Structural and functional studies of pseudomurein peptide ligases in methanogenic archaea

A dissertation presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Biochemistry

Massey University, Manawatu,

New Zealand.

Bishwa Prakash Subedi

2018

Abstract

Prokaryotes are classified as *Archaea* and *Bacteria* in the tree of life and have several distinguishing characteristics, among which the cell wall is one of the most essential and early evolving. Cell walls serve a number of essential functions including protection against osmotic stress, maintenance of cell shape, reduction of lateral gene transfer, and protection from viruses. The cell walls in *Bacteria* are predominantly comprised of peptidoglycan (murein) whereas *Archaea* contain a wide range of cell wall types, none of them being murein. However, methanogens of the order *Methanobacteriales* and *Methanopyrales* contain pseudomurein that shares an overall architectural structure similar to that of murein with a glycan backbone that is cross-linked by a peptide. Understanding the enzymatic steps for pseudomurein pentapeptide biosynthesis and structural information of these enzymes, could be key to resolving the evolutionary history of cell wall synthesis and was the focus of this project.

Analysis of the sequences and gene clusters of the murein peptide ligase genes suggested that analogous putative pseudomurein peptide ligases exist in methanogens. Moreover, the structures of two pseudomurein peptide ligases, pMurE and pMurC, the first of any archaeal peptide ligase, have been determined and their structural homology with bacterial murein ligase MurE and MurC, respectively, was analysed. The structures of pMurE from *Methanothermus fervidus* DSM 2088 (Mfer762) and *Methanothermobacter thermautotrophicus* ΔH DSM 1053 (Mth734), and pMurC (Mfer336), also from *M. fervidus* were determined to a resolutions of 1.7, 2.7, and 2.5 Å, respectively. The pseudomurein peptide ligase structures share a similar overall three domain arrangement and one shows a rigid-body rotation of the C-terminal domain as observed for murein peptide ligases. The ATP-binding sites in both pMurE and pMurC have been identified based on structure

homology. The N^{α}-UDP-Glu$^{\gamma}$-Ala-binding site for pMurE peptide ligase has been proposed based on the UDP-binding position suggesting a similar peptide ligation mechanism as that of MurE peptide ligase. The study thereby suggests a proposed functional role of the pseudomurein peptide ligases and proposes an evolutionary pathway for both murein and pseudomurein peptide ligases from common ancestral genes.

Acknowledgements

First of all, I would like to thank my supervisors Assoc. Prof. Andrew Sutherland-Smith and Dr. Ron Ronimus for giving me this opportunity to work on this project and excel my knowledge and experience in the field of molecular biology and biochemistry. The valuable time by both of you to supervise this project, insightful discussions and helpful advices at different stages have brought this project to this point.

I would like to thank my co-supervisor Assoc. Prof. Gill Norris for providing much-appreciated encouragement and guidance. Also, special thanks to Dr. Vince Carbone, who helped me at different steps of protein structure determination experiments, and Dr. Linley Schofield for the practical guidance and advice during biochemical activity and molecular mass determination.

A special thank you for support and advice in practical matter goes to Dr. Santosh Panjikar for his helpful suggestions in solving structures, Mr. Maximilian Wolf, Ms. Denise Schafer and Mr. Youri Van Nuland who had previously worked with pseudomurein peptide ligases and prepared constructs that were used in this project.

My gratitude to the Marsden Fund for providing my scholarship and travel fund to attend international conference and Massey University Conference Presentation Grants and NZSG travel grant. Also my gratitude to Australian synchrotron for allowing the CAP and Rapid access for data collection.

Finally, special thanks to my family members for your love, support and patience, especially my daughter Ebisha, parents Ishwor P. Subedi and Bhagawati Subedi, siblings Amrit/Kopila and Srijana/Nabin, nephew and niece Yokisha and Yokesh. My wife Ishara, you equally deserve this honour as without you, your support and motivation, I would have never made this happen, thanks a lot.

Dedication

To the loveliest couple who have always been there for me at every ups and downs of my life, my grandparents, Nityananda and Laxmi Subedi. You are the most amazing and supportive grandparents anyone can imagine. All this journey from my first step with support to your hands, first word written with my hand on your hand, first schooling with a ride on your back, I have come to a point where I achieved my dream. This accomplishment is dedicated to you for your continuous cheer ups, encouragement and blessings despite the physical distance and to express my gratitude for the teaching and guidance you initiated.

गुरुर्ब्रह्मा गुरुर्विष्णु र्गुरुर्देवो महेश्वरः

गुरु साक्षात परब्रह्मा तस्मै श्रीगुरवे नमः

Table of Contents

List of Tables

List of Figures

Abbreviations

°C	Temperature in centigrade
Å	Ångström (10^{-10} m)
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
AMPD	2-Amino-2-methyl-1, 3-propanediol
AMP-PNP	Adenylyl-imidodiphosphate
ANP	Phosphoaminophosphonic acid-adenylate ester
ATP	Adenosine tri-phosphate
Bis-tris	Bis (2-hydroxyethyl) aminotris (hydroxymethyl) methane.
BLAST	Basic Local Alignment Search Tool
BME	β-mercaptoethanol
BSA	Bovine Serum Albumin
GalNAc	*N*-acetylgalactosamine
GlcA	Glucuronic acid residue
GlcNAc	*N*-acetylglucosamine
GlyGly	Glycylglycine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Temperature in Kelvin
KCl	Potassium chloride
kDa	Kilodalton
MOPS	3-(N-Morpholino) propane sulfonic acid
MCWO	Molecular weight cut off
mM	Millimolar (10^{-3} Molar concentration)

Mur	Murein ligase
MurNAc	*N*-acetylmuramic acid residues
NAcAltNA	*N*-acetylaltrosaminuronic acid
NAcTalNA	*N*-acetyl-L-talosaminuronic acid
NDSB	Non detergent sulfobetaines
PAGE	Polyacrylamide gel electrophoresis
pH	Measurement of acidity or basicity of an aqueous solution
P_i	Inorganic phosphate
PSI-BLAST	Position-specific iterated BLAST
PMSF	Phenylmethylsulfonyl fluoride
pMur	Pseudomurein peptide ligase
RMSD	Root-mean-square deviation
POP	Pyrophosphate representation in the protein structure
SDS	Sodium dodecyl sulfate
TCEP	Tris (2-carboxyethyl) phosphine hydrochloride
TEA	Triethanolamine
UAG	UDP-*N*-acetylmuramoyl-L-Ala-D-Glu
UDP	Uridine 5′-diphosphate
UMA	UDP-*N*-acetylmuramoyl-L-Ala
v/v	Volume per volume

Chapter 1. Introduction

1.1 Archaea

Carl Woese first discovered the third domain of life, *Archaea*, in 1977 which was later accepted universally after 1990. The distinction of the three domains of life *Bacteria*, *Archaea* and *Eukarya* is mostly based on small ribosomal sub-unit rRNA (16S rRNA) sequence analyses and is used extensively to construct phylogenetic trees that depict the relationships between each domain of life (Woese & Fox, 1977). Woese and colleagues analysed a wide range of prokaryotes based on the 16S rRNA and observed that there were distinct groups of organisms which survived in extreme conditions and some which produced methane. These particular groups of organisms were grouped together and proposed as a separate domain, initially named Archaebacteria that was later changed into "*Archaea*". Classification of this domain started with two kingdoms, *Euryarchaeota* and *Crenarchaeota*; however, as of 2014 *Archaea* have been grouped into five superphyla namely: *Euryarchaeota*, *Nanoarchaeota*, *Crenarchaeota*, *Korachaeota* and *Thaumarchaeota* (Woese *et al.*, 1990; Guy & Ettema, 2011; Cohen, 2014).

After the classification of *Archaea* as a separate domain of life several other features unique from *Bacteria* have been revealed and studied in detail. For example, the cell wall is one of the distinguishing features that has been extensively studied in recent years. Distinct differences in the cell wall composition and biosynthetic mechanisms have been studied between *Archaea* and *Bacteria*, which has helped to establish their existence as separate domains in the phylogenetic tree of life (Hartmann & König, 1990; Albers & Meyer, 2011). Initially, archaeal growth was studied in extreme environments but it has been discovered that they are found widely in other environmental niches similar to ones suitable for bacterial

growth (Cavicchioli, 2011). This wide range of variation in the habitats of archaea has been linked with the diverse range of cell wall envelopes present in different archaea (König, 1988; Albers & Meyer, 2011). In addition, the cell membranes of archaea differ in their lipid constituents compared to those of *Bacteria* and *Eukarya*. The lipids found in archaeal cell membrane have isoprenoid side-chains and are ether-linked to a glycerol-1-phosphate moiety, which is unique to and universally distributed in *Archaea* (Figure 1.1) (Jain *et al.*, 2014; Carbone *et al.*, 2015).

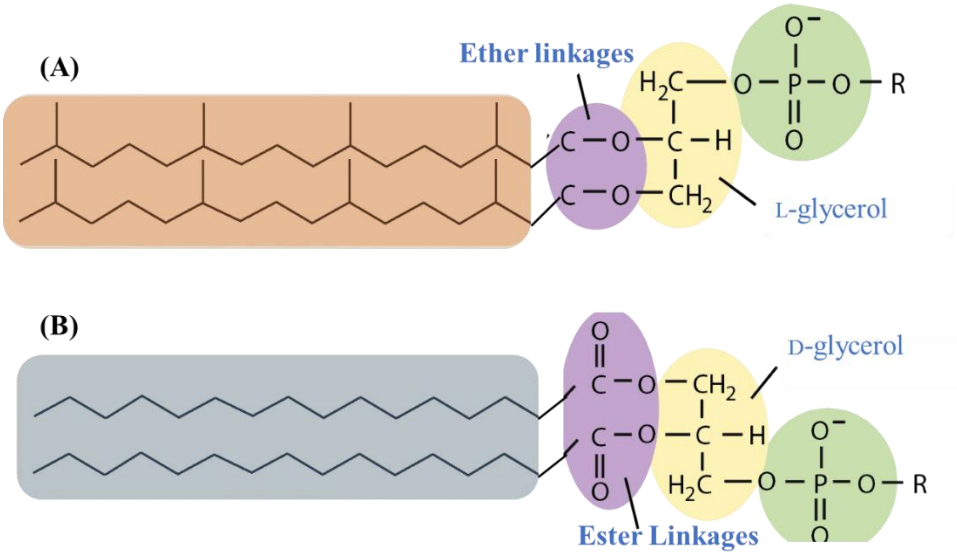


Figure 1.1 Comparison of phospholipids of *Archaea* (A) and *Bacteria* (B) (Caforio & Driessen, 2017).

1.2 Cell walls in Archaea

The archaeal cell wall is represented by a wide range of chemical compositions and primary structures of polymers. The components of archaeal cell walls range from proteinaceous or glycoproteinaceous layers, also known as the S-layer, to methanochondroitins, proteinaceous sheaths, polysaccharides, glutaminylglycans, pseudomurein and halomucin. Furthermore, some *Archaea* possess additional cell wall structure types as shown in the

phylogenetic tree obtained from the alignment of the full length 16S rRNA sequences and most of the cell wall components in certain archaeal types as shown in Figure 1.2 are not well understood (Claus & König, 2010; Albers & Meyer, 2011).

Among various cell wall constituents, the proteinaceous or glycoprotein surface layer (S-layer) cell wall structures are the most abundantly expressed protein type. S-layers are predominant in *Archaea* representing one of the unique features of archaeal cell wall and are proposed to have a role in cell shape and osmo-protection (Sleytr *et al.*, 2014). The S-layer is considered to be the earliest evolved cell wall due to its wide distribution and simple composition and is present solely on its own or in combination with other polymers such as polysaccharides, pseudomurein and methanochondroitin (Albers & Meyer, 2011). The archaeal S-layers are formed by only one type of protein or glycoprotein (molecular size 40-200 kDa) and show a para-crystalline structure and are linked to the cytoplasmic membrane (Albers & Meyer, 2011).

Methanochondroitin is another archaeal cell wall component consisting of two *N*-acetylgalactosamine residues (GalNAc) and one glucuronic acid residue (GlcA) in a repeating chain of UDP-GalNAc-GalNAc-GlcA and is synthesised from the nucleotide-activated UDP-GalNAc and UDP-GlcA precursors. This cell wall type has been observed in aggregating *Methanosarcina* species along with an S-layer (Figure 1.2) whereas the S-layer is observed independently in individual cells (Kreisl & Kandler, 1986; Sowers *et al.*, 1993; Kandler & König, 1998; Klingl, 2014).

Proteinaceous sheaths are found in *Methanospirillum hungatei* and *Methanosaeta concilii* and consist of amino acids and neutral sugars having a structure reminiscent of paracrystalline S-layers (Zeikus & Bowen, 1975). Similarly, highly sulfated

heterosaccharides, a mixture of neutral and amino sugars, uronic acids, gulosaminuronic acid, acetate, glycine and sulfate form a rigid cell wall of the extremely halophilic *Halococcus morrhuae* (Schleifer *et al.*, 1982). Another studied archaeal cell wall component is glutaminylglycan found in *Natronococcus occultus* that consists of a novel glycoconjugate (Kandler & König, 1998). In addition, the extra surface protein present in *Haloquadearum walsbyi* (on top of the S-layer) is assumed to have a role in adaptation towards high salt concentration (2.0 M $MgCl_2$) and is proposed to be the largest (9,159 amino acids long) archaeal protein known to date (Kandler & König, 1998; Albers & Meyer, 2011; Klingl, 2014).

Finally, pseudomurein, which is specific to the orders Methanobacteriales and Methanopyrales, is the cell wall polymer that has roles in cell shape and defence mechanism against viruses and osmotic stress and is present independently or in combination with the S-layer protein as in *Methanothermus fervidus* (Figure 1.2) (Claus & König, 2010). Pseudomurein is ~ 15-20 nm thick and is composed of chemical moieties similar to bacterial murein (peptidoglycan) (Claus & König, 2010). Even though structural similarity is observed between the murein and pseudomurein, certain differences exist between them and these will be discussed in Section 1.4.1.

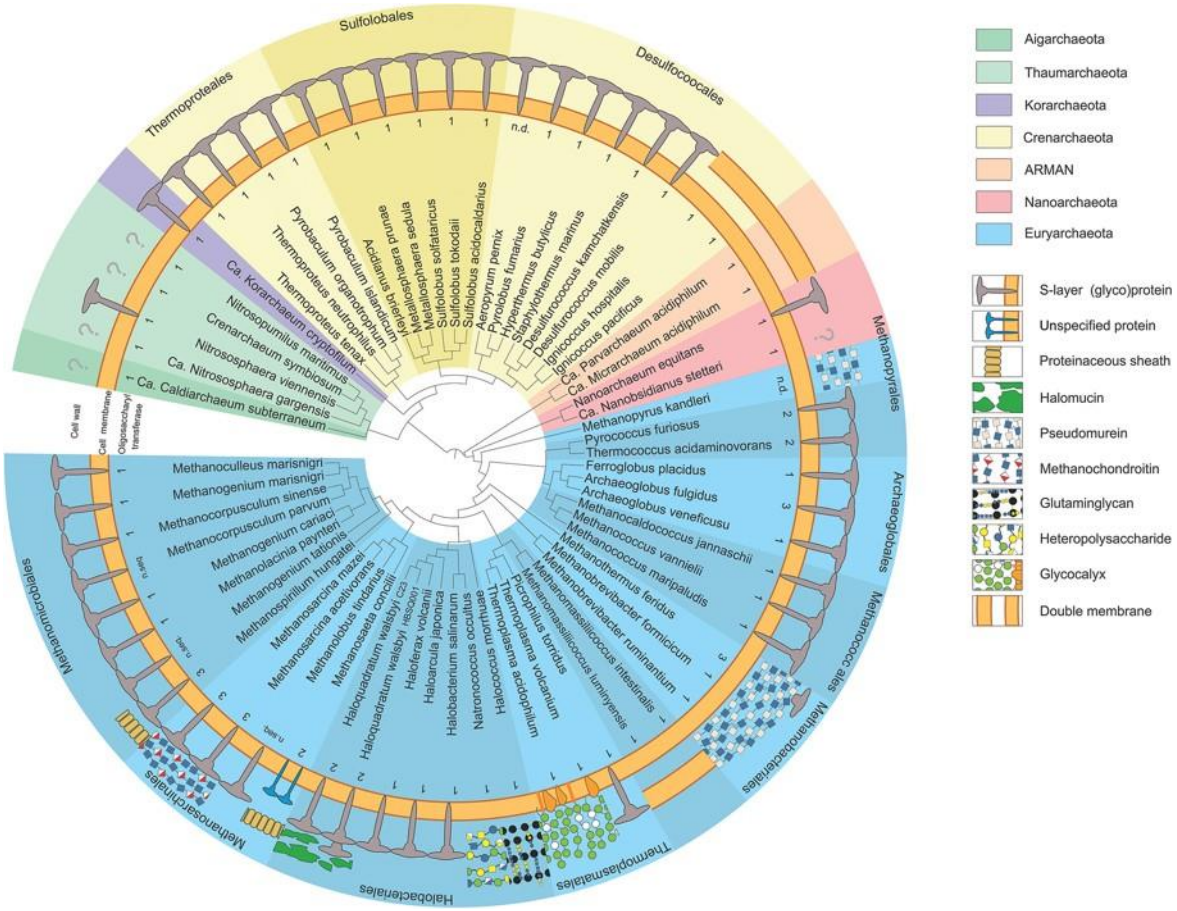


Figure 1.2 Distribution of cell wall types in seven different types of super-phyla of the archaeal domain (Albers *et al.*, 2015), copyright to Cold Spring Harbor Laboratory Press.

1.3 Cell wall in Bacteria

Murein, commonly known as peptidoglycan, is a major component that forms cell walls in the *Bacterial* domain. With the rare exception of *Mycoplasma* and a few other species that do not contain cell walls, all bacteria contain murein as a cell wall component. The murein cell wall is formed by the polymerisation of a building block with a glycan-pentapeptide subunit (Figure 1.3A). The glycan backbone is formed by alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid residues (MurNAc) that are linked together by β-1,4 glycosidic bonds. These glycan moieties are cross-linked by short peptides that are mostly contributed by L- and D-amino acids and a single diamino acid (*meso*-diamoinopimelic acid

(*m*-DAP) in the case of Gram negative bacteria. The addition of the pentapeptide occurs at the UDP-MurNAc of the glycan backbone by a series of four murein peptide ligases (MurC, D, E and F) through a series of reaction with a common catalytic mechanism (Barreteau *et al.*, 2008). The composition of the cross-linking pentapeptide appears as L-Ala-γ-D-Glu-*m*DAP/L-Lys-D-Ala-D-Ala in nascent peptidoglycan where the last D-Ala is removed during cross-linking (Schleifer & Kandler, 1972). However, some alternative sequences for the cross-linking peptide subunits, while retaining conservation in the basic features, have been reported for certain organisms (Vollmer *et al.*, 2008).

The structural properties of the murein are determined by the length of the glycan and the proportion of the peptide subunit cross-linking the glycan backbones. The cell wall of Gram positive bacteria are multilayered and relatively thick compared to that of Gram negative bacteria. Thus, the murein sacculi of the Gram positive bacteria contains higher cross-linkage, providing them with a stronger and stiffer cell wall whereas Gram-negative bacteria possess a lower level of cross-linkage and this contributes to mostly a monolayer arrangement of the sacculus. In addition, the length of the glycan strands is also another key feature determining the properties of the cell wall and follows a wide distribution among bacteria (Vollmer, *et al.*, 2008).

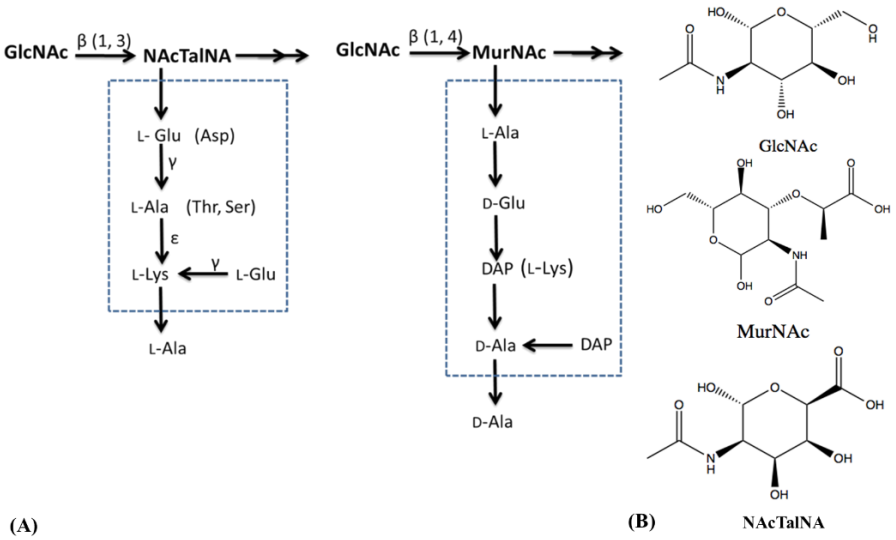


Figure 1.3 (A) The building blocks of pseudomurein (left) and murein (right) cell wall types (Schofield *et al.*, 2015), (B) chemical structures of sugars present in these cell wall types.

1.3.1 Murein biosynthesis

The murein biosynthesis pathway is completed in three different stages at three different locations: starting in the cytoplasm (synthesis of nucleotide precursors); at the inner membrane (synthesis of lipid-linked intermediates); and finally at the outer membrane of cytoplasmic membrane (polymerization reaction) (Figure 1.4).

The initial step in the cytoplasm begins with the formation of the glycan backbone. UDP-GlcNAc is synthesised from fructose-6-phosphate which involves four successive reactions catalysed by: glucosamine-6-phosphate synthase (GlmS), phosphoglucosamine mutase (GlmM), glucosamine-1-phosphate acetyltransferase and *N*-acetylglucosamine-1-phosphate uridyltransferase (GlmU, a bifunctional enzyme) (Barreteau, *et al.*, 2008). The product of this reaction (UDP-GlcNAc) is then used to produce UDP-MurNAc utilizing the MurA and MurB enzymes of the murein biosynthesis pathway. MurA catalyses the initial step where enoylpyruvate from phosphoenolpyruvate is transferred to the C3-OH of a UDP-GlcNAc

and forms UDP-GlcNAc-enolpyruvate, which further undergoes a reduction catalysed reaction mediated by MurB (UDP-*N*-acetyl-enolpyruvylglucosamine reductase) to produce UDP-MurNAc with the use of NADPH and a solvent-derived proton (Barreteau, *et al.*, 2008). Chemical structures of sugars used in murein and pseudomurein cell wall types are shown in Figure 1.3B. Sugars present in other archaeal cell wall types and the chemical structures of key intermediates and substrates of murein cell wall biosynthesis pathway discussed in this thesis are shown in Appendix 1.

The cross-linking pentapeptide is then formed by the stepwise addition of the amino acid moiety to the UDP-MurNAc in a sequential manner by a series of murein ligases (MurC, MurD, MurE and MurF). Each of the enzymes is specific for the addition of a specific amino acid to the cross-linking peptide subunit. The synthesis of the pentapeptide starts with the reaction catalysed by the MurC peptide ligase, which adds the first amino acid L-Ala to the UDP-MurNAc, and forms a UDP-MurNAc-L-Ala as a product. The product of the MurC is then used by the MurD peptide ligase as a substrate to produce UDP-MurNAc-L-Ala-D-Glu. The bacterial murein ligases seem to be stereospecific to the amino acid they bind (Pratviel-Sosa *et al.*, 1994; Liger *et al.*, 1995). The third amino acid of the murein peptide subunit is L-lysine in Gram positive bacteria and meso-diaminopimelic acid (mDAP), a derivative of lysine, in Gram negative bacteria. The binding of the relevant amino acids by MurE is highly specific and incorporation of an inappropriate amino acid could lead to cell death (Mengin-Lecreulx *et al.*, 1999). Hence, the MurE peptide ligase forms a UDP-MurNAc-tripeptide to which the MurF peptide ligase adds the D-Ala-D-Ala dipeptide to form the UDP-MurNAc-pentapeptide. D-Ala is produced from L-Ala by an alanine racemase, which uses a covalently bound pyridoxal 5′-phosphate cofactor to catalyse the reaction (Doublet *et al.*, 1993). This

D-Ala is utilised by the D-Alanine:D-Alanine ligase (ddl) through two half reactions to produce D-Ala-D-Ala dipeptide (Walsh, 1989).

The synthesised nucleotide precursor from the cytoplasmic step is transferred onto the membrane-associated step and is catalysed by the integral membrane protein enzyme, MraY. The undecaprenyl phosphate carrier (C_{55}-P) on the cytoplasmic side of the cell membrane accepts the UDP-MurNAc-pentapeptide and in the presence of MraY forms undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide (PP-MurNAc-pentapeptide), also known as Lipid I (Chung *et al.*, 2013). The translocase enzyme (MurG) subsequently catalyses the transfer of the GlcNAc moiety from UDP-GlcNAc to Lipid I (PP-MurNAc-pentapeptide) to yield undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide-GlcNAc (PP-MurNAc-pentapeptide-GlcNAc), commonly known as Lipid II. Amidation of α-D-isoglutamic acid in position 2 of the Lipid II stem peptide in some pathogenic Gram-positive bacteria are suggested to play a decisive role in the polymerization of cell wall building blocks. The amidation of the α-carboxyl group of D-glutamic acid to form iso-D-glutamine is catalysed by a recently identified MurT enzyme, homologous to murein ligases, in complex with GatD (Noldeke *et al.*, 2018). Finally, the orientation of Lipid II is inverted with respect to the membrane using a flippase enzyme mechanism and is then used as a substrate for polymerisation outside of the cytoplasmic membrane (Bouhss *et al.*, 2008; Sham *et al.*, 2014).

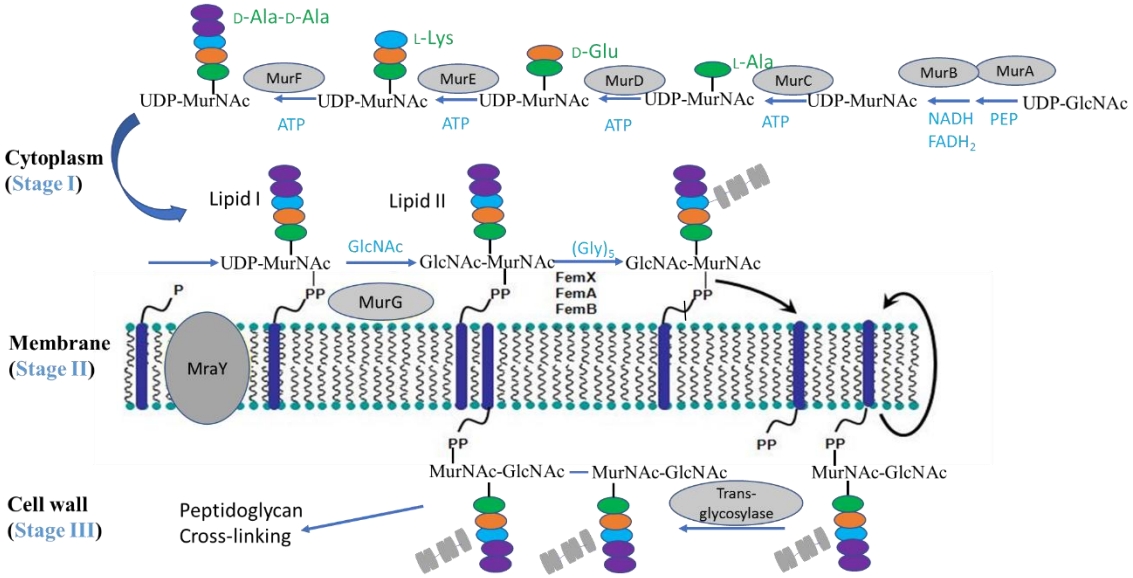


Figure 1.4 Summary of murein biosynthesis pathway in bacteria adapted and modified from Bugg *et al.*, (2011) and enzymes catalysing each step during the pathway.

1.3.2 Murein peptide ligases

Murein peptide ligases that are involved in the formation of the UDP-MurNAc-pentapeptide subunit are ATP-dependent and have a similar catalytic mechanism for all four peptide ligases (MurC, D, E and F) (El Zoeiby *et al.*, 2003). The characteristic activity of murein peptide ligase involves an activation of the C-terminal amino acid of the growing nucleotide precursor to an acyl-phosphate intermediate, followed by a nucleophilic attack by the NH_2 group of the added amino acid leading to release of phosphate and formation of a new amide or peptide bond (Bouhss *et al.*, 2002). The four peptide ligases show limited sequence identity among each other (15-25%), yet contains a common structural motif and an overall domain arrangement. All the murein ligases contain three structural domains that include an N-terminal Rossmann-like α/β fold domain, a middle ATPase domain, and a C-terminal Rossmann-type α/β fold domain (Smith, 2006).

The three domains of the murein ligases have a similar three-dimensional structural arrangement with the middle domain at the centre flanked by the N- and C-terminal domains such that the interface of the domains forms a crescent-shaped cavity (Figure 1.5 and Figure 1.6A). The N- and C-terminal domains are linked to the middle domain by hinge regions which allow these domains to move 'to' and 'from' the middle domain and provides flexibility to the peptide ligases to adopt 'closed' and 'open' conformations (Perdih *et al.*, 2007). The structures of murein ligases usually adopt an 'open' conformation in the apo state and undergo a conformational change to adopt a 'closed' conformation when the substrate is bound (Smith, 2006). The ordered kinetic mechanism for the substrate binding in MurC peptide ligase begins with the addition of ATP, followed by the nucleotide substrate (UDP-MurNAc) and eventually the amino acid (Emanuele *et al.*, 1997) and is understood to be the same for all the murein ligases.

The ATP-binding site in all murein peptide ligases is present at the interface of the middle- and C-terminal domains and contains invariant residues. In addition, the UDP-MurNAc-based growing substrate of the consecutive murein ligases binds at the crescent-shaped cavity where the nucleoside of the UDP-MurNAc is bound to the N-terminal domain followed by the phosphate groups and muramyl sugar, which extends via the crescent-shaped cavity towards the C-terminal domain where the amino acid gets ligated (Mol *et al.*, 2003; Basavannacharya *et al.*, 2010b). Hence, the crescent-shaped cavity would require a gradual increase in size from MurC-MurF as the size of the substrate they bind increases with sequential addition of amino acids by preceding enzymes.

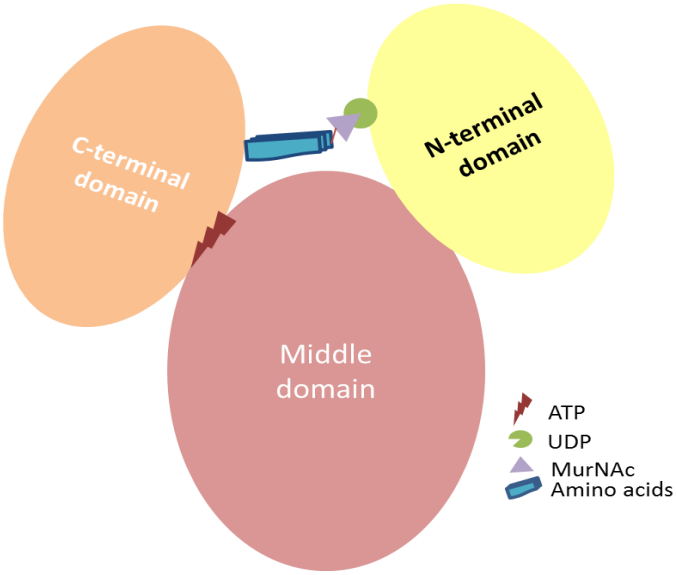


Figure 1.5 Simplified diagrammatic representation of bacterial murein ligases and the binding site for UDP-MurNAc-pentapeptide and ATP. The structures are coloured based on each separate domain; yellow for N-terminal, terracotta for middle and orange for C-terminal domain.

The N-terminal domain is responsible for the binding of the UDP head-group of UDP-MurNAc, a common substrate motif for all murein peptide ligases to which the amino acids are ligated sequentially by their respective ligase enzymes. This domain is structurally similar in MurC and MurD peptide ligases, but differs to those in MurE and MurF and vice-versa, which has led to the classification of the N-terminal domain in the murein ligases as MurC/D type and MurE/F type (Smith, 2006). Thus, based on the type of N-terminal domain possessed the bacterial murein ligases can be grouped into MurC/D and MurE/F types. The N-terminal domains within each group share a similar overall secondary structure topology, but have differences in their connectivity (Smith, 2006). The conserved glycine-rich motif is reported for both MurC and MurD at the N-terminal domain that is responsible for UDP-binding (Mol, *et al.*, 2003), but MurE and MurF do not share a similar conserved UDP binding motif at the N-terminal domain. Moreover, the MurE binds the UDP head-group solely at the N-terminal domain whereas the MurF binds the UDP at the interface of middle

and N-terminal domains, for example, the *E. coli* MurF structure contains a part of the middle domain, the β9-α6 loop, which extends out towards the N-terminal domain to bind pyrophosphate (Yan *et al.*, 2000).

The middle domain is the most conserved of the three domains in all murein ligases and shares a common catalytic mechanism in using the phosphate bond energy to catalyse the addition of amino acids onto the growing peptide chain of the cross-linking peptide stem. The middle domain of murein ligases contains six invariant residues, in addition to an ATP-binding consensus sequence, commonly referred to as the P-loop (Eveland *et al.*, 1997). The P-loop is a signature primary structure in adenine and guanine nucleotide-binding proteins that is rich in glycine residues followed by a conserved lysine and a serine or threonine residue (Saraste *et al.*, 1990). Moreover, the six conserved residues and P-loop are a signature for the Mur ligase family as they are also present in three other enzymes, which are not directly related to peptidoglycan biosynthesis: folylpoly-γ-L-glutamate synthase (FPGS) (Sheng *et al.*, 2000), the C-terminal region of cyanophycin synthase CphA (Ziegler *et al.*, 1998), and poly-γ-glutamate synthase CapB (Candela & Fouet, 2006).

The C-terminal domains of the murein ligases have a structural fold resembling dihydrofolate reductase and are involved in amino acid residue binding (Smith, 2006). The C-terminal domain in all four murein peptide ligases contains two conserved residues, histidine and aspartate, of which the aspartate residue is involved in ribose sugar binding and the histidine residue is involved in tri-phosphate binding (Smith, 2006). Depending on the amino acid specificity of the MurE, conserved motifs have been identified: a consensus DNPR motif for *m*-DAP and a D(D,N)P(N,A) motif for those MurE peptide ligases that bind L-lysine (Boniface *et al.*, 2006). The product-bound structure for the *H. influenzae* MurC has been studied, which shows that the L-Ala portion of the product (UDP-MurNAc-L-Ala) is

positioned between the arginine residues (377 and 380) and the specificity for this interaction derives from a shallow pocket formed by His348, His376, Tyr346 and Ala459 (Mol, *et al.*, 2003) (Figure 1.6B).

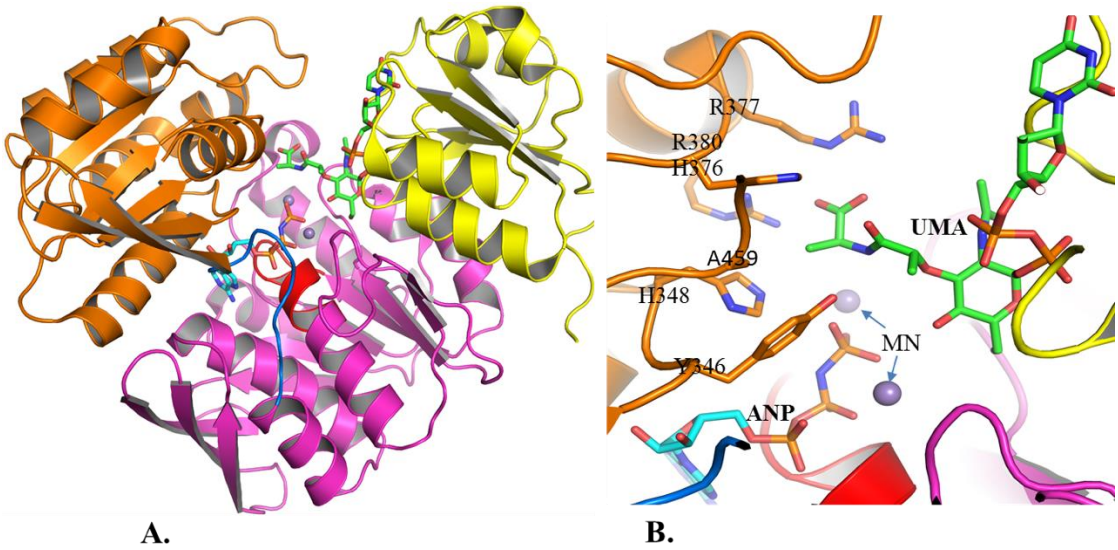


Figure 1.6 Structure of the *H. influenzae* MurC highlighting the three domain architecture of the murein ligase and the product-binding region, (A) Cartoon representation of MurC with N-terminal domain in yellow, middle domain in magenta and C-terminal domain in orange. The UMA (green) and ANP (cyan) are represented as sticks, P-loop is coloured red and bound manganese ions shown as sphere. (B) Amino acid binding region of the *H. influenzae* MurC at the C-terminal domain.

1.3.3 Folylpolyglutamate synthase (FPGS)

FPGS is a member of the Mur ligase-like family with a highly specialised variation. Unlike murein ligases that adds an individual amino acid to the growing peptide of the amino acid substrate in a sequential pattern by consecutive ligases, FPGS initially binds a folate substrate and then repetitively binds up to ten additional L-Glu residues to the carboxyl terminal of the growing poly-glutamate chain (Sun *et al.*, 2001). The catalytic mechanism for the FPGS involves the usage of the product from one reaction as a substrate for the second

step with appropriate movement to the correct position for phosphoryl transfer from a new ATP (Smith, 2006). The key characteristics for the Mur ligase family, as discussed earlier in Section 1.3.2**,** are present in FPGS. Further, the overall structure of FPGS shows similarities to the middle and C-terminal domains of the murein ligases but does not contain an N-terminal equivalent domain. The comparative analysis of FPGS with the *E. coli* MurD structure reveals an insertion of 50 residues at the N-terminal domain of the FPGS structure (Sheng, *et al.*, 2000) that contains the Ω-loop, which is important in the formation of the active site and is essential for enzyme function (Sun, *et al.*, 2001; Smith *et al.*, 2006). The available FPGS sequences are approximately 50 amino acids shorter than the pseudomurein ligases, which have an insertion of approximately 50 amino acids at their N-termini and corresponds to the middle domain of the *E. coli* MurD structure (Sheng, *et al.*, 2000). A similar insertion was observed while comparing FPGS with other murein ligases and the pseudomurein ligases determined during this project. FPGSs are mostly present in bacteria but have also been reported in several archaeal genomes that are suggested to have been acquired from bacteria through lateral gene transfer (Leipe *et al.*, 2003).

1.4 Methanogens

Methanogens are distinct archaeal cell types that are grouped in the *Euryarchaeota*, one of the five superphyla of the *Archaeal* domain, and are shown to be physiologically and phylogenetically separate from other existing archaea (Cohen, 2014; Enzmann *et al.*, 2018). Methanogens are best known for their role in methane production and possess a unique metabolism and energy conservation strategy (Garcia *et al.*, 2000). They are found in a wide range of natural habitats including anoxic fresh water swamps, ocean and lake sediments, hydrothermal vents, digestive tracks of animals and the cytoplasm of anaerobic protozoa (Cavicchioli, 2011; Enzmann, *et al.*, 2018). Methanogens are diverse in their morphology as

well as in terms of phylogenetics and have been divided into seven orders: *Methanobacteriales*, *Methanococcales*, *Methanomicrobiales*, *Methanosarcinales*, *Methanopyrales*, *Methanocellales* and *Methanoplasmatales* (Kletzin, 2007; Paul *et al.*, 2012; Borrel *et al.*, 2013; Enzmann, *et al.*, 2018).

1.4.1 Pseudomurein cell wall

Out of these seven orders of methanogens, the orders *Methanobacteriales* and *Methanopyrales* contain a unique cell wall envelope, pseudomurein, which is unique among genera that grow under diverse conditions ranging from mesophilic to hyperthermophilic temperatures. Pseudomurein is 15-20 nm thick with a density of 1.39-1.46 g/cm^3 and each elementary disaccharide peptide subunit has a cell dimension of 4.5 x 10 x 21-22.5 Å - suggesting an overall structural similarity to bacterial murein (Figure 1.3A). Despite the overall structural similarity of the murein and pseudomurein present in bacteria and methanogens, respectively, significant differences are found: i) in the chemical composition of the glycan backbone, ii) in the stereo-specificity of amino acids forming the cross-linking pentapeptide, and iii) in the peptide bonds and inter-peptidyl cross-link (Kandler & Konig, 1978; Leps *et al.*, 1984; Kandler & König, 1993; Claus & König, 2010).

Pseudomurein contains *N*-acetyl-L-talosaminuronic acid (NAcTalNA) as a sugar moiety in the glycan backbone that is typically linked to *N*-acetyl-D-glucosamine (GlcNAc) through *β*-1,3-glycosidic bonds. The presence of the NAcTalNA sugar moiety is unique for pseudomurein whereas the bacterial murein consists of *N*-acetylmuramic acid (MurNAc) (Leps *et al.*, 1984). Further, pseudomurein consists of the polypeptide that is formed by only L-amino acids, usually glutamate, lysine and alanine in a molar ratio of 2:1:1 and cross-links the parallel glycan subunits formed by NAcTalNA β(1→3) GlcNAc (Kandler & König,

1978). The activated glutamic acid residue and the three activated peptides without the glycan components were isolated from trichloroacetic acid extracts of *M. thermautotrophicus*, which shows that UDP is linked to the α-amino group of the glutamic acid (Hartman *et al.*, 1990). Hence, the cross-linking polypeptide in pseudomurein is formed through the ligation of L-amino acids to the activated glutamic acid residue (N^{α}-UDP-Glu^{γ}-P). The pentapeptide subunits of pseudomurein have been synthesised *in vitro* in seven steps using cell extracts of *Methanothermobacter marburgensis* (Hartmann & König, 1994). This suggests that the cytoplasmic peptide precursor formation of pseudomurein could vary compared to that of murein. The peptide moiety of pseudomurein exhibits a high frequency of ε- and γ-isopeptide bonds making it insensitive towards common proteases (Kandler & König, 1993). This pentapeptide is attached to the carboxyl group of NAcTalNA via the *N*-amino group of glutamic acid (Hartmann & König, 1990).

1.4.1.1 Pseudomurein biosynthesis

The information on the pseudomurein biosynthesis pathway of methanogenic archaea is limited compared to the murein biosynthesis pathway. The pseudomurein biosynthesis pathway is proposed to occur in three different parts of the cell, the cytoplasm and the inner and outer layers of the cell membrane as that shown for murein biosynthesis (Figure 1.7) (Hartmann & König, 1990; Claus & König, 2010). Pseudomurein biosynthesis is initiated in the cytoplasm where the production of the glycan backbone and the pentapeptide structure are carried out separately. Pseudomurein is unique with a glycan moiety, UDP-NAcTalNA, to which the pentapeptide subunit is attached. The biosynthesis of UDP-NAcTalNA involves two steps. It starts with the epimerisation of UDP-*N*-acetylgalactosamine (UDP-GalNAc) to produce UDP-*N*-L-acetylaltrosaminuronic acid (UDP-L-NAcAltNA) as an intermediate product, which is then oxidised to UDP-NAcTalNA. The UDP-NAcTalNA and GlcNAc are

then linked to form the disaccharide with a β-1,3 glycosidic bond which acts as a glycan intermediate (Figure 1.7) (Claus & König, 2010). Production of the pentapeptide commences at the same time with a glutamic acid that is phosphorylated at the *N*-α-amino group and converted into N^{α}-UDP-glutamic acid (Hartmann *et al.*, 1989) followed by phosphorylation of the α-carboxyl group of the glutamate moiety to form N^{α}-UDP-Glu^{γ}-P. The stepwise addition of L-amino acids to the activated glutamate moiety requires the presence of different peptide ligase enzymes (to be studied in this project), and also occurs in the cytoplasm. The pentapeptide subunit and the NAcTalNA glycan unit of the glycan backbone are attached by the γ-amino group of glutamic acid and then transported across the cell membrane to the outside of the cell membrane at stage III (König *et al.*, 1983).

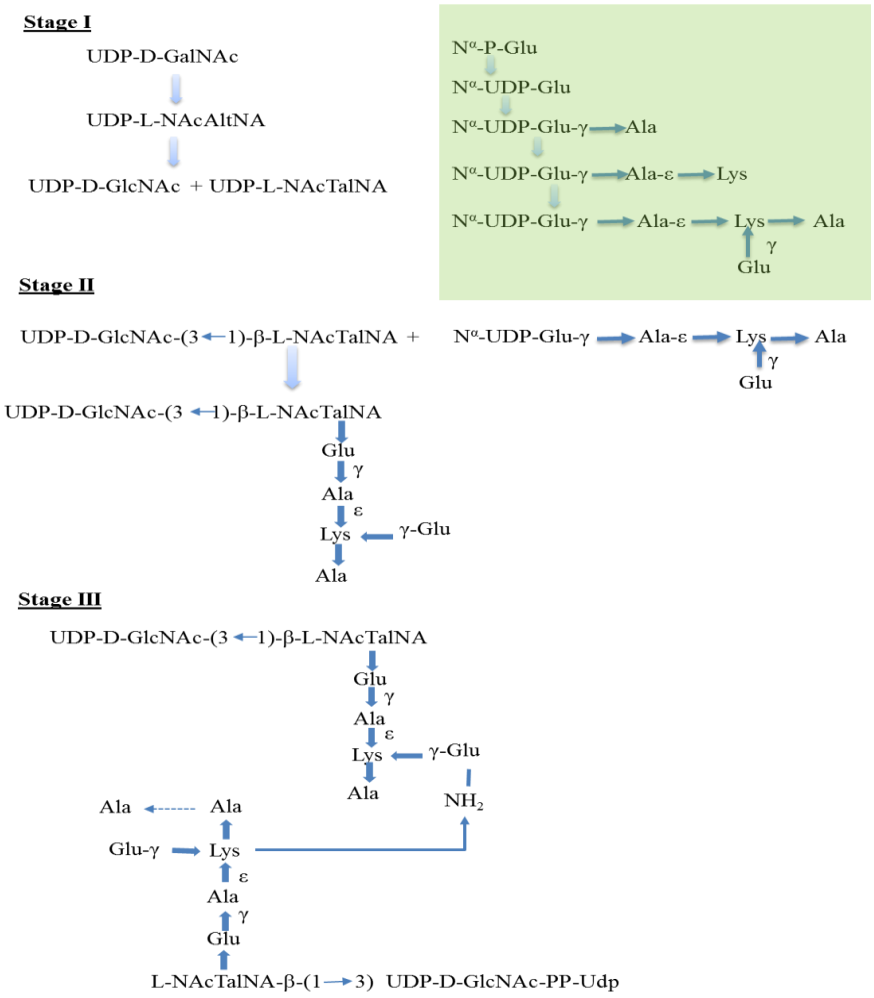
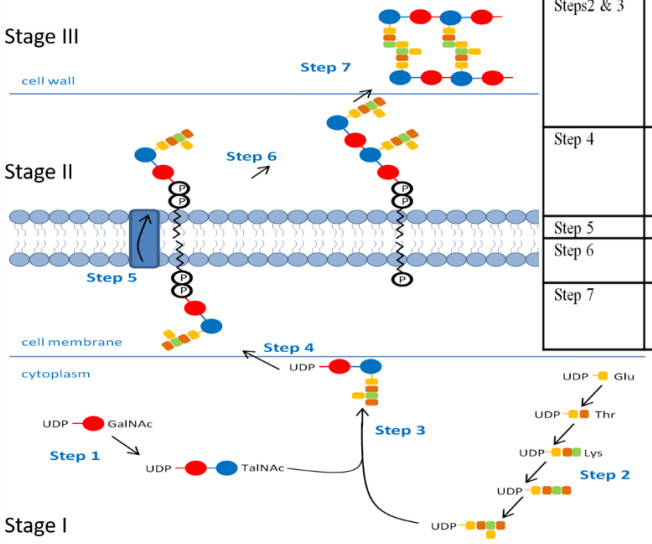


Figure 1.7 Proposed pseudomurein biosynthesis pathway (Hartmann & König, 1990). The steps for synthesis of the pentapeptide are highlighted in green and are suggested to be performed in parallel with glycan backbone formation.

The genome sequences of several pseudomurein-containing methanogens have been completed in recent years, which has added to a hypothesis about how pseudomurein biosynthesis could occur. A depiction of the proposed pseudomurein pathway in *M. ruminantium* shows that the biosynthesis of pseudomurein might follow the same pattern, and likely involves similar mechanisms and analogous steps, to murein biosynthesis as shown in Figure 1.8 (Leahy *et al.*, 2010). Based on the sequence homology with the bacterial murein peptide ligases and their distribution among analysed pseudomurein-containing

methanogens, it is proposed that similar enzymes are involved in the sequential addition of L-amino acids to the pentapeptide subunit and involves the enzymes encoded by the genes in Steps 2 and 3 in Figure 1.8.

Identification of MraY and MurG encoding gene homologues in pseudomurein-containing methanogens suggests that a similar pattern of nucleotide precursor transfer onto the inner membrane layer occurs, which is then followed by the flipping the cell wall subunit to the outside of the cell membrane. The lipid-linked intermediates of pseudomurein biosynthesis pathway has been isolated and characterised (Hartmann & Konig, 1990) and the presence of MraY genes suggest that a similar mechanism to that found in the bacterial murein biosynthesis pathway. The role of MurG is, however, thought to be different, or at least occur in a different location, due to the formation of glycan disaccharide already during the cytoplasmic stage. Finally, each of these pentapeptide units bound to the disaccharide are attached to the growing glycan stand outside the cell membrane during the final stage (stage III) of pseudomurein biosynthesis.



Step	*Locus tag*	*Annotation*
Step1	mru2126	cell wall biosynthesis protein UDP-glycosyltransferase family
Steps2 & 3	mru0707	cell wall biosynthesis protein Mur ligase family
	mru1042	cell wall biosynthesis protein Mur ligase family
	mru1118	cell wall biosynthesis protein Mur ligase family
	mru1745	cell wall biosynthesis protein Mur ligase family
	mru2091	cell wall biosynthesis protein Mur ligase family
	mru2092	cell wall biosynthesis protein Mur ligase family
Step 4	mru0964	cell wall biosynthesis protein phospho-N-acetylmuramoyl-pentapeptide-transferase family
	mru1041	cell wall biosynthesis protein phospho-N-acetylmuramoyl-pentapeptide-transferase family
Step 5	mru2136	polysaccharide biosynthesis protein
Step 6	mru0470	conserved hypothetical transmembrane protein
	mru2175	glycosyl transferase GT2 family
Step 7	mru0824	adhesin-like protein with transglutaminase domain
	mru1499	adhesin-like protein with transglutaminase domain
	mru1604	adhesin-like protein with transglutaminase domain

Figure 1.8 Depiction of pseudomurein biosynthesis in *M. ruminantium* (Leahy, *et al.*, 2010). It has been observed that, similar to the murein biosynthesis mechanism, pseudomurein biosynthesis also involves three different stages of cell compartment. In addition, the genes involved at each stage of the proposed pseudomurein biosynthesis pathway agree and support the role of hypothesised enzymes during this project.

1.4.2 Pseudomurein pentapeptide synthesis

Of most interest to this project, both the murein and pseudomurein biosynthesis pathways show similar stepwise addition of amino acids to the cross-linking peptide chain (Figure 1.9A and Figure 1.9B, respectively). In addition to the putative enzymes involved in amino acid ligation during pentapeptide biosynthesis, the candidate methanogen genes encoding enzymes for different steps of pseudomurein biosynthesis have been identified and are distantly related to the enzymes of the core murein biosynthesis pathway. Unlike murein pentapeptide biosynthesis, pseudomurein pentapeptide biosynthesis involves a separate process for the glycan backbone formation and pentapeptide biosynthesis in the early

cytoplasmic step. However, it is observed that the addition of the L-amino acids to the UDP-linked activated glutamic acid residue (N^{α}-UDP-Glu) follows a similar mechanism (Figure 1.9B). It has been suggested that a separately formed UDP-activated disaccharide and UDP-linked pentapeptide intermediate together form a UDP-activated disaccharide pentapeptide during the stage II at the inner cell membrane as proposed in Figure 1.7 and depicted in Figure 1.8. The analogous murein peptide ligase genes, except for *murA* and *murB*, have been identified in methanogens and the enzymes encoded by the genes are hypothesised to be involved in catalysing similar reactions. The MurA and MurB enzymes catalyse the initial two reactions of the murein pentapeptide biosynthesis and have no role in the pseudomurein pentapeptide biosynthesis (Figure 1.9A).

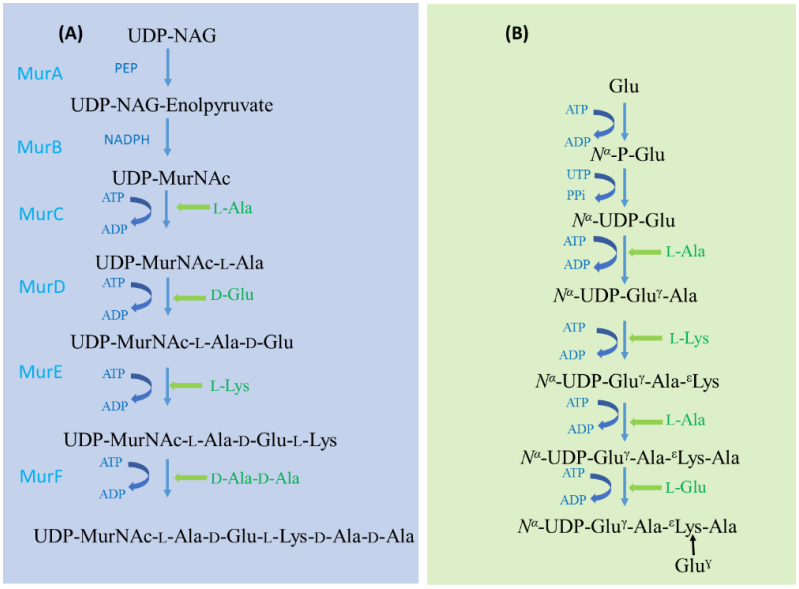


Figure 1.9 Pentapeptide biosynthesis pathway for murein (A) and pseudomurein (B).

1.5 Hypothesis and scope of the study

The project contributes to the first investigation of the enzymatic steps of pseudomurein biosynthesis and structural information of the enzymes involved in the pathway, thus leading to a significant increase in our understanding of how cell walls evolved. Identification of the

candidate methanogen genes encoding enzymes for the pseudomurein biosynthesis and in particular, four putative pseudomurein ligases that are thought to be involved in the addition of L-amino acids during cross-linking pentapeptide biosynthesis in pseudomurein-containing methanogens suggest a similar peptide ligase activity in both murein and pseudomurein biosynthesis pathway. Only the major differences have been suggested between murein and pseudomurein in the past (Hartmann & König, 1990; Albers & Meyer, 2011), but this study finds a low level of sequence identify (< 25%) to analogous key enzymes in the murein biosynthesis pathway and for the first time reports specific homologies to each murein peptide ligase type (MurC-F). This suggests a relationship between murein and pseudomurein peptide ligases and structure determination of the pseudomurein peptide ligases could aid in phylogenetic analyses though structure-based alignment and demonstrating structural and functional similarity to their bacterial homologues.

The genes encoding for the homologous MurC ligases that catalyses the addition of an alanine residue in murein pentapeptide biosynthesis pathway have been identified in pseudomurein-containing methanogens. Similarly, homologous genes encoding for MurD, MurE and MurF ligases responsible for the ligation of D-glutamate, L-lysine and D-ala-D-ala, respectively, in bacteria have also been observed. The results based on bioinformatics analyses (multiple sequence alignment and homology-based structure predictions) showed that the identified candidate gene-encoded proteins possess the key features required for pentapeptide biosynthesis (ATP- and nucleotide precursor-binding sites) and support the possibility that the proteins are highly likely to be involved in pseudomurein biosynthesis. The evidence for the homology of these enzymes in the pseudomurein biosynthesis pathway with the murein peptide ligases will therefore be the basis of the project. For this project, it is proposed that the putative pseudomurein ligases - pMurC, pMurD1/D2 and pMurE,

homologous to respective murein ligases (MurC, D and E), are involved in the addition of alanine, glutamate and lysine amino acid residues to form the cross-linking peptide subunit. The sequence homology of bacterial MurF peptide ligase with archaeal pMurF has also been analysed using PSI-BLAST; however, the pMurF distribution is not limited to pseudomurein-containing methanogens suggesting a role other than as a peptide ligase. Recently, MurF in archaea has been shown to be involved in F_{430} co-factor synthesis (Zheng *et al.*, 2016). In addition, there are numbers of homologous genes encoding for enzymes (pGlmU, ddl ligase, pMurG, pMraY) present among pseudomurein-containing methanogens that could possibly be involved in similar catalytic activity of the pseudomurein biosynthesis pathway to that performed by homologous enzymes of the murein biosynthesis pathway (Table 6.1). However, the detailed study and characterisation of these other enzymes are beyond the scope of this study.

The distinct relationship between murein and pseudomurein ligases and the similarity of their biosynthetic mechanisms suggests a common origin for cell wall evolution in bacterial murein and pseudomurein-containing methanogens. To further verify this hypothesis, structural and functional characterisation of pseudomurein ligases will be carried out during this project. The putative pseudomurein ligases show sequence homology to the bacterial murein ligases, even though the sequence identity is less than 30%. For this reason 3D structural analysis of putative pseudomurein ligases is essential for comparative analysis, as structures are more conserved than sequence (Kim & Lee, 2007). There are numbers of murein ligase structures available in the Protein Data Bank (PDB) (Berman *et al.*, 2000) that can be used for comparative analysis and structure-based alignments can be used to analyse distant evolutionary relationships. The structural and functional characterisation of the putative pseudomurein ligases will help to resolve the elusive and complex evolutionary

history of bacterial and archaeal cell wall biosynthesis. If the structures of the putative pseudomurein ligases targeted share any structural homology with the bacterial murein ligases, it would show that pseudomurein ligases are derived from common ancestral genes that encode the ligase domains. However, if large structural differences are identified, this could imply an independent evolution of the methanogen enzymes.

In addition to the structural information, the *in vitro* synthesis of pentapeptide using the methanogen cell extracts as conducted by Hartmann and Kong (Hartmann & König, 1994) will be attempted as reference experiment to test our proposed candidate methanogen ligases in the pentapeptide subunit biosynthesis pathway. Hence, this study will contribute in understanding a critical step in the early evolution of life and thereby provide new insights into the detailed structural knowledge of bacterial murein ligases that are important for antibiotic development (Schneider & Sahl, 2010). The results from this project could also contribute towards developing methods to control methane emissions from ruminants by targeting a specific group of methanogens in the rumen which contain pseudomurein (Leahy *et al.*, 2010).

1.6 Objectives

The main objective of this thesis was to identify the key enzymes involved in biosynthesis of pseudomurein and, specifically, the enzymes that catalyse the biosynthesis of the pentapeptide. Therefore, after the successful completion of the project it would be possible to identify the enzymes that catalyse individual steps of pentapeptide biosynthesis in the pseudomurein biosynthesis pathway. I will also test the hypothesis that the biosynthesis of bacterial murein and archaeal pseudomurein share a common evolutionary history. To achieve this main goal the project is sub-categorised into different parts with specific goals.

1. The candidate genes involved in the pseudomurein biosynthesis pathway will be identified using bioinformatics tools and sequence search algorithms. The identified genes will be compared with their respective encoded enzymes involved in murein biosynthesis to analyse their relationships to one another and if they share an evolutionary history. Ten candidate genes involved in the pseudomurein biosynthesis that share similarity with bacterial genes have been identified. This finding supports the hypothesis that cell wall evolution originated from common ancestral genes. However, low sequence identities between the encoded proteins and a lack of structural evidence is preventing the construction of reliable phylogenetic trees.
2. The biochemical characteristics of the identified enzymes will be studied to confirm their role in pseudomurein biosynthesis, thereby identifying the function of each candidate enzyme. The specific step catalysed by each enzyme in the biosynthetic pathway will be confirmed. The biochemical properties of the enzymes will be compared to bacterial murein synthesising enzymes that will help to elucidate any ancestral linkages.
3. The 3D structure of each class of enzymes will be determined which will enable us to make structural comparisons with bacterial murein ligases (more than 30 structures of different bacterial murein ligases have been deposited in the PDB database). Successful pseudomurein ligase structure determination will be a novel achievement and will be potentially helpful in producing robust structure-based phylogenetic comparisons. This will help overcome the limitations of the low sequence identity issue in constructing reliable phylogenetic trees and therefore help to clarify evolutionary links between cell wall biosynthesis in *Bacteria* and *Archaea*.

Chapter 2 Materials and Methods

2.1 Equipment and reagents

2.1.1 Water

Double-distilled water was used for media preparation whereas MilliQ water (Millipore) was used for experiments related to molecular biology, protein crystallography and biochemical assays.

2.1.2 Chemicals

Chemicals used for the experimental work were of analytical grade or higher, unless otherwise indicated.

2.1.3 *E. coli* strains used

All the pseudomurein peptide ligases studied during this project were expressed in the *E. coli* BL21 Star™ (DE3) strain. In addition, *E. coli* TOP10 strain was used for plasmid preparation. Both of these strains were obtained from Invitrogen (USA).

2.1.4 DNA source material

The identified putative pseudomurein peptide ligase genes from *Methanothermobacter thermautotrophicus* strain ΔH (Mth) and *Methanothermus fervidus* (Mfer) were mostly targeted for the study. The *M. thermautotrophicus* strain ΔH is an extensively studied model organism that represents pseudomurein-containing methanogens for which certain biochemical information and a complete genome sequence is available (König *et al.*, 1983;

Smith *et al.*, 1997). Moreover, *M. fervidus* is a hyper-thermophilic methanogen that was initially isolated from a hot spring in Iceland (Stetter *et al.*, 1981). In addition, putative pseudomurein peptide ligases from *Methanothermobacter marburgensis* (formerly *Methanothermobacter thermautotrophicum* strain Marburg) (MBB) were also targeted for structural determination. The DNA source materials that were used during this project are summarized in Table 2.1. The plasmids used for protein expression of Mth532 and Mth734 were available from the previous work of Maximilian Wolf (Wolf, 2010). These plasmids had been constructed by cloning the targeted gene into the pET100D vector (Invitrogen, USA). The plasmids of other targets were designed and obtained from Genscript (USA) using *de novo* synthesis of the target genes. All the plasmids obtained from Genscript used pET-15b as an expression vector and used NdeI and BamHI for sub-cloning sites and contained a thrombin cleavage site. The obtained plasmids were codon optimised for bacterial (*E. coli*) expression using OptimumGene™.

Table 2.1 DNA source material used during the project.

Organism	Gene number	Accession Number	Vector	Selection	Description
MTH	Mth530	WP_010876169.1	pET15b and pET100D	Amp^R	Putative pseudomurein peptide ligase type C (pMurC)
MTH	Mth531	WP_010876170.1	pET15b and pET100D	Amp^R	Putative pMurD1
MTH	Mth532	WP_010876171.1	pET100D	Amp^R	Putative pMurD2
MTH	Mth734	B69198	pET100D	Amp^R	Putative pMurE
MFER	Mfer336	WP_013413417.1	pET15b and pET100D	Amp^R	Putative pMurC
MFER	Mfer337	WP_013413418.1	pET15b	Amp^R	Putative pMurD1
MFER	Mfer340	WP_013413421.1	pET15b	Amp^R	Putative pMurD2
MFER	Mfer762	WP_013413839.1	pET15b	Amp^R	Putative pMurE
MFER	Mfer1205	WP_013296078.1	pET15b	Amp^R	Putative pMurF
MBB	Mbb918 (1-500	(Wolf, 2010)	pET151D	Amp^R	Putative pMurD1
MBB	Mbb1268	WP_013296078.1	pET15b and pET101D	Amp^R	Putative pMurF

2.1.5 Expression vectors

Vectors used for expression of the methanogenic pseudomurein peptide ligases in *E. coli* used during this work are listed in Table 2.2.

Table 2.2 List of expression vectors used to clone the gene of interest.

Name	**Details**	**Selection**	**Supplier**
pET100D	TOPO cloning with N-terminal 6xHis-tag, T7 promoter, IPTG inducer	Ampicillin (Amp^R)	Invitrogen® Cat# K100-01
pET101D	TOPO cloning with C-terminal 6xHis tag, T7 promoter, IPTG inducer	Amp^R	Invitrogen® Cat# K101-01
pET151D	TOPO cloning with N-terminal 6xHis tag, T7 promoter, IPTG inducer	Amp^R	Invitrogen® Cat# K151-01
pET15b	N-terminal 6xHis tag, T7 lac promoter, IPTG inducer	Amp^R	Genscript

2.1.6 Crystallisation supplies and screens

Most of the protein crystallisation experiments were performed manually using multi-channel pipettes to replicate the crystallisation screens into the 96-well plates and a multi-dispensing Hamilton glass syringe to dispense 1.0 µL of protein during each ejection (50 µL volume capacity). Automated setup of crystallisation screens were performed with a Mosquito® crystal liquid handling robot (TTP Labtech, Cambridge, UK). The fourth generation MRC 96-well 2 drop UV crystallisation plates for high-throughput screening and ClearVue sheets to seal the plates were obtained from Molecular Dimensions and used for both manual and robotic experiments. Crystallisation screens Structure Screen 1 + 2 HT-96

(SS1+2), JCSG-plus™ (JCSG+), ShotGun™ (SG1), MIDAS™, MORPHEUS® and MORPHEUS II® were also obtained from Molecular Dimensions (UK). In addition, crystallisation screens INDEX, PEG/Ion™ and SaltRx HT and the optimisation screens Additive screen, Silver Bullets Bio and Silver Bullets were obtained from Hampton Research (USA). The seed beads for microseeding, crystal mounts, cryo loops, cryo tools, pre-crystallisation test kit and Izit crystal dye were obtained from Hampton Research (USA).

2.1.7 Cryo-protectants

Three different cryo-protectant solutions, 25% (v/v) glycerol and ethylene glycol and perfluoropolyether cryo oil (PFO), were used to harvest the protein crystals. Both 25% (v/v) glycerol and ethylene glycol were prepared using the mother liquor that produced the protein crystals. In addition, the crystallisation conditions that contained more than 30% (v/v) precipitant were also directly tested as a cryo-protectant. In general, it was observed that 25% (v/v) glycerol seemed to be the best cryo-protectant for both Mfer762 and Mth734. During the course of experiments, 25% (v/v) glycerol was used as a cryo-protectant for the crystals during soaking experiments and especially when there were not many spare crystals.

2.2 Methods and methodology

Multiple approaches of bioinformatics, biochemistry, molecular biology and structural biology methods were implemented to understand and achieve the research goals. Bioinformatics tools were used to study the relevant gene clusters, homologous genes of interest and phylogenetic relationships. The archaeal genes that are homologous to murein peptide ligase genes were identified and analysed further. Molecular and biochemical techniques were used to clone, express and purify the targeted proteins. X-ray crystallography was implemented as the main technique to obtain structural details of the

proteins. The purified proteins were used for crystallisation experiments and the obtained crystals were further used for structure determination experiments. Structural details obtained from the high resolution enzyme structures were used for function prediction and analysing the structural homology with bacterial murein peptide ligases. In addition, biochemical testing was performed using purified pMurE ligase and the bacterial substrate (UDP-MurNAc-Ala) to see if the substrate was recognised by archaeal enzyme. The detailed research techniques and experimental procedures described above are elaborated in the sections below.

2.2.1 Bioinformatics analyses

2.2.1.1 Homologous gene search

The amino acid sequences for murein peptide ligases of *Bacillus subtillis* were initially used to identify the homologous genes in pseudomurein-containing methanogens (*Methanobacteriales* and *Methanopyrales*) by implementing the highly sensitive PSI-BLAST algorithm (Altschul *et al.*, 1997). *B. subtillis* genes were selected from prominent species among the Firmicutes, a phylum of *Bacteria*, which are deeply rooted in the phylogenetic tree of life to have association with *Archaea* (Lake & Sinsheimer, 2013). Among the list of genes in different pseudomurein-containing methanogens, the putative pseudomurein biosynthesis genes from *M. thermautotrophicus* strain ΔH (Mth) were further used as query sequences to identify and analyse homologous genes in other pseudomurein-containing organisms. The search was extended to include most of the available pseudomurein-containing methanogens whose genome sequences have been published. The list of putative genes that are involved in pseudomurein cross-linking peptide formation

(pMurC-F) that were identified and are presented in Figure 6.2. The list of organisms used for the analysis are listed in Appendix 2.

2.2.1.2 Gene cluster analysis

The GenBank database was used to extract the genome files of pseudomurein-containing methanogens used for analysis during this project. The Artemis programme (Rutherford *et al.*, 2000) was used to visualise genome sequences in a graphical format and analyse the location and the neighbouring genes of the putative pseudomurein peptide ligases (pMurC-F) identified from the homologous gene searches. Genes located close to the putative pseudomurein peptide ligase genes, especially those that were not annotated (i.e. annotated as hypothetical proteins), were analysed for their occurrence in other pseudomurein-containing organisms and to relate their presence to potential involvement in pseudomurein biosynthesis. The amino acid sequences of these translated genes were used as query sequences for PSI-BLAST to examine their presence in taxonomic lineages.

In addition, the KEGG database was primarily used to gather information regarding protein sequence, their predicted involvement in pathways and possible functions. KEGG SSDB database was also used to study orthologs, paralogs, conserved gene clusters and sequence motifs. Further, STRING, the online bioinformatics tool (Franceschini *et al.*, 2013), was useful in identifying other genes that were present in the genomic neighbourhood, their known and predicted interactions, and probability of co-expression, co-occurrence and fusion of the gene type with the predicted functional partners in predicted networks. Another online bioinformatics tool, SYNTAX (Oberto, 2013), was used to search for orthologous genes in other pseudomurein-containing methanogens and to check if any neighbouring genes were conserved.

2.2.1.3 Constructing phylogenetic trees

2.2.1.3.1 Sequence searches

The amino acid sequences for the bacterial murein ligases enzymes were obtained using the KEGG database (Tanabe & Kanehisa, 2012). The list of organisms used for analysis are listed in Appendix 2. The amino acid sequences were retrieved using the peptidoglycan biosynthesis pathway of selected organisms and selecting the murein peptide ligases (MurC-F) from the KEGG database (Tanabe & Kanehisa, 2012). In addition, the amino acid sequences of putative pseudomurein peptide ligases (pMurC-F), as explained in Section 2.2.1.1, were retrieved from the NCBI protein database (Database resources of the National Center for Biotechnology Information, 2015).

2.2.1.3.2 Alignments

The multiple sequence alignments used to build the phylogenetic trees were performed using the MUSCLE algorithm (Edgar, 2004) in MEGA7 (Kumar *et al.*, 2016). In addition, multiple sequence alignments together with the available structures were performed using the PROMALS3D multiple sequence and structure alignment server (Pei & Grishin, 2014).

2.2.1.3.3 Generating phylogenetic trees

The alignments obtained from either the MUSCLE algorithm within MEGA or the PROMALS3D server were opened with MEGA7 and used to construct phylogenetic trees. The phylogenetic trees were generated using the maximum likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992). The robustness of the phylogeny analyses was assessed using the bootstrap method and used 100 replicates.

2.2.2 Cloning, expression and purification

The pseudomurein peptide ligase genes from selected organisms identified from the bioinformatics analysis were cloned into a suitable vector (pET151D or pET100D) as listed in Table 2.1. Alternatively, some of gene targets were *de novo* synthesised (GenScript, USA). Plasmid DNA from sequence-verified clones were transformed into appropriate competent cell lines and expressed and purified in sufficient quantity from multi-litre cultures to perform, crystal screening, and optimisation and enzyme characterisation experiments.

2.2.2.1 Transformation

All the plasmids, both acquired from the previous work and synthesized commercially, that were used during this project were confirmed by DNA sequencing to ensure that the gene inserts were correct. The sequence-verified plasmids were then used for transformation of target genes into OneShot® BL21 Star™ (DE3) (Invitrogen) competent cells. The chemical transformation method was used during this project. A vial of competent cells stored at -80 °C was thawed on ice and then aseptically split into two reactions (25 µL each). An aliquot (2.0 µL) of plasmid (~2.0 ng) was added to the aliquots of the competent BL21 Star™ (DE3) (25 µL). The mixture was incubated on ice for approximately 15 min and then heat-shocked for 1 min at 42 °C without shaking followed by immediate transfer into ice. The mixture was then added with 250 µL SOC medium (Invitrogen), and incubated at 37 °C with shaking at 150 rpm for approximately an hour. The transformed bacterial cell culture was spread onto pre-warmed LB agar plates and incubated at 37 °C overnight. Ampicillin at a concentration of 100 µg/mL was used as a selector in the agar plate.

2.2.2.2 Expression

A single colony of each transformed bacterial cells was inoculated into 10 mL of LB broth with ampicillin at a concentration of 100 µg/mL. The 10 mL inocula were then grown overnight at 37 °C in a shaker at 150 rpm. The overnight grown cultures were then transferred into larger 2.0 litre flasks with 750 mL 2×YT broth containing ampicillin of the same concentration (100 µg/mL). The cultures were grown at 37 °C with shaking at 150 rpm until mid-log phase was attained. The mid-log phase of the culture was determined by measuring the optical density (OD) of the culture at 600 nm between 0.4 and 0.6. Once the cultures reached the mid-log phase, protein expression was induced with 1.0 mM IPTG and then the temperature was reduced to 27 °C and the cultures were grown overnight at 150 rpm.

The cultures were centrifuged at 6000 *g* for 15 min to harvest the cells pellets and growth medium discarded (autoclaved before disposing). The cell pellets were collected into sterile Falcon tubes and frozen at -20 °C.

2×YT medium has been used as an alternative to LB medium during the experiments for effective growth and expression. In addition to LB and 2×YT, Auto Induction (AI) medium as described by Studier (Studier, 2005) was also trialled. The AI medium does not require induction by IPTG and therefore does not require measurement of optical density to identify the mid-log phase of culture growth.

2.2.2.3 Cell lysis

Cell pellets were thawed and refrozen at least three times to enhance cell lysis. Lysis buffer was prepared as shown in Table 2.3. Approximately three times volume of lysis buffer was

added to the cell pellets and the samples were pipetted/vortexed to mix well. The cOmplete™ ULTRA tablets (Roche, USA) were used as a protease inhibitor cocktail during cell lysis. A single tablet was dissolved in 5.0 mL MQ and 40 µL per 1.0 mL lysis buffer used was added, followed by the addition of lysozyme to a concentration of 1.0 mg/mL, with mixing. The cells were incubated on ice with gentle agitating for ~30 minutes minimum, or until obvious lysis had occurred (as indicated by the decreased viscosity of the cell slurry).

Phenylmethylsulfonyl fluoride (PMSF) and β-mercaptoethanol (BME) were also added to a concentration of 1.0 and 2.0 mM, respectively in the mixture. Finally, DNase and RNase were added along with slight agitation to a concentration of 5.0 µg/mL. The mixtures were then incubated at 4 °C with gentle mixing overnight.

The cell debris were removed by centrifugation at 10,000 *g* at 4 °C for 15 min. The supernatants were carefully decanted into a syringe and filter-sterilised using a 0.22 µm filter, and were then used for subsequent protein purification.

Table 2.3 Lysis buffer composition.

Reagent	**Concentration used**
Tris-HCl (pH 7.5)	50 mM
Imidazole	10 mM
Glycerol	10% (v/v)
NaCl	300 mM
Triton X-100	1% (v/v)

2.2.2.4 Protein purification

The clarified cell lysates were subjected to Immobilized Metal Affinity Chromatography (IMAC) using nickel (Ni^{2+}) nitrilotriacetic acid agarose resin (Ni-NTA). 6% Ni-NTA agarose was used to purify the His-tagged proteins using the gravity purification technique.

The nickel-containing affinity resin was obtained from Jena Bioscience (Germany) that contained a 50% aqueous suspension with 30% (v/v) ethanol. The gravity flow columns of 12 mL total capacity were filled with approximately 2.0 mL bed volume of Ni-NTA agarose. The packing of the columns and protein purification during this project were performed at room temperature with the samples being placed on ice before and immediately after elution. The packed columns were first washed with 10 column volumes of MQ water and then equilibrated with five column volumes of Buffer A (Table 2.4). The cell lysates were added to the columns and then allowed to flow through the resin under gravity. The flow through (FT) was collected and re-applied to the column to allow effective binding. The column was then washed with three column volumes of Buffer A to remove unbound protein. The wash step was also repeated three times. Targeted proteins that were bound to the columns were eluted with three column volumes of Buffer B that varies only in the concentration of imidazole (250 mM) compared to Buffer A (Table 2.4) and repeated three times. Eluates were collected separately in Falcon tubes and analysed using SDS-PAGE. The concentration of the proteins were determined using the Bradford protein determination method (Kruger, 1994) as explained in Section 2.2.2.5.

Table 2.4 Composition of Buffer A and Buffer B

Reagents	**Concentration**
Tris (pH 7.5)	20 mM
Imidazole	20 mM (250 mM)
Glycerol	10% (v/v)
NaCl	300 mM
BME	2.0 mM

Note: Buffer A and Buffer B were filtered using a 0.22 µm filter before use.

2.2.2.5 Bradford protein determination method

The Bradford protein determination method was used to analyse sample protein concentration by measuring the absorbance at a wavelength of 595 nm. An aliquot (1.0 mL) of Bradford reagent (0.01% (w/v) Coomassie® Brilliant Blue G-250, 4.75% (v/v) ethanol, 8.5% (w/v) phosphoric acid, (filtered through Whatman #1 paper) was added with an appropriate volume of sample, until the colour changed blue. MQ water was then added so as to prepare a total added sample volume of 100 µL. The spectrometer reading was brought to zero using 1.0 mL Bradford reagent and 100 µL of MQ. The absorbance values were used to calculate the final concentration of the protein using the standard calibration curve generated using bovine serum albumin (BSA). The final protein concentrations were calculated with Microsoft Excel using the Bradford calibration curve (Figure 2.1). If the absorbance was higher than 0.7 the samples were re-analysed using less sample.

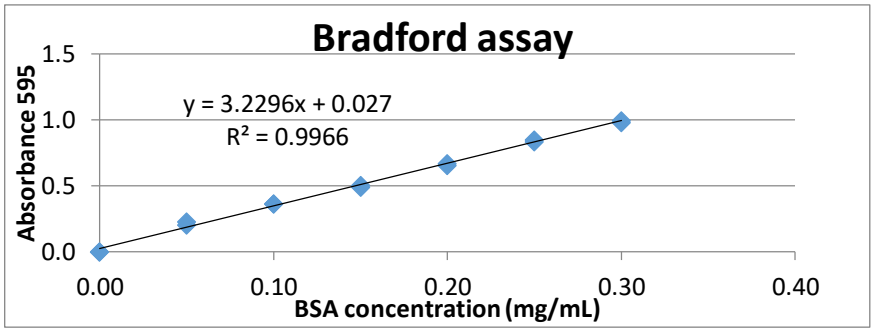


Figure 2.1. Bradford calibration curve. The y-axis represents absorbance at 595 nm, the x-axis indicates BSA concentrations. A trend-line was drawn and the equation obtained was used for calculating protein concentration.

2.2.2.6 SDS-PAGE

For every purification step, SDS-PAGE was performed to analyse the results. BIO-RAD ® low range protein standard was used to determine and verify the molecular masses of the

purified proteins. Total protein (TP) (before cell lysis), pellet (after cell lysis), supernatant of lysate, flow-through (FT) from the Ni^{2+}-column, wash from the Ni^{2+}-column and eluted protein samples were analysed by SDS-PAGE to check the level of expression and purity of the proteins. Standard, TP, pellet, supernatant and FT were diluted with MQ water to make the final volume of 10 µL as shown below (Table 10) and 5.0 µL of gel loading buffer (GLB) was added. The GLB was prepared as specified in Table 2.5. Wash and eluent fractions were run through the gel as collected whereas for concentrated protein samples an appropriate volume of MQ was added to dilute the samples to approximately 2.0 mg/mL and avoid over-loading of the sample. The samples were then heated at 95 °C for 6.0 min to allow for denaturation of the proteins.

The denatured samples were loaded on to a 4% (w/v) stacking and 12% (w/v) running precast gel (BIO-RAD Mini-PROTEAN® TGX™, 12%, 10-wells, Cat# 456-1043) and run in 1× SDS-PAGE running buffer (0.144% (w/v) glycine, 1.9 mM Tris base, 0.01% (w/v) SDS). The gel was run at 120 V for 1 hour (until the dye front reached the bottom of the gel).

Table 2.5 Composition of gel loading buffer.

Reagent	Final Concentration
0.5 M Tris HCl, pH 6.8	41.5 mM
Glycerol	5% (v/v)
20% SDS	8% (w/v)
EDTA	10 mM
Bromophenol Blue	0.04%

2.2.2.7 Desalting and concentrating protein

Desalting was carried out by dialysis using SnakeSkin™ dialysis tubing membrane, 3.5 kilo-Dalton (kDa) molecular weight cut off (MWCO), obtained from ThermoFisher scientific

(NZ). An appropriate length of the membrane was taken, one end of the membrane was sealed with dialysis tubing clips and the protein samples were loaded and sealed. The dialysis was carried out at 4 °C for at least 6-8 hours. The storage buffer contained 20 mM MOPS buffer at pH 7.0 and 2.0 mM TCEP in common, along with a varying concentration of KCl. The concentration of KCl used ranged from 50 mM to 600 mM depending on protein stability, and has been specified within the respective result sections. The ultrafiltration method using Amicon membrane filters, MWCO 10 kDa, was used to concentrate the proteins to a concentration suitable for crystallisation experiments.

2.2.2.8 Molecular mass determination

The native molecular masses of purified proteins were determined by gel filtration chromatography (BioRad BioLogic LP). A standard curve was generated using standard proteins (Sigma-Aldrich) including aprotinin (6.5 kDa), cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (150 kDa) and β-amylase (200 kDa). A filtered sample of the enzyme (250 µL) at a concentration of 2.5 mg/mL was applied to a Superdex 200 column (1 x 59 cm) using a 0.5 mL loop. The proteins were eluted with a buffer (20 mM MOPS pH 7.0, 2 mM TCEP, 0.2 M KCl) at a flow rate of 0.6 mL/min. The molecular mass was estimated from a standard curve prepared by plotting the elution time versus the log of the molecular mass of the standards. The experiments were performed on the same column with a constant flow rate.

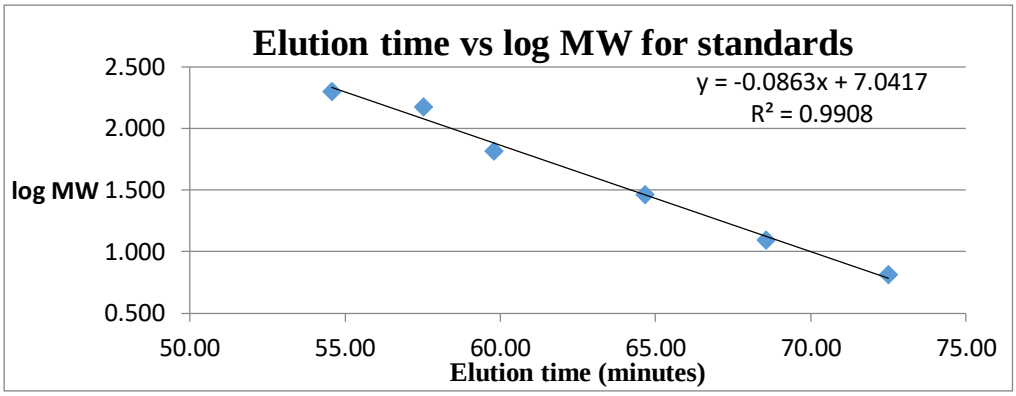


Figure 2.2 Standard curve to calculate molecular mass. The y-axis represents log molecular weight, the x-axis indicates elution time. A trend-line was drawn and the equation obtained was used for determining molecular mass. Proteins used as standards are aprotinin (6.5 kDa), carbonic anhydrase (29 kDa), alcohol dehydrogenase (150 kDa), cytochrome C (12.4 kDa), BSA (66 kDa) and β-amylase (200 kDa).

2.2.3 Methanogen peptide ligase structure determination

This project aimed at determining the 3D structures of putative archaeal pseudomurein peptide ligases (pMur). Macromolecular X-ray crystallography has been used during this project to determine the structures of three proposed pseudomurein peptide ligases. Obtaining good quality diffraction crystal is the bottle-neck of this method, therefore, crystallisation experiments were conducted on multiple members of the same pMur type. Thus, a major aim was to obtain good diffracting protein crystals for each pseudomurein peptide ligase. The list of proteins attempted for structural determination are summarised in Appendix 3. The production of good diffracting protein crystals requires highly purified homogeneous proteins, which was achieved during the project as described earlier in Section 2.2.2.4.

2.2.3.1 Crystallisation of proteins

The purified proteins obtained from Ni-affinity gravity columns (Section 2.2.2.4) were analysed using SDS-PAGE (Section 2.2.2.6) before setting up the crystallisation screens. The purified proteins were filtered through a 0.22 µm ultrafree-MC micro-centrifuge filter using the benchtop centrifuge to ensure that no insoluble materials were present in the crystallising protein samples. Homogeneous proteins of appropriate concentrations (> 8.0 mg/mL for protein size greater than 40 kDa) were used to set up the initial screens. All the primary screens specified earlier in Section 2.1.6 were used to examine the crystallisation condition for each protein sample. The sitting drop vapour diffusion method was used as a primary screening method.

2.2.3.1.1 Sitting drop method

MRC 96-well 2 drop UV plates were used for setting up the sitting drop experiments. Aliquots (75 µL) of each condition from initial screens were replicated into the reservoir compartments of 96-well plates using a multi-channel pipette. An aliquot (1.0 µL) of each target protein was pipetted into the sitting drop compartment using a Hamilton multi-dispensing pipette and mixed with 1.0 µL of crystallising condition which was added directly from each respective reservoir well. The crystallisation plates were sealed with transparent sealing ClearVue sheets using a sealing applicator to increase sheet-to-surface contact. The sealing sheet applicator provides maximum sealing that prevents evaporation and possible cross-contamination. 24-well plates were used for crystal optimisation in which the drop size was doubled (i.e. 2.0 µL of protein was mixed with 2.0 µL of crystallisation buffer). The crystallisation plates were incubated in vibration-free incubator at 21 °C and observed under polarised microscope, typically every day for the first week and every alternative day for a

month. Promising conditions were noted and optimisation experiments were performed in attempts to obtain single large crystals for each target protein.

2.2.3.1.2 Hanging drop method

The hanging drop method was used as an optimising strategy for obtaining larger crystals. In most cases, the method was used to replicate the crystallisation condition identified in 96-well sitting drop experiments. Siliconised glass cover slides were used for the preparation of drops that helps to maintain the spherical shape of the droplet. An equal proportion of protein to precipitant solution method was used as a standard protocol. In cases where crystals were not observed, the protein to precipitant solution ratio were altered. A spherical shape of the crystallisation drops was aimed as it can improve crystallisation of proteins and has an effect on vaporisation behaviour, and thereby, could affect nucleation and consequently crystal size (Sobac *et al.*, 2015). The cover slides with the droplets were placed upside down over a depression in a tray which was partially filled with 800 µL of precipitant solution. The cover slides were sealed using silicone grease around the surface of each wells before placing the cover slides to prevent evaporation of the precipitant solution.

2.2.3.1.3 Randomised microseeding

Crystals which could not be further improved by optimisation of initial crystal conditions were also, in some cases, used as a seed stock for microseeding experiments. The randomised microseeding method was used where the seed stocks were used together with protein samples on to a random screen (Till *et al.*, 2013). This method takes crystals from any promising condition and extends the seeds to other crystal conditions so as to obtain better crystals.

As a part of this method the crystals were crushed with a probe to make sure they were not salt crystals as well as to mix the crystals in the reservoir solution so that fine seeds could be obtained. The additional reservoir solution was dispensed and sucked back to transfer it into the seed bead tube, keeping it on ice. The process was repeated until the well was clear with no crystal debris remaining. The seed bead tube was vortexed for two minutes with approximately 60 µL of reservoir solution, stopping every 30 seconds to cool the tube on ice. Thus produced seed stock was used to screen the crystal on to a random screen by using 0.3 µL of protein, 0.2 µL of reservoir, and 0.1 µL of seed stock with robotic dispensing. In addition, increased volumes of 1.5 µL of protein, 1.0 µL of reservoir, and 0.5 µL of seed stock were used while setting up the plate manually. The seed stocks were agitated to re-suspend any crystals that may have settled in the tubes. The seed stocks were prepared immediately before the crystallisation experiments.

2.2.3.2 Co-crystallisation experiments

All the pseudomurein peptide ligases are indicated to use adenosine triphosphate (ATP) for their activity as noticed from the seven step *in vitro* pentapeptide synthesis experiment (Hartmann & König, 1994). Hence, all the protein targets used for crystallisation experiments were also tested for co-crystallisation in the presence of ATP. Any crystal conditions that produced promising crystals were also repeated in the presence of ATP to check if the crystals were repeated in the corresponding condition. Co-crystallisation experiments in the presence of ATP were also attempted as an optimisation strategy to obtain better diffracting crystals. The concentration of ATP used in all the co-crystallisation experiments was four-fold by molarity of the protein concentration to enable effective binding. Magnesium chloride ($MgCl_2$) was used as a source of Mg^{2+} which is typically the cation found in ATP binding sites. Mg^{2+} has also been studied in bacterial murein peptide

ligases and is bound to the gamma phosphate group of ATP molecule (Bertrand *et al.*, 1999; Dementin *et al.*, 2001). The protein mixture that used ATP as a substrate for co-crystallisation also contained $MgCl_2$ to an approximately same molar concentration as that of the ATP.

The ATP was added to the protein followed by $MgCl_2$, the sample mixed well and incubated on ice for about an hour before adding to the crystallisation plate. The substitution of ATP by adenosine diphosphate (ADP) and the non-hydrolysable analogue of ATP (AMP-PNP) was also tested at later stages with the pMurE type pseudomurein ligases and is described in Chapter 3. After analysis of the first structure of Mfer762, it was observed that presumed phosphate molecules (PO_4^{3-}) were present in the crystal structure at the proposed ATP binding site from structural comparison analysis with bacterial structures. Hence, non-hydrolysable ATP (AMP-PNP) was assumed to be an effective non-hydrolysable substrate for structure determination purposes.

Co-crystallisation experiments using a common chemical group to the murein and pseudomurein peptide ligases (UDP), dipeptides (γ-Glu-Ɛ-Lys), possible amino acids that the enzymes could potentially bind (L-Ala, L-Glu, L-Lys) were also tested. In addition to the co-crystallisation experiments, all the aforementioned putative substrates along with bacterial murein ligase substrate (UDP-MurNAc-Ala) were used in soaking experiments.

2.2.3.3 Crystal testing

Crushing crystals with a sharp object can be used to test if the obtained crystal is protein or salt. A cracking sound on application of slight force and dispersion of crystals far apart from each other displays the characteristics of a salt crystal whereas protein crystals typically break easily without sound and behave more like an ordered gel (Krengel & Imberty, 2007).

In addition, to test if the obtained crystals were protein or salt, the well was cut open through the ClearVue cover sheet and crystals were also on occasion stained with Izit crystal dye to see if there was a change in colour of the crystals. Protein crystals usually absorb the blue colour of the dye quickly and turn dark blue with the addition of dye, whereas salt crystals do not (Krengel & Imberty, 2007). However, the best method to confirm whether a crystal is salt or protein is to test the crystals by X-ray diffraction and analyse the resulting patterns. The In-house Rigaku MicroMax-007 X-ray generator available at Massey University was used to test some crystals and to ensure that salt crystallising conditions were not optimised. The available detectors at the Massey University, both Spider and R-AxisIV++, were used during the entire period of the project to test crystals. The crystal testing was performed at 120 K using the crystals, both straight from the drops, as well as those stored in liquid nitrogen at the In-house X-ray source.

2.2.3.4 X-ray diffraction and data collection

The protein crystals were looped from the original drops and quickly run through cryo-protectant solutions as listed in Section 2.1.7 using a nylon cryo-loop, and snap-frozen using liquid nitrogen. The frozen crystals were stored either in canes or a 96-position cassette. The crystals were mounted onto a goniometer manually at the in-house X-ray source, whereas the sample mounting robot was used at the synchrotron beamline source (Australia) to expose the crystals to the X-ray source. For the crystals that showed protein diffraction patterns, X-ray diffraction data were collected at MX1 and MX2 beamlines at the Australian Synchrotron, Melbourne.

The X-ray diffraction data at the MX1 beamline was collected using the ADSC Quantum 210r detector. Three structures, Mfer762 apo and UDP-bound and Mfer336 that are

discussed during this thesis were collected using the detector. In addition, the X-ray diffraction data for the structure of Mth734 and one of the Mfer762 structures included in this thesis were collected at MX2 beamline using the EIGER detector (Casanas *et al.*, 2016). The data collection using the ADSC Quantum detectors were performed with 70% attenuation, 0.5° oscillation angle and 360° overall oscillation range. Moreover, with the EIGER detector an oscillation angle of 0.1° was used by default and each of the datasets was collected with an oscillation range of 360° unless otherwise specified. The exposure time and detector distance were varied during the use of both the ADSC Quantum and the EIGER detector depending on the crystal diffraction and data quality. The crystal testing as well as the data collection on both the beamlines (MX1 and MX2) at the synchrotron beamlines were performed on crystals frozen at 100 K.

The X-ray diffraction data were collected at the Australian synchrotron and indexed using XDS software (Kabsch, 2010) while collecting the data. Further, the Aimless program (Evans & Murshudov, 2013) was used for data scaling and the resulting log files were examined to analyse the mosaicity, resolution, completeness and other statistics of the collected datasets. The Ctruncate program (Evans, 2011) was used to ensure unique data and to generate a FreeR data subset for 5% of the data that could represent the full dataset with respect to the distribution of structure factor amplitudes and the distribution of reflection resolutions, which will be used for unbiased refinement. The Matthews' coefficient program (Kantardjieff & Rupp, 2003) was used to calculate the solvent content and indicate the number of molecules in the asymmetric unit.

2.2.3.5 Phasing, model building and refinement

All the pseudomurein peptide ligases have sequence identities less than 30% compared to their respective bacterial murein ligases suggesting that it would be challenging to solve the structures using molecular replacement. However, the sequence alignment of pseudomurein peptide ligases with murein peptide ligases reveal that the middle domain is the most conserved of the three domains. The molecular replacement tools were successful in solving the conserved middle domain of the pseudomurein peptide ligases. The molecular replacement tools including MrBUMP (Keegan & Winn, 2008), BABLES and MORDA, were accessed online using the CCP4 online server. The output files were downloaded and monitored individually to select the best solution based on the model to electron density map correlation. The partial solutions from the molecular replacement tools were used for C-α tracing and density modification using SHELXE (Thorn & Sheldrick, 2013) in the CCP4i suite (Winn *et al.*, 2011). The application of the SHELXE program is limited to favourable combinations of resolution range (better than 2.1 Å) and solvent content. The output files from SHELXE were used as an input for automated model building based on the amino acid sequence using the automated model building program Buccaneer (Cowtan, 2006). The model building using Buccaneer was performed based on molecular replacement phases and used to extend the partial model obtained from SHELXE. Fortunately, the crystals for Mfer762 diffracted to a high resolution of 1.7 Å resulting in successful auto-building, ultimately resulting in complete model building for Mfer762.

In addition to the CCP4i suite, the Phenix software (Adams *et al.*, 2010) and the tools inbuilt within the program suite have been also used during this project. AutoBuild program (Terwilliger *et al.*, 2008) has been used for automatic model building and refinement, Phaser-MR (He *et al.*, 2016) and MRage-automated pipeline (Bunkoczi *et al.*, 2013) were

also tested for molecular replacement. The refinement program REFMAC5 (Murshudov *et al.*, 2011) in CCP4i and phenix.refine (Afonine *et al.*, 2012) in Phenix were used for structure refinement. Visualisation and rebuilding of the refined models were performed using COOT (Emsley *et al.*, 2010). The COOT program was also used for manual model building based on electron density in cases where the final models from auto-building programs were not complete. The molecular graphics programs Chimera (Pettersen *et al.*, 2004; McNicholas *et al.*, 2011), CCP4MG (McNicholas, *et al.*, 2011) and PyMOL (Schrodinger, 2015) have been used in addition to COOT for visualisation and comparative analyses. The molecular graphics images used in this thesis were prepared using the aforementioned three programs.

2.2.3.6 Structure validation and analysis

The quality of the structures were initially validated using the validation tools in COOT. The MolProbity (Davis *et al.*, 2004) program inbuilt within the Phenix software was also used to identify and resolve the problematic region such as clashes, basic geometry and protein restraints. In addition, the MolProbity validation web server used for protein structure deposition was used to analyse the structure quality and refinement statistics. PDBePISA (Paxman & Heras, 2017) was used to analyse the protein structure interface and assemblies. The PDBsum (Laskowski, 2009) web server was used to analyse the topology of the protein structure, particularly the protein secondary structure arrangement motifs and amino acids contributing to those structures, protein-substrate interactions, cavities or grooves formed by the structure and their volumes. All the available structures in the PDB database were searched for structural similarity to the newly-determined structures using the Dali web server (Holm & Rosenström, 2010).

2.2.4 Structure-based sequence analysis and phylogenetic analyses

The structure of a protein is more conserved than its sequence (Kim & Lee, 2007); hence, the experimental 3D structures obtained during the project have been used to make comparative analysis of the hypothetical pseudomurein peptide ligases with murein peptide ligases. Structure-based phylogenetic comparisons have been carried out in an attempt to help elucidate the ancient linkages and test hypotheses of early evolution of cell wall biosynthesis.

The amino acid sequences for each type of pseudomurein peptide ligases were extracted from genome sequences as explained in Section 2.2.1.1. For uniform comparison and effective analysis, amino acid sequences of homologous murein peptide ligases from representative organisms in the two super-kingdoms, *Bacteria* (both Gram positive and negative bacteria) and *Archaea* (in addition to pseudomurein-containing methanogens), were selected and included in list of the organisms, as specified in Appendix 2. The structures of the available bacterial murein ligases of the respective types were downloaded from the Protein Data Bank (PDB) (Berman, *et al.*, 2000). In addition, the structures of the recently solved pseudomurein peptide ligases (pMurC and pMurE) from this thesis were also used for the analysis.

All the available structures and sequence files of each pseudomurein peptide ligase types and their homologues in bacteria were entered into the PROMALS3D multiple sequence and structure alignment server (Pei & Grishin, 2014). The alignment file obtained from the server was downloaded and visualised using MEGA 7 (Kumar, *et al.*, 2016). MEGA 7 was also used to generate phylogenetic trees. The Neighbour-Joining and Maximum likelihood algorithms were tested for the bootstrap analysis to validate branching topologies.

2.2.5 Biochemical assay

2.2.5.1 ADP quantification assay

The specificity of the pseudomurein peptide ligases towards the bacterial murein peptide ligase substrate UDP-MurNAc-alanine (UMA) was tested using the purified Mfer762 pMurE enzyme. A spectrophotometric method was used to measure the amount of ADP released during the reaction (R1 below, N^{α}-UDP-Glu^{γ}-Ala substituted by UMA) by using the ADP quantification protocol as provided by Sigma-Aldrich (NZ) and explained by Malhotra *et al.* (Malhotra *et al.*, 1996). The rate of reactions catalysed by lactate dehydrogenase (LDH) is equal to that of the coupled reaction (R2). Hence, the amount of NADH oxidised to NAD^+ in the LDH catalysed reaction (R3) is proportional to the amount of pyruvate released by the pyruvate kinase (PK) catalysed reaction (R2). The magnitude of the decrease in absorbance at 340 nm correlates to the oxidation of NADH, which represents the concentration of pyruvate formation by the PK catalysed reaction (R2) and therefore indicates the amount of ADP produced in reaction R1.

$$N^{\alpha}\text{-UDP-}Glu^{\gamma}\text{-Ala+ATP+Lys+}MgCl_2 \xrightarrow{pMurE} N^{\alpha}\text{-UDP-}Glu^{\gamma}\text{-Ala-}^{\varepsilon}\text{Lys + ADP+ Pi}\ldots\ldots\ldots \text{(R1)}$$

$$\text{PEP + ADP} \xrightarrow{PK} \text{Pyruvate + ATP}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(R2)}$$

$$\text{Pyruvate + ß-NADH} \xrightarrow{LDH} \text{Lactate + ß-NAD}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(R3)}$$

Samples with the purified Mfer762 enzyme, putative pMurE, were set up in a final volume of 100 µL containing 4.0 mM Mg^{2+}, 0.6 M KCl, 2.0 mM TCEP, 2.0 mM ATP and 5.0 mM L-lysine (Sigma-Aldrich; NZ) in 50 mM Bis-Tris propane (BTP) buffer (pH 7.5 at 25 °C), 1.0 mM UMA, 0.2 mM NADH (Sigma-Aldrich; NZ), 4 mM phosphoenolpyruvate (PEP) (Sigma-Aldrich; NZ) and 0.02 unit/reaction of PK/LDH enzymes. The mixture without the

Mfer762 enzyme was prepared and incubated at the sample holder at 50 °C for 3 min. The reaction was initiated by the addition of 1.0 mg/mL of purified Mfer762 enzyme and the absorbance was measured at 340 nm using a Cary 100 UV-vis spectrophotometer (Agilent Technologies, USA). Measurements were made using 1.0 cm path length stoppered quartz cuvettes keeping them in a thermostated cuvette holder to maintain the temperature of the reaction mixture.

Furthermore, a stopped reaction for ADP analysis was also performed. During this method, the sample mixture was prepared with 4.0 mM $MgCl_2$, 0.6 M KCl, 2.0 mM TCEP, 2.0 mM ATP and 5.0 mM L-lysine, in 50 mM BTP (pH 7.5 at 25 °C), 1.0 mM UMA, 0.2 mM NADH and 1.0 mg/mL Mfer762 enzyme and incubated at 80 °C. Two different experiments, one with 10 min incubation and another with overnight incubation were performed. The mixture after 10 min incubation was brought to 50 °C and then 0.02 units of PK/LDH enzymes per reaction added. The reaction was finally initiated with 0.4 mM PEP and absorbance was observed at 340 nm. The experiment was performed at 37 °C for the mixture that was incubated at 80 °C overnight. The controls in absence of only UMA and both UMA and Mfer762 enzymes were also performed using the same protocol. The reaction mixture were saved for inorganic phosphate assays and therefore stored at -20 °C after measuring the absorbance.

2.2.5.2 Phosphate assay

To quantify the concentration of phosphate in R1, a phosphate assay was performed. A calibration curve was obtained by adding 1.0 mL of different PO_4^{3-} standards (containing 0.04 M H_2SO_4) to 160 µL fresh reaction reagent (Figure 2.3). The reaction solution was

prepared by mixing 30% (v/v) 0.1 M ascorbic acid and 70% (v/v) molybdate reagent (Sigma-Aldrich; USA).

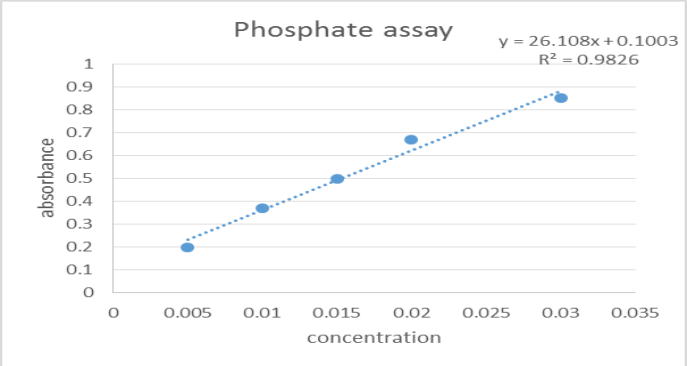


Figure 2.3 Calibration curve for the P_i assay; Y-axis presents the absorbance intensity, X-axis the P_i concentration [mM]; trendline and equation are shown.

The samples that were stored at -20 °C from the ADP quantification assay were thawed on ice and used as samples for the phosphate assay. The samples were set up to a final volume of 200 µL using MilliQ water. The samples were further added with 100 µL of 0.5 M H_2SO_4, followed by 700 µL of MilliQ water and 160 µL of reaction reagent. A blank was prepared with 1.0 mL MilliQ water and 160 µL of reaction reagent. Standards, blank and samples were incubated for 20 min and absorbance was measured at 880 nm.

Chapter 3 Pseudomurein peptide ligase type E

3.1 Introduction

MurE is a class of murein peptide ligase that is responsible for the addition of the third residue to the pentapeptide side-chain of peptidoglycan. This is the only peptide ligase type in bacteria that has specificity towards two different substrates in Gram positive and Gram negative bacteria. In Gram positive bacteria, MurE has specificity towards L-lysine, whereas in Gram negative bacteria it is responsible for the addition of meso-diaminopimelic acid (mDAP), a derivative of lysine, to the pentapeptide cross-linking peptide (Paradis-Bleau *et al.*, 2009). Among all the murein peptide ligase types, MurE plays the most crucial role in the pentapeptide formation as the third residue is involved in glycan cross linkage and addition of an incorrect amino acid can result in morphological changes, or even cell lysis (Mengin-Lecreulx, *et al.*, 1999; Consaul *et al.*, 2005).

Bioinformatics analysis indicated that homologous MurE-encoding genes were present in methanogens and for these to be present only among the pseudomurein-containing methanogens (details in next section). This suggested that the protein encoded by the gene could be involved in pseudomurein biosynthesis, and more specifically, with the addition of L-lysine based on sequence homology and presence of a lysine residue in the cross-linking peptide of pseudomurein. Hence, it is proposed that MurE homologues in pseudomurein-containing methanogens, putative pseudomurein peptide ligase type E (pMurE), could be involved in the addition of L-lysine to the cross-linking pentapeptide chain in the pseudomurein biosynthesis pathway. The sequence identities between bacterial MurEs and pMurEs are only up to 28%, which is not sufficient to support a clear evolutionary history, but is an indication that they might be involved in a similar function. To clarify the relation

between the bacterial MurE and archaeal pMurE the structures of pMurE have been solved for the first time and analysed to predict its function.

3.2 Overview

The structures of pMurE from two different methanogenic archaea (*M. fervidus* (Mfer762) and *M. thermoautotrophicus* ΔH (Mth734)) and their function predictions are presented in this chapter. Three different structures of Mfer762 and one structure of Mth734 have been solved and are presented. Among the three different structures of Mfer762, one is an apo form and two are UDP-bound structures, with UDP-bound in two completely different positions. Each Mfer762 structure determined has been analysed and compared to identify any conformational changes adopted by the enzyme upon UDP binding. Further, the structures of Mfer762 have been compared to the Mth734 structure to examine any differences within the pMurE members in pseudomurein-containing methanogens. The structures of pMurE, particularly their substrate-binding sites, have been compared using homology analysis. In addition, the structural details have been used to study their structural homology with bacterial murein peptide ligases (MurC-F) to determine if the two groups of proteins share a similar structure-function relation through similarity search and analysis of substrate-binding sites. This was undertaken to help establish their evolutionary relationship to one another, an overall aim of the thesis.

3.3 Bioinformatics analysis

3.3.1 Homologous protein search

The amino acid sequence for MurE from *Bacillus subtillis* (WP_003245326.1) was used to search for homologous protein sequences in pseudomurein-containing methanogen using the

highly sensitive PSI-BLAST algorithm (Altschul, *et al.*, 1997). The results obtained from the search were then used to search for homologous proteins within the archaeal superphylum. The amino acid sequence of putative pMurE from *M. thermautotrophicus* (Mth734) (WP_010876373.1) was identified and then used as a query sequence to search for orthologous proteins among *Archaea*. The results showed that genes encoding homologous protein sequences were highly conserved in the pseudomurein-containing methanogens and were not found in other non-pseudomurein-containing methanogens or other members of the *Archaeal* superphylum. This led to a hypothesis that the proteins encoded by the genes are involved in pseudomurein cell wall biosynthesis and were probably homologous in function to MurE in bacterial murein cell wall synthesis.

3.3.2 Gene cluster analysis

The Artemis programme (Rutherford, *et al.*, 2000) was used to visualise the genome sequences of pseudomurein-containing methanogens in a graphical format and analyse the location and neighbouring genes of the putative pseudomurein peptide ligase genes. The KEGG SSDB database indicated that the putative pMurE in pseudomurein-containing methanogens was in a gene cluster with genes encoding D-ala-D-ala ligase and MraY-like enzymes of the bacterial murein biosynthesis pathway. The conserved neighbouring genes in pseudomurein-containing methanogens were also confirmed by implementing the online bioinformatics tools SYNTAX (Oberto, 2013) and STRING (Franceschini, *et al.*, 2013).

The STRING bioinformatics tool was useful to examine the potential interactions of the pMurE-encoding gene with neighbouring genes and their co-occurrence, thereby enhancing information supporting the roles of the neighbouring genes relating to the pMurE's function in pseudomurein peptide biosynthesis (Figure 3.1). It was observed that the pMurEs, and Mth734 in particular, formed a network with other enzymes indicated to be involved in cell

wall synthesis. For example, the Mth735, annotated as a phospho-*N*-acetylmuramoyl-peptapeptide transferase, is located next to Mth734 and could be inferred based on homology to encode a putative MraY in pseudomurein cell wall synthesis. In bacteria, MraY adds the undecaprenyl phosphate to the product of MurF (Mth837 equivalent bacterial protein) and is the first step of the lipid cycle reaction in the biosynthesis of cell wall peptidoglycan (Bouhss *et al.*, 2004). Another protein Mth736, analogous to D-ala-D-ala ligase in murein cell wall synthesis, is found in all pseudomurein-containing methanogens. The D-ala-D-ala ligase in murein cell wall biosynthesis catalyses the formation of D-alanyl-D-alanine with the use of two D-alanine amino acids and an ATP molecule. Similarly, the putative pMurD2 (Mth532) and UDP-*N*-acetylmuramyl tripeptide synthetase-like protein (proposed pMurF, Mth873) were also present in the network (Figure 3.1).

Further, Mth232, homologous to UppS in bacteria, is involved in the sequential condensation of isopentenyl diphosphate (IPP) with geranylgeranyl diphosphate (GGPP) to yield seven diphosphates and tritrans, hepta-cis-undecaprenyl diphosphate, probably the precursor of glycosyl carrier lipids (Guo *et al.*, 2005). Mth590 is predicted to have role in phospho-*N*-acetylmuramoyl-pentapeptide-transferase activity. Mth1676 (FtsZ) encodes an essential cell division protein that forms the contractile ring structure (Z-ring) at the future cell division site (Lutkenhaus, 1993) and is predicted to have functional role associated with Mth734. Mth142 protein is predicted to be an inosine-5'-monophosphate dehydrogenase that catalyses the conversion of 5'-phosphate (IMP) to xanthosine 5'-phosphate (XMP), the first committed and rate-limiting step in the *de novo* synthesis of guanine nucleotides, and therefore plays an important role in the regulation of cell growth. Mth1608 (FtsY) is a signal recognition particle protein (docking protein) involved in targeting and insertion of nascent membrane proteins into the cytoplasmic membrane. It acts as a receptor for the complex formed by the signal recognition particle (SRP) and the ribosome-nascent chain (RNC) (Draycheva *et al.*,

2016). Mth1334 (DapF) is a diaminopimelate epimerase that catalyses the stereoinversion of LL-2,6-diaminoheptanedioate (L,L-DAP) to meso-diaminoheptanedioate (meso-DAP), a precursor of L-lysine (Liechti *et al.*, 2018). The wide range of aforementioned proteins involved in the Mth734 network, except for Mth532, are involved in cellular processes apparently beyond the scope of the cross-linking peptide formation and hence not discussed in detail during the thesis.

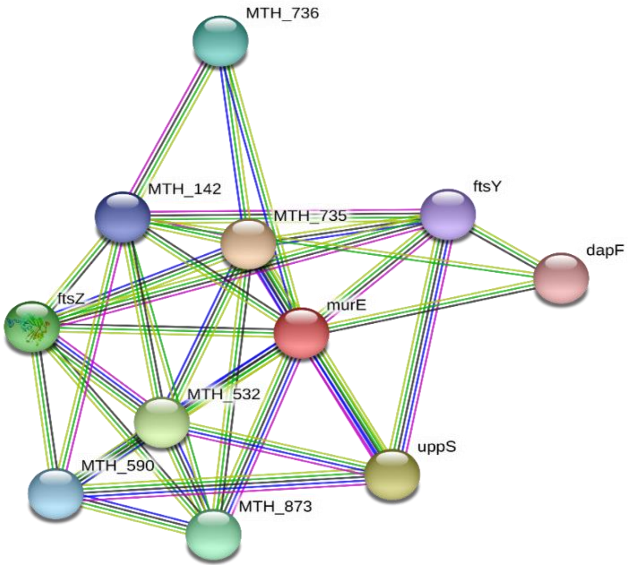


Figure 3.1 Summary view of Mth734 and its network to show potential interactions. The Mth734 (query protein) is annotated as MurE in the figure and indicated by a pink sphere. The lines in green indicates gene neighbourhood, blue indicates gene co-occurrence, magenta indicates interaction that is experimentally determined and light green indicates potential interactions based on text-mining. Each node and the values in the network refer to the proteins putative MraY (Mth735), putative MurF (Mth873), putative D-ala-D-ala ligase (Mth736), pMurD2 (Mth532), undecaprenyl diphosphate (UppS), *N*-acetylglucosamine-1-phosphate transferase (Mth590), signal recognition particle protein, (Mth1608, FtsY), diaminopimelate epimerase, (Mth1334, DapF), inosine-5'-monophosphate

dehydrogenase (Mth142), cell division protein (Mth1676, FtsZ). The figure is obtained from the STRING webserver (Franceschini, *et al.*, 2013).

The SYNTAX bioinformatics tool also indicated a conserved gene cluster for Mth734 as obtained from the KEGG SSDB database and graphically represents the positions of orthologous genes in other pseudomurein-containing methanogens. It was clearly observed from the SYNTAX result that the neighbouring genes encoding for Mth735 and Mth736 were conserved and co-occurred with pMurE genes (Figure 3.2). The SYNTAX result also indicated that Mth873, homologous to MurF in bacteria, is also conserved but is not included in the figure as it lies more distant from the gene cluster. Moreover, the enzyme MurF in methanogens has been recently studied and shown to encode a methanogen co-factor biosynthesis enzyme F_{430} synthase (Zheng, *et al.*, 2016; Moore *et al.*, 2017).

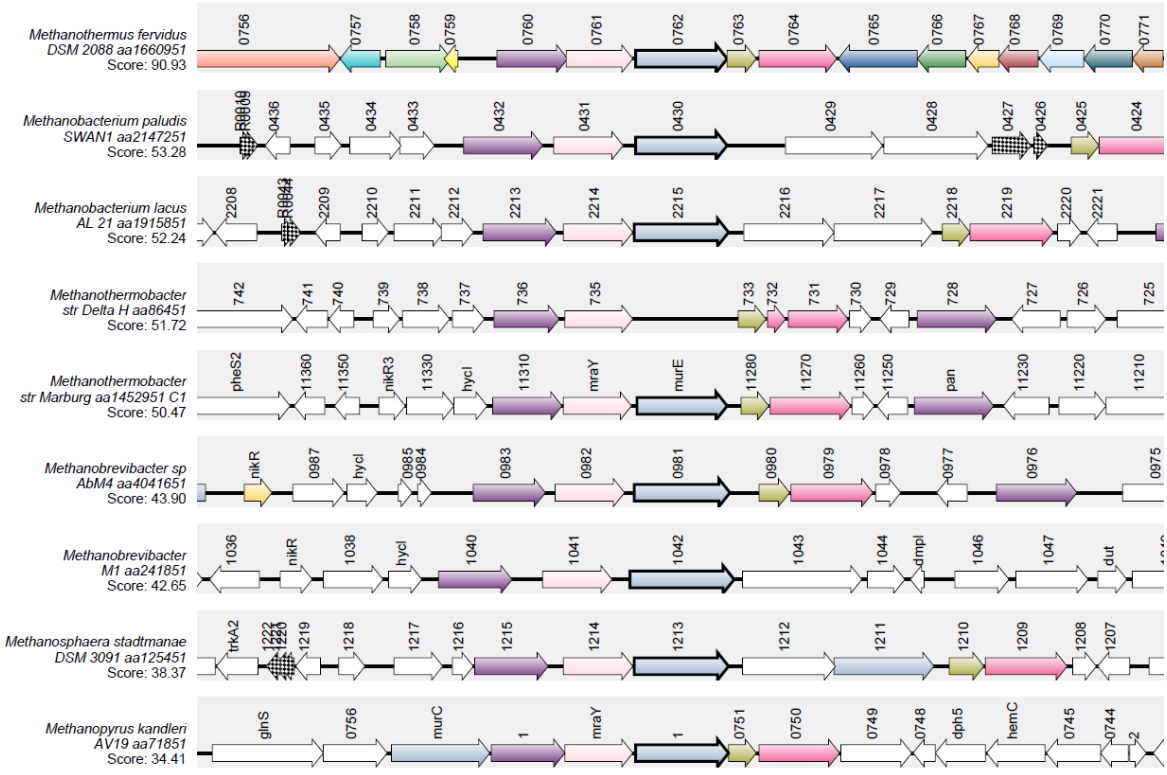


Figure 3.2 SYNTAX result showing the presence of pMurE in representative pseudomurein-containing methanogens. In addition to the pMurE, the orthologues of Mth735 (shown in light pink)

and Mth736 (in purple) are also well conserved and present as a gene cluster among all studied pseudomurein containing methanogens. The Mfer762 (equivalent to Mth734) amino acid sequence was used as query sequence to obtain the alignment. For brevity, only the top nine pseudomurein–containing methanogen gene clusters are shown.

3.3.3 Phylogenetic trees

The amino acid sequences for the bacterial MurE and putative pMurE enzymes were obtained using the KEGG database and protein databases, respectively, as explained in Section 2.2.1.3.1. The amino acid sequences were retrieved from the same source for all the bacterial enzymes and used the peptidoglycan biosynthesis pathway (ko00550) in the KEGG database. The MurE (EC: 6.3.2.13) was selected in each case and sequences were retrieved in *fasta* format for multiple sequence alignment. Both the MurE and pMurE amino acid sequences were used to perform multiple sequence alignment using the MUSCLE algorithm (Edgar, 2004) in MEGA7 (Kumar, *et al.*, 2016). The aligned multiple sequences were used to build the phylogenetic tree based on the Maximum Likelihood method. The phylogeny analysis was assessed using the bootstrap method and used 100 replicates. The phylogenetic tree for MurE and pMurE is shown in Figure 3.3. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones, *et al.*, 1992). Initial tree for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value (Kumar, *et al.*, 2016). Even though all the sequences collected for the list of organisms listed in Appendix 2 were considered for the evolutionary history, the current analysis (Figure 3.3) utilised 83 amino acid sequences and considered 380 positions after elimination of gaps. The information obtained from both the trees were similar. It was observed that both the MurE and pMurE

clade separately in the phylogenetic tree. *Methanopyrus kandleri* roots deeply in the pMurE clade, similarly *Caldicericum exile* and *Thermotaga maritima* root deep in MurE clade. A similar evolutionary history was inferred using the Neighbour joining method assessed using the bootstrap method and used 1000 replicates (tree not shown). The result from both these trees suggested an individual clade of both murein and pseudomurein peptide ligases.

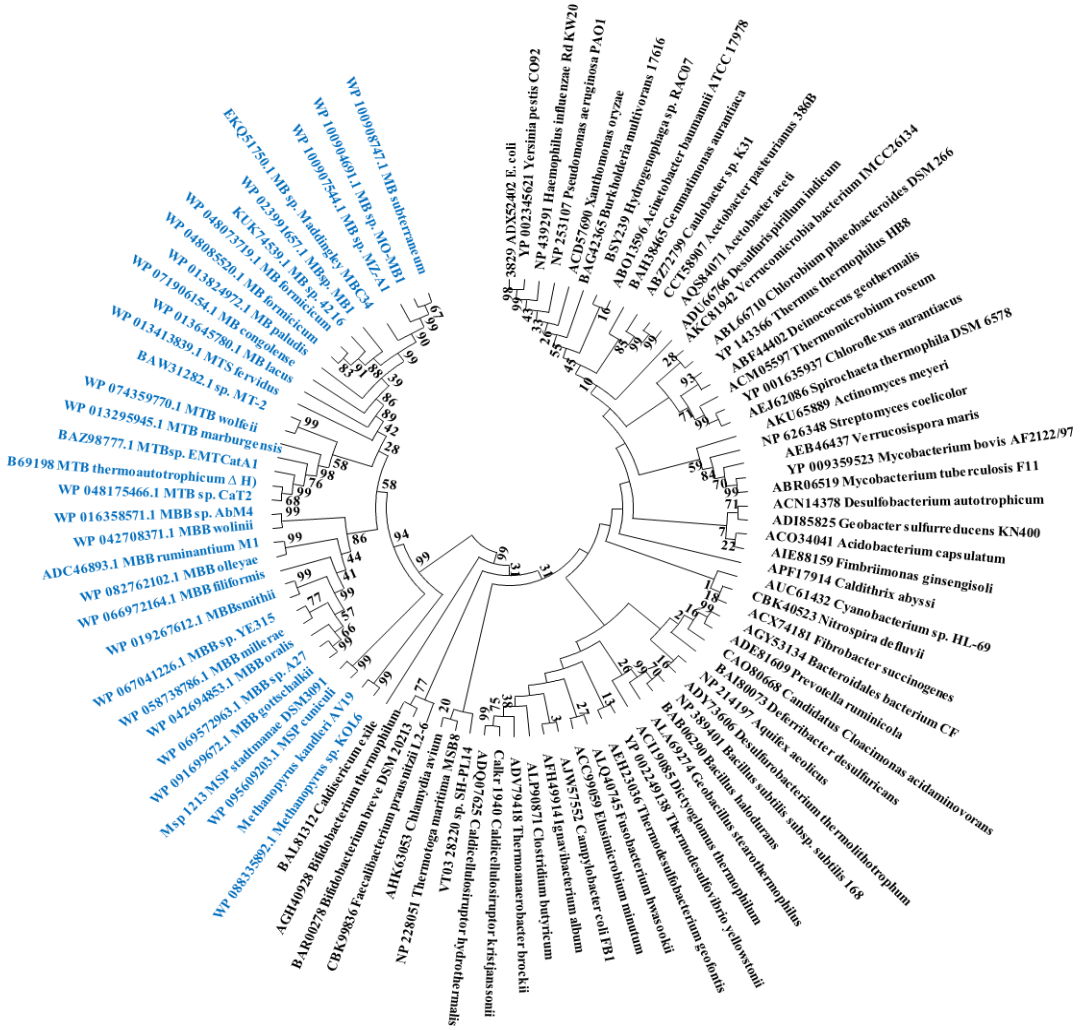


Figure 3.3 Molecular phylogenetic analysis of the murein and pseudomurein peptide ligase type E using the Maximum Likelihood method.

3.4 Cloning, expression and purification

The pMurE from *M. fervidus* (Mfer762) and *M. thermoautotrophicus* ΔH (Mth734) have been the focus of analysis during this project. The Mfer762-gene was synthesised and obtained from GenScript, USA and was cloned into the pET-15b expression vector between the NdeI and BamHI sites and codon-optimised using OptimumGene™ for expression in bacterial cell lines (*E. coli*). The Mth734 gene was cloned into expression vector pET100D and the resulting plasmid was obtained from the previous work by Maximilian Wolf (Wolf, 2010). The plasmids were used for transformations of *E. coli* BL21 Star competent cells (Section 2.2.2.1). The bacterial colonies with transformed plasmids were selected by Amp^R and individual colonies were picked and grown in 10 mL 2×YT medium. Two alternative media, 2×YT and auto-induction (AI) medium, were tested for the growth and expression of *E. coli* cultures. Expression of Mth734 in *E. coli* BL21 star was higher on 2×YT medium compared to AI medium. Hence, 2×YT medium was used for larger volume protein production. Expression of protein in 2×YT medium was carried out as explained in Section 2.2.2.2. His-tagged Mth734 protein was purified using the Ni-NTA gravity column chromatography technique at room temperature. The Mth734 protein eluted from the Ni-resin was analysed by SDS-PAGE against low molecular weight standard from BioRad (Figure 3.4). The protein was dialysed against 20 mM MOPS buffer, 50 mM KCl and 2.0 mM TCEP, pH 7.0. It was noticed that the protein was unstable at 4 °C in the storage buffer. The stability of the protein was finally improved by increasing the KCl concentration to 200 mM in the storage buffer. Two separate additional preparations of the protein were prepared to obtain an adequate amount of purified Mth734 protein for crystal screening and optimisation. Each time the experimental parameters and methodology were kept the same and dialysed directly into storage buffer with 20 mM MOPS buffer, 200 mM KCl and 2.0

mM TCEP, pH 7.0. The purified Mth734 protein from each preparation were concentrated to a final concentration of approximately 10 mg/mL. The concentration of the KCl in the storage buffer for Mfer762 was also kept the same (200 mM).

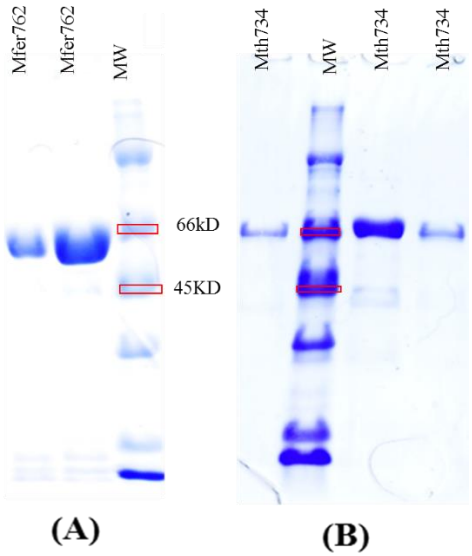


Figure 3.4 SDS-PAGE analysis for the Mfer762 (A) and Mth734 (B). Low molecular weight standard (BioRad) has been used in both gels. Proteins collected at each elution step during the protein purification process were loaded separately in the gels.

3.5 Protein structure determination

The Mfer762 apo structure and structures with the UDP-bound were determined using different crystal forms. UDP, a common chemical group to both the substrate in murein peptide ligases and pseudomurein peptide ligases, was used for co-crystallisation and soaking experiments. The substrate for bacterial murein ligase (MurE) is UDP-MurNAc-L-Ala-D-Glu (Emanuele, *et al.*, 1997), whereas for the putative pseudomurein peptide ligase (pMurE) it is N^{α}-UDP-Glu$^{\gamma}$-Ala (Hartmann & König, 1994). Hence, UDP was used for binding experiments and enabled comparative analysis of the UDP-binding interaction to be performed in two different peptide ligases, and will be referred to as a partial substrate

hereon. The structure determination experiments performed during the project for the pMurE peptide ligase are summarised in Appendix 4 and Appendix 5.

3.5.1 Protein crystallisation

3.5.1.1 Crystallisation of Mfer762

Five primary screens from Molecular Dimensions, UK (SS1+2, JCSG+, SG1, MIDAS and MORPHEUS) and INDEX screen from Hampton Research, USA, as listed in Section 2.1.6, were used for screening with purified Mfer762. Out of 576 crystal conditions tested, only two crystallisation conditions, one each from ShotGun screen (SG1™) and JCSG-plus™ from Molecular Dimensions (UK) produced protein crystals. The mother liquor contained 0.2 M ammonium sulfate and 30% w/v PEG4000, pH 5.4 from the SG1™ screen, hereon referred as SG1_E8, and produced a single crystal in the primary screening plate and required no further optimisation. The crystal obtained in SG1_E8, and used for the Mfer762 apo structure determination, was observed after two weeks from the date of plating and kept growing for four weeks until it was harvested directly from the primary 96-well plate sitting drop. The crystal was looped in a 0.2 mm nylon loop, passed quickly through 25% ethylene glycol prepared in mother liquor and snap-frozen with liquid nitrogen. This crystal diffracted to 1.9 Å and is the first apo structure for Mfer762. The second crystallisation condition, with 3.2 M ammonium sulfate, 0.1 M citrate, pH 5.0, from the JCSG-plus™ screen (JCSG_F2), produced bundled thin-needle shaped crystals and required optimisation to obtain a larger single crystal. The JCSG_F2 crystallisation condition contained a high salt concentration and had no precipitant. Hence, the condition was eventually discarded and instead, crystals from the SG1_E8 condition were used for soaking experiments and data collection.

The Mfer762 crystals in the SG1_E8 condition reproduced with similar morphology and produced good size crystals that fitted within 0.2-0.3 mm nylon loops. The crystal condition also reproduced crystals in co-crystallisation experiments. A common chemical group to the bacterial MurE substrate (UDP), essential cofactors required for Mur family to function (ATP and $MgCl_2$), and putative amino acids that the enzyme could bind (L-Glu and L-Lys), individually, as well as collectively, were used for co-crystallisation experiments. The selection of amino acids (L-Lys and L-Glu) was made based on the enzyme homology with the respective bacterial types, which supports a specific role of pMurE during pentapeptide biosynthesis pathway in pseudomurein-containing methanogens.

Interestingly, the crystals in SG1_E8 condition kept growing for a month but slowly dissolved after 3 months and then re-appeared with a different morphology (flat long plates) in the same drop (Figure 3.5C). The solved co-crystallised crystal datasets did not contain ATP or amino acids in the crystal structure and therefore further soaking experiments were conducted. The first soaking experiment was carried out in the presence of only UDP. A 20 mM UDP solution was prepared using mother liquor from the crystallisation condition. The crystal from the crystallisation drop was looped and dropped into the solution and left for approximately 10 mins. After 10 minutes the crystal was taken from the UDP-containing solution and quickly run through 25% v/v glycerol and snap-frozen in liquid nitrogen. The soaking experiment resulted in the Mfer762 crystal structure with UDP bound to it for the first time. The same experimental procedure was applied to perform soaking experiments using additional substrates. The soaking experiments were also carried out on both morphological protein crystals types (that are similar to crystals as shown in Figure 3.5A and C), using the same substrate concentration. The concentrations of substrates in the solution used for extended substrate soaking experiments were 25 mM ADP/ATP, 25 mM Glu/Lys, and 20 mM UDP.

Further, soaking experiments were also performed using the bacterial substrate UDP-*N*-acetylmuramic acid-Ala (UMA). The experiment was initially attempted with 10 mM UMA but the crystal dissolved in a couple of minutes, hence, a lower concentration (2.0 mM) UMA was used. The crystals were soaked with the 2.0 mM UMA for either 20 minutes or 24 hours and appeared stable during this treatment. The crystals from both the experiments were used for data collection. The soak solutions used in all the soaking experiments were prepared using the mother liquor of the SG1_E8 condition. All the Mfer762 protein structures determined during the project were obtained from the crystal forms that grew up in SG1_E8 condition.

3.5.1.2 Crystallisation of Mth734

Mth734 protein was screened using all the crystallisation screens as listed in Section 2.1.6 at different stages. Five primary screens from Molecular Dimensions, UK (SS1+2, JCSG+, SG1, MIDAS and MORPHEUS) were screened each time for three different experimental preparations of Mth734 protein. Mth734 protein, together with approximately four-fold ATP with 0.2 mM Mg^{2+} (both presence and absence), was used to set up the crystallisation experiments for each screen. The first protein crystals were obtained using the MORPHEUS screen in several crystallisation conditions. The morphology of the crystals from MORPHEUS_F7 and _F9 appeared most promising. Both of these conditions contained 0.12 M monosaccharides mixture (D-glucose; D-mannose; D-galactose; L-fucose; D-xylose; *N*-acetyl-D-glucosamine); F7 contained buffer system 2 (sodium HEPES; MOPS), pH 7.5 whereas F9 contained buffer system 3 (tris (base); bicine), pH 8.5. In addition, the precipitant used in these two conditions was a combination of glycerol and PEG4000 for F7 and PEGMME 550 and PEG20000 for F9. The crystals were harvested directly from the primary screen wells, and crystallisation was also repeated with the exact same conditions.

Optimising the pH and precipitant concentration of the crystallising condition did not improve the crystal size and/or morphologies. All of these crystals were tested at the Australian synchrotron MX1 beamline but none of them diffracted.

Eventually, with the introduction of MORPHEUS II (Gorrec, 2015), a follow-up to the original MORPHEUS screen (Gorrec, 2009), crystallisation of the Mth734 protein was successful. The MORPHEUS II_H3, MORPHEUS II_H7 and MORPHEUS II_H9 conditions produced multiple needle-shaped crystals as shown in Figure 3.5D. All the crystallisation conditions contained 40 mM of polyamine (10 mM each of spermine tetra-hydrochloride, spermidine tri-hydrochloride, 1,4-di-aminobutane di-hydrochloride, DL-ornithine monohydrochloride). Each condition contains 100 mM buffer of MOPSO/bis-tris pH 6.5, BES/TEA pH 7.5 and Gly-Gly/AMPD pH 8.5 for MORPHEUS II_H3, _H7 and _H9, respectively. The precipitant includes 10% w/v PEG8000, 20% v/v 1,5-pentanediol) for MORPHEUS II_H3, 10% w/v PEG8000, 20% v/v 1,5-pentanediol) for MORPHEUS II_H7 and 15% w/v PEG3000, 20% v/v 1,2,4-butanetriol, 1% w/v NDSB 256 for MORPHEUS II_H9. The crystals appeared within 3 weeks of screening and the crystals from the MORPHEUS II_H7 grew quickly and gained the most optimal shape of the three conditions. The crystals from this condition were used to prepare seed stock, explained in Section 2.2.3.1.3 for the randomised microseeding (rMMS) technique (Till, *et al.*, 2013) and screened against SG1 and MORPHEUS II screens. The SG1 screen produced no distinct protein crystals but MORPHEUS II produced similar multi-needle crystals in a wide range of conditions that contained a mixture of monosaccharides (xylitol, myo-inositol, D-fructose, L-rhamnose monohydrate, D-sorbitol), amino acids (DL-arginine hydrochloride, DL-threonine, DL-histidine mono-hydrochloride monohydrate, DL-5-hydroxylysine hydrochloride, trans-4-hydroxy-L-proline) and polyamines (spermine tetra-hydrochloride, spermidine tri-hydrochloride, 1, 4-di-aminobutane di-hydrochloride, DL-ornithine mono-

hydrochloride) within one to four weeks. No new crystal conditions were observed after 4 weeks; however, nice robust crystals were observed after 6 months in the same drops where the multiple needles initially appeared, and with the same nucleation point (Figure 3.5E). The Mth734 crystals from MORPHEUS II_G9 and _G10 were snap-frozen in liquid nitrogen in the presence of different cryo-protectants (25% v/v glycerol or 25% v/v ethylene glycol both prepared in mother liquor or PFO). Among several crystals from the aforementioned crystal conditions only one crystal from the MORPHEUS II_G10 condition (100 mM amino acids mix; 0.1 M Gly-Gly/AMPD buffer, pH 8.5; 50% v/v precipitant mix of 25% w/v PEG4000 and 40% w/v 1, 2, 6-hexanetriol) diffracted to a solvable resolution of 2.7 Å. The crystal was flat plate-shaped and was harvested in the presence of 25% v/v glycerol (Figure 3.5E). The crystallisation condition in MORPHEUS II_G9 contains amino acids rather than polyamines compared to MORPHEUS II_H9 while the rest of the condition composition remained the same. Summary of crystallisation experiments performed to obtain the good diffracting pMurE crystals is included in Appendix 4.

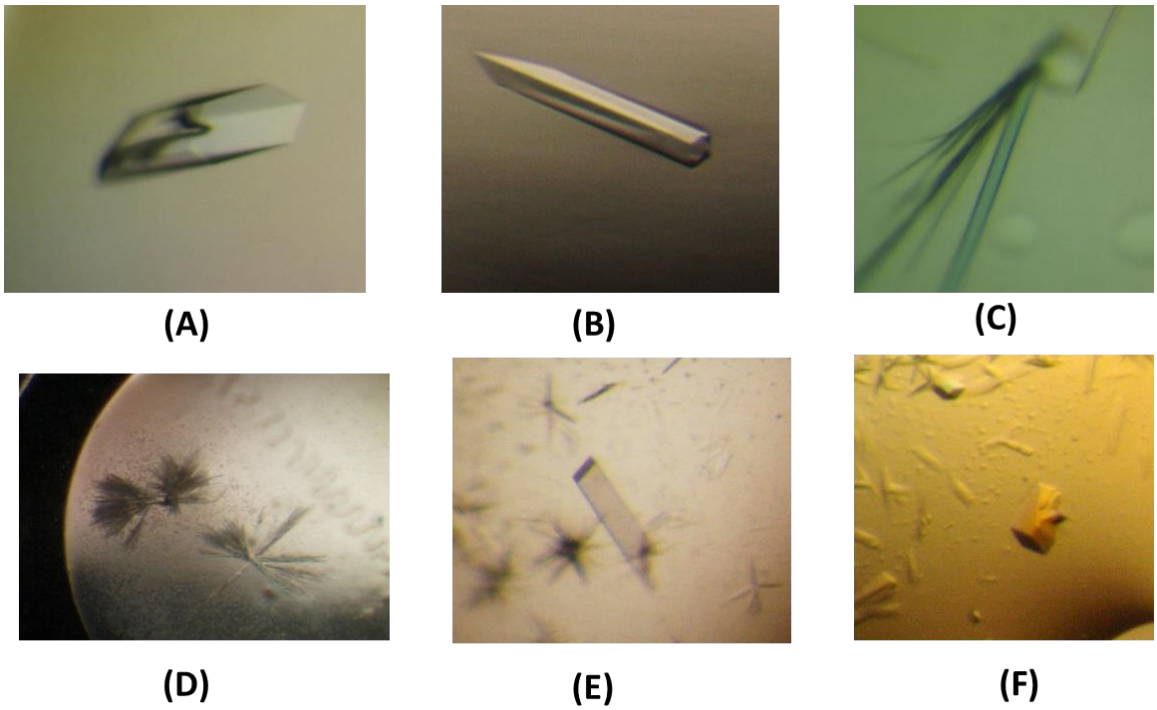


Figure 3.5 Protein crystals of Mfer762 and Mth734. **(A)** Mfer762 crystal obtained from the primary screen (SG1_E8 condition); (**B)** Mfer762 crystal from the SG1_E8 repeat drop; (**C)** Mfer762 crystal that dissolved and reappeared with different morphology; (**D**) Mth734 protein crystal in MORPHEUS II_H7 condition that was used for seeding experiment; (**E**, **F)** Mth734 protein crystal that appeared after six months on the same drop that initially had multiple needles (MORPHEUS II_G10 and _G9 condition, respectively). The magnification for all the crystal images is 80× except for D (45×).

3.5.2 Crystal testing

Salt crystals and their precipitates typically exhibit extreme birefringence under the microscope in the presence of polarised light whereas protein crystals change their colour slowly as the polarising filter is rotated. It is also one of the characteristics that can differentiate amorphous precipitate from microcrystals in a drop. Mfer762 crystals changed their colour slowly when polarising filter of the microscope was rotated. In addition, the

edges of Mfer762 and Mth734 crystals were regular and resembled the typical characteristics of protein crystals. Moreover, the Mfer762 crystal from SG1_E8 condition of the SG1™ screen was the only single crystal out of 576 conditions and therefore was not destroyed with the crush test or protein stain test. The crystal was exposed to the MX1 beamline at the Australian synchrotron and a protein diffraction pattern with resolution up to 1.9 Å resolution was obtained (Section 3.5.3.1 below).

Furthermore, several crystallisation conditions for Mth734 produced crystals but only one condition from the microseeding technique diffracted to a reasonable resolution. The Mth734 crystals were crushed using acupuncture needles to examine if they produced cracking sounds, a characteristics associated with salt crystals. The crystals that did not produce distinct cracking sounds and were further stained with crystal violet dye (Hampton "Izit" dye) and checked if the crystals retained the blue colour of the dye. More than 50 crystals of Mth734 from different crystallisation conditions were tested and diffracted as salt crystals before obtaining protein crystals diffracting at 2.7 Å resolution (Section 3.5.3.2 below).

3.5.3 X-ray diffraction and data collection

3.5.3.1 Mfer762 data collection

Multiple datasets for Mfer762 were collected at the MX1 and MX2 beamlines at the Australian synchrotron. The first dataset for apo Mfer762 was collected at the MX1 beamline using the SG1_E8 condition crystal straight for the primary screen plate and is shown in Figure 3.5A. The crystal diffracted to a resolution of 1.9 Å. A complete data set was collected at 100 K with an oscillation angle of 0.5° over a 360° oscillation range. All together 720

frames were collected with crystal to detector distance of 200 mm, attenuation 90% and a double-crystal monochromator (DCM) energy of 13 keV.

In addition, five different datasets for Mfer762 using the crystals obtained from co-crystallisation experiments and one dataset for a crystal from a soaking experiment were also collected at the MX1 beamline. Most of the datasets diffracted within the resolution range of 1.7 Å - 2.1 Å. Different detector distances were used for the data collection. However, other parameters for data collection such as attenuation, oscillation angle, overall oscillation range were kept the same (70%, 0.5° and 360°, respectively).

A further four datasets of Mfer762 were collected at the MX2 beamline for crystals with soaking experiments. The image collection at the MX2 beamline was performed using an EIGER detector (Casanas, *et al.*, 2016) where the oscillation angle by default was set to 0.1° with a complete dataset consisting of 3,600 frames. The total exposure time for all of the datasets collected using the EIGER detector was 72.0 secs. The energy used for data collection at the MX2 beamline for all four datasets was 13 keV and detector distance varied from 250 to 300 mm to optimise the data quality.

3.5.3.2 Mth734 data collection

The Mth734 crystal from MORPHEUS II_G10 (100 mM amino acids mix; 0.1 M buffer system 6, pH 8.5; 50% v/v precipitant mix of 25% w/v PEG4000 and 40% w/v 1, 2, 6-hexanetriol) that was obtained using randomised microseeding technique, as shown in Figure 3.5E, diffracted to a resolution of 2.7 Å. The data collection was performed with the same crystals at the MX2 beamline using the EIGER detector with a detector distance of 250 mm. An oscillation angle of 0.1° and total oscillation range of 360° as used by default at MX2 beamline. Hence, a total of 3,600 frames were collected for each of the datasets. Similarly, the second dataset was collected for the crystal from MORPHEUS II_G9 condition (Figure

3.5F) that diffracted to a resolution of 3.06 Å. The crystal to detector distance for the second dataset was also set to 250 mm. The total exposure time for both datasets was set to 72 secs. The dataset with a resolution of 2.7 Å was used for experimental structure determination. The summary of various datasets collected for pMurE targets during the project is shown in Appendix 5.

3.5.4 Phasing, model building and refinement

3.5.4.1 Solving the Mfer762 structure

The apo Mfer762 dataset was auto-processed using inbuilt XDS (Kabsch, 2010) software at the Australian Synchrotron. The Pointless and Aimless (CCP4i) programs (Evans & Murshudov, 2013) were used to scale and average the intensities and define the Rfree reflection set for crystallographic structure determination. An initial partial model for Mfer762 was obtained using the molecular replacement technique as implemented in MrBUMP (Keegan & Winn, 2008). The Aimless mtz file and amino acid sequence for Mfer762 were provided for the molecular replacement model search. The program selected MurF from *Acinetobacter baumannii* (4QDI) as the best homologous template model to solve the middle domain of Mfer762. The partial model from MrBUMP was then used as input into SHELXE (Thorn & Sheldrick, 2013) within CCP4i for model building. The output poly-alanine model from the program was used as further input into Buccaneer-autobuild/refine protein (Cowtan, 2006) within CCP4i. The output structure file from Buccaneer program contained a complete model for Mfer762. Visualisation and manual rebuilding was performed using COOT (Emsley, *et al.*, 2010) and refinement was conducted in REFMAC5 (Murshudov, *et al.*, 2011). The automatic water building and refinement option in REFMAC5 was used to place water molecules in the model. Further refinement

was later carried out using phenix.refine (Afonine, *et al.*, 2012) program in the Phenix software package (Adams, *et al.*, 2010). The MolProbity (Davis, *et al.*, 2004) results obtained after refinement run and validation tools included in COOT were used for model validation and the regions indicated as requiring further examination by the tools were inspected closely to improve the structure quality and statistics.

The solved and refined structure of apo Mfer762 was later used as a template for molecular replacement to solve the Mfer762 substrate-bound structures.

3.5.4.2 Solving the Mth734 structure

Mth734 has 54% sequence identity with Mfer762. The solved structure of Mfer762 was used as a search model for Mth734 for molecular replacement using PHASER (McCoy *et al.*, 2007) in CCP4i (Collaborative Computational Project, 1994). Matthew's coefficient calculation (Kantardjieff & Rupp, 2003) indicated the presence of two molecules in the asymmetric unit and suggested a solvent content of 55%. However, only one molecule could be solved using PHASER. The Phaser-MR tool (Bunkoczi, *et al.*, 2013) in Phenix resulted in a model with two molecules in the asymmetric unit, however, the C-terminal domain of the second molecule did not correlate with the electron density and was built manually.

The web server for automated crystal structure determination platform, Auto-Rickshaw, (Panjikar *et al.*, 2005) was implemented to perform MRSAD phasing method (Panjikar *et al.*, 2009) and used Mfer762 as a search model. Even though no contribution from SAD was observed for phasing, the program was selected to run iterative model building using Buccaneer (Cowtan, 2006) and ARP/wARP (Langer *et al.*, 2008) automatically and to check if the final model could be improved. However, the output from the program was also a single molecule model which was manually inspected and again used as a model for molecular replacement in PHASER. Only a single copy search for ensemble produced a

successful result. The initial output statistics from PHASER indicated a promising model (LLG and TFZ values of 2206 and 44.1, respectively, and RFZ value of 9.7), but a search for two copies gave a warning message with the top solution having a TFZ score below the cut-off and was unable to pack the solution model. This single molecule model was used as a search model in Phaser-MR (simple one component interface) in Phenix (He, *et al.*, 2016). The statistics for the output model indicated a reliable model (LLG and TFZ values of 2737.8 and 27.5, respectively) and contained two molecules in the asymmetric unit. However, the C-terminal domains of both the molecules clashed with one another. The output from Phaser-MR was used as an input file to run the AutoBuild program (Terwilliger, *et al.*, 2008) in Phenix and used to check if it could improve the model and build the C-termini in different orientations. There was no difference in the model, however, and it was clear that the C-terminus of the second molecule (B-chain) did not agree with the electron density map. The C-terminus of the second molecule and other regions in the structure that did not correlate to the electron density map were removed before structure refinement using phenix.refine. The positive density for the C-terminus of the second molecule reappeared in a completely new position resulting in a new orientation for the entire C-terminal domain in the second molecule. The manual rebuilding of the C-terminal residues of the second molecule in COOT (Emsley, *et al.*, 2010) was performed based on positive electron density map that reappeared after each round of refinement in phenix.refine. The final structure of the second molecule of Mth734 was completed and validated using the validation tools in COOT. The C-terminal domain of the final model of second molecule of Mth734 aligned neither with the corresponding domain of the first molecule nor the Mfer762 structure, which was used as a template structure for molecular replacement.

3.6 Structural studies of pMurE

3.6.1 Mfer762 structure

The structure of a methanogenic archaeal pseudomurein peptide ligase type E (pMurE) has been determined for the very first time. The overall structure of Mfer762 consists of three domains, an N-terminal Rossmann-fold like domain, a nucleotide-binding middle domain and another Rossmann-fold like C-terminal domain (Figure 3.6). Each domain is linked to one another by a connecting loop, which has been explained as a hinge region in bacterial murein peptide ligases (Smith, 2006). The connecting loop between the middle and C-terminal domain has been confirmed as a hinge region in another member of pMurE, Mth734, through domain movement (discussed in Section 3.6.3 below) and will be referred to as a hinge region for pseudomurein peptide ligases type E (pMurE). The first hinge region between the N-terminal and middle domains in the Mfer762 structure is short and consists of four residues, Ala101, Pro102, Asn103 and Thr104, whereas, the second hinge region between the middle and C-terminal domains is formed by seven residues and contains Tyr319, three arginine residues (Arg320, 324, 325), Pro321, Leu322, Lys323 and Phe326. All the Mfer762 structures determined during this project adopt cis-conformation for two proline residues (Pro333 and Pro474), which are located close to the two Cys residues forming the disulfide bond.

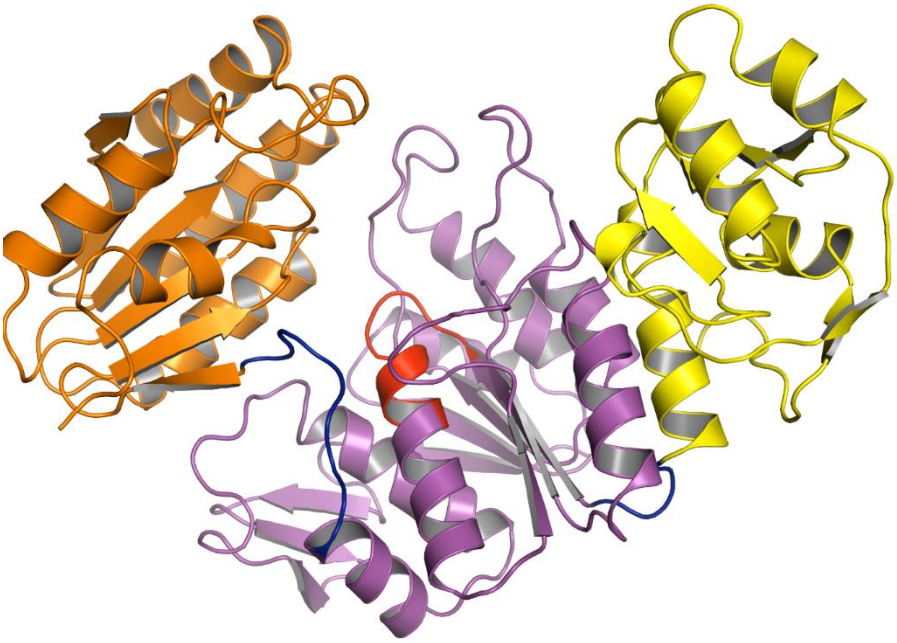


Figure 3.6 Cartoon representation of the overall Mfer762 structure. The N-terminal domain is represented in yellow, middle domain in purple and C-terminal domain in orange. Each domain is linked by a hinge region that is represented in blue. The signature P-loop is shown in red. The overall orientation of the three domains is maintained throughout the chapter (right to left) unless otherwise specified. The figure has been prepared using PyMOL (Schrodinger, 2015).

The Mfer762 structure exhibits a three domain architecture with the middle domain at the centre flanked by N- and C-terminal domains and contains 476 amino acid residues in total. The N-terminal domain is formed by the first 100 amino acid residues of the amino acid sequence (residues 1-100). It contains a mixed β-sheet formed by five β-strands (β1-β5) and four α helices (α1-α4) (Figure 3.7).

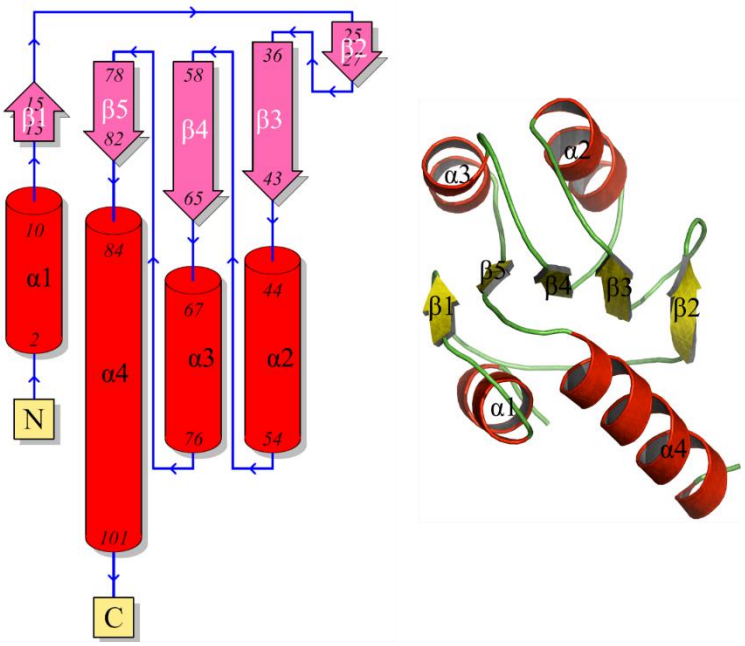


Figure 3.7 Topology and cartoon representation of the N-terminal domain of Mfer762 structure. The topology diagrams were obtained from the PDBsum webserver (Laskowski, 2009) using the co-ordinate file for domain separately. In the topology diagram, the red cylinders represent α-helices; the magenta arrows represent β-strands with the arrow head indicating the direction (N- to C-terminus) of the strands, blue lines indicate the loops with the arrow head representing their directions. Both α-helices and β-strands are numbered based on their connectivity and position in the secondary structure arrangement. The cartoon representation has been prepared using PyMOL.

The middle domain contains 214 amino acid residues (residues 105-318), which form three β-sheets and eight α helices (α5-α12). Among the three β-sheets, one forms a parallel sheet with six strands (β6-β8 and β11-β13), whereas the other two β-sheets are antiparallel and formed by two (β9 and β10) and three (β14-β16) strands, respectively (Figure 3.8). The beta hairpin structure formed by two antiparallel β-strands (β9 and β10) contains two amino acids in each strand (Gly 176, Trp177 and Arg180, Leu181, respectively) and is a flexible region in the Mfer762 structure and observed to have a role in substrate-binding (Section 3.6.1.2.1.1).

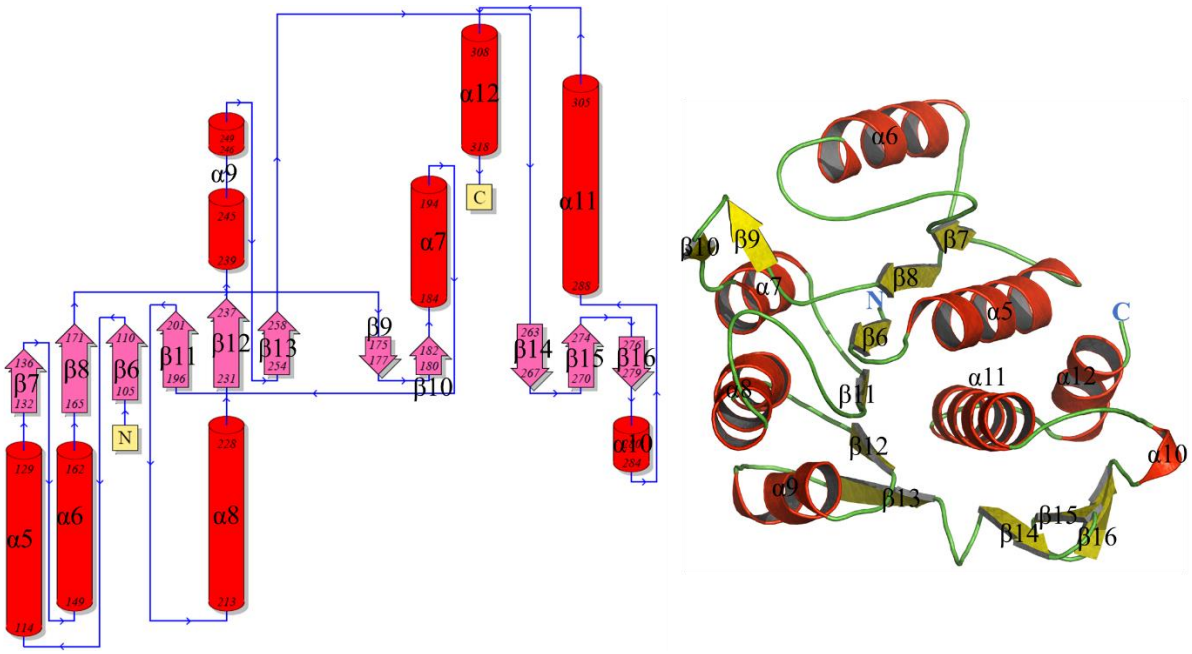


Figure 3.8 Topology and cartoon representation of the middle domain of Mfer762 structure. The topology diagrams were obtained from the PDBsum webserver (Laskowski, 2009) using the co-ordinate file for each domain separately. In the topology diagram, the red cylinders represent α-helices; the magenta arrows represent β-strands with the arrow head indicating the direction (N to C-terminus) of the strands, blue lines indicate the loops with the arrow head representing their directions. Both α-helices and β-strands are numbered based on their connectivity and position in the secondary structure arrangement. The cartoon representation has been prepared using the PyMOL visualisation software.

Finally, the C-terminal domain consists of 150 amino acid residues (residues 327-476) and contains one mixed β-sheet, formed by six β-strands (β17-β22) and five α-helices (α13-α17). In addition, all the Mfer762 structures determined during this study contain a disulfide bond in the C-terminal domain that involves the residues Cys330 and Cys475. The positions of these Cys residues are at two different ends of the C-terminal domain and are represented as sticks in the Figure 3.9. The function of the disulfide bond in the Mfer762 structure is not

clear, however, the position at which the disulfide bond is formed suggests that it could possibly help lock the C-terminal domain of Mfer762 at the observed conformation.

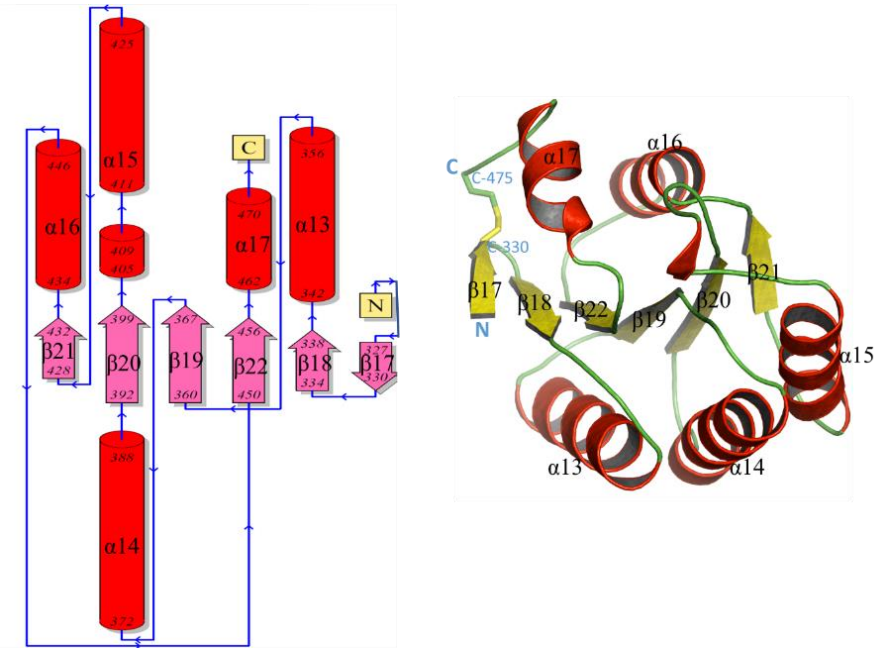


Figure 3.9 Topology and cartoon representation of the C-terminal domain of Mfer762 structure. The topology diagrams were obtained from the PDBsum webserver (Laskowski, 2009) using the co-ordinate file for each domain separately. In the topology diagram, the red cylinders represent α-helices; the magenta arrows represent β-strands with the arrow head indicating the direction (N- to C-terminus) of the strands, blue lines indicate the loops with the arrow head representing their directions. Both α-helices and β-strands are numbered based on their connectivity and position in the secondary structure arrangement. The cartoon representation has been prepared using PyMOL. The disulfide bond formed by Cys330 and Cys475 is represented as a stick and appears at the N- and C-termini of the C-terminal domain.

Three different structures of Mfer762 that were determined during this project will be discussed sequentially in the following sections. These structures include an apo and two UDP-bound structures of Mfer762. Among the two UDP-bound structures, each structure has UDP-bound at a different position and this will be discussed in detail in Section 3.6.1.2

and its sub-sections. The overall topology and secondary structure arrangement in all the Mfer762 structures analysed in this study are the same and therefore are discussed only once in this section. Any differences in the conformation or secondary structure arrangement will be discussed in the respective sections. The structural similarity search and comparative analysis suggest that Mfer762 is similar to MurE/F rather than MurC/D peptide ligase type in bacteria and this will be discussed in detail later in Section 3.11.

3.6.1.1 Mfer762 apo structure

The Mfer762 apo structure was obtained using the initial crystal from SG1_E8 condition (Figure 3.5A) and solved to 1.9 Å resolution. This is the first solved structure of Mfer762 and was obtained by molecular replacement technique using a combination of several programs as explained in Section 3.5.4.1 above.

The Mfer762 apo structure belongs to the $P6_1$ space group with a unit-cell parameter of a=b=119.77, c= 70.73 Å and contains a single molecule in the asymmetric unit. The final model for the Mfer762 apo structure was refined and rebuilt to achieve statistics as shown in Table 3.1. The R_{free} and R_{work} for the structure are 0.20 and 0.18, respectively. The structure contains two presumed PO_4^{3-} ions. In addition, the structure contains one glycerol molecule and Mg^{2+} ion. The average B-factor for the protein structure is 26 $Å^2$, whereas that for heteroatoms is 41 $Å^2$. The Mfer762 apo structure has an overall domain architecture, as explained earlier in Section 3.6.1 that resembles the overall structure and three domain arrangement of bacterial murein ligases (Smith, 2006; Kouidmi *et al.*, 2014) (Figure 3.10A). The three domains are arranged such that they form a crescent-shaped cavity at their interfaces as observed in bacterial homologues. This large cavity is required to bind a large substrate as observed for the MurE in bacteria (Basavannacharya, *et al.*, 2010b). The signature P-loop required for the Mur family activity is present in the middle domain and

possesses a similar conserved consensus amino acid sequence as observed in bacterial MurE at the putative ATP-binding region (Basavannacharya, *et al.*, 2010b). The residues across the P-loop and presumed phosphate ion bound close to the region along with their electron density are shown in Figure 3.10B.

In contrast to bacterial MurE, the Mfer762 apo structure is unique with the presence of two cis-proline residues (Pro333 and Pro474) that are close to the two cysteine residues involved in disulfide bond formation (Cys330 and Cys475). The disulfide bridge in the C-terminal domain where Cys330 is close to the hinge region and Cys475, which is the second last residue of the protein amino acid sequence. Even though the functions of the cis-proline residues and the disulfide bond in the Mfer762 structure are not clear, it can be speculated from the positioning that the disulfide bond could stabilise the conformation of the C-terminal domain (Figure 3.9).

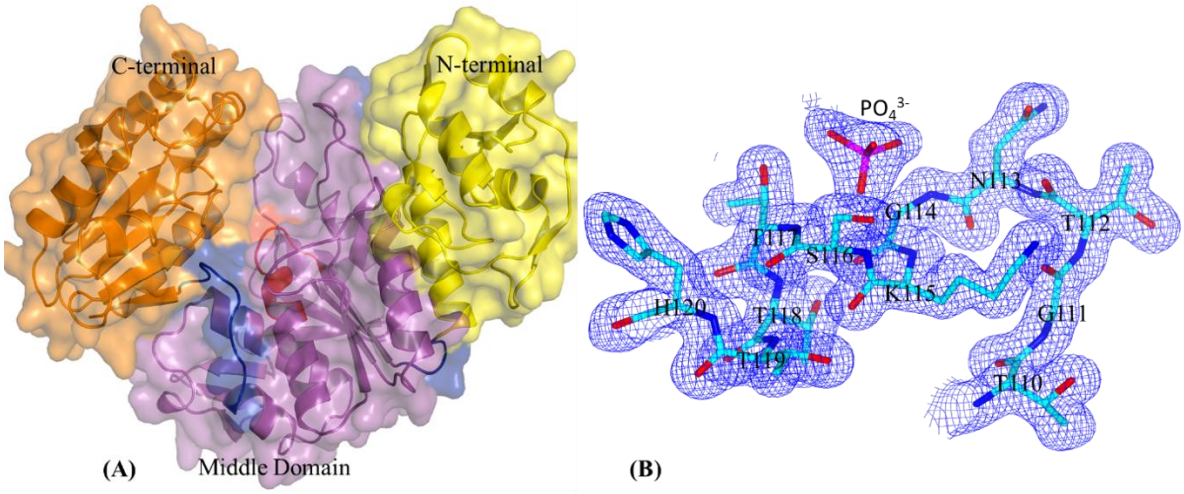


Figure 3.10 Mfer762 apo structure. (A) Transparent surface view with the N-terminal domain represented in yellow, the middle domain in purple and the C-terminal domain in orange. Each domain is linked by a hinge region that are represented in blue. The conserved P-loop is shown in red (buried). (B) Electron density map showing the residues across the P-loop. The presumed

phosphate ion and the amino acids are represented as sticks. The electron density map for the P-loop residues and phosphate ion is contoured at 1.0-σ.

3.6.1.2 UDP-bound Mfer762 structure

Two structures of Mfer762 in complex with partial substrate (UDP) have been solved at high resolutions (1.8 Å and 2.0 Å). Two different UDP-binding sites at different positions of the Mfer762 structure have been identified. The first structure has UDP bound on the exterior surface of the protein with the residues interacting from both N-terminal and middle domains and is referred to as Mfer762-UDP_1. Further, the second structure has UDP-bound on the inner surface, at the crescent-shaped cavity, but interacting solely with the residues of N-terminal domain and is referred as Mfer762-UDP_2 throughout this thesis.

Table 3.1 Data collection and refinement statistics. The values in parentheses represent the values for the highest resolution range. The table was generated using the 'Generate table' tool in Phenix.

	Mfer762-APO	**Mfer762-UDP_1**	**Mfer762-UDP_2**	**Mth734**
Resolution range (Å)	34.3 - 1.7 (1.8 - 1.7)	41.81 - 1.8 (1.9 - 1.8)	41.34 - 2.0 (2.1 - 2.0)	46.2 - 2.7 (2.8 - 2.7)
Space group	$P6_1$	$P6_1$	$C2$	$P\ 2_1\ 2_1\ 2_1$
Unit cell (Å / °)	119.8 119.8 70.8 90 90 120	119.4 119.4 71 90 90 120	126.8 45.0 98.1 90 98 90	66.0 127.2 134.7 90 90 90
Total reflections	1522905 (141585)	1234378 (115095)	248692 (18725)	433047 (41750)
Unique reflections	67101 (6643)	54413 (5429)	37078 (3578)	31878 (3130)
Multiplicity	22.7 (21.3)	22.7 (21.2)	6.7 (5.2)	13.6 (13.3)
Completeness (%)	99.85 (98.97)	99.90 (99.83)	99.61 (96.72)	99.97 (99.97)
Mean I/sigma(I)	31.86 (1.94)	35.94 (2.21)	15.91 (1.71)	15.57 (0.86)
Wilson B-factor ($Å^2$)	20.66	22.86	28.96	67.41
R-merge	0.18 (1.94)	0.22 (1.8)	0.16 (0.75)	0.251 (1.61)
R-meas	0.18 (1.99)	0.23 (1.84)	0.17 (0.83)	0.26 (1.68)
R-pim	0.04 (0.43)	0.05 (0.40)	0.06 (0.34)	0.07 (0.46)
CC1/2	0.99 (0.59)	0.99 (0.67)	0.99 (0.26)	0.99 (0.52)
CC*	0.99 (0.86)	0.99 (0.90)	0.99 (0.65)	0.99 (0.83)
Reflections used in refinement	67074 (6636)	54375 (5429)	37044 (3571)	31876 (3130)
Reflections used for R-free	3478 (319)	2692 (278)	1834 (159)	1532 (144)
R-work	0.18 (0.27)	0.19 (0.31)	0.18 (0.26)	0.20 (0.31)
R-free	0.20 (0.31)	0.23 (0.33)	0.23 (0.29)	0.26 (0.33)
CC(work)	0.96 (0.72)	0.96 (0.77)	0.95 (0.71)	0.93 (0.70)
CC(free)	0.96 (0.7)	0.94 (0.72)	0.93 (0.67)	0.91 (0.60)

Number of non-hydrogen atoms	4346	4333	3903	6846
Macromolecules	3905	3845	3700	6826
Ligands	27	58	40	
Solvent	414	430	163	20
Protein residues	482	482	469	897
RMS(bonds)	0.017	0.019	0.019	0.012
RMS(angles)	1.33	1.95	1.94	1.6
Ramachandran favoured (%)	96.25	95.62	96.54	95.3
Ramachandran allowed (%)	3.75	3.96	3.02	4.48
Ramachandran outliers (%)	0	0.42	0.43	0.22
Rotamer outliers (%)	2.93	2.55	2.42	9.08
Clash score	3.72	1.38	2.22	10.93
Average B-factor ($Å^2$)	27.4	28.41	36.43	73.73
Macromolecules	26.13	27.23	36.38	73.78
Ligands	45.44	38.57	46	
Solvent	38.2	37.57	35.16	54.96

3.6.1.2.1 Mfer762-UDP_1 structure

The UDP-bound structure (Mfer762-UDP_1) was the first Mfer762 structure to be determined in complex with UDP, representing a portion of the native substrate (N^{α}-UDP-Glu-Ala). The crystal structure was obtained from a soaking experiment using the similar morphology crystal as that used for the apo Mfer762 structure (Figure 3.5B) and was obtained from the same crystallisation condition (SG1_E8). The soak solution contained solely UDP that was prepared in the crystallisation buffer.

The Mfer762-UDP_1 structure also belongs to the $P6_1$ space group and has similar cell dimensions as that of the apo structure, a=b=119.37 Å and c= 71.10 Å. The Matthews' coefficient for Mfer762-UDP_1 is 2.73 $Å^3$/Da and the estimated solvent content is 55% with a single molecule in the asymmetric unit. The final refined model has 482 residues that includes seven extra amino acid residues at the N-terminus end of the structure (-7 to 0) whereas the last amino acid residue (Ser492) of the protein sequence could not be modelled. The final model contains one single molecule of Mfer762, a UDP molecule, two presumed PO_4^{3-} ions, one Mg^{2+} ion and a glycerol molecule. The electron density for the presumed PO_4^{3-} was initially modelled with the SO_4^{2-} ions as the crystal was grown in ammonium sulfate condition, however, the electron density and the B-factor improved after modelling as a phosphate ion. Moreover, the position for the presumed PO_4^{3-} is at the putative ATP-binding site and has B-factor of 25.4 $Å^2$ and is approximately at the same position as that for the presumed PO_4^{3-} in the Mfer762 apo structure (Figure 3.10B). The model contains seven residues with alternative conformations. The structure of Mfer762-UDP_1 has been refined with statistics as shown in Table 3.1. The final R_{free} and R_{work} values for the structure are 0.23 and 0.19, respectively.

The overall structure of Mfer762-UDP_1 is nearly identical to the Mfer762 apo structure with a RMSD value of 0.30 Å for 475 Cα atoms. The structure varies slightly at the N-terminal domain with movement of the domain away from the plane of Mfer762 apo structure (refer Figure 3.16). The conformational differences of the Mfer762-UDP_1 structure compared to the Mfer762 apo structure will be discussed further in Section 3.7.

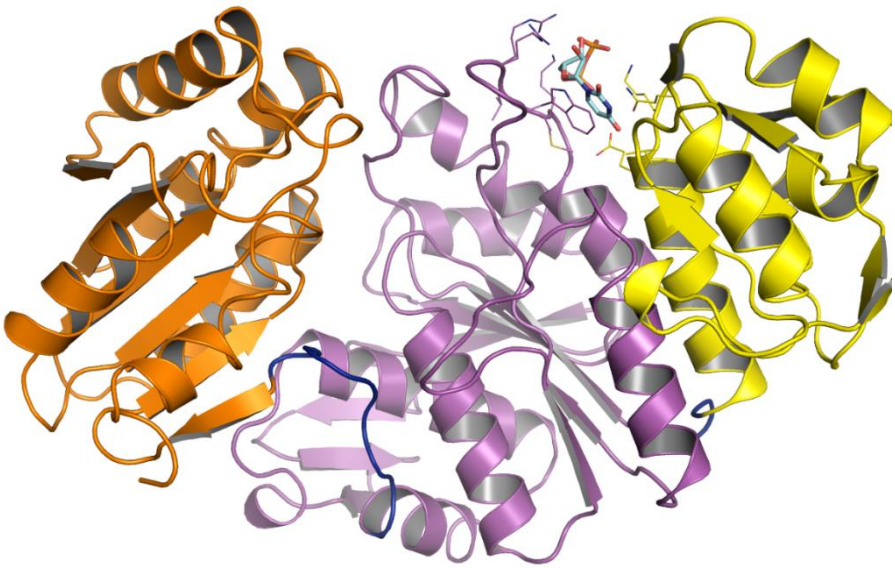


Figure 3.11 Structure of Mfer762_UDP_1. Individual domains are coloured separately, N-terminal domain in yellow, middle domain in purple and C-terminal domain in orange. The bound UDP is shown as sticks and the residues within 4.0 Å distance are shown as lines.

3.6.1.2.1.1 UDP-binding site of the Mfer762-UDP_1 structure

The UDP molecule for the Mfer762-UDP_1 structure is bound at the interface of the N-terminus and middle domains at the exterior surface of the protein molecule (Figure 3.11). The UDP forms interactions with the surrounding amino acid residues from both the N-terminal and the middle domains (Figure 3.12A). The oxygen atoms of the uracil ring of the UDP molecule form hydrogen bonds with the side-chain nitrogen atom of Gln62 and main chain nitrogen atom of Glu86. Similarly, the Gln62 side-chain oxygen atom also forms a

hydrogen bond with the (N3) nitrogen atom of the uracil ring. The side-chain nitrogen of Lys84 in the N-terminal domain interacts with the β-phosphate of the UDP molecule. Moreover, Lys183 from the middle domain interacts with the α-phosphate of the UDP molecule through the main chain as well as the side-chain nitrogen atom. Similarly, the main chain oxygen atom of Leu181 also forms a hydrogen bond with the phosphate at the α position and Arg180 forms an interaction with the (O2) oxygen atom of the sugar moiety (Figure 3.12B). This is a novel peptide ligase UDP-binding site which is formed through the involvement of the β-hairpin formed by β9 and β10 of the middle domain, (Figure 3.8), and has not been observed for bacterial murein peptide ligase structures. In addition to the aforementioned residues, Trp177 and Met182 also contribute to the UDP-binding pocket through hydrophobic interactions (Figure 3.12C). The uracil ring stacks between the Trp177 and Lys84 in this interaction. The electron density for the UDP molecule is well-defined and strongly supports the interaction of the UDP molecule at the site. The electron density around the residue Arg180 suggests an alternative conformation for the residue. Similarly, the electron density between C2 and C3 carbon atoms in the ribose sugar of UDP molecule is not connected (Figure 3.12C).

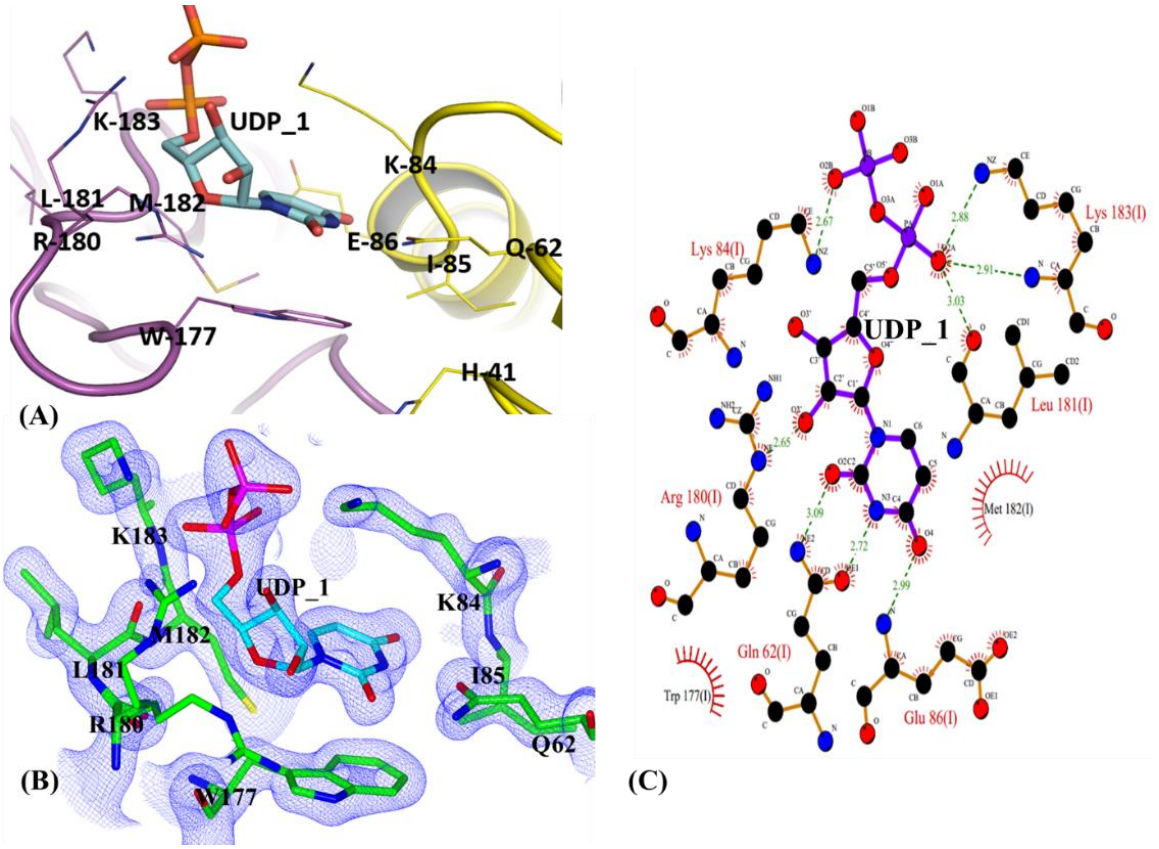


Figure 3.12 UDP-binding site in the Mfer762-UDP_1 structure. (A) The UDP-binding pocket is formed by the residues from both N-terminal and middle domains, shown in yellow and magenta, respectively. The residues within 4.0 Å distance have been shown and are represented as lines. (B) Electron density map for the UDP_1 molecule and the neighbouring residues within 4.0 Å is contoured at 1.0-σ. (C) The UDP interaction with the Mfer762 residues in the Mfer762-UDP_1 structure as obtained from the PDBsum webserver (Laskowski, 2009). The hydrogen bonds are shown as green dashed lines and hydrophobic interactions are represented by arches with spokes radiating toward the UDP molecule.

3.6.1.2.2 Mfer762-UDP_2 structure

The crystal structure of Mfer762-UDP_2 was obtained from a soaking experiment using a different morphology crystal (Figure 3.5C) than that used for the Mfer762 apo and Mfer762-UDP_1 structures, yet was grown in the same crystallisation condition (SG1_E8). This

crystal morphology was obtained in the same drop after the initial crystal dissolved, as described earlier in Section 3.5.1 above. The soak solution contained 20 mM UDP, 25 mM ADP and 25 mM L-glutamic acid prepared in the crystallisation buffer. The dataset was collected to 2.0 Å and solved by molecular replacement technique using the previously solved Mfer762 apo structure.

The Mfer762-UDP_2 structure belongs to the *C*2 space group with the cell dimensions of 126.76 44.98 98.05 90.00 98.19 90.00. The Matthews' coefficient for Mfer762-UDP_1 is 2.59 $Å^3$/Da and the estimated solvent content is 52.5% with a single molecule in the asymmetric unit. The final refined model of Mfer762-UDP_2 has 469 residues, which includes eight extra amino acid residues at the N-terminus end of the structure (-7 to 0); the last amino acid residue (Ser492) was not modelled. In addition, the structure contains two gap regions, 139-150 and 175-178, due to the residues within the regions not being able to be modelled in the final structure owing to limited electron density. Both these gaps are close to the UDP-binding site (UDP_2). The first longer gap between 139 and 150 is deep at the crescent-shaped cavity formed at the interface of three domains whereas the second gap between 174 and 179 is on the surface closer to the UDP-binding site (both represented as broken lines in Figure 3.13). The second short gap include residues from the β-hairpin formed by β9 and β10 in the middle domain (Figure 3.8) and contributes to UDP-binding site in the Mfer762-UDP_1 structure. Moreover, the final model contains one molecule of UDP, three presumed phosphate (PO_4^{3-}) ions and 165 water molecules, however, no ADP molecule or amino acid residue (L-Glu) used in soaking experiments was bound to the structure. The position of the presumed PO_4^{3-} is close to the putative ATP-binding region and aligns with the α phosphate of the bacterial ATP in structural comparison with bacterial MurE and has a B-factor of 23 $Å^2$. The structure of Mfer762-UDP_2 has refined with

statistics as shown in Table 3.1. The final R_{free} and R_{work} values for the structure are 0.23 and 0.18, respectively.

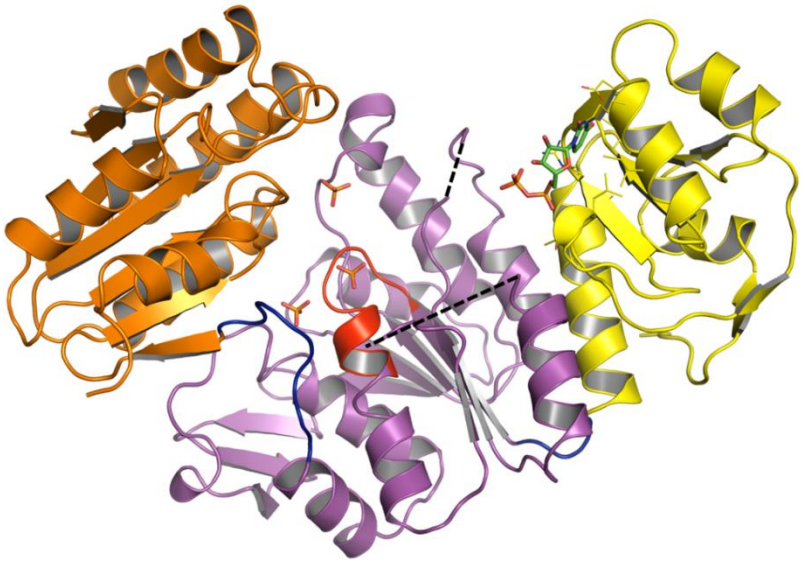


Figure 3.13 Structure of Mfer762-UDP_2 with each domain coloured differently, N-terminal domain in yellow, middle domain in purple and C-terminal domain in orange. The UDP is bound to the N-terminal domain and is represented as sticks with carbon atoms coloured in green. The residues within 4.0 Å distance from UDP are represented as lines. The regions that are flexible and could not be modelled in the final model are indicated by dashed lines (black). The presumed phosphate molecules are also represented as sticks and are close to the putative ATP-binding site. The conserved P-loop motif is shown in red.

3.6.1.2.2.1 UDP-binding site of the Mfer762-UDP_2 structure

Mfer762-UDP_2 is the Mfer762 protein in complex with UDP molecule where the UDP-binding pocket is completely different compared to that of the Mfer762-UDP_1 structure. The position of the UDP binding in Mfer762-UDP_2 is located at the inner cavity of the structure and comprises residues solely from the N-terminal domain. The position of the UDP in Mfer762-UDP_2 structure is similar to the UDP-binding in murein peptide ligases

(particularly MurF from *Acinetobacter baumannii* (4QDI) (Cha *et al.*, 2014)) and will be discussed later in Section 5.9.1 during comparative analysis.

The UDP in the Mfer762-UDP_2 structure binds at the crescent-shaped cavity of the protein structure with the uridine ring buried into the N-terminal domain whereas the negatively charged phosphate end is suspended freely in the cavity. The UDP molecule binds to the region surrounded by residues Thr28, Leu29, Gly30, Arg40, Ile43, Asp44, Lys46, Gly47 and Ile50 within 4.0 Å (Figure 3.14). Among these residues, Asp44, Arg40 and Gly30 are involved in hydrogen bonding to the UDP molecule whereas other residues form hydrophobic interactions. Similarly, Arg40 and main chain nitrogen of Gly30 also form hydrogen bonds with the phosphate of the UDP molecule (Figure 3.14B). The position of the UDP in the Mfer762-UDP_2 structure favours the assumption for the presence of a large substrate-binding pocket that could incorporate N^{α}-UDP-Glu$^{\gamma}$-Ala. The disordered loop region 139-150 suggests a conformational change in the Mfer762 structure near the crescent-shaped cavity formed at the interface of three domains during substrate binding.

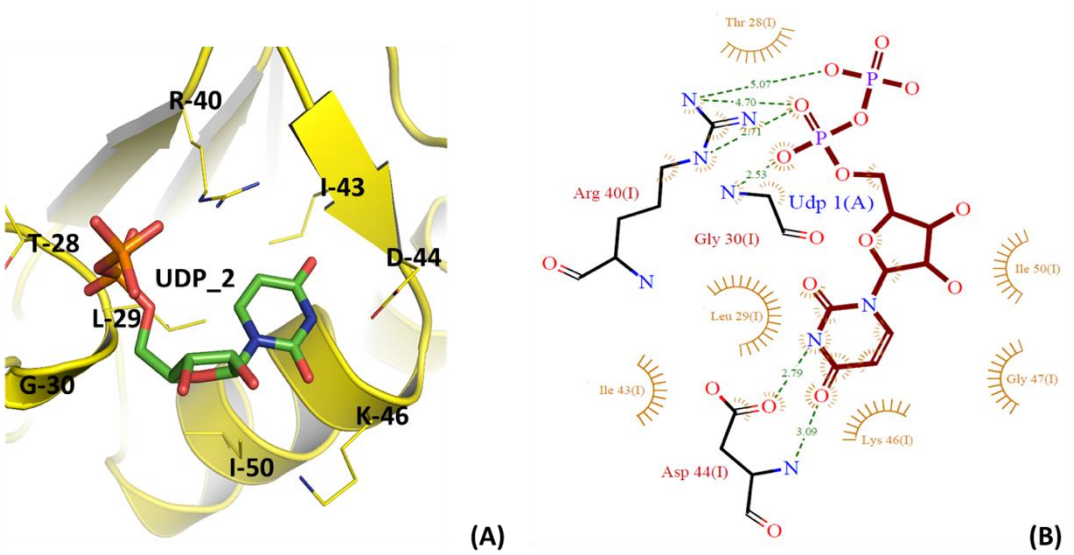


Figure 3.14 UDP_2 binding mode in the Mfer762 structure. **(A)** The UDP-binding site in Mfer762-UDP_2 structure. The binding site is entirely within the N-terminal domain. The residues within 4.0

Å were selected and are represented as lines and are involved in interactions with the UDP molecule. **(B)** Interaction of UDP with surrounding amino acid residues. The UDP makes hydrogen bonds with side-chains of Arg40 and Asp44 and main chains of Gly30 and Asp44. The hydrogen bonding is shown as green dashed lines and hydrophobic contacts are represented by arches with spokes radiating toward the UDP molecule. The cartoon representation for the UDP-binding pocket was prepared using PyMOL and the interaction figure was obtained from the PDBsum webserver and uses the LIGPLOT program (Wallace *et al.*, 1995).

3.6.1.2.3 Proposed ATP-binding site in Mfer762

A structure of Mfer762 in complex with ATP could not be obtained despite several experimental attempts. Experiments targeting the co-crystallisation of ATP, ADP and a non-hydrolysable structural analogue of ATP, AMP-PNP (adenylyl-imidodiphosphate), were also attempted but none of the final resulting electron density maps showed evidence for ATP or ATP derivatives bound to the structure(s). In addition, soaking experiments with all the aforementioned ATP derivatives were attempted but they too did not contain ATP or derivative forms of ATP molecules in the final model(s).

It has been observed that a presumed PO_4^{3-} ion is observed at the proposed ATP-binding region and aligns with the α phosphate of a bacterial bound ATP molecule. The structure superposition of Mfer762 with MurF from *Acinetobacter baumannii* (4QDI) (Cha, *et al.*, 2014) suggests that the putative ATP-binding site in the Mfer762 structure is conserved. Also, it is important to note that the ATP-binding site for all the available murein peptide ligases is located near to the conserved P-loop and at the interface of the middle and C-terminal domains. The ATP-binding residues are well-conserved for all the murein ligases (Smith, 2006). Further, the PO_4^{3-} molecule in Mfer762 aligns well with the α-phosphate of the ATP molecule in the 4QDI structure and is shown in Figure 3.15. The density for the

PO_4^{3-} ion approximate to this position appears with most of the Mfer762 structures solved during this project. The positive electron density peak that appeared after a few rounds of refinement at this position was initially modelled by fitting in SO_4^{2-}, as found in the crystallisation conditions, but the electron density map and refinement improved when modelling a PO_4^{3-} ion instead.

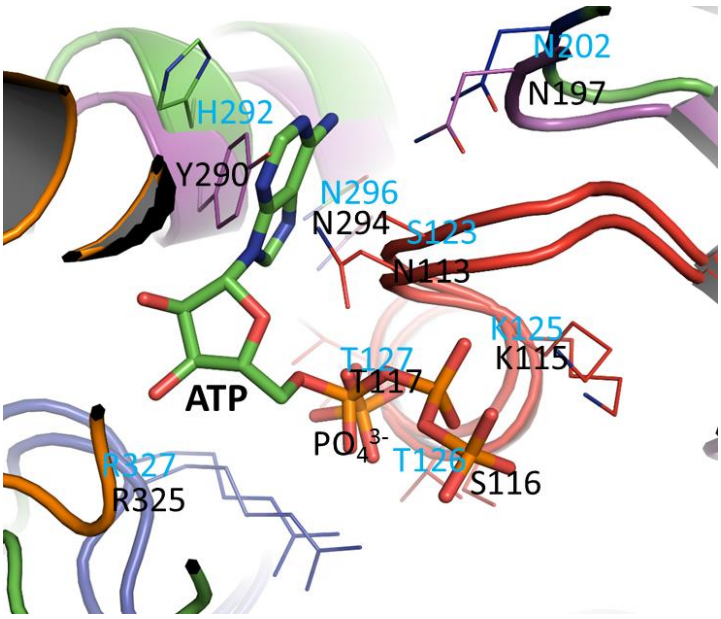


Figure 3.15 The putative ATP-binding site in the Mfer762 structure. The binding site is close to the conserved P-loop (red). The Mfer762 structure is represented as purple and the bacterial MurF (4QDI) in green. The amino acid residues within the 4.0 Å from the ATP of the 4QDI structure are shown as lines and labelled in blue and the respective residues in Mfer762 structure are labelled black. The Arg325 from the hinge region of Mfer762 also appears to interact with the ATP as observed for bacterial structure. The presumed PO_4^{3-} ion in Mfer762 lies at the α position of ATP molecule and is shown as stick.

3.6.1.3 Mfer762 structure comparison

The most interesting difference observed among the Mfer762 structures is the presence of two different UDP-binding sites. The UDP-binding site for the Mfer762-UDP_1 structure is

present at the external surface of the protein at the interface of the N-terminal and the middle domains. A loop extends from the middle domain and projects slightly towards the N-terminal domain and interacts with the UDP molecule. In contrast, the UDP-binding site for the Mfer762-UDP_2 structure is present at the crescent-shaped cavity formed by three domains yet solely interacting with the residues from the N-terminal domain. This position for UDP is the one expected for pseudomurein peptide ligase activity and has been observed for the bacterial murein peptide ligases, and will be explained in Section 3.11 below.

All three Mfer762 structures presented in this chapter were compared and demonstrated that they share the same overall domain architecture and secondary structure arrangement. The secondary structure superposition of two Mfer762 structures (Mfer762 apo and Mfer762-UDP_1) shows no significant differences other than slightly higher RMSD values at the UDP-binding region in the overall Mfer762-UDP_1 structure. However, the Mfer762-UDP_2 structure has a relatively higher RMSD around the crescent cavity surface and undergoes a conformational change close to the UDP-binding site. The secondary structure based superposition of the Mfer762-UDP_2 and the Mfer762 apo structures has a RMSD value of 1.3 Å for 460 Cα atoms. The comparative analysis of each domain indicates that the middle domain of the two UDP-bound Mfer762 structures contributes to most of the structural differences (RMSD value of 0.89 for 196 residues out of 215 residues used for alignment). The RMSD value for the N-terminal domain (100 Cα atoms), and the C-terminal domain (150 Cα atoms) are both 0.3 Å. All the amino acids used for the N- and C-terminal domains comparison are used in the analysis and display highly similar alignment. The shift of the N-terminal domain away from Mfer762 apo structure is clearly observed in the secondary structure based superposition of all the three Mfer762 structures and is shown in Figure 3.16. The movement of the N-terminal domain of Mfer762-UDP_2 in reference to the Mfer762 apo structure supports the evidence for the flexibility of the N-terminal domain

to move towards and away from the middle domain to adopt open and closed conformations. The opening and closing of the murein ligase is required to bind the substrate and perform the ligation of upcoming amino acid to the peptide unit and is enabled by the flexibility of the hinge region between each domain (Perdih, *et al.*, 2007; Sink *et al.*, 2016).

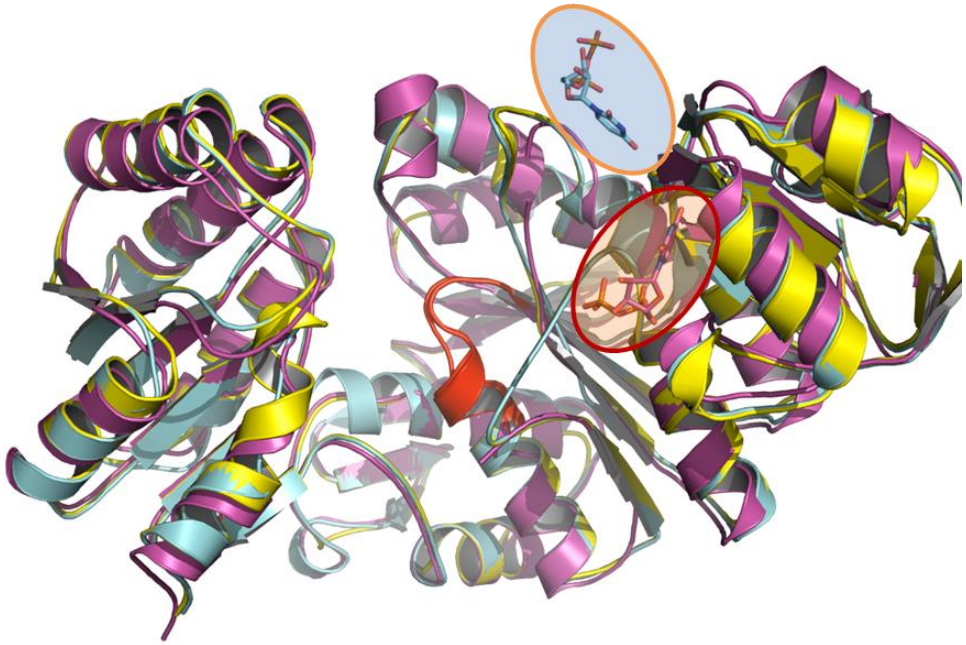


Figure 3.16 Structural comparison of Mfer762. The Mfer762 apo structure is shown in yellow, the Mfer762-UDP_1 in cyan and the Mfer762-UDP_2 in magenta. The conserved P-loop is represented in red and the two different UDP molecules are circled using different colours. The UDP from the Mfer762-UDP_1 structure is circled in orange whereas the bound UDP from the Mfer762-UDP_2 structure is in red. It can be clearly observed that the position of the two UDP binding sites are at two completely different positions of the protein structures. Moreover, the structures of Mfer762 apo and Mfer762-UDP_1 are relatively similar in comparison to Mfer762-UDP_2.

The major differences in the structural alignment (resulting in relatively higher RMSD value at the middle domain) of the two UDP-bound Mfer762 structures were observed in regions 139-150, 172-184 and 207-217 of the middle domain. Among these, the amino acid residues in region 139-150 could not be modelled in the Mfer762-UDP_2 structure, presumably

disordered, that provides additional space at the crescent-shaped cavity. The region 172-184 contributes the β-hairpin structure in the Mfer762 overall structure and contains two β-strands formed by two residues each (Figure 3.8). This is the region in the overall Mfer762 structure that extends from the middle domain slightly towards the N-terminal domain and is involved in binding of UDP in the Mfer762-UDP_1 structure. This region will be referred to as an 'extended loop' region and four residues (175-178) of this region are missing in the Mfer762-UDP_2 structure indicative of another disordered region in the structure. The residue Trp177 of the region adopts different side-chain conformations in the Mfer762 apo and Mfer762-UDP_1 structures. Trp177 together with His41 from the N-terminal domain adopts a 'gateway-like' conformation formed by the side-chains of these residues and will be discussed further in Section 3.7.1.1.1 and a similar conformation is adopted in Mth734 structure (detail in Section 3.6.3). Both of the aforementioned residues are structurally conserved in the recently determined pMurE structures and are conserved among most of the pMurE sequence of the pseudomurein-containing methanogens (Figure 3.38). Finally, the third difference observed with the structural alignment of two UDP-bound structures is in region 207-217. This is the loop region adjacent to the 'extended loop' and suggests structural flexibility that could have been propagated as a result of flexibility in the 'extended loop' region.

In addition, there is a switch of the Trp42 rotamer from m95 to p-90 in the Mfer762-UDP_2 structure. This conformation of the Trp42 is only observed upon UDP_2 binding and appears to be a key conformational change that would be required by pMurE for activity. The switch in the side-chain rotamer conformation of Trp42 in Mfer762-UDP_2 allows the binding of UDP and prevents steric hindrance as expected through p-90 rotamer adopted by the other two structures (Figure 3.17). Hence, based on the structural comparison it could be speculated that the disordered 139-150 region and rotameric switch of the Trp42 side-chain

could be two major conformational changes that pMurE might adopt for the substrate-binding at the putative catalytic cavity as seen in the Mfer762-UDP_2 structure.

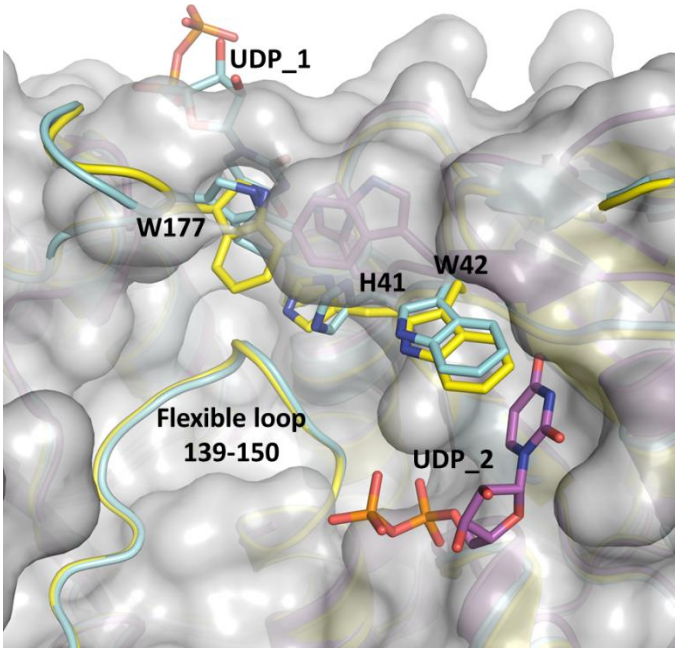


Figure 3.17 Surface view of Mfer762-UDP_2 structure in superposition with Mfer762 apo (yellow) and Mfer762-UDP_1 (cyan) structures. The putative conformational change required for the binding of the UDP_2 into the inner cavity (shown in magenta) suggests the need for the switch of Trp42 side-chain rotamer (m95, yellow and cyan) to p-90 (magenta) to avoid steric hindrance. The conformation of the flexible loop (139-150) that is located deep in the crescent cavity also suggests that it may hinder UDP-binding at the inner UDP_2 site and therefore would require conformational flexibility. The residues Trp177 and His41 that contribute to the 'gateway-like' conformation are also shown to indicate their position in the structure and are represented as sticks.

3.6.2 Mth734 structure

The Mth734 structure was obtained from a crystal produced implementing a microseeding technique that used the initial needle clusters grown in MORPHEUS II_H7 condition that were plated back into MORPHEUS II screen using the protocol explained in Section 2.2.3.1.3. The crystals observed in MORPHEUS II_G10 condition after six months were

tested at the MX2 beamline and data collected leading to the only dataset for Mth734 at 2.7 Å. None of the other crystals diffracted to this resolution from the same or any other crystal condition. Most of the crystals diffracted to a resolution lower than 3.5 Å. A complete dataset was collected on the crystal at the MX2 beamline using an EIGER detector at a distance of 250 mm, 0.1° oscillation angle and an overall oscillation range of 360°.

The Mth734 structure was solved by molecular replacement using the previously solved pMurE (Mfer762 apo) structure (Section 3.5.4.2). The Mth734 structure belongs to the $P2_12_12_1$ space group with unit-cell dimensions of 65.97, 127.22 and 134.62 Å. The final model contains two protein molecules, one sulfate molecule and 30 water molecules. The Matthew's coefficient for Mth734 is 2.82 $Å^3$/Da and the estimated solvent content is 56.5% with two molecules in the asymmetric unit. The Mth734 shares a similar domain arrangement and topology as explained for the Mfer762 structures (Section 3.6.1). The three domains are arranged together to form a crescent-shaped cavity as observed for Mfer762 and for bacterial murein ligases. The Mth734 protein consists of 450 amino acid residues, of which two amino acids at the C-terminus (Arg449 and Met450) could not be modelled in the final structure. The Mth734 structure has a shorter N-terminal domain compared to Mfer762 and contains the first 80 residues of the amino acid sequence (residues 1-80). The domain consists of one parallel β-sheet, formed by four β-strands (β1-β4), and four α-helices (α1-α4). The middle domain consists of 213 residues (residues 85-297) that forms three β-sheets and seven α-helices (α5-α11). Among three β-sheets, one forms a parallel sheet with six strands (β5-β7 and β10-β12) where as two other sheets are antiparallel and contain two (β8 and β9) and three (β13-β15) strands, respectively. The antiparallel β-sheet with β8 and β9 strands forms a β-hairpin that contains two amino acids in each strand (Gly 156, Trp157 and Arg160, Val161, respectively). This β-hairpin resembles the β-hairpin of the Mfer762 structure formed by β9 and β10 and referred as an 'extended region' that contributes to

interaction with the UDP molecule at the exterior surface (Section 3.6.1.2.1.1). Finally, the C-terminal domain consists of 143 amino acid residues (306-449) and contains one mixed β-sheet, formed by six β-strands (β16-β21) and six α-helices (α12-α17). The N-terminal domain is linked to the middle domain by a connecting loop that contains four residues; Ala81, Pro82, Ser83 and Ser84. The connecting loop between the N-terminal domain and the middle domain has been referred as the hinge region in published bacterial murein ligase structures (Smith, 2006). Similarly, the second connecting loop between middle domain and C-terminal domain consist of seven residues consisting of Tyr298, three Arg299/303/304, Pro300, Leu301, Lys302 and Phe305 and has been identified as a hinge region in pMurE.

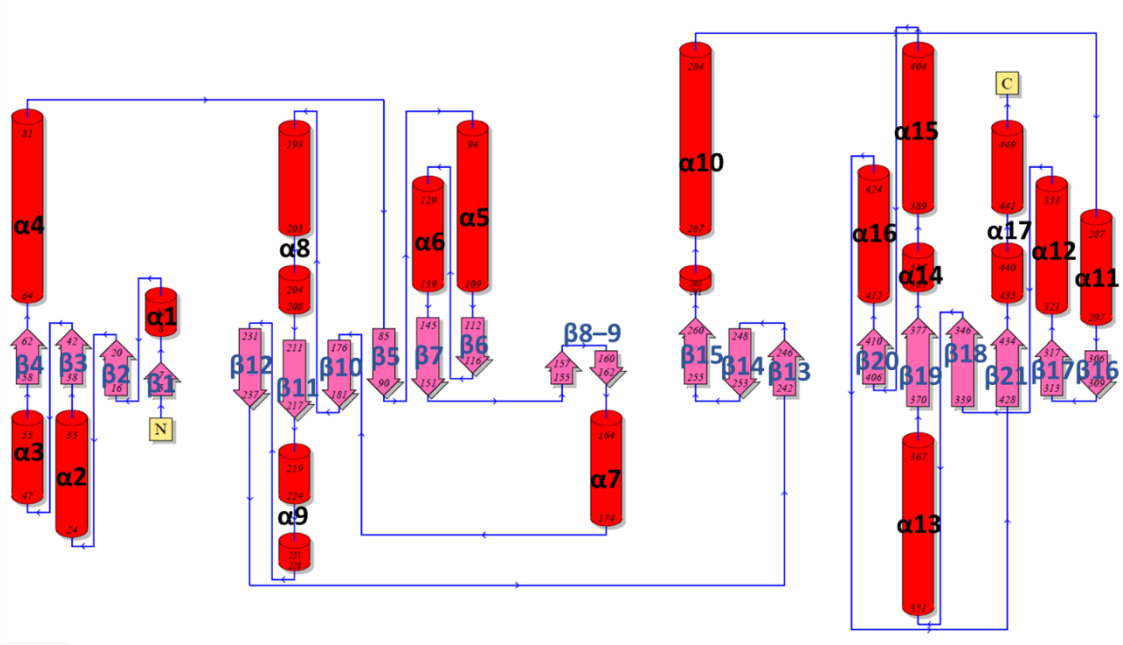


Figure 3.18 The topology diagram for the Mth734 structure. The α-helices and β-strands are represented as cylinders (red) and arrow (magenta) with their respective numbering in black and light blue, respectively. The structure adopts the same topology as shown for the Mfer762 structure except

for a shorter N-terminal domain resulting, therefore in one less β-strand compared to the Mfer762 structure.

The final model for the Mth734 structure has been refined and rebuilt to achieve the statistics as indicated in Table 3.1. The R_{free} and R_{work} for the Mth734 structure are 0.26 and 0.20, respectively. The structure contains a cis-proline residue (Pro312) at the N-terminus of the C-terminal domain of both the Mth734 molecules. The signature P-loop required by the Mur family is present in the middle domain and the amino acid sequences of the P-loop in the Mfer762 and the Mth734 structures are strictly conserved, as are the residues in bacterial MurE. The final structure of Mth734 has two molecules in the asymmetric unit that exhibit different domain conformations, mostly with respect to the C-terminal domain. Each molecule of Mth734 is shown separately in Figure 3.19. The C-terminal domain of the second molecule makes a rigid body rotation compared to the first molecule that has been used as a reference structure for comparison and exhibits the same conformation as that studied for Mfer762.

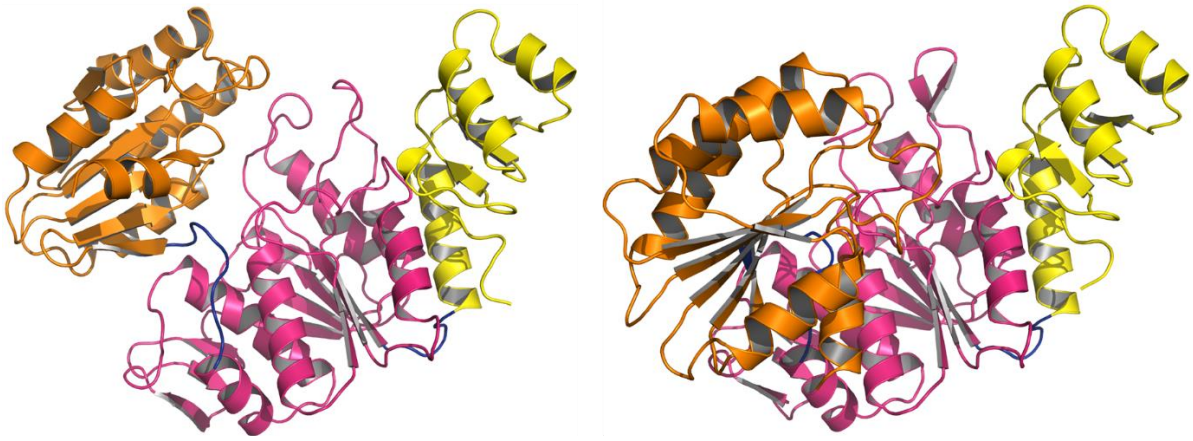


Figure 3.19 Cartoon representation of the Mth734 structure. The two molecules of the structures are presented in the same orientation, as represented for Mfer762 structures (N-terminal-domain (yellow) to the right, middle domain (magenta) at centre and C-terminal domain (orange) to the left).

(A) The first molecule of Mth734 (Mth734_A), (B) the second molecule (Mth734_B) arranged in reference to Mth734_A to highlight the rigid body motion of the C-terminal domain.

3.6.3 Intramolecular comparison of Mth734

The two individual molecules within the asymmetric unit of the Mth734 structure show conformational differences in the overall structure. As explained in the previous section and demonstrated in Figure 3.19, the arrangement of the C-terminal domain in the Mth734 structure appears to limit access to the crescent-shaped cavity as shown in Figure 3.19B. These two different domain arrangements within the asymmetric unit could have been a result of crystal contacts and demonstrate the possible 'open' and 'closed' conformation of the Mth734 structure. The first molecule of Mth734 in an open conformation will be referred to as Mth734_A (Figure 3.19A) and the second molecule with relatively closed conformation as Mth734_B (Figure 3.19B) throughout this chapter/thesis.

The comparative analysis of the two Mth734 molecules show that they share the same domain architecture and secondary structure arrangement. The structural superposition of the two molecules aligns residues 1-299 that contribute for the N-terminal and middle domains of the Mth734 structure. The structures are misaligned at the C-terminal domain due to its rotation. The C-terminal domain in Mth734_B makes a 91.5° rigid body rotation from the C-terminal domain of the Mth734_A (Figure 3.20A) that results in large conformational change compared to Mth734_A. The conformational difference between the two molecules is mostly due to the rotation of the C-terminal domain that originates from the hinge region. The secondary structure superposition of the C-terminal domains itself has a RMSD value of 0.35 Å for 148 Cα atoms suggesting that they are structurally identical. The RMSD for the N-terminal and middle domain is 0.70 Å for 292 Cα atoms. The major differences observed within the aligned regions include residues 118-123, 153-163 and 186-

194 where the alignment of the structures is relatively poorer (RMSD values greater than 1.0 Å) and are discussed further as a flexible region. Moreover, the distinct difference observed between the two molecules of Mth734 is at the C-terminal domain.

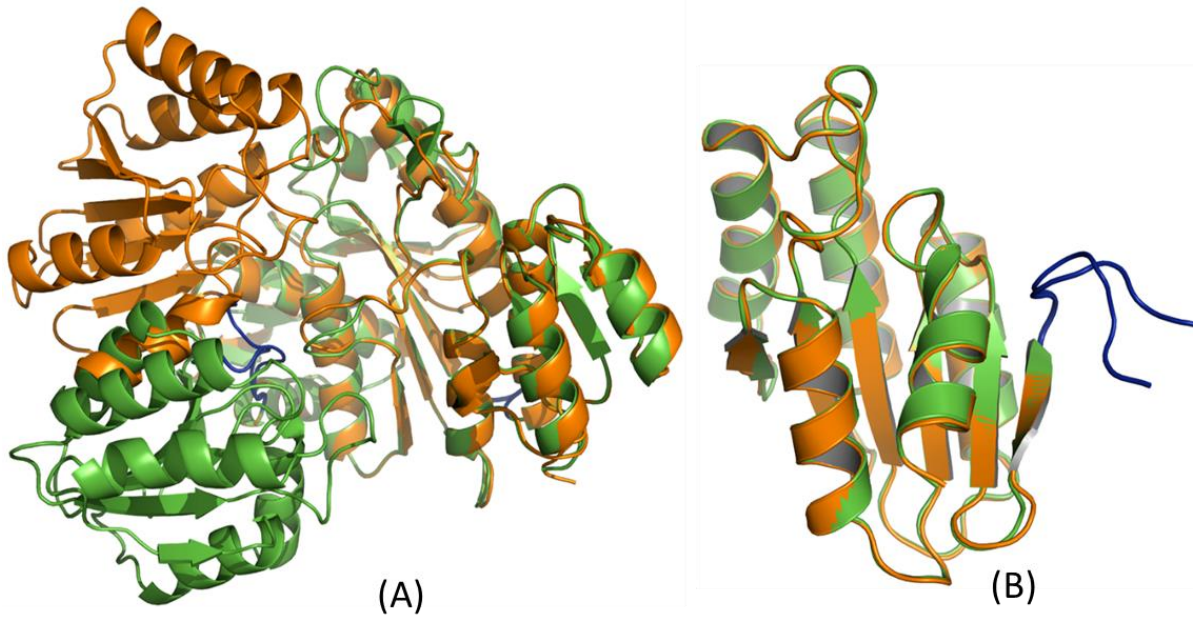


Figure 3.20 Structural comparison of two molecules of Mth734 highlighting the rigid body rotation of the C-terminal domain, (A) The Mth734_B (green) and Mth734_A (orange) structures display a large conformational change in the overall structure which originates from the hinge region between the middle domain and the C-terminal domain (shown in blue). The orientation of the Mth734 molecules in this figure is a 90° horizontal rotation from the reference orientation used in Figure 3.19. (B) The comparison of only the C-terminal domains in the two different molecules shows no difference other than in the hinge region.

The analysis of the aforementioned flexible regions of Mth734 molecules can potentially help to explain their involvement in the rotation of C-terminal domain in the Mth734_B molecule. The conformation acquired by the side-chain of His21 in Mth734_B (His21|B) clashes with the conformation of Trp157|A suggesting an alternative conformation for the Trp157|B. This could be associated with the movement of the loop 153-163 in Mth734_B

away from Mth734_A (indicated as 1 in Figure 3.21) that eventually propagates on to the next flexible loop 186-194 (indicated as 2 in Figure 3.21). This observation in the flexibility of the loops suggests the need for rotation of the C-terminal domain away from the conformation as adopted by the C-terminal domain of Mth734_A. The rotation of the C-terminal domain initiates from the connecting loop between the middle and the C-terminal domains (298-305), and therefore, has been discussed as a hinge region in pMurE enzyme type (indicated as 4 in Figure 3.21). The residues in this hinge region are conserved for both Mfer762 and Mth734. While two of the flexible loops are on the surface of the protein structure, the third loop 118-128 (indicated as 3 in Figure 3.21) lies deep in the crescent cavity formed by the three domains and has been described as a disordered region in the Mfer762 structure.

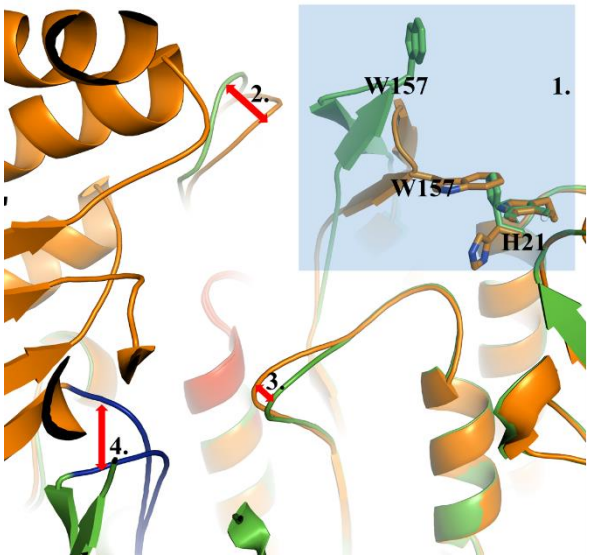


Figure 3.21 The secondary structure superposition of two molecules of Mth734 highlighting the key conformational changes. 1. The flexible region 153-163 in Mth734 adopts a 'gateway-like' conformation through Trp157 and His21 (blue box), 2. Conformational change adopted by second flexible loop (186-194), 3. The flexible loop region 118-128 that is buried deep in the crescent cavity and close to the internal UDP-binding site. 4. Difference in conformation at the hinge region that allows the C-terminal domain to rotate. (1, blue box) is the same region that could have a role in

UDP recognition and may form a ‘gateway-like’ for the UDP to enter into the enzyme surface as discussed previously for Mfer762.

It is interesting that the ‘gateway-like’ conformation adopted by His21 and Trp157 (equivalent to Trp177 and His41 in Mfer762) is similar to that studied with different Mfer762 structures. The residues involved and the conformations adopted by them in the two molecules of the Mth734 structure are similar to those of the Mfer762 structures and are highlighted in the blue box and indicated as (1) in Figure 3.21. The movement of the Trp157 in the Mth734_B molecule (green) favours the movement of the entire flexible loop 153-163 away from the N-terminal domain suggesting a wider opening at the region in the Mth734_B molecule. These two conformations adopted by the Mth734 molecules at the interface of the N-terminal and middle domains could be crucial in acquiring ‘open’ and ‘closed’ state for pMurE structure. Such a conformational switch from an ‘open’ to ‘closed’ state is required by murein ligases for ligase activity (Sink, *et al.*, 2016) and could reflect to the conformations likely possessed by pMurEs for their activity.

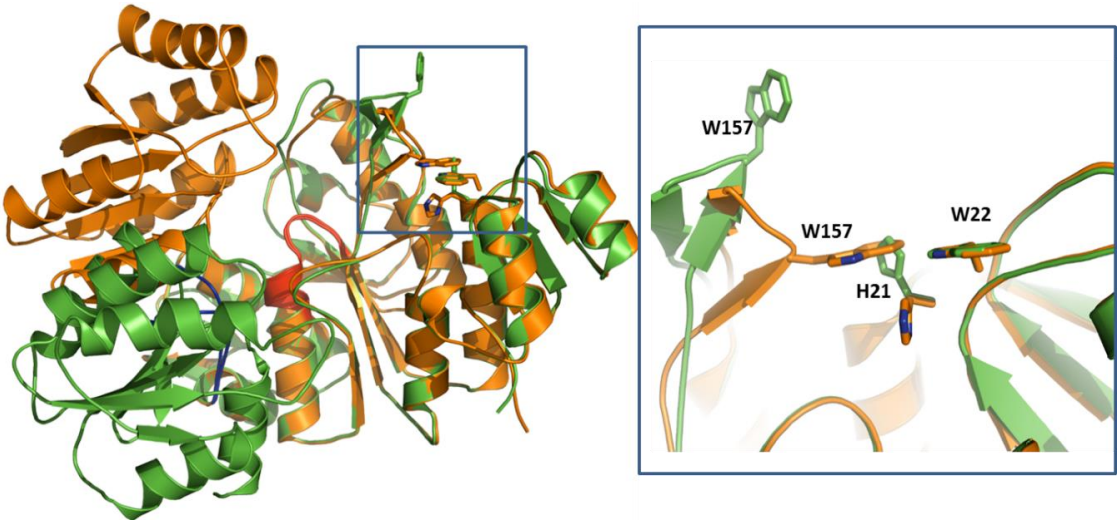


Figure 3.22 Comparative analysis of two separate molecules of Mth734. The ‘gateway-like’ conformation adopted by His21 and Trp157 resembles the structural analysis for the Mfer762 structures and are shown as sticks within the blue box. The Mth734_A is shown as orange and

Mth734_B as green. The closer view of the region forming a 'gateway-like' conformation is shown at the right and residues are labelled as single letter amino acid representation.

3.7 Structural comparison of pMurE

Mfer762 and Mth734, both members of the pMurE peptide ligase type, are structurally similar to one another, and share 54% sequence identity. The overall domain architecture of both enzymes is the same suggesting that the three domain architecture of the pMurE in pseudomurein-containing methanogens and MurE in bacteria is conserved. Among the two recently determined pMurE structures, the Mth734 protein is slightly smaller in size compared to Mfer762, an estimated molecular weight of 48.8 kDa versus 53.5 kDa, respectively, and the difference is contributed through an additional 26 amino acids at the N-terminus of Mfer762. The N-terminal domain in Mth734 contains one less β strand resulting in a parallel β-sheet flanked by alternating α-helices (Figure 3.23A). Topology analysis of the N-terminal domain of the Mth734 and the Mfer762 structures shows that the antiparallel β1-strand of Mfer762 is absent in the Mth734 structure, which appears to be an effect of a shorter amino acid sequence at the N-terminal domain (Figure 3.23). The N-terminal domains of the available bacterial MurE structures also contain a β-sheet with five β strands of which one strand is arranged in antiparallel fashion. It has been studied that the truncation of initial 24 amino acids at the N-terminal domain of the *M. tuberculosis* MurE does not affect the activity or the substrate specificity (Basavannacharya, *et al.*, 2010b), which could be the case for pMurE as well. The presence of the naturally occurring shorter N-terminal domain in Mth734 when taken together with Mfer762, yet catalysing the similar reaction, suggests that the initial 20 amino acid residues in the Mfer762 might not be crucial for enzyme activity or substrate specificity.

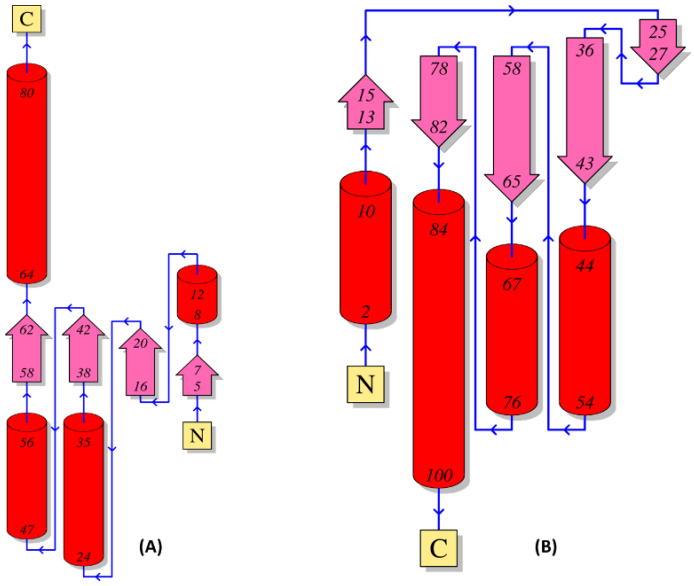


Figure 3.23 Topology diagram of the N-terminal domain of the Mth734 and Mfer762 structures. (A) The topology diagram of the N-terminal domain of the Mth734 structure, (B) The topology diagram of the N-terminal domain of the Mfer762 structure.

The Mth734_A is more suitable for comparative structural analysis with the Mfer762 structure as both the structures adopt a similar C-terminal domain conformation and this is supported by a lower RMSD value (1.2 Å for 442 Cα atoms) (Figure 3.24). Each domain of the Mth734 and Mfer762 structures was compared separately and it was observed that the structures were essentially similar, with the RMSD value between the N-terminal domains of two structures being 0.6 Å for 77 Cα atoms. Further, the middle and C-terminal domains in both the structures are also well conserved with a RMSD values of 1.1 Å for 208 Cα atoms residues and 0.8 Å for 142 Cα atoms, respectively. The structural analysis and structure-based sequence alignment suggest that the UDP-binding sites observed in the Mfer762 structure are well-conserved in the Mth734 structure and will be discussed in the next Section 3.7.1.1. Furthermore, the Mth734 structure has identical residues in the P-loop as well as the hinge regions between the domains. The structural-based sequence alignment confirms that the hinge region between the middle domain and C-terminal domain is

conserved in both the structures (Figure 3.36). The structures of both Mth734 and Mfer762 contain cis-proline residues (Pro312, Mth734 numbering) exactly in the same position and this proline residue is conserved (presumably cis) for all pMurE sequences in pseudomurein-containing methanogens.

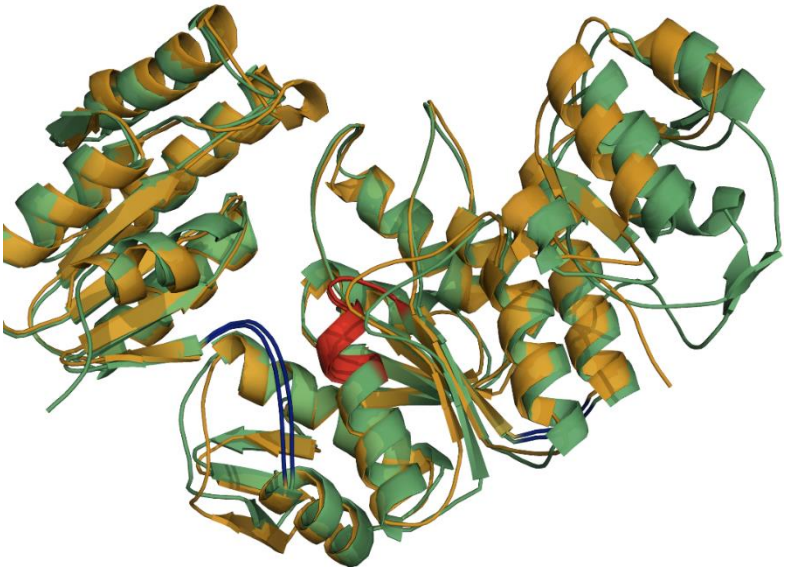


Figure 3.24 Secondary structure-based superposition of Mfer762 and Mth734_A (green and orange, respectively). The hinge regions between each domain are shown in blue and the P-loop is shown in red for both structures.

3.7.1.1 UDP-binding site in pMurE

The overall structure of Mth734_A is similar to Mfer762 and is structurally conserved along the middle domain and C-terminal domain. It was observed that the N-terminal domain in Mth734 is slightly different than Mfer762 and is 20 amino acids shorter in length (Section 3.7). Moreover, it is observed that both the pMurE structures from separate organisms determined during this project have conserved UDP-binding sites. The UDP-binding site at the exterior surface of the enzyme also appears to be conserved in both the structures. Even though the Mth734 structure in complex with UDP could not be determined during this

project, the structural superposition of the UDP-bound Mfer762 on to Mth734 suggests a nearly equivalent interacting environment for UDP in the Mth734 structure. The structural comparison of the Mfer762-UDP_1 structure with the Mth734_A indicates that the amino acid residues involved in the UDP binding at the exterior of the enzyme surface have similar amino acid residues, Lys183 and Lys84 of the Mfer762 structure are substituted by Arg163 and Arg64 in the Mth734 structure, respectively. In addition, Leu181 is substituted by Val161, and other than that surrounding residues within 4.0 Å distance from the UDP molecule of Mfer762-UDP_1 are conserved in the Mth734 structure and are shown in Figure 3.25.

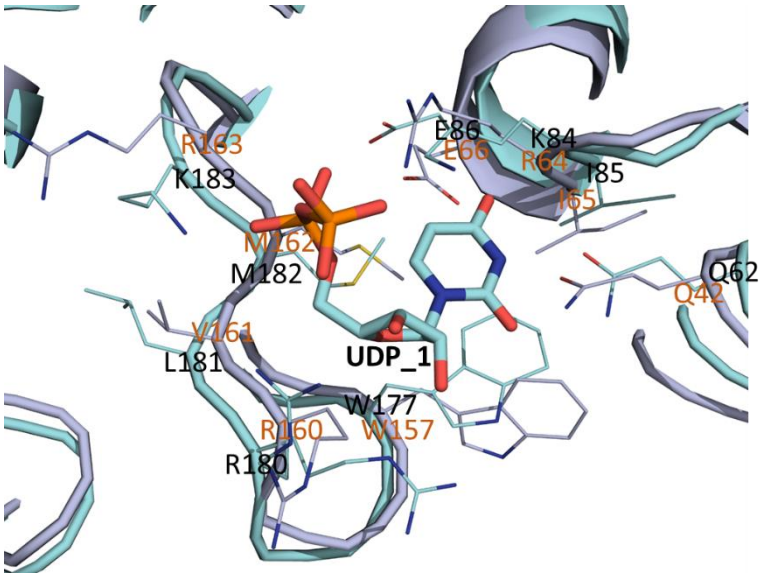


Figure 3.25 Proposed external UDP-binding site in the Mth734 structure. Structural comparison between Mth734 (light blue) and Mfer762-UDP_1 (cyan) shows that the UDP-binding sites observed in the Mfer762-UDP_1 structure appear to be conserved in the Mth734 structure. The structural superposition compared the UDP-binding pocket in the two structures and the UDP molecule from the Mfer762-UDP_1 structure is represented as sticks in the figure (UDP_1). The residues within 4.0

Å from the UDP molecule are shown as lines and labelled orange and black for the Mth734 and the Mfer762 structures, respectively.

Moreover, the amino acid residues in the UDP-binding pocket at the inner surface of the Mth734 structure remain completely conserved when compared to the Mfer762-UDP_2 structure and are shown in Figure 3.26. Structurally conserved amino acid residues at the inner surface of the Mth734 structure suggest that it is the conserved binding site that could be required for substrate binding in pMurE. There is an additional residue, Trp22, within the environmental contact distance of 3.5 Å from the UDP molecule of the Mfer762-UDP_2 structure in the Mth734 structure, which is another conserved residue of pMurE and corresponds to the Trp42 in the Mfer762 structures. This is the residue that goes through a rotamer switch between the apo and UDP-bound states in the Mfer762-UDP_2 structure. Despite the large conformational change between the two molecules of the Mth734 structure, the presence of the same rotamer conformation of the Trp42 (Mfer762 numbering) as that for the Mfer762 apo or the Mfer762-UDP_1 structures in the Mth734 structure suggests that the residue might be involved in UDP recognition at the inner UDP_2-binding site. This could support the hypothesis that the Trp42 might undergo a rotamer switch only in the presence of UDP or the complete presumed substrate, N^{α}-UDP-Glu$^{\gamma}$-Ala.

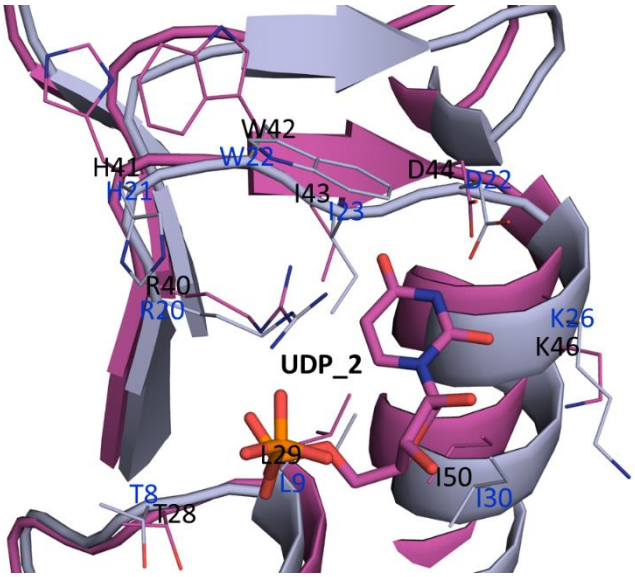


Figure 3.26 Proposed internal UDP-binding site at the crescent cavity of the Mth734 structure. The residues within 4.0 Å from the UDP molecule of the Mfer762-UDP_2 structure (magenta) are compared with the Mth734_A structure (light blue) and shows conserved UDP_2-binding site in Mth734. The UDP molecule is represented as sticks and the residues within 4.0 Å from the UDP molecule are shown as lines and labelled blue and black for the Mth734 and the Mfer762 structures, respectively. An alternative conformation of Trp22 is observed in the Mth734 structure unlike the Mfer762-UDP_2 structure and is similar to Mfer762 apo structure.

3.7.1.1.1 'Gateway-like' conformation among Mfer762 structures

The observation of two different UDP-binding sites in separate Mfer762 structures reveals that these UDP-binding sites are adjacent to one another, separated by a flexible 'extended loop' that extends slightly from the middle domain towards the N-terminal domain. The 'extended loop' contains polar and charged residues that are mostly conserved among the pMurE sequences in pseudomurein-containing methanogens. Among them the side-chain of Trp177 flanks with the side-chain of His41 at the N-terminal domain in both apo and UDP_1-bound structures. These two residues (Trp177 and His41) are the key residues that form the 'gateway-like' structure through their side-chain conformations (Figure 3.27).

The ‘gateway-like’ conformation appears to be a control mechanism for the binding of the substrate into the putative active site, represented by a UDP molecule in this study (UDP_2). The side-chain of Trp177 forms a face-to-face π-π stacking interaction with the side-chain of His41 in the Mfer762 apo structure. Whereas in the Mfer762-UDP_1 structure, the side-chain of Trp177 changes to form a face-to-face π-π stacking interaction with the uracil ring of the UDP molecule and slightly disrupts the face-to-face arrangement of two side-chains. This rearrangement of the t-90 rotamer of Trp177 results in the switch of rotamer for His41 from m-70 (Mfer762 apo) to p80 that prevents steric hindrance. Both the Mfer762 apo and Mfer762-UDP_1 structures have the residues forming the ‘gateway-like’ conformation in a closed fashion so that a distinct cavity at the inner surface to bind the UDP molecule is not formed (Figure 3.27).

Further, the structural comparison with the Mfer762-UDP_2 structure suggest that the N-terminal domain moves slightly away from the middle domain compared to that observed with apo and UDP_1-bound structures. In addition, four amino acid residues, including Trp177, of the ‘extended loop’ in the Mfer762-UDP_2 structure could not be modelled suggesting its flexible behaviour and conformational change upon substrate binding. This observation suggests that the Mfer762 requires the movement of the ‘extended loop’ away from the N-terminal domain resulting in the disrupting the π-π stacking interaction with both the ‘gateway-like’ conformation forming His41 of N-terminal domain and the uracil head of the UDP molecule and therefore allowing access to the substrate for binding into the putative inner catalytic cavity. The movement of the N-terminal domain away from the Mfer762 apo structure and the missing of the extended loop in the Mfer762-UDP_2 structure suggest that the ‘gateway-like’ conformation could also be contributing to the open and closed conformations of the enzyme. It can further be suggested that the UDP_1-binding site could be a molecular-regulatory site rather than a catalytic site. It is not clear if the regulatory site

is specific to UDP, but the current structural observation suggests a possibility of a nucleotide compound.

Hence, it is hypothesised that the highly conserved ‘extended loop’ in the pMurE structure is responsible for the formation of the ‘gateway-like’ conformation, which is required for the structural stability and to maintain the open versus closed conformation that are essential for the ligase activity.

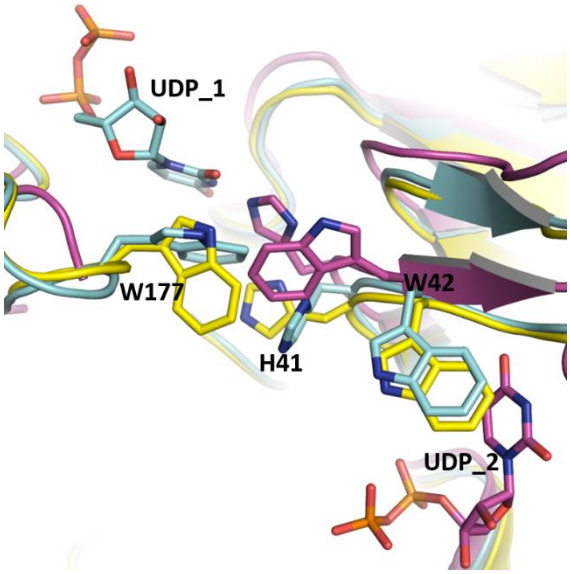


Figure 3.27 The two different modes of UDP-binding in Mfer762 structures. The Mfer762-UDP_1 is shown in cyan, Mfer762-UDP_2 in magenta and apo Mfer762 in yellow. The side-chain of Trp177 and His41 forms the ‘gateway-like’ structure that forms face-to-face π-π stacking interaction with each other in the Mfer762 apo structure, whereas the Trp177 adopts the conformation favouring a similar face-to-face π-π stacking interaction with the uracil ring of UDP in the Mfer762-UDP_1 structure. The side-chain of Trp42 that is observed to contribute to generating a larger space for UDP

interaction in Mfer762-UDP_2 structure is also shown as sticks. All residues are labelled using single letter amino acid representation.

3.7.1.2 Proposed ATP-binding site for pMurE

None of the four pMurE structures determined during this study contained either an ATP or adenosine nucleotide molecule bound to the structures. All mono-nucleotide binding enzymes contain the P-loop motif, GxxxxGK(S/T), where x denotes variable residues. This motif is responsible for α- and β-phosphate binding (Pathak *et al.*, 2014). The middle domain is the most conserved domain of all bacterial murein ligases and possesses conserved ATP-binding residues (Smith, 2006). In addition to the invariant residues, multiple sequence alignment and structural analyses shows that the P-loop in pMurEs share a consensus sequence similar to that present in bacterial MurE type ligases (Smith, 2006). Structure superposition analyses have been used to compare the ATP-binding site of the two pMurE structures. It was observed that the proposed ATP-binding site is structurally conserved with a similar location and forming similar interactions at the interface of middle and C-terminal domain as studied for available murein ligase structures. In addition to the conserved P-loop, other residues putatively involved in ATP-binding identified in Mfer762 appear to be conserved in Mth734 as well (Figure 3.28). Except for Tyr290, all other residues of the proposed Mfer762 ATP-binding site remain invariant in the Mth734 structure. The Tyr290 in Mfer762 is substituted by His269 in Mth734 (as in the bacterial ATP-binding site, Figure 3.15). The presence of i) conserved P-loop and its involvement in ATP-binding, ii) structural homology and conserved ATP-binding site (further to P-loop) compared with bacterial MurE/F ligases and among pMurE structures and iii) the presence of a presumed phosphate ion (PO_4^{3-}) at the β-phosphate position of the ATP analogues in bacterial murein ligase

structures suggest that the ATP-binding site is conserved in both murein ligase and pseudomurein ligase structures.

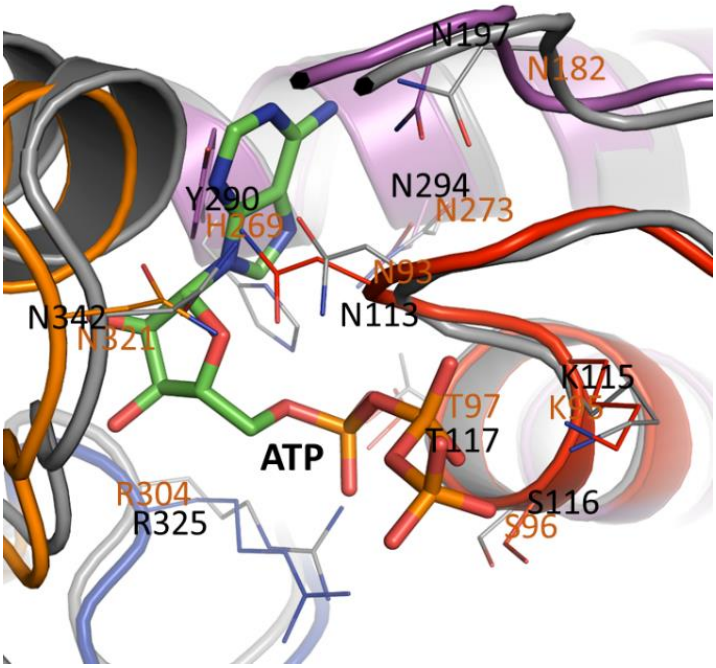


Figure 3.28 Proposed ATP-binding site of the Mth734 structure. The ATP-binding site for Mfer762 has been proposed based on homology with bacterial structures (Section 3.6.1.2.3, Figure 3.15). The proposed ATP-binding site in Mfer762 is used as a reference to model the putative ATP-binding residues in Mth734. The Mth734 structure is coloured grey and the proposed ATP-binding residues are labelled in blue.

3.8 Structure similarity analysis

The Dali server (Holm & Rosenström, 2010) was used to search for structural homologues of pMurE members, Mfer762 and Mth734, against the PDB database. Both of the structures showed similar results, therefore, only the Mfer762 results will be referred in the text for clarity and brevity. The overall structure comparison shows that the Mfer762 structure best aligns with the bacterial MurE/F type peptide ligases. Among the murein ligases, the MurF from *Acinetobacter baumannii* in complex with ADP (4QF5) showed the highest structural

similarity with Mfer762 with a Z-score of 31.8 and a RMSD value of 3.5 Å for 416 aligned Cα atoms and a sequence identity of 18%. The highest sequence identity was exhibited by the MurE of *Staphylococcus aureus* (4C13) and was calculated to be 20%. In addition to the murein ligases MurC-MurF, several folylpolyglutamate synthases also appeared with a significant Z-score. Among these, the folylpolyglutamate synthases from *Yersinia pestis* CO92 (2N2A) showed the highest structural similarity with a Z-score of 25.3 (RMSD of 3.5 Å for 343 Cα atoms and a sequence identity of 18%). The output from the Dali search was perplexing during initial inspection because a few bacterial MurC/D types also compare well with relatively similar RMSD values (Table 3.2).

However, separate structure-based alignment of the two different types (MurC/D and MurE/F) show that MurC/D types are more conserved within their middle and C-terminal domains, whereas the MurE/F types are conserved throughout the entire structure. It is therefore reasonable to have the MurC/D type showing structural similarity with the MurE/F type as the ATPase middle domain comprises the largest percentage of the structure and is the most conserved domain of all the murein peptide ligases. The Mfer762 structure is distinct from the MurC/D types at the N-terminal domain where the structures do not clearly align with one another showing that Mfer762 and Mth734 are E/F type peptide ligases. The N-terminal domain is also the most distinct domain of bacterial murein peptide ligases and has been categorised as MurC/D type and MurE/F type based on their structural details (Smith, 2006).

Table 3.2 Structure similarity statistics for the Mfer762 structure and its individual domains obtained using Dali. The values in each cell represents the Z-score/RMSD (Å) units and number of aligned Cα atoms in parenthesis. The Z-score indicates the degree of confidence in structural similarity – the higher the Z-score, the more the confidence. The organism of which the enzyme types are used have been abbreviated as *Yersinia pestis* (Yp); *Thermotoga maritima* (Tm); *E. coli* (Ec); *Pseudomonas aeruginosa* (Pa); *Streptococcus agalactiae* (Sa); *Staphylococcus aureus* (Sca); *Acinetobacter baumannii* (Ab); *Mycobacterium tuberculosis* (Mt); *Lactobacillus casei* (Lc); *Haemophilus influenzae* (Hf)

	Structural similarity with bacterial MurC				
	Hf-MurC (1P3D)	**Hf-MurC-apo (1GQY)**	**Yp-MurC (4HV4)**	**Tm-MurC (1J6U)**	**Ec-MurC (2F00)**
Mfer762	28.9/3.5 (354)	28.7/3.4 (353)	28.2/3.6 (350)	21.4/4.0 (332)	27.8/3.5 (355)
N-terminal domain	N/A	N/A	N/A	N/A	N/A
Middle domain	20.3/2.3 (178)	19.9/2.3 (172)	20.2/2.3 (171)	17.8/2.7 (168)	20.4/2.3 (176)
C-terminal domain	14.5/2.5 (131)	14.4/2.5 (131)	14.5/2.6 (133)	13.2/2.2 (115)	12.3/2.6 (126)
	Structural similarity with bacterial MurD				
	Ec-MurD (1EEH)	**Ec-MurD (1UAG)**	**Ec-MurD (2JFF)**	**Sa-MurD (3LK7)**	**Tm-MurD (4BUC)**
Mfer762	21.5/4.8 (242)	24.1/3.1 (341)	24.4/3.1 (344)	21.8/4.8 (272)	18.7/4.8 (253)
N-terminal domain	N/A	N/A	N/A	N/A	N/A
Middle domain	20.3/2.7 (182)	20.8/2.4 (178)	21.2/2.4 (179)	21.1/2.3 (185)	17.9/2.6 (177)
C-terminal domain	11.3/2.5 (116)	11.2/2.6 (116)	11.3/2.6 (117)	11.4/2.4 (118)	11.9/2.5 (116)
	Structural similarity with bacterial MurE				
	Mt-MurE (2XJA)	**Tm-MurE (4BUB)**	**Sca-MurE (4C12)**	**Mt-MurE (2WTZ)**	**Ec-MurE (1E8C)**

Mfer762	27.6/3.7 (422)	26.2/3.4 (395)	26.5/3.2 (407)	26.7/3.7 (435)	28.0/3.6 (427)
N-terminal domain	10.3/2.4 (93)	9.2/2.4 (87)	8.8/2.6 (87)	10.2/2.5 (95)	6.6/2.9 (81)
Middle domain	21.2/2.6 (196)	19.7/2.7 (188)	22.2/2.4 (196)	20.8/2.5 (195)	20.3/2.6 (194)
C-terminal domain	13.5/2.9 (132)	15.3/2.2 (129)	13.8/2.8 (130)	13.5/2.8 (133)	15.2/2.4 (132)
	Structural similarity with bacterial MurF				
	Tm-MurF (3ZL8)	**Ec-MurF (1GG4)**	**Pa-MurF (4CVK)**	**Ab-MurF (ATP-bound) (4QF5)**	**Ab-MurF(ATP-UDP-bound) (4QDI)**
Mfer762	30.8/3.0 (427)	26.3/5.0 (372)	27.9/4.4 (379)	31.7/3.5 (453)	28.3/4.8 (379)
N-terminal domain	4.8/3.0 (69)	7.7/2.7 (89)	10.3/2.4 (89)	N/A	8.1/2.7 (90)
Middle domain	21.9/2.0 (191)	21.4/2.2 (182)	23.5/2.1 (187)	25.2/1.9 (193)	24.4/2.1 (195)
C-terminal domain	11.7/2.7 (117)	12.1/2.2 (113)	11.9/2.6 (116)	12.3/2.7 (122)	11.9/2.7 (122)
	Structural similarity with bacterial FolC				
	Mt-FolC (2VOR)	**Lc-FolC (1JBW)**	**Yp-FolC (3QCZ)**	**Ec-FolC (1W78)**	**Tm-FolC (1O5Z)**
Mfer762	23.7/3.4 (342)	24.3/3.5 (333)	25.0/3.7 (344)	24.6/3.6 (341)	21.6/4.3 (338)
N-terminal domain	NA	NA	NA	NA	NA
Middle domain	16.0/2.5 (187)	17.5/2.6 (190)	15.8/2.8 (183)	16.0/2.7 (181)	16.7/2.5 (183)
C-terminal domain	12.5/2.8 (126)	11.1/2.6 (106)	13.9/2.3 (122)	13.8/2.3 (123)	12.8/2.5 (121)

Folylpolyglutamate synthetase (FPGS) also shares structural similarity with the middle and C-terminal domains of murein ligases in bacteria (Smith, 2006). The structures of FPGS lack an N-terminal equivalent domain compared to the murein ligases and instead includes an extra amino acid insertion sequence (approximately 50 residues) that is positioned at the middle domain compared to murein ligases (Sheng, *et al.*, 2000). The Dali webserver result suggested that the members of FPGS also share structural similarity with the Mfer762 structure with a RMSD values of 3.1 to 3.5 Å and are close to other murein ligases (Table 3.2). In addition, the FPGSs show similar insertion of amino acids when compared with MurD at the middle domain, approximately 25 amino acids away from the conserved P-loop motif (Figure 3.29) (Sheng, *et al.*, 2000). A similar observation was obtained while comparing pMurE with the FPGS structures suggesting a similar relation of FPGS to the pMurE as that with the murein ligases.

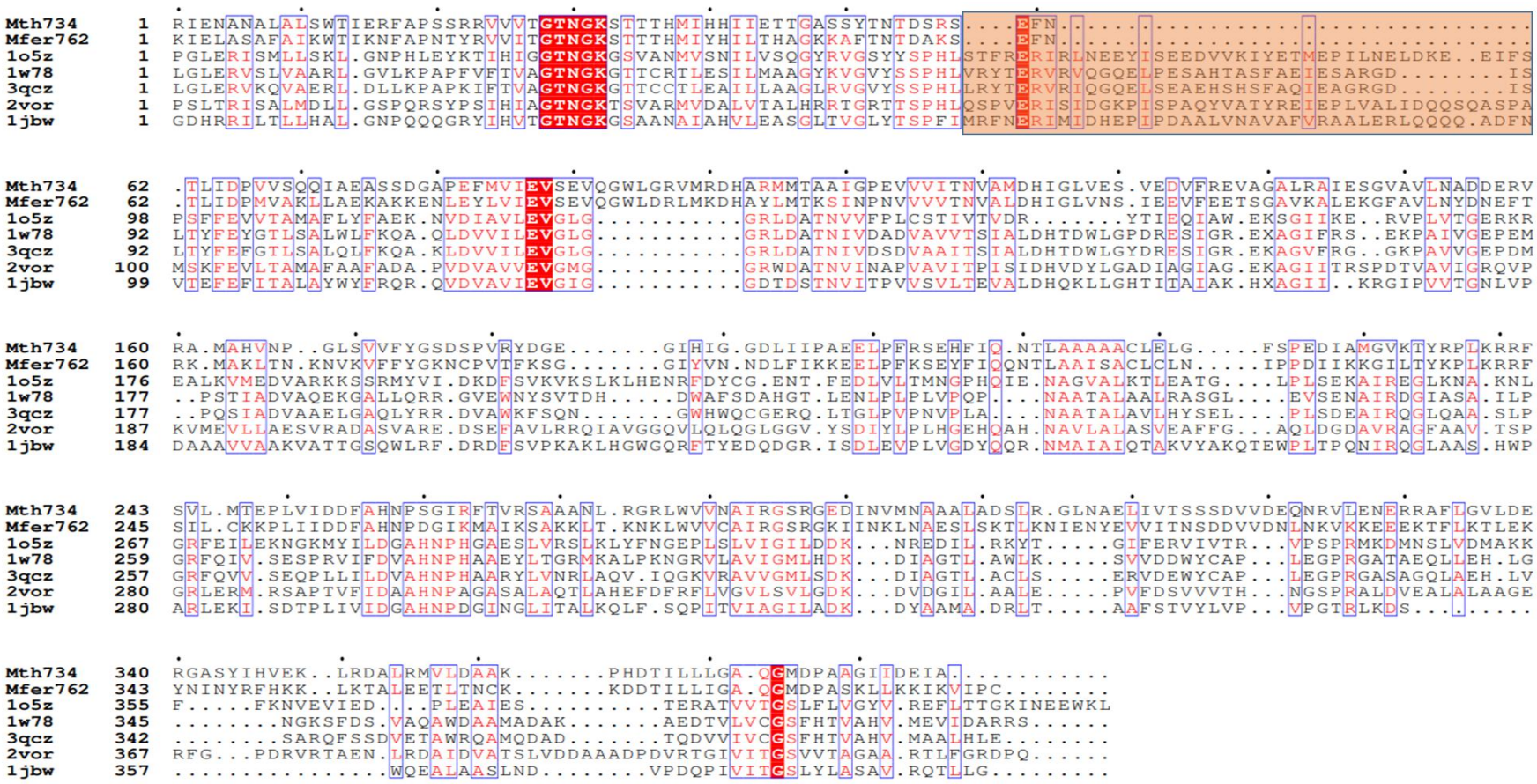


Figure 3.29 Structure-based sequence alignment of the pMurE and different FPGS from bacterial species. The insertion region is highlighted with orange box. Multiple structure based alignment was performed using PROMALS3D webserver (Pei & Grishin, 2014) and figure was created using the ESPrit3.0 webserver (Robert & Gouet, 2014). PDB codes for bacterial FPGS are same as that used in Table 3.2 for structure similarity analysis.

Each domain of the Mfer762 structure was analysed separately to identify the highest scoring structural homologues using the Dali server. The N-terminal domain analysis of the Mfer762 shows that it shares structural similarity with the N-terminal domain of the available MurE and MurF structures. The MurF N-terminal domain from *Pseudomonas aeruginosa* (4CVK) structure shares the highest structure similarity with a Z-score of 10.3 and a RMSD value of 2.4 Å for 89 aligned Cα atoms and a sequence identity of 20%. The N-terminal domain in pseudomurein ligases also appears to be the most distinct of three domains as studied in murein peptide ligases which can be classified as pMurC/D and pMurE/F types (Smith, 2006). The N-terminal domain of MurC/D type does not appear in the list of structure with the Z-score value as low as 2.0 indicating that the Mfer762 structure is clearly more related to a MurE/F type ligase.

The structural similarity search result with the middle domain suggests that the domain is significantly conserved to the middle domain in bacterial murein peptide ligase structures (MurC-F). The highest structural similarity was exhibited by the MurF from *Acinetobacter baumannii* (4QF5) with a Z-score of 25.2 and a RMSD value of 1.9 Å for 193 aligned Cα atoms and a sequence identity of 22%. The middle domain of available MurF structures showed higher structural similarity followed by MurE, compared to that of MurC/D as indicated by the Dali result (Table 3.2). The obtained result indicates that the Mfer762 middle domain share a higher level of conservation to the bacterial MurE/F peptide ligase type. In addition, the middle domain also shows structural similarity to FPGS.

Similarly, the C-terminal domain also showed structural similarity with a wide range of murein peptide ligases and FPGS. The C-terminal domain of the Mfer762 structure exhibits the highest structural similarity with the C-terminal domain of the MurE from *T. maritima* (4BUB) with a Z-score of 15.3 and a RMSD value of 2.2 Å for 129 aligned Cα atoms and a sequence identity of 19%. Further, among the FPGS structures, the C-terminal domain of

FPGS from *Yersinia pestis* C092 (3NRS) shares the highest structural similarity to the C-terminal domain of Mfer762 with a RMSD value of 2.3 Å for 193 aligned Cα atoms and a sequence identity of 22% and a Z-score of 14.5. This suggests that the C-terminal domain in Mfer762 is related to the bacterial MurE and further structural similarity with the C-terminal domain of FPGS enzyme supports an early evolutionary link. The structural similarity search result for the Mth734 structure showed similar results with slight differences in the Z-scores and RMSD values (not shown).

3.9 Structure comparison with bacterial murein type

3.9.1 Comparison with MurF

The bacterial MurF structure from *Acinetobacter baumannii* (4QDI) was used as a model for molecular replacement to solve the structure of Mfer762. The final model of Mfer762 has structural similarity to the 4QDI with the middle domain possessing the most conserved secondary structure of the overall structure. In addition, 4QDI contains UDP and ATP molecules bound to the structure. The presence of UDP as a partial substrate in Mfer762 makes the comparison with 4QDI more sensible in analysing the UDP-binding site and identifying any conserved residues in pseudomurein peptide ligases in comparison to murein peptide ligases. Structure-based similarity search using Dali suggests that MurF from *A. baumannii* (4QF5) as the most similar of solved bacterial murein ligase structure. The 4QF5 structure has only ATP bound compared to 4QDI that binds both ATP and UDP. It is important to note that the ATP-binding mode in 4QDI is incomplete due to the absence of Mg^{2+} and a carbamoylated lysine (Cha, *et al.*, 2014). All bacterial murein ligase types except for MurC contain a carbamolyated lysine which is required for substrate binding (Dementin,

et al., 2001; Smith, 2006), however, no carbamolyated lysine has been observed in either of the solved structures of pMurE.

It has been reported that MurE/F peptide ligases share similar domain architectures, especially considering the N-terminal domain (Smith, 2006) and therefore, both murein peptide ligase types have been used for comparative analysis. The comparative analysis with the MurE type will be discussed in the next section (Section 3.9.2). In bacterial MurF, ATP binding is the first event of the catalysed reaction, which is followed by binding of the UDP-*N*-acetylmuramyl-tripeptide at the crescent cavity (Cha, *et al.*, 2014). The binding of UDP in the bacterial MurF structure leads to a conformation such that the uridine base is buried into the N-terminal domain with the phosphate pointing towards the large cavity in the interface of three domains. A similar conformation of UDP binding is observed in the Mfer762-UDP_2 structure.

The secondary structure superposition of Mfer762 and bacterial MurF (4QDI) structures shows that the N- and C-terminal domains in both structures have higher RMSD values compared to that of the middle domain. The middle domain is the most conserved domain in all the murein types and this also applies for Mfer762 (Figure 3.30). Even though no ATP bound pseudomurein ligase structure has been solved, homology modelling with the 4QDI structure has been successful in predicting the putative ATP-binding site in Mfer762 and Mth734 structure. The presence of a presumed PO_4^{3-} ion bound at the α-phosphate position of the ATP molecule present in the 4QDI structure and conserved amino acids forming the ATP-binding site in Mfer762 support the presence of a similar binding site in pseudomurein ligase structures (refer Sections 3.6.1.2.3 and 3.7.1.2). The amino acid residues involved in ATP interaction in all the bacterial murein ligases are conserved and have been highlighted (Smith, 2006). These residues are also conserved in both Mfer762 and Mth734 and adopt a

similar position of interaction - at the interface of middle and C-terminal domains, as observed for 4QDI and other murein ligases.

The UDP-binding mode at the external surface (UDP_1) is entirely different and unique to Mfer762 whereas the UDP-binding site at the internal surface (UDP_2) shares similar position for UDP binding, yet adopts a slightly different conformation by the UDP molecule compared to the 4QDI structure. These observations are similar with other members of the bacterial MurE/F whose structures have been determined.

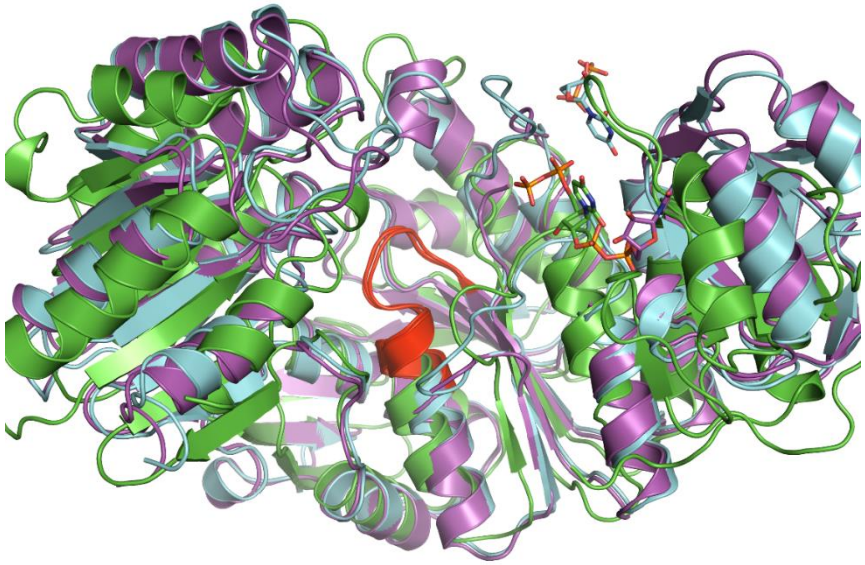


Figure 3.30 Structural superposition of Mfer762-UDP_1 (cyan), Mfer762-UDP_2 (magenta) and MurF structure from *A. baumannii*, 4QDI, (green) showing bound UDP. The UDP in all three structures binds close to the N-terminal – middle domain boundary but at different positions. The P-loop of Mfer762 is well-conserved compared to the bacterial MurF structure; shown in red. The middle domain is the most conserved of the three domains, however, the N- and C- terminal domains also indicate structural homology of Mfer762 to bacterial MurF.

The comparative analysis of the Mfer762 and 4QDI structures suggests that an ‘extended loop’ in the Mfer762 structure is formed by an insertion of the amino acid sequence that extends from the middle domain slightly towards the N-terminal domain. This ’extended

loop' in Mfer762 contributes to UDP binding at the external surface of the Mfer762-UDP_1 structure. A similar insertion is observed at the N-terminal domain of the 4QDI structure, amino acids 57-62, which provides a deep cavity conformation for the UDP-binding pocket in 4QDI and is shown in Figure 3.31. These insertion regions are adjacent to one another and are present in two different domains, yet both are contributing in UDP-binding. It is interesting that the insertion region in Mfer762 creates a narrow space between the N-terminal and middle domains, whereas the 4QDI structure has a relatively larger space around the interface of the two domains.

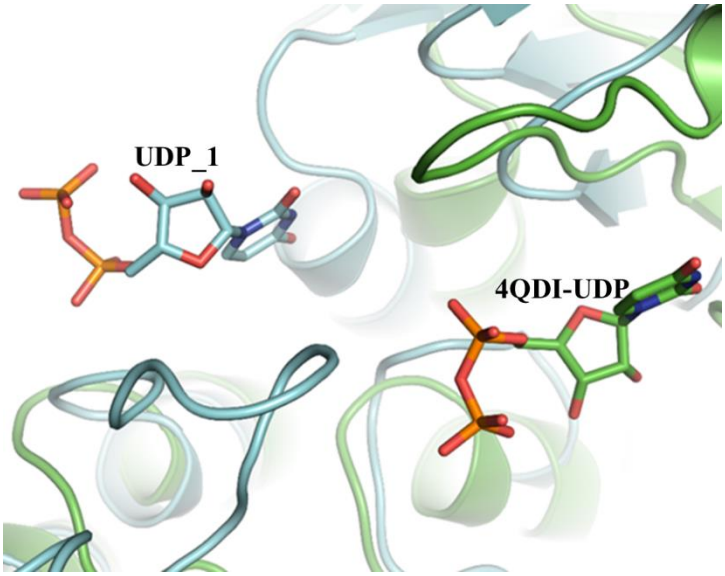


Figure 3.31 Structural comparison of Mfer762-UDP_1 (cyan) and 4QDI (green) demonstrating the different positions for binding of the UDP in two structures. The UDP-binding site in Mfer762-UDP_1 is different from the 4QDI and is formed in part by the 'extended loop' (cyan at left). Similarly, the insertion region is also observed in 4QDI (green at right) compared to Mfer762-UDP_1. The position of these UDP-binding sites are on two different sides of the enzyme surface.

Moreover, the Mfer762-UDP_2 structure has the UDP-binding mode in the position similar to that of the 4QDI structure, yet possesses different interacting residues and a slightly different conformation as shown in Figure 3.32. The binding of the UDP in the Mfer762-

UDP_2 structure points the uridine base deep into the N-terminal domain whereas the phosphate end is extended towards the crescent cavity as explained earlier in Section 3.6.1.2.2.1 above. In addition, the regions 139-150 is presumably disordered and could not be modelled in the Mfer762-UDP_2 structure suggesting a conformational change may occur in the structure enabling the binding of a large substrate in the cavity.

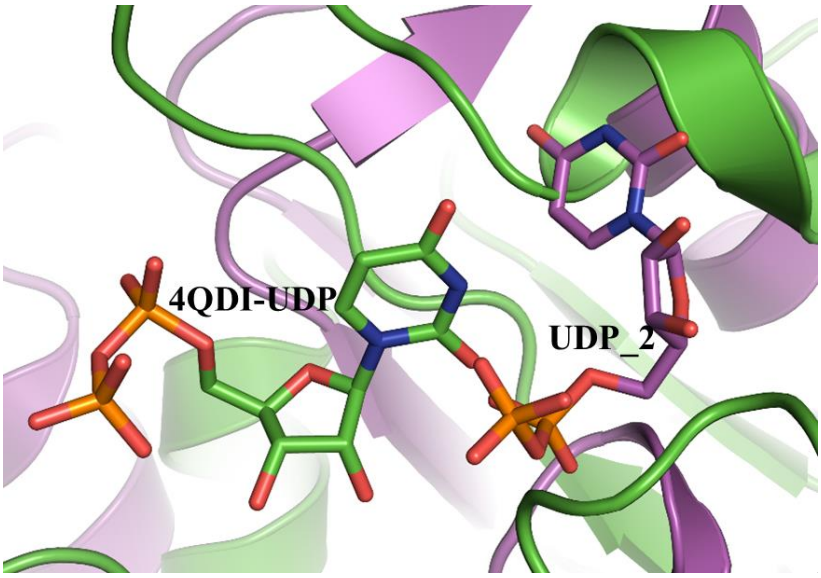


Figure 3.32 Structural comparison of Mfer762-UDP_2 (magenta) and 4QDI (green) demonstrating the position for UDP binding in the two structures. The UDP-binding site in Mfer762-UDP_2 is closer and in a similar conformation as compared to 4QDI. The UDP-binding site in Mfer762-UDP_2 is deeper into the N-terminal domain compared to the 4QDI structure whereas the extended loop (green at right) provides a deep cavity for the UDP binding in the 4QDI structure. The position of these UDP-binding sites are on same side of the enzyme surface.

3.9.2 Comparison with MurE

The structural similarity search from Dali webserver (Holm & Laakso, 2016) suggests that Mfer762 shares the highest structural similarity to MurE from *Mycobacterium tuberculosis* (2XJA) of the bacterial MurE type. The Mfer762 structure shows structure similarity to the 2XJA structure based on overall structure as well as individual domains (Table 3.2). The

structural similarity searches for individual Mfer762 domains suggest that the Mfer762 is more like MurE among the MurE/F types. This analysis is supported by the structural similarity result with the Mfer762 C-terminal domain having structural similarity with more MurE structures compared to the MurF structures. Due to the close structural similarity of Mfer762, the 2XJA structure was considered as the template structure for representing the bacterial MurE peptide ligases and performing structural analysis in this study. The 2XJA structure contains the substrate for bacterial MurE (UDP-*N*-acetylmuramoyl-L-Ala-D-Glu (UAG)) and an ADP molecule bound to it. The structural comparison of individual domains of the Mfer762 and 2XJA structures show structural homology to one another (Figure 3.33).

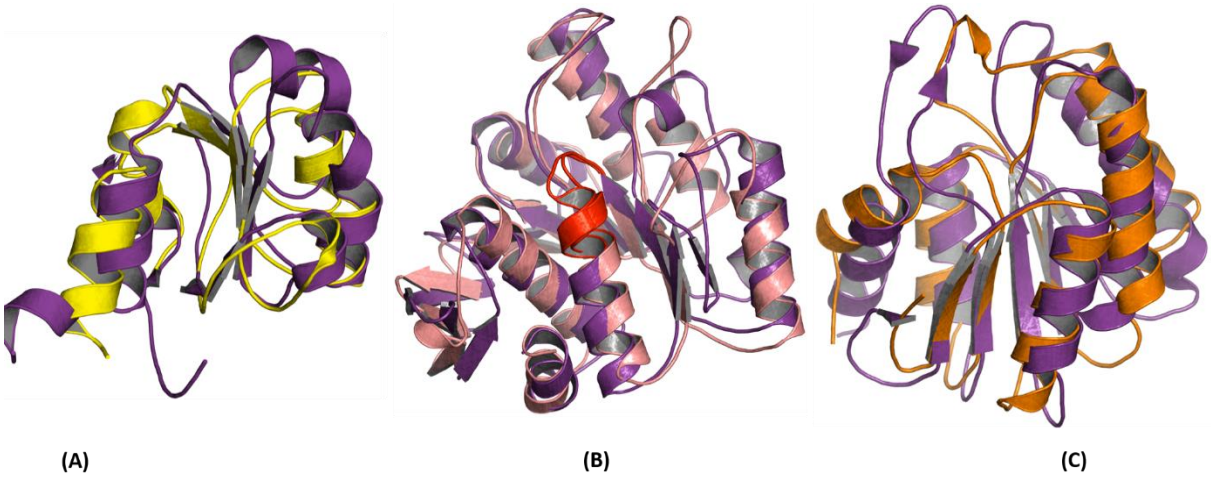


Figure 3.33 Domain comparison of Mfer762 and bacterial MurE (2XJA). (A) The Mfer762 N-terminal domain (yellow) compared to the N-terminal domain of 2XJA. (B) The middle domain comparison with the conserved P-loop shown in red. (C) The C-terminal domain comparison with Mfer762 domain shown in orange.

The structural analysis of Mfer762 in comparison to 2XJA and other available bacterial MurE confirms that the insertion (as observed in comparison to MurF (4QDI)) in the middle domain of Mfer762 structure also applies to the MurE type. The insertion of amino acid residues that form the UDP-binding site in the Mfer762-UDP_1 structure is conserved in

both Mfer762 and Mth734. The residues in this region are also well conserved among pseudomurein-containing methanogens (Figure 3.38). Moreover, the insertion of amino acid sequence in the N-terminal domain of bacterial MurE, in comparison to pMurE, seems to be structurally conserved as indicated in Figure 3.34 and Figure 3.36.

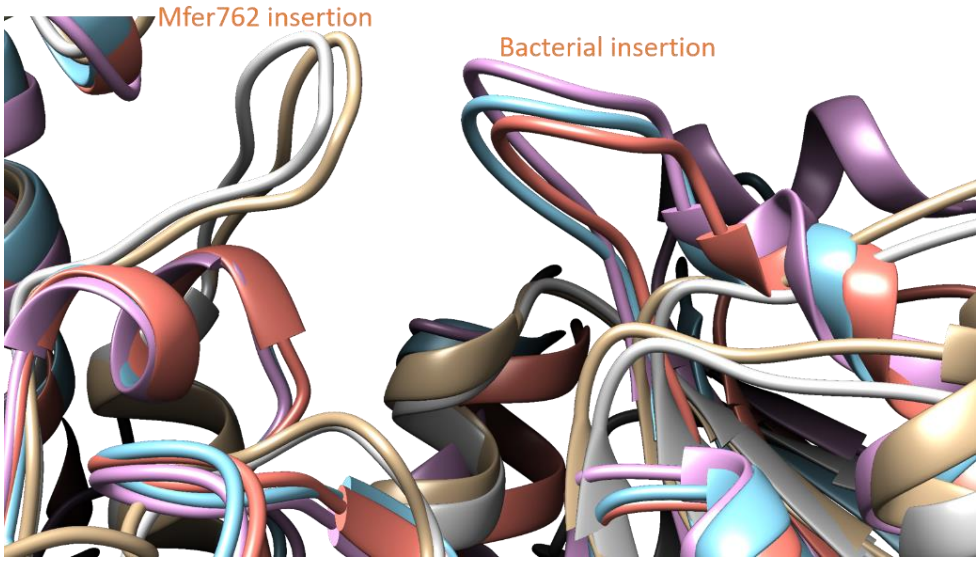


Figure 3.34 Structural comparison of pMurE and MurE, highlighting the insertion region (refer label). The insertion (Mfer762 insertion) for pMurE (Mth734 and Mfer762, indicated as grey and light yellow) creates a UDP-binding pocket as observed in the Mfer762-UDP_1 structure, whereas the inserted loop on the right (bacterial insertion) for MurE (showing 2XJA, 1E8C and 4C12, indicated as cyan, light purple and light orange, respectively) forms the substrate binding pocket and is mostly responsible for incorporating the UDP portion of the substrate (also applies to 4QDI see Figure 3.32).

The signature P-loop required by the Mur family is well-conserved in the middle domain of both pMurE and MurE structures. The P-loop is involved in ATP-binding in murein peptide ligases and binds ADP in the 2XJA structure. In addition, the 2XJA structure binds UAG with the UDP binding to the N-terminal domain whereas the *N*-acetylmuramic acid ring and peptide end extends in the crescent cavity forming interaction with the residues in the middle

domain (Osman *et al.*, 2012). The addition of the L-lysine by MurE occurs at the end of the UAG and occurs at the C-terminal domain, as observed in the MurE structure from *Staphylococcus aureus* (PDB ID- 4C12/4C13) (Ruane *et al.*, 2013). The structural comparison of these bacterial structures with the analysed Mfer762-UDP_2 structure was used to model the bacterial substrate (UDP-MurNAc) into the pMurE putative substrate/product-binding cavity and is shown in Figure 3.35. This mimics the substrate and product binding mechanism as discussed earlier for bacterial structures. The UDP base of the bacterial substrate clashes with the pMurE structure at residues Arg40, His41 and Trp42 and suggests a different conformation for the UDP base as adopted by the UDP_2 structure. The conformation adopted by the Mfer762-UDP_2 structure provides a larger space for the substrate (and product) binding compared to the other two Mfer762 structures analysed in this study and has been discussed earlier in Section 3.6.1.3 and is shown in Figure 3.17.

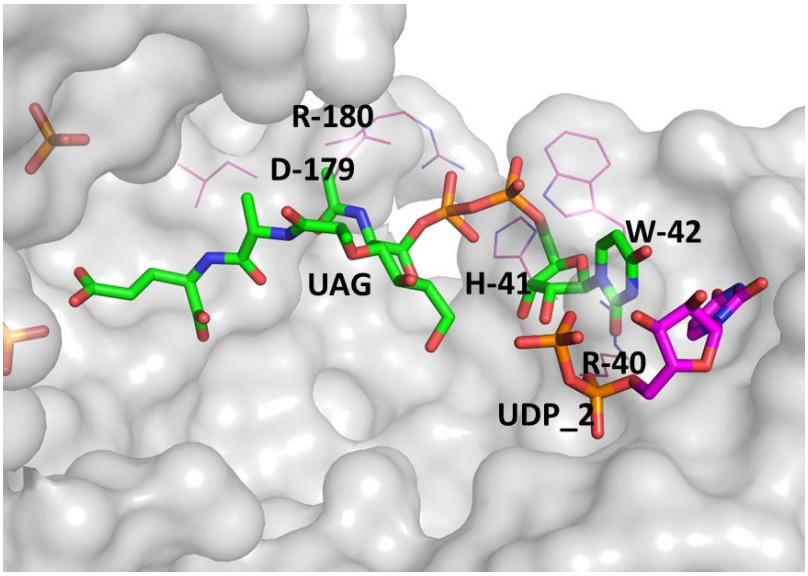


Figure 3.35 Modelling of the bacterial substrate (UAG in green) of the bacterial MurE peptide ligase into the Mfer762-UDP_2 structure (shown as surface representation). The residues that could sterically hinder the position and conformation of the bacterial substrate (UAG) are represented as lines (magenta). In addition, the residues close to the flexible region 174-179 are also shown as lines.

The UDP_2 from the Mfer762-UDP_2 structure is represented as sticks and shown in magenta. The presumed phosphate molecules in the Mfer762-UDP_2 structure are also represented as sticks.

3.10 Biochemical Assay

3.10.1 ADP quantification assay

Based on the proposed pathway for pseudomurein biosynthesis and seven step *in vitro* synthesis of the pseudomurein pentapeptide (Hartmann & König, 1994; Claus & König, 2010), pMurE might be responsible for the addition of the L-lysine to the end of the cross-linking peptide chain as represented in reaction 1 (R1). Despite dedicated attempts, both biochemically using putative enzymes *in vitro* (Sauter, 2017) and synthesis through commercial companies, the initial substrate (N^{α}-UDP-Glu) to which the L-Ala gets ligated could not be obtained, hence, limiting the biochemical testing of the putative pseudomurein ligases. However, based on structural information and analysis of structural homology for pMurE to bacterial MurE, biochemical assay using the bacterial substrate was performed to test the substitution of the pseudomurein peptide ligase substrate (N^{α}-UDP-Glu$^{\gamma}$-Ala) in pMurE by UMA. Moreover, the estimated size of UMA is close to that of N^{α}-UDP-Glu$^{\gamma}$-Ala (Figure 3.37) and successful use of UMA by pMurE could suggest an evolutionary relationship.

$$N^{\alpha}\text{-UDP-Glu}^{\gamma}\text{-Ala+ATP+Lys+MgCl}_2 \xrightarrow{\text{pMurE}} N^{\alpha}\text{-UDP-Glu}^{\gamma}\text{-Ala-}^{\varepsilon}\text{Lys + ADP+ Pi}\ldots\ldots\ldots \text{(R1)}$$

$$\text{PEP + ADP} \xrightarrow{\text{PK}} \text{Pyruvate + ATP}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(R2)}$$

$$\text{Pyruvate + NADH} \xrightarrow{\text{LDH}} \text{Lactate + NAD}^{+}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{(R3)}$$

Both the continuous and stopped reaction methods for quantification of ADP catalysed by reaction (R1) were performed. No significant detection of ADP was observed in either reaction. During the continuous method, the reaction was initiated using 1.0 mg/mL enzyme

(Mfer762). The experiments were performed at 50°C. The absorbance at 340 nm started to increase and did not decrease for up to five mins as required to obtain evidence for activity of the enzyme towards the substrate. The mixture solution was then incubated at 80 °C for 10 mins and lowered to 50 °C to measure the absorbance at 340 nm. The enzyme did not show noticeable activity suggesting that the binding Mfer762 does not recognise the UMA as substrate.

3.10.2 Phosphate assay

A phosphate assay (P_i assay) was performed on the reaction mixture from the ADP quantification assay to test for any noticeable difference in free phosphate between the control reaction and putative enzyme catalysed reaction (R1). The control reaction is the same as R1 but without the substrate (UMA) added to the mixture. The detection of free phosphate would imply the activity of the enzyme in the presence of substrate, yet not detected by the ADP quantification assay. However, the calculated amount of free phosphate released during the putative enzyme catalysed reaction was negligible (only 0.013 mM out of 2.0 mM ATP used). Hence, it was concluded that the enzyme did not recognize the bacterial substrate and therefore had no activity under these conditions.

3.11 Conclusions and discussion

Structures of archaeal pseudomurein peptide ligases have been solved for the first time. The structure of two orthologues from *M. fervidus* and *M. thermautotrophicus* ΔH (Mfer762 and Mth734, respectively) have been determined at high resolution and are best described as structural homologues of MurE peptide ligases in bacteria. Based on the structural comparisons it was found that both Mfer762 and Mth734 share a similar domain arrangement as studied for all available murein ligases (MurC-MurF). Moreover, the

invariant common amino acids that are studied to be present in all the murein peptide ligases are also present with two pseudomurein enzymes (Bouhss *et al.*, 1997). These invariant residues in addition to the glycine-rich conserved P-loop, required to be Mur family enzymes, are mostly involved in ATP binding in the murein ligases (MurC-MurF) and are structurally conserved and invariant in pMurE as well (Figure 3.36). The presence of these invariable residues in all the orthologues of Mfer762 and the structural alignment of these residues in Mfer762 and Mth734 with the murein ligase structures support their role as putative pseudomurein peptide ligases and a member of the Mur family. Further, the structure-based similarity search suggested that the Mfer762 structure is similar to the bacterial MurE/F types, mostly based on the N-terminal domain comparison. The individual domain-based structural similarity search suggests Mfer762 to be more similar to MurE (Section 3.8), and hence, support the enzyme being a pseudomurein peptide ligase type E (pMurE).

The structural comparison of the pMurE and bacterial peptide ligase shows that the core structure of the N-terminal domain is consistent to that studied for the MurE/F peptide ligase type. Moreover, the secondary structure connectivity for the pMurE structures is similar to that for bacterial MurE. The Mth734 structure is similar to Mfer762 but has a shorter N-terminal domain. The presence of a short N-terminal domain in the orthologue of pMurE (Mth734), yet performing the same function suggests that initial 20 amino acids residues of the Mfer762 N-terminal domain might not be significant for pMurE function. This can be explained from the proposed internal UDP-binding site in Mth734 (Figure 3.26), where the first conserved amino acid presumably involved in UDP interaction is immediately at the start of the primary structure of Mth734 compared to Mfer762 (Thr8 versus Thr28 (Mfer762)). The truncation of the N-terminal 24 amino acid residues at the N-terminal domain is not associated with a functional difference in the *M. tuberculosis* MurE

(Basavannacharya, *et al.*, 2010b). Thus, the comparative analysis of the two pMurE enzymes could be used to explain Mth734 as a naturally truncated pMurE.

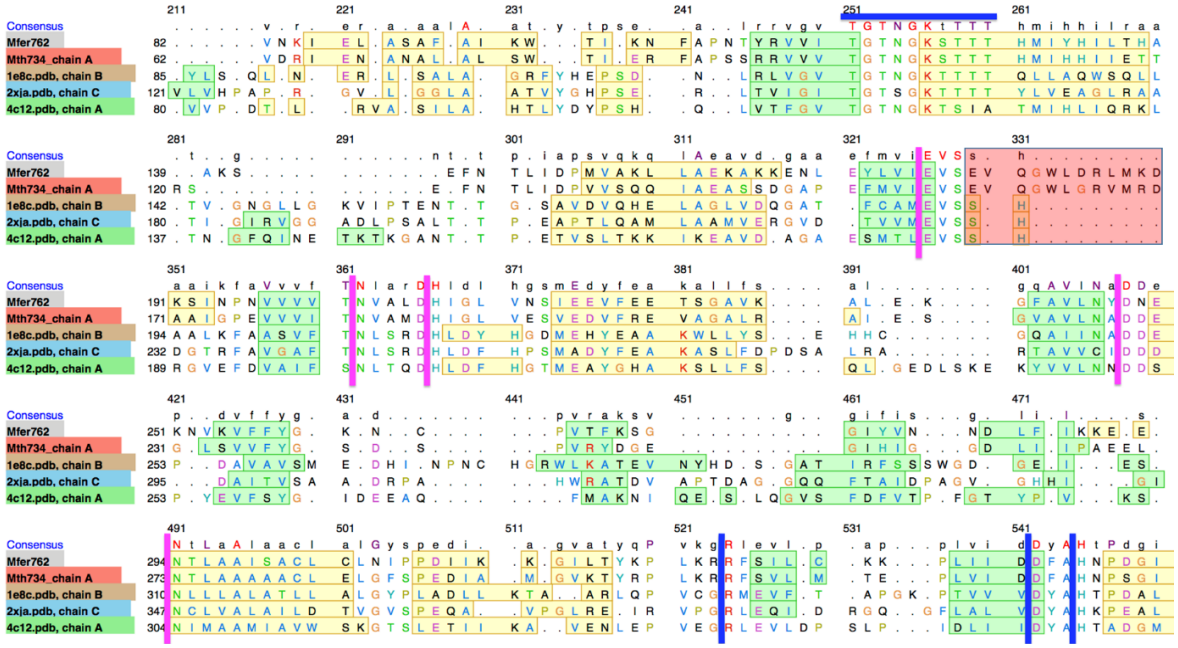


Figure 3.36 Structure-based alignment of pMurE (first two, Mfer762 and Mth734) and MurE from *E. coli* (1E8C), *M. tuberculosis* (2XJA) and *S. aureus* (4C12) highlighting the ATP-binding residues conserved in MurE and their existence in pMurE (magenta line). The residues that contribute to the interaction from the C-terminal domain are indicated by vertical blue lines. Arg325 (Mfer762 numbering) is a residue of the hinge region that is conserved both in pMurE and MurE. The signature P-loop residues are indicated by a horizontal blue line and the insertion region forming UDP-binding site at exterior surface (Figure 3.31) is highlighted in a red box.

Further, two UDP-binding sites have been identified in two Mfer762 structures from different crystal forms. The UDP-binding site observed in Mfer762-UDP_1 structure is present at the exterior surface of the protein and is unique to pMurE. The binding site at the exterior surface appears to fit the UDP with binding interaction from the 'extended loop' of the middle domain and region of the N-terminal domain, yet is not of sufficient size for the expected substrate (N^{α}-UDP-Glu$^{\gamma}$-Ala). The 'extended loop' appears to be a result of an

amino acid sequence insertion compared to MurE or MurF structures (as shown previously in Figure 3.34 and Figure 3.31) and is conserved among pseudomurein-containing methanogens (Figure 3.38). This region forms a β-hairpin structure and contributes to a 'gateway-like' conformation with the N-terminal domain with two structurally conserved residues Trp177 and His41. The side-chains of these residues flank each other forming a π-π interaction as observed with the Mfer762 apo structure. The distance between the two side-chains of the residues are within the range (3.8 Å) as suggested for such interactions (Janiak, 2000; Liao *et al.*, 2013). A similar observation exists for the Mfer762-UDP_1 structure, however, the side-chain of Trp177 shifts to form a face-to-face stacking interaction with the uracil ring of the UDP molecule. The change in Trp177 side-chain results in the switch of His41 rotamer from m-70 to p80, yet regaining the π-π stacking interaction. The structural conformation as exhibited by both the Mfer762 apo Mfer762-UDP_1 structures have the region in a closed-'gateway-like' conformation. Thus, it could be proposed that a closed-'gateway-like' conformation formed by the residues in 'extended loop' and N-terminal domain holds the structure intact and might be contributing for the structure to remain in the open conformational state.

The mechanism with which the bacterial murein ligases function is with an open conformation at the apo state and then closure of the cavity through the movement of C-terminal domain towards the middle domain upon substrate binding (Smith, 2006). Moreover, it has been postulated that neither the nucleotide nor the substrate binding could be sufficient for the murein ligases to adopt a closed conformation but could be limited by the incoming ligand (Favini-Stabile *et al.*, 2013). This could be the reason for the Mfer762-UDP_2 structure to adopt a slightly opened conformation through the movement of the N-terminal domain away from the middle domain upon binding the partial substrate (UDP). The intermediate conformational state is also observed in the bacterial MurD ligase and is

present both either in the presence of small molecules, or even in their absence (Sink, *et al.*, 2016), and could be applicable for pseudomurein ligases as well. This could also explain the reason for further open conformation at the N-terminal domain compared to the apo and the Mfer762-UDP_1 structure.

Moreover, the UDP-binding site observed in the Mfer762-UDP_2 structure is present at the inner surface of the enzyme that interacts solely with the N-terminal domain. A similar UDP-binding position has been observed in bacterial murein ligases and comparative analysis indicates that the position of the UDP-binding site, as observed in the Mfer762-UDP_2 structure, is required to fit the putative large substrate (N^{α}-UDP-Glu$^{\gamma}$-Ala) for the pMurE peptide ligase. The UDP binding in MurE is solely at the N-terminal domain whereas that for MurF is at the border of N- and C-terminal domains (Cha, *et al.*, 2014). Thus, the UDP-binding mode in the Mfer762-UDP_2 structure is more similar to the UDP binding in bacterial MurE peptide ligase than compared to the MurF peptide ligase. The modelling of the bacterial substrate, UDP-*N*-acetylmuramoyl-L-Ala-D-Glu (UAG), as exhibited by MurE of *M. tuberculosis* (2XJA) using structural superposition suggests that the conformation adopted by the Mfer762-UDP_2 structure is required for the complete substrate-binding. The binding of the substrate/product in Mfer762 structure appears to adopt a slightly different conformation, yet approximately similar location, as compared to bacterial MurE (Figure 3.35). Hence, it could be proposed that the UDP-binding site at the inner surface as exhibited by the Mfer762-UDP_2 structure is the catalytic active site required for entire substrate binding and product formation.

The modelling of UAG using the 2XJA structure into the Mfer762 structure, based on structural superposition, suggests that the UDP head of UAG would clash via steric hindrance with structurally conserved residues Arg40, His41 and Trp42 (Figure 3.35). This suggests that the conformation as observed with the UDP molecule at the putative catalytic

cavity of Mfer762 (UDP_2-binding site) might be adopted by the N^{α}-UDP-Glu$^{\gamma}$-Ala. Moreover, the steric hindrance suggested by the residues of the disordered region (139-150) and flexible 'extended loop' between the β8 and α7 including Trp177 when compared to the Mfer762 apo structure, indicates that the conformation as adopted by the Mfer762-UDP_2 structure is structurally favourable for putative substrate binding. This observation indicates that both disordered loops of the Mfer762-UDP_2 structure could be required to adopt the conformation for the Mfer762 structure to bind the substrate in the inner cavity. It is important to note that UAG is not the substrate for pMurE or any archaeal pMur ligase and is used for comparative purpose as the length of putative product for the enzyme type (N^{α}-UDP-Glu$^{\gamma}$-Ala-$^{\varepsilon}$Lys) is almost the same size and has UDP in common (Figure 3.37).

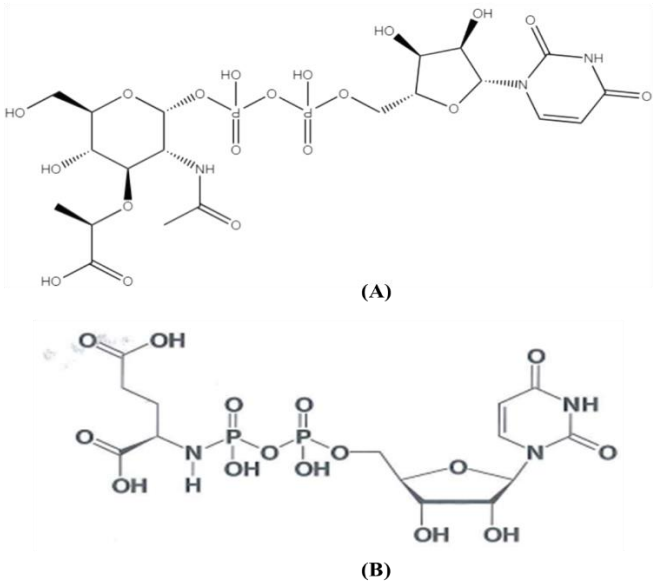


Figure 3.37 Structural comparison of the initial substrate for murein and pseudomurein polypeptide biosynthesis pathway. (A) UDP-*N*-acetylmuramic acid, substrate for MurC, (B) UDP-N-glutamic acid (free acid)

The closed 'gateway-like' conformation adopted by both the Mfer762 apo and the Mfer762-UDP_1 structures prevents the UDP at the exterior surface from entering into the inner cavity. Moreover, the exterior UDP-binding cavity (UDP_1-binding cavity) appears to be

limited for the UDP, hence, will not be able to accommodate the entire substrate expected to bind the enzyme. In addition, the UDP_1-binding site is positioned far apart from the ATP-binding site. The mechanism for bacterial murein peptide ligase requires the binding of the ATP followed by the addition of UDP-MurNAc and appropriate amino acid in a sequential manner with the arrangement such that the end of the UDP-MurNAc comes close to the γ-phosphate end of the ATP molecule (Emanuele, *et al.*, 1997; Perdih, *et al.*, 2007). The addition of the first amino acid in bacteria is L-Ala, catalysed by MurC, where the terminal carboxyl group of UDP-MurNAc is activated by ATP-dependent phosphorylation that generates an acyl phosphate intermediate used by the amino acid (Mol, *et al.*, 2003). The structural analysis of bacterial murein ligases indicate that the ATP- and the UDP-MurNAc-binding sites are located in close proximity and positioned within the crescent shaped cavity formed at the interface of three domains (Gordon *et al.*, 2001; Deva *et al.*, 2006). This does not seem to be applicable with the 762 UDP_1-binding site suggesting that it could possibly be an allosteric site.

In contrast, the soaked solution used for the Mfer762-UDP_1 structure contained only the UDP and lacks the first substrate (ATP) for putative pseudomurein ligase activity, which could limit the conformational change required by the enzyme to adopt for the next step. The binding of the ATP has been suggested to be required for inducing or stabilising the appropriate crescent-shaped cavity at the interface of the three domains (Yan, *et al.*, 2000). Moreover, the study on *E. coli* MurF indicates a lack of a characteristic nucleotide-binding fold at the N-terminal domain and contains an extended loop from the central domain towards the N-terminal domain that is involved in pyrophosphate binding (Yan, *et al.*, 2000). The Mfer762-UDP_1 structure shows detailed interactions through side-chain rearrangement and a switch in rotamer of Trp177 and His41, respectively, the 'gateway-like' conformation forming residues, in response to the UDP binding. In addition, the highly

conserved amino acid residues form a flexible ‘extended loop’ that contributes to UDP interaction. This specific interaction at the external surface of the Mfer762-UDP_1 structure could possibly represent an allosteric regulation and might be limiting the binding of substrate in the absence of ATP. An allosteric site has not been reported for any available murein peptide ligase structures and its presence in pseudomurein peptide ligase, pMurE, indicates a slight difference in the pMurE ligation activity, which could be possibly through a controlled regulatory mechanism that might be present in methanogens.

The binding of the UDP molecule in Mfer762-UDP_2 structure (UDP_2-binding site) is present at the crescent-shaped cavity and located solely at the N-terminal domain as found in the bacterial MurE. Also, the binding cavity is large enough to accommodate the entire substrate and product length as expected by the enzyme. In addition, the position of the putative ATP-binding site as proposed and explained in Section 3.6.1.2.3 and the UDP_2-binding site come close in proximity when extended in the presence of complete substrate. This supports the mechanism required by the murein ligases to add an appropriate amino acid in sequential order which is expected to occur in pseudomurein peptide ligases as well. Hence, it could be proposed that the UDP_2-binding site is the putative catalytic cavity for the enzyme and possibly adopts the conformation as suggested by the UDP molecule, which is in a slightly different orientation as adopted by the bacterial substrate (Figure 3.35).

A structurally conserved residue (Trp42) is present close to the putative catalytic site of the pMurE. The residue is not involved in interaction with the UDP molecule as revealed by the Mfer762-UDP_2 structure, however, a switch of the Trp42 rotamer to m95 appears to be important in providing a larger binding cavity around the UDP. The rotamer p-90 for Trp42 appears to be a favourable conformation for the Mfer762 apo and the Mfer762-UDP_1 structure (Figure 3.27). The switch is not even observed in the Mth734 structure with both the molecules that have two different conformations. Hence, the switch in conformation

could be required by pMurE to recognise and bind substrate into the cavity and seems to be facilitated by the movement of the N-terminal domain slightly away from middle domain. The rotamer for Trp42 in Mfer762-UDP_1 structure is arranged between the side-chain of Arg40 and Lys65 and does not favour any alternative rotamer other than p-90. The m95 rotamer for the Trp42 in the Mfer762-UDP_1 structure clashes with Asn63 and is favoured only in the Mfer762-UDP_2 structure where the N-terminal domain moves away from the middle domain. In contrast, the p-90 rotamer for Trp42 in the Mfer762-UDP_2 structure sterically hinders Arg40, a conserved residue among pMurE sequences that interacts with the UDP molecule at the inner cavity, in the Mfer762-UDP_2 structure. This implies that the switch in the rotamer is feasible only with the N-terminal domain movement.

The movement of the N-terminal domain away from the Mfer762 apo structure initiates from the hinge region between the middle and N-terminal domains and ends near the first residue that makes a hydrophobic interaction with the UDP molecule (Thr28). Hence, the Mfer762-UDP_2 structure suggests that an intermediate conformation is required by the pMurE to bind the substrate into the catalytic cavity. It is also important to note that the ATP binding is the first step in the murein peptide ligase mechanism and this is expected to be likely for pseudomurein peptide ligase activity. The modular rearrangement has been studied for bacterial murein ligases and is well understood using MurD structures (Sink, *et al.*, 2016).

The conclusions derived based on the structural detail of Mfer762 are supported by the Mth734 structure. The Mth734 structure adopts a similar structure and topology as explained for the Mfer762 and provides information on ‘open’ and ‘closed’ structural state of pMurE that could be required for the enzyme function. The structure of Mth734 is important in supporting the proposed ATP-binding site. In addition, the UDP-binding site at the exterior surface of the Mfer762-UDP_1 structure, putatively an allosteric site, as well as the catalytic site at the interface of three domains as exhibited by the Mfer762-UDP_2 structure both

appear to be structurally, as well as sequentially conserved, among the pMurE. Moreover, the ‘gateway-like’ conformation formed by Trp177 of the ‘extended loop’ and His41 of the N-terminal domain as observed in the Mfer762 structure is also conserved with the Mth734_A structure. Further, the Mth734_B structure contains the ‘extended loop’ in an open-‘gateway-like’ conformation through the movement of Trp157 (equivalent to Trp177 in Mfer762), away from His21 (41) resulting in the movement of the flexible ‘extended loop’ into the core middle domain away from the N-terminal domain. The ‘extended loop’ in the Mth734_B structure makes interactions with the middle domain of the Mth734_A suggesting that the loop region is flexible and involves adopting two different conformations as exhibited by the two molecules in the asymmetric unit of the Mth734 crystal. Further with respect to the ‘extended loop’-region, residues from the middle domain of two molecules form a static interaction as a result of crystal packing to hold the two Mth734 asymmetric unit molecules at two different conformations. The protein interfaces, surfaces and assemblies service PISA at the European Bioinformatics Institute (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (Krissinel & Henrick, 2007), shows abundant crystal contacts between two molecules, yet not significant enough to suggest complex formation (Table 3.3).

Table 3.3. Interface interactions of the two asymmetric molecules of the Mth734 structure as calculated using the PISA analysis.

Mth734_A	Distance (Å)	Mth734_B
Salt bridges		
Asp218 [OD1]	3.84	Arg160 [NE]
Asp364 [OD1]	3.20	His187 [ND1]
Asp364 [OD2]	3.75	His187 [ND1]
Glu372 [OE1]	3.01	Arg347 [NH2]
Glu372 [OE2]	3.07	Arg347 [NH2]
Glu372 [OE2]	2.44	Arg350 [NH1]
Arg367 [NH1]	2.93	Glu122 [OE1]
Arg418 [NH2]	3.46	Asp353 [OD2]
Hydrogen bonds		
Ala334 [O]	3.17	Trp22 [NE1]
Ala334 [O]	3.70	Arg45 [NH1]
Ala334 [O]	3.67	Arg45 [NH2]
Glu268 [OE2]	2.57	Leu158 [N]
Asp218 [OD1]	3.84	Arg160 [NE]
Asp364 [OD1]	3.20	His187 [ND1]
Asp401 [O]	2.21	Arg304 [NH2]
Glu402 [O]	3.05	Arg304 [NH2]
Tyr407 [OH]	3.61	Ser323 [OG]
Ser406 [OG]	3.47	Arg347 [NH2]
Glu372 [OE2]	2.44	Arg350 [NH1]
Arg367 [NH1]	2.93	Glu122 [OE1]
Arg326 [NH1]	3.50	Gln155 [OE1]
Arg326 [NH2]	3.53	Gln155 [OE1]
Arg326 [NH1]	3.23	His187 [O]
Arg326 [NH1]	3.61	Ile188 [O]
Arg221 [NH1]	3.02	Leu190 [O]
Arg339 [NH2]	3.32	Ser349 [O]
Arg418 [NH2]	3.46	Asp353 [OD2]
Arg339 [NH1]	3.48	Gln385 [OE1]
Asn370 [ND2]	3.03	Gln385 [OE1]
Arg339 [NH1]	2.98	Gln385 [O]

In addition to the residues listed in Table 3.3, the residue Trp157 in the Mth734_B structure (Trp157|B) forms a π-π stacking interaction with Tyr245|A, unlike the Mth734_A structure that forms the stacking interaction with His21 of the N-terminal domain of the same

molecule to form a ‘gateway-like’ conformation as observed for Mfer762. The conformational difference exhibited by two molecules in the Mth734 asymmetric unit suggests structural flexibility exist in the apo state of pMurE. Similar flexibility in the C-terminal domain orientation revealed by two molecules of the asymmetric unit has been observed with the bacterial MurD of *Thermotoga maritima* (Favini-Stabile, *et al.*, 2013).

The multiple sequence alignment of all the pMurE peptide ligases and the structures of two members of the pMurE (Mfer762 and Mth734) suggest that the propositions derived based on the available structures might be true for all the members of the pMur enzyme family (Figure 3.38). Firstly, the P-loop in the middle domain appears to be well-conserved among all the pseudomurein-containing methanogens and is required by the proteins of the Mur family. Second, the ‘extended loop’, that has been observed and discussed as an amino acid sequence insertion in the middle domain and indicated to contribute to UDP-binding in the Mfer762-UDP_1 structure, also seems to contain highly conserved amino acid residues among all the members of pMurE. The Trp177 residue (based on Mfer762 numbering) within this loop is conserved in all the pMurE sequences except for pMurE in *Methanopyrus*. This is the key residue that forms the ‘gateway-like’ structure as explained in Section 3.7.1.1.1 and discussed earlier in this Section. Further, another residue in the gateway formation, His41, is also conserved among all the members of the pMurE. The substitution of Trp by Leu in *Methanopyrus* at position 177 (numbering for Mfer762) could be to avoid the possible steric hindrance due to the presence of the bulky Trp side-chain in the N-terminal domain instead of His41. Hence, the substitution of the Trp residue by a non-aromatic residue at position 177 suggest that *Methanopyrus* pMurE lacks the π-π stacking interaction as observed with Mfer762 and Mth734 structure and could stabilise the structure in an open conformation through different interactions. Moreover, residue Arg40, involved in UDP-binding in the Mfer762-UDP_2 structure is also conserved in all pMurE sequences.

This residue together with His41 and Trp42 sterically hinders the binding of the modelled UAG and suggests that the orientation of the final substrate for pMurE follows the path as shown for UDP_2 (Figure 3.35). UDP is a partial mimic for the complete pMurE substrate and is extended from the phosphate end towards the C-terminal domain in the complete substrate. The conservation of Arg40 is crucial in depicting the orientation of the substrate binding to pMurE due to its direct interaction with the UDP_2 molecule through its side-chain nitrogen atom (Figure 3.14). It was observed that the residues involved in UDP-binding at both positions, as revealed by the Mfer762 structure, are well conserved in all other pMurEs (Figure 3.38).

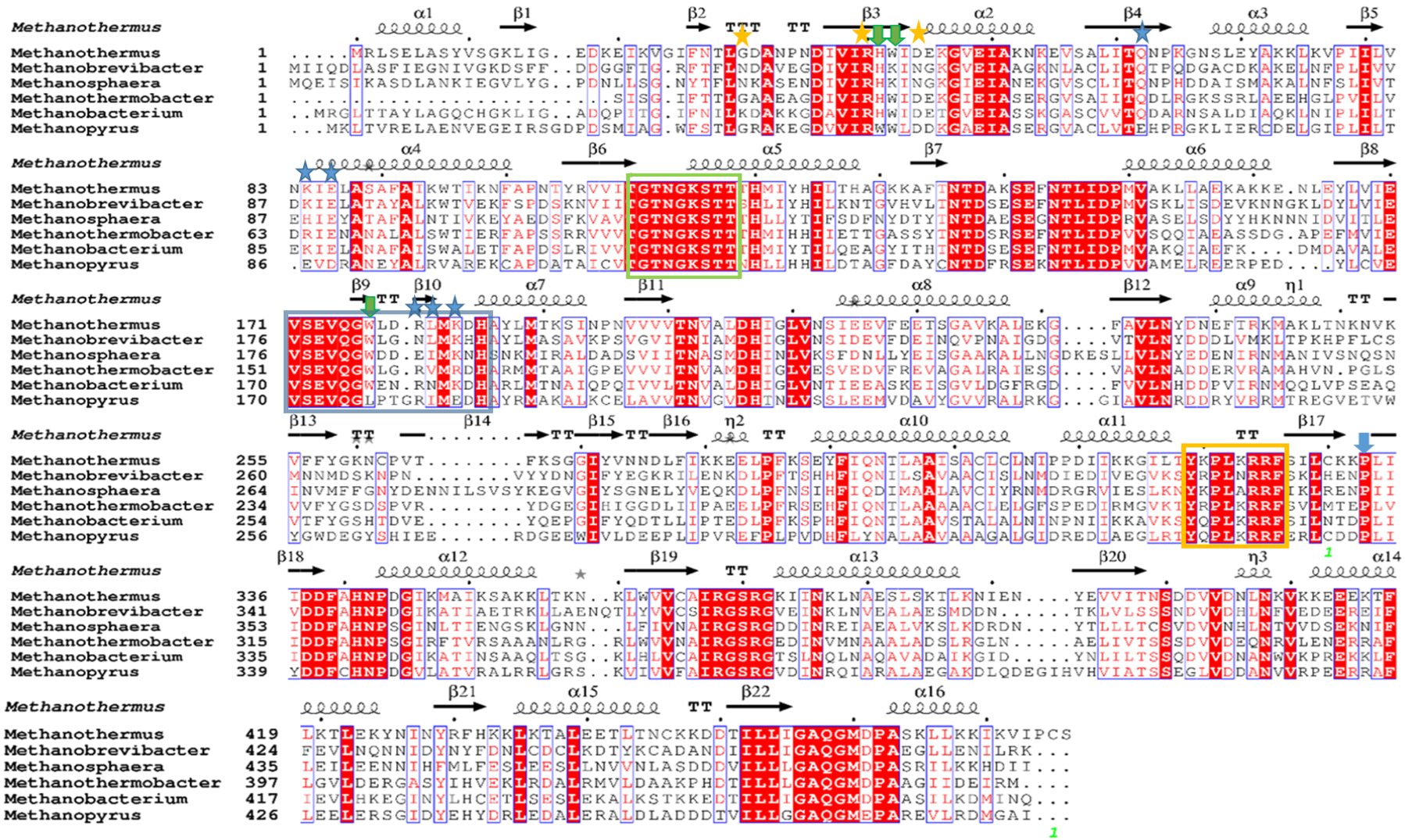

Methanothermus
Methanothermobacter
Methanobrevibacter
Methanosphaera
Methanobacterium
Methanopyrus

Figure 3.38 Structure-based multiple sequence alignment of pMurE peptide ligases. The multiple sequence alignment highlights the key conserved residues of the enzyme class and provides insights to the conserved ATP and substrate binding sites. For clarity, only a few sequences of representative species are shown. The list of organisms included are *Methanothermus fervidus, Methanosphaera stadtmanae* DSM3091, *Methanothermobacter thermautotrophicus* ΔH, *Methanobacterium* sp. MB1, *Methanobrevibacter* sp. AbM4 and *Methanopyrus kandleri* AV19. The conserved P-loop, represented by a green box, which contributes the residues for ATP binding. In addition, the residues involved in UDP binding in Mfer762 are highlighted with a star (yellow for internal UDP-binding site and blue for external binding site). The conserved residues forming the 'gateway-like' structure in pMurE structures are represented by green down arrows. The conserved hinge region of Mth734 and Mfer762 appears highly conserved among pseudomurein-containing methanogens and is represented by an orange box. The amino acids forming the 'extended loop' of the Mfer762 and the Mth734 structures are also highly conserved and represented by a blue box. The conserved proline that is observed in cis-conformation in both the Mfer762 and Mth734 structures is highlighted by a blue down arrow.

Finally, Pro333 in the Mfer762 structures determined during this project all adopt a cis-conformation and is observed close to Cys330 that forms a disulfide bond with Cys475. The proline at this position is structurally conserved and adopts a cis-conformation in the Mth734 structure, and is also well-conserved sequentially among all the members of pMurE. The presence of the cysteine residues and formation of a disulfide bond appears to be the case only with *M. fervidus* (Figure 3.38). The reason for the presence of disulfide bond only in Mfer762 at the position such that it stabilises two ends of the C-terminal domain is not understood but could be related to the structural stability at the high temperature and lower ionic strength at which Mfer762 enzyme functions. Moreover, the way the disulfide bond holds two ends of the C-terminal domain could be related to the conserved structural fold

that might be required by the enzyme to insure the accuracy with appropriate amino acid binding at its physiological temperature. The bacterial MurE ligase that adds the L-Lys to the growing cross-linking peptide subunit contains a conserved motif D(D,N)P(N,A) which is absent in pMurE. The C-terminal domain of the pMurEs has a well conserved sequence motif with unknown function. The first sequence motif immediately after the hinge region with consensus sequence DDFxHNPxG extends from the β18 to α12 (based on Mfer762 structure) where the second aspartate and histidine residues correspond to an invariant residues conserved among all murein ligases (Smith, 2006). These two residues are involved in ribose sugar and tri-phosphate binding of the ATP molecule in the bacterial murein ligase structures, respectively (Smith, 2006). Similarly, the conserved motifs, IRGSRG and ILLxGAQGMD/EPA, are present between β19 and α13 and extended from β22 to α16, respectively in the pMurE structures. Bacterial cell lysis has been related to the incorporation of the inappropriate amino acid to the cross-linking peptide end (Mengin-Lecreulx, *et al.*, 1999). The need for specific folding might also be needed for pMurE as they lack any identifiable conserved motif at the C-terminal domain as possessed by the bacterial MurE to identify the specific amino acid they bind (Boniface, *et al.*, 2006; Ruane, *et al.*, 2013).

Hence, the study was successful in establishing structural homology between the pMurE and the MurE peptide ligase. In addition, the determined pMurE structures provide insight to the UDP-binding site at the crescent-shaped cavity formed at the interface of three domains and provides structural evidence to support the presence of the putative N^{α}-UDP-Glu^{γ}-Ala-binding site that could extend further from the UDP_2-binding site. Although a striking difference has been observed in the pMurE through the identification of a novel putatively allosteric site (UDP_1-binding site), the mechanism for the ligation of L-lysine by the pMurE is still presumed to be similar to that of the bacterial MurE.

Chapter 4 Pseudomurein peptide ligase type C

4.1 Introduction

MurC in bacteria is responsible for the addition of the first amino acid (L-Ala) to UDP-MurNAc during the biosynthesis of the cross-linking pentapeptide (Deva, *et al.*, 2006). The initial substrate to which the L-Ala is added in pseudomurein is different to that of murein (Figure 6.1). The addition of L-Ala, in the cross-linking pentapeptide of the pseudomurein cell wall, occurs to the N^{α}-UDP-glutamic acid-γ-phosphate, which is synthesised from N^{α}-P-glutamic acid. Furthermore, during biosynthesis of the glycan backbone, UDP-MurNAc in murein is substituted by UDP-*N*-acetyl L-talosaminuronic acid (UDP-NAcTalNA) that is biosynthesised though the epimerisation of UDP-*N*-acetylgalactosamine followed by the oxidation of the intermediate product (UDP-*N*-acetyl L-altrosaminuronic acid).

Homologous gene searches using PSI-BLAST showed that the genes encoding MurB are absent in pseudomurein-containing methanogens, however, *murA* from *E. coli* (WP_000357259.1) shows homology with *aroA* gene from *Methanobrevibacter smithii* and *Methanopyrus* sp KOL6 and *aroA* genes are present in most of the *Archaea* suggesting a role other than just in pseudomurein biosynthesis. Moreover, genes homologous to those responsible for encoding MurC in bacteria have been identified and are present only in pseudomurein-containing methanogens. The presence of L-Ala as the second amino acid in the cross-linking peptide of pseudomurein and the presence of the homologous enzyme in organisms containing this cell wall type led us to a hypothesis that pseudomurein ligase type C (pMurC) in pseudomurein-containing methanogens is involved in the addition of L-Ala to the cross-linking peptide chain. The sequence identities between bacterial MurC and their homologues in pseudomurein-containing methanogens is low, < 25%. This sequence

identity is not sufficient to support a clear evolutionary history but is an indication that these enzymes might be involved in a similar function. To clarify the relationship between bacterial MurC and archaeal pMurC, the first structure of pMurC has been solved and used to predict the function of the enzyme. Moreover, the structural details helped to resolve whether MurC and pMurC are homologous and thereby helped to understand their evolutionary relationship.

4.2 Bioinformatics analysis

4.2.1 Homologous protein search

The amino acid sequence for MurC from *Bacillus subtillis* (WP_003229274.1) was used to search for homologous protein sequences in pseudomurein-containing methanogens using the highly sensitive PSI-BLAST algorithm (Altschul, *et al.*, 1997). The pseudomurein-containing methanogen proteins that shared significant sequence alignment with the query sequence as computed by the PSI-BLAST algorithm were then used to search for the homologous proteins within the *Archaeal* domain. It was observed that the pMurC gene was present in a few of the pseudomurein-containing methanogens, for example *Methanothermobacter marburgensis,* but was found to be approximately double the sequence length. The average gene sequence length encoding for a pseudomurein ligase is approximately 1500 bp. The sequence alignment indicated that the gene was a result of the fusion of two genes, pMurC and pMurD1. The possible fusion of two genes, *mth_530* (WP_010876169.1) and *mth_531* (WP_010876170.1), was studied earlier in our laboratory with *M. thermautotrophicus* (ΔH) that contains the related gene as currently annotated for other *Methanothermobacter* (Wolf, 2010). Based on the support of sequence alignment the merged gene was manually divided into two genes of approximately the same length

(*mth_530_new* and *mth_531_new*). The *mth_530_new* is the gene that is proposed to encode pMurC in *M. thermautotrophicus* (ΔH) and this assignment is supported by the PSI-BLAST result. The *mth_531_new* is proposed to encode pMurD1 and will be discussed further in Chapter 5.

The amino acid sequence of the newly annotated Mth530_new, putative pMurC, was used as a query sequence to search for orthologous proteins in the genomes of pseudomurein-containing methanogens. The results showed that the genes encoding the protein sequences were highly conserved in the pseudomurein-containing methanogens and were not found in other methanogens or archaeal groups. This lead to the hypothesis that the proteins encoded by the genes are involved in pseudomurein cell wall biosynthesis and were probably homologous to MurC in bacterial murein cell wall synthesis. The results of pMurC sequences in other pseudomurein-containing methanogens were verified using the amino acid sequence of Mfer336, the putative pMurC in *Methanothermus fervidus* (WP_013413417.1).

4.2.2 Gene cluster analysis

It was overserved that the pMurC genes in most of pseudomurein-containing methanogens were found neighbouring to the pMurD1 or even fused, as discussed in previous section, as observed within the genus *Methanothermobacter*. Any unannotated genes close to the putative pseudomurein biosynthesis genes were then analysed for their occurrence in other pseudomurein-containing methanogens. The amino acid sequences of these genes were extracted and used to identity if their presence was only in pseudomurein-containing methanogens or were widely present among *Archaea* using PSI-BLAST. The presence of pMurC-encoding genes only in pseudomurein-containing methanogens could imply their putative role in pseudomurein peptide biosynthesis.

The SYNTAX bioinformatics tool was also used to search for similar genes and any conserved neighbouring genes among the pseudomurein-containing methanogens (Oberto, 2013). It was observed that the genes responsible for encoding pMurC and a neighbouring pMurD1 genes were widely conserved among the target organisms. It was also clear from the SYNTAX result that the merging of two genes (possibly 'fusion') was common among the available genomes of *Methanothermobacter* (Figure 4.1). The pMurD2 genes were present within the gene cluster in *Methanothermobacter*, however, they are not always present and adjacent to pMurD1 and pMurC in other pseudomurein-containing methanogens. The SYNTAX result using Mth530_new amino acids sequence as a query search also displays a similar result and is shown in Figure 4.1.

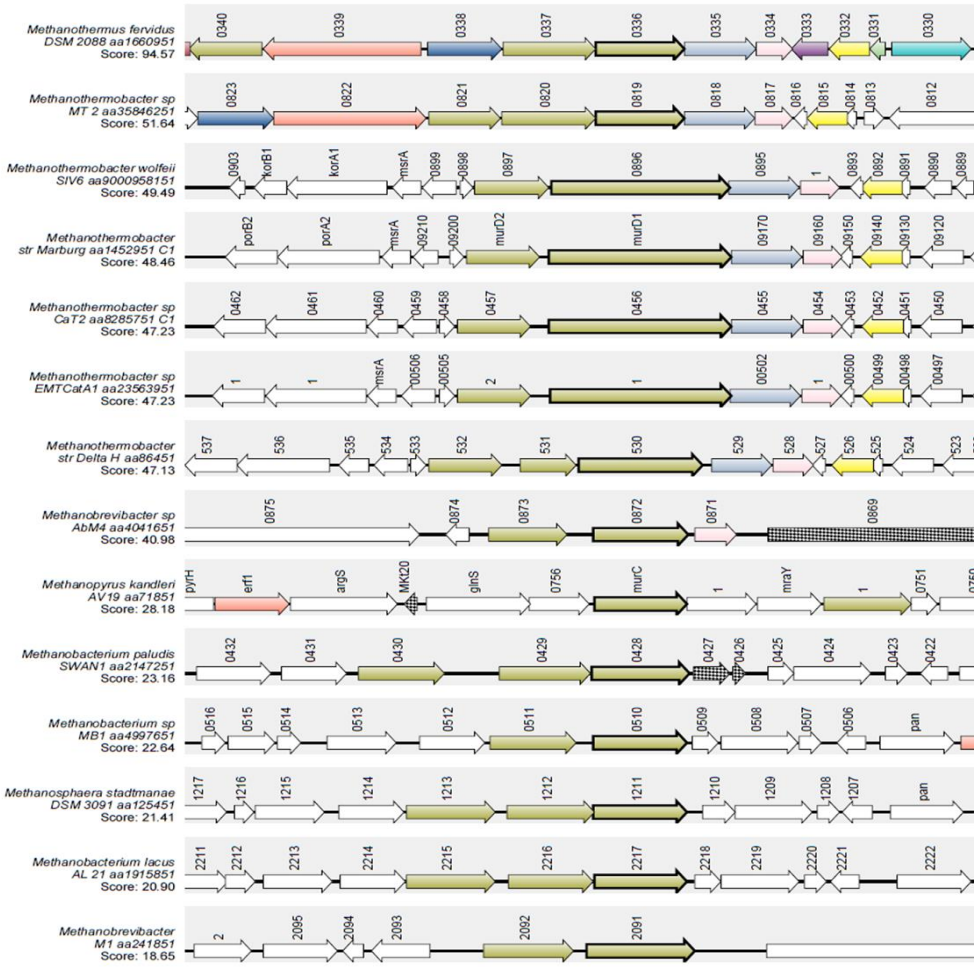


Figure 4.1 SYNTAX analysis showing the presence of pMurC in pseudomurein-containing methanogens. In addition to the pMurC, the orthologues of pMurD1 are also conserved and are present adjacent to one another in a gene cluster of other pseudomurein-containing methanogens. In *Methanothermobacter* sp and strains pMurC and pMurD1 are annotated as being merged as indicated by the highlighted brown arrow of approximately double the size of pMurC present in other listed organisms; pMurD2 type is also present in the gene cluster.

Similarly, the STRING bioinformatics tool was also useful to examine the interaction of the gene with neighbouring genes and their co-occurrence in the genomes. The STRING output for *mth_530* (WP_010876169.1) indicates that the gene is fused with *mth_531* (WP_010876170.1) (Figure 4.2), suggesting the amino acid sequence Mth530_new (as explained in Section 4.2.1) is more reliable. Moreover, the neighbouring genes to the

mth_530 were identified as being involved in pseudomurein biosynthesis pathway during this study. These include: *mth_531* that shows homology with bacterial MurD and proposed as putative pMurD1; *mth_532* that is again proposed as a putative pMurD2; and Mth735, which shows homology with gene encoding MraY in bacteria that catalyses the first step of the lipid cycle reaction in the biosynthesis of cell wall peptidoglycan (Bouhss, *et al.*, 2004). In addition, two more genes *mth_529* and *mth_528* that are present close the query gene have been expressed and purified to test the activity of the enzymes encoded by these genes in the formation of the initial substrate (N^{α}-UDP-Glu^{γ}-P) for the addition of amino acids by the putative pseudomurein peptide ligases (parallel project for studying the pseudomurein biosynthesis pathway) (Sauter, 2017). Moreover, surface protease-like proteins encoded by *mth_75* and *mth_87*, cell surface glycoprotein (S-layer protein)-like protein encoded by *mth_1513* two genes encoding for the protein with unknown characteristics, *mth_542* and *mth_938* also co-occur with *mth_530*.

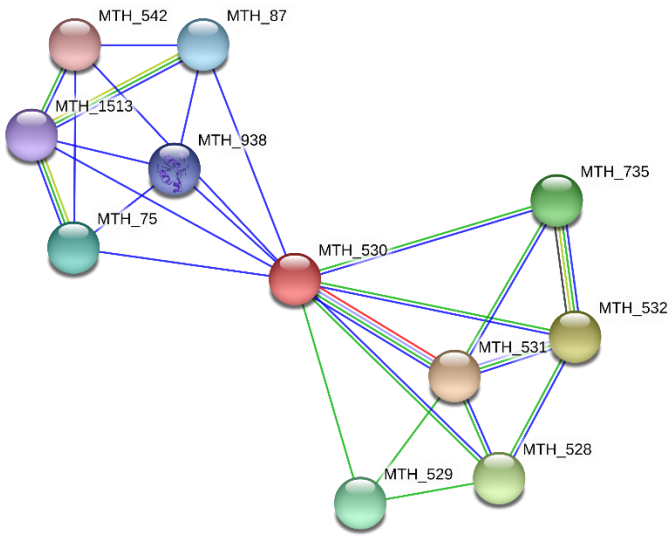


Figure 4.2 The summary view of *mth_530* and its network to show potential interactions with other genes in the *M. thermautotrophicus* (ΔH) genome. The *mth_530* (query gene) is indicated by a pink sphere, which is indicated to be fused with *mth_531* represented by the red line. The lines in green

indicate neighbouring genes, blue indicates gene co-occurrence and light green indicates that interactions occur based on text-mining. Each node and the values in the network refer to the gene numbers. The genes at the lower right side appear to be the neighbouring genes that have been identified as putative pseudomurein biosynthesis genes in this project where as those at the upper left are the genes that co-occur with *mth_530* gene.

4.2.3 Phylogenetic trees

The amino acid sequences for the bacterial MurC and putative pMurC enzymes were obtained as explained in Section 2.2.1.1. The amino acid sequences were retrieved from the same source for all the bacterial enzymes and were obtained using links via the peptidoglycan biosynthesis pathway in the KEGG database (Tanabe & Kanehisa, 2012). MurC (EC: 6.3.2.8) was selected in each case and sequences were retrieved in *fasta* format for multiple sequence alignment. Multiple sequence alignment was performed using the MUSCLE alignment algorithm (Edgar, 2004) in MEGA7 (Kumar, *et al.*, 2016). The phylogeny analysis was validated based on the bootstrap method and used 100 replicates (Felsenstein, 1985). The tree used to analyse the evolutionary history is shown in Figure 4.3. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 82 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 320 positions in the final dataset (Kumar, *et al.*, 2016).

The phylogenetic tree shows an individual clade for both the murein and pseudomurein peptide ligases type C. The high bootstrap values support the presence of pMurC as a separate clade with the *Methanopyrus kandleri* sequence deeply rooted among the pMurC clade. Similarly, *Thermus thermophillus* and *Thermotoga maritima* are observed deeply

rooted in the MurC clade. The bootstrap values at each nodes are not supportive to suggest evolutionary history and only supports individual clade.

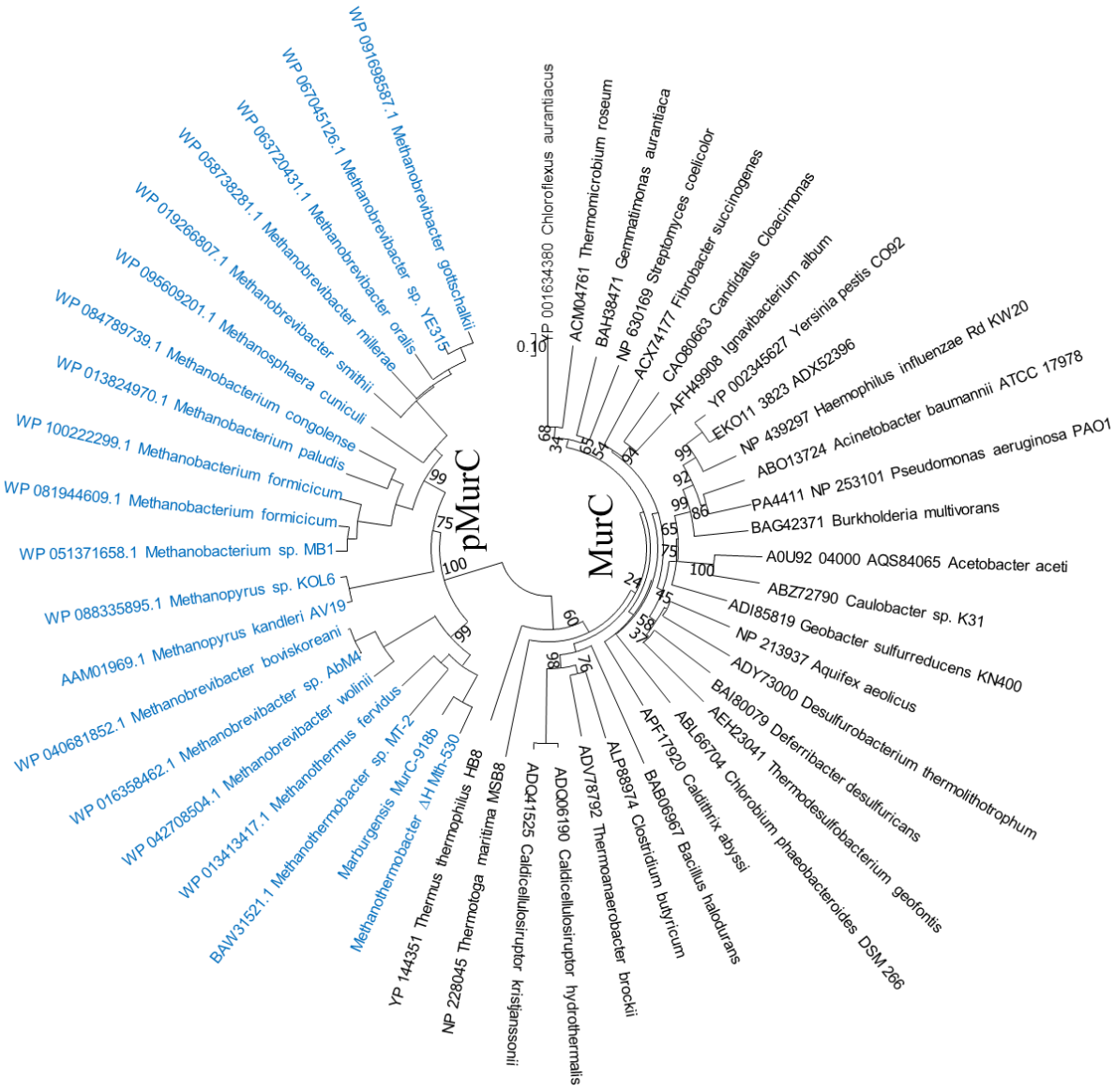


Figure 4.3 Molecular phylogenetic analysis of bacterial MurC and pseudomurein-containing methanogen pMurC by the maximum likelihood method. The pMurC sequences are labelled in blue.

4.3 Cloning, expression and purification

The putative pseudomurein peptide ligase type C (pMurC) from *M. fervidus* (Mfer336), accession WP_013413417.1, and from *M. thermautotrophicus* ΔH (Mth530_new, the amino

acid sequence encoded by the edited *mth_530_new* as in Section 4.2.2), were used for the structure determination experiments during this project. The sequence-verified pET100D-based expression plasmid for *mth_530_new* and pET151D-based plasmid for *mfer_336,* obtained from previous work of Maxillian Wolf (Wolf, 2010) and Youri Van Nuland (Van Nuland, 2011), respectively, were used initially. Later, constructs based on the pET-15b expression vector with sequences that were codon optimised (using OptimumGene™) were obtained for both the targets from GenScript, USA.

The pET-15b-based expression plasmids for the Mfer336 were transformed into *E. coli* BL21 Star™ (DE3) competent cells with ampicillin as selection (Section 2.2.2.1). The transformed bacterial colonies were individually picked and incubated at 37 °C overnight in 10 mL 2×YT medium with 100 µg/mL ampicillin and then transferred into fresh 750 mL of the same growth medium. pMurC expression was induced with 0.1 M IPTG and expression analysis of recombinant protein was carried out as described in Section 2.2.2.2. The 6×His-tagged pMurC protein was purified using a Ni-NTA agarose column affinity chromatography at room temperature as described in Section 2.2.2.4. The Mfer336 protein that was eluted from the Ni-resin was analysed by SDS-PAGE (BIO-RAD Mini-PROTEAN® TGX™, Cat# 4561043) against BIO-RAD ® low range protein standards and used for crystallisation experiments after dialysis and concentration to approximately 8.0 mg/mL. The buffer used for dialysis contained 20 mM MOPS buffer at pH 7.0; 2.0 mM TCEP and 300 mM KCl.The SDS-PAGE analysis for the Mfer336 protein is shown in Figure 4.4A. The protein was dialysed against 20 mM MOPS buffer (pH 7.0), 300 mM KCl and 2.0 mM TCEP. The purified Mfer336 protein was concentrated to a final concentration of approximately 8.0 mg/mL as explained in Section 2.2.2.7 and immediately used for crystallisation experiments.

The expression of Mth530 using the pET15b expression vector was also performed. The Mth530 protein from the Ni-resin was analysed under SDS-PAGE and is shown in Figure 4.4B. The purified Mth530 protein was also dialysed against 20 mM MOPS buffer (pH 7.0), 200 mM KCl and 2.0 mM TCEP. The dialysed Mth530 protein was concentrated to a final concentration of approximately 8.0 mg/mL and immediately used for crystallisation experiments.

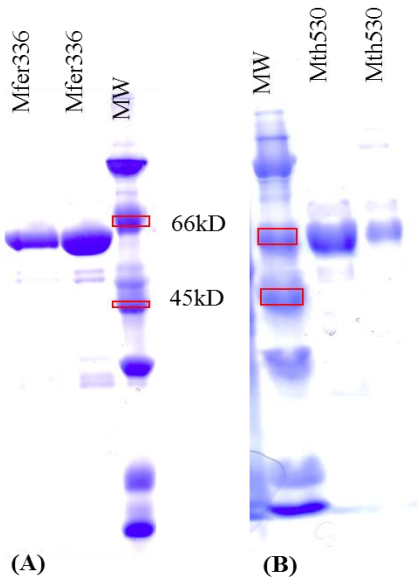


Figure 4.4 SDS-page analysis of Mfer336 and Mth530 prteins eluted from the Ni-resin. (A) Mfer336 protein expressed and purified using the Mfer336-pET15b plasmid (GenScript synthesised and codon optimised) (B) Mth530 protein expressed and purified using the Mth530-pET15b plasmid. Proteins collected from each elution step during purification process were loaded separately in the gels.

4.3.1 Molecular mass determination

The apparent molecular mass of the purified Mfer336 protein as determined by gel filtration chromatography was 46.5 kDa which was close to the estimated molecular weight of 50.0 kDa using the ProtParam tool in ExPASy (Wilkins *et al.*, 1999). The molecular mass

determined by gel filtration chromatography corresponded to the molecular mass observed from SDS-PAGE and indicated that pMurC is monomeric.

4.4 Mfer336 structure determination

4.4.1 Crystallisation

Crystallisation experiments were performed using freshly prepared Mfer336 and Mth530 proteins. Among all the crystal conditions from six different screens (576 crystallisation conditions in total) tested for the crystallisation experiments, the five conditions JCSG+_H7, JCSG+_H9, SG1_F5, SG1_E8 and INDEX_G3 produced crystals in the primary screen plates. All these crystallisation conditions contained a sulfate salt (lithium or ammonium) in common. Except for SG1_E8 (0.2 M ammonium sulfate; 30% w/v PEG4000) all other crystallisation conditions contain Bis-Tris, either pH 5.5 or 6.5, as the crystallisation buffer and PEG3350 as a precipitant. All the crystals in these crystallisation conditions, except from INDEX_G3, were thin plates (Figure 4.5). The SG1_E8 produced very small envelope-shaped crystals (Figure 4.5) on the surface of the drop which appeared within a week time and then slowly disappeared after a month. In other cases, the crystals were observed after the second week, which continued to grow for months and remained in the drop for more than six months.

In addition, INDEX_G3 produced comparatively thicker plates in the presence of ATP at the bottom of the drop (Figure 4.5). The crystal was stuck to the wall of the well, and hence, broke while harvesting. However, this was the only crystal from this condition that diffracted to a reasonable resolution (3.5 Å). The crystals were reproduced on 24-well plates using both the hanging and sitting drop methods with a larger drop size which produced similar

morphology crystals but did not diffract to the same resolution as the initial crystal from the primary screen.

The Mfer336 protein crystallised with Bis-Tris at pH 5.5 and 6.5, and therefore, attempts were made to optimise the crystal condition with the pH range of 4.5 to 7.5. However, only the pH range between 5.5 and 6.5 produced protein crystals, but these repeated experiments did not improve the crystal morphology or X-ray diffraction quality. Several other none additive-based optimisation strategies such as altering the sulfate salt, altering the protein to buffer concentration ratio, varying salt and precipitant (PEG3350) concentrations and addition of ATP and/or the putative substrate L-Ala were attempted but none contributed to the improvement of the crystal morphology or diffraction quality. All of the optimised crystal conditions that produced crystals produced similar cluster of plates as shown in Figure 4.5. The crystals were slightly elastic in nature and stuck to the walls of the plates. The co-crystallisation of Mfer336 in the presence of ATP was also repeated with all the screens but that too did not add any improvement in the crystallisation results other than with the aforementioned INDEX_G3.

Finally, the use of an optimisation screen, the Silver Bullets, along with JCSG+_H9 crystal condition (0.2 M lithium sulfate; 0.1 M BIS-Tris pH-5.5; 25% w/v PEG3350) produced different morphology crystals in the A1 condition (0.33% w/v 1,5-naphthalenedisulfonic acid disodium salt; 0.33% w/v 2,5-pyridinedicarboxylic acid; 0.33% w/v3,5-dinitrosalicylic acid; 0.02 M HEPES sodium pH 6.8) as shown in Figure 4.5F. This was the only crystal of Mfer336 which diffracted to 2.5 Å and is presented in the following section. The crystal could not be repeated and therefore, is the only successful hit for structure determination of pMurC type peptide ligases.

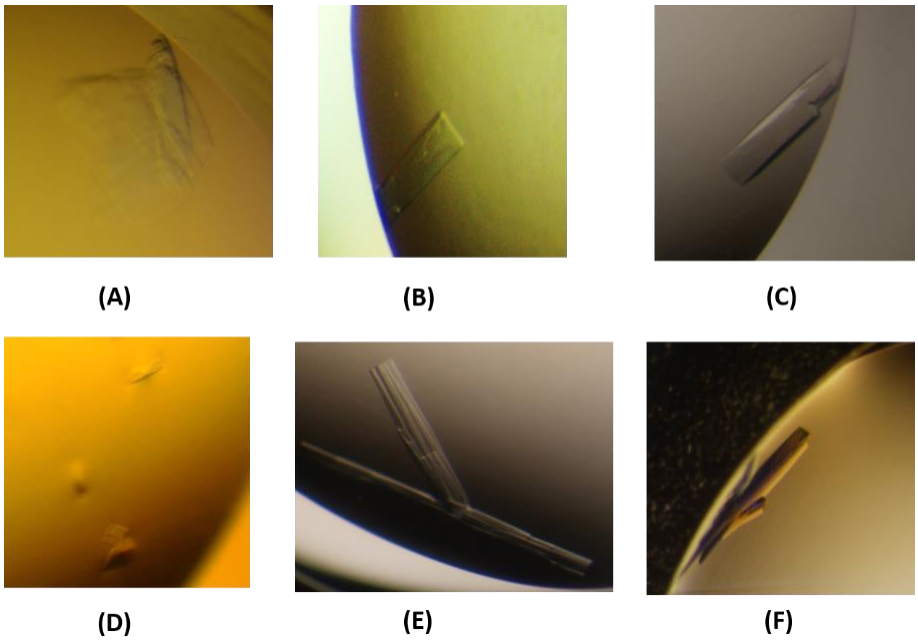


Figure 4.5 Crystal images of pMurC (Mfer336 protein). (A-D) Crystal images of Mfer336 protein grown at SG1_F5, JCSG+_H9, INDEX_G3, SG1_E8, respectively (E) crystal obtained from pH optimisation (range 4.5-7.5) at pH 5.6; similar morphology crystals were obtained from optimised conditions (refer to the text above) (F) crystal obtained on optimisation screen (Silver Bullets_A1) with JCSG+_H9 that diffracted to 2.5 Å.

4.4.2 Crystal testing

The Mfer336 crystal grown in INDEX_G3 was harvested in the presence of 25% (v/v) glycerol, prepared in the crystallisation buffer, as a cryo-protectant and tested using the X-ray source at the MX1 beamline at the Australian Synchrotron. The crystal diffracted to 3.5 Å, however, the repeat crystals from the same crystallisation condition did not diffract to that resolution. Other aforementioned crystallisation conditions for Mfer336 protein crystals were initially judged based on their morphology and birefringence properties under polarised light. Even the optimised crystallisation condition produced twinned or clusters of plates, and attempts were made to separate crystals using the loops and sharp needles, but they

remained intact. It was noticed that the crystals were elastic and, when not present on the surface of the drop, stuck firmly to the wall of the crystallisation plate causing the crystal to bend while attempting to harvest it. The crystals present at the surface of the drop were covered under a thin layer of skin, hence making the harvesting of Mfer336 difficult. Most of the crystals for Mfer336 were tested at either the MX1 or the MX2 beamlines at the Australian synchrotron X-ray source but none of the crystals other than INDEX_G3 and JCSG+_H9 optimised with Silver Bullets_A1 condition diffracted to a resolution higher than 3.5 Å.

4.4.3 X-ray diffraction and data collection

Two datasets for Mfer336 were collected at the MX1 beamline at the Australian synchrotron. The first dataset was collected using the crystal from INDEX_G3 (the broken fragment) that was co-crystallised with ATP. A complete data set of 3.5 Å was collected for the crystal. Similarly, another dataset was obtained on a crystal grown from the JCGS+_H9 condition optimised with the Silver Bullet_A1 condition. A complete dataset of 2.5 Å was collected from the crystal. Both the crystals used for the data collection were harvested in the presence of 25% (v/v) glycerol used as a cryo-protectant. The data sets were then collected at 100 K with an oscillation angle of 0.5° over a 360° oscillation range. All together 720 frames were collected with the detector at distance of 200 mm from the crystal, beam attenuation at 90% and using a double-crystal monochromator (DCM) energy of 13 keV.

The data set collected at 2.5 Å resolution has been used for structure determination and analysis during this thesis and belongs to $P4_22_12$ space group with unit cell parameters of a=b=97.8 Å and c=138.2 Å. The Matthews' coefficient for Mfer336 is 3.03 $Å^3$/Da and the estimated solvent content is 59.5% with a single molecule in the asymmetric unit.

Table 4.1 Data quality statistics for Mfer336.

Resolution range (Å)	34.58 - 2.5 (2.6 - 2.5)
Space group	$P4_22_12$
Unit cell (Å / °)	97.8 97.8 138.2 90 90 90
Total reflections	221902 (22257)
Unique reflections	30487 (3011)
Multiplicity	7.3 (7.4)
Completeness (%)	99.82 (100.00)
Mean I/sigma(I)	16.02 (1.87)
Wilson B-factor ($Å^2$)	38.58
R-merge	0.28 (1.32)
R-meas	0.30 (1.42)
R-pim	0.11 (0.52)
CC1/2	0.95 (0.77)
CC*	0.99 (0.93)

4.4.4 Phasing, model building and refinement

The Mfer336 structure was obtained by the molecular replacement technique as implemented in MORDA (Vagin & Lebedev; Vagin & Lebedev, 2015) to obtain a partial model. The truncated mtz file and sequence file were used as an input files for a molecular replacement model search. The program determined a partial solution based on the structure of MurF from *Pseudomonas aeruginosa* (PDB-4CVK_A) as the best homologous template model for the middle and C-terminal domains of Mfer336. The Buccaneer - autobuild/refine protein program (Cowtan, 2006) within CCP4i was used for further model building and side-chain fitting based on the amino acid sequence. The N-terminal domain for Mfer336 could not be modelled by any of the available automated model building programs and was eventually built manually in COOT (Emsley, *et al.*, 2010) relying on the positive 2Fo-Fc map obtained from a subsequent refinement run in REFMAC5 (Murshudov, *et al.*, 2011).

The automatic water building and refinement option in REFMAC5 was used to place water molecules in the model. Refinement was later carried out using phenix.refine (Afonine, *et al.*, 2012) within the Phenix software package (Adams, *et al.*, 2010). The MolProbity (Davis, *et al.*, 2004) results from the refinement run were used for model validation and closely inspected to improve the structure quality during subsequent model building.

Table 4.2 The Mfer336 structure refinement statistics.

Reflections used in refinement	30449 (3011)
Reflections used for R-free	1493 (138)
R-work	0.24 (0.40)
R-free	0.30 (0.42)
CC (work)	0.90 (0.45)
CC (free)	0.86 (0.36)
Number of non-hydrogen atoms	3906
Macromolecules	3776
Ligands	20
Solvent	110
Protein residues	486
RMS(bonds in Å)	0.008
RMS(angles in °)	1.62
Ramachandran favored (%)	91.7
Ramachandran allowed (%)	6.64
Ramachandran outliers (%)	1.66
Rotamer outliers (%)	12.95
Clash score	12.57
Average B-factor (Å^2)	53.7
Macromolecules	53.93
Ligands	61.39
Solvent	43.33

4.5 Overall structure of Mfer336

The first structure of a methanogenic archaeal putative pseudomurein peptide ligase type C (pMurC), from *M. fervidus* (Mfer336), has been determined using the X-ray crystallography technique during this project. The structure has been solved at 2.5 Å using the molecular replacement method as explained in Section 4.4.4. The final model of Mfer336 enzyme has a modular multi-domain structure as observed in another putative pseudomurein peptide ligase type E (pMurE), recently determined during this project, and other available bacterial murein ligases. The final model for Mfer336 structure has been refined and rebuilt to achieve refinement statistics as indicated in Table 4.2. The R_{free} and R_{work} for Mfer336 structure are 0.30 and 0.24, respectively.

The Mfer336 structure consists of three domains; an N-terminal Rossmann-fold like domain, a nucleotide-binding middle domain and another Rossmann-fold like C-terminal domain. Each domain is connected to one another by a connecting loop, which in bacterial murein ligases acts as a hinge region. The hinge regions in murein ligases are crucial for the N-and C-terminal domain movements that allow the structure to adopt open and closed conformations which are potentially required for the substrate-binding and product formation during the ligase activity, assuming that the archaeal pMurs have a similar mechanism to the bacterial Murs (Smith, 2006). The arrangement of the domains creates a crescent-shaped cavity at the interface of the three domains as shown in Figure 4.6. This crescent-shaped cavity in bacterial murein ligases is used to bind the substrates to add the additional amino acid to the product and is assumed to be a similar mechanism for the pseudomurein ligases.

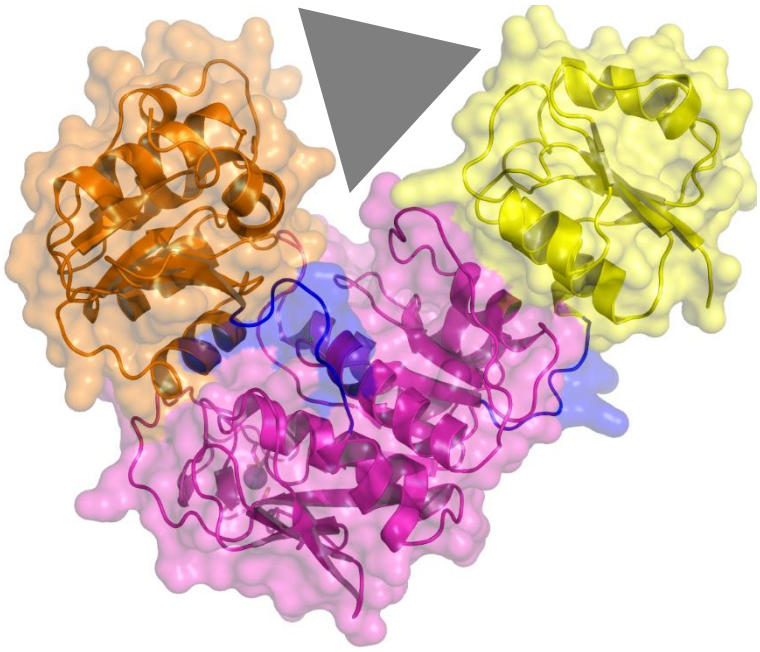


Figure 4.6 Cartoon representation of the Mfer336 structure within a transparent surface view. Each domain is coloured separately (yellow for N-terminal, magenta for middle and orange for C-terminal) with the connecting loops, presumed hinge regions, coloured in blue. The crescent-shaped cavity is present at the interface of three domains and is highlighted by a grey triangle

Mfer336 consists of 492 amino acid residues in the primary structure out of which the first five amino acids residues (1-5) and last amino acid (Lys492) could not be modelled in the final Mfer336 structure owing to limited electron density. The N-terminal domain contains the initial 97 residues (6-102) that form a parallel β-sheet (β1-β5), flanked by alternating α-helices (α1-α5) as shown in Figure 4.7. Such an arrangement of parallel β-sheet and alternating α-helices is typical with Rossmann-like fold proteins and has been observed in bacterial MurC structures (Deva, *et al.*, 2006) and is common among the N-terminal domain of all available murein ligase structures (MurC-F) (Smith, 2006).

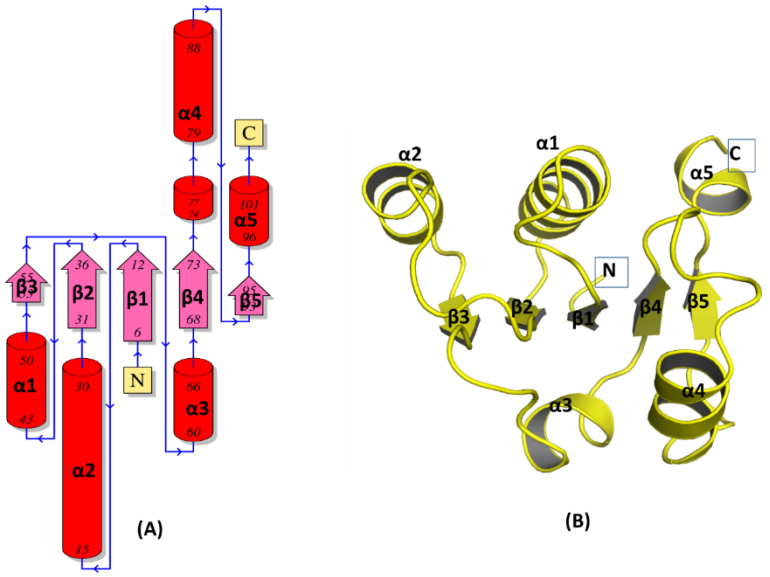


Figure 4.7 Topology and cartoon representation of the N-terminal domain of the Mfer336 structure. (A) The topology diagram was obtained from the PDBsum webserver (Laskowski, 2009) using the co-ordinate file for just the N-terminal domain. The red cylinders represent α-helices; the magenta arrows represent β-strands with the arrow head indicating their direction (N to C-terminus), blue lines indicate the loops with the arrow head representing their directions. Both α-helices and β-strands are numbered based on their connectivity and position in the secondary structure arrangement. (B) The cartoon representation for the N-terminal domain with the numbering of the α-helices and β-strands.

The middle domain contains 241 residues (108-348) and is comprised of one parallel β-sheet formed by β6 to β10 and two small antiparallel β-sheets (β11 - β12 and β13 - β15). The middle domain contains seven α-helices (α6-α12) as shown in Figure 4.8. The signature P-loop required for Mur family proteins is present between ~~the~~ β6 and α6 and formed by a conserved consensus amino acid sequence GTNGKTTT (114-121), similar to the consensus sequence GTHGKTTT found in bacterial MurC (Smith, 2006). The Mfer336 structure contains an inserted region in the middle domain between the β10 and β13 strands that projects away from the core middle domain and includes four cysteine residues that form a

tetrahedral, presumed, Zn^{2+}-binding pocket, with density consistent with Zn^{2+}. The presence of this Zn^{2+}-binding pocket is unique to this putative pseudomurein ligase and has not been observed in MurC or any other available murein ligase structures. This sequence between β10 and β13 mostly contributes to a loop structure except for a β-hairpin structure formed by antiparallel strands β11 and β12. Except for the two β-strands at the insertion region, the antiparallel β-strands β13 - β15 fold back on the core parallel β-sheet to form a similar expanded open barrel as observed in bacterial MurC (Deva, *et al.*, 2006) curving around α-helix α11.

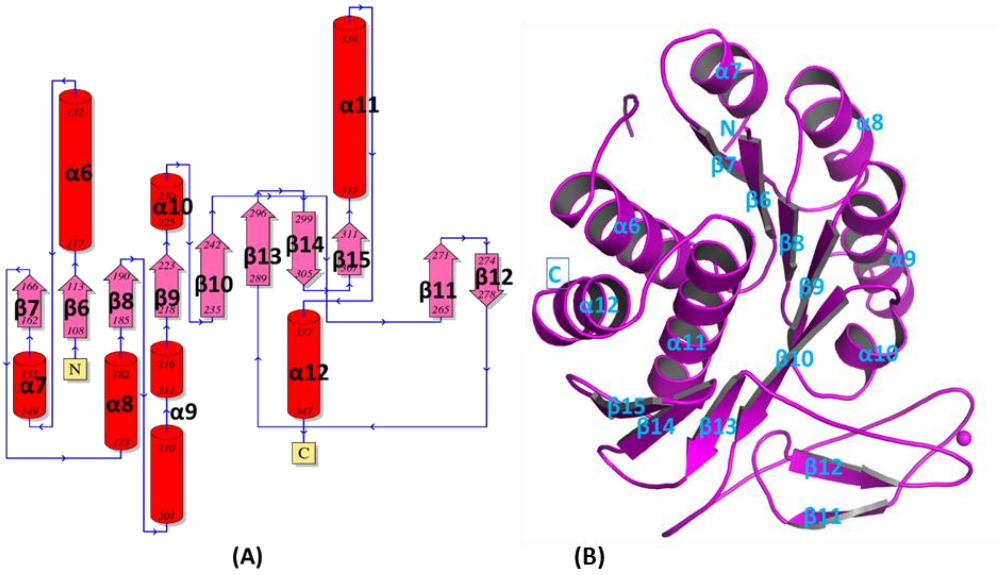


Figure 4.8 (A) Topology and (B) cartoon representation of the middle domain of the Mfer336 structure (Zn^{2+} is shown as a sphere).

The C-terminal domain contains 137 residues (355-491) and folds to form a Rossmann-type fold, comprised of a six stranded β-sheet containing five parallel (β17-β21) and one antiparallel (β16) strands flanked by six α-helices (α13-α18) (Figure 4.9). The connecting loop between the N-terminal and the middle domain consist of five residues (Tyr103,

Lys104, Val105, Lys106 and Lys107), whereas the loop between the middle and the C-terminal domains contains Lys349, Lys350, Val351, Pro352, Gly353 and Arg354.

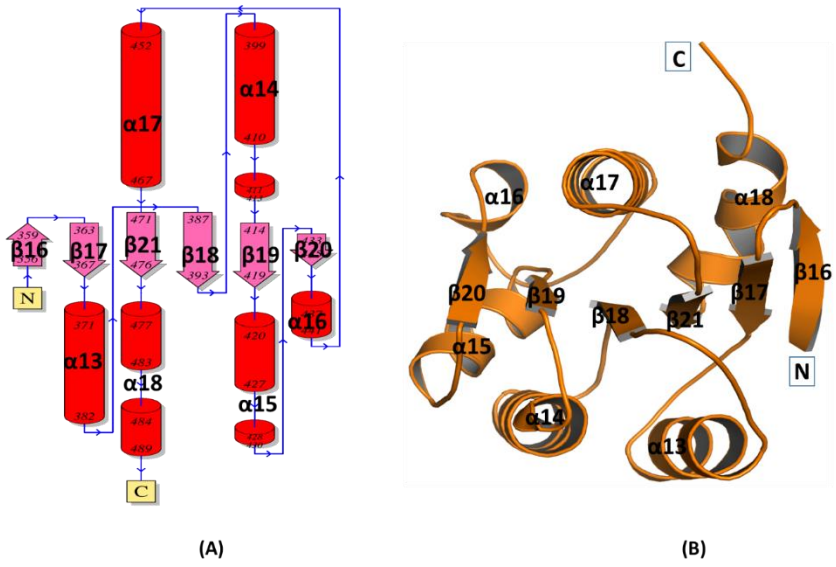


Figure 4.9 Topology and cartoon representation of the C-terminal domain of the Mfer336 structure.

The final Mfer336 structure contains a presumed pyrophosphate (POP) at the putative ATP-binding site. The electron density for the presumed POP was initially modelled with a sulfate ion (SO_4^{2-}) due to the crystal grown in a SO_4^{2-} condition, but the positive electron density (Fo-Fc) continued to be present and was later substituted to phosphate. Even the modelling of the phosphate did not completely improve the electron density causing the positive electron density to extend towards the side-chain of the Phe319 residue. Hence, a pyrophosphate was eventually modelled, which helped to improve the electron density (Figure 4.10).

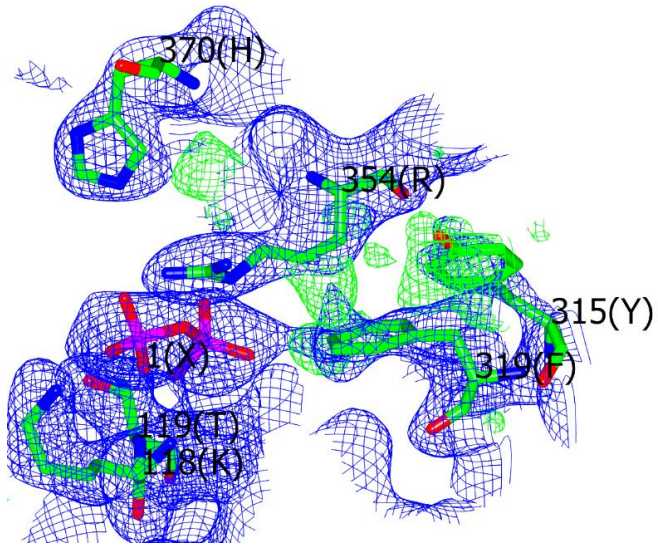


Figure 4.10 The pyrophosphate (POP)-binding site of the Mfer336 structure. The electron density between the POP and Phe319 is connected and has a positive electron density (Fo-Fc) shown as green. The alternative conformation by Phe319 is indicated by the Fo-Fc electron density around the side-chain, but could not be modelled due to steric hindrance with side-chain of Tyr315. The electron density map (2Fo-Fc) is contoured at 1.0-σ and that for Fo-Fc (green) is at 3.0-σ. The residues within 3.5 Å of the POP molecule (labelled as 1(X)) are labelled using single amino acid representations.

The N-terminal domain of the Mfer336 structure has a loop region (residues 36-41), which appears to be flexible and is indicated by the high B-factors for the residues within this region (Figure 4.11). The residue Lys39, due to the limiting electron density within this loop, has been truncated to Cβ. Moreover, the loop comes close to another loop between β3 and α3 (residues 56-59) within the same domain. The B-factors for residues in this region are also high again indicating that the loop is also flexible (Figure 4.11). The conformation adopted by the flexible loop region 36-41, propagates the conformational flexibility on to the loop region between α6 and α7 in the middle domain. The residues Gln144 and Gly145 could not be modeled into the final Mfer336 structure due to limited electron density. The loop between α6 and α7 is positioned at the crescent-shaped cavity formed at the interface of the three domains and is the putative site for substrate binding. Furthermore, the loop and, in

particular, the missing residues of the loop are close to the presumed POP-binding site of the putative ATP-binding site.

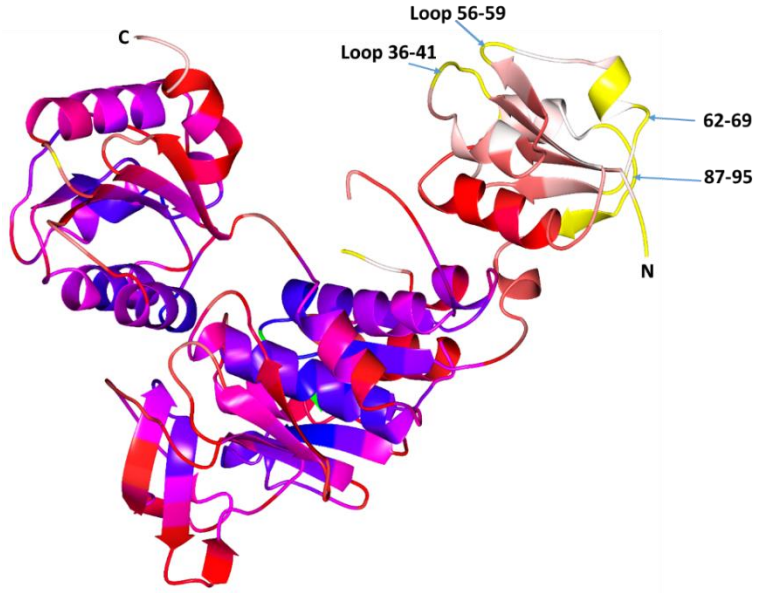


Figure 4.11 Cartoon representation of Mfer336 structure illustrating the flexible N-terminal domain. The figure was prepared using CCP4MG and coloured based on the temperature factors. The region in blue represents the region with low (25 $Å^2$), red with average (50 $Å^2$) and white with high (100 $Å^2$) B-factors. Interpolation around colour wheel feature of CCP4MG led to the other colours (magenta, purple and salmon) seen in the figure. The B-factors below and above the range are shown in green and yellow, respectively.

4.5.1 Proposed ATP-binding site

In spite of several attempts for co-crystallisation and soaking experiments using ATP and its derivatives, no structure of the Mfer336 structure in complex with the ATP or any of its derivatives such as ADP or AMP could be solved during this project. However, a presumed POP-binding site is identified close to the conserved P-loop at the interface between the middle and C-terminal domains. This is the location for the ATP-binding site in bacterial structures and the presumed POP aligns with the α- and β-phosphate position of the ATP

molecule when compared to the ATP-bound MurC structure and will be explained further in the next section. The region close to the presumed POP-binding site contains positive electron density (Fo-Fc) that extends continuously from the O2 atom of the POP molecule towards the side-chain of Phe319 as shown in Figure 4.10. In addition, the residue has a high B-factor at the side-chain atom (highest 70.7 $Å^2$ for CZ atom) and the conformation adopted by the side-chain prevents the modeling of adenosine nucleotide binding. The switch of Phe319 side-chain from t80 rotamer to p90 would be required to fit the adenosine nucleotide into the cavity based on homology with the bacterial ATP-containing Mur structures but would result in steric hindrance with Tyr315. Hence, it appears that an appropriate conformation of the Phe319 side is required for ATP binding in Mfer336 structure and is prevented in the current apo structure crystal form. Further, Arg354, from the connecting loop between the middle and C-terminal domains, is in close proximity to the POP-binding site. The arrangement of the middle and C-terminal domains creates a narrow cavity for ATP binding into the site though the side-chain conformation of the Phe319 and Tyr315 pointing towards the putative ATP-binding site (Figure 4.12).

The structure of bacterial MurC containing the phosphoaminophosphonic acid-adenylate ester (ANP), analogue of ATP, from *Haemophilus influenzae* (1P3D) was used to compare and model the binding of the molecule into the Mfer336 structure based on structural superposition. Structural comparison with the 1P3D structure suggested that the Mfer336 structure has a similar ATP-binding site as seen for the bacterial MurC (Figure 4.12). Similar to the bacterial MurC, the ATP-binding site is present at the interface of the middle and C-terminal domains with key interactions with the adenine rings and α- and β- phosphate provided by the residues of the middle domain, and specifically the conserved P-loop of the domain (Mol, *et al.*, 2003). The consensus sequence for the P-loop required by the Mur

family protein is present between the β6 and α6 of the Mfer336 structure and closely matches the sequences observed in bacterial MurC. Bacterial MurCs have a consensus sequence GTHGKTTT (Smith, 2006), whereas the Mfer336 structure contains the closely similar P-loop formed by GTNGKTTT and is conserved in all other members of the putative pMurC family (Figure 4.19). The histidine residue (His127 of 1P3D) of the bacterial P-loop is substituted by Asn116 in the Mfer336 structure. The structure superposition of Mfer336 with the ANP-containing *H. influenzae* MurC (1P3D) structure suggests that the ATP-binding residues in MurC are conserved in the Mfer336 structure (Figure 4.12). This proposition for the conserved ATP-binding site in Mfer336 is based on structural homology with bacterial MurC and is also supported by the presence of a presumed pyrophosphate (POP) at the α- and β-phosphate position of the ATP analogues in bacterial MurC (both ANP in the 1P3D and AMPPCP in the 1GQY structure).

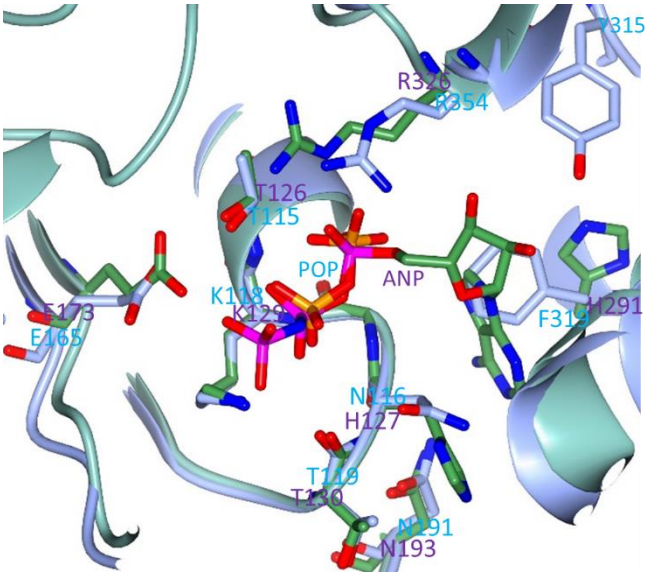


Figure 4.12 Putative ATP-binding site of the Mfer336 structure. The ANP molecule from the *H. influenzae* MurC (1P3D) structure (shown in light green) is modelled based on structure superposition to identify the putative ATP-binding site in the Mfer336 structure (light blue). The residues within 4.0 Å from the ANP molecule in 1P3D structure are shown as sticks and coloured

based on atom type. The carbon atoms are coloured green and phosphorous atoms as magenta for ANP molecule and phosphorous atoms for POP are coloured orange. The Tyr315 residue that hinders the alternative conformation for Phe319 is also shown as sticks at the top right.

4.5.2 Zinc-binding site

The Mfer336 structure contains a metal-binding site in the middle domain of the enzyme. The amino acid residues between Tyr239 to Thr289 in the middle domain extends away from the conserved core middle domain and appears as an insertion region compared to the middle domain of bacterial murein ligases. The insertion region is approximately 45 amino acid residues long and contains two antiparallel strands (β11 and β12) resulting in a β-hairpin loop and has four Cys residues arranged such that each two Cys residues lie on two different ends of the β-hairpin structure (Figure 4.13A).

The cysteine residues contributing to the Zn^{2+}-binding pocket are present at the end of the antiparallel sheet suggesting that the binding of Zn^{2+} could be stabilising the surface exposed loop region and the antiparallel strands to maintain the conformation. The position of the Zn^{2+}-binding region is at the solvent exposed surface of the protein structure, close to the interface with the symmetry molecules. The side-chains of the four cysteine residues, Cys257, Cys260, Cys279 and Cys281, within the insertion region are arranged in a tetrahedral orientation (Figure 4.13B). The distance between the Zn^{2+} and the sulphur atoms of the Cys residues are at approximately 2.3 Å distance, thus favouring the interactions as seen in other available Zn^{2+}-containing structures (Zheng *et al.*, 2008). The zinc-binding Cys residues are conserved among all other members of pMurC (Figure 4.19), thereby, suggesting that similar Zn^{2+}-binding motif and extended region in the middle domain might

be structurally conserved in pMurC peptide ligase in all pseudomurein-containing methanogens.

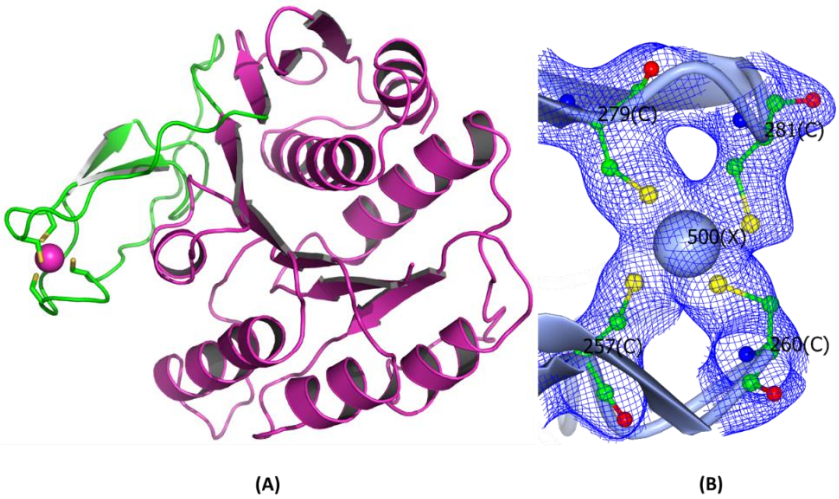


Figure 4.13 The insertion region of the middle domain containing the zinc-binding site in the Mfer336 structure. (A) The Mfer336 middle domain highlighting the insertion region forming the zinc-binding site (green). The P-loop is shown in blue. (B) Four cysteine residues forming the zinc-binding site through tetrahedral arrangement of the side-chain. The electron density (2Fo-Fc) for the Cys residues are shown at 1.5σ using the CCP4MG graphical program (McNicholas, *et al.*, 2011). The Zn^{2+} ion is represented as sphere and labelled 500(X) and cysteine residues are shown as balls and sticks.

4.5.3 Substrate-binding site comparison

The Mfer336 structure has been solved in an apo state and limits the information related to the putative substrate binding site. However, the structure-based sequence alignment with the bacterial MurC and MurD leads to a suggestion that the putative substrate-binding site for Mfer336 is similar to that of bacterial MurC. The putative ATP-binding site has been proposed and discussed earlier in Section 4.5.1. The proposed ATP-binding site is formed by mostly invariant residues that are present in all murein ligases (MurC-F) (Bouhss, *et al.*,

1997). Furthermore, the crystal structure of fully assembled substrate- and product-bound complex of *H. influenzae* MurC is available and comparative analysis using the structures suggest that Mfer336 could possibly possess a similar binding cavity (Figure 4.15). The Mfer336 structure shares an N-terminal domain similar to that of MurC and MurD. Moreover, the residues that are involved in UDP recognition of the UDP-*N*-acetylmuramic acid by these ligases are also conserved in the Mfer336 structure (Mol, *et al.*, 2003) and are shown in Figure 4.14.

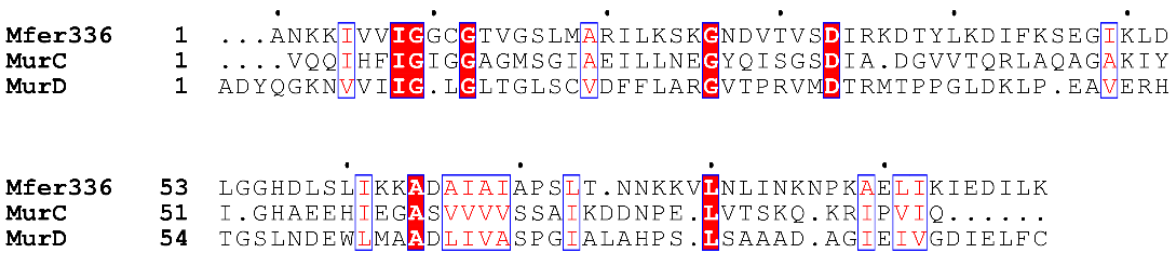


Figure 4.14 Structure-based sequence alignment of the Mfer336 N-terminal domain in comparison to the bacterial MurC and MurD. The residues that are conserved in both MurC and MurD (Mol, et al., 2003) also appear to be conserved in Mfer336.

UDP-*N*-acetylmuramoyl-L-Ala (UMA) is the final product for bacterial MurC. The UMA-bound MurC (1P3D) was used for comparative analysis to identify and propose the putative substrate/product-binding site in the Mfer336 structure. Based on the hypothesis that Mfer336, a member of pMurC, performs a similar function to that of bacterial MurC, the putative substrate binding to the Mfer336 structure is N^{α}-UDP-Glu-P and the product is N^{α}-UDP-Glu-L-Ala. Both the MurC and pMurC have a UDP group in common as part of their substrates and it is suggested that the binding of the substrate to Mfer336 would require either a conformational change in the structure or the substrate to follow a different path during binding. The flexible loop between the α6 and α7 in the middle domain appears to

hinder possible substrate-binding at the putative binding site proposed from structural comparison analysis (Figure 4.15). It was also noticed that the Mfer336 structure exhibited an open conformational state relative to the 1P3D structure, as indicated by the large binding cavity with additional binding space, even after the modelling of product of suitable length bound to the cavity (represented by black line, Figure 4.15). It has been noted that the murein ligases remain in the open or intermediate conformation, which later adopt a closed conformation upon binding of the substrate and/or the product (Sink, *et al.*, 2016). This suggests that the conformational change in the flexible loop between α6 and α7 together with re-arrangement of the N- and C- terminal domains are required for substrate binding in the Mfer336 structure.

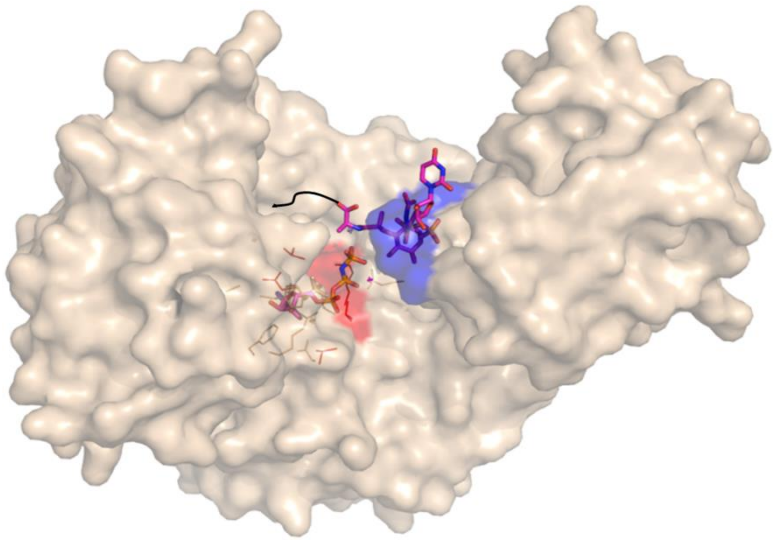


Figure 4.15 Surface view of the Mfer336 structure. The product of MurC (UMA) is modelled into the crescent-shaped cavity using structural superposition supporting the open conformation exhibited by the Mfer336 structure. The region in blue represents the flexible loop between α6 and α7 and the possible steric hindrance with the putative incoming UDP-based substrate. The black line indicates the additional space in the cavity. The putative ATP-binding site is located deep in the cavity with the proposed interacting residues represented as lines and close to the P-loop which is indicated by a red surface.

4.6 Mfer336 structure analysis

The PDBePISA webserver (Krissinel & Henrick, 2007) (PISA in short) was used to analyse the structural interface characteristics of the Mfer336 structure. The monomer analysis of the Mfer336 structure from the PISA server suggests that out of 487 residues modelled in the final structure 426 residues are surface exposed and form a solvent accessible area of 21,968.5 $Å^2$. In addition, the interface area formed by the solvent accessible residues of the Mfer336 structure with the –X+1/2, Y-1/2, -Z+1/2 symmetry molecule is approximately 515 $Å^2$ and contributes to a total of 1.8% of the interface area. The interfacing residues Lys348 and Lys462 form hydrogen bonds with Ser230 and Glu207. The residues within the zinc-binding site (257-261 and 280-283) also appear as interface residues in addition to residues in the C-terminal domain (Lys439, Glu441, Lys442, Phe444, Lys445, Phe454, Gln455, Glu458, Lys461 and Tyr466). The interfacing residues present in the Mfer336 structure are not significant to suggest dimer formation, and instead supports that Mfer336 is a monomer. The complex formation significance score (CSS) for the Mfer336 structure calculated by the PISA interface tool scored zero (0.0) value, which implies that the interface residues in the Mfer336 structure play no role in complex formation.

4.6.1 Structure similarity search

The Mfer336 structure was also used to search for similar structures against all available protein structures in the PDB database through the Dali webserver (Holm & Rosenström, 2010). The MurF from *T. maritima* in complex with ADP (3ZL8) showed the highest structural similarity with Mfer336 with a Z-score of 28.8 and a RMSD value of 3.0 Å for 353 Cα atoms and a sequence identity of 21%. The highest sequence identity was exhibited by the MurE of *T. maritima* and was calculated to be 25%. In addition to the murein ligases

MurC-MurF, several folylpolyglutamate synthases also appeared with a significant Z-score, among which the folylpolyglutamate synthase from *M. tuberculosis* (2VOR_A) showed the highest structural similarity with a Z-score of 22.1 (RMSD of 3.0 Å for 320 Cα atoms and a sequence identity of 22%). The result from Dali, based on the overall structure of Mfer336, showed a similar result as that observed for the pMurE structure and includes all available murein ligases (MurC-MurF) and FPGSs, which is assumed to be a result of the conserved middle domain that contributes the majority of the structure and the relatively conserved C-terminal domain. The middle domain of the Mfer336 structure shows similarity to all the murein peptide ligases (MurC-F) suggesting that the domain is the most conserved, as shown for murein ligases (Smith, 2006).

Each domain was individually used to analyse the structure similarity using the Dali server, which showed that the N-terminal domain of Mfer336 is most similar to the N-terminal domain of MurC. A heuristic PDB search using the Dali server for the N-terminal domain of the Mfer336 structure suggested that the N-terminal domain of *Streptococcus agalactiae* MurD (3LK7) showed the highest structural similarity with a Z-score of 12.1 (RMSD of 1.8 Å for 95 aligned Cα atoms with a sequence identity of 25%). Moreover, a range of bacterial MurC and MurD N-terminal domains showed structural similarity to the Mfer336 N-terminal domain with significant Z-scores and RMSD values, but none of the available MurE and MurF structures had structural similarity, indicating that the N-terminal domain of Mfer336 is more like a MurC/D type and does not align with the MurE/F type. In addition to the N-terminal domain of bacterial MurC/D types, the eugenol synthase 1 from *Ocimum basilicum* of the plant kingdom (PDB-3C3X) showed structural similarity with a Z-score of 11.7 and a RMSD of 2.7 Å for 98 aligned Cα atoms and a sequence identity of 23%. This enzyme is found in plants and synthesises the volatile phenylpropene compounds eugenol

and isoeugenol through the use of coniferyl acetate and NADPH (Koeduka *et al.*, 2008). Further, siroheme synthase from *Salmonella typhimurium* (PDBID-1PJQ) also shared structural similarity with the N-terminal domain of the Mfer336 structure with a Z-score of 10.6 and a RMSD of 2.5 Å for 94 aligned Cα atoms and a sequence identity of 15%. The 1PJQ structure is for a CysG protein, which results from a homo-dimeric gene fusion product containing two structurally independent modules: a bismethyltransferase and a dual-function dehydrogenase-chelatase that synthesizes siroheme from uroporphyrinogen III (Stroupe *et al.*, 2003).

The middle domain of the Mfer336 structure shows structural similarity with the middle domain of all murein ligases (MurC-MurF), with the highest structural similarity exhibited by the MurF from *Acinetobacter baumannii* (4QF5) with Z-score of 23.1 and RMSD of 1.8 Å for 189 aligned Cα atoms and sequence identity of 28%. The middle domain of available MurE and MurF structures shows a higher structural similarity compared to that of MurC and MurD as indicated by the Dali result. The obtained result indicates that the Mfer336 middle domain is more conserved to the bacterial MurE/F.

Further, the heuristic search using Dali server suggested that the C-terminal domain of the Mfer336 structure separately has highest structural similarity with the *Yersinia pestis* C092 folylpolyglutamate synthetase (3NRS) with a Z-score value of 14.8 and a RMSD of 2.1 Å for 116 aligned Cα atoms and a sequence identity of 20%. The structural similarity analysis based on the Dali result suggests that the C-terminal domain of murein ligases (MurC-F) share structural similarity with approximately similar statistics (Table 4.3). Hence, the result suggests that the C-terminal domain of the Mfer336 structure is also conserved among the murein ligases and even among the bifunctional dihydrofolate and folylpolyglutamate synthase.

Table 4.3 Structure similarity statistics for the Mfer336 structure and its individual domains obtained using Dali. The values in each cell represents the Z-score/RMSD (Å) and the number of aligned Cα atoms are in parentheses. The organism of which the enzyme types are used have been abbreviated as *Yersinia pestis* (Yp); *Thermotoga maritima* (Tm); *E. coli* (Ec); *Pseudomonas aeruginosa* (Pa); *Streptococcus agalactiae* (Sa); *Staphylococcus aureus* (Sca); *Acinetobacter baumannii* (Ab); *Mycobacterium tuberculosis* (Mt); *Lactobacillus casei* (Lc) and *Haemophilus influenzae* (Hf).

	Structural similarity with bacterial MurC					
	Hf-MurC (1P31)	**Hf-MurC-apo (1GQQ)**	**Yp-MurC (4HV4)**	**Tm-MurC (1J6U)**	**Ec-MurC (2F00)**	**Pa-MurC (5VVW)**
Mfer336	28.4/3.1 (343)	27.3/3.4 (383)	26.8/3.3 (349)	21.8/4.4 (350)	26.9/3.4 (351)	23.5/4.3 (272)
N-terminal domain	11.7/2.2 (94)	11.6/2.1 (93)	11.6/2.2 (94)	10.7/2.2 (91)	11.3/2.1 (92)	12.1/1.8 (95)
Middle domain	13.0/2.2 (114)	19.7/1.6 (159)	20.3/2.0 (170)	18.1/2.6 (173)	20.1/2.1 (175)	20.5/2.2 (176)
C-terminal domain	13.2/2.2 (121)	13.0/2.2 (121)	12.7/2.1 (120)	12.1/2.3 (113)	12.4/2.2 (121)	4.9/3.4 (96)

	Structural similarity with bacterial MurD				
	Ec-MurD (1EEH)	**Ec-MurD (4UAG)**	**Ec-MurD (2JFG)**	**Sa-MurD (3LK7)**	**Tm-MurD (4BUC)**
Mfer336	16.9/8.3 (229)	23.2/4.5 (334)	22.7/3.3 (327)	17.6/6.3 (297)	13.6/5.8 (224)
N-terminal domain	11.6/1.9 (94)	11.6/2.0 (94)	11.4/2.0 (94)	12.1/1.8 (95)	11.5/1.9 (94)
Middle domain	17.9/2.4 (174)	18.6/2.4 (175)	19.0/2.4 (178)	18.3/2.2 (176)	14.6/2.8 (168)
C-terminal domain	12.6/2.1 (115)	12.3/2.3 (115)	12.1/2.3 (115)	12.4/2.3 (113)	13.0/2.2 (114)

	Structural similarity with bacterial MurE and MurF				
	Mt-MurE (2WTZ)	**Sca-MurE (4C13)**	**Ab-MurF (4QF5)**	**Tm-MurE (4BUB)**	**Tm-MurF (3ZL8)**
Mfer336	26.6/3.6 (331)	27.1/3.2 (355)	25.1/2.6 (349)	24.6/3.3 (338)	28.8/3.0 (353)
N-terminal domain	NA	NA	NA	NA	NA

Middle domain	21.0/2.1 (193)	20.3/2.3 (193)	23.1/1.8 (189)	19.5/2.2 (187)	20.7/2.2 (189)
C-terminal domain	13.0/2.4 (120)	13.9/2.0 (118)	11.9/2.3 (117)	13.4/2.1 (120)	12.7/2.4 (119)
	Structural similarity with bacterial FolC				
	Mt-FolC (2VOR)	**Lc-FolC (1JBW)**	**Yp-FolC (3QCZ)**	**Ec-FolC (1W78)**	**Tm-FolC (1O5Z)**
Mfer336	22.1/3.0 (320)	22.1/3.1 (309)	21.4/3.5 (319)	21.4/3.6 (323)	18.0/4.5 (313)
N-terminal domain	NA	NA	NA	NA	NA
Middle domain	14.3/2.5 (179)	15.2/2.6 (177)	13.5/3.0 (172)	13.6/2.8 (173)	14.0/2.7 (176)
C-terminal domain	13.4/2.2 (116)	12.3/2.2 (101)	14.0/2.2 (116)	14.2/2.2 (117)	11.9/2.5 (112)

4.7 Structural comparison

The structure similarity search against all the available structures in the PDB database using the Dali server suggested that Mfer336 has structural similarity with the murein ligases and is most closely related to the MurC and MurD types. In addition to these ligase types, the Mfer336 structure also showed structural similarity with the middle and C-terminal domains of the MurE and MurF types as well as that of the folylpolyglutamate synthetase as explained in Section 4.6.1. Based on the Dali result, the N-terminal domain of the Mfer336 structure does not show structural similarity with the N-terminal domain of MurE/F type, indicating that the Mfer336 could be a structural homologue of either MurC or MurD. To further investigate the similarities of the Mfer336 structure based separately on individual domains, the structure of Mfer336 has been compared with each of the MurC/D peptide ligase type and are discussed separately in the following sub-sections.

4.7.1 Structural comparison with MurC

The Mfer336 structure showed the highest overall structural similarity with the bacterial MurC from *H. influenzae* among the MurC peptide ligase. The MurC structure from *H. influenzae* has been solved in both apo and a substrate-bound form. Comparative analysis has been performed using both the structures, however, the analysis based on the apo MurC structure (1GQQ) has been included in this section corresponding to apo-pMurC Mfer336. The substrate bound structure (1P3D) was used to compare and propose the ATP-binding region of the Mfer336 structure earlier in Section 4.5.1. Each domain of the 1GQQ structure was compared with the respective domain of the Mfer336 structure and analysed showing that the two structures share structural homology and is shown in Figure 4.16.

The N-terminal domain in both the Mfer336 and 1GQQ structures contain a Rossmann-like α/β fold β-sheet formed by five parallel strands flanked by alternating α-helices. This is one of the distinguishing characteristics of MurC/D ligases in bacteria (Deva, *et al.*, 2006; Smith, 2006) and is also observed with the Mfer336 structure (Figure 4.16A). Further, the core middle domain of the Mfer336 structure also resembles the structural arrangement as observed in the 1GQQ structure except for the Zn^{2+} insertion region as explained earlier in Section 4.5.2. This insertion region is unique to the Mfer336 structure and is not observed with the 1GQQ middle domain (Figure 4.16) or any other available murein ligase structure. Despite the arrangement of two β-strands at the insertion region projecting away from the core middle domain, the small antiparallel β-sheet (β13- β15) folds back on the core β-sheet to form an expanded open barrel with α11 helix in the centre that has also been observed with bacterial MurC (Deva, *et al.*, 2006). Finally, the C-terminal domain of the Mfer336 structure shares similar topology and secondary structure arrangement to that of the 1GQQ C-terminal domain as shown in Figure 4.16C. Both the structures have a six-stranded Rossmann-like fold with the first strand of the N-terminus in antiparallel fashion to the remaining five parallel strands, with the five parallel stands flanked by five α-helices.

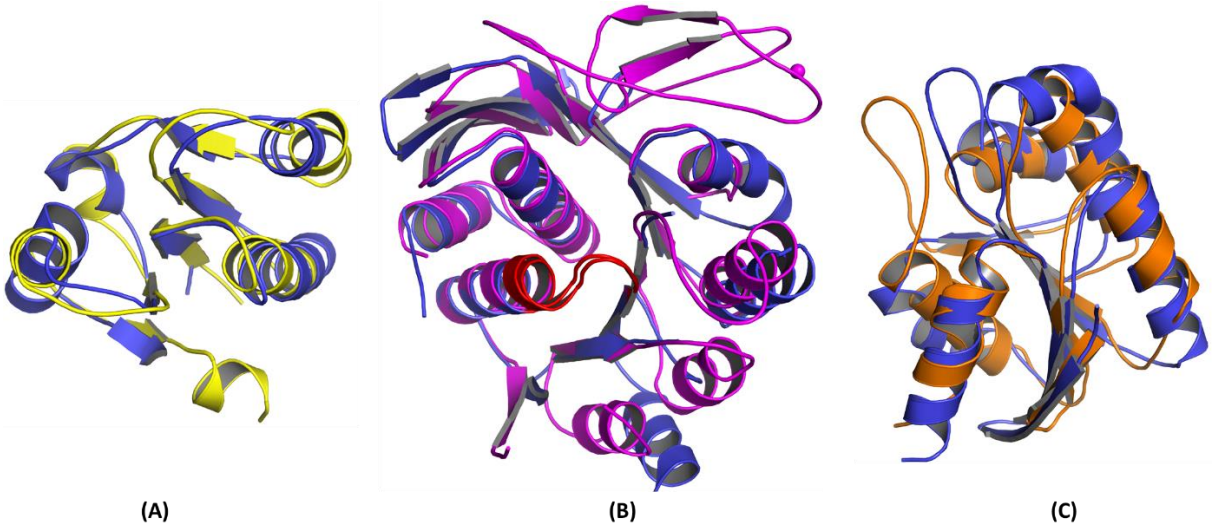


Figure 4.16 Domain comparison of Mfer336 and *H. influenzae* MurC (1GQQ). (A) The Mfer336 N-terminal domain (yellow) compared to the N-terminal domain of the 1GQQ structure (blue). (B) The middle domain comparison with the conserved P-loop is shown in red. The Mfer336 middle domain is shown in magenta and the Zn^{2+} ion as a sphere. (C) The C-terminal domain comparison; Mfer336 shown in orange.

4.7.2 Structural comparison with MurD

The *E. coli* MurD structure is the most extensively studied murein ligase and contains the largest number of structures in the PDB database. Among them the overall structural similarity of Mfer336 is close to the product-bound (UDP-*N*-acetylmuramoyl-L-Ala-D-Glu (UMAG)) *E. coli* MurD structure (4UAG) as indicated in Table 4.3. Table 4.3 indicates that the Mfer336 structure is slightly less similar to the MurD, compared to MurC, and mostly varies at the C-terminal domain. The overall structural comparison of the Mfer336 structure with the 4UAG showed differences in the position of the N- and C-terminal domains (Figure 4.17).

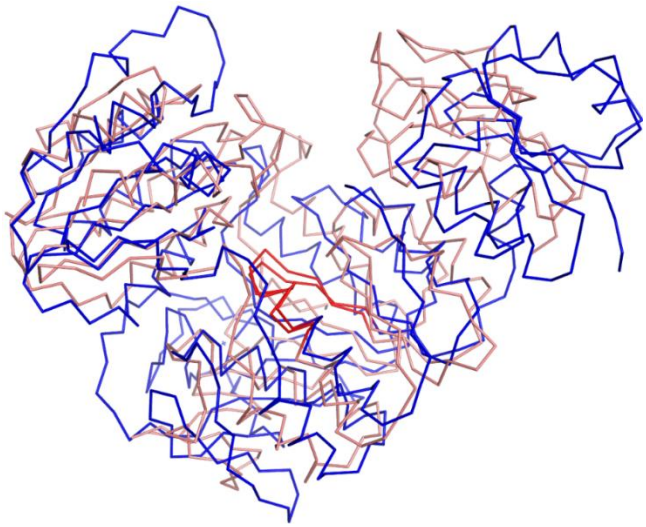


Figure 4.17 Structural comparison of the Mfer336 structure (blue) with *E. coli* MurD (4UAG) shown in light pink. Both the structures are represented as ribbons as represented in PyMOL and the conserved P-loop required by the murein ligases is represented as red.

The domains in murein ligases adopt different relative conformations based on the substrates bound to the structure. The differences observed in Figure 4.17 could be a result of different domain arrangements, hence, the structure of each domain of the Mfer336 structure and the 4UAG structure were compared individually based on the secondary structure matching using the SUPERPOSE tool in CCP4i (Krissinel & Henrick, 2004). The N-terminal domains of the two structures had a RMSD difference of 2.2 Å for 88 Cα atoms. The secondary structure arrangement between the domains were similar to one another as shown in Figure 4.18A. Similarly, the RMSD value for the middle domain comparison was calculated to be 1.8 Å for 162 Cα atoms (Figure 4.18B). Finally, the C-terminal domain showed structural similarity with a RMSD value of 2.2 Å for 109 Cα atoms and is shown in Figure 4.18C.

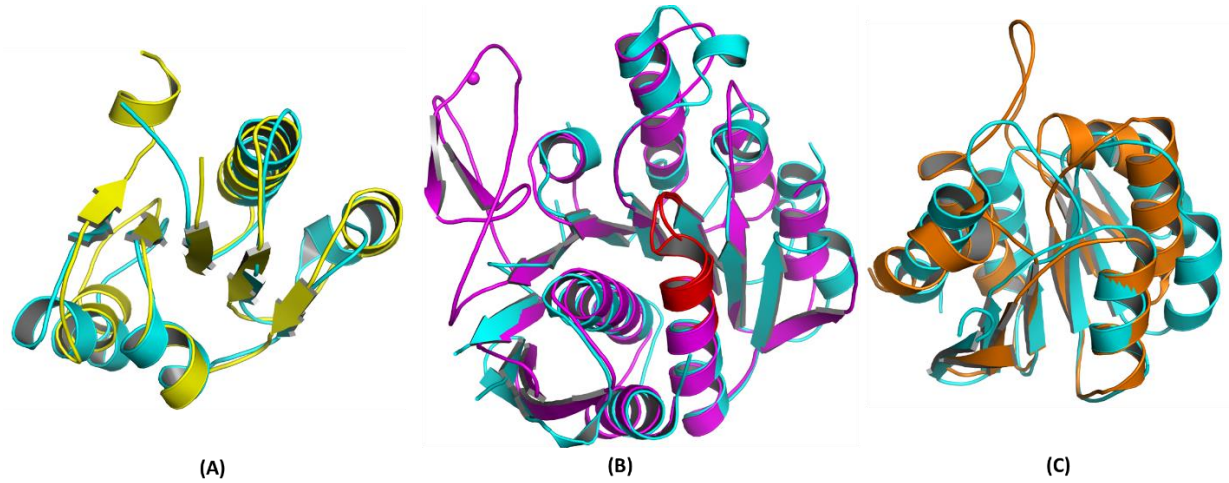


Figure 4.18 Domain comparison of Mfer336 and *E. coli* MurD (4UAG). (A) The Mfer336 N-terminal domain (yellow) compared to the N-terminal domain of the 4UAG structure (cyan). (B) The middle domain comparison with the conserved P-loop is shown in red. The Mfer336 middle domain is shown in magenta. (C) The C-terminal domain comparison (Mfer336 shown in orange).

4.8 Conclusions and discussion

The structure of a putative pseudomurein peptide ligase from *M. fervidus* (Mfer336) has been determined at 2.5 Å resolution and is best described as a structural homologue of MurC peptide ligases in bacteria. The overall Mfer336 structure shares a similar domain arrangement as exhibited by all the available murein ligases (MurC-MurF). The Mfer336 structure contains three domains arranged together to form a crescent-shaped cavity at their interface. A similar domain arrangement has been shown in bacterial murein ligases (MurC-F) and the cavity formed at the interface has been shown to accommodate the substrate and product for the respective bacterial enzyme types. The hinge region between each domain allows the flexible rearrangement of the N- and C-terminal domains around the middle domain, which is required to accommodate the appropriate sized substrate into the cavity. Murein ligases switch between the open and closed conformation through domain

movements mediated by the hinge regions (Deva, *et al.*, 2006; Smith, 2006). Recently, the intermediate conformation for the apo and substrate-bound MurD has also been reported (Sink, *et al.*, 2016). Due to the limitation of a single crystal structure determined for the pMurC type, it is not clear from the available data to show that the connecting loop between Mfer336 contributes to domain movement. However, the structural homology with the MurC ligase type and their putative role in catalysing a similar amino acid ligation reaction suggests that the connecting loop in the Mfer336 structure most likely performs as a hinge region. Moreover, the connecting loop in putative pMurE types of pseudomurein-containing methanogens are involved in domain rearrangement and has been shown to represent an analogous hinge region.

Among the three domains of the Mfer336 structure, the ATP-binding middle domain contributes to the majority of the structure and is the most conserved domain in both murein (Smith, 2006) and pseudomurein peptide ligases as discussed further in (Figure 6.3). The middle domain shares high sequence as well as structure similarity among all murein ligases (MurC-F) and contains invariant amino acids that are involved in ATP binding (Bouhss, *et al.*, 1997; Smith, 2006). These amino acid residues are also conserved in the Mfer336 structure and are proposed to be involved in the formation of the ATP-binding site at the interface of middle and C-terminal domains of the Mfer336 structure. The position at which the ATP-binding site is present in all available murein ligase structures shows that the binding of the ATP occurs at the interface of middle and C-terminal domains, as proposed for the Mfer336 structure, where the C-terminal domain could contribute to capping the ATP-binding site as explained by (Mol, *et al.*, 2003).

The N-terminal domain in the Mfer336 structure shows the highest structural similarity with the N-terminal domains of MurC and MurD peptide ligases. This is the most distinct domain

and does not show obvious structure similarity with the N-terminal domain of MurE and MurF peptide ligases. This observation clearly indicates that the pseudomurein peptide ligases also contain two distinct types of N-terminal domain (pMurC/D and pMurE/F) and Mfer336 belongs to the pMurC/D type. Moreover, the structural comparison of N-terminal domains with the bacterial MurC and MurD shows that the Mfer336 N-terminal domain contains conserved residues that are involved in UDP-substrate recognition by bacterial MurC as shown in Figure 4.14 (Mol, *et al.*, 2003). The overall Mfer336 structure similarity statistics from the Dali result and N-terminal domain analysis suggest that the Mfer336 is more similar to MurC compared to MurD (Table 4.3). Hence, based on observation obtained from i) structure similarity statistics from Dali, ii) comparative structural analysis and iii) initial bioinformatics analysis used to group the enzyme, suggest that Mfer336 is a member of pMurC in pseudomurein-containing methanogens and could putatively involve in catalysing a similar function as MurC, ligation of L-Ala to the cross-linking pentapeptide stem in pseudomurein cell wall biosynthesis.

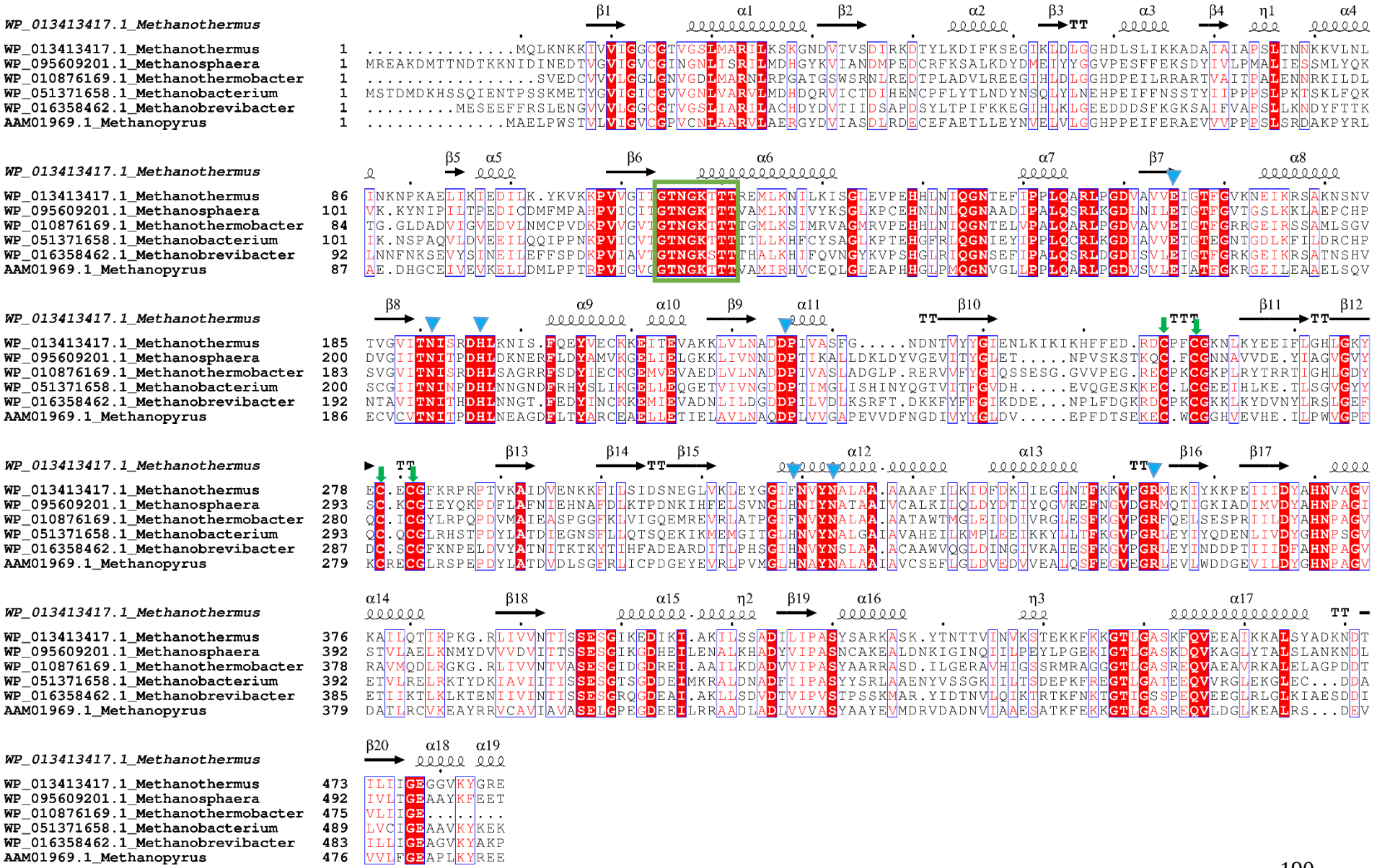

WP_013413417.1_Methanothermus
WP_095609201.1_Methanosphaera
WP_010876169.1_Methanothermobacter
WP_051371658.1_Methanobacterium
WP_016358462.1_Methanobrevibacter
AAM01969.1_Methanopyrus

Figure 4.19 Multiple sequence alignment of pMurC. The multiple sequence alignment highlights the key conserved residues of the enzyme class and provides insight into the conserved ATP and substrate binding sites. For clarity, only representative species from respective genera of pseudomurein-containing methanogens have been included. Organisms used in the alignment are *Methanothermus fervidus*, *Methanosphaera stadtmanae* DSM3091, *Methanothermobacter thermautotrophicus* ΔH, *Methanobacterium* sp. MB1, *Methanobrevibacter* sp. AbM4 and *Methanopyrus kandleri* AV19. The conserved P-loop is highlighted in a green box. In addition, the residues that are conserved in all murein ligases and involved in ATP- and Mg^{2+}-binding (Smith, 2006) and also conserved in pMurC are indicated by blue triangles. The conserved cysteine residues forming the Zn^{2+}-binding site are indicated by green arrows.

The Mfer336 structure is unique compared to the bacterial MurC at the ATP-binding middle domain in that it contains an additional β-strand in the core β-sheet. Even though the topology of the Mfer336 structure is similar, the overall structure shows an insertion of approximately 40 amino acid residues region extending away from the core middle domain. This region contributes to an antiparallel β-sheet formed by two stands and contains four Cys residues forming a Zn^{2+}-binding site. The Cys residues are conserved in all other members of the pMurC as shown in Figure 4.19. The role of the zinc binding site in pseudomurein peptide ligase is not clear and has not been found in murein ligases. The presence of four Cys residues forming an interaction to bind the Zn^{2+} within a short linear sequence (approximately 25 amino acids) and absence of water molecules in the interaction suggest that it is a structural zinc site (Auld, 2001). The role of a structural zinc site in general is understood to be involved in maintaining the structure of the protein around this region and has been reported to be present in a number of alcohol dehydrogenases, protein kinases, matrix metalloproteinases and tRNA synthases (Auld, 2001). Of most interest, a similar zinc-binding site has been studied in the RNA polymerase subunit RPB10 of *M.*

thermautotrophicus (1EF4), a pseudomurein-containing methanogen, which has a metal ion bound within an atypical CX_2CX_nCC sequence motif (Mackereth *et al.*, 2000). The Mfer336 structure contains closely related CX_2CX_nCXC sequence motif with an additional amino acid between the third and fourth Cys of the motif and has a slightly different secondary structure. Unlike, the α-helix stabilised by the zinc binding in the 1EF4 structure, the Mfer336 structure forms two antiparallel strands (β11 and β12) between the Cys residues forming the tetrahedral arrangement (Cys260 and Cys279) and appears to stabilise this short β-sheet ultimately the structure around the insertion region (Figure 4.13).

Moreover, the zinc-binding site in Mfer336 structure is formed at the interface of the N-terminal domain of the Y+1/2, -X+1/2, Z+1/2 symmetry molecule and C-terminal domain of the X+1/2, -Y+1/2, -Z+1/2 symmetry molecule (Figure 4.20) suggestive of a potential protein interface mediated by the Zn^{2+}-Cys complex. The molecular mass determination for Mfer336 and PISA analysis for the Mfer336 structure co-ordinates both suggest that Mfer336 is monomeric, and hence, indicating that the Mfer336 structure could form a protein-protein complex with another protein rather than forming a homodimer. This proposition is supported by the gene fusion suggested for Mth530 (Mfer336 equivalent in Mth) and Mth531 by STRING analysis and the large protein size for equivalent genes in other members of the genus *Methanothermobacter* as explained earlier in Section 4.2.2.

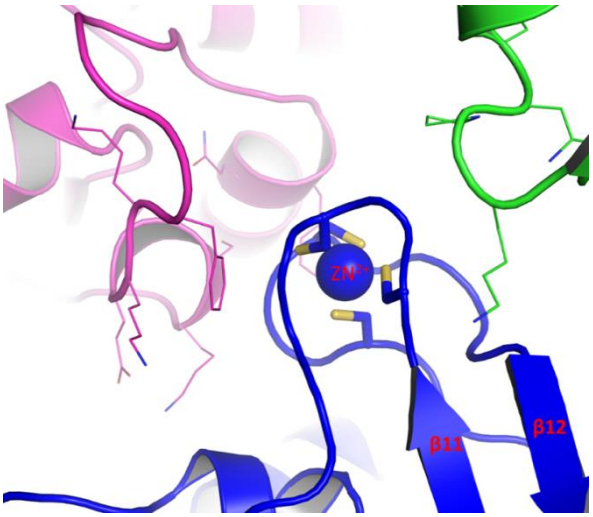


Figure 4.20 Symmetry arrangement of the Mfer336 structure (blue). The structure in green refers to the section of N-terminal domain of the Y+1/2, -X+1/2, Z+1/2 symmetry molecule and that in magenta refers to section of the C-terminal domain of X+1/2, -Y+1/2, -Z+1/2 symmetry molecule. The zinc ion and the antiparallel β-strands stabilised by the zinc-binding site are labelled in red.

Even though the zinc-binding motif of the murein-related protein family has been reported for the first time through the Mfer336 structure, similar CX_2CX_nCXC sequence motifs, referred to as a zinc-bundle (Mackereth, *et al.*, 2000), seem to be present in a few bacterial murein ligases from different bacterial clades including *Firmicutes* and *Staphylococcus* in particular. Mur ligase family proteins that showed sequence similarity with the insertion region and contained a similar sequence motif had sequence homology with MurE ligases in other bacterial species. A PSI-BLAST search against the pseudomurein-containing methanogens showed specific hits with MurE family and a DUF1727 domain at the C-terminus but appeared to be in members of the putative pMurT ligases or putative pMurC among those that do not possess pMurT. In bacteria, the MurT ligase in complex with glutamine amidotransferase-like protein (GatD) catalyses the amidation of the α-carboxyl group of D-glutamic acid of cell wall precursor stem peptides. Further, MurT shares sequence

similarity with MurE and is able to substitute for MurE in an *in vitro* UDP-MurNAc-pentapeptide synthesis assay (Munch *et al.*, 2012).

Moreover, the amino acid sequence of the extra domain in the Mfer336 structure showed homology with the MurT, whose structure has been recently reported (Morlot *et al.*, 2018). The structure of the MurT is reported to have structural similarity with the middle and C-terminal domains of the murein peptide ligases with an additional cysteine-rich insertion region that is assumed to be involved in Zn^{2+}-binding. The structure and sequence analysis presented here was performed using the available MurT structure (6FQB) and indicate that it is structurally similar to the Mfer336 structure but does not have the N-terminal domain (Figure 4.21). The structural comparison of the MurT structure with the bacterial murein ligases shows a cysteine-rich insertion region that corresponds to the Zn^{2+}-binding site in the Mfer336 structure. The structures compare with RMSD value of 2.0 Å for 210 Cα atoms from the middle domains and shows a rigid body rotation of approximately 10° from the middle domain. The sequence identity of the Mfer336 with the *S. pneumoniae* MurT is 24%.

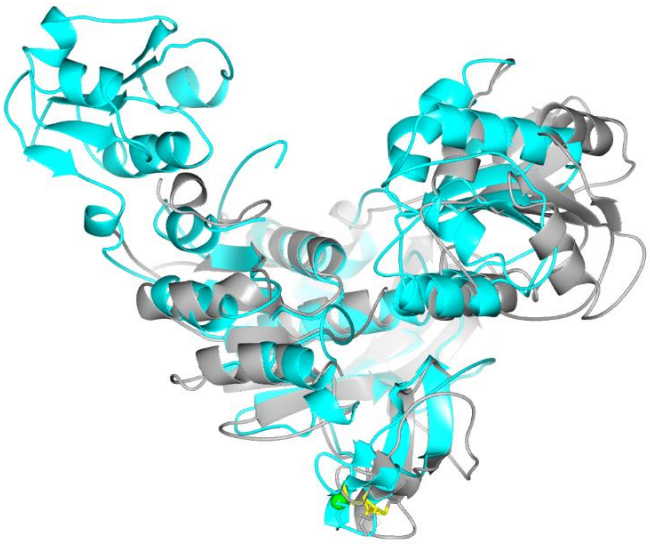


Figure 4.21 Structure comparison of Mfer336 with the *S. pneumoniae* MurT. The structure of MurT is shown as grey and does not contain the N-terminal domain equivalent of the Mfer336 structure.

The insertion region that contains the Zn^{2+}-binding site in Mfer336 structure is shown at the bottom with zinc ion is represented as sphere and the Cys residues represented as sticks. The Cys-rich insertion region of MurT structurally aligns well with the one in the Mfer336 structure.

Hence, based on the observation that the homologous CX_2CX_nCXC sequence motif is present in murein ligase homologue (MurT) and also forms a similar structural motif in the overall structure compared to that of the Mfer336 structure, it could be proposed that the Zn^{2+}-binding domain could be involved in similar function in both the structures. It can be proposed that the Zn^{2+}-binding domain is involved in maintaining the structure of the enzyme that would facilitate protein-protein interactions and that the cysteine-rich Zn^{2+}-binding site might be involved in the structure stability of such additional domain. Protein-protein interactions have been reported among the enzymes involved in the murein cell wall biosynthesis where the adjacent genes in the gene cluster are fused by a linker contributed by approximately 20 residues in a number of bacterial species (Laddomada *et al.*, 2016). Similar observations have been found in the pseudomurein ligases, and especially with the pMurC and pMurD1 of the genus *Methanothermobacter*, where the two genes start immediately after the first one and the linker encoded between these genes is not defined.

Chapter 5 Pseudomurein ligase type D

5.1 Introduction

MurD in bacteria is responsible for the addition of the second amino acid (D-Glu) to a growing peptide (UDP-MurNAc-Ala), a product of the MurC catalysed reaction, during the biosynthesis of the cross-linking pentapeptide (Sink *et al.*, 2013; Sink, *et al.*, 2016). The MurD structure also possesses a three domain architecture with the middle domain at the centre flanked by N- and C-terminal domains (Bertrand *et al.*, 1997). The amidation of the D-Glu residue of the monomeric peptide subunit to form iso-D-Gln in some common Gram-positive bacterial clades through the homologous murein ligase, MurT, in complex with the GatD enzyme (Munch, *et al.*, 2012). It has been suggested that the MurD catalysed reaction during murein biosynthesis is the biosynthesis step which could regulate the maintenance optimal murein cell wall thickness during the cell growth in Gram-negative bacteria (Walsh *et al.*, 1999).

Homologous gene searches using PSI-BLAST showed that the genes encoding MurD in *Bacteria* are also present in pseudomurein-containing methanogens. Moreover, two types of pMurD, referred to as pMurD1 and pMurD2, have been identified and both are present only in pseudomurein-containing methanogens. Further, the gene cluster analysis suggests that the pMurD1 and pMurD2 are usually present close to genes encoding pMurC in the genome, but not always together except for the available genomes of *Methanothermobacter* species (Section 4.2.2). The *Methanothermobacter* sps have the pMurC and pMurD1 putatively fused or overlapping one another and other pseudomurein containing methanogen genomes have these genes adjacent to one another.

5.2 Bioinformatics analysis

5.2.1 Homologous protein search

The amino acid sequences of pMurD1 and pMurD2 from *M. fervidus*, Mfer337 (WP_013413418.1) and Mfer340 (WP_013413421.1), respectively, were used to perform the searches for homologous protein sequences in pseudomurein-containing methanogens using PSI-BLAST (Altschul, *et al.*, 1997). It was observed that the pMurD1 homologous gene in available genomes of the genus *Methanothermobacter* were approximately double the typical sequence length for a peptide ligase and corresponded to the same gene as that for the *pMurC* search result. The sequence alignment indicates that the gene was a result of the fusion of two genes, *pMurC* and p*MurD1*. The Mth531_new amino acid sequence was obtained by merging two genes, *mth_530* (WP_010876169.1) and *mth_531* (WP_010876170.1), and manually dividing them again into two genes *mth530_new* and *mth531_new* based on similar gene length and sequence alignment (Wolf, 2010). Hence, Mth531_new is a pMurD1 ligase encoded by *mth531_new* in *M. thermautotrophicus* ΔH which is different to the Mth531 (WP_010876170.1) available at the NCBI protein database. The PSI-BLAST result using the amino acid sequence of the Mth531_new supports the result as obtained from the Mfer337 sequence.

5.2.2 Phylogenetic analysis

A phylogenetic analysis to infer the evolutionary history of pMurD1 and pMurD2 in relation to bacterial MurD sequences was performed using the maximum likelihood method (Figure 5.1). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino

acid substitutions per site. The analysis involved 65 amino acid sequences that were retrieved as explained in Section 2.2.1.3. There were a total of 270 positions in the final dataset that were considered for evolutionary analyses using MEGA7. It was observed that both the pseudomurein peptide ligases pMurD1 and pMurD2 clade separately in the phylogenetic tree and so do the MurD sequences. The pMurD2 from the *Methanopyrus kandleri* roots deeply among the pseudomurein-containing methanogens, as are the *M. fervidus* and *M. thermautotrophicus* along with *M. kandleri* in the pMurD1 clade. *Thermus thermophilus* MurD roots deeply in the MurD clade.

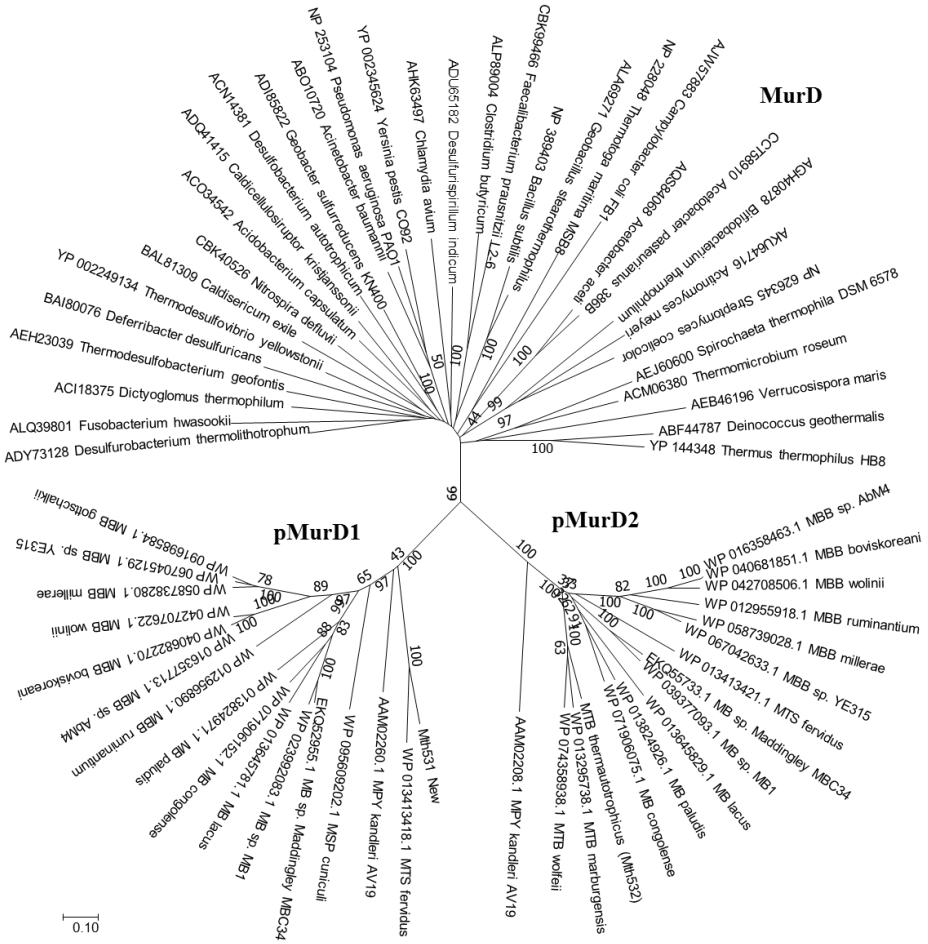


Figure 5.1 . Evolutionary relationships of bacterial MurD and archaeal pMurD1 and pMurD2 ligases.

5.3 Cloning, expression and purification

The plasmids used for cloning of pMurD targets from *M. fervidus* (Mfer337 and Mfer340) were initially obtained from the previous work of Youri Van Nuland (Van Nuland, 2011). The protein yields using the pET151D-based expression plasmids of both *mfer_337* and *mfer_340* were low, hence, at later point these genes were synthesised and obtained from GenScript (USA). The pET15b-based expression plasmids obtained from GenScript were codon optimised using the OptimumGene™. Further, the *mth_531_new* and *mth_532* genes from *M. thermautotrophicus* ΔH that were cloned into pET151D and pET101D expression vectors, respectively, were also obtained from the previous work of Maximilian Wolf (Wolf, 2010). The sequence-verified plasmids were transformed into *E. coli* BL21 Star™ (DE3) competent cells with ampicillin as a selectable marker as explained in Section 2.2.2.1. The transformed bacterial colonies were grown in 2×YT medium and induced with 0.1 M IPTG for protein expression Section 2.2.2.2. AI media was used for expression of the Mth532 as described by Studier (Studier, 2005). All the proteins were purified using Ni-NTA agarose gravity column affinity chromatography at room temperature and is described in Section 2.2.2.4. All the proteins eluted from the Ni-resin were analysed by SDS-PAGE (BIO-RAD Mini-PROTEAN® TGX™, Cat# 4561043) against BIO-RAD ® low range protein standards and used for crystallisation experiments.

The SDS-PAGE analysis of the purified pMurD targets produced during this project are shown in Figure 5.2. The Mfer337 and Mfer340 proteins started to precipitate when dialysed with the storage buffer containing 200 mM of KCl. However, the precipitation problem was solved by dialysing the protein in the presence of a higher salt concentration (600 mM KCl) in the storage buffer. Hence, the final storage buffer used for dialysis of Mfer337 and Mfer340 contained 20 mM MOPS buffer (pH 7.0), 600 mM KCl and 2.0 mM TCEP. In

contrast, the purified Mth532 protein was stable with a lower salt concentration and was dialysed with 20 mM MOPS buffer (pH 7.0), 200 mM KCl and 2.0 mM TCEP. All the purified proteins were concentrated to a final concentration of approximately 8.0 - 10 mg/mL as explained in Section 2.2.2.7. A small aliquot of each concentrated protein was immediately used for crystallisation experiments. Remaining proteins were aliquoted and snap-frozen in liquid nitrogen and stored at -80 °C. Also, an aliquot of each concentrated protein was stored separately for biochemical assays in the presence of 30 % (v/v) glycerol and snap-frozen using liquid nitrogen.

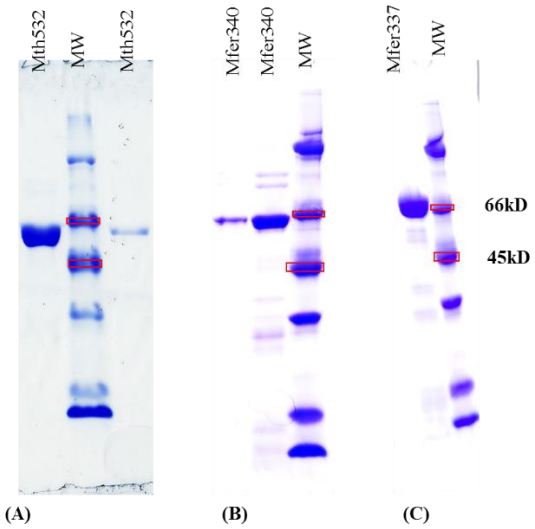


Figure 5.2 SDS-page analysis for the pMurD1 and pMurD2 peptide ligases. (A, B) Mth532 and Mfer340, members of pMurD2, (C) Mfer337, a member of pMurD1. Proteins collected from each elution step during purification process were loaded separately in the gels.

5.4 Crystallisation of pMurD targets

In addition to the aforementioned purified pMurD1 and pMurD2 targets, purified Marb918 (1-500) and Mth531 proteins were available from the previous work by Denise Schafer and Carrie Sang, colleagues working on the project in the past, and were used to screen against

all the primary crystallisation screens as summarised in Table 5.1. The proteins had been stored at -80 °C for more than three years for Marb918 (1-500) and more than two years for Mth531. Both the proteins were analysed using the SDS-page before performing crystallisation experiments. All the primary screens listed in Section 2.1.6 were used to screen the crystallisation conditions for the pMurD types. The crystallisation experiments were performed in apo as well as well as in the presence of ATP, ADP, Mg^{2+}, partial substrate (UDP), amino acid (L-Glu) and dipeptide (γ-Glu-Ɛ-Lys) and are summarised in Table 5.1.

Table 5.1 Summary of the crystallisation experiments performed and screens used for various pMurD targets.

Protein	Screens Tested								Remark
	SS1+2	JCSG+	MIDAS	MORPHEUS	SG1	INDEX	MORPHEUS II	PEG/ION	
Mth531	✓	✓	✓	✓	✓	✓	✓	✓	Apo, ± ATP, $MgCl_2$ (protein obtained from previous work)
Mth531					✓	✓			Apo protein (GenScript synthesised)
Marb918 (1-500)	✓	✓	✓	✓	✓	✓	✓	✓	± ATP, $MgCl_2$, L-Glu
Mth532	✓	✓	✓	✓	✓	✓	✓		± ATP, $MgCl_2$
Mfer337	✓	✓	✓	✓	✓	✓	✓		± ATP, $MgCl_2$, UDP, dipeptide (γ-Glu-ε-Lys)
Mfer340	✓	✓	✓	✓	✓	✓	✓		± ATP, $MgCl_2$, UDP, dipeptide (γ-Glu-ε-Lys)

A few crystal conditions were obtained during the primary screening, but subsequently displayed they had diffraction patterns consistent with salt crystals upon exposure to the X-ray beam at the Australian Synchrotron. Fresh Mth531_new protein was produced during this project to perform the crystallisation experiments and to test if the long term storage was the reason for not being able to obtain crystals despite attempts to test numerous crystal conditions and additives to enhance crystallisation. The yield of Mth531 using the expression system as explained in Section 2.2.2.2 was low and not sufficient to perform screening with

all the available screens tested during the project. However, the INDEX and SG1 screens were selected based on their successful results with other targets and used for screening. The freshly prepared protein also did not show any significant difference in initial crystallisation behaviour and failed to yield any protein crystals.

The purified Mfer337 and Mfer340 proteins, produced using both the plasmids - obtained from Youri and GenScript synthesised, were also used for crystallisation experiments. More than three different preparations of Mfer337 and Mfer340 protein were produced and each time the purified proteins were produced, a fresh aliquot was taken and used immediately from crystallisation experiments. None of the screening experiments produced robust crystals that diffracted with a protein crystal pattern. Fine needle-shaped crystals were obtained for both Mfer337 and Mfer340 in several crystallisation conditions but none of the crystal conditions could be optimised to produce better robust crystals.

5.5 Structure prediction

The structure of pMurD could not be determined during this project using the X-ray crystallography technique due to the absence of protein crystals of pMurD targets. The structure of the two different pMurD types, pMurD1 and pMurD2, have been predicted using the homology modelling technique. The models for Mfer337 and Mfer340, proposed as putative pMurD1 and pMurD2, respectively, have been predicted. The success of the project in determining the structures of pMurC and pMurE from *M. fervidus* during this project was the main reason for selecting the targets from the same organism to model the pMurD structures.

Appropriate templates were searched based on amino acid sequences of the targets (Mfer337 and Mfer340) using BLAST and HHblits incorporated for target search under model

building pipeline in the SWISS-MODEL server (Arnold *et al.*, 2006). Lists of murein peptide ligases, inclusive of all types – MurC/D and MurE/F, and folylpolyglutamate synthase (FolC) were selected as targets for model building. In addition, the models of Mfer337 and Mfer340 were also obtained using the Mfer336 and the Mfer762 structures as a template. The best model among the several models using different templates were selected based on GMQE and QMEAN values (Benkert *et al.*, 2008).

The predicted model of Mfer337 based on the Mfer762 structure as homology template did not include the N-terminal domain. Similar results were observed when modelling Mfer337 using MurE, MurF and FolC as templates. The models built using MurC and MurD were complete and were selected as the best models based on the GMQE and QMEAN values. This was already indicated while selecting the appropriate template for homology modelling through the query sequence coverage. Among 18 different models built for Mfer337 based on available bacterial murein peptide ligases, the models based on MurC from *Haemophilus influenza* (1P3D_A) and MurD from *Streptococcus agalactiae* (3LK7_A) have been selected for analysis and are referred to as Mfer337_model-Hf and Mfer337_model-Sa, respectively (Figure 5.3A and B). The Mfer337_model-Hf and Mfer337_model-Sa were selected based on the GMQE, QMEAN values and query coverage. The models have sequence identity of 23% - 26%. Moreover, the model of Mfer337 based on the Mfer336 structure was also built and is complete with 34% sequence identity. This model is referred to as Mfer337_model-336 (Figure 5.3C) and has also been used for the analysis.

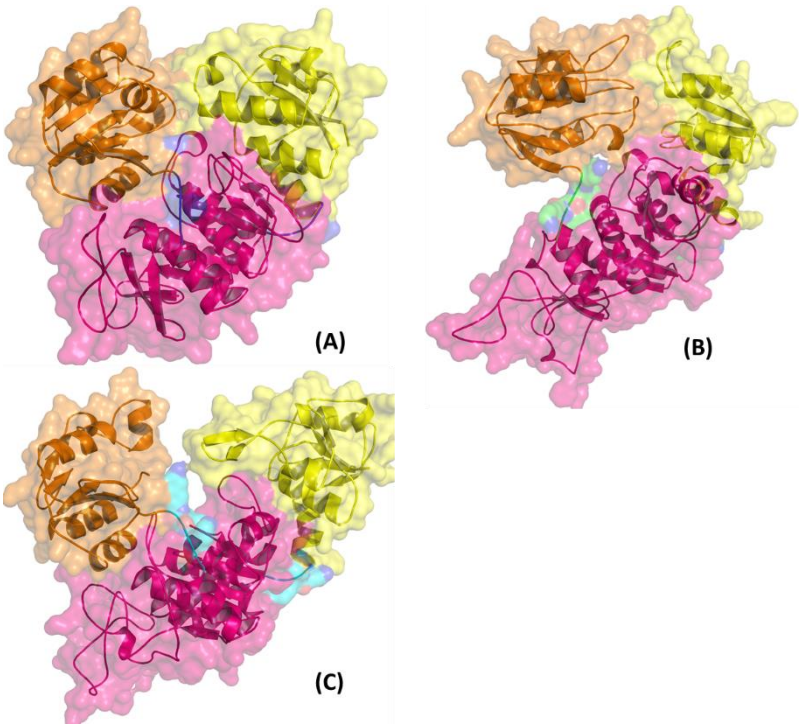


Figure 5.3 Models of Mfer337 obtained by homology modelling. (A) Mfer337_model-Hf - the model of Mfer337 based on MurC of *Haemophilus influenza* (1P3D chain A), (B) Mfer337_model-Sa - the model of Mfer337 based on MurD of *Streptococcus agalactiae* (3LK7 chain A), (C) Mfer337_model-336 - the model of Mfer337 based on Mfer336 (pMurC). Each of the models have been coloured based on the domains, yellow for N-terminal domain, magenta for middle domain, orange for C-terminal domain and have been oriented in the same orientations resembling the structure of pMurC and pMurE in Chapters 3 and 4 with N-terminal domain to the right, middle domain at centre and the C-terminal domain to the right.

Furthermore, 19 different models of Mfer340 were built against 19 different templates that included all different types of murein peptide ligases and FolC. Among the models obtained, two models that were built based on MurD from *Streptococcus agalactiae* (3LK7_A) and MurC from *Haemophilus influenza* (1P3D_A) were selected as the best models based on the statistical values and wide query coverage. These models are referred to as Mfer340_model-Sa and Mfer340_model-Hf, respectively (Figure 5.4A and B). The sequence identities of the

models to the templates are 25% and 22%, respectively. In addition, the model based on the Mfer336 structure too has also been built and shares a sequence identity of 27%. The model has been referred to as Mfer340_model-336 (Figure 5.4C).

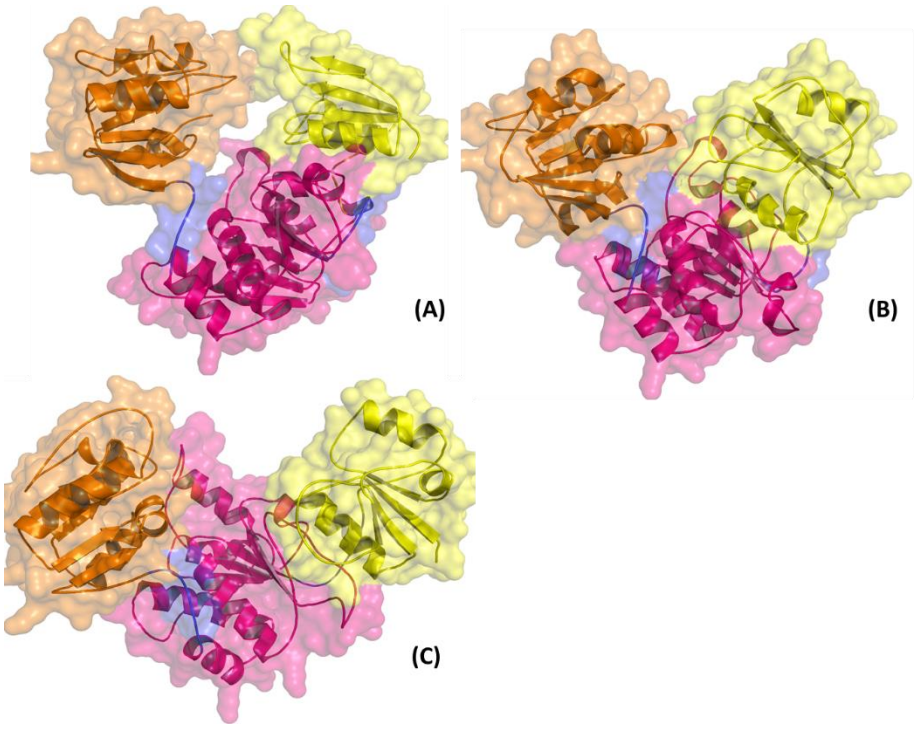


Figure 5.4 Models of Mfer340 obtained by homology modelling. (A) Mfer340_model-Sa- the model of Mfer337 based on MurD of *Streptococcus agalactiae* (3LK7 chain A), (B) Mfer340_model-Hf- the model of Mfer340 based on MurC of *Haemophilus influenza* (1P3D chain A), (C) Mfer340_model-336 - the model of Mfer340 based on Mfer336 (pMurC). Each of the models have been coloured based on the domains and is oriented as explained in previous figure.

5.6 Conclusions and discussion

The structure determination of the pseudomurein peptide ligase type D, pMurD1 and pMurD2, was not successful due to time constraints and lack of suitable diffracting protein crystals. However, the homology-based structure modelling method was used to predict the structures of the Mfer337 and Mfer340. The predicted Mfer336 and Mfer340 models were

useful in explaining the additional domain and its putative contribution to evolutionary history. It was observed that the models for Mfer337 contained a cysteine-rich insertion region at the middle domain as observed with the Mfer336. The insertion region for all the models were flexible whereas the cysteine residues in all cases were arranged such that they could form a disulfide bond aiding to the structural stability. However, the cysteine residues were not sequentially conserved in all the pseudomurein-containing methanogens. Moreover, the structure prediction using the same templates, including the Mfer336 structure, showed no insertion region is likely present in the Mfer340 models. This is also supported by a relatively low molecular weight of the pMurD2 compared to other pMur ligases. Thus, based on homology models and sequence analysis, it seems that pMurD2 do not contain the cysteine-rich amino acid sequence region, which indicates that the pMurD2 could perhaps have lost the insertion region during the course of evolution from pMurC or acquired the insertion region to form pMurC. Hence, three different scenarios can be generalised from the sequence and structural information available for three putatively related pseudomurein peptide ligases, pMurC, pMurD1 and pMurD2: i) the middle domain of pMurC contains a highly conserved cysteine-rich insertion region, also determined from the crystal structure and shown to have zinc-binding role (Chapter 4); ii) the cysteine-rich amino acid sequence region is present in the pMurD1, however, the cysteine residues are not strictly conserved in all the pseudomurein-containing methanogens. Also the predicted models suggest that a similar insertion region is present in the middle domain in all the models as observed in the Mfer336 structure; iii) relatively low molecular weight pMurD2 without the cysteine-rich amino acid sequence and also forming no insertion region in the models modelled using the same templates as that used for predicting pMurD1. Hence, these three different scenarios of these likely related pseudomurein peptide ligases suggest that the

function of the insertion region at the middle domain could have gradually either been lost or acquired during their evolution.

Furthermore, the phylogenetic analysis based on pMurD1 and pMurD2 sequences together with selective MurD sequences has been helpful to show that each of the murein and pseudomurein peptide ligase type D clade separately in the phylogenetic tree and suggest that these enzymes could have evolved from the same origin a long time ago.

Chapter 6 Discussion and summary

6.1 Understanding the pseudomurein biosynthesis pathway

The cell walls of methanogens in the orders *Methanobacteriales* and *Methanopyrales* are comprised of pseudomurein that is structurally similar to the murein cell wall in *Bacteria* (Claus & König, 2010; Cohen, 2014). The pathway for the biosynthesis of the pseudomurein was proposed by Konig *et al.* (1994) and is shown in Figure 6.1. The pathway exhibits similarities with the murein biosynthesis pathway, yet both of them display some striking differences as explained earlier in Section 1.4.1. Of most interest to this project, the amino acid residues present in the peptide subunit of pseudomurein (glutamic acid, lysine and alanine) are similar to those present in the murein peptide subunit and cross-link the glycan backbone. The addition of amino acids to the cross-linking pentapeptide of pseudomurein follows a sequential order similar to bacterial murein pentapeptide biosynthesis as shown by *in vitro* synthesis experiments using cell extracts of *M. thermautotrophicus* ΔH (Hartmann & König, 1994) (Figure 6.1). Here, it is proposed that the biosynthesis of the pseudomurein cross-linking peptide could involve similar enzymes as involved in murein pentapeptide formation and that homology analyses would help elucidating the evolutionary relationship of the murein and pseudomurein biosynthesis pathways. Bioinformatics analysis has been successful in identifying candidate methanogen enzymes homologous to the key enzymes that are involved in the murein biosynthesis pathway in bacteria. The limited presence of the homologous pseudomurein peptide ligases in only the pseudomurein-containing archaea supports their putative role in pseudomurein biosynthesis.

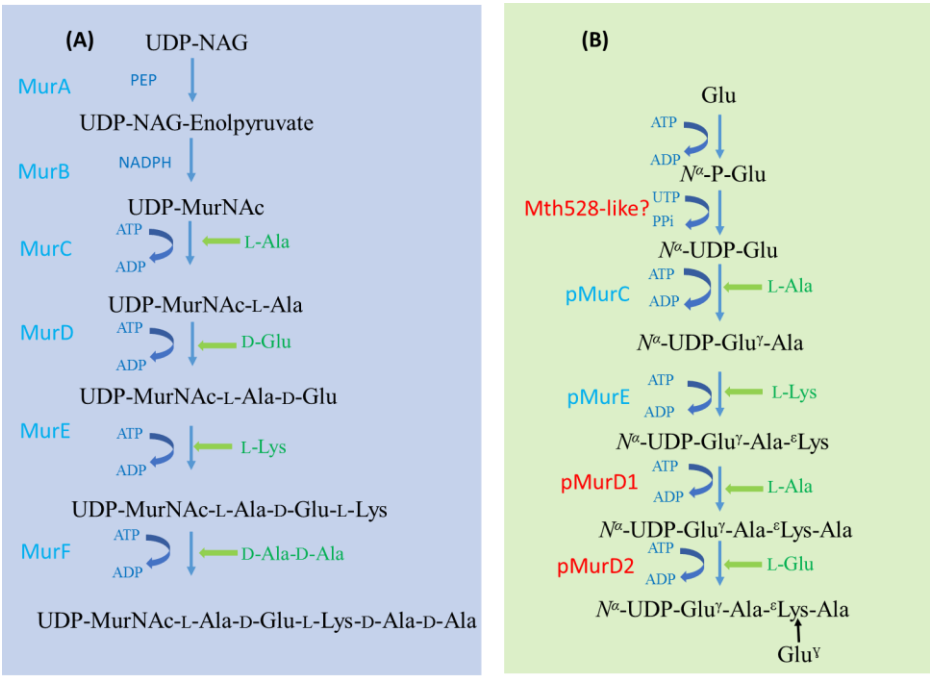


Figure 6.1 The steps for synthesis of the cross-linking peptide in murein (A) and pseudomurein (B) biosynthesis pathways. The peptide ligases for which structural information is available are labelled in blue whereas those that are hypothesised to catalyse the step based on sequence homology are labelled red.

MurA and MurB are two enzymes of the core murein biosynthesis pathway, and interestingly, their homologues are not present in pseudomurein-containing methanogens. These enzymes catalyse the formation of the initial substrate (UDP-MurNAc) to which the amino acids are ligated in sequential order by murein peptide ligases (Barreteau, *et al.*, 2008). The absence of candidate MurA and MurB enzymes in the pseudomurein biosynthesis pathway supports the proposition of pseudomurein glycan backbone formation occurring through an epimerase and additional oxidation step (Hartmann, *et al.*, 1989). The fourth peptide ligase of the murein biosynthesis pathway, MurF that is involved in the ligation of the dipeptide (D-Ala-D-Ala), is present in a wider range of methanogens, in addition to those that contain pseudomurein. The presence of a homologous pseudomurein gene encoding

MurF was perplexing at first instance as the pseudomurein peptide subunit does not contain the D-amino acid or a dipeptide. However, the presence of both homologous D-ala-D-ala ligase (ddl ligase) and MurF among wide range of methanogens led to a suggestion that these enzymes might not be limited to pseudomurein biosynthesis and could have an alternative function. This notion was recently verified, during the period of this study, that the homologue of MurF in methanogens (pMurF) is involved in co-factor F_{430} synthesis (Zheng, *et al.*, 2016; Moore, *et al.*, 2017).

Based on homology searches using PSI-BLAST, it has been proposed that an additional pseudomurein peptide ligase (pMurD2), that shares homology with MurD and pMurD1, is present in pseudomurein-containing methanogens. Hence, different to what was proposed at the start of the project, pseudomurein pentapeptide biosynthesis likely involves four peptide ligases pMurC, pMurD1 and pMurD2, and pMurE that share homology with MurC, MurD and MurE, respectively. The gene cluster analysis of the pseudomurein peptide ligases suggests that pMurC and pMurD1 are present close to one another in most pseudomurein-containing methanogens. These genes in the genus *Methanothermobacter* were found to be overlapped with one another, or adjoining as shown in Figure 6.2 for *M. thermautotrophicus* ΔH and *M. marburgensis*, respectively. Further, two genes with unknown function (*mth_529* and *mth_528* in *M. thermautotrophicus* ΔH) are found in a gene cluster with pMurC (*mth530*) and are present only among pseudomurein-containing methanogens (Figure 6.2). This indicates that these genes might be involved in the pseudomurein biosynthesis pathway and catalyse a reaction adjacent to pMurC. Homology searches using PSI-BLAST showed that Mth528 contains a putative conserved domain to that of MocA, which is a molybdenum cofactor cytidylyltransferase that releases pyrophosphate and functions in coenzyme transport and metabolism whereas Mth529 does not show a specific hit to any known super

families (Marchler-Bauer *et al.*, 2017). Hence, the putative Mth528-like enzyme might be involved in the catalysis of the N^{α}-UDP-Glu formation from N^{α}-P-Glu whereas the subsequent step of phosphorylation of the N^{α}-UDP-Glu at the γ-carboxyl group could be catalysed by putative Mth529-like enzyme (Hartmann & König, 1994).

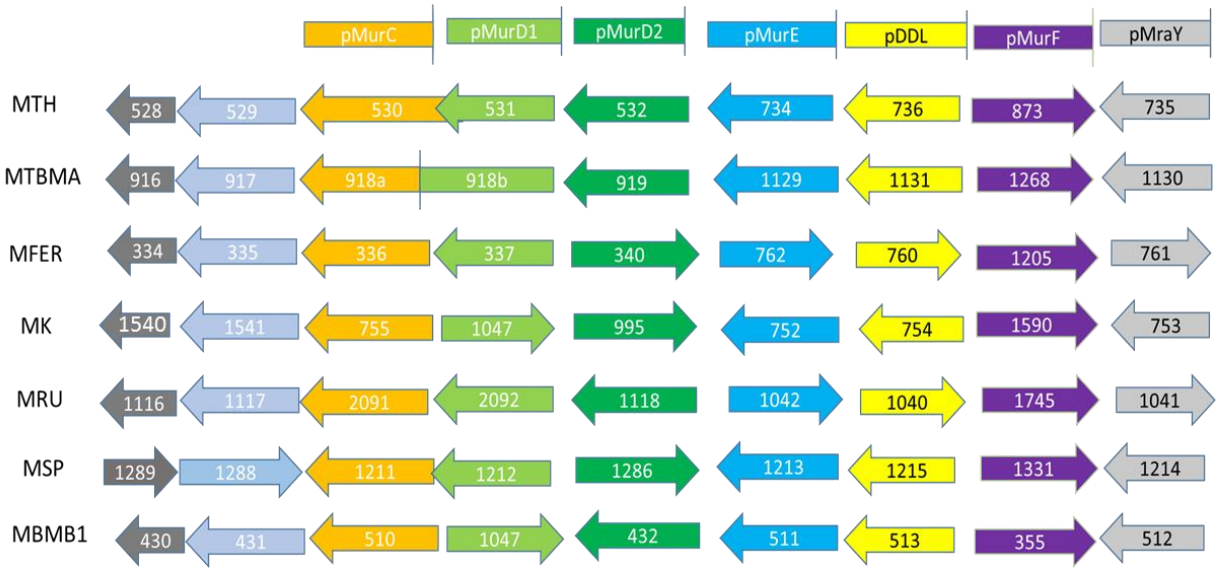


Figure 6.2 Gene cluster analysis of representative pseudomurein-containing methanogen genes putatively involved in pseudomurein biosynthesis. The arrow head of each gene represents the direction in the genome. For clarity, only a representative organism genome from each genus has been used and abbreviated: MTH; *Methanothermobacter thermautotrophicus* ΔH, MTBMA; *Methanothermobacter marburgensis*, MFER; *Methanothermus fervidus*, MK; *Methanopyrus kandleri*, MRU; *Methanobrevibacter ruminantium*, MSP; *Methanosphaera stadtmanae*, MBMB1; *Methanobacterium* sp. MB1.

Based on homology it is hypothesised that pMurC in pseudomurein biosynthesis might be involved in L-Ala ligation to the initial substrate (N^{α}-UDP-Glu). However, N^{α}-UDP-Glu$^{\gamma}$-P could be an intermediate product catalysed by the Mth529-like enzyme together with pMurC. The next step in pseudomurein pentapeptide formation is the addition of L-Lys, which again based on homology, appears to be catalysed by pMurE. The study has been

successful in establishing structural homology between the murein and pseudomurein peptide ligase C and E types. No structure for pMurD1 and pMurD2 could be determined. The putative role for these ligases could only be proposed based on the sequence homology with bacterial MurD peptide ligases and the gene cluster analysis of the available pseudomurein-containing methanogen genomes. However, the presence of two L-Ala residues, the first and the third, during the peptide ligase catalysed steps suggests that one of the pMurD (pMurD1 based on sequence analysis) could be catalysing the addition of L-Ala in the pseudomurein cross-linking peptide biosynthesis pathway. Similarly, based on the sequence homology the pMurD2 could be involved in the addition of the L-Glu as depicted in Figure 6.1.

In addition to the peptide ligases, homologues of key enzymes involved in the murein biosynthesis pathway were also found to be present in the pseudomurein biosynthesis pathway and are shown in Table 6.1. The homology analysis and identification of the enzymes further to the pseudomurein peptide ligases is beyond the scope of the study and has been presented to support the relationship between the two cell wall biosynthesis pathways. Hence, this study for the first time reports the homologous genes of the murein biosynthesis pathway and in particular the homologous peptide ligases in genomes of pseudomurein-containing methanogens which suggests that the pseudomurein biosynthesis pathway could have been a result of divergent evolution from the murein biosynthesis pathway.

Table 6.1 Enzymes identified in pseudomurein-containing methanogens that share homology with the key enzymes of the murein biosynthesis pathway. The naming is based on the homology with the bacterial enzymes. The value in each cell represents the gene number in the respective genome. The colour classification has been used for representing different genus types. Abbreviations have been used for each genus MTB; *Methanothermobacter*, MTS; *Methanothermus*, MPY; *Methanopyrus*, MB; *Methanobacterium*, MBB; *Methanobrevibacter*, MSP; *Methanosphaera*.

	pGlmU	**pMurC**	**pMurD1**	**pMurD2**	**pMurE**	**pMurF**	**pMraY1**	**pMraY2**	**pDdl**	**pMurG**	**pMurT**	**pGlu-AT**	**pMatE**
MTB ΔH	1589	530	531	532	734	873	590	735	736	884	NA	787	1471
MTB CaT2	1448	456b	456a	457	655	791	513	656	657	1503	NA	705	1337
MTB marburgensis	175	918b	918a	919	1129	1268	970	1130	1310	1279	NA	1183	550
MTS fervidus	129	336	337	340	762	1205	609	761	760	468	1209	1210	1062
MPY kandleri AV19	892	755	1047	995	752	1590	NA	753	754	977	NA	1414	1626
*MB maddingely*_MBC34	199	2106	1491	217	2107	145	1564	2108	2109	2018	1366	1367	
MB plaudis (SWAN-1)	358	428	429	381	430	316	NA	431	432	1822	1029	1030	663
MB sp. MB1	406	510	947	432	511	355	1079	512	513	1436	1172	1173	684
MB lacus AL-21	2293	2217	2216	2264	2215	2343	NA	2214	2213	249	1171	1170	1956
MB formicicum BRM9	1856	1783	1121	1829	1782	1909	1283	1781	1780	663	1377	1378	1512
MBB boviskoreani JH1	1048	522	522	521	1165	1592	NA	410	409	1318	NA		1124
*MBB_wolinii*_SH	125	998	27	999	870	1411	709	869	867	139	NA		1563
MBB ruminantium M1	456	2091	2092	1118	1042	1745	964	1041	1040	715	707	708	2136
MBB sp AbM4	226	872	87	873	981	460	622	982	983	448	NA		42
MBB smithii ATCC35601	655	1190	1191	118	359	880	66	360	361	423	1139	1138	2108
MSP stadtmanae DSM3091	1301	1211	1212	1286	1213	1331	80	1214	1215	645	581	580	165

6.2 Overall pseudomurein peptide ligase structure

The structural determination of the two putative pMur ligase types from *M. fervidus* revealed a three-domain architecture with an ATP-binding middle domain flanked by N- and C-terminal domains. In addition, the structure of the pMurE from the *M. thermautotrophicus* provides additional confirmation of the overall structure and the observations derived from the pMurE structure of *M. fervidus*. A similar domain arrangement as observed with the pMur ligase structures has been observed in available murein ligase structures with an active site assembling at the interface of each domain (Mol, *et al.*, 2003; Ruane, *et al.*, 2013). The protein structure analysis for pMurC and pMurE based on evolutionary conservation using the ConSurf webserver suggests a similar observation in pseudomurein peptide ligases and results are presented later in Section 6.3. The N- and C-terminal domains flanking the middle domain provide structural flexibility for the enzyme to accommodate the growing substrate into the crescent-shaped cavity at their interface (Deva, *et al.*, 2006; Smith, 2006). The discovery of similar pseudomurein ligase architectures suggests a presumably similar binding mechanism for the substrates and for product formation. The Mfer762-UDP_2 structure, as discussed earlier in Chapter 3, indicates the presence of a similar binding pocket for UDP as that described for the *Acinetobacter baumannii* MurE and MurF structures (Cha, *et al.*, 2014). The structural and functional relationships between the four murein ligases have been reviewed (Smith, 2006) and this publication was used for guiding comparative analysis and to interpret information regarding the pseudomurein peptide ligases determined during this project.

6.2.1 Domain movement in pseudomurein ligases

The presence of the hinge regions in pseudomurein ligases suggests that the domain rearrangement, as studied for the murein ligases (Sink, *et al.*, 2016), and exhibited by the Mth734 structure, could be expected with the pseudomurein ligases. The N- and C-terminal domains are able to move with respect to the middle domain as exhibited by the pMurE structures, and these movements are enabled by the hinge regions. The N-terminal domain of the Mfer762-UDP_2 structure moves away for the middle domain when compared to the apo structure. Similarly, a rigid body rotation of approximately 91° from the open conformation structure has been observed with the C-terminal domain of the Mth734 structure in the apo state. This rotation in the C-terminal domain supports the flexibility that pMurE could exhibit and suggests that a relatively low difference in the energy level exists for the two conformational states. In addition, this result shows a conformational change as exhibited by the C-terminal domain of *T. maritima* MurD, further indicating flexibility of the C-terminal domain to move with respect to the middle domain in the apo state (Favini-Stabile, *et al.*, 2013).

6.2.2 Distinguishing N-terminal domain in pseudomurein ligases

The N-terminal domain, the key domain that distinguishes MurC and MurD from MurE and MurF in *Bacteria*, shares a similar spatial arrangement of the secondary structure elements among the respective homologous MurC/D and MurE/F ligase types. The basic observation from the topology of the N-terminal structure in MurC and MurD in murein peptide ligases is that they contain five parallel β strands that are flanked by alternating α-helices. In contrast, the N-terminal domain of MurE and MurF structures contain five β strands with one arranged in an antiparallel fashion. The arrangement of α helices is not ordered as

observed with C/D types and predominantly contains three α helices. In general, the C/D type domain starts with a β strand whereas the E/F type begins with α helix. Even though the connectivity between the secondary structures of each murein ligase type among the two groups (C/D and E/F) is slightly different, they clearly form two distinct groups suggesting a different origin (Smith, 2006). It is observed that the pseudomurein peptide ligase structures determined during the study also exhibit similar characteristics. For example, the structure similarity search using Dali clearly indicates that the N-terminal domain of pMurE is different from the MurC and MurD structures and does not allow for any obvious secondary structure superposition. A similar result was observed with the N-terminal domain of the pMurC structure where no structural similarity was found with the N-terminal domain of available MurE and MurF structures. Hence, this study suggests that the pseudomurein peptide ligases also contain at least two distinct C and E types of N-terminal domain. The distinct characteristics of the N-terminal domain in murein peptide ligases has been the key feature to categorise the determined pMur ligase structures and hypothesise the roles of the enzyme types based on their structural similarities.

6.2.2.1 UDP-binding site in pseudomurein ligases

The pseudomurein and murein ligases have in common UDP as a chemical group as part of the substrates, which has been determined in complex with the pMurE of *M. fervidus*. This has allowed a detailed comparative analysis to that of the bacterial murein ligase substrate-binding site. A conserved UDP-binding site for all the murein ligase types is not observed and could be explained by the existence of the different N-terminal domain types as discussed in the previous section. Moreover, each murein ligase type possesses substantial differences in the shape and chemical environment of the UDP-binding sites in their structures (Mol*, et al.,* 2003). This further indicates that the UDP-binding site is variable

among the murein ligases and this appears to be applicable for the pseudomurein ligases as well. Among the two different types of pseudomurein ligases determined, pMurC possesses the N-terminal domain similar to MurC/D and pMurE similar to MurE/F type. It has been observed that the pMurE binds the UDP molecule solely at the N-terminal domain as exhibited by the Mfer762-UDP_2 structure and infers the putative catalytic active site based on the position at which it binds compared to bacterial MurE and MurF (Yan, *et al.*, 2000). The residues of the UDP-binding site are not conserved among the structural homologue, MurE, but suggest that the pMurE structure will adopt a similar binding cavity as seen for the Mfer762-UDP_2 structure due to the presence of conserved residues within its class. In addition to the binding site residues, a key residue, Arg40, is observed as an invariant residue among all the pMurEs and appears to contribute to the conformation of the UDP at the catalytic binding site. The UDP-binding position of Mfer762-UDP_2 closely resembles that seen in *A. baumannii* MurF (4QDI). Hence, the presence of a relatively similar N-terminal domain and the binding position for the UDP molecule in pMurE suggests that the putative substrate-binding site in pMurE could be similar to that of MurE.

The Mfer336 structure has been determined in the apo form, which limits the structural details regarding the putative substrate-binding site or a UDP molecule. However, the structural similarity of the N-terminal domain with its respective bacterial homologues and the conserved UDP recognition residues in the N-terminal domain of the Mfer336 structure compared to the bacterial MurC and MurD (Section 4.5.3) helps to support the hypothesis that the substrate binding mode in pMurC could be similar to that of MurC.

When examining the position at which the substrate for the murein ligases binds, it has been shown that the UDP head binds to the N-terminus, and the sugar moiety extends through the crescent shaped cavity towards the C-terminal domain where the L-amino acids would be

added by ligase activity (Smith, 2006). The substrates for pseudomurein ligases do not contain the sugar moiety, however, they do contain a glutamate residue to which the L-amino acids get ligated (Hartman, *et al.*, 1990). Hence, the pseudomurein peptide ligase substrate binding site needs to accommodate a substrate with a relatively similar length, but with a comparatively more linear configuration (Figure 3.37).

6.2.3 Conserved middle and C-terminal domains

The pseudomurein ligases contain an ATP-binding middle domain that shows structural similarity with the middle domain of all available murein ligase structures. The murein ligases have a conserved P-loop and additional six invariant residues that are used to define Mur ligase as a separate enzyme family (Bouhss, *et al.*, 1997), which are structurally conserved in the recently determined pseudomurein peptide ligases. Moreover, it is observed that invariant residues present in the pseudomurein ligases contribute to form a similar ATP-binding site, which was predicted based on the homology-based analysis with available murein ligase structures (Figure 6.3). The proposed ATP-binding site in pseudomurein ligases is located at the interface of the middle and the C-terminal domains and is close to the conserved P-loop. This result resembles the ATP-binding site in murein ligases (Mol, *et al.*, 2003; Smith, 2006) and suggests that both the murein and pseudomurein ligases share a similar mechanism for ATP binding. Hence, the current structural data suggests that both pseudomurein and murein peptide ligases fall into the Mur ligase family.

The C-terminal domains of the pMurE structures have structural similarity to the C-terminal domains of the murein ligases, as well as the C-terminal domain of FPGS (Section 3.8). The absence of a substrate- and product-bound pMurE structure has been a limitation to identify the amino acid binding motif of pMurE. The sequence homology and structural

superposition also do not support an obvious sequence motif for L-lysine-binding in pMurE as identified for MurEs (Section 3.11). The C-terminal domain is seen to move 'to' and 'from' the middle domain to adopt closed and open conformations as exhibited by the two different Mth734 structures of the crystallographic asymmetric unit. Such a mechanism involving the movement of the N- and C-terminal domains might be expected for pMur ligase activity and has been well studied among the murein ligases (Sink, *et al.*, 2016). Flexibility of the C-terminal domain in Mth734 is mediated by the rigid body rotation of the domain and was calculated to be 91°, however, the protein fold adopted by the C-terminal domain is similar in both the pMurE structures and shares similarity with the C-terminal domains of murein ligases (MurC-F).

MurEs in murein biosynthesis possess variation in the conserved motif at the C-terminal domain that is associated with its ability to bind varying substrates enabling them to complete their respective reactions. For example, the conserved motif for *m*-DAP adding MurE was found to have a consensus DNPR motif, whereas that for L-lysine addition have a D(D,N)P(N,A) motif (Basavannacharya, *et al.*, 2010b; Basavannacharya *et al.*, 2010a). The major difference between the two motifs was attributed to the arginine residue, which was involved in the formation of hydrogen bonding interaction with the carboxyl group of *m*-DAP (Mengin-Lecreulx, *et al.*, 1999; Gordon, *et al.*, 2001). Based on the substrate specificity of the putative pMurE towards L-lysine, a D(D,N)P(N,A) sequence motif, or similar sequence motif, was expected among the pMurE but could not be observed suggesting that the amino acid binding motif for the pMurE is unique compared to that for the MurE.

Moreover, the C-terminal domain of Mfer336 also shows structural similarity, mostly with the C-terminal domains of murein ligases and folylpolyglutamate synthetases (FPGS)

(Section 4.6.1). Unlike pMurE, the C-terminal domain of the pMurC (Mfer336) structure shares the highest structural similarity with *Yersinia pestis* C092 FPGS (3NRS) followed by the C-terminal domain of available murein ligases. The conserved amino acid binding motif for the pMurC could not be structurally identified due to the absence of substrate- and product-bound structure of pMurC. The structure-based superposition was also not successful in identifying the conserved motif for presumed L-alanine binding in pMurC. However, based on the conserved middle and C-terminal domains it appears that pseudomurein peptide ligases, like murein ligases, have a conserved ancient ATP-binding middle domain and a FPGS-related C-terminal domain.

6.2.3.1 Conserved ATP-binding site

The ATP-binding site in murein ligases (MurC-F) is conserved at the interface of the middle and C-terminal domains and is contributed mostly by the residues in the P-loop and other invariant residues (Smith, 2006). It has been presumed that a similar binding position and similar interactions by the invariant residues are present in the pseudomurein ligases (Figure 6.3). In addition to the residues of the middle domain, two residues, Arg325 and Arg338 (Mfer762 numbering) from the hinge region and C-terminal domain, are also strictly conserved in both murein and pseudomurein ligases. The residue corresponding to Arg325 of the Mfer762 structure is involved in presumed phosphate and pyrophosphate interaction in pMurE and pMurC structures, respectively. The residues of the conserved P-loop are located close to the presumed ATP molecule such that they could form hydrogen bonds with the triphosphate of the ATP molecule as observed in bacterial murein ligases (Deva, *et al.*, 2006). This observation is even clearer from the pyrophosphate (POP)-binding in the Mfer336 structure. This shows that the conserved ATP-binding residues form a similar fold to form a catalytic active site at the interface of the middle and the C-terminal domains as

that observed in murein ligases. Hence, based on the presence of invariant binding site residues and a similar binding position for the proposed ATP molecule, it can be predicted that pseudomurein ligases might adopt a similar ATP-binding mechanism as studied for murein ligases.

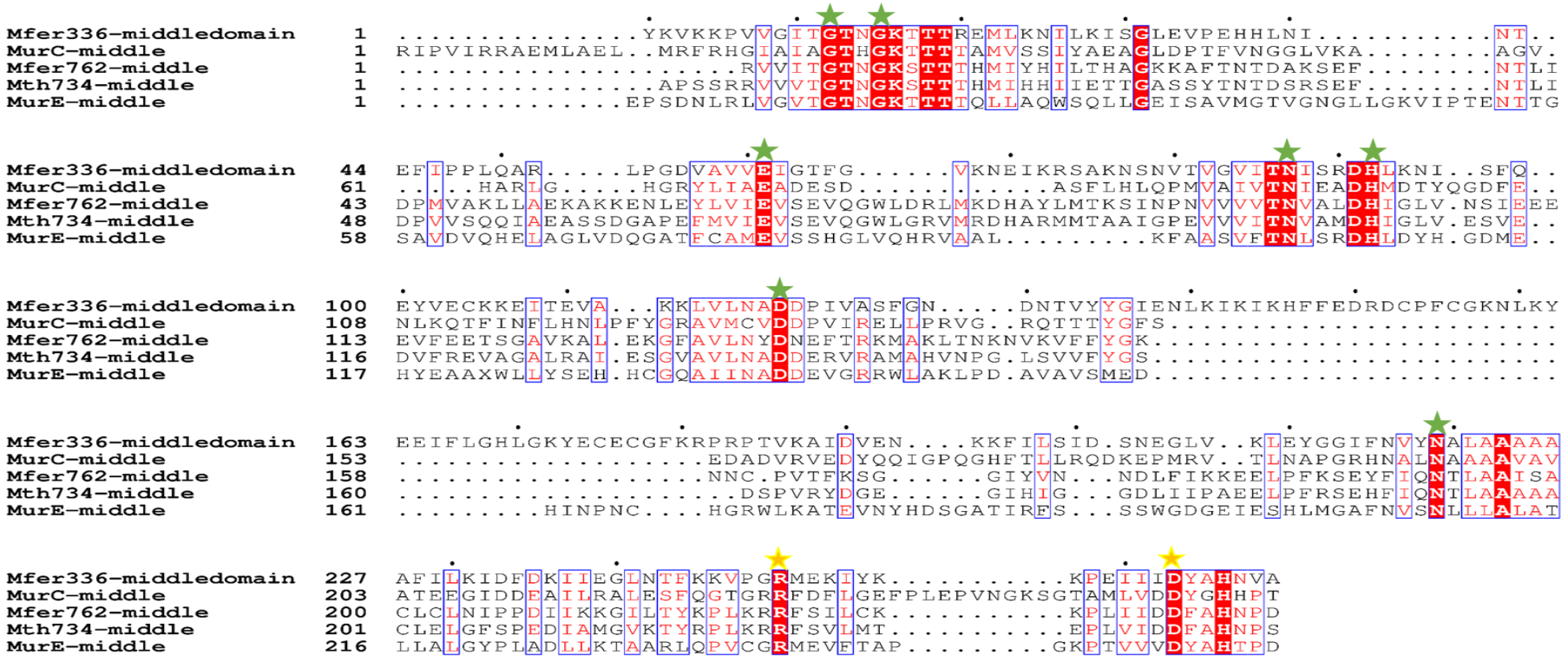


Figure 6.3 Structure-based sequence alignment of the middle domains of pMurC and pMurE and bacterial homologues. The middle domain of MurC and MurE from *E. coli* have been compared to recently determined pMurC and pMurE middle domains. The residues from the hinge region and a small portion of the C-terminal domains are also included in the alignment to show the last two invariant residues that putatively interact with the ATP molecule in pMur enzymes. The invariant residues conserved in the middle domain are represented by green stars and those from the hinge region and the C-terminal domain are represented by yellow stars.

6.2.3.2 Conserved zinc-binding site in pMurC

It has been observed that the Mfer336 structure has an additional small domain that contains a unique zinc-binding site formed in part by four conserved cysteine residues in the middle domain. The Cys residues are conserved among all pMurC peptide ligases in pseudomurein-containing methanogens including the extreme thermophile *M. kandleri*, which is suggested to be the most ancient of the methanogens based on 16s rRNA (Wong *et al.*, 2007). The additional small domain within the middle domain of the Mfer336 structure has a consensus CX_2CX_nCXC sequence motif where the conserved Cys residues form a tetrahedral binding site for a zinc ion. The presence of the zinc-binding site among all the available pMurC sequences and even within the most deeply rooted methanogen (*Methanopyrus kandleri*) in the evolutionary tree, suggests that this domain and Zn^{2+}-binding site in particular is a common characteristic of pMurC.

Even though the Zn^{2+}-binding site is reported for the first time and the function of the Zn^{2+} ion or the region binding the ion is not yet clearly defined, the sequence homology search indicated that a similar sequence motif is present in a few bacterial murein ligase homologues and DUF1727 domain-containing proteins. These proteins show homology with MurE in bacteria and pMurT in pseudomurein-containing methanogens. MurT in the murein biosynthesis pathway catalyses the amidation of the α-carboxyl group of D-Glu and functions in complex with GatD (Munch, *et al.*, 2012). Interestingly pMurT are not present in all the pseudomurein-containing methanogens, yet could potentially be helpful in elucidating the shared evolutionary history of the murein and pseudomurein biosynthesis pathways.

MurT and its homologues in pseudomurein-containing methanogens have shorter amino acid sequences compared to those of murein or pseudomurein ligases and lack the N-terminal

domain equivalent of peptide ligases as observed with the FPGS. The structure of the first MurT has recently been published and a conserved sequence motif, CX_2CX_nCXC, appears to form a similar structural fold with the conserved cysteines making a similar tetrahedral arrangement to that seen in pMurC peptide ligase (Morlot *et al.*, 2018). However, the MurT structure did not contain a zinc ion in the binding site and also the structure includes an additional β-strand compared to the Mfer336 structure, yet suggests that the region might be involved in protein-complex formation with GatD. Hence, this observation further supports our assumption that the conserved cysteine residues act to create a structural zinc binding site leading to speculation of a role in protein complex formation (Section 4.8).

6.3 Proposed peptide ligase mechanism in pMur

The conserved three domain architecture of the pseudomurein peptide ligases determined during this project suggests that the murein and pseudomurein peptide ligases might both share a similar substrate-binding site and a common ligation mechanism. The crescent-shaped cavity formed at the interface of the three domains is used to bind the growing peptide in murein ligases and is observed to be similar in the pseudomurein ligases. The binding of the growing peptide by consecutive murein ligases in the step-wise addition of amino acids supports a gradually wider opening of the crescent-shaped cavity in the murein ligase structures from MurC to MurF (El Zoeiby, *et al.*, 2003). The opening and closing of the ligase structures are favoured by the flexibility of the hinge regions present between each domain (Smith, 2006). The comparative structure analysis suggest that the key features of the murein peptide ligases are conserved in pseudomurein ligases.

Even though the complete substrate bound structure could not be determined for any pseudomurein peptide ligases, structural comparison analysis have been used to propose that

pseudomurein ligases could adopt a similar mechanism as adopted by murein ligases for the ATP molecule binding and phosphorylation of the amino acid carboxylate (Bouhss, *et al.*, 2002). Due to the availability of the pMurE structure in complex with UDP, the pMurE structural detail can be used as a reference to compare it with MurE catalysis steps. In the murein biosynthesis pathway, MurE catalyses the ligation of *m*-DAP or L-lysine to UMAG through four key steps: (i) bringing the UMAG and ATP together; (ii) orientating UMAG and ATP accurately such that the acyl-phosphate intermediate can be formed; (iii) orienting the upcoming amino acid for nucleophilic attack; and (iv) stabilizing the tetrahedral intermediate, thereby lowering the activation barrier and accelerating catalysis (Gordon, *et al.*, 2001).

During this study, it has been predicted that the binding of the ATP molecule could be at approximately the similar position for both the pMurC and pMurE in respect to their bacterial counterparts (Section 3.7.1.2 and Section 4.5.1). Moreover, the structural details from the pMurE structure (Mfer762-UDP_2) provides clear information about the catalytic UDP-binding site and infers the putative N^{α}-UDP-Glu^{γ}-Ala-binding site which could be an extension of the determined UDP-binding site. The information derived from the pMurE structures could be used to propose a similar aforementioned four-step ligation mechanism as explained for the MurE to occur in the L-lysine ligation during the pseudomurein biosynthesis pathway.

In addition, the ConSurf webserver that maps the amino acid position in the protein structure based on evolutionary conservation was used to analyse the pMurE structure based on archaeal sequences (Landau *et al.*, 2005). It was observed that the crescent shaped cavity in the pMurE was evolutionary conserved and so is the UDP-binding site at the exterior surface as studied for the Mfer762-UDP_1 structure (Section 1.1.1.1.1). The proposed binding site

for the putative substrate (N^{α}-UDP-Glu$^{\gamma}$-Ala) and the ATP appear to be evolutionarily conserved (Figure 6.4). The conservation of the pMurE residues from the N-terminal domain all the way through the middle domain to the C-terminal domain at the crescent-shaped cavity suggests that a similar binding mode is present for N^{α}-UDP-Glu$^{\gamma}$-Ala and ATP as seen in bacteria. For clarity purposes, the bacterial MurE substrate, UAG, has been modelled in the region using structure homology with MurE structure from *M. tuberculosis*. The observation supports the proposition made for substrate-binding in the pMurE structure and that could extend from the UDP_2 (represented as yellow) in Figure 6.4.

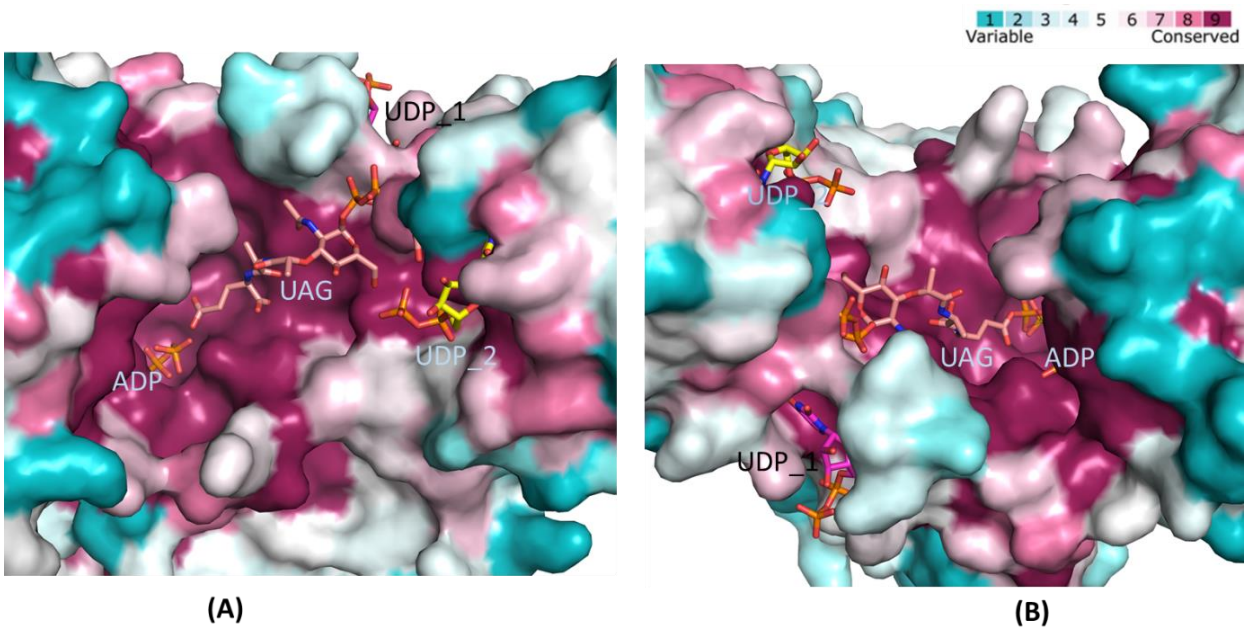


Figure 6.4 ConSurf analysis to show evolutionarily conserved residues of the Mfer762 structure. (A) ConSurf analysis result along with the UMAG and ADP from MurE of *M. tuberculosis* modelled using structure superposition to depict the binding-site in pMurE. (B) Top view of the figure in (A) for clarity and to show the evolutionarily conserved UDP_1-binding site (magenta). The colour used

in the figure refers to conservation as calculated by the ConSurf tool (Landau et al., 2005) and is shown in the top right corner.

Furthermore, Mfer336 is the only structure determined for the pMurC type for which a limited amount of structural detail could be depicted, yet demonstrates the conserved three domain architecture, similar presumed ATP-binding site and structurally conserved substrate recognition residues at the N-terminal domain indicating that pMurC is homologous to MurC. In addition, a proposition for the ligation of L-alanine to the N^{α}-UDP-Glu could be derived for the pMurC based on the structure comparison with bacterial MurC. The putative ATP-binding site for the Mfer336 has been proposed based on the structural homology with the bacterial MurC (Section 4.5.1). The structural detail on the UDP molecule or the substrate–binding site (N^{α}-UDP-Glu) could not be determined for the pMurC, however, the structural similarity search and homology with the bacterial MurC/D N-terminal domain infers a similar substrate-binding site for the pMurC (Section 4.5.3). The N-terminal domain comparison of the Mfer336 structure with the MurC and MurD from *H. influenzae* shows conserved UDP-binding residues for Mfer336 as that studied for MurC (Mol, *et al.*, 2003). Hence, based on the proposed ATP- and UDP- binding site in the Mfer336 structure, it could be predicted that the amino acid ligation mechanism for the pMurC could be similar to that in MurC. This proposition is also supported by the information of the conserved residues in the Mfer336 structure as determined by the ConSurf analysis (Figure 6.5). The residues at the crescent-shaped cavity of the Mfer336 structure are observed to be evolutionarily conserved and supports the hypothesis made for substrate-binding in pMurC (Section 4.5.3, Figure 4.15).

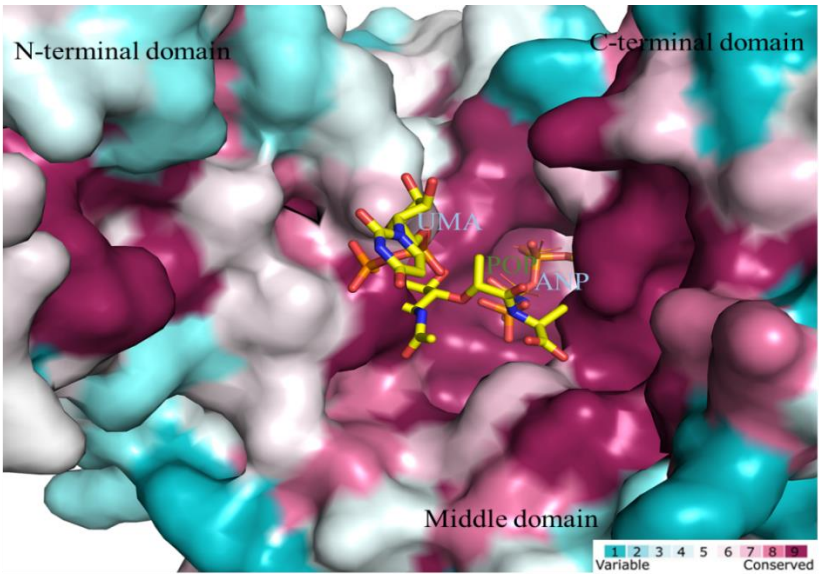


Figure 6.5 ConSurf analysis to show evolutionarily conserved residues of the Mfer336 structure. The bacterial MurC substrate, UMA and ANP (represented as sticks), from the *H. influenzae* MurC (1P3D) structure have been modelled using structure superposition to depict the binding-site in pMurC. The pyrophosphate (POP) bound to the Mfer336 structure is shown as lines and possesses a position similar to the α- and β-phosphate positions of ANP. The colour used in the figure refers to conservation as calculated by the ConSurf tool (Landau et al., 2005) and is shown in the bottom right corner.

The biochemical assays could have been a determinant for identification of the pseudomurein ligase towards specific amino acids and their ligation mechanism and was also an objective of the study. However, despite dedicated attempts, both biochemically using putative enzymes *in vitro* (Sauter, 2017) and synthesis through commercial companies, the initial substrate (N^{α}-UDP-Glu) to which the L-Ala gets ligated could not be obtained, hence, limiting the biochemical testing of the putative pseudomurein ligases. The biochemical assays to test the bacterial substrate (MurNAc-Ala) on pMurE (Mfer762) were also performed to see if the enzymes possessed activity towards the bacterial substrate. The experiment showed no activity of the pMurE towards the bacterial substrate despite being of

a similar length, with a similar proximal UDP moiety and a terminal Ala residue, indicating a requirement for a specific pseudomurein-related substrate for activity, thereby indicating a distinct difference in the evolutionary relationship of the enzyme types catalysing murein and pseudomurein biosynthesis.

6.4 Phylogenetic analysis

Phylogenetic analyses were aimed at understanding the molecular evolution of pseudomurein and murein peptide ligases. The analysis was expected to help elucidate the evolutionary history of the two cell wall biosynthesis pathways. The phylogenetic trees based on the sequence alignment of each ligase type have been analysed within each chapter and it was observed that both the murein and pseudomurein peptide ligases clade separately into different groups. All the four murein ligases and five putative pseudomurein ligases, including the pMurF, have been analysed to see if any of the ligase type clades among or close to one another. It was observed that each of the pseudomurein ligase clade separately in the phylogenetic tree with strong bootstrap support (Figure 6.6). The pMurF was included in the analysis with the assumption that being a homologous enzyme of MurF it could provide insight to the evolutionary history of the peptide ligases in the two cell wall biosynthetic pathways. It was observed that the pMurF and its bacterial homologues also clade separately in the phylogenetic tree, supported by the strong bootstrap values (Figure 6.7). Peptide ligase sequences from *Methanopyrus kandleri* AV19 are deeply rooted among all pseudomurein ligase clades. However, despite several attempts using different combinations of murein and pseudomurein peptide ligase sequences Appendix 6 none of the phylogenetic analyses were helpful in elucidating their evolutionary relationships due to limited bootstrap support (<40). Similarly, phylogenetic analyses using both Maximum Likelihood and Neighbour-joining methods were tested but both methods could only support

the separate clades of each peptide ligase group and no branching patterns that could be taken as supporting an evolutionary association was found. Hence, only the molecular phylogenetic trees obtained using the Maximum Likelihood method have been used here and only the bootstrap values supporting their clades have been included.

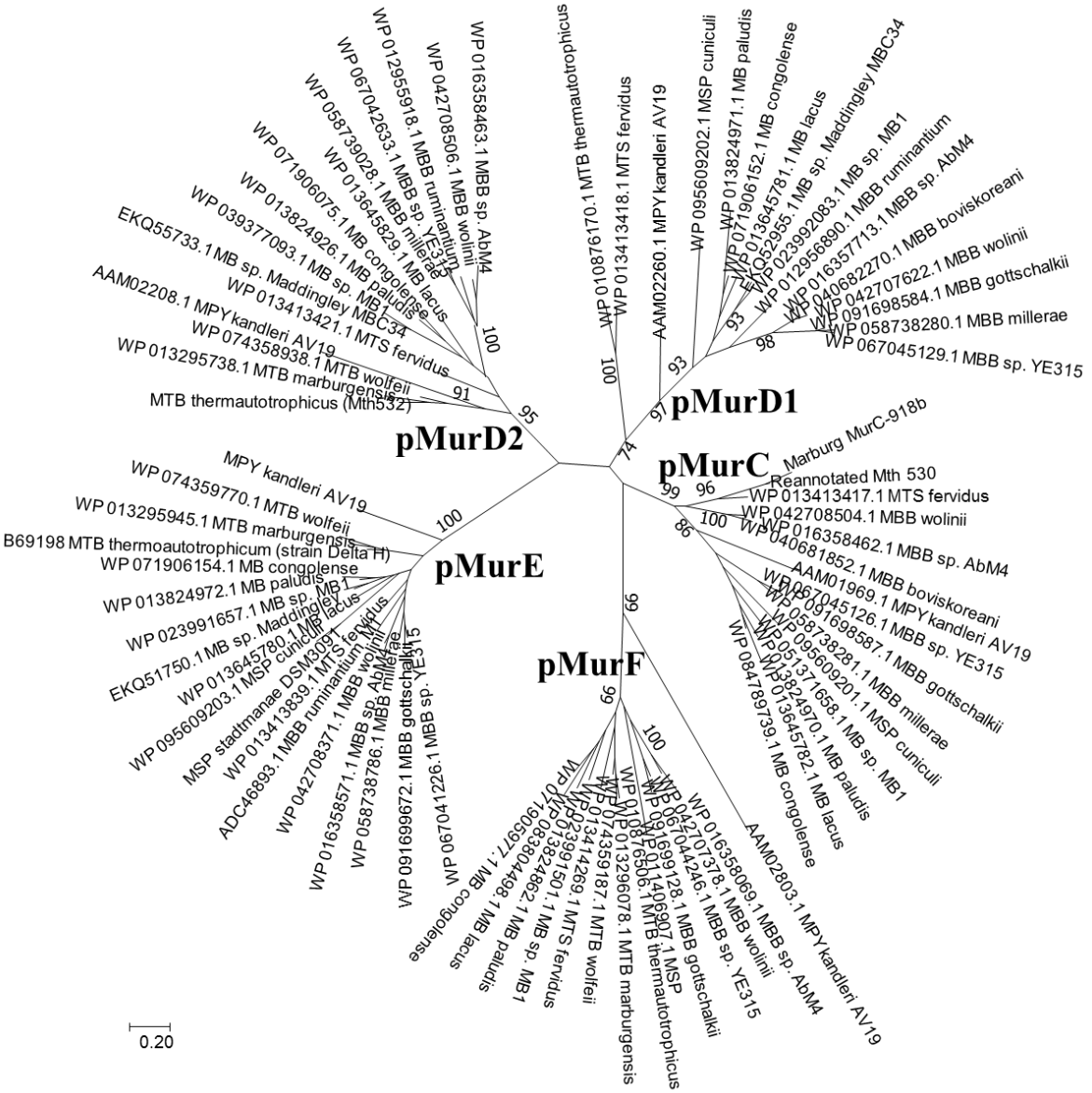


Figure 6.6 Molecular phylogenetic analysis of pseudomurein peptide ligases and pMurF by the Maximum Likelihood method using representative pseudomurein-containing methanogens

sequences (Kumar, *et al.*, 2016). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 80 amino acid sequences.

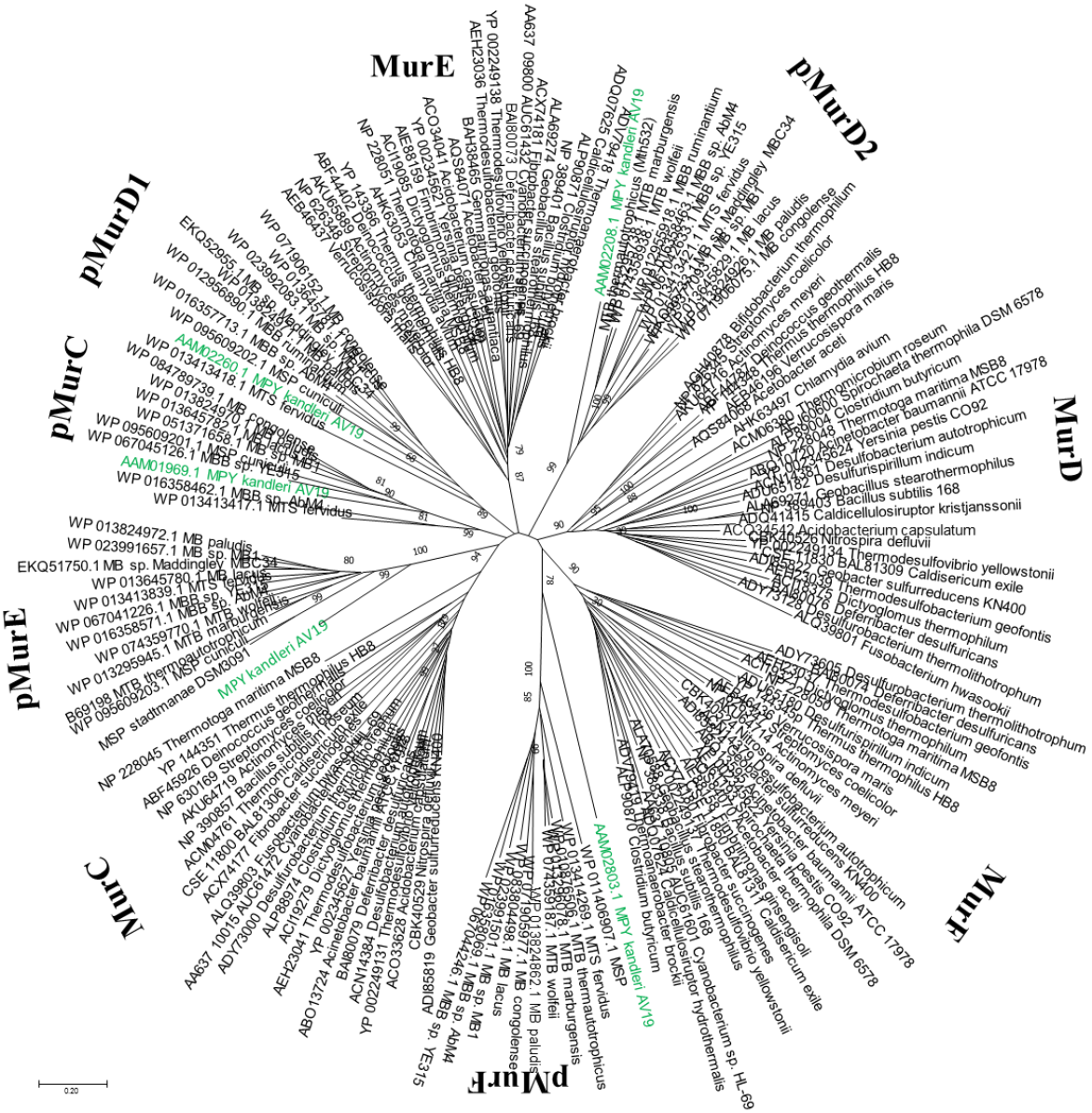


Figure 6.7 Molecular phylogenetic analysis of murein peptide ligases and their homologues in pseudomurein-containing methanogens by Maximum Likelihood method using representative sequences (Kumar, *et al.*, 2016). *M. kandleri* AV19 sequences are deep rooting in all pseudomurein peptide ligase clades and are highlighted in green.

Furthermore, structure-based phylogenetic trees were also constructed based on available murein ligase structures and recently determined pseudomurein peptide ligase structures. Due to significantly low bootstrap values, the phylogenetic relationship between the murein and pseudomurein peptide ligases could not be inferred (Supp. Figure 1). Interestingly, based on the structure-guided information of the pseudomurein and murein peptide ligases available, the phylogenetic tree built to analyse the evolution of the pMurT, pMurC, pMurF and FPGS suggested that the pMurC and pMurT are closely related which was supported by relatively high bootstrap values (Figure 6.8). This is the only phylogenetic tree that showed an evolutionary relationship between the pMurC and its homologues. The branch lengths of the pMurC, FPGS and pMurT sequences suggest that the pMurT could be late evolving enzyme even though they are more closely related (sharing a common node), and FPGS could be earlier evolving enzyme compared to pMurC. The pMurF clade seems to be independently evolving of the four types of homologues analysed and is indicated by the separate clade from the point of origin. The attempt to extend the analysis within the murein peptide ligases resulted in low bootstrap values (data not shown).

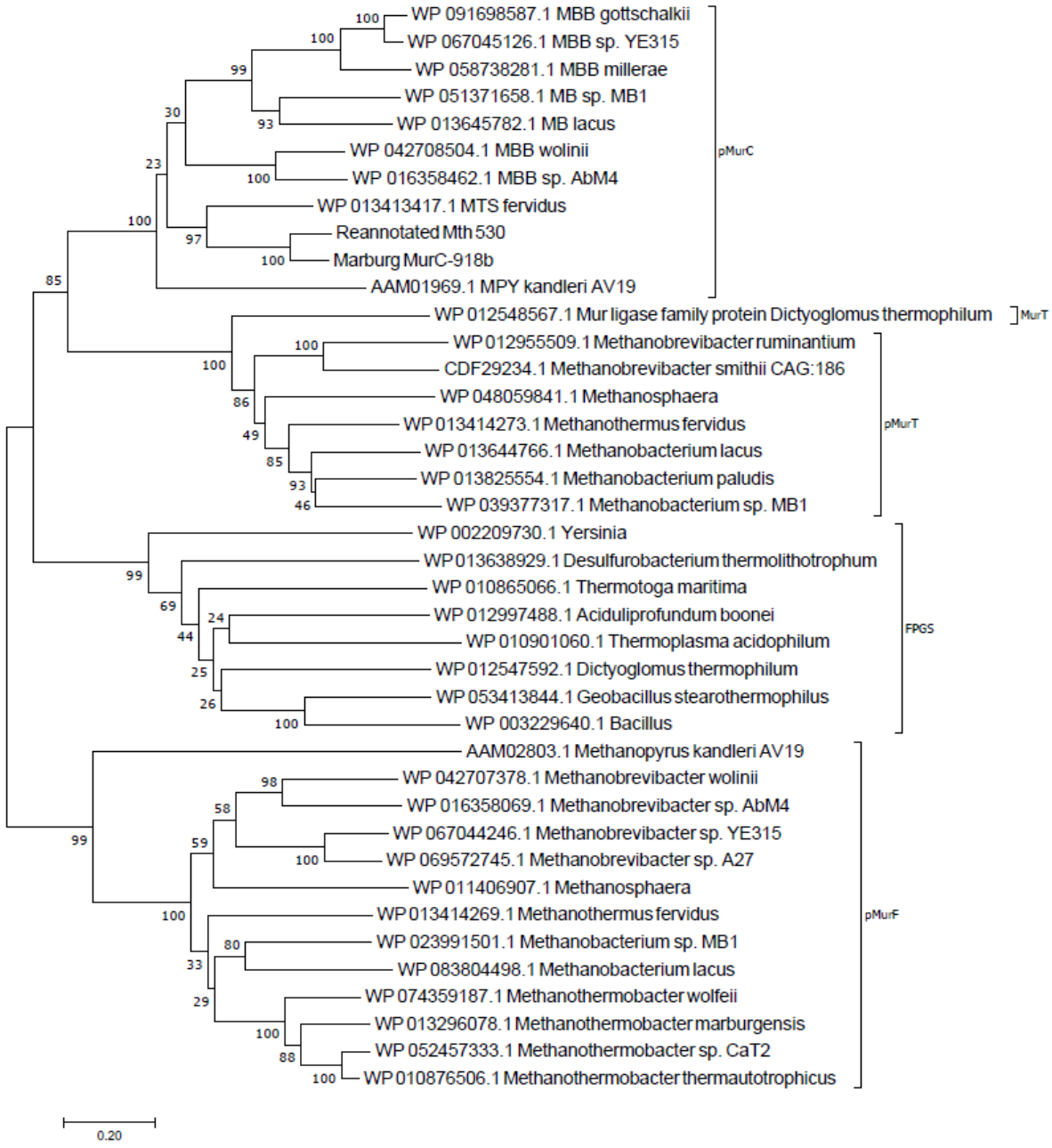


Figure 6.8 Molecular phylogenetic analysis of pMurC, pMurF, pMurT and FPGS by Maximum Likelihood method using representative sequences. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Kumar, *et al.*, 2016)

6.5 Proposition for peptide ligase evolution

A potential scenario depicting the evolutionary pathway for the bacterial and pseudomurein-containing archaeal cell wall peptide ligases is summarised in Figure 6.9 and is one of the

scenarios that is supported by the results of this study. Based on the structural information available for the murein ligases and the observations derived from the recently determined pseudomurein ligases, it is clear that the peptide ligases in both cell wall biosynthesis pathways possess a highly conserved ancient ATPase-like middle domain and an amino acid binding DHFR-like C-terminal domain (Smith, 2006). During an early phase of evolution, these two domains could have fused together to form an early FPGS (pre-FPGS) that initially would not have contained an additional amino acid insertion region in the N-terminus. The FPGS could have acquired the insertion region at the N-terminal domain to adopt the substrate-binding mechanism required by the enzyme. The current study shows that with exception to the insertion region, FPGS shares similar middle and C-terminal domains compared to pseudomurein ligases but lacks the typical cell wall peptide ligase N-terminal domain. This leads to the possibility that pre-FPGS in bacteria could have acquired either of the two structurally different N-terminal domains (C/D and E/F) types to form separate MurC/D and MurE/F ligases which clearly clade separately in the phylogenetic tree (Figure 6.7).

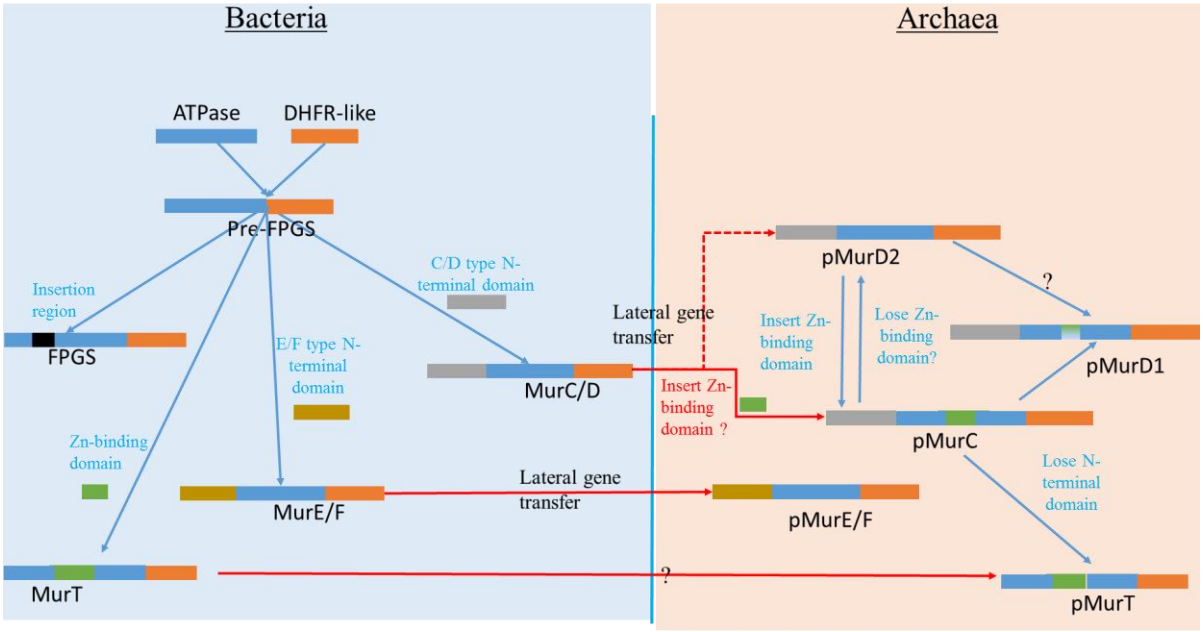


Figure 6.9 Schematic diagram depicting a potential scenario for the evolution of peptide ligases in bacteria and archaea. The evolution of murein ligases in bacteria is highlighted in the blue box and pseudomurein ligases in archaea in the orange box. The solid arrows in blue represent the most parsimonious steps. The route of lateral gene transfer for pMurE and pMurC are shown with solid red arrows and that for pMurD2 with a broken red arrow. For brevity the representation of the murein peptide ligases have been kept as two types based on their structure differences (MurC/D and MurE/F).

The evolution of pseudomurein peptide ligases is proposed to partially be a result of lateral gene transfers of full-length murein peptide ligases from bacteria. The structures obtained for pMurE and pMurC and their structural homology with the respective murein peptide ligases suggest that pMur peptide ligases could have come from ancestral genes from bacteria. Moreover, the identification of both the C/D and E/F types of N-terminal domains in pseudomurein peptide ligases suggest that lateral gene transfer of two murein peptide ligases with each type of N-terminal domain could have occurred.

Taken together, the structural similarity and putative peptide ligation step proposed for pMurE peptide ligase suggests that the MurE peptide ligase could have led to pMurE in pseudomurein-containing archaea. Even though the sequence analysis suggests homology between pMurF and MurF, the pMurF is identified as F_{430} synthase, which suggests that pMurF could have resulted from modification of the pMurE to bind a different substrate. Further, based on the fact that pMurF is widely present among other methanogens, it could be possible that pMurF could have been an early evolving among the pMurE/F group and pMurE could have been a result of modification from pMurF. However, this scenario is limited by a lack of structural information for the pMurF.

Similarly, lateral gene transfer of the MurC peptide ligase could have resulted in the pMurC peptide ligases (indicated by red elbow arrow Figure 6.9). However, the structural difference in the middle domain due to the presence of a cysteine-rich insertion forming a zinc-binding domain in pMurC suggests that it would require an additional insertion of the zinc-binding domain. Hence, the most parsimonious scenario would be the lateral gene transfer of MurD to form pMurD2. Further, pMurD2 could have led to pMurC though the insertion of a zinc-binding domain. The homology-based models of pMurD2 and sequence analysis show that pMurD2 is similar to pMurC except for a cysteine-rich insertion region. This supports the likely lateral gene transfer of MurD rather than MurC in pseudomurein-containing methanogens. Furthermore, the evolution of the pMurD1 could have resulted through a modification to pMurC, where the activity of the cysteine-rich zinc-binding domain could have been lost. The pMurD1 peptide ligase contains the cysteine-rich insertion region but the cysteines are not conserved.

Finally, the MurT homologues in *Archaea*, that are suggested to be late evolving from the phylogenetic analysis (Figure 6.8), could have resulted via loss of the N-terminal domain of

pMurC. Alternatively, pMurT could have evolved through lateral gene transfer from bacteria where, MurT could have been formed from pre-FPGS by acquiring a zinc-binding domain (Figure 6.9).

Hence, the project has been successful in proposing an evolutionary pathway for the murein and pseudomurein peptide ligases involved in cross-linking pentapeptide formation in murein and pseudomurein biosynthesis pathway.

Ideas for future research

This project has been successful in determining the structure of two different pseudomurein ligase types, pMurC and pMurE, and in developing a proposal for the ATP-binding site in both ligase types. The study has also successfully established the structural homology of pMurC and pMurE with MurC and MurE, respectively. The conserved three domain arrangement in pseudomurein ligases and the flexibility of the N- and C-terminal domains to rotate around the conserved middle domain have been analysed and suggested to be a common characteristics of all available murein and pseudomurein ligases. Moreover, the study provides structural details to support the presence of a similar ligation mechanism by both the murein and pseudomurein peptide ligases.

However, due to the time limitation and absence of good diffracting crystals, the structure of two putative pseudomurein ligases, pMurD1 and pMurD2, could not be determined. In addition, the structure determination of pMurF would have been useful in understanding the key differences in the two enzymes (MurF and pMurF) that were predicted to be homologous based on amino acid sequence information. Most importantly, the challenges with the production of the initial substrate for the pseudomurein peptide ligases and the lack of

suitable substrates to test biochemical activity using the pseudomurein ligases also limited the study.

In addition to the aforementioned primary objectives that are yet to be completed, the following further work could aid in the understanding of the pseudomurein biosynthesis pathway:

1. Obtaining structures with the archaeal substrates and amino acids would be useful in determining the binding mechanism, thereby enhancing our understanding about the ligation mechanism of pseudomurein ligases.
2. Additional work will be required to identify the enzymes responsible for producing the initial substrate for the pseudomurein ligase (N^{α}-UDP-Glu$^{\gamma}$-P). This would allow testing biochemical activity using the available ligases to identify the steps catalysed by the enzymes in the pseudomurein biosynthesis pathway.
3. Mutagenesis of the conserved 'extended loop' region of the pMurE structure could be informative in aiding information regarding its role in proposed structural stability and if the closed 'gateway-like' conformation was required to adopt the 'open' conformation of the pMurE as observed in all the Mfer762 structures as well as Mth734_A, whereas, Mth734_B with movement of the 'extended loop' adopts a 'closed' conformation. In addition, the mutation of the region would also be useful in understanding if the region could play a role in regulating the ligation reaction.
4. The structure of a homologous murein ligase (MurT) that shares a similar zinc-binding motif with the middle domain of the pMurC structure has been studied. The corresponding region in bacterial MurT is suggested to be involved in protein complex formation. The presence of the Zn^{2+}-binding site and its putative role for complex formation in the pseudomurein ligase needs to be investigated. Gel filtration could be

used to test complex formation by pMurC with other pseudomurein peptide ligases and especially pMurD1 as both of these genes are found in a gene cluster and are also annotated as a single gene in the recently available *Methanothermobacter* species.

5. The putative role of the zinc-binding domain-like structure in the pMurC structure to form a protein-protein complex and the presence of Mth529-like unknown protein only in the pseudomurein-containing methanogens arise a new hypothesis for the two proteins form the complex. In addition, the gene cluster analysis suggest that both the genes encoding these proteins are present adjacent to one another in the genome.

Chapter 7 References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-21.

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 4), 352-67.

Albers, S., Eichler, J., & Aebi, M. (2015). Archaea. In rd, A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar & P. H. Seeberger (Eds.), *Essentials of Glycobiology* (pp. 283-92). Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press

Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved.

Albers, S. V., & Meyer, B. H. (2011). The archaeal cell envelope. *Nat Rev Microbiol, 9*(6), 414-26.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*(17), 3389-402.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics, 22*(2), 195-201.

Auld, D. S. (2001). Zinc coordination sphere in biochemical zinc sites. *Biometals, 14*(3-4), 271-313.

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., & Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol Rev, 32*(2), 168-207.

Basavannacharya, C., Robertson, G., Munshi, T., Keep, N. H., & Bhakta, S. (2010a). ATP-dependent MurE ligase in *Mycobacterium tuberculosis*: biochemical and structural characterisation. *Tuberculosis (Edinb), 90*(1), 16-24.

Basavannacharya, C., Moody, P. R., Munshi, T., Cronin, N., Keep, N. H., & Bhakta, S. (2010b). Essential residues for the enzyme activity of ATP-dependent MurE ligase from *Mycobacterium tuberculosis*. *Protein Cell, 1*(11), 1011-22.

Benkert, P., Tosatto, S. C., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins, 71*(1), 261-77.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res, 28*(1), 235-42.

Bertrand, J. A., Auger, G., Fanchon, E., Martin, L., Blanot, D., van Heijenoort, J., & Dideberg, O. (1997). Crystal structure of UDP-N-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *EMBO J, 16*(12), 3416-25.

Bertrand, J. A., Auger, G., Martin, L., Fanchon, E., Blanot, D., Le Beller, D., van Heijenoort, J., & Dideberg, O. (1999). Determination of the MurD mechanism through crystallographic analysis of enzyme complexes. *J Mol Biol, 289*(3), 579-90.

Boniface, A., Bouhss, A., Mengin-Lecreulx, D., & Blanot, D. (2006). The MurE synthetase from *Thermotoga maritima* is endowed with an unusual D-lysine adding activity. *J Biol Chem, 281*(23), 15680-86.

Borrel, G., O'Toole, P. W., Harris, H. M., Peyret, P., Brugere, J. F., & Gribaldo, S. (2013). Phylogenomic data support a seventh order of Methylotrophic methanogens and provide insights into the evolution of methanogenesis. *Genome Biol Evol, 5*(10), 1769-80.

Bouhss, A., Crouvoisier, M., Blanot, D., & Mengin-Lecreulx, D. (2004). Purification and characterization of the bacterial MraY translocase catalyzing the first membrane step of peptidoglycan biosynthesis. *J Biol Chem, 279*(29), 29974-80.

Bouhss, A., Trunkfield, A. E., Bugg, T. D., & Mengin-Lecreulx, D. (2008). The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol Rev, 32*(2), 208-33.

Bouhss, A., Mengin-Lecreulx, D., Blanot, D., van Heijenoort, J., & Parquet, C. (1997). Invariant amino acids in the Mur peptide synthetases of bacterial peptidoglycan synthesis and their modification by site-directed mutagenesis in the UDP-MurNAc:L-alanine ligase from *Escherichia coli*. *Biochemistry, 36*(39), 11556-63.

Bouhss, A., Dementin, S., van Heijenoort, J., Parquet, C., & Blanot, D. (2002). MurC and MurD synthetases of peptidoglycan biosynthesis: borohydride trapping of acyl-phosphate intermediates. *Methods Enzymol, 354*, 189-96.

Bunkoczi, G., Echols, N., McCoy, A. J., Oeffner, R. D., Adams, P. D., & Read, R. J. (2013). Phaser.MRage: automated molecular replacement. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 11), 2276-86.

Caforio, A., & Driessen, A. J. M. (2017). Archaeal phospholipids: Structural properties and biosynthesis. *Biochim Biophys Acta Mol Cell Biol Lipids, 1862*(11), 1325-39.

Candela, T., & Fouet, A. (2006). Poly-gamma-glutamate in bacteria. *Mol Microbiol, 60*(5), 1091-98.

Carbone, V., Schofield, L. R., Zhang, Y., Sang, C., Dey, D., Hannus, I. M., Martin, W. F., Sutherland-Smith, A. J., & Ronimus, R. S. (2015). Structure and evolution of the archaeal lipid synthesis enzyme sn-glycerol-1-phosphate dehydrogenase. *J Biol Chem, 290*(35), 21690-704.

Casanas, A., Warshamanage, R., Finke, A. D., Panepucci, E., Olieric, V., Noll, A., Tampe, R., Brandstetter, S., Forster, A., Mueller, M., Schulze-Briese, C., Bunk, O., & Wang, M. (2016). EIGER detector: application in macromolecular crystallography. *Acta Crystallogr D Struct Biol, 72*(Pt 9), 1036-48.

Cavicchioli, R. (2011). Archaea-timeline of the third domain. *Nat Rev Microbiol, 9*(1), 51-61.

Cha, S. S., An, Y. J., Jeong, C. S., Yu, J. H., & Chung, K. M. (2014). ATP-binding mode including a carbamoylated lysine and two Mg(2+) ions, and substrate-binding mode in *Acinetobacter baumannii* MurF. *Biochem Biophys Res Commun, 450*(2), 1045-50.

Chung, B. C., Zhao, J., Gillespie, R. A., Kwon, D. Y., Guan, Z., Hong, J., Zhou, P., & Lee, S. Y. (2013). Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science, 341*(6149), 1012-16.

Claus, H., & König, H. (2010). Cell envelopes of methanogens. In H. König, H. Claus & A. Varma (Eds.), *Prokaryotic Cell Wall Compounds* (pp. 231-51): Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-05062-6_7.

Cohen, G. N. (2014). Methanogens and Methylotrophs. In *Microbial Biochemistry* (pp. 153-77). The Netherlands: Springer.

Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr, 50*(Pt 5), 760-63.

Consaul, S. A., Wright, L. F., Mahapatra, S., Crick, D. C., & Pavelka, M. S., Jr. (2005). An unusual mutation results in the replacement of diaminopimelate with lanthionine in the peptidoglycan of a mutant strain of *Mycobacterium smegmatis*. *J Bacteriol, 187*(5), 1612-20.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 9), 1002-11.

Database resources of the National Center for Biotechnology Information. (2015). *Nucleic Acids Res, 43*(Database issue), 6-17.

Davis, I. W., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res, 32*(Web Server issue), 615-9.

Dementin, S., Bouhss, A., Auger, G., Parquet, C., Mengin-Lecreulx, D., Dideberg, O., van Heijenoort, J., & Blanot, D. (2001). Evidence of a functional requirement for a carbamoylated lysine residue in MurD, MurE and MurF synthetases as established by chemical rescue experiments. *Eur J Biochem, 268*(22), 5800-7.

Deva, T., Baker, E. N., Squire, C. J., & Smith, C. A. (2006). Structure of *Escherichia coli* UDP-N-acetylmuramoyl:L-alanine ligase (MurC). *Acta Crystallogr D Biol Crystallogr, 62*(Pt 12), 1466-74.

Doublet, P., van Heijenoort, J., Bohin, J. P., & Mengin-Lecreulx, D. (1993). The murI gene of *Escherichia coli* is an essential gene that encodes a glutamate racemase activity. *J Bacteriol, 175*(10), 2970-79.

Draycheva, A., Bornemann, T., Ryazanov, S., Lakomek, N. A., & Wintermeyer, W. (2016). The bacterial SRP receptor, FtsY, is activated on binding to the translocon. *Mol Microbiol, 102*(1), 152-67.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics, 5*, 113.

El Zoeiby, A., Sanschagrin, F., & Levesque, R. C. (2003). Structure and function of the Mur enzymes: development of novel inhibitors. *Mol Microbiol, 47*(1), 1-12.

Emanuele, J. J., Jr., Jin, H., Yanchunas, J., Jr., & Villafranca, J. J. (1997). Evaluation of the kinetic mechanism of *Escherichia coli* uridine diphosphate-N-acetylmuramate:L-alanine ligase. *Biochemistry, 36*(23), 7264-71.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501.

Enzmann, F., Mayer, F., Rother, M., & Holtmann, D. (2018). Methanogens: biochemical background and biotechnological applications. *AMB Express, 8*(1), 1.

Evans, P. R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 282-92.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr, 69*(Pt 7), 1204-14.

Eveland, S. S., Pompliano, D. L., & Anderson, M. S. (1997). Conditionally lethal *Escherichia coli* murein mutants contain point defects that map to regions conserved among murein and folyl poly-gamma-glutamate ligases: identification of a ligase superfamily. *Biochemistry, 36*(20), 6223-29.

Favini-Stabile, S., Contreras-Martel, C., Thielens, N., & Dessen, A. (2013). MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis. *Environ Microbiol, 15*(12), 3218-28.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution, 39*(4), 783-91.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., & Jensen, L. J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res, 41*(Database issue), 808-15.

Garcia, J. L., Patel, B. K., & Ollivier, B. (2000). Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea. *Anaerobe, 6*(4), 205-26.

Gordon, E., Flouret, B., Chantalat, L., van Heijenoort, J., Mengin-Lecreulx, D., & Dideberg, O. (2001). Crystal structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli*. *J Biol Chem, 276*(14), 10999-1006.

Gorrec, F. (2009). The MORPHEUS protein crystallization screen. *J Appl Crystallogr, 42*(Pt 6), 1035-42.

Gorrec, F. (2015). The MORPHEUS II protein crystallization screen. *Acta Crystallogr F Struct Biol Commun, 71*(Pt 7), 831-37.

Guo, R. T., Ko, T. P., Chen, A. P., Kuo, C. J., Wang, A. H., & Liang, P. H. (2005). Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis. *J Biol Chem, 280*(21), 20762-74.

Guy, L., & Ettema, T. J. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology, 19*(12), 580-87.

Hartman, E., König, H., Kandler, O., & Hammes, W. (1990). Isolation of nucleotide activated amino acid and peptide precursors of the pseudomurein of *Methanobacterium thermoautotrophicum*. *FEMS Microbiol Lett, 69*(3), 271-75.

Hartmann, E., & König, H. (1990). Comparison of the biosynthesis of the methanobacterial pseudomurein and the eubacterial murein. *Naturwissenschaften, 77*(10), 472-75.

Hartmann, E., & Konig, H. (1990). Isolation of lipid activated pseudomurein precursors from *Methanobacterium thermoautotrophicum*. *Arch Microbiol, 153*(5), 444-47.

Hartmann, E., & König, H. (1994). A novel pathway of peptide biosynthesis found in methanogenic Archaea. *Arch Microbiol, 162*(6), 430-32.

Hartmann, E., Konig, H., Kandler, O., & Hammes, W. (1989). Isolation of nucleotide activated amino acid and peptide precursors of the pseudomurein of *Methanobacterium thermoautotrophicum*. *FEMS Microbiol Lett, 57*(3), 271-75.

He, H., Fang, H., Miller, M. D., Phillips, G. N., Jr., & Su, W. P. (2016). Improving the efficiency of molecular replacement by utilizing a new iterative transform phasing algorithm. *Acta Crystallogr A Found Adv, 72*(Pt 5), 539-47.

Holm, L., & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res, 38*(Web Server issue), 545-49.

Holm, L., & Laakso, L. M. (2016). Dali server update. *Nucleic Acids Res, 44*(W1), 351-55.

Jain, S., Caforio, A., & Driessen, A. J. (2014). Biosynthesis of archaeal membrane ether lipids. *Front Microbiol, 5*, 641-57.

Janiak, C. (2000). A critical account on π-π stacking in metal complexes with aromatic nitrogen-containing ligands. *J. Chem. Soc., Dalton Trans*(21), 3885-96.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci, 8*(3), 275-82.

Kabsch, W. (2010). XDS. *Acta Crystallographica Section D, 66*(2), 125-32.

Kandler, O., & König, H. (1978). Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria. *Arch Microbiol, 118*(2), 141-52.

Kandler, O., & König, H. (1998). Cell wall polymers in Archea (Archaebacteria). *Cell Mol Life Sci, 54*(4), 305-08.

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci, 12*(9), 1865-71.

Keegan, R. M., & Winn, M. D. (2008). MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 119-24.

Kim, C., & Lee, B. (2007). Accuracy of structure-based sequence alignment of automatic methods. *BMC Bioinformatics, 8*, 355.

Kletzin, A. (2007). General characteristics and important model organisms. In *Archaea*: American Society of Microbiology. Retrieved from http://www.asmscience.org/content/book/10.1128/9781555815516.ch02.

Klingl, A. (2014). S-layer and cytoplasmic membrane - exceptions from the typical archaeal cell wall with a focus on double membranes. *Front Microbiol, 5*, 624-30.

Koeduka, T., Louie, G. V., Orlova, I., Kish, C. M., Ibdah, M., Wilkerson, C. G., Bowman, M. E., Baiga, T. J., Noel, J. P., Dudareva, N., & Pichersky, E. (2008). The multiple phenylpropene synthases in both *Clarkia breweri* and *Petunia hybrida* represent two distinct protein lineages. *Plant J, 54*(3), 362-74.

König, H., Kandler, O., Jensen, M., & Rietschel, E. T. (1983). The primary structure of the glycan moiety of pseudomurein from *Methanobacterium thermoautotrophicum*. *Hoppe-Seyler's Z. Physiol Chem, 364*(6), 627-36.

Kouidmi, I., Levesque, R. C., & Paradis-Bleau, C. (2014). The biology of Mur ligases as an antibacterial target. *Mol Microbiol, 94*(2), 242-53.

Kreisl, P., & Kandler, O. (1986). Chemical structure of the cell wall polymer of *Methanosarcina*. *Syst Appl Microbiol, 7*(2), 293-99.

Krengel, U., & Imberty, A. (2007). Crystallography and Lectin Structure Database. In C. L. Nilsson (Ed.), *Lectins* (pp. 15-50). Amsterdam: Elsevier Science B.V. Retrieved from http://www.sciencedirect.com/science/article/pii/B978044453077650003X.

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2256-68.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol, 372*(3), 774-97.

Kruger, N. J. (1994). The Bradford method for protein quantitation. *Methods Mol Biol, 32*, 9-15.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol, 33*(7), 1870-74.

Laddomada, F., Miyachiro, M. M., & Dessen, A. (2016). Structural insights into protein-protein interactions involved in bacterial cell wall biogenesis. *Antibiotics (Basel), 5*(2).

Lake, J. A., & Sinsheimer, J. S. (2013). The deep roots of the rings of life. *Genome Biol Evol, 5*(12), 2440-48.

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res, 33*(Web Server issue), 299-302.

Langer, G., Cohen, S. X., Lamzin, V. S., & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols, 3*(7), 1171-79.

Laskowski, R. A. (2009). PDBsum new things. *Nucleic acids research, 37*(Database issue), 355-59.

Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., Li, D. C., Kong, Z., McTavish, S., Sang, C., Lambie, S. C., Janssen, P. H., Dey, D., & Attwood, G. T. (2010). The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLoS One, 5*(1), Article No. e8926.

Leipe, D. D., Koonin, E. V., & Aravind, L. (2003). Evolution and classification of P-loop kinases and related proteins. *J Mol Biol, 333*(4), 781-815.

Leps, B., Labischinski, H., Barnickel, G., Bradaczek, H., & Giesbrecht, P. (1984). A new proposal for the primary and secondary structure of the glycan moiety of pseudomurein. Conformational energy calculations on the glycan strands with talosaminuronic acid in 1C conformation and comparison with murein. *Eur J Biochem, 144*(2), 279-86.

Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W., & Huang, R. B. (2013). The multiple roles of histidine in protein interactions. *Chem Cent J, 7*(1), 44.

Liechti, G., Singh, R., Rossi, P. L., Gray, M. D., Adams, N. E., & Maurelli, A. T. (2018). Chlamydia trachomatis dapF Encodes a Bifunctional Enzyme Capable of Both d-Glutamate Racemase and Diaminopimelate Epimerase Activities. *MBio, 9*(2).

Liger, D., Masson, A., Blanot, D., van Heijenoort, J., & Parquet, C. (1995). Over-production, purification and properties of the uridine-diphosphate-N-acetylmuramate:L-alanine ligase from *Escherichia coli*. *Eur J Biochem, 230*(1), 80-7.

Lutkenhaus, J. (1993). FtsZ ring in bacterial cytokinesis. *Mol Microbiol, 9*(3), 403-9.

Mackereth, C. D., Arrowsmith, C. H., Edwards, A. M., & McIntosh, L. P. (2000). Zinc-bundle structure of the essential RNA polymerase subunit RPB10 from Methanobacterium thermoautotrophicum. *Proc Natl Acad Sci U S A, 97*(12), 6316-21.

Malhotra, O. P., Ambasht, P. K., Prabhakar, P., Lal, A. K., & Kayastha, A. M. (1996). An assay procedure for determining the rate of an enzyme reaction lacking an optical signal: validity of coupled enzyme assays. *Biochem Edu, 24*(1), 56-59.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., & Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res, 45*(D1), D200-d203.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-74.

McNicholas, S., Potterton, E., Wilson, K. S., & Noble, M. E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 386-94.

Mengin-Lecreulx, D., Falla, T., Blanot, D., van Heijenoort, J., Adams, D. J., & Chopra, I. (1999). Expression of the *Staphylococcus aureus* UDP-N-acetylmuramoyl- L-alanyl-D-glutamate:L-lysine ligase in *Escherichia coli* and effects on peptidoglycan biosynthesis and cell growth. *J Bacteriol, 181*(19), 5909-14.

Mol, C. D., Brooun, A., Dougan, D. R., Hilgers, M. T., Tari, L. W., Wijnands, R. A., Knuth, M. W., McRee, D. E., & Swanson, R. V. (2003). Crystal Structures of Active Fully Assembled Substrate- and Product-Bound Complexes of UDP-N-Acetylmuramic Acid:L-Alanine Ligase (MurC) from *Haemophilus influenzae*. *J Bacteriol, 185*(14), 4152-162.

Moore, S. J., Sowa, S. T., Schuchardt, C., Deery, E., Lawrence, A. D., Ramos, J. V., Billig, S., Birkemeyer, C., Chivers, P. T., Howard, M. J., Rigby, S. E., Layer, G., & Warren, M. J. (2017). Elucidation of the biosynthesis of the methane catalyst coenzyme F430. *Nature, 543*(7643), 78-82.

Morlot, C., Straume, D., Peters, K., Hegnar, O. A., Simon, N., Villard, A. M., Contreras-Martel, C., Leisico, F., Breukink, E., Gravier-Pelletier, C., Le Corre, L., Vollmer, W., Pietrancosta, N., Havarstein, L. S., & Zapun, A. (2018). Structure of the essential peptidoglycan amidotransferase MurT/GatD complex from Streptococcus pneumoniae. *Nat Commun, 9*(1), 3180.

Munch, D., Roemer, T., Lee, S. H., Engeser, M., Sahl, H. G., & Schneider, T. (2012). Identification and in vitro analysis of the GatD/MurT enzyme-complex catalyzing lipid II amidation in *Staphylococcus aureus*. *PLoS Pathog, 8*(1), e1002509.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 355-67.

Noldeke, E. R., Muckenfuss, L. M., Niemann, V., Muller, A., Stork, E., Zocher, G., Schneider, T., & Stehle, T. (2018). Structural basis of cell wall peptidoglycan amidation by the GatD/MurT complex of Staphylococcus aureus. *Sci Rep, 8*(1), 12953.

Oberto, J. (2013). SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics, 14*, 4.

Osman, K., Evangelopoulos, D., Basavannacharya, C., Gupta, A., McHugh, T. D., Bhakta, S., & Gibbons, S. (2012). An antibacterial from Hypericum acmosepalum inhibits ATP-dependent MurE ligase from Mycobacterium tuberculosis. *Int J Antimicrob Agents, 39*(2), 124-29.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., & Tucker, P. A. (2005). Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr, 61*(Pt 4), 449-57.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., & Tucker, P. A. (2009). On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr, 65*(Pt 10), 1089-97.

Paradis-Bleau, C., Lloyd, A., Sanschagrin, F., Maaroufi, H., Clarke, T., Blewett, A., Dowson, C., Roper, D. I., Bugg, T. D., & Levesque, R. C. (2009). *Pseudomonas aeruginosa* MurE amide ligase: enzyme kinetics and peptide inhibitor. *Biochem J, 421*(2), 263-72.

Pathak, E., Atri, N., & Mishra, R. (2014). Analysis of P-Loop and its Flanking Region Subsequence of Diverse NTPases Reveals Evolutionary Selected Residues. *Bioinformation, 10*(4), 216-20.

Paul, K., Nonoh, J. O., Mikulski, L., & Brune, A. (2012). "Methanoplasmatales," Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl Environ Microbiol, 78*(23), 8245-53.

Paxman, J. J., & Heras, B. (2017). Bioinformatics Tools and Resources for Analyzing Protein Structures. *Methods Mol Biol, 1549*, 209-20.

Pei, J., & Grishin, N. V. (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol Biol, 1079*, 263-71.

Perdih, A., Kotnik, M., Hodoscek, M., & Solmajer, T. (2007). Targeted molecular dynamics simulation studies of binding and conformational changes in *E. coli* MurD. *Proteins, 68*(1), 243-54.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-12.

Pratviel-Sosa, F., Acher, F., Trigalo, F., Blanot, D., Azerad, R., & van Heijenoort, J. (1994). Effect of various analogues of D-glutamic acid on the D-glutamate-adding enzyme from *Escherichia coli*. *FEMS Microbiol Lett, 115*(2-3), 223-28.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res, 42*(Web Server issue), 320-24.

Ruane, K. M., Lloyd, A. J., Fulop, V., Dowson, C. G., Barreteau, H., Boniface, A., Dementin, S., Blanot, D., Mengin-Lecreulx, D., Gobec, S., Dessen, A., & Roper, D. I. (2013). Specificity determinants for lysine incorporation in Staphylococcus aureus peptidoglycan as revealed by the structure of a MurE enzyme ternary complex. *J Biol Chem, 288*(46), 33439-48.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A., & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics, 16*(10), 944-5.

Saraste, M., Sibbald, P. R., & Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci, 15*(11), 430-34.

Sauter, J. (2017). *Investigation into enzymes involved in the biosynthesis of pseudomurein in thermophilic methanogens.* Masters thesis, University Duisburg-Essen, Germany.

Schleifer, K. H., & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev, 36*(4), 407-77.

Schleifer, K. H., Steber, J., & Mayer, H. (1982). Chemical composition and structure of the cell wall of *Halococcus morrhuae*. *Zentralblatt für Bakteriologie Mikrobiologie und Hygiene: I. Abt. Originale C: Allgemeine, angewandte und ökologische Mikrobiologie, 3*(2), 171-78.

Schneider, T., & Sahl, H. G. (2010). An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. *Int J Med Microbiol, 300*(2-3), 161-69.

Schofield, L. R., Beattie, A. K., Tootill, C. M., Dey, D., & Ronimus, R. S. (2015). Biochemical Characterisation of Phage Pseudomurein Endoisopeptidases PeiW and PeiP Using Synthetic Peptides. *Archaea, 2015*, 12.

Schrodinger, LLC. (2015). The PyMOL Molecular Graphics System, Version 1.8.

Sham, L. T., Butler, E. K., Lebar, M. D., Kahne, D., Bernhardt, T. G., & Ruiz, N. (2014). Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. *Science, 345*(6193), 220-22.

Sheng, Y., Sun, X., Shen, Y., Bognar, A. L., Baker, E. N., & Smith, C. A. (2000). Structural and functional similarities in the ADP-forming amide bond ligase superfamily: implications for a substrate-induced conformational change in folylpolyglutamate synthetase. *J Mol Biol, 302*(2), 427-40.

Sink, R., Barreteau, H., Patin, D., Mengin-Lecreulx, D., Gobec, S., & Blanot, D. (2013). MurD enzymes: some recent developments. *Biomol Concepts, 4*(6), 539-56.

Sink, R., Kotnik, M., Zega, A., Barreteau, H., Gobec, S., Blanot, D., Dessen, A., & Contreras-Martel, C. (2016). Crystallographic Study of Peptidoglycan Biosynthesis Enzyme MurD: Domain Movement Revisited. *PloS one, 11*(3), e0152075.

Sleytr, U. B., Schuster, B., Egelseer, E. M., & Pum, D. (2014). S-layers: principles and applications. *FEMS Microbiol Rev, 38*(5), 823-64.

Smith, C. A. (2006). Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J Mol Biol, 362*(4), 640-55.

Smith, C. A., Cross, J. A., Bognar, A. L., & Sun, X. (2006). Mutation of Gly51 to serine in the P-loop of *Lactobacillus casei* folylpolyglutamate synthetase abolishes activity by altering the conformation of two adjacent loops. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 5), 548-58.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J. N., & et al. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J. Bacteriol. , 179*(22), 7135-7155.

Sobac, B., Talbot, P., Haut, B., Rednikov, A., & Colinet, P. (2015). A comprehensive analysis of the evaporation of a liquid spherical drop. *J Colloid Interface Sci, 438*, 306-17.

Sowers, K. R., Boone, J. E., & Gunsalus, R. P. (1993). Disaggregation of *Methanosarcina spp*. and growth as single cells at elevated osmolarity. *Applied and Environmental Microbiology, 59*(11), 3832-39.

Stetter, K. O., Thomm, M., Winter, J., Wildgruber, G., Huber, H., Zillig, W., Jané-Covic, D., König, H., Palm, P., & Wunderl, S. (1981). *Methanothermus fervidus*, sp. nov., a novel extremely thermophilic methanogen isolated from an Icelandic hot spring. *ZBL BAKT MIK HYG I C, 2*(2), 166-78.

Stroupe, M. E., Leech, H. K., Daniels, D. S., Warren, M. J., & Getzoff, E. D. (2003). CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis. *Nat Struct Biol, 10*(12), 1064-73.

Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif, 41*(1), 207-34.

Sun, X., Cross, J. A., Bognar, A. L., Baker, E. N., & Smith, C. A. (2001). Folate-binding triggers the activation of folylpolyglutamate synthetase. *J Mol Biol, 310*(5), 1067-78.

Tanabe, M., & Kanehisa, M. (2012). Using the KEGG database resource. *Curr Protoc Bioinformatics, Chapter 1*, Unit1.12.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., & Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 61-69.

Thorn, A., & Sheldrick, G. M. (2013). Extending molecular-replacement solutions with SHELXE. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 11), 2251-56.

Till, M., Robson, A., Byrne, M. J., Nair, A. V., Kolek, S. A., Shaw Stewart, P. D., & Race, P. R. (2013). Improving the success rate of protein crystallization by random microseed matrix screening. *J Vis Exp*(78).

Vagin, A., & Lebedev, A. MoRDa, an automatic molecular replacement pipeline.

Vagin, A., & Lebedev, A. (2015). MoRDa, an automatic molecular replacement pipeline. *Acta Crystallographica Section A, 71*(a1), s19.

Van Nuland, Y. (2011). *Investigation into the evolution and biosynthesis of pseudomurein in thermophilic methanogens.* Masters thesis, University of Wageningen, The Netherland.

Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiol Rev, 32*(2), 149-67.

Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng, 8*(2), 127-34.

Walsh, A. W., Falk, P. J., Thanassi, J., Discotto, L., Pucci, M. J., & Ho, H. T. (1999). Comparison of the D-glutamate-adding enzymes from selected gram-positive and gram-negative bacteria. *J Bacteriol, 181*(17), 5395-401.

Walsh, C. T. (1989). Enzymes in the D-alanine branch of bacterial cell wall peptidoglycan assembly. *J Biol Chem, 264*(5), 2393-96.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol, 112*, 531-52.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 235-42.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A., 74*(11), 5088-90.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria* and *Eucarya*. *Proc. Natl. Acad. Sci. U.S.A., 87*, 4576-79.

Wolf, M. (2010). *Discovering the pathway of methanogen pseudomurein biosynthesis and the evolutionary correlation to bacterial murein.* Bachelor thesis, University of Duisburg, Germany.

Wong, J. T., Chen, J., Mat, W. K., Ng, S. K., & Xue, H. (2007). Polyphasic evidence delineating the root of life and roots of biological domains. *Gene, 403*(1-2), 39-52.

Yan, Y., Munshi, S., Leiting, B., Anderson, M. S., Chrzas, J., & Chen, Z. (2000). Crystal structure of *Escherichia coli* UDPMurNAc-tripeptide d-alanyl-d-alanine-adding enzyme (MurF) at 2.3 A resolution. *J Mol Biol, 304*(3), 435-45.

Zeikus, J. G., & Bowen, V. G. (1975). Fine structure of *Methanospirillum hungatii*. *Journal of Bacteriology, 121*(1), 373-80.

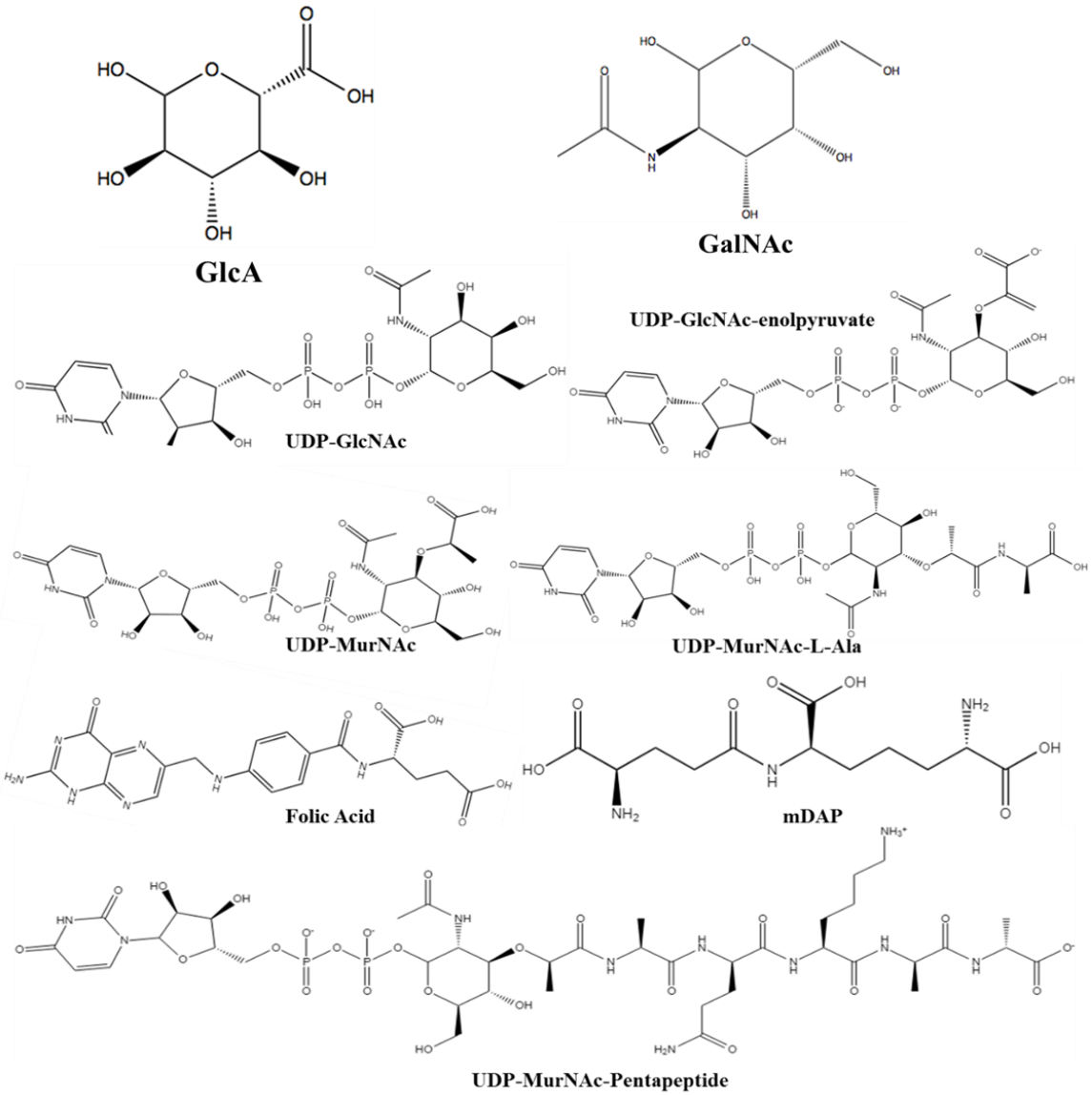
Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., & Minor, W. (2008). Data mining of metal ion environments present in protein structures. *J Inorg Biochem, 102*(9), 1765-76.

Zheng, K., Ngo, P. D., Owens, V. L., Yang, X. P., & Mansoorabadi, S. O. (2016). The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science, 354*(6310), 339-42.

Ziegler, K., Diener, A., Herpin, C., Richter, R., Deutzmann, R., & Lockau, W. (1998). Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *Eur J Biochem, 254*(1), 154-59.

Chapter 8 Appendices

Appendix 1 Chemical structures of various sugars found in bacterial peptidoglycan and archaeal cell wall types and the substrates and intermediate products discussed in the thesis. Labelling of each sugars are based on the abbreviations used in the thesis; Glucuronic acid (GlcA), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), *N*-acetylmuramic acid residues (MurNAc).



Appendix 2 List of organisms used for the phylogenetic analysis

Bacteria

Thermus thermophilus HB8
Thermotoga maritima MSB8
Aquifex aeolicus
Dictyoglomus thermophilum
Verrucomicrobia bacterium IMCC26134
Verrucosispora maris
Xanthomonas oryzae pv. oryzae PXO99A
Caulobacter sp. K31
Burkholderia multivorans ATCC 17616
Desulfobacterium autotrophicum
Campylobacter coli FB1
Acidobacterium capsulatum
Bacteroidales bacterium CF
Prevotella ruminicola
Chloroflexus aurantiacus
Spirochaeta thermophila DSM 6578
Thermodesulfovibrio yellowstonii
Nitrospira defluvii
Caldicellulosiruptor kristjanssonii
Caldicellulosiruptor hydrothermalis
Thermoanaerobacter brockii
Clostridium butyricum
Bacillus subtilis subsp. subtilis 168
Bacillus halodurans
Geobacter sulfurreducens KN400
Geobacillus stearothermophilus
Deinococcus geothermalis
Actinomyces meyeri
Bifidobacterium breve DSM 20213
Bifidobacterium thermophilum
Fusobacterium hwasookii
Faecalibacterium prausnitzii L26
Thermomicrobium roseum
Hydrogenophaga sp. RAC07
Mycobacterium bovis AF2122/97
Mycobacterium tuberculosis F11
Haemophilus influenzae Rd KW20
Streptococcus pneumoniae A026
Yersinia pestis CO92
Pseudomonas aeruginosa PAO1
Acinetobacter baumannii ATCC 17978
Streptomyces coelicolor
Acetobacter pasteurianus 386B
Acetobacter aceti
Desulfurobacterium thermolithotrophum
Fimbriimonas ginsengisoli
Caldithrix abyssi
Caldisericum exile
Chlamydia avium
Chlamydia muridarum Nigg
Chlorobium phaeobacteroides DSM 266
Desulfurispirillum indicum
Candidatus Cloacimonas acidaminovorans
Cyanobacterium sp. HL69
Deferribacter desulfuricans
Elusimicrobium minutum
Fibrobacter succinogenes
Gemmatimonas aurantiaca
Ignavibacterium album
Planctomyces sp. SHPL14
Thermodesulfobacterium geofontis

Methanogens

Methanothermobacter thermautotrophicus ΔH
Methanothermobacter thermautotrophicus CaT2
Methanothermobacter marburgensis
Methanothermobacter wolfeii
Methanothermus fervidus
Methanosphaera stadtmanae
Methanobrevibacter ruminantium Mru
Methanobrevibacter millerae SM9
Methanobrevibacter cuticularis
Methanobrevibacter wolinii
Methanobrevibacter filiformis
Methanobrevibacter boviskorani
Methanobrevibacter oralis
Methanobrevibacter sp. AbM4
Methanobrevibacter sp. YE315
Methanobrevibacter curvatus
Methanobrevibacter smithii DSM 2375
Methanobrevibacter smithii DSM 2374
Methanobrevibacter olleyae
Methanobrevibacter arboriphilus
Methanobacterium lacus
Methanobacterium paludis
Methanobacterium sp. SMA27
Methanobacterium sp. Maddlingley MBC34
Methanobacterium formicicum BRM9
Methanobacterium MB1 (MBMB1)
Methanobacterium sp. 42_16
Methanobacterium formicicum DSM1535
Methanobacterium formicicum
Methanobacteriaceae archaeon 41_258

Appendix 3 List of proteins attempted for structure determination experiments.

Proteins	Putative ligase type	Accession Number	Protein purification	Crystal Screening	Crystal optimisation	Co-crystallisation	Crystal testing	Structure solved
Mth530	pMurC	WP_010876169.1	Ni-NTA	√	√	NA	NA	NA
Mfer336	pMurC	WP_013413417.1	Ni-NTA	√	√	√	√	√
Mth531	pMurD1	WP_010876170.1	Ni-NTA	√	√	√	√	NA
Mbb918 (1-500)	pMurD1	(Wolf, 2010)	(Wolf, 2010)+ion exchange	√	√	√	√	NA
Mth532	pMurD2	WP_010876171.1	Ni-NTA	√	√	√	√	NA
Mfer337	pMurD1	WP_013413418.1	Ni-NTA	√	√	√	√	NA
Mfer340	pMurD2	WP_013413421.1	Ni-NTA	√	√	√	√	NA
Mfer762	pMurE	WP_013413839.1	Ni-NTA	√	√	√	√	√
Mth734	pMurE	B69198	Ni-NTA+Gel filtration	√	√	√	√	√
Mfer1205	pMurF	WP_013296078.1	Ni-NTA	√	NA	NA	NA	NA
Mbb1268	pMurF	WP_013296078.1	Ni-NTA	√	NA	NA	NA	NA

Appendix 4 Summary of crystallisation experiments performed to obtain pMurE crystals and substrate bound pMurE structures.

Proteins	**Screening**								**Remarks**
	SS1+2	JCSG+	SG1	MIDAS	MORPHEUS	INDEX	MORPHEUS II	PEG/ION	
Mfer762	✓	✓	✓	✓	✓	✓	NA	NA	SG1_E8 produced single good diffracting crystal, no optimisation required
Mth734	✓	✓	✓	✓	✓	✓	✓	✓	No single crystal produced, Morpheus II_H7 multiple needle cluster used for seeding

Optimisation							
	Silver Bullet Bio	Additive screen	pH optimisation	PEG optimisation	Protein buffer ratio alteration	rMMS	Only rMMS improved protein crystal
Mth734	✓	✓	✓	✓	✓	✓	

Co-crystallisation							
Mfer762	ATP/ADP ± $MgCl_2$	dipeptide (γ-Glu-ε-Lys) ±ATP/ADP ± $MgCl_2$	ATP/ADP + L-Glu/L-Lys ± $MgCl_2$	UDP ± ATP/ADP +L-Glu and L-Lys + $MgCl_2$	± UDP +dipeptide (γ-Glu-ε-Lys)	UDP + AMPPNP +dipeptide (γ-Glu-ε-Lys) + $MgCl_2$	Similar morphology crystal as that of Apo protein produced
Mth734	ATP/ADP ± $MgCl_2$		ATP/ADP + L-Glu/L-Lys ± $MgCl_2$		UDP ± ATP/ADP + MgCl2		Did not help crystallisation

Soaking experiments				
Mfer762	ATP/ADP + L-Glu/L-Lys ± $MgCl_2$	UDP ± ATP/ADP + L-Glu/L-Lys ± $MgCl_2$	UMA + AMPPNP ± $MgCl_2$	Only UDP structure present in final structure. Soak only with UDP produced UDP_1 structure and that with UDP+ADP+$MgCl_2$+L-Glu produced UDP_2 structure
Mth734	UMA ± dipeptide (γ-Glu-ε-Lys) ±ATP/ADP ± $MgCl_2$		UDP + ATP/ADP+dipeptide (γ-Glu-ε-Lys) + $MgCl_2$	No solvable resolution dataset

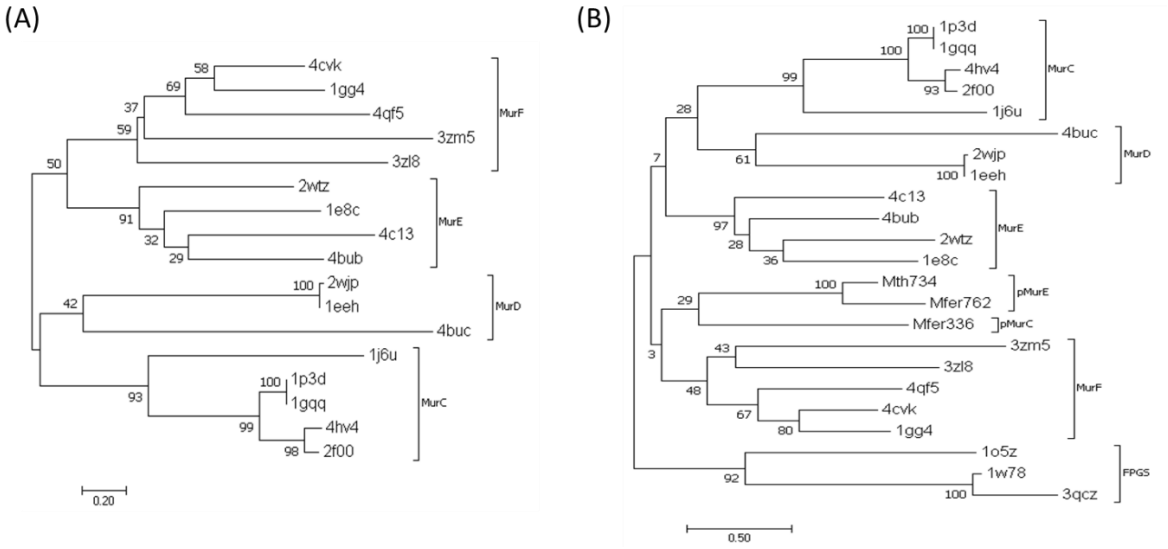
Appendix 5 Summary of various datasets collected for pMurE targets during the project. Structures referred in the thesis chapter are highlighted in bold text.

Proteins	**Experiment**	**Detector**	**Space group**	**Res (Å)**	**Structure solving**	**Refinement**	**Ligand**	**Remark**
Mfer762	Apo	ADSC Quantum 210r	*P*6$_1$	1.90	Combination of MR programs, used 4QID as template	Both Refmac5 and phenix.refine	NA	Used for MR; first apo structure solved
Mfer762	UDP soaked	ADSC Quantum 210r	P61	1.80	MR using Apo Mfer762	Both Refmac5 and phenix.refine	UDP	**Mfer762-UDP_1 structure**; UDP at exterior surface
Mfer762	Lys co-crystallised	ADSC Quantum 210r	C*2*	1.83	MR using Apo Mfer762	Both Refmac5 and phenix.refine	NA	No substrate bound and structure similar to Apo
Mfer762	Lys, ATP, Mg^{2+} co-crystallised	ADSC Quantum 210r	*P*6$_1$	1.67	MR using Apo Mfer762	Both Refmac5 and phenix.refine	NA	Did not contain soaks; R_{free} imported from apo and used as **Apo structure**
Mfer762	Glu co-crystallised	ADSC Quantum 210r	*P*6$_1$	1.93	MR using Apo Mfer762	Both Refmac5 and phenix.refine	NA	Structure similar to Apo; no soaks present
Mfer762	Glu, ATP, Mg^{2+} co-crystallised	ADSC Quantum 210r	*P*6$_1$	1.80	MR using Apo Mfer762	Both Refmac5 and phenix.refine	NA	Structure similar to Apo; no soaks present
Mfer762	UDP, ADP, Lys soaked	Eiger	C*2*	2.6	MR using Apo Mfer762	Both Refmac5 and phenix.refine	NA	Structure similar to Mfer762-UDP_2
Mfer762	UDP, ADP, Glu soaked	Eiger	C*2*	2.0	MR using Apo Mfer762	Both Refmac5 and phenix.refine	UDP	**Mfer762-UDP_2 structure**; UDP at internal surface
Mfer762	UMA	Eiger	*P*6$_1$	3.2	MR using Apo Mfer762	phenix.refine	NA	No substrate bound
Mfer762	UMA, AMPPNP, $MgCl_2$	Eiger	*P*6$_1$	3.5	MR using Apo Mfer762	phenix.refine	NA	No substrate bound
Mth734	Apo	Eiger	*P*2$_1$2$_1$2$_1$	2.7	MR using Apo Mfer762	Refmac5	NA	**Mth734 structure**
Mth734	Apo	Eiger	*P*2$_1$2$_1$2$_1$	3.0	MR using Apo Mfer762	Refmac5	NA	Structure similar to Mth734; no additional information

Appendix 6 Phylogenetic analyses attempts using various combinations of murein and pseudomurein peptide ligases.

Pseudomurein ligases										
pMurC	✓					✓		✓		✓
pMurD1		✓	✓			✓		✓		✓
pMurD2			✓			✓		✓		✓
pMurE				✓		✓			✓	✓
pMurF					✓	✓			✓	✓
Murein ligases										
MurC	✓						✓	✓		✓
MurD		✓	✓				✓	✓		✓
MurE				✓			✓		✓	✓
MurF					✓		✓		✓	✓
FPGS						+/-	+/-	+/-	+/-	+/-

Supplementary figures



Supp. Figure 1 Molecular phylogenetic analysis using the available peptide ligase structures. (A) Maximum likelihood tree based on murein peptide ligases only (reduced to avoid duplication). (B) Maximum likelihood tree based on both murein and pseudomurein peptide ligases structures The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Kumar, *et al.*, 2016).