# Enzyme molecular choreography — Studies of soluble inorganic pyrophosphatases

Esko Oksanen

Institute of Biotechnology and
Department of Chemistry, Laboratory of Organic Chemistry
Faculty of Science
University of Helsinki

and

National Graduate School in Informational and Structural Biology

**ACADEMIC DISSERTATION**

To be presented with the permission of the Faculty of Science of the University of Helsinki, for public criticism in the Auditorium LS2404 in Biocenter 3, Viikinkaari 1, on December 11th at 12 noon

**Helsinki 2009**

***Supervised by***
Professor Adrian Goldman
Structural Biology and Biophysics
Institute of Biotechnology
University of Helsinki
Finland

***Reviewed by***
Professor Peter C. Kahn
Department of Biochemistry and Microbiology
Rutgers University
New Jersey, United States of America

***and***

Associate Professor Carrie Wilmot
Department of Biochemistry, Molecular Biology and Biophysics
University of Minnesota
Minnesota, United States of America

***Opponent***
Adjunct Professor John Kozarich
Chairman & President, ActivX Biosciences, Inc.
Department of Chemical Physiology
The Scripps Research Institute, La Jolla
California, United States of America

***Custos***
Professor Ilkka Kilpeläinen
Laboratory of Organic Chemistry
Department of Chemistry
University of Helsinki
Finland

ISBN 978-952-10-5894-3 (paperback)
ISBN 978-952-10-5895-0 (PDF)
ISSN 1795-7079

Helsinki University Press
Helsinki 2009

"*Science is part of culture. Culture isn't only art and music and literature, it's also understanding what the world is made of and how it functions. People should know something about stars, matter and chemistry. People often say that they don't like chemistry but we deal with chemistry all the time. People don't know what heat is, they hardly know what water is. I'm always surprised how little people know about anything. I'm puzzled by it.*"

–Max F. Perutz

## ABBREVIATIONS

| | |
|---|---|
| ADP | Adenosine diphosphate |
| ATP | Adenosine triphosphate |
| CCD | Charge coupled device |
| DEAE | Diethylaminoethyl |
| DHFR | Dihyrofolate reductase |
| DNA | Deoxyribonucleic acid |
| EDTA | Ethylene diamine tetra-acetic acid |
| EGTA | Ethylene glycol tetra-acetic acid |
| EPR | Electron paramagnetic resonance |
| ESI-TOF | Electrospray ionisation – time-of-flight |
| ESRF | European Synchrotron Radiation Facility |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonate |
| HIV | Human immunodeficiency virus |
| HMBP | Hydroxymethylene bisphosphonate |
| HPLC | High pressure liquid chromatography |
| ILL | Institut Laue Langevin |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| J-PARC | Japan Proton Accelerator Research Center |
| JRR | Japan Research Reactor |
| LADI | Laue Diffractometer (ILL) |
| LB | Luria-Bertani |
| LBHB | Low-barrier hydrogen bond |
| LFER | Linear free energy relationship |
| MANDI | Macromolecular Neutron Diffractometer (SNS) |
| MBP | Methylene bisphosphonate |
| MES | 2-morpholinoethane sulphonic acid |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MPD | 2-methyl-2,4-pentanediol |
| NADPH | Nicotine adenine dinucleotide phosphate |
| NMR | Nuclear magnetic resonance |
| PCS | Protein Crystallography Station (LANSCE) |
| PNP | Imidodiphosphate |
| PPase | Pyrophosphatase |
| Y-PPase | Yeast PPase |
| E-PPase | *Escherichia coli* PPase |
| TmPPase | *Thermotoga maritima* PPase |
| BsPPase | *Bacillus subtilis* PPase |
| $PP_i$ | Inorganic pyrophosphate |
| QM/MM | Quantum mechanics/molecular mechanics |
| r.m.s.d. | Root mean square deviation |
| RNA | Ribonucleic acid |

| | |
|---|---|
| SDS-PAGE | Sodium dodecyl sulphate polyacrylamide gel electrophoresis |
| SNS | Spallation Neutron Source |
| Tris | Tris-hydroxymethylaminomethane hydrochloride |
| Å | $10^{-10}$ m |

## AMINO ACID ACRONYMS

| | | |
|---|---|---|
| A | Ala | Alanine |
| C | Cys | Cysteine |
| D | Asp | Aspartic acid |
| E | Glu | Glutamic acid |
| F | Phe | Phenylalanine |
| G | Gly | Glycine |
| H | His | Histidine |
| I | Ile | Isoleucine |
| K | Lys | Lysine |
| L | Leu | Leucine |
| M | Met | Methionine |
| N | Asn | Asparagine |
| P | Pro | Proline |
| Q | Gln | Glutamine |
| R | Arg | Arginine |
| S | Ser | Serine |
| T | Thr | Threonine |
| V | Val | Valine |
| W | Trp | Tryptophan |
| Y | Tyr | Tyrosine |

## ABSTRACT

Inorganic pyrophosphatases (PPases, EC 3.6.1.1) hydrolyse pyrophosphate in a reaction that provides the thermodynamic 'push' for many reactions in the cell, including DNA and protein synthesis. Soluble PPases can be classified into two families that differ completely in both sequence and structure. While Family I PPases are found in all kingdoms, family II PPases occur only in certain prokaryotes.

The enzyme from baker's yeast (*Saccharomyces cerevisiae*) is very well characterised both kinetically and structurally, but the exact mechanism has remained elusive. The enzyme uses divalent cations as cofactors; in vivo the metal is magnesium. Two metals are permanently bound to the enzyme, while two come with the substrate. The reaction cycle involves the activation of the nucleophilic oxygen and allows different pathways for product release.

In this thesis I have solved the crystal structures of wild type yeast PPase and seven active site variants in the presence of the native cofactor magnesium. These structures explain the effects of the mutations and have allowed me to describe each intermediate along the catalytic pathway with a structure. Although establishing the 'choreography' of the heavy atoms is an important step in understanding the mechanism, hydrogen atoms are crucial for the mechanism. The most unambiguous method to determine the positions of these hydrogen atoms is neutron crystallography. In order to determine the neutron structure of yeast PPase I perdeuterated the enzyme and grew large crystals of it. Since the crystals were not stable at ambient temperature, a cooling device was developed to allow neutron data collection.

In order to investigate the structural changes during the reaction in real time by time-resolved crystallography a photolysable substrate precursor is needed. I synthesised a candidate molecule and characterised its photolysis kinetics, but unfortunately it is hydrolysed by both yeast and *Thermotoga maritima* PPases.

The mechanism of Family II PPases is subtly different from Family I. The native metal cofactor is manganese instead of magnesium, but the metal activation is more complex because the metal ions that arrive with the substrate are magnesium – different from those permanently bound to the enzyme. I determined the crystal structures of wild type *Bacillus subtilis* PPase with the inhibitor imidodiphosphate and an inactive H98Q variant with the substrate pyrophosphate. These structures revealed a new trimetal site that activates the nucleophile. I also determined that the metal ion sites were partially occupied by manganese and iron using anomalous X-ray scattering.

# 1. INTRODUCTION

## 1.1 ENZYME MECHANISMS

Enzymes are biological catalysts that mediate almost all reactions in living cells. The mechanisms of enzymes have been studied for decades, but our understanding of the origin of their catalytic power is still incomplete. What do we then actually mean by understanding enzyme action? Because the reaction can be described at various levels of detail, requiring different experimental and computational approaches, I will also introduce the central concepts of enzyme mechanisms, as well as some methods to study them, particularly the changes in molecular structure in an enzymatic reaction. The ultimate aim is to fully characterise the potential energy surface of the reaction, allowing us to predict rate constants quantitatively from transition state theory (1), but enzymatic reactions are sufficiently complex that many questions have to be answered before addressing the nature of the elementary step. I will not attempt to cover all of enzymology, but will concentrate on phosphoryl transferases.

Phosphoryl transferases are an ubiquitous class of enzymes in all organisms. They catalyse the transfer of a phosphate moiety from a donor to an acceptor nucleophile (2) (Figure 1.1). The donor may be a phosphate ester such as a sugar phosphate or a phosphoanhydride such as adenosine triphosphate (ATP) or even a phosphonate. In the case of ATP synthethases the substrate is phosphoric acid itself. The acceptor may be an alcoholic (serine-threonine kinases) or phenolic hydroxyl (tyrosine kinases), another phosphate moiety (phosphoglycerate kinase) or water (phosphate hydrolases, including inorganic pyrophosphatases).

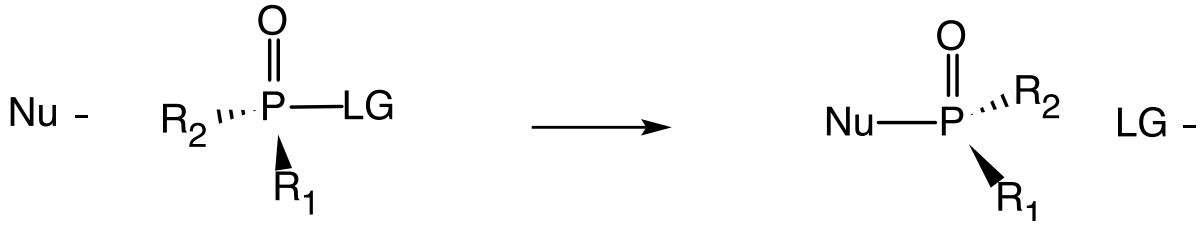


**Figure 1.1. The generalised reaction catalysed by phosphoryl transferases, where Nu is the nucleophile, LG the leaving group and the substituent groups $R_{1\text{-}2}$ .**

The biological significance of phosphoryl transferases can hardly be overestimated; nucleotide phosphates are the basic energy ‘currency’ of the cell and hence the energy transducing machinery of the cell must involve phosphoryl transfer. The key enzymes of glycolysis are phosphoryl transferases – from hexokinase that phosphorylates glucose to phosphoglycerate kinase that harvests the chemical energy in the form of a high-energy phosphoanhydride bond in ATP. ATP is also

generated by the $F_1F_0$-ATPase (3), in a rotary motion powered by the proton gradient across the mitochondrial membrane.

On the other hand many thermodynamically unfavourable biosynthetic reactions such as protein or nucleic acid biosynthesis are coupled to phosphoanhydride hydrolysis. Phosphoryl transferases are indeed involved in most of the fundamental metabolic processes of the cell.

The role of phosphoryl transferases is not limited to basic metabolism. Phosphorylation of proteins is essential for the regulation of protein function and cell signalling in eukaryotes. Proteins are specifically phosphorylated at serine, threonine or tyrosine hydroxyls by kinases and dephosporylated by phosphatases. Finally many kinases involved in signalling are interesting drug targets (4) against a number of diseases.

All enzymatic reactions must involve the formation of the enzyme-substrate complex and the dissociation of the enzyme-product complex, so from a kinetic point of view enzymatic reactions are always multi-step reactions. The phosphoryl transfer reaction itself involves a nucleophilic attack on the electrophilic phosphorous, followed by the loss of the leaving group (Figure 1.1). These events may be simultaneous or may include intermediates.

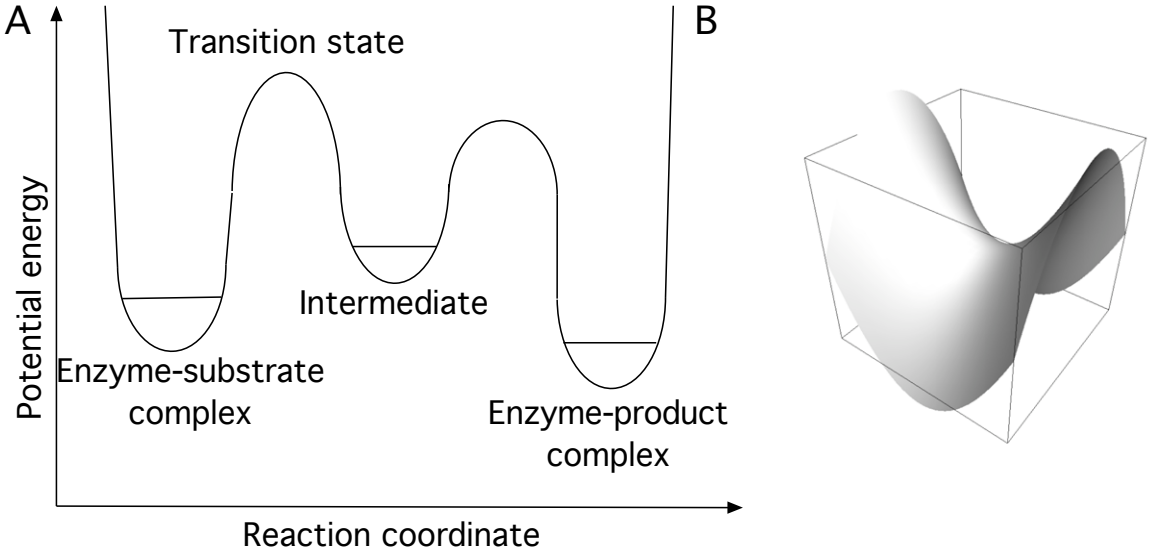


**Figure 1.2. A) A schematic potential energy plot along the reaction coordinate illustrates the difference between a transition state (a maximum) and an intermediate (a minimum). B) A hyperboloid surface with a saddle point: a maximum in one coordinate and a minimum in others.**

In this context it is important to make the distinction between a transition state and an intermediate (Figure 1.2A). A transition state is a saddle point (Figure 1.2B) in the potential energy surface of the reaction, so while it is a minimum in all other coordinates, it is also a maximum along the reaction coordinate. This means that the lifetime of the transition state is one vibrational period (in the order of femtoseconds) and it can never be isolated. An intermediate, on the other hand, is a potential energy minimum, but only a local one. Therefore reaction intermediates can be described as metastable species and can be isolated if their lifetime is long enough.

Changing the conditions either to increase the barrier toward the product or lower the thermal energy so that the barrier can no longer be crossed can also trap intermediates. The number and nature of intermediates and the rates of their interconversion define the *kinetic mechanism* of the reaction. This establishes the elementary steps of the reaction. The slowest step is called rate-determining; the rate constants for steps after the rate-determining step cannot be distinguished by following the rate of product formation.

### 1.1.1 Kinetic mechanism and its structural interpretation — atomic choreography

To establish the kinetic mechanism a number of individual rate constants must in principle be measured. The simplest way to study enzyme reactions is to measure the overall rate constant as a function of the substrate concentration and interpret the data in terms of some model. The typical starting point is the Michaelis-Menten model (5) where the formation of the enzyme-substrate complex is in pre-equilibrium with the rate limiting, irreversible conversion of the enzyme-substrate complex to product. Such *steady-state kinetics* can exclude certain mechanisms particularly when more complex models are used, but it is generally hard to prove that one particular model would be better than other possible models. As discussed below (1.2.1), clever experiments including the kinetics of the reverse reaction can sometimes yield all the microscopic rate constants of a reaction, but this is unusual.

In order to measure the individual rate constants, pre-steady state measurements are indispensable. Usually rapid mixing and sampling are necessary because in most enzymatic reactions the system reaches the steady state very quickly. Techniques such as continuous flow, stopped flow and quenched flow are commonly used. Kinetic studies of enzymes are discussed in detail in various textbooks, such as Fersht (6)

Although the kinetic mechanism is necessary for understanding the catalytic power of an enzyme, it is not quite sufficient. What the kinetic mechanism does not establish, first of all, is the structure of the intermediates. As the chemical composition of each intermediate is often known from various experiments, some intermediate structures can be inferred from the structures of the reactants and products. The phosphoenzyme intermediates discussed below (section 1.1.2) – where the site of enzyme phosphorylation more or less characterises the intermediate – are a good example. Typically however, the structures of intermediates have to be somehow trapped; this is discussed in more detail in section 1.3. The combination of the kinetic mechanism with the structures of each intermediate yield in principle the ‘choreography’ of atoms during the reaction. As the principal method of investigating enzyme structures at sufficient accuracy is X-ray crystallography, only the non-hydrogen atoms are included in these structures, as discussed in section 1.4. While even this level of information is very difficult to achieve, it is the starting point for truly understanding the sources of catalytic power in terms of electronic structure.

The choice of experimental methods for investigating electronic structure is limited. Optical spectroscopy can be applied in cases where an extended π-orbital system is present, and electron paramagnetic resonance (EPR) spectroscopy is useful with systems containing unpaired electrons. Cytochrome c oxidase – an enzyme of the respiratory chain that generates a proton gradient across the mitochondrial membrane (7) – is a good example of a system that has been studied by both optical (8) and EPR spectroscopy. Another way to investigate the changes in electronic structure in enzyme catalysis is computational simulation. Sufficiently accurate electronic structure calculations with *ab initio* methods are, however, computationally very demanding, often limiting their application only to the active site. The rest of the protein can be treated classically; such quantum mechanics/classical mechanics (QM/MM) methods are becoming more and more powerful (9, 10). Computational methods still require accurate structural information as input, as well as experimental measurements to validate the computed results.

### 1.1.2 Phosphoenzyme intermediates

One of the most obvious ways in which phosphoryl transferases could facilitate the reaction is to first transfer the phosphate moiety to a functional group on the enzyme (such as a serine or histidine side chain) and then, in a second step to the final acceptor. The formation of such a covalent intermediate, analogous to the acyl-enzyme intermediate (11) of serine proteases (12), is an example of nucleophilic catalysis. For this to be effective the nucleophilic group on the enzyme must be both a better nucleophile and a better leaving group than the final nucleophile of the reaction. It turns out that while some phosphoryl transferases – such as the phospholipase D superfamily (13) or protein tyrosine phosphatases (14) – do rely on this type of catalysis, not all of them do.

The detection of phosphoenzyme intermediates is not always easy. If the lifetime of the intermediate is sufficient it can be isolated from a quenched reaction mixture, but if the intermediate is less stable, it has to be detected directly during catalytic turnover, for example by $^{31}P$ nuclear magnetic resonance spectroscopy (NMR) (15, 16). For example the transfer of a radioactive label from the substrate to the enzyme can prove or exclude (17) a phosphoenzyme intermediate. Perhaps the most sensitive probe for the presence of a phosphoenzyme intermediate along the reaction path is the stereochemistry of the reaction. A single nucleophilic substitution at the phosphorous (see section 1.1.3) typically results in an inversion of configuration, because the nucleophile approaches from the opposite side to the leaving group and the peripheral substituents stay in place. This is the case even in a dissociative mechanism if the intermediate remains bound in the enzyme active site, as discussed below. Another substitution, such as the hydrolysis of a phosphoenzyme intermediate causes another inversion, so that the overall reaction results in retention of configuration.

### 1.1.3 The elementary step

Regardless of how many phosphoryl transfer steps are involved, the next level of understanding the reaction is the mechanism of each elementary step, namely the structures of the respective transition states. Of most interest is the mechanism of the rate-determining step. As discussed above, only indirect information of the transition state is usually available. Kinetic isotope effects (18) are a particularly powerful tool to study the structure of the transition state. The potential energy surface of the reaction remains in principle identical upon isotopic substitution, but the zero-point energy of molecular vibrations changes, including the reaction coordinate (which can be considered a vibration with imaginary frequency). This means that the rate of reaction changes upon isotopic substition, and the magnitude of this change reflects the degree of bond breaking and formation in the transition state. Therefore any mechanistic postulate has to be consistent with the observed isotope effects.

Phosphorous, the reaction center in phosphoryl transferases, is a third period element and hence has available *d*-orbitals. It is therefore geometrically much more flexible than, for example, carbon. Carbon can have a maximum valency of four, and a pentacoordinate species is only possible as a fleeting transition state, not as a metastable intermediate. Phosphorous, on the other hand, can accept the extra electrons of a pentavalent state, expanding the possible reaction pathways. If the potential energy is plotted against the nucleophile-electrophile and electrophile-leaving group distances, known as a More O'Ferrall-Albery-Jencks plot (Figure 1.3), the reaction path can be traced along the line of the lowest potential energy.

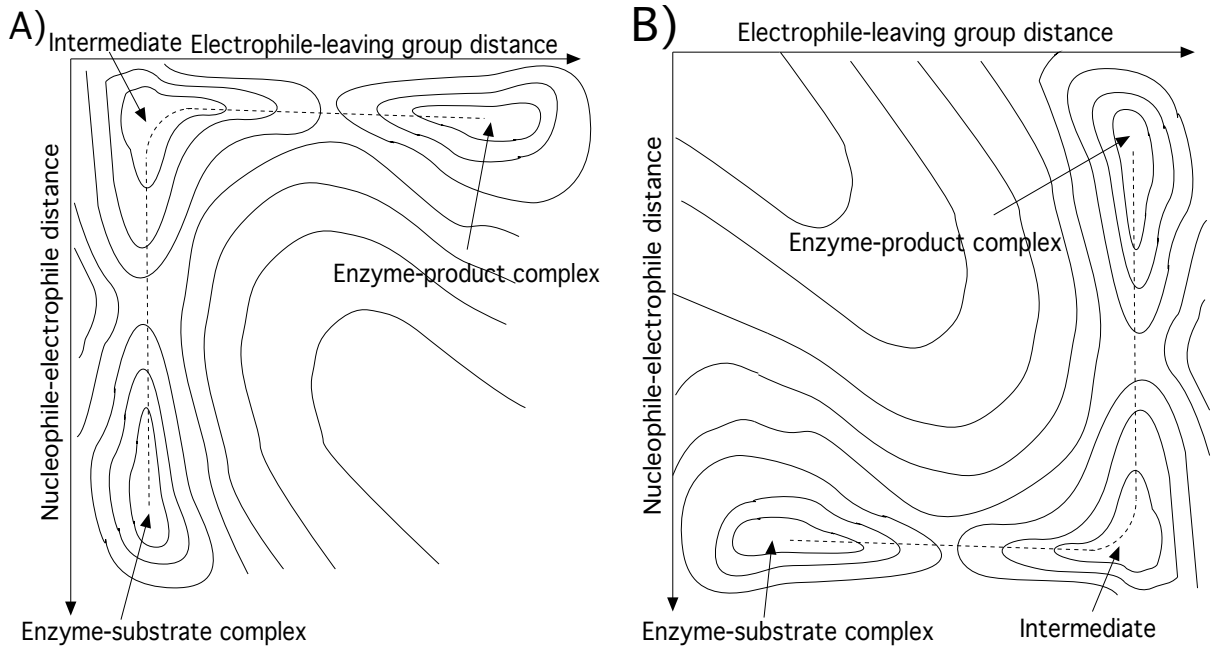


**Figure 1.3. The isocontours of potential energy as a function of the nucleophile-electrophile and electrophile-leaving group distances (More O'Ferrall-Albery-Jencks plot). Completely associative (A) and dissociative (B) pathways are shown.**

This is the simplest definition of a reaction coordinate for a nucleophilic substitution. The two extremes of possible mechanism are a completely associative pathway (Figure 1.3A), where the intermediate is an anionic pentacoordinate phosphorane, and the completely dissociative pathway (Figure 1.3B), where the leaving group leaves before the nucleophilic attack on the tricoordinate cationic metaphosphate. The symmetric $S_N2$-like pathway lies in between the two extremes. For the associative and dissociative extremes there is an added distinction whether there is actually a phosphorane or metaphosphate *intermediate* (Figure 1.4).

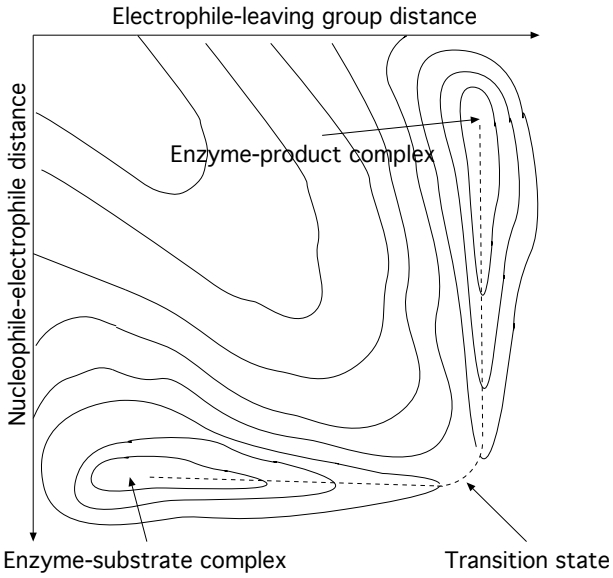


**Figure 1.4. A More O'Ferrall-Albery-Jencks plot of a dissociative pathway with no intermediate.**

In the dissociative case there is little difference between a true metaphosphate and an $S_N2$-type mechanism with longer nucleophile-electrophile and electrophile-leaving group distances, because, unlike in free solution, the metaphosphate is probably not free to rotate in the active site and is hence susceptible to nucleophilic attack only by a prepositioned nucleophile. Another complication arises from the geometry of a trigonal bipyramidal pentacoordinated phosphorous, where the apical and peripheral oxygens are not equivalent. If the intermediate is sufficiently long-lived, pseudorotation (Figure 1.5) may occur.

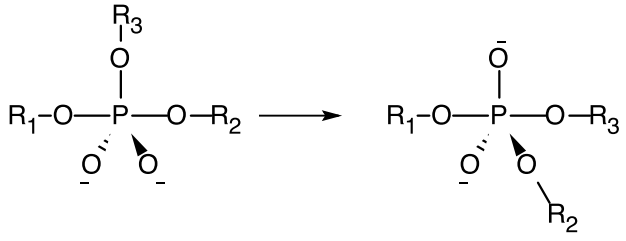


**Figure 1.5. In pseudorotation an apical substituent ($R_2$) moves into a peripheral position and a peripheral substituent ($R_3$) into an apical one.**

This leaves four classes of pathways (as in Knowles (2), Figure 1.6): the dissociative pathway (A) and the three associative pathways with a pentacoordinate transition state (B), a pentacoordinate intermediate (C) and a pentacoordinate intermediate with pseudorotation. The mechanisms of the elementary step in phosphoryl transferases can be described in terms of these pathways.

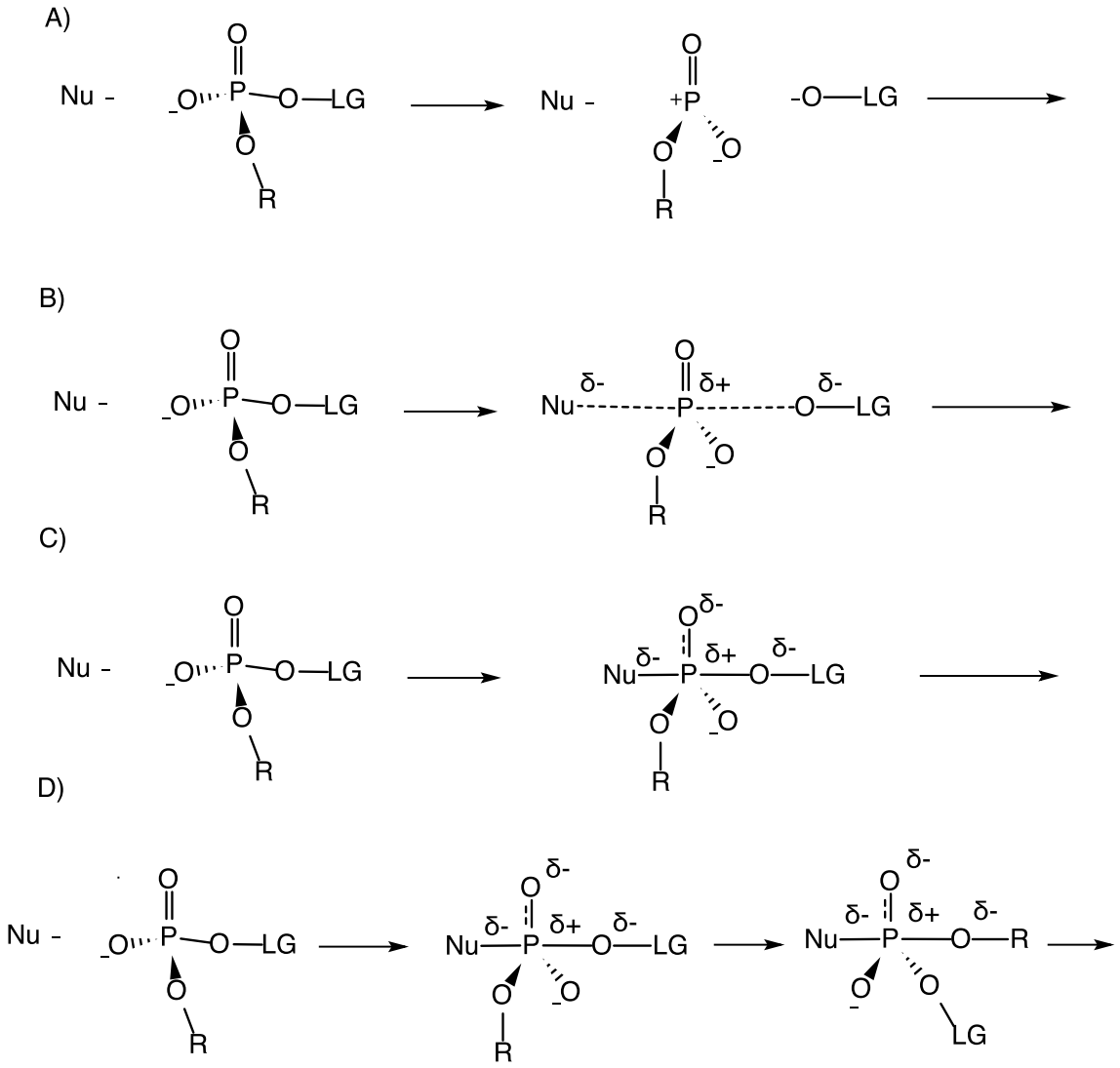


**Figure 1.6. The four possible pathways of phosphoryl transfer after Knowles (2): A) dissociative pathway with a metaphosphate intermediate; B) associative pathway with a pentacoordinate transition state; C) associative pathway with a phosphorane intermediate; D) associative pathway with a phosphorane intermediate and pseudorotation. Nu is the nucleophile and LG is the leaving group.**

A more quantitative approach to classifying the mechanism of the elementary step in the cases A and B above has been proposed by Mildvan (19). The bond order between two atoms can be calculated as proposed by Pauling (20) (Equation 1) if the single bond distance is known, such as 1.73 Å for P-O.

$$D(n) = D_1 - 0.71\ \text{Å} \log n$$

**Equation 1. The bond length D as a function of bond order n, where $D_1$ is the length of a single bond (n=1).**

Assuming, for simplicity, a symmetric transition state, only the nucleophile-electrophile bond order has to be considered. If the *axial distance* between the nucleophile and electrophile in the trigonal bipyramidal transition state is estimated and the bond order calculated, the fractional associativity is obtained. So a 100 % associative reaction would have an axial bond distance of 1.73 Å and and a 50 % associative one (the $S_N2$ case) would have an axial bond distance of 1.91 Å.

The problem with this approach is that we have no direct structural information about the transition state. We can therefore only estimate the bond distances from the available structural information, which may include structures of substrate, product or intermediate complexes. Since the transition state is a high-energy species compared to any of these complexes, it seems unlikely that the nucleophile would move further away from the electrophile when the reaction takes place. Therefore the overall distance between the nucleophile and the leaving group is presumably not longer in the transition state. If we then also assume the transition state to be symmetrical, the nucleophile-leaving group distance divided by two gives us a bond length that can be used to calculate a *minimum associativity* as above.

So how do enzymes lower the activation energy of the rate-limiting step so spectacularly? To answer this question we must first try and quantitate exactly *how* good catalysts they are. Obviously the comparison is to the uncatalysed reaction rate in solution, but the enzymatic reaction inevitably has a more complicated kinetic scheme due to the formation of the enzyme-substrate complex and the dissociation of the enzyme-product complex. The simplest way is to compare pseudo-first order reaction rates $k_{sol}$ (in solution) and $k_{cat}$ (from steady-state kinetics) extrapolated to the same conditions (19). The catalytic power is simply the ratio $k_{cat}/k_{sol}$, which for phosphoryl transferases can easily exceed $10^{15}$. Such rate enhancements are unlikely to result from any single factor or catalytic residue.

A major factor in any enzyme is simply bringing the substrates together in the right orientation. This is similar to surface catalysis of solution reactions in the sense that it reduces the degrees of freedom available for the substrates, thus increasing the probability of a ‘productive encounter’.

It has been postulated that the catalytic power of enzymes stems from being able to bind the transition state tighter than either the reactants or products. While this is probably true for most enzymes, it is not the whole truth (21). The transition state is not sufficiently different from the reactants or products in terms of electronic structure for the difference in binding energy to be large enough to account for the catalytic power (22); if the binding is too tight the product dissociation becomes rate limiting. According to Knowles (2) the key issue is to have $K_{int}$, the equilibrium constant between the enzyme-substrate complex and the enzyme-product complex as close to one as possible (Figure 1.7). This means that the potential energy barrier between the substrate and product is crossed in smaller steps.

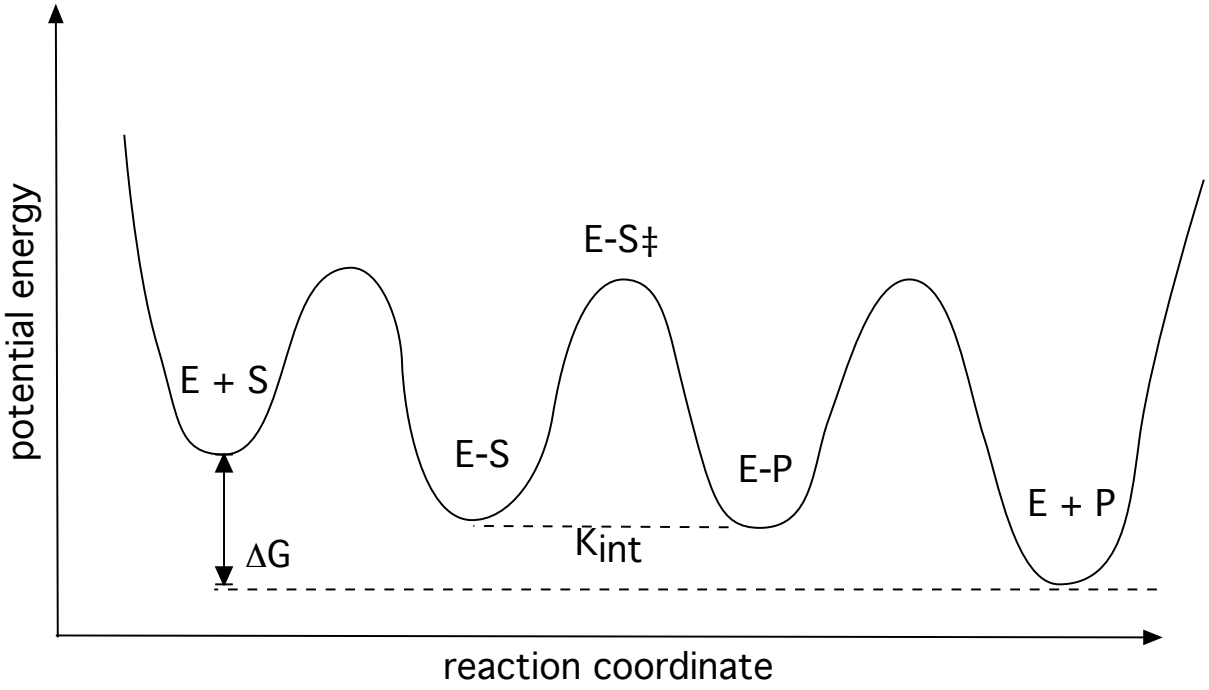


**Figure 1.7. A schematic depiction of the internal thermodynamics of enzyme catalysis.**

Another theory to explain part of the catalytic power of enzymes is conformational flexibility coupled to the reaction coordinate. Proteins, including enzymes, are highly dynamic molecules that exist in more than one conformation. These are in rapid equilibrium in solution. The relative energy of each conformation may change as a function of the environment, including substrate binding, thereby changing the position of the equilibrium. It is thus not necessarily so that substrate binding directly *causes* a conformational change, but that the substrate binds better to one conformation, which in turn increases the population of that conformation when the substrate is present. The change from substrate to product, the actual catalysis, is similar to other conformational changes in the sense that the conformation of the enzyme with the product is thermodynamically not far away. This is logical if one considers the multidimensional free energy ' landscape' of the reaction. It is likely to be very rugged since many conformations are available to the enzyme, but low energy pathways are available between them.

The enzyme may also have motions that correspond to the reaction coordinate – vibrations that may be identifiable, for example, by normal mode analysis. Such vibrations are not necessarily harmonic, since their amplitude is so large, but they 'push' the substrates along the reaction coordinate (22). One of the most studied examples is dihydrofolate reductase (DHFR), where changes in the main chain flexibility (achieved by mutating glycine to more rigid residues) far from the active site significantly affect reaction rates (23).

These proposals to explain the catalytic power of enzymes have been questioned by Warshel (24, 25), who argues based on a re-interpretation of kinetic isotope effects (26) and computational studies (27) that enzyme catalysis is based on electrostatics and that other effects are negligible. While the details of this rather

involved debate are beyond the scope of this dissertation, they illustrate that enzyme catalysis remains an incompletely understood subject.

### 1.1.4 Means of catalysis

There are various ways to facilitate a phosphoryl transfer (Figure 1.8). The rate of a nucleophilic substitution obviously increases if the nucleophile is activated by general base catalysis. This does not necessarily mean complete deprotonation; just donating electron density to the proton allows for more electron density in the nucleophile's attacking lone pair orbitals. General bases in enzymes are typically histidine imidazoles or carboxylates from aspartate or glutamate. The p$K_a$ of the general base can be estimated from the pH-dependence of the reaction rate; carboxylates acting as general bases often have abnormal p$K_a$s (28). Given that linear free energy relationships (LFERs) for solution reactions suggest that phosphoryl transfer is not very sensitive to nucleophilicity (29), the effect of such groups may not be entirely due to general base catalysis. In addition to nucleophile activation, they also position and orient the nucleophile for attack. This interpretation of the LFERs has also been questioned (26) by the Warshel group (see above).

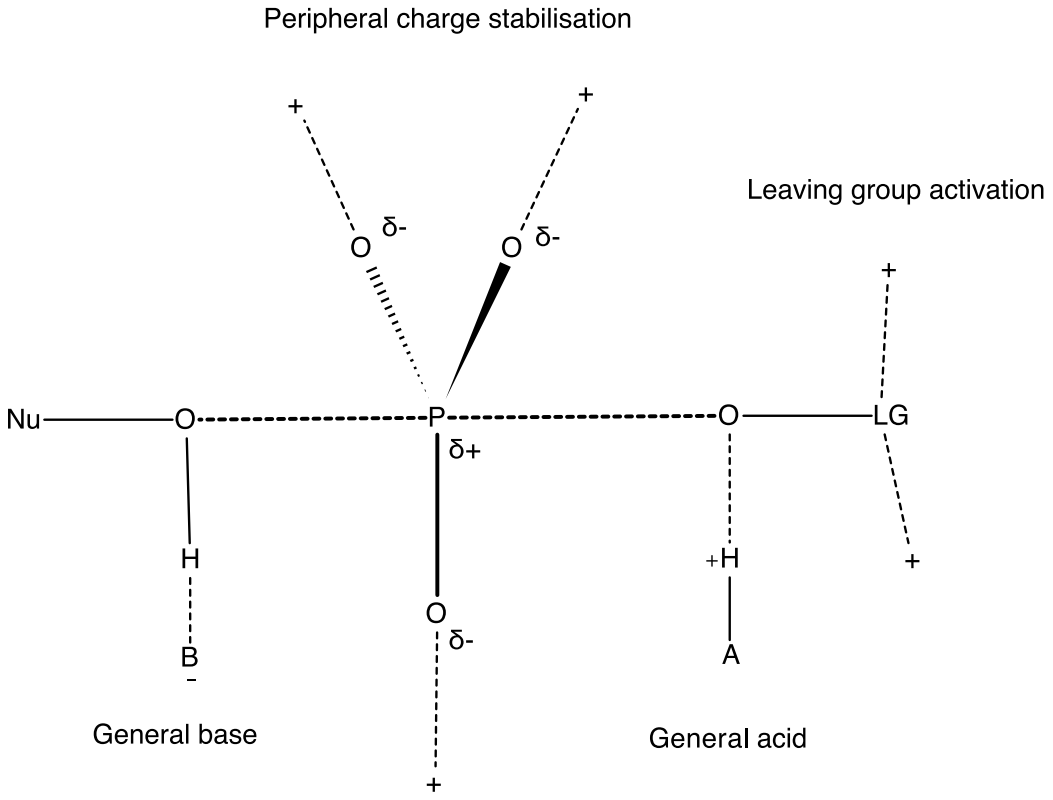


**Figure 1.8. The means to catalyse the elementary step in the case of phosphoanhydride hydrolysis: general base and acid, balancing the charge of the peripheral oxygens and leaving group activation.**

Similarly, general acid catalysis does not necessarily mean full protonation of the leaving oxygen. Even partial donation of a proton helps to compensate the

charge build-up. General acid catalysts are similar groups to general bases, which is understandable because both have to have p$K_a$s close to the 'effective pH' of the active site. Since the active sites are usually well ordered and excluded from bulk solution, the electrostatic environment is not very similar to that of bulk water. Consequently, changes the protonation equilibria of the protolysable groups can be very different from those on the protein surface. Water molecules coordinating metal ions such as $Zn^{2+}$ can also be acidic enough to act as general acid catalysts.

A slightly different effect is achieved by compensating the charge of the peripheral oxygens in the trigonal bipyramidal transition state (Figure 1.8). The electrostatics of the transition state depends very much on the nature of the nucleophile, since the leaving group will be negatively charged anyway unless it is completely protonated by a general acid. If the nucleophile has a formal negative charge, the build-up of charge in the transition state is quite significant. It is then advantageous to compensate this charge by positively charged residues and/or hydrogen bond donation. The localisation of charge also depends on the associativity of the transition state. The more associative it is, the more negative charge is localised on the peripheral oxygens. Another way of understanding this mode of catalysis is that withdrawing electrons from the peripheral oxygens withdraws in turn the electrons from the electrophilic phosphorous, thus making it a better electrophile. In enzymes this is usually accomplished by positively charged lysine or arginine residues or metal ions.

Also the leaving group can be activated outside the actual leaving oxygen by general acid catalysis. This applies particularly to phosphoanhydride hydrolases, where the leaving group is a phosphate moiety. Compensating negative charges by hydrogen bond donation makes it a better leaving group.

Metal ions are almost always present in phosphoryl transferases, as indeed in many other enzymes that recognise a phosphate moiety. Their role can be to simply bind and orient the substrate or they may have a more direct role in nucleophile activation. The nucleophile is often a water molecule bridging two metal ions (30), as for example in DNA polymerases (31). It is interesting to note that the change of the metal ion from the native $Mg^{2+}$ to $Mn^{2+}$ also changes the specificity of the polymerase. This probably results from the increased flexibility in coordination geometry for $Mn^{2+}$ which, unlike $Mg^{2+}$, has accessible *d*-orbitals.

### 1.1.5 Associative *vs.* dissociative

Phosphoryl transfer reactions in solution are generally thought to have dissociative transition states. This is based on LFER data showing that the reaction rates are not very sensitive to the nucleophilicity of the incoming nucleophile, but the effect of the leaving group acidity is significant. In enzymes the situation is more complicated. LFER studies are difficult to perform in enzymes due to their specificity for certain substrates, so the comparison to solution data is not straightforward (26, 32, 33).

While there is almost certainly wide variation in the fractional associativity of phosphoryl transferases, the general trend appears to be that the structural

information points mostly towards associative mechanisms (19). The biophysical data suggest more dissociative mechanisms (29), although a consensus over the interpretation of the biophysical experiments is lacking. In phosphoryl transferases a truly dissociative pathway is hard to conceive, as all the reactants typically remain bound in the active site during the chemical step. But the degree of associativity as defined in section 1.1.3 is difficult to determine, as the structure of the transition state is not known. The method suggested by Mildvan (19) (section 1.1.3) gives a lower limit to the associativity of the mechanism, given that some reasonable model of the substrate complex exists. There are several structures of phosphoryl transferases with a mimic of a pentacovalent intermediate such as vanadate bound in the active site. A classic example is ribonuclease A, discussed in more detail below (section 1.5.4). In the structure of β-phosphoglucomutase a trigonal bipyramidal phosphorane intermediate is even observed directly (34). In this case the 1.2 Å resolution allowed the authors to exclude the possibility of this being an artefact of *e.g.* multiple conformations.

Crystal structures of dissociative, metaphosphate-like transition state analogues are also known. A typical analogue in this case is the nitrate ion, which is trigonal planar and roughly the same size as metaphosphate. Structures of arginine kinase (35) and creatine kinase (36) in the presence of the ADP, nitrate and arginine or creatine respectively showed that the nitrate is perfectly positioned to mimic the γ-phosphate of ATP – the true substrate – for an in-line attack of the nitrogen nucleophile. In these cases, too, the nitrate and the ADP β-phosphate were bridged by a $Mg^{2+}$-ion and held tightly in place, in contrast to a truly dissociative pathway with free metaphosphate.

## 1.2 PYROPHOSPHATASES

Pyrophosphatases (PPases, EC 3.6.1.1) are essential enzymes for any living cell. They hydrolyse inorganic pyrophosphate ($PP_i$) that is produced in a huge number of biochemical reactions in the cell. Many reactions that would otherwise be thermodynamically unfavourable, such as DNA or protein biosynthesis, are driven by the release and consequent hydrolysis of $PP_i$, which has a large and negative standard free energy difference of 17 kJ/mol (37).

PPases can be divided into three families: soluble Family I (38, 39) and II PPases (40) and the integral membrane PPases (41). As the soluble PPases have more or less a simple ‘ housekeeping’ function to hydrolyse all available pyrophosphate, they are usually not under much regulation. The exceptions are some plant PPases, which are involved for example in pollen tube elongation (42). Since plants also have vacuolar membrane bound PPases, the entire responsibility for $PP_i$ hydrolysis does not rest on the soluble PPase. The membrane bound PPases on the other hand have quite a different function; they use the energy of $PP_i$ hydrolysis to pump protons or sodium ions (43, 44) across a membrane against an electrochemical gradient.

All eukaryotes as well as most eubacteria and archaea have a Family I soluble PPase. Family II PPases are limited to prokaryotes with the exception of the protozoan *Giardia lamblia*. The two families have no similarity at all either in sequence or structure, but the active sites do have a number of similarities.

All PPases use divalent metal ions as cofactors; the metal ion preferences of the soluble PPases are discussed below.

### 1.2.1 Family I PPases

The Family I PPases are by far the best-studied PPases. PPase activity was first identified in animal tissues in 1928 (45) and the pure enzyme was isolated from baker's yeast (*Saccharomyces cerevisiae*) in 1952 (38); since then a wealth of kinetic information has accumulated, especially on the *S. cerevisiae* (46-48) and *Escherichia coli* (49, 50) enzymes, referred to below as Y-PPase and E-PPase respectively. It was established early on that no phosphoenzyme intermediate is formed (51) and that though the native cofactor is $Mg^{2+}$, $Mn^{2+}$, $Co^{2+}$ and $Zn^{2+}$ can also activate the enzyme (52). $Ca^{2+}$ on the other hand is a competitive inhibitor. The enzyme is very specific for $PP_i$ in the presence of $Mg^{2+}$ (17). Because the reaction can be easily run in the direction of $PP_i$ synthesis, the internal equilibrium constant $K_{int}$ could be evaluated.

Structural studies were started early on (53), and the first high-resolution structure was solved in 1981 (54). It did not, however, contain the metal cofactors or the product. The first structure of the E-PPase product complex appeared in 1994 (55). Since then numerous structures of family I PPases have been published (56-60), including a number of active site variants (61, 62).

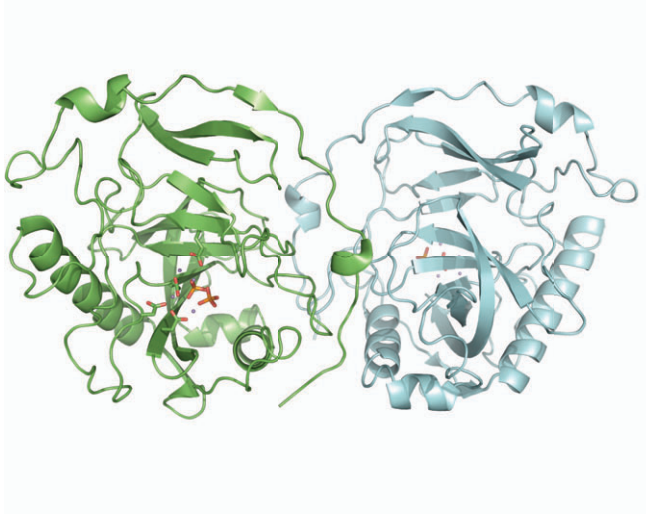

**Figure 1.9. The Y-PPase dimer from the product complex structure (PDB-ID 1E9G (56)) represented as a cartoon. α-helices are shown as ribbons and β-sheets as arrows. The active site residues E58, D115, D120 and D152, as well as the product phosphates are shown as sticks. The $Mn^{2+}$-ions are shown in magenta.**

The overall fold of Family I PPases is a twisted five-stranded β-barrel flanked by two a-helices (Figure 1.9). The active site is located in a groove on the surface.

Family I PPases are similar enough to each other in sequence to be easily identified by similarity searches and their structures are very similar. The E-PPase sequence for example is shorter and lacks some loops of Y-PPase, so the sequence identity between Y-PPase and E-PPase is only 22-28 % depending on the choice of the alignment parameters (63). The structures are essentially identical for the consensus part, though, with a $C_\alpha$ root mean square deviation (r.m.s.d.) 1.45 Å (57). Also the conservation of the active site residues is quite strict (Figure 1.10). The quarternary structure on the other hand differs significantly; while eukaryotic PPases (such as Y-PPase) are obligate dimers (57), prokaryotic PPases (like E-PPase) are hexamers with 32 (D3 in the Schönflies notation) symmetry (55, 64).

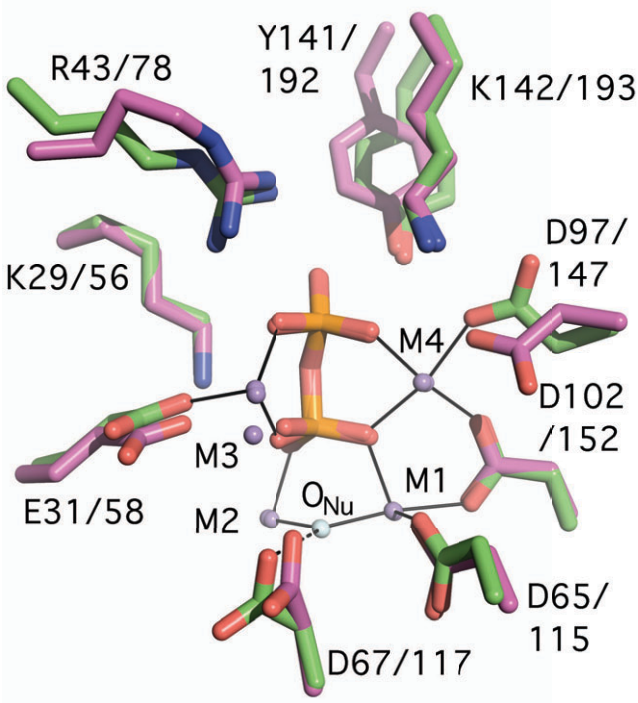


**Figure 1.10. A superposition of the E-PPase (magenta) and Y-PPase (green) active sites shows a remarkable similarity.**

At the optimal pH of 6.5-7 (48) three metal ions are required for catalysis, two of which are bound to the enzyme throughout the catalytic cycle (Figure 1.11) while the third comes in with the substrate (46). At lower pH (< 6), where most of the crystals are grown, a four-metal-ion mechanism dominates (48). The metal ions are labelled in the order of binding affinity; M1-2 are permanently enzyme-bound and M3-4 come in with the substrate. Prokaryotic PPases contain an additional metal ion that has a purely structural role; it binds between the monomers so that there are three per hexamer.

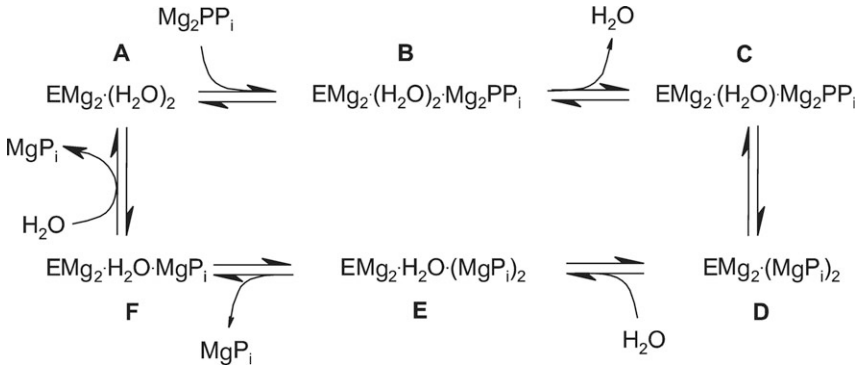


**Figure 1.11. The catalytic cycle of Y-PPase with the intermediates detected by kinetic studies.**

There have been several postulates for the nucleophile of the reaction (65, 66); by now it has been reasonably well established that the nucleophile is located between M1 and M2 (56, 57). The nucleophile appears to be a hydroxide ion, as discussed below.

After substrate binding Y-PPase undergoes a conformational change before chemical catalysis occurs. In the ‘resting enzyme’ structure, two water molecules bridge the two permanently bound metal ions, whereas in the product complex there is just one (Figure 1.12) (see Results section 3.1.). The atomic resolution structure of the product complex contains two alternative conformations for the product phosphates; the immediate product, where the nucleophilic oxygen has become a phosphate oxygen but still occupies the same position between the M1 and M2 and the relaxed product where the phosphates have rotated (56).

Fluoride is a potent inhibitor, especially of Y-PPase (67), because it is isoelectronic with the hydroxide ion nucleophile. The fluoride inhibition of E-PPase and Y-PPase is intriguingly different; for E-PPase it is rapid and reversible, but for Y-PPase there is first a fast phase followed by a slower phase to produce a fluoride complex that can even be isolated chromatographically (67).

Isotope exchange experiments showed that in the presence of $Mg^{2+}$, the electrophilic phosphate is the first one to be released and second one to bind (46). It is hence less tightly bound; therefore it is called P2. The more tightly bound leaving group phosphate is labelled P1. In the presence of $Mn^{2+}$, however, the order is reversed (68) and the release of product instead of the chemical step is rate limiting.

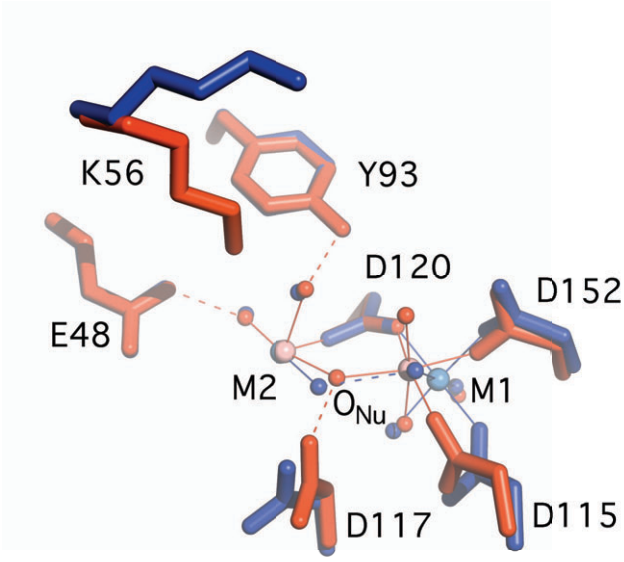


**Figure 1.12. The superposition of the open conformation (blue) with the two-water bridge, and the closed conformation (red) with the one-water bridge, of Y-PPase.**

The pH-dependence of the reaction revealed two important ionisation constants: a general acid at a p$K_a$ of 7.5 and an essential base at 5.9 (69). The identity of these ionisable groups is not clear, but in the reverse reaction of pyrophosphate synthesis, the p$K_a$ of the essential base is lowered to 5.5. This would be consistent with a water molecule, although in a rather unusual environment. A similar change in p$K_a$ also occurs after the substrate binds and the enzyme conformation changes (48)(Figure 1.12). This issue is discussed in more detail in section 4.1.1. It was postulated that a low barrier hydrogen bond (LBHB) between the nucleophile and D117 is involved in the activation of the nucleophile (56).

### 1.2.2 Family II PPases

The discovery of family II PPases (40, 70) was somewhat surprising, as all PPases were thought to be very similar in terms of sequence, structure and mechanism. They are now known to belong to the 'DHH' superfamily of phosphoesterases that share a DHH catalytic motif (SCOP classfication 64181 (71)). While Family I PPases have only one domain, as discussed above, Family II PPases have two (Figure 1.13) (72). Despite the lack of structural similarity in terms of the fold, the active site residues are positioned in a similar fashion, which was initially interpreted as an example of convergent evolution.

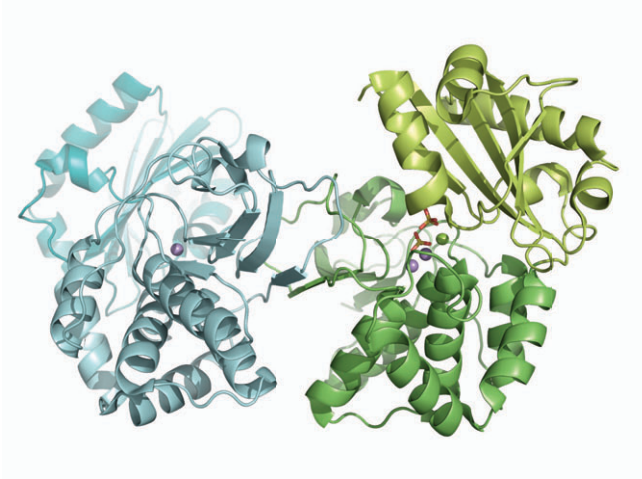

**Figure 1.13. A cartoon representation of a *Streptococcus mutans* pyrophosphatase dimer, showing the active site containing two $Mn^{2+}$-ions (magenta), a $Mg^{2+}$-ion (green) and two sulphate ions located between the N-terminal (green) and C-terminal (lime) domains.**

The active site is located at the interface of the domains; the tightly bound metal cofactors bind to the N-terminal domain but substrate binding requires residues from both domains. A number of structures of Family II PPases have been solved since 2001 (72-75). Depending on the contents of the active site, the orientation of the two domains varies widely. In the absence of product or substrate they are wide open in an almost linear configuration (PDB-ID 1K23 (74)), whereas in the presence of product they adopt a more closed orientation (Figure 1.14). This has led to the suggestion that substrate binding triggers the closure of the two domains, and after the chemical step they re-open, separating the two product phosphates (73).

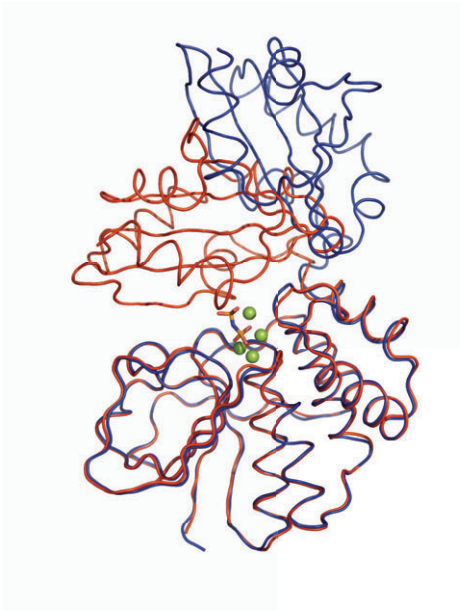

**Figure 1.14. A superposition of the N-terminal domains of the open (blue) and closed (red) structures of *Bacillus subtilis* PPase (BsPPase). The active site metals of the closed structure are shown in green.**

The metal activation of Family II PPases is rather more complex than in Family I. With $Mn^{2+}$, the native cofactor the catalytic rates ($k_{cat}$) are ten times those of Family I PPases, but with $Mg^{2+}$ the rates are comparable to those of Family I enzymes (76). However, even the $Mn^{2+}$ activated enzyme makes use of $Mg^{2+}$ that comes in with the substrate. In the presence of substrate there are four metal ions present in the active site (75). The tightest binding metal is actually denoted M2 in the structures, due to initial comparison to the Family I active site (Figure 1.15) (72). $Mn^{2+}$ binds to this site with nanomolar affinity, whereas $Mg^{2+}$ binding is in the micromolar range (77). Binding to the second site (M1) is even weaker, in the millimolar range (74).

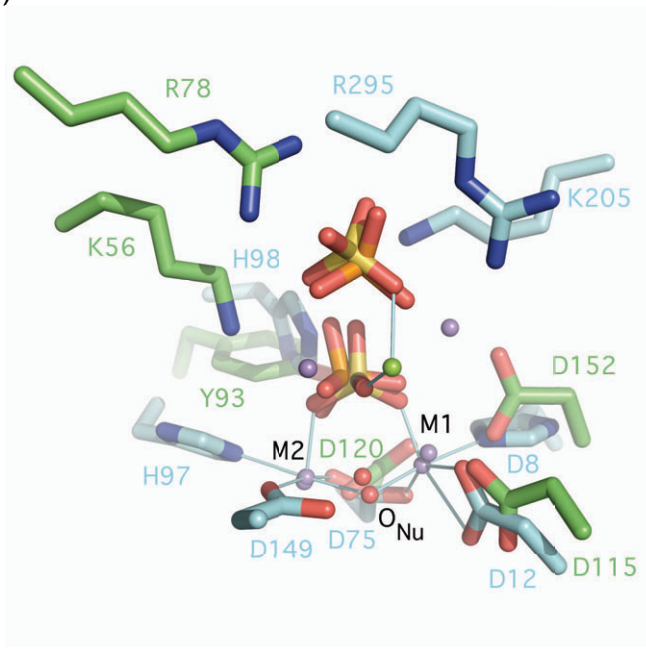


**Figure 1.15. The active sites of the Y-PPase product complex (PDB-ID 1E9G, green (56)) and S. mutans PPase sulphate complex (PDB-ID 1I74, cyan (72)) superposed by the phosphates and metals M1 and M2. $Mn^{2+}$-ions are shown in magenta and a $Mg^{2+}$-ion in green.**

The preference of Family II PPases for $Mn^{2+}$ over $Mg^{2+}$ is at least partly explained by the presence of histidines as opposed to aspartates as metal-binding ligands. The $Mg^{2+}$-ion is a relatively hard electrophile that prefers harder, oxygen nucleophiles such as water or carboxylates, wheras the imidazole group of histidine is a softer nucleophile and therefore preferred by softer electrophiles such as $Mn^{2+}$ or $Co^{2+}$ (78). There is also a more subtle reason for the $Mn^{2+}$-preference: the metal coordination geometry. In the absence of bound substrate (or product) the M2 metal ion tends to be only five-coordinated and in a distorted geometry somewhere between trigonal bipyramidal and square planar (74). Upon substrate binding, the coordination number changes to six and the geometry to octahedral. This moves the nucleophile into a better position for attack and additionally causes some conformational strain, which presumably helps in product release. Therefore $Mg^{2+}$, which can only adopt an octahedral geometry, binds poorly compared to transition metal ions, which can tolerate larger deviations of the ideal geometry due to their available *d*-orbitals.

$Zn^{2+}$, on the other hand, binds similarly to $Mn^{2+}$, but displays interesting activation behaviour. $Zn^{2+}$ activates the enzyme, but only when bound to the lower affinity M1 site, because it, in turn, cannot assume the octahedral geometry required for catalysis when the substrate binds (74).

## 1.3 METHODS TO STUDY REACTION INTERMEDIATES BY CRYSTALLOGRAPHY

Crystallography is often mistakenly believed to provide only static information on protein structure. Even though a crystal structure is an average over all the molecules in the crystal and the measurement time, it does not imply that the protein molecules in the crystals are either static or all in the same conformation during the measurement. Indeed a protein crystal is more like an ordered jelly than a true solid state. This flexibility combined with the high solvent content of protein crystals means that enzymes are sometimes active in crystals and the enzymatic reaction can be followed *in crystallo* (79).

The aforementioned constraints of time and space averaging require that data can be collected within the lifetime of the intermediate of interest and that the reaction is well *synchronised* in the crystal (*i.e.* all molecules are in step). The first requirement can be fulfilled either be collecting the data very fast with the polychromatic Laue method or by somehow extending the lifetime of the intermediate (80).

The latter requirement is a result of the space averaging in crystallography – the electron density maps represent the mean over all molecules in the crystal. Even if the data could be collected faster than the intermediate lifetime, the information will be averaged out if the molecules in the crystal are not simultaneously in the same phase of the reaction.

### 1.3.1 Initiating a reaction in the crystal

The requirement of synchronisation means in practice that the reaction must be initiated an order of magnitude faster than the actual reaction rate. The process of growing a diffraction quality protein crystal takes from hours to months, whereas even slow enzymatic reactions happen in seconds. (Exceptions include clock enzymes that can have time constants of $10^{-3}$-$10^{-4}$ $s^{-1}$ (81).) It is therefore clear that the substrate cannot be included in the crystallisation conditions. The classical way to introduce small molecules into protein crystals is soaking. Protein crystals have a high solvent content and the channels of disordered solvent allow almost unhindered diffusion of small molecules into the enzyme active site. This diffusion process typically happens in seconds, which means that initiating a reaction by simply soaking in the substrate only works for very slow reactions, or when an intermediate accumulates, like compound I in the case of cytochrome c peroxidase (82).

For faster reactions an inactive precursor of the substrate has to be co-crystallised with the enzyme and then somehow activated sufficiently rapidly. In simple organic reactions temperature, pressure or pH jumps are usually used, but

even though a pH jump approach was used with trypsin (83), proteins are generally stable over such a limited range of pH and temperature that such jumps are not practical. Laser photolysis of photolabile compounds typically occurs very rapidly, on the microsecond or faster timescales (84, 85). A key requirement for photolytic triggering is a sufficient quantum yield to convert all the photolabile precursor molecules to substrate. The extremely high brilliance of laser radiation allows this to happen in the protein crystal, where the concentration of the protein is in the millimolar range. There is usually only one substrate molecule per protein so the concentration of the photolysable molecule and the protein are the same.

"Caged compounds" containing a nitrophenylethyl chromophore (86) (Figure 1.16) have been used successfully in triggering reactions (87) involving *e.g.* nucleotides (86). The 1-(2-nitrophenyl)-ethyl moiety absorbs at 265 nm, the fourth harmonic frequency of the neodymium-YAG laser. The electronic excitation allows an intramolecularly assisted release of the pyrophosphate group. The rate of photolysis is pH dependent, as the protonation state of the phosphate moiety affects the probability of its release. At physiological pH, the time scale of the photolysis is a few milliseconds.

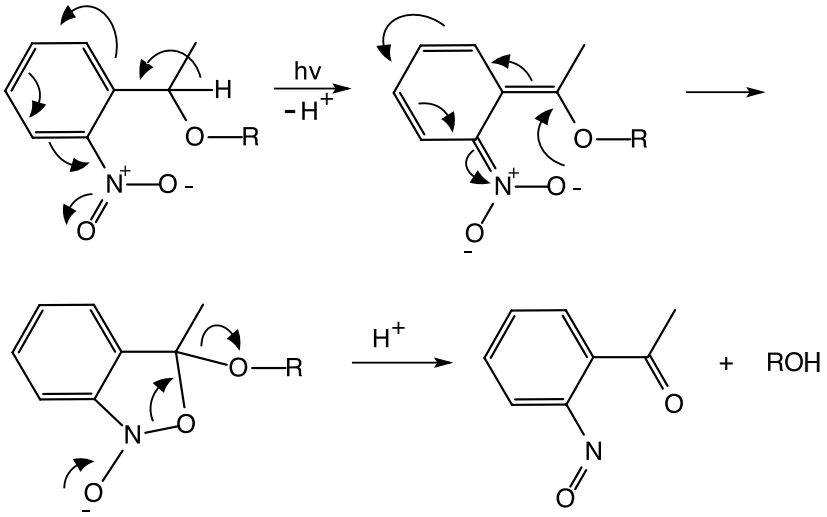


**Figure 1.16. The photolytic mechanism of *o*-nitrophenylethyl caging groups.**

The actual photochemistry is much faster than this, but the dark reactions are rate limiting (88). Cage groups with faster photolysis rates have been used for example for studying ion channels (89-92), but they tend to be bulkier and hence less suitable to be bound in the active site. It is also possible to photo-cage the active site residues themselves by some rather elaborate molecular biology (93), but so far this approach has not been used for time-resolved crystallography.

### 1.3.2 Following the reaction *in crystallo*

In order to observe the structural changes, it is essential to identify the intermediates formed as a function of time and link them to those observed in free solution. Optical spectroscopy is the method of choice, since the measurement is

relatively easy to perform both in solution and in the crystal during X-ray data collection. Protein crystals used for X-ray diffraction are typically less than one mm in any dimension, much smaller than the cuvette in a typical optical spectrophotometer. The spectrum acquisition has to be done *while* the X-ray data are being collected at a synchrotron beamline (94, 95). Thus *in crystallo* spectroscopy requires specific equipment. Dedicated microspectrophotometers have been constructed and installed at laboratory sources or synchrotron beamlines. Obviously the first requirement for *in crystallo* spectroscopy is a suitable spectroscopic probe for the progress of the reaction in the UV-visible region. The majority of enzymatic reactions, however, lack such a probe. In these cases the use of Raman spectroscopy may provide a viable alternative, even though the time resolution is rather more modest (96).

### 1.3.3 Time resolved Laue crystallography

In regular protein crystallography the monochromatic oscillation method (97) has become almost the sole method of data collection. In each diffraction image only those reflections that fall on the Ewald sphere during the oscillation are recorded on the diffraction image. In order to cover all of reciprocal space the crystal has to be rotated and even at the fastest modern synchrotron beamlines this takes several minutes (98), which is orders of magnitude slower than the timescale of most enzymatic reactions.

In the Laue method the incident radiation is not monochromatised, so the Ewald sphere is replaced by a much larger volume (Figure 1.17) and a far larger number of reflections are simultaneously in diffracting position. Hence many more reflections can be measured on a single image. In principle an entire diffraction data set can be collected on a single image, albeit at reduced completeness.

The Laue method requires highly brilliant polychromatic synchrotron radiation. Compared to the monochromatic oscillation method where the majority of the X-ray photons are rejected by the monochromator, the incident X-ray intensity is significantly higher, which allows exposure times as low as nanoseconds (99). Therefore reactions in the crystal can be followed in real time, solving the structure from each Laue image.

Time resolved Laue crystallography is technically very challenging, requiring good, low mosaicity crystals that retain their quality even when a reaction is initiated in the crystal. With so many reflections on one image overlaps are unavoidable. There are both harmonic overlaps, where the reflection of the indices $h$, $k$, $l$ at wavelength $\lambda$ is overlapped for example by a reflection $2h$, $2k$, $2l$ at $2\lambda$, and spatial overlaps that occur by coincidence. Separating the spots on the detector surface and integrating their intensities is a non-trivial task compared to monochromatic data processing. If the mosaicity of the crystal increases this becomes even more difficult.

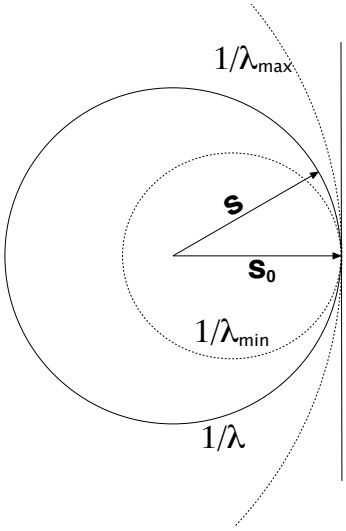


**Figure 1.17 The Ewald construction for Laue diffraction. The monochromatic Ewald construction is depicted for one wavelength λ, where $s_0$ and s are the incident and scattered wavevectors, respectively. In the Laue method all reflections that fall between the Ewald spheres of the shortest ($\lambda_{min}$) and longest ($\lambda_{max}$) wavelengths are in diffracting position simultaneously.**

The number of proteins studied by time-resolved Laue crystallography remains small due to the technical challenges outlined above (see also Ren *et al.* (100)). Perhaps the most spectacular example is myoglobin, where the photolysis of a carbon monoxide complex and subsequent migration of the CO has been followed with nanosecond resolution (99). Enzymes that have been studied this way include glycogen phosphorylase (101), the GTPase Ha-ras p21 (102), α-chymotrypsin (103), isocitrate dehydrogenase (104) and hydroxymethylbilane synthase (105).
The Laue method is inherently plagued with low completeness, both because not all of reciprocal space is in diffracting position and because of the overlap problem. This low completeness adversely affects the quality of the electron density maps, making their interpretation difficult. This problem has been partly addressed by a singular value decomposition approach in which the electron density is separated to time-dependent and time-independent components. Since the interesting changes are in the time-dependent component, the time-independent component can be considered essentially background. Therefore the time-dependent electron density is easier to interpret. This method has been applied for example with photoactive yellow protein (106).

### 1.3.4 Physical trapping — freeze trapping and trap freezing intermediates

The temperature dependence of enzymatic reactions is very steep, so that a drastic drop in temperature effectively stops the reaction. This provides a convenient way of extending the lifetime of intermediates almost indefinitely. Since X-ray data are nowadays almost always collected at 100 K to avoid radiation damage, it is natural to cryo-cool the crystal in the middle of the reaction. It is crucial to know the extent of reaction at the time of cryo-cooling; spectroscopic observation as discussed

in section 1.3.2 is required. This approach is known as trap-freezing. It was applied for example with the hammerhead ribozyme, where the reaction is comparatively slow and can be initiated with a change in pH or $Mg^{2+}$ concentration (107). The time resolution in trap-freezing is also limited by the cooling rate of the crystal; it can take as long as one second to cool a large protein crystal to 77 K (108).

The second option is freeze-trapping, in which the reaction is initiated photochemically in the cryo-cooled crystal (87). This obviously requires that the reaction rate is fast enough to observe the reaction at the cryo-temperature but slow enough for intermediates to accumulate.

### 1.3.5 Chemical trapping — mimicking intermediates

As discussed in section 1.1, the lifetime of the transition state of a chemical reaction is so short that it can only be directly observed by femtosecond laser spectroscopy (109), which is not yet practical for enzyme systems. Also, the reaction intermediates are very short-lived compared to the timescales involved in experimental structure determination. However, the transition state must lie between the structures of the substrate and the product, so if the structures of the intermediates are known, it is possible to postulate what its structure should be like and synthesise molecules that mimic the transition state (110). Such molecules should not react, so that experimental structure determination is possible.

Analogues of substrates, products or intermediates of phosphoryl transferases are often phosphonates or phosphoramidates, which are geometrically and electrostatically quite similar to phosphate esters or anhydrides. Such compounds are helpful in elucidating the molecular choreography as discussed for Y-PPase in section 3.1, but they do not really address the question of the transition state. Transition state analogues for associative phosphoryl transfer reactions should have a similar geometry to a pentacoordinate phosphorous; mimics of dissociative intermediates should be trigonal planar so that the axial ligands are not actually covalently bonded. Typical mimics of an associative transition state are vanadates (111, 112), which feature a pentacovalent geometry and negatively charged peripheral oxygens. The planar tetrafluoroaluminate ion ($AlF_4^-$) has also been used as a dissociative transition state analogue, the fluorine atoms mimicking the charged peripheral oxygens (113). A common mimic of a dissociative transition state is the nitrate ion in combination with some analogues of the nucleophile and leaving group (35, 114).

## 1.4 Methods to study protonation states

Hydrogen has only one proton and one electron. As the X-rays that are used for structure determination by crystallography are scattered from electrons, the contribution of hydrogens to the electron density maps is so small that they can be observed only in the very highest resolution X-ray structures (115). Yet the positions

of hydrogens – namely the protonation states of acids and bases – are crucial to understanding any reaction mechanism.

Neutron crystallography (see section 1.5) is the most unambiguous method for protonation state determination. The technical difficulties are so great, however, that if alternative methods can provide the same information, they are often preferred. In small molecule crystallography, hydrogen atoms are routinely refined as part of the structure (116), but this is rarely done with proteins. What kind of X-ray data are actually needed to observe the hydrogens? The question is not easy to answer (117), as the interpretation of electron density near the noise level is always somewhat ambiguous. In crystal structures at sub-atomic resolution (< 0.9 Å) a large number of hydrogens can be seen in the $F_o$-$F_c$ difference maps. The peak height of a hydrogen (just like any other atom) in an electron density map is strongly dependent on the atomic displacement factor or B-factor. As the hydrogens 'ride' (118) with the heavier atom they are bound to, the B-factor of that atom largely defines whether the hydrogen is visible in the map or not.

An indirect way to observe the protonation state of a functional group – most commonly carboxylate – is the analysis of bond lengths. In a deprotonated carboxylate the C-$O^-$ bond lengths are equal (1.249 Å) (119) while protonation of one oxygen leads to a single C-OH bond of 1.304 Å and a double C=O bond of 1.208 Å. In crystallographic refinement, bond lengths are usually restrained to one of these values since the data-to-parameter ratio is insufficient to refine all atoms independently. Therefore the analysis of bond lengths is only useful at atomic resolution (< 1.2 Å) when unrestrained refinement is possible (120). In order to decide whether the differences in bond lengths are statistically significant, a full matrix inversion (121) has to be performed to estimate the standard uncertainties in the atomic coordinates. This is only feasible at atomic resolution and may present computational problems if more than ~1000 atoms are involved (121).

Another possible technique to determine protonation states directly is nuclear magnetic resonance spectroscopy (NMR) (122, 123). NMR spectroscopy provides information about the electronic environment of nuclei, which tends to change as a function of protonation. Measuring and interpreting NMR spectra from protein samples, however, is far from trivial. Perhaps the most significant limitation is the relaxation rate of the nuclei, which is related to the tumbling rate of the molecule in solution. If the tumbling is too slow the NMR signals cannot be observed in practice. An additional complication is that the NMR-active isotopes of *e.g.* carbon and nitrogen ($^{13}C$ and $^{15}N$) constitute only a small fraction of the atoms in the sample, so that elaborate isotope labelling schemes are often required to measure useful spectra. The tumbling rate in turn is related to the size of the protein, imposing a practical size limit. Even though NMR experiments typically observe protons, the proton signals from exchangeable protons are often broad and hard to observe. Protonation also changes the resonances of carbon and nitrogen atoms down the side chain (124, 125), which can be more easily observed by for example heteronuclear single quantum coherence (HSQC) spectroscopy. The change in the signals as a function of pH gives the p*K*a; this is known as NMR titration (126). The

disadvantage of this technique is that it cannot observe the protonation states of ordered water or ligand (substrate, product or analogue) molecules.

## 1.5 NEUTRON CRYSTALLOGRAPHY

In contrast to X-rays that are scattered from the electron clouds of atoms – X-ray scattering lengths (also called scattering factors) hence being proportional to the number of electrons – neutron scattering is a purely nuclear process. The scattering lengths are different for each nucleus, *i.e.* each isotope of each element, and they are completely unrelated to the nuclear mass (127) (Table 1). Therefore the scattering length of hydrogen is of the same order as those of heavier atoms found in proteins, such as carbon, nitrogen and oxygen. This means that neutron crystallography is the most unambiguous method for determining the protonation states of proteins and their bound ligands or water molecules.

**Table 1. Neutron scattering properties of elements common in proteins. The coherent scattering length $b_c$, the coherent ($\sigma_{coh}$) and incoherent ($\sigma_{incoh}$) scattering cross sections, as well as the number of electrons for comparison, are listed.**

| Nucleus | $b_c$ ($10^{-15}$ m) | $\sigma_{coh}$ ($10^{-28}$ m$^2$) | $\sigma_{incoh}$($10^{-28}$ m$^2$) | # of electrons |
|---|---|---|---|---|
| $^{1}$H | -3.74 | 1.76 | 82.03 | 1 |
| $^{2}$H | 6.67 | 5.59 | 7.64 | 1 |
| $^{12}$C | 6.65 | 5.56 | 0 | 6 |
| $^{14}$N | 9.37 | 11.03 | 0.50 | 7 |
| $^{16}$O | 5.81 | 4.23 | 0 | 8 |

### 1.5.1 Technical challenges — neutron sources and instruments

Neutron crystallography remains technically so demanding that its use has been limited to only a few proteins. The first problem is that of neutron sources. Modern synchrotron sources have revolutionised X-ray crystallography by providing highly brilliant (photon flux at sample $10^{13}$-$10^{23}$ photons s$^{-1}$ mm$^{-2}$ mrad$^{-2}$ 0.1% bandwidth) and collimated (natural divergence 0.2-0.5 mrad) radiation (116), although the facilities are very large and expensive. The two principal methods for producing neutrons require equally or even more expensive facilities and yet the achievable brilliance at the sample is lower than that of laboratory X-ray sources (photon flux at sample $10^{7}$-$10^{9}$ photons s$^{-1}$ mm$^{-2}$ mrad$^{-2}$ 0.1% bandwidth) (128).

The classical method of producing neutron beams is nuclear fission. The neutron-initiated splitting of the fissile isotope $^{235}$U of uranium produces two to five neutrons. Of the 2.5 neutrons produced on average, 1.5 are needed to propagate the chain reaction but the rest can in principle be used for the neutron beam. Because the source is completely divergent, the only way to increase the brilliance is to make the source smaller. The fundamental design problem of reactor based neutron sources is heat transfer, because if the energy density of the reactor increases too much, the nuclear reaction can no longer be controlled. This limits the achievable brilliance of reactor sources, so the reactor source at the Institut Laue-Langevin (ILL)

in Grenoble, France (maximal flux density ~$10^{15}$ n $cm^{-2}$) has remained the brightest neutron source in the world for more than 30 years.

The other method for producing bright neutron beams is spallation, where high-energy protons (in the GeV range) are directed upon a heavy metal target made of uranium, tungsten or mercury. The nucleus is excited and emits the extra energy as neutrons. Because the incident protons come in pulses, typically at a frequency of 50 Hz, the resulting neutron beam also has a well-defined time structure. As discussed below, this time structure can be used to resolve reflections by the neutron time of flight. Recently two new, high brilliance spallation sources have been built: the Spallation Neutron Source (SNS) in Oak Ridge, Tennesee, United States and the Japan Proton Accelerator Research Center (J-PARC) in Tokai-mura, Japan.

Both fission and spallation initially produce fast neutrons, the wavelength of which is far too short for protein crystallography. Therefore the neutrons have to be moderated in order to match their wavelength with the interatomic spacings in proteins. When neutrons collide inelastically with light nuclei, their speed (and hence wavelength) distribution is moderated to a Maxwellian distribution corresponding to the temperature of the moderator. Thermal neutrons (energy ~25 meV or 1.8 Å) are typically produced from $D_2O$ at ambient temperature; hot neutrons (>100 meV or 0.9 Å) from a graphite rod at 2000 K and cold neutrons (<10 meV or 2.85 Å) from liquid $D_2$ at 20 K. Cold neutrons are usually used for protein crystallography, partly because the detectors are more sensitive to slower (longer wavelength) neutrons (129).

There are currently only few instruments in the world suitable for neutron protein crystallography. Due to the limitations of the sources and detectors and also the wavelengths used, the design concepts of neutron diffractometers differ somewhat from that of X-ray diffractometers. The BIX-3 (130) and BIX-4 (131) instruments at the Japanese Research Reactor (JRR) in Tokai-mura, Japan, and D19 at the Institut Laue Langevin (ILL) in Grenoble, France use a monochromatic neutron beam and hence resemble a conventional X-ray diffractometer, but the Japanese instruments have a cylindrical detector for better coverage of reciprocal space. Even though the monochromatisation facilitates the data processing immensely and the data quality tends to be better, the loss of overall flux is significant. The Laue diffractometer LADI at the ILL (129) – recently replaced by an improved version LADI-III – uses a 'pink Laue' method; a broadband wavelength filter excludes the extremes of the neutron spectrum, which reduces the overlap issues, but the vast majority of the available neutron flux is used to advantage.

Instruments at spallation sources invariably use the Laue method, but the time structure of the incoming neutron beam makes it possible to resolve overlaps by the neutron time-of-flight that is related to the velocity and hence wavelength. An additional benefit of the time-of-flight resolution is that the background is evenly distributed over the wavelength dimension, effectively improving the signal-to-noise ratio of the Bragg reflections. The only operational instrument to date is the Protein Crystallography Station (PCS) (132) at the Los Alamos Neutron Science Center (LANSCE) at Los Alamos, United States. Two others are in construction at brighter spallation sources; the iBIX (133) at the J-PARC, and the Macromolecular Neutron

Diffractometer (MANDI) (134) at SNS. All of these instruments feature an array of area detectors with a high time resolution to cover as much reciprocal space as feasible.

### 1.5.2 Requirements for the sample

Even the brightest neutron sources are very weak, so the data collection times are measured in days or weeks (135). Even then the crystal size required to collect data is about of 1 $mm^3$: roughly a thousand times that used routinely for X-ray crystallography. For a crystal of a given size, the scattering intensity in each reflection is inversely proportional to the third power of the unit cell volume. This means that larger crystals are needed when the unit cell is large. The size requirement is also highly dependent on the crystalline order, so data have been collected from crystals as small as 0.15 $mm^3$ (136). These are aldose reductase (see section 1.5.4) crystals that diffract X-rays to 0.66 Å. They are very robust, and their mosaicity is low.

While it is never really easy to grow protein crystals (137, 138), the difficulties increase exponentially with the crystal size. The key to understanding crystal growth is the phase diagram (Figure 1.18) where the solubility of the protein is plotted as a function of some physical variable, such as precipitant concentration or temperature.

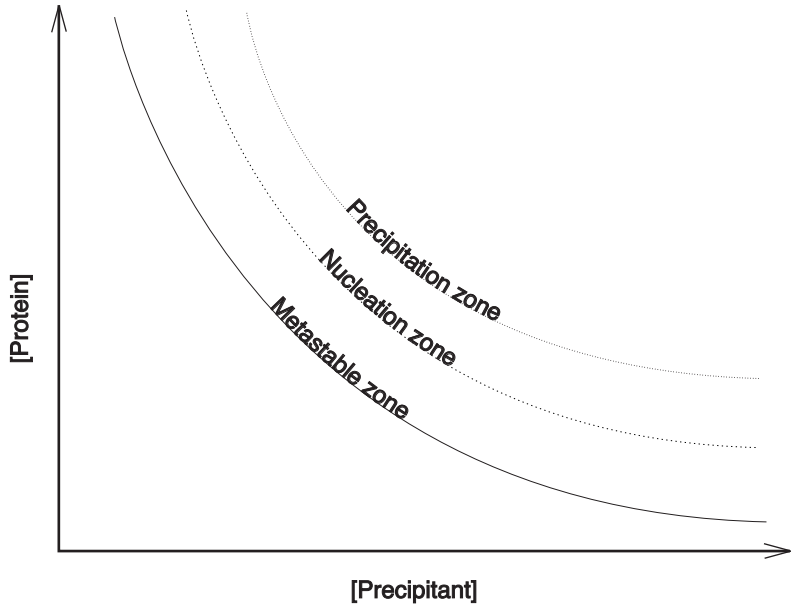


**Figure 1.18 A phase diagram for protein solubility as a function of precipitant concentration depicting the supersaturation, nucleation and precipitation regions.**

Below the solubility curve the protein is undersaturated and existing crystals will dissolve. Above the solubility curve the protein is supersaturated, so crystals will grow; but no new crystals will nucleate at the vicinity of the curve, because there is an activation energy associated with nucleation. Once the degree of supersaturation reaches the nucleation limit, new crystals will nucleate. If crystal nuclei are introduced to a supersaturated solution below the nucleation limit, there are only a

limited number of protein molecules available for the growth of crystals. The product of the liquid volume and the concentration difference to the solubility limit defines this number. Once the solubility curve is reached the crystals stop growing. If the starting protein concentration is too high, new nuclei form and consume all the available protein, the result being more but smaller crystals.

So some method is needed to move along the phase diagram. Although large crystals can sometimes be grown by vapour diffusion by repeated microseeding (139, 140) or dialysis (141), temperature is the only variable where the rate of change can be reliably controlled. The temperature-controlled batch crystallisation method of Budaoyva-Spano *et al.* (142) allows large crystals to be grown in a reproducible manner provided that the initial crystallisation conditions and the temperature response of the system are known well enough.

While the objective of neutron crystallography is to locate hydrogen atoms, one of the biggest technical problems is related to hydrogen. As the neutrons are scattered by the nuclei, each isotope of the same element behaves differently. The light isotope of hydrogen, $^{1}H$, is particularly problematic, because it has, due to its spin, a very high incoherent scattering cross section (Table 1). The incoherent scattering does not contribute to the Bragg reflections at all and even though it can be used to study protein dynamics (143), it only adds to the background in crystallographic data collection. As the signal to noise ratio in the diffraction images is poor, this factor can decide whether a useful data set can be collected or not. $^{1}H$ also has a negative scattering length (Table 1), or in other words the phase change upon scattering is 180° different from the phase change for $^{12}C$, $^{14}N$ and $^{16}O$. In practice this means that protons give rise to negative peaks in the nuclear density map (analogous to the electron density map in X-ray crystallography). Because most other atoms common in proteins, such as carbon or nitrogen have positive scattering lengths, the peaks cancel each other out to some extent. This seriously complicates the interpretation of the maps. It is therefore imperative that the sample contains as little $^{1}H$ as possible.

Deuterium ($^{2}H$ or D), the heavier isotope of hydrogen does not suffer from these problems. It has a positive scattering length comparable to carbon, nitrogen or oxygen and a much lower incoherent scattering cross-section than $^{1}H$ (Table 1), so the logical approach is to replace $^{1}H$ with D. Growing large crystals of proteins in deuterated conditions is not as easy as it may sound, because the protein solubility is different in $D_2O$ (144). Worse still, this procedure exchanges only the exchangeable protons to deuterons, leaving all the aliphatic protons untouched. This often does not reduce the incoherent scattering background sufficiently. The solution is perdeuteration (145); recombinant protein is expressed in bacteria that are grown in $D_2O$ and with a deuterated carbon source (146). Perdeuteration is particularly important for crystals with large unit cells (115, 147).

The relative difference in mass between the two isotopes of hydrogen is larger than between *e.g.* $^{12}C$ and $^{13}C$ or $^{14}N$ and $^{15}N$, and most importantly the energies of hydrogen bonds change, which disturbs the normal function of a cell. Higher eukaryotes can not grow at all in 100 % $D_2O$ and only certain strains of *E. coli* survive and grow relatively well, which limits the choice of expression system. $D_2O$

and particularly the deuterated carbon sources are not cheap, but there is a significant difference in price between deuterated glucose and glycerol, so glycerol is preferred as a carbon source. The expression protocols are often optimised for fermentors allowing a high cell density in order to minimise the use of expensive $D_2O$.

Neutron data are usually collected at ambient temperature. The cylindrical image plate detector in instruments such as LADI-III or BIX-3 does not allow a conventional cryostream to be used. Although a setup based on a liquid He displex cryostat is available for LADI-III, only one structure has been solved at a cryogenic temperature (44 K) (148). At PCS a cryostream is available but no cryo-cooled structures have been reported. The reason for this is that successful cryo-cooling of sufficiently large crystals for neutron work is very difficult. Finally, the Laue method is very sensitive to crystal mosaicity (100), which tends to increase upon cryo-cooling.

### 1.5.3 Structure refinement

Once the data are processed to a list of reflection intensities, structure solution and refinement work in principle just as in X-ray crystallography. The only change is that the corresponding neutron scattering lengths (Table 1) replace the atomic X-ray scattering factors. The phase problem is rather trivial, as the X-ray structure is always well known beforehand and X-ray and neutron data can be collected on the same crystal (or at least crystals grown from identical conditions). There is however an issue with the data-to-parameter ratio, because including the hydrogens in the refinement roughly doubles the number of atomic coordinates and B-factors to be refined. The number of reflections on the other hand does not increase and the resolutions typically obtained in neutron data remain rather modest (> 1.5 Å). This problem can be addressed by refining the structure against a combined neutron-X-ray target function, implemented for example in the program packages PHENIX (149) and nCNS (150) (a modification of CNS (151)). This requires that the X-ray data be collected at the same temperature and that the unit cell be exactly the same. This is of course best achieved by collecting an X-ray data set of the same crystal used for the neutron data collection. This has the added benefit of providing the correct unit cell dimensions (the unit cell may change slightly *e.g.* because of deuteration) for Laue data processing, where the unit cell has to be known in order to index the image. Joint refinement significantly improves the quality of the neutron maps (152) and the X-ray map can be used to accurately position the non-hydrogen atoms, which in turn aids in defining the hydrogen atom positions.

### 1.5.4 Examples of neutron structures

The first structure investigated by neutron crystallography was myoglobin in 1969 (153), but since then the neutron structures of just 17 different proteins have been solved, underlining the technical difficulties involved. In recent years there has been a surge of neutron structures due to developments in instrumentation (135, 154). Even so, over half of the available structures are of just a few proteins, such as

lysozyme (2 structures), concanavalin A (3 structures), insulin (3 structures) and rubredoxin (2 structures). I will briefly discuss the most relevant neutron structures of enzymes reported in the literature.

The serine protease endothiapepsin (155) contains a pair of catalytic aspartates, D32 and D215, one of which was proposed to be protonated. Their actual protonation states in the presence of a transition state analogue were determined from the neutron structure. D215 was found to be protonated, while D32 was not. While shedding light on the nucleophile activation, the interpretation of these results was complicated by indications of a low-barrier hydrogen bond between D32 and the substrate hydroxyl. The presence of LBHBs was later confirmed by proton NMR. Later a neutron structure in the presence of a *gem*-diol inhibitor that mimics the transition state even better was solved (156). This structure together with high-resolution X-ray studies further supported the proposed mechanism.

One of the early neutron structures was that of ribonuclease A (111), a phosphoryl transferase. It was crystallised with uridine vanadate, a transition state analogue. The reaction involves a nucleophilic attack by the 2'-OH of the ribose to form a cyclic phosphodiester intermediate, which is then hydrolysed. The enzyme therefore catalyses two distinct phosphoryl transfers: the cleavage of the RNA chain and the hydrolysis of the cyclic intermediate. In both reactions it is assumed that the leaving group oxygen has to be apical in the transition state, which means that the respective nucleophiles come from different sides. Hence they have to be activated by two different histidine side chains, H12 and H119, both of which were protonated in the neutron structure. This result is not very compatible with the mechanistic model, but later work suggested that H12 might play a different role, the general base being a lysine (157).

One of the recent highlights of neutron crystallography has been aldose reductase. This enzyme reduces a wide variety of substrates, including glucose, using reduced nicotine adenine dinucleotide phosphate (NADPH) as a reductant. In the presence of the inhibitor IDD594 the X-ray structure was solved at 0.66 Å resolution and yet only 54 % of the protons could be located. The polar hydrogens in particular were poorly visible. The neutron structure of aldose reductase (136) confirmed the protonation states of the Asp-Lys-Tyr catalytic triad, with a proton shared between Asp and Lys. The combination of neutron and high resolution X-ray data served as a starting point for QM/MM simulations of the reaction trajectory. This suggested that a deprotonated Tyr stabilised by the rest of the catalytic triad is involved in the proton transfer step.

Diisopropyl fluorophosphatase, also a phosphoryl transferase, hydrolyses a number of organophosphorous compounds, including numerous nerve agents such as Sarin, Tabun and Soman. It is activated by $Ca^{2+}$, which is coordinated by four carboxylates and two water molecules. Two competing catalytic mechanisms had been proposed, one involving nucleophilic attack by a deprotonated aspartate, D229, and another by a hydroxide ion coordinating the calcium. The neutron structure (152) – refined using a joint X-ray-neutron target function – clearly excluded the second

mechanism, since D229 was deprotonated and the putative hydroxide was a water molecule with two hydrogens visible in the nuclear density map.

Xylose isomerase, an aldose-ketose isomerase that catalyses the interconversion between xylose and xylulose or alternatively fructose and glucose, crystallises in one of the largest unit cells (a=94 Å b=100 Å c=103 Å) studied by neutron crystallography (158). The reaction involves first the opening of the sugar ring and a subsequent hydrogen transfer between adjacent carbon atoms required to convert an aldose to a ketose or *vice versa*. Even though the first neutron structure (159) did not contain a substrate or analogue, comparisons to substrate-containing X-ray structures identified a water molecule occupying the position of the ring oxygen of glucose. This water was in turn hydrogen bonded to the doubly protonated H54, which would act as a general acid catalyst in the ring opening. For the second step three possible mechanisms had been proposed: a hydride shift, a *cis*-ene diol intermediate or a metal-mediated hydride shift. The putatively catalytic, metal-coordinating water molecule is indeed a water in this structure, not a hydroxide. Its lone pairs are oriented in such a way that it could function as the proton-abstracting group in the *cis*-ene diol mechanism. Later a neutron structure with a linear xylulose molecule bound to the active site was solved (160) where that water was replaced by hydroxide. The O2 of the xylulose was identified in the neutron structure to be a carbonyl, confirming that a proton transfer from O2 to O1 had occurred. Based on the neutron structures the *cis*-ene diol mechanism seems most likely, even though a hydride shift could not be ruled out.

Dihydrofolate reductase (DHFR) is an enzyme involved in many biosynthetic pathways. Using NADPH as a hydride donor, it catalyses the reduction of 7,8-dihydrofolate to 5,6,7,8-tetrahydrofolate. The neutron structure of DHFR was solved (161) in the presence of the inhibitor methotrexate, which is known to bind to several different conformations of the enzyme. Conformational mobility is considered essential for the catalytic activity of DHFR (23). The structure contains two monomers in different conformations, which allowed the comparison between hydrogen-deuterium exchange and hence dynamics in these two conformations. The N1 of methotrexate was shown to be protonated and the neighbouring aspartate (D27) to be unprotonated. This confirmed the ionic nature of the interaction as suggested by previous biochemical results (162). Previous theoretical calculations (21) on the other hand had suggested a neutral dipole-dipole interaction instead of an ionic interaction – a good reminder that theoretical calculations have their limitations.

HIV-1 protease is an aspartyl protease that is essential for the replication of the human immunodeficiency virus (HIV), because it cleaves the viral polyprotein to functional proteins. Inhibitors of this enzyme are effective drugs against HIV. Two aspartates, D25 and D125, one of which is protonated, activate the nucleophilic water. The neutron structure with the transition state analogue KNI-272 (163) showed that D25 is protonated and the *gem*-diol oxygen of the inhibitor donates a hydrogen bond to D125. This confirms some earlier postulations (164, 165) but contradicts computational studies, (166) which suggested that a proton is shared between the carboxylates.

# 2. MATERIALS AND METHODS

## 2.1 ACTIVE SITE MUTATIONS OF YEAST PPASE

*S. cerevisiae* PPase is a well known enzyme, as discussed in section 1.2. A number of conservative active site variants have been previously characterised (167), but the structural explanations of why they were inactive and what they could reveal about the catalytic mechanism could not be deduced without solving the structures. Wild type Y-PPase and seven variants were crystallised in the presence of $Mg^{2+}$ in order to explain the effects of the mutations in terms of structure and shed more light on the mechanism.

### 2.1.1 Crystallisation, data collection and structure solution

The production and mutagenesis of yeast PPase in E. coli are described in Heikinheimo et al. (168). The purification protocol consists of anion exchange with a diethyl aminoethyl (DEAE) resin followed by size exclusion chromatography. The protocol is very similar to the purification of the perdeuterated Y-PPase described in section 2.4.1. In addition to the wild type enzyme, the variants E48D, Y93F, D115E, D117E, D120E, D120N and D152E were investigated. Crystallisation, data collection and refinement procedures are described in Oksanen et al. (61) but are repeated here for convenience. Crystals were grown at 277 K (+4 °C) in 8 $\mu$l hanging drops with a protein concentration of ~20 mg/ml. The drop solution contained 5 mM $MgCl_2$, 1 mM $KH_2PO_4$, 30 mM 2-morpholinoethane sulphonic acid (MES) buffer at pH 6.0 and 16-19 % 2-methyl-2,4-pentanediol (MPD). The drops were equilibrated against a well solution with a 2-3 % higher concentration of MPD. The thin, plate-like crystals took two to four weeks to grow. For data collection the crystals were cryoprotected in 32 % MPD, 30 mM MES (pH 6.0), 10 mM $MgCl_2$ and 1 mM $KH_2PO_4$ at 277 K before plunging into liquid nitrogen. Data for the wild type crystal were collected at the Swiss Norwegian beamline BM01 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, while the data for the variants E48D and D115E were collected at a laboratory source with Cu Ka radiation (Rigaku R3HU generator with Osmic optics operating at 50 kV voltage and 100 mA current and an R-AXIS IV++ detector). The other data (Y93F, D117E, D120E, D120N and D152E) were collected at a laboratory source in Tromsø with Cu Ka radiation (Nonius generator operating at 40 kV and 170 mA and an R-AXIS IV detector). The data quality was good (Table 2) and the data were not even collected to the diffraction limit of the crystals. Data were processed with the HKL package (169) and Scala (170). The space group of all the crystals was P21. The structures were phased by molecular replacement in the program Molrep (171) using PDB-ID 1HUK (Swaminathan et al., unpublished) as the model. This model was chosen because it was known not to be exactly correct, so the model bias could be more easily estimated by inspecting the areas known to be mistraced in the model. ARP/wARP (172) was used for automated model completion;

refinement (Table 3, Table 4) was performed with Refmac5 (173) and manual model building in O (174) and Coot (175). The superpositions were performed with RgbSup (Fitzgerald, P.D.M., unpublished program) that performs an iterative alignment of the Ca-atoms to be used in the fit. A cut off of 0.2 Å was used.

**Table 2. Data processing statistics for the Y-PPase variants**

| | YWT | E48D | Y93F | D115E | D117E | D120E | D120N | D152E |
|---|---|---|---|---|---|---|---|---|
| Unit cell (Å) | a=51.92 b=93.55 c=69.17 β=99.66° | a=51.71 b=93.20 c=69.29 β=99.47° | a=51.77 b=93.54 c=69.44 β=99.88° | a=51.58 b=93.37 c=69.78 β=99.86° | a=51.76 b=93.23 c=69.22 β=99.58° | a=51.65 b=93.19 c=69.38 β=99.62° | a=51.65 b=93.00 c=69.27 β=99.33° | a=51.77 b=93.22 c=70.11 β=100.01° |
| Wavelength (Å) | 0.8727 | 1.5418 | 1.5418 | 1.5418 | 1.5418 | 1.5418 | 1.5418 | 0.8000 |
| Resolution (Å) [a] | 100-1.5 (1.53-1.50) | 50-1.7 (1.76-1.70) | 100-1.7 (1.74-1.70) | 50-1.8 (1.86-1.80) | 100-1.8 (1.85-1.80) | 100-1.8 (1.85-1.80) | 100-1.9 (1.94-1.90) | 100-1.5 (1.55-1.50) |
| Unique reflections [a] | 103964 (6946) | 68800 (3877) | 67570 (4572) | 58354 (3832) | 59558 (4078) | 59922 (4148) | 50505 (2671) | 102428 (9238) |
| Redundancy | 7.4 | 3.5 | 4.1 | 3.2 | 3.5 | 4.1 | 3.5 | 8.3 |
| Completeness (%)[a] | 99.8 (100.0) | 96.8 (84.0) | 94.4 (94.4) | 96.8 (93.6) | 99.3 (98.1) | 99.9 (99.9) | 99.3 (96.6) | 97.7 (88.6) |
| $R_{sym}$ (%)[a,b] | 4.1 (32.1) | 6.7 (25.3) | 6.5 (34.6) | 6.7 (34.4) | 5.6 (32.3) | 5.9 (38.8) | 9.1 (54.2) | 3.5 (15.4) |
| I/σ [a] | 21.7 (3.2) | 21.8 (4.4) | 7.9 (2.2) | 18.5 (2.8) | 9.8 (2.4) | 9.5 (1.9) | 7.7 (1.4) | 33.2 (6.5) |

[a] Values for the highest resolution shell are given in parentheses.
[b] $R_{sym} = \Sigma_{hkl}|I(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}I(hkl)$ where $\langle I(hkl)\rangle$ is the mean of the symmetry equivalent reflections of I(hkl).

**Table 3. Refinement statistics for the Y-PPase variants**

| | YWT | E48D | Y93F | D115E | D117E | D120E | D120N |
|---|---|---|---|---|---|---|---|
| $D_{min}$ (Å) | 1.5 | 1.7 | 1.7 | 1.8 | 1.8 | 1.8 | 1.9 |
| Total unique reflections/test set | 103685/ 5172 | 68738/ 3472 | 67533/ 3417 | 58286/ 2933 | 59484/ 3002 | 59834/ 3024 | 46553/ 2370 |
| $R_{work}/R_{free}$ [a,b] | 0.162/ 0.180 (0.175/ 0.208) | 0.159/ 0.199 (0.247/ 0.318) | 0.167/ 0.190 (0.310/ 0.356) | 0.175/ 0.217 (0.293/ 0.304) | 0.162/ 0.190 (0.233/ 0.279) | 0.167/ 0.203 (0.229/ 0.265) | 0.154/ 0.210 (0.252/ 0.294) |
| Bond r.m.s.d. from ideality (Å) | 0.011 | 0.015 | 0.013 | 0.016 | 0.013 | 0.015 | 0.015 |
| Angle r.m.s.d. from ideality (°) | 1.445 | 1.531 | 1.536 | 1.565 | 1.410 | 1.470 | 1.503 |
| Waters | 670 | 782 | 505 | 657 | 437 | 520 | 736 |
| Ramachandran plot, most favoured | 89.9 % (435) | 89.6 % (429) | 90.0 % (433) | 89.0 % (431) | 90.1 % (435) | 90.5 % (430) | 89.5 % (434) |

[a] Values for the highest resolution shell are given in parentheses.
[b] $R_{work} = \Sigma_{hkl}(|F_o(hkl)| - |F_c(hkl)|)/\Sigma_{hkl}|F_o(hkl)|$ for F(hkl) not belonging to the test set (5%);
$R_{free} = \Sigma_{hkl}(|F_o(hkl)| - |F_c(hkl)|)/\Sigma_{hkl}|F_o(hkl)|$ for F(hkl) in the test set (5%)

**Table 4. Average temperature factors (B-factors) in Å$^2$.**

| | YWT | E48D | Y93F | D115E | D117E | D120E | D120N | D152E |
|---|---|---|---|---|---|---|---|---|
| main chain | 15.3 | 16.9 | 24.1 | 20.6 | 25.0 | 24.8 | 27.3 | 15.0 |
| side chain | 16.6 | 18.5 | 26.0 | 21.7 | 26.7 | 26.3 | 28.3 | 17.5 |
| all protein | 15.9 | 17.7 | 25.0 | 21.2 | 25.8 | 25.5 | 27.8 | 16.2 |
| Mg | 21.1 | 18.5 | 22.5 | 30.6 | 24.5 | 42.2* | 28.0 | 19.0 |
| Waters | 26.6 | 28.5 | 32.4 | 31.9 | 31.1 | 32.7 | 38.5 | 27.4 |
| All atoms | 17.4 | 19.3 | 25.8 | 22.6 | 26.3 | 26.3 | 29.3 | 17.9 |

*Likely two conformations

### 2.1.2 Animations

The amount of structural information available on Y-PPase alone is such that its mechanistic interpretation is not evident. As discussed in sections 1.2.1 and 3.1 all of the intermediates in Scheme can now be correlated to a structure. The most intuitive way of visualising this information is an animation. As described in the previous section, a reaction cycle involves seven distinct intermediates. All the structures were superimposed on the model for intermediate A (2IHP MolA) using only the $C_\alpha$-atoms of one monomer. A crucial requirement for the animation to make sense is that the error from the superposition be smaller than the movements that take place during the reaction.

The simplest way to generate an animation would be to just use these seven structures as frames, but in practice this is rather jerky. Therefore some kind of interpolation between the intermediates helps the visualisation, even if it is not the true trajectory of the reaction. This is called morphing. It has to be kept in mind that we have no direct information about these trajectories, only the static frames. A chemically sensible way of interpolating protein motions would be to assume that the movement is essentially that of bond torsions and interpolate in torsion angle space. For PPase, however, some of the most interesting motions involve metal ions, the water molecules bound to them, and the pyrophosphate that is converted to two phosphates. When covalent bonds are broken and formed, the torsion angle coordinate system becomes impractical.

Another possibility would be to estimate the trajectory by performing a molecular mechanics minimisation at points along the linear interpolation path. This approach was not used, since the available force fields do not accurately represent the metals and phosphates in the Y-PPase active site.

I therefore used the simplest approach: linear interpolation. While this almost certainly does not accurately represent the trajectory as the bond lengths and angles are distorted throughout the molecule, it is useful as long as the changes between the intermediates are not too drastic.

Some of the models contained $Mn^{2+}$ and some $Mg^{2+}$, some were variants and the numbering of waters and metal ions was not consistent between the different structures. Also the naming of equivalent atoms, for instance the choice of $O_{\delta1}$ and $O_{\delta2}$ in Asp is completely arbitrary. This causes a rotation of the carboxylate if the choice differs between two structures being interpolated. I therefore modified the

superposed structures to have the same nomenclature for the atoms in and around the active site. The starting position of the $PP_i$ with M3-4 was 50 Å outside the active site. In the ‘ resting enzyme’ structure in the absence of substrate or product (intermediate A) the positively charged side chains are in somewhat different conformations, compared to structures containing at least P1. To reflect the notion that $PP_i$ must first bind before the side chains of the active site can close on it, the incoming $PP_i$ molecule was morphed to bind to intermediate A. This unphysical structure was then morphed to intermediate B without further movement of the pyrophosphate.

The linear interpolation was performed in LSQMAN (176) with 30 steps per morph. For the chemical step the atom names for the pyrophosphate and the nucleophile were changed at step 15 to those of two phosphates. The resulting 240 coordinate files (in pdb format) were then loaded as frames into PyMol (DeLano Scientific LLC, Palo Alto, CA, USA), in which the actual animation clips were generated. The clips were combined and effects were added in iMovie (Apple Computer Inc., Cupertino, CA, USA).

## 2.2 SYNTHESIS AND CHARACTERISATION OF ‘CAGED’ $PP_i$

As a first step towards time-resolved structural studies a ‘caged pyrophosphate’ was synthesised (Figure 2.1). Its photolysis kinetics were investigated to verify that it could be used for triggering purposes. Also its binding to Y-PPase was assayed by inhibition measurements.

### 2.2.1 Synthesis

NMR spectra were recorded with Varian Unity Inove 500 MHz and Varian Mercury+ 300 MHz instruments. The UV-spectra were recorded with a Cary 5E UV-VIS-NIR spectrophotometer. The mass spectra were recorded with an Applied Biosystems Mariner 5114 ESI-TOF. A Waters 501 pump system combined with a Waters 990 UV-detector was used for analytical HPLC, whereas for preparative HPLC a combination of an ISCO 2350 pump and Shimadzu SPD-6A UV-detector was used.

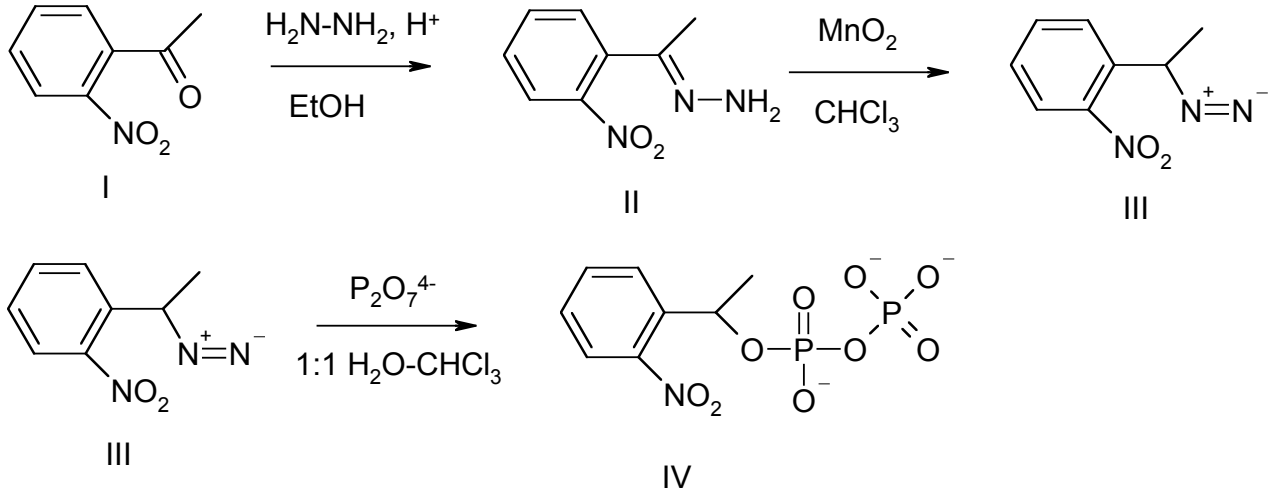


**Figure 2.1 Synthesis scheme**

*2.2.1.1 Hydrazone of 2-nitroacetophenone*

The hydrazone (II, Figure 2.1) was prepared as described by Walker *et al*. (86). 8.26 g of 2-nitroacetophenone was dissolved in 100 ml of 99% ethanol. 5.62 g of hydrazine hydrate was added, followed by 3.2 ml of glacial acetic acid. The reaction mixture was refluxed for 3 h, after which the solvent was removed by rotary evaporation. The residue was dissolved in 100 ml 1:1 chloroform-water. The aqueous phase was back-extracted with 20 ml of $CHCl_3$ and the combined organic phases washed with 50 ml of water. The solvent was removed by rotary evaporation. The yield was 52 %. The $^1$H-NMR spectrum measured in $CDCl_3$ was consistent with that reported by Walker *et al.* (86). The product was used without further purification.

*2.2.1.2 1-(2-nitrophenyl)-diazoethane*

The oxidation of the hydrazone was carried out following Walker *et al*. (86). 2.14 g of the hydrazone was dissolved in 100 ml of chloroform. The following steps were carried out in a dark room illuminated only by red light. 8.55 g of activated grade $MnO_2$ was added to the solution and the mixture stirred for 5 min. The $MnO_2$ was removed by filtration and the product (III, Figure 2.1) used immediately and without further purification. The mass spectrum had a peak at m/z 177 and the $^1$H-NMR spectrum at $\delta$ 2.10 as reported by Walker *et al.* (86).

*2.2.1.3 1-(2-nitrophenyl)-ethyl pyrophosphate*

A 150 mM solution of $Na_4P_2O_7$ was adjusted to pH 4.0 with HCl. 100 ml of the $PP_i$ solution was added to the diazoethane (III, Figure 2.1) solution from the previous step and stirred vigourously in the dark for 24 h. After the reaction the aqueous phase was separated and washed twice with $CHCl_3$. The product (IV, Figure 2.1) was purified by preparative HPLC on a $C_{18}$ reversed-phase column. The eluent was 1:1 water-methanol. The retention time was established with UV detection at 280 nm and the detector was then turned off to avoid photolysis during the purification. The

solvent was removed by rotary evaporation. The overall yield of the purified product starting from I was 43 %.

#### *2.2.1.4 Verification of the product*

A mass spectrum was recorded using electrospray ionisation and time-of-flight mass filter (Applied Biosystems Mariner). A peak at m/z 328 was observed, and the fragmentation was consistent with the expected structure. The $^{1}$H-NMR spectrum was also consistent with the expected structure. Further evidence of a pyrophosphoric acid ester was obtained by a $^{31}$P decoupling experiment, upon which the fine structure of the benzylic proton changed from a quartet of doublets (appearing as a quintet due to the small coupling constant to the phosphosrus) to a pure quartet. The UV-VIS spectrum contained a broad absorption band in the UV with a maximum at 266 nm, as was expected for nitrophenyl ester chromophores.

### 2.2.2 Photolysis

1-(2-nitrophenyl)-ethyl pyrophosphate was photolysed in a 10 mm x 10 mm quartz cuvette using the fourth harmonic wavelength of an Nd-YAG laser (Continuum Powerlite 9010) at 266 nm. This was convenient, as it coincides with the absorption maximum of the compound.

The initial experiments were monitored with an Ocean Optics SD2000 fiber optics spectrometer and Top Sensor Systems DH-2000 broadband light source. In the actual data collection two Thorlabs DET2-Si high-speed silicon detectors were used for triggering and measurement. An optical filter was used for wavelength selection between 300 nm and 400 nm, and the data were stored on a Tektronics TDS520B fast storage oscilloscope. The sample concentrations used were 0.5 mM, 1 mM and 2 mM. The photolysis was performed at three different pH values: 7.1, 8.0 and 9.1. The solutions were buffered with 10 mM Tris-HCl, adjusted to the desired pH with NaOH. Initial photolysis experiments monitored with a CCD-based UV spectrometer established that the process was too fast for that instrument (accumulation + readout ~50 ms) and that a transient absorption with a maximum at 350 nm occurs during the process. This was presumed to correspond to the aci-nitro intermediate. The decay of this absorption was then followed with a photodiode and a fast storage oscilloscope. An optical filter was used to select the appropriate wavelength range. The data collection was synchronised to the laser pulse with a second photodiode.

The data were converted to absorbance units and plotted as $\ln(A/A_0)$ *versus* time. The value of $A_0$ was chosen so that the intercept was zero. This plot yields a pseudo first-order rate constant. The data were fitted to the equation $A/A_0 = ae^{-kt} + b$, where a and b are constants.

### 2.2.3 Inhibition measurements

In order to determine the usefulness of the synthesised compound in triggering a reaction in PPase crystals, the first step was to measure the inhibition

constant ($K_i$) values. Assays of PPase activity are based on the colorimetric detection of the released phosphate as a phosphomolybdate complex that absorbs visible light (177). The most common one involves detection at 400 nm, but since the caged $PP_i$ absorbs significantly at this wavelength, this detection method was ruled out. An automatic phosphate analyser (178) based on an assay that uses similar chemistry but detection at 660 nm was therefore used. The rate of phosphate formation in the presence and absence of caged $PP_i$ was measured both for Y-PPase (concentration 0.025 mg/ml) and for membrane bound PPase from *Thermotoga maritima* (TmPPase, concentration 18.8 mg/ml). The measurements were carried out at 303 K in a buffer containing 100 mM 3-(N-morpholino)propanesulfonic acid (MOPS) at pH 7.2, 10 mM NaCl and 40 μM ethylene glycol tetraacetic acid (EGTA).

## 2.3 *BACILLUS SUBTILIS* PPASE

To investigate the transition state in family II PPases, *Bacillus subtilis* PPase (BsPPase) was crystallised in the presence of the inhibitors imidodiphosphate (PNP) and fluoride. The identities of the metal cofactors and their occupancies were determined by anomalous scattering.

### 2.3.1 Fluoride inhibition and non-hydrolysable analogue

A substrate or substrate analogue containing structure of a Family II PPase has been elusive. PPases are very specific to pyrophosphate as a substrate, so carbon-containing pyrophosphate analogues like methylene bisphosphonate (Figure 2.2) or hydroxymethylene bisphosphonate do not bind sufficiently tightly for crystallographic studies (Fabrichniy I., *personal communication*).

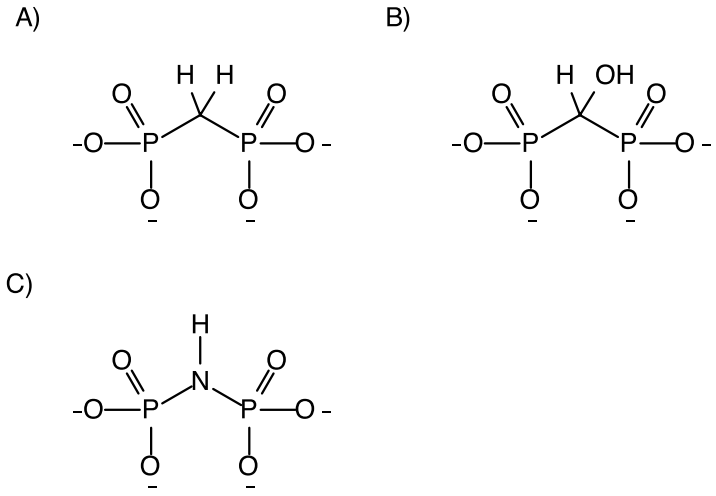


**Figure 2.2 Inhibitors of PPase A) methylene bisphosphonate (MBP) B) hydroxymethylene bisphosphonate (HMBP) C) imidodiphosphate (PNP).**

Imidodiphosphate (PNP, Figure 2.2) on the other hand is hydrolysed by both Family I (Oksanen, E. *unpublished*) and II PPases (Fabrichniy I., *personal communication*), albeit slowly. As discussed in section 1.2, fluoride is an inhibitor of soluble PPases. It displaces the nucleophilic water and thus prevents reaction. In order to prevent reaction BsPPase was crystallised in the presence of both PNP and NaF. The proteins were expressed, purified and treated to remove any bound metal ions as described in Halonen *et al.* (179) To minimise the hydrolysis of PNP by the wild type enzyme, the fluoride-PNP complex was prepared by first adding 5 mM $MgCl_2$ and 10 mM NaF, followed by 1 mM PNP after a 10 min incubation. The crystals were grown at 277 K from 0.1 M potassium 4-(2-hydroxyethyl)-1-piperazineethanesulfonate (HEPES) buffer at pH 7.5 and 2.3-2.5 M $(NH_4)_2SO_4$. The $(NH_4)_2SO_4$ was ultra-pure (Sigma-Aldrich Inc.) to avoid contamination by transition metals. The same crystallisation conditions were used for the H98Q variant (75).

**Table 5. Data processing and refinement statistics for the BsPPase-F-PNP complex and the H98Q variant.**

| | WT-Bs-PPase | H98Q-Bs-PPase |
|---|---|---|
| **Data collection** | | |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell dimensions | | |
| *a*, *b*, *c* (Å) | 60.01, 115.9, 147.9 | 59.91, 115.2, 148.1 |
| Resolution (Å) [a] | 50-1.75 (1.80-1.75) | 50-2.15 (2.19-2.15) |
| $R_{sym}$ [a, b] | 0.052 (0.452) | 0.086 (0.399) |
| $I / \sigma I$ [a] | 14.9 (2.8) | 15.0 (3.3) |
| Completeness (%)[a] | 99.9 (100) | 98.9 (97.9) |
| Redundancy[a] | 4.2 (3.2) | 3.7 (3.4) |
| **Refinement** | | |
| Resolution (Å) | 20 - 1.75 | 30 - 2.15 |
| No. reflections | 99323 | 50598 |
| $R_{work}$ / $R_{free}$ [c] | 0.166/0.186 | 0.169/0.205 |
| No. atoms | | |
| Protein | 4940 | 4771 |
| Ligand/ion | 109 | 74 |
| Water | 436 | 486 |
| *B*-factors (Å$^2$) | | |
| Protein | 44.4 | 33.5 |
| Ligand/ion | 53.9 | 44.7 |
| Water | 47.6 | 35.4 |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.016 | 0.015 |
| Bond angles (°) | 1.52 | 1.47 |

[a] Highest resolution shell is shown in parentheses.

[b] $R_{sym} = \Sigma_{hkl}|I(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}I(hkl)$ where $\langle I(hkl)\rangle$ is the mean of the symmetry equivalent reflections of I(hkl).

[c] $R_{work} = \Sigma_{hkl}(|F_o(hkl)| - |F_c(hkl)|)/\Sigma_{hkl}|F_o(hkl)|$ for F(hkl) not belonging to the test set (5%);
$R_{free} = \Sigma_{hkl}(|F_o(hkl)| - |F_c(hkl)|)/\Sigma_{hkl}|F_o(hkl)|$ for F(hkl) in the test set (5%)

Data for structure refinement were collected at the European Synchrotron Radiation Facility (ESRF) beamline ID14-1 ($\lambda$=0.934 Å) for the wild type and at a laboratory source with Cu $K_\alpha$ radiation, $\lambda$=1.54 Å (Rigaku R3HU generator with

Osmic optics operating at 50 kV voltage and 100 mA current and an R-AXIS IV$^{++}$ detector) for the H08Q variant. Data were processed with the HKL (169) and XDS (180) packages and refined with CNS (151) and Refmac5 (173).

### 2.3.2 Identification of metals by anomalous scattering

The metal binding to Family II PPases is in the nanomolar range (179) and metal ions therefore are difficult to exchange; even small impurities in the reagents used may result in mixed metallation states (74, 75). The metal activation is rather complex (section 1.2.2) and the metal content has subtle effects on the mechanism as discussed in section 4.2, so it is important to verify the identity of the metal cofactors actually present.

**Table 6. Data collection and processing statistics for the data sets at Fe and Mn absorption edges and the heights of the anomalous Fourier map peaks in multiples of the standard deviation σ.**

| | H98Q-Bs-PPase above iron K-edge | H98Q-Bs-PPase below iron K-edge | H98Q-Bs-PPase above manganese K-edge | H98Q-Bs-PPase below manganese K-edge |
|---|---|---|---|---|
| Wavelength (Å) | 1.722 | 1.771 | 1.881 | 1.910 |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| **With fluoride** | | | | |
| Cell dimensions, *a, b, c* (Å) | 59.9, 115.2, 146.8 | 59.9, 115.3, 146.8 | 60.1, 116.0, 147.6 | 60.1, 115.8, 147.5 |
| Resolution (Å) | 2.5 (2.65-2.5) | 2.5 (2.65-2.5) | 2.5 (2.65-2.5) | 2.5 (2.65-2.5) |
| $R_{sym}$[a,b] | 0.057 (0.164) | 0.061 (0.185) | 0.103 (0.365) | 0.095 (0.313) |
| I/σ[a] | 14.2 (5.2) | 13.4 (4.7) | 8.4 (2.6) | 8.6 (2.9) |
| Completeness (%)[a] | 98.6 (97.8) | 98.6 (97.7) | 95.4 (95.6) | 95.7 (97.4) |
| Redundancy[a] | 2.5 (2.5) | 2.5 (2.5) | 2.6 (2.4) | 2.6 (2.4) |
| **Peak heights (σ)** | | | | |
| MolA: M1 and M2 | 23 and 24 | 8 and 13 | 8 and 15 | 3 and 7 |
| MolB: M1 and M2 | 24 and 26 | 8 and 14 | - and 11 | 3 and - |
| **Without fluoride** | | | | |
| Cell dimensions, *a, b, c* (Å) | 59.7, 115.4, 147.0 | 59.7, 115.3, 147.0 | 60.1, 115.9, 147.7 | 60.5, 116.5, 148.5 |
| Resolution (Å) | 2.5 (2.65-2.5) | 2.5 (2.65-2.5) | 2.7 (2.86-2.70) | 2.7 (2.87-2.70) |
| $R_{sym}$[a,b] | 0.072 (0.267) | 0.080 (0.331) | 0.131 (.482) | 0.183 (0.735) |
| I/σ[a] | 11.1 (3.2) | 10.3 (2.6) | 8.1 (2.3) | 6.7 (1.4) |
| Completeness (%)[a] | 99.5 (98.6) | 99.5 (98.2) | 96.3 (98.9) | 96.6 (97.8) |
| Redundancy[a] | 2.5 (2.5) | 2.5 (2.5) | 2.6 (2.4) | 2.5 (2.4) |
| **Peak heights (σ)** | | | | |
| MolA: M1 and M2 | 25 and 21 | 7 and 12 | 4 and 8 | 4 and 3 |
| MolB: M1 and M2 | 26 and 24 | 10 and 13 | 5 and 9 | - and - |

[a]Highest resolution shell is shown in parenthesis.

[b] $R_{sym} = \Sigma_{hkl}|I(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} I(hkl)$ where $\langle I(hkl)\rangle$ is the mean of the symmetry equivalent reflections of I(hkl)

It cannot be assumed that the metal ions present originate solely from the crystallisation reagents. Even determining the metal content of the crystals by *e.g.* atomic absorbance spectroscopy is not sufficient if multiple metal sites with different catalytic roles are present, as in BsPPase (74, 76). The least ambiguous method for determining the identity of individual metal ions in a crystal structure is anomalous scattering (181). This method is applicable to most transition metals, which have absorption edges at energies accessible at synchrotron beamlines (~6-15 keV) adapted for protein crystallography, but not to lighter elements like Mg whose absorption edge is at 1.3 keV.

To identify the metals present in the crystals, fluorescence scans at the absorption edges of zinc, cobalt, iron and manganese were performed. This suggested the presence of iron and manganese in the crystal, so data sets above and below the relevant peak energies of 7.200 keV and 6.591 were collected. Anomalous difference Fourier maps were then calculated for each data set and the height difference of the peaks above and below the absorption edge were used to estimate the occupancy of the metal in question.

## 2.4 PERDEUTERATED YEAST PPASE

Even though yeast PPase is well studied (see in section 1.2), the proton positions are unknown and the most suitable method to determine them is neutron crystallography. Without perdeuteration (section 1.5.2) even large crystals diffracted neutrons only to ~6 Å resolution (139). Perdeuteration was therefore absolutely necessary. The production of perdeuterated yeast PPase as described by Tuominen *et al.* (139) was not reliable, so I had to modify the procedure.

### 2.4.1 Production, purification and metal exchange

The original pYWT plasmid used by Heikinheimo *et al.* (168) for Y-PPase expression contains ampicillin resistance, which is not ideal for expression in $D_2O$. The β-lactamase that confers the resistance is secreted to the medium and destroys the antibiotic and hence the selection in less than a day (182). Therefore the ampicillin resistance gene in the plasmid (168) was changed to a kanamycin resistance gene using a cassette exchange method. The plasmid was purified with the MiniPrep kit (Qiagen Inc.) and transformed into chemically competent *E. coli* cells of the BLR strain. The transformed bacteria were plated on an agar plate with Luria-Bertrani (LB) broth and 100 mg/l of kanamycin. The plates were incubated overnight at 310 K. One of the transformant colonies was inoculated into 10 ml of LB medium containing 50 mg/l kanamycin in a 100 ml Erlenmeyer flask. After 2-3 hours of growth in a shaker at 310 K, 2 ml of the culture was used to inoculate 250 ml of defined medium in $D_2O$ (6.86 g/l $(NH_4)_2SO_4$, 0.49 g/l ammonium citrate, 1.56 g/l $KH_2PO_4$, 1 mM $MgSO_4$, 6.48 g/l $Na_2HPO_4$) with 5 g/l of perdeuterarated glucose (Eurisotop) as a carbon source. The medium was supplemented with a mixture of trace elements and thiamine and adjusted to pD 7 with NaOD.

The culture was incubated at 310 K in a shaker until the optical density at 600 nm reached a plateau, which took 2-4 days. Even though the expression of Y-PPase by the pYWT-plasmid is in principle under the control of the tac-promoter, the construct is very ' leaky'. Induction of expression by isopropyl-β-D-thiogalactopyranoside (IPTG) thus had little effect on the expression levels. As IPTG is slightly toxic to the bacteria, the cultures were not induced at all.

The cells were harvested by centrifugation at 6000 g and washed with a small quantity of the lysis buffer containing 25 mM tris-hydroxymethylaminomethane hydrochloride (Tris-HCl) at pH 8.4 and 15 mM $MgCl_2$. The cells were broken with a cell distruptor (Continuous Systems Inc.) and DNAse was added to reduce the viscosity of the lysate. The cell debris was removed by centrifugation for 45 min at 18000 g.

The supernatant was loaded onto a DEAE sepharose ion exchange column and eluted with the same buffer as above. Y-PPase is not permanently bound to the column material in these conditions, but is retained by the resin to some extent. It therefore elutes (Figure 2.3) after the flowthrough that contains DNA and other anionic components without the use of a salt gradient that is often used in ion exchange chromatography.

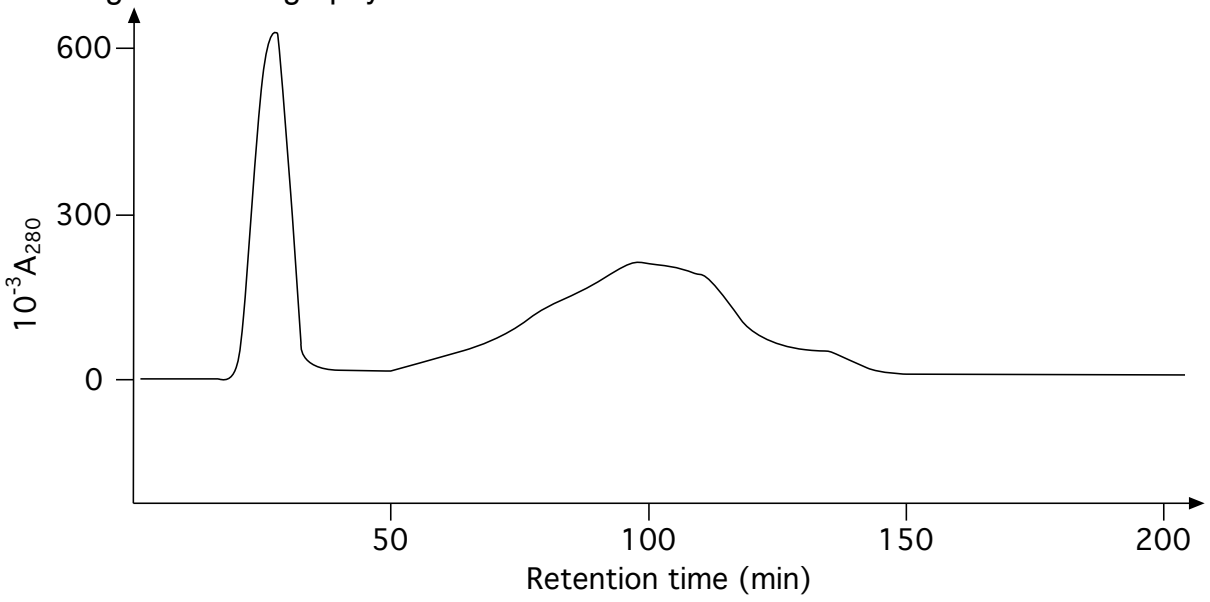


**Figure 2.3. A typical chromatogram of the ion-exchange step.**

The fractions containing Y-PPase were identified by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), combined and concentrated with centrifugal concetrators (Amicon Centriprep, molecular weight cutoff 10 kDa). The concentrate was loaded to a gel filtration column and eluted with a buffer containing 150 mM TrisHCl (pH 7.2), 15 mM $MgCl_2$ and 100 mM NaCl. Y-PPase elutes in two peaks, monomer and dimer, the ratio of which depends on the protein concentration in the loaded sample. As above, the fractions containing PPase were identified by SDS-PAGE, combined and concentrated with centrifugal concentrators.

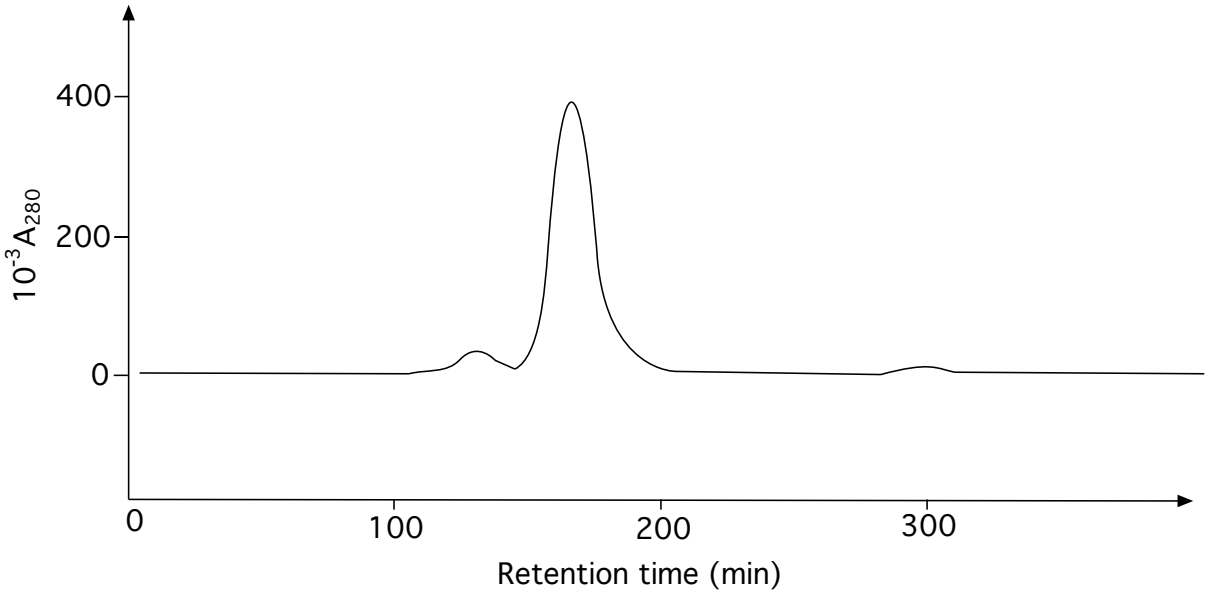


**Figure 2.4. A typical chromatogram of the gel filtration step.**

The metal ion present up to this point in the purification has been $Mg^{2+}$, instead of $Mn^{2+}$ that is used for the crystallisation. The metal affinities are so high that in practice the exchange of the tightly bound metal ions M1-2 is slow even at large excesses (Heikinheimo, P. *personal communication*). Incomplete metal exchange can lead to serious problems in crystallisation, so the purified Y-PPase was diluted with a ten-fold excess of metal exchange buffer (60 mM MES at pH 6.0, 5 mM $MnCl_2$ and 0.1 mM ethylenediaminetetra-acetic acid (EDTA)) and incubated overnight at 277 K. To complete the metal exchange and to exchange the protein into a deuterated buffer for crystallisation, the sample in the metal exchange buffer was reconcentrated to a tenth of the volume and diluted tenfold with the crystallisation buffer, containing 60 mM MES at pD 6.0 and 2 mM $MnCl_2$ in $D_2O$. This procedure was repeated three times. The level of perdeuteration was verified by electrospray mass spectrometry to be 100 %.

### 2.4.2 Optimisation of crystallisation

A crystal is the thermodynamically most stable solid-state form of a substance. For proteins, however, the potential energy difference between the crystalline form and an amorphous precipitate is very small and the precipitation is often irreversible. This means that protein crystals have to be grown by decreasing the solubility very slowly to be as close as possible to an equilibrium process. The most usual way of achieving this is vapour diffusion where both the protein and precipitant concentrations increase as water is evaporating from the crystallisation drop. The crystallisation conditions of Y-PPase are slightly different in the sense that the precipitant is MPD, which has a higher vapour pressure than water and is therefore the first species to equilibrate through the vapour phase. Hence the protein in the drop is actually diluted as the MPD concentration increases.

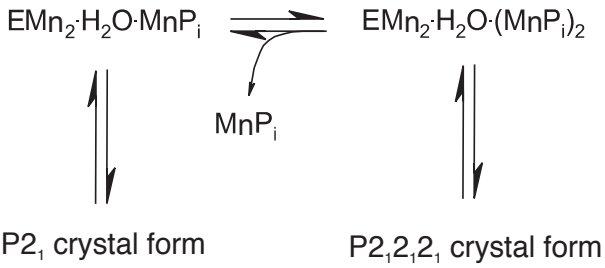


**Figure 2.5. The equilibrium between Y-PPase crystal forms**

The crystal form of Y-PPase in the presence of $Mn^{2+}$ depends on the contents of the active site (Figure 2.5). The orthorhombic form that diffracts to highest resolution has all four metals and two phosphates bound to the active site. A monoclinic form similar to the $Mg^{2+}$ containing crystals described in section 2.1.1 grows when insufficient $MnHPO_4$ is available. Indeed the limiting factor for the growth of the orthorhombic form is the availability of manganese phosphate (Figure 2.5). Growing the very large crystals (~1$mm^3$) required for neutron diffraction in hanging drops required repeated macroseeding into relatively large drop volumes (30 $\mu$l) and obtaining the desired crystal form was never very reproducible (Tuominen, V., *personal communication*). Hence this method was not very economical in terms of protein usage. In a batch setup the volumes are larger, but this also means that the availability of $MnHPO_4$ is higher and therefore the orthorhombic form is almost exclusively nucleated at 277 K. The condition used throughout in the batch crystallisation was 30 mM MES at pD 6.0, 1 mM $MnCl_2$, 0.5 mM $NaH_2PO_4$ and 15 % (v/v) perdeuterated MPD. The batch volume was 400 μl. Initially the protein was used at concentrations around 20 mg/ml, where spontaneous nucleation occurred. After the crystallisation batch had reached equilibrium the protein concentration was measured by absorbance at 280 nm to be 14 mg/ml. These spontaneously nucleated crystals were macroseeded to a new batch with a protein concentration of 15 mg/ml. The crystals grow very slowly and it takes several months for them to reach equilibrium.

The orthorhombic crystals are very sensitive to temperature and start to crack at ~283 K (10 °C). This rules out neutron data collection at ambient temperature. Cryo-cooling of the crystals was attempted in order to collect data with the cryostat setup at the LADI-III instrument (135) at ILL. Y-PPase crystals of moderate size can be successfully cryoprotected in 30-35 % (v/v) MPD in addition to the mother liquor components. In the absence of protein it was determined that 32 % (v/v) d-MPD was required to avoid ice rings. The crystals are not, however, indefinitely stable in such a cryoprotectant solution. Upon slowly increasing the MDP concentration in a batch containing hydrogenated Y-PPase crystals the crystals started to crack at 20% (v/v). Another effect of increasing the MPD concentration is that the monoclinic crystal form starts to nucleate, so it was not possible to grow the crystals originally in higher MPD. Therefore the procedure for cryocooling large crystals of PPase is essentially based on the crystals spending only a sufficiently time in the cryoprotectant solution, in which they gradually degrade. This proved to be a major drawback, especially as the cryostat setup on the LADI-III instrument required a special copper goniometer

head (see section 3.4) that did not allow the cryocooled crystals to be mounted on a laboratory X-ray source to verify the success of cooling.

The neutron data were collected at 277 K with the LADI-III (135), with the cooling device described in more detail in section section 3.4. The data were indexed and integrated with Lauegen (183), the wavelength scaling was performed in LSCALE (184) and scaling and merging in Scala (170).

# 3. RESULTS

## 3.1 Y-PPase variants

I will first describe the overall structures of the wild type (wt) and the variants (E48D, Y93F, D115E, D117E, D120E, D120N and D152E) and compare them to the previous structures of Y-PPase. Based on the contents of the active site (61), each variant can be classified as a structural mimic of an intermediate along the catalytic path (Figure 1.11).

### 3.1.1 Overall structures of the variants

The asymmetric unit in the monoclinic crystal form contains a PPase dimer, but the two molecules are not quite identical. The crystal contacts that the two molecules are involved in are somewhat different; as a result one of the molecules (MolA) is free to adopt any conformation, whereas the second molecule (MolB) is forced into a closed conformation. In the orthorhombic crystal form of the earlier structures, on the other hand, both molecules are in the closed conformation and the active sites are very similar. This is also reflected in the main chain r.m.s.d (Table 7); the conformation of MolA is clearly closer to the open ‘resting enzyme’ structure (1WGI (57)). MolB is even more similar to the closed substrate (1E6A (56)) or product bound state (1E9G (56)). The iterative fitting in RgbSup (Fitzgerald, P.D.M., *unpublished program*) also showed that MolB has a larger region than MolA that superposes well with these closed structures, indicating that the open-closed transition involves some 100 residues (Table 7).

**Table 7. Main chain superpositions of wild type (PDB-ID 2IHP) on previous PPase structures. 1WGI is the ‘resting enzyme’ structure (57), 1E6A is the substrate complex (56) and 1E9G is the high resolution product complex (56)**

| Stationary molecule | WT A | WT A | WT B | WT B | WT A | WT A | WT B | WT B | WT A | WT A | WT B | WT B |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Moving molecule | 1WGI A | 1WGI B | 1WGI A | 1WGI B | 1E6A A | 1E6A B | 1E6A A | 1E6A B | 1E9G A | 1E9G B | 1E9G A | 1E9G B |
| RMS deviation (Å) for all atoms | 0.52 | 0.61 | 0.49 | 0.63 | 0.96 | 0.89 | 0.33 | 0.30 | 0.90 | 0.79 | 0.31 | 0.31 |
| RMS deviation (Å) for atoms not used in fit | 0.93 | 1.10 | 0.84 | 1.11 | 1.4 | 1.32 | 0.83 | 0.93 | 1.34 | 1.16 | 0.81 | 0.88 |
| Number of $C_\alpha$ atoms used in fit | 207 | 200 | 201 | 193 | 153 | 158 | 248 | 259 | 158 | 157 | 251 | 253 |

MolA is in an open conformation in most of the variant structures and MolB in a closed one for all. The main criterion for classifying the conformation as open or closed is the number of water molecules between M1 and M2, even though this correlates well with the main chain changes seen between the conformations (Table 8), namely the movement of the helix 180-195 and the ordering of the loop 105-114. The exceptions to this were the D120E/N and D152E variants, where the two molecules were more similar and the conformation could not be so clearly classified as “open” or “closed”.

**Table 8. Main chain superpositions of molecule A on B of the Y-PPase variants**

| Stationary molecule | WT B | E48D B | Y93F B | D115E B | D117E B | D120E B | D120N B | D152E B |
|---|---|---|---|---|---|---|---|---|
| Moving molecule | WT A | E48D A | Y93F A | D115E A | D117E A | D120E A | D120N A | D152E A |
| RMS deviation for all atoms | 0.76 | 0.67 | 0.80 | 0.66 | 0.63 | 0.56 | 0.34 | 0.34 |
| RMS deviation for atoms not used in fit | 0.92 | 0.81 | 0.96 | 0.93 | 0.82 | 0.71 | 0.54 | 0.54 |
| Number of $C_a$ atoms used in fit | 88 | 85 | 84 | 140 | 111 | 104 | 170 | 174 |

### 3.1.2 Assigning the mutants to reaction intermediates

The six-state catalytic cycle of Y-PPase (Figure 1.11) inferred from the kinetic data does not describe the events after the chemical step, which must include product release, the regeneration of the nucleophile and the regeneration of the two-water bridge between M1 and M2. A revised reaction scheme (Figure 3.1) based on the structures presented here accounts for all of these. Based on the contents of the active site, the structures can be assigned as mimics of a particular intermediate (Table 9).

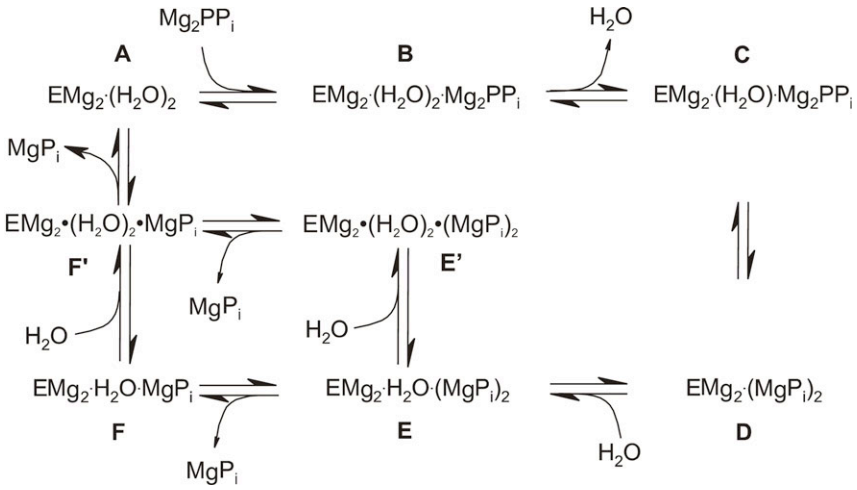


**Figure 3.1. The revised catalytic cycle of Y-PPase based on the structures presented here.**

**Table 9. Active site contents of Y-PPase and possible role in the reaction scheme in.**

| Variant (PDB code) | Active site content (MolA) | Intermediate in Figure 3.1 | Active site content (MolB) | Intermediate in Figure 3.1 |
|---|---|---|---|---|
| *wt* (2IHP) | M1, M2, $P_{exit}$ | **A** | M1, M2, M3, P1 | **F** |
| E48D (2IK0) | M1, M3, P1 | **F'** | M1, M3, P1 | **F** |
| Y93F (2IK1) | M1, M2 | **A** | M1 | |
| D115E (2IK2) | M1, M2, M3, M4, P1, P2 | **E'** | M1, M2, M3, M4, P1, P2 | **E** |
| D117E (2IK4) | M1, M2 | **A** | M1, M2, M4, P2 | **C** |
| D120E (2IK6) | M1, M3, P1 | | M1, M3*, P1 | |
| D120N (2IK7) | M1, M3, P1 | | M1, P1 | |
| D152E (2IK9) | M2, $P_{exit}$ | | M2, M9 | |
| *wt* (1WGI)§ | M1, M2 | **A** | | |
| *wt* (1E9G)§ | M1, M2, M3, M4, P1, P2 | **E, D** | | |
| *wt* (1E6A)§ | M1, M2, M3, M4, F-$PP_i$ | **C** | | |
| D117E(117E)§ | M1, M2, M3, M4, P1, P2 | **E** | | |

*Likely two conformations

§ $P2_12_12_1$ crystal form has two equivalent active sites featuring the same intermediate of the reaction scheme.

A structural model for the starting point of the reaction cycle, the empty enzyme (A,Figure 1.11) exists both with $Mn^{2+}$ (1WGI (57)) and $Mg^{2+}$ (2IHP (61)) in the wild type enzyme. The r.m.s.d. for $C_\alpha$s is 0.5 Å, which probably reflects mostly the lower resolution (2.2 Å vs. 1.5 Å) of the earlier $Mn^{2+}$-containing structure (57). MolA of 2IHP was chosen simply because the resolution is better; otherwise the models are equivalent.

Intermediate B has proven to be the most elusive; it is also unique to eukaryotic PPases, which show biphasic fluoride inhibition kinetics (67) The fluoride ion first rapidly binds to the open intermediate B replacing one of the waters between M1 and M2. The remaining water molecule is then ejected in a slower but essentially irreversible step, leading to a fluoride-$PP_i$-complex (see also section 1.2.1). Presumably the binding of pyrophosphate or phosphate drives the closing of the active site; it is therefore difficult to observe an open conformation with phosphate or pyrophosphate bound. Fluoride can only bind effectively in the closed active site where it bridges M1 and M2. In the D115E variant all the metal ions and two product

phosphates are bound in the active site (Table 9), but MolA is still in the open conformation. The reason for this is that E115 binds not only to M1 as in wild type, but also to M4. Although the M4 binding in MolA is through a water molecule, the increase in phosphate affinity is sufficient to observe both phosphates despite the open conformation. The structure is therefore a mimic of the product release intermediate E'. Replacing the two phosphates by a pyrophosphate molecule in the conformation observed in the substrate complex (1E6A) provides a reasonable model of intermediate B. It is noteworthy that no other modelling was performed to fit pyrophosphate in the active site.

The model for intermediate C is the fluoride-inhibited substrate complex. This structure was solved in the presence of $Mn^{2+}$, so the coordination distances of the metals are ~0.3 Å longer than with $Mg^{2+}$, which represents the level of uncertainty in the superposition.

Intermediates D and E, the 'immediate' and 'relaxed' product complexes, are alternate conformations observed in the same structure (1E9G (56)). In the chemical step (C→D) the nucleophilic oxygen ($O_{Nu}$) becomes a phosphate oxygen, but remains between M1 and M2. This is presumably not a very stable arrangement, since all the other product complex structures contain the 'relaxed' conformation where the nucleophile has been regenerated. It is possible that this alternate conformation is indeed present in the other structures as well, but was not observed due to the limited resolution.

Whether the first phosphate is released before the second water between M1 and M2 is reintroduced is unclear. In principle both pathways are possible and structural model exist for both intermediate F in which P2 has left first and intermediate E' in which the two-water bridge is regenerated first. The best model for the F intermediate is the $Mg^{2+}$-containing wild-type structure (2IHP), which also contains M3. There is a general trend (Table 9) for M3 to be present with P1 and M4 with P2. Therefore it seems reasonable that M3 is indeed the metal ion associated with P1 and M4 with P2. The best model for E' is again MolA from the D115E variant.

For the intermediate F', where P1 has dissociated and the two-water bridge has been regenerated, the best available model is MolA from the variant E48D. Even though M2 is not present in the active site, the two waters next to M1 are at well-conserved positions.

### 3.1.3 Structural reaction scheme – heavy atom choreography

As discussed in section 1.2.1, yeast PPase is a very well characterised enzyme with a well-estabilished kinetic mechanism (Figure 3.1). With the native magnesium cofactor, the rate-limiting step is chemical catalysis, so the relaxation and product release steps do not feature in the kinetic equations. Therefore it is impossible to distinguish which product release pathway in Figure 3.1 is followed. It is indeed possible that the product release is not ordered, *i.e.* multiple pathways operate simultaneously.

The starting point of the catalytic cycle is the open conformation of the active site with M1-2 bound. The flexible positively charged side chains in the P1 end of the active site (K56, R78 and K193) are relatively mobile. When the substrate binds, presumably with M3-4 bound, these side chains close on the substrate, particularly around P1. This arrangement is strained and the open-closed conformational transition (section 3.1.2) follows rapidly (67). This step generates the activated nucleophile in Y-PPase (discussed in detail in section 4.1.1) by ejecting the second water molecule between M1 and M2 (Figure 3.2 B→C). It also creates the sites on the enzyme where M3 and M4 bind, which is possibly also a driving force for the closure.

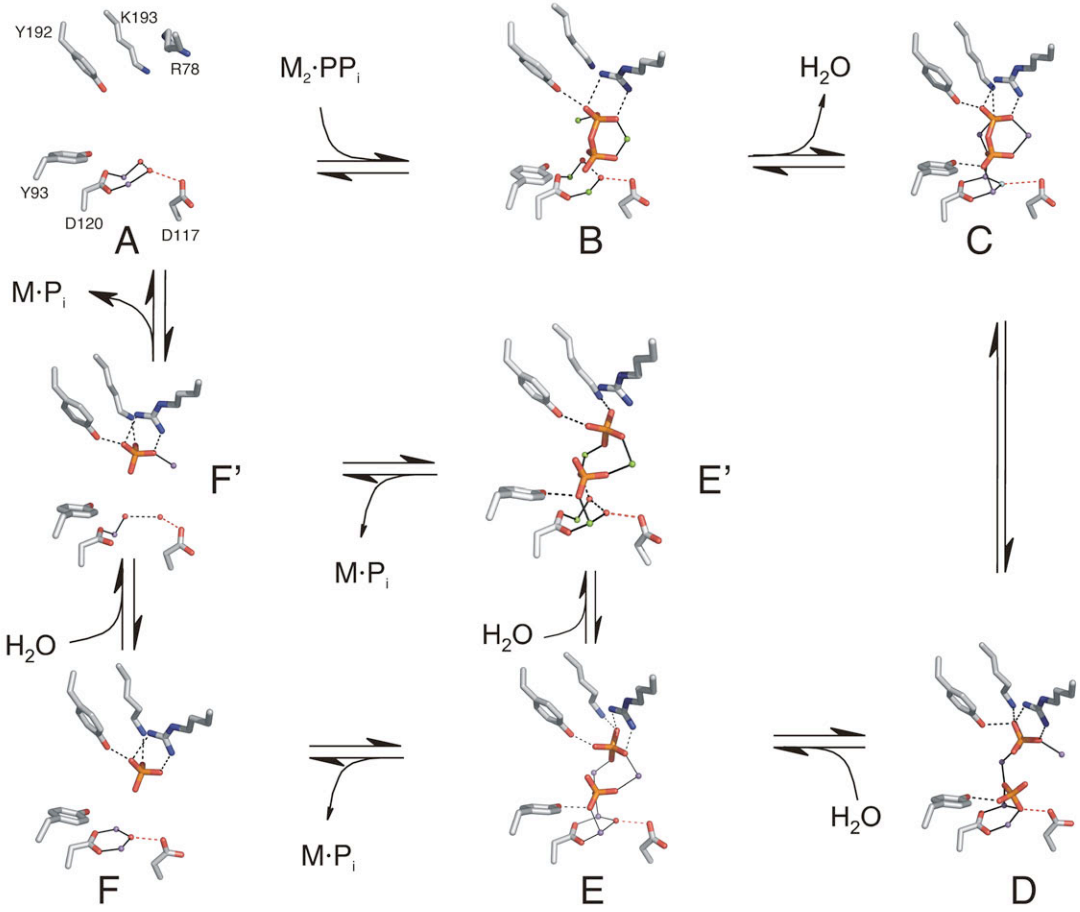


**Figure 3.2. The structural reaction scheme.**

In the actual chemical step the nucleophilic oxygen must attack the electrophilic phosphorous and form a covalent bond while the bond to the leaving group is broken. This means in practice that either the leaving group as a whole must move or it must rotate to allow the former bridging oxygen to move away from the electrophile. In the tightly organised active site there is no room for the entire leaving group (P1) to move back, so the only option is rotation (Figure 3.2 C→D). As essentially each oxygen of pyrophosphate is either hydrogen bonded to a side chain or coordinates a metal ion, this is not as simple as it may seem. M3 rotates together

with P1, while M4 acts as a pivot. The electrophilic phosphorous of P2 moves little; in the immediate product complex only P1 and its associated metal ion M3 have moved. But a phosphate oxygen between M1 and M2 is also a strained configuration, and can only relax by a rotation of P2 accompanied by the replacement of the nucleophilic oxygen (Figure 3.2 D→E). This rotation also includes M3, which returns to its prior position.

Product release can then proceed along multiple paths. Either the electrophilic phosphate P2 leaves first (E→F), propelled by the strong electrostatic repulsion, or the nucleophilic hydroxide between M1 and M2 recruits another water molecule to reform the two-water bridge(E→E'). When P2 leaves, M4 leaves with it. While the expulsion of P2 is electrostatically quite favourable since there are two strongly charged phosphates next to each other in the active site, the dissociation of P1 is not as easy. The active site is, after all, adapted to bind negatively charged phosphates and the positively charged side chains (K56, R78 and K198) all bind in P1.

### 3.1.4 Understanding the effects of mutations

Even conservative mutations, such as Asp→Glu can cause an order of magnitude reduction in the catalytic efficiency of PPases (Table 10, (167)). The Y-PPase active site is very delicately ordered to position the substrate and metal ions for catalysis. Almost all of the mutations introduced also change the pH optimum by several units, underlining the importance of the proper electrostatic environment and hydrogen-bonding network.

**Table 10. Catalytic properties of the Y-PPase variants**

| Variant | $k_{cat}/K_m$ ($10^6$ $M^{-1}s^{-1}$) | p$K_{a1}$ | p$K_{a2}$ | $k_{cat}$ ($s^{-1}$) | $K_m$ (μM) |
|---|---|---|---|---|---|
| wt | 136 ± 14 | 5.5 ± 0.1 | 7.7 ± 0.1 | 197 ± 9 | 1.45 |
| E48D | 65 ± 6 | 7.1 ± 0.05 | > 10 | 78 ± 3 | 1.2 |
| Y93F | 3.7 ± 0.4 | 7.2 ± 0.05 | > 10 | 15.5 ± 0.4 | 4 |
| D115E | 39 ± 7 | 5.9 ± 0.05 | 7.9 ± 0.1 | 23 ± 1 | 0.6 |
| D117E | 35 ± 8 | 8.9 ± 0.1 | > 10 | 28 ± 1 | 0.8 |
| D120E | 0.61 ± 0.16 | 9.2 ± 0.05 | > 10 | 3.0 ± 0.1 | 5 |
| D152E | 0.11 ± 0.01 | 8.1 ± 0.05 | > 10 | 1.91 ± 0.05 | 17 |

[a] Data taken from Pohjanjoki *et al.* (167) The D120N variant did not show measurable activity (185).

The metal binding affinities of each variant were also measured previously in solution and were assumed to correspond to M1 and M2 as in wild type, but this is not necessarily the case for all the variants. These affinities can now be assigned to individual sites based on the contents of the active site (Table 11) as discussed below (section 3.1.4) for each variant.

**Table 11 Metal binding affinities of Y-PPase variants**

| Variant | $K_{M1}$ μM[a] | likely M | $K_{M2}$ μM[a] | likely M |
|---|---|---|---|---|
| wt | 0.7 ± 0.1 | | 34 ± 6 | |
| E48D | 0.7 ± 0.1 | M1 | 5000 ± 1000 | M3 |
| Y93F | 0.7 ± 0.1 | M1 | 60 ± 10 | M2 |
| D115E | 0.8 ± 0.1 | M1 | 400 ± 100 | M2 |
| D117E | 1.4 ± 0.1 | M1 | 170 ± 50 | M2 |
| D120E | 24 ± 2 | M1 | 2800 ± 700 | M3 |
| D120N | 11 ± 1 | M1 | 1400 ± 200 | M3 |
| D152E | 2.8 ± 0.2 | M2 | 1700 ± 500 | M9[b] |

[a] Data taken from Pohjanjoki *et al.* (167).

[b] In the D152E structure an additional metal ion is bound outside the active site

In terms of main chain conformation the variant structures (61) are very similar to the wild type, even though the variation in the more flexible MolA (Table 12) is larger than in MolB (Table 13). There is no indication that the mutations cause any large-scale structural perturbation, so their sometimes rather drastic effects on the catalysis (Table 10) must be due to changes in the active site only. The somewhat surprising structural consequences of each individual mutation in the active site are discussed below.

**Table 12 Main chain superpositions of molecule A of the mutants on wild type**

| Stationary molecule | WT A | WT A | WT A | WT A | WT A | WT A | WT A |
|---|---|---|---|---|---|---|---|
| Moving molecule | E48D A | Y93F A | D115E A | D117E A | D120E A | D120N A | D152E A |
| RMS deviation for all atoms | 0.28 | 0.16 | 0.44 | 0.28 | 0.40 | 0.78 | 0.62 |
| RMS deviation for atoms not used in fit | 0.63 | 0.46 | 0.75 | 0.64 | 0.67 | 0.96 | 0.85 |
| Number of $C_\alpha$ atoms used in fit | 222 | 247 | 185 | 226 | 179 | 92 | 133 |

**Table 13 Main chain superpositions of molecule B of the mutants on wild type**

| Stationary molecule | WT B | WT B | WT B | WT B | WT B | WT B | WT B |
|---|---|---|---|---|---|---|---|
| Moving molecule | E48D B | Y93F B | D115E B | D117E B | D120E B | D120N B | D152E B |
| RMS deviation for all atoms | 0.16 | 0.10 | 0.18 | 0.13 | 0.16 | 0.18 | 0.41 |
| RMS deviation for atoms not used in fit | 0.51 | 0.29 | 0.34 | 0.44 | 0.45 | 0.43 | 0.60 |
| Number of $C_\alpha$ atoms used in fit | 257 | 263 | 218 | 257 | 249 | 241 | 152 |

### *3.1.4.1 E48D*

The role of E48 in the wild type active site does not at first sight seem very central. It forms a hydrogen bond to one of the coordinating waters of M2. The kinetic studies (167), however, showed that E48 is indeed an important constituent of the catalytic machinery, since the $K_m$ changes by a factor of two (Table 10) and the pH optimum shifts. The metal binding constants determined by equilibrium dialysis are also altered (Table 11); the second binding constant decreases 150-fold. The variant structure (61) contains only M1 and M3, meaning that the second metal binding constant describes M3, not M2 binding.

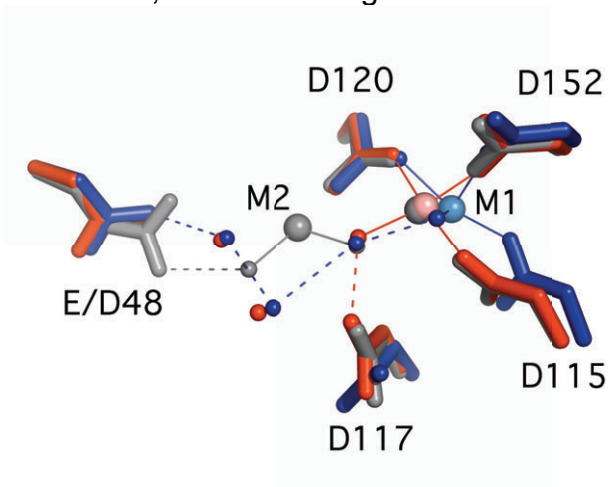


**Figure 3.3. The reorganisation of the hydrogen-bonding network in the active site of the E48D variant destroys the biding site of M2. MolA is shown in blue, MolB in red and wild type in grey.**

So why is E48 so important in forming the M2 site even if it does not participate directly in the metal binding? The answer is probably the conservation of the hydrogen bonding pattern and the comparative rigidity of the active site. The water molecule hydrogen-bonded to E48 is still present in the variant, but it is simply too far away to coordinate M2. As there is no room to accommodate a second water *and* M2, only the second water is present, replacing M2. This shows how delicately ordered the active site is and how important is the correct hydrogen-bonding network.

### *3.1.4.2 Y93F*

Y93 hydrogen bonds to an oxygen of the electrophilic phosphate P2 with a relatively short (2.5 Å) hydrogen bond (56, 57). It is also hydrogen bonded to one of the coordinating waters of M2. The variant structure (61) (Table 9) contains M2 in the open MolA, but not in MolB.

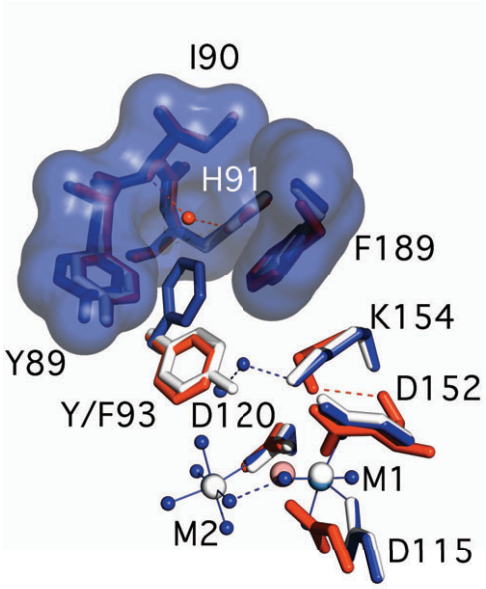


**Figure 3.4. In the Y93F variant the closing of a cryptic hydrophobic pocket (blue surface) present in the open MolA (blue) prevents M2 binding to the closed conformation of MolB (red). Wild type is shown in grey.**

In MolA the hydrophobic sidechain of F93 turns out of the active site into a pocket enlarged by a small movement of residues Y89 and F189. The phenyl side chain also replaces a water molecule buried in the pocket. M2 and the water molecules coordinating it are present. In MolB, on the other hand, the F93 side chain is positioned as in wild type, but M2 is absent even though all the coordinating carboxylates are present.

This variant illustrates particularly well the difference in flexibility between the open and closed conformations. In the open conformation the phenyl side chain that is electrostatically incompatible with the hydrophilic environment of the active site has room to turn away to a hydrophobic pocket opened by main chain motions. When the substrate binds and the conformation closes, this is no longer possible. The F93 side chain is forced to the active site, but cannot form a hydrogen bond with the M2 coordinating water, so the binding site for M2 is lost. This is reflected in the catalytic efficiency, $k_{cat}/K_m$, which is only 3 % of the wild type (167) (Table 10). Even though the open conformation appears catalytically competent and so the variant should have catalytic constants similar to the wild type, the effects of this mutation only become apparent after substrate binding.

### *3.1.4.3 D115E*

D115 coordinates M1, the tightest binding metal ion, and is involved in the open-closed conformational transition. Increasing the length of the side chain should then destroy the delicately ordered active site, but the structure (61) of the variant differs surprisingly little from the wild type (Table 12, Table 13). Both monomers contain the full complement of metals and phosphates in the active site (Figure 3.5). In the open MolA, E115 still coordinates M1 and in addition forms a hydrogen bond

to a water molecule that coordinates M4. This is probably the reason why both metals and phosphates are present in this variant even in the open conformation. Consistent with this the $K_m$ of the variant is actually lower than wild type. Because $K_m$ is a function of $k_{cat}$, a lowering of $K_m$ could also be due to a faster $k_{cat.}$ The change reflects a genuine increase in substrate (and product) binding affinity, because $k_{cat}$ is an order of magnitude slower (Table 10) than wild type.

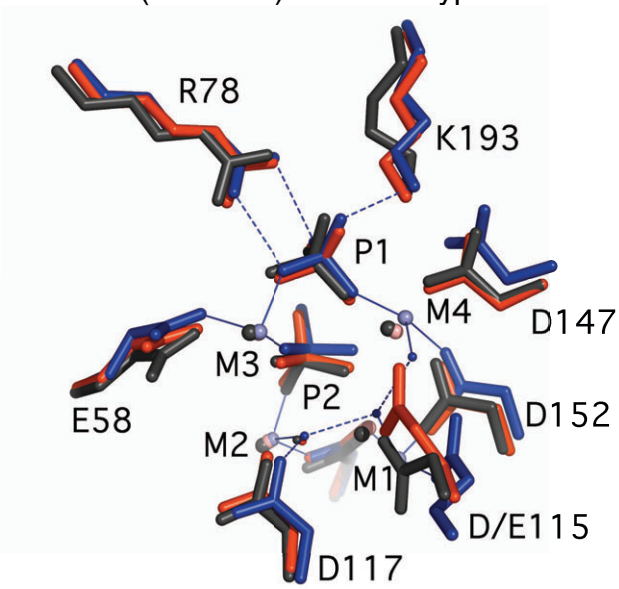


**Figure 3.5. The active site of the D115E contains the full complement of metals and phosphates not only in the closed MolB (red), but also in the open MolA (blue). Both resemble closely the high-resolution product complex (grey, PDB-ID 1E9G(56)).**

In the closed conformation, the full complement of metals and phosphates is also present, with E115 now directly coordinating M4. The reason that the enzyme can adapt to the change in the carboxylate position easily compared to D120 (see below) is that even though M1 is the tighter binding metal (Table 11), the M1-binding loop is the 'moving part' in the open-closed transition and can move out of the way of the longer Glu side chain.

#### *3.1.4.4 D117E*

D117 is one of the most interesting residues in Y-PPase (section 1.2.1) because it forms a low barrier hydrogen bond to the nucleophilic oxygen $O_{Nu}$ (56). The effect of this mutation on the catalytic constants is also only apparent after substrate binding. While the $K_m$ is slightly lower than wild type, $k_{cat}$ is significantly lowered, indicating that the chemical step is hindered (Table 10).

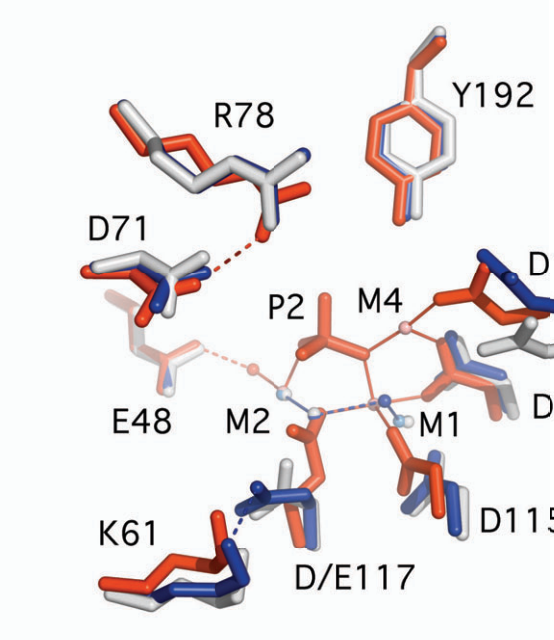


**Figure 3.6. In the open MolA (blue) of the D117E variant, E117 turns away from the active site and forms an ion pair with K61 similarly to wild type (grey). In the closed MolB (red) E117 is trapped in the nucleophile site between M1 and M2, trapping the electrophilic phosphate P2 with it.**

The reason for this is clear from the structure (61); in the open MolA the longer side chain of E117 is turned away from the active site and forms an ion pair with K61. In the closed MolB, on the other hand, a carboxylate oxygen replaces $O_{Nu}$ between M1 and M2. This configuration is obviously catalytically unproductive. In addition, P2 is found in the absence of P1, and M4 in the absence of M3. This indicates that the measured binding affinities (167) correspond to different ions (Table 11). It also clearly illustrates that M4 binds with P2 like M3 with P1. When the charged carboxylate group acts as a mimic of the hydroxide ion between M1 and M2, it brings the two metal ions closer together, which presumably increases the affinity to P2.

### *3.1.4.5 D120E/N*

D120 is the residue that coordinates both M1 and M2, so mutations of this residue decrease the metal affinities from micromolar to millimolar (Table 11). Consequently the activity of the enzyme is destroyed; both variants have less than 0.1% of the wild type activity (Table 10). In terms of the structure (61) these mutations disturb the active site architecture more seriously than any of the others. The binding of M2 is abolished in both cases, which by itself is enough to inactivate the enzyme. For D120N it is easy to understand that the amide group can only coordinate one metal ion. It is actually surprising that even M1 is conserved, as the carbonyl group of asparagine is a poor ligand for $Mg^{2+}$. The reason why the enzyme cannot adapt to the D120E mutation is not as clear. Presumably main chain motions cannot compensate for the longer side chain as with D115E (3.1.4.3) and the active site becomes so seriously distorted (Table 12, Table 13) that these variant structures

cannot really be described in terms of the open and closed conformations (see also section 3.1.2). M2 is absent from both of the variant structures while M1 is bound to both, but its position does not correspond to either the open or closed conformation (Figure 3.7).

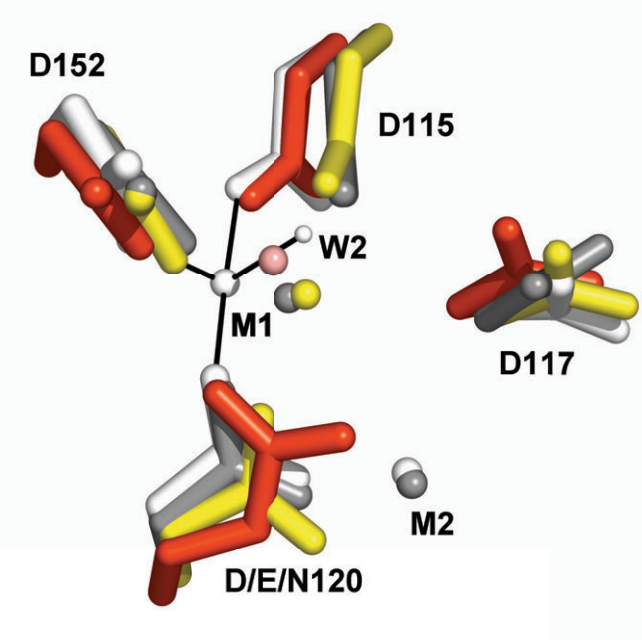


**Figure 3.7. The active sites of D120E (red), D120N (yellow), wild type MolA (white) and MolB (grey) superposed. W2 is the second water molecule beside the nucleophile bridging M1 and M2.**

Instead it is in the same position as a water molecule in the wild type structure, meaning that the nucleophile is not only insufficiently activated due to the lack of the second metal ion, but also mispositioned. Even though the M1 site has higher metal affinity than the M2 site, it is better able to compensate distortions because the main chain of D115 is next to a flexible loop. The site M2, on the other hand, is formed by residues with less conformational freedom.

#### *3.1.4.6 D152E*

D152, along with D115 and D120, coordinates M1 in the wild type but the D152E mutation is the only one to destroy the M1 site. It also causes more significant distortions of the main chain conformation than the other mutations; this is due to its position at the end of a mobile loop. This motion does, however, affect the rest of the structure unlike the motion of D115 (section 3.1.4.3), so the main chain conformation cannot be classified as either open or closed (section 3.1.2, Table 12, Table 13), and unlike the other mutants the two molecules have similar active site conformations. This shows the importance of M1 in ordering the active site. Even though M2 is present and the nucleophile can be identified due to its hydrogen bonding to D117, the enzyme is not appreciably active; $k_{cat}/K_m$ is only 0.01% of wild type (Table 10). An interesting finding outside the active site was a phosphate ion trapped in a putative exit channel in MolA (discussed in more detail in section 4.1.5). This is presumably because the active site is distorted to the point that it can no

longer bind phosphate, and so the next highest affinity site in the exit pathway is occupied.

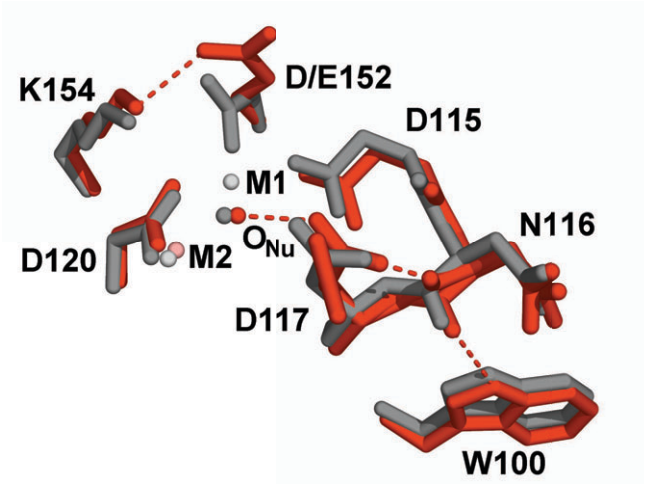


**Figure 3.8. The active site of MolB of the D152E variant (red) superposed on the wild type MolB (grey).**

In MolB an alternate conformation of the N116 main chain carboxylate forms a hydrogen bond to W100. A similar phenomenon is observed in the high resolution product complex structure (56). Even though the D152E variant represents an off-pathway structure, it samples conformations that are also thermodynamically accessible to the wild type enzyme.

## 3.2 CAGED $PP_i$

The preparation of the hydrazone II is rather standard chemistry, as well as its oxidation to the diazoethane III. The alkylation step is a two-phase reaction and proper stirring seems to be essential. The pH also plays a role in selectivity; *i.e.* higher pH favors overalkylation. Overalkylation may be avoided by using sufficiently low pH and excess of the nucleophile.

The synthetic methodology presented by Walker *et. al.* (86) was fairly well suited for my purpose and allowed a rather straightforward synthesis of caged pyrophosphate. The characterisation of the product was harder, but the mass spectra along with the $^{31}P$ decoupling of the $^{1}H$ NMR conclusively demonstrated that the product is indeed the desired one.

The photolytic process is fairly fast. The rate constant at pH 7.1 was ~500 $s^{-1}$, and decreasing the pH by one unit appeared to increase the rate approximately ten-fold. This is consistent with the 350 nm band being associated with the aci-nitro intermediate, the decay rate of which is expected to be first order with respect to $[H^+]$ because the reaction involves a protonation.

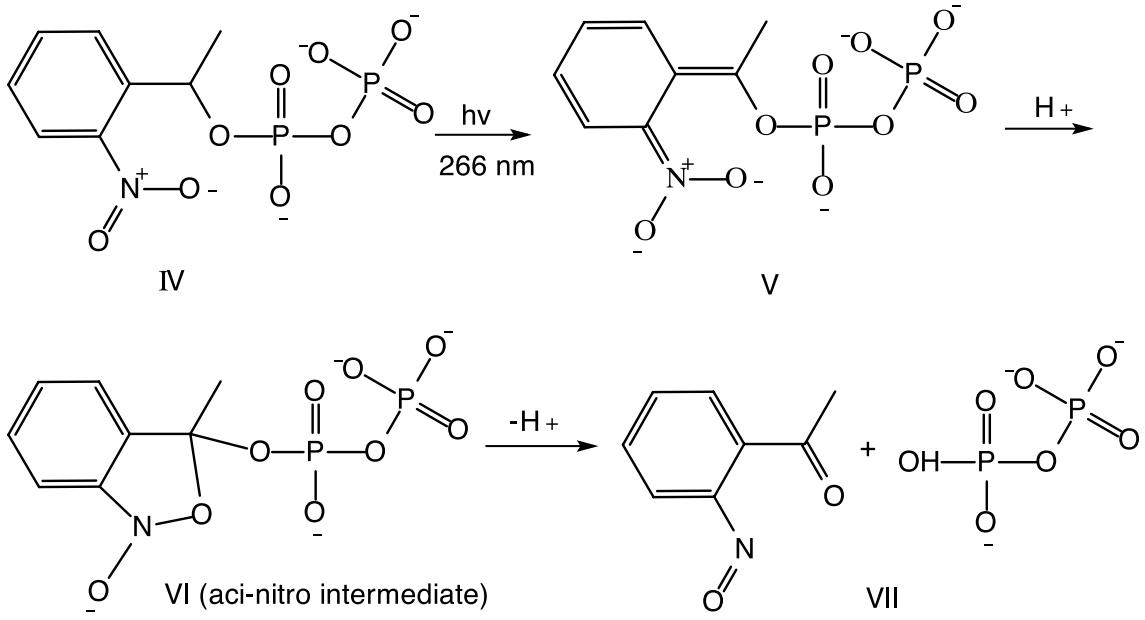


**Figure 3.9. Photolysis scheme**

At higher pHs the kinetic measurements were of quite satisfactory quality, but at pH 7.1 the luminescence from the cuvette began to significantly interfere with the measurement. Table 14 shows the fitted first order rate constants for different conditions. The fact that the rate constants seem to show a concentration dependence suggests that the first-order approximation is in fact not valid. This is reasonable, since the proton concentration is actually much lower than that of caged $PP_i$. For a photolysis reaction in a crystal, the situation might be quite different. However, these measurements do give order-of-magnitude information about the photolysis rate and suggest that the reaction would indeed be sufficiently rapid for triggering purposes.

**Table 14. Kinetic parameters from a non-linear (exponential) fit.**

| pH | [caged $PP_i$] (mM) | $k$ ($s^{-1}$) | $R^2$ of fit |
|---|---|---|---|
| 9.1 | 2.0 | 17.9 ± 0.5 | 0.99 |
| 8.0 | 2.0 | 132 ± 1.9 | 0.98 |
| 7.1 | 2.0 | 599 ± 30 | 0.80 |
| 9.1 | 1.0 | 15.9 ± 0.4 | 0.96 |
| 8.0 | 1.0 | 116± 2.0 | 0.98 |
| 7.1 | 1.0 | 496 ± 35 | 0.69 |
| 9.1 | 0.5 | 13.0 ± 2.3 | 0.79 |
| 8.0 | 0.5 | 144 ± 7 | 0.83 |

The next test for the suitability of ‘ caged’ pyrophosphate for time-resolved crystallography is that it actually binds to PPase and acts as an inhibitor. Unfortunately it is clear from Table 15 that this is not the case. The compound is actually hydrolysed by both Y-PPase and Tm-PPase. Even though PPases are quite selective for pyrophosphate (17), they do hydrolyse other phosphoanhydrides. These results would suggest that the nitrophenylethyl group does not change the properties of the attached pyrophosphate moiety sufficiently to prevent hydrolysis; the bridging atom must be changed. This, on the other hand, would be difficult to achieve in such a way that the compound could still bind to the active site – and Family I PPases are known to slowly hydrolyse inhibitors with N bridging atom substitutions such as imidodiphosphate.

**Table 15. Measured reaction rates with pyrophosphate and ‘caged’ pyrophosphate as substrates of Y-PPase and Tm-PPase.**

| Sample | Volume (μl) | Activity (μmol/min) | Specific activity (μmol/min/mg) |
|---|---|---|---|
| Y-PPase, 1$\mu$M caged-$PP_i$ | 5 | 2.26 | 90.34 |
| Y-PPase, 1$\mu$M $Mg_2PP_i$ | 5 | 4.49 | 179.62 |
| Y-PPase, 10$\mu$M caged-$PP_i$ | 5 | 8.58 | 343.03 |
| Y-PPase, 10$\mu$M $Mg_2PP_i$ | 5 | 8.35 | 333.99 |
| Y-PPase, 10$\mu$M $Mg_2PP_i$ | 10 | 7.12 | 284.71 |
| no enzyme, 1$\mu$M caged-$PP_i$ | 0 | 0 | |
| Tm-PPase, 1$\mu$M caged-$PP_i$ | 10 | 0.46 | 0.025 |
| Tm-PPase, 1$\mu$M $Mg_2PP_i$ | 10 | 1.03 | 0.055 |
| Tm-PPase, 10$\mu$M caged-$PP_i$ | 10 | 2.14 | 0.114 |
| Tm-PPase, 10$\mu$M $Mg_2PP_i$ | 10 | 2.53 | 0.134 |

## 3.3 TRIMETAL SITE AND HYDROXIDE NUCLEOPHILE IN *B. SUBTILIS* PPASE

The structures of wild type BsPPase and its H98Q variant (75) are very similar; they contain two molecules in the asymmetric unit, both in the closed conformation. This is hardly surprising, as the substrate analogue brings the two domains together (Figure 1.14). The N-terminal domain is very similar to other known BsPPase structures ($C_\alpha$ r.m.s.d. 0.15 – 0.39 Å), as is the structure of the active site with small but significant exceptions.

Equilibrium dialysis experiments indicate (77) that besides one tightly binding metal ion, family II PPases bind two others with significantly lower affinity even in the absence of substrate or product. The previous structures of Family II PPases have contained a maximum of three metal ions (72-74), but one of them (M3) has no protein ligands, which makes it an unlikely candidate as the third metal observed in the equilibrium dialysis. Also the identity of the metal ions has not been unambiguously verified. Manganese is thought to be the natural activator (77), but only one of the sites needs to be occupied by $Mn^{2+}$ or another transition metal ion capable of assuming both five and six-coordinated, octahedral geometries. I observed a fourth metal ion (M4) in the new structures. It is coordinated by two aspartates, D13 and D147, a phosphate oxygen, two water molecules and the

fluoride/hydroxide ion (75). The height of the electron density peak in the $2F_o$-$F_c$ map, octahedral coordination and the lack of a peak in the anomalous difference Fourier map (see below) identify it as $Mg^{2+}$.

Based on the X-ray fluorescence scans (see 2.3.2) iron (presumably $Fe^{3+}$) and manganese ($Mn^{2+}$) were present in the crystals. In the anomalous difference Fourier map from data collected at the Fe absorption edge both Fe and Mn have a significant anomalous scattering contribution, because the Fe edge is at higher energy. At the Mn edge the anomalous peaks are due to Mn only, so the relative occupancies of the metals can be estimated by comparing the heights of the peaks at the two wavelengths (Table 6). The comparison of peak heights revealed that the higher affinity M2 site is occupied 66 % with $Fe^{3+}$ and 33 % with $Mn^{2+}$. The lower affinity M1 site is occupied 85 % $Fe^{3+}$ and 15 % $Mn^{2+}$. This suggests that the transition metals bind essentially irreversibly to these sites, whereas the M3 and M4 sites exchange rapidly to $Mg^{2+}$, which was present at 5 mM concentration (75).

A combination of different types of inhibition – PNP that mimics the substrate and fluoride that replaces the nucleophile – was needed to inhibit Bs-PPase sufficiently for crystallographic structure determination. Fluoride, which is an efficient inhibitor of Family I PPases, does not bind as effectively to Family II PPases, because fluoride is a very hard nucleophile and binds less effectively to the softer metal ion electrophile $Mn^{2+}$ in the Family II active site. It is also possible that since Family II PPases are intrinsically more active, a more effective inhibitor is needed to prevent the reaction.

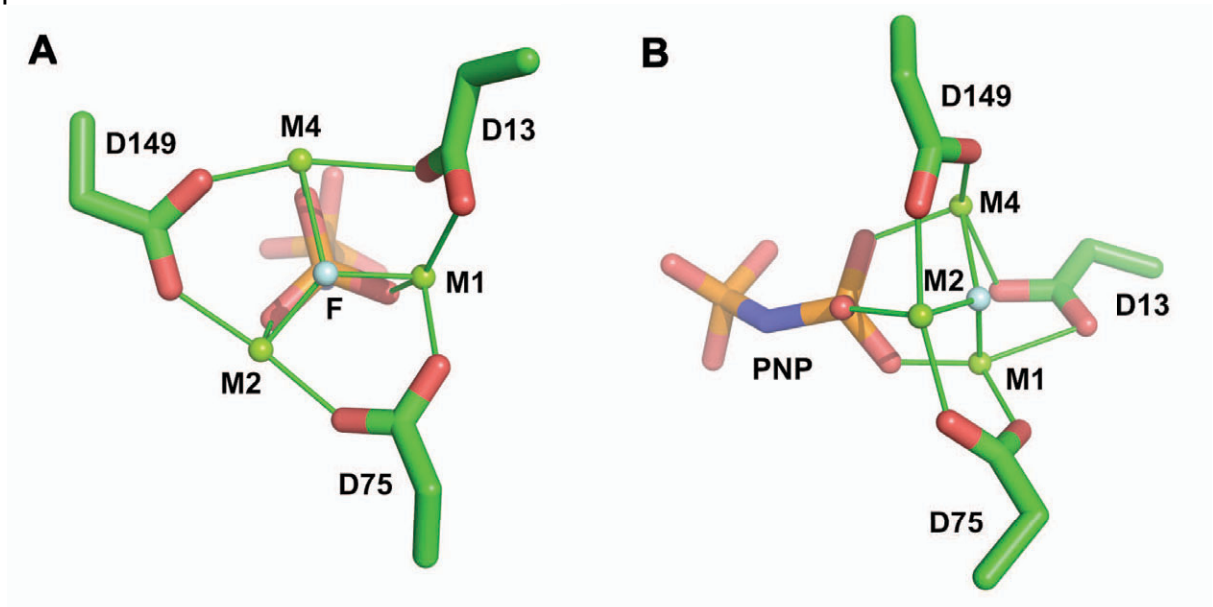


**Figure 3.10. The trimetal site of Bs-PPase with PNP viewed A) perpendicular to the trimetal plane B) from the side.**

The fluoride ion is bound between three metal ions, M1, M2 and M4 in the center of an equilateral triangle (Figure 3.10). The metal ions in turn are each coordinated by one of the oxygens of the electrophilic phosphate, and each pair of metals is bridged by an aspartate (D13, D75 and D149), forming a highly symmetrical arrangement that orients the nucleophile for in-line attack. Fluoride is isoelectronic

with hydroxide, so the binding of fluoride to the nucleophile site strongly suggests that the real nucleophile is a hydroxide. Further evidence was provided by the H98Q variant crystallised in the absence of fluoride, where the nucleophile site was identical (75). Also the fact that the nucleophile coordinates three metal ions suggests that it has to have three lone pairs, which excludes water.

A further way to inactivate the enzyme was the H98Q mutation. While H9 and H97 coordinate the metal cofactors, H98 is actually positioned close to the bridging oxygen and so may act as a general acid. The glutamine side chain in the variant can donate a hydrogen bond to the bridging atom with amide nitrogen. The amide group, however, is far less acidic, so it cannot act as a general acid catalyst.

In the wild type structure PNP replaces $PP_i$ in the active site. PNP is protonated at the bridging atom (Figure 2.2), which means that it can only donate a hydrogen bond, not accept one like $PP_i$. Therefore H98 is very likely to be deprotonated in the substrate complex. In the fluoride-inhibited BsPPase PNP was partly hydrolysed, so the PNP was modelled at 85 % occupancy and a sulphate ion was modelled at the P2 site with 15 % occupancy. In MolA the PNP was also distorted from ideal geometry. The P-N-P angle was 143º compared to 134º in the H98Q variant structure. In addition the P-N bonds are no longer equal in length (1.68 Å) as in the variant. The N-P1 bond elongates to 1.83 Å and the N-P2 bond shortens to 1.47 Å (75).

## 3.4 Neutron diffraction of yeast PPase

Neutron data collection on the LADI diffractometer and its successor LADI-III at the ILL has mostly been performed at ambient temperature. Even though a setup based on a liquid helium displex cryostat exists, it is far from straightforward to use. I tried many times to collect neutron data of perdeuterated Y-PPase crystals at a cryogenic temperature, but no diffraction was observed even for crystals of reasonable size (~0.5 $mm^3$). Whether this was due to insufficient diffracting volume or damage to the crystal in the cooling is hard to determine, as the mounting system is specific for LADI-III and the crystal cannot be recovered for X-ray data collection.

Cryo-temperature data collection was therefore abandoned. Data collection at 278 K was attempted previously with a device that blows cold air around the capillary-mounted crystal (F. Dauvergne, *unpublished*). This had not been successful because a temperature gradient of > 5 K forms between the crystal and the drop of mother liquor that keeps the crystal humid. Since the vapour pressure of MPD is still significant at this temperature, the MPD equilibrates through the vapour phase. This leads to the accumulation of MPD on the crystal, which destroys the crystalline order. In previous tests the crystal had diffracted for 24 h, but then became white and stopped diffracting (M. Spano, *personal communication*). This temperature gradient is difficult to avoid because the airflow, once released from the tubing, becomes turbulent and fails to maintain a precise temperature along the capillary. To counter this effect an aluminium tube was installed around the crystal itself (Figure 3.11).

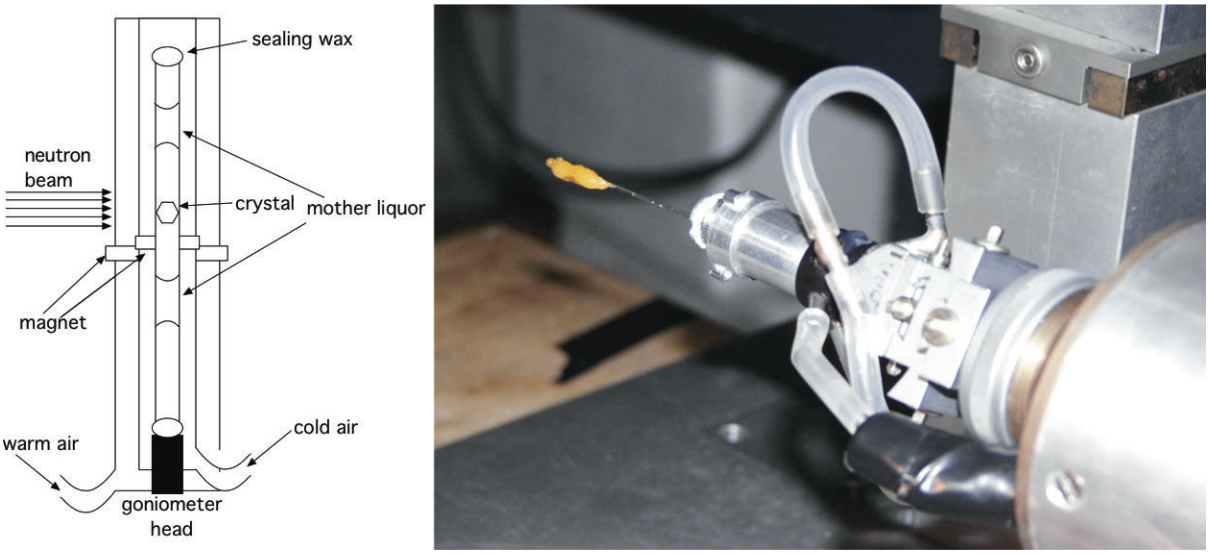


**Figure 3.11 A) A schematic drawing of the crystal cooling device B) A crystal being aligned; the second part of the aluminium tubing is not yet in place.**

With this setup the temperature gradient in the ~5 cm distance of capillary that contains the crystal and the two drops of mother liquor was measured by a thermocouple to be less than 0.5 K. Aluminium was chosen as the material because it is almost transparent to neutrons, and therefore does not contribute significantly to the background. The disadvantage is that the centering of crystal on the rotation axis requires that the crystal can be observed by a microscope, which is obviously not possible through the aluminium. To circumvent this problem a detachable aluminium tube was added with a magnetic ring to hold it in place (Figure 3.11A). The tube covering the crystal is inserted after the crystal has been centered on the $\phi$-axis. This system was tested for two weeks and the temperature stayed at 278 K without condensation in the capillary.

A crystal of perdeuterated Y-PPase was grown as described above (section 2.4.2) and reached a volume of 0.7 $mm^3$. It was mounted in a capillary in the cold room (278 K) and taken to the LADI-III instrument in a box of ice. The crystals had suffered a serious thermal shock due to an air conditioning failure, raising the temperature to at least 287 K, which is enough to damage the crystals (see section 2.4.2). Even though the damage (Figure 3.12) was clear, the optical quality of the crystals showed some recovery after several days. Unfortunately the crystalline order was irreversibly damaged, as evidenced by the X-ray diffraction pattern of a smaller crystal that no longer showed any visible signs of damage.

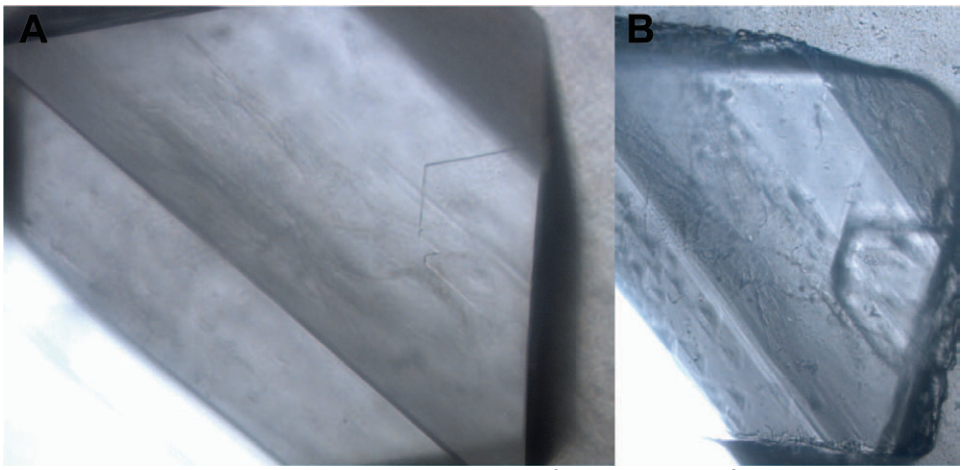


**Figure 3.12 The crystal of perdeuterated Y-PPase a) before and b) after the thermal shock.**

Despite this, the crystal diffracted neutrons to ~2.7 Å resolution, though the spots were split and elongated (Figure 3.13). Seven images at 12 h exposure were collected with a 7º step angle. Unfortunately the diffraction was anisotropic so a full data set could not be collected, but the diffraction at the best orientation remained unchanged after one week, proving the new cooling device to be a significant improvement. The data were eventually processed to 3 Å resolution (Table 16). This experiment also established that the background from the aluminium tube did not affect data quality.

**Table 16. Neutron data processing statistics for the damaged crystal**

| | 47.4-3.0 Å | 47.3-9.5 Å | 3.16-3.0 Å |
|---|---|---|---|
| $R_{sym}$ [a] | 0.167 | 0.113 | 0.189 |
| $I/\sigma$ | 4.1 | 4.1 | 3.6 |
| Completeness (%) | 45.6 | 47.6 | 32.5 |
| Multiplicity | 1.5 | 1.4 | 1.3 |

[a] $R_{sym} = \Sigma_{hkl}|I(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}I(hkl)$ where $\langle I(hkl)\rangle$ is the mean of the symmetry equivalent reflections of I(hkl)

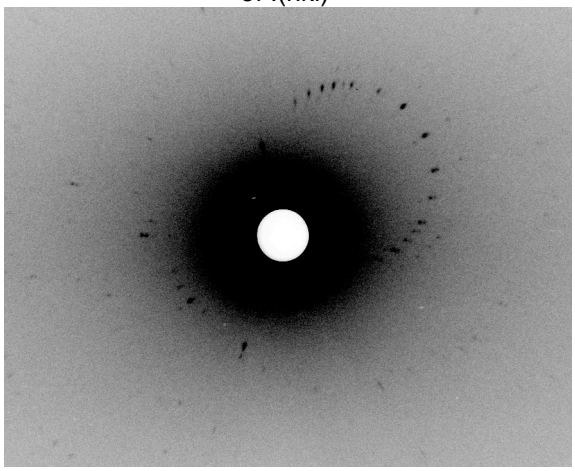

**Figure 3.13. The low-resolution region of a neutron Laue diffraction pattern of a damaged crystal of perdeuterated Y-PPase recorded at LADI-III (12h exposure).**

Later a new crystal of approximately the same volume was grown and mounted in a similar manner. This crystal showed spots to ~2.5 Å resolution (Figure 3.14). The crystal continued diffracting and 17 images of 18-24 h exposures were collected with a 7° step angle. I could process the data to 2.8 Å resolution (Table 17).

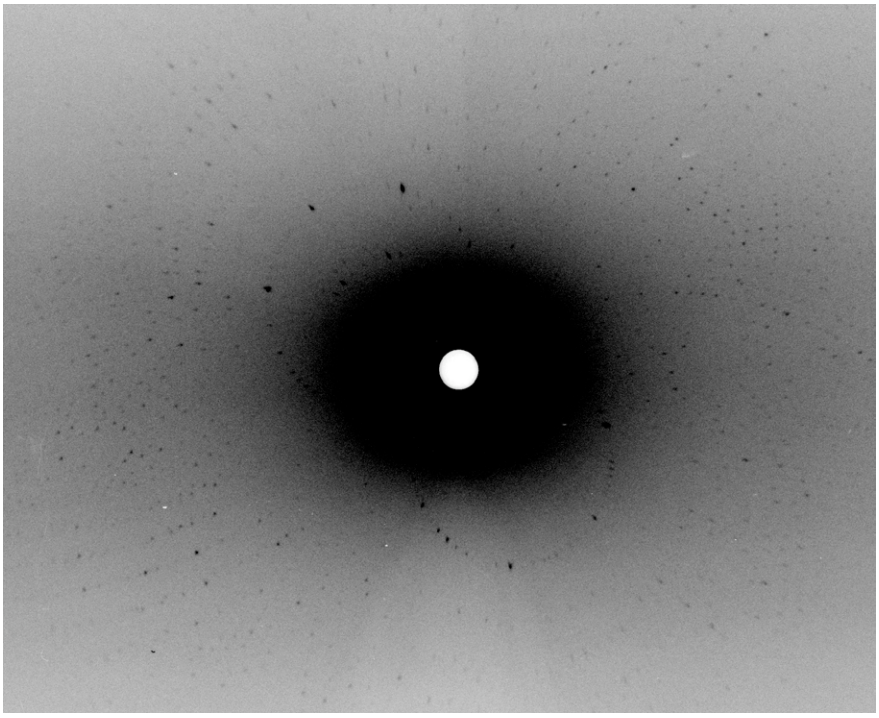

**Figure 3.14. The central region of a neutron Laue diffraction pattern of perdeuterated Y-PPase recorded at LADI-III (24h exposure).**

**Table 17. Neutron data processing statistics for the undamaged crystal.**

| | 52.6-2.8 Å | 52.6-8.9 Å | 2.95-2.8 Å |
|---|---|---|---|
| $R_{sym}$[a] | 0.156 | 0.106 | 0.169 |
| I/σ | 5.9 | 6.1 | 4.5 |
| Completeness (%) | 61.0 | 79.0 | 41.5 |
| Multiplicity | 2.4 | 2.2 | 1.6 |

[a] $R_{sym} = \Sigma_{hkl}|I(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}I(hkl)$ where $\langle I(hkl)\rangle$ is the mean of the symmetry equivalent reflections of I(hkl)

When the data from the two crystals are merged together in Scala (170), the statistics (Table 18) are comparable to other neutron Laue data sets (115, 186). The analysis of these data is in progress

**Table 18. Neutron data statistics after merging the data from the two crystals.**

| | 52.6-2.8 Å | 52.6-8.9 Å | 2.95-2.8 Å |
|---|---|---|---|
| $R_{sym}$[a] | 0.213 | 0.146 | 0.197 |
| I/σ | 4.5 | 5.6 | 2.8 |
| Completeness (%) | 70.4 | 85.1 | 41.5 |
| Multiplicity | 2.7 | 2.7 | 1.7 |

[a] $R_{sym} = \Sigma_{hkl}|I(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}I(hkl)$ where $\langle I(hkl)\rangle$ is the mean of the symmetry equivalent reflections of I(hkl).

# 4. DISCUSSION

## 4.1 MECHANISM OF FAMILY I PPASES

The catalytic power of Family I PPases cannot be attributed to any one residue; even small, charge-conserving changes in the active site (section 3.1.4) are enough to prevent efficient catalysis. The importance of the metal cofactors is indisputable, but is not limited to just one function. Essentially all the possible modes of catalysis described in section 1.1.4 are present in Family I PPases.

A major factor in any enzymatic reaction is preorganisation (22); bringing the substrates together in a productive orientation is in itself catalytic and the accurate positioning of the substrate is an essential prerequisite for other means of catalysis. To bind and orient negatively charged phosphate groups the active site needs to be positively charged. The active sites of family I PPases contain relatively few lysines or arginines (and usually no histidines); instead the active site is packed with aspartates that in turn bind positively charged $Mg^{2+}$-ions. Why does the enzyme make use of metals instead of basic side chains?

One reason may be that phosphate esters and particularly anhydrides, due to their high charge density, rarely exist in cytosolic solution without associated metal ions; therefore the true substrate is a $PP_i$-$Mg^{2+}$-complex. The coordination bond between the metal ion and a phosphate oxygen is essentially a covalent bond and is more effective in transferring charge than a hydrogen bond to an amine (lysine) or guanidino (arginine) group. The covalent character also means that the coordination bonds are much more direction sensitive than hydrogen bonds, not to mention simple charge-charge electrostatic interactions (salt bridges) and thus allow more precision in positioning the substrate. This probably also contributes to the specificity of the enzyme. It is remarkable how the specificity changes with the metal cofactor (187); in the presence of $Mn^{2+}$ the selectivity for pyrophosphate *vs.* other phosphoanhydrides such as ATP is much less marked than with $Mg^{2+}$, even though the geometry is very similar. The electronic properties, namely the presence of available *d*-orbitals in $Mn^{2+}$, provides more flexibility in the coordination geometry. There must also be other factors in the electronic structure of the active site that make it specific for phosphate moieties, since geometrical analogues of phosphate esters, such as vanadates, bind poorly to PPases.

### 4.1.1 Nucleophile generation

The key way in which PPases facilitate pyrophosphate hydrolysis is by activating the nucleophile. It is fairly clearly established that the nucleophilic species is an hydroxide ion. Firstly the isoelectronic fluoride ion can be trapped in the nucleophile position (56). Secondly the environment excludes a water molecule; two electron pairs are taken up by the coordination bonds to M1-2 and the only other plausible hydrogen bond the nucleophile can make is to D117 (56). The distance between the nucleophile and the electrophilic phosphorous atom is 2.9 Å in both the

substrate (1E6A) and product (1E9G, 2IK2) structures, but a hydrogen bond between a water and a phosphorous is very unlikely. As the distance between D117 and the nucleophile is unusually short (2.3 Å) in the immediate product complex (alternative conformation B in 1E9G) where the nucleophile has become a phosphate oxygen, it has been postulated that a low barrier hydrogen bond is involved (56). The role of a low barrier hydrogen bond would be to further deprotonate the hydroxide ion, increasing its nucleophilicity. However, even though the structure of the product complex was solved at atomic resolution (1.15 Å) (56) so the expected coordinate error is low, the active site has two alternate conformations. The alternate conformations significantly increase the uncertainty of the coordinates, so the short distance may also be a refinement artefact. Interestingly, the alternate conformation also involves the main chain of D117 and the neighbouring N116, where Oδ1 moves 0.9 Å to form a hydrogen bond to the indole Nε1 of W100. Mutation of W100 also affects the catalysis (Rantanen, M. *unpublished*). It is also possible that this movement is related to the removal of a proton from the immediate product complex (see below).

Regardless of whether the hydrogen bond is unusual or not, D117 does orient the nucleophile's lone pair for attack; in the substrate complex (1E6A) the angle between the electrophilic phosphorous, the $F^-$ and D117 $O_\delta1$ is 104° (Figure 4.1). For an $OH^-$ this would be very close to the optimal angle between the hydrogen and the lone pair.

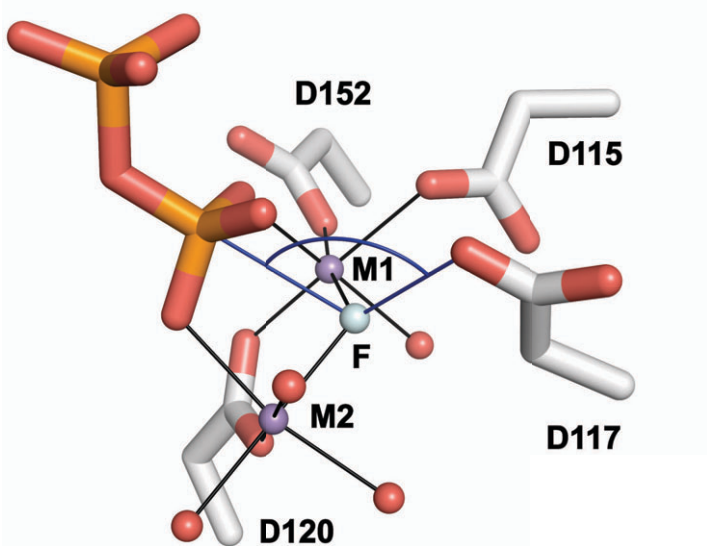


**Figure 4.1. D117 orients the lone pair of the nucleophile for attack. The 104° angle between D117 $O_\delta1$, the fluoride ion**

PPases have to function at the cytosolic pH of ~7, where a hydroxide ion is a very rare species. So how does the enzyme generate the hydroxide in the first place? Here there is an interesting difference between prokaryotic (E-PPase) and eukaryotic (Y-PPase) enzymes. In Y-PPase the substrate binds to the open conformation of the enzyme where two water molecules bridge M1 and M2. The

substrate binding triggers the closure of the active site in which the loop binding M1 moves ~1 Å. This small motion of the main chain is sufficient to push away the second water molecule *with* the proton between the two. This proton has an accessible path to the bulk solvent (discussed in section 4.4).

E-PPase and other prokaryotic, hexameric PPases (such as *Mycobacterium tuberculosis* (188), *Helicobacter pylori* (189) PPases) apparently lack the conformational change described above; all the structures contain only one water (hydroxide) between M1 and M2 and the fluoride inhibition is rapid and reversible unlike Y-PPase (67) (section 1.2.1). This conformational change allows Y-PPase to activate the nucleophile more effectively than E-PPase, and also explains why the pH optimum of Y-PPase is lower.

A recent computational study on E-PPase catalysis (190) supports the role of the metal ions in orienting the substrate and stabilising a trigonal bipyramidal transition state. The orienting effect of D67 (D117 in Y-PPase) was also clear, but a deprotonated model led to an irreversible reaction, contrary to experimental evidence. Yang *et al.* (190) therefore proposed a model in which D67 was protonated and accepts a hydrogen bond from the nucleophilic hydroxide. This does not seem reasonable: the protonation state of the pyrophosphate/phosphate, discussed below (section 4.4) is likely to affect the relative energies strongly.

### 4.1.2 Electrophilic assistance

Electrophilic assistance for $PP_i$ hydrolysis comes in three different varieties: general acid catalysis, leaving group activation and compensation of charge on the peripheral oxygens. All three are to some extent present in Family I PPases. As mentioned in section 1.2.1, the pH dependence of the reaction rate in Family I PPases revealed a group of p$K_a$ 7.5: a general acid. Such a general acid catalyst could at least partly protonate the bridging oxygen of pyrophosphate. However, its identity is not evident from the structure. In the Y-PPase substrate complex structure (1E6A (56)) there are no enzyme side chains within hydrogen bonding distance of the bridging oxygen. There are, however, two water molecules; one is bound between M4 and K56 and the other one between Y89 and Y93 (Figure 4.2). Both water molecules are activated for proton donation, but since the protonation states are unknown, it is difficult to establish whether one or the other or both act as the general acid. In the E-PPase substrate complex structure (PDB-ID 2AU6 (191)), K29 (analogous to K56 in Y-PPase) hydrogen bonds directly to the bridging oxygen. The other water molecule is still present between Y51 and Y55 equivalent to Y89 and Y93.

The positively charged residues R78 and K193 around the leaving group phosphate P1 contribute to the leaving group activation. Interestingly, in the 'nucleophile end' of the active site lysines or arginines are absent, even though the negative charge in the transition state should be concentrated on the peripheral oxygens of the electrophilic phosphate (P2).

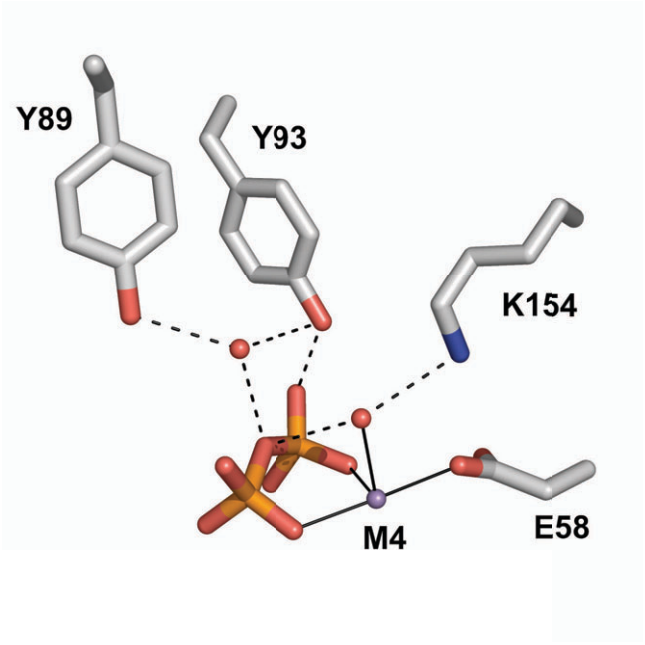


**Figure 4.2. The two putative general acid waters.**

This is related to the discussion above concerning the role of metal ions; complexed metal ions bind the phosphate oxygens efficiently but do not compensate for the electrostatic effect of the positive charges surrounding P1.

The third type of electrophilic catalysis as defined in section 1.1.4, charge compensation of the peripheral oxygens on the electrophilic phosphate moiety, is also present in Family I PPases, although not as marked as in Family II. In the substrate complex (1E6A (56)), all of the non-bridging oxygens of the electrophilic phosphate coordinate a metal ion. The amine group of K56 is also close, 3.2 Å from the nearest oxygen and 3.5 Å from the phosphorous, but it is one of the most flexible residues, so it may well have other roles. The electron withdrawing effect of the coordination bonds is not as efficient as that of hydrogen bonds, so charge compensation does not seem to be a major factor in catalysis. All in all electrophilic catalysis appears less important than nucleophile activation, a notion that gives hints about the nature of the transition state.

### 4.1.3 Transition state

Family I PPases have generally been considered to act through an associative transition state (57, 75). The features discussed above do not as such distinguish between an associative and dissociative transition state, but they do give hints. For example strong activation of the nucleophile is more typical for associative mechanims, because in a truly dissociative mechanism even a weak nucleophile such as water is sufficient to attack the metaphosphate ion. The activation of the nucleophile is more necessary when nucleophilic attack is the initiating event in the chemical step.

The net charge of the transition state and the distribution of that charge are crucial for understanding the mechanism. Unfortunately the charge is tied to the

protonation state, which is not even known for substrate or product. The pH optimum of the enzyme is ~7 (48), but the crystal structures are at a lower pH. It is however likely that the incoming pyrophosphate is highly charged in any case, even if the charge is somewhat compensated by the accompanying metal ions. The neutron structure should shed more light on these questions.

### 4.1.4 Conformational flexibility

As discussed in section 1.1.4, a number of conformational states are thermodynamically accessible to enzymes and changes in their relative energies can affect catalysis profoundly. The Y-PPase variant structures presented here (3.1.4) neatly illustrate this principle. The change between the open and closed conformation has been previously observed as a function of substrate/product presence: the open conformation in the absence of phosphates and the closed in their presence. In the $P2_1$ crystal form reported here (Table 9), the crystal contacts force one molecule into the closed conformation but the other one lacks such constraints. This demonstrates that both conformations are accessible even in the absence of substrate binding; probably they are in rapid equilibrium in solution. Nucleophile activation appears to be a very important catalytic strategy of Y-PPase, and it seems that the generation of the hydroxide nucleophile is driven by this conformational change (61). This feature is, however, absent in E-PPase, which also lacks the mobile loop that in Y-PPase becomes ordered in the closed conformation (61). E-PPase is thus always in the closed conformation, and the lack of nucleophile activation may explain the higher pH optimum (~8) of E-PPase (48).

### 4.1.5 Product release

The active site of Family I PPases is adapted to bind phosphate functionalities, be it pyrophosphate or phosphate, fairly tightly. Ideally an enzyme should bind substrate tighter than product, and transition state tightest of all, but significant differences in the binding energy are difficult to achieve if the substrate or product are chemically similar (see section 4.2 for another example). The electrostatic repulsion between the two charged phosphates pushes one product phosphate away rather easily. It is remarkable that this happens to be the electrophilic phosphate P2, which is more buried than P1 and is surrounded by metal ions rather than positively charged side chains. In the presence of a softer metal ion such as $Mn^{2+}$ that binds more tightly to the phosphate oxygens the order of product release is reversed, because P2 has many more bonds to metals than P1.

The P1 site, however, is much better adapted to expel a single bound phosphate ion. The positively charged side chains that bind the phosphate are quite flexible, and it appears that they undergo correlated motions to ‘thr ow’ the P1 phosphate out of the active site (61). This ‘ waving’ motion is evident from a superposition of the open and closed structures. In the D152E variant a phosphate ion is even observed in this exit channel (Figure 4.3).

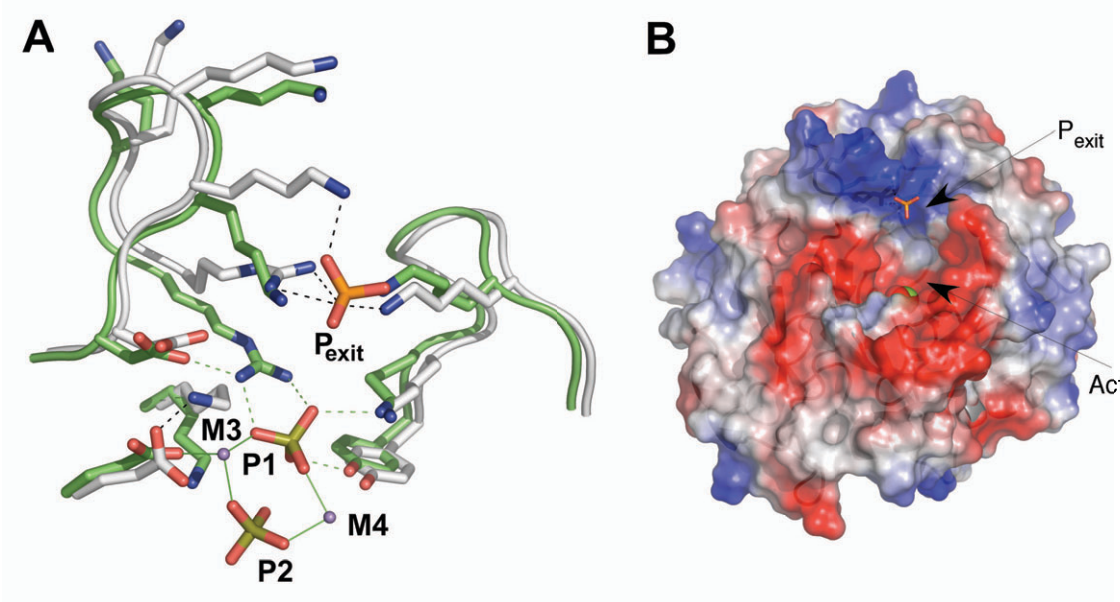


**Figure 4.3 A) The superposition of the high resolution product complex (56) (grey, phosphorous lime) and MoIB of the D152E variant (61) (green, phosphorous orange) shows how the waving motion of lysines expels the P1 phosphate. B) A representation of the surface charge of a Y-PPase monomer showing the positively charged exit channel.**

One open question regarding product release is the protonation state of the product phosphates discussed in section 4.4.

## 4.2 Similarities and differences between family I and II PPase mechanisms

Despite the lack of similarity in either sequence or structure, the mechanisms of Family I and II pyrophosphatases were at first thought to be similar because of an activated water molecule between two metal ions. The metal activation of Family II PPases has since then turned out to be much more complex (section 1.2.2), and the two domains of Family II PPases undergo large motions, making the catalytic cycle quite different from Family I PPases.

Another significant difference is the pronounced role of general acid catalysis in Family II PPases. Whereas the identity of the general acid in Family I PPases remains unclear from the structural studies (see 4.1.2), in the family II active site a histidine hydrogen bonds to the bridging oxygen. Furthermore mutation of this histidine (H98 in BsPPase) to glutamine effectively inactivates the enzyme (75) Glutamine can form similar hydrogen bonds, but is a much weaker acid than protonated histidine, so it cannot act as a general acid. This confirms the importance of general acid catalysis; there are no similar mutations for Family I PPases.

The more prominent role of general acid catalysis also suggests that the mechanism of Family II PPases may be more dissociative than in Family I. In a dissociative mechanism the electron pair of the O-P1 bond is first transferred to the bridging oxygen, concentrating the charge in the transition state there.

The activation of the nucleophile is also rather different, even though both are based on a hydroxide ion between divalent metal ions. While in Family I the nucleophile is oriented for an in-line attack by an aspartate (56, 61), in Family II it is positioned by three metal ions in a roughly equilateral triangle (75). Each of the metal ions is coordinated by a phosphate oxygen, thus positioning the substrate and the nucleophile in line. Compared to the position in the absence of substrate, the nucleophile is actually further back, which seems rather surprising at first. The most likely explanation is that the binding of substrate generates some conformational strain, ' cocking' the nucleophile for hydrolysis. The backward movement of the nucleophile upon substrate binding can also be interpreted as evidence for a more dissociative mechanism, because the nucleophile is so tightly bound by the three metal ions that it is actually less activated than in Family I. Therefore a more activated electrophile is needed.

The more open and asymmetric conformation of the PNP molecule in the fluoride complex could also be due to partial dissociation of the P1-N bond, which would suggest that the leaving group is at least partly leaving before nucleophilic attack.

The problem with this configuration is that the nucleophile is on the opposite side of the tri-metal plane from the electrophile and all three lone pairs of the hydroxide are involved in metal coordination. Nucleophilic attack thus requires a change in coordination at at least one of the metal ions. The most likely candidate for this is M2, which must be a transition metal cation able to adopt both five- and six-coordinated states (74) (section 1.2.2). The conformational strain of the six-coordinated state in M2 in the presence of substrate could help in the nucleophilic attack in addition to speeding product release.

PPases must bind phosphate moieties relatively tightly, but ideally they should bind pyrophosphate – the substrate – tighter than phosphate – the product. The difference between them, however, is small enough to make this difficult for the enzyme. The solutions adopted by Family I and II PPases respectively are quite different. The electrostatic repulsion between the two product phosphates makes the expulsion of the first phosphate quite easy, but the second one binds more tightly. In both familes the electrophilic phosphate is the first to leave (46, 76).

The Family I enzyme only has one domain, so the main chain motions are relatively small, but the movement of the positively charged side chains at the 'leaving group end' of the active site sweeps the phosphate into a groove lined with positive charge. This groove probably provides multiple thermodynamically accessible binding sites for phosphate, facilitating its exit.

In Family II PPases, on the other hand, electrostatic repulsion also propels the opening of the two domains (74), separating the phosphate binding sites. The individual phosphate binding affinities of each site would then be low enough that the phosphates can dissociate, even in the presence of $Mn^{2+}$.

## 4.3 THE ROLE OF HYDROGEN BOND NETWORKS

The traditional approach to understanding enzyme mechanism has been to identify the 'catalytic' functionalities, whether amino acid residues or cofactors, and to dissect their influence by mutating each one in turn. As discussed in section 3.1.4, the effects of such mutations can be rather surprising and far-reaching. This points towards a more synergistic view of the catalytic power of enzymes; the active site is more than the sum of its parts. Hydrogen bonding networks are crucial not only in holding proteins together and positioning the substrate, but also in relaying and delocalising the charges that form during catalysis. The positions of water molecules are remarkably conserved in the different structures of Family I PPase, which is not very surprising as these waters are relatively well ordered. When mutations disturb these networks, they tend to rearrange in a way that provides hydrogen-bonding partners to all the polar residues (61). This is hardly surprising from a thermodynamic point of view, because even though the active site is not buried from the solvent, the solvent molecules are held tightly in place by the many charged groups. Therefore the mutation of an aspartate to glutamate, for example, cannot simply move the water molecule that is hydrogen-bonded to the carboxylate, but the whole network has to re-arrange. This means that the side chains, metal ions and water molecules crucial to the catalytic machinery have to move as well, seriously disrupting the function of the active site.

Another important function of hydrogen-bond networks is the shuffling of protons. The catalytic mechanism involves changes of protonation states of the active site side chains and water molecules, as well as the substrate/product itself. Protons diffuse in bulk water by the Grotthus mechanism and move much faster than the water molecules themselves. This requires a chain or a network of hydrogen bonds. In the Y-PPase active site this is an issue in the nucleophile activation (section 4.1.1) since a water molecule becomes a hydroxide when an another water is expelled. The leaving water molecule that is coordinating M2 is connected to an extended network of ordered water molecules, which can quickly delocalise the charge of a proton.

Identifying hydrogen bonds and hydrogen-bonding networks is usually relatively simple from an X-ray crystal structure – provided that the resolution is sufficient to define the side chains positions and ordered water molecules clearly ($\leq$2.5 Å). Hydrogen bonds can be located by 2.5-3.2 Å distances between potential donor and acceptor atoms in suitable geometries. This, however, does not yet define whether a given atom is a hydrogen bond acceptor or donor. Some atoms, like amide nitrogens, can be confidently assigned to only act as donors while others, such as amide carbonyls, only act as acceptors. For most hydrogen bonds, however, it is not evident which atom is the donor and which the acceptor. In order to understand the role of the hydrogen bonding network, these proton positions would have to be known.

## 4.4 OPEN QUESTIONS – WHERE ARE THE PROTONS?

Some of the protonation states of active site residues can of course be readily inferred, such as the aspartates that coordinate a metal ion and must therefore be deprotonated. The most interesting carboxylate, D117, which orients the nucleophile and possibly also deprotonates it (section 4.1.1), is also likely to be deprotonated at least in the closed conformation. If we accept that the nucleophile $O_{Nu}$ is an hydroxide ion (section 4.1.1), its only hydrogen bonding partner is D117, which therefore has to act as an acceptor. The basic side chains of lysines and arginines in the active site, such as R78 and K193 are probably protonated, particularly because they form hydrogen bonds to the phosphate oxygens, where they act as donors. The exception is K56, which is also among the most mobile of the active site side chains based on the variant structures. It is hydrogen bonded to E58, which in turn coordinates M3, so E58 must therefore be deprotonated. Although deprotonated lysine has been observed in the neutron structure of xylose isomerase at pH 8 (159), this possibility seems unlikely in Y-PPase, as a protonated lysine could form an ion pair with E58.

The protonation state of the pyrophosphate/phosphates is an interesting question, since the dominant solution species is not the fully deprotonated $P_2O_7^{4-}$ or $PO_4^{3-}$. In the active site, however, there are no obvious oxygens that could be protonated based on their hydrogen bonding partners, if the above assumptions about lysine, arginine and metal-coordinating aspartate protonation are correct. In the substrate complex (1E6A) there are only two phosphate oxygens hydrogen bonded to tyrosines: Y93 and Y192. While tyrosine is unlikely to be deprotonated, the phenolic oxygen may act as a hydrogen bond acceptor if the phenolic proton forms a hydrogen bond to another partner. Both of these hydrogen bonds between tyrosine and phosphate are very short (2.5 Å), which indicates that the proton may be shared between the oxygens. What other acceptors could these two tyrosines then form hydrogen bonds with? Close to Y93 there are no reasonable candidates, the closest being the D120 carboxylate at a distance 3.2 Å. The geometry is not very favourable for a hydrogen bond either, because the phenolic oxygen is perpendicular to the carboxylate plane. Y192 on the other hand forms a hydrogen bond to a water molecule, which could act as an acceptor. If any of the oxygens of the leaving group phosphate are protonated, this oxygen is the most likely candidate.

# 5. CONCLUSIONS

Pyrophosphatases are thoroughly studied enzymes, but our understanding of their mechanism remains incomplete – as indeed of any other enzyme. My work presented here has deepened that understanding by providing a unified view of the catalytic cycle of eukaryotic Family I PPases and their differences from the prokaryotic Family I PPases. Analysis of the effects of mutations has also revealed a number of aspects of the catalytic machinery. My work shows how important it is to solve the structures of the variant enzymes, as the changes caused by even conservative mutations turned out to be rather unexpected.

The structure of *B. subtilis* PPase with the substrate analogue revealed a novel aspect of the mechanism of Family II PPases and has allowed a better comparison between the two families. A more detailed analysis of the metal content was instrumental in interpreting the structure.

Time-resolved experiments are a logical next step in the investigation of enzyme mechanism. They require in practice a photolysable substrate precursor. I synthesised a candidate molecule and the photolysis kinetics were suitable, but unfortunately it was hydrolysed by both yeast and *T. maritima* PPases.

The neutron diffraction of Y-PPase is a very challenging project that stretched the limits of the available instrumentation. Collecting a data set required extensive effort, and the methodological developments that arose from this project will certainly benefit the field of neutron protein crystallography in pursuing more and more difficult targets.

# ACKNOWLEDGEMENTS

This work was carried out between 2005 and 2009 at the Institute of Biotechnology, University of Helsinki and the European Molecular Biology Laboratory (EMBL) Grenoble Outstation. I would like to thank Prof. Mart Saarma, the director of the Insitute of Biotechnology for providing the excellent research facilities and a stimulating working environment. My thanks also go to Dr. Stephen Cusack, the head of the EMBL Grenoble Outstation, for making possible my visit to Grenoble and kindly covering the chemicals expenses for the PPase perdeuteration. The National Graduate School of Informational and Structural Biology (ISB) financed the work, and perhaps even more importantly, has organised a number of excellent meetings. I would like to thank Prof. Mark Johnson and all the members of ISB for creating a very friendly and supportive atmosphere.

My greatest thanks go to my supervisor Prof. Adrian Goldman, who has always having faith in me as well as the project despite the many setbacks. I am honoured to have been his student. I am also grateful for Dr. Monika Budayova-Spano for all her efforts to make the PPase neutron project work and all the great discussions we have had. I have always admired Monika's determination and kindness.

I would like to thank Prof. Ilkka Kilpeläinen for always arranging things for me at the Chemistry department and for being a very insightful member of my thesis committee. I thank also Prof. Pirkko Heikinheimo, the third member of my thesis committee, for all her support, her interest in my work and good advice.

I thank Prof. Peter Kahn and Prof. Carrie Wilmot for reviewing this thesis and their helpful comments.

During these years I have had many great colleagues in the Goldman and Heikinheimo labs, and their help and support and the good working atmosphere in the group has been extremely important to me. I would specially like to thank Veli-Pekka for introducing me to crystallography; Igor, Veli-Matti and Ansko for sharing an office with me and Lari for help with computers and synchrotron trips. I am grateful for the people I have had the pleasure to work with: Andrzej. Arnab, Chiara, Danielle, Ellen, Heli, Heidi, Jaana, Juho, Katja, Maria, Marika, Mika, Sanjay, Saurabh, Sirpa, Tommi, Vimal and Wesley.

I am grateful to Vesa Tuominen, my predecessor in the PPase neutron project for introducing me to the subject and all the good advice. I also thank Heidi Tuominen for all the help with PPases.

During my longer and shorter visits to Grenoble I have also had the chance to work with many wonderful people. I would like to specially thank Dr. Matthew Blakeley for all the help with the many attempts at neutron data collection, Marie-Thérèse Dauvergne for helping me with the wet lab work and François Dauvergne for the all the technical help. I thank everybody in the D-Lab: Jean-Baptiste, Maxime, Michael, Ricardo, Suzy, Trevor and everyone else for a great time in France.

I owe my deepest gratitude to my parents Ulpu and Pekka, who have always supported me in all my studies and set an example for me.

Finally I would like to express my deepest love and gratitude to Lotta, who has supported me throughout the writing process.

Helsinki, November 2009

# REFERENCES


(1) Eyring, H. "The activated complex and the absolute rate of chemical reactions". *Chem. Rev.* (1935), **17**, 65-77.

(2) Knowles, J. "Enzyme-catalyzed phosphoryl transfer reactions". *Annu. Rev. Biochem.* (1980), **49**, 877-919.

(3) Abrahams, J. P., Leslie, A. G. W., Lutter, R., Walker, J. E. "Structure at 2.8 Å resolution of $F_1$-ATPase from bovine heart mitochondria". *Nature* (1994), **370**, 621-628.

(4) Noble, M. E. M., Endicott, J. A., Johnson, L. N. "Protein kinase inhibitors: insights into drug design from structure". *Science* (2004), **303**, 1800-1805.

(5) Michaelis, L., Menten, M. L. "Die kinetik der invertinwirkung". *Biochem. Z.* (1913), **49**, 148.

(6) Fersht, A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*; WH Freeman: New York, NY, 1998;

(7) Wikström, M. "Cytochrome c oxidase: 25 years of the elusive proton pump". *BBA-Bioenergetics* (2004), **1655**, 241-247.

(8) Belevich, I., Verkhovsky, M. I., Wikström, M. "Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase". *Nature* (2006), **440**, 829-832.

(9) Friesner, R. A., Guallar, V. "*Ab initio* quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis". *Ann. Rev. Phys. Chem.* (2004), **56**, 389-427.

(10) Hu, H., Yang, W. "Free energies of chemical reactions in solution and in enzymes with *ab initio* quantum mechanics/molecular mechanics methods". *Ann. Rev. Phys. Chem.* (2008), **59**, 573-601.

(11) Hartley, B. S., Kilby, B. A. "The reaction of *p*-nitrophenyl esters with chymotrypsin and insulin". *Biochem. J.* (1954), **56**, 288.

(12) Hedström, L. "Serine protease mechanism and specificity". *Chem. Rev.* (2002), **102**, 4501-4524.

(13) Snijder, H. J., Ubarretxena-Belandia, I., Blaauw, M., Kalk, K. H., Verheij, H. M., Egmond, M. R., Dekker, N., Dijkstra, B. W. "Structural evidence for dimerization-regulated activation of an integral membrane phospholipase". *Nature* (1999), **401**, 717-721.

(14) Burke Jr., T. R., Zhang, Z. Y. "Protein-tyrosine phosphatases: structure, mechanism, and inhibitor discovery". *Biopolymers* (1998), **47**,

(15) Okar, D. A., Kakalis, L. T., Narula, S. S., Armitage, I. M., Pilkis, S. J. "Identification of transient intermediates in the bisphosphatase reaction of rat liver 6-phosphofructo-2-kinase/fructose-2, 6-bisphosphatase by $^{31}$P-NMR spectroscopy". *Biochem. J.* (1995), **308**, 189.

(16) Denu, J. M., Lohse, D. L., Vijayalakshmi, J., Saper, M. A., Dixon, J. E. "Visualization of intermediate and transition-state structures in protein-tyrosine phosphatase catalysis". *Proc. Natl. Acad. Sci. USA* (1996), **93(6)**, 2493-2498.

(17) Sperow, J. W., Moe, O. A., Ridlington, J. W., Butler, L. G. "Yeast Inorganic Pyrophosphatase VI. STUDIES ON SPECIFICITY AND MECHANISM". *J. Biol. Chem.* (1973), **248**, 2062-2065.
(18) Cleland, W. W. "Use of isotope effects to determine enzyme mechanisms". *J. Labelled Compd. Radiopharm.* (2007), **50**,
(19) Mildvan, A. S. "Mechanisms of signaling and related enzymes". *Proteins: Structure, Function, and Genetics* (1997), **29**,
(20) Pauling, L. *The nature of the chemical bond and the structure of molecules and crystals: An introduction to modern structural chemistry*; Cornell University Press: Ithaca, NY, 1960;
(21) Cannon, W. R., Singleton, S. F., Benkovic, S. J. "A perspective on biological catalysis". *Nat. Struct. Biol.* (1996), **3**, 821-833.
(22) Rajagopalan, P. T., Benkovic, S. J. "Preorganization and Protein Dynamics in Enzyme Catalysis". *The Chemical Record* (2002), **2**, 24-36.
(23) Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T., Benkovic, S. J., Hammes-Schiffer, S. "Network of coupled promoting motions in enzyme catalysis". *Proc. Natl. Acad. Sci. USA* (2002), **99**, 2794-2799.
(24) Warshel, A., Sharma, P. K., Kato, M., Xiang, Y., Liu, H., Olsson, M. H. "Electrostatic basis for enzyme catalysis". *Chem. Rev.* (2006), **106**, 3210-3235.
(25) Warshel, A. "Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites". *J. Biol. Chem.* (1998), **273**, 27035-27038.
(26) Klahn, M., Rosta, E., Warshel, A. "On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins". *J. Am. Chem. Soc.* (2006), **128**, 15310-15323.
(27) Warshel, A. "Computer simulations of enzyme catalysis: methods, progress, and insights". *Ann. Rev. Biophys. Biomol. Struct.* (2003), **32**, 425-443.
(28) McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., Withers, S. G. "The $pK_a$ of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a $^{13}C$-NMR study of *Bacillus circulans* xylanase". *Biochemistry* (1996), **35**, 9958-9966.
(29) Admiraal, S. J., Herschlag, D. "Mapping the transition state for ATP hydrolysis: implications for enzymatic catalysis". *Chem. Biol.* (1995), **2**, 729-739.
(30) Mitic, N., Smith, S. J., Neves, A., Guddat, L. W., Gahan, L. R., Schenk, G. "The catalytic mechanisms of binuclear metallohydrolases". *Chem. Rev.* (2006), **106**, 3338-3363.
(31) Brautigam, C. A., Steitz, T. A. "Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes". *Curr. Opin. Struct. Biol.* (1998), **8**, 54-63.
(32) Rosta, E., Kamerlin, S. C., Warshel, A. "On the interpretation of the observed linear free energy relationship in phosphate hydrolysis: a thorough computational study of phosphate diester hydrolysis in solution". *Biochemistry* (2008), **47**, 3725-3735.
(33) Kamerlin, S. C., Florian, J., Warshel, A. "Associative versus dissociative mechanisms of phosphate monoester hydrolysis: on the interpretation of activation entropies". *Chemphyschem* (2008), **9**, 1767-1773.

(34) Lahiri, S. D., Zhang, G., Dunaway-Mariano, D., Allen, K. N. "The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction". *Science* (2003), **299**, 2067-2071.
(35) Yousef, M. S., Fabiola, F., Gattis, J. L., Somasundaram, T., Chapman, M. S. "Refinement of the arginine kinase transition-state analogue complex at 1.2 Å resolution: mechanistic insights". *Acta Cryst.* (2002), **D58**, 2009-2017.
(36) Lahiri, S. D., Wang, P. F., Babbitt, P. C., McLeish, M. J., GL. "The 2.1 Å Structure of *Torpedo californica* Creatine Kinase Complexed with the ADP-$Mg^{2+-}$ $NO_3^{2"\cdot}$ *Biochemistry* (2002),
(37) Flodgaard, H., Fleron, P. "Thermodynamic parameters for the hydrolysis of inorganic pyrophosphate at pH 7.4 as a function of ($Mg^{2+}$), ($K^{+}$), and ionic strength determined from equilibrium studies of the reaction". *J. Biol. Chem.* (1974), **249**, 3465-3474.
(38) Kunitz, M. "Crystalline Inorganic Pyrophosphatase Isolated from Baker's Yeast". *J. Gen. Physiol.* (1952), **35**, 423-450.
(39) Baykov, A. A., Cooperman, B. S., Goldman, A., Lahti, R. "Cytoplasmic inorganic pyrophosphatase". *Prog. Mol. Subcell. Biol.* (1999), **23**, 127-150.
(40) Shintani, T., Uchiumi, T., Yonezawa, T., Salminen, A., Baykov, A. A., Lahti, R., Hachimori, A. "Cloning and expression of a unique inorganic pyrophosphatase from *Bacillus subtilis*: evidence for a new family of enzymes". *FEBS Lett.* (1998), **439**, 263-266.
(41) Baltscheffsky, M., Schultz, A., Baltscheffsky, H. "$H^{+}$-proton-pumping inorganic pyrophosphatase: a tightly membrane-bound family". *FEBS Lett.* (1999), **452**, 121-127.
(42) de Graaf, B. H. J., Rudd, J. J., Wheeler, M. J., Perry, R. M., Bell, E. M., Osman, K., Franklin, F. C. H., Franklin-Tong, V. E. "Self-incompatibility in Papaver targets soluble inorganic pyrophosphatases in pollen". *Nature* (2006), **444**, 490.
(43) Belogurov, G. A., Malinen, A. M., Turkina, M. V., Jalonen, U., Rytkonen, K., Baykov, A. A., Lahti, R. "Membrane-bound pyrophosphatase of *Thermotoga maritima* requires sodium for activity". *Biochemistry* (2005), **44**, 2088-2096.
(44) Malinen, A. M., Belogurov, G. A., Baykov, A. A., Lahti, R. "$Na^{+}$-Pyrophosphatase: A Novel Primary Sodium Pump". *Biochemistry* (2007), **46**, 8872-8878.
(45) Kay, H. D. "The phosphatases of mammalian tissues: Pyrophosphatase.". *Biochem. J.* (1928), **22**, 1446-1448.
(46) Springs, B., Welsh, K. M., Cooperman, B. S. "Thermodynamics, Kinetics, and Mechanism in Yeast Inorganic Pyrophosphatase Catalysis in Inorganic Pyrophosphate: Inorganic Phosphate Equilibration". *Biochemistry* (1981), **20**, 6384-6391.
(47) Halonen, P., Baykov, A. A., Goldman, A., Lahti, R., Cooperman, B. S. "Single-Turnover Kinetics of *Saccharomyces cerevisiae* Inorganic Pyrophosphatase". *Biochemistry* (2002), **41**, 12025-12031.
(48) Belogurov, G. A., Fabrichniy, I. P., Pohjanjoki, P., Kasho, V. N., Lehtihuhta, E., Turkina, M. V., Cooperman, B. S., Goldman, A., Baykov, A. A., Lahti, R.

"Catalytically Important Ionizations along the Reaction Pathway of Yeast Pyrophosphatase". *Biochemistry* (2000), **39**, 13931-13938.
(49) Baykov, A., Shestakov, A. S., Kasho, V. N., Vener, A. V., Ivanov, A. I. "Kinetics and thermodynamics of catalysis by the inorganic pyrophosphatase of *Escherichia coli* in both directions". *Eur. J. Biochem.* (1990), **194**, 879-887.
(50) Fabrichniy, I. P., Lahti, R., Baykov, A. A. "Effect of Asp-97->Glu substitution on pH dependence of catalysis by inorganic pyrophosphatase of *Escherichia coli*". *Biochemistry - Moscow* (1997), **62**, 946-950.
(51) Cooperman, B. S. "The Mechanism of Action of Yeast Inorganic Pyrophosphatase". *Methods Enzymol.* (1982), **87**, 526-548.
(52) Cooperman, B. S., Panackal, A., Springs, B., Hamm, D. J. "Divalent metal ion, inorganic phosphate, and inorganic phosphate analogue binding to yeast inorganic pyrophosphatase". *Biochemistry* (1981), **20**, 6051-6060.
(53) Bunick, G., McKenna, G. P., Colton, R., Voet, D. "The X-ray Structure of Yeast Inorganic Pyrophosphatase, crystal properties". *J. Biol. Chem.* (1974), **249**, 4647-4649.
(54) Harutyunyan, E., Terzyan, S. S., Voronova, A. A., Kuranova, I. P., Smirnova, E. A., Vainshtein, B. K., Höhne, W., Hansen, G. "An X-ray study of yeast inorganic pyrophophosphatase at 3 Å resolution". *Dokl Akad Nauk SSSR* (1981), **258**, 1481-1485.
(55) Teplyakov, A., Obmolova, G., Wilson, K. S., Ishii, K., Kaji, H., Samejima, T., Kuranova, I. "Crystal structure of inorganic pyrophosphatase from *Thermus thermophilus*". *Prot. Sci.* (1994), **3**, 1098-1107.
(56) Heikinheimo, P., Tuominen, V., Ahonen, A. K., Teplyakov, A., Cooperman, B. S., Baykov, A. A., Lahti, R., Goldman, A. "Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase". *Proc. Natl. Acad. Sci. USA* (2001), **98**, 3121.
(57) Heikinheimo, P., Lehtonen, J., Baykov, A., Lahti, R., Cooperman, B. S., Goldman, A. "The structural basis for pyrophosphatase catalysis". *Structure* (1996), **4**, 1491-1508.
(58) Samygina, V. R., Popov, A. N., Rodina, E. V., Vorobyeva, N. N., Lamzin, V. S., Polyakov, K. M., Kurilova, S. A., Nazarova, T. I., Avaeva, S. M. "The structures of *Escherichia coli* inorganic pyrophosphatase complexed with $Ca^{2+}$ or $CaP_i$ at atomic resolution and their mechanistic implications". *J. Mol. Biol.* (2001), **314**, 633-645.
(59) Avaeva, S. et al. "Crystal sructure of *Escherichia coli* inorganic pyrophosphatase complexed with $SO_4^{2-}$". *FEBS Lett.* (1997), **410**, 502-508.
(60) Harutyunyan, E. H., Kuranova, I. P., Vainshtein, B. K., Höhne, W. E., Lamzin, V. S., Dauter, Z., Teplyakov, A. V., Wilson, K. S. "X-ray structure of yeast inorganic pyrophosphatase complexed with manganese and phosphate". *Eur. J. Biochem.* (1996), **239**, 220-228.
(61) Oksanen, E., Ahonen, A. K., Tuominen, H., Tuominen, V., Lahti, R., Goldman, A., Heikinheimo, P. "A complete structural description of the catalytic cycle of yeast pyrophosphatase". *Biochemistry* (2007), **46**, 1228-1239.

(62) Tuominen, V., Heikinheimo, P., Kajander, T., Torkkel, T., Hyytiä, T., Käpylä, J., Lahti, R., Cooperman, B. S., Goldman, A. "The R78K and D117E active-site variants of *Saccharomyces cerevisiae* soluble inorganic pyrophosphatase: structural studies and mechanistic implications". *J. Mol. Biol.* (1998), **284**, 1565-1580.
(63) Lahti, R., Kolakowski, L. F., Jr., Heinonen, J., Vihinen, M., Pohjanoksa, K., Cooperman, B. S. "Conservation of Functional Residues Between Yeast and *E. coli* Inorganic Pyrophosphatase". *Biochim. Biophys. Acta* (1990), **1038**, 338-345.
(64) Salminen, T., Teplyakov, A., Kankare, J., Cooperman, B. S., Lahti, R., Goldman, A. "An unusual route to thermostability disclosed by the comparison of *Thermus thermophilus* and *Escherichia coli* inorganic pyrophosphatases". *Prot. Sci.* (1996), **5**, 1014-1025.
(65) 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. "Crystal structure of holo inorganic pyrophosphatase from *Escherichia coli* at 1.9 angstrom resolution. Mechanism of hydrolysis". *Biochemistry* (1997), **36**, 7754-7760.
(66) Salminen, T., Käpylä, J., Heikinheimo, P., Kankare, J., Goldman, A., Heinonen, J., Baykov, A., Cooperman, B., Lahti, R. "Structure and function analysis of *Escherichia coli* inorganic pyrophosphatase: Is an hydroxide ion the key catalytic residue?". *Biochemistry* (1995), **34**, 782-791.
(67) Baykov, A. A., Fabrichniy, I. P., Pohjanjoki, P., Zyryanov, A. B., Lahti, R. "Fluoride Effects along the Reaction Pathway of Pyrophosphatase: Evidence for a Second Enzyme. Pyrophosphate Intermediate". *Biochemistry* (2000), **39**, 11939-11947.
(68) Zyryanov, A. B., Pohjanjoki, P., Kasho, V. N., Shestakov, A. S., Goldman, A., Lahti, R., Baykov, A. A. "The electrophilic and leaving group phosphates in the catalytic mechanism of yeast pyrophosphatase". *J. Biol. Chem.* (2001), **276**, 17629-17634.
(69) Knight, W. B., Fitts, S. W., Dunaway-Mariano, D. "Investigation of the catalytic mechanism of yeast inorganic pyrophosphatase.". *Biochemistry* (1981), **20**, 4079-4086.
(70) Young, T. W., Kuhn, N. J., Wadeson, A., Ward, S., Burges, D., Cooke, G. D. "*Bacillus subtilis* ORF yybQ encodes a manganese-dependent inorganic pyrophosphatase with distinctive properties: the first of a new class of soluble pyrophosphatase?". *Microbiology* (1998), **144**, 2563-2571.
(71) Murzin, A. G., Brenner, S. E., Hubbard, T., Chothia, C. "SCOP: A structural classification of proteins database for the investigation of sequences and structures". *J. Mol. Biol.* (1995), **247**, 536-540.
(72) Merckel, M. C., Fabrichniy, I. P., Salminen, A., Kalkkinen, N., Baykov, A. A., Lahti, R., Goldman, A. "Crystal structure of *Streptococcus mutans* pyrophosphatase: a new fold for an old mechanism". *Structure* (2001), **9**, 289-297.

(73) Ahn, S., Milner, A. J., Futterer, K., Konopka, M., Ilias, M., Young, T. W., White, S. A. "The "open" and "closed" structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii*". *J. Mol. Biol.* (2001), **313**, 797-811.
(74) Fabrichniy, I. P., Lehtiö, L., Salminen, A., Zyryanov, A. B., Baykov, A. A., Lahti, R., Goldman, A. "Structural Studies of Metal Ions in Family II Pyrophosphatases: The Requirement for a Janus Ion". *Biochemistry* (2004), **43**, 14403-14411.
(75) Fabrichniy, I. P., Lehtiö, L., Tammenkoski, M., Zyryanov, A. B., Oksanen, E., Baykov, A. A., Lahti, R., Goldman, A. "A trimetal site and substrate distortion in a family II inorganic pyrophosphatase". *J. Biol. Chem.* (2007), **282**, 1422-1431.
(76) Zyryanov, A. B., Vener, A. V., Salminen, A., Goldman, A., Lahti, R., Baykov, A. A. "Rates of elementary catalytic steps for different metal forms of the family II pyrophosphatase from *Streptococcus gordonii*". *Biochemistry* (2004), **43**, 1065-1074.
(77) Parfenyev, A. N., Salminen, A., Halonen, P., Hachimori, A., Baykov, A. A., Lahti, R. "Quaternary structure and metal ion requirement of family II pyrophosphatases from *Bacillus subtilis*, *Streptococcus gordonii*, and *Streptococcus mutans*". *J. Biol. Chem.* (2001), **276**, 24511-24518.
(78) Bock, C. W., Katz, A. K., Markham, G. D., Glusker, J. P. "Manganese as a Replacement for Magnesium and Zinc: Functional Comparison of the Divalent Ions". *J. Am. Chem. Soc.* (1999), **121**, 7360-7372.
(79) Stoddard, B. L., Farber, G. K. "Direct measurement of reactivity in the protein crystal by steady-state kinetic studies". *Structure* (1995), **3**, 991-996.
(80) Stoddard, B. L. "Intermediate trapping and Laue X-ray diffraction: Potential for enzyme mechanism, dynamics, and inhibitor screening". *Pharmacol. Therapeut.* (1996), **70**, 215-256.
(81) Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T., Kondo, T. "Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro". *Science* (2005), **308**, 414-415.
(82) Fulop, V., Phizackerley, R. P., Soltis, S. M., Clifton, I. J., Wakatsuki, S., Erman, J., Hajdu, J., Edwards, S. L. "Laue Diffraction Study on the Structure of Cytochrome c Peroxidase Compound-I". *Structure* (1994), **2**, 201-208.
(83) Singer, P. T., Smalås, A., Carty, R. P., Mangel, W. F., Sweet, R. M. "The Hydrolytic Water Molecule in Trypsin, Revealed by Time-Resolved Laue Crystallography". *Science* (1993), **259**, 669-673.
(84) McCray, J. A., Herbette, L., Kihara, T., Trentham, D. R. "A new approach to time-resolved studies of ATP-requiring biological systems; laser flash photolysis of caged ATP". *Proc. Natl. Acad. Sci. USA* (1980), **77**, 7237-7241.
(85) Walker, J. W., Reid, G. P., Trentham, D. R. "Synthesis and properties of caged nucleotides". *Methods Enzymol.* (1989), **172**, 288-301.
(86) Walker, J. W., Reid, G. P., McCray, J. A., Trentham, D. R. "Photolabile 1-(2-nitrophenyl)ethyl phosphate esters of adenine nucleotide analogs. Synthesis and mechanism of photolysis". *J. Am. Chem. Soc.* (1988), **110**, 7170-7177.

(87) Ursby, T., Weik, M., Fioravanti, E., Delarue, M., Goeldner, M., Bourgeois, D. "Cryophotolysis of caged compounds: a technique for trapping intermediate states in protein crystals". *Acta Cryst.* (2002), **D58**, 607-614.
(88) Il'ichev, Y. V., Schworer, M. A., Wirz, J. "Photochemical Reaction Mechanisms of 2-Nitrobenzyl Compounds: Methyl Ethers and Caged ATP". *J. Am. Chem. Soc.* (2004), **126**, 4581-4595.
(89) Ramesh, D., Wieboldt, R., Niu, L., Carpenter, B. K., Hess, G. P. "Photolysis of a protecting group for the carboxyl function of neurotransmitters within 3 microseconds and with product quantum yield of 0.2". *Proc. Natl. Acad. Sci. USA* (1993), **90**, 11074-11078.
(90) Wieboldt, R., Gee, K. R., Niu, L., Ramesh, D., Carpenter, B. K., Hess, G. P. "Photolabile precursors of glutamate: synthesis, photochemical properties, and activation of glutamate receptors on a microsecond time scale". *Proc. Natl. Acad. Sci. USA* (1994), **91**, 8752-8756.
(91) Gee, K. R., Kueper, L. W., Barnes, J., Dudley, G., Givens, R. S. "Desyl Esters of Amino Acid Neurotransmitters. Phototriggers for Biologically Active Neurotransmitters". *J. Org. Chem.* (1996), **61**, 1228-1233.
(92) Banerjee, A., Grewer, C., Ramakrishnan, L., Jager, J., Gameiro, A., Breitinger, H.-G. A., Gee, K. R., Carpenter, B. K., Hess, G. P. "Toward the Development of New Photolabile Protecting Groups That Can Rapidly Release Bioactive Compounds upon Photolysis with Visible Light". *J. Org. Chem.* (2003), **68**, 8361-8367.
(93) Wu, N., Deiters, A., Cropp, T. A., King, D., Schultz, P. G. "A genetically encoded photocaged amino acid". *J. Am. Chem. Soc.* (2004), **126**, 14306-14307.
(94) Royant, A., Carpentier, P., Ohana, J., McGeehan, J., Paetzold, B., Noirclerc-Savoye, M., Vernéde, X., Adam, V., Bourgeois, D. "Advances in spectroscopic methods for biological crystals. 1. Fluorescence lifetime measurements". *J. Appl. Cryst.* (2007), **40**, 1105–1112.
(95) Owen, R. L., Pearson, A. R., Meents, A., Boehler, P., Thominet, V., Schulze-Briese, C. "A new on-axis multimode spectrometer for the macromolecular crystallography beamlines of the Swiss Light Source". *J. Synchrotron Rad.* (2009), **16**, 173-182.
(96) Katona, G., Carpentier, P., Niviere, V., Amara, P., Adam, V., Ohana, J., Tsanov, N., Bourgeois, D. "Raman-assisted crystallography reveals end-on peroxide intermediates in a nonheme iron enzyme". *Science* (2007), **316**, 449-453.
(97) Arndt, U. W., Wonacott, A. J. *The Rotation method in crystallography : data collection from macromolecular crystals*; Elsevier/North-Holland: Amsterdam, 1977;
(98) Flot, D. et al. "The care and nurture of undulator data sets.". *Acta Cryst.* (2006), **D62**, 65-71.
(99) Schmidt, M., Pahl, R., Srajer, V., Anderson, S., Ren, Z., Ihee, H., Rajagopal, S., Moffat, K. "Protein kinetics: structures of intermediates and reaction mechanism from time-resolved x-ray data". *Proc. Natl. Acad. Sci. USA* (2004), **101**, 4799-4804.

(100) Ren, Z., Bourgeois, D., Helliwell, J. R., Moffat, K., Srajer, V., Stoddard, B. L. "Laue crystallography: coming of age". *J. Synchrotron Rad.* (1999), **6**, 891-917.
(101) Duke, E. M. H., Wakatsuki, S., Hadfield, A., Johnson, L. N. "Laue and monochromatic diffraction studies on catalysis in phosphorylase b crystals". *Prot. Sci.* (1994), **3**, 1178-1196.
(102) Schlichting, I. et al. "Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis". *Nature* (1990), **345**, 309-315.
(103) Stoddard, B. L., Koenigs, P., Porter, N., Petratos, K., Petsko, G. A., Ringe, D. "Observation of the light-triggered binding of pyrone to chymotrypsin by Laue x-ray crystallography". *Proc. Natl. Acad. Sci. USA* (1991), **88**, 5503-5507.
(104) Bolduc, J. M., Dyer, D. H., Scott, W. G., Singer, P., Sweet, R. M., Koshland, D. E., Stoddard, B. L. "Mutagenesis and Laue structures of enzyme intermediates: Isocitrate dehydrogenase". *Science* (1995), **268**, 1312-1318.
(105) Helliwell, J. R. et al. "Time-resolved structures of hydroxymethylbilane synthase (Lys59Gln mutant) as it is loaded with substrate in the crystal determined by Laue diffraction". *J. Chem. Soc. Faraday Trans.* (1998), **94**, 2615-2622.
(106) Rajagopal, S., Schmidt, M., Anderson, S., Ihee, H., Moffat, K. "Analysis of experimental time-resolved crystallographic data by singular value decomposition". *Acta Cryst.* (2004), **D60**, 860-871.
(107) Scott, W. G., Murray, J. B., Arnold, J. R. P., Stoddard, B. L., Klug, A. "Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme". *Science* (1996), **274**, 2065-2069.
(108) Teng, T.-Y., Moffat, K. "Cooling Rates During Flash Cooling". *J. Appl. Cryst.* (1998), **31**, 252–257.
(109) Williamson, J. C., Zewail, A. H. "Structural femtochemistry: experimental methodology". *Proc. Natl. Acad. Sci. USA* (1991), **88**, 5021-5025.
(110) Wolfenden, R. "Analog approaches to the structure of the transition state in enzyme reactions". *Acc. Chem. Res.* (1972), **5**, 10-18.
(111) Wlodawer, A., Sjoelin, L. "Structure of ribonuclease A: results of joint neutron and x-ray refinement at 2.0-Å resolution". *Biochemistry* (1983), **22**, 2720-2728.
(112) Davies, D. R., Hol, W. G. "The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes". *FEBS Lett.* (2004), **577**, 315-321.
(113) Fisher, A. J., Smith, C. A., Thoden, J. B., Smith, R., Sutoh, K., Holden, H. M., Rayment, I. "X-ray Structures of the Myosin Motor Domain of *Dictyostelium discoideum* Complexed with MgADP.$BeF_x$ and MgADP.$AlF_4^{-"}$. *Biochemistry* (1995), **34**, 8960-8972.
(114) Fauman, E. B., Yuvaniyama, C., Schubert, H. L., Stuckey, J. A., Saper, M. A. "The X-ray crystal structures of *Yersinia* tyrosine phosphatase with bound tungstate and nitrate. Mechanistic implications". *J. Biol. Chem.* (1996), **271**, 18780-18788.
(115) Blakeley, M. P., Mitschler, A., Hazemann, I., Meilleur, F., Myles, D. A., Podjarny, A. "Comparison of hydrogen determination with X-ray and neutron

crystallography in a human aldose reductase-inhibitor complex". *Eur. Biophys. J.* (2006), **35**, 577-583.
(116) Giacovazzo, C., Monaco, H. L., Artioli, G., Viterbo, D., Ferraris, G., Gilli, G., Zanotti, G., Catti, M. *Fundamentals of Crystallography*; Oxford University Press: Oxford, 2002;
(117) Fisher, S. J., Helliwell, J. R., Khurshid, S., Govada, L., Redwood, C., Squire, J. M., Chayen, N. E. "An investigation into the protonation states of the C1 domain of cardiac myosin-binding protein C". *Acta Cryst.* (2008), **D64**, 658-664.
(118) Sheldrick, G. M. "A short history of *SHELX*". *Acta Cryst.* (2008), **A64**, 112–122.
(119) Engh, R. A., Huber, R. "Accurate Bond and Angle Parameters for X-Ray Protein-Structure Refinement". *Acta Cryst.* (1991), **A47**, 392-400.
(120) Ahmed, H. U., Blakeley, M. P., Cianci, M., Cruickshank, D. W., Hubbard, J. A., Helliwell, J. R. "The determination of protonation-states in proteins". *Acta Cryst.* (2007), **D63**, 906-922.
(121) Cruickshank, D. W. "Remarks about protein structure precision". *Acta Cryst.* (1999), **D55**, 583-601.
(122) Rule, G. S., Hitchens, T. K. *Fundamentals of protein NMR spectrosopy*; Springer: Dordrecht, 2006;
(123) Cavanagh, J. *Protein NMR spectroscopy : principles and practice*; Academic Press: Amsterdam ; Boston, 2007;
(124) Poon, D. K., Schubert, M., Au, J., Okon, M., Withers, S. G., McIntosh, L. P. "Unambiguous determination of the ionization state of a glycoside hydrolase active site lysine by $^{1}H$-$^{15}N$ heteronuclear correlation spectroscopy". *J. Am. Chem. Soc.* (2006), **128**, 15388-15389.
(125) Schubert, M., Poon, D. K., Wicki, J., Tarling, C. A., Kwan, E. M., Nielsen, J. E., Withers, S. G., McIntosh, L. P. "Probing electrostatic interactions along the reaction pathway of a glycoside hydrolase: histidine characterization by NMR spectroscopy". *Biochemistry* (2007), **46**, 7383-7395.
(126) Sondergaard, C. R., McIntosh, L. P., Pollastri, G., Nielsen, J. E. "Determination of electrostatic interaction energies and protonation state populations in enzyme active sites". *J. Mol. Biol.* (2008), **376**, 269-287.
(127) Sears, V. F. "Scattering lengths for neutrons". *International Tables for Crystallography* (2006**)**, Prince, E, **C**, 444-454.
(128) Baruchel, J. *Neutron and synchrotron radiation for condensed matter studies*; Springer-Verlag: Berlin, 1994;
(129) Cipriani, F., Castagna, J.-C., Lehmann, M. S., Wilkinson, C. "A large image-plate detector for neutrons". *Physica B: Condensed Matter* (1995), **213-214**, 975-977.
(130) Tanaka, I., Kurihara, K., Chatake, T., Niimura, N. "A high-performance neutron diffractometer for biological crystallography (BIX-3)". *J. Appl. Cryst.* (2002), **35**, 34–40.
(131) Kurihara, K., Tanaka, I., Refai, M., Muhammad, Ostermann, A., Niimura, N. "A new neutron single-crystal diffractometer dedicated for biological macromolecules (BIX-4)". *J. Synchrotron Rad.* (2004), **11**, 68–71.

(132) Langan, P., Greene, G., Schoenborn, B. P. "Protein crystallography with spallation neutrons: the user facility at Los Alamos Neutron Science Center". *J. Appl. Cryst.* (2004), **37**, 24–31.
(133) Ichiro, T., Katsuhiro, K., Katsuaki, T., Nobuo, N., Takashi, O., Kazuo, K., Takaaki, H., Tomoji, O. "Overview of a new biological neutron diffractometer (iBIX) in J-PARC". *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (2009), **600**, 161 - 163.
(134) Schultz, A. J., Thiyagarajan, P., Hodges, J. P., Rehm, C., Myles, D. A. A., Langan, P., Mesecar, A. D. "Conceptual design of a macromolecular neutron diffractometer (MaNDi) for the SNS". *J. Appl. Cryst.* (2005), **38**, 964–974.
(135) Blakeley, M., Langan, P., Niimura, N., Podjarny, A. "Neutron crystallography: opportunities, challenges, and limitations". *Curr. Opin. Struct. Biol.* (2008), **18**, 593-600.
(136) Blakeley, M. P., Ruiz, F., Cachau, R., Hazemann, I., Meilleur, F., Mitschler, A., Ginell, S., Afonine, P., Ventura, O. N., Cousido-Siah, A. "Quantum model of catalysis based on a mobile proton revealed by subatomic x-ray and neutron diffraction studies of h-aldose reductase". *Proc. Natl. Acad. Sci. USA* (2008), **105**, 1844.
(137) Ducruix, A., Giegé, R. *Crystallization of nucleic acids and proteins : a practical approach*; Oxford University Press: Oxford ; New York, 1999;
(138) Bergfors, T. M. *Protein crystallization*; International University Line: La Jolla, CA, 2009;
(139) Tuominen, V. U., Myles, D. A., Dauvergne, M. T., Lahti, R., Heikinheimo, P., Goldman, A. "Production and preliminary analysis of perdeuterated yeast inorganic pyrophosphatase crystals suitable for neutron diffraction". *Acta Cryst.* (2004), **D60**, 606-609.
(140) Chatake, T., Mizuno, N., Voordouw, G., Higuchi, Y., Arai, S., Tanaka, I., Niimura, N. "Crystallization and preliminary neutron analysis of the dissimilatory sulfite reductase D (DsrD) protein from the sulfate-reducing bacterium *Desulfovibrio vulgaris*". *Acta Cryst.* (2003), **D59**, 2306–2309.
(141) Maeda, M., Chatake, T., Tanaka, I., Ostermann, A., Niimura, N. "Crystallization of a large single crystal of cubic insulin for neutron protein crystallography". *J. Synchrotron Rad.* (2004), **11**, 41–44.
(142) Budayova-Spano, M., Dauvergne, F., Audiffren, M., Bactivelane, T., Cusack, S. "A methodology and an instrument for the temperature-controlled optimization of crystal growth". *Acta Cryst.* (2007), **D63**, 339-347.
(143) Teixeira, S. C. M. et al. "New sources and instrumentation for neutrons in biology". *Chem. Phys.* (2008), **345**, 133 - 151.
(144) Budayova-Spano, M., Lafont, S., Astier, J.-P., Ebel, C., Veesler, S. "Comparison of solubility and interactions of aprotinin (BPTI) solutions in $H_2O$ and $D_2O$". *J. Cryst. Growth* (2000), **217**, 311 - 319.
(145) Berns, D. S. "Studies of Completely Deuteriated Proteins.* II. Thermal Denaturation in $D_2O$". *Biochemistry* (1963), **2**, 1377-1380.

(146) Meilleur, F., Weiss, K. L., Myles, D. A. "Deuterium labeling for neutron structure-function-dynamics analysis". *Methods Mol. Biol.* (2009), **544**, 281-292.
(147) Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van, D., A., Mitschler, A., Myles, D. A. A., Podjarny, A. "High-resolution neutron protein crystallography with radically small crystal volumes: application of perdeuteration to human aldose reductase". *Acta Cryst.* (2005), **D61**, 1413–1417.
(148) Blakeley, M. P., Kalb, A. J., Helliwell, J. R., Myles, D. A. "The 15-K neutron structure of saccharide-free concanavalin A". *Proc. Natl. Acad. Sci. USA* (2004), **101**, 16405-16410.
(149) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Terwilliger, T. C. "PHENIX: building new software for automated crystallographic structure determination". *Acta Cryst.* (2002), **D58**, 1948-1954.
(150) Mustyakimov, M., Langan, P. "Copyright C-06, 104 Patch for CNS; nCNS an open source distribution patch for CNS for macromolecular structure refinementistry 30:1211–1221". (2007),
(151) Brunger, A. T. et al. "Crystallography & NMR system: A new software suite for macromolecular structure determination". *Acta Cryst.* (1998), **D54**, 905-921.
(152) Blum, M. M., Mustyakimov, M., Rüterjans, H., Kehe, K., BP. "positions and protonation states of diisopropyl fluorophosphatase by joint neutron and X-ray.". *Proc. Natl. Acad. Sci. USA* (2009),
(153) Schoenborn, B. P. "Neutron diffraction analysis of myoglobin". *Nature* (1969), **224**, 143-146.
(154) Myles, D. A. "Neutron protein crystallography: current status and a brighter future". *Curr. Opin. Struct. Biol.* (2006), **16**, 630-637.
(155) Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A. A., Cooper, J. B. "A Neutron Laue Diffraction Study of Endothiapepsin: Implications for the Aspartic Proteinase Mechanism". *Biochemistry* (2001), **40**, 13149-13157.
(156) Coates, L., Tuan, H. F., Tomanicek, S., Kovalevsky, A., Mustyakimov, M., Erskine, P., Cooper, J. "The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-ray Diffraction". *J. Am. Chem. Soc.* (2008), **130**, 7235-7237.
(157) Wladkowski, B. D., Svensson, L. A., Sjolin, L., Ladner, J. E. "Structure (1.3 Å) and Charge States of a Ribonuclease A- Uridine Vanadate Complex: Implications for the Phosphate Ester Hydrolysis Mechanism". *J. Am. Chem. Soc.* (1998),
(158) Hanson, B. L., Langan, P., Katz, A. K., Li, X., Harp, J. M., Glusker, J. P., Schoenborn, B. P., Bunick, G. J. "A preliminary time-of-flight neutron diffraction study of *Streptomyces rubiginosus* D-xylose isomerase". *Acta Cryst.* (2004), **D60**, 241-249.
(159) Katz, A. K., Li, X., Carrell, H. L., Hanson, B. L., Langan, P., Coates, L., Schoenborn, B. P., Glusker, J. P., Bunick, G. J. "Locating active-site hydrogen atoms in d-xylose isomerase: Time-of-flight neutron diffraction". *Proc. Natl. Acad. Sci. USA* (2006), **103**, 8342-8347.

(160) Kovalevsky, A. Y. et al. "Hydrogen location in stages of an enzyme-catalyzed reaction: time-of-flight neutron structure of D-xylose isomerase with bound D-xylulose". *Biochemistry* (2008), **47**, 7595-7597.
(161) Bennett, B., Langan, P., Coates, L., Mustyakimov, M., Schoenborn, B., Howell, E. E., Dealwis, C. "Neutron diffraction studies of *Escherichia coli* dihydrofolate reductase complexed with methotrexate". *Proc. Natl. Acad. Sci. USA* (2006), **103**, 18493.
(162) Basran, J., Casarotto, M. G., Barsukov, I. L., Roberts, G. C. "Role of the active-site carboxylate in dihydrofolate reductase: kinetic and spectroscopic studies of the aspartate 26-->asparagine mutant of the *Lactobacillus casei* enzyme". *Biochemistry* (1995), **34**, 2872-2882.
(163) Adachi, M. et al. "Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography". *Proc. Natl. Acad. Sci. USA* (2009), **106**, 4641-4646.
(164) Wang, Y. X., Freedberg, D. I., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y., Torchia, D. A. "Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272". *Biochemistry* (1996), **35**, 9945-9950.
(165) Smith, R., Brereton, I. M., Chai, R. Y., Kent, S. B. "Ionization states of the catalytic residues in HIV-1 protease". *Nat. Struct. Biol.* (1996), **3**, 946-950.
(166) Piana, S., Sebastiani, D., Carloni, P., Parrinello, M. "*Ab initio* molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site". *J. Am. Chem. Soc.* (2001), **123**, 8730-8737.
(167) Pohjanjoki, P., Lahti, R., Goldman, A., Cooperman, B. S. "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* (1998), **37**, 1754-1761.
(168) Heikinheimo, P., Pohjanjoki, P., Helminen, A., Tasanen, M., Cooperman, B. S., Goldman, A., Baykov, A., Lahti, R. "A site-directed mutagenesis study of *Saccharomyces cerevisiae* pyrophosphatase - Functional conservation of the active site of soluble inorganic pyrophosphatases". *Eur. J. Biochem.* (1996), **239**, 138-143.
(169) Otwinowski, Z., Minor, W. "Processing of X-ray diffraction data in oscillation mode". *Methods Enzymol.* (1997), **276**, 307-326.
(170) Evans, P. "Scaling and assessment of data quality". *Acta Cryst.* (2006), **D62**, 72-82.
(171) Vagin, A., Teplyakov, A. "MOLREP: an automated program for molecular replacement". *J. Appl. Cryst.* (1997), **30**, 1022-1025.
(172) Perrakis, A., Harkiolaki, M., Wilson, K. S., Lamzin, V. S. "ARP/wARP and molecular replacement". *Acta Cryst.* (2001), **D57**, 1445-1450.
(173) Murshudov, G. N., Vagin, A. A., Dodson, E. J. "Refinement of macromolecular structures by the maximum-likelihood method". *Acta Cryst.* (1997), **D53**, 240-255.

(174) Jones, T. A., Zou, J. Y., Cowan, S. W., Kjeldgaard, M. "Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models". *Acta Cryst.* (1991), **A47**, 110-119.
(175) Emsley, P., Cowtan, K. "Coot: model-building tools for molecular graphics". *Acta Cryst.* (2004), **D60**, 2126-2132.
(176) Kleywegt, G. J. "Experimental assessment of differences between related protein crystal structures". *Acta Cryst.* (1999), **D55**, 1878-1884.
(177) Heinonen, J. K., Lahti, R. J. "A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase". *Anal. Biochem.* (1981), **113**, 313-317.
(178) Baykov, A. A., Avaeva, S. M. "A simple and sensitive apparatus for continuous monitoring of orthophosphate in the presence of acid-labile compounds". *Anal. Biochem.* (1981), **116**, 1-4.
(179) Halonen, P., Tammenkoski, M., Niiranen, L., Huopalahti, S. "Effects of Active Site Mutations on the Metal Binding Affinity, Catalytic Competence, and Stability". *Biochemistry* (2005), **44**, 4004-4010.
(180) Kabsch, W. "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants". *J. Appl. Cryst.* (1993), **26**, 795-800.
(181) Tereshko, V., Wilds, C. J., Minasov, G., Prakash, T. P., Maier, M. A., Howard, A., Wawrzak, Z., Manoharan, M., Egli, M. "Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments". *Nucleic Acids Res.* (2001), **29**, 1208-1215.
(182) Studier, F. W., Rosenberg, A. H., Dunn, J. J., Dubendorff, J. W. "Use of T7 RNA polymerase to direct expression of cloned genes". *Methods Enzymol.* (1990), **185**, 60-89.
(183) Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D., Wilkinson, C. "LAUEGEN version 6.0 and INTLDM". *J. Appl. Cryst.* (1998), **31**, 496–502.
(184) Arzt, S., Campbell, J. W., Harding, M. M., Hao, Q., Helliwell, J. R. "LSCALE -- the new normalization, scaling and absorption correction program in the Daresbury Laue software suite". *J. Appl. Cryst.* (1999), **32**, 554–562.
(185) Pohjanjoki, P., Fabrichniy, I. P., Kasho, V. N., Cooperman, B. S., Goldman, A., Baykov, A. A., Lahti, R. "Probing essential water in yeast pyrophosphatase by directed mutagenesis and fluoride inhibition measurements". *J. Biol. Chem.* (2001), **276**, 434-441.
(186) Budayova-Spano, M., Bonnete, F., Ferte, N., El Hajji, M., Meilleur, F., Blakeley, M. P., Castro, B. "A preliminary neutron diffraction study of rasburicase, a recombinant urate oxidase enzyme, complexed with 8-azaxanthin". *Acta Cryst.* (2006), **F62**, 306-309.
(187) Zyryanov, A. B., Shestakov, A. S., Lahti, R., Baykov, A. A. "Mechanism by which metal cofactors control substrate specificity in pyrophosphatase". *Biochem. J.* (2002), **367**, 901-906.
(188) Tammenkoski, M., Benini, S., Magretova, N. N., Baykov, A. A., Lahti, R. "An unusual, His-dependent family I pyrophosphatase from *Mycobacterium tuberculosis*". *J. Biol. Chem.* (2005), **280**, 41819.

(189) Wu, C. A., Lokanath, N. K., Kim, D. Y., Park, H. J., Hwang, H. Y., Kim, S. T., Suh, S. W., Kim, K. K. "Structure of inorganic pyrophosphatase from *Helicobacter pylori*". *Acta Cryst.* (2005), **D61**, 1459-1464.
(190) Yang, L., Liao, R. Z., Yu, J. G., Liu, R. Z. "DFT Study on the Mechanism of *Escherichia coli* Inorganic Pyrophosphatase". *J. Phys. Chem. B* (2009), **113**,
(191) 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. "Reversible inhibition of *Escherichia coli* inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies". *J. Mol. Biol.* (2007), **366**, 1305-1317.