INSIGHTS INTO THE STRUCTURE-ACTIVITY RELATIONSHIPS OF ENZYMES IMPLICATED IN PHYTATE METABOLISM

Emma Jade Thompson

PhD Thesis

University of East Anglia

School of Biological Sciences

May 2024

This thesis is dedicated to my partner, Andrew Cole, without whom I'm fairly certain I would not have survived to the end.

Abstract

Phytases catalyse the stepwise hydrolysis of IP6 (phytate) to release inorganic phosphate along with a set of inositol polyphosphate intermediates. However, the low abundance of endogenous phytases in the gastrointestinal tracts of monogastric livestock limits the bioavailability of dietary phosphate. Exogenous microbial phytases play a key role in monogastric nutrition, where their augmentation to the feed improves phosphorus bioavailability thereby reducing environmental burden arising from excessive inorganic phosphate excretion. However, limitations of commercial phytases – particularly IP_4 accumulation and incomplete dephosphorylation - renders this an active area of research.

The overarching aim of this research was to bioengineer phytase variants with superior hydrolysis characteristics, with particular emphasis on improved processing of IP_4 to ameliorate this bottleneck. Several approaches were implemented, including building upon previous work on Multiple Inositol Polyphosphate Phosphatases (MINPPs) and investigating novel phytases for structure-activity insights.

Previously shown to be specificity hotspots, 3 different B-pocket residues of *Bt*MINPP (R275, Q276, K280) were mutagenized to generate 7 variants each, which were analysed for their relative IP_6 and IP_4 activities by measurement of released phosphate and by HPLC of the degradation profiles. Several variants exhibited statistically significant differential activities and/or specificities to the wild-type.

The structure of MβLp01, a postulated novel bacterial phytase, was solved at 1.95A° (R_{Free} = 0.23), with the aim of investigating the structure-function relationship. Unexpectedly, whilst limited activity against ADP and ATP suggests this enzyme may be a phosphatase, the IP_6 data strongly indicate it is not a phytase. However, phosphodiesterase activity was revealed. Along with the indicated β-lactamase activity (carbapenemase and penicillinase), it is plausible that this enzyme is a promiscuous phosphodiesterase belonging to the metallo β-lactamase (MβL) superfamily. The findings in this chapter illustrate the profound importance of critically assessing published results for their scientific credibility.

This thesis also extends to an enzyme integral to IP_6 biosynthesis - ITPK1 from *Solanum tuberosum*. Attempts to elucidate the structure of *St*ITPK1 as a trapped substrate complex were unfruitful due to the pathologies of the crystals which limited the quality of the X-ray diffraction data.

List of Contents

List of Figures

List of Tables

List of Abbreviations

A_{Xnm}	Absorbance at X nm
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
AMP-PCP	Adenosine 5'-[β,γ-methylene] triphosphate
AMP-PNP	Adenosine 5'-[β,γ-imido] triphosphate
AppA	*E. coli* periplasmic phytase, AppA
AST	Antimicrobial susceptibility testing
ASU	Asymmetric unit
Bis-pNPP	Bis (4-nitrophenyl) phosphate
bp	Base pair
***Bt*MINPP**	MINPP from *Bacteroides thetaiotaomicron*
cAMP	Cyclic adenosine monophosphate
CC1/2	Half dataset correlation coefficient
CFE	Cell-free extract
CV	Column volumes
Da	Dalton
DLS	Diamond light source
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
FT	Flow-through
FTU	Phytase units
G1P	Glucose 1-phosphate
G6P	Glucose 6-phosphate
HAP	Histidine acid phosphatase
HAPhy	Histidine acid phytase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
HP-Trx	His-patch thioredoxin
ICP-MS	Inductively coupled plasma mass spectrometry
IMAC	Immobilised metal affinity chromatography
$IP_{1\text{-}6}$	*Myo*-inositol mono/di/tri/tetra/penta/hexakisphosphate
IPK	Inositol polyphosphate kinase
IPMK	Inositol phosphate multikinase
IPPK	IP_5-2K; Inositol pentakisphosphate 2-kinase
IP_3-3K	IP_3-3K; Inositol(1,4,5)P_3-3 kinase
IP6K	Inositol hexakisphosphate kinase
ITPK	inositol 1,3,4-triphosphate 5/6 kinase
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
$I/\sigma(I)$	Signal to noise ratio
kDa	Kilodalton
LB	Luria–Bertani medium
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
MβL	Metallo β-lactamase
MES	2-(N-morpholino)ethanesulfonic acid)
MIC	Minimum inhibitory concentration
MINPP	Multiple inositol polyphosphate phosphatase
MWCO	Molecular weight cut-off
MW	Molecular weight
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at 600 nm
o/n	Overnight
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
P_i	Inorganic phosphate
PNGM-1	Papua New Guinea metallo β-lactamase
pNPP	4-nitrophenyl phosphate
pNP	4-nitrophenol
PP-InsP	Inositol pyrophosphate
QB	Quantum Blue
RNA	Ribonucleic acid
R_{Free}	Residual factor (based on a subset of diffraction data)
R_{work}	Residual factor
R_{merge}	A measure of the internal agreement of diffraction data
RMSD	Root mean square deviation
RPM	Revolutions per minute
RT	Room temperature
SEC	Size exclusion chromatography
SDS-PAGE	Sodium dodecyl polyacrylamide gel electrophoresis
***St*ITPK**	ITPK from *Solanum tuberosum*
SOC	Super optimal medium with catabolic repressor
SDM	Site-directed mutagenesis
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TOPO	Topoisomerase based (cloning)
Tris	Tris(hydroxymethyl)aminomethane
tRNA	
Trx	Thioredoxin
T_a	Annealing temperature
T_R	Retention time
T	Time
WT	Wild-type
WCE	Whole cell extract

'When the day becomes the night and the sky becomes the sea, when the clock strikes heavy and there's no time for tea; and in our darkest hour, before my final rhyme, she will come back home to Wonderland and turn back the hands of time.'

Acknowledgements

First and foremost, I'd like to express my gratitude to my supervisor, Andrew Hemmings, for taking a chance on me, and for his patience and guidance during my challenging yet immensely fulfilling PhD journey. Thanks also extend to my co-supervisor Charles Brearley, and Hayley Whitfield, for their HPLC expertise and provision of materials. I'd like to express my appreciation to the members of Lab 2.30 for being an easy-going and diverse group of individuals, which rendered the lab a pleasant atmosphere and somewhere I enjoyed working; and particularly members of Lab 2.27 for the moral support and friendly chats. I am also grateful to the NRP DTP for their continued support and funding throughout the programme.

I'd like to thank my iCASE partner, AB Vista Ltd - particularly Mike Bedford - without whom this project would not be possible. I am especially grateful for the opportunity to spend time working with Dr Jose Francisco Perez Hernandez at the UAB in Barcelona, where I strengthened my Spanish skills and had a highly valuable and memorable experience. Thanks also go to Dr Gemma Gonzalez Ortiz for overseeing this placement.

Special thanks go to Emma Ainger-McKillop for being a continued source of pastoral and wellbeing support – for over a decade now! Thanks also to my mentor Sam and for the continued support from Sarah Yeates – especially during the later stages of my PhD.

I am grateful for the friendship with Simon Ellis both on and off campus and for his assistance with various errands.

A special thank you to my mum for always being there for me and believing in me no matter what.

Most importantly, I would not have been able to do this without the unwavering support from my partner, Andrew Cole. Words cannot describe the value of your contribution in terms of facilitating me to function through the emotional and domestic stability you provide. You have never doubted me, have always encouraged and reassured me and have witnessed me flourish over the past 4.5 years, both in my self-confidence and as a researcher.

I also need to mention Karl Pilkington – whose podcasts have kept me company throughout the past 4.5 months of thesis writing!

"Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."

CHAPTER ONE

Introduction

1.1 Phytate, Phytases and the Livestock Feed Industry: A Global Perspective

More from less – this is the ethos which is required to underpin the future global agricultural system in order to ameliorate the inevitable worldwide food deficit[1]. The world population is projected to reach 9 billion by 2040, yet even sustaining the current 7 billion people is already putting the global infrastructure under immense strain – the fact that 805 million of those people suffer from chronic hunger is testament to this. With rising incomes and changing food consumption patterns in tandem with the population growth, protein demand is set to increase to levels which will place the security of its supply at risk. The increasing popularisation of the Western diet by the emerging markets in the developed countries of Asia and Africa is a substantial factor affecting global food security. The consensus is that the world's existing food system will not be able to sustain 9 billion people on such an extremely resource intensive diet[2-4].

Much concern is focussed on the perceived competition for food resources - particularly grains - between the human and livestock populations[5,6]. In addition to direct allocation of land to livestock production, a considerable proportion of land - about half of global agricultural land – is used indirectly in the production of animal feedstuffs[2,7]. Many argue that starvation in the third world countries could be alleviated by the reattribution of land currently committed to feed grain production for livestock to that in which the product is suitably human edible[7]. Generally, the output from one acre of land is either 250 pounds of beef or 53,000 pounds of potatoes. Whilst the amount, in weight, of the latter is much greater, only 4 million acres are devoted to growing this type of produce, yet the land attributed to livestock is 14 times greater. However, there are a number of justifications supporting the traditional agricultural system of which the influencing factor is the variation in land quality constituting the world's surface area[5].

Of the fraction (less than half) of land which is arable, only a small portion is able to provide high quality food, such as wheat and rice, for human consumption; the remainder yielding that which is less than favourable with regard to digestibility, palatability and quality[5,6]. Corn, barley and oats are some such examples[6]. Indeed, only a quarter of the feed fed to livestock – and even

less of that fed to ruminants such as cattle - is potentially consumable by humans[5,7]. This represents a substantial portion of the total resources available for human food production which is inedible directly by humans. This indigestible fraction is comprised primarily of crop residues, by-products and wastes - from the harvesting of grain, vegetable, fruit and nut crops - which require processing through animals to produce high quality human food in the form of meat and milk[5,6]. By-products comprise that part of the crop which remains once the high quality material has been obtained for human consumption, often being imbalanced in major nutrients[5,6,8]. In addition to these arable land-derived resources, animals can also utilise roughages from non-arable land which adds to the amount of total human food resources[5,8]. Thus whilst the maintenance of a viable livestock industry does require more land per calorie or unit of protein than plant-based equivalents, much land is unsuitable for anything other than animals[9]. Since around 75% of the world's agricultural land produces forage that can be utilized only by ruminants, these domestic animals represent an integral link in the human food production chain[6,7].

However, although these low-quality feedstuffs are sufficient for their maintenance, they must be combined with inputs of other food sources in order to maximise animal productivity in terms of meat and milk, necessitating greater dietary energy concentrations and better quality feedstuffs hence the diversion of a quantity of human edible crop products to livestock feed[5,6]. Elements of agricultural economics influence the source of nutrients fed to livestock and therefore the proportion of grain which is redirected to livestock feed. Therefore, when the price of grain is less than other sources of appropriate nutrients, it will be used to augment low quality animal feedstuffs in order to realise maximum output efficiency[6]. Indeed, grain is often lower in terms of cost per calorie of feed energy than harvested roughages[7]. Since feed costs are the main variable cost in livestock production, feed and economic efficiency exist in synergy[5]. For example, total feed efficiency of the animal – another major concern - is largely determined by growth rate which is itself attributed to energy intake from their feed which in turn links back to the cost of energy-yielding feedstuffs[5]. Since the world's cattle consume the equivalent of the caloric requirements of the entire human population, this system of human food production has been criticized due to its apparent inefficiencies[6,10]. Despite the reduction in available energy and protein compared to the direct consumption of the plant feedstuff, as pointed out previously, much of the feed consumed by animals is either inedible initially, requiring processing through animals, or of low acceptability as human food[5,6].

Although society could feasibly be sustained on a completely plant-based diet, the existence of a livestock sector expands the resource base from which cereal grains produced in developed economies can be redeployed to human needs in times of food shortages[7,8]. With regard

to human dietary requirements, animal products are a source of great nutritional value due to the high availability of iron and other vitamins and minerals, and the appropriateness of the protein composition in meeting the requirements of human essential amino acids[5,11]. Indeed, there are certain amino acids and vitamins which can only be obtained from food of animal origin. It is clear that there is an imbalance in nutrition across the population, with malnutrition from lack of dietary sustenance in some under-developed countries and on the opposite end of the spectrum, health problems as a result of excessive intake of certain nutrients in the economically advanced parts of the world[12]. In these developed countries dietary intake is characterised by a greater contribution from animal products[5]. Certainly, the animal agriculture sector provides a substantial contribution to the economy and to worldwide total human food production[5].

As demonstrated here, there are a multitude of interrelated elements, the consideration of which represents a complex topic. Attention to, for example, geographical, environmental, economic and cultural factors amongst others, and complete comprehension of their relationships with one another, is necessary in any critical analysis or argument[5,11].

This is of course beyond the scope of this overview. The ultimate goal is the maximisation of the global total food resource, whereby output is largely dictated by the input of nutrients to the animals which in turn is influenced by a complex matrix of factors, as discussed. Because land is the overarching limiting factor, efforts must be concentrated on optimising the total available world resources – including animals – to generate *more from less*. This demands an appreciation of the science of animal nutrition.

1.2 Aspects of Non-ruminant Animal Nutrition

Analysis of animal efficiency in conversion of livestock feed to consumable animal products requires a review of animal dietary requirements, digestion and metabolism. An overview of animal nutrition must begin with an outline of the differences between ruminants, such as cattle and sheep, and non-ruminants such as swine and poultry. The fundamental difference is that concerning the digestion of cellulose[8]. Whilst the symbiotic relationship between the ruminant and the microflora which inhabit their digestive tract facilitates the animal's utilisation of cellulose as a source of energy, for monogastric animals lacking extensive microbial digestion, the digestion of the cellulosic constituents of plants is significantly limited by the nature of their digestive capabilities[8]. This ability of polygastric animals to exploit a greater proportion of human inedible feeds of low quality thus results in higher returns from human edible inputs than from monogastric animals[5]. To obtain maximum productivity from a particular feed resource, feed material more suitable for ruminants i.e. those with a high cellulose content can be preferentially given instead of wasting a valuable nutrient resource on those simple-stomached animals which cannot process it[8]. A pre-requisite for this is information on the composition of feed ingredients, this being facilitated by the 'Weende' feed analysis system developed in 1865[8]. This has now been replaced by more au courant classification methods.

1.2.1 Digestion in Monogastric Systems

The physiology of the cattle's digestive system is significantly different to that of monogastric livestock, comprising the rumen –or 'first stomach' - as the primary site of microbial fermentation[12]. Digestion in swine as compared to poultry, however, is very comparable, although the gizzard of poultry has no counterpart in the pig. In these single stomached animals, microbial activity occurs in the large intestine as a result of such gut microflora as lactobacilli, streptococci, coliforms, bacteroides, clostridia and yeasts[12]. Whilst not the main site of digestion, some digestion may occur in the large intestine attributed to enzyme activity resulting from the feed ingredients as well as from the microbes. The major site of digestion is in the small intestine[12].

Porcine

The pig's digestive system is analogous to that of a human comprising, in order of passage of feed: the mouth (where salivary glands secrete digestive enzymes to mix with the feed), pharynx, oesophagus, stomach, small intestine (which consists of the duodenum at the first part, the jejunum in the centre and at the terminal end, the ileum), large intestine, caecum, colon, rectum and anus[12,13]. Mature swine have a stomach capacity of about eight litres with a pH of about two resulting from the secreted acidic (~0.1M) gastric juice[12]. In addition to the hydrochloric acid, secretions include digestive enzymes, primarily pepsin which is activated by the acidic conditions. The caecum is the area of the intestinal tract responsible for the digestion of cellulose[13].

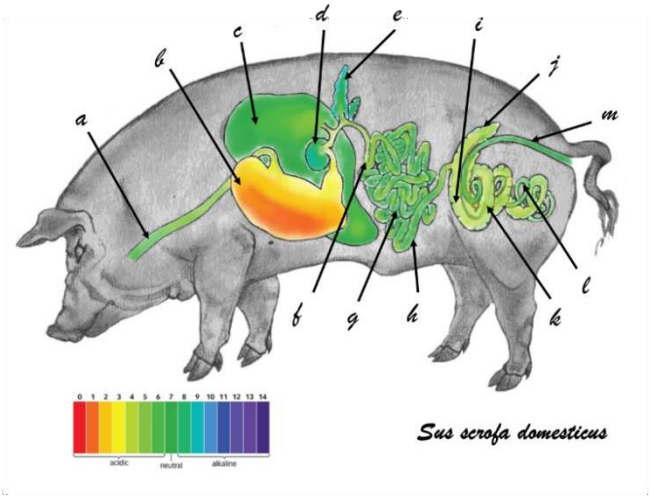


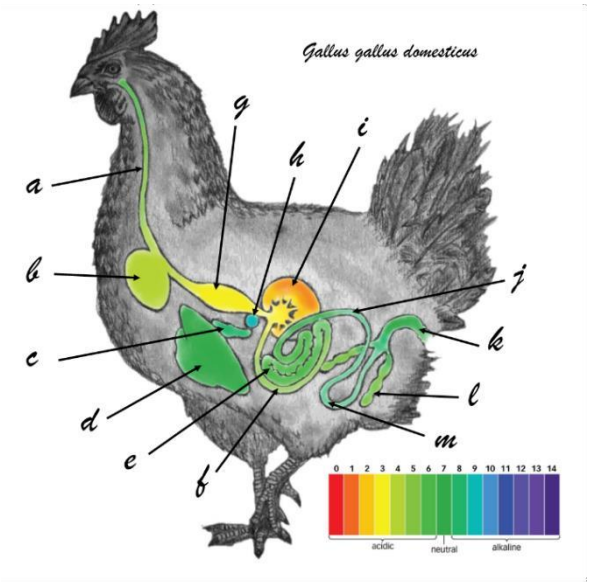
Figure 1.1 ***Sus scrofa domesticus.*** Figure illustrating porcine gastrointestinal tract with corresponding colour- scaled pH indicator. Stomach depicts the lowest pH neutralised on entry to duodenal loop via relatively basic conditions provided by bile secreted from the gall bladder. pH in liver and small intestine have stable neutral pH's for optimal enzymatic action. Image produced by Cole, A. (2020)

Avian

In poultry the stomach is divided into two chambers – the proventriculus and the ventriculus - or as the latter is most commonly known, the gizzard. At the base of the chicken's neck, leading from the oesophagus, there is an expandable storage reservoir - the crop – which enables the bird to accumulate a great quantity of feed in a short space of time for digestion within the next twelve hours. The proventriculus is situated between the crop and the gizzard and is the glandular region of the stomach, secreting the digestive enzyme pepsin and hydrochloric acid. The feed is ground and mixed with this enzyme mixture in the muscular compartment of the stomach, the gizzard, which removes the need for teeth. This process is the reverse of that in the ruminant where the exposure to enzymes follows the grinding of feed[12,14]. It is thought that the digesta only spends 60 – 90 minutes in the anterior region of the digestive tract (crop, proventriculus and gizzard), restricting the time for enzyme activity (Table 1.1)[15]. Nutrient absorption occurs in the small intestine which is composed of the duodenum, jejunum and ileum. The undigested residue subsequently passes via the pair of caeca, facilitating further

breakdown with the aid of bacteria, to the large intestine which sees absorption of water and dehydration of indigestible matter. The large intestine connects to the cloaca, the orifice through which the combined faecal/urinal waste is voided[14]. At three to four hours, the average retention time of feed in the digestive tract - between ingestion and the lower ileum - is relatively short[15].

It is important to consider that the present systems of poultry feeding are not conducive to the maintenance of the crop and gizzard functionality. The currently employed *ad libitum* arrangement is detrimental to the nature of the crop as the birds have constant access to food thus storage is no longer necessary. Similarly, pelleted, finely ground feed does not support gizzard development and leads to diminished grinding function and therefore reduced retention time. It has been demonstrated that the gizzard's proficiency is restored with the feeding of coarse and whole grains; and with the resultant enhanced grinding ability, larger gizzard volume and increased reversed peristalsis of digesta, retention time is lengthened[15].



As tabulated in Table 1.1, the distinct regions which constitute the digestive tract exist in different states of pH, from the mildly acidic crop, the acidic proventriculus and gizzard to the range of pH's exhibited by the small intestine – from mildly acidic at the proximal end to mildly alkaline and neutral at the distal end[15,16].

Figure 1.2 ***Gallus gallus domesticus.*** Image describing the complete digestive tract representational of avian species – specifically the domesticated chicken. Proventricular and ventricular regions are depicted with the lowest pH's to provide optimal range for pepsin activity. This pH is rapidly increased to around just below neutral at the duodenal loop and continues to increase to mildly basic conditions at the ileum terminus and colon. Alkalinity of bile stored in gall bladder serves to increase pH in duodenal regions to provide optimal pH environments for enzymatic activity. Image produced by Cole, A. (2020)

Segment	pH	Transit time min
Crop	5.5	10 – 50
Proventriculus/gizzard	2.5 -3.5	30 – 90
Duodenum	5-6	5 – 10
Jejunum	6.5 - 7.0	20 – 30
Ileum	7.0 – 7.5	50 – 70
Cecum/colon	8.0	20 - 30

Table 1.1 Corresponding pH and average digesta transitional times for particular gastrointestinal regions of domesticated chickens[15].

Piscine

The digestive systems of fish vary significantly across species, with the different feeding requirements of carnivorous, omnivorous, herbivorous, detritivorous and planktivorous fish resulting in differently adapted gut morphologies[17,18]. The digestive tract of carnivorous fish tends to be short with a large, muscular and elastic stomach for accommodating larger prey, whereas some non-predacious fish such as planktivores lack a distinct stomach but overall have a longer gut. In these fish, a constant stream of small particles flows directly from the oesophagus into the intestine, as there is little requirement for storage and/or gastric pre-digestion of food. Meanwhile, omnivorous fish have a larger digestive tract which has morphological elements of both carnivorous and herbivorous fish. There are some herbivorous species in which the stomach has adapted to act as the gizzard in avian species, crushing the vegetal material which constitutes the majority of their diet. Absent in stomach-less fish, pyloric caeca are present in various size and number in carnivorous species, with the purpose of increasing the overall surface area of the intestines and therefore increasing the absorption capacity of nutrients. Whilst predatory fish usually have short, straight intestines, herbivorous fish have longer, coiled intestines, compensating for the fewer caeca they possess. Unlike other higher vertebrates, there is less clear differentiation between the intestinal sections, with the terms proximal, middle and terminal being used instead, where proximal represents the duodenum equivalent[17-19].

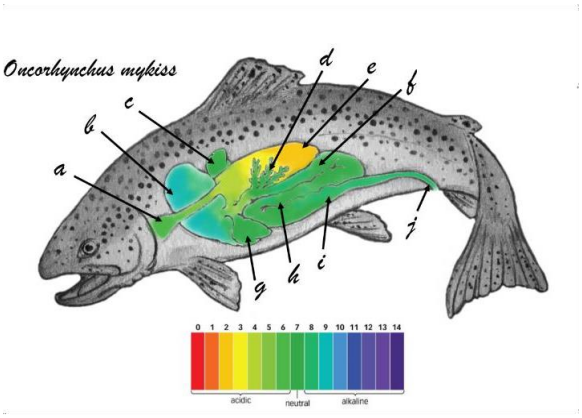


Figure 1.3 ***Oncorhynchus mykiss***. Image illustrating the gut morphology of a typical carnivorous fish. The digestive tract of rainbow trout is approximately the same as its body length[20]
Image produced by Cole, A. (2020)

Pictured is an example of a representative bony fish – *Oncorhynchus mykiss* (Rainbow trout), illustrating a stomach adapted for a carnivorous diet. Their good swimming ability is well

placed for capturing prey and the intestines are short since their food contains only minimal amounts of indigestible material (unlike herbivorous fish)[21]. Their digestive tract is also characterised by the presence of a true stomach. Like other fish and terrestrial monogastric animals, rainbow trout have a full set of digestive enzymes[20].

The nutrient composition of feed consumed by domestic animals is concomitant with their efficiency of conversion to animal products such as meat, milk and eggs[6,12]. As such, the value of a diet in meeting the needs of the animal is based on the concentration of nutrients in the diet and also the amount of feed ingested[22]. The digestibility of feed determines the quantity of nutrients which are actually absorbed and hence available for use in maintenance and/or production[12]. Nutrient intake is thus associated with digestibility in addition to food intake and utilisation such that the higher the digestibility of a given feed, the higher its intake by the animal and the greater its efficiency of utilisation of metabolisable energy[22]. The primary function of energy supplied by the nutrients in the feed is to meet the maintenance requirement of the animal i.e. its basal metabolism. Energy can only be channelled into growth and milk production when the energy supplied is in excess of that required for maintenance. As such the nutritional value of feed is based on its energy content[12]. The energy available for maintenance or production purposes is referred to as the 'net energy' and is the fraction of dietary energy available after certain losses have been accounted for. It is calculated from gross energy minus the energy lost in faeces (this being 'digestible energy') and also from urine (this being 'metabolisable energy'), with further subtraction equating to the value of heat loss. In this context, digestibility is more accurately referred to as 'apparent digestibility' since it is merely an assumption that anything not excreted has been absorbed and that the faeces are entirely composed of undigested food residues – it is not a measure of true digestibility[12,22]. In poultry it is easier to determine metabolisable energy than digestible energy due to the nature of their excretion in which faeces and urine exit from the same orifice, the cloaca[12,22]. The metabolisable energy of ruminants is further decreased due to the additional loss of energy in the form of methane. Metabolisable energy is utilised more efficiently by non- ruminant animals than the former[12]. The target of optimum feed formulation sought by manufacturers is thus in the consideration of an animal's nutritional need in the most efficient manner.

1.2.2 Phosphorus in Monogastric Animal Nutrition

1.2.2.1 Biological Functions and Requirements of Phosphorus

In addition to energy supplying nutrients, minerals, one of the six classes of essential nutrients, represent a quota of the nutritional dietary requirements of animals in maintenance and growth/production. There are at least forty minerals classed as essential, including trace elements. Phosphorus is one of those minerals required in relatively large amounts in the diet, as such classified as a macromineral. A multitude of diverse functions in the animal's body are attributed to this mineral element and as such phosphorus is a systemically abundant inorganic component, having the second greatest mineral concentration after calcium. Along with calcium, phosphorus is a fundamental constituent of bone, with 80% of the body's phosphorus accreted in bones and teeth. It also plays a key role in energy metabolism, and forms part of the structures of phosphoproteins, phospholipids and nucleic acids[12,16]. A synergistic relationship exists between phosphorus and calcium and as a result there is a certain ratio which must be attained, through the diet, in order to maintain physiological balance and bone growth and repair. Deviation from this ratio may be detrimental to the same extent as a deficiency of either element in the diet, since an excess of either calcium or phosphorus interferes with the absorption and activity of the other. For example, as phosphorus is complexed with calcium in bone, if bone resorption occurs in animals receiving calcium-inadequate diets to meet the calcium requirements of the animal, phosphorus is simultaneously liberated from the bone and excreted[12]. Egg laying hens, however, require a ratio containing a higher proportion of calcium due to the large amount of this element used in the synthesis of the eggshell. As such, poultry are excluded from the recommended range of the calcium: phosphorus ratio 1:1 to 2:1 for livestock animals[12]. Given the universal role of phosphorus in many physiological functions, it is unsurprising that its deficiency in livestock is considered as the most prevalent and economically significant of the range of mineral deficiencies. Phosphorus deficiency is linked to such conditions and effects as: rickets, osteomalacia, pica, stiff joints and muscular weakness, poor fertility, reduced milk yields, poor feed efficiency and subnormal growth. As such, livestock producers must adhere to feeding standards in order to prevent deficiencies of phosphorus and other nutrients in their herds and these standards have been updated over the decades as more information is accumulated from technological advancements[12]. An example of the standards for phosphorus allowances in the poultry sector are tabulated below (Table 1.2)[23].

Since the dissolved phosphorus concentrations of aquatic environments are low, fish are unable to meet phosphorus requirements through gill membrane and GI tract absorption alone, necessitating a relatively high dietary source of phosphorus to meet their growth requirements[24].

	Phosphorus %	**Phosphorus (available) %**
Broiler starter (0 - 5 weeks)	0.70	0.50
Broiler finisher (5 – 9 weeks)	0.60	0.45
Capon finisher (9 – 14 weeks)	0.60	0.40
Roaster finisher	-	0.40
Replacement starter	-	0.50
Chick grower (laying type)	-	0.45
Chick grower (broiler breeder)	-	0.45
Layer	-	0.50
Broiler breeder	-	0.50

Table 1.2 Recommended phosphorus allowances for rations for different categories of poultry; adapted from info in[23].

Phosphorus is a vital component of the major plant-derived sources used in livestock feeds, namely cereal grains, nuts, oilseeds and legumes. Within these plant seeds it is present as a constituent of phytate – the mineral bound salt form of the phosphoric acid derivative commonly referred to as phytic acid, the latter being the free acid form[12]. Phytate represents the primary storage form of phosphorus, with 60- 90% of total organic phosphorus content being attributed to this molecular species – myo-inositol-1,2,3,4,5,6-hexakisphosphate (Fig 1.4)[25-27]. Since it accumulates during seed development until maturation, the highest concentration of phytate is in the seeds - where it is present mostly as the insoluble calcium-magnesium phytate salt, phytin – and acts as a mineral reserve and the stored phosphates are used as an energy source and antioxidant for the germinating seed (Fig 1.4c)[12,27]. This explains the prevalence of phytic acid in plant-based foods[28]. In contrast, the leaves of plants consist of at least 70% of phosphorus in the inorganic form, with the remaining organic forms comprising phospholipids, ribonucleic acids and phosphate esters[26]. Phytic acid/phytate is also the primary storage form of the organic component of this molecule: inositol[28].

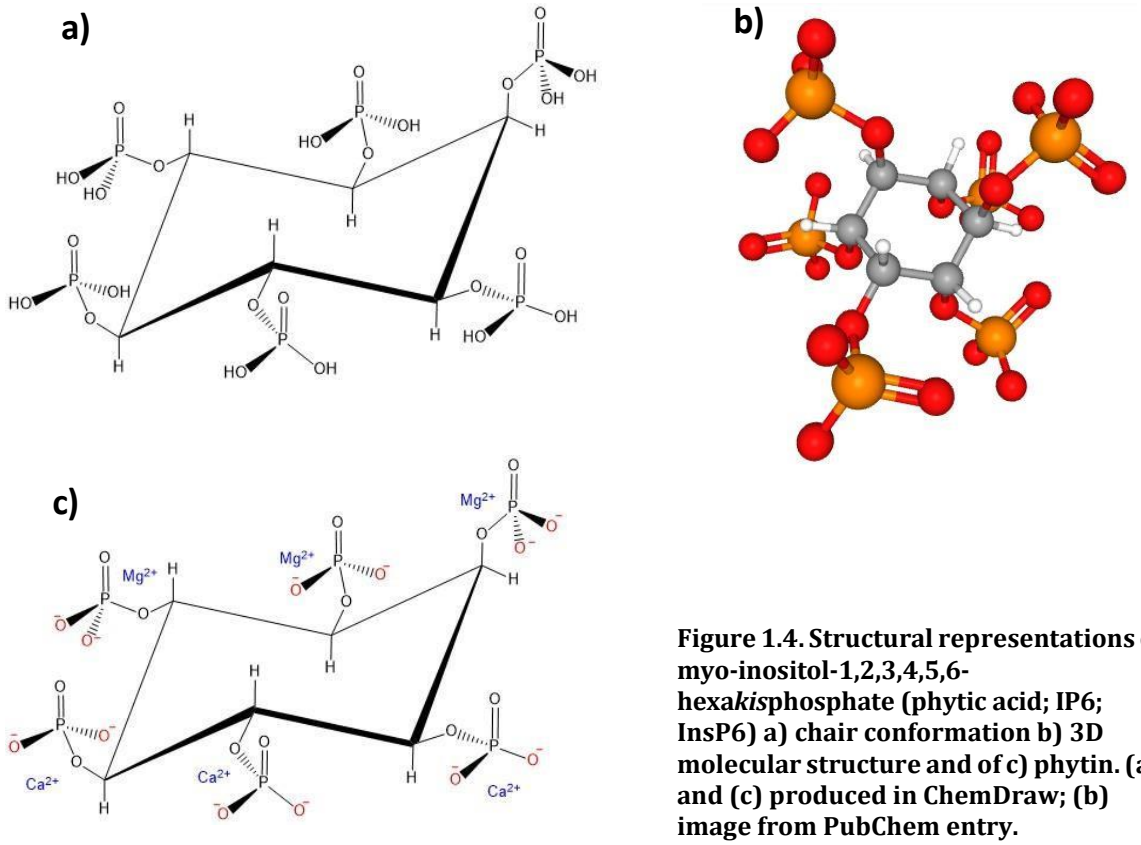


Figure 1.4. Structural representations of myo-inositol-1,2,3,4,5,6-hexa*kis*phosphate (phytic acid; IP6; InsP6) a) chair conformation b) 3D molecular structure and of c) phytin. (a) and (c) produced in ChemDraw; (b) image from PubChem entry.

1.2.2.2 Structure and Chemistry of Inositol Phosphates

Myo-inositol 1,2,3,4,5,6 hexa*kis*phosphoric acid (phytic acid; also abbreviated to 'IP6' or 'InsP6' and referred to as 'IP6' hereafter) is one of a family of several phosphorylated inositol derivatives present in the natural environment which consist of a hexahydroxycyclohexane scaffold – the inositol ring - esterified to up to six phosphate groups and abbreviated as IPx (where x is equivalent to the number of phosphate moieties)[29,30]. In addition to the degree of phosphorylation, the inositol phosphates vary also in isomeric form based on the conformations of the hydroxyl groups of the unsubstituted inositols, although myo-inositol is certainly the most prevalent of the nine* naturally occurring forms (Fig 1.5)[29,30]. For example, in terrestrial environments myo-inositol is the major nutritionally relevant form of inositol in plant material, with the *myo*-IP6 often constituting 100% of the total inositol phosphate content of plant seeds[27,29]. Indeed, inositol phosphate stereoisomers other than the '*myo'* form are found almost exclusively in soil, with the *myo* form of IP6 for example representing up to 90% of total IP6, the scyllo form 20-50%, the D-chiro form ~10% and the neo form ~1%[29].

*Unlike the rest of the literature, in reference [31] it is stated that there are actually 13 stereoisomers: 'The maximum number of stereoisomers for a molecule with more than one stereogenic centre is given by the 2^n rule (where n= number of stereogenic centres), which in the case of inositol should be 64 (2^6). The symmetry inherent in inositol reduces the possible distinct stereoisomers to 16. Of these, 6 are made of 3 enantiomeric pairs while the rest are *meso* compounds leading to only 13 theoretically possible isomers of inositol as enantiomeric pairs have identical energies'[31].

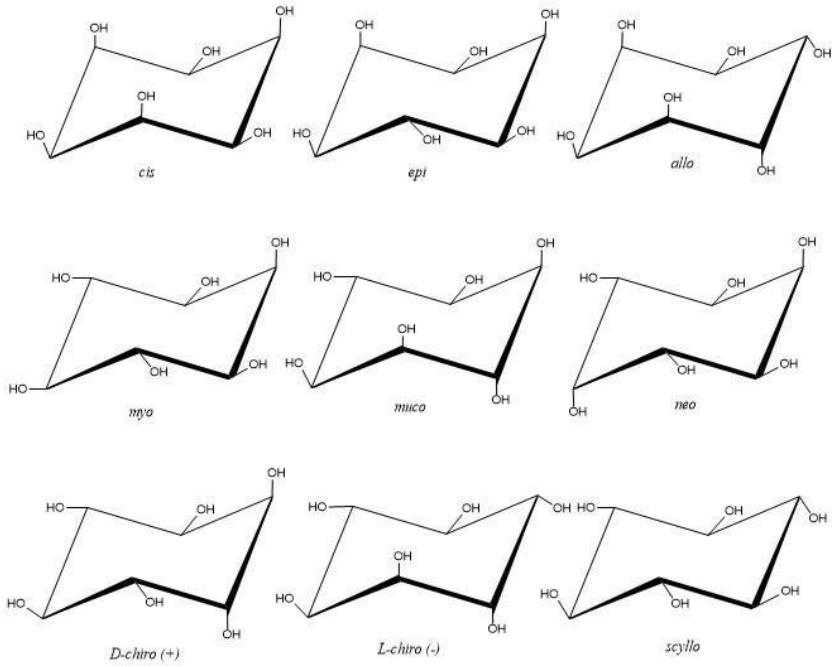


Figure 1.5. The nine stereoisomers of inositol displayed in chair conformation: cis-, epi-, allo-, *myo*-, muco-, neo-, D-chiro(+)-, L-chiro(-)- and scyllo- inositols. Produced in ChemDraw. Information from [29, 31, 32]

Stereochemical considerations of myo-IP6: Phosphorus is stereochemically significant since it renders the achiral *myo*-inositol scaffold chiral upon substitution of a hydroxyl moiety with a phosphate group at a stereogenic position (C1, C3, C4, C6)[30]. The *myo* form of IP6 represents just one of the nine different stereoisomeric forms that the six phosphoric ester groups can exist in, where the proportion of axial groups are between none and three[30,32]. This correlates to 66 possible isomers of phosphorylated *myo*-inositol[30]. However, Agranoff himself stated that there are 63 possible phosphomonoesters of *myo*-inositol: 6 each of IP1 and IP5, 15 each of IP2 and IP4, 20 of IP3, and IP6[33]. Nevertheless, considering just *myo*-IP6, there are several different conformations which the molecule can exist in, transiently or otherwise. Nomenclature is inherently linked to stereochemistry and this must be unambiguous in any analysis or communication involving stereochemistry. Thanks to B.W. Agranoff in his aptly titled paper 'Cyclitol confusion', the conformation of *myo*-inositol - and therefore *myo*-IP6 - may be viewed, provided one disregards the hydrogens, as analogous to the shape of a turtle, with the turtle's head representing the axial C2 group and the co-planar limbs and tail representing the remaining equatorial groups (Fig 1.6). In terms of numbering the carbon positions, C1 is in the position of the turtle's right limb and numbering is in the counterclockwise (D) direction[34]. The nomenclature and thus stereochemistry of inositol derivatives is a potential danger zone owing to an official nomenclature reversal of 'D' and 'L'

designations of the inositol ring back in the 1970's. This is a source of much uncertainty when reading papers from a range of time periods since those written before this time will be using the outdated, 'inverted' nomenclature[35]. As a result, it also means that the nomenclature of phosphoester-hydrolysing enzymes is different before and after the change (e.g. 4-phytase vs 6-phytase, section 1.3.1.2). The conformation of *myo*-IP6 is pH-dependent such that between pH 0.5 – 9.0 the sterically unhindered and stable 1-axial/5-equatorial conformation dominates and between pH 10–13 the conformational inversion to the sterically hindered 5- equatorial/1-axial form supersedes (Fig 1.7)[36,37]. The extent of phosphorylation of IP6 underpins its strong complexation with soil, the stability of which is preferential over other organic phosphorus compounds. This renders it relatively insusceptible to hydrolytic enzymes such that these inositol compounds accumulate, representing the major fraction of organic phosphorus in most soils[29].

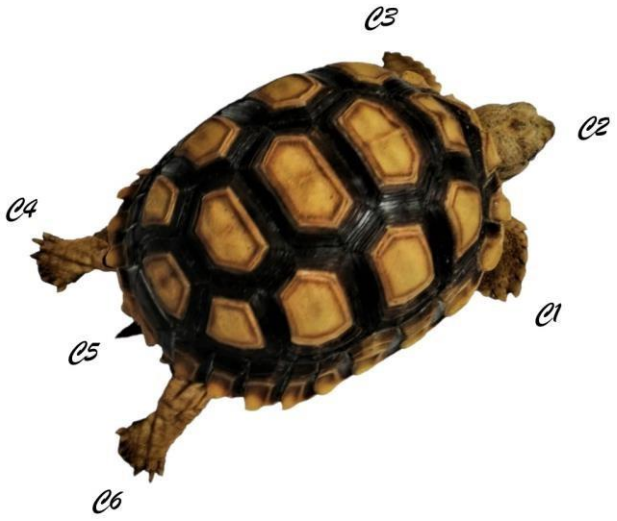


Figure 1.6. 'Agranoff's turtle'- visual mnemonic for inositol phosphate stereochemistry, with carbon numbering. Model: Benzo the tortoise. Image edited by Cole, A 2020. Information from Agranoff's paper[34].

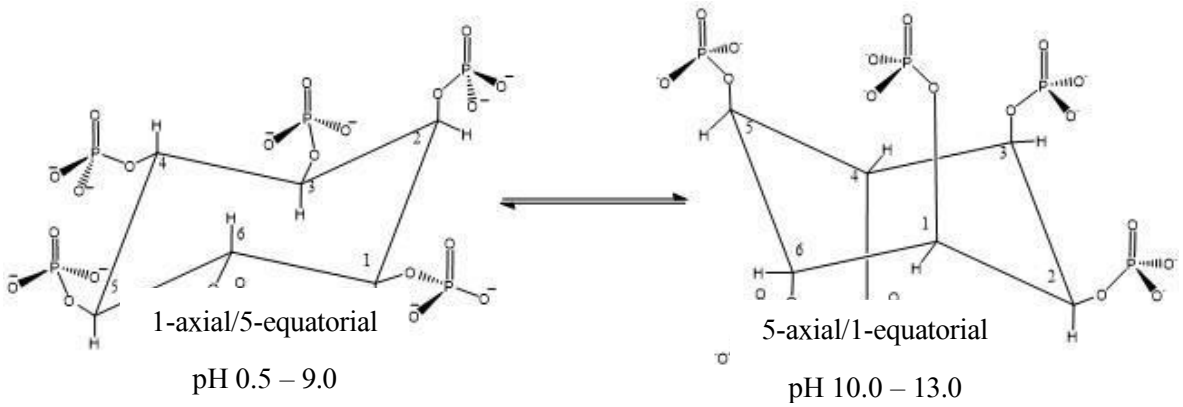


Figure 1.7. The equilibrium between the two conformations of *myo*-IP6 within different pH ranges. Below pH 9.0, all phosphates bar one are in the equatorial conformation (1-axial/5-equatorial), whereas above pH 10.0 all bar one are in the axial conformation. At pH 9.0 - 9.5 both conformations are in dynamic equilibrium. Image produced in ChemDraw. Information from [36,37]

1.2.2.3 Biological Functions of Inositol Phosphates

The 'discovery' of IP6 in 1903 by Posternak – following the first successful preparation of phytin - established this family of inositol phosphoric esters that was populated over the ensuing decades following identifications of various inositol polyphosphates[38-40]. Of particular significance is the triphosphorylated inositol-1,4,5-triphosphate, identified in 1983 as a Ca^{2+}-mobilising second messenger, which initiated the identification of the key role of other lower phosphorylated inositols in cellular signalling pathways[35]. The inositol mono- to triphosphates are constituents of the phosphoinositides which represent ~2-9% of the total phospholipid content of plant and animal tissues[29]. As hydrolytic products of phospholipids, these lower myo-inositol phosphates are also involved in a myriad of diverse cellular activities such as: stress responses, membrane biogenesis, protein folding and trafficking, endo- and exocytosis, oocyte maturation, gene regulation, cell division and differentiation, efficient export of mRNA, RNA editing and DNA repair[27,30]. Despite the significance of these lower *myo*- inositol phosphates in eukaryotic metabolic processes, their existence in the biosphere is scarce (or undetectable), with their presence merely as transient intermediates of biochemical reactions. Nevertheless, the inositol phosphate family as a whole are in fact prevalent in terrestrial and aquatic environments, although this is attributed largely to the *myo*-IP6 congener, which constitutes approximately 83% of the inositol phosphate species in soil[29]. The proportions of the mono- through hexa- forms in soil correlate with the extent of phosphorylation, with IP5 representing ~12%; IP4 ~4%; IP3 ~1% and IP2 only trace amounts.

This correlation with number of phosphate moieties is due to preferential stabilization in the soil of the most phosphorylated inositol esters as a result of the highly adsorptive nature of the phosphate species and subsequently results in the culmination of IP6 and, to a lesser extent, IP5. Their strong interaction with soil also prevents their hydrolysis by hydrolytic enzymes, which contributes to the accumulation of these organic phosphates in terrestrial environments[29].

1.2.2.4 Mineral Bioavailability

In the free acid form, the six phosphate functionalities of phytic acid equate to twelve ionisable protons which correspond to a high anionic charge of the phytate molecule[27,29]. As such, the phosphate groups of phytate and other inositol phosphates act as chelating agents and have a propensity to complex strongly with the polyvalent cations of calcium, iron, magnesium, manganese, potassium, zinc and copper. The solubility – a prerequisite for absorption - of the resultant salt complexes is influenced by pH, with lower levels exhibiting increased solubility. The cation: phytate ratio also affects solubility, such that a very low or very high metal ion: phytate ratio increases the solubility of the salt. Thus for the complexes to be soluble the phytic acid must be in excess, enabling 1:1 stoichiometries, as when metal ions are in excess, insoluble metal ion: phytate complexes of 6:1 dominate. The type of individual cation is a further factor affecting solubility, since salts of Ca^{2+}, Cd^{2+}, Zn^{2+} and Cu^{2+} tend to be soluble below pH 4-5 and the Mg^{2+} salt up to pH 7.5, whereas for $Fe^{2+/3+}$ solubility is increased above pH 4[27,41]. In general, the divalent cations of zinc, calcium and magnesium usually form insoluble penta- and hexa-substituted salts[28]. The insensitivity of monovalent cations such as sodium to the effects of phytic acid are attributed to the weaker bonds between the two species than that which form with divalent cations, rendering solubility unaffected[42]. The order of affinity for the complex formation at particular pH ranges is not categorically verified, with varying reports and differences between studies in plants and in the human GI tract[29,41]. At human gastrointestinal pH, these mineral-phytate complexes are thus mostly insoluble and consequently not absorbed, rendering essential minerals unavailable as nutritional factors. Studies in humans have demonstrated in particular the inhibition of iron, zinc, calcium, magnesium and manganese absorption by phytic acid and since these are dietary minerals, phytic acid is classified as an antinutrient[27,41]. Zinc reportedly forms the most insoluble of the mineral complexes at physiological pH, to the detriment of bioavailability. As such, individuals with a high dietary intake of phytic acid are at risk of developing mineral-related deficiencies, particularly of zinc[41]. Since phytate-rich foods constitute the staple source of nutrition in developing countries, widespread human nutritional deficiencies of zinc, calcium and iron are inevitable in these regions[27]. Additionally, mineral sub-deficiencies are common in the Westernised world,

especially with the recent culture of whole-grain diets and the growing popularity of veganism, which see an increase in the proportion of phytate in the diet[43]. Since the less phosphorylated inositol phosphates have a lower binding capacity for iron, zinc and calcium they have a lesser contribution to the decreased bioavailability of these minerals[43]. It has been demonstrated that IP3 and IP4 have no effect on absorption of iron, zinc and calcium unless in the presence of higher phosphorylated inositol phosphates, where they were found to synergistically contribute to the negative impact of IP6 on absorption. For example, the absorption of zinc is negatively correlated with the sum of IP3 through IP6. The lower inositol phosphate-mineral complexes also exhibit lower solubility and stability than IP6, with these factors decreasing in proportion to the level of phosphorylation thus contributing to the reduced impairment of mineral uptake by the lower substituted inositol phosphates[28]. In poultry diets rich in phytic acid, the bioavailability of minerals such as zinc and manganese have been shown to be increased upon addition of other chelating agents such as EDTA – presumably these complexes have a higher binding affinity than with phytic acid and are soluble and thus can be absorbed[12]. Phytic acid also binds to and impedes the functions of some dietary amino acids, proteins, carbohydrates and lipids thereby affecting enzyme activity and protein solubility and digestibility. As a result of the insolubility of protein-phytate complexes above pH 3.5, dietary proteins undergo less processing by proteolytic and digestive enzymes – including pepsin, trypsin and amylase -thus preventing their absorption and attesting to the antinutrient characteristics of phytate[41]. Human studies have however also demonstrated an antioxidant role for phytate, as it chelates and subsequently inhibits non-haem iron absorption, which would otherwise give rise to undesirable free radical formation through Fe^{3+} catalysis[35,41]. In fact, the prevalence of phytic acid in eukaryotic cellular systems can be mostly attributed to its properties as a chelating agent, which in turn are ascribed to its high negative charge over a wide pH range[27]. In general, the net consequence of dietary phytic acid is that calcium and various other nutritional biomolecules are lost by sequestration and phosphorus is not absorbed.

1.2.2.5 Phytate Content of Plant Feedstuffs

As stated in section 1.2.2.1, phytate is abundant in the staple plant-based sources (cereals, legumes, oilseeds and nuts), comprising 1-5% of their weight, which constitute the foundations of the human diet worldwide - both directly, and indirectly through animal-sourced products[44]. As a direct source of human food, these feedstuffs contribute around 40% of total calorie intake in developed countries and 60% in lesser developed countries[43].

The average phytate intake in the United Kingdom and America ranges between 631 mg and 746 mg per day although on a vegetarian diet, subject to the quantity and processing of plant sources in the diet, intake of phytate may be up to 4500 mg daily and on average is 2000-2600 mg; whilst for rural inhabitants of developing countries on mixed diets, daily intake is on average 150- 1400mg[28.41,45]. Much of the phytate in cereals and other plants is in the form of insoluble magnesium and calcium phytates, in particular the phytin form, which is the Ca^{2+} - Mg^{2+} salt of phytic acid. Given that the percentage of phosphorus in every phytate molecule is 28.2%, of the 1 - 5% of phytate comprising the major plant feed sources, around 0.28 - 1.4% is thus attributed to phosphorus, yet much less is bioaccessible by the animal[46]. Like humans, simple-stomached species such as swine, poultry and fish are unable to efficiently utilise the phosphorus in their plant-based feed which is present predominantly as phytate, since any hydrolase activity in their upper digestive tract is insufficient to liberate the phosphate molecules from phytate. As such, although the phosphorus content of grain-based feeds is generally adequate to support the needs of the animal**,** the presence of phytate in swine and poultry feed is undesirable and an inorganic phosphate supplement is used to augment the feed in order to satisfy mineral requirements[16,43]. The bioavailability of phosphorus is generally about half as great in swine than poultry with less than 15% bioavailable phosphorus in swine and around 30% (but ranging 15% - 40%) in poultry[46,47]. Fish tend to retain around 40% of phosphorus in modern commercial fish feeds[48]. However, it must be noted that phosphorus availability is a relative value compared to a reference so is not an absolute quantitative representation and further, is highly variable as a result of many influencing factors, including endogenous plant phosphohydrolytic activity, the proportion of phytate phosphorus in the feed matter and the technological procedures the feeds are subjected to. For example, it was originally established that the availability of plant phosphorus is 30%, however this is now considered an almost arbitrary figure[46]. Due to this broad range of bioavailability values which depend on both the animal and the feed ingredient, the dietary phosphorus requirement of egg-laying hens is difficult to judge, so in this context the requirements are based solely on supplemental inorganic phosphorus[47]. The author of this thesis considers that this allows the potential for significant overestimations which not only compound the excreta pollution concern but is also a very unsustainable use of natural rock phosphate. In animal feeds where there are specific nutritional requirements to be met, percentages of required phosphorus are given in terms of available phosphorus, so the amount of the mineral in the feedstuffs must be expressed similarly, rather than in terms of total phosphorus content which is redundant. There is high variability in the amount of phytate in cereal grains, oilseeds, legumes and nuts – on average 60 – 82% of total phosphorus - owing to variation in plant variety, climate, growing conditions, harvesting techniques, processing methods, storage conditions, testing methods and the maturity of the seed. For example, foods grown under high phosphate fertiliser will

have inherently accumulated more phytic acid[49]. Phytate phosphorus content within a given feedstuff is a further factor of inconsistency. In general seeds and the bran portion of cereal grains boast the largest phytic acid manifestation[26,27]. Feed formulations of livestock feeding programmes in established economies typically consist of a cereal grain basis supplemented with protein, vitamins and minerals. Soybean meal is the standard protein supplement in livestock diets worldwide – particularly in poultry and swine diets - since it boasts the highest nutritional value of all the plant protein sources. It is a major source of highly digestible protein with desirable amino acid composition which is superior to the other oilseeds with regard to the limiting amino acids[46,50]. It is perhaps appropriate at this point to highlight that up until 1996, livestock obtained dietary protein from animal sources such as meat- and bone-meal (MBM) but this was outlawed in Britain in response to the *Bovine spongiform* encephalopathy (BSE) crisis which arose from cattle consuming animal protein from infected livestock*[51]. Since global production of soybeans is greater than all other major oilseeds combined, soybean meal is readily available. In fact, three million tonnes of soybean were imported to the UK in 2023 for this purpose[50]. Soybean and other oilseed meals are residual by-products from the production of vegetable oil which are indispensable in feeding systems and a key factor in sustaining modern levels of meat production and in the maintenance of a viable livestock industry[8]. In combination with maize, soybean meal forms the primary energy source in swine and poultry diets, and although phosphorus levels of the soybean meal are high, more than half is unavailable for uptake by these monogastric animals, necessitating the inclusion of a mineral supplement[46]. Bioavailability for these feedstuffs is reported to range between 10-30% in poultry. The composition of a typical swine diet, and the total phosphorus content of components, is tabulated below (Table 1.3)[8] Since total phosphorus levels are relatively insignificant without knowing the fraction of phosphorus which is derived from phytate, the author has used phytate phosphorus values from other sources to augment this information. Approximately 70% of the P in a corn-SBM diet is unavailable to pigs[60].

Feed ingredient	Level in diet (g/Kg)	Phosphorus (g/Kg)	Phytate phosphorus (average)
Maize	820	0.82	0.92 g/100g[53]; (raw) 9.58 g/kg[54]; (fresh mature) 1.71 g/kg[55]; (dry) 7.15-7.60 g/kg[55]; 75-80% of total P[56]
Soybean meal (protein supplement)	150	0.35	1.43 g/100g[53]; ~67% of total P[57]; 0.34%[58]; 71% of total P[59]
Mineral supplements (Calcium and Phosphorus)	30	2.83	N/A

Table 1.3 Phosphorus contribution of maize, soybean meal and mineral supplement to a typical swine feed.

*In the UK, the original feed ban was introduced in 1988 to prevent ruminant protein being fed to ruminants. In addition, it has been illegal to feed ruminants with all forms of mammalian protein (with specific exceptions) since November 1994 and to feed any farmed livestock, including fish and horses, with mammalian meat and bone meal (mammalian MBM) since 04 April 1996. The ban was expanded in January 2001 to include the use of all processed animal protein (PAP) in the feed of animals farmed for food [52].

The percentage of phytate phosphorus of the main cereals, legumes, oilseeds and nuts commonly used in livestock feed are presented in Table 1.4. In general, the oilseeds and resultant meals have relatively high phytic acid levels, with levels of phytate phosphorus in excess of 80% of total phosphorus in groundnut (peanut) and sesame seed, for example (although the author of this report notes that groundnut is technically a legume)[26]. Along with soy protein concentrate – maximum phytate content of 10.7% - groundnuts and sesame seeds are well known to have exceptionally high levels of phytate, sometimes amounting to up to 5% of the dry weight in the case of the former[26,43]. The various oilseed meals on average contain phytate at 60-77% of total phosphorus and the seeds 1.0 - 5.4% by dry weight[26,43]. Although high in phosphorus, cottonseed meal is more suitable for ruminants. Linseed meal, a by-product of flax oilseed, is rarely administered to poultry. Canola – Canadian low acid rapeseed is an exemplary transgenic plant-derived feed, being a cross between two different species of rapeseed, and has widespread use in all stages of production for swine and poultry in addition to its common usage in fish feed[12]. As the predominant energy source, cereal grains supply insufficient nutrients required by the animal and as such they are augmented by the protein sources in the diet. Their phosphorus bioavailability is low (5-15%), in which the phytate content (64-85% of total phosphorus) exacerbates the already low mineral content[43]. With regard to the cereal grains commonly included in animal diets, and their phytate levels, refer to Table 1.4. Maize is well-established as the primary cereal component of livestock feed, particularly for swine and poultry, being superior to the other grains in terms of energy provision. Whilst wheat is comparable in nutritive value to maize, it is less economically favourable and is preferred for human food use. The high fibre content of oats and barley restricts their application and least commonly employed is rye, all of which must not exceed certain fractions of the total diet[12]. Ranging between 0.6 - 1.03% (dry weight) phytate, grain legumes are usually lower in amount of phytate phosphorus than cereals and oilseeds, this comprising approximately 60-75% of total phosphorus[26].

	Total P (%)	Phytate P (%)	Phytate P as a % of total P
Cereals			
Maize	0.23	0.18	78
Oats	0.29	0.17	59
Rye	0.34	0.20	59
Wheat	0.29	0.23	79
Barley	0.31	0.19	61
Legume Seeds			
Peas	0.43	0.24	56
Lupines	0.33	0.16	48
Beans	0.39	0.08	21
Soybeans	0.73	0.33	45
Oilseeds			
Linseed	0.60	0.34	57
Rapeseed	1.05	0.76	72
Cereal by-products			
Wheat bran	1.16	0.88	76
Rye bran	0.96	0.73	76
Oat bran	0.83	0.68	82

Table 1.4 Phytate P relative to total P in common feed ingredients for monogastric livestock. From [61].

1.2.2.6 Environmental Considerations

Since monogastric animals are unable to utilise phytate phosphorus through lack of endogenous enzyme activity which serves to release the phosphate molecules from phytate, feed for swine and poultry are traditionally supplemented with a source of inorganic phosphate (e.g. calcium phosphate) to ensure their nutrient requirements for optimal growth are adequately met. However, the rock phosphate employed in production of the supplements is an expensive, finite natural resource[12,16]. In fact, phosphorus is the third most expensive component of poultry diets in general[46]. Meanwhile, the ingested, unresorbed phytate phosphorus is excreted, accumulating in the soil through manure application – and adding to the phosphorus from mineral fertilisers - to levels in excess of that utilised by plants[47]. Some decades ago, Barrow and Lambourne (1962) reported that only ~0.06g of organic phosphorus is excreted per 100g of feed ingested, whilst the remainder is excreted as the inorganic form; and that the higher the phosphorus content of the feed, the greater the quantity of inorganic phosphorus in excreta[62]. More recently, the percentage of total phosphorus intake lost in excreta has been reported to be around 70%[29,43]. Once in soil, the phosphorus compounds then infiltrate into water bodies which jeopardises the health of the aquatic ecosystem through eutrophication[46,47]. Indeed, phosphorus, which has a high affinity for soil and sediment particles, is the primary cause of eutrophication in freshwater, contributing to the induction of rapid accumulation of aquatic plant life and subsequent

oxygen depletion[29]. The culmination of this is the establishment of algal blooms (including toxin-producing cyanobacteria), hypoxia and aquatic organism death, whilst the resulting nitrous oxide released contributes to greenhouse gas emissions[43]. This has significant implications for drinking water quality and purification costs[46]. Some of the phosphorus is in the form of inositol phosphates (other than IP6) which are abundant not only in the terrestrial environment but in aquatic systems too, and are a contributing factor in eutrophication, since they provide a potential phosphorus source for algal growth[29]. Anthropogenic factors have exacerbated the imbalance in the natural phosphorus cycle, namely by mining of phosphate rock for use in the production of fertilisers and – although not in the past fifty years – detergents (reader is referred to the phosphate rock market); and through the intensification of the livestock industry in response to the increased global meat demand – all of which generate excessive inputs of phosphorus into the environment[47]. Minimising these environmental penalties and reaping the concomitant economic benefits is thus one of the main rationales for the reduction of phytic acid.

1.2.2.7 Phytic Acid Abatement

As a result of the aforementioned detrimental impacts of phytic acid on human and animal nutritional health, ecological systems and the livestock economy, the reduction of phytic acid is thus a reasonable objective with tangible benefits and a clear rationale. Such anticipated gains include the improvement of water quality and state of aquatic ecosystems, compliance with environmental regulations with regard to phosphorus pollution, economic advantage (since the inorganic phosphate supplements would subsequently be redundant), circumvention of micronutrient malnutrition and the improvement of livestock production. For example, research has predicted that in the absence of phytic acid the absorption of zinc and magnesium in humans would be approximately 20% and 60% higher, respectively[63]. It is considered by some nutritionists that for 'best health' the phytic acid content should be no more than 0.03% of the phytate-containing food. Attempts to eliminate or lower phytic acid levels in food have been an enduring concept, with the time-consuming preparation of grains and legumes commonplace in traditional non-industrialised societies.

There are now several potential approaches to dietary manipulation, and these may be regarded as enzymatic (exogenous and/or endogenous); genetic; and livestock management strategies. In terms of the latter, such options involve the accurate estimation of dietary phosphorus requirements to minimise unnecessary excesses – this requires knowledge of the phosphorus bioavailability of feed sources in each species; phase feeding (which is essentially the first point but with estimates re-evaluated at the various stages of animal development); and diet optimisation – for example, the inclusion of supplements (non-enzymatic) such as

vitamin D to augment phosphorus retention[47].

Genetic manipulation approaches feature the generation of low phytic acid mutant seeds and biofarming, the latter concerning transgenic plants and transgenic livestock[27]. Although not a strategy to reduce phytic acid itself, biofortification of staple crops is employed to counteract its adverse effects on nutrition, mitigating the micronutrient malnutrition compounded by high-phytate intake[43]. Low phytic acid mutant seeds are produced through suppression of phytic acid biosynthesis, based on knock-out gene biotechnology[27]. In recent decades, there have been reports of improved nutritional properties with low phytate mutants of maize, barley, rice and soybeans[64]. The alternative to inhibiting phytic acid biosynthesis is to enhance its degradation. In the context of genetic modification, this entails inserting a gene encoding a phytate-degrading enzyme into the plant genome, the overexpression of which during seed development reduces phytate in mature seeds. Recent work in this area has given rise to transgenic soybean producing one of these enzymes[65]. <TRANSGENIC APPA SEAWEED> Biofarming of phytate-degrading enzymes is considered to be a cost-effective approach to their production. A further genetic-based approach to the reduction of phytic acid by monogastric animals is to generate transgenic livestock. A recent study demonstrated the possibility of introducing phytate-hydrolysing enzymes into the saliva of transgenic mice[66].

A less biotechnologically advanced yet fundamentally established strategy, particularly in more traditional populations of lesser developed countries, is the processing of feed and food grains to reduce phytic acid content. These well-documented techniques are largely the product of a substantial amount of time, heat and water and include soaking, cooking, malting (sometimes referred to as sprouting or germination), milling and fermentation[28,41,43]. Although milling eliminates the majority of phytic acid due to the association of phytic acid with the bran layer, it also removes beneficial nutrients too which is undesirable. All of these processing methods, aside from the physical process of milling, activate intrinsic enzymes in the grain which are able to hydrolyse the phosphomonoester bonds of phytic acid such that IP6 levels are reduced, and inorganic phosphate is rendered bioavailable. The natural process of germination serves to convert phytate into a source of inorganic phosphorus and mineral cations for the emerging seedling through the activity of these phosphohydrolases, and this is what these processing methods endeavour to emulate. It is well-documented that the soaking of food grains in warm acidic water for a considerable amount of time prior to cooking is sufficient to eliminate a significant proportion of the phytate. Cooking alone is redundant since the high temperature deactivates the desired enzymes, whereas the temperature and pH of the water used for soaking can be adjusted to that for optimal enzyme activity. However, if the grains and legumes have naturally low phytate-hydrolysing enzyme activity, soaking is an inadequate technique to lower phytic acid content[28]. There are numerous reports of significant reductions in phytate levels by combinations of these processing techniques however the author of this report

surmises that the substantial heat and time required to achieve these reductions – let alone the consumption of energy - in addition to possible limits on equipment and automation/labour render these not economically and commercially viable in industry as the sole solution. Indeed, enzymatic hydrolysis of phytate is the gold-standard approach of all of these discussed since it is able to most effectively break down phytic acid without jeopardising other valuable dietary nutrients[43]. As such**,** these phytate-degrading enzymes – phytases - are the focus of the following section, and attention shall be given there to the enzymes intrinsic within plant-derived seeds and grains. As this report is concerned with enzymatic reduction of phytic acid, these alternative methods and reports based on them, regardless of viability, shall not be discussed henceforth.

1.3 Enzymatic Degradation of Phytate

1.3.1 Phytases: Overview, Classification & Prevalence

'Any meaningful discussion of the kinetics and mechanism of enzymic catalysis must be based on a sound knowledge of the type of reaction concerned in the absence of catalysis' (Kirby and Varvoglis, 1967)[67]. The phosphoric ester bonds constituting the phytate species are quite hydrolytically unreactive, a property which renders the phosphate ester bond-containing nucleic compounds exclusively suitable as components of genetic material. Since it is essential that genetic material be stable for the length of an organism's life-span, the requirement is that these macromolecules be relatively resistant to hydrolysis. The hydrolytic stability of the phosphoester bond is attributed to the ability of phosphoric acid to link two nucleotides yet still ionise, such that the resultant negative charge stabilises the diesters against hydrolysis[68]. The incomplete hydrolysis of phosphate ester linkages in chemical, non-enzymatic systems, is testament to their stability. For example, even under harsh conditions such as concentrated acidic medium (optimum activity at pH 4.5), high temperature (100°C) and an extended reaction time, hydrolysis is slow and completion is not accomplished (Fig 1.8)[29]. Despite undergoing slow hydrolysis in the absence of enzymes, phosphate ester bonds are subject to rapid hydrolysis under enzyme catalysis, which explains the dominance of this type of molecular species in biological systems[68].

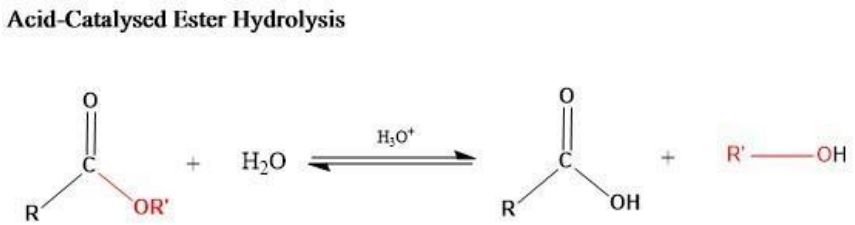


Figure 1.8. General reaction of chemical hydrolysis of phosphate esters, where R' in this case represents a phosphate group. Scheme produced in ChemDraw.

Phosphatases are a diverse class of phosphomonoesterase enzyme which catalyse the hydrolysis of a broad spectrum of phosphorylated organic compounds[69]. As members of this hydrolase family, phytases - myo-inositol (1,2,3,4,5,6) hexakisphosphate phosphohydrolases – are phytate-specific phosphatases, specifically hydrolysing IP6 in a stepwise manner to yield orthophosphate and lower substituted myo-inositol phosphates. They have sometimes been defined as a class of phosphatases with the *in vitro* capability to release at least one phosphate

from phytic acid, thereby liberating phosphate and lowering phosphorylation levels of inositol phosphates and potentially chelated minerals[27]. In accordance with the descending order of preference of phytases for IP6 through IP1, all of the fully phosphorylated IP6 present is first degraded to penta-esters of inositol before these then become the next substrate for hydrolysis to lower phosphorylated intermediates[70]. In addition, bound mineral cations such as magnesium and calcium and other nutritional factors are released and made available for absorption by monogastric livestock which are unable to access adequate phosphorus in the form of phytate, thereby improving nutritional quality of a given phytate-rich feedstuff and ameliorating the need for supplementation with inorganic phosphorus. Ruminant animals on the other hand, are host to their own in-house phytase producers – their anaerobic ruminal microflora – which facilitate the biological abstraction of phytate phosphorus. The gastro-intestinal tracts of non-ruminants such as swine, poultry and fish are mostly deficient in these analogous endogenous phytases, warranting intervention – i.e. by exploiting the blueprints of nature through biotechnology. Biotechnologically produced phytases thus have industrial applicability as livestock feed additives, and their incorporation represents a financial and environmental incentive.

Before defining the categories and nomenclature of this diverse enzyme family, it is pertinent to outline the different means by which monogastric species obtain phytase activity; and in particular to differentiate between the different derivations of endogenous phytases and exogenous phytases, although some of this may overlap with subsequent parts of section 1.3. Firstly, the (plant-based) feed source itself exhibits intrinsic phytase activity, the extent being highly variable and dependent on species. Additionally, the animal has two innate sources of phytase activity – that arising from their intestinal mucosa and that produced by their gut microflora – although it is well-documented that these are negligible and insufficient to improve phosphorus digestibility[43,46]. As such, the author of this thesis describes the exogenous dietary phytase – which can either be implemented during feed pre-treatment or as a dietary supplement - as the fourth 'layer' of phytase activity. These shall be duly attended to, after the following classification points are outlined.

Phytases can be classified by three different bases: by initiation site of hydrolysis; broadly by pH optimum/activity-pH profile; and thirdly by catalytic attributes and structural motifs. With the exception of the latter class, there are variations in catalytic mechanism within a given class (i.e. when classified by pH or preferred site of hydrolysis), as a result of structural variances[27]. Nevertheless, all currently known phytases share a common pronounced stereo- specificity and a strong preference for equatorial rather than axial phosphate appendages, generating *myo*-inositol monophosphate as the end product of hydrolysis[47].

1.3.1.1 Classification by pH Optima

The acidic, neutral and alkaline phytases are known to exhibit optimal activities at pH 5.0, 7.0, and 8.0, respectively[71]. Acidic phytases are predominant in fungi, plants and bacteria and due to the acidic conditions at the site of gastrointestinal absorption, are the more physiologically significant group. In contrast to the distribution of acidic phytases, the alkaline phytase group are primarily composed of extracellular β-propeller phytases of the gram-positive bacteria of the Bacillus genus. An alkaline phytase with a pH optimum of 8.0 is one of just two phytases which initiate dephosphorylation at the C5 inositol position. Variations in optimum pH for phytase activity occur across different plant species. In contrast to cereals, the phytases of which usually have greatest activity between pH 4.5 – 5.6, some legumes have optimum phytase activity at neutral or alkaline pH[72]. For example, legume seeds have also been reported to exhibit alkaline phytase activity, with a pH optimum of 8.0. Broadly speaking, phytases representative of the Histidine Phosphatase, Purple Acid Phosphatase and Cysteine Phosphatase groups are usually referred to as acidic and those of the β-propeller phytases as alkaline. In other words, acid phosphatases with specific activity for phytic acid are subdivided into three structurally distinct classes: HAPhy, βPPhy and PAPhy[73].

1.3.1.2 Classification by Site of Dephosphorylation Initiation

At the present time there are four types of phytases within this classification system, systematically termed *myo*-inositol hexakisphosphate-X-phosphohydrolases, X being '3', '6', '4' or '5'. They are more commonly referred to as 3-phytases, 6-phytases, 4-phytases and 5- phytases (although only the former two phytases are recognised by the IUPAC-IUB), whereby hydrolysis of the phosphoric ester bond is initiated at the D-3-, D-6-, D-4- and D-5- carbon positions of inositol, respectively, and is indicated by the resultant D-*myo*-inositol pentakisphosphate isomer[74]. Their initial hydrolytic products are as such: D-*myo*-inositol (1,2,4,5,6) pentakisphosphate (often abbreviated to Ins(1,2,4,5,6)P5), Ins(1,2,3,4,5)P5, Ins(1,2,3,5,6)P5 and Ins(1,2,3,4,6)P5, respectively. These phytases therefore exhibit differing phytate dephosphorylation pathways, giving rise to varying inositol phosphate degradation isomeric products. Disconcertingly, the 3- and 6-phytases are sometimes ambiguously (i.e. without clarity on which nomenclature system is used) termed 1-and 4-phytases, respectively, this being a result of the change in D/L nomenclature in the 1970's, as previously discussed in section 1.2.2.2[75]. In other words, a D-6-phytase and an L-4-phytase produce the same product, as do a D-3-phytase and an L-1-phytase. In this report they shall be referred to, as per the recommendation of the IUPAC-IUB to explicitly use the D system, as 3-phytases, microbial 6-phytases and plant 4-phytases – the latter often being ambiguously annotated as '6-phytases' without specifying that these are actually L-6-phytases and therefore, by the recommended D system, 4-phytases.

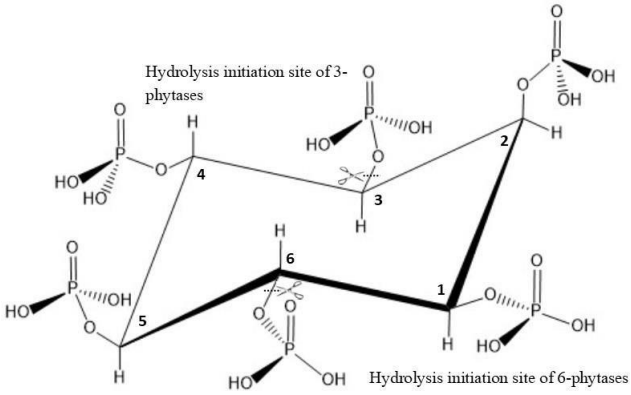


Figure 1.9. Structure of IP6 depicting the scissile bonds cleaved by 3- and 6-phytases as a general example of this phytase classification type. Image produced in ChemDraw.

The most abundant group in existence are the 3-phytases which are common products of bacteria and fungi, whilst the 4-phytases are generally of plant – particularly higher plants - origin[27]. However, there are a number of exceptions to this theme, such as the presence of 3-phytase activity in some plant seeds and 6-phytase activity in bacterial species, namely *E. coli* and *Paramecium* - the former giving rise to one of the most recognised phytase due to its establishment in commercial application, the topic of which will be discussed in more detail later in this report[74]. To date, there are merely two accounts of a 5-phytase – an alkaline phytase from Lily pollen and an alkaline phytase, PhyAsr, from *Selenomonas ruminantium* subsp. Lactilytica[76,77]. The latter is an anaerobic bacterium which resides in the rumen of polygastric animals and the sequence of this phytase is unlike other microbial phytases, having some catalytic features similar to tyrosine phosphatases. These phytases were reported to yield Ins(1,2,3)P3 as the end product of hydrolysis, whereby the hydrolysis of adjacent phosphates is preferential. Structurally, the Lily pollen 5-phytase is analogous in conformation to a Histidine acid phosphatase (HAP) phytase (see section 1.3.2), although the amino acid sequence homology of all but the active site is higher towards Multiple Inositol Polyphosphate Phosphatase (MINPP; section 1.3.6) from humans or rats[27]. Since no other HAP phytase initiates dephosphorylation at the C-5 position of inositol, the author of this thesis considers that this may be the reason for having more sequence similarity to the more promiscuous MINPPs. The majority of the 3-phytases are structurally homologous to two of the catalytic classes of phytase described below – β-propeller phytases (βPPhy) and the aforementioned HAP phytases[27]. Based on experimental reports, the sole end product of catalysis is reported as Ins(2,4,6)P3,, since the hydrolysis of non-adjacent, alternate phosphates is favoured[78]. The familiar baker's yeast, *Saccharomyces cerevisiae*, produces a now well-studied 3-phytase and its sequential phytate dephosphorylation pathway is illustrated overleaf (Fig 1.10)[74].

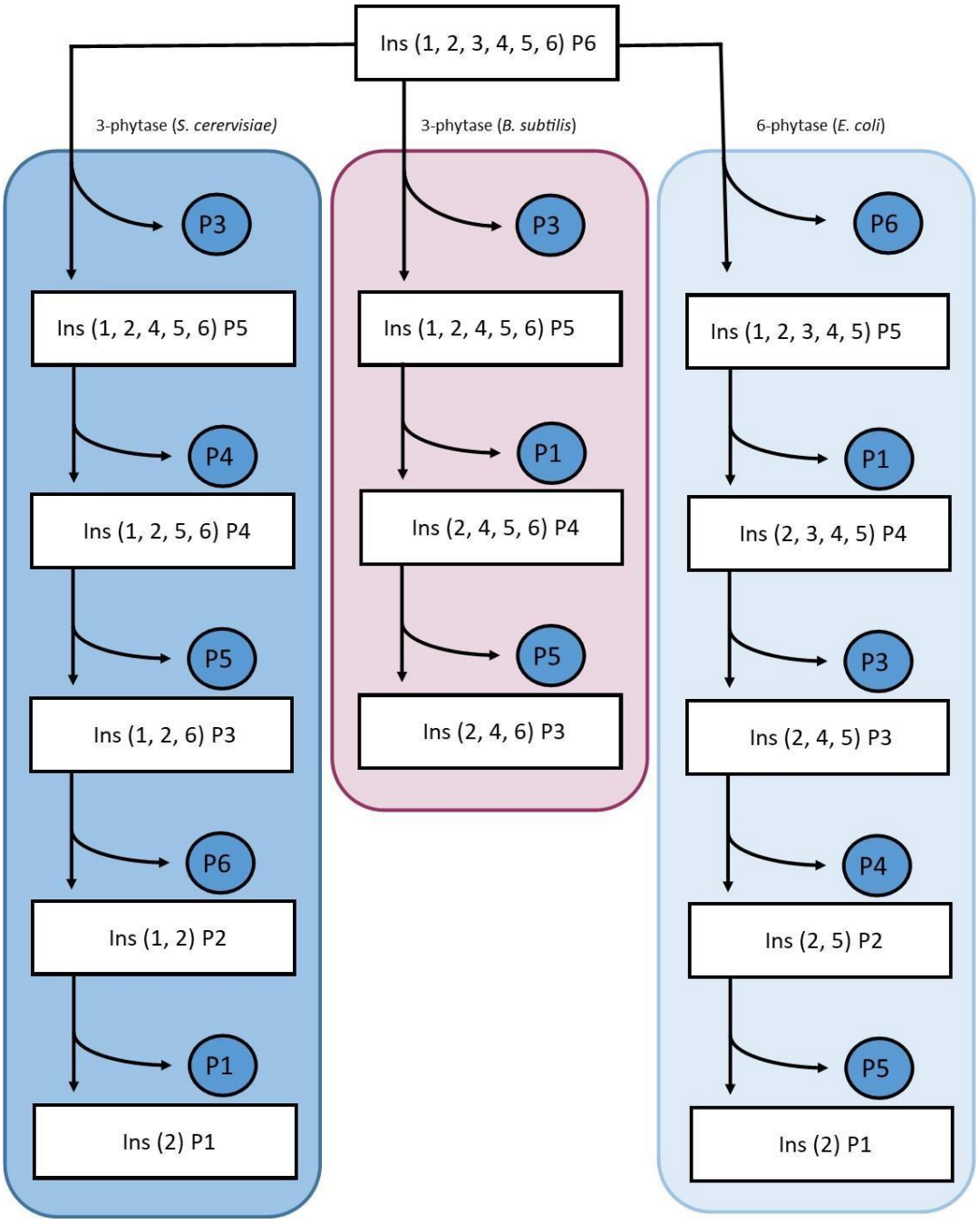


Figure 1.10. Visual representation of the IP6 degradation pathways by representatives of different phytase classes – a yeast HAPhy 3-phytase, a bacterial βPPhy 3-phytase and a bacterial HAPhy 6-phytase (AppA). The pathway for βPPhy is specifically for the Ca^{2+}-phytate substrate. Figure produced in a Microsoft graphics suite. Information from [74].

1.3.1.3 Classification by Catalytic Properties

Phytases are categorised as Histidine Acid Phosphatases (HAPhy), β-Propeller Phytases (βPPhy), Cysteine Phytases (CPhy) and Purple Acid Phosphatases (PAPhy) based on their active site residues and specific sequence attributes, and which differ in their structure, catalytic mechanism and biochemical properties. Most bacterial, fungal and plant phytases belong to the HAPhy family – a sub-family within the Histidine Phosphatase (HP) superfamily and commonly termed HP2 phytases (HP2P) since they belong to clade 2 of the HP superfamily[79].

1.3.2 Histidine Acid Phosphatases (HAPhy)

The histidine phosphatase (HP) superfamily is a large and functionally diverse group of enzymes, divided into two clades, whereby the second, smaller clade (HP2) is composed of phytases and acid phosphatases. There is limited sequence similarity between the two clades and even within each clade, and whilst clade 1 consists of more diverse intracellular enzymes of largely (two thirds) bacterial origin, clade 2 comprises predominantly eukaryotic extracellular enzymes. These are channelled to the secretory pathway, and unlike their clade 1 counterparts, contain disulphide bonds. A conserved catalytic core is common among all members of this superfamily, comprising an integral histidine residue and three other invariant amino acids (a pair of flanking arginine residues and another histidine), forming the 'phosphate pocket'. Phosphorylation and dephosphorylation of one of these histidine residues forms the basis of the two-step catalytic mechanism (Fig 1.11). During catalysis the key histidine residue of the conservative heptapeptide RHGXRXP motif (clade 1 HPs = RHG only) acts as a nucleophile, anchored in position by the carbonyl of a glycine (part of the same characteristic N-terminal RHG motif), whilst the interaction of the positively charged arginine in the RHG tripeptide with the substrate phosphorus renders it more susceptible to nucleophilic attack. In the first stage, this histidine attacks the substrate-bound phosphorus generating a phosphohistidine intermediate, whilst in the second mechanistic stage an aspartic acid of the catalytic C-terminal dipeptide (HD) – which is responsible for binding the substrate and release of the hydrolytic product - acts as a proton donor for the oxygen atom. This thereby stabilises the leaving group as transfer of phospho group from substrate to enzyme/hydrolysis of the phosphohistidine intermediate occurs. The three core residues of the phosphate pocket function to maintain correct orientation of substrate, facilitating hydrolysis. A third conserved feature is the cysteine motif which is associated with the formation of disulphide bridges, these being integral to the thermo- and conformational stability, and catalytic activity. This family of phosphoesterases initiate dephosphorylation at either the C-3 or C-6 position of inositol[68,73,79].

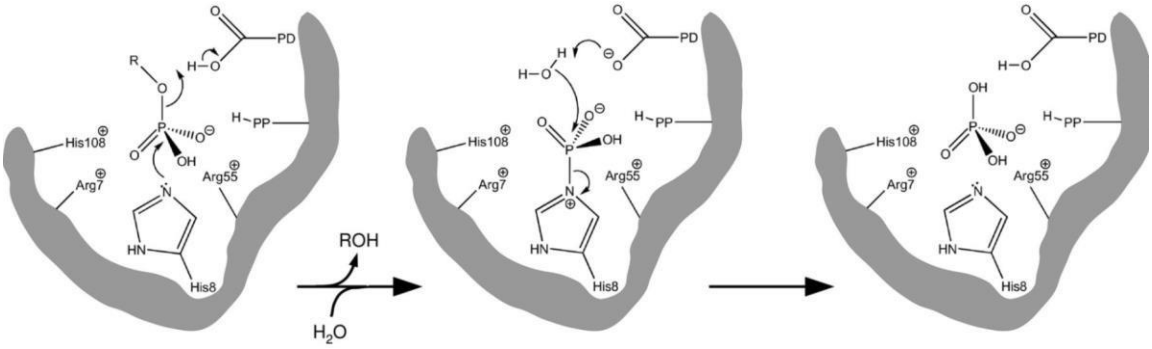


Figure 1.11. Catalytic mechanism of the histidine phosphatase superfamily. Residue numbering from *E. coli* SixA, a phosphohistidine phosphatase. Image as found in [79].

1.3.3 β-Propeller Phytases (βPPhy)

Phytases designated βPPhy constitute a more structurally distinct group of phytases, having an unusual six-bladed beta-propeller architecture, and attributes unlike those of other phosphatase families. Their Ca^{2+}-dependent catalytic activity, high thermostability, activity at non-acidic pH and absolute substrate specificity - accepting the calcium-phytate complex only
- is testament to this. Predominantly of microbial origin, the majority of these phytases have been isolated from the Bacillus genus, some examples of structurally characterised ones being phyC from *Bacillus subtilis* and TS-Phy from *Bacillus amyloliquefaceiens* DS11, the latter being the representative βPPhy in terms of delineating the catalytic mechanism of these phytases[73,80]. The inherent desirable biotechnologically relevant properties of particular Bacillus species is such that they are well-suited to industrial application. βPP-producing microbes are prevalent in the natural environment, isolated from such sources as water ecosystems, soil, sediments and animal gastrointestinal tracts[81]. An important difference of βPPhy phytases in contrast to other phytases is their stability over a broad pH range (3.0 -9.0), which gives them advantages over other phytases in certain applications. In general, the optimum pH range for activity of Bacillus βPPhy phytases is 6.0-7.5, yet they still exhibit minor activity above pH 8[81]. The distinctive thermostability and catalytic properties of βPPhys are characteristically dependent on the binding of Ca^{2+} ions which, acting as activators, render the catalytic environment electrostatically favourable for substrate binding by modifying the active centre conformation to promote the substrate's approach. The three tightly bound calcium ions, in the high-affinity site, serve to maintain conformation and stability whilst three low affinity calcium binding sites comprise the catalytic triad responsible for activity[81]. The strict substrate specificity is another attribute of the bound calcium, whereby the recognition of bidentate chelation of calcium ions by two neighbouring phosphate groups of IP6 is essential

for hydrolysis of the substrate (Ca^{2+}-phytate) and is a substrate specificity determinant. In fact, the molar ratio of Ca^{2+}: phytate in the substrate complex, as well as the medium's pH, influences the enzymatic activity, being maximal at a ratio of four calcium ions for every phytate molecule[81]. This particular substrate recognition facilitates the stereospecific dephosphorylation of phytate by the successive detachment of each second phosphate group, generating as the final hydrolytic product, *myo*-inositol (2,4,6) triphosphate. Although initially it was thought that the end product was either Ins(1,3,5)P3 or Ins(2,4,6)P3, more recent research has concluded that the latter is the sole product[81,82]. Aside from the bound calcium ions - both as part of the substrate complex and bound to the phytase - the prerequisite for hydrolysis is the binding of two adjacent phosphate moieties, in the so called cleavage (or degrading) and affinity sites, respectively, which are both located in the region of the low-affinity calcium binding sites[73]. The former phosphate-binding site, involving the same catalytic residues which bind the calcium ions, is responsible for hydrolysis of phosphate ester bonds in the substrate complex – preferentially at the D3 position of inositol – and is the site of nucleophilic attack on the phosphorus by a calcium-bound water molecule. βPPhys are as such classified as (D)3-phytases. The role of the affinity site is to increase the affinity of substrate binding. Intermediate products generated during the successive dephosphorylations at C3 followed by C1 then C5 are *myo*-Ins(1,2,4,5,6)P5 and *myo*-Ins(2,4,5,6)P4 before the final product, *myo*-Ins(2,4,6)P3, respectively. Further hydrolysis is prevented by the lack of neighbouring phosphate groups whose proximity is necessary for the formation of the strategic bidentate chelate with calcium[80]. Only substrates that simultaneously fill both binding sites are hydrolyzed by βPPhys which explains why these enzymes can only remove three phosphates from IP6[81].

1.3.4 Purple Acid Phytases (PAPhy)

A sub-family of the metallohydrolase superfamily - the metallophosphoesterase family - includes the dinuclear purple acid phosphatases. These enzymes catalyse the hydrolysis of a broad phosphomonoester bond-containing substrate range and have been identified in both plant, animal and fungal sources, but are unlikely to occur in most microorganisms[83]. They function to aid in phosphate acquisition in plants and have diverse roles in animals, in particular within mammalian bone resorption, and exhibit bifunctionality in mammalian species, catalysing both hydrolytic and peroxidation reactions. Despite low shared sequence identity (<20%), size and oligomeric state between animal and plant PAPs, the catalytic active sites and domains are in contrast highly conserved. The key feature of the catalytic core, which is representative of all metallohydrolases, is the presence of two closely spaced metal ions – Fe^{3+} and M^{2+}- which represent a mixed-valent metal centre and constitute the chromophoric

and redox-active sites, respectively. In plants, M^{2+} represents divalent zinc or manganese whereas in animals it is always divalent iron[83,84]. The seven residues/ligands which co-ordinate the dinuclear centre, encompassed within five key motifs, are also conserved across kingdoms. A particular distinction of PAPs is their characteristic purple colour when in solution owing to a ligand to metal charge-transfer transition around 515-560 nm from the conserved tyrosine ligand to the chromophoric ferric ion it coordinates[83,84]. Additional characteristics of the PAPs are substantial glycosylation and resistance to inhibition by L-tartrate. The presence of isoforms – different forms of the same enzyme - are common in plant and animal PAPs and are distinguished by their molecular weights. These often occur as a 35 kDa form and a 55-60 kDa form and plants tend to have multiple copies of each. The catalytic hydrolysis of activated phosphoric acid esters and anhydrides involves eight postulated steps and two nucleophilic hydroxide ions, one of which is coordinated to Fe^{3+} and attacks the phosphate after an interaction between the divalent metal and the substrate. The majority of characterised PAPs are designated as non-specific acid phosphatases that catalyse phosphate hydrolysis from a broad spectrum of phosphate esters and as such, only certain isoforms accept phytate as a substrate, referred to as PAP phytases or PAPhy[83]. Up until 2005[85] the soybean (*Glycine max*) PAP isoform (*Gm*Phy) was the only PAP phytase which had been characterised, although these are indicated, as per sequence homology, to be common amongst plants[73]. Since then, PAP phytases have been identified in *Medicago sativa* (alfalfa), *Nicotiana tabacum* (tobacco) and multiple isoforms within *Arabidopsis thaliana* (thale cress)[83].

1.3.5 Protein Tyrosine Phosphatase like Cysteine Phytases (CPhy)

Also known as cysteine phytases due to their catalytic residue, protein tyrosine phosphatase-like phytases are found exclusively in microbes with the first member – Phy*Asr* - identified in the anaerobic ruminal bacterial species *S. ruminantium*[86,87]. In fact, these phytases are the major phytate-degrading enzymes resident in the rumen microbiome of cattle[88], negating the need for addition of exogenous phytases to cattle diets. Of the four phytase classes, phytases of the PTP-like/cysteine phytases are relatively the least well characterised[89]. With the adoption of the protein fold and catalytic mechanism of protein tyrosine phosphatases, PTP-like phytases are composed of a large core domain and a smaller domain. A four-stranded β-sheet, sandwiched on both sides by several α-helices, constitutes the larger domain whilst a partial β-barrel comprised of a five-stranded β-sheet constitutes the smaller domain[90]. At the interface of the two domains, the active site forms a loop which functions as a phytate-binding pocket[86]. This P loop hosts the active site signature sequence (CX5R(S/T))[90]. Catalysis is initiated by a nucleophilic cysteine which results in a thiophosphate enzyme intermediate,

and an aspartic acid residue on the WPD/general acid loop acts as the proton donor to hydrolyse the scissile phosphoester bond[90,91]. PTP-like phytases exhibit high levels of hydrolytic activity towards phytate with pH optima ranging from pH 4-6. Despite this, the nature of the catalytic mechanism, in which potential irreversible oxidation would inactive the enzyme, limits their commercial application[90].

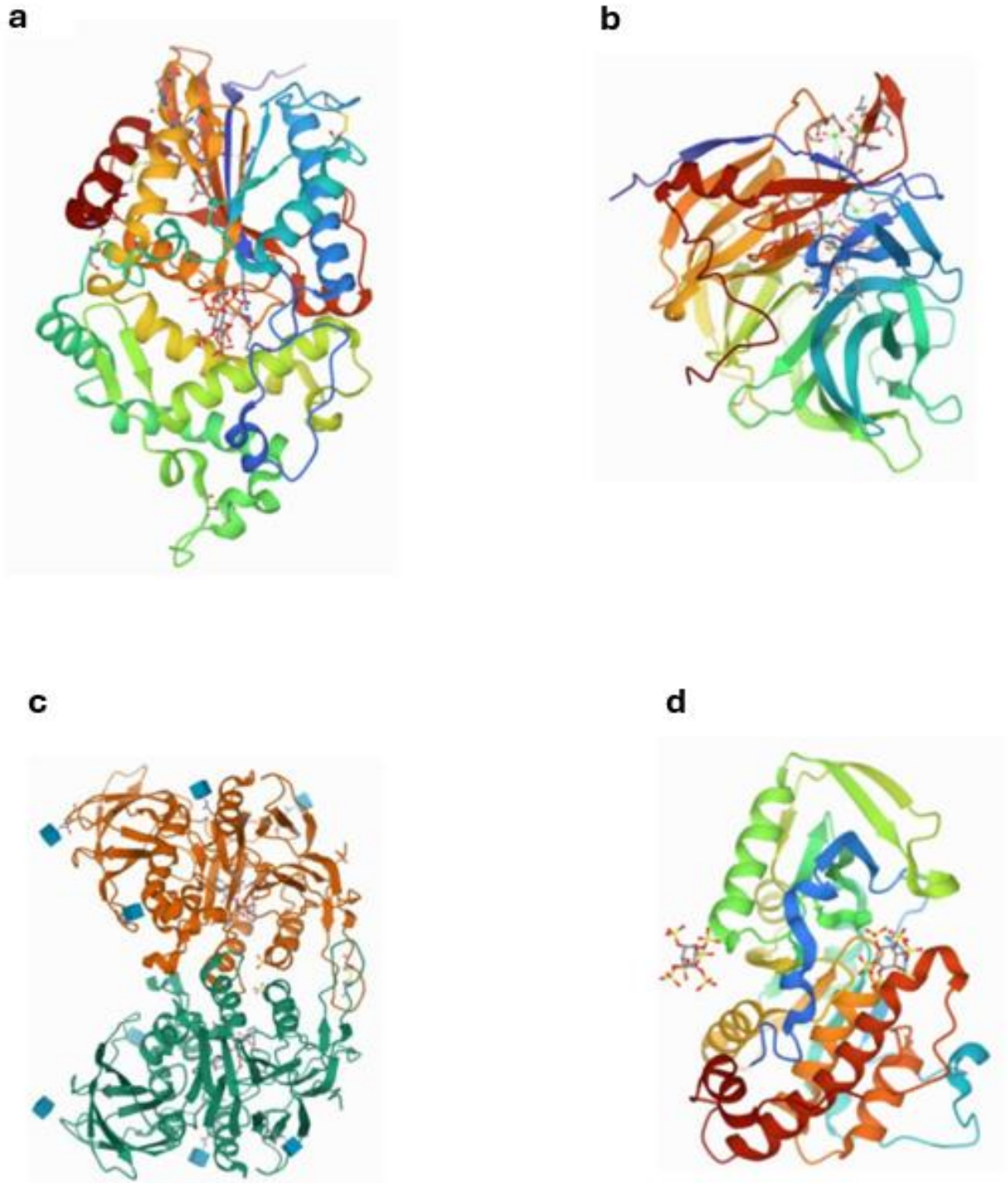


Figure 1.12 Crystal structures of a) HAPhy from *Escherichia coli* AppA in complex with phytate. Phytate is bound with its 3-phosphate in the active site. Image from PDB entry 1DKQ; b) β-propeller phytase from *Bacillus amyloliquefaciens* in complex with phosphate and Ca^{2+} ions. Four Ca^{2+} ions are involved in catalysis and creation of a favourable electrostatic potential; three stabilise the enzyme. Image from PDB entry 1H6L. c) PAP from red kidney bean (*Phaseolus vulgaris)* with bound sulphate. Blue cubes represent N-acetylglucosamine (NAG) residues. Image from PDB entry 2QFR. d) CPhy from *Selenomonas ruminantium* complexed with myo- inositol hexasulphate in the standby position and in the active site.

1.3.6 Multiple Inositol Polyphosphate Phosphatases (MINPP)

The so-called Multiple Inositol Polyphosphate Phosphatases – MINPP – are a family of IP6-degrading enzymes, belonging to clade 2 of the HP superfamily, yet distinct from HAP phytases. Their positional promiscuity towards IP6 specifically, with regards to the site of initial dephosphorylation, is the distinguishing feature of this family in contrast to other members of the HAPhy family. Whereas the latter follow a set sequential pattern of phosphate hydrolysis, and are classified by a distinct preference for position of first phosphate removal, MINPPs generate an array of IP5 isomers[92]. The isolation and characterisation of a MINPP from rat liver marked the first report of a MINPP, in 1997[93]. Whilst having no preference for initial site of hydrolysis amongst the equatorial phosphates of IP6, this phytase is highly specific with Ins(1,3,4,5,6)P_5 , converting it into the key Ca^{2+}-mobilising inositol triphosphate Ins(1,4,5)P_3 via Ins(1,4,5,6)P_4[94]. The activity of mammalian MINPPs against IP6 is relatively very low, and phytases from a range of microorganisms exhibit between 10^3-fold and 10^4-fold higher specific activities compared to the rat MINPP[95]. Nevertheless, they are the only mammalian enzymes to hydrolyse IP5 and IP6, with the attack at D-3, D-5 and D-6 of IP6 yielding a mixture of IP5s[96]. Within a year of identification of the rat MINPP, the chick MINPP HiPER1 (Histidine Phosphatase of the Endoplasmic Reticulum-1) was characterised[97]. By the end of the 90's, a number of MINPPs had been sequenced – *Homo sapiens, Mus musculus*, two from *Drosophila melanogaster*, *Rattus norvegicus* and *Gallus gallus*. Whilst avian MINPP is considerably more active towards phytate than its mammalian homologues, the specific activity is still 30-300-fold lower than that of microbial phytases[95]. To date, MINPPs have been identified from various species of Bacteria and Eukarya. Whilst the physiological function of several of these are yet to be confirmed, they are assumed to have a key role in providing bioavailable phosphate to developing cells and simultaneously generating valuable physiologically downstream inositol phosphate metabolites[98]. For example, the MINNPs from barley and wheat were presumed to contribute significantly to the endogenous phytase activity of the developing seed. The most recent plant MINPP characterised to date is the MINPP from *Arabidopsis thaliana* (*At*MINPP) which, although sequenced in the late 90's, has only now been shown to be involved in leaf senescence and has acid phosphatase activity in addition to phytase activity. Experiments revealed the essentialness of *At*MINPP for growth and development[99].

Homology modelling of the avian MINPP suggested that the enzyme has a conserved, phytase-like active site, despite sharing a low sequence similarity of 20% with phytases[95]. The MINPP from the human gut bacterium *Bacteroides thetaiotamicron*, however, represents the first MINPP to be structurally characterised by crystallographic analysis[92]. Like other HP2P enzymes, the active site motif RHGxRxP is present in MINPPs, however, instead of an HD proton donor motif, they have a HAE triplet which is assumed to provide an equivalent function[92]. In both cases, the acidic residue functions as the proton donor, breaking the phosphomonoester

bond to release a phosphate group from phytate[92]. *Bt*MINPP has provided a good model on which to base structural mutagenic investigations, and this is covered in more detail in Chapter 3. To date, several MINPPs have been identified as containing HAE proton donor motif outliers, including from maize, zebrafish, mosquito, the amoeba *Dictyostelium discoideum*, a yeast and two from *Drosophila melanogaster*. Further investigations of MINPP structures revealed that some contain a large polypeptide insertion in the α-domain – the U-loop – which acts as a cap and undergoes a substantial movement upon substrate binding[100]. The exemplary phytase of this type is the membrane-anchored extracellular MINPP from the Gram-positive bacterium *Bifidobacterium longum* subsp. infantis ATCC 15697 (*Bl*MINPP), which is prevalent in the human gut. *Bl*MINPP generates the 4/6-OH IP5 as the major IP5 isomer, and studies suggest that dephosphorylation does not proceed beyond IP3[101]. Along with the MINPP from *Bifidobacterium pseudocatenulatum* ATCC 27919 (*Bp*MINPP), these enzymes represent the first bacterial MINPP homologues identified (in 2012). Studies showed that the preference of *Bp*MINPP is to generate first either the 4/6-OH or the 5-OH IP5 isomer, and subsequently the other of the two, generating Ins(1,2,3,4)P_4; it then proceeds to generate any/all of Ins(1,2,3)P_3, Ins(1,2,4)P_3, Ins(1,3,4)P_3 as the final product (similar to *Bl*MINPP)[102]. The alkaline 5-phytase from *Lilium longiflorum* (mentioned previously) is another example of deviation from the standard, whereby it has not been strictly classified as a MINPP as such, yet is more closely related by sequence to human (25%) and rat (23%) MINPPs than to other HAPs[103]. Regardless, unlike MINNPs, it shows a distinct preference for the 5-phosphate for initiation of hydrolysis. Yet like the bacterial MINPP homologues, IP3 appears to be the final product of hydrolysis, specifically Ins(1,2,3)P_3[103]. A recently identified MINPP from soil Acinetobacter sp., AC1-2, displaying 5-phytase and 6-phytase activity, is one of the first MINPPs to be isolated from the soil environment and has a slightly different heptapeptide sequence, RHGSRGL[104].

1.3.7 Sources & Distribution

The occurrence of phytases is diverse and well documented in plants, bacteria, fungi, yeast and to a lesser degree, in animals. Eukaryotic phytases were recognised initially, evidenced first by the identification of phytase activity in rice bran in 1907 (by Suzuki, Yoshimura and Takaishi) and in calf blood in 1908 (by McCollum and Hart)[41,105,106]. Microbial phytase activity was first discovered in the fungus *Aspergillus niger* mycelium over a century ago[107], with the earliest report by Dox and Golden in 1911 who demonstrated that the fungus could convert organic phosphorus into available phosphorus through the action of a secreted enzyme[108]. Following Jackman and Black's report detailing the detection of low levels of phytase activity in several soils in 1952, several years later in 1959 Casida observed phytase activity from a number of

Aspergillus soil isolates[109,110]. It wasn't until 1967 that a microbial phytase – produced from a laboratory culture of *Aerobacter aerogenes* - was more rigorously investigated, with Greaves, Anderson and Webley yielding information on the hydrolytic pathway and biochemical characteristics[111]. 1969 saw reports of phytase activity from a species of Pseudomonas soil bacteria, by Cosgrove, Irving and Bromfield[112,113], and characterisation of an extracellular phytase from *Aspergillus ficuum* NRRL 3135 by Shieh, Wodzinski and Ware[114]. The pathway of IP6 degradation was further analysed in 1972 by Irving and Cosgrove which identified the major IP5 intermediate generated[115]. This phytase would go on to become the first commercial phytase – Natuphos® – launched in 1991.

Aside from exogenous phytase preparations, sources of endogenous phytases are: intrinsic plant phytases; intestinal mucosa phytases; and phytase-producing gut microbiota, as illustrated in Fig 1.13. Phytate may therefore be subject to a multifaceted degradation approach through both endogenous and exogenous enzyme action. Although not beyond the scope of this thesis, due to time restrictions these various sources of phytase activity shall not be elaborated on, however there are multiple comprehensive review articles covering this, one example being that by Rizwanuddin et al[117],

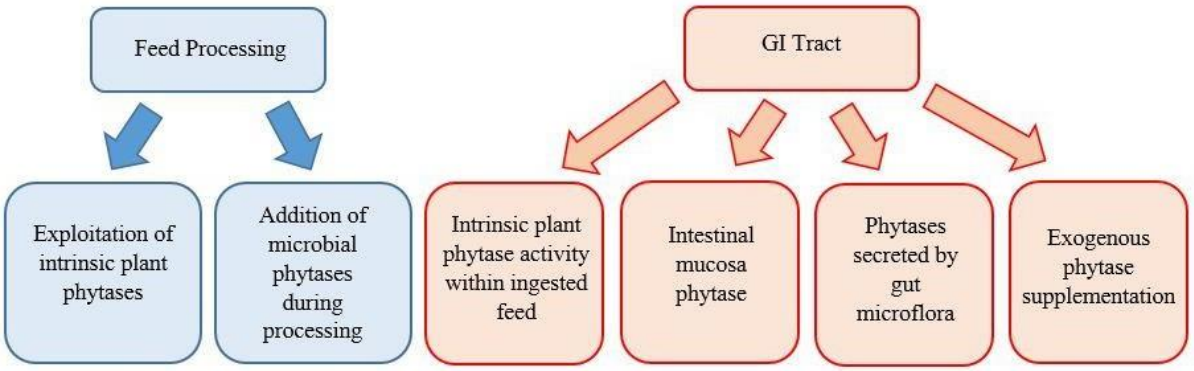


Figure 1.13. Flow diagram created by author as a representation of the various origins of phytase activity in monogastric animals. Compiled using information from sources cited in current chapter.

1.4 Phytase Application in the Feed Industry

1.4.1 Rationale for Phytase Supplementation

The fundamental role of the livestock industry is the optimisation of animal production – sustainably – necessitating the reiteration of the ethos from section 1.1 - *more from less*. In order to sustain the production of animal derived consumables (meat, milk, eggs) in accordance with the heightened demand for animal protein, it is in the best interests of producers and the global food supply to maximise output, which is highly dependent on and therefore facilitated by the feed inputs. However, the economic pressures which the livestock producers face – emphasized in developing countries - demand the use of lower quality feed, necessitating the capability to enhance the naturally low nutritive quality by technological means[118]. Nevertheless, the benefits of the nutritional value of even the highest quality feed are only seen upon effective digestion and absorption of said nutrients. A significant proportion of plant material in feed is inherently indigestible by monogastric animals and as such, exogenous enzymes are incorporated into livestock diets to improve animal performance through improvement of nutrition. The inclusion of preparations of carbohydrases, proteases and phytases all work to enhance the availability of nutrients and therefore improve animal growth and productivity. Phosphorus is absorbed as the inorganic orthophosphate form, rendering the utilization of phytate phosphorus dependent on the capability of a species to hydrolyse phytate to liberate phosphate[118,119]. Phosphorus in the form of phytate, as in animal feed, is poorly available to monogastric animals since they lack significant endogenous phytase activity in - and have inadequate phytase-secreting microflora populations in - the upper part of the digestive tract. When added to animal diets, exogenous phytase augments the limited range of endogenous phytate-degrading enzymes present in the gastrointestinal tract, releasing bound inorganic phosphate and other important dietary nutrients from phytate, making them available for use by the animal thereby obviating the need to supplement diets with inorganic phosphorus, the price of which is anticipated to rise over time in accordance with the decreasing global phosphorus reserves. It also increases the energy of the animal since they no longer have to expend energy on the secretion of endogenous phytases. Phytase has been shown to replace 50-60% dicalcium phosphate in feed[44]. The resultant financial savings encourage the use of these lucrative biological tools whilst protecting the environment through reduced manure phosphorus outputs. For example, the use of phytase has been demonstrated to reduce phosphorus excretion by up to 50%, facilitating compliance with environmental regulations (see section 1.2.2.6)[47]. Since the feed is the most costly input in livestock production – accounting for e.g. 70% (poultry) of the total cost - not only does the

reduced inorganic phosphorus consumption generate monetary savings but also the amount of, for example, soybean meal in the diet can be reduced, therefore increasing profit, since phytase improves its amino acid digestibility. Feed enzymes therefore offer significant economic benefits both by saving on supplement/feed costs and by improving the 'bang for buck' (feed conversion ratios; FCR) e.g. production of a given quantity of meat in a shorter amount of time and at lower cost, or greater productive output from a given animal/number of animals, resulting in greater profit margins. Feed enzymes are thus attractive investments.

1.4.2 The Feed Enzyme Sectors

The history of the commercial application of enzymes to feed for livestock is relatively brief, spanning less than forty years[120,121]. Despite many decades of, for example, phytase research, following its discovery at the start of the 20th century[105,106], due to the technological limitations of that era, their commercial application did not become viable until the 1990s, when the large-scale production of feed enzymes became possible at relatively low cost, owing to biotechnological advancements such as the development of heterologous microbial expression systems[44]. In addition to the practical limitations, prior to the 'enzyme revolution', there was little pressure from environmental regulations, coupled with the perception that supplemental enzyme solutions were cost-intensive and therefore not economically viable for livestock production. Around one hundred years prior to the enzyme industry's take-off in the latter part of the 20th Century, a Japanese individual -adapting a Japanese technology to a western industry -patented a process, licensed under 'Taka-diastase', for the production of α-amylase from *Aspergillus oryzae*, marking the first patent on a microbial enzyme in the USA (1894; U.S. Patent No. 525,823)[122]. However, the earliest report of the use of an enzyme in animal feed was in 1925 when a fungal enzyme was applied to poultry diets[121]. Swine and poultry do not have the necessary enzymes in their endogenous arsenal to break down and utilize *all* components of their diet, which led to the development of enzymes targeting specific dietary components, these enzymes being β-glucanases, xylanases, phytases, proteases, lipases and galactosidases. The introduction of the first commercially established feed enzymes marked the first phase of the industry's development and began with the glycanases. Encompassing arabinoxylanases and β-glucanases, glycanases address the degradation of the indigestible fibrous material of grain feed – the anti-nutritive non-starch polysaccharides (NSP). NSPs are components of the cell walls of viscous grains such as wheat, rye, barley and triticale and are not degradable by non-ruminant livestock. They include cellulose, hemicellulose, pectins, α-galactosidase, β-glucans, mannan and xylans[120,122-124]. Carbohydrases hydrolyse these NSPs and thereby aid the release of nutritional constituents such as proteins, starch, lipids and other minerals that are trapped within the cell wall matrix[124]. It is presumed that by 1996 at least 80% of all European

broiler diets containing a viscous cereal would have included a fibre-degrading enzyme[125]. Driven by the increasing pressure to comply with environmental standards, efforts were later focussed on reducing the amount phosphorus entering the environment as a direct result of livestock production. Subsequently, the market introduction of phytases at the beginning of the 1990s represented the second wave of the feed enzyme generation and paved the way for further generations of phytases boasting enhanced properties. As such, livestock production in today's modern era sees 90% (poultry)/70% (swine) of feed for monogastric animals supplemented with phytase, this having risen in response to the introduction of TSE (Transmissible Spongiform Encephalopathies) feed controls during the mid-1990s which outlawed processed animal protein (PAP) in feed for livestock (see pg 31). This meant more undigested phytate phosphorus was released into the environment due to diet composition[120]. Feed enzymes are now established as an indispensable tool in livestock production and their overarching benefits are overt, both maximising profitability through allowing flexibility in feed formulation (i.e. least cost feed formulation) without compromising nutritional value, providing savings on nutritional supplements and contributing to efforts to minimize discharge of environmentally damaging byproducts associated with intensive livestock production[120,122]. The global feed enzymes market size was valued at USD 1.47 billion in 2024 and is projected to reach USD 2.76 billion by 2031, growing at a compound annual growth rate (CAGR) of 8.20%[126]. Relatively recent (2018) figures report that 55-60% is attributed to food and feed applications, with xylanase, β-mannanase, β-glucanase, α-amylase and phytase constituting the majority of feed enzymes used currently[127,128]. In fact, the global phytase market was estimated (at the first international phytase summit in 2010) to account for over 60% of the total feed enzyme market[129]. As per the past decade, the animal feed sector is currently dominated by poultry, the industry leader, which accounted for the largest share (by value) at 44.2% of the market in 2023[130,131]. Whilst the aquafeed enzyme sector represents the lowest value, valued at USD 483.0 million in 2024[132], it is forecasted to be the fastest growing segment in the feed enzymes market segmented by livestock at a CAGR of 10.05% during 2023-2028[133]. This is illustrated in Fig 1.14 which represents the 2023 feed enzymes market shares by livestock type[130]. In terms of enzyme type, whilst carbohydrase is the fastest-growing segment, phytase holds the largest share, with the global feed phytase market valued at USD 450 – 727 million (depending on market research source) in 2023[133-137] and thereby accounting for up to 43%. Overall, it is thought that the best-selling enzymes for non-ruminants are phytases with 60% of the sale market, carbohydrases with 30%, and proteases and lipases representing the remaining 10%, with 80% of the carbohydrase share corresponding to xylanase[138].

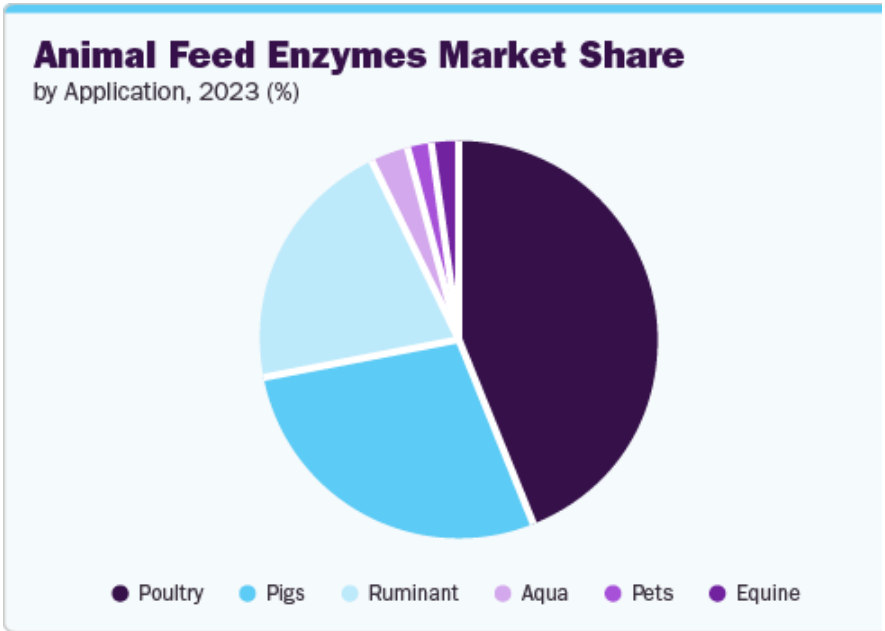


Figure 1.14 Breakdown of the feed enzymes market share by sector in 2023

1.4.2.1 Poultry

Currently one of the largest and rapidly expanding agricultural production sector globally – with particular emphasis in developing countries - the poultry division of the feed industry was the pioneer of exploiting the use of enzymes in animal feed, instigated by the production of the commercial preparation 'Protozyme' - a combination of fungal enzyme products of *Aspergillus oryzae* – as early as in the 1920s[46,139]. This was succeeded by experiments in the late 1960s and early 1970s by Nelson *et al* who reported the phosphorus utilization improvements of introducing crude phytase from cultures of *Aspergillus ficuum* into poultry feed[119]. Similarly, it was noted by Warden and Schaible in 1962 that the cellular material of lysed *E. coli* improved poultry bone growth and development[140]. Nevertheless, it was the requirement of counteracting the detrimental effects of the poorly digestible NSPs that ignited the commercial development of carbohydrase enzymes, which have improved the nutritive value of barley for poultry since the 1980s and are today a standard dietary component. This is testament to their contribution to the fundamental nutritional advancements that feed enzymes offered. NSPs, the degradative enzymes for which are non-existent in poultry, were identified as responsible for gut viscosity and correspondingly, impaired nutrient digestion, as well as poor excreta consistency, such that their elimination represents the most common use of feed enzymes in general over the last three decades. The resulting low feed efficiency and less than ideal animal health has financial implications for the producers as output and hence profit is lowered. Depolymerisation of these NSPs into smaller constituents prevents their many detrimental effects, resulting in increased digesta passage rate, more available energy through generation

of usable carbohydrate forms, improved nutrient digestibilities and reduced excreta moisture content. As such, the use of NSP-degrading enzymes has enabled the use of less expensive, lower quality feed whilst circumventing the consequences of the higher percentage of NSPs present in these feedstuffs[120,122]. This is of great benefit since feed accounts for the greatest expense in poultry production systems, at around 70% of total production cost per individual bird[122]. Of the feed components, supplemental inorganic phosphate is one of the most expensive thus replacement of this by phytase inclusion results in more economic poultry production, since phytase has been shown to increase the digestibility of phytate from around 25% to 50-70%[46]. The growth of the global poultry sector is anticipated to continue to increase in accordance with the demand for meat and eggs arising from the expanding populations, incomes and urbanisation.

1.4.2.2 Aquaculture

The commercial aquaculture industry has been undergoing rapid expansion since 2011 and is often cited as the fastest developing food-producing sector at present. Within the past decade, its market share has been increasing by 17% annually, with an average annual growth rate of 6.7% over the past three decades, which is expected to decelerate as aquaculture matures and production figures rise. This is owing primarily to the increasing demand for fish for human consumption, driven by development in the less economically advanced regions of the world and shifting consumer preferences and perceptions of aquaculture worldwide[131,141]. Global aquaculture is unevenly distributed, with Asia being the main producer representing 91.6% of global production (and 85% of value) in 2020[141]. According to the Food and Agriculture Organization of the United Nations (FAO), global consumption of farmed fish exceeds that of beef on a weight basis[131]. It is then unsurprising that the aquafeed production sector is proliferating concurrently - propelled by the exponential growth of the aquaculture industry in the last decade - and had been expanding at a rate of more than 30% per year*, which is more than triple that of terrestrial livestock production. *Since 2000, the average annual growth rate of feed production has been 10.3 % per year[142]. The growth of the industry is expected to be maintained by the expanding middle class populations in developing regions. However, the industry is outgrowing its traditional feed supply, this being wild fish extracted from the ocean to feed farmed fish. Fishmeal as a feed (for both farmed fish and domestic animals) is becoming increasingly less acceptable since it is responsible for the depletion of global ocean fish stocks and the imbalance of marine ecosystems, attributed to sustained mismanagement over the years. This lack of sustainability has resulted in over-exploitation of over a third of the global fish resources and inevitably induced conflict between the demands for fish for human versus non-human consumption. This has understandably put further pressure on the aquaculture

industry – the principle consumer of fish-based feed - to meet the demand, whilst maintaining long-term sustainability. The industry has lagged behind its terrestrial counterparts in terms of technological solutions and production scale, such that the use of feed enzymes in aquaculture is at present disproportionately under-exploited in contrast to the customary enzyme supplementation in the other livestock industries. However, the inevitable shift away from the use of fishmeal as feed is creating growth opportunities for the aquafeed enzymes market, as the alternative to fish meal - plant-based feeds - contain many poorly digestible components which the use of enzymes exploit. Indeed, the demand for plant-based aquafeed is increasing in parallel with the global consumption of fish, since plant-based ingredients are increasingly constituting a substantial portion of the diet in accordance with sustainability and cost factors. Formulated aquaculture feeds generally account for 50 to 60 % of total production costs[143]. In order to meet the future demand, for the mutually exclusive consumers and fish feed, productivity from the remaining available resources must be optimised in a sustainable manner, which is where enzymes play a key role[144]. As with poultry and swine production systems, the addition of phytase to the diet helps to alleviate the environmental concerns, specifically excessive phosphorus discharge in water from intensive fish farming. OptiPhos® was introduced to the aquafeed market by Huvepharma in 2008 which has, under diets of reduced fishmeal inclusion, proven efficacy in reducing phosphorus excretion through better utilisation and improves growth and feed conversion[145]. It is currently predicted that the global aquafeed additives market will reach US$2.2Bn within the next decade, almost double that of 2019. For a detailed review of the aquaculture industry, the reader is referred to the recent article by Naylor et al[146].

1.4.2.3 Swine

Exogenous enzymes have seen commercial use in swine diets for the past 30 years[147]. As with other monogastric livestock, the supplementation of pig diets with exogenous enzymes represents an alternative solution to improve productive performance, by increasing dietary energy and fibre digestibility at a lower cost to both the environment and producers. Same as for poultry, feed for pigs constitutes a large proportion of NSPs (namely arabinoxylans, cellulose and β-glucans) which are poorly metabolised by pigs, which, as monogastric animals, lack the specific endogenous enzymes for their degradation[124]. Taken together with the fact that pig feed accounts for 55-75% of total production costs[148], the addition of supplementary carbohydrases enables lower cost feed containing higher levels of NSPs (e.g. barley, wheat and grain co-products such as dried distillers' grains) to be used[124]. The carbohydrases most commonly used in swine diets are β-glucanase and xylanase, although β-mannanase, α-amylase, cellulase, pectinase, α-galactosidase and others are commercially available[149].

In fact, in a recent systematic review by Aranda-Aguirre et al, β-mannanase was the carbohydrase most widely used. Pigs can produce endogenous digestive proteases (e.g. pepsin, trypsin, chymotrypsin, carboxypeptidases) although a fraction of their dietary protein substrates end up undigested in the excreta, which presents a market for commercial protease preparations[124]. Nevertheless, proteases represent a much smaller proportion of total sales relative to carbohydrases and phytases and studies have often described inconsistent/inconclusive correlations between digestibility and performance effects[147]. They are often added in combination with carbohydrases although individual proteases are also on the market. Indeed, combined enzymes are commonly used in pig diets for all productive stages, and carbohydrase mixtures are often superior to the use of the individual enzymes[124]. Once microbial phytase entered the market in 1991, their inclusion in swine diets surpassed the carbohydrases[124]. The systematic review by Aranda-Aguirre found that in pig diets, phytases are the most supplemented enzymes at all productive stages of pig nutrition. During the weaning and growing stages, phytases and mannanases are most commonly used, whilst xylanases and, less frequently, proteases have been reported as being used throughout all stages[124]. The specific effects of the enzymes depend on the productive stage of the pig i.e. weaning, growth, finishing. For example, whilst carbohydrase inclusion in young pigs is an essential dietary intervention due to their intestinal incapacity and the negative impacts of high fibre levels, their use in sows is much less common[138]. Newly weaned pigs face particular challenges as a result of their immature digestive system's inability to digest complex carbohydrates, potentially impairing growth and health, and this is why many products on the market specifically target this growth stage of pigs[150]. A systematic review by Torres-Pitarch et al found that, in post-weaning pig diets, phytase supplementation presented the most consistent improvements in piglet growth, phosphorus digestibility and bone mineralisation, whilst supplementation with xylanase alone or in combination with β-glucanase produced inconsistent responses to growth and nutrient digestibility. When multi-enzyme complexes were used, results were more consistent when protease and/or mannanase were components of the complex[151]. The same authors, in a systematic review on grow-finisher pigs, reported that xylanase alone or in combination with β-glucanase did not improve feed efficiency, despite their widespread use in wheat and barley based diets. Mannanase was observed to increase feed efficiency with maize-based diets whilst multi-enzyme complexes worked best with maize-, wheat-, barley- and co-product-based diets. Therefore, it was determined that enzyme supplementation responses are influenced by the major cereal source in the diet[152]. Indeed, another important factor dictating the efficacy of enzyme supplementation is the substrate profile of the feed ingredients. Careful consideration must be given to the type and level of enzyme supplementation in relation to the type and level of substrates constituting the feed ingredients. Since the most abundant NSPs in wheat are arabinoxylans/xylans, wheat-based

diets would respond better to xylanase, whereas β-glucanase would be better suited to diets based on barley[153]. However, even within a particular feed ingredient, the content of the various NSPs varies significantly so expectations of efficacy may not always be met[154]. Indeed, it is widely reported that increasing the level of enzyme inclusion does not necessarily result in a linear improvement in nutrient utilization[155]. A recent study by Junior et al suggests a possible explanation for lack of performance enhancement in cases where enzyme supplementation improves nutrient digestibility or intestinal health. It was proposed that under the intense conditions presented by commercial farms, the energy released by enzyme action may not necessarily be translated into animal growth, instead being diverted to other processes[128].

1.4.3 Commercially Established Phytases

To date, no phytase class other than the HAPhys have seen their applicability to the animal feed industry realized[70,118]. Established commercial phytase products are derived from either a 3-phytase *(A. niger, P. funiculosum*) or a 6-phytase (*E. coli, P. lycii, C. braakii*, Buttiauxella, Yersinia)[156], with 3-phytases being of fungal origin and 6-phytases being bacterial. These are detailed in Table X.X. In general, the fungal and bacterial HAPhys exhibit pH and temperature optima in the range pH 2.5-5.0/55-60°C and pH 3.5-5.0/40-60°C, respectively[156]. *E. coli* phytases in particular demonstrate a single pH optimum between pH 3.5 and pH 4.5. Phytase from *P. lycii* on the other hand is optimum between pH 4.0-5.0[156]. Fungal phytases dominated the market in its early days, before in 1999, bacterial *E. coli* phytases were recognised to be more effective and were therefore the source of the second generation of phytases which were superior in practice[70].

1.4.3.1 Fungal Representative: PhyA

Setting the benchmark for modern day industrial phytases, the first phytase which saw commercialisation was the fungal 3-phytase, PhyA, from *Aspergillus niger* NRRL 3135, marketed by BASF under the proprietary name Natuphos® since 1991 (Table 1.5). Its proven efficacy as a feed additive arises from its high specific activity for phytic acid among other enzymatic properties. Five disulfide bridges and substantial glycosylation endow this 85 kDa monomer with thermostability, with optimum activity at 58°C. PhyA exhibits optimum activity at two pH values (2.5 and 5.0). *A. niger* also produces the 65 kDa PhyB phytase, and although this has not received commercial interest to date, together with PhyA forms the most extensively

characterised representatives of the HAPhy class[44]. PhyA and PhyB differ in the amino acid landscape around the catalytic site, resulting in different net charges, and this is the cause for the lack of activity of PhyB at pH 5.0, instead being active at the other optimum of PhyA, pH 2.5. They are both secretory proteins[44].

1.4.3.2 Bacterial Representative: AppA

AppA is the prototypical bacterial 6-phytase [EC 3.1.3.2] from *E. coli* which has been extensively and iteratively engineered since its introduction to the feed industry to produce one of the best-selling commercial feed phytases of all time – Quantum Blue (AB Vista)[158]. Exhibiting both acid phosphatase and phytase activity, AppA is a periplasmic phytase with high catalytic efficiency and stereospecificity for the phosphate at the 1D-6 position of the inositol backbone, whose catalytic mechanism was elucidated in 1992, five years after its discovery[159,160]. The wild-type AppA was marketed as Phyzyme® XP (Table X) and represented the first bacterial phytase to enter the industry[161]. Efforts then focussed on improving the thermal tolerance of these enzymes for animal feed applications, with the rapid growth of the industry sparking multiple competitors. To this end, native AppA underwent eight amino acid mutations, yielding a variant – Phy9X/NOV9X - with a melting temperature of 12°C above that of the parental enzyme[162]. This variant formed the basis of Quantum Blue®, being identical in sequence[163]. Several more engineered variants have been generated in subsequent years, focussed primarily on enhancing thermal and gastric stability properties. For example, in 2021 an additional non-consecutive disulfide bond was incorporated into AppA to generate a new variant, ApV1, having a total of five disulfide bonds. Although this variant demonstrated increased residual activity after heat treatment, whilst maintaining an identical biochemical profile to the wild-type enzyme, its performance did not surpass the benchmark of Quantum Blue®[163]. Nevertheless, Quantum Blue® does suffer from a marked deterioration in rate beyond liberation of the first two phosphate groups of phytate, which presents an as yet unconquered challenge. This drastically reduced hydrolysis as the reaction proceeds culminates in substantial amounts of lower phosphorylated InsP species, IP4 and IP3 in particular[164-166.] This is observed in other bacterial phytases from *E. coli* and *C. braakii* too, whilst phytases from *P. lycii* and *A. niger* only accumulate IP3[156]. The recent work by Hermann et al represents, for the first time, efforts targeted at improving hydrolysis of IP4 and IP3 by AppA. It was demonstrated that mutations in the active site to optimize the hydrolysis of IP4 and IP3 may be mutually exclusive, that is, variants with superior IP4 hydrolysis will more than likely be sub-optimal for IP3[167]. Very recent research has extended into attempts to engineer AppA to be displayed on the cell surface of *E. coli*, as a biomembrane-immobilised enzyme[168].

Product	**Year**	**Donor/Source Organism**	**Production Organism**	**Manufacturer**	**Positional specificity**	**pH**	**Temp optimum** °C
Natuphos®	1991	*Aspergillus niger var ficuum* (PhyA)	*Aspergillus niger*	BASF	3-phytase	2; 5-5.5	58
Allzyme® SSF	?	*Aspergillus niger*	*Aspergillus niger* (non-recombinant)	Alltech	3-phytase	6.0	-
Phyzyme®	2001			DuPont			
Phyzyme® XP	2003	*Escherichia coli* (AppA)	*Schizosaccharomyces pombe* (ATCC 5233)	Danisco/DuPont	6-phytase	4.5	55
Rovabio™ PHY	Pre 2007	*Penicillium funiculosum*	*Penicillium funiculosum* CBS 111 433	Adisseo	3-phytase		
Quantum™	2006-2008?	*Escherichia coli* (NOV9X)	*Pichia pastoris* sNOV9Xpp27	Syngenta	6-phytase	4.5	65?
Ronozyme® (NP)	2007?	*Peniophora lycii*	*Aspergillus oryzae*	Novozymes/DSM	6-phytase	4-4.5	50-55
Optiphos®	2008 (2006 US)	*Escherichia coli* (AppA2)	*Komagataella phaffii DSM 23036 (formerly known as Pichia pastoris)*	Huvepharma (originally Enzyvia)	6-phytase	3.4, 5.0	58
Finase® EC	2009?	*Escherichia coli* (AppA)	*Trichoderma reesei*	AB Vista	6-phytase	4.5	55
Finase P/L		*Aspergillus niger* (PhyB)	*Trichoderma reesei*	AB Vista	3-phytase	2.5	
Ronozyme® HiPhos	2011?	*Citrobacter braakii*	*Aspergillus oryzae* DSM 33699	Novozymes/DSM	6-phytase	4.0	50
Quantum® Blue	2012	*Escherichia coli*	*Trichoderma reesei* CBS 126897	AB Vista	6-phytase	3.5-5.0	-
Axtra® PHY	2013	*Buttiauxella spp.*	*Trichoderma reesei* SD-6528	Danisco/Dupont	6-phytase	3.5-4.5	60
Natuphos® E	2016	Hybrid of *Hafnia sp, Yersinia sp* and *Buttauxiella sp.*	*Aspergillus niger* (DSM 25770)	BASF	6-phytase	4.5	
Axtra® PHY Gold	2020	*???*	*Trichoderma reesei* CBS 146250	Danisco/Dupont	6-phytase	3.0	
Nutrase® P		*??*	*Komagataella phaffii* (CGMCC 7.19)	Nutrex	6-phytase		
Rovabio PhyPlus	2021	*????*	*Trichoderma reesei*	Adisseo	6-phytase		
HiPhorius™	2022	*Citrobacter braakii* ATCC 51113	*Aspergillus oryzae DSM* 33737	DSM	6-phytase		
VTR-phyatse		*??*	*Komagataella phaffii CGMCC 7.370*	Victory Enzymes GmbH	6-phytase		

Table 1.5 Commercial phytase products on the market and associated information. Information from multiple sources [70;129;169-187]

1.4.3.3. Functionality of Selected Commercial Phytases in Poultry – *in vitro* and *in vivo* studies

Commercial phytase preparations are supplemented according to their activity as reported by the manufacturer; however, the standard conditions under which activities are determined are pH 5.5; 37°C; 5 mmol/L sodium phytate, the pH of which is not reflective of the range of physiological conditions present in the digestive tract of the animal (see section 1.2.1)[70]. The activity of different phytases – of differing pH optimum, proteolytic stability and kinetic properties - vary with transit as conditions along the digestive tract change, with certain conditions suiting a particular phytase over others, for example. The activity *in vivo* is thus different from the standard phytase activity measurement and therefore prompted investigations into the relative performance of exogenous phytases in the digestive tract of the animal. There are, however, a multitude of factors which affect biological efficacy of supplemental phytases *in vivo* such that the most accurate - albeit cost- and time-intensive - picture is gained from direct feeding trials. Since *in vivo* performance results from a combination of all enzymatic properties including endogenous phosphatase activities, determination of enzymatic properties are merely guidelines for potential functionality in the digestive tract.

Menezes and co-workers used an *in vitro* model of poultry digestive tract to compare seven commercial phytases: Quantum™, Quantum® Blue, Phyzyme®XP, Axtra®PHY, Ronozyme® HiPhos, Ronozyme® NP and Natuphos®[16]. Zeller and co-workers in conjunction with AB Vista analysed the digesta of different segments of poultry digestive tract of birds on diets without and with supplemental phytases – Finase®, Quantum™ and Quantum® Blue[188]. This enabled characterisation of the *in vivo* IP6 degradation patterns and comparison to known *in vitro* properties. This latter study, unlike others, also analysed the positional isomers of IP6 hydrolysis. Regardless of exogenous phytase inclusion, endogenous phytate-degrading activity occurs through both gut microbiota and intestinal mucosa, necessitating determination of the characteristics of IP6 hydrolysis at the basal level in order to more accurately attribute to/differentiate from activities of commercial phytases. Both studies included Quantum™ and Quantum® Blue but these two studies were distinctly different thus not comparable to one another. Since the study by Menezes and co was based on an *in vitro* simulation of the digestive tract, the behaviour of the phytases studied does not accurately reflect their performance *in vivo*, whereas Zeller and co gained a 'snapshot' of the composition of IP hydrolytic products in the different regions of the digestive tract at a given point in time. Nevertheless, endogenous phytase activities inevitably complicated the delineation of the hydrolytic pathway of the phytase tested, leaving much yet to categorically determine.

Since pH optima and activity profiles of phytases indicate likelihood of catalytic activity in the various gastrointestinal compartments, Menezes and co compared activities of the seven phytase subjects at pH 3.0 (representative of pH in the proventriculus and gizzard) and pH 7.0 (representative of pH in the small intestine) as a percentage relative to 100% activity at pH 5.5. As well as not being reflective of *in vivo* conditions, reported properties of the commercial products are directly incomparable since different assay conditions are used by the manufacturer, which is the rationale for the aforementioned study. The *in vitro*-determined acidic pH optima and lack of activity above pH 7.0 of all of the phytases is well-aligned to conditions of the anterior segments of the digestive tract (crop, proventriculus, and gizzard). This indicates efficient phytate dephosphorylation in the GI tract whilst the intestinal pH conditions render the phytase less/(in)active, with the crop being reported in the literature as the primary site of exogenous phytase action. Despite this theoretical similarity, the pH activity profiles at pH 3.0 of the seven phytases exhibited significant variance, with the fungal derived Ronozyme® NP having the lowest relative activity and the bacterial Axtra®PHY having the highest. The three *E. coli* origin phytases (Quantum™, Quantum® Blue and Phyzyme®XP) were similar in relative activity to that at pH 5.5, whereas bacterial Ronozyme® HiPhos (and Axtra®PHY) were significantly higher and fungal Natuphos (and Ronozme®NP) significantly lower. In contrast, at pH 7.0 all phytases had less than 8% relative activity. Zeller and co found that in the crop, Finase® had a much greater effect on IP6 hydrolysis (64%) compared to the basal diet/control than did Quantum™ (31%) and Quantum® Blue (44%), but in the duodenum/jejunum and lower ileum, exogenous phytases had no significant effect. *E. coli* phytases are more active than fungal phytases at pH 3.0 – the pH of the proventriculus/gizzard – which is thought to also explain the lack of difference in activity between exogenous phytases in the duodenum/jejunum compared to in the crop. Both studies concluded that the primary site of exogenous phytase action is the proventriculus/gizzard, but significant activity occurs also in the crop, with 80-90% of dephosphorylation taking place within these three components. However, without supplemental phytase, IP6 hydrolysis in the crop is very low (9%). Hydrolysis of IP6 and net phosphorus absorption was found by Zeller and co to be more effective with Quantum™/Quantum® Blue when in the lower ileum and caeca, compared to Finase® and the non-supplemented control. For example, all three phytases equally increased net phosphorus absorption compared to the non-supplemented diet up to the duodenum/jejunum, but then up to the lower ileum Finase® (56%) was comparative to the basal (57%) whereas Quantum™ and Quantum® Blue were more efficient, with 60% and 64% absorption, respectively. Regarding the IPx isomers resulting from breakdown of IP6, Zeller and co were the first to demonstrate that in the crop of chickens the degradation patterns with these three commercial phytases are very similar to that determined *in vitro.* For example, the predominant IP5 isomer with Finase® was Ins(1,2,4,5,6)P5, whereas with the Quantum

phytases, it was Ins(1,2,3,4,5)P5, which conforms with the knowledge that fungal phytases initiate dephosphorylation at C3 of inositol and bacterial phytases at C6. They also reported relatively high concentrations of Ins(1,2,5,6)P4 and IP3 in the crop, with IP4 concentrations being much higher for Quantum™ (57%)/Quantum® Blue (74%), compared to 14% for Finase®; and IP3 concentrations much higher for Finase® (69%) compared to Quantum™ (15%) and Quantum® Blue (13%), the former of which generated only one IP4 isomer - Ins(1,2,5,6)P4. This accumulation of IP4 by the *E. coli* phytases corresponds to *in vitro* results and in these, IP4 was slowly hydrolyses further to IP3 by endogenous phosphatases. In the proventriculus/gizzard, Ins(1,2,5,6)P4 was significantly higher for the *E. coli* phytases as in the crop, however, unlike in the crop, IP3 was also higher than for Finase®, which had a higher proportion of IP5 compared to IP4/IP3. In the duodenum and jejunum of the small intestine, supplemental phytases - unlike samples of the non-supplemented treatment - generated the Ins(1,2,5,6)P4 isomer, although the Ins(1,2,3,4)P4 isomer was the predominant IP4 isomer in this region of the tract. The predominant IP5 isomers for the *E. coli* phytases and Finase® were the same as in the crop and all had a lower proportion of Ins(1,2,3,4,6)P5 IP5 isomer compared with the non-supplemented control. IP3 was not seen neither with nor without exogenous phytases in this gut segment. In the lower ileum the concentration difference in Ins(1,2,3,4,5)P5 between the fungal and bacterial phytases lost significance. The caeca was the segment which saw less distinct differences in the isomeric pattern; for all phytases the predominant IP5 isomer was Ins(1,2,3,4,5)P5, although Finase® still had a higher proportion of Ins(1,2,4,5,6) compared to Quantum™ and Quantum® Blue. The authors point out that there is a possibility that concentrations of IPx isomers anterior to the caeca may be affected by the retrograde movement of digesta and microbiota.

Returning to the other study, Menezes and co found that all phytases with the exception of the fungal Ronozyme® NP had substantial stability at the pH of the proventriculus/gizzard, and that bacterial phytases had greater ability to withstand proteolytic cleavage by pepsin, unlike the fungal phytases which were susceptible. This is in agreement with the literature, in which *E. coli* phytases are reported to be more resistant to pepsin and pancreatin and have higher activity at pH 3.0 than Aspergillus phytases. Based on kinetic constants, all the phytases had better enzymatic characteristics at pH 5.0 compared to pH 3.0. They also identified that the bacterial phytases were more affected by buffer volume in the *in vitro* simulation hence potentially by viscosity *in vivo*. Overall, they reported significant variation in the relative performances of the seven phytases, but that there were too many influencing parameters to be able to attribute these differences to thus their model cannot be used in ranking phytases by bio-efficacy but rather to determine their suitability as feed additives. The major insight from Zeller and co's study is the *in vivo* evidence for IP6 degradation pathways previously reported from *in vitro* analyses[16,188].

1.5 Reports of Complete Enzymatic IP6 Degradation

It is well documented in the literature that the stability of the axially orientated phosphate moiety at C2 of the *myo*-inositol core of IP6 prevents both its detachment and consequently complete degradation of IP6 to free *myo*-inositol and orthophosphate. There are very few reports in the literature of phytases or phytate-degrading phosphatases which are able to cleave the C2 phosphate, and of these none initiate dephosphorylation at the C2 position (or is unknown) and thus to the author's knowledge, no phytase has ever been classified as a 2-phytase. Inositol monophosphate is almost always the final degradation product by reported phytases other than those of the βPPhy class, with the sole remaining phosphate group being the axial C2 phosphate, indicating preferential hydrolysis of equatorial groups[189]. Phytases classified as βPPhy on the other hand generate inositol triphosphate as the final product. A phytase of the βPPhy class was once thought to act via a degradation pathway resulting in Ins(1,3,5)P3 as one of the final IP3 products, nevertheless, the evidence otherwise served to further establish the equatorial preference of phytases[82]. There are just three reports in the literature to date which report the dephosphorylation of all six phosphate groups and liberation of *myo*-inositol: two from yeasts and one from bacteria (see **1.5.1**). The livestock production industries are continually seeking more proficient phytases to generate more profit and outperform their commercial competitors; and there's still room for improvement with regard to the efficacy of current feed phytases. It is thus well-justified to strive towards the production of a phytase – by enzyme tailoring or seeking novel phytase-producing isolates – which achieves the comprehensive hydrolysis of IP6 to release *myo*-inositol and inorganic phosphate, in the most rapid, efficient and economically competitive manner.

1.5.1 *Klebsiella pneumoniae* Phy9-3B phytase

Escobin-Mopera and co-workers in 2012 reported the isolation and characterisation of a phytate-degrading enzyme produced by *Klebsiella pneumoniae* 9-3B, a soil isolate from a Japanese spinach field[189]. The authors stated that to their knowledge, this was the first bacterial phytase to be shown to have the capability to completely hydrolyse phytate, releasing free *myo*-inositol. Of four strains which exhibited high phytase activity, this isolate was chosen due to it having the highest intracellular phytase activity. The culture medium employed featured IP6 as the sole carbon and phosphorus source, which stimulated phytase production in the absence of a readily utilisable energy source, in accordance with the requirement for the hydrolysis products of phytate for growth. Detection of *myo*-inositol and phytase activity in the assay mixture after enzyme treatment established that isolate 9-3B was able to release phosphate and *myo*-inositol from the phytate source in the medium and utilise these for growth, with the

latter used as an energy source. The authors hypothesise that the dependence of the *Klebsiella pneumoniae* isolate on degradation of IP6 for survival under the culturing conditions employed is the reason for its ability to completely hydrolyse IP6, with production of phytase as a survival mechanism. Results of characterisation studies indicate that this phytase is comparable in biochemical properties to the HAP phytases (section 1.3.2) – and in particular to other bacterial phytases of Klebsiella spp. - with pH and temperature optima of pH 4.0 and 50°C, respectively, and a broad pH stability range of pH 2.0 – 7.0. However, the activity dropped sharply by 60°C with zero activity at 70°C. Similar to HAPhys, the phytase was inhibited by Zn^{2+} and Fe^{2+} and was estimated to be a 45 kDa monomeric intracellular protein. Despite exhibiting broad substrate specificity – which is usually synonymous with a low specific activity – this phytase demonstrated a high specific activity for phytate, which is the highest reported among phytases of Klebsiella spp. Further evidence for complete hydrolysis was the exact molar ratio of 1:6 (phytate: phosphate) detected during measurement of liberated phosphate, and the hydrolysis of *myo*-insoitol-2-monophosphate as a substrate at a 1:1 ratio. Lower phosphorylated *myo*-inositol phosphate intermediates were also detected by HPLC prior to complete hydrolysis of IP6. Many properties of this phytase render it suitable for application as a supplemental feed phytase, in particular its broad substrate specificity and pH activity profile, the latter of which is aligned to the conditions in the stomach and small intestine of monogastric species, enabling it to elicit catalytic activity in those regions[189].

The authors published further work on this phytase - now named Phy9-3B – in 2014, in which the gene sequence enabled the accurate determination of the molecular weight (43.4 kDa) and revealed similarity to a known Klebsiella HAPhy phytase, PhyK, having only nine amino acid differences, enabling its classification as such[190]. The identification of these differences being located on the outer surface prompted investigations into dimerisation as a novel approach to increase activity and specificity, since these differences between residues of Phy9-3B and PhyK were postulated to enable Phy9-3B to homodimerise. This in turn causes the biochemical differences between the two phytases, such as ten-fold greater specific activity with Phy9-3B and its distinguishing ability to cleave the axial C2 phosphate of inositol. They hint to future mutagenesis and crystallography analyses in order to identify key residues which the unique features are attributed to[190].

1.5.2 *Debaryomyces castellii* CBS 2923 PhytDc phytase

The phytase from *Debaryomyces castellii* CBS 2923, PhytDc, is one of just two yeast species capable of complete hydrolysis of phytic acid. It was first reported in 2008 before the publishing of its structure a year later[191-194]. The authors state this is the first yeast phytase described to date as capable of achieving full phytic acid hydrolysis whereby the complete hydrolytic sequence has been elucidated. It is classified as a 3-phytase since the first product of hydrolysis is Ins(1,2,4,5,6)P5. The hydrolysis sequence was established by ^{1}H NMR monitoring and is as follows: Ins(1,2,4,5,6)P5 – Ins(1,2,5,6)P4 – Ins(1,2,6)P3 – Ins(1,2)P2 - Ins(1 or 2)P1 – inositol (Fig 1.15). IP3 was found to accumulate before further hydrolysis, indicating weaker affinity of the phytase for IP3 compared to IP6/IP5/IP4. The final two ester bonds were simultaneously hydrolysed to yield inositol. As such, data from HPIC and NMR analyses confirmed the capacity of PhytDc to cleave the phosphate groups of inositol monophosphates i.e. Ins(1)P1 and Ins(2)P1 since all six phosphate groups are evidently hydrolysed. For example, the amount of free phosphate measured after 5 hours was 100% of the potential phosphate available from phytate. Both the 74 kDa glycosylated form of PhytDc and the 53 kDa deglycosylated form are active; in this study the phytase was deglycosylated post-purification. The biochemical profile is similar to other fungal and yeast phytases, exhibiting thermostability up to 66^{o}C with 55-60^{o}C the optimum and pH optimum 4.0 – 4.5 within the range 2.5 – 6.5. Following elucidation of the crystal structure and sequence analyses, it was classified as a HAPhy, having the characteristic sequence motif RHGXRXG and structural similarity to other HAPhys – a large α/β domain and an additional smaller α domain – particularly the acid phosphatase from *A. niger*. Structural analyses also revealed its tetrameric biological form. It displayed a wide activity spectrum with preferential hydrolysis of five phosphorylated compounds including sodium phytate, in addition to a range of others. It demonstrated preferential hydrolysis of the inositol monophosphate isomer Ins(2)P1 (84% relative to sodium phytate) compared to Ins(1)P1 (58%)[191-194].

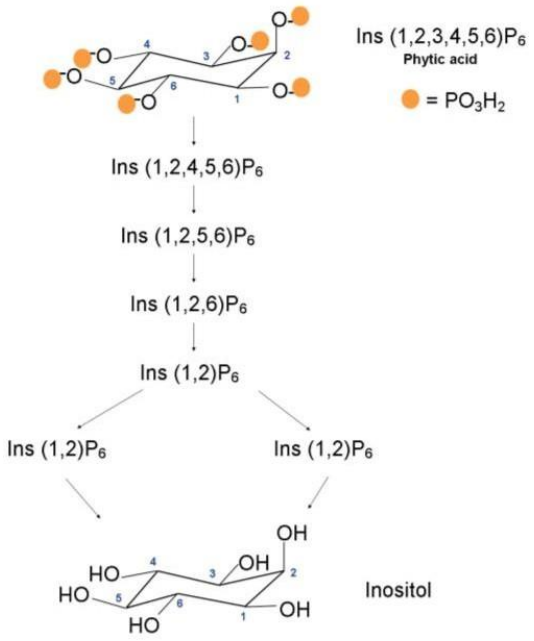


Figure 1.15. Pathway of IP6 degradation by PhytDc determined by Ragon et al. Figure from reference [193].

1.5.3 *Schwanniomyces castellii* CBS 2863 phytase

Schwanniomyces castelli CBS 2863 (AKA *Debaryomyces occidentalis*?) is the other of the two phytase-producing yeast species capable of stepwise removal of all six phosphate groups of IP6. Reported in 1992 as a tetrameric glycoprotein, it was described as having ten-fold greater affinity for phytate than the previously discussed phytase from *Debaryomyces castelli* CBS 2923 and, like Phy9-3B from *Klebsiella pneumoniae*, phytate is its preferential substrate despite broad specificity, making it ideal for use as a feed additive[195,196]. As with PhytDc, IP3 degradation was slower than for IP4 through IP6. Whilst characteristic of acid phosphatases, the pH optimum (pH 4.4) is somewhat lower than those commonly reported for bacterial or fungal HAP phytases (around 5.5) and is much more similar to *Saccharomyces cerevisiae* phytase (4.6); it is active between pH 3.4 -6.0. It has an unusual preference for high temperatures, with optimal activity at 77°C and is much larger than other microbial phytases in as much that the deglycosylated enzyme – which retains activity - is tetrameric in structure, consisting of one subunit (125,000 kDa) and three identical subunits (70,000 kDa). Inhibition studies implicated the involvement of SH groups in the catalytic site of the enzyme[195].

1.6 Research Aims and Objectives

The work constituting this thesis overall sought to develop a phytase with a superior set of characteristics to the current gold standard on the market, with particular focus on extending the limits of phytate dephosphorylation. Two core approaches were taken: extending the previous fundamental work on MINPPs with the objective of bioengineering variants with enhanced phytate hydrolysis properties; and exploring novel phytases for insights into their structure-function relationship which could enable their structure-informed engineering.

Within the MINPP sphere, efforts were focussed on an improved hydrolytic capacity for IP4, to overcome the limitation that commercial phytases based on AppA present. The B-pocket of *Bt*MINPP was elected as the target for generating an array of phytase variants by mutagenesis. The activity of these variants against both IP6 and IP4 were evaluated to identify drivers of an increased rate of hydrolysis.

The approach based around novel phytases entailed investigating the activity, biochemical and structural characteristics of two structurally metallo-β-lactamase-presenting enzymes proposed to constitute a novel class of bacterial phytase - MβLp01 and MβLp02.

The work described in the final chapter represents efforts to determine the structure of an enzyme critical to the biosynthesis of phytate in plants – ITPK1. Although disparate to the primary aim of this thesis, this work provided ample opportunity to develop crystallographic insight and expertise and was important for collaborative work with a closely linked research group.

"When you've understood this scripture, throw it away. If you can't understand this scripture, throw it away. I insist on your freedom."

CHAPTER TWO

General Experimental Protocols

This chapter provides an overview of the routine protocols used in recombinant protein expression, purification and crystallisation along with a brief background to a method where appropriate.

2.1. Recombinant Protein Expression in *E. coli*

2.1.1 Host Bacterial Strains

Escherichia coli is the most popular, well-established choice of host for recombinant protein expression due to the ease of genetic manipulation and fast growth afforded by this species. Host strains can be divided into those for DNA propagation and those for protein expression. The requirements for overexpression of a foreign protein in *E. coli* are significantly different to those of the propagation and maintenance of plasmid integrity and as such, strains have been engineered for specific applications. Whereas the main purpose of DNA propagation strains is the preservation of heterologous DNA with a low mutation risk, protein expression strains are designed to accommodate increased protein levels and circumnavigate the many perils that the overexpression of a non-native protein can pose to the host cell. As such, it is not advised to perform plasmid extraction from an expression strain since many of these strains do not have the genotypic traits to maintain chromosomal stability so the plasmid may have undergone mutations. Equally, it is ill-advised to express a heterologous protein in a DNA propagation strain, as these do not harbour the necessary engineered genetic mutations to withstand the pressures that protein overexpression put on the host cell. Most commercial *E coli* strains originate from either the 'B' or 'K-12' strains, the latter of which was initially isolated in 1920 as a human pathogen. The common lab strain DH5alpha is a derivative of the K-12 strain whereas the B strain gave rise to today's routinely used BL21 line.

2.1.1.1 DNA propagation strains

Increased plasmid yield and quality is the main feature of DNA propagation strains which are commonly used for cloning applications. Plasmid yield/quality is improved by a knock-out mutation in Endonuclease 1 (endA/endA1) which eliminates non-specific cleavage of dsDNA.

Increased plasmid stability and reduced recombination are achieved by mutations in several genes: recBCD mutation abolishes exonuclease V activity; mutations in recA, recA1 or recA13 impair a DNA-dependent ATPase which is essential for recombination and general DNA repair. Inactivation of the pathway responsible for cleavage of methylated cytosine DNA facilitates the uptake of foreign (i.e. methylated) DNA - attributed to the genes mcrA and mcrBC – thereby enabling cloning. Problems with methylation-sensitive restriction enzymes are mitigated by mutations in dam/dcm, which prevents adenine and cytosine methylation at specific recognition sequences, and $hsdS_B$ ($r_B^-m_B^-$) which inactivates the native restriction/methylation system. This enables cloning involving methylation-sensitive restriction enzymes.

2.1.1.2 Protein expression strains

Mutations which result in reduced proteolysis of expressed heterologous protein are an integral feature of all expression strains. The key genes responsible are ompT (mutation) and lon (deletion), the latter being conferred by all B strains. Resultant strains are deficient in outer membrane protease VII and Lon ATPase-dependent protease, respectively. There is a plethora of commercial protein expression strains available, each engineered for specific applications. For example, the Rosetta lines (Table 2.1) are tailored to the expression of eukaryotic proteins by the provision of a set of codons not commonly supplied by *E. coli.* In addition to the extra codons, the pLysS derivatives of the Rosetta range carry a plasmid containing T7 lysozyme which further suppresses basal expression of T7 RNA polymerase (RNAP) prior to induction – good for potentially toxic proteins. Many host strains use the T7 expression system - as lysogens of bacteriophage DE3, they carry a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter/lac repressor, which is inducible by the addition of the lactose analog IPTG. Induction of T7 RNAP expression subsequently drives transcription of the target gene. However, the lac promoter often gives rise to a background level of expression in the absence of inducer, which poses a challenge for toxic protein products since they may reduce the rate of growth of the culture. Some expression strains confer tighter control of basal expression by the incorporation of the T7 lysozyme gene-containing pLys plasmid, such as Rosetta (DE3) pLysS and BL21 (DE3) pLysE. pLysE offers stronger repression of the promoter pre-induction than pLysS. Included in Table 2.1 are the common *E. coli* expression strains used in the work in this thesis, along with their key features.

Expression Strains				
Strain	**Antibiotic resistance**	**Primary Use**	**Features**	**Genotype**
SHuffle T7 Express	Spectinomycin; Streptomycin (low)	Enhanced capacity to correctly fold **proteins with multiple disulfide bonds** in the cytoplasm.	Deficient in proteases Lon and OmpT Resistant to phage T1 (fhuA2) Expresses disulfide bond isomerase DsbC	fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10 --TetS) endA1 Δgor Δ(mcrC-mrr)114::IS10
Rosetta (DE3)	Chloramphenicol	BL21 derivates designed to enhance the **expression of eukaryotic protein**s rarely used in *E. coli*.	supply tRNAs for 6 rare codons (AGG, AGA, AUA, CUA, CCC, GGA) on a compatible chloramphenicol-resistant plasmid (pRARE)	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR)
Rosetta 2 (DE3)	Chloramphenicol	BL21 derivates designed to enhance the **expression of eukaryotic proteins** rarely used in *E. coli.*	supply tRNAs for 7 rare codons (the same 6 as carried by pRARE plus CGG) on a compatible chloramphenicol-resistant plasmid (pRARE2)	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE2 (CamR)
Rosetta (DE3)pLysS	Chloramphenicol	BL21 derivates designed to enhance the **expression of eukaryotic proteins** rarely used in *E. coli* along with pLysS for tighter expression control.	supply tRNAs for 6 rare codons (AGG, AGA, AUA, CUA, CCC, GGA) on the T7 lysozyme gene-containing plasmid (pLysSRARE). T7 lysozyme further suppresses basal expression of T7 RNA polymerase prior to induction.	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysSRARE (CamR)
Rosetta 2(DE3)pLysS	Chloramphenicol	BL21 derivates designed to enhance the **expression of eukaryotic proteins** rarely used in *E. coli* along with pLysS for tighter expression control.	supply tRNAs for 7 rare codons (the same 6 as carried by pRARE plus CGG) on the T7 lysozyme gene-containing plasmid (pLysSRARE2). T7 lysozyme further suppresses basal expression of T7 RNA polymerase prior to induction.	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysSRARE2 (CamR)
BL21 (DE3)	None	Transformation and routine protein expression General purpose T7 expression	Deficient in proteases Lon and OmpT Expression from T7 promoter Basic IPTG-inducible strain containing T7 RNAP (DE3)	F- ompT lon hsdSB (rB- mB-) gal dcm (DE3)
Plasmid propagation/storage and cloning strains				
DH5a	None	General cloning and storage of common plasmids		F- Phi80lacZDeltaM15 Delta(lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 tonA
TOP10	Streptomycin	General cloning and storage of plasmids		F- mcrA Delta(mrr-hsdRMS-mcrBC) Phi80lacZM15 Delta-lacX74 recA1 araD139 Delta(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG

Table 2.1 Overview of the common *E. coli* expression and DNA storage strains used in this work

2.1.2 Expression Systems and Plasmids

A wide range of expression systems based on different promoters are available for recombinant protein expression. Both the pET and pBAD expression systems and their corresponding vectors were used throughout this work.

Developed in the late 80's/early 90's, the pET expression system is one of the most popular systems for recombinant protein expression in *E. coli* and has become a workhorse of molecular biology. It comprises both a T7 or T7lac expression vector and a corresponding host bacterial strain genetically engineered to contain a lysogenised lambda DE3 phage fragment encoding the T7 RNA polymerase (typically *E. coli* BL21 (DE3)). pET expression vectors are derivatives of the pBR322 backbone and enable the induction, by IPTG, of the chromosomally integrated, lacUV5 promoter-driven bacteriophage T7 RNA polymerase. Also carried by the lambda DE3 lysogen, the product of lacI – the lac repressor – regulates the promoter by being bound to the lac operator region between the lacUV5 promoter and the T7 RNA polymerase gene, thereby blocking its transcription. Upon induction the highly active T7 polymerase out-competes transcription by the host RNA polymerase such that the majority of the cell's resources are dedicated to target gene expression and as a result the target protein constitutes the majority of the cellular protein expressed a few hours post-induction. Since the bacteriophage-encoded RNA polymerase is highly specific to its cognate T7 promoter sequences, the host cell RNA polymerase is unable to recognise the T7 promoter thereby rendering the target genes transcriptionally silent until induction – yet once induced, the sole gene transcribed. This lack of transcription pre-induction, is beneficial if the target gene is cytotoxic to the host. However, despite these features designed to be robust, in practice the promoter is known to be 'leaky' giving rise to undesirable basal expression, particularly in DE3 strains. This is because the lac repressor and the lac operator region are in equilibrium such that the operator site is not occupied 100% of the time therefore there is always a small amount of T7 RNA polymerase present. This poses the risks of reduced growth rates, cell death or plasmid instability if a gene encoding a toxic protein is cloned downstream of the T7 promoter. Where there are concerns over the toxicity of a recombinant protein, expression strains can be used which are engineered to suppress basal expression. These strains co-express the T7 RNA polymerase inhibitor, T7 lysozyme, from the plasmids pLysS, pLysE or pLysY. In the case of the latter, the T7 lysozyme is a variant which has the benefit of lacking lysozyme activity against the *E. coli* host cell wall. The key difference between pLysS and pLysE is the level of T7 lysozyme activity: pLysE offers a much higher level relative to pLysS. However, as a result of this, there are some disadvantages such as the potential for a significant decrease in host cell growth, cell lysis, reduced expression level of recombinant protein and a greater lag between addition of inducer and expression of

gene. There are just over one hundred pET vectors in the series, differing in, for example, the stringency of suppression of basal expression levels.

Expression from pBAD vectors is more tightly controlled than from pET vectors. The pBAD expression system is based on the prokaryotic araBAD operon which regulates the arabinose metabolic pathway in *E. coli* and is under the control of the araBAD promoter. The operon contains the araC gene for stringent control of the L-arabinose-regulated promoter which facilitates dose dependent regulation by inducer. In the absence of L-arabinose, transcription is prevented since AraC blocks the promoter region, but it undergoes a conformational change in the presence of L-arabinose such that transcription is activated. The araBAD promoter-based expression plasmids, which drive expression of the target gene in response to L-arabinose, are designed to enable precise modulation of the levels of target protein expression whilst maintaining negligible basal expression. Addition of glucose to the growth media further represses basal expression levels by lowering cAMP levels to decrease transcriptional activation and, regardless of L-arabinose levels, expression of genes under the control of the araBAD promoter will be robustly inhibited in the presence of glucose. This is because high cAMP levels are required in order to facilitate promoter activity by the association of a cAMP-CRP complex with the pBAD promoter. The main advantage of this system, thanks to the titratability of induction, is the facilitation of problematic protein expression such as those with toxicity or insolubility concerns. The *E. coli* expression host for pBAD vectors should be TOP10 since this strain lacks the genes for L-arabinose catabolism, unlike the wild-type. However if tight suppression of basal expression is necessary, *E. coli* LMG194 should be used as growth in minimal media better allows repression of araBAD by glucose. At the time of writing there are 9 pBAD vectors in the series, of which one (pBAD202/D-TOPO) was used throughout the work in Chapter 4. Table 2.2 lists the vectors used during the recombinant expression procedures in this thesis.

Vector	Expression System	Antibiotic resistance	Tags/additional features	Cleavage	Size (bp)
pET15b	T7	Ampicillin (Amp^R)	His		5708
pDEST17	T7	Ampicillin (Amp^R)	His		6354
pOPINF	T7				
pOPINB	T7				
pBAD202/D-TOPO	pBAD	Kanamycin (Kan^R)	His-patch thioredoxin	EK	4448

Table 2.2 Vectors used throughout the work in this thesis.

2.2 Cloning

2.2.1 (Directional) TOPO Cloning

TOPO based cloning negates the need to use restriction enzymes or DNA ligases by exploiting the reaction of the type I DNA topoisomerase from *Vaccinia virus* which functions both as a restriction endonuclease and a ligase by cleaving and resealing supercoiled DNA ends to release supercoils and therefore facilitate replication. This particular topoisomerase catalyses a reversible site-specific DNA strand cleavage and re-ligation by the formation of a transient 3'-phosphotyrosyl linkage between the cleaved DNA strand and a tyrosyl residue (Tyr-274) of the enzyme. Topoisomerase I recognises the pentameric motif 5'-(C or T)CCTT in duplex DNA and cleaves the phosphodiester backbone of one strand after this site. Topoisomerase is released by the attack on the formed phospho-tyrosyl bond via the 5' hydroxyl of the cleaved strand. There are three types of TOPO cloning: the bidirectional TA and blunt TOPO cloning, in which inserts are able to ligate in either direction, and directional TOPO cloning, which is the method employed in this work. With the directional TOPO cloning system, blunt-ended PCR products are cloned by the addition of four bases (CACC) to the forward primer which subsequently anneal to the GTGG overhang in the vector (Figure 2.1). Inserts are cloned successfully in the correct orientation in at least 90% of cases. Unlike most vectors, commercial TOPO cloning vectors are linearised with topoisomerase attached at both ends.

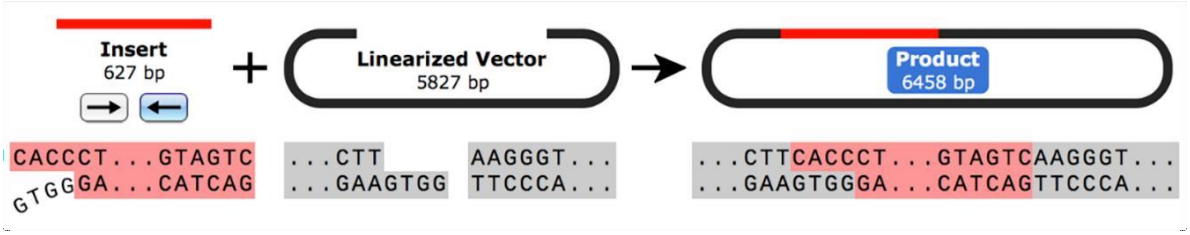


Figure 2.1 Overview of directional TOPO cloning.

2.3 DNA Agarose Gel Electrophoresis

Electrophoresis is a technique which uses electrical current to separate DNA, RNA and proteins based on the physical properties of size and charge. Agarose gel electrophoresis separates nucleic acid fragments based on their size, by drawing molecules through a gel mesh of tiny pores. When an electrical current is applied, the inherently negatively charged nucleic acid molecules migrate through the gel pores towards the positive electrode. Smaller fragments

migrate faster and hence further than larger fragments since they are less retarded by the size limitations of the apertures. The resultant bands are then visualized under UV light.

The percentage of agarose – a component of agar which generates a 3D gel matrix of helical agarose molecules in supercoiled bundles creating channels - in the gel dictates the pore sizes and therefore the size and rate of the molecules which can enter and pass through, which in turn controls the resolution. The higher the percentage of agarose, the smaller the pores in the gel matrix and vice versa. As such, a higher percentage agarose resolves smaller fragments better whilst larger fragments are better suited to a lower agarose percent. A 0.7-1% agarose gel is the typical agarose percent range for standard DNA separations, which differentiates fragments in the 0.2-10 kb range[197,198].

Since DNA is not inherently visible, an intercalating dye is used which binds to the DNA and fluoresces under UV light, enabling visualization of the DNA fragments. The dye is incorporated during the casting of the liquid gel, before it solidifies. Ethidium bromide has been routinely used since the 1960's[199] however, recent concerns regarding its safety due to its mutagenic/cytotoxic nature have led to many laboratories phasing out its use and instead using safer alternatives such as Gel Red[200], SBYR™ Green[201] and SYBR™ Safe[202]. The concentration of ethidium bromide in a gel is typically 0.2-0.5ug/mL and this results in a 20-fold increase in brightness upon DNA binding[203]. With regard to the work detailed in this thesis, ethidium bromide was used for the majority of DNA agarose gels (all except those in Chapter 3), before a transition to Gel Red. Gel Red is also visualized using UV light and can be used as either an in- gel or post gel stain. The dye used will be detailed in each chapter's specific methods section. Nucleic acid samples are mixed with a loading dye which contains a tracking dye for visualization of the extent of migration and glycerol to introduce density to the sample in order to prevent sample loss to the buffer. Whilst the gel is submerged in a running buffer, samples are loaded into wells at one end of the gel. The running buffer is either TAE (tris-acetic acid- EDTA) or TBE (tris-borate-EDTA) which serve to maintain the appropriate ion concentration and pH during the electrophoresis run – which is important for maintaining the net charge of the nucleic acid samples - and to facilitate the flow of electric current through the gel. Both contain EDTA to prevent nucleases potentially degrading the samples. When considering which buffer is most suitable for a particular application, it should be noted that borate is an enzyme inhibitor so TBE may not be compatible with downstream cloning involving ligases[204]. TBE is more suitable for longer runs for high resolution of smaller fragments since it has greater buffering capacity and is more resistant to overheating, whereas TAE is better suited for the separation of larger fragments (>2kb) and/or subsequent cloning. The voltage is typically set to between 80-150V, with a balance being sought between resolution and time. That being, resolution can be improved by lowering the voltage but at the expense of time. DNA bands are usually sufficiently

separated once the tracking dye is around 70-80% down the gel. Samples are run alongside a DNA ladder composed of DNA fragments of known sizes such that the sizes of the sample fragments can be extrapolated, since the relationship between size and migration distance is non-linear.

Specific parameters will be detailed in each chapter's methods section, however some conditions were universal. Samples were mixed in the appropriate ratio with either GelPilot 5X (Qiagen) or Purple gel loading dye 6X (NEB) before loading 5-10 μL in a 0.8% or 1% agarose gel containing 0.5ug/mL EtBr. Gels were run in TAE buffer, at 120V for 45-60 min alongside a 1kb plus DNA ladder (NEB).

2.4 SDS-PAGE (Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis)

SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis is essentially analogous to agarose gel electrophoresis, but for the separation of proteins. Both involve the separation of molecules based on differential rates of migration through a porous gel matrix under the influence of an applied electrical field. In SDS-PAGE, the gel is composed of polyacrylamide, which is chemically inert, and as with agarose gels, the percentage of polyacrylamide dictates the pore sizes. For larger proteins, lower percent gels are used and vice versa. A rough guide is 7% for 50-500kDa proteins, 10% for 20-300kDa proteins, 12% for 10-200kDa proteins and 15% for 3-100kDa proteins. For a broader separation range, gradient gels can be used which have layers of increasing polyacrylamide concentration. Unlike DNA which is inherently negatively charged, the net charge of a protein is determined by its amino acid composition. To make the protein migration rate proportional to molecular weight only, the influence of charge and molecular radius must be eliminated and this is achieved by denaturing the protein and masking the intrinsic net charge. The detergent SDS is added to the protein sample via the sample buffer and boiled in the presence of a reducing agent. As well as unfolding the protein, it coats the now linear molecule with a net charge which is mostly uniform – around 1.4g SDS binds to 1g protein[205] – such that charge becomes approximately proportional to length and thereby molecular weight. As such, SDS-coated proteins have the same charge to mass ratio which eliminates differential migration as a result of charge. SDS-PAGE incorporates a discontinuous buffer system which involves a stacking gel and different pHs – the stacking gel at pH 6.8, the running gel at pH 8.8 and the electrode buffer at pH 8.3. The stacking gel, composed of a low percentage of agarose, sits above the running gel and uses

the different pHs to sandwich proteins in a sample between glycine and Cl^-, so that they all start at the interface of the stacking and running gel at the same time. Once the electrical current is applied and the proteins enter the running gel, the pH change of the running buffer causes glycine to become negatively charged and accelerate ahead of the proteins. Differential rates of migration are exaggerated by the high friction of the gel matrix such that larger proteins migrate slower and therefore less far through the gel. Proteins are thus separated by size from large to small and after staining to reveal the bands, comparison to a protein molecular weight marker facilitates estimation of size of a protein in a sample. Several different protein staining systems are available, with those based on Coomassie blue the most commonly used[206,207].

For the work in this thesis, pre-cast Bolt™ Bis-Tris plus, 4-12% gels (Invitrogen™) were used for SDS-PAGE in conjunction with Bolt™ MES running buffer (Invitrogen™). These gels are buffered to pH 6.4. Appropriately diluted protein samples were mixed with 4X Bolt™ LDS sample buffer, containing lithium dodecyl sulphate pH 8.5 instead of SDS and two tracking dyes (Coomassie G250 and Phenol Red) in addition to glycerol such that the final concentration of sample buffer is 1X. 10X Bolt™ sample reducing agent was added to a final concentration of 1X which provides 50 mM DTT). Samples were heated at 70^{o}C for 10 min before loading 30uL per well. PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific™) containing nine protein standards (10 -50 kDa) was run alongside protein samples. Run parameters were 165V for 40 min. Gels were stained by submerging in InstantBlue® Coomassie Protein Stain overnight, before destaining by rinsing in water then imaging with a white light filter on a G:BOX Chemi XX6/XX9 gel documentation system using GeneSys software.

2.5 Protein Purification Methods

2.5.1 Purification of His-tagged recombinant proteins by Immobilised Metal Affinity Chromatography (IMAC)

Recombinant proteins expressed in fusion with a peptide affinity tag facilitate purification by affinity chromatography through biorecognition of the tag. Immobilised-metal affinity chromatography (IMAC) is a chromatographic technique which exploits the affinity between transition metal divalent ions and certain amino acid side chains, which in the case of His-tagged proteins is the imidazole side chain of histidine. The fundamental principle is that a histidine-tagged protein will be bound to the metal ions immobilised on a matrix whilst contaminating host cell/endogenous proteins with surface exposed histidine clusters will

either bind weakly or not at all, allowing the desired protein to subsequently be selectively eluted. Imidazole – an analog of histidine - is used to elute the target protein since it competes for binding to the metal.

The resin consists of highly cross-linked agarose beads to which a chelating ligand has been immobilised and pre-charged with metal^{2+} ions – either nickel, cobalt, zinc or copper. The most popular IMAC resin for his-tagged proteins is perhaps Ni-NTA in which nickel is chelated to a nitrilotriacetic acid matrix. The tetradentate NTA ligand securely co-ordinates the Ni^{2+} through four valencies, leaving two co-ordination sites available to interact with imidazole rings of histidine (image). The metal ion therefore is in a state of octahedral co-ordination (from four matrix ligands and two histidine residues). Co-CMA (carboxymethylaspartate) resin is also widely used. Iminodiacetic acid (IDA) resin is an alternative to NTA, although usually suffers from considerable metal leaching and lower purity compared to NTA-based purification.

The most common his-tag is a hexahistidine consisting of six consecutive histidine residues (His_6) although polyhistidine tags of His_4 – His_{10} may be used. Although a decahistidine tag provides the strongest affinity for the medium, hexahistidine is more than sufficient since the six metal-binding sites it provides are enough to shift the equilibrium in favour of association. Thanks to the relatively high affinity and specificity of the his-tag, the level of purity is such that a one-step purification in many cases is adequate. In addition, the tag may improve the solubility and stability of the protein. It is not always necessary to remove the tag, unless it poses interference in downstream applications such as X-ray crystallography.

Despite the fact that histidine represents just 2% of all protein residues, some host cell proteins contain two or more adjacent histidine residues which may co-elute with the recombinant target protein[208]. To reduce this non-specific binding of untagged host cell proteins, a low concentration of imidazole (20 – 40 mM) is included in the binding and wash buffers. Sodium chloride (up to 500 mM) is also routinely included for this purpose. Disulfide bond formation between the protein of interest and other cell proteins, and therefore co-elution, is a possibility, so the inclusion of a reducing agent – at a concentration within the limits of the resin – in all buffers is sensible. Inclusion of a low level (<1%) of a nonionic detergent such as Triton X-100 or Tween-20 also serves to reduce non-specific hydrophobic interactions. Glycerol (<20%) is also commonly employed to help stabilise the target protein.

In brief, the workflow consists of four steps: **equilibration** (of the resin), **binding** (of the target protein), **washing** (to remove weakly bound contaminating proteins), **elution** (of the target protein). Once the resin is equilibrated in binding buffer, the clarified bacterial lysate is loaded onto the column and the unbound proteins collected as flow-through. The application of 5-10 CV (or until the baseline absorbance is stable) of wash buffer serves to detach weakly bound proteins which may otherwise co-purify with the target protein. Ultimately, one must achieve

a balance between high purity and high yield I.e. decreased binding of host cell proteins whilst retaining strong binding of his-tagged target protein. By increasing the concentration of free imidazole (above 100 mM but usually around 500 mM for Ni-NTA; above 50 mM for Co-CMA), the imidazole out-competes the his-tagged target protein for binding to the matrix resulting in selective elution of the recombinant protein. The elution can be performed with either a step (isocratic) or linear gradient; the latter requires the use of an automated FPLC system such as the ÄkTA otherwise gravity flow or the use of a peristaltic pump are applicable. A step elution would usually require 5 CV whereas for a linear gradient elution 10 - 20 CV is recommended. An alternative elution method is that of decreasing the pH such that the imidazole nitrogen atom of the histidine residue (pKa 6.0) becomes protonated and subsequently disrupts the co-ordination bond between the histidine and the transition metal. For Ni-NTA this pH is between 5.3 and 4.5, whereas for Co-CMA it is pH 6.0. However, elution with imidazole poses less risk of negatively affecting the target protein and so is generally the chosen option.
In this work, the resin type and specific IMAC buffers and parameters will be specified for each chapter. Other than when using Ni-NTA agarose superflow (Qiagen) in a benchtop procedure (Chapter 3), an ÄkTA Pure instrument was used, which consists of two system pumps, an injection valve, a UV monitor, a fraction collector (F9-C), and sample inlet, column and loop valves.

2.5.2 Size Exclusion Chromatography (SEC)

To further purify samples following IMAC, size exclusion (SEC)/gel filtration (GF) is often used, which separates molecules based on their hydronamic radius, thereby isolating the target recombinant protein from other co-eluting proteins of different sizes. The resin consists of a porous matrix of spherical hollow beads which are inert and lack adsorptive properties, unlike resins used in affinity, IEX, and hydrophobic chromatography in which the biomolecules actually interact with the resin. Molecule size (and shape) dictates the degree of penetration into the gel matrix and thereby the retention time. As such, larger molecules elute first since they are unable to enter the resin beads and flow straight through the column, whereas smaller molecules diffuse into the pores to varying degrees based on their size which delays their passage through the column. Molecules which are larger than the largest pores in the gel beads elute in the void volume which is equal to the volume of the mobile phase[209]. The size of molecules that elute in the void volume is indicated by the exclusion limit of the resin[210]. The volume of stationary phase is equal to the volume of solvent inside the gel which is available

to smaller molecules and this, minus the void volume, represents the elution volume of those small molecules which distribute freely between the mobile and stationary phases[209]. Each resin has a specified fractionation range which defines the molecular weight range of molecules which have partial access to the resin pores. For example, the Superdex 75 pg resin separates globular proteins in the molecular weight range 3000 to 70,000 Daltons, with optimal performance between 8000 and 50,000 Daltons. Superdex resin is a composite gel consisting of dextran chains which are covalently bonded to a highly cross-linked porous agarose gel bead matrix[209]. The high physical and chemical stability is primarily a result of the agarose matrix whilst the gel filtration properties are determined by the dextran chains[209]. Superdex 75 prep grade has an average wet bead diameter of 34 μm with a range of 24-44 μm. For this resin sample volume should be 1-2% of the total bed volume and ionic strength should be at least 20 mM[209]. For the majority of separations, the overarching aim is to achieve the required resolution in the shortest possible time, and this is determined by the flow rate. Resolution decreases with increased flow rate thus for large molecules a slower flow rate is superior[209]. Regardless of parameters such as flow rate and resin, an ionic strength of at least 0.15 M must be maintained so as to prevent unwanted ionic interactions between solute biomolecules and the gel matrix[209]. Unlike in IMAC procedures, elution is isocratic, that is, the concentration of buffer components remain constant. SEC separates monomers from dimers and higher aggregates and is therefore usually used as a polishing step[209]. SEC can also be used for group separation (desalting/buffer exchange) and in this case, the sample volume can be can be up to 30% of the bed volume.

2.6 Molybdenum Blue Determination of Orthophosphate – Phosphatase Activity Determination

The spectrophotometric molybdenum blue method is a well-established protocol for quantitative determination of inorganic phosphate with predominant application in environmental and water quality analyses, its popularity arising from its simplicity and sensitivity. There was an intensity of communication around the experimental procedure since the 1920s, although the earliest report dates back to 1826 and is widely attributed to Berzelius[211]. Gmelin has also been noted as the first to observe the reaction between phosphoric acid and ammonium molybdate, which forms the yellow precipitate phosphomolybdate, in the 19th Century[212].

By the 1940s there was already a milieu of reported modifications to the method, as compiled in a review by Woods and Mellon, although Fiske and Subbarow's was deemed the most popular at the time as well as the first quantitative and reproducible method and has been the basis on which subsequent variations were made over the decades; summarized in Table 2.3[213-215]. The colorimetric determination of phosphorus report by Fiske and Subbarow has in fact been described as one of the most highly cited papers in the history of biochemistry[216]. Earlier assays for phosphorus relied on the reduction of phosphomolybdic acid by hydroquinone to form a blue colour, the reduction of which was slow and susceptible to interference[216].

The chemical basis of the reaction is the formation of a blue coloured complex - Molybdenum blue - which is formed by the selective reduction of the molybdenum in a heteropoly acid – such as that formed by heteropolymolybdates of phosphate, silicate, arsenate or germanate - in the presence of excess molybdate. When phosphate is the heteroatomic species under analysis, as in this report, the phosphate-containing sample is treated with a pre-acidifed and pre-mixed reagent solution comprising a source of Mo^{VI} - such as ammonium molybdate - in excess to form the heteropoly acid 12-molybdophosphoric acid (12-MPA), the controlled reduction of which forms molybdenum blue (Fig 2.2). Since the amount of this resultant complex is proportional to the amount of phosphate present in the sample solution, the corresponding intensity of the blue colour subsequently serves for the spectrophotometric quantification of phosphate[211]. The spectrophotometric measurement is taken at the wavelength at which absorbance is maximal; this has been predominantly 700 nm although some more relatively recent reports identified different wavelengths such as 655 nm[217], 840 nm[218] and 850 nm[219] at which absorbance of molybdenum blue was greater than at 700 nm (Table 2.3). One variant of the molybdenum blue method features the direct detection of the yellow phosphomolybdic acid without reduction to molybdenum blue and as such this is monitored at 355 nm[220]. Inorganic orthophosphate is subsequently quantified from a calibration curve generated with phosphate standards of known concentration.

The first stage involves the formation of a Keggin ion around the analyte anion, forming 12-MPA:

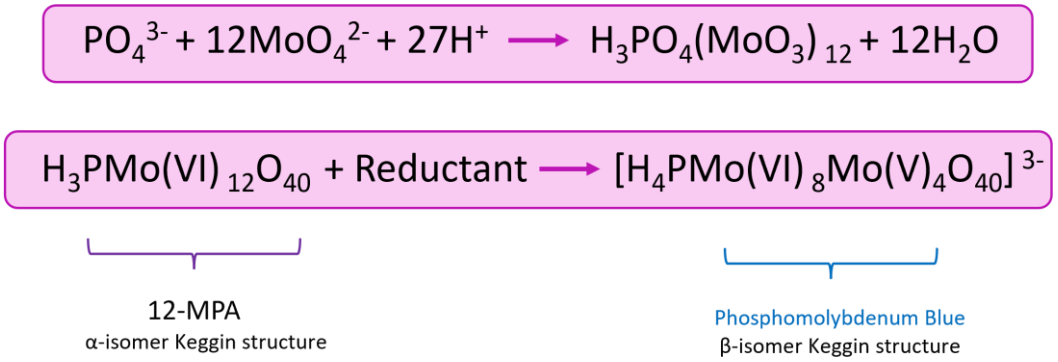


Figure 2.2 Two-stage Molybdenum Blue-forming reaction.

Experimental parameters and considerations: The underlying chemistry is deceptively complex - involving redox and equilibria chemistry - with many experimental parameters which must be considered when optimising the method. In general, all molybdenum blue reactions require a strong acid, a source of Mo(VI) and a reductant. A large excess of Mo(VI) over P_i is needed to avoid leaving significant residual amounts of 11-MPA by driving the phosphomolybdate equilibria to form 12-MPA almost exclusively[211]. Once reduced, 12-MPA depends on excess reductant to protect the phosphomolybdenum blue from re-oxidation by dissolved O_2 [211]. However, it is well-documented that excessive reductant concentrations induce direct Mo(VI) reduction after 12-MPA reduction is apparently complete. As such, the reduction of 12-MPA must be controlled since it is this potential of excess molybdate to be itself reduced which is responsible for the common undesirable issue of reagent blank absorbance, which should be minimised[213,218]. The redox behaviour of 12-MPA is highly acid-dependent therefore concentrations of acid and molybdate are important for controlling its reduction, such that the optimum ratio of molybdate to acid is that which effects maximum reduction of the 12-MPA whilst mitigating direct reduction of the uncombined molybdate, in a given time[211,213]. For example, a greater concentration of Mo(VI) requires a greater acidity to mitigate the effect on the reagent blank, however, this is at the expense of 12-MPA decomposition and resultant loss of sensitivity at higher acidities. Thus, the lowest possible acidity should be used in order to minimise phosphomolybdenum blue decomposition whilst avoiding unsolicited Mo(VI) reduction[211]. The pH 0 – 1 range facilitates the formation of suitable amounts of stable reduced product without excessive direct reduction of Mo(VI) thus orthophosphate determination procedures are usually performed as such[211]. Hence, the intensity of the phosphomolybdenum blue complex colour is influenced significantly by acidity[218]. The vast majority of published MB

methods utilise sulphuric acid, despite some inhibition of formation of 12-MPA with this acid[211]. A multitude of variations in reductants and reaction conditions and combinations of such have been tested and this has resulted in the observation that different reductants or different conditions with the same reductant generate products with deviations in UV-visible spectra and apparent molar absorptivities. There are two specific absorption bands corresponding to the reduced molybdenum blue complex and these are strong inter-valence charge transfer bands resulting from electron exchange between Mo(V) and Mo(VI) centres. The positions of these two intensely absorbing bands are influenced by the degree of reduction, the protonation state of the absorbing species and the reductant type (metallic vs organic)[211]. With all these factors taken into consideration, the overarching goal of method optimisation is to achieve a balance between maximum formation and reduction of 12-MPA yet minimisation of isopoly molybdenum blue formation (species formed through through the reduction of excess molybdate). An in-depth review of the fundamental chemistry of the molybdenum blue reaction, beyond the scope of this report, is provided in reference [211]. This inherently provides chemical rationalisation for the perpetual modifications to this method for its optimisation.

Enzymatic assays using the Molybdenum Blue method: Phytase activity determinations are often based on analysing released phosphate by colorimetric analysis such as the Molybdenum Blue method. When using this method in enzymatic studies requiring P determination, there are several caveats which arise from an integral aspect of the application of this method - the monitoring of product formation (i.e. released P_i) in contrast to substrate degradation[221]. The measurement of free P_i has long been known to result in overestimations of P_i concentration and this is due to the inability of the method to distinguish between the different phosphorus compounds in the molybdate-reactive phosphorus pool, which gives rise to errors pertaining to non-enzymatic (and non-specific enzymatic) hydrolysis of P [219,221]. One of the contributing factors is the strong acid conditions required for conventional molybdenum blue methods which interferes with Pi determination due to rapid hydrolysis of labile organic P compounds (P_o) and also causes precipitation of enzymes[211,219,221]. The time-dependent nature of labile P_o hydrolysis culminates in a progressive increase in intensity of blue colour since errors associated with this are not corrected. Thus for enzymatic assays this effect can be reduced by adaptation to a mild acid buffer (pH 4-5) which slows the rate of acid-labile P hydrolysis[219]. A method was developed by Dick and Tabatabai (1977) to ameliorate the interference from of non-enzymatic hydrolysis of Po, in which excess molybdate ions are complexed with added citrate-arsenite to prevent further formation of blue colour from acid labile P hydrolysis-derived Pi –since subsequent Pi is preventing from complexing with the already complexed

molybdate[222]. However, it is not only the acid conditions which induce unwanted P hydrolysis and subsequent false positives in enzymatic Pi determination, since there is also Pi-production catalyzed by non-phytase inositol phosphatases. Firstly, some released phosphate may be derived from the degradation of phosphate containing compounds other than phytate by non-phytase-specific phosphatases, which are often present in microbial and plant sample preparations. Additionally, the products of IP6 hydrolysis are also substrates for the same enzyme, and also for the non-specific phosphatases, which degrade the lower inositol phosphates (from IP5 downwards)[221]. In fact, even commercial IP6 may contain contamination of lower inositol phosphates, with reports of up to 12% IP5 present[223]. A study by Qvirist et al provided evidence for the contribution of Pi hydrolysis by the lower inositol phosphates by comparison with HPIC. This revealed that the released Pi at 15 min of the colorimetric assay originated from degradation of several lower inositol phosphates in addition to IP6. It also concluded that phytase activity determination based on released Pi results in activities up to 386% higher than from IP6 data[221]. All of these factors generate false positives from the erroneous Pi measurement and as such, the contribution of IP6 hydrolysis in relation to the total inositol phosphatase reactions cannot be accurately determined by assessing Pi. Therefore, measuring released Pi per time does not represent only the enzymatic activity on the specific substrate IP6[221]. Since the molybdenum method only detects total molybdate-reactive P, it is more appropriately used when total P is the sought after measurement, such as in environmental analyses[219]. As discussed in their comparative study of phytase activity determination methods – comparing substrate (IP6) versus product (Pi) based assays - Qvirist et al advised the use of HPIC for determining phytase activity, since this monitors IP6 hydrolysis exclusively and mitigates the overestimations of Pi from methods based on product detection[221]. For enzymatic activity determinations based on phosphate release, they also recommend the usage of distinct terms to differentiate between total inositol phosphatase activity; total phosphatase activity; and phytase activity, where these are defined as, respectively: the degradation of the total mixed inositol phosphate pool (the sum of all IPx hydrolysis) catalyzed by phytase during the assay in assays where IP6 is the only substrate and phytase is the only enzyme; the phosphate released from all possible sources by all phosphatases during the assay, for use when there are other phosphate sources in addition to IP6 present as substrates, and/or when the extract to be assayed may contain non-phytase phosphatases; the term 'phytase activity' reflects the released product per time from IP6 only, which by definition is performed only by phytases, and which is assessed by analysing IP6 concentration[221].

Authors	Year	Modification	Λ (nm)
Fiske & Subbarow[214]	1925	The predecessor upon which subsequent methods were founded: acid ammonium molybdate solution was added to a phosphate containing sample to form the yellow phosphomolybdic acid, subsequently reduced by 1-amino-2-naphtol-4-sulfonic acid (ANSA) to molybdenum blue	600-700
Holman[224]	1943	Potassium iodide as reducing agent; ~3x more sensitive than Fiske & Subbarow method; indistinguishability between inorganic and organic phosphate compounds	N/A – used colorimeter
Dick & Tabatabai[222]	1977	Complexed excess molybdate ions with citrate-arsenite to prevent further formation of blue colour derived from labile P_o hydrolysis	700
Peterson[225]	1978	Increased stability over time and decreased sensitivity to interfering substances; based on ANSA reduction; used SDS	700
Heinonen and Lahti[220]	1981	Direct detection of the yellow phosphomolybdic acid, without reduction to molybdenum blue	355
Shand & Smith	1997	Reduced interference of protein precipitation by addition of dimethyl sulfoxide	?
Bae et al [226]	1999	Ferrous sulphate as reductant; trichloroacetic acid	700
Katewa & Katyare[215]	2003	A mixture of hydrazine sulphate and ascorbic acid as reductant; optimised stability of the colour complex; applicability in both enzymatic and nonenzymatic reactions	820
He & Honeycutt[219]	2005	Based on Dick & Tabatabai method; 0.2% SDS for prevention of enzyme precipitation; readings at 850nm improved sensitivity	700 & 850
Central Pollution Control Board and Agilent[221]	2011	For reproducible analyses of phosphate in water and wastewaters	700
Pradhan & Pokhrel[218]	2013	Hydrazine hydrate as reducing agent; advantage is blank is colourless	840
Qvirist L, Carlsson NG, Andlid [221]	2015	Based on the Peterson protocol with ANSA exchanged for ascorbic acid	700
McKie & McLeary[217]	2016	High-throughput method for the measurement of total phosphorus and phytic acid in nonprocessed foods and feeds; sulphuric acid-ascorbic acid solution as the reductant.	655

Table 2.3 **Chronological collation of selected modified molybdenum blue methods including a brief summary of the modification and the wavelength of molybdenum blue complex detection. By all means not an exhaustive list; this is just a representative cross-section of the existing reports.**

2.7 Separation of Inositol Phosphates by Anion Exchange HPLC (HPIC) – Phytase Activity Determination

HPLC is a technique to separate, identify and quantify components within a sample utilizing a mobile phase pumped through a stationary phase. The most common form of HPLC used to separate inositol phosphates is based on anion exchange[227], whereby the stationary phase consists of a positively charged ion exchange resin with an affinity for molecules having a net negative surface charge, the degree of charge being that which forms the basis of separation of

the molecules. At a given pH, different proteins will bind to the resin to varying extents governed by their PI value, facilitating their separation when a salt gradient is applied. The elution is such that proteins with PI values closer to the working pH will elute at lower ionic strength and contrariwise.

Ion exchange chromatography is a long-established method of choice for fractionation and purification of myo-inositol phosphates and inevitably the technique has undergone much refinement over the years[228], since the early days of stepwise elution with HCl[229]. Prior to this, quantitative analysis of phytic acid was achieved by precipitation with ferric chloride as early as 1914. The interest in achieving separation of the lower phosphorylated inositol phosphates was initially instigated by calcium metabolism studies[230]. Early procedures were based on isocratic ion chromatography or by ion-pair reversed phase HPLC, however, these failed to resolve the mono- and diphosphates[228]. It wasn't until 1988 that the ion chromatographic method (suitable for phytic acid) was transformed into a gradient elution procedure[231,232], which allows ions of widely different retention behaviours to be eluted in the same run[233]. This isomer- specific gradient elution protocol with post column detection was able to resolve the positional isomers of the spectrum of myo-inositol phosphoesters in 30 min. Subsequently, Rounds and Nielsson in 1993 reported an improvement on resolution by a method which utilised a polystyrene-based strong anion exchange column in conjunction with gradient elution at pH 4[228]. Nevertheless, standard anion exchange chromatography still cannot resolve the optical isomers of inositol polyphosphates – for example, the enantiomeric pair Ins(1,4,5,6)P4 and Ins(3,4,5,6) cannot be differentiated, and similarly the Ins(1,5,6)P3 and Ins(3,4,5) pair[227,234]. This is because positions 1 and 3 are chemically equivalent, as are positions 4 and 6. The discrimination of these enantiomers is only possible with the use of a chiral-based column or extensive chemical degradation techniques[227]. Theoretically there are a staggering 66 possible inositol phosphate isomers – 3 of which are cyclic[233]. Thus, aside from the enantiomers, there are theoretically 39 inositol phosphates which can be separated by anion exchange: 4 x IP5 isomers, 9 x IP4 isomers, 12 x IP3 isomers, 9 x IP2 isomers, 4 x IP1 isomers in addition to IP6[234]. Skoglund et al established the elution order of 25 inositol phosphate peaks in their 1997 paper, and this was accomplished by using two combined HPIC systems[233]. A positive correlation was observed between the number of phosphate groups and their retention on the column[233]. Several years later, Chen and Li achieved separation of 35 inositol phosphates into 27 peaks by way of a linear gradient elution (65 min) with HCl on a CarboPac PA-100 column, followed by detection at 295 nm after post-column complexation with iron (III)[234].

Due to the absence of a characteristic absorption spectra in the ultraviolet or visible region of inositol polyphosphates, post column derivatization is necessary before ultraviolet absorbance

Detection[234]. Early methods described the use of the Wade reagent (iron chloride and sulphosalicylic acid) – in which the detection method involves a ligand-exchange reaction between the Fe^{3+}-sulfosalicylic acid complex and eluting inositol phosphates[228,235] which results in a decrease in absorbance at 500 nm. The use of nitric acid as an eluent is necessary to prevent iron-phytate precipitation (favourable at higher pH and low phytate: iron molar ratios) by providing an acidic environment[236]. At the time, gradient anion exchange HPLC in conjunction with post-column detection via the Wade reagent was sufficient, so long as the discrimination of isomers was not crucial[228]. Modern methods now use post-column complexation with Fe^{3+} ions in perchloric acid solution and UV detection at 290nm, as employed by Phillipy and Bland[16] and subsequently Skoglund et al[233].

Instrumentation: An anion exchange HPLC system consists of a sample injection pump, two anion exchange columns in series within a column oven, a second pump and a UV detector. Usually a guard column is placed before the analytical columns to prevent sample contaminants from eluting onto the analytical columns. Effluent from the analytical column is then mixed with a colour reagent in post-column reactor before UV absorbance detection.

In this work, the reaction products of IP6 or IP4 (2.5 mM) incubated with *Bt*MINPP mutants (10 nM) at 25°C for various lengths of time were separated on a CarboPac PA-200 column (two columns in series; 3x50mm and 3x250mm) which contains a resin of hydrophobic, polymeric pellicular anion exchange resin. Samples were injected onto the column (preceded by a guard cartridge) by a sample injection pump (Jasco PU-2089 I Plus quaternary insert pump) at a flow rate of 0.2 mL/min. The elution protocol consisted of a 25 min methane sulfonic acid gradient (0-0.6M) followed by 14 min at 0.6M and finally 11 min of water, at 0.4 mL/min. The effluent was then pumped by a Jasco PU-1585 intelligent HPLC pump for mixing with a post column colour reagent (0.1% ferric nitrate and 2% perchloric acid). UV detector: Jasco UV 1575 intelligent UV/Vis detector, 16uL cell. Autosampler. ChromNAV(Jasco) software.

2.8 Macromolecular X-ray Crystallography

2.8.1 Protein structure solution by X-ray crystallography

The prerequisite for X-ray crystallography experiments is the successful production of diffraction quality protein crystals – yet this is no mean feat, often requiring multiple attempts and representing a potential bottleneck. The workflow from purified protein to an experimental crystal structure involves several distinct stages - once a protein crystal is

obtained from crystallisation experiments, it is subjected to X-ray beams to generate X-ray diffraction data. The collected diffraction data, along with a homology model, is used to generate an electron density map into which the protein structure can be built. Multiple rounds of refinement ultimately result in a structural model whose coordinates have a satisfactory fit to the experimental electron density, reflected by the R factor value. The quality of the model is then determined through various scores/parameters.

2.8.2 Fundamentals of protein crystal architecture

Protein crystallization results in the formation of structured, ordered lattices which are free of contaminants and large enough to provide a diffraction pattern when hit with an X-ray beam. The unit cell is the fundamental repeating unit of the crystal which extends in all three dimensions of space to form the crystal lattice, whereas the asymmetric unit (ASU) is the smallest unique portion of a crystal structure which, when subjected to crystallographic symmetry operations (rotations, translations and screw axes; the latter being a combination of the former two operations) can generate the complete unit cell. Put more simply, it has no internal symmetry. Translations of the unit cell in three dimensions makes up the entire crystal lattice, such that it is an almost perfectly symmetrical arrangement of asymmetric units. The ASU does not automatically represent the biological assembly, that is, the macromolecular assembly which is purported to be the functional form of the molecule. The biological assembly can thus be either the ASU itself, a subset of the ASU, or the structure resulting from the application of certain symmetry operations to the ASU. The unit cell is defined by crystal axes which are three distances a, b, c and three inter-axial angles α, β, γ. This underpins the classification of the unit cell into seven crystal systems – cubic, triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal – which further give rise to fourteen Bravais lattices (Fig 2.3). That is, within the 7 crystal systems there may be extra translational symmetry in the unit cell – centring – such that a lattice can either be primitive (P) i.e. an absence of centring, body-centred (I), face-centred (F), or base-centred (C). In a primitive lattice, the lattice points lie only at the vertices of the unit cell. Based on their symmetry operators, unit cells are classified into space groups which represent a complete description of the symmetry of a crystal. Combinations of these symmetry operators give rise to 32 point groups. Taken together, the space group and the contents of the ASU define the positions of all atoms in the protein crystal, with knowledge of the former being essential in order to solve the protein structure. The combination of the point group and lattice symmetries yields a total of 230 different space groups, although this number is reduced to 65 enantiomorphic space groups for chiral macromolecules[237]. The most frequently obtained space groups are $P2_12_12_1$, which accounts for around 30% of all protein crystals, and $P2_1$. It has been demonstrated that these

are the least restrictive to packing and therefore there are more ways they can be generated[238]. Some are less common, accounting for less than 1% of all observed crystals, such as $P4_1$ and $P6_522$. In the space group symbol, the capital letter indicates the unit cell centre (P, C, I, F or R), followed by the symmetry along one (triclinic and monoclinic) or three directions (all other crystal systems). It is important to note that different pHs can yield different packing orientations.

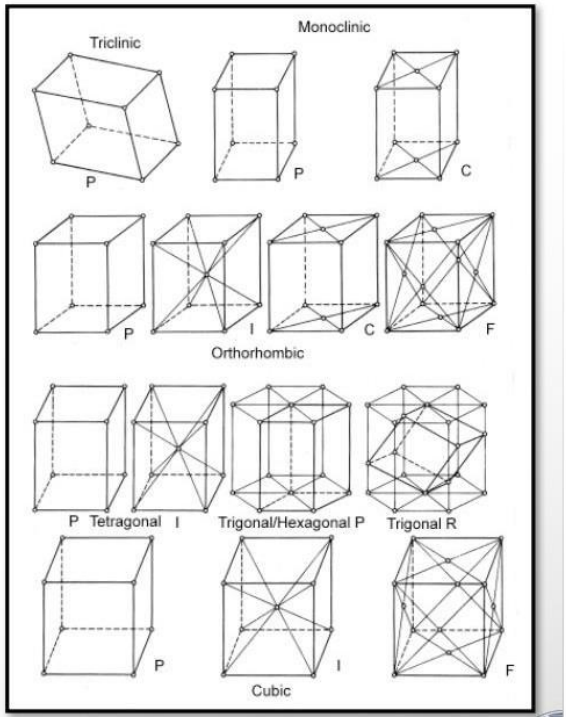


Figure 2.3 The 14 Bravais lattices.

Crystal System	Point Group	Space Group	Edge Length	Interfacial Angles
Triclinic	C_1	P1	a≠b≠c	α≠β≠γ≠90°
Monoclinic	C_2	P2, $P2_1$, C2	a≠b≠c	α=γ =90°C≠β
Orthorhombic	D_2	P222, $P222_1$, $P2_12_12$, $P2_12_12_1$, $C222_1$, C222, F222, I222, $I2_12_12_1$	a≠b≠c	α=β=γ = 90°
Tetragonal	C_4	P4, $P4_1$, $P4_2$, $P4_3$, I4, $I4_1$	a=b≠c	α=β=γ = 90°
	D_4	P422, $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$, I422, $I4_122$,		
Trigonal	C_3	P3, $P3_1$, $P3_2$, R3	a=b=c	α=β=γ≠90°
	D_3	P312, P321, $P3_112$, $P3_121$, $P3_212$, $P3_221$, R32		
Hexagonal	C_6	P6, $P6_1$, $P6_5$, $P6_2$, $P6_4$, $P6_3$	a=b≠c	α=β=90°,γ=120°
	D_6	P622, $P6_122$, $P6_522$, $P6_222$, $P6_422$, $P6_322$		
Cubic	T	P23, F23, I23, $P2_13$, $I2_13$	a=b=c	α=β=γ = 90°
	O	P432, $P4_232$, F432, $F4_132$, I432, $P4_332$, $P4_132$, $I4_132$		

Table 2.4. The 7 crystal systems and their associated point groups which yield 65 space group

2.8.3 Protein crystallisation – Theory and practice

The requirements of a protein for crystallisation is that it is of very high purity (at least 97-99% pure), stable and monodisperse. In order to form the crystalline lattice, individual protein molecules must associate in a consistent and ordered orientation such that they are aligned in a series of repeating unit cells, which are held together by non-covalent interactions. To transition from a dissolved, disordered state to this crystalline form, the solubility of the protein must be reduced, but in a specific controlled way such as to avoid the otherwise inevitable formation of amorphous precipitated protein, phase separation or aggregation. For a protein to crystallize, the protein solution must be brought into a supersaturated state, which is thermodynamically unstable. Essentially, the crystallisation process is a transition from the supersaturated state to a regular saturated state. Crystal growth consists of two distinct phases – first, an energetically unfavourable nucleation phase in which individual molecules form small initial aggregates, followed by the energetically favourable growth phase. Within the supersaturated phase there are two zones as shown in the phase diagram (Fig 2.4) – the so called metastable and labile zones. The protein solution must first be pushed into the labile zone for crystal nucleation to occur[239-241]. A nucleus has historically been defined as the minimum amount of a new phase capable of independent existence[242]. However, if the protein solution remains in the labile zone for too long, rapid growth of crystal nuclei will occur, resulting in an excess of small crystals - which are undesirable given that large protein crystals (~500 μm) are usually required for successful X-ray diffraction. The key is to allow the protein/precipitant solution to approach the nucleation zone very slowly such that developing nuclei have sufficient time to grow. Once nucleation is induced such that crystal nuclei reach a critical size of around 100-1000 molecules, further nucleation must be restricted so that controlled crystal growth may instead be promoted. Crystal growth requires a lower level of supersaturation than for nucleation and this is represented by the metastable zone, in which nuclei will develop into crystals, but no nucleation will occur. Unlike nucleation, the growth phase is energetically favourable, resulting in an increase in entropy since molecules in the crystal are able to partially shed their hydration shell. Ultimately, there is a fine line between the different processes of crystal nucleation and growth and attempting to control these is an arduous task involving patience, knowledge and sometimes a little luck[239-241].

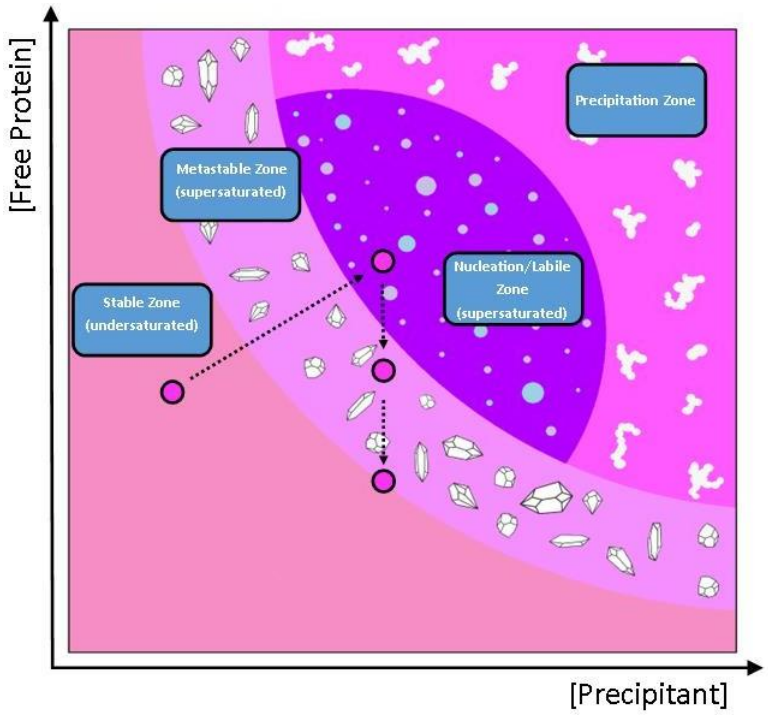


Figure 2.4 The Ostwald-Miers phase diagram for protein crystal growth. Adapted from McPherson & Gavira 2014[239].

In practice, in order to achieve the transition from a stable solution to a supersaturated solution, the concentration of either the protein or the precipitant must be increased such as to arrive at a desired position in the phase diagram (Fig 2.4). The most common technique for protein crystallisation is vapour diffusion, in the form of either sitting or hanging drops. The basis of this method is that of a closed system equilibrium between a protein-containing droplet and a reservoir solution containing the same components but at higher concentrations. The consequent concentration gradient results in an increase in concentration of both precipitant and protein in the protein-containing droplet as water vaporises and transfers to the reservoir. Once the concentration of the protein reaches supersaturation, if all other parameters (pH, ionic strength, buffer, additives) are suitable, crystallisation may be induced. The precipitant plays an integral role in the generation of the concentration gradient, the nature of the precipitant usually being ammonium sulphate or polyethylene glycols of various lengths[239-241].

The protein crystallisation process is influenced by a number of complex interactions between variables such that the aim of identifying a set of conditions which results in crystals of the correct morphology, purity and growth rate is often empirical. To aid in the laborious task of screening for appropriate crystallisation conditions, commercial crystallisation screens are available which include pre-mixed sets of hundreds of solutions with differing combinations of

buffers, pH, ionic strength, salts, precipitants, additives and metal ions. These screens may be set up using specialised liquid handling robotics which, as well as increasing the amount of screens which can be set up in a given time, conserve valuable purified protein, since nanoscale volumes are accurately dispensed, allowing the usage of minimum amounts of protein. The initial conditions identified as yielding crystals often require further optimisation before diffraction-quality crystals are obtained. When crystals of sufficient size and quality are obtained, they are harvested and cryoprotected before freezing in liquid nitrogen to protect them from the radiation damage they are subsequently subject to[239-241].

2.8.4 Macromolecular X-ray diffraction

In an X-ray diffraction experiment, a cryoprotected crystal is mounted and aligned on a goniometer into a high intensity X-ray beam at a synchrotron, a facility which exploits the light produced by high energy particles undergoing acceleration to generate a tunable collimated beamline. A robotic sample changer performs the loading of the sample and tools are used to optically align the sample based on strength of diffraction. The X-ray diffraction procedure is performed at cryo-temperatures (100 K) to decrease the extent of radiation damage and to prevent thermal motion of the protein which risks the quality of the diffraction images. Another strategy for minimising radiation damage to the crystal is to attenuate the X-ray beam in order to decrease the dose of radiation delivered. A balance has to be achieved between the advantage of using a higher dose to obtain a greater signal to noise ratio and the risk of not being able to complete data collection due to rendering the sample unusable. In practice, the user specifies the percentage transmission, and the data acquisition system calculates the number of foils – inserted into the beam – to use to provide the closest transmission for a given wavelength[243]. Data collection consists of a series of single images of the X-ray diffraction pattern that are obtained as the crystal is rotated 360 degrees, one degree at a time, about an axis, ideally perpendicular to the monochromatic X-ray beam, until a complete data set is acquired, usually consisting of thousands of reflections. This is called the Arndt-Wonacott rotation method of data collection[244]. The incoming X-rays get scattered into discrete waves by the electrons of the atoms in the crystal and add together constructively or destructively resulting in the formation of a diffraction pattern which is captured on a detector as a pattern of dark spots. The crystal lattice acts as a grating which generates the constructive interference of the diffracted X-ray photons and therefore results in the observed spacing between the spots. This is dictated by Bragg's Law. Diffraction images therefore represent two-dimensional sections through a three-dimensional lattice of diffraction maxima. The lattice of diffraction maxima is reciprocally related to the crystal lattice in real space, such that large crystal axes generate short distances

between diffraction spots and vice versa. The complete diffraction data set represents the three-dimensional reciprocal lattice. The structure of the macromolecule thus becomes encoded in the distribution of diffracted waves. The information contained in an X-ray diffraction experiment is the position and amplitude of the diffraction maxima (reflections). The position of diffraction maxima is dictated by the *dimensions* of the unit cell, whilst the amplitudes reflect the *contents* of the unit cell. Although diffraction patterns do not possess translational symmetry, they do exhibit a centre of symmetry defined by Friedel's law, which reduces the 32 point groups to 11 Laue groups.

For the work reported in this thesis, X-ray diffraction data was collected at the medium energy synchrotron at the Diamond Light Source (DLS; Oxfordshire, UK), which produces a 3 GeV electron beam. Irradiation of crystals was performed on the I24 microfocus mx beamline which has a beam-size of 5x5-50x40 (variable aspect ratio) and a wavelength of 0.62–1.77A°. This beamline incorporates a Pilatus3 6M detector. Cryo-protected crystals were sent in a dry shipper under liquid nitrogen and the diffraction experiments were performed at 100K.

2.8.5 X-ray diffraction data processing

Following data collection, the data must be processed. The individual diffraction images need to be processed into a single dataset containing intensity information about all the contributing X-ray reflections. That is, every single diffraction maximum must be indexed and integrated[237]. Previously, the output from the data collection procedure was a set of raw diffraction data, but advances in data reduction software over recent decades has facilitated the acquisition of a set of scaled intensity data as the output[243]. The data processing stage comprises several steps, which is commenced by locating a selection of diffraction spots on the images. First, indexing involves assigning each reflection an unambiguous identifier which mark their position in the reciprocal lattice in Fourier Space. These are described by the Miller Indices (h,k,l) which are the integer coordinates of the reflections in the reciprocal lattice. This is executed by projecting the 2D detector coordinates onto scattering vectors on the 3D Ewald sphere (Fig 2.5), corresponding to lattice points in the reciprocal lattice[238,245]. Reciprocal space is related to real space by a Fourier transform and is a way of visualising the result of the Fourier transform. Indexing is usually automated thus termed 'autoindexing'. In addition to providing an index for each reflection, it also yields information on the crystal orientation relative to the diffractometer and estimated unit cell dimensions along with an indication of crystallographic symmetry[245]. The pre-requisite for these is knowledge of the crystal to detector distance, the direct beam position on the detector and the radiation wavelength[237]. The following step is

parameter refinement which results in accurate (close to the true value) and precise (small standard deviation) values for the unit cell parameters and crystal orientation angles with respect to the diffractometer, since autoindexing yields only an approximate set of parameters. Performed in parallel, positional refinement optimises the fit of observed to calculated spot co-ordinates on the detector, whereas post-refinement optimises the distance between the reciprocal lattice points and the Ewald sphere, by analysing the relative intensities of the different partials belonging to the same reflection, and can only be performed after the intensities have been measured[237,245]. Next, the integration process extracts the intensity information of the reflections which is dependent on both their amplitude and phase. This involves removal of the background contribution, which is performed by either summation integration or profile-fitting integration[237,245]. Used for strong reflections, summation integration adds the contributing pixel values together and subtracts the assumed background which is underneath. For weak spots whose pixels may be obscured by the background, it is better to use the profile-fitting method which uses the spot shapes of adjacent reflections[245]. The final steps, scaling and merging, provide the most useful statistics about the quality and resolution of the data collection and the processing, and gives a good indication of the true symmetry of the crystal[237]. Scaling corrects for variation between images by putting the measured intensities onto a common scale which accounts for imperfections in the experiment. Merging is the process of merging partial and symmetry related reflections to produce a complete set of reflections which are easier for downstream computational handling. Therefore, only after scaling and merging are the most useful agreement indices on the self-consistency of the data i.e. merging R values, half-dataset correlation coefficient ($CC_{1/2}$) etc obtained[245]. Nevertheless, although a self-consistent dataset is a pre-requisite for subsequent structure solution, it may not necessarily be sufficient. Ultimately, regardless of how good a processing program is, a particularly bad dataset will not be able to be successfully integrated[245]. In practice, automatic data processing software performs the indexing, integration, merging and scaling of each reflection in a diffraction image, and generates a single text file from thousands of images.

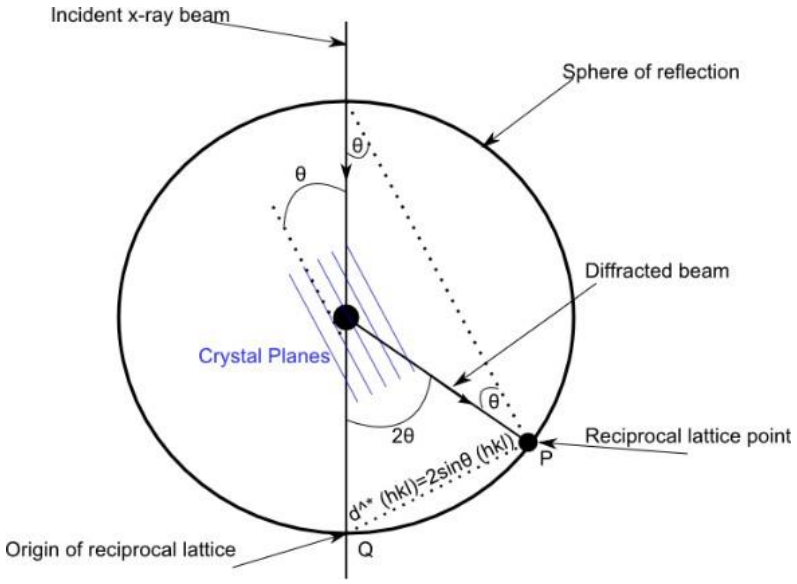


Figure 2.5 The 3D Ewald Sphere, demonstrating the relationship between the crystal's reciprocal lattice, the X-ray wavelength and the angle of diffraction for a given reflection. Image from University of Warwick, Dept of Physics.

xia2 is one of the most common data processing pipelines employed at synchrotrons and is usually triggered to run automatically by the data collection program on the macromolecular beamlines[245]. xia2 applies existing processing software to generate Miller indices along with their corresponding averaged and corrected intensities[243,246]. xia2 comprises three modes, one of which is 3dii, which is useful in cases of weak diffraction since it uses all available images for the initial characterisation of the diffraction pattern. Since it is not possible to know which data reduction strategy will yield the greatest quality results, xia2 is run in all three modes at the Diamond Light Source (DLS). It applies the DIALS package for indexing and integration[243,246]. fast_dp was developed at DLS as a faster and more streamlined version of xia2, which outputs almost real-time feedback on data quality such as merging statistics and merged intensities in the correct point group[243,246]. However, xia2 is more thorough and as such, more applicable for direct downstream analysis without additional re-processing by the user. Subsequent analysis of the processed diffraction data is important to assess quality and detect any pathologies that may complicate structure solution, such as twinning. A program which computes this is xtriage within the Phenix suite[247].

During data collection on a macromolecular crystallography (MX) beamline at DLS, the DISTL package records the total number of reflections, the number classified as good Bragg candidates and a resolution estimate, which are displayed in the ISPyB information management system as a real-time graph. Approximately constant numbers of spots and

resolution estimates would be expected for a well-diffracting sample with no visible anisotropy, fully immersed in the X-ray beam[243]. Diffraction datasets were automatically processed with the xia2, xia2 3dii, autoPROC, fast_dp, and autoPROC+STARANISO DLS-integrated pipelines.

2.8.6 Quality indices of X-ray diffraction data

The quality of the diffraction data inherently dictates the quality of the macromolecular crystal structure solved from it[248]. Therefore, data processing statistics are executed to infer the quality of both the crystal and the data collection[245]. There are several metrics in use and values are given for both the overall data and data from the highest resolution shell only. At this point, it is appropriate to provide rationalisation for this. Resolution and signal intensity are correlated in that the higher the resolution of a reflection (and the further away from the centre of the detector), the weaker the signal[249]. In theory, the resolution limit of the dataset is the point at which the signal is too weak to be useful[249]. To mitigate the issue of a weak signal being mixed with noise and inevitably incorporating that noise into the calculated electron density map, the data is truncated to omit those weak high-resolution reflections[249]. The point at which it is truncated, however, is a source of much debate[249,250]. The rule of thumb until relatively recently was to truncate the data at the threshold where the signal-to-noise ratio was equal to 2, such that all reflections with a signal less than twice the intensity of the estimated noise would be discarded - although this has since been relaxed to 1.5 as a result of underestimation of the information in the excluded data[249,251]. The highest resolution shell can thus be defined as the last shell where completeness is ~100% and the signal-to-noise ratio is 1.5. In practice, the resolution cut-off is automatically determined by the processing program, although the user can manually modify this. Results are thus generated for the highest resolution/outer shell separate to the rest of the data which lacks inclusion of the outer shell data, and this is to aid in the decision of whether to include the outer shell data i.e. if it contributes more information about the structure than excluding it does. Nevertheless, this is a much debated topic and beyond the scope of this chapter. The key indicators of data quality are resolution, completeness, I/σ (signal-to-noise ratio), and R_{merge}, with the latter two parameters also commonly used to determine the resolution cut-off[248,249]. Usually reported as I/σ(I), the average signal-to-noise ratio of the reflection intensities, is of course a case of the higher the value the better. R_{merge} is an indicator of precision and is a measure of the agreement between multiple measurements of the same reflection i.e. the extent to which measurements of the same reflection differ in intensity from the average intensity of that reflection. Lower values are better - a value of 40-60% for R_{merge} is usually an indication of the highest resolution shell, since measurements are deemed error-prone at values above this[249,250]. Completeness measures the

coverage of all theoretically possible unique reflections within the dataset thus the higher the value the better. The rule of thumb for decent data is a completeness value above 93% overall, and above 70% in the outer high resolution shell[248]. Resolution is defined as the minimum spacing (d) of crystal lattice planes which are still able to generate measurable X-ray diffraction. The smaller the spacing, the more independent reflections there are available to define the structure therefore the higher the resolution. The standard terms to describe resolution are 'low, 'medium', 'high' and 'atomic'[248].

2.8.7 Quality indices of the refined model/validation parameters

The R-factor is possibly the most important indicator of structure quality. It is a measure of the global relative agreement between the experimentally obtained structure factor amplitudes, F_{obs}, and the calculated structure factor amplitudes, F_{calc} obtained from the model. It is also called the residual, since a residual in mathematical terms is the difference between predicted and observed values of a particular variable[248]. Throughout structure refinement, the R-factor is re-calculated to assess progress. For perfect agreement with the data, the R-factor would have a value of 0%, whereas for a random model the value would be ~60%[250]. The established guideline is that a well-refined structure should have a R-factor of less than 20%[248]. When the diffraction data is of very high quality, the R-factor essentially reflects errors in the model[248].

In order to mitigate bias, a small subset of around 1000 randomly selected reflections (or around 10% of the experimental observations) omitted in model refinement are used to analogously calculate another R-factor, R_{Free}[248,250]. It is prudent to note that these reflections may have still influenced model definition. The R_{Free} is then calculated by measuring how well the model predicts the set of reflections that were not used in refinement. The R-factor which reflects the larger 'working set' of reflection data is thus termed $R_{work.}$ R_{Free} will always be larger than $R_{work,}$ however it should not exceed R_{work} by more than 7%, since this indicates over-fitting of the experimental data or a serious defect in the model[248,250]. For an ideal model without over-interpretation, the R_{Free} will be only slightly higher than the R_{work} value. $CC_{1/2,}$ the half-dataset correlation coefficient, is the correlation coefficient between the intensities of two arbitrarily divided halves of the dataset, with greater values being desirable. Essentially, it is a measure of how well one half of the data predicts the other half, with a value of 1 indicating a perfect correlation and a value of 0 indicating no correlation. At low resolution, the correlation is around 1, and this decreases as the resolution increases, such that the useful range of $CC_{1/2}$ is considered to be within the range of 0.1 to 0.5[250,251].

2.8.8 Molecular replacement

Diffraction generates the Fourier Transform of an object and through its reversion, the structure of the object can be recalculated, that is, the electron density function is calculated via a Fourier Transform of the collected diffraction data of a crystal, specifically, of the structure factors. This process yields a three-dimensional map of the distribution of electron densities in the asymmetric unit. Therefore, one must work backwards from the diffraction pattern to obtain the electron density map. The process of converting the reciprocal space-representation of the crystal into an interpretable electron density map is known as phasing. However, the phase angle of reflection not contained in the image represents what is commonly known as the 'phase problem'.

2.9 ICP-MS (Inductively Coupled Plasma Mass Spectrometry)

ICP-MS is mass spectrometry-based elemental analysis technique used for determining trace elements and is the method of choice for metal cofactor identification in protein analysis. It is highly sensitive, exhibiting a limit of detection at parts per billion levels and is able to detect many elements at lower than the parts per trillion level[252,253].

Since the primary elements constituting proteins (C,H,N,O) cannot be measured with ICP-MS due to their high ionization potentials and vast background signals, for protein analysis, instead, any ICP-detectable element associated with a particular protein is used for quantification. Sulphur is used as it is a constituent of methionine and cysteine thus naturally present in proteins. The ICP-MS response is directly proportional to the mass concentration of the detected element and is therefore independent of protein concentration, molecular weight, structure or charge state. As such, the elemental quantity must then be converted to protein abundance using the stoichiometry of sulphur to protein which requires knowledge of the number of sulphur-containing amino acid residues in the protein sequence. It is also necessary to correlate the sulphur signal to concentration by standardization. In addition, other protein-associated elements can be measured, such as co-ordinated metal co-factors in metalloproteins and elements incorporated during chemical modification of a protein[252,253].

In short, an ICP-MS instrument uses an atmospheric pressure, high temperature argon plasma to ionize the sample before measurement in a mass spectrometer. Seven fundamental components comprise an ICP-MS procedure: first, the sample introduction system containing a nebuliser which produces a fine aerosol mist from a liquid sample; the inductively coupled

plasma then converts elements in the sample to ions; an interface extracts these ions into the vacuum chamber; a set of electrostatic lenses (the ion optics) focusses the ion beam and resolves analyte ions from background signals; a collision reaction cell (CRC) filters the analyte ions from interfering ions; the quadrupole mass analyser separates the analyte ions by mass; and finally, the electron multiplier detector to convert the response into an electrical signal. There is also a scanning quadrupole mass filter between the ion lens and the CRC. A triple quadrupole ICP-MS (ICP-QQQ-MS) includes an additional quadrupole mass filter to lower detection limits[252,253].

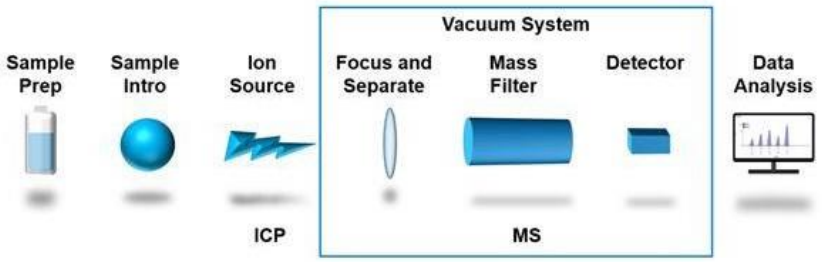


Figure 2.6 Overview of the main components and stages of ICP-MS. Image from Agilent.com.

CHAPTER THREE

Investigations of the basis for the specificity of HP2 phytases

3.1 Chapter Introduction

As introduced in chapter 1.3.6, MINPPs are a distinct sub-family of HAPhy enzymes with conserved structures which are homologous in sequence and structure to HAPhys. However, where HAPhys are stereospecific phytases, MINPPS display positional hydrolytic flexibility whilst retaining catalytic efficiency[92,100]. The detection in ileal digesta of raised levels of IP4 and IP3 revealed that cleavage of the first phosphate is not the sole limiting step in IP6 degradation[188,254]. It is now well established that it is this positional specificity of commercial HAPhys which results in accumulation of IP4 intermediates during hydrolysis of IP6[158]. The catalytic flexibility of MINPPs could therefore be exploited in the design of a next-generation phytase derivative which overcomes the bottleneck of IP4 accretion[100,158]. The MINPP from *Bacteroides thetaiotaomicron*, *Bt*MINPP, is one such potential candidate for engineering existing commercial phytases.

Originally isolated from human faecal matter[255], *B. thetaiotaomicron* is a prevalent bacterial endosymbiont of human gut flora – the second most common species isolated[256] - and it has previously been discovered that it expresses a MINPP enzyme – *Bt*MINPP - homologous to that of the human cell-signalling MINPP1, and which is enclosed and secreted in outer membrane vesicles[92]. This feature was proposed to be a protection mechanism of the enzyme from proteolysis by gastrointestinal enzymes and in addition, their delivery to human colonic epithelial cells was revealed to promote intracellular Ca^{2+} signalling[92]. *Bt*MINPP, like human MINPP1, and distinctive of MINPPS in general, has the ability to attack IP6 at various positions, and also like mammalian MINPPs, has increased catalytic activity towards the 3-phosphate from Ins(1,3,4,5)P4, as well as exhibiting very high catalytic activity in general[92]. The pH optima of *Bt*MINPP is consistent with the pH profile of the GI tract of its host – 2.5, 4.0 and 7.5 – indicating its evolutionary adaptation to the conditions provided by the host[92]. Published in 2014, and representing the first structure of a MINPP, the structure of *Bt*MINPP in complex with the non-hydrolysable persulfated IP6 analogue (HIS; PDB:4FDU) facilitated the explanation of how this enzyme is able to permit alternate orientations of IP6[92]. Like other HAP members, *Bt*MINPP folds into a structurally conserved α/β domain (also referred to as the core domain)

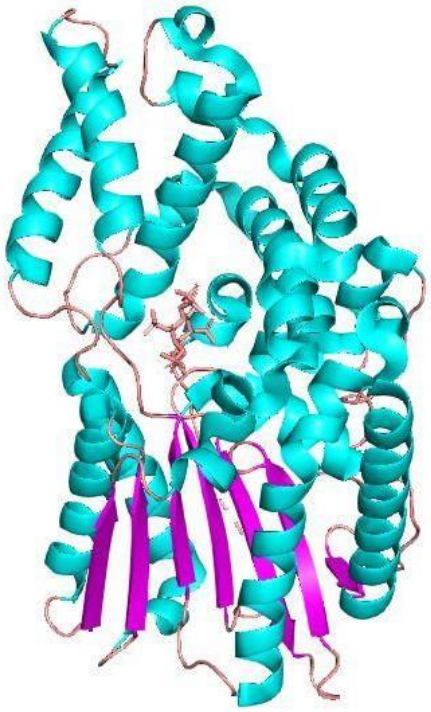


Figure 3.1 Cartoon representation of the structure of *Bt*MINPP, coloured by secondary structure (α-helices cyan, β-strands magenta, loops coral). Oriented with the α/β domain at the bottom and the α domain at the top. Image rendered in Pymol.

and a more variable α domain (referred to as the cap domain), with the active site located in the cleft between the two domains, which are anchored together by disulphide bridges (Fig 3.1)[79]. The active site is comprised predominantly of basic amino acids to facilitate acceptance of the highly charged IP6, forming six phosphate specificity pockets which accommodate the six phosphate groups of the substrate (Fig 3.2). The catalytic triad responsible for hydrolysis is positioned in pocket A and includes the nucleophilic histidine, H59, which is responsible for the generation of the phosphohistidine intermediate[92]. Consistent with the substitution of the HD dipeptide in HAPs with the HAE tripeptide in MINPPs, the glutamic acid E325 in the HAE tripeptide acts as the proton donor for the leaving group[92]. The HAE tripeptide of MINPPs is purported to provide an equivalent function to the HD proton donor motif of HAPs[92]. Comparison with the structure of *A. niger* PhyA, with which *Bt*MINPP shares close structural homology, provided justification for the lower positional discrimination exhibited with IP6 and this was determined to be a result of its larger and less polar ligand-binding pocket[92]. Indeed, the substitution of the HD motif in a classic HAP such as *An*PhyA with the HAE tripeptide is responsible for the increased volume and decreased polarity of the B and F specificity pockets of *Bt*MINPP[92,257]. An alanine instead of a tyrosine in the B and C pockets also contributes to the volume of the pocket relative to PhyA[92,257]. Identification of the amino acid residues which determine catalytic flexibility is a prerequisite for the engineering of existing phytases. Previous research to rationalise the basis of the low stereospecificity displayed by *Bt*MINPP revealed that the B-pocket is a key specificity locus, largely controlled by residues T30 and Q276[257]. The hypothesis was that changes in the composition of residues in the specificity pockets of the enzyme is the driver of deviations from wild-type stereospecificity, and this was investigated by analysis of reciprocal mutants of *Bt*MINPP/AnPhyA[92,257]. Four residues constituting the B pocket of *Bt*MINPP were identified – A30, T31, Q276 and K280 – as being determinants of positional stereospecificity (Fig 3.2)[257]. The conclusion that T31 and

Q276 are the principal positional determinants was reached by further analysis of triple mutants[257]. The F-pocket was also identified during previous studies as hosting some key residues involved in specificity[258]. R183 was exposed as being crucial for the catalytic flexibility of *Bt*MINPP[92,258]. This was determined by the altered specificity upon substitution with the equivalent *An*PhyA residue, in which the fungal-like positional preference of hydrolysis initiation was instead displayed[92,258]. Meanwhile, D186 was concluded not to contribute to specificity, and although the majority of V147 variants retained the wild-type distribution of IP5 intermediates, one substitution (V147F) did result in a change of specificity profile[258]. The overwhelming majority of variants exhibited statistically significantly different IP6 activities relative to the wild-type[258].

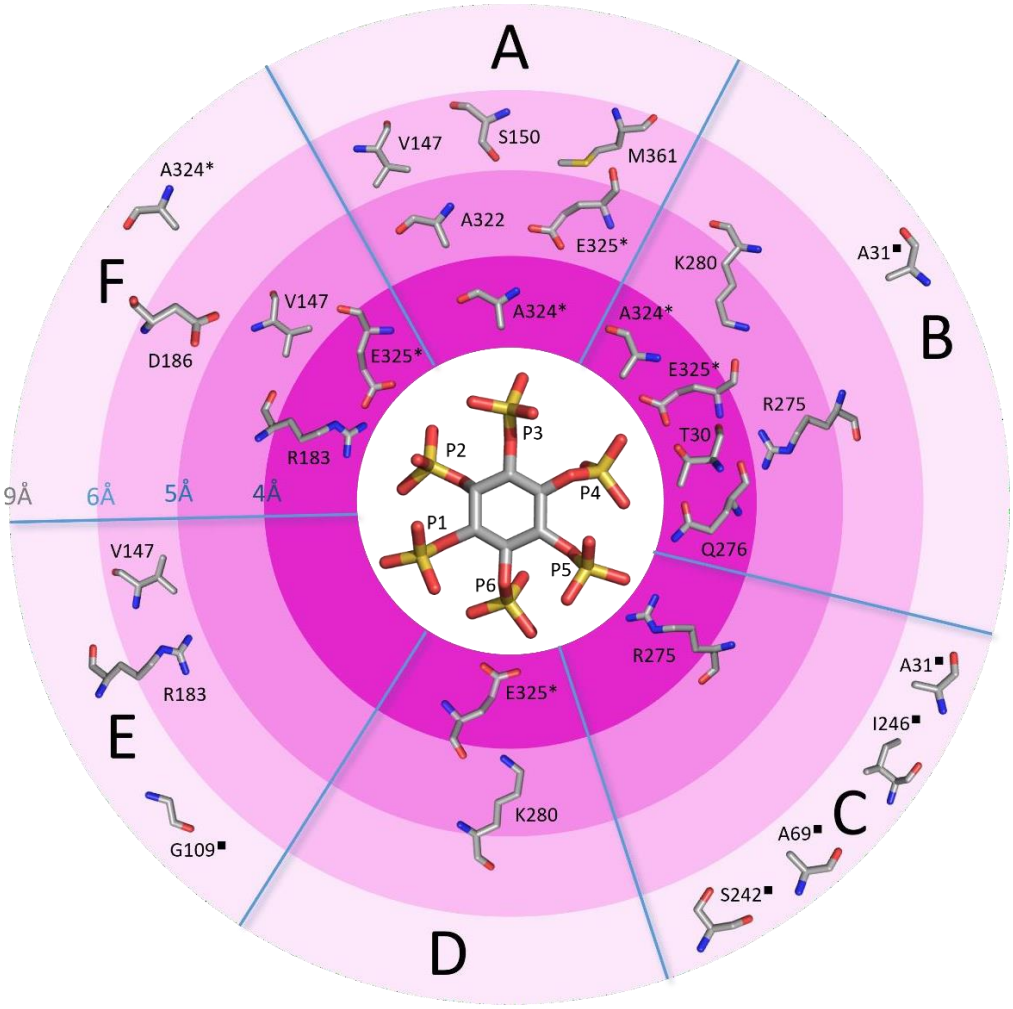


Figure 3.2 A 2D representation of the residues constituting the six pockets (A-F) of *Bt*MINPP. Pockets are labelled on the basis that pocket A contains the catalytic histidine, with the remainder of the pockets labelled according to increasing sulphate number of IHS.

3.1.2 Objectives

With the established knowledge of the individual specificity subsites, the purpose of this work was to generate different sets of *Bt*MINPP variants for the interrogation of their IP6 and IP4 activities, with particular focus on the latter. Three residues from the B-pocket were elected – R275, Q276 and K280 – and a restricted set of seven variants were generated for each. The positions of these residues relative to the substrate analogue, IHS, in the wild-type enzyme are highlighted in Fig 3.3. Based on coverage of a range of amino acid properties, the seven residues selected for substitution were leucine, alanine, serine, tyrosine, asparagine, aspartic acid and methionine.

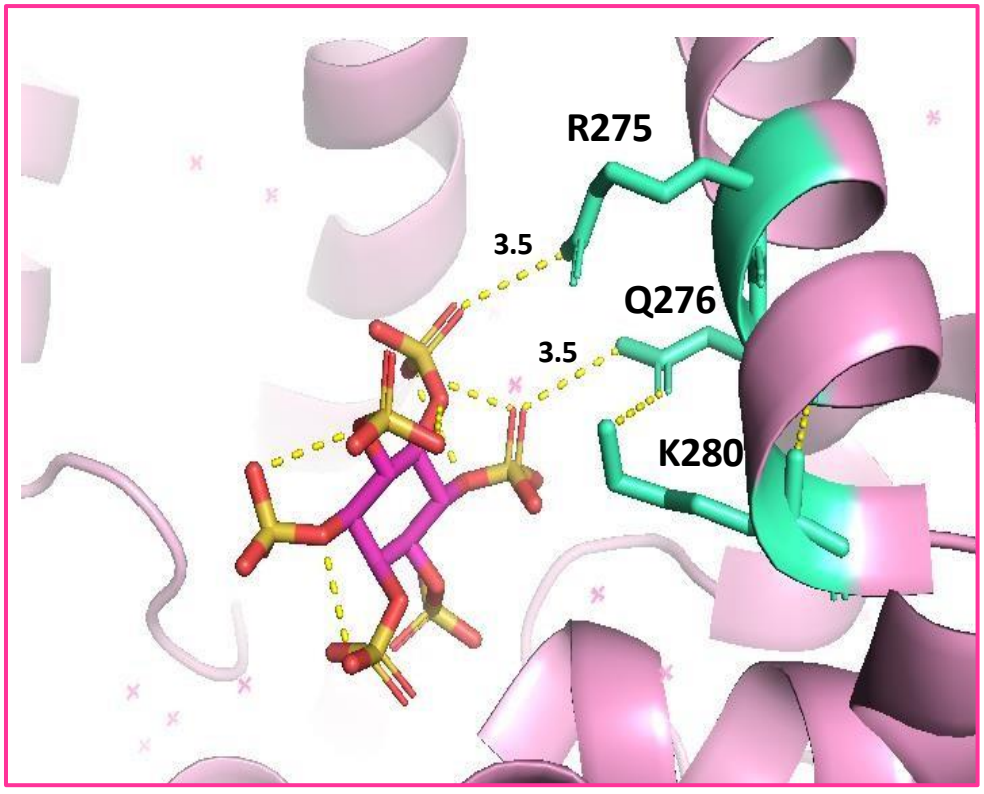


Figure 3.3 The three residues within the B-pocket of *Bt*MINPP chosen for restricted mutagenesis, with yellow dashed lines representing intermolecular interactions and distances, in Å, between each residue and IHS included.

3.2 Experimental Procedures

3.2.1 Site-directed mutagenesis of *Bt*MINPP residues R275, Q276 and K280

PCR primers were designed according to the Liu and Naismith variation of the QuikChange mutagenesis method such that each primer pair contains an overlapping region and a non-overlapping stretch in which individual primers extend in opposite directions[259]. Primer sequences are detailed in the Appendices (A3.2.1). PCR reactions of 25uL were set up as detailed in Table 3.1.

Component	Volume (µL)	Final concentration
5x Phusion HF buffer	5	1x
10uM forward primer	1	
10uM reverse primer	1	
10mM dNTPs	1	
DMSO	1	
Phusion polymerase	0.2	
Template DNA	1	
Water		

Table 3.1 Composition of the mutagenic PCR reactions set up for site-directed mutagenesis of *Bt*MINPP.

Using His-tagged *Bt*MINPP_pET15b as the template, the thermal cycling protocol consisted of 98°C for 3 min, followed by 25 cycles of [98 °C for 30 sec, 50 °C for 1 min, 68 °C for 8 min] before a final 68 °C for 10 min. The presence of a product of the correct size was evidenced by 0.8% agarose TAE gel electrophoresis (120 V, 45 min) with Gel Red staining. Negative control (no primers) lacks any band. The methylated template DNA was digested by incubation (37 °C, 2h) with DpnI (NEB; 0.5µL). A further 0.8% agarose TAE gel electrophoresis (120 V, 45 min) was run to check the correct size PCR products were still present before proceeding to transformation of XL10-Gold competent cells. Single colonies of *E.coli* XL10-Gold successfully transformed with the mutagenised PCR product were grown in LB overnight (37 °C, 180 rpm) with carbenicillin selection (100 µg/mL) in order for extraction of the propagated plasmid DNA by miniprep. Plasmid samples separated on a 0.8% agarose TAE gel were visualized with Gel Red to confirm the plasmid size before verification by sequencing.

3.2.2 Expression of *Bt*MINPP residues R275, Q276 and K280

Plasmids carrying the respective desired mutations were used to transform, by heat shock, *E. coli* Rosetta 2 (DE3) pLysS competent cells in preparation for expression. For each mutant, a single transformant was used to inoculate LB (10 mL) containing carbenicillin (100 µg/mL) and the cells were grown overnight (37°C, 180 rpm). Fresh LB (50mL) supplemented with carbenicillin (100 µg/mL) was then inoculated with the overnight culture (0.5 mL) and incubated (37°C, 180 rpm) until the OD_{600} reached circa 0.6. At this point, cells were induced by the addition of IPTG (0.1mM) before overnight incubation (20 °C, 180 rpm). Cells were subsequently harvested by centrifugation (4500 *x*g, 10 min) and frozen (-80 °C) for a brief period (1-2h) to aid lysis.

3.2.3 Purification of *Bt*MINPP residues R275, Q276 and K280

Cell pellets were resuspended (0.8 mL) in a solution of lysis buffer (Tris-HCl 50 mM, NaCl 500 mM, Imidazole 20 mM, glycerol 10% v/v and Tween-20 1% v/v, pH 8) supplemented with lysozyme (1 mg/mL), TCEP (1 mM) and half of a protease inhibitor tablet (cOmplete EDTA free). After a short period of incubation (30°C, 180 rpm, 10 min) DNase I was added (10 ug/mL) followed by a longer incubation (30 °C, 180 rpm, 1h). Cell lysates were clarified by centrifugation (21,000 *xg*, 10 min, 4 °C) and the supernatant applied to Ni-NTA superflow agarose (100 µL resin; Qiagen) pre-equilibrated with the resuspension solution (200 µL). The resin suspension was mixed in a rotary format (30 min, 4 °C) to allow binding of the His-tagged proteins to the NTA matrix. The flow-through was collected after centrifugation (700 *xg*, 2 min, 4 °C). The resin was washed twice (200 µL x 2) with wash buffer (Tris-HCl pH 8.0 50 mM, NaCl 500 mM, Imidazole 20 mM, glycerol 10% and TCEP 1mM) and the washes collected after centrifugation (700 *xg*, 2 min, 4 °C). To elute the His-tagged *Bt*MINPP mutants, elution buffer (100 µL) consisting of Tris-HCl pH 8.0 (50 mM), NaCl (500 mM), Imidazole (250 mM), glycerol (10% v/v) and TCEP (1mM) was added to the resin and this subjected to rotary mixing (10 min, 4 °C). His-tagged *Bt*MINPP mutant proteins were eluted by a final centrifugation step (700 *xg*, 2 min, 4 °C). Protein concentrations were estimated using a NanoDrop One spectrophotometer (Thermo Scientific) and purity was assessed by 4-12% Bis-Tris SDS-PAGE gel electrophoresis (Invitrogen) using MES running buffer (165 V, 40 min). Protein purity after IMAC was such that a further purification step was not necessary. For storage, glycerol was added to 30% (v/v) and protein samples flash frozen in LN2 before storage at -80 °C.

3.2.4 Estimation of relative IP6 and IP4 activities

*Bt*MINPP variants at an approximate concentration of 10 nM were incubated with IP6 or IP4 (2.5 mM) at 25 ^{o}C with a range of incubation times (50 – 120 min) to achieve ~10% of total phosphate released from the substrate. Inorganic phosphate was quantified by the molybdenum blue reaction as detailed in Chapter 2.6.

3.2.5 HPLC analysis of IP6 hydrolysis reaction products

Reaction products generated from the incubation of IP6 or IP4 (2.5 mM) with *Bt*MINPP variants (10 nM), as detailed in section 3.2.4, were separated on a CarboPac PA-200 column (two columns in series; 3x50mm and 3x250mm) which contains a resin of hydrophobic, polymeric pellicular anion exchange resin. Appropriately diluted samples of 20 µL were injected onto the column (preceded by a guard cartridge) by a sample injection pump (Jasco PU-2089 I Plus quaternary insert pump) at a flow rate of 0.2 mL/min. The elution protocol consisted of a 25 min methane sulfonic acid gradient (0-0.6M) followed by 14 min at 0.6M and finally 11 min of water, at 0.4 mL/min. The effluent was then pumped by a Jasco PU-1585 intelligent HPLC pump for mixing with a post column colour reagent (0.1% ferric nitrate and 2% perchloric acid). Detection was by a Jasco UV 1575 intelligent UV/Vis detector containing a 16µL cell. Data was processed using ChromNAV software (Jasco).

3.3 Results

3.3.1 PCR mutagenesis of three sets of *Bt*MINPP variants: R275, Q276 and K280

For the R275 and Q276 sets of variants, the PCR mutagenesis procedure was attempted and subsequently modified a number of times before success was attained. Figure 3.4 is an example of a successful gradient PCR protocol performed for five variants which had been unsuccessfully produced by previous attempts. All except the Q276D, 68°C product were successful.

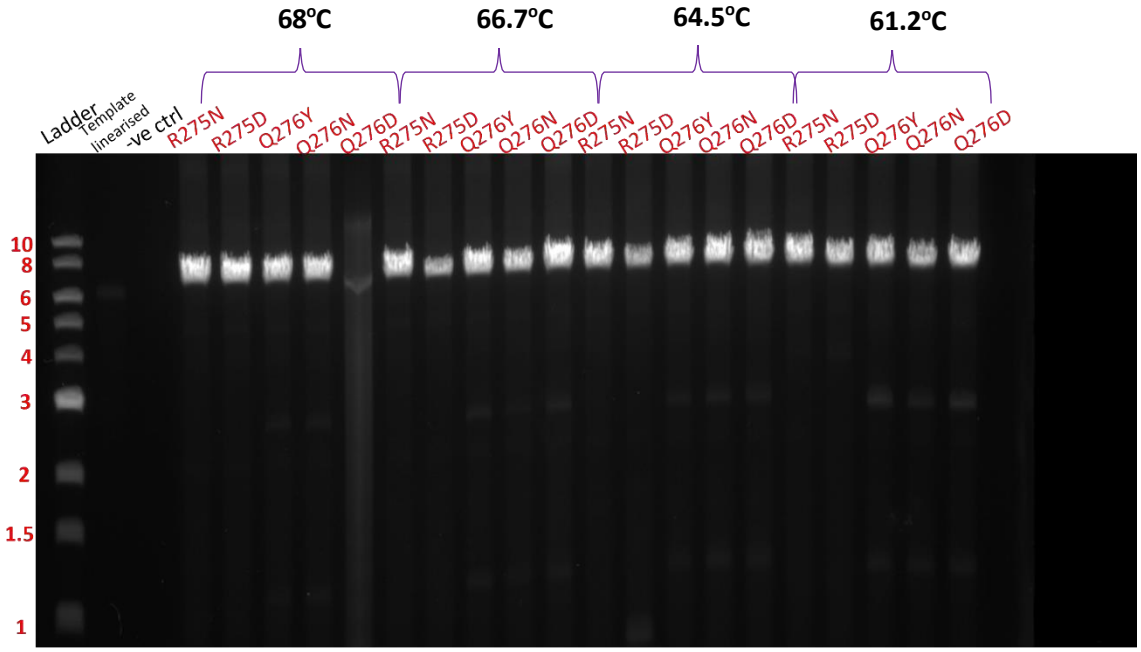


Figure 3.4 DNA agarose gel of the PCR mutagenesis products of a subset of the R275 and Q276 variants for which PCR protocols had been previously unsuccessful: R275N, R275D, Q276Y, Q276N and Q276D.

Figure 3.5 is an example of a DNA agarose gel of the set of K280 variants after successful PCR mutagenesis and transformation of *E. coli* XL10-Gold. The agarose gel of the plasmids extracted from XL10-Gold transformants reveals that one of the two colonies from which the K280M plasmid was extracted yielded an incorrectly sized plasmid therefore the other correct plasmid preparation was used for storage and downstream work. Note that K280A is absent from this particular gel, since this particular mutant necessitated additional PCR mutagenesis attempts before obtaining the desired product. All correctly sized plasmids were sent for DNA sequencing to verify the successful incorporation of the mutation.

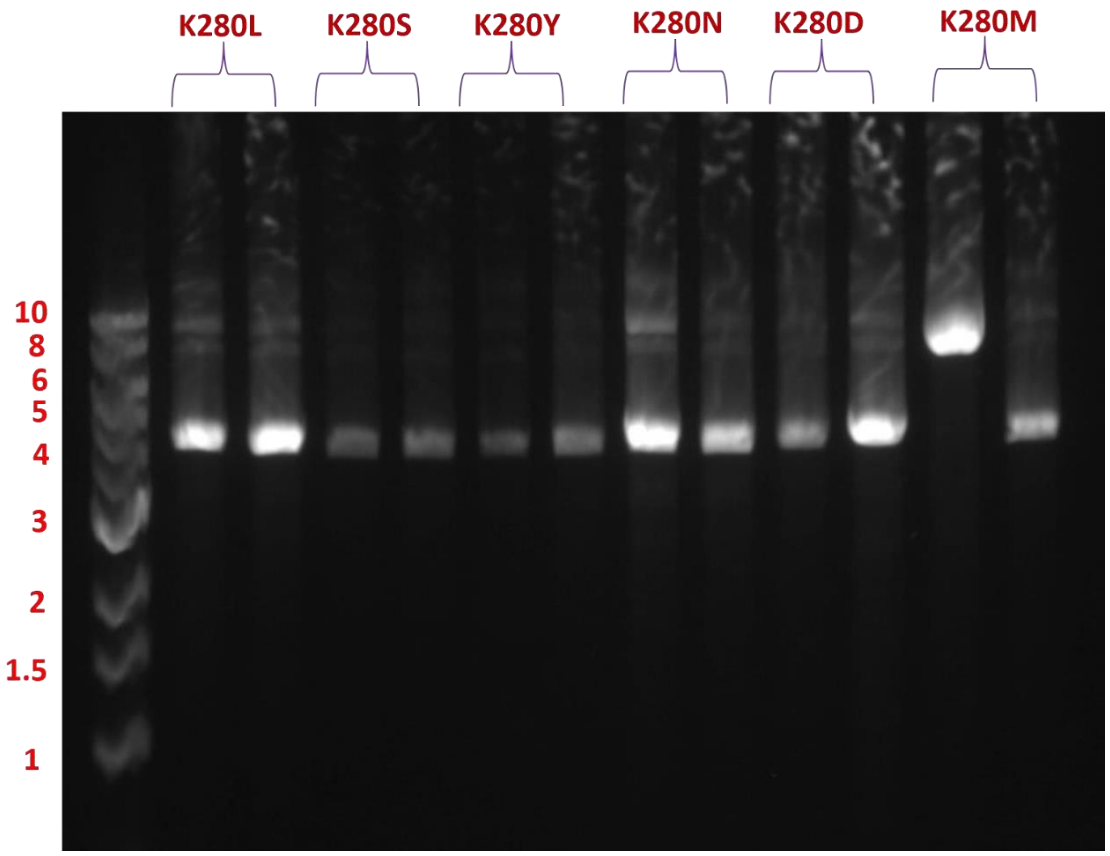


Figure 3.5 DNA agarose gel of the mutagenized plasmids extracted from XL10-Gold transformants, for the K280 set of variants. Plasmids were each extracted from two different colonies per mutant.

3.3.2 Expression and purification of R275, Q276 and K280 variants

The seven variants of each of the three *Bt*MINPP residues were expressed and purified in parallel on a small-scale by Ni-NTA affinity chromatography. This one-step purification generated enzymes of sufficient yield and purity for the purposes of the intended downstream analyses (Fig 3.6). Two of the Q276 variants, Q276S and Q276M, were not amenable to expression, despite conducting an expression trial after initial expression attempts proved unsuccessful (Fig 3.7). As such, these two variants were not continued further, and the set was consequently reduced to five variants.

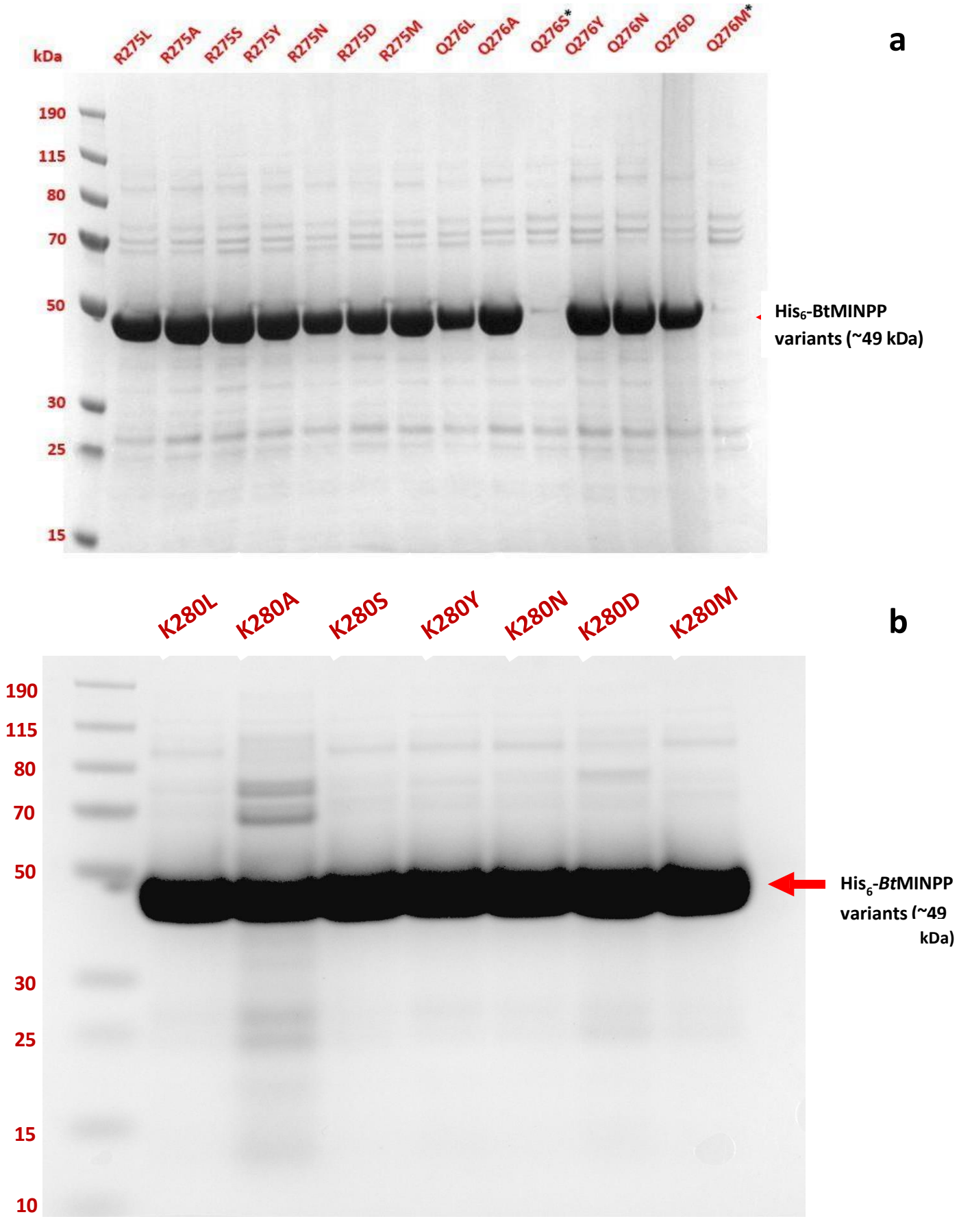


Figure 3.6 SDS-PAGE of the Ni-NTA purified *Bt*MINPP variants a) The seven variants of the R275 and Q276 sets b) The seven variants of the K280 set.

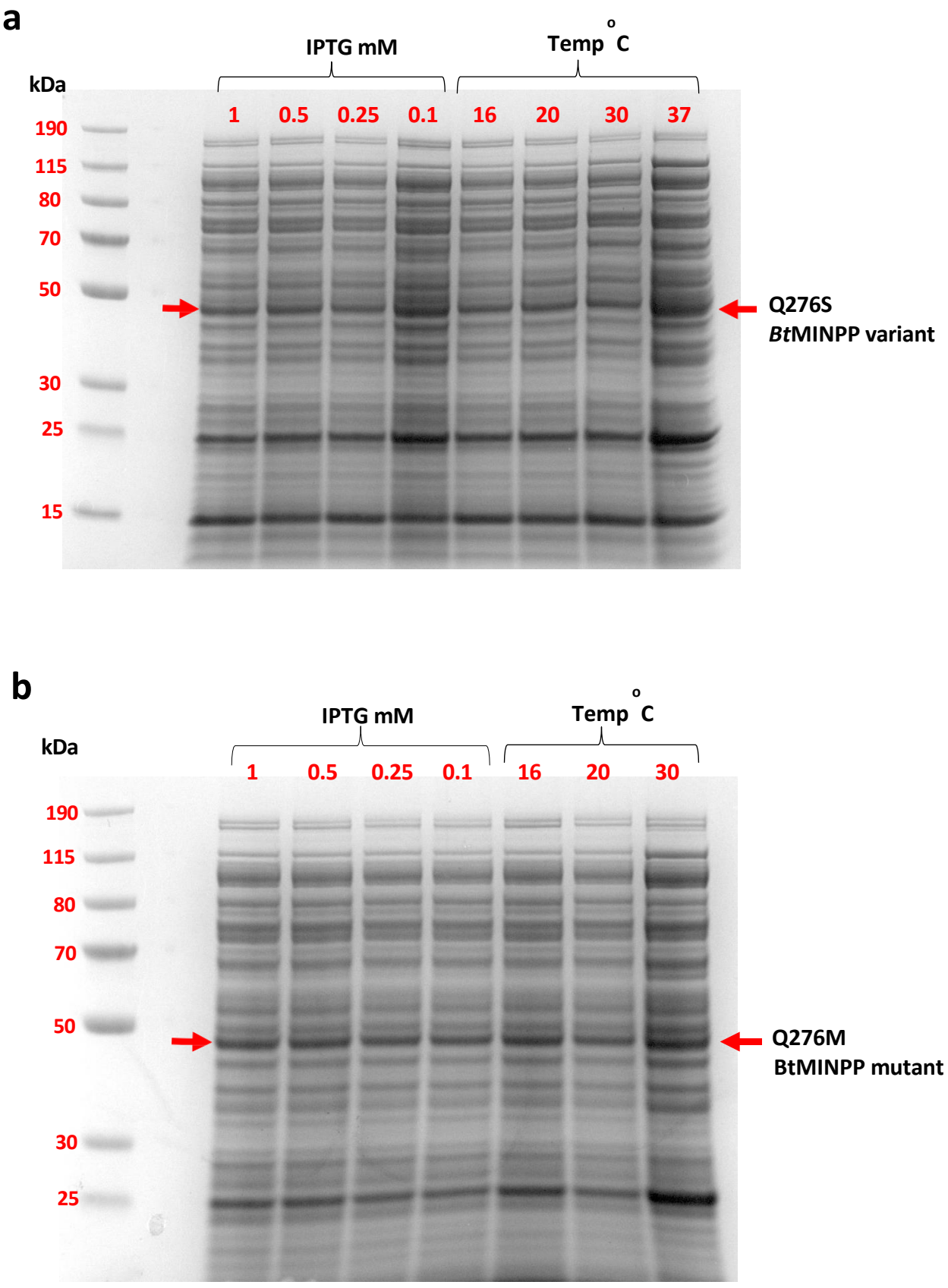


Figure 3.7 SDS-PAGE of the expression trials conducted for a) the Q276S and b) the Q276M variants of *Bt*MINPP.

3.3.3 Activity towards IP6 and IP4 of *Bt*MINPP R275 variants

Seven variants of the R275 residue of *Bt*MINPP were generated by site-directed mutagenesis before investigation of their hydrolytic activities against IP6 and Ins(2,3,4,5)P_4 relative to the wild-type and the established prototypical phytase, the periplasmic phytase AppA from *E. coli*. The residues comprising this 'restricted set' were leucine, alanine, serine, tyrosine, asparagine, aspartic acid and methionine. The results from these assays are displayed as a function of inorganic phosphate liberated from either IP6 or IP4 (Fig 3.8). Quantity of phosphate released is referred to interchangeably as 'activity' for the purposes of this chapter.

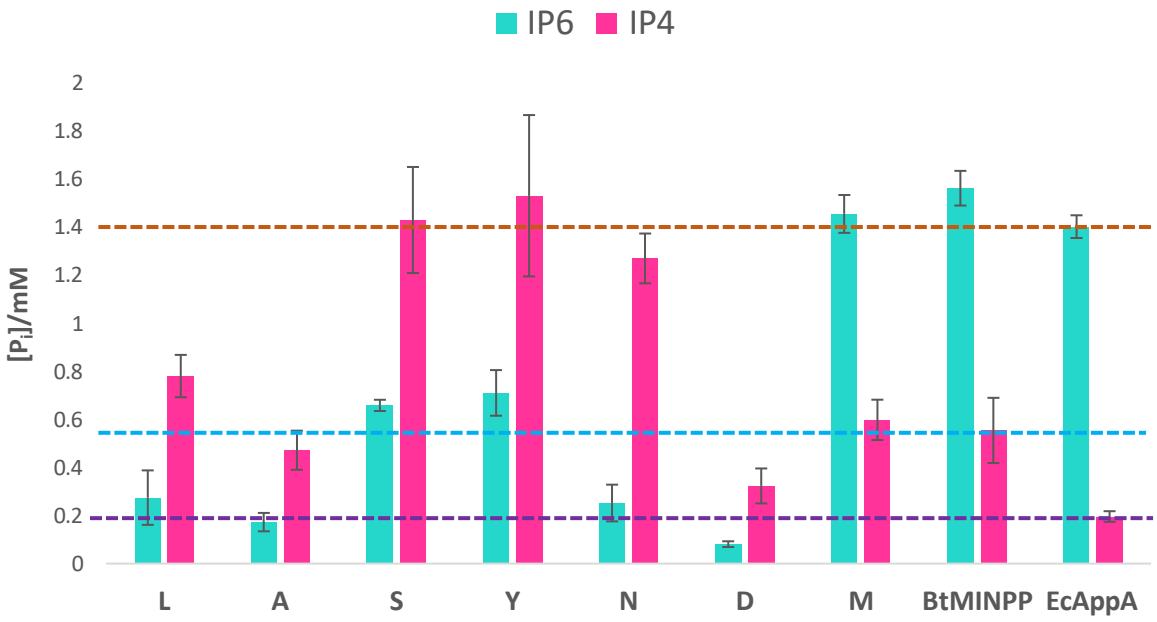


Figure 3.8 Relative levels of phosphate released from IP6 (teal) or (2,3,4,5)IP4 (magenta) by seven *Bt*MINPP R275 variants. Dashed lines highlight the level of phosphate released from IP6 by AppA (brown); from IP4 by AppA (purple) and from IP4 by *Bt*MINPP (blue). Error bars are based on standard deviation.

*Bt*MINPP and AppA exhibit rather similar activity levels towards IP6 (*Bt*MINPP marginally greater) whereas for IP4, the activity of *Bt*MINPP is approximately three times greater. Therefore several of the R275 variants display greater activities towards either IP6 or IP4 relative to AppA, since the 'baseline' is much lower, whereas the occurrence of a variant possessing greater activity than *Bt*MINPP was less frequent. Indeed, for IP4 activity, all variants were more active than AppA. Both *Bt*MINPP and AppA are considerably more active against IP6

than IP4, with *Bt*MINPP being almost three times - and AppA more than seven times - more active towards IP6 than IP4. Interestingly, all of the R275 mutants – except for R275M – have considerably greater activity towards IP4 than IP6. The methionine variant was almost identical to the wild-type in terms of the phosphate released in this assay. Indeed, it was the only variant of the set which exceeded the activity of both the wild-type and AppA in terms of IP6 activity. Although arginine and methionine are spatially similar, their chemical properties are very different. Whilst arginine is a highly polar, positively charged residue – often found in the active sites of proteins which bind phosphorylated substrates – methionine is a hydrophobic residue not usually involved in the active site chemistry of enzymes. The similarity in enzyme activity would be expected for an arginine to lysine mutant, since these two residues have similar chemical properties, but is interesting that the substitution to methionine had little effect. When the positively charged arginine was mutated to the negatively charged aspartic acid, activity against IP6 was essentially abolished (~5% of WT activity; Fig 3.9) and for IP4 it was reduced compared to the wild-type. This could be rationalised by the fact that this is considered a radical mutation. Substitution with alanine resulted in activity data moderately similar to that for R275D, albeit with higher values for both IP4 and IP6 activity. IP6 activity was just 11% relative to the wild-type (Fig 3.9). The R275S and R275Y variants exhibited, relative to one another, very similar activities towards both substrates, although the standard deviation (represented by error bars) for the IP4 values is very large. The values produced by R275S and R275Y with IP6 were similar to that of IP4 by the wild-type, and vice versa – the values produced by R275S and R275Y with IP4 were similar to that of IP6 by the wild-type. Of the variants, R275N has the highest ratio of IP4:IP6 activity, with IP4 activity five times greater than that towards IP6. R275L and R275N displayed similar IP6 activities (16% and 17% relative to WT; Fig 3.9) but IP4 activity was much greater in the latter. It is very interesting that the wild-type and AppA activities towards IP6 were much greater than for IP4, yet for the mutants, in general the IP4 activity was much greater.

Whilst the IP4 in these assays was the isomer produced by Quantum Blue™ (QB) - the Ins(2,3,4,5)P_4 -activities of the R275 variants against a different IP4 isomer, the (1,4,5,6) IP4 produced by the fungal PhyA phytase, were also determined. A comparison of the QB versus PhyA IP4 activities are displayed in Fig 3.10. The wild-type is 3.7 times more active against the IP4 produced by Quantum Blue than it is the IP4 from PhyA. For all the R275 variants, the activity levels are higher for QB IP4 than PhyA IP4, except for R275A, however this is not significant as the error bars overlap such that the highest error bar value of QB IP4 is higher than the lowest error bar value of PhyA IP4. R275S and R275N exhibit activities against QB IP4 much greater than the wild-type (2.6 and 2.3 times greater, respectively), but almost no activity

against PhyA IP4. R275Y is similar with regard to QB IP4 activity (2.8 times greater relative to the wild-type), but displays a moderately greater activity against PhyA IP4 than the wild-type (1.7 times greater). Levels of QB versus PhyA IP4 activity are very similar for the R275M variant, and these levels are around the same value as for the wild-type QB IP4 activity. R275L is approximately the same as for R275M for PhyA IP4 activity, but is slightly more active (1.4 times) than the wild-type for QB IP4. R275D results in the lowest QB IP4 activity (1.7 times less than the wild-type) and almost no PhyA activity.

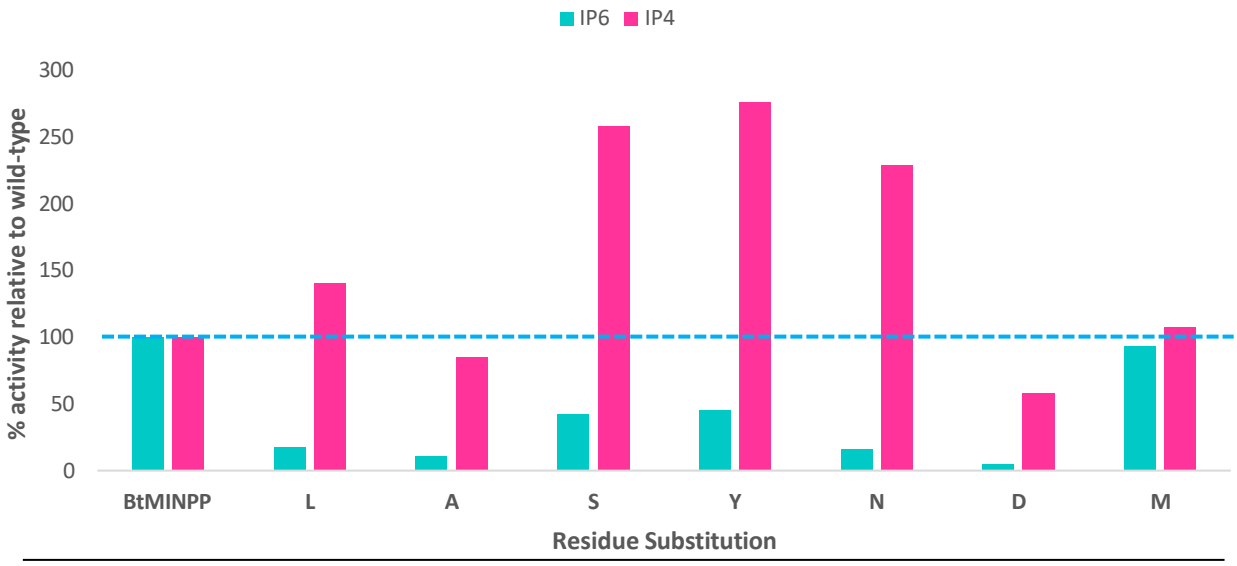


Figure 3.9 Activities against IP6 and IP4 of the set of *Bt*MINPP R275 variants displayed as a percentage relative to the wild-type activity.

There are several metrices which have been developed over the years in attempts to quantitate the effect of different amino acid substitutions on the function of a protein. These are based on quantitating different properties of the amino acids. One of the most established of these is the Grantham's distance, which attempts to predict the evolutionary distance between two amino acids[260]. Grantham scores range from 5 (leucine/isoleucine) to 215 (cysteine/tryptophan), with a lower score reflecting a smaller evolutionary distance and vice versa. The greater the distance between two amino acids, the less likely they are to be substituted by one another naturally, because the effect would be detrimental. The amino acid properties that Grantham's distance is based on are composition, polarity and molecular volume[260]. An exchange between two amino acids separated by a large physiochemical distance is thereby considered a radical substitution and a smaller psychochemical distance a conservative substitution. The set of Grantham scores are tabulated in the original publication, reference #260. Based on the

Grantham score, within the set of R275 variants, substitution by alanine represents the largest physiochemical distance, and tyrosine the smallest (Table 3.3.1). R275A does result in the second lowest (R275D being the lowest) percent activity relative to the wild-type, so correlates fairly well with the Grantham score, although there is no correlation in the other variants.

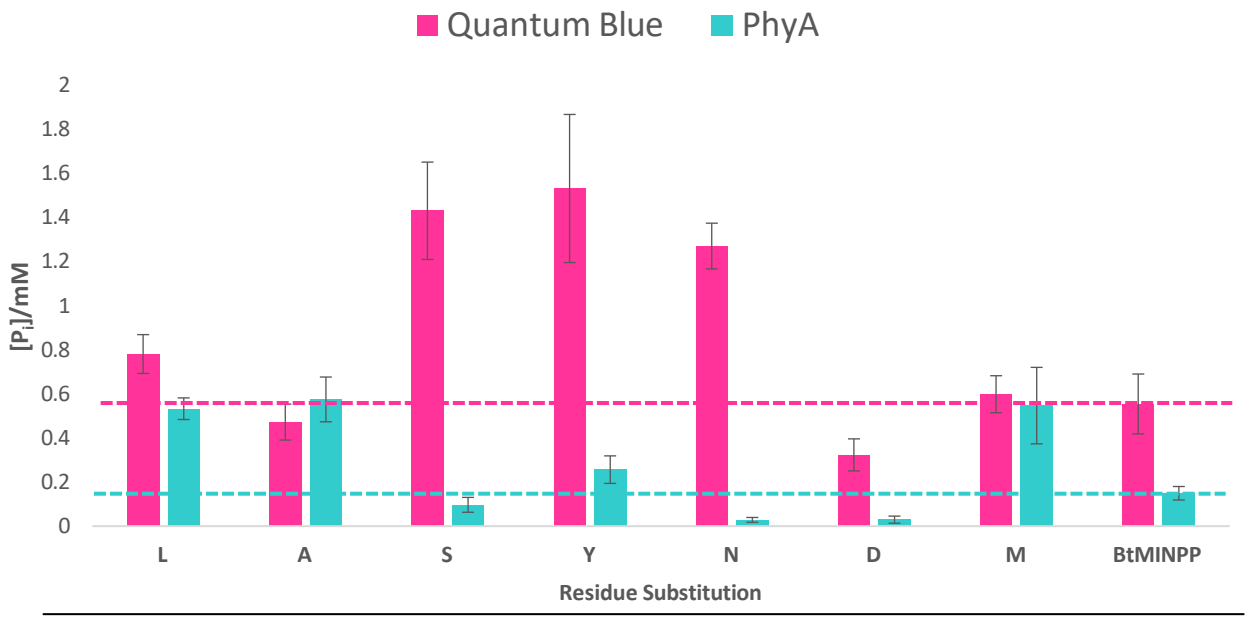


Figure 3.10 Relative levels of phosphate released from QB IP4 (magenta) or PhyA IP4 (teal) by seven *Bt*MINPP R275 variants. Dashed lines highlight the level of phosphate released by *Bt*MINPP from QB IP4 (magenta) and PhyA IP4 (teal), respectively. Error bars are based on standard deviation.

Whilst the Grantham's distance is based on an evolutionary perspective, Yampolsky and Stoltzfus developed the measure of Experimental Exchangeability which evaluates the mean effect of amino acid exchange on protein activity *in vitro,* based on the analysis of 9671 reported amino acid exchanges in 12 different target proteins[261]. The table of exchangeability scores is published in the original article, reference #261. Based on experimental exchangeability, overall, alanine was determined to be the best replacement for other residues, and lysine the most easily replaced. According to the experimental exchangeability matrix, within the set of R275 variants, alanine is the most exchangeable amino acid, and methionine and asparagine

are the least. However, the value for methionine and asparagine are based on the mean of just one and three, respectively (Table 3.2). Therefore these values cannot be awarded too much weight.

Sneath's index measures the overall resemblance between the 20 amino acids, taking into account 134 categories of physiochemical properties[262]. The result is a value of dissimilarity – the dissimilarity index, Ds[I,j] – which is calculated as the percentage of the sum of all properties not shared between two amino acids (Fig 3.11)[262]. According to Sneath's index, the similarity of residues (relevant to the mutagenesis of R275) to arginine, in increasing order of dissimilarity are: methionine > leucine and asparagine > alanine, serine, aspartic acid and tyrosine (Table 3.2). Methionine being closest in similarity to arginine, out of the seven residues in question, correlates well with the results in Fig 3.8. Indeed, except for the serine and tyrosine substitutions, the results presented here reflect the Sneath's index, depicted in Fig 3.11b. It could therefore be predicted that exchanging arginine for lysine would result in even less of a difference in IP6 and IP4 activity.

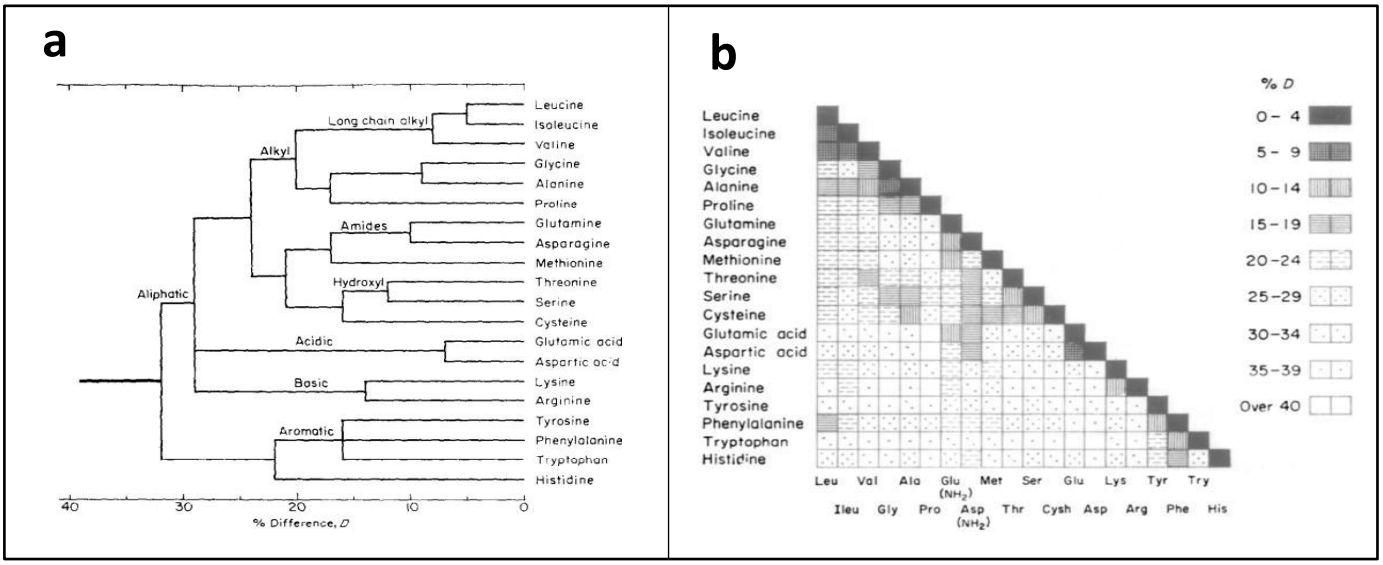


Figure 3.11 a) Dendrogram of amino acid relationships based on Sneath's measure of dissimilarity. b) Sneath's matrix of resemblance between amino acids. Darker shades represent close resemblance, and fully black squares represent each amino acid compared with itself i.e complete resemblance. The dendrogram in (a) is essentially a summary of, and complementary to, (b). Image from Sneath[262].

Where Sneath's index is based on a large, unweighted number of chemical and physical properties of amino acids, Epstein's coefficient of difference is based on just size and polarity, so it is not unexpected that the coefficients vary considerably between these two measures[263]. The assumptions are that the polarity of an amino acid side chain is the most critical determinant of its effect on protein conformation; and that the size of a side chain may influence conformation, particularly if a larger residue is replaced by a smaller one in the protein interior. There are two different equations to distinguish between two different cases:
1) when a smaller hydrophobic residue is replaced by a larger hydrophobic or polar residue;
2) when a polar residue is exchanged or when a larger residue is replaced by a smaller one.
The coefficients of difference values were implemented by Epstein to calculate indices of difference for amino acid exchanges between homologous proteins[263]. The coefficients can be located in the original article, reference #263.

Similarly to Epstein's coefficients of difference, Miyata's distance is based on amino acid volume and polarity and the author states that these are the primary representative property parameters and consequently are the key determinants for protein conformation[264]. Miyata's method is essentially a modification of Grantham's physiochemical distances. The distance ranges from 0.06 for the most similar pair, alanine and proline, to 5.13 for the least similar pair, glycine and tryptophan and these are reported in the original article in reference #264. Miyata also categorised the amino acids into six well defined groups by distance and this is tabulated in Table 3.3. For the set of R275 variants, all seven pairs have moderately similar Miyata distances, ranging from 2.02 (tyrosine) to 2.92 (alanine). This is in accordance with the Grantham distances.

Substitution	Grantham Score	Exchangeability Score	Sneath's Index (% *D*)	Epstein's coefficients of difference	Miyata's pair distance (D)
R > L	102	0.24 (43)	30-34	1.00	2.62
R > A	112	0.46 (44)	35-39	0.62	2.92
R > S	110	0.27 (46)	35-39	0.24	2.74
R > Y	77	0.27 (30)	35-39	0.80	2.02
R > N	86	0.07 (3)	30-34	0.08	2.04
R > D	96	0.12 (4)	35-39	0.08	2.34
R > M	91	0.07 (1)	25-29	1.00	2.29

Table 3.2 Amino acid difference 'scores' based on five different methods for the seven arginine to L/A/S/Y/N/D/M residue exchanges. N.B In the column Exchangeability Score, in parentheses is the number from which the mean was derived.

Included in Table 3.2 are the amino acid difference scores for the seven substitutions of arginine 275 relevant to this section, according to the five methods discussed previously. They cannot be directly compared with one another as they are based on completely different functions.

To assess the effects, if any, on specificity, reactions were subjected to HPLC to reveal the relative proportions of the different IP_5 isomers. As a pre-requisite, the variants were each incubated with IP6 for varying lengths of time before determination of phosphate released by the molybdenum blue assay. This facilitated the estimation of reaction time required to generate an appropriate IP6 degradation profile i.e. one in which the IP_5 isomers were visible. If the degradation does not proceed far enough, only the IP6 peak would be present, and conversely, if the degradation proceeds too far, the IP_5 peaks will be absent as lower inositol phosphate peaks appear. The percentage of IP6 degradation aimed for was approximately 10%. The HPLC traces for the seven R275 *Bt*MINPP variants are presented in Fig 3.12, and the analogous IP_5 isomer compositions in Fig 3.13. Interestingly, all of the variants generated a larger proportion of the 1/3-OH IP_5 isomer, whereas the major isomer of the wild-type is the 5-OH isomer. It is possible that this has a link to the much higher IP4 versus IP6 activity observed in these variants. For all the variants, the percentage represented by the 1/3-OH isomer ranged from 43-51% with a mean of 46%, whilst in the wild-type this was just 23%. The 4/6-OH isomer proportion was consistent across the variants, ranging from 21-24%, and this was similar to the wild-type (27%). Whilst the wild-type generated 50% 5-OH IP_5, the variants yielded 27-35%, with an average of 31%. Therefore, the specificity has flipped from 5- OH to 1/3-OH IP_5 in these variants.

Group					
1 (special case)	cysteine				
2 (small & neutral)	Proline	alanine	glycine	serine	threonine
3a (hydrophilic & relatively small)	glutamine	glutamic acid	asparagine	aspartic acid	
3b (hydrophilic & relatively large)	histidine	lysine	arginine		
4a (hydrophobic & relatively small)	valine	leucine	isoleucine	methionine	
4b (hydrophobic & relatively large)	phenylalanine	tyrosine	tryptophan		

Table 3.3 The six groups of amino acids as defined by Miyata[264]

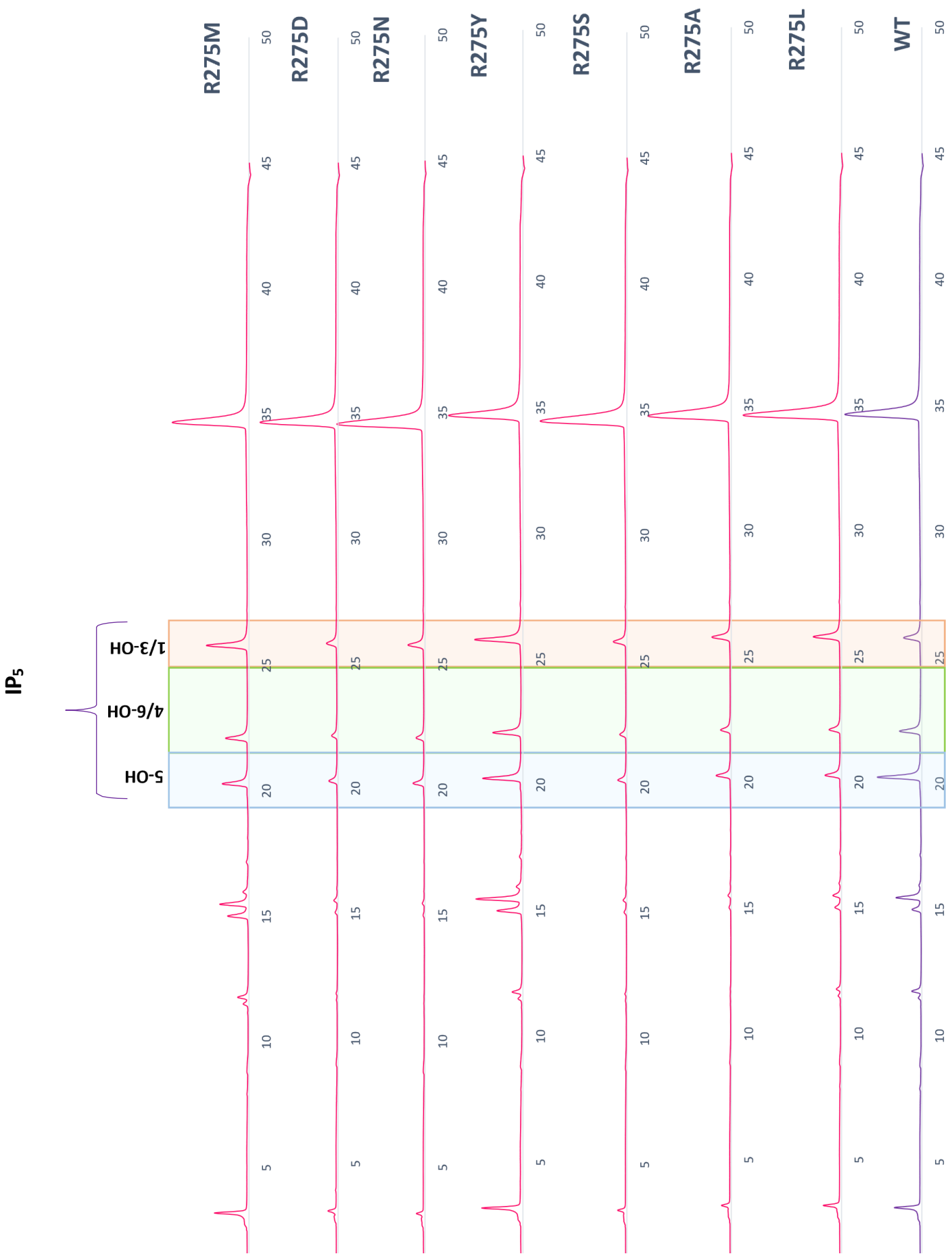


Figure 3.12 HPLC traces of the seven R275 variants (magenta) compared to the wild-type *Bt*MINPP (purple). The IP5 regioisomers generated are highlighted by a coloured box: 1/3-OH in orange, 4/6-OH in green and 5-OH in blue.

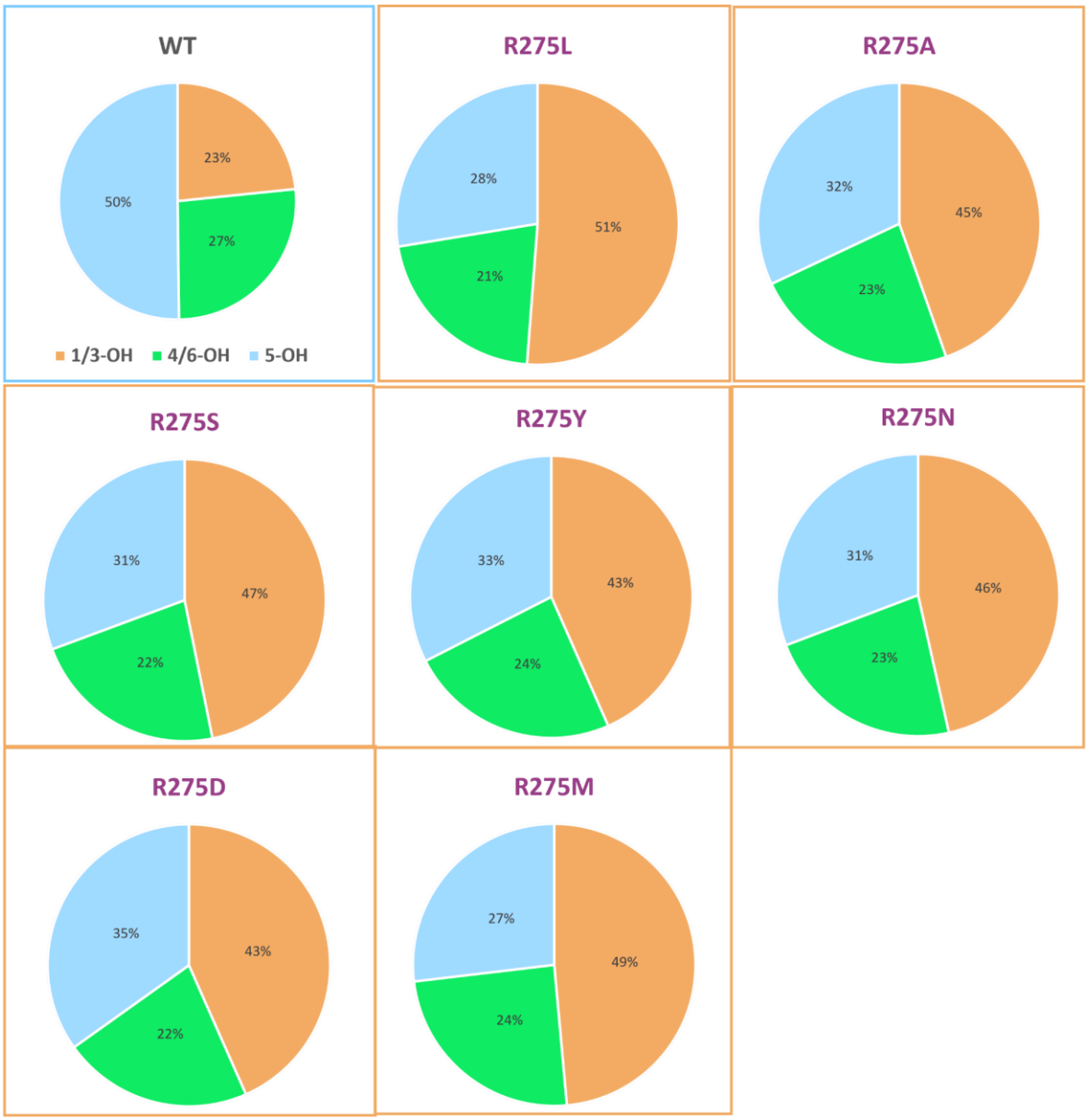


Figure 3.13 The proportion of the three different IP5 regioisomers generated by the set of R275 *Bt*MINPP variants. The 5-OH isomer is represented by blue, the 1/3-OH isomer by orange and the 4/6-OH isomer by green.

As a visual accompaniment, the R275 position was mutated in Pymol to each of the substituting residues (Fig 3.14).

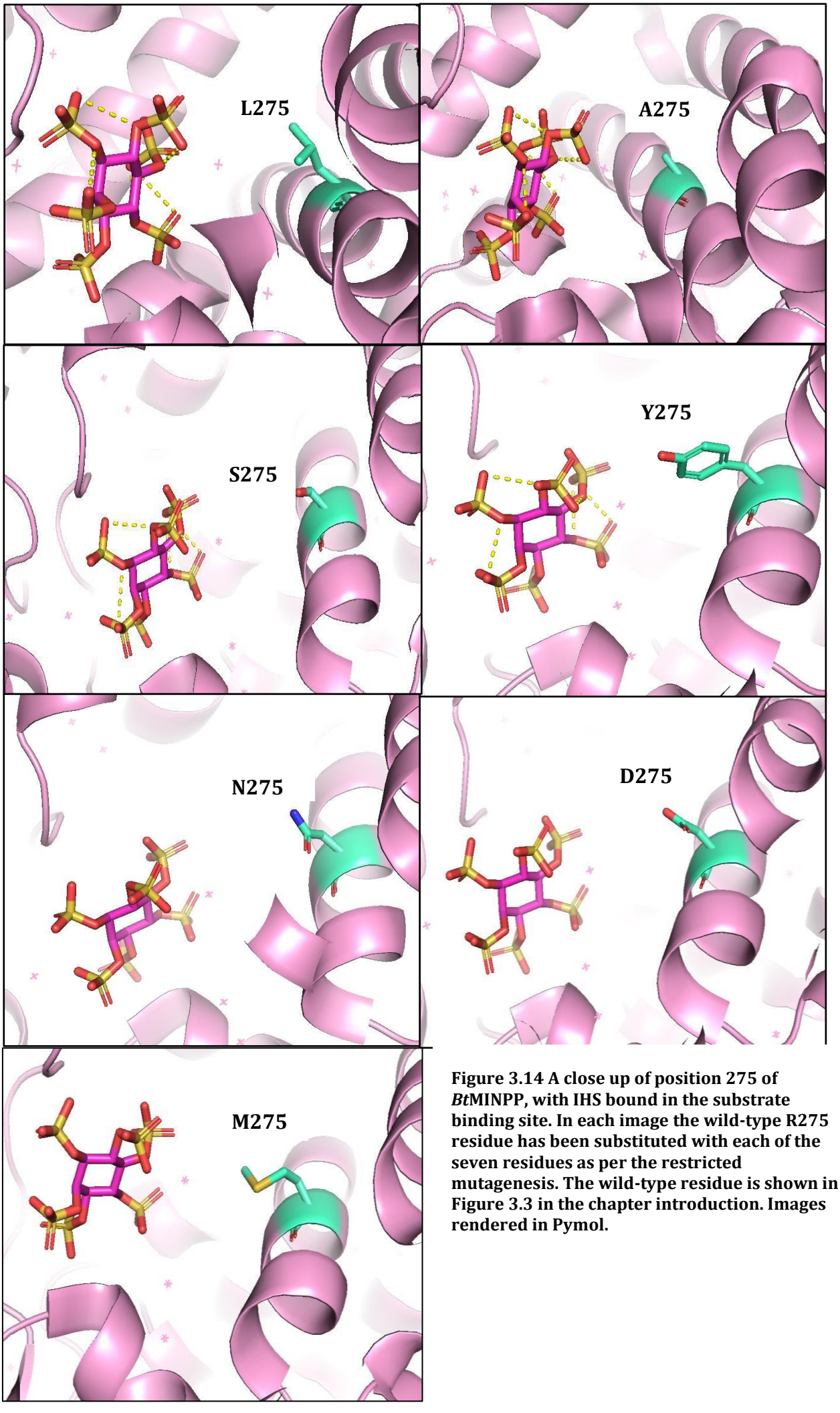


Figure 3.14 A close up of position 275 of *Bt*MINPP, with IHS bound in the substrate binding site. In each image the wild-type R275 residue has been substituted with each of the seven residues as per the restricted mutagenesis. The wild-type residue is shown in Figure 3.3 in the chapter introduction. Images rendered in Pymol.

3.3.4 Activity towards IP6 and IP4 of *Bt*MINPP Q276 variants

Activities of the Q276 variants against IP6 and Ins(2,3,4,5)P_4 are presented in Figure 3.15.

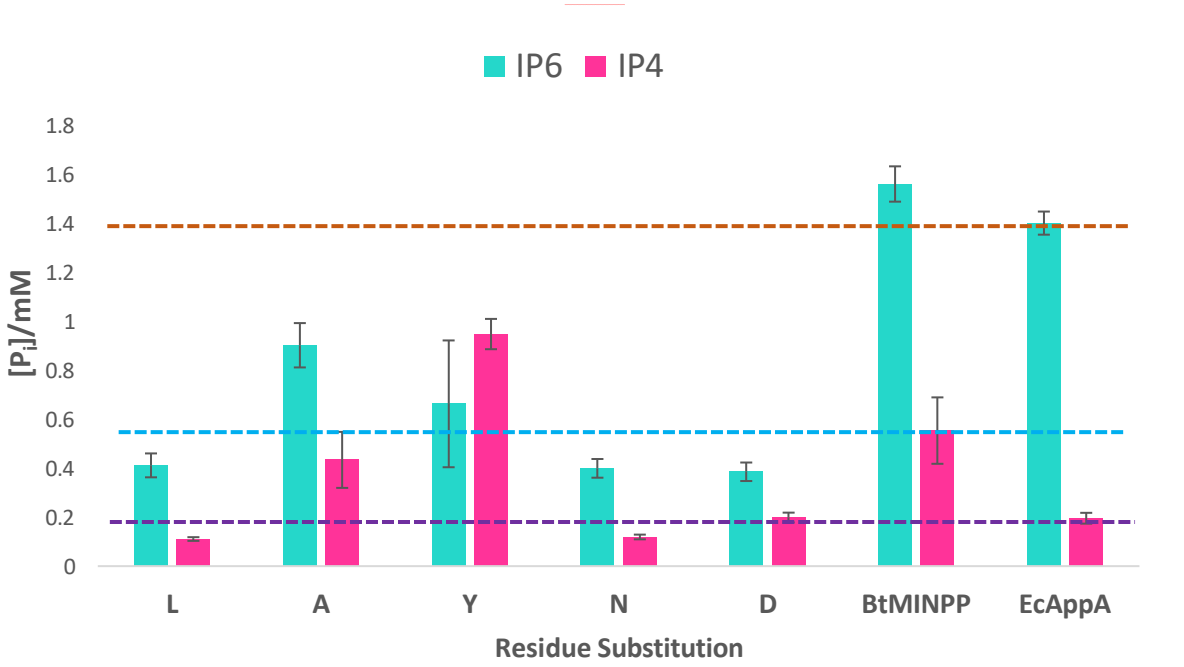


Figure 3.15 Relative levels of phosphate released from IP6 (teal) or IP4 (magenta) by five *Bt*MINPP Q276 variants. Dashed lines highlight the level of phosphate released from IP6 by AppA (brown); from IP4 by AppA (purple) and from IP4 by *Bt*MINPP (blue). Error bars are based on standard deviation.

Overall, levels of activity against both IP6 and IP4 are much lower than for the R275 variant set, with all levels significantly lower than the wild-type and AppA. Interestingly, whilst all except one of the seven R275 variants displayed greater IP4 than IP6 activity, for the Q276 set, for all except one mutant – Q276Y – the activity against IP6 was higher than for IP4. Further, the Q276Y variant was the only one for which IP4 activity out-performed that of the wild-type (and AppA). The tyrosine substitution also resulted in the highest IP4 activity in the R275 set. Q276L and Q276N were even less active than AppA for IP4 activity, whilst in the R275 variants, all were substantially more active than AppA – the lowest level still being 1.65 times greater than AppA (R275D). The L, N and D substitutions of Q276 exhibit very similar activity profiles for both IP6 and IP4, with 26.4, 25.6, and 24.7% IP6 activity relative to the wild-type, and 20.0, 21.5, 35.9% IP4 activity relative to the wild-type (Figure 3.16).

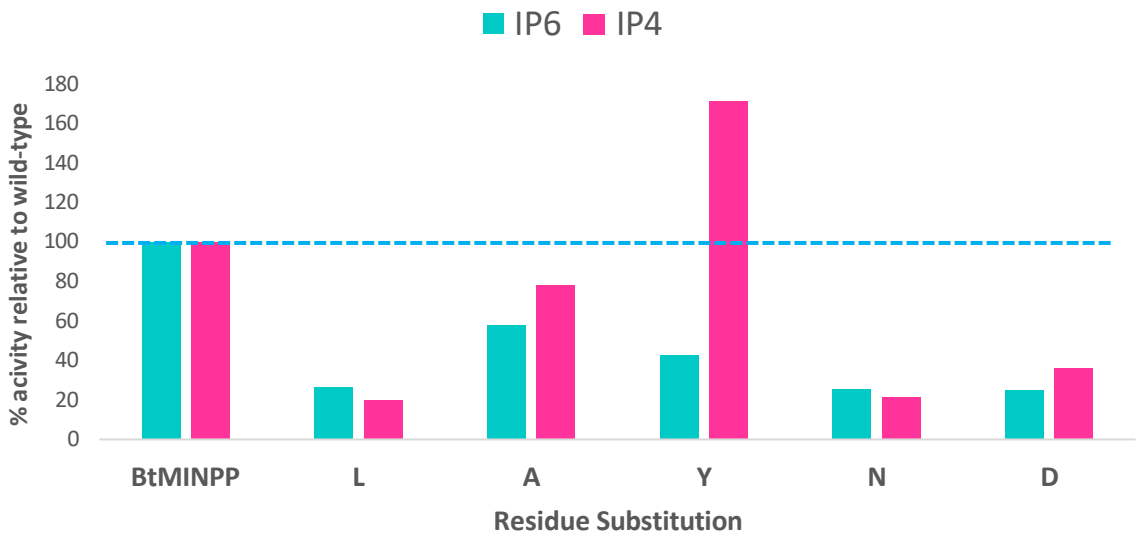


Figure 3.16 Activities against IP6 and IP4 of the set of *Bt*MINPP Q276 variants displayed as a percentage relative to the wild-type activity.

Comparison of the activities against QB IP4 and PhyA IP4 is represented in Figure 3.17. Q276Y is the only variant of the set which outperforms the wild-type in activity against QB IP4, with 1.7 times greater activity. It is also the only variant of this set for which the QB IP4 activity is greater than the corresponding PhyA IP4 activity (2.8 times greater) and is therefore similar to the wild-type in terms of IP4 preference. Unlike the previous set of variants, levels of PhyA IP4 activity are all higher than the wild-type. Q276L displays the greatest difference between the two IP4 substrates, with 3.8 times greater activity against PhyA IP4 than QB IP4. Q276D has the least difference, with PhyA IP4 activity only 1.5 times greater than QB IP4 activity, and it's PhyA IP4 activity is also the lowest of the set, excluding the wild-type. Q276A exhibits the greatest PhyA IP4 activity, at 4.1 times greater than the corresponding wild-type activity, although the standard deviation is rather large for Q276A.

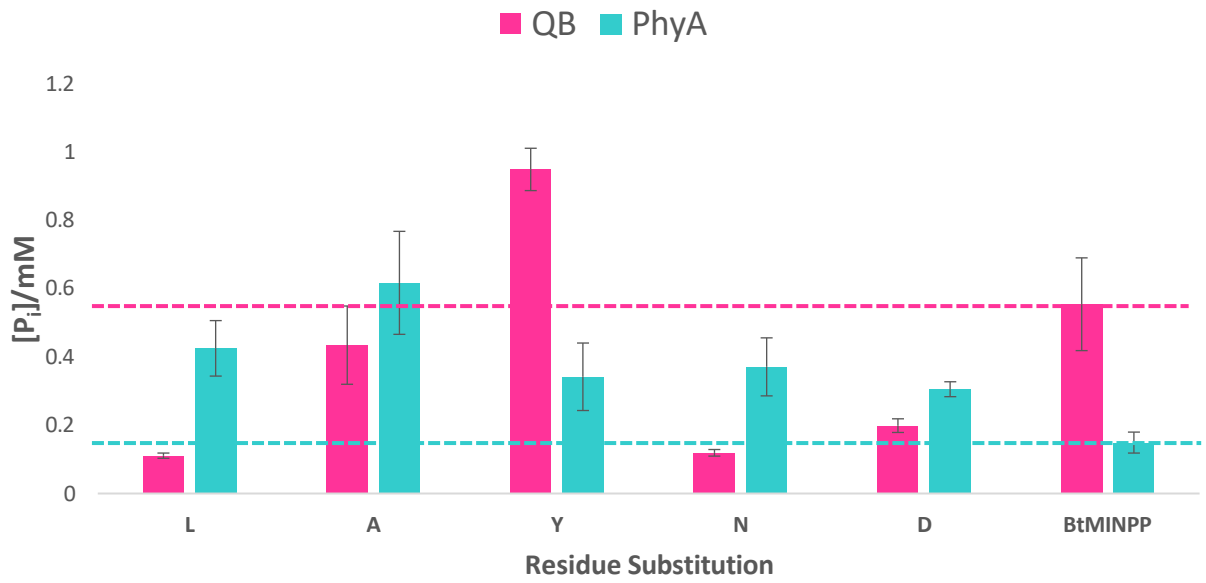


Figure 3.17 Relative levels of phosphate released from QB IP4 (magenta) or PhyA IP4 (teal) by five *Bt*MINPP Q276 variants. Dashed lines highlight the level of phosphate released by *Bt*MINPP from QB IP4 (magenta) and PhyA IP4 (teal), respectively. Error bars are based on standard deviation.

According to Sneath's Index[262], Grantham distances[260] and Miyata pair distances[264] (Table 3.4), the five Q276 variants overall are closer amino acid pair exchanges than the R275 variants. Based on the Grantham distance[260], Sneath's index[262], Miyata pair distances[264] and Epstein's coefficients of difference[263], the residue most similar to glutamine is asparagine. It would thus be expected, based purely on this, that Q276N would have the highest percentage activity relative to the wild-type, yet this is not the case – aside from Q276Y which has a considerably higher IP4 activity relative to the wild-type, Q276A has the highest relative activity, at 57.8% for IP6 and 78.4% for IP4.

Substitution	Grantham Score	Exchangeability Score	Sneath's Index (% *D*)	Epstein's coefficients of difference	Miyata's pair distance (D)
Q > L	113	0.39 (46)	20-24	1.00	2.70
Q > A	91	0.50 (44)	25-29	0.61	1.92
Q > Y	99	0.36 (30)	25-29	0.80	2.48
Q > N	46	0.34 (3)	10-14	0.03	0.99
Q > D	61	0.07 (1)1	20-24	0.03	1.47

Table 3.4 Amino acid difference 'scores' based on five different methods for the five glutamine to L/A/Y/N/D residue exchanges. N.B In the column Exchangeability Score, in parentheses is the number from which the mean was derived.

The HPLC traces for the five Q276 variants are presented in Figure 3.18. In terms of extent of degradation, it can be seen that Q276L, Q276A and Q276N yielded IP4 peaks thereby proceeded further than Q276Y and Q276D. However, the obtainment of 10% IP6 degradation was only approximate since the proportion of different IP_5 isomers was the primary focus of the HPLC analysis. Two variants – Q276L and Q276A – exhibit a similar IP_5 profile to the wild-type, whereby 5-OH is the major IP5 isomer produced, and levels of 4/6-OH and 1/3-OH are approximately equal (27% and 23%, respectively in the wild-type). Indeed, whilst the wild-type yielded 50% 5-OH IP5, Q276L and Q276A produced 63% and 56%, respectively. On the other hand, Q276Y, Q276N and Q276D deviate from this. Q276N produced more 1/3-OH (55%), followed by 5-OH (28%) then 4/6-OH (17%). For Q276Y and Q276D, the IP5 peaks were all considerably smaller, and therefore the proportions aren't readily visible from the HPLC traces, but the data revealed that the levels of the different isomers were similar to Q276N – 52/55% 1/3-OH, 26/24% 5-OH and 22/21% 4/6-OH for Q276Y and Q276D, respectively. This data is represented as pie charts in the corresponding Figure 3.19.

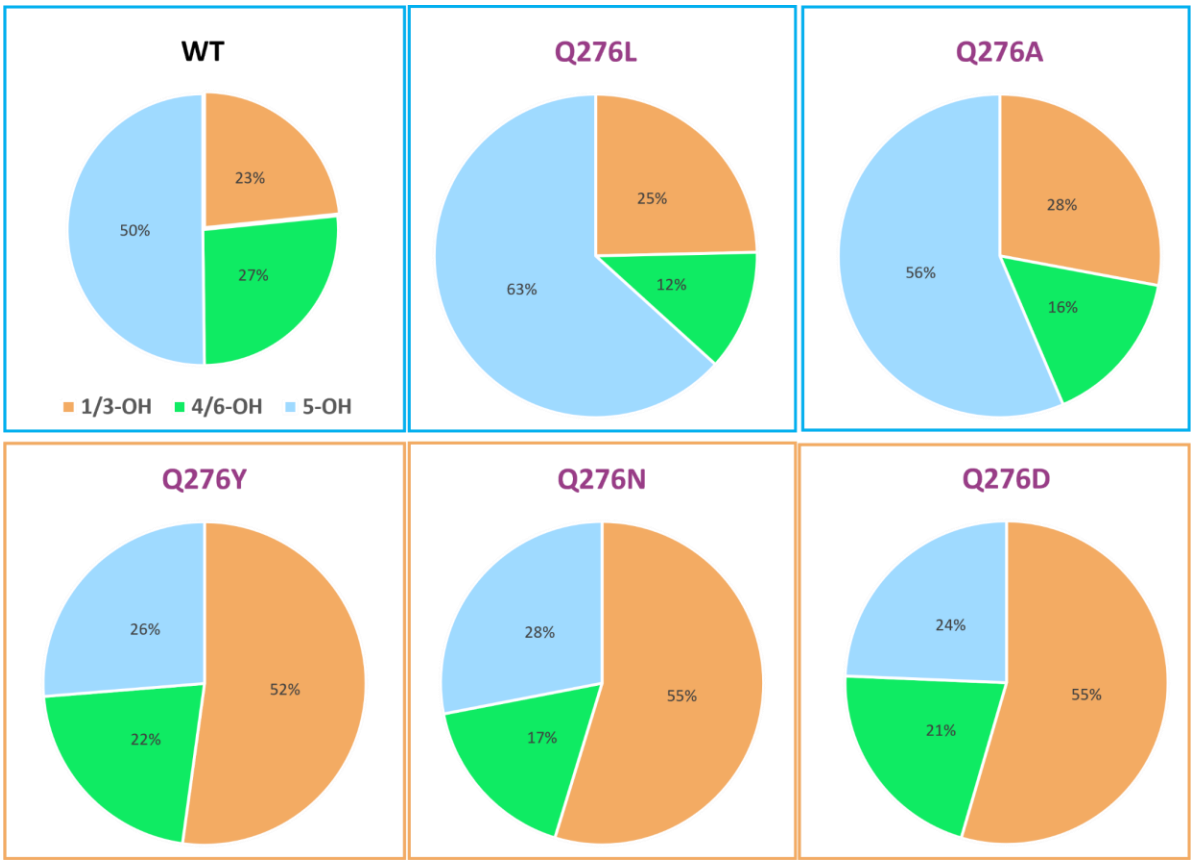


Figure 3.19 The proportion of the three different IP5 regioisomers generated by the set of Q276 *Bt*MINPP variants. The 5-OH isomer is represented by blue, the 1/3-OH isomer by orange and the 4/6-OH isomer by green.

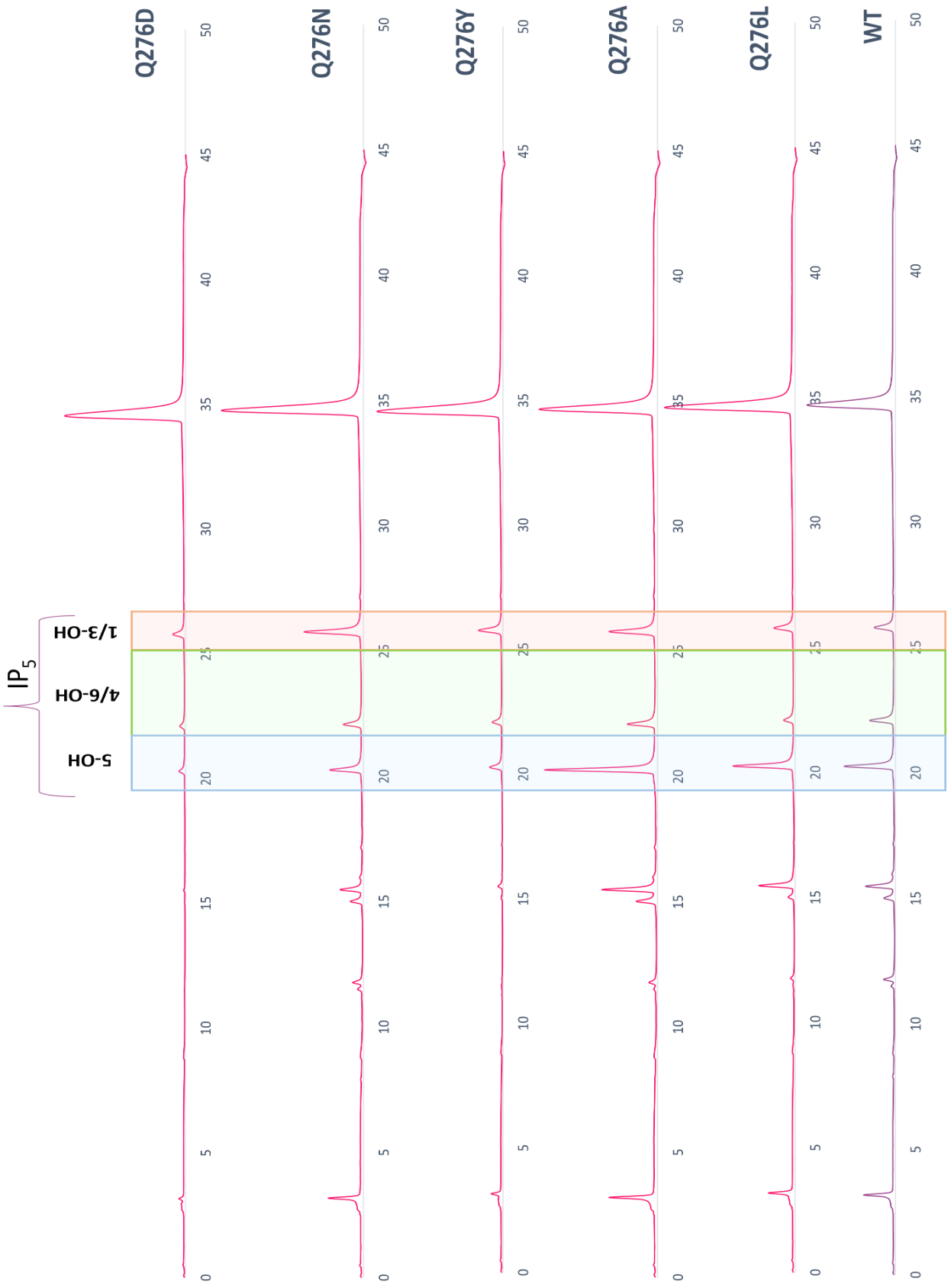


Figure 3.18 HPLC traces of the five Q276 variants (magenta) compared to the wild-type *Bt*MINPP (purple). The IP5 regioisomers generated are highlighted by a coloured box: 1/3-OH in orange, 4/6-OH in green and 5-OH in blue.

3.3.5 Activity towards IP6 and IP4 of *Bt*MINPP K280 variants

Presented in Figure 3.20 are the activities of the K280 variants against IP6 and Ins (2,3,4,5)P_4.

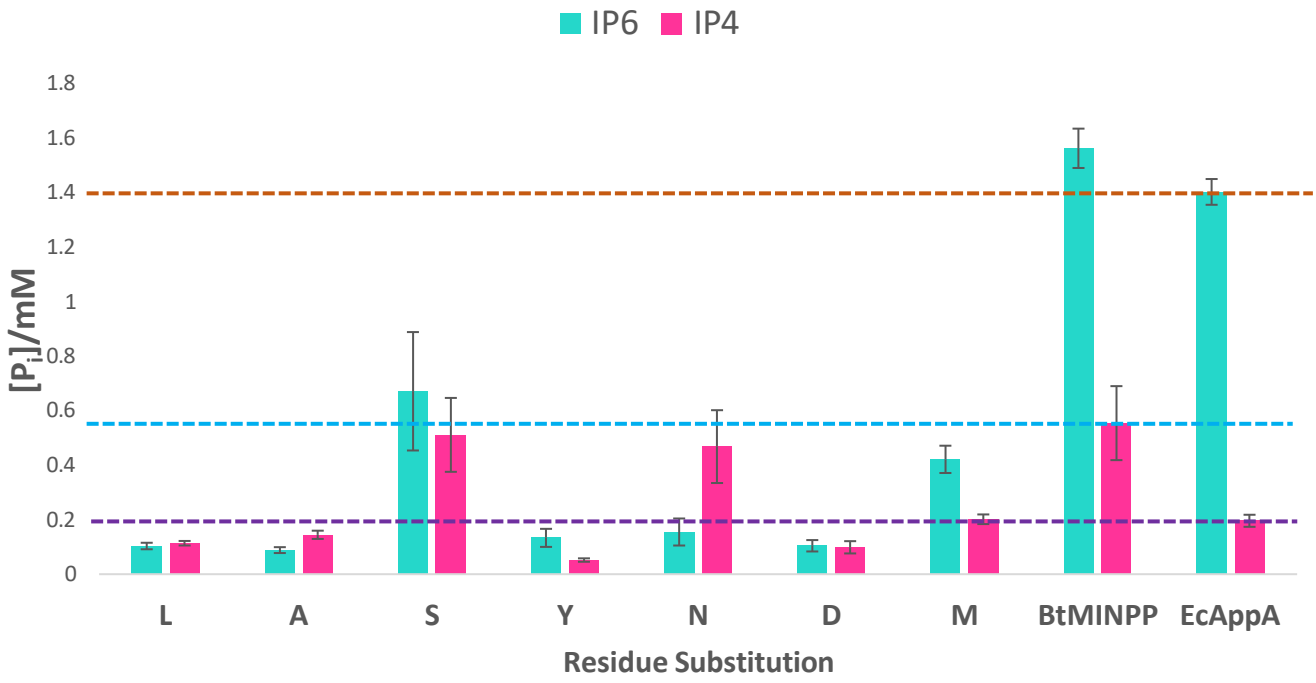


Figure 3.20 Relative levels of phosphate released from IP6 (teal) or IP4 (magenta) by seven *Bt*MINPP K280 variants. Dashed lines highlight the level of phosphate released from IP6 by AppA (brown); from IP4 by AppA (purple) and from IP4 by *Bt*MINPP (blue). Error bars are based on standard deviation.

It is very apparent that the levels of activity against both IP6 and IP4 are considerably lower than the R275 and Q276 sets of variants. None of the amino acid exchanges implemented resulted in a level of IP6 activity anywhere near equal to the wild-type. The variant with the highest activity for both IP6 and IP4 was K280S, however, the standard deviation was very large with very inconsistent triplicate readings. Nevertheless, relative to the wild-type, activity was 43% for IP6 and 92.2% for IP4 (Figure 3.21). K280N also exhibited a level of IP4 activity almost as high as the wild-type, at 84.5%, but it's activity against IP6 was much lower, at 9.9%. Activities of K280M against IP6 and IP4 were similar, at 27.0% and 36.4%, respectively, relative to the wild-type. K280L, K280A, K280Y and K280D are all essentially inactive compared to the wild-type, with 6.6%, 5.7%, 8.5%, and 6.7% relative IP6 activity, respectively. The analogous IP4 activities were 20.5%, 26.1%, 9.4% and 17.8% respectively, relative to the wild-type, and were all even lower than that of AppA.

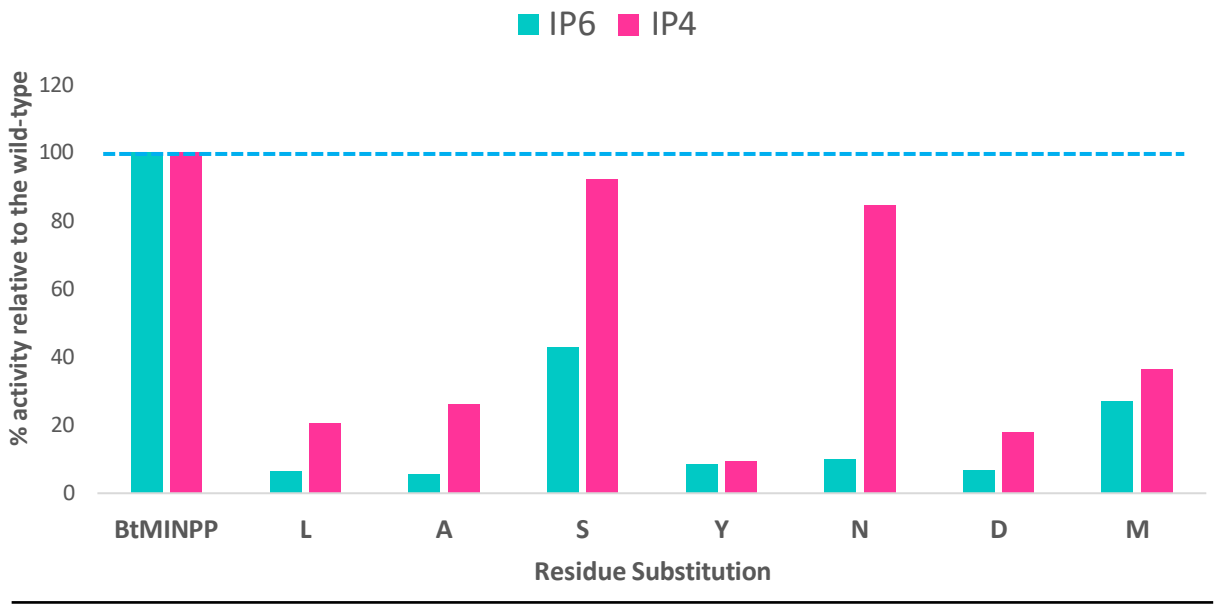


Figure 3.21 Activities against IP6 and IP4 of the set of *Bt*MINPP K280 variants displayed as a percentage relative to the wild-type activity.

The variant which stands out most in this set is K280M, having an activity against PhyA IP4 considerably greater than all other variants - including the wild-type - against either IP4 isomer (Figure 3.22). In addition to being 6.3 times more active than the wild-type for PhyA IP4, its activity for PhyA IP4 is also 1.7 times greater than for the wild-type against QB IP4. It is significantly more active (4.7 times) against the IP4 from PhyA than the IP4 produced by QB. K280L is the only other variant of this set which has higher PhyA IP4 activity relative to the corresponding QB IP4, although the difference is not as profound (3.4 times). The tyrosine mutation at the K280 position appears to have rendered the enzyme inactive against both IP4 substrates, as is the case for K280D and K280A with PhyA IP4. Both K280S and K280N are almost as active as the wild-type in terms of QB IP4 activity and they also exhibited similar activities against PhyA IP4 – levels both 1.5 times greater relative to the wild-type.

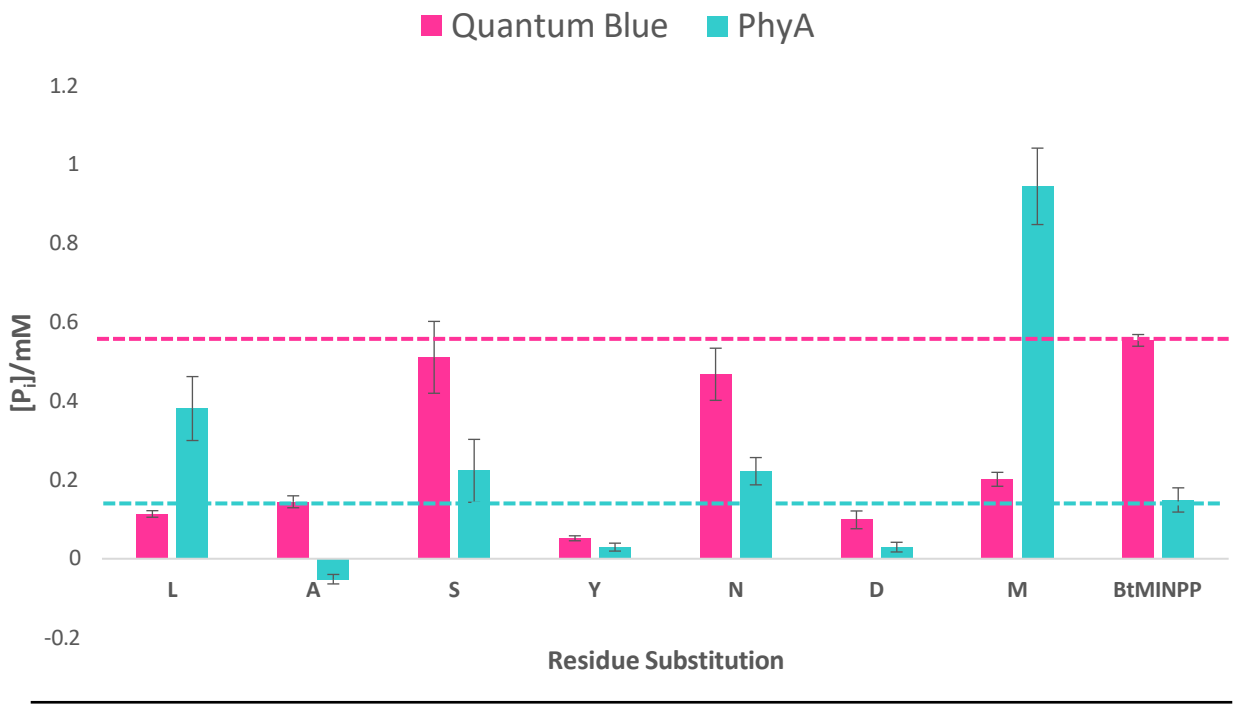


Figure 3.22 Relative levels of phosphate released from QB IP4 (magenta) or PhyA IP4 (teal) by seven *Bt*MINPP K280 variants. Dashed lines highlight the level of phosphate released by *Bt*MINPP from QB IP4 (magenta) and PhyA IP4 (teal), respectively. Error bars are based on standard deviation.

Substitution	Grantham Score	Exchangeability Score	Sneath's Index (% *D*)	Epstein's coefficients of difference	Miyata's pair distance (D)
K > L	107	0.30	20-24	1.00	2.98
K > A	106	0.60	25-29	0.61	2.96
K > S	121	0.38	30-34	0.22	2.71
K > Y	85	0.34	30-34	0.80	2.42
K > N	94	0.46	25-29	0.03	1.84
K > D	101	0.47	30-34	0.03	2.05
K > M	95	0.41	20-24	1.00	2.63

Table 3.5 Amino acid difference 'scores' based on five different methods for the seven lysine to L/A/S/Y/N/D/M residue exchanges. N.B In the column Exchangeability Score, in parentheses is the number from which the mean was derived.

Overall, the Grantham[260] and Miyata[264] distances of the K280 variant set are higher than between the amino acids in the R275 and Q276 sets, therefore indicating relatively less similarity; Sneath's index[262] suggests that the similarity of the lysine to X amino acid exchanges is in between the R275 and Q276 sets – more similar than the R > X set but less so than the Q > X set. According to the exchangeability scores, the K > X substitutions are more easily exchangeable than the analogous R275 and Q276 sets (Table 3.5).

The significance of the differences between the IP6 and IP4 activities of the variants relative to the wild-type are tabulated in the Appendices (A3.2). A negative t-value indicates higher activity of a variant relative to the wild-type and vice versa, whereas a t-value close to zero is indicative of similarity to the wild-type and therefore a non-significant value. Out of all 19 variants, the only variant which was not significantly different to the wild-type in terms of activity against IP6 was R275M. All others were significantly different ($p=<0.01$), whether that was increased or decreased relative activity. For activity against IP4, more than half the variants (11) had values which were significantly different ($p=0.05$) to the wild-type. The only variants which were statistically significantly more active relative to the wild-type were: R275S, R275Y and R275N.

3.4 Discussion

Phytases employed in feed enzyme applications suffer from a significant bottleneck which is the accumulation of the IP4 intermediate and corresponding reduced rate of hydrolysis[158]. The engineering of an established phytase to alleviate this issue would be very valuable to the monogastric feed sector due to optimisation of efficiency and resources[100]. There has previously been extensive analysis of the activities against IP6 and corresponding stereospecificities of a number of *Bt*MINPP variants, but investigation of IP4 activities is somewhat less explored[257,258]. The work in this chapter demonstrates the distinct differences effected by mutating the R275, Q276 and K280 positions of the catalytically flexible *Bt*MINPP, with single mutations eliciting changes in both IP6 and IP4 activities, and positional specificity.

Overall, there was no mutation which resulted in a variant that was more effective at IP6 dephosphorylation than the wild-type. However, several variants were able to process IP4 more efficiently than the wild-type. For the IP4 isomer produced by Quantum Blue, these consisted of five R275 variants (Y,S,N,L,M) and one Q276 variant (Y). Unfortunately, the improved IP4 activity was gained at the expense of IP6 activity, with IP6 activity levels significantly decreased relative to the wild-type. As for the PhyA IP4 isomer, since the wild-type activity is very low, multiple variants yielded improved levels – all of the Q276 variants and more than half of the R275 (L,A,Y,M) and K280 (L,S,N,M) variants. For variants of K280, although generally these were all much inferior to the wild-type for all substrates, one variant yielded a substantial improvement relative to the wild-type -K280M was more than six times more active against the PhyA IP4. The R275 and Q276 variants display differing preferences for the two IP4 substrates – variants of R275 are more effective at processing QB IP4 whereas Q276 variants are better with PhyA IP4. Aside from one variant (R275M), all R275 variants were much more active against IP4 (QB) than IP6, whilst for Q276 variants, aside from one variant (Q276Y), activity was greater for IP6 than IP4 (QB).

In terms of stereospecificity, mutations to the R275 position resulted in a change in the predominant IP5 isomer generated – from the 5-OH to the 1/3-OH IP5. The variant with the largest proportion of 1/3-OH IP5 was R275L, at 51%, which is slightly higher than the proportion of the 5-OH IP5 produced by the wild-type. With the exception of two variants, the Q276 variants also resulted in an interchange of specificity from the 5-OH to the 1/3-OH IP5, and was even more profound a change than in the R275 variants, with the 1/3-OH IP5 isomer constituting as much as 55%. On the contrary, the R275L and R275A variants maintained the same specificity as the wild-type, but the proportion of the predominant IP5 isomer was increased to 63% and 56%, respectively.

It is well established that during protein evolution, amino acid mutations eliciting fewer physiochemical changes are significantly more frequent than those involving relatively large changes, and this is because substitutions which are compatible with the retention of the existing conformation of the protein are favoured[264]. Yampolsky[261] indicated that the approach taken by Grantham[260] and the subsequently derived Miyata's distances[264], correlate particularly well with observed evolutionary propensities. Attempts to rationalise the observations in this work using a number of amino acid exchange indices based on physiochemical properties were somewhat inconclusive in that there was no definitive correlation with the experimental data. For example, according to the Grantham[260] and Miyata pair distances[264] and Sneath's index[262], of all the variants the Q to N substitution is the closest in terms of evolutionary distance and therefore would theoretically be the least negatively impacted, however in reality Q276 was one of the variants of the Q276 set in which activity was much decreased relative to the wild- type. Another example is that the K to A substitution is the closest according to the exchangeability scores, however the experimental data demonstrates that this variant was essentially inactive, with only ~5% activity relative to the wild-type. It must be acknowledged that some of these exchangeability scores are based on averages from a very limited number of samples therefore are not good representations. On the whole, the Q276 set of mutations are physiochemically closer pairs, according to Sneath's index, than the exchanges between R275X and K280X (where X is L, A, S, Y, N, D or M), and the R275 set are the most distant pairs, although this did not correspond to the results here.

Previous studies demonstrated that replacing the largely basic active site residues with acidic residues significantly decreased IP6 activity and suggested that a basic pocket is essential for substrate binding[257]. Since R275 and K280 are basic residues, it would be expected that the R275D and K280D variants result in a much decreased activity against IP6, relative to the wild-type. This is indeed observed in the data in Figures 3.9 and 3.21, in which the activities of R275D and K280D against IP6 were only ~5% and 6.7% of the wild-type activity, respectively. There is also a similar trend for the alanine, leucine and asparagine substitutions of R275 and K280. The alanine variant resulted in 11% (R275A) and 5.7% (K280A), the leucine variant 17.6% (R275L) and 6.6% (K280L), and the asparagine variant 16.1% (R275N) and 9.9% (K280N), relative to the wild-type. The reduction in activity is more pronounced in the K280 variants of these three substitutions. It is somewhat surprising that an alanine substitution decreases the activity so much, since alanine is supposed to be the residue which can most easily be a replacement for other residues. The corresponding Q276 variants were less impacted, with Q276D still maintaining 24.7% activity relative to the wild-type, which is not surprising since glutamine is not a basic residue like arginine and lysine. However, with IP4

(QB), there is a distinct difference – the R275 variants exhibit much higher activity, some much higher than the wild-type, yet the analogous K280 variants still have limited activity: 58.3% R275D vs 17.8% K280D; 140.8% R275L vs 20.5% K280L; 85.1% R275A vs 26.1% K280A; 229.2% R275N vs 84.5% K280N. It is plausible to suggest that since IP4 is less negatively charged than IP6, mutations from R275 or K280 to a non-basic residue would have less of a negative impact on IP4 activity relative to the wild-type, and may explain why some variants of R275 have much greater IP4 activity than the wild-type.

An particularly interesting trend observed in this work is that a mutation to tyrosine appears to effect a considerable change on activity: R275Y and Q276Y both exhibited the greatest activity against IP4 within their respective set of variants, yet, in contrast K280Y was essentially inactive. In previous studies on the F pocket of *Bt*MINPP (unpublished), R183Y and D186Y both displayed increased IP6 activities relative to the wild-type. This is somewhat surprising, since the partially hydrophobic tyrosine is usually found buried in the hydrophobic protein core, and is a disfavoured substitution for both arginine and glutamine, which are commonly involved in substrate binding due to their charge and/or polarity[265] In fact, the complex guanidinium moiety on the side chain of arginine is well suited for binding the negatively charged groups on phosphates[265]. The outcome observed with K280Y is much more expected, having similar properties to arginine.

3.5 Conclusion and Future Perspectives

Based on the differences in activity levels of R275, Q276 and K280 variants of *Bt*MINPP, these three residues are clearly involved in substrate binding, as already postulated. Since only seven amino acid substitutions were generated for each of these three positions due to time limitations, the next sensible step would be to perform saturation mutagenesis, which could reveal variants possessing greater levels of both IP6 and IP4 activity relative to the wild-type. Subsequently, residues could be identified which could facilitate the generation of variants with double mutations – for example, a single substitution of one of R275/Q276/K280 to improve IP6 activity, along with a single substitution in one of the other positions to improve IP4 activity. Already, based on the results reported here, a polyvalent approach may be implemented without investigating further mutations, since there are several variants within the limited set studied, which display considerably improved activity against both IP4 isomers. One of these may be combined with *Bt*MINPP in a formulation such that the wild-type phytase dephosphorylates IP6 down to IP4, before the *Bt*MINPP variant continues to rapidly degrade IP4. This would go some way towards solving the bottleneck of IP4 accumulation. One such potential combination could be *Bt*MINPP with R275Y. Of course, the thermostability of the variants potentially chosen would first need to be determined to ensure they are suitable for use as an animal feed additive. Indeed, even combinations of variants with differing stereospecificities could be exploited to increase the rate at which IP6 becomes dephosphorylated, i.e. a polyvalent approach. Additionally, solving the structure of a variant with a non-hydrolysable IP4 analogue such as IS4 could yield further insights into the nature of the substrate binding and provide rationalisation for the improved IP4 dephosphorylation of certain variants. Finally, the exploration of enzymes from an early point of evolution could serve as starting points for protein engineering efforts, since it is now known that the ancestral members of enzyme superfamilies sharing a common protein scaffold often display catalytic promiscuity and can easily be engineered to catalyse different reactions or accept different substrates[266].

CHAPTER FOUR

Structure-Activity Investigations of two MβLPhy Enzymes

4.1 Chapter Abstract

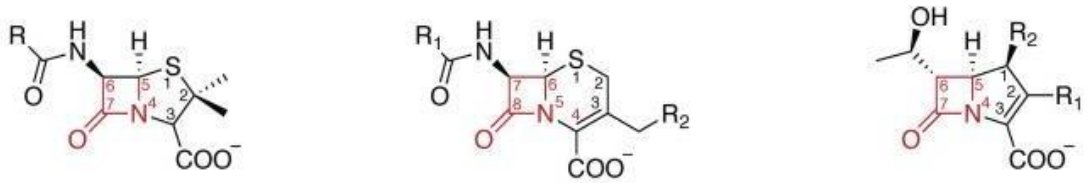
A novel class of phytate-degrading enzyme was reported in 2019 through functional metagenomic screening of forest soil[267]. Comprising just two enzymes, MβLp01 and MβLp02, this class is characterised by a metallo β-lactamase (MβL) fold which to date has never been associated with phytase activity[267]. The intention was to crystallise and solve the structure of at least one of these enzymes in order to investigate the structure-activity relationship with particular emphasis on the binding mode of IP6 and how this influences the stereospecificity of these enzymes in comparison with classical HAP phytases. This could reveal novel features which could be exploited in future phytase engineering projects. The X-ray crystal structure of MβLp01 was successfully solved at 1.95Å with an R_{free} of 0.23, however, remarkably, analysis of enzymatic activities of both MβLp01 and MβLp02 revealed lack of activity against IP6, suggesting that these enzymes are in fact not phytases. Investigation of phosphohydrolase activity revealed phosphodiesterase activity of MβLp01, which is plausible given the au courant literature on the MβL superfamily. There was also some low-level activity against ATP. MβLp02 on the other hand exhibited phosphatase activity against both ADP and ATP. Due to their structural classification as MβL superfamily members, these enzymes were also tested for β-lactamase activity using different methods. MβLp01 exhibited carbapenemase and penicillinase activity but no activity against the tested cephalosporin. The findings in this chapter illustrate the profound importance of critically assessing published results for their scientific credibility and how the reporting of an incorrect conclusion can rapidly be disseminated throughout the scientific community, influencing subsequent studies. In the words of Thomas Edison: *'Just because something doesn't do what you planned it to do doesn't mean it is useless'.* Although phytases were the actual focus of this thesis, what has resulted from the work in this chapter is the identification of an incorrectly reported novel enzyme class, the impact of which should not be underestimated.

The following chapter necessitates a brief narrative on the structure, function and characteristics of the MβL superfamily, since this was not covered in the general thesis introduction yet is a pre-requisite for the discussion section of this chapter.

4.2 The Metallo β-Lactamases: A Brief Introduction

4.2.1 Classification

The β-lactamase family constitutes a large and diverse group of enzymes which catalyse the hydrolysis of the four membered heterocyclic β-lactam ring of the most commonly used antibiotics - the β-lactams[268]. Accounting for 60% of all current antibiotics[269], the extensively and indiscriminately used β-lactam antibiotics encompass four chemical scaffold groups: penicillins, cephalosporins, carbapenems, and monobactams (Fig 4.1). This cleavage at the amide bond by these hydrolytic enzymes thus renders the antibiotic ineffectual against the pathogenic bacterial strain harbouring the enzyme, permitting the spread of bacterial infection[269-271]. The production of these extracellular or periplasmic enzymes by pathogenic bacteria therefore represents the most common mechanism of antibiotic resistance[270-273], the increasing incidence of which is a rapidly mounting global threat to public health[274]. They are of particular concern because they possess the ability to inactivate all β-lactam antibiotics – including the last-line carbapenems - apart from monobactams[266], and as yet there are no effective and clinically available inhibitors against MβLs in pathogenic bacteria[269,271]. With over 50,000 β-lactamase sequences accumulated to date since their discovery in the 1960's, at least 100 variants have been identified in pathogens[273]. There are two concurrent classification systems for β-lactamases, with Ambler's classification[275] based on molecular and structural characteristics whilst the Bush-Jacoby functional classification[276] is based on the clinically relevant phenotypic factors. The latter - consisting of group 1 (class C) cephalosporinases; group 2 (classes A and D) broad spectrum, inhibitor resistant and extended spectrum B-lactamases and carbapenemases; and group 3 (class B) metallo-β-lactamases - is beyond the scope of this chapter and therefore the reader is referred to Bush & Jacoby (2009)[276]. Ambler's classification meanwhile divides β-lactamases into four classes, where A, C, and D employ serine in their catalysis whilst class B are zinc-dependent metallo-hydrolases (EC 3.5.2.6)[275]. The latter subclass, the eubacterial class B metallo-β-lactamases, belong to an overwhelmingly diverse and extensive superfamily of metallo-β-lactamase (MβL) fold proteins which are not limited to bacteria[271,277,278]. Named after the identification of the founding member, BcII**,** they are characterised by a distinctive αββα core fold, usually of 200-300 residues[273], and this is the group of enzymes with which this chapter is concerned[269,271].



Penicillin **Cephalosporin** **Carbapenem**

Figure 4.1. Chemical scaffolds of the three main classes of clinically used β-lactam antibiotics hydrolysed by MβLs.

4.2.2 Class B MβLS

The first zinc-dependent MβL (BcII) was isolated from a clinically insignificant *Bacillus cereus* strain in 1966 by Sabath and Abraham, and marked the start of the myriad of MβLs discovered since[279]. To date, at least 6000[280] MβL enzymes have been identified in more than 50 bacterial species including gram-negative and non-fermentative bacteria[277]. At least 325 variants exist in bacteria, grouped into 63 MβL types[277]. Such was the scope that class B MβLs were further divided into 3 sub-classes – B1, B2, B3[281] – according to zinc(II) stoichiometry, metal co-ordinating ligands and the primary sequence homology (albeit limited – as low as <20%[272,273]) between them[271]. With the extensively studied BcII as the founding member, the B1 subclass represents the largest number of recognised MβL enzymes (around 2800 sequences[273])- with members most usually associated with the spread of antibiotic resistance in gram- negative human pathogens[271] - and is therefore the most clinically relevant and most characterised MβL subclass[272]. The most notorious of these are NDM-1 (New Delhi metallo β-lactamase 1) from *Klebsiella pneumoniae*[282], IMP-1 (imipenemase 1) from *Serratia marcescens*[283] and VIM-2 (Verona imipenemase 2) from *Pseudomonas aeruginosa*[284]. The B2-type MβLs constitute the smallest group of MβLs[271] (around 140 sequences[273]), representing just 3% of all known MβLs[269], and they are not produced by primary human pathogens, being found only in *Serratia* and *Aeromonas* species[285]. The representative B2 MβL is CphA[271] (*carbapenem- hydrolysing Aeromonas hydrophila*) originating from *Aeromonas hydrophila*[286]. Whilst MβLs belonging to the B3 subclass (~3400 sequences[273]) have usually originated from environmental bacteria, they are increasingly being identified in clinically significant pathogenic organisms[271] such as the MβLs AIM-1 from *Pseudomonas aeruginosa*[287] and SMB-1 from *Serratia marcescens*[288].

In general, MβL enzymes are soluble periplasmic enzymes of 240-310 amino acids before removal of a 17-30 amino acid N-terminal signal peptide[274] by type 1 signal peptidase (SPase 1)[289]. Based on sequence, subclass B3 is relatively more evolutionarily distant from B1 and B2

which are more closely related (14-24%) than subclass B3 is to [B1+B2] (2-14%)[271,274]. Among each subclass, the B1, B2 and B3 MβLs have sequence identities of 31.8, 60.2 and 33.0%, respectively[269]. Despite being highly divergent in sequence, the overall scaffold is well conserved among all three subgroups[269,274] - particularly between B1 and B2 MβLs which have higher similarity to each other than B3 MβLs have to either B1 or B2, despite the difference in number of catalytic zinc ions[269]. B3 MβLs also display considerable diversity within the subgroup, with uncharacteristic active site variations not witnessed in the other subclasses, and pairwise similarities as low as 15%[290], rendering them the most divergent group phylogenetically and structurally[271]. This core scaffold consists of 5 α-helices and 13 β-strands which constitute the distinctive αβ/βα sandwich fold[269]. Residues of four loops (L1-4) protruding from the two central β sheets co-ordinate the catalytic zinc ion(s), which are located at the centre, whilst three external loops (eL1-3) protruding above the canonical sandwich fold form and shape the substrate binding pocket[269]. This pocket is dissimilar between the three subgroups thereby dictating substrate binding and specificity[269]. The substrate binding pocket of B1 MβLs is medium sized, in B2 MβLs it is narrow whereas B3 MβLs have a wide-open pocket[269]. In all cases, the upper left side of the pocket accommodates the various bulky R groups at the five- or six-membered ring side of the core β-lactam scaffold whilst the lower face of the pocket binds the R groups on the β-lactam ring side, permitting limited structural substitutions[269]. The lower area of the pocket is particularly restrictive in B2 and B3 MβLs[269]. The substrate profile is broad across the three sub-classes, although none of the MβLs are active against monobactams[271,274]. B1 enzymes exhibit a broad-spectrum (hydrolysing penicillins, cephalosporins and carbapenems) activity profile whilst B3 enzymes are broad-spectrum with a putative preference for cephalosporins[272,274]. Members of the B2 group on the other hand display the highest degree of substrate selectivity within the MβLs, exhibiting a very narrow substrate spectrum in which they are strict carbapenemases[269,271,274]. In addition to the variations in substrate binding pocket, this diversity in substrate profile arises from the differences in catalytic mechanism[269,271-274]. This is attributed in large part to the number of zinc ions bound in the active site, which are central to catalysis, serving to stabilise a hydroxide ion (formed from deprotonation of a catalytic water molecule) which is activated to initiate β-lactam hydrolysis[269,271,291]. Although all class B MβLs host two zinc binding sites, enzymes of the B2 series are active only in the mono-zinc state and are inhibited by the binding of a second zinc ion[271,272,274]. Despite some former controversy, it is believed that the di-zinc state of subclasses B1 and B3 is the only catalytically relevant form[274] and this is evidenced by positive co-operativity of zinc binding[292]. Located in close proximity to one another, the zinc binding sites are designated Zn1 and Zn2[274,290] with the distance between them being almost the same across all subclasses[269], although there are some reports of these ranging from 3.5-4.6Å[274]. Indeed, these metal binding sites in B1 and

B3 MβLs have differing affinities for zinc ions[285,293]. In subclass B1, Zn1 is also referred to as the 3H site and Zn2 as the DCH site since the strictly conserved zinc ligand residues are 3 histidines (116, 118, 196) and Asp120-Cys221-His263, respectively[271-274]. Both metal ions are coordinated by a Asp221[273]. MβLs of subclass B2 and B3 have variations in the co-ordinating residues. In subclass B2 – in which the single zinc ion binds to the Zn2 site[274] - whilst the zinc ligands in the Zn2 site are conserved, the histidine residue at 116 in the Zn1 site is replaced by an asparagine (Asn116-His118-His196)[272,274,290]. The canonical B3 MβL active site motif is H/G116-H118-H196 and D120-H121-H263 although there is considerable diversity among this group of enzymes[272,274,290]. These different zinc binding motifs can be used as a distinguishing feature and can be summarised as a conserved consensus motif of HXHXD(X)iH(X)jC(X)kH (where i = 55–74, j = 18–24 and k = 37–41[294]. Although *in vivo*, MβLs function only with zinc, other 2^+ metals such as Cd, Co, Ni, Cu and Mn can give rise to activity *in vitro*[274]. The reader is referred to a review on the role of zinc in metallo β-lactamases[274]. Under physiological conditions, studies have shown that the metalation state of MβLs depends largely on the Zn(II) availability in the external milieu, since they are secreted as unfolded polypeptides which bind Zn(II) in the periplasm[289]. Earlier investigations however, led to the concept that substrate binding induces the binding of Zn(II)[295]. It was thought that in the absence of substrate, MβLs exist in the apo- state due to the very low concentrations of free zinc ions in the cellular environment. In this same theory, the binuclear zinc state was considered a potentially artificial state, being stabilised only at unphysiologically high Zn(II) concentrations[295].

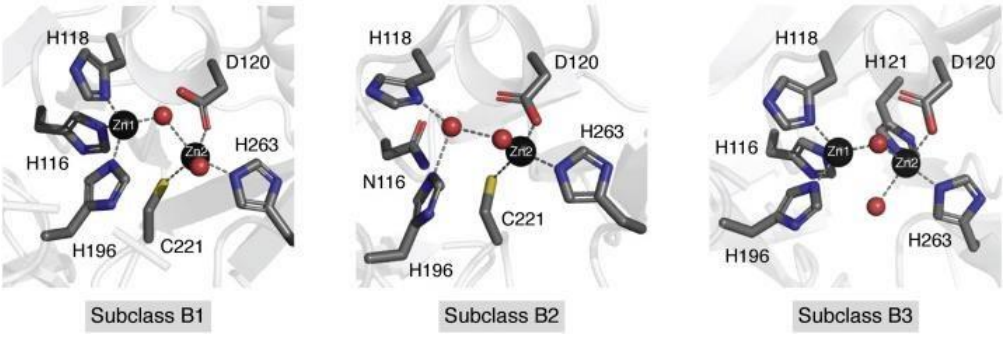


Figure 4.2. Zinc coordination geometry in the three MβL subclasses. Image from reference #269.

Although a detailed summary of the B1 and B2 subclasses is beyond the remit of this chapter, it is useful to compare the active site architecture across the different subclasses. The reader is referred to a detailed review of the structures of MβLS among the three subclasses[269].

4.2.3 Structural characteristics of the B3 MβL subclass

In comparison to the B1 and B2 groups, MβLs of the B3 subclass are more diverse, having significant structural differences in the overall scaffold, with further variation in the active site geometry, mechanism, reactivity, substrate selectivity and conformation of the external loops[269,271]. They also display substantial intra-subgroup diversity[271]. In terms of the core scaffold, the first three β strands in the first β-sheet are very short with an extended N-terminal tail[269]. This affords the first external loop (eL1) with a flexible conformation[269]. The second β-sheet comprises five β-strands rather than the six in subclasses B1 and B2[269]. The C-terminal β13 strand is instead a helix (α3) and in some B3 MβLS there is an additional α3' helix following the α3 helix, before the β7 strand[269]. The external loops which influence the shape of the substrate binding pocket are distinctly different from those of B1 and B2 MβLs, having a long eL1 and eL2, with a short eL3[269]. Since the inner loop L3 lacks a zinc coordinating residue, eL3 from the second β-sheet is shifted across to the right, exposing a large cavity in the upper left lip which is sheltered only on the upper and right sides by eL3[269]. The open shape of the substrate binding pocket accommodates bulky R groups on β-lactam substrates, unlike in B1 and B2 MβLs[269]. It also renders both zinc ions substantially exposed to solutions[269]. The zinc binding sites are more diverse both among B3 MβLs and relative to the other subclasses[269,290]. Instead of a cysteine residue from the L3 Zn2-coordination loop, a compensatory histidine residue from the L1 loop is recruited to bind the Zn2 ion from the underside position[269]. The L3 is thus of a shorter, different conformation[269]. In addition, in some B3 enzymes the first histidine residue of the Zn1 site is replaced by a glutamine residue[269,271,290]. The nucleophilic hydroxide oxygen bridges the two zinc ions and, along with the three Zn1 histidine ligands, coordinates Zn1 tetrahedrally[271]. Meanwhile, the Zn2 site adopts a distorted trigonal bipyramidal geometry, with the equatorial plane formed by the two Zn2 histidine residues along with the metal ion-bridging hydroxide and the axial positions occupied by the aspartic acid (D120) and an additional water ligand[271,274]. In addition, D221 is replaced by a S221 which no longer coordinates the metal directly[273].

There is now emerging evidence of B3 MβL variants with different zinc-coordinating residues in both sites, indicating the active site diversity among this subclass[271,290]. According to sequence analyses, in addition to the canonical form hosting the HHH/DHH motif, there are at least three distinct but less abundant variants which have evolved within this clade[271,290]. Two of these are characterised by QHH/DHH and EHH/DHH and referred to as B3-Q and B3-E, respectively[271,290]. With substitutions in both zinc binding sites, B3 variants with the HRH/DQK motif, termed B3- RQK, are the most divergent and least represented members of the B3 clade[271,290].

The evolutionary landscape of the B3 subgroup, as revealed by recent studies, indicates that B3 MβLs are related to enzymes with distinct functions such as nucleases[271,277,296].This has led

to the philosophy that they have evolved from a discrete ancestral functionality and independently acquired β-lactamase activity as a result of environmental pressure[271].
For a discussion of recent advances in structure-function relationships of B3-type MβLs, the reader is referred to ref #271.

4.2.4 Functional diversity of the MβL superfamily

Comprising around 34,000 known proteins to date[277,296], and over 480,000 sequences in the UniProt database[273], with sequence similarities as low as 5%[296], the MβL superfamily represents an extensive set of highly diverse and multi-functional enzymes which are ubiquitously distributed across Eukarya, Bacteria and Archaea[266,277,278]. Ancestral studies have shown that the highly conserved MβL motif is one of the oldest in existence[277] and based on this and other evidence, the MβL superfamily is thought to have ancient origins[266]. This motif, HxHxDH, along with residues H196 and H263, form the catalytic site which is ancestrally shared by the superfamily[277,296,297]. Due to the accrual of new insight based on the discovery of novel MβLs exhibiting diverse functions, it is now believed that there are at least 24[296] specialised functions among members of the family, in addition to the canonical β-lactamase activity, including: ribonuclease, nuclease, glyoxalase II, lactonase, alkyl- and arylsulfatase, phosphodiesterase, phytase (see discussion), ascorbic acid degradation, anti-cancer drug degradation, membrane transport, pesticide degradation, dehalogenase, oxidase, flavoproteins, phospholipase, DNA repair and uptake and CMP-NeuAc hydroxylase[271,273,277,278,296]. In fact, the true β-lactam-degrading members represent only a small proportion of the superfamily[271] and enzymes belonging to the MβL superfamily which lack β-lactamase activity are defined as structurally representative MβL fold proteins[297]. To this end, Daiyasu et al (2001) categorised the MβL superfamily into 16 groups based on biological function[278]. Groups 1-7 (Class B β-lactamase, glyoxalase II, flavoproteins and oxidoreductases, arylsulphatase, type II polyketide synthase, enzymes involved in the processing of mRNA 3-ends, enzymes involved in DNA repair) were already established as MβL superfamily members, with the classic metallo β-lactamases of Ambler Class B clustered in Group 1, whilst groups 8-16 (proteins involved in DNA uptake, teichoic acid phosphorylcholine esterase, phnP, CMP-*N*-acetylneuraminate hydroxylase, the *romA* gene product, alkylsulphatase, carbofuran hydrolase, methyl parathion hydrolase, Class II 3′, 5′- cyclic-nucleotide phosphodiesterases) have been more recently introduced as a result of database searching[278]. In fact, the class B β-lactamases constitute just 1.5% of the MβL superfamily, with around 6,300 sequences[273]. There are even some MβL members which have non-hydrolytic (nitric oxidoreduction and sulphur di-oxygenation) or non-enzymatic functions[271,277,296]. There is increasing evidence which indicates these enzymes are promiscuous

in their functions, catalysing more than one specialised reaction[271]. Indeed, on average, MβLs have been shown to catalyse 1.5 reactions in addition to their native one[296]. It is thought that the voluminous active site cavity provided by the MβL fold facilitates the binding of multiple substrates of various complexities which is an important factor determining enzyme promiscuity[266]. Another factor influencing promiscuity is flexibility, although this is debated[298], since natural protein evolution is believed to begin with a stable protein backbone and promiscuous activity, evolving over evolutionary timescales to decreased stability and corresponding increased substrate specificity[299]. Nevertheless, catalytic promiscuity is often found in members of mechanistically diverse superfamilies that share a common protein scaffold, just as is the case with the MβL superfamily[298] ,with the promiscuous activities manifesting connections between different catalytic landscapes[296]. It is thought that the conservation of the structural framework in MβLs facilitates this multifunctionality via a unifying catalytic mechanism and this was exemplified by the extensive engineering of human glyoxalase II to a functional MβL[300]. Indeed, a study of 24 enzymes in the metallo-β-lactamase superfamily showed an overwhelming display of catalytic promiscuity against substrates for enzymes in distinct families[298]. The recently identified and characterised hyperthermostable Igni18 is postulated to embody the core of multiple modern MβLs among prokaryotes and eukaryotes, such that it may serve as a prototype for the study of evolution within this superfamily[266]. This study also defined the non-conserved, so-called protein variable regions which dictate the multifunctionality of MβL fold enzymes[266]. Recent evolutionary studies have suggested that the last common ancestor of the MβL superfamily was in fact a phosphodiesterase involved in nucleotide processing[296].

4.3 Experimental Procedures

4.3.1 Gene constructs and cloning

The genes coding for MβLp01 and MβLp02 were codon optimised for *E. coli* expression and each cloned into the vector pET15b (with the signal peptides omitted) by Genscript. The genes were amplified by PCR using Phire® Green Hot Start II polymerase (Thermo Scientific). PCR reactions were incubated in a Thermal cycler C1000 Touch (Bio-Rad) and their composition and thermocycling details are detailed in Table 4.1. PCR was carried out employing an annealing temperature gradient (50, 55, 60, 65°C) and two different primer pairs (sequences tabulated below) to maximise success. Oligonucleotides were obtained in the dry, desalted form from Merck. Generation of a PCR product of the expected size was verified by DNA agarose gel electrophoresis (chapter 2.3). Following PCR amplification of the genes, directional TOPO cloning (chapter 2.2.1) was performed to transfer each gene into the pBAD202/D-TOPO® vector (Invitrogen) for cytoplasmic expression as HP-thioredoxin fusion proteins. This vector encodes kanamycin resistance and provides an N-terminal cleavable His-patch thioredoxin tag (HP-Trx) followed by an enterokinase cleavage site (Figure 4.3). Competent *E. coli* One-shot TOP10 cells were transformed with the recombinant plasmids and used as both the plasmid storage strain and expression strain. Glycerol stocks of *E. coli* TOP10 transformants cultured overnight in LB media at 37°C were prepared and the pBAD202/D-TOPO_MβLp01 and pBAD202/D-TOPO_MβLp02 plasmids were extracted using a QIAprep Spin Miniprep kit (Qiagen) for plasmid stock storage at -20°C. Plasmids were sequenced by sanger sequencing at Eurofins using their Mix2seq kit along with the TrxFus forward and pBAD reverse sequencing primers included in the pBAD202/D-TOPO® cloning kit (sequences included below).

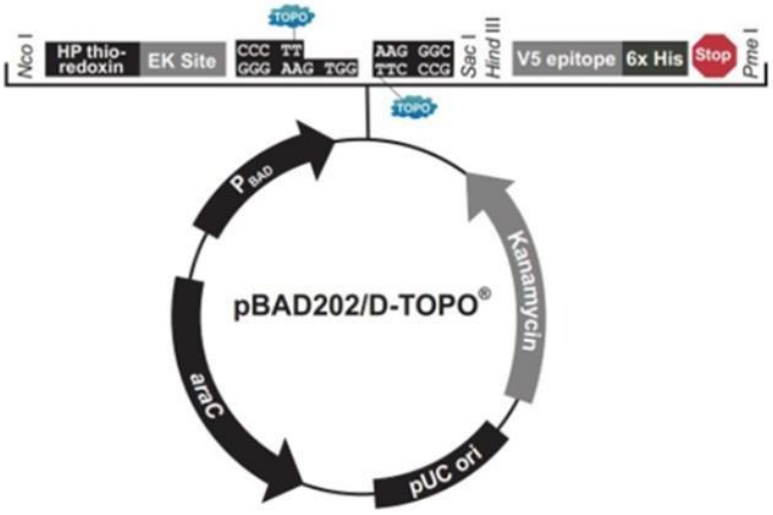


Figure 4.3 Features of the pBAD202/D-TOPO® vector, including the cloning insert site.

PCR Reaction Components

Component	Volume (μL)	Final Concentration
Phire® Green Hot Start II polymerase	0.4	
5X Phire® Green buffer	4	1X
dNTPs (10 mM)	0.4	200 μM each
Fwd Primer	1	0.5 μM
Rev Primer	1	0.5 μM
Template DNA	1	
DMSO	0.6	3%
H_2O	11.6	

Thermocycling Parameters

Cycle step	Temperature	Time	Cycles
Initial denaturation	98°C	30s	1
Denaturation	98°C	5s	
Annealing (Gradient)	50/55/60/65°C	5s	30
Extension	72°C	9s	
Final extension	72°C	60s	1

Table 4.1 Composition of the PCR reactions for amplification of MβLp01 and MβLp02

MβLp01 Forward 1	**5'-CACCCATATGGCTTCACCCGTAACACAAG-3'**
MβLp01 Reverse 1	**5'-GGATCCTTAGTAAACATCCAGATCATGC-3'**
MβLp01 Forward 2	**5'-CACCCATATGGCTTCACCCGTAACACAAGTTG-3'**
MβLp01 Reverse 2	**5'-GGATCCTTAGTAAACATCCAGATCATGCG-3'**
MβLp02 Forward	
MβLp02 Reverse	
TrxFus Forward	**5´-TTCCTCGACGCTAACCTG-3´**
pBAD Reverse	**5´-GATTTAATCTGTATCAGG-3´**

Table 4.2 Sequences of the oligonucleotides used in the PCR amplification of MβLp01 and MβLp02 and of the sequencing primers.

Prior to sequencing of the plasmids, a diagnostic restriction digest was carried out to verify the correct size and orientation of the insert. Reactions (components as detailed in Table 4.3) were incubated at 37°C for 15 min before the products were separated by DNA agarose gel electrophoresis and visualised under UV light.

Component	Volume	Final Concentration/amount
DNA	3-5uL	200ng
10X CutSmart Buffer	1uL	1X
HindIII-HF	0.5uL	10 units
Pm1I	0.5uL	10 units
H_2O	Up to 10uL	

Table 4.3 Composition of restriction digests of pBAD202/D-TOPO_MβLp01

4.3.2 Expression trials of MβLp01 and MβLp02 (small-scale)

Expression of MβLp01 and MβLp02 was trialled from pET15b and pBAD202/D-TOPO, with induction by IPTG or L-arabinose, respectively. Upon receipt of the MβLp01/MβLp02 harbouring pET15b plasmids, before cloning into pBAD202/D-TOPO, the plasmids were used to transform *E. coli* BL21 and SHuffle® T7 Express expression strains and prepare glycerol stocks of these. For the pET15b constructs, expression from *E. coli* BL21 and SHuffle® T7 Express were tested on a small scale (10 mL culture volume), with IPTG concentrations and temperature ranging from 0.2 -1 mM and 16-37°C, respectively. The pBAD202/D-TOPO constructs were expressed from *E. coli* TOP10 with arabinose concentrations of 0.2%, 0.02%, 0.002% and 0.0002% and temperature conditions of 16°C, 23°C for 16-20h or 30°C, 37°C for 4.5h. Growth media used throughout was LB. The protein extraction reagent BugBuster™ was used for the quick and convenient extraction of soluble proteins prior to analysis by SDS-PAGE. Cells were also lysed by sonication depending on requirements.

4.3.3 Large-scale expression and purification of de-tagged MβLp01

For large-scale expression, 6L (1L x 6) LB supplemented with kanamycin (50ug/mL) was inoculated with an appropriate volume of an overnight culture of a single colony of *E. coli* TOP10 transformed with pBAD202/D-TOPO_MβLp01. The culture was maintained at 37°C with shaking (180rpm) until the OD_{600} reached ~0.6, at which point arabinose (0.2%) was added. After overnight growth at 18°C, 180rpm, cells were harvested (5000 x g, 10 min, 4°C) and then resuspended (after discarding the spent media) in a lysis buffer consisting of 50 mM NaH_2PO_4 pH 7.5, 300 mM NaCl, 10 mM Imidazole, 10% glycerol and 0.1% Triton X-100. After 1-2h at -80°C, the cell suspension was thawed and added to it was a protease inhibitor tablet (cOmplete EDTA-free, Roche), bovine DNase I (~10 µg/mL) and lysozyme (chicken egg white, 0.2 mg/mL). After 30 min on ice with agitation, cells were subjected to mechanical lysis by cell disruption at 18,000 psi on an LM20 microfluidizer (Analytik Ltd). The lysate was centrifuged (48,000 x g, 30 min, 4°C) and the clarified lysate, after filtration through a 0.45µm filter (Sartorius), was loaded onto a Ni^{2+}-NTA affinity column (HisTrap HP 5mL, Cytiva) using an AKTA Pure (Cytiva) chromatography system. The resin was first equilibrated with X CV of Ni^{2+}-NTA wash buffer (50mM NaH_2PO_4 pH 7.5, 300mM NaCl, 20mM Imidazole) at a flow rate of 5 mL/min, before application of the sample. Unbound proteins were washed off the resin with X CV of wash buffer, before elution of the target protein with a gradient from 20 to 500 mM imidazole, at a flow rate of 1 mL/min. Fractions containing the target protein were identified by SDS-PAGE (4-12% Bis-Tris Bolt™ gels) and pooled, [concentrated] after which the sample was desalted using a HiPrep 26/10 column pre-packed with Sephadex G-25 resin (Cytiva) into an enterokinase cleavage buffer consisting of 50 mM Tris-HCl pH 7.4, 50 mM NaCl, 2mM $CaCl_2$.

MβLp01 was incubated at 25°C with enterokinase at a ratio of 1000:1 for 1h before assessing the extent of cleavage by SDS-PAGE. The cleaved sample was then desalted back into IMAC wash buffer and subjected to a reverse IMAC step in which the flow through was retained and all fractions corresponding to chromatogram peaks at 280nm reviewed by SDS-PAGE. Fractions containing tag-free MβLp01 were again pooled and then concentrated in a spin concentrator device (3000 or 5000 kDa MWCO; Amicon) until the volume was 2-2.5 mL. This was then injected onto a HiLoad 16/600 Sepharose 75 pg column (Cytiva) for separation by size exclusion chromatography, in a buffer composed of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1mM DTT, circulated at 1 mL/min for 1 CV. SDS-PAGE was again used to determine the fractions containing the target protein, which were pooled and after further concentration (3000 kDa MWCO; Amicon, pure MβLp01 was obtained at a final concentration of 1.4 mg/mL. Some of this was used immediately for crystallisation screens whilst the remainder was cryoprotected (30% glycerol and flash frozen in LN_2) and stored for later use in aliquots at -80°C.

4.3.4 Crystallisation of de-tagged MβLp01

The OryxNano protein crystallisation robot (Douglas Instruments Ltd) was used to set up several commercial screens with the sitting drop vapour diffusion technique – Index screen (Hampton Research), LMB screen (Molecular Dimensions) and 3 plates of PEG/ion screen (Hampton Research). Drops of 1μL were composed of 1:1 protein and reservoir solution. Crystals grew in a variety of conditions within three days at 16°C. 16 crystals were harvested from the PEG/ion and LMB screens using a loop size of 0.08mm and cryoprotected in a solution of mother liquor and 20% (v/v) glycerol before being stored in LN_2 and sent for X-ray data collection in a LN_2 filled- dry shipper.

4.3.5 Data collection and processing

Single wavelength anomalous X-ray diffraction was carried out on beamline I24 of the synchrotron at the Diamond Light Source (DLS; Oxfordshire, UK). Images from a single crystal were collected on a Pilatus3 6M detector with 2400 frames of 0.15° oscillation and an exposure time of 0.02 s at an X-ray wavelength of 0.9795 A° and beamsize 30 x 30 am which was set to 100% transmission. Resolution 2.20Å. The X-ray diffraction dataset was automatically integrated, merged and scaled by the DLS pipelines. Datasets with the best diffraction statistics were chosen for structure solution.

4.3.6 Molecular replacement

Molecular replacement with PHASER utilised the previously solved 2.10 A° resolution structure of PNGM-1 - deposited in the PDB under accession code 6J4N by M.K. Hong et al in 2019[301] - as a search model. The solution was then subjected to iterative rounds of automatic refinement using phenix.refine and manual model refinement using WinCoot.

4.3.7 Large scale expression and purification of MβLp01 and MβLp02 as HP-Trx fusion proteins

For large-scale expression, 6L (1L x 6) LB supplemented with kanamycin (50ug/mL) was inoculated with an appropriate volume of an overnight culture of a single colony of *E. coli* TOP10 transformed with pBAD202/D-TOPO_MβLp01 or pBAD202/D-TOPO_MβLp02. The culture was maintained at 37°C with shaking (180rpm) until the OD_{600} reached ~0.6, at which point L-arabinose (0.2%) was added. After overnight growth at 16-18°C with shaking (180rpm), cells were harvested (5000 x g, 10 min, 4°C) and then resuspended (after discarding the spent media) in a lysis buffer consisting of 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 20 mM Imidazole and 0.5 mM $ZnSO_4$. After 1-2h at -80°C, the cell suspension was thawed and added to it was a protease inhibitor tablet (cOmplete EDTA-free, Roche), bovine DNase I (~10 µg/mL) and lysozyme (chicken egg white, 0.2 mg/mL). After 30 min on ice with agitation, cells were subjected to mechanical lysis by cell disruption either at 18,000 psi on an LM20 microfluidizer (Analytik Ltd), or 10,000 psi on a French Press. The lysate was centrifuged (48,000 x g, 30 min, 4°C) and the clarified lysate, after filtration through a 0.45um filter, was loaded onto either a Ni^{2+}-NTA affinity column (HisTrap HP 5mL, Cytiva) or a Zn^{2+}-NTA column (1mL, Cube Biotech) using an AKTA Pure (Cytiva) chromatography system. The resin was first equilibrated with X CV of IMAC wash buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 20mM Imidazole, 0.5 mM $ZnSO_4$) at a flow rate of 5 mL/min, before application of the sample. Unbound proteins were washed off the resin with X CV of wash buffer, before elution of the target protein with a gradient from 20 to 500 mM imidazole, at a flow rate of 1 mL/min. Fractions containing the target protein as identified by SDS-PAGE (4-12% Bis-Tris Bolt™ gels) were pooled and then concentrated in a spin concentrator device (3000 or 5000 kDa MWCO; Amicon) until the volume was 2-2.5 mL. This sample was then injected onto a HiLoad 16/600 Sepharose 75 pg column (Cytiva) for separation by size exclusion chromatography, in a buffer composed of 50 mM Tris-HCl pH 7.5, 150 mM NaCl, either including or excluding 0.5 mM $ZnSO_4$, circulated at 1 mL/min for 1 CV. SDS-PAGE was again used to identify the target protein-containing fractions, which were pooled and concentrated (3000 kDa MWCO) before usage and storage. Concentration was estimated using

a NanoDrop™ One spectrophotometer. For storage, glycerol was added to a final concentration of 30% and aliquots were flash frozen in LN_2 before storage at -80°C.

4.3.8 Cloning, expression, and purification of HP-Thioredoxin

PCR mutagenesis was carried out on the pBAD202/D-TOPO_MβLp01 plasmid to convert the start codon for MβLp01 translation to a stop codon. PCR was carried out in the same way as for MβLp01 and MβLp02 but with mutagenic primers (Table 4.4) and corresponding appropriate thermocycling conditions. Primers were obtained in the dry, desalted form from Eurofins. Once DNA agarose gel electrophoresis verified the presence of a PCR product of the expected size, *E. coli* One-shot TOP10 cells were transformed with the PCR product and single colonies from overnight growth at 37°C with kanamycin (50ug/mL) selection were used to generate glycerol stocks and extract propagated plasmid for stocks.

XL LB supplemented with kanamycin (50ug/mL) was inoculated with an inoculum from an overnight culture of a single colony of *E. coli* TOP10 carrying pBAD202/D-TOPO_HP-Trx and incubated at 37°C with shaking (180 rpm). When the culture reached OD_{600} ~0.6, L-arabinose was added to a final concentration of 0.2% before incubation at 16-18°C with shaking (180 rpm) overnight. Cells were harvested by centrifugation (5000 x g, 10 min, 4°C) and the spent media discarded. Cell pellets were resuspended in a lysis buffer consisting of 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 20 mM Imidazole and 0.5 mM $ZnSO_4$ and stored at -80°C for 1-2h. Once thawed, a protease inhibitor tablet (cOmplete EDTA free), lysozyme (chicken egg white, 0.2 mg/mL) and bovine DNase I (~10ug/mL) were added. After 30 min on ice with agitation, cells were subjected to mechanical lysis by cell disruption at 18,000 psi on an LM20 microfluidizer (Analytik Ltd) before clarification by centrifugation (48,000 x g, 30 min, 4°C). The clarified lysate was filtered through a 0.45um filter before being loaded onto a Ni^{2+}-NTA affinity column (HisTrap HP 5mL, Cytiva) using an AKTA Pure (Cytiva) chromatography system. The resin was first equilibrated with X CV of Ni^{2+}-NTA wash buffer (50mM Tris-HCl pH 7.4, 150mM NaCl, 20mM Imidazole, 0.5 mM $ZnSO_4$) before application of the sample. The resin was washed with X CV of wash buffer, before gradient elution of the target protein with an elution buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 500 mM Imidazole and 0.5 mM ZnSO4. Fractions containing the target protein were identified by SDS-PAGE (4-12% Bis-Tris Bolt™), pooled, and then concentrated in a spin concentrator device (3000 or 5000 kDa MWCO; Amicon). Once the volume was 2-2.5 mL, it was injected onto a HiLoad 16/600 Sepharose 75 pg column (Cytiva). The size exclusion buffer was composed of 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.5 mM $ZnSO_4$ and was run through the column at 1 mL/min for 1 CV. SDS-PAGE was

again used to identify the target protein-containing fractions, which were finally pooled and concentrated (3000 kDa MWCO) before usage and storage. Concentration was estimated using a NanoDrop™ One spectrophotometer. For storage, glycerol was added to a final concentration of 30% and aliquots were flash frozen in LN_2 before storage at -80°C.

Primer Name	Primer Sequence
MβLp01 Mutagenic Forward	**5'-CTTCACCCATTAAGCTTCACCCGTAACACAAGTTGTCG-3'**
MβLp01 Mutagenic Reverse	**5'GGTGAAGCTTAATGGGTGAAGGGATCAATTCC-3'**

Table 4.4 Sequences of the oligonucleotides used in the PCR mutagenesis of MβLp01

4.3.9 Determination of phosphomonoesterase activity of MβLp01 and MβLp02 HP-Trx fusions

4.3.9.1 Molybdenum Blue Method

Phosphatase activity (including phytase activity) was determined by detection of released inorganic phosphate according to the molybdenum blue method (chapter 2.6). Substrates tested at a concentration of 10 mM were phytic acid dipotassium salt (also tested at 2.5 mM), IP4 (produced though hydrolysis of IP6 by Quantum Blue), ATP, ADP, glucose-6-phosphate and glucose-1-phosphate. Reactions were performed in triplicate in a volume of 50 µL. Purified MβLp01 or MβLp02 (various concentrations), plus samples from different stages of purification, were pre-incubated for 3 min at the specified temperature in 0.2M sodium acetate buffer pH 5 (or 5.5) before addition of substrate. Reactions were then incubated for 30 min at various temperatures (25, 30, 37, 40, 50°C) before inactivation by incubation for 10 min at 100°C. Reaction solutions were then transferred to a 96 well plate before the addition of 50 µL freshly prepared developing solution (4:1 ammonium molybdate solution: acidified ferrous sulphate solution). Absorbance was recorded at 700 nm on a compact microplate reader (Byony Absorbance 96) after 20 min incubation at RT. Reactions of non-recombinant *E. coli* TOP10 cells (clarified lysate) and of purified HP-thioredoxin were run alongside as background controls. Substrate only and buffer only reactions were run to measure substrate background contribution. The average A_{700nm} value for substrate only was subsequently subtracted from each experimental value to reflect the P_i response which was not resulting from the substrate itself. A phosphate calibration curve was generated by measuring the A_{700nm} of potassium phosphate standards (7.8 – 1000 µM). Incubation of reaction solutions was performed in a Thermocycler C1000 Touch (Bio-Rad).

4.3.9.2 pNPP Hydrolysis Assay

One 'Phosphatase substrate' 5 mg tablet (Sigma) was dissolved in 1.35 mL dH_2O to generate a 10 mM working stock which was used immediately. Excess solution was discarded after completion of each assay due to the short (12h) stability of the solution. Reactions of 50 µL volume consisted of the pNPP substrate at a final concentration of 1 mM, and MβLp01 at concentrations of 1, 2, 5 and 10 µM, in 0.2M sodium acetate pH 5.4 buffer. The positive control (*Bt*MINPP) was therefore tested at the same concentrations. Reaction solutions were incubated in PCR tubes in a Thermocycler C1000 Touch (Bio-Rad) for 30 min at 37°C before transferring to a 96 well plate and addition of 2.5N NaOH. After 10 min at RT, the absorbance at 405 nm was measured using a compact microplate reader (Byony absorbance 96). In parallel, blanks of buffered substrate, assay buffer only and MβLp01 SEC buffer only were subjected to the same assay conditions. A_{405nm} values were corrected for the substrate background absorbance.

4.3.10 Determination of phosphodiesterase activity of MβLp01 HP-Trx fusion

Reactions of 50 µL volume consisted of bis-pNPP at a final concentration of 10 mM, and MβLp01 at concentrations of 1, 2 and 5 µM, in 0.2M sodium acetate pH 5.4 buffer. The positive control (*E. coli* AppA) was therefore tested at the same concentrations. Reaction solutions were incubated in PCR tubes in a Thermocycler C1000 Touch (Bio-Rad) for 30 min at 37°C before transferring to a 96 well plate and addition of 2.5N NaOH. After 10 min at RT, the absorbance at 405 nm was measured using a compact microplate reader (Byony absorbance 96). In parallel, blanks of buffered substrate, assay buffer only and MβLp01 SEC buffer only were subjected to the same assay conditions. A_{405nm} values were corrected for the substrate background absorbance.

4.3.11 Investigation of β-lactamase activity of MβLp01 and MβLp02

4.3.11.1 Nitrocefin hydrolysis assay

Nitrocefin is a chromogenic cephalosporin, the hydrolysis of which generates a coloured product which is detectable at 490 nm and is directly proportional to the amount of β-lactamase activity. This assay was performed using a commercial kit (Abcam ab197008) which provides a positive control and was carried out according to the manufacturer's instructions. Reactions were performed using the proprietary buffer included. First, a standard curve of hydrolysed nitrocefin was generated using the provided hydrolysis buffer. Purified MβLp01 was tested along with *E. coli* TOP10 cell lysates expressing MβLp01 or MβLp02. The empty

vector pBAD202/D-TOPO and non-recombinant *E. coli* TOP10 cell lysates were run alongside as controls, as was the proprietary positive control. Absorbance was continuously monitored at 490 nm over the course of 60 min at 25°C. The A_{490nm} value corresponding to the 0 nm hydrolysed nitrocefin standard was subtracted from all the experimental values to correct for background absorbance.

4.3.11.2 Antimicrobial susceptibility testing

Disc diffusion method

For the disc diffusion method, sterile discs (Oxoid) were impregnated with 20µL of a β-lactam antibiotic – carbenicillin (Formedium), ceftazidime (MedChem Express) or doripenem (hydrate; Acros Organics) at various concentrations (Table 4.5), either with or without an aliquot of purified MβLp01 (5 µL), and incubated at 37°C overnight. The following day, a culture of *E. coli* DH5α was grown until an OD_{600} of approximately 0.5 and 100 µL spread onto LB agar plates (all 25 mL agar) in triplicate. Immediately, the pre-incubated sterile discs were positioned using sterile tweezers into separate quadrants of the surface of the plate such that each plate contained four discs consisting of different concentrations of the same β-lactam, plus a control disc in the centre containing pre-incubated buffer only. This was performed in triplicate for each concentration. Plates were incubated at 37°C overnight (~18h). After incubation, the zones of inhibition were measured manually and the surface of the agar plates were imaged in a gel documentation system (G:BOX Chemi XRQ, Syngene).

B-lactam antibiotic	Quantities tested (µg)
Carbenicillin	10, 20, 30, 40, 50
Ceftazidime	0.75, 0.50, 0.25, 0.2, 0.15, 0.1
Doripenem	2, 1, 0.5, 0.25

Table 4.5 Quantities of each β-lactam antibiotic tested in final disc diffusion experiments, after refinement through preliminary tests.

Agar dilution method

For the agar dilution technique, cultures of MβLp01 or MβLp02 in *E. coli* TOP10 were grown until an OD_{600} of approximately 1.0, diluted 1:1 and 100/200 µL spread onto LB agar plates (all 25 mL agar) containing several different concentrations of β-lactam antibiotics along with 10 µg/mL kanamycin and various L-arabinose concentrations (0, 0.002, 0.02, 0.2%), in duplicate.

Plates were incubated at 37°C overnight. In parallel the same was set up for a culture of *E. coli* TOP10 transformed with pBAD202/D-TOPO, and of non-recombinant *E. coli* TOP10. After incubation, the surface of the agar plates were imaged in a gel documentation system (G:BOX Chemi XRQ, Syngene) and colonies were counted visually.

B-lactam antibiotic	Concentrations tested (µg/mL)
Carbenicillin	1, 5, 10, 15, 20, 25, 30
Ceftazidime	0.5, 0.75, 1, 2, 4, 8
Doripenem	0.0125, 0.025, 0.05, 0.075, 0.1, 0.5, 1, 2, 4, 8

Table 4.6 Concentrations of each β-lactam antibiotic tested in the agar dilution method

4.3.12 Metal ion cofactor analysis of MβLp01

Pure MβLp01 samples were subjected to ICP-MS-QQQ to quantify the $Ni^{2+}$and Zn^{2+} present. Both de-tagged MβLp01 and HP-thioredoxin tagged MβLp01 were analysed. Samples were thermally digested by addition of high purity concentrated nitric acid (1mL) before being run through an Ultrawave digester on a program ramped for 15 min to 240°C, held for a further 15 min then cooled to 40°C. 1 ml of internal standard solution was added and the volume made up to 10 ml. ICP-MS settings:- Power 1550W, Cooling gas 14 l/min, Nebulizer 1.1 l/min, Auxiliary 0.8 l/min, Sample depth 15 mm, Spray Chamber 2.7°C and Detector Voltage 1630 V.

4.3.13 Mass spectrometry of MβLp01

4.3.13.1 MALDI-TOF-MS analysis

Both de-tagged MβLp01 and the HP-thioredoxin fusion were run. Protein at 1-10 pmol/uL in acetonitrile/H_20/TFA 30:70:0.1. The target was pre-coated with 0.5 uL matrix solution (sinapinic acid in a saturated solution of acetonitrile:H_2O:TFA 50:50:0.1) for 10 sec before the excess solvent was removed. Samples of 0.5uL were then mixed 1:1 with matrix solution.

4.3.13.2 LC-QTOF analysis

4.4 Results

4.4.1 Trial expression of MβLp01 from pET15b

The MβLp01 gene (MH367836.1) was modified by removal of the signal peptide and codon optimised for *E. coli* expression before gene synthesis and cloning into pET15b by Genscript (Figure 4.4a). First, this construct was used to transform *E. coli* chemically competent One shot™ TOP10 cells for plasmid propagation and preparation of glycerol stocks. To assess expression of MβLp01 as a hexahistidine tagged protein from the pET15b vector (32.7 kDa), the *E. coli* IPTG-inducible T7 expression strains, BL21 and SHuffle® T7 Express (NEB), were transformed with the pET15b_MβLp01 construct, and a small-scale (100 mL) expression was conducted. Attempts to purify MβLp01 from BL21 lysate by Ni^{2+}-NTA IMAC were unsuccessful due to the lack of soluble protein (as shown by SDS-PAGE; Figure 4.4b) therefore the decision was to proceed to clone MβLp01 into the pBAD202/D-TOPO vector, as per the Villamizar group[267]. Upon retrospection, the reason for the apparent insolubility of MβLp01 expressed from pET15b in BL21 could be a fault of the lysis procedure for this particular experiment, which consisted of chemical lysis by the BugBuster® protein extraction reagent. Lysis by a combination of chemical and physical methods (such as cell disruption or sonication) may have yielded a higher ratio of soluble to insoluble target protein, enabling subsequent purification.

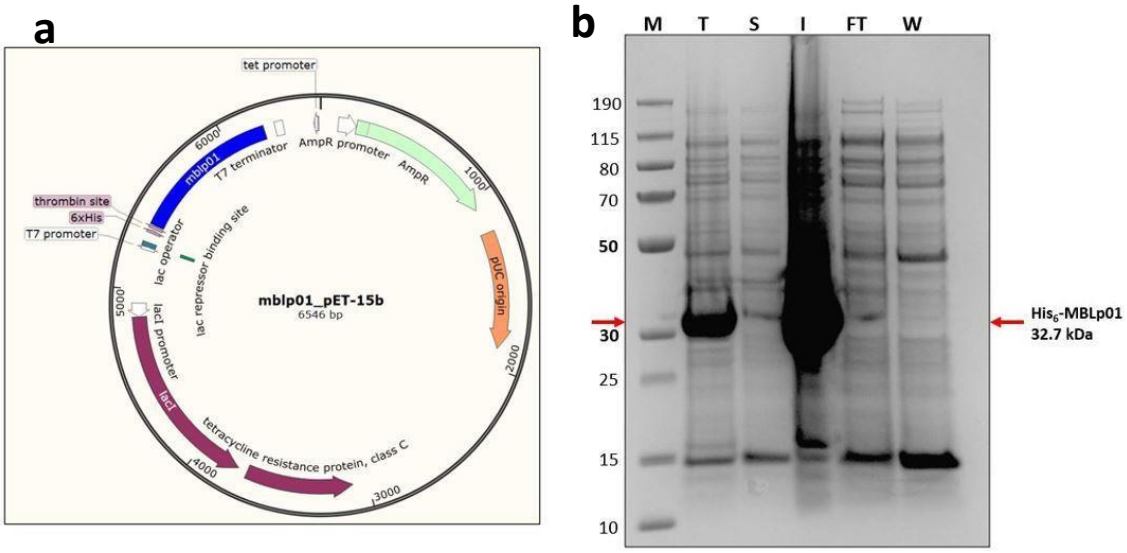


Figure 4.4 a) Plasmid map of MβLp01 as cloned into the pET15b vector; b) SDS-PAGE of the protein samples from the small-scale trial expression of MβLp01 from pET15b, in *E. coli* BL21.
M= molecular weight marker; T = total protein; S = soluble protein; I = insoluble protein; FT = IMAC flow-through; W = IMAC wash

4.4.2 Cloning of MβLp01 into pBAD202/D-TOPO

The pBAD202/D-TOPO vector encodes an N-terminal His-patch thioredoxin peptide which is commonly used as a fusion partner in recombinant protein expression in *E. coli* (Figure 4.5). Thioredoxin is the 11.7 kDa protein product of the *E. coli* trxA gene. It has multiple properties which facilitate high-level production of soluble fusion proteins in the *E. coli* cytoplasm. Fundamentally, it has been shown to accumulate up to 40% of the total cellular protein whilst remaining fully soluble, and this property - a product of particularly efficient translation - is transferred to the protein fused to thioredoxin at the N-terminus[302-304]. His-patch thioredoxin is an engineered form of the protein in which a cluster of surface amino acid residues (E30, Q62, S1) have been mutated to histidine in order to introduce affinity for metal chelate column matrices, facilitating purification by IMAC[305]. Upon folding of the 11.9 kDa protein, the two mutant histidine residues interact with a native histidine at residue 8 to form the histidine patch, in which the three residues can coordinate a divalent cation[305]. In the pBAD202/D-TOPO vector, the trxA translation termination codon is replaced by DNA encoding a ten-residue peptide linker sequence which includes an enterokinase (enteropeptidase) cleavage site, before the target gene sequence. This permits cleavage of the His-patch thioredoxin tag from the target protein during purification using the enterokinase protease. Enterokinase is a calcium dependent, highly specific serine protease which recognises the amino acid sequence Asp-Asp-Asp-Asp-Lys and cleaves after the lysine. Although 26.3 kDa is the theoretical molecular weight, the apparent molecular weight is 31 kDa[306]. Expression from pBAD202/D-TOPO is driven by the *ara*BAD promoter (P_{BAD}), which is positively regulated by the araC gene product[307]. This pBAD vector is inducible by L-arabinose and carries a kanamycin resistance marker. Although the vector also supplies a V5 epitope and a hexahistidine tag at the C-terminus, a stop codon was inserted before the V5 epitope during the cloning since these were unnecessary and would need to be removed before crystallisation, regardless.

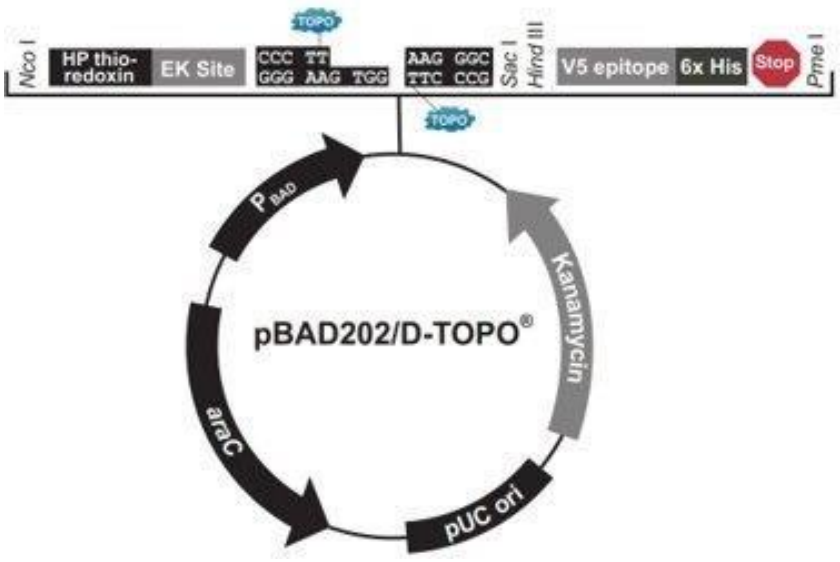


Figure 4.5 Plasmid map overview of the pBAD202/D-TOPO® vector, highlighting the key features. Expressed proteins are fused at the N-terminus to a His-patch thioredoxin leader peptide which is followed by an enterokinase cleavage site before the target protein. The vector also encodes a C-terminal V5 epitope followed by a hexahistidine tag, and harbours Kan^R.

MβLp01 was cloned into the pBAD202/D-TOPO vector by directional TOPO cloning (protocol chapter 2.2.1), using the pBAD202 Directional TOPO™ expression kit (Invitrogen™), according to the manufacturer's instructions. This one-step cloning method facilitates the insertion of a DNA sequence into the vector with the correct orientation and high efficiency. First, the target gene is amplified by PCR with appropriately designed primers which must contain a 5'-CACC overhang and a blunt 3' end. The resulting PCR product is then used directly in the TOPO cloning reaction, which ligates the insert and the linearised topoisomerase I-activated vector. Two pairs of PCR primers were designed and used (sequences included in chapter 4.3.1). Gradient PCR yielded PCR products of the expected size (855/846 bp) for both primer pairs and all four temperatures (50, 55, 60, 65°C) employed (Figure 4.6a). These were then used in the cloning reaction and the resulting product was propagated by transformation of *E. coli* TOP10 and extracted from transformants by a commercial miniprep kit (Qiagen). To evaluate the success of the cloning, a diagnostic restriction digest was carried out in which the pBAD202/D-TOPO_MβLp01 plasmid was cut with restriction endonucleases in both a single and a double digest and the products separated on a DNA agarose gel and visualised (Figure 4.6b). A single digest (in which the plasmid is cut with one restriction enzyme only) linearises the plasmid such that its size can be estimated relative to the marker - which is composed only of linear DNA fragments. A double digest indicates the presence, size and correct orientation of the insert by generating bands of characteristic sizes visible on a DNA agarose gel. Restriction enzymes were therefore chosen based on location of restriction sites in the plasmid such that

the plasmid is cut once within the vector and once within the insert. HindIII and Pm1I were chosen for the diagnostic restriction analysis of pBAD202/D-TOPO_MβLp01 - HindIII for the single digest and HindIII/Pm1I for the double digest. Pm1I cuts in the first 300 bp of the insert which allows the orientation of the insert to be indicated – however, absolute verification must be obtained through sequencing.

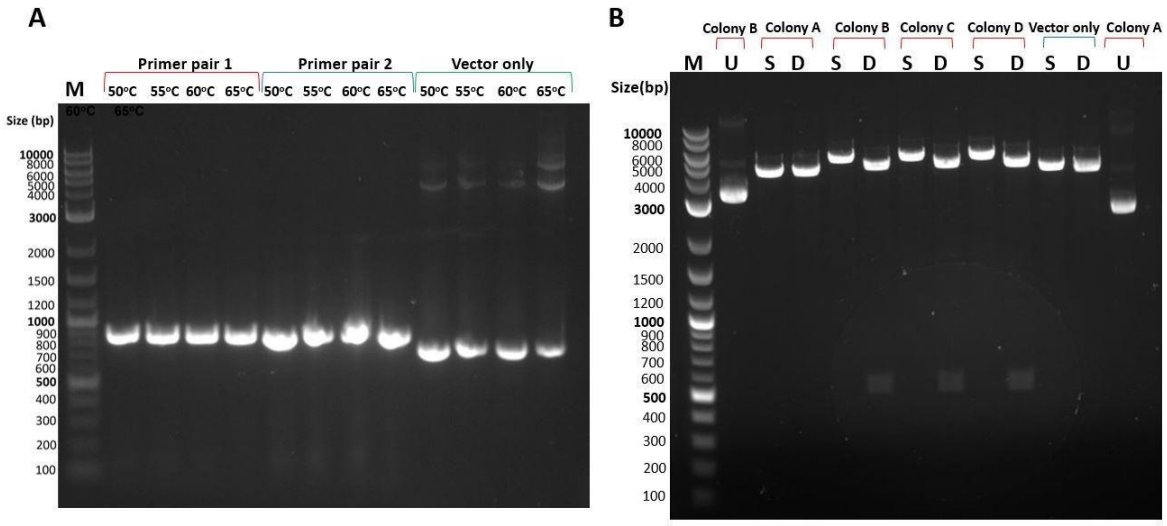


Figure 4.6 Separation of PCR products and extracted plasmids by DNA agarose gel electrophoresis. A) PCR products generated by gradient PCR with two different primer pairs and four different temperatures, alongside a vector only control. Expected size of MβLp01 insert: 855/846 bp; negative control: 755 bp. B) Diagnostic restriction digest of the MβLp01-containing pBAD202/D-TOPO plasmid extracted from *E. coli* TOP10, using HindIII and Pm1I. A+B) Ladder: Quick-load purple 1kb plus (NEB); 1% agarose in 1X TAE buffer containing 0.5 mg/mL EtBr; 1X TAE running buffer; samples mixed with GelPilot loading dye (Qiagen) and 5-10 µL loaded; gels run at 100-120 V, ~60 min. M= DNA size marker; U = uncut; S = single digest; D = double digest.

Four different *E. coli* TOP10 transformants were used for plasmid extraction and of these, three were successful, as indicated by the double digest (Figure 4.6b). The plasmids linearised by digestion with HindIII ran at the expected size of 5303 bp (lanes 'S' for colonies B, C, D) and for these three samples the double digest (lanes 'D' for colonies B, C, D) produced the two predicted fragments of 4742 bp and 561 bp as expected on the basis of the restriction sites of HindIII and Pm1I (Figure 4.7). It is evident that colony A yielded the vector lacking the insert, because the linearised plasmid band (lane 'S' for colony A) appeared at the same size as the linearised vector only control - both at 4448 bp (lane 'S' for vector only). Further, the double digest for this sample (lane 'D' for colony A) produced only the one same fragment as in the single digest which shows that it lacked the insert and therefore lacked the Pm1I restriction site. The uncut plasmid has an apparent size of ~3500bp (lane 'U' for colony B) and the difference in migration between the uncut and linearised plasmid can be seen by comparison of lanes 'U' and 'S' for

colony B. The three plasmids identified as having the correct restriction pattern were verified by sequencing (Appendices A4.1).

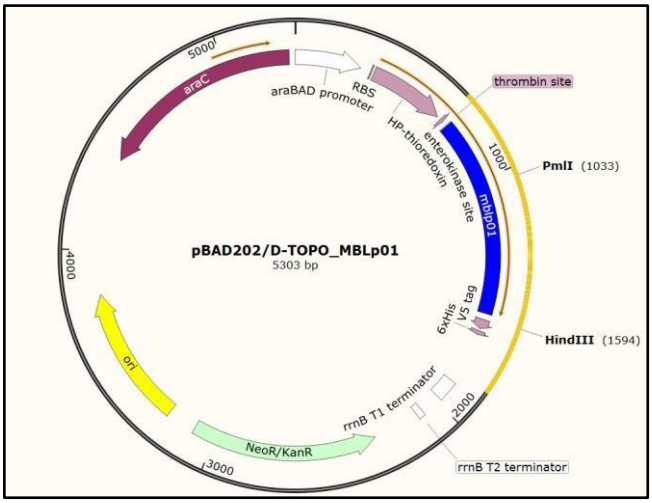


Figure 4.7 Plasmid map of MβLp01 as cloned into the pBAD202/D-TOPO vector. The restriction sites of HindIII and Pm1I are indicated. The difference in size between these two restriction sites is 561 bp, which correlates with the smaller fragment produced in the double digest.

When checking the size of a plasmid on a DNA agarose gel, it is advisable to first linearise the plasmid by a single restriction digest since plasmid DNA is present in multiple distinct conformations which are revealed by bands of different sizes on a gel. This concept is illustrated in Figure 4.8. Uncut plasmid consists primarily (around 95%)[308] of supercoiled (covalently closed circular) DNA and this form of DNA has the greatest electrophoretic mobility of the different forms, migrating faster by around 30% relative to linear DNA[309]. As such, this form of DNA cannot be compared to a standard DNA ladder since the predominant supercoiled plasmid band will have a much smaller apparent size. Plasmid samples also contain a small amount of nicked DNA which migrates the slowest as it is the most 'relaxed' form. All these different forms can be seen in Figure 4.8. An alternative to linearising the plasmid (for example to conserve plasmid stock) is to use a supercoiled DNA ladder such as that which is available from New England Biolabs. Perhaps an overkill, both a standard ladder and a supercoiled ladder can be run alongside plasmids to be analysed.

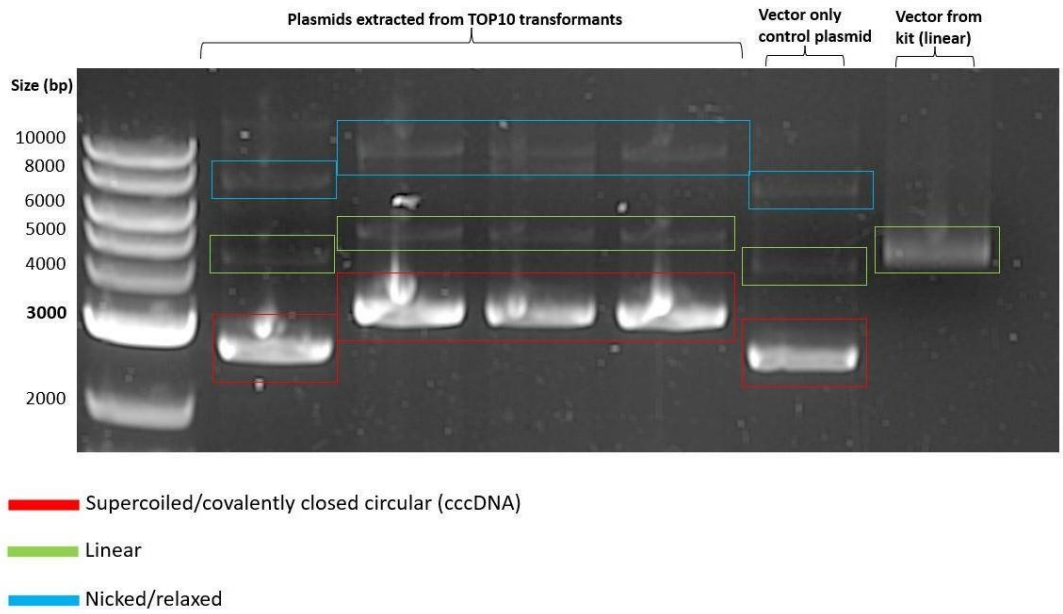


Figure 4.8 Example of the distinct DNA conformations present within a plasmid DNA sample.

4.4.3 Expression trials of MβLp01

The recommendation for expression of proteins using the arabinose inducible pBAD system is to use an *E. coli* strain which is araBADC- and araEFGH+ to permit transportation of L-arabinose without metabolizing it. The *E. coli* strains LMG194 and TOP10 are two such suitable strains. Since LMG194, a K-12 derivative, grows well in minimal media, it can be used to lower basal expression levels of potentially toxic genes by the supplementation of RM media with glucose which serves to further repress P_{BAD}[307]. The genotype of LMG194 is:

F⁻ Δ*lac*X74 *gal*E *thi* *rps*L(Str^R) Δ*pho*A (*Pvu*II) Δ*ara*714 *leu*::Tn*10*

and it is streptomycin and tetracycline resistant. However, this strain is not endA- or recA- which may decrease the stability of glycerol stocks[310]. TOP10 on the other hand may be used for expression in addition to general cloning, which has the advantage of time saved in the ability to transition directly from cloning to expression without additional transformation steps. The genotype of TOP10 is:

F⁻*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻*rps*L(Str^R) *end*A1 *nup*G

which is almost identical to the DH10B strain[310]. The difference between the two strains is essentially the stringency of inducible expression. Thus, if a heterologous protein is suspected to be toxic to the cell, then LMG194 should be used, otherwise TOP10 is equally effective[307].

For expression of MβLp01, due to complications with the competency of LMG194, and the observed lack of host toxicity, TOP10 was chosen as the expression strain. Preliminary expression trials were conducted to ascertain the optimum conditions for expression, varying L-arabinose concentration and temperature of induction. Figure 4.9 represents an expression trial in which three L-arabinose concentrations (0.2%, 0.02%, 0.002%) were tested each at two temperatures (18°C and 37°C), alongside an uninduced sample. The cultures which were induced at 18°C were lysed by two different methods for comparison, since the level of soluble overexpressed protein may be obscured by the inefficiency of a particular lysis method. The two lysis methods trialled were chemical lysis by the BugBuster protein extraction reagent, and a combination of chemical and physical lysis by incubation with lysozyme followed by sonication. There are multiple protein bands in the expected molecular weight region of 44.2 kDa which renders the comparison of the different induction conditions challenging and therefore it was difficult to choose a temperature/concentration combination based on this. The large proportion of lysozyme in the SDS-PAGE samples which were lysed by sonication/lysozyme further conceals the target protein, since the lysozyme takes up a large proportion of the total protein, resulting in less intense bands for the other proteins in the sample. The only method to unambiguously identify which of the bands are the target protein is immunodetection using an anti-Trx antibody for binding to the HP-thioredoxin tag. However, a larger scale expression test at 18°C and 0.2% L-arabinose was successful and so these were the expression parameters selected to proceed with. In general, lower temperatures result in more soluble target protein as the cell machinery has more time to properly fold the protein, whereas higher temperatures (30-37°C) often result in the target protein being produced in insoluble inclusion bodies. Therefore, 18°C was chosen as a sensible temperature.

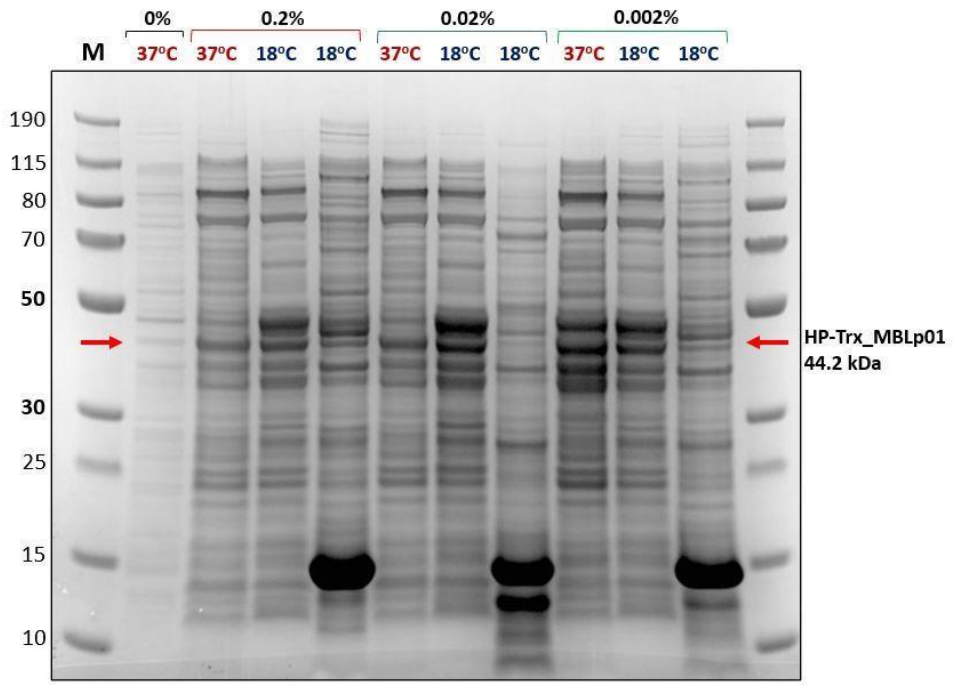


Figure 4.9 Assessment by SDS-PAGE of the expression trial conducted for MβLp01 expressed from *E. coli* TOP10, at combinations of temperature (37°C, 18°C) and L-arabinose concentrations (0, 0.2, 0.02, 0.002%).

As for the concentration of inducer, it is important to explain the nature of the araBAD system. The promoter P_{BAD} is subject to all-or-none induction whereby intermediate concentrations of L-arabinose result in subpopulations of cells that are either fully induced or completely uninduced, as such reflecting the *proportion* of cells that are fully induced rather than the level of intermediate expression in any individual cell[312,313]. Therefore, gene expression is not uniform with respect to individual cells.[314] It was demonstrated by Guzman et al that expression of a target protein from this system could be modulated over several orders of magnitude in cultures grown in the presence of sub-saturating concentrations of L-arabinose[307]. Since the population of cells is a mixture of induced and uninduced, only the fraction of induced cells increases with increasing L-arabinose concentrations[313]. Regardless, the overall expression level of the culture is known to correlate with increasing L-arabinose concentrations and is reported to provide a linear induction response and corresponding titratable expression levels which can be used to precisely modulate desired protein levels for optimum yield of soluble protein[307]. However, from expression trials of MβLp01, this was not observed for the soluble protein fraction. An SDS-PAGE comparing the total, soluble and insoluble protein fractions, however, did show an increasing intensity of suspected target protein proportionate to inducer concentration, but only in the total protein and insoluble protein fractions (Figure 4.10). The disparity in intensity between the overexpressed band in the total and insoluble samples in

comparison to that of the soluble sample suggests that either the MβLp01 fusion is insoluble under this range of induction conditions, or the cell lysis was not optimum. For the subsequent larger scale preparations, the cells were lysed using a high-pressure cell homogenizer which proved much more effective. With regard to the intense bands observed in both Figure 4.9 and Figure 4.10, around 15 kDa, it is likely that this is lysozyme (14.6 kDa) which was added to aid cell lysis (and this is absent from the total protein sample, since the lysozyme sample hadn't yet been added).

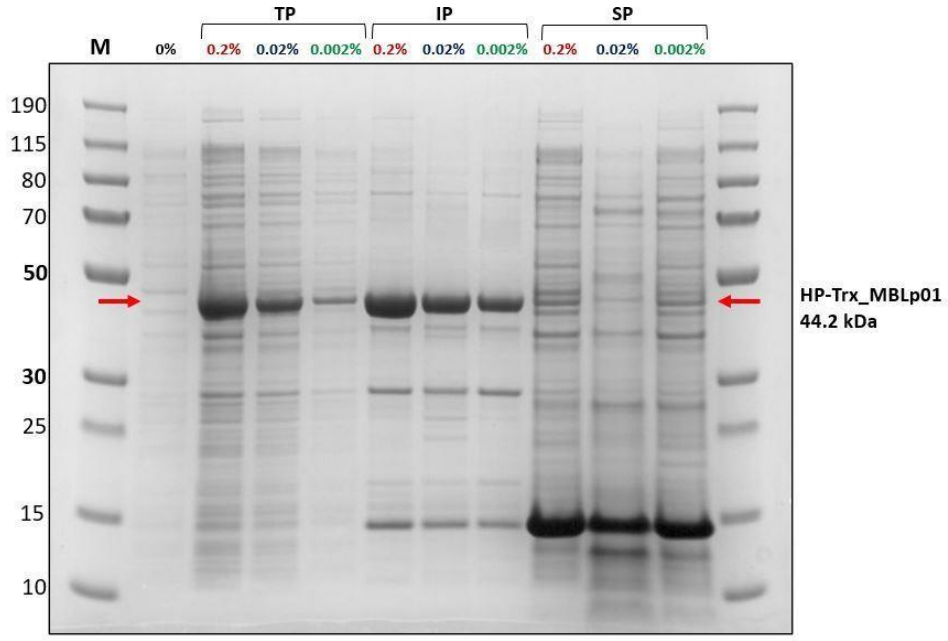


Figure 4.10 Assessment by SDS-PAGE of the solubility trial conducted for MβLp01 expressed from *E. coli* TOP10. Solubility was evaluated as a function of L-arabinose concentration, at a temperature of 18°C. Protein samples tested were total cellular protein (TP), insoluble protein (IP) and soluble protein (SP) i.e. clarified cell lysate.

4.4.4 Large-scale expression and purification of MβLp01

Multiple purifications were performed for MβLp01 in order to optimise yield and purity, with various buffer compositions and chromatographic steps employed, yielding different outcomes with regard to yield and purity. First, the purification procedure which gave rise to protein crystals will be reported.

MβLp01 was overexpressed and purified as described in chapter 4.3.3. A representative SDS-PAGE gel of fractions from Ni^{2+}-NTA IMAC is illustrated in Figure 4.11, alongside the corresponding chromatogram. It can be observed that MβLp01 elutes as a single band at the expected molecular weight (44.2 kDa). The intensity of the overexpressed MβLp01 bands

correlate with the intensity of absorption in the peak observed in the chromatogram. Here, 9 of the 11 fractions were pooled, omitting the 2 fractions with the least intense overexpressed band (the first two lanes) so as not to unnecessarily dilute the sample pool. A small amount of target protein eluted in the wash and this is likely to be because the column capacity was exceeded. To increase yield, the wash could have been recirculated through the resin. After IMAC, the yield was approximately 8mg but some of this represented non-specific host proteins.

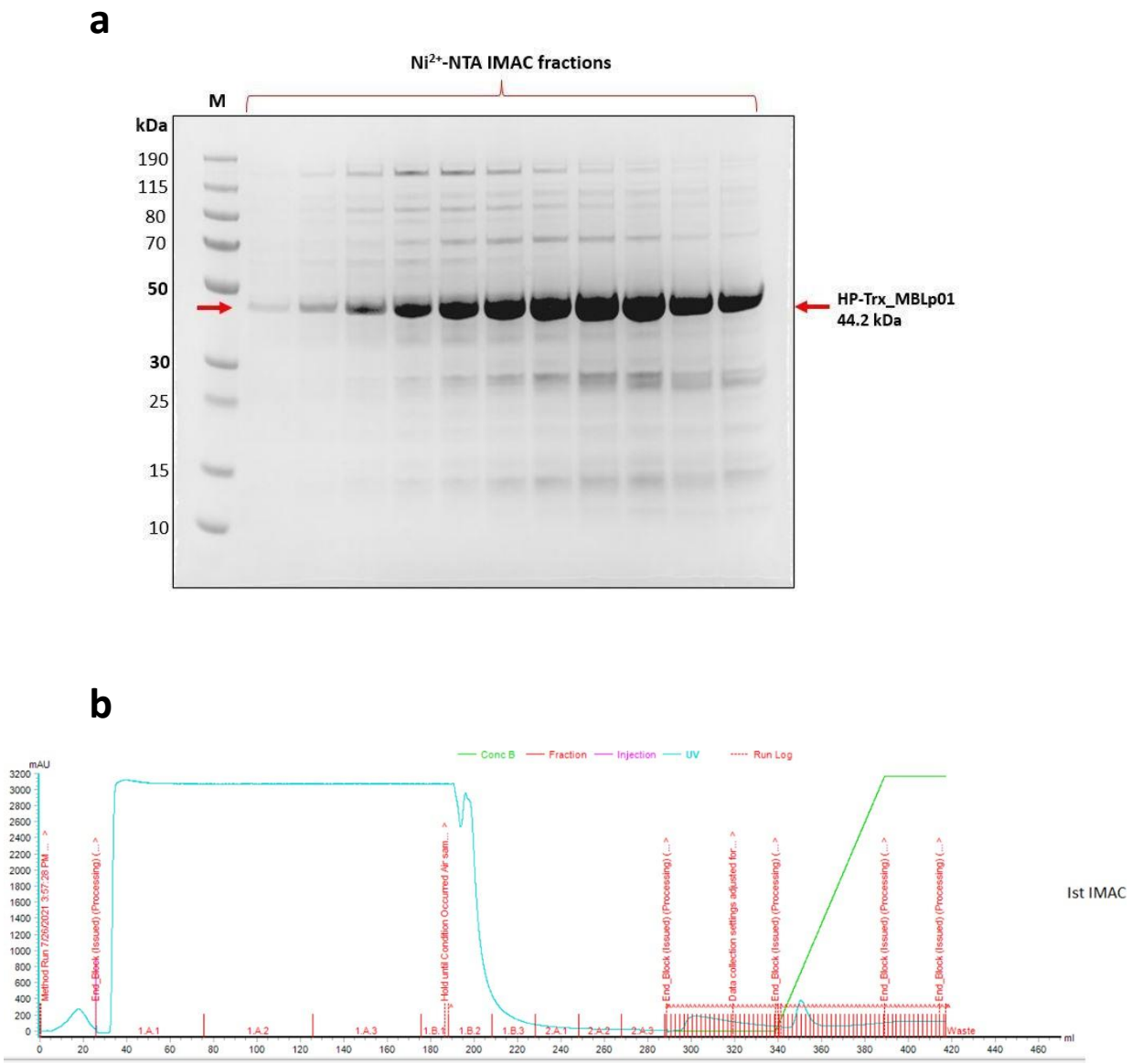


Figure 4.11 Fractions from Ni^{2+}-NTA IMAC of MβLp01, separated by SDS-PAGE (a), along with the corresponding UV_{280nm} chromatogram trace (b).

As a pre-requisite to cleavage of the HP-thioredoxin tag, a cleavage trial was set up to ascertain the parameters required to successfully cleave the tag, i.e. amount of enterokinase per protein weight, incubation temperature and time. The manufacturer of the enterokinase product used recommends performing the cleavage reaction at 25°C although the temperature may be increased to increase the rate of reaction. However, increasing the temperature runs the risk of denaturing the enterokinase and/or the target protein. On the other hand, the reaction could be performed at a lower temperature such as 4°C in an attempt to retard proteolytic degradation, however this would then require a longer incubation time to compensate for the reduced activity of enterokinase at lower temperature. As such, a balance has to be sought between rate of cleavage (desired) and rate of degradation of target protein (undesired). Ultimately the aim is to accomplish complete cleavage of the fusion protein in the minimum time frame at an ambient temperature. For the cleavage trial, three different ratios of MβLp01 to enterokinase were set up and reactions were incubated at 25°C for 24h taking samples at specific time points. The result of this cleavage trial is presented in Figure 4.12 – there did not appear to be any difference between the MβLp01: enterokinase ratio or between the time points. As such, there was no advantage to incubating the reaction for longer than 1h. There is a very faint band in all samples for the remaining uncleaved HP-Trx fusion, but this is negligible compared to the intensity of the cleaved fusion protein band, and since this did not decrease in 24h, it's highly unlikely that any further cleavage would occur if incubated for upwards of 24h. To conserve the enterokinase, the minimum amount of enterokinase was chosen to proceed with.

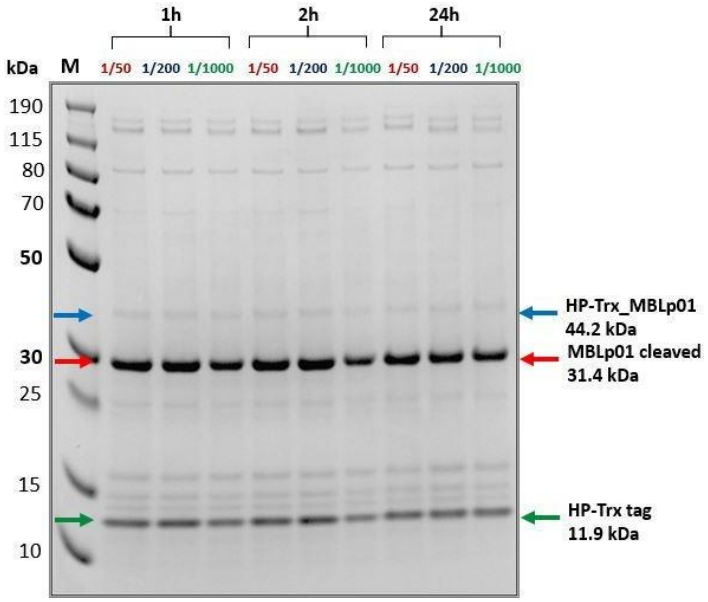
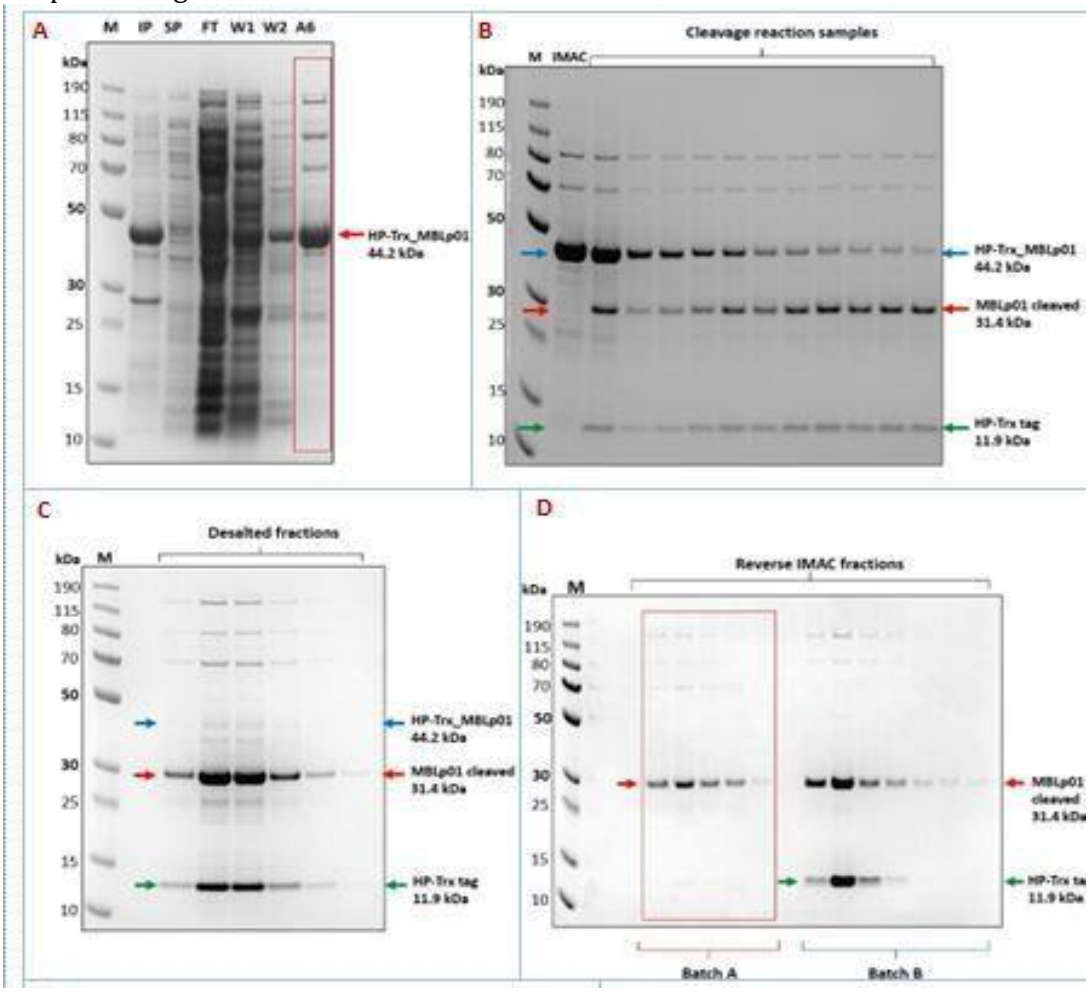


Figure 4.12 Results of the trial cleavage of MβLp01 with enterokinase to ascertain suitable conditions.

Once suitable conditions for cleavage by enterokinase were established, these were implemented using the protein harvest from the IMAC stage to remove the HP-Trx tag from MβLp01. Around 90% of the fusion protein was successfully de-tagged, with a small amount of MβLp01 fusion remaining. This was not an issue since running a second IMAC serves to separate the cleaved recombinant protein from the His-tagged thioredoxin tag and any remaining His-tagged fusion protein. During the reverse IMAC step, the HP-Trx tag was not completely separated from cleaved MβLp01, since it co-eluted with a significant portion of MβLp01 (Figure 4.13D). This could be related to a decrease in metal binding affinity of the tag upon separation from the fusion protein. As such, the fractions in which the tag did not co- elute were collected and kept separate from the remaining MβLp01-containing fractions. These were designated 'Batch A' and 'Batch B', respectively. Both batches were separately subjected to size exclusion chromatography in order to remove contaminating higher molecular weight proteins and to enable separation of MβLp01 and HP-Trx in batch B. This resulted in successful isolation of MβLp01 in both batches to close to 100% purity (Figure 4.13E+F). These were then combined and concentrated yielding 140 ug pure MβLp01 at a concentration of 1.4 mg/mL (Figure 4.13H). SDS-PAGE gels representing the entire purification workflow are displayed in the panel of figures below.



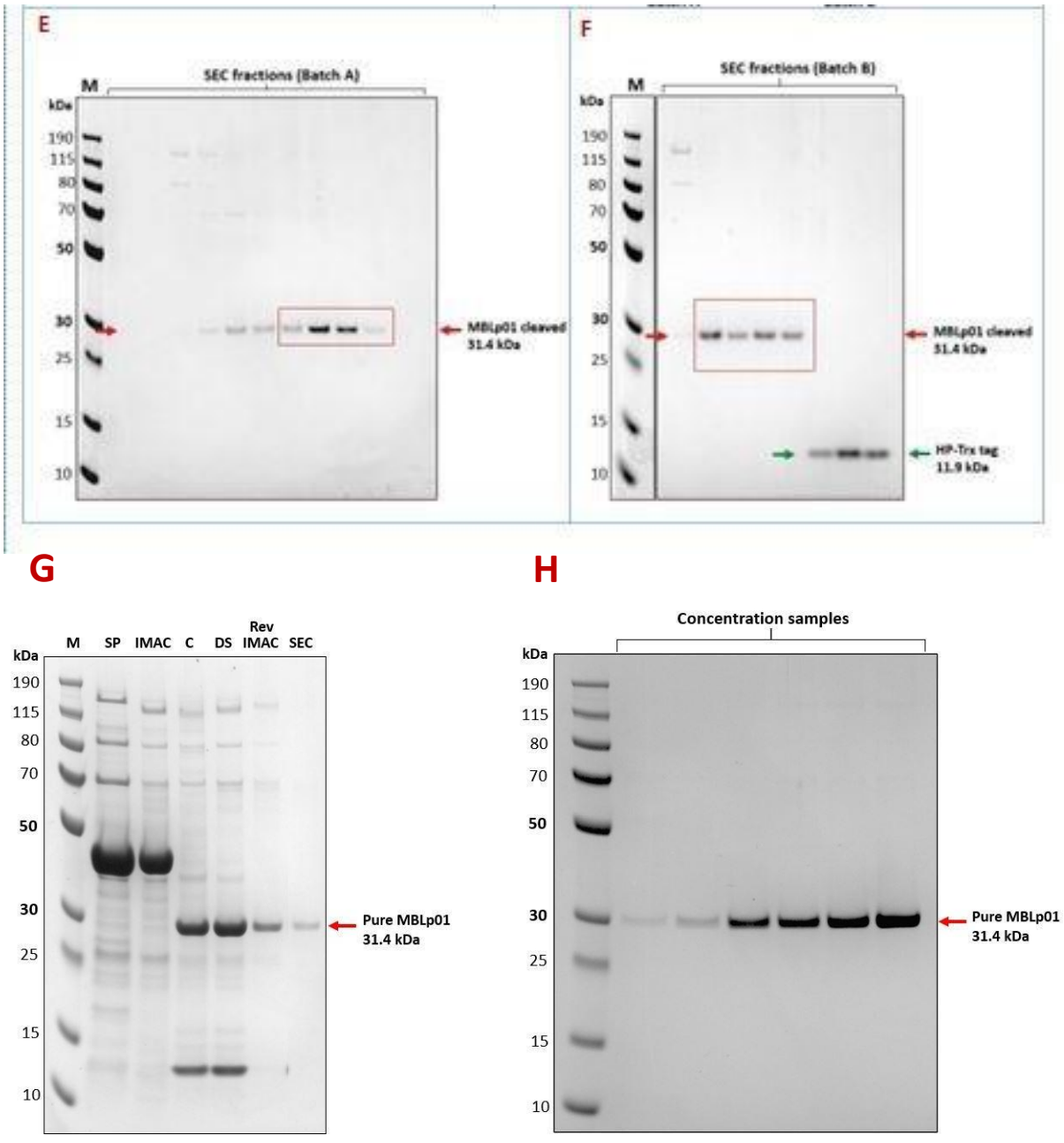


Figure 4.13 Panel of SDS-PAGE gels representing an overview of the multi-step purification procedure of de-tagged MβLp01. A) IMAC step B) HP-Trx tag cleavage step C) Desalting step D)Reverse IMAC step E+F) SEC steps G)Samples from each stage run on one representative gel H) Samples of purified MβLp01 over the course of the concentration procedure.

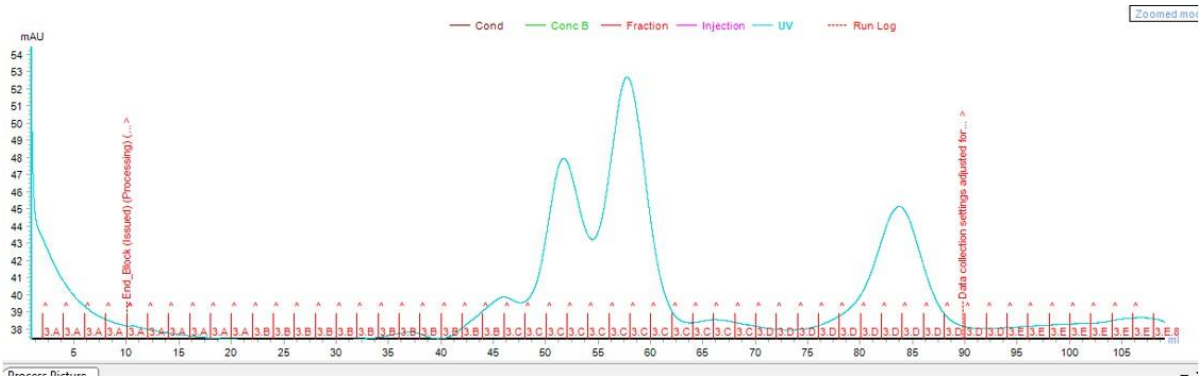


Figure 4.14. SEC chromatogram of MβLp01 corresponding to the SDS-PAGE gel in Figure 4.13F. SEC successfully separated MβLp01 from the HP-Trx tag and some higher molecular weight proteins, yielding pure MβLp01.

4.4.5 The X-ray Crystal Structure of MβLp01

Diffraction quality crystals of MβLp01 grew in three days in multiple conditions in a variety of sizes and morphologies including single crystals and large needle clusters. An example of some of the needle clusters is shown in Figure 4.15b. A total of 16 crystals were harvested and cryoprotected before X-ray diffraction data collection was carried out on beamline I24 of the synchrotron at the Diamond Light Source (UK).

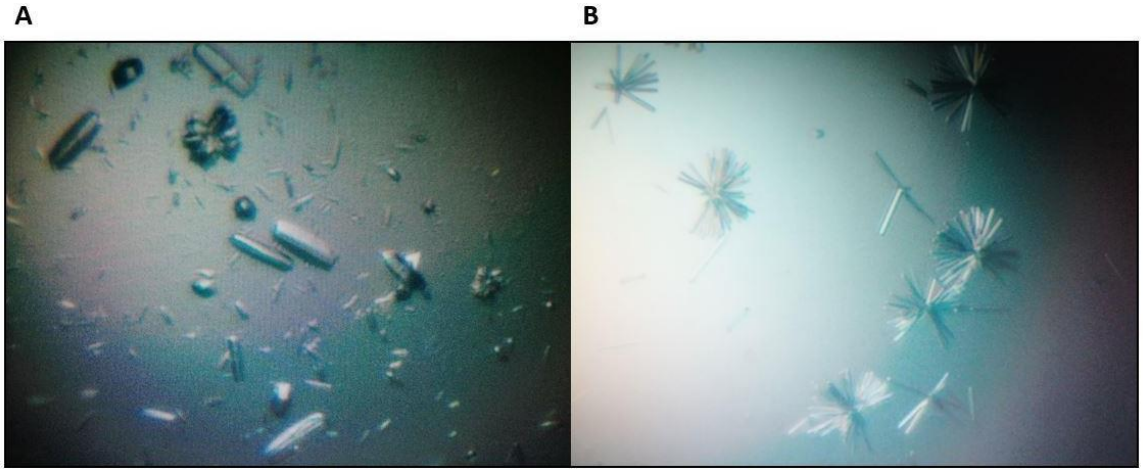


Figure 4.15 Images of some of the MβLp01 crystals formed, viewed under a microscope. a) A mixture of crystal morphologies formed in PEG/ion screen b) An example of the needle cluster morphology present in many conditions of the screens.

The X-ray diffraction images were automatically indexed, integrated, scaled and merged by the DLS servers immediately upon collection using several different pipelines – xia2, xia2 3dii,fast_dp, autoPROC and autoPROC+STARANISO. Two crystals yielded datasets with suitable diffraction statistics, (tabulated in the Appendices A4.4.2) with both crystals being classified as belonging to the orthorhombic space group I222. Of the two, the dataset in Table 4.7, processed by xia2 3dii, had the best diffraction statistics (100% completeness, $CC_{1/2}$ of 1.0, $I/\sigma(I)$ of 6.4) and was therefore chosen for structure solution by molecular replacement. This crystal had grown in 0.2M sodium acetate trihydrate and 20% PEG 3350, pH 8.0 (Figure 4.15a), with unit cell parameters of a= 53.26, b=87.61, c=118.42, α=90°, β=90°, γ=90°. The X- ray diffraction image generated by this crystal, along with the graphical representation of the resultant reflections, is shown below (Figure 4.16a and b, respectively),

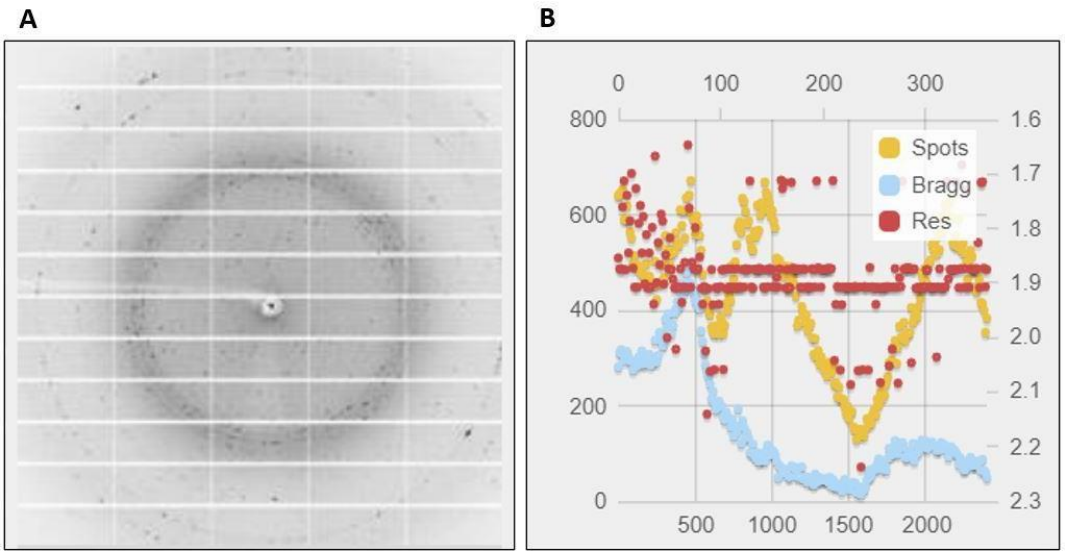


Figure 4.16 X-ray diffraction of MβLp01 a) The resultant X-ray diffraction image and b) The graphical output from the ISPyB interface, representing the individual reflections generated.

Molecular replacement was carried out using a homology model based on the crystal structure of the metallo β-lactamase PNGM-1 (PDB: 6J4N). The final structure of MβLp01 was solved and refined at 1.95Å, with R_{work} of 19.5% and R_{free} of 23.5%. These values indicate that the agreement between the model and the experimental data is within the accepted range, and that the data has not been 'overfitted' – as shown by a difference between R_{work} and R_{free} of < 7%. The full set of diffraction and structure refinement statistics are tabulated below (Table 4.4.1).

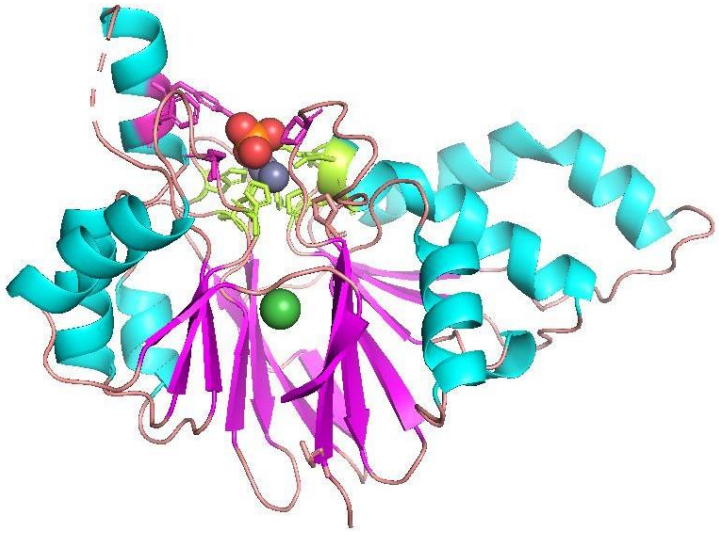


Figure 4.17 The X-ray crystal structure of MβLp01, solved by molecular replacement. Coloured by secondary structure. Binding pocket residues represented as pink sticks, zinc coordinating residues as lime sticks and loops shown in salmon pink. The bound phosphate is coloured red and the additional bound metal ion in green. Image generated using Pymol.

Data collection statistics		Structure refinement statistics	
Wavelength		**Reflections used in refinement**	20617 (2033)
Resolution range	**35.99 - 1.95 (2.02 - 1.95)**	**Reflections used for R-free**	970 (99)
Space group	**I 2 2 2**	**R-work**	**0.1955 (0.3029)**
Unit cell	53.26 87.61 118.42 90 90 902	**R-free**	**0.2348 (0.3421)**
Total reflections	233206 (12886)	**CC(work)**	0.950 (0.726)
Unique reflections	20628 (2033)	**CC(free)**	0.925 (0.447)
Multiplicity	11.3 (6.3)	**# of non-hydrogen atoms**	2274
Completeness (%)	99.89 (100.00)	**Macromolecules**	2058
Mean I/sigma(I)	6.42 (1.65)	**Ligands**	8
Wilson B-factor	20.60	**Solvent**	208
R-merge	0.4844 (1.954)	**Protein residues**	268
R-meas	0.5066 (2.131)	**RMS(bonds)**	0.014Å
R-pim	0.1462 (0.8324)	**RMS(angles)**	2.35°
CC1/2	**0.984 (0.4)**	**Ramachandran favoured (%)**	95.45
CC*	0.996 (0.756)	**Ramachandran allowed (%)**	4.17
		Ramachandran outliers (%)	0.38
		Rotamer outliers (%)	0.00

Table 4.7 The set of X-ray diffraction data and structure refinement statistics for the structure of MβLp01 solved by molecular replacement.

Matthews coefficient (V_M) calculations suggested the presence of 1 molecule in the asymmetric unit, corresponding to a solvent content of ~44.3%. 8 copies of the ASU in the unit cell (see Appendices A4.4.4).

The structure of MβLp01, as solved by molecular replacement, is presented in Figure 4.17. There is a phosphate ion bound in the substrate binding pocket and a pair of metal ions located in the active site. There is also an additional metal ion bound elsewhere on the protein surface. Two different methods were performed to verify the identity of the di-metal centre. ICP-MS was run on both the tag-free MβLp01 and the His-patch thioredoxin fusion, however the signal to noise ratio was very poor for the tag-free sample (a result of protein quality degradation throughout two years' storage at -80°C) so this data was not considered further. Data for the HP-Trx fusion (see Appendices A4.4.5) indicated the presence of zinc, with an amount of nickel which was negligible. The ratio of Zn:S calculated suggests the presence of 5.4 zinc atoms in the fusion protein. ICP-MS was also run on the purified HP-Thioredoxin protein in case some of the zinc binding was attributed to this. X-ray diffraction data was also collected at the zinc K absorption edge and the calculated anomalous difference fourier map revealed the two highest peaks overlaying with the 2 active site metal ions (Figure 4.18), and the 3rd peak lies at the position of the additional surface metal ion. This is necessary but not conclusive evidence that the metal identity is zinc as opposed to nickel, as nickel sites may also give rise to large peaks at the wavelength of the zinc k-edge. A dataset taken at the nickel K-edge would be more conclusive, since the nickel k-edge is lower in energy than the zinc k-edge, resulting in absence of peaks in the anomalous difference fourier map at this wavelength. Due to the size of the crystal and radiation damage, a nickel K-edge was unable to be obtained. The metal binding site validation server 'Check my metal' (https://cmm.minorlab.org/)[315] indicates that the supposition of a di-zinc centre is acceptable. Nevertheless, due to concerns around the nickel from the Ni^{2+}-NTA column replacing one or both of the zinc ions in MβLp01, the use of a Zn^{2+}-NTA column was implemented for subsequent purifications of this metalloprotein.

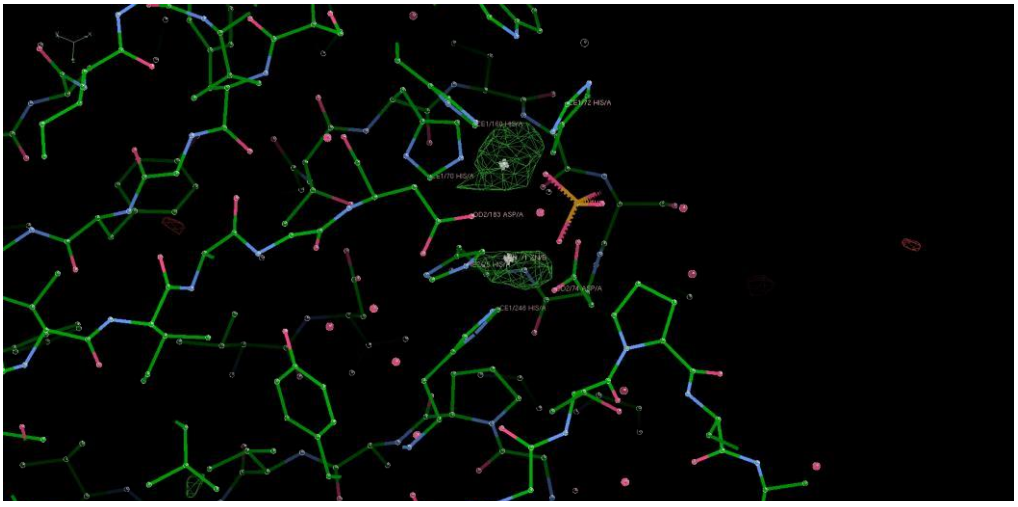


Figure 4.18 Difference Fourier map indicating the presence of two metal ions (green mesh)

Characteristic of the MβL superfamily, the core scaffold of MβLp01 consists of a distinctive αβ/βα sandwich fold in which the active site is located at the interface of the two domains. In MβLp01, this fold consists of 14 β-strands and 6 α-helices. The two central β-sheets of the sandwich are each comprised of 7 β-strands: β1-β6 with β14 and β7-β13. There are 3 α-helices on either side of each β-sheet; α1-3 flanking the strands β1-β6/β14 and α4-6 flanking β7-β13. A highly simplified, diagrammatical representation of this is illustrated in Figure 4.19. The zinc-coordinating residues originate from the 4 loops (L1-4) which protrude from the 2 central β-sheets. There are 3 external loops which shield the zinc sites from the outer surface and form the substrate-binding pocket. It is these external loops which are integral to the substrate specificity of MβLs. In Figure 4.20, the model of MβLp01 is orientated to demonstrate this.

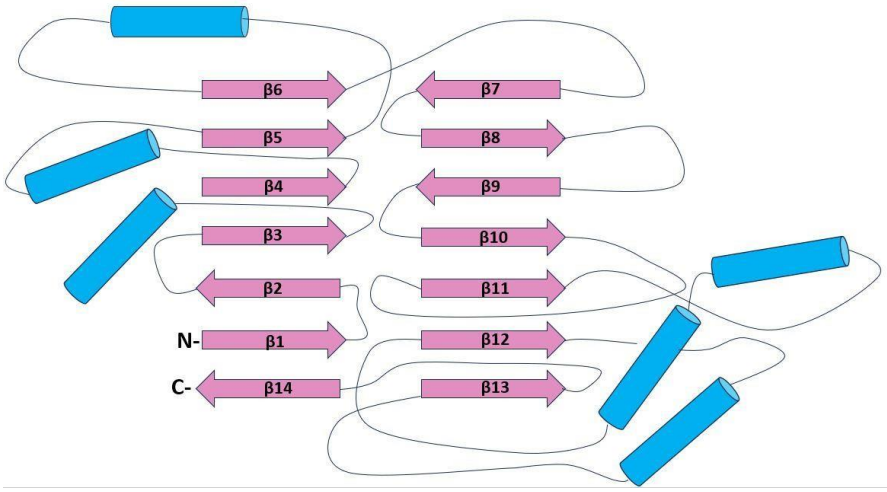


Figure 4.19 Simplified diagrammatic representation of the overall topology of MβLp01, depicting the 14 β-strands (pink arrows) constituting the 2 central β-sheets and the 3 α-helices (blue cylinders), connected by loops.

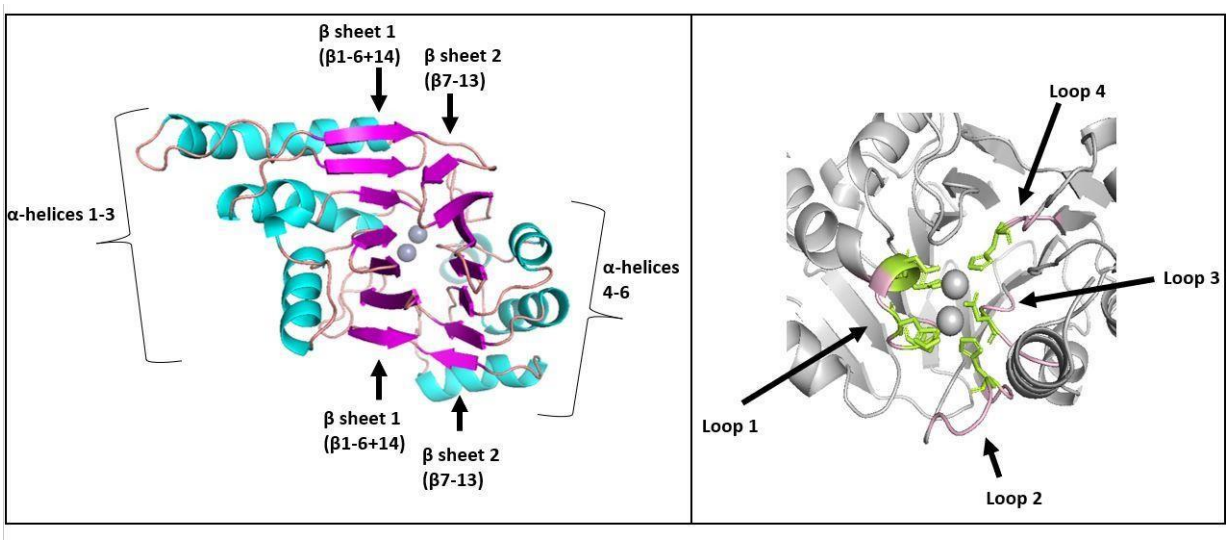


Figure 4.20 Different orientations of the MβLp01 structure to highlight a) the overall αβ/βα topology (as in Figure 4.19) and b) the zinc binding site, with particular focus on the four loops (pale pink) which anchor the zinc coordinating residues (lime green). Images generated using Pymol.

The binding pocket as calculated by the CASTp 3.0 server (Computed Atlas of Surface Topography of Proteins)[316] has a surface area of 40.75Å^2 and volume of 10.31Å^3 (Figure 4.21). The residues involved in the substrate binding pocket are predicted to be Pro15, His72, Asp74, His160, Asp183, Tyr221, Phe222, His248 and Met250 (coloured magenta).

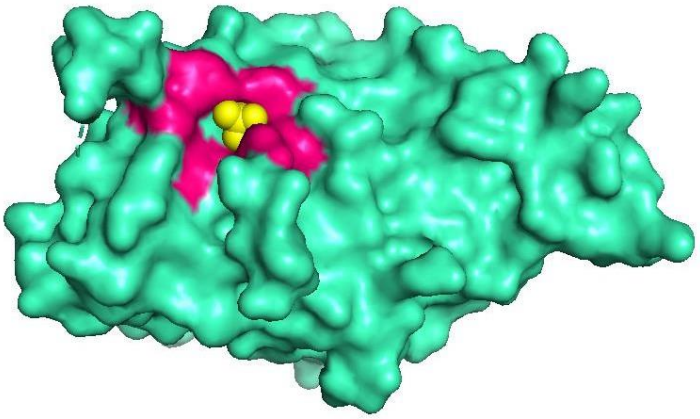


Figure 4.21 Surface representation of the experimentally determined MβLp01, with the predicted binding pocket determined by the CASTp server. Predicted binding pocket residues – Pro15, His72, Asp74, His160, Asp183, Tyr221, Phe222, His248 and Met250 - shown in magenta and the bound phosphate in yellow. Image generated using Pymol.

To predict the metal-binding residues in MβLp01, a metal ion-binding site prediction and modelling server, MIB2, was used (http://combio.life.nctu.edu.tw/MIB2/)[317], which utilises both the (PS)[2] method and the AlphaFold protein structure database to obtain 3D information to execute metal ion docking and predict binding residues. This predicted Asp74, His75, Asp183 and His248 as the most likely metal binding residues. This appears to be the Zn2 site only.

A comparison of protein structures through the Dali server[318] indicated that the closest homologue of MβLp01 was an arylsulfatase from *Pseudoalteromonas atlantica* T6c (PaAst; PDB: 8GYG)[319], a member of the MβL superfamily, sharing a Z-score of 33.6 with a sequence identity of 29% and an RMSD of 2.1Å. Up until this structure was recently published (late 2023), the closest structural homologue to MβLp01 - and the one which was used as the starting model for molecular replacement - was PNGM-1 from a deep-sea sediment metagenome[320,321]. This shares with MβLp01 a Z-score of 31.9, a sequence identity of 28% and an RMSD of 1.9 (according to the DALI server; Pymol reports RMSD of 1.3). Figure 4.22 represents the overall structures of MβLp01 and PNGM-1 superimposed. The zinc binding residues (lime green sticks) appear to support the zinc ions from underneath, whilst the substrate binding residues (magenta) sit above, with the phosphate ion located at the centre of these residues, directly above the zinc ions.

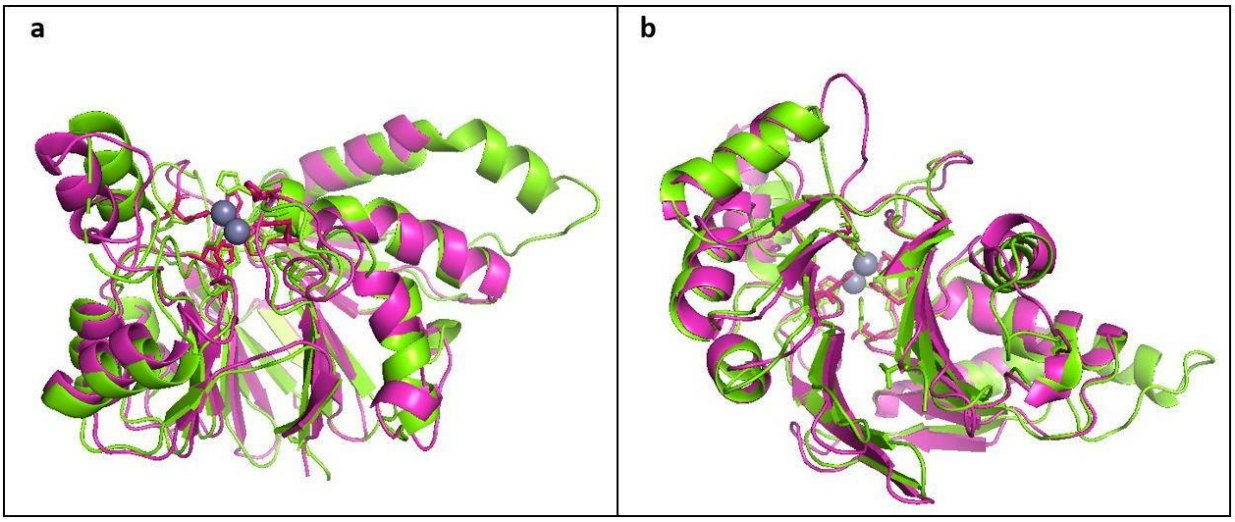


Figure 4.22 The experimentally determined structure of MβLp01 (magenta) overlaid with that of a close structural homologue, PNGM-1 (green). a) and b) illustrate two alternate orientations, rotated along the z axis. The zinc coordinating residues are displayed as sticks. Images generated in Pymol.

Superimposition of MβLp01 and PNGM-1 illustrates that the position of the two zinc ions and of the zinc coordinating residues in both structures are well conserved (Figure 4.23). The RMSD value for the 7 zinc-coordinating residues is 0.2. When the RMSD values for the two zinc binding sites, Zn1 and Zn2, were calculated separately, the Zn1 site has a lower value, of 0.06, compared to 0.18 for Zn2. However, the Zn1 residue set has 5 atoms less than the Zn2 set, which is partly responsible for the lower value. This higher value for the Zn2 site is in accordance with the literature on the B3 MβL subclass.

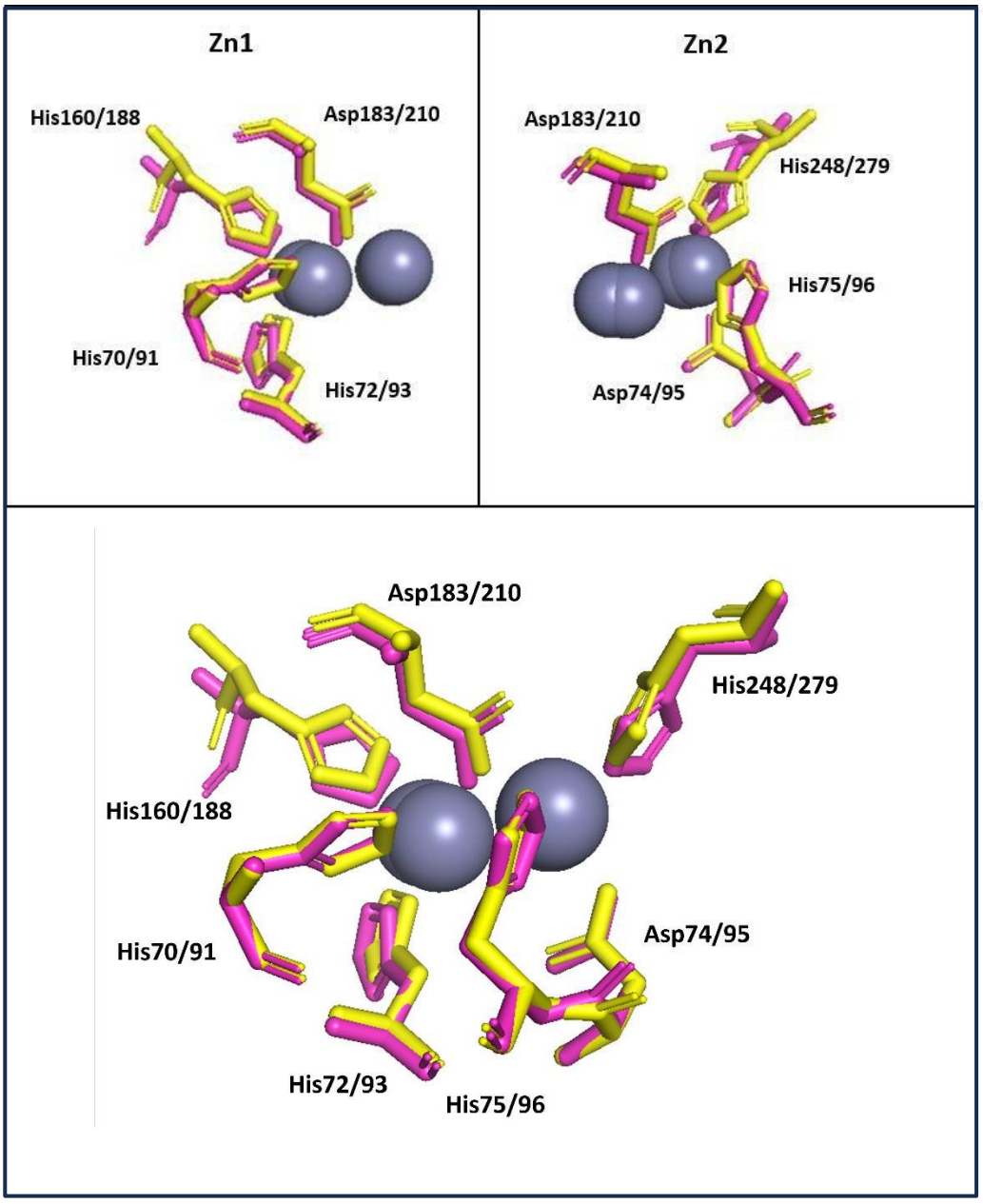


Figure 4.23 Superimposition of the zinc coordinating residues of MβLp01(yellow) and PNGM-1 (magenta). The Zn1 and Zn2 sites are displayed both separately (upper) and combined (lower). Asp183/210 bridges the two zinc ions.

4.4.6 Investigation of the biological function of MβLp01

4.4.6.1 Phytic acid-degrading ability of MβLp01

When purified MβLp01 was first assayed against phytic acid, the lack of activity demonstrated by the molybdenum blue assay (Chapter 2.6) was somewhat unexpected – based on the reported results in the paper *'Functional Metagenomics Reveals a New Catalytic Domain, the Metallo-β-Lactamase Superfamily Domain, Associated with Phytase Activity'* published in mSphere in 2019 by the Daniel group[267]. The reader is encouraged to read this publication prior to reading the following section, to appreciate the results herein reported. This outcome led to countless purifications of MβLp01 with varying buffer compositions and a myriad of assays incorporating a range of parameters such as incubation temperature and time, enzyme and substrate concentrations and inclusion of $ZnSO_4$ as an additive. Nevertheless, activity against phytic acid was never observed. Only HP-thioredoxin tagged-MβLp01 was used in assays (whereas tag-cleaved MβLp01 was used for crystallisation). Included in the following chapter are some representative results of various assays performed.

Figure 4.24 is a standard example of the results obtained from assays against phytic acid using the molybdenum blue method for phosphate determination; in addition to the purified MβLp01, fractions from all stages of purification were also assayed alongside, as were controls of non-recombinant *E. coli* TOP10 cells and the purified HP-thioredoxin peptide (as the 'empty vector' control). The A_{700nm} values shown are the values following correction from the substrate background contribution. Incubation temperature was the reported temperature for optimal activity of MβLp01, 50°C. This particular assay was performed with MβLp01 at a final assay concentration of ~0.34 μM and ~3.4 μM (MβLp01 was purified to a final concentration of ~6.8 μM in this particular purification). It is expected that the samples from earlier stages of the purification i.e. the lysate (CFE; cell-free extract), IMAC flow-through, and to some extent, the post-IMAC sample, exhibit higher A_{700nm} values as a result of the endogenous *E. coli* host phytases present. Nevertheless, these results clearly illustrate the decrease in phosphate release as purity of MβLp01 increases, to the extent that the pure MβLp01 has an average A_{700nm} of around 0, even when including the range of error. A_{700nm} values were 0.011, -0.0006 and 0.002 for the two different batches of SEC-purified MβLp01 and the 10-fold higher concentration sample, respectively. Desalting results in dilution of a sample so this is likely the reason for the lower absorbance of the sample which was desalted between IMAC and SEC. For the other purification-stage samples, it was total protein which was being measured so concentrations could not be standardised. The HP-thioredoxin peptide tag was purified in order to be used as an assay control, such as to rule out any contribution from this tag since assays were performed with the HP-thioredoxin-MβLp01 fusion. Contribution from this

peptide was negligible, however it was still significantly higher relative to pure MβLp01. Endogenous phytase activity from *E. coli* TOP10 produced an A_{700nm} value of 0.24, which was the same value as that produced by the MINPP used as a positive control. This data indicates that purified MβLp01 was not active against phytic acid under these assay conditions.

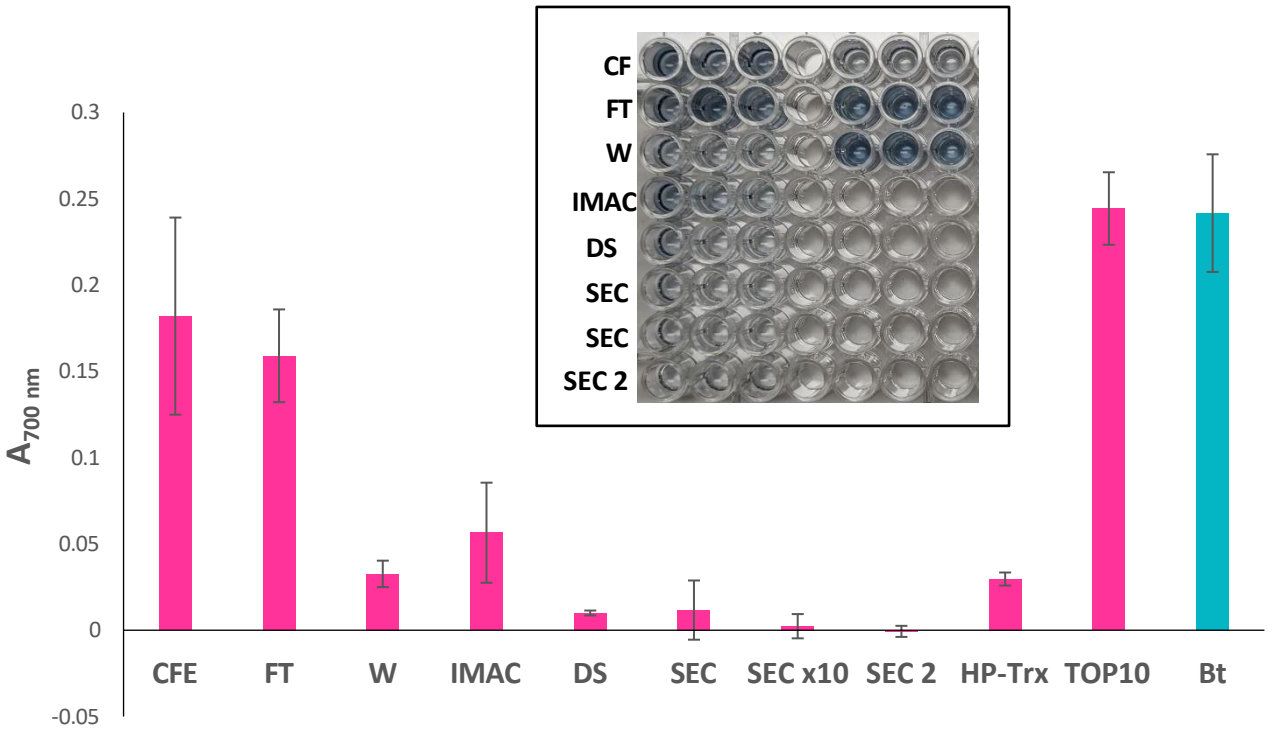


Figure 4.24 Relative amount of phosphate released from samples at each stage of MβLp01 purification incubated with phytic acid at 50°C for 30 min. Absorbance at 700nm of the resulting blue phosphomolybdate complex was measured after addition of developing solution. Corrected for substrate background contribution. Sample stage/type is shown on the x-axis. Abbreviated x-axis labels are as follows: CFE – cell free extract; FT – IMAC flow-through; W – IMAC wash; IMAC – IMAC elution pool; DS – desalted sample; SEC – SEC-purified MβLp01; SEC 2 – a different batch of SEC-purified MβLp01; SEC x10 – the same sample as in 'SEC' but with 10-fold increased sample volume; HP-Trx – purified HP-thioredoxin peptide; TOP10 – *E. coli* TOP10 host cells; Bt – *Bt*MINPP positive control. Inset: an image of the reactions in the 96 well plate, following colour development.

Similarly, with an incubation temperature of 40°C (all other parameters held constant, and the MβLp01 samples were from the same purification as used in the 50°C assay), there was no activity against phytic acid (Figure 4.25). 40°C was the assay temperature employed in the aforementioned paper[267] which was the reason for this temperature choice. The results are unambiguous – the purified MβLp01 has zero activity against phytic acid under these conditions. The IMAC purified sample had a fraction of activity (A_{700nm} = 0.04) but when compared to the positive control (A_{700nm} = 0.97), this is negligible. The positive control has four-fold higher activity at 40°C compared to 50°C which is in line with previous work on this

phytase. *Nota bene*: At the time of performing this particular assay, the earlier purification samples had been consumed in other assays, therefore only the IMAC fraction and pure samples were included.

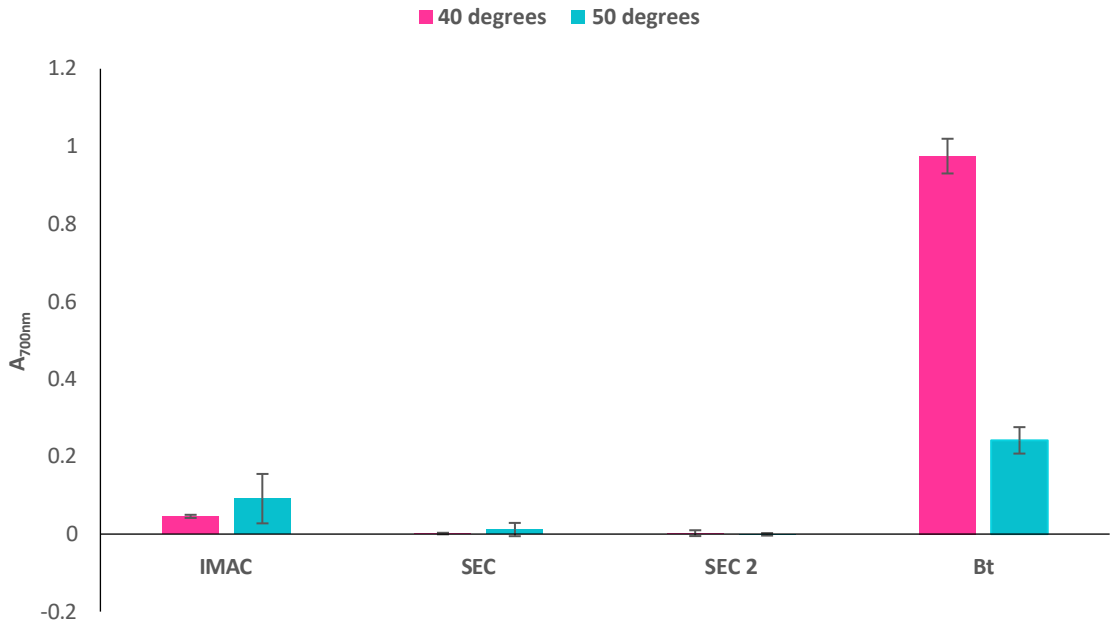


Figure 4.25 Comparison of phosphate released from MβLp01 samples (partially and fully purified) incubated with phytic acid at 40°C or 50°C for 30 min. Absorbance at 700nm of the resulting blue phosphomolybdate complex was measured after addition of developing solution. Corrected for substrate background contribution. Enzyme samples are shown on the x-axis. Abbreviated x-axis labels are as follows: IMAC – IMAC elution pool; SEC – SEC-purified MβLp01; SEC 2 – a different batch of SEC-purified MβLp01; Bt – *Bt*MINPP positive control.

On one occasion, there was a non-zero response (after substrate background correction) with three SEC fractions of MβLp01, 'E8', 'E9' and 'E10' incubated with both phytic acid and the purer myo-IP6 dodecasodium salt source. However, HPLC run subsequently on each of these reactions (repeated the same but omitting the colour development step) established that there was in fact no hydrolysis of IP6 (Appendices A4.4.6). Subsequently, one of these fractions, E10, was incubated with IP6 for 4 hours at 37°C -to ensure ample incubation time - before subjecting the reaction mixture to HPLC (Figure 4.26). This data is further evidence that IP6 is not a substrate of MβLp01, since all of the starting IP6 peak (retention time ~35 min) remained whereas in the positive control (the MINPP P15) the IP6 had been completely hydrolysed to lower inositol polyphosphate intermediates. These samples were then incubated *without* IP6 and, interestingly, still produced the same colour development, despite the absence of a phosphate-containing substrate. Therefore, to determine if this colorimetric response was of

enzymatic origin, the E8, E9, and E10 samples were incubated without substrate over 4 hours at 25°C (Figure 4.27). The E8, E9 and E10 samples produced A_{700nm} values of 0.23-0.25, 0.41-0.43 and 0.32-0.34, respectively. The values over the 4 hours were invariant, neither increasing nor decreasing, which suggests that the observed colour development was not enzymatic in origin. It is of course prudent to note the relatively large error for the 4h time point of E9.

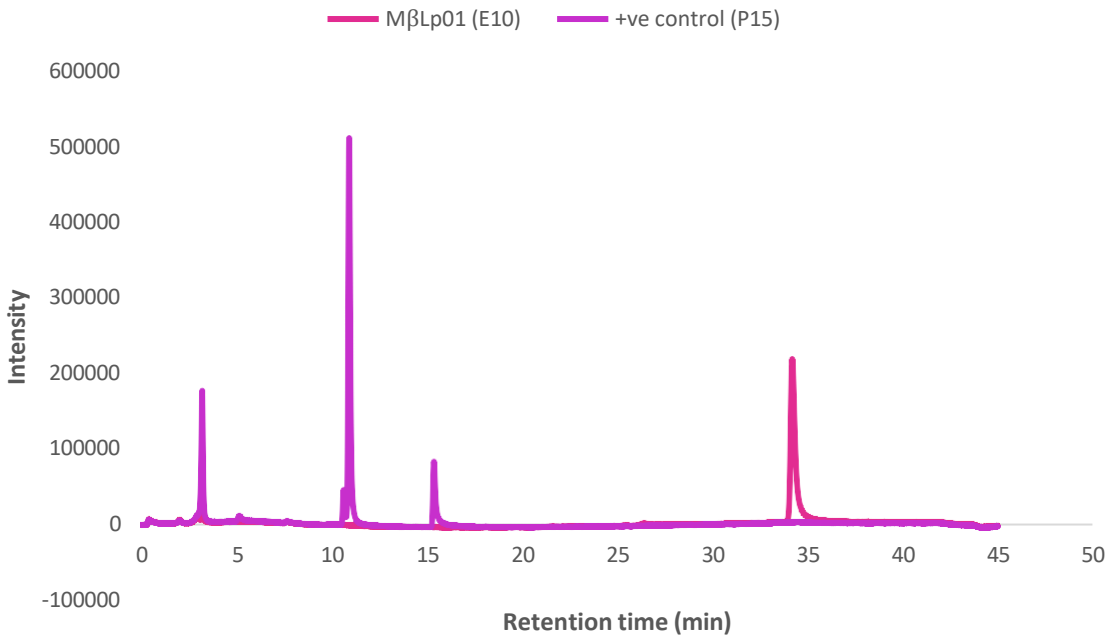


Figure 4.26 HPLC trace of the IP6 hydrolysis profile generated by MβLp01 in comparison to a positive control (P15). HPLC traces overlayed of MβLp01 (magenta) and P15 (purple). MβLp01 and P15 were each incubated with IP6 for 4h at 37°C before the reaction product mixture were subjected to HPLC. Traces of IP6 only and the IP6 hydrosylate standard are omitted for clarity (data in Appendices A4.4.7).

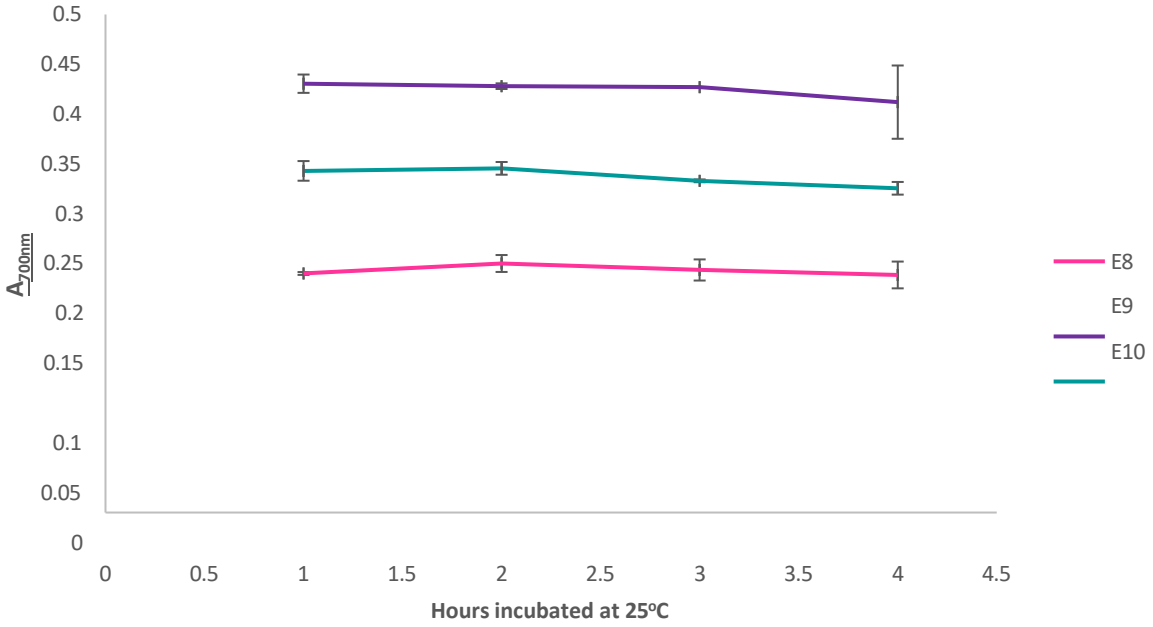


Figure 4.27 Three MβLp01 SEC samples (E8, E9, E10) incubated without substrate over 4 hours at 25ºC before the molybdenum blue colour development.

4.4.6.2 Phosphatase activity of MβLp01

Since assays with phytic acid revealed absence of phytase activity of MβLp01, a number of substrates containing a phosphomonoester bond were tested to ascertain if MβLp01 exhibits phosphohydrolase activity. Substrates included were adenosine triphosphate (ATP), adenosine diphosphate (ADP), glucose 1-phosphate (G1P) and glucose-6-phosphate (G6P). IP6 and also IP4 were also run alongside for comparison. Incubated at 50°C for 30 min, each substrate was incubated with and without MβLp01 (at a final concentration of 0.34μM), in order to quantify the contribution from the substrate itself. Since background-corrected A_{700nm} values were mostly negative, the substrate only data is presented alongside the substrate + MβLp01 reactions (Figure 4.28). it is evident that in the majority of substrates tested, there is no difference between the 'substrate only' samples and the substrates incubated with MβLp01. The only exception is ATP, in which there is a difference between the 'substrate only' (A_{700nm} = 0.55) and 'substrate + MβLp01' (A_{700nm} = 0.73) samples. Errors for all data points are relatively small which indicates the consistency of this data.

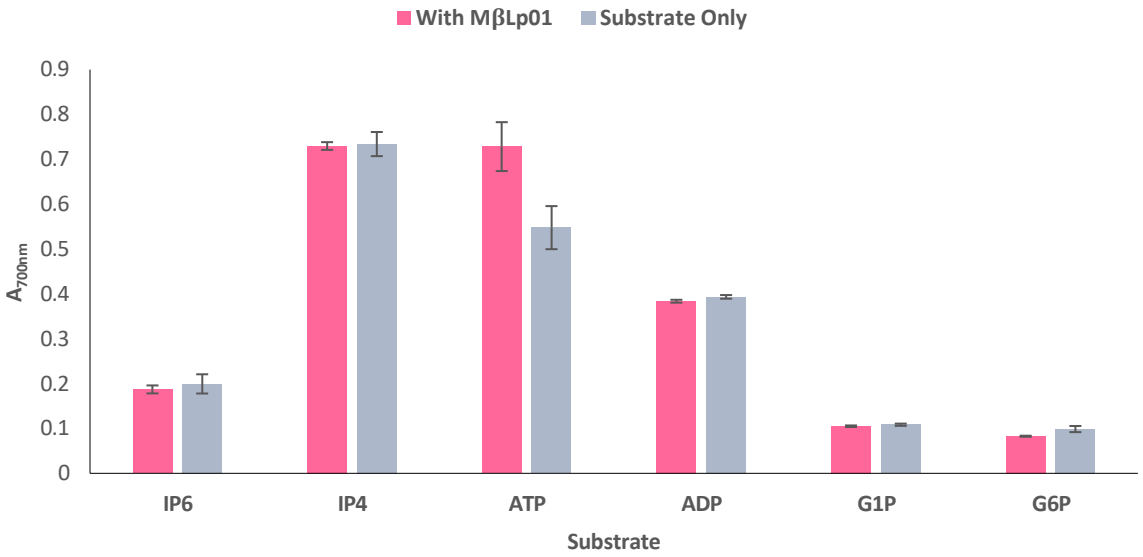


Figure 4.28. Relative amount of phosphate released from 6 different monophosphate bond-containing substrates incubated either with or without MβLp01 at 50°C for 30 min. Abbreviated x-axis labels are as follows: G1P – glucose 1-phopshate; G6P – glucose 6-phosphate.

To further illustrate the lack of activity against IP6, IP4, ADP, G-1-P and G-6-P, the assay was repeated this time with a positive control for phytase activity, *Bt*MINPP, run alongside (Figure 4.29). The A_{700nm} values of 'substrate only' and for 'substrate + MβLp01' were almost exactly the same as in the previous assay – further evidence of the reliability of these results – whilst *Bt*MINPP exhibited activity against all substrates tested with the exception of glucose-1-phosphate. There was the possibility (albeit unlikely) for example, that all the ATP had hydrolysed to ADP upon storage, leaving low amounts of available substrate for MβLp01. However, *Bt*MINPP was able to produce a 2.8-fold higher A_{700nm} value, indicating that this was not the case. The A_{700nm} values attributed to *Bt*MINPP for IP6, IP4, ATP, ADP, G1P, and G6P were 6.7, 1.77, 2.8, 2.1, 1.3 and 5.2 times higher than for MβLp01/substrate only, respectively. It is important to bear in mind that since *Bt*MINPP is so much more active than MβLp01, the absorbance at 700nm is a little beyond the linear range (~X – X) according to the phosphate calibration. However, for the purposes of this assay, i.e. demonstrating the distinct differences in amount of phosphate released (as determined by the molybdenum blue assay) between an established MINPP and MβLp01, this was insignificant. It is also prudent to note that this assay was performed at 50°C which is not the optimum temperature for *Bt*MINPP activity, yet despite this, *Bt*MINPP still yielded higher A_{700nm} values than MβLp01. Unsurprisingly, the A_{700nm} values of the 'substrate only' samples were all slightly higher in the second assay including *Bt*MINPP, since the reconstituted substrates in solution are prone to chemical hydrolysis over time.

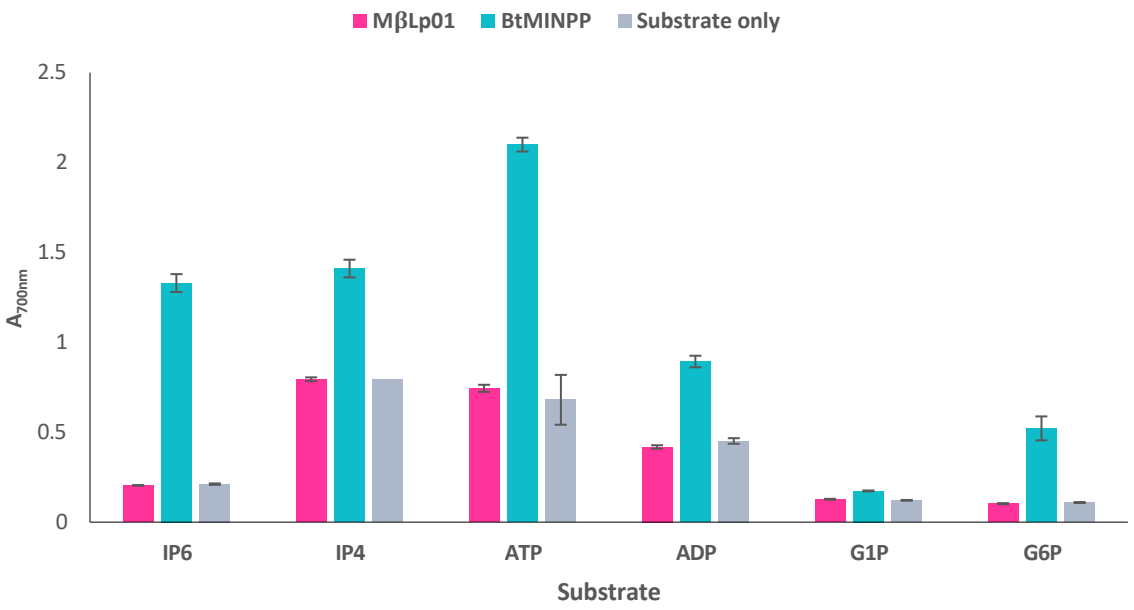


Figure 4.29 Relative amount of phosphate released from 6 different monophosphate bond-containing substrates incubated with either MβLp01, *Bt*MINPP (positive control) or without any enzyme, at 50°C for 30 min.

To evaluate the statistical significance of these data, the p values were calculated from a one-tailed or two-tailed t-test for two independent means (Table 4.8). For 'substrate only' vs 'MβLp01 + substrate', the data from the assay in Figure 4.28 were used. It would be expected that the p values for 'MβLp01 + substrate' vs 'substrate only' were insignificant for all except ATP. Unexpectedly, the p value was also significant for ADP and G6P too. However, the A_{700nm} values were *higher for the substrate only* relative to with MβLp01, and apparently significantly so. Regardless, with the exception of ATP, there was no significant *increase* in A_{700nm} values when MβLp01 was included with the substrates. As expected, for MβLp01 vs *Bt*MINPP, the difference in A_{700nm} values were highly significant.

	MβLp01 vs *Bt*MINPP		MβLp01 vs substrate only	
Substrate	**p value**	**Significant? <0.05)**	**p value**	**Significant? (<0.05)**
IP6	< 0.00001	Yes	0.205	No
IP4	0.000474	Yes	0.399	No
ATP	< 0.00001	Yes	0.006	Yes
ADP	0.000016	Yes	0.016	Yes
G1P	0.000016	Yes	0.066	No
G6P	0.000407	Yes	0.0080	Yes

Table 4.8 Significance of the data from phosphatase assays of MβLp01 relative to either substrate alone or the positive control. p values generated with the use of an online t-test calculator for two independent means; one tailed for MβLp01 vs substrate only; two tailed for MβLp01 vs *Bt*MINPP.

The fact that the difference between MβLp01 and the positive control across all tested substrates is statistically significant is strong evidence against the null hypothesis, where the null hypothesis is that *there is no difference in A_{700nm} values between MβLp01 and the positive control.* Quantity of phosphate liberated from the tested substrates with MβLp01 was therefore statistically significantly lower relative to the positive control, as illustrated in Figure 4.28. Conversely, the null hypothesis of *there is no difference in A_{700nm} values between 'substrate + MβLp01' and the 'substrate only'* can be accepted for IP6, IP4 and glucose-1-phosphate whilst rejected for ATP, ADP and glucose-6-phosphate. However, for ADP and glucose-1-phosphate, although there is a difference with regard to the inclusion of MβLp01, this difference is not representing the positive net colour response of MβLp01 relative to the substrate alone; rather, the fact that the substrate alone generates a colour response which is statistically significantly greater than with the inclusion of MβLp01. Since both treatments utilised the same source of substrate, this supposedly significant difference is likely to be attributed to random/human error in the performance of the assay. Undeniably, this assay suffers from multiple sources of potential error, as discussed in Chapter 2.6. Nevertheless, the data in question suggests that whilst IP6, IP4, ADP, glucose-1-phosphate and glucose-6-phosphate are not substrates of MβLp01, it is plausible that ATP is a potential substrate.

To further investigate the potential phosphatase activity of MβLp01, and to circumnavigate the pitfalls of the molybdenum blue assay, the classic artificial non-specific phosphatase substrate, 4-para-nitrophenyl phosphate (pNPP) was employed in a colorimetric hydrolysis assay with MβLp01, which is useful for rapid assessment of total phosphatase activity in a sample. An advantage of this assay over assays based on detection of free phosphate is the lack of interference of phosphate originating from sources other than the enzyme catalysed reaction, since the pNPP assay is monitored at a different wavelength. The reaction product of pNPP hydrolysis is the yellow chromogenic compound, para-nitrophenolate (pNP), the maximal absorbance of which, under alkaline conditions, is at 405nm. As a discontinuous determination of activity, four different concentrations of MβLp01 were each incubated with pNPP in a pH 5.4 buffer at 37°C for 30 min before the reaction mixtures were alkalised with NaOH. Reactions of *Bt*MINPP, as the positive control, were run in parallel at the same four concentrations. The formed p-nitrophenolate ion (ε=18300 $M^{-1}cm^{-1}$ -at pH X.X) was measured spectrophotometrically at 405nm. The raw A_{405nm} values were compared as displayed in Figure 4.30. This assay is recounted to have a wide linear range since the substrate concentration is not limiting, with saturation represented by an A_{405nm} value of around 4. Whilst increasing *Bt*MINPP concentration does not result in significantly increased pNP concentrations, MβLp01 exhibits a clear positive correlation between enzyme and substrate concentration – between

MβLp01 concentration increments, pNP concentration was increased either 1.5- or 2-fold. In contrast, *Bt*MINPP appears to produce similar pNP concentrations regardless of increased enzyme concentration, as if it is saturated. For example, the difference in A_{405nm} between 1uM and 10uM *Bt*MINPP was just 0.5 whereas for MβLp01 it was 1.65. It appears that *Bt*MINPP saturates the colour response . The difference between MβLp01 and *Bt*MINPP is most distinct at the lowest enzyme concentration tested, with a difference in A_{405nm} of ~2.4, whereas at the highest enzyme concentration tested the difference was only ~1.2. Based on the Beer Lambert law, and the extinction coefficient for pNP of 18,300 $M^{-1}cm^{-1}$ (pH X.X), MβLp01 produced 25.7, 53.0, 77.6 and 115.8 µM pNP at concentrations of 1, 2, 5 and 10 µM, respectively. At the same enzyme concentrations, *Bt*MINPP yielded pNP at concentrations of 154.1, 151.9, 161.2 and 179.8µM, respectively.

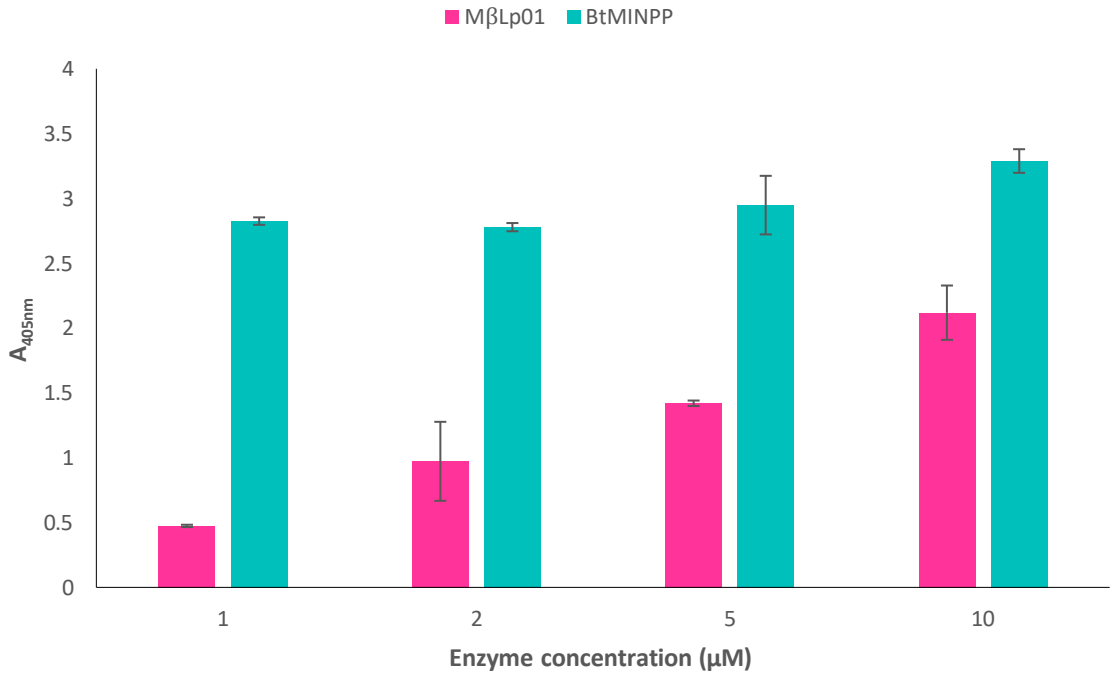
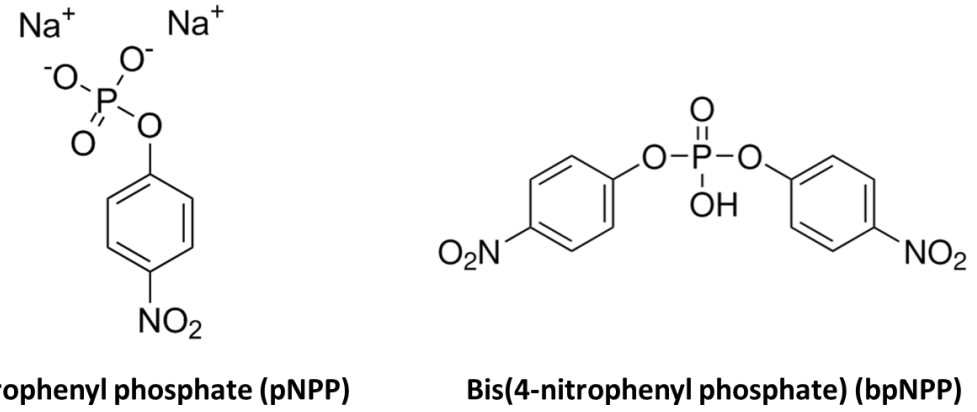


Figure 4.30 Hydrolysis of pNPP by MβLp01. pNPP (5mM) was incubated with four different concentrations of either MβLp01 or *Bt*MINPP (positive control) at pH 5.4, 37°C for 30 min. Relative amount of pNP released from pNPP is represented by the background corrected spectrophotometric intensity at 405nm.

4.4.6.3 Phosphodiesterase activity of MβLp01

Based on the literature (see chapter 4.2.4), it was reasonable to screen MβLp01 for phosphodiesterase activity, therefore an assay was performed with the phosphodiester substrate bis(4-nitrophenyl) phosphate (bpNPP; Figure 4.31). Hydrolysis of bpNPP yields the same product as in the hydrolysis of pNPP, the 4-nitrophenolate anion, which is monitored spectrophotometrically at 405nm. There is no free phosphate liberated in this assay.



4-nitrophenyl phosphate (pNPP) **Bis(4-nitrophenyl phosphate) (bpNPP)**

Figure 4.31 Chemical structures of the phosphatase substrates pNPP and bis pNPP.

Three different concentrations of MβLp01 were incubated with bis-pNPP at pH 5.4 and 37°C for 30 min before rendering the reaction alkaline for spectrophotometric measurements. A phosphodiesterase positive control was not available for comparison, however the *E. coli* phytase AppA was tested against bis-pNPP alongside, in lieu of a negative control. It is very apparent that AppA does not possess phosphodiesterase activity, which is as expected. Meanwhile, the data indicates that MβLp01 was able to hydrolyse the phosphodiester substrate, with A_{405nm} values approximately doubling upon increasing the MβLp01 concentration 2- or 2.5-fold (Figure 4.32+4.33). Using the extinction coefficient of pNP (18,300 $M^{-1}cm^{-1}$), MβLp01 hydrolysed bis-pNPP to yield 9.6, 16.4 and 35.5 µM pNP at corresponding concentrations of 1, 2 and 5 µM.

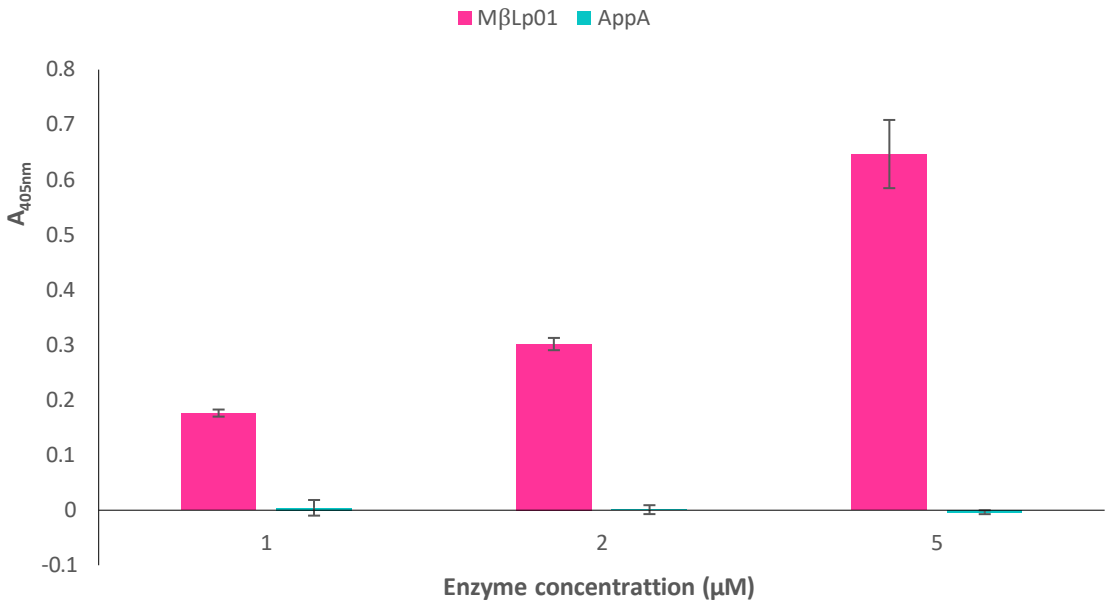


Figure 4.32 Hydrolysis of bis-pNPP by MβLp01. bis-pNPP (10 mM) was incubated with three different concentrations of either MβLp01 or AppA (positive control) at pH 5.4, 37°C for 30 min before addition of NaOH to terminate the reaction. Relative amount of pNP released from pNPP is represented by the background corrected A_{405nm} values.

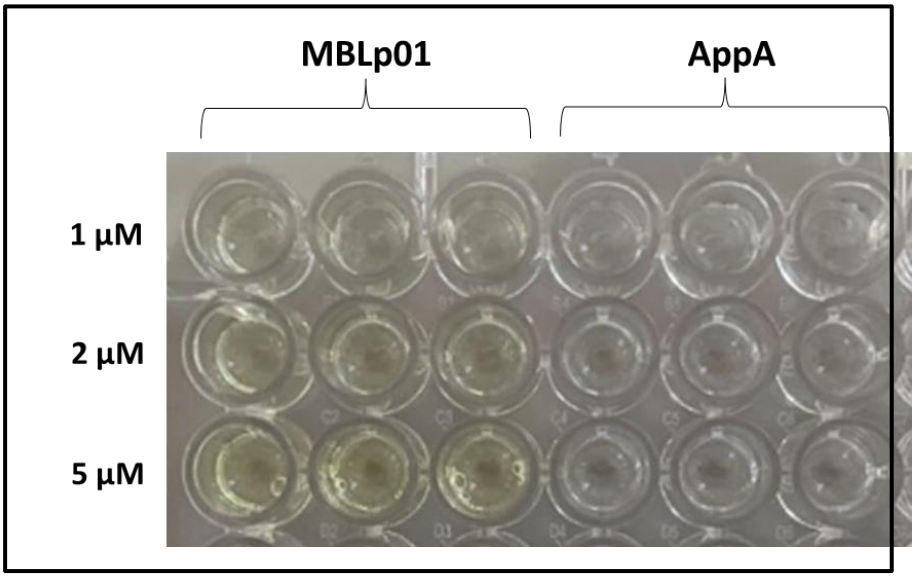


Figure 4.33 Qualitative/visual representation of the data in Figure 4.32. Colour development of reactions 10 min after the addition of NaOH.

4.4.6.4 β-lactamase activity of MβLp01

Multiple methods for assessing β-lactamase activity were trialled, including a colorimetric hydrolysis assay and several different methods based on minimum inhibitory concentration (MIC). Tests were performed using either purified MβLp01, whole *E. coli* TOP10 cells expressing MβLp01, or the lysate of *E. coli* TOP10 cells expressing MβLp01, depending on the type of method. For the MIC tests, a representative antibiotic from each of the three β-lactam classes was used: carbenicillin (penicillins), ceftazidime (cephalosporins) and doripenem (carbapenems).

4.4.6.4.1 Nitrocefin hydrolysis assay

Nitrocefin is an inherently chromogenic cephalosporin which undergoes a distinctive colour change from yellow (λ_{max} = 390 nm at pH 7.0) to red (λ_{max} = 486 nm at pH 7.0) upon hydrolysis, which is an indicator of β-lactamase activity. Preparations of MβLp01 – 1) lysed *E. coli* TOP10 cells expressing MβLp01 and 2) purified MβLp01 – were each incubated with nitrocefin at 25°C and pH 7.0* for 60 min. Also run in parallel were uninduced *E. coli* TOP10 cells harbouring the MβLp01 gene; *E. coli* TOP10 cells harbouring the 'empty' pBAD202/D-TOPO vector; purified HP-thioredoxin; and a positive control – a β-lactamase included in the commercial kit. Absorbance was monitored at 490nm over the course of the reaction and a standard curve of hydrolysed nitrocefin was used both to correct for the background absorbance and to quantify the nitrocefin hydrolysed based on the A_{490nm} values. For the MβLp01 lysate, four different sample volumes were run in duplicate and then the average was calculated for all eight readings for each time point. The colour development of purified MβLp01 along with the intermediary purification fractions and controls are displayed in Figure 4.34. For clarity, in Figure 4.35 only the two MβLp01 traces are shown in comparison to the positive control. The full set of data is included in the Appendices (A4.4.8). These results are somewhat contradictory, since the 'CFE' in the 96 well plate at the end of the reaction (Figure 4.34) is dark orange/red in colour, signifying a level of nitrocefin hydrolysis not too dissimilar from the positive control (as expected for a sample consisting of a milieu of endogenous *E. coli* proteins), yet the data of the same sample in the time course (Figure 4.35) suggests a similar level of activity to the purified enzyme, which is around seven times less active than the positive control. However, despite the similarity of this quantitative data, the two samples manifest completely different colours in the visual data, with the pure MβLp01 remaining yellow (inactive) and the crude lysate having undergone the characteristic colour change to dark orange. **This particular assay utilised a commercial kit; the pH of the provided proprietary assay buffer was not indicated but is assumed to be pH 7.0 since the definition of β-lactamase activity according to the manufacturer's instructions is quoted as at pH 7.0.*

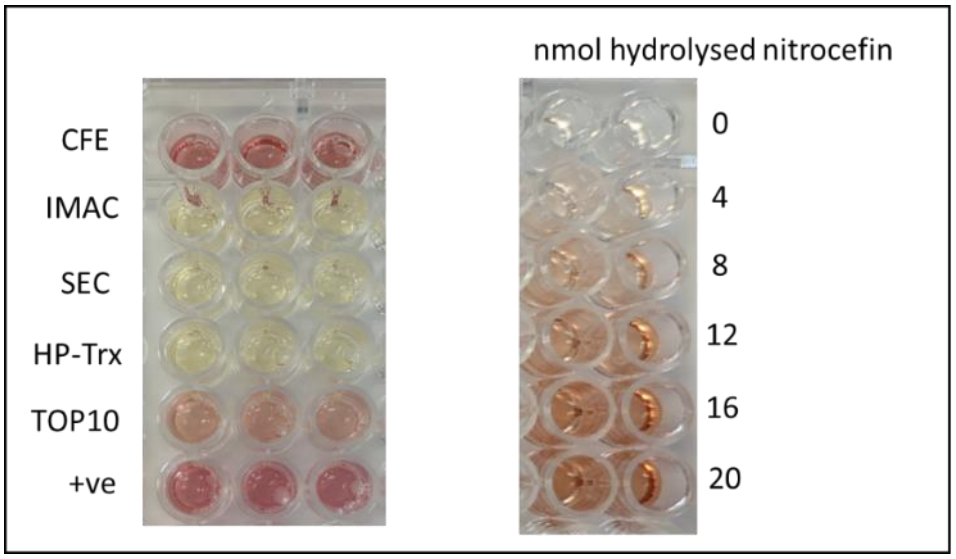


Figure 4.34 Colour development of MβLp01 samples and controls at the end of the course of reaction with nitrocefin (left) and of hydrolysed nitrocefin standards (right).

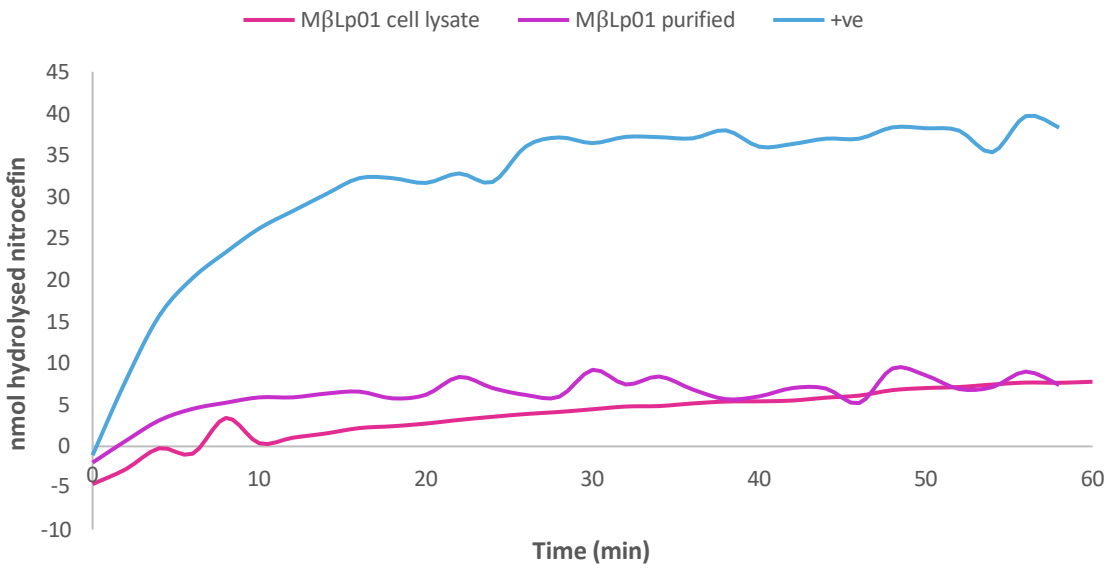


Figure 4.35 Hydrolysis of nitrocefin by MβLp01-expressing *E. coli* cells and purified MβLp01 in comparison to the positive control supplied in the commercial kit. N.B It was not possible to display error bars due to a technical issue in Microsoft Excel.

As such, from the data, it is unclear if MβLp01 actually hydrolysed nitrocefin or whether the data merely represent noise. There is a substantial difference between the positive control and MβLp01 – 38.3 nmol vs 7.4-7.6 nmol hydrolysed nitrocefin, respectively. The reaction time was empirical and it may be the case that MβLp01 was indeed hydrolysing nitrocefin, albeit much slower than the positive control. Since this assay is only an indicator of the enzymatic activity against one type of β-lactam substrate, a different approach was taken to investigate the full spectrum of β-lactamase activity of MβLp01 further.

4.4.6.4.2 β-lactamase susceptibility tests

Determination of the MICs of β-lactam antibiotics for *E. coli* incubated with or without a suspected β-lactamase is one approach for identification of β-lactamase activity. Whilst the presence of a β-lactam antibiotic will inhibit or retard the growth of susceptible bacteria, addition of a β-lactamase will hydrolyse and therefore deactivate the antibiotic, resulting in increased bacterial growth relative to a β-lactamase-free equivalent. There are a few different techniques for MIC determination along with variations on these, the core methods being disc diffusion, agar dilution and broth dilution. In this investigation, variations of both disc diffusion and agar dilution methods were performed. It is important to note that in clinical applications, there are specific antibiotic susceptibility testing (AST) guidelines (e.g. EUCAST) which must be strictly adhered to that in this work were less important since this work was not involving patients or contributing to any clinical data. Of course, as with any experiment, efforts were made to maintain consistency in procedure, account for all variables and generally conduct the investigation in a scientifically robust manner.

To determine the MIC of an antibiotic by the agar dilution method, a sterile preparation of antibiotic at varying concentrations is added directly into the molten agar before pouring plates. A bacterial culture (of a specific cell density) is plated out and incubated at 37°C overnight. The number of bacterial colonies theoretically should be negatively correlated with antibiotic concentration, that is, the number of bacterial colonies should decrease with increasing antibiotic concentration. The approximate MIC is the concentration at which bacterial growth is fully inhibited. An example of the results of an agar dilution experiment performed in this work is included in Figure 4.36.

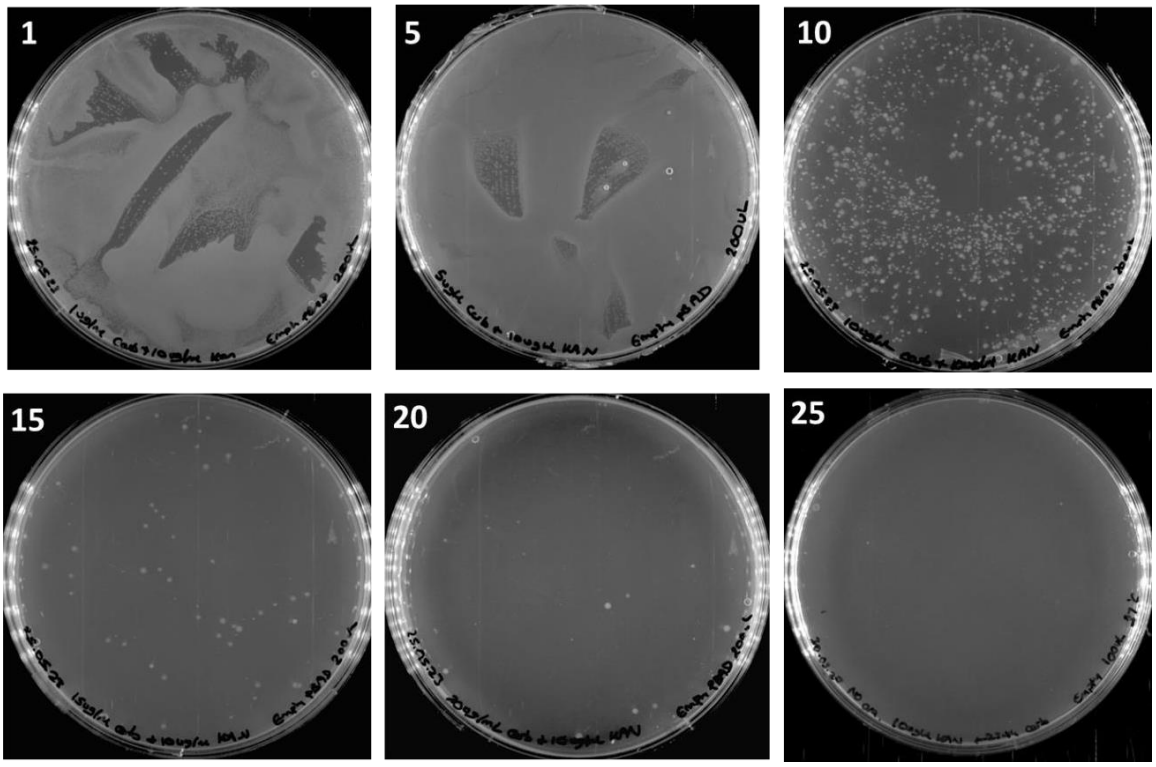


Figure 4.36 Example of the agar dilution method of β-lactamase activity determination. *E. coli* TOP10 cells expressing the empty vector were spread onto LB agar plates containing 10µg/mL kanamycin and 1, 5, 10, 15, 20 or 25 µg/mL carbenicillin.

The activity of MβLp01 against three different β-lactam antibiotics – carbenicillin, ceftazidime and doripenem – was investigated using the agar dilution procedure. Ballpark MIC values were sought from the literature for each antibiotic against *E. coli* (no specific strain) to identify suitable preliminary concentrations to test, which were then refined as necessary. In duplicate, *E.coli* TOP10 cells hosting the MβLp01 gene were spread on LB agar plates containing arabinose (various concentrations) and kanamycin in addition to the β-lactam antibiotic tested. Colonies formed after overnight incubation at 37°C were counted (unless there was confluent growth) and compared to the control - *E. coli* TOP10 cells carrying the empty pBD202/D-TOPO vector. Unfortunately there was no access to a positive control due to limitations on resources (time and finances). The results were somewhat challenging to interpret, especially given the lack of positive control. For example, preliminary tests with varying arabinose levels revealed that at 37°C, the inclusion of 0.2% arabinose prevented any bacterial growth, yet at room temperature, the same level of arabinose yielded colonies. It was possible that the higher arabinose concentration and higher temperature was slowing the growth of bacteria, potentially due to greater expression of MβLp01. It is sensible to suppose that there were too many influencing factors to enable conclusions to be drawn. A particular obscuring factor was the requirement of kanamycin inclusion for ensuring only recombinant *E. coli* cells harbouring

either MβLp01 or the empty vector were able to grow. Due to concerns about interference with the bacteria's response to the β-lactam antibiotic, the kanamycin level was lowered to one fifth of the standard working concentration. In hindsight, an additional control of non-recombinant *E. coli* TOP10 could have been used to ascertain the approximate MICs in the absence of additional compounding factors. It can be seen that there is a small difference between the MIC for ceftazidime between MβLp01 and the empty vector – for the empty vector, the MIC was around 0.75 μg/mL ceftazidime whereas for MβLp01 it was somewhere between 0.5 and 0.75 μg/mL (Table 4.9). It would be expected that if MβLp01 indeed has cephalosporinase activity, the MIC would be raised relative to the empty vector, as the bacteria expressing MβLp01 would be less susceptible to ceftazidime.

	MβLp01	**Empty pBAD202/D-TOPO**
Carbenicillin		25 μg/mL
Ceftazidime	0.5-0.75 μg/mL	0.75 μg/mL
Doripenem		

Table 4.9 Approximate MICs of three β-lactam antibiotics in the presence of either MβLp01 or the analogous empty vector, ascertained by agar dilution tests.

Since agar dilution tests did not produce conclusive results, additional AST-based methods were carried out. A set of disc diffusion experiments using purified MβLp01 were more fruitful and revealed a consistent difference between MβLp01 and the negative control. In this procedure, sterile discs were impregnated with specific amounts of a β-lactam antibiotic and incubated overnight – with or without MβLp01 - at 37°C before placing onto LB agar plates onto which *E. coli* DH5a cells had just been spread. On each plate a control disc saturated with MβLp01's buffer and treated the same as the experimental discs was included, Each treatment was tested in triplicate. After overnight incubation at 37°C, the zones of inhibition around each disc were measured. In theory, the presence of β-lactamase activity would result in a decrease in the zone of inhibition around a disc containing the enzyme.

Figure 4.37 highlights the clear difference of with versus without MβLp01 for doripenem, most pronounced for the 0.25μg/mL doripenem treatment. For each of the four doripenem concentrations tested, MβLp01 exhibited a consistently smaller zone of inhibition. The analogous quantitative data is presented in Figure 4.38.

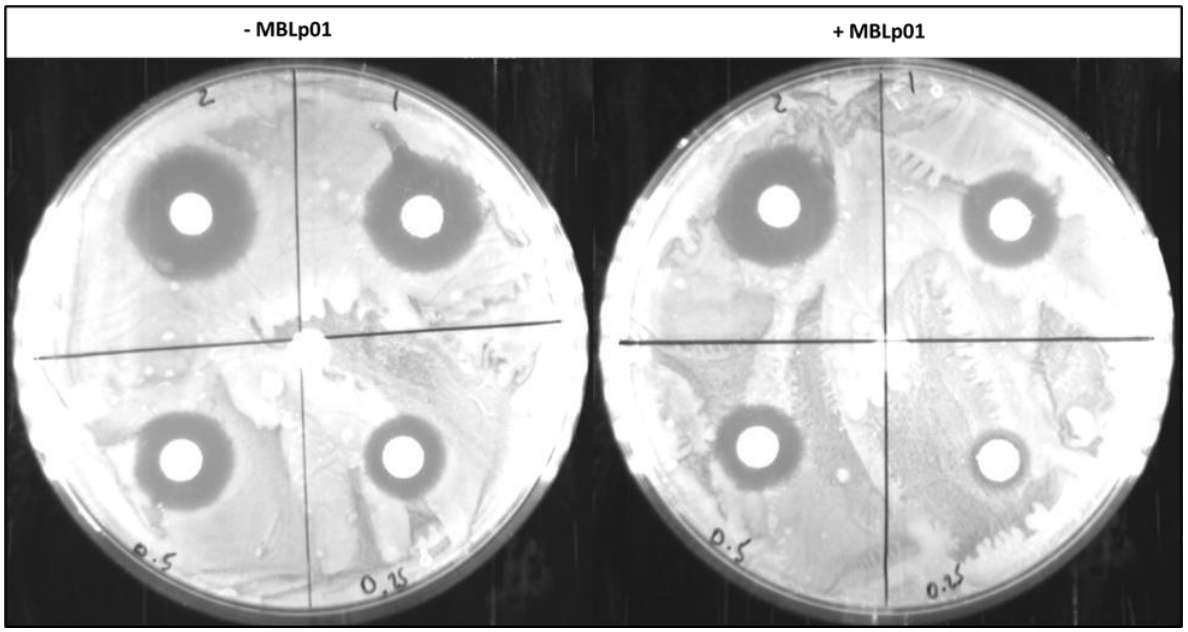


Figure 4.37 Zones of inhibition generated by doripenem at 2, 1, 0.5 and 0.25 µg/mL, upon incubation with or without purified MβLp01, on agar plates spread with *E. coli* DH5α. The 2, 1, 0.5 and 0.25 µg/mL discs are in the upper left; upper right; lower left; and lower right quadrants of each plate, respectively. Control disc containing buffer only is in the centre of each plate.

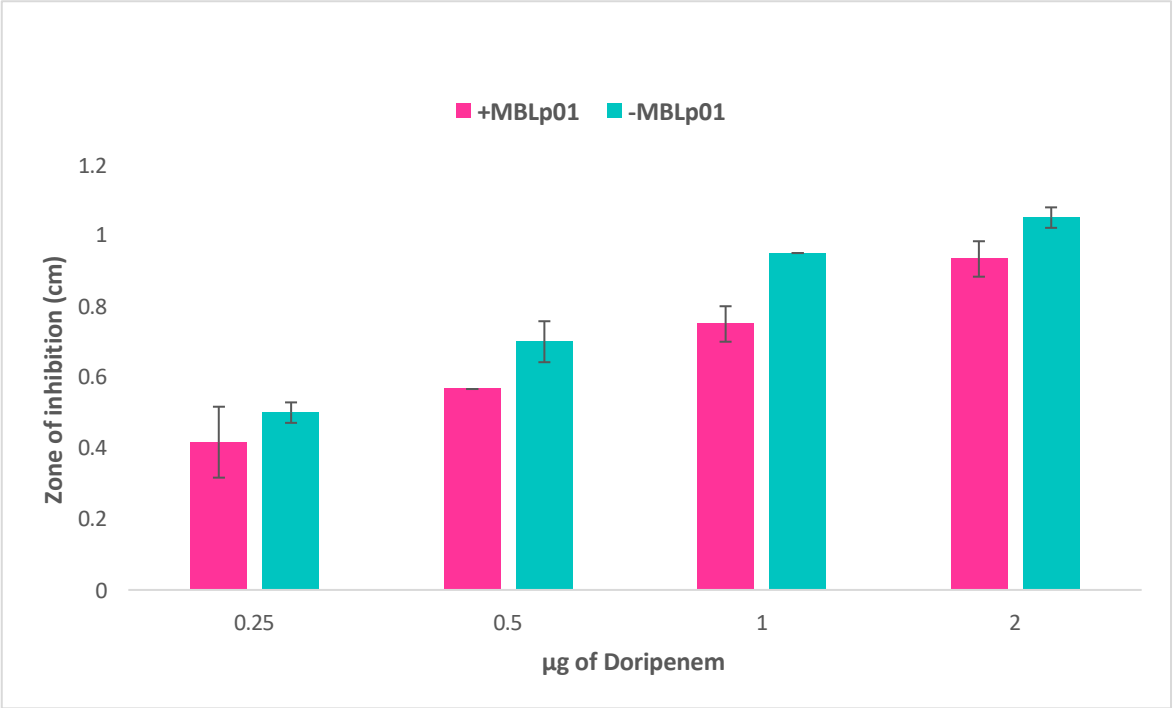


Figure 4.38 Sizes of inhibition zones produced by discs impregnated with four different concentrations of doripenem with or without purified MβLp01, on agar plates of *E. coli* DH5α.

It is clear from this data that MβLp01 has hydrolytic activity against doripenem. At 0.25, 0.5, 1 and 2 µg/mL doripenem, the presence of MβLp01 made a difference of 18.2%, 21%, 23.5% and 11.8%, respectively – an average difference of almost 20%.

For carbenicillin, the difference was more significant, with the inclusion of MβLp01 resulting in a substantially decreased inhibition zone (Figures 4.39 and 4.40). Indeed, at a carbenicillin level of 20 µg/mL, the addition of MβLp01 abolishes the inhibition zone entirely, effectively rendering the *E. coli* strain resistant to the penicillin. At 10µg/mL carbenicillin, the *E. coli* is resistant to the β-lactam regardless of MβLp01 presence. The differences were 200% at 20µg/mL carbenicillin, 51.9% at 30µg/mL carbenicillin and 20.2% at 40µg/mL carbenicillin.

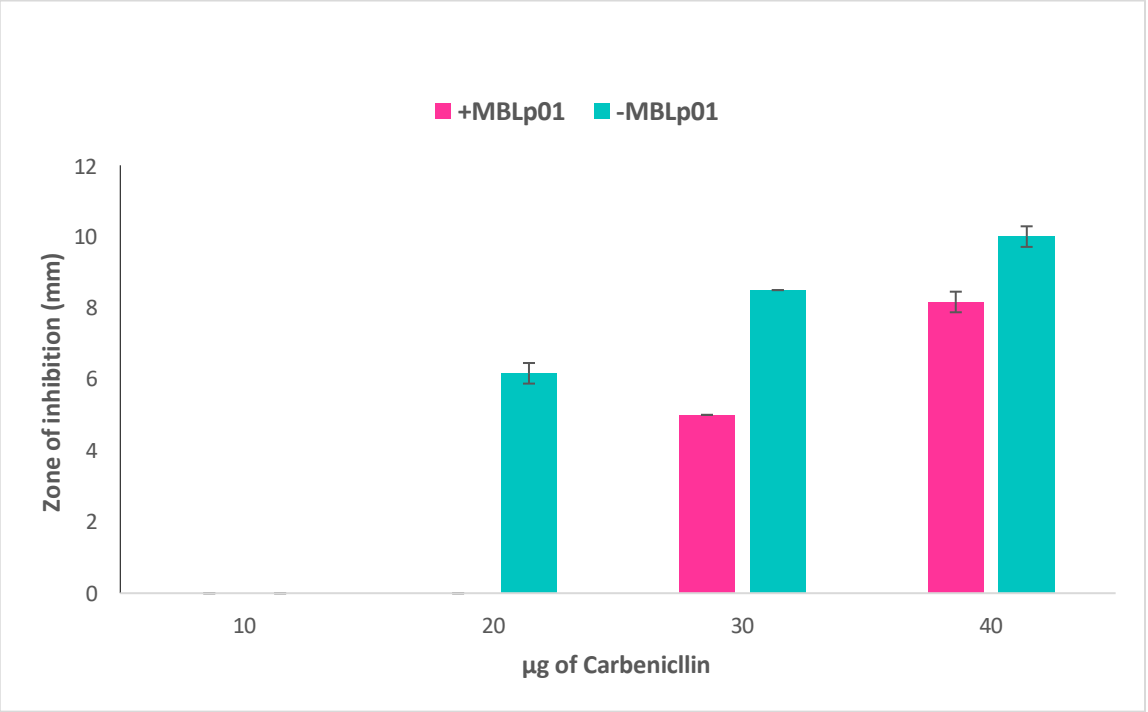


Figure 4.39 Sizes of inhibition zones produced by discs impregnated with three different concentrations of carbenicillin with or without purified MβLp01, on agar plates of *E. coli* DH5α.

Ceftazidime on the other hand did not appear to be affected by the inclusion of MβLp01, since there was either a decrease of 1.8% (0.5µg/mL ceftazidime), or increases of 7.4%, 4.4% and 1.6% for 0.15µg/mL, 0.25µg/mL and 0.75µg/mL ceftazidime, respectively, in the sizes of the zones of inhibition (Figures 4.41 and 4.42). Since the differences are either relatively small or non-existent, and inconsistent, it can be presumed, from this data at least, that MβLp01 does not have activity against ceftazidime.

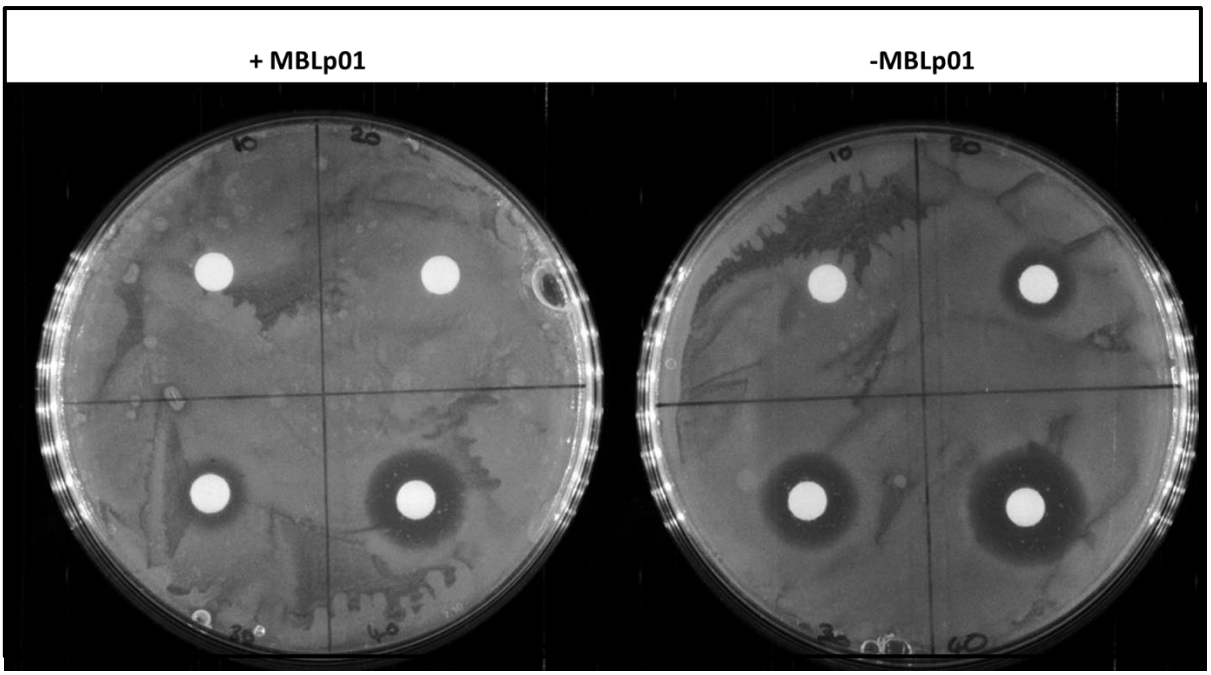


Figure 4.40 Zones of inhibition generated by carbenicillin at 10, 20, 30 and 40 µg/mL, upon incubation with or without purified MβLp01, on agar plates of *E. coli* DH5α. The 10, 20, 30 and 40 µg/mL discs are in the upper left, upper right, lower left and lower right quadrants of each plate, respectively.

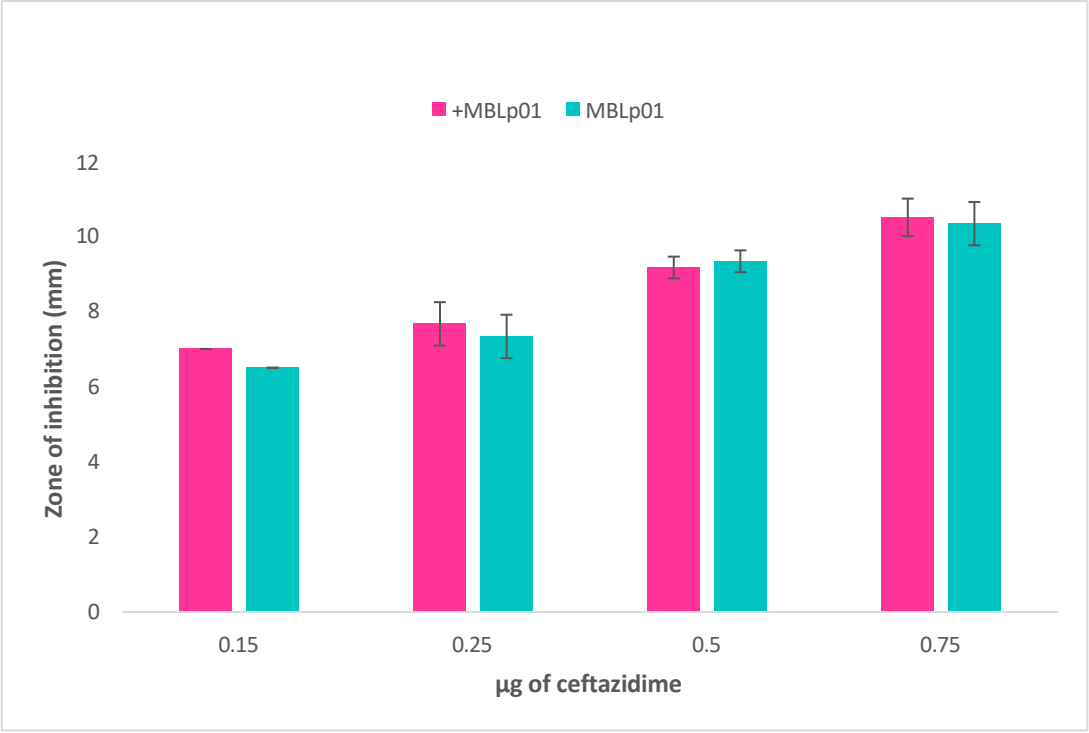


Figure 4.41 Sizes of inhibition zones produced by discs impregnated with four different concentrations of ceftazidime with or without purified MβLp01, on agar plates of *E. coli* DH5α.

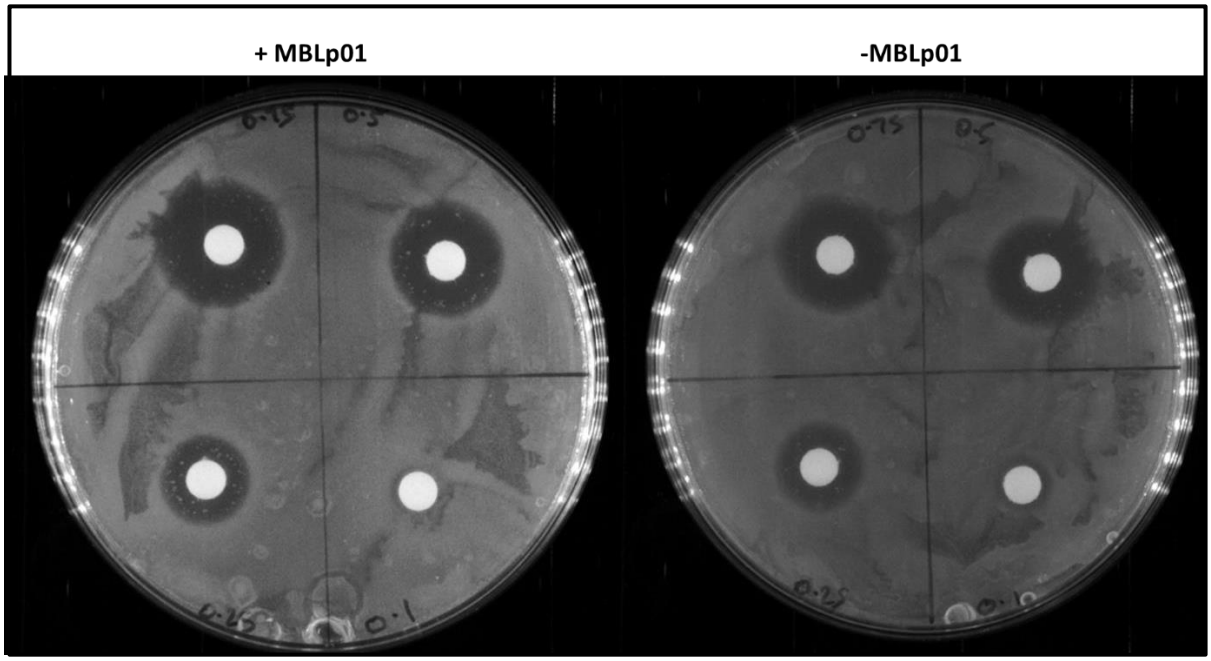


Figure 4.42 Zones of inhibition generated by ceftazidime at 0.15, 0.25, 0.5 and 0.75 µg/mL, upon incubation with or without purified MβLp01, on agar plates of *E. coli* DH5α. The 0.15, 0.25, 0.5 and 0.75 µg/mL discs are in the lower right, lower left, upper right and upper left quadrants of each plate, respectively.

The conclusions that can thus be drawn from this set of data is that MβLp01 has carbapenemase and penicillinase - but not cephalosporinase - activity. Therefore, one could speculate that MβLp01 is indeed a β-lactamase. Based on recent literature surrounding the evolution of the MβL superfamily (Chapter 4.2.4), it is plausible to suppose that MβLp01 is a β-lactamase with promiscuous phosphodiesterase activity.

4.5 Discussion

The work in this chapter presents a number of points for deliberation and speculation. First and foremost, activity assays revealed the unexpected lack of phytase activity of both MβLp01 and MβLp02 – that is, the ability to catalyse the hydrolysis of IP6 specifically. This bore two key questions 1) what is/are the actual substrate/s of these enzymes? 2) Why did the authors of the published paper, *'Functional Metagenomics Reveals a New Catalytic Domain, the Metallo-β-Lactamase Superfamily Domain, Associated with Phytase Activity'*, report phytase activity, or rather, what gave rise to the phytase activity in their investigations that led to the reporting of MβLPhy as a novel phytase class[267]? The latter question shall be addressed first. *N.B* when 'this work' is referred to, it means the work performed for this thesis chapter, and not the work performed by the Daniel group in the aforementioned paper.

4.5.1 Possible reasons for source of phytase activity in published paper

In the published paper, MβLp01 and MβLp02 were purified only by Ni^{2+}-NTA IMAC, without a further size exclusion step. With any *E. coli* expression strain there are a milieu of endogenous host proteins which have some extent of affinity for either the Ni-NTA matrix or the target protein itself, and inevitably co-elute with the target recombinant protein. If assays are performed on an IMAC purified sample, it is not just one enzyme which is responsible for any enzymatic activity, but all of the co-eluting enzymes too. AppA is a periplasmic *E. coli* phytase which may co-elute with the target protein, highlighting the importance of assay controls and accounting for any background phytase or phosphatase activity, bearing in mind that assays based on ammonium molybdate are not phytase specific and are detection methods of total inorganic phosphate in a sample – whether liberated through enzymatic activity or already present in the sample. Indeed, in the experimental work in this thesis chapter, samples from each stage of purification were assayed for phytase and phosphatase activity – the total protein in the cell lysate, the soluble protein in the clarified cell lysate, the insoluble protein in the cell pellet, the IMAC flow through, wash, elution, and the pure protein after size exclusion chromatography. It is clear from the experiments in this chapter that any phytase activity resulted from the less pure samples containing co-eluting host cell proteins, or from the background phosphate contribution from the substrate itself. Since the published paper did not give any indication of the purity or yield of MβLp01 and Mβlp02, it is impossible to infer that any phosphate released from phytic acid is attributed to the enzymatic activities of MβLp01 and MβLp02. Further, the paper does not indicate if any substrate background or empty vector controls were included in the substrate specificity assay. The data presented in this chapter revealed that the phosphate released in the molybdenum blue assay arose from the substrate

itself – that is, subtraction of the A_{700nm} value of the substrate only control from the value of the MβLp01 or MβLp02 samples resulted in a value of 0 or less. In personal communications, the authors of the paper stated that they 'never got an ultra-high level of enzyme purity' and that 'the activity of those proteins was hard to get for a long time, because they seemed to lose their activity very fast with the time[322]'.

To ensure no factors were overlooked, every step of the workflow from the gene sequence to the final purified enzyme used in activity assays – in both the published method and the methods used in this thesis - was scrutinised to check for any disparities which may affect the behaviour of the MβLPhy enzymes or give rise to sources of free orthophosphate in assays. In addition, once absence of phytase activity was established, a number of parameters were varied and tested, such as composition of purification buffers and inclusion versus exclusion of zinc sulphate.

4.5.1.1 Differences between the purification methods

First, in this work, the genes for MβLp01 and MβLp02 were codon optimised for expression in *E. coli* whereas in the published work they were not. Since the protein sequence is unaffected by codon optimisation, this was ruled out as a cause of the difference in observed results. Indeed, it is established that neither the guanine-cytosine content, codon usage nor mRNA stability are critical when determining the resistance phenotype or expression levels in different hosts[323]. Second, in the published method, *E. coli* LMG194 was used as the expression host strain, however, attempts to render this strain chemically competent were unsuccessful and since expression from the transformation host *E. coli* TOP10 was satisfactory, this strain was instead used herein. The genotypes of these two strains are very similar with the main difference being that basal expression can be better suppressed from LMG194, useful for potentially cytotoxic proteins. Growth of LMG194 is most suitable in Minimal Media so this was used by the Daniel group, whereas in this work, LB was used for the culturing of *E. coli* TOP10. Although minimal media contains phosphate salts, this should not affect downstream analysis. In terms of expression, the Daniel group harvested induced cells after 5h incubation at 37°C whereas in this work, once cultures had been induced, they were incubated at 16-18°C overnight. This is because the consensus is that heterologous proteins express better at lower temperatures since they have time to fold properly and don't have the tendency to aggregate into insoluble inclusion bodies.

The next differences lie in the composition of the purification buffers. For resuspension of cells and binding/washing the IMAC resin, the Daniel group used a buffer composed of 50 mM HEPES pH 8.0, 250 mM NaCl and 0.5 mM $ZnSO_4$. Two points of note: Firstly, there is no

imidazole which is likely to result in greater levels of co-eluting host cell proteins, at least in the earlier stages of elution. A concentration of 20-40 mM imidazole is widely used in the binding/wash buffer to limit binding of non-specific host proteins to the column. Other than when testing the buffers employed by the Daniel group, buffers used in the work in this chapter consistently contained 20 mM imidazole. However, their use of a Ni-TED IMAC resin is the reason for the absence of imidazole in the lysis/equilibration/wash buffer, since the strong pentadentate tris(carboxymethyl)ethylenediamine (TED) chelating group has greater specificity than NTA or IDA and therefore the manufacturer recommends 0 mM imidazole. Whereas NTA occupies four of the six binding sites of the $Ni2^{+}$ ion leaving two sites available for protein binding, TED occupies five of the binding sites in the $Ni2^{+}$ coordination sphere, leaving just one site available for protein binding[324]. Of course, using 0 mM imidazole with a resin other than Ni-TED will result in a much lower level of purity. Second, $ZnSO_4$ is insoluble between pH 7.5 -11 in aqueous solution, with solubility increasing with acidity[325]. It has been demonstrated that raising the pH from 7.0 to 8.0 decreases solubility by a factor of 100[326]. Indeed, when performing this purification as per the published protocol (N.B. as far as could be achieved – several parameters were unable to be changed in this case, such as expression strain and accompanying media, as explained previously), precipitation (white) was observed upon addition of $ZnSO_4$ to the HEPES buffer. The likely cause would have been precipitation of the zinc hydroxide formed upon adjusting the pH of the solution with sodium hydroxide. However, according to Krezel and Maret, 'a relatively high solubility of $Zn(OH)_{2(s)}$ is the reason why addition of $ZnSO_4$ or $ZnCl_2$ at sub-millimolar concentrations to a HEPES buffer (at a pH close to the pK_a of the buffer) does not result in precipitation'[327]. However, the subsequent reaction of zinc hydroxide with sodium hydroxide results in sodium zincate which is completely insoluble. Regardless, precipitation will lower the actual amount of $ZnSO_4$ in solution. The pH range of HEPES is 6.8 – 8.2, therefore using HEPES at a pH of 6.8 would have been more appropriate to prevent precipitation of the $ZnSO_4$. An alternative would be to use zinc acetate since this has greater solubility at pH 8. Nevertheless, the reader is referred to a comprehensive review of the biological chemistry of zinc, which is particularly relevant to this chapter[327]. As for the elution buffer, as recommended by the manufacturer of the Protino® Ni-TED kit, the Daniel group used a buffer consisting of 50 mM HEPES pH 8.0, 250 mM NaCl, 250 mM imidazole and 0.5 mM $ZnSO_4$. Again, other than when this published procedure was replicated, the elution buffers used for the purification in this work contained 500 mM imidazole.

Following collection of the target protein-containing elution fractions, (as determined by SDS-PAGE) the Daniel group used a spin concentrator with a molecular weight cut-off (MWCO) of 30 kDa. Since the molecular weight of the MβLp01 fusion is 44.2 kDa, using a cut-off of 30 kDa

is not recommended, since the general rule of thumb is to not use a MWCO greater than a third of the mass of the target protein. As such, in the purification procedures performed in this chapter, spin concentrators were used with MWCO of 3000, 5000 or 10,000 kDa, and even when performing as close as possible the specific purification procedure reported by the Daniel group, a MWCO of 10,000 kDa or less was used – since all that using a 30 kDa MWCO concentrator would serve to do is jeopardise yield. Following concentration, in this work, the sample was then subjected to desalting on a HiPrep desalting column to remove imidazole, which then necessitated further concentration of the now-diluted sample pool. The Daniel group used an enterokinase cleavage capture kit for removal of the HP-thioredoxin tag, whereas in this work, enterokinase (recombinant human protein, active) was used to cleave the tag before performing reverse IMAC to separate the cleaved protein from the HP-thioredoxin – which is essentially what the enterokinase cleavage capture kit serves to do in gravity flow format, but in addition, it separates the enterokinase protein from the target protein by its proprietary enterokinase-binding resin. The co-eluting enterokinase in the procedure implemented in this work was removed during the later SEC stage. Due to limitations with enterokinase cleavage efficiency, the majority of the MβLp01 and MβLp02 samples purified – not including the MβLp01 samples used for crystallisation and structure solution – were not subject to tag cleavage since they were used in assays alongside a purified HP-thioredoxin protein as a control. At this point, after cleavage of the tag, the Daniel group proceeded to activity assays after a final concentration on a 10 kDa MWCO. In their published paper, an SDS-PAGE of the cleaved MβLPhy enzymes was not absent so the level of purity is unknown, as is concentration, which was not reported. In the work in this thesis, following desalting and concentration, the samples were subjected to SEC to further separate the target protein from contaminants and also from enterokinase in the cleaved MβLP01 procedure. Concentration was then estimated by UV/Vis spectroscopy on a NanoDrop One. As exemplified in chapter 4.4.4, the resultant purified MβLp01 was of high purity with negligible levels of contaminating proteins.

4.5.1.2 Differences between determination of phytase activity

For detection of released orthophosphate, the Daniel group used a variant of the molybdenum blue assay in which it is the bright yellow phosphomolybdate complex which is detected, rather than the molybdenum blue product formed upon reduction. Absorbance is thus measured at 355 nm whereas 700 nm is used for detection of the resultant molybdenum blue. The phosphomolybdate method is based on a solution of acetone, sulphuric acid and ammonium molybdate (AAM), whereas in the molybdenum blue protocol the developing solution consists of ammonium molybdate, sulphuric acid and ferrous sulphate. Aside from these fundamental

differences in developing solution, all other reaction parameters in this work were aligned closely to the protocol in the paper, such as pre-incubation of enzyme for 3 min at the specified temperature, an incubation time of 30 min, the buffer and pH, and the concentration of substrates tested. In terms of controls and reliability of the results, both the Daniel group and the work in question performed reactions in triplicate. The Daniel group measured samples against blanks in which the AAM solution was added prior to addition of the enzyme. This is essentially the equivalent of measuring the developing solution alone as a blank in the molybdenum blue protocol. In this work, in addition to measuring the absorbance of the developing solution, standards were included to estimate the non-enzymatic hydrolysis of substrate and any other contributing phosphate. As such, triplicate measurements were taken of solutions of substrate only, assay buffer only, enzyme only, IMAC buffer only and SEC buffer only and all of these underwent the same procedure as the experimental samples i.e. pre-incubation followed by 30 min incubation and then stopping the reaction by thermal deactivation (100°C, 10 min). The outcome of these assays highlighted the importance of the substrate only control, since it was revealed that the majority of detected phosphate originated from non-enzymatic hydrolysis of the substrate itself. In addition to these 'background contribution controls', a positive control - an in-house purified MINPP - and a 'background' control – the in-house purified HP-thioredoxin peptide - were also included. The latter was necessary since the majority of assays utilised the HP-thioredoxin MβLp01 fusion rather than the enzyme post-cleavage. Aside from the developing solution blank, the published paper does not reference any additional controls. This adds further justification to the hypothesis of this thesis that the phytic acid-degrading activity observed by the Daniel group was not associated with MβLp01/MβLp02. Further, HPLC of reaction products of ultra- pure *myo*-IP6 incubated with MβLp01 established the lack of phytase activity in this work.

4.5.2 Substrate specificities of MβLp01 (and Mβlp02)

The matter of substrate specificity shall now be addressed, taking into account the *au courant* literature.

This thesis has presented evidence which contests the previous reporting of MβLp01 and MβLp02 as metallo-β-lactamases with phytase activity. In the published paper, the set of phosphorylated compounds tested were: ADP, ATP, NADP, glucose-6-phosphate, glycerol phosphate, pyrophosphate, 1-naphthyl phosphate, pyridoxal phosphate and phytic acid, with ATP and glucose 6-phosphate being the two substrates yielding the highest activities for both enzymes[1]. Activities were stated as percentage activity relative to ATP, with the value of ATP

being set to 100%. Therefore, when MβLp01 is reported to have an activity against phytic acid of 18% +/- 3.1% relative to ATP (11% +/- 1.2% for MβLp02), it is not actually known what the absolute activity against ATP was, and correspondingly the absolute activity against phytic acid irrespective of ATP. Indeed, the authors implicate the significantly lower catalytic efficiencies of MβLp01 and MβLp02 towards phytate than of other cultured single microorganism-derived phytases. As such, it is impossible to compare activities determined in this thesis to the published values, since the values published are only comparable among each other. In this thesis, the set of phosphorylated compounds which were subjected to testing for phosphohydrolase activity by MβLp01 were: phytic acid, ATP, ADP, glucose 6-phosphate, glucose 1-phosphate, pNPP and bis-pNPP. Therefore, the only common substrates were phytic acid, ATP, ADP and glucose 6-phosphate. The activities of phytic acid, ADP and glucose 6-phosphate relative to ATP could be determined and compared to the values reported in the published paper. However, of these substrates, only ATP was shown to be a possible substrate of MβLp01, with no observed activity against ADP, glucose 6-phosphate or phytic acid. However, the true extent of activity cannot be compared due to the relative nature of the published values.

In terms of β-lactamase activity, MβLp01 and MβLp02 were reported to render the bacterial strains expressing them less susceptible to all three types of β-lactam antibiotics – penicillins, cephalosporins and carbapenems. However, investigations in this thesis indicated a lack of effect on a β-lactam belonging to the cephalosporin class. It is appropriate to now address one important point – when assessing β-lactamase activity by AST-based methods, it must be appreciated that it is the resistance phenotype of the recombinant *E. coli* which is actually being determined – unless, as in some of the experiments in this thesis, the purified MβL is used instead. The resistance phenotype is quantitated based on the MIC of an antibiotic that inhibits bacterial growth, and it is important to note that these vary between hosts as a result of changes in cell permeability, the presence of complementary resistance mechanisms and the different environments present in each host which impact on expression levels, processing and activity of MβLs[289]. Indeed, studies have demonstrated a key role of the signal peptide in host specificity, in that the signal peptide processing could give rise to the host-dependent phenotypes[328]. Accordingly, the *in vitro* study of a purified mature protein overlooks the impact of mutations in the signal peptide that may affect the translocation and levels of functional protein in the periplasm, their cellular localisation, and the correct folding or stability.[289] For more extensive discussion of these considerations, the reader is referred to the recent publication by Lopez et al[289]. Of course, in this set of studies, MβLp01 and MβLp02 were expressed without their signal peptides, and for the most part were analysed at the purified protein – rather than host-

dependent - level. For example, although preliminary disc diffusion tests were performed using recombinant *E. coli* expressing MβLp01/MβLp02, these were superseded by a modified method in which purified enzyme was incubated with the β-lactam antibiotic on the disc. This method was used in the work on Igni18 and provided the general principle for the protocol used in this thesis[266]. The inclusion/exclusion of the signal peptide was not addressed in the published work by the Daniel group and since their β-lactamase tests were performed at the phenotypic level, this is an important aspect to acknowledge. Nevertheless, the Daniel group concluded that MβLp01 and MβLp02 are promiscuous enzymes exhibiting both phytase and β-lactamase activities[267]. In contrast, the work in this thesis indicates a distinct lack of phytase activity and a less broad spectrum β-lactamase activity profile, with the additional presence of phosphodiesterase activity also revealed, which was not screened for in the published paper.

It was reported that the closest structural homologue of MβLp01 was the *E. coli* zinc phosphodiesterase, ZiPD (PDB entry: 2CBN), encoded by the elaC gene[267,329]. ZiPD is involved in the tRNA maturation process, where it acts as a 3' tRNA-processing endonuclease and belongs to the tRNase Z family of enzymes – within the MβL superfamily - which have been identified in a wide variety of organisms[329-331]. At the time of solving the structure of MβLp01 by molecular replacement (a few years after the publishing of the MβLp01/MβLp02 paper), the closest structural homologue was the B3 MβL PNGM-1 from a deep-sea sediment metagenome which predated the era of antibiotics[320,321]. Dual activity of PNGM-1, having both metallo β-lactamase and ribonuclease activities, was demonstrated by Lee et al in 2019 and contributed – along with observed structural similarities to tRNase Z enzymes - to the postulation that PNGM-1 evolved from a tRNase Z, and in turn, gave rise to the subclass B3 MβLs via a promiscuous metallo β-lactamase activity[297]. MβLs and phosphatases are both considered promiscuous enzymes, with a deep and complex evolutionary relationship between them. Indeed, in both ZiPD and PNGM-1, the metallo β-lactamase activity is thought to be promiscuous, with ribonuclease activity being the native function. Therefore, it is conceivable, given the close alignment of the metal binding site geometries of MβLp01 and of PNGM-1, that MβLp01 is a promiscuous, dual-function β-lactamase and phosphodiesterase. This is in alignment with the suggestion in recent literature that a phosphodiesterase involved in nucleotide processing was the last common ancestor of the MβL superfamily[296]. Indeed, the large active site volume afforded by the MβL fold is a key determinant of enzyme promiscuity, facilitating the binding of various substrates of a range of complexities. Further *in vitro* investigations would need to be performed to explore the physiological phosphodiester substrate/s of MβLp01, since bis-pNPP is a purely synthetic compound. For example, DNA, RNA and cyclic nucleotide phosphodiester substrates such as cAMP should be tested. Ideally, more extensive β-lactamase

screening would have been completed, to reinforce the results obtained from the disc diffusion-based tests and the nitrocefin hydrolysis assays. Orthogonal methods would complement one another well, for example, colorimetric assays, broth microdilution[332], the acidimetric method[333] and the iodometric method[334].

Due to time restraints, it was not possible to test MβLp01 with all the compounds identified as being potential substrates. In addition to testing phosphorylated compounds, the use of a random substrate screen, such as Taxa Profile E plates (Merlin), would have been very convenient since they contain a range of chromogenic substrates.[329] Nuclease activity could have been tested by incubating DNA with MβLp01 before analysis of the reaction products (if any) by DNA agarose gel electrophoresis.[329] There is also the colorimetric methyl green method for determination of DNase activity, in which hydrolysis of DNA generates unbound methyl green and an absorbance decrease at 620 nm[335]. Ribonuclease activity could have been probed by analysis of the absorption change at 260 nm after addition of MβLp01 to, for example, yeast RNA Type II[329]. Since an arylsulfatase was determined by the DALI server[318] as a close homologue of MβLp01, arylsulfatase activity could have been tested using the substrates *p*-nitrocatechol sulphate and *p*-nitrophenyl sulphate.[329] Hydrolysis of cyclic mononucleotides could have been investigated by incubation of the MβLp01 reaction products with 5'-nucleotidase to liberate phosphate from the open nucleoside monophosphate before detection of the inorganic phosphate by the molybdenum blue assay[329]. Another compound often used for the assessment of phosphodiesterase activity, specifically 5'-nucleotide phosphodiesterase activity, is thymidine-5'-monophosphate-*p*-nitrophenylester (TpNPP), which could have been used in addition to bis-pNPP[336]. In addition to the limitation on testing a more thorough breadth of substrates, financial restriction meant that positive controls (purchased) were not present in the β-lactamase tests (except for the nitrocefin hydrolysis assay) and the phosphodiesterase assay performed. In addition, the aforementioned time restrictions prevented the in-house expression and purification of a known β-lactamase - such as, for example, the B3 MβL L1 from *Stenotrophomonas maltophilia* - and a known phosphodiesterase. For the same reasons, it was not possible to generate a standard curve of *p*-nitrophenol for the quantification of product formed from the hydrolysis of pNPP and bis-pNPP by MβLp01. However, for the purposes of this thesis, this was not critical, since the primary goal was not absolute quantification, but merely for comparison of relative values.

4.5.3 Additional Points

It was noted that the inclusion of zinc sulphate in the SEC buffer resulted in a change in the retention time of MβLp01, causing it to elute after the column volume. It appears that the presence of zinc in the buffer effectively rendered MβLp01 'sticky' such that its mobility through the column was retarded.

The Daniel group had reported that DTT reduced activity of both MβLp01 and MβLp02 to less than 15%, therefore DTT was never included in buffers.

4.5.4 Concluding Remarks

There are currently only a limited number of MβL structures available for the investigation of structure-function relationships, therefore, the structure of MβLp01 solved in this thesis may contribute to future studies involving enzymes of the MβL superfamily. Further, at the time of writing, there are no available structures of any non-hydrolysed substrate β-lactam bound MβL – only structures in which the substrate is hydrolysed exist in the PDB[269]. One valuable aspect of future work would be to solve the structure of MβLp01 with a non-hydrolysable β-lactam analogue, which would be useful in the area of β-lactamase inhibitor development in addition to studies of MβL enzymes. As an alternative or interim approach, molecular docking could be used to investigate the binding of β-lactam substrates into the active site of MβLp01.

Recently (July 2023), there have been review articles published which have incorporated into their narrative the concept that MβLp01 and MβLp02 represent the first MβL superfamily enzymes with phytase activity[271,273,277]. The original paper by the Daniel group is thus influencing studies on the evolutionary origins of MβL enzymes which could be very detrimental given the insights yielded by this thesis. It is therefore imperative that 1) the structure of MβLp01 and 2) the lack of phytase activity are reported because the structure is a very valuable resource which can be used by other researchers and dissemination of the current evidence of its activity will prevent the waste of time and resources for researchers solely interested in phytases. To reiterate - ***'Just because something doesn't do what you planned it to do doesn't mean it is useless'.***

CHAPTER FIVE

Structural insights into a plant inositol triphosphate kinase, ITPK1

5.1 Introduction

Note that, in lieu of coverage in the main thesis introduction, this chapter introduction serves as a chapter-specific literature review.

Whilst the core theme of this thesis is the enzymatic degradation of IP6, this chapter is focussed on a different - yet equally important – aspect of inositol phosphate metabolism in plants. Ultimately responsible for the level of phytate in food crops, it is therefore of substantial agricultural interest to establish the regulatory mechanisms which govern inositol phosphate signalling in plants[339,340]. Cellular inositol phosphate levels are maintained by an elaborate network of phosphate interchange, multiplexed by a sequence of kinase and phosphatase reactions and based on the combinatorial attachment of phosphate groups to the inositol ring. This yields an arsenal of messenger species, some of which are involved in cell signalling pathways utilised by all eukaryotes to regulate a diverse range of fundamental physiological process such as energy metabolism, phosphate homeostasis and immune response[341-343]. Indeed, the metabolic role of some of the higher inositol polyphosphates are well established, since knockout of genes involved in their biosynthesis in mice is lethal, indicating their indispensability[344]. The molecular complexity is further exacerbated by pyro-phosphorylation. Inositol pyrophosphates (PP-InsP) – also referred to as diphospho-*myo*-inositol polyphosphates - are inositol phosphate derivatives formed by the phosphorylation of IP6 to yield diphosphoinositol-pentakisphosphate (IP_{7}) or bisdiphosphoinositol-tetrakisphosphate (IP_{8}), and contain at least one energy-rich phosphoanhydride bond[342,345]. As molecular messengers, they play a key role in the regulation of a wide range of eukaryotic cellular processes such as energy metabolism, phosphate homeostasis and immune responses[339,342,345]. For example, they contribute to the defence against necrotrophic fungi and insect herbivores in plants, and are involved in lipid metabolism and nutrient sensing[345]. Indeed, IP_7 and IP_8 reportedly act as high- energy non-enzymatic phosphate donors in protein phosphorylation[341]. As a result, in recent years, inositol pyrophosphates have demanded considerable attention, dominating the inositol phosphate research domain[342]. The enzymes responsible for increasing the phosphate substitution state of inositol belong to a large multifaceted family of inositol phosphate kinases, of which a comprehensive description is beyond the scope of this chapter. Readers are thus referred to the review by Laha et al which is constructed from an evolutionary

perspective[342], and that by Wang and Shears which provides coverage of the wider scope of inositol phosphate kinase families, from a structural viewpoint[346]. In brief, there are four distinct kinase families prevalent in eukaryotes – inositol polyphosphate kinase (IPK); inositol 1,3,4-triphosphate 5/6 kinase (ITPK); IPPK (IP_5-2K) and $PPIP_5K$ – with the extensively characterised IPK family further divided into three subgroups: I(1,4,5)P_3-3 kinase (IP_3-3K), inositol phosphate multikinase (IPMK) and inositol hexakisphosphate kinase (IP6K)[342,346]. In most multicellular eukaryotes, there are two disparate routes of inositol phosphate synthesis which utilise different sources of cellular inositol. Thus, exogenously acquired inositol proceeds via the canonical lipid dependent pathway – common to both plants and animals[347] - whilst endogenously synthesised inositol, originating from the isomerisation of glucose-6P to inositol-3P, is directed through the cytosolic route[341,342] (Figure 5.1). Only the latter is subject to the action of inositol phosphate kinase enzymes[342]. Inositol phosphate biosynthesis has a distinct configuration in plants, yeast and animals/metazoan[341], for example, the ITPK enzymes are absent in yeast[348] whilst plants host many more ITPK members than animals[342]. Indeed, until recently, the consensus was that plants and yeast lack the Ins (1,4,5)P_3 3-kinases of the metazoan which generate the Ins (1,3,4,5)P_4 precursor of Ins(1,3,4)P_3[340], the latter being integral to IP_6 synthesis in animals, where it serves as a precursor for the synthesis of other higher inositol phosphate isomers. As such, it was thought that plants and mammals harbour redundant pathways for IP_6 biosynthesis, with plants lacking Ins(1,4,5)P_3 3-kinases and instead hosting Ins(1,3,4)P_3 5/6-kinases (ITPK). However, gene products of the Ins(1,4,5)P_3 3-kinases have since been identified in plants and fungi[349]. Nevertheless, in higher plants, biosynthesis of IP_6 is achieved by the actions of members of the ITPK family, which constitute the pathway to Ins(1,3,4,5,6)P_5 synthesis by the phosphorylation and dephosphorylation of various inositol phosphate intermediates, before phosphorylation by a selective 2-kinase to produce the IP_6 end-product[339]. A key regulatory point and rate-limiting step in the biosynthesis of higher inositol polyphosphates in mammalian cells is the phosphorylation of the intermediate Ins(1,3,4)P_3 to generate Ins (1,3,4,6)P_4 which represents the entry point for the synthesis of all other tetrakis and higher inositol polyphosphates in animals[341,350]. Identified in 1987 was the founding 6-kinase member of the group of enzymes responsible for catalysis of this pivotal step - the inositol 1,3,4- triphosphate 5/6-kinase family (ITPK), which, structurally, form a subgroup of a larger set of ATP-grasp fold proteins[339,348,351]. This family was later reclassified as inositol 3,4,5,6-tetrakisphosphate 1-kinase/inositol 1,3,4,5,6-pentakisphosphate 1-phosphatase based on the emergence of new experimental data[352]. ITPK1, one of the 4-6 isoforms expressed in higher plants[1], is a fundamental member of the ITPK family and is the focus of this chapter.

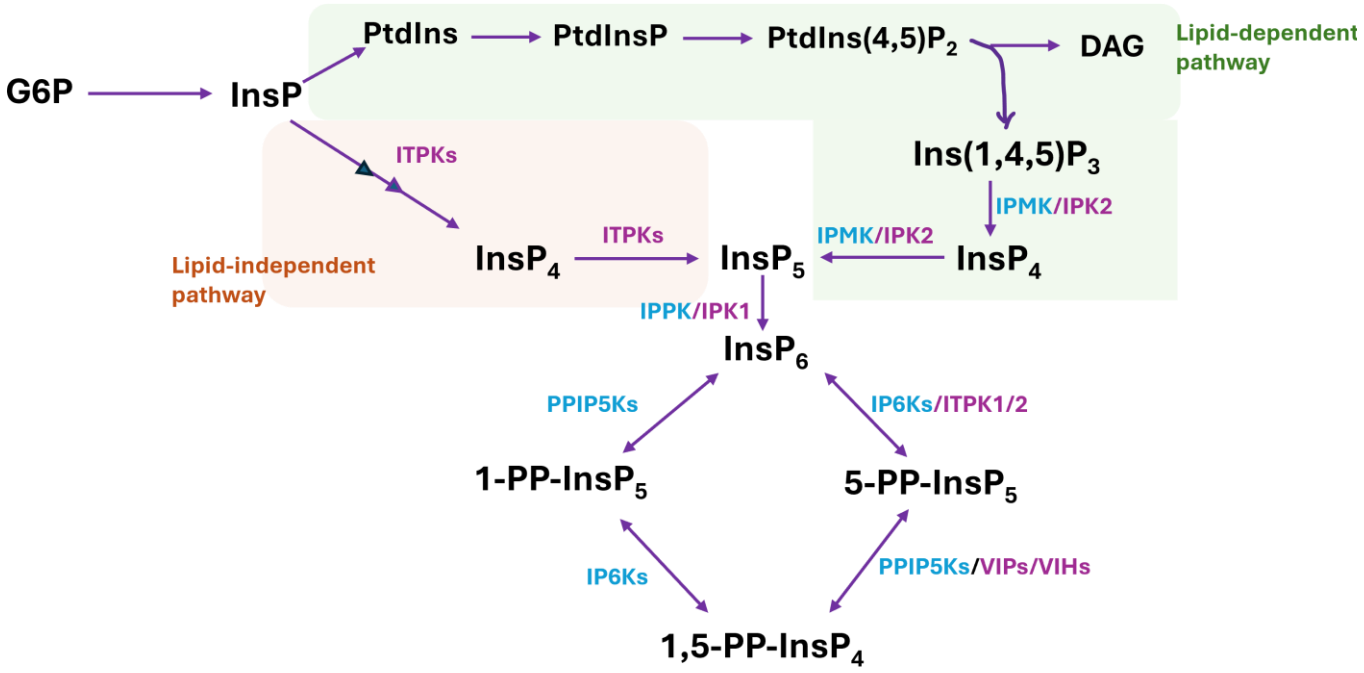


Figure 5.1 The currently established biosynthetic pathway to higher inositol polyphosphates in plants, including pyrophosphates. Schematic adapted from Molecules 2020[347].

Human ITPK1 (*h*ITPK1) has been well characterised yet, until recently, comparatively little had been known of the plant counterparts[340]. Indeed, the recently reported *Zea mays* ITPK1 structure was the first structure of an ITPK1 from plants[339]. To date, there are structures of ITPK1 from *Entamoeba histolytica* (2005; PDB 1Z2P)[341], *Homo sapiens* (2007; PDB 2OTD)[343], *Zea mays* (2022; PDB 7TN5)[339] and *Solanum tuberosum* (2023; PDB 8OXE)[353].

The structure of the *E. histolytica* ITPK1 (*Eh*ITPK1) revealed the multifunctionality of the enzyme, which, in addition to generating both Ins(1,3,4,5) and Ins(1,3,4,6) tetrakisphosphates from Ins(1,3,4)P_3, can: 1) perform the reverse reaction i.e. dephosphorylation of these same two inositol tetrakisphosphates to Ins(1,3,4)P_3 2) dephosphorylate Ins(1,3,4,5,6)P_5 to Ins(3,4,5,6)P_4 3) interconvert the IP_4 isomers Ins(1,3,4,5)P4 and Ins(1,3,4,6)P_4 4) act as a 1-kinase, phosphorylating Ins(3,4,5,6)P_4 to Ins(1,3,4,5,6)P_5[341]. This is an example of the promiscuity with which a number of kinases phosphorylate multiple substrates[341,343]. Structurally, *Eh*ITPK1 boasts a highly versatile catalytic cleft lacking in stereospecificity constraints, in which it was possible to model up to eighteen inositol phosphate substrates into the active site[341]. It was postulated that many of these characteristics also applied to human ITPK1, which is topologically the same as *Eh*ITPK1[343]. However, *h*ITPK1 was indicated to possess a much greater ability to discriminate between substrates, in part due to its much more refined binding pocket[343]. Interestingly, it has been shown that in higher organisms specifically, one of the alternate substrates of ITPK1, Ins(1,3,4)P3, generated from Ins(1,4,5)P3[354], actually stimulates dephosphorylation of Ins(1,3,4,5,6)P5[355], thereby promoting increased cellular concentrations of Ins(3,4,5,6)P4[354] rather than acting as a competitive inhibitor[343]. Ultimately, the key findings from studies of human ITPK1 was that this enzyme performs ADP-dependent inter-substrate phosphate transfer and that this serves to rationalise 1) the ability of competing substrates to stimulate one another's catalysis by ITPK1; 2) the interactions between these two substrates (Ins(3,4,5,6)P4 and Ins(1,3,4)P3) given that they exist in discrete branches of the metabolic pathway[343]. The fact that this phenomenon was at the time restricted to the human ITPK1 was thought to highlight the distinction, from an evolutionary perspective, between animal IP kinases and their plant and protozoan orthologs[343]. However, it was since discovered that the ITPK1 from *Solanum tuberosum* (*St*ITPK1) also retains this ability to catalyse reciprocal reactions and as such, inter-substrate transfer of phosphate groups between inositol phosphates is now presumed to be an evolutionary conserved function among the ITPK1 enzymes[340].

Although the ITPK1 from *Zea mays* (*Zm*ITPK1) yielded the first structure of a plant ITPK1, the catalytic activities of *St*ITPK1 had already been investigated, and it demonstrated mostly the same functions as its human and protozoan orthologs[340]. The predominant activity of *St*ITPK1 at

physiological ADP levels is as an $InsP_5$-ADP phosphotransferase, transferring a phosphate from Ins(1,3,4,5,6)P_5 to either Ins(1,3,4)P_3 or ADP[340]. Additional functions are as a phosphomutase acting on Ins(1,3,4,5) and Ins(1,3,4,6)P_4, like the protozoan ITPK1 with which it shares 21% sequence identity, and as a kinase acting on the following substrates: Ins(3,4)P_2; Ins (1,3,4)P_3; Ins(1,4,6)P_3; Ins(3,4,5)P_3; Ins(3,4,6)P_3; Ins(3,4,5,6)P_4[340]. The researchers sought to determine the relative affinities of *St*ITPK1 for the substrates Ins(1,3,4)P_3 and Ins(3,4,5,6)P_4, concluding that the latter is the preferred substrate[340]. Indeed, the order of substrate preference was established as follows: Ins(3,4,5,6)P_4 ~ Ins(1,4,6)P_3 ~ Ins(3,4,6)P_3 >> Ins(1,3,4)P_3 > Ins(3,4,5)P_3[340]. At the time of publication the authors of this work postulated that the phosphotransfer to Ins(1,3,4)P_3 was not a physiologically relevant function in plants, given that the mammalian 3-kinases which generate the precursor of Ins(1,3,4)P_3 were thought not to be present in plants and yeast[340].

Recent studies of plant ITPK enzymes has led to the knowledge that they not only phosphorylate lower inositol polyphosphates but also contribute to inositol pyrophosphate synthesis, generating 5-$InsP_7$ – a precursor to the plant immunity-enhancer 1,5-$InsP_8$ - by the addition of a β-phosphate to IP6[339,345]. Nevertheless, until recently, the physiological roles of plant inositol pyrophosphates and the catalytic diversity of the family of enzymes responsible for their synthesis had remained somewhat elusive due to the lack of structural information[345,356] - until the first plant ITPK1 structure (*Zm*ITPK1) was published in 2022[339]. Despite the fact that the researchers were unable to obtain a crystal complex containing either nucleotide or substrate in addition to wild-type *Zm*ITPK1, they were able to soak IP6 into crystals of a 'hinge mutant', *Zm*ITPK1^{H192A}. The resultant structure highlights the greater overall structural similarity to the human ITPK1 in contrast to the protozoan ITPK1, however there are still considerable variations[339]. A significant identification, through additional studies with *Glycine max* ITPK1, was that of a catalytic specificity element which governs the variation in relative rates of IP6 kinase and Ins(1,3,4,5,6) phosphatase activities among individual members of the family[339]. Comparison of IP6 kinase activities between the human, maize and soybean ITPK1 enzymes led the authors to conclude that this is a specialised function of the plant ITPK1 family[339]. This is in disagreement with the concept of Laha et al that the IP_6 kinase activities of ITPK enzymes are evolutionary conserved among higher eukaryotes including humans[342]. Nevertheless, it was the subsequent work by Laha et al on Arabidopsis ITPK1 and ITPK2 that led to the conclusion that kinase activity by these enzymes form the symmetric IP_7 isomer, 5-IP_7[345].

Hitherto, the lack of bound nucleotide and substrate in crystal structures of plant ITPK enzymes has obscured the mechanism of phosphotransfer, and offered a clear rationale for attempts to

solve the structure of a complex of a plant ITPK1 with nucleotide and inositol phosphate. Indeed, whilst the structure of *At*ITPK4 (PDB: 7PUP) boasts bound nucleotide, it lacks bound substrate[355], and contrariwise, the structure of *Zm*ITPK1 (PDB: 7TN8) benefits from bound substrate but lacks the nucleotide[339]. Previous efforts with the ITPK1 from potato, *St*ITPK1, proved unfruitful* yet paved the way for subsequent studies which are the focus of this chapter. Those initial experiments revealed the necessity of truncating a portion of the C-terminus of *St*ITPK1 to facilitate crystallisation[257]. Crystallisation experiments with the truncated *St*ITPK1 plus ADP or ATP yielded several diffraction quality crystals, however, areas of poor electron density due to suspected twinning rendered the structure solution a considerable challenge[357].
**see discussion*

It was therefore the purpose of the work presented in this chapter to elaborate on the experiments completed by earlier researchers (Li[257] and Rodriguez[357]) in order to obtain a trapped substrate (IP_6 and/or IP_4) complex of *St*ITPK1 co-crystallised with a non-hydrolysable ATP analogue. This would provide novel insights into the orientation and nuances of substrate binding by a plant ITPK1 which also has cofactor bound. *In silico* docking studies were planned to be used in the potential case of inadequate X-ray diffraction data quality. Ultimately, delineating the interactions which stabilise the substrate-bound complex will offer important insights into the structural basis for phosphate signalling by an ITPK1 in plants.

5.2 Experimental Procedures

5.2.1 Expression and purification of *St*ITPK1

The *Solanum tuberosum* ITPK1 gene (*St*ITPK1; previously also termed *St*IPKα) had previously been cloned into pOPINF as a C-terminal truncated construct so this plasmid was therefore used for expression and purification. The T7 promoter-driven pOPINF vector is based on the backbone of pTriEx2 (Novagen®) and carries an ampicillin resistance gene cassette[358]. The gene is flanked at the N-terminus by a hexahisidine-tag followed by a 3C cleavage site for its removal (Figure 5.2).

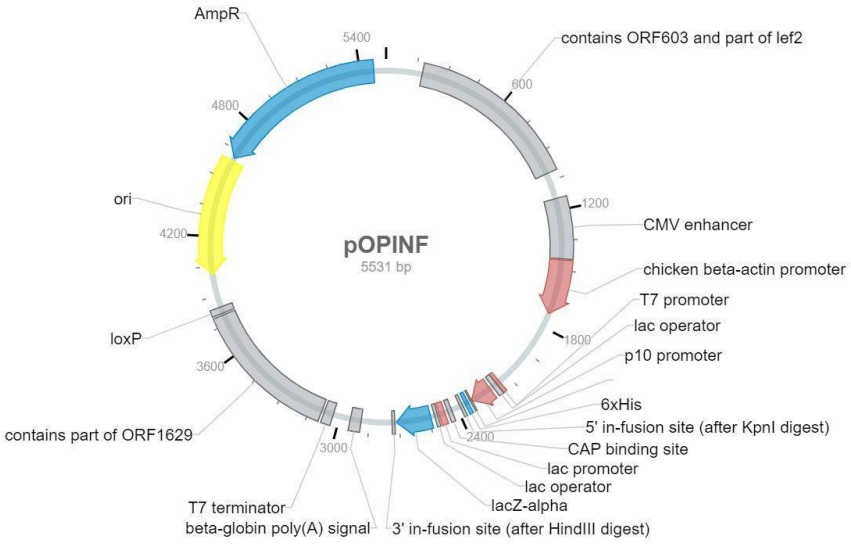


Figure 5.2 Plasmid map of the pOPINF vector used for recombinant expression of *St*ITPK1.

Recombinant *St*ITPK1 was expressed on a large scale (XL) from *E. coli* Rosetta 2 (DE3) cells in LB media. A single colony from a freshly glycerol stock-streaked LB agar plate containing carbenicillin (100 µg/mL), which had been incubated overnight at 37 °C, was used to inoculate a pre-culture of carbenicillin-supplemented LB (50 mL) for incubation overnight at 37 °C with shaking (180 rpm). X x 1L LB supplemented with carbenicillin (100 µg/mL) was inoculated with an aliquot of the starter culture (5 mL) and shaken with incubation at 37 °C, 180 rpm until an OD_{600} of approximately 0.6. At this point IPTG was added to a final induction concentration of 0.25 mM and the cultures were left to express overnight at 25 °C with shaking (180 rpm). Cells were harvested by centrifugation (5500 xg, 20 min, 4 °C) before resuspension in a lysis buffer consisting of NaH_2PO_4 (50 mM) pH 7.5, NaCl (300 mM), imidazole (20 mM) and Triton X-100 (0.5%) with added protease inhibitor tablet (cOmplete™ Mini EDTA-free protease

inhibitor cocktail, Roche) and DNase I (~20 μg/mL). The resuspended cells were then subjected to mechanical lysis at 18,000 psi by a microfluidizer (LM20, Analytik). The clarified supernatant was obtained by centrifugation of the lysate (48,000 xg, 30 min, 4 °C) and was then filtered through a membrane filter (either PES or cellulose acetate; 0.45 μm, Minisart® Sartorius). The sample was applied to a Ni-sepharose 6 Fast Flow column (HisTrap FF 5mL, Cytiva) after which the resin was washed with a buffer composed of NaH_2PO_4 (50 mM) pH 7.5, NaCl (300 mM), and imidazole (20 mM). Recombinant *St*ITPK1 was eluted with a linear gradient of increasing percentage of an analogous buffer containing an elevated imidazole concentration of 500 mM and excluding the Triton X-100. The fractions corresponding to peaks on the UV trace were assessed by SDS-PAGE and those identified as containing *St*ITPK1 were pooled. The sample pool was incubated with 3C protease (3 mg) for 68h at 4 °C during simultaneous dialysis against the lower imidazole concentration IMAC wash buffer. After verification of cleavage by SDS-PAGE, reverse IMAC was performed to separate the cleaved His-tag from *St*ITPK1. Identification of *St*ITPK1-containing fractions was facilitated by SDS-PAGE and the flow-through was therefore pooled. The sample was then desalted to remove imidazole, into a buffer of Tris-HCl pH 7.5 (10 mM), NaCl (150 mM) and DTT (2 mM). The sample was concentrated down to 9 mL at 5.8 mg/mL. Due to precipitation upon concentrating beyond this, SEC was performed over multiple runs applying 2 mL sample each time, at a flow rate of 0.5 mL/min. The final yield after combining the *St*ITPK1-containing SEC peaks was 14.7 mg/mL

5.2.2 Crystallisation screens

Crystallisation screens were set up in the sitting drop vapour diffusion format by an OryxNano crystallisation robot (Douglas Instruments Ltd). A number of commercial screens (Molecular Dimensions) were set up (Structure Screen 1 and 2, PACT Premier, MIDASplus, JCSG Plus, LMB) in addition to a screen around previously identified conditions. This screen consisted of 0.2M $CaCl_2$, 0.1M MES with pH ranging from 5.4 – 6.4 in increments of 0.2 and PEG6000 ranging from 16-22% in increments of 2%. $MgCl_2$ (5 mM) was added to all screens, both commercial and custom. Each screen was set up with 1) AMP-PNP; 2) β-γ-imido ATP; 3) AMP-PCP at concentrations of 1 mM. IP6 was added at a 10-fold molar excess to one of each pair of drops such that each condition was prepared both with (1:1:2 protein: screen: IP6 5mM) and without (1:1 protein: screen) the inclusion of IP6. Two co-crystals of *St*ITPK1 with AMP-PNP grown in 0.2M $CaCl_2$, 0.1M MES pH 5.6/5.4, 22/18% PEG6000 and 5mM MgCl2 were harvested after 6 days at 16 °C and subsequently soaked in an IP6 solution before cryoprotection. A refined screen using only the β-γ-imido ATP analogue was subsequently prepared based on the previous screen and this consisted of 0.2M $CaCl_2$, 0.1M MES with a pH ranging from 5.4 to 6.5

differing by increments of 0.1 and either PEG6000 or PEG3350 over a range of 18-25% in increments of 1%. $MgCl_2$ (5 mM) was included throughout. 9 crystals were harvested in total, after 9 weeks at 16 °C, however 4 dissolved upon cryoprotection. Details of the substrate soak and cryoprotection of the remaining 5 are tabulated below (Table 5.1).

Condition crystal harvested from	Substrate soak solution	Cryoprotectant solution
0.2M $CaCl_2$, 0.1M MES pH 6.4, 24% PEG6000 and 5 mM $MgCl_2$	N/A	Screen solution + 25% ethylene glycol
0.2M CaCl2, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM MgCl2	Screen solution + IP4 (1 mM)	Screen solution + 10% glycerol
0.2M CaCl2, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM MgCl2	Screen solution + IP6 (1 mM)	Screen solution + 10% glycerol
0.2M CaCl2, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM MgCl2	40% PEG3350 + 1 mM IP4	
0.2M CaCl2, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM MgCl2	40% PEG3350 + 1 mM IP4	

Table 5.1 Composition of the crystallisation conditions and corresponding substrate soak and cryoprotectant solutions used.

5.2.3 X-ray diffraction data collection and processing

X-ray diffraction data was collected on beamline i24 of the Diamond Light Source (Oxfordshire, UK). For the first 2 crystals, X-ray diffraction was executed with a beam size of 30 x 30 μm, wavelength of 0.9795 Å, transmission of 5.6% and an exposure time of 0.01s. 3600 images were collected over oscillation increments 0.10°. For the latter 5 crystals, beam parameters were varied as necessary, such that beam size was 7 x 7μm, wavelength was 0.6199 Å, transmission was either 25.43%, 49.89% or 100%, exposure time varied between 0.01s and 0.05s. 3600 images were collected in oscillation increments of 0.10°. The only 2 datasets which were potentially adequate for downstream use were automatically processed by xia2 dials. Xtriage (Phenix) and Phaser (Phenix) were run for crystal quality analysis and molecular replacement, respectively.

Each of these parameters or factors is then carefully incremented in additional trial matrices encompassing a range spanning the condition which gave the 'hit'.

5.3 Results

5.3.1 Expression and purification of *St*ITPK1

Purification of recombinant *St*ITPK1 consistently yielded ample protein of at least 10 mg/mL. A representative SDS-PAGE of the purification fractions is displayed in Figure 5.3.

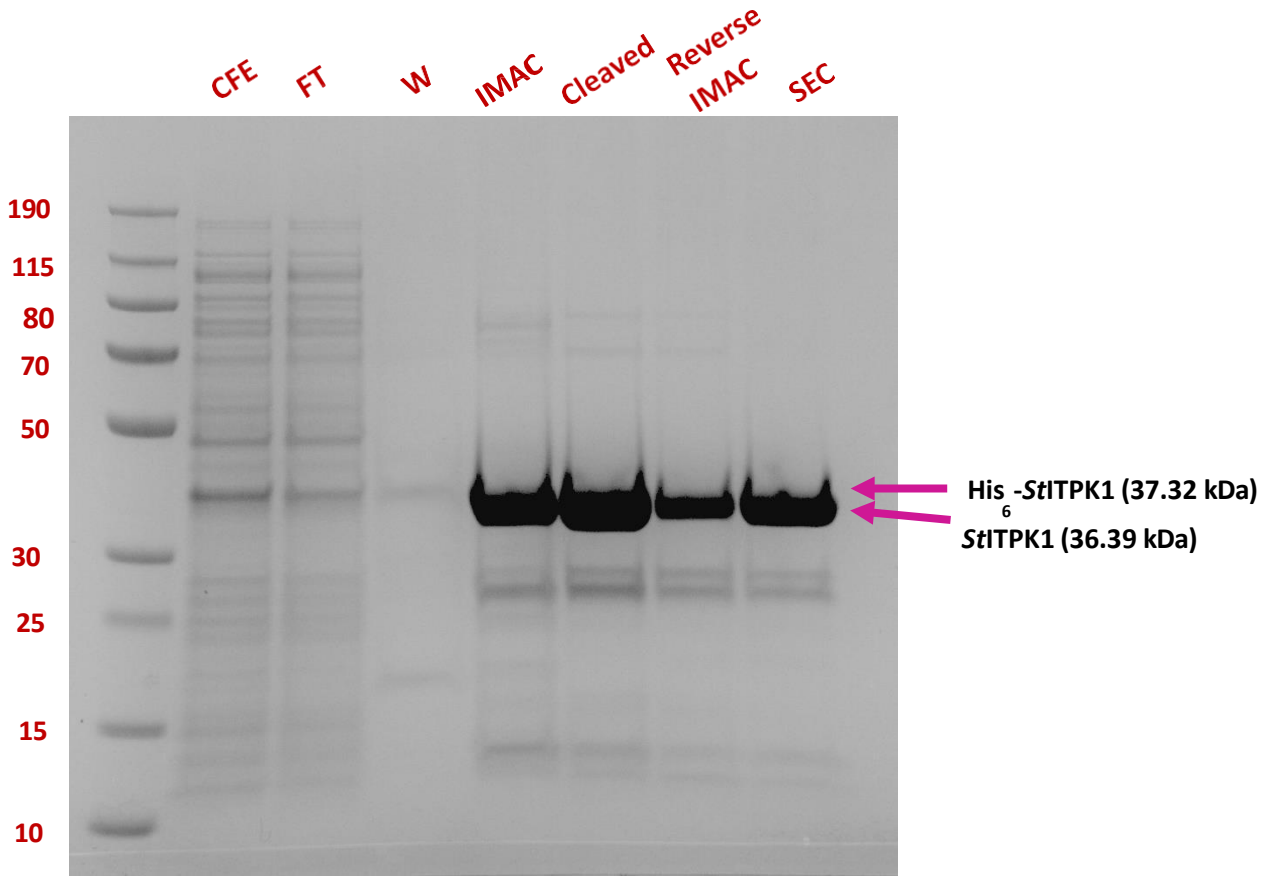


Figure 5.3 4-12% SDS-PAGE representing the purification workflow of *St*ITPK1

5.3.2 Co-crystallisation trials of *St*ITPK1

A multitude of sitting drop crystallisation screens were prepared incorporating 3 different ATP analogues: AMP-PNP, AMP-PCP and β-γ-imido ATP. In addition to several commercial screens, a trial matrix encompassing a range spanning the conditions which yielded the best crystals in previous work by Rodriguez (0.2M $CaCl_2$, 0.1 MES pH 5.8/6.2, 20% (w/v) PEG6000) were optimised based on pH and PEG molecular weight and concentration. This resulted in a trial matrix comprising 0.2M $CaCl_2$, 0.1M MES with pH ranging from 5.4 – 6.4 in increments of 0.2 and PEG6000 ranging from 16-22% in increments of 2%. $MgCl_2$ (5 mM) was included

throughout all screens. Of the total number of plates set up (26), crystals grew in a large number of conditions and morphologies, however, due to their very small size and inherent fragility, only 2 crystals were able to be harvested and sent for X-ray diffraction data collection (Figure 5.4), and these had not been co-crystallised with substrate. However, they had been co-crystallised with AMP-PNP and subsequently soaked with IP6. The relatively large square bipyramidal crystal in Figure 5.4a emerged to be a salt crystal, whilst for the protein crystal (Figure 5.4b), the autoprocessing of the resultant diffraction data either failed or had multiple warnings such as presence of ice rings, higher crystallographic symmetry and/or non-crystallographic symmetry.

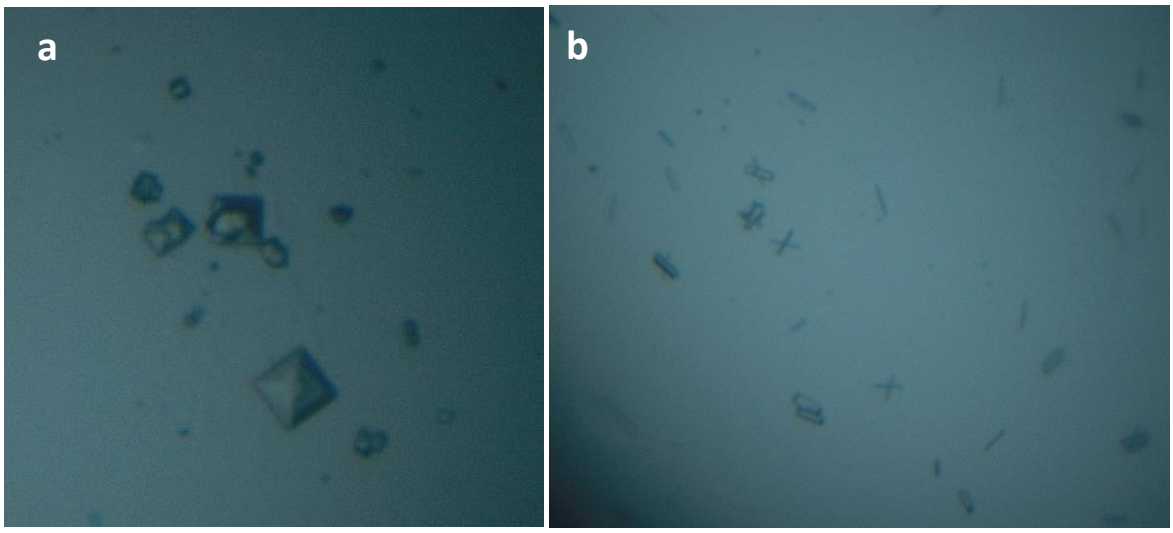


Figure 5.4. Crystals of *St*ITPK1 co-crystallised with AMP-PNP which formed in a) 0.2M $CaCl_2$, 0.1M MES pH 5.6, 22% PEG6000 and 5mM $MgCl_2$; b) 0.2M $CaCl_2$, 0.1M MES pH 5.4, 18% PEG6000 and 5mM $MgCl_2$.

Based on the observations of sizes and morphologies of crystals in the screens, conditions were further refined with a pH range of 5.4-6.5 differing by increments of 0.1 and either PEG6000 or PEG3350 in a range of 18-25% in increments of 1%. $MgCl_2$ (5 mM) was also included throughout all crystallisation plates. However, the protein used in this batch of screens was at around one third of the concentration than that used in the previous set of screens (~4 mg/mL versus ~14 mg/mL). Nevertheless, there were some interesting crystal formations observed, albeit only in screens with the β-γ-imido ATP analogue. No crystals of potential diffraction quality grew in the analogous plates containing AMP-PNP, and the AMP-PCP analogue was not used due to resource limitations. For example, there were numerous long 'branched' structures, in which portions of the structure appear to sprout/peel off from the core body, some of which

could only be described as looking akin to a piece of 'farfalle' pasta. Some examples of these are displayed in the following figures, with Figure 5.5 comprising examples of the 'better' (more 'glassy' and with more straight edges) type and Figure 5.6 representing the extensively branched 'farfalle' form, the latter being deemed unsuitable for X-ray diffraction due to the aggregation and lack of straight edges.

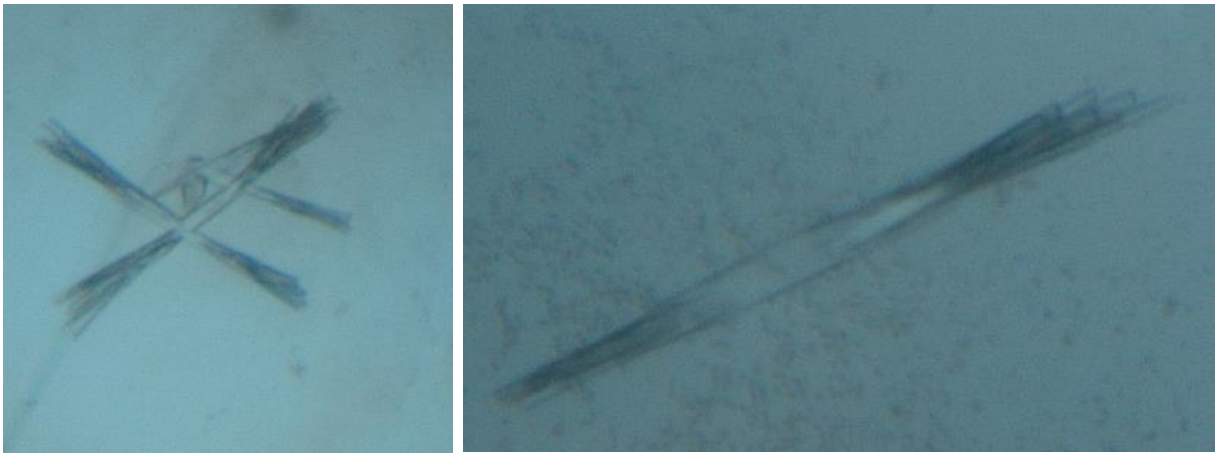


Figure 5.5. Crystals of *St*ITPK1 co-crystallised with IP6 and β-γ-imido ATP, formed in the following conditions: Left= 0.2M $CaCl_2$, 0.1M MES pH 5.4, 18% PEG3350 and 5 mM $MgCl_2$); Right = 0.2M $CaCl_2$, 0.1M MES pH 6.5, 25% PEG3350 and 5 mM $MgCl_2$.

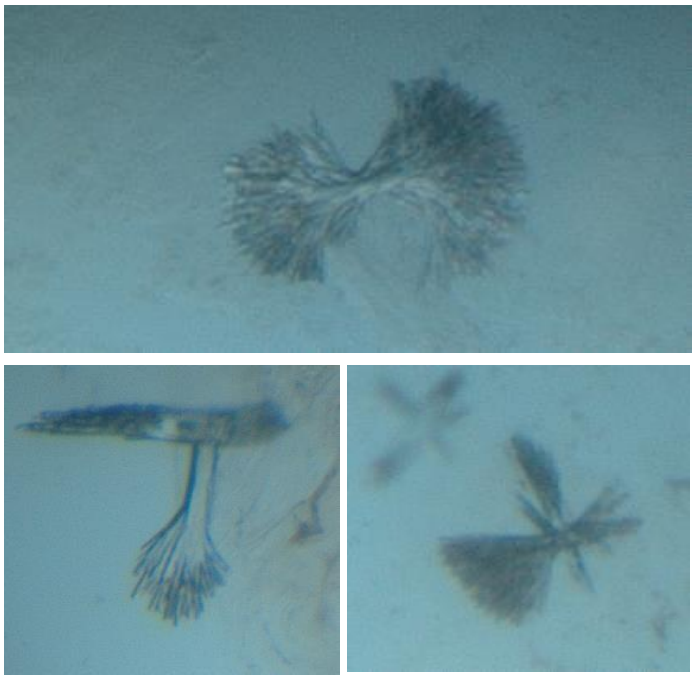


Figure 5.6 Crystals of *St*ITPK1 co-crystallised with IP6 and β-γ-imido ATP, formed in the following conditions: Top = 0.2M $CaCl_2$, 0.1M MES pH 6.5, 23% PEG3350 and 5 mM $MgCl_2$; Lower left = 0.2M $CaCl_2$, 0.1M MES pH 6.0, 25% PEG6000 and 5 mM $MgCl_2$; Lower right = 0.2M $CaCl_2$, 0.1M MES pH 5.5, 25% PEG3350 and 5mM $MgCl_2$.

Another predominant crystal morphology which occurred in this screen was long thin needles as in Figure 5.7, with some conditions resulting in a ring of these needles around the entire inside edge of the drop, as in Figure 5.8. This is a common occurrence in vapour diffusion experiments since the rate of evaporation is greater around the perimeter which in turn results in faster supersaturation. These long needle-like crystals ended up being the morphology of the majority which were harvested.

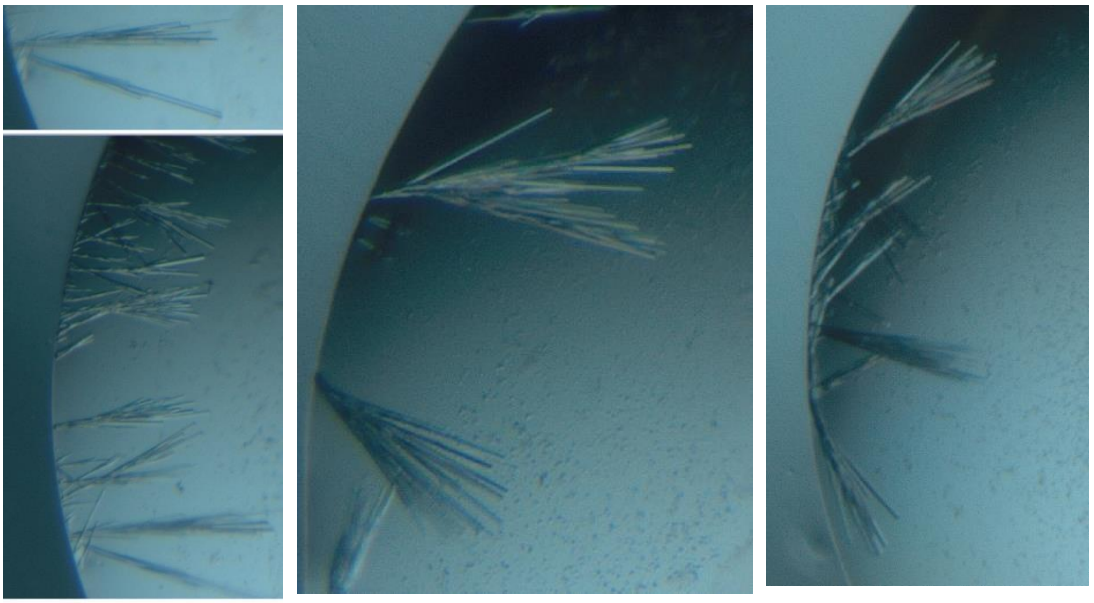


Figure 5.7 Crystals of *St*ITPK1 co-crystallised with IP6 and β-γ-imido ATP, formed in the following conditions: Left (upper and lower) = 0.2M $CaCl_2$, 0.1M MES pH 6.4, 24% PEG6000 and 5 mM $MgCl_2$; Centre = 0.2M $CaCl_2$, 0.1M MES pH 5.5, 25% PEG6000 and 5 mM $MgCl_2$; Right = 0.2M $CaCl_2$, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM $MgCl_2$.

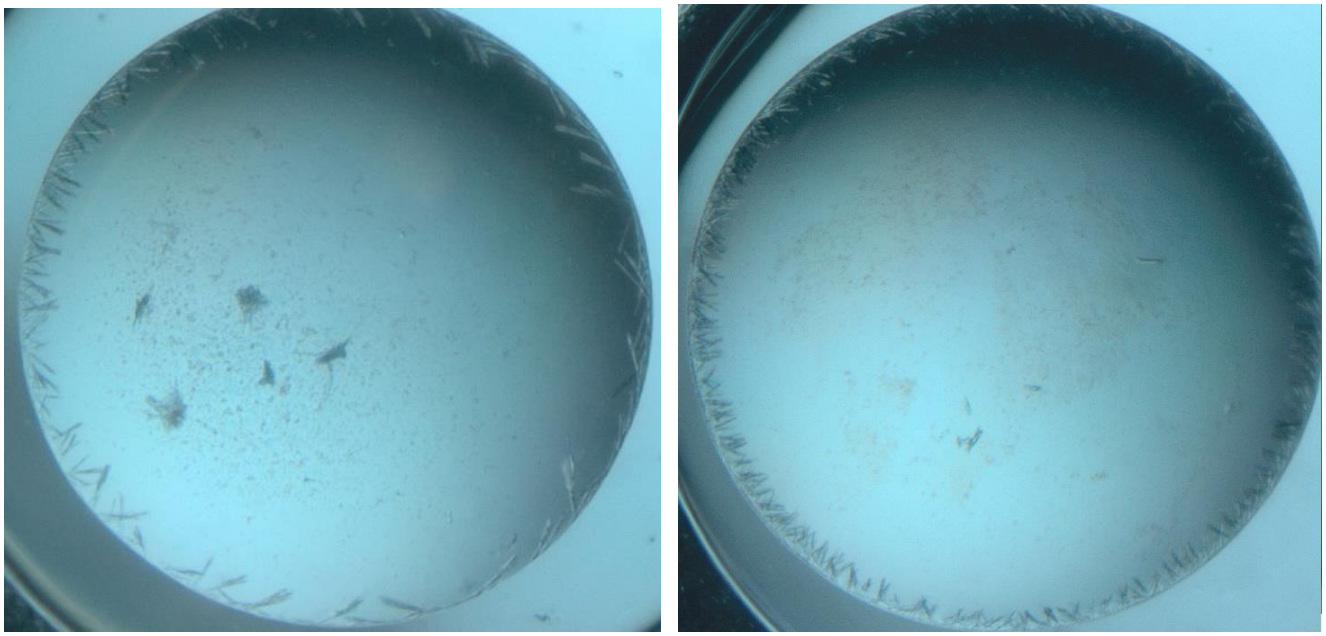


Figure 5.8 Crystals of *St*ITPK1 co-crystallised with IP6 and β-γ-imido ATP, formed in the following conditions: Left = 0.2M $CaCl_2$, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM $MgCl_2$; Right = 0.2M $CaCl_2$, 0.1M MES pH 5.4, 18% PEG3350 and 5 mM $MgCl_2$.

Initially, the crystals identified as most suitable for X-ray diffraction experiments were those as in Figure 5.9, provided the daughter crystals were broken off during harvesting such that individual needle-like fragments could be isolated and subjected to X-ray diffraction (an example circled in Figure 5.9). 9 crystals, all crystallised with β-γ-imido ATP, were harvested and immediately soaked in a solution of cryoprotectant plus either IP6 or IP4 but unfortunately 4 crystals dissolved upon addition. Nevertheless, 5 crystals were successfully soaked, however they did not prove very amenable to X-ray diffraction, being so particularly small (< 100 μm) and fragile, and as such, X-ray diffraction data was only able to be obtained from 2 of the needle crystals, shown in Figure 5.10.

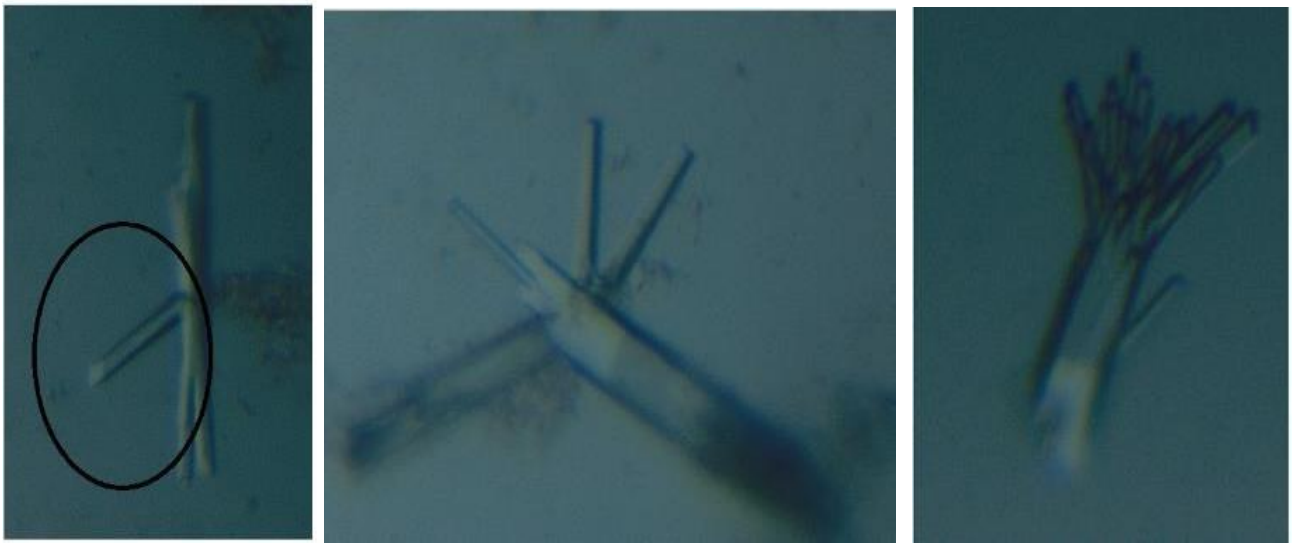


Figure 5.9 Crystals of *St*ITPK1 co-crystallised with β-γ-imido ATP, formed in 0.2M $CaCl_2$, 0.1M MES pH 6.3, 25% PEG6000 and 5 mM $MgCl_2$(H10 bottom).

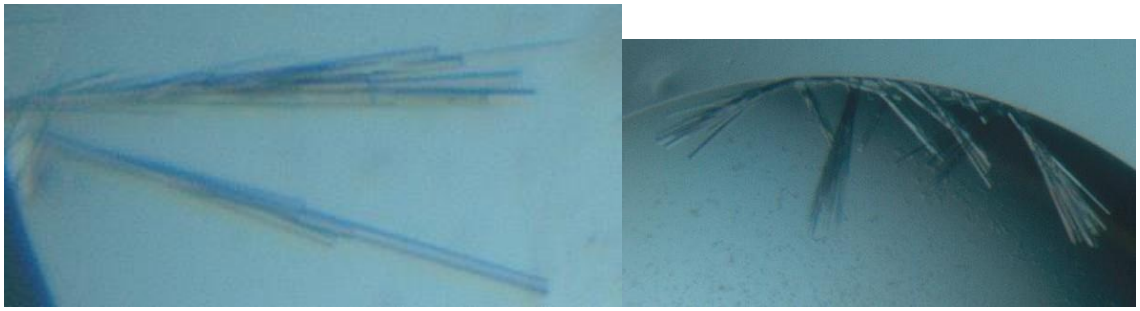


Figure 5.10 Co-crystal complexes of *St*ITPK1 with β-γ-imido ATP. Top: Crystal was cryoprotected in a solution of mother liquor (0.2M $CaCl_2$, 0.1M MES pH 6.4, 24% PEG6000 and 5 mM $MgCl_2$) and 25% ethylene glycol without prior soaking with substrate. Bottom: Crystal was cryoprotected in a solution of mother liquor (0.2M $CaCl_2$, 0.1M MES pH 6.4, 25% PEG3350 and 5 mM $MgCl_2$) and 10% glycerol after a soak in IP4.

5.3.3 Structural analysis of *St*ITPK1

Despite multiple attempts and with parameters such as exposure time and transmission percentage varied, only one dataset obtained from one of these two crystals was potentially suitable for molecular replacement (MR). This was auto processed by xia2 dials in space group C 2 2 2, with cell parameters of 85.02 (90.00), 149.53 (90.00), 56.54 (90.00). An image of this crystal mounted in the loop is represented in Figure 5.11, along with the corresponding X-ray diffraction pattern and reflections. Whilst several aspects of the dataset are reasonable – such as 99.9% completeness and a resolution down to 2.80Å – there is a suspicion of twinning, rendering structure solution a challenging feat. This is indicated by the intensity statistics and by the twin operators determined by Xtriage. Attempts to perform molecular replacement using Phaser, with 8OXE as the search model, were to no avail, since the top MR solution had dubious log-likelihood gain (LLG) and translation function Z (TFZ) scores of -20.2 and 4.2, respectively. Given that the rule of thumb for an incorrect MR solution is represented by TFZ scores of below 5, and LLG values of below 25, it was clear that a satisfactory solution was not obtained. As such, structure solution was not possible using this dataset.

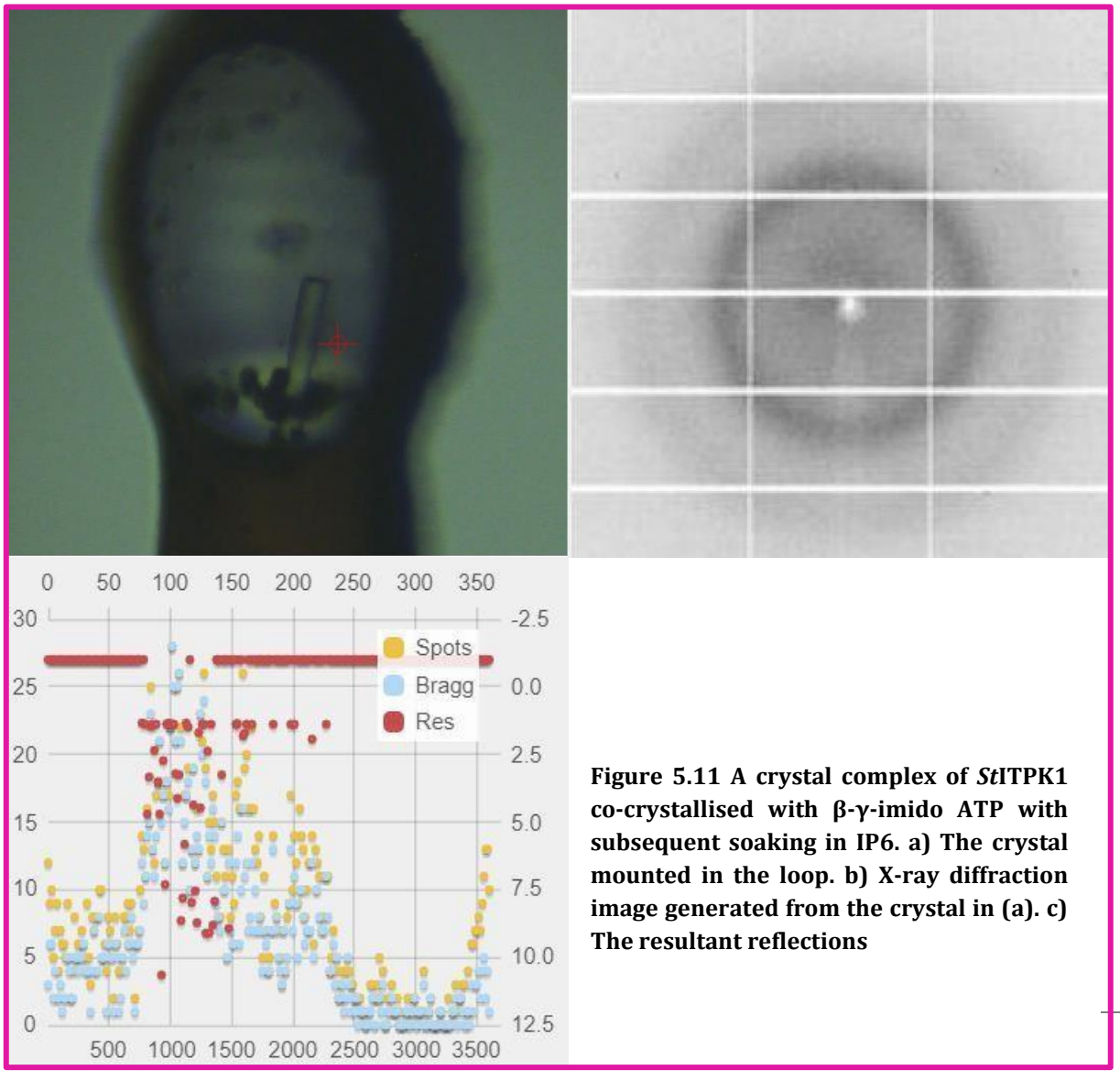


Figure 5.11 A crystal complex of *St*ITPK1 co-crystallised with β-γ-imido ATP with subsequent soaking in IP6. a) The crystal mounted in the loop. b) X-ray diffraction image generated from the crystal in (a). c) The resultant reflections

5.3.4 Comparison of the IP6 binding sites in plant ITPK1 enzymes: *St*ITPK1 and *Zm*ITPK1

Since completion of the *in vitro* aspect of this project, the structure of *St*ITPK1 in complex with ATP solved at 2.26 Å resolution was reported in Biochemistry at the end of 2023 by a related research group[353]. It was therein established that *St*ITPK1 assumes the characteristic ATP-grasp kinase fold, or α-β-α fold, comprised of three conserved lobes – N-terminal, central and C-terminal, in accordance with the nomenclature assigned to previously reported ITPK1 orthologs, *Eh*ITPK1 and *Hs*ITPK1. The structure solution of *St*ITPK1 revealed a flexible segment, termed a 'tether', which links the central and C-terminal domains by bridging across the top of the active site cavity and provides cofactor/substrate binding pocket-contributing residues. This is the region which is considerably different between the potato and maize ITPK1* enzymes, since the binding of nucleotide, as in the structure of *St*ITPK1, induces a shift which facilitates stabilization of this region by interactions with a neighbouring copy of the protein in the crystal lattice, whilst in the maize enzyme this region (between β11 and α8, 231-256) is disordered due to lack of bound nucleotide. Meanwhile, the soaking in of IP6 in the maize ITPK1 stabilises the region immediately before the disordered polypeptide (residues 216-230), forming a four residue β- strand (β11) which is absent in both of the substrate-lacking *St*ITPK1 and *Zm*ITPK1 structures. The nucleotide-binding ATP-grasp region is formed from two sets of anti-parallel β-sheets distributed across the C-terminal and central subdomains. Whilst in the human and protozoan orthologs these β-sheets are built from four β-strands each in the C-terminal lobe and central lobe, in comparison the central lobe of *Zm*ITPK1 lacks one β-strand. Interestingly, relatively to *Zm*ITPK1, the structure of *St*ITPK1 displays an extra β-strand in the central lobe, similar to *Hs*ITPK1 and *Eh*ITPK1. Whilst the β8-β9 loop of *Zm*ITPK1, representing a 'hinge' between the central and C-terminal subdomains, is proposed to govern the productive orientation of the nucleotide, it is speculated that the tether insertion of *St*ITPK1 may influence substrate recognition,

Upon superimposing the structures of ATP-bound *St*ITPK1 (PDB: 8OXE) and IP6-bound *Zm*ITPK1^{H192A} (PDB: 7TN8), it is clear that the N-terminal subdomain is very well aligned between the two structures, (RMSD = 0.577Å, 73 atoms), whilst the central and C-terminal subdomains are much less so, since they include the flexible 'tether' polypeptide linkage which is either stabilised or disordered in the *St*ITPK1 and *Zm*ITPK1^{H192A} structures, respectively (Figure 5.12). The two structures have an overall RMSD of 5.0Å (1951 atoms of 1989). For the apo-*Zm*ITPK1, this is decreased to 2.8AÅ (1898 atoms of 1945). Nevertheless, predictably, the two plant ITPK apo-enzymes are much more structurally similar to one another than either is to the human or protozoan orthologs: with regard to the human ITPK1 (PDB: 2QB5; complex

with Mn^{2+} and ADP), the potato and maize ITPK1 enzymes share an RMSD of 4.5Å and 4.4Å, respectively whilst with the protozoan ITPK1 (PDB: 1Z2N; complex with Mg^{2+} and ADP), they share an RMSD of 7.3Å and 9.3Å. This is not unexpected given that Zong et al reported that the overall topology of *Zm*ITPK1 is more similar to the human ITPK1 than the protozoan equivalent[339]. **Referring specifically to the ZmITPK1^{H192A} 'hinge mutant' in complex with IP6 **Note that RMSD was calculated in Pymol as an all-atom RMSD which does not do outlier rejection.*

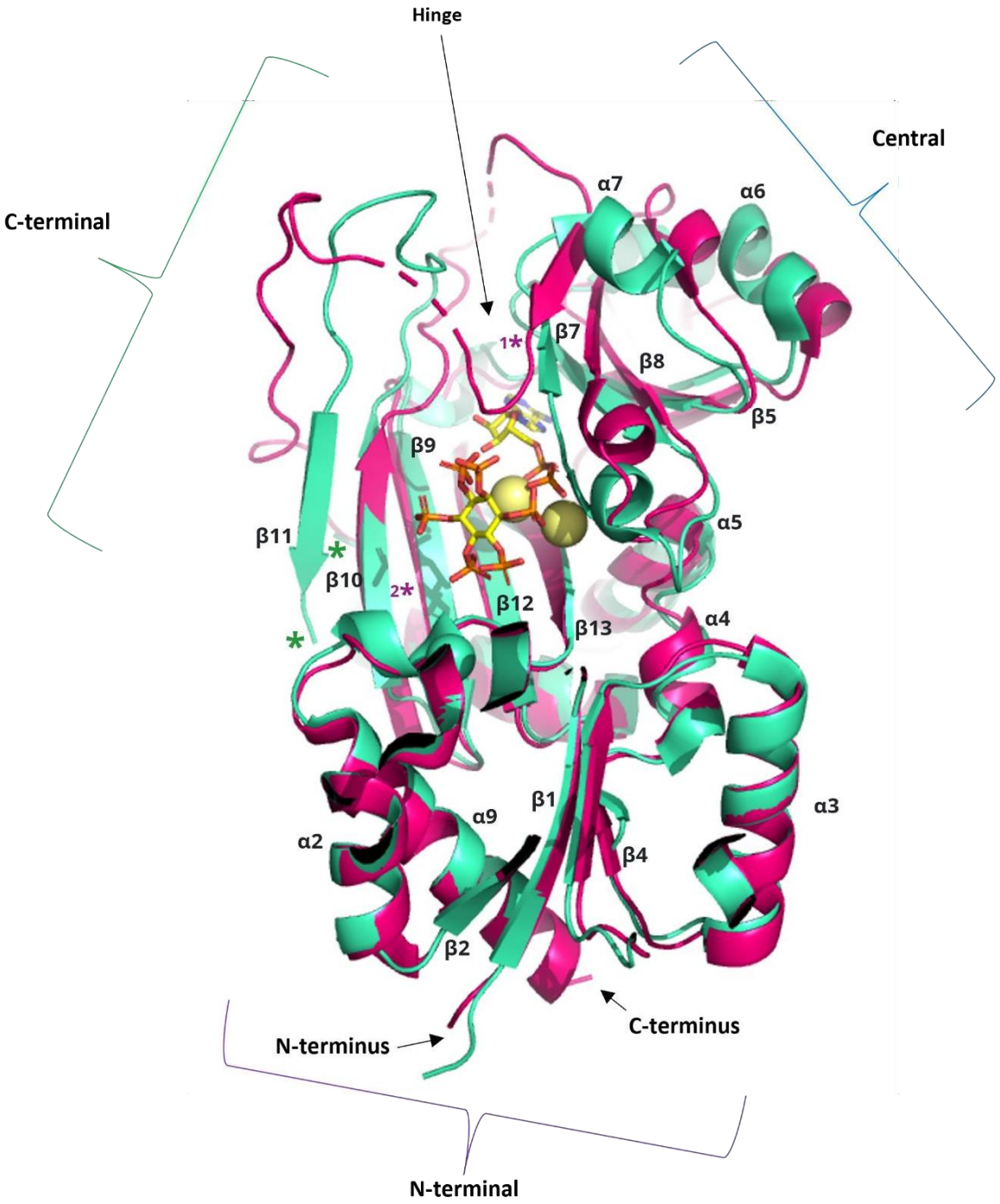


Figure 5.12 Superimposition of ATP-bound *St*ITPK1 (PDB: 8OXE; magenta) and IP6-bound *Zm*ITPK1^{H192A} (PDB: 7TN8; teal). The N-terminal, central and C-terminal lobes are loosely indicated by coloured brackets.

β-strands and α-helices are labelled based on the *Zm*ITPK1 structure, where it does not impinge on clarity. Residues missing between β11 and α8 in *Zm*ITPK1 due to disorder are indicated by green asterisks. The α8 helix which runs behind the β-sheet of the C-terminal lobe is obscured by clipping in Pymol. The beginning (the C-terminal β-strand of the central lobe) and end (helix α8) of the 'tether' polypeptide of *St*ITPK1 are indicated by dark pink 1* and 2*, respectively. The six-residue hinge polypeptide between β8 and β9 connecting the central and C-terminal lobes, respectively, is indicated by a black arrow but is also obscured by clipping. The loop between β10 and β11 of *Zm*ITPK1 is the region which is stabilised due to IP6 binding. Image generated in Pymol.

Comparison of the residues in *St*ITPK1 which are equivalent to the residues which interact with IP6 in *Zm*ITPK1^{H192A} reveals inevitable differences. Whilst some residues – such as K70/K61, K29/K20, Y200/Y191 and N280/271 – are very well aligned, others such as R211/202, K306/298, K164/155 and K167/158 are distinctly different in orientation (Figure 5.13). The most pronounced difference is observed with R211/202, in which the side chain rotamers are oriented in opposing directions, such that whilst in *Zm*ITPK1 this residue interacts with the 3-phosphate, in *St*ITPK1 the equivalent residue interacts with the 4-phosphate. As such, in *St*ITPK1, the 4-phosphate participates in direct polar contacts with the side chains of two residues – R202 and K189 – whilst the 3-phosphate lacks interactions to any residues. These differences may explain the variations in substrate preference between the potato and maize enzymes. In *Zm*ITPK1, K306 forms interactions to the 6-phosphate, whereas the equivalent residue (K298) in *St*ITPK1 is further away (13.2Å vs 4.8Å) from the 6-phosphate. Nevertheless, the N-terminal lobe residues K29/20 and K70/61 are positioned to interact with the 6-phosphate in both enzymes, and this is via two water molecules. Meanwhile, the residues forming interactions with the 5-phosphate (a tyrosine and a glycine) are more conserved between the two enzymes. Another significant difference is that, whilst K164 and K167 in *Zm*ITPK1 participate in van der waals reactions with the 1- and 2-phosphates, respectively, the equivalent *St*ITPK1 residues, K155 and K158, are located much further away. Of course, binding of IP6 to *St*ITPK1 would likely induce a shift in the positioning of these residues, and so this rudimentary analysis merely serves to provide a comparison of the binding pocket residues both in the presence and absence of IP6.

Whilst the *St*ITPK1 active site has been described as narrow, the binding pocket of *Zm*ITPK1 is shallow and open. Like *St*ITPK1, the structure of *At*ITPK4 has a bound nucleotide and lacks substrate, but the binding pocket has been described as more similar to that found in *Zm*ITPK1.

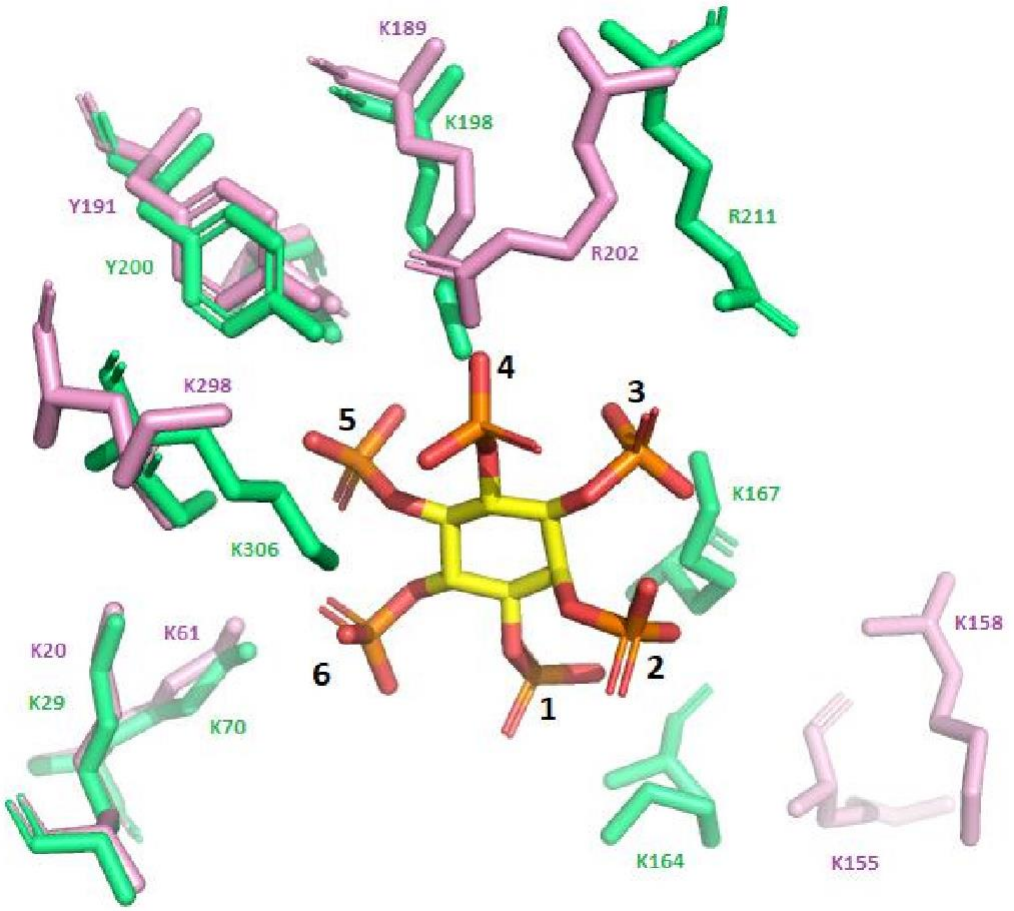


Figure 5.13 The residues which interact with IP6 in *Zm*ITPK1 (green), and the equivalent *St*ITPK1 residues (pink), are displayed along with IP6, in a structural alignment of *Zm*ITPK1 and *St*ITPK1 in Pymol. The phosphates of IP6 are numbered accordingly. Atom colours of IP6 are as follows: yellow = carbon; red = oxygen; orange = phosphorus.

5.4 Discussion

The successful structure solution of a protein is entirely dependent on the quality of the X-ray diffraction data generated from the crystal, and therefore the size and physical properties of the crystal ultimately determine the likelihood of success[239,359,360]. In turn, the crystallisation of a protein is dependent on a multiplex system of factors, yet due to the empirical nature of crystal growth, serendipity (and perseverance) often plays a significant part[239,361]. As Bob Cudney stated, '*There is only one rule in the crystallization. And that rule is, there are no rules*'[361]. The amount and combination of variables to screen are almost infinite and as such, despite increasing technological advances in automation, one is limited by amount of target protein and patience[239,360]. Indeed, it is estimated that 70-80% of soluble, non-membrane, proteins are recalcitrant to crystallisation[362]. Obtaining a crystal of an appropriate size and morphology is only the first battle – whether it produces adequate X-ray diffraction data is another challenge entirely and therefore optimisation of crystal growth represents the most common obstacle to successful structure solution[360].

This project was entirely dependent on the acquisition of diffraction quality co-crystal complexes and was inherently 'high risk, high reward' in nature. As with all macromolecular crystallography experiments, the process was considerably time-consuming with no guarantee of useable data, hence the high risk. This was compounded by the challenge of obtaining a complex with bound co-factor and substrate. The binding of a cofactor or ligand often induces a conformational change in a protein which may promote a more stable, monodisperse conformation through contribution of beneficial cross-links in the crystal[363] and as such, when the other molecule is not bound, the protein, being more dynamic, may be less likely to form the homogenous ordered solution necessary for the formation of a crystal lattice[360]. This was exemplified in the work of Zong et al on the *Zm*ITPK1 structure, whereby crystals were not obtained of the apoprotein, yet soaking of a ligand into a single mutant causes stabilisation of a region in the protein and consequently forms a secondary structural element which is absent in the apoprotein[339]. However, these researchers were unable to resolve electron density for the nucleotide analogue[339]. It is unclear whether the crystals of *St*ITPK1 in the work reported in this chapter contained bound nucleotide analogue, since the electron density map generated was poor.

The morphology of the *St*ITPK1 crystals formed were primarily very fine needles which were inherently more difficult to harvest and handle. The fact that they did not diffract very well could be due to their being more disposed to radiation damage as a result of their small volume. These type of crystals are notorious for cracking during the cryo-cooling stage, due to over-hydration[362]. Indeed, several harvested crystals dissolved when introduced to the

cryoprotectant solution. It is plausible therefore that the reason for poor X-ray diffraction data of the crystals which made it as far as the synchrotron was due to crystal defects developed during handling. In retrospect, a solution to overcome this obstacle could be to use a harvesting loop with an elliptical aperture specific for needles, which provides more support for the crystal and serves to minimise the surrounding excess liquid which, in standard loops, often gives rise to background X-ray scatter. Another strategy to provide support for needle crystals during harvesting is to dip the cryoloop into a solution of 0.5% Formvar, the resulting thin film of which offers support for fragile crystals. Although this can cause a slightly higher background, the reflections generated are much sharper[364]. Another option would have been to use some of the needles for seeding, in order to optimise subsequent crystallisation attempts[239,360]. A seed stock was prepared but due to time limitations, further purification and crystallisation experiments were not completed. Quality metrics of the X-ray diffraction data did indicate the possibility of twinning in the crystal, rendering molecular replacement an arduous challenge, therefore, if time permitted, attempts could have been made to 'de-twin' the data, or more simply, to return to optimisation of crystal growth. Since completing the *in vitro* aspect of the *St*ITPK1 structural studies, a structure of *St*ITPK1 in complex with ADP was reported by Whitfield et al[353]. Like the crystals reported in this thesis, the crystal which ultimately led to the elucidation of the structure was of the needle morphology and C 2 2 2 space group, and the data was particularly demanding to delineate (personal communication)[365].

Another predominant crystal morphology obtained in crystallisation screens of *St*ITPK1 was what is termed a crystal bouquet, as pictured in Figure 5.6. These clusters of needles, which share a common nucleus, are formed when a single nucleus yields an active crystal, but one of the crystal faces becomes nutrient-deprived as a result of the nature of the surface attachment (heterogeneous epitaxy). As such, whilst the other faces of the crystal continue to extend, this face is unable to develop and introduces stress into the lattice, which is relieved by the splintering of the crystal. Lattice strain can also be caused by the accumulation of impurities in the growing crystal. Again, seeding can be employed in an attempt to improve the morphology[360].

Despite multiple crystallisation attempts, the *St*ITPK1 crystals in this thesis only generated one potentially useable dataset, yet the scores returned from Phaser indicated that molecular replacement was unsuccessful. It is possible that this was a result of the flexible nature of the protein, for example, if the relative orientation of domains were different in the crystal to the model. The model used for molecular replacement was the recently solved *St*ITPK1/ADP complex (PDB: 8OXE)[353]. Initially, the *Zm*ITPK1 structure (7TN5) was used[339]. Despite being the

same protein, the conformation must have been different between the *St*ITPK1 crystals here, and the crystals which yielded the 8OXE model, rendering the electron density map unusable. This could have been the case if the ATP analogues used in crystallisation didn't actually bind to *St*ITPK1, unlike in 8OXE where the ADP was resolved, meaning that the two would adopt different conformations in certain regions. This is certainly the case where the lack of bound nucleotide in *Zm*ITPK1 resulted in a markedly different orientation of the central domain relative to that in the *St*ITPK1 co-crystal structure[339,353]. Given the combination of a difficult dataset, substantial time restrictions and the fact that the 8OXE structure had recently been published the decision was made to not proceed with attempts to improve molecular replacement. Otherwise, the sensible approach would be to split the model into the three different domains and perform molecular replacement on these individually.

As with *At*ITPK1, enantiomeric IP_4 substrate binding was modelled by Whitfield et al, to justify the preference displayed by *St*ITPK1 for Ins $(3,4,5,6)P_4$ over the Ins $(1,4,5,6)P_4$ enantiomer[353,366]. However, as Zong et al pointed out, the uncertainty with regard to modelling must be appreciated, since side chains of catalytic residues are likely to undergo ligand binding-induced changes in orientation[339]. Therefore, predicting the binding of different substrates with a static structural model may not reflect the true situation. Indeed, in the IP_6-liganded structure of *Zm*ITPK1, the R211 residue interacts with the 3-phosphate however, the equivalent residue in *St*ITPK1 is oriented away from the 3-phosphate, instead interacting with the 4-phosphate. Of course, in reality, the binding of IP_6 into the substrate binding site of *St*ITPK1 would likely induce movement of the involved residues. As such, the pursuit of *St*ITPK1 crystals in complex with these different IP_4 substrates is still required in order to solve the structures of the complexes such that a more accurate reflection is obtained. It would also be interesting to obtain structures of the *Zm*ITPK1 in complex with different IP_4 substrates and subsequently determine if the substrate preferences are conserved between these two plant ITPK1 orthologs.

Nevertheless, due to the work by Whitfield et al, it is now established that *St*ITPK1, like its *Arabidopsis thaliana* ortholog *At*ITPK1, displays a preference for IP_4 over IP_6 and pyrophosphate substrates, with the IP_4 substrate $Ins(3,4,5,6)P_4$ actually inhibiting IP_6 kinase activity[353]. This is most likely in agreement with the conclusion of Zong et al that the relative rates of IP_6 kinase and Ins $(1,3,4,5,6)P_5$ phosphatase activities are governed by a catalytic specificity element among individual members of the plant ITPK1 family[339]. Similarly, in *A thaliana*, the combination of ITPK1 and IPK1 drives 5-$InsP_7$ synthesis from $Ins(3,4,5,6)P_4$ in an ATP-dependent manner[366]. Ins $(3,4,5,6)P_4$ is a recognised physiological substrate of human ITPK1 playing a key role in cell signalling, and as such, the predominant ITPK1 activity in

humans is the generation of Ins (3,4,5,6)P_4 by 1-phosphatase activity[339,341]. It is thought that mammals have evolved a mechanism for regulating the cell signalling of Ins (3,4,5,6)P_4 and this is reflected by differences in ITPK1 homologs across different kingdoms[343]. Despite mammals displaying some level of IP_6 kinase activity, the au courant literature indicates that IP_6 phosphokinase activity by ITPK enzymes is primarily restricted to plants[339]. A much larger number of ITPK1 enzymes are encoded in plant relative to metazoan genomes, and despite the presence of an ITPK1 homolog in mammals, canonical IP6K enzymes – absent in plant lineages -fulfil the analogous function[342]. Nevertheless, the lack of IP6K enzymes in plants has hindered elucidation of the function and biosynthesis of inositol pyrophosphates, whilst in animals and yeast, knowledge of the metabolic pathways yielding inositol pyrophosphates is relatively much more advanced[356]. Ultimately, there are still aspects of the inositol phosphate metabolic pathway in plants, namely the physiological role and metabolic contributions of some ITPKs, which remain elusive and necessitate further *in vitro* and *in vivo* investigation[342]. For example, previously the consensus had been that Ins (1,3,4)P_3 was not present in plants, and neither was an enzyme which synthesised it[343]. However, it has since been shown that Ins (1,3,4)P_3 is a substrate of all isoforms of *A. thaliana* and *G. max* ITPK1 enzymes[342,367]. It is clear that new knowledge of this complex system across different domains of life is continually being accumulated, thereby changing the landscape of the field. The reader is thus referred to recent reviews by Tu-Sekine et al and by Xiong et al for au courant coverage of the pathways of inositol phosphate synthesis[349,368].

Whilst the work in this chapter has not yielded successful contribution to the inositol phosphate kinase field, it may at least serve to highlight the importance of optimising the factor which everything downstream is governed by – the crystal. Indeed, one cannot synthesise a fine wine with rotten grapes.

"Would you tell me, please, which way I ought to go from here?" "That depends a good deal on where you want to get to", said the Cat.

CHAPTER SIX

Final Discussion

As with most theses, this thesis tells a research story; one which ends up following a different path to that projected at the start, with some unexpected revelations along the way. Ultimately, the aim of the overall project was to make progress towards the goal of enzymatically unlocking the maximum amount of phosphate from phytate**,** with an early working title *of 'Towards the complete enzymatic dephosphorylation of phytic acid'.* The obvious approach was to identify in the literature any promising candidates which could be used as a starting point for bioengineering. A phytase from *Klebsiella pneumoniae* 9-3B – Phy9-3B – was reported in 2012 to complete the dephosphorylation of IP6 (detailed in section 1.5.1). The fact that its properties were suitable for the hydrolysis of phytate for industrial applications and no other researchers seemed to be working on it rendered it an ideal contender for this work**.** Unfortunately, it was not possible to access the gene sequence and as such, this project came to a dead end. A further promising candidate was based on the yeast phytase PhytDc from *Debaryomyces castelli* CB3, however attempts to produce the recombinant protein were unfruitful due to practical limitations. Since these were the only 2 enzymes (to the author's knowledge) which were claimed to degrade phytate to myo-inositol, alternative approaches were turned to. One strategy was to build upon previous groundwork on the MINPP enzyme family, testing MINPP variants with promising phytate-degrading properties and iteratively introducing further mutations until the arrival at a phytase with enhanced features; this reflects the first data chapter of this thesis. Another was to investigate the enzymatic and structural properties of novel or unusual phytate-degrading enzymes in order to shed light on structural features which may dictate functional characteristics; this constitutes the second data chapter of this thesis. The wider sphere of phytate metabolism in plants encompasses both the addition and removal of phosphate groups, whereby phytate is continually synthesized and degraded in a highly regulated equilibrium. Therefore, whilst the majority of this thesis is focussed on the degradation of phytate, the third data chapter is concerned rather with its biosynthesis.

Ultimately, whilst none of this work resulted in progress towards the complete dephosphorylation of phytate, what it did achieve was unveil error within the field of phytase research. Rather than disregard this unanticipated observation, adhering rigidly to the original agenda, the author personally was committed to solving this mystery which offered invaluable information for the enzyme research community. Indeed, the fundamental criterion for a doctoral degree demands an original contribution to knowledge. This is tangibly illustrated

through the ‘MβLp01 saga’, in which not only was the experimentally determined structure presented for the first time, but the discoveries of 1) incorrect evidence in the scientific literature and 2) a novel enzyme function previously unreported were made. Fundamentally, understanding evolves more from disagreement than confirmation. Thus, whilst it was unfortunate to not have contributed more in the remainder of the thesis work, the author is grateful to the challenges of the MβLp01 saga for the opportunity it presented to make one a better researcher.

Appendices

A3.2.1

Mutant	Forward Primer Sequence	Reverse Primer Sequence
R275L	AATTTGCTACAATATATGAGTAAGAGCTCTGC	TATATTGTAGCAAATTCTGGGTTTGCCAATAAC
R275A	GAATTTGGCACAATATATGAGTAAGAGCTCTGC	TATATTGTGCCAAATTCTGGGTTTGCCAATAAC
R275S	AATTTGAGCCAATATATGAGTAAGAGCTCTGC	TATATTGGCTCAAATTCTGGGTTTGCCAATAAC
R275Y	GAATTTGTATCAATATATGAGTAAGAGCTCTGCACC	CATATATTGATACAAATTCTGGGTTTGCCAATAACGATG
R275N	GAATTTGAATCAATATATGAGTAAGAGCTCTGCACC	CATATATTGATTCAAATTCTGGGTTTGCCAATAACGATG
R275D	GAATTTGGATCAATATATGAGTAAGAGCTCTGCACC	CATATATTGATCCAAATTCTGGGTTTGCCAATAACGATG
R275M	AGAATTTGATGCAATATATGAGTAAGAGCTCTGCACC	CATATATTGCATCAAATTCTGGGTTTGCCAATAACGATGC
Q276L	GAATTTGCGACTATATATGAGTAAGAGCTCTGCACC	CATATATAGTCGCAAATTCTGGGTTTGCCAATAACG
Q276A	GAATTTGCGAGCATATATGAGTAAGAGCTCTGCACC	CATATATGCTCGCAAATTCTGGGTTTGCCAATAACG
Q276S	GAATTTGCGATCATATATGAGTAAGAGCTCTGCACC	CATATATGATCGCAAATTCTGGGTTTGCCAATAACG
Q276Y	GAATTTGCGATATTATATGAGTAAGAGCTCTGCACCTGTC	CTCATATAATATCGCAAATTCTGGGTTTGCCAATAACG
Q276N	GAATTTGCGAAATTATATGAGTAAGAGCTCTGCACCTGTC	CTCATATAATTTCGCAAATTCTGGGTTTGCCAATAACG
Q276D	GAATTTGCGAGATTATATGAGTAAGAGCTCTGCACCTGTC	CTCATATAATCTCGCAAATTCTGGGTTTGCCAATAACG
Q276M	GAATTTGCGAATGTATATGAGTAAGAGCTCTGCACCTGTCGG	ACTCATATACATTCGCAAATTCTGGGTTTGCCAATAACGATGCC
K280L		
K280A	ATGAGTGCCAGCTCTGCACCTGTCGGTAAG	AGAGCTGGCACTCATATATTGTCGCAAATTCTGG
K280S	ATGAGTAGCAGCTCTGCACCTGTCGGTAAG	AGAGCTAGCACTCATATATTGTCGCAAATTCTGG
K280Y		
K280N		
K280D		
K280M	ATGAGTATGAGCTCTGCACCTGTCGGTAAG	AGAGCTATGACTCATATATTGTCGCAAATTCTGG

A3.2.2

Sequence results of the *Bt*MINPP variants

R275L (T7 Forward)

TTGGGGGACATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATC
ATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAG
GGACGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAA
ATCATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTG
GTATCAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCA
GATTATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGT
ATGATAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAG
CATAAATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAG
GAAAACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACT
GGCTTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATC
CGGAACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGAT
ACAAGCATCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTT
GCTACAATATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTT
CTGAATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCAT
GCAGAAACAGTTATACCTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATT
CCGGTTTCAGTCTATTGGAAGGATTATGAAATATCTCCATGGGCCGCTAAAGG

R275L (T7 Reverse)

CCGCTCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATTGTAGCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATG

R275A (T7 Forward)

GGGACATTCCCCTCTAGGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCC
ATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGA
CGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATC
ATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTAT
CAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATT
ATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGA
TAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATA
AATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAA
AACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGC
TTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGG
AACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACA
AGCATCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGGC

ACAATATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTG
AATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCA
GAAACAGTTATACCTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTCCG
TTTCAGTCTATTGGAAGGATTATGAAATATCTCCAATGGCCGCTAAGGTACA

R275S (T7 Forward)

GAAAATTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCGCAGCAGCCATCATCCATC
ATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACG
GCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATC
TGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATCA
GCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATT
TGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGATAA
GGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAAT
AGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAAAC
AATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCC
TATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACA
ATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCA
TCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGAGCCAA
TATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGAATT
CATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCAGAA
ACAGTTATACCTTTTGTGTCGTTGAAGGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTCCGTT
TCAGTCTATTGGAAGGATATGAAATATCTCCATGGGCCGCTAATGGAA

R275Y (T7 Reverse)

CGCTAGCTTCTTTCGGGCTTTGTTAGCAGCCGGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATTGATACAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGAGGAT

R275N (T7 Forward)

TTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGACATATACCATGGGCAGCAGCCATCATCATCATCATCAC
AGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACGGCCATGCC
CTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATCTGGGAAG
ACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATCAGCCCAACA
AGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATTTGACGGTC
AATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGATAAGGAACTAT
CCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAATAGTATGGA
CGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAAACAATATAATC
ATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCCTATATACAA
AGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACAATATCTGG
ATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCATCCCATTA
AATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCCAAACCCAGAAATTTGAATCAATATATG

AAGTAAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGAAATTCA
TTCCGTTCTGC

R275N (T7 Reverse)

CAGGCCTGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCC
ATTTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGG
AAGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACT
GTACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTG
TACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATTGATTCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGATGATTTATATAAAATGGGGTCATACCATCC
CAAAAAGAATGGACGAAT

R275D (T7 Reverse)

CCGCTCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATTGATCCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCGAAGATT

R275M (T7 Forward)

GAAATAATTTTGTTTATCTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATACATTCAATCATCACAGCA
GCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACGGCCATGCCCTATC
CTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATCTGGGAAGACATG
GAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATCAGCCCAACAAGAG
AATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATTTGACGGTCAATG
GGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGATAAGGAACTATCCTC
AACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAATAGTATGGACGCT
TTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAAACAATATAATCATAT
CCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCCTATATACAAAGC
TTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACAATATCTGGATAA
AGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCATCCCATTAAATT
TGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGATGCAATATATGAGTAAG
AGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGAATTCATTCGTTCTGC

TCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCAGAAACAGTTATACCTT
TTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTCCGTTTCAGTC

Q276L (T7 Forward)

GGGGTCATTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCCA
TTCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGA
CGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATC
ATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTAT
CAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATT
ATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGA
TAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATA
AATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAA
AACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGC
TTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGG
AACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACA
AGCATCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCG
ACTATATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTG
AATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCA
GAAACAGTTATACCTTTTGTGTCGTTGATGGGCCAAAA

Q276A (T7 Forward)

GCCTCGGCCTGGAGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACGGCCATGCCC
TATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATCTGGGAAGA
CTTGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGCATCAGCCCAACA
AGAGAATGGACTGACCTCAGAAGGGATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATTTGACGGTC
AATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGTGAATTGCCGGACGTATGATAAGGAACTA
TCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAATAGTATGG
ACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGCACAGCGAAGTGAAGGAAAACAATATAA
TCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCCTATATAC
AAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACAATATCT
GGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCATCCCAT
TAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCGAGCATATATG
AGTAAGAGCTCTGCACCTGCCGC

Q276A (T7 Reverse)

GGCATCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATGCTCGCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCATGTCTTCCCAATGATTTATTAA

Q276S (T7 Forward)

GGGGGGAAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCA
TCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGG

GACGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAA
TCATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGG
TATCAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAG
ATTATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTA
TGATAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGC
ATAAATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGG
AAAACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTG
GCTTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCC
GGAACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATA
CAAGCATCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTG
CGATCATATATGAGTAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCT
GAATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGC
AGAAACAGTTATACCTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTCC
GTTTCAGTCTATTGGAAGGATTATGAAATATCTCCAAGGGCCGCTAATGTACAGTGGCTTTTCTATCGTGAAG

Q276Y (T7 Forward)

GAGGTAATTCCCCTCTAGATTAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCCT
TCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGAC
GGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCA
TCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATC
AGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTA
TTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGAT
AAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAA
ATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAA
ACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTT
CCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAA
CAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAG
CATCCCATTAAATTTGGAAGACCTTTTTACACATGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCGAT
ATTATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGGCTTTGCTTTCTGAA
AT

Q276Y (T7 Reverse)

ACGCTCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATAATATCGCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGATGAATTTAAATAAAATGGGGTCATACCAT
CCCGAAAAGTAATGGACGAATC

Q276N (T7 Forward)

TTTGGGGAAATTCCCTCTAGAATAATTTTAGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCAT
CATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGG
ACGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAAT
CATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGT

ATCAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGA
TTATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTAT
GATAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCA
TAAATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGG
AAAACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTG
GCTTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCC
GGAACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATA
CAAGCATCCCATTACATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTG
CGAAATTATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTC
TGAATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAGATCCGATTATCAGGCTAACTTTCGTTTTGCTCATG
CAGAAACAGTTATACCTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTC
CGTTTCAGTCTATTGGAAGGATTATGAAATATCTCCAATGGCCGCTAATGGACAG

Q276N (T7 Reverse)

CCCGCTCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCA
TTTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATAATTTCGCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAACGGGCTC

Q276D (T7 Forward)

TTTGGGGACATTCCCTCTAGCAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCAT
CCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGG
GACGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAA
TCATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGG
TATCAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAG
ATTATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTA
TGATAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGC
ATAAATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGG
AAAACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTG
GCTTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCC
GGAACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATA
CAAGCATCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTG
CGAGATTATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGACTTTGCTTTC
TGAATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCAT
GCAGAAACAGTTATACCTTTTGTGTCCTTGATGGAGATAAAAAAACTGATGTACAGGTTTGCCGGACCAATA

Q276D (T7 Reverse)

GCATCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCATT
TCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGAA
GCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTGT
ACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGTA
CATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAGC
CTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGATT

GCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATAATCTCGCAAATTCTGGGTTTGCCAATA
ACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGCA
ACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCTT
CATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTATA
ATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCTG
TAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGTT
GCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTCT
TGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGAG
AGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTAG
GGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCGAAGAA

Q276M (T7 Forward)

TTTGGGGAAATTCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATC
ATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGA
CGGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATC
ATCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTAT
CAGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATT
ATTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGA
TAAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATA
AATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAA
AACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGC
TTCCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGG
AACAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACA
AGCATCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCG
AATGTATATGAGAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGA
ATTCATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCAG
AAACAGTTATACCTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTCCGT
TTCAGTCTATTGGAAGGATTATGAAATATCTCCAATGGCCGCTAATGAAAGTG

K280L (T7 Reverse)

ACCTGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCATTT
CGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGAAG
CGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTGTA
CATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGTAC
ATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAGCCT
GATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGATTGC
AACCGGTAACATCTTACCGACAGGTGCAGAGCTCAGACTCATATATTGTCGCAAATTCTGGGTTTGCCAATAAC
GATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGCAAC
AGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCTTCA
TTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTATAAT
TGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCTGTA
AAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGTTGC
TATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTCTTG
TTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGAGA
GTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTAGG
GCTTTTCTTGAAGTTGGGAAGCGGGCTC

K280A (T7 Reverse)

CGCTCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCATT
TCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGAA
GCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTGT
ACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGTA
CATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAGC
CTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAACGATT
GCAACCGGTAACATCTTACCGACAGGTGCAGAGCTGGCACTCATATATTGTCGCAAATTCTGGGTTTGCCAAT

AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGATGAT

K280S (T7 Forward)

GGGTAATTCCCCTCTAGGAGTAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCAT
CATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACG
GCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATC
TGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATCA
GCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATT
TGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGATAA
GGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAAT
AGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAAAC
AATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCC
TATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACA
ATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCA
TCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCGACAA
TATATGAGTAAGAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGAATT
CATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCAGAA
ACAGTTATACCTTTTGGGGTC

K280S (T7 Reverse)

CCGTCGCTTCTTTCGGGCTTTGTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCATT
TCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGAA
GCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTGT
ACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGTA
CATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAGC
CTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGATT
GCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATTGTCGCAAATTCTGGGTTTGCCAATA
ACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGCA
ACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCTT
CATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTATA
ATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCTG
TAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGTT
GCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTCT
TGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGAG
AGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTAG
GGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGATGATTTAA

K280Y (T7 Reverse)

CGATCGCTTCTTTCGGGCTTTGTTAGCAGCCGGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTGTAACTCATATATTGTCGCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT

TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGATGATTTAAATAAA

K280N (T7 Forward)

CGGAAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCAT
CATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACG
GCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATC
TGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATCA
GCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATT
TGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGATAA
GGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAAT
AGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAAAC
AATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCC
TATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACA
ATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCA
TCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCGACAA
TATATGAGTAACAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGAATT
CATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCAGAA
ACAGTTATACCTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATTCCGTTT
CAGTCTATTGGAAGGATTATGAAATATCTC

K280N (T7 Reverse)

CCATTGCTTCTTTCGGGCTTTGTTAGCAGCCGGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCAT
TTCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGA
AGCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTG
TACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGT
ACATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAG
CCTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGAT
TGCAACCGGTAACATCTTACCGACAGGTGCAGAGCTGTTACTCATATATTGTCGCAAATTCTGGGTTTGCCAAT
AACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGC
AACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCT
TCATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTAT
AATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCT
GTAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGT
TGCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTC
TTGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGA
GAGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTA
GGGCTTTTCTTGAAGTTGGGAAACGGGCTCCATGTCTTCCCAGATGAT

K280D (T7 Forward)

GGTAATTCCCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCAT
CATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGACGGC
CATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCATCTG
GGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATCAGC
CCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTATTTG
ACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGATAAG
GAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAAATA
GTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAAACA
ATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTTCCT
ATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAACA

ATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAGCA
TCCCATTAAATTTGGAAGACCTTTTTACACTTGATGAATGGCATCGTTATTGGCAAACCCAGAATTTGCGACAA
TATATGAGTGATAGCTCTGCACCTGTCGGTAAGATGTTACCGGTTGCAATCGCTTGGCCTTTGCTTTCTGAATT
CATTCGTTCTGCTCAAGAGGTGATAAGTGGGAAATCCGATTATCAGGCTAACTTTCGTTTTGCTCATGCAAAA
ACAGTTATACTTTTGTGTCGTTGATGGGCATAGAAAAAACTGATGTACAGGTTTGCCGGCCCGATCCCGTTTC
AGCCTATG

K280D (T7 Reverse)

TATGTTTAGCAGCCGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCATTTCGATTCGCTGATTAAAG
AATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGAAGCGCTGCTGCTTCCTCAT
TTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTGTACATTAGCGGCCATTGGA
GATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGTACATCAGTTTTTTCTATGCC
CATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAGCCTGATAATCGGATTTCCCA
CTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGATTGCAACCGGTAACATCTTA
CCGACAGGTGCAGAGCTATCACTCATATATTGTCGCAAATTCTGGGTTTGCCAATAACGATGCCATTCATCAAG
TGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGCAACAGAAAATAAAGCCATG
ACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCTTCATTATAGGAACTGGAGAT
ATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTATAATTGACATAAGATTTATTCA
GATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCTGTAAAGCTGGATTATGCCTTA
TCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGTTGCTATTGCTTCAATCTTGGC
AGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTCTTGTTCCGTCTCGCCCAACTT
TGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGAGAGTAAAGCCATGCCTTCTG
AGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGGCTAGGGTTTTCTTGAGATTGGGA
AACGGG

K280M (T7 Forward)

TAGGGAAATTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCA
TCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCAAACTAAGATACAGAAGTATGCAGGGAC
GGCCATGCCCTATCCTAATAGGACAGACTCGTCCATTACTTTTCGGGATGGTATGACCCCATTTTATATAAATCA
TCTGGGAAGACATGGAGCCCGCTTCCCAACTTCAAGAAAAGCCCTAGACAAAGTGGAGAAGGTATTGGTATC
AGCCCAACAAGAGAATGGACTGACCTCAGAAGGCATGGCTTTACTCTCTATGATACGACGTTTATCCAGATTA
TTTGACGGTCAATGGGGAAAATTGTCAAAGTTGGGCGAGACGGAACAAGAAGGAATTGCCGGACGTATGAT
AAGGAACTATCCTCAACTATTCAGCAATTCTGCCAAGATTGAAGCAATAGCAACCTATGTTCCCCGCAGCATAA
ATAGTATGGACGCTTTTCTTTCTTGCATGATAAGGCATAATCCAGCTTTACAGGTACAGCGAAGTGAAGGAAA
ACAATATAATCATATCCTTCGTTTTTTTGATCTGAATAAATCTTATGTCAATTATAAAGAGAAAGGTGACTGGCTT
CCTATATACAAAGCTTTTGTACACAAGAAGATATCTCCAGTTCCTATAATGAAGAAGTTTCTTCTTAATCCGGAA
CAATATCTGGATAAAGAGGCAGAGGAATTTGTCATGGCTTTATTTTCTGTTGCGGCTATTCTTCCTGATACAAG
TATCCCATTAAATTTGGAAGACCTTTT

K280M (T7 Reverse)

ACCTGCTTCTTTCGGGCTTTGTTAGCAGCCGGGATCCCTATTCATTGAAAACTGAAAGTGTTTTTTTAGCCATT
TCGATTCGCTGATTAAAGAATATACGAGTTTTTTCCCAAGAGTAATATGGGAAACAAGCGGTTGATATAGGAA
GCGCTGCTGCTTCCTCATTTAACAGAATCTTAACCCAAATTCTTTGGTCTCTGTCACGATAGAAAAGCCACTGT
ACATTAGCGGCCATTGGAGATATTTCATAATCCTTCCAATAGACTGAAACGGAATCGGGCCGGCAAACCTGTA
CATCAGTTTTTTCTATGCCCATCAACGACACAAAAGGTATAACTGTTTCTGCATGAGCAAAACGAAAGTTAGC
CTGATAATCGGATTTCCCACTTATCACCTCTTGAGCAGAACGAATGAATTCAGAAAGCAAAGGCCAAGCGATT
GCAACCGGTAACATCTTACCGACAGGTGCAGAGCTCTTACTCATATATTGTCGCAAATTCTGGGTTTGCCAATA
ACGATGCCATTCATCAAGTGTAAAAAGGTCTTCCAAATTTAATGGGATGCTTGTATCAGGAAGAATAGCCGCA
ACAGAAAATAAAGCCATGACAAATTCCTCTGCCTCTTTATCCAGATATTGTTCCGGATTAAGAAGAAACTTCTT
CATTATAGGAACTGGAGATATCTTCTTGTGTACAAAAGCTTTGTATATAGGAAGCCAGTCACCTTTCTCTTTATA
ATTGACATAAGATTTATTCAGATCAAAAAAACGAAGGATATGATTATATTGTTTTCCTTCACTTCGCTGTACCTG
TAAAGCTGGATTATGCCTTATCATGCAAGAAAGAAAAGCGTCCATACTATTTATGCTGCGGGGAACATAGGTT
GCTATTGCTTCAATCTTGGCAGAATTGCTGAATAGTTGAGGATAGTTCCTTATCATACGTCCGGCAATTCCTTCT
TGTTCCGTCTCGCCCAACTTTGACAATTTTCCCCATTGACCGTCAAATAATCTGGATAAACGTCGTATCATAGAG

AGTAAAGCCATGCCTTCTGAGGTCAGTCCATTCTCTTGTTGGGCTGATACCAATACCTTCTCCACTTTGTCTAG
GGCTTTTCTTGAAGTTGGGAAGCGGGCTCCATGTCTTCCCAGATGATTTAAATAAA

A3.2.3-4

	2 tailed (either direction)			1 tailed		
	***t*-value**	***p*-value**	**Significance**	**t-value**	**p-value**	**significance**
R275L	11.05	0.000381	Significant at <0.01		0.000191	Significant at <0.01
R275A	22.44	0.000023	Significant at <0.01		0.000012	Significant at <0.01
R275S	15.55	0.000578	Significant at <0.01		0.000289	Significant at <0.01
R275Y	10.55	0.000456	Significant at <0.01		0.000228	Significant at <0.01
R275N	15.10	0.000112	Significant at <0.01		0.000056	Significant at <0.01
R275D	31.03	<0.00001	Significant at <0.01		<0.00001	Significant at <0.01
R275M	1.47	0.238087	Not significant		0.119043	Not significant
Q276L	19.63	0.00004	Significant at <0.01		0.00002	Significant at <0.01
Q276A	8.71	0.000956	Significant at <0.01		0.000478	Significant at <0.01
Q276Y	4.61	0.009986	Significant at <0.01		0.004993	Significant at <0.01
Q276N	21.90	0.000026	Significant at <0.01		0.000013	Significant at <0.01
Q276D	22.18	0.000024	Significant at <0.01		0.000012	Significant at <0.01
K280L	31.50	<0.00001	Significant at <0.01		<0.00001	Significant at <0.01
K280A	31.84	<0.00001	Significant at <0.01		<0.00001	Significant at <0.01
K280S	5.41	0.005652	Significant at <0.01		0.002826	Significant at <0.01
K280Y	23.13	0.000021	Significant at <0.01		0.00001	Significant at <0.01
K280N	18.73	0.000048	Significant at <0.01		0.000024	Significant at <0.01
K280D	27.20	0.000011	Significant at <0.01		<0.00001	Significant at <0.01
K280M	19.30	0.000042	Significant at <0.01		0.000021	Significant at <0.01

Table A3.2.3 Statistical significance of the relative activities against IP_6 of the *Bt*MINPP variants. Values are based on the raw A_{700nm} values and were calculated by a t-test for independent means (https://www.socscistatistics.com/tests/studentttest/default2.aspx).

	2 tailed (either direction)			1 tailed		
	***t*-value**	***p*-value**	**Significance**	**t-value**	**p-value**	**significance**
R275L	-1.88	0.133441	Not significant		0.000191	Significant at <0.01
R275A	0.68	0.535925	Not significant		0.000012	Significant at <0.01
R275S	-3.88	0.017831	Significant at <0.05		0.000289	Significant at <0.01
R275Y	-4.10	0.026321	Significant at <0.05		0.000228	Significant at <0.01
R275N	-2.92	0.043256	Significant at <0.05		0.000056	Significant at <0.01
R275D	1.89	0.131099	Not significant		<0.00001	Significant at <0.01
R275M	-0.36	0.73419	Not significant		0.119043	Not significant
Q276L	4.28	0.012828	Significant at <0.05		0.00002	Significant at <0.01
Q276A	0.86	0.438122	Not significant		0.000478	Significant at <0.01
Q276Y	-2.85	0.065376	Not significant		0.004993	Significant at <0.01
Q276N	4.19	0.013825	Significant at <0.05		0.000013	Significant at <0.01
Q276D	3.38	0.027842	Significant at <0.05		0.000012	Significant at <0.01
K280L	4.25	0.013151	Significant at <0.05		<0.00001	Significant at <0.01
K280A	3.91	0.017386	Significant at <0.05		<0.00001	Significant at <0.01
K280S	0.29	0.785318	Not significant		0.002826	Significant at <0.01
K280Y	4.82	0.008513	Significant at <0.05		0.00001	Significant at <0.01
K280N	0.58	0.593098	Not significant		0.000024	Significant at <0.01
K280D	4.17	0.014057	Significant at <0.05		<0.00001	Significant at <0.01
K280M			Significant at <0.01		0.000021	Significant at <0.01

Table A3.2.4 Statistical significance of the relative activities against IP_4 of the *Bt*MINPP variants. Values are based on the raw $_{A700nm}$ values and were calculated by a t-test for independent means (https://www.socscistatistics.com/tests/studentttest/default2.aspx).

A4.4.1

MβLp01 sequencing of the colonies B C and D which were correct

A4.4.2-3

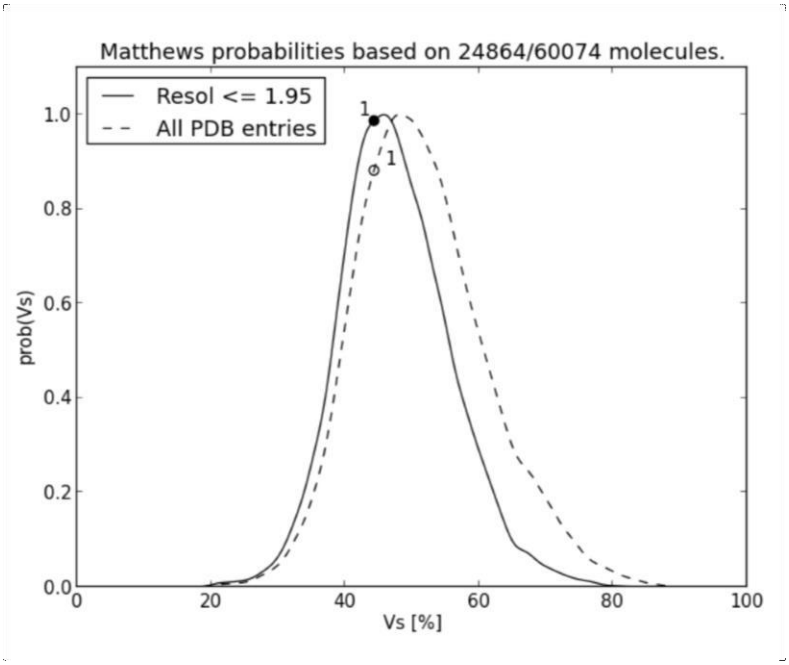
Pipeline	xia2 dials	**xia2 3dii**	autoPROC	fast_dp	autoPROC+STARANISO
a, b, c	53.23, 87.50, 118.19	**53.26, 87.61, 118.42**	53.22, 87.58, 118.40	53.28, 87.51, 118.42	53.22, 87.58, 118.40
α, β, γ	90, 90, 90	**90, 90, 90**	90, 90, 90	90, 90, 90	90, 90, 90
Resolution	48.53-2.05	**59.21-1.95**	70.41-1.70	29.61-2.52	70.41-1.76
I/σ(I)	3.4	**6.4**	4.9	9.6	5.5
R_{meas} Inner	0.195	**0.132**	0.125	0.112	0.116
R_{meas} Outer	1.226	**2.139**	0.908	0.618	0.535
Completeness	99.9	**100.0**	73.7	99.0	73.4
$CC_{1/2}$	0.9	**1.0**	1.0	1.0	1.0

Table A4.4.2 Data processing statistics for the MβLp01 crystal, including the dataset processed by xia2 3dii which yielded the structure by molecular replacement.

Pipeline	xia2 dials	**xia2 3dii**	autoPROC	fast_dp	autoPROC+STARANISO
Space Group	I 2 2 2	I 2 2 2	I 2 2 2	P1	I 2 2 2
a, b, c	53.56, 87.81, 118.48	53.56, 87.78, 118.56	53.59, 87.82, 118.56	53.04, 77.64, 77.77	53.59, 87.82, 118.56
α, β, γ	90, 90, 90	90, 90, 90	90, 90, 90	68.11, 70.14, 70.26	90, 90, 90
Resolution	45.73-2.18	48.81-2.03	70.57-1.88	29.38-3.10	70.57-1.84
I/σ(I)	0.1	6.5	7.7	2.1	8.3
R_{meas} Inner	0.718	0.125	0.073	0.164	0.073
R_{meas} Outer	-3.029	3.945	2.057	0.528	0.880
Completeness	99.6	99.5	79.7	92.6	78.0
$CC_{1/2}$	0.4	1.0	1.0	0.5	1.0

Table A4.4.3 Data processing statistics for the other of the two MβLp01 crystals.

A4.4.4


Matthews probabilities based on 24864/60074 molecules.
Resol <= 1.95
All PDB entries
1
1
prob(Vs)
Vs [%]
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100

A4.4.5 – ICP-MS data for MβLp01

	Sample List	32S | 32S.16O (S-TQ-O2)	58Ni (S-SQ-KED)	64Zn (S-SQ-KED)		Ni ratio to S	Zn ratio to S		Sample List	32S | 32S.16O (S-TQ-O2)	58Ni (S-SQ-KED)	64Zn (S-SQ-KED)
Start Time		ppb	ppb	ppb				Start Time	Label	ppb	ppb	ppb
2023-12-14 11:24:33	0.5mg/ml Emma OLD	4769.752	101.5225	338.9388		0.0213	0.0711	2023-12-14 10:40:29	mq2	-4.9973	0.006071	-0.65167
								2023-12-14 10:45:47	mq3	-4.94674	0.033751	-0.29931
2023-12-14 11:29:52	0.35 mg/ml Emma Fresh	14587.44	71.54562	6093.328		0.004905	0.41771	2023-12-14 10:51:05	mq4	-5.03133	0.028346	-0.21564
								2023-12-14 10:56:23	mq5	-5.67541	0.02078	-0.19277
								2023-12-14 11:01:41	mq6	-5.3791	0.029922	-0.20965
									Sdev	0.3129	0.0110	0.1933
									LOD	**0.939**	**0.033**	**0.580**
								2023-12-14 11:07:00	LMS-2A+4-	12.2721	10.3312	9.417859
								2023-12-14 11:12:17	LMS-2A+4-	10.61829	10.37491	9.387807
								2023-12-14 11:17:35	LMS-2A+4-	10.43494	10.23412	9.225614
								2023-12-14 11:35:11	LMS-2A+4-	11.43634	10.00956	9.018206
								2023-12-14 11:40:28	LMS-2A+4-	11.26614	10.08902	9.097603
								2023-12-14 11:45:45	LMS-2A+4-	11.0725	9.946638	9.054887
									mean	10.966	10.131	9.157
									precision	**0.426**	**0.174**	**0.151**
									rsd	3.884	1.714	1.649
									accuracy +/-	**0.966**	**0.131**	**0.843**

A4.4.6 – HPLC traces of 3 MβLp01 SEC fractions (E8, E9, E10)

[need to locate data on HPLC PC in Brearley lab]

A4.4.7 – The full data from Figure 4.4.23. HPLC traces of the IP6 hydrolysis profile generated by MβLp01 in comparison to a positive control (P15); IP6 only; and the IP6 hydrosylate standard. HPLC traces overlayed of MβLp01 (magenta), P15 (purple), IP6 only (colour?) and the IP6 hydrosylate (colour?). MβLp01 and P15 were each incubated with IP6 for 4h at 37°C before the reaction product mixture were subjected to HPLC.

A4.4.8 – The full data from Figure 4.4.32. Hydrolysis of nitrocefin by MβLp01-expressing *E. coli* cells and purified MβLp01 in comparison to the positive control supplied in the commercial kit. Also included alongside were uninduced *E. coli* TOP10 cells harbouring the MβLp01 gene; *E. coli* TOP10 cells harbouring the 'empty' pBAD202/D-TOPO vector and the purified HP-thioredoxin peptide.

List of References

1. Pioneering Animal Nutrition Products & Services [Internet]. AB Vista. Available from: https://www.abvista.com/
2. The feed behind our food [Internet]. Forum for the Future. Available from: https://www.forumforthefuture.org/feed-compass
3. The Protein Challenge 2040: Shaping the Future of Food [Internet]. Forum for the Future. Available from: https://www.forumforthefuture.org/Handlers/Download.ashx?IDMF=026a4509-ecf6-400a-b176-7e33ad649813
4. Gaffney J, Bing J, Byrne PF, Cassman KG, Ciampitti I, Delmer D, et al. Science-based intensive agriculture: Sustainability, food security, and the role of technology. Global Food Security. 2019 Dec;23:236–44.
5. Bywater AC, Baldwin RL. Alternative Strategies in Food-Animal Production. In: Baldwin RL, editor. Animals, Feed, Food and People (AAAS Selected Symposia; 42). Boulder, Colorado : Westview Press, Inc.; 1980. p. 1–30.
6. Smith NE. Opportunities for Forage, Waste and By-Product Conversion to Human Food by Ruminants. In: Baldwin RL, editor. Animals, Feed, Food and People (AAAS Selected Symposia; 42). Boulder, Colorado : Westview Press, Inc.; 1980. p. 53–63.
7. Whittaker JK, Brokken RF, Eisgruber LM. Past, Present and Future Resource Allocation to Livestock Production. In: Baldwin RL, editor. Animals, Feed, Food and People (AAAS Selected Symposia; 42). Boulder, Colorado : Westview Press, Inc.; 1980. p. 81–102.
8. Bayley HS. Opportunities for Waste and By-Product Conversion to Human Food by Non-Ruminants. In: Baldwin RL, editor. Animals, Feed, Food and People (AAAS Selected Symposia; 42). Boulder, Colorado : Westview Press, Inc.; 1980. p. 65–79.
9. Bringezu S, Schütz H, O'Brien M, Kauppi L, Howarth RW, McNeely J. Assessing Biofuels: Towards Sustainable Production and Use of Resources. A Report of the Working Group on Biofuels of the International Panel for Sustainable Resource Management. Paris: United Nations Environment Programme; 2009.
10. Gold M. The Global Benefits of Eating Less Meat: A Report for Compassion in World Farming Trust. New Delhi: Navodanya in collaboration with Compassion in World Farming Trust; 2004.
11. Young VR. Animal Foods, Past, Present and Future: A Nutritionist's View. In: Baldwin RL, editor. Animals, Feed, Food and People (AAAS Selected Symposia; 42). Boulder, Colorado : Westview Press, Inc.; 1980. p. 121–49.
12. McDonald P, Edwards RA, Greenhalgh JFD. Animal Nutrition. Essex: Longman Scientific & Technical; 1988.
13. Muirhead M, Alexander TJL. Managing pig health: a reference for the farm. 2nd ed. Carr J, editor. Sheffield: 5M Publishing; 2013.
14. Konig HE, Korbel R, Liebich H. Avian Anatomy. 2nd ed. Sheffield: 5m Publishing; 2016.
15. Ravindran V. Feed enzymes: The science, practice, and metabolic realities. Journal of Applied Poultry Research. 2013 Oct;22(3):628–36.
16. Menezes-Blackburn D, Gabler S, Greiner R. Performance of Seven Commercial Phytases in an in Vitro Simulation of Poultry Digestive Tract. Journal of Agricultural and Food Chemistry. 2015 Jul 2;63(27):6142–9.
17. Rodríguez JL. Physiological differences in digestion between carnivorous fishes and omnivorous fishes. Veterinaria Digital [Internet]. 2022 Nov 07; Aquaculture. Available from: https://www.veterinariadigital.com/en/articulos/physiological-differences-in-digestion-between-carnivorous-fishes-and-omnivorous-fishes/
18. Arnette SD, Simonitis LE, Egan J, Cohen KE, Kolmann MA. True grit? Comparative anatomy and evolution of gizzards in fishes. Journal of Anatomy. 2023 Sep 28; 244(2):260-273
19. Biovet S.A. Feeding and digestion in fish. Veterinaria Digital [Internet]. 2020 Mar 05; Aquaculture, digestives, preservatives. Available from: https://www.veterinariadigital.com/en/articulos/feeding-and-digestion-in-fish/

20. FAO: Natural food and feeding habits [Internet]. Available from: https://www.fao.org/fishery/affris/species-profiles/rainbow-trout/natural-food-and-feeding-habits/en/
21. Phillips AMJr. Nutrition, digestion and energy utilization. In: Hoar WS, Randall DJ, editors. Fish physiology, Volume 1. New York: Academic Press; 1969. p. 391–432.
22. Raymond WF. The Nutritive Value of Herbage. In: Abrams JT, editor. Recent Advances in Animal Nutrition. London: J. & A. Churchill LTD; 1966. p. 81–118.
23. Lewis D. Aspects of Poultry Nutrition. In: Abrams JT, editor. Recent Advances in Animal Nutrition. London: J. & A. Churchill LTD; 1966. p. 119–142.
24. Lall SP. Nutrition and health of fish. Cruz -Suárez LE, Ricque-Marie D, Tapia-Salazar M, Olvera-Novoa MA, Civera-Cerecedo R, editors. Avances en Nutrición Acuícola. 2019;(2000: Memorias del Quinto Simposium Internacional de Nutrición Acuícola).
25. Cosgrove DJ. Inositol Phosphates: Their Chemistry, Biochemistry and Physiology. New York: Elsevier Scientific Publishing Company; 1980.
26. Ravindran V, Ravindran G, Sivalogan S. Total and phytate phosphorus contents of various foods and feedstuffs of plant origin. Food Chemistry. 1994 Jan;50(2):133–6.
27. Bohn L, Meyer AS, Rasmussen SK. Phytate: impact on environment and human nutrition. A challenge for molecular breeding. Journal of Zhejiang University SCIENCE B. 2008 Mar;9(3):165–91.
28. Mittal A, Gupta V, Singh G, Yadav A, Aggarwal NK. Phytase: A Boom in Food Industry. Octa Journal of Biosciences. 2013 Dec;1(2):158–69.
29. Turner BL, Papházy MJ, Haygarth PM, Mckelvie ID. Inositol phosphates in the environment. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2002 Apr 29;357(1420):449–69.
30. Parthasarathy R, Eisenberg F. Biochemistry, Stereochemistry and Nomenclature. In: Reitz AB, editor. Inositol Phosphates and Derivatives: Synthesis, Biochemistry, and Therapeutic Potential. Washington D.C: American Chemical Society; 1991. p. 1–19.
31. Siddiqui N, Singh V, Deshmukh MM, Gurunath R. Structures, stability and hydrogen bonding in inositol conformers. Physical Chemistry Chemical Physics. 2015 Jun 10;17(28):18514–23.
32. Murthy P. Structure and Nomenclature of Inositol Phosphates, Phosphoinositides, and Glycosylphosphatidylinositols. In: Majumder AL, Biswas BB, editors. Biology of Inositols and Phosphoinositides Subcellular Biochemistry, vol 39. Boston, MA: Springer; 2006. p. 1–19.
33. Agranoff BW. Turtles All the Way: Reflections on myo-Inositol. Journal of Biological Chemistry. 2009 Aug 7;284(32):21121–6.
34. Agranoff BW. Textbook Errors: Cyclitol Confusion. Trends in Biochemical Sciences. 1978;3:283–5.
35. Irvine RF, Schell MJ. Back in the water: the return of the inositol phosphates. Nature Reviews Molecular Cell Biology. 2001 May;2:327–38.
36. Bauman AT, Chateauneuf GM, Boyd BR, Brown RE, Murthy PPN. Conformational inversion processes in phytic acid: NMR spectroscopic and molecular modeling studies. Tetrahedron Letters. 1999 Jun;40(24):4489–92.
37. Turner BL, Mahieu N, Condron LM. Quantification of Myo-inositol Hexakisphosphate in Alkaline Soil Extracts by Solution 31P NMR Spectroscopy and Spectral Deconvolution. Soil Science. 2003 Jul;168(7):469–78.
38. Anderson RJ. PHYTIN AND PHOSPHORIC ACID ESTERS OF INOSITE. Journal of Biological Chemistry. 1912 Jun;11(5):471–88.
39. Posternak S. Sur un nouveau principe phosphor-organique d origine vegetale, la phytine. Comptes Rendus des Seances de la Societe de Biologie et de Ses Filiales. 1903;55:1190–2.
40. Posternak S. Sur la synthese de l'ether hexaphosphorique de l'inosite avec le principe phospho-organique de reserve des plantes vertes. Comptes rendus de l'Académie des Sciences. 1919;169:138–40.
41. Kumar V, Sinha AK, Makkar HPS, Becker K. Dietary roles of phytate and phytase in human nutrition: A review. Food Chemistry. 2010 Jun;120(4):945–59.
42. Humer E, Schwarz C, Schedle K. Phytate in pig and poultry nutrition. Journal of Animal Physiology and Animal Nutrition. 2014 Nov 18;99(4):605–25.
43. Gupta RK, Gangoliya SS, Singh NK. Reduction of phytic acid and enhancement of bioavailable micronutrients in food grains. Journal of Food Science and Technology. 2013 Apr 24;52(2):676–84.

44. Vats P, Banerjee UC. Production studies and catalytic properties of phytases (myo-inositolhexakisphosphate phosphohydrolases): an overview. Enzyme and Microbial Technology. 2004 Jul;35(1):3–14.
45. Ellis R, Kelsay JL, Reynolds RD, Morris ER, Moser PB, Frazier CW. Phytate:zinc and phytate X calcium:zinc millimolar ratios in self-selected diets of Americans, Asian Indians, and Nepalese. Journal of the American Dietetic Association. 1987 Aug;87(8):1043–7.
46. Abbasi F, Fakhur-un-Nisa T, Liu J, Luo X, Abbasi IHR. Low digestibility of phytate phosphorus, their impacts on the environment, and phytase opportunity in the poultry industry. Environmental Science and Pollution Research. 2019 Feb 20;26(10):9469–79.
47. Lei XG, Porres JM. Phytase and inositol phosphates in animal nutrition: dietary manipulation and phosphorus excretion by animals. In: Turner BL, Richardson AE, Mullaney EJ, editors. Inositol Phosphates: Linking Agriculture and the Environment. CABI eBooks; 2006. p. 133–49.
48. Sugiura SH. Phosphorus, Aquaculture, and the Environment. Reviews in Fisheries Science & Aquaculture. 2018 May 24;26(4):515–21.
49. Srivastava BN, Biswas TD, Das NB. Influence of fertilizer and manures on the content of phytin and other forms of phosphorus in wheat and their relation to soil phosphorus. Journal of the Indian Society of Soil Science. 1955 Aug 1;3(1):33–40.
50. Department for Environment, Food & Rural Affairs. United Kingdom Food Security Report 2024. London, UK: Department for Environment, Food & Rural Affairs; 2024 Dec. Available from: https://www.gov.uk/government/statistics/united-kingdom-food-security-report-2024
51. Rubio LA, Molina E. Las leguminosas en alimentación animal. Arbor. 2016 Jun 30;192(779):a315.
52. BSE feed controls | Department of Agriculture, Environment and Rural Affairs [Internet]. DAERA. 2015. Available from: https://www.daera-ni.gov.uk/articles/bse-feed-controls
53. Hídvégi M, Lásztity R. Phytic acid content of cereals and legumes and interaction with proteins. Periodica Polytechnica Chemical Engineering. 2002 Jan 1;46(1).
54. Nsabimana S, Ismail T, Lazarte CE. Enhancing iron and zinc bioavailability in maize (Zea mays) through phytate reduction: the impact of fermentation alone and in combination with soaking and germination. Frontiers in Nutrition. 2024 Dec 2;11.
55. Khan N, Zaman R, Elahi M. Effect of heat treatments on the phytic acid content of maize products. Journal of the Science of Food and Agriculture. 1991 Jan 1;54(1):153–6.
56. Raboy V, Gerbasi PF, Young KA, Stoneberg SD, Pickett SG, Bauman AT, et al. Origin and Seed Phenotype of Maize low phytic acid 1-1 and low phytic acid 2-1. Plant Physiology. 2000 Sep 1;124(1):355–68.
57. Nelson TS, Shieh TR, Wodzinski RJ, Ware JH. The Availability of Phytate Phosphorus in Soybean Meal Before and After Treatment With a Mold Phytase. Poultry Science. 1968 Nov;47(6):1842–8.
58. Powers WJ, Fritz ER, Fehr W, Angel R. Total and water-soluble phosphorus excretion from swine fed low-phytate soybeans. Journal of Animal Science. 2006 Jul 1;84(7):1907–15.
59. Oltmans SE, Fehr WR, Welke GA, Raboy V, Peterson KL. Agronomic and Seed Traits of Soybean Lines with Low-Phytate Phosphorus. Crop Science. 2005 Mar;45(2):593–8.
60. Nutrient Requirements of Swine: *10th Revised Edition*. Washington, D.C.: National Academies Press; 1998.
61. Humer E, Schwarz C, Schedle K. Phytate in pig and poultry nutrition. Journal of Animal Physiology and Animal Nutrition. 2014 Nov 18;99(4):605–25.
62. Barrow NJ, Lambourne LJ. Partition of excreted nitrogen, sulphur, and phosphorus between the faeces and urine of sheep being fed pasture. Australian Journal of Agricultural Research. 1962 Jan;13(3):461-71.
63. Nävert B, Sandström B, Cederblad A. Reduction of the phytate content of bran by leavening in bread and its effect on zinc absorption in man. British Journal of Nutrition. 1985 Jan;53(1):47–53.
64. Raboy V. Seeds for a better future: "low phytate" grains help to overcome malnutrition and reduce pollution. Trends in Plant Science. 2001 Oct;6(10):458–62.
65. Zhao Y, Zhu L, Lin C, Shen Z, Xu C. Transgenic soybean expressing a thermostable phytase as substitution for feed additive phytase. Scientific Reports. 2019 Oct 7;9(14390).
66. Golovan SP, Hayes MA, Phillips JP, Forsberg CW. Transgenic mice expressing bacterial phytase as a model for phosphorus pollution control. Nature Biotechnology. 2001 May;19(5):429–33.
67. Kirby AJ, Varvoglis AG. The Reactivity of Phosphate Esters. Monoester Hydrolysis. Journal of the American Chemical Society. 1967 Jan;89(2):415–23.
68. Westheimer FH. Why nature chose phosphates. Science. 1987 Mar 6;235(4793):1173–8.

69. Mullaney EJ, Ullah AHJ. Phytases: Attributes, Catalytic Mechanisms and Application. In: Turner BL, Richardson AE, Mullaney EJ, editors. Inositol Phosphates: Linking Agriculture and the Environment. CABI eBooks; 2006. p. 97-110.
70. Dersjant-Li Y, Awati A, Schulze H, Partridge G. Phytase in non-ruminant animal nutrition: a critical review on phytase activities in the gastrointestinal tract and influencing factors. Journal of the Science of Food and Agriculture. 2014 Dec 19;95(5):878–96.
71. Kumar V, Sinha A. General aspects of phytases. In: Kumar V, Nunes C, editors. Enzymes in Human and Animal Nutrition. 2018. p. 53–72.
72. Sandberg A. Bioavailability of minerals in legumes. British Journal of Nutrition. 2002 Dec;88(S3):281–5.
73. Mullaney EJ, Ullah AHJ. The term phytase comprises several different classes of enzymes. Biochemical and Biophysical Research Communications. 2003 Dec 5;312(1):179–84.
74. Greiner R. Phytate-degrading Enzymes: Regulation of Synthesis in Microorganisms and Plants. In: Turner BL, Richardson AE, Mullaney EJ, editors. Inositol Phosphates: Linking Agriculture and the Environment. CABI eBooks; 2006. p. 78-96.
75. Numbering of atoms in myo-inositol. Recommendations 1988. Nomenclature Committee of the International Union of Biochemistry. Biochemical Journal. 1989 Feb 15;258(1):1–2.
76. Barrientos L, Scott JJ, Murthy PP. Specificity of Hydrolysis of Phytic Acid by Alkaline Phytase from Lily Pollen. Plant Physiology. 1994 Dec;106(4):1489–95.
77. Puhl AA, Greiner R, Selinger LB. A protein tyrosine phosphatase-like inositol polyphosphatase from Selenomonas ruminantium subsp. lactilytica has specificity for the 5-phosphate of myo-inositol hexakisphosphate. The International Journal of Biochemistry & Cell Biology. 2008;40(10):2053–64.
78. Shin S, Ha NC, Oh BC, Oh TK, Oh BH. Enzyme Mechanism and Catalytic Property of β Propeller Phytase. Structure. 2001 Sep;9(9):851–8.
79. Rigden DJ. The histidine phosphatase superfamily: structure and function. Biochemical Journal. 2008 Jan 15;409(2):333–48.
80. Balaban NP, Suleimanova AD, Valeeva LR, Shakirov EV, Sharipova MR. Structural characteristics and catalytic mechanism of Bacillus β-propeller phytases. Biochemistry (Moscow). 2016 Aug;81(8):785–93.
81. Kumar V, Yadav AN, Verma P, Sangwan P, Saxena A, Kumar K, et al. β-Propeller phytases: Diversity, catalytic attributes, current developments and potential biotechnological applications. International Journal of Biological Macromolecules. 2017 May;98:595–609.
82. Greiner R, Lim BL, Cheng C, Carlsson NG. Pathway of phytate dephosphorylation by β-propeller phytases of different origins. Canadian Journal of Microbiology. 2007 Apr;53(4):488–95.
83. Tran HT, Hurley B, Plaxton WC. Feeding hungry plants: The role of purple acid phosphatases in phosphate nutrition. Plant Science. 2010 Jul;179(1):14–27.
84. Schenk G, Mitić N, Hanson GR, Comba P. Purple acid phosphatase: A journey into the function and mechanism of a colorful enzyme. Coordination Chemistry Reviews. 2013 Jan;257(2):473–82.
85. Zhu H, Qian W, Lu X, Li D, Liu X, Liu K, et al. Expression Patterns of Purple Acid Phosphatase Genes in Arabidopsis Organs and Functional Analysis of AtPAP23 Predominantly Transcribed in Flower. Plant Molecular Biology. 2005 Nov;59(4):581–94.
86. Rizwanuddin S, Kumar V, Singh P, Naik B, Mishra S, Chauhan M, et al. Insight into phytase-producing microorganisms for phytate solubilization and soil sustainability. Frontiers in Microbiology. 2023 Apr 11;14:1127249.
87. Yanke LJ, Selinger LB, Cheng K-J. Phytase activity of Selenomonas ruminantium: a preliminary characterization. Letters in Applied Microbiology. 1999 Jul 1;29(1):20–5.
88. Li Z, Huang H, Zhao H, Meng K, Zhao J, Shi P, et al. Genetic diversity and expression profiles of cysteine phytases in the sheep rumen during a feeding cycle. Letters in Applied Microbiology. 2014 Dec 1;59(6):615–20.
89. Kumar V, Agrawal S. Short Communication: An insight into protein sequences of PTP-like cysteine phytases. Nusantara Bioscience. 2014 May;6(1):102–6.
90. Chen CC, Cheng KJ, Ko TP, Guo RT. Current Progresses in Phytase Research: Three-Dimensional Structure and Protein Engineering. ChemBioEng Reviews. 2015 Jan 9;2(2):76–86.
91. Chu HM, Guo RT, Lin TW, Chou CC, Shr HL, Lai HL, et al. Structures of Selenomonas ruminantium Phytase in Complex with Persulfated Phytate. Structure. 2004 Nov;12(11):2015–24.

92. Stentz R, Osborne S, Horn N, Li Arthur WH, Hautefort I, Bongaerts R, et al. A Bacterial Homolog of a Eukaryotic Inositol Phosphate Signaling Enzyme Mediates Cross-kingdom Dialog in the Mammalian Gut. Cell Reports. 2014 Feb;6(4):646–56.
93. Craxton A, Caffrey JJ, Burkhart W, Safrany TS, Shears BS. Molecular cloning and expression of a rat hepatic multiple inositol polyphosphate phosphatase. Biochemical Journal. 1997 Nov 15;328(1):75–81.
94. Dijken PV, de Haas JR, Craxton A, Erneux C, Shears SB, Van Haastert PJ. A Novel, Phospholipase C-independent Pathway of Inositol 1,4,5-Trisphosphate Formation in Dictyostelium and Rat Liver. Journal of Biological Chemistry. 1995 Dec 15;270(50):29724–31.
95. Cho J, Choi K, Darden T, Reynolds PR, Petitte JN, Shears SB. Avian multiple inositol polyphosphate phosphatase is an active phytase that can be engineered to help ameliorate the planet's "phosphate crisis." Journal of Biotechnology. 2006 Nov;126(2):248–59.
96. Dionisio G, Holm PB, Brinch-Pedersen H. Wheat (Triticum aestivum L.) and barley (Hordeum vulgare L.) multiple inositol polyphosphate phosphatases (MINPPs) are phytases expressed during grain filling and germination. Plant Biotechnology Journal. 2007 Feb 19;5(2):325–38.
97. Romano PR, Wang J, O'Keefe RJ, Puzas JE, Rosier RN, Reynolds PR. HiPER1, a phosphatase of the endoplasmic reticulum with a role in chondrocyte maturation. Journal of Cell Science. 1998 Mar 15;111(6):803–13.
98. Chi H, Tiller GE, Dasouki MJ, Romano PR, Wang J, O'Keefe RJ, et al. Multiple Inositol Polyphosphate Phosphatase: Evolution as a Distinct Group within the Histidine Phosphatase Family and Chromosomal Localization of the Human and Mouse Genes to Chromosomes 10q23 and 19. Genomics. 1999 Mar 15;56(3):324–36.
99. Peng X, Li H, Xu W, Yang Q, Li D, Fan T, et al. The AtMINPP Gene, Encoding a Multiple Inositol Polyphosphate Phosphatase, Coordinates a Novel Crosstalk between Phytic Acid Metabolism and Ethylene Signal Transduction in Leaf Senescence. International Journal of Molecular Sciences. 2024 Aug 17;25(16):8969.
100. Acquistapace IM, Zietek MA, Li AWH, Salmon M, Kühn I, Bedford MR, et al. Snapshots during the catalytic cycle of a histidine acid phytase reveal an induced-fit structural mechanism. Journal of Biological Chemistry. 2020 Dec 18;295(51):17724–37.
101. Tamayo-Ramos JA, Sanz-Penella JM, Yebra MJ, Monedero V, Haros M. Novel Phytases from Bifidobacterium pseudocatenulatum ATCC 27919 and Bifidobacterium longum subsp.infantisATCC 15697. Applied and Environmental Microbiology. 2012 May 11;78(14):5013–5.
102. Haros M, Carlsson NG, Almgren A, Larsson-Alminger M, Sandberg AS, Andlid T. Phytate degradation by human gut isolated Bifidobacterium pseudocatenulatum ATCC27919 and its probiotic potential. International Journal of Food Microbiology. 2009 Sep 30;135(1):7–14.
103. Mehta BD, Jog SP, Johnson SC, Murthy PPN. Lily pollen alkaline phytase is a histidine phosphatase similar to mammalian multiple inositol polyphosphate phosphatase (MINPP). Phytochemistry. 2006 Sep;67(17):1874–86.
104. Rix GD, Sprigg C, Whitfield H, Hemmings AM, Todd JD, Brearley CA. Characterisation of a soil MINPP phytase with remarkable long-term stability and activity from Acinetobacter sp. PLOS ONE. 2022 Aug 31;17(8):e0272015.
105. McCollum EV, Hart EB. ON THE OCCURRENCE OF A PHYTIN-SPLITTING ENZYME IN ANIMAL TISSUES. Journal of Biological Chemistry. 1908 Jun 1;4(6):497–500.
106. Suzuki U, Yoshimura K, Takaishi M. Bulletins of the College of Agriculture vii. Tokyo; 1907 p. 503–12.
107. Gessler NN, Serdyuk EG, Isakova EP, Deryabina YI. Phytases and the Prospects for Their Application (Review). 2018 Jul 13;54(4):352–60.
108. Dox AW, Golden R. PHYTASE IN LOWER FUNGI. Journal of Biological Chemistry. 1911 Oct;10(3):183–6.
109. Jackman R, Black C. HYDROLYSIS OF PHYTATE PHOSPHORUS IN SOILS. Soil Science. 1952 Mar;73(3):167–72.
110. Casida LJ. PHOSPHATASE ACTIVITY OF SOME COMMON SOIL FUNGI. Soil Science. 1959 May;87(5):305–10.
111. Greaves MP, Anderson G, Webley DM. The hydrolysis of inositol phosphates by aerobacter aerogenes. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 1967 Mar 1;132(2):412–8.
112. Cosgrove D, Irving G, Bromfield S. INOSITOL PHOSPHATE PHOSPHATASES OF MICROBIOLOGICAL ORIGIN.* THE ISOLATION OF SOIL BACTERIA HAVING INOSITOL PHOSPHATE PHOSPHATASE ACTIVITY . Australian Journal of Biological Sciences. 1969 Sep 17;23:339–43.

113. Cosgrove D. INOSITOL PHOSPHATE PHOSPHATASES OF MICROBIOLOGICAL ORIGIN. INOSITOL PHOSPHATE INTERMEDIATES IN THE DEPHOSPHORYLATION OF THE HEXAPHOSPHATES OF MYO-INOSITOL, SCYLLO-INOSITOL, AND D-CHIRO-INOSITOL BY A BACTERIAL (PSEUDOMONAS SP.) PHYTASE*. Australian Journal of Biological Sciences. 1970 Apr 20;23:1207–20.
114. Shieh TR, Wodzinski RJ, Ware JH. Regulation of the Formation of Acid Phosphatases by Inorganic Phosphate in *Aspergillus ficuum*. Journal of Bacteriology. 1969 Dec 1;100(3):1161–5.
115. Irving, Cosgrove DJ. Inositol Phosphate Phosphatases of Microbiological Origin: the Inositol Pentaphosphate Products of *Aspergillus ficuum* Phytases. Journal of Bacteriology. 1972 Oct 1;112(1):434–8.
116. Lei XG, Porres JM, Mullaney EJ, Brinch-Pedersen H. Phytase: Source, Structure and Application. Industrial Enzymes. :505–29.
117. Rizwanuddin S, Kumar V, Naik B, Singh P, Mishra S, Rustagi S, et al. Microbial phytase: Their sources, production, and role in the enhancement of nutritional aspects of food and feed additives. Journal of Agriculture and Food Research. 2023 Mar 24;12(1):100559.
118. Greiner R, Konietzny U. Phytases: Biochemistry, Enzymology and Characteristics Relevant to Animal Feed Use. In: Bedford MR, Partridge GG, editors. Enzymes in Farm Animal Nutrition. CAB International; 2010. p. 96–128.
119. Nelson TS. The Utilization of Phytate Phosphorus by Poultry—A Review. Poultry Science. 1967 Jul 1;46(4):862–71.
120. Choct M. Enzymes for the feed industry: past, present and future. World's Poultry Science Journal. 2006 Mar 1;62(1):5–16.
121. Bedford MR. The evolution and application of enzymes in the animal feed industry: the role of data interpretation. British Poultry Science. 2018 Jul 27;59(5):486–93.
122. Khattak FM, Pasha TN, Mahmud A, Hayat Z. Enzymes in poultry nutrition. Journal of Animal and Plant Sciences. 2006 Jan;16(1).
123. Slominski BA. Recent advances in research on enzymes for poultry diets. Poultry Science. 2011 Sep;90(9):2013–23.
124. Aranda-Aguirre E, Robles-Jimenez LE, Osorio-Avalos J, Vargas-Bello-Pérez E, Gonzalez-Ronquillo M. A systematic-review on the role of exogenous enzymes on the productive performance at weaning, growing and finishing in pigs. Veterinary and Animal Science. 2021 Dec;14:100195.
125. C. Sheppy. The current feed enzyme market and likely trends. CABI Publishing eBooks. 2001 Jan 1;1–10.
126. Global Feed Enzymes Market By Product Type (Phytase, Carbohydrase, Protease), By Form (Liquid, Dry), By Livestock (Ruminants, Poultry, Swine, Aquatic Animals), By Source (Microorganism, Plant, Animal), By Geographic Scope and Forecast [Internet]. Washington D.C: Verified Market Research; 2024 Nov. Available from: https://www.verifiedmarketresearch.com/product/feed-enzymes-market/
127. Guerrand D. Economics of food and feed enzymes: Status and prospectives. In: Nunes C, Kumar V, editors. Enzymes in Human and Animal Nutrition. London: Academic Press; 2018. p. 487–514.
128. Junior DTV, Genova JL, Kim SW, Saraiva A, Rocha GC. Carbohydrases and Phytase in Poultry and Pig Nutrition: A Review beyond the Nutrients and Energy Matrix. Animals (Basel). 2024 Jan 11;14(2):226.
129. Lei XG, Weaver JD, Mullaney E, Ullah AH, Azain MJ. Phytase, a New Life for an "Old" Enzyme. Annual Review of Animal Biosciences. 2013 Jan;1(1):283–309.
130. Animal Feed Enzymes Market Size, Share & Trends Analysis Report By Product (Phytases, Proteases) By Formulation, By Application (Poultry, Pigs), By Region And Segment Forecasts, 2024 - 2030 [Internet]. San Francisco, CA: Grand View Research; 2024 Sep. Available from: https://www.grandviewresearch.com/industry-analysis/animal-feed-enzymes-industry#:~:text=Poultry%20secured%20the%20largest%20share,shifts%20in%20consumer%20dietary%20preferences.
131. Coffey D, Dawson KA, Ferket P, Connolly AJ. Review of the feed industry from a historical perspective and implications for its future. Journal of Applied Animal Nutrition. 2016;4(e3).
132. Global Aquafeed Enzymes Market to Exceed a Valuation of USD 804.5 million by 2034: Future Market Insights, Inc. GlobeNewswire [Internet]. 2024 Sep 5; Available from: https://www.globenewswire.com/news-release/2024/09/06/2941916/0/en/Global-Aquafeed-Enzymes-Market-to-Exceed-a-Valuation-of-USD-804-5-million-by-2034-Future-Market-Insights-Inc.html

133. Feed Enzymes Market 2024 – By Type (Protease, Phytase, Carbohydrase, Other Types), By Form (Dry, Liquid), By Livestock (Swine, Poultry, Ruminants, Aquatic Animals, Other Livestock), By Source (Microorganism, Plant, Animal), And By Region, Opportunities And Strategies – Global Forecast To 2033 [Internet]. London: The Business Research Company; 2024 Nov. Available from: https://www.thebusinessresearchcompany.com/report/feed-enzymes-market
134. More A. Phytase Market. Bhat S, editor. Ontario, CA: Dataintelo; 2024. Available from: https://dataintelo.com/report/global-phytase-market
135. Animal Feed Phytase Market Size, Share & Trends Analysis Report By Type (Granular Phytases, Powder Phytases, Liquid Phytases, Thermostable Phytases), By Applications (Swine, Poultry , Ruminants, Aquatic Animals, Others) and By Region(North America, Europe, APAC, Middle East and Africa, LATAM) Forecasts, 2024-2032. India: Straits Research; 2024 Apr. Available from: https://straitsresearch.com/report/animal-feed-phytase-market
136. Animal Feed Phytase Market Size - By Form (Liquid, & Granules/Powder), Livestock (Poultry (Broiler, Layer, Turkey), Swine (Starter, Grower, Sow), Cattle (Dairy, Calf), Aquaculture (Salmon, Trout, Shrimp), Regional Outlook & Global Forecast, 2024 – 2032 [Internet]. Delaware, USA: Global Market Insights; 2024 Jan. Available from: https://www.gminsights.com/industry-analysis/animal-feed-phytase-market
137. Kumar M. Feed Phytase Market Size, Share, Growth, Trends, and Global Industry Analysis: Product (Histidine Acid Phosphatases (HAPs), B propeller Phytases, Purple Acid Phosphatases, Protein Tyrosine Phosphatase like Phytases, and Others), By Animal (Swine, Poultry, Ruminants, Aquatic Animals, and Others), and Region Forecast 2019-2030. India: Precision Business Insights. Available from: https://www.precisionbusinessinsights.com/market-reports/feed-phytase-market
138. Adeola O, Cowieson AJ. BOARD-INVITED REVIEW: Opportunities and challenges in using exogenous enzymes to improve nonruminant animal production. Journal of Animal Science. 2011 Oct 1;89(10):3189–218.
139. Clickner FH, Follwell EH. Application of “Protozyme” (Aspergillus orizae) to Poultry Feeding. Poultry Science. 1926 Jun;5(5):241–7.
140. Warden WK, Schaible PJ. Action of Antibiotics in Stimulating Growth of Poultry. Poultry Science. 1962 May;41(3):725–32.
141. FAO. The State of World Fisheries and Aquaculture 2022. Rome: FAO; 2022.
142. Tacon AGJ, Metian M. Feed Matters: Satisfying the Feed Demand of Aquaculture. Reviews in Fisheries Science & Aquaculture. 2015 Jan 2;23(1):1–10.
143. IAFFD [Internet]. Iaffd.com. 2021 Available from: https://www.iaffd.com/aquaoverview.html
144. Lehane S. Fish for the Future: Aquaculture and Food Security. Dalkeith, Western Australia: World Health Organisation; 2013.
145. OptiPhos® | Huvepharma [Internet]. Huvepharma.com. 2025. Available from: https://www.huvepharma.com/products/product/optiphos/
146. Naylor RL, Hardy RW, Buschmann AH, Bush SR, Cao L, Klinger DH, et al. A 20-year retrospective review of global aquaculture. Nature. 2021 Mar 24;591(7851):551–63.
147. Gonzalez-Ortiz G, Walk C. Protease supplementation in commercial swine diets. Pig333 [Internet]. 2019; Available from: https://www.pig333.com/articles/protease-supplementation-in-commercial-swine-diets_14728/
148. Nguyen DH, Park JW, Kim IH. Effect of crumbled diet on growth performance, market day age and meat quality of growing-finishing pigs. Journal of Applied Animal Research. 2016 Jul 13;45(1):396–9.
149. Thacker PA. Alternatives to antibiotics as growth promoters for use in swine production: a review. Journal of Animal Science and Biotechnology. 2013 Sep 14;4(35).
150. Jang KB, Kim YI, Duarte ME, Kim SW. Effects of β-mannanase supplementation on intestinal health and growth of nursery pigs. Journal of Animal Science. 2024 Jan 3;102.
151. Torres-Pitarch A, Hermans D, Manzanilla EG, Bindelle J, Everaert N, Beckers Y, et al. Effect of feed enzymes on digestibility and growth in weaned pigs: A systematic review and meta-analysis. Animal Feed Science and Technology. 2017 Nov;233:145–59.
152. Torres-Pitarch A, Manzanilla EG, Gardiner GE, O’Doherty JV, Lawlor PG. Systematic review and meta-analysis of the effect of feed enzymes on growth and nutrient digestibility in grow-finisher pigs: Effect of enzyme type and cereal source. Animal Feed Science and Technology. 2019 May;251:153–65.

153. Jacela JY, DeRouchey JM, Tokach MD, Goodband RD, Nelssen JL, Renter DG, et al. Feed additives for swine: Fact sheets – carcass modifiers, carbohydrate-degrading enzymes and proteases, and anthelmintics. Kansas Agricultural Experiment Station Research Reports. 2009 Jan 1;(10):325–32.
154. Towers L. Understanding True Impact of Non-Starch Polysaccharides in Poultry and Swine Diets [Internet]. Thepigsite.com. 2015 Mar 10. Available from: https://www.thepigsite.com/articles/understanding-true-impact-of-nonstarch-polysaccharides-in-poultry-and-swine-diets
155. da Silva CA, Callegari MA, Dias CP, Bridi AM, Pierozan CR, Foppa L, et al. Increasing doses of phytase from Citrobacter braakii in diets with reduced inorganic phosphorus and calcium improve growth performance and lean meat of growing and finishing pigs. Loor JJ, editor. PLOS ONE. 2019 May 24;14(5):e0217490.
156. Greiner R, Konietzny U. Update on Characteristics of Commercial Phytases. 2nd International Phytase Summit. 2015 Jan 1; Available from: https://www.researchgate.net/publication/322301882_Update_on_Characteristics_of_Commercial_Phytas
157. Greiner R, Bedford M. Phytase analysis, pitfalls and interpretation of FTU for efficacy in the animal. In: Proceedings of the 1st International Phytase Summit. Washington D. C; 2010.
158. Acquistapace IM, Thompson EJ, Kühn I, Bedford MR, Brearley CA, Hemmings AM. Insights to the Structural Basis for the Stereospecificity of the Escherichia coli Phytase, AppA. International Journal of Molecular Sciences. 2022 Jun 6;23(11):6346.
159. K Ostanin, Harms EH, P.E. Stevis, R Kuciel, Zhou MM, Etten RLV. Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase. Journal of Biological Chemistry. 1992 Nov 1;267(32):22830–6.
160. Boquet PL, C Manoil, Beckwith J. Use of TnphoA to detect genes for exported proteins in Escherichia coli: identification of the plasmid-encoded gene for a periplasmic acid phosphatase. Journal of Bacteriology. 1987 Apr 1;169(4):1663–9.
161. Feedinfo - 20 Years of Phytase Optimisation, The DuPont Story - INDUSTRY PERSPECTIVES [Internet]. Feedinfo.com. 2021. Available from: https://www.feedinfo.com/our-content/20-years-of-phytase-optimisation-the-dupont-story-industry-perspectives/219474
162. Garrett JB, Kretz KA, O'Donoghue E, Kerovuo J, Kim W, Barton NR, et al. Enhancing the Thermal Tolerance and Gastric Performance of a Microbial Phytase for Use as a Phosphate-Mobilizing Monogastric-Feed Supplement. Applied and Environmental Microbiology. 2004 May;70(5):3041–6.
163. Navone L, Vogl T, Luangthongkam P, Blinco JA, Luna-Flores CH, Chen X, et al. Disulfide bond engineering of AppA phytase for increased thermostability requires co-expression of protein disulfide isomerase in Pichia pastoris. Biotechnology for Biofuels. 2021 Mar 31;14(1).
164. Greiner R, Carlsson NG, Alminger ML. Stereospecificity of myo-inositol hexakisphosphate dephosphorylation by a phytate-degrading enzyme of Escherichia coli. Journal of Biotechnology. 2000 Nov;84(1):53–62.
165. Konietzny U, Greiner R. Molecular and catalytic properties of phytate-degrading enzymes (phytases). International Journal of Food Science and Technology. 2002 Oct;37(7):791–812.
166. Greiner R, U. Konietzny, Jany KD. Purification and Characterization of Two Phytases from Escherichia coli. Archives of Biochemistry and Biophysics. 1993 May 1;303(1):107–13.
167. Herrmann KR, Brethauer C, Siedhoff NE, Hofmann I, Eyll J, Davari MD, et al. Evolution of E. coli Phytase Toward Improved Hydrolysis of Inositol Tetraphosphate. Frontiers in Chemical Engineering. 2022 Mar 9;4.
168. Muñoz-Muñoz PLA, Terán-Ramírez C, Mares-Alejandre RE, Márquez-González AB, Madero-Ayala PA, Meléndez-López SG, et al. Surface Engineering of Escherichia coli to Display Its Phytase (AppA) and Functional Analysis of Enzyme Activities. Current Issues in Molecular Biology. 2024 Apr 17;46(4):3424–37.
169. Natuphos ® Internet]. Available from: https://images.proultry.com/files/company/4143/BASF_Animal_Nutrition_Enzyme_Natuphos.pdf
170. ALLZYME ® SSF Product Specifications [Internet]. Available from: https://madbarn.com/wp-content/uploads/2023/06/Allzyme-SSF-Alltech-Guaranteed-Analysis.pdf?srsltid=AfmBOooL1Dzv3f0ZjzEXmoR2b1PgsToeE3UKmxUOfKtbJsZVL_hBi5Kz
171. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Assessment of the application for renewal of authorisation of PHYZYME® XP 5000 G/L (6-phytase) for

chickens for fattening, laying hens, turkeys for fattening, ducks for fattening, weaned piglets, pigs for fattening and sows for reproduction. EFSA Journal. 2019 May 1;17(5).
172. Official Journal of the European Union. COMMISSION REGULATION (EC) No 1141/2007 of 1 October 2007 concerning the authorisation of 3-phytase (ROVABIO PHY AP and ROVABIO PHY LC) as a feed additive. 2007.
173. Safety and efficacy of the product Quantum™ Phytase 5000 L and Quantum™ Phytase 2500 D (6-phytase) as a feed additive for chickens for fattening, laying hens, turkeys for fattening, ducks for fattening and piglets (weaned) - Scientific Opinion of the Panel on Additives and Products or Substances used in Animal Feed and the Panel on Genetically Modified Organisms. EFSA Journal. 2008 Feb 8;6(2):627.
174. Safety and efficacy of RONOZYME® NP (6-phytase) as a feed additive for pigs for fattening. EFSA Journal. 2016 Feb 1;14(2).
175. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Safety and efficacy of a feed additive consisting of 6-phytase (produced by Komagataella phaffii DSM 23036) (OptiPhos®) for poultry for fattening, chickens reared for laying, laying hens, turkeys reared for breeding, weaned piglets, pigs for fattening and sows for the renewal of their authorisation and for the new use in breeding hens and turkeys, ornamental birds, suckling piglets and minor pig species for fattening and breeding (Huvepharma EOOd). EFSA Journal. 2022 May 5;20(5).
176. FEEDAP. Safety and efficacy of OPTIPHOS® (6-phytase) as a feed additive for finfish. EFSA Journal. 2017 Apr 1;15(4).
177. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Assessment of the application for renewal of authorisation of 6-phytase produced by Trichoderma reeseiCBS 122001 as a feed additive for pigs and poultry, from Roal Oy. EFSA Journal. 2020 Dec 15;18(12).
178. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Safety and efficacy of the feed additive consisting of 6-phytase (produced by Aspergillus oryzae DSM 33699) (RONOZYME® Hiphos GT/L) for poultry, pigs for fattening, weaned piglets and sows (DSM Nutritional Products Ltd). EFSA Journal. 2023 Jan 11;21(1).
179. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Scientific Opinion on the safety and efficacy of Quantum® Blue (6-phytase) as a feed additive for laying hens and minor laying poultry species. EFSA Journal. 2013 Oct 30;11(10).
180. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Safety and efficacy of Axtra® PHY 20000 TPT2 (6-phytase) as a feed additive for poultry and porcine species. EFSA Journal. 2016 Nov 14;14(11).
181. BASF. The new phytase molecule [Internet]. Available from: https://agroktinotrofiki.gr/wp-content/uploads/2023/10/NPE-introduction-2020-BASF_-BTC.pdf
182. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Safety and efficacy of a feed additive consisting of 6-phytase produced by Trichoderma reesei CBS 146250 (Axtra® PHY GOLD 30L, Axtra® PHY GOLD 30T, Axtra® PHY GOLD 65G) for all poultry species and all pigs (Danisco (UK) ltd). EFSA Journal. 2022 Nov 8;20(11).
183. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Safety and efficacy of the feed additive consisting of 6-phytase (produced by Komagataella phaffii CGMCC 7.19) (Nutrase P) for chickens for fattening, other poultry for fattening or reared for laying and ornamental birds (Nutrex N.V.). EFSA Journal. 2022 Jun 7;20(6).
184. REF FOR ROVABIO HERE
185. Thorsen M, Nielsen LA, Zhai HX, Zhang Q, Wulf-Andersen L, Skov LK. Safety and efficacy profile of a phytase produced by fermentation and used as a feed additive. Heliyon. 2021 Jun 4;7(6):e07237–7.
186. EFSA Panel on Additives and Products or Substances used in Animal Feed (EFSA FEEDAP Panel). Safety and efficacy of a feed additive consisting of 6-phytase produced by Komagataella phaffii CGMCC 7.370 (VTR-phytase powder/liquid) for all pigs and all avian species (Victory Enzymes GmbH). EFSA Journal. 2022 Dec 20;20(12).
187. Urgessa OE, Koyamo R, Dinka H, Tefese K, Gemeda MT. Review on Desirable Microbial Phytases as a Poultry Feed Additive: Their Sources, Production, Enzymatic Evaluation, Market Size, and Regulation. International Journal of Microbiology. 2024 Jun 7;2024(1).

188. Zeller E, Schollenberger M, Kühn I, Rodehutscord M. Hydrolysis of phytate and formation of inositol phosphate isomers without or with supplemented phytases in different segments of the digestive tract of broilers. Journal of Nutritional Science. 2015 Jan 26;4(e1).
189. Escobin-Mopera L, Ohtani M, Sekiguchi S, Sone T, Abe A, Tanaka M, et al. Purification and characterization of phytase from Klebsiella pneumoniae 9-3B. Journal of Bioscience and Bioengineering. 2012 May;113(5):562–7.
190. Sherman K, Sone T, Asano K. 2P-038 A Histidine Acid Phytase Homodimer from Klebsiella pneumoniae 9-3B that Liberates myoinositol from Phytic Acid. In: Japan Society of Bioengineering Conference Abstracts. Japan: Japan Society of Bioengineering; 2014. p. 116.
191. Ragon M, Aumelas A, Chemardin P, Galvez S, Moulin G, Boze H. Complete hydrolysis of myo-inositol hexakisphosphate by a novel phytase from Debaryomyces castellii CBS 2923. Applied Microbiology and Biotechnology. 2008 Feb 1;78(1):47–53.
192. Ragon M, Neugnot-Roux V, Chemardin P, Moulin G, Boze H. Molecular gene cloning and overexpression of the phytase from Debaryomyces castellii CBS 2923. Protein Expression and Purification. 2008 Apr;58(2):275–83.
193. Ragon M, Hoh F, Aumelas A, Chiche L, Moulin G, Boze H. Structure of *Debaryomyces castellii* CBS 2923 phytase. Acta crystallographica Section F: Structural Biology and Crystallization Communications. 2009 Mar 25;65(4):321–6.
194. Boze H, Moulin G, Aumelas A, inventors; ADISSEO FRANCE S.A.S., assignee. Debaryomyces castellii phytase. United States patent US 8,003,360 B2. 2011-08-23.
195. Segueilha L, Lambrechts C, Boze H, Moulin G, Galzy P. Purification and properties of the phytase from Schwanniomyces castellii. Journal of Fermentation and Bioengineering. 1992 Jan;74(1):7–11.
196. Mochizuki D, Tokuda J, Suzuki T, Shimada M, Tawaki S, inventors; Mitsui Toatsu Chemicals, Inc., assignee. Phytase. United States patent US 5,830,732.1998-11-03.
197. Koontz L. Agarose Gel Electrophoresis. Methods in Enzymology. 2013;529:35–45.
198. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose Gel Electrophoresis for the Separation of DNA Fragments. Journal of Visualized Experiments. 2012 Apr 20;3923(62).
199. Borst P. Ethidium DNA agarose gel electrophoresis: How it started. IUBMB Life (International Union of Biochemistry and Molecular Biology: Life). 2005 Nov 1;57(11):745–7.
200. Huang QFW, Baum L, Fu WL. Simple and practical staining of DNA with GelRed in agarose gel electrophoresis. Clinical Laboratory. 2010;56(3-4):149–52.
201. Dragan AI, Pavlovic R, McGivney JB, Casas-Finet JR, Bishop ES, Strouse RJ, et al. SYBR Green I: fluorescence properties and interaction with DNA. Journal of fluorescence. 2012 Apr 26;22(4):1189–99.
202. Evenson WE, Boden LM, Muzikar KA, O'Leary DJ. 1H and 13C NMR Assignments for the Cyanine Dyes SYBR Safe and Thiazole Orange. The Journal of Organic Chemistry. 2012 Nov 8;77(23):10967–71.
203. Olmsted J, Kearns DR. Mechanism of ethidium bromide fluorescence enhancement on binding to nucleic acids. Biochemistry. 1977 Aug 9;16(16):3647–54.
204. Stellwagen NC, Gelfi C, Righetti PG. DNA and buffers: The hidden danger of complex formation. Biopolymers. 2000 Jun 13;54(2):137–42.
205. Smith BJ. SDS Polyacrylamide Gel Electrophoresis of Proteins. Methods in Molecular Biology. 1984;1:41–55.
206. Brunelle JL, Green R. One-dimensional SDS-Polyacrylamide Gel Electrophoresis (1D SDS-PAGE). Methods in Enzymology. 2014 Mar 25;541:151–9.
207. Kielkopf CL, Bauer W, Urbatsch IL. Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis of Proteins. Cold Spring Harbor Protocols. 2021 Dec 1;2021(12).
208. Bornhorst JA, Falke JJ. Purification of Proteins Using Polyhistidine Affinity Tags. Methods in Enzymology. 2000;326:245–54.
209. Pharmacia LKB Biotechnology. Gel Filtration: Principles and Methods. 5th ed. Uppsala, Sweden: Amersham Pharmacia Biotech; 1998.
210. Cytiva. Size Exclusion Chromatography: Principles and Methods. Cytiva Life Sciences; 2020.
211. Nagul EA, McKelvie ID, Worsfold P, Kolev SD. The molybdenum blue reaction for the determination of orthophosphate revisited: Opening the black box. Analytica Chimica Acta. 2015 Aug;890:60–82.
212. Kanno S, Cuyas L, Javot H, Bligny R, Gout E, Dartevelle T, et al. Performance and Limitations of Phosphate Quantification: Guidelines for Plant Biologists. Plant and Cell Physiology. 2016 Feb 9;57(4):690–706.

213. Woods J, Mellon M. Molybdenum Blue Reaction: A Spectrophotometric Study. Industrial & Engineering Chemistry Analytical Edition. 1941 Nov 1;13(11):760–4.
214. Fiske CH, Subbarow Y. THE COLORIMETRIC DETERMINATION OF PHOSPHORUS. Journal of Biological Chemistry. 1925 Dec;66(2):375–400.
215. Katewa SD, Katyare SS. A simplified method for inorganic phosphate determination and its application for phosphate analysis in enzyme assays. Analytical Biochemistry. 2003 Dec;323(2):180–7.
216. Simoni RD, Hill RL, Vaughan M. The Determination of Phosphorus and the Discovery of Phosphocreatine and ATP: the Work of Fiske and SubbaRow. Journal of Biological Chemistry. 2002 Aug;277(32):e1–2.
217. McKie VA, McCleary BV. A Novel and Rapid Colorimetric Method for Measuring Total Phosphorus and Phytic Acid in Foods and Animal Feeds. Journal of AOAC INTERNATIONAL. 2016 May 1;99(3):738–43.
218. Pradhan S, Pokhrel MR. Spectrophotometric Determination of Phosphate in Sugarcane Juice, Fertilizer, Detergent and Water Samples by Molybdenum Blue Method. Scientific World. 2013 Nov 11;11(11):58–62.
219. He Z, Honeycutt CW. A Modified Molybdenum Blue Method for Orthophosphate Determination Suitable for Investigating Enzymatic Hydrolysis of Organic Phosphates. Communications in Soil Science and Plant Analysis. 2005 May;36(9-10):1373–83.
220. Heinonen J, Lahti R. A new and convenient colorimetric determination of inorganic orthophosphate and its application to the assay of inorganic pyrophosphatase. 1981 May 15;113(2):313–7.
221. Qvirist L, Carlsson NG, Andlid T. Assessing phytase activity–methods, definitions and pitfalls. Journal of Biological Methods. 2015 Jun 4;2(1):16.
222. Dick WA, Tabatabai MA. Determination of Orthophosphate in Aqueous Solutions Containing Labile Organic and Inorganic Phosphorus Compounds. Journal of Environmental Quality. 1977 Jan;6(1):82–5.
223. Carlsson, N, et al. Rapid Analyzes of Inositol Phosphates. Chalmers University of Technology. 2000. Doctoral Thesis.
224. Holman WIM. A new technique for the determination of phosphorus by the molybdenum blue method. Biochemical Journal. 1943 Jul 1;37(2):256–9.
225. Peterson GL. A simplified method for analysis of inorganic phosphate in the presence of interfering substances. Analytical Biochemistry. 1978 Jan;84(1):164–72.
226. Bae HD, Yanke LJ, Cheng KJ, Selinger LB. A novel staining method for detecting phytase activity. Journal of Microbiological Methods. 1999 Dec;39(1):17–22.
227. Barker CJ, Illies C, Berggren PO. HPLC Separation of Inositol Polyphosphates. Methods in molecular biology. 2010;645:21–46.
228. Rounds MA, Nielsen SS. Anion-exchange high-performance liquid chromatography with post- column detection for the analysis of phytic acid and other inositol phosphates. Journal of Chromatography A. 1993 Oct 29;653(1):148–52.
229. Smith DH, Clark FE. Chromatographic Separations of Inositol Phosphorus Compounds. Soil Science Society of America Journal. 1952 Apr;16(2):170–2.
230. Irvine RF, Änggård E, Letcher AJ, Downes CP. Metabolism of inositol 1,4,5-trisphosphate and inositol 1,3,4-trisphosphate in rat parotid glands. Biochemical Journal. 1985 Jul 15;229(2):505–11.
231. Phillippy BQ, Bland JM. Gradient ion chromatography of inositol phosphates. Analytical Biochemistry. 1988 Dec;175(1):162–6.
232. Mayr GW. A novel metal-dye detection system permits picomolar-range h.p.l.c. analysis of inositol polyphosphates from non-radioactively labelled cell or tissue specimens. Biochemical Journal. 1988 Sep 1;254(2):585–91.
233. Skoglund E, Carlsson NG, Sandberg AS. Determination of Isomers of Inositol Mono- to Hexaphosphates in Selected Foods and Intestinal Contents Using High-Performance Ion Chromatography. Journal of agricultural and food chemistry. 1997 Feb 17;45(2):431–6.
234. Chen QC, Li BW. Separation of phytic acid and other related inositol phosphates by high- performance ion chromatography and its applications. Journal of Chromatography A. 2003 Nov 7;1018(1):41–52.
235. Wade HE, Morgan DM. Fractionation of phosphates by paper ionophoresis and chromatography. Biochemical Journal. 1955 Jun;60(2):264–70.
236. Marolt G, Kolar M. Analytical Methods for Determination of Phytic Acid and Other Inositol Phosphates: A Review. Molecules. 2020 Dec 31;26(1):174.
237. Powell HR. A beginner's guide to X-ray data processing. The Biochemist. 2021 May 28;43(3).

238. Wukovitz SW, Yeates TO. Why protein crystals favour some space-groups over others. Nature Structural & Molecular Biology. 1995 Dec;2(12):1062–7.
239. McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallographica Section F Structural Biology Communications. 2013 Dec 24;70(1):2–20.
240. Myerson AS, Erdemir D, Lee AY. Crystal Nucleation. In: Handbook of Industrial Crystallization. Cambridge University Press; 2019. p. 76–114.
241. Rhodes G. Crystallography Made Crystal Clear. San Diego: Academic Press; 1993.
242. Khamskii E. CRYSTALLIZATION FROM SOLUTIONS AND MELTS. Springer US; 1969.
243. Winter G, McAuley KE. Automated data collection for macromolecular crystallography. Methods. 2011 Sep;55(1):81–93.
244. Kabsch W. *The rotation method in crystallography* edited by U. W. Arndt and A. J. Wonacott. Acta crystallographica Section B, Structural crystallography and crystal chemistry. 1978 Mar 1;34(3):1049–50.
245. Powell HR. X-ray data processing. Bioscience Reports. 2017 Oct 6;37(5).
246. Winter G. xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography. 2010;43:186–90.
247. Zwart PH, Grosse-Kunstleve RW, Adams PD. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation .
248. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS Journal. 2007 Nov 23;275(1):1–21.
249. Dubach VRA, Guskov A. The Resolution in X-ray Crystallography and Single-Particle Cryogenic Electron Microscopy. Crystals. 2020 Jul 5;10(7):580.
250. Karplus PA, Diederichs K. Linking Crystallographic Model and Data Quality. Science. 2012 May 24;336(6084):1030–3.
251. Karplus PA, Diederichs K. Assessing and maximizing data quality in macromolecular crystallography. Current Opinion in Structural Biology. 2015 Jul 24;34:60–8.
252. Nosti AJ, Barrio LC, Celis FC, Soldado A, Encinar JR. Absolute quantification of proteins using element mass spectrometry and generic standards. Journal of Proteomics. 2022 Mar 30;256:104499.
253. Wilschefski SC, Baxter MR. Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects. Clinical Biochemist Reviews. 2019 Aug 28;40(3):115–33.
254. Walk CL, Kühn I, Stein HH, Kidd MT, Rodehutscord M. Phytate Destruction - Consequences for Precision Animal Nutrition. Wageningen Academic Publishers; 2016.
255. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, et al. A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. Science. 2003 Mar 28;299(5615):2074–6.
256. Mishra S, Imlay JA. An anaerobic bacterium, Bacteroides thetaiotaomicron, uses a consortium of enzymes to scavenge hydrogen peroxide. Molecular Microbiology. 2013 Oct 24;90(6):1356–71.
257. Li AWH. Structure-function Studies of Multiple Inositol Polyphosphate Phosphatases from Gut Commensal Bacteria [PhD Thesis]. [University of East Anglia]; 2014. p. 58–102.
258. Salmon M. Unpublished data.
259. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnology. 2008;8(1):91.
260. Grantham R. Amino Acid Difference Formula to Help Explain Protein Evolution. Science (New York, NY). 1974 Sep 6;185(4154):862–4.
261. Yampolsky LY, Stoltzfus A. The Exchangeability of Amino Acids in Proteins. Genetics. 2005 Aug 1;170(4):1459–72.
262. Sneath PHA. Relations between Chemical Structure and Biological Activity in Peptides. Journal of Theoretical Biology. 1966 Nov;12(2):157–95.
263. EPSTEIN CJ. Non-randomness of Ammo-acid Changes in the Evolution of Homologous Proteins. Nature. 1967 Jul 22;215(5099):355–9.
264. Miyata T, Miyazawa S, Yasunaga T. Two types of amino acid substitutions in protein evolution. Journal of Molecular Evolution. 1979 Mar;12(3):219–36.
265. Betts MJ, Russell RB. Amino Acid Properties and Consequences of Substitutions. In: Bioinformatics for Geneticists. Wiley; 2003. p. 289–316.

266. Perez-Garcia P, Kobus S, Gertzen CGW, Hoeppner A, Holzscheck N, Strunk CH, et al. A promiscuous ancestral enzyme´s structure unveils protein variable regions of the highly diverse metallo-β-lactamase family. Communications biology. 2021 Jan 29;4(132).
267. Villamizar GAC, Funkner K, Nacke H, Foerster K, Daniel R. Functional Metagenomics Reveals a New Catalytic Domain, the Metallo-β-Lactamase Superfamily Domain, Associated with Phytase Activity. mSphere. 2019 Jun 19;4(3):e00167-19.
268. Bush K, Bradford PA. β-Lactams and β-Lactamase Inhibitors: An Overview. Cold Spring Harbor Perspectives in Medicine. 2016 Aug 1;6(8):a025247.
269. Yun Y, Han S, Park YS, Park H, Kim D, Kim Y, et al. Structural Insights for Core Scaffold and Substrate Specificity of B1, B2, and B3 Metallo-β-Lactamases. Frontiers in Microbiology. 2022 Jan 13;12.
270. Bush K. Past and Present Perspectives on β-Lactamases. Antimicrobial Agents and Chemotherapy. 2018 Sep 24;62(10):e01076-18.
271. Krco S, Davis SJ, Joshi P, Wilson LA, Pedroso MM, Douw A, et al. Structure, function, and evolution of metallo-β-lactamases from the B3 subgroup—emerging targets to combat antibiotic resistance. Frontiers in Chemistry. 2023 Jun 20;11:1196073.
272. Palzkill T. Metallo-β-lactamase structure and function. Annals of the New York Academy of Sciences. 2012 Nov 16;1277:91–104.
273. Fröhlich C, Chen JZ, Gholipour S, Erdogan AN, Tokuriki N. Evolution of β-lactamases and enzyme promiscuity. Protein engineering, design & selection: PEDS. 2021 May 3;34:1–12.
274. Karsisiotis AI, Damblon CF, Roberts GCK. A variety of roles for versatile zinc in metallo-β-lactamases. Metallomics : integrated biometal science. 2014 Mar 26;6(7):1181–97.
275. Hall BG, Barlow M. Revised Ambler classification of β-lactamases. Journal of Antimicrobial Chemotherapy. 2005 Jun 1;55(6):1050–1.
276. Bush K, Jacoby GA. Updated Functional Classification of -Lactamases. Antimicrobial Agents and Chemotherapy. 2009 Dec 7;54(3):969–76.
277. Diene SM, Pontarotti P, Azza S, Armstrong N, Pinault L, Chabrière E, et al. Origin, Diversity, and Multiple Roles of Enzymes with Metallo-β-Lactamase Fold from Different Organisms. Cells. 2023 Jun 30;12(13):1752.
278. Daiyasu H, Osaka K, Ishino Y, Toh H. Expansion of the zinc metallo-hydrolase family of the β- lactamase fold. FEBS Letters. 2001 Aug 14;503(1):1–6.
279. Sabath LD, Abraham EP. Zinc as a Cofactor for Cephalosporinase from Bacillus cereus 569. Biochemical Journal. 1966 Jan;98(1):11C13C.
280. Rodríguez MM, Herman R, Ghiglione B, Kerff F, González GD, Bouillenne F, et al. Crystal structure and kinetic analysis of the class B3 di-zinc metallo-β-lactamase LRA-12 from an Alaskan soil metagenome. PLoS ONE. 2017 Jul 27;12(7):e0182043.
281. Galleni M, Lamotte-Brasseur J, Rossolini GM, Spencer J, Dideberg O, FrèreJM, et al. Standard Numbering Scheme for Class B β-Lactamases. Antimicrobial Agents and Chemotherapy. 2001 Mar;45(3):660–3.
282. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, et al. Characterization of a New Metallo- β-Lactamase Gene, blaNDM-1, and a Novel Erythromycin Esterase Gene Carried on a Unique Genetic Structure in Klebsiella pneumoniae Sequence Type 14 from India. Antimicrobial Agents and Chemotherapy. 2009 Sep 21;53(12):5046–54.
283. Osano E, Arakawa Y, Wacharotayankun R, Ohta M, Horii T, Ito H, et al. Molecular characterization of an enterobacterial metallo beta-lactamase found in a clinical isolate of Serratia marcescens that shows imipenem resistance. Antimicrobial Agents and Chemotherapy. 1994 Jan;38(1):71–8.
284. Poirel L, Naas T, Nicolas D, Collet L, Bellais S, Cavallo JD, et al. Characterization of VIM-2, a Carbapenem-Hydrolyzing Metallo-β-Lactamase and Its Plasmid- and Integron-Borne Gene from a Pseudomonas aeruginosa Clinical Isolate in France. Antimicrobial Agents and Chemotherapy. 2000 Apr;44(4):891–7.
285. Bebrone C. Metallo-beta-lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily. Biochemical pharmacology. 2007 Jun 2;74(12):1686–701.
286. Hernandez Valladares M, Felici A, Weber G, Adolph HW, Zeppezauer M, Rossolini GM, et al. Zn(II) Dependence of the Aeromonas hydrophila AE036 Metallo-β-lactamase Activity and Stability. Biochemistry. 1997 Sep 23;36(38):11534–41.
287. Yong D, Toleman MA, Bell J, Ritchie B, Pratt R, Ryley H, et al. Genetic and Biochemical Characterization of an Acquired Subgroup B3 Metallo-β-Lactamase Gene,blaAIM-1, and Its Unique Genetic Context in

Pseudomonas aeruginosa from Australia. Antimicrobial Agents and Chemotherapy. 2012 Sep 17;56(12):6154–9.
288. Wachino J, Yoshida H, Yamane K, Suzuki S, Matsui M, Yamagishi T, et al. SMB-1, a Novel Subclass B3 Metallo-β-Lactamase, Associated with ISCR1and a Class 1 Integron, from a Carbapenem-Resistant Serratia marcescens Clinical Isolate. Antimicrobial Agents and Chemotherapy. 2011 Aug 29;55(11):5143–9.
289. López C, Delmonti J, Bonomo RA, Vila AJ. Deciphering the evolution of metallo-β-lactamases: A journey from the test tube to the bacterial periplasm. Journal of Biological Chemistry. 2022 Mar;298(3):101665.
290. Pedroso MM, Waite DW, Melse O, Wilson L, Mitić N, McGeary RP, et al. Broad spectrum antibiotic-degrading metallo-β-lactamases are phylogenetically diverse. Protein & Cell. 2020 Jun 15;11(8):613–7.
291. Park YS, Kim TY, Park H, Lee JH, Nguyen DQ, Hong MK, et al. Structural Study of Metal Binding and Coordination in Ancient Metallo-β-Lactamase PNGM-1 Variants. International journal of molecular sciences. 2020 Jul 12;21(14):4926.
292. Jacquin O, Balbeur D, Damblon C, Marchot P, De Pauw E, Roberts GCK, et al. Positively Cooperative Binding of Zinc Ions to Bacillus cereus 569/H/9 β-Lactamase II Suggests that the Binuclear Enzyme Is the Only Relevant Form for Catalysis. Journal of Molecular Biology. 2009 Oct 9;392(5):1278–91.
293. Orellano EG, Girardini JE, Cricco JA, Ceccarelli EA, Vila AJ. Spectroscopic Characterization of a Binuclear Metal Site in Bacillus cereus β-Lactamase II. Biochemistry. 1998 Jul 14;37(28):10173–80.
294. Wang Z, Fast W, Valentine AM, Benkovic SJ. Metallo-β-lactamase: structure and mechanism. Current Opinion in Chemical Biology. 1999 Oct;3(5):614–22.
295. Wommer S, Rival S, Heinz U, Galleni M, Frère JM, Franceschini N, et al. Substrate-activated Zinc Binding of Metallo-β-lactamases. Journal of Biological Chemistry. 2002 Apr 19;277(27):24142–7.
296. Baier F, Tokuriki N. Connectivity between Catalytic Landscapes of the Metallo-β-Lactamase Superfamily. Journal of Molecular Biology. 2014 Jun 26;426(13):2442–56.
297. Lee JH, Takahashi M, Jeon JH, Kang LW, Seki M, Park KS, et al. Dual activity of PNGM-1 pinpoints the evolutionary origin of subclass B3 metallo-β-lactamases: a molecular and evolutionary study. Emerging Microbes & Infections. 2019 Nov 21;8(1):1688–700.
298. Copley SD. Shining a light on enzyme promiscuity. Current Opinion in Structural Biology. 2017 Dec;47:167–75.
299. Wheeler LC, Lim SA, Marqusee S, Harms MJ. The thermostability and specificity of ancient proteins. Current Opinion in Structural Biology. 2016 Jun 9;38:37–43.
300. Park HS, Nam SH, Lee JK, Yoon CN, Mannervik B, Benkovic SJ, et al. Design and Evolution of New Catalytic Activity with an Existing Protein Scaffold. Science. 2006 Jan 27;311(5760):535–8.
301. PDB: 6J4N
302. LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the E. coli Cytoplasm. Nature Biotechnology. 1993 Feb 1;11(2):187–93.
303. LaVallie ER, DiBlasio-Smith EA, Collins-Racie LA, Lu Z, McCoy JM. Thioredoxin and Related Proteins as Multifunctional Fusion Tags for Soluble Expression in E. coli. In: Vaillancourt PE, editor. E coli Gene Expression Protocols Methods in Molecular Biology. Humana Press; 2003. p. 119–40.
304. McCoy J, LaVille E. Expression and Purification of Thioredoxin Fusion Proteins. Current Protocols in Protein Science. 1997 May 1;10(1).
305. Zhijian L, DiBlasio-Smith EA, Grant KL, Warne NW, LaVallie ER, Collins-Racie LA, et al. Histidine Patch Thioredoxins:MUTANT FORMS OF THIOREDOXIN WITH METAL CHELATING AFFINITY THAT PROVIDE FOR CONVENIENT PURIFICATIONS OF THIOREDOXIN FUSION PROTEINS . Journal of Biological Chemistry. 1996 Mar 1;271(9):5059–65.
306. Kitamoto Y, Yuan X, Wu Q, McCourt DW, Sadler JE. Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. Proceedings of the National Academy of Sciences of the United States of America. 1994 Aug 2;91(16):7588–92.
307. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. Journal of Bacteriology. 1995 Jul;177(14):4121– 30.
308. Levy MS, Lotfian P, O'Kennedy R, Lo-Yim MY, Shamlou PA. Quantitation of supercoiled circular content in plasmid DNA solutions using a fluorescence-based method. Nucleic Acids Research. 2000 Jun 15;28(12):e57.

309. Calladine CR, Collis CM, Drew HR, Mott MR. A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. Journal of Molecular Biology. 1991 Jun 20;221(3):981–1005.
310. Yau SY, Keshavarz-Moore E, Ward J. Host strain influences on supercoiled plasmid DNA production inEscherichia coli: Implications for efficient design of large-scale processes. Biotechnology and Bioengineering. 2008 Apr 3;101(3):529–44.
311. Casali N. Escherichia coli Host Strains. In: Casali N, Preston A, editors. E coli Plasmid Vectors Methods in Molecular Biology. Humana Press; 2003. p. 27–48.
312. Khlebnikov A, Risa O, Skaug T, Carrier TA, Keasling JD. Regulatable Arabinose-Inducible Gene Expression System with Consistent Control in All Cells of a Culture. Journal of Bacteriology. 2000 Dec 15;182(24):7029–34.
313. Széliová D, Krahulec J, Šafránek M, Lišková V, Turňa J. Modulation of heterologous expression from P BAD promoter in Escherichia coli production strains. Journal of Biotechnology. 2016 Oct 20;236:1–9.
314. Siegele DA, Hu JC. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. Proceedings of the National Academy of Sciences. 1997 Jul 22;94(15):8168–72.
315. Gucwa M, Lenkiewicz J, Zheng H, Cymborowski M, Cooper DR, Murzyn K, et al. CMM—An enhanced platform for interactive validation of metal binding sites. Protein science. 2023 Jan 1;32(1):e4525.
316. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Research. 2018 Jul 2;46(W1):W363–7.
317. Lu CH, Chen CC, Yu CS, Liu YY, Liu JJ, Wei ST, et al. MIB2: metal ion-binding site prediction and modeling server. Martelli PL, editor. Bioinformatics. 2022 Jul 29;38(18):4428–9.
318. Holm L, Sander C. Dali: a network tool for protein structure comparison. Trends in Biochemical Sciences. 1995 Nov;20(11):478–80.
319. Dong P, Yang W, Sun L, Jing D, Zhang H, Yang J, et al. Characterization and Structural Insights of a Novel Arylsulfatase from Pseudoalteromonas atlantica T6c. Catalysts. 2023 Apr 15;13(4):759.
320. Park KS, Kim TY, Kim JH, Lee JH, Jeon JH, Karim AM, et al. PNGM-1, a novel subclass B3 metallo-β-lactamase from a deep-sea sediment metagenome. Journal of Global Antimicrobial Resistance. 2018 Sep;14:302–5.
321. Park KS, Hong MK, Jeon JW, Kim JH, Jeon JH, Lee JH, et al. The novel metallo-β-lactamase PNGM-1 from a deep-sea sediment metagenome: crystallization and X-ray crystallographic analysis. Acta crystallographica Section F, Structural biology communications. 2018 Sep 19;74(10):644–9.
322. Villamizar GAC. MβLp01. Personal communication. 2021.
323. Mehlhoff JD, Stearns FW, Rohm D, Wang B, Tsou EY, Dutta N, et al. Collateral fitness effects of mutations. Proceedings of the National Academy of Sciences. 2020 May 8;117(21):11597–607.
324. Petty KJ. Metal-chelate affinity chromatography. In: Current Protocols in Molecular Biology. John Wiley & Sons Ltd; 1996.
325. Effect of pH Values on the Formation and Solubility of Zinc Compounds. Bulletin of the Institute for Chemical Research, Kyoto University. 1978 Dec;56(5):242–6.
326. Reed P, Richey D, Roseboom D. Acute Toxicity of Zinc to Some Fishes in High Alkalinity Water . 1980;4.
327. Krężel A, Maret W. The biological inorganic chemistry of zinc ions. Archives of Biochemistry and Biophysics. 2016 Dec 1;611:3–19.
328. Socha RD, Chen J, Tokuriki N. The Molecular Mechanisms Underlying Hidden Phenotypic Variation among Metallo-β-Lactamases. Journal of Molecular Biology. 2019 Feb 13;431(6):1172–85.
329. Vogel A, Schilling O, Niecke M, Bettmer J, Meyer-Klaucke W. ElaC Encodes a Novel Binuclear Zinc Phosphodiesterase. Journal of Biological Chemistry. 2002 Aug 9;277(32):29078–85.
330. Kostelecky B, Pohl E, Vogel A, Schilling O, Meyer-Klaucke W. The Crystal Structure of the Zinc Phosphodiesterase from Escherichia coli Provides Insight into Function and Cooperativity of tRNase Z-Family Proteins. Journal of Bacteriology. 2006 Feb;188(4):1607–14.
331. Späth B, Canino G, Marchfelder A. tRNase Z: the end is not in sight. Cellular and Molecular Life Sciences. 2007 Jun 29;64:2404–12.
332. EUCAST. Determination of minimum inhibitory concentrations (MICs) of antibacterial agents by broth dilution. Clinical Microbiology and Infection. 2003 Aug;9(8):ix–xv.
333. Shannon K, Phillips I. β-Lactamase detection by three simple methods: Intralactam, nitrocefin and acidimetric. Journal of Antimicrobial Chemotherapy. 1980 Sep 1;6(5):617–21.
334. Kuo SS, Feng TY. Iodometric method for detection of β-lactamase activity in yeast cells carrying ampicillin resistance gene in chimeric plasmids. Analytical Biochemistry. 1989 Feb 15;177(1):165–7.

335. Sinicropi D, Baker DL, Prince WS, Shiffer K, Shak S. Colorimetric determination of DNase I activity with a DNA-methyl green substrate. Analytical Biochemistry. 1994 Nov 1;222(2):351–8.
336. Lee SY, Lévesque SA, Sévigny J, Müller CE. A highly sensitive capillary electrophoresis method using p-nitrophenyl 5′-thymidine monophosphate as a substrate for the monitoring of nucleotide pyrophosphatase/phosphodiesterase activities. Journal of chromatography B. 2012 Dec 12;911:162–9.
337. Naas T, Oueslati S, Bonnin RA, Dabos ML, Zavala A, Dortet L, et al. Beta-lactamase database (BLDB) – structure and function. Journal of Enzyme Inhibition and Medicinal Chemistry. 2017 Jul 19;32(1):917–9.
338. Carfi A, Pares S, Duee E, Galleni M, Duez C, Frère JM, et al. The 3-D structure of a zinc metallo-beta-lactamase from Bacillus cereus reveals a new type of protein fold. EMBO. 1995 Oct 16;14(20):4914–21.
339. Zong G, Shears SB, Wang H. Structural and catalytic analyses of the InsP 6 kinase activities of higher plant ITPKs. The FASEB Journal. 2022 Jul 30;36(7).
340. Caddick SEK, Harrison CJ, Stavridou I, Mitchell JL, Hemmings AM, Brearley CA. A Solanum tuberosum inositol phosphate kinase (StITPK1) displaying inositol phosphate-inositol phosphate and inositol phosphate-ADP phosphotransferase activities. FEBS Letters. 2008 Apr 28;582(12):1731–7.
341. Miller GJ, Wilson MP, Majerus PW, Hurley JH. Specificity Determinants in Inositol Polyphosphate Synthesis: Crystal Structure of Inositol 1,3,4-Trisphosphate 5/6-Kinase. Molecular Cell. 2005 Apr;18(2):201–12.
342. Laha D, Portela-Torres P, Desfougères Y, Saiardi A. Inositol phosphate kinases in the eukaryote landscape. Advances in Biological Regulation. 2021 Jan;79:100782.
343. Chamberlain PP, Qian X, Stiles AR, Cho J, Jones DH, Lesley SA, et al. Integration of Inositol Phosphate Signaling Pathways via Human ITPK1. Journal of Biological Chemistry. 2007 Jul 6;282(38):28117–25.
344. Majerus PW, Wilson DB, Zhang C, Nicholas PJ, Wilson MP. Expression of inositol 1,3,4-trisphosphate 5/6-kinase (ITPK1) and its role in neural tube defects. Advances in Enzyme Regulation. 2010;50(1):365–72.
345. Laha D, Parvin N, Hofer A, Giehl RFH, Fernandez-Rebollo N, von Wirén N, et al. Arabidopsis ITPK1 and ITPK2 Have an Evolutionarily Conserved Phytic Acid Kinase Activity. ACS Chemical Biology. 2019 Sep 16;14(10):2127–33.
346. Shears SB, Wang H. Inositol phosphate kinases: Expanding the biological significance of the universal core of the protein kinase fold. Advances in Biological Regulation. 2019 Jan;71:118–27.
347. Cridland C, Gillaspy G. Inositol Pyrophosphate Pathways and Mechanisms: What Can We Learn from Plants? Molecules. 2020 Jun 17;25(12):2789.
348. Sweetman D, Stavridou I, Johnson S, Green P, Caddick SEK, Brearley CA. Arabidopsis thalianainositol 1,3,4-trisphosphate 5/6-kinase 4 (AtITPK4) is an outlier to a family of ATP-grasp fold proteins from Arabidopsis. FEBS Letters. 2007 Jul 30;581(22):4165–71.
349. Xiong T, Zhang Z, Fan T, Ye F, Ye Z. Origin, evolution, and diversification of inositol 1,4,5- trisphosphate 3-kinases in plants and animals. BMC genomics. 2024 Apr 8;25(1).
350. Zhang C, Majerus PW, Wilson MP. Regulation of inositol 1,3,4-trisphosphate 5/6-kinase (ITPK1) by reversible lysine acetylation. Proceedings of the National Academy of Sciences of the United States of America. 2012 Jan 30;109(7):2290–5.
351. Shears SB, Schenk TMH, Zhou Y. ITPK1 (Inositol 1,3,4-Triphosphate 5/6 Kinase). Choi S, editor. Springer eBooks. 2012;984–90.
352. Ho MWY, Yang X, Carew MA, Zhang T, Hua L, Kwon YU, et al. Regulation of Ins(3,4,5,6)P4 Signaling by a Reversible Kinase/Phosphatase. Current Biology. 2002 Mar 19;12(6):477–82.
353. Whitfield HL, Rodriguez RF, Shipton ML, Li AWH, Riley AM, Potter BVL, et al. Crystal Structure and Enzymology of Solanum tuberosum Inositol Tris/Tetrakisphosphate Kinase 1 (StITPK1). Biochemistry. 2023 Dec 26;63(1):42–52.
354. Saiardi A, Cockcroft S. Human ITPK1: A Reversible Inositol Phosphate Kinase/Phosphatase That Links Receptor-Dependent Phospholipase C to Ca2+-Activated Chloride Channels. Science Signaling. 2008 Jan 29;1(4):pe5–5.
355. Whitfield HL, He S, Gu Y, Sprigg C, Kuo HF, Chiou TJ, et al. Diversification in the inositol tris/tetrakisphosphate kinase (ITPK) family: crystal structure and enzymology of the outlier AtITPK4. Biochemical Journal. 2023 Mar 10;480(6):433–53.
356. Laha NP, Giehl RFH, Riemer E, Qiu D, Pullagurla NJ, Schneider R, et al. INOSITOL (1,3,4) TRIPHOSPHATE 5/6 KINASE1-dependent inositol polyphosphates regulate auxin responses in Arabidopsis. Plant Physiology. 2022 Sep 19;190(4):2722–38.

357. Rodriguez RF. Engineering Metallophytases for the Animal Feed Industry [Research Report]. [University of East Anglia]; 2015. p. 25–39.
358. Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Research. 2007 Feb 1;35(6):e45.
359. McPherson A. Science in Pictures: Macromolecular Crystals. Scientific American. 1989 Mar;260(3):62–9.
360. McPherson A, Cudney B. Optimization of crystallization conditions for biological macromolecules. Acta Crystallographica Section F Structural Biology Communications. 2014 Oct 31;70(11):1445–67.
361. Cudney B. Protein Crystallisation and Dumb Luck. Rigaku Journal. 1999;16(1).
362. Bergfors TM. Protein crystallization. Third. Vol. 10. La Jolla, Calif.: International University Line; 2021.
363. McPherson A, Cudney B. Searching for silver bullets: An alternative strategy for crystallizing macromolecules. Journal of Structural Biology. 2006 Dec;156(3):387–406.
364. Evdokimov AG. Cryoprotection of Delicate crystals, in 2 Easy Steps (and 16 Hard steps) [Internet]. Available from: https://ccr.cancer.gov/
365. Hemmings A. Personal communication. 2022.
366. Whitfield HL, White G, Sprigg C, Riley AM, Potter BVL, Hemmings AM, et al. An ATP-responsive Metabolic Cassette Comprised of Inositol tris/tetrakisphosphate Kinase 1 (ITPK1) and Inositol Pentakisphosphate 2-kinase (IPK1) Buffers Diphosphosphoinositol Phosphate Levels. Biochemical Journal. 2020 Jul 31;477(14):2621–38.
367. Stiles AR, Qian X, Shears SB, Grabau EA. Metabolic and signaling properties of an Itpk gene family in Glycine max. FEBS Letters. 2008 May 12;582(13):1853–8.
368. Tu-Sekine B, Kim SF. The Inositol Phosphate System—A Coordinator of Metabolic Adaptability. International Journal of Molecular Sciences. 2022 Jun 16;23(12):6747.