Transport and catabolism of murein tripeptide in *Escherichia coli* K-12

Abbas Maqbool

PhD

University of York

Department of Biology

December 2011

Abstract

In bacteria, extracellular peptides are not only a source of nutrients but also play important roles in cell-cell communication. The primary mode of uptake of these peptides in bacteria is via ATP Binding Cassette (ABC) transporters. The substrate-binding protein (SBP) of these transporters captures extracellular peptides and delivers them to membrane-bound components for transport. Bacterial peptide-binding SBPs that function in ABC transporters have evolved predominantly to function as generalists, recognising peptides of particular lengths but with low or no sequence specificity. However, within the peptide-specific SBPs are examples that appear to have evolved subsequently to recognise specific peptides.

Here we have provided the first biophysical evidence, using native electrospray mass spectrometry, for the binding of a specific peptide to an *Escherichia coli* SBP. This peptide is a cell wall component named murein tripeptide (Mtp, L-Ala-γ-D-Glu-*meso*-Dap) and is transported by an ABC transporter containing the SBP MppA. We demonstrate that MppA recognises Mtp specifically and with high affinity (K_D ~ 250 nM), as determined by protein fluorescence spectroscopy and isothermal titration calorimetry. The crystal structure of MppA in complex with Mtp has been solved in order to understand the structural basis of specific binding of Mtp to MppA. Comparison of the structure of MppA-Mtp with structures of general tripeptides bound to oligopeptide binding protein OppA, reveals close similarity in the protein chains which fold to form an enclosed interdomain pocket in which the respective peptides reside. The peptide ligands superimpose remarkably closely given the profound differences in their structure. Strikingly, the effect of the D-stereochemistry, which projects the side chain of residue 2 in Mtp in the direction of the main chain in a conventional tripeptide, is compensated by the side chain γ-linkage of the carboxylate of the D-Glu to the amino group of diaminopimelic acid. The resulting amide linkage mimicks the peptide bond between residues 2 and 3 of a conventional tripeptide. Specificity for Mtp is conferred by a series of ionic and polar residues of MppA which make charge-charge and dipole-charge interactions with the highly ionised peptide.

Analysis of multiple bacterial genomes revealed that *mppA* is often situated near a gene called *mpaA* which encodes a putative amidase enzyme which is thought to hydrolyze murein tripeptide. This genetic organization is conserved among γ-proteobacteria and suggests that *mppA* and *mpaA* function together and are a part of peptide catabolic pathway. The MpaA enzyme from *E. coli* was purified to homogeneity as a His-tagged form, and its kinetic properties and parameters were determined. MpaA was found to hydrolyze Mtp efficiently but it did not cleave murein tetrapeptide. Replacement of *meso*-Dap by L-Lys in the tripeptide resulted in much lower efficiency. Mass spectrometric analysis of purified MpaA suggested that it co-purifies with a zinc ion and exists as a dimer in solution. The first crystal structure of an MpaA protein from *Vibrio harveyi* was solved to 2.17Å. The structural fold of *V. harveyi* MpaA was similar to eukaryotic carboxypeptidases. Analysis of the structure showed an obvious dimer interface between two monomers of MpaA and presence of a zinc ion in the active centre of each monomer consistent with the mass spectrometric data. The structure provides the basis for future modelling and mutational studies for extensive functional characterization of MpaA.

Contents

List of tables

List of figures

Acknowledgements

In the first place I would like to thank Dr. Gavin Thomas and Prof. Anthony Wilkinson for providing me an opportunity to undertake PhD under their supervision and for all of their help, guidence and tolerance. I am also grateful to my thesis advisory panel, Prof. Gideon Davies and Dr. Christoph Baumann, who gave guidance and encouragement at regular intervals.

I thank former and present members of Thomas group for their help and nice company with special thanks to Judith. I must say thanks to Vladimir and Elena for helping me fishing and testing crystals in the beginning of my work.

I would like to thank Professor Eleanor Dodson for her help and expertise with the model building of the structure *Vh*_MpaA. I thank Dr. Andrew Leech and Berni Strongitharm in the Molecular Interactions Lab for the time I took up.

My sincere thanks go to Amna for her continuing love and support. I would also like to thank my family members especially the army of young nephews and nieces for their immense love and support. To all my friends, you know who you are, a big thanks for your friendship, support, patience and being there for me in the good and the not so good moments.

Last but certainly not least, I dedicate this thesis to my mother whose prayers always lifted me up and enabled me to achieve the highest goals of my life.

Author's declaration

I declare that all work presented in this thesis is based on investigations conducted by myself. The work presented in section 6.6 was done by Dr. Mireille Herve in the University of Paris-Sud, France. Any contribution from others is truthfully acknowledged and indicated in the bibliography. Part of the thesis has been published in *J Biol Chem* **286** (36): 31512–31521 (Maqbool *et al*., 2011). Results from Chapter 6 are being prepared for publication.

Abbas Maqbool

Chapter 1

1 Introduction

The chapter begins by explaining the need for transporters in living cells and a brief overview on the major types of transporters present in living organisms. A detailed description is given on the ATP binding cassette (ABC) transporters using the *E. coli* maltose transporter as a model. There is a separate section on substrate-binding proteins mainly discussing their classification into different groups based upon their structural similarity. Transport of peptides in bacteria is described with the main focus on ABC transporters of peptides in *E. coli*. In the later part of this chapter I introduce the peptidoglycan, including its structure, synthesis and turnover. The section also explains the existing pathway for recycling of cell wall derived murein tripeptide in *E. coli*. Finally, an overview of protein crystallography as a technique to obtain three dimensional structures of proteins is given.

1.1 The need for transporters

Living cells are surrounded by a membrane composed of phospholipids, separating the inside from the outside. In Gram-positive bacteria, this membrane is surrounded by a thick layer of protective peptidoglycan, together making up the cell envelope (Figure 1.1). Gram-negative bacteria possess a second membrane above a much thinner peptidoglycan layer, producing another distinct compartment for the cell called the periplasm (Figure 1.1).

The phospholipids are arranged in a bilayer, with their polar, hydrophilic phosphate heads facing outwards, and their non-polar, hydrophobic fatty acid tails facing inwards towards the middle of the bilayer (Wilkins *et al*., 1971). Cell membranes are a barrier to most substances, and this property allows materials to be concentrated inside cells, excluded from cells, or simply separated from the outside environment. This is compartmentalisation and is essential for life, as it enables metabolic reactions to take place that would otherwise be impossible. Obviously materials need to be able to enter and leave the cells, and thus transporters have evolved to move exogenous and endogenous substrates across this barrier (Saier Jr, 1994, Saier Jr, 2000b). These act as gateways controlling what compounds are allowed to enter and leave the cell.

Various types of membrane transporters have evolved to suit environment and function. They can be divided into two superfamilies. First, the transporters that function independent of an energy source and allow the unimpeded movement of substrates

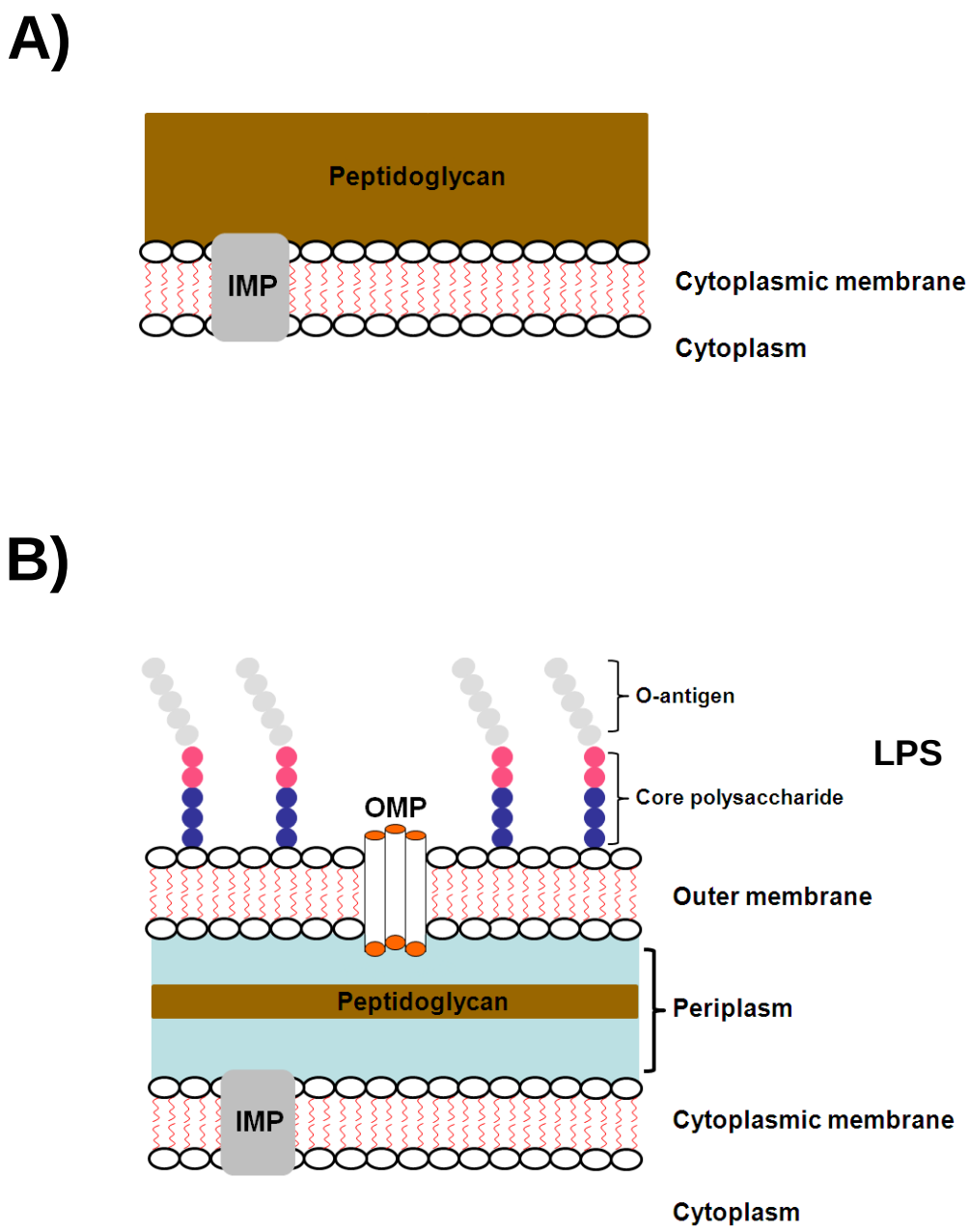


Figure 1.1 Components of a typical bacterial cell envelope

Cell envelope of Gram-positive bacteria **(A)** and Gram-negative bacteria **(B)** indicating the relative locations of the cytoplasmic (inner) membrane, periplasm, peptidoglycan, outer membrane and lipopolycassharide (LPS). IMP represents inner membrane protein and OMP stands for outer membrane protein.

across the membrane down a concentration gradient, which are traditionally known as "channels". The second superfamily consists of the transporters that utilize energy to enable active transport, often against a concentration gradient. This energy comes either from an 'energy currency' such as ATP (primary transporters) or via the proton gradient, in which case the transporters are classed as secondary transporters. These superfamilies, and all substrate uptake or export systems more generally, have been defined by Milton Saier, and can be divided up into five accepted classes of the Transport Classification (TC) system based on similar mechanisms of function and sequence (Busch and Saier, Jr., 2002). All transport systems are included in the Transport Classification Database, TCDB available at http://www.tcdb.org/ (Saier *et al.*, 2009).

1.2 Major types of transporters

Diffusion of a substrate on either side of a membrane via a channel is the least complex method of transport. This process does not involve any energy and is called passive transport. The channels constitute the first class of the transport classification system. An example of this phenomenon is glycerol facilitator, GlpF (TC1.A.8.1.1) of *Escherichia coli* (Figure 1.2) (Sweet *et al.*, 1990, Fu *et al.*, 2000).

The second class of TC system are ion gradient driven transporters, called secondary active transporters where transport of the solute is driven by its coupling with ion flow down an electrochemical gradient, such as H^+ or Na^+ (Saier Jr, 2000b) (Figure 1.2). These ion driven transporters can be subdivided into three classes; uniporters, antiporters and symporters. Uniporters transport one substrate in one direction and the latter two are involved in the transport of two different substrates simultaneously. Symporters transport two substrates in the same direction whereas antiporters transport two substrates in the opposite directions. The classical example of this class of transporter is the *E. coli* lactose permease (TC2.A.1.5.1). This is a lactose:H^+ symporter where the translocation of extracellular protons into the low proton environment of the cytoplasm drives the accumulation of intracellular lactose (Wong *et al.*, 1970).

The third class of transporters are primary active transporters that use a primary source of energy to drive active transport of a solute against a concentration gradient. Primary

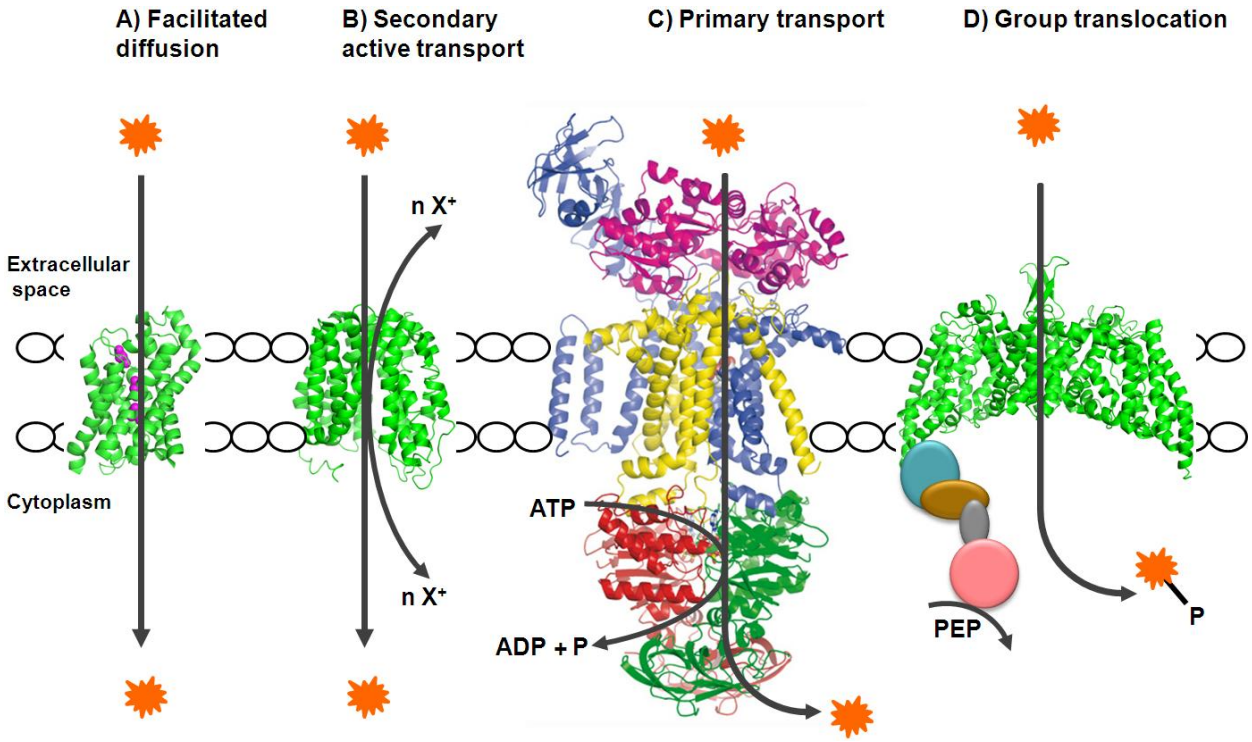


Figure 1.2 Major types of transporters in bacteria

A) Substrate specific channels that equalise the concentration of the substrate across the membrane. A monomer of the *E. coli* glycerol facilitator (GlpF) is shown with three molecules of bound glycerol as coloured spheres. **B**) Secondary active transporters couple translocation of the substrate with ion flow down an electrochemical gradient. The example given here is the *E. coli* lactose permease, LacY. **C**) Primary transporters use ATP energy to translocate the substrate. The example given here is a catalytic intermediate of the *E. coli* maltose transporter, $MalEFGK_2$ where ATP hydrolysis drives substrate translocation across the membrane. **D**) Group translocators. This group catalyse transport of the substrate at the same time as its phosphorylation. The example shown here is of a phosphorylation-coupled saccharide transporter of *Bacillus cereus*. In each example substrate is represented by a orange star.

sources known to be coupled to transport are chemical or solar energy. A well studied example of a primary transporter that uses chemical energy are the P-type ATPases. These adenosine triphosphate (ATP) driven ion pumps are multisubunit complexes found ubiquitously from bacteria to animal cells (Kühlbrandt, 2004). P-type ATPase complexes pump a number of different substrates including H^+, K^+, Na^+ and Ca^{2+} forming electrochemical gradients. The formation of these electrochemical gradients can power transport (Doeven *et al*., 2004). The primary transporters most appropriate to the current project are the ATP-binding cassette trasnporters that use ATP hydrolysis to drive substrate translocation across the membrane. In the case of importers, the substrate is often delivered to the permease by a substrate-binding protein, which binds the substrate with high specificity and affinity (Figure 1.2) (Ames and Nikaido, 1978). ATP is hydrolysed by ATP-binding cassettes, hence the superfamily name of ABC transporters. The classical example of this family is the *E. coli* maltose transporter (TC3.A.1.1.1) (Oldham *et al*., 2007).

Class four transporters are called group translocators and they alter the substrate during translocation as in the bacterial phosphotransferase system (PTS) (Figure 1.2) (Saier Jr *et al*., 2005). The range of substrates for these transporters is limited to sugars including glucose, ascorbate, glucosides, maltose and *N*-acetylglucosamine (Saier Jr *et al*., 2005). The substrate is transported into the cell by an integral membrane protein (Cao *et al*., 2011) in the complex and then phosphorylated by a complex of cytoplasmic enzymes using phosphoenol pyruvate (PEP) as a substrate (Saier Jr *et al*., 2005).

It is also worth mentioning that different classes of transporters exist for the inward translocation of substrates through the outer membrane of Gram-negative bacteria. These include porins (non specific channels), specific channels and TonB-dependent receptors, which are powered transporters and are capable of active transport, accumulating their substrate against a concentration gradient (Nikaido, 2003).

1.3 ATP binding cassette (ABC) transporters

The ATP-binding cassette (ABC) transporters are widespread transport system among living organisms with representatives in all three domains of life from bacteria to higher eukaryotes, including humans (Holland, 2003). ABC transporters constitute one of the largest superfamilies of paralogous sequences. A lot of ABC transporters have been implicated in various human diseases, including cystic fibrosis (Gadsby *et al*., 2006) and multidrug resistance in chemotherapeutic treatment (van de Water *et al*., 2007).

The ABC transporters were first discovered and characterized in prokaryotes in the early 1970 (Davidson *et al*., 2008). The most extensively analyzed systems are the high affinity maltose and histidine uptake systems of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium, respectively. The first nucleotide sequence of all the components of an ABC transporter was determined in 1982 for histidine transporter from *S.* Typhimurium (Higgins *et al*., 1982). The term ABC transporter was coined as a common name for transporters of this superfamily in 1990 (Hyde *et al*., 1990).

ABC transporters are called primary active transporters because it has been found that they couple the hydolysis of ATP forming ADP and P_i to the transport of the substrate across the membrane. They perform active transport on a number of diverse substrates resulting in the formation of a high substrate gradient up to (10^6:1) (Dippel and Boos, 2005).

Although the composition of ABC transporter components is quite variable, they can be generally divided into two categories; ABC importers and ABC exporters. ABC importers catalyze the unidirectional uptake of a wide range of substrates including sugars (Dean *et al*., 1989), a number of amino acids (Higgins *et al*., 1982), peptides (Tame *et al*., 1994), vitamins (Borths *et al*., 2002) and metal chelate complexes (Pinkett *et al*., 2007). ABC exporters catalyze the extrusion of virulence factors, for example haemolysin (Dinh *et al*., 1994) and hydrophobic compounds, the majority of which have anti-microbial action.

The core structure of an ABC transporter contains two components; a hydrophobic integral membrane domain that forms the translocation channel also known as transmembrane domain (TMD) and a soluble, cytoplasmic nucleotide binding domain (NBD), also known as ATPase domain that binds and hydrolyses ATP. In the functional

ABC transporter, the TMD and NBD are both dimeric and have been found in a number of configurations. Often the four domains are independent polypeptide chains that come together to make a multimeric complex. However, they have also been found to have one TMD and one ATPase domain fused into one polypeptide and the functional transporter composed of two of these so-called "half transporter"(Dawson *et al*., 2007). A third structural component of ABC importers is an extracytoplasmic solute receptor also known as substrate-binding protein (SBP). This component is found in the periplasm of Gram-negative bacteria and attached to the cytoplasmic membrane by a lipid moiety in Gram-positive bacteria (Sutcliffe and Russell, 1995). The SBP binds the substrate of the transporter with high affinity (sub micromolar) and delivers it to the membrane components of the transporter for translocation. The SBP is thought to confer specificity, high affinity and also unidirectionality to the transporter (Holland, 2003).

1.4 The *E. coli* ABC transporter for maltose as a model system

The *E. coli* genome contains 65 predicted ABC transporters, taking up over 5% of the total genome (Moussatova *et al*., 2008). Fifteen of these transporters lack SBP and are presumed to be ATP-driven exporters, leaving the remaining 50 as importers. The maltose ABC transporter in *E. coli* is the most comprehensively studied ABC transporter in nature. It belongs to the carbohydrate uptake transporter family (CUT) as defined by Saier (Saier Jr, 2000a) and has been shown to transport di and oligosaccharides. The maltose uptake system in *E. coli* is composed of a periplasmic SBP (maltose binding protein; MBP, MalE), two integral membrane domains (MalF and MalG) and a homodimer of the ATPase domain ($MalK_2$). These components are discussed in detail in the following sections.

1.4.1 Maltose binding protein MalE

MalE is a soluble protein that is found in the periplasm of Gram-negative bacteria or anchored to the cytoplasmic membrane as a lipoprotein in Gram-positive bacteria. Like other SBPs, MalE, is synthesized in the cytoplasm and then translocated to the periplasm by the Sec translocon that recognizes an N-terminal signal peptide. The signal peptide is cleaved off during the translocation process producing the mature protein.

In vitro studies using purified MalE has shown that it binds maltose and maltodextrin with a high affinity and dissociation constant (K_D) of 1 μM (Kellermann and Szmelcman, 1974). MalE is also known to have been recruited for a use as a chemotaxis receptor. The high resolution structure of MalE has been determined in both the liganded and unliganded states (Sharff *et al*., 1992, Spurlino *et al*., 1991). These structures revealed that MalE is composed of two globular domains connected by a flexible hinge region (Figure 1.3). The ligand binding site is in the middle of the globular domains which is freely accessible to the solvent prior to binding the ligand. Upon binding ligand, the protein undergoes a large conformational change which results in complete occlusion of the ligand from the bulk solvent (Figure 1.3). Residues from both globular domains contribute to the formation of the binding site and make a number of van der Waals interactions and hydrogen bonding interactions.

1.4.2 ATPase domain (MalK)

MalK is the NBD subunit of the maltose ABC transporter and is the most comprehensively studied of all the ABC transporter subunits. MalK was experimentally confirmed to be the ATPase subunit because when expressed and purified alone, MalK spontaneously hydrolyses ATP (Morbach *et al*., 1993), whereas, in the full transporter ATP hydrolysis is tightly coupled to the presence of the substrate-binding protein MalE (Davidson *et al*., 1992).

The first structure of a MalK homologue was from the archaeon *Thermococcus litoralis*. This structure was determined to be non-physiological due to the small dimer interface and peculiar arrangement of sub-domains (Diederichs *et al*., 2000). More recent structures of MalK in isolation (Chen *et al*., 2003, Lu *et al*., 2005) and in complex with the rest of the complex ($MalEFGK_2$) from *E. coli* have been solved that are consistent with the biochemical data for MalK (Oldham and Chen, 2011, Oldham *et al*., 2007). The structure of MalK subunit has revealed that it is composed of two domains, an N-terminus nucleotide binding domain (NBD) and a C-terminus regulatory domain (RD) (Figure 1.4). NBD can be further divided into two subdomains: RecA like subdomain present in many ATPases (Story and Steitz, 1992), containing the Walker A and Walker B motifs (Walker *et al*., 1982) and a helical subdomain that is specific to ABC transporters and contains the ABC family signature motif, LSGGQ (Hyde *et al*., 1990).

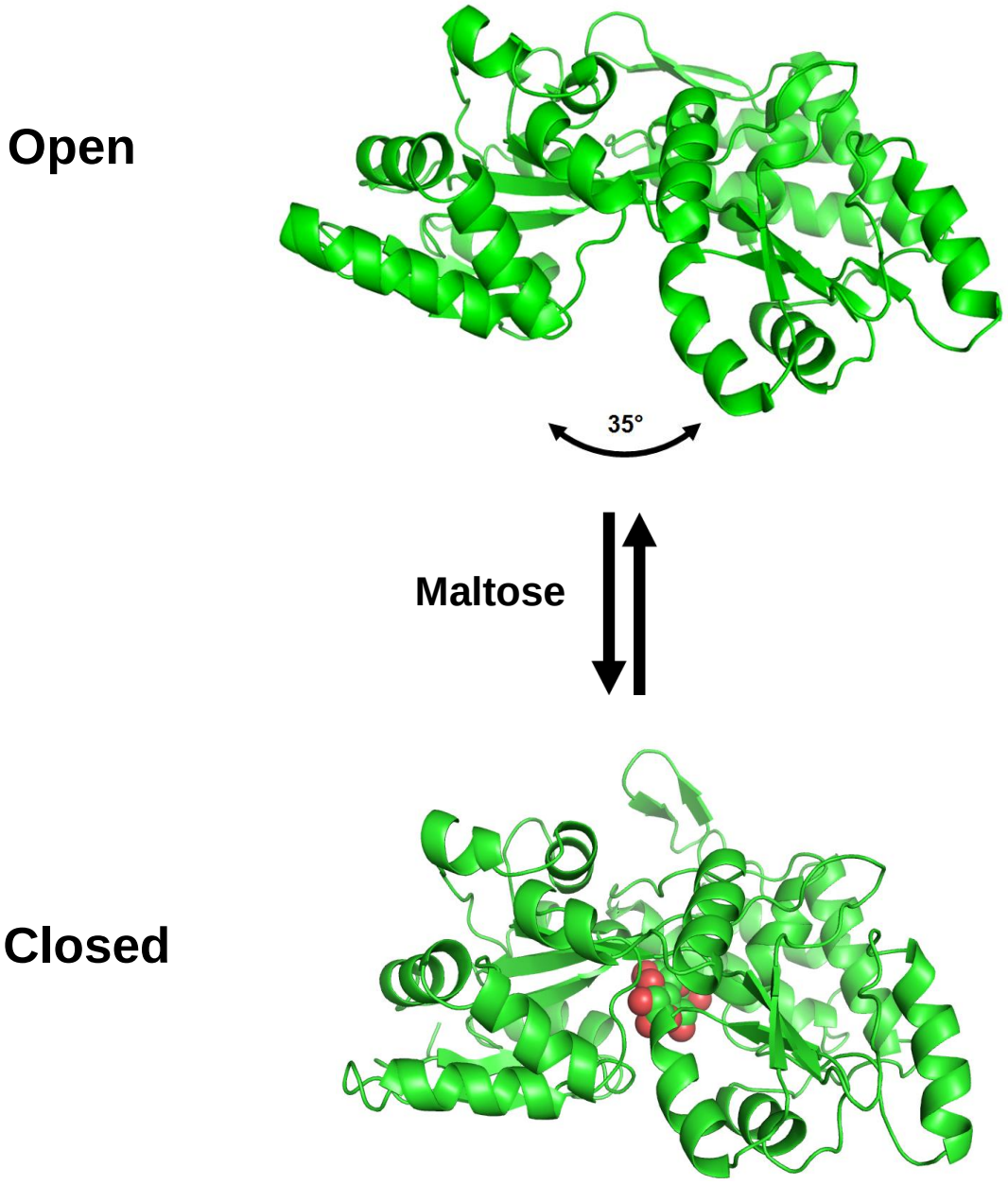


Figure 1.3 Crytal structures of MalE

High resolution structure of MalE in open, unliganded form (PDB accession code 1OMP) and closed maltose bound from (PDB accession 1ANF). The figures were prepared using the program PyMOL (DeLano Scientific).

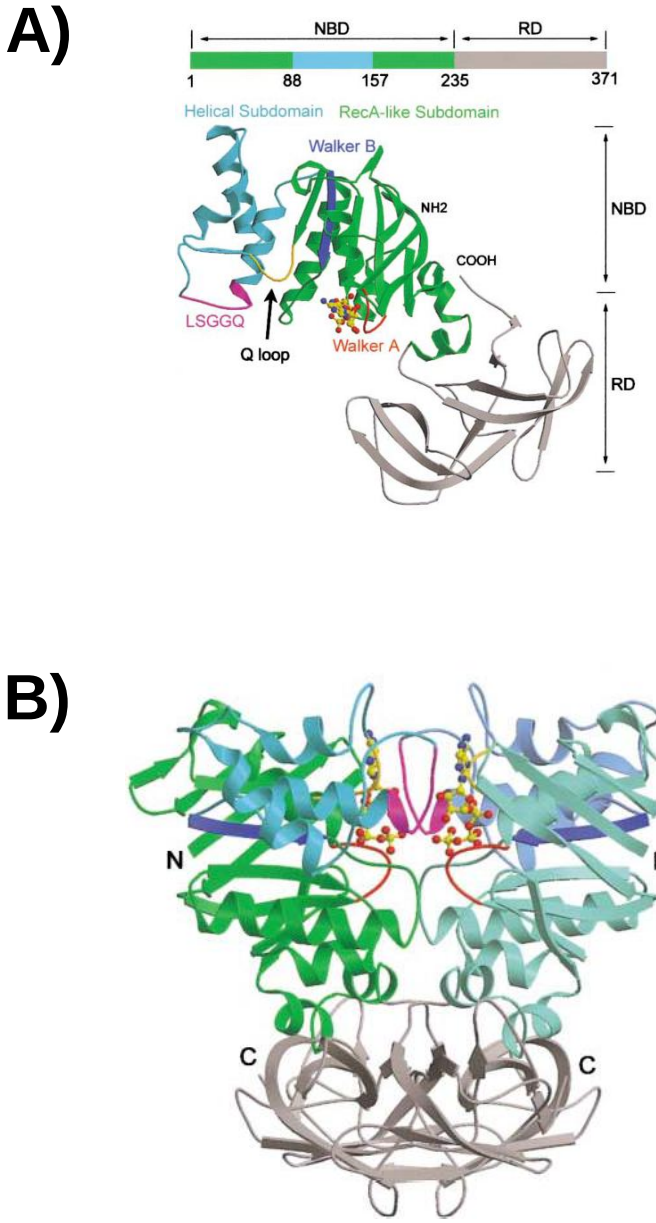


Figure 1.4 The structure of MalK homodimer with bound ATP

A) Crystal structure of a MalK monomer showing the nucleotide binding domain (NBD), the regulatory domain (RD), LSGGQ motif, Q loop and Walker A/B motifs. **B)** Crystal structure of a dimer of MalK with two molecules of ATP bound at the dimeric interface. The ATP is represented in ball-and-stick model (O atom, red; N atom, blue; and P atom, yellow). Both N and C termini of each monomer are labelled. Figure taken from (Chen *et al*., 2003).

MalK from *E. coli* is not a typical ABC transporter NBD/subunit for two reasons. Firstly, it contains an additional C-terminal domain present in MalK homologues, but not in other NBDs/subunits (Oldham *et al*., 2007, Diederichs *et al*., 2000) and secondly, MalK can be purified as a dimer in the presence and absence of nucleotide. This feature of MalK is not shared with other NBDs (Diederichs *et al*., 2000, Kennedy and Traxler, 1999). The structure reveals that MalK is a homodimer and two ATP molecules are bound along the dimeric interface of the nucleotide binding domains and the ATP binding sites are buried(Chen *et al*., 2003) . There are a number of contacts between the two protomers and also a number of contacts between both subunits of the dimer with each ATP molecule. Each ATP molecule interacts with the Walker A motif from one protomers and the LSGGQ motif from the other promoter, explaining the requirement for a dimer.

Mutational and phylogenetic analysis has revealed that MalK, and indeed all NBDs, contain several conserved motifs, some of which are characteristic of all ATPases and some are specific to ABC proteins. The Walker A and Walker B motifs are found in a number of different nucleotide-binding proteins. The Walker B motif is preceded by LSGGQ ("ABC signature sequence") motif (Hyde *et al*., 1990). This motif is found in all NBD subunits in ABC transporters. There is also a conserved glutamine, called the Q loop between Walker A and LSGGQ motifs.

The proposed functions of these motifs and folds have been deduced from the observed phenotypes produced upon mutation of residues within the motifs. This mutational analysis has shown that any variants of the Walker A or Walker B motifs prevent ATP binding and/or hydrolysis and therefore also inhibit transport.

The Q loop has been implicated in signalling between the NBD and the TMDs because mutations in this motif abolished transport, but left nucleotide binding unaffected. The postulate that the Q loop facilitates the interaction of the NBD with the TMD is further strengthened by the production of suppressor mutations in MalK that restore transport activity in response to mutation of a conserved region of the TMD (Hunke *et al*., 2000, Mourez *et al*., 1997). Mutations in the LSGGQ motif have been found to abolish transport, but leave nucleotide binding unaffected indicating that this motif is also involved in signalling between the NBD and TMD.

1.4.3 Transmembrane domains: MalFG

The transmembrane component of the maltose ABC transporter is composed of two integral membrane proteins, MalF and MalG. These proteins make a transmembrane channel for the translocation of substrate delivered by the SBP. MalF and MalG are made up of 8 and 6 transmembrane helices, respectively.

MalF is unusual in that it has an extended periplasmic loop (MalF P2) that reaches around MalE in the protein crystal structure. This interaction with MalE is apparently stable, since this loop, when expressed alone, can associate with MalE in solution (Jacso *et al*., 2009).

Phylogenetic analysis of the amino acid sequence of MalG revealed a conserved motif in a cytoplasmic loop that was present in MalF and other ABC transporter TMDs (Dassa and Hofnung, 1985). The consensus sequence of this motif is $EAAX_3GX_9IXLP$ (where X is any amino acid) and is referred to as the EAA motif. Due to the conserved nature of the EAA motif it was thought essential for functional transport and, because of its cytoplasmic location, was infered to interact with the NBD of MalK. This was later confirmed using site-specific chemical crosslinking (Hunke *et al*., 2000) and by mutating the most highly conserved residues in the EAA motif in MalF and/or MalG (Mourez *et al*., 1997). This latter study revealed that mutation of the highly conserved residues leads to decreased or complete abolition of transport activity. In the mutants where transport was abolished, MalK was no longer membrane associated and was localized in the cytoplasm indicating that the EAA motif forms an interaction with MalK.

1.4.4 The structure of complete $MalEFGK_2$ complex

Following the determination of the crystal structure of *E. coli* vitamin B_{12} transporter BtuCD (Locher *et al*., 2002) a large number of crystal structures of intact ABC transporter complexes have been solved (Rees *et al*., 2009).

The first structure of $MalEFGK_2$ was solved to a resolution of 2.8Å in 2007 (Oldham 2007) (Figure 1.5). In order to obtain the structure of the intact complex, a single mutation was introduced into MalK (E159Q). This mutation permitted binding of ATP to NBD but prevented its hydrolysis.

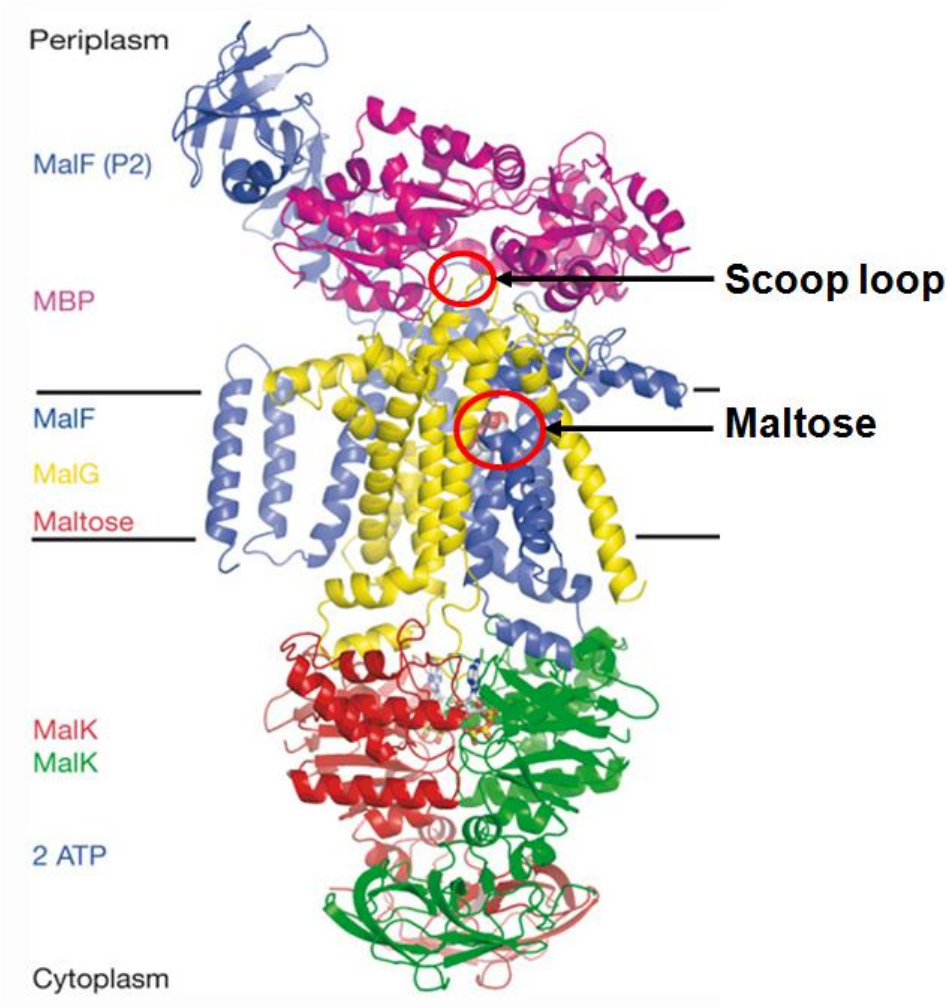


Figure 1.5 Crystal structure of the intact MalEFGK$_2$ complex

Each polypeptide is coloured and labelled accordingly on the left hand side. The structure is of a catalytic intermediate able to bind, but not hydrolyze ATP. In the structure the NBDs are closed and bound to ATP. MalE (MBP) is open and maltose is bound in the translocation channel of MalF and MalG. A loop from MalG "scoop loop" is inserted into the binding pocket of MalE. Figure taken from (Oldham *et al*., 2007)

An overview of the structure reveals that the MalK dimer is essentially identical to the crystal structure of isolated MalK previously obtained (Chen *et al*., 2003). It is a closed dimer with two ATP molecules bound. The only difference between this structure and the structure of $MalK_2$ crystallized on its own is that the Q loops are ordered into *ß*–strand structures. In the structure of $MalK_2$ alone, this region was disordered, but in the structure of the complete transporter the Q loops interact with the TMDs and form an ordered structure. MalF and MalG form the translocation channel. The SBP, MalE is unliganded with both globular domains forming interactions with both MalF and MalG.

As predicted from previous work, the EAA motif of the TMDs is prominent in the interaction face between the MalFG dimer and the MalK dimer. The EAA motifs from both MalF and MalG form 2 small helices and each of these helices is docked into a cleft of corresponding MalK (Figure 1.6A). This MalK cleft consists of the helical subdomain, the Q loop and the helix after the Walker A motif.

The substrate binding pathway is the second region forming important interactions in this structure. This pathway is composed of MalF, MalG and MalE and at the intersection of these three proteins there is a large solvent-filled cavity that extends from the periplasmic side to the mid-point of the lipid bilayer (Figure 1.6B). This cavity is sufficiently large to accommodate the longest $MalEFGK_2$ substrate, maltoheptaose. At the bottom of this cavity, which is at the mid-point of the lipid bilayer, a maltose molecule is observed making a number of contacts with several residues from MalF, but none from MalG. The specific interactions made between protein and substrate, and previous biochemical evidence, suggests that this is a genuine interaction and not an artefact of purification or crystallization (Figure 1.6B). This is the first structural evidence of a binding site in the TMD of an ABC transporter.

At the other end of this cavity, MalE is in its open conformation, covering the top of the translocation channel. There is no maltose in the binding cleft of MalE, therefore binding to MalFG presumably stabilizes the open conformation allowing release of maltose into the MalFG binding pocket. One of the most striking features from this structure is the presence of a periplasmic loop (P3) from MalG in the binding cleft of MalE (Figure 1.5). This so-called “scoop loop” either dislodges maltose from MalE and/or sterically hinders its re-entry therefore facilitating unidirectional transport.

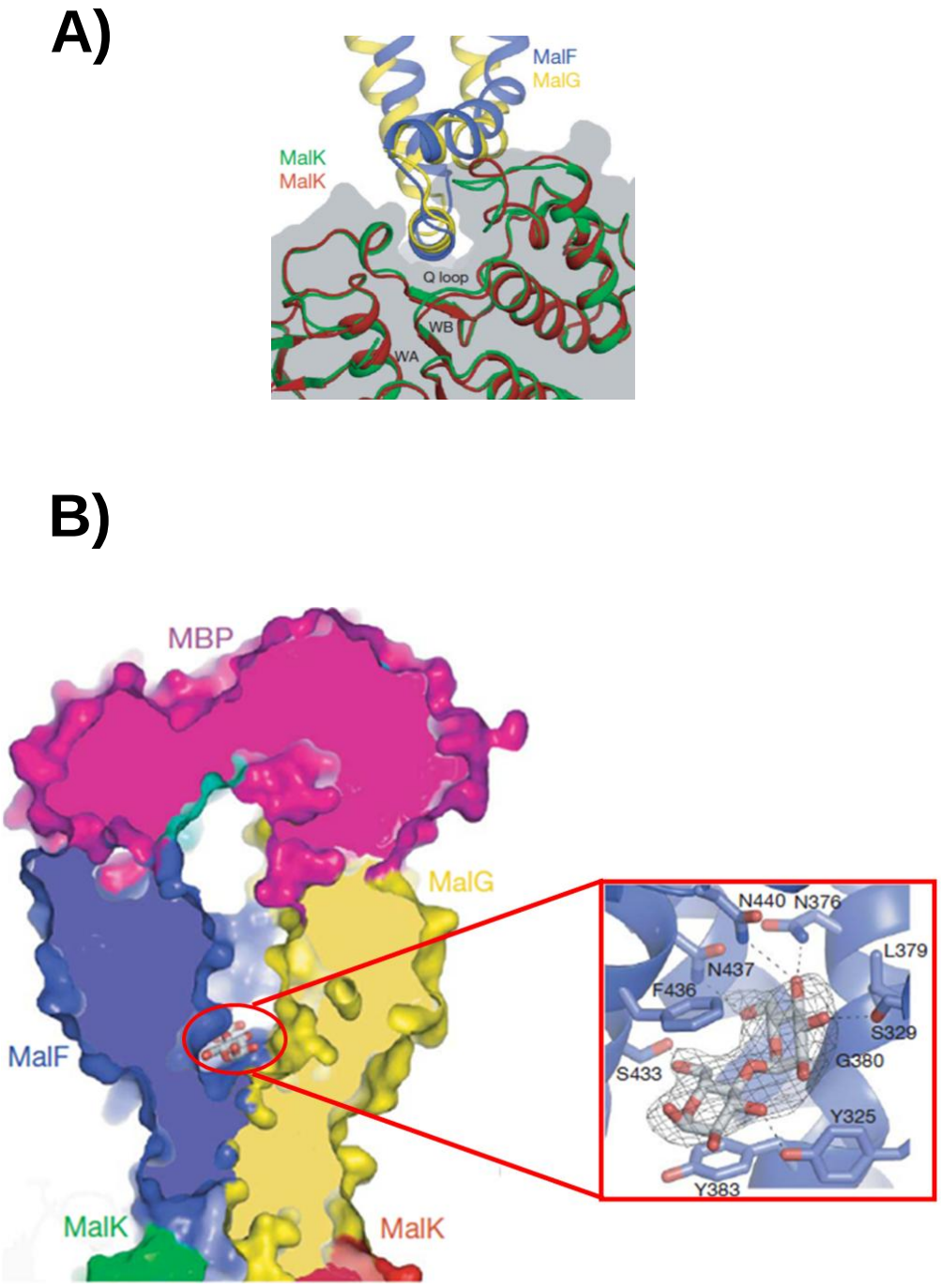


Figure 1.6 Domain interactions in the MalEFGK$_2$ complex

A) The conserved EAA loop of MalF and MalG docking with a surface cleft in MalK. Positions of the EAA motif from MalF (blue) and MalG (yellow) are compared by superposing the MalK monomers (red and green). The Q loop, Walker A (W/A) and Walker B (W/B) are shown. **B)** Cross-section of a space-filling model showing the open conformation of MalE, the solvent filled cavity and the positioning of maltose at the bottom. The maltose binding site within the TMD is enlarged showing interactions between maltose and amino acids.

The structure of $MalEFGK_2$ described above is a catalytic intermediate of the maltose transporter in the outward-facing conformation with a TM maltose binding site exposed to the periplasm. Later on, another structure of the complex was obtained in the absence of maltose binding protein (Khare *et al*., 2009). Analysis of this structure shows that in the absence of MalE, $MalFGK_2$ forms an inward-facing conformation with the TM maltose-binding site exposed to the cytoplasm. Comparison of this structure to that of outward-facing conformation reveals that there is a rigid body rotation of the TMDs which is coupled to the opening of NBD dimer interface in the inward-facing conformation of $MalFGK_2$.

The inward-facing conformation represents the resting state where the transporter has a very low ATPase activity and the outward-facing conformation represents a catalytic intermediate where ATP is ready for hydrolysis. More recently, another structure of $MalEFGK_2$ has been solved where MalE interacts with the resting state conformation to give a pre-translocation complex (Oldham and Chen, 2011). This complex represents an intermediate state between inward and outward-facing conformations (Figure 1.7A). Analysis of this structure reveals that MalE is present in the closed liganded conformation and two lobes of MalE interact with MalF and MalG. The interface between MalE and MalFG is smaller than observed in the open outward-facing complex. A maltose molecule was also identified in the TM site of this structure which is likely nonphysiological and an artefact of crystallization. Access to the TM binding site was restricted from both sides of the membrane. The MalK dimer is semi-open as the distance between the two NBDs is between those of open and closed forms (Figure 1.7B).

A key feature of this structure is that the conserved D-loops are present at the MalK dimer interface making contacts to the residues of Walker A of the opposite subunit. Walker A residues interact with the gamma phosphate of ATP in the closed dimer form of outward-facing state. Thus the binding of ATP to MalK of the pre-translocation state causes local conformational changes of Walker A and the D loop and initiates progression towards the outward-facing conformation. In contrast, in the inward-facing resting state, where MalE is absent, the Walker A and D loops are distant from the dimer interface (Figure 1.7B); thus, ATP binding would not induce such a global structural rearrangement necessary for dimerization of MalK and ATP hydrolysis. In

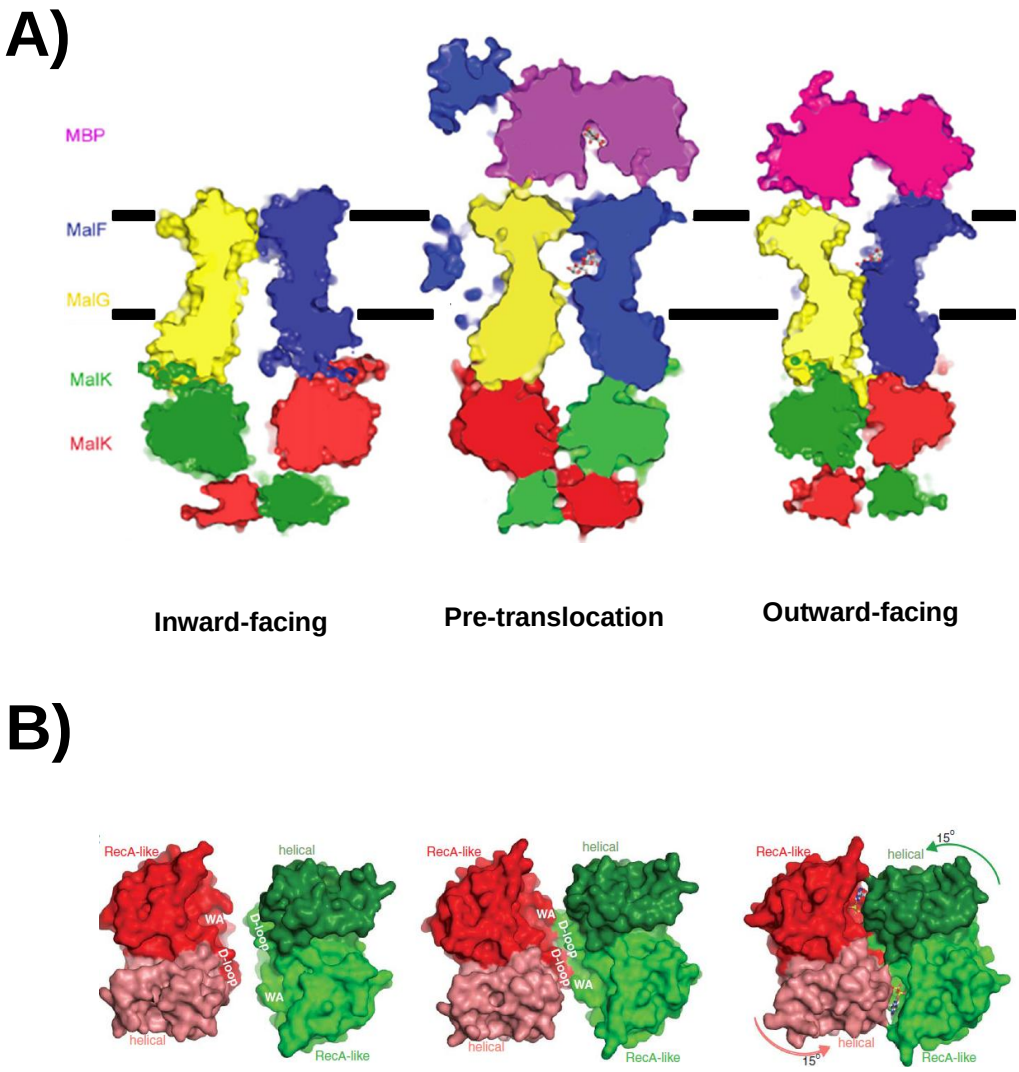


Figure 1.7 Pre-translocation state is an intermediate conformation between inward and outward facing states

A) Surface slab view of three different states of the transport cycle. Each component of the complex is colour coded and labelled accordingly on the left hand side. **B)** Surface representations of the nucleotide-binding domains (NBDs) in the three states. The RecA-like and the helical subdomains are differentiated by color shade. The D loop and theWalker A (WA) motifs are labeled. The rotation of the helical subdomain with respect to the RecA-like domain during transition from the pretranslocation state to the outward-facing conformation is indicated. Figure adapted from (Khare *et al*., 2009, Oldham and Chen, 2011).

pre-T state MalE induces rotation of TMDs and partial closure of the MalK dimer and lowers the energy barrier for closure of the MalK dimer necessary for ATP hydrolysis.

1.4.5 Proposed mechanism of transport of maltose

The wealth of information available from previous biochemical and genetic data and recent biophysical data has allowed proposing a model for the transport of maltose by $MalEFGK_2$ (Figure 1.8).

Once maltose diffuses into the periplasm through the outer membrane channel, it binds MalE and stabilizes it. The ligand bound MalE interacts with the inward-open state of $MalFGK_2$, in which the TM channel is closed to periplasm and the NBD is in the open form. Binding of MalE to $MalFGK_2$ induces conformational changes in the TMD which in turn brings the NBDs closer and the complex is now in a pre-translocation state. Binding of ATP to MalK promotes a concerted motion of MalK closure, reorientation of the TM subunits and opening of MalE. This results in the formation of the outward-facing conformation and transfers maltose from MalE to the TM binding site. This is the time for ATP hydrolysis to occur. Although the crystal structure of the post hydrolysis state of the complete transporter is lacking, analysis of the isolated NBD structures suggest that ADP is not sufficient to stabilize the closed dimer conformation (Lu *et al*., 2005). Thus once ATP is hydrolyzed, the TMDs will likely reorient towards the cytoplasm and the substrate will be released into the cytoplasm. MalE is released, $MalFGK_2$ reverts to the resting state and the transport cycle is ready to begin again.

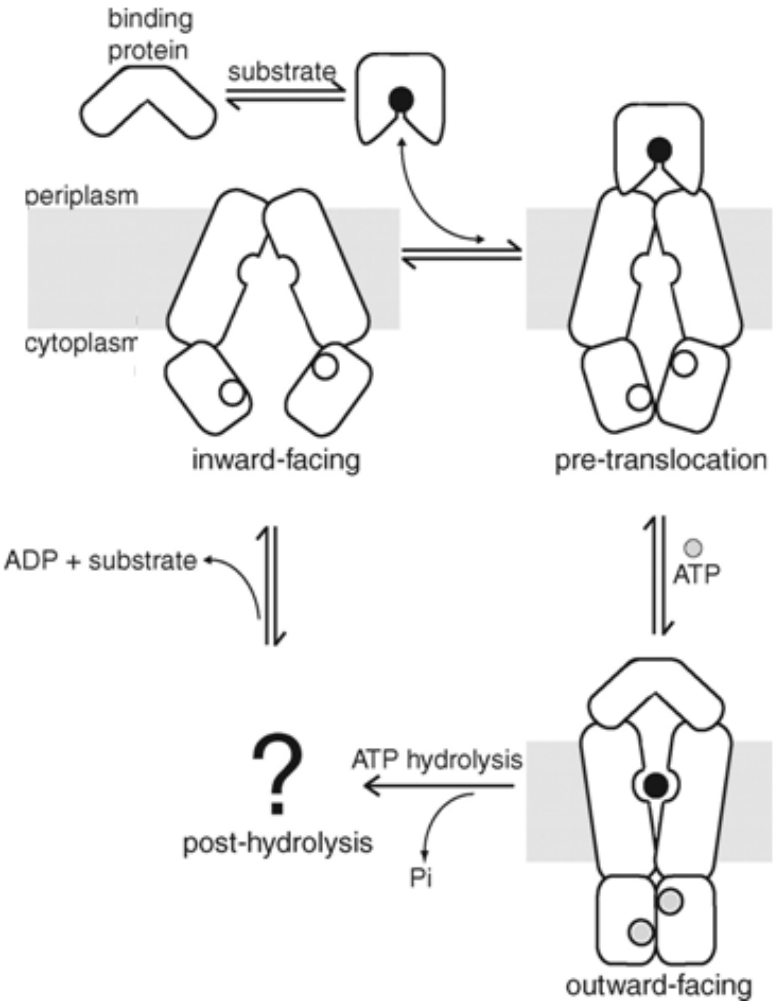


Figure 1.8 Proposed mechanism for transport of maltose by MalEFGK$_2$

Maltose binds to the SBP and interacts with TMD and NBD complex at inward-facing state (resting state). Interaction of liganded SBP triggers the conformational changes within the TMD and NBD bringing NBDs closer (pre-translocation). Binding of ATP to NBDs causes dimerization of NBDs and reorientation of TMDs resulting the opening of the periplasmic side of TMDs and release of substrate from SBP. The hydrolysis of ATP occurs leading to opening of NBDs and subsequent opening of the cytoplasmic side of the TMD releasing substrate into the cytoplasm. Figure adapted from (Oldham and Chen, 2011).

1.5 Substrate-binding Proteins

Substrate-binding proteins (SBPs) are a group of proteins that are often associated with membrane protein complexes for transport of substrate. They have also been shown to perform a receptor function in chemoreception and signal transduction (Tam and Saier, 1993, Neiditch *et al*., 2006). The SBPs were originally found to be associated with ABC transporters (Holland, 2003), but recently they have been shown to be associated with other membrane protein complexes, such as tripartite ATP independent periplasmic (TRAP)-transporters (Mulligan *et al*., 2009).

The SBP is the most variable of the subunits of ABC transporters and this is due to the fact that this subunit binds the ligand directly and therefore confers specificity on the system. As stated ABC transporters that contain an SBP normally function as substrate importers while those systems that lack the SBP tend to function in substrate export and this fact can almost be used as a rule in defining the two major classes of ABC systems (Higgins and Linton, 2001).

The SBPs are found within the periplasm of Gram-negative bacteria, often in large excess over the TMD and NBD subunits, with the SBPs found at around 0.1-1 mM concentration within the periplasm (Holland, 2003). In Gram-positive bacteria and Archaea, organisms without a periplasm, the SBPs are anchored to the outer surface of cell membrane via an N-terminal lipid moiety (Sutcliffe and Russell, 1995) or an N-terminal TM segment in Archaea (Albers *et al*., 1999).

In addition to the SBP dependent ABC transporters, a new family of prokaryotic ABC transporters has been identified that does not depend upon SBP for the uptake of ligand. This is called energy coupling factor transporter (Rodionov *et al*., 2006, Rodionov *et al*., 2009). Here, an integral membrane protein binds the substrate with high affinity (nM) and translocates it into the cytoplasm (Zhang *et al*., 2010).

In the last decade new membrane protein complexes were discovered that contained a SBP. This family of transporters was called tripartite ATP-independent periplasmic (TRAP)-transporters (Kelly and Thomas, 2001). They are secondary transporters but rely on SBPs to import their ligands. The membrane component of these proteins usually consists of one larger and one smaller subunit. The transport of substrate is driven by an electrochemical ion gradient (H^+ or Na^+ gradient) (Mulligan *et al*., 2009).

Many SBPs have been characterized in the last fifty years and shown to bind a diverse variety of ligands including amino acids, sugars, peptides, vitamins and metal ion or

their chelates (Magnusson *et al*., 2004, Tame *et al*., 1995, Borths *et al*., 2002, Clarke *et al*., 2000, Heddle *et al*., 2003). The size of the ligands accommodated by SBPs range from a single ion such as Zn^{2+} (Bound by TroA) with a volume of 1.5 $Å^3$ to large peptides of up to 26 amino acids long with the volume up to 4900 $Å^3$ (Bound in OppA) (Lee *et al*., 1999, Levdikov *et al*., 2005, Berntsson *et al*., 2009).

In bacteria, ABC transporters are generally encoded in operons containing all the components needed for the transport of substrates. These components include a single gene encoding the SBP, genes encoding TM domains and NBDs which bind and hydrolyze ATP. The maltose ABC transporter from *E. coli* is an example of this type of transporter. However, there are examples of SBPs which are encoded by genes which are not part of an operon and are present at remote locations on the genome. They are called orphan SBPs (Thomas, 2010). An example of orphan SBP from the model bacterium *E. coli* K-12 is MppA that is homologous to the oligopeptide SBP OppA which recognizes the murein tripeptide and functions with TMD/NBD proteins (OppBCDF) of the oligopeptide transport system (Park *et al*., 1998). The orphan SBPs are widespread among different groups of bacteria: for example, a quick look in TransportDB (Ren *et al*., 2006) shows that *Enterococcus faecalis* V583 has 10 genes encoding orphan SBPs suggesting that bacteria use SBP dependent transporters to increase substrate range at minimal cost of energy.

1.5.1 Overall structures of SBPs and their classification

The first SBP crystal structure was solved in 1974 for L-arabinose binding protein (Quiocho *et al*., 1974). Since then, many more have been solved revealing a wealth of information about this family of proteins. The core of all SBPs consists of two structurally conserved domains (Quiocho and Ledvina, 1996), although some members contain an extra domain as an exception from the common rule (Tame *et al*., 1994). The two domains are connected by a hinge region, with the ligand binding site buried in between two domains usually in a manner that sequesters them completely from the solvent (Holland, 2003). In the absence of ligand, the protein is normally flexible, with the two domains rotating around the hinge and exists largely in the open conformation. Upon substrate binding, the closed conformation is stabilized and the ligand is trapped between the domains. This process is called the "Venus Fly-Trap" mechanism (Mao *et al*., 1982).

Based upon the available structural information, SBPs have been observed in four distinct structural states (Figure 1.9); open unliganded, open liganded, closed unliganded and closed liganded (Berntsson *et al*., 2010). As stated above, SBPs mainly exist in the open-unliganded state in the absence of substrate, with possibly a small fraction in a closed-unliganded state (Bermejo *et al*., 2010). In the presence of a substrate, the equilibrium between open and closed conformation shifts towards the closed-liganded state.

Crystal structures of SBPs have shown that they generally act as monomers ranging in size from 25 to 60 kDa (Holland, 2003). However, recently it was shown that TakP, a SBP from a tripartite ATP-independent transporter, is a dimer (Gonin *et al*., 2007). Another example of dimeric SBP is of the iron-binding protein FitE (Shi *et al*., 2009). It is clear that these proteins dimerize in solution, but whether these dimers are of physiological relevance remains unclear.

In the past 20 years several groups have performed phylogenetic analysis of the SBPs from specific organisms or more widely with the aim of creating subgroups or classes that bind to specific ligand or types of ligand (Saurin and Dassa, 1994, Tam and Saier, 1993). In 1999 Fukami-Kobayashi classified SBPs into two different classes based on the connectivity of secondary structure elements and especially the topology of the beta-sheets core in the domains (Fukami-Kobayashi *et al*., 1999). Examples of proteins belonging to Class I are the leucine/isoleucine/valine-binding protein (LIVBP) and of class II are the histidine-binding protein (HisJ) and the oligopeptide-binding protein (OppA). Later on, a third class of SBPs was defined, with proteins containing B12 binding protein and Zinc binding protein. The size of SBPs does not correlate with the class to which it belongs, e.g., both OppA (65 kDa) and HisJ (27 kDa) fall in Class II.

For the purposes of this thesis the classification of Ronnie Berntsson was chosen (Berntsson *et al*., 2010). In this classification scheme, SBPs were grouped into six clusters on the basis of their structural similarity.

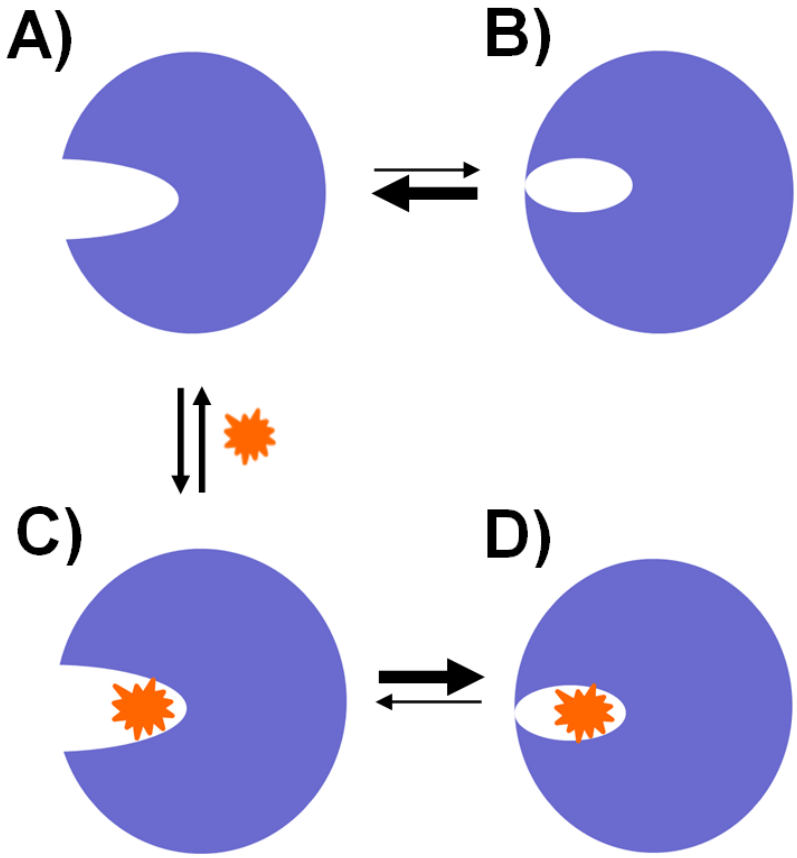


Figure 1.9 Different structural states of the monomeric SBPs

A) Open-unliganded **B)** Closed-unliganded **C)** Open-liganded and **D)** Closed-liganded. SBPs are represented by blue circles and ligand is represented by red star. In the absence of a substrate, the equilibrium is towards the open-unliganded conformation. Upon binding of substrate, the open-liganded conformation is formed and the equilibrium is shifted to closed liganded.

Cluster A consists of SBPs that play a role in metal binding. The distinguishing characteristic of Cluster A is an α-helix serving as the hinge between the two domains (Figure 1.10A). This helix ensures a rigid overall structure. An example of this group is BtuF, the SBP of vitamin B_{12} ABC importer from *Escherichia coli*. This cluster is further divided into into two sub clusters (A-I and A-II) based on their cognate ligands. In A-I the SBPs bind metal ions (iron, zinc, or manganese) via direct interactions with the metal ion and in A-II they bind chelated metals by siderophores like enterobactin, catecholate and hydroxamate.

Cluster B consists of SBPs that bind mainly carbohydrates such as ribose, glucose and arabinose but also amino acids. The SBPs of this cluster interact with ABC-transporters and the receptors involved in signal transduction. The SBPs of this cluster contain three hinges between the two domains (Figure 1.10B) as shown in the structure of ribose binding protein (RBP) (Figure 1.10B).

The SBPs belonging to cluster C bind very different ligands, such as di- and oligopeptides, arginine, nickel ions and cellobiose. A characteristic feature of SBPs in this cluster is their large size ranging from 55 to 70 kDa. This large size is due to presence of an extra domain in these proteins (Figure 1.10C).

Cluster D contains SBPs which are characterized by a unique hinge region consisting of two short strands, 4-5 amino acids long (Figure 1.10D). This group of proteins binds to a large variety of substrates; carbohydrates, thiamine, tetrahedral oxyanion and iron etc.

Cluster E consists of SBPs which are part of TRAP transporters. The unique feature of this class of SBPs is a long helix that spans both domains and an additional β-strand connecting both domains (Müller *et al*., 2006) (Figure 1.10E).

Cluster F contain SBPs, with a distinguishing feature in the form of a hinge containing domains I and II. However, the hinges of the cluster F proteins are significantly longer, 8–10 amino acids (Figure 1.10F) than the hinges of 4–5 amino acids typically observed in SBPs of cluster D. Cluster F SBPs bind a large variety of substrates ranging from trigonal planar anions (nitrate and bicarbonate) to amino acids and compatible solutes such as glycine and betaine.

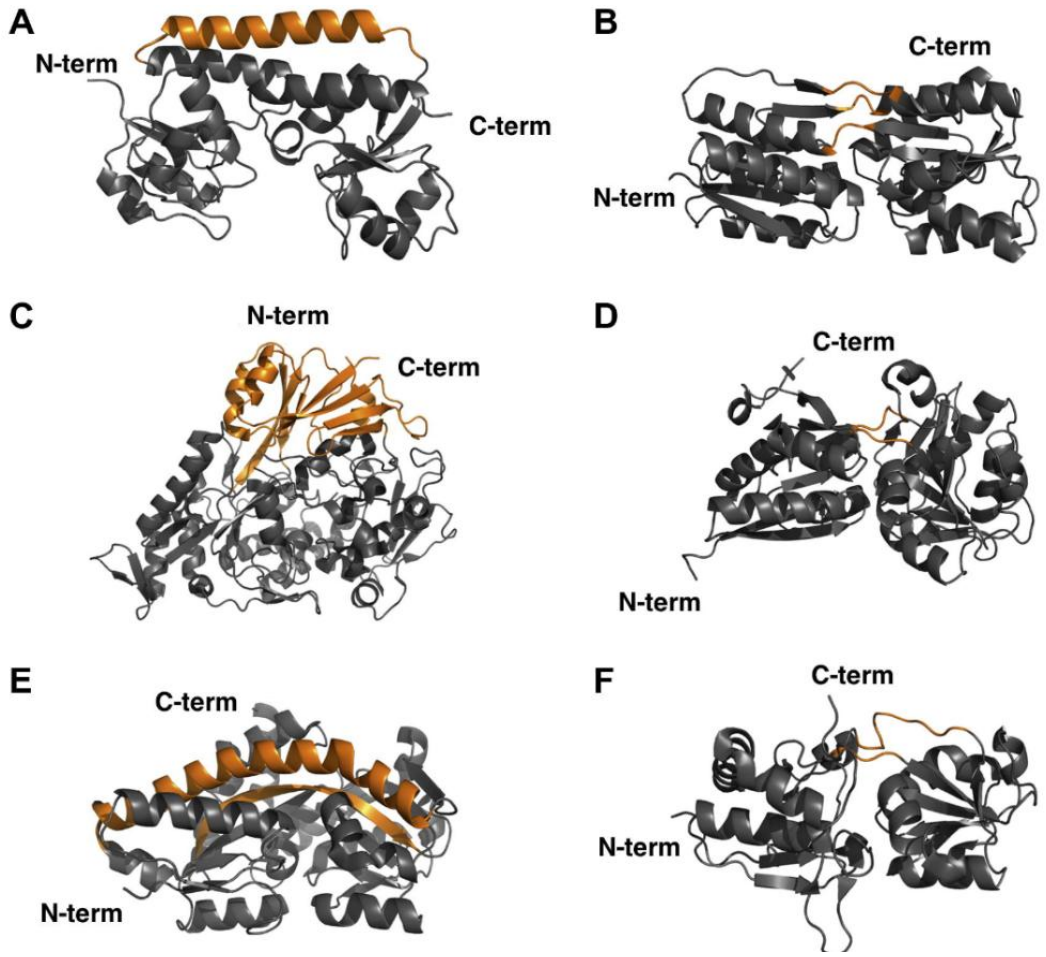


Figure 1.10 The different clusters of SBPs are shown with their distinct structural feature colored in orange.

(A) Cluster A contains proteins having a single connection between the two domains in the form of a rigid helix. **(B)** Cluster B contains SBPs with three interconnecting segments between the two domains. **(C)** Cluster C contains SBPs that have an extra domain and are significantly larger in size when compared with the others. **(D)** Cluster D contains SBPs with two relative short hinges. **(E)** Cluster E contains SBPs associated with TRAP-transporters which all contain a large helix functioning as hinge region. **(F)** Cluster F contains SBPs with two hinges similar like cluster D, however, these hinges have almost double the length creating more flexibility inside the SBP. The proteins used to illustrate the features in clusters A–F are BtuF (PDB code: 1N2Z), RBP (PDB code: 1DRJ), OppA (PDB code: 3DRF), ModA (PDB code: 1ONR), UehA (PDB code: 3FXB) and HisJ (PDB code: 1HSL), respectively (Taken from Berntsson *et al* 2010)

1.6 Peptide transport

Peptides transport refers to the uptake of peptides of a chain length of two or more amino acids. Peptide transporters are integral membrane components that mediate cellular uptake of peptides. They are found in bacteria, yeast, plants, invertebrates and vertebrates (Wilkinson, 1996). Peptides are widespread in nature and serve a variety of functions, for example they serve as a source of nutrients for a cell and in microorganisms they are also involved in various cellular signaling processes like sporulation, DNA transfer by conjugation and competence development (Lazazzera, 2001)

Free-living bacteria have evolved multiple routes for the uptake of peptides from environment (Detmers *et al*., 2001). High-affinity bacterial transporters for di-, tri-, tetra-, and oligopeptides are well studied in bacteria and allow them to take up any peptide of a particular length (Detmers *et al*., 2001, Doeven *et al*., 2005). The high-affinity peptide transporters are all members of the ATP-binding cassette (ABC) transporter family and utilize a substrate-binding protein, (Guyer *et al*., 1986) which defines the substrate specificity of the transporter.

Escherichia coli has two well characterized high-affinity peptide transporters with broad specificity for substrates: the oligopeptide system (Opp) that recognizes peptides between 2 and 5 amino acid residues in length, irrespective of their amino acid sequence, with highest affinity for tripeptides (Guyer *et al*., 1986) and the Dpp system, which has a preference for dipeptides (Smith *et al*., 1999) (Figure 1.11).

The sequencing and annotation of the *E. coli* genome revealed the presence of a number of orthologues of the *oppA* and *dppA* genes on the chromosome including a number of orphan SBP coding sequences. Genetic data suggest that the orphan SBP, MppA, which is 46% identical to OppA, functions with the OppBCDF ABC transporter (Figure 1.11) (Park *et al*., 1998). The *mppA* gene was identified in 1998 as an additional locus in *E. coli* that was required for growth on the cell wall-derived murein tripeptide (L-Ala-γ-d-Glu-*meso*-Dap, where *meso*-Dap is *meso*-diaminopimelate) as a source of Dap for

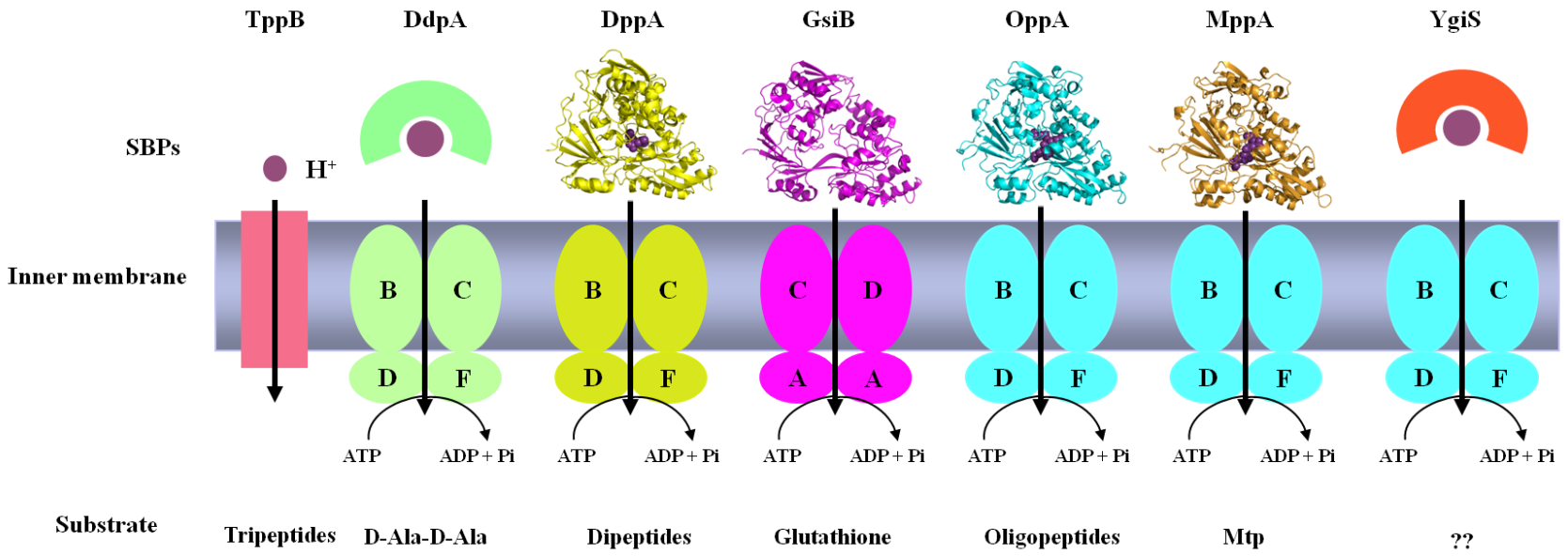


Figure 1.11 Schematic figure of the known peptide transporters present in *E. coli* K-12.

This bacterium has a wide range of peptide transporters, mainly high-affinity SBP-dependent ABC systems, with a single secondary carrier TppB for tripeptides. The complete ABC transporters identified are a putative stationary phase-induced dipeptide permease (DdpABCDF), the dipeptide permease (DppABCDF), the glutathione-specific permease (GsiABCD), and the oligopeptide permease (OppABCDF) that can accept peptides from 2 to 5 residues in length. In ABC transporters, the SBP captures peptides from the periplasm and delivers them to membrane components for translocation. DdpA, DppA, GsiB, and OppA are the SBPs that work in conjunction with their cognate membrane components, whereas MppA and YgiS are orphan SBPs and are thought to use the OppBCDF permease components for solute translocation (Park *et al.*, 1998).

peptidoglycan biosynthesis (Park *et al*., 1998). This MppA/OppBCDF transporter presumably allows *E. coli* to recycle murein tripeptide (Mtp) from either exogenous sources or Mtp that is released by the AmiD outer membrane amidase into the periplasm during cell growth (Park and Uehara, 2008). YgiS is another putative peptide-binding orphan SBP of unknown specificity (Figure 1.11) which has 44% amino acid sequence identity to OppA.

E. coli also contains a third characterized peptide ABC transporter, the Ddp system, which is thought to have evolved from the general dipeptide (Dpp) transporter system to specifically recognize D-Ala-D-Ala, another component of the peptidoglycan that can be recycled by the cell (Figure 1.11) (Lessard *et al*., 1998).

Finally, a glutathione-specific transporter of the oligopeptide permease family has been characterized genetically in *E. coli*, the GsiABCD system (Suzuki *et al*., 2005) (Figure 1.11). Glutathione is a tripeptide (γ-glutamylcysteinylglycine) which is taken up by bacteria as part of the cysteine salvage pathway. Genetic studies show that transport of glutathione was not observed when the gene encoding the glutathione binding protein (GsiB) was deleted.

Therefore it is clear that the cluster C oligopeptide SBP family includes proteins that recognize general peptides, as well as others that have evolved to recognize specific peptides. The SBPs recognizing specific peptides are not restricted to Gram-negative bacteria. There is genetic evidence in the literature about specific peptide-binding proteins from Gram-positive bacteria too. It is worth mentioning that in a Gram-negative bacterium like *E. coli* there exists an outer membrane that limits the size of peptides that can be taken up. Peptides up to five or six amino acids long can freely diffuse across the outer membrane (Nikaido, 2003) of Gram-negative bacteria. In Gram-positive bacteria there is no such restriction and peptides up to 35 amino acids long are transported (Doeven *et al*., 2004).

Conjugation is the transfer of certain plasmids from donor cells to recipient cells. Pheromone dependent conjugation is an interesting phenomenon studied in *Enterococcus faecalis* (Leonard *et al*., 1996). Pheromones are small hydrophobic peptides of seven or eight amino acids secreted by plasmid free recipient cells. Each of these pheromones triggers the transfer of a specific plasmid from a donor cell to a recipient cell. The pheromones cPD1, cAD1, and cCF10 are specific for plasmids

pPD1, pAD1 and pCF10 respectively and are recognised by a cell surface binding protein (designated TraC for pPD1 and pAD1, and PrgZ for pCF10) (Chandler and Dunny, 2004). Studies conducted on pCF10 show that import of peptide into donor cells depends upon the concerted action of the membrane components of the Opp system and PrgZ (Leonard *et al*., 1996). Mutation of PrgZ decreased the sensitivity of donor cells to pheromone. In mutants without PrgZ, but with a functional Opp system, signalling still occurs but higher concentration of the pheromone is necessary. This indicates that OppA is able to bind the pheromone but with lower affinity than PrgZ. Both PrgZ and TraC are orphan SBPs and likely use membrane components of *E. faecalis* Opp system. Both PrgZ and TraC proteins exhibit a very high binding affinity, the dissociation constants for peptides are in the nanomolar range. PrgZ and TraC are highly similar but they seem to be dedicated to specific hydrophobic peptides. On the other hand, the general binding protein (OppA) of the Opp systems have dissociation constants in the micromolar range for a wide variety of peptides, differing in length and amino acid composition. This correlates with the function of these peptide-binding proteins. OppA plays a role in nutrition and should be able to accept peptides irrespective of their sequence and composition, whereas the binding proteins involved in conjugation should only respond to a specific signal (Detmers *et al*., 2001).

1.6.1 The structural basis of sequence independent peptide binding

The peptide-binding SBPs of ABC transporters determine the substrate specificity of the system. A wealth of structural data is now available for these peptide-binding SBPs. The first structure of a peptide-binding SBP (OppA of *S.* Typhimurium) was solved by Prof. Tony Wilkinson and colleagues in 1994 (Tame *et al*., 1994). The structures of two liganded forms of OppA were determined, the first containing trilysine (KKK) and the second containing a co-purified (presumably heterogenous) peptide ligand. OppA has high affinity for its ligands and as a result co-purifies with bound peptides, a phenomenon referred as the retention effect (Guyer *et al.*, 1986, Silhavy *et al.*, 1975).

A unique feature of the overall structure of liganded forms of OppA was the organization of the polypeptide chain into three domains instead of two (Figure 1.12). Two of the domains (I and II) are structurally analogous to those present in other SBPs (Figure 1.10). Domain III has no counterpart among other SBPs and its function is not clear.

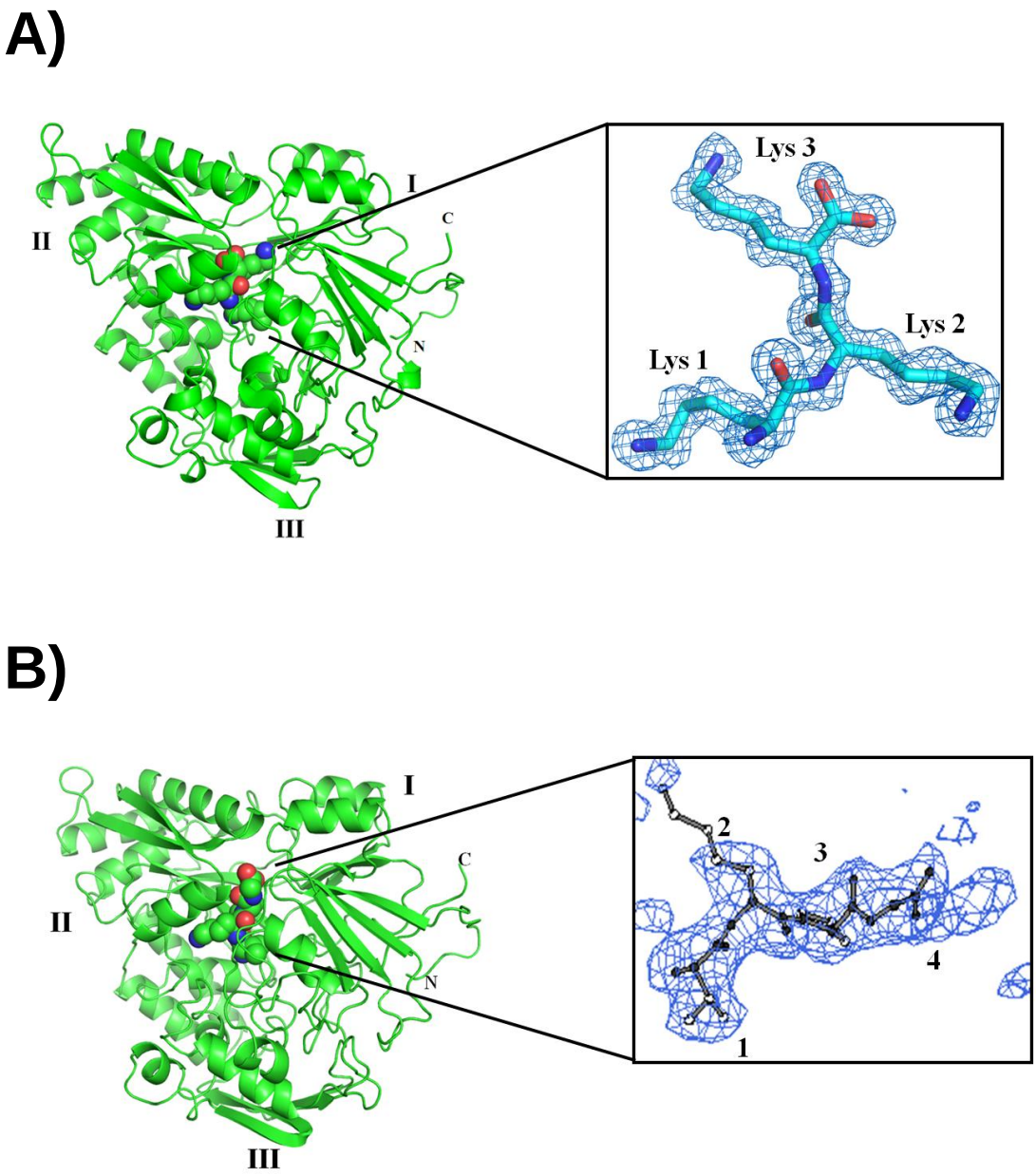


Figure 1.12 Crystal structure of OppA in complex with peptide ligands

A) Crystal structure of OppA in complex with trilysine (KKK). Domain I, II and III are labeled. The electron density of the ligand ($2F_o$-F_c map) is displayed as blue mesh (right) and contoured at 1 σ. Image is drawn in PyMol using the coordinates from PDB (2OLB). B) Crystal structure of OppA in complex with an undefined mixture of co-purified peptides. The electron density of the ligand ($2F_o$-F_c map) is displayed as blue mesh (right) and contoured at 1 σ. A tetrapeptide (VKPG) is modeled into the density.

The structure of OppA in complex with a tripeptide (KKK) shows a very clear electron density for the tripeptide in the binding pocket (Figure 1.12A). The peptide adopts an extended conformation. Tight peptide binding is mainly due to main-chain to main-chain contacts. The main-chain interactions between the protein and a tripeptide are shown in Figure 1.13. The charges at the N and C termini of the peptide are countered in the complex by oppositely charged side chains, Asp419 and Arg413.

The structure of OppA in complex with endogenously co-purified ligand also shows extra electron density in the binding pocket of OppA, which could not be accounted for by protein atoms (Figure 1.12B). However the additional electron density was not well defined and it was not possible to model a peptide that fits well into the electron density. The true sequence of the peptide ligand remains unidentified in the structure and it was attempted to model the tetrapeptide (VKPG) into the electron density (Figure 1.12). The ambiguity in defining the electron density was perhaps due to heterogeneous population of peptides that were co-purified with OppA. This feature is consistent with the structures of all peptide-binding SBPs that are thought to bind a large variety of peptides (Berntsson *et al*., 2009, Levdikov *et al*., 2005, Klepsch *et al*., 2011).

Later on, structure of OppA was solved in complex with tripeptides of varying amino acid sequences (Sleigh *et al*., 1999) proving the versatility of OppA in terms of accommodating various peptides irrespective of their amino acid sequence. Crystallographic analysis of the structures showed that the side chains of the amino acids are accommodated in large hydrated pockets. These pockets can accept any side chain, which is the basis of lack of specificity for the ligand's amino acid sequence. The hydrogen bonds and salt bridges formed by the peptide backbone alone can drive binding and there is only a small contribution of the amino acid side chains to the binding affinity.

Specificity for peptides of a particular length is determined by salt bridges that constrain the positions of their N and C termini in the binding site. This is consistent with the architecture of Gram-negative bacteria where an outer membrane restricts the size of peptides that can diffuse into the periplasm.

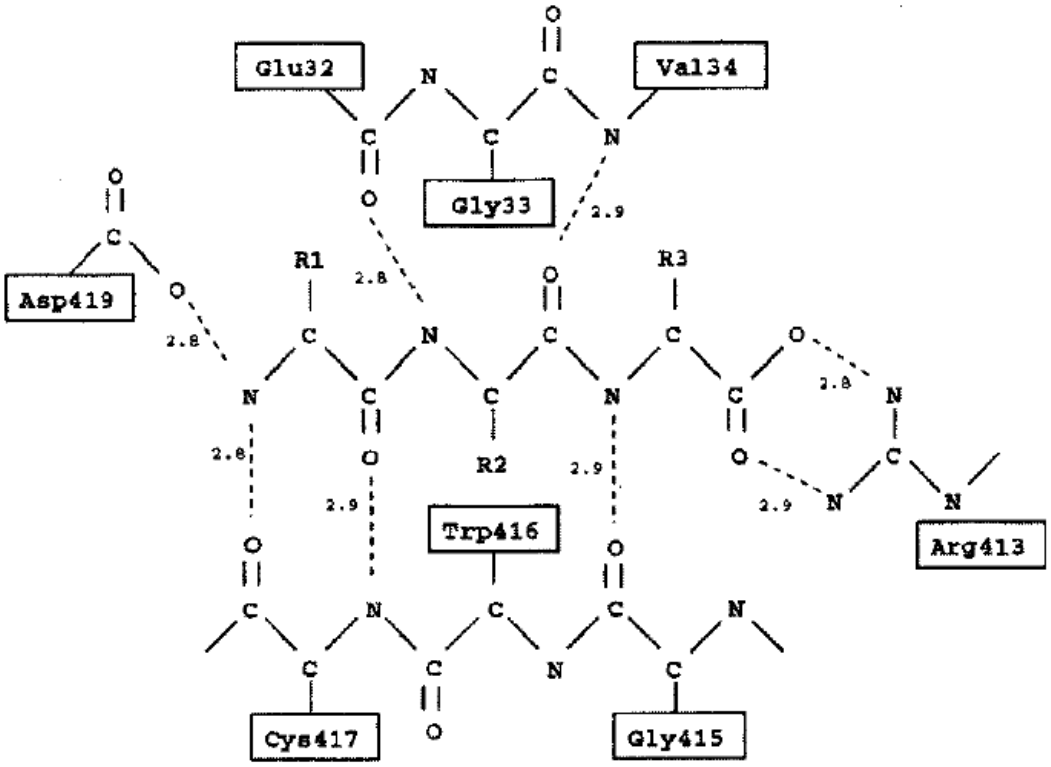


Figure 1.13 Schematic diagram illustrating the interactions made by the main chain of the tri-lysine ligand with OppA

Protein residues are labelled. Hydrogen bonding and electrostatic interactions are indicated by the dotted lines. R1, R2 and R3 indicate the ligand side chains. Figure taken from Tame *et al* 1995.

In Gram-positive bacteria there is no outer membrane and thus large peptides are accessible to peptide-binding SBPs. A crystal structure of a peptide-binding SBP (AppA) from *Bacillus subtilis* has shown that it can accommodate a nonapeptide (Levdikov *et al*., 2005). Similarly the Opp system of *Lactococcus lactis* has been shown to bind and transport peptides as long as 35 residues of any composition though it has a clear preference for nonapeptide (Doeven *et al*., 2004). The basis of binding larger peptides has recently been shown by solving the crystal structure of OppA from *L. lactis*. A salient feature of the structure is the enormous volume (4900Å^3) of its ligand binding cavity (Berntsson *et al*., 2009). The volume of cavity is bigger than those of AppA (2600Å^3) and OppA (1000Å^3) from *B. subtilis* and *S.* Typhimurium, respectively. This greatly expanded cavity is created due to structural rearrangements of the loops and thus responsible for accommodating larger peptides.

1.7 Peptidoglycan

Peptidoglycan (or murein) is the major structural component of bacterial cell walls, which conserves the turgor pressure and shape of the cells. Peptidoglycan (PG) is a macromolecular structure that forms a mesh like layer outside the cytoplasmic membrane (van Heijenoort, 2001). It consists of glycan strands cross-linked by short peptides referred to as stem peptides. The glycan strands are made up of alternating units of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), which are linked together by β- 1,4 linkage (Vollmer *et al*., 2008, Park and Uehara, 2008) (Figure 1.14). The glycan moiety of the PG is remarkably uniform among bacteria (Schleifer and Kandler, 1972). Short peptides are attached to MurNAc sugars and expand in all directions. In *E. coli*, a pentapeptide (L-Ala-γ-D-Glu-m-Dap-D-Ala- D-Ala) is attached at the carboxyl group of MurNAc in nascent peptidoglycan (Figure 1.14). Mature peptidoglycan subsequently loses one or both D-Ala residues (van Heijenoort, 2001). About one half of the stem peptides are involved in cross linking between neighbouring glycan strands. This-cross linking occurs between amino group of m-Dap residue at position 3 of one peptide and carboxy terminus of the D-Ala residue at position 4 of adjacent peptide (Navarre and Schneewind, 1999).

In contrast to the uniform structure of the glycan, the peptide moiety shows considerable variation. The peptide subunits are attached at the carboxyl group of muramic acid by their N-terminal amino acid. The amino acid linked to muramic acid is

usually L-Ala, but in some cases it can be replaced by L-Ser (Schleifer and Kandler, 1972). The second amino acid is usually D-Glu. The γ-carboxyl group of D-Glu is linked to the next amino acid in the peptide subunit. The α-carboxyl group is either free or substituted. In many organisms it is amidated for example in *S. aureus*, *S. pneumonia* and *Mycobacteria* (Waldmer review and 1972 review). In some bacteria like Micrococus leuteus it is substituted by Gly (Schleifer and Kandler, 1972). The greatest variation occurs at position 3, where usually a diamino acid is present. The most widely present diamino acid is *meso*-diaminopimelic acid (m-Dap). It is present in most Gram-negative bacteria, Mycobacteria and Bacilli. In most Gram-positive bacteria L-Lys is present instead. Less frequently L-Orn and LL-Dap are present. Since all these amino acids possess an additional amino group, they are excellent anchoring point for the cross linking of peptides. L-homoserine may be present in rare cases (Schleifer and Kandler, 1972, Patin *et al*., 2010) but as it lacks additional amino group, cross linking occurs in a different way. Since all the amino acids at position 3 possess an additional amino group, they are an excellent anchoring point for the crosslinking of the peptide subunits. Indeed, almost all peptide subunits containing one of these diamino acids in position 3 are crosslinked by means of these diamino acids. Positions 4 and 5 are almost always occupied by D-Ala, with very little variation (Schleifer and Kandler, 1972).

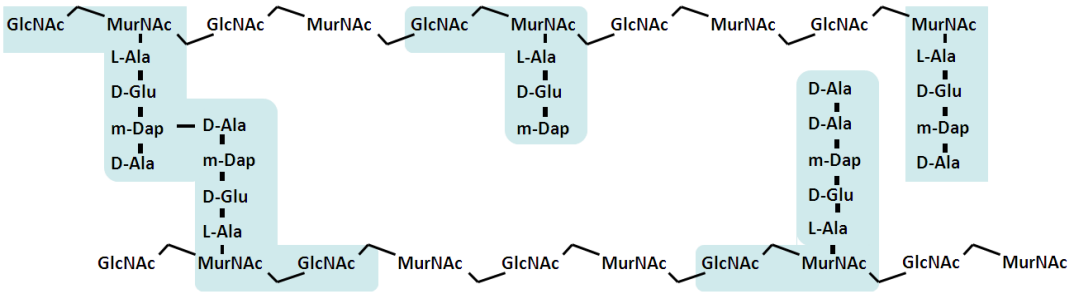


Figure 1.14 Basic structure of peptidoglycan of *E. coli*

1.7.1 Synthesis of peptidoglycan

The biosynthesis of peptidoglycan is a complex process and takes place in three compartments of the cell: cytoplasm, cytoplasmic membrane and periplasm. The first step involves the assembly of its monomer unit by the enzymes located in the cytoplasm or at the inner side of cytoplasmic membrane (van Heijenoort, 2001). The subunits of the PG polymer are synthesized in the cytoplasm as UDP-activated precursors (UDP-MurNAc-pentapeptide). These precursors are then translocated as lipid-linked compounds across the cytoplasmic membrane and finally inserted into the preexisiting PG by murein synthases (Holtje, 1998).

1.7.2 Turnover of PG and recycling of Mtp

The cell wall of bacteria is not a static structure since permanent remodelling occurs during cell growth and division; for example, in *E. coli* 40-50% of peptidoglycan is broken down and reused each generation for the synthesis of new murein (Park, 1995). In this process, murein is degraded by the murein hydrolases which include lytic transglycosylases, amidases and endopeptidases. The main turnover products, which have been identified as GlcNAc-MurNAc(anhydro)-tetra- and tripeptides, are imported into the cytoplasm by a specific permease (AmpG) and are reused in a process that has been termed the recycling pathway (Park and Uehara, 2008) (Figure 1.15). This process involves a large set of enzymes including the amidase (AmpD), the LD-carboxypeptidase (LdcA) and the β-*N*-acetylglucosaminidase (NagZ), that catalyze the stepwise breakdown of anhydro-muropeptides, yielding GlcNAc, anhydro- MurNAc, D-Ala, and the tripeptide L-Ala-γ-D-Glu-*meso*-Dap. The murein tripeptide L-Ala-γ-D-Glu-*meso*-Dap is then directly ligated to UDP-MurNAc by a dedicated murein peptide ligase (Mpl) and thus re-enters the biosynthesis pathway for *de novo* peptidoglycan synthesis (Herve *et al*., 2007, Park and Uehara, 2008).

1.7.3 Peptidoglycan amidases

Amidases are one of the most numerous peptidoglycan degrading enzymes in bacteria. Almost all peptidoglycan amidases are peptidases. In most species there exists a specific amidase for almost every peptide linkage present in the peptidoglycan. Often the number of amidases actually exceeds the number of peptide bonds. For example in *E. coli* 3 different proteins have D-Ala-D-Ala amidase activity (Firczuk and Bochtler, 2007). This multiplicity can be understood as functional redundancy.

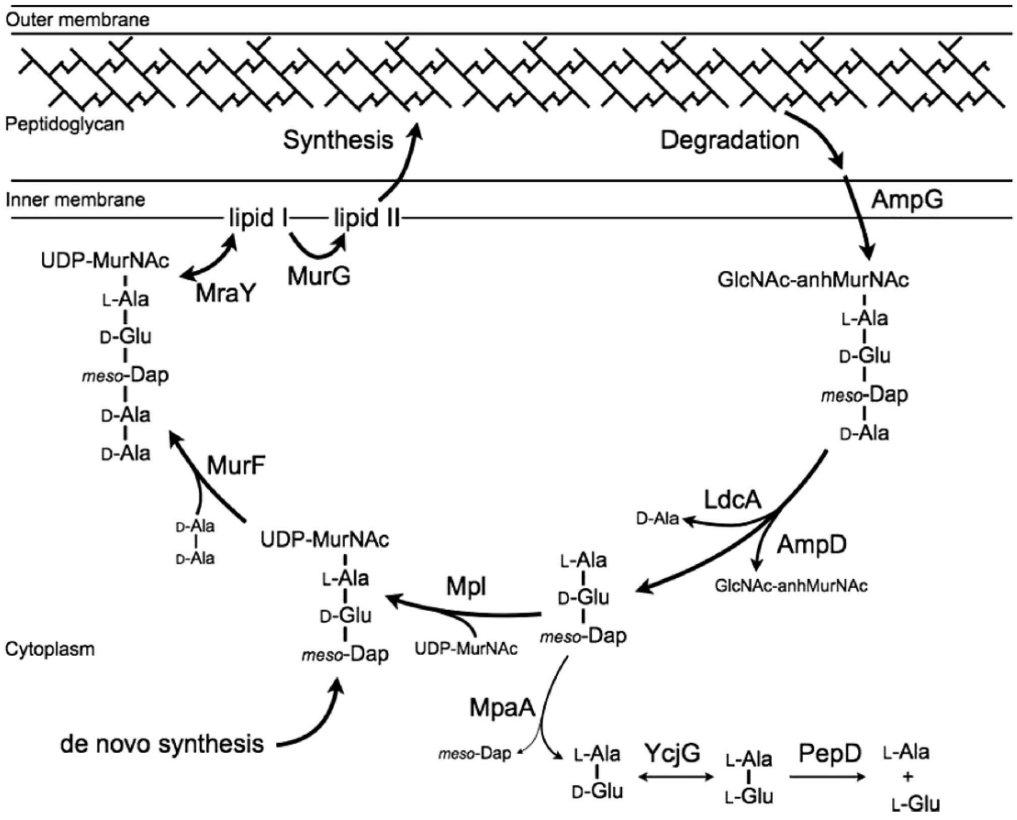


Figure 1.15 Turnover of peptidoglycan and pathway for recycling of murein tripeptide
The figure is adapted from Park *et al*., 2008

Another explanation is that they are involved in different cellular processes at different stages of bacterial life cycle. Peptidoglycan amidases represent a very diverse group of enzymes in terms of their sequence similarities and structural criteria. They can be grouped into three main families; metallopeptidases, cysteine peptidases and serine peptidases and into nine different fold groups (Firczuk and Bochtler, 2007). MpaA is a putative D-Glu-*meso*-Dap amidase of *E. coli* and has been grouped into metallopeptidase family and carboxypeptidase T fold group on the basis of its sequence similarity to characterized amidases (Firczuk and Bochtler, 2007). Chapter 6 describes biochemical and structural characterization of MpaA.

1.8 Protein crystallography

Protein crystallography is a technique used to determine three dimensional structures of proteins and their complexes to atomic resolution level. During this process a crystalline form of a single protein or protein complexes is produced which is then exposed to an X-ray beam resulting in a diffraction pattern on the X-ray detector. Using a combination of mathematical functions, the diffraction pattern is translated into an electron density map. A three dimensional model of the protein structure is then produced by interpreting the electron density map (Blow, 2002, Glusker and Trueblood, 1985) .

The method is similar to light microscopy in many ways. In case of light microscopy light is transmitted through an object. In both cases scattered light is focused by a series of lenses to form an image of object. Similarly X-rays are scattered by the atoms of macromolecule. However X-rays cannot be focused by lenses to form an image of a molecule. Instead X-rays are scattered from a single crystal to give a pattern that represents macromolecular order and structure. The structure is constructed using mathematics as a lens to transform the pattern back into the original structure (Van Holde *et al*., 1998). Since the crystals are imperfect and some information is lost during the transformation, the structure we get is not a true image, as from microscopy, but a model of the structure. Although a number of higher resolution microscopes have been developed, the only method that provides structure of macromolecule at atomic resolution is X-ray diffraction. The closest distance between two atoms in space is the length of a covalent bond. Since the typical covalent bond is approximately 0.12 nm, we need to see two atoms separated by this distance as distinct particles. There are some practical limitations to get a molecular structure solved to atomic resolution. First, the

system must have the atoms of its molecules held rigidly and secondly each molecule must have identical conformations. Any fluctuation in the positions of the atoms in the molecule or any deviation of molecules from a single conformation would result in an averaging of the structure, which would reduce the resolution of the technique.

1.8.1 Protein crystallization

Proteins solidify to form crystals under certain conditions. In the process of crystallization individual molecules adopt identical orientations. The resulting crystal is an ordered three dimensional array of molecules held together by noncovalent interactions. Crystals of protein are grown by slow, controlled precipitation from aqueous solution under conditions that do not denature the protein. A number of substances cause proteins to precipitate. Ionic compounds (salts) precipitate proteins by a process called "salting out." Organic solvents also cause precipitation, but they often interact with hydrophobic portions of proteins and thereby denature them. The water-soluble polymer polyethylene glycol (PEG) is widely used because it is a powerful precipitant and a weak denaturant. One simple means of causing slow precipitation is to add denaturant to an aqueous solution of protein until the denaturant concentration is just below that required to precipitate the protein. Then water is allowed to evaporate slowly, which gently raises the concentration of both protein and denaturant until precipitation occurs. Whether the protein forms crystals or instead forms a useless amorphous solid depends on many properties of the solution, including protein concentration, temperature, pH, and ionic strength. Finding the exact conditions to produce good crystals of a specific protein often requires many careful trials and is perhaps more art than science (Rhodes, 2000).

1.8.2 X-ray diffraction

When an X-ray beam is directed towards a crystal, the X-rays are scattered from a regular repeating array of molecules in a single crystal. As a result a diffraction pattern is obtained on the X-ray detector. Each reflection can be assigned three coordinates or indices (named as *h*, *k* and *l*) in imaginary 3D space of the diffraction pattern. The position of spots from the diffraction experiment gives information about the unit cell dimensions and the intensities of spots give information on the arrangement of atoms in that unit cell. Each diffracted X-ray that arrives at the film to produce a recorded reflection can also be described as the sum of the contributions of all scatterers in the

unit cell. The sum that describes a diffracted ray is called a structure-factor equation. The computed sum for the reflection *hkl* is called the *structure factor* $\boldsymbol{F_{hkl}}$. The structure factor is a wave created by the superposition of many individual waves, each resulting from diffraction by an individual atom (Rhodes, 2000).

1.8.3 Phase problem

In the diffraction experiment we measure the intensities of waves scattered from planes (denoted by *hkl*) in the crystal. The amplitude of the wave $|F_{hkl}|$ is proportional to the square root of the intensity I_{hkl} measured on the detector. To calculate the electron density at a position (xyz) in the unit cell of a crystal requires us to perform the following summation over all the *hkl* planes, which in words we can express as: electron density at (xyz) which is equal to the sum of contributions to the point (xyz) of waves scattered from plane (*hkl*) whose amplitude depends on the number of electrons in the plane, added with the correct relative phase relationship. Mathematically this can be expressed as in equation 1.1, where V is the volume of the unit cell and alpha $_{hkl}$ is the phase associated with the structure-factor amplitude $|F_{hkl}|$ (Taylor, 2003). We can measure the amplitudes, but the phases are lost in the diffraction experiment. This is called the phase problem.

$$\rho(xyz) = 1/V \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz) \qquad \textbf{(1.1)}$$

There are different ways to determine the phases. One of the methods to solve the phase problem is molecular replacement (MR). This method was first described by Michael Rossmann and David Blow (Blow, 2002). MR utilizes the solved structure of a homologous protein as a model source of the initial phases for the new structure. Only structures showing high similarilty can be considered as a suitable model. As a rule of thumb, a sequence identity >25% is normally required between the model and new structure (Taylor, 2003). The phases are obtained from the model structure to produce an initial electron density map. MR is a fast and relatively straight forward approach provided a suitable homology model is available. In the absence of any homology model, experimental phasing is the only way to solve the phase problem (Taylor, 2003). Since the detectors can only detect the intensity of the reflection, the phase information

has to be measured through variations in the intensities. This is achieved through interference between a diffracted wave of unknown phase and a reference wave of known phase. There are two possible sources of such a reference wave that can be introduced into the structure, a heavy atom (isomorphous replacement method) or anomalous atoms such as selenium, which resonate with incoming X-ray radiation (anomalous scattering method) (Blow, 2002).

1.9 Aims of this study

This thesis focuses on the biochemical characterization of peptide specific substrate-binding proteins (SBPs) which are thought to recognize specific peptides. The project uses three peptide-binding SBPs namely MppA, GsiB and YgiS from *Escherichia coli* and involves a combination of biophysical and biochemical methods to examine specific peptide binding and also a structural approach to investigate how specific peptides are coordinated by the SBPs. The study also investigates the catabolism of a peptidoglycan derived component, murein tripeptide, by an amidase enzyme MpaA in the context of conserved scavenging pathway for murein tripeptide in γ-proteobacteria.

Chapter 2

2 Materials and Methods

2.1 Media and Antibiotics

2.1.1 Luria-Bertani broth and agar

Luria-Bertani (LB) broth was made in distilled water using 10 g/l tryptone, 5 g/l yeast extract (trace elements) and 10 g/l NaCl. LB agar was made in smaller volumes of 200 ml using 4g tryptone, 2g yeast extract, 4g NaCl and 3g agar. Volume was made up with distilled water.

2.1.2 M9 minimal medium

20x M9 minimal medium (Neidhardt *et al*., 1974) was made up from 68g Na_2HPO_4 (Fisher Scientific), 30g KH_2PO_4 (Fisher Scientific) and 10g NH_4Cl (Melford) with distilled water to 500 ml, pH adjusted to 7.4 and autoclaved. 1 M $MgSO_4$ (Fisher Scientific) and 12.5 mg/ml $FeSO_4.7H_20$ (AnalaR) were sterilized separately. The M9 minimal medium salts were supplemented with 2 mM $MgSO_4$, 25 μg/ml $FeSO_4.7H_20$ and 0.4 % glucose as a carbon source.

2.1.3 Enhanced M9 minimal medium

Enhanced minimal medium was composed of same components as M9 minimal medium with the addition of following supplements in order to enhance bacterial growth. A mix of 19 amino acids (-Met) at 40 μg/ml, seleno-L-methionine at 40 μg/ml and mix of vitamins at 1 μg/ml.

2.1.4 Antibiotics

Antibiotic selection, where appropriate, was used at 100 μg/ml ampicillin (Melford) and 30 μg/ml kanamycin (Sigma-Aldrich), each from 1000-fold stock, 0.22 μm filter sterilized and stored at 4ºC.

2.1.5 IPTG and Arabinose (Inducers)

10 ml aliquots of 100 mM isopropyl-beta-D-thiogalactopyranoside (IPTG) were routinely made. 238 mg IPTG was weighed and made up with sterile distilled water to 10 ml. This was then filter sterilized using a 0.22 nm filter. All aliquots were stored in sterile universal, wrapped in aluminium foil to prevent photodegredation and stored at 4ºC. Large scale protein expression cultures needed larger amounts of IPTG and thus 1 M IPTG was prepared for that as described above. IPTG was added to cells such that its

final concentration was 0.5 mM. Similarly 5 ml aliquots of 10% L-arabinose were made in sterile water, filter sterilized and stored at 4ºC. L-arabinose was added to cells such that its final concentration became 0.005%.

2.1.6 Strains and plasmids

The genotypes of bacterial strains that were used for this work are described in Table 2.1 while Table 2.3 lists all the plasmids used and constructed during this work.

2.2 Recombinant DNA techniques / Gene cloning

2.2.1 Agarose gel electrophoresis

Separation of DNA fragments by electrophoresis was performed using 1% (w/v) agarose gel in TBE buffer. TBE buffer was made up with 1.62 g/l Tris (Invitrogen), 2.75 g/l boric acid (Fisher Scientific) and 0.95 g/l EDTA (Fisher Scientific). 1 % agarose gels were made by pouring molten agarose directly into a horizontal electrophoresis tank, adding combs to produce appropriate sized wells and then allowing the gel to set. SYBR safe DNA stain (Invitrogen) was added at 10 μl per 100 ml of molten agarose before pouring into the tank. Upon setting of agarose gel the comb and plastic blockers were removed and the gel was submerged in TBE buffer. Samples were prepared by mixing 15 μl DNA with 5x stock of sample loading buffer (Bioline) and loading into a well. Each gel contained at least one DNA ladder (Bioline Hyperladder I) to act as a marker (Figure 2.3). DNA separation was performed by applying a potential difference of 80 Volts for 40 minutes. SYBR safe stained DNA was visualised on a transilluminator (Syngene Imaging System).

2.2.2 Table of primers

The primers used in this work are shown in Table 2.2. All primers were ordered from a biotech company (Eurofins MWG Operon).

2.2.3 Polymerase Chain Reaction (PCR)

PCR was used for the amplification of genes for cloning. High fidelity DNA polymerase (KOD from Novagen) was used throughout. A standard PCR reaction of 50 μl consisted of KOD PCR buffer (1x), 0.2 mM each dNTPs, 1 mM $MgSO_4$, 25 pmoles

Table 2.1 Bacterial strains

Strains	Genotype	Source
DH5α	K-12 F ϕ80d*lacZ*ΔM15 *recA1 endA1 gyrA26 thi-1 supE44 relA1 deoR* Δ(*lacZYA-argF*)*U169*	Invitrogen
BL21 (DE3)	F^- *ompT hsdSB* (r_B^-, m_B^-) *gal dcm* (DE3)	Novagen
MC1061	*araD139* Δ(*ara-leu*)*7696* Δ*lacX74 galU*$^-$ *galK*$^-$ *hsdR2* (r_k^- m_k^+) *mcrB1 rpsL* (F^-)	(Casadaban and Cohen, 1980)
CH212	*hsdS2* (r_B^- m_B^-) *recA13 ara-14 proA2 lacY1 galK2 rpsL20* (Sm^r) *xyl-5 mtl-1 supE44 oppA462*	(Higgins and Hardie, 1983)
B834 (DE3)	F– *ompT hsdS*B (r_B^- m_B^-) *gal dcm met* (DE3)	Novagen

Table 2.2 Details of nucleotide primers used for amplification of DNA fragments

Name of Primer	Sequence of primer (5′-3′)
*mppA*F	CCAGGGACCAGCAATGGCAGAAGTTCCGAGCGGCACAGTACTGG
*mppA*R	GAGGAGAAGGCGCGTTATCAATGCTTCACAATATACATAGTCCGACTGTACG
*gsiB*F	CCAGGGACCAGCAATGGCCAAAGATGTGGTGGTGGCGGTAGG
*gsiB*R	GAGGAGAAGGCGCGTTATTGCAAATCCGCGTCTTCAAAGCTG
*ygiS*F	CCAGGGACCAGCAATGGCTGACGTTCCCGCCAACAC
*ygiS*R	GAGGAGAAGGCGCGTTATCAATGTGCCTTGATATACAACTCT
mppA-NF	ATGGTGAGAATTTATATTTTCAAGGTGCAGAAGTTCCGAGCGGCA
mppA-NR	TGGGAGGGTGGGATTTTCATCAATGCTTCACAATATACATA
mppA-CF	ATGGGTGGTGGATTTGCTATGTCGGTTAGAGGGAAACT
mppA-CR	TTGGAAGTATAAATTTTCATGCTTCACAATATACATAG
gsiB-NF	ATGGTGAGAATTTATATTTTCAAGGTGCCAAAGATGTGGTGGTGG
gsiB-NR	TGGGAGGGTGGGATTTTCATTATTGCAAATCCGCGTCTTC
gsiB-CF	ATGGGTGGTGGATTTGCTATGGCAAGAGCTGTACACCGT
gsiB-CR	TTGGAAGTATAAATTTTCTTGCAAATCCGCGTCTTCAA
*mppAA274E*F	TATTATGAATTTAACACGCAAAAGGGCCCGACG
*mppAA274E*R	GTTAAATTCATAATAATAGGTCCCGAGCTGCGG
*mppAR402A*F	GATAGCGCGAACACCGGCAATTTTGATGTTATC
*mppAR402A*R	GGTGTTCGCGCTATCGATATAGGTCTTCCATTCCTG
*Ec_mpaA*F	CCAGGGACCAGCAATGACCGTAACCCGCCCACG
*Ec_mpaA*R	GAGGAGAAGGCGCGTTATCACGACGGGCGAATTGCATCT
*Vh_mpaA*F	CCAGGGACCAGCAATGAGTCTAATTCCAAGAACG
*Vh_mpaA*R	TAACGCGCCTTCTCCTCCAGCACGATCCAGATCTTTAA
*Vh_mppA*F	CCAGGGACCAGCAATGGATTCTTTGCCTGATGGTGT
*Vh_mppA*R	TAACGCGCCTTCTCCTCAGACCTCTACATCAAGAAGTAG

Table 2.3 Plasmids used for vector construction and expression studies

Plasmid	Description	Resistance	Source
pET-YSBLIC3C	High copy number expression plasmid based on pET28a	Kan	YSBL
pBADnLIC2005	pBAD vector modified with LIC cassette	Amp	Dr. Eric Geertsma
pBADcLIC2005	pBAD vector modified with LIC cassette	Amp	This study
pAM6091	*mppA* in pET-YSBLIC3C with N-terminal H_6	Kan	This study
pAM6092	*gsiB* in pETY-SBLIC3C with N-terminal H_6	Kan	This study
pAM6093	*ygiS* in pET-YSBLIC3C with N-terminal H_6	Kan	This study
pBAD6094	*mppA* in pBADnLIC2005 with N-terminal H_{10}	Amp	This study
pBAD6095	*gsiB* in pBADnLIC2005 with N-terminal H_{10}	Amp	This study
pBAD6096	*mppA* in pBADcLIC2005 with C-terminal H_{10}	Amp	This study
pBAD6097	*gsiB* in pBADcLIC2005 with C-terminal H_{10}	Amp	This study
pAM6091A274E	*mppA* in pET-YSBLIC3C with N-terminal H_6 *mppA* (A274E) mutant	Kan	This study
pAM6091R402A	*mppA* in pET-YSBLIC3C with N-terminal H_6 *mppA* (R402A) mutant	Kan	This study
pEC6091	*Ec_mpaA* in pET-YSBLIC3C with N-terminal hexahistidine	Kan	This study
pVH6092	*Vh_mpaA* in pET-YSBLIC3C with N-terminal hexahistidine	Kan	This study
pVH6093	*Vh_mppA* in pET-YSBLIC3C with N-terminal H_6	Kan	This study
pCH19	*St_oppA* in pCH19	Amp	(Hiles and Higgins, 1986)

of forward and reverse oligonucleotide primers, 1 unit KOD polymerase of Novagen and either genomic or plasmid DNA template. The standard reaction volume was 50 μl and the final volume was made up with sterile distilled deionised water. The whole process was performed on ice in 0.5 ml PCR tubes. Following a 2 minute, 95 °C denaturation step, the target DNA was amplified by 40 cycles of 94 °C for 1 minute, 55 °C for 1 minute and 72 °C for 1 minute per kilobase of target. Following a final extension step of 10 minutes at 72 °C, the PCR product was removed from the thermal cycler (Techne) and frozen.

2.2.4 Site-directed mutagenesis

2.2.4.1 Mutagenic primer design

Oligonucleotide primers for site-directed mutagenesis were designed according to the published protocol (Zheng *et al*., 2004). This used 30-50 base primers with a 5′ overlap to reduce the primer-primer annealing temperature. The overlap covered a fifteen base region centred on the target codon and each primer was extended in the 3′ direction for at least 15 bases or until the expected melting temperature reached 70 °C. Using Webcutter 2.0 (http://users.unimi.it/~camelot/tools/cut2.html), a silent restriction site was introduced into one of the overhanging regions that would produce a unique restriction digestion pattern.

2.2.4.2 Mutagenic PCR conditions

The 50 μl PCR reaction mixture was made up as in Section 2.2.3. Following a 5 minute, 95 °C denaturation step, the whole target plasmid was amplified by 16 cycles of 94 °C for 1 minute, 55 °C for 1 minute and 68 °C for 3 minute per kilobase of target. Following a final extension step of 1 hour at 68 °C, the PCR product was held at 10 °C until it was removed from the thermal cycler and frozen. Before transformation in to competent *E. coli* DH5α, the PCR product was subjected to DpnI restriction enzyme digestion to destroy the template DNA.

2.2.5 Preparation of plasmid DNA

Plasmid DNA was isolated from overnight grown bacterial cultures using a miniprep kit (Machery-Nagel) that employs the alkaline lysis method for plasmid recovery. The manufacturers recommended protocol was used.

2.2.6 Gel extraction

DNA components were purified by 1% agarose gel electrophoresis and extracted from the gel using a Qiagen kit, before switching to one provided by Machery-Nagel.

2.2.7 Ligation independent cloning (LIC)

Ligation independent cloning was used to ligate PCR products of the target genes with a pET-YSBLIC3C (Fogg and Wilkinson, 2008) vector or with pBAD vectors; pBADnLIC2005 and pBADcLIC2005. pET-YSBLIC3C is a cleavable N-terminal His tagged vector that leaves 3 additional N-terminal amino acids after cleavage of His tag by HRV3C protease. It is based on pET28a (Figure 2.1) such that all the restriction sites in the MCS of pET28a are intact. pBADnLIC2005 and pBADcLIC2005 (Figure 2.2) are modified versions of the commercially available pBAD vectors with the inclusion of a LIC cassette at the site of cloning (modified by Dr. Eric Geertsma).

Primers were designed such that a dedicated LIC tail was added to the PCR product, which was designed to be complementary to the LIC cassette in vectors. The gene of interest was amplified using PCR and purified using gel extraction. The vectors were linearised by digesting with the appropriate enzyme, *BseRI* for pET-YSBLIC3C and *SwaI* for pBAD vectors. The linearised vectors were purified by gel extraction. Both the insert and vector DNA were treated with T4 DNA polymerase (Novagen). The $3' \rightarrow 5'$ exonuclease activity of T4 DNA polymerase produces long, complementary overhangs on the vector and the insert. To do this, 200 ng of linearised vector was made up to 10μl with sterile water and treated with T4 DNA polymerase in the presence of dTTP for pET-YSBLIC3C and dCTP for pBAD vectors (1.5 μl 25 mM dCTP / dTTP, 3 μl 5x buffer and 0.5 μl T4 DNA polymerase). The reaction was incubated at room temperature for 30 minutes, after which T4 DNA polymerase was deactivated by incubation at 75 °C for 20 minutes. A volume of insert DNA corresponding to a 1:1 molar ratio with vector was treated with T4 DNA polymerase but this time in the presence of dGTP (for cloning into pBAD vectors) or dATP (for cloning into pET-YSBLIC3C). Both the LIC ready insert and vectors can be stored at -20 °C for several months.

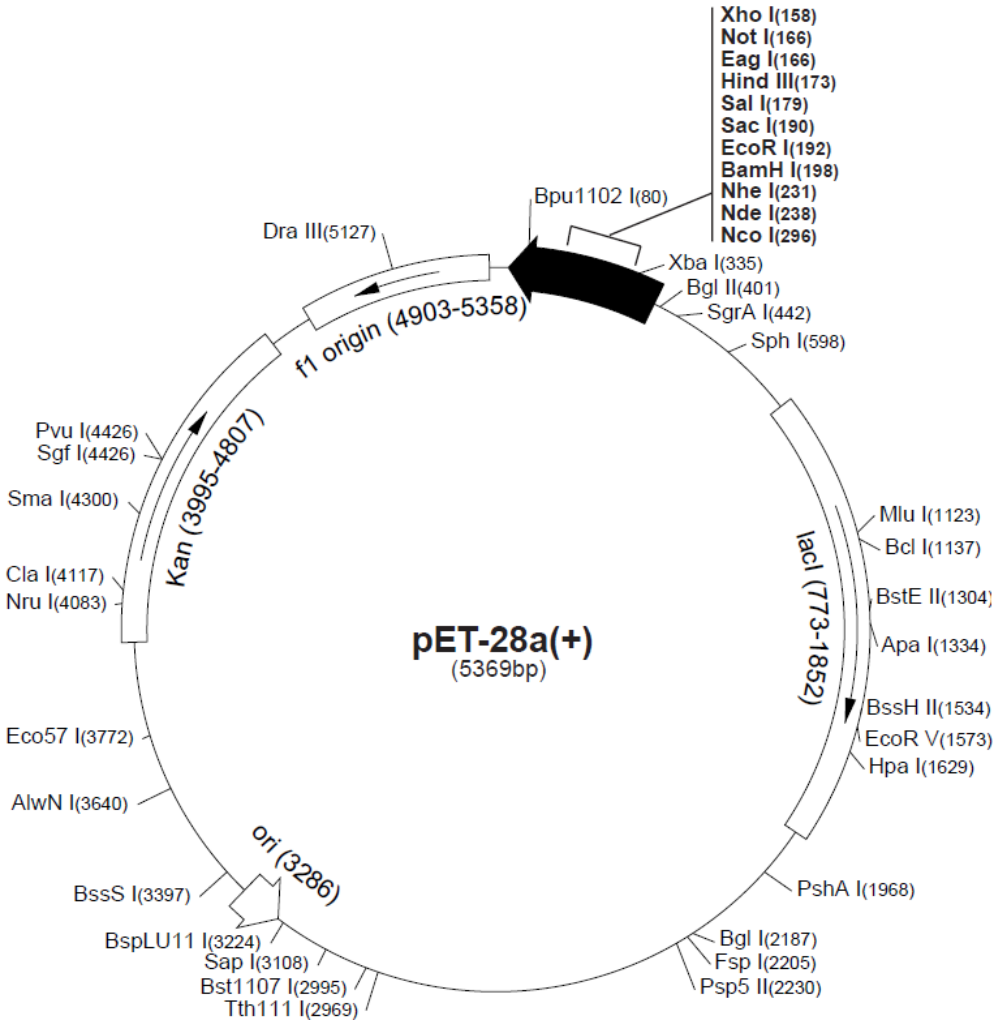
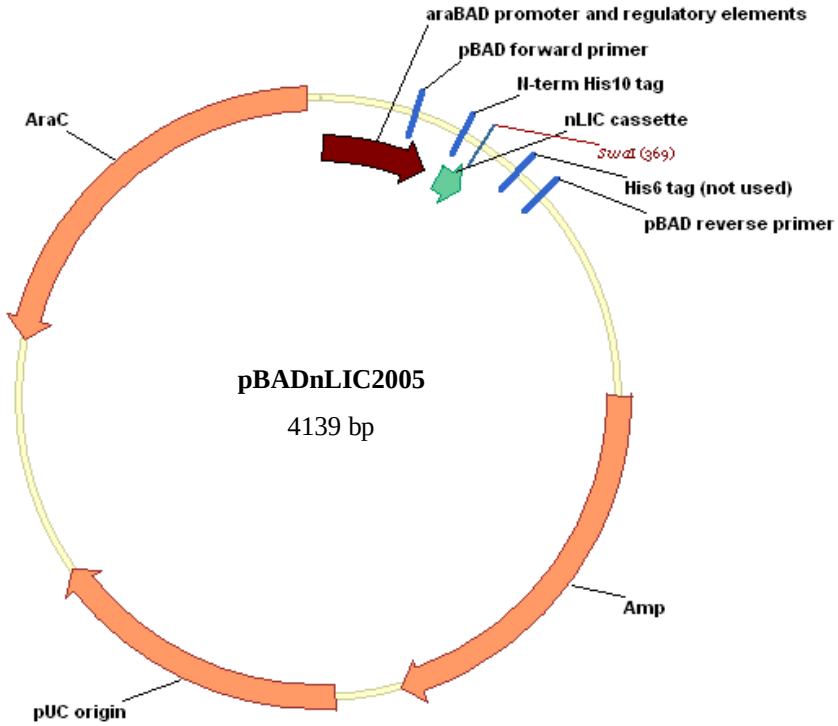


Figure 2.1 Map of pET28a vector

The kanamycin resistance protein is used as a selectable marker. The map is taken from Novagen.



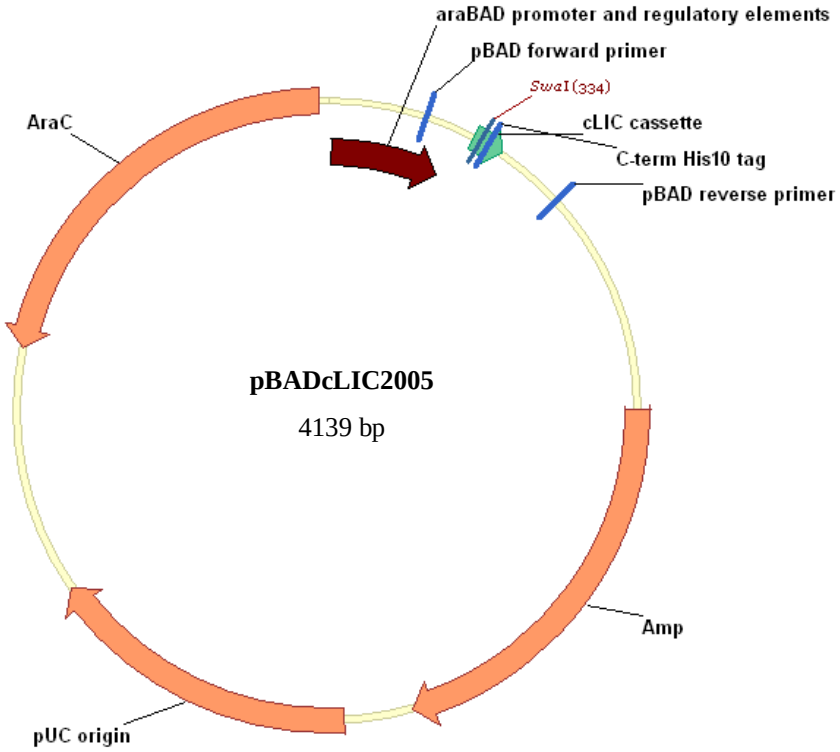


Figure 2.2 Physical maps of pBADnLIC2005 and pBADcLIC2005

Maps were kindly provided by Dr. Christoph Mulligan

The LIC-ready components were mixed in a 1:2 ratio of vector to insert and incubated at room temperature for 20 min. 1.5 μl of 25 mM EDTA was added and the reaction was incubated at room temperature for another 15 min. 2 μl of the reaction was used to transform into competent cells using the heat shock method. Transformation of pBAD based constructs was done into competent cells of MC1061 strain of *E. coli* while transformation of pET based vectors was made into DH5α strain of *E. coli*.

2.2.8 Preparation of competent *E. coli*

To chemically induce competence in *E. coli*, a 10 ml LB culture of the desired strain was incubated overnight. 1 ml of this culture was diluted 10-fold and allowed to grow to an optical density (OD_{650}) of 0.4-0.5 at which point the cells were kept on ice. The cells were separated into 1 ml aliquots in sterile 1.5 ml tubes. The cells were harvested by centrifugation at 4430 x g for 10 minutes and then resuspended in 1 ml ice-cold 100 mM $CaCl_2$. The cells were incubated on ice for 1 hour and then harvested by centrifugation at 4430 x g for 10 min. The aliquots of cells were resuspended in 200 μl ice-cold 100 mM $CaCl_2$ each and incubated on ice for 20 minutes. The chemically competent cells could be used immediately or, by addition of 20 % glycerol in the final resuspension, could be snap frozen with liquid nitrogen and stored at -80°C.

2.2.9 Transformation of competent cells: heat shock method

This method of transformation can be used to transform the desired *E. coli* strain with plasmid DNA or ligation mixtures. This method requires a stock of ready-made competent cells as described above. An aliquot of chemically competent cells was allowed to thaw on ice. 1-5 ng of plasmid DNA or a volume of ligation mixture was added to the cells. The mixtures was mixed by tapping and incubated on ice for further 30 minutes. The mixture was incubated for 30 seconds at 42 °C, and then 1 ml of LB was added and the mixture was incubated at 37 °C for 90 minutes. After that 100 μl of the cell suspension was spread on selective LB agar plate, the remaining cell suspension was harvested by centrifugation and resuspended in 100 μl LB and spread on a second selective LB agar plate. All plates were incubated at 37 °C overnight.

2.2.10 Restriction digestion of plasmid DNA

Restriction endonuclease digestions were used to ensure the insert was correctly cloned into vectors. Restriction digests are normally done using two enzymes, which are normally chosen based on sequence they cut at and intercompatibility. For example NdeI and NcoI are compatible with each other and can be used in NEB buffer 4. Typically a 10 μl reaction of restriction digestion reaction consists of 0.5 μl of each restriction enzyme, 1 μl of appropriate buffer, 1 μl of mini preprating DNA and 7 μl of sterile distilled water. The reaction was incubated at 37°C for a minimum of two hours. After incubation 2 μl of loading dye was added into reaction tube and analysed on a 1% agarose gel.

2.2.11 Automated DNA sequencing

Where necessary, the sequence of specific target region (gene of interest) in recombinant plasmid was confirmed by DNA sequencing done at Technology Facility, University of York. The sequence data were verified on the basis of corresponding fluorescence electropherogram.

2.3 Expression of recombinant proteins

2.3.1 Whole cell small scale expression trials

For expression of recombinant genes, 1-2 ng of purified plasmid was introduced into the expression strain given in Table 2.1. A single colony was used to inoculate 5 ml LB medium and the culture was grown aerobically at 37°C overnight. This was used to inoculate 25 ml of fresh LB medium (supplemented with antibiotic) to OD_{650} of 0.05. Cultures were grown aerobically at 37°C with shaking (180 rpm) to mid-log phase (OD_{650} of 0.5) and induced (IPTG / L-arabinose). 1 ml of culture samples were taken every hour after induction for four hours. A sample was also taken from the culture after overnight growth. It was done for each of four constructs. Cells were sedimented by centrifugation at 38000 x g for 5 min. Cells were resuspended in 25 × b μl of 4x sample buffer (where b μl is the OD_{650} of the cells when harvested), boiled for 5 minutes at 97.5 °C on the heating block and centrifuged at 38000 x g for 5 minutes to get the supernatant. Accumulation of protein was checked by loading the supernatant on sodium dodecyl sulphate-polyacrylamide gel and running the gel at 200 Volts for 50 minutes.

2.3.2 Solubilization trials

Multiple samples (1 ml) were taken from small scale expression trials, cells were pelleted by centrifugation and the supernatant was discarded. The cells were resuspended in a particular buffer (for example, 50 mM potassium phosphate pH 7.8 or 50 mM Tris pH 8) and lysed using a microtip sonicator (3×3 sec). A small portion of total cell lysate was taken out and the rest was centrifuged at 38000 x g for 5 min to get a clear/soluble lysate. Total and soluble fractions were run on SDS-PAGE to check the solubility of over produced protein in that buffer.

2.3.3 Large scale expression

625 ml LB medium (supplemented with antibiotic) was inoculated to OD_{650} of 0.05 with overnight cultures of the particular *E. coli* strain. Cultures were grown to mid log phase (OD_{650} of 0.5) at 37 ºC with shaking (180 rpm), at which point expression was induced. After induction, cells were grown overnight and were harvested by centrifugation at 4430 x g for 15 min and resuspended in appropriate buffer (For example, 50 mM KPi, pH 7.8 supplemented with 20% glycerol). Cells were lysed by sonication and cell debris was removed by centrifugation at 4430 x g. Soluble fraction, containing the over produced protein was stored at -20ºC.

2.4 Selenomethionine labeling of recombinant proteins

To produce selenomethionyl protein, the plasmid containing the cloned gene was transformed to BL 834 (DE3) strain of *E. coli*. This strain is a derivative of the commonly used BL21 (DE3) strain, and is methionine-requiring auxotroph containing the T7 RNA polymerase gene under *lac* control on the host chromosome. The T7 RNA polymerase is turned on by the addition of IPTG to 0.5 mM. Growth of these cells in selenomethionine ensures that all the methionine in the protein is in the form of selenomethionine. Growth in selenomethionine is slow in minimal media (M9) and yield can be low. This problem can be addressed by using a synthetic medium (enhanced M9 minimal medium) containing a cocktail of the other 19 amino acids and L-selenomethionine.

2.4.1 Small scale expression trials

10 ml of enhanced minimal M9 medium (supplemented with kanamycin) was inoculated to OD_{650} of 0.1 with overnight grown culture in LB. Cultures were grown

aerobically at two temperature (37°C and 16°C) at 180 rpm. All cultures were grown to mid-log phase (OD_{650} of 0.5) and induced with 0.5 mM IPTG. 1 ml of culture samples were taken every hour after induction for three hours. Cells were sedimented by centrifugation at 38000 x g for 5 min. Cells were resuspended in 100 μl of resuspension buffer (50 mM Tris pH 8) and were lysed by using microtip sonicator for 3×3sec. A portion (10 μl) of total lysate was taken out and rest was spun in microcentrifuge at 38000 x g for 5 min in order to get clear lysate containing soluble protein. Total and soluble fractions were run on SDS-PAGE to check accumulation and solubility of over produced protein.

2.4.2 Large scale expression

500 ml enhanced minimal M9 medium (supplemented with kanamycin) was inoculated to OD_{650} of 0.1 with overnight culture (inoculum) of the particular *E. coli* strain in LB that had been washed twice with M9 media. Cultures were grown to mid-log phase (OD_{650} of 0.5) at 37 ºC with shaking (180 rpm), at which point expression was induced by the addition of 0.5 mM IPTG. After induction, cells were grown for four hours and were harvested by centrifugation at 4430 x g for 15 min and resuspended in 50 mM Tris buffer of pH 8. Cells were lysed by sonication and cell debris was removed by centrifugation. The soluble fraction, containing the over produced protein was stored at -20 ºC and later on used for purification of over produced protein.

2.5 Preparation of bacterial extracts

2.5.1 Periplasmic fraction preparation

The cells were harvested by centrifugation at 4430 x g for 15 minutes at 4 ºC (SLC 6000 rotor in Sorvall Evolution centrifuge). The subsequent cell pellet was resuspended in ice-cold SET buffer (5 mM EDTA, 0.5 M sucrose and 50 mM KPi pH 7.8) in a ratio of 10 ml per litre of original culture. Chicken egg white lysozyme (Sigma) was added to cell resuspension at 600 μg/ml and incubated at 30 ºC for 2 hours. The subsequent sphaeroplasts were harvested by centrifugation at 38000 x g for 30 minutes at 4 ºC (SS34 rotor in Sorvall Evolution centrifuge); the supernatant from this centrifugation was the periplasmic fraction.

2.5.2 Cytoplasmic protein recovery

Induced cells were harvested by centrifugation at 4430 x g for 15 minutes at 4 ºC (SLC 6000 rotor in Sorvall Evolution centrifuge). The cells were resuspended in 35 ml of resuspension buffer (for 1250 ml of culture) and sonicated on ice for a total of 10 minutes to rupture the cells. Insoluble protein and cell debris were removed by centrifugation at 38000 x g for 10 minutes at 4 ºC and the supernatant collected.

2.6 Protein purification techniques

2.6.1 Nickel-affinity chromatography using gravity flow columns

The soluble fraction containing the protein of interest was mixed with Ni^{2+}-NTA resin (Qiagen) equilibrated with buffer A (50 mM KPi, pH 7.8, 200 mM NaCl, 20% glycerol) supplemented with 10 mM imidazole. The mixture was incubated for 1 h at room temperature in a disposable polystyrene column (Pierce). Subsequently, the resin was washed with buffer A containing 50 mM imidazole, for 20 column volumes (CV) to remove weakly binding contaminants. Bound His-tagged protein was eluted with elution buffer A containing 500 mM imidazole.

2.6.2 Nickel-affinity chromatography using HisTrap HP columns

The soluble fraction containing recombinant protein was filtered through a 0.2 μm filter. A 5 ml HisTrap column (GE Healthcare) was equilibrated with buffer A (50 mM Tris pH 8, 200 mM NaCl and 10 mM imidazole) and connected to an AKTA Purifier P-900. Filtered supernatant was then injected into a pre-equilibrated column. The column was washed with 7 CV of wash buffer (50 mM Tris, pH 8, 200 mm NaCl and 10 mM imidazole) to remove weakly bound contaminants. A gradient was applied over 20 CV from 10 to 500 mM imidazole to elute His-tagged bound protein. Purification was monitored by detecting the absorbance of the column eluent at 280 nm (A_{280}). Fractions were collected with an automated F-950 Fraction Collector (GE Healthcare) and were run on SDS-PAGE gel for analysis.

2.6.3 Ion exchange chromatography

The soluble fraction containing over produced protein was dialysed into 50 mM Tris buffer of pH 8 and filtered through a 0.2 μm filter. A 5 ml MonoQ column (GE Healthcare) was equilibrated with buffer A (50 mM Tris pH 8) and connected to an

AKTA Purifier P-900. The protein was injected on the column using a flow rate of 2 ml/min. The column was then washed with 10 CV of buffer A. A linear gradient was applied over 20 CV from 0 to 500 mM NaCl with a flow rate of 2 ml/min. Purification was monitored by detecting the absorbance of the column eluent at 280 nm. Fractions were collected with an automated F-950 Fraction Collector (GE Healthcare) and were run on SDS-PAGE gel for analysis.

2.7 Purification of ligand free protein

2.7.1 Guanidine hydrochloride (GnHCl) denaturation

2.7.1.1 Buffers

The wash buffer was made up to 50 mM KPi buffer pH 7.8, 200 mM NaCl, 20% glycerol, 10 mM imidazole from higher concentration stocks. Guanidine hydrochloride (GnHCl) was added to 50 mM KPi buffer, 200 mM NaCl, 20% glycerol, 20 mM imidazole to give wash buffers containing 2 M, 1.5 M, 1 M and 0.5 M GnHCl. The elution buffer was made up of 50 mM KPi buffer, 200 mM NaCl, 20% glycerol, 500 mM imidazole.

2.7.1.2 In-column denaturation and refolding

The soluble fraction containing protein was mixed with pre-equilibrated Ni-NTA resin and incubated for 1 hour at room temperature in gravity flow columns. After the resin was sedimented in a disposable column, the bound protein was denatured with 30 CV of buffer containing 2 M GnHCl, followed by 4 CV of each of the wash buffer containing decreasing GnHCl concentrations. Finally, the column was washed with 50 mM KPi buffer, 200 mM NaCl, 20% glycerol, 20 mM imidazole pH 7.8 and then the protein was eluted with 50 mM KPi buffer, 200 mM NaCl, 20% glycerol, 500 mM imidazole pH 7.8.

2.7.2 Cation exchange chromatography

The soluble fraction containing protein was dialysed into buffer A (50 mM sodium acetate pH 5) and filtered through a 0.2 μm filter. A 5 ml MonoS column (GE Healthcare) was equilibrated with buffer A and connected to an AKTA Purifier P-900. The protein was injected into the column using a flow rate of 2 ml/min. The column was then washed with 25 CV of buffer A. A linear gradient was applied over 20 CV from 0

to 500 mM NaCl with a flow rate of 2 ml/min. Purification was monitored by detecting the absorbance of the column eluent at 280 nm. Fractions were collected with an automated F-950 Fraction Collector (GE Healthcare) and were run on SDS-PAGE gel for analysis.

2.8 SDS- Polyacrylamide gel electrophoresis (SDS-PAGE)

2.8.1 Buffers and gel

The resolving polyacrylamide gel was cast from 12% acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% sodium dodecyl sulphate (SDS; Melford), 0.1% APS, 0.01% TEMED, while the stacking gel was cast from 4% acrylamide, 125 mM Tris/HCl pH 6.8, 0.1% SDS, 0.1% APS, 0.01% TEMED. The running buffer in the tank was made up of 3 g/l Tris, 14 g/l glycine and 1 g/l SDS and pH was adjusted to 8.8 with HCl.

2.8.2 Sample preparation

The desired amount of protein was mixed with sterile water and 4x sample buffer consisting of 100 mM Tris pH 6.8, 5 % SDS, 0.4 % bromophenol blue, 20% glycerol and 30 μl/ml β-mercaptoethanol. This mixture was then heated to 97.5 °C for 5 minutes and spun at 38000 x g for 5 minutes. NEB prestained protein marker (Figure 2.3B) was used to allow the molecular weight of proteins to be estimated.

2.8.3 Electrophoresis conditions

SDS-PAGE gels were run using a Mini-PROTEAN 3 Cell system (Bio-Rad). The manufacturer's guidelines were used for the assembly of the SDS-PAGE apparatus. Once the gel was submerged in running buffer, the samples were loaded and a 200 volt potential difference was applied for 50 minutes. When finished, the gels were rinsed with distilled water.

2.8.4 Staining/destaining of SDS-PAGE gels

Coomassie Brilliant Blue (CBB) dye was used to detect proteins on SDS-PAGE gels. The stain was prepared by mixing 2.5 g CBB, 450 ml methanol, 450 ml sterile water and 100 ml acetic acid. Gels were removed from apparatus, rinsed with water and submerged in stain for 5 minutes with gentle agitation. The excess stain was poured off

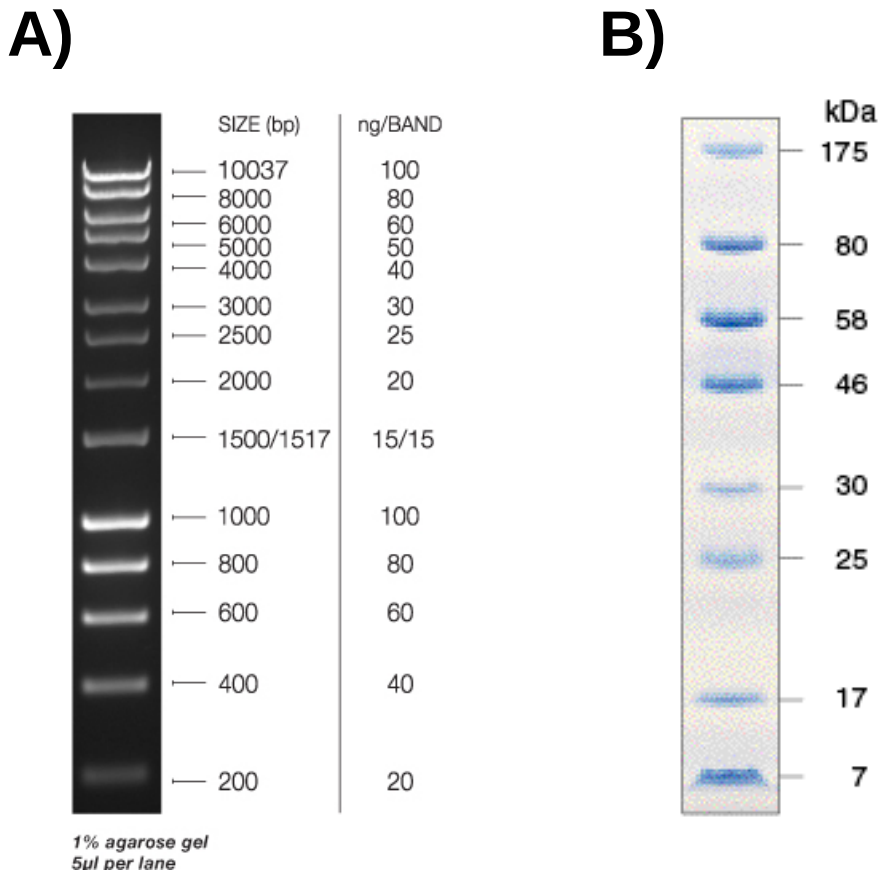


Figure 2.3 Markers for DNA and protein gels

A) DNA hyperladder I from Bioline **B)** NEB prestained protein marker

and the gel was briefly rinsed with water to remove the residual dye. Approximately 100 ml of destain containing 10% acetic acid and 10% ethanol was added and heated at full power in a microwave for 90 seconds. The gel was left on a rocker with gentle agitation for 30 minutes before an image was captured using the Syngene Imaging System.

2.9 Protein concentration determination

2.9.1 Bradford assay

A set of protein standards were produced by serially diluting a stock solution of bovine serum albumin (BSA) from 400 μg/ml to 6.25 μg/ml. The protein of unknown concentration was serially diluted by at least 6 times. 50 μl of each of dilutions of BSA and unknown sample was added to 1 ml of Bradford reagent (Bio-Rad) in a cuvette. The absorbance was measured at 595 nm wavelength. A graph of protein concentration versus absorbance (A_{595}) was plotted and the concentrations of the unknown samples were calculated using the standard curve.

2.9.2 Using A_{280}/extinction coefficient

A NanoDrop ND-1000 spectrometer was used to confirm the results from the Bradford assay by measuring the absorbance at a wavelength of 280nm. The molecular weight and the extinction coefficient of purified proteins were calculated using the peptide property calculator tool available on Northwestern University website (http://www.basic.northwestern.edu/biotools/proteincalc.html). Protein concentration was determined using the Beer-Lambert law.

2.10 Dialysis, concentrating and storage of protein

Dialysis was used to change the buffer solution of both proteins and cell fractions. Dialysis was performed using 0.33 or 3.33 ml/cm dialysis tubing that is activated by "moisturising" using distilled water. Mediclips were used to seal the end of the tubing and then the sample pipetted into the dialysis tubing which was placed in a beaker of the new buffer at least 100x the volume to be dialysed. A minimum of three changes of buffer was used and each dialysis step left for a minimum three hours and at least one dialysis step was done overnight, with all dialysis performed at 4ºC. Protein concentration was increased using a MWCO 5000 Vivaspin columns. These columns

use centrifugation to reduce the amount of solvent while preventing the protein from passing through the membrane. Concentrated protein was stored at –80ºC for long periods, or at 4ºC for routine use.

2.11 General biochemical and biophysical techniques

2.11.1 Mass spectrometry

A Micromass LCT Premier XE mass spectrometer using a nanospray source was used for electrospray mass spectrometry (ES-MS). For analysis of protein under denaturing conditions, purified protein (after removal of the His-tag) was dialysed into 50 mM KPi buffer pH 7.8 or 10 mM Tris pH 7.5 and concentrated to 100 μM. Concentrated protein was diluted to 1 μM in 1:1 acetonitrile-water containing 0.1% formic acid. The sample was passed into the ES-MS and the data were collected in positive mode over 3 minutes within an m/z range of 500-2000. To investigate the mass of any endogenous ligand bound to protein, a spectrum was obtained under native conditions. For analysis under native conditions purified MppA was dialysed into sterile water or 200 mM ammonium acetate and then concentrated to 100 μM. Data were collected over 5 minutes within an m/z range of 500-10000. The raw m/z data were deconvoluted to mass spectra using the MaxEnt1 routine in the MassLynx software provided by the manufacturer (Waters).

2.11.2 Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS / FT-MS)

A CoEMS's Burker apex ultra 9.4 Tesla spectrometer was used for Fourier transform ion cyclotron resonace mass spectrometry of protein complexes. Purified protein after removal of the His-tag was dialysed into water and concentrated to 200 μM and given to Adam Dowle in the Technology Facilty, University of York, for analysis using the CoEMS's mass spectrometry equipment. Samples were diluted to between 10 and 100 μM to give a final concentration of 50 % (*v;v*) aqueous methanol, 1% (*v;v*) formic acid, before introduction into the mass spectrometer.

FT-ICR-MS was performed using an Apex ultra Fourier Transform Ion Cyclotron Resonance mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with a 9.4 T superconducting magnet, interfaced to a Triversa NanoMate nanoESI source (Advion BioSciences, Ithaca, NY).

Instrument control and data acquisition were performed using Compass 1.3 for Apex software (Bruker Daltonics). Nitrogen served as both the nebulizer gas and the dry gas for nanoESI. Observed precursors were manually selected for sustained off-resonance irradiation (SORI) fragmentation in the infinity ICR cell, with argon used as both the cooling and collision gas.

Bruker dataAnalysis software (version 4.0) was used to perform the spectral processing and peak picking for both MS and SORI fragmentation spectra. Monoisotopic masses of proteins were calculated using the SNAP 2.0 algorithm in dataAnalysis (repetitive building block: C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583).

2.11.3 Circular Dichroism (CD) spectroscopy

CD spectra of MppA mutants were obtained using a J-810 spectropolarimeter (Jasco) controlled by the supplied software SpectraManager version 1.53.00 (Jasco). Protein was dialysed into 10 mM KPi buffer pH 7.8 and diluted to a concentration of 5 μM. The spectrum was recorded at 20 °C in a 1 mm pathlength quartz cuvette (Starna) between 180 – 240 nm at 100 nm/minute with 1 nm pitch.

2.11.4 Tryptophan fluorescence spectroscopy

Tryptophan fluorescence spectroscopy was performed using a FluoroMax 4 fluorescence spectrometer (Horiba Jobin-Yvon) with connecting water bath to maintain temperature. Fluorescence experiments were performed in 3 ml quartz cuvettes (Starna) and purified protein was used at a concentration of 0.05 μM in 50 mM KPi, pH 7.8. The sample was excited at the 297 nm with slit widths of 3.5 nm and emission was monitored at 330 nm with slit widths of 3.5 nm. Increasing concentrations of ligand was added to protein solution in 3 ml cuvette and the fluorescence change was measured. The cumulative fluorescence change was plotted against the cumulative concentration of ligand in SigmaPlot 11. The data were fitted to a single rectangular hyperbola, using the equation 2.1 and non-linear regression was used to solve the equation and produce values for dissociation constant (K_D).

$$y = \frac{ax}{b + x} \qquad \textbf{(2.1)}$$

Here, *y* represents the change in fluorescence, *x* represents concentration of ligand, *b* represents dissociation constant (K_D) and *a* represents fluorescence when curve flattens at high concentration of *x*.

The change in fluorescence for addition of ligand was rescaled before plotting in SigmaPlot. A fluorescence change of 10,000 counts per second was scaled to 1. The data used to calculate dissociation constant (K_D) of Mtp to MppA is given below in Table 2.4 as an example.

Table 2.4 Fluorescence data for binding of Mtp to MppA

Concentration of ligand (Mtp) added (μM)	Cumulative concentration of Mtp added (μM)	Change in fluorescence upon addition of Mtp (cps)	Change in fluorescence upon addition of Mtp (10,000 cps = 1)	Cumulative change in fluorescence upon addition of Mtp
0	0	14,88,000	0	0
0.0676	0.0676	15,00,000	1.2	1.2
0.0879	0.1555	15,35,000	3.5	4.7
0.1488	0.3043	15,55,000	2	6.7
0.2706	0.5749	15,70,000	1.5	8.2
0.676	1.2509	15,85,000	1.5	9.7
1.3491	2.6	15,85,000	0	9.7

2.11.5 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a solution based technique to determine the thermodynamic parameters of molecular interactions. In the simplest case, a ligand (L) is introduced into a solution of protein (P) that can bind to form a simple complex (PL). The binding parameters; enthalpy change (ΔH), entropy change (ΔS), change in Gibb's free energy (ΔG), association constant (K_a) and stoichiometry (n) can be determined from a single ITC experiment by measuring the heat evolved or absorbed during an interaction. The heat change is a direct result of the protein-protein interactions.

Calorimetry experiments were carried out using a VP-ITC instrument (MicroCal Inc., GE Health Sciences). In a typical experiment, the calorimeter cell contained 1.4 ml of protein, and the syringe contained 300 μl of peptide ligand. The concentration of ligand in the syringe was typically 10 times that in the cell, whereas the cell concentration was chosen according to *c* value of greater than 10, where *c* = [macromolecule]/ (predicted) K_D. Experiments were carried out in 50mM KPi buffer, pH 7.8 at 25°C. Both cell and syringe solutions were degassed at 20 °C for 10 min before use. The titrations were performed as follows. A single preliminary injection of 5 μl of ligand solution was followed by 28 injections (10 μl), delivered at an injection speed of 10 μl s^{-1}. Injections were made at 3 min intervals with a stirring speed of 307 rpm. For titrations involving peptide PFK a heat of dilution control experiment was performed, where equivalent volumes of peptide were injected into buffer. This ruled out the possibility of heat of dilution masking weak binding. Raw titration data were integrated and binding isotherms fitted by an iteration process based on the 'one-set of sites' model using MicroCal Origin version 7.0. The goodness of the fit was determined by an algorithm in the software (least square method). Values for K_a can and ΔH can be determined by fitting the data obtained from the ITC experiment with the equation (2.2), where q_i is heat absorbed or released at time interval *i* for the amount of ligand added at that interval, T is temperature and V is volume of the sample.

$$q_i = \Delta H(T)V[P_T]\left(\frac{K_a[L]_i}{1+K_a[L]_i} - \frac{K_a[L]_{i-1}}{1+K_a[L]_{i-1}}\right) \qquad \textbf{(2.2)}$$

Once K_a and ΔH are determined, ΔG can be calculated as $\Delta G = -RT \ln K_a$ and ΔS from the standard equation $\Delta G = \Delta H - T\Delta S$.

2.11.6 Size Exclusion Chromatography Multi-Angle Laser Light Scattering (SEC-MALLS)

Size exclusion chromatography (SEC) coupled with multi angle laser light scattering is a technique, which allows to determine the molecular weight of proteins and their complexes in solution (Folta-Stogniew and Williams, 1999). In a typical measurement the fractionation device, a light scattering (LS) detector and a concentration detector are

connected in series. The concentration detector is usually a refractive index (RI) or ultraviolet (UV) absorption detector. When light passes through a solution containing the macromolecule (protein), interaction with the macromolecule leads to scattering of light beam. The scattering of light can directly determine the molecular mass, molecular root mean square (rms) radius size and aggregation state of macromolecule and is measured by LS detector. Because accurate concentration determination is vital for estimation of molar masses, the system also includes a refractive index detector. The molecular mass is determined for each eluting fraction separately.

For SEC-MALLS experiments, sample of protein (100 μl of 1 mg/ml in 20 mM MES buffer pH 6.5, 200 mM NaCl) was injected and separated on a Superdex 200 10/300 GL SEC column (GE Healthcare) at 0.5 ml/min with an HPLC system (Shimadzu). Light scattering data were recorded on an in-line Wyatt Dawn Heleos LS detector with an in-line Wyatt Optilab rEX refractive index detector and SPD-20A UV detector. Molecular weights were calculated by analysing the LS and RI data with the Wyatt program ASTRA version 5.3.4.14. A value of 0.19 ml/g was used for protein refractive index increment (dn/dc).

2.12 Crystallization of protein

Purified proteins were maintained at a concentration of 7-30 mg/ml in different buffers. Preliminary crystallization screening was performed by the Mosquito Nanolitre Pipetting robot (TTP Labtech) to set up 96-well plates at room temperature, using the sitting-drop vapour diffusion method in which 150 nl of protein were mixed with 150 nl of reservoir solutions. Various commercially available screens were used: Index, Hampton I and II, JCSG (Hampton Research, Molecular Dimensions), PACT (Newman *et al*., 2005), CSS I and II (Brzozowski and Walton, 2001). The plates were sealed to avoid evaporation and then incubated at 20 ºC and examined regularly for the appearance of crystals. Conditions where small crystals grew were further optimized either using the same robotic procedure or done manually in 24 well plate using hanging-drop vapour diffusion and a 1:1 mixture of protein solution with the reservoir solution (1 μl each).

Crystals were harvested by nylon loops of various sizes, transferred to a precipitant solution containing glycerol for cryoprotection and flash cooled in liquid nitrogen. The

crystals were then checked for diffraction at the YSBL X-ray source and test images were collected using a Rigaku RU-H3R X-ray generator with rotating anode equipped with Osmic multilayer optics and a MAR Research mar345 imaging plate detector. Test data was collected in-house at 0° and 90° in the two crystal dimensions orthogonal to the incident X-ray beam to check for anisotropic diffraction and to get an indication of crystal quality, exclude salt crystals and determine unit cell dimensions. In case the sample changer was used, data was collected on an R-Axis IV++ detector. This significantly reduced the time invested on screening of dozens of crystals. The best crystal was frozen in liquid nitrogen and was shipped to European Synchrotron Radiation Facility (ESRF), Grenoble, or the DIAMOND light source, Didcot, for X-ray data collection

.

Chapter 3

3 Expression, purification and biophysical characterization of the putative peptide-binding SBPs, MppA, GsiB and YgiS from *E. coli*

3.1 Cloning of *mppA, gsiB* and *ygiS* genes

A common approach for expression and purification of substrate-binding proteins is to clone the protein encoding genes in-frame with short tags like hexahistidine, into a plasmid vector suited for over pruction of the proteins. In the present study three genes encoding putative peptide-binding proteins were cloned into two different expression vectors, an IPTG inducible pET-YSBLIC3C vector and an arabinose inducible pBAD vector.

The pET-YSBLIC3C expression vector is based on pET28a (Figure 2.1 in methods section) and has an N-terminal hexahistidine tag which can be cleaved by HRV 3C protease (Fogg and Wilkinson, 2008). Regions of the *E. coli* K-12 genome were amplified by PCR using primers *mppA*F and *mppA*R for *mppA, gsiB*F and *gsiB*R for *gsiB,* and *ygiS*F and *ygiS*R for *ygiS* amplification. These were cloned into pET-YSBLIC3C using ligation independent cloning (LIC) so that the coding sequence was in-frame with a 5′ nucleotide sequence encoding a hexahistidine tag (Methods section 2.2.7). This resulted in three different expression vectors; pAM6091, pAM6092 and pAM6093 expressing an N-terminally hexahistidine tagged MppA, GsiB and YgiS, respectively. The vectors were made such that they encode a version of protein lacking its native signal peptide and resulting in cytoplasmic accumulation of protein.

The pBAD system has been used successfully in the Thomas lab to obtain high yield of soluble SBPs (Mulligan *et al.*, 2009). It allows addition of the decahistidine tag to either the N or C-terminus of the protein. Regions of the *E. coli* K-12 genome were amplified by PCR using primers *mppA*-NF with *mppA*-NR and *mppA*-CF with *mppA*-CR for *mppA* and *gsiB*-NF with *gsiB*-NR and *gsiB*-CF with *gsiB*-CR for *gsiB* and cloned into pBAD using LIC so that the gene sequence was in-frame with a 5′- or 3′- nucleotide sequence encoding a decahistidine tag (Methods section 2.2.7). This resulted in four different expression vectors; pBAD6094, pBAD6095, pBAD6096 and pBAD6097 expressing an N- or C- terminally decahistidine tagged MppA and GsiB. pBAD6094 and pBAD6095 resulted in cytoplasmic accumulation of MppA and GsiB whereas pBAD6096 and pBAD6097 encode a version of protein with its native signal peptide and resulting in periplasmic accumulation of MppA and GsiB, respectively.

Construction of the desired recombinant plasmid was confirmed by restriction analysis with the appropriate restriction endonuclease (Figure 3.1). Recombinant plasmids that showed the correct insert size, ~1.5 Kb for *mppA*, *gsiB* and *ygiS*, were sequenced in the Technology Facility (University of York) and the sequence analysis confirmed in frame cloning of genes with no mutation.

3.2 Expression trials of selected binding proteins

If the putative peptide-binding SBP genes appear to be correctly inserted into the expression vectors, a series of small-scale expression trials were performed to determine whether the recombinant gene is expressed and is of the correct molecular weight.

3.2.1 Expression trials of *mppA*, *gsiB* and *ygiS* genes cloned in pET-YSBLIC3C

The recombinant plasmids (pAM6091, pAM6092 and pAM6093) were transformed into the expression strain BL21 of *E. coli* and used in small scale expression trials. The expression trials were performed using 25 ml LB media in 100 ml flask and cultures of *E. coli* BL21 strains were grown at two temperatures, 25°C and 37°C. Cultures were induced using 0.5 mM IPTG at an OD_{650} of 0.4-0.6. Growth was monitored by measuring OD_{650} after each hour starting from the time of induction and was continued for four hours post induction with a final reading taken after 24 hours. At each time point 1 ml of culture was pelleted for later analysis by SDS-PAGE. The rate of growth for each target grown at 25°C and 37°C can be seen in (Figure 3.2). None of the strains failed to grow or showed an obvious reduction in growth after induction so this shows that the recombinant proteins are not toxic to the cell, assuming they were induced. The strains carrying recombinant plasmids grew better than the control strain (BL21 transformed with empty vector).

Accumulation of protein was checked by re-suspending the cell pellets in sample buffer and separating the proteins using SDS-PAGE as described in methods section 2.8. The gel from Figure 3.3A shows the protein accumulation profile of the strain carrying the *mppA* construct, an IPTG inducible band is seen at around 54 kDa (whereas the expected size of MppA is 60 kDa) with almost same expression at all stages when grown at 25°C and 37°C . However the culture reaches a higher OD_{650} after 24 hours

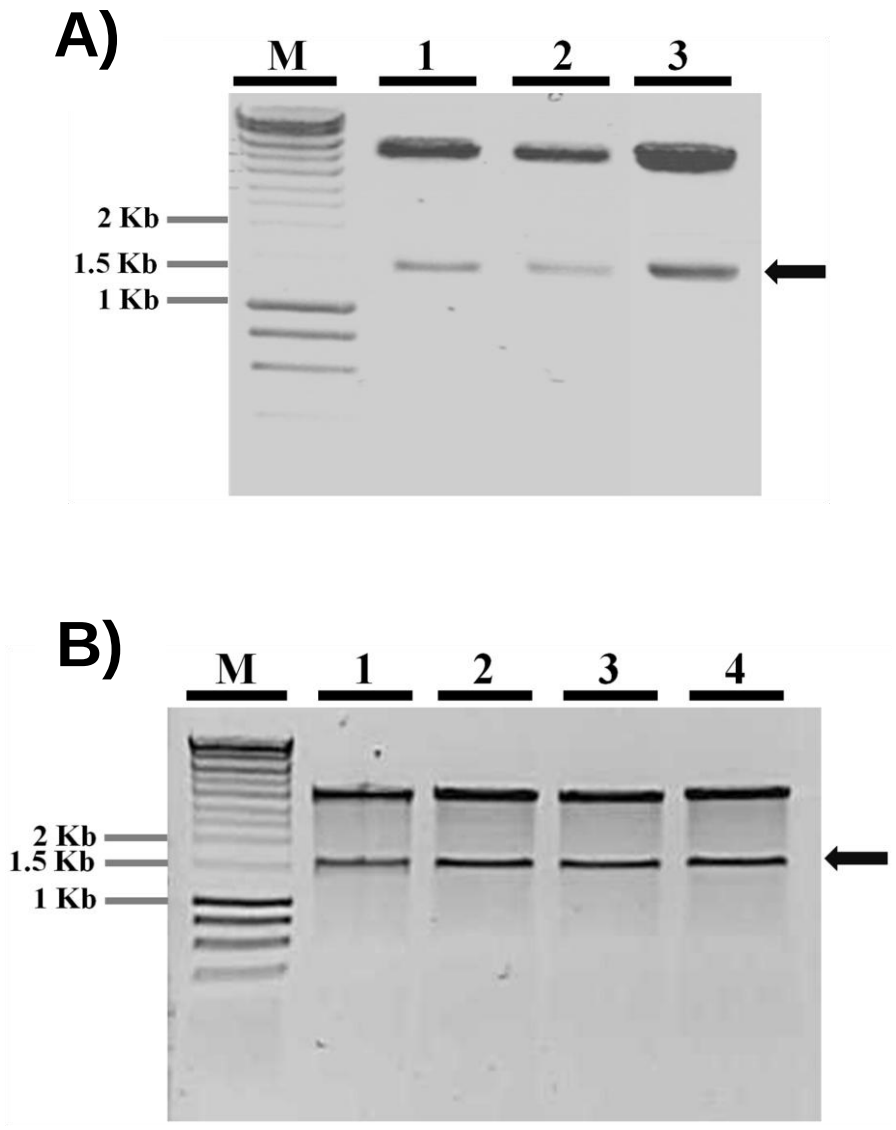


Figure 3.1 Agarose gel to verify recombinant plasmids

A) Agarose gel (1 %) stained with SYBR safe showing products of a restriction digestion of pET vector constructs after gel electrophoresis. Plasmids were doubly digested with NdeI and NcoI. M is Hyperladder I. Lanes 1, 2 and 3 show the products of restriction digestion of the vectors; pAM6091, pAM6092 and pAM6093 respectively and arrow represent the excised fragment of DNA resulting from double digestion of vectors. **B)** Agarose gel (1 %) stained with SYBR safe showing products of a restriction digestion of pBAD vector constructs after gel electrophoresis. Plasmids were doubly digested with BamHI and EcoRI. M is Hyperladder I. Lanes 1, 2, 3 and 4 show the products of a restriction digestion of the vectors; pBAD6094, pBAD6095, pBAD6096 and pBAD6097 respectively and arrow represent the excised fragment of DNA resulting from double digestion of vectors.

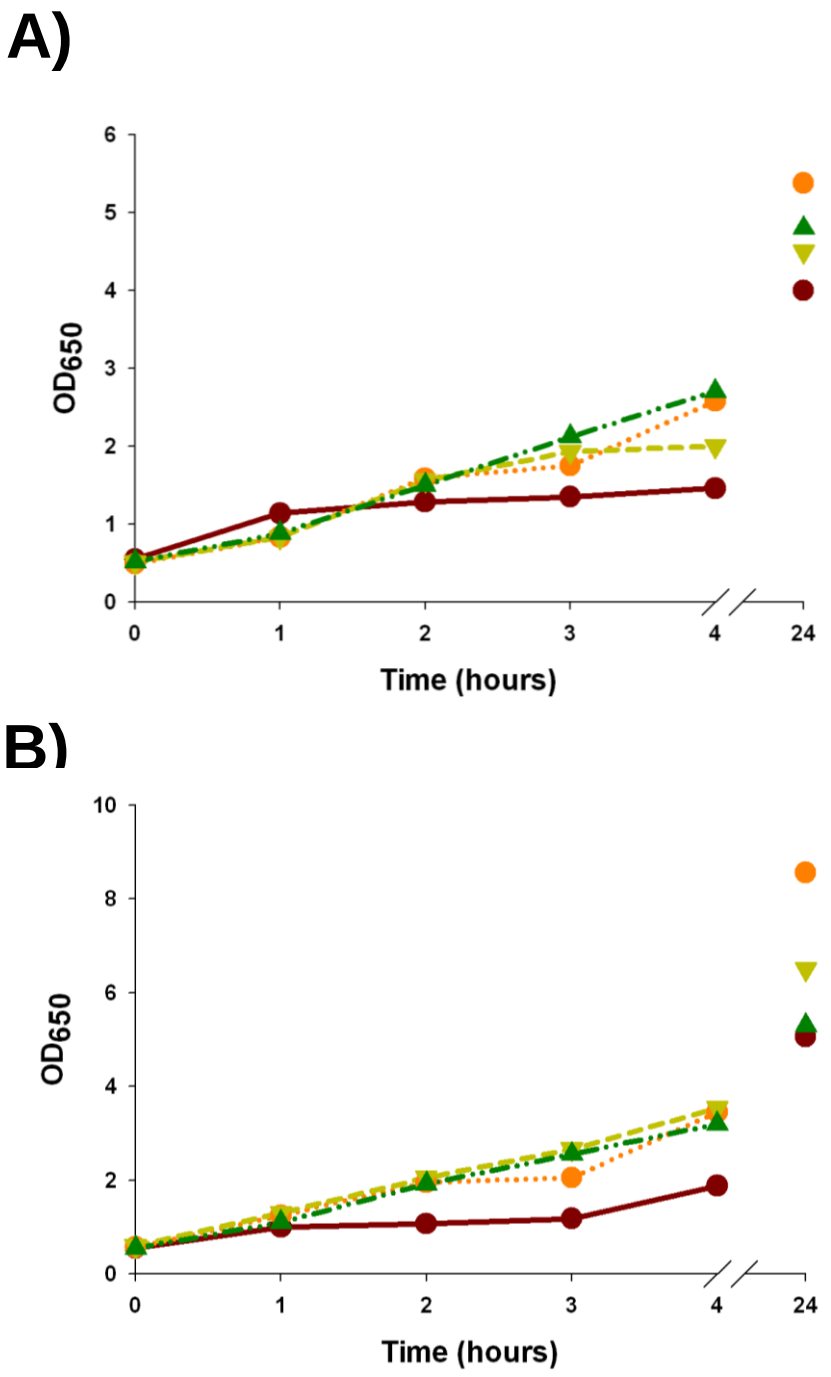


Figure 3.2 Growth curves of *E. coli* BL21 (DE3) strains used in expression trials

BL21 (DE3) containing plasmid was grown in a single growth experiment at 25°C **(A)** and 37°C **(B)** pAM6091 (MppA), pAM6092 (GsiB), pAM6093 (YgiS) and vector only (pET-YSBLIC3C)

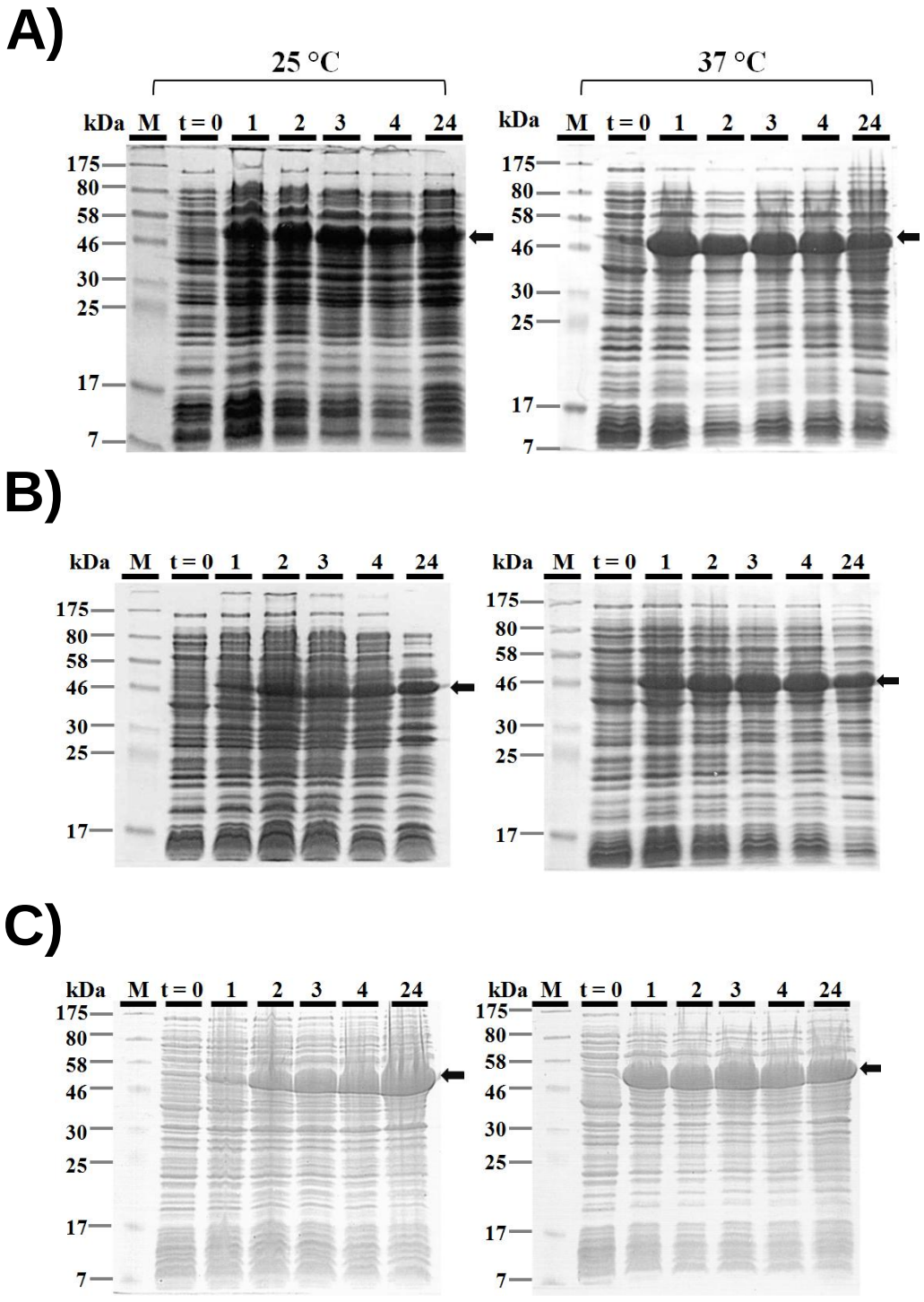


Figure 3.3 Expression of MppA, GsiB and YgiS from pET based constructs

Coomassie stained SDS-PAGE gels from expression trials conducted at 25°C (left) and 37°C (right). M is Pre-stained protein ladder, time (t) is equal to number of hours post induction and arrow indicates the position of an over produced protein on the gel.

A) Expression of MppA from BL21 (DE3) pAM6091 **B)** Expression of GsiB from BL21 (DE3) pAM6092 **C)** Expression of YgiS from BL21 (DE3) pAM6093

when grown at 37°C (Figure 3.2), so these conditions were chosen for future over expression of MppA to maximise yield. The SDS polyacrylamide gel of the strain carrying GsiB construct (Figure 3.3B) shows induced synthesis of a 50 kDa protein (whereas the expected size of GsiB is 56 kDa) with maximum accumulation of protein at 4 hours post-induction when grown at 37°C. The expression of GsiB was lower at each stage when grown at 25°C. YgiS shows an intense band around 54 kDa (whereas the expected size of YgiS is 60 kDa) in Figure 3.3C, after induction with IPTG. Maximum accumulation of YgiS occurs at 25°C with the overnight time point.

3.2.2 Expression trials of *mppA* and *gsiB* genes cloned in pBAD vectors

The recombinant plasmids (pBAD6094, pBAD6095, pBAD6096 and pBAD6097) were transformed into the expression strain MC1061 of *E. coli* and used in small scale expression trials. The expression trials were performed at two growth temperatures, 25°C and 37°C. Cultures were induced using 0.005% L-arabinose at an OD_{650} of 0.4-0.6. Growth was monitored by measuring OD_{650} after each hour starting from the time of induction and was continued for four hours post-induction with a final reading taken after 24 hours. At each time point 1 ml of culture was pelleted for later analysis by SDS-PAGE. The rate of growth for each target grown at 25°C and 37°C can be seen in Figure 3.4. It is evident that cultures of *E.coli* carrying constructs based on pBAD vector grow better and reach a higher OD_{650} than those carrying pET based constructs.

Accumulation of protein was checked by re-suspending the cell pellets in sample buffer and separating the proteins using SDS-PAGE as described in the methods section 2.8. The gel from Figure 3.5A shows the protein accumulation profile of the strain carrying the MppA construct (pAM6094) for its cytoplasmic expression, an arabinose inducible band is seen at around 54 kDa (where as the expected size of MppA is 60 kDa) with almost the same expression at all stages. Periplasmic accumulation of MppA was checked from the SDS-PAGE gel of the strain carrying the construct pBAD6096 (Figure 3.5B). Maximum level of protein accumulation can be seen in the culture grown for overnight post-induction. However the level of MppA synthesis resulting from the construct pAM6094 was lower than resulting from the construct pAM6096. The gel from Figure 3.5C shows cytoplasmic accumulation of GsiB in an *E. coli* strain carrying the

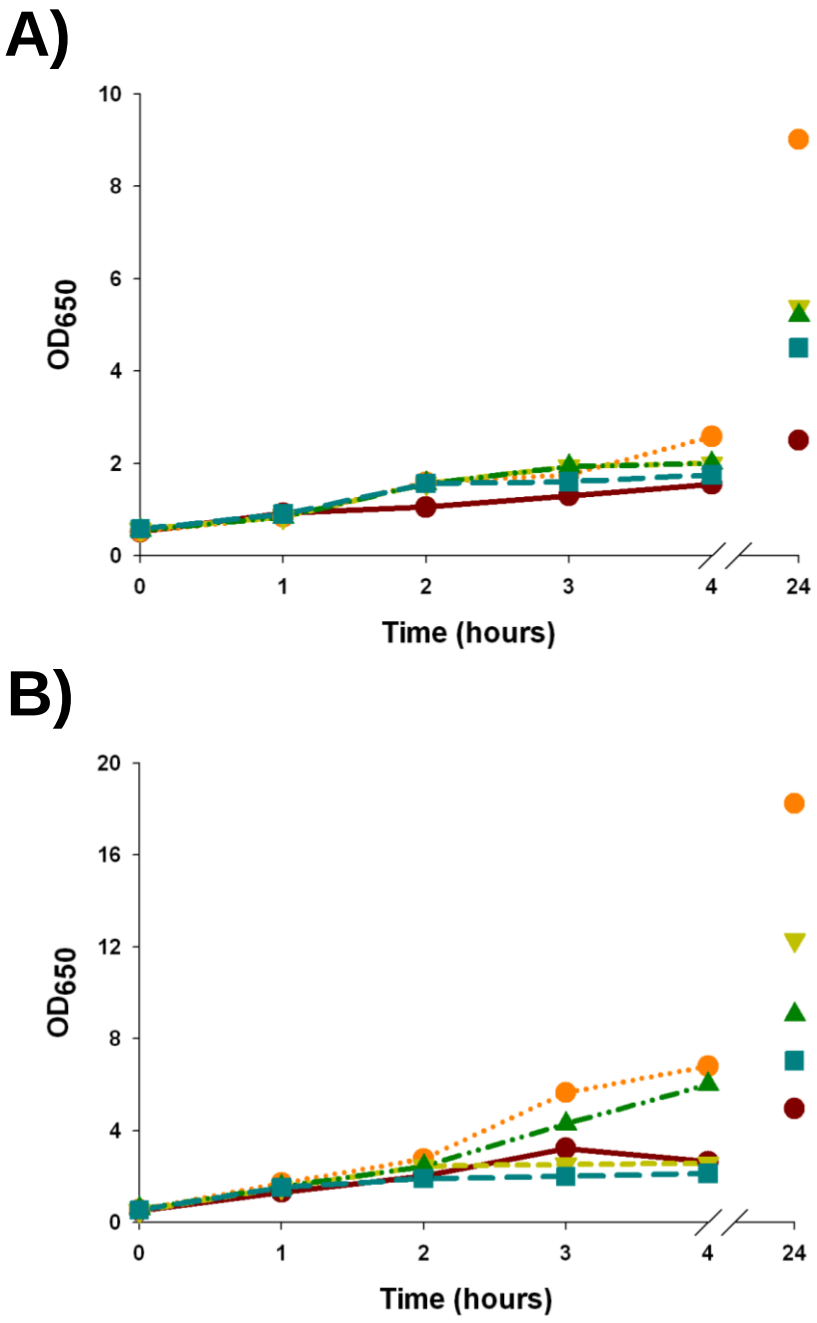


Figure 3.4 Growth curves of *E. coli* MC1061 strains used in expression trials

MC1061 containing plasmid was grown in a single growth experiment at 25°C **(A)** and 37°C **(B)** pAM6094 (Cytoplasmic MppA), pAM6096 (Periplasmic MppA), pAM6095 (Cytoplasmic GsiB), pAM6097 (Periplasmic GsiB) and vector only (pBADnLIC)

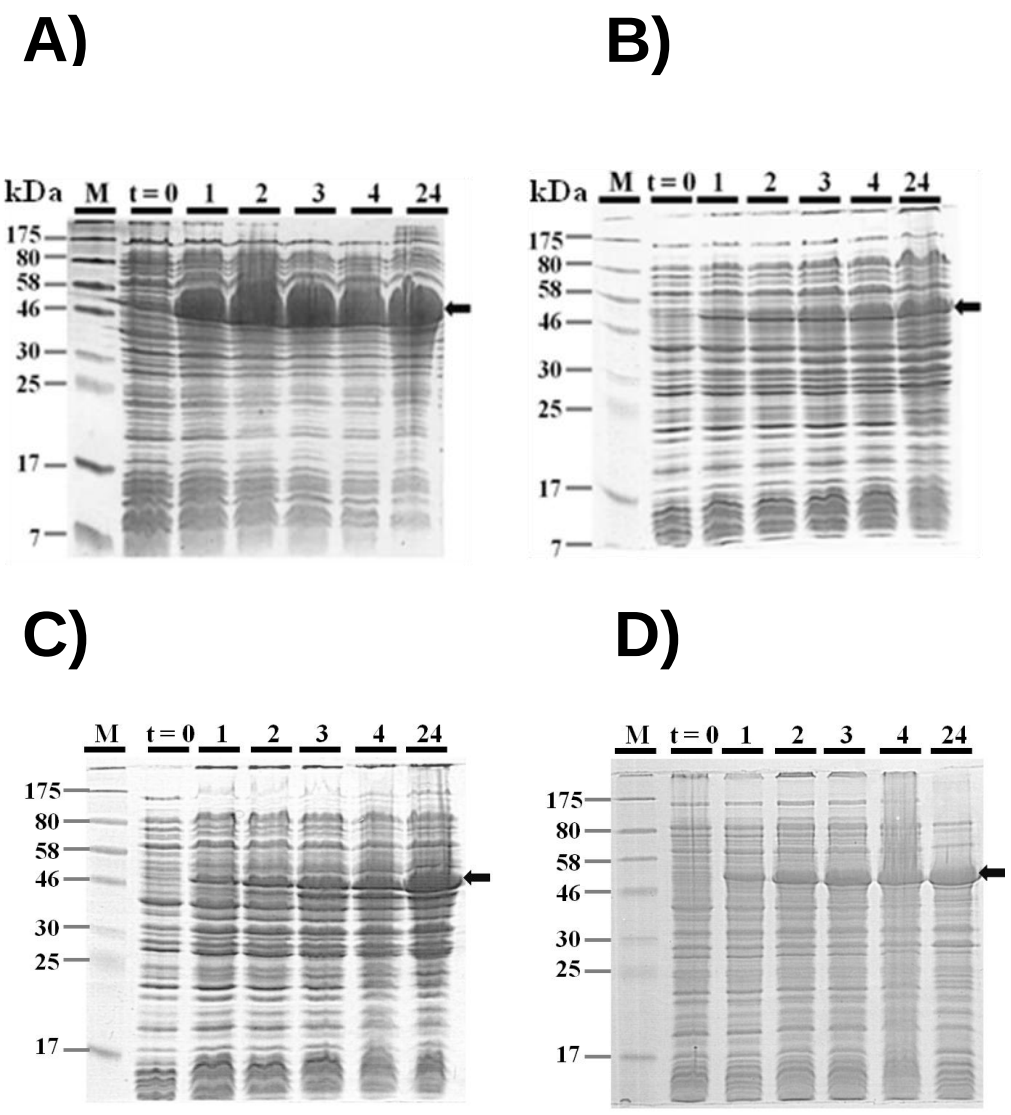


Figure 3.5 Expression of MppA and GsiB from pBAD vectors

Coomassie stained SDS-PAGE gels from expression trials conducted at 37°C. M is pre-stained protein ladder, time (t) is equal to number of hours post induction and arrow indicates the position of an over produced protein on the gel.

A) Cytoplasmic expression of MppA from MC10601 (pAM96094) **B)** Periplasmic expression of MppA from MC10601 (pAM96096) **C)** Cytoplasmic expression of GsiB from MC10601 (pAM96095) **D)** Periplasmic expression of GsiB from MC10601 (pAM96097)

construct (pAM6094). The band of over produced protein can be seen at around 50 kDa (whereas the expected size of GsiB is 56 kDa) with maximum accumulation from overnight grown culture at 37°C. Similar expression was also observed for GsiB targeted to the periplasm with a His tag on its C-terminus (Figure 3.5D).

Analysis of MppA and GsiB accumulation at 25°C demonstrated less accumulation of MppA and GsiB than at 37°C (data not shown).

The small scale expression trials show that strains of *E. coli* carrying pBAD vectors grow better and achieve higher OD_{650} than those carrying pET vectors. Among pBAD vectors, constructs producing cytoplasmic production of protein (pBAD6094 and pBAD6095) give higher level of accumulated protein than constructs producing periplasmic accumulation of protein (pBAD6096 and pBAD6097). Thus we decided to grow strains of *E. coli* carrying pBAD6094 and pBAD6095 on large scale for purification of recombinant protein.

3.3 Large scale expression and purification of MppA and GsiB using pBAD based expression constructs

Large scale expression was performed in 2 L flasks and cultures of *E. coli* MC1061 strains were grown under the optimum conditions determined during the small scale expression trials for each target. After induction, cells were grown overnight and harvested by centrifugation at 4430 x g for 15 min and resuspended in 50 mM KPi, pH 7.8. Cells were lysed by sonication and cell debris was removed by centrifugation at 38000 x g for 30 min at 4°C. The whole cell lysate containing over produced protein was collected. Nickel affinity chromatography was used for the purification of recombinant protein (MppA and GsiB) using gravity-flow columns (Methods section 2.6.1). The whole cell lysate containing the recombinant protein was incubated with equilibrated Ni^{2+}-NTA resin for 1 hour and weakly binding contaminants were eluted by washing the resin with a wash buffer containing 50 mM imidazole. Bound His-tagged protein was eluted with elution buffer containing 500 mM imidazole.

Analysis of the fractions taken during purification of protein using Coomassie-stained SDS-PAGE revealed an intense band in elution fraction 2 that corresponded to purified protein (Figure 3.6). An intense band was present in the whole cell lysate fraction,

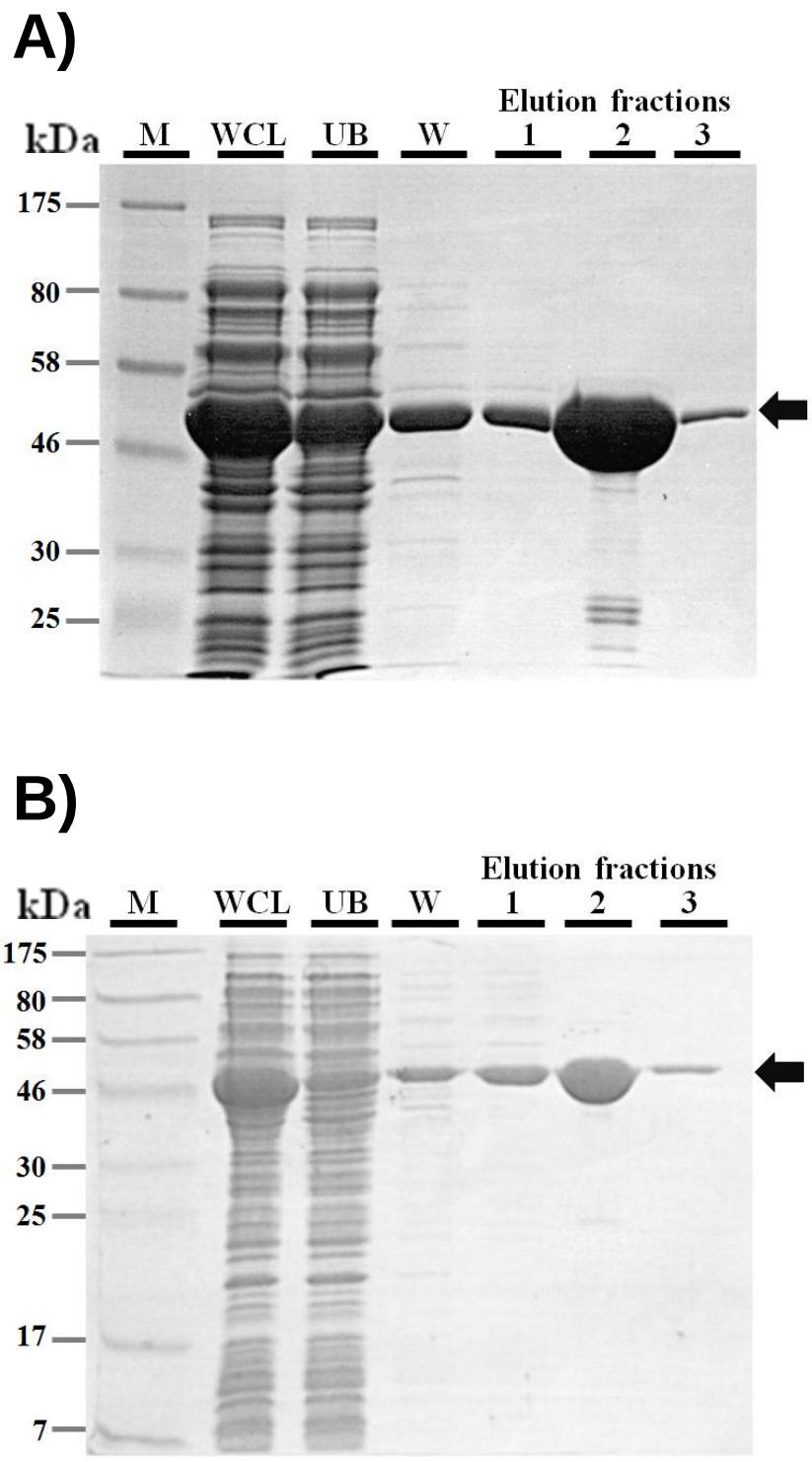


Figure 3.6 Purification of MppA and GsiB from cultures of *E. coli* MC1061

Coomassie stained SDS-PAGE of purification of MppA **(A)** and GsiB **(B)** by Ni-affinity chromatography showing clarified whole cell lysate (WCL), the flowthrough (UB), the 50 mM imidazole washing step (W) and the elution fractions containing purified proteins indicated by arrows. M is pre-stained protein ladder.

which was less intense in the flow through from the column, indicating that this band was over produced protein and was abundant enough to be visible in the whole cell lysate (Figure 3.6). The entire elution fractions containing purified protein were pooled together and the concentration of protein was determined using a NanoDrop ND-1000 spectrometer as described in methods section 2.9.2. The total yield obtained after Ni^{2+}-affinity chromatography was 35 mg/l and 20 mg/l for MppA and GsiB, respectively.

The ultimate objective of the project was to solve the crystal structure of the selected binding proteins using X-ray crystallography which requires production of protein crystals. It was advised by Prof. Tony Wilkinson (YSBL) that the His tag should be removed from purified protein. The recombinant proteins produced in pBAD vectors have a cleavage site of TEV protease to remove decahistidine tag. Purified proteins were incubated with increasing concentration of TEV protease for different time intervals and at different temperatures. His tag cleaved protein being smaller in size migrate faster than uncleaved protein on SDS-PAGE gels and can be distinguished from protein with decahistidine tag. Thus cleavage of decahistidine tag was checked by running SDS-PAGE gels (data not shown). The gels showed no cleavage of His tag. A possible reason could be that the proteins expressed in pBAD vectors are not good substrates for TEV protease.

Therefore, it was decided to purify proteins from strains of *E. coli* carrying pET based expression vectors. Recombinant proteins expressed in pET-YSBLIC3C vector have a cleavage site of HRV 3C protease to remove His tag. HRV 3C was routinely used in YSBL to cleave His tag from protein.

3.4 Large scale expression and purification of MppA, GsiB and YgiS using pET based expression constructs

Large scale expression was performed by growing *E. coli* cultures in 625 ml LB using two litre flasks under the optimum conditions determined during the small scale expression trials. N-terminal hexahistidine tagged proteins (MppA, GsiB and YgiS)

were purified using nickel affinity chromatography (Figure 3.7) as described previously in the section 3.3.

Analysis of the fractions taken during protein purification using Coomassie-stained SDS-PAGE revealed an intense band in elution fraction 2 that corresponded to purified protein (Figure 3.7A, B and C). An intense band was present in the whole cell lysate fraction, which was absent in the flow-through from the column, indicating that this band was recombinant protein and was abundant enough to be visible in the whole cell lysate (Figure 3.7).

The entire elution fractions containing purified protein were pooled together and the concentration of protein was determined using a NanoDrop ND-1000 spectrometer or by performing Bradford assay as described in methods section 2.9. The total yield obtained after Ni^{2+}-affinity chromatography was 26 mg/l, 15 mg/l and 23 mg/l for MppA, GsiB and YgiS, respectively.

3.5 Cleavage of hexahistidine tag by HRV 3C protease and purification of cleaved protein

Fractions containing purified protein were mixed with HRV 3C protease in a ratio of 1 to 100 while dialysing into 50 mM KPi overnight. Cleavage of the His tag was checked by running SDS-PAGE of cleaved and uncleaved protein. Cleaved protein being smaller migrated faster than un-cleaved protein on SDS-PAGE. An example of this phenomenon is shown for MppA in Figure 3.8 where cleaved protein can be distinguished from uncleaved sample. HRV 3C protease has His-tag on N-terminal and it binds with Ni^{2+}-NTA resin. This property of protease was exploited to separate it from cleaved protein using 5 ml His Trap column (GE Healthcare). At this stage proteins were approximately 99 % pure as judged by SDS-PAGE (Figure 3.9A, inset for MppA). Four residues from the affinity tag (GPAM) remain attached to the protein following HRV 3C cleavage. Fractions containing purified protein were pooled together and dialysed in an appropriate buffer for mass spectrometry.

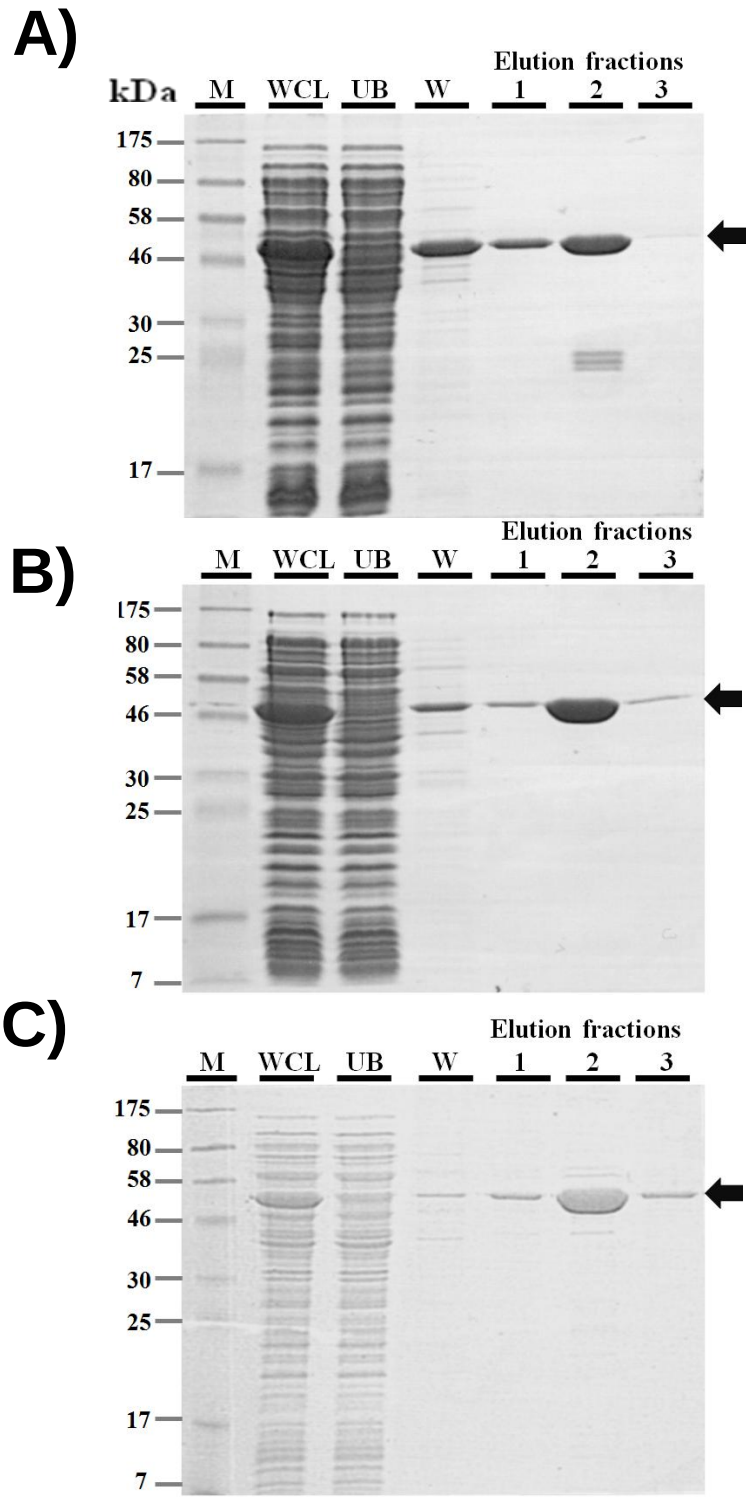


Figure 3.7 Purification of MppA, GsiB and YgiS from cultures of *E. coli* BL21 (DE3)

Coomassie stained SDS-PAGE of purification of MppA **(A)**, GsiB **(B)** and YgiS **(C)** by Ni-affinity chromatography showing clarified whole cell lysate (WCL), the flowthrough (UB), the 50 mM imidazole washing step (W) and the elution fractions containing purified proteins indicated by arrows. M is pre-stained protein ladder.

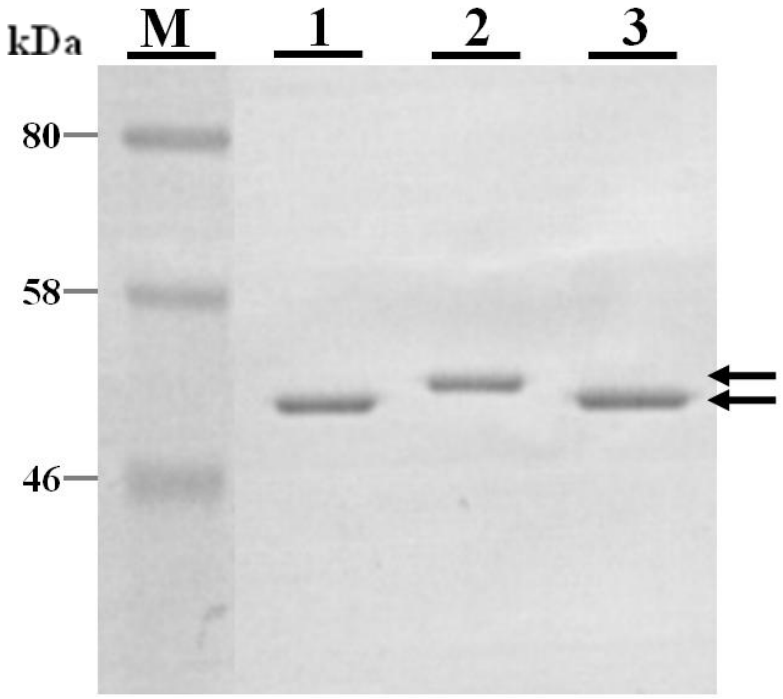


Figure 3.8 Cleavage of His tag from MppA

Coomassie-stained SDS-PAGE showing migration of cleaved and uncleaved MppA. M is pre-stained protein ladder. Lanes 1 and 3 represent MppA after cleavage of the His tag and lane 2 represent MppA with hexahistidine tag. Arrows on right show the position of cleaved and uncleaved MppA on gel.

3.6 Mass spectrometry of purified SBPs

Mass spectrometry is used to ensure that the recombinant protein produced and purified is the one expected and that it is not modified in any way during synthesis. As purified MppA, GsiB and YgiS migrated faster on SDS-PAGE gels than their expected sizes (Figure 3.3) suggesting they had been modified, electrospray mass spectrometry (ES-MS) was therefore used to ensure that these proteins expressed and purified are of the correct size and there is no proteolysis occuring. A Micromass LCT Premier XE mass spectrometer using a nanospray source was used for ES-MS of the proteins as described in the methods. Apart from determining the molecular weight of proteins, mass spectrometry was also used to check the presence of any co-purified ligand. These experiments used native electrospray mass spectrometry analysis, that is under conditions in which SBPs often retain endogenously bound ligands (Severi *et al*., 2005, Thomas *et al*., 2006).

3.6.1 Analysis of MppA by mass spectrometry

The spectrum recorded under denaturing conditions show a peak at 57979 ± 6 Da (Figure 3.9A). This is very close to the predicted molecular weight of MppA after removal of hexahistidine tag (57972 Da). The result shows that the recombinant protein was MppA and is of correct molecular weight. Previous genetic studies had suggested indirectly that MppA binds the murein tripeptide (Mtp, L-Ala-γ-D-Glu-*meso*-Dap). A mutant strain of *E. coli* was isolated which failed to grow in minimal media when Dap was replaced by Mtp, and this phenotype could be rescued by complementation with a plasmid carrying the *mppA* gene (Park *et al*., 1998). We used ES-MS under native conditions to identify if MppA purifies with an endogenous prebound ligand from *E. coli*. A native spectrum for MppA shows a clear peak in the mass spectrum that was not present in the denaturing spectrum (Figure 3.9B). This additional peak accounts for nearly half of the total protein signal. This species has a mass of 58361 Da, being 389 mass units larger than that of the ligand-free protein (57972 mass units in this sample). Given that the mass of Mtp is 389.38 Da, this is strong biophysical evidence that the protein binds a single molecule of Mtp selectively over other peptides present in the cytoplasm.

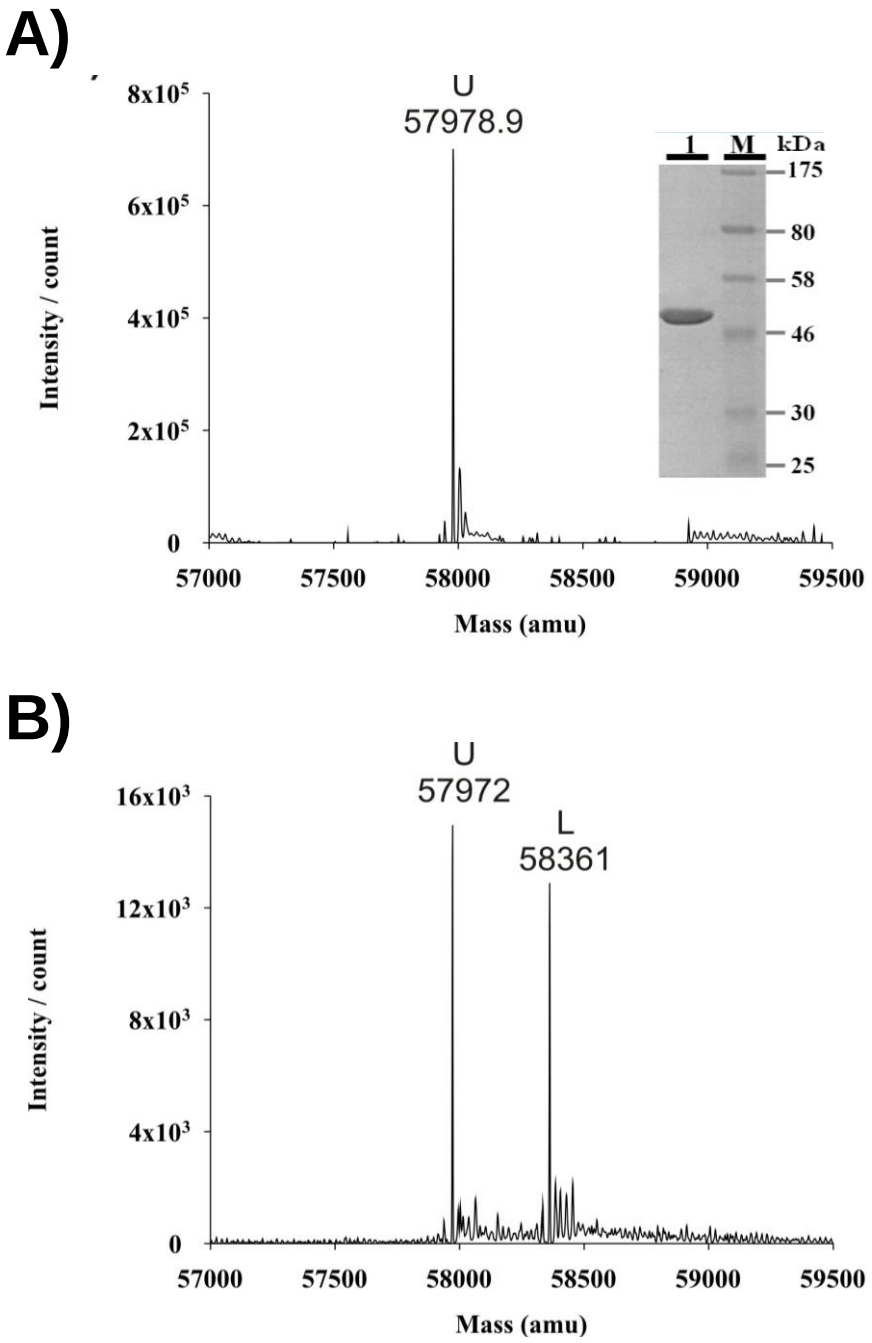


Figure 3.9 ES-MS spectrum of MppA

A) ES-MS spectrum of 5 μM MppA under denaturing conditions. The molecular mass of the unbound protein (U) is indicated at 57978.9 atomic mass units (amu). Inset: Coomassie stained SDS-PAGE gel containing 3 μg of purified MppA (lane 1) and molecular weight markers (M). **B)** ES-MS spectrum of MppA under native conditions in water. Peak (U) gives the molecular weight of ligand free MppA and peak (L) gives the molecular weight of MppA bound to a ligand (Mtp) of 390 Da. Other peaks are adducts of phosphates.

3.6.2 Analysis of GsiB by mass spectrometry

Using denaturing electrospray mass spectrometry the molecular weight of GsiB was determined to be 54261 ± 6 Da (Figure 3.10A) which is consistent with the predicted mass of the protein after removal of the hexahistidine tag (54255 Da). This clearly shows that purified protein is full length GsiB.

Genetic studies suggest that GsiB facilitates transport of glutathione, a tripeptide (γ-glutamylcysteinylglycine; GSH) (Suzuki *et al.*, 2005). We again used mass spectrometry under native conditions to check the presence of any endogenously bound ligand to GsiB. A native spectrum of purified GsiB shows an additional peak that was not present in the denaturing spectrum (Figure 3.10B). The peak has a mass of 54564 Da, being 309 Da larger than that of ligand free protein. This result shows that GsiB has co purified with ligand of 309 Da. Given that the mass of GSH is 307.3 Da, this is biophysical evidence that supports the hypothesis that GsiB binds GSH selectively in a myriad of peptides in the cytoplasm.

3.6.3 Analysis of YgiS by mass spectrometry

The molecular weight of YgiS was determined to be 58782 ± 6 Da using denaturing ES-MS (Figure 3.11A) which is in agreement with the theoretical mass of YgiS after removal of the hexa histidine tag (58784 Da). Unlike MppA and GsiB, YgiS did not show any additional peak in the spectrum recorded under native conditions (Figure 3.11B). This means that either YgiS does not bind a peptide or it binds to a specific non-endogenous peptide not seen in the growth medium.

3.7 Detection of protein ligand complexes using FT-ICR-MS

Although results from ES-MS gave an indication of the ligands bound to SBPs, these were further confirmed using the high mass accuracy of FT-ICR-MS. The accuracy of this technique results from the extremely high resolution achievable (> 100,000 at full width half maximum). At these resolutions it is possible to measure the monoisotopic masses of proteins to well within half a Dalton, whereas conventional ES-MS will typically give the average mass of proteins, also called apex mass, with precisions of

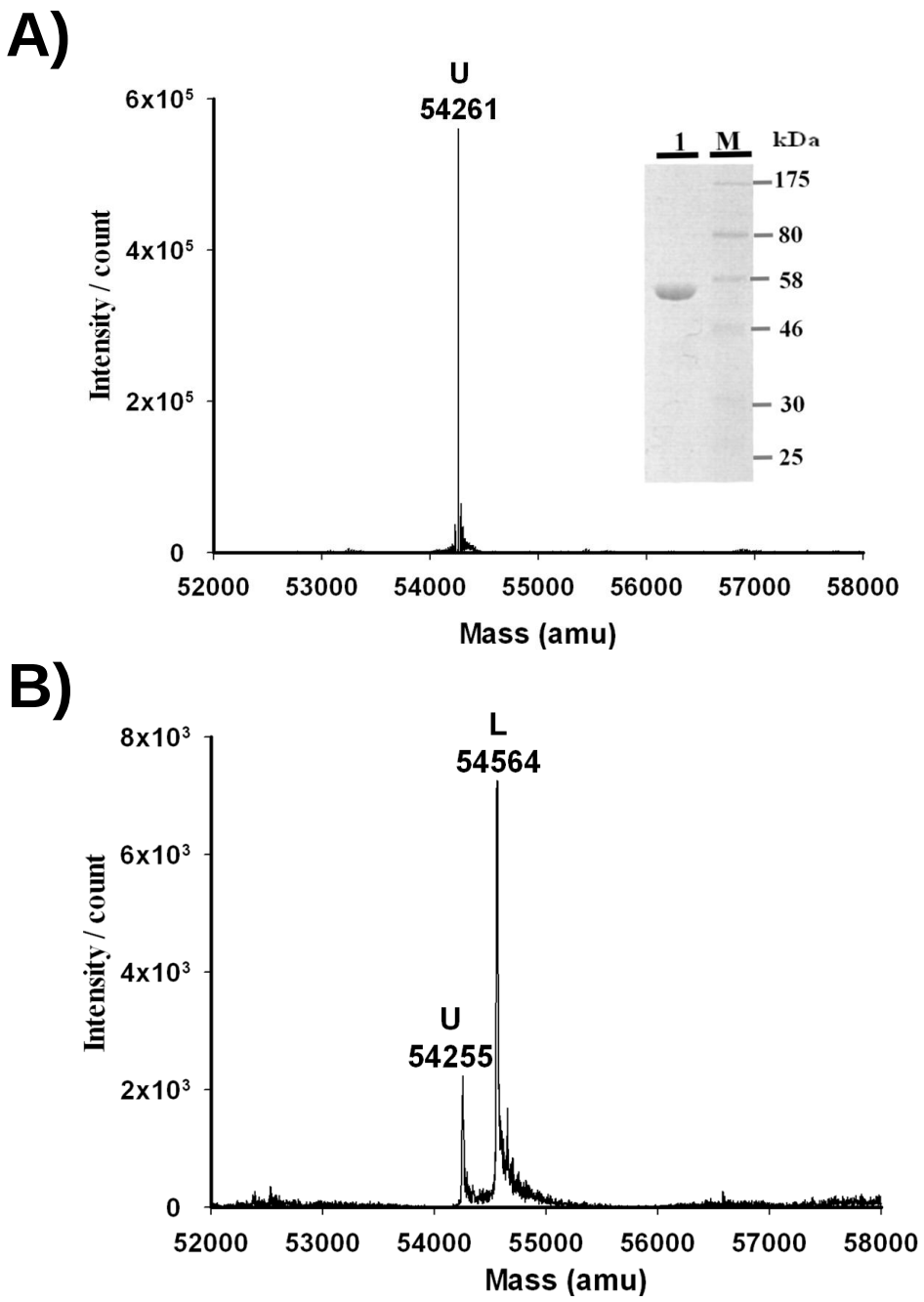


Figure 3.10 ES-MS spectrum of GsiB

A) ES-MS spectrum of 5 μM GsiB under denaturing conditions. The molecular mass of the unbound protein (U) is indicated at 54261 atomic mass units (amu). Inset: Coomassie stained SDS-PAGE gel containing 3 μg of purified GsiB (lane 1) and molecular weight markers (M). **B)** ES-MS spectrum of GsiB under native conditions in water. Peak (U) gives the molecular weight of ligand free GsiB and peak (L) gives the molecular weight of GsiB bound to a ligand (GSH) of 307 Da. Other peaks are adducts of phosphates.

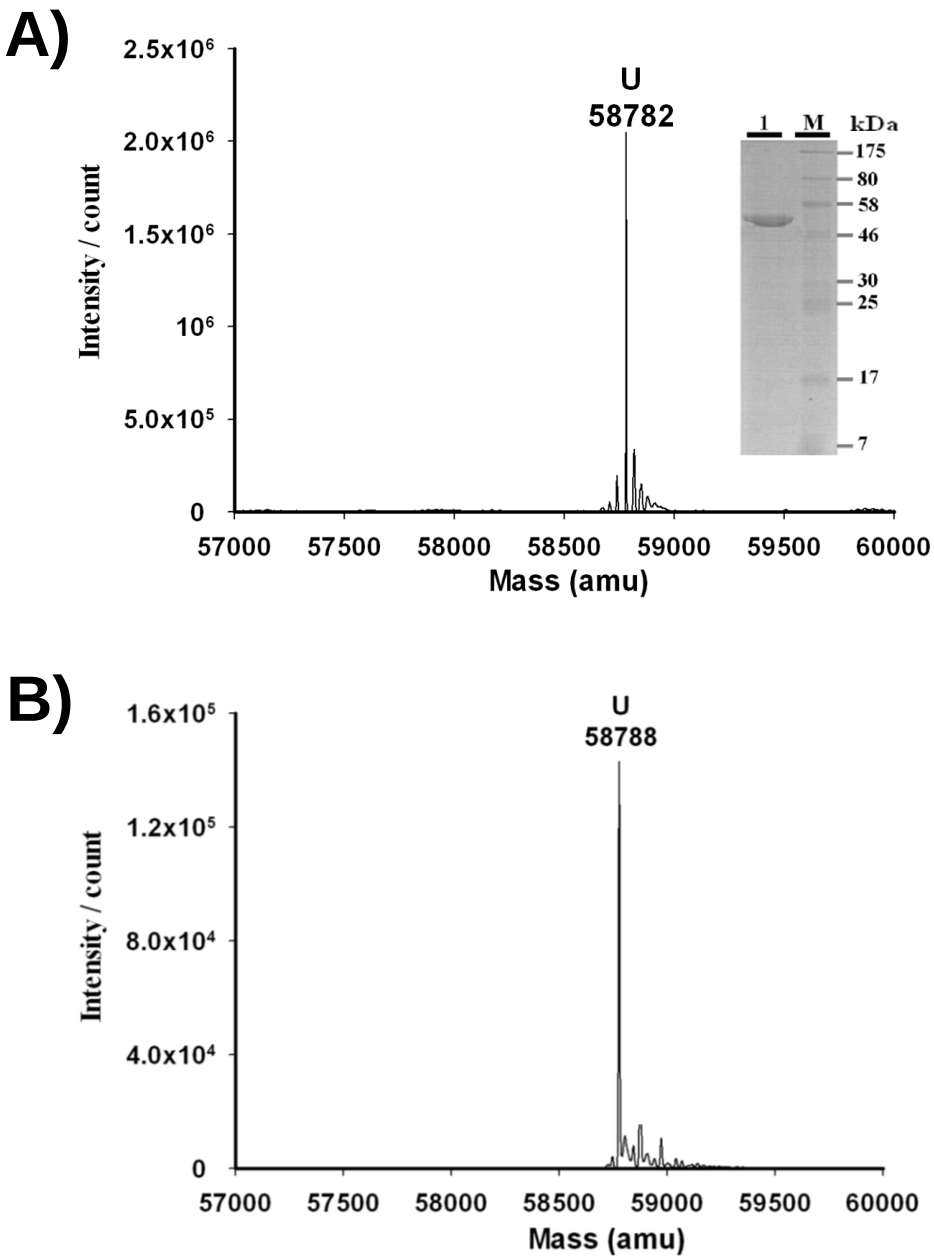


Figure 3.11 ES-MS spectrum of YgiS

A) ES-MS spectrum of 5 μM YgiS under denaturing conditions. The molecular mass of the unbound protein (U) is indicated at 58782 atomic mass units (amu). Inset: Coomassie stained SDS-PAGE gel containing 3 μg of purified YgiS (lane 1) and molecular weight markers (M). **B)** ES-MS spectrum of YgiS under native conditions in water. Peak (U) gives the molecular weight of ligand free YgiS.

less than one Dalton being difficult to achieve. FT-ICR-MS affords even greater precision with lower mass analytes such as peptides, where a mass accuracy of sub milli Dalton is routinely achievable.

We first used FT-ICR-MS to study MppA and obtained an accurate mass for the protein ($[M+H]^+$ = 57940.10 *m/z*, mass error = 0.2 Da). A clear signal for the released ligand was also observed (Figure 3.12A). The accurately measured mass of the ligand ($[M+H]^+$ = 391.1827 *m/z*) is consistent with only one possible empirical formula (considering: carbon, hydrogen, nitrogen and oxygen atoms) within a mass tolerance of 3 ppm, that of Mtp (Figure 3.12B) (mass error = 0.4 mDa). Fragmentation of this peptide using sustained off-resonance irradiation (SORI) enabled detection of the C-terminal diaminopimelate (Dap) residue (observed at 191.1027 *m/z*, mass error = 0.1 mDa), further confirming the identification of Mtp (Figure 3.12B).

We also demonstrated that GsiB is able to bind to GSH using the same method. In a typical FT-ICR-MS experiment the molecular weight of GsiB was accurately measured at $[M+H]^+$ = 54222.532 *m/z*, mass error = 0.03 Da. We saw two clear peaks for the released ligands; (Figure 3.13B) a peak at 308.09129 *m/z* is in agreement with the predicted monoisotopic mass of GSH ($[M+H]^+$ = 308.09108 *m/z*, mass error = 0.21 mDa). This is consistent with ES-MS, which shows a ligand of mass 309 Da bound to GsiB. A second peak represents another species at 613.15918 *m/z*, which was not seen by conventional ES-MS. This species was identified as glutathione sulphide (GSSG) on the basis of its accurately measured mass (mass error = 0.14 mDa). GSSG is an oxidized form of glutathione derived from two glutathione molecules. GSSG may either result from oxidation of two glutathione molecules during the FT-ICR-MS experiment, or alternatively be prebound with GsiB. The later possibility suggests that GsiB may also bind to GSSG.

For these peptide-binding SBPs the use of FT-ICR-MS has been shown to be a much more powerful technique than conventional ES-MS. This is primarily because we were able to measure the masses of ligands with a sub milli Dalton level of accuracy.

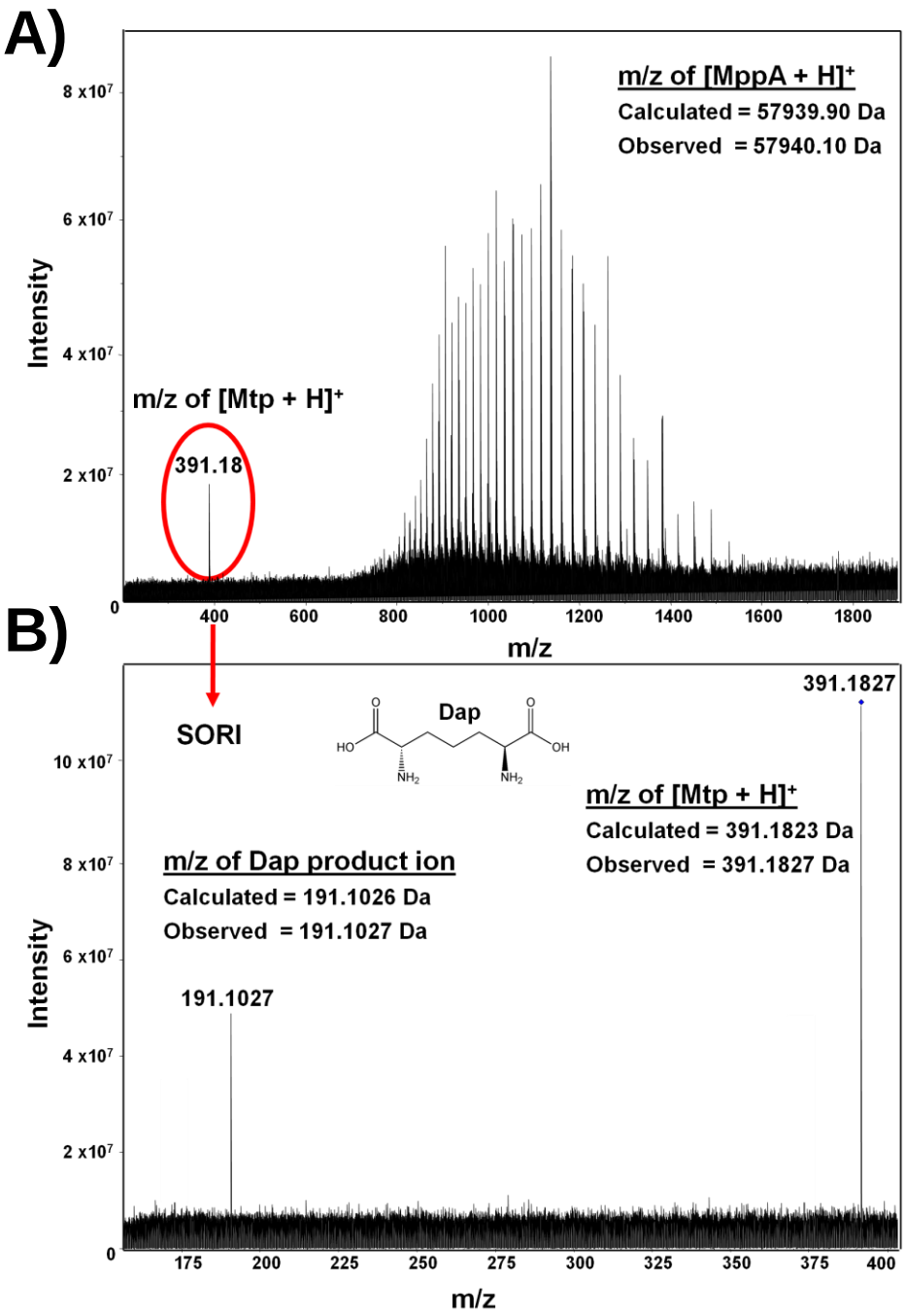


Figure 3.12 FT-ICR-MS of MppA-Mtp complex

A) FT-MS spectrum of 10 μM MppA. The peaks on right side represent multiple charge species of MppA. A small peak on left side represents the released peptide ligand (Mtp). **B)** Quad isolation and sustained off-resonance irradiation of Mtp to establish the presence of Dap upon fragmentation.

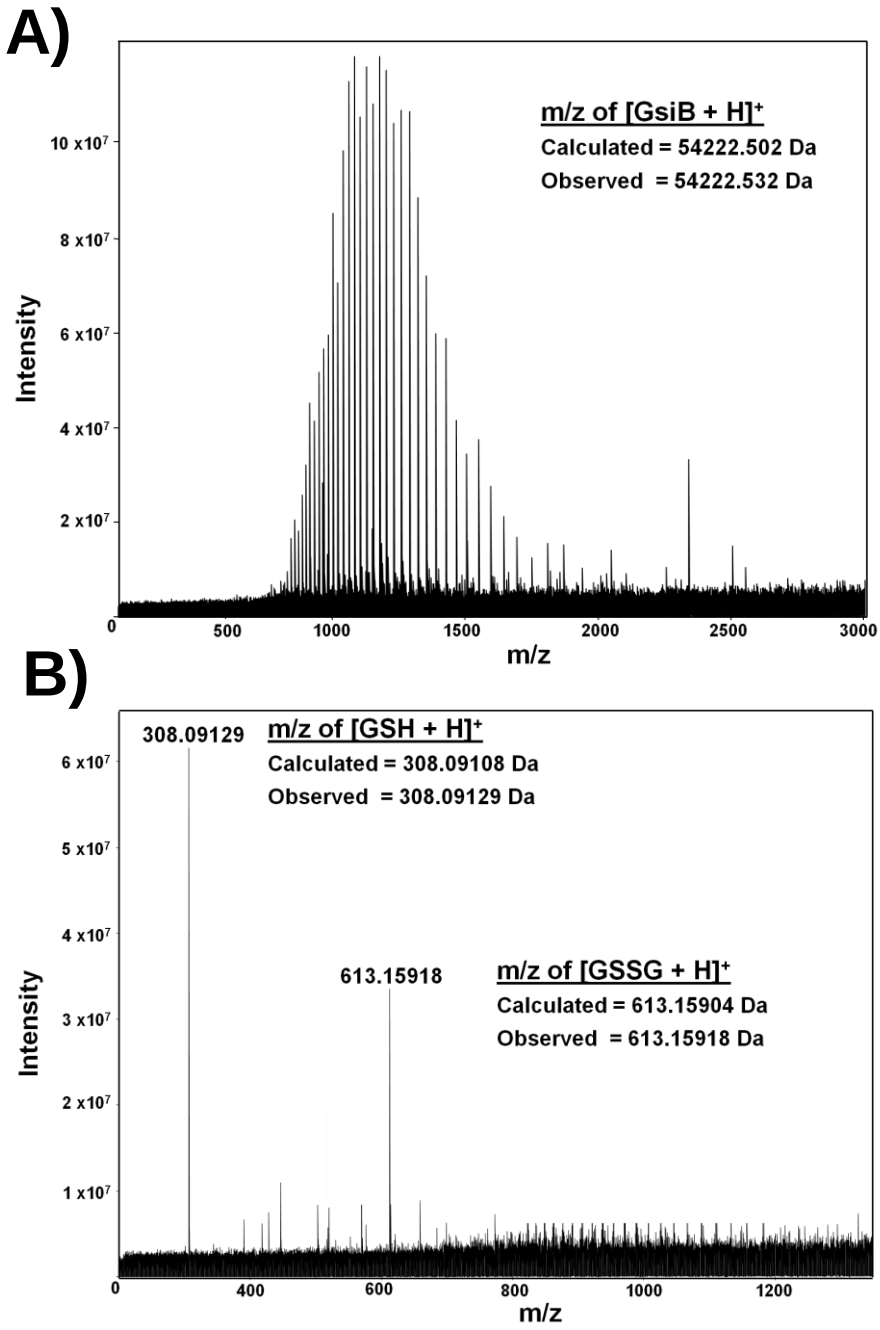


Figure 3.13 FT-ICR-MS of GsiB

A) FT-MS spectrum of 10 μM GsiB. The peaks represent multiple charge species of GsiB. **B)** FT-MS spectrum of GsiB optimized for low mass showing the released peptide ligands (GSG and GSSG).

3.8 Conclusion

The genes encoding SBPs MppA, GsiB and YgiS were successfully cloned into expression vectors. Growth conditions were optimized for achieving stable and higher production of proteins and all the targets were successfully purified using Ni-affinity chromatography. Histidine tags were cleaved off from purified proteins and tag free proteins were purified to almost 99% purity. Mass spectrometric analysis showed that MppA and GsiB were purified with their cognate ligands and thus selected for crystallization experiments. However we could not identify the ligand for YgiS and the protein was not taken forward for crystallization studies.

Chapter 4

4 Crystallization of SBPs and solving the X-ray structure of MppA

4.1 Protein crystallography

As the ultimate aim of my project was to understand the structural basis of ligand binding in peptide-binding proteins, I used protein crystallography to solve the structures of peptide-binding proteins in the presence of their cognate ligands. Mass spectrometric analysis of purified protein from chapter 3 suggested that MppA and GsiB were co-purified with their cognate peptide ligands. These proteins were dialysed into appropriate buffers and concentrated to between 10-60 mg/ml for crystallization.

4.1.1 Growing MppA crystals

For crystallization, MppA was concentrated to 16 mg/ml in 20 mM MES, pH 6.5. Preliminary crystallization trials were performed using a range of commercially available screens; PACT, Hampton I, II, Index and Clear Strategy Screens CSSI and CSSII, by hanging-drop vapour diffusion procedures in 96-well plates setup with a Mosquito nanolitre pipetting robot as described in methods. The plates were sealed to avoid evaporation and then incubated at 20 ºC and examined regularly for the appearance of crystals. Initial crystallization trials resulted in crystals obtained under three different conditions. Two of these conditions where the crystals were very small in size (Figure 4.1A) were further optimized. Crystals grown with 0.1 M Tris pH 8, 10 mM zinc chloride (PACT condition 48) did not reproduce while optimization of crystals grown with 0.1 M bis-Tris pH 6.5, 2M ammonium sulphate (Index Condition 4) led to crystals with poor diffraction quality of 4.5 Å resolution. Optimisation experiments involved varying the pH, precipitant concentrations, volume ratio of protein and crystallization conditions, as well as the MppA concentration which was increased from 16 mg/ml to 27 mg/ml. One condition produced single crystals that were big enough for X-ray data collection. These crystals grew with 50 mM zinc acetate and 20% w/v PEG3350 (Index condition 93) and had dimensions of 0.1 x 0.1 x 0.1 mm^3 (Figure 4.2A). One of the crystals was transferred to a precipitant solution containing 25% glycerol for cryoprotection, flash cooled in liquid nitrogen and checked for diffraction (Figure 4.2). The crystal diffracted to 2.6 Å when tested for X-ray diffraction at the YSBL home source (Rigaku MicroMax 007HF generator) and was stored in liquid nitrogen in order to collect X-ray data on a beamline at a synchrotron facility.

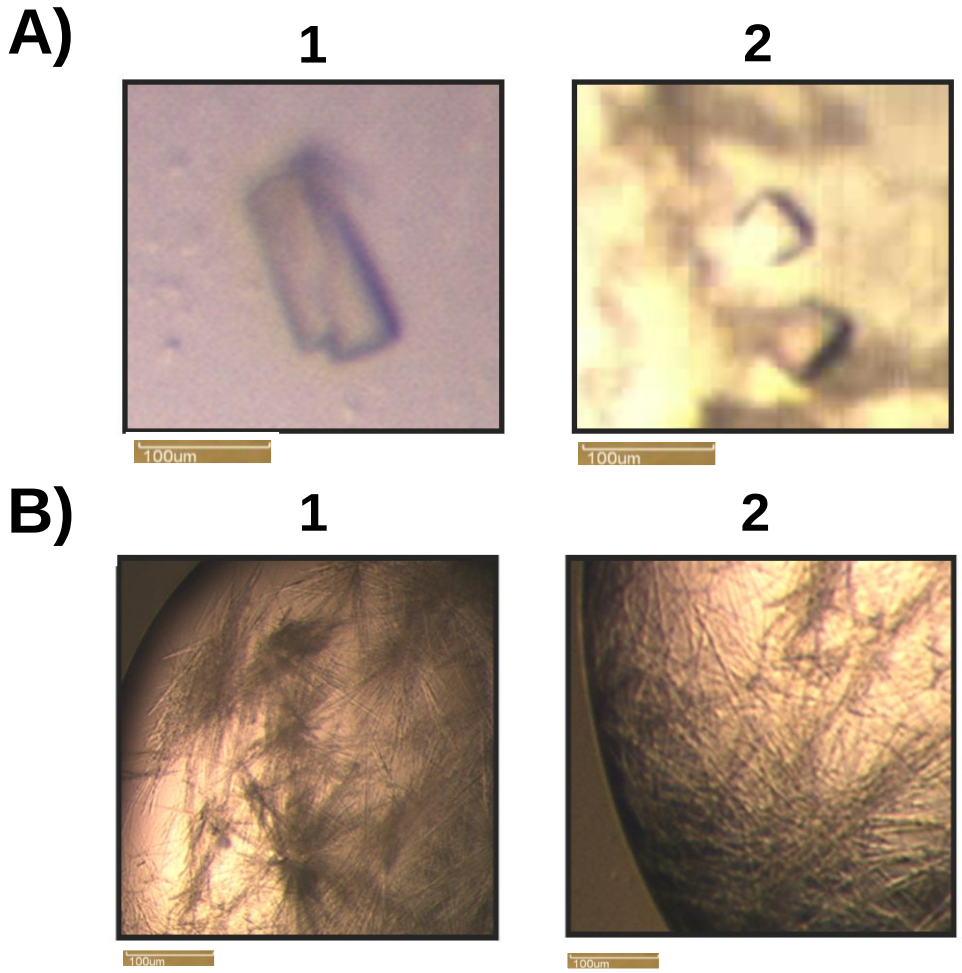


Figure 4.1 Crystallization trials of MppA and GsiB

Images of crystals for target proteins MppA (A) and GsiB (B) taken while crystals growing in 96 well plates

A1 = Crystals of MppA in PACT condition 48
A2 = Crystals of MppA in Index condition 93
B1 = Crystals of GsiB in PACT condition 7
B2 = Crystals of GsiB in HCS II condition 25

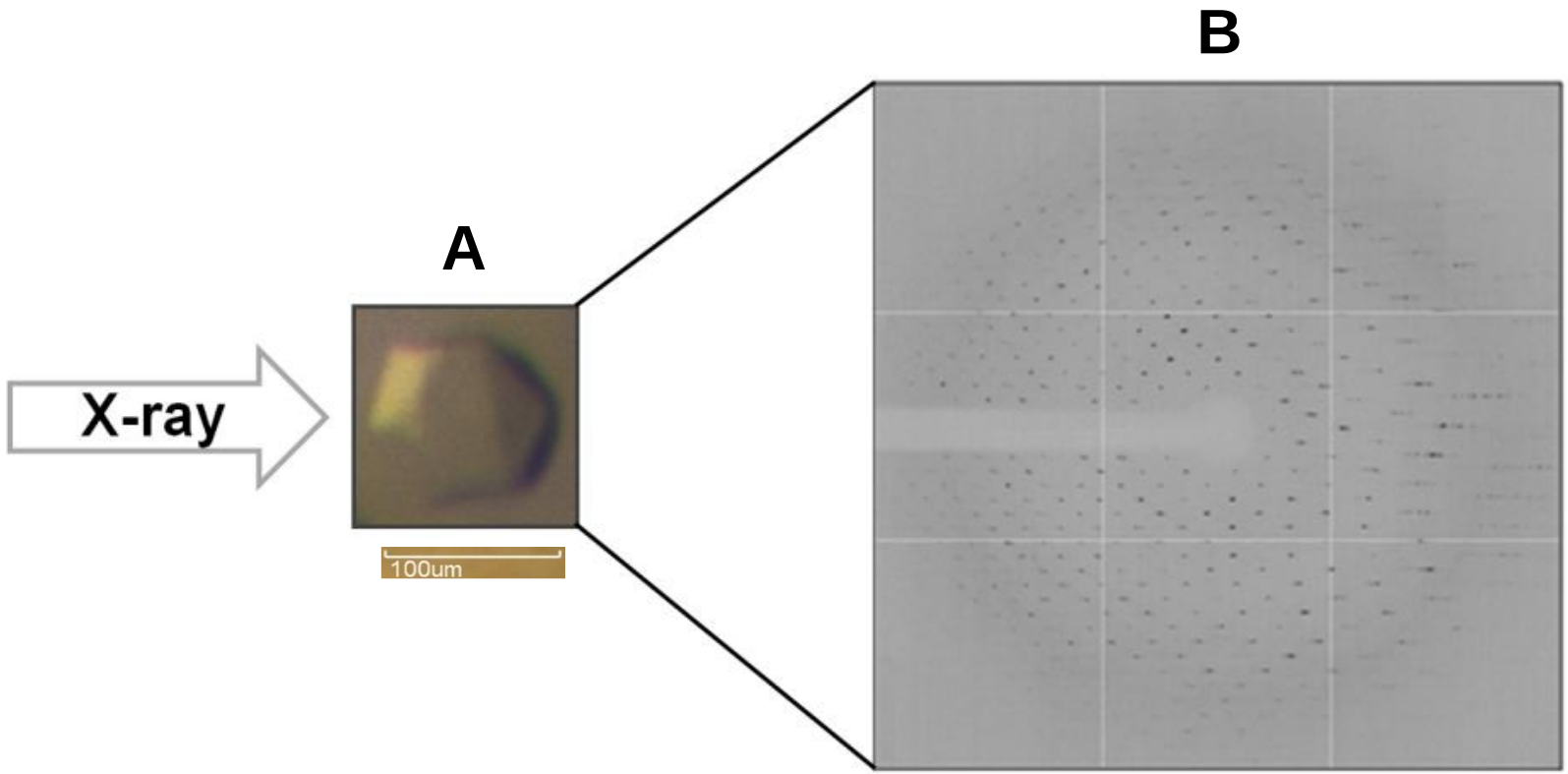


Figure 4.2 X-ray diffraction from MppA crystal

A) The best crystal of MppA which grew in 50 mM zinc acetate and 20% w/v PEG3350. **B)** The diffraction image obtained at the ESRF station ID14-4.

4.1.2 Growing GsiB crystals

GsiB was concentrated to 60 mg/ml in 20 mM MES, pH 6.5 and subjected to crystallization trials using Mosquito nanolitre pipetting robort. Commercial screens (PACT, Hampton, CSS, and Index) were employed initially and crystallization trials were conducted using two concentrations of protein, i.e. 30 mg/ml and 60 mg/ml. This resulted in needle like crystals of GsiB under two different conditions; 0.1 M sodium acetate, pH 5 and 20% w/v PEG6000 (PACT condition 7) and 1.8 M ammonium sulphate, 0.1 M MES, pH 6.5 and 0.01 M cobaltous chloride hexahydrate (Hampton Crystal Screen II condition 25). These crystals were too small to be used for diffraction experiment (Figure 4.1B). Much subsequent effort was made to improve the crystals by trying various modifications like seeding, varying the concentration of the protein, varying buffer composition and buffer pH, etc. None of these strategies produced good enough crystals to be used for diffraction. Some crystallization experiments were also conducted at 4 °C but with no success.

4.2 Solving the X-ray structure of MppA

The crystal of MppA that diffracted to 2.6 Å at YSBL home source was sent to European Synchrotron Radiation Facility (ESRF), Grenoble for data collection. For MppA X-ray data were collected at ESRF station on beam line ID14-4 from a single crystal. Data collection was performed at 100 K using an oscillation range of 0.5° per image with a total crystal rotation of 180°. The diffraction images were indexed and integrated with DENZO and scaled with SCALEPACK (Otwinowski and Minor, 1997). Indexing permitted the determination of the crystal system, the unit cell dimensions and assignment to each spot on the image an index, quoted as three integers: *h*, *k*, and *l*. Data collection and processing statistics are summarised in Table 4.1.

A homology model was available for MppA in the form of oligopeptide binding protein (OppA) from *Yersinia pestis*, thus the structure of MppA was determined by the molecular replacement method using these coordinates (PDB code 2Z23) as a search

Table 4.1 X-ray data statistics for the crystal of MppA

Data collection and processing statistics	
X-ray source	ID14-4, ESRF
Wavelength (Å)	0.9335
Temperature (K)	100
Space group	P6
Unit cell parameters (Å)	a=b=165.1 Å, c=38.2 Å α=β=90, γ=120
Resolution range (Å)	47.65-2.07
Number of unique reflections, overall/outer shell[a]	34597 (633)
R_{merge}[b] (%), overall/outer shell[a]	10.7 (55.3)
Completeness (%), overall/outer shell[a]	92 (53.5)
Redundancy , overall/outer shell[a]	9.7 (4.3)
$I/\sigma(I)$, overall/outer shell[a]	20.6 (1.6)

[a]The outer shell corresponds to 2.10 – 2.06 Å.

[b]$R_{merge} = \sum_{hkl}\sum_i|I_i - \langle I\rangle| / \sum_{hkl}\sum_i \langle I\rangle$ where I_i is the intensity of the ith measurement of a reflection with indexes *hkl* and $\langle I\rangle$ is the statistically weighted average reflection intensity.

model. Molecular replacement was carried out using the MOLREP program (Vagin and Teplyakov, 2000). The first peak of rotation function (RF) had a sigma value of 6.67 which was distinctive from other RF peaks (for example second peak was at 4.3 and all others were below 4). Similarly, the first peak of the translation function had a sigma value of 38.52 with all other sigma values less than 5. The correctness of the solution was also reflected in the high contrast factor of 12.77. Manual model building was carried out in the program COOT (Emsley and Cowtan, 2004) and the structure was refined using REFMAC5 (Murshudov *et al.*, 1997) to give final R_{factor} and R_{free} values of 16.7% and 21.9% respectively. The prefinement statistics are given in the Table 4.2.

The quality of our model was checked by PROCHECK analysis from PDBSum page. The Ramachandran plot thus generated put more than 90% of amino acid residues of protein in the most favoured regions. The statistics of Ramachandran plot are given in Table 4.2. The final set of coordinates and structure factors were deposited in the RCSB Protein Data Bank under the accession code 309P.

4.3 Overall structure of MppA

The open reading frame coding for *E. coli* MppA encodes a 537 residue protein that contains a 22 residue N-terminal signal peptide that is cleaved off after export through the cytoplasmic membrane. Thus mature protein contains 515 amino acids starting from Ala 1 to the C-terminal residue His 515. As we crystallized the mature form of MppA, the model of MppA structure was built from Ala 1 to His 515. The structure of MppA shows a continuous polypeptide chain from Thr-7 to His-515 (Figure 4.3). The first 6 residues (AEVPSG) are not represented in the electron density maps and are assumed to be disordered. The protein is organized into three structural domains, with domain 1 encompassing residues 7-46, 165-263 and 487-515, domain 2 encompassing residues 264-486 and domain 3 encompassing residues from 47-164 (Figure 4.3). The overall fold of MppA is similar to that of the oligopeptide binding protein as discussed below. A DALI search (Holm *et al.*, 2008) of the protein data bank (PDB) revealed that MppA has highest structural similarity to the oligopeptide binding proteins (OppAs) of *Yersinia pestis* (PDB code 2Z23; Z-score = 55.3, with 508 equivalent Cα atoms superimposing with an rmsΔ of 1.1 Å) and *Salmonella* Typhimurium (PDB code 2OLB; Z-score = 55.2, with 509 equivalent Cα atoms superimposing with an rmsΔ of 1.0 Å). This structural similarity spans the full length of the protein (Figure 4.4).

Table 4.2 MppA structure refinement statistics

Refinement statistics	
R_{factor}[a] (R_{free}[b])	16.7 (21.9)
Reflections (working / free)	32581 / 1743
Outer shell[c] R_{factor}[a] (R_{free}[b])	24.2 / (30.4)
Outer shell[c] reflections (working / free)	1697 / 74
Molecules / asymmetric unit	1
Number of protein nonhydrogen atoms	4046
Number of ligand atoms	50
Number of water molecules	366
rmsd bond lengths(Å)[d]	0.02
rmsd bond angles(Å)[d]	1.84
Average B-factor (Å^2)	25
Ramachandran plot[e]	91.3/8.5/0.2/0.0

[a]$R_{factor} = \Sigma||F_o|-|F_c|| / \Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

[b]R_{free} is the R_{factor} calculated with 5% of the reflections chosen at random and omitted from refinement.

[c]Outer shell for refinement corresponds to 2.073–2.126 Å.

[d]Root mean square deviation of bond lenths and bond angles from ideal geometry

[e]Percentage of residues in most favored/additionally allowed/generously allowed/disallowed regions of the Ramachandran plot, according to PROCHECK.

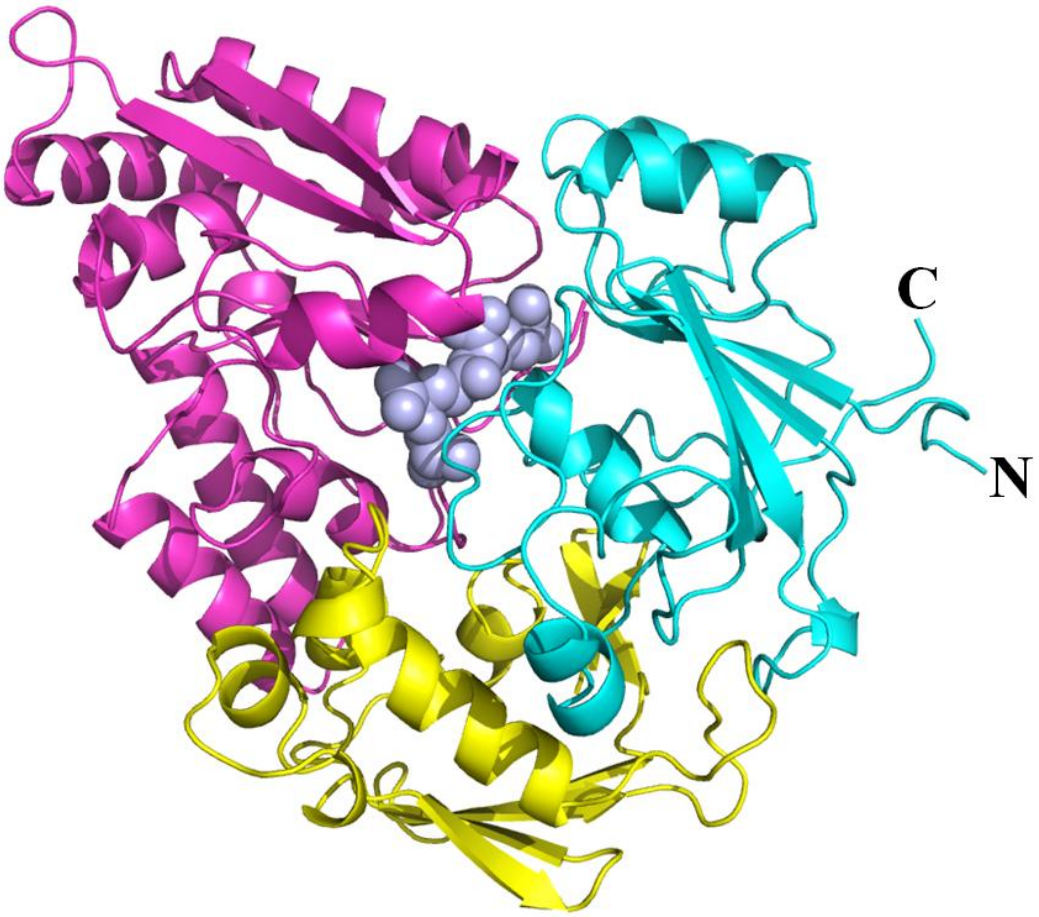


Figure 4.3 Structure of MppA in the closed ligand bound state

Ribbon representation of the MppA structure. The three-domain organisation; domains I, II and III are coloured in cyan, magenta and gold, respectively. The amino [N] and carboxyl [C] termini are indicated. The Mtp ligand is displayed in space-filling format.

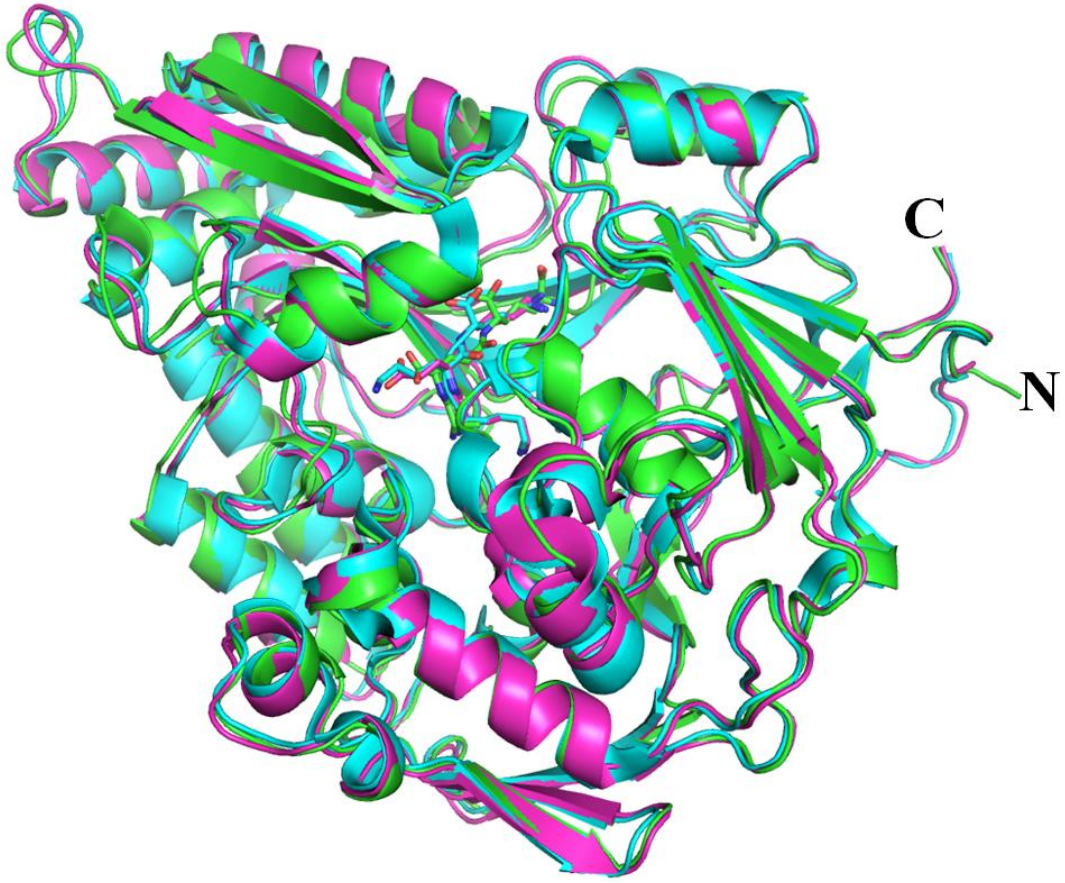


Figure 4.4 Structural superposition of MppA with OppAs

The proteins were superposed in PyMol and are shown in cartoon representation with MppA in green, OppA *Salmonella* Typhimurium (OppA_*St*) in cyan and OppA *Yersinia pestis* (OppA_*Yp*) in magenta. The ligands, KEK for OppA_*St*, KKK for OppA_*Yp* or Mtp for MppA are shown in stick representation.

Four other high scoring matches are DppA from *E. coli* (PDB code 1DPP, Z-score = 41.2, with 471 equivalent Cα atoms superimposing with an rmsΔ 2.1 Å), AppA from *B. subtilis* (PDB code 1XOC, Z-score = 39.4, with 471 equivalent Cα atoms superimposing with an rmsΔ 2.3 Å), GsiB from *E. coli* (PDB code 1UQW, Z-score = 35.2, with 472 equivalent Cα atoms superimposing with an rmsΔ 4.8 Å) and NikA from *E. coli* (PDB code 3DP8, Z-score = 34.6, with 469 equivalent Cα atoms superimposing with an rmsΔ 2.8 Å). A sequence alignment of MppA with OppA, DppA, AppA, GsiB and NikA is shown in Figure 4.5. The peptide and nickel-binding proteins are larger than other ABC transporter receptors because the latter lack domain II (Figure 4.3).

4.4 The structure of MppA contains bound Mtp

Examination of the interdomain region in MppA revealed electron density that could not be accounted for by protein atoms. Usually the electron density for bound peptides in peptide-binding proteins is ambiguous because a heterogeneous mixture of the peptides is present (Tame *et al.*, 1994, Berntsson *et al.*, 2009, Levdikov *et al.*, 2005), but here we were able to unambiguously model and refine Mtp into the electron density (Figure 4.6), This showed that MppA co-crystallized with its cognate ligand (Mtp) consistent with the idea that MppA binds Mtp selectively.

4.5 The Mtp binding site in MppA

The crystal structure reveals MppA in a ‘closed’ conformation with a single molecule of Mtp completely buried within the protein, as seen with other peptides in oligopeptide binding protein structures (Levdikov *et al.*, 2005, Berntsson *et al.*, 2009, Tame *et al.*, 1994, Dunten and Mowbray, 1995). The Mtp sits in a clear binding pocket, forming multiple interactions with the protein (Figure 4.7). The *N*-terminal amino group of the peptide forms a salt bridge with the side chain carboxylate of Asp-417 and a hydrogen bond to the main chain carbonyl of Val-415 in MppA. A prominent interaction is a bivalent salt bridge that is formed between the free α-carboxylate from the γ-linked D-Glu component of Mtp and Arg-402 in MppA. Additional charge-charge and charge-dipole interactions are made to the first carboxylate group in the *meso*-Dap component via Lys-20, Arg-411 and Ser-413 in the protein. At the C-terminus of the tripeptide, the second carboxylate of the Dap is coordinated by two water molecules. Meanwhile the

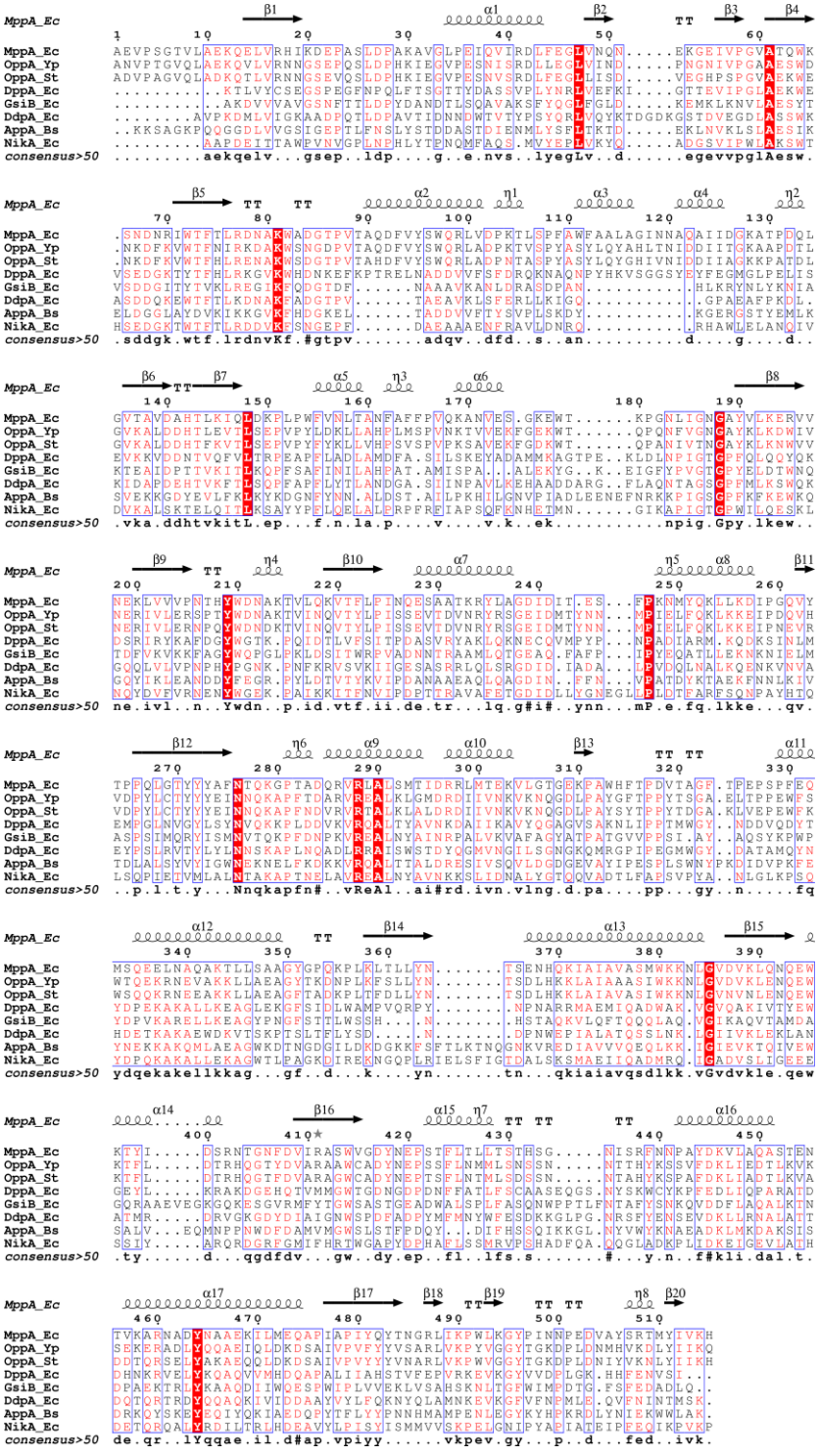


Figure 4.5 Multiple sequence alignment of selected binding proteins

Sequence alignment of MppA of *E. coli* (MppA_Ec), OppA of *Y. pestis* (OppA_Yp), OppA of *S.* Typhimurium (OppA_St), DppA of *E. coli* (DppA_Ec), GsiB of *E.coli* (GsiB_Ec), DdpA of *E. coli* (DdpA_Ec), AppA of *B. subtilis* (AppA-Bs) and NikA of *E.coli* (NikA_Ec) based on an alignment of the 3D structures of the proteins generated using ESPript.

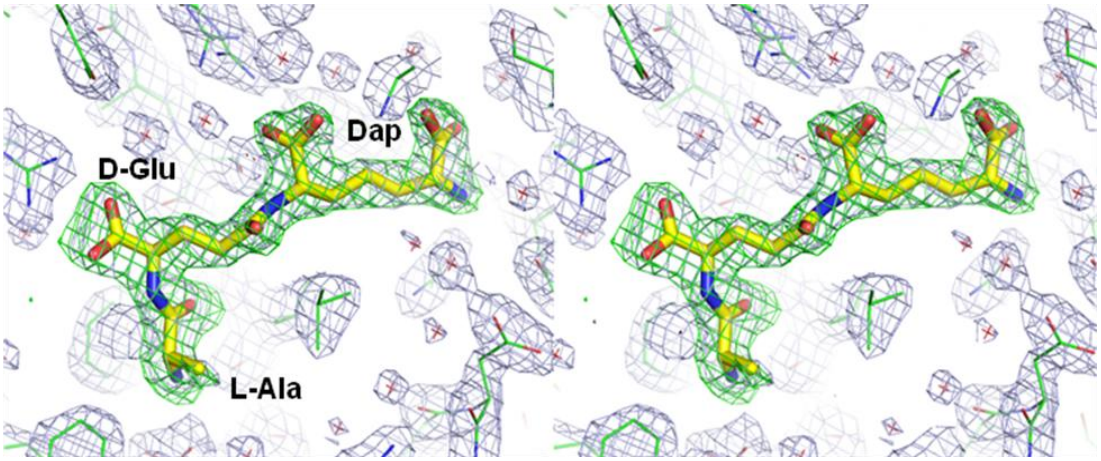


Figure 4.6 The structure of Mtp modelled into its associated electron density

Stereo image of OMIT maps calculated with coefficients $F_{obs} - F_{calc}$ (green) contoured at the 3 σ level and with coefficients $2F_{obs} - F_{calc}$ (light blue) contoured at 1 σ level,and displayed in the vicinity of the Mtp ligand. The Mtp peptide was left out from the refined coordinate set and several cycles of refinement in REFMAC were performed prior to the calculation of an OMIT map. The component amino acids of Mtp are labeled. Temperature factors for Mtp atoms in the final model were refined with full occupancies and are in the range of 19 - 33 $Å^2$, similar to that of surrounding protein atoms.

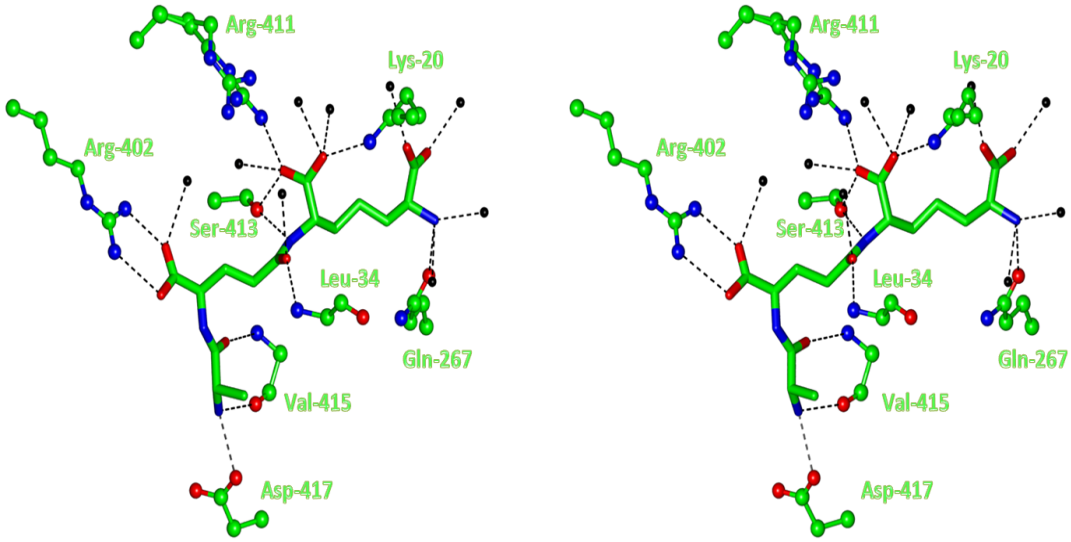


Figure 4.7 Binding pocket of MppA with the ligand Mtp

The amino acids involved in coordinating the ligand Mtp are represented in ball and stick format and are labeled, and the Mtp is represented in cylinder format. Hydrogen bonding interactions are indicated with dashed lines. Water molecules interacting with ligand are represented as black spheres.

C-terminal amino group interacts with two water molecules and the side chain carbonyl of Gln-267 (Figure 4.7).

4.6 Comparison of the binding sites in MppA to OppA

To compare the binding site of a general oligopeptide binding protein with that of MppA, we superposed the backbones of OppA and MppA and analysed the binding of the two ligands, KEK for OppA and Mtp for MppA (Figure 4.8 and Figure 4.5 for a structure based sequence alignment). The proteins are 85% similar overall at the sequence level and hence certain parts of the binding site are very well conserved between these proteins, the first being the Asp-417 residue in OppA (Asp-419) which forms a salt bridge to the α-ammonium group of the peptide. This 'anchors' the *N*-terminus of the peptide and in OppA it is known that a protonated *N*-terminal amino group on the ligand is essential for transport, suggesting that this is also likely to be a key interaction in MppA (Payne, 1971). It is apparent that the overall shape of Mtp in the binding pocket of MppA is surprisingly similar to that of KEK in OppA, such that the γ-linked D-Glu adopts a similar conformation to the more typical peptide linkage seen in the KEK tripeptide (Figure 4.8B).

The Arg-402 residue that binds to the carboxylate group of the γ-linked D-Glu is also conserved in OppA (Arg-404). However in OppA, it is positioned differently as a result of its interaction with Glu-276 (Figure 4.8) so that it does not form a direct contact to the peptide. In OppA, Arg-404 coordinates H_2O molecules that in turn form polar contacts to the ligand (Tame *et al.*, 1994). Water molecules, that are abundant in the OppA binding pocket, allow the protein to accommodate variability in its ligand side chains. This is because the water network can be adapted/altered according to the particular ligand present. However, in MppA, where Glu-276 is replaced by alanine (Ala-274), the Arg-402 side chain is more prominent in the binding pocket so that its guanidine group forms what are expected to be strong ion-pairing interactions with the α-carboxylate group of the D-Glu in Mtp (Figure 4.8). The substitution of Ala for Glu at this position is a clear distinguishing feature of the MppA and OppA proteins (Figure 5.7).

A second conserved residue Arg-411 (Arg-413 in OppA) also interacts strongly with the ligand. In OppA, Arg-413 either forms a salt bridge to the C-terminal carboxylate of the

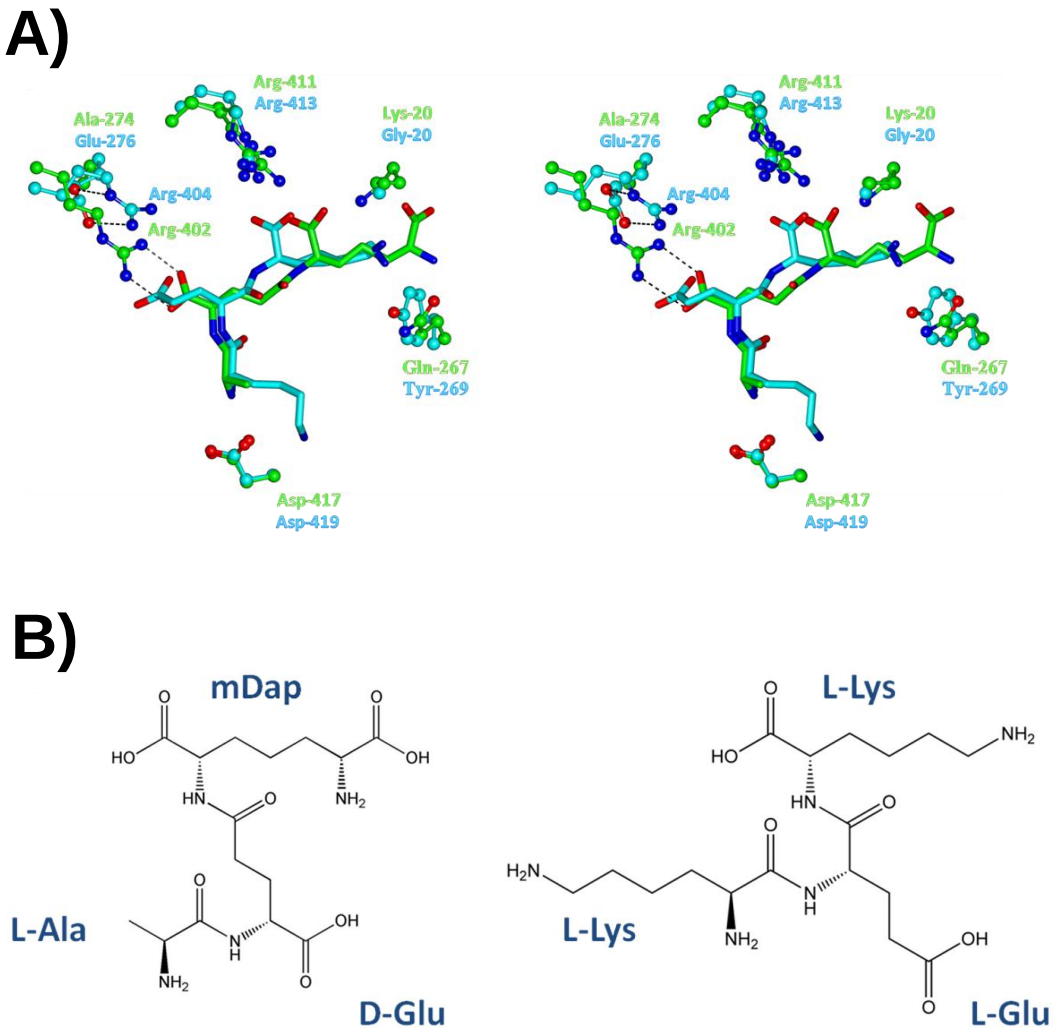


Figure 4.8 Ligand binding in MppA and OppA

A) Structural superposition of the ligand binding sites of MppA and OppA. The amino acid residues in the region of substrate binding sites of MppA (green) and OppA (cyan) are compared and labelled by their positions in the PDB files 3O9P and 1JEU respectively. Mtp and KEK are represented in cylinder format, and amino acid residues of MppA and OppA are represented in ball and stick format. **B)** Chemical structures of Mtp and the KEK tripeptide. The structures were provided by Dr. Christoph Baumann

tripeptide KEK or a hydrogen bond to the main chain carbonyl of residue 3 in the tetrapeptide KKKA (Tame *et al*., 1995). In MppA where there is not a typical amino acid at the third position in the peptide, it is the first carboxylate of the *meso*-Dap that now forms a salt bridge to the corresponding Arg residue. The side chain ε-NH_3^+ group of Lys-20 also forms electrostatic interactions with this carboxylate in MppA. The corresponding residue is a glycine in OppA so this particular contact to the ligand is unique to MppA (Figure 4.8).

4.7 Size Exclusion Chromatography Multi-Angle Laser Light Scattering (SEC-MALLS

The MppA crystal had an unusual P6 space group, often seen in proteins that are hexamers (Tsai *et al*., 2007, Satyshur *et al*., 2007, Yu *et al*., 1998), which prompted us to determine the native molecular weight of MppA using SEC-MALLS (Figure 4.9). The molecular weight measured by light scattering was 60.6 ± 1.5 kDa which is close to the calculated molecular weight for the monomer of 58 kDa, based on the amino acid sequence and suggests that MppA, like other peptide-binding SBPs, is a monomer in solution.

4.8 Conclusion

The structure of MppA was solved in the presence of its cognate ligand. Analysis of the binding cavity of MppA has revealed that Mtp is bound in a relatively extended conformation with its three carboxylates projecting from one side of the molecule and its two amino groups projecting from the opposite face. Comparison of the structure of MppA-Mtp with structures of conventional tripeptides bound to OppA, reveals that the peptide ligands superimpose remarkably closely given the profound differences in their structures. Strikingly, the effect of the D-stereochemistry, which projects the side chain of the D-Glu residue at position 2 in the direction of the main chain in a conventional tripeptide, is compensated by the formation of a γ-linkage to the amino group of diaminopimelic acid, mimicking the peptide bond between residues 2 and 3 of a conventional tripeptide.

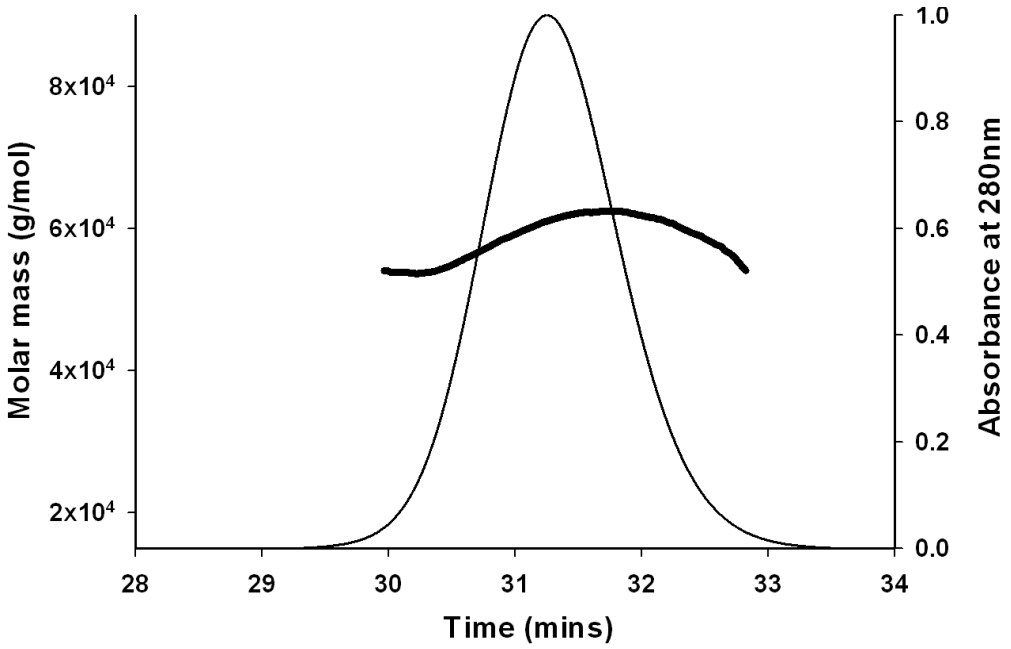


Figure 4.9 Molecular mass of MppA determined by SEC-MALLS

A SEC-MALLS trace of MppA showing the relative absorbance at A_{280} trace normalized to 1 (thin line) and the molar mass of the species present in the sample (thick line).

Chapter 5

5 Investigation of the ligand binding properties of MppA

So far we have provided a biophysical evidence to show that MppA binds a unique peptide, Mtp and presented a crystal structure of MppA in complex with Mtp. As a next step we decided to determine the binding properties of MppA with related ligands. This necessitates the purification of ligand free protein. Co-purified ligand (Mtp) was removed from MppA using guanidinium HCl (GdmHCl) as described previously (Lanfermeijer *et al*., 1999) to enable the ligand-free protein (denoted MppA*) to be used for binding studies with peptide ligands. The protein was partially unfolded while bound to the nickel-nitrilotriacetic acid resin, by including additional wash steps with buffer containing GdmHCl as described in methods section 2.7.1. The protein (MppA*) was eluted with elution buffer as described previously in section 3.3 and the absence of ligand being confirmed using native electrospray mass spectrometry (data not shown). All the binding experiments used ligand free MppA. The ligand, Mtp was kindly prepared and supplied by Dr. Mireille Herve (University of Paris-Sud, France).

5.1 MppA binds Mtp with high affinity

We investigated Mtp binding to the protein using tryptophan fluorescence spectroscopy, a method that exploits the significant conformational change that accompanies ligand binding. Binding of Mtp to MppA causes ~ 7% enhancement in fluorescence upon binding. A signal increase can be seen on addition of Mtp to MppA in Figure 5.1A. The affinity of MppA was determined by florescence titration where MppA was titrated with increasing concentrations of Mtp and the corresponding fluorescence change was monitored in time-acquisition mode. The cumulative fluorescence change was plotted in SigmaPlot and the K_D was calculated from the hyperbolic fit of the binding curve (Figure 5.1B) using equation 2.1. Titration experiments yielded a K_D of 0.24 ± 0.02 μM.

We also assessed the binding of Mtp to MppA using isothermal titration calorimetry (ITC). 50 μM MppA was titrated with 10 μl additions of 500 μM Mtp in 50 mM KPi pH 7.8 at 37 °C (Figure 5.2). This gave a K_D of 0.30 ± 0.06 μM, consistent with the value obtained from tryptophan fluorescence spectroscopy (Figure 5.1A). The ITC data also confirm the 1 : 1 Mtp : MppA binding stoichiometry (0.99 ± 0.01) and demonstrate a positive enthalpy change upon Mtp binding ($\Delta H°$ of 2092 ± 27 kJ mol^{-1} and $\Delta S°$ of 36.9 $J.K^{-1}mol^{-1}$) (Figure 5.2), suggesting that binding is entropy driven as observed previously for OppA (Sleigh *et al*., 1999).

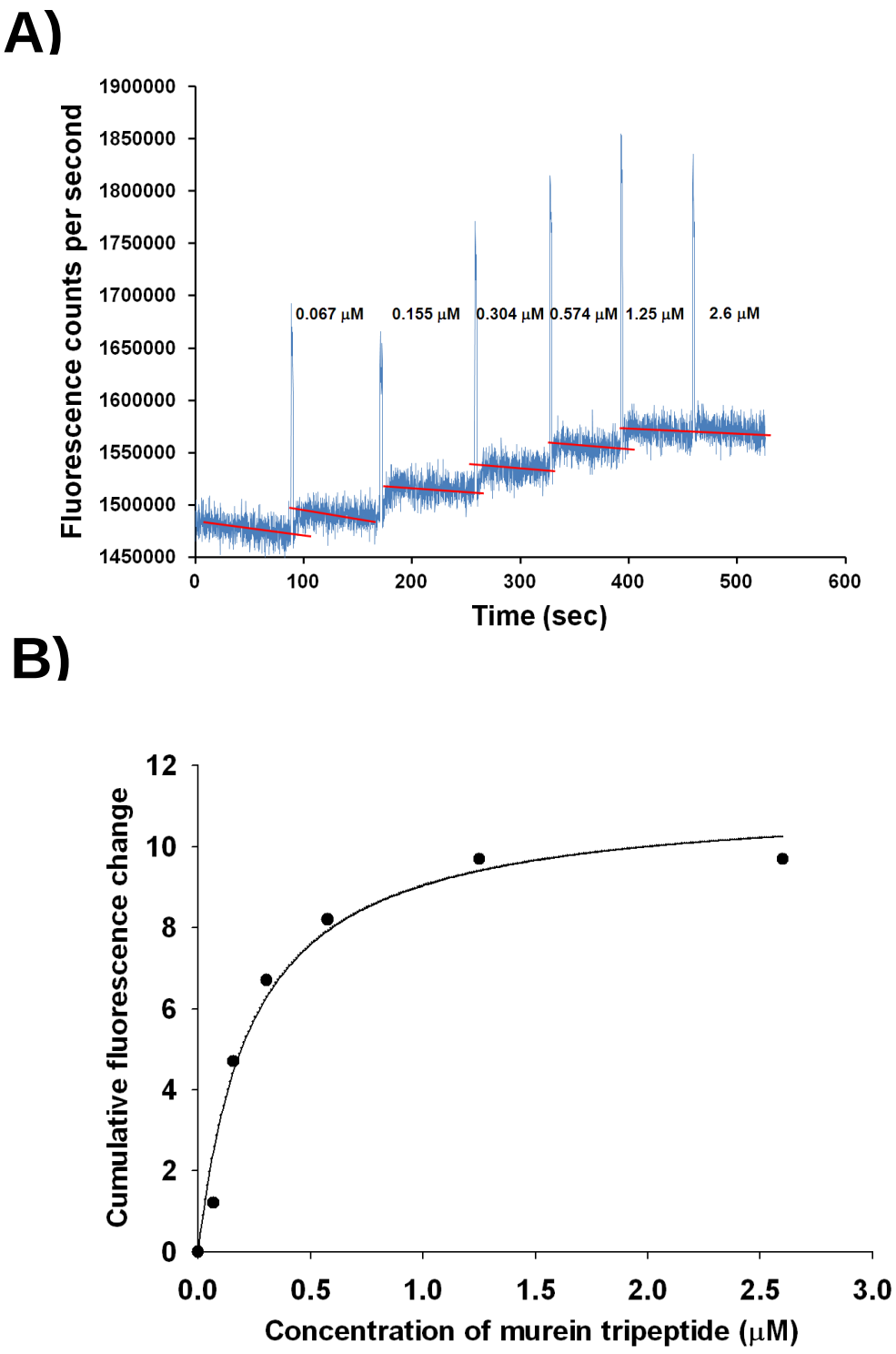


Figure 5.1 Mtp binding to MppA

A) Fluorescence change upon addition of Mtp to MppA, cumulative concentration of Mtp shown. **B)** Cumulative changes in tryptophan fluorescence of MppA upon titration with Mtp.

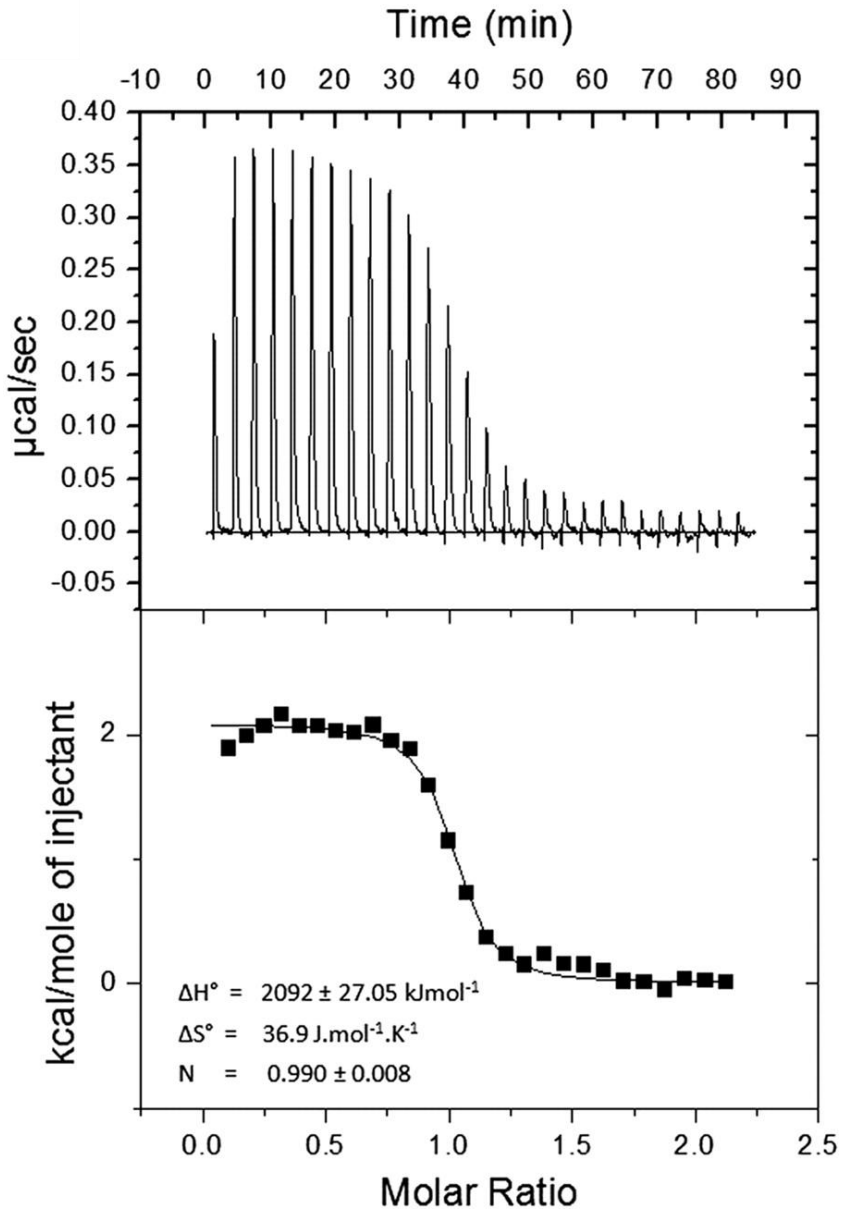


Figure 5.2 Binding isotherms for the interaction of MppA with Mtp

The top panels show heat differences upon injection of ligands and lower panels show integrated heats of injection (■) and the best fit (solid line) to a single site binding model using Microcal Origin software.

5.2 Specific recognition of Mtp by MppA

To determine the specificity of the interaction between Mtp and MppA we assessed whether MppA was able to bind a number of closely related ligands (Table 5.1). We first examined the binding of the synthetic peptide, L-Ala-γ-D-Glu-L-Lys (AEK), in which the Dap in Mtp is replaced by the closely related amino acid lysine. As measured by tryptophan fluorescence spectroscopy (Figure 5.3),AEK binds to MppA with a K_D of 130 ± 10 μM, which is ~500-fold lower affinity than for Mtp. The binding affinity was also confirmed by ITC (Table 5.1). This points to an important role for the side chain carboxylate of Dap in high affinity binding to MppA. We also investigated a peptide L-Ala-D-Glu-*meso*-Dap where the D-stereochemistry of residue 2 is not accompanied by a γ peptide linkage. We were unable to detect binding of this ligand to MppA by tryptophan fluorescence spectroscopy (Table 5.1).

It has been demonstrated previously, using genetic methods, that a peptide transporter using MppA as the SBP is able to support growth of an *E. coli* proline auxotroph on the tripeptide L-Pro-L-Phe-L-Lys (PFK) (Park *et al*., 1998).We examined direct binding of this tripeptide to MppA, but were unable to detect binding even using concentrations up to 1 mM using tryptophan fluorescence (data not shown). It is possible that PFK binding to MppA does not result in a net change in the tryptophan fluorescence of the protein, and so we measured PFK binding to MppA using isothermal titration calorimetry (Figure 5.4). We were able to detect an interaction between MppA and PFK, although the binding was weak with a K_D greater than 300 μM and the binding isotherm was markedly different from that observed for Mtp binding (ΔH of -3100 ± 43 kJ mol^{-1} and ΔS of 5.72 $J.K^{-1}mol^{-1}$) (Figure 5.4), suggesting that the binding mechanism is different, perhaps indicative of non-specific binding to the protein. It is certainly not consistent with a previous report that suggested a K_D of MppA for PFK less than 50 μM (Letoffe *et al*., 2006). Together these data suggest that high-affinity binding in the submicromolar range is seen only for Mtp itself and even small modifications of the ligand structure result in significant reductions in affinity.

We next examined a range of conventional tripeptides that have been used in studies on OppA including KEK, KKK, AAA, KKKA and AKK. For none of these peptides were we able to detect binding to MppA (Table 5.1). It has also been reported that MppA can bind heme and its precursor δ-aminolevulinic acid and its binding is competitive with

Table 5.1 Binding affinities of MppA with different ligands

Ligand	Binding affinity (K_D)	
	Tryptophan fluorescence	**ITC**
L-Ala-γ-D-Glu-mDap (Mtp)	0.24 ± 0.02 μM	0.30 ± 0.06 μM
L-Ala-γ-D-Glu-Lys (AEK)	132 ± 10 μM	240 ± 40 μM
L-Ala-D-Glu-mDap (AEDap)	Binding not detected	Not checked
Pro-Phe-Lys (PFK)	Binding not detected	> 300 μM
δ aminolevulinic acid	Binding not detected	Not checked
KEK	Binding not detected	Not checked
KKK	Binding not detected	Not checked
AKK	Binding not detected	Not checked
AAA	Binding not detected	Not checked
KKKA	Binding not detected	Not checked

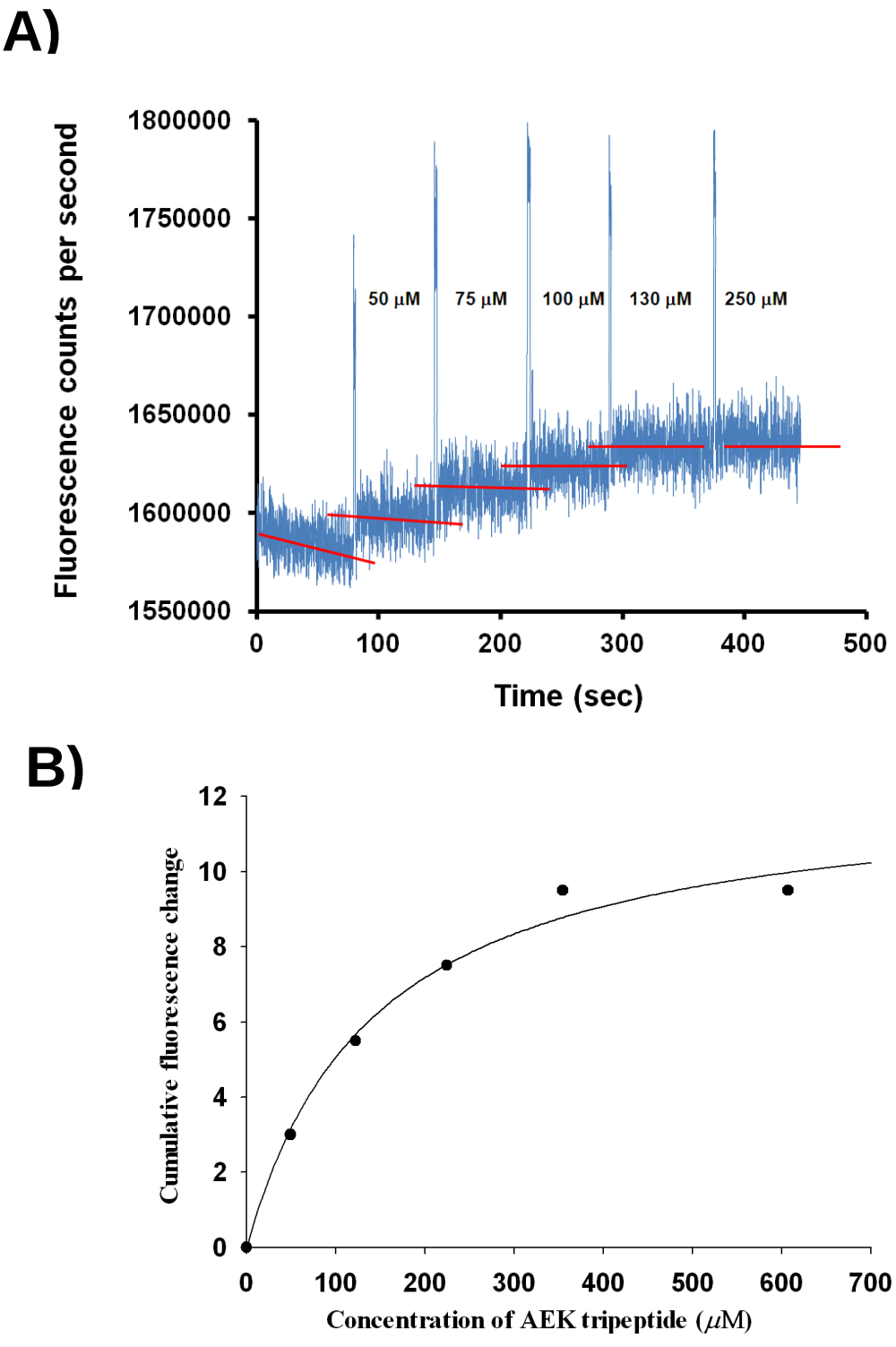


Figure 5.3 AEK binding to MppA

A) Fluorescence change upon addition of AEK to MppA, cumulative concentration of AEK shown. **B)** Cumulative changes in tryptophan fluorescence of MppA upon titration with AEK.

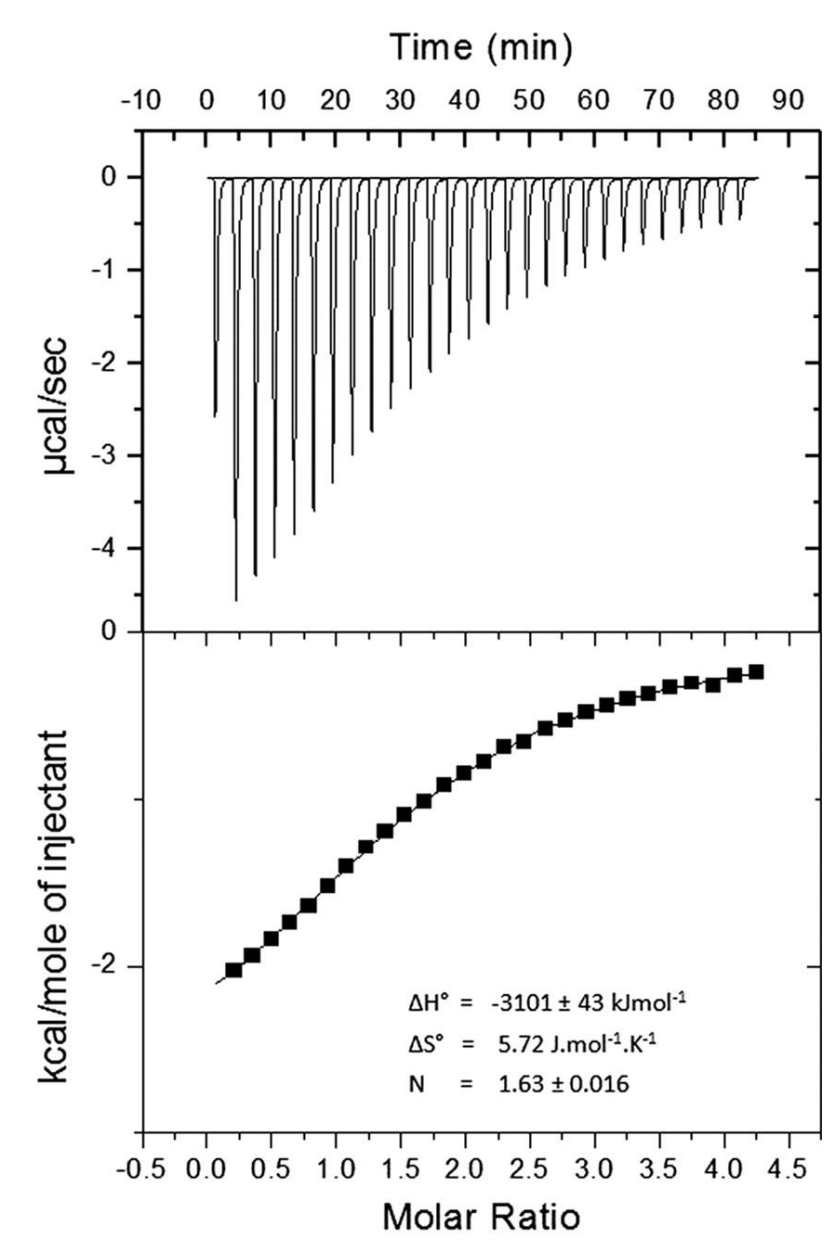


Figure 5.4 Binding isotherms for the interaction of MppA with PFK

The top panels show heat differences upon injection of ligands and lower panels show integrated heats of injection (■) and the best fit (solid line) to a single site binding model using Microcal Origin

PFK, which we have already demonstrated binds much more weakly to MppA than Mtp. Indeed we were unable to observe binding of δ-aminolevulinic acid to MppA via tryptophan fluorescence spectroscopy (Table 5.1).

5.3 Binding of Mtp to OppA

After establishing the specificity of MppA to Mtp, we assessed binding of the Mtp to OppA to investigate whether this unusual peptide ligand can be accommodated by a general peptide-binding protein. We purified *Salmonella* Typhimurium OppA from an *E. coli* strain CH212 (Hiles and Higgins, 1986) carrying a plasmid pCH19. pCH19 is a multicopy plasmid which expresses the *S.* Typhimurium OppA protein from its native promoter. A glycerol stock of the strain CH212 was previously received as a gift from Chris Higgin's lab and was maintained in YSBL. Culture of *E. coli* CH212 was grown in LB medium containing ampicillin 30 μg/m at 37°C. Cells were harvested by centrifugation and the pellet was resuspended in ice-cold SET buffer (methods). Chicken egg white lysozyme (Sigma) was added to cell resuspension at 600 μg/ml and incubated at 30 ºC for 2 hours. The subsequent sphaeroplasts were harvested by centrifugation and the supernatant was stored. The supernatant constituted the periplasmic fraction and contained the over produced OppA. The periplasmic fraction was dialysed into 50 mM Tris buffer (pH 8) and then loaded onto a pre-equilibrated 5 ml ion exchange column (MonoQ from GE Healthcare) on an AKTA Purifier P-900. The OppA protein remained bound to the column during washing with 50 mM Tris buffer (pH 8). Proteins were eluted using a continuous 20 CV gradient of 0 - 500 mM NaCl in buffer (50mM Tris pH 8, 0.5M NaCl) (Figure 5.5A). OppA protein eluted at about 150 mM NaCl. The fractions containing the OppA protein were analysed by separating the protein using SDS-PAGE gel (Figure 5.5B).

As OppA is high affinity peptide-binding protein and co-purifies with bound peptides. Peptides were removed by extensive washing of protein with sodium acetate (pH 5) on a Mono S (GE Healthcare) cation exchange column. Ligand free protein (denoted OppA*) was eluted using a continuous 20CV gradient of 0 - 500 mM NaCl. Protein was 99% pure at this stage as judged on SDS gel (Figure 5.5C) and was checked for the absence of peptides by native mass spectrometry.

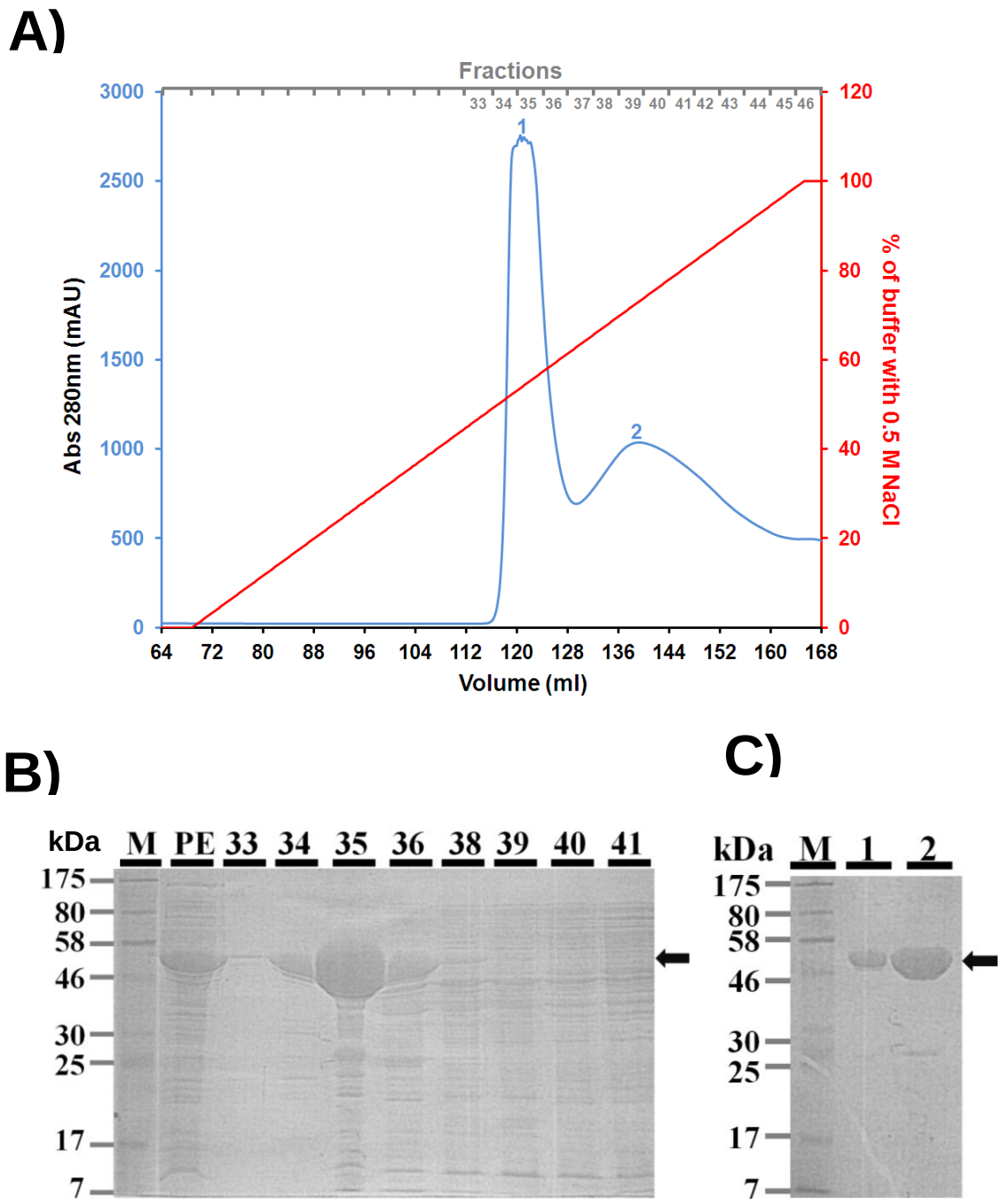


Figure 5.5 Purification of *Salmonella* Typhimurium OppA

A) Chromatogram of the purification of OppA from anion exchange column (Abs 280 nm in blue and % of buffer containing 0.5 M NaCl). **B)** Coomassie stained SDS-PAGE of the periplasmic extract (PE) loaded to the column and elution fractions (33-36) for peak 1 and (38-41) for peak 2 in the chromatogram. Fractions 33-36 contain OppA where as fractions 38-41 contain contaminant proteins. **C)** Coomassie stained SDS gel containing 5 μg (Lane 1) and 20 μg Lane2) of ligand free purified OppA. M is prestained protein ladder.

We used isothermal titration calorimetry to assess whether OppA can bind Mtp and similar peptides. The protein was able to bind a conventional tripeptide, KEK with submicromolar affinity but we could not detect binding of Mtp (Figure 5.6).

5.4 Site directed mutagenesis of MppA

5.4.1 Creation of A274E and R402A mutants

As explained in section 4.5, that Arg-402 is a potentially important residue in MppA that binds to α carboxylate group of the γ-linked D-Glu of Mtp. However, In OppA, this residue (Arg-404) is interacting with water and positioned away from the ligand because of its interaction with Glu-276 (Figure 4.8A). In MppA, Glu-276 is replaced by Ala-274 (Figure 5.7B) and thus allows Arg-402 to make a direct interaction with Mtp. In order to investigate the role of the Arg-402 residue in MppA, we mutated this residue to Ala and investigated the effects on ligand binding. We also mutated Ala-274 to Glu to assess whether an OppA-like residue at this position could sequester the Arg-402 side chain and prevent it from making a direct contact with Mtp.

To construct A274E and R402A mutants of MppA, site-directed mutagenesis was carried out using the one-step site-directed and site-saturation mutagenesis protocol of Zheng *et al*., 2004. Briefly, this uses 30–50 base primers with a 5′ overlap of about half their length to reduce the primer-primer annealing temperature and a silent mutation to add a restriction endonuclease cleavage site to allow for screening of potential mutants. The long overhang allows the primers to anneal across the opposite end of the plasmid PCR product, so introducing PCR-like properties to the reaction. This improves the yield by allowing amplification of the entire plasmid and the high melting temperature of the overhang region removes the need for a plasmid ligation step before transformation. The population of transformants bearing the template plasmid was reduced still further by digestion of the PCR product with DpnI restriction endonuclease. The expression vector pAM6091 was used as a template with the following primers: *mppAA274E*F and *mppAA274E*R for making A274E mutant and *mppAR402A*F and *mppAR402A*R for R402A mutant. The sequences of mutants were confirmed by DNA sequencing of the plasmids (pAM6091A274E and pAM6091R402A). The plasmids carrying the desired mutants were transformed into BL21 (DE3) strain of *E. coli* for over production of the mutants.

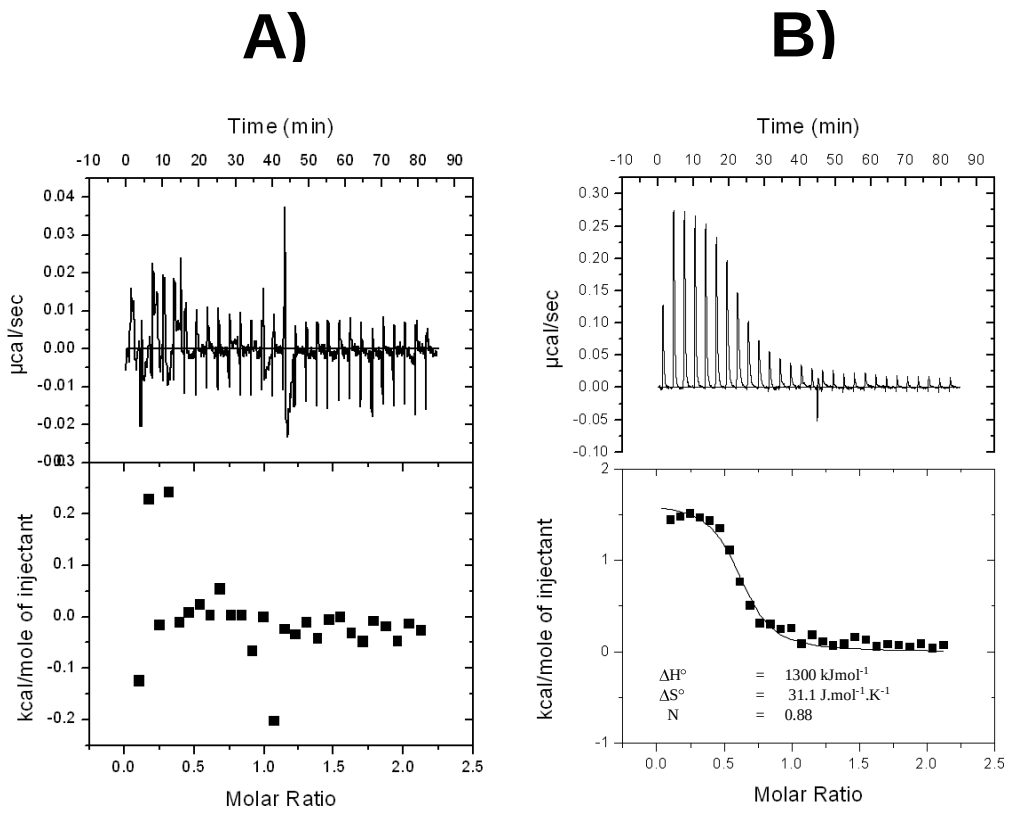


Figure 5.6 Binding isotherms for the interaction of OppA with Mtp (A) and KEK (B)

The top panels show heat differences upon injection of ligands and lower panels show integrated heats of injection (■) and the best fit (solid line) to a single site binding model using Microcal Origin software.

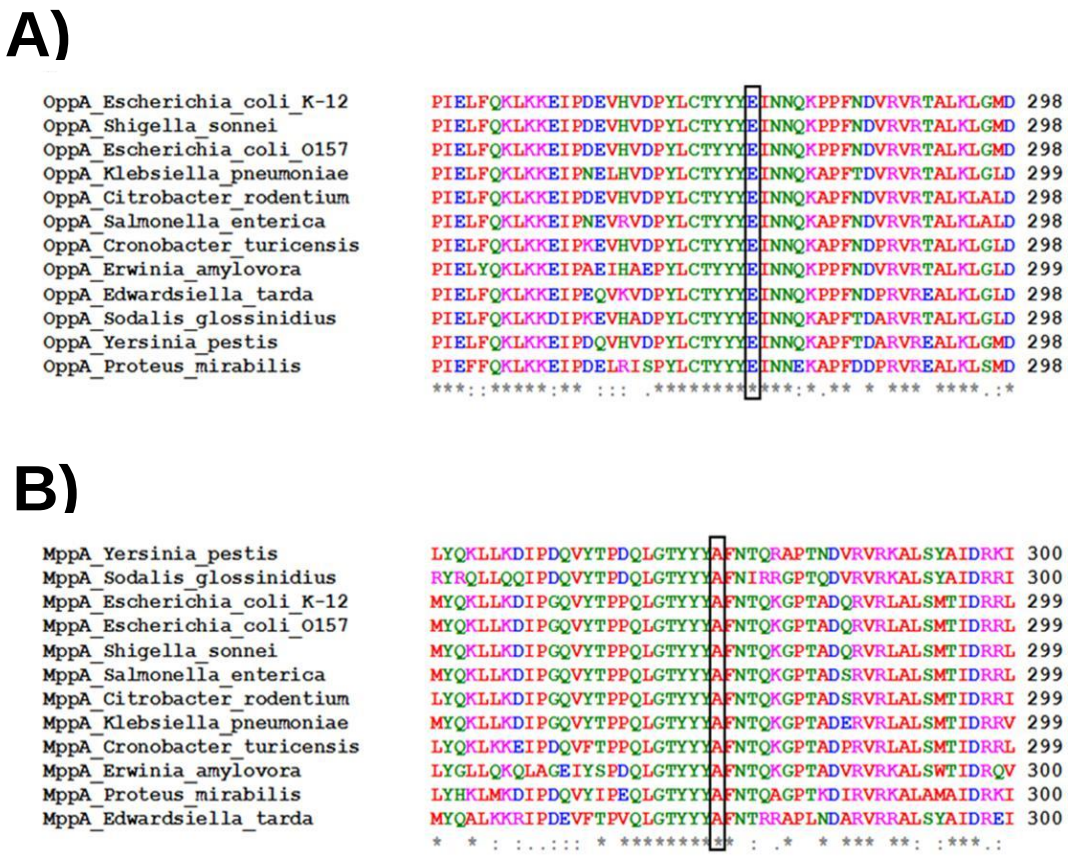


Figure 5.7 Conservation of Glu 276 and Ala 274 in MppA

Sequence alignments of MppA and OppA in different γ-proteobacteria showing that Glu 276 is conserved in all OppAs **(A)** and is replaced by Ala 274 in all MppAs **(B)**.

5.4.2 Expression and purification of A274E and R402A

For each mutant, protein expression was tested in small scale expression trials. Both mutants accumulated to a similar level as the native protein (data not shown). For large scale production cultures of *E. coli* BL21 (DE3) strains were grown in LB medium at 37 °C using 2L flasks. Expression was induced at OD_{650} 0.4–0.6 with 0.5 mM IPTG and the cells were left growing overnight. The cells were harvested by centrifugation and resuspended in 50 mM KPi, pH 7.8. Cells were lysed by sonication and cell debris was removed by centrifugation at 38000 x g for 30 min at 4°C. The whole cell lysate containing over produced protein was collected. Nickel affinity chromatography was used for the purification of protein using gravity-flow columns. The whole cell lysate containing the over produced protein was incubated with equilibrated Ni^{2+}-NTA resin for 1 hour and weakly binding contaminants were eluted by washing the resin with a wash buffer containing 50 mM imidazole. Bound His-tagged protein was eluted with elution buffer containing 500 mM imidazole. Analysis of the fractions taken during purification of protein using Coomassie-stained SDS-PAGE revealed an intense band in elution fraction 2 that corresponded to purified protein (Figure 5.8). An intense band was present in the whole cell lysate fraction, which was less intense in the flow through from the column, indicating that this band was over produced protein and was abundant enough to be visible in the whole cell lysate (Figure 5.8). The entire elution fractions containing purified protein were pooled together and dialysed into 50 mM KPi buffer pH 7.8 and stored.

5.4.3 Analysis of A274E and R402A for ligand binding

As native protein co-purifies with endogenously bound ligand, the mutant proteins were checked by ES-MS for the presence of a ligand. The mass spectrum of both the mutant proteins recorded under native conditions showed no peak for a ligand bound protein (Figure 5.9) suggesting that purified mutant proteins did not purify as complexes with Mtp. Purified mutant proteins were also examined using tryptophan fluorescence spectroscopy for their ability to bind Mtp. Neither of the mutant proteins showed any binding to Mtp suggesting that both residues are important for MppA recognition of Mtp. As we could not detect binding of Mtp to either mutant, we recorded the circular dichroism spectra of the mutant proteins. Both mutants give very similar spectra to the

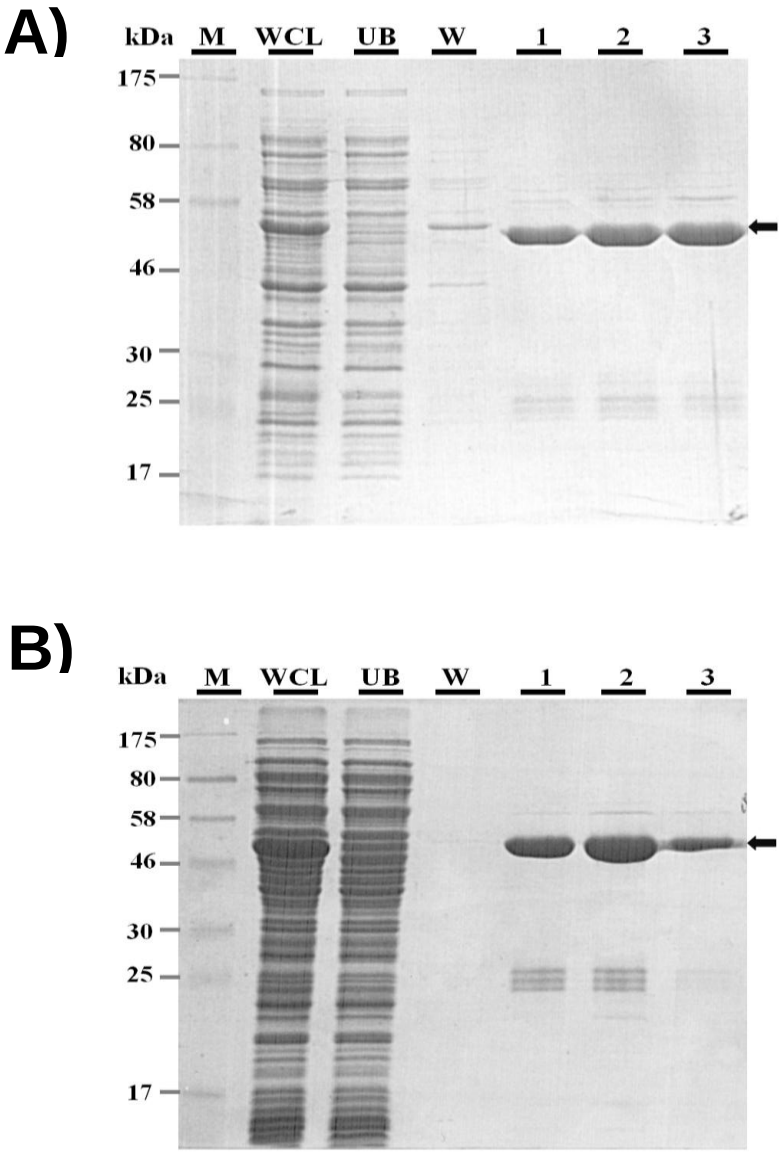


Figure 5.8 Purification of A274E and R402A mutants of MppA

Coomassie stained SDS-PAGE gels showing the purification of MppA: A274E **(A)** and MppAR402A **(B)** by Ni-affinity chromatography showing clarified whole cell lysate (WCL), the flowthrough (UB), the 50 mM imidazole washing step (W) and the elution fractions containing purified proteins indicated by arrows. M is pre-stained protein ladder.

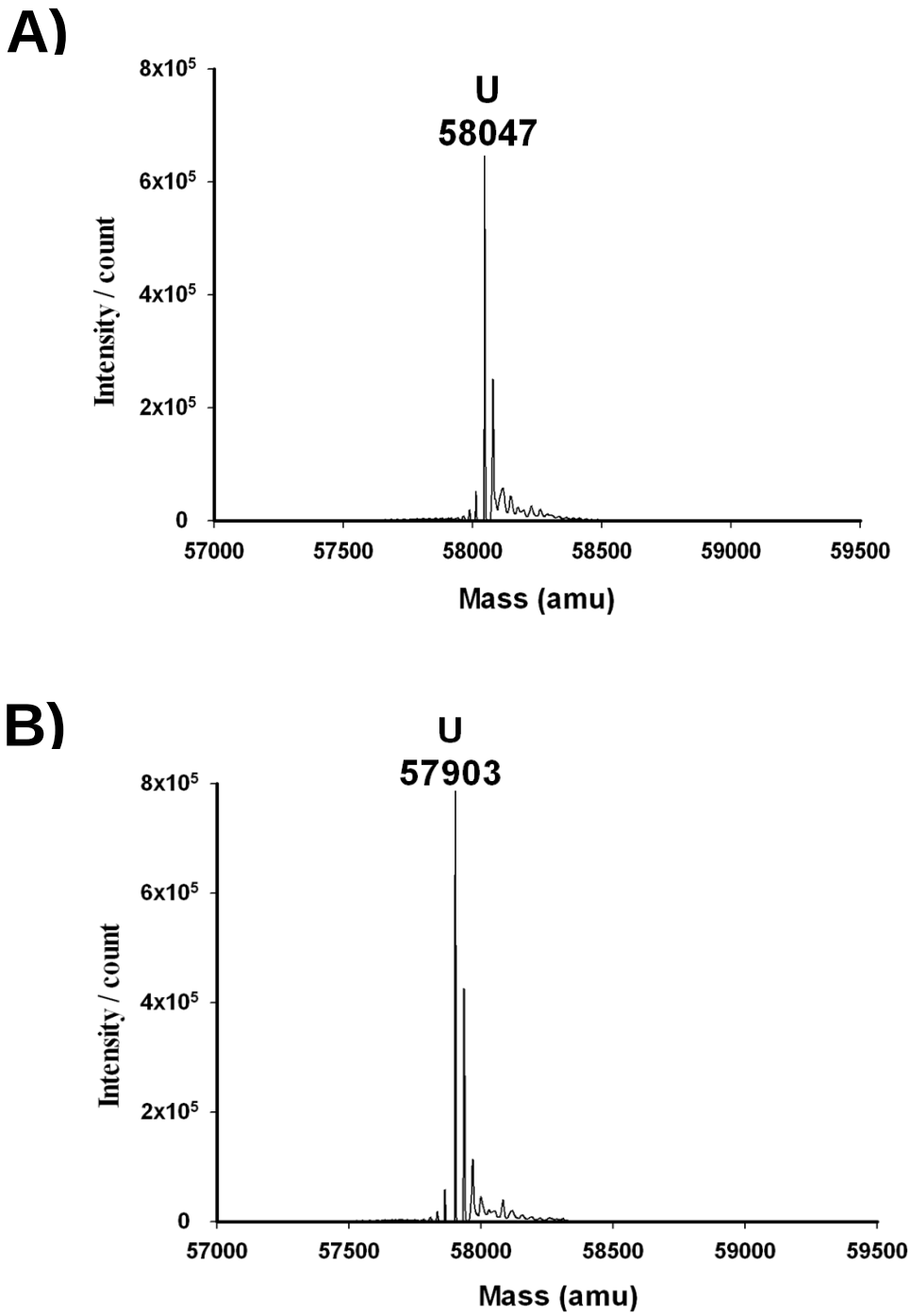


Figure 5.9 ES-MS of MppA mutants

(A) ES-MS spectrum of A274E under native conditions. The molecular mass of the unbound protein (U) is indicated at 58047 atomic mass units (amu). **(B)** ES-MS spectrum of R402A under native conditions. The molecular mass of the unbound protein (U) is indicated at 57903 atomic mass units (amu). Other peaks are adducts of phosphates.

wild-type MppA (data not shown) indicating that failure to bind Mtp is probably not caused by protein misfolding or instability. The binding properties of the mutants were further investigated using tryptophan fluorescence spectroscopy for other peptide ligands. A274E mutant showed a slight quench in signal upon addition of PFK upto 10 mM (Figure 5.10), however it was not possible to titrate this small change in fluorescence. This intrigued us to check binding of PFK by ITC and to check if binding affinity was improved for this ligand. We therefore used ITC to check whether the change in fluorescence represents true binding of A274E to PFK. Although we could detect an interaction between A274E and PFK but this was similar to that observed in case of wild type MppA (data not shown). This confirmed that the quench in fluorescence signal was an artefact and there is no or very weak binding (K_D > 300 μM) between A274E and PFK. We could not detect any binding of the mutant proteins with rest of the peptide ligands used in Table 5.1.

5.5 A possible physiological function of MppA

It has been suggested that MppA allows *E. coli* to scavange Mtp that is either released during turnover of its own cell wall or that is available in the environment. This suggestion is supported by the fact that *E. coli* is known to secrete Mtp during growth (Goodell and Schwarz, 1985). *E. coli* turns over one third of its cell wall each generation and efficiently reuses Mtp for the synthesis of new wall in a process termed cell wall recycling (Park, 1995). In this pathway cell wall is degraded to *N*-acetylglucosaminyl-1,6-anhydro-*N*-acetylmuramyl-L-alanyl-γ-D-glutamyl-*meso*-diaminopimelate (GlcNAc-anhMurNAc-tripeptide also called muropeptide) by the action of lytic enzymes which are present in the periplasm. The muropeptide is transported into the cytoplasm via the membrane-bound AmpG permease (Jacobs *et al*., 1994). The tripeptide is released from the muropeptide by the cytoplasmic amidase AmpD (Jacobs *et al*., 1995) and thus reenters the cell wall biosynthetic pathway (Figure 5.11A).We proposed a new route for the direct uptake of Mtp via MppA and OppBCDF (Figure 5.11B). This route may give the bacterium an extra advantage of scavenging all the Mtp present in the periplasm. An amidase enzyme (AmiD) present in the outer membrane liberates free Mtp, present in the peptidoglycan (Uehara and Park, 2007). However we realized that Mtp is only 7.41% of the total peptide content of cell wall and

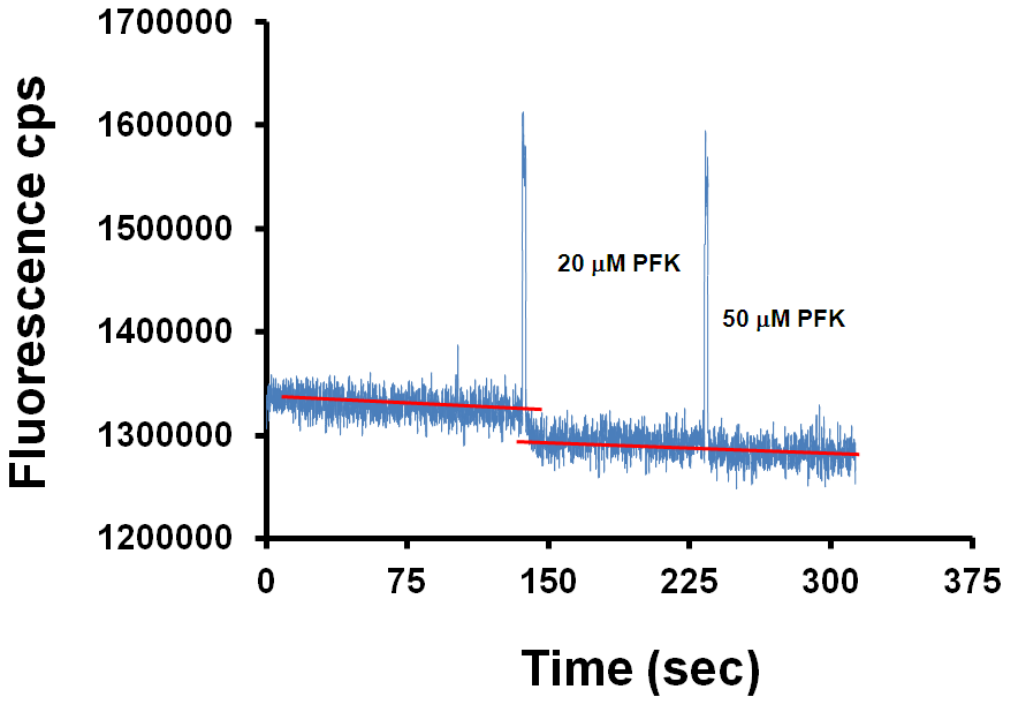
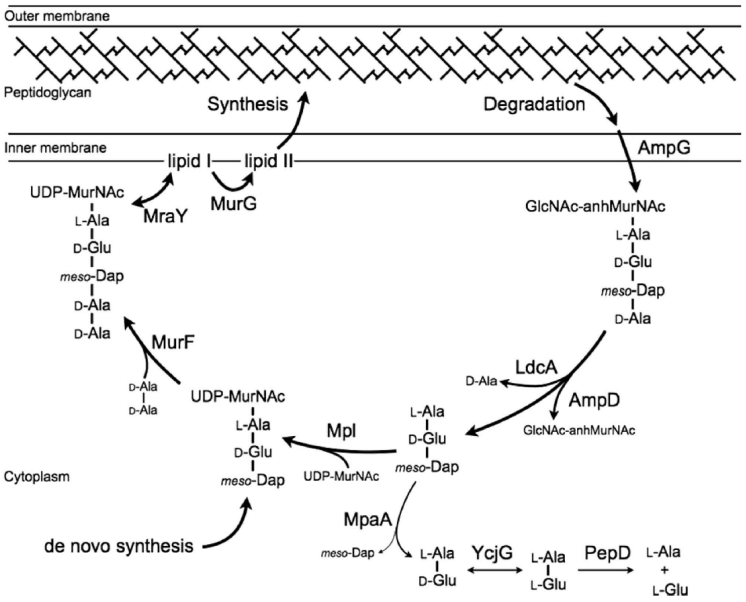


Figure 5.10 PFK binding to A274E

Fluorescence change upon addition of PFK to A274E, cumulative concentration of PFK is shown.

A)



B)

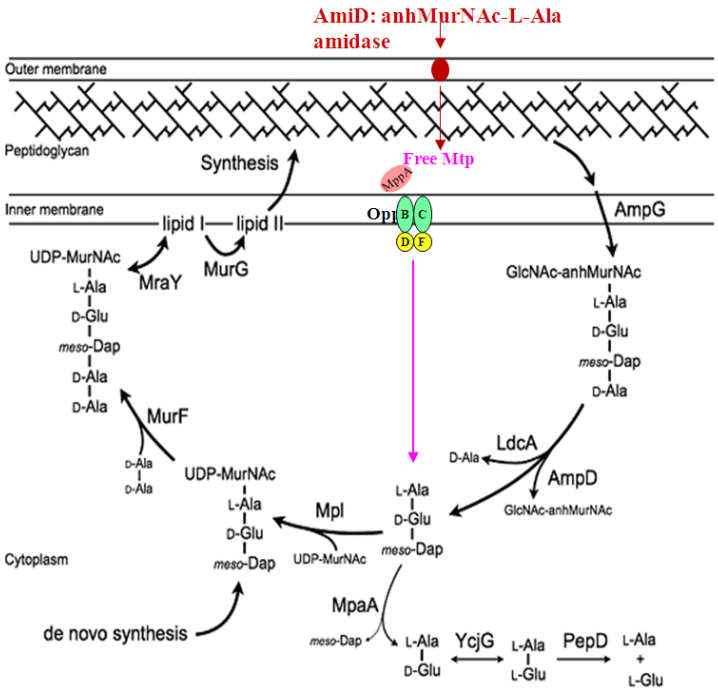


Figure 5.11 Schematic of peptidoglycan biosynthesis and recycling

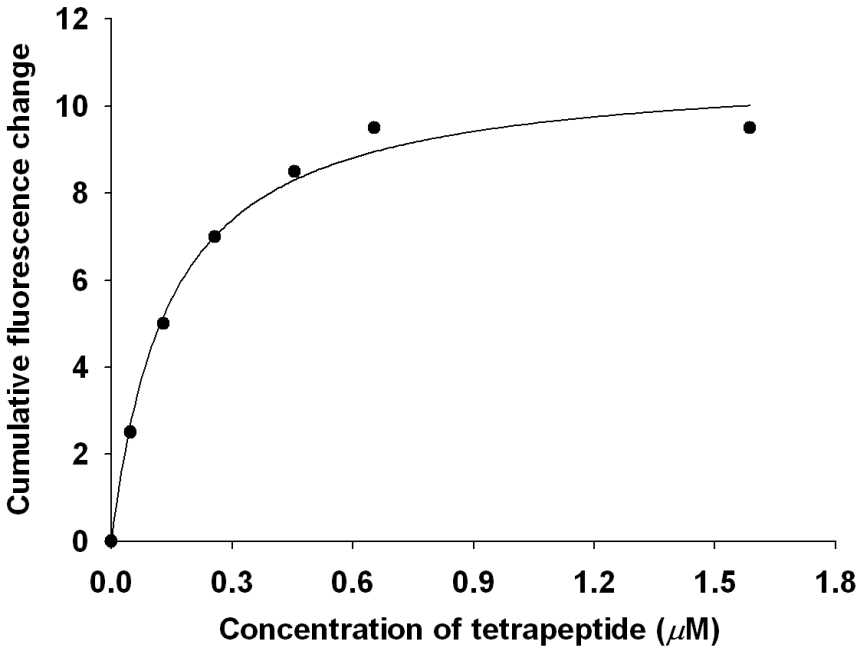
A) Existing pathway for recycling murein tripeptide (Park and Uehara, 2008) **B)** Proposed role of MppA in the uptake of Mtp as a part of cell wall recycling pathway.

the most abundant peptide in cell wall is tetrapeptide (35.9 % of the total peptide content) (Glauner *et al*., 1988). The tetrapeptide is an extension of D-Ala to the C-terminus of Mtp. Considering the abundance of tetrapeptide in cell wall we proposed that bacterium must uptake this as well and reuse it too in the process of recycling. Considering that Mtp and tetrapeptides are not the substrates for AmpG (Cheng and Park, 2002), we speculated that transport of tetrapeptide might occur via the MppA-OppBCDF pathway. We first analysed the structure of MppA to see any room in the binding pocket for accommodating a tetrapeptide. Modelling of a tetrapeptide into the binding cavity of MppA showed no steric clashes with MppA. Binding was confirmed *in vitro* where we showed binding of tetrapeptide to MppA using tryptophan fluorescence spectroscopy. Addition of tetrapeptide to MppA caused approximately 5.5 % quench in the fluorescence. Titration of this change gave a K_D of 0.140 μM ± 0.01311 (Figure 5.12A). To our surprise tetrapeptide binds to MppA with 2 times higher affinity than Mtp. Other breakdown products of cell wall which are present in the periplasm were checked for their binding to MppA. This included a pentapeptide (Mtp-D-Ala-D-Ala) and muropeptide. However we could not see any binding of MppA to these products (data not shown). Thus we propose that uptake of Mtp and tetrapeptide via MppA-OppBCDF is a part of cell wall recycling independent of the major recycling of muropeptide via AmpG (Figure 5.12B).

5.6 Conclusion

This chapter has focussed on the interactions between MppA and peptide ligands, particularly Mtp. MppA binds Mtp with high affinity (submicromolar) and specificity. The investigation of roles of A274 and R402 residues by site directed mutagenesis showed that they make key interactions with Mtp and are essential for MppA binding to Mtp. OppA did not show any binding to Mtp and thus disapproving the previous notion that OppA can accommodate Mtp (Goodell and Higgins, 1987). It was proposed that MppA plays role in cell wall recycling by uptake of Mtp from the periplasm. It was shown for a first time that MppA can also bind to a murein tetrapeptide.

A)



B)

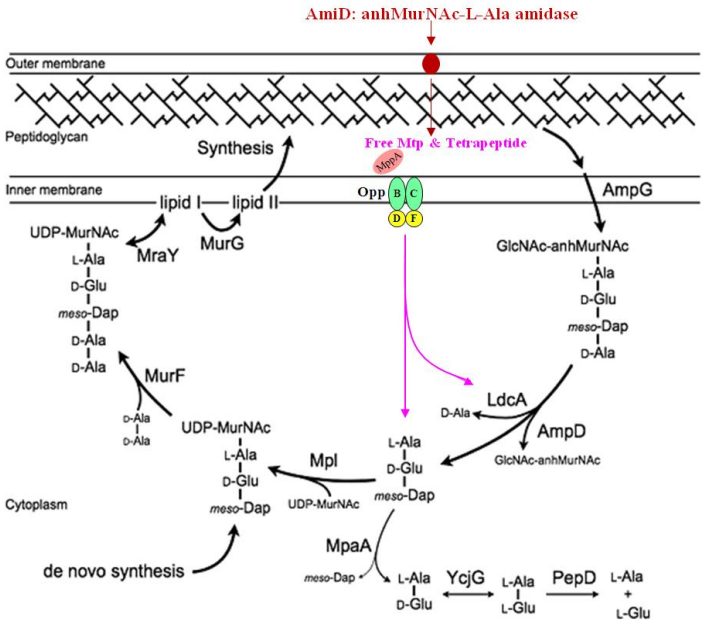


Figure 5.12 Tetrapeptide binding to MppA

A) Cumulative changes in tryptophan fluorescence of MppA upon titration with tetrapeptide**. B)** Proposed function of MppA for uptaking of Mtp and tetrapeptide in cell wall recycling pathway.

Chapter 6

6 Biochemical and structural characterization of the MpaA amidase as part of a conserved scavenging pathway for peptidoglycan derived peptides in γ-proteobacteria

6.1 Genomic context of *mppA* and discovery of a conserved scavenging pathway for peptidoglycan derived peptides

After proposing that MppA is involved in the uptake of murein tripeptide (Mtp) and tetrapeptide as a part of cell wall recycling, the next step was to test the hypothesis by performing growth experiments. The idea was to grow bacteria on cell wall derived peptides as a sole source of carbon and nitrogen and compare the growth of wild type and *mppA* mutant bacteria. This experiment needed large amounts of murein tri and tetrapeptides (μg) and we did not have enough peptides in the stock to do growth experiments.

Meanwhile, an interesting observation was made concerning the genomic context of *mppA* in *E. coli*. *mppA* is situated near a gene called *mpaA*, which was suggested to cleave Mtp (Uehara and Park, 2003). Three uncharacterized genes (*ycjZ*, *ycjY* and *ymjC*) separate *mppA* and *mpaA* in *E. coli*. *ycjZ* is a putative DNA binding transcriptional regulator, *ycjY* is a predicted hydrolase and *ymjC* is a small stretch of DNA that codes for only 46 amino acid residues with no similarity to known proteins. The genetic organization of *mppA* and *mpaA* was checked in other bacteria including *Salmonella*, *Klebsiella*, *Yersinia*, *Erwinia* and *Vibrio*. The comparative genomics analysis carried out with "The Seed" (Overbeek *et al*., 2005) revealed that the two genes are always present together and divergently transcribed (Figure 6.1). The clustering of *mppA* and *mpaA* genes suggests that they are co-regulated and may function together. In *Salmonella*, *Klebsiella* and *Yersinia*, the *mpaA* is adjacent to a gene (*ycjG*), encoding an epimerase enzyme (Figure 6.1) which epimerizes L-Ala-D-Glu to L-Ala-L-Glu (Schmidt *et al*., 2001). A myriad of murein hydrolases are known to degrade peptidoglycan to the product dipeptide L-Ala-D-Glu. No known bacterial peptidase can cleave this dipeptide. However, if L-Ala-D-Glu could be epimerized to L-Ala-D-Glu, hydrolysis could occur with known dipeptidases. The clustering of *mppA*, *mpaA* and epimerase genes led us to propose a peptide scavenging pathway that is conserved in γ-proteobacteria. We propose that MpaA cleaves murein tripeptide and tetrapeptide that are taken up via MppA-OppBCDF thus enabling bacteria to scavange murein tripeptide and tetrapeptide present in the periplasm. The dipeptide L-Ala-D-Glu resulting from the action of MpaA is epimerized to L-Ala-L-Glu, further degraded by a peptidase (PepD) and utilized by bacteria as a nutrient (Figure 6.2). Thus MppA may be involved in a

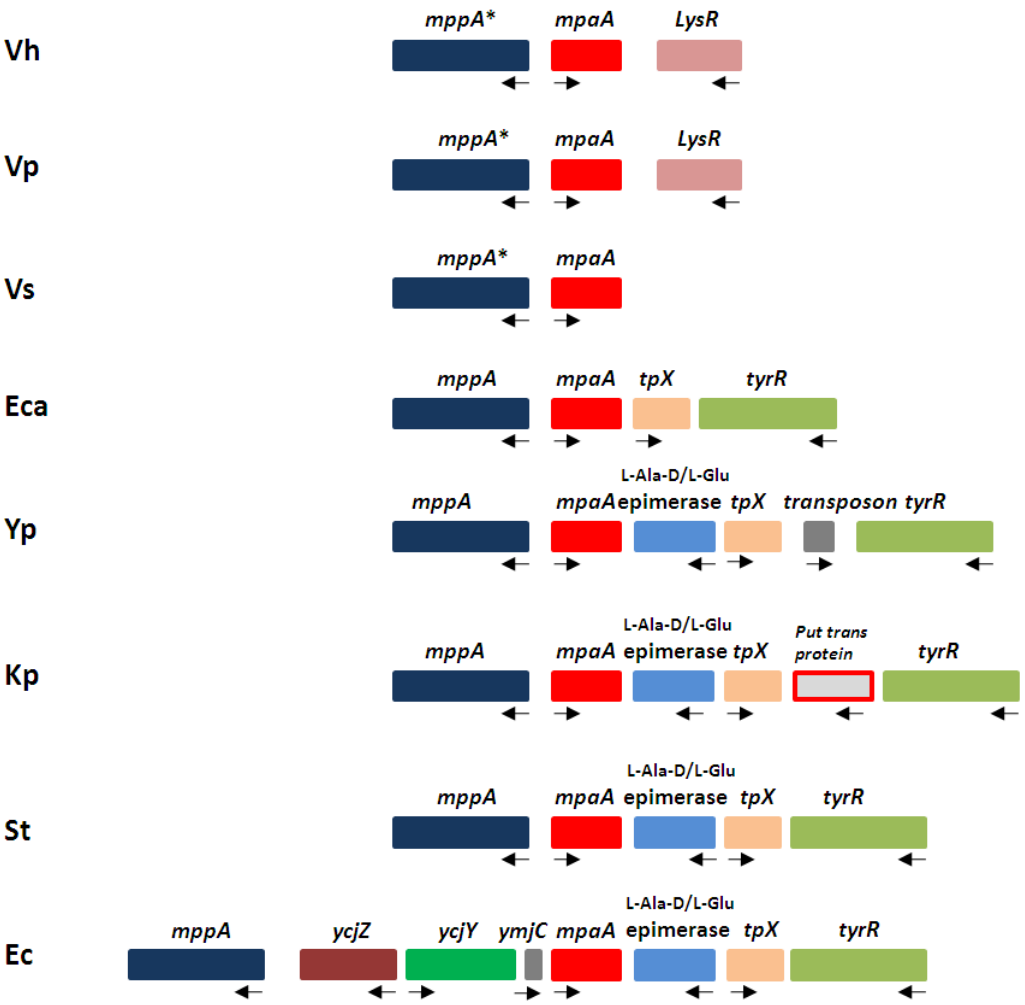


Figure 6.1 Genetic organization of *mppA* and *mpaA* in different bacteria

Putative *mppA* and *mpaA* genes are represented by dark blue and red boxes resepectively. The gene encoding the epimerase is coloured light blue. Genes with similar putative function are represented by boxes coloured appropriately. Ec, *Escherichia coli* K-12; St, *Salmonella enterica* serovar Typhimurium LT2; Kp, *Klebsiella pneumoniae* subsp. Pneumoniae MGH 78578; Yp, *Yersinia pestis* KIM; Eca, *Erwinia carotowora* subsp. Atroseptica SCRI1043; Vs, *Vibrio sp.* MED 222; Vp, *Vibrio parahaemolyticus* AQ3810; Vh, *Vibrio harveyi* ATCC BAA-1116. *mppA** belongs to *Vibrio* and is different from other bacteria.

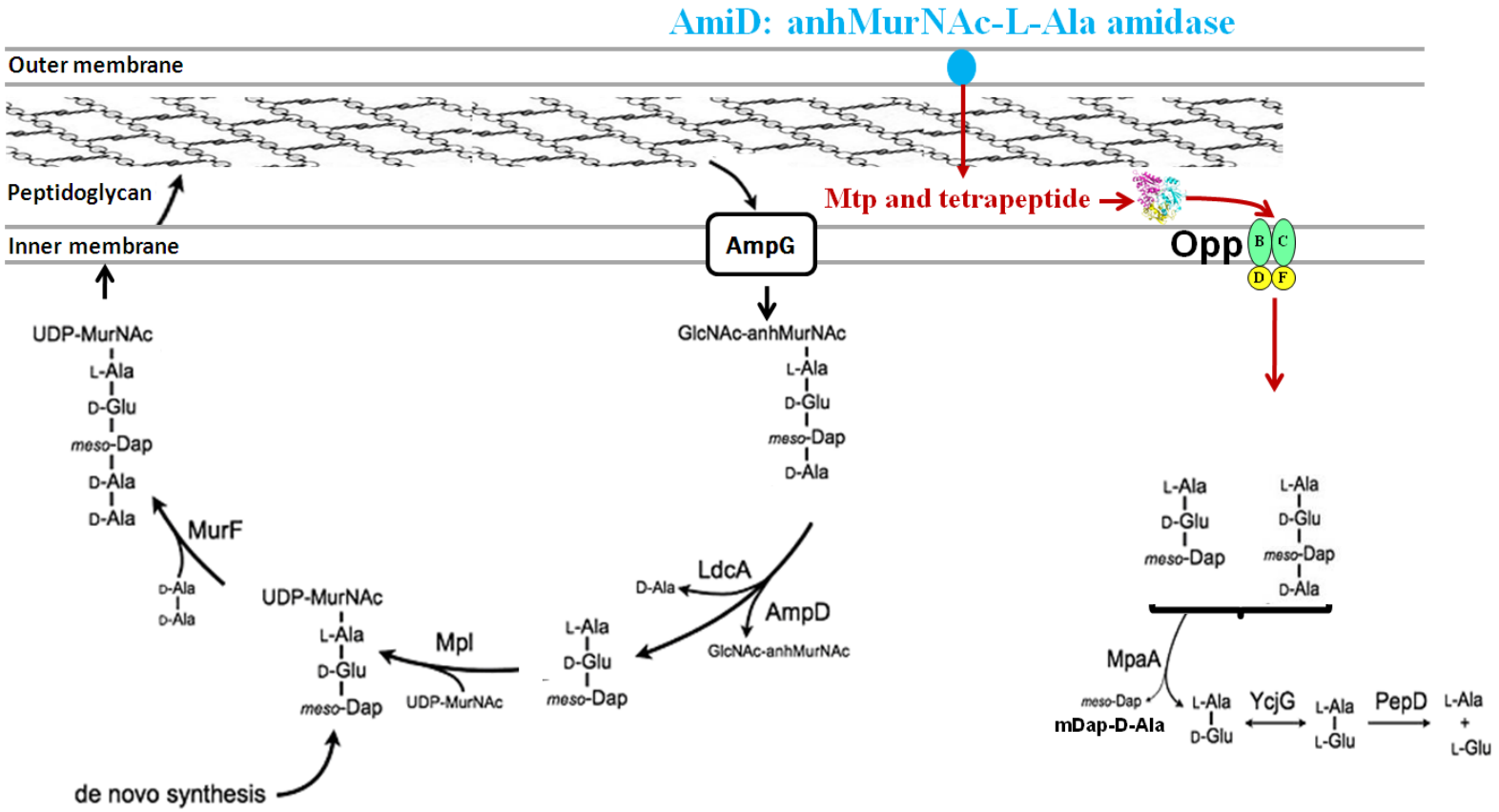


Figure 6.2 Proposed function of MppA and MpaA in a peptide catabolic pathway

MppA binds free Mtp and tetrapeptide present in the periplasm and delivers them to cytoplasm via OppBCDF membrane components. The peptides are cleaved by MpaA in the cytoplasm and the resulting dipeptide L-Ala-D-Glu is epimerized by an epimerase gene (*ycjG*) and cleaved by a peptidase PepD in the cytoplasm.

peptide catabolic pathway which is different from the recycling pathway proposed initially (Figure 6.1and Figure 5.12).

A project was designed to biochemically characterize the enzyme MpaA in terms of its activity and substrate specificity for a range of peptides. As a first step, it was decided to clone the gene encoding MpaA from *E. coli* K-12 in an expression vector.

During the analysis of *mppA* genome organization it was observed that MppA of *Vibrio harveyi* (*Vh*_MppA) is different from MppA of *E. coli* (*Ec*_MppA) as a simple BLAST search using amino acid sequence of *Ec*_MppA did not identify its orthologue from *Vibrio harveyi*. It rather pulled the OppA of *V. harveyi*. This was true also for *Vibrio sp*. MED 222 and *Vibrio parahaemolyticus*. This observation led us to hypothesize that *mppA* may have evolved twice. Thus we also decided to express *Vh_mppA* and *Vh_mpaA* in order to characterize them biochemically and to test if they are functional orthologues of *Ec_mppA* and *Ec_mpaA*.

6.2 Cloning and expression trial of candidate *mpaA* and *mppA* homologues using the pET-YSBLIC3C vector

To characterize the proteins, it was necessary to clone the genes *Ec_mpaA*, *Vh_mpaA* and *Vh_mppA* into a plasmid vector suitable for over production of the proteins. As for *E. coli mppA*, we used the pET-YSBLIC3C expression vector for the expression of recombinant proteins. Regions of the *V. harveyi* ATCC BAA-1116 genome were amplified by PCR using primers *Vh_mpaA*F and *Vh_mpaA*R for *Vh_mpaA* and *Vh_mppA*F and *Vh_mppA*R for *Vh_mppA*. Region of the *E. coli* K-12 genome encoding *mpaA* were PCR amplified using *Ec_mpaA*F and *Ec_mpaA*R primers. These were cloned into pET-YSBLIC3C using ligation independent cloning (LIC) so that the coding sequence was in-frame with a 5′ nucleotide sequence encoding a hexa histidine tag. This resulted in three different expression vectors; pEC6091, pVH6092 and pVH6093, expressing an N-terminally hexa histidine tagged *Ec*_MpaA, *Vh*_MpaA and *Vh*_MppA, respectively. Construction of the desired recombinant plasmid was confirmed by the cleavage pattern analysis with the appropriate restriction endonuclease (data not shown). Recombinant plasmids were also sequenced in the Technology Facility (University of York) to confirm in frame cloning of genes with no mutation.

While the genes appear to be correctly inserted into the expression vector, the expressed protein is the actual target. To ensure that the target is expressed and is of correct molecular weight, a small-scale expression trial was performed. The recombinant plasmids (pEC6091, pVH6092 and pVH6093) were transformed into the *E. coli* expression strain BL21 and used in small scale expression trials. The expression trials were performed using 25 ml LB media in 100 ml flasks and cultures of *E. coli* BL21 strains were grown at two temperatures, 25°C and 37°C. Expression of recombinant protein was induced by addition of 0.5 mM IPTG at an OD_{650} of 0.4-0.6. Growth was monitored by measuring OD_{650} after each hour starting from the time of induction and was continued for four hours post induction with a final reading taken after 24 hours (data not shown). None of the strains failed to grow or showed any obvious reduction in growth rate after induction indicating that the recombinant proteins are not toxic to the cell, assuming they were induced. At each hourly time point, 1 ml of culture was pelleted for analysis by SDS-PAGE.

Synthesis of protein was checked by re-suspending the cell pellets in sample buffer and separating the proteins using SDS-PAGE as described in methods section 2.8. The protein accumulation profile of the strain carrying the *Ec_mpaA* construct, (Figure 6.3 A) shows an IPTG inducible band at around 27 kDa as expected, with maximum accumulation of protein at 4 hours post-induction when grown at 37°C. The synthesis of *Ec*_MpaA was lower at each stage when grown at 25°C.

The SDS polyacrylamide gel of the strain carrying *Vh_mpaA* construct (Figure 6.3B) shows induced synthesis of a 27 kDa protein with maximum accumulation of protein at 24 hours post-induction when grown at 37°C.

The protein accumulation profile of the strain carrying the *Vh_mppA* construct shows an intense band at around 58 kDa after induction with IPTG with almost the same expression at all stages when grown at 25°C and 37°C (Figure 6.3C). However the culture reaches a higher OD_{650} after 24 hours when grown at 37°C, so these conditions were chosen for future over production of *Vh*_MppA to maximise yield.

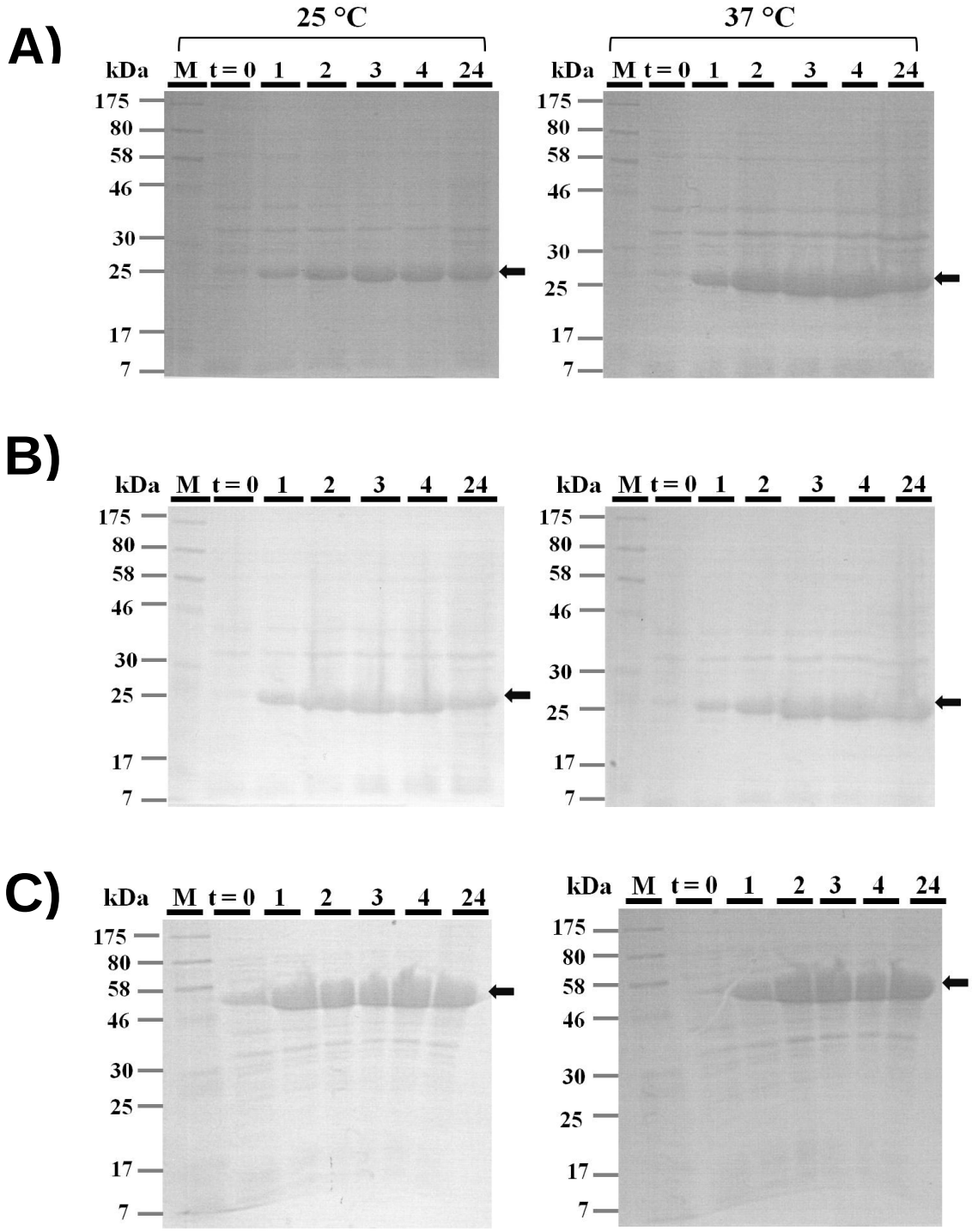


Figure 6.3 Expression of *Ec*_MpaA, *Vh*_MpaA and *Vh*_MppA

Coomassie stained SDS-PAGE gels from expression trials conducted at 25°C (left) and 37°C (right). M is pre-stained protein ladder, time (t) is equal to number of hours post induction and arrow indicates the position of an over produced protein on the gel.

A) Expression of *Ec*_MpaA from BL21 (DE3) pEC6091 **B)** Expression of *Vh*_MpaA from BL21 (DE3) pVH6092 **C)** Expression of *Vh*_MppA from BL21 (DE3) pEC6093

6.3 Large scale expression and purification of *Ec*_MpaA, *Vh*_MpaA and *Vh*_MppA

Large scale production was performed in two litre flasks and cultures of *E. coli* BL21 were grown under the optimum conditions determined during the small scale expression trials for each target. After induction, cells continued to grow for a specific time interval optimised in small scale expression trials and were harvested by centrifugation at 4430 x g for 15 min. The cell pellet was resuspended in buffer A (50 mM Tris pH 8, 500 mM NaCl and 10 mM imidazole) and lysed by sonication. Cell debris was separated from the soluble fraction by centrifugation at 38000 x g for 30 min at 4°C. The supernatant was loaded onto a 5 ml nickel affinity column (GE Healthcare) connected to an AKTA purifier P-900. Recombinant protein containing the hexahistidine tag on the N-terminus was attached to Ni-NTA resin. Weakly bound contaminants were removed by washing the column with 20 CV of buffer A and the recombinant protein was eluted by applying a linear (0-500mM) gradient of imidazole. Fractions were collected throughout and the resulting A_{280} spectrum was used to identify the location of eluted protein. Selected fractions were analysed on SDS-PAGE gels.

The chromatogram of elution of *Ec*_MpaA shows a clear peak (Absorption 280 nm) spanning fractions 25 to 30 (Figure 6.4A). Analysis of the selected elution fractions (25-30) revealed an intense band in fractions 27 and 28 that corresponded to purified protein of around 28 kDa as expected (Figure 6.4B). An intense band was also present at around 28 kDa in the whole cell lysate fraction, indicating that this band was over produced protein and was abundant enough to be visible in the whole cell lysate (Figure 6.4B). Similarly, analysis of the selected elution fractions (19-24) collected during purification of *Vh*_MpaA also revealed an intense band corresponding to purified protein and thus suggesting successful purification of *Vh*-MpaA (Figure 6.5).

The chromatogram of elution of *Vh*_MppA shows a small peak (Absorption 280 nm) corresponding to fractions 30-34 (Figure 6.6A). Analysis of the selected fractions (30-34) using SDS-PAGE showed no band for the purified protein (Figure 6.6B). An intense band was however present in the whole cell lysate fraction at around 58 kDa suggesting that synthesis of protein occurred after induction with IPTG. This suggested either that the protein was insoluble in buffer A or that it was unable to bind to the Ni^{2+}-column

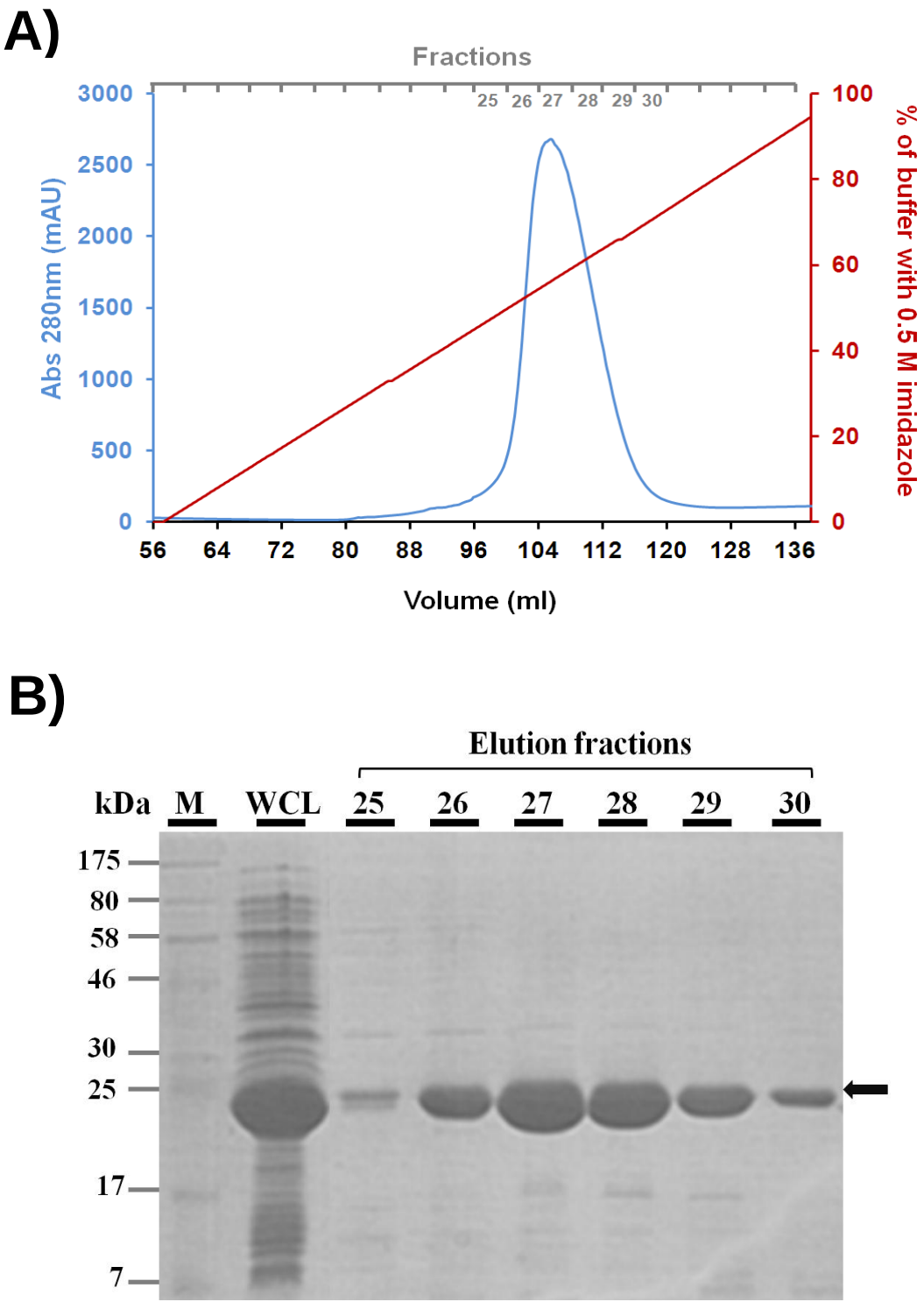


Figure 6.4 Purification of *Ec*_MpaA

A) Chromatogram of the elution of *Ec*_MpaA from Ni^{2+}-column (Abs 280 nm in blue and % of buffer B containing 0.5 M imidazole is in red). **B)** Coomassie stained SDS-PAGE of the whole cell lysate (WCL) and elution fractions (25-29) for the peak in the chromatogram.

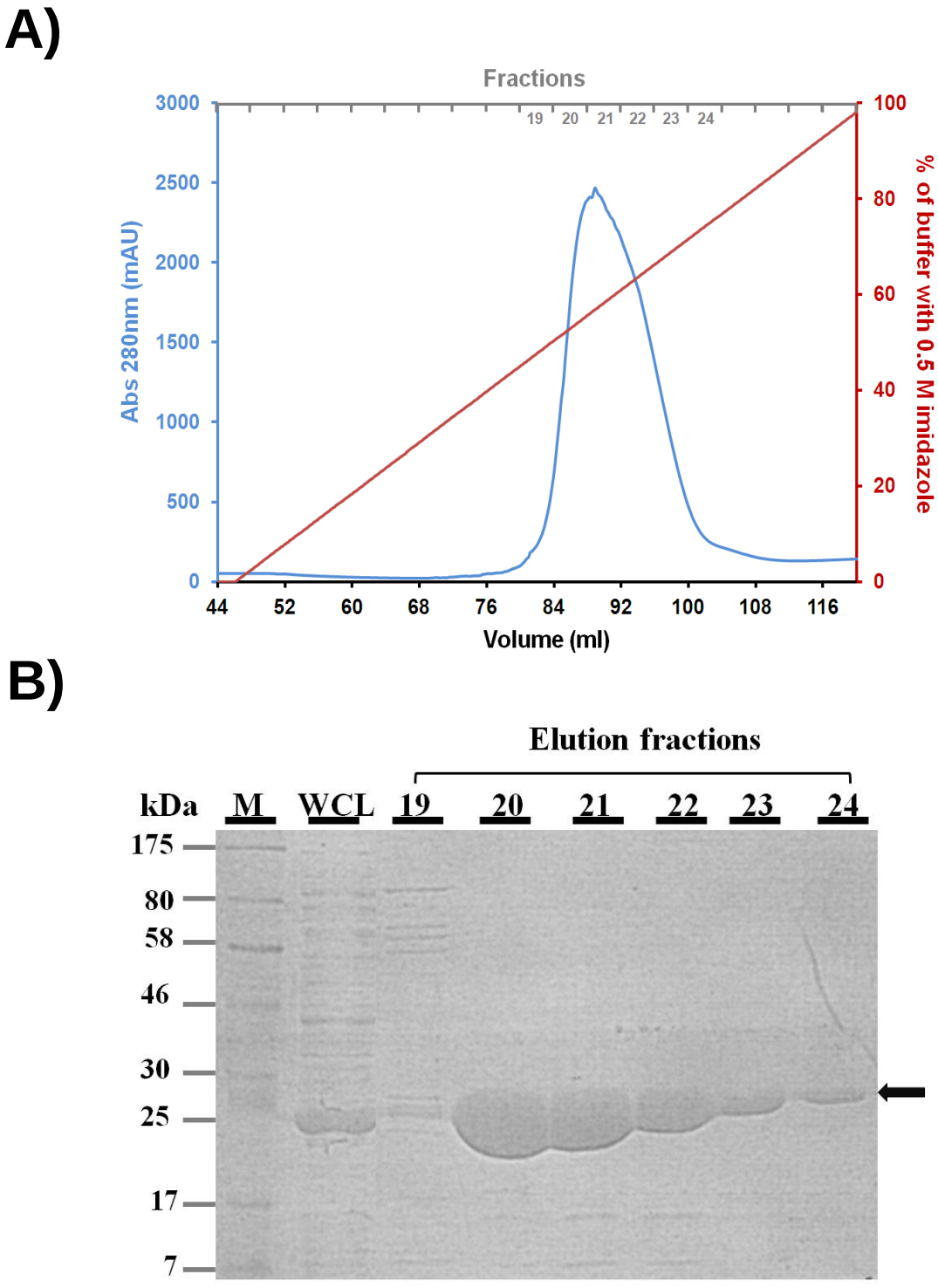


Figure 6.5 Purification of *Vh*_MpaA

A) Chromatogram of the elution of *Vh*_MpaA from Ni^{2+}-column (Abs 280 nm in blue and % of buffer B containing 0.5 M imidazole is in red). **B)** Coomassie stained SDS-PAGE of the whole cell lysate (WCL) and elution fractions (19-24) for the peak in the chromatogram.

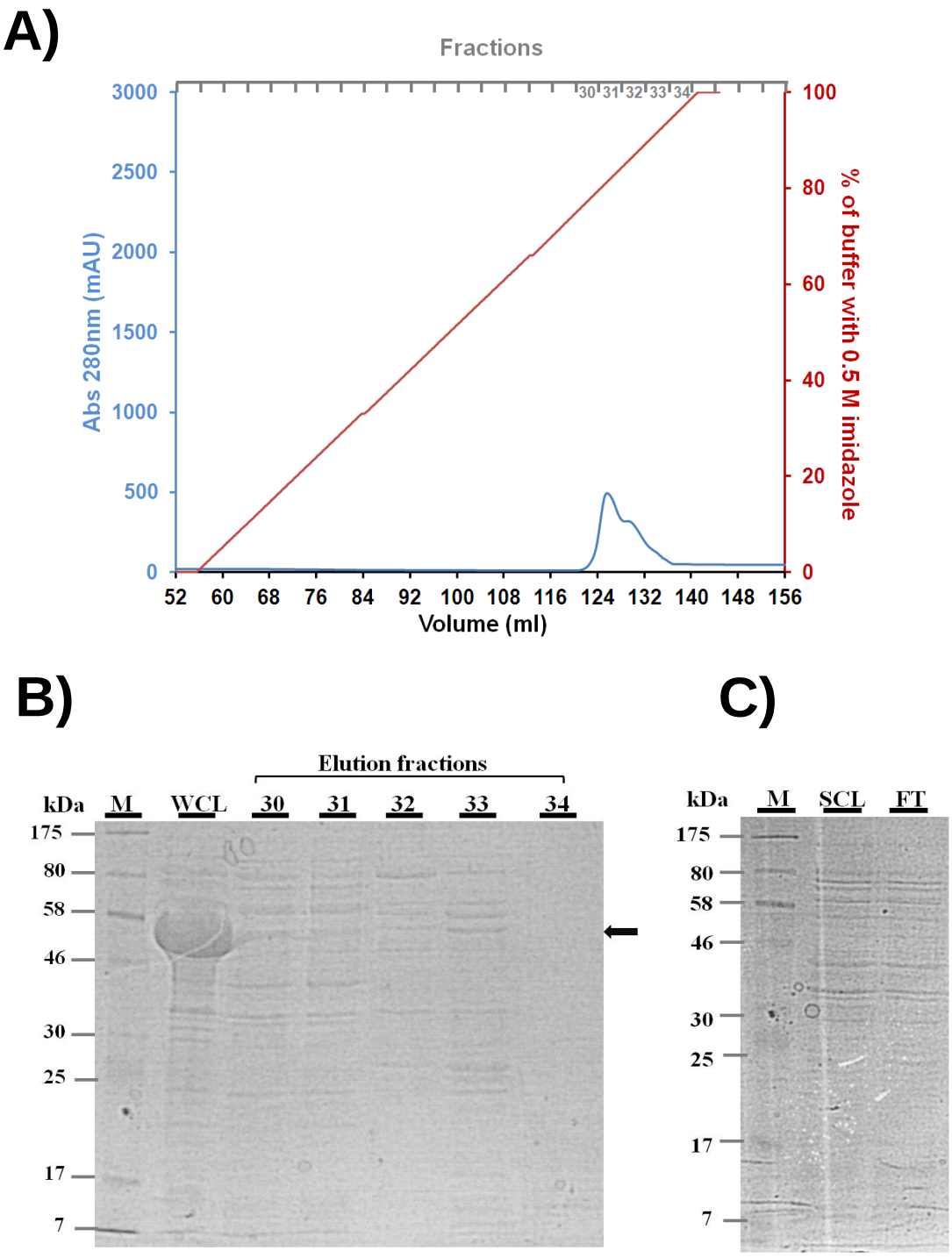


Figure 6.6 Purification of *Vh*_MppA

A) Chromatogram of the elution of *Vh*_MppA from Ni^{2+}-column (Abs 280 nm in blue and % of buffer B containing 0.5 M imidazole is in red). **B)** Coomassie stained SDS-PAGE of the whole cell lysate (WCL) and elution fractions (30-34) for the peak in the chromatogram. **C)** Coomassie stained SDS-PAGE of soluble cell lysate (SCL) and flow through (FT) from column

and flowed through the column. This was investigated by analysing the soluble cell lysate fraction and flow-through fractions. SDS-PAGE gel of the soluble cell lysate showed that an intense band for over produced protein was absent proving that the recombinant protein was insoluble in buffer A (Figure 6.6C). Many efforts were made to improve the solubility of *Vh*_MppA by trying different buffers of various compositions without success. Thus we decided to drop *Vh*_MppA from the list of targets and focussed on the putative amidase enzyme (MpaA).

The elution fractions containing purified protein (*Ec*_MpaA and *Vh*_MpaA) were pooled together and the concentration of protein was determined using a NanoDrop ND-1000 spectrometer as described in methods section 2.9.2. The total yield obtained after Ni^{2+}-affinity chromatography was 9 mg/l and 13 mg/l for *Ec*_MpaA and *Vh*_MpaA, respectively. At this stage proteins were approximately 99 % pure as judged by SDS-PAGE (Figure 6.7, inset for *Ec*_MpaA). Four residues from the affinity tag (GPAM) remain attached to the protein following HRV 3C cleavage. Fractions containing purified protein were pooled and dialysed in an appropriate buffer for mass spectrometry.

6.4 Mass spectrometry of purified protein

Mass spectrometry was used to ascertain whether the recombinant protein expressed and purified is the one expected and it is not modified in any way during synthesis. ES-MS was used to determine the molecular weight of the proteins as described previously (Result section 3.6). MpaA is suggested to cleave Mtp from its C-terminus liberating a free dipeptide (L-Ala-γ-D-Glu) and Dap as reaction products (Uehara and Park, 2003). Thus we also used mass spectrometry to check the presence of any co-purified reaction product. This was done by recording the spectrum under native conditions, where proteins remain associated to non-covalently attached peptide species (Maqbool *et al*., 2011).

For *Ec*_MpaA, the spectrum recorded under denaturing conditions shows a peak at 26783 ± 1 Da. This is very close to the predicted molecular weight of *Ec*_MpaA after removal of the hexahistidine tag (26782 Da) (Figure 6.7A).

We used ES-MS under native conditions to see if *Ec*_MpaA co-purifies with its substrate or a reaction product as was previously observed for a related protein (Xu *et*

al., 2010). We could not see any peptide species bound to *Ec*_MpaA under native conditions. However we obtained a peak at 26846 Da, this being 64 Da larger than the predicted mass of *Ec*_MpaA (26782 Da) suggesting that a species of 64 Da is bound to protein (Figure 6.7B). As MpaA belongs to a family of metal dependent peptidoglycan amidases and these proteins are Zn^{2+}dependent, with a single Zn^{2+} in the active site (Firczuk and Bochtler, 2007), we propose that the *Ec*_MpaA has co-purified with a Zn^{2+} (molar mass of Zn^{2+} being 65.38 Da) and the 26846 peak in the native condition corresponds to *Ec*_MpaA bound to a Zn^{2+}. Under native conditions we also saw an additional small peak at 53692 Da, which is very close to the predicted mass of a dimer of *Ec*_MpaA bound to two Zn^{2+} (53694.76 Da) (Figure 6.7B) suggesting that *Ec*_MpaA may exist as a dimer in solution. The dimerization of *Ec*_MpaA was further investigated by SEC-MALLS.

The molecular weight of *Vh*_MpaA was determined to be 26313 ± 1 Da using denaturing ES-MS which is in agreement with the theoretical mass of *Vh*_MpaA after removal of the hexahistidine tag (26314 Da) (Figure 6.8 A). A native spectrum of purified *Vh*_MpaA showed that like *Ec*_MpaA, it is also bound to a species of 64 Da. Again we assigned this species a Zn^{2+} which was possibly co-purified with protein. The native spectrum also showed a peak at 52756 Da corresponding to a dimer of *Vh*_MpaA bound to two Zn^{2+} (Figure 6.8B)

6.5 SEC-MALLS of MpaA to study dimerization

As mass spectrometry data suggested that *Ec*-MpaA can also exist as a dimer, we investigated the ability of *Ec*_MpaA to form a dimer in solution by another method called SEC-MALLS. A sample of *Ec*_MppA (120 μl of 1.5 mg/ml in 20 mM Tris buffer pH 8) was applied to a Superdex 200 10/300 GL SEC column (GE Healthcare) at 0.5 ml/min with an HPLC system (Shimadzu), linked to a Wyatt Dawn Heleos MALLS detector and a Wyatt Optilab rEX unit, to determine the refractive index. The buffer was 150 mM NaCl, 20 mM Tris, pH 8. Data were analyzed with the program Wyatt ASTRA version 5.3.4.14. The molecular weight measured by light scattering was 54 kDa ± 1kDa which is close to the calculated molecular weight for the dimer of 53.5 kDa, based on the amino acid sequence and suggests that *Ec*_MpaA exists as a dimer in solution (Figure 6.9A). Dimers of *Vh*_MpaA were also observed using the same procedure (Figure 6.9B).

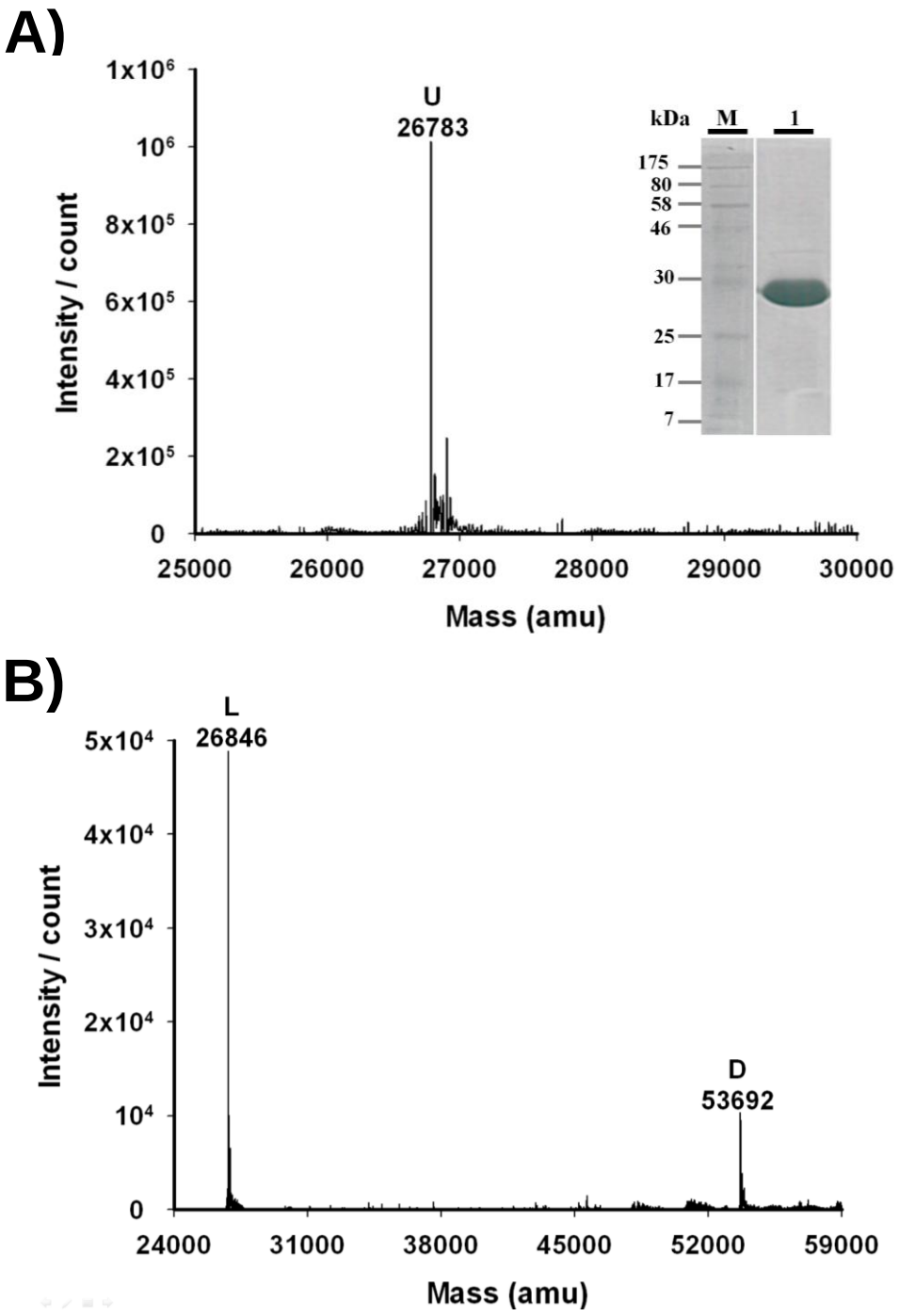


Figure 6.7 ES-MS spectrum of *Ec*_MpaA

A) ES-MS spectrum of 5 μM *Ec*_MpaA under denaturing conditions. The molecular mass of the Zn^{2+} free protein (U) is indicated at 26783 atomic mass units (amu). Inset: Coomassie stained SDS-PAGE gel containing 6 μg of purified *Ec*_MpaA (lane 1) and pre-stained molecular weight markers (M). **(B)** ES-MS spectrum of *Ec*_MpaA under native conditions in water. Peak (L) gives the molecular weight of *Ec*_MpaA bound to zinc and peak (D) gives the molecular weight of *Ec*_MpaA dimer bound to two zinc.

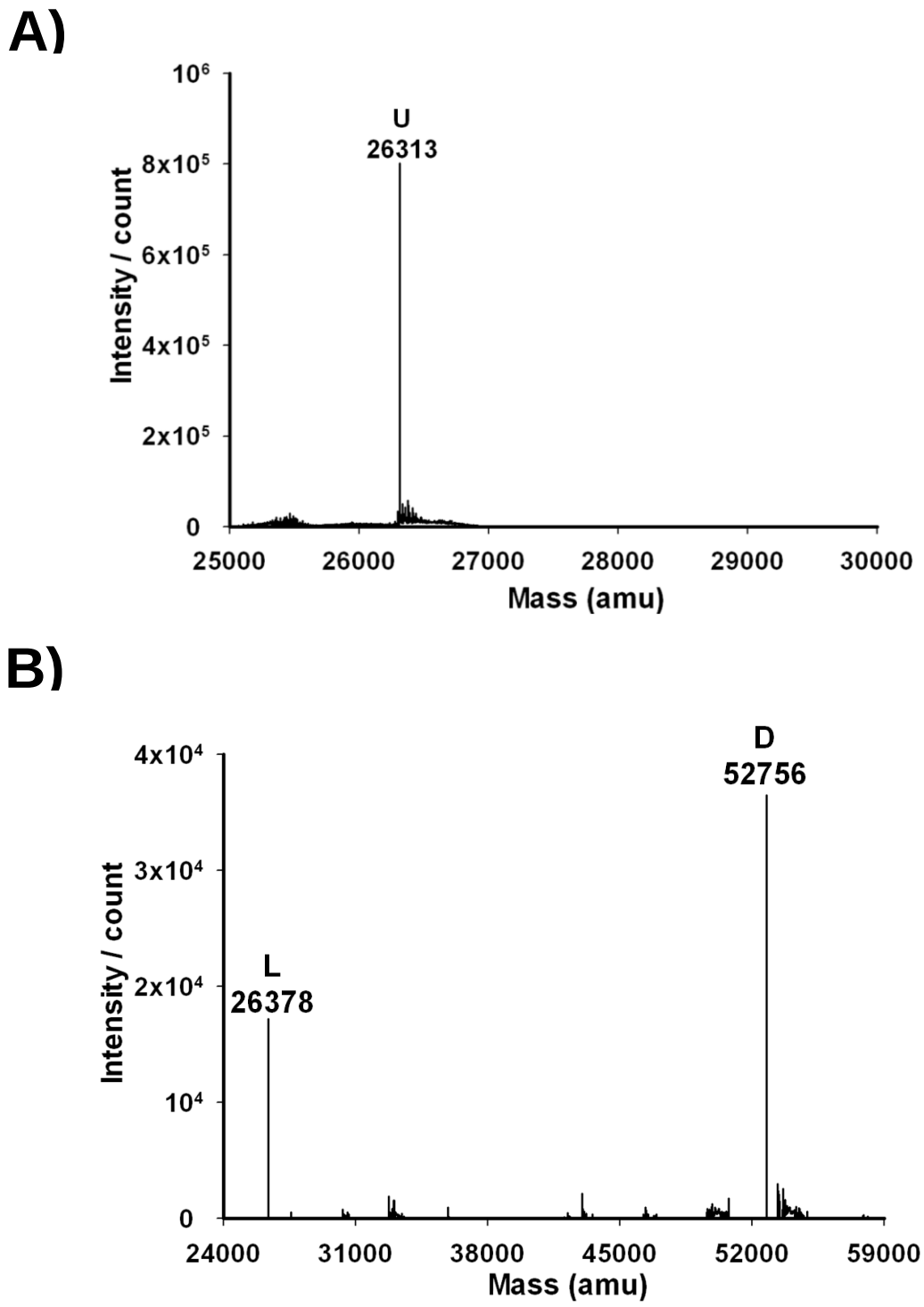


Figure 6.8 ES-MS spectrum of *Vh*_MpaA

A) ES-MS spectrum of 5 μM *Vh*_MpaA under denaturing conditions. The molecular mass of the Zn^{2+} free protein (U) is indicated at 26313 atomic mass units (amu). **(B)** ES-MS spectrum of *Vh*_MpaA under native conditions in water. Peak (L) gives the molecular weight of *Vh*_MpaA bound to zinc and peak (D) gives the molecular weight of *Vh*_MpaA dimer bound to two zinc.

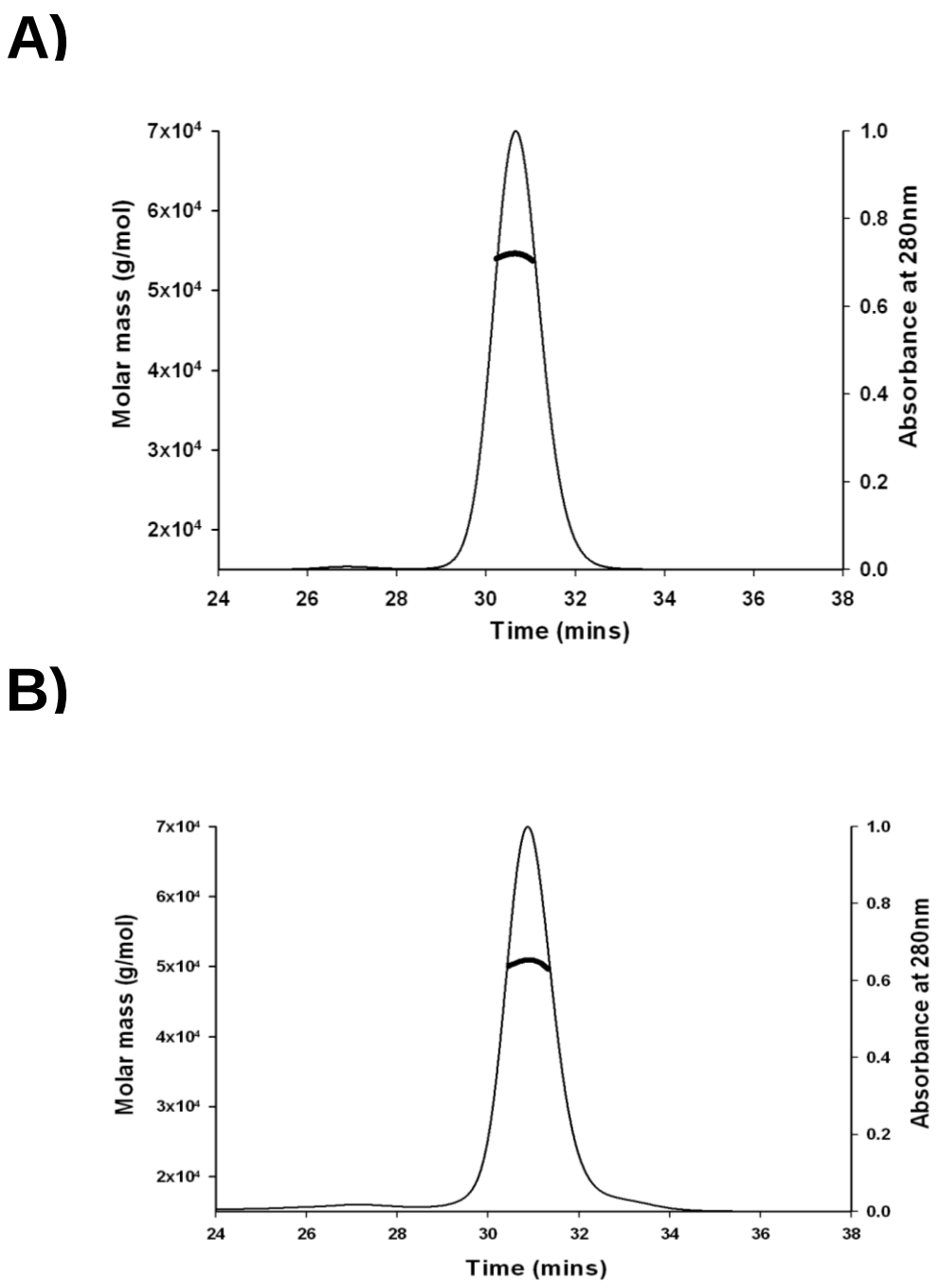


Figure 6.9 Molecular mass of *Ec*_MpaA and *Vh*_MpaA determined by SEC-MALLS

A SEC-MALLS trace of *Ec*_MpaA **(A)** and *Vh*_MpaA **(B)** showing the relative absorbance at A_{280} trace normalized to 1 (thin line) and the molar mass of the species present in the sample (thick line).

6.6 Biochemical characterization of MpaA

The work in section 6.6 was performed by Dr. Mireille Herve (University of Paris-Sud, France) as a part of collaboration. Purified enzyme was sent to her lab and she performed the enzymatic assays to check the substrate specificity and to measure activity of the enzyme.

6.6.1 Enzymatic assays

MpaA is an amidase that hydrolyses the γ-D-Glu-*meso*-Dap bond in L-Ala-γ-D-Glu-*meso*-Dap tripeptide. The standard MpaA activity assay consisted in following the hydrolysis of the tripeptide in a reaction mixture (25 μl) containing: 100 mM Tris-HCl buffer, pH 8, 0.2 mM tripeptide L-[^{14}C]Ala-γ-D-Glu-*meso*-Dap (40 Bq) and MpaA enzyme (5 ng of protein in 5 μl of 50 mM Tris buffer, pH 8, containing 0.5 mg.ml^{-1} of BSA). After 30 min of incubation at 37°C, the reaction was stopped by adding 5 μl of acetic acid. Radiolabeled substrate and reaction product L-Ala-γ-D-Glu-*meso*-Dap and L-Ala-D-Glu were separated by thin layer chromatography (TLC) on silica gel plates LK6D (Whatman), using ethanol/1M ammonium acetate pH 3.8 (5/2) as a mobile phase. The radioactive spots were located and quantified with a radioactivity scanner (model Multi Tracemaster LB285; Berthold). When other compounds were tested as substrates, assay conditions were modified: the amount of MpaA enzyme was appropriately adjusted (from 5 ng to 10 μg), as well as the time of incubation (from 30 min to 16 h). Some enzymatic assays used unlabelled compounds as substrates; in that case, reaction mixtures were analyzed by HPLC, using a C_{18} ODS Hypersil 3 μm column (250 × 4.6 mm; Grace). Elution was performed either using isocratic conditions (0.05% TFA) or by applying a gradient of acetonitrile in 0.05% TFA (from 0 to 20% acetonitrile in 30 min) at a flow rate of 0.6 ml.min^{-1}. Detection was at 214 nm. In some cases, the enzymatic digestion of tripeptide derivative compounds was also analyzed by measuring the release of the C-terminal amino acid residue (Dap or lysine) using a Hitachi L8800 amino acid analyzer (ScienceTec).

6.6.2 Enzymatic activity and kinetic parameters of MpaA

Preliminary assays were carried out to determine the optimal conditions for enzyme activity. Determination of *Ec*_MpaA activity as a function of pH in the standard assay conditions (substrate: tripeptide L-Ala-γ-D-Glu-*meso*-Dap) showed a low pH

dependence between pH 7.2 and pH 9, with an optimum value around 8.0 (Figure 6.10). *Ec*_MpaA seems to require a divalent cation since the activity of the enzyme was inhibited by incubating it with 10 mM EDTA. Although we did not perform any metal analysis to exactly determine the nature of divalent cation, however, considering our mass spec data we proposed it to be a Zn^{2+}. The kinetic parameters of *Ec*_MpaA towards L-Ala-γ-D-Glu-*meso*-Dap were determined at substrate (Mtp) concentrations ranging from 0.07 to 0.7 mM. The data were fitted to the equation $v = V_{max}S/(K_m + S)$, using the Simplex algorithm (MDFitt software developed by M. Desmadril, IBBMC, Orsay, France). Based on a 26,453 molecular mass value for MpaA, the k_{cat} was estimated at 2300 ± 600 min^{-1} and the K_m at 0.52 ± 0.05 mM.

6.6.3 Substrate specificity of MpaA

The activity of *Ec*_MpaA was tested on a series of potential substrates including peptidoglycan precursors and their analogues, as well as intermediates generated during maturation of the cell-wall polymer. As indicated in Table 6.1, *Ec*_MpaA was able to hydrolyze a tripeptide containing L-Lys instead of *meso*-Dap at the third position, although with a much lower efficiency (specific activities of 0.17 and 12 μmoles.min^{-1}. mg^{-1}, respectively). *Ec*_MpaA does not hydrolyze a synthetic tripeptide in which *meso*-Dap is bound to the α-carboxyl group of D-Glu, indicating that the enzyme specifically cuts γ-D-Glu-*meso*-Dap bonds. The dipeptide γ-D-Glu-*meso*-Dap was very poorly hydrolyzed by *Ec*_MpaA (0.4 nmole. min^{-1}. mg^{-1}), indicating that the presence of the N-terminal L-Ala residue was essential but not absolutely required for activity. The tetrapeptide L-Ala-γ-D-Glu-*meso*-Dap-D-Ala was not digested, showing that *Ec*_MpaA does not exhibit endopeptidase activity. A very low activity was detected with MurNAc-tripeptide (0.3 nmole. min^{-1}. mg^{-1}) but neither the nucleotide precursor UDP-MurNAc-tripeptide nor the peptidoglycan recycling product GlcNAc-MurNAc (anhydro)-tripeptide, whose size is greater, were hydrolyzed by *Ec*_MpaA. The activity of *Vh*_MpaA was tested in the same way as that for *Ec*_MpaA. *Vh*_MpaA was able to hydrolyze Mtp, liberating free Dap and a dipeptide (L-Ala-γ-D-Glu) (specific activity of 9.4 μmoles.min^{-1}. mg^{-1}) but it could not hydrolyze tetrapeptide proving that *Vh*_MpaA is a functional orthologue of *Ec*_MpaA.

Table 6.1 Substrate specificity and kinetic parameters of *Ec*_MpaA

Substrate	**Specific Activity** nmoles. $min^{-1}.mg^{-1}$	K_m (mM)	k_{cat} min^{-1}
L-Ala-γ-D-Glu-*meso*-Dap (Mtp)	12,000	0.52 ± 0.05	2300± 600
L-Ala-γ-D-Glu-L-Lys (AEK)	170		
L-Ala-D-Glu-*meso*-Dap	NC		
L-Ala-γ-D-Glu-*meso*-Dap-D-Ala	NC		
GlcNAc-MurNAc(anhydro)-tripeptide	NC		
GlcNAc-MurNAc(anhydro)-tetrapeptide	NC		
MurNAc-tripeptide	0.3		
UDP-MurNAc-tripeptide	NC		
UDP-MurNAc-pentapeptide	NC		
L-Lys-L-Glu-L-Lys (KEK)	NC		
L-Pro-L-Phe-L-Lys (PFK)	NC		
γ-D-Glu-*meso*-Dap	0.4		

NC-No cleavage detected

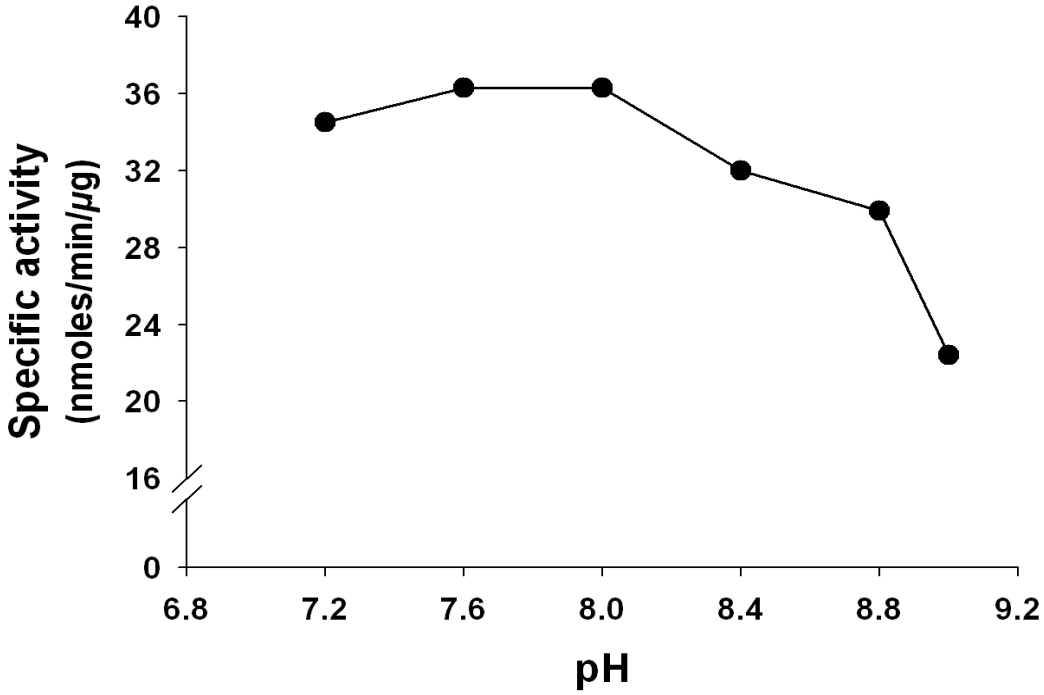


Figure 6.10 L-Ala-γ-D-Glu-*meso*-Dap hydrolyzing activity of *Ec*_MpaA as a function of pH

The optimum pH for MpaA activity was evaluated by incubating 26 ng of MpaA with 0.5 mM [^{14}C] tripeptide in Tris-HCl buffer at pH from 7.2 to 9.0, for 30 min at 37°C, in a final volume of 25 μl.

6.7 Crystallization of MpaA

Crystals of *Ec*_MpaA and *Vh*_MpaA were grown in 96 well plates using the hanging drop vapour diffusion method. Protein was dialysed in 50 mM Tris pH 8 and concentrated to 30 mg/ml. Initially crystallization trials were set up at two concentrations of protein (15 mg/ml and 30 mg/ml) using a Mosquito Nanolitre Pipetting robot. Both His-tagged and untagged versions of the proteins were used for crsytallization. A range of commercially available screens; Hampton I, II, PACT, Index and Clear Strategy Screens CSSI and CSSII were used in the crystallization trials. The plates were sealed to avoid evaporation and then incubated at 20 ºC and examined regularly for the appearance of crystals. No crystals were observed for *Ec*_MpaA but the *Vh*_MpaA (His-tagged version only) produced crystals with 1 M lithium sulphate and 2% PEG 8000 (Hampton I condition 49). However these crystals were in the form of clusters (Figure 6.11A,B,C) and efforts to obtain single crystals by crushing the cluster were not successful due to their small size and poor diffraction. Thus the condition was optimised in order to get single crystals. Optimization experiments were performed by setting crystal trials manually in 24 well plates as described in methods section 2.12. Optimization involved varying the concentration of lithium sulphate (0.8 M to 1.3 M) and adding dioxane (1 %) which is known to reduce clustering of crystals (Personal communication with Elena Blagova). These efforts resulted in production of single crystals (Figure 6.11D). One of the crystals was transferred to a precipitant solution containing 30 % glycerol for cryoprotection, flash cooled in liquid nitrogen and tested for X-ray diffraction on the YSBL home source (Rigaku MicroMax 007HF generator). The crystal diffracted to 3Å spacing and was stored in a Dewar flask containing liquid nitrogen and later on sent to the Diamond Light Source Synchrotron facility, Oxford, UK for X-ray data collection. The native crystal diffraction data were collected to 2.3 Å at Diamond Light Source on the beam line I03 at 100 K at a wavelength of 0.97630 Å.

As native mass spectrometry showed that zinc is bound to the protein, a fluorescence scan was also performed at the synchrotron to establish whether zinc is present in the crystal. It could allow collecting anomalous data for phasing, the fluorescence scan showed no zinc bound.

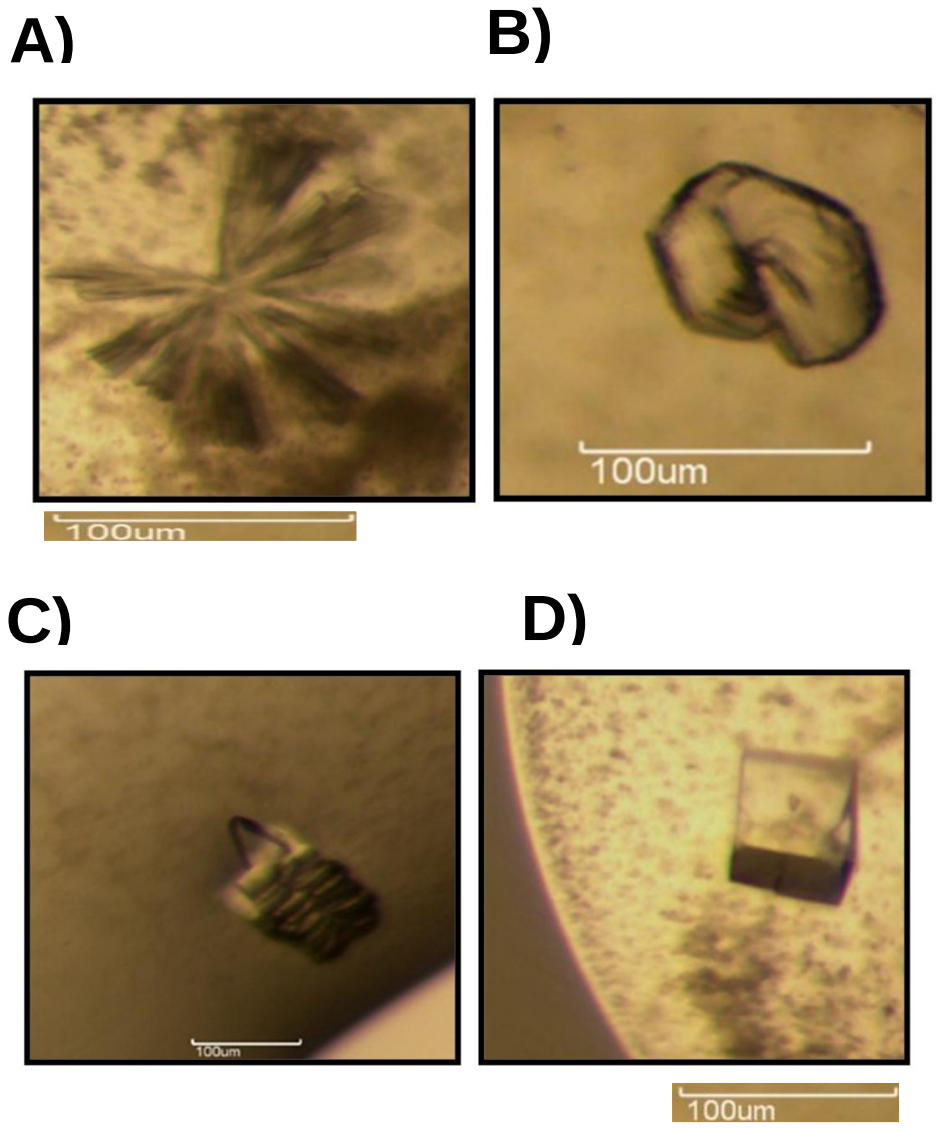


Figure 6.11 Crystals of *Vh*_MpaA

A-C) Images of crystals for target protein *Vh*_MpaA obtained from 96-well plate preliminary screens (well condition 1 M lithium sulphate and 2 % PEG 8000). D) Image of the crystal for target protein *Vh*_MpaA obtained from 24 well plate grown with 0.9 M lithium sulphate 2 % PEG 8000 and 1% dioxane)

Although there are structures of carboxypeptidases in the PDB, none had sufficient amino acid sequence identity with *Vh*_MpaA to allow solution of phase problem using a molecular replacement method. Thus it became a priority to produce the selenomethionine derivative (SeMet) of *Vh*_MpaA in order to get phases and solve the structure.

6.8 Selenomethionine derivative of *Vh*_MpaA

6.8.1 Expression trials of SeMet*Vh*_MpaA

Selenomethionine labelled *Vh*_MpaA was produced in a methionine auxotroph strain, B834 (DE3) of *E. coli*. The recombinant plasmid (pVH6092) was transformed into the expression strain B834 (DE3) and used in small scale expression trials. The expression trials were performed using 25 ml enhanced minimal medium containing selenomethionine as sole source of methionine in 100 ml flasks and cultures of *E. coli* B834 (DE3) were grown at two temperatures, 25 °C and 37 °C. Expression of recombinant protein was induced by adding 0.5 mM IPTG at an OD_{650} of 0.4-0.6. 1 ml of culture was pelleted after 4 hours of induction for analysis by SDS-PAGE. The cells was re-suspended in buffer A (50 mM Tris pH 8, 500 mM NaCl and 10 mM imidazole) and lysed by sonication while using a micro tip. The lysed cells were centrifuged at 38000 x g for 5 minutes to get a soluble fraction. Total and soluble fractions were analysed by running SDS-PAGE gels (Figure 6.12). An intense band could be seen at around 28 kDa in both fractions suggesting that like native protein SeMet labelled protein is also soluble in buffer A. The cells achieve high OD_{650} after 4 hours when grown at 37 °C, so these conditions were chosen for future over production of SeMet*Vh*_MpaA to maximise yield.

6.8.2 Large scale production and purification of SeMet*Vh*_MpaA

Large scale production was performed in two litre flasks and cultures of *E. coli* B834 (DE3) strain were grown under the optimum conditions determined during the small scale expression trials. After induction, cells were grown for 4 hours and harvested by centrifugation at 4430 x g for 15 min. The purification of the SeMet*Vh*_MpaA was performed using a 5 ml nickel affinity column and following same procedure as for native protein. Selected fractions obtained during the course of purification were analysed on SDS-PAGE gels (Figure 6.13).

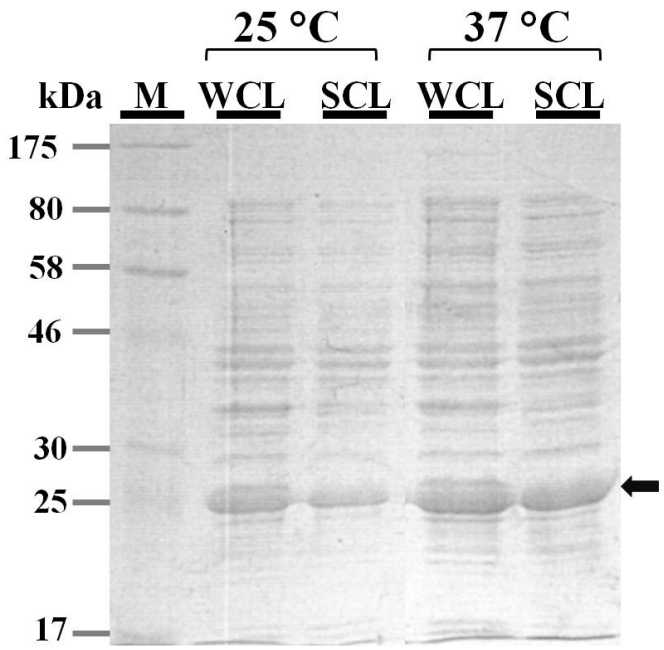


Figure 6.12 Expression of SeMet*Vh*_MpaA

Coomassie stained SDS-PAGE gel showing the accumulation of SeMet*Vh*_MpaA from B834 (DE3) grown at 25°C and 37°C for four hours after induction with IPTG. M is pre-stained protein ladder, WCL is whole cell lysate after sonication and SCL represents the soluble cell lysate. Arrow shows the position of over produced protein on the gel.

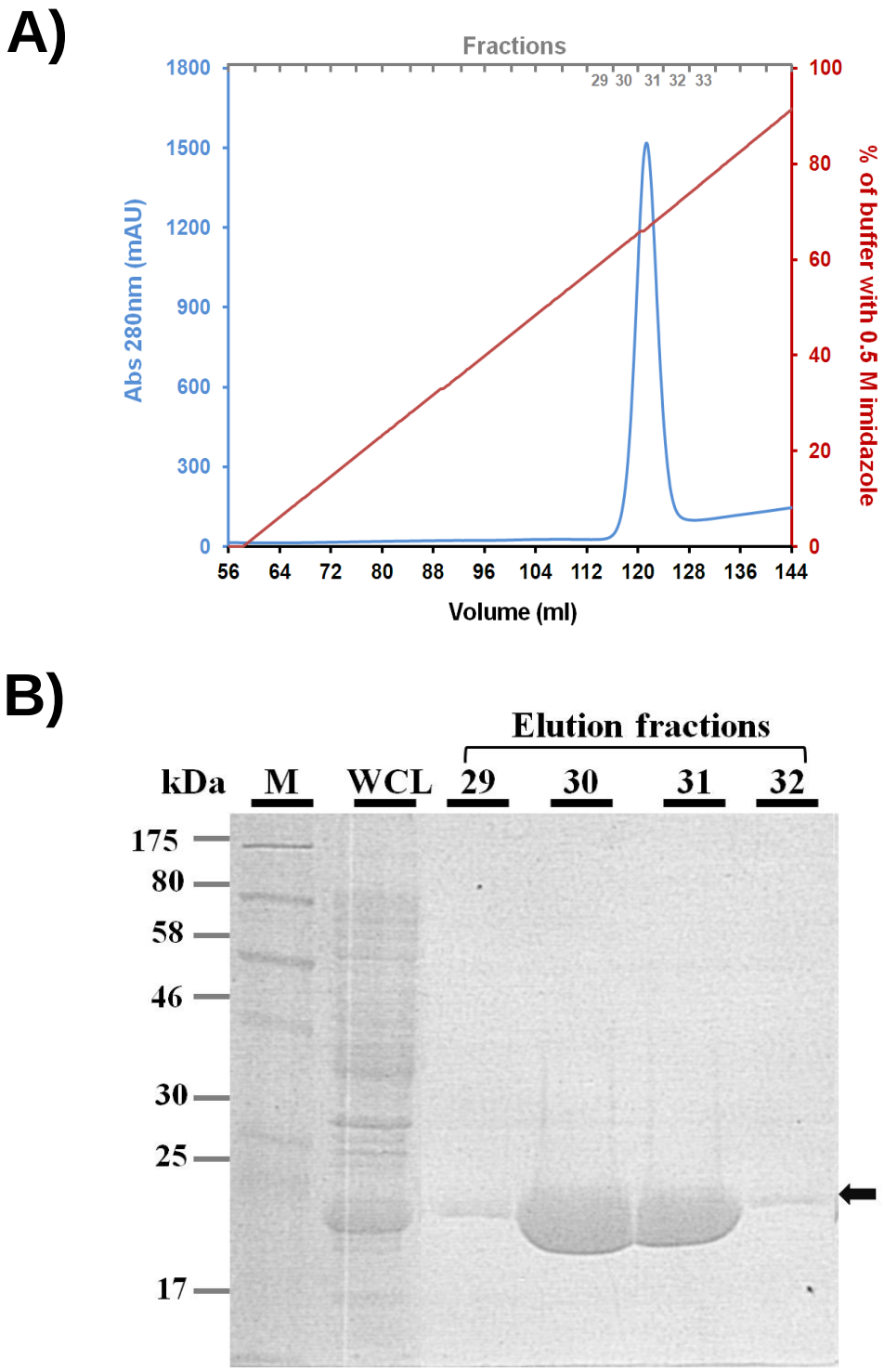


Figure 6.13 Purification of SeMet *Vh*_MpaA

A) Chromatogram of the elution of SeMet*Vh*_MpaA from Ni^{2+}-column (Abs 280 nm in blue and % of buffer B containing 0.5 M imidazole is in red). **B)** Coomassie stained SDS-PAG of the whole cell lysate (WCL) and elution fractions (29-32) for the peak in the chromatogram.

6.9 Crystallization of SeMet*Vh*_MpaA and data collection

Purified SeMet*Vh*_MpaA was dialysed into 50 mM Tris pH 8 and concentrated to 20 mg /ml for crystallization. Initial crystallization trials of purified, His-tagged SeMet *Vh*_MpaA were carried out in sitting-drop 96-well plates using the Mosquito robot. A range of commercially available screens; Index, Hampton I and II, Newcastle and PACT were used in crystallization trials. Crystallization trials were set at two concentrations of protein (10 mg/ml and 20 mg/ml). Due to the absence of crystallization hits from initial crystallization trials, similar conditions where native crystals were obtained were repeated for the SeMet derivative crystals. This gave very small crystals which could not be used in a diffraction experiment (Figure 6.14A). Optimization experiments by changing concentration of precipitant, protein etc gave no success in terms of getting large crystals. At this point all the screens available in the laboratory were used for setting up of crystallization experiment. Fortunately small crystals appeared with 0.12 M MES/imidazole pH 6.5, 0.02 M mix of monosaccharides and a mix of PEG 8000 and ethylene glycol (Morpheus condition F2) but they were also very small (Figure 6.14B) and gave no / poor diffraction. The exact concentration of reagents in Morpheus condition F2 are not available, so optimization experiments were performed using various concentrations of PEG 8000 and ethylene glycol as a mix of precipitant. This was done manually in 24 well plates as described in methods section. This exercise resulted in the production of single crystals with well conditions of 0.12 M MES/imidazole pH 6.5, 0.01 M D-glucose, 8% PEG 8000 and 20 % ethylene glycol (Figure 6.14C). Ten crystals were transferred to a cryoprotectant solution (0.12 M MES/imidazole pH 6.5, 8 % PEG 8000 and 40% ethylene glycol), flash cooled in the liquid nitrogen and tested for X-ray diffraction at YSBL home source (Rigaku MicoMax 007HF) using actor sample changer. Two of the crystals gave diffraction to 2.6 Å and were chosen for X-ray data collection at synchrotron.

Selenomethionine labelled crystals diffracted to 2.17 Å at Diamond Light Source and a SAD dataset was collected on the beam line I02 at 0.97970 Å wavelength. Diffraction data were processed by Xia2 autoprocessing (Winter, 2010). The crystal belongs to point group P6/mmm and absences were consistent with space group $P6_122$ or $P6_522$ with the unit cell dimensions a=b=73.6 Å, c=208.7 Å, α=β=90°, γ=120°. Data collection

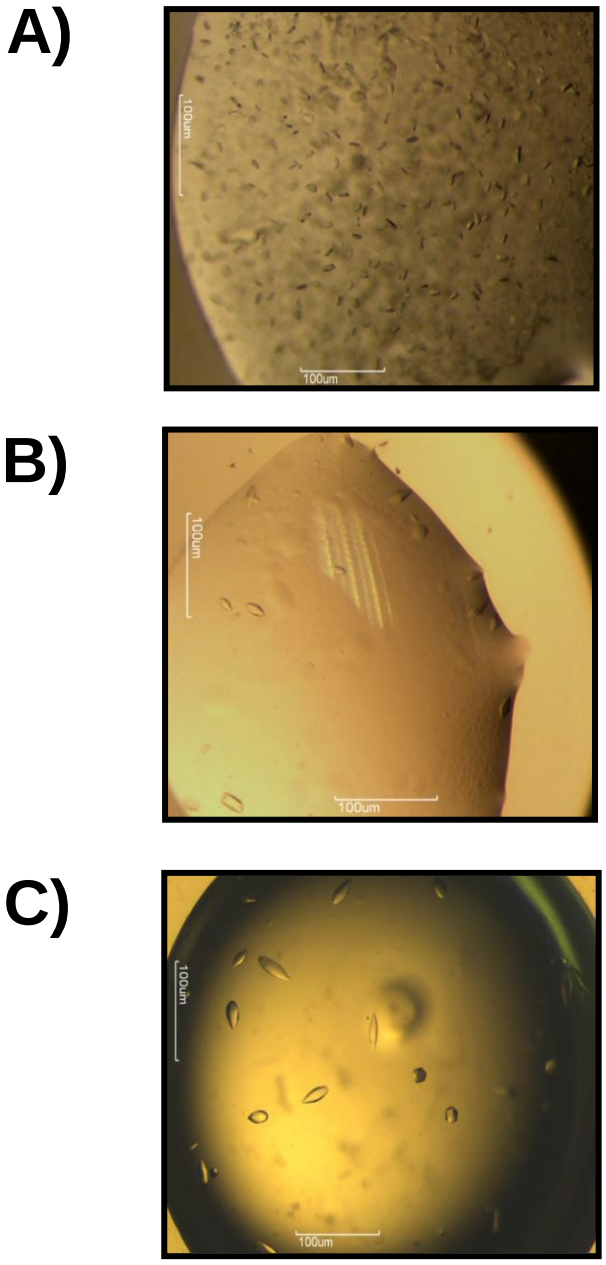


Figure 6.14 Crystallization of SeMet*Vh*_MpaA

Images of the crystals of SeMet*Vh*_MpaA grown during the process of crystallization
A) 0.9 M lithium sulphate 2 % PEG 8000 and 1% dioxane
B) 0.12 M MES/imidazole pH 6.5, 0.02 M mix of monosaccharides and a mix of PEG 8000 and ethylene glycol (Morpheus condition F2)
C) 0.12 M MES/imidazole pH 6.5, 0.01 M D-glucose, 8% PEG 8000 and 20 % ethylene glycol

and processing statistics are summarised in Table 6.2. A Mathews coefficient of 3.14 corresponds to one protein molecule in asymmetric unit plus 60 % solvent (After structure solution we found that the crystalllographic symmetry generated a dimer as expected from the SEC-MALLS results).

6.9.1 Structure solution

ShelxCDE automatic pipeline (Sheldrick, 2010) was used to solve the structure of *Vh*_MpaA using the SAD dataset collected from a SeMet crystal. ShelxD could locate the two Se atom positions expected from one molecule of *Vh*_MpaA. The location of the two Se atoms allowed phasing of the *Vh*_MpaA structure and the space group ambiguity was clearly resolved in the density modification step performed by ShelxE, showing a better result for $P6_522$. The resultant electron density map was of sufficient quality for Buccaneer (Cowtan, 2006) to build 244 residues out of a total of 258 residues. The *Vh*_MpaA model produced by Buccaneer was then refined by subsequent cycles of model building and refinement in REFMAC5 (Murshudov *et al*., 1997) and COOT (Emsley and Cowtan, 2004), respectively to give R_{factor} and R_{free} values of 19.6% and 23.2% respectively. The refinement statistics are given in the Table 6.3.

Table 6.2 X-ray data statistics for the crystal of SeMet*Vh*_MpaA

Data collection and processing statistics	
X-ray source	I02, Diamond
Wavelength (Å)	0.97970
Temperature (K)	100
Space group	$P6_122$ / $P6_522$
Unit cell parameters (Å)	a=b=73.6 Å, c=208.7 Å α=β=90°, γ=120°
Resolution range (Å)	69.57-2.17
Number of unique reflections, overall/outer shell[a]	18606 (1339)
R_{merge}[b] (%), overall/outer shell[a]	10.5 (78.5)
Completeness (%), overall/outer shell[a]	100 (100)
Multiplicity / Redundancy, overall/outer shell[a]	40.8 (42.7)
$I/\sigma(I)$, overall/outer shell[a]	33.5 (7.4)

[a]The outer shell corresponds to 2.23 – 2.17 Å.

[b]$R_{merge} = \sum_{hkl}\sum_i|I_i - \langle I\rangle| / \sum_{hkl}\sum_i \langle I\rangle$ where I_i is the intensity of the ith measurement of a reflection with indexes *hkl* and $\langle I\rangle$ is the statistically weighted average reflection intensity.

Table 6.3 *Vh*_MpaA structure refinement statistics

Refinement statistics	
R_{factor}[a] (R_{free}[b])	19.6 (23.2)
Reflections (working / free)	17580 / 948
Outer shell[c] R_{factor}[a] (R_{free}[b])	20.3 / (26)
Outer shell[c] reflections (working / free)	1093 / 61
Molecules / asymmetric unit	1
Number of protein nonhydrogen atoms	1933
Number of zinc	1
Number of water molecules	46
rmsd bond lengths(Å)[d]	0.02
rmsd bond angles(Å)[d]	2.11
Average B-factor (Å^2)	33
Ramachandran plot[e]	90/10/0.0/0.0

[a]$R_{factor} = \Sigma||F_o|-|F_c|| / \Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively.

[b]R_{free} is the R_{factor} calculated with 5% of the reflections chosen at random and omitted from refinement.

[c]Outer shell for refinement corresponds to 2.172–2.228 Å.

[d]Root mean square deviation of bond lenths and bond angles from ideal geometry

[e]Percentage of residues in most favored/additionally allowed/generously allowed/disallowed regions of the Ramachandran plot, according to PROCHECK.

6.10 Overall structure of *Vh*_MpaA

We crystallized a hexahistidine tagged version of *Vh*_MpaA which consists of 258 amino acid residues out of which the first 22 residues belong to the His tag. The structure of *Vh*_MpaA shows continuous electron density from Val-16 to Asp-257. The first 15 residues (MGSSHHHHHHSSGLE) are not visible in the electron density maps and are assumed to be disordered. There is one Zn^{2+} and 46 water molecules associated with each molecule of *Vh*_MpaA. A ribbon drawing of the *Vh*_MpaA molecule is shown in the Figure 6.15.

The overall fold is similar to eukaryotic carboxypeptidases. A DALI search of the protein data bank revealed that *Vh*_MpaA shares the highest structural similarity with duck carboxypeptidase D (PDB code 1H8L; Z-score = 20.5, with 207 equivalent Cα atoms superimposing with an rmsΔ of 2.7 Å). Two other high scoring matches are the catalytic domain of human carboxypeptidase N (PDB code 2NSM; Z-score = 20.2, with 200 equivalent Cα atoms superimposing with an rmsΔ of 2.6 Å) and *Drosophila melanogaster* carboxypeptidase D (PDB code 3MN8; Z-score = 20, with 206 equivalent Cα atoms superimposing with an rmsΔ of 2.8 Å).

A sequence alignment of *Vh*_MpaA with duck CPD, human CPN and *D. melanogaster* CPD is shown in Figure 6.16. The three eukaryotic structures have an additional C-terminal domain which is absent in *Vh*_MpaA. The human CPN structure consists of two domains, namely a classical CP domain and a transthyretin (TT) domain which is characteristic of CPN/E subfamily members and is absent in *Vh*_MpaA (Keil *et al*., 2007). *Vh*_MpaA superposes on the CP domain of human CPN that contains the catalytic residues. A loop in *Vh*_MpaA spanning residues from Thr-124 to Lys-152 (Figure 6.15) folds away from the human CPN structure and is involved in forming a dimer interface.

Another important high scoring match from the DALI search is human carboxypeptidase A (PDB code 2PCU; Z-score = 19.4, with 213 equivalent Cα atoms superimposing with an rmsΔ of 2.9 Å). The structure of human capboxypeptidase A (human_CPA) has been solved in the presence of a product complex with a cleaved hexapeptide (Bayés *et al*., 2007).

Figure 6.15 Crystal sructure of *Vh*_MpaA

A ribbon representation of the *Vh*_MpaA structure. The amino (N) and carboxyl (C) termini are indicated. A zinc is shown as grey sphere. The loop proposed to be involved in dimerization is coloured black.

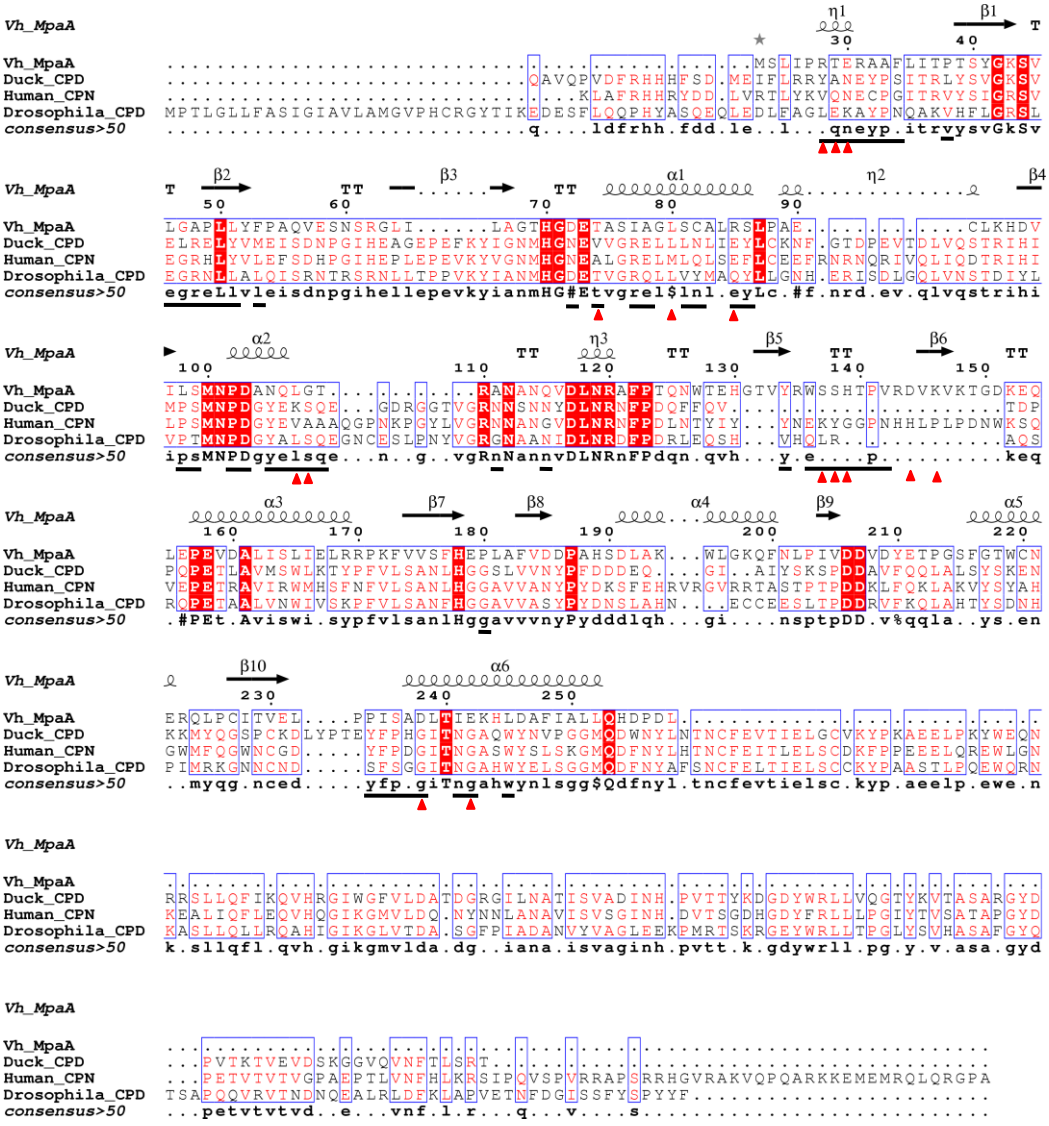


Figure 6.16 Sequence alignment and topological superposition of *Vh*_MpaA with other CP domain-containing proteins

Sequence alignment of *Vh*_MpaA, duck carboxypeptidase D (Duck_CPD), human carboxypetidase N (Human_CPN) and drosophila carboxypeptidase D (Drosophila_CPD) based on an alignment of the 3D structures of the proteins generated using ESPript. The residues involved in dimer interface are underlined. The residues involved in hydrogen bond interactions for stabilizing two subunits in a dimer are marked by red triangles.

6.10.1 Dimerization of *Vh*_MpaA

A striking feature of *Vh*_MpaA is its isolation with higher order quaternary structure. Consistent with mass spec and SEC-MALLS data, analysis of the crystal structure of *Vh*_MpaA showed that it is a dimer (Figure 6.17). Two molecules interact to bury ~ 2052 $Å^2$ of surface area. The dimer interface includes residues which are highly conserved among the closely related structures of CPs and also the residues which are unique to *Vh*_MpaA (Figure 6.16). However out of 25 hydrogen-bond interactions that stabilize the dimer interface none involves highly conserved residues (Figure 6.16).

6.10.2 Active centre

Residues important for catalysis and binding of substrate have been determined in a number of studies on carboxypeptidases A (Bayés *et al*., 2007). The active side residues of human carboxypeptidase A include Arg ^{127}, Arg ^{145}, Glu ^{270} and the Zn coordinating triad His^{69}, Glu^{72} and His^{196}. Superposition of human_CPA and *Vh*_MpaA revealed that the structural organization of the active centre is essentially the same in both proteins (Figure 6.18). Amino acid residues directly involved in catalysis and substrate binding are strictly conserved suggesting that the catalytic mechanism proposed for human_CPA is also applicable to *Vh*_MpaA.

6.11 Conclusion

The genome context of *mppA* reveals that it is closely linked to a gene encoding a putative Mtp carboxypeptidase (MpaA) and thus involved in a peptide scavenging pathaway, which is conserved in γ proteobacteria. The *E. coli* putative carboxypeptidase (*Ec*_MpaA) and its orthologue from *Vibrio harveyi* were purified and characterized biochemically for their activity and substrate specificity. Our mass spectrometric data shows that this enzyme binds to zinc and belongs to family of metallocarboxypeptidase. The structure of *Vh*_MpaA was solved to 2.17 Å. The structure shows that zinc is co-ordinated by a catalytic triad which is conserved in closely related carboxypeptidases. The structure also shows that the enzyme was crystallized as dimer which is in agreement of our data of native mass spectrometry and SEC-MALLS.

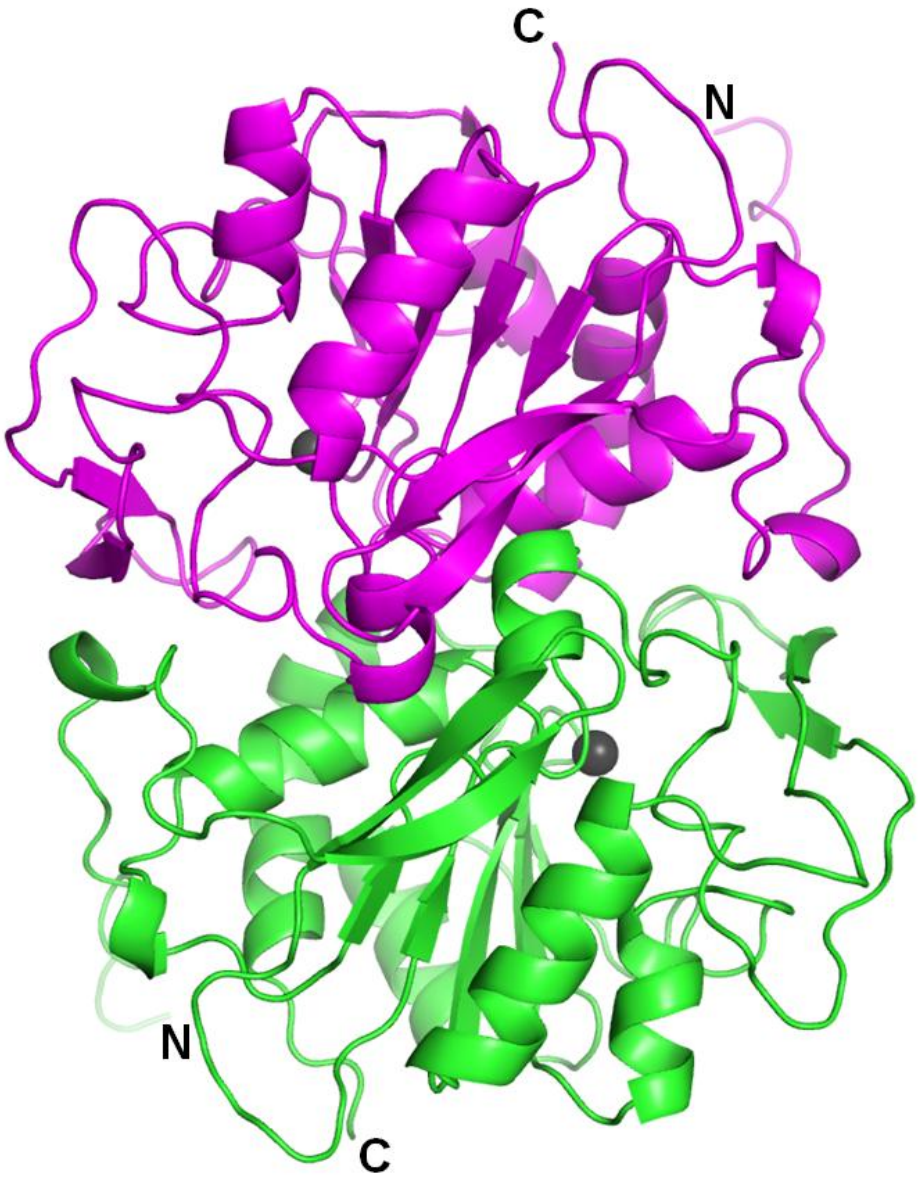


Figure 6.17 Cartoon representation of the *Vh*_MpaA dimer.

Two monomers of *Vh*_MpaA make a dimer. Monomers A and B are coloured as green and magenta respectively.The N and C termini of both monomers are labeled. A zinc is present in each monomer and represented by a grey sphere.

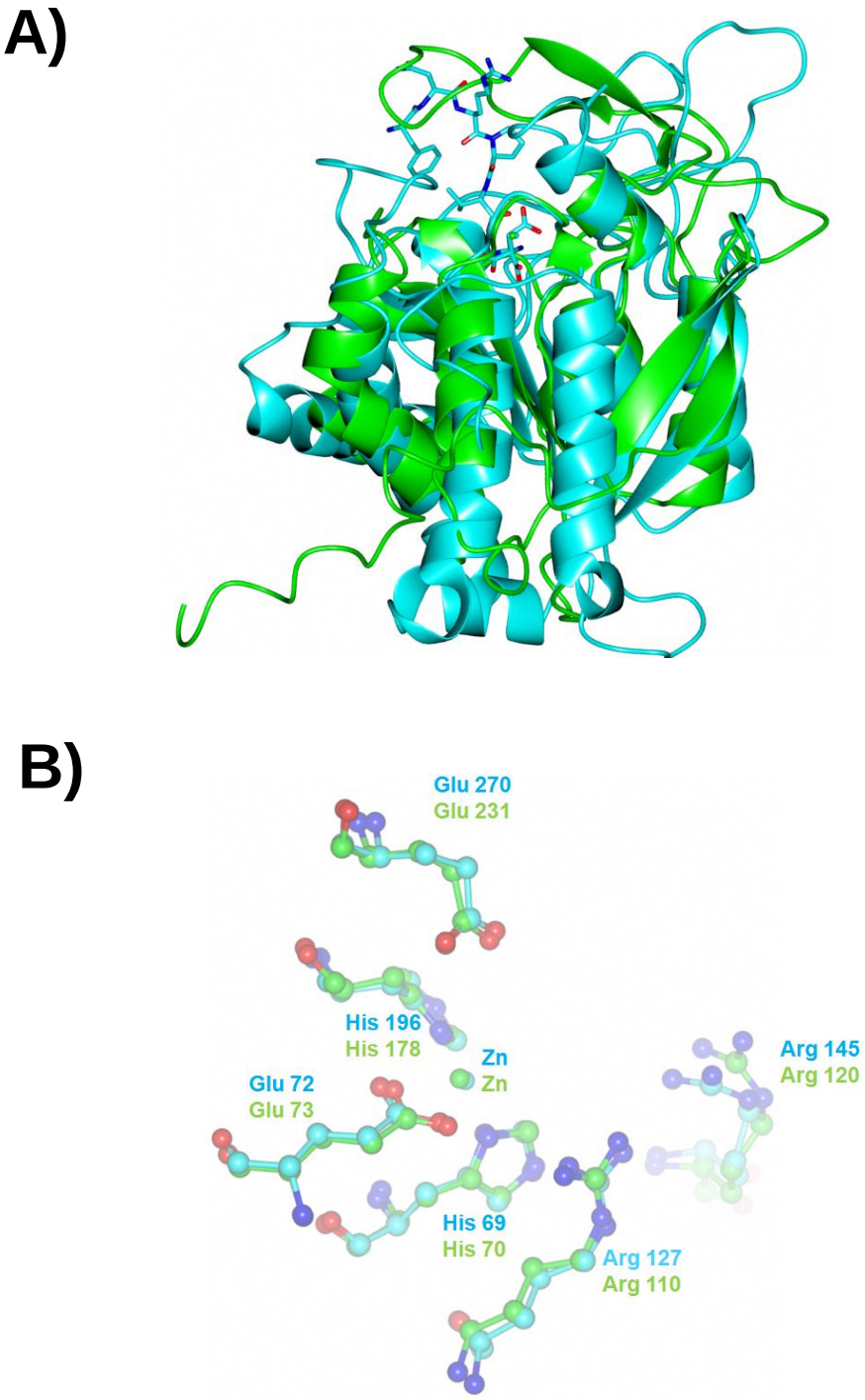


Figure 6.18 Structural superposition of *Vh*_MpaA with human_CPA

A) The proteins were superposed in ccp4mg and are shown in cartoon representation with *Vh*_MpaA in green and human_CPA in cyan. The peptide ligand in human_CPA is shown in stick representation. **B)** The active centre residues of human_CPA are compared to that of *Vh*_MpaA. The active centre residues are represented as ball and stick format, numbered and are coloured same as in A. Zinc is shown as small spheres and coloured according to proteins.

Chapter 7

7 Discussion and Future Work

Oligopeptides present in the environment of bacteria are often recognized and transported by ABC transporters. OppABCDF is a well known ABC transporter of oligopeptides in bacteria where a peptide-binding SBP (OppA) captures the extracellular peptides and delivers them to membrane components (OppBCDF) for translocation (Detmers *et al.*, 2001). It has been established from previous studies that showed that OppA binds oligopeptides irrespective of their amino acid sequence and acts as a general peptide-binding SBP (Tame *et al.*, 1994, Sleigh *et al.*, 1999, Berntsson *et al.*, 2009). However genetic studies suggested the presence of peptide-binding SBPs which are involved in the transport of a specific peptide. These SBPs are paralogues of OppA, for example *E. coli* has two, namely MppA and YgiS, which are both thought to use the OppBCDF subunits. MppA was thought to transport murein tripeptide, a cell wall derived component (Park *et al.*, 1998) whereas YgiS is a putative peptide-binding SBP with unknown substrate specificity. *E. coli* also possesses an ABC transporter for the uptake of the tripeptide, glutathione. There is genetic evidence suggesting that the SBP of this transporter, GsiB is required for the uptake of glutathione (Suzuki *et al.*, 2005).

We investigated a number of peptide transporters that had been predicted to bind and transport specific peptides. In this study we cloned expressed and purified three putative peptide-binding SBPs from *E. coli* namely MppA, YgiS and GsiB in order to characterize their ligand specificity both biochemically and structurally.

Recombinant MppA was expressed intracellularly at very high levels which is often possible with soluble proteins. Co-purification of MppA with a unique peptide (Mtp) represents the first biophysical evidence of a peptide-binding protein binding to a specific peptide. Previously OppA from *L. lactis* was shown to co-purify with a mixture of peptides ranging in length from 7-26 amino acids (Berntsson *et al.*, 2009). Recently, OppA from *E. coli* has been shown to co-purify a variety of peptides as well (Klepsch *et al.*, 2011).

The structure of MppA in complex with its cognate ligand allowed us to compare its binding mode with that of a general peptide-binding protein OppA. It has been

proposed by Tame *et al*. that OppA achieves sequence independent peptide binding by minimising the formation of direct interactions between the protein and the ligand side chains (Tame *et al*., 1994). This is because favourable interactions with one side chain type are likely to be unfavourable with a different side chain type leading to ligand discrimination (Tame *et al*., 1994). This binding model has been borne out in the structures of other general peptide-binding proteins (Levdikov *et al*., 2005, Berntsson *et al*., 2009). It is striking that in MppA the peptide is accommodated in an almost identical location in the protein to that used by OppA with conservation of some important protein/ligand interactions (Figure 4.8). Despite this overall conservation, there is in addition a specific electrostatic interaction between Mtp and the protein, namely the interaction between Arg-402 and the carboxylate of the D-Glu. MppA has most likely evolved from a general peptide-binding protein OppA, and one of the sequence changes that appears to be important in the transformation from a generalist to a specialist is the replacement of Glu-276 by Ala, that would release Arg-404 (now Arg-402 in MppA) so that it can adopt a conformation that would enable it to form a direct side-chain interaction with the ligand and so restrict the range of peptides that can be recognised by this SBP. In addition, the Lys-20 residue that coordinates the first *meso*-Dap carboxylate via its side chain is a glycine in OppA so this particular contact to the ligand is unique to MppA. Our mutagenesis studies demonstrate that residues Ala 274 and Arg 402 are essential for the binding of Mtp by MppA, which supports the predictions that the contacts to the free carboxylate of the D-Glu are important for binding. Further studies can be done by mutagenizing other amino acids involved in ligand binding, e.g. Lys-20 and Ser 413. It would be interesting to mutagenize Glu-276 of OppA to Ala and analyzing the mutant for binding to Mtp. Moreover some other residues of OppA can be mutated in order to investigate the transformation of a generalist (OppA) to a specialist (MppA).

Despite the fact that MppA has unique interactions with the ligand which are absent in OppA, the most striking feature of the MppA:Mtp complex is that Mtp adopts a similar conformation to a conventional α-linked peptide, KEK in OppA (Figure 4.8). Mtp is an unusual peptide ligand as it contains an L-Ala linked to a

D-Glu, which in turn has a γ-linkage to the diaminopimelate (*meso*-Dap) moiety at the C-terminus. It is synthesised non-ribosomally using the MurC-F ligases (Park and Uehara, 2008). Intuitively, one would not expect a peptide with D-stereochemistry at position 2 to be accommodated within an OppA-like protein fold. A change in amino acid chirality at an internal position in a peptide is a major alteration which would be expected to have dramatic structural consequences. In Mtp, the residue-2 carboxylate projects in the direction otherwise occupied by the side chain and *vice versa*. In the MppA:Mtp complex, extension of such a peptide through a γ-linkage would lead to clashes with Arg-402. Similarly, a peptide with a γ-linkage to a Glu will differ markedly from a conventional peptide. In Mtp, the γ-linkage to Glu-2 on its own would cause clashes with the MppA in the vicinity of Arg-402. Remarkably, as the structure of Mtp in the MppA:Mtp complex shows, the combination of these two changes within one peptide represents a much smaller structural perturbation than either change alone, and the murein tripeptide is readily accommodated in the OppA protein framework. The changes are in fact compensating. The free main chain carboxylate of Glu-2 is accommodated by MppA in what would be the position-2 side chain binding pocket of OppA, while the side chain of this Glu takes the direction followed by the main chain in OppA.

As Mtp is slightly longer than a conventional tripeptide chain due to the presence of *meso*-Dap at position 3, the pocket of MppA has evolved to accommodate the longer peptide compared to OppA, which is reflected by the small displacement of a common loop lining the binding site to accommodate Mtp (Figure 7.1). Very recently it has been revealed that OppA of *E. coli* has a strong preference for positively charged peptides and does not bind or binds very weakly to peptides containing acidic residues (Klepsch *et al*., 2011). Clearly Mtp has more acidic residues and thus does not bind to OppA. This is consistent with our results demonstrating that OppA does not bind to Mtp.

Our *in vitro* analysis of ligand binding to MppA, clearly supports the hypothesis that Mtp is the physiological ligand for this protein, however, a number of other studies have suggested additional substrates for MppA using a range of genetic

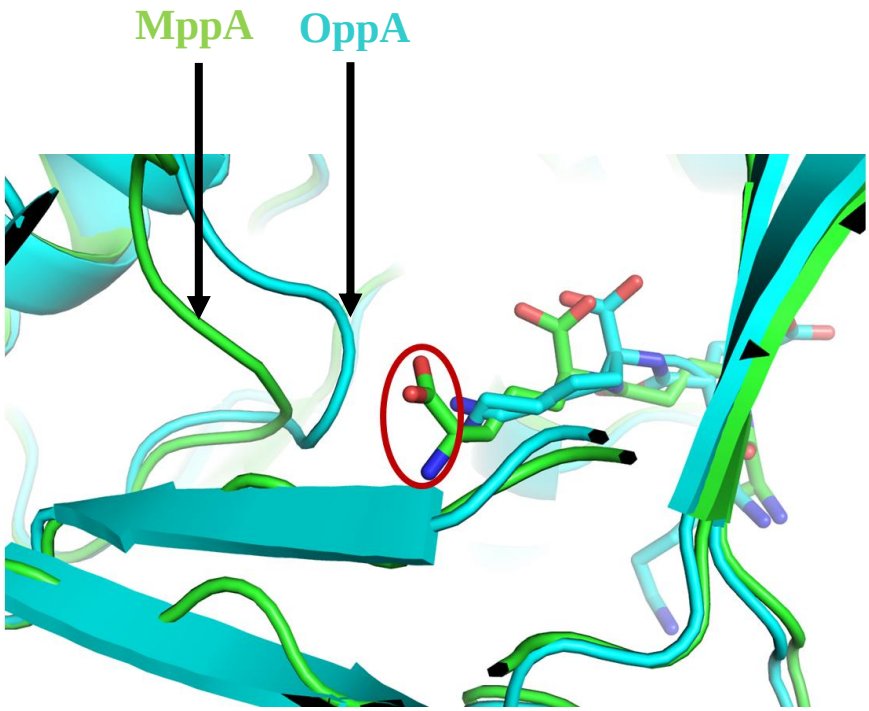


Figure 7.1 Structural superposition of MppA with OppA and displacement of a loop in MppA to accommodate Mtp

The proteins were superposed in PyMol and are shown in cartoon representation with MppA in green and OppA *Salmonella* Typhimurium in cyan. The ligands Mtp for MppA and KEK for OppA are shown in stick representation. Mtp is longer than a conventional tripeptide (KEK) and highlighted by red circle at its C-terminus. Clearly a loop in MppA structure lining the binding cavity around the C-terminus of Mtp is positioned away to accommodate a longer peptide as indicated by the arrow. The position of corresponding loop in OppA is also indicated by arrow. Only the part of structural superposition around binding cavity is shown and the rest is omitted for clarity.

and biochemical approaches. It has been suggested that MppAOppBCDF can transport the standard tripeptide Pro-Phe-Lys (PFK) (Park *et al*., 1998, Letoffe *et al*., 2006). However, we were unable to detect binding of PFK by fluorescence spectroscopy and could detect only weak binding by ITC. It is possible that PFK binds to a site on the protein distinct from that to which Mtp is bound as the thermodynamics of PFK binding were quite different from that which we saw for Mtp. Whereas Mtp binding to MppA and tripeptide binding to OppA are entropy driven, PFK binding to MppA is enthalpy driven. Certainly, we could see no evidence for high affinity binding of PFK to MppA which does not agree with the previous studies (Letoffe *et al*., 2006, Park *et al*., 1998). Also the binding stoichiometry was odd (N = 1.63) indicative of weak interactions between MppA and PFK. Probably PFK is sticking to MppA on its surface away from a defined binding cavity. It has also been reported that MppA can bind haem and its precursor δ aminolevulinic acid (ALA), however, whether this is of physiological relevance is unclear as haem cannot diffuse across the outer membrane of *E. coli* and its binding is competitive with PFK, which we have already demonstrated binds much more weakly to MppA that Mtp. Indeed we were unable to observe binding of ALA to MppA using tryptophan fluorescence spectroscopy. Also, the significant drop in binding affinity that we observed with the AEK peptide, which alters only the water mediated hydrogen bond network in the binding site, suggests that recognition of Mtp is very specific. Together these data suggest that MppA substrates are restricted to Mtp under normal conditions in the cell, which is consistent with our simple observation that MppA co-purifies from *E. coli* with Mtp bound.

While the specificity of MppA for Mtp is clearly established, the physiological role of the MppA protein is not so well understood, although it most likely serves to allow *E. coli* to scavenge Mtp that is either released during turnover of its own cell wall or which is available in the environment. This suggestion is supported by the fact that *E. coli* is known to secrete Mtp during growth (Goodell and Schwarz, 1985). It is also clear that there must be a free pool of Mtp inside *E. coli* as cytoplasmically produced recombinant MppA was purified in high yield as a complex with Mtp. This is consistent with the suggestion that Mtp is the ultimate

breakdown product of murein which is recycled as a building block for further synthesis (Park and Uehara, 2008). While the majority of murein is recycled via AmpG in the form of the muropeptide, there is a known outer membrane localised amidase, AmiD, that releases free tetrapeptide and tripeptide (Uehara and Park, 2007).

Our efforts to find a possible physiological role of MppA revealed that it can also bind to a murein tetrapeptide with two fold higher affinity than Mtp. A cytoplasmic peptidase called LdcA is known to degrade murein tetrapeptide into Mtp thus converting the free tetrapeptide pool into Mtp. As we produced MppA in the cytoplasm, it is possible that LdcA utilizes all the tetrapeptide so that only Mtp is available to bind to MppA. This hypothesis can be checked by expressing MppA in an LdcA deficient strain and then analysing the purified protein. Another possibility is that the MppA:tetrapeptide complex ionizes poorly and is unable to fly during native mass spectrometry. However latter possibility was ruled out by *in vitro* mixing of ligand free MppA and murein tetrapeptide and observing MppA:tetrapeptide complex under native conditions during mass spectrometry (data not shown).

It will be interesting to crystallize MppA in presence of murein tetrapeptide and solve the structure to see how a tetrapeptide is accommodated in the binding cavity. We could not perform this due to limited substrate availability.

Thus it is realistic to propose that MppA scavenges free Mtp and tetrapeptide present in the periplasm. MppA can combine with OppBCDF to form a complete transporter for the translocation of cell wall derived peptides. There is genetic evidence that it can function with both the Opp and Dpp membrane components (Park *et al*., 1998, Letoffe *et al*., 2006). This is an interesting phenomenon where a substrate-binding protein interacts with membrane components of multiple transporters and could be investigated further biochemically.

It is likely that *mppA* evolved by gene duplication and diversification from an ancestral *oppA* gene. To identify orthologues of MppA, the amino acid sequence of *E. coli* K-12 MG1655 MppA was used with a BlastP search. The matches were

found within the Enterobacteriaceae family members including *Salmonella*, *Shigella*, *Yersinia*, *Klebsiella* and *Erwinia* species. This suggested that MppA is restricted to the Enterobacteriaceae family. However a quick look in TransportDB revealed that orphan peptide-binding SBPs were found in distantly related bacteria including *Vibrio* and *Bacillus*. These SBPs being orphan could be considered as homologues of MppA but the high sequence identity with OppA complicates their identification. To try to determine if these are true homologues of MppA, The Seed was used to identify their genome context. The orphan peptide SBPs from *Vibrio* species were present near an amidase encoding gene *mpaA* and were considered true homologues of MppA as the genes *mppA* and *mpaA* were always present together in Enterobacteriaceae. However in Bacillus the gene encoding the amidase is absent and thus the criteria of genome context analysis to identify MppA homologues are not applicable in Bacillus. These orphan peptide SBPs may be homologues of MppA, or of YgiS, or a similar protein with different specificity. As MppA of *Vibrio* is different from MppA of the Enterobacteriaceae it can be speculated that MppA evolved multiple times during the process of evolution (Figure 7.2).

We were able to achieve high levels of expression and purification of GsiB suggesting that like MppA, it is a soluble protein. We demonstrated that GsiB co-purified with endogenously bound glutathione, a finding consistent with the previous genetic data suggesting transport of GSH by an ABC transporter takes place in a GsiB-dependent manner (Suzuki *et al.*, 2005). Cells maintain high concentrations of GSH as it plays an important role in protecting against reactive oxygen species and peroxides. It also acts against xenobiotics and drugs by the formation and excretion of glutathione conjugates. Within the cell glutathione exists predominantly as the reduced form (GSH) (>98%) (Dalle-Donne *et al.*, 2009). Interestingly our FT-MS data showed that an oxidized form of glutathione (GSSG) can also bind to GsiB. Unlike the reduced GSH the oxidized form GSSG does not act as antioxidant and thus its uptake may not be advantageous for cell. Thus the evidence of GSSG binding to GsiB can either be an artefact or it suggests that GSSG performs a different biological role in the

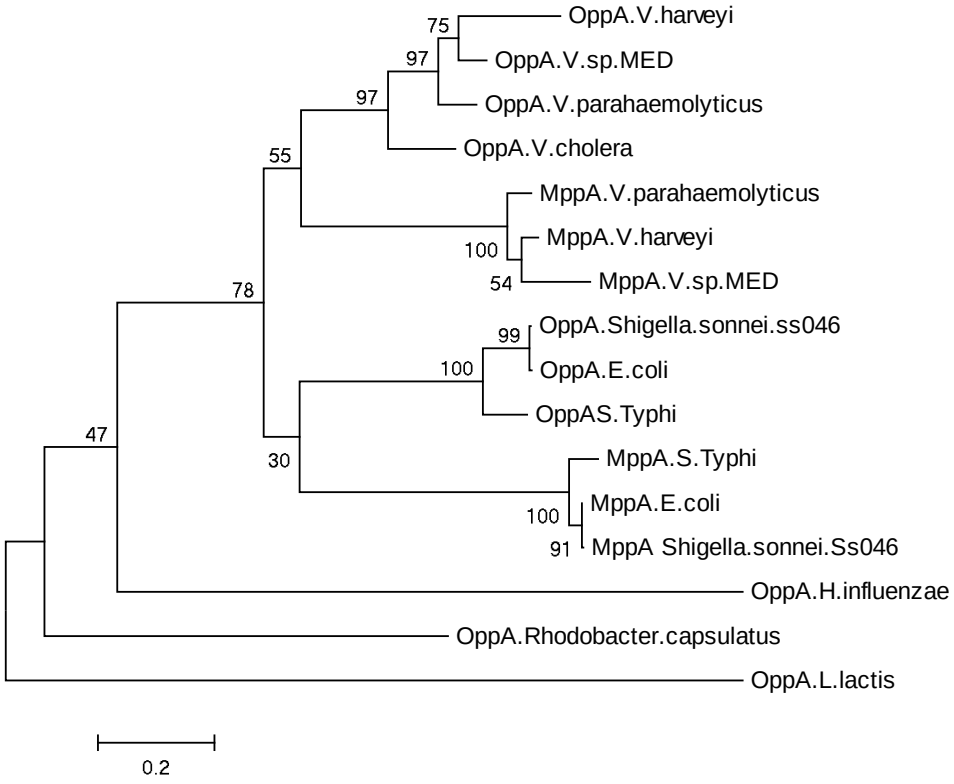


Figure 7.2 Phylogenetic tree based on maximum likelihood method using MppA and OppA protein sequences

The amino acid sequences were collected from NCBI and aligned in ClustaW. Phylogenetic tree was made using MEGA5 software. Clearly OppAs of *Vibrio* and Enterobacteriaceae family fall into separate clusters suggesting that they evolve independently from a common ancestor. Similarly MppAs of *Vibrio* and Enterobacteriaceae fall into separate clusters suggestive of their independent evolution. Node support is provided by bootstrap values.

cytoplasm. Recently *Haemophilus influenzae* glutathione binding protein has been shown to bind and facilitate the uptake of GSH and GSSG to provide a source of sulphur (Vergauwen *et al*.). It can be speculated that *E. coli* uses GSSG as source of nutrients under starvation conditions.

The *ygiS* gene encodes an orphan SBP that has 44 % sequence identity with OppA and 42 % sequence identity with MppA. The identity between YgiS and MppA or OppA is similar, providing no clues to the specificity of YgiS. This suggests that YgiS binds a specific peptide or at the very least a different set of peptides to OppA so that the physiological ligand specificity of the three SBPs, OppA, MppA and YgiS, are not overlapping. Our mass spectrometric analysis showed no ligand bound to YgiS. This suggests that either YgiS is not a peptide-binding protein or it binds to a specific non-endogenous peptide not present in the medium. The ligand could be a signaling peptide from a different strain or species of bacteria present in the environment.

Further work to identify the ligand of YgiS may involve growing bacteria using different media and growth environments and then purifying the protein for biophysical analysis. A library of oligopeptides can also be used to screen peptide ligands.

It would also be interesting to study the interactions of the two orphan SBPs with transmembrane components of Opp and Dpp. Do these SBPs interact with transmembrane components of one transporter or multiple transporters and do they have equal affinity or does the general peptide-binding protein, OppA, take precedence over the other two.

The peptide-binding SBPs can potentially be used in the diagnosis of disease and infection. Many pathogens and diseases like prostate cancer produce specific short peptides that are indicative of their presence. These peptides can be detected using SBPs fused to a reporter gene that will be able to measure the levels of peptides in biological samples. The peptide-binding SBPs can also be used in the elimination of toxic peptides from biological samples for example, several venoms are toxic peptides (Favreau *et al*., 2007). The SBPs can be immobilized on a gel or some

other material from which they can capture and sequester the toxin from a contaminated sample.

Catabolism of murein tripeptide

Cell wall turnover is an enzymatic process that results in the loss of peptidoglycan components and has been reported in many bacteria (Park and Uehara, 2008). The turnover products are generally re-utilized through a process known as cell wall recycling (Park and Uehara, 2008). In *E. coli* the main turn over product (Muropeptide, GlcNAc-anhMurNAc-L-Ala-γ-D-Glu-Dap-D-Ala) is known to be taken up by AmpG permease and further processed by lytic enzymes to release free murein tripeptide (Mtp) in the cytoplasm which then re-enters the peptidoglycan biosynthetic pathway (Park and Uehara, 2008). In addition we propose that free Mtp and tetrapeptide present in the preiplasm can be taken up via MppA:OppBCDF transporter for recycling.

We discovered another fate of free murein tri- and tetrapeptide by studying the genome context of *mppA* in *E. coli* and then carrying out a comparative genomics study. Under nutrient limiting conditions, the bacterium is likely to catabolise Mtp and murein tetrapeptide and uses them as a source of nutrient. Genetic linkage of *mppA* with the gene encoding a putative amidase (*mpaA*) in several γ-proteobacteria is suggestive of a related function for *mppA* and *mpaA* in transport and catabolism of Mtp and murein tetrapeptide in these organisms. There are many examples in nature where the genes encoding a complete transporter are situated next to the genes involved in the downstream metabolism of the solute such that transport of a ligand is coupled to the expression of genes involved in the metabolism of that ligand (Severi *et al.*, 2007). Here we show that a gene encoding an orphan SBP is situated next to one encoding an amidase and the two genes appear to be divergently transcribed from an overlapping promoter. This genome organization where a gene encoding an orphan SBP is situated next to one encoding an amidase enzyme is the first example in the literature.

In the present work, the *E. coli* MpaA protein was successfully over produced and the recombinant protein purified to homogeneity and its substrate specificity and

kinetic parameters were investigated in detail. Previously efforts were made to overproduce MpaA in *E. coli* with no success (Uehara and Park, 2003). The possible reason for this failure was selection of wrong start codon which resulted in the expression of an additional 20 residues upstream of the MpaA sequence. We chose the correct start codon as determined from the EcoGene database of *E. coli* (Rudd, 2000) resulting in successful over production of MpaA.

Our results show that MpaA is a carboxypeptidase that specifically hydrolyzes the γ-D-Glu-Dap amide bond of Mtp, a finding consistent with previous data (Uehara and Park, 2003), which suggested that the MpaA activity was involved only at a very late step in the peptidoglycan recycling process, *i.e.* for the degradation of the free tripeptide into L-Ala-D-Glu and *meso*-Dap.

As the dipeptide *meso*-Dap-D-Ala had been detected in cell extracts of *E. coli* (Goodell and Schwarz, 1985), it was suggested that MpaA would also be able to hydrolyze murein tetrapeptide (Uehara and Park, 2003) and may act as an endopeptidase as well. Our results show that MpaA does not hydrolyze murein tetrapeptide and cannot act as an endopeptidase. The presence of *meso*-Dap-D-Ala in cell extracts of *E. coli* strongly suggests that hydrolysis of murein tetrapeptide is catalyzed by another as yet un-identified endopeptidase in *E. coli*, which warrants further investigation

Bacillus subtilis produces a protein named YkfC which contains a C-terminal NIpC/P60 cysteine peptidase domain. Characterized NIpC/P60 enzymes are almost all γ-D-Glu-Dap or γ-D-Glu-Lys endopeptidases (Xu *et al*., 2010). In a search of the *E. coli* database to find a homologue of *Bs*_YkfC, a region of YdhO was found to display significant homology to the catalytic domain of *Bs*_YkfC. YdhO is encoded by an uncharacterized open reading frame and appears to have a signal peptide (Misra *et al*., 2005). It is predicted that YdhO is responsible for production of m-Dap-D-Ala in the periplasm by cleaving murein tetrapeptide.

Uehara and Park tested the activity of cell extracts prepared from *E. coli mpaA*-over producing cells on several substrates and observed that MpaA was not active on UDP-MurNAc-pentapeptide (Uehara and Park, 2003). We confirmed that the

purified MpaA does not act on the tri-and pentapeptide-containing nucleotide precursor. This confirms that MpaA does not interfere with the peptidoglycan biosynthetic pathway. Uhera and Park also observed that the over production of *mpaA* did not abolish the accumulation of MurNAc (anhydro)-tripeptide that occurs in *ampD* mutant cells, suggesting that MpaA was not active when the tripeptide was bound to a MurNAc residue. We show here that MpaA has very low activity towards MurNAc-tripeptide (specific activity being 0.3 *vs* 12,000 nmoles.min^{-1}.mg^{-1}) and is inactive towards a disaccharide (anhydro)-tripeptide.

It has been reported previously that MpaA is inactive against synthetic tri-, tetra- and pentapeptides containing L-Lys at the third position (Schmidt *et al*., 2001). However here we observe that MpaA is active towards L-Ala-γ-D-Glu-L-Lys, although the activity is much lower (70-fold) than that for the Dap-containing tripeptide. This suggests that MpaA can also hydrolyze the Lys-containing cell wall tripeptide present in most Gram-positive bacteria. L-Lys is present instead of m-Dap in the cell wall peptides of most Gram-positive bacteria (Schleifer and Kandler, 1972). Whether this property of MpaA provides any physiological benefit to *E. coli* needs to be investigated. It can be speculated that *E. coli* may get growth advantages over other Gram-positive bacteria present by catabolizing their cell wall peptides. In the light of our experimental data we were able to propose a new scavenging pathway for peptidoglycan derived peptides in γ-proteobacteria (Figure 7.3).

On the basis of amino acid sequence similarity, MpaA was grouped in the family of metalloproteases which require a divalent cation (zinc) for function. This is consistent with our mass spectrometry data suggesting that MpaA co-purifies with zinc. The requirement of a divalent metal ion for MpaA function was further confirmed when its activity was inhibited in the presence of EDTA.

In the study we also cloned, expressed and purified a homologue of MpaA from *Vibrio harveyi*. Our results showed that *Vh*_MpaA hydrolyzes Mtp but not murein tetrapeptide and thus is a true functional orthologue of *Ec*_MpaA.

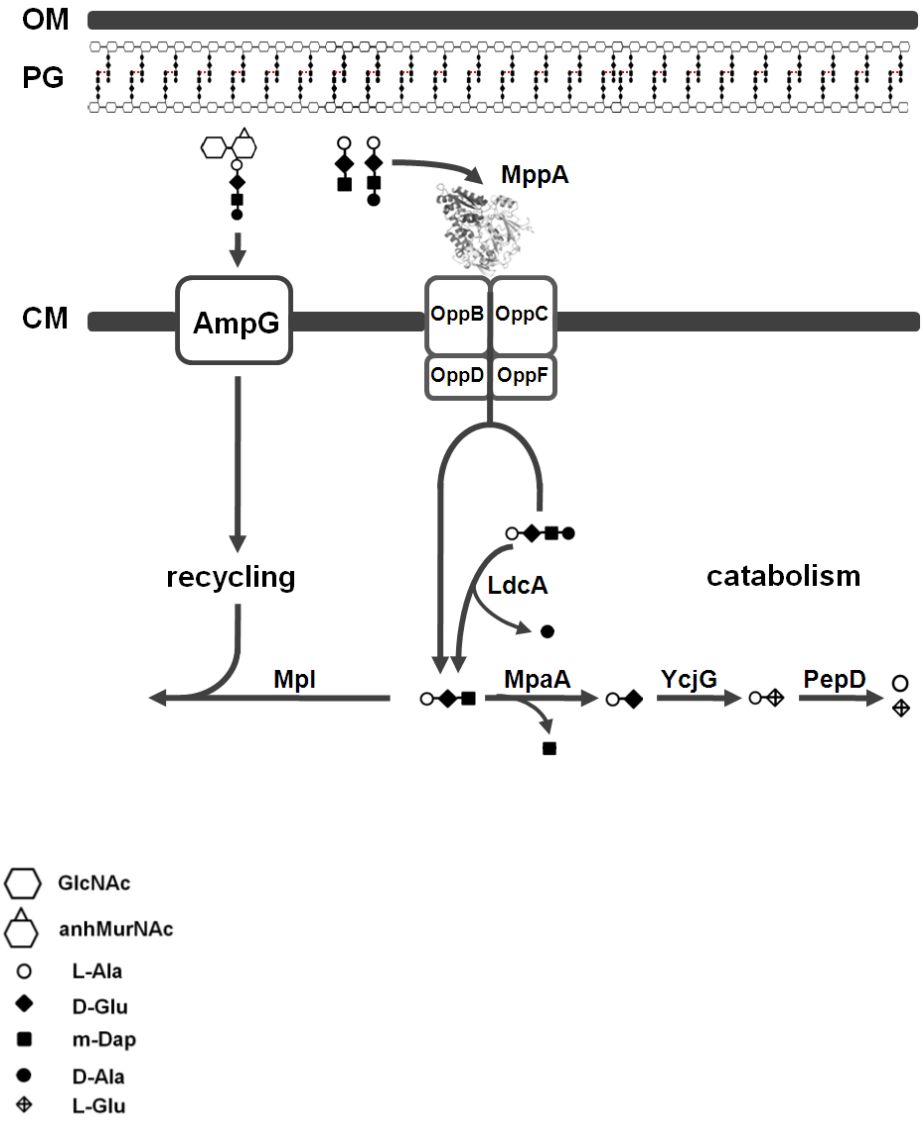


Figure 7.3 Proposed pathway for uptake and catabolism of Mtp and murein tetrapeptide

Murein hydrolases break down peptidoglycan into muropeptide (GlcNAc-anhMurNAc-L-Ala-γ-D-Glu-mDap-D-Ala), Mtp and murein tetrapeptide. Muropeptides are taken up by AmpG permease and enter the recycling pathway. Mtp and murein tetrapeptide are taken up via MppA:OppBCDF. Murein tetrapepide is hydrolyzed by LdcA to Mtp in the cytoplasm. Mtp either enters into the recycling pathway by the action of murein peptide ligase enzyme (Mpl) or hydrolyzed into its component amino acids by the concerted actions of an amidase (MpaA), an epimerase (YcjG) and a peptidase (PepD).

The crystal structure of *Vh*_MpaA revealed that, in spite of the relatively low amino acid similarity, the polypeptide fold of *Vh*_MpaA is the same as those of the eukaryotic carboxypeptidases. This confirms the notion that structure and fold space is more limited compared to sequence space and nature tends to adapt conserved structural modules for different functions. The conserved nature of the active site residues suggests that the mechanism proposed for human CPA is applicable to *Vh*_MpaA and probably to the whole family of Zn carboxypeptidases. The difference in the substrate specificity of MpaA might be ascribed to a loop in *Vh*_MpaA that covers the active site and restricts the length of peptide that can extend into the catalytic centre (Figure 7.4). Identification of residues likely to be involved in binding metal and substrate and those involved in dimerization provide a basis for further experimentation to explore the structure and function of MpaA.

Although MpaA seems to have true orthologues in γ-proteobacteria, there are enzymes in Bacillus and Cyanobacteria that are unrelated to MpaA in amino acid sequence that have equivalent function. An example is YkfC from *Bacillus subtilis*. YkfC contains a C-terminal NIpC/P60 cysteine peptidase domain and is known to cleave tri, tetra and penta peptides containing L-Lys at the third position. The genome context of *ykfC* in *B. subtilis* reveals that it is present next to the dipeptide ABC transporter operon and an L-Ala-D/L-Glu epimerase gene. This gene clustering suggests that a similar catabolic pathway for cell wall peptides may exist in *B. subtilis* and other bacteria albeit one that uses a different set of proteins.

The proposed pathway for uptake and catabolism of cell wall derived peptides (Figure 7.3) can be studied and validated by performing *in vivo* experiments. A simple experiment can be done by growing wild type and mutant strains of a bacterium (*E. coli*) in M9 minimal medium containing cell wall derived peptides as a sole source of carbon and nitrogen, and comparing the growth of wild type and mutant strains. The mutant strains can be constructed by deleting the genes involved in uptake and catabolism of cell wall derived peptides, *e.g. ampG*, *mppA*, *oppD*, *mpaA*, *ldcA*, *ycjG* and *pepD* etc. The mutant strains being unable to uptake

and/or catabolise cell wall derived peptides should exhibit a growth phenotype. The phenotype exhibited by mutant strains can be restored by providing a copy of the deleted gene to the mutants.

In another experiment a double mutant strain of *E. coli* (that have deletions of two genes viz *mpl* and *mpaA*) can be grown in enhanced M9 minimal medium and accumulation of Mtp in the cytosol can be measured using the analytical procedures described previously. (Mengin-Lecreulx *et al.*, 1982). Later on, cleavage of Mtp can be studied by providing a copy of functional *mpaA* to the double mutant strain. Various mutants of *mpaA* can be constructed to produce partially active MpaA or one with low turnover rate. These mutants can be transformed into the double mutant strain in order to study the turnover of Mtp.

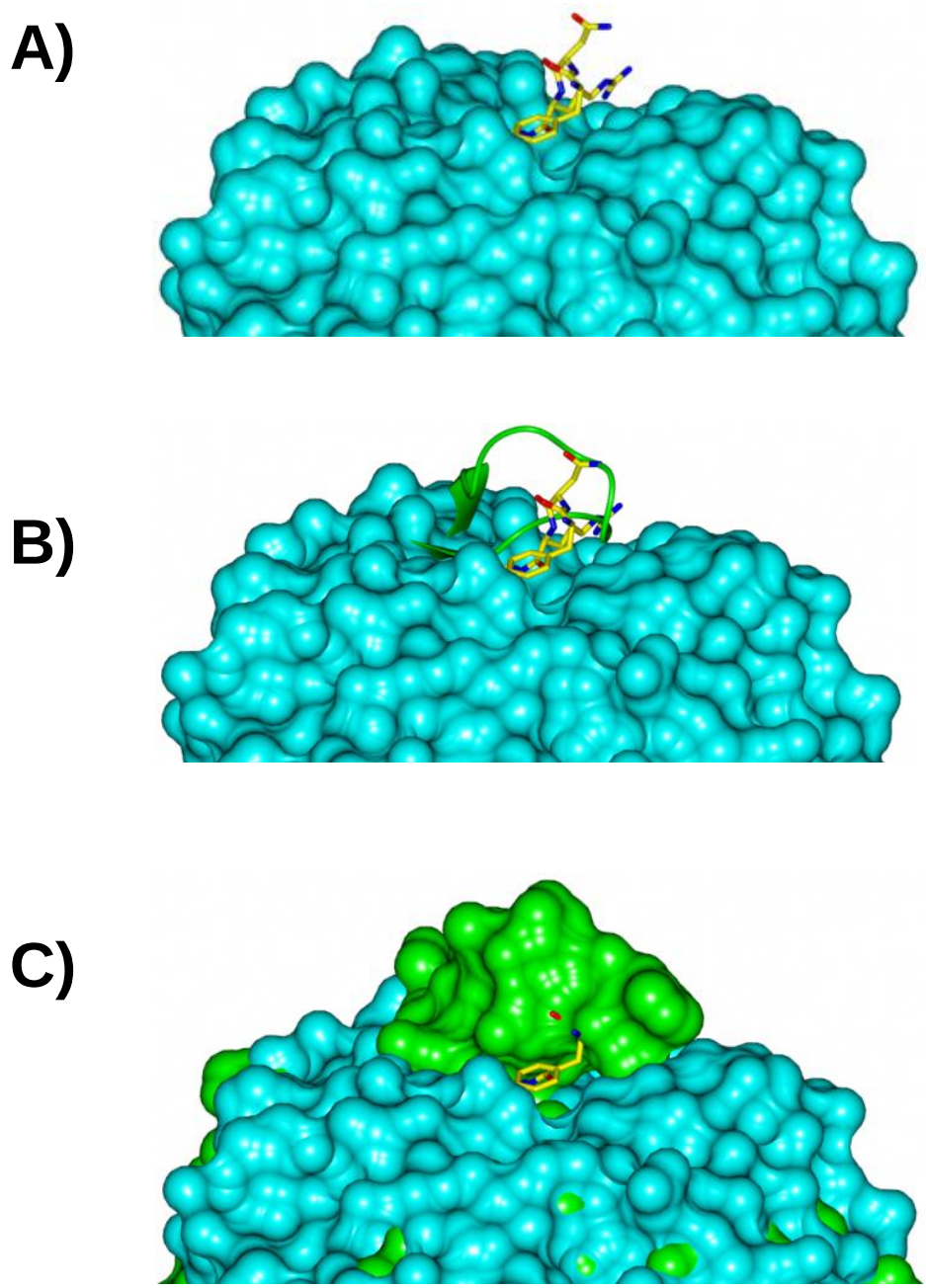


Figure 7.4 Surface representation of human_CPA and *Vh*_MpaA

A) Surface representation of human_CPA (coloured in cyan) in complex with a cleaved hexapeptide. The peptide is present in a groove above active site and represented as stick format. **B)** The structure of human_CPA is superposed on *Vh*_MpaA. Human_CPA is shown as cyan surface and *Vh*_MpaA is shown as cartoon representation (green). A loop of *Vh*_MpaA covers the groove above active site and possibly restricts the length of peptide that enters the catalytic centre for cleavage.**C)** same as **B** except *Vh*_MpaA is represented by surface coloured in green.

Abbreviations

ABC – ATP Binding Cassette

ATP – Adenosine 5′-triphosphate

bp – base pairs

BSA – Bovine Serum Albumin

CBB – Coomassie Brilliant Blue

CD – Circular Dichroism

CV – Column Volume

Da – Dalton

DNA – Deoxyribonucleic Acid

EDTA – Ethylenediaminetetraacetic acid

ES-MS – Electrospray Mass Spectrometry

ESRF – European Synchrotron Radiation Facility

FT – Fourier Transform

HRV – Human rhinovirus

ICR – Ion Cyclotron Resonance

IPTG – Isopropyl β-D-1-thiogalactopyranoside

ITC – Isothermal Titration Calorimetry

KPi – Potassium Phosphate

LB – Luria-Bertani

LIC – Ligation Independent Cloning

MALLS – Multi-Angle Laser Light Scattering

m-Dap – *meso* diaminopimelate

MES – 2-(N-morpholine)-ethane sulphonic acid

Mtp – Murein tripeptide

MWCO – Molecular Weight Cut-Off

NDB – Nucleotide Binding Domain

PAGE – Polyacrylamide Gel Electrophoresis

PCR – Polymerase Chain Reaction

PEG – Polyethyleneglycol

SBP – Substrate-binding Protein

SDS – Sodium Dodecyl Sulphate

SEC-MALLS – Size Exclusion Chromatography in line with MALLS

SeMet – Selenomethionine derivative

TMD – Transmembrane domain

YSBL – York Structural Biology Laboratory

References

ALBERS, S. V., ELFERINK, M. G. L., CHARLEBOIS, R. L., SENSEN, C. W., DRIESSEN, A. J. M. & KONINGS, W. N. 1999. Glucose transport in the extremely thermoacidophilic Sulfolobus solfataricus involves a high-affinity membrane-integrated binding protein. *J. Bacteriol.,* 181**,** 4285-4291.

AMES, G. & NIKAIDO, K. 1978. Identification of a membrane protein as a histidine transport component in Salmonella Typhimurium. *Proc. Natl. Acad. Sci. USA,* 75**,** 5447-5451.

BAYÉS, À., FERNÁNDEZ, D., SOLA, M., MARRERO, A., GARCÍA-PIQUÉ, S., AVILÉS, F. X., VENDRELL, J. & AND F. XAVIER GOMIS-RÜTH 2007. Caught after the Act: A human A-type metallocarboxypeptidase in a product complex with a cleaved hexapeptide. *Biochemistry,* 46**,** 6921-6930.

BERMEJO, G. A., STRUB, M. P., HO, C. & TJANDRA, N. 2010. Ligand-free open-closed transitions of periplasmic binding proteins: the case of glutamine-binding protein. *Biochemistry,* 49**,** 1893-902.

BERNTSSON, R. P., DOEVEN, M. K., FUSETTI, F., DUURKENS, R. H., SENGUPTA, D., MARRINK, S. J., THUNNISSEN, A. M., POOLMAN, B. & SLOTBOOM, D. J. 2009. The structural basis for peptide selection by the transport receptor OppA. *EMBO J.,* 28**,** 1332-1340.

BERNTSSON, R. P., SMITS, S. H., SCHMITT, L., SLOTBOOM, D. J. & POOLMAN, B. 2010. A structural classification of substrate-binding proteins. *FEBS Lett.,* 584**,** 2606-17.

BLOW, D. M. 2002. *Outline of crystallography for biologists,* Oxford University Press, USA.

BORTHS, E. L., LOCHER, K. P., LEE, A. T. & REES, D. C. 2002. The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter. *Proc. Natl. Acad. Sci. USA,* 99**,** 16642-7.

BRZOZOWSKI, A. M. & WALTON, J. 2001. Clear strategy screens for macromolecular crystallization. *J. Appl. Cryst.,* 34**,** 97-101.

CAO, Y., JIN, X., LEVIN, E. J., HUANG, H., ZONG, Y., QUICK, M., WENG, J., PAN, Y., LOVE, J., PUNTA, M., ROST, B., HENDRICKSON, W. A., JAVITCH, J. A., RAJASHANKAR, K. R. & ZHOU, M. 2011. Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature,* 473**,** 50-54.

CASADABAN, M. J. & COHEN, S. N. 1980. Analysis of gene control signals by DNA fusion and cloning in Escherichia coli. *J. Mol. Biol.,* 138**,** 179-207.

CHANDLER, J. R. & DUNNY, G. M. 2004. Enterococcal peptide sex pheromones: synthesis and control of biological activity. *Peptides,* 25**,** 1377-88.

CHEN, J., LU, G., LIN, J., DAVIDSON, A. L. & QUIOCHO, F. A. 2003. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell,* 12**,** 651-661.

CHENG, Q. & PARK, J. T. 2002. Substrate specificity of the AmpG permease required for recycling of cell wall anhydro-muropeptides. *J. Bacteriol.,* 184**,** 6434.

CLARKE, T. E., KU, S. Y., DOUGAN, D. R., VOGEL, H. J. & TARI, L. W. 2000. The structure of the ferric siderophore binding protein FhuD complexed with gallichrome. *Nat. Struct. Biol.,* 7**,** 287-91.

COWTAN, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.,* 62**,** 1002-1011.

DALLE-DONNE, I., ROSSI, R., COLOMBO, G., GIUSTARINI, D. & MILZANI, A. 2009. Protein S-glutathionylation: a regulatory device from bacteria to humans. *Trends Biochem. Sci.,* 34**,** 85-96.

DASSA, E. & HOFNUNG, M. 1985. Sequence of gene malG in E. coli K12: homologies between integral membrane components from binding protein-dependent transport systems. *EMBO J.,* 4**,** 2287-93.

DAVIDSON, A. L., DASSA, E., ORELLE, C. & CHEN, J. 2008. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.,* 72**,** 317-64.

DAVIDSON, A. L., SHUMAN, H. A. & NIKAIDO, H. 1992. Mechanism of maltose transport in Escherichia coli: transmembrane signaling by periplasmic binding proteins. *Proc. Natl. Acad. Sci. USA,* 89**,** 2360-2364.

DAWSON, R. J., HOLLENSTEIN, K. & LOCHER, K. P. 2007. Uptake or extrusion: crystal structures of full ABC transporters suggest a common mechanism. *Mol. Microbiol.,* 65**,** 250-7.

DEAN, D. A., DAVIDSON, A. L. & NIKAIDO, H. 1989. Maltose transport in membrane vesicles of Escherichia coli is linked to ATP hydrolysis. *Proc. Natl. Acad. Sci. USA,* 86**,** 9134-9138.

DETMERS, F. J., LANFERMEIJER, F. C. & POOLMAN, B. 2001. Peptides and ATP binding cassette peptide transporters. *Res. Microbiol.,* 152**,** 245-58.

DIEDERICHS, K., DIEZ, J., GRELLER, G., MÜLLER, C., BREED, J., SCHNELL, C., VONRHEIN, C., BOOS, W. & WELTE, W. 2000. Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis. *EMBO J.,* 19**,** 5951-5961.

DINH, T., PAULSEN, I. T. & SAIER JR, M. H. 1994. A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. *J. Bacteriol.,* 176**,** 3825-3831.

DIPPEL, R. & BOOS, W. 2005. The maltodextrin system of Escherichia coli: metabolism and transport. *J. Bacteriol.,* 187**,** 8322-31.

DOEVEN, M. K., ABELE, R., TAMPE, R. & POOLMAN, B. 2004. The binding specificity of OppA determines the selectivity of the oligopeptide ATP-binding cassette transporter. *J. Biol. Chem.,* 279**,** 32301-7.

DOEVEN, M. K., KOK, J. & POOLMAN, B. 2005. Specificity and selectivity determinants of peptide transport in Lactococcus lactis and other microorganisms. *Mol. Microbiol.,* 57**,** 640-9.

DUNTEN, P. & MOWBRAY, S. L. 1995. Crystal structure of the dipeptide binding protein from Escherichia coli involved in active transport and chemotaxis. *Protein Sci.,* 4**,** 2327-34.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.,* 60**,** 2126-2132.

FAVREAU, P., CHENEVAL, O., MENIN, L., MICHALET, S., GAERTNER, H., PRINCIPAUD, F., THAI, R., MÉNEZ, A., BULET, P. & STÖCKLIN, R. 2007. The venom of the snake genus Atheris contains a new class of peptides with clusters of histidine and glycine residues. *Rapid Commun. Mass Spectrom.,* 21**,** 406-412.

FIRCZUK, M. & BOCHTLER, M. 2007. Folds and activities of peptidoglycan amidases. *FEMS Microbiol. Rev.,* 31**,** 676-91.

FOGG, M. J. & WILKINSON, A. J. 2008. Higher-throughput approaches to crystallization and crystal structure determination. *Biochem. Soc. Trans.,* 36**,** 771-775.

FOLTA-STOGNIEW, E. & WILLIAMS, K. 1999. Determination of molecular masses of proteins in solution: implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J. Biomol. Tech.,* 10**,** 51-63.

FU, D., LIBSON, A., MIERCKE, L. J. W., WEITZMAN, C., NOLLERT, P., KRUCINSKI, J. & STROUD, R. M. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science,* 290**,** 481-486.

FUKAMI-KOBAYASHI, K., TATENO, Y. & NISHIKAWA, K. 1999. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol.,* 286**,** 279-290.

GADSBY, D. C., VERGANI, P. & CSANÁDY, L. 2006. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature,* 440**,** 477.

GLAUNER, B., HOLTJE, J. V. & SCHWARZ, U. 1988. The composition of the murein of Escherichia coli. *J. Biol. Chem.,* 263**,** 10088-95.

GLUSKER, J. P. & TRUEBLOOD, K. N. 1985. *Crystal structure analysis: a primer,* Oxford University Press.

GONIN, S., ARNOUX, P., PIERRU, B., LAVERGNE, J., ALONSO, B., SABATY, M. & PIGNOL, D. 2007. Crystal structures of an extracytoplasmic solute receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for alpha-keto acid binding. *BMC Struct. Biol.,* 7.

GOODELL, E. W. & HIGGINS, C. F. 1987. Uptake of cell wall peptides by Salmonella typhimurium and Escherichia coli. *J. Bacteriol.,* 169**,** 3861-5.

GOODELL, E. W. & SCHWARZ, U. 1985. Release of cell wall peptides into culture medium by exponentially growing Escherichia coli. *J. Bacteriol.,* 162**,** 391-397.

GUYER, C. A., MORGAN, D. G. & STAROS, J. V. 1986. Binding specificity of the periplasmic oligopeptide-binding protein from Escherichia coli. *J. Bacteriol.,* 168**,** 775-779.

HEDDLE, J., SCOTT, D. J., UNZAI, S., PARK, S. Y. & TAME, J. R. 2003. Crystal structures of the liganded and unliganded nickel-binding protein NikA from Escherichia coli. *J. Biol. Chem.,* 278**,** 50322-9.

HERVE, M., BONIFACE, A., GOBEC, S., BLANOT, D. & MENGIN-LECREULX, D. 2007. Biochemical characterization and physiological properties of Escherichia coli UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase. *J. Bacteriol.,* 189**,** 3987-95.

HIGGINS, C., HAAG, P., NIKAIDO, K., ARDESHIR, F., GARCIA, G. & AMES, G. F. L. 1982. Complete nucleotide sequence and identification of

membrane components of the histidine transport operon of S. typhimurium. *Nature,* 298**,** 723-727.
HIGGINS, C. F. & HARDIE, M. M. 1983. Periplasmic protein associated with the oligopeptide permeases of Salmonella typhimurium and Escherichia coli. *J. Bacteriol.,* 155**,** 1434-8.
HIGGINS, C. F. & LINTON, K. J. 2001. Structural biology. The xyz of ABC transporters. *Science,* 293**,** 1782-4.
HILES, I. D. & HIGGINS, C. F. 1986. Peptide uptake by Salmonella typhimurium. The periplasmic oligopeptide-binding protein. *Eur. J. Biochem.,* 158**,** 561-7.
HOLLAND, I. B. 2003. *ABC proteins: from bacteria to man*, Academic Press.
HOLM, L., KÄÄRIÄINEN, S., ROSENSTRÖM, P. & SCHENKEL, A. 2008. Searching protein structure databases with DaliLite v.3. *Bioinformatics,* 24**,** 2780-2781.
HOLTJE, J. V. 1998. Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiol. Mol. Biol. Rev.,* 62**,** 181-203.
HUNKE, S., MOUREZ, M., JÉHANNO, M., DASSA, E. & SCHNEIDER, E. 2000. ATP modulates subunit-subunit interactions in an ATP-binding cassette transporter (MalFGK2) determined by site-directed chemical cross-linking. *J. Biol. Chem.,* 275**,** 15526.
HYDE, S. C., EMSLEY, P., HARTSHORN, M. J., MIMMACK, M. M., GILEADI, U., PEARCE, S. R., GALLAGHER, M. P., GILL, D. R., HUBBARD, R. E. & HIGGINS, C. F. 1990. Structural model of ATP-binding proteing associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature,* 362**,** 362-365.
JACOBS, C., HUANG, L. J., BARTOWSKY, E., NORMARK, S. & PARK, J. T. 1994. Bacterial cell wall recycling provides cytosolic muropeptides as effectors for beta-lactamase induction. *EMBO J.,* 13**,** 4684-94.
JACOBS, C., JORIS, B., JAMIN, M., KLARSOV, K., VAN BEEUMEN, J., MENGIN-LECREULX, D., VAN HEIJENOORT, J., PARK, J. T., NORMARK, S. & FRERE, J. M. 1995. AmpD, essential for both beta-lactamase regulation and cell wall recycling, is a novel cytosolic N-acetylmuramyl-L-alanine amidase. *Mol. Microbiol.,* 15**,** 553-9.
JACSO, T., GROTE, M., DAUS, M. L., SCHMIEDER, P., KELLER, S., SCHNEIDER, E. & REIF, B. 2009. Periplasmic loop P2 of the MalF subunit of the maltose ATP binding cassette transporter is sufficient to bind the maltose binding protein MalE. *Biochemistry,* 48**,** 2216-25.
KEIL, C., MASKOS, K., THAN, M., HOOPES, J. T., HUBER, R., TAN, F., DEDDISH, P. A., ERDOS, E. G., SKIDGEL, R. A. & BODE, W. 2007. Crystal structure of the human carboxypeptidase N (kininase I) catalytic domain. *J. Mol. Biol.,* 366**,** 504-516.
KELLERMANN, O. & SZMELCMAN, S. 1974. Active transport of maltose in Escherichia coli K12. *Eur. J. Biochem.,* 47**,** 139-149.
KELLY, D. J. & THOMAS, G. H. 2001. The tripartite ATP independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol. Rev.,* 25**,** 405-424.

KENNEDY, K. A. & TRAXLER, B. 1999. MalK Forms a Dimer Independent of Its Assembly into the MalFGK2 ATP-binding Cassette Transporter ofEscherichia coli. *J. Biol. Chem.,* 274**,** 6259-64.

KHARE, D., OLDHAM, M. L., ORELLE, C., DAVIDSON, A. L. & CHEN, J. 2009. Alternating access in maltose transporter mediated by rigid-body rotations. *Mole. Cell,* 33**,** 528-536.

KLEPSCH, M. M., KOVERMANN, M., LOW, C., BALBACH, J., PERMENTIER, H. P., FUSETTI, F., DE GIER, J. W., SLOTBOOM, D. & BERNTSSON, R. P. 2011. Escherichia coli peptide binding protein OppA has a preference for positively charged peptides. *J. Mol. Biol.,* 414**,** 75-85.

KÜHLBRANDT, W. 2004. Biology, structure and mechanism of P-type ATPases. *Nature Rev. Mol. Cell Biol.,* 5**,** 282-295.

LANFERMEIJER, F. C., PICON, A., KONINGS, W. N. & POOLMAN, B. 1999. Kinetics and consequences of binding of nona- and dodecapeptides to the oligopeptide binding protein (OppA) of Lactococcus lactis. *Biochemistry,* 38**,** 14440-50.

LAZAZZERA, B. A. 2001. The intracellular function of extracellular signaling peptides. *Peptides,* 22**,** 1519-27.

LEE, Y. H., DEKA, R. K., NORGARD, M. V., RADOLF, J. D. & HASEMANN, C. A. 1999. Treponema pallidum TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat. Struct. Mol. Biol.,* 6**,** 628-633.

LEONARD, B. A., PODBIELSKI, A., HEDBERG, P. J. & DUNNY, G. M. 1996. Enterococcus faecalis pheromone binding protein, PrgZ, recruits a chromosomal oligopeptide permease system to import sex pheromone cCF10 for induction of conjugation. *Proc. Natl. Acad. Sci. USA,* 93**,** 260-4.

LESSARD, I. A., PRATT, S. D., MCCAFFERTY, D. G., BUSSIERE, D. E., HUTCHINS, C., WANNER, B. L., KATZ, L. & WALSH, C. T. 1998. Homologs of the vancomycin resistance D-Ala-D-Ala dipeptidase VanX in Streptomyces toyocaensis, Escherichia coli and Synechocystis: attributes of catalytic efficiency, stereoselectivity and regulation with implications for function. *Chem. Biol.,* 5**,** 489-504.

LETOFFE, S., DELEPELAIRE, P. & WANDERSMAN, C. 2006. The housekeeping dipeptide permease is the Escherichia coli heme transporter and functions with two optional peptide binding proteins. *Proc. Natl. Acad. Sci. USA,* 103**,** 12891-6.

LEVDIKOV, V. M., BLAGOVA, E. V., BRANNIGAN, J. A., WRIGHT, L., VAGIN, A. A. & WILKINSON, A. J. 2005. The structure of the oligopeptide-binding protein, AppA, from Bacillus subtilis in complex with a nonapeptide. *J. Mol. Biol.,* 345**,** 879-92.

LOCHER, K. P., LEE, A. T. & REES, D. C. 2002. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science,* 296**,** 1091.

LU, G., WESTBROOKS, J. M., DAVIDSON, A. L. & CHEN, J. 2005. ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci. USA,* 102**,** 17969.

MAGNUSSON, U., SALOPEK-SONDI, B., LUCK, L. A. & MOWBRAY, S. L. 2004. X-ray structures of the leucine-binding protein illustrate

conformational changes and the basis of ligand specificity. *J. Biol. Chem.*, 279**,** 8747-52.

MAO, B., PEAR, M., MCCAMMON, J. & QUIOCHO, F. 1982. Hinge-bending in L-arabinose-binding protein. The" Venus's-flytrap" model. *J. Biol. Chem.,* 257**,** 1131-1133.

MAQBOOL, A., LEVDIKOV, V. M., BLAGOVA, E. V., HERVE, M., HORLER, R. S. P., WILKINSON, A. J. & THOMAS, G. H. 2011. Compensating stereochemical changes allow murein tripeptide to be accommodated in a conventional peptide-binding protein. *J. Biol. Chem.*, 286**,** 31512-31521.

MENGIN-LECREULX, D., FLOURET, B. & VAN HEIJENOORT, J. 1982. Cytoplasmic steps of peptidoglycan synthesis in Escherichia coli. *J. Bacteriol.,* 151**,** 1109-1117.

MISRA, R. V., HORLER, R. S., REINDL, W., GORYANIN, II & THOMAS, G. H. 2005. EchoBASE: an integrated post-genomic database for Escherichia coli. *Nucleic Acids Res.,* 33**,** D329-33.

MORBACH, S., TEBBE, S. & SCHNEIDER, E. 1993. The ATP-binding cassette (ABC) transporter for maltose/maltodextrins of Salmonella typhimurium. Characterization of the ATPase activity associated with the purified MalK subunit. *J. Biol. Chem.,* 268**,** 18617-18621.

MOUREZ, M., HOFNUNG, M. & DASSA, E. 1997. Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.,* 16**,** 3066-3077.

MOUSSATOVA, A., KANDT, C., O'MARA, M. L. & TIELEMAN, D. P. 2008. ATP-binding cassette transporters in Escherichia coli. *Biochim. Biophys. Acta,* 1778**,** 1757-1771.

MÜLLER, A., SEVERI, E., MULLIGAN, C., WATTS, A. G., KELLY, D. J., WILSON, K. S., WILKINSON, A. J. & THOMAS, G. H. 2006. Conservation of structure and mechanism in primary and secondary transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus influenzae. *J. Biol. Chem.,* 281**,** 22212-22.

MULLIGAN, C., GEERTSMA, E. R., SEVERI, E., KELLY, D. J., POOLMAN, B. & THOMAS, G. H. 2009. The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc. Natl. Acad. Sci. USA,* 106**,** 1778-83.

MURSHUDOV, G. N., VAGIN, A. A. & DODSON, E. J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.,* 53**,** 240-255.

NAVARRE, W. W. & SCHNEEWIND, O. 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.,* 63**,** 174-229.

NEIDHARDT, F. C., BLOCH, P. L. & SMITH, D. F. 1974. Culture medium for Enterobacteria. *J. Bacteriol.,* 119**,** 736-747.

NEIDITCH, M. B., FEDERLE, M. J., POMPEANI, A. J., KELLY, R. C., SWEM, D. L., JEFFREY, P. D., BASSLER, B. L. & HUGHSON, F. M. 2006. Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell,* 126**,** 1095-108.

NEWMAN, J., EGAN, D., WALTER, T. S., MEGED, R., BERRY, I., BEN JELLOUL, M., SUSSMAN, J. L., STUART, D. I. & PERRAKIS, A. 2005. Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr. D Biol. Crystallogr.,* 61**,** 1426-1431.

NIKAIDO, H. 2003. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.,* 67**,** 593-656.

OLDHAM, M. L. & CHEN, J. 2011. Crystal Structure of the Maltose Transporter in a Pretranslocation Intermediate State. *Science,* 332**,** 1202-1205.

OLDHAM, M. L., KHARE, D., QUIOCHO, F. A., DAVIDSON, A. L. & CHEN, J. 2007. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature,* 450**,** 515-521.

OTWINOWSKI, Z. & MINOR, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.,* 276**,** 307-326.

OVERBEEK, R., BEGLEY, T., BUTLER, R. M., CHOUDHURI, J. V., CHUANG, H. Y., COHOON, M., DE CRÉCY-LAGARD, V., DIAZ, N., DISZ, T. & EDWARDS, R. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.,* 33**,** 5691-5702.

PARK, J. 1995. Why does Escherichia coli recycle its cell wall peptides? *Mol. Microbiol.,* 17**,** 421-426.

PARK, J. T., RAYCHAUDHURI, D., LI, H., NORMARK, S. & MENGIN-LECREULX, D. 1998. MppA, a periplasmic binding protein essential for import of the bacterial cell wall peptide L-alanyl-gamma-D-glutamyl-meso-diaminopimelate. *J. Bacteriol.,* 180**,** 1215-23.

PARK, J. T. & UEHARA, T. 2008. How bacteria consume their own exoskeletons (turnover and recycling of cell wall peptidoglycan). *Microbiol. Mol. Biol. Rev.,* 72**,** 211-27, table of contents.

PATIN, D., BONIFACE, A., KOVAČ, A., HERVÉ, M., DEMENTIN, S., BARRETEAU, H., MENGIN-LECREULX, D. & BLANOT, D. 2010. Purification and biochemical characterization of Mur ligases from Staphylococcus aureus. *Biochimie,* 92**,** 1793-1800.

PAYNE, J. 1971. The requirement for the protonated α-amino group for the transport of peptides in Escherichia coli. *Biochem. J.,* 123**,** 245-253.

PINKETT, H., LEE, A., LUM, P., LOCHER, K. & REES, D. 2007. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science,* 315**,** 373-377.

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

QUIOCHO, F. A., PHILLIPS JR, G. N., PARSONS, R. G. & HOGG, R. W. 1974. Crystallographic data of an L-arabinose-binding protein from Escherichia coli. *J. Mol. Biol.,* 86**,** 491-493.

REES, D. C., JOHNSON, E. & LEWINSON, O. 2009. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.,* 10**,** 218-227.

REN, Q., CHEN, K. & PAULSEN, I. T. 2006. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.,* 35**,** D274-D279.

RHODES, G. 2000. *Crystallography made crystal clear: a guide for users of macromolecular models*, Academic Press.

RODIONOV, D. A., HEBBELN, P., EUDES, A., TER BEEK, J., RODIONOVA, I. A., ERKENS, G. B., SLOTBOOM, D. J., GELFAND, M. S., OSTERMAN, A. L., HANSON, A. D. & EITINGER, T. 2009. A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.*, 191**,** 42-51.

RODIONOV, D. A., HEBBELN, P., GELFAND, M. S. & EITINGER, T. 2006. Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J. Bacteriol.*, 188**,** 317-27.

RUDD, K. E. 2000. EcoGene: a genome sequence database for Escherichia coli K-12. *Nucleic Acids Res.*, 28**,** 60-4.

SAIER JR, M., HVORUP, R. & BARABOTE, R. 2005. Evolution of the bacterial phosphotransferase system: from carriers and enzymes to group translocators. *Biochem. Soc. Trans.*, 33**,** 220-224.

SAIER JR, M. H. 1994. Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiol. Mol. Biol. Rev.*, 58**,** 71-93.

SAIER JR, M. H. 2000a. Families of transmembrane sugar transport proteins. *Mol. Microbiol.*, 35**,** 699-710.

SAIER JR, M. H. 2000b. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.*, 64.

SAIER, M. H., JR., YEN, M. R., NOTO, K., TAMANG, D. G. & ELKAN, C. 2009. The Transporter Classification Database: recent advances. *Nucleic Acids Res.*, 37**,** D274-8.

SATYSHUR, K. A., WORZALLA, G. A., MEYER, L. S., HEINIGER, E. K., AUKEMA, K. G., MISIC, A. M. & FOREST, K. T. 2007. Crystal structures of the pilus retraction motor PilT suggest large domain movements and subunit cooperation drive motility. *Structure*, 15**,** 363-376.

SAURIN, W. & DASSA, E. 1994. Sequence relationships between integral inner membrane proteins of binding protein-dependent transport systems: evolution by recurrent gene duplications. *Protein Sci.*, 3**,** 325-344.

SCHLEIFER, K. H. & KANDLER, O. 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.*, 36**,** 407-77.

SCHMIDT, D. M., HUBBARD, B. K. & GERLT, J. A. 2001. Evolution of enzymatic activities in the enolase superfamily: functional assignment of unknown proteins in Bacillus subtilis and Escherichia coli as L-Ala-D/L-Glu epimerases. *Biochemistry*, 40**,** 15707-15.

SEVERI, E., HOOD, D. W. & THOMAS, G. H. 2007. Sialic acid utilization by bacterial pathogens. *Microbiology*, 153**,** 2817-22.

SEVERI, E., RANDLE, G., KIVLIN, P., WHITFIELD, K., YOUNG, R., MOXON, R., KELLY, D., HOOD, D. & THOMAS, G. H. 2005. Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol. Microbiol.*, 58**,** 1173-1185.

SHARFF, A. J., RODSETH, L. E., SPURLINO, J. C. & QUIOCHO, F. A. 1992. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry,* 31**,** 10657-10663.

SHELDRICK, G. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.,* 66**,** 479-485.

SHI, R., PROTEAU, A., WAGNER, J., CUI, Q., PURISIMA, E. O., MATTE, A. & CYGLER, M. 2009. Trapping open and closed forms of FitE: a group III periplasmic binding protein. *Proteins,* 75**,** 598-609.

SILHAVY, T. J., SZMELCMAN, S., BOOS, W. & SCHWARTZ, M. 1975. On the significance of the retention of ligand by protein. *Proc. Natl. Acad. Sci. USA,* 72**,** 2120-4.

SLEIGH, S. H., SEAVERS, P. R., WILKINSON, A. J., LADBURY, J. E. & TAME, J. R. 1999. Crystallographic and calorimetric analysis of peptide binding to OppA protein. *J. Mol. Biol.,* 291**,** 393-415.

SMITH, M. W., TYREMAN, D. R., PAYNE, G. M., MARSHALL, N. J. & PAYNE, J. W. 1999. Substrate specificity of the periplasmic dipeptide-binding protein from Escherichia coli: experimental basis for the design of peptide prodrugs. *Microbiology,* 145**,** 2891-2901.

SPURLINO, J. C., LU, G. Y. & QUIOCHO, F. 1991. The 2.3-A resolution structure of the maltose-or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.,* 266**,** 5202.

STORY, R. M. & STEITZ, T. A. 1992. Structure of the recA protein–ADP complex. *Nature,* 355**,** 374-376.

SUTCLIFFE, I. C. & RUSSELL, R. R. 1995. Lipoproteins of gram-positive bacteria. *J. Bacteriol.,* 177**,** 1123-8.

SUZUKI, H., KOYANAGI, T., IZUKA, S., ONISHI, A. & KUMAGAI, H. 2005. The yliA, -B, -C, and -D genes of Escherichia coli K-12 encode a novel glutathione importer with an ATP-binding cassette. *J. Bacteriol.,* 187**,** 5861-7.

SWEET, G., GANDOR, C., VOEGELE, R., WITTEKINDT, N., BEUERLE, J., TRUNIGER, V., LIN, E. & BOOS, W. 1990. Glycerol facilitator of Escherichia coli: cloning of glpF and identification of the glpF product. *J. Bacteriol.,* 172**,** 424-430.

TAM, R. & SAIER, M. H., JR. 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.,* 57**,** 320-46.

TAME, J. R., DODSON, E. J., MURSHUDOV, G., HIGGINS, C. F. & WILKINSON, A. J. 1995. The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure,* 3**,** 1395-406.

TAME, J. R., MURSHUDOV, G. N., DODSON, E. J., NEIL, T. K., DODSON, G. G., HIGGINS, C. F. & WILKINSON, A. J. 1994. The structural basis of sequence-independent peptide binding by OppA protein. *Science,* 264**,** 1578-81.

TAYLOR, G. 2003. The phase problem. *Acta Crystallogr. D Biol. Crystallogr.,* 59**,** 1881-1890.
THOMAS, G. H. 2010. Homes for the orphans: utilization of multiple substrate binding proteins by ABC transporters. *Mol. Microbiol.,* 75**,** 6-9.
THOMAS, G. H., SOUTHWORTH, T., LEÓN-KEMPIS, M. R., LEECH, A. & KELLY, D. J. 2006. Novel ligands for the extracellular solute receptors of two bacterial TRAP transporters. *Microbiology,* 152**,** 187-198.
TSAI, Y., SAWAYA, M. R., CANNON, G. C., CAI, F., WILLIAMS, E. B., HEINHORST, S., KERFELD, C. A. & YEATES, T. O. 2007. Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *PLoS Biol.,* 5**,** 1345-1354.
UEHARA, T. & PARK, J. T. 2003. Identification of MpaA, an amidase in Escherichia coli that hydrolyzes the gamma-D-glutamyl-meso-diaminopimelate bond in murein peptides. *J. Bacteriol.,* 185**,** 679-82.
UEHARA, T. & PARK, J. T. 2007. An anhydro-N-acetylmuramyl-L-alanine amidase with broad specificity tethered to the outer membrane of Escherichia coli. *J. Bacteriol.,* 189**,** 5634-41.
VAGIN, A. & TEPLYAKOV, A. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.,* 56**,** 1622-1624.
VAN DE WATER, F. M., BOLEIJ, J. M., PETERS, J. G. P., RUSSEL, F. G. M. & MASEREEUW, R. 2007. Characterization of P-glycoprotein and multidrug resistance proteins in rat kidney and intestinal cell lines. *Eur. J. Pharm. Sci.,* 30**,** 36-44.
VAN HEIJENOORT, J. 2001. Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat. Prod. Rep.,* 18**,** 503-19.
VAN HOLDE, K. E., JOHNSON, W. C. & HO, P. S. 1998. *Principles of Physical Biochemistry,* Prentice Hall Upper Saddle River, NJ.
VERGAUWEN, B., ELEGHEERT, J., DANSERCOER, A., DEVREESE, B. & SAVVIDES, S. N. Glutathione import in Haemophilus influenzae Rd is primed by the periplasmic heme-binding protein HbpA. *Proc. Natl. Acad. Sci. USA,* 107**,** 13270-5.
VOLLMER, W., BLANOT, D. & DE PEDRO, M. A. 2008. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.,* 32**,** 149-67.
WALKER, J. E., SARASTE, M., RUNSWICK, M. J. & GAY, N. J. 1982. Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.,* 1**,** 945-951.
WILKINS, M., BLAUROCK, A. & ENGELMAN, D. 1971. Bilayer structure in membranes. *Nat. New Biol.,* 230**,** 72-76.
WILKINSON, A. J. 1996. Accommodating structurally diverse peptides in proteins. *Chem. Biol.,* 3**,** 519-524.
WINTER, G. 2010. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.,* 43**,** 186-190.
WONG, P., KASHKET, E. & WILSON, T. 1970. Energy coupling in the lactose transport system of Escherichia coli. *Proc. Natl. Acad. Sci. USA,* 65**,** 63.
XU, Q., ABDUBEK, P., ASTAKHOVA, T., AXELROD, H. L., BAKOLITSA, C., CAI, X., CARLTON, D., CHEN, C., CHIU, H. J., CHIU, M.,

CLAYTON, T., DAS, D., DELLER, M. C., DUAN, L., ELLROTT, K., FARR, C. L., FEUERHELM, J., GRANT, J. C., GRZECHNIK, A., HAN, G. W., JAROSZEWSKI, L., JIN, K. K., KLOCK, H. E., KNUTH, M. W., KOZBIAL, P., KRISHNA, S. S., KUMAR, A., LAM, W. W., MARCIANO, D., MILLER, M. D., MORSE, A. T., NIGOGHOSSIAN, E., NOPAKUN, A., OKACH, L., PUCKETT, C., REYES, R., TIEN, H. J., TRAME, C. B., VAN DEN BEDEM, H., WEEKES, D., WOOTEN, T., YEH, A., HODGSON, K. O., WOOLEY, J., ELSLIGER, M. A., DEACON, A. M., GODZIK, A., LESLEY, S. A. & WILSON, I. A. 2010. Structure of the gamma-D-glutamyl-L-diamino acid endopeptidase YkfC from Bacillus cereus in complex with L-Ala-gamma-D-Glu: insights into substrate recognition by NlpC/P60 cysteine peptidases. *Acta Crystallogr. F Struct. Biol. Cryst. Commun.*, 66**,** 1354-64.

YU, R. C., HANSON, P. I., JAHN, R. & BRÜNGER, A. T. 1998. Structure of the ATP-dependent oligomerization domain of N-ethylmaleimide sensitive factor complexed with ATP. *Nat. Struct. Mol. Biol.,* 5**,** 803-811.

ZHANG, P., WANG, J. & SHI, Y. 2010. Structure and mechanism of the S component of a bacterial ECF transporter. *Nature,* 468**,** 717-20.

ZHENG, L., BAUMANN, U. & REYMOND, J. L. 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, 32**,** e115.