Nazarova, T. I., Vorobyeva, N. N., Kurilova, S. A., Huber, R. & Mather, T. (1997). Crystal structure of holo inorganic pyrophosphatase from *Escherichia coli* at 1.9 A resolution. Mechanism of hydrolysis. *Biochemistry* **36**, 7754-60.

[3]Oksanen, E., Ahonen, A.-K., Tuominen, H., Tuominen, V., Lahti, R., Goldman, A. & Heikinheimo, P. (2007). A complete structural description of the catalytic cycle of yeast pyrophosphatase. *Biochemistry* **46**, 1228-1239.

[4]Heikinheimo, P., Lehtonen, J., Baykov, A., Lahti, R., Cooperman, B. S. & Goldman, A. (1996). The structural basis for pyrophosphatase catalysis. *Structure* **4**, 1491-508.

[5]Chao, T. C., Huang, H., Tsai, J. Y., Huang, C. Y. & Sun, Y. J. (2006). Kinetic and structural properties of inorganic pyrophosphatase from the pathogenic bacterium *Helicobacter pylori*. *Proteins* **65**, 670-80.

[6]Liu, B., Li, X., Gao, R., Zhou, W., Xie, G., Bartlam, M., Pang, H., Feng, Y. & Rao, Z. (2004). Crystallization and preliminary X-ray analysis of inorganic pyrophosphatase from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3. *Acta Crystallogr D Biol Crystallogr* **60**, 577-9.

[7]Leppanen, V. M., Nummelin, H., Hansen, T., Lahti, R., Schafer, G. & Goldman, A. (1999). Sulfolobus acidocaldarius inorganic pyrophosphatase: structure, thermostability, and effect of metal ion in an archael pyrophosphatase. *Protein Sci* **8**, 1218-31.

[8]Baykov, A. A., Cooperman, B. S., Goldman, A. & Lahti, R. (1999). Cytoplasmic inorganic pyrophosphatase. *Prog Mol Subcell Biol* **23**, 127-50.

[9]Harutyunyan, E. H., Kuranova, I. P., Vainshtein, B. K., Hohne, W. E., Lamzin, V. S., Dauter, Z., Teplyakov, A. V. & Wilson, K. S. (1996). X-ray structure of yeast inorganic pyrophosphatase complexed with manganese and phosphate. *Eur J Biochem* **239**, 220-8.

[10]Avaeva, S. M. (2000). Active site interactions in oligomeric structures of inorganic pyrophosphatases. *Biochemistry (Mosc)* **65**, 361-72.

[11]Baykov, A. A., Volk, S. E. & Unguryte, A. (1989). Inhibition of inorganic pyrophosphatase of animal mitochondria by calcium. *Arch Biochem Biophys* **273**, 287-91.

[12]Baykov, A. A., Hyytia, T., Volk, S. E., Kasho, V. N., Vener, A. V., Goldman, A., Lahti, R. & Cooperman, B. S. (1996). Catalysis by *Escherichia coli* inorganic pyrophosphatase: pH and Mg^{2+} dependence. *Biochemistry* **35**, 4655-61.

[13]Baykov, A. A., Shestakov, A. S., Kasho, V. N., Vener, A. V. & Ivanov, A. H. (1990). Kinetics and thermodynamics of catalysis by the inorganic pyrophosphatase of *Escherichia coli* in both directions. *Eur J Biochem* **194**, 879-87.

[14]Baykov, A. A., Alexandrov, A. P. & Smirnova, I. N. (1992). A two-step mechanism of fluoride inhibition of rat liver inorganic pyrophosphatase. *Arch Biochem Biophys* **294**, 238-43.

[15]Welsh, K. M., Armitage, I. M. & Cooperman, B. S. (1983). Yeast inorganic pyrophosphatase. Functional and 113Cd2+ and 31P nuclear magnetic resonance studies of the Cd2+-enzyme. *Biochemistry* **22**, 1046-54.

[16]Lee, H. S., Cho, Y., Kim, Y. J., Lho, T. O., Cha, S. S., Lee, J. H. & Kang, S. G. (2009). A novel inorganic pyrophosphatase in *Thermococcus onnurineus* NA1. *FEMS Microbiol Lett* **300**, 68-74.

[17]Baykov, A. A., Volk, S. E. & Unguryte, A. (1989). Inhibition of inorganic pyrophosphatase of animal mitochondria by calcium. *Arch Biochem Biophys* **273**, 287-91.

[18]Avaeva, S. M., Vorobyeva, N. N., Kurilova, S. A., Nazarova, T. I., Polyakov, K. M., Rodina, E. V. & Samygina, V. R. (2000). Mechanism of Ca^{2+}-induced inhibition of *Escherichia coli* inorganic pyrophosphatase. *Biochemistry (Mosc)* **65**, 373-87.

[19]Baykov, A. A., Artjukov, A. A. & Avaeva, S. M. (1977). Fluoride inhibition of inorganic pyrophosphatase. III. Dependence on the nature of substrate and metal ion cofactor. *Biochimica et biophysica acta* **481**, 195-201.

[20]Pohjanjoki, P., Lahti, R., Goldman, A. & Cooperman, B. S. (1998). Evolutionary conservation of enzymatic catalysis: Quantitative comparison of the effects of mutation of aligned residues in *Saccharomyces cerevisiae* and *Escherichia coli* inorganic pyrophosphatases on enzymatic activity. *Biochemistry* **37**, 1754-61.

[21]Avaeva, S., Ignatov, P., Kurilova, S., Nazarova, T., Rodina, E., Vorobyeva, N., Oganessyan, V. & Harutyunyan, E. (1996). *Escherichia coli* inorganic pyrophosphatase: Site-directed mutagenesis of the metal binding sites. *FEBS Lett* **399**, 99-102.

[22]Hyytia, T., Halonen, P., Salminen, A., Goldman, A., Lahti, R. & Cooperman, B. S. (2001). Ligand binding sites in *Escherichia coli* inorganic pyrophosphatase: Effects of active site mutations. *Biochemistry* **40**, 4645-53.

[23]Avaeva, S., Kurilova, S., Nazarova, T., Rodina, E., Vorobyeva, N., Sklyankina, V., Grigorjeva, O., Harutyunyan, E., Oganessyan, V., Wilson, K., Dauter, Z., Huber, R. & Mather, T. (1997). Crystal structure of *Escherichia coli* inorganic pyrophosphatase complexed with SO4(2-). Ligand-induced molecular asymmetry. *FEBS Lett* **410**, 502-8.

[24]Yang, L., Liao, R. Z., Yu, J. G. & Liu, R. Z. (2009). DFT study on the mechanism of *Escherichia coli* inorganic pyrophosphatase. *J Phys Chem B* **113**, 6505-10.

[25]Liu, B., Bartlam, M., Gao, R., Zhou, W., Pang, H., Liu, Y., Feng, Y. & Rao, Z. (2004). Crystal structure of the hyperthermophilic inorganic pyrophosphatase from the archaeon *Pyrococcus horikoshii*. *Biophys J* **86**, 420-7.

[26]Cleland, W. W., Frey, P. A. & Gerlt, J. A. (1998). The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* **273**, 25529-32.

[27]Blow, D. M. (1969). The study of alpha-chymotrypsin by x-ray diffraction. The Third CIBA Medal Lecture. *The Biochemical Journal* **112**, 261-8.

[28]Tamada, T., Kinoshita, T., Kurihara, K., Adachi, M., Ohhara, T., Imai, K., Kuroki, R. & Tada, T. (2009). Combined high-resolution neutron and X-ray analysis of inhibited elastase confirms the active site oxyanion hole but rules against a low-barrier hydrogen bond. *J Am Chem Soc* **131**, 11033-40.

[29]Perrin, C. L. & Nielson, J. B. (1997). "Strong" hydrogen bonds in chemistry and biology. *Annu Rev Phys Chem* **48**, 511-44.

[30]Cassidy, C. S., Lin, J. & Frey, P. A. (1997). A new concept for the mechanism of action of chymotrypsin: the role of the low-barrier hydrogen bond. *Biochemistry* **36**, 4576-84.

[31]Yamaguchi, S., Kamikubo, H., Kurihara, K., Kuroki, R., Niimura, N., Shimizu, N., Yamazaki, Y. & Kataoka, M. (2009). Low-barrier hydrogen bond in photoactive yellow protein. *Proc Natl Acad Sci U S A* **106**, 440-4.

[32]Steiner, H., Jonsson, B. H. & Lindskog, S. (1975). The catalytic mechanism of carbonic anhydrase. Hydrogen-isotope effects on the kinetic parameters of the human C isoenzyme. *European journal of biochemistry / FEBS* **59**, 253-9.

[33]Tu, C. K., Silverman, D. N., Forsman, C., Jonsson, B. H. & Lindskog, S. (1989). Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry* **28**, 7913-8.

[34]Silverman, D. N. & McKenna, R. (2007). Solvent-mediated proton transfer in catalysis by carbonic anhydrase. *Accounts of Chemical Research* **40**, 669-75.

[35]Fisher, S. Z., Kovalevsky, A. Y., Domsic, J. F., Mustyakimov, M., McKenna, R., Silverman, D. N. & Langan, P. A. (2010). Neutron structure of human carbonic anhydrase II: Implications for proton transfer. *Biochemistry* **49**, 415-21.

[36]Mikulski, R. L. & Silverman, D. N. (2010). Proton transfer in catalysis and the role of proton shuttles in carbonic anhydrase. *Biochimica et biophysica acta* **1804**, 422-6.

[37]de Kok, A., Hengeveld, A. F., Martin, A. & Westphal, A. H. (1998). The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria. *Biochimica et Biophysica Acta* **1385**, 353-66.

[38]Mattevi, A., Obmolova, G., Schulze, E., Kalk, K. H., Westphal, A. H., de Kok, A. & Hol, W. G. (1992). Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science* **255**, 1544-50.

[39]Frank, R. A., Titman, C. M., Pratap, J. V., Luisi, B. F. & Perham, R. N. (2004). A molecular switch and proton wire synchronize the active sites in thiamine enzymes. *Science* **306**, 872-6.

[40]Jordan, F. (2004). Biochemistry. How active sites communicate in thiamine enzymes. *Science* **306**, 818-20.

[41]Cleland, W. W. (2000). Low-barrier hydrogen bonds and enzymatic catalysis. *Arch Biochem Biophys* **382**, 1-5.

[42]Blum, M. M., Mustyakimov, M., Ruterjans, H., Kehe, K., Schoenborn, B. P., Langan, P. & Chen, J. C. (2009). Rapid determination of hydrogen positions and protonation states of diisopropyl fluorophosphatase by joint neutron and X-ray diffraction refinement. *Proc Natl Acad Sci U S A* **106**, 713-8.

[43]Kawamura, K., Yamada, T., Kurihara, K., Tamada, T., Kuroki, R., Tanaka, I., Takahashi, H. & Niimura, N. (2011). X-ray and neutron protein crystallographic analysis of the trypsin-BPTI complex. *Acta crystallographica. Section D, Biological crystallography* **67**, 140-8.

[44]Tomanicek, S. J., Wang, K. K., Weiss, K. L., Blakeley, M. P., Cooper, J., Chen, Y. & Coates, L. (2011). The active site protonation states of perdeuterated Toho-1 beta-lactamase determined by neutron diffraction support a role for Glu166 as the general base in acylation. *FEBS letters* **585**, 364-8.

[45]Pikuta, E. V., Marsic, D., Itoh, T., Bej, A. K., Tang, J., Whitman, W. B., Ng, J. D., Garriott, O. K. & Hoover, R. B. (2007). *Thermococcus thioreducens* sp. nov., a novel hyperthermophilic, obligately sulfur-reducing archaeon from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* **57**, 1612-8.

[46]Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-91.

[47]Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51**, 263-73.

[48]Marsic, D., Hughes, R., Byrne-Steele, M. & Ng, J. (2008). PCR-based gene synthesis to produce recombinant proteins for crystallization. *BMC Biotechnology* **8**, 44.

[49]Inoue, H., Nojima, H. & Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-8.

[50]Marsic, D., Hughes, R. C., Byrne-Steele, M. L. & Ng, J. D. (2008). PCR-based gene synthesis to produce recombinant proteins for crystallization. *BMC Biotechnol* **8**, 44.

[51]Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). Molecular cloning: A laboratory manual. *2nd edn. Col Spring Harbor, NY*.

[52]van Alebeek, G. J., Keltjens, J. T. & van der Drift, C. (1994). Purification and characterization of inorganic pyrophosphatase from Methanobacterium thermoautotrophicum (strain delta H). *Biochim Biophys Acta* **1206**, 231-9.

[53]Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: The PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr* **61**, 1426-31.

[54]Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326.

[55]French, S. A. W., K. . (1978). On the treatment of negative intensity observations. *Acta Cryst.* **A34**, 517-525.

[56]Brunger, A. T. (1997). Free R value: Cross-validation in crystallography. *Methods Enzymol* **277**, 366-96.

[57]Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. . *Journal Of Applied Crystallography* **30**, 1022-1025

[58]Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

[59]McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

[60]Adams, P. D., Grosse-Kunstleve, R.W. Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read R.J., Sacchettini, J.C., Sauter, N.K. & Terwilliger. T.C. (2002). PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallographica Section D* **58**, 1948-1954.

[61]Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60**, 2184-95.

[62]Emsley, P. & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126-2132.

[63]Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

[64]Garcia-Ruiz, J. M., Gonzalez-Ramirez, L. A., Gavira, J. A. & Otalora, F. (2002). Granada Crystallisation Box: A new device for protein crystallisation by counter-diffusion techniques. *Acta Crystallogr D Biol Crystallogr* **58**, 1638-42.

[65]Ng, J. D., Gavira, J. A. & Garcia-Ruiz, J. M. (2003). Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. *J Struct Biol* **142**, 218-31.

[66]García-Ruiz, J. M. & Charles W. Carter, J. A. R. M. S. (2003). Counterdiffusion Methods for Macromolecular Crystallization. In *Methods in Enzymology* **368**, 130-154.

[67]Ng, J. D., Stevens, R. C. & Kuhn, P. (2008). Protein crystallization in restricted geometry: Advancing old ideas for modern times in structural proteomics. In *Methods in Molecular Biology :Structural Proteomics* **426**, 363-376.

[68]Langan, P. & Greene, G. (2004). Protein crystallography with spallation neutrons: Collecting and processing wavelength-resolved Laue protein data. *Journal of Applied Crystallography* **37**, 253-7.

[69]Pflugrath, J. W. (1999). The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **55**, 1718-25.

[70]Blakeley, M. P., Teixeira, S. C., Petit-Haertlein, I., Hazemann, I., Mitschler, A., Haertlein, M., Howard, E. & Podjarny, A. D. Neutron macromolecular crystallography with LADI-III. *Acta Crystallogr D Biol Crystallogr* **66**, 1198-205.

[71]Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J. C., Cipriani, F., Hoghoj, P., Lehmann, M. S. & Wilkinson, C. (1997). Neutron Laue diffractometry with an imaging

plate provides an effective data collection regime for neutron protein crystallography. *Nat Struct Biol* **4**, 909-14.

[72]Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82.

[73]Otwinowski, Z., Borek, D., Majewski, W. & Minor, W. (2003). Multiparametric scaling of diffraction intensities. *Acta Crystallogr A* **59**, 228-34.

[74]Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80.

[75]Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-30.

[76]Ng, J. D. & Garcia-Ruiz, J. M. (2006). Counter-diffusion capillary crystallization for structural genomics. *Trends in Drug Discovery* **52,** 54-72

[77]Gavira, J. A., Toh, D., Lopez-Jaramillo, J., Garcia-Ruiz, J. M. & Ng, J. D. (2002). Ab initio crystallographic structure determination of insulin from protein to electron density without crystal handling. *Acta Crystallogr D Biol Crystallogr* **58**, 1147-54.

[78]Ostwald, W. (1897). Besprechung der Arbeit von Liesenganga A-Linien. Z. *Physical Chemistry* **23**, 365.

[79]Tomanicek, S. J., Blakeley, M. P., Cooper, J., Chen, Y., Afonine, P. V. & Coates, L. (2010). Neutron diffraction studies of a class A beta-lactamase Toho-1 E166A/R274N/R276N triple mutant. *J Mol Biol* **396**, 1070-80.

[80]Avaeva, S., Grigorjeva, O., Mitkevich, V., Sklyankina, V. & Varfolomeyev, S. (1999). Interaction of Escherichia coli inorganic pyrophosphatase active sites. *FEBS Lett* **464**, 169-73.