Allosteric Regulation of the First Enzyme in Histidine Biosynthesis

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science with Honours in Chemistry

In the Department of Chemistry
At the University of Canterbury

By Emma Kathrine Livingstone

2015

Abstract

The ATP-PRTase enzyme catalyses the first committed step of histidine biosynthesis in archaea, bacteria, fungi and plants.[1] As the catalyst of an energetically expensive pathway, ATP-PRTase is subject to a sophisticated, multilevel regulatory system.[2] There are two families of this enzyme, the long form ($HisG_L$) and the short form ($HisG_S$) that differ in their molecular architecture. A single $HisG_L$ chain comprises three domains. Domains I and II house the active site of $HisG_L$ while domain III, a regulatory domain, forms the binding site for histidine as an allosteric inhibitor. The long form ATP-PRTase adopts a homo-hexameric quaternary structure.[3,4] $HisG_S$ comprises a similar catalytic core to $HisG_L$ but is devoid of the regulatory domain and associates with a second protein, HisZ, to form a hetero-octameric assembly.[5]

This thesis explores the allosteric regulation of the short form ATP-PRTase, as well as the functional and evolutionary relationship between the two families. New insight into the mode allosteric inhibition of the short form ATP-PRTase from *Lactococcus lactis* is reported in chapter two. A conformational change upon histidine binding was revealed by small angle X-ray scattering, illuminating a potential mechanism for the allosteric inhibition of the enzyme. Additionally, characterisation of histidine binding to HisZ by isothermal titration calorimetry, in the presence and absence of $HisG_S$, provided evidence toward the location of the functional allosteric binding site within the HisZ subunit.

Chapter three details the extensive effort towards the purification of the short form ATP-PRTase from *Neisseria menigitidis,* the causative agent of bacterial meningitis. This enzyme is of particular interest as a potential target for novel, potent inhibitors to combat this disease. The attempts to purify the long form ATP-PRTase from *E. coli,* in order to clarify earlier research on the functional multimeric state of the enzyme, are also discussed.

Chapter four reports the investigation of a third ATP-PRTase sequence architecture, in which *hisZ* and *hisG*$_S$ comprise a single open reading frame, forming a putative fusion enzyme. The engineering of two covalent linkers between HisZ and HisG$_S$ from *L. lactis* and the transfer of the regulatory domain from HisG$_L$ to HisG$_S$, is also discussed, in an attempt to delineate the evolutionary pathway of the ATP-PRTase enzymes. Finally, the *in vivo* activity of each functional and putative ATP-PRTase was assessed by *E. coli* BW25113Δ*hisG* complementation assays.

Acknowledgements

Firstly, I would like to thank my wonderful supervisor Emily Parker. Thank you for your support and encouragement, wine and coffee. I truly appreciate your enthusiasm and tireless work ethic. I am going to miss working with you.

Secondly I would like to thank the Parker group for being such fantastic people to work with. In particular, thank you to Gerd Mittelstädt for your support, patience, and many laughs throughout this project. Thank you to Annette Steward for help with molecular biology, Nathan Cowieson, Moritz Lasse, Nicky Blackmore, and Penel Cross for assistance with SAXS experiments, Wanting Jiao for computing prowess, and Fiona Given and Gert-Jan Moggre for critiquing various drafts of this thesis. I would also like to thank Anja Stampfli and Effie Fan for their encouragement and friendship. Thank you to the Biomolecular Interaction Centre for funding and the Australian Synchrotron for travel to the Synchrotron School and for use of the SAXS/WAXS and MX beamlines.

Finally I would like to thank my family and friends, for their love and support. I am not sure how I would have got through this without you.

Contents

Abbreviations

ACT domain – regulatory domain with a βαββαβ fold

AICAR – 5'-phosphoribosyl-4-carboxamide-4-aminoimidazole

AMP – adenosine monophosphate

ATP – adenosine triphosphate

ATP-PRTase – adenosine triphosphate phosphoribosyl transferase

Cje – *Campylobacter jejuni*

***Cje*ACT** – the ACT domain from *Campylobacter jejuni*

$\boldsymbol{D_{max}}$ – maximum dimension of the particle

DNA – deoxyribonucleic acid

Eco – *Escherichia coli*

EDTA – ethylenediaminetetraacetic acid

gDNA – genomic DNA

GST – glutathione S-transferase

HisG$_{core}$ – domains I and II of HisG$_L$

HisG$_L$ – the long form ATP-PRTase

HisG$_S$ – the catalytic subunit of the short form ATP-PRTase

HisZ – the regulatory subunit of the short form ATP-PRTase

HisZG$_S$ – the short form ATP-PRTase

HisZG$_{fusion}$ – a putative ATP-PRTase with a covalently linked HisZ and HisG$_S$

IAP – imidazole acetol phosphate

IGP – imidazole glycerol phosphate

IMAC – immobilised metal affinity chromatography

IPTG – isopropyl β-D-1-thiogalactopyranoside

ITC – isothermal titration calorimetry

$\boldsymbol{k_{cat}}$ – turnover number

$\boldsymbol{K_D}$ – dissociation constant

K_i – inhibition constant

K_M – Michaelis constant

LB – Lysogeny broth

Lla – *Lactococcus lactis*

Lme – *Leuconostoc mesenteroides*

MES – 2-(N-morpholino) ethanesulfonic acid

MilliQ – water purified with a Millipore MilliQ system

MOPS – 2-(N-morpholino) propanesulfonic acid

Mtu – *Mycobacterium tuberculosis*

Nme – *Neisseria meningitidis*

NMR spectroscopy – nuclear magnetic resonance spectroscopy

PCR – polymerase chain reaction

PDB – Protein Data Bank

PPase – inorganic pyrophosphatase

PR-AMP – 5'-phosphoribosyl-AMP

PR-ATP – 5'-phosphoribosyl-ATP

Pro-FAR – *N'*-[(5'-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide

PRPP – 5-phosphoribosyl 1-pyrophosphate

PRTase – phosphoribosyl transferase

RMSD – root mean square deviation

RNA – ribonucleic acid

R_g – radius of gyration

SANS – small angle neutron scattering

SAXS – small angle X-ray scattering

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC – size exclusion chromatography

SOC – Super optimal broth with catabolite repression

TAE – Tris-acetic acid-EDTA buffer

TCEP – tris (2-carboxyethyl) phosphine

TEV protease – tobacco etch virus protease

Tma – *Thermotoga maritima*

tRNA – transfer RNA

List of Figures

List of Tables

Chapter One

Introduction

The catalytic power of enzymes makes nature an exceptional chemist. Unravelling the secrets of how these large biomolecules control metabolic and cellular signalling pathways is necessary to understand the chemistry of life. This thesis explores the structure and function of the first enzyme in histidine biosynthesis, an enzyme that initiates a metabolically essential but energetically expensive reaction sequence. Understanding the tight regulation and diverse structures of this enzyme will help to probe the different modes of allosteric regulation and provide insight into the evolution of enzymes.

1.1 L-Histidine Biosynthesis

L-Histidine (hereafter referred to as histidine) is an essential amino acid that often plays a critical role in enzyme active sites. The imidazole side chain of histidine can be reversibly protonated at physiological pH, a property that allows it to act as a proton shuttle. Histidine residues often function as metal co-ordinating ligands in metallo-proteins and the free amino acid is able to be decarboxylated to form histamine, a neurotransmitter. The histidine biosynthetic pathway is present in archaea, bacteria, plants, and some lower eukaryotes. The absence of this pathway from mammalian metabolism highlights it as a potential target for herbicide[6] and antibiotic design.[7,8]

There are eight or nine genes that encode the enzymes that catalyse the ten reactions of histidine biosynthesis.[2,9] These genes are usually arranged in an operon, referred to as the his operon. Early investigation into the histidine biosynthetic pathway aided the definition and refinement of operon theory[10] and has unravelled some fundamental biological processes, including attenuation and feedback inhibition.[2] Three of the genes of the his operon (*hisD*, *hisB*, and *hisI*) encode bifunctional enzymes and two (*hisH* and *hisF*) or sometimes four (*hisZ* and *hisGs* – discussed later) encode single polypeptide chains that assemble into a heteromeric enzyme that catalyses a single reaction.[2]

The substrates of the first committed step of histidine biosynthesis, adenosine triphosphate (ATP) and 5-phosphoribosyl 1-pyrophosphate (PRPP) play a key role in intermediary and energy metabolism. Using ATP and PRPP, the first product *N*'-[5-phosphoribosyl]-ATP (PR-ATP) is formed via a phosphoribosyl transferase reaction (Figure 1.1). This step is followed by hydrolysis, resulting in *N*'-[5-phosphoribosyl]-AMP (PR-AMP). Hydrolytic opening of the adenosine ring results in the production of *N*'-[(5'-phosphoribosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (ProFAR). This intermediate undergoes ribose ring opening via an Amadori rearrangement to produce 5'-phosphoribosyl-4-carboxamide-4-aminoimidazole (AICAR) and imidazole glycerol phosphate (IGP). Glutamine is deaminated to provide the ammonia required for this step. IGP is further dehydrated to give imidazole acetolphosphate (IAP). After transamination to histidinol phosphate and dephosphorylation, histidinol undergoes two sequential oxidation steps to produce histidine. The histidine biosynthetic pathway is closely linked to the biosynthesis of tryptophan, folates, pyrimidines, and is also linked to purine

biosynthesis via the AICAR cycle. The focus of this thesis is the phosphoribosyl transferase enzyme that catalyses the first committed step of this pathway, ATP-phosphoribosyl transferase (ATP-PRTase).[2]

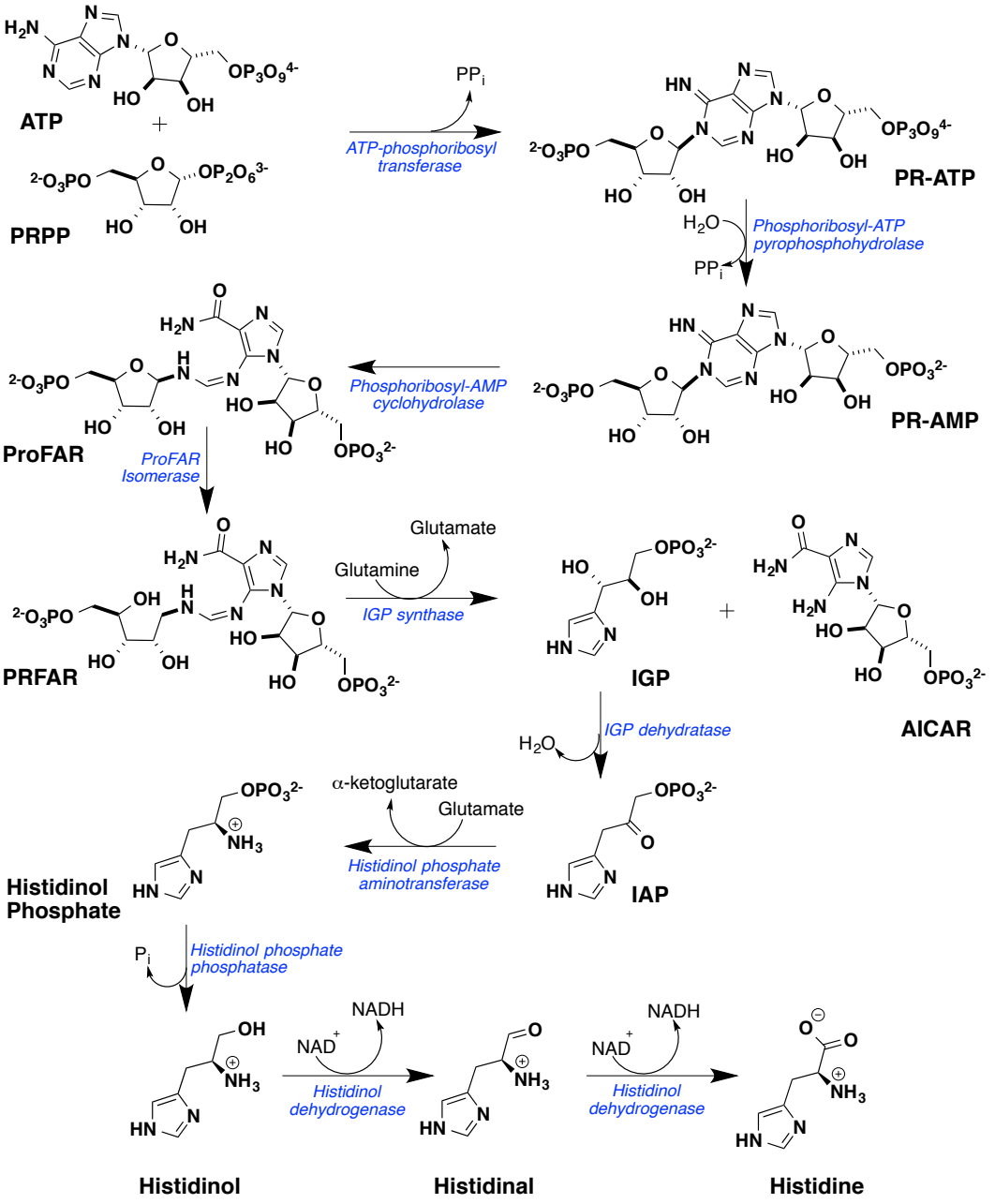

Figure 1.1 The histidine biosynthetic pathway. The names of the enzymes that catalyse each step are indicated in blue while the names of each substrate and product are shown in black.

1.2 Phosphoribosyl Transferases

Phosphoribosyl transferase (PRTase) enzymes are of critical metabolic importance as they catalyse essential reactions in the biosynthesis of purines, pyrimidines, and the amino acids, tryptophan and histidine.[11,12] These enzymes are structurally diverse and are organised into four classes based on their protein fold (Figure 1.2).[4,11,13] PRTases are characterised by the attack of a nitrogenous and generally aromatic base on PRPP, an activated form of ribose 5-phosphate. PRTase reactions require a divalent metal ion for catalysis and proceed with inversion of stereochemistry at the anomeric centre.[12] The diversity of sequence and structure between the four types of PRTase suggests a convergent evolution of this enzyme family.

The type I PRTases are the most well defined class of PRTase and are typically involved in nucleotide synthesis and salvage pathways. These include adenine, orotate, and hypoxanthine PRTase. Type I PRTases generally consist of four domains: a core region that binds both PRPP and the nucleotide substrate, a flexible loop that closes over PRPP, a variable hood domain that recognises the specific nucleotide substrate, and a C-terminal region that facilitates dimerisation.[9,10] Type I PRTases comprise a parallel β-sheet catalytic fold flanked by α-helices and they are identified by a 13 amino acid PRPP binding motif that includes two acidic amino acids, four hydrophobic amino acids, and seven variable amino acids that usually include glycine and threonine.[13,14]

The type II PRTases lack the distinct PRPP binding motif, which is limited to the type I PRTases. Type II PRTases include quinolate PRTase[15] and nicotinic acid PRTase[16] and are formed from a mixed α/β

N-terminal domain that resembles an open β-sandwich, and a $\beta_7\alpha_6$ C-terminal barrel.[15] The quaternary structures of the type II PRTases vary from the hexameric *Homo sapiens* quinolate PRTase[17] to the monomeric nicotinic acid PRTase of *Saccharomyces cerevisiae,*[16] although many exist as dimers.[15]

The tryptophan biosynthetic enzyme, anthranilate PRTase, is an example of a type III PRTase. Anthranilate PRTase consists of a large mixed α/β C-terminal domain linked to a smaller N-terminal α-helical domain via a hinge region. The quaternary structure of anthranilate PRTase is described as a head-to-head dimer.[18] The type IV PRTase enzymes are represented solely by ATP-PRTase. These enzymes consist of a mixed α/β bilobal catalytic domain with a PRPP binding motif that is conserved within the ATP-PRTase family and differs from that of the type I PRTases (Figure 1.3).[3,4,19,20]

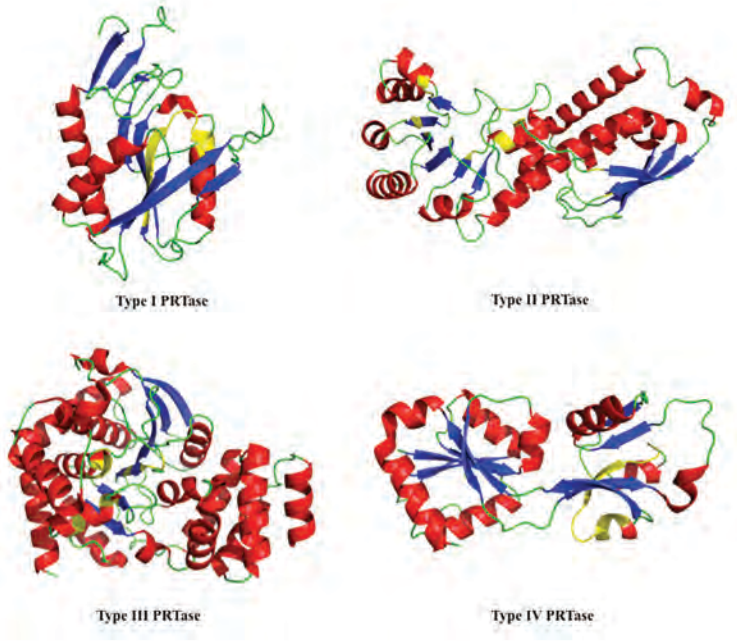

Figure 1.2 The array of PRTase protein folds. Single chains of each of the four PRTases are coloured by secondary structure with the PRPP binding regions coloured yellow. The type I PRTases are represented by hypoxanthine PRTase from *Trypanosoma cruzi* (PDB 1TC2). Quinolate PRTase from *Mycobacterium tuberculosis* represents the type II PRTases (PDB 1QPR). *M. tuberculosis* anthranilate PRTase (PDB 3QSA) and *Lactococcus lactis* ATP-PRTase ($HisG_S$) (PDB 1Z7M) represent the type III and type IV PRTases respectively.

Several PRTases have complex quaternary structures and sophisticated regulatory mechanisms, including glutamine PRPP amidotransferase, which catalyses the first committed step of purine biosynthesis,[21] and ATP-PRTase.[22,23] Both of these enzymes exhibit pathway end product inhibition and regulation by metabolites that reflect cellular energy levels, although the mechanism of this regulation is not well understood for either enzyme.

Figure 1.3 Multiple sequence alignment of several ATP-PRTase enzymes. An asterisk indicates conserved residues and a colon indicates residues with conserved functionality. The PRPP binding motif is coloured yellow. Organism names are as follows: Eco – *Escherichia coli*, Mtu – *Mycobacterium tuberculosis*, Cje – *Campylobacter jejuni*, Lla – *Lactococcus lactis*, Tma – *Thermotoga maritima*, Nme – *Neisseria meningitidis*. The sequence identity across this family is typically 20-30%.

1.3 ATP-Phosphoribosyl Transferase

The ATP-phosphoribosyl transferase enzyme (ATP-PRTase, HisG) encoded by the *hisG* gene is responsible for initiating histidine biosynthesis.[1,23] ATP-PRTase catalyses the reversible Mg^{2+}-dependent

condensation of PRPP and ATP to yield PR-ATP and inorganic pyrophosphate (Figure 1.4).[23] ATP-PRTase has been studied since the 1950s and this early effort, predominantly by Martin *et al.*, resulted in a basic understanding of its reaction chemistry.[1,22] Kinetic characterisation of the ATP-PRTase enzymes from *Escherichia coli, Salmonella typhimurium, Lactococcus lactis* and *Mycobacterium tuberculosis* revealed that the reaction follows an ordered bi-bi kinetic mechanism with ATP binding first and PR-ATP leaving last.[3,24-26] The mechanistic details of the substitution reaction are yet to be deduced. ATP-PRTase has also been shown to be feedback regulated competitively by adenosine monophosphate (AMP) and allosterically by histidine.[22,27]

The enzymes of histidine biosynthesis are potential pharmaceutical targets for a wide range of pathogens including *Neisseria meningitidis,* the causative agent of bacterial meningitis, and *M. tuberculosis,* which causes tuberculosis. The essentiality of the ATP-PRTase reaction and the histidine biosynthetic pathway to the survival of *M. tuberculosis* has been confirmed by high resolution transposon mutagenesis[28] and by genetic knockouts of other enzymes in the pathway.[29] Furthermore, the absence of this pathway from mammalian metabolism may indicate no *a priori* toxicity associated with potential inhibitors targeting these enzymes. There are two distinct subfamilies of ATP-PRTase that differ in their molecular architecture,[3,20] and an understanding of the structure and function of these enzymes could inform the design of potent antimicrobial compounds.

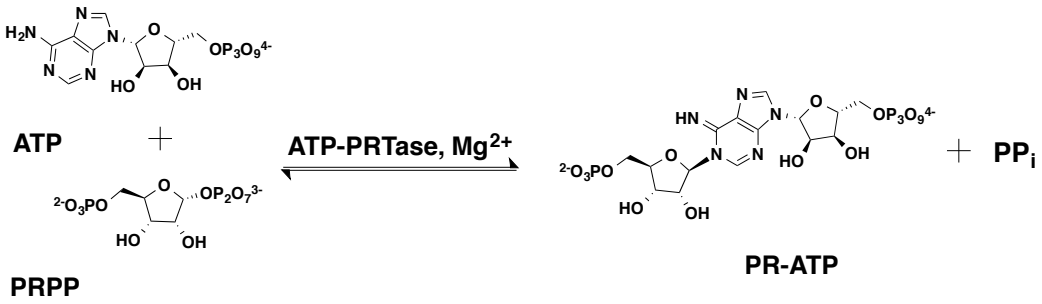

Figure 1.4 The reaction catalysed by the ATP-PRTase enzyme.

1.3.1 The Long Form of ATP-PRTase

The long form of ATP-PRTase ($HisG_L$) is present in most bacteria, as well as fungi and plants.[2] A single polypeptide chain of $HisG_L$ consists of 280-310 amino acids and comprises three domains (Figure 1.5 A).[3] Domains I and II form the catalytic core of $HisG_L$, with Domain II topographically inserted into Domain I (Figure 1.5 C).

The active site lies in the cleft between Domains I and II and includes a conserved PRPP binding motif.[13] Domain III is a regulatory domain that adopts the classic βαββαβ ACT domain fold[30] and forms the binding site for histidine as an allosteric inhibitor.[3] The ACT domain is a common regulatory motif found in many enzymes involved in amino acid metabolism.[31] The ACT domain was named after the three enzymes in which it was first discovered: aspartate kinase, chorismate mutase, and TyrA (prephenate dehydrogenase).[32] X-ray crystal structures have revealed a homo-hexameric quaternary structure of the long form ATP-PRTase (Figure 1.5B).[3,4] The long form ATP-PRTase enzymes from *M. tuberculosis*,[33] *E. coli*,[4] and *S. typhimurium*[34] have been well characterised. *Arabidopsis thaliana*

encodes two isozymes of $HisG_L$, which share 82% sequence identity, but the structure and function of these isozymes is not well understood.[35,36]

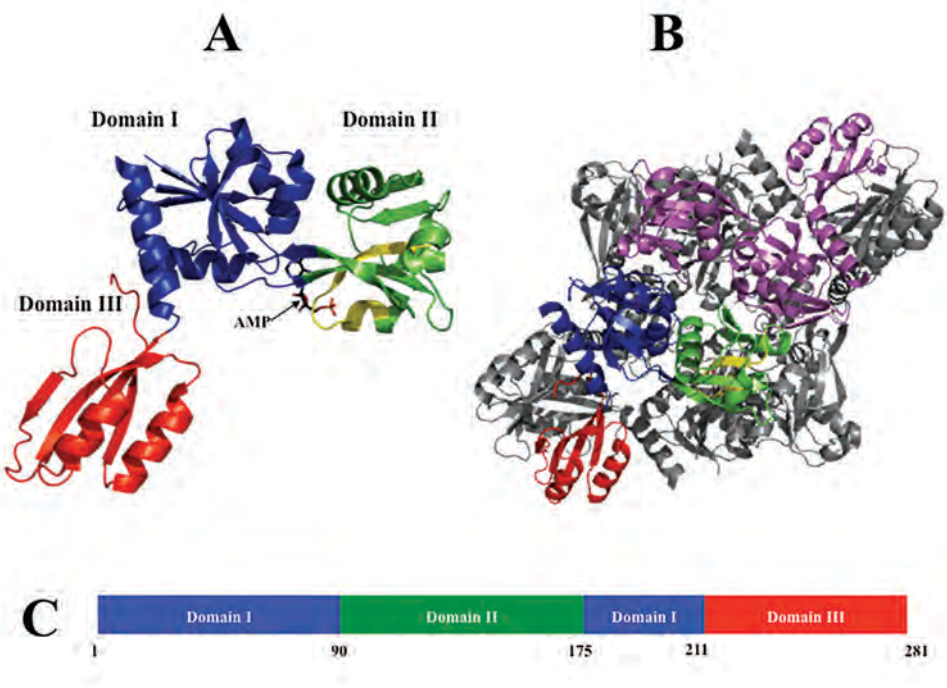

Figure 1.5 A) Single $HisG_L$ chain from *M. tuberculosis* (PDB 1NH8) with domain I coloured blue, domain II coloured green, and domain III coloured red. The PRPP binding motif is coloured yellow and AMP is shown bound in the active site. B) The hexameric quaternary structure with a single chain of $HisG_L$ coloured by domain, a second chain of $HisG_L$ coloured purple to show the dimer interface, and the remaining four chains shown in grey. C) A schematic of the $HisG_L$ amino acid sequence with *M. tuberculosis* $HisG_L$ numbering shown.

1.3.2 The Short Form of ATP-PRTase

The short form, denoted $HisG_S$ to illustrate its relationship to $HisG_L$, is approximately 80 amino acids shorter than $HisG_L$ at the C-terminus of the protein (Figure 1.6).[5,37] $HisG_S$ is restricted to proteobacteria and some bacteria and bioinformatic analysis has indicated that it is likely to be the phylogenetically older form of ATP-PRTase.[19,37] Structural characterisation of the short form enzymes from *L. lactis* and *Thermotoga maritima* has been reported.[19,20] $HisG_S$ has approximately 25 % sequence identity to $HisG_L$ and consists of two domains comparable to domains I and II of $HisG_L$,

but is devoid of the ACT regulatory domain (Figure 1.6 B). Surprisingly, $HisG_S$ appears to be catalytically inactive on its own[19,20,38] and instead forms a non-covalent complex with a second protein, HisZ, to generate a functional enzyme.[5]

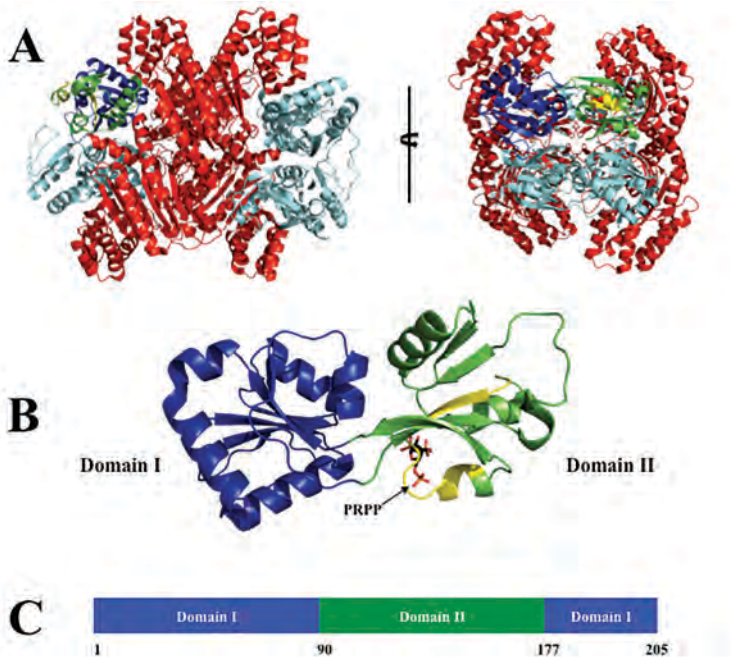

Figure 1.6 A) Two views of the $HisZG_S$ complex from *L. lactis* (PDB 1Z7M). One $HisG_S$ chain is coloured by domain, the remaining four $HisG_S$ chains are shown in cyan and the four HisZ chains are coloured red. B) A single $HisG_S$ chain coloured by domain. The cleft between domains I (blue) and II (green) houses the active site and the conserved PRPP binding region is shown in yellow. PRPP is shown bound in the active site. C) Schematic of the $HisG_S$ polypeptide chain coloured by domain and numbered according to the *L. lactis* sequence.

HisZ is a paralogue of class II histidyl-tRNA synthetase.[5] Aminoacyl-tRNA synthetases are enzymes responsible for correctly matching amino acids with their cognate tRNA to maintain the fidelity of the genetic code. This relationship reflects a clear evolutionary link between the synthesis of proteins and their amino acid building blocks.[5] The correlation between the absence of the ACT domain and the presence of HisZ, as well as the evolutionary relationship between HisZ and histidyl-tRNA synthetase, suggests that HisZ plays a role in the allosteric regulation by histidine.[19,20] HisZ from *L. lactis* does

exhibit non-specific RNA binding properties but this no longer appears to be functionally significant.[5] Although HisZ does not display phosphoribosyl transferase activity itself, it is believed to be required for catalysis, as an *E. coli hisG* knockout strain required co-transformation of plasmids containing both *hisG*$_S$ and *hisZ* from *L. lactis* in order to grow on histidine deficient media.[5] The molecular basis for this requirement is unknown. The HisZ and HisG$_S$ complex from *L. lactis* and *T. maritima* adopts a hetero-octameric quaternary structure (Figure 1.6A).[19,20,39]

1.3.3 Catalytic Mechanism

The ATP-PRTase reaction follows an ordered bi-bi kinetic mechanism with ATP binding first and PR-ATP leaving last.[24] Pre-steady state kinetics showed that product release, rather than chemical reaction, is likely to be rate-limiting.[33] The reaction was initially believed to proceed via the formation of a covalent intermediate, however, detailed kinetic analysis has refuted the involvement of such an intermediate.[40]

The reaction chemistry is a nucleophilic substitution that proceeds with inversion of stereochemistry.[41] The mechanism of the ATP-PRTase reaction is yet to be determined, although two mechanisms can be proposed (Figure 1.7). The reaction may proceed via a dissociative (S_N1-like) mechanism in which the pyrophosphate leaving group of PRPP dissociates first, resulting in an oxocarbenium ion intermediate. This is followed by nucleophilic attack of the 1-N of the adenosine ring of ATP on the oxocarbenium ion. Alternatively, the reaction may proceed via an associative (S_N2-like) mechanism in which

nucleophilic attack and dissociation of the leaving group occur in the same step, without the formation of a discrete intermediate. The degree of S_N1-like or S_N2-like character can be discerned by determining the structure of the transition state. The transition state structure is likely to vary between enzymes of different organisms.[42]

Parsons *et al.* have used kinetic isotope effects to study the mechanism of all four classes of PRTase from *S. typhimurium*, including ATP-PRTase.[43] Types I, II and III PRTases were clearly shown to follow an S_N1-like mechanism, with large ^3H kinetic isotope effects observed, consistent with the formation of an oxocarbenium ion intermediate. The results of the analysis for ATP-PRTase also indicated an S_N1-like mechanism, however these results were less conclusive due to the reverse reaction potentially supressing the overall kinetic effect. It is probable that, in accordance with the other three classes of PRTase, ATP-PRTase proceeds via an S_N1-like reaction, however the observation that ATP binds first, and that the mechanism is sequential, means that possible participation of the ATP nucleophile cannot be excluded.

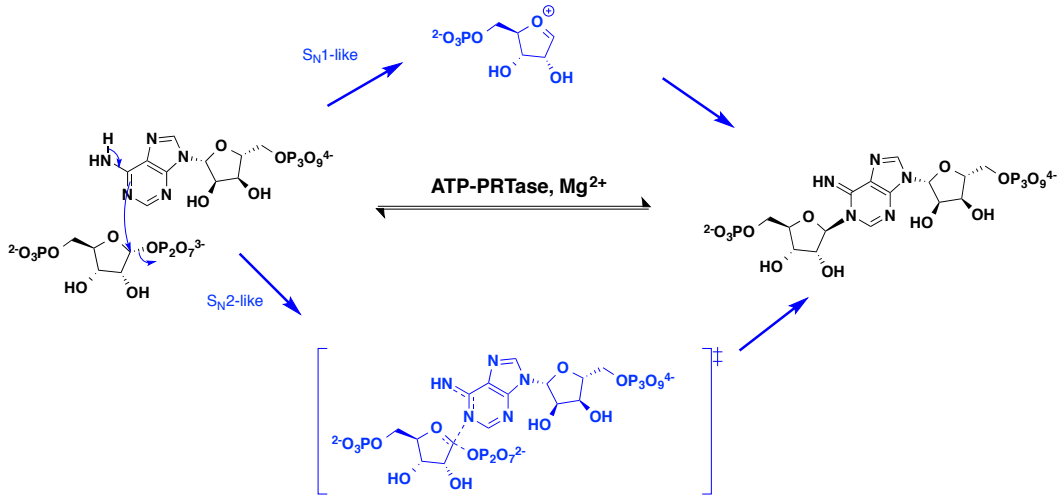

Figure 1.7 Reaction catalysed by ATP-PRTase with the two proposed mechanisms indicated in blue.

1.3.4 ATP-PRTase Structure

The $HisG_L$ and $HisG_S$ chains both possess a bilobal catalytic domain that is homologous to the periplasmic binding proteins that bind sulfate, phosphate, and other small molecules.[44,45] Early biochemical work on ATP-PRTase focused on the relationship of the quaternary structure to the catalysis and regulation of the enzyme.[23,46-48] There are a total of 12 crystal structures available in the protein data bank (PDB).[49] These include the long form structures from *E. coli* (PDB 1Q1K and 1H3D),[4] *M. tuberculosis* (PDB 1NH7 and 1NH8),[3] and *Methanobacterium thermoautotrophicum* (PDB 2VD3), the short form complex from *L. lactis* (PDB 1Z7M and 1Z7N)[20] and *Thermotoga maritima* (PDB 1USY),[19] and the $HisG_S$ dimer from *Bacillus subtilis* (PDB 2VD2) and *Thermus thermophilus* (PDB 1VE4) .

Long Form ATP-PRTase Structure

In 2003 Cho *et al.* provided the first glimpse of the ATP-PRTase structure with the publication of two X-ray crystal structures of the enzyme from *M. tuberculosis* (*Mtu*$HisG_L$).[3] *Mtu*$HisG_L$ was solved to 1.8 Å in the presence of the inhibitors AMP and histidine (PDB 1NH8), and to 2.7 Å in the absence of ligands (PDB 1NH7). This was quickly followed by the publication of a 2.7 Å crystal structure of the AMP bound ATP-PRTase from *E. coli* (*Eco*$HisG_L$) (PDB 1H3D), and a second *Eco*$HisG_L$ structure at a resolution of 2.9 Å with the product, PR-ATP, partially modelled into the active site* (PDB 1Q1K).[4] A single chain of *Mtu*$HisG_L$ (PDB 1NH7) aligns to *Eco*$HisG_L$ (PDB 1H3D) with an RMSD of 1.6 Å, reflecting a high degree of similarity between the overall folds the two structures. As described in section 1.3.1, the monomeric unit of

* The ribosyl-triphosphate could not be resolved.

$HisG_L$ consists of three domains.[3] Domain I is formed from an antiparallel β-sheet flanked by four α-helices, one of which links to domain III, the ACT regulatory domain. Domain II consists of a twisted β-sheet core, surrounded by α-helices. Domains I and II are connected by a twisted β-ribbon and the crevice between the two domains houses the active site, which includes the conserved PRPP binding motif (Figure 1.8 A).

Early biochemical work on *Eco*$HisG_L$ suggested that the long form ATP-PRTase adopts an active dimeric structure (Figure 1.8 B), and that binding of histidine, as an allosteric inhibitor, facilitates the transition to an inactive hexameric species.[46] This dimer-hexamer model was supported by analytical size exclusion chromatography, performed by Cho *et al.,* on the long form enzyme from *M. tuberculosis,* although the experimental data were not presented.[3] This model persisted despite the revelation of the hexameric crystal structures of ATP-PRTase from both organisms in the absence of ligands and in the presence of PR-ATP, histidine, and AMP (Figure 1.9).[3,4] Lohkamp *et al.* attributed the hexameric crystal structure to the high concentration of protein in the crystal, citing the propensity of the enzyme to aggregate at high concentrations.[4,50] Recent work by Pedreno *et al.*,[33] on the ATP-PRTase from *M. tuberculosis*, has challenged the dimer-hexamer model of the long form ATP-PRTase and has been supported by further studies, using a variety of biophysical techniques, on $HisG_L$ from *Campylobacter jejuni* (Gerd Mittelstädt personal communication). Contrary to previous reports, it appears more likely that the active and inactive states are both hexameric. Koshland *et al.* in 1974, also suggested that the physiologically relevant form of ATP-PRTase from *S. typhimurium* is hexameric, although a variety of oligomeric states were observed.[51]

The dimer interface of $HisG_L$ is antiparallel, with domain I of one chain of $HisG_L$ interacting with domain II of the second $HisG_L$ chain (Figure 1.8 B). Domain I joins the ACT domain via a loop connected to a long α-helix. Interestingly, this loop is unresolved in the *Mtu*$HisG_L$ crystal structures, suggesting that this region is quite flexible. The contacts responsible for formation of the homo-hexameric quaternary structure are almost exclusively between the ACT domains.

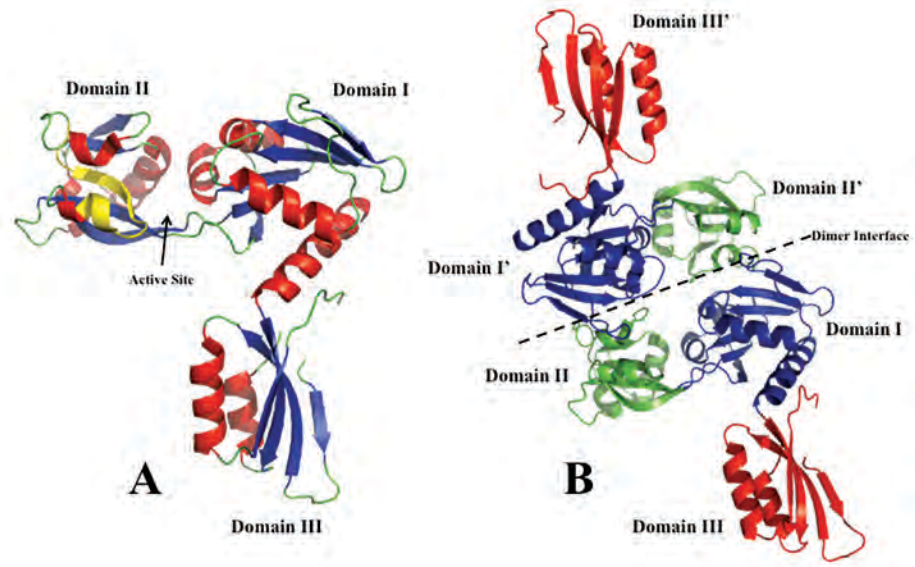

Figure 1.8 A) A single chain of *Mtu*$HisG_L$ (PDB 1NH7) coloured by secondary structure. The active site is situated in the cleft between domains I and II and the PRPP binding motif is coloured yellow. B) Two chains of *Mtu*$HisG_L$ coloured by domain with the dimer interface indicated by a dotted line.

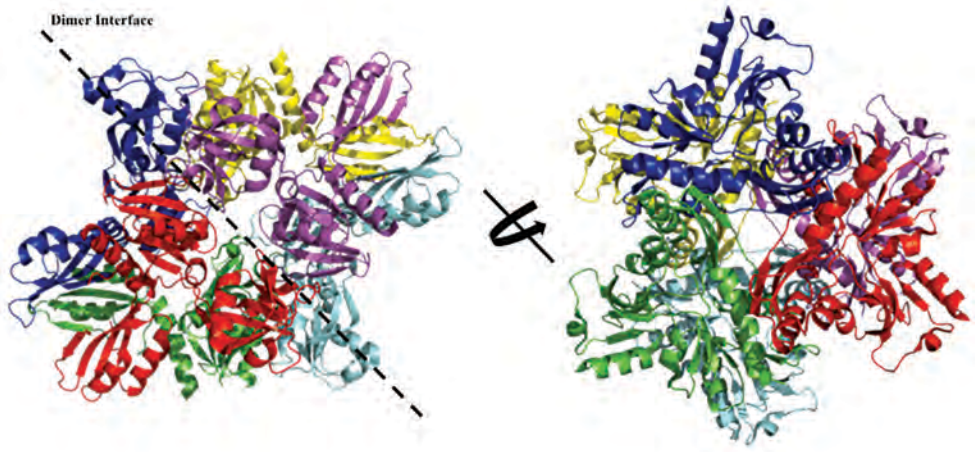

Figure 1.9 The hexameric quaternary structure of *Mtu*$HisG_L$ coloured by chain.

Short Form ATP-PRTase Structure

X-ray crystal structures of the short form ATP-PRTase from *L. lactis* (*Lla*HisZG$_S$) and *T. maritima* (*Tma*HisZG$_S$) have revealed a hetero-octameric quaternary structure that is unusual among both the phosphoribosyl transferases and the aminoacyl-tRNA synthetases.[19,20] Despite the similarity in the folds of the individual chains, the *Tma*HisZG$_S$ complex (PDB 1USY) aligns to *Lla*HisZG$_S$ (PDB 1Z7M) with an RMSD of 20 Å. It is unclear whether this difference is due to ligand-induced conformational changes, crystal packing, or sequence differences. The comparison of *Lla*HisZG$_S$ and *Tma*HisZG$_S$ is explored further in chapter two.

The short form complex consists of four subunits of HisZ arranged in an X-shaped core with two large cavities, each of which binds two subunits of HisG$_S$ (Figure 1.11). HisZ typically consists of 300-350 amino acids and comprises a seven-stranded antiparallel β-sheet flanked by α-helices. HisZ features the three domains characteristic of histidyl-tRNA synthetases[52] and class II aminoacyl-tRNA synthetase paralogues,[53,54] but possesses an α-helical distal domain with a different topology (Figure 1.10).[19] In the absence of HisG$_S$, *Tma*HisZ has been reported to exist as a dimer-tetramer equilibrium in solution, however no data were shown to support this observation.[19] HisZ usually lacks the anticodon binding domain of the histidyl-tRNA synthetase and is therefore devoid of amino acylation activity.[5] The HisZ proteins from the *Geobacter* clade, however, are formed from approximately 550 amino acids and sequence analysis has attributed the additional amino acids to the retention of the anticodon binding domain.[38]

The bacteria of the *Geobacter* clade are the only organisms reported to date whose genome encodes both a long and short form ATP-PRTase. Gene knockout studies demonstrated that the short form is the predominant ATP-PRTase, while the long form functions mainly under nitrogen fixation conditions.[38] Surprisingly, Aklujkar *et al.* also observed that HisZ is essential to the survival of *Geobacter sulfurreducens,* both in the presence and absence of an exogenous source of histidine. This unexpected result may be due to a poorly constructed knockout or could potentially indicate another essential function of HisZ in the cell.[38]

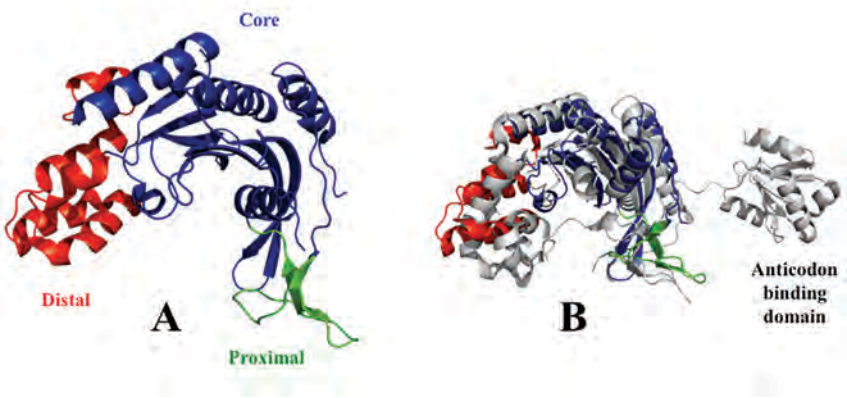

Figure 1.10 A) A single chain of *Tma*HisZ (PDB 1USY) coloured by domain. B) *Tma*HisZ aligned with the histidyl-tRNA synthetase (grey) from *Thermus thermophilus* (PDB 1H4D) with an RMSD of 4.108 Å. *Tma*HisZ is coloured by domain as in A. The folds of the core and proximal domains are similar. The distal subdomain, responsible for interaction between HisZ and $HisG_S$, differs in topology. *Tma*HisZ also lacks the anticodon binding domain.

It has been noted that HisZ is required for catalysis, despite $HisG_S$ housing the active site, however, the molecular details that account for this activation are unclear. The crystal structures of $HisG_S$ from *B. subtilis* (PDB ID 2VD2) and *T. thermophilus* (PDB ID 1VE4) reveal a dimeric structure of the protein in the absence of HisZ. Both of these structures readily superimpose with the $HisG_S$ chains from the *Lla*$HisZG_S$ hetero-octameric complex, indicating no major conformational changes of $HisG_S$ in the presence or absence of HisZ

(Figure 1.12). Further research into the structure and dynamics of the short form ATP-PRTase is required to determine the role of HisZ in catalysis and to understand how this catalytic activity evolved via collaborative adaptation of two distinct ancestral proteins.

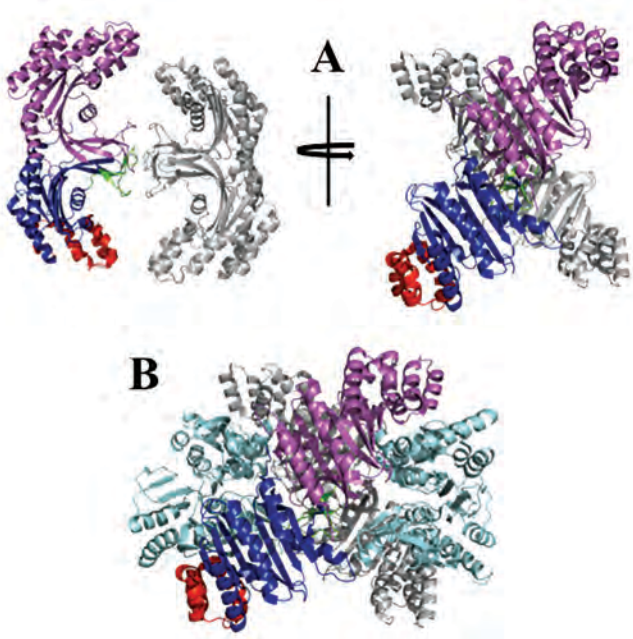

Figure 1.11 A) Two views of the *Tma*HisZ core (PDB 1USY). One chain of HisZ is coloured by domain, a second HisZ chain is coloured purple to show the dimer interface, and the remaining two chains are coloured grey. B) The *Tma*HisZG$_S$ hetero-octameric complex. The HisZ subunits are coloured as in A and oriented as shown on the RHS. The four HisG$_S$ chains are coloured cyan.

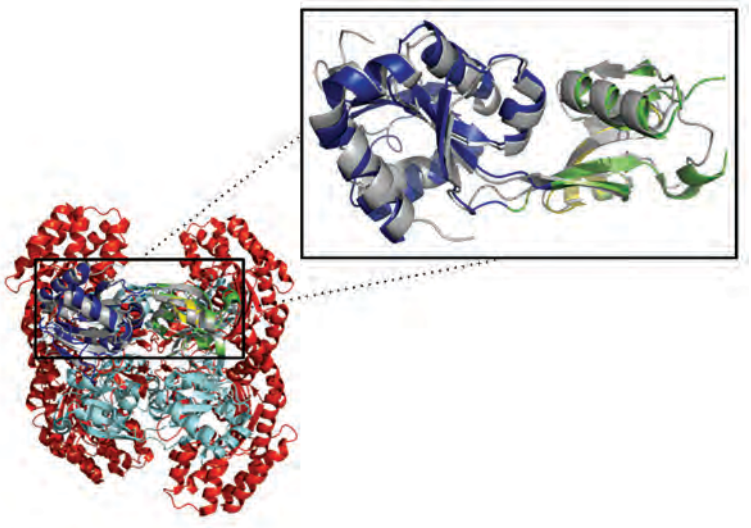

Figure 1.12 *Lla*HisZG$_S$ hetero-octameric structure (PDB 1Z7N) with a single chain of *Lla*HisG$_S$ (coloured by domain) aligned with the HisG$_S$ crystal structure from *T. thermophilus* (PDB 1VE4), coloured grey (RMSD 1.3 Å).

1.4 Regulation of ATP-PRTase

Histidine biosynthesis is one of the most energetically expensive metabolic pathways and ATP-PRTase, as the catalyst of the first committed step of this pathway, represents the commitment of 41 ATP equivalents to the synthesis of one molecule of histidine. Cells that exhibit unregulated histidine biosynthesis have been shown to dissipate up to 2.5% of metabolic energy.[55] As a result; ATP-PRTase is subject to an elaborate multilevel regulatory system.

1.4.5 Regulation on the Level of Transcription

A high concentration of histidine inhibits the transcription of the histidine operon, mediated by the attenuation effect of histidyl-tRNA, thereby preventing the synthesis of the histidine biosynthetic enzymes.[2] Promoter substitution in *Corynebacterium glutamicum*, whereby the histidine promoter region was exchanged for the stronger tac promoter, disrupted transcriptional regulation, resulting in increased levels of histidine biosynthesis.[56]

1.4.6 Feedback Inhibition

Adenosine monophosphate (AMP) and adenosine diphosphate (ADP) act as competitive inhibitors of ATP-PRTase with respect to both ATP and PRPP (Figure 1.13).[27] AMP and ADP reflect a low overall energy status of the cell and a need to down regulate energetically expensive pathways such as histidine biosynthesis. The product PR-ATP also inhibits the enzyme, as does inorganic pyrophosphate, and ATP when it is present in excess.[25] $HisG_L$ also loses specific

activity at high protein concentrations, although this may reflect a tendency of the enzyme to aggregate at high concentrations – a phenomenon that is unlikely to be biologically significant.[50] The most prominent control of ATP-PRTase activity is the overall synergistic inhibition by AMP and the allosteric inhibitor histidine.[27]

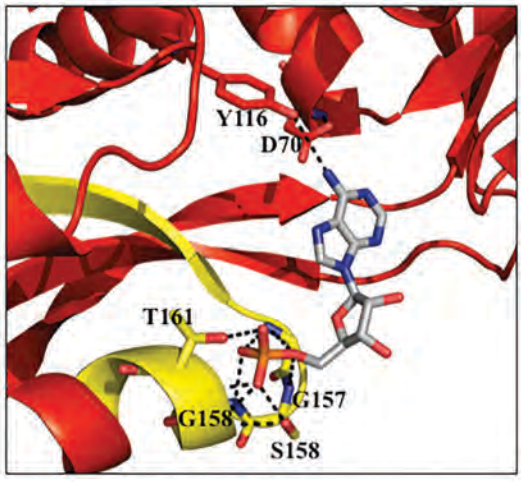

Figure 1.13 Crystal structure of *Mtu*HisG$_L$ (1NH8) with AMP bound in the active site. The PRPP binding motif is coloured yellow. AMP is shown to partially occupy both the ATP and PRPP binding regions, which provides a molecular basis for the competitive nature of the inhibition with respect to both substrates.

1.4.7 Allosteric Regulation by Histidine

The two families of ATP-PRTase differ significantly in their quaternary structure and likely transition between active and inactive states via two very different mechanisms.

Long Form of ATP-PRTase

The X-ray crystal structure of the histidine and AMP-bound *Mtu*HisG$_L$ (PDB 1NH8), reveals the location of the histidine binding sites at the cleft between two ACT domains in the hexamer, approximately 30 Å from the active site.[3] One histidine molecule binds

per ACT domain, for a total of six in the hexamer. The interaction of histidine with conserved residues suggests that these binding sites are likely to be biologically functional (Figure 1.14). Although Cho *et al.* attributed the mode of regulation to oligomerisation from an active dimer to an inactive hexamer, facilitated by histidine binding, recent work has suggested a more subtle mechanism of allosteric regulation.[33] It appears that with a combination of AMP and histidine *Mtu*HisG$_L$ exists as a tight hexamer, caused by a slight twist in the ACT domains relative to domains I and II (Figure 1.15). Domains I and II of the unliganded and histidine/AMP bound *Mtu*HisG$_L$ crystal structures align with an RMSD of 1.4 Å, whereas the relative positioning of the ACT domains differs by approximately 13 Å, indicative of a conformational change. This is consistent with the observation of Tebar *et al.* that the radius of the *Eco*HisG$_L$ enzyme decreases in the presence of histidine.[57] Further research is required to determine how this allosteric inhibition is communicated. Tightening of the hexamer may not be enough to attribute inhibition to steric hindrance, as there remains a cavity of sufficient size for substrate and product diffusion. To this end, molecular dynamics simulations may prove to be invaluable, as it has been well established that dynamics and conformational changes are both important for allosteric signal transduction.[58]

The proposed histidine binding residues have been supported by recent work on *C. glutamicum* HisG$_L$. Schendzielorz *et al.* developed a method for the *in vivo* identification of mutant strains that over produce histidine.[59] In the process they discovered that the HisG$_L$ mutants GT233HQ, N216R, and N216I had significantly impaired histidine inhibition. Homology modelling and mutagenesis by Zhang *et al.* demonstrated that N216, L231, T235, and A271 play a role in

histidine inhibition in *C. glutamicum* and possibly serve as histidine binding residues.[60] pH dependence of allosteric inhibition of *Mtu*HisG$_L$ has also been observed. Pedreno *et al.* measured histidine binding to *Mtu*HisG$_L$ using saturation transfer difference NMR. A neutral α-amino group was shown to achieve the tightest binding to the enzyme, however superior inhibition is achieved with a neutral imidazole group and an ionised α-amino group.[33]

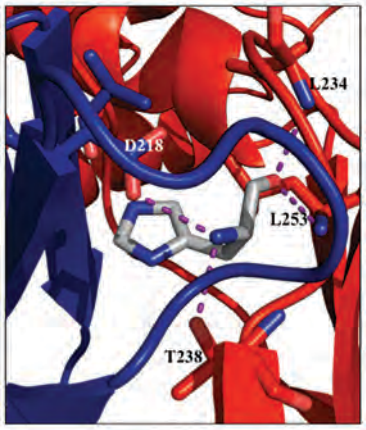

Figure 1.14 The histidine binding site of *Mtu*HisG$_L$ (PDB 1NH8) between two ACT domains (coloured red and blue). Hydrogen bonds are indicated by purple dotted lines.

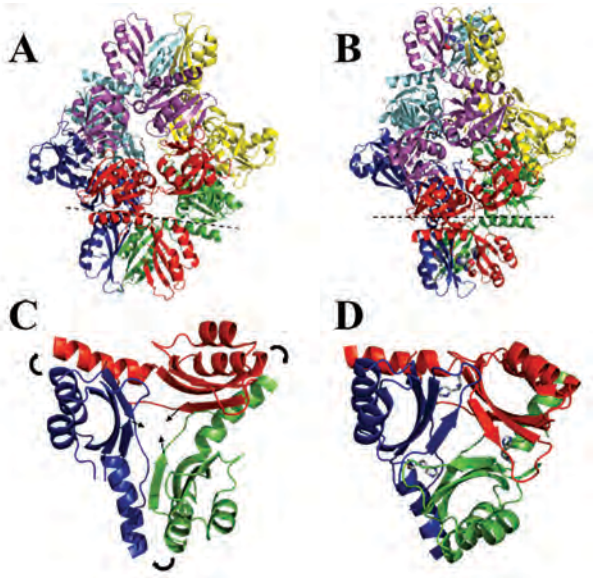

Figure 1.15 Crystal structures of *Mtu*HisG$_L$ depicting A) the unliganded hexamer (PDB 1NH7), and B) the histidine and AMP bound hexamer (PDB 1NH8). The ACT domain interfaces are shown (truncated at the dotted lines) to demonstrate the transition from unliganded (C) to histidine bound (D) structures.

Short Form of ATP-PRTase

The mode of allosteric inhibition of the short form ATP-PRTase is unclear. Analytical ultracentrifugation experiments on *Lla*HisZG$_S$ showed a change in the sedimentation coefficient in the presence of histidine.[39] This may be interpreted as a change in the conformation or subunit composition of the protein. In contrast, dynamic light scattering on *Tma*HisZG$_S$ indicated that histidine binding had no effect on the hydrodynamic radius of the protein.[19]

The X-ray crystal structure of *Tma*HisZG$_S$ has been solved in the presence of histidine to 2.5 Å.[19] Eight histidine binding sites, located at each of the HisZ-HisG$_S$ interfaces, are shown in the hetero-octamer (Figure 1.16). It was later revealed that mutation of conserved residues in *Lla*HisZ, homologous to the active site of the histidyl-tRNA synthetase, resulted in a loss of allosteric inhibition without affecting phosphoribosyl transferase activity.[26] This observation cast doubt on the biological relevance of the histidine binding sites reported in the *T. maritima* structure. It should be noted that it was the loss of histidine inhibition that was detected, as opposed to the loss of histidine binding; therefore mutation of these residues may have interfered with transmission of the allosteric signal rather than histidine binding. Nevertheless, any of the three potential binding sites represents a different mode of allosteric inhibition to the long form ATP-PRTase. The discrepancy between the histidine binding sites is discussed in chapter two.

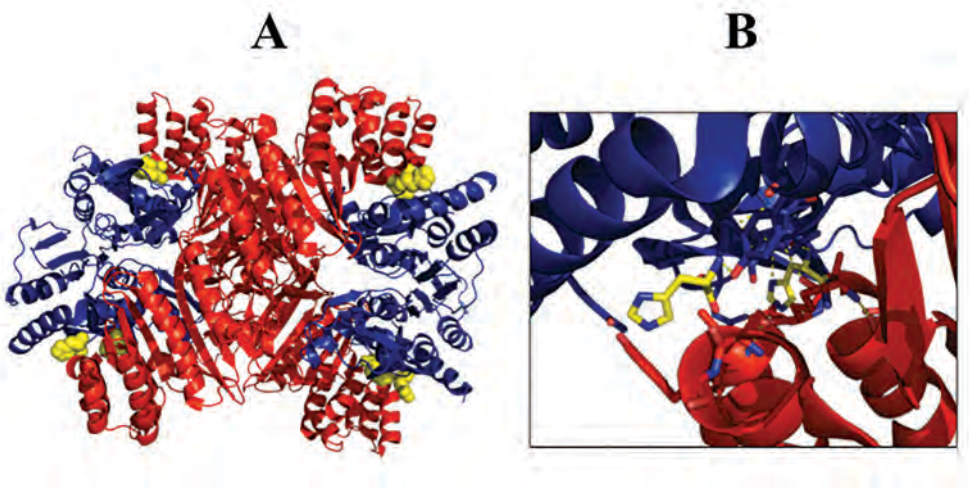

Figure 1.16 Crystal structure of *Tma*HisZG$_S$ with histidine bound (PDB 1USY). The HisZ chains are shown in red, HisG$_S$ chains in blue, and the bound histidine ligands are coloured yellow. A) The hetero-octameric structure with the histidine ligands shown as spheres. B) Two histidine binding sites at the interface of HisZ and HisG$_S$.

1.5 Objectives of this Thesis

The aim of this research is to investigate the allosteric regulation of the short form ATP-PRTase by histidine and to explore the functional and evolutionary relationship between the two families of ATP-PRTase. Chapter two describes the cloning, expression, and purification of the short form ATP-PRTase from *L. lactis* and the use of small angle X-ray scattering and isothermal titration calorimetry to characterise the allosteric regulation of the enzyme.

Chapter three details the attempts to clone, express, and purify the short form ATP-PRTase from *N. meningitidis*, the causative agent of meningitis, as an understanding of the structure and function of this enzyme could inform the design of antibiotics to combat this disease. The attempts to clone, express, purify and characterise the long form ATP-PRTase from *E. coli*, with the intent of clarifying the previous work on the dimer-hexamer model of allosteric inhibition, are also

described, as is the synthesis of a truncation mutant that is devoid of the ACT regulatory domain.

Chapter four reports the first study of a third molecular architecture of ATP-PRTase from *Leuconostoc mesenteroides,* in which HisZ and $HisG_S$ are covalently linked. This architecture was investigated both by cloning, expression and homology modelling of the wild-type fused enzyme and by engineering a covalent linker between HisZ and $HisG_S$ from *L. lactis*. The transfer of the ACT domain from $HisG_L$ of *C. jejuni* to $HisG_S$ of *L. lactis* is also reported, in the attempt to model the domain transfer event that conferred both allosteric regulation and phosphoribosyl transferase activity to the short form HisG, allowing the reaction to be catalysed and inhibited by a single gene product.

Chapter Two

2 Allosteric Inhibition of *Lla*HisZG$_S$

2.1 Introduction

Allostery, whereby ligand binding at one site affects catalysis at another remote functional site, is a common feature of key metabolic enzymes. A spectrum of general allosteric mechanisms has been proposed, ranging from the traditional paradigm of large conformational changes, to perturbations of backbone and side chain dynamics, to local unfolding of the protein.[58] Histidine acts as an allosteric inhibitor of ATP-PRTase. The histidine binding sites at the interfaces of the ACT domains in the long form ATP-PRTase have been well established, and an allosteric mechanism involving subtle conformational changes has been proposed.[3,33,56] A lack of structural information has precluded the development of a convincing model for the allosteric inhibition of the short form ATP-PRTase. The HisZG$_S$ complexes from *L. lactis* and *T. maritima* represent the only structurally characterised short form enzymes, and discrepancies exist between the reported response to histidine binding and the location of the functional allosteric sites within the two enzymes (Figure 2.1).[19,20,26,39]

Analytical ultracentrifugation experiments, reported by Bovee *et al.*, indicated a change in the sedimentation coefficient of *Lla*HisZG$_S$ upon the addition of histidine.[39] This may be interpreted as a change in the subunit composition or the conformation of the protein. In contrast,

Vega *et al.* reported that dynamic light scattering suggested the hydrodynamic radius of *Tma*HisZG$_S$ was unaffected by histidine binding, although no data were shown to support this observation.[19]

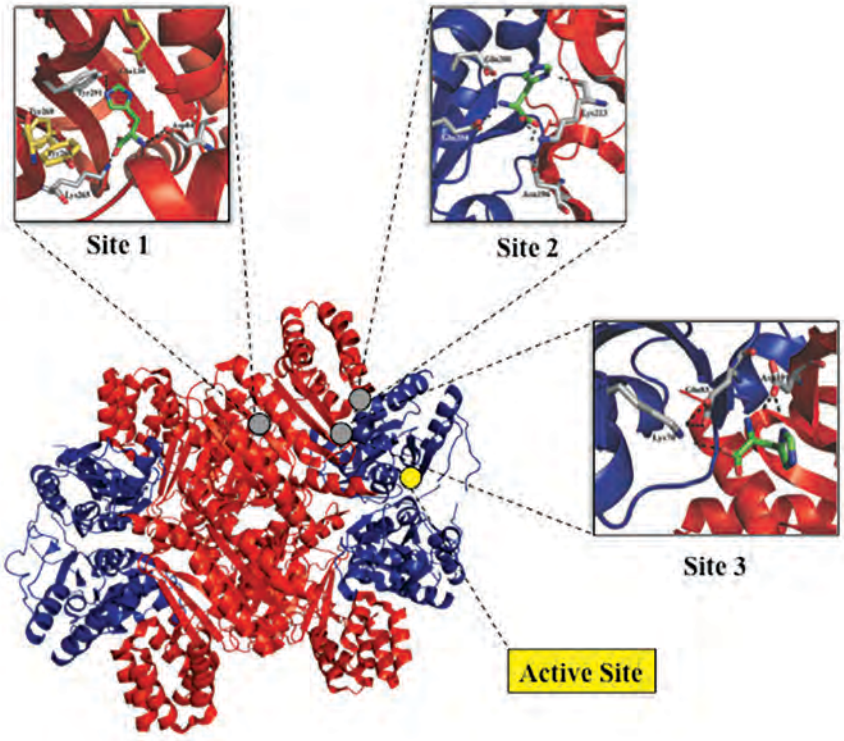

Figure 2.1 The crystal structure of *Lla*HisZG$_S$ (PDB 1Z7N) with each of the three proposed histidine binding sites shown. The active site is indicated by a yellow dot. Histidine (green) is shown modelled into each of the three potential binding sites. Site 1 is the binding site proposed by Champagne *et al.*, equivalent to the active site of histidyl-tRNA synthetase,[20] while the remaining two sites are proposed based on comparison with the *Tma*HisZG$_S$ structure (PDB 1USY).[19]

The *Tma*HisZG$_S$ crystal structure (PDB 1USY) revealed two histidine binding sites at each HisZ-HisG$_S$ interface, a total of eight in the hetero-octamer.[19] The histidine binding residues at the HisZ-HisG$_S$ interfaces are not conserved among the short form ATP-PRTase family, which is unusual for an important functional binding site. Champagne *et al.* reported that the mutation of highly conserved residues in *Lla*HisZ, equivalent to the histidine binding residues of histidyl-tRNA synthetase from *T. thermophilus*, disrupted allosteric inhibition of

*Lla*HisZG$_S$ without impacting catalysis.[20] This represents an entirely different histidine binding site to those observed in the *Tma*HisZG$_s$ crystal structure, and leads to the question of which site(s) is the functional allosteric site? (Figure 2.1) Mutation of the histidine binding residues proposed by Champagne *et al.* may have simply disrupted the communication pathway between the allosteric and active sites. Alternatively, the lowest energy state of *Tma*HisZG$_S$, stabilised by the crystal lattice, may not represent the biologically relevant inhibited state.

Champagne *et al.* drew comparisons between the *Tma*HisZG$_S$ and *Lla*HisZG$_S$ crystal structures in an attempt to develop a model for the allosteric activation and inhibition of the enzyme.[20] The *Tma*HisZG$_S$ structure was assigned to the inhibited state, as the enzyme was purified in the presence of the synergistic inhibitors histidine (0.4 mM) and AMP (0.6 mM), and the RMSD of the alignment between *Tma*HisG$_S$ and histidine-bound *Mtu*HisG$_L$ (PDB 1NH8, RMSD 1.15 Å) was lower than the RMSD of the alignment with the apo *Mtu*HisG$_L$ structure (PDB 1NH7, RMSD 1.35 Å). Similarly, both crystal structures of *Lla*HisZG$_S$ were assigned to the active state based on co-crystallisation with ATP or its analogue *N*-1-methyl-ATP and PRPP, and alignment with the apo and histidine bound *Mtu*HisG$_L$ structures (RMSD of 1.24 Å and 1.27 Å respectively). This assignment is tentative and illustrates the requirement for crystal structures of the active and inactive HisZG$_S$ enzymes from the same organism, to distinguish sequence differences from ligand-induced conformational changes.

Champagne *et al.* noted the presence of an interstitial phosphate ion at two of the four *Lla*HisZ-*Lla*HisG$_S$ interfaces that is absent from

the histidine-bound *Tma*HisZG$_S$ structure.[20] It should be noted that phosphate buffers were used in the purification of both *Tma*HisZG$_S$ and *Lla*HisZG$_S$. However, a Tris-HCl buffer was employed for the crystallisation of both enzymes.[19,20] The phosphate binding residues are conserved among the HisZG$_S$ enzyme family, which suggests that the presence of the phosphate ion is not merely a crystallisation artefact.[20] The switch between inhibited and active states was proposed to involve the rearrangement of the HisG$_S$ dimer interface, driven by binding of the interstitial phosphate ion and histidine dissociation. The interstitial phosphate ion was proposed to operate in concert with HisZ to cause this rearrangement, which is believed to recruit catalytic residues to the active site.[20] Notably, an inorganic phosphate has also been implicated in the activation of both bacterial and mammalian PRPP synthetases, which are similar in structure to the type I PRTases.[61,62] Champagne *et al.* suggested that an inorganic phosphate ion may represent a cellular signal to stimulate energy requiring pathways, although the origin of such a phosphate *in vivo* was not discussed.[20]

The RMSD of the alignment between the hetero-octameric *Lla*HisZG$_S$ and *Tma*HisZG$_S$ is 20 Å, which reflects a significant conformational difference (Figure 2.2). Alignment of the HisZ and HisG$_S$ components individually suggests that conformational differences in HisZ may be the largest contributor to structural differences between the complexes (Table 2.1), although it is unclear what, if any, the functional significance of these differences is. The debate that has arisen over the functionality of the histidine binding sites in the *Tma*HisZG$_S$ crystal structure calls into question the relevance of this comparison. If histidine is not bound in the functional allosteric sites, does this structure truly represent the inhibited form of

HisZG$_S$? The inhibited conformation may be stabilised, fortuitously, by crystal lattice interactions, or the lowest energy state in the crystal lattice may not represent the inhibited form of HisZG$_S$.

Table 2.1 RMSD of the Structural Alignment Between Chains of *Lla*HisZG$_S$ (PDB 1Z7N) and *Tma*HisZG$_S$ (PDB 1USY)

Alignment Chain	RMSD (Å)
*Lla*HisG$_S$-*Tma*HisG$_S$ (monomer)	1.9
*Lla*HisG$_S$-*Tma*HisG$_S$ (dimer)	2.1
*Lla*HisZ-*Tma*HisZ (monomer)	7.2
*Lla*HisZ-*Tma*HisZ (dimer)	5.9
*Lla*HisZ-*Tma*HisZ (tetramer)	9.3
*Lla*HisZG$_S$-*Tma*HisZG$_S$ (octamer)	20

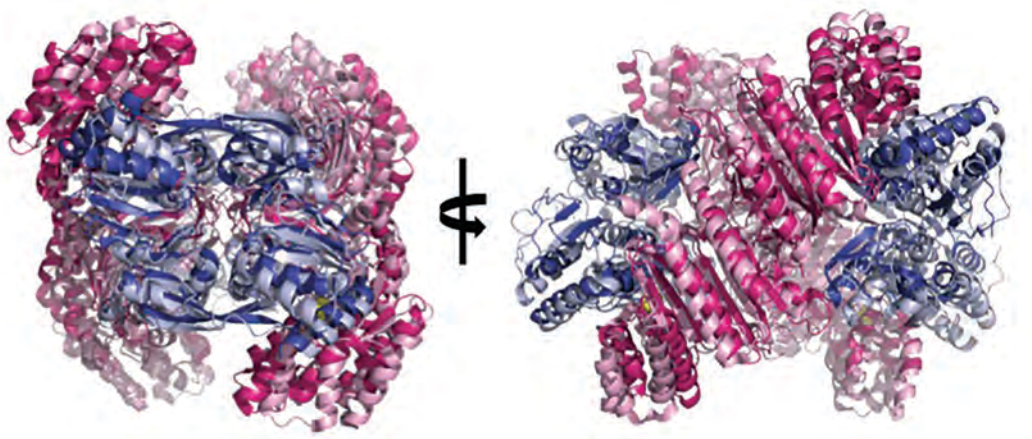

Figure 2.2 Alignment of *Lla*HisZG$_S$ (PDB 1Z7N) and *Tma*HisZG$_S$ (PDB 1USY). *Tma*HisZ and *Tma*HisZG$_S$ are shown in dark pink and dark blue respectively whilst *Lla*HisZ and *Lla*HisG$_S$ are coloured light pink and light blue. Interstitial phosphate ions are shown as yellow spheres.

The formation of a definitive model for the allosteric regulation of the short form ATP-PRTase requires an understanding of the crystal structure, solution behaviour, and dynamics of the active and inhibited enzyme. This chapter reports the determination of the solution

structure of both the active and inhibited *Lla*HisZG$_S$ and further evidence towards the definition of the functional histidine binding sites.

2.2 Cloning, Expression and Purification of *Lla*HisZG$_S$

The purification method of *Lla*HisZG$_S$ described by Bovee *et al*. reported that co-expression of *Lla*HisZ and *Lla*HisG$_S$ results in a decreased expression level of *Lla*HisG$_S$ relative to *Lla*HisZ.[39] The work towards the purification of HisZG$_S$ from *N. meningitidis,* detailed in chapter three, supports this observation. To this end the decision was made to utilise the approach taken by Vega *et al.* who expressed *Tma*HisZ and *Tma*HisG$_S$ separately, and purified the *Tma*HisZG$_S$ hetero-octameric complex by co-lysing the combined cell pellets.[19] Furthermore, the observation by Sissler *et al.* that the untagged *Lla*HisG$_S$ protein co-purified with (His)$_6$-tagged *Lla*HisZ, via immobilised metal affinity chromatography (IMAC), was exploited in the development of this purification scheme.[5]

2.2.1 Cloning of *LlahisZGs*

Synthetic genes encoding *Lla*HisG$_S$ and *Lla*HisZ, codon optimised for expression in *E. coli*, were purchased from Epoch Life Science. The synthetic genes were provided in standard pBSK vectors and sequence verified prior to delivery. *LlahisG$_S$* and *LlahisZ* were subcloned into pET expression vectors using the In-Fusion® HD Cloning Kit (Clontech). The In-Fusion® enzyme catalyses the recombination of a linearised vector and a PCR product by the precise

recognition of a 15 base pair overlap sequence, complementary to the sequence of the vector (Figure 2.3).

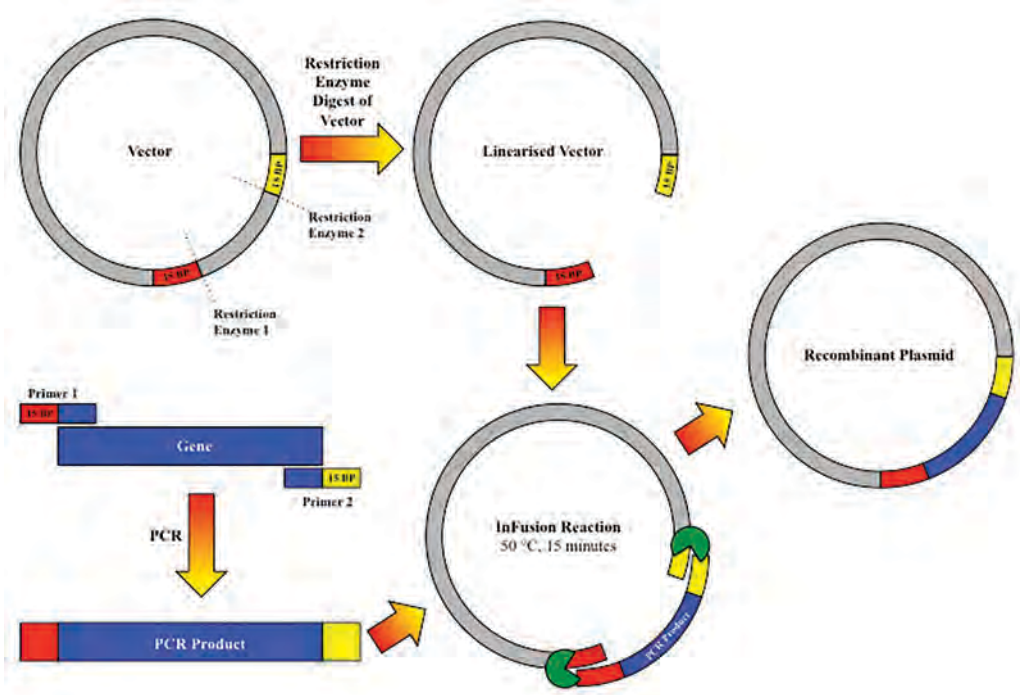

Figure 2.3 Schematic of the In-Fusion® HD cloning protocol. The destination vector can be linearised by restriction enzyme digest or by inverse PCR. The gene of interest is amplified with primers that include a sequence of 15 base pairs complementary to the vector sequence (represented here by red and yellow extensions). The In-Fusion® enzyme (green) recombines the PCR product with the linearised vector, resulting in the formation of a recombinant plasmid. The In-Fusion® reaction mixture is then transformed into competent *E. coli* cells and plated on agar plates containing the appropriate media and antibiotics.

LlahisZ and *LlahisGs* amplified beautifully from their respective pBSK vectors, using primers designed to include the 15 base pair recombination sequence. *LlahisZ* was subcloned into pET-28a, linearised with EcoRI and NdeI, to produce a construct with an N-terminal $(His)_6$-tag. The recognition site for tobacco etch virus (TEV) protease was incorporated into the synthetic gene sequence of *LlahisZ,* to allow subsequent cleavage of the $(His)_6$-tag. *LlahisGs* was subcloned into pET-21a, digested with NdeI and XhoI, to form an untagged construct. *LlahisGs* was also subcloned into pET-28a with an N-terminal $(His)_6$-tag, to allow purification of the protein by IMAC in

the absence of *Lla*HisZ. Each of the clones that were screened via colony PCR was found to be positive. The linearised vectors were gel purified prior to incubation with the In-Fusion® enzyme, and religation of the vector in the absence of insert is unlikely, as the In-Fusion® system is a ligase-free cloning procedure.

2.2.2 Expression and Purification of the *Lla*HisZG$_S$ Complex

An appreciable yield of soluble *Lla*HisZ was obtained from expression in *E. coli* BL21*(DE3) cells, in the absence of *Lla*HisG$_S$. In contrast, *Lla*HisG$_S$ was predominantly found in the cellular debris following cell lysis. A common strategy for improving the solubility of recombinant proteins is co-expression with chaperonins, to promote chaperone-mediated folding of the protein.[63,64] Expression of *Lla*HisG$_S$ in *E. coli* BL21*(DE3) pBB540/pBB542 (Chaperone 3) cells led to a significant improvement in the yield of soluble protein. Chaperone 3 cells contain plasmid-borne genes encoding groESL, clpB, grpE, dnsK, and dnaJ. Cultures of both *Lla*HisG$_S$ and *Lla*HisZ were grown at 37 °C until an optical density of 0.4-0.6. At this point, protein expression was induced by the addition of isopropyl-1-thiol-β-D-galactopyranoside (IPTG) to a final concentration of 0.5 mM. Following induction, the cultures were incubated at 23 °C overnight and harvested by centrifugation.

The buffers utilised in the purification of the *Lla*HisZG$_S$ complex were adapted from the previously published protocol.[†,39] The *Lla*HisZG$_S$ complex was purified by the combination of cell pellets of *Lla*HisZ and *Lla*HisG$_S$ in a 1:1 (v/v) ratio of cell cultures.

[†] For a detailed description of this and other methods, refer to chapter six.

Unsurprisingly, ATP-PRTase activity was detected in the cell lysate of the combined *Lla*HisZ and *Lla*HisG$_S$ but was absent from the cell lysate of *Lla*HisG$_S$ expressed independently. This finding is consistent with previous reports that HisZ is required for catalysis.[5]

The supernatant was applied to an IMAC column and the (His)$_6$-tagged *Lla*HisZG$_S$ enzyme eluted, as a single peak, with an imidazole gradient. TEV protease was employed to cleave the (His)$_6$-tag from *Lla*HisZ. Several attempts were made to cleave the (His)$_6$-tag, varying both the temperature (4 °C to 37 °C) and duration (1 hour to 2 days) of the incubation, as well as the ratio of TEV protease to *Lla*HisZG$_S$. Cleavage of the tag was incomplete under all conditions tested and the resulting mixture of tagged and untagged protein could not be resolved by IMAC or size exclusion chromatography (SEC). The decision was made to omit the (His)$_6$-tag cleavage step and proceed directly to SEC, the final polishing step of the purification. The presence of the (His)$_6$-tag had no impact ATP-PRTase activity (section 2.3).

SEC separates proteins based on size, with large proteins eluting earlier than small proteins. *Lla*HisZG$_S$ elutes as three peaks from the preparative SEC (Figure 2.4). SDS-PAGE analysis showed that the first and second peaks both contain *Lla*HisZ and *Lla*HisG$_S$ in approximately equal proportion (Figure 2.4). The first peak is unlikely to represent a biologically relevant species. The early elution volume is consistent with a complex significantly larger than the hetero-octamer, which is believed to be the physiological form. In addition, the band at 60 kDa on the SDS-PAGE gel indicates the presence of a stoichiometric amount of a chaperonin protein, suggesting that *Lla*HisZG$_S$ may be misfolded, as chaperonins have a high affinity for misfolded

proteins.[65,66] The third peak corresponds to an excess of *Lla*HisZ, however the presence of trace amounts of *Lla*HisG$_S$ prohibited the use of this protein for the independent characterization of *Lla*HisZ. The first and third peaks were excluded from the purification and the fractions corresponding to the second peak were pooled to yield 50-100 mg of *Lla*HisZG$_S$ per litre of cell culture.

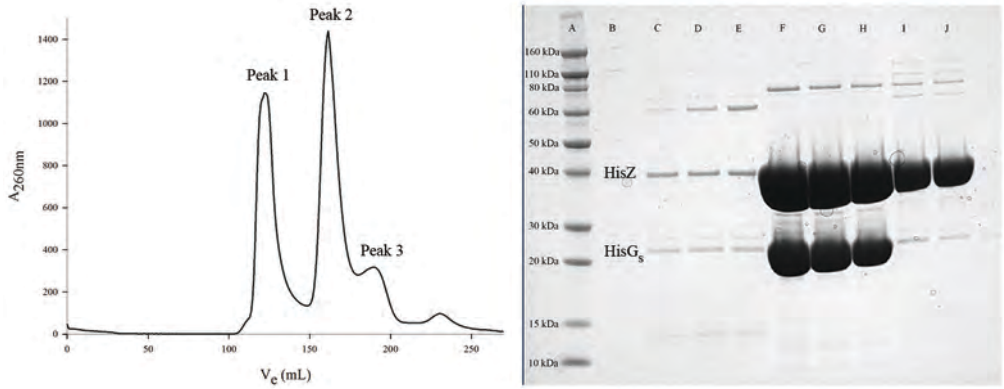

Figure 2.4 SEC trace of *Lla*HisZG$_S$ (left) showing three peaks. Note that the A_{260nm} trace is shown instead of A_{280nm}, which exceeded the detection limit of the instrument. An SDS-PAGE gel (right) shows the fractions corresponding to each peak on the SEC trace. Lane A is a molecular weight standard, lanes C-E show peak 1, lanes F-H represent peak 2, and lanes I and F correspond to peak 3. The band corresponding to 60 kDa in lanes C-E is indicative of a chaperonin protein, which suggests that *Lla*HisZG$_S$ is not correctly folded. The high molecular weight band present in lanes F-H is due to the presence of HisZ dimer, caused by insufficient heat treatment of the sample.

2.2.3 Expression and Purification of *Lla*HisZ

The purification of *Lla*HisZ was relatively facile (Figure 2.5). The same methodology was employed as for *Lla*HisZG$_S$, described in section 2.2.2, with the omission of *Lla*HisG$_S$. The final yield of *Lla*HisZ was 36 mg per litre of cell culture.

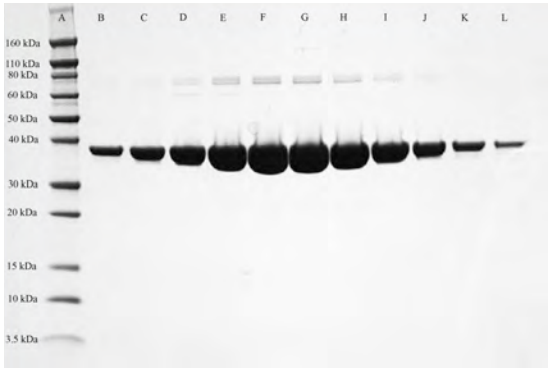

Figure 2.5 SDS-PAGE gel showing the purified *Lla*HisZ protein. The higher molecular weight band visible in lanes C-J corresponds to HisZ dimer, due to insufficient heat treatment of the sample.

2.2.4 Attempted purification of *Lla*HisG$_S$

The purification of *Lla*HisG$_S$ was attempted both untagged, via anion exchange chromatography, and with an N-terminal $(His)_6$-tag via IMAC. Chromatography was performed both at room temperature and at 4 °C and the composition of the purification buffers was varied by altering the concentration of NaCl (0-500 mM), TCEP (200 μM – 1 mM), and by the addition of glycerol (5 % v/v). The protein precipitated during the first chromatography step under all conditions tested. Optimisation of the purification protocol was not pursued due to time constraints.

2.3 Kinetic Characterisation of *Lla*HisZG$_S$

Kinetic characterisation of *Lla*HisZG$_S$ was performed using an established enzyme assay that detects the formation of PR-ATP at 290 nm.[26] The K_M^{PRPP}, K_M^{ATP}, and k_{cat} values determined for this preparation are comparable to the kinetic parameters reported by Champagne *et al.* (Table 2.2) (Figure 2.6).[26] The specific activity was also evaluated in the presence of varying concentrations of histidine,

and was found to be consistent with the K_i^{His} value of 81 μM reported by Champagne *et al* (Figure 2.7).[26]

Table 2.2 Kinetic Parameters of *Lla*HisZGs

	K_M^{ATP} (mM)	K_M^{PRPP} (μM)	k_{cat} (s^{-1})	Reference
*Lla*HisZG$_S$	2.9 ± 0.4	16 ± 3	2.5 ± 0.2	This work
*Lla*HisZG$_S$	2.7 ± 0.3	18 ± 4	2.7 ± 0.3	Champagne *et al*.[26]

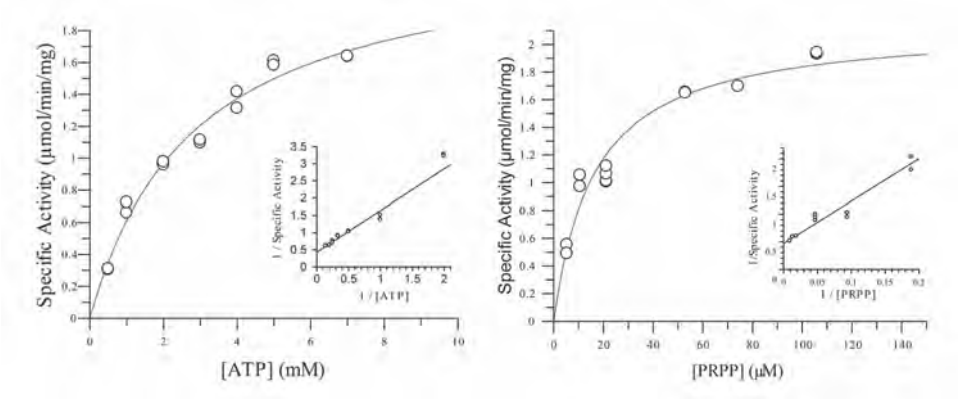

Figure 2.6 Michaelis-Menten curves of *Lla*HisZG$_S$. The K_M^{ATP} (left) was determined by holding the [PRPP] constant at 100 μM while varying the [ATP]. To determine the K_M^{PRPP} (right) the [ATP] was held constant at 10 mM and the [PRPP] was varied. The Lineweaver-Burk plots are shown as inserts.

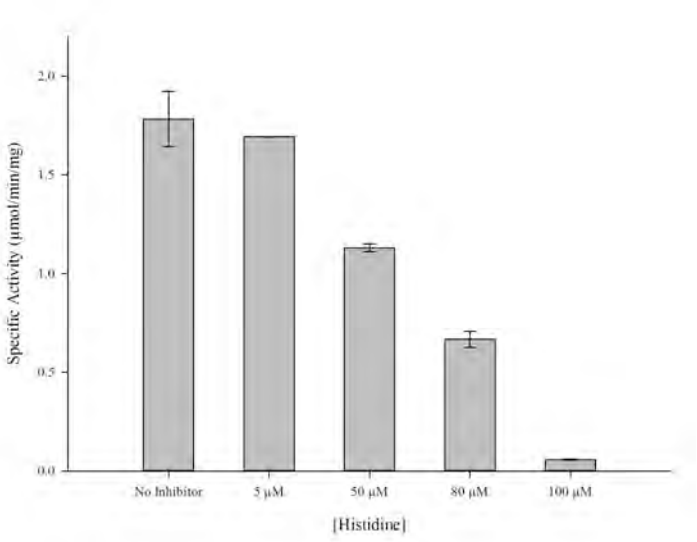

Figure 2.7 The variation in the specific activity of *Lla*HisZG$_S$ with the addition of the allosteric inhibitor histidine. The [PRPP] and [ATP] were held constant at 50 μM and 10 mM respectively.

2.4 Determination of the Quaternary Structure of *Lla*HisZG$_S$ and *Lla*HisZ by Analytical SEC

Analytical SEC is a common technique for determining the molecular weight of proteins in solution.[67] The protein standards thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), conalbumin (75 kDa), and ovalbumin (44 kDa) were used to construct a calibration curve by plotting the log of the molecular weight against the normalised elution volume (V_N) (Figure 2.8). V_N is equal to the quotient of the elution volume of the sample and the void volume of the column, defined by the elution volume of blue dextran (2,000 kDa).

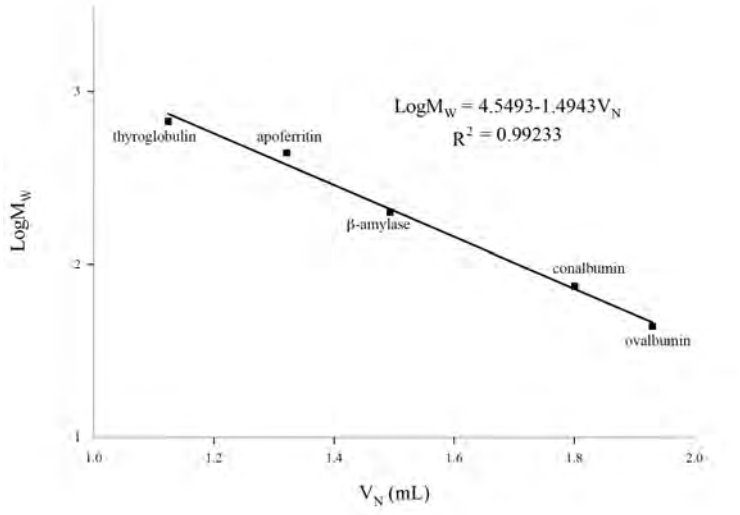

Figure 2.8 Analytical SEC calibration curve formed by plotting the log of the molecular weight against the normalised elution volume of the protein standards.

2.4.1 Characterisation of *Lla*HisZG$_S$ by Analytical SEC

In the absence of ligands, *Lla*HisZG$_S$ eluted as two peaks. The main peak has a molecular weight consistent with the *Lla*HisZG$_S$ hetero-octamer, with a small shoulder, corresponding to a HisZ dimer

(Table 2.3) (Figure 2.9). To determine whether the presence of the shoulder reflects an equilibrium between oligomeric states, the protein corresponding to the hetero-octamer was collected and reapplied to the column, without dilution or concentration. The sample eluted as a single peak, consistent with the *Lla*HisZG$_S$ hetero-octamer (Figure 2.9). Thus the presence of the HisZ dimer was attributed to a slight excess of HisZ in the purification due to overlapping peaks in the preparative SEC. This could potentially be alleviated by the addition of a third chromatography step, such as hydroxyapatite chromatography.

Table 2.3 *Lla*HisZG$_S$ Analytical SEC Data

	V_e (mL)	V_N	LogM_w	M_w (kDa)	Oligomeric State
*Lla*HisZG$_S$ main peak	11.3	1.47	2.35	224	Hetero-octamer
*Lla*HisZG$_S$ shoulder	13.4	1.74	1.94	88.0	HisZ dimer
*Lla*HisZG$_S$ collected main peak	11.3	1.47	2.35	222	Hetero-octamer
*Lla*HisZG$_S$ + Histidine	11.4	1.49	2.33	214	Hetero-octamer

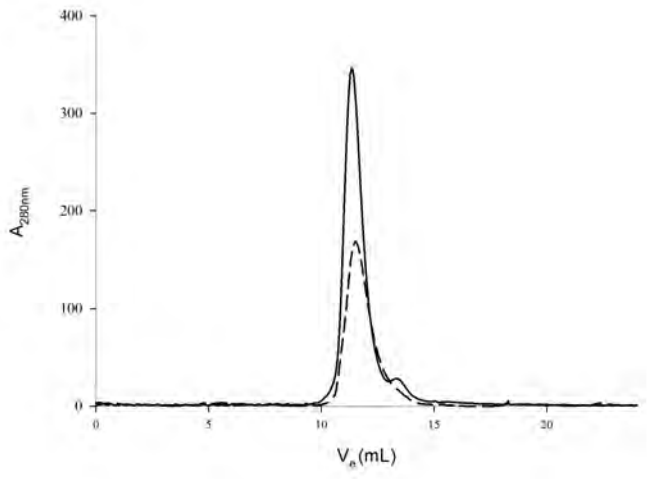

Figure 2.9 Analytical SEC trace of *Lla*HisZG$_S$ at a concentration of 1mg/mL (solid line). The fractions corresponding to the main peak were reapplied to the column, and eluted as a single species (dashed line).

In the presence of 1 mM histidine, *Lla*HisZG$_S$ eluted as a single, hetero-octameric species. Comparison with the data collected in the absence of histidine revealed a slight shift in the peak (Table 2.3)

(Figure 2.10). This suggests that histidine binding may induce a conformational change, without affecting the subunit composition of the protein.

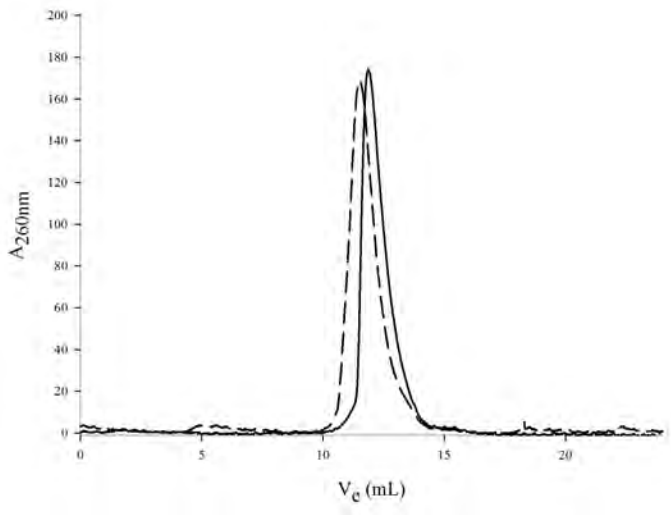

Figure 2.10 Analytical SEC traces of *Lla*HisZG$_S$ (1mg/mL) in the presence (solid line) and absence (dashed line) of 1 mM histidine. A slight shift in the peak suggests that histidine binding cause a slight compression of the *Lla*HisZG$_S$ complex relative to the unliganded species.

2.4.2 Characterisation of *Lla*HisZ by Analytical SEC

Contrary to previous reports of a *Tma*HisZ dimer-tetramer equilibrium, described by Vega *et al.*,[19] *Lla*HisZ eluted as a single species, consistent with a HisZ dimer, at concentrations of both 1 mg/mL and 4 mg/mL (Table 2.4) (Figure 2.11). This supports the previous proposal by Sissler *et al.*[5] of a dimeric *Lla*HisZ and suggests that the X-shaped core of four *Lla*HisZ chains, found in the hetero-octameric structure, is stabilised only in the presence of *Lla*HisG$_S$.

Table 2.4 *Lla*HisZ Analytical SEC Data

	V_e (mL)	V_N	LogM_w	M_w (kDa)	Oligomeric State
*Lla*HisZ (1mg/mL)	13.2	1.72	1.98	94.6	Dimer
*Lla*HisZ (4 mg/mL)	13.1	1.71	1.99	97.6	Dimer

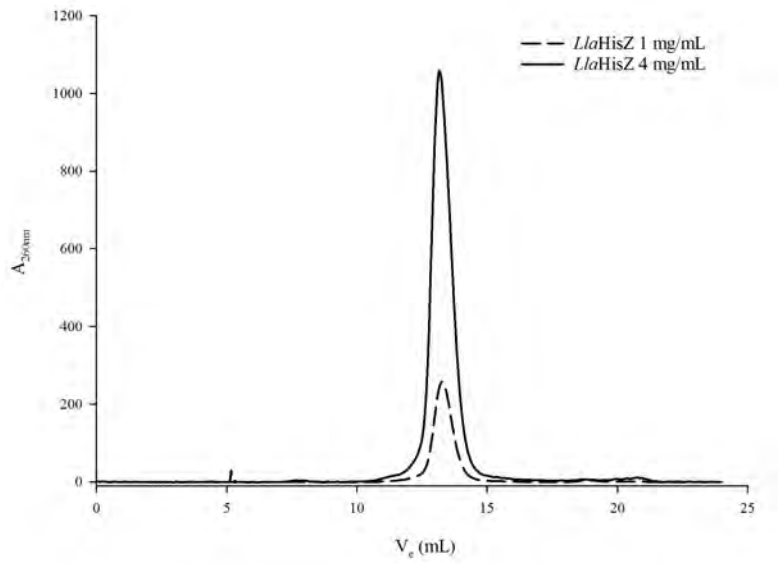

Figure 2.11 Analytical SEC trace of *Lla*HisZ at a concentration of 4mg/mL (solid line) and 1 mg/mL (dashed line). Note that the A_{260nm} trace is shown as the A_{280nm} trace for the 4 mg/mL sample exceeded the detection limit, resulting in a flattened peak. There is no evidence of a tetrameric species observed in either sample.

2.5 Characterisation of *Lla*HisZG$_S$ by Small Angle X-ray Scattering

Small angle X-ray scattering (SAXS) was employed to verify the solution structure of *Lla*HisZG$_S$ (1 mg/mL) in the presence and absence of 1 mM histidine. SEC-SAXS was performed to ensure that the scattering data was collected from the hetero-octameric species. Guinier analysis of the *Lla*HisZG$_S$ scattering profile in the absence of histidine revealed a radius of gyration (R_g) of 41.4 ± 0.2 Å.[68] The pair distribution (P(r)) function was determined via indirect Fourier transformation. The R_g and the maximum dimension (D_{max}) of the particle were defined by P(r) as 41.0 ± 0.2 Å and 146 Å respectively. The particle volume defined by the Porod invariant (Porod volume) was equal to 435000 Å^3.

Overlay of the SAXS scattering profiles revealed a notable difference in the presence of histidine (Figure 2.12). The R_g of the *Lla*HisZG$_S$ + histidine scattering profile was calculated to be 40.9 ± 0.2 Å by Guinier approximation and 40.4 ± 0.1 Å by the P(r) function. The protein in the presence of histidine has a D_{max} of 117 Å and a Porod volume of 323000 Å^3. This is consistent with the histidine induced change observed by analytical SEC (Section 2.4.1).

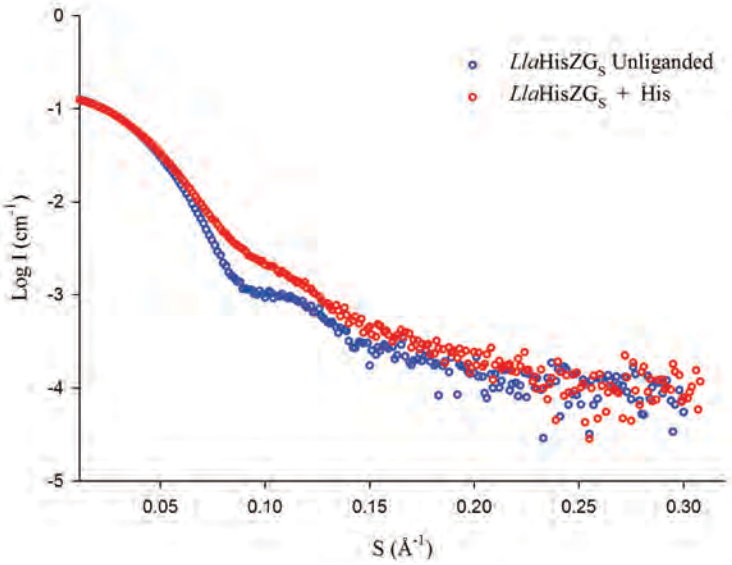

Figure 2.12 Overlay of the SAXS scattering profiles of *Lla*HisZG$_S$ in the presence (red) and absence (blue) of 1 mM histidine.

.

2.5.1 Crysol Fitting of the *Lla*HisZG$_S$ SAXS Profiles to the X-ray Crystal Structure

The difference between the protein conformation in the presence and absence of histidine is more pronounced when compared to the X-ray crystal structure. Crysol was used to deduce the theoretical scattering curve from the X-ray crystal structure of *Lla*HisZG$_S$ (PDB 1Z7N) and calculate the fit to the experimentally determined SAXS profiles.[69] In the absence of histidine, the *Lla*HisZG$_S$ scattering profile

fits the theoretical scattering profile calculated from the X-ray crystal structure (PDB 1Z7N, χ^2 = 1.1) (Figure 2.13). This suggests that the X-ray crystal structure is a reasonable representation of the conformation adopted by aqueous *Lla*HisZG$_S$ (1 mg/mL) in the absence of ligands.

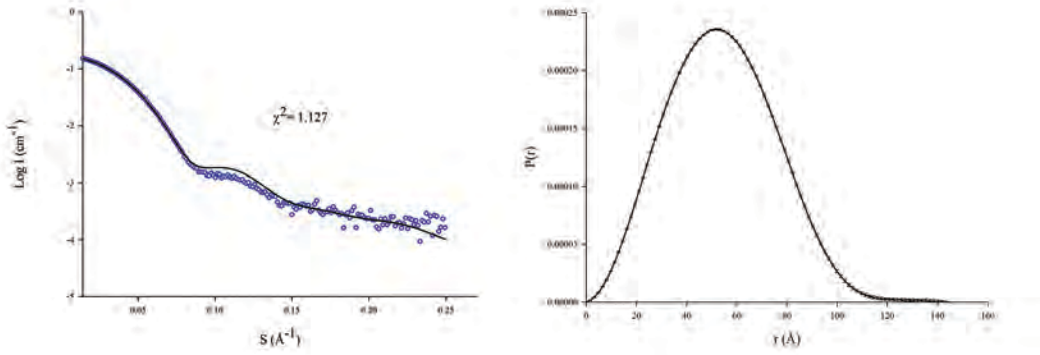

Figure 2.13 (left) Crysol fit of the *Lla*HisZG$_S$ SAXS profile (blue circles) to the theoretical scattering profile calculated from the crystal structure (PDB 1Z7N) (black line). The χ2 value of 1.1 indicates a good fit. The P(r) function of the *Lla*HisZG$_S$ SAXS profile is shown on the right.

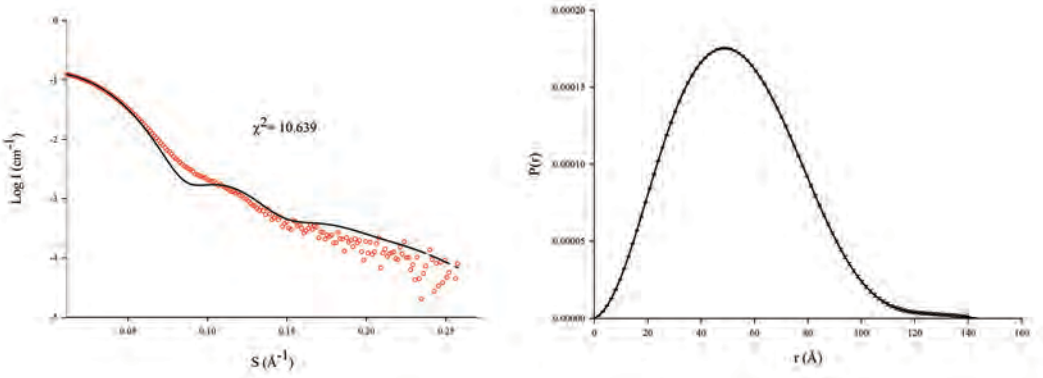

Figure 2.14 (left) Crysol fit of the *Lla*HisZG$_S$ + histidine SAXS profile (red circles) to the theoretical scattering profile calculated from the crystal structure (PDB 1Z7N) (black line). The χ2 value of 10.6 demonstrates that the crystal structure is not good representation of the histidine-inhibited *Lla*HisZG$_S$. The P(r) function of the *Lla*HisZG$_S$ + histidine SAXS profile is shown on the right.

In the presence of histidine, however, the experimental scattering profile of *Lla*HisZG$_S$ differs significantly to the theoretical scattering curve calculated from the crystal structure (χ^2 = 10.6) (Figure 2.14). This observation is consistent with a conformational

change in response to histidine binding, and hints at a potential mechanism for allosteric inhibition of the enzyme.

Slight differences in the *Lla*HisZG$_S$ scattering data compared to the crystal structure may be attributed to the presence of an N-terminal (His)$_6$-tag on *Lla*HisZ. It must also be noted that the crystal structure is incomplete, as the first five N-terminal and last five C-terminal residues are missing from *Lla*HisZ. The structure reported by Champagne *et al.* also has three mutations in *Lla*HisG$_S$ (A51P, D86Y, and R110H) and four in *Lla*HisZ (N73K, H174Q, E209G, and V303F) compared to the reported sequence of *Lla*HisZG$_S$ from *L. lactis* subsp. *lactis* KF147, from which the genes were originally cloned. The synthetic genes provided by Epoch Life Sciences are consistent with the reported sequence of HisZG$_S$ from *L. lactis* subsp. *lactis* KF147.

The experimental SAXS data was also compared to the other *Lla*HisZG$_S$ crystal structure (PDB 1Z7M) as well as the *Tma*HisZG$_S$ structure (PDB 1USY) (Table 2.5).‡ The results suggest that the available *Lla*HisZG$_S$ crystal structures are reasonable representations of the *Lla*HisZG$_S$ envelope in the absence of histidine, but are unlikely to reflect the structure adopted by histidine bound *Lla*HisZG$_S$. The histidine bound *Tma*HisZG$_S$ structure resembles neither the average conformation adopted by *Lla*HisZG$_S$ in the presence or the absence of histidine.

‡ See Appendix I for plots of these fits.

Table 2.5 χ^2 Values Obtained by Fitting the *Lla*HisZG$_S$ SAXS Data to the Available HisZG$_S$ Crystal Structures

Crystal Structure (PDB)	*Lla*HisZG$_S$	*Lla*HisZG$_S$ + Histidine
*Lla*HisZG$_S$ (1Z7N)	1.1	11
*Lla*HisZG$_S$ (1Z7M)	0.87	6.8
*Tma*HisZG$_S$ (1USY)	5.5	6.2

2.5.2 *Ab Initio* Modelling to the *Lla*HisZG$_S$ + Histidine SAXS Data

In lieu of a histidine bound *Lla*HisZG$_S$ crystal structure, *ab initio* modelling was performed to glean information about the shape of the protein in the presence of histidine. *Ab initio* modelling generates low resolution envelope structures by optimising the agreement between the experimental data and the theoretical scattering curves (or P(r) functions), calculated from the proposed models.[70] In order to reduce the search space, the shape of the protein is modelled as a continuous object.

Gasbor was used to generate *ab initio* models of the histidine bound *Lla*HisZG$_S$. The *ab initio* models produced by Gasbor comprise chain-line ensembles of dummy residues, whose spatial positioning approximates the Cα atoms of the protein backbone.[71] 20 *ab initio* models with P1 symmetry were generated by Gasbor. These models were averaged with the Damaver program suite.[72] Supcomb was used to superimpose the crystal structure of *Lla*HisZG$_S$ (PDB 1Z7N) on to the most typical model; identified by Damaver as having the lowest normalised spatial discrepancy (Figure 2.15).[73] The process was repeated to generate models with P2 and P4 symmetry, as imposing

symmetry on *ab initio* models may improve the quality of the model.[70] Crysol was employed to compare the average models to the scattering profile of *Lla*HisZG$_S$ + histidine (Figure 2.16). Crysol fits indicated no significant difference between the models generated with P1 ($\chi^2 = 0.35$), P2 ($\chi^2 = 0.36$), or P4 symmetry ($\chi^2 = 0.29$), and the low χ^2 values indicate a requirement for further refinement.

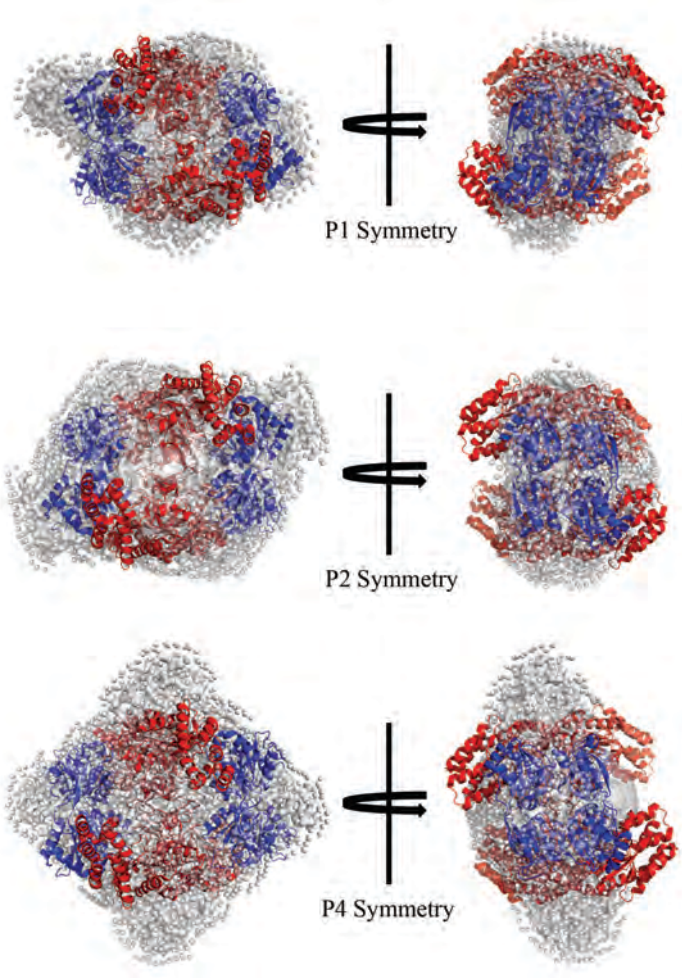

Figure 2.15 *Ab initio* models of histidine bound *Lla*HisZG$_S$ generated with Gasbor and averaged using Damaver (grey surface representation). The *Lla*HisZG$_S$ crystal structure (PDB 1Z7N) is shown superimposed on these models (cartoon representation with four *Lla*HisG$_S$ chains coloured blue and four *Lla*HisZ chains coloured red).

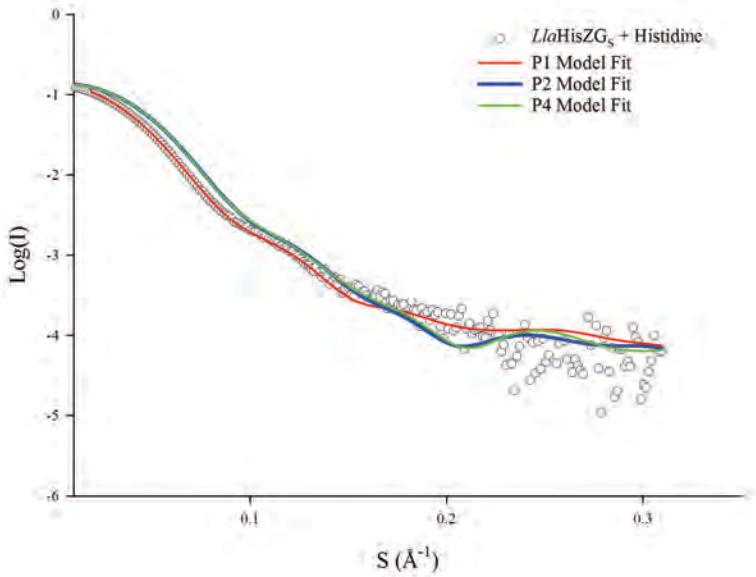

Figure 2.16 Crysol fits of the Gasbor models of P1 symmetry (red), P2 symmetry (blue) and P4 symmetry (green) to the experimental scattering (black circles).

2.5.3 Rigid Body Modelling to the *Lla*HisZG$_S$ + Histidine SAXS Data

The difference between the proposed shapes of the histidine bound *Lla*HisZG$_S$, and the unliganded enzyme, may reflect a change in the relative orientation of the *Lla*HisZ and *Lla*HisG$_S$ subunits. Atomic resolution structures of the *Lla*HisZ and *Lla*HisG$_S$ chains are available; therefore the relative position of these chains could potentially be resolved by rigid body modelling. The general method is reminiscent of *ab initio* modelling, whereby the final model is calculated by optimising the agreement between the experimental scattering and the theoretical scattering from the proposed structures.[70] In rigid body modelling, however, the atomic resolution structures are used in place of dummy residues. These structures are considered to be static objects and only their relative spatial arrangement is altered.

Coral was used to construct models of the histidine bound *Lla*HisZG$_S$.[74] Individual PDB files of each of the eight chains in the *Lla*HisZG$_S$ crystal structure were created and the *Lla*HisZ and *Lla*HisG$_S$ dimer interfaces were fixed. These interfaces are strong, $\Delta G = -102$ kJmol^{-1} and -58 kJmol^{-1} respectively, as calculated by PDBePISA.[75] It is a reasonable first assumption that these interfaces remain largely intact upon histidine binding, although slight conformational changes at this interface are possible. The most likely conformations of *Lla*HisZG$_S$ were obtained by running Coral with P1 or P2 symmetry (Figure 2.17) as imposing P4 symmetry resulted in disruption of the *Lla*HisZG$_S$ complex and χ^2 values greater than 10 when fitted to the experimental data. Analytical SEC revealed that, under the conditions of the SAXS experiment, the hetero-octameric complex remains intact upon histidine binding.

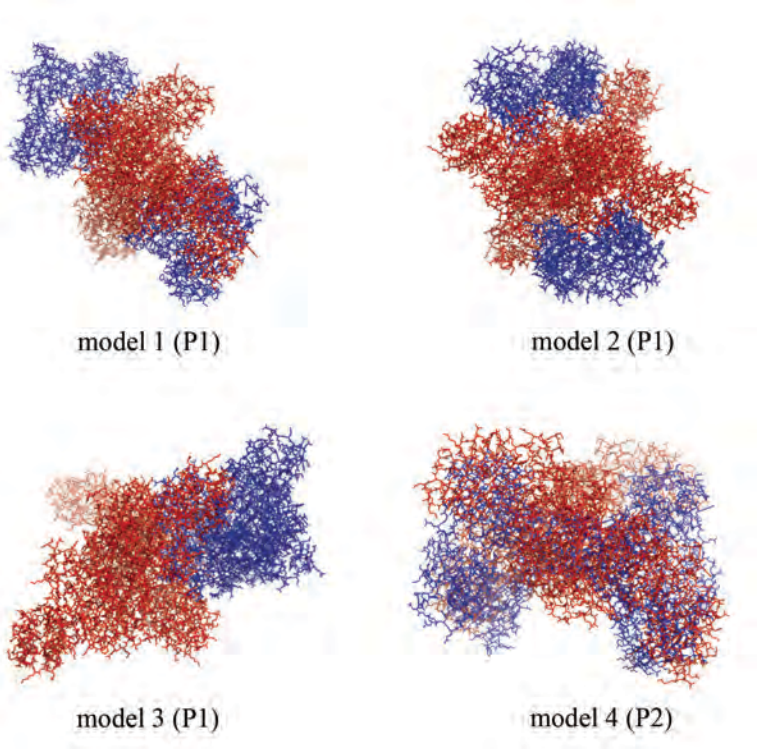

Figure 2.17 The four best models of histidine-bound *Lla*HisZG$_S$ generated by Coral. The *Lla*HisZ chains are coloured red and the *Lla*HisG$_S$ chains are coloured blue.

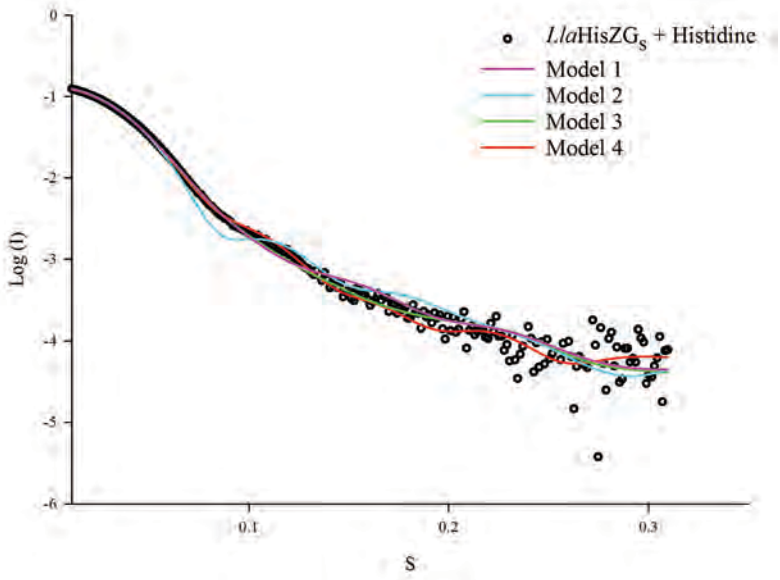

Figure 2.18 Crysol fitting of the rigid body models produced by Coral to the experimental scattering of *Lla*HisZGs. The best fit defined by Coral is Model 1 (pink, $\chi 2 = 0.79$) followed by Model 3 (green, $\chi 2 = 0.47$), Model 4 (red, $\chi 2 = 0.42$) and Model 2 (blue, $\chi 2 = 11$).

Crysol revealed a large variation in the quality of the fit of the theoretical scattering profiles to the experimental data, between each of the four models (Figure 2.18). This signifies that, while rigid body modelling has proved successful for the structural characterisation of many proteins,[70] further information is required to construct a convincing model of the histidine bound *Lla*HisZG$_S$. Model 2 most closely resembles the known conformation of the hetero-octameric HisZG$_S$, however the fit to the experimental data is poor, as evidenced by a χ^2 value of 11. The model that best reflects the experimental data, model I ($\chi^2 = 0.79$), indicates a possible change in the relative position of the *Lla*HisZ subunits, as well as a change in the *Lla*HisZ-*Lla*HisG$_S$ interface. This is consistent with elongation of the protein complex in response to histidine binding, however significant disruption of protein-protein interfaces unlikely. Rigid body modelling does not account for conformational changes within the individual protein subunits in response to ligand binding and work towards

understanding the conformational response to histidine binding is continuing.

2.6 Induced Fit Docking of Histidine into Three Potential Binding Sites on *Lla*HisZG$_S$

Three distinct histidine binding sites have been proposed for *Lla*HisZG$_S$. The first (designated site one) is located in a region highly conserved between HisZ and histidyl-tRNA synthetase. Champagne *et al*. used mutagenesis to attribute histidine binding to three conserved residues at this site (Glu130, Tyr268 and Tyr269).[20] Whether these represent the histidine binding residues, or if mutation of these residues disrupted the allosteric signal transduction pathway, is yet to be deduced. Sites two and three reside at the interface between *Lla*HisZ and *Lla*HisG$_S$ and were proposed based on comparison with the histidine-bound crystal structure of *Tma*HisZG$_S$ (PDB 1USY).[19] Induced fit docking of histidine into each of the three potential binding sites on the *Lla*HisZG$_S$ complex was performed by Dr Wanting Jiao, with the intention of identifying the biologically relevant histidine binding sites(s). An ensemble of histidine poses was generated and Prime structure prediction was used to accommodate the ligand by reorienting side chains in the proposed binding site. The ligand was then re-docked into the lowest energy protein structure and poses were ranked by their Glide docking score. The Glide score is an empirical measure that approximates the free energy of ligand binding.[76,77] Analysis of these models is presented in this section.

2.6.1 Histidine Docking into Site One of *Lla*HisZG$_S$

Preparation of the *Lla*HisZG$_S$ complex for induced fit docking resulted in Glu130 being assigned to a neutral state, instead of the negatively charged state expected of this residue. It is unclear why this occurred, therefore the induced fit docking of histidine into site one was evaluated under both negative and neutral charge states. The top scored pose of histidine docked into site one with a neutral Glu130 has a Glide docking score of –6.009. This pose reveals no significant interaction between histidine and the histidine-binding residues proposed by Champagne *et al.*[20] (Figure 2.19). The Tyr268 and Tyr269 side chains both orient away from the bound histidine. Glu130 lies in close proximity to histidine, however there is no evidence to suggest that this residue is involved in stabilisation of the ligand in the binding pocket. This pose highlights Lys265, Asp84, and Tyr291 as potential histidine binding residues. The carboxylate group of histidine forms both a hydrogen bond and a salt bridge with the Lys265 side chain. A salt bridge is also formed between the amino group of histidine and the side chain of Asp84 and pi-pi stacking interactions are established with the side chain of Tyr291.

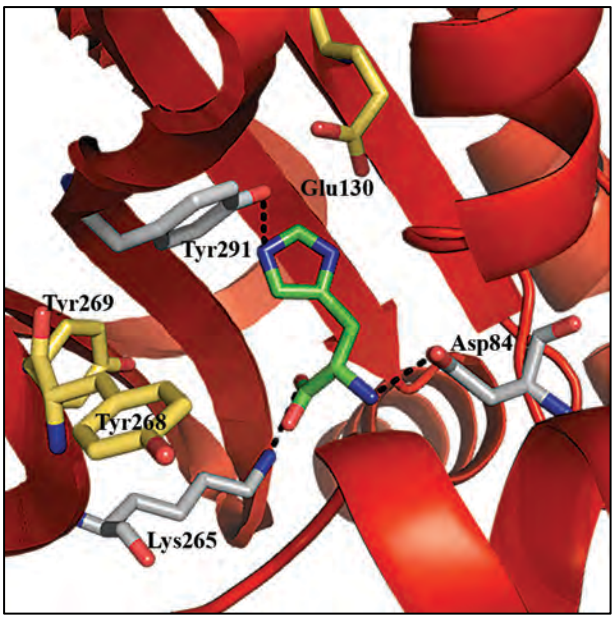

Figure 2.19 The highest scored docking pose of histidine in site one of *Lla*HisZG$_S$ with a neutral Glu130. The proposed histidine binding residues determined by mutagenesis (yellow sticks) show no interaction with the histidine ligand (green). Grey sticks are used to show the histidine binding residues in this pose. The cartoon structure of *Lla*HisZ is coloured red.

When histidine was docked into site one with a negative Glu130, the highest scored pose had a docking score of –6.041. Histidine adopts a different conformation in this pose, compared to the previously described pose, which suggests that this site is sufficiently spacious to accommodate histidine binding in different orientations (Figure 2.20). The imidazole ring of histidine forms pi-pi stacking interactions with Tyr291 and the side chain of Asp84 forms a salt bridge with the histidine amino group, as described previously. The carboxylate group of histidine also interacts with the side chains of Lys265 and Gln126. The three proposed histidine binding residues again show no interaction with the ligand.

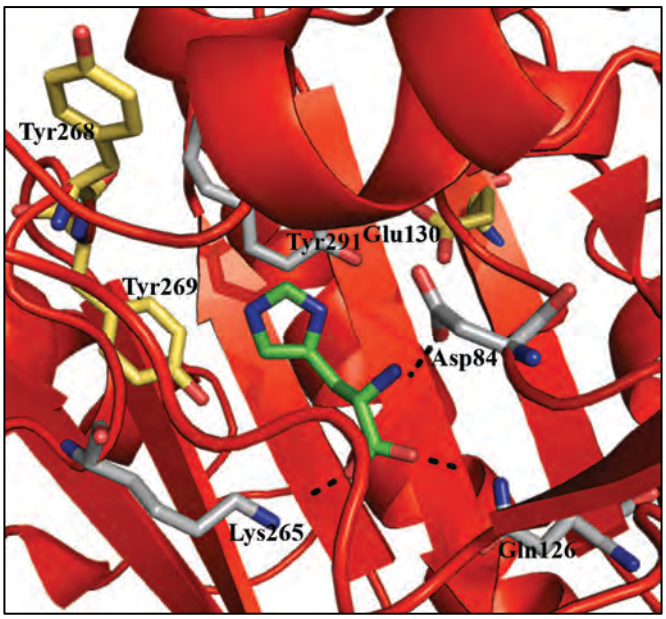

Figure 2.20 The highest scored docking pose of histidine docked into site one with a negative Glu130. *Lla*HisZ is coloured red, the previously proposed histidine binding residues are shown in yellow, and grey sticks represent the histidine binding residues.

2.6.2 Histidine Docking into Site Two of *Lla*HisZG$_S$

Docking of histidine into potential binding site two of *Lla*HisZG$_S$, at the interface of *Lla*HisZ and *Lla*HisG$_S$, resulted in a pose with a docking score of -8.071 (Figure 2.21). Hydrogen bonding is observed between the carboxylate group of histidine and the side chains of Lys213 and Asn196 from *Lla*HisZ. The backbone amino group of Lys213 also appears to stabilise the imidazole group of histidine by hydrogen bonding. The amino group of histidine forms salt bridges with Glu200 and Glu203 from *Lla*HisG$_S$.

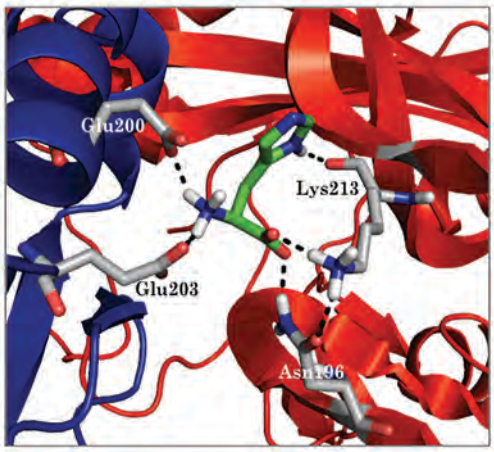

Figure 2.21 Histidine (green) docked into potential binding site two at the interface of *Lla*HisG$_S$ (blue) and LlaHisZ (red). The histidine binding residues are coloured grey.

2.6.3 Histidine Docking into Site Three of *Lla*HisZG$_S$

Docking of histidine into potential binding site three of *Lla*HisZG$_S$, also located at the *Lla*HisZ–*Lla*HisG$_S$ interface, resulted in a pose with a docking score of -5.813 (Figure 2.22). In this pose, the amino group and the imidazole group of histidine establish hydrogen bonds with the side chain of Asp191 from *Lla*HisZ. The Glu83 and Lys70 of *Lla*HisG$_S$ also form salt bridges with the carboxylate group of histidine.

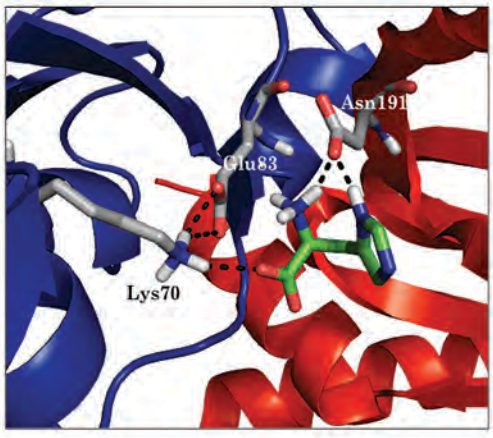

Figure 2.22 Histidine (green) docked into potential binding site three at the interface of *Lla*HisG$_S$ (blue) and *Lla*HisZ (green). The histidine binding residues are represented as grey sticks.

2.6.4 Where Does Histidine Bind?

Induced fit docking of histidine into potential binding site two, at the interface of *Lla*HisZ and *Lla*HisG$_S$, gave the best (most negative) docking score. This binding site, along with potential binding site three, was identified by comparison with the histidine-bound *Tma*HisZG$_S$ crystal structure. It is tempting to conclude that this is the functional allosteric site and that the Glu130, Tyr268 and Tyr269 residues of site one are critical for allosteric signal transmission, but not involved in ligand binding. Induced fit docking, however, was performed on the unliganded *Lla*HisZG$_S$ structure and binding of histidine has been shown to trigger a conformational change of the protein. The algorithm used for induced-fit docking compensates for changes in side chain positioning but does not account for the large conformational changes observed by SAXS.

2.7 Characterisation of Histidine Binding to *Lla*HisZG$_S$ by Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) provides a direct measure of the heat changes associated with binding events in solution. ITC was employed to assess histidine binding to both the *Lla*HisZG$_S$ complex, and the *Lla*HisZ protein independently, in order to identify the functional allosteric site.

2.7.1 Histidine Binding to the *Lla*HisZG$_S$ Complex

When a 2 mM solution of histidine was titrated into 50 μM *Lla*HisZG$_S$, the resulting binding curve fit a one-site model with a

K_D = 26.0 ± 0.9 μM and stoichiometry of 1.13 ± 0.02 (Figure 2.23). *Lla*HisZ and *Lla*HisG$_S$ were treated as a combined unit in the determination of concentration; therefore the stoichiometry reveals a total of four histidine binding sites in the hetero-octameric complex.

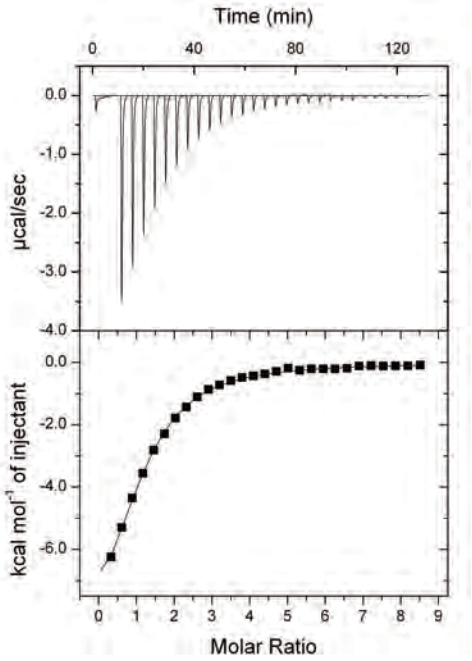

Figure 2.23 ITC data of histidine binding to the *Lla*HisZG$_S$ complex. The stoichiometry is 1.13 ± 0.02 sites and K_D = 26.0 ± 0.9 μM. ΔH = -41.4 ± 0.8 kJmol^{-1} and ΔS = -50.6 Jmol^{-1}K^{-1}

2.7.2 Histidine Binding to *Lla*HisZ

*Lla*HisZ is a histidyl-tRNA synthetase paralogue, and the sequence of proposed histidine binding site one is highly conserved (Figure 2.30). If *Lla*HisZ houses the histidine binding site, then it is conceivable that histidine may bind to *Lla*HisZ in the absence of *Lla*HisG$_S$. A 0.8 mM solution of histidine was titrated into 47 μM *Lla*HisZ, and the resulting binding curve fit a one-site model with a K_D = 16 ± 1 μM and stoichiometry of 0.46 ± 0.03. The concentration was determined per *Lla*HisZ chain; therefore the stoichiometry suggests one histidine molecule binds for every two chains. This could be due to

the low sigmoidicity of the curve, which may be improved by optimising the relative concentrations of protein and ligand. *Lla*HisZ is dimeric in solution; therefore the observed stoichiometry could alternatively indicate that histidine binds to only one chain of each dimer. Further investigation using SAXS is required to determine the conformational response of *Lla*HisZ upon histidine binding.

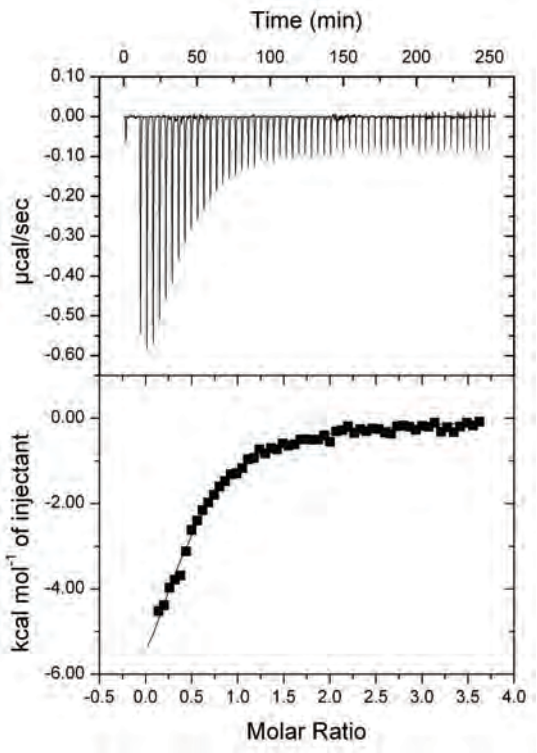

Figure 2.24 ITC data of histidine binding to *Lla*HisZ in the absence of *Lla*HisG$_S$. The stoichiometry is 0.46 ± 0.03, K_D = 16 ± 1, ΔH = -39 ± 3 kJmol^{-1} and ΔS= -40 Jmol^{-1}K^{-1}

The observation that *Lla*HisZ binds histidine in the absence of *Lla*HisG$_S$ and the similarity in the K_D values obtained for histidine binding to *Lla*HisZ and *Lla*HisZG$_S$ provides strong evidence to support site one as the functional histidine binding site.

2.8 Crystallisation Trials of *Lla*HisZG$_S$ + Histidine

The complementarity of SAXS and X-ray crystallography has been well documented, and the availability of high-resolution protein crystal structures has proved invaluable to the identification of allosteric mechanisms.[70] Numerous attempts were made to obtain an X-ray crystal structure of the histidine-bound *Lla*HisZG$_S$. Adaptation of the conditions reported for the apo-*Lla*HisZG$_S$, along with replication of the conditions reported for the histidine-bound *Tma*HisZG$_S$ proved unsuccessful. A variety of commercial screens were employed to detect conditions under which the protein crystallised. Crystals were tested for diffraction and the most promising conditions were optimised, via hanging drop vapour diffusion, by systematically varying the concentration of protein, precipitant, ligand, as well as pH, drop volume, temperature, incubation time, and the cryoprotectant. Multiple diffraction quality crystals were obtained, however the best crystals diffracted to just 4.7 Å on the MX1 beamline at the Australian Synchrotron (Figure 2.25).

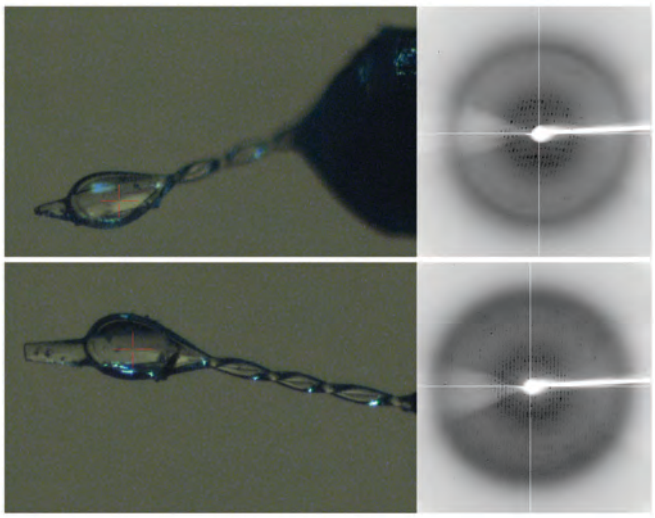

Figure 2.25 Images of the two best diffracting crystals and their corresponding diffraction patterns, taken on the MX1 beamline at the Australian Synchrotron.

2.9 Discussion

The mechanism of allosteric regulation of *Lla*HisZG$_S$ is still unclear, however the results of SAXS and ITC suggest that histidine binding at HisZ is associated with a conformational change of the protein complex. Site one has been identified as the likely functional allosteric site through a combination of the ITC presented here and the mutagenesis studies reported by Champagne *et al.*[20] This site is highly conserved between HisZ and histidyl-tRNA synthetase (Figure 2.30) and serves as the histidine binding site of *T. thermophilus* histidyl-tRNA synthetase (Figure 2.26). In contrast, Vega *et al.* attributed functional allostery to the HisZ-HisG$_S$ interface sites, as opposed to the former histidyl-tRNA synthetase active site, due to the presence of eight histidine ligands in the crystal structure, and the location of the HisZ active site buried at the core of the protein (Figure 2.27).[19] The biological relevance of the interface binding sites in *Tma*HisZG$_S$ was inferred from the proximity to the active site in HisG$_S$, presumably providing direct contacts leading to rearrangement of the active site. It has been established that transmission of allosteric signals can occur over phenomenal distances and close contact between the allosteric site and the active site is unnecessary.[78,79]

Induced fit docking favoured histidine binding to potential binding site two, an interface site, over site one. The induced fit docking algorithm does not account for the conformational change observed by SAXS. In addition, docking was performed on the unliganded X-ray crystal structure and the histidine-binding site may not be properly formed in this structure. Crystal packing may cause proteins to adopt conformations that are not biologically relevant.[70]

Allosterically regulated proteins have multiple stable conformations that are separated by small energy changes and are particularly affected by lattice forces.[70] Despite the apparent affinity of the interface binding sites for histidine, it appears unlikely that the histidine binding sites observed in the *Tma*HisZG$_S$ structure are functional.

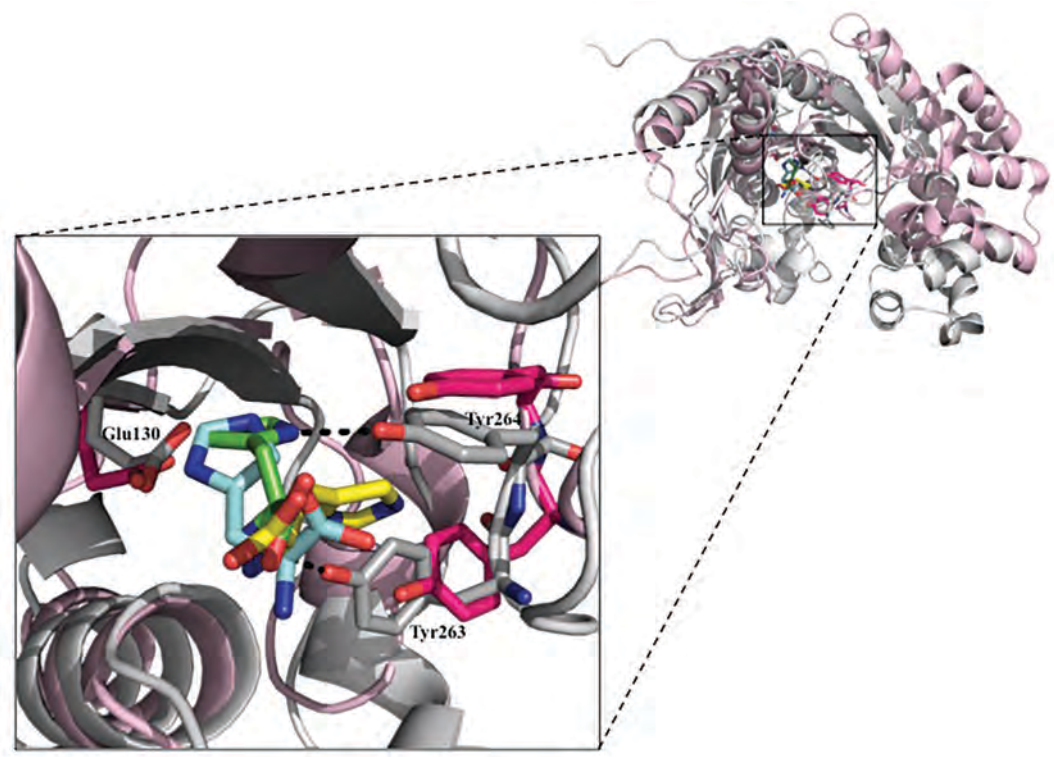

Figure 2.26 Structural alignment of *Lla*HisZ (pale pink) with histidine-bound histidyl-tRNA synthetase from *T. thermophilus* (PDB 1ADJ). The histidine ligands are shown docked into site one of *Lla*HisZ (blue and yellow sticks) with either a negative or neutral Glu130. Histidine is also shown bound in the histidyl-tRNA synthetase active site (green). The histidine binding residues (grey sticks) are shown with the comparable residues from *Lla*HisZ (pink).

Comparison of the *Lla*HisZ histidine binding site with the histidine binding site of histidyl-tRNA synthetase from *T. thermophilus* reveals a striking similarity (Figure 2.26). Although it appears that *Lla*HisZ houses the histidine binding site, the binding residues proposed by Champagne *et al.* have not been confirmed. The proposed residues (Glu130, Tyr268 and Tyr269) are highly conserved

within both the HisZ and histidyl-tRNA synthetase families and have been implicated in histidine binding based on the histidine bound *T. thermophilus* histidyl-tRNA synthetase structure.[80] The high sequence conservation and confirmed role of these residues in allosteric inhibition points to these residues as strong candidates for histidine coordination. The induced fit docking experiments, however, highlighted additional highly conserved residues as potential histidine binding residues (Tyr291, Asp84 and Gln126) that merit further investigation. Lys265 was also implicated in histidine binding however the lack of conservation of this residue within the HisZ family suggests that a crucial function of this residue is unlikely. The definition of the histidine-binding residues of this enzyme will require the generation of enzymes with loss of function point mutations and characterisation of histidine binding by ITC or saturation transfer difference NMR.

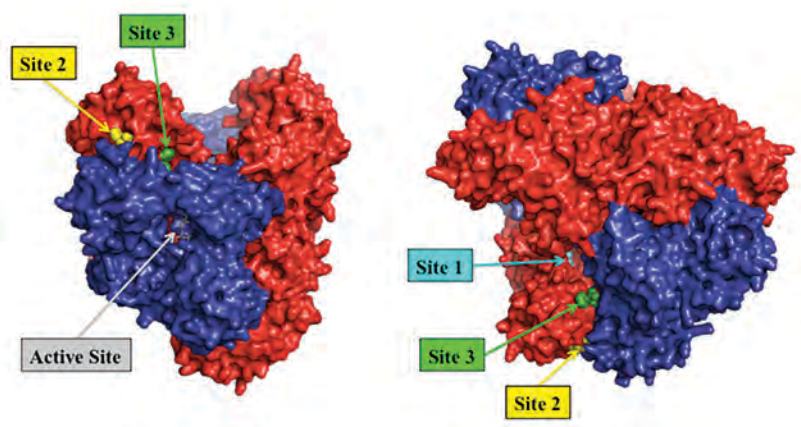

Figure 2.27 Surface representation of *Lla*HisZG$_S$ with histidine docked into each of the proposed binding sites. The four *Lla*HisZ chains are coloured red and the four *Lla*HisG$_S$ chains are shown in blue. The histidine residues are represented as spheres coloured as per the labels indicated in the figure. Sites two and three are clearly visible near the surface of the protein whilst site one is buried within the core of the protein.

The solution conformation of *Lla*HisZG$_S$ in the absence of histidine resembles the previously published *Lla*HisZG$_S$ crystal structure. A change in the SAXS scattering profile in the presence of

histidine reflects a conformational change of the enzyme associated with allosteric inhibition. In contrast, the SAXS profile of the long form ATP-PRTase from *C. jejuni* is largely unaffected by the addition of histidine (Gerd Mittelstädt personal communication). Crysol fitting demonstrated that neither the *Lla*HisZG$_S$ nor the histidine-bound *Tma*HisZG$_S$ crystal structure adequately represents the shape of the inhibited enzyme in solution. This casts doubt on the relevance of the comparison between the *Lla*HisZG$_S$ and *Tma*HisZG$_S$ structures to devise a model of allosteric regulation. It should be noted that the Crysol fit of the *Lla*HisZG$_S$ SAXS scattering profile in the absence of histidine to the histidine-bound *Tma*HisZG$_S$ crystal structure was better than the same fit with the *Lla*HisZG$_S$ + histidine scattering data. The structure adopted by *Tma*HisZG$_S$ may represent a conformation accessible to *Lla*HisZG$_S$, but not the average shape of the enzyme in solution.

The shape of the SAXS profile of histidine-bound *Lla*HisZG$_S$ suggests a more elongated structure of the inhibited enzyme. The attempts to obtain a crystal structure of the histidine-bound *Lla*HisZG$_S$ were unsuccessful and further optimisation is required. In lieu of a high resolution structure, both *ab initio* and rigid body modelling methods were employed to extract structural information from the SAXS data. The shape of the SAXS envelopes generated by Gasbor and the rigid body models generated by Coral are both consistent with an extended conformation of the protein, however neither method provided sufficient insight into the molecular details of the domain rearrangements. Model 2 generated by Coral, which most closely resembles the known HisZG$_S$ conformations, has the highest χ^2 value when compared to the experimental data. This suggests that model 2 is not an accurate representation of the inhibited enzyme. The

remaining models appear unlikely to represent an accessible conformation, as they would require significant disruption of multiple protein-protein interfaces. Analytical SEC, under the same conditions as SEC-SAXS, indicated that the inhibited *Lla*HisZG$_S$ complex is hetero-octameric. The strongest interfaces in the complex, determined by PDBePISA, are the *Lla*HisZ and *Lla*HisG$_S$ dimer interfaces, with estimated ΔG values of -96 kJmol^{-1} and -54 kJmol^{-1} respectively. These interfaces may undergo a conformational change in response to histidine binding, but are unlikely to be significantly disrupted. Rigid body modelling works by altering the relative positioning of individual domains or subunits and does not account for conformational changes within the individual chains. The weakest functional interface in the complex is the interface between the four *Lla*HisZ chains (estimated ΔG = -4 kJmol^{-1}). This interface, along with the heteromeric interfaces between HisG$_S$ and HisZ, is more likely to be susceptible to large conformational changes.

The histidyl-tRNA synthetase enzymes from *E. coli* and *T. thermophilus* are both dimeric in solution, however they crystallise with four chains in the asymmetric unit.[52,80] Interestingly, the non-biological interface structurally resembles the interface between the four HisZ chains in the *Lla*HisZG$_S$ complex (Figure 2.28). The angle between the two histidyl-tRNA synthetase enzymes in the asymmetric unit differs significantly from the angle between the four HisZ chains in the hetero-octameric crystal structure. This observation may suggest that the HisZ proteins rotate relative to each other, like a pair of scissors, resulting in an extended conformation of the allosterically inhibited enzyme. Future investigation should focus on the interactions between the HisZ and HisG$_S$ proteins via chemical crosslinking and fluorescence resonance energy transfer.

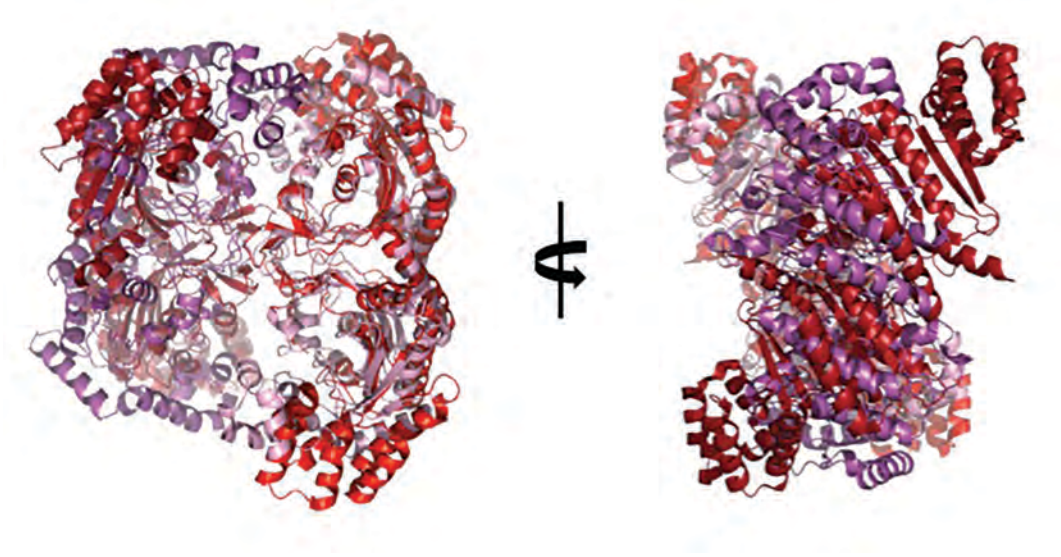

Figure 2.28 Alignment of four chains of histidyl-tRNA synthetase from *T. thermophilus* (Purple) with *Lla*HisZ (red). Note that the anticodon binding domains of histidyl-tRNA synthetase have been omitted for clarity.

Notably, HisZ is requiried for catalysis although the molecular basis for this requirement is unknown. The interaction between HisZ and $HisG_S$ may provide contacts that are favourable for catalysis. One possible mechanism of allosteric inhibition is that binding of histidine to HisZ may disrupt these interactions and inhibit the enzyme. It is unclear how histidine binding to HisZ inhibits catalysis, however the large, open nature of the active site suggests that it is unlikely to be the result of steric hinderance. ITC or saturation transfer difference NMR could also be utilised to investigate substrate binding in the presence of inhibitory concentrations of histidine. Molecular dynamics simulations could also be invaluable to determine dynamic contributions to allostery, however the absence of a histidine-bound crystal structure complicates the use of this technique.

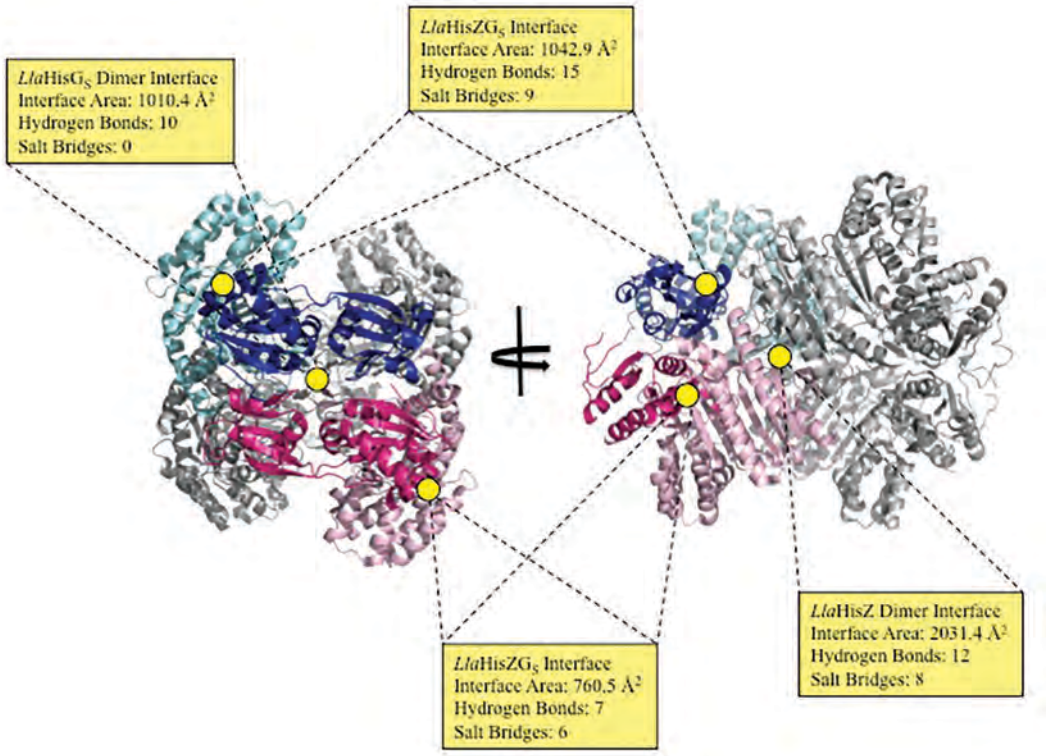

Figure 2.29 Analysis of the *Lla*HisZG$_S$ (PDB 1Z7N) interfaces using PDBePISA reveals asymmetry across the HisG$_S$ dimer interface, of the interfaces between HisG$_S$ domain I and HisZ. Two *Lla*HisZ-*Lla*HisG$_S$ interface interactions are coloured by chain. The larger of the two interfaces has *Lla*HisG$_S$ coloured dark blue and *Lla*HisZ coloured cyan. The second interface has *Lla*HisG$_S$ coloured dark pink and *Lla*HisZ coloured light pink. There is symmetry of the asymmetry, with equivalent interfaces found in the second HisG$_S$ dimer of the complex.

Further analysis of the *Lla*HisZG$_S$ crystal structure with PDBePISA revealed an asymmetry of the interactions between HisZ and domain I of HisG$_S$, across the HisG$_S$ dimer interface (Figure 2.29). Each pair of HisG$_S$ chains has one large HisZ-HisG$_S$ interface (1040 Å^2) and one small HisZ-HisG$_S$ interface (760 Å^2). The small interface provides the binding site for the interstitial phosphate ion that was proposed to be involved in activation of HisZG$_S$.[20] This may be indicative of alternating catalysis, where pairs of HisG$_S$ at opposing ends of the X-shaped HisZ core are active, while their dimer partners are not. Such a mechanism could be investigated by using fluorescence techniques in combination with crystallography and computational analysis.

In conclusion, investigation into the allosteric regulation of *Lla*HisZG$_S$ provided further evidence to support the histidine binding site proposed by Champagne *et al.*[20], comparable to the active site of histidyl-tRNA synthetase. SAXS revealed a conformational change to a more extended structure in response to allosteric inhibition, which may be the result of rotation of the HisZ subunits relative to each other. Further work is required to understand the nature of this conformational change and the role of HisZ in both the allosteric inhibition and activation of the short form ATP-PRTase.

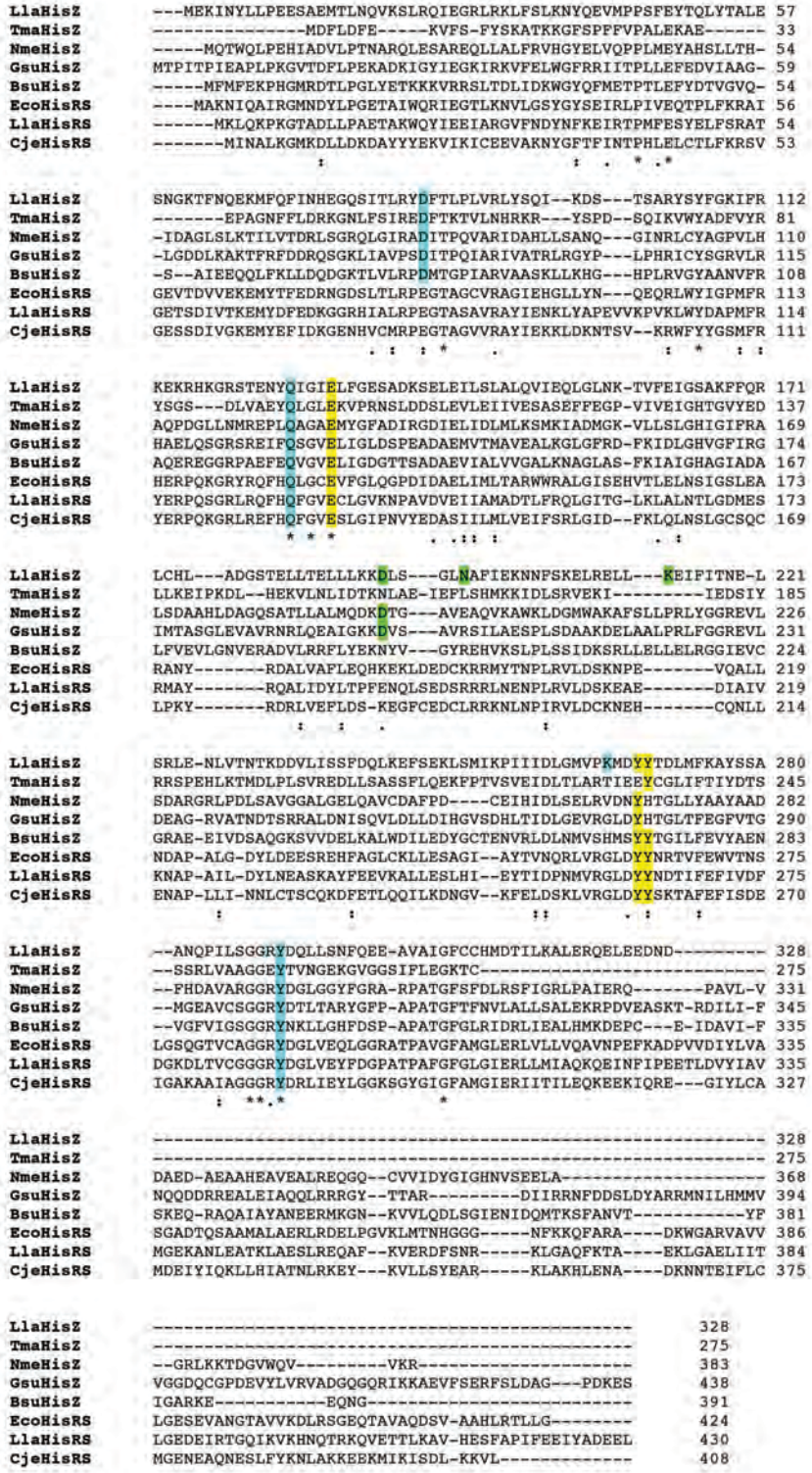

Figure 2.30 Sequence alignment of (in order from the top) HisZ from *L. lactis*, *T. maritima*, *N. meningitidis*, *G. sulfurreducens*, *B. subtilis*, and histidyl-tRNAsynthetase from *E. coli*, *L. lactis*, and *C. jejuni*. The histidine binding residues proposed by Champagne et al. are coloured yellow. The site one histidine binding residues proposed by induced fit docking experiments are coloured blue, and the histidine binding residues corresponding to sites two and three are coloured green.

Chapter Three

3 Cloning and Purification *Eco*HisG$_L$ and *Nme*HisG$_S$

3.1 Introduction

Structural and biochemical characterisation of the long form ATP-PRTase enzymes from *S. typhimurium, M. tuberculosis,* and *E. coli* has been progressing since the 1950s.[1,4,22] Despite decades of research, the mechanistic details of the allosteric regulation of HisG$_L$ are yet to be fully deduced. Prior to the elucidation of the X-ray crystal structures of *Eco*HisG$_L$ and *Mtu*HisG$_L$, investigation focused primarily on the quaternary structure of the enzyme in the presence of natural substrates and inhibitors.[3,4,47,51] A variety of oligomeric states have been described for the long form ATP-PRTases.[3,23,47] The most widely reported model of the allosteric inhibition is the histidine-induced oligomerisation from an active dimer to an inactive hexamer, initially described by Tebar *et al.*[47] Tebar *et al.* also noted the presence of a tetrameric species of unknown function.

The dimer-hexamer model was supported by Klungsoeyr and Kryvi, who reported a change in the sedimentation coefficient of *Eco*HisG$_L$ from a 12.6 S species in the absence of ligands to an 8.9 S species in the presence of inhibitors,[81] as well as the characterisation of *Mtu*HisG$_L$ by analytical SEC described by Cho *et al.*[3] The dimer-hexamer model of allosteric inhibition persisted despite the revelation of the hexameric crystal structures, as both Lohkamp *et al.* and Cho *et*

al. attributed the hexameric quaternary structure to crystal packing.[3,4] Recent investigation into *Mtu*HisG$_L$[33] and HisG$_L$ from *C. jejuni* (*Cje*HisG$_L$) (Gerd Mittelstädt, unpublished results) has challenged the dimer-hexamer model of allosteric inhibition.

Recent investigation has shown that removal of the ACT regulatory domain from *Cje*HisG$_L$ results in an active, albeit diminished, homodimeric enzyme (Gerd Mittelstädt, unpublished results), despite the observation that HisG$_S$ proteins only exhibit catalytic activity in the presence of HisZ.[5] This chapter describes the work towards the purification of *Eco*HisG$_L$ with the intent to clarify the earlier model of the allosteric regulation, as well as the generation of a truncation mutant devoid of the ACT domain to explore whether catalytic activity of such a mutant is a general feature of long form ATP-PRTases.

This chapter also reports extensive effort towards the purification of ATP-PRTase from *N. meningitidis* (*Nme*HisZG$_S$), the causative agent of bacterial meningitis. In 2012 there were 1.2 million reported cases of meningococcal disease, leading to 135,000 deaths worldwide.[82] The absence of the histidine biosynthetic pathway from mammalian metabolism highlights ATP-PRTase as an attractive drug target. Shramm *et al.* have presented PRTase enzymes as viable candidates for the synthesis of transition state analogues that serve as highly potent inhibitors targeting these enzymes.[12,42] Structural and biochemical characterisation of *Nme*HisZG$_S$ may inform the design of potent antibiotics to combat this disease. Sequence analysis has identified *N. meningitidis* ATP-PRTase as a short form ATP-PRTase. The sequence identities between *Nme*HisG$_S$ and the previously characterised *Tma*HisG$_S$ and *Lla*HisG$_S$ are 37% and 39% respectively.

*Nme*HisZ has 20% sequence identity with *Tma*HisZ and 23 % sequence identity with *Lla*HisZ. Biochemical characterisation of *Nme*HisZG$_S$ has not been previously reported.

3.2 Cloning, Expression and Preliminary Characterisation of *Eco*HisG$_L$

EcohisG$_L$ was amplified from *E. coli* genomic DNA (gDNA) using two rounds of nested PCR (Figure 3.1). The first round amplified the gene from gDNA and the second installed the TEV protease cleavage site and the *attB* sites for cloning using the Gateway® system (Life Technologies). The modified *EcohisG$_L$* gene was cloned into pDONR221 using the standard procedure for BP clonase™ (Figure 3.1). An *E. coli* TOP 10 competent cell line was used for the selection and proliferation of positive constructs. LR clonase™ was employed to subclone the sequence verified *EcohisG$_L$*pDONR221 construct into pDEST-17 with an N-terminal (His)$_6$-tag (Figure 3.2).

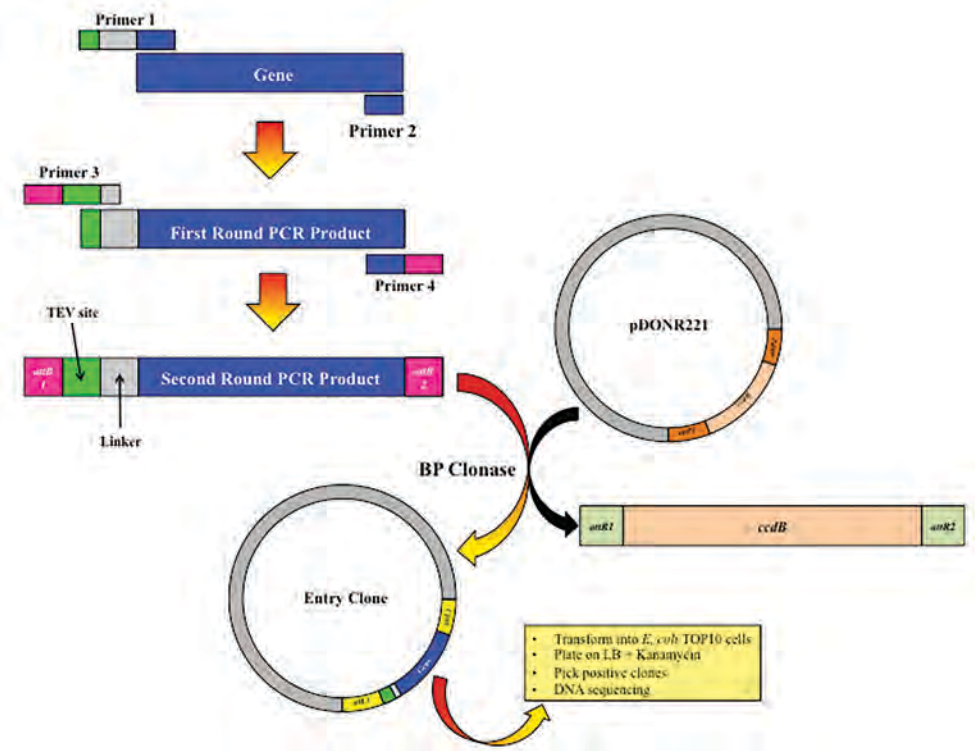

Figure 3.1 Schematic of the general protocol for amplification of a gene via nested PCR and cloning of a gene into pDONR221 using BP clonase™.

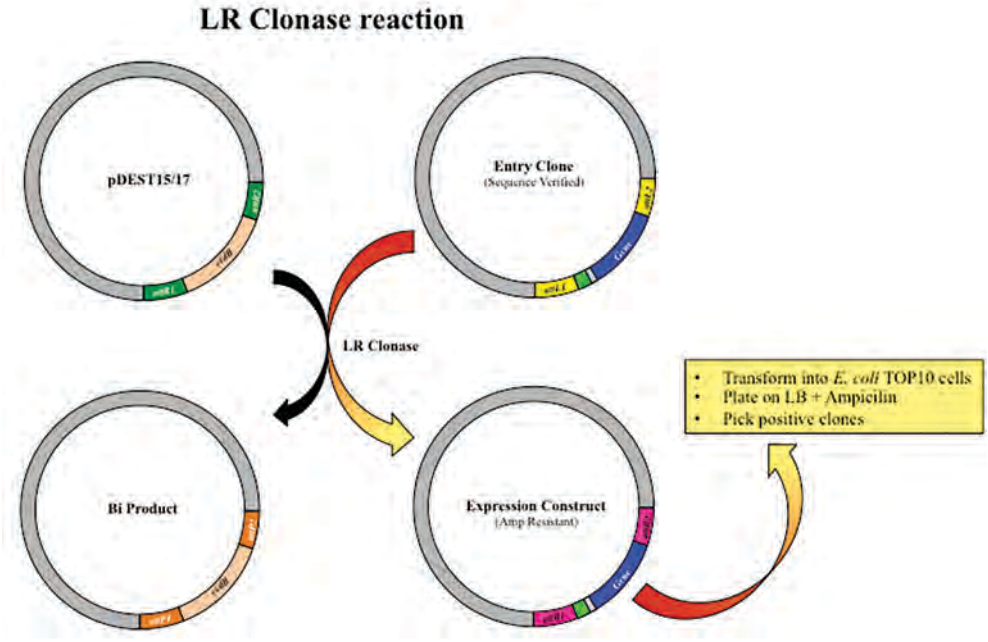

Figure 3.2 Schematic of the general protocol for cloning into a destination vector using LR clonase™.

Limited soluble protein was obtained from expression in BL21*(DE3) cells. Protein expression conditions were optimised by systematically varying the cell line, IPTG concentration, media, temperature and duration of incubation. As discussed in chapter two, the solubility of *Eco*HisG$_L$ was improved by expression in Chaperone 3 cells. The purification buffers were adapted from a previously published protocol.[83] Purification of *Eco*HisG$_L$ was achieved via two chromatography steps. The (His)$_6$-tagged protein eluted as a single peak from IMAC. The pooled fractions were desalted prior to the addition of TEV protease to cleave the N-terminal (His)$_6$-tag. Complete cleavage required incubation at 37 °C for one hour, followed by 4 °C for three days. Samples were taken throughout the duration of incubation to ensure the retention of catalytic activity. The untagged protein was then purified via SEC. *Eco*HisG$_L$ eluted from the preparative SEC as a single peak, with a small shoulder that likely corresponds to an aggregate species (Figure 3.3). Fractions corresponding to the main peak were pooled to yield 8 mg of *Eco*HisG$_L$ per litre of cell culture.

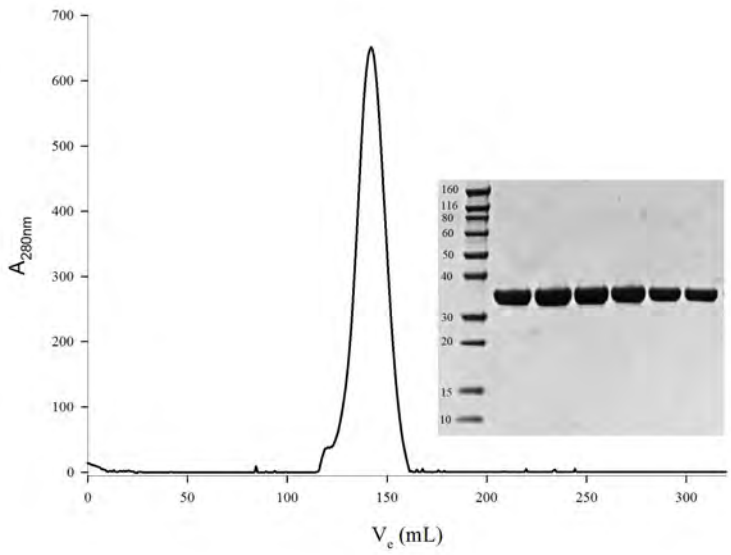

Figure 3.3 A_{280nm} trace of tagless *Eco*HisG$_L$ following purification via SEC. Insert: SDS-PAGE gel of purified *Eco*HisG$_L$ following SEC.

3.3 Determination of the Quaternary Structure of *Eco*hisG$_L$ by Analytical SEC

Analytical SEC was used to determine the quaternary structure of *Eco*HisG$_L$ (1 mg/mL) in solution. The protein eluted as a single peak at the void volume of the column, consistent with a higher order species (molecular weight > 669 kDa) (Figure 3.4). There was no evidence of a hexameric (267 kDa), tetrameric (133 kDa), or dimeric (66.7 kDa) species as previously reported by Tebar *et al.*[47]

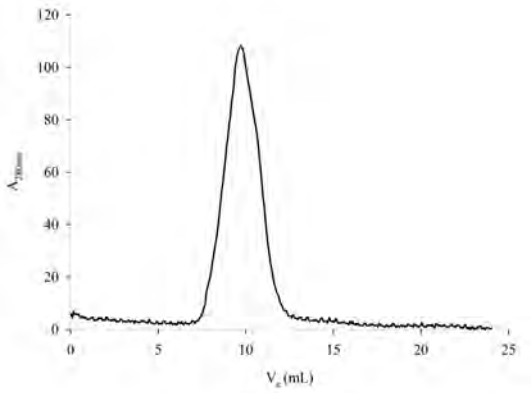

Figure 3.4 A_{280nm} trace of *Eco*HisG$_L$ following analytical SEC. A single species was observed with an elution volume consistent with the void volume of the column. The volume of elution of blue dextran (2000 kDa) was 8.05 mL and the elution volume of *Eco*HisG$_L$ was 8.15 mL. The expected molecular weight of the hexameric *Eco*HisG$_L$ is 267 kDa.

3.4 Determination of the Quaternary Structure of *Eco*HisG$_L$ by SAXS

In order to confirm the results of analytical SEC, SAXS experiments were performed on *Eco*HisG$_L$ in the absence of ligands at concentrations of 0.3 mg/mL, 0.75 mg/mL, 1.5 mg/mL and 3 mg/mL. SAXS experiments were also conducted with the addition of 1 mM ATP, 100 μM PRPP, 1 mM histidine, and the combination of 1 mM histidine and 1 mM AMP (Figure 3.5). Each scattering profile was consistent with a higher order species. The theoretical R_g values

calculated from the hexameric crystal structures were 23.2 Å (PDB 1H3D) and 22.9 Å (PDB 1Q1K). The R_g values determined from the experimental SAXS data by the Guinier approximation and P(r) function ranged from 60.4 ± 0.1 Å to 85.2 ± 0.8 Å (Table 3.1). Crysol was used to fit the SAXS profiles to each crystal structure, and the fits were sufficiently poor that χ^2 values could not be determined (Figure 3.5).

Table 3.1: The R_g, D_{max} and Porod Volume Determined for *Eco*HisG$_L$ SAXS Data in the Presence and Absence of Natural Substrates and Inhibitors

Sample	R_g (Guinier) (Å)	R_g (P(r)) (Å)	Porod Volume (Å^3)	D_{max} (Å)
*Eco*HisG$_L$	60.4 ± 0.1	62.0 ± 0.1	1045850	242
*Eco*HisG$_L$ + His + AMP	77.6 ± 0.2	73 ± 1	1292950	241
*Eco*HisG$_L$ + ATP	77.2 ± 0.1	72.3 ± 0.9	1553690	261
*Eco*HisG$_L$ + His	77.3 ± 0.3	69 ± 1	1360820	244
*Eco*HisG$_L$ + PRPP	77.1 ± 0.2	85.2 ± 0.8	1661890	356

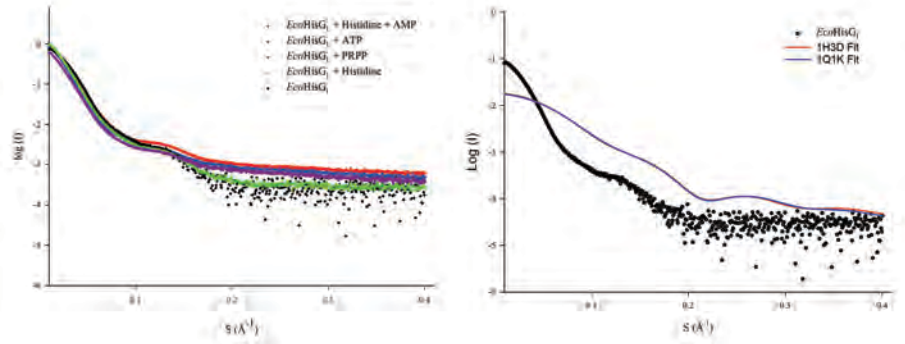

Figure 3.5 Overlay of *Eco*HisG$_L$ SAXS scattering profiles (left). Crysol was used to fit the experimental scattering to the theoretical scattering profile from the two available crystal structures (PDB 1Q1K and 1H3D). The crysol plot for unliganded *Eco*HisG$_L$ is shown on the right, the χ^2 values could not be determined.

Investigation into the quaternary structure of HisG$_L$ from *S. typhimurium*, *E. coli*, and *M. tuberculosis* has revealed the presence of dimeric, hexameric, and potentially tetrameric species, which has led to debate over the functional oligomeric state(s).[3,23,47] In addition, higher order oligomers of these enzymes have been reported, however

these are unlikely to be biologically significant.[46,50] The higher order species of *Eco*HisG$_L$ detected by both analytical SEC and SAXS is surprising, as the elution volume of *Eco*HisG$_L$ from the preparative SEC is consistent with the elution volume of the hexameric *C. jejuni* HisG$_L$ from the same column (Gerd Mittelstädt personal communication). This suggests that oligomerisation may be the result of downstream handling of the protein. Typically *Eco*HisG$_L$ was concentrated to 2-10 mg/mL and flash frozen following purification. Analysis of *Eco*HisG$_L$ by analytical SEC prior to concentration and storage was also consistent with a higher order species. Alteration of buffer composition by the variation of the pH, concentration of NaCl (0-500 mM) and the addition of TCEP and glycerol (5 % v/v) had no effect on the oligomeric state of the protein. In addition, chromatography was performed at both room temperature and at 4 °C to determine whether oligomerisation was temperature dependent.

Preliminary kinetic analysis, using the standard enzyme kinetics assay, revealed that the purified *Eco*HisG$_L$ is catalytically active following storage at -80 °C, however the estimated specific activity (0.2 μmol/min/mg) at 25 °C with 50 μM PRPP and 10 mM ATP, was approximately ten fold lower than the previously reported activity.[50] The estimated k_{cat} (approximately 0.1 s^{-1}) was 10-20 fold lower than the k_{cat} reported for the long form ATP-PRTase from *S. typhimurium*.[25] The diminished activity suggests that a portion of the *Eco*HisG$_L$ is not catalytically active. Time constraints precluded optimisation of the purification protocol and further characterisation of this protein.

3.5 Cloning and Expression of an *Eco*HisG$_L$ Truncation Mutant

In order to explore the function of the regulatory ACT domain in *Eco*HisG$_L$ a truncation mutant, devoid of the ACT domain, was generated (*Eco*HisG$_{core}$), in parallel with the studies exploring the wild-type protein. The gene sequence corresponding to domains I and II of *Eco*HisG$_L$ up to E225 was amplified using nested PCR, as previously described for the wild type *Eco*HisG$_L$, and cloned into pDEST-17 with an N-terminal (His)$_6$-tag using the Gateway system (Figure 3.6). Expression trials revealed soluble protein in Chaperone 3 cells, however purification and characterisation was not pursued due to the absence of wild-type *Eco*HisG$_L$ as a suitable comparison (Figure 3.7). The *EcohisG$_{core}$*pDEST-17 construct was shown to express catalytically active protein *in vivo* by *E. coli* EW25113Δ*hisG* complementation studies reported in chapter four.

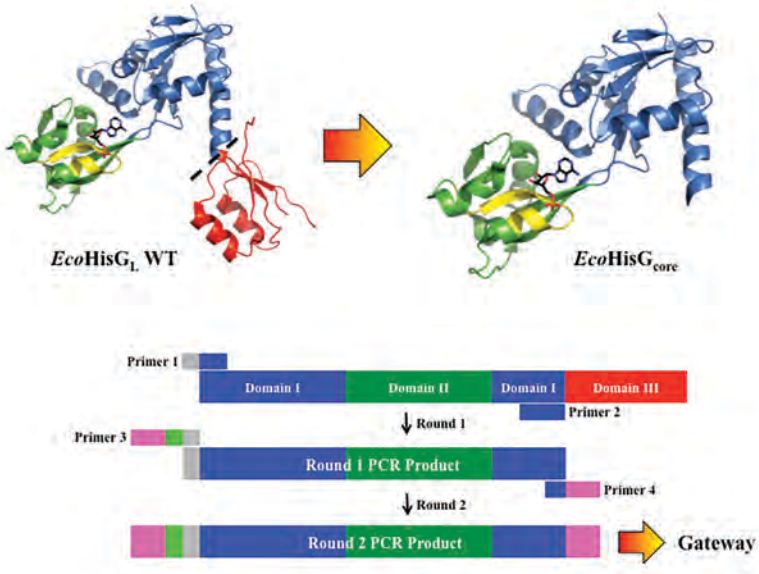

Figure 3.6 (Top) A single chain of *Eco*HisG$_L$ (PDB 1H3D) coloured by domain. The PRPP binding motif is coloured yellow and AMP is shown bound in the active site. The truncation site to form *Eco*HisG$_{core}$ is represented by a dashed line. (Bottom) a schematic diagram of the amplification of *EcohisG$_{core}$* via nested PCR. Primers were designed to amplify domains I and II of *EcohisG$_L$* and install a linker (grey), TEV protease cleavage site (bright green) and the *attB* sites for gateway cloning (pink).

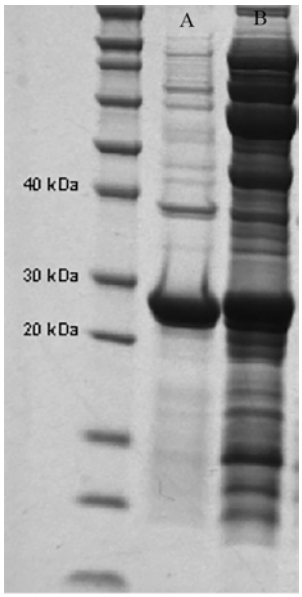

Figure 3.7 SDS-PAGE gel following expression trial of *Eco*HisG$_{core}$ lane A consists of the insoluble fraction and lane B is the soluble fraction.

3.6 Cloning, Expression and Purification of ATP-PRTase from *N. meningitidis* (*Nme*HisZG$_S$)

Numerous attempts were made to clone, express, and purify *Nme*HisZG$_S$, with limited success. The gene encoding *Nme*HisG$_S$ was cloned from *N. meningitidis* gDNA and the *NmehisZ* gene was purchased from Epoch Life Science. Initially the decision was made to co-express *Nme*HisZ and *Nme*HisG$_S$ to allow formation of the *Nme*HisZG$_S$ complex *in vivo*. The observation that HisG$_S$ co-purifies as a hetero-octameric complex[5] was exploited in the development of this purification scheme, as discussed in chapter two. *NmehisZ* was cloned into multiple cloning site I of the pRSFDuet-1 vector with an N-terminal (His)$_6$-tag and *NmehisGs* was cloned into multiple cloning site II of the *NmehisZ*pRSFDuet-1 vector without a tag. Standard methods of restriction enzyme digest followed by ligation using either T4 DNA ligase or the Quick Ligation Kit (New England Biolabs) were unsuccessful. Various methods of DNA purification were tested

including E-Gel® CloneWell™ (Life Technologies), the High Pure® PCR Product Purification Kit (Roche) and the Nucleospin® PCR Product Cleanup Kit (Clontech) as buffer components, such as ethanol, from DNA purification kits are known to inhibit ligation reactions.[84] Ligation was unsuccessful under all conditions tested. The In-Fusion® HD Cloning Kit (Clontech) was subsequently employed and both genes were cloned sequentially into the pRSFDuet-1 vector using the standard procedure described in chapter two.

The expression level of *Nme*HisZ from the *NmehisZGs*pRSFDuet-1 construct in both BL21*(DE3) and Chaperone 3 cell lines greatly exceeded the expression level of *Nme*HisG$_S$ (Figure 3.8 A). Purification of the (His)$_6$-tagged *Nme*HisZG$_S$ complex via IMAC followed by SEC, yielded less than 1 mg of per litre of cell culture. Optimisation of expression conditions by varying the cell line, IPTG concentration, media, temperature, and duration of incubation had no effect on the relative expression levels. The protein eluted from the preparative SEC as two peaks (Figure 3.9). The *Nme*HisZG$_S$ complex eluted as a small peak at the void volume of the column, which suggests that the complex exists as a soluble aggregate species. The fractions from the second peak, which corresponded to excess *Nme*HisZ, were pooled to yield 15 mg of *Nme*HisZ per litre of cell culture.

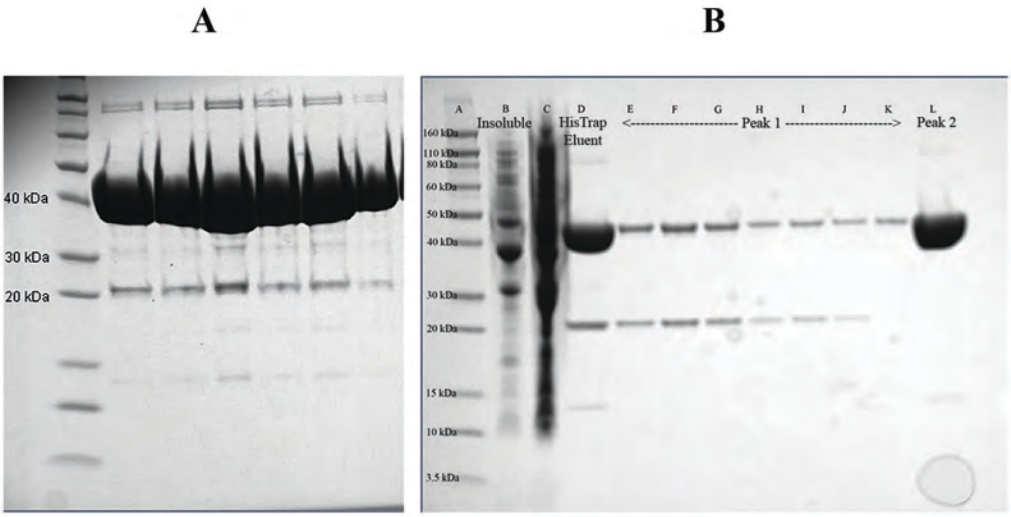

Figure 3.8A) SDS-PAGE gel following the IMAC purification of *Nme*HisZG$_S$ via expression of the *NmehisZG$_S$*pRSFDuet-1 construct revealed a significantly greater expression level of *Nme*HisZ (band at approximately 40 kDa) relative to *Nme*HisG$_S$ (band at approximately 20 kDa). B) SDS-PAGE gel following the purification of *Nme*HisZG$_S$ by co-lysing the pellets of *NmehisZ*pRSFDuet-1 BL21*(DE3) and *NmehisG$_S$*pDEST-14 Chaperone 3. Lanes B and C correspond to the insoluble and soluble fractions respectively, lane D is the protein eluted from IMAC, lanes E-K correspond to peak 1 from the preparative SEC column and lane L represents purified *Nme*HisZ.

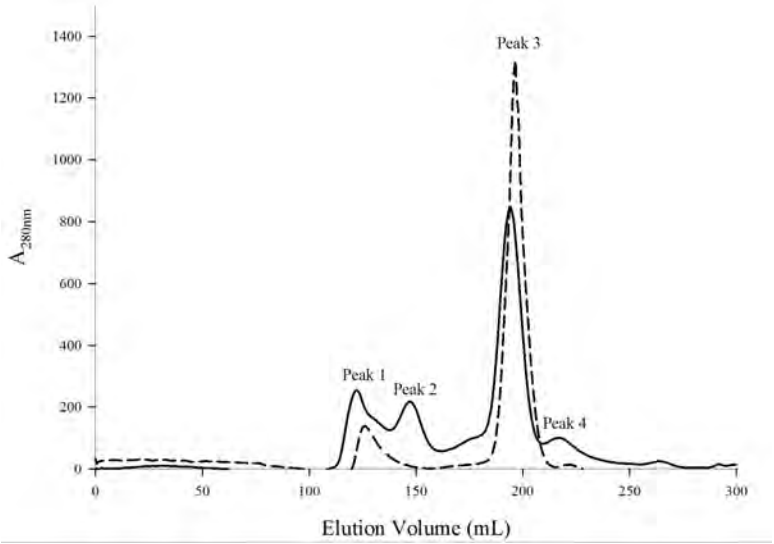

Figure 3.9 Two SEC traces following *Nme*HisZG$_S$ purification. The dotted line represents the trace following early attempts to purify the enzyme and the solid line represents the trace where a small amount of hetero-octameric protein was obtained. Peak 1 corresponds to higher order misfolded protein, peak 2 is consistent with an *Nme*HisZG$_S$ hetero-octamer, peak 3 represents uncomplexed NmeHisZ and peak 4 is consistent with *Nme*HisG$_S$.

Alternative approaches to purify the *Nme*HisZG$_S$ complex were explored. Attempts were made to purify *Nme*HisZ and *Nme*HisG$_S$

independently and combine the two proteins *in vitro* to form the complex. This approach potentially had the additional benefit of allowing *Nme*HisZ and *Nme*HisG$_S$ to be characterised independently. The Gateway® system (Life Technologies) was utilised to clone *NmehisG$_S$* into pDEST-17 with an N-terminal (His)$_6$-tag, pDEST-15 with an N-terminal Glutathione S-transferase (GST)-tag, and pDEST-14 without a tag. Multiple expression trials were performed, varying the cell line, IPTG concentration, media, temperature, duration of incubation, and the composition of the lysis buffer. Initially the (His)$_6$-tagged and untagged proteins were insoluble under all conditions tested, however soluble (GST)-tagged protein was obtained from Chaperone 3 cells. Purification of the GST-tagged protein was attempted, however TEV cleavage of the GST-tag following the first chromatography step resulted in precipitation of the protein under all conditions tested. Reconstitution of the *Nme*HisZG$_S$ complex was attempted by the addition of purified *Nme*HisZ, in a stoichiometric ratio with the GST-tagged *Nme*HisG$_S$, to the TEV cleavage mixture. Samples were tested for catalytic activity using the standard enzyme kinetics assay, however ATP-PRTase activity was not detected after 12 hours incubation at 4 °C and precipitation of the protein was observed. It is imperative that the GST tag should be cleaved given its size (25.7 kDa) relative to the size of *Nme*HisG$_S$ (23.3 kDa) and the propensity of the GST-tag to dimerise, which is unlikely to be conducive to formation of the hetero-octameric complex.

Optimisation of lysis buffer composition eventually resulted in soluble (His)$_6$-tagged *Nme*HisG$_S$ and untagged *Nme*HisG$_S$, however both the tagged and untagged protein precipitated during the first chromatography step. This suggests that *Nme*HisG$_S$ is relatively

unstable in the absence of *Nme*HisZ. This observation is consistent with the attempted purification of *Lla*HisG$_S$ discussed in chapter two.

The decision was made to utilise the pellet mixing technique described by Vega *et al.*[19], as discussed in chapter two. Cell pellets of *NmehisZ*pRSFDuet-1 BL21*(DE3) and *NmehisGs*pDEST-14 Chaperone 3 cells were combined and co-lysed via sonication. The expression level of *Nme*HisG$_S$ improved in comparison to the previously described expression level from the *NmehisZGs*pRSFDuet-1 construct. The *Nme*HisZG$_S$ complex, however, eluted at the void volume of the preparative SEC (Figure 3.8). As previously reported for *Lla*HisZG$_S$, ATP-PRTase activity was detected in the cell lysate of the combined *Nme*HisZ and *Nme*HisG$_S$ but was absent from the cell lysate of *Nme*HisG$_S$ expressed independently. The information gleaned from the purification of *Lla*HisZG$_S$ eventually led to the decision to subclone *NmehisGs* into pET-21a. The expression level improved significantly in comparison to the pDEST-14 construct. Purification of the *Nme*HisZG$_S$ complex was achieved by co-lysing the cell pellets from *NmehisZ*pRSFDuet-1 BL21*(DE3) and *NmehisGs*pET-21a chaperone 3, followed by IMAC and SEC. A new peak was observed on the preparative SEC A_{280nm} trace (Figure 3.8). SDS-PAGE analysis revealed an approximately equal ratio of *Nme*HisZ and *Nme*HisG$_S$ (Figure 3.10). The fractions corresponding to this peak were pooled to yield 5 mg of *Nme*HisZG$_S$ per litre of cell culture. The detection of a fourth peak on the A_{280nm} trace, corresponding to *Nme*HisG$_S$ in the absence of HisZ, was also noted and the corresponding fractions were pooled to yield 1 mg of *Nme*HisG$_S$ per litre of cell culture.

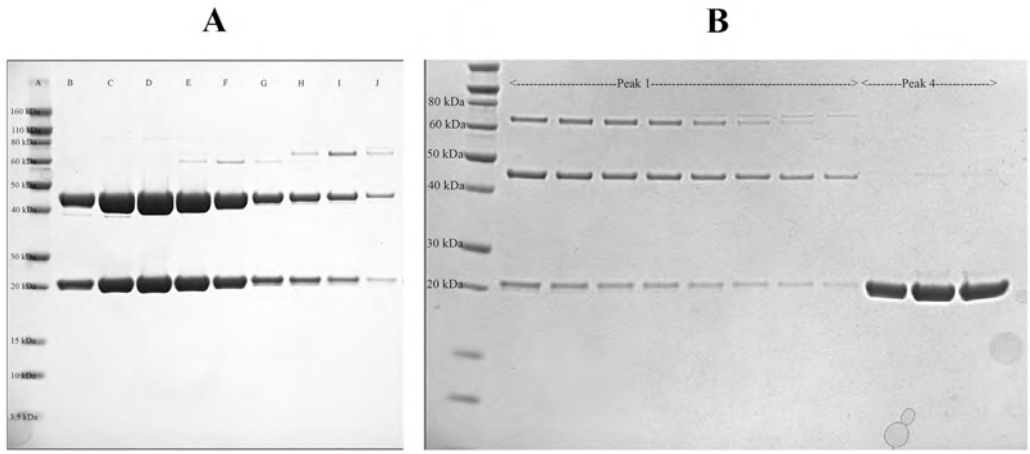

Figure 3.10 A) SDS-PAGE gel of peak 2 from the preparative SEC showing the *Nme*HisZG$_S$ enzyme complex. B) SDS-PAGE gel showing the fractions corresponding to peak 1 and peak 4 from the preparative SEC. Peak 1 is likely to be an aggregated, misfolded species, indicated by the early elution volume and the presence of chaperonin protein (band at approximately 60 kDa). Peak 4 corresponds to purified *Nme*HisG$_S$.

3.7 Determination of the Quaternary Structure of *Nme*HisZG$_S$ by Analytical SEC

3.7.1 Characterisation of *Nme*HisZG$_S$ by Analytical SEC

Analytical SEC was employed to determine the quaternary structure of *Nme*HisZG$_S$ in solution. The purified protein from the co-lysis of *NmehisZ*pRSFDuet-1 BL21*(DE3) and *NmehisGS*pDEST-14 Chaperone 3 cell pellets eluted as two peaks from the analytical SEC. The first peak, eluted at the void volume of the column, consistent with an aggregate species, whilst the second peak had an elution volume, consistent with an *Nme*HisZ dimer (Figure 3.11). The protein purified via the combination of *NmehisZ*pRSFDuet-1 BL21*(DE3) and *NmehisGS*pET-21a Chaperone 3 cell pellets also eluted as two peaks (Figure 3.11). The first peak, again corresponded to an aggregate species, however the second peak had an elution volume consistent with the *Nme*HisZG$_S$ hetero-octamer. This suggests that the protein

has a propensity to aggregate and further optimisation of the purification and experimental conditions is required.

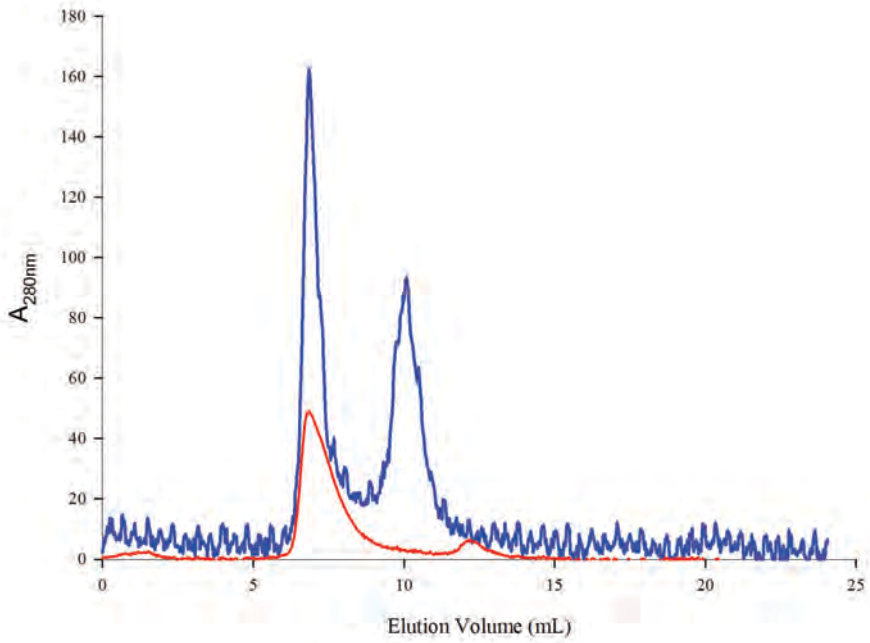

Figure 3.11 Analytical SEC of *Nme*HisZG$_S$ purified from the combination of *Nmehis*ZpRSFDuet-1 BL21*(DE3) and *Nmehis*GspDEST-14 Chaperone 3 cell pellets (red line) shows two peaks. The first peak is consistent with an aggregate species and the second peak is consistent with an *Nme*HisZ dimer. The analytical SEC trace of *Nme*HisZG$_S$ purified via the combination of *Nmehis*ZpRSFDuet-1 BL21*(DE3) and *Nmehis*GspET-21a Chaperone 3 (blue line) also shows two peaks. The first is an aggregate species and the second is consistent with the *Nme*HisZG$_S$ hetero-octamer.

In order to determine whether the *in vitro* combination of the purified *Nme*HisG$_S$ and *Nme*HisZ to form the hetero-octameric *Nme*HisZG$_S$ complex was feasible, 0.04 mM *Nme*HisZ was combined with 0.04 mM *Nme*HisG$_S$ and loaded onto the analytical SEC. This protein eluted as two peaks, the first at the void volume of the column, consistent with an aggregate species, and the second was consistent with an *Nme*HisZ dimer (Figure 3.12). This protein preparation was also catalytically inactive, which suggests that *in vitro* complex formation is not feasible under the conditions tested.

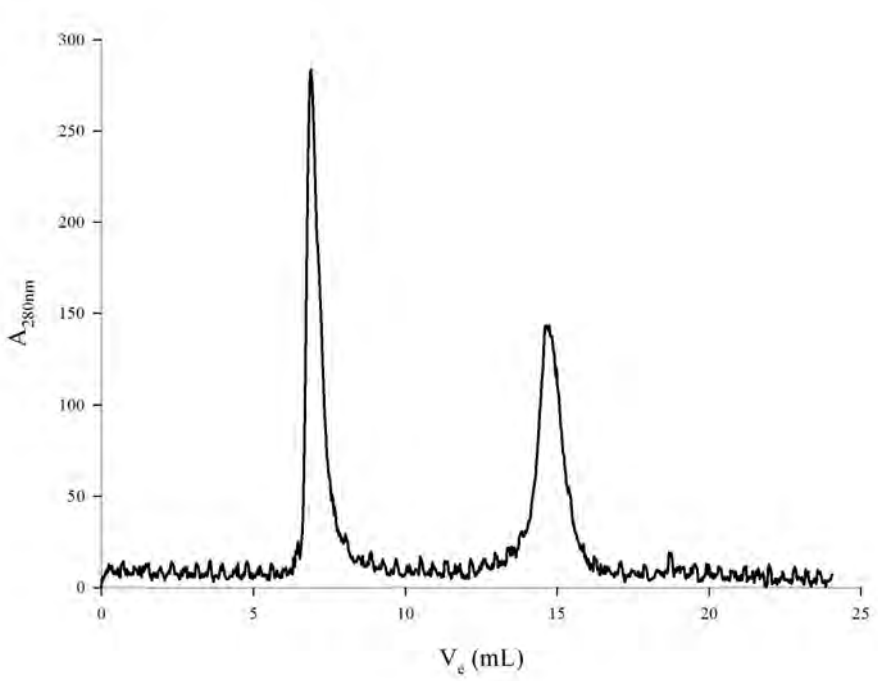

Figure 3.12 Analytical SEC following the *in vitro* combination of purified *Nme*HisZ and *Nme*HisGs revealed two peaks. The first peak is consistent with a higher order aggregate species and the second is consistent with an *Nme*HisZ dimeric species.

3.7.2 Characterisation of *Nme*HisZ by Analytical SEC

*Nme*HisZ eluted from the analytical SEC as a single peak, consistent with an *Nme*HisZ dimer, at concentrations ranging from 0.2 mg/mL to 9 mg/mL (Figure 3.13). This is in line with the *Lla*HisZ dimeric species discussed in chapter two and contradicts the observation of a *Tma*HisZ dimer-tetramer equilibrium reported by Vega *et al.*[19]

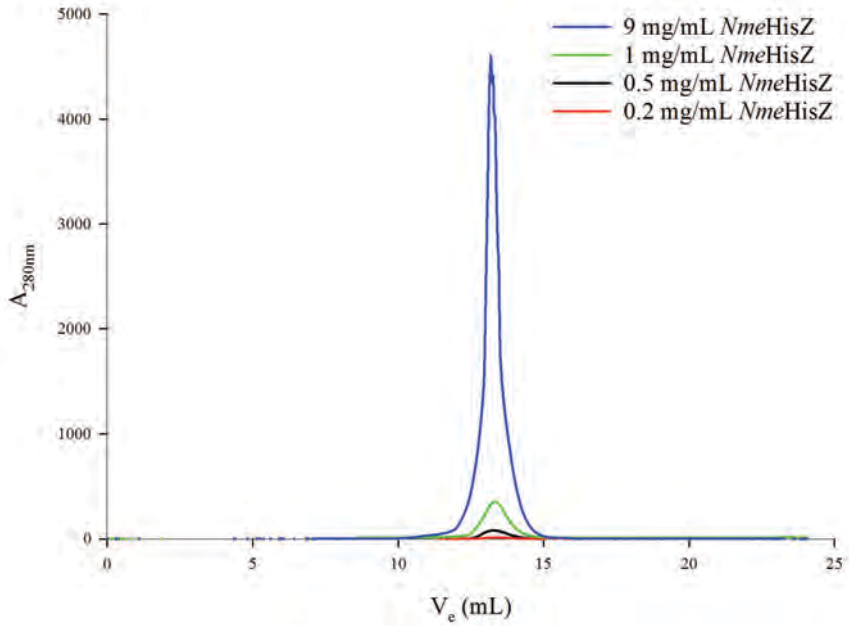

Figure 3.13 Analytical SEC trace of *Nme*HisZ at concentrations ranging from 0.2 mg/mL to 9 mg/mL revealed a single species with a molecular weight consistent with an *Nme*HisZ dimer.

3.7.3 Analytical SEC Studies on *Nme*HisGS

*Nme*HisGS also eluted as a single peak from analytical SEC, corresponding to a dimeric species (Figure 3.14). This is consistent with the dimeric HisGS crystal structures from *T. thermophilus* (PDB 1VE4) and *B. subtilis* (PDB 2VD2).

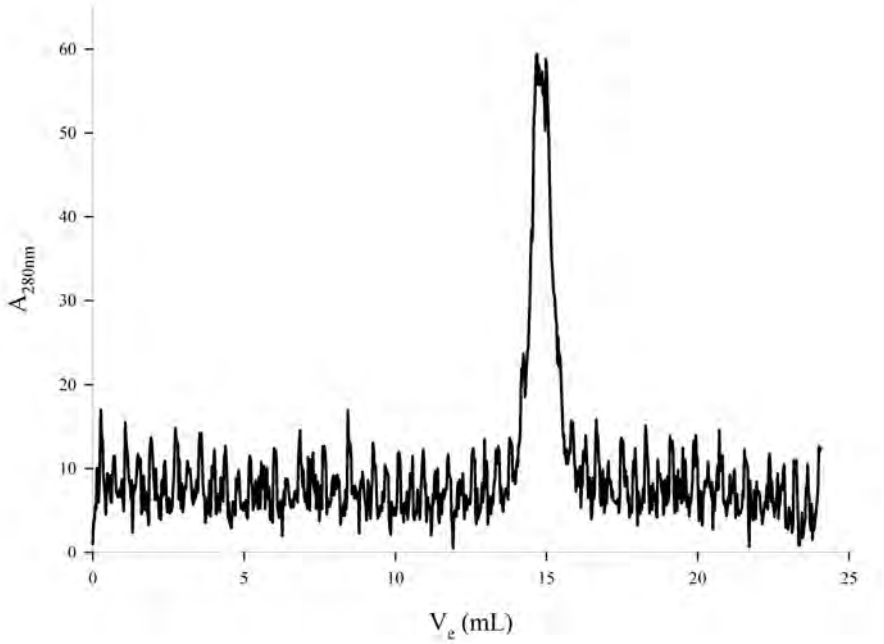

Figure 3.14 Analytical SEC trace of *Nme*HisGS (0.5 mg/mL) showing a single species with a molecular weight corresponding to a dimer.

3.8 Crystallisation Trials of *Nme*HisZ

Multiple commercial screens were employed to determine crystal conditions for the *Nme*HisZ in the absence of *Nme*HisGS. Several crystal conditions were found, however no diffraction quality crystals were obtained and further optimisation is required (Figure 3.15).

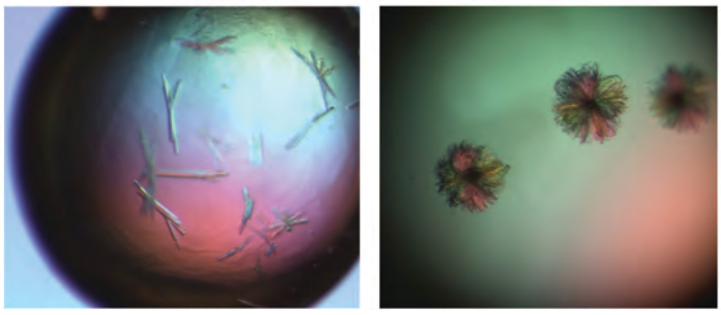

Figure 3.15 Crystals of *Nme*HisZ obtained from the manual optimisation of conditions from the *ProPlex* commercial crystal screen.

3.9 Preliminary Kinetic Analysis

Preliminary kinetic characterisation of the purified *Nme*HisZG$_S$ revealed a K_M^{PRPP} value of 18 ± 1 μM, which is consistent K_M^{PRPP} value of 18 ± 4 μM reported for *Lla*HisZG$_S$ (Figure 3.16).[26] The specific activity, however, is lower than that of *Lla*HisZG$_S$. The k_{cat} determined for *Nme*HisZG$_S$ (0.91 ± 0.02 s^{-1}) was approximately 3 fold less than the k_{cat} of *Lla*HisZG$_S$ (2.7 ± 0.3 s^{-1}). This may reflect a lower catalytic activity of *Nme*HisZG$_S$ compared to *Lla*HisZG$_S$, however this observation, in combination with the large aggregate peak observed by analytical SEC, suggests that the aggregate species may not be catalytically active and further optimisation of the purification protocol is required, which precluded further kinetic and biochemical characterisation of *Nme*HisZG$_S$.

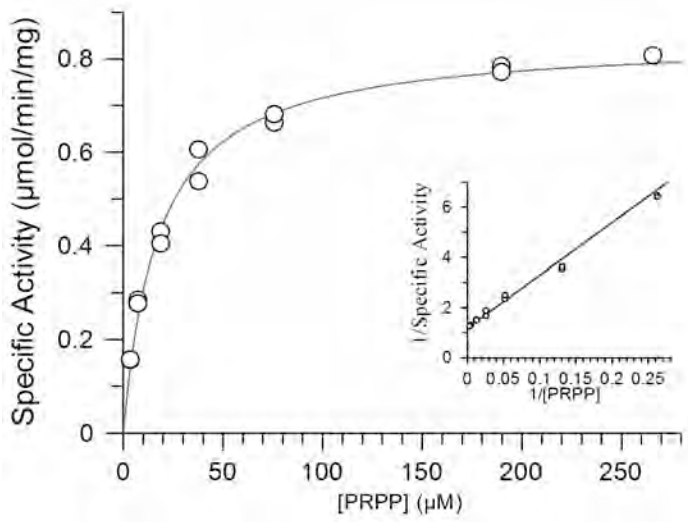

Figure 3.16 The K_M^{PRPP} for *Nme*HisZG$_S$ determined by holding the [ATP] constant at 10 mM and varying the [PRPP]. The lineweaver-burk plot is shown as an insert.

3.10 Discussion

Attempts to clone, express and purify the long form ATP-PRTase from *E. coli* and the short form ATP-PRTase from *N. meningitidis* proved largely unsuccessful due to the tendency of both enzymes to aggregate during or following purification. Aggregate species of *Eco*HisG$_L$ have previously been described, however the complete absence of the biologically relevant oligomeric state(s) of *Eco*HisG$_L$ is surprising.[47,50] Kryvi and Klungsoyer have reported loss of activity of *Eco*HisG$_L$ at high concentration, which they attributed either to hexamer formation or to the formation of non-specific aggregates, however Dall-Larsen suggested that this inhibition was due to an early build up of PR-ATP, thereby inhibiting the enzyme.[50,85] Analytical SEC and SAXS experiments were consistent with higher order *Eco*HisG$_L$ species in the presence and absence of inhibitors. SAXS revealed no clear trend in the R_g or $D_{max,}$ which suggests that the higher order species is not uniform, a further indication that it is not biologically functional. Variation of the concentration of NaCl and the addition of a reducing agent and glycerol to the purification buffers had no impact on aggregation. The buffers used in this purification were adapted from the method described by Lohkamp *et al.,* however the published method utilised anion exchange chromatography to purify the untagged protein, as opposed to the IMAC purification of the (His)$_6$-tagged protein described in this chapter.[83] Further investigation should focus on purification of the untagged protein by anion exchange, as it is possible that the presence of the (His)$_6$-tag, or the three day incubation required to cleave the tag, may be responsible for aggregation of the protein. It should be noted that Lohkamp *et al.* reported significant loss of *Eco*HisG$_L$ during purification due to precipitation.[83] Further optimisation of the purification scheme could

include additional buffer components such as detergents or bovine serum albumin, which may help to mitigate aggregation.

Numerous techniques were employed to clone, express, and purify *Nme*HisZG$_S$, with minimal success. *Nme*HisZ expressed and purified readily whilst the expression level of *Nme*HisG$_S$ was significantly lower. *Nme*HisG$_S$ has a tendency to precipitate in the absence of *Nme*HisZ. This suggests that *Nme*HisZ plays a role in stabilising *Nme*HisG$_S$ in the formation of the active hetero-octamer. *Nme*HisZ appears to activate *Nme*HisG$_S$, as previously reported for *Tma*HisZG$_S$ and *Lla*HisZG$_S$. It is interesting, therefore, that truncation mutants of HisG$_L$ from both *E. coli* and *C. jejuni,* lacking the ACT domain, retain diminished catalytic activity. Further investigation into the structure, dynamics, and biochemistry of both the wild type and mutant proteins is required to understand this retention of catalytic activity.

*Nme*HisZG$_S$ primarily adopts a higher order species during purification, a phenomenon that has not been previously reported for the short form ATP-PRTase enzymes.[19,20,26] Further optimisation of the purification procedure is required to minimise aggregation of the protein. This could include the purification of the untagged complex, or the addition of bovine serum albumin, or detergents to the purification buffers. The *in vitro* combination of purified *Nme*HisZ and *Nme*HisG$_S$ is unlikely to be a viable purification method as it was shown to result in inactive protein, whilst the presence of cellular contents during lysis of the combined cell pellets appears to promote complex formation. The small amount of hetero-octameric *Nme*HisZG$_S$ obtained was likely due to the improved expression level of *Nme*HisG$_S$ following subcloning into pET-21a. This increase in expression level was unexpected given

that pDEST vectors are derived from pET vectors, however it may simply be due to the difference between distance of the ribosome binding site from the start codon in the two constructs. Although a small amount of hetero-octameric *Nme*HisZG$_S$ was eventually obtained, the observation of the higher order species by analytical SEC and the low k_{cat} value relative to the k_{cat} reported for *Lla*HisZG$_S$, suggests that the higher order species may not be catalytically active. Further optimisation of the purification is required before proceeding with characterisation of the enzyme.

Contrary to the reports of a *Tma*HisZ dimer-tetramer equilibrium described by Vega *et al.*,[19] characterisation of the quaternary structure of *Nme*HisZ by analytical SEC revealed a dimeric structure of the protein at concentrations ranging from 0.2 mg/mL to 9 mg/mL. This suggests that the X-shaped HisZ core, assuming that the structure of *Nme*HisZG$_S$ is similar to the structures reported for *Tma*HisZG$_S$ and *Lla*HisZG$_S$, is stabilised only in the presence of *Nme*HisG$_S$. *Nme*HisG$_S$ also appears to adopt a dimeric structure in the absence of *Nme*HisZ. This is consistent with the dimeric crystal structures of the HisG$_S$ proteins from *B. subtilis* and *T. thermophilus*, however further investigation is required as the low yield following the serendipitous *Nme*HisG$_S$ purification precluded in depth biochemical characterisation of the protein.

In conclusion, the tendency of both *Eco*HisG$_L$ and *Nme*HisZG$_S$ to aggregate during or following purification precluded extensive biochemical characterisation of these proteins. Future investigation should focus on optimisation of the purification protocols. *Nme*HisZG$_S$ is of particular importance given its identification as a potential target for the development of potent antibiotics.

Chapter Four

4 Cloning and Purification of *ATP-PRTase from Leuconostoc mesenteroides* and the Construction of Three Chimeric Proteins

4.1 Introduction

Allosteric regulation is a critical feature of metabolism. Binding of a ligand at an allosteric site affects activity at a remote functional site, exploiting protein flexibility and altering the shape of the protein energy landscape.[79] The importance of allostery is underpinned by its ubiquity and diversity. A myriad of allosteric mechanisms have been unravelled, which likely evolved through several independent mechanisms.[86] Three strategies have been proposed for the evolutionary acquisition of functional allostery.[86] The first involves the structural modification of an existing domain to form an allosteric effector site. The second involves the development of homomeric or heteromeric assemblies, with the formation of allosteric binding sites at the subunit interfaces. Thirdly, allostery may be bestowed by the recruitment of a distinct regulatory domain, via gene fusion, to form a multidomain protein.

The ATP-PRTase family likely acquired allosteric regulation by a combination of all three evolutionary mechanisms. There is

unambiguous evidence to support the common ancestry of $HisG_S$ and $HisG_L$, and the structural and sequential similarity between HisZ and the histidyl-tRNA synthetases shows a clear evolutionary relationship between histidine biosynthesis and the use of histidine in protein synthesis by its cognate tRNA.[5,19] Assuming that aminoacyl-tRNA synthetases were among the earliest enzymes to emerge, it is plausible that the short form ATP-PRTase evolved subsequently. In this scenario, HisZ and histidyl-tRNA synthetase share a common evolutionary ancestor and the contemporary $HisG_S$ emerged independently from an unidentified precursor. The subsequent recruitment of an ACT domain may have given rise to the $HisG_L$ family, although how this domain fusion event could confer both catalytic activity and allostery to the protein remains to be explored.

The ACT domain has been identified as a mobile unit of allostery.[31,32] Most enzymes in which the ACT domain has been identified catalyse early reactions in amino acid metabolism, and ligand binding is typically associated with the interface of two or more ACT domains.[32] Recent work has shown that the simple transfer of an ACT domain from one protein is sufficient to confer functional allostery to its unregulated orthologue from a distinct species. This observation suggests that the acquisition of functional allostery may be remarkably facile. This chapter reports the formation of a chimeric protein in which the ACT domain from $HisG_L$ of *C. jejuni* (*Cje*ACT) was transferred to *Lla*$HisG_S$, in an attempt to replicate the ACT domain recruitment event that led to the long form ATP-PRTase.

Two molecular architectures of ATP-PRTase have been biochemically characterised, the homo-hexameric long form and the hetero-octameric short form.[3,4,19,20] A search of the Pfam database

revealed a third sequence structure in which *hisZ* and *hisG*$_S$ constitute a single open reading frame, encoding a putative covalently linked short form ATP-PRTase (Figure 4.1).[87] This chapter describes the cloning and purification of such a protein from *Leuconostoc mesenteroides* (*Lme*HisZG$_{Fusion}$). The generation of two covalently linked *Lla*HisZG$_S$ proteins, in an attempt to mimic the gene fusion event that led to the third ATP-PRTase molecular architecture, is also discussed.

Figure 4.1 Schematic representation of each of the three sequence structures of ATP-PRTase. The short form of ATP-PRTase is assembled from the non-covalent association of HisZ (red) and HisG$_S$ (blue). The long form is composed of HisG$_L$ chains, which consist of a similar catalytic domain to HisG$_S$ (blue) and a regulatory ACT domain (pink). The covalent short form is comprised of HisZ (red) and HisG$_S$ (blue) domains joined by a short, flexible linker (green).

4.2 Cloning, Expression, Purification and Homology Modelling of *Lme*HisZG$_{Fusion}$

A search of the Pfam database revealed 35 ATP-PRTase sequences in which HisZ and HisG$_S$ comprise a single open reading frame.[87] Expression of this open reading frame is expected to generate a polypeptide chain with an N-terminal HisZ domain covalently linked to a C-terminal HisG$_S$ domain. Analysis of the amino acid sequence of this hypothetical fusion protein revealed the presence of conserved ATP-PRTase motifs, including the PRPP binding motif. The bacterial

genomes that possess the HisZG$_{Fusion}$ sequence also encode the complete his operon. It is therefore a reasonable assumption that these organisms are histidine prototrophs and therefore express a functional ATP-PRTase.

4.2.1 Homology Model of *Lme*HisZG$_{Fusion}$

Despite the development of high throughput methods for protein crystallography, the rate of protein structure determination has been eclipsed by the rapid advancement of genome sequencing.[88-90] As a result, the number of protein sequences available greatly exceeds the availability of structural and biochemical information. Homology modelling methods are therefore useful to predict the three dimensional structure of a protein from its sequence, using information derived from the structure of a homologous protein.[91-93] Homology models have proven useful for structure-based drug design and for the development of biochemical experiments, such as site directed mutagenesis, to probe the structure-function relationship.

The Schrödinger program suite was employed by Dr Wanting Jiao to construct a homology model of *Lme*HisZG$_{Fusion}$. The X-ray crystal structure of *Lla*HisZG$_S$ (PDB 1Z7M) was used as a template. *Lme*HisZG$_{Fusion}$ shares 22 % sequence identity with *Lla*HisZ (Figure 4.2) and 39 % sequence identity with *Lla*HisG$_S$ (Figure 4.3). These values are consistent with the identities across the short form ATP-PRTase family.[19] The homology model shows the *Lme*HisZG$_{Fusion}$ chains arranged in a homo-tetramer, reminiscent of the hetero-octameric structure of *Lla*HisZG$_S$ (Figure 4.5). The *Lme*HisZ and *Lme*HisG$_S$ domains are connected by a flexible loop (Figure 4.4). An overlay of the

homology model of *Lme*HisZG$_{Fusion}$ and the *Lla*HisZG$_S$ crystal structure revealed a similarity in both the histidine binding sites and the position of the PRPP binding motif (Figure 4.6).

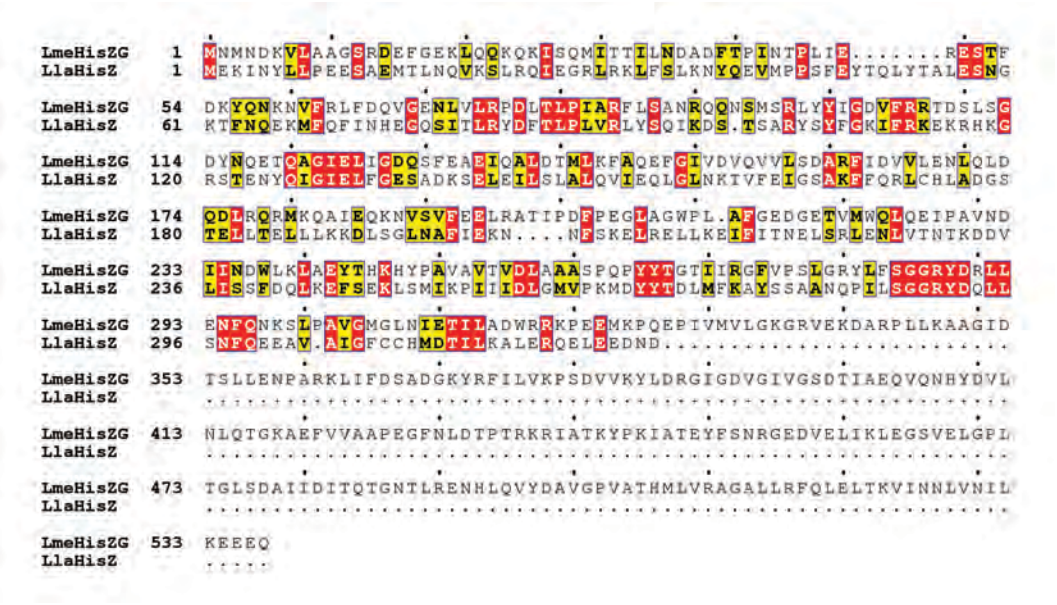

Figure 4.2 Sequence alignment of *Lme*HisZG$_{Fusion}$ and *Lla*HisZ. Conserved residues are coloured red and conserved functionalities are shown in yellow. The two proteins share 22 % sequence identity.

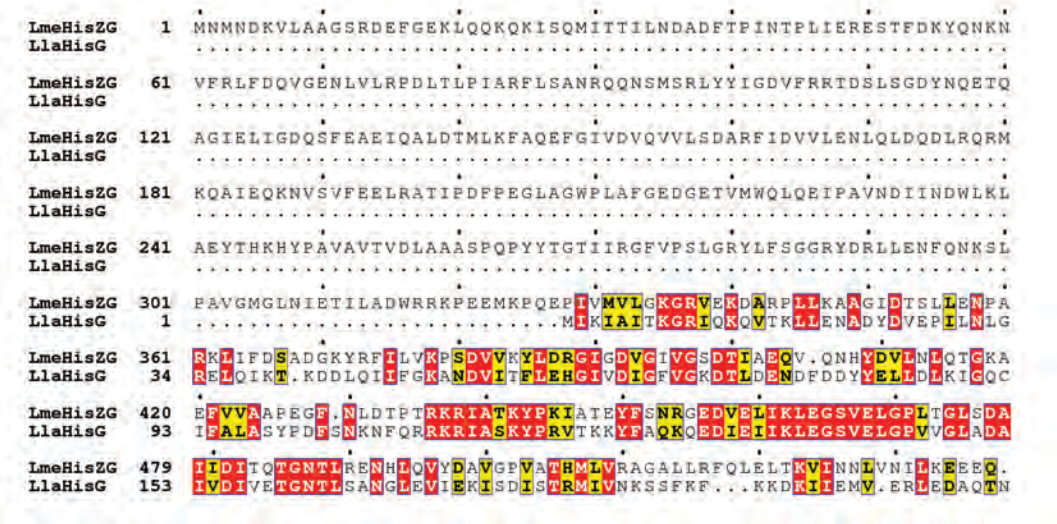

Figure 4.3 Sequence alignment of *Lme*HisZG$_{Fusion}$ and *Lla*HisG$_S$. Conserved residues are highlighted in red and conserved functionalities are shown in yellow. The sequence identity between *Lla*HisG$_S$ and *Lme*HisZG$_{Fusion}$ is 39 %.

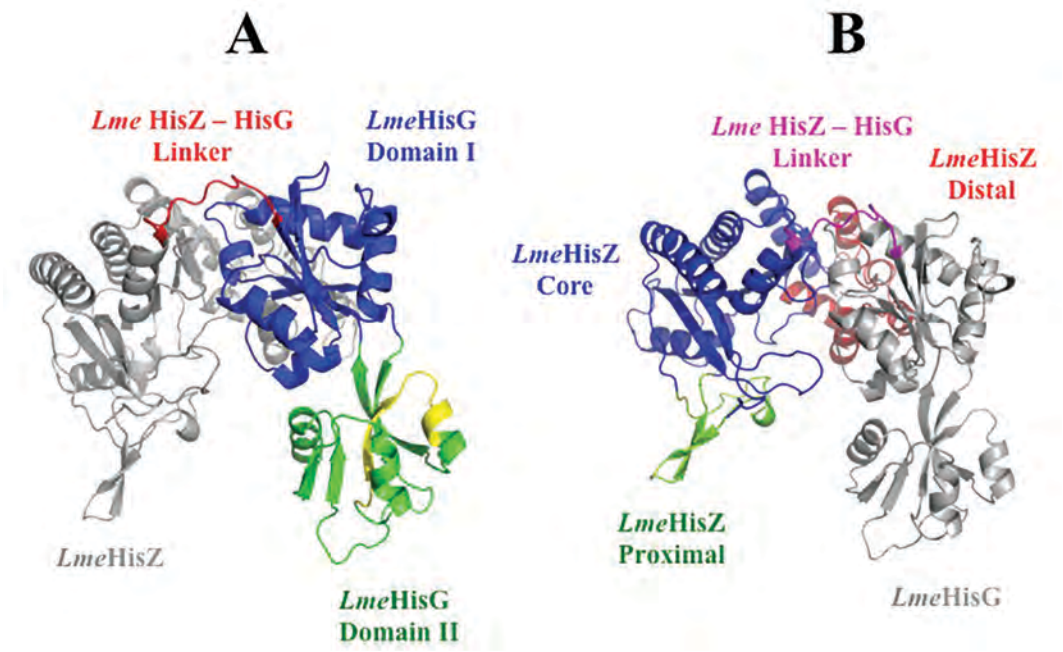

Figure 4.4A) A single chain of *Lme*HisZG$_{Fusion}$ showing *Lme*HisG$_S$ coloured by domain and *Lme*HisZ coloured grey. The PRPP binding motif is shown in yellow and the linker region is coloured red. B) The monomeric unit of *Lme*HisZG$_{Fusion}$ with *Lme*HisZ coloured by domain, *Lme*HisG$_S$ coloured grey and the linker region shown in magenta. Note that the HisZ-HisG$_S$ linker is likely to be highly flexible and the relative positioning of *Lme*HisZ and *Lme*HisG$_S$ is due to the constraints imposed by the formation of a tetrameric assembly.

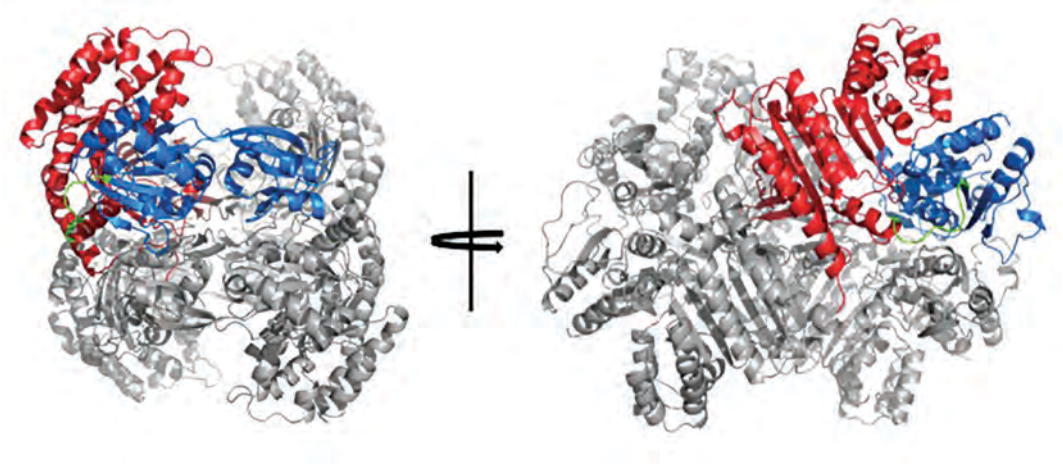

Figure 4.5 Two views of the tetrameric homology model of *Lme*HisZG$_{Fusion}$ with one HisZ domain coloured red, one HisG$_S$ domain coloured blue, and the linker between the two domains coloured green.

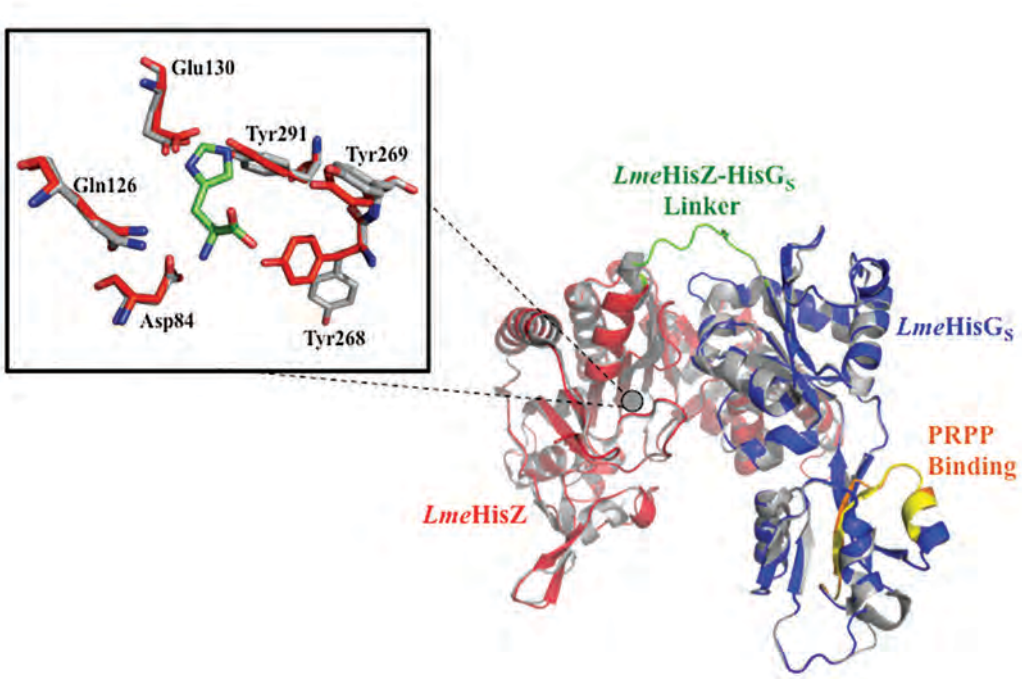

Figure 4.6 Alignment of a single chain of *Lme*HisZG$_{Fusion}$ coloured by domain, with single chains of *Lla*HisZ and *Lla*HisG$_S$ (grey) (PDB 1Z7M). The PRPP binding motifs of *Lla*HisG$_S$ and *Lme*HisZG$_{Fusion}$ are shown in yellow and orange respectively. Conserved residues from the proposed histidine binding site are shown as an insert. The *Lme*HisZG$_{Fusion}$ residues are coloured red and the *Lla*HisZ residues are coloured grey. *Lla*HisZ numbering is shown. Each of the proposed histidine binding residues overlay well with the equivalent residues from *Lla*HisZ, with the exception of Tyr268, which adopts a more favourable conformation for histidine binding in the *Lme*HisZG$_{Fusion}$ homology model.

4.2.2 Cloning, Expression and Purification of *Lme*HisZG$_{Fusion}$

A synthetic gene encoding *LmehisZG$_{Fusion}$*, codon optimised for expression in *E. coli,* was purchased from Epoch Life Science. *LmehisZG$_{Fusion}$* was subcloned into pET-28a, with an N-terminal (His)$_6$-tag, using the standard procedure for cloning using the InFusion® HD cloning kit (Clontech), described in chapter two. Expression trials of the sequence verified construct in *E. coli* BL21*(DE3) cells revealed the presence of soluble *Lme*HisZG$_{Fusion}$, however no ATP-PRTase activity was detected in the cellular lysate. Purification of *Lme*HisZG$_{Fusion}$ via IMAC (Figure 4.7), followed by SEC (Figure 4.8), yielded approximately 20 mg of *Lme*HisZG$_{Fusion}$ per litre of cell culture. The elution volume from the preparative SEC is consistent

with the elution volume of the hetero-octameric *Lla*HisZG$_S$ from the same column. This suggests a tetrameric assembly of *Lme*HisZG$_{Fusion}$. ATP-PRTase activity was not detected in either the eluent from IMAC or the purified protein, using the standard enzyme kinetics assay with 10 mM ATP, 50 μM PRPP and up to 1 mg/mL *Lme*HisZG$_{Fusion}$. The purified *Lme*HisZG$_{Fusion}$ was concentrated to 5 mg/mL and stored at -80 °C. Precipitation of the protein precluded the analysis of secondary structure via circular dichroism and quaternary structure via analytical SEC. It is therefore unclear whether the lack of catalytic activity is the result of incorrect folding of the protein or whether *L. mesenteroides* does not express a functional ATP-PRTase. It should be noted that co-expression with chaperonins in *E. coli* Chaperone 3 cells decreased the yield of *Lme*HisZG$_{Fusion}$ and did not result in detectable ATP-PRTase activity.

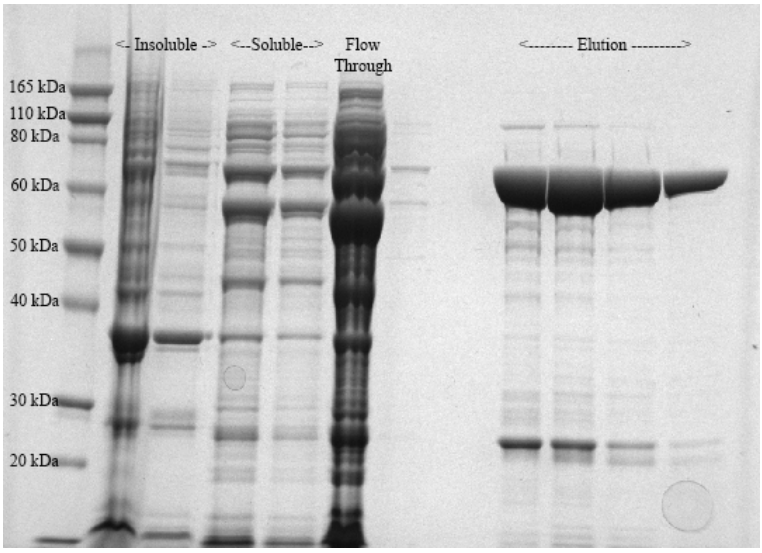

Figure 4.7 SDS-PAGE gel showing the purification of *Lme*HisZG$_{Fusion}$ by IMAC. The first lane is a molecular weight marker. The second and third lanes indicate the insoluble fraction; the soluble fraction is represented by lanes four and five, and the flow through is shown in the sixth lane. The final four lanes correspond to the peak following IMAC.

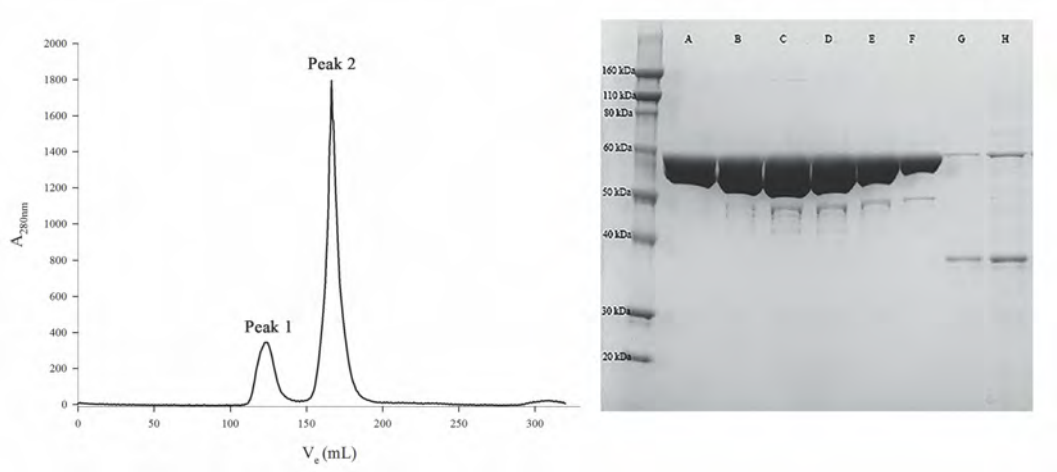

Figure 4.8 A_{280nm} trace and SDS-PAGE gel following purification of *Lme*HisZG$_{Fusion}$ by SEC. The SEC trace shows two peaks. Peak 1 corresponds to lanes G and H on the SDS-PAGE gel, which indicates the presence of two protein species. The second peak corresponds to lanes A-F on the SDS-PAGE gel and represents purified *Lme*HisZG$_{Fusion}$. There are additional lower molecular weight species present in these lanes. The retention time following preparative SEC is similar to the retention time of the hetero-octameric *Lla*HisZG$_S$, which suggests that *Lme*HisZG$_{Fusion}$ may be adopting a tetrameric quaternary structure.

4.3 Engineering a covalent linker between *Lla*HisZ and *Lla*HisG$_S$

L. mesenteroides shares a recent common ancestor with *L. lactis*,[94] which is known to express a functional short form ATP-PRTase. This section reports the engineering of a flexible linker between *Lla*HisZ and *Lla*HisG$_S$, in an attempt to mimic the domain fusion event that may have given rise to the wild-type HisZG$_{Fusion}$ proteins.

4.3.1 Linker design

There is no sequence identity between the linker regions of the wild-type HisZG$_{Fusion}$ proteins, which is unsurprising given the limited number of organisms featuring this sequence architecture. Two *Lla*HisZG$_{Fusion}$ constructs were designed, one with a short linker, denoted *Lla*HisZG$_{Fusion}$(Short) and one with a long linker (*Lla*HisZG$_{Fusion}$(Long)). Fusion sites were chosen based on regions of

conservation observed in the structure-based sequence alignment of the *Lme*HisZG$_{Fusion}$ homology model and the crystal structure of *Lla*HisZG$_S$ (PDB 1Z7M) (Figure 4.9). The *Lla*HisZG$_{Fusion}$(Short) construct was formed by the insertion of a ten amino acid *Lme*HisZG$_{Fusion}$ sequence between *Lla*HisZ D26 and *Lla*HisG$_S$ K8. The *Lla*HisZG$_{Fusion}$(Long) construct was generated by inserting a 21 amino acid sequence of *Lme*HisZG$_{Fusion}$ between *Lla*HisZ L315 and *Lla*HisG$_S$ K8.

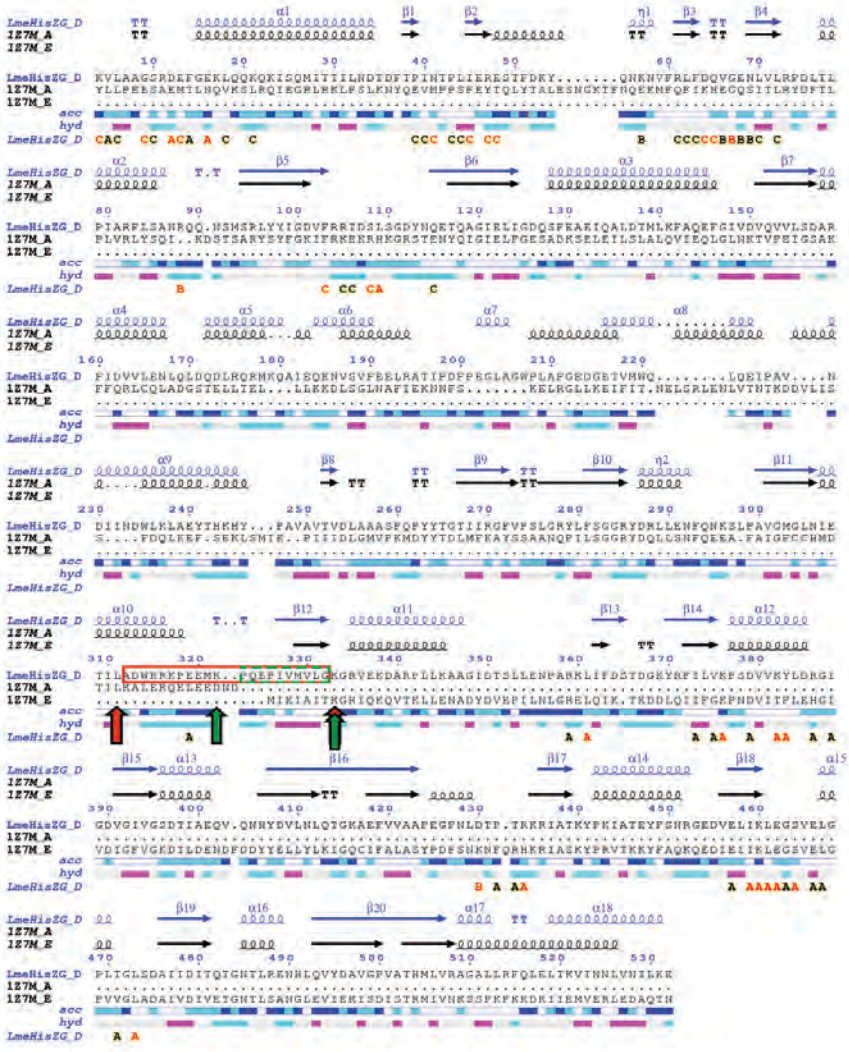

Figure 4.9 Structure-based sequence alignment of a single chain of *Lme*HisZG$_{Fusion}$ with *Lla*HisZ (1Z7M_A) and *Lla*HisG$_S$ (1Z7M_E) generated using the ENDscript server.[95] The *Lla*HisZG$_S$ long linker was constructed by inserting the *Lme*HisZG$_{Fusion}$ sequence indicated by the red box, between *Lla*HisZ L315 (red arrow) and *Lla*HisG$_S$ K8 (red arrow). The *Lla*HisZG$_{Fusion}$ short linker was constructed by inserting the *Lme*HisZG$_{Fusion}$ sequence, shown by the green box, between *Lla*HisZ D326 (green arrow) and *Lla*HisG$_S$ K8 (green arrow).

4.3.2 Cloning and expression

The *LlahisZ* and *LlahisG*$_S$ genes amplified beautifully from their respective pBSK constructs. The linker regions were added by the stepwise extension of the 3' end of the *LlahisZ* PCR product and the 5' end of the *LlahisG*$_S$ PCR product (Figure 4.10), resulting in a 23 base pair overlap between *LlahisZ* and *LlahisG*$_S$ for generation of the *LlahisZG*$_{Fusion}$*(Long)* PCR product, and a 17 base pair overlap for formation of the *LlahisZG*$_{Fusion}$*(Short)* PCR product. The final PCR reaction installed the 15 base pair recognition sites for cloning using the InFusion® HD Cloning Kit (Clontech). Each PCR product was cloned into pET-28a with an N-terminal $(His)_6$-tag. Expression of the sequence verified constructs in *E. coli* BL21*(DE3) cells yielded soluble protein (Figure 4.11). ATP-PRTase activity was not detected in either the cellular lysate or the eluent from IMAC, using the standard enzyme kinetics assay with 50 μM PRPP and 10 mM ATP. Purification and characterisation of these proteins was not pursued due to the paucity of catalytic activity.

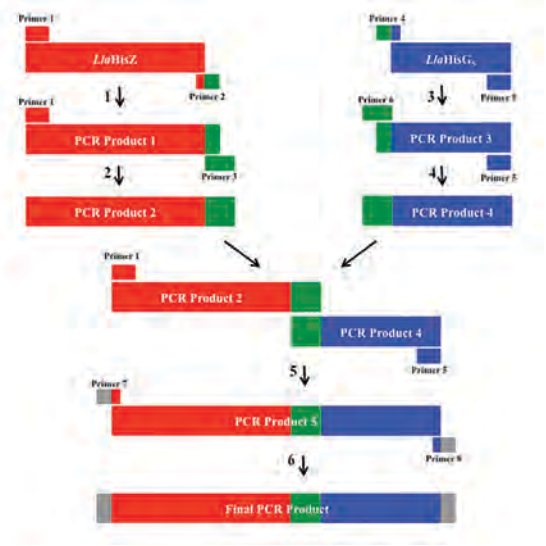

Figure 4.10 Schematic representation of the linker formation (green) between *Lla*HisZ (red) and *Lla*HisG$_S$ (blue) via five PCR reactions. The final PCR step installed the 15 base pair recombination sequence for the InFusion® reaction (grey). Note that the *LlahisZG*$_{Fusion}$*(Short)* PCR product was formed via four PCR steps, as the linker region was sufficiently short to generate an overlapping region by a single PCR amplification of *LlahisZ*.

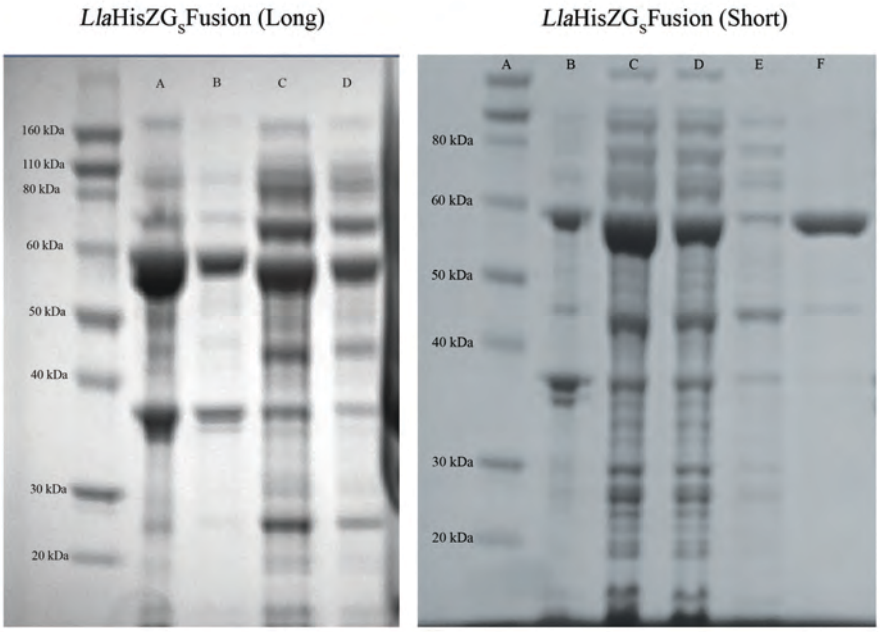

Figure 4.11 (left) SDS-PAGE gel following the expression trial of *Lla*HisZG$_{Fusion}$(Long) in BL21*(DE3) cells. Lanes A and B show the insoluble fractions and lanes C and D show the soluble fraction. A band at approximately 60 kDa indicates that *Lla*HisZG$_{Fusion}$(Long) is present in both the soluble and insoluble fractions. (right) SDS-PAGE gel following the expression trial of *Lla*HisZG$_{Fusion}$(Short). Lane A shows the insoluble fraction, lane B shows the soluble fraction, lane C shows the flow through from a spin HisTrap column, lane D shows the wash step of the spin HisTrap and lane E shows the eluent. A band at approximately 60 kDa indicates that *Lla*HisZG$_{Fusion}$(Short) is predominantly found in the soluble fraction and is able to bind the IMAC column.

4.4 Transfer of the ACT regulatory domain from *Cje*HisG$_L$ to *Lla*HisG$_S$

The ACT domain is a common regulatory motif found in enzymes involved in amino acid metabolism.[31,32] The simple transfer of an ACT domain has been shown to confer functional allostery to an unregulated enzyme.[96] The HisZG$_S$ complex is considered to be the phylogenetically older form of ATP-PRTase and HisG$_L$ may have subsequently evolved via the recruitment of an ACT domain. In order to explore this relationship, a chimeric protein was constructed whereby the ACT domain of HisG$_L$ from *C. jejuni* (*Cje*HisG$_L$) was fused to the C-terminus of *Lla*HisG$_S$ via PCR.

4.4.1 Choice of Fusion Site

The long form ATP-PRTase enzymes from *M. tuberculosis,*[3] *E. coli*,[4] *C. jejuni* (Gerd Mittelstädt personal communication) and the short form ATP-PRTase enzymes from *L. lactis*[20] and *T. maritima*[19] have been structurally characterised. Removal of the ACT domain from *Cje*HisG$_L$ results in a homo-dimeric enzyme with diminished ATP-PRTase activity, and preliminary investigation has suggested that the *Cje*ACT domain is trimeric in the absence of the catalytic core of *Cje*HisG$_L$ (Gerd Mittelstädt, personal communication). *Lla*HisG$_S$ has been structurally characterised as a component of the *Lla*HisZG$_S$ complex. Preliminary results from *E. coli* BW25113Δ*hisG* complementation studies, reported in section 4.5, suggest that *Lla*HisG$_S$ may exhibit marginal ATP-PRTase activity in the absence of *Lla*HisZ. *Lla*HisG$_S$ was chosen as the recipient of the ACT domain as it is mesophilic, unlike *Tma*HisG$_S$, which minimises the chance of incompatibility of the optimal functional temperature between the catalytic core and the ACT domain. *Cje*HisG$_L$ was chosen as the ACT domain donor. The observation that the ACT domain may independently associate to form a trimer is particularly important, as it suggests that transfer of the ACT domain may be sufficient to form a hexameric ATP-PRTase.

The overall fold of the *Lla*HisG$_S$ and *Cje*HisG$_L$ catalytic domains are very similar.[§] The most prominent differences are the unresolved β11-α6 loop in *Cje*HisG$_L$, which is helical in *Lla*HisG$_S$, and the length of the *Cje*HisG$_L$ α6 helix connecting domain I to the ACT domain,

[§] A single chain of *Cje*HisG$_{core}$ aligns to a single chain of *Lla*HisG$_S$ from the *Lla*HisZG$_S$ crystal structure with an RMSD of 1.2 Å (PDB 1Z7N) and 1.3 Å (PDB 1Z7N).

which is seven residues longer than the equivalent helix in *Lla*HisG$_S$ (Figure 4.12). *Lla*HisG$_S$ shares 28 % sequence identity with *Cje*HisG$_L$ and the choice of fusion site was informed by a structure-based sequence alignment of the two proteins. Two possible fusion sites were considered. The first was the replacement of the *Lla*HisG$_S$ sequence after β11 with the equivalent *Cje*HisG$_L$ sequence. This would introduce the β11-α6 loop sequence from *Cje*HisG$_L$, in addition to α6 and the ACT domain. This method was not pursued due to the possibility that introduction of the loop sequence from *Cje*HisG$_L$ in place of the kinked helix of *Lla*HisG$_S$, may result in incorrect protein folding. The decision was made to extend α7 of *Lla*HisG$_S$ (equivalent to α6 of *Cje*HisG$_L$) by the addition of the *Cje*HisG$_L$ sequence after residue R202 (R216 in *Cje*HisG$_L$). R216/R202 was chosen as the fusion point due to the conservation of this residue.

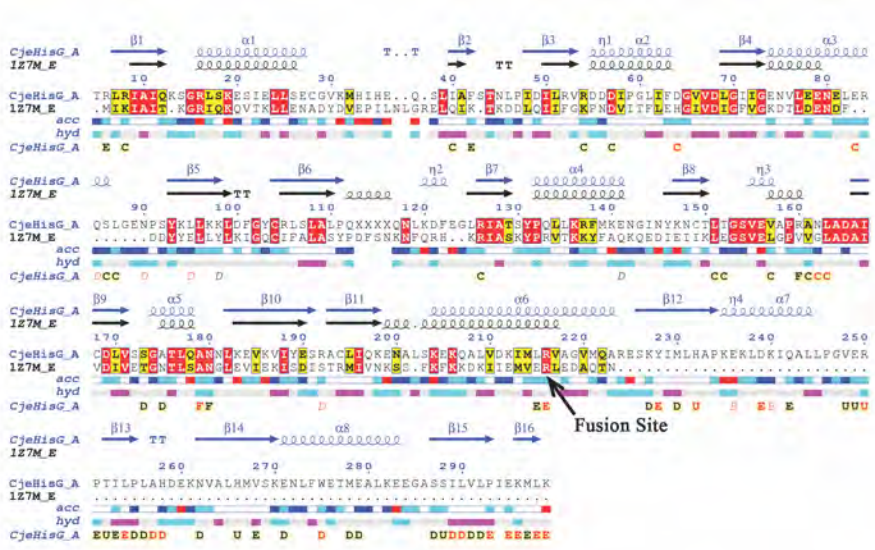

Figure 4.12 Structure-based sequence alignment of *Cje*HisG$_L$ chain A and *Lla*HisG$_S$ (1Z7M_E) generated by the ENDscript server. The fusion site at R216 (*Cje*HisG$_L$ numbering) is indicated.

4.4.2 Cloning and Expression

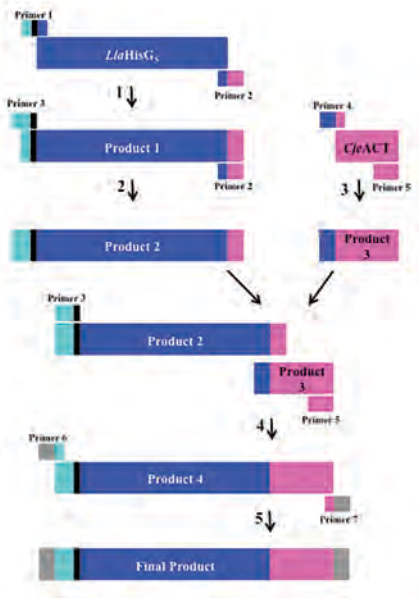

Figure 4.13 Schematic of the construction of the *Lla*HisG*S*-*Cje*ACT chimera using PCR. Step 1 amplified *LlahisGS* (blue) from the template vector and partially installed the 5' sequence corresponding to the TEV cleavage site (cyan) and a linker (black) as well as a 3' extension complementary to the sequence of *CjeACT* (pink). The second reaction installed the remainder of the TEV cleavage site. The third reaction amplified *CjeACT* with a 5' extension complementary to *LlahisGS*. The fourth reaction stitched the two products together and the fifth installed the recognition sequence for the InFusion enzyme (grey).

The *LlahisGS-CjeACT* chimera was constructed using five PCR reactions (Figure 4.13). The first reaction amplified *LlahisGS* from the *LlahisGS*pBSK vector with partial installation of the TEV cleavage site at the 5' end, and a 3' extension complementary to the sequence of *CjeACT*. The remainder of the TEV cleavage site was incorporated in the second PCR reaction. The third step amplified the *CjeACT* sequence from the *CjehisGL*pDONR221 vector and installed a 5' extension complementary to the *LlahisGS* sequence. The fourth PCR step fused the two genes together and the fifth installed the 15 base pair recognition sequence for the InFusion® reaction. The *LlahisGS-CjeACT* PCR product was cloned into pET-28a with an N-terminal $(His)_6$-tag using the standard procedure for cloning using the InFusion® HD cloning kit (Clontech). Expression trials of the sequence

verified construct in *E. coli* Chaperone 3 cells resulted in soluble *Lla*HisG$_S$-*Cje*ACT (Figure 4.14), however catalytic activity was not observed in the cellular lysate, or the eluent following IMAC, using the standard enzyme kinetics assay with up to 1 mg/mL protein.

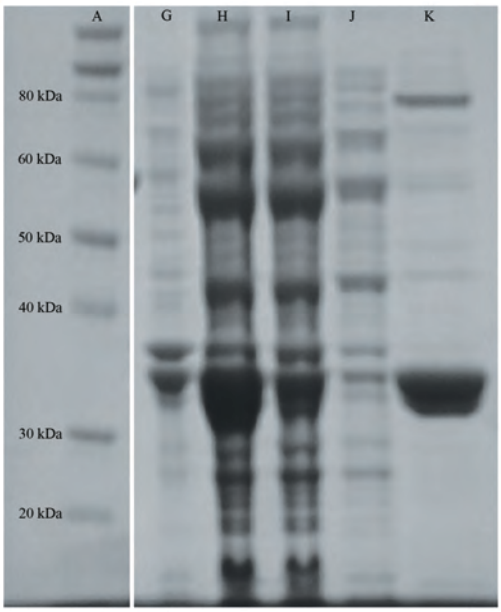

Figure 4.14 SDS PAGE gel of the (His)$_6$-tagged *Lla*HisG$_S$-*Cje*ACT chimeric protein purification via IMAC. Lane A corresponds to a molecular weight marker, Lanes G and H show the insoluble and soluble fractions respectively, Lane I corresponds to the IMAC flow through, Lane J shows the IMAC wash step and Lane K shows the IMAC eluent. The band at approximately 35 kDa corresponds to the (His)$_6$-tagged *Lla*HisG$_S$-*Cje*ACT protein.

4.5 *E. coli* BW25113Δ*hisG*::kan Complementation Studies

Complementation studies of *E. coli* BW25113Δ*hisG*::kan cells from the Keio collection[97] with plasmids encoding *Eco*HisG$_L$, *Cje*HisG$_L$, *Mtu*HisG$_L$, *Lla*HisZG$_S$, *Nme*HisZG$_S$, *Eco*HisG$_{core}$, *Cje*HisG$_{core}$, *Lme*HisZG$_{Fusion}$, *Lla*HisG$_S$, *Nme*HisG$_S$, *Lla*HisZ+*Nme*HisG$_S$, *Nme*HisZ+*Lla*HisG$_S$, *Lla*HisZG$_{Fusion}$(short), *Lla*HisZG$_{Fusion}$(long) and *Lla*HisG$_S$-*Cje*ACT were used to probe the ATP-PRTase activity of each of these proteins *in vivo*. The *E. coli* BW25113Δ*hisG*::kan cells lack the

hisG gene that encodes ATP-PRTase, and are therefore auxotrophic for histidine. Growth of these cells is supported only by the addition of histidine to the growth medium or by complementation with a plasmid encoding a functional ATP-PRTase. Complementation studies allow the visualisation of proteins that may exhibit minimal catalytic activity, not detected spectroscopically, and are not confounded by purification methods or assay conditions.

Table 4.1 Growth of the *E. coli* BW25113 Δ*hisG* strain transformed with plasmids encoding functional and putative ATP-PRTases on minimal media agar plates

	Detection of Growth	
Protein Encoded By Transformed Plasmid	+ Histidine	+ IPTG
Untransformed Strain	✓	✗
*Eco*HisG$_{L}$	✓	✓
*Cje*HisG$_{L}$	✓	✓
*Mtu*HisG$_{L}$	✓	✓
*Lla*HisZG$_{S}$	✓	✓
*Nme*HisZG$_{S}$	✓	✓
*Eco*HisG$_{core}$	✓	*
*Cje*HisG$_{core}$	✓	✗
*Lla*HisG$_{S}$	✓	*
*Nme*HisG$_{S}$	✓	*
*Lme*HisZG$_{Fusion}$	✓	✗
*Lla*HisZ+*Nme*HisG$_{S}$	✓	*
*Nme*HisZ+*Lla*HisG$_{S}$	✓	*
*Lla*HisZG$_{Fusion}$(Short)	✓	✗
*Lla*HisZG$_{Fusion}$(Long)	✓	✗
*Lla*HisG$_{S}$-*Cje*ACT	✓	✗

Key: ✓ indicates definite growth, ✗ indicates no growth, and * indicates marginal growth (Figure 15).

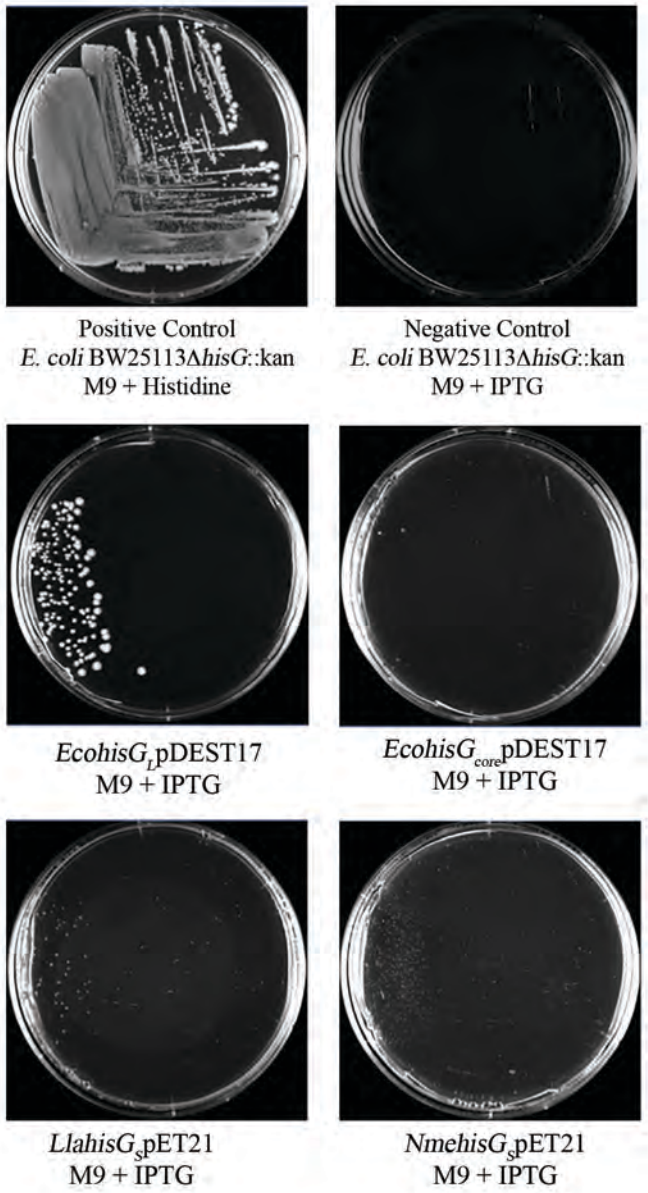

Figure 4.15 Examples of *E. coli* EW25113Δ*hisG* knockout complementation plates showing growth (✓) (positive control and *Eco*HisG$_L$), no growth (✗) (negative control) and marginal growth (*) (*Eco*HisG$_{core}$, *Nme*HisG$_S$, and *Lla*HisG$_S$).

Transformed and untransformed (control) *E. coli* BW25113ΔhisG::kan cells were plated on M9 minimal medium supplemented with the appropriate antibiotics and either IPTG (0.1 mM), to induce protein expression, or histidine (13 μM) (Table 4.1)

(Figure 4.15).** Plasmids encoding wild type *Eco*HisG$_L$, *Cje*HisG$_L$, *Mtu*HisG$_L$, *Lla*HisZG$_S$, and *Nme*HisZG$_S$ supported growth on histidine deficient media, whilst *Lme*HisZG$_{Fusion}$ was unable to rescue the strain. This observation, combined with the absence of observed catalytic activity of the purified *Lme*HisZG$_{Fusion}$, suggests that the *LmehisZG$_{Fusion}$*pET-28a construct does not express a functional ATP-PRTase. Growth of the cell line was supported by *Eco*HisG$_{core}$ but not by *Cje*HisG$_{core}$, which is expressed as a GST-tagged construct. Interestingly, *Nme*HisG$_S$ and *Lla*HisG$_S$ both appeared to rescue the strain, albeit poorly, in the absence of their respective HisZ partners. This observation contradicts both the previously reported complementation studies, and the absence of ATP-PRTase activity detected *in vitro*.[5] Mix and match of HisZ and HisG$_S$ from *N. meningitidis* and *L. lactis* rescued the strain, however neither the *Lla*HisZG$_{Fusion}$ proteins nor the *Lla*HisG$_S$-*Cje*ACT chimera were able to restore ATP-PRTase activity.

4.6 Discussion

Allostery can arise from the structural modification of a protein via the reshaping or repurposing of existing domains, formation of multimeric assemblies, or domain recruitment.[86] A combination of all three mechanisms likely gave rise to the contemporary ATP-PRTases. The work described in this chapter aimed to characterise a third putative ATP-PRTase architecture and to facilitate the acquisition of allostery by mimicking domain fusion events.

** See Appendix II for figures.

The bacteria from which the 35 putative $HisZG_{Fusion}$ sequences originate are typically associated with the human gut and fermentation. Despite colonising a nutrient rich environment, each of these organisms encodes the complete histidine biosynthetic pathway. Comparative genomics studies on lactic acid producing bacteria, including *L. mesenteroides* and *L. lactis,* noted a reduction of the genome, presumably in response to occupation of a nutrient rich ecological niche. It should be noted that *L. lactis*, has retained a functional, metabolically essential ATP-PRTase.[5,26,39,98] Few bacterial species have been shown to exhibit complete or partial deletion of the his operon, which is testament to the essentiality of this pathway.[10] Some bacteria devoid of the genes responsible for histidine biosynthesis, such as *Mycobacterium genitalium* and *Helicobacter pylori,* are able to colonise and persist in a host organism.[99-102] Juliao *et al.* observed that strains of *Haemophilus influenzae* that persist in the throat are auxotrophic for histidine, whilst those that colonise the nutrient deficient environment of the middle ear express a functional histidine biosynthetic pathway.[103] Although these organisms are niche limited, the energy conservation afforded by histidine auxotrophs may be selectively advantageous. To date there have been no bacteria reported that are both auxotrophic for histidine and encode the complete histidine biosynthetic pathway, however the proportion of bacteria that have been studied with respect to histidine biosynthesis is limited compared to the number of bacterial species that exist.

Several enzymes involved in histidine metabolism consist of multiple domains or multiple subunits, although not all of the apparent gene fusion events have resulted in catalytically competent enzymes.[2,104] Ino *et al.* observed that in some strains of *S. typhimurium, hisG*$_L$ and *hisD* comprise a single open reading frame

that translates to a putative bifunctional enzyme.[104] Expression and purification of the gene product revealed a protein that displayed neither ATP-PRTase (*hisG*$_L$) nor histidinol dehydrogenase (*hisD*) activity. Ino *et al.* subsequently proposed that the fusion protein is cleaved by endopeptidase activity *in vivo*. Bifunctional enzymes typically catalyse sequential reactions, whilst HisG$_L$ and HisD represent the first and last steps of the pathway respectively.[2,105,106] The functional integrity of the HisG$_L$-HisD fusion protein is less conceivable than that of the HisZG$_{Fusion}$, as the structure of the hexameric HisG$_L$ differs significantly from the dimeric HisD. HisG$_L$ and HisD both function independently and association of the two enzymes has not been reported. In contrast, HisZ and HisG$_S$ are known to assemble non-covalently, it a 1:1 ratio, to form a functional ATP-PRTase.[39]

The observation that *Lme*HisZG$_{Fusion}$ is devoid of ATP-PRTase activity, both *in vitro* and by complementation *in vivo* is perplexing. Sequence alignments revealed that the putative fusion protein possesses the known sequence motifs characteristic of the short form ATP-PRTase family. *Lme*HisZG$_{Fusion}$ shares 39 % sequence identity with *Lla*HisG$_S$ and 22 % sequence identity with *Lla*HisZ, values that are comparable to the identities across the functional short form ATP-PRTase family. The genome of *L. mesenteroides* encodes the entire cohort of histidine biosynthetic enzymes, however biochemical characterisation of this pathway has not been reported. Jin *et al.* noted that addition of histidine to the growth media of *L. mesenteroides* is not essential for proliferation, but appears to enhance the growth rate.[107] Taken together, these observations suggest that *L. mesenteroides* may procure histidine from a nutrient rich environment, however the histidine biosynthetic pathway is expected to be functional to support

survival of the bacteria in the absence of an exogenous source of histidine.

Further investigation is required to understand why the *LmehisZG*$_{Fusion}$pET-28a construct does not appear to express a functional ATP-PRTase. Precipitation of the protein following storage precluded the analysis of the secondary structure; therefore the absence of catalytic activity cannot be definitively attributed to inactivity of the HisZG$_{Fusion}$ molecular architecture. The elution volume from the preparative SEC is consistent with a tetrameric quaternary structure, however it is unclear whether the protein was correctly folded. Optimisation of the buffer composition may prevent precipitation of the protein. Circular dichroism and analytical SEC may subsequently be employed to characterise the secondary and quaternary structures respectively. There is some evidence to suggest that the addition of a $(His)_6$-tag could lead to improper folding and diminished catalytic activity of an enzyme.[108] *E. coli* complementation studies with a tagless *LmehisZG*$_{Fusion}$ construct may also be used to probe the potential impact of the $(His)_6$-tag on catalysis, without being confounded by assay conditions. Additionally *LmehisZG*$_{Fusion}$ knockout studies may be employed to assess the essentiality of this gene product to the survival of the strain. Future work could also assess whether fission of *Lme*HisZG$_{Fusion}$ into its constituent HisZ and HisG$_S$ subunits leads to a functional non-covalent complex, resembling the short form ATP-PRTases.

Neither of the engineered *Lla*HisZG$_{Fusion}$ proteins exhibited catalytic activity. The choice of fusion site was relatively ambiguous due to the low sequence identity between the linker regions of the native HisZG$_{Fusion}$ proteins and the absence of the five C-terminal

*Lla*HisZ residues from the *Lla*HisZG$_S$ crystal structure. As with *Lme*HisZG$_{Fusion}$ it is unclear whether the loss of ATP-PRTase activity is the result of improper folding. The wild type *Lla*HisZ is involved in the activation of *Lla*HisG$_S$ via assembly into a hetero-octameric complex, although the mechanism for this activation is yet to be deduced. It is possible that installation of a covalent linker affected the relative domain orientation, the formation of crucial contacts, or the protein dynamics that facilitate catalysis. Future work should focus on understanding the function of the wild type *Lla*HisZG$_S$ and *Lme*HisZG$_{Fusion}$, which may subsequently inform the construction of a chimeric protein.

Previous research has shown that the simple transfer of an ACT domain from one enzyme is sufficient to confer functional allostery to an unregulated ortholog.[96] It is crucial to note that in this instance the regulated and unregulated enzymes are catalytically active, adopt similar structures, and allostery was conferred by the transfer of a distinct domain that is responsible for a large conformational change. In contrast, sequence analysis suggests that all ATP-PRTase enzymes possess a regulatory element in the form of either an ACT domain or a HisZ subunit. The observation of diminished activity of a truncated HisG$_L$ that has lost the ACT domain, and marginal activity of HisG$_S$ in the absence of HisZ, suggests that these regulatory elements also promote catalysis. The ACT domain is largely responsible for the formation of the homo-hexameric quaternary structure of HisG$_L$ therefore the ACT domain transfer is required to not only confer functional allostery but alter the quaternary structure of the HisG$_S$ protein.[3,4] Cross *et al.* also noted the importance of the choice of fusion site.[96] One key difference between the available structures of HisG$_S$ and HisG$_L$ is that the β11-α6 linker is unresolved in all HisG$_L$

structures, which implies flexibility of this region, whilst the equivalent linker in $HisG_S$ adopts helical character. Flexibility of this region may be required for functionality of the long form ATP-PRTase and future work could investigate whether inclusion of this flexible region in the engineered chimera could generate a functional ATP-PRTase.

Complementation of the *E. coli* BW25113Δ*hisG* strain with plasmids encoding wild type $HisG_L$ and $HisZG_S$ restored ATP-PRTase activity, facilitating growth on minimal media plates supplemented with IPTG. The observation that the *Mtuhis*G_LpET-24a construct rescues the strain is of particular importance as this construct is yet to yield catalytically active protein *in vitro* (Gerd Mittelstädt personal communication). This observation suggests that *in vivo* complementation may be used to detect enzymes whose apparent inactivity is the result of purification or assay conditions. The lack of catalytic activity, both *in vivo* and *in vitro*, of the native and engineered fusion proteins discussed in this chapter suggests that the gene products are completely devoid of ATP-PRTase activity.

Interestingly, *Nme*$HisG_S$ and *Lla*$HisG_S$ both appeared to rescue the strain, albeit poorly, in the absence of their respective HisZ partners. The untransformed strain failed to proliferate without an external source of histidine and the experiments were repeated in triplicate, which suggests that the observed growth is not an artefact. Sissler *et al.* reported that an *E. coli* knockout strain required complementation of both *LlahisZ* and *LlahisGS* in order to grow on histidine deficient media.[5] They also observed low expression levels of *Lla*$HisG_S$. One explanation for this discrepancy is that *Lla*$HisG_S$ may exhibit marginal catalytic activity in the absence of *Lla*HisZ and that

the increased expression level of the *Llahis*G_SpET-21a construct utilised in this experiment was sufficient to rescue the strain. Alternatively, $HisG_S$ may form an assembly with another protein in the cell that fulfils the role of HisZ.

The observed activity of the $HisG_{core}$ proteins gives credence to the hypothesis of a marginally active $HisG_S$. The $HisG_{core}$ enzymes from *C. jejuni* and *E. coli*, discussed in chapter three, are truncated forms of $HisG_L$ that are missing the ACT domain and show diminished ATP-PRTase activity. *Cje*$HisG_{core}$ is homo-dimeric (Gerd Mittelstädt personal communication) as is $HisG_S$ in the absence of HisZ.[39] Complementation of BW25113Δ*hisG* cells with *Ecohis*G_{core}pDEST-17 is sufficient to restore ATP-PRTase activity, however *Cjehis*G_{core}pDEST-15 was unable to rescue the strain. *Cje*$HisG_{core}$ expresses as a fusion protein with an N-terminal GST-tag. The inability to complement the strain is likely the result of dimerisation of GST, which is a similar size to *Cje*$HisG_{core}$, impeding catalysis.

Molecular mix and match between HisZ and $HisG_S$ of *L. lactis* and *N. meningitidis* also appeared to restore ATP-PRTase activity to the cell. It is unclear if this is due to the formation of a mixed species $HisZG_S$ complex or if it is a reflection of $HisG_S$ activity in the absence of HisZ. Catalytic activity was not observed with a standard kinetic assay following the *in vitro* incubation of *Lla*HisZ with *Nme*$HisG_S$, however the *in vitro* combination of *Nme*HisZ and *Nme*$HisG_S$ also failed to form a functional ATP-PRTase (as described in chapter three). Co-purification of *Nme*HisZ with *Lla*$HisG_S$, and vice versa, may be employed to determine whether HisZ and $HisG_S$ from two genetically distinct species will form a catalytically active ATP-PRTase.

Chapter Five

5 Summary of Thesis and Future Work

This project aimed to further the current understanding of the allosteric regulation of ATP-PRTase, and to explore the functional and evolutionary relationship between the two molecular architectures of this enzyme.

5.1 *Lla*HisZG$_S$ Undergoes a Conformational Change in Response to Histidine Binding

Elucidation of the solution conformation of *Lla*HisZG$_S$, in the presence and absence of histidine, strongly suggested a potential mechanism of allosteric regulation involving a conformational change. In the absence of ligands, the SAXS envelope is consistent with the theoretical scattering calculated from the X-ray crystal structures of *Lla*HisZG$_S$ (PDB 1Z7M & 1Z7N), while the arrangement adopted by the histidine inhibited enzyme bears little resemblance to any available ATP-PRTase structure. *Ab initio* and rigid body modelling revealed elongation of the enzyme in the presence of histidine. However a more rigorous approach is required to glean further insight into the domain movements associated with this rearrangement.

Histidine binding to *Lla*HisZ, in the presence and absence of *Lla*HisG$_S$, was demonstrated by ITC. This observation supports the allosteric binding site proposed by Champagne *et al.*,[20] over the

interface sites proposed by comparison with the *Tma*HisZG$_S$ structure. Induced fit docking revealed three additional candidates for the histidine binding residues (Tyr271, Asp84, Gln126). Generation of point mutants, targeting each of the proposed histidine binding residues, coupled with ITC or saturation transfer difference NMR and kinetics, is required to confirm the functional histidine binding residues.

5.2 Progress Has Been Made Towards the Purification and Characterisation of *Nme*HisZG$_S$ and *Eco*HisG$_L$

Purification of *Nme*HisZG$_S$ and *Eco*HisG$_L$ proved difficult due to the propensity of each enzyme to adopt higher order multimeric states, as evidenced by both analytical SEC and SAXS. Although Numerous purification conditions for *Eco*HisG$_L$ were tested, no evidence of a hexameric, tetrameric, or dimeric species was observed. Extensive effort towards the optimisation of the *Nme*HisZG$_S$ purification produced a low yield of hetero-octameric *Nme*HisZG$_S$, but this species could not be reliably separated and may be in equilibrium with a higher order species. The tendency of each enzyme to adopt a non-biological assembly precluded the biochemical characterisation of either protein and further optimisation of the two purification protocols is required.

5.3 *L. mesenteroides* Encodes a Putative Covalently Linked Short Form ATP-PRTase

A search of the Pfam database[87] revealed 35 sequences in which HisZ and $HisG_S$ comprise a single open reading frame, encoding a putative $HisZG_{Fusion}$ ATP-PRTase. The bacteria that encode this third putative ATP-PRTase architecture possess complete his operons and are therefore expected to be prototrophic for histidine. The $HisZG_{Fusion}$ protein from *L. mesenteroides* was cloned, expressed, and purified but no ATP-PRTase activity was detected using the standard enzyme kinetics assay or by *in vivo* complementation studies.

5.4 Domain Fission Results in an Enzyme with Diminished Activity and Domain Fusion is Insufficient to Confer Functionality

It is clear from the *E. coli* BW25113Δ*hisG* complementation studies of *Eco*$HisG_{core}$, as well as the biochemical characterisation of *Cje*$HisG_{core}$ (Gerd Mittelstädt, personal communication), that truncation of $HisG_L$, to remove the ACT domain, results in a catalytically competent ATP-PRTase. This result is surprising in light of previous research suggesting that $HisG_S$ requires HisZ to reconstitute a functional ATP-PRTase. Catalytic activity of *Nme*$HisG_S$ and *Lla*$HisG_S$ was not detected *in vitro,* however the observation that both $HisG_S$ constructs complement the *E. coli* BW25113Δ*hisG* strain, albeit poorly, warrants further investigation.

Engineering two covalent linkers between *Lla*HisZ and *Lla*$HisG_S$, as well as transfer of the ACT domain from *Cje*$HisG_L$ to

*Lla*hisG$_S$, was not sufficient to form a functional ATP-PRTase. Previous research has shown that such domain transfer events can mimic the evolutionary gene fusion events, however the choice of fusion site is crucial for functionality.[96] Further investigation into the structure and function of the wild-type ATP-PRTase enzymes is required to inform the construction of functional chimeric enzymes.

5.5 Future Work

Allostery plays an essential role in metabolism and presents a promising target for drug design. Understanding allostery at the molecular level could shed light on cellular function and may lead to new therapeutics to combat a range of diseases. This project has opened several avenues for further investigation. The first is the maturing model of the allosteric inhibition of *Lla*HisZG$_S$. Secondly, there is scope for the development of potent inhibitors of *Nme*HisZG$_S$ that may lead to novel antibiotics to combat bacterial meningitis. Finally, the evolutionary origin of the ATP-PRTases and the functional relationship between the two families warrants further exploration.

The model of the allosteric regulation of *Lla*HisZG$_S$ described in this thesis was developed by comparison of SAXS profiles of *Lla*HisZG$_S$ in the presence and absence of histidine. SAXS may be employed to characterise the conformational response to substrate binding, as well as the synergistic inhibition of AMP and histidine. The complementarity of SAXS and small angle neutron scattering (SANS) may also be exploited to develop a more rigorous model of the conformational change in solution, utilising contrast variation with a deuterated *Lla*HisZ or *Lla*HisG$_S$. As with SAXS, there are several methods to reconstruct protein molecular architectures from SANS

profiles, using either *ab initio* methods or crystallographic data.[74,109,110] Optimisation of the crystal conditions to obtain a structure of the inhibited *Lla*HisZG$_S$ should also be pursued in order to gain high-resolution insight into the histidine-induced structural rearrangement. This may inform the design of mutagenesis experiments to probe the allosteric signal transduction pathway. Molecular dynamics simulations could also be employed to model the transition from the active to inhibited state and to understand the role of dynamics in the allosteric regulation.

*Nme*HisZG$_S$ is of particular interest as a potential target for antibiotic drug design. Linus Pauling proposed that an enzyme binds tightest to its transition state, therefore transition state analogues are expected to be potent enzyme inhibitors.[111,112] Pioneering work by Vern Schramm *et al.* has led to the determination of transition state structures and the subsequent development of tight binding inhibitors for several PRTases.[42,113-115] This technique has recently been employed to investigate the transition state of *Cje*HisG$_L$ and may be applied to *Nme*HisZG$_S$ in the future. In addition, the allosteric sites may also be exploited for the purposes of antibiotic design.

In addition, there is scope for rigorous bioinformatics analysis to probe the evolutionary relationships among the ATP-PRTases, aminoacyl-tRNA synthetases, and the periplasmic binding proteins.

Chapter 6

6 Experimental Methods

6.1 General Methods

6.1.1 Protein Structure Figures

All protein structure figures were created using the PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

6.1.2 Amino Acid Sequence Alignments

Amino acid sequence alignments were generated using the Clustal Omega server (http://www.ebi.ac.uk/Tools/msa/clustalo/). Alignment figures 4.2 and 4.3 were constructed using ESPript 3.0 server (http://espript.ibcp.fr).[95] Structure based sequence alignments were generated using ENDscript 2.0 (http://endscript.ibcp.fr).[95]

6.1.3 Purified Water

MilliQ water (referred to as MilliQ), which has been purified with a Millipore MilliQ water system, was used in all experimental procedures. The MilliQ used for molecular biology was autoclaved prior to use.

6.1.4 pH Determination

The pH of all buffer solutions was determined at room temperature using the Denver UB-10 Ultrabasic pH meter. The pH was adjusted by the addition of either NaOH (1 M or 10 M) or HCl (1 M or 10 M).

6.1.5 Antibiotic Stocks

Ampicillin (100 mg/mL), kanamycin (50 mg/mL) and spectinamycin (10 mg/mL) stock solutions were prepared in MilliQ while chloramphenicol (30 mg/mL) stocks were made up in analytical reagent ethanol. All stocks were filter sterilised and stored in microcentrifuge tubes at -80 °C.

6.1.6 Media

Super optimal broth with catabolite repression (SOC) media (2 % (w/v) tryptone, 0.5 % (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$ and 20 mM glucose) was prepared by the addition of all components, with the exception of $MgCl_2$ and glucose, to MilliQ water, which was subsequently autoclaved at 121 °C, 15 psi for 20 minutes. The media was separated into 500 μL aliquots in sterile microcentrifuge tubes. Filter sterilised glucose and $MgCl_2$ was added prior to use.

Cell cultures were grown in Lysogeny Broth (LB) media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl). LB media was prepared in MilliQ and autoclaved at 121 °C, 15 psi for 20 minutes prior to use. LB agar plates were made by the addition of agar (25 g/L) to LB (100

mL) before sterilisation. Alternatively, LB agar premix (Invitrogen) was used. The LB agar was incubated at 60 °C and appropriate antibiotics were added immediately before pouring.

6.1.7 Chemically Competent Cells

Chemically competent cells, with the exception of the commercial *E. coli* Stellar™ chemically competent cells (Clontech), were prepared using the following procedure. 100 mL of sterile LB was inoculated with 5 mL of an overnight preculture of *E. coli* BL21*(DE3), Chaperone 3, TOP 10, or BW25113Δ*hisG* cells. The LB media was supplemented with spectinomycin and chloramphenicol for the growth of Chaperone 3 cells and kanamycin for BW25113Δ*hisG* cell growth. The cultures were incubated at 37 °C with shaking (180 rpm) until an optical density of 0.4, at which point they were harvested and the pellets resuspended in 25 mL of a sterile solution of $CaCl_2$ (25 mM). The cells were incubated on ice for 30 minutes, harvested, then resuspended in 2.5 mL of the $CaCl_2$ solution and incubated on ice for two to four hours before the addition of 15 % (v/v) glycerol. The cells were separated into 50 µL aliquots and snap frozen with liquid nitrogen prior to storage at -80 °C. All tubes, pipette tips, media, and solutions were sterilised and stored at 4 °C prior to use.

6.1.8 Transformation

Chemically competent cells (50 µL aliquots) were thawed on ice prior to the addition of purified plasmid (2 µL), In-Fusion reaction mixture (5 µL), or LR reaction mixture (2 µL). The cells were incubated on ice for a further 30 minutes then subjected to a 30-45 second heat shock at 42 °C. The cells were incubated on ice for a further 2 minutes

prior to the addition of 250 µL of SOC medium and incubated at 37 °C for one hour with shaking (180 rpm). 50-200 µL of the culture was plated on LB agar supplemented with the appropriate antibiotics and incubated at 37 °C over night.

6.1.9 Polymerase Chain Reaction

PCR reactions, with the exception of colony PCR, were performed using the standard procedure for Phusion® High Fidelity DNA Polymerase (Thermo Scientific). PCR reaction mixtures (50 µL) consisted of 1× HF buffer, 200 µM deoxyribonucleotide triphosphates (dNTPs), forward and reverse primers (0.5 µM), Template DNA (50-250 ng gDNA or 1 pg-10 ng plasmid), and 10 µL of Phusion® High Fidelity DNA Polymerase. Each reaction employed either a two-step or three-step thermocycling protocol.

Colony PCR was performed using Taq DNA polymerase (New England Biolabs). PCR reaction mixtures (50 µL) consisted of 1× reaction buffer, 200 µL dNTPs, 0.5 µM gene specific forward primer, 0.5 µM reverse sequencing primer, and a toothpick scraping of a colony. The thermocycling protocol consisted of 94 °C for 5 mins followed by cooling to 80 °C at which point the method was paused and 0.2 µL Taq DNA polymerase was added. The method was subsequently resumed. 35 cycles of 94 °C for 45 seconds, 45 °C for 30 seconds, and 72 °C for 90 seconds preceded a 10 minute extension at 72 °C.

6.1.10 Agarose Gel Electrophoresis

Agarose gels were prepared using a 1% (w/v) solution of LE agarose (Seakem) in Tris-acetic acid-EDTA buffer (TAE) (40 mM Tris,

20 mM glacial acetic acid, 2 mM EDTA) heated to 60 °C prior to the addition of SYBR® Safe DNA Gel Stain (Invitrogen). Samples were mixed with 6× loading buffer (60 mM Tris-HCl, 60 mM EDTA, 0.02 % (w/v) Orange G, 0.05 % (w/v) xylene cyanol FF, 60 % glycerol). Gels were run in 1× TAE buffer using a Mini-sub® Cell GT (Bio-Rad) at 85 V for 45 minutes. A Molecular Imager® Gel Doc™ XR was used to capture images under UV (302 nm). Alternatively, DNA samples were separated on an E-Gel® 1.2 % precast agarose gel (Invitrogen) and run for 27 minutes using the E-Gel® iBase™ Power System.

DNA fragments were purified using a self cast agarose gel, prepared as above, run in 1× TAE buffer using a Mini-sub® Cell GT (Bio-Rad) at 85 V until the dye front migrated sufficiently to separate the bands (typically 20 minutes). Images were not captured in order to minimise exposure of the samples to DNA damaging UV. The gels were visualised with either the Molecular Imager® Gel Doc™ XR at 302 nm or the E-Gel® Safe Imager™ Transilluminator at 480 nm. The appropriate bands were excised and the DNA was purified using the Nucleospin® Gel and PCR Clean-up kit (Clontech). Alternatively, DNA fragments were purified with an E-Gel® CloneWell™ 0.8 % precast agarose gel (Invitrogen) using the E-Gel® iBase™ Power System. The gel was visualised using the E-Gel® Safe Imager™ Transilluminator at 480 nm and a pipette was used to extract DNA when it reached the well.

6.1.11 Restriction Enzyme Digest

Vectors were linearised by digestion with high fidelity restriction enzymes (New England Biolabs). 1 μg of plasmid DNA was combined with 10 × Cutsmart® Buffer, 1 μL of each restriction enzyme,

and MilliQ for a total volume of 50 µL. The restriction enzyme digest mixture was incubated at 37 °C for at least one hour followed by 80 °C for 20 minutes to denature the restriction enzymes. The digested vector was gel purified prior to use.

6.1.12 In-Fusion® Reaction

The procedure for the In-Fusion® reaction was adapted from the standard protocol in order to utilise half the recommended volume of the In-Fusion® HD enzyme premix (Clontech). 1 µL of 5× In-Fusion® HD enzyme premix was combined with the purified PCR product (5-100 ng), purified linearised vector (20-100 ng), and sterile MilliQ for a total volume of 5 µL. The reaction mixture was incubated at 50 °C for 15 minutes and stored on ice prior to transformation into either *E. coli* TOP 10 chemically competent cells or *E. coli* Stellar™ chemically competent cells (Clontech). Colonies were screened via colony PCR and positive colonies were grown in 5 mL LB cultures, plasmid prepped, and sequence verified.

6.1.13 Plasmid Extraction and Purification

A 5 mL LB preculture, supplemented with the appropriate antibiotics, was inoculated with a single colony of *E. coli* TOP 10 or Stellar™ chemically competent cells (Clontech) cells that had been transformed with the desired plasmid, and incubated at 37 °C with shaking (180 rpm) overnight. The cells were harvested and the plasmids extracted using the High Pure Plasmid Isolation Kit (Roche), the ChargeSwitch®-Pro Plasmid Miniprep Kit (Invitrogen) or the Nucleospin® Plasmid Easy Pure kit (Clontech). The concentration of purified plasmid was determined by measuring the absorption at 260 nm with a Nanodrop-1000 spectrophotometer.

6.1.14 DNA Sequencing

Either Canterbury Sequencing and Genotyping or the Massey Genome service provided DNA sequencing services. Approximately 250 ng of plasmid and 3.2 μM sequencing primers were supplied for the sequencing reactions.

Sequencing Primers for pDONR221

M13_F tgtaaaacgacggccagt

M13_R caggaaacagctatgacc

Sequencing Primers for pET-21 and pET-28

T7_F taatacgactcactataggg

T7_R gctagttattgctcagcgg

Sequencing Primers for pRSFDuet-1 MCSI

ACYC_Duet_Up1: ggatctcgacgctctccct

Duet_Down1: gattatgcggccgtgtacaa

Sequencing Primers for pRSFDuet-1 MCSII

Duet_UP2: ttgtacacggccgcataatc

T7_R: gctagttattgctcagcgg

6.1.15 Glycerol Stocks

Glycerol stocks were prepared by combining a sample of an overnight preculture with 20 % (v/v) glycerol (sterile) in a sterile 1.5 mL microcentrifuge tube. The stocks were snap frozen in liquid nitrogen prior to storage at -80 °C.

6.1.16 *E. coli* Culture Growth

For the preparation of an overnight preculture 50 mL LB media, supplemented with the appropriate antibiotics, was inoculated with a scraping from a glycerol stock and incubated at 37 °C with 180 rpm

shaking overnight. The 50 mL preculture was used to inoculate 1 L LB media supplemented with antibiotics. Cultures were grown in baffled flasks at 37 °C with 180 rpm shaking until mid-logarithmic growth phase (OD 0.4-0.8). The pDEST, pET, and pRSFDuet plasmids are under control of the *lac* promotor and are therefore induced by lactose. To induce protein expression IPTG, a synthetic analogue of lactose that is not metabolised by the cell, was added to a final concentration of 0.5. Note that 0.1 mM IPTG was used for *Eco*HisG$_L$ cultures as induction with higher IPTG concentrations resulted in insoluble protein. Following induction cultures, with the exception of *Eco*HisG$_L$, were incubated at 23 °C with 180 rpm shaking overnight. *Eco*HisG$_L$ cultures were incubated at 37 °C following induction and harvested after four hours. Cells were harvested by centrifugation at 12,000 g for 15 minutes and the pellets were stored at -80 °C.

6.1.17 Cell Lysis

Cells were lysed by sonication using an Omni-Ruptor 4000 Ultrasonic Homogenizer. Cell pellets were resuspended in 20-40 mL of lysis buffer and sonicated on ice with 4-6 repetitions of 5 minutes at 80% power and 40 % pulsar. Cell debris was removed by centrifugation at 24,000 *g* or 40,000 *g* and 4 °C for 30 minutes. The supernatant was incubated with 2 µL Benzonase® Nuclease (Sigma Aldrich) for 5 minutes before the first chromatography step.

6.1.18 Fast Protein Liquid Chromatography

An ÄKTApurifier™ (GE Healthcare) or a Bio-Rad Biologic Protein Chromatography system at 4 °C was employed to purify the proteins used in this project. All buffers and solvents were filtered

using a 0.2 μm filter (Millipore). Protein samples were also filtered prior to loading onto a 10 mL or 50 mL Superloop™ (GE Healthcare). Proteins with an N-terminal $(His)_6$-tag were purified by running a 5 mL HisTrap or a 5 mL HiTrap Talon Column (GE Healthcare) with a gradient of 20 mM to 500 mM imidazole (HisTrap) or 5 mM to 250 mM imidazole (HiTrap Talon). 2 mL fractions were collected in a 96 well plate. Fractions that displayed absorbance at 280 nm were analysed by SDS-PAGE. Fractions corresponding to the correct protein were buffer exchanged and loaded onto a Superdex S200 26/60 (GE Healthcare) column for size exclusion chromatography.

6.1.19 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed using a Bolt® 10 % Bis-Tris Plus precast protein gel (Invitrogen) run in a Mini Gel Tank (Invitrogen). Samples were boiled before application to the gel. Novex® Sharp prestained protein standards (Invitrogen) were used as a molecular weight marker. Gels were run in either MES ($HisZG_S$ or $HisG_L$) or MOPS buffer ($HisZG_{Fusion}$). Gels were run at 165 V for either 35 minutes (MES) or 45 minutes (MOPS) and stained with a solution of Coomassie Brilliant Blue R250 in 50 % (v/v) methanol, 10 % (v/v) acetic acid for approximately 15 minutes. Gels were then destained by the addition of 40 % (v/v) methanol and 10 % (v/v) acetic acid for 30 minutes.

6.1.20 Concentration and Buffer Exchange of Protein

Proteins were concentrated before and after SEC using a Vivaspin™ 20 mL, 2 mL or Vivaspin™ Turbo 15 mL 10 kDa molecular

weight cut off centrifugal concentrator (Sartorius Stedim Biotech). All concentrators were washed with 2-3 volumes of MilliQ before use. Proteins were buffer exchanged by alternating concentration followed by dilution with the desired buffer.

6.1.21 Determination of Protein Concentration

To determine the concentration of a purified protein a 2 µL protein sample was applied to the Nanodrop-1000 spectrophotometer and the absorbance at 280 nm was detected. The extinction coefficient was determined from the amino acid sequence using the ExPASy ProtParam tool (http://web.expasy.org/protparam/).

6.1.22 Protein Storage

Following purification, proteins were separated into aliquots of 50 µL, 100 µL, or 200 µL, snap frozen in liquid nitrogen and stored at -80 °C.

6.1.23 Enzyme Kinetics Assay

ATP-PRTase activity was monitored using a standard enzyme assay.[24] The formation of PR-ATP was monitored at 290 nm (ε= 3600 M^{-1} cm^{-1}) at 25 °C using a Varian Cary 100 UV Visible Spectrophotometer. Measurements were made in triplicate using a quartz cuvette with a 1 cm pathlength. Stock substrate solutions contained 250 mM ATP + 250 mM $MgCl_2$ and 10 mM PRPP + 10 mM $MgCl_2$. The standard assay buffer consisted of 100 mM Tris-HCl pH 8.5 with 150 mM NaCl, 10 mM $MgCl_2$, and 200 µM tris (2-carboxyethyl) phosphine (TCEP) and was made with Chelex-treated MilliQ. An inorganic pyrophosphatase (PPase) from *E. coli* was added in excess

(0.026 mg/mL) to ensure the reaction went to completion. Assay buffer, ATP, ATP-PRTase and PPase were incubated at 25 °C before the reaction was initiated by the addition of PRPP. Initial rates were determined by least squares fitting across the initial rate data. K_M and k_{cat} values were determined by fitting initial rates to the Michaelis-Menten equation using the Grafit software (Erithacus).

6.1.24 Determination of Substrate Concentration

The enzyme kinetics assay was used to determine the concentration of PRPP and ATP stock solutions. For example, the concentration of PRPP was determined, in duplicate, by incubating an assay mixture with a saturating concentration of ATP at 25 °C and the reaction was initiated with PRPP of unknown concentration. The reaction was allowed to go to completion and the ΔA_{290nm} was determined. The Beer-Lambert Law was then applied with $\varepsilon = 3600\ M^{-1}\ cm^{-1}$ and path length 1 cm to determine the concentration of PR-ATP formed (and therefore the concentration of PRPP, the limiting reagent, in the cuvette). The concentration of the PRPP stock was then determined.

6.2 Chapter Two Methods

6.2.1 PCR Amplification and Cloning of *LlahisGS* and *LlahisZ*

LlahisGS was amplified from the *LlahisGS*pBSK vector purchased from Epoch Life Science. Primers were designed to incorporate the 15 base pair overlap for cloning into pET-21a digested with NdeI and XhoI using the InFusion enzyme. A three-step thermocycling protocol was employed with a 60 °C annealing

temperature and an extension time of 30 seconds. The pET-21a vector was digested with NdeI-HF and XhoI-HF (New England Biolabs).

Primers:

*LlahisG*_NdeI_Fwd aaggagatatacatatgatcaagatcgctattactaaag

*LlahisG*_XhoI_Rev ggtggtggtgctcgagtcatcagttggtctgagc

LlahisZ was amplified from the *LlahisZ*pBSK vector purchased from Epoch Life Science. Primers were designed to incorporate the 15 base pair overlap for cloning into pET-28a digested with NdeI and EcoRI using the InFusion system. A three-step thermocycling protocol was employed with a 65 °C annealing temperature and an extension time of 60 seconds. The pET-28a vector was digested with NdeI-HF and EcoRI-HF (New England Biolabs).

Primers:

*LlahisZ*_NdeI_Fwd cgcgcggcagccatatggagaatctgtacttccagg

*LlahisZ*_EcoRI_Rev gacggagctcgaattctcatcaatcattatcttcttcc

LlahisZ was also cloned into multiple cloning site I of the pRSFDuet-1 vector for use in the *E. coli* BW25113Δ*hisG* complementation studies. The *LlahisZ*pBSK vector was used as a template for amplification with InFusion primers using a three-step thermocycling protocol with 60 °C annealing and a 30 second extension time. The pRSFDuet-1 vector was digested with BamHI-HF and HindIII-HF (New England Biolabs).

Primers used to subclone *LlahisZ* into pRSFDuet-1:

LlaHisZ_BamHI_F accacagccaggatccagagaatctgtacttccag

LlaHisZ_HindIII_R atgcggccgcaagctttcatcaatcattatcttcttcc

LlahisGS was also amplified from the *LlahisGS*pBSK vector using primers designed to partially install the TEV protease cleavage sequence at the 5' end of the PCR product (reaction 1). The purified PCR product was used as a template for installation of the remainder of the TEV cleavage site as well as the recognition sequence for the InFusion reaction. A three-step thermocycling protocol with 60 °C annealing was used for both amplifications. The pET-28a vector was digested with EcoRI-HF and NdeI-HF (New England Biolabs).

Primers used to amplify *LlahisGS* for cloning into pET-28a:

Reaction 1

LlaHisG_pET-28_rd1_F ctgtacttccagggaatcaagatcgctattactaaag
LlaHisG_pET-28_rd1_R gaattctcatcagttggtctgagc

Reaction 2

LlaHisG_pET-28_rd2_F cgcgcggcagccatatggagaatctgtacttccagg
LlaHisG_pET-28_rd2_R gacggagctcgaattctcatcagttggtctgagc

6.2.2 Purification of *Lla*HisZGS

The *Lla*HisZGS complex was purified by combining the cell pellets from 500 mL cultures (1 L cultures were harvested in two 500 mL batches and stored as separate pellets) of *LlahisZ*pET-28a in BL21*(DE3) and *LlahisGS*pET-21a in Chaperone 3 cells. The combined pellets were resuspended in 20 mL lysis buffer (50 mM potassium phosphate, pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$, 200 μM TCEP, 20 mM imidazole) and lysed by sonication. The supernatant was applied to a 5 mL HisTrap IMAC column (GE Healthcare). The $(His)_6$-tagged *Lla*HisZGS eluted as a single peak with an imidazole gradient of 20 mM to 500 mM over ten column volumes. The fractions containing *Lla*HisZGS were pooled and buffer exchanged into SEC buffer (50 mM potassium phosphate, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 200 μM TCEP) and loaded onto the SEC column. The purified protein was concentrated to 5-10 mg/mL and either stored at -80 °C or used for

crystal screens. The same methodology was employed for the purification of *Lla*HisZ, with the omission of *Lla*HisG$_S$. This protocol was also employed to purify (His)$_6$-tagged *Lla*HisG$_S$ in the absence of *Lla*HisZ however the protein precipitated during the first chromatography step.

6.2.3 Determination of the Quaternary Structure of *Lla*HisZG$_S$ and *Lla*HisZ by Analytical SEC

Analytical SEC, using a Superdex 200 10/300 GL column (GE Healthcare), was employed to determine the multimeric states of *Lla*HisZG$_S$ and *Lla*HisZ. Samples (500 µL) and protein standards (Sigma Aldrich and GE Healthcare) were eluted in *Lla*HisZG$_S$ SEC buffer.

6.2.4 SAXS Data Collection

SAXS data was collected at the Australian Synchrotron SAXS/WAXS beamline equipped with a Pilatus detector. The wavelength of the X-rays was 1.0332 Å and the sample detector distance was 1600 mm. Scattering data was collected from *Lla*HisZG$_S$ (approximately 1 mg/mL) following elution from a Superdex 200 5/150 SEC equilibrated with SEC buffer with or without the addition of 1 mM histidine. 2D intensity plots from the SEC peak were radially averaged, normalised to sample transmission, and the background was subtracted.

6.2.5 SAXS Data Processing

Primary data reduction was performed using the scatterbrain software (Australian Sycnchrotron). Guinear fits were made using

Primus.[68] All plots were checked for increasing intensity at low *s*, which is indicative of an aggregated sample. Indirect Fourier transformation was performed using GNOM to generate the *P(r)* function, which shows the relative probabilities of distances between scattering centres and D_{max}. Crysol[69] was used to generate the theoretical scattering curves from the atomic coordinates (PDB 1Z7N, 1Z7M, and 1USY) compared to the experimental scattering curves. The fits were evaluated by comparing the χ^2 values.

6.2.6 *Ab initio* Modelling

Ab initio shaped determination of the histidine bound *Lla*HisZG$_S$ was performed by running GASBOR.[72] The ATSAS server was used to perform 20 iterations of GASBOR with P1, P2, and P4 symmetry imposed. The *ab initio* models were averaged using the Damaver program suite. Supcomb was employed to superimpose the X-ray crystal structures onto the most probable models. Crysol[69] was used to generate the theoretical scattering curves from the average models compared to the experimental scattering curve. The fits were evaluated by analysing the χ^2 values.

6.2.7 Rigid Body Modelling

Coral was used to construct models of histidine-bound *Lla*HisZG$_S$.[74] Individual PDB files of each of the eight chains in the *Lla*HisZG$_S$ crystal structure (PDB 1Z7N) were created and the *Lla*HisZ and *Lla*HisG$_S$ dimer interfaces were fixed. Models were generated with imposed P1, P2, and P4 symmetry. Crysol[69] was used to generate the theoretical scattering curves from the rigid body models compared to

the experimental scattering curve. The fits were evaluated by analysing the χ^2 values.

6.2.8 Isothermal Titration Calorimetry

Histidine binding to *Lla*HisZ and *Lla*HisZG$_S$ was monitored by ITC. ITC experiments were performed using a VP-ITC microcalorimeter (MicroCal). Contamination of the MilliQ resulted in precipitation of the potassium phosphate buffers normally used in *Lla*HisZG$_S$ experiments. *Lla*HisZ and *Lla*HisZG$_S$ samples were therefore buffer exchanged into 50 mM Tris-HCl pH 7.5 with 150 mM NaCl, 5 mM $MgCl_2$, and 200 µM TCEP. Histidine samples were prepared in the identical buffer. The samples of histidine and protein were degassed under vacuum for 15 minutes prior to use. The sample cell was filled with 1.46 mL of protein sample (50 µM *Lla*HisZG$_S$ or 47 µM *Lla*HisZ) and the syringe was filled with 250 µL of histidine (2 mM for binding to *Lla*HisZG$_S$ and 0.8 mM for binding to *Lla*HisZ). To monitor histidine binding to *Lla*HisZG$_S$ 10 µL of ligand was injected into the sample cell every 600 seconds, and 5 µL injections every 300 seconds were used to monitor histidine binding to *Lla*HisZ. The heat of dilution was determined by injecting the histidine sample into buffer and was subtracted from the experimental data. All experiments were performed at 25 °C. K_A and n values were determined by fitting the data to a one site model in Origin.

6.2.9 Crystallisation Trials

Freshly purified *Lla*HisZG$_S$ was used in sitting drop vapour diffusion crystal trials with a Mosquito® LCP crystallisation robot. A range of commercial screens, including JCSG-*plus*™ *HT-96*, PACT

premier^{TM} *HT-96,* and Clear *Strategy*^{TM} Screens 1 & 2 *HT-96* (Molecular Dimension) were employed with protein concentrations of 5 mg/mL, 8 mg/mL, and 10 mg/mL and 400 nL drop size. In addition crystal trials were performed using hanging drop vapour diffusion in a 24 well format with the *ProPlex* (Molecular Dimensions) screen. Crystals were screened for diffraction at the MX1 beamline at the Australian Synchrotron. The best crystal conditions were optimised by systematically altering the concentrations of precipitant, protein, salt, and the addition of histidine or histidine and AMP. Optimisation of crystal conditions was performed using hanging drop vapour diffusion in a 24 well format with a drop size of 2 µL or 4 µL. The most promising crystals were obtained by hanging drop vapour diffusion with 4 µL drops, 4 mg/mL *Lla*HisZG$_S$ (concentration in the drop) and a well solution of 100 mM Tris-HCl pH 7.3-8, 1.5 mM sodium acetate, and 2-8 % (v/v) PEG 4000, 1 mM histidine. Diffraction of 4.6 Å to 10 Å was measured at the MX1 beamline at the Australian Synchrotron.

6.2.10 Kinetic Characterisation

The standard enzyme kinetics assay, with a protein concentration of 3.2 µM, was employed to determine the K_M^{PRPP} and K_M^{ATP}. The response to histidine inhibition was determined with a PRPP concentration of 50 µM and an ATP concentration of 10 mM.

6.3 Chapter Three Methods

6.3.1 PCR amplification of *EcohisG*$_L$

EcohisG$_L$ was amplified from *E. coli* gDNA using two rounds of nested PCR. The first round amplified the *EcohisG*$_L$ gene from gDNA and incorporated part of the TEV cleavage site at the 5' end. The

purified PCR product was used as a template for the second amplification, which installed the remainder of the TEV recognition site as well as the *attB* sites for cloning into pDONR221 using Gateway® Technology (Invitrogen). Both PCR reactions employed a three-step thermocycling protocol with a 55 °C annealing temperature and a 30 second extension time.

Primers designed to amplify *EcohisG*$_L$

Reaction 1

EcoHisG_F ggcagcggcgcgatgacagacaacactcgt

EcoHisG_R gaaagctgggtgtcactccatcatcttctcaatcgg

Reaction 2

Gateway_rd2_F ggggacaagtttgtacaaaaaagcaggcttcgaaaacctgtattttcagggc agcggcgcg

Gateway_rd2_R ggggaccactttgtacaagaaagctgggt

6.3.2 Gateway® cloning of *EcohisG*$_L$

The purified *EcohisG*$_L$ PCR product was introduced into pDONR221 using the standard procedure for BP clonase™. The BP clonase reaction consisted of 100 ng of PCR product, 150 ng of pDONR221 and Tris-EDTA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and 2 μL BP clonase™ enzyme mix, made up to a total volume of 10 μL. The reaction mixture was incubated at 25 °C for 60 minutes followed by the addition of 1 μg of Proteinase K, which was incubated at 37 °C for ten minutes to terminate the reaction.

EcohisG$_L$ was cloned into pDEST-17, with an N-terminal $(His)_6$-tag using the standard procedure for LR clonase™. The sequence verified *EcohisG*$_L$pDONR221 construct (100 ng) was incubated with 150 ng pDEST-17, and 2 μL LR clonase™ in Tris-EDTA buffer (total

volume 10 μL). The reaction mixture was incubated at 25 °C for 60 minutes followed by the addition of 1 μg of Proteinase K, which was incubated at 37 °C for ten minutes to terminate the reaction.

6.3.3 Purification of *Eco*HisG$_L$

The cell pellet from a 1 L culture of *Eco*HisG$_L$ in *E. coli* Chaperone 3 cells was resuspended in 20 mL lysis buffer (50 mM potassium phosphate pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 5 mM imidazole) and lysed by sonication. The supernatant was applied to a 5 mL HiTrap Talon Crude IMAC column (GE Healthcare) and eluted with an imidazole gradient of 5 mM to 250 mM. The fractions containing *Eco*HisG$_L$ were pooled and buffer exchanged into SEC buffer (20 mM Tris-HCl pH 8.5, 50 mM NaCl, 5 mM $MgCl_2$). The sample was diluted to approximately 1 mg/mL in SEC buffer and the N-terminal $(His)_6$-tag was cleaved by the addition of TEV protease in a 10 : 1 *Eco*HisG$_L$: TEV (w/w) ratio. The TEV cleavage mixture was incubated at 37 °C for one hour followed by 4 °C for three days. SEC was used to purify the untagged *Eco*HisG$_L$. The purified *Eco*HisG$_L$ was concentrated to 5-10 mg/mL and stored at -80 °C. To test for ATP-PRTase activity 1-20 μL samples 5 mg/mL *Eco*HisG$_L$ were introduced into the standard enzyme kinetic assay with 10 mM ATP and 1 mM PRPP.

6.3.4 Determination of the Quaternary Structure of *Eco*HisG$_L$ by Analytical SEC

Analytical SEC, using a Superdex 200 10/300 GL column (GE Healthcare), was employed to determine the multimeric state of

*Eco*HisG$_L$. Samples (1 mL) and protein standards (Sigma Aldrich and GE Healthcare) were eluted in *Eco*HisG$_L$ SEC buffer.

6.3.5 SAXS Data Collection

SAXS data was collected at the Australian Synchrotron SAXS/WAXS beamline equipped with a Pilatus detector. The wavelength of the X-rays was 1.0332 Å and the sample detector distance was 1600 mm. Static solution sample detection was performed using a 96 well plate. Samples were prepared at concentrations of 0.3 mg/mL, 0.75 mg/mL, 1.5 mg/mL, and 3 mg/mL in SEC buffer (20 mM Tris-HCl pH 8.5, 50 mM NaCl, 5 mM $MgCl_2$, 500 μM TCEP) with or without the addition of histidine, AMP, PRPP, or ATP (1 mM).

6.3.6 SAXS Data Processing

Primary data reduction was performed using the scatterbrain software (Australian Sycnchrotron). Guinear fits were generated using Primus.[68] All plots were checked for increasing intensity at low *s*, which is indicative of an aggregated sample. Indirect Fourier transformation was performed using GNOM to calculate the *P(r)* function, which shows the relative probabilities of distances between scattering centres and D_{max}. Crysol[69] was used to generate the theoretical scattering curves from the atomic coordinates (PDB 1H3D and 1Q1K) compared to the experimental scattering curves. The fits were evaluated by comparing the χ^2 values.

6.3.7 Cloning and Expression of *Eco*HisG$_{core}$

EcohisG$_{core}$ was amplified from the *EcohisG$_L$*pDONR221 construct using two rounds of nested PCR. The first round amplified

the gene and part of the TEV cleavage site and the second round installed the *attB* sites for cloning using Gateway® technology. The purified PCR product was cloned into pDONR221 using the standard procedure for BP clonase™. Following sequence verification the *EcohisG*$_{core}$ was cloned into pDEST-17 using the standard procedure for LR clonase™.

Primers used to amplify *EcohisG*$_{core}$

Reaction 1

EcoHisG_F ggcagcggcgcgatgacagacaacactcgt

EcoHisG_core_Rd1_Rev gaaagctgggtgttattcgcgcgcctggatcacac

Reaction 2

Gateway_rd2_F ggggacaagtttgtacaaaaaagcaggcttcgaaaacctgtattttcagggc agcggcgcg

Gateway_rd2_R ggggaccactttgtacaagaaagctgggt

Expression trials of both *Eco*HisG$_{core}$ were performed in *E. coli* BL21*(DE3) and Chaperone 3 cells grown in 500 mL LB cultures at 37 °C and 23 °C. The cell pellets were resuspended in 20 mL of lysis buffer (50 mM potassium phosphate pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$) and lysed by sonication. To test for ATP-PRTase activity 5-50 µL samples of the soluble fractions were introduced into the standard enzyme kinetic assay with 10 mM ATP and 1 mM PRPP.

6.3.8 Cloning of *NmehisZ* and *NmehisG*$_S$

NmehisG$_S$ was amplified from the *N. meningitidis* gDNA using two rounds of nested PCR. The first round amplified the gene and part of the TEV cleavage site and the second round installed the *attB* sites for cloning using Gateway® technology. PCR amplification was achieved using a three-step thermocycling protocol with an annealing

temperature of 45 °C for round one and 50 °C for round 2. The purified PCR product was cloned into pDONR221 using the standard procedure for BP clonase™. Following sequence verification, *NmehisGS* was cloned into pDEST-17 and pDEST-15 using the standard procedure for LR clonase™.

Primers used to amplify *NmehisGS* for Gateway® cloning:

Reaction 1

NmeHisG_rd1_F ggaagtggagctatgcaggataatgctttg

NmeHisG_rd1_R gaaagctgggtgttacttcgccttcac

Reaction 2

Gateway_rd2_F ggggacaagtttgtacaaaaaagcaggcttcgaaaacctgtattttcagggc agcggcgcg

Gateway_rd2_R ggggaccactttgtacaagaaagctgggt

The Gateway® system was also used to clone *NmehisGS* into pDEST-14 with no tag. The pDEST-14 vector does not contain a ribosome binding site. The *NmehisGS* gene was amplified from the *NmehisGS*pDONR221 construct via two rounds of nested PCR. The first round incorporated the Shine-Delgarno sequence and the second round installed the *attB* sites for cloning using Gateway® technology. A three step thermocycling protocol with a 55 °C annealing temperature was employed for both reactions. The *NmehisGS* PCR product was introduced into pDONR221 using the standard procedure for BP clonase™. The sequence verified *NmehisGS*pDONR221 construct was used to clone *NmehisGS* into pDEST-14 using the standard procedure for LR clonase™.

Primers used to clone *NmehisGS* into pDEST-14:

Reaction 1

NmeHisG_pDEST14_rd1_F gcttcgaaggagatagaaccatgcaggataatgctttg

NmeHisG_rd1_R gaaagctgggtgttacttcgccttcac

Reaction 2

Gateway_pDEST14_rd2_F ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaccatg

Gateway_pDEST14_rd2_R ggggaccactttgtacaagaaag

NmehisGS was also subcloned into pET-21a using the InFusion® HD cloning kit. *NmehisGS* was amplified from the *NmehisGS*pDONR221 construct using primers designed to incorporate the overlapping regions for the InFusion® reaction. The pET-21a vector was digested with NdeI-HF and XhoI-HF (New England Biolabs).

Primers designed to clone *NmehisGS* into pET-21a

NmeHisG_pET21_F aaggagatatacatatgcaggataatgctttgaccatcgcct
NmeHisG_pET21_R ggtggtggtgctcgaggacgtctcattacttcgccttcactgcgccgc

NmehisZ was cloned into multiple cloning site I of pRSFDuet-1 with an N-terminal $(His)_6$-tag. *NmehisZ* was amplified from the *NmehisZ*pBSK vector purchased from Epoch Life Science using primers designed for cloning using the InFusion® enzyme. A two step thermocycling protocol with a 60 second extension time was employed. The pRSFDuet-1 vector was digested with BamHI-HF and HindIII-HF. Following sequence verification; *NmehisGS* was cloned into multiple cloning site II of the *NmehisZ*pRSFDuet-1 construct. *NmehisGS* was amplified from the *NmehisGS*pDONR221 construct using primers designed to incorporate the complementary region for the InFusion® reaction. A two step thermocycling protocol was employed with a 30 second extension time. The *NmehisZ*pRSFDuet-1 vector was digested with NdeI-HF and AatII-HF (New England Biolabs).

Primers used to clone *NmehisZ* and *NmehisGS* into pRSFDuet-1

NmehisZ

NmeHisZ_MCSI_F accacagccaggatccagagaatctgtacttccag

NmeHisZ_MCSI_R atgcggccgcaagctttcatcagcgtttcacgacc

NmehisG$_S$

NmehisG_MCSII_F aaggagatatacatatgcaggataatgctttgaccatcgcct

NmehisG_MCSII_R gactcgagggtaccgacgtctcattacttcgccttcactgcgccgc

6.3.9 Purification of *Nme*HisZG$_S$

Numerous purification protocols were employed, which are discussed in chapter three. This section reports the most successful purification protocol. The pellets from 500 mL cultures of *NmehisZ*pRSFDuet-1 in BL21*(DE3) cells and *NmehisGS*pET-21a in Chaperone 3 cells were combined and resuspended in 20 mL lysis buffer (50 mM potassium phosphate pH 7.0, 300 mM NaCl, 5 mM $MgCl_2$, 500 µM TCEP, 20 µM imidazole) and lysed by sonication. The supernatant was applied to a 5 mL HisTrap (GE Healthcare) and eluted with an imidazole gradient of 20 mM to 500 mM imidazole. The fractions containing *Nme*HisZG$_S$ were pooled, buffer exchanged, and loaded onto the size exclusion column pre equilibrated with SEC buffer (50 mM potassium phosphate pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 500 µM TCEP). The fractions corresponding to *Nme*HisZG$_S$ and *Nme*HisG$_S$ were pooled separately and stored at -80 °C.

*Nme*HisZ was purified serendipitously while attempting to purify the *Nme*HisZG$_S$ complex. The pellet from a 1 L culture of *NmehisZGS*pRSFDuet-1 was resuspended in 20 mL lysis buffer (20 mM Tris-HCl pH 8.5, 100 mM NaCl, 5 mM $MgCl_2$, 5 mM imidazole) and lysed by sonication. The supernatant was applied to a 5 mL HiTrap Talon Crude (GE Healthcare) IMAC column. The $(His)_6$-tagged protein was eluted over an imidazole gradient of 5 mM to 250 mM. The fractions containing *Nme*HisZG$_S$ were pooled, buffer exchanged, and

loaded onto the size exclusion column pre equilibrated with SEC buffer (20 mM Tris-HCl pH 8.5, 100 mM NaCl, 5 mM $MgCl_2$, 5 mM imidazole). The fractions corresponding to *Nme*HisZ were pooled, concentrated to 10 mg/mL and stored at -80 °C.

6.3.10 Determination of the Quaternary Structure of *Nme*HisZG$_S$, *Nme*HisZ, and *Nme*HisG$_S$ by Analytical SEC

Analytical SEC, using a Superdex 200 10/300 GL column (GE Healthcare), was employed to determine the multimeric state of *Nme*HisZG$_S$, *Nme*HisZ and *Nme*HisG$_S$. Samples (1 mL) and protein standards (Sigma Aldrich and GE Healthcare) were eluted in *Nme*HisZG$_S$ SEC buffer.

6.3.11 Crystallisation Trials

Freshly purified *Nme*HisZ was used in sitting drop vapour diffusion crystal trials with a Mosquito® LCP crystallisation robot. A range of commercial screens, including JCSG-*plus™ HT-96*, PACT *premier™ HT-96,* and Clear *Strategy™* Screens 1 & 2 *HT-96* (Molecular Dimension) were employed with protein concentrations of 5 mg/mL and 10 mg/mL and 400 nL drop size. In addition crystal trials were performed using hanging drop vapour diffusion in a 24 well format with the *ProPlex* (Molecular Dimensions) screen. Crystals were screened for diffraction at the MX1 beamline at the Australian Synchrotron. No diffraction quality crystals were obtained.

6.3.12 Preliminary Kinetic Characterisation

The standard enzyme kinetics assay was employed to determine the K_M^{PRPP} for *Nme*HisZG$_S$ (6 μM). The ATP concentration was held constant at 10 mM and the PRPP concentration was varied from 1 μM to 50 μM. The decision was made not to continue kinetic characterisation as the estimated k_{cat} was three fold lower than that of *Lla*HisZG$_S$ and analytical SEC revealed aggregation of *Nme*HisZG$_S$.

6.4 Chapter Four Methods

6.4.1 PCR Amplification of *LmehisZG$_{Fusion}$*

LmehisZG$_{Fusion}$ was amplified from the *LmehisZG$_{Fusion}$*pBSK vector purchased from Epoch Life Science. Primers were designed to clone *LmehisZG$_{Fusion}$* into pET-28a, digested with NdeI and EcoRI, using the InFusion enzyme. A two-step thermocycling protocol was employed with an extension time of 90 seconds.

Primers:

LmeHisZG_Inf_pET28_F cgcgcggcagccatatggagaatctgtacttccaag
LmeHisZG_Inf_pET28_R gacggagctcgaattctcattactgctcctcttctt

The pET-28a vector was digested with NdeI-HF and EcoRI-HF (New England Biolabs). Both the linearised vector and PCR products were gel purified prior to incubation with the InFusion enzyme.

6.4.2 Purification of *Lme*HisZG$_{Fusion}$

The pellet from a 1 L culture of *Lme*HisZG$_{Fusion}$ in *E. coli* BL21*(DE3) cells was resuspended in 20 mL lysis buffer (50 mM potassium phosphate pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 200 μM

TCEP, 20 mM imidazole) and lysed by sonication. The supernatant was applied to a 5 mL HisTrap column (GE Healthcare) and the $(His)_6$-tagged *Lme*HisZG$_{Fusion}$ was eluted by running an imidazole gradient from 20 mM to 500 mM over ten column volumes. The fractions corresponding to the A_{280nm} peak were analysed by SDS-PAGE. The fractions containing *Lme*HisZG$_{Fusion}$ were pooled, buffer exchanged, and loaded onto the Superdex S200 26/60 column (GE Healthcare) pre equilibrated with SEC buffer (50 mM potassium phosphate pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 200 µM TCEP). The purified protein was concentrated to 5 mg/mL and stored at -80 °C.

6.4.3 Construction of *Lla*HisZG$_{Fusion}$ by PCR

The amino acid sequence of the *Lla*HisZG$_{Fusion}$(long) construct is shown below. The TEV cleavage site is highlighted in cyan, the linker between the TEV site and *Lla*HisZ is underlined, and the linker between *Lla*HisZ and *Lla*HisG$_S$ is highlighted in red.

MENLYFQGSSGAEKINYLLPEESAEMTLNQVKSLRQIEGRLRKLFSLKNYQEVMP
PSFEYTQLYTALESNGKTFNQEKMFQFINHEGQSITLRYDFTLPLVRLYSQIKDS
TSARYSYFGKIFRKEKRHKGRSTENYQIGIELFGESADKSELEILSLALQVIEQL
GLNKTVFEIGSAKFFQRLCHLADGSTELLTELLLKKDLSGLNAFIEKNNFSKELR
ELLKEIFITNELSRLENLVTNTKDDVLISSFDQLKEFSEKLSMIKPIIIDLGMVP
KMDYYTDLMFKAYSSAANQPILSGGRYDQLLSNFQEEAVAIGFCCHMDTILADWR
RKPEEMKPQEPIVMVLGKGRIQKQVTKLLENADYDVEPILNLGRELQIKTKDDLQ
IIFGKANDVITFLEHGIVDIGFVGKDTLDENDFDDYYELLDLKIGQCIFALASYP
DFSNKNFQRRKRIASKYPRVTKKYFAQKQEDIEIIKLEGSVELGPVVGLADAIVD
IVETGNTLSANGLEVIEKISDISTRMIVNKSSFKFKKDKIIEMVERLEDAQTN

To generate the *LlahisZG$_{Fusion}$(long)* PCR product, *LlahisZ* was amplified from the *LlahisZ*pBSK vector with primers designed to incorporate 21 base pairs of the 63 base pair linker at the 3' end of the PCR product (reaction 1). The purified product from reaction 1 was used as a template for the second PCR reaction that installed a further

33 base pairs of the linker (reaction 2). The same forward primer was used in both PCR reactions

Primers used to amplify *LlahisZ*:

Reaction 1

LlaHisZG_Fusion_HisZ_f catatggagaatctgtacttccagg

LlaHisZG_FusionL_1_R cggtttacgacgccaatccgccagaatggtgtccatgtggcaaca

Reaction 2

LlaHisZG_Fusion_HisZ_f catatggagaatctgtacttccagg

LlaHisZG_FusionL_rxn2_R ggttcctgcggtttcatttcttccggtttacgacgccaatcc gcc

LlahisGs was amplified from the *LlahisGs*pBSK vector using primers designed to incorporate 21 base pairs of the 63 base pair linker at the 5' end of the PCR product (reaction 3). The purified PCR product was used as a template for an additional PCR reaction that installed a further 21 base pairs of the linker region (reaction 4). The same reverse primer was used in both amplifications.

Primers used to amplify *LlahisGs*:

Reaction 3

LlaHisZG_FusionL_rxn3_F ccgattgtgatggtgctgggcaaaggccgtatccagaaacag gtg

LlaHisZG_Fusion_HisG_R gaattctcatcagttggtctgagc

Reaction 4

LlaHisZG_FusionL_rxn4_F gaagaaatgaaaccgcaggaaccgattgtgatggtgctgggc aaa

LlaHisZG_Fusion_HisG_R gaattctcatcagttggtctgagc

Each of the PCR reactions used to amplify *LlahisZ* employed a two-step thermocycling protocol with a 60 second extension. PCR amplification of *LlahisGs* was achieved using a two-step thermocycling protocol with an extension time of 30 seconds.

The *LlahisZ* and *LlahisGs* PCR products were stitched together by the addition of both purified PCR products, in an equimolar ratio, to

a standard PCR reaction, without primers. A two-step thermocycling protocol with a 90 second extension was run for 10 cycles prior to the addition of primers. The reaction proceeded for a further 30 cycles following the addition of the *LlahisZ* forward primer and *LlahisG*$_S$ reverse primer (reaction 5). The purified *LlahisZG*$_{Fusion}$*(long)* PCR product was used as a template for the final PCR step that installed the recognition sequence for cloning into pET-28a using the InFusion enzyme (reaction 6).

Primers for amplification of *LlahisZG*$_{Fusion}$*(long)*

Reaction 5

LlaHisZG_Fusion_HisZ_F catatggagaatctgtacttccagg

LlaHisZG_Fusion_HisG_R gaattctcatcagttggtctgagc

Reaction 6

LlaHisZG_Fusion_Inf_F cgcgcggcagccatatggagaatctgtacttccagg
LlaHisZG_Fusion_Inf_R gacggagctcgaattctcatcagttggtctgagcgtcttc

The pET-28a vector was linearised with NdeI-HF and EcoRI-HF (New England Biolabs). The *LlahisZG*$_{Fusion}$*(long)* PCR product was introduced into pET-28a using the standard protocol for the InFusion system.

The amino acid sequence of *Lla*HisZG$_{Fusion}$(short), coloured in accordance with the sequence of *Lla*HisZG$_{Fusion}$(long) is shown below.

MENLYFQGSSGAEKINYLLPEESAEMTLNQVKSLRQIEGRLRKLFSLKNYQEVMP
PSFEYTQLYTALESNGKTFNQEKMFQFINHEGQSITLRYDFTLPLVRLYSQIKDS
TSARYSYFGKIFRKEKRHKGRSTENYQIGIELFGESADKSELEILSLALQVIEQL
GLNKTVFEIGSAKFFQRLCHLADGSTELLTELLLKKDLSGLNAFIEKNNFSKELR
ELLKEIFITNELSRLENLVTNTKDDVLISSFDQLKEFSEKLSMIKPIIIDLGMVP
KMDYYTDLMFKAYSSAANQPILSGGRYDQLLSNFQEEAVAIGFCCHMDTILKALE
RQELEEDPQEPIVMVLGKGRIQKQVTKLLENADYDVEPILNLGRELQIKTKDDLQ
IIFGKANDVITFLEHGIVDIGFVGKDTLDENDFDDYYELLDLKIGQCIFALASYP
DFSNKNFQRRKRIASKYPRVTKKYFAQKQEDIEIIKLEGSVELGPVVGLADAIVD
IVETGNTLSANGLEVIEKISDISTRMIVNKSSFKFKKDKIIEMVERLEDAQTN

The *LlahisZ*pBSK construct was used as a template for the amplification of *LlahisZ*. Primers were designed to incorporate 17 bases of the 30 base pair linker at the 3' end of the PCR product (reaction 1). *LlahisG$_S$* was amplified from the *LlahisG$_S$*pBSK vector using primers designed to install 23 base pairs of the linker region at the 5' end of the PCR product (reaction 2). The purified product from reaction 2 was used as a template for the third PCR reaction that introduced a further 17 base pairs of the linker at the 5' end of the PCR product.

Reaction 1

LlaHisZG_Fusion_HisZ_f catatggagaatctgtacttccagg

LlaHisZG_FusionS_rxn1_R acaatcggttcctgcggatcttcttccagctcctggcgtt

Reaction 2

LlaHisZG_FusionS_rxn2_F aaccgattgtgatggtgctgggcaaaggccgtatccagaaac
aggtg

LlaHisZG_Fusion_HisG_R gaattctcatcagttggtctgagc

Reaction 3

LlaHisZG_FusionS_rxn3_F ccgcaggaaccgattgtgatggtgctgggcaa

LlaHisZG_Fusion_HisG_R gaattctcatcagttggtctgagc

The purified *LlahisZ* and *LlahisG$_S$* PCR products were combined, in an equimolar ratio, to amplify the fusion construct, as described for the formation of the *LlahisZG$_{Fusion}$(long)* PCR product. The *LlahisZG$_{Fusion}$(short)* PCR product was used as a template in the final PCR reaction, which installed the overlapping sequences for cloning into pET-28a using the InFusion enzyme.

Reaction 4

LlaHisZG_Fusion_HisZ_F catatggagaatctgtacttccagg

LlaHisZG_Fusion_HisG_R gaattctcatcagttggtctgagc

Reaction 5

LlaHisZG_Fusion_Inf_F cgcgcggcagccatatggagaatctgtacttccagg

LlaHisZG_Fusion_Inf_R gacggagctcgaattctcatcagttggtctgagcgtcttc

The pET-28a vector was linearised with NdeI-HF and EcoRI-HF (New England Biolabs). The *LlahisZG*$_{Fusion}$*(short)* PCR product was introduced into pET-28a using the standard protocol for the InFusion enzyme.

6.4.4 Expression of *Lla*HisZG$_{Fusion}$

Expression trials of both *Lla*HisZG$_{Fusion}$ constructs were performed in *E. coli* BL21*(DE3) and Chaperone 3 cells grown in 500 mL LB cultures at 37 °C and 23 °C. The cell pellets were resuspended in 20 mL of lysis buffer (50 mM potassium phosphate pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 200 μM TCEP, 20 mM imidazole) and lysed by sonication. A 600 μL sample of each supernatant was applied to a His SpinTrap (GE Healthcare) and the $(His)_6$-tagged protein was eluted with 500 mM imidazole. SDS-PAGE was used to analyse the fractions from the His SpinTrap in addition to the soluble and insoluble fractions from the lysis trial. To test for ATP-PRTase activity 5-50 μL samples of the soluble fractions and the His SpinTrap eluents were introduced into the standard enzyme kinetic assay with 10 mM ATP and 1 mM PRPP.

6.4.5 PCR Amplification of *LlahisGs-CjeACT*

The amino acid sequence of the *Lla*HisG$_S$-*Cje*ACT chimera is shown below. The TEV protease cleavage site is highlighted in cyan, the linker between the TEV site and *Lla*HisG$_S$ is underlined, the *Lla*HisG$_S$ sequence is highlighted in green and the *Cje*ACT domain sequence is highlighted in yellow.

MENLYFQGSSGAIKIAITKGRIQKQVTKLLENADYDVEPILNLGRELQIKTKDDL
QIIFGKANDVITFLEHGIVDIGFVGKDTLDENDFDDYYELLDLKIGQCIFALASY
PDFSNKNFQRRKRIASKYPRVTKKYFAQKQEDIEIIKLEGSVELGPVVGLADAIV
DIVETGNTLSANGLEVIEKISDISTRMIVNKSSFKFKKDKIIEMVERVAGVMQAR
ESKYIMLHAPKEKLDKIQALLPGVERPTILPLAHDEKNVALHMVSKENLFWETME
ALKEEGASSILVLPIEKMLK

The *LlahisGs* gene was amplified from the *LlahisGs*pBSK vector using primers designed to install part of the TEV protease cleavage site at the 5' end of the PCR product and a 14 base pair sequence complementary to the *CjeACT* sequence at the 3' end (reaction 1). The purified PCR product was used as a template for the second PCR reaction to install the remainder of the TEV cleavage site (reaction 2).

Primers designed to amplify *LlahisGs*:

Reaction 1

LlaCje_Chimera_rxn1_F gatccagcggcgcgatcaagatcgctattactaaaggccg

LlaCje_Chimera_Lla_R attactcctgctacacgttcaaccatttcgataatcttgt

Reaction 2

LlaCje_Chimera_rxn2_F gagaatctgtacttccagggatccagcggcgcgatcaagat

LlaCje_Chimera_Lla_R attactcctgctacacgttcaaccatttcgataatcttgt

The *CjeACT* sequence was amplified from the *CjehisGL*pDONR221 sequence using primers designed to include a 24 base pair overlap with the *LlahisGs* sequence (reaction 3). A two-step thermocycling protocol with a 30 second extension time was employed for all amplifications.

Primers designed to amplify the *CjeACT* sequence:

Reaction 3

LlaCje_Chimera_ACT_F cgaaatggttgaacgtgtagcaggagtaatgcaagcaaga

LlaCje_Chimera_ACT_R tcatcactttaacatttttttctataggcag

The purified *LlahisGS* and *CjeACT* PCR products were introduced into a PCR reaction without primers. A two-step thermocycling protocol was employed with a 30 second extension. The thermocycling protocol was run for 10 cycles at which point the protocol was paused and primers were added to amplify the *LlahisGS-CjeACT* chimera (reaction 4). The purified product from reaction four was used as a template for the final amplification to install the recognition sequence for cloning into pET-28a using the InFusion enzyme (reaction 5).

Primers designed to amplify *LlahisGS-CjeACT*:

Reaction 4

LlaCje_Chimera_rxn2_F gagaatctgtacttccagggatccagcggcgcgatcaagat

LlaCje_Chimera_ACT_R tcatcactttaacatttttctataggcag

Reaction 5

Lla_Cje_Chimera_Inf_F cgcgcggcagccatatggagaatctgtacttccagggatcc
Lla_Cje_Chimera_Inf_R gacggagctcgaattctcatcactttaacatttttctataggc
ag

The pET-28a vector was digested with NdeI-HF and EcoRI-HF (New England Biolabs). The final purified *LlahisGS* PCR product was cloned into pET-28a using the standard protocol for the InFusion enzyme.

6.4.6 Expression of the *Lla*HisGS-*Cje*ACT Chimera

Expression trials of the *Lla*HisGS-*Cje*ACT chimera were performed in *E. coli* BL21*(DE3) and Chaperone 3 cells, grown in 500 mL LB cultures at 37 °C and 23 °C. The cell pellets were resuspended in 20 mL of lysis buffer (50 mM potassium phosphate pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 200 μM TCEP, 20 mM imidazole) and lysed by sonication. A 600 μL sample of each supernatant was applied to a His SpinTrap (GE Healthcare) and the $(His)_6$-tagged protein was eluted with 500 mM imidazole. SDS-PAGE was used to analyse the fractions

from the His SpinTrap in addition to the soluble and insoluble fractions from the lysis trial. To test for ATP-PRTase activity 5-50 μL samples of the soluble fractions and the His SpinTrap eluents were introduced into the standard enzyme kinetic assay with 10 mM ATP and 1 mM PRPP.

6.4.7 *E. coli* BW25113Δ*hisG* Complementation Studies

E. coli BW25113Δ*hisG* cells from the Keio collection[97] were provided by Dr Wayne Patrick from the University of Otago. This strain is devoid of the *hisG* gene and is auxotrophic for histidine. A kanamycin resistance cassette has replaced the hisG gene. The strain may therefore be grown selectively in the presence of kanamycin and either an exogenous supply of histidine or complementation of a functional ATP-PRTase. The cells were made chemically competent prior to use.

Sterile solutions of agar, M9 salts, $MgSO_4$, $CaCl_2$, glucose, antibiotics, and either IPTG or histidine, were combined to make M9 minimal media plates. 1.5 g of agar was mixed with 80 mL MilliQ and autoclaved. The agar was melted and incubated at 60 °C prior to use. A 5× solution of M9 salts consisting of Na_2HPO_4 (34g/L), KH_2PO_4 (15 g/L), NaCl (2.5 g/L), and NH_4Cl (0.5 g/L) was prepared in MilliQ and autoclaved at 121 °C, 15 psi for 20 minutes. Solutions of $MgSO_4$ (1 M), $CaCl_2$ (0.1 M) and 20 % (w/v) glucose, IPTG (100 mM), and histidine (2 g/L) were filter sterilised prior to use. M9 minimal media plates were made as described in Table 6.1.

Table 6.1 Recipe for M9 Minimal Media Agar Plates

Component	Stock Concentration	Volume Added	Final Concentration
M9 Salts	5×	20 mL	1×
$MgSO_4$	1 M	200 µL	2 mM
$CaCl_2$	0.1 M	100 µL	0.1 mM
Glucose	20 %	2 mL	0.4 %
Ampicillin*	100 g/L	100 µL	0.1 g/L
Kanamycin*	50 g/L	100 µL	0.05 g/L
IPTG**	100 mM	1 mL	0.1 mM
Histidine**	2 g/L	1 mL	0.002 g/L

* Antibiotics were added as appropriate for the strain. The *E. coli* BW25113Δ*hisG* strain is kanamycin resistant therefore all M9 plates included kanamycin. The pET-28, pET-24 and pRSFDuet-1 vectors are kanamycin resistant, therefore cells that were transformed with these plasmids were grown on M9 plates containing only kanamycin. The pDEST-14/15/17 and pET-21 vectors are ampicillin resistant; therefore cells that were transformed with these plasmids were grown on M9 plates containing both kanamycin and ampicillin. **M9 plates were made with either IPTG, to induce protein expression, or histidine as a positive control.

The *E. coli* BW25113Δ*hisG* cells were transformed with plasmids encoding functional and putative ATP-PRTase enzymes as per Table 6.2. An untransformed strain was used as both a positive and negative control. M9 minimal media plates were streaked with 5 µL of transformation mixture and incubated at 37 °C. Plates were inspected each day for a maximum of three days to detect growth. The experiments were performed in triplicate.

Table 6.2 Plasmids used for Complementation Studies in the *E. coli* BW25113Δ*hisG* Strain

ATP-PRTase	Plasmid(s)
*Eco*HisG$_{L}$	*Ecohis*G_{L}pDEST-17
*Cje*HisG$_{L}$	*Cjehis*G_{L}pDEST-17
*Mtu*HisG$_{L}$	*Mtuhis*G_{L}pET-24a
*Eco*HisG$_{core}$	*Ecohis*G_{core}pDEST-17
*Cje*HisG$_{core}$	*Cjehis*G_{core}pDEST-15
*Nme*HisG$_{S}$	*Nmehis*G_{S}pET-21a
*Lla*HisG$_{S}$	*Llahis*G_{S}pET-21a
*Nme*HisZG$_{S}$	*Nmehis*G_{S}pET-21a + *Nmehis*ZpRSFDuet-1
*Lla*HisZG$_{S}$	*Llahis*G_{S}pET-21a + *Llahis*ZpRSFDuet-1
*Nme*HisG$_{S}$+*Lla*HisZ	*Nmehis*G_{S}pET-21a + *Llahis*ZpRSFDuet-1
*Lla*HisG$_{S}$+*Nme*HisZ	*Llahis*G_{S}pET-21a + *Nmehis*ZpRSFDuet-1
*Lme*HisZG$_{Fusion}$	*LmehisZ*G_{Fusion}pET-28a
*Lla*HisZG$_{Fusion}$(long)	*LlahisZ*G_{Fusion}*(long)*pET-28a,
*Lla*HisZG$_{Fusion}$(short)	*LlahisZ*G_{Fusion}*(short)*pET-28a,
*Lla*HisG$_{S}$-*Cje*ACT	*Llahis*G_{S}-*CjeACT*pET-28a

Appendix I

*Lla*HisZG$_S$ SAXS Crysol Fits

Crysol was used to fit the *Lla*HisZG$_S$ SAXS data to the X-ray crystal structures of *Lla*HisZG$_S$ (PDB 1Z7N), shown in Chapter Two, *Lla*HisZG$_S$ (PDB 1Z7M) (Figure I.1) and *Tma*HisZG$_S$ (PDB 1USY) (Figure I.2).

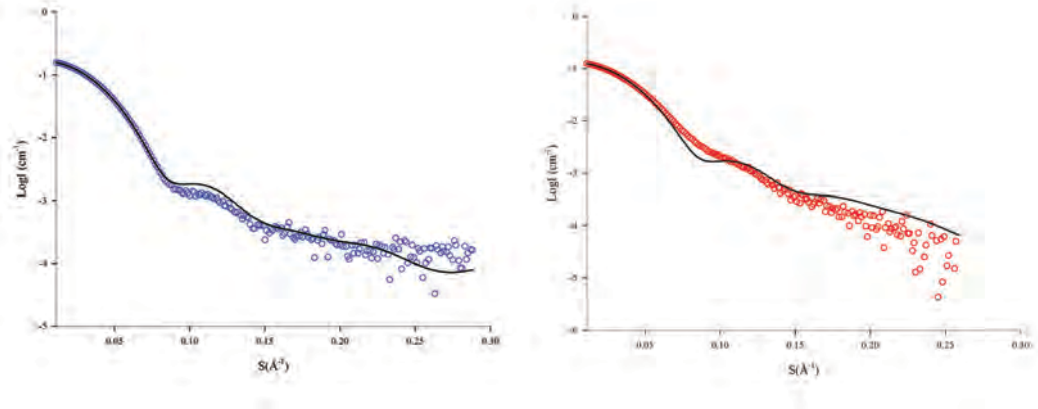

Figure I.1 (Left) Crysol fit of the *Lla*HisZG$_S$ SAXS data (blue) to the theoretical scattering obtained from the X-ray crystal structure of *Lla*HisZG$_S$ (PDB 1Z7M) (χ^2 = 0.87). (Right) Crysol fit of the *Lla*HisZG$_S$ + histidine SAXS data (red) to the theoretical scattering obtained from the X-ray crystal structure of *Lla*HisZG$_S$ (PDB 1Z7M) (χ^2 = 11).

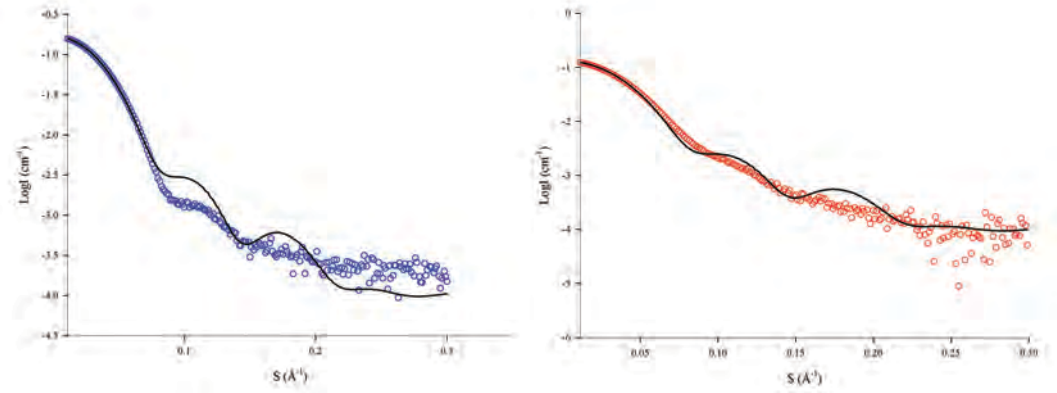

Figure I.2 (Left) Crysol fit of the *Lla*HisZG$_S$ SAXS data (blue) to the theoretical scattering obtained from the X-ray crystal structure of *Tma*HisZG$_S$ (PDB 1USY) (χ^2 = 5.5). (Right) Crysol fit of the *Lla*HisZG$_S$ + histidine SAXS data (red) to the theoretical scattering obtained from the X-ray crystal structure of *Tma*HisZG$_S$ (PDB 1USY) (χ^2 = 6.2).

Appendix II

Complementation Studies with *E. coli* BW25113Δ*hisG*::kan

E. coli BW25113Δ*hisG*::kan cells from the Keio collection were transformed with plasmids encoding both functional and putative ATP-PRTase enzymes, and plated on M9 agar plates supplemented with either histidine or IPTG, as discussed in chapter four. The following photographs were taken following incubation at 37 °C for three days.

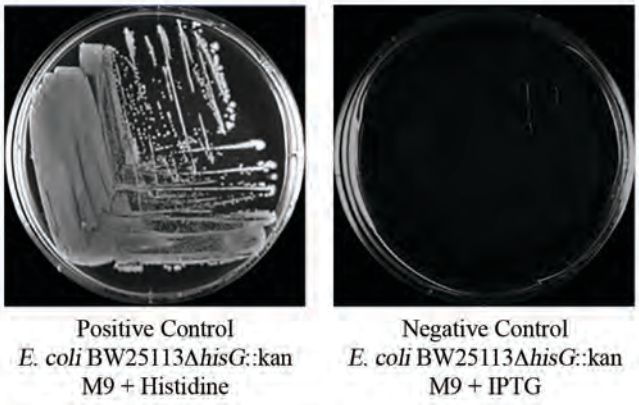

Figure II.1 The untransformed E. coli BW25113Δ*hisG*::kan strain plated on M9 + histidine agar plates and M9 + IPTG plates to form a positive and negative control respectively. Cell growth was observed on the positive control plate and no colonies were observed on the negative control plate.

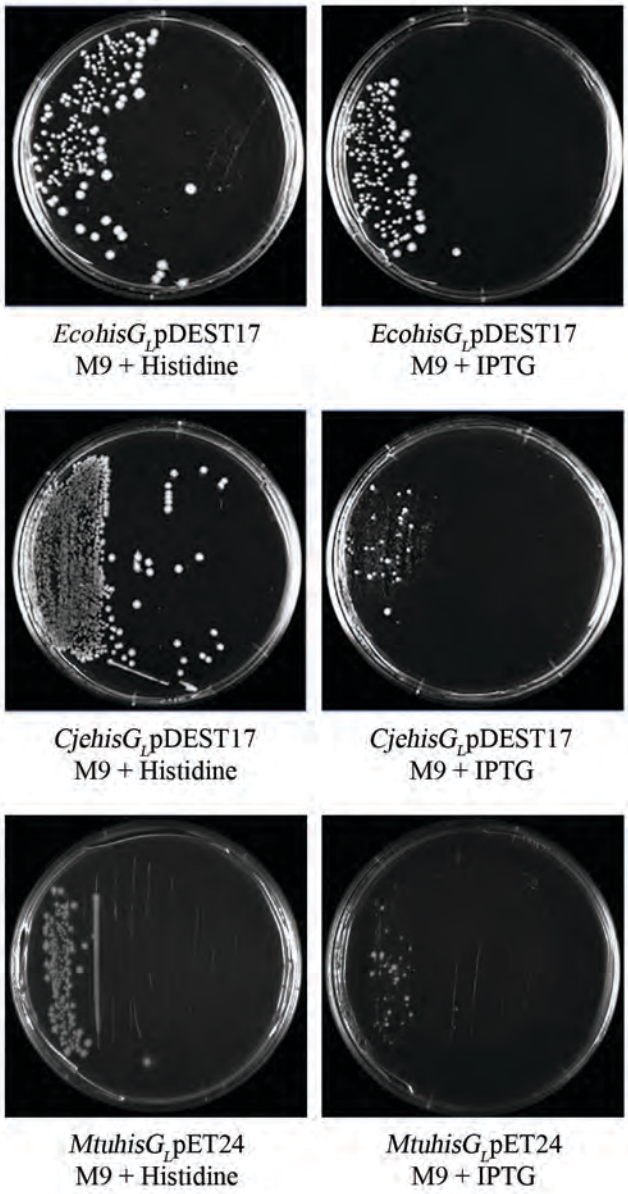

Figure II.2 Complementation studies of the wild type long form ATP-PRTase proteins. Colonies were observed on all three experimental plates with *Eco*HisG$_L$ resulting in more growth than either *Cje* or *Mtu*HisG$_L$. This shows that all three constructs express functional ATP-PRTase proteins *in vivo*.

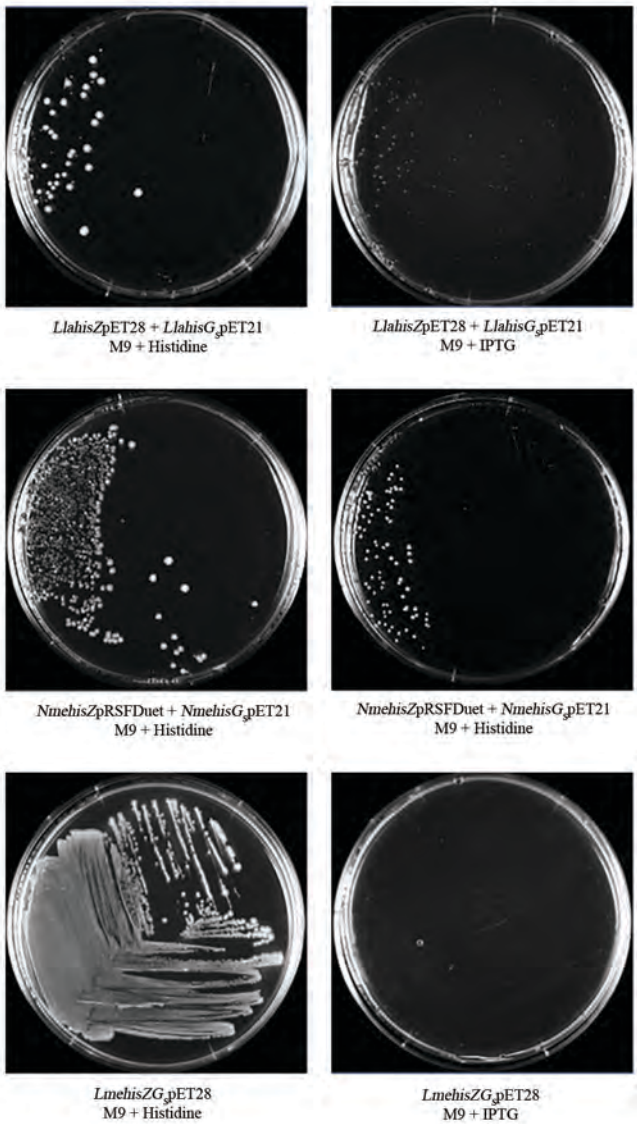

Figure II.3 Complementation studies of the wild type short form ATP-PRTase proteins. Growth was observed on the *Lla*HisZG$_S$ and *Nme*HisZG$_S$ plates, indicating that these are both functional ATP-PRTase proteins. *Lme*HisZG$_S$ was unable to rescue the strain.

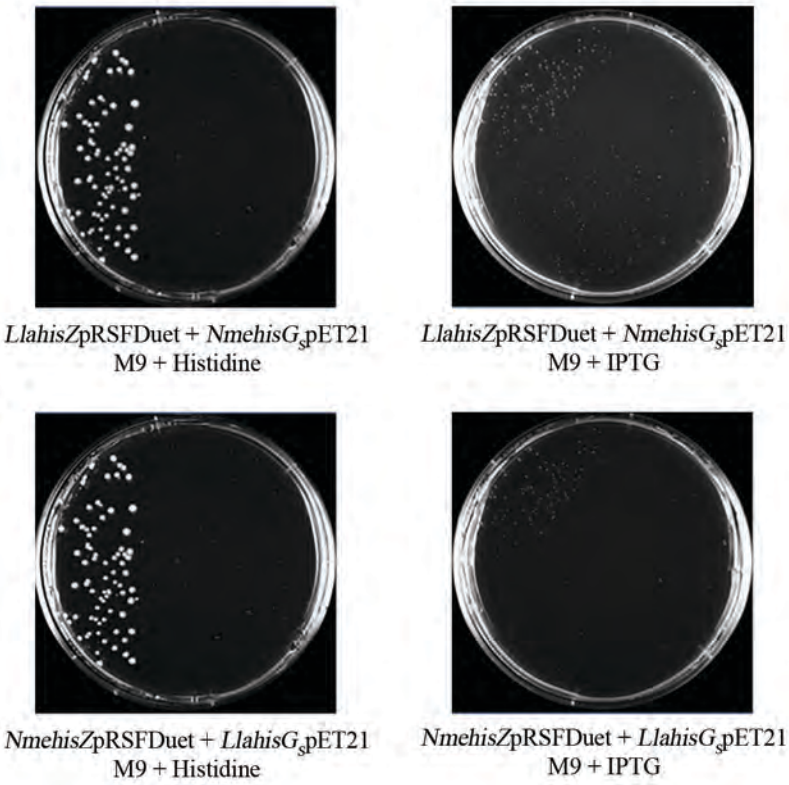

Figure II.4 Complementation studies of the combination of *Nme*HisZ with *Lla*HisG$_S$ and *Lla*HisZ with *Nme*HisG$_S$, both of which were able to restore ATP-PRTase activity.

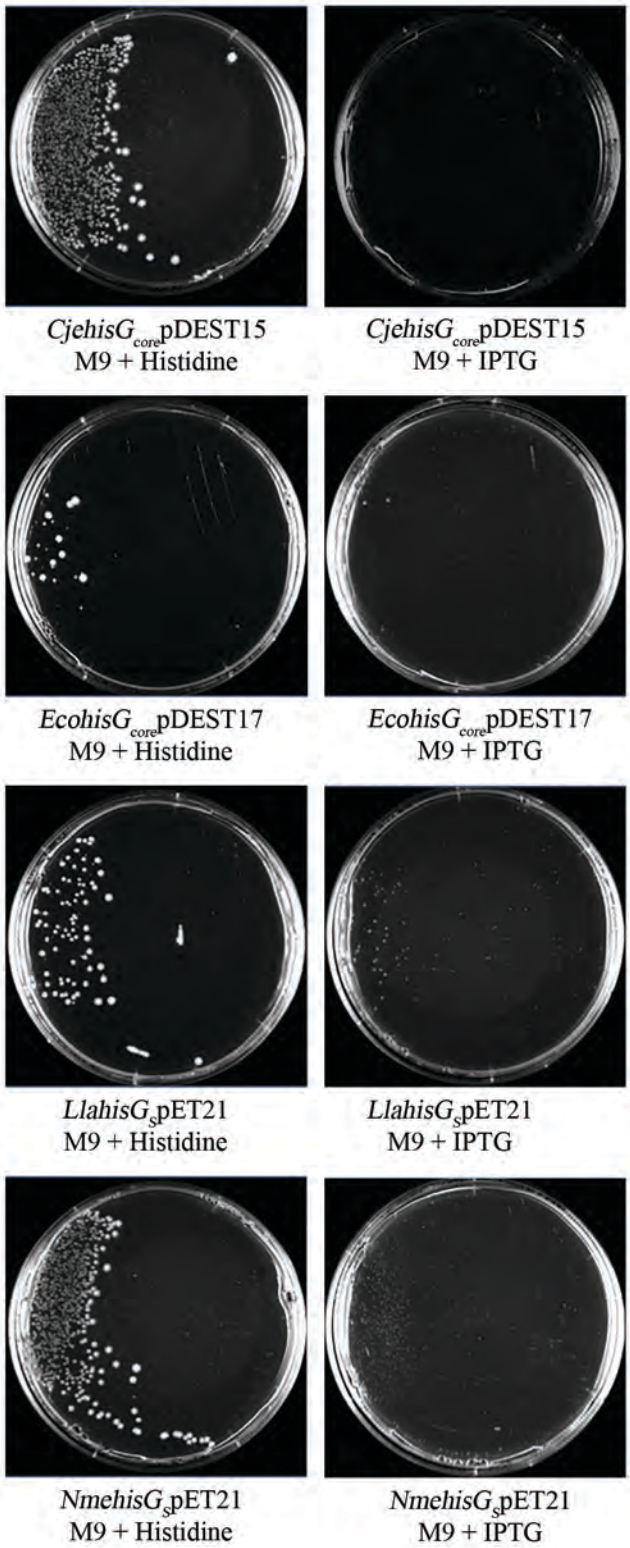

Figure II.5 Complementation studies of *Nme*HisG$_S$, *Lla*HisG$_S$ and the HisG$_{core}$ proteins. The HisG$_S$ and *Eco*HisG$_{core}$ proteins were able to rescue the strain, however *Cje*HisG$_{core}$ was not, although this may be due to the pDEST15 construct expressing a GST-tagged protein.

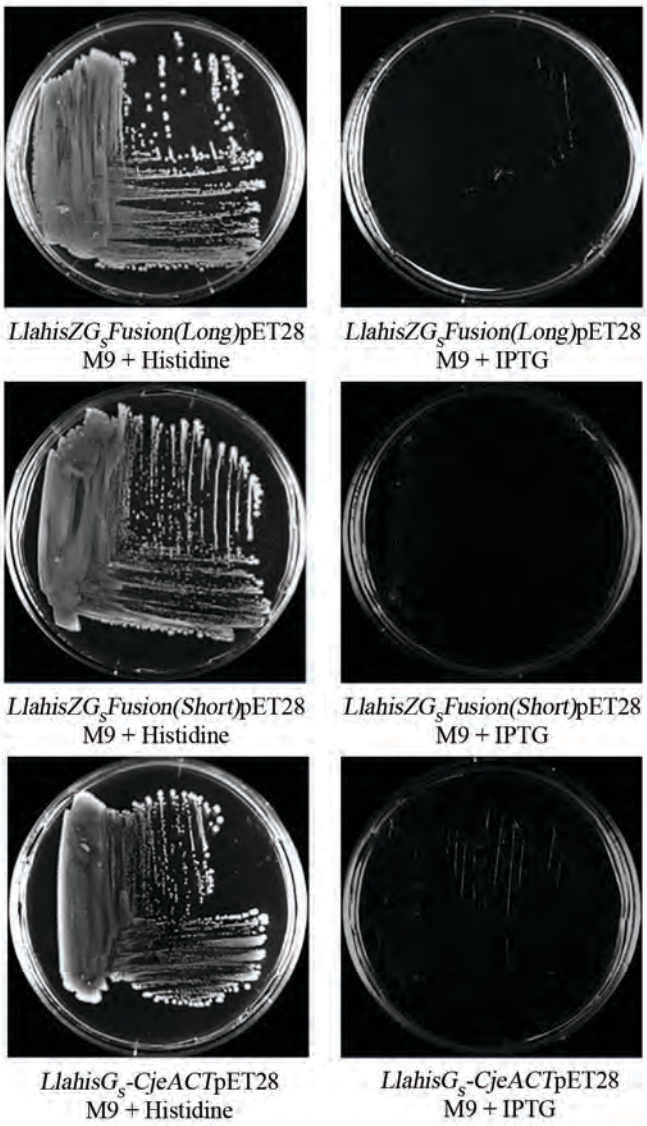

Figure II.6 Complementation studies of the *Lla*HisZG$_S$Fusion proteins and the *Lla*HisG$_S$-*Cje*ACT chimera, none of which were able to restore ATP-PRTase activity.

References

1. Ames, B. N.; Martin, R. G.; Garry, B., First step of histidine biosynthesis. *J. Biol. Chem.* **1961,** *236*, 2019-2026.

2. Alifano, P.; Fani, R.; Lio, P.; Lazcano, A.; Bazzicalupo, M.; Carlomagno, M. S.; Bruni, C. B., Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol. Rev.* **1996,** *60*, 44-69.

3. Cho, Y.; Sharma, V.; Sacchettini, J. C., Crystal structure of ATP phosphoribosyltransferase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **2003,** *278*, 8333-8339.

4. Lohkamp, B.; McDermott, G.; Campbell, S. A.; Coggins, J. R.; Lapthorn, A. J., The structure of *Escherichia coli* ATP-phosphoribosyltransferase: identification of substrate binding sites and mode of AMP inhibition. *J. Mol. Biol.* **2004,** *336*, 131-144.

5. Sissler, M.; Delorme, C.; Bond, J.; Ehrlich, S. D.; Renault, P.; Francklyn, C., An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96*, 8985-8990.

6. Mori, I.; Fonne-Pfister, R.; Matsunaga, S.-i.; Tada, S.; Kimura, Y.; Iwasaki, G.; Mano, J.-i.; Hatano, M.; Nakano, T.; et, a., A novel class of herbicides: specific inhibitors of imidazoleglycerol phosphate dehydratase. *Plant Physiol.* **1995,** *107*, 719-723.

7. Mousdale, D. M.; Coggins, J. R. *Amino acid synthesis*, Plenum, **1991**; 29-56.

8. Cho, Y.; Ioerger, T. R.; Sacchettini, J. C., Discovery of novel nitrobenzothiazole inhibitors for *Mycobacterium tuberculosis* ATP phosphoribosyl transferase (HisG) through virtual screening. *J. Med. Chem.* **2008,** *51*, 5984-5992.

9. Stepansky, A.; Leustek, T., Histidine biosynthesis in plants. *Amino Acids* **2006,** *30*, 127-142.

10. Fani, R.; Brilli, M.; Lio, P., The origin and evolution of operons: the piecewise building of the proteobacterial histidine operon. *J. Mol. Evol.* **2005,** *60*, 378-390.

11. Musick, W. D. L., Structural features of the phosphoribosyltransferases and their relationship to the human deficiency disorders of purine and pyrimidine metabolism. *CRC Crit. Rev. Biochem.* **1981,** *11*, 1-34.

12. Schramm, V. L.; Grubmeyer, C., Phosphoribosyltransferase mechanisms and roles in nucleic acid metabolism. *Prog. Nucleic Acid Res. Mol. Biol.* **2004,** *78*, 261-304.

13. Sinha, S. C.; Smith, J. L., The PRT protein family. *Curr. Opin. Struct. Biol.* **2001,** *11*, 733-739.

14. Tomchick, D. R.; Turner, R. J.; Switzer, R. L.; Smith, J. L., Adaptation of an enzyme to regulatory function: structure of *Bacillus subtilis* PyrR, a pyr RNA-binding attenuation protein and uracil phosphoribosyltransferase. *Structure (London)* **1998,** *6*, 337-350.

15. Eads, J. C.; Ozturk, D.; Wexler, T. B.; Grubmeyer, C.; Sacchettini, J. C., A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. *Structure (London)* **1997,** *5*, 47-58.

16. Chappie, J. S.; Canaves, J. M.; Han, G. W.; Rife, C. L.; Xu, Q.; Stevens, R. C., The Structure of a eukaryotic nicotinic acid phosphoribosyltransferase reveals structural heterogeneity among type II PRTases. *Structure (Cambridge)* **2005,** *13*, 1385-1396.

17. Okuno, E.; White, R. J.; Schwarcz, R., Quinolinic acid phosphoribosyltransferase: purification and partial characterization from human liver and brain. *J. Biochem.* **1988,** *103*, 1054-1059.

18. Lee, C. E.; Goodfellow, C.; Javid-Majd, F.; Baker, E. N.; Lott, J. S., The crystal structure of TrpD, a metabolic enzyme essential for lung colonization by *Mycobacterium tuberculosis*, in complex with its substrate phosphoribosylpyrophosphate. *J. Mol. Biol.* **2006,** *355*, 784-797.

19. Vega, M. C.; Zou, P.; Fernandez, F. J.; Murphy, G. E.; Sterner, R.; Popov, A.; Wilmanns, M., Regulation of the hetero-octameric ATP phosphoribosyl transferase complex from *Thermotoga maritima* by a tRNA synthetase-like subunit. *Mol. Microbiol.* **2005,** *55*, 675-686.

20. Champagne, K. S.; Sissler, M.; Larrabee, Y.; Doublie, S.; Francklyn, C. S., Activation of the hetero-octameric ATP phosphoribosyl transferase through subunit interface rearrangement by a tRNA synthetase paralog. *J. Biol. Chem.* **2005,** *280*, 34096-34104.

21. Smith, J. L.; Zaluzec, E. J.; Wery, J. P.; Niu, L.; Switzer, R. L.; Zalkin, H.; Satow, Y., Structure of the allosteric regulatory enzyme of purine biosynthesis. *Science.* **1994,** *264*, 1427-1433.

22. Martin, R. G., The first enzyme in histidine biosynthesis; the nature of feedback inhibition by histidine. *J. Biol. Chem.* **1963,** *238*, 257-268.

23. Voll, M. J.; Appella, E.; Martin, R. G., Purification and composition studies of phosphoribosyladenosine triphosphate: pyrophosphate phosphoribosyltransferase, the first enzyme of histidine biosynthesis. *J. Biol. Chem.* **1966,** *242*, 1760-1767.

24. Tebar, A. R.; Ballesteros, A. O., Kinetic properties of ATP phosphoribosyltransferase of *Escherichia coli*. *Mol. Cell. Biochem.* **1976,** *11*, 131-136.

25. Morton, D. P.; Parsons, S. M., Biosynthetic direction substrate kinetics and product inhibition studies on the first enzyme of histidine biosynthesis, adenosine triphosphate phosphoribosyltransferase. *Arch. Biochem. Biophys.* **1976,** *175*, 677-686.

26. Champagne, K. S.; Piscitelli, E.; Francklyn, C. S., Substrate recognition by the hetero-octameric ATP phosphoribosyltransferase from *Lactococcus lactis*. *Biochemistry* **2006,** *45*, 14933-14943.

27. Morton, D. P.; Parsons, S. M., Inhibition of ATP phosphoribosyltransferase by AMP and ADP in the absence and presence of histidine. *Arch. Biochem. Biophys.* **1977,** *181*, 643-648.

28. Griffin, J. E.; Gawronski, J. D.; DeJesus, M. A.; Ioerger, T. R.; Akerley, B. J.; Sassetti, C. M., High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog.* **2011,** *7*, e1002251.

29. Parish, T., Starvation survival response of *Mycobacterium tuberculosis*. *J. Bacteriol.* **2003,** *185*, 6702-6706.

30. Schuller, D. J.; Grant, G. A.; Banaszak, L. J., The allosteric ligand site in the V_{max}-type cooperative enzyme phosphoglycerate dehydrogenase. *Nat. Struct. Biol.* **1995,** *2*, 69-76.

31. Lang, E. J. M.; Cross, P. J.; Mittelstadt, G.; Jameson, G. B.; Parker, E. J., Allosteric ACTion: the varied ACT domains regulating enzymes of amino-acid metabolism. *Curr. Opin. Struct. Biol.* **2014,** *29*, 102-111.

32. Grant, G. A., The ACT domain: a small molecule binding domain and its role as a common regulatory element. *J. Biol. Chem.* **2006,** *281*, 33825-33829.

33. Pedreno, S.; Pisco, J. P.; Larrouy-Maumus, G.; Kelly, G.; de, C. L. P. S., Mechanism of feedback allosteric inhibition of ATP phosphoribosyltransferase. *Biochemistry* **2012,** *51*, 8027-8038.

34. Johnston, H. M.; Roth, J. R., Histidine mutants requiring adenine: selection of mutants with reduced *hisG* expression in *Salmonella typhimurium*. *Genetics* **1979,** *92*, 1-15.

35. Ohta, D.; Fujimori, K.; Mizutani, M.; Nakayama, Y.; Kunpaisal-Hashimoto, R.; Munzer, S.; Kozaki, A., Molecular cloning and

characterization of ATP-phosphoribosyl transferase from *Arabidopsis*, a key enzyme in the histidine biosynthetic pathway. *Plant Physiol.* **2000,** *122*, 907-914.

36. Ingle, R. A., Histidine biosynthesis. *Arabidopsis Book* **2011,** *9*, 141.

37. Bond, J. P.; Francklyn, C., Proteobacterial histidine-biosynthetic pathways are paraphyletic. *J. Mol. Evol.* **2000,** *50*, 339-347.

38. Aklujkar, M., Two ATP phosphoribosyltransferase isozymes of *Geobacter sulfurreducens* contribute to growth in the presence or absence of histidine and under nitrogen fixation conditions. *Can. J. Microbiol.* **2011,** *57*, 547-558.

39. Bovee, M. L.; Champagne, K. S.; Demeler, B.; Francklyn, C. S., The quaternary structure of the HisZ-HisG N-1-(5'-Phosphoribosyl)-ATP transferase from *Lactococcus lactis*. *Biochemistry* **2002,** *41*, 11838-11846.

40. Brashear, W. T.; Parsons, S. M., Evidence against a covalent intermediate in the adenosine triphosphate phosphoribosyltransferase reaction of histidine biosynthesis. *J. Biol. Chem.* **1975,** *250*, 6885-6890.

41. Chelsky, D.; Parsons, S. M., Stereochemical course of the adenosine triphosphate phosphoribosyltransferase reaction in histidine biosynthesis. *J. Biol. Chem.* **1975,** *250*, 5669-5673.

42. Ringia, E. A. T.; Schramm, V. L., Transition states and inhibitors of the purine nucleoside phosphorylase family. *Curr. Top. Med. Chem.* **2005,** *5*, 1237-1258.

43. Goitein, R. K.; Chelsky, D.; Parsons, S. M., Primary carbon-14 and α secondary tritium substrate kinetic isotope effects for some phosphoribosyltransferases. *J. Biol. Chem.* **1978,** *253*, 2963-2971.

44. Quiocho, F. A.; Ledvina, P. S., Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **1996,** *20*, 17-25.

45. Wang, Z.; Luecke, H.; Yao, N.; Quiocho, F. A., A low energy short hydrogen bond in very high resolution structures of protein receptor-phosphate complexes. *Nat. Struct. Biol.* **1997,** *4*, 519-521.

46. Tebar, A. R.; Fernandez, V. M.; Martin, D. R. R.; Ballesteros, A. O., Studies on the quaternary structure of the first enzyme for histidine biosynthesis. *Experientia* **1973,** *29*, 1477-1479.

47. Tebar, A. R.; Fernandez, V. M.; Martin, D. R.; Ballesteros, A. O., Quaternary structure of the first enzyme for histidine biosynthesis. *Experientia* **1973,** *29*, 1477-1479.

48. Bell, R. M.; Koshland, D. E., Jr., Phosphoribosyl-enzyme covalent intermediate in the first enzyme of histidine biosynthesis. *Biochem. Biophys. Res. Commun.* **1969,** *38*, 539-545.

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

50. Kryvi, H.; Klungsoeyr, L., Kinetic properties of phosphoribosyladenosine triphosphate synthetase. Inhibition by aggregation at high enzyme concentrations. *Biochim. Biophys. Acta, Enzymol.* **1971,** *235*, 429-434.

51. Bell, R. M.; Parsons, S. M.; Dubravac, S. A.; Redfield, A. G.; Koshland, D. E., Jr., Characterization of slowly interconvertible states of phosphoribosyladenosine triphosphate synthetase dependent on temperature, substrates, and histidine. *J. Biol. Chem.* **1974,** *249*, 4110-4118.

52. Arnez, J. G.; Harris, D. C.; Mitschler, A.; Rees, B.; Francklyn, C. S.; Moras, D., Crystal structure of histidyl-tRNA synthetase from

Escherichia coli complexed with histidyl-adenylate. *EMBO J.* **1995,** *14*, 4143-4155.

53. Nakatsu, T.; Kato, H.; Oda, J. i., Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase. *Nat. Struct. Biol.* **1998,** *5*, 15-19.

54. Carrodeguas, J. A.; Theis, K.; Bogenhagen, D. F.; Kisker, C., Crystal structure and deletion analysis show that the accessory subunit of mammalian DNA polymerase functions as a homodimer. *Mol. Cell* **2001,** *7*, 43-54.

55. Kulis-Horn, R. K.; Persicke, M.; Kalinowski, J., Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*. *Microbial Biotechnology* **2014,** *7*, 5-25.

56. Cheng, Y.; Zhou, Y.; Yang, L.; Zhang, C.; Xu, Q.; Xie, X.; Chen, N., Modification of histidine biosynthesis pathway genes and the impact on production of L-histidine in *Corynebacterium glutamicum*. *Biotechnol. Lett.* **2013,** *35*, 735-741.

57. Tebar, A. R.; Ballesteros, A.; Soria, J., Spin label studies of ATP phosphoribosyltransferase of *E. coli*. *Experientia* **1977,** *33*, 1014-1016.

58. Motlagh, H. N.; Wrabl, J. O.; Li, J.; Hilser, V. J., The ensemble nature of allostery. *Nature,* **2014,** *508*, 331-339.

59. Schendzielorz, G.; Dippong, M.; Gruenberger, A.; Kohlheyer, D.; Yoshida, A.; Binder, S.; Nishiyama, C.; Nishiyama, M.; Bott, M.; Eggeling, L., Taking control over control: use of product sensing in single cells to remove flux control at key enzymes in biosynthesis pathways. *ACS Synth. Biol.* **2014,** *3*, 21-29.

60. Zhang, Y.; Shang, X.; Deng, A.; Chai, X.; Lai, S.; Zhang, G.; Wen, T., Genetic and biochemical characterization of *Corynebacterium glutamicum* ATP phosphoribosyltransferase and its three mutants resistant to feedback inhibition by histidine. *Biochimie* **2012,** *94*, 829-838.

61. Willemoes, M.; Hove-Jensen, B.; Larsen, S., Steady state kinetic model for the binding of substrates and allosteric effectors to *Escherichia coli* phosphoribosyl-diphosphate synthase. *J. Biol. Chem.* **2000,** *275*, 35408-35412.

62. Eriksen, T. A.; Kadziola, A.; Bentsen, A.-K.; Harlow, K. W.; Larsen, S., Structural basis for the function of *Bacillus subtilis* phosphoribosyl-pyrophosphate synthetase. *Nat. Struct. Biol.* **2000,** *7*, 303-308.

63. Amrein, K. E.; Takacs, B.; Stieger, M.; Molnos, J.; Flint, N. A.; Burn, P., Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL. *Proc. Natl. Acad. Sci. U. S. A.* **1995,** *92*, 1048-1052.

64. Mitsuda, M.; Iwasaki, M., Improvement in the expression of CYP2B6 by co-expression with molecular chaperones GroES/EL in *Escherichia coli. Protein Expr Purif* **2006,** *46*, 401-405.

65. Lucent, D.; Stumpe, M. C.; Pande, V. S. In *Chaperonins: the machines which fold proteins*, World Scientific Publishing, **2011**; pp 87-98.

66. Willardson, B. M.; Tracy, C. M., Chaperone-mediated assembly of G protein complexes. *Subcell. Biochem.* **2012,** *63*, 131-153.

67. Whitaker, J. R., Determination of molecular weights of proteins by gel filtration of sephadex. *Analytical Chemistry* **1963,** *35*, 1950-1953.

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

69. Svergun, D.; Barberato, C.; Koch, M. H. J., CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl, Crystallogr.* **1995,** *28*, 768-773.

70. Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A., X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quart. Rev. Biophys.* **2007,** *40*, 191-285.

71. Svergun, D. I.; Petoukhov, M. V.; Koch, M. H. J., Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **2001,** *80*, 2946-2953.

72. Volkov, V. V.; Svergun, D. I., Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.* **2003,** *36*, 860-864.

73. Kozin, M. B.; Svergun, D. I., Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **2001,** *34*, 33-41.

74. Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I., New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* **2012,** *45*, 342-350.

75. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372*, 774-797.

76. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** *47*, 1739-1749.

77. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006,** *49*, 6177-6196.

78. Feher, V. A.; Durrant, J. D.; Van Wart, A. T.; Amaro, R. E., Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* **2014,** *25*, 98-103.

79. Nussinov, R.; Tsai, C.-J., Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struct Biol* **2014,** *30C*, 17-24.

80. Aberg, A.; Yaremchuk, A.; Tukalo, M.; Rasmussen, B.; Cusack, S., Crystal structure analysis of the activation of histidine by *Thermus thermophilus* histidyl-tRNA synthetase. *Biochemistry* **1997,** *36*, 3084-3094.

81. Klungsoeyr, L.; Kryvi, H., Sedimentation behavior of phosphoribosyladenosine triphosphate synthetase. Effects of substrates and modifiers. *Biochim. Biophys. Acta, Enzymol.* **1971,** *227*, 327-336.

82. Jafri, R. Z.; Ali, A.; Messonnier, N. E.; Tevi-Benissan, C.; Durrheim, D.; Eskola, J.; Fermon, F.; Klugman, K. P.; Ramsay, M.; Sow, S.; Shao, Z.; Bhutta, Z. A.; Abramson, J., Global epidemiology of invasive meningococcal disease. *Population Health Metrics* **2013,** *11*.

83. Lohkamp, B.; Coggins, J. R.; Lapthorn, A. J., Purification, crystallization and preliminary x-ray crystallographic analysis of ATP-phosphoribosyltransferase from *Escherichia coli*. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000,** *D56*, 1488-1491.

84. Troubleshooting Guide For Cloning, New England Biolabs. http://www.neb.com/tools-and-resources/troubleshooting-guides/troubleshooting-guide-for-cloning (accessed May 2014).

85. Dall-Larsen, T., Stopped flow kinetic studies of adenosine triphosphate phosphoribosyl transferase, the first enzyme in the histidine biosynthesis of *Escherichia coli*. *Int. J. Biochem.* **1988,** *20*, 811-815.

86. Peracchi, A.; Mozzarelli, A., Exploring and exploiting allostery: Models, evolution, and drug targeting. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2011,** *1814*, 922-933.

87. Finn, R. D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; Sonnhammer, E. L. L.; Tate, J.; Punta, M., Pfam: the protein families database. *Nuc. Acids Res.* **2014,** *42*, D222-D230.

88. Pusey, M. L.; Liu, Z. J.; Tempel, W.; Praissman, J.; Lin, D. W.; Wang, B. C.; Gavira, J. A.; Ng, J. D., Life in the fast lane for protein crystallization and X-ray crystallography. *Prog. Biophys. Mol. Biol.* **2005,** *88*, 359-386.

89. Luft, J. R.; Snell, E. H.; DeTitta, G. T., Lessons from high-throughput protein crystallization screening: 10 years of practical experience. *Expert Opinion on Drug Discovery* **2011,** *6*, 465-480.

90. Su, X.-D.; Zhang, H.; Terwilliger, T. C.; Liljas, A.; Xiao, J.; Dong, Y., Protein crystallography from the perspective of technology developments. *Crystallog. Rev.* **2015,** *21*, 122-153.

91. Forster, M. J., Molecular modelling in structural biology. *Micron* **2002,** *33*, 365-384.

92. Jacobson, M. P.; Kalyanaraman, C.; Zhao, S. W.; Tian, B. X., Leveraging structure for enzyme function prediction: methods, opportunities, and challenges. *Trends in Biochemical Sciences* **2014,** *39*, 363-371.

93. Schwede, T., Protein Modeling: What Happened to the "Protein Structure Gap"? *Structure* **2013,** *21*, 1531-1540.

94. Makarova, K.; Slesarev, A.; Wolf, Y.; Sorokin, A.; Mirkin, B.; Koonin, E.; Pavlov, A.; Pavlova, N.; Karamychev, V.; Polouchine, N.; Shakhova, V.; Grigoriev, I.; Lou, Y.; Rohksar, D.; Lucas, S.; Huang, K.; Goodstein, D. M.; Hawkins, T.; Plengvidhya, V.; Welker, D.; Hughes, J.; Goh, Y.; Benson, A.; Baldwin, K.; Lee, J. H.; Díaz-Muñiz, I.; Dosti,

B.; Smeianov, V.; Wechter, W.; Barabote, R.; Lorca, G.; Altermann, E.; Barrangou, R.; Ganesan, B.; Xie, Y.; Rawsthorne, H.; Tamir, D.; Parker, C.; Breidt, F.; Broadbent, J.; Hutkins, R.; O'Sullivan, D.; Steele, J.; Unlu, G.; Saier, M.; Klaenhammer, T.; Richardson, P.; Kozyavkin, S.; Weimer, B.; Mills, D., Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103*, 15611-15616.

95. Robert, X.; Gouet, P., Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014,** *42*, W320-W324.

96. Cross, P. J.; Allison, T. M.; Dobson, R. C. J.; Jameson, G. B.; Parker, E. J., Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110*, 2111-2116, S2111/2111-S2111/2117.

97. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H., Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2006,** *2*.

98. Delorme, C.; Ehrlich, S. D.; Renault, P., Histidine biosynthesis genes in *Lactococcus lactis* subsp. *lactis*. *J Bacteriol* **1992,** *174*, 6571-6579.

99. Reynolds, D. J.; Penn, C. W., Characteristic of *Helicobacter pylori* growth in a defined medium and determination of its amino acid requirements. *Microbiol.* **1994,** *140*, 2649-2656.

100. Tomb, J.-F.; White, O.; Kerlavage, A. R.; Clayton, R. A.; Sutton, G. G.; Fleischmann, R. D.; Ketchum, K. A.; Klenk, H. P.; Gill, S.; Dougherty, B. A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E. F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H. G.; Glodek, A.; McKenney, K.; Fitzegerald, L. M.; Lee, N.; Adams, M. D.; Hickey, E. K.; Berg, D. E.; Cocayne, J. D.; Utterback, T. R.; Peterson, J.

D.; Kelley, J. M.; Cotton, M. D.; Weidman, J. M.; Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes, W. S.; Borodovsky, M.; Karp, P. D.; Smith, H. O.; Fraser, C. M.; et, a., The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **1997,** *388*, 539-547.

101. Alm, R. A.; Ling, L.-S. L.; Moir, D. T.; King, B. L.; Brown, E. D.; Doig, P. C.; Smith, D. R.; Noonan, B.; Guild, B. C.; DeJonge, B. L.; Carmel, G.; Tummino, P. J.; Caruso, A.; Uria-Nickelsen, M.; Mills, D. M.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S. D.; Jiang, Q.; Taylor, D. E.; Vovis, G. F.; Trust, T. J., Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **1999,** *397*, 176-180.

102. Fraser, C. M.; Gocayne, J. D.; White, O.; Adams, M. D.; Clayton, R. A.; Fleischmann, R. D.; Bult, C. J.; Kerlavage, A. R.; Sutton, G.; et, a., The minimal gene complement of *Mycoplasma genitalium*. *Science* **1995,** *270*, 397-403.

103. Juliao, P. C.; Marrs, C. F.; Xie, J.; Gilsdorf, J. R., Histidine auxotrophy in commensal and disease-causing nontypeable *Haemophilus influenzae*. *J. Bacteriol.* **2007,** *189*, 4994-5001.

104. Ino, I.; Hartman, P. E.; Hartman, Z.; Yourno, J., Deletions fusing the *hisG* and *hisD* genes in *Salmonella typhimurium*. *J. Bacteriol.* **1975,** *123*, 1254-1264.

105. Nagradova, N. K., Polyfunctional Enzymes. *Biochemistry-Moscow* **1992,** *57*, 659-675.

106. Fani, R.; Lio, P.; Lazcano, A., Molecular evolution of the histidine biosynthetic pathway. *J. Mol. Evol.* **1995,** *41*, 760-774.

107. Kim, Y. J.; Eom, H.-J.; Seo, E.-Y.; Lee, D. Y.; Kim, J. H.; Han, N. S., Development of a chemically defined minimal medium for the exponential growth of *Leuconostoc mesenteroides*. *J. Microbiol. Biotechnol.* **2012,** *22*, 1518-1522.

108. Halliwell, C. M.; Morgan, G.; Ou, C.-P.; Cass, A. E. G., Introduction of a (Poly)histidine tag in L-lactate dehydrogenase produces a mixture of active and inactive molecules. *Anal. Biochem.* **2001,** *295*, 257-261.

109. Svergun, D. I., Small-angle scattering studies of macromolecular solutions. *J. Appl. Crystallogr.* **2007,** *40*, s10-s17.

110. Heller, W. T., Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010,** *66*, 1213-1217.

111. Wolfenden, R. V., Transition state analogues for enzyme catalysis. *Nature* **1969,** *223*, 704-705.

112. Pauling, L., Molecular architecture and biological reactions. *Chem. Eng. News* **1946,** *24*, 1375-1377.

113. Burgos, E. S.; Vetticatt, M. J.; Schramm, V. L., Recycling nicotinamide. The transition-state structure of human nicotinamide phosphoribosyltransferase. *J. Am. Chem. Soc.* **2013,** *135*, 3485-3493.

114. Parikh, S.; Schramm, V.; Clinch, K.; Frohlich, R.; Furneaux, R.; Harvey, J.; Tyler, P. *Design, synthesis, and testing of ADP-ribose inhibitors: Chromogenic substrates of bacterial ADP-ribosylating toxins*, American Chemical Society: 2012; pp 129.

115. Ho, M.-C.; Shi, W.; Rinaldo-Matthis, A.; Tyler, P. C.; Evans, G. B.; Clinch, K.; Almo, S. C.; Schramm, V. L., Four generations of transition-state analogues for human purine nucleoside phosphorylase. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107*, 4805-4812