The evolution of AroA and MurA enzymes from *Bacillus*

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Biological Sciences

at

The University of Waikato

by

TIFANY OULAVALLICKAL

The University of Waikato

2016

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Abstract

MurA and AroA are important antibacterial targets due to their essentiality in microorganisms and the absence of their respective pathways within mammals. Although much research has focussed on these enzymes, little is known about the evolution of AroA and MurA. Ancestral sequence reconstruction (ASR), a technique used to infer sequences of ancestral proteins and study their properties in the laboratory, was used to study the evolution of AroA and MurA enzymes from *Bacillus*. The ancestral AroA and MurA enzymes were functionally and structurally characterised and their properties compared to those of contemporary AroA and MurA.

AroA and MurA from the last common ancestor (LCA) of *Bacillus* along with three intermediate *Bacillus* MurA ancestors were inferred using ASR. The reconstructed AroA and MurA ancestral enzymes show comparable kinetic properties to the contemporary enzymes. The thermal properties of *Bacillus* AroA LCA, *Bacillus* MurA LCA and the remaining three MurA ancestors were found to be moderately thermophilic. However, the contrasting thermal profiles observed for AroA, MurA and LeuB ancestral enzymes from *Bacillus* at the same period of time led to the hypothesis that reconstructed ancestral enzymes provide a snapshot of the evolving host at a given point in time.

The exclusive inhibition of AroA and MurA by their inhibitors glyphosate and fosfomycin respectively remains unchanged for AroA LCA and MurA LCA. On the other hand, the high affinity for PEP and increased glyphosate sensitivity exhibited by AroA LCA indicate AroA LCA to be intermediate between class I (glyphosate sensitive) and class II (glyphosate insensitive) AroA. Characterisation of AroA LCA *in vivo* resulted in AroA LCA posing a fitness cost to the cells carrying this enzyme on a plasmid in a Δ*aroA* background. AroA LCA and MurA LCA remain highly structurally similar to the contemporary enzymes, with minor differences predominantly located on the protein surface.

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my primary supervisor Professor Vic Arcus. Thank you for providing me with the opportunity to work on this project, and for your constant encouragement and guidance throughout the study. I also would like to thank my co-supervisor, Dr Jo Hobbs, for teaching me everything in the lab. Thank you for all the help you have provided throughout this project and for being there for me every step of the way.

I extend my sincere gratitude to everyone in the Proteins and Microbes Laboratory, for all your help and incredible support throughout this journey and special thanks to everyone in C2.03 lab, TRU, and chemistry lab that helped me with my research. I would like to thank all my colleagues, past and present, specially: Dr Ali Ruthe, Dr Emma Andrews, Dr Emma Summers, Dr Jo McKenzie, Dr John Steemson, Dr Judith Burrows, Dr Magdalena Bereza, Dr Mark Liddament, Dr Mathew Cumming, Abby, Chelsea, Claire, Emily, Erica, Heng, Jo D, Joel, Konny and Vikas. I would like to thank Emma A, Emma S and Associate Professor Ian McDonald for taking time to help me with my chapters. Thanks to Cheryl Ward for helping me getting through the final stages of shaping up my thesis. Emma A, thank you for always being there for me, Judith, thank you for your motherly care and for always looking out for me, Emma S, thank you for our great chats, Erica, thank you for being patient through my endless questions and for all your help, and Magda, thank you for your constant support and for providing me with a great writing space.

Thanks to all my friends and family for the unconditional support and encouragement over the past few years. I want to specially thank Jay, who has helped me so much through this journey. I would not have gotten this far without your help and I cannot thank you enough. Special thanks to Rebecca for all the hugs and Florence for our catch up coffees and relaxing trips.

Finally, I would like to thank Papa and Mamma for your constant love, support, patience and prayers. Words cannot express how grateful I am for always being there for me and for making this venture possible. Your prayers for me sustained me thus far.

I would like to thank University of Waikato for my scholarship and travel funds, and Maurice Wilkins Centre for funding the project. Also, special thanks to our collaborators at the University of Otago.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

SI (Systeme Internationale d'Unités) abbreviations for units and standard notations for chemical elements and formulae are used throughout this thesis. Other abbreviations are listed below.

°C	degrees Celsius
3′	three prime
5′	five prime
aa	amino acid
Abs	absorbance
ACN	Acetonitrile
ADP	adenosine-5′ diphosphate
Ala (A)	Alanine
ANC	ancestor
Arg (R)	Arginine
AroA	5-enolpyruvylshikimate-3-phosphate synthase
Asn (N)	Asparagine
Asp (D)	Aspartic acid (Aspartate)
ASR	ancestral sequence reconstruction
ATP	adenosine-5′ triphosphate
BI	Bayesian inference
BLAST	basic local alignment search tool
bp	base pair(s)
BSA	bovine serum albumin
C-terminal	carboxyl terminus of peptide chain
CV	column volume
Cys (C)	Cysteine

Da	Daltons
DAHP	3-Deoxy-D-arabinoheptulosonate-7-phosphate
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
EC	Enzyme Commission number
EDTA	ethylene diamine tetraacetic acid
EF-Tu	elongation factor Tu
EPSP	5-enolpyruvylshikimate-3-phosphate
F	forward
g	times the force of gravity
Gln (Q)	Glutamine
Glu (E)	Glutamic acid (Glutamate)
Gly (G)	Glycine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	poly histidine tag
His (H)	Histidine
HPLC	high performance liquid chromatography
Ile (I)	Isoleucine
IMAC	immobilised metal affinity chromatography
IPTG	isopropylthio-β-D-galactosidase
kb	kilobase
kDa	kiloDaltons
KO	knock out
LB	Luria Bertani broth
LCA	last common ancestor
Leu (L)	Leucine
LeuB	3-isopropylmalate dehydrogenase gene

Lys (K)	Lysine
mA	milliamps
mAU	milli absorbance units
MESG	2-amino-6-mercapto-7-methylpurine riboside
Met (M)	Methionine
ML	maximum likelihood
MP	maximum parsimony
MQ	milli Q water - ion exchanged purified water
MurA	UDP-N-acetylglucosamine enolpyruvyl transferase
MurA ANC1	MurA reconstructed from ANC1 node
MurA ANC2	MurA reconstructed from ANC2 node
MurA ANC3	MurA reconstructed from ANC3 node
MW	molecular weight
Myr	million years
NAD+	nicotinamide adenine dinucleotide (oxidised)
NADH	nicotinamide adenine dinucleotide (reduced)
NADPH	Nicotinamide adenine dinucleotide phosphate
NCBI	National Centre for Biotechnology Information
NMR	nuclear magnetic resonance
N-terminal	amino terminus of peptide chain
OD	optical density
OGT	optimal growth temperature
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PEP	phosphoenolpyruvate

Phe (F)	Phenylalanine
pI	isoelectric point
Pi	inorganic phosphate
PNP	purine nucleoside phosphorylase
Pro (P)	Proline
R	reverse
RMSD	root mean square deviation
rpm	revolutions per minute
RT	room temperature
rTEV	recombinant tobacco etch virus
S3P	shikimate-3-phosphate
SCOP	Structural Classification of Proteins
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
Ser (S)	Serine
SSM	secondary structure matching
TAE	tris acetate ethylene diamine tetraacetic acid
TB	tuberculosis or terrific broth
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
Thr (T)	Threonine
T_m	melting temperature
T_{opt}	optimal temperature for activity
Trp (W)	Tryptophan
Tyr (Y)	Tyrosine
UDP	uridine diphosphate

UMP	uridine monophosphate
UNAG	UDP-*N*-acetylglucosamine
UV	ultraviolet light (280 nm)
V	volts
v/v	volume per volume
Val (V)	Valine
w/v	weight per volume

Chapter 1

Introduction

1.1 Antibiotic Resistance and Multi-target Drug Development

Infectious diseases are the second leading cause of death worldwide and will remain a dominant aspect of future health considerations (Fauci 2001). Pathogenic bacteria are prokaryotes with structural features and metabolic characteristics significantly different from eukaryotes, making it relatively simple to find and develop antibacterial drugs with minimal side effects (National Institutes of Health 2007). Antibiotics have played a critical role in the fight against infectious diseases caused by bacteria. However, the development of resistance against antibacterial drugs is becoming a global concern. Alarmingly, many strains have also developed resistance to multiple drugs (National Institutes of Health 2007). Therefore, there is a growing need to find new targets for the development of new antibiotics (McDevitt et al. 2002; Livermore 2004).

Many bacterial targets have been identified and extensively evaluated, and research is ongoing to find promising inhibitors. Examples of some pathways of interest include the shikimate pathway responsible for aromatic amino acid biosynthesis and the peptidoglycan biosynthesis pathway (McDevitt et al. 2002; Amer et al. 2008). The inhibition of any enzyme involved in any of the above pathways may prove lethal to the organism and thus make very good antibiotic targets. There are many antibacterial agents that target single enzymes. However, it has been observed that drugs that interact with multiple molecular sites, the structures of which are determined by multiple genes, have a low likelihood of developing a high level of endogenous resistance (Silver 2007). This observation favours the development of multi-target drugs over single-target compounds for the development of novel antibiotics. This hypothesis is also supported by network models, suggesting that partial inhibition of a small number of targets can be more efficient than the complete inhibition of a single target (Csermely et al. 2005). Many researchers have studied the shikimate pathway and peptidoglycan pathway, and pointed out that the respective enzymes AroA and MurA, are structurally and mechanistically similar to each other (Eschenburg et al. 2003; Oberdorfer et al. 2012). Such similarities offer the opportunity to find multi-target

inhibitors. Consequently, understanding the evolution of these enzymes imparts us with information that could be used for the design of multi-target inhibitors as well as understanding the origin of antibiotic resistance (Nobeli et al. 2009).

1.2 Enolpyruvyl Transferases

The peptidoglycan synthesis pathway and the shikimate pathway are absent from mammals but are essential for the survival of microorganisms (McDevitt et al. 2002; Eschenburg et al. 2003). This makes the enzymes of these pathways good antibacterial targets. The enolpyruvyl transferase family is a small family of enzymes consisting of AroA (5-enolpyruvylshikimate-3-phosphate synthase, EC 2.5.1.19), MurA (UDP-*N*-acetylglucosamine enolpyruvyl transferase, EC 2.5.1.7) (Eschenburg et al. 2003) and the putative enolpyruvyl transferase NikO (SCOP number: 55209) (Lauer et al. 2000; Oberdorfer et al. 2012). AroA is the sixth enzyme of the shikimate pathway (Funke 2008), MurA catalyses the first committed step of peptidoglycan biosynthesis (Skarzynski et al. 1996; Du et al. 2000) and NikO is a key enzyme in the biosynthesis of nikkomycins (Oberdorfer et al. 2012). The characteristic reaction carried out by members of the enolpyruvyl transferase family is the transfer of the enolpyruvyl moiety from phosphoenolpyruvate (PEP) to their respective substrates (Figure 1.1). AroA and MurA have always been reported as the only two known members of the enolpyruvyl transferase family (Skarzynski et al. 1996; Eschenburg et al. 2003; Biery 2007) until the characterisation of NikO which showed significant sequence identity to AroA and MurA. For example, MurA shares ~30 % sequence identity with NikO (Lauer et al. 2000; Oberdorfer et al. 2012). NikO was also found to be inhibited by MurA's unique inhibitor fosfomycin (Oberdorfer et al. 2012).

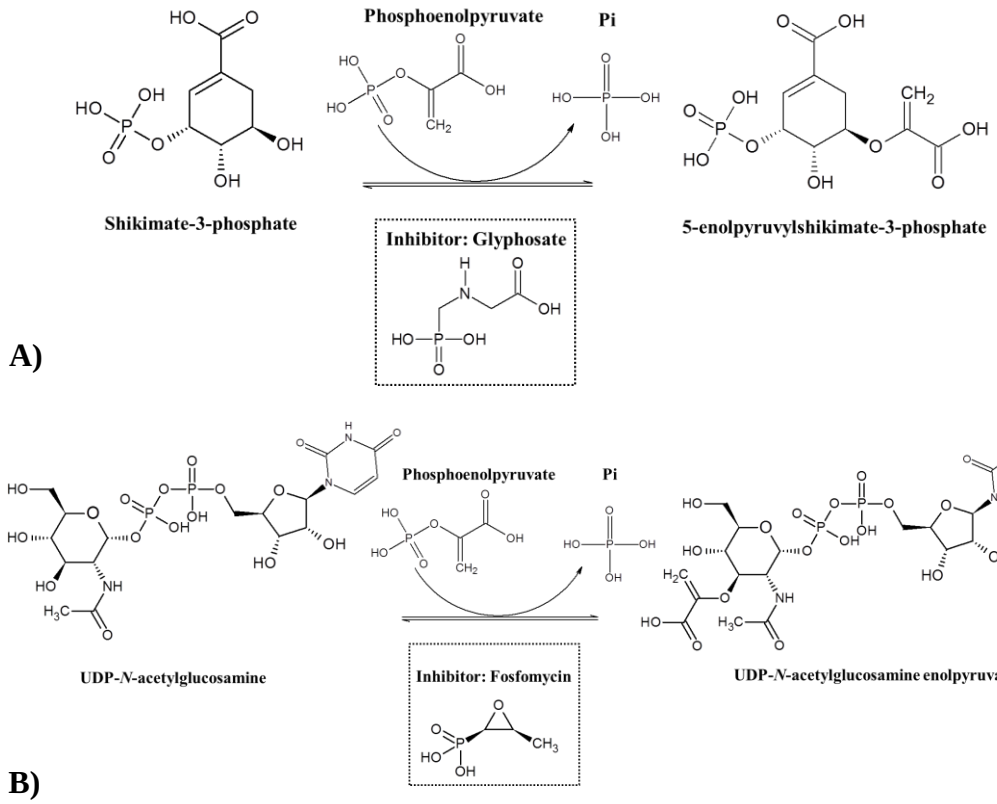


Figure 1.1: The enolpyruvyl transferase reaction catalysed by AroA and MurA
A) The reaction mechanism of AroA. B) The reaction mechanism of MurA.

The analysis of MurA sequences has led to the identification of two genes encoding for UDP-*N*-acetylglucosamine enolpyruvyltransferase in Gram positive bacteria, *murA* and *murZ* (sometimes called *murA2*), which possibly originated from a gene duplication event, whereas only one gene, *murA*, is found in Gram negative bacteria (Du et al. 2000; Blake et al. 2009). The *in vitro* characterisation of MurA and MurZ from *Streptococcus pneumoniae* showed that both enzymes catalyse the same enolpyruvyl transfer reaction from PEP to UNAG, have similar catalytic parameters and are both inhibited by fosfomycin (Du et al. 2000). However, in *Staphylococcus aureus*, MurA is the primary enzyme with five times greater expression than MurZ and a 25% reduction in peptidoglycan upon MurA inactivation as opposed to minimal change with MurZ inactivation (Blake et al. 2009).

Independent inactivation of MurA and MurZ in *S. pneumoniae* and *S. aureus* shows that either of them can sustain viability, however double deletions are unachievable showing that at least one UDP-*N*-acetylglucosamine enolpyruvyltransferse enzyme is essential to these organisms (Du et al. 2000; Blake et al. 2009). In contrast, MurA is an essential enzyme in *Bacillus subtilis*

and *Bacillus anthracis* as MurZ is unable to compensate for deletion of MurA (Kock et al. 2004; Kedar et al. 2008).

The essentiality of AroA has also been tested in different pathogenic organisms. The deletion of the *aroA* gene from *S. pneumoniae* caused the strain to be entirely attenuated in terms of virulence (McDevitt et al. 2002). In 2002, Parish and Stoker disrupted the gene that encodes shikimate kinase, *aroK*, and demonstrated that the shikimate pathway is essential for the viability of *Mycobacterium tuberculosis* (Parish & Stoker 2002).

1.3 AroA (5-enolpyruvylshikimate-3-phosphate synthase)

1.3.1 Pathway

The shikimate pathway is known to be the common aromatic biosynthetic pathway, conserved in plants, algae, fungi and bacteria (McConkey 1999; Priestman et al. 2005b). In bacteria, the shikimate pathway is almost exclusively used for the synthesis of aromatic amino acids for protein synthesis (Herrmann 1995b; Tzin & Galili 2010).

The synthesis of aromatic amino acids involves two processes: (i) the synthesis of chorismate from PEP and erythrose-4-phosphate (called the shikimate pathway), and (ii) the synthesis of the three aromatic amino acids (phenylalanine, tyrosine and tryptophan) and *p*-amino and *p*-hydroxy benzoate from chorismate (Herrmann 1995a; Herrmann 1995b; Parish & Stoker 2002). The shikimate pathway involves seven enzymes (Figure 1.2). The first step of the shikimate pathway is catalysed by 3-Deoxy-D-arabinoheptulosonate-7-phosphate (DAHP) synthase causing the condensation of PEP and erythrose-4-phosphate (Herrmann 1995a). The second step involves the formation of the cyclohexane derivative 3-dehydroquinate using divalent cations and NAD^+ cofactors (Herrmann 1995b; Maeda & Dudareva 2012). The last five steps are responsible for the introduction of one side chain and two double bonds to the ring. The first double bond is introduced through the dehydration of 3-dehydroquinate to 3-dehydroshikimate by 3-dehydroquinate dehydratase. This is followed by the reversible reduction of 3-dehydroshikimate into shikimate using NADPH by shikimate dehydrogenase. Shikimate kinase, during the fifth step of the shikimate pathway, produces shikimate-3-phosphate *via* phosphorylation of shikimate using ATP. A side chain is introduced in the sixth step catalysed by 5-enolpyruvyl-shikimate-3-phoshate synthase (AroA)

(Herrmann 1995b; Maeda & Dudareva 2012). The two most studied enzymes on the pathway are shikimate kinase and AroA. The final step is catalysed by chorismate synthase by converting 5-enolpyruvyl-shikimate-3-phoshate to chorismate, thereby introducing the second double bond in the ring (Priestman et al. 2005b; Maeda & Dudareva 2012).

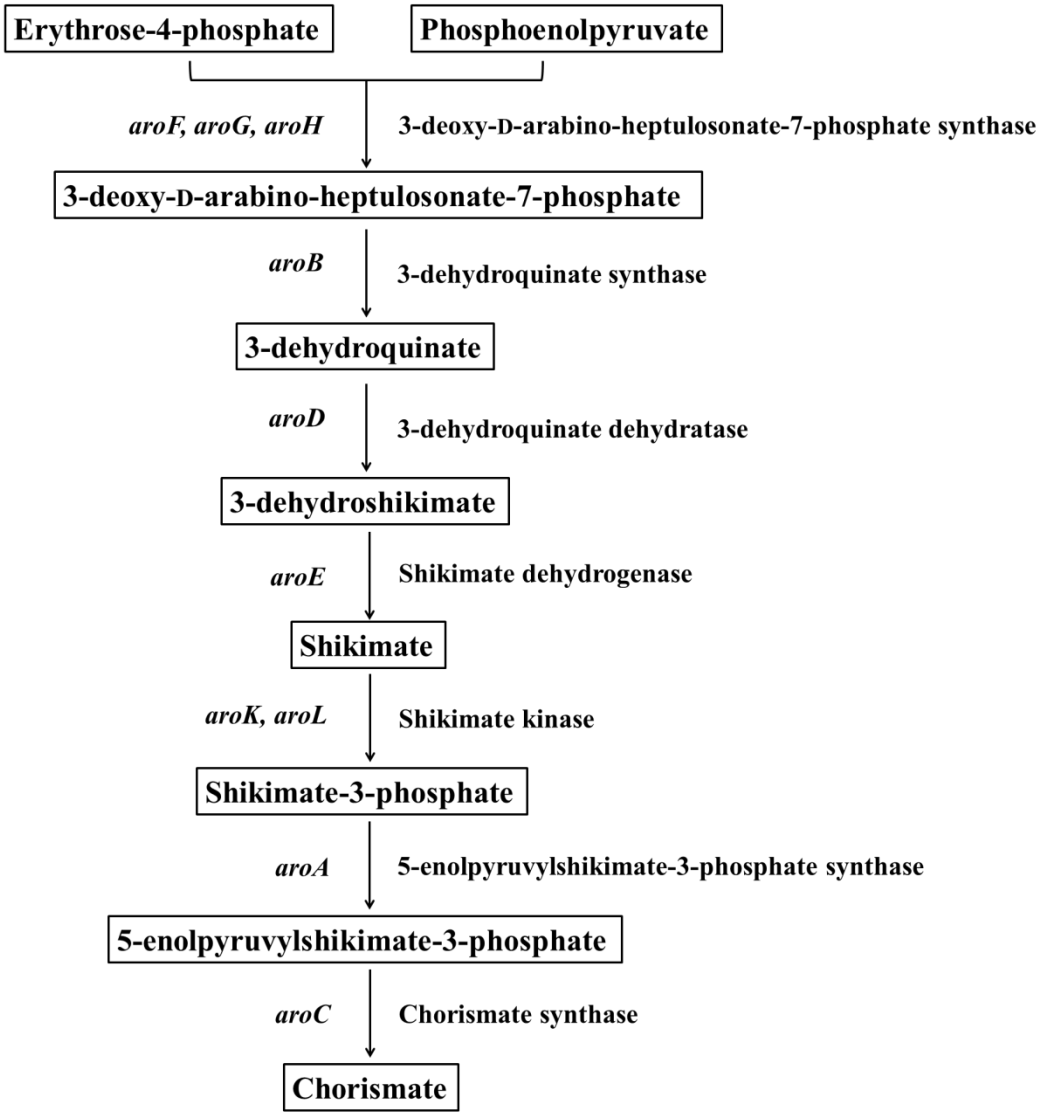


Figure 1.2: Flow diagram showing the steps involved in the shikimate pathway

1.3.2 Mechanism of AroA

Huge interest has been directed to the study of the shikimate pathway and its enzymes because the pathway is absent in mammals and thus make an attractive target for the development of new antimicrobials (McConkey 1999; McDevitt et al. 2002; Amer et al. 2008).

AroA, the sixth enzyme of the shikimate pathway, catalyses the reversible transfer of the intact enolpyruvyl moiety of PEP to the 5′ hydroxyl group of shikimate-3-phosphate (S3P) to produce 5-enolpyruvylshikimate-3-phosphate (EPSP) and inorganic phosphate (Lewendon & Coggins 1983; Steinrücken & Amrhein 1984a; Schönbrunn et al. 2001; Eschenburg et al. 2003) (Figure 1.1). Unlike typical PEP-utilizing enzymes, AroA catalyses the reaction by the cleavage of the C-O bond of PEP instead of the high energy P-O bond (Schönbrunn et al. 2001; Eschenburg et al. 2003).

A reversible addition-elimination reaction was put forward in 1964 (Levin & Sprinson 1964). The nucleophilic attack of the 5′ hydroxyl group of shikimate-3-phosphate on the C2 position of PEP forms a non-covalent tetrahedral intermediate at the active site of the enzyme. This mechanism was later unequivocally confirmed using NMR (Anderson et al. 1988a). The mechanism catalysed by AroA *via* its intermediate was further investigated by chemically synthesising both diastereomers of the tertrahedral intermediate, (*R*)-TI and (*S*)-TI (An et al. 2003). The conversion of (*S*)-TI to EPSP, PEP and S3P conclusively identified the tetrahedral intermediate as the (*S*)-ketal phosphate, thereby confirming the catalytic mechanism proceeds *via* an *anti* addition followed by a *syn* elimination (An et al. 2003).

The role of the catalytic residues in the addition-elimination steps of the reaction was examined by a novel approach called partitioning analysis (Mizyed et al. 2003). It was found that many amino acid residues contribute to catalysis, however no single residue is responsible for only addition or elimination. In the addition step of the overall reaction, the amino acid residue Lys22 acts as a general base catalyst and Glu341 acts as a general acid catalyst. The same residues act as general acid or general base catalysts of the tetrahedral intermediate breakdown in both the forward and reverse reactions (Mizyed et al. 2003). The identification of Lys22 and Glu341 as general acid/base catalytic residues is backed up by a later study performed by Berti and Chindemi (2009). In addition to the acid/base catalysis role, they also assigned Glu341 to have an electrostatic catalysis role. This is because Glu341 and Asp313 stabilize the cationic intermediates/transition states by forming an electrostatic sandwich around the positive charge at C2. Mutations of Glu341 and Asp313 greatly affect the k_{cat} thus confirming their importance as catalytic residues whereas mutation of

Lys22 mainly affects substrate binding (*E.coli* numbering) (Berti & Chindemi 2009). The position of Lys22, Asp313 and Glu341 within the active site is shown in Figure 1.3.

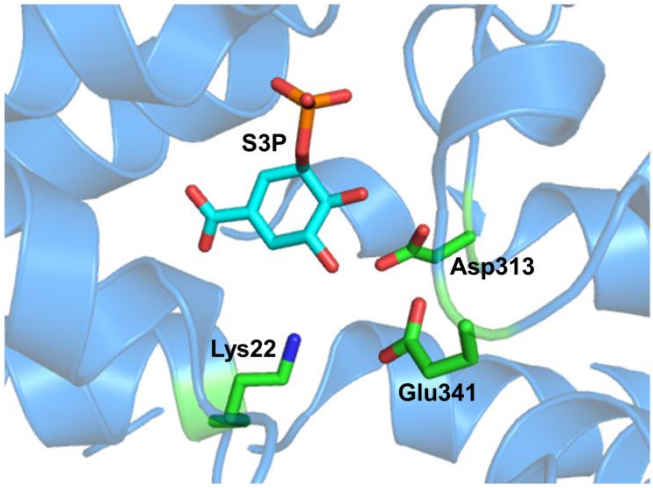


Figure 1.3: Positioning of catalytic residues Lys22, Asp313 and Glu341 in the active site of AroA enzyme
The position of catalytic residues Lys22, Asp313 and Glu341 from *E.coli* AroA (PDB 1G6T) are shown in relation to the AroA substrate shikimate-3-phosphate (S3P) within the active site.

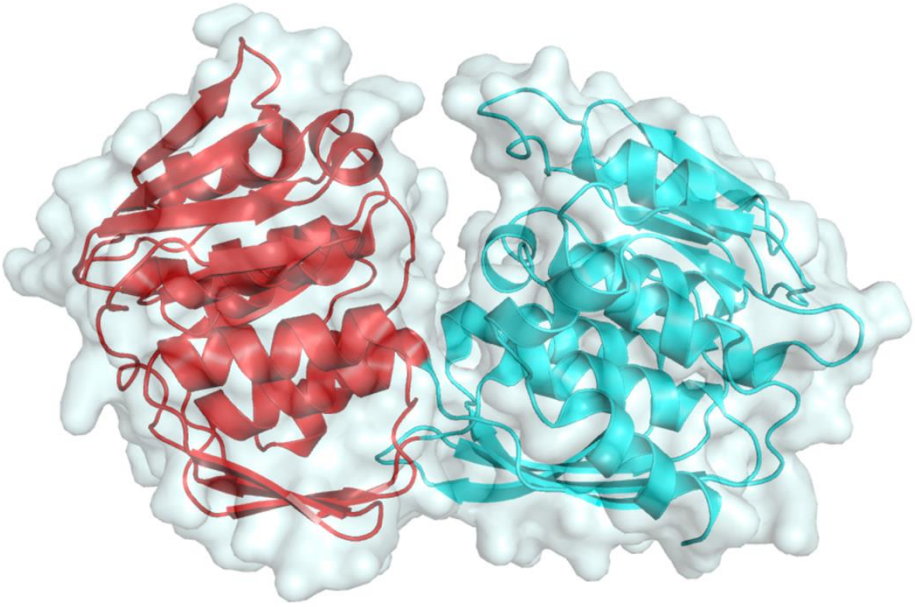
Steady-state kinetics of *B. subtilis* AroA showed allosteric behaviour (Majumder et al. 1995). Kinetic analyses performed by site-directed mutagenesis of Pro105 and His385 suggest that PEP has a second binding site which is non catalytic. It was then demonstrated that the non-catalytic PEP binding resides at His385 (Shuttleworth & Evans 1994). It was also observed that *B. subtilis* AroA could be activated using ammonia. Mutations at Arg24 and His385 affected the ammonium interaction sites as they nearly eliminated the ammonia activation (Majumder et al. 1995).

Therefore, AroA enzymes from different species have specific characteristics and their mechanism and catalytic activity may be controlled differently in different species.

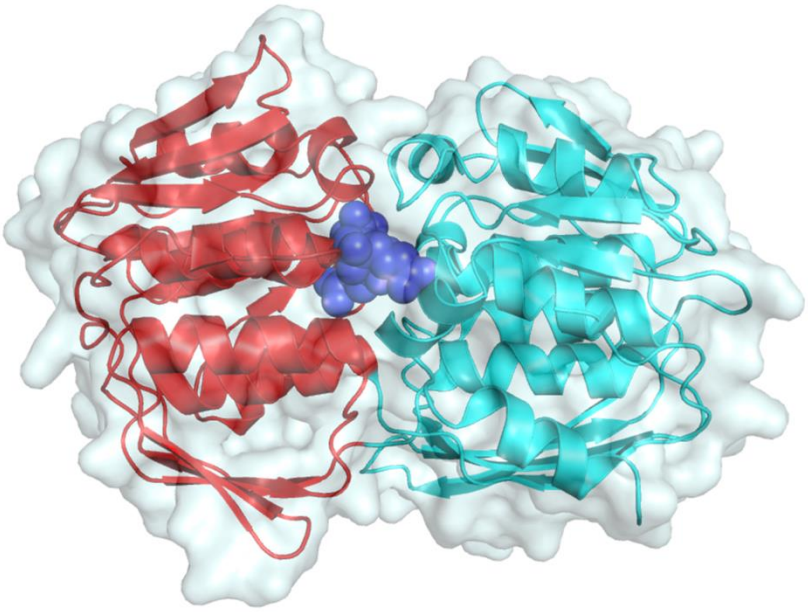
1.3.3 Structure

The polypeptide chain of AroA folds into two globular domains connected by a short hinge region. The mainchain fold of each domain is very similar with each other consisting of three repeats of βαβαββ motif (Skarzynski et al. 1996; Eschenburg et al. 2003; Funke 2008). Upon binding of the AroA substrate S3P, the enzyme undergoes a conformational change from an ‘open’ to ‘closed’

conformation resulting from the two domains approaching each other in a screw-like movement (Figure 1.4) (Schönbrunn et al. 2001). This leads to the formation of the active site in the interdomain cleft where the second substrate PEP and the inhibitor glyphosate bind (Eschenburg et al. 2003; Funke et al. 2009).



A)



B)

Figure 1.4: *Streptococcus pneumoniae* AroA in its open and closed conformations
A) *S. pneumoniae* AroA (PDB 1RF5) in its unliganded open conformation. B) *S. pneumoniae* AroA (PDB 1RF6) bound to S3P and glyphosate in a closed conformation. The figure shows the cartoon representation of the N-terminal domain (domain I) (red) and C-terminal domain (domain II) (blue). The protein surface (transparent light blue) in A) clearly shows the gap between the N- and C-terminal domains in open conformation in comparison to closed. S3P and glyphosate complex is represented with spheres in dark blue.

The structure of the free enzyme AroA, in the 'open' conformation, from *E. coli* was reported by Stallings et al. (1991). The structure consisted of two remarkably similar globular domains with 131 residues out of 208 in each domain equivalent to the other. In addition to this overall twofold topological symmetry, the structure also showed an additional threefold axis of symmetry in each domain. Thus each domain can further be divided into three more subunits with a total of six subunits in the entire structure (Figure 1.5). Out of these six subunits, even though only three subunits are formed from a continuous polypeptide chain, there is a possibility that the structure may have evolved from a sixfold replication of the gene responsible for the folding unit (Stallings et al. 1991).

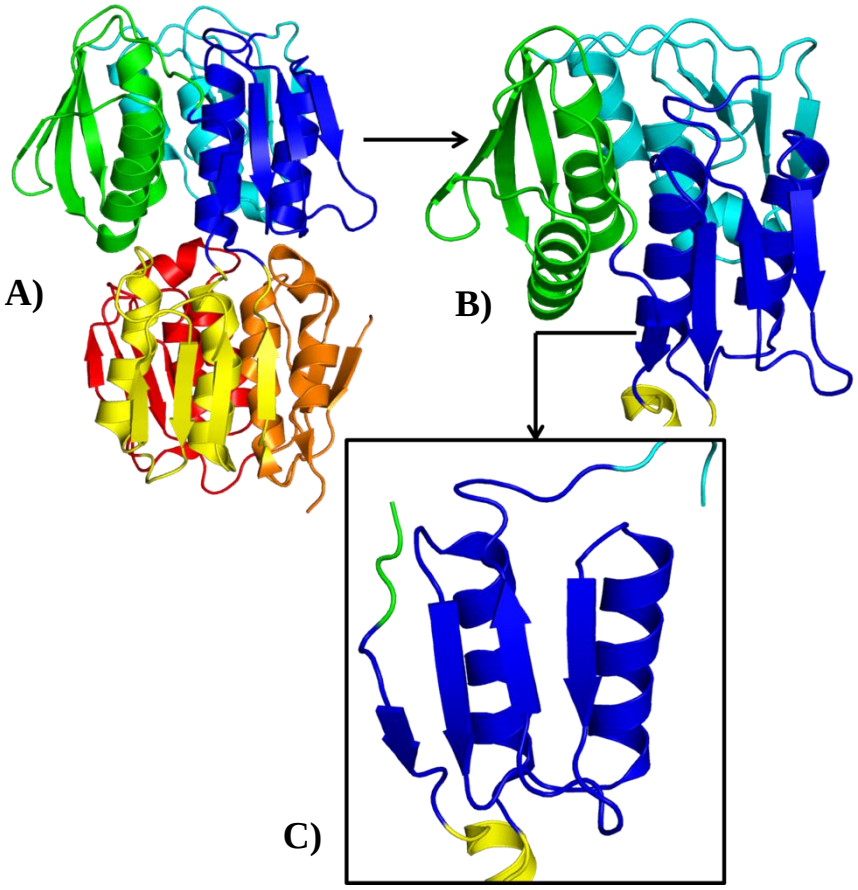


Figure 1.5: Cartoon representation of the two globular domains and individual subunits constituting the AroA enzyme
A) The full AroA enzyme from *E. coli* (PDB 1G6T) showing its two domains, the N-terminal domain (domain I) and C-terminal domain (domain II). The colour spectrum from dark blue to red goes from N-terminal to C-terminal. The structure consists of a total of six subunits, each subunit coloured differently. B) The N-terminal domain consisting of three individual subunits. C) The individual subunit consisting of two α helices and four β strands. The subunit shown is formed from discontinuous polypeptide chain. The whole AroA structure consists of a six fold repetition of this βαβαββ motif.

Stallings et al. (1991) proposed the "mushroom button" model for the overall structure of AroA with parallel interacting helices acting as the stem and the extensive domain surface covered with β sheets as the cap of the mushroom. This model is also referred to as the 'inside-out α/β-barrel' (Stallings et al. 1991) (Figure 1.6).

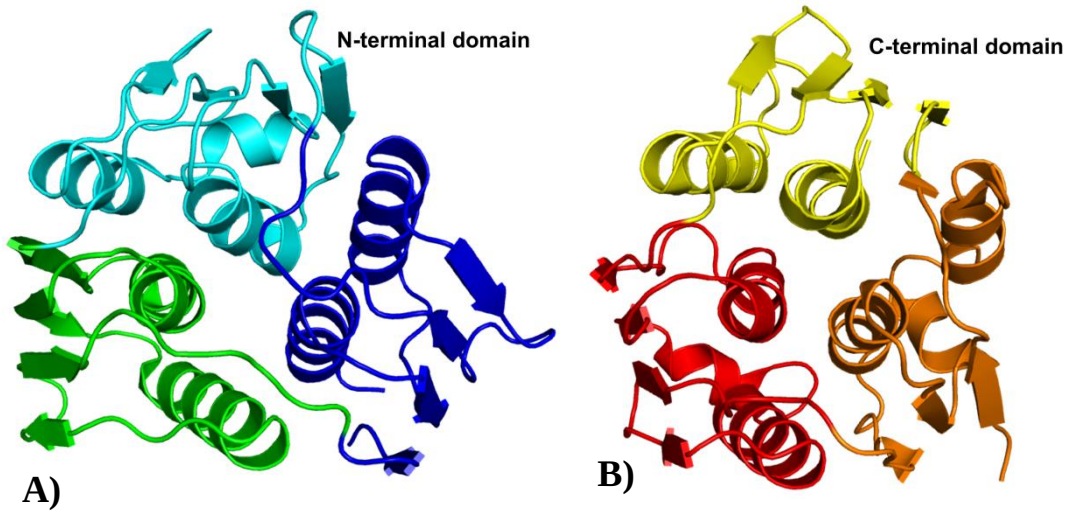


Figure 1.6: Cartoon representation of "inside-out α/β barrel" configuration of AroA enzymes
The two domains of *E. coli* (PDB 1G6T) A) N-terminal domain (domain I) and B) C-terminal domain (domain II).

Homology among the bacterial, fungal and plant AroA sequences has been observed (Gasser et al. 1988) and this suggests the possibility that the three dimensional structure of all these AroAs may be similar to that of *E. coli* AroA. Structural information for AroA from a different species, *S. pneumoniae*, showed similar topology with six βαβαββ motifs, in spite of only having 25% amino acid sequence identity with AroA from *E. coli.* It also has approaching domains for ligand binding and local structural changes of certain highly flexible regions from disordered to ordered helical structure upon binding of the ligand (Park et al. 2004).

The structure of AroA from *S. pneumoniae* showed minor differences when compared to *E. coli* AroA with the presence of an additional short helix (3_{10}-helix) in certain motifs. A significant decrease in the B-factor of the C-terminal domain in the closed conformation was observed in comparison to the open state, as a consequence of ligand binding which causes a highly flexible region (331-342 in *S. pneumoniae* and 406-415 in *E. coli*) to transition to an ordered helical structure.

Even though the overall domain closure mechanism is similar for both *S. pneumoniae* and *E. coli*, they differ in rotation angles and axis angles. Salt bridges and hydrogen bonds are responsible for the stabilization of the interdomain interaction (Park et al. 2004).

Therefore, even though the overall structure of AroA from all organisms may be expected to be similar, each structure shows distinct variations.

Substrate binding and reaction mechanism:

The interdomain cleft, formed by the two domains of AroA, harbours the amino acid residues responsible for ligand binding (Park et al. 2004). It was proposed that a macrodipolar effect was responsible for attracting the approaching anionic ligands such as S3P, PEP, reaction products, tetrahedral intermediate and the inhibitor glyphosate, and guiding them to an active site near a highly positively charged interdomain pocket (Stallings et al. 1991). Upon comparison of unliganded AroA with liganded AroA, Schönbrunn et al. (2001) explained that positive charge accumulates at the cleft as a result of the conformational change from 'open' to 'closed' states initiated by S3P binding.

Glyphosate is known to occupy the same binding site as PEP at the active site. The interaction of the charged residues from the two domains of *E. coli*, Lys22, Arg124, and Lys411, with glyphosate is very likely to be similar to that of PEP. Conservative residues from MurA which interact with PEP (Lys22, Arg120, Asp305, Arg331, Arg371, Arg397) were also found to correspond to the residues present in the glyphosate binding site of AroA (Lys22, Arg124, Asp313, Arg344, Arg386, Lys411) (Schönbrunn et al. 2001). The structures of *S. pneumoniae* AroA in complex with S3P and glyphosate were compared with AroA in its tetrahedral intermediate form in complex with S3P and F-PEP (pseudo substrate (Z)-3-fluoro-PEP for the formation of stable tetrahedral intermediate), agreeing that PEP and glyphosate share a common binding site (Park et al. 2004). The PEP/glyphosate binding site residues common to *E. coli* and *S. pneumoniae* AroA are shown in Figure 1.7.

Analysis of the PEP binding site provides more information about the residues involved in the catalytic reaction of AroA. One such residue, Glu341, was suggested to act as the proton donor for the methylene group of PEP and involved in the stabilisation of PEP oxocarbenium ion. The interactions of the backbone nitrogen of Glu341 and interactions of the Lys411 and His385 sidechains, holds

the side chain of Glu341 in place. His385 may act as a proton source for Glu341, shifting up the pKa for Glu341. We already know that the reaction mechanism involves the deprotonation of S3P 5′ hydroxyl group to attack C2 of PEP. The structure shows that the S3P 5′ hydroxyl group interacts with Asp313 which hence may act as a proton acceptor (Schönbrunn et al. 2001).

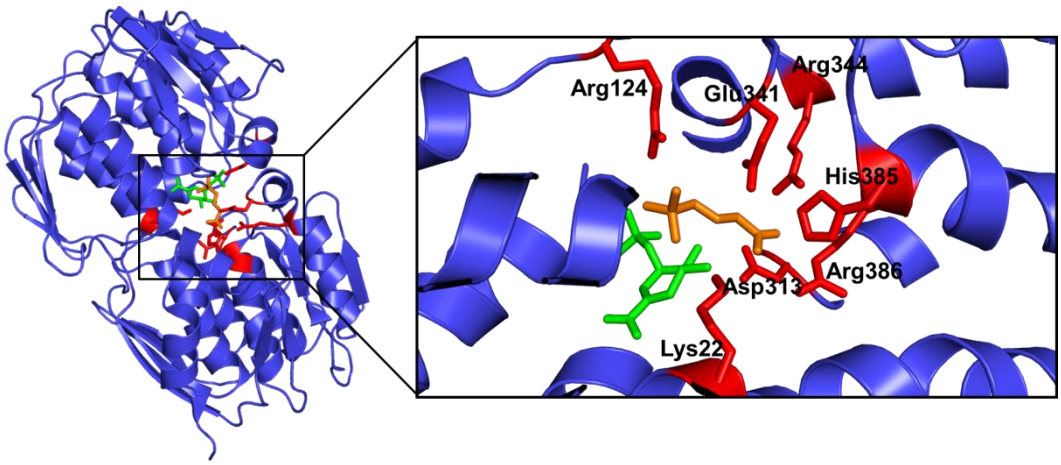


Figure 1.7: The residues present in the PEP/glyphosate binding site of AroA enzyme
The residues (red) identified at the PEP/glyphosate binding site, common to *E.coli* AroA and *S. pneumoniae* AroA are shown on *E. coli* AroA structure (PDB1G6S) (blue). Substrate S3P (green);inhibitor glyphosate (orange).

The ligand binding process adopts both the lock and key model and an induced fit model (Park et al. 2004). The binding of S3P follows a lock and key model as the binding site remains unchanged irrespective of the protein conformation (Park et al. 2004). The N-terminal domain of the structure possesses the S3P binding site. NMR spectroscopy and isothermal titration calorimetry identified a stable conformational change of the isolated N-terminal domain upon binding with S3P (Stauffer et al. 2001). The residues responsible for binding PEP or glyphosate are equally distributed over the C- and N-terminal domain (Stauffer et al. 2001). Upon binding of S3P, a conformational change of the protein brings the two domains together, where parts of the binding site are reorganized to accommodate the binding of PEP or glyphosate *via* the induced-fit model (Park et al. 2004).

1.3.4 Inhibition by glyphosate

AroA is inhibited by the broad-spectrum herbicide glyphosate. The inhibition of AroA by glyphosate is competitive versus PEP, and noncompetitive versus S3P (Sun et al. 2005). AroA enzymes can be categorised into two classes based on their glyphosate sensitivity. Class I enzymes (e.g. from *E. coli*) are naturally

sensitive to glyphosate. The generation of class I mutants with tolerance to glyphosate causes a decrease in the affinity of AroA for PEP. Class II AroA enzymes (e.g. from *B. subtilis*) have a high affinity for PEP and are naturally tolerant to glyphosate (Sun et al. 2005).

Glyphosate is the main ingredient of the broad-spectrum herbicide Roundup (Steinrücken & Amrhein 1984b; Priestman et al. 2005b). The proposed ordered mechanism for glyphosate binding is that S3P has to bind to the enzyme first in order for glyphosate to bind. This is supported by the fact that upon mutation of Arg27, a residue that is important in the binding of S3P, to alanine, AroA was not able to bind either S3P or glyphosate (Anderson et al. 1988b; Schönbrunn et al. 2001). It was determined that the AroA structure changes its conformation from an open to a closed state due to the binding of S3P and that glyphosate is not responsible for the structural conformational change (Schönbrunn et al. 2001). Lys22, Arg124 and Lys411 are residues implicated in PEP binding. A comparison of the interaction of the carboxyl moieties of PEP and glyphosate with Lys22, Arg124 and Lys411, and a comparison of the PEP binding site of MurA with the glyphosate binding region of AroA, provides evidence that glyphosate binds at the same binding site as PEP (Schönbrunn et al. 2001).

The determination of the mechanism and residues responsible for glyphosate resistance is important, especially in cases where we want to achieve inhibition. One such modification which gave glyphosate resistance, but also maintained a desirable substrate affinity, was the double mutation of T97I/P101S in *E. coli* AroA (Funke et al. 2009). Similarly, substitution of Gly96 with alanine (numbering corresponding to *E. coli*), was found to confer glyphosate tolerance. Mutation at Pro101 made the enzymes less sensitive towards glyphosate (Padgette et al. 1991; Priestman et al. 2005b). It was also found that the use of shikimate as an alternative substrate for AroA induces a conformational change that does not allow efficient binding of glyphosate thus making it insensitive to glyphosate (Priestman et al. 2005b).

Based on the sensitivity towards glyphosate, AroA enzymes are therefore classified as Class I and Class II. Class I and Class II AroA enzymes share less than 30% amino acid identity (Sun et al. 2005; Funke et al. 2009; Tian et al. 2013). Class I AroA are found in plants and bacteria, including *Escherichia coli, Salmonella typhimurium, Aeromonas salmonicida, Bordetella pertussis, Petunia*

petunia, and *Arabidopsis thaliana* (Tian et al. 2012; Zhang et al. 2014). Class II AroA have been identified in *Pseudomonas* strain PG2982, *Agrobacterium tumefaciens* strain CP4, *Streptococcus pneumoniae*, *Staphylococcus aureus* (Sun et al. 2005; Funke et al. 2009; Tian et al. 2013), *Bacillus cereus* (Tian et al. 2013), *Alcanivorax* sp. L27 (Zhang et al. 2014), *Halothermothrix orenii* H168 (Tian et al. 2012) and *Bacillus subtilis* (Priestman et al. 2005a).

1.4 MurA (UDP-*N*-acetylglucosamine enolpyruvyl transferase)

1.4.1 Pathway

Peptidoglycan (also called murein), is a polymer consisting of sugars and amino acids which forms a mesh-like layer outside the plasma membrane to form the cell wall. Its biosynthesis is a complex two-stage process. The first stage occurs in the cytoplasm and involves a series of Mur enzymes for the formation of the building block UDP-*N*-acetylmuramyl pentapeptide (Figure 1.8). This part of the pathway begins with the transfer of an enolpyruvate moiety of PEP to the 3′ hydroxyl group of UDP-*N*-acetylglucosamine, catalyzed by MurA. The enolpyruvate moiety is then reduced to D-lactate by MurB to give UDP-*N*-acetylmuramate. A stepwise addition of amino acids by MurC to MurF forms UDP-*N*-acetylmuramyl pentapeptide (Molina-López et al. 2006; Yoon et al. 2008).

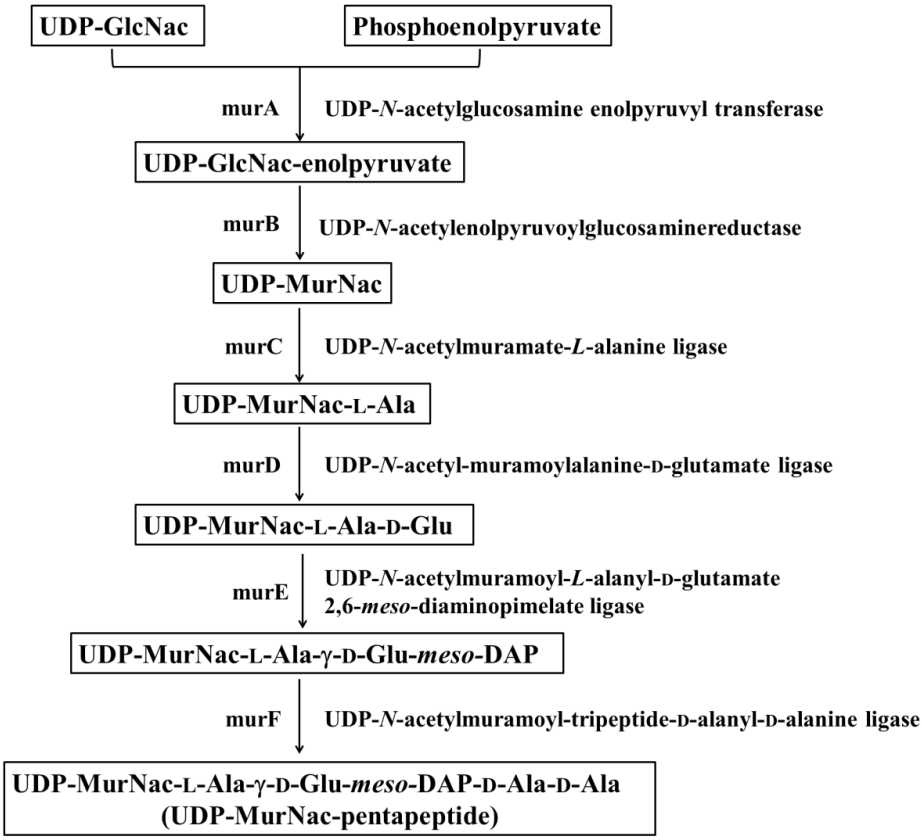


Figure 1.8: Flow diagram showing the cytoplasmic steps of the peptidoglycan biosynthesis pathway involved in the formation of UDP-*N*-acetylmuramyl pentapeptide
UDP-GlcNac: Uridine diphosphate-*N*-acetylglucosamine; UDP-MurNac: UDP-*N*-acetylmuramic acid; DAP: diaminopimelic acid.

The absence of the peptidoglycan biosynthesis pathway in mammals makes it a very good target for the development of antibacterial agents (Du et al. 2000; Eschenburg et al. 2005b). The inhibition of any of the enzymes MurA to MurF causes the bacterial cells to loose shape and integrity, and ultimately to undergo cell lysis and death (Molina-López et al. 2006)

1.4.2 Mechanism of MurA

MurA catalyses the first committed step of peptidoglycan biosynthesis which involves the transfer of an enolpyruvate residue from PEP to UDP-*N*-acetylglucosamine (Figure 1.1)(El Zoeiby et al. 2003). The MurA reaction proceeds *via* an addition-elimination reaction (Marquardt et al. 1993) in which a tetrahedral intermediate is formed by the direct attack of 3′-OH of UDP-*N*-acetylglucosamine on the C-2 position of PEP, followed by the elimination of inorganic phosphate to produce UDP-*N*-acetylenolpyruvylglucosamine (Skarzynski et al. 1996). Early studies on the mechanism of MurA suggested

many possible intermediates. One such study suggested a mechanism which involves the formation of two intermediates, with a phospholactoyl-enzyme tetrahedral intermediate preceding a phospholactoyl-UNAG intermediate (Brown et al. 1994).

It is now known that the reaction starts by the binding of UNAG followed by PEP in the active site of the enzyme. A proton is transferred to PEP forming an oxocarbenium ion which is then joined to UNAG. This gives rise to the tetrahedral intermediate with both substrates covalently attached to each other. (Gautam et al. 2011). Here the C2 of the PEP moiety assumes the tetrahedral configuration whereas the C3 becomes a methyl group. A proton is then removed from C3 and inorganic phosphate is released to yield the vinyl ether product, thus re-establishing the double bond between C2 and C3 (Walsh et al. 1996; Eschenburg et al. 2003). Unlike most PEP utilizing enzymes, MurA cleaves the C-O bond of PEP to achieve the transfer of the enolpyruvyl moiety instead of the high energy P-O bond (Walsh et al. 1996; Eschenburg et al. 2003).

Many studies have observed that the antibiotic fosfomycin, considered as an analog of PEP, irreversibly inhibits MurA through its covalent bonding to a cysteine residue of the enzyme. Therefore it was believed that PEP may bind to this cysteine residue to form a stable intermediate as part of the enzymatic reaction (Cassidy & Kahan 1973). This cysteine residue was identified to be at position 115 from the amino acid sequence of MurA that had been cloned and overexpressed from *Enterobacter cloacae* (Wanke et al. 1992). To prove that Cys115 is essential for MurA enzyme activity, an experiment was conducted where all the cysteine residues with the exception of one, were changed to serine by site-directed mutagenesis. Out of the mutants obtained, the only mutant resulting in an inactive enzyme was C115S, thus demonstrating the essential role of Cys115 in MurA enzyme activity (Wanke & Amrhein 1993). Another variant of MurA, C115A, expressed in *E. coli* at normal levels, however it did not show any enzyme activity. Such evidence obtained from mutagenesis studies, along with a number of NMR studies, strongly indicates that Cys115 is the active site nucleophile responsible for the formation of a covalent phospholactyl-enzyme adduct derived from the substrate PEP (Brown et al. 1994). Contrary to the above studies, site directed mutagenesis of Cys115 to aspartate (C115D) was performed in *E. coli* and resulted in a highly active MurA enzyme, in addition to full

resistance to its inhibitor fosfomycin (Kim et al. 1996a). Comparing the studies that give inactivity upon C115S and C115A mutations and the active C115D suggests a role for the carboxylate side chain of aspartate in catalytic function. The kinetics, thermodynamics and stereochemistry of the reaction of wild type MurA has been previously studied in the presence of the pseudosubstrate (*Z*)-FPEP (Kim et al. 1994; Kim et al. 1995a; Kim et al. 1995b). The results from studies on the C115D mutant suggest that Cys115 functions as an active site acid involved in the initial protonation of PEP as well as the active site base in the elimination step. It was also observed that the difference in the nucleophilicity between the carboxylate side chain of Asp115 and the thiolate group of Cys115 suggests that the formation of a covalent adduct is not required for catalytic activity (Kim et al. 1996a).

1.4.3 Inhibition by fosfomycin

Peptidoglycan is essential for bacterial cell wall synthesis and thus any inhibitor of the enzymes involved in the peptidoglycan synthesis would make good antibiotic candidates. Out of the many enzymes involved in the peptidoglycan pathway, MurA is the only one which is inhibited by a clinically validated antibiotic: fosfomycin. Fosfomycin was first isolated from a *Streptomyces* species (Christensen et al. 1969; Hendlin et al. 1969). It was found that fosfomycin was bactericidal in action by affecting the integrity of the bacterial cell wall. Initially its oral administration to mice proved to be effective against Gram negative and Gram positive organisms. Its oral and intravenous administration in humans proved to be effective and non-toxic upon clinical and pharmacological trials marking its potential as a broad spectrum antibiotic (Hendlin et al. 1969). The antimicrobial action of fosfomycin, particularly against uropathogenic strains, and its notable tolerability makes it effective for treating urinary tract infections (UTIs). UTIs are the most common bacterial infection in women and the second most common reason for antibiotic prescription in the world (Fadda et al. 2006; Keating 2013).

Fosfomycin mimics and competes with the MurA substrate PEP in a stereo-specific manner covalently and irreversibly *via* formation of a thioether bond to Cys115 (Skarzynski et al. 1996; Fadda et al. 2006). Near this bond, a pocket present between fosfomycin and Arg331 and Arg371 is strongly believed to be the PEP binding site (Figure 1.9) because these two arginines are conserved in

structural alignments of AroA enzymes and MurA enzymes. Fosfomycin forms various bonds with the enzyme and UNAG. The phosphonate oxygen atom of fosfomycin forms hydrogen bonds with positively charged residues of MurA (Lys22, Arg120, Arg397) and with the amide nitrogen of UNAG (Figure 1.9). The fosfomycin hydroxyl group also forms hydrogen bonds with the C3 hydroxyl of the UNAG sugar ring. These hydrogen bonds ensure the tight packing of fosfomycin in between the enzyme and UNAG (Skarzynski et al. 1996). The presence of UNAG was also found to accelerate the time-dependent inactivation of MurA by fosfomycin. The enzyme was tested in the presence of the unreactive UNAG analog, 3-deoxy-UNAG, to see whether the UNAG was participating in the chemistry of the reaction to cause inactivation. Both UNAG and 3-deoxy-UNAG caused acceleration of the inactivation process, although not to the same extant. This suggests that UNAG may not participate in the reaction, but instead, it may influence the conformation of the active site necessary for the inactivation event (Marquardt et al. 1994).

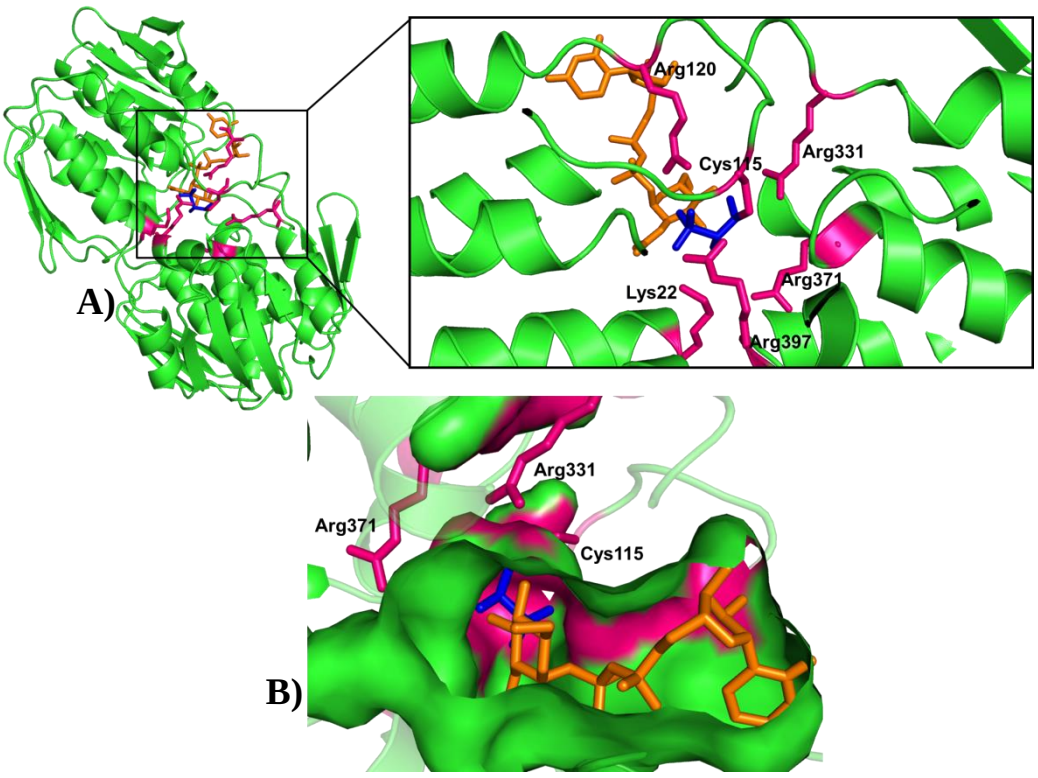


Figure 1.9: The fosfomycin and PEP binding site of MurA enzyme
A) *E. coli* MurA monomer (green) with UNAG (orange) and fosfomycin (blue) bound to it (PDB 1UAE). The structure shows the active site with residues involved in interaction with fosfomycin (magenta). B) The pocket that is present between fosfomycin and residues Arg331 and Arg371 believed to be the PEP binding site.

One of the major concerns with the use of fosfomycin is the rising development of fosfomycin resistance. Fosfomycin resistance can be obtained in two ways: chromosome alterations in the strains, or encoded by plasmids. The strains which are chromosomally altered to gain fosfomycin resistance can have impairment in fosfomycin uptake, low affinity MurA or overproduction of the enzyme (El Zoeiby et al. 2003). Fosfomycin resistance genes can also be aquired through plasmids, however this constitutes only a low percentage of fosfomycin-resistant strains (Arca et al. 1997; Castañeda-García et al. 2013). Various mutations can also lead to fosfomycin resistance. *Mycobacterium tuberculosis* is naturally resistant to fosfomycin. A closer look at the MurA amino acid sequence from *Mycobacterium tuberculosis* shows that the site corresponding to C115 of *Escherichia coli* is replaced by an aspartate residue in *M. tuberculosis* (D117). The replacement of D117 with cysteine reversed the resistance making the strain sensitive to fosfomycin (De Smet et al. 1999). A similar result was observed in *Escherichia coli*, where C115D generated an active MurA enzyme that was completely resistant to fosfomycin. The mutation of C115 to glutamate (C115E) was also resistant to fosfomycin, however much less active in comparison to C115D (Kim et al. 1996a).

The rise of fosfomycin resistance has been a motivation to find new inhibitors for MurA. The knowledge on the mode of inhibition by fosfomycin and its interaction with the enzyme and substrate will help the development of new antibiotics against MurA.

1.4.4 Structure

The overall structure of MurA comprises two very similar globular domains linked *via* two strands of four amino acids each. The structure is commonly described as an 'inside-out α/β barrel and has six repetitive βαβαββ motifs with three in each domain. Each domain consists of six helices and three four stranded β sheets (Schönbrunn et al. 1996; Skarzynski et al. 1996; Eschenburg et al. 2003; Oberdorfer et al. 2012). Out of the six-fold subdomain repetition of the MurA structure, the only repetitive element in the amino acid sequence is a short Leu-X_3-Gly(Ala) motif (Marquardt et al. 1992). This motif is responsible for the attachment of the folding units *via* hydrogen bond interactions forming three hydrophobic pockets per domain. The two domains are connected *via* a hinge region with high degree of rigidity and a salt bridge (Schönbrunn et al. 1996). The

enzyme is also known to be in an 'open' conformation in the unliganded state and acquires a 'closed' conformation upon ligand binding with the active site situated at the interdomain cleft (Skarzynski et al. 1996) (Figure 1.10).

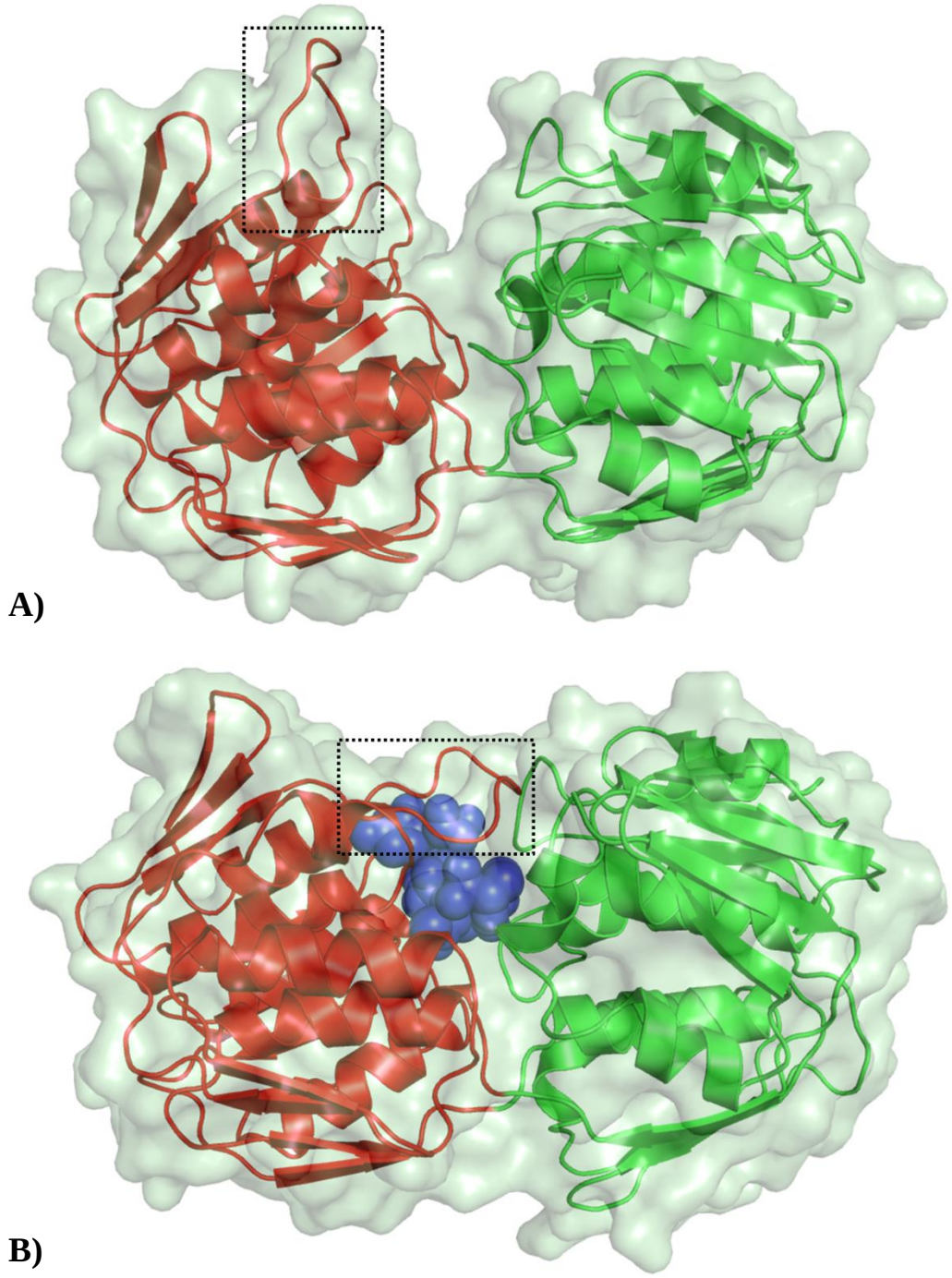


Figure 1.10: Cartoon representation of MurA in its open and closed forms
A) *E. cloacae* MurA (PDB: 1NAW) in its unliganded 'open' conformation and B) *E. coli* MurA (PDB: 1UAE) bound to UNAG and fosfomycin in a 'closed' conformation. N-terminal domain (red); C-terminal domain (green); UNAG and fosfomycin complex shown in spheres (blue). The surface represented with transparency clearly shows the distinction between open and closed conformations. The dashed box shows the 'lid'-like loop over the active site.

The open conformation of the unliganded MurA has been studied from *Enterobacter cloacae.* A network of water molecules has been identified in the interdomain cleft which is believed to be important for the catalytic mechanism of the enzyme (Eschenburg & Schönbrunn 2000). These water molecules stabilise the catalytic residues and are displaced upon ligand binding. The open conformation of the enzyme can be described as a resting stage for the enzyme in the absence of substrates and it was observed that the open state did not have a specific phosphate binding site suggesting that this site may arise only in the closed conformation (Eschenburg & Schönbrunn 2000). The open conformation of the enzyme is believed to be stabilised through the repulsive forces between lysine-arginine residue pair (Arg397 and Lys48) in the interdomain cleft. The repulsive force between the two domains is neutralized upon UNAG binding thereby allowing the two domains to move towards each other (Schönbrunn et al. 1996).

The structure of MurA from *Escherichia coli* was successfully complexed with its substrate UNAG and inhibitor fosfomycin by Skarzynski et al. (1996). The MurA substrate UNAG comprises two parts: one uridinyl ring and an N-acetylglucosamine moiety. The specific interactions and the hydrogen bonds involved in the binding of the uridinyl ring in between two hydrophobic surfaces provide it with a good fit and specificity. Whereas, the N-acetyglucosamine moiety interacts with a number of water molecules present in the active site pocket, and makes three hydrogen bonds involving the Asn23 and Asp305 residues of the protein (Skarzynski et al. 1996) (Figure 1.11). The residues Asn23 and Asp305 were found to be essential active site residues with Asp305 being essential for binding of UNAG and Asn23 involved in the stabilization of the transition states (Samland et al. 2001). Fosfomycin was found to be packed in between the enzyme and UNAG.

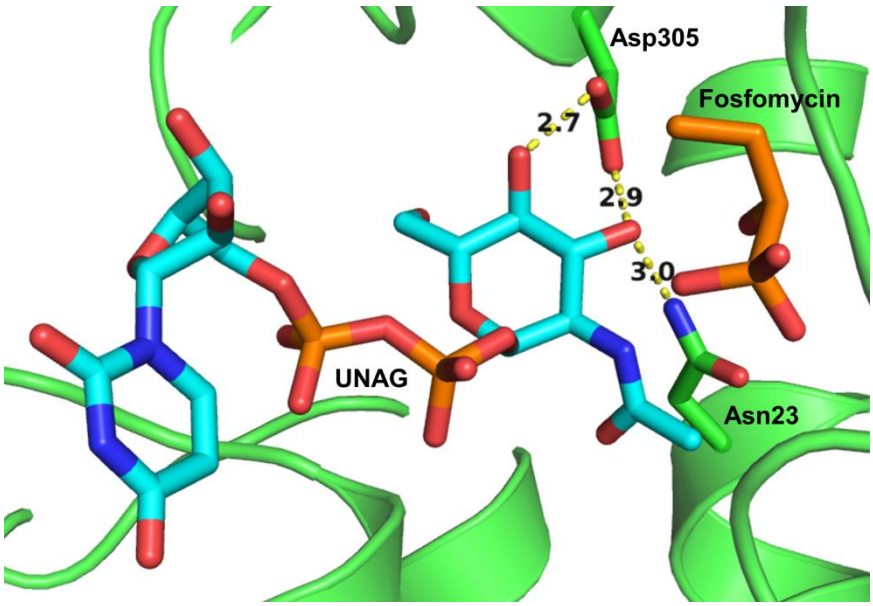


Figure 1.11: Essential MurA active site residues interacting with UNAG
The figure shows the interaction of UNAG (cyan) with the essential active site residues Asn23 and Asp305 in *E. coli* MurA structure (green)(PDB 1UAE). The position of fosfomycin (orange) with respect to these residues and UNAG are also shown. The bond distances are shown in yellow dashes. Colour key for atoms: red-oxygen; dark blue-nitrogen.

It was proposed that the enzyme recognises the substrate UNAG and its binding induces a conformational change of the domains from an 'open' to 'closed' states accompanied with a conformational change of the highly flexible loop 111-121 which allows the required positioning of the residue C115 to react with PEP or fosfomycin. Ten direct interdomain hydrogen bonds were identified which stabilise the closed conformation of the enzyme (Schönbrunn et al. 1996; Skarzynski et al. 1996). It is also observed that in a 'closed' conformation, a specific shielding of regions and unspecific overall protection associated with reduced flexibility, protects the enzyme towards proteolytic degradation (Krekel et al. 1999). Two very similar structures of MurA from *Haemophilus influenzae*, a binary complex with UNAG and a ternary complex with UNAG and fosfomycin, however, display a half-open conformation for the active site loop (Yoon et al. 2008). Therefore, the active site loop potentially adopts three conformations: the fully open, the half-open and the closed conformation (Figure 1.12).

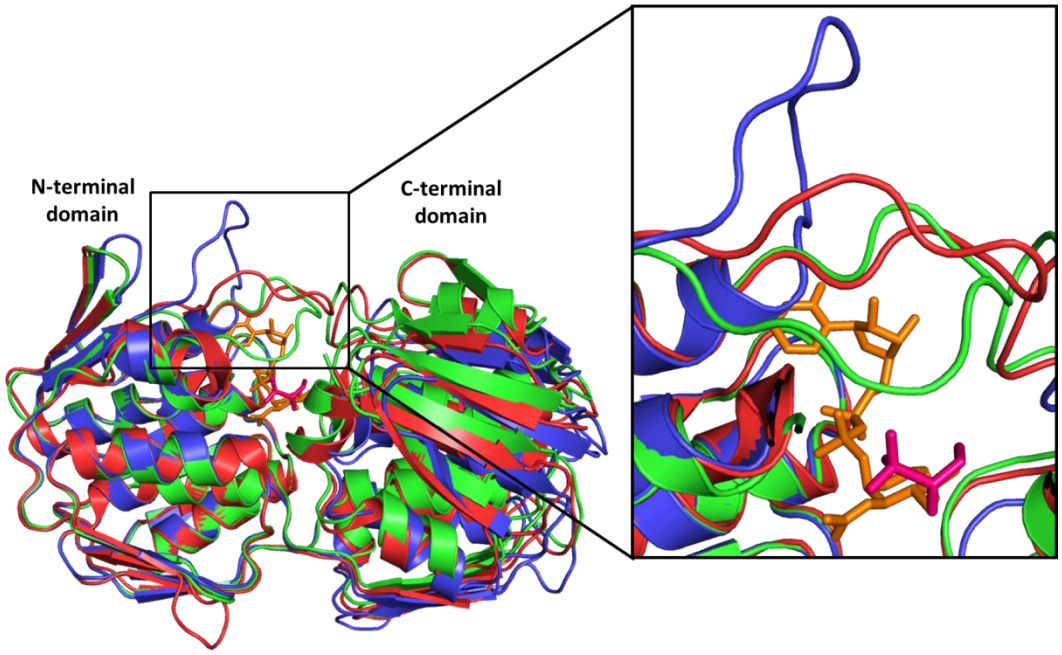


Figure 1.12: Cartoon representation of the three conformations adopted by the active site loop of MurA
The N-terminal domain of the three MurA structures was superimposed to show the different conformations. The position of UNAG (orange) and fosfomycin (magenta) are shown in the active site. The loop of unliganded *E. cloacae* MurA (PDB 1NAW) (blue) adopts an open conformation, the loop of liganded *Haemophilus influenza* MurA (PDB 2RL2) (red) adopts a half-open conformation and the loop of liganded *E. coli* MurA (PDB 1UAE) (green) adopts a closed conformation.

The MurA structure follows the induced-fit mechanism of substrate binding. It is the interaction of the UNAG sugar nucleotide with the free enzyme that triggers the conformational change necessary for the binding of the second substrate PEP (Schönbrunn et al. 1998; Gautam et al. 2011). UNAG interacts with Arg91, Arg120, Ser162, Val163 and Gly164 upon binding. The oxygen atoms on the pyrophosphate bridge of UNAG form hydrogen bonds with the nitrogen atoms of Val163 and Gly164, the hydroxyl oxygen of Ser162 and arginine residues Arg91 and Arg120 (Skarzynski et al. 1996).

All the MurA structures determined so far share many similarities such as the two domain topology connected by double stranded linker, and the catalytically important residue C115 present on the active site loop, Pro111 to Pro121 (*E. coli* numbering).

1.5 Similarities of AroA and MurA and significance of ASR study on AroA/MurA

The structural and functional characterisations of many AroA and MurA enzymes have been reported and despite only sharing ~25% (Schönbrunn et al. 1996; Eschenburg et al. 2003) amino acid sequence identity, these two enolpyruvyl enzymes share many unique mechanistic and structural similarities that could suggest their evolution from a single common ancestor.

The reactions catalysed by MurA and AroA both proceed *via* an addition-elimination mechanism (Skarzynski et al. 1996). The stereochemical analysis of AroA and MurA reactions indicate that the addition-elimination steps proceed with opposite stereochemistry, i.e., *syn*/*anti* or *anti*/*syn* (Kim et al. 1996b). There are typically two reactions that PEP can undergo: P-O cleavage or C-O cleavage. P-O cleavage of PEP makes the PEP fulfil the dual role of being a phosphorylating agent and versatile biosynthone as a precursor of a carbanion equivalent at C3. C-O cleavage achieves the net aldol condensation of PEP and C-O cleavage for enol ether transfer. The latter is only achieved by MurA and AroA (Schönbrunn et al. 1996; Walsh et al. 1996).

The forward enolpyruvyl transfer reaction catalysed by AroA and MurA results in the production of phosphate and enolpyruvyl products. In the reverse reaction, phosphates increase the reactivity of the enolpyruvyl products for the formation of PEP and S3P/UNAG. Various phosphate analogues were used to study the reverse reaction of AroA and MurA and it was found that the same three analogues promoted catalysis of the AroA and MurA reverse reaction (Zhang & Berti 2006).

In addition to their mechanistic similarities, AroA and MurA also share unique protein architecture. The polypeptide chain forms two similar domains with a six fold repetition of βαβαββ motif forming an "inside-out α/β barrel" and the two domains are connected with a double stranded hinge (Figure 1.9). This unusual structural topology is a characteristic feature of enolpyruvyl transferases (Schönbrunn et al. 1996; Schönbrunn et al. 2001; Oberdorfer et al. 2012).

Both AroA and MurA follow an induced-fit mechanism of ordered substrate binding which allows the initial binding of S3P and UNAG respectively to trigger a conformation change causing the two domains to approach each other to expose the active site in the interdomain cleft (Eschenburg et al. 2005a). The unliganded

enzyme is in an ‘open conformation and the liganded enzyme is generally in a ‘closed’ state, both enzymes protected from proteolytic degradation in the ‘closed’ conformational state (Krekel et al. 1999). Twelve 100% conserved residues (majority of them within the active site and not all catalytic residues) and 17 other residues conserved in 90% of the sequences were identified upon sequence alignment of AroA and MurA primary structures (Schönbrunn et al. 1996; Eschenburg et al. 2003).

However, it should also be noted that the length and conformation of several loops connecting the secondary structure as well as the length and relative orientation of individual β strands and helices differ between the structures of AroA and MurA (Skarzynski et al. 1996).

Thus the stereochemical similarity, along with mechanistic and structural similarity reinforces the evolutionary relationship of AroA and MurA. However, despite all these similarities, it is very important to note that both AroA and MurA are exclusively inhibited by separate molecules, i.e., AroA is not inhibited by fosfomycin and MurA is not inhibited by glyphosate (Schönbrunn et al. 1996; Eschenburg et al. 2003). Such an inhibitor specificity can be explained by comparing AroA and MurA active sites which shows that positioning glyphosate into the active site of MurA causes clashes with Arg120 and Arg371 (*E. cloacae* numbering), is sterically hindered by the N-acetyl moiety of UNAG (Figure 1.13) and lacks certain residues for keeping glyphosate in place Eschenburg, 2003 #43}.

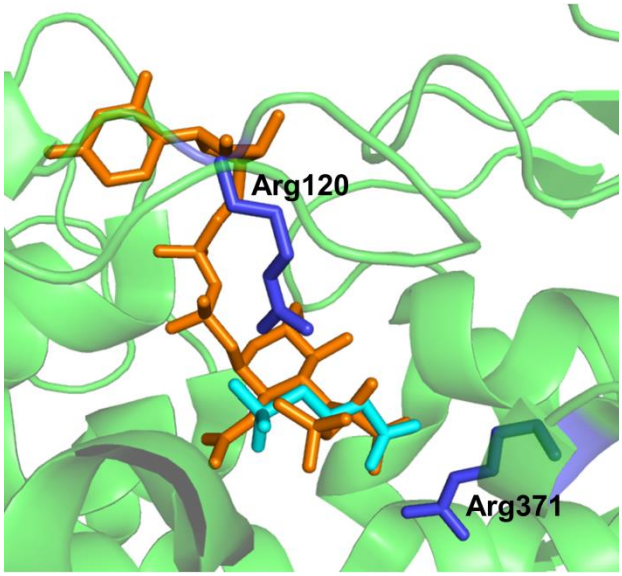


Figure 1.13: Sterical hindrance observed upon placing glyphosate within the active site of MurA
Glyphosate (cyan) from the *E. coli* AroA structure (PDB 1G6S) was placed onto the active site of *E. coli* MurA (green) (PDB 1UAE). The figure shows how the position of glyphosate within MurA active site clashes with that of UNAG (orange) and the position of glyphosate with respect to residues Arg120 and Arg371 (dark blue).

The similarities of the enolpyruvyl transferase enzymes AroA and MurA and their essentiality for bacterial survival make the idea of designing a compound that could inhibit both enzymes very appealing. Bacteria will not be able to easily develop resistance against such an inhibitor and thus will make an invaluable antibiotic candidate. However, one major obstacle for the development of such an inhibitor is the ordered substrate binding mechanism followed by both AroA and MurA. Therefore, the inhibitor should either be able to disrupt the catalytic reaction after the binding of S3P or UNAG or it should be able to mimic these first substrates and disrupt domain closure (Eschenburg et al. 2003).

The relationship between AroA and MurA has been previously noted and suggested to be a result of gene duplication and divergence of function (Nicholas et al. 2001). AroA enzyme, being present in archaea, bacteria and eukaryotes, may have been present in the last universal common ancestor (LUCA) (Richards et al. 2006; Zhi et al. 2014). MurA, on the other hand, an enzyme from the peptidoglycan biosynthesis pathway which is absent in archaea and eukaryotes, is likely to have arisen more recently than AroA. This may indicate that, following the gene duplication of AroA and MurA from the LUCA, the MurA enzyme may have specialised its function only later during the evolution of bacteria. There has not been any experimental study on the evolution of these enzymes. However, the

technique of ancestral sequence reconstruction (ASR) provide us with the opportunity to reconstruct ancestral AroA and MurA enzymes and compare their structural and functional characteristics to present day AroA and MurA enzymes to gain insight on their evolutionary mechanism.

1.6 Ancestral Sequence Reconstruction

1.6.1 ASR Technique

Ancestral sequence reconstruction (ASR) is a technique which uses the sequences of contemporary proteins to infer ancestral sequences (Pauling & Zuckerkandl 1963; Thornton 2004; Harms & Thornton 2010). Such a technique is useful for understanding evolutionary processes and to identify functional aspects of ancestral proteins (Harms & Thornton 2010).

In 1963, Pauling and Zuckerkandl hypothesised that homologous proteins originate from a common ancestor following a gene duplication event (Pauling & Zuckerkandl 1963). The daughter genes exhibit different characteristics based on their independent mutations. Based on this idea they suggested that it is possible to determine the sequence of the common ancestor by comparing sequences of different homologous proteins. Comparison of homologous protein sequences can also give information on the relative time at which the common ancestor existed based on their degree of difference. The amino acid differences between the homologous proteins can be used to determine where the mutations occurred and where different species diverged from the common ancestor. The identification of conserved residues in a protein, within distant organisms, emphasises their role in maintaining the function of the molecule (Pauling & Zuckerkandl 1963).

Important questions such as how protein sequences determine the structure and function of proteins and how contemporary proteins develop such diversity in their structure and function have been examined by using site-directed mutagenesis to swap functional residues between family members. Such a ‘horizontal’ approach has many limits that can be overcome by studying the ‘vertical’ evolution of proteins from its ancestor (Harms & Thornton 2010). Such studies were most commonly achieved using statistical analysis of modern day genes to infer evolutionary patterns until the recent advances in technology led to the development of ASR techniques. ASR allows us to reconstruct the ancestors

using phylogenetics and test the ancestors in laboratories to test evolutionary hypotheses and study their function and structure.

The process of inferring an ancestral protein sequence using ASR involves five steps (Thornton 2004; Harms & Thornton 2010; Hobbs et al. 2012). 1) Extant protein sequences are collected and aligned along with outgroup sequences. A phylogenetic tree that describes the evolutionary relationship of the extant sequences is inferred. A statistical model of amino acid substitution is also identified. Phylogenetic methods such as Maximum Likelihood (ML) are used to infer the sequence at any ancestral node of the phylogenetic tree which has the highest probability to generate the modern day protein sequences, given the ancestral state, model of evolution and phylogenetic tree. 2) The DNA sequence coding for the ancestral protein is inferred. Ambiguous sites obtained from the inference can be resolved by choosing the amino acids that do not incur bias towards any particular feature and that are least likely to contribute to any structural or physiochemical effects to the protein (Hobbs et al. 2012). 3) The DNA sequence coding for the ancestral protein is synthesised *de novo*. 4) The synthesised ancestral gene is then cloned into an expression vector and transformed into bacterial cells. 5) The ancestral protein is expressed and purified and functionally characterised using enzyme assays and/or structurally characterised by crystallography. The enzyme activity can also act as internal control for the ancestral inference because error in inference would most likely result in an inactive enzyme (Thornton 2004; Harms & Thornton 2010; Hobbs et al. 2012). A schematic of the steps involved in experimental ancestral sequence reconstruction studies is given in Figure 1.14.

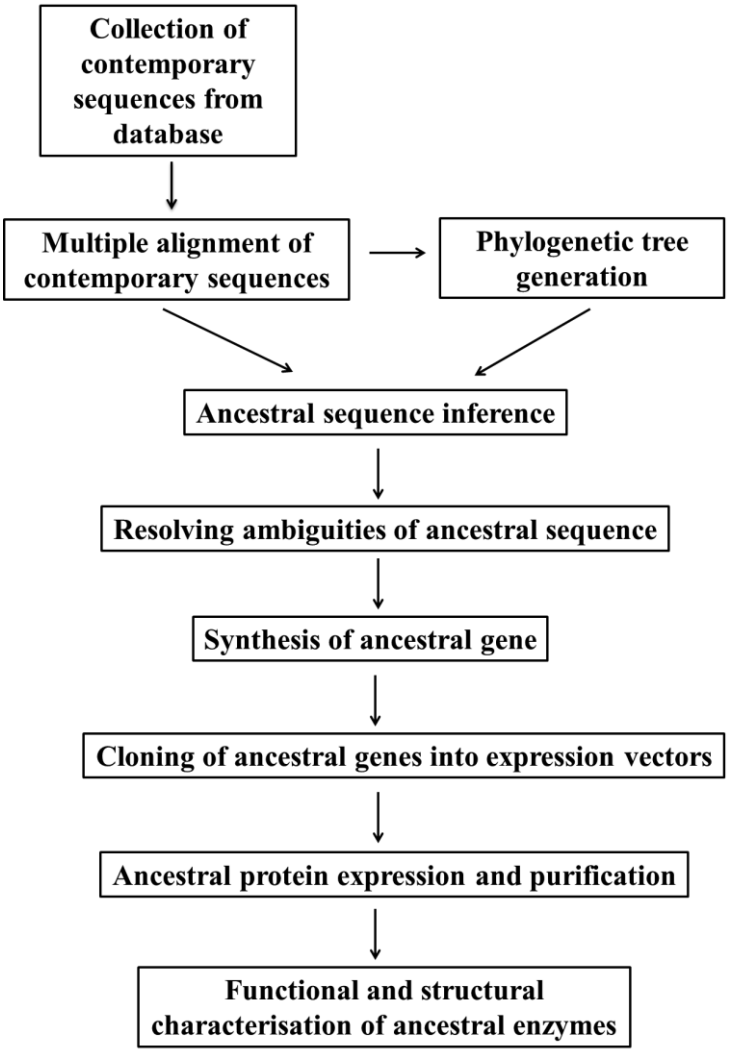


Figure 1.14: Flow diagram showing the steps involved in ancestral sequence reconstruction studies

1.6.2 Methods used for ASR

Three different methods are used for ancestral sequence reconstruction: maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI) methods. Most of the initial ancestral sequences were inferred by MP, which was later superseded by ML, and then followed by BI (Hall 2006; Williams et al. 2006; Hanson-Smith et al. 2010; Hobbs et al. 2012).

1.6.2.1 Maximum Parsimony Method

The Maximum parsimony method was one of the first methods used for ASR studies (Hanson-Smith et al. 2010). A parsimonious evolution represents the ancestral relationship in which the descent to the present day character states are defined by a minimum number of evolutionary changes. Given a sequence alignment of extant species and a phylogenetic tree describing the ancestral

relationship, the MP method assigns the possible ancestral sequences at each node of the tree to be consistent with the minimum number of nucleotide replacements necessary for the descendant to evolve from the common ancestor (Fitch 1971).

During the construction of a phylogenetic tree, the determination of correct topology and branch length is very important. However, since the principle of MP method is to minimize the number of changes required, the method is known to underestimate the number of amino acid and nucleotide substitutions and as a result underestimates branch lengths. This problem does not affect the reliability of the phylogenetic tree as long as the evolutionary distance is less than 0.2 nucleotide substitutions per nucleotide site for each branch (Saitou 1989). The study of the evolution of a particular character hugely benefits from the knowledge obtained from character state ancestral reconstruction. The MP method is found to show compositional bias during such reconstruction especially in cases where there is an unequal proportion of taxa with different states (Collins et al. 1994). Another major drawback of the MP method is that it cannot distinguish equally parsimonious reconstructions (Yang et al. 1995; Cai et al. 2004). The accuracy of the reconstruction was defined to be inversely proportional to the number of equally parsimonious pathways (Zhang & Nei 1997).

1.6.2.2 Maximum Likelihood Method

To infer ancestral sequences using the maximum likelihood (ML) method, we need an alignment of extant sequences, a phylogenetic tree showing the historical relationship of these sequences and a model of evolution. ML method calculates the likelihood of each ancestral state at each node of the phylogeny and each amino acid of the sequence based on the probability of obtaining all the extant states given that ancestral state, tree and model. The ancestral sequence determined by this method will be the state with the maximum likelihood (Hanson-Smith et al. 2010). For all the results on evolutionary characteristics and structural and functional properties of ancestral proteins to be reliable, it is very important that we are confident of the accuracy of the reconstructed ancestral molecule. One of the major disadvantages of the MP method was its inability to quantify uncertainty (Yang 2006). However, ML method can measure the confidence of the ancestral state inference by calculating the posterior probability (Yang et al. 1995).

Unlike MP method, another advantage of ML is that it takes into account branch lengths of the tree and nucleotide and amino acid substitution models to calculate the likelihood function (Yang 2006). However, ML fails to incorporate sample uncertainty. It assumes that the alignment, phylogenetic tree and model of evolution used for the reconstruction to be correct and rules out the possibility of any alternatives to the tree and parameters. As true phylogeny is seldom known with certainty, the fact that ML does not accommodate sampling errors in parameter estimates is problematic particularly in the case of small datasets (Yang 2006; Hanson-Smith et al. 2010).

1.6.2.3 Hierarchical Bayesian Inference Method

Rannala and Yang (1996) proposed a Bayesian model for the estimation of phylogenetic trees using a birth-death process to specify the prior distribution of tree topologies and branch lengths of the extant species and a Markov process to model nucleotide substitution (Rannala & Yang 1996). Maximum likelihood is used to estimate the parameters of the birth-death process and substitution model. The tree with highest posterior probability is chosen as the best estimate of the evolutionary relationship among species. This process is called empirical Bayesian analysis. It differs from ML method as it considers topologies and branch lengths as random variables instead of parameters (Rannala & Yang 1996; Yang & Rannala 1997). However, the calculations involved in this model restricted the number of species that could be used for the analysis. Hierarchical Bayesian analysis, on the other hand, overcomes this restriction and has an improved model for prior distribution of trees and speciation times (Yang & Rannala 1997). Hierarchical Bayesian analysis, unlike ML method, incorporates uncertainty by summing the likelihoods over a distribution of possible trees and parameter values (Huelsenbeck & Bollback 2001; Hanson-Smith et al. 2010). Various ways of incorporating uncertainty into the Baysian method have been proposed such as integration of topological uncertainty (Pagel et al. 2004), uncertainty about the parameters of the evolutionary model (Schultz & Churchill 1999), and uncertainty about the tree topology, branch lengths and model parameters (Huelsenbeck & Bollback 2001) into inference of ancestral states (Hanson-Smith et al. 2010).

1.6.3 Accuracy of ASR

Ancestral sequence reconstruction has become a valuable tool to study the evolution of proteins, structurally and functionally, and for testing various evolutionary hypotheses. All these studies are only relevant if the reconstructions are accurate, and if the reconstructed ancestors are evolutionarily relevant. However, we do not have any way to be certain of the reconstruction accuracy as we do not have actual ancestral sequences to compare them against. Only a few studies on bacteriophages enabled such a comparison of actual ancestor versus inferred ancestor, although the relevance of results obtained from artificially accelerated bacteriophage phylogenies to biomolecules that slowly evolve over millions of years is not known (Hillis et al. 1992; Oakley & Cunningham 2000).

Experimental studies of organismal phylogenetic history are limited as they undergo genetic differentiation and produce phylogenies over very long periods of time, decades to centuries. However, bacteriophages can be manipulated to accelerate their evolution and thus can be used to study their phylogenies. An experimental analysis of the ML method accuracy was conducted by Sanson et al. (2002) where an SSU rDNA ancestor sequence was evolved using a PCR-based method and the resulting ancestor and terminal sequences were determined. This method is an improvement over previous bacteriophage experiments as the mutations were allowed to occur naturally instead of using mutagenic agents. The ancestors inferred using ML were more than 99% accurate and the inaccuracies were mainly at the base of the phylogenetic tree due to insertions, deletions and multiple substitutions. This provided the first biochemical experimental support for ancestral phylogeny inference using ML method (Sanson et al. 2002).

Due to large biases caused by the MP method, ML has become one of the most popular ancestral inference methods. However, ML also suffers certain limitations and many researchers are of the opinion that Bayesian inference methods are better at analyzing complex evolutionary models and large data sets as it is computationally faster and more efficient than ML (Krishnan et al. 2004). A reconstruction of primate mitochondrial cytochrome *b* and cytochrome oxidase subunit I by ML showed changes in nucleotide frequencies within the tip sequences in comparison to the ancestral states, whereas the BI reconstructions resembled the extant sequences more closely. In this study, increased computational speed and feasibility of incorporating more complex evolutionary

models was achieved by using a 'conditional pathway' approach, which limits the number of substitutions allowed per site along each branch. Krishnan et al (2004) suggests that such differences in the nucleotide frequency between inferred and actual ancestor is the influence of bias in ML ancestral reconstruction method. To see if such biased reconstructions would affect inferences of functional properties, primate mitochondrial tRNA ancestors were inferred and evaluated *in silico* resulting in the BI inferred ancestor to be more theoretically functional than the ML inferred ancestor. Hence, bias in ancestral reconstruction does not solely affect the nucleotide frequencies but can lead to inaccuracies in functional predictions (Krishnan et al. 2004). Therefore, it was pointed out by Krishnan et al. (2004) that conclusions drawn from the study performed by Gaucher et al. (2003) on elongation factors using ML needs to be reevaluated. They suggest that the thermophilic ancestors might have been obtained as a result of ML method bias towards the amino acids responsible for thermostability. Williams et al. (2006) agrees that reconstructions using MP and ML methods could generate misleading results and the conclusions drawn from such experiments should be reevaluated. He supports this by performing an *in silico* evolutionary simulation study based on a realistic phylogenetic tree obtained from RNase A sequences. The protein fitness was measured as the amount of time the protein remains correctly folded. The proteins were pre-equilibrated before the simulation started, making it easy to find out the consistent biases as a deviation from the equilibrium properties. Thus the thermodynamics of the reconstructed ancestors were compared to the actual ancestors and it was observed that ML overestimated thermostability whereas BI did not. This is because ML tends to choose the amino acids that are not detrimental to the protein stability (Williams et al. 2006). A similar conclusion regarding the preference of BI method for the reconstruction process was obtained from a study using a sequence evolution simulation program (Hall 2006).

Many studies have demonstrated contradictory results to the above mentioned conclusions and indicate that ancestors are more accurately inferred using ML as opposed to BI. For instance, one of the best known advantages of BI is its incorporation of uncertainty about tree topology, branch lengths and model of evolution. This is considered an advantage over ML, as ML ignores any uncertainty and assumes that the datasets are correct (Hanson-Smith et al. 2010). Simulation based experiments were performed to study the effect of incorporating

phylogenetic uncertainty on ASR accuracy. The ancestors were inferred using ML method and BI method incorporated with topological uncertainty and the inferred ancestors were compared to actual ancestors. The results showed that ML produced slightly higher accuracy and indicate that the use of BI with topological uncertainty is neither necessary nor beneficial. Thus ML can produce accurate ancestral reconstruction and was found to be robust to phylogenetic uncertainty (Hanson-Smith et al. 2010). Reconstruction of ancestral sequences of elongation factors across two bacterial phylogenies showed that ML reconstructions are robust to uncertainties and potential biases associated with ancestral inference. The resulting theory that ancient environments cooled progressively was supported by an identical cooling trend exhibited by the ancient oceans inferred from deposition of oxygen isotopes (Gaucher et al. 2008).

All the above studies comparing the ASR methods are theoretical and the accuracy of the methods can be more reliably measured by testing the reconstructed ancestors *in vitro* or *in vivo*. Only one such study was performed experimentally to compare the accuracies of ML versus BI inferred ancestors. Hobbs et al. (2012) reconstructed the ancestors of structurally complex core metabolic enzyme LeuB from *Bacillus* species. Any error in the ancestral inference could potentially generate inactive enzymes or enzymes with biologically unrealistic properties. Thus, enzyme activity can be used as a measure of ancestral inference accuracy. Therefore, to compare the BI and ML reconstructions, the thermodynamic and kinetic properties of the enzymes reconstructed by both methods were determined. It was observed that the ancestors reconstructed using ML method were fully functional and the ancestral enzyme kinetics were comparable to the contemporary LeuB, whereas BI method exhibited a greater number of errors in the inferred sequences with biologically unrealistic thermodynamic and kinetic values (Hobbs et al. 2012). This study highlights the accuracy of ML in comparison to BI, however, the result is based on a single enzyme and similar experimental trials on other enzymes need to be performed to strengthen this conclusion.

The topology of the phylogenetic tree used for reconstruction of the ancestral sequence of its extant taxa, and the number of taxa in the phylogenetic tree are important factors to consider for accurate ancestral reconstruction. Depending on the method of reconstruction and model of evolution used for ancestral inference,

the number of taxa used for the phylogenetic tree can influence the accuracy of the reconstruction. With some trees it is better to have a subset of species rather than having all species (Li et al. 2008). On the other hand, MP method can give more accurate reconstructions with increased sampling of terminal taxa (Salisbury & Kim 2001; Li et al. 2008). Recombination is widespread and depending on each case, it is advisable that recombination is accommodated in the ancestral sequence inference process, as ignoring recombination can bias phylogenetic estimation (Arenas & Posada 2010).

With the exception of experiments conducted on bacteriophages, the relevance of which is questionable, it has to be noted that all the studies comparing the accuracy of various methods of ASR are theoretical. Even though simulation provides considerable insight, it incorporates many untested evolutionary assumptions. Therefore, future work is necessary to find out which method reconstructs ancestors that represent the true biological ancestor. More laboratory experiments on reconstructed proteins need to be conducted as opposed to simulation comparisons. Based on many studies mentioned above, ML method proves to be reconstructing biologically realistic ancestral molecules and should be used for future ancestral sequence reconstructions.

1.7 *In Vivo* enzyme characterisation

At a molecular level, the evolution of proteins with respect to their sequence, expression, structure and biophysical properties such as stability, aggregation and degradation, impact organismal fitness (DePristo et al. 2005; Couñago et al. 2006). However, enzyme evolution is often studied by focussing on single proteins and by characterising the enzymes *in vitro*. By doing so, the complex interactions occurring in a cellular context are ignored. The important evolutionary processes are influenced by the properties of the cellular metabolic network (Vitkup et al. 2006). There are various uncertainties that are not taken into consideration while solely focussing on *in vitro* characteristics of proteins, such as, the influence of other metabolites within the system, the difference in overall conditions within the living cells and *in vitro*, and the enzyme concentration (Teusink et al. 2000). Therefore, it is beneficial to integrate the fields of molecular evolution and biochemistry for a better understanding of protein evolution (Harms & Thornton 2013; Hobbs et al. 2015).

While in some instances the protein properties determined *in vitro* directly reflect on the cell behaviour and organismal fitness, it is not always the case (Teusink et al. 2000; DePristo et al. 2005). For example, it has been observed that while half of the enzymes from yeast glycolysis described the *in vivo* activity satisfactorily using the *in vitro* kinetics, the rest of the enzymes showed discrepancies between the *in vitro* and *in vivo* kinetic characteristics of yeast glycolysis (Teusink et al. 2000).

Furthermore, the technique of ASR provides an opportunity to reconstruct ancestral enzymes. *In vivo* properties of reconstructed ancestral enzymes have been studied and it was observed that superior *in vitro* properties did not translate into organismal fitness *in vivo* (Hobbs et al. 2015). Therefore, the use of ASR for resurrection of ancestral enzymes and the study of their biochemical properties *in vitro* and organismal fitness *in vivo,* would not only provide more information on the protein evolution, but also help us have a better understanding of the relationship between organismal fitness and protein properties.

1.8 Using ASR to study evolution of thermophily

One of the most intriguing questions of ancient life is understanding the role of thermophily in the history of Earth. There are two main proposed hypotheses for the evolution of thermophily. One of the early hypotheses suggests that thermophily is a primitive trait, as all the deep branches of the eubacterial tree are predominantly thermophilic groups (Woese 1987; Pace 1991; Stetter 1996). The second hypothesis proposes a mesophilic origin of life with the present day prokaryotes originating from a mesophilic last common ancestor (Forterre 1995). A possible way to test these hypotheses is to use ASR to reconstruct ancestral proteins from Precambrian bacteria, and to find their thermostabilities, as it would give a direct estimation of the environmental temperature the bacteria lived in at that time (Gaucher et al. 2003). Studies on elongation factors from *Escherichia, Thermus* and *Thermotoga* have confirmed such a correlation between the thermal optima and organism growth temperature (Arai et al. 1972; Nock et al. 1995; Sanangelantoni et al. 1996). A theoretical analysis performed by Gromiha et al. on 56 globular proteins from 16 different families also confirms a direct relationship between environment temperature of host organisms and protein melting temperature with a correlation coefficient of 0.91. The temperatures in this study

were assigned to the proteins using either the optimal growth of the species involved or from *in vitro* experiments as stated in literature (Vogt et al. 1997; Gromiha et al. 1999).

Many studies have been previously done on the evolution of thermophily using the technique of ASR. Elongation factors EF-Tu proteins were good candidate proteins to perform such a study because of their very slow rates of sequence divergence allowing robust reconstructions (Gaucher et al. 2003). Two phylogenetic trees were used to support the robustness of the interpretations and 'most probabilistic ancestral sequences' were reconstructed. The most recent common ancestor of the mesophilic lieneage was also reconstructed. The ability of EF-Tu proteins to bind GDP at different temperatures was used as a measure of the thermostabilities of the reconstructed ancestor proteins and proteins from representative contemporary organisms. The results suggested that the paleoenvironment of ancient bacteria containing these resurrected proteins was at approximately 65 °C. It was also inferred that the ancestor of modern mesophiles lived at higher temperatures (Gaucher et al. 2003). Williams et al. expressed his concern about this study on elongation factors, as the reconstruction was achieved using maximum likelihood, and conclusions drawn from such studies can be incorrect if ML reconstruction overestimates or underestimates thermostability (Williams et al. 2006). These concerns were resolved by the future work of Gaucher et al. on elongation factors (EF) giving results robust to uncertainties and potential biases. They reconstructed ancestral EF sequences using one phylogeny having the hyperthermophilic lineage at the base of the bacterial tree and a second phylogeny placing the hyperthermophilic lineage at a more derived portion of the tree. The thermostabilities of ancestral EF were determined across both phylogenies and the results showed similar temperature trends (Gaucher et al. 2008).

Perez-Jimenez et al. reconstructed seven Precambrian thioredoxin enzymes (Trx) dating back between ~1.4 and ~4 billion years (Gyr). The conserved active site CXXC and Trx fold marks the existence of Trx enzymes in primitive forms of life. The results agree with studies performed on elongation factors (Gaucher et al. 2003; Gaucher et al. 2008) and also indicate a primitive thermophilic environment and adaptation to progressive cooling (Perez-Jimenez et al. 2011).

Study on the evolution of thermophily has also been performed on a structurally complex core metabolic enzyme by reconstructing four ancestors (ANC1 to ANC4) of 3-isopropylmalate dehydrogense (LeuB) of *Bacillus* species (Hobbs et al. 2012). The ancestral reconstruction was proved to be accurate as the enzymes were functional with biologically realistic kinetics. The last common ancestor of LeuB from *Bacillus* species (ANC4) and ANC1 exhibited thermophilic characteristics. However, the difference in the thermophily mechanism of ANC4 and ANC1 suggests that thermophily can be readily gained and lost through independent evolutionary paths and that thermophily is not a strictly primitive trait. The fluctuating trend in thermophily observed in this study is in contrast to the decreasing thermophily trend proposed by Gaucher et al. (2008), however this could be due to lack of enough samples (Hobbs et al. 2012). The conclusions drawn from Hobbs et al. (2012) are based on one enzyme. Therefore, conducting the study on a different enzyme would be ideal to confirm the results.

1.9 Research Objectives

AroA and MurA are considered to be very important antibacterial targets due to the essentiality of their pathways to bacteria and the absence of the pathways in mammals. AroA and MurA, have been widely studied for the synthesis of new antibiotics and both being enolpyruvyl transferase enzymes, exhibit very high structural and mechanistic similarities. Such similarities led to the hypothesis that AroA and MurA originated from a common ancestor followed by a gene duplication and functional diversification event. Therefore, by going back in evolutionary time, we may be able to see characteristic features that prove such an origin of AroA and MurA.

This study is the first experimental investigation of the evolution of AroA and MurA from *Bacillus* using the technique of ASR (Figure 1.15). This study aims at answering three major questions:

1) How different the ancestral AroA and MurA enzymes are to their contemporary enzymes with respect to their kinetic, thermal, inhibitory and structural properties?

2) Do ancestral AroA and MurA enzymes show any evidence of promiscuous inhibition in contrast to their present day exclusive inhibition by glyphosate and fosfomycin respectively?

3) How robust is the technique of ASR in studying the evolution of thermophily?

This is achieved through the following steps:

1) Ancestral sequence reconstruction of AroA and MurA from *Bacillus* ancestors

2) Determination of kinetic parameters and inhibitory properties of both ancestral and contemporary AroA and MurA enzymes

3) Studies on the thermal properties of MurA ancestors, to investigate the evolution of thermophily in the *Bacillus* genus

4) *In vivo* characterisation of AroA enzyme activity

5) Determination of the structure of AroA and MurA enzymes from the last common ancestor of *Bacillus* genus

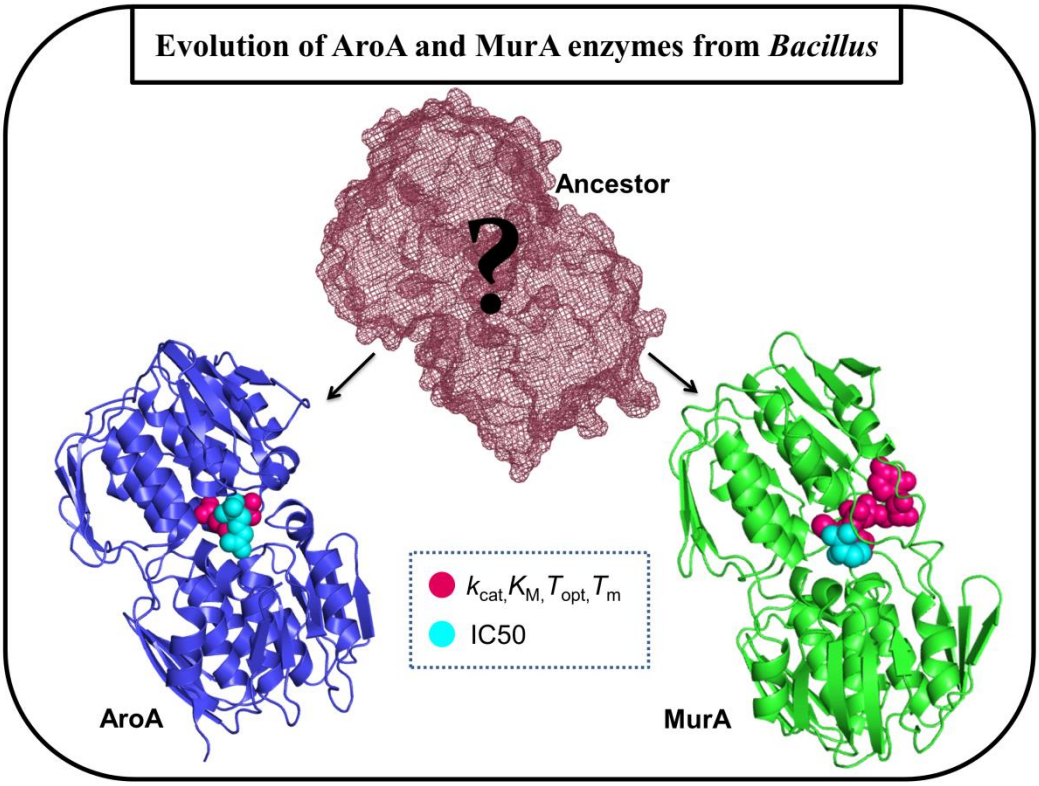


Figure 1.15: Schematic diagram depicting the aims of this thesis
AroA: *E. coli* AroA (PDB 1G6S); MurA: *E. coli* MurA (PDB 1UAE). The respective substrates are shown in magenta and inhibitors shown in cyan.

Chapter 2

Materials and Methods

2.1 Phylogenetics

2.1.1 Sequences and Alignment

All the available amino acid sequences for AroA and MurA of the *Bacillus* genus were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/) and the nucleotide sequences were obtained from GenBank. Two of the sequences from the *Clostridium* genus were also collected to act as the outgroup. The gene accession numbers and strain information are listed in the Appendix A. All the sequences were collated in Geneious version 8.0.2 (http://www.geneious.com) (Kearse et al. 2012) and the nucleotide sequences were verified to ensure that they exactly matched the amino acid sequences.

Multiple sequence alignments of the amino acid and nucleotide sequences were generated in Geneious using the built in ClustalW algorithm (Larkin et al. 2007). The protein alignment was manually checked and adjusted wherever necessary by moving and/or rearranging the existing gaps. The nucleotide alignment was adjusted to exactly match the protein alignment. For use in different programs, both alignments were exported from Geneious in PHYLIP format.

2.1.2 Determination of Models of Evolution

The most appropriate model of protein evolution was determined using the program ProtTest version 2.4 (Abascal et al. 2005) based on the protein alignment of the contemporary sequences. The most appropriate model of nucleotide evolution was determined using jModelTest version 0.1.1 (Posada 2008) based on the nucleotide alignment of the contemporary sequences.

2.1.3 Phylogenetic Tree Construction

2.1.3.1 Maximum Likelihood Tree

The programme GARLI version 1.0 (Zwickl 2006) was used to build phylogenetic trees under the Maximum Likelihood framework. To determine the final phylogeny, at least five trees were generated using GARLI and the topologies of

the resulting trees were compared. The ancestral root position of these trees was defined using Geneious by rooting the node that leads only to the outgroup *Clostridium* sequences. The best tree out of all the GARLI generated trees was chosen based on the most consistent topology and the highest log likelihood (lnL) score.

2.1.3.2 Bootstrapping

The bootstrap percentage gives the frequency with which a given branch occurs within the pseudoreplicate datasets and allows the reliability of each branch within the phylogenetic tree to be estimated. The phylogenetic tree generated by GARLI was bootstrapped by running GARLI version 1.0 (Zwickl 2006) and setting the field of 'bootstraprep' to 1024. The output file, containing 1024 trees, was opened in Geneious and a consensus tree was generated using a 50 % support threshold and 10 % burn-in. The bootstrap percentages were transposed onto the branches of the 'best' tree if those branches exist.

2.1.4 Ancestral Sequence Inference

2.1.4.1 Phylogenetic Analysis by Maximum Likelihood (PAML)

PAML was used to perform the ancestral inference under the maximum likelihood criterion. The protein alignment was saved with an .aa extension and the nucleotide alignment was saved as a .nuc file. The 'best tree' obtained as in section 2.1.3.1 was exported from Geneious in Nexus format, manually edited into the Newick format and saved as a .tre file.

CODEML performs ancestral inference based on amino acid and codon information whereas BASEML uses nucleotide information for the same. For this purpose, the 'best tree' along with the protein alignment was used by CODEML for the amino acid inference, while nucleotide alignment was used by CODEML for codon inference and by BASEML for nucleotide inference. The configuration files for CODEML and BASEML were edited to mainly contain the protein alignment filename, the best tree filename, output filename, and model of evolution. The program was run using command prompt and the output 'rst' file was saved for later use.

2.1.4.2 Generating Consensus Ancestral Sequences

The 'rst' files from each of the inference runs were opened in a text editing program for analysis. The tree description given in the 'rst' file was copied into a new file with .tre extension. This tree file was then opened in Geneious and the node numbers assigned by PAML for each ancestral node was displayed. The ancestral sequences corresponding to the node numbers given in the tree were obtained from the amino acid, codon and nucleotide inference 'rst' files and Geneious was used to analyse them. The overall posterior probabilities (a measure of confidence of the ancestral state), also given in the 'rst' files, were noted down to confirm the choices of ancestral nodes.

The ancestral nodes that were reconstructed were chosen based on the position of the node in the tree, the bootstrap percentage and posterior probabilities of the nodes, and sequence identity between the node and the extant sequences. Once the node of interest was chosen, the inferred sequences corresponding to that node from the three inference methods were translated into protein sequences in Geneious and compared to generate the final ancestral sequence. Ambiguous sites within the inference were resolved as described below.

2.1.4.3 Resolving Sequence Inference Ambiguities

The three versions of each sequence were compared to resolve sequence ambiguities between the three inference methods. The sites that are identical for amino acid and nucleotide methods but different from the codon method, and the sites that are chosen differently by all the three methods, were considered as ambiguous sites and certain criteria were followed to resolve the ambiguities. If the amino acid and nucleotide methods chose the same sites but differed from the codon method, then generally the original alignment was considered to find out the amino acid which is most conserved in that particular site. In cases where all three methods chose a different amino acid, the following criteria were used:

(i) The distributions of the amino acids in the original alignment were considered and the amino acid present in the majority of the sequences was used. If not, the amino acid present in a functional extant protein was used.

(ii) Amino acids with conserved physicochemical characteristics between different contemporary species were chosen.

(iii) The structures of the extant proteins were considered and the implication of each ambiguous amino acid in the structural environment was analysed.

In all cases of ambiguous site determination, sequence bias from contemporary species branching close to the node of interest and sequence bias with respect to thermal adaptation was avoided.

2.1.5 Aging Ancestral Nodes

The programme r8s version 1.71 (Sanderson 2003) was used to estimate the age of each node of the tree and to convert the phylogram into a chronogram. Calibration points were taken from the literature (Battistuzzi et al. 2004). The chronogram for prokaryotic evolution, from which these calibration points were taken, was generated using 32 protein sequences, and molecular data and geologic calibration points were used to estimate their phylogenetic relationship and divergence times (Battistuzzi et al. 2004). To select calibration points, the node where two species of interest diverged was traced back and the timescale along the bottom of the tree was used to determine the calibration point for divergence of these two species. In this case, using the chronogram for prokaryotic evolution (Battistuzzi et al. 2004), the point of divergence of *Bacillus subtilis* and *Bacillus halodurans,* and the point of divergence of the *Bacillus* and *Clostridium* genera were determined as the calibration points. R8s fixed these calibration points to estimate the ages of the remaining nodes and the resulting chronogram was opened as a .tre file in FigTree version 1.3.1 (Rambaut 2009). Further image editing was performed in CorelDraw.

2.2 Cloning

2.2.1 Gene Synthesis

The genes encoding for the *Bacillus* AroA LCA, *Bacillus* MurA ancestors and *Bacillus coagulans* MurA were synthesised by GeneArt (Life Technologies). The genes were codon optimized for expression in *Escherichia coli* and the restriction

sites *Bam*HI and *Xho*I were incorporated on the 5′ and 3′ ends of the genes, respectively.

2.2.2 Contemporary Genes

Genomic DNA from *Bacillus subtilis* Marburg 168 was provided by Joel McMillan (University of Waikato) and used for the amplification of *B. subtilis murA* and *aroA*.

2.2.2.1 Polymerase Chain Reaction (PCR)

Amplification primers were designed to be 28 to 38 bp long and included the appropriate restriction sites at the 5′ and 3′ ends for the subsequent steps of cloning (Table 2.1). The primers were supplied by Integrated DNA Technologies (USA) and primer stocks were made up to 100 pmol/µl with ultrapure water.

Table 2.1: Primer sequences

Primer		Primer Sequence
Bacillus subtilis AroA	F	5′-TGGAAGGATCCAAACGAGATAAGGTGCAG-3′
	R	5′-AATCCTCGAGTAAAACTTCAGGATTTTTTCGAAAGC-3′
Bacillus subtilis MurA	F	5′-CATGCCATGGAATTGGAAAAAATCATCGTCCGCGGCG-3′
	R	5′-AACTGCAGTTATGCATTTAAGTCAGAAACGACT-3′
pPROEX	F	5′-AGCGGATAACAATTTCACACA -3′
	R	5′-TATCAGGCTGAAAATCTTCTC-3′

*F – Forward primer; R – Reverse primer

Genes *aroA* and *murA* were amplified from the genomic DNA of *B. subtilis* Marburg168 using iProof High-Fidelity PCR Kit (Bio-Rad) according to the manufacturer's instructions. The reactions were set up to 15 µl or 50 µl final volume.

The following are the PCR cycling conditions and reaction concentrations:

Steps	**Temperature**	**Time**	
Initial Denaturation	98 °C	2 minutes	
Denaturation	98 °C	10 seconds	29 cycles
Annealing	62 °C	30 seconds	
Elongation	72 °C	45 seconds	
Final Elongation	72 °C	5 minutes	

DNA template	**50-250 ng**
Forward primer	0.5 μM
Reverse primer	0.5 μM
Deoxynucleotide triphosphate mix	0.2 mM
$MgCl_2$	1 mM
Reaction buffer	1x
iPROOF DNA Polymerase	1 U

The PCR products were purified using the Qiaquick PCR Product Purification Kit (Qiagen) according to manufacturer's instructions.

2.2.3 DNA Quantification

The Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies) was used to measure the concentration of DNA samples at 260 nm absorbance.

2.2.4 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments. The agarose gels (1 % w/v) were made up using 1x TAE buffer (Appendix E1) and stained with 1x SYBR Safe DNA gel stain (Invitrogen). The DNA samples were mixed with 10x DNA loading dye (Invitrogen) prior to loading onto the gel. The DNA fragments were separated by running the gels at 100 V for 40 minutes and visualised on a blue light box. The DNA size was determined by comparison with a 1kb-Plus DNA ladder (Invitrogen).

2.2.5 Plasmid Preparation

The expression plasmid pPROEX HTb (Invitrogen) was isolated from DH5α cells which were grown overnight in LB media at 37 °C in the presence of antibiotic ampicillin at a final concentration of 100 μg/ml. The QIAprep Spin Miniprep Kit (Qiagen) was used to isolate the plasmid from the cells.

2.2.6 Restriction Digestion

The GeneArt genes and the *B. subtilis* AroA gene amplified by PCR were digested with *Bam*HI and *Xho*I. The PCR product for *B. subtilis* MurA was digested with *Nco*I and *Pst*I. Restriction digestion was followed by a purification step using the Qiaquick PCR Product Purification Kit (Qiagen).

The plasmid pPROEX HTb was digested with either *Bam*HI and *Xho*I or *Nco*I and *Pst*I and run on an agarose gel to separate the digested fragment from the rest of the plasmid. The band corresponding to the linearised plasmid was then excised from the gel using a scalpel and purified using the QIAquick Gel Extraction Kit (Qiagen). These restriction sites were used to allow an N-terminal His tag to the insert.

2.2.7 Ligation

Ligation reactions contained a 3:1 molar ratio of insert to plasmid, 1 U of T4 DNA ligase, and 1x ligase buffer in a 20 μl total volume (Invitrogen). The reaction was incubated at 16 °C overnight.

A negative control was also set up in the same way as described above that contained digested plasmid.

2.2.8 Transformation

2.2.8.1 Electrocompetent Cell Preparation

DH5α cells were grown on a LB agar plate in the absence of antibiotic at 37 °C overnight. A single colony from this plate was used to set up a 10 ml inoculum culture in LB broth in the absence of antibiotic at 37 °C overnight in an incubator shaking at 200 rpm. One litre of LB broth was inoculated with the 10 ml inoculum and the growth of the culture was monitored using a spectrophotometer at 600 nm until the optical density (O.D.) reached between 0.5 and 0.7. The flask containing the grown cells was cooled on ice for 20 minutes then pelleted using a Heraeus Multifuge 3SR Plus centrifuge at 4000 g for 20 minutes at 4 °C. The supernatant

was discarded and the cells resuspended in 1 L of cold, sterilised 10 % glycerol and centrifuged again. This step was repeated with decreasing volumes (500 ml, 20 ml and 2 ml) of 10 % glycerol. The final amount to resuspend the cells was determined based on the cell density of the initial culture. Aliquots of 50 µl of the cell suspension were transferred into microcentrifuge tubes and immediately stored at -80 °C.

2.2.8.2 Electroporation

Electroporation cuvettes (Bio-Rad Laboratories, cuvette path length = 0.2 cm) were chilled on ice before transferring 20 µl of the ligation reaction, 50 µl of electrocompetent DH5α cells and 30 µl of 10 % glycerol into them. Transformation of the ligation products was achieved by electroporation using the Bio-Rad Gene Pulser (Bio-Rad Laboratories) with 2.5 kV at 25 µF capacitance and 200 Ω resistance. One millilitre of LB broth was added to the cells immediately after electroporation and left to incubate at 37 °C for one hour shaking at 200 rpm. The cells were then plated onto LB agar plates containing the required antibiotic and incubated overnight at 37 °C.

2.2.8.3 Screening of Transformants

The number of colonies on the ligation plates was compared with the number of colonies on the negative control plate. If it was considerably different, with more colonies on the ligation plate than the control plate, the following procedures were used to screen the transformants:

(i) Plasmid Restriction digestion or PCR

A number of colonies selected from the ligation plates were grown in 5 ml of LB broth along with the respective antibiotic overnight at 37 °C shaking at 200 rpm. The plasmid from the cultures was purified as described in section 2.2.5.

The plasmids were either digested (section 2.2.6) or a diagnostic PCR using *Taq* polymerase was performed as per the manufacturer's instructions using pPROEX primers (Table 2.1) which flank either side of the multiple cloning sites. The reaction was run onto an agarose gel to check for the presence of the insert.

(ii) Colony PCR

Colonies from the ligation plates after transformation were selected and resuspended in a dilute solution of the appropriate antibiotic. The resuspended colony acted as the DNA template in the *Taq* DNA polymerase PCR reaction using pPROEX primers. The PCR was performed with an annealing temperature of 52.5 °C due to the T_m of the pPROEX primers. The colonies of positive transformants were then grown overnight at 37 °C with appropriate antibiotic and the plasmid-insert complex was purified as described in section 2.2.5.

The presence of the insert of interest in the purified plasmids was confirmed by sequencing using either gene primers or pPROEX HTb primers (Table 2.1).

2.2.8.4 Glycerol Stocks

Glycerol stocks were made up by combining 900 µl of culture grown from positive transformants with 100 µl of 80 % glycerol. The glycerol stocks were then stored at -80 °C.

2.3 Protein Purification and *In Vitro* Enzyme Characterisation

2.3.1 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples were analysed by separating them on a 15 % SDS-polyacrylamide gel (Appendix E4). The protein samples were mixed with 4x SDS loading dye (Appendix E1) and incubated at 95 °C for 5 minutes prior to loading 10 µl of the mix onto the gel. To determine the protein size, 10 µl of protein ladder, Precision Plus Protein Unstained Ladder (Bio-Rad Laboratories), was run alongside the samples. The SDS-PAGE gels were run in 1x SDS-PAGE running buffer (Appendix E1) at 15 mA until the samples passed the stacking gel layer, and at 30 mA through the resolving gel layer until the dye reached the bottom of the gel.

After running, the gels were visualised by staining with Coomassie stain. The gels were transferred into containers, covered with Coomassie stain (Appendix E1) and microwaved for 30 seconds. Gels were then transferred onto a shaker and left shaking at room temperature (RT) for about 30 minutes. The stain was decanted and replaced with a destain solution of 10 % v/v acetic acid. After microwaving for 30 seconds, the gels were left to destain until the bands were well defined.

2.3.2 Measurement of Protein Concentration (Bradford Assay)

The concentration of AroA and MurA proteins were measured using the Bio-Rad Protein Assay (Bio-Rad Laboratories). The Bio-Rad reactions were set up in a 96 well microplate (Greiner Bio-One) and contained 5 µl of protein sample, 40 µl of Bio-Rad protein assay dye (Bio-Rad Laboratories). The buffer in which the protein is stored was added last into the Bio-Rad reaction to make up the reaction volume to 200 µl. The absorbance was read at 600 nm using a Multiskan GO microplate spectrophotometer (Thermo Scientific) immediately after mixing the buffer in.

A standard curve was generated using bovine serum albumin (BSA). Ten dilutions of BSA ranging from 0 to 1 mg/ml were made up and the absorbance measured as mentioned above immediately after the addition of MQ water instead of buffer. A standard curve was generated using the absorbance data and this curve provided a relative measurement of protein concentration.

2.3.3 Small-scale Protein Expression Trial

The glycerol stocks made up after successful cloning were used to grow 1 ml inocula of the strains in LB broth with appropriate antibiotic at 37 °C overnight shaking at 200 rpm. The inocula were used to inoculate 100 ml LB broth containing appropriate antibiotic and were incubated at 37 °C with shaking at 200 rpm until the OD_{600nm} reached between 0.5 and 0.7. A Helios spectrophotometer (Thermo Fisher Scientific) was used to measure the OD_{600nm} at various intervals. Samples were taken from this main culture for the following methods of protein expression analysis.

2.3.3.1 Protein Expression Trial: Effect of time and pH

One control sample of 10 ml volume was taken from the above mentioned culture once it reached the appropriate OD and was pelleted by centrifuging the cells at 4500 rpm at 4 °C for 20 minutes. The supernatant was discarded and the pellet stored at -80 °C. Protein expression was induced in the remaining culture by the addition of IPTG to a final concentration of 1 mM and incubated at 37 °C with shaking of 200 rpm. Samples of 10 ml volume were taken periodically from the culture and treated in the same way as the control sample. After collecting all the samples, the cell pellets were defrosted and resuspended in 300 µl of Lysis Buffer

(50 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM imidazole). The resuspended cells were transferred into microcentrifuge tubes and sonicated on ice using the Ultra sonic processor XL 2020 sonicator (Misonix incorporation) (micro tip, setting 5, 4 x 15 seconds bursts with 30 second interval in between). The sonicated cells were centrifuged at 13,000 rpm at 4 °C for 20 minutes to separate the soluble and insoluble fractions. The protein concentration present in the supernatant was measured using a Nanodrop ND-1000 and approximately 10 μg of total protein was loaded onto a 15 % SDS-PAGE gel along with 1 μl of the insoluble pellet diluted with Lysis Buffer for analysis.

This method of protein expression testing was also used to perform a lysis buffer screen. Instead of taking samples periodically, seven samples of 10 ml volume were taken from the main culture after incubating the culture overnight and pelleted as previously described. Each of these pellets were resuspended with lysis buffers of different pHs (Appendix E2) and then treated in the same way as the time course samples.

2.3.3.2 Ni^{2+}-Sepharose Bead Protein Expression Trial

Ni^{2+}-sepharose beads were used to test for the expression and his-tag binding affinity of the proteins. A pellet from a 10 ml sample taken from a culture grown overnight (post induction with IPTG) was resuspended with 300 μl of Lysis Buffer. The cells were sonicated as previously mentioned and centrifuged at 13,000 rpm at 4 °C for 20 minutes. The Ni^{2+}-sepharose high performance beads (GE Healthcare) were equilibrated with Lysis Buffer by mixing the beads with the buffer, allowing the beads to settle down and then centrifuging at 3000 rpm at RT for 30 seconds. The buffer supernatant was removed and the beads were loaded with 300 μl of the supernatant from the centrifuged cell lysate. The beads with the load were incubated in an Eppendorf ThermoMixer comfort at 22 °C with shaking at 1000 rpm for 15 minutes. The beads were then allowed to settle before centrifuging for 30 seconds at RT at 3000 rpm. The supernatant was removed and kept aside. The beads were then washed with 1 ml Lysis Buffer three times. For the analysis of the protein expression, the pellet obtained after sonication (pellet), the supernatant used as the load (load), the supernatant after the incubation with

the beads (flowthrough), the first wash (wash 1) and the Ni^{2+}-sepharose beads (beads) were loaded onto a 15 % SDS-PAGE gel.

2.3.4 Large-scale Protein expression

The glycerol stocks of the strains of interest were used to grow inoculum cultures in 10 ml LB broth in the presence of appropriate antibiotic overnight at 37 °C. The inocula were used to inoculate 1 L of LB broth in the presence of antibiotic and left to grow in a shaking incubator at 200 rpm at 37 °C until the OD_{600nm} reached between 0.5 and 0.7. Once the cultures reached the right OD, protein expression was induced with the addition of IPTG at a final concentration of 1 mM. AroA protein expression was achieved overnight at 37 °C while MurA protein expression was achieved overnight at 18 °C (Du et al. 2000).The cells were then pelleted by centrifugation at 4 °C at 4500 rpm for 20 minutes. The pellets were stored at -80 °C.

2.3.5 Immobilised Metal Affinity Chromatography (IMAC)

Cell pellets obtained from the large-scale protein expression were thawed, resuspended in 20 ml of Lysis Buffer and sonicated on ice using a microtip (setting 5) for 6x 15 second bursts with 30 seconds interval in between. The insoluble protein fraction was separated out by centrifugation at 4 °C for 20 minutes at 13,000 rpm. A 5 ml HisTrap FF nickel affinity column (GE Healthcare) was washed with four column volumes (CV) of MQ water and primed with four CV of Lysis Buffer. The supernatant from the lysed cells was filtered through 1.2, 0.45 and 0.2 μm Minisart filters (Sartorius AG) prior to being loaded onto the HisTrap column which was in turn attached to an ÄKTA Prime, ÄKTA Basic or ÄKTA Purifier FPLC system (GE Healthcare). Unbound protein was removed by washing the column with Lysis Buffer at a flow rate of 1 ml/min until the absorbance at 280 nm dropped to baseline. Bound proteins were eluted with a gradient from 0-100 % elution buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl and 1 M imidazole) at a flow rate of 1 ml/min over 50 ml and collected in 2 ml fractions. The fractions corresponding to the peak/s at 280 nm on the resulting chromatogram were analysed using SDS-PAGE for the protein of interest (section 2.3.1).

2.3.6 Size Exclusion Chromatography (SEC)

Size exclusion chromatography was used to further purify proteins following nickel affinity chromatography. The size exclusion column to be used, either a Superdex 200 10/300 or a Superdex 75 10/300, depending on the size of the protein (GE Healthcare), was equilibrated with SEC buffer (50 mM Tris-HCl pH 7.5 and 300 mM NaCl). The protein fractions from the nickel purification were concentrated by centrifugation in 2 ml or 20 ml Vivaspin concentrators (Sartorius AG) with a 10 kDa cut off at 4 °C at 3300 rpm and passed through a 0.2 μm filter prior to loading onto the size exclusion column. Proteins were eluted from the size exclusion column with SEC buffer at a flow rate of 0.5 ml/min. Fractions of 0.5 ml were collected and fractions corresponding to the peak/s at 280 nm on the resulting chromatogram were analysed using SDS-PAGE for the protein of interest (section 2.3.1).

2.3.7 Enzyme Assays

2.3.7.1 Standard Malachite Green Assay

Malachite green assay reactions were set up in 96 well microplate (Greiner Bio-One) (Baykov et al. 1988). The reaction mixture consisting of the substrates and assay buffer (50 mM Tris-HCl pH 7.5 for MurA and 300 mM Tris-HCl pH 7.5 for AroA) was pre-equilibrated by incubating the plates at the reaction temperature for 10 minutes. Reactions were started with the addition of enzyme, thus making the total volume of 100 μl, and incubated at the reaction temperature for a set reaction time before being stopped by the addition of 30 μl of stop solution (9.4 % (w/v) malachite green, 13 % (v/v) sulphuric acid, 1.5 % (w/v) ammonium molybdate and 0.16 % Tween 80). Colour development was achieved by incubating the plates at RT for 5-10 minutes and the absorbance at 620 nm measured using a Multiskan GO microplate spectrophotometer (Thermo Scientific). All the reactions were performed in triplicate along with blank reactions containing no protein.

2.3.7.2 Phosphate Standard

A 10 mM stock of KH_2PO_4 was prepared and diluted to make a working stock of 100 μM. The working stock was diluted serially six times in assay buffer in a 1:1 ratio to give a concentration range of 50 μM, 25 μM, 12.5 μM, 6.25 μM, 3.13 μM

and 1.56 μM in a total volume of 100 μl. The standard solutions were prepared in duplicate and the malachite green assay was performed as described in section 2.3.7.1. The absorbance measured at 620 nm for 100 μl of assay buffer (Blank: 0 μM KH_2PO_4 concentration) was subtracted from the absorbance measured for the phosphate samples. The resulting data were plotted in Excel and a standard curve was generated.

2.3.7.3 Time Course Assay

A time course assay was performed to determine the reaction time, which is the time for which the reaction was allowed to proceed before adding the stop solution. A master mix was prepared in a microcentrifuge tube which consisted of a non-limiting amount of substrate (Substrate K_M of MurA enzyme was taken from Blake et al (2009) and substrate K_M of AroA enzyme was taken from Priestman et al (2005)) and assay buffer. The reaction mix was incubated at 37 °C for 10 minutes before adding the enzyme. Samples (100 μl) were taken at different time points ranging from 0-10 minutes, stopped with 30 μl of stop solution and absorbance measured at 620 nm. A graph of absorbance at 620 nm versus time was plotted in Excel and the reaction time was determined manually from the graph.

2.3.7.4 Michaelis-Menten Kinetics

Preliminary trials were performed to determine the appropriate enzyme concentration and concentration range of the substrates for the determination of Michaelis-Menten kinetics using the malachite green assay. Stock solutions of substrates were prepared in assay buffer and enzymes were diluted to appropriate concentration for the assays and stored in SEC buffer. The AroA and MurA enzymes catalyse reactions in the presence of two substrates. Hence, to determine the K_M for each of the substrates, the concentration of one of the substrates was maintained in excess while altering the concentration of the other. The total reaction volume was maintained at 100 μl and the malachite green assay was performed as explained section 2.3.7.1.

The Michaelis-Menten kinetics were initially determined at 37 °C and the K_M obtained for each substrate was used as a reference to determine how much excess

(generally 5x K_M or 10x K_M whenever possible) should be added in assays performed at higher temperatures.

The optimum temperature for enzyme activity (T_{opt}) was measured as below in section 2.3.7.5. The Michaelis-Menten kinetics were also measured at the T_{opt} of each enzyme, to obtain comparable data between enzymes. Graphpad Prism version 5.01 (Graphpad Software) was used to fit the Michaelis-Menten equation to the data and determine the K_M and V_{max} values. The V_{max} generated by Prism, in absorbance/s format, was converted to concentration/s format using the phosphate standard curve generated in section 2.3.7.2 prior to calculating k_{cat} of the reaction.

2.3.7.5 Determination of Optimum Temperature (Thermoactivity T_{opt})

The optimum temperature of enzyme activity was determined by performing the AroA/MurA reactions at different temperatures and their activity measured using malachite green assay as described in section 2.3.7.1. The reactions were set up with non-limiting amounts of substrate based on the Michaelis-Menten kinetics measured at 37 °C. Initial trials were performed at a broad temperature range from 20 °C to 80 °C at 10 °C intervals. A narrower temperature range was then investigated. Blank reactions were set up for every temperature to eliminate any contribution of temperature to free phosphate. The optimum temperature of enzyme activity was obtained by plotting the reaction rate at each temperature. The assays were repeated multiple times to confirm the validity of the results.

2.3.7.6 Half Maximal Inhibitory Concentration (IC50)

The AroA/MurA enzyme reactions in the presence of their respective inhibitors were set up and enzyme activity was quantified using the malachite green assay to measure the amount of inhibitor needed to inhibit the enzyme reaction by 50 %. Inhibitor stocks were made up in assay buffer and serially diluted to provide stocks of varying concentrations. The 100 µl reaction was set up with constant amounts of substrates, varying concentrations of the inhibitor and assay buffer. The enzyme was added last to start the reaction. The concentration of enzyme added was kept constant at approximately 6 µg/ml in all the IC_{50} assays. Graphpad Prism version 5.01 (Graphpad Software) was used to plot the log inhibitor concentration against enzyme activity and found the IC_{50} using the '*log(inhibitor) vs. response -- Variable slope'* function in Prism.

2.3.8 Determination of Thermostability

2.3.8.1 Differential Scanning Calorimetry (DSC)

Thermostabilities of enzymes were measured by differential scanning calorimetry (DSC) in a TA Instruments Nano DSC at the University of Canterbury, Christchurch. The sample and reference cells were loaded with their respective samples and the system was set to ramp up from 20 °C to 90 °C at 1 °C/min. For the blank runs, both the sample and reference cells were loaded with the buffer the protein was stored in. For the sample runs with protein, 0.5 mg/ml of enzyme was loaded in to the sample cell and its respective buffer in the reference cell. The NanoAnalyze software was used to set the baseline and find the peak representing the T_m in power input for each enzyme. The raw data were exported to Excel and a graph was plotted after subtracting the baseline data from the analysis data.

2.3.8.2 Protein Melt Analysis

The estimation of T_m was also achieved for some enzymes using a RotorGene-6000 Real-time PCR machine (Corbett Research). To prepare the real-time melt reaction mixtures, buffer solutions of different pHs ranging from pH 4 to pH 10 were prepared (Appendix E2). The 5000x SYPRO orange protein stain dye (Sigma Aldrich) was diluted to 300x using each different pH buffer. A volume of 7.5 µl diluted dye was transferred into PCR tubes. Protein was added to the dye at a final concentration of 0.1 mg/ml, 0.2 mg/ml or/and 0.4 mg/ml in total volume of 25 µl and the reaction was made up to the total volume with the appropriate buffers. Blanks were set up containing dye and buffer of different pH at a volume of 25 µl.

The tubes were placed into a 36-tube rotor in the RotorGene-6000 Real-time PCR machine. The run was set up to increase the temperature from 25 °C to 99 °C with an increment of 0.2 °C with 5 seconds delay between each increment. The fluorescence of the samples at an excitation wavelength of 470 nm and an emission wavelength at 555 nm was monitored. The resulting data were exported from the RotorGene software version 1.7 (Corbett Research) to an Excel spreadsheet, blank data subtracted from sample data and fluorescence vs. temperature data plotted. The first derivative of each dataset was taken and

plotted using Prism. The T_m was then determined manually to be the middle of the peak, which is equivalent to the inflection point of the original melt slope.

2.3.9 HPLC Purification of S3P

The column used for the HPLC purification of shikimate-3-phosphate (S3P) was a DiscoveryBIO wide pore C18 HPLC column (Sigma-Aldrich). The column was maintained at 40 °C with a flow rate of 1 ml/min. The injection volume was 200 µl and S3P was monitored at 215 nm. Multiple mobile phases were attempted: 0.1 % trifluoroacetic acid (TFA), 0.1 % TFA in the presence of an ion pairing reagent (5 mM tetramethylammonium bromide) and 10% acetonitrile (ACN).

2.4 *In Vivo* Enzyme Characterisation

The parent strain *E. coli* K-12 BW25113 and its Δ*aroA* derivative were obtained from the Keio collection (Baba et al. 2006). Electrocompetent cells were made using the protocol described in section 2.2.8.1. The plasmid pPROEX HTb containing the *Bacillus* AroA LCA gene and pPROEX HTb with *B. subtilis* AroA gene were transformed into the Δ*aroA* strain as described in section 2.2.8. As controls, the empty pPROEX HTb plasmid was transformed into the parent strain and the Δ*aroA* strain.

To test the complementation of the Δ*aro*A strain by *B. subtilis aroA* and *Bacillus* LCA *aroA*, the strains were grown in 1 ml of LB broth with appropriate antibiotic at 37 °C shaking at 200 rpm. A final concentration of 100 µg/ml of ampicillin and 30 µg/ml of kanamycin was added to all the cells except the parent strain with empty pPROEX HTb which was grown in absence of kanamycin. The cells were pelleted by centrifugation at 4 °C for 10 minutes at 4000 rpm and the supernatent discarded. The cells were washed with 1 ml of M9 minimal media (Appendix E3) in the presence of antibiotic. Finally, the cells were pelleted again and resuspended in 1 ml of M9 minimal media to remove all the residual LB.

The cells were serially diluted to 1/10, 1/100, 1/1000 and 1/10,000 concentrations using M9 minimal media containing appropriate antibiotic, a final concentration of 1 mM IPTG and glucose as the sole carbon source. The undiluted culture and each dilution was spotted onto an M9 agar plate containing antibiotic, a final concentration of 1 mM IPTG and glucose as the carbon source. Each culture was

spotted by pipetting 20 μl drops onto the plates. The plates were grown for 24 to 36 hours at 37 °C and the growth monitored.

2.5 Protein Crystallography

2.5.1 Determination of Initial Crystallisation Conditions

For the purpose of laying down crystal trays to determine the initial crystallisation conditions for the protein, purified protein after size exclusion chromatography was concentrated to approximately 200 μl. Four sets of crystallisation screens, PEGRx HT - HR2-086, Crystal Screen HT - HR2-130, Index HT - HR2-134, and SaltRx HT - HR2-136 (Hampton Research), were used to test for the best condition by the sitting drop method. Each screen contained 96 crystallisation conditions and 100 μl of each of these solutions was transferred into the reservoirs of four 96-2 low profile Intelli-Plate protein crystallisation plates (Hampton Research). A Mosquito crystallisation robot (TTP LabTech Ltd.) was used to pipette 100 nl of protein and 100 nl of mother liquor before dispensing onto the sitting drop wells of the Intelli-Plate. The plates were sealed using a ClearSeal film (Hampton Research) and incubated at 18 °C. The drops were checked regularly for crystal growth.

2.5.2 Fine Screening to Optimize Crystallisation Conditions

2.5.2.1 Standard Fine Screens

The hanging drop vapour diffusion method of crystallisation was used to optimise the crystallisation conditions. The top of each well of a 24-well VDX plates (Hampton Research) was greased using glissealN grease (Borer Chemie) and based on the crystallisation conditions that grew the best crystals among the four robot screens, 500 μl of mother liquor varying in pH, and concentration of other components were added to the wells. One microlitre of protein was laid onto the center of a 22 mm siliconised cover slip (Hampton Research) and mixed with 1 μl of mother liquor taken from each respective well. The cover slip was then inverted over the pre-greased well containing the mother liquor and pressed down to seal. Fine screens plates were incubated at 18 °C and checked for crystal growth as for robot screens.

2.5.2.2 Additive Fine Screens

Once an optimised crystallisation condition had been determined, a screen of 14 different additives was performed by adding each additive to the crystallisation condition and setting up the drops as described in section 2.5.2.1. The additives used and their final concentration are given in (Table 2.2).

Table 2.2: Additives used in additive screens

Additive	Final Concentration
$Co(II)Cl_2.6H_2O$	0.01 M
$MgCl_2.6H_2O$	0.01 M
$Ni(II)Cl_2.6H_2O$	0.01 M
$CsCl_2$	0.1 M
Glycine	0.1 M
EDTA	0.01 M
Glucose	3 %
Sucrose	3 %
Glycerol	3 %
PEG 400	5 %
Ethanol	3 %
Methanol	3 %
Acetonitrile	4 %
Butanol	0.7 %

2.5.2.3 Trypsin treatment

Purified protein was mixed with trypsin (1 mg/ml stock) at a 1:500 mass ratio and immediately used for the hanging drop diffusion method (section 2.5.2.1). The drops were set up by mixing 1 µl of protein-trypsin mixture with 1 µl of mother liquor from respective wells.

2.5.2.4 Seeding

The batch seeding method was used to improve the quality of the crystals. A fine screen drop containing crystals or microcrystals was transferred into a microcentrifuge tube, mixed with 18 µl of mother liquor, thus diluting it 10-fold, and thoroughly vortexed to produce a seed stock. The seed stock was then serially diluted to make stocks of 1:10, 1:100 and 1:1000. Finally, the seed stock and the diluted stocks were mixed with protein in a 1:1 ratio and placed over the reservoir for hanging drop diffusion method (section 2.5.2.1).

2.5.3 Data Collection

Cryo-protectant solutions of the mother liquor containing 5 %, 10 %, 15 % and 20 % (v/v) glycerol were prepared. Protein crystals were picked up from their original drops using a cryo-loop (Hampton research) and soaked through each of these cryo-protectants or directly into the mother liquor containing 20 % (v/v) glycerol. The cryo-protectant soaked crystals were snap frozen using liquid nitrogen and transferred into pucks.

X-ray diffraction data were collected at the Australian Synchrotron, Melbourne, Australia using the MX1 or MX2 beam line. An ADSC Quantum 210r detector (Area Detector Systems Corp.) was used to measure reflections. The strategy function of MOSFLM (Leslie & Powell 2007) was used in certain cases to assist optimal data collection. As the data was collected, each dataset was auto processed at the Australian Synchrotron.

2.5.4 Data Processing

The X-ray diffraction data collected were indexed using XDS and scaled using Aimless (Evans & Murshudov 2013) at the Australian Synchrotron. The resulting log file was examined for the completeness, mosaicity and resolution of the collected dataset. Once these parameters were verified, the .mtz file was directly used for the following steps.

2.5.4.1 Truncate

The programme Truncate (French & Wilson 1978) in the CCP4 suite (Winn et al. 2011) was used to ensure unique data and to generate a FreeR set for 5 % of the data.

2.5.4.2 Matthews Coefficient

The output file from Truncate was used to determine the number of monomers present in the asymmetric unit using MATTHEWS_COEF programme (Matthews 1968) in the CCP4 suite (Winn et al. 2011). The programme calculated the solvent occupancy of the unit cell which was used to determine the number of monomers in the asymmetric unit.

2.5.4.3 Molecular Replacement

Molecular replacement was performed using Phaser (McCoy et al. 2007) within the CCP4 suite (Winn et al. 2011). Suitable models for molecular replacement were obtained by finding structural homologues *via* NCBI BLAST (Johnson et al. 2008). Molecular replacement for *Bacillus* MurA LCA structure was performed using the structure of *Bacillus anthracis* MurA (PDB 3SG1) and molecular replacement for *Bacillus* AroA LCA structure was performed using the structure of *Bacillus halodurans* AroA (PDB 3RMT). For datasets with high Rmerge, a resolution cutoff was set up during this step.

2.5.4.4 Model Building and Refinement

Automated model building was achieved using Buccaneer (Cowtan 2006) in the CCP4 suite (Winn et al. 2011) for *Bacillus* MurA LCA and using Autobuild (Terwilliger et al. 2008) within PHENIX (Adams et al. 2010) for *Bacillus* AroA LCA. Iterative rounds of manual building was performed using COOT (Emsley & Cowtan 2004) with electron density maps contoured to 1 σ. Model refinement was achieved using refmac5 (Murshudov et al. 2011) in CCP4 suite (Winn et al. 2011) and/or PHENIX Refine (Afonine et al. 2012) within PHENIX (Adams et al. 2010).

2.5.4.5 Structural Analysis

All structural images illustrated in this thesis were generated using PyMOL version 2.5.4 (Schrödinger 2010). The average B factors and total number of atoms in the structure were obtained using the Baverage programme in the CCP4 suite (Winn et al. 2011). PDBePISA (Krissinel & Henrick 2007) was used to determine the interface interaction between two subunits involved in an oligomeric formation. PDBsum (Laskowski 2001) was used to determine the bonds involved in the protein-substrate interaction. The closest structural

homologues from the PDB archive were identified using PDBeFOLD (Krissinel & Henrick 2004).

Chapter 3

MurA

3.1 Introduction

Many studies have focused on exploring the structure and function of the MurA enzyme as it is a very attractive antibiotic target. In spite of all these studies, very little is known about the evolution of MurA. I chose ancestral sequence reconstruction (ASR), as a powerful tool to study protein evolution of MurA. This has previously been used for the reconstruction of ancestral enzymes such as β-lactamases (Risso et al. 2013) and structurally complex core metabolic enzymes such as LeuB (Hobbs et al. 2012).

Comparing the structural and functional properties of ancestral MurA enzymes reconstructed by ASR to those of MurA contemporary enzymes can provide insight into the evolution of MurA. Functional characterisation of the MurA enzymes includes determination of the enzyme's kinetic parameters, inhibitory properties and thermal properties such as thermostability and thermoactivity. The most common way of determining such properties for MurA is by monitoring the inorganic phosphate released by the MurA reaction. This can be done using either a MurA-MurB coupled assay (Biery 2007), continuous spectroscopic assay using purine nucleoside phosphorylase (PNP) and 2-amino-6-mercapto-7-methylpurine riboside (MESG) (Molina-López et al. 2006; Blake et al. 2009) and/or the most commonly used method, the malachite green assay (Lanzetta et al. 1979; Du et al. 2000; Baum et al. 2001; Jin et al. 2009). Due to its simplicity, effectiveness and sensitivity, a modified malachite green method was used to characterise the MurA enzymes (Baykov et al. 1988).

The thermal properties of MurA enzymes also provide information on the host organism growth temperature (Gromiha et al. 1999). Such a correlation can be exploited to study the evolution of thermophily using ancestral MurA enzymes reconstructed by ASR. The *Bacillus* genus is an ideal model to study the evolution of thermophily as the contemporary species live in a wide range of thermal habitats. Furthermore, ancestral reconstruction was found to have been previously successful using the *Bacillus* genus (Hobbs et al. 2012). Therefore *Bacillus* MurA

ancestral and contemporary enzymes were used as the candidates for the evolutionary study of MurA.

This chapter describes the kinetic, inhibitory and thermal properties of ancestral *Bacillus* MurA in comparison to selected contemporary *Bacillus* MurA enzymes to shed some light on the evolution of MurA. Additionally, the thermal properties of reconstructed ancestral MurA enzymes are used to study the evolution of thermophily.

3.2 Results and Discussion

3.2.1 Phylogenetic Analysis and ASR

3.2.1.1 Sequence Alignment

All the amino acid and nucleotide sequences of *Bacillus* species available as at November 2011 were collected from online databases (Accession numbers, species and strain details are provided in Appendix A). A total of 20 MurA sequences from *Bacillus* species, inhabiting mesophilic and moderately thermophilic environments, were retrieved. The alignment of these sequences, along with two *Clostridium* MurA sequences (outgroup sequences) has an overall 33.3 % of identical sites. The exclusion of the *Clostridium* MurA outgroup sequences from the alignment revealed that the *Bacillus* MurA sequences are well conserved (50.2 % identical sites). The inclusion of the outgroup in the alignment only decreases identical sites by 16.9 %, thus proving that the outgroup chosen is appropriate by being close enough to the ingroup sequences to provide a robust alignment but at the same time distant enough to be a clear outgroup (Salemi & Vandamme 2003). Out of the 20 sequences used for the alignment, seven MurA sequences were taken from *Bacillus* species that inhabit a temperature range of 25-30 °C, four sequences from 45-50 °C and nine sequences from 37 °C.

Various conserved residues found in MurA have been identified in the literature. The *Bacillus* MurA alignment showed the presence of many universally recognised conserved residues (the alignment given in Appendix B). For example, Arg401 contributes to the phosphate binding site, Asp50 is involved in a salt bridge in *E. coli* MurA (Skarzynski et al. 1996) and Leu34 forms a hydrophobic pocket with other residues to connect two of the MurA structural subunits

(Schönbrunn et al. 1996). Many conserved residues that have been identified to be involved in ligand binding interactions such as Lys22, Arg124, Asp373, Leu374, Cys119 and Asp309 (Schönbrunn et al. 1996; Skarzynski et al. 1996; Takahata et al. 2010; Shahab et al. 2014) are also conserved among *Bacillus* MurA sequences (numbering corresponds to the sequence alignment in Appendix B).

The high sequence identity between the *Bacillus* MurA sequences implies a higher accuracy for the phylogeny that is inferred from them (Kumar & Filipski 2007). This also confirms that these sequences have evolved from a common ancestor and phylogenetic analysis based on these sequences will result in a biologically relevant phylogenetic tree (Brown 2002).

3.2.1.2 Phylogenetic Analysis

The characterisation of protein evolution requires the determination of an appropriate model of evolution which gives the probability of change from one amino acid to another. ProtTest version 2.4 was used to choose the best model of amino acid evolution for MurA sequences based on the alignment of the MurA sequences (Abascal et al. 2005). ProtTest chose the LG model as the best model for evolution, which is a sophisticated model that incorporates evolutionary rate variability across sites in the matrix estimation (Le & Gascuel 2008).

The model of evolution and amino acid alignment were used in GARLI 1.0 (Zwickl 2006) to generate five phylogenetic trees based on the ML criterion. Each phylogenetic tree was rooted with *Clostridium* sequences as the outgroup. Four out of the five phylogenetic trees had equal log likelihood score of -6055 with identical positions of tree branches, suggesting it to be the most plausible branching pattern and therefore phylogeny.

The phylogeny of *Bacillus* MurA was confirmed with the help of bootstrap percentages and by comparison to other *Bacillus* phylogenetic trees from the literature (Nakamura & Jackson 1995; Fritze & Pukall 2001; Nogi et al. 2005; Alcaraz et al. 2010; Hobbs et al. 2012). Bootstrapping was achieved as described in section 2.1.3.2. The bootstrap values ≥50% are displayed in Figure 3.1.

The MurA phylogenetic tree (Figure 3.1) reproduced the previously observed phylogenetically conserved clades: the soil clade, the pathogenic clade and the halophile clade (Alcaraz et al. 2010). The positions of species in the MurA

phylogenetic tree are consistent with phylogenetic trees constructed by Alcaraz et al. (2010) and Hobbs et al. (2012). The phylogenetically conserved clades were identified in both *Bacillus* MurA and LeuB phylogenetic trees (Figure 3.1 and Figure 3.2). *Bacillus atrophaeus* is very closely associated with *Bacillus subtilis* and is therefore grouped within the soil clade (Fritze & Pukall 2001). *Bacillus cereus* and *Bacillus thuringiensis* are grouped together and the position of *Bacillus mycoides* in the pathogenic clade is consistent with previous taxonomic classifications (Nakamura & Jackson 1995). Based on 16s rRNA analysis, *Bacillus pseudofirmus* has been shown to be part of the halophile clade along with *Bacillus halodurans* and *Bacillus clausii* (Nogi et al. 2005).

Four ancestors were chosen from the phylogenetic tree (Figure 3.1) for sequence reconstruction with each ancestor progressively going back in evolutionary time. LCA represents MurA from the last common ancestor of the *Bacillus* genus. ANC3 was chosen to be reconstructed as it only differs from ANC2 by the presence of *Bacillus selenitireducens,* yet it is more distant from ANC2 and closer to LCA. Therefore, analysis of the properties of ANC3 would not only provide information on the evolution of MurA, but allows us to determine any bias imposed by the *B. selenitireducens* branch. The absence of such bias would also concur with realistic ancestral reconstruction.

The age of the nodes was determined by converting the phylogram to a chronogram. Unlike a phylogram, in which branch lengths represent the number of amino acid changes per site, the branch lengths in a chronogram represent evolutionary time. The program R8s version 1.71 (Sanderson 2003) was used to generate the chronogram. Fixing one or more time nodes, based on information obtained from fossils or other sources, allows R8s to generate absolute rates and ages (Sanderson 2003). The chronogram depicting the timescale of prokaryote evolution (Battistuzzi et al. 2004) was used as a reference to find the node times necessary to fix in R8s program. The divergence point of *Clostridium* and *Bacillus* was fixed to be 2650 myr and *B. halodurans* and *B. subtilis* diverge at 950 myr. Fixing these two node times allowed R8s to estimate age of the remaining nodes (Figure 3.1).

Comparing the MurA ancestral nodes to the reconstructed LeuB ancestors (Hobbs et al. 2012), MurA LCA (950 myr) can be considered to occupy a temporally and

phylogenetically equivalent position to the LeuB4 (Figure 3.2). MurA ANC2 could also be considered to occupy a phylogenetically similar position to LeuB3 as they both represent the pathogenic clade, soil clade and *Bacillus megaterium*. The high bootstrap percentages of 100 % and 92 % for MurA LCA and MurA ANC2 also suggest the robustness of these phylogenies.

The optimal growth temperatures of each *Bacillus* species in the MurA tree were taken from the literature (Gordon 1977; Nakamura 1998; Fritze & Pukall 2001; Denizci et al. 2004; Nogi et al. 2005; Cerritos et al. 2008; Ju et al. 2009) and are represented by coloured circles as in Figure 3.1. It was observed that, in the phylogenetic tree, species with the same growth temperature are not grouped together but instead species with different growth temperatures are dispersed evenly throughout the phylogenetic tree.

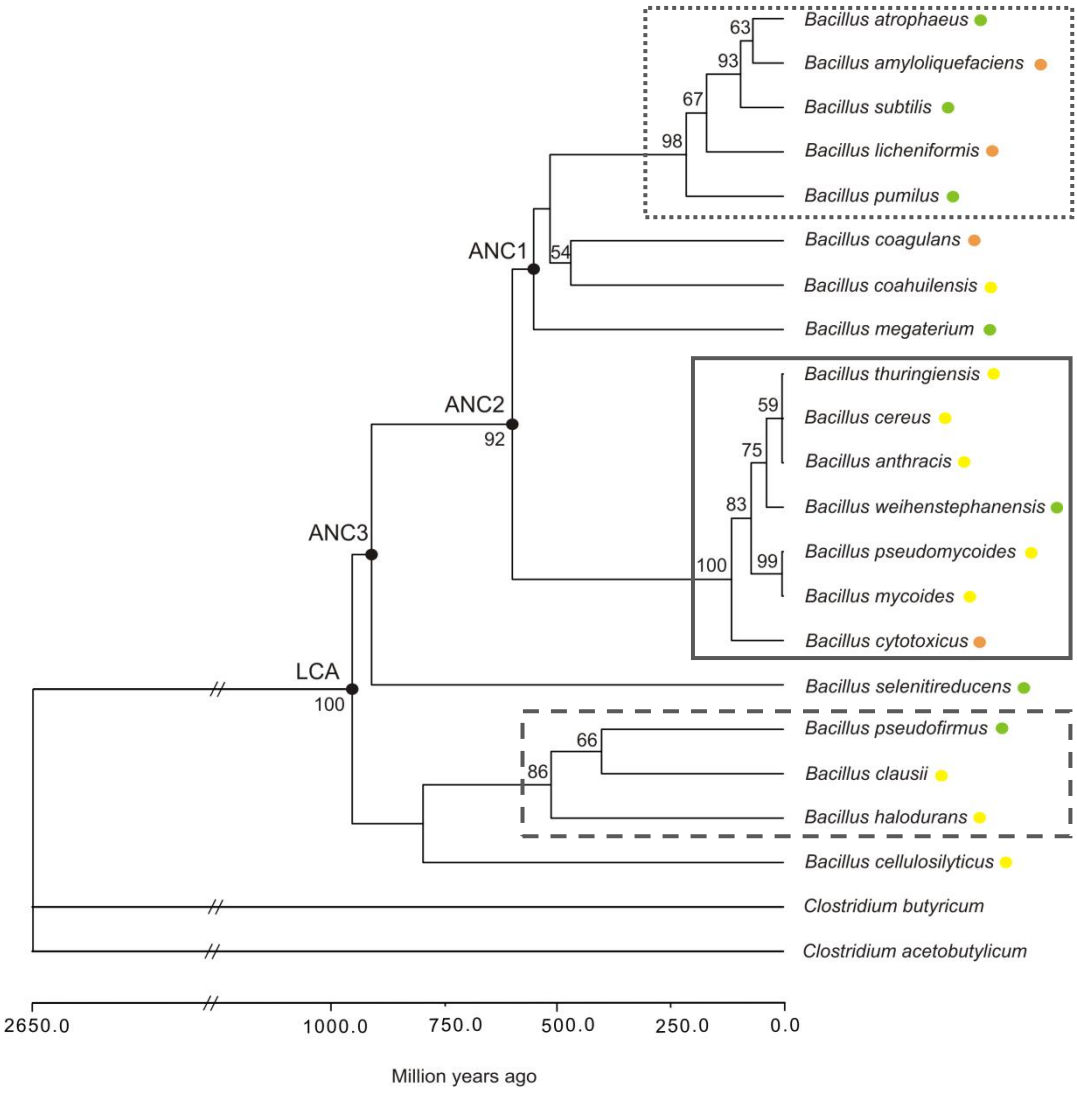


Figure 3.1: ML chronogram of *Bacillus* species based on MurA amino acid sequences.
The phylogenetic tree was rooted with *Clostridium* sequences as outgroup. The reconstructed ancestral enzymes are indicated by black circles. The coloured circles next to each *Bacillus* species relates to their optimal growth temperatures: 45-50 °C (orange), 37 °C (yellow) and 25-30 °C (green). The phylogenetically conserved clades are indicated in boxes: the pathogenic clade (solid line), the soil clade (dotted line) and the halophile clade (dashed line) (Alcaraz et al. 2010). The bootstrap values ≥ 50%, assessed from 1,024 bootstrap replicates are also shown.

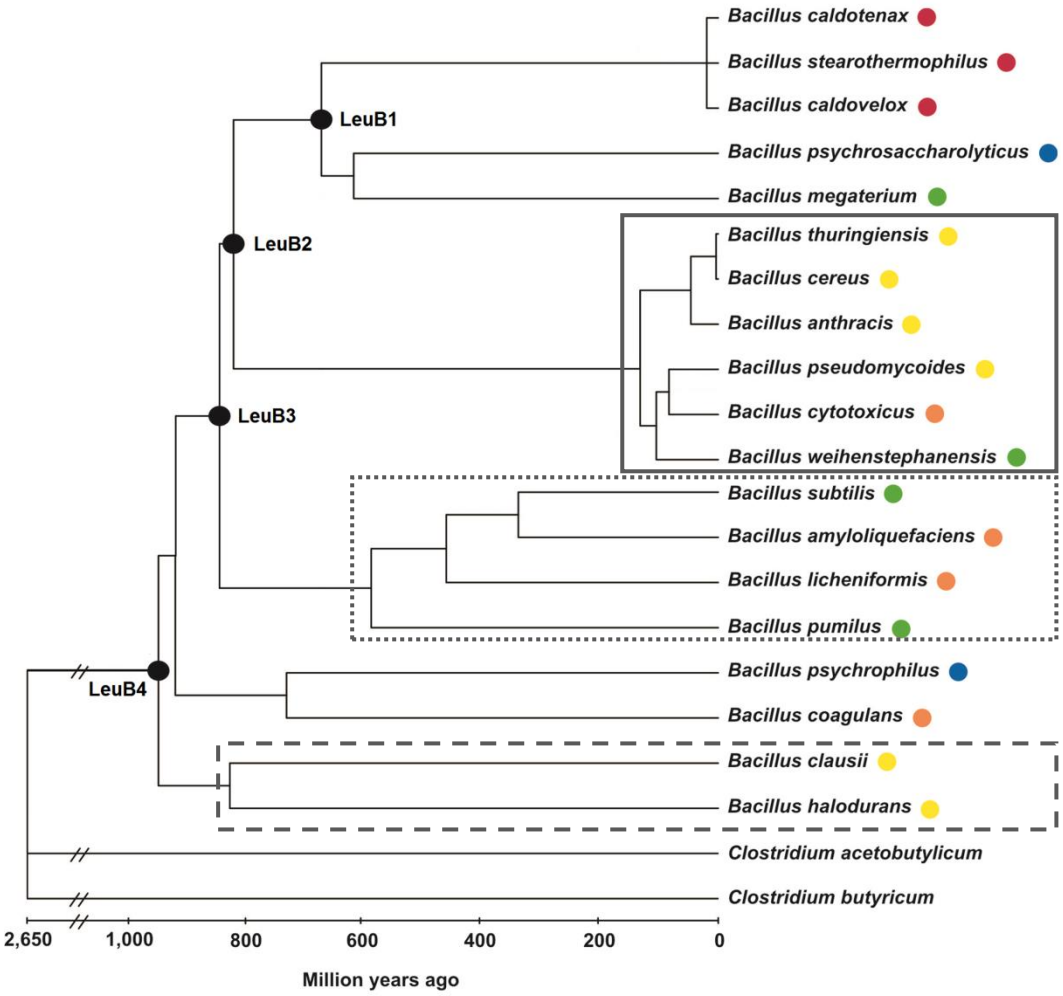


Figure 3.2: ML chronogram of *Bacillus* species based on LeuB amino acid sequences (Hobbs et al. 2012).
The reconstructed ancestral enzymes are indicated by black circles. LeuB1-LeuB4 were named as ANC1-ANC4 in the LeuB study (Hobbs et al. 2012). The coloured circles next to each *Bacillus* species relates to their optimal growth temperatures: 60-80 °C (red), 45-50 °C (orange), 37 °C (yellow), 25-30 °C (green) and 20 °C (blue). The phylogenetically conserved clades are indicated in boxes: the pathogenic clade (solid line), the soil clade (dotted line) and the halophile clade (dashed line) (Alcaraz et al. 2010).

The generation of an ML chronogram for MurA was succeeded by the inference of the sequences corresponding to MurA ancestral nodes.

3.2.1.3 Ancestral Inference

Inference of sequences for the four ancestral nodes was performed by the ML criterion. Inference was achieved using three methods: amino acid, nucleotide and codon inference methods. The most appropriate model of nucleotide evolution was determined by jModelTest (Posada 2008). The best model was chosen to be the general time reversible model (GTR model) of nucleotide substitution for the

nucleotide inference (Tavaré 1986) and the Jones amino acid rate file was employed for amino acid and codon inference (Jones et al. 1992).

The ancestral inference was performed using PAML version 4.3 (Yang 2007). Geneious was used to translate the sequences inferred by the codon and nucleotide methods and together with sequence from the amino acid inference, the three amino acid sequences were then aligned using ClustalW (Larkin et al. 2007). For all the four nodes, the three inference methods were in agreement for ≥89% of sites, and for a further 7-10% of sites two out of the three methods were in agreement. Any remaining ambiguous sites, which accounted for ≤8 sites, were resolved as per the criteria employed by Hobbs et al. (2012), by 1) considering the distribution of amino acids in the original alignment and choosing the amino acid present in majority of the sequences, or the amino acid present in a functional extant protein, 2) choosing amino acids with conserved physico-chemical characteristics between different contemporary species and by 3) considering the structure of extant proteins and analysing the implication of each ambiguous amino acid within the structural environment (section 2.1.4.3). The average posterior probability at each site from the amino acid method for all the ancestral sequences was >0.95. An alignment of the inferred ancestral sequences is given in Figure 3.3.

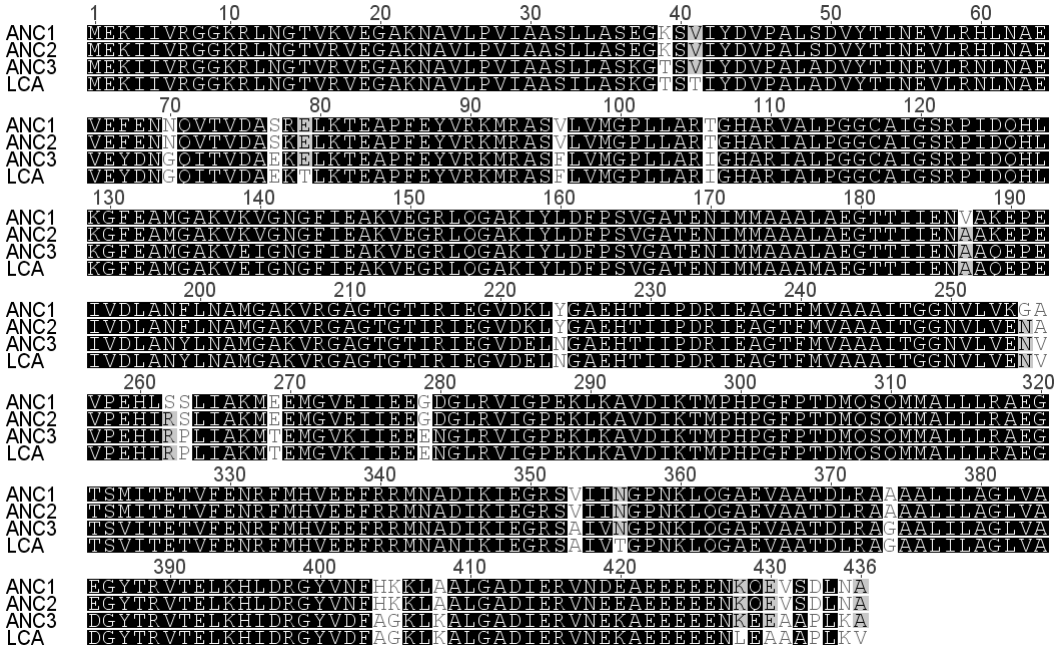


Figure 3.3: Alignment of inferred ancestral MurA sequences
The different shades indicate the percentage similarity at each position in the alignment. Black = 100% similar, dark grey = 80 to 100% similar, light grey = 60 to 80% similar and white = less than 60% similar.

The pairwise identities of the above inferred ancestral sequences to the extant protein sequences are given in Table 3.1. As expected from the phylogenetic tree, ANC1 and ANC2 are more similar to each other than they are to ANC3 and LCA and vice versa. The sequence identities between the ancestral sequences and the contemporary sequences range from 76 % to 90 %. Even though the similarities are high, the fact that the ancestral sequences are not significantly more similar to one contemporary sequence than the others suggests that there is no bias on any of the ancestral nodes by extant sequences. It was also noted that ANC3 is most similar to *Bacillus cellulosilyticus* and not to *B. selenitireducens*, thus suggesting that ANC3 is not biased by the *B. selenitireducens* branch. The high percentage identity of ancestral sequences to the extant sequences suggests a structural and functional conservation and therefore greater chances of resulting in functional ancestral proteins.

Table 3.1: Pairwise sequence identities (%) between extant and ancestral protein sequences.

	ANC1	ANC2	ANC3	LCA
ANC1	-------	98	89	87
ANC2	98	-------	91	89
ANC3	89	91	-------	98
LCA	87	89	98	------
Bacillus amyloliquefaciens	87	85	79	78
Bacillus anthracis	86	88	82	80
Bacillus atrophaeus	87	85	79	78
Bacillus cellulosilyticus	80	82	**89**	**90**
Bacillus cereus	86	88	82	80
Bacillus clausii	78	**78**	81	82
Bacillus coagulans	83	82	**78**	**76**
Bacillus coahuilensis	84	83	**78**	**76**
Bacillus cytotoxicus	88	**89**	82	81
Bacillus halodurans	79	79	83	85
Bacillus licheniformis	87	86	80	78
Bacillus megaterium	**89**	88	81	80
Bacillus mycoides	86	87	81	80
Bacillus pseudomycoides	86	87	81	80
Bacillus pseudofirmus	78	**78**	83	84
Bacillus pumilus	86	84	79	77
Bacillus selenitireducens	**76**	**78**	84	84
Bacillus subtilis	87	85	79	78
Bacillus thuringiensis	86	88	82	80
Bacillus weihenstephanensis	86	87	81	79

Blue cell: Highest pairwise identity between ancestral and extant sequences
Green cell: Lowest pairwise identity between ancestral and extant sequences

The high posterior probability and low number of ambiguous sites of the ancestral sequences indicate a higher confidence in the ancestral inference given the model of evolution used. Furthermore, the high sequence identity between ancestral sequences and contemporary sequences suggest slow evolutionary rates and

therefore, a higher chance for the successful reconstruction of active ancestral MurA enzymes. Therefore, the process of characterising the ancestral enzymes was initiated by cloning the genes encoding these ancestral proteins followed by their expression and purification.

3.2.2 Cloning, Protein Expression and Purification

3.2.2.1 Cloning

The genes encoding for *Bacillus* MurA ANC1, ANC2, ANC3, LCA and *Bacillus coagulans* MurA proteins were cloned into pPROEX HTb using the restriction sites *Bam*HI and *Xho*I, resulting in expression of the protein with an N-terminal His-tag. *B. subtilis* MurA gene was amplified from genomic DNA of *B. subtilis* 168 and cloned into pPROEX HTb using the restriction sites *Nco*I and *Pst*I. pPROEX HTb containing the gene of interest was used to transform either *E.coli* DH5α or BL21 cells and screened for positive transformants as described in section 2.2.8.

3.2.2.2 Small-scale Protein Expression

The small scale expression trial of *Bacillus* MurA LCA (Lysis Buffer: 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM imidazole) showed that the protein expressed but most was found in the insoluble fraction rather than the soluble fraction (Figure 3.4). A range of different buffers at a variety of pHs tested did not increase the solubility of the expressed MurA LCA protein. However, the amount of soluble protein obtained was enough to perform the enzyme characterisation assays. An increase in the amount of protein expressed with time was observed in trials performed for *Bacillus* MurA ANC1, ANC2, ANC3, LCA and *Bacillus coagulans* MurA (Figure 3.4). The expression of *B. subtilis* MurA was confirmed *via* an expression trial and subsequent Ni^{2+}-sepharose pull down (Figure 3.4) (section 2.3.3.2).

Protparam (Gasteiger et al. 2005) predicted the molecular weight of the monomeric form of each of the six MurA enzymes to be approximately 47.0 KDa. This molecular weight was used as a reference while analysing the SDS-PAGE gel bands.

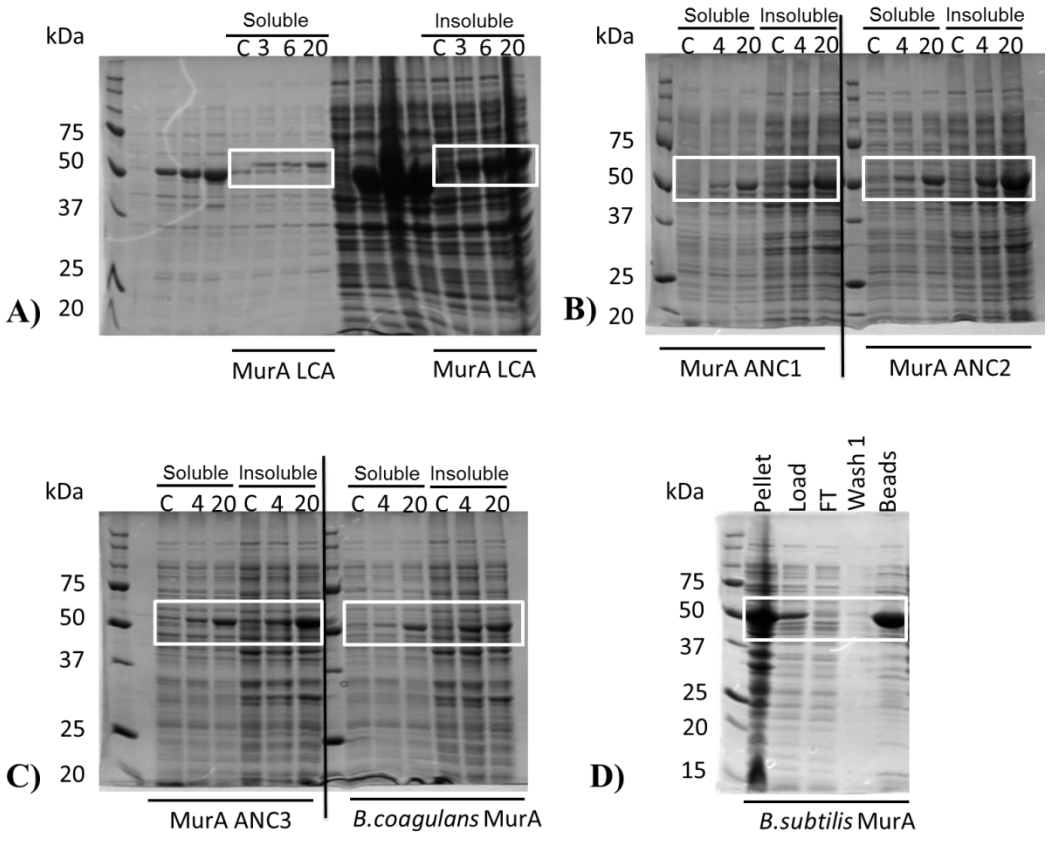


Figure 3.4: SDS-PAGE gels of small-scale expression trials for MurA enzymes.
A) *Bacillus* MurA LCA soluble and insoluble expression trial B) MurA ANC1 and ANC2 soluble and insoluble expression trial C) MurA ANC3 and *Bacillus coagulans* MurA soluble and insoluble expression trial D) *Bacillus subtilis* MurA Ni^{2+}-sepharose bead expression trial. White boxes indicate the expected band size. Labels for A), B) and C): C - control sample taken prior to induction; Numbered lanes indicate the number of hours after which the samples were taken post induction (3, 4, 6 and 20 hours). Labels for D): FT- flow through.

3.2.2.3 Protein Purification

Large scale protein expression of *Bacillus* MurA enzymes was achieved overnight at 18 °C. All the *Bacillus* MurA proteins were purified using Ni^{2+} affinity chromatography (IMAC). The bound MurA proteins were eluted using a gradient of increasing concentrations of elution buffer (50mM Tris, 300mM NaCl, 1M imidazole pH 7.5). All the six MurA proteins started eluting at around 14 % of elution buffer (140 mM imidazole). The fractions corresponding to the chromatographic peaks were run on an SDS-PAGE gel to confirm the size and purity of the proteins. A representative UV trace and SDS-PAGE gel for Ni column purification of MurA proteins is given in Figure 3.5.

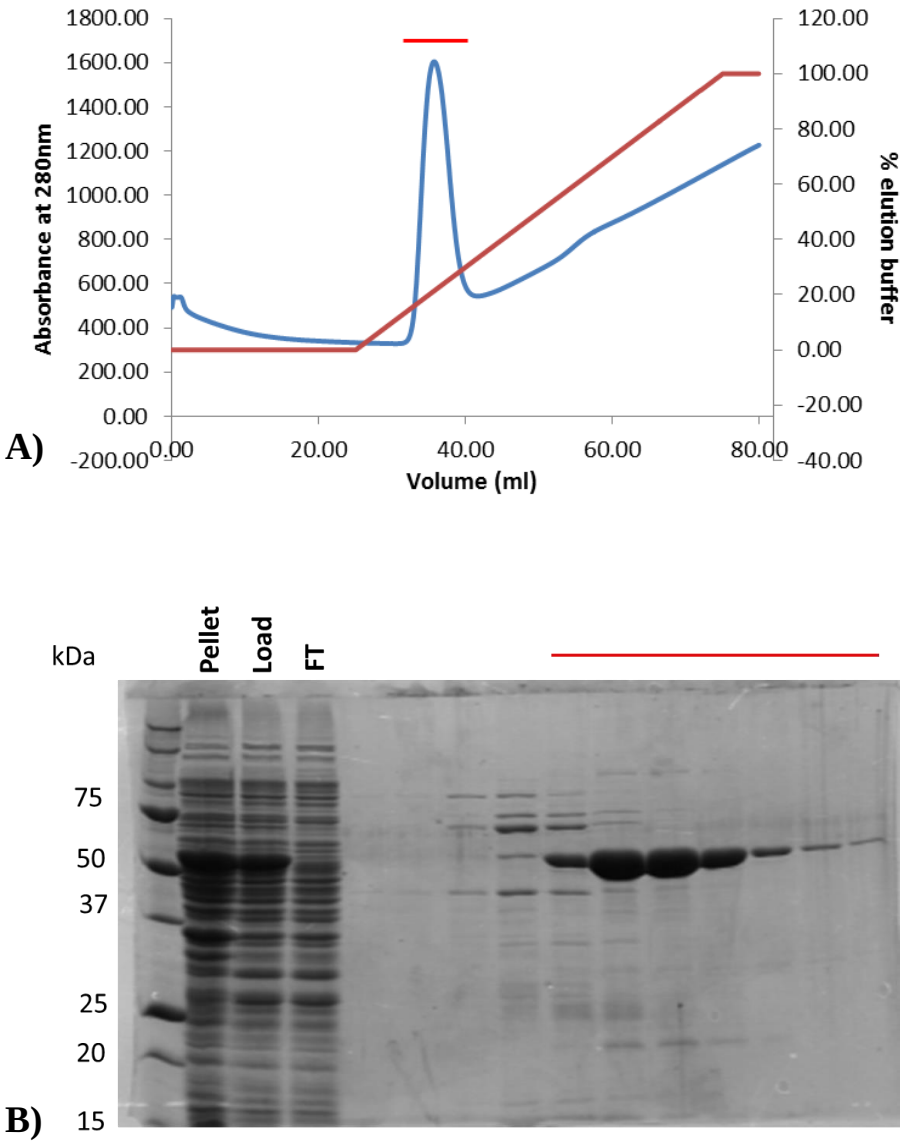


Figure 3.5: IMAC purification of a recombinant *Bacillus* MurA enzyme
A) The chromatogram shows the UV absorbance at 280nm. The peak represents the elution of His-tagged *Bacillus* MurA LCA from the IMAC column. B) SDS-PAGE gel demonstrates the size and purity of *Bacillus* MurA LCA. The SDS-PAGE gel shows the insoluble pellet, the column load, flow through (FT) and the MurA LCA fractions eluted during IMAC. MurA LCA fractions are indicated by the red bar.

All the MurA proteins were successfully over expressed in large scale cultures, mostly soluble protein with minimal amount of impurities.

Size exclusion chromatography (SEC) was used to further purify the enzymes (50 mM Tris-HCl pH 7.5 and 300 mM NaCl) and separate the protein monomers and other oligomers. The chromatographic separation for SEC column S200 10/300 (GE Healthcare) using reference samples of different molecular weights, elutes proteins with a molecular weight of 75 KDa at ~13 ml and 44 KDa at 15 ml,

which is comparable to the dimeric and monomeric forms of MurA protein respectively (Protparam predicted a MW of approximately 47.0 KDa for monomeric MurA enzymes). Fractions corresponding to each SEC peak were run on an SDS-PAGE gel to confirm the presence of MurA protein and its purity (Figure 3.6 and Figure 3.7). One example SDS-PAGE gel showing the purity of protein is given in Figure 3.7.

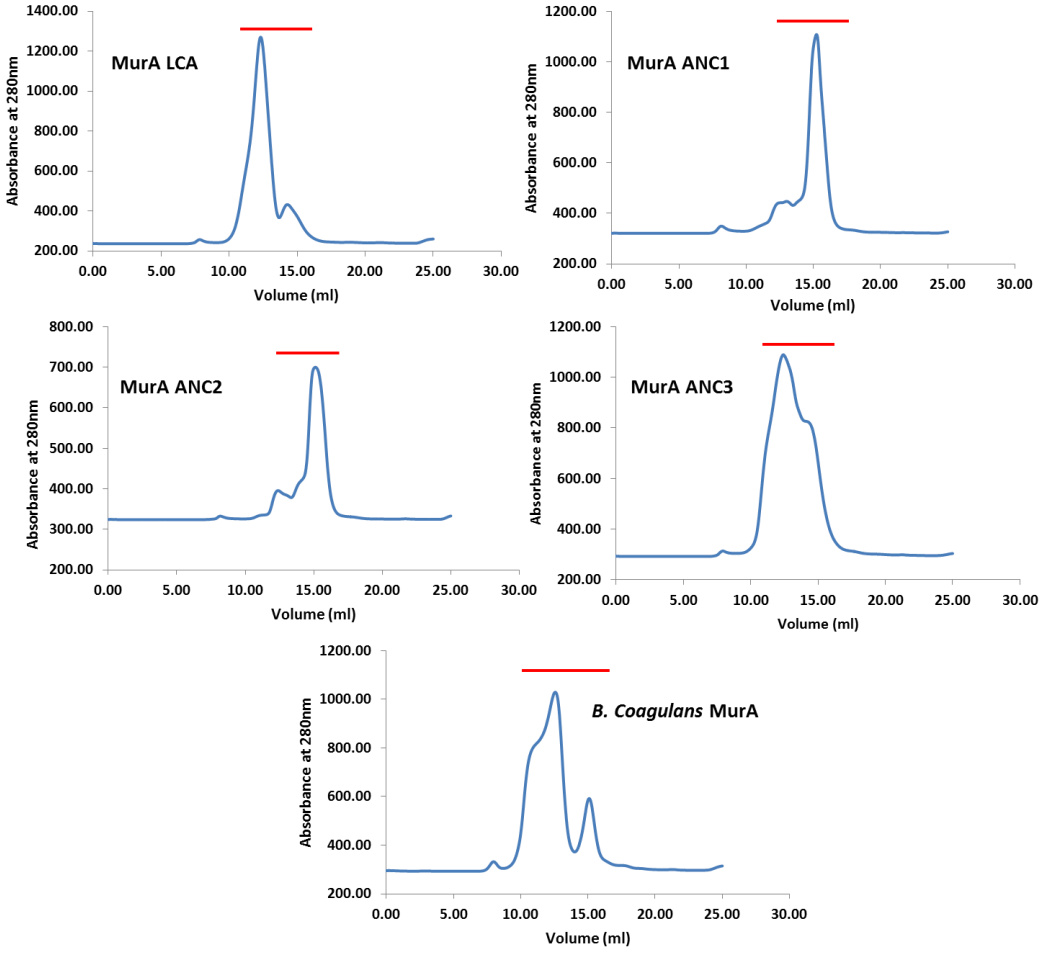


Figure 3.6: Size exclusion chromatography (SEC) purification of recombinant MurA enzymes.
The chromatogram depicts the UV absorbance at 280nm for MurA proteins and the red bar represents their elution position as well as the fractions run on SDS-PAGE gel to confirm their size and purity.

B. subtilis MurA expression was not codon optimised for expression in *E. coli*, which can help explain its low yield in comparison to the other MurA proteins. Initial purification of *B. subtilis* MurA using the generic purification buffers resulted in highly aggregated protein (Figure 3.7) Therefore, 5 % glycerol was

incorporated in all the purification buffers which drastically reduced protein aggregation (Figure 3.7).

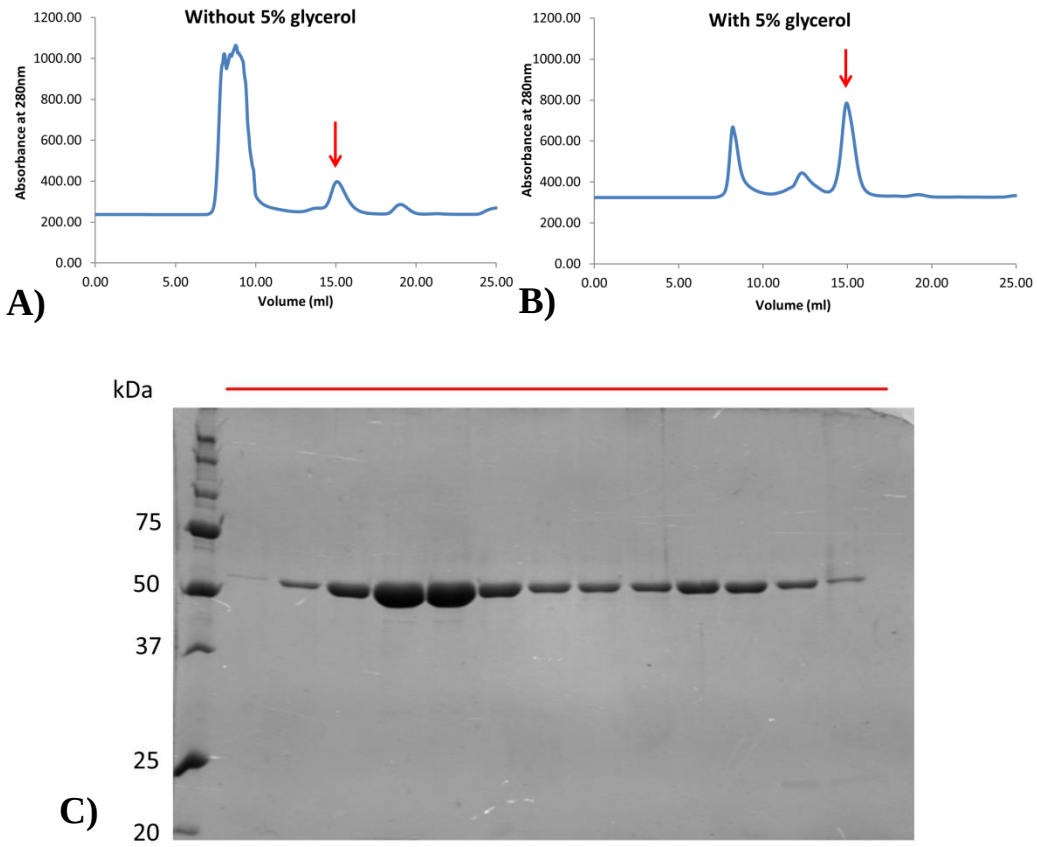


Figure 3.7: Size exclusion chromatography (SEC) purification of *Bacillus subtilis* MurA enzyme and SDS-PAGE gel
A) SEC of *Bacillus subtilis* MurA without glycerol in the purification buffers. B) SEC of *Bacillus subtilis* MurA in the presence of 5 % glycerol in purification buffers to reduce protein aggregation. The red arrow indicates the monomeric peak of *Bacillus subtilis* MurA which increases approximately 2-folds with the addition of 5 % glycerol. The red bar represents the elution position of the protein as well as the fractions run on SDS-PAGE gel to confirm their size and purity. C) The SDS-PAGE gel demonstrates the size and purity of *Bacillus* MurA LCA. The SDS-PAGE gel shows MurA LCA fractions eluted during SEC. The SDS-PAGE gel relatively corresponds to the fractions indicated by the red bar in SEC chromatograms.

The purified proteins were quantified using a Bradford assay (section 2.3.2) prior to enzyme characterisation assays.

3.2.3 Enzyme Characterisation

The evolution of organisms to adapt to new environments involves evolutionary changes to their core metabolic enzymes, in order for their necessary metabolic processes to remain functional (Hobbs et al. 2012). Therefore, such evolutionary

changes can be studied by measuring the kinetic parameters of reconstructed ancestral enzymes and thus grant insight into the evolution of enzymatic catalysis. The structural and functional aspects of enzyme activity are so complex that any error in the ancestral inference may result in either inactivity or biologically unrealistic properties of the enzyme. Therefore, enzyme activity could also be used as a proxy for accurate ancestral inference (Hobbs et al. 2012).

The enzyme reaction catalysed by MurA converts the substrates PEP and UNAG to the product enolpyruvyl UDP-*N*-acetyl glucosamine and inorganic phosphate. For the biochemical characterisation of MurA enzymes, the production of inorganic phosphate was measured using a modified malachite green assay (Baykov et al. 1988). The complex formed during the reaction between malachite green, molybdate and free phosphate, results in an absorbance band between 600 and 660nm, which can be measured using a spectrophotometer.

3.2.3.1 Activity assay

Various preliminary tests were performed to optimize the malachite green assay conditions. As the assay is highly sensitive, it was important to determine the contribution of reaction components to free phosphate in order to differentiate it from the free phosphate generated by the reaction. Therefore, the assay buffer only, assay buffer + enzyme, and substrates + buffer were tested using the malachite green assay. It was observed that the assay buffer and enzyme did not add to the free phosphate, whereas, the substrates' contribution to the free phosphate in the reaction is high. Therefore, to ensure the kinetics obtained for MurA were entirely due to the activity of MurA, a control/blank reaction was carried out for all the substrate concentrations tested which was subtracted from the MurA reaction readings to measure the free phosphate generated by MurA.

Among the SEC purifications of the six MurA enzymes, *B. coagulans* MurA and *Bacillus subtilis* MurA UV traces show two distinctive peaks corresponding to the monomeric and dimeric forms of MurA protein (Figure 3.6 and Figure 3.7) (section 3.2.2.3). Therefore, activity tests were performed to verify which peak/protein state, is active or whether there is any difference in activity between them. It was observed that, for all the ancestral enzymes, both the dimer and monomer, were equally active. Whereas, for the contemporary enzymes *B. subtilis*

MurA and *B. coagulans* MurA, the monomer peak was found to be much more active than the dimer . Therefore, for consistency, fractions from the monomeric peak were selected for all the enzymes for characterisation assays. The activity tests were generally performed with 1mM of PEP and 1mM UNAG final concentration in assay buffer 50mM Tris pH 7.5 at 37 °C. Due to the high sensitivity of malachite green assay, activity assays with various concentrations of enzymes were performed to find out an optimum enzyme concentration which would not result in precipitation.

3.2.3.2 Time course assay

A time course assay was performed as described in section 2.3.7.3 and a graph with absorbance at 620 nm versus time was plotted to determine the most appropriate duration for the assays (Figure 3.8).

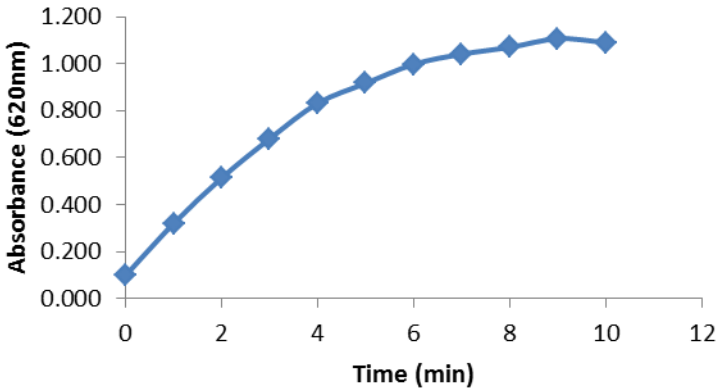


Figure 3.8: Time course assay from which the reaction time was manually determined to be ≤5 minutes.

3.2.3.3 Phosphate standard

A phosphate standard curve was generated as explained in section 2.3.7.2 using KH_2PO_4. An example for one such phosphate standard is given in Figure 3.9. The standard curve was used to calculate the amount of free phosphate produced during the MurA reactions.

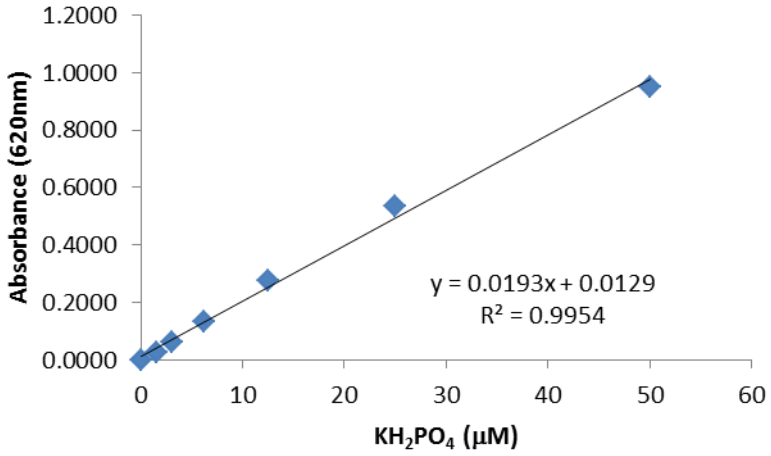


Figure 3.9: Standard curve obtained for 0 µM - 50 µM KH_2PO_4.

3.2.3.4 Michaelis-Menten kinetics determination

The malachite green assay was used to perform kinetic analysis of MurA enzymes. For PEP K_M measurements, UNAG was maintained at a constant concentration of 200 µM. Once the PEP K_M was determined, five to ten times the PEP K_M was added for measurements of the UNAG K_M. Therefore, k_{cat} values were calculated from UNAG K_M values at V_{max} (more reliable than the V_{max} from PEP K_M with UNAG being rate limiting). In cases where ten times the PEP K_M was not added during a UNAG K_M determination, the apparent V_{max} was calculated, that is, the V_{max} which would have been obtained had there been ten times the PEP K_M.

The substrate affinity represented by the *Michaelis Menten* kinetic constant K_M, typically increases with an increase in temperature (Griffith et al. 2001). Therefore the K_M for each substrate, PEP and UNAG, were initially determined at 37 °C (Table 3.2) so that the temperature optima of the enzymes (T_{opt}), and K_M of substrates at the enzymes respective T_{opt}, could be accurately determined at saturating amounts of substrate.

Table 3.2: Kinetic constants for contemporary and ancestral MurA enzymes at 37°C.

Enzyme	PEP K_M (μM)	UNAG K_M (μM)	k_{cat} (sec^{-1})	k_{cat}/K_M (UNAG) ($sec^{-1}M^{-1}$)
B. subtilis	430 ± 90	170 ± 20	5.7 ± 0.2	3.4 (±0.4)x10^4
B. coagulans	40 ± 10	130 ± 30	4.5 ± 0.3	3.0 (±1.0) x10^4
ANC1	240 ± 60	150 ± 20	3.8 ± 0.1	2.5 (±0.5) x10^4
ANC2	90 ± 10	110 ± 10	6.0 ± 0.2	5.5 (±0.9) x10^4
ANC3	50 ± 10	80 ± 20	4.7 ± 0.3	6.0 (±2.0) x10^4
LCA	120 ± 10	150 ± 20	1.58 ± 0.06	1.1 (±0.2)x10^4

The kinetic values were obtained from three technical replicates.
The errors represent the standard error of the mean.

It was observed that the PEP K_M of *B. subtilis* MurA measured at 37 °C is >6-fold higher than previously reported PEP K_M of 66 μM from *Bacillus subtilis* MurA measured at 25 °C (Biery 2007). The UNAG K_M from the study performed by Biery et al (2007) is comparable to the above measured UNAG K_M. By comparison, data reported for *Staphylococcus aureus* MurA (no incubation temperature specified) only shows a 4-5 fold difference in k_{cat} (1.13 sec^{-1}) (Blake et al. 2009). Differences may be expected between different assays as seen with the catalytic efficiency, k_{cat}/K_M (UNAG), measured for *S. aureus* MurA using a MurA-MurB coupled assay (12 $sec^{-1}M^{-1}$) (Biery 2007) differs by 750 fold from the k_{cat}/K_M (UNAG) of *S. aureus* MurA obtained using a PNG assay (9000 $sec^{-1}M^{-1}$) (Blake et al. 2009).

The optimum temperature of enzyme activity (T_{opt}) was measured as described in section 2.3.7.5 and the kinetic parameters PEP K_M, UNAG K_M, k_{cat} and k_{cat}/K_M were determined at each enzyme's respective T_{opt} (Table 3.3; Figure 3.10 and Figure 3.11). The T_{opt} data is described in section 3.2.5.

Table 3.3: Kinetic constants for contemporary and ancestral MurA enzymes at their respective optimal temperatures (Topt).

Enzyme	PEP K_M (μM)	UNAG K_M (μM)	k_{cat} (sec^{-1})	k_{cat}/K_M (UNAG) ($sec^{-1}M^{-1}$)	T_{opt} (°C)
B. subtilis	390 ± 70	490 ± 60	6.0 ± 0.3	1.2 (±0.2) $x10^4$	42
B. coagulans	320 ± 60	190 ± 40	32 ± 2	17.0 (±4.0) $x10^4$	59
ANC1	150 ± 20	100 ± 40	6.2 ± 0.7	6.0 (±3.0) $x10^4$	62
ANC2	100 ± 20	200 ± 30	72 ± 3	36.0 (±7.0) $x10^4$	62
ANC3	400 ± 100	240 ± 30	58 ± 2	24.0 (±4.0) $x10^4$	62
LCA	220 ± 20	140 ± 20	7.1 ± 0.3	5.1 (±0.7) $x10^4$	58

The kinetic and thermoactivity values were obtained from three technical replicates.
The errors represent the standard error of the mean.

The k_{cat} values obtained for *B. subtilis* MurA at 37 °C (Table 3.2) and at its T_{opt} temperature of 42 °C (Table 3.3) were found to be very similar as the temperature difference is minimal. *B. coagulans* MurA was found to have >2-fold higher affinity for UNAG than *B. subtilis* MurA. *B. coagulans* MurA is also 5-fold faster than *B. subtilis* MurA with more than 12-fold its catalytic efficiency. It was also observed that the measured k_{cat} for *B. subtilis* MurA (6.7 sec^{-1}) was found to be comparable with previously reported k_{cat} for *E. coli* MurA (8.9 sec^{-1}) at 37 °C (Dai et al. 2002).

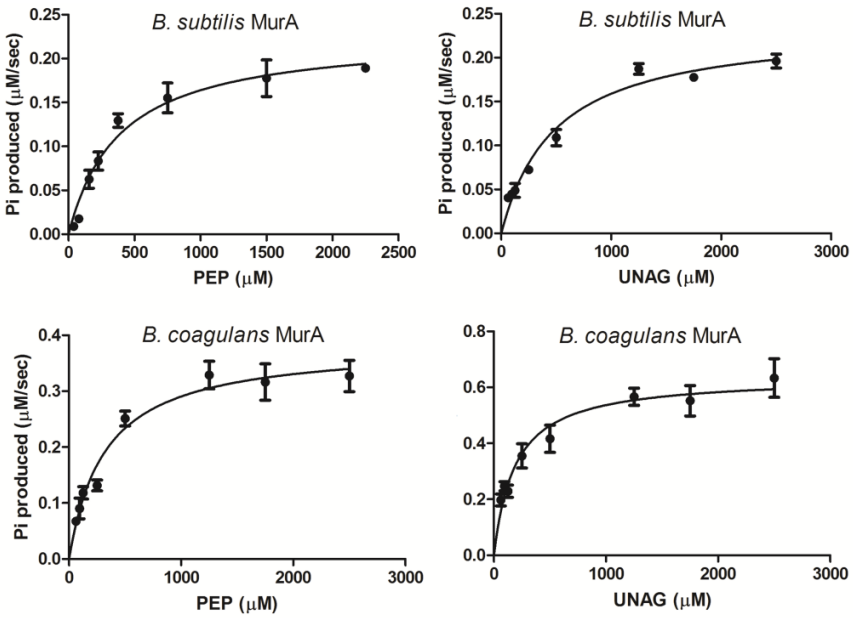


Figure 3.10: Michaelis-Menten plots for contemporary MurA enzymes at their respective T_{opt} values
Results are the mean of three technical replicates, the error bars representing the standard error of the mean (SEM).

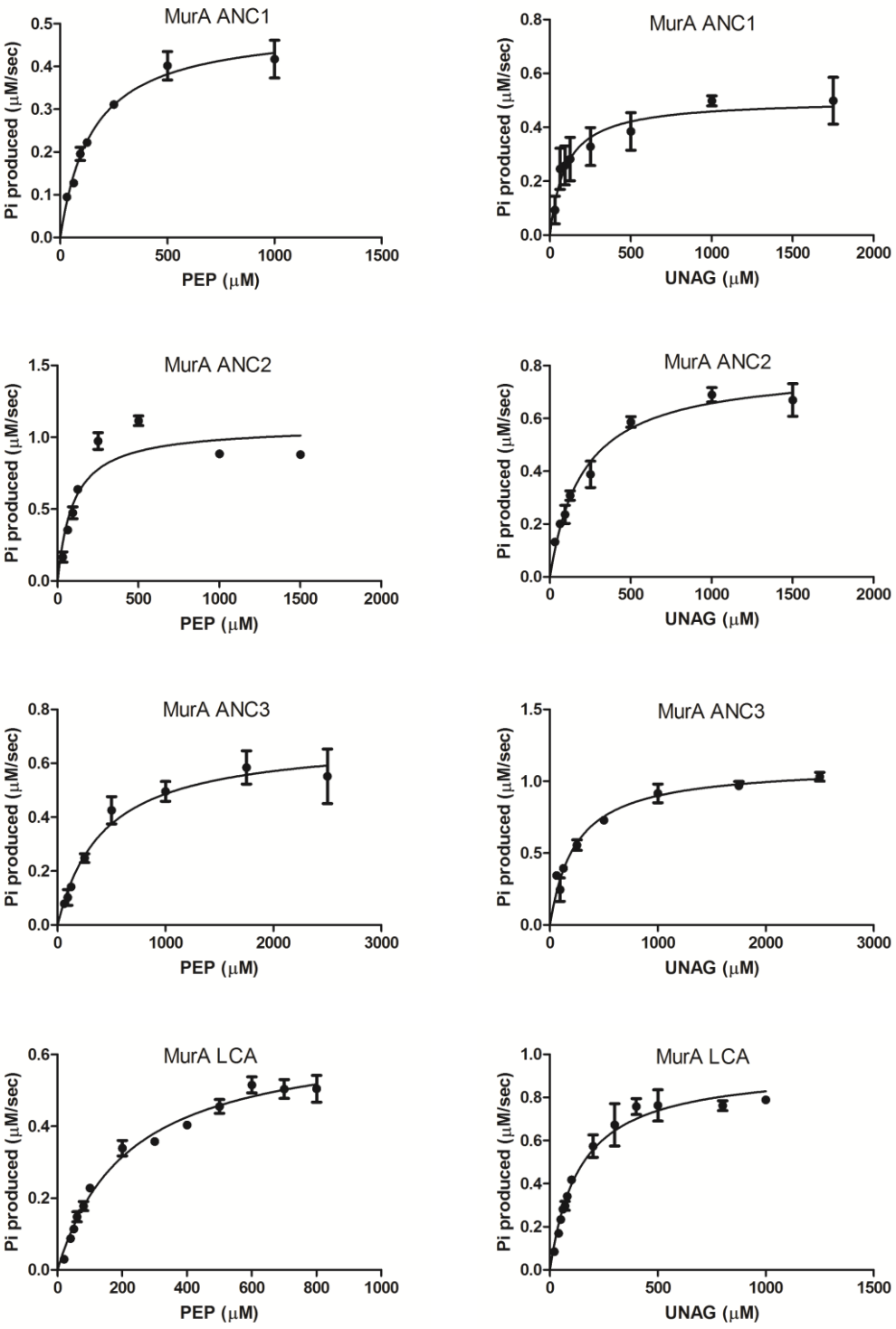


Figure 3.11: Michaelis-Menten plots of ancestral MurA enzymes at their respective T_{opt} values
Results are the mean of three technical replicates, the error bars representing the SEM.

All four ancestral MurA enzymes were catalytically active and their kinetic parameters were observed to be proportionate with those of the contemporary enzymes (Table 3.3). It should be noted that in comparison to the other ancestral

and contemporary enzymes, MurA ANC2 displays the highest *k*cat and catalytic efficiency, with more than twice the catalytic efficiency of contemporary *B. coagulans*. It was also observed that *Bacillus* MurA LCA showed very similar kinetic values to that of MurA ANC1.

Another characteristic property of MurA is its susceptibility to the antibiotic fosfomycin. Therefore, the inhibitory properties of selected MurA enzymes were measured using the malachite green assay.

3.2.4 IC_{50} determination

MurA, being an important antibiotic target, has been focussed on for the development of new inhibitors and for identification of evolutionary aspects leading to inhibitor resistance, such as Cys-Asp single mutation (Kim et al. 1996a; De Smet et al. 1999) leading to fosfomycin resistance. Therefore, studying the inhibitory properties of ancestral MurA in comparison to contemporary MurA enzymes may also provide insight into the evolution of MurA.

3.2.4.1 Inhibition by Fosfomysin

Fosfomycin is the only clinically available antibiotic which inhibits MurA (Chang et al. 2014). However, many microorganisms have evolved to be resistant to fosfomycin. Therefore, finding the difference in IC50 for fosfomycin of *Bacillus* MurA LCA and contemporary *B. subtilis* MurA would allow us to determine how much of the evolution of MurA has been affected by its inhibitory properties. The IC50 for fosfomycin measured as described in section 2.3.7.6 are given in Table 3.4 and the corresponding dose-response curves are given in Figure 3.12. It was observed that IC50 of fosfomycin for *B. subtilis* MurA is slightly lower than that of *Bacillus* MurA LCA. However, the values are still in a comparable range.

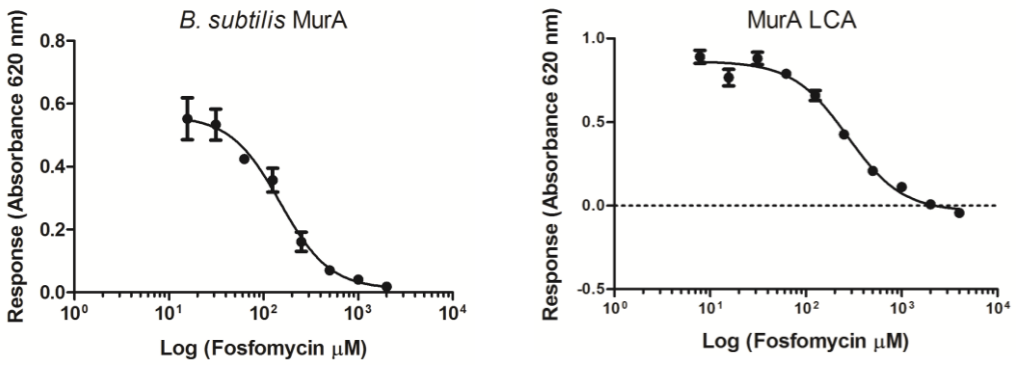


Figure 3.12: The IC50 dose-response curves obtained for different concentration of fosfomycin
Results are the mean of three technical replicates, the error bars representing the SEM

The IC50 for fosfomycin measured for *B. subtilis* MurA (Table 3.4) was observed to be >30-fold higher than that previously observed for *B. subtilis* MurA (3.8±0.9 μM) (Biery 2007). However, the values better compared with fosfomycin IC50 measured for *Haemophilus influenza* MurA (40 μM) (Jin et al. 2009). Both studies by Biery et al. (2007) and Jin et al (2009) were performed at 25 °C, whereas, the IC50 values here were measured at 37 °C. The increase in temperature can drastically increase IC50 values (Beavis & Powers 2004). This may explain the difference between the measured and previously reported IC50 values.

3.2.4.2 Inhibition by Glyphosate

As outlined in chapter 1 (section 1.5), MurA and AroA are known to be exclusively inhibited by fosfomysin and glyphosate respectively. However, the structural and functional similarities of MurA and AroA lead us to hypothesise that MurA and AroA might have originated from a common ancestor. Glyphosate specifically interacts with active site residues of AroA to cause inhibition (Eschenburg et al. 2003). The difference in amino acid composition and spatial arrangement between AroA and MurA makes it impossible for glyphosate to specifically inhibit MurA. Glyphosate could inhibit MurA only if appropriate conformational changes occurred in order to avoid any distance clashes, steric interference and to allow proper positioning of glyphosate molecule (Eschenburg et al. 2003). Therefore, measuring the IC50 of glyphosate for *Bacillus* MurA LCA

and *Bacillus subtilis* MurA would provide us with information on whether any such changes have occurred to the ancestral enzyme.

The glyphosate IC50 graphs show that inhibition of MurA occurs with very high concentrations of glyphosate (Figure 3.13). However, such high concentrations of glyphosate suggest that it does not specifically inhibit *Bacillus* MurA LCA and *B. subtilis* MurA (Table 3.4). The IC50 of glyphosate for *Bacillus* MurA LCA is almost 2-fold higher than that of *B. subtilis* MurA. However, considering nonspecific inhibition, the difference between the IC50 values is still comparable.

Table 3.4: Fosfomycin and glyphosate IC50 values for *Bacillus subtilis* MurA and *Bacillus* MurA LCA

	***B. subtilis* MurA**	***Bacilus* MurA LCA**
Fosfomycin	150 μM	270 μM
Glyphosate	2800 μM	4800 μM

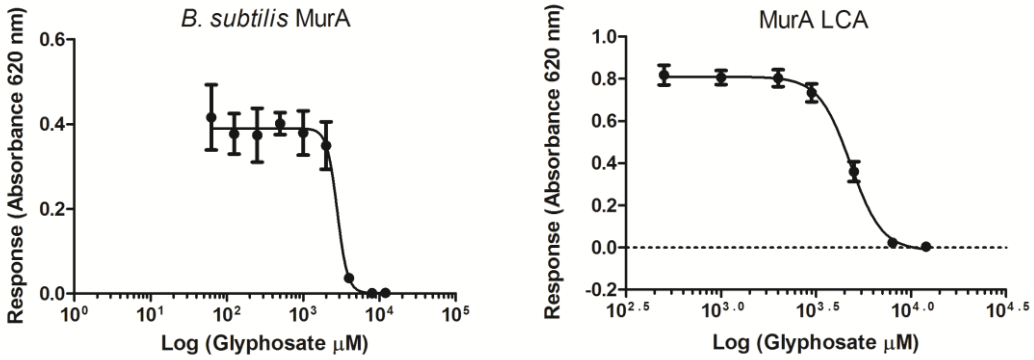


Figure 3.13: The IC50 dose-response curves obtained for different concentration of glyphosate
Results are the mean of three technical replicates, the error bars representing the SEM

The IC50 measurements for fosfomycin indicate that both LCA of *Bacillus* genus and contemporary *Bacillus* MurA enzymes are sensitive to fosfomycin to relatively the same extent. Fosfomycin, mimics the MurA substrate PEP and competes with it to form an irreversible thioether bond to Cys115 (*E.coli* numbering) (Skarzynski et al. 1996; Fadda et al. 2006). Upon binding of fosfomycin, several hydrogen bonds are made between fosfomycin and UNAG as

well as between fosfomycin and the polypeptide chain, among which, Lys22, Arg120 and Arg397 provide strong electrostatic interactions (in *E. coli* MurA) (Skarzynski et al. 1996). All the above amino acids which are identified in binding of fosfomycin are conserved within the MurA LCA. Such a conservation of residues, along with comparable IC50 values obtained from MurA LCA and *B. subtilis* MurA, may suggest that the amino acids and respective conformational changes involved in the fosfomycin inhibition of *B. subtilis* MurA, remains the same within MurA LCA.

The glyphosate IC50 for *B. subtilis* MurA suggests that the enzyme cannot be specifically inhibited by glyphosate as might be expected for ancestral enzymes. Glyphosate targets the PEP binding site in AroA enzymes (Eschenburg et al. 2003). The PEP binding site involves the residues Lys22, Arg124, Asp313, Arg344 and Arg386 in AroA (*E. coli* numbering); and Lys22, Arg120, Asp305, Arg331, and Arg371 in MurA (*Enterobacter cloacae* numbering), which are also conserved within MurA LCA. The interaction of these residues observed with the PEP moiety of the tetrahedral intermediate, are identical within AroA and MurA enzymes (Eschenburg et al. 2003). However, the residues Lys22, Arg124/Arg120 and Arg344/Arg331 exhibit geometrical deviations, causing distortions within the PEP binding site. Therefore, these distortions in the PEP binding site between MurA and AroA, as well as the substantial differences in MurA and AroA binding sites due to the variation in size, shape, polarity and charge of UNAG and S3P, may explain why glyphosate cannot inhibit MurA, in spite of glyphosate targeting the PEP binding site in a similar way as fosfomycin (Eschenburg et al. 2003). Therefore, the lack of specific inhibition of MurA LCA by glyphosate suggests that the ancestral enzyme does not possess any structural changes that could allow the correct positioning of glyphosate to allow inhibition.

Another characteristic feature involved in the study of evolution of MurA enzymes involves determining their thermal properties such as thermoactivity and thermostability. The next section describes the evolutionary significance of MurA enzymes' thermal properties as well as the significance of these thermal characteristics in the evolution of thermophily.

3.2.5 Evolution of Thermophily

Thermophily is believed to be a primitive trait. The study on elongation factors (EF-Tu) by Gaucher et al. (2008) suggested that evolution shows a gradual decrease in thermophily from that of the last common ancestor. On the other hand, a temporal fluctuation in respect to thermophily was observed with *Bacillus* genus from the ancestral LeuB enzymes reconstructed using ASR (Hobbs et al. 2012). However, the sampling involved in both studies differ and the fluctuations in the temporal trend in thermophily observed by Hobbs et al. (2012) may have been obtained from the period that was not sampled in the study on EF-Tu (Gaucher et al. 2008). These findings are based on the data obtained from a single enzyme. It is not known whether the reconstructions of a different protein would result in the same temporal trend in thermophily. Therefore, this could be addressed by repeating the study on the evolution of thermophily within the *Bacillus* genus using a different enzyme, such as MurA.

As organisms adapt to new environments *via* evolutionary selection pressures such as temperature changes, core metabolic enzymes, such as MurA, adapt accordingly to maintain their function (Hobbs et al. 2012). The MurA phylogenetic tree was built using sequences from *Bacillus* species inhabiting different thermal environments and hence may prove to be ideal to investigate the evolution of thermophily within *Bacillus* genus. The study of the evolution of thermophily using ASR exploits the correlation between the organism growth temperature and the enzyme thermoactivity (T_{opt}) and thermostability (T_m) (Gromiha et al. 1999; Akanuma et al. 2013). The T_{opt} and T_m of reconstructed enzymes were used to represent their thermal adaptation (Hobbs et al. 2012). Using MurA to study evolution of thermophily with the *Bacillus* genus would not only provide information on thermal properties of ancestral MurA, but also help us to test the robustness of using ASR as a tool to study the evolution of thermophily.

Thermoactivity of MurA enzymes (T_{opt}):

The thermoactivity measured for mesophilic *B. subtilis* MurA and moderately thermophilic *B. coagulans* MurA is consistent with their growth temperature and hence confirms the correlation (Figure 3.14; Table 3.5).

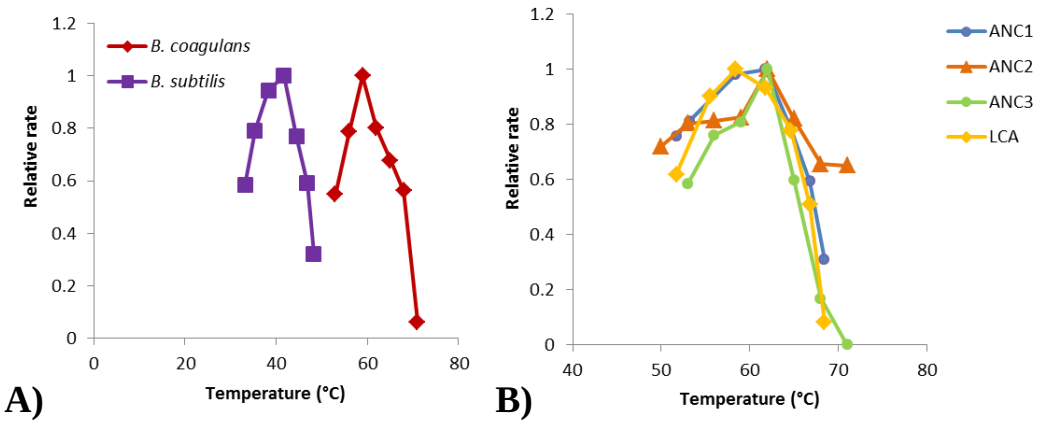


Figure 3.14: Thermoactivity profiles for contemporary and ancestral MurA enzymes
Thermoactivity data for A) contemporary and B) ancestral MurA enzymes shown are the mean of at least three replicates and are expressed as a proportion of the highest activity (at T_{opt}).

It was observed that the T_{opt} obtained for *B. subtilis* MurA was slightly higher than its growth temperature, however, such a trend is common in mesophilic organisms (Georlette et al. 2004; Hobbs et al. 2012).

Thermostability of MurA enzymes (T_m):

It was observed that the measured thermostability is consistent with thermoactivity of MurA enzymes (Figure 3.15; Table 3.5). The thermoactivity and thermostability of ancestral MurA enzymes shows that the temporal trend is consistently moderately thermophilic. However, this trend differs from the fluctuating trend in thermal evolution observed for LeuB (Hobbs et al. 2012). The finer resolution thermostability data provided by DSC for ancestral MurA enzymes also indicate a slight decrease in thermophily from ANC1 through to LCA (Figure 3.15; Table 3.5). This contradicts the trend observed by Gaucher et al. (2008) and Perez-Jimenez et al. (2011) which show a general increase in thermophily with increasing evolutionary time.

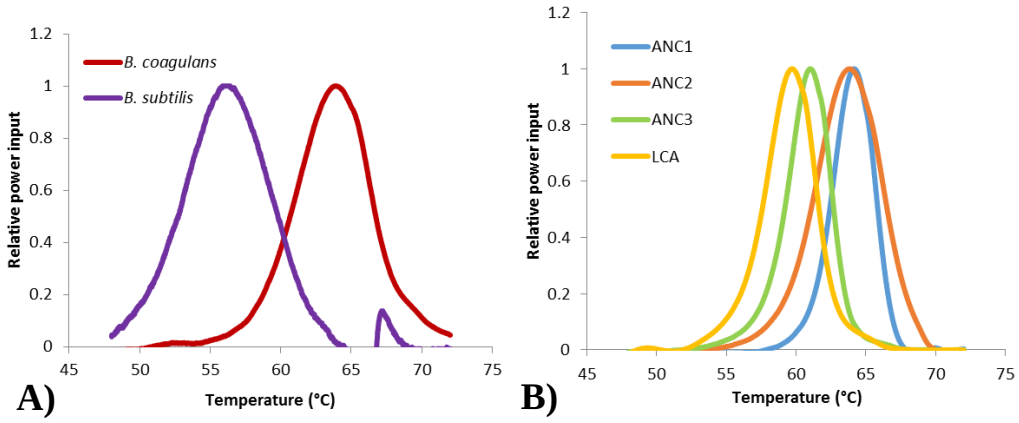


Figure 3.15: Thermostability profiles for contemporary and ancestral MurA enzymes
Thermostability data for A) contemporary and B) ancestral MurA enzymes shown are processed DSC power inputs expressed as a proportion of the peak value which equates to the T_m.

Table 3.5: Thermoactivities and/or thermostabilities of contemporary and ancestral MurA and LeuB enzymes along with their respective ages

Enzyme		Age (Myr)	T_{opt} (°C)	T_m (°C)
MurA	*B.subtilis*	-	42	55.4
	B.coagulans	-	59	63.9
	ANC1	550	62	64.2
	ANC2	600	62	63.8
	ANC3	910	62	61.1
	LCA	950	58	59.7
LeuB	*B. psychrosaccharolyticus*	-	47	-
	B. subtilis	-	53	-
	B. caldovelox	-	69	61.0
	LeuB1	670	73	64.7
	LeuB2	820	49	47.6
	LeuB3	850	60	55.5
	LeuB4	950	70	65.3

It must be noted that there is a difference of 11 °C between the T_{opt} of MurA and LeuB from *B. subtilis* (Table 3.5). This suggests that even though the correlation

between T_{opt} and T_m with organism growth temperature exists, the T_{opt} and T_m properties are protein dependent. Therefore, for accurate comparison of thermal trends between proteins, the temperatures need standardization using the enzyme temperatures from the common contemporary species *B. subtilis* (Figure 3.16). Standardization was achieved by subtracting the temperature of *B. subtilis* from all other temperatures. All MurA ancestors are moderately thermophilic with a <5 °C difference between ancestors in T_{opt} and T_m with a slight temporal decrease, while LeuB ancestors show a fluctuating trend in thermophily as it goes back in time (Figure 3.16 and Table 3.5). MurA ancestors are more clustered around time periods 550-600myr (ANC1 and ANC2) and 910-950myr (ANC3 and LCA). The LeuB ancestor responsible for the prime fluctuation in LeuB thermal trend is LeuB2 (820myr), the temperature of which cannot be confirmed as there is no MurA ancestor equivalent in age. In terms of age, LeuB3 (850myr) could be considered to be approximately the same age as MurA ANC3 (910myr) with similar thermal properties. LeuB1 (670myr) and MurA ANC2 (600 myr) differ in their thermoactivity by 11 °C but is consistent with its thermostability. In terms of phylogenetic equivalence, LeuB3 could be considered equivalent to MurA ANC2, both consisting of the pathogenic and soil clades, showing similar moderate thermophily. However, it was observed that MurA LCA and LeuB4, both aged 950myr, do not agree on their thermal adaptation. Even though both MurA LCA and LeuB4 agree that the *Bacillus* LCA is thermophilic, there is a considerable difference in the range of thermophily displayed by the T_{opt} of both enzymes.

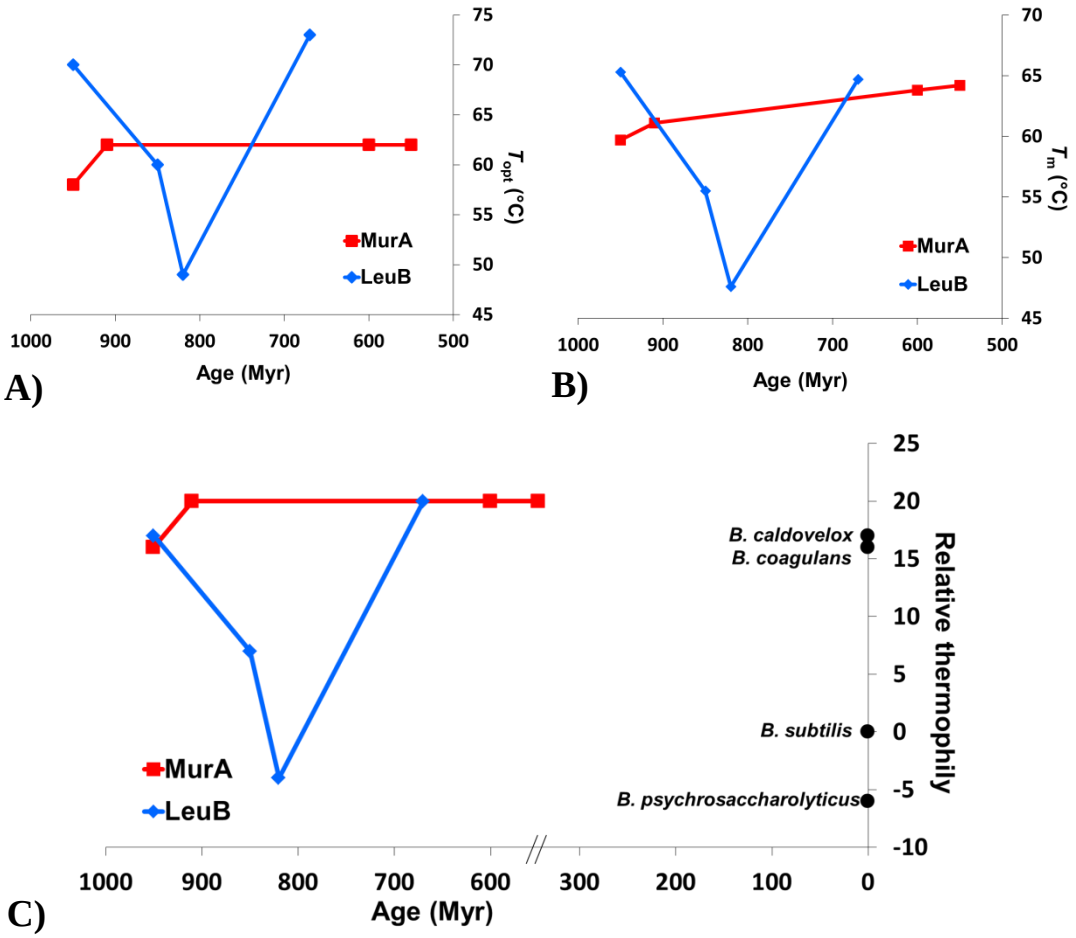


Figure 3.16: Comparison of age versus thermal adaptation data for LeuB and MurA
Data shown are age estimates for ancestral nodes against either T_m or T_{opt} values for each ancestral enzyme. A) T_{opt} data for MurA and LeuB B) T_m data for LeuB and MurA and C) standardised data obtained after subtracting value for *B. subtilis* enzyme from LeuB T_{opt} and MurA T_{opt}. Thermal data for contemporary enzymes are shown for comparison at age 0 Myr.

As previously mentioned, thermoactivity and thermostability is protein dependent and does not solely depend on the organism growth temperature. This may be one explanation for the temperature differences between MurA and LeuB ancestors of a similar age. Another explanation for such differences in thermal trends is the hypothesis that each reconstructed ancestral enzyme may only represent a snapshot of the evolving host organism at a particular period of time. Our present day *Bacillus* genus includes a wide variety of species adapted to different environment conditions. Similarly, it is logical to assume that a reconstructed ancestral enzyme from *Bacillus* species may give information pertaining to a single member of the species consortium present at that particular point of time. This explanation is in agreement with the evolutionary model proposed by Woese,

in which the universal ancestor is not a discrete entity but a community (Woese 1998). The observations made in this study also support the hypothesis made by Hobbs et al. (2012) that the thermal profiles obtained for the ancestral enzymes are most likely to reflect changes of the host microenvironments rather than representing a global temperature.

The main differences between MurA and LeuB lie in the set of contemporary species used in the study as well as the phylogenetic tree topology. However, both MurA and LeuB ancestral inference was achieved using the same method based on ML criteria. The reconstructed MurA enzymes were all active and showed comparable catalytic constants as the contemporary MurA enzymes, therefore their thermal properties are also to be trusted and hence the thermal trends they imply. One source of error associated with ML inference is the overestimation of thermophily (Williams et al. 2006) of reconstructed ancestral proteins, which is not supported by MurA as the data does not indicate any considerable increase in thermophily nor does it show a trend of increasing thermophily with evolutionary time. The thermal properties of the MurA ancestors show a consistent moderate thermophily, with a slight decrease in thermal properties with increasing evolutionary time, considering the absolute values measured. It was also noted that *B. selenitireducens* branch (OGT of 26 °C) in the phylogenetic tree did not influence the ANC3 temperature as ANC3 remains thermophilic. There is also confidence in the thermal properties of the ancestral LeuB enzymes reconstructed by Hobbs et al. (2012) as their thermal properties do not simply represent averages of their extant descendants. Therefore, it is clear that the two core metabolic enzymes, MurA and LeuB, from the same genus show different thermal profiles. This suggests that the choice of protein candidate may influence the ancestral inference.

Overall, all the ancestors of *Bacillus* inferred using MurA sequences are consistently moderately thermophilic. Thermostability and thermoactivity of an enzyme is correlated to the host organism growth temperature, however, the absolute T_{opt} and T_m are protein dependant. The temporal differences in thermophily observed by different enzymes may either be explained by the hypothesis that the reconstructed ancestors only represent a single snapshot of the evolving host at a given point of time, or the choice of protein candidate may

influence biological conclusions. Further, as the reconstructed MurA ancestors demonstrate rather similar properties as that of the contemporary MurA enzymes, it may be necessary to go back further in time in the reconstruction of ancestors to be able to see some differences which in turn would facilitate the study of MurA evolution.

The functional characterisation of ancestral and contemporary AroA enzymes was studied in a similar way as MurA and is described in chapter 4.

Chapter 4

AroA

4.1 Introduction

The enzyme AroA, otherwise also known as 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), is an attractive antibiotic target, and is the target of the herbicide glyphosate. Although most studies on AroA focus on genetic modifications to synthesise glyphosate tolerant or sensitive strains, there is an increased interest to design novel inhibitors for their use as broad-spectrum antibiotics as the pathogenicity of many microorganisms depend on a functional AroA enzyme and other shikimate pathway enzymes (Funke et al. 2007). Similar to that of MurA, previous studies focused on the function and structure of AroA, yet there is little information available on the evolution of AroA. Studying the evolution of AroA could contribute to better inhibitor design. Therefore, the evolution of AroA could be explored in a similar way as MurA, by using ASR to reconstruct the LCA of the *Bacillus* genus using AroA and compare its functional and structural properties to contemporary AroA.

AroA catalyses a similar reaction as MurA, involving the transfer of the enolpyruvyl moiety to its substrate accompanied by the release of inorganic phosphate. Therefore, the kinetic, inhibitory and thermal properties of AroA could also be determined using the malachite green assay. The AroA substrate, shikimate-3-phosphate (S3P), although commercially available, is high-priced which limits the efficiency with which the experiments could be carried out. Therefore, it was necessary to synthesize S3P in the laboratory for its use in AroA characterisation. In addition to the *in vitro* properties, the behavior of ancestral AroA *in vivo* in contrast to contemporary AroA was also examined using knockout strains.

This chapter describes the reconstruction of *Bacillus* AroA LCA and compares the *in vitro* and *in vivo* characteristics of AroA LCA against the *in vitro* and *in vivo* characteristics of contemporary *Bacillus subtilis* AroA enzyme. The process involved in the synthesis of the AroA substrate S3P, necessary for AroA *in vitro* characterisation, is also described in this chapter.

4.2 Results and Discussion

4.2.1 Phylogenetic Analysis and ASR

4.2.1.1 Sequence Alignment

The AroA amino acid and nucleotide sequences were collected from the same 20 *Bacillus* species as for MurA, along with AroA from the same two *Clostridium* species as outgroup sequences (Accession numbers, species and strain details are given in Appendix A). The amino acid alignment of *Bacillus* AroA sequences along with the two outgroup sequences has an overall 14.3% of identical sites. Overall, the protein alignment without the *Clostridium* sequences has 35.0% identical sites. The low percentage of identical sites may suggest a high rate of evolutionary change from the LCA of *Bacillus* AroA to contemporary *Bacillus* species, which may be reflected in the *in vitro* and *in vivo* properties of the enzyme measured in this chapter.

Various catalytically important residues and conserved residues have been previously identified for AroA. The active site residues reported from *E. coli* include Lys22, Arg124, Asp313, Glu341, Arg344, Arg386 and Lys411 (*E. coli* numbering) (Eschenburg et al. 2003). His385 from *E. coli* has been suggested to provide a proton to Glu341, which in turn acts as proton donor for PEP (Schönbrunn et al. 2001). These catalytically important and conserved residues were also identified in the *Bacillus* AroA alignment, except for Lys411 which is replaced by Valine in all the *Bacillus* AroA sequences at that position. The numbering of the above residues from *E. coli* corresponds to the numbering in the AroA alignment as follows: Lys22 → Lys25, Arg124 → Arg131, Asp313 → Asp334, Glu341 → Glu362, Arg344 → Arg365, His385 → His407, Arg386 → Arg408 and Lys411 → Val434 (shown in AroA alignment in Appendix B). The conservation of these residues observed in the alignment may suggest that it would be very likely that these residues will also be conserved in the ancestral AroA sequence that is inferred from these *Bacillus* sequences.

4.2.1.2 Phylogenetic Analysis

The model of evolution was identified using ProtTest version 2.4 (Abascal et al. 2005) as performed for MurA. Similar to MurA, the LG model was chosen to be the most suitable model of amino acid evolution. Five phylogenetic trees were then generated using GARLI 1.0 (Zwickl 2006) based on ML criterion, each tree

rooted with the outgroup *Clostridium* AroA sequences. All five phylogenetic trees had equal log likelihood scores of -9718 with identical clade groupings as well as branch positions. Therefore, this generated phylogram may define the evolutionary relationship between the *Bacillus* species in the best possible way.

The *Bacillus* AroA phylogenetic tree was bootstrapped to provide confidence for the clades in the phylogenetic tree. The bootstrap values ≥50% are displayed in Figure 4.1. The phylogeny for *Bacillus* AroA also shows the presence of distinct phylogenetically conserved clades: the soil clade, the pathogenic clade and the halophilic clade (Alcaraz et al. 2010) (Figure 4.1). The topology of these conserved clades in *Bacillus* AroA phylogram is also consistent with that of the *Bacillus* LeuB phylogram (Hobbs et al. 2012) (Figure 3.2). Yet, it was observed that in the AroA phylogram, *Bacillus megaterium*, a non-pathogenic *Bacillus* species, was categorised within the pathogenic clade. However, *Bacillus megaterium* shares a phylogenetic relationship to *Bacillus anthracis* and *Bacillus cereus* (Dib et al. 2003; Ramasamy et al. 2013). The halophile clade described by Alcaraz et al (2010) is comprised of *Bacillus halodurans* and *Bacillus clausii*. In addition to these species, the halophile clade in the *Bacillus* AroA phylogenetic tree includes *Bacillus selenitireducens*, *Bacillus cellulosilyticus* and *Bacillus pseudofirmus* which are also halophilic species (Nielsen et al. 1995; Blum et al. 1998; Nogi et al. 2005).

AroA from the last common ancestor (LCA) of the *Bacillus* genus was chosen for sequence reconstruction. The AroA phylogram was converted to a chronogram in the same manner as was done for MurA, by fixing time nodes at 2650 myr and 950 myr corresponding to the divergence point of *Clostridium* and *Bacillus* and *Bacillus halodurans* and *Bacillus subtilis* respectively using R8s version 1.71 (Sanderson 2003) (Figure 4.1). The optimal growth temperatures of each *Bacillus* species indicated in Figure 4.1 remains the same as shown in MurA chronogram (Figure 3.1).

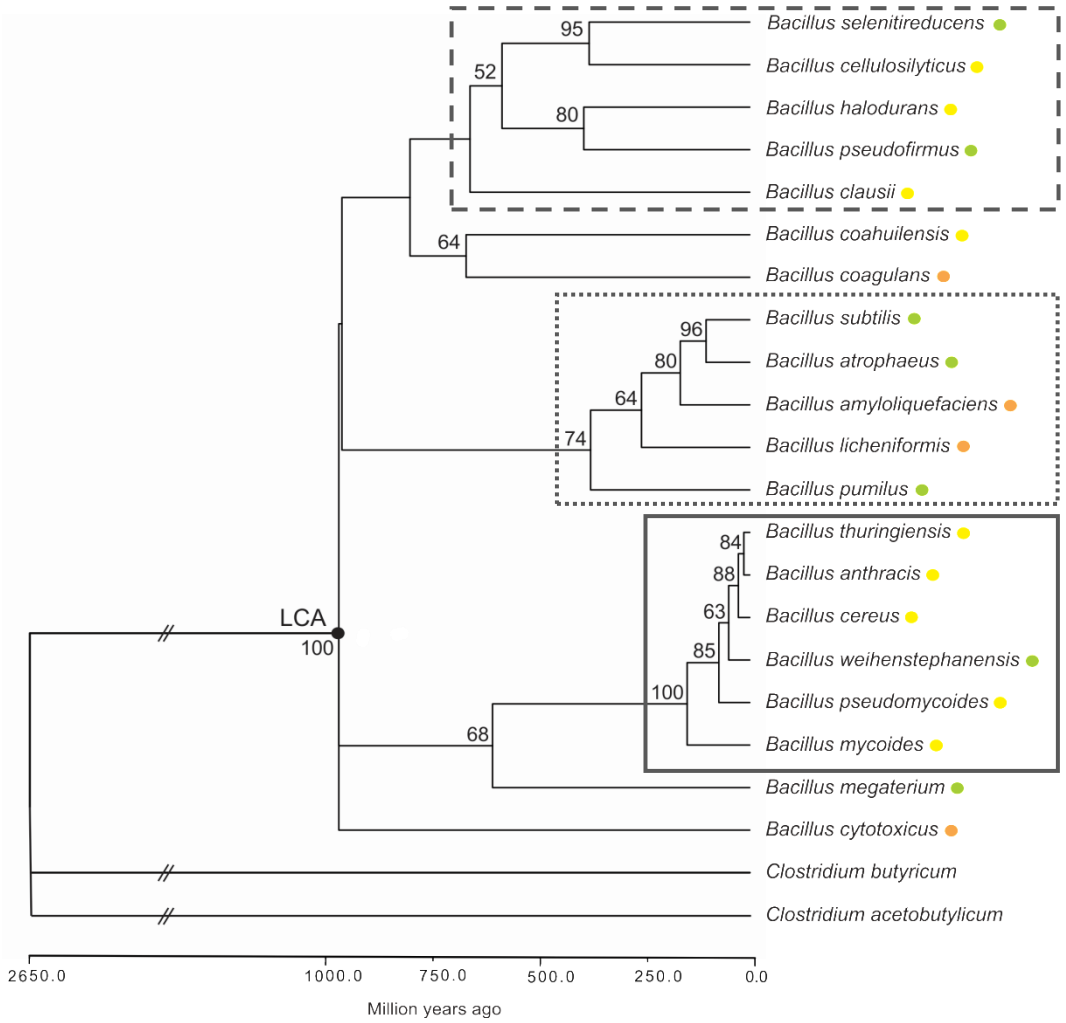


Figure 4.1: ML chronogram of *Bacillus* species based on AroA amino acid sequences. The phylogenetic tree was rooted with *Clostridium* sequences as outgroup. The reconstructed ancestral enzyme is indicated by a black circle. The coloured circles next to each *Bacillus* species relates to their optimal growth temperatures: 45-50 °C (orange), 37 °C (yellow) and 25-30 °C (green). The phylogenetic conserved clades are indicated in boxes: pathogenic clade (solid line), soil clade (dotted line) and halophile clade (dashed line) (Alcaraz et al. 2010). The bootstrap values ≥ 50%, assessed from 1,024 bootstrap replicates are also shown.

4.2.1.3 Ancestral Inference

The inference of the ancestral sequence for *Bacillus* AroA LCA was achieved using amino acid, nucleotide and codon inference methods by ML criterion. The nucleotide model of evolution for the nucleotide inference was determined by jModelTest version 0.1.1 (Posada 2008) to be the general time reversible model (GTR model) of nucleotide substitution (Tavaré 1986) and the Jones amino acid rate file was employed for amino acid and codon inference.

PAML version 4.3 (Yang 2007) inferred the AroA LCA sequence for the three methods and ClustalW (Larkin et al. 2007) was used to align the three translated

sequences. The three inference methods were in agreement for 90% of the sites and two out of three methods were in agreement for a further 9% of sites. The remaining 8 ambiguous sites were resolved as per the criteria employed by Hobbs et al. (2012) and as performed for MurA ancestral inference (section 2.1.4.3, section 3.2.1.3). The posterior probability at each site from the amino acid method for *Bacillus* AroA LCA sequence was 0.92. The inferred ancestral sequence is provided below:

LCA

MKEKTINTAEKGLNGTIRIPGDKSISHRAVMFGAIAEGTTTVKNFLPGEDCLSTI
ACFRKLGVEIEQNGDDVTINGKGLDGLKEPKDVLDVGNSGTTIRLMLGILANRPF
HSTIIGDESIAKRPMKRVTDPLREMGAQIDGREDGKFTPLSIRGGNLKGIDYNSP
VASAQVKSAILLAGLQAEGTTTVTEPHKSRDHTERMLEAFGVKVEEEGQTVSIEG
GQTLKGTDIEVPGDISSAAFFLVAGAIVPNSRIVLKNVGLNPTRTGIIDVLKKMG
ANLEIDQVREEGAEPYGDLTIETSSLKGIEIGGDLIPRLIDEIPIIALLATQAEG
TTVIKDAEELKVKETNRIDTVVSELKKLGANIEATDDGMIIHGKTTLKGGATVDS
HGDHRIGMMLAIASCIAEGEVEIEDTDAVSVSYPKFFEHLESLKK

The conserved and catalytically important residues mentioned in section 4.2.1.1 were also identified in the LCA sequence (highlighted residues in the above sequence).

The pairwise identity of *Bacillus* AroA LCA to the extant *Bacillus* sequences shows that the LCA shares the highest pairwise sequence identity of 82% to *Bacillus cytotoxicus* and the lowest of 64% to *B. cellulosilyticus* (Table 4.1). This shows that AroA LCA is not disproportionately affected by the deep-branching *B. cytotoxicus*. As the ancestral sequence is not significantly more similar to one contemporary sequence than the others, it is improbable that any extant sequence would pose bias on the ancestral node.

Table 4.1: Pairwise sequence identities (%) between extant and ancestral protein sequences.

	LCA
Bacillus amyloliquefaciens	77
Bacillus anthracis	75
Bacillus atrophaeus	75
Bacillus cellulosilyticus	64
Bacillus cereus	76
Bacillus clausii	71
Bacillus coagulans	67
Bacillus coahuilensis	65
Bacillus cytotoxicus	82
Bacillus halodurans	73
Bacillus licheniformis	75
Bacillus megaterium	76
Bacillus mycoides	75
Bacillus pseudomycoides	74
Bacillus pseudofirmus	74
Bacillus pumilus	80
Bacillus selenitireducens	67
Bacillus subtilis	77
Bacillus thuringiensis	75
Bacillus weihenstephanensis	75

Blue cell: Highest pairwise identity between ancestral and extant sequences
Green cell: Lowest pairwise identity between ancestral and extant sequences

The reconstruction of *Bacillus* AroA LCA resulted in high posterior probability as well as a low number of ambiguous sites. This indicates a high probability of reconstructing a functional ancestral enzyme. Therefore, as performed for MurA, the process of characterising *Bacillus* AroA LCA, for its comparison with a contemporary AroA enzyme, was started by cloning the genes encoding these proteins followed by their expression and purification.

4.2.2 Cloning, Protein expression and purification

4.2.2.1 Cloning

Bacillus AroA LCA gene was obtained from GeneArt (Life Technologies) and *B. subtilis* AroA was amplified from *B. subtilis* 168 genomic DNA using gene specific primers (section 2.2.2.1). Both *Bacillus* AroA LCA and *B. subtilis* AroA genes were then cloned into pPROEX HTb vector with an N-terminal His-tag using *Bam*HI and *Xho*I restriction sites. pPROEX HTb containing the gene was used to transform either *E. coli* DH5α or BL21 cells and screened for positive transformants as explained in section 2.2.8.

4.2.2.2 Small-Scale Protein Expression

Small-scale expression of *Bacillus* AroA LCA protein (Lysis buffer: 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM imidazole) resulted with part of the protein in the soluble fractions and majority of protein remaining in the insoluble fraction. An overall increase in protein expression was observed with increase in time (Figure 4.2). A lysis buffer screen was performed to improve the solubility of protein and did not result in increased protein solubility.

The expression of *B. subtilis* AroA in either *E. coli* DH5α or BL21 cells grown in LB media resulted in very low protein expression in soluble and insoluble fractions. Therefore, the *E. coli* cells with *B. subtilis* AroA were grown and protein expressed in terrific broth (TB media) (Appendix E3) to increase the total amount of expressed protein obtained from cultures. A Ni^{2+} bead affinity test confirmed the Ni affinity to the N-terminal His tag and protein expression of *B. subtilis* AroA in TB media (Figure 4.2). The molecular weight of the monomeric form of *Bacillus* AroA LCA and *B. subtilis* AroA were predicted to be 45.9 KDa and 45.2 KDa respectively by Protparam (Gasteiger et al. 2005) and was used as reference while analysing the SDS-PAGE gel bands.

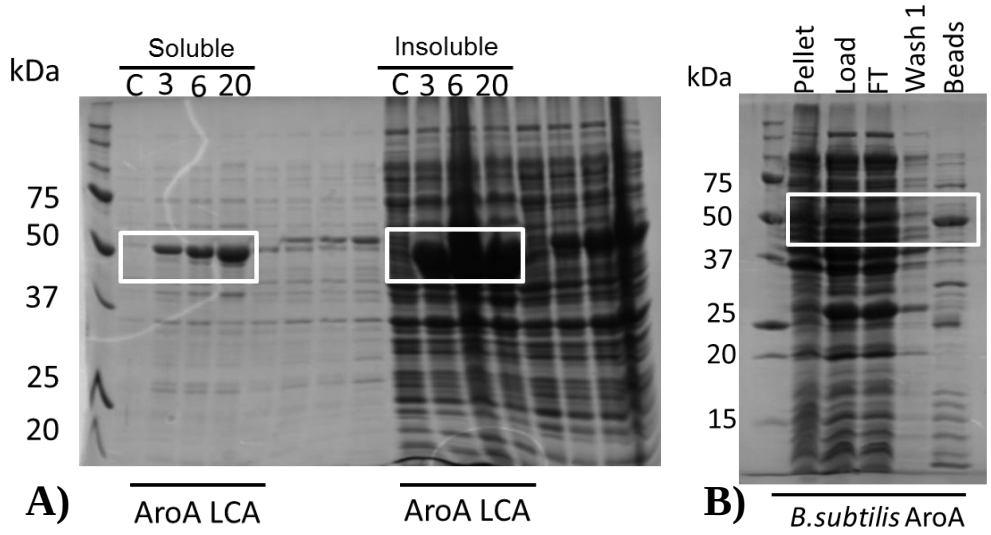


Figure 4.2: SDS-PAGE gels of small-scale expression trials for AroA enzymes.
A) *Bacillus* AroA LCA soluble and insoluble expression trial B) *B. subtilis* AroA Ni^{2+}-sepharose bead expression trial. White boxes indicate the expected band size. Labels for A): C - control sample taken prior induction; Numbered lanes indicate the number of hours after which the samples were taken post induction (3, 4, 6 and 20 hours). Labels for B): FT- flow through.

4.2.2.3 Protein Purification

Protein expression of *Bacillus* AroA LCA was scaled up to 1 L LB cultures and was initially purified using Ni affinity chromatography. Protein was eluted at 16 % of the elution buffer (50 mM Tris, 300 mM NaCl, 1 M imidazole pH 7.5), that is approximately 160 mM imidazole. Fractions corresponding to the peak were then run on an SDS-PAGE gel to confirm the size of the *Bacillus* AroA LCA protein (Figure 4.3).

B. subtilis AroA, unlike *Bacillus* AroA LCA, was not codon optimized for *E. coli* protein expression and therefore expressed in low quantities even when scaled up (Figure 4.3). The optimum procedure to produce as much *B. subtilis* AroA protein as required for enzyme characterisation assays, was to grow the cells in 500 ml cultures of TB media and then pooling together protein purified using Ni affinity chromatography from two such cultures for further SEC purification. *B. subtilis* AroA protein was found to be eluted at 34 % of elution buffer (50 mM Tris, 300 mM NaCl, 1 M imidazole pH 7.5) at 340 mM imidazole (Figure 4.3). The region indicated with a red bar on the *B. subtilis* AroA chromatogram corresponds to the fractions in which *B. subtilis* AroA protein is present (Figure 4.3).

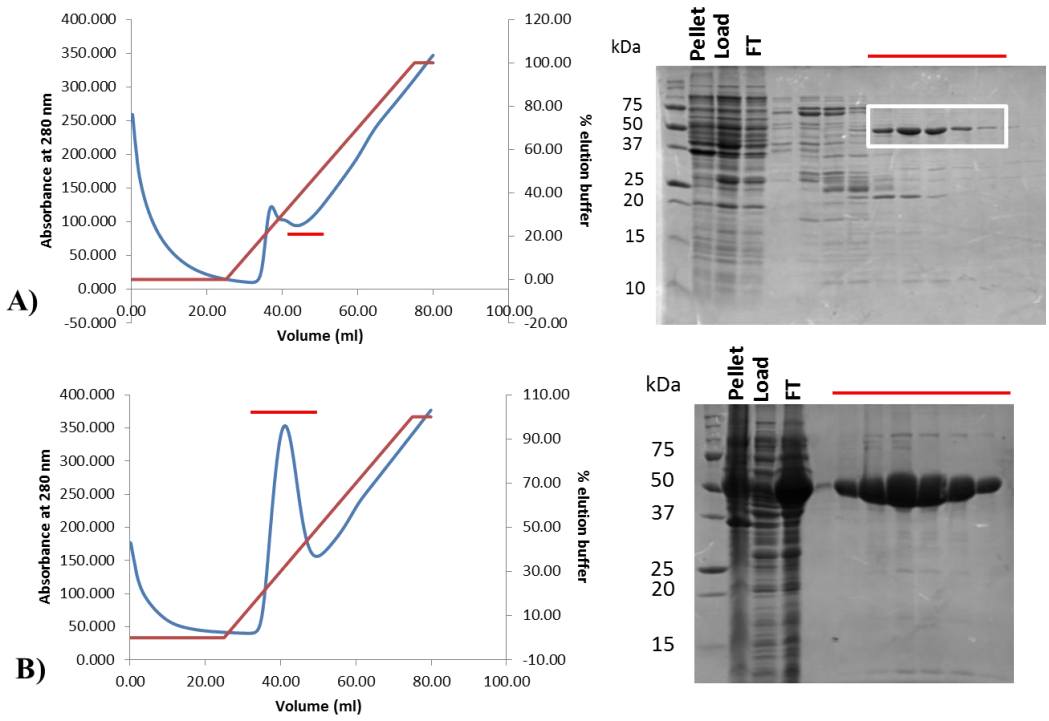


Figure 4.3: IMAC purification of recombinant AroA enzymes.
The IMAC chromatograms and corresponding SDS-PAGE gels for A) *Bacillus subtilis* AroA and B) *Bacillus* AroA LCA. The chromatograms show the UV absorbance at 280 nm. The SDS-PAGE gel demonstrates the size and purity of AroA enzymes. The SDS-PAGE gel shows the insoluble pellet, the column load, flow through (FT) and AroA fractions eluted during IMAC. The white box indicates *Bacillus subtilis* AroA protein. The fractions and their elution position are indicated by the red bar.

The fractions of AroA proteins obtained from Ni affinity purification were further purified using size-exclusion chromatography (SEC) to separate the protein monomers and other oligomers. It was observed that *Bacillus* AroA LCA had a prominent peak, whereas, *B. subtilis* AroA resulted in a chromatogram with multiple peaks (Figure 4.4). The fractions corresponding to peaks on the chromatogram were run on an SDS-PAGE gel to identify the protein fractions and their purity (Figure 4.4).

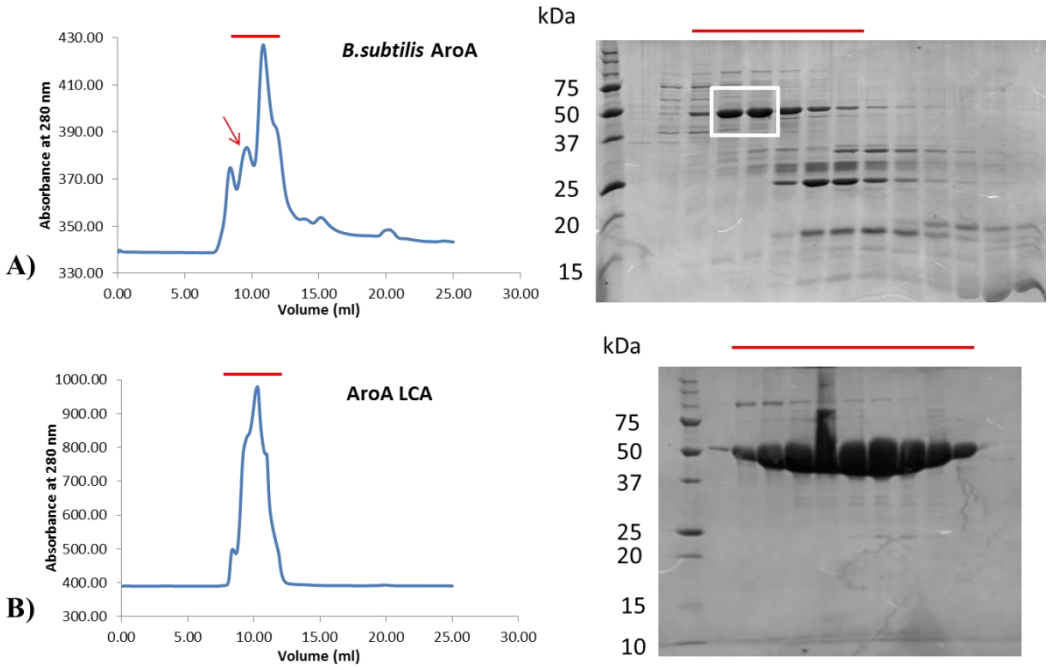


Figure 4.4: Size-exclusion chromatography (SEC) purification of recombinant AroA enzymes.
The SEC chromatograms and corresponding SDS-PAGE gels for A) *B. subtilis* AroA and B) *Bacillus* AroA LCA. The chromatogram shows the UV absorbance at 280 nm. The SDS-PAGE gel demonstrates the size and purity of AroA enzymes. The SDS-PAGE gel shows the AroA fractions eluted during SEC. The white box indicates the two pure *B. subtilis* AroA fractions used for enzyme characterisation assays, which corresponds to the peak indicated in the chromatogram by the red arrow. The fractions and their elution position in the chromatogram are marked by the red bar.

A Bradford assay was performed to quantify AroA protein (section 2.3.2). The purified AroA protein was then used for respective enzyme characterisation assays. However, before the enzyme assays were performed, the AroA enzyme substrate shikimate-3-phosphate (S3P) was required to be synthesised in the laboratory due to the high costs of buying the substrate. The synthesis of substrate S3P is described in the following section.

4.2.3 S3P synthesis

Shikimate-3-phosphate is one of the substrates utilized by AroA onto which the enolpyruvyl moiety from PEP is transferred. AroA's preceding step is the formation of shikimate-3-phosphate by shikimate kinase from shikimate. This reaction mechanism can be utilized for the enzymatic synthesis of shikimate-3-phosphate and its further purification.

4.2.3.1 Shikimate kinase preparation and preliminary S3P synthesis tests

Shikimate kinase cloning, protein expression and purification:

The gene encoding shikimate kinase II (*aroL*), was obtained from GeneArt (Life Technologies), and was cloned into an in-house modified pET28b-*Pst*I using the restriction sites *Pst*I and *Xho*I with a C-terminal His-tag. The cloned *aroL* was used to transform *E. coli* BL21 cells as explained in section 2.2.8. A Ni^{2+} bead small scale expression trial in shikimate kinase lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 20 mM imidazole) (Akiyama et al. 1999; Oliveira et al. 2001) was performed to confirm the expression of shikimate kinase in LB media at 37 °C overnight (Molecular weight: 19.2 kDa) (Figure 4.5). Shikimate kinase was then expressed in large scale cultures with LB media at 37 °C and nucleic acids precipitated from the culture supernatant by the addition of 1% w/v streptomycin sulfate (final concentration) (Oliveira et al. 2001; Oliveira et al. 2003). Shikimate kinase was then purified using Ni-affinity chromatography (Shikimate kinase lysis buffer: 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 20 mM imidazole; shikimate kinase elution buffer: 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 1 M imidazole) and size exclusion chromatography (Shikimate kinase SEC buffer: 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 50 mM KCl). The fractions containing shikimate kinase protein were identified using SDS-PAGE (Figure 4.5). The purified protein was quantified using Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA).

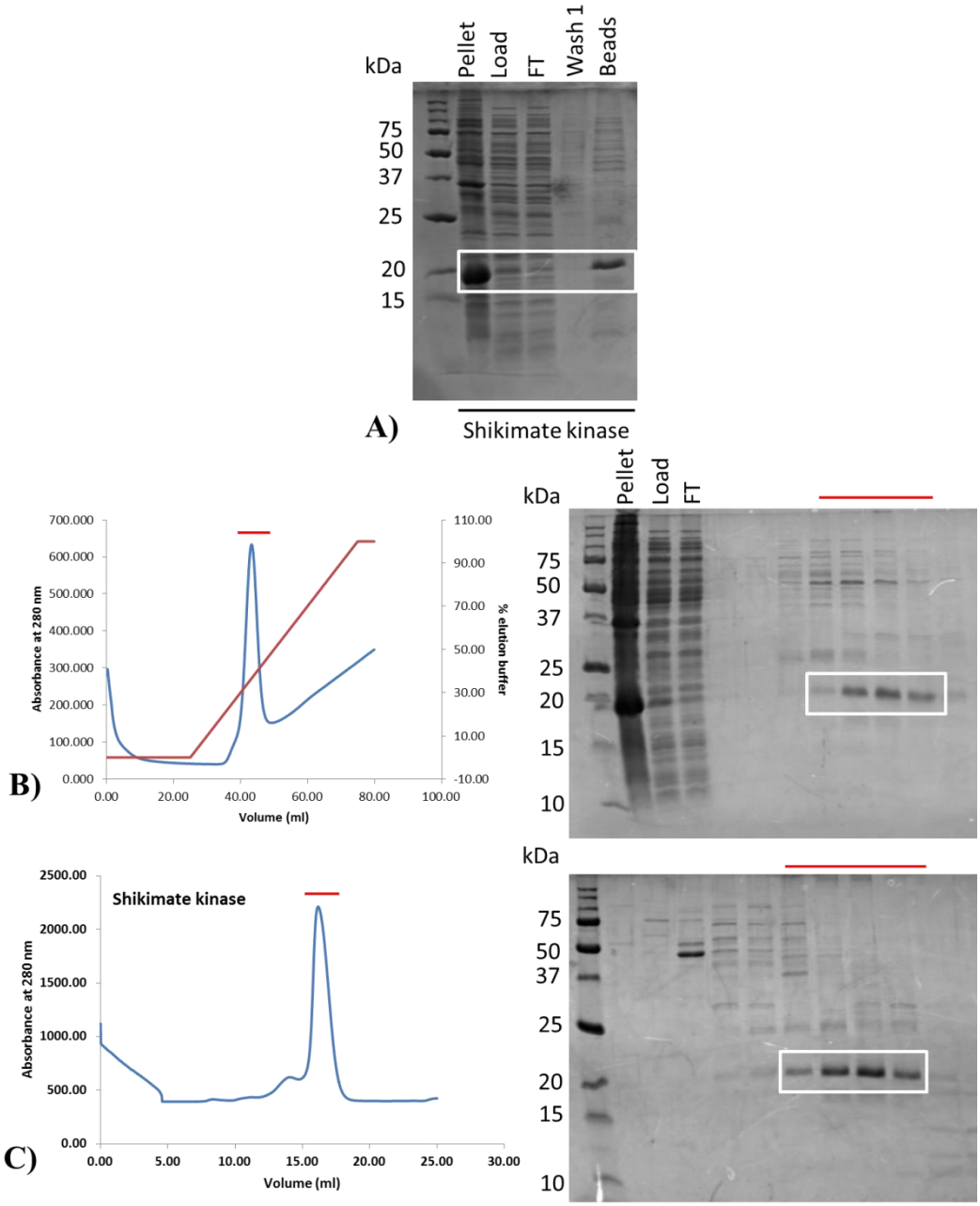


Figure 4.5: Shikimate kinase expression trial and purification.
A) Shikimate kinase Ni^{2+} bead expression trial. B) Shikimate kinase Ni affinity purification chromatogram showing UV absorbance at 280 nm along with SDS-PAGE gel demonstrating the size and purity of shikimate kinase. C) SEC purification chromatogram of shikimate kinase enzyme showing absorbance at 280 nm and the SDS-PAGE gel confirming the size and purity of the shikimate kinase fractions. Fractions and their elution positions are marked with the red bar. White boxes indicate the expected band size. Label for A) and B): FT- flowthrough.

Shikimate kinase activity assay:

The activity of shikimate kinase was assayed at room temperature using a three reaction coupled assay , with shikimate dependant oxidation of NADH monitored at 340 nm (Millar et al. 1986). The reactions are as follows:

$$\text{Shikimate} + \text{ATP} \xrightarrow{\text{Shikimate kinase}} \text{S3P} + \text{ADP}$$

$$\text{PEP} + \text{ADP} \xrightarrow{\text{Pyruvate kinase}} \text{Pyruvate} + \text{ATP}$$

$$\text{Pyruvate} + \text{NADH} \xrightarrow{\text{Lactate dehydrogenase}} \text{Lactate} + \text{NAD}^+$$

The tests were carried out with 50 mM Tris pH 7.0, 50 mM KCl, 5 mM $MgCl_2$, 2.5 mM shikimate, 0.125 mM ATP, 2.5 mM PEP, 0.25 mM NADH, 0.006 mg/ml pyruvate kinase, 0.007 mg/ml lactate dehydrogenase and 0.125 mg/ml shikimate kinase final concentration in 400 µl final volume reaction mixture. The reaction cascade was started with the addition of shikimate and compared to a blank reaction lacking the presence of shikimate. Shikimate kinase activity was confirmed by monitoring the decrease in absorbance at 340 nm as NADH gets oxidised to NAD^+ (Figure 4.6).

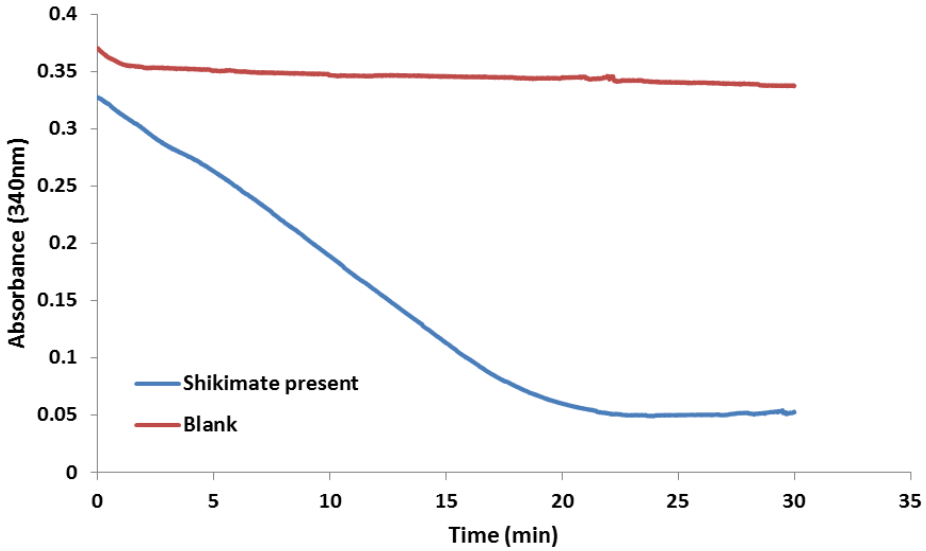


Figure 4.6: Shikimate kinase activity monitored at 340 nm.
The shikimate dependant activity of shikimate kinase in a three reaction coupled assay is illustrated through the decrease in absorbance at 340 nm as NADH gets oxidised to NAD^+ (Blue). The blank reaction with no shikimate shows a constant absorbance at 340 nm (Red).

S3P reaction activity tests using HPLC:

Preparation of S3P and its purification has previously been successful using HPLC (Akiyama et al. 1999; Okunuki et al. 2003). Therefore, preliminary attempts for the synthesis and purification of S3P were based on these previous procedures. The reaction mixtures were set up by modifying the reaction used for the shikimate kinase activity test (Millar et al. 1986). The S3P reaction mixture contained a final concentration of 50 mM Tris pH 7.0, 50 mM KCl, 5 mM $MgCl_2$, 1.6 mM shikimate, 2.5 mM ATP and 0.2 mg/ml shikimate kinase enzyme in a final volume of 400 µl; while the control reaction was set up with no shikimate kinase enzyme. The mixtures were left overnight at 25 °C followed by deproteination using a protein concentrator with a 5 KDa or 10 KDa molecular weight cutoff. Attempts to purify S3P were performed using a DiscoveryBIO wide pore C18 HPLC column (Sigma-Aldrich) (Figure 4.7) (section 2.3.9). Albeit the differences between the S3P reaction and control reaction chromatograms, the resolution of the chromatographic separation was too low to purify S3P from the reaction. The difference in HPLC columns, with the DiscoveryBIO wide pore C18 HPLC column used in this attempt versus the TSK gel CARBON-500 column (Tosoh Co.) used in the previously mentioned papers, may have been one of the reasons for poor separation of the reaction components.

Multiple tests were performed to identify each peak and to confirm the presence of S3P in the reaction. Individual runs of each component separately resulted in peaks at the following retention times: ATP: 3.1 and 3.3 minutes, ADP: 3.3 and 3.5 minutes, Shikimate: 3.675 minutes. Therefore, Figure 4.7 suggests the disappearance of shikimate and ATP to produce ADP and S3P, as the reaction is meant to proceed. Although the chromatograms of S3P reaction and control reaction suggest the presence of S3P in the reaction, the poor separation of the components makes it inconclusive.

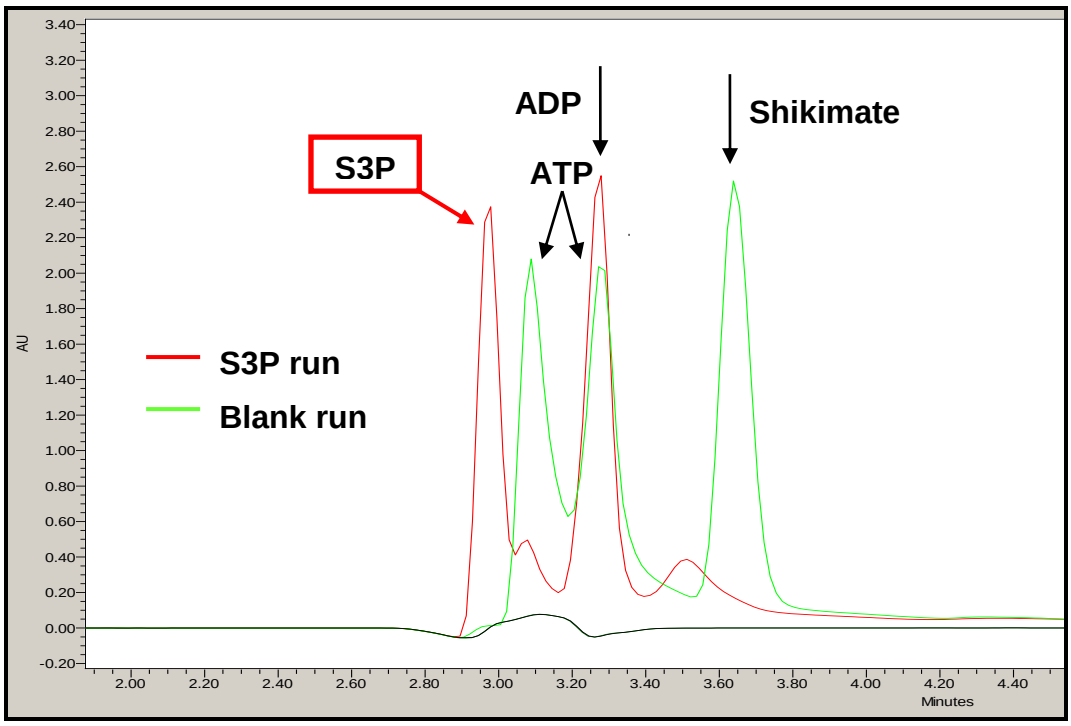


Figure 4.7: HPLC purification of S3P reaction and control reaction.
The HPLC chromatograms of S3P reaction run (red) and control reaction run (green) are shown in the figure. The peak labels were assigned according to the retention time of the peaks obtained from HPLC runs of individual components. The chromatograms suggest a successful reaction between ATP and shikimate producing S3P and ADP, with the S3P peak as labelled, however, inconclusive.

4.2.3.2 MonoQ

Synthesis of S3P was achieved by optimising and modifying enzymatic S3P synthesis protocols (Funke 2008; Pedroso et al.). Shikimate kinase protein was expressed and purified as previously described (section 4.2.3.1) and the protein was thoroughly dialysed against 300 mM Tris pH 7.6, 100 mM KCl and 5 mM $MgCl_2$ (S3P buffer) post SEC purification.

For large scale synthesis of S3P, 2 ml reactions were set up containing a final concentration of 50 mM ATP, 50 mM shikimate, 0.2 mg/ml shikimate kinase enzyme and S3P buffer. The pH of the reaction was adjusted to 6.6 by adding 10 M sodium hydroxide and the reaction was left overnight at 25 °C. A control reaction was also set up with no shikimate kinase enzyme. Prior to the MonoQ purification, S3P reactions were deproteinated by passing the mixture though a protein concentrator with 10 KDa molecular weight cutoff.

S3P was purified by anion exchange with a MonoQ 4.6/100 PE column (GE Healthcare Life Sciences) using a gradient ranging from 0-0.5 M NH_4HCO_3 pH 9.0. Each run was carried out by loading 1 ml of deproteinated S3P reaction

mixture. An overlay of the S3P reaction and control reaction chromatograms showing UV absorbance at 254 nm indicates a successful reaction as observed from the decrease in ATP and shikimate and increase in ADP (Figure 4.8).

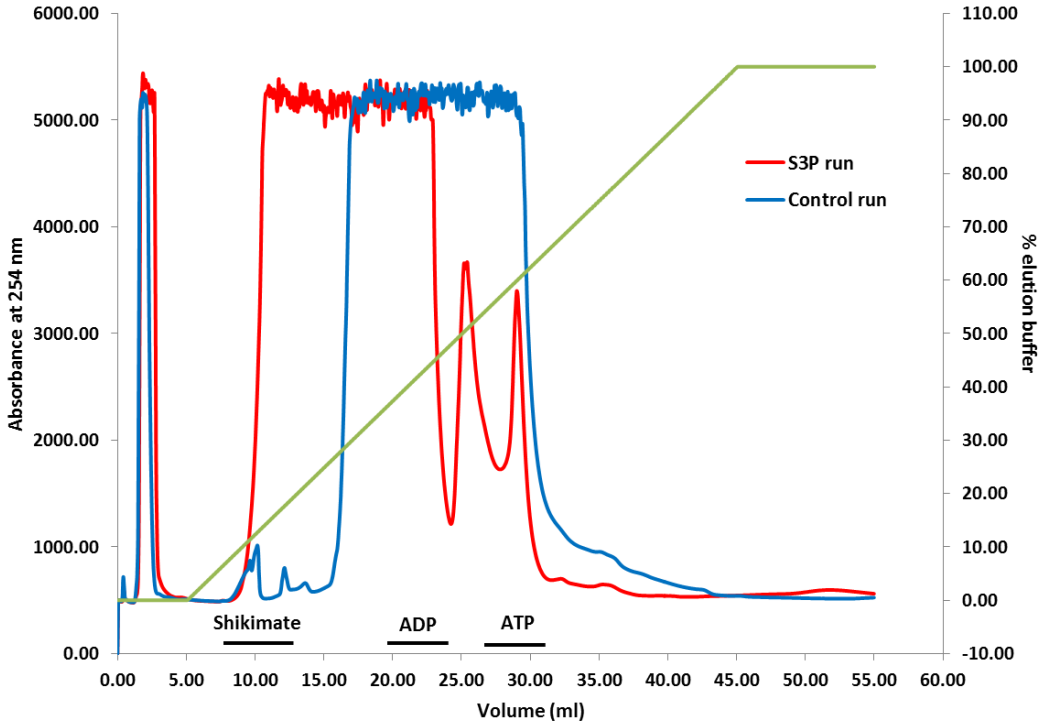


Figure 4.8: Overlay of MonoQ purification of S3P reaction and control reaction chromatograms.
The chromatogram shows absorbance at 254 nm. The black bars identify the respective positions of shikimate, ADP and ATP as determined from separate MonoQ purifications of individual reaction components. The high amounts of ATP and ADP in the reactions have resulted in wide peaks in addition to surpassing the absorbance detection range. Although the separation between each individual component is not very clear, there is a clear decrease in ATP and increase in ADP illustrated by the peak shifts between control (blue) and S3P (red) reactions.

The presence of S3P in the fractions was detected by catalytic activity of AroA using the malachite green assay. The fractions containing S3P, were pooled and freeze dried. The purity and identity of the sample was inspected using NMR and UV spectroscopy methods.

4.2.3.3 NMR

The identity of S3P was reconfirmed by performing an NMR analysis. A sample of the freeze dried S3P was dissolved in deuterated water (D_2O) to perform ^{1}H and ^{31}P (coupled) NMR analysis on Bruker Avance III spectrometers. The ^{1}H NMR spectra (400.14 MHz) was referenced internally using the residual protio solvent resonance relative to tetramethylsilane (δ 0) and ^{31}P (161.97 MHz) externally to 85% H_3PO_4 (δ 0). Although inconclusive, the ^{1}H spectrum obtained

suggested the possible presence of shikimate, ADP as well as shikimate-3-phosphate in the sample. The ^{31}P NMR analysis identified two peaks for ADP of shift -6.2169 ppm and -10.6805 ppm from the β and α phosphates of ADP respectively (Figure 4.9, Table 4.2). This observed shift for ADP peaks is consistent with the literature (Mascaros et al. 2011; Yu et al. 2013). The minor difference in shift from the β phosphate of ADP measured by Mascaros et al. (2011) may be due to difference in pH of the sample. The presence of ADP in the sample was expected as S3P co-eluted with ADP, similar to the S3P purification performed by Funke et al. (2008). Another peak was found at a shift of 4.3989 ppm which corresponds to the phosphate from S3P (Figure 4.9, Table 4.2). This is consistent with the literature which reports S3P showing a chemical shift of ~4.5 ppm (Ge et al. 2010) and AroA enzyme bound S3P (3 ppm) (Jakeman et al. 1998) to be 2 ppm upfield from unbound S3P (5 ppm) (Christensen & Schaefer 1993). Minor shift differences may be a result of pH differences of samples. Although the presence of S3P was established using the catalytic activity of AroA, NMR analysis aimed to ensure the same, as well as estimate the purity of the sample by determining the different species present in the sample. Therefore, the S3P stocks used for AroA enzyme characterisation contained ADP and possibly shikimate.

Table 4.2: ^{31}P NMR peak properties

	No: of peaks	Shift (ppm)	Multiplicity	Coupling constant
S3P	1	4.3989	Doublet	9 Hz (P-H)
ADP	2	-6.2169	Doublet	22 Hz (P-P)
		-10.6805	Doublet	22 Hz (P-P)

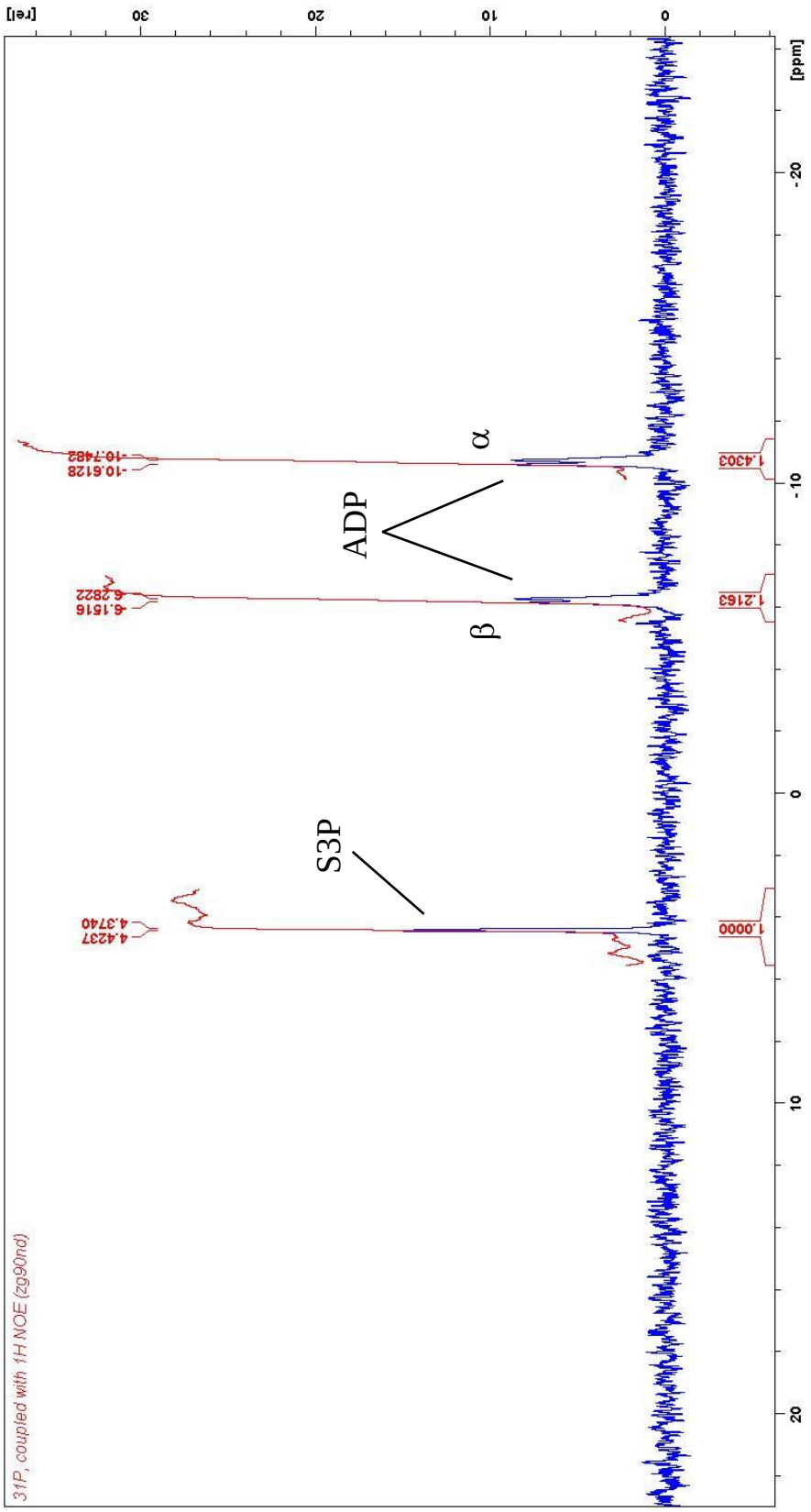


Figure 4.9: The 161.97 MHz ^{31}P NMR spectra of S3P and ADP.
The sample is freeze dried S3P dissolved in D_2O. The ^{31}P chemical shift of S3P is 4.3989 ppm and the ^{31}P chemical shift of β and α phosphate of ADP are -6.2169 ppm and -10.6805 ppm respectively. The shift reference is 85% H_3PO_4 (δ 0) (external).

4.2.3.4 Concentration and purity

The purity and concentration of S3P within the stock was further examined using AroA enzymatic reaction. As the amount of inorganic phosphate produced is proportional to the amount of initial S3P present in the reaction, the absolute amount of S3P present in freeze dried powder was determined by allowing the AroA reaction to reach completion and calculating the amount of inorganic phosphate produced using a phosphate standard graph (section 2.3.7.2). This was in turn used to estimate that the S3P in the freeze dried 'S3P powder' was 12.7% pure. The AroA enzymatic reaction was used to determine the exact concentration of S3P in every stock solution used for the AroA enzyme characterisation assays.

4.2.4 Enzyme characterisation

4.2.4.1 Activity test and Time course assay

Similar to MurA enzymatic assays, preliminary tests were performed to optimize the assay conditions. To eliminate the contribution of substrates towards the inorganic phosphate within the reaction, AroA reactions were carried out with control/blank reactions for all the substrate concentrations and temperatures.

Fractions from the prominent peak of *Bacillus* AroA LCA SEC purification and the purest fractions from the *B. subtilis* AroA SEC purification were chosen for enzymatic characterisation assays (Figure 4.4). These fractions, based on the reference chromatograms for SEC column S200 10/300 and S75 10/300 (GE Healthcare) were found to correspond to the molecular weight of a dimeric AroA protein. A time course assay was performed and a reaction time of 2 minutes, ie. time for which the reaction was allowed to proceed before adding stop solution, was manually confirmed from the graph to remain within the linear phase of the reaction.

4.2.4.2 Michaelis Menten kinetics

The malachite green assay was used to determine the kinetic properties of ancestral and contemporary AroA enzymes. The substrates were added in the most feasible ways to determine the results as accurately as possible.

Whenever possible, the substrate maintained at a constant concentration was added at ten times its K_M to ensure non-limiting amounts of substrate during K_M determinations. Otherwise, the maximum amount of substrate possible was added. All the assays were initially performed at 37 °C to accurately determine the T_{opt}

and K_M of substrates at the enzyme's T_{opt} with non-limiting amounts of substrate (Table 4.3). It was observed during PEP K_M assays that increased concentrations of PEP exhibited substrate inhibition. It has previously been proposed that substrate inhibition may occur as a result of competitive binding of PEP at the S3P binding site (Gruys et al. 1992). Therefore, assays were performed considering the factor of substrate inhibition. All k_{cat} values were calculated using V_{max} from S3P K_M. Under circumstances where non-limiting amounts of PEP were not added in S3P K_M assays due to substrate inhibition factors, the apparent V_{max} was used for k_{cat} determination.

Table 4.3: Kinetic constants for contemporary and ancestral AroA enzymes at 37°C

Enzyme	PEP K_M (μM)	S3P K_M (μM)	k_{cat} (sec^{-1})	k_{cat}/K_M (S3P) ($sec^{-1}M^{-1}$)
B. subtilis	110 ± 30	800 ± 200	9.0 ± 0.8	1.1 (±0.3)x10^4
LCA	90 ± 10	80 ± 10	3.0 ± 0.1	3.8 (±0.7)x10^4

The kinetic values were obtained from three technical replicates.
The errors represent the standard error of the mean

Wildtype *E. coli* AroA has a PEP K_M (88 μM, 140 μM) similar to that of the measured *Bacillus subtilis* AroA PEP K_M, yet with approximately six fold lower S3P K_M (120 μM, 140 μM) and with 3-6 times higher k_{cat} (28.8 sec^{-1}, 51 sec^{-1}) (Eschenburg et al. 2002; Priestman et al. 2005a).

The enzymatic activity of AroA has been shown to be influenced by monovalent cations such as potassium and ammonium ions with increased catalytic turnover (Majumder et al. 1995; Priestman et al. 2005a). In the case of AroA assays performed in this chapter, it is possible that potassium and ammonium ions are present in the assay mixtures (Potassium ions from KCl used in shikimate kinase purification buffers and ammonium ions from the NH_4HCO_3 buffer used in the MonoQ purification of S3P). Such residual amounts of potassium and ammonium ions may have been enough to activate the AroA enzymes. The measured *Bacillus subtilis* AroA K_M values were more comparable to the ammonium activated *Bacillus subtilis* AroA PEP K_M (90 μM) and S3P K_M (330 μM) (Majumder et al. 1995). *Staphylococcus aureus* AroA activated by KCl has a comparable k_{cat} (4.2 sec^{-1}) to the measured k_{cat} of *Bacillus subtilis* AroA and a comparable PEP

K_M (180 μM), though about six fold lower S3P K_M (140 μM) (Priestman et al. 2005a). All these kinetics values from the literature were measured by determining the amount of inorganic phosphate produced in the reaction using the malachite green assay (Lanzetta et al. 1979) at 20-37 °C.

Substrate affinity (K_M), depends on a variety of factors including temperature, and an increase of K_M values with increase in temperature has previously been reported (Griffith et al. 2001; Barrett 2011). Therefore, the kinetic properties of AroA enzymes were measured at the enzymes respective optimum temperature (T_{opt}) for accurate comparison. The optimum temperatures for both AroA enzymes were measured as described in section 4.2.6. The kinetic parameters for both *Bacillus* AroA LCA and *Bacillus subtilis* AroA determined at their respective T_{opt} are shown in Table 4.4 and Figure 4.10.

Table 4.4: Kinetic constants for contemporary and ancestral AroA enzymes at their respective optimal temperatures (T_{opt})

Enzyme	PEP K_M (μM)	S3P K_M (μM)	k_{cat} (sec^{-1})	k_{cat}/K_M (S3P) ($sec^{-1}M^{-1}$)	T_{opt} (°C)
B. subtilis	570 ± 90	1500 ± 200	21.0 ± 2.0	1.4 (±0.4)x10^4	46
LCA	130 ± 30	340 ± 80	10.0 ± 1.0	3.0 (±1.0)x10^4	51

The kinetic values were obtained from three technical replicates.
The errors represent the standard error of the mean

As previously seen for *B. subtilis* AroA at 37 °C, the K_M of S3P is more than two folds higher than the K_M obtained for PEP. While having a lower k_{cat} than *B. subtilis* AroA, it must be noted that AroA LCA exhibits twice the catalytic efficiency of contemporary AroA.

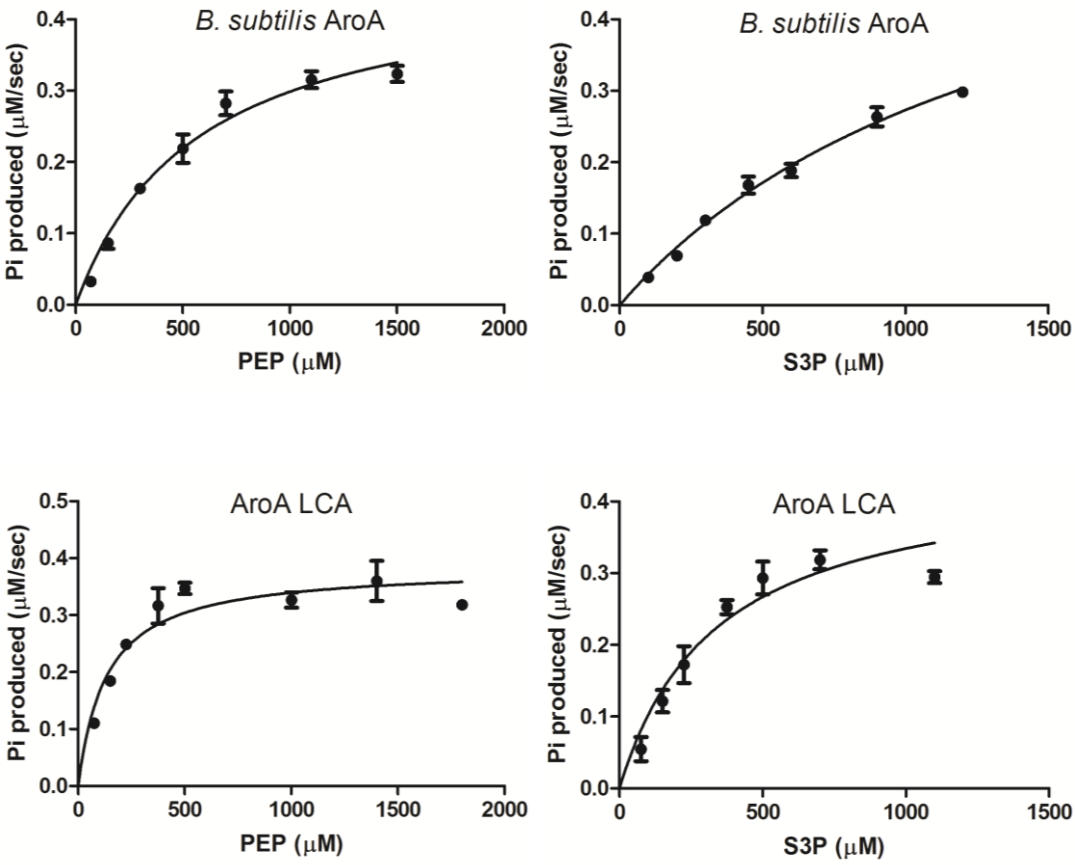


Figure 4.10: Michaelis-Menten plots for contemporary and ancestral AroA enzymes at their respective T_{opt} values.
Results are the mean of three technical replicates, the error bars representing the SEM.

Overall, the kinetic parameters determined for *Bacillus* AroA LCA were comparable to the contemporary AroA enzyme kinetic parameters, thus validating an accurate reconstruction of AroA LCA enzyme. It must be noted that, although PEP K_M of *B. subtilis* AroA and AroA LCA were comparable, the S3P K_M of *B. subtilis* AroA is 3-6 folds higher than the S3P K_M for AroA LCA (Table 4.4). Further information on why the AroA LCA has a higher affinity towards S3P than *B. subtilis* AroA may have to be determined *via* structural analysis of AroA LCA bound to S3P.

The inhibitory properties of AroA enzymes were also measured in a similar way as done for the MurA enzymes in chapter 3.

4.2.5 Inhibitory properties: IC50

Multiple studies have focused on the inhibition of AroA, due to the rise in tolerance towards its most commonly used inhibitor glyphosate, the prime

ingredient of the herbicide Roundup. Inhibition of AroA in pathogenic organisms has also been of increased interest as disruption of AroA activity or shikimate pathway has been associated with attenuated virulence (McDevitt et al. 2002) and viability of microorganisms (Parish & Stoker 2002; Funke et al. 2007). The evolution of AroA enzyme may be investigated by determining the inhibitor sensitivity of *Bacillus* AroA LCA in comparison to the contemporary AroA enzyme.

4.2.5.1 Inhibition by glyphosate

Depending on the resistance of AroA towards its inhibitor glyphosate, AroA enzymes are classified into two groups, Class I and Class II. Studying the glyphosate sensitivity of AroA LCA may provide information pertinent to the evolution of these two classes of AroA enzymes.

Class I AroA are naturally sensitive to glyphosate, yet they can be modified to acquire glyphosate tolerance which comes often at the expense of having a decrease in PEP affinity (Sun et al. 2005). From different studies, the IC50 value for glyphosate for wild type *E. coli*, a known class I AroA, may range from as low as 8.6 μM (Priestman et al. 2005a) to as high as 55 μM (Tian et al. 2013). Mutagenesis of class I *E. coli* to P101S conferred glyphosate tolerance exhibiting IC50 of 5 mM (Stallings et al. 1991).

Class II AroA are naturally glyphosate tolerant with high affinity for PEP. Less than 30% sequence identity is observed between class I and class II AroA (Sun et al. 2005). While the presence or absence of cations was shown to affect the property of certain AroA enzymes, *Staphylococcus aureus* AroA was classified as true class II AroA with IC50 of 0.9 mM to 1.6 mM (Priestman et al. 2005a). *Bacillus cereus* AroA was classified as class II AroA with IC50 of 0.5 mM (Tian et al. 2013).

Bacillus subtilis AroA is also classified as class II AroA (Barry et al. 1992; Sun et al. 2005). The IC50 values of *B. subtilis* AroA and AroA LCA were measured as explained in section 2.3.7.6. The dose-response curves and the IC50 values are given in Figure 4.11 and Table 4.6. The IC50 of *B. subtilis* AroA was found to be very high (4.4 mM), confirming its glyphosate tolerance. *B. subtilis* AroA also showed high affinity to PEP at 37 °C (110±30 μM) which agrees with the characteristic of class II enzymes.

Bacillus AroA LCA, on the other hand, has an IC50 of 0.4 mM, which is in the range of class II *S. aureus* AroA and *Bacillus cereus* AroA, however, approximately 12 fold more sensitive in comparison to *B. subtilis* AroA (Note that fosfomycin sensitivity of MurA obtained by malachite green assay was 0.2-0.3 mM). However, AroA LCA exhibits equivalent affinity for PEP as *B. subtilis* AroA at 37 °C and a four fold higher affinity for PEP than *B. subtilis* AroA at its T_{opt} (Table 4.3 and Table 4.4).

Although the results of assays performed under different experimental conditions may vary, comparison to values reported in the literature may provide a range to which values may be compared (Table 4.5).

Table 4.5: Class I and Class II AroA enzymes and their IC50 values

Enzyme	Classification	IC50 (mM)	Reference
Wild type *E.coli*	Class I	0.009-0.06	(Priestman et al. 2005a; Tian et al. 2013)
E.coli P101S	Class I with glyphosate tolerance	5	(Stallings et al. 1991)
Staphylococcus aureus	Class II	0.9-1.6	(Priestman et al. 2005a)
Bacillus cereus	Class II	0.5	(Tian et al. 2013)
Bacillus subtilis	Class II	4.4*	(Barry et al. 1992; Sun et al. 2005)**
Bacillus AroA LCA	-	0.4*	-

* Measured IC50 values for enzymes
** *Bacillus subtilis* classified as class II AroA (Barry et al. 1992; Sun et al. 2005)

Various mutagenesis studies have identified different amino acid residues which affect glyphosate sensitivity. Nevertheless, the specific amino acid residues that represent the signature of class I or class II enzymes are not known. Therefore, categorization of *Bacillus* AroA LCA cannot be performed with confidence. From the phylogenetic tree used to infer the LCA of *Bacillus*, *B. subtilis* and *B. cereus* are classified as class II AroA enzymes. However, the class of the rest of the *Bacillus* species is not known. AroA LCA shows characteristics of class II

enzymes in the following ways: 1) Class I and class II AroA share less than 30% amino acid identity (Tian et al. 2012), yet *Bacillus* AroA LCA shares ≥64% sequence identity to the *Bacillus* species (Table 4.1) and only 28.3% sequence identity to class I *E. coli* AroA, 2) AroA LCA exhibits an identical high affinity towards PEP as class II *B. subtilis* AroA, 3) The mutation of Pro101 in *E. coli* to Ser101 confers glyphosate resistance which may be the result of small alterations around the binding pocket for glyphosate (Priestman et al. 2005a). Sequence alignment revealed that the Pro101 residue of *E. coli* is substituted by a leucine residue in class II *S. aureus* AroA and all *Bacillus* AroA sequences used for ancestral inference as well as in *Bacillus* AroA LCA, and 4) All the residues identified by Park et al (2004), involved in glyphosate interaction from class II *S. pneumoniae* AroA (K20, S21, R25, G92, T93, R120, R124, S166, A167, Q168, D312, K339, E340, R343, H384 and R385) were also conserved in *Bacillus* AroA LCA, suggesting that the difference in sensitivity may not be due to major residue differences, but rather subtle changes of the glyphosate binding site caused by other amino acid changes. Although the above evidence suggests *Bacillus* AroA LCA to be a class II AroA, the higher glyphosate sensitivity of LCA in comparison to *B. subtilis* AroA does not concur. Therefore, either the present day class II *Bacillus* species with natural glyphosate tolerance evolved from the glyphosate sensitive *Bacillus* AroA LCA; or, if the reconstructed ancestral enzyme only represents a snapshot of the evolving host at a given point in time (as describe in section 3.2.5), then the LCA may only represent one species with glyphosate sensitivity, which is neither class I nor class II.

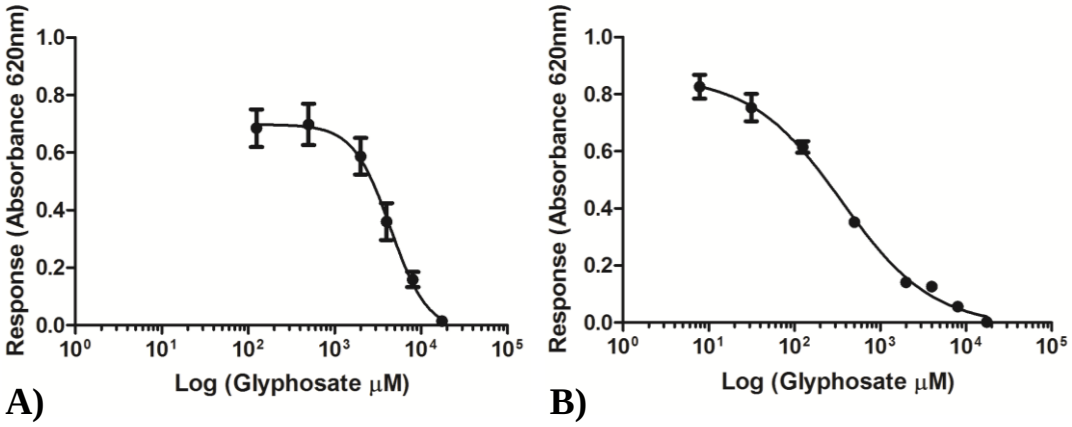


Figure 4.11: The IC50 dose-response curves obtained for different concentration of glyphosate A) *Bacillus subtilis* AroA B) *Bacillus* AroA LCA
Results are the mean of three technical replicates, the error bars representing the SEM

4.2.5.2 Inhibition by fosfomycin

As performed for MurA, the exclusive inhibition of MurA and AroA by fosfomycin and glyphosate respectively, was tested on the *Bacillus* AroA LCA enzyme. Inspite of the structural similarity between AroA and MurA, due to the spatial arrangement of residues in the enzyme's active site, the inhibitors are specific to contemporary AroA and MurA. Therefore, determining the sensitivity of AroA LCA towards the MurA inhibitor fosfomycin, would allow us to find out the presence of any structural or mechanistic properties in the ancestor that would indicate a similarity in active sites of AroA and MurA.

The IC50 for fosfomycin measured as explained in section 2.3.7.6 are given in Table 4.6 and the corresponding dose-response curves are given in Figure 4.12. *B. subtilis* AroA was found to be resistant to fosfomycin inhibition as expected. *Bacillus* AroA LCA also exhibited a similar non-specific inhibition, thereby suggesting the absence of any characteristic changes in LCA that would indicate similarity in inhibition to MurA.

Table 4.6: Glyphosate and fosfomycin IC50 values for *Bacillus subtilis* AroA and *Bacillus* AroA LCA

	***B.subtilis* AroA**	***Bacillus* AroA LCA**
Glyphosate	4.4 mM	0.4 mM
Fosfomycin	55.3 mM	48.4 mM

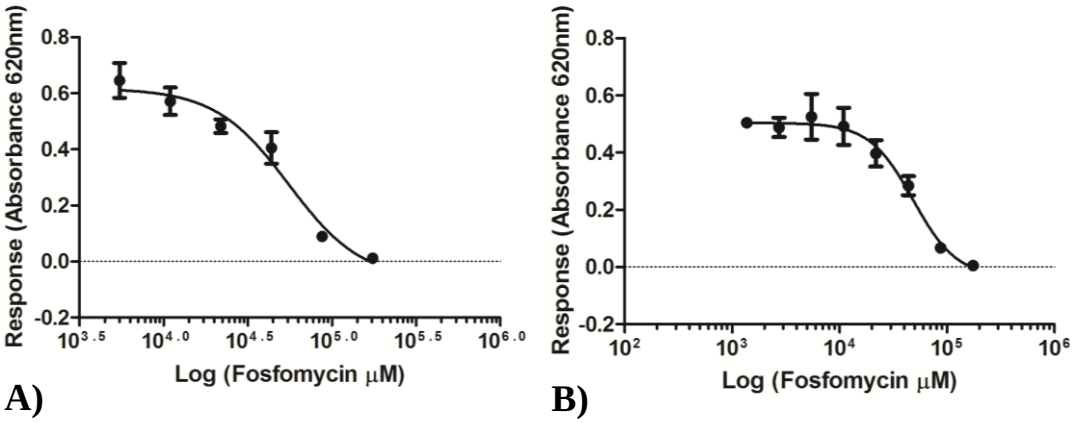


Figure 4.12: The IC50 dose-response curves obtained for different concentration of fosfomycin A*) Bacillus subtilis* AroA B) *Bacillus* AroA LCA.
Results are the mean of three technical replicates, the error bars representing the SEM

The optimum temperature of enzyme activity and enzyme melting temperature of ancestral and contemporary AroA enzymes were also measured for the study of the enzyme evolution.

4.2.6 Thermal properties

Core metabolic enzymes undergo certain changes to facilitate adaptation of their host organisms to the new environments resulting from evolutionary processes (Hobbs et al. 2012). Therefore, the thermal characteristics of AroA ancestral enzyme would also provide us with key evolutionary information.

The thermoactivity (T_{opt}) of AroA enzymes were measured over a broad temperature range followed by a narrow temperature range as performed for MurA enzymes. All the assays were performed by adding the maximum amount of substrate possible, considering substrate inhibition and malachite green assay sensitivity. Therefore, it was not possible to add ten times the substrate K_M and hence may have resulted in substrate limitation. The T_{opt} of *B. subtilis* AroA was found to be 46 °C which could still be considered to be within the mesophilic range as it is common for mesophilic organisms to show thermoactivity higher than their growth temperature (Georlette et al. 2004; Hobbs et al. 2012). Similarly, the T_{opt} of *Bacillus* AroA LCA was measured to be 51 °C falling within the moderately thermophilic range (Figure 4.13, Table 4.4).

The thermostability (T_m) of AroA enzymes was measured by performing real-time protein melt assays as described in section 2.3.8.2. The T_m values determined using real-time protein melt method has been shown to be in reasonable concordance with values obtained by DSC method (Easter 2010). The T_m of *B. subtilis* AroA was found to be 48 °C and that of *Bacillus* AroA LCA was found to be 62 °C (Figure 4.13).

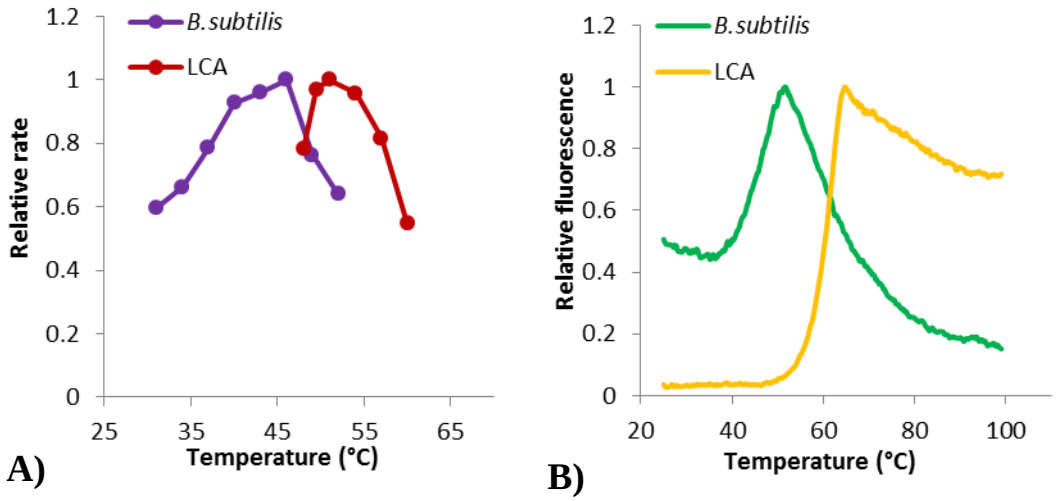


Figure 4.13: Thermoactivity and thermostability profiles for contemporary and ancestral AroA enzymes.
A) Thermoactivity data shown are the mean of at least three replicates and are expressed as a proportion of the highest activity (at T_{opt}) B) Thermostability data shown are real-time melts of *B. subtilis* AroA and *Bacillus* AroA LCA at pH 7.5 and pH 7 respectively.

The T_{opt} and T_m values were found to be consistent with each other. Gromiha et al (1999) indicates the correlation between the enzyme thermoactivity (T_{opt}) and thermostability (T_m) to the host organism growth temperature, which is hence attested by *B. subtilis*. Therefore, based on this correlation, the thermal properties of *Bacillus* AroA LCA suggest its host to have inhabited a moderately thermophilic environment.

The temperature profile exhibited by *B. subtilis* AroA (T_{opt} of 46 °C) and *B. subtilis* LeuB (T_{opt} of 53 °C) differ from each other, thus indicating the protein dependency of such properties (Hobbs et al. 2012). Similarly *Bacillus* AroA LCA (T_{opt} of 51 °C; T_m of 62 °C) and LeuB ancestor LeuB4 (T_{opt} of 70 °C; T_m of 65.3 °C), aged 957 myr and 950 myr respectively, differ in the extent of their thermophilicity substantially with respect to their T_{opt}, though similar in their T_m (Hobbs et al. 2012). Such variation in temperature profiles between enzymes may also concur with the hypothesis that the thermal properties are likely to reflect the changes in the microenvironment of the organism and does not indicate global environmental changes (Hobbs et al. 2012). The difference in thermoactivity between *Bacillus* AroA LCA and *Bacillus* LeuB4 from the same time period, may also point towards the idea that each reconstructed ancestor only represents a single candidate from the consortium of *Bacillus* species present at that particular point of time (refer to section 3.2.5).

Therefore, thermal characterisation of *Bacillus* AroA LCA revealed that the ancestral AroA enzyme has considerable thermostability. It is however difficult to precisely ascertain the amino acids or structural characteristics of *Bacillus* AroA LCA that impart such a thermophilic nature.

Lastly, the behaviour of *Bacillus* AroA LCA *in vivo* in comparison to contemporary AroA was investigated.

4.2.7 *In vivo* characterisation

The *in vitro* parameters of *B. subtilis* AroA and *Bacillus* AroA LCA determined in section 4.2.4.2, were comparable to each other with two notable characteristics being: lower affinity of *B. subtilis* AroA towards S3P and AroA LCA exhibiting twice the catalytic efficiency of *B. subtilis* AroA. The growth rates of these enzymes were used to investigate the effect of the enzyme's *in vitro* parameters on their fitness *in vivo*.

E. coli, with its AroA gene knocked out, fails to grow on M9 minimal media, but may be complemented with the help of a plasmid containing the *aroA* gene (Padgette et al. 1987; Jin et al. 2007). Therefore, *in vivo* characterisation of AroA enzymes was carried out using the *E. coli* Δ*aroA* from the Keio collection. This system was used as a preliminary way of investigating the effect of ancestral AroA and contemporary AroA enzymes on growth rate. The parent *E. coli* cells (strain BW25141) comprising of intact *aroA* gene, was used as a positive control and Δ*aroA* complemented with empty pPROEX HTb plasmid was used as a negative control. The AroA fitness study was performed by complementing *ΔaroA* with *B. subtilis* AroA as well as *Bacillus* AroA LCA. The relative growth rates of complemented cells was determined as explained in section 2.4, by growing the cells in M9 minimal media such that the growth of the cells results only from the activity of the complemented genes. *B. subtilis* AroA successfully rescued the Δ*aroA* cells, while *Bacillus* AroA LCA did not (Figure 4.14). The decreased fitness of *Bacillus* AroA LCA *in vivo* possibly explains the loss of this enzyme over evolutionary time.

Although, the fitness cost caused by AroA LCA needs to be further confirmed with the help of growth experiments and competition assays, *Bacillus* AroA LCA may need to be on the *Bacillus* chromosome, under the control of *Bacillus* machinery to be able to rescue the cells. Also, the k_{cat} of *Bacillus* AroA LCA at

37 °C (3.0±0.1 sec^{-1}) is three times lower than *B. subtilis* AroA k_{cat} (9.0±0.8 sec^{-1}), which may not have been sufficient activity for the rescue of the Δ*aroA* cells. Furthermore, if evolutionary changes occurred to all the enzymes in the shikimate pathway to adapt to its present evolutionary state, then the insertion of an ancestral AroA enzyme within the modern shikimate pathway may result in uncoupling, and thereby preventing AroA LCA from rescuing the cells. However, further work needs to be done to explore the *in vivo* fitness cost imposed by AroA LCA.

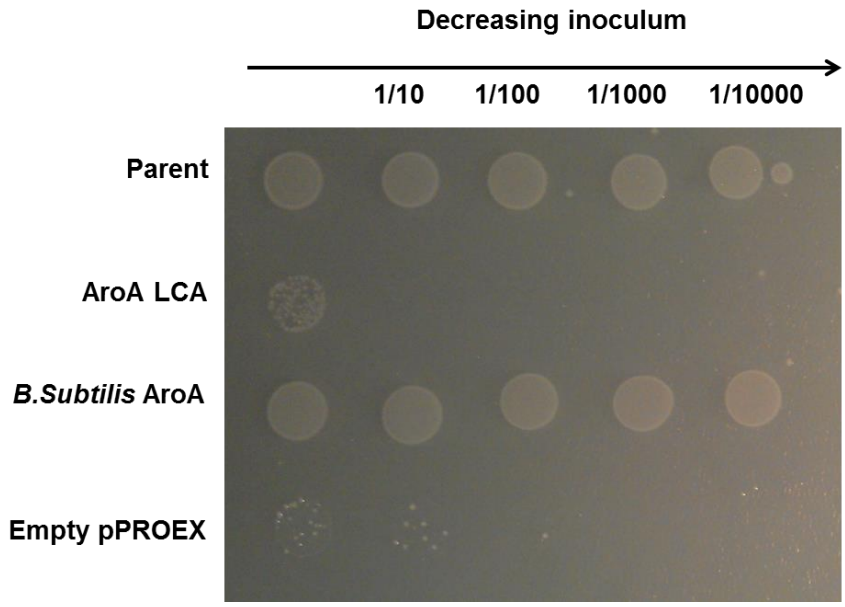


Figure 4.14: Relative growth rate of Keio collection knockout cells complemented with contemporary and ancestral *aroA* genes on M9 agar.

To summarize, the reconstructed AroA LCA from *Bacillus* genus, exhibits biologically realistic kinetic parameters comparable to the contemporary enzyme kinetics. AroA LCA was found to be moderately thermophilic in nature and demonstrated more sensitivity towards inhibitor glyphosate in comparison to the glyphosate tolerant class II *B. subtilis* AroA. Although, AroA LCA has twice the catalytic efficiency of *B. subtilis* AroA, unlike the contemporary AroA, AroA LCA was not able to rescue Δ*aroA E. coli* cells. However, further growth experiments may need to be performed in order to study the fitness of these enzymes more thoroughly.

The evolution of AroA and MurA enzymes was further explored by their structural characterisation as described in chapter 5.

Chapter 5

Structures of *Bacillus* MurA LCA and *Bacillus* AroA LCA

5.1 Introduction

The second approach used to study the evolution of the enolpyruvyl transferase enzymes MurA and AroA was X-ray crystallography. X-ray crystallography relies on the protein to crystallise into a well ordered lattice to determine its 3-dimensional structure. This technique was used to determine the structures of *Bacillus* MurA LCA and *Bacillus* AroA LCA for comparison to their respective contemporary structures. Such a detailed comparison between the ancestral and contemporary enzymes would allow us to identify any structural changes that happened over the last 950 myr and whether or not those differences are significant enough to affect the enzyme's functional characteristics.

5.2 Results and Discussion

5.2.1 MurA

5.2.1.1 Crystallisation of Bacillus MurA LCA

A crystallisation screen of 384 conditions (Hampton Research) was laid down with *Bacillus* MurA LCA as explained in section 2.5.1. Out of these 384 crystallization conditions, many conditions proved to be suitable for crystal formation, with nine conditions producing crystals overnight and a further 16 conditions with crystals formed at the end of one month. The conditions with the best crystals were advanced through hanging drop crystallisation screens (section 2.5.2.1). Besides the modifications to the original conditions, fine screens were also carried out in the presence of additives. The final condition from which crystals gave good diffraction consisted of 0.2 M ammonium acetate, 0.1 M Tris, pH 8.5, 25% PEG 3,350 along with the addition of 0.2 M glycine (in the mother liquor) as an additive and 1 mM UDP-*N*-acetylglucosamine (UNAG) (final concentration) in the hanging drop. This condition led to the formation of needle like crystals as shown in Figure 5.1.

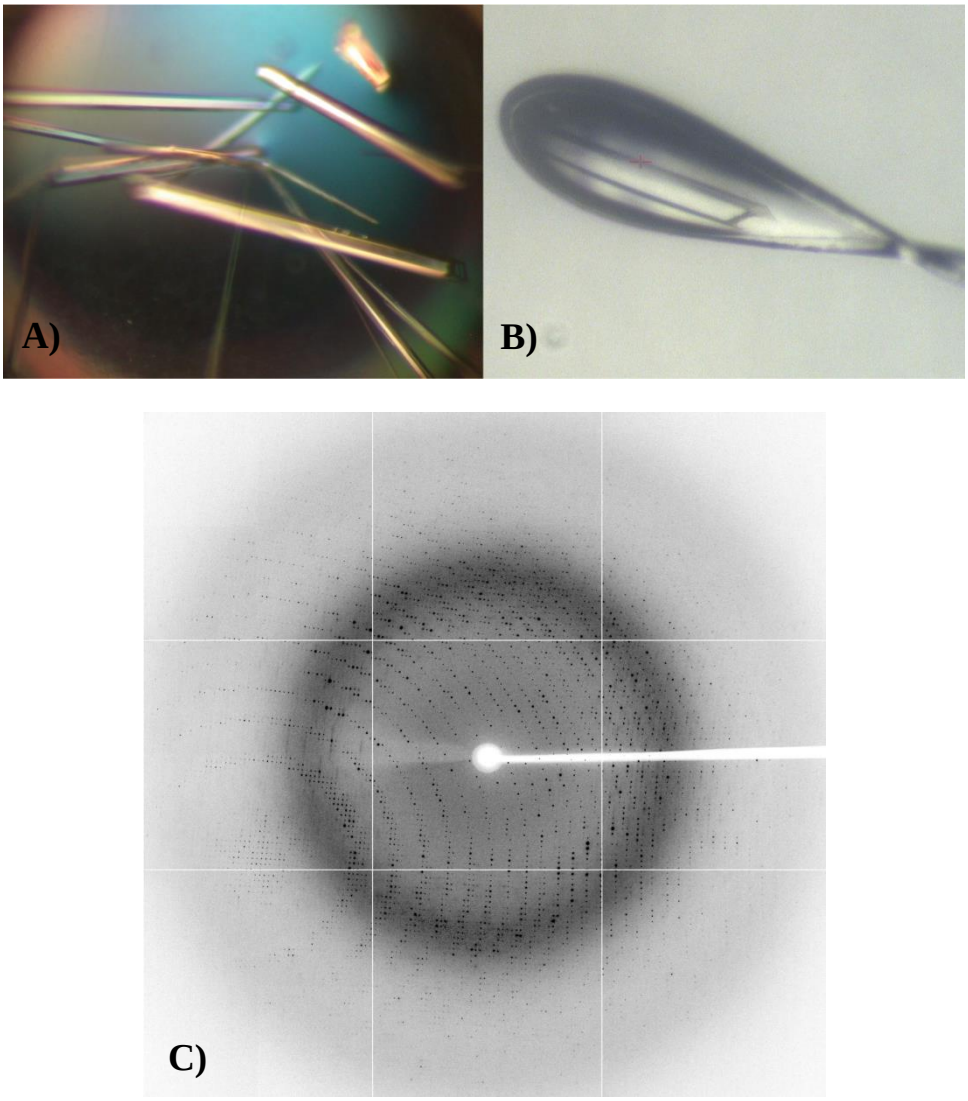


Figure 5.1: *Bacillus* MurA LCA crystals and their diffraction pattern
A) The crystals grown in a drop consisting of mother liquor with 0.2 M ammonium acetate, 0.1 M Tris, pH 8.5, 25% PEG 3,350 along with glycine and UNAG as additives. B) The crystal from which diffraction data was collected. C) X-ray diffraction pattern for *Bacillus* MurA LCA to 2.11 Å resolution.

5.2.1.2 X-ray Data Collection

The crystals were sent to the Australian Synchrotron for data collection on the MX2 beamline. Data was collected from different spots of the crystal needle. The best data was collected to a resolution of 2.11 Å with detector distance of 300 mm and 0.5° oscillation angles for a total of 200 frames.

5.2.1.3 Data Processing

The data collected was indexed using XDS and scaled using the program Aimless (Evans & Murshudov 2013). The space group was found to be $P2_12_12_1$ with data

at a resolution of 2.11 Å. 'Truncate' (French & Wilson 1978) from the CCP4 suite (Winn et al. 2011) was used to ensure unique data and to generate a FreeR set for 5% of the data. Matthews coefficient analysis (Matthews 1968) predicted the presence of four molecules in the asymmetric unit. The data collection statistics are given in Table 5.1.

Table 5.1: Data collection statistics for *Bacillus* MurA LCA

Data Statistics	**Overall**	**Outer Shell**
Space group	$P2_12_12_1$	
Wavelength (Å)	0.95370	
Cell dimensions		
a b c (Å)	96.06 / 122.85 / 153.53	
α β γ (°)	90.00 / 90.00 / 90.00	
Mosaicity	0.13	
Monomers in asymmetric unit	4	
Resolution range (Å)	48.03-2.11	2.14-2.11
Number of observations	431273	20057
Number of unique reflections	105421	5020
R-merge	0.067	0.636
Mean I/σI	11.8	2.1
Completeness %	99.8	97.2
Multiplicity	4.1	4.0

The structure of *Bacillus* MurA LCA was solved using molecular replacement (using Phaser (McCoy et al. 2007) in CCP4 (Winn et al. 2011)) with the structure of *Bacillus anthracis* MurA (PDB code 3SG1) as the model. *Bacillus* MurA LCA and *B. anthracis* MurA share 82% sequence identity. The *Bacillus* MurA LCA sequence was then automatically built into the electron density map using Buccaneer (Cowtan 2006) in CCP4. The structure was refined multiple times using refmac5 (Murshudov et al. 2011) in CCP4, imposing non-crystallographic symmetry (NCS) restraints, and manual model building using COOT (Emsley & Cowtan 2004). The amino acid residues from 356-363 and 415-418 from all the four chains of *Bacillus* MurA LCA structure were not properly modelled by

Buccaneer and hence were built manually in COOT. Each domain of a *Bacillus* MurA LCA monomer was defined as an individual TLS group for the final round of refinement

The *Bacillus* MurA LCA structure was co-crystallised with the substrate UNAG. UNAG is present in the structure, bound to the active site of the enzyme along with a phosphate molecule. The refinement statistics are given in Table 5.2.

Table 5.2: Refinement and model statistics for *Bacillus* MurA LCA

Refinement	
R-factor (%)	16.5
R_{free} (%)	21.3
Total number of atoms	13603
Total number of protein atoms	12549
Chain A	3133
Chain B	3133
Chain C	3150
Chain D	3133
Total number of ligand atoms	176
Total number of solvent atoms	878
RMSD	
Bond lengths (Å)	0.0153
Bond angles (°)	1.72
Average B factors (Å^2)	
Chain A	38.1
Chain B	43.2
Chain C	37.09
Chain D	35.7
Ramachandran analysis	
Preferred regions	97.8 % (1626 residues)
Allowed regions	1.9 % (31 residues)
Outliers	0.3 % (5 residues)

5.2.1.4 Overall Structure

Bacillus MurA LCA enzyme consists of 436 amino acids plus an additional 27 amino acids including the N-terminal His-tag. There was no electron density for the N-terminal His-tag nor for the first amino acid, nor the last 18 amino acids of the *Bacillus* MurA LCA sequence. There are four chains in the asymmetric unit of the *Bacillus* MurA LCA crystal structure.

The *Bacillus* MurA LCA monomer consists of two globular domains, with two individual strands connecting the two domains. Domain I consists of residues 23-228 and domain II, containing the N- and C-terminal ends, consists of residues 2-17 and 234-418. The linker region links the two domains by two strands of five amino acids each (residues 18-22 and 229-233). The two domains are also connected *via* a salt bridge between Arg234 and Glu192 (Costantini et al. 2008).

Each domain is made up of three repetitive subunits with a total of 6 repetitive subunits within the monomeric structure (Figure 5.2). Each subunit is typically made of four β-strands and two α-helices with the exception of subunit II which has long loop followed by a one turn helix (α3) instead of a β-strands and subunit VI which has its second helix replaced by two 2-turn helices (α13 and α14) (Figure 5.3). The irregularity of subunits II and VI of *Bacillus* MurA LCA differs from the previously solved structures of MurA (Schönbrunn et al. 1996; Skarzynski et al. 1996; Eschenburg & Schönbrunn 2000). Similar to previously solved MurA structures, the core of each domain consist of helices surrounded by three clusters of β sheets to give the inside out α/β barrel which is typical of enolpyruvyl transferases (Figure 5.2) (Schönbrunn et al. 1996; Skarzynski et al. 1996; Eschenburg et al. 2003; Oberdorfer et al. 2012).

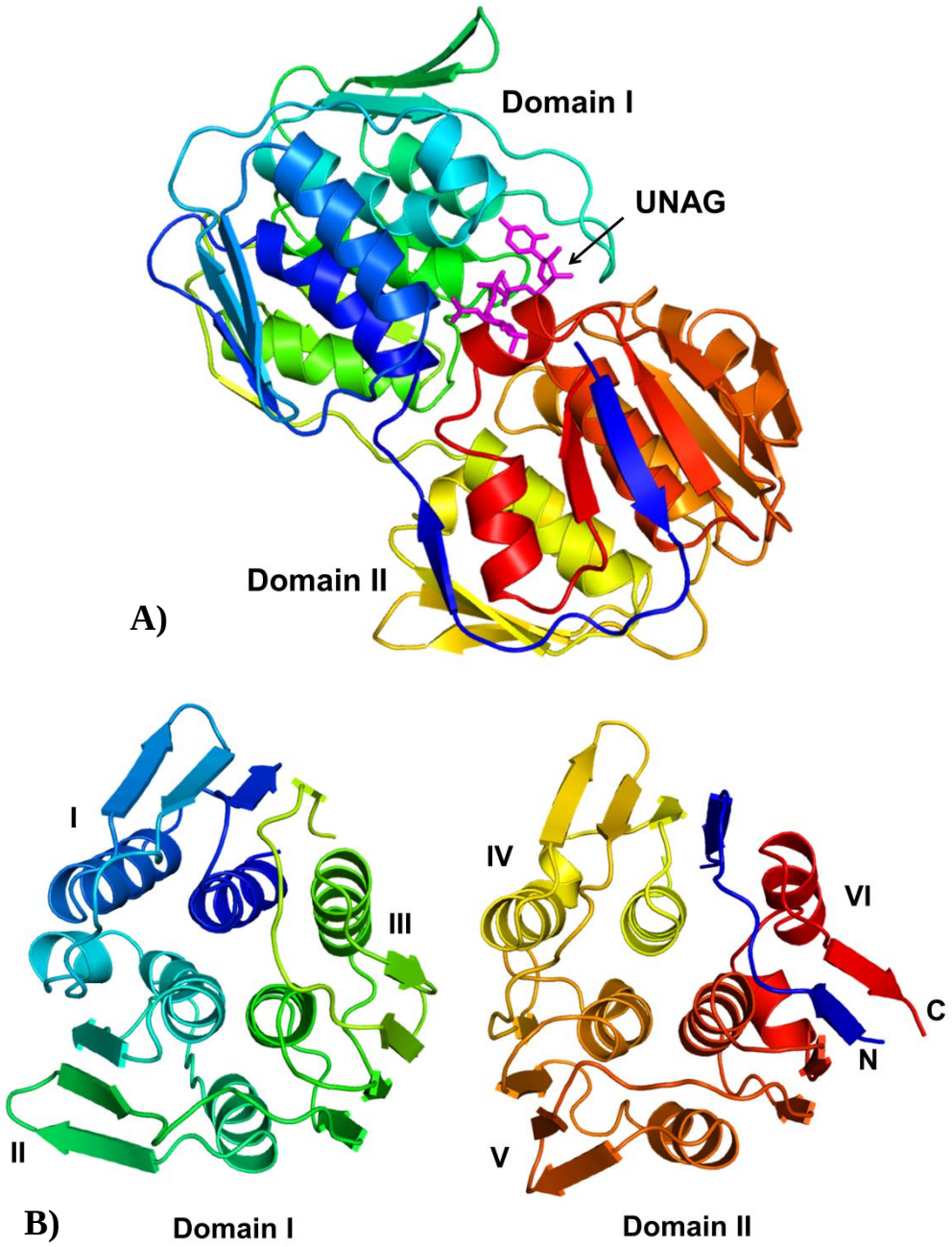


Figure 5.2: *Bacillus* MurA LCA monomer and its inside out α/β barrel conformation
A) Cartoon representation of monomeric *Bacillus* MurA LCA bound to its substrate UNAG (magenta). B) Cartoon representation of the two domains I and II with its six subunits labelled (I-VI). The orientation is chosen to show the inside-out α/β barrel configuration of *Bacillus* MurA LCA.

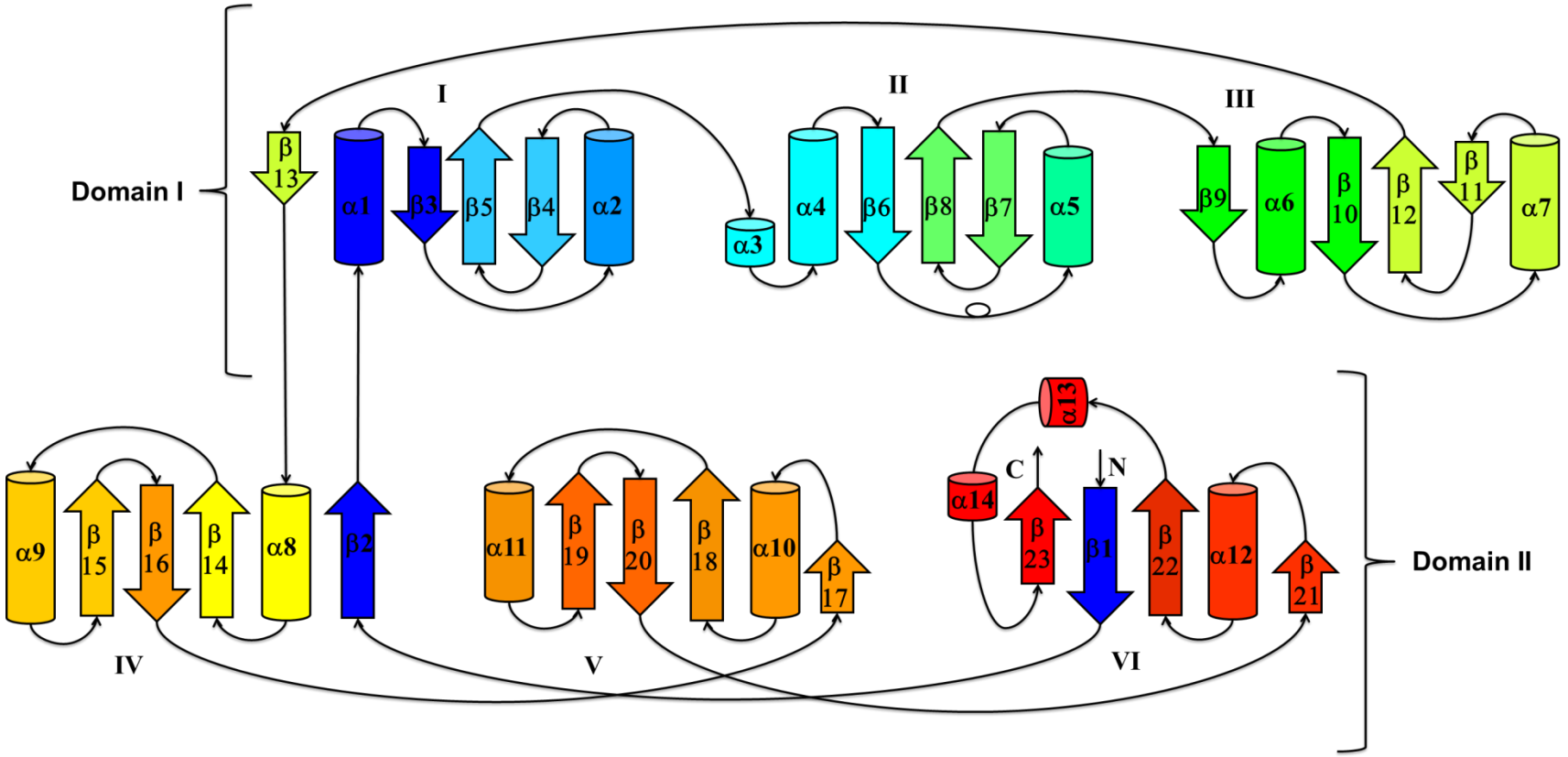


Figure 5.3: Topology sketch of *Bacillus* MurA LCA
The figure displays the two globular domains, domain I and II, each consisting of three subunits (I-VI). Each helix and beta strand is labelled sequentially. The twisted loop in subunit II represents the long loop Ala112-Ile124. The components are coloured corresponding to the colouration used in Figure 5.2.

The size exclusion chromatography purification of *Bacillus* MurA LCA protein showed two peaks, with the dominant peak corresponding to a dimer (Figure 3.6 and section 3.2.3.1) used for crystallography. The protein crystallised with four molecules in the asymmetric unit. Chain C and chain D form a dimer with an interface interaction of 6.6% of the total surface area (1056.3 $Å^2$) (Figure 5.4) (PDBePISA (Krissinel & Henrick 2007)). The remaining two molecules (chain A and chain B) require symmetry neighbours to form their respective dimers with a contact area of 6.7% (1067$Å^2$) (PDBePISA (Krissinel & Henrick 2007)). Therefore the four chains A, B, C and D form three dimers split between three unit cells.

The presence of UNAG in the crystallisation condition of *Bacillus* MurA LCA allowed the protein to be crystallised in its substrate bound form. It was observed that domain I of *Bacillus* MurA LCA shows a general higher B factor than that of domain II, indicating that domain I is less ordered than domain II. It was also noted that chain B has a higher average B factor than the remaining three chains (Table 5.2).

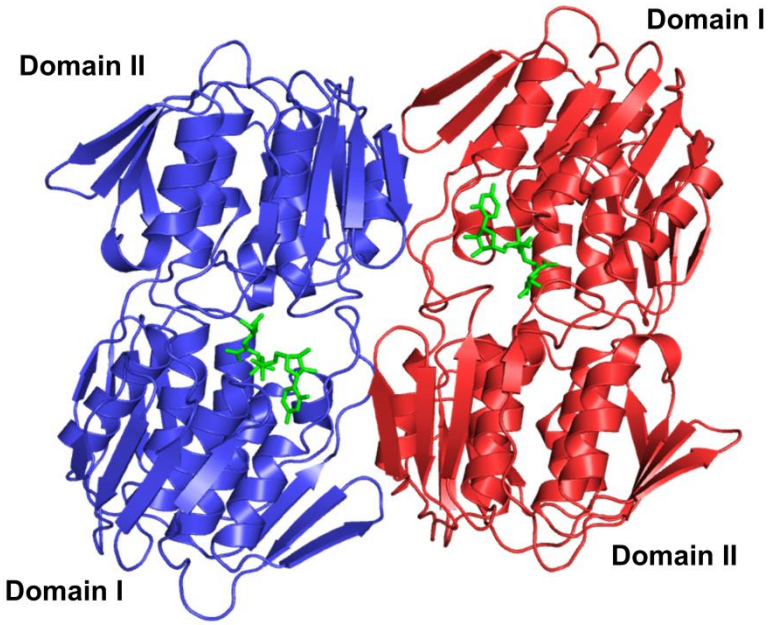


Figure 5.4: *Bacillus* MurA LCA dimer bound to UNAG
Cartoon representation of *Bacillus* MurA LCA dimer bound to UNAG (green). The dimer is formed by chain C and D sharing an interface area of 1056.3Å^2 within the same unit cell.

5.2.1.5 Substrate binding site

The catalytic site of *Bacillus* MurA LCA is situated in between the two domains. A total of 10 bonds are formed between the substrate UNAG and amino acid residues of *Bacillus* MurA LCA Figure 5.5 (PDBsum (Laskowski 2001)). Out of these 10 residues involved in ligand interaction, only two residues are located within domain II (Asp306 and Val328). The remaining eight residues are located on domain I (Asn23, Val165, Ser164, Gln126, Asp125, Ser121, Arg122 and Gly166) with residues Ser121 and Arg122 positioned on the loop between β6 and α5 (Figure 5.3). Asn23, from *Enterobacter cloacae*, was recognised to be an essential active site residue involved in the stabilisation of the MurA transition states (Samland et al. 2001).The second substrate for MurA, phosphoenolpyruvate (PEP), is not present in the crystal structure of *Bacillus* MurA LCA. According to *E. cloacae* and *E.coli* MurA numbering, the strictly conserved residue Asp305 is essential for the addition-elimination reaction of the enzyme. Residues Arg331, Arg371,

Lys22 and Arg120 are responsible for PEP binding and the transfer of the enolpyruvyl moiety. Other conserved residues such as Arg397 (involved in a salt bridge), contribute to the phosphate binding site and Cys115 is involved in the protonation of PEP during the reaction (Skarzynski et al. 1996; Eschenburg et al. 2003). All the afore mentioned catalytically important active site residues correspond to residues Asp306, Arg332, Arg372, Lys22, Arg122, Arg398, Cys117 and Asn23 respectively of *Bacillus* MurA LCA and are illustrated in Figure 5.5.

Residues Cys117, Ser121 and Arg122 are present on the long loop between β6 and α5 (Ala112-Ile124). The section of this loop from Pro114-Pro123 may correspond to the previously identified loop from *E.coli* and *Enterobacter cloacae* (Pro112 to Pro121) which is believed to have a role in stabilising the enzyme in its unliganded form and is known to undergo conformational changes upon substrate binding (Schönbrunn et al. 2000). The conformational change after substrate binding brings the domains closer together to be in a 'closed state'. Skarzynski et al. (1996) describes this loop as a 'lid' over the catalytic pocket when the enzyme is in its closed state. The loop of *Bacillus* MurA LCA exhibits a similar conformation by forming a lid over the catalytic pocket, unlike the orientation of the loop seen in unliganded MurA (PDB 1NAW) (Figure 5.6).

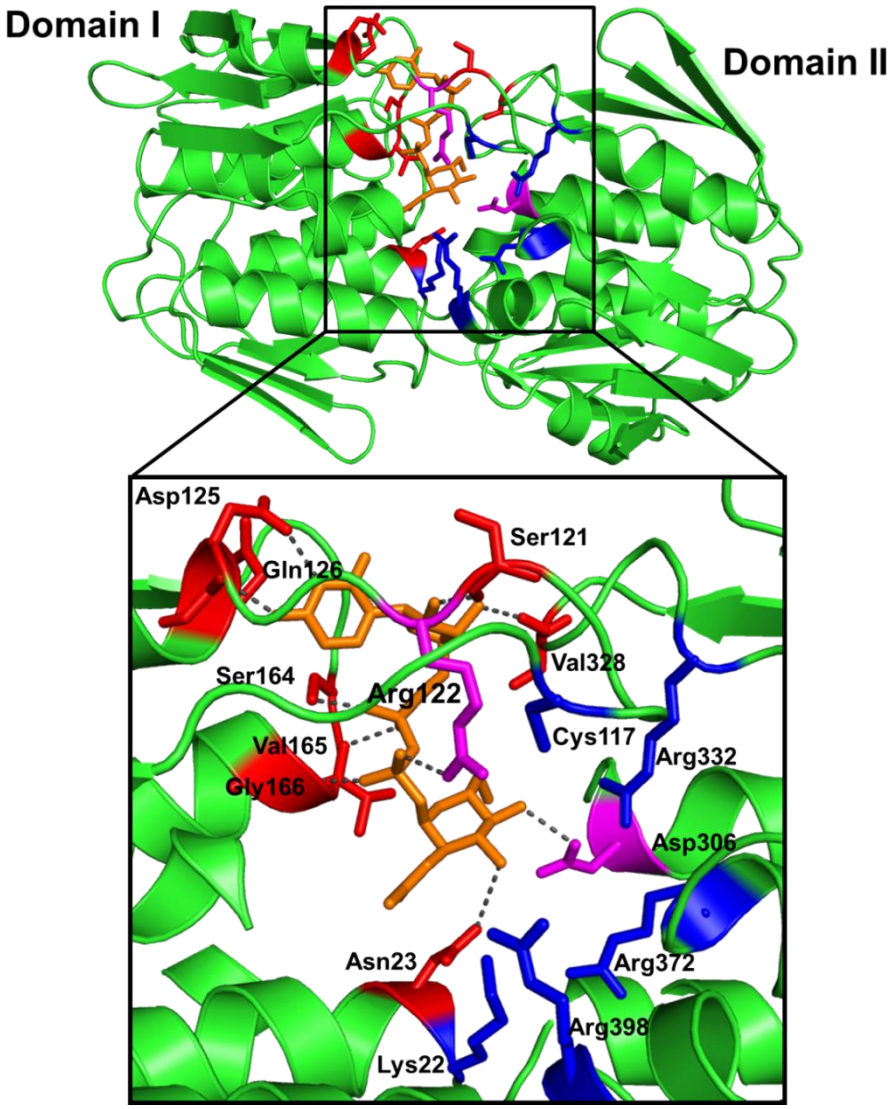


Figure 5.5: Cartoon representation of *Bacillus* MurA LCA showing the active site residues

The residues interacting with UNAG are shown in red, residues interacting with PEP are coloured dark blue and the residues involved in both UNAG and PEP interaction are coloured in magenta. UNAG (orange) is bound to the LCA structure and the bonds formed between UNAG and residues are shown in dashes. The labels correspond to *Bacillus* MurA LCA structure numbering.

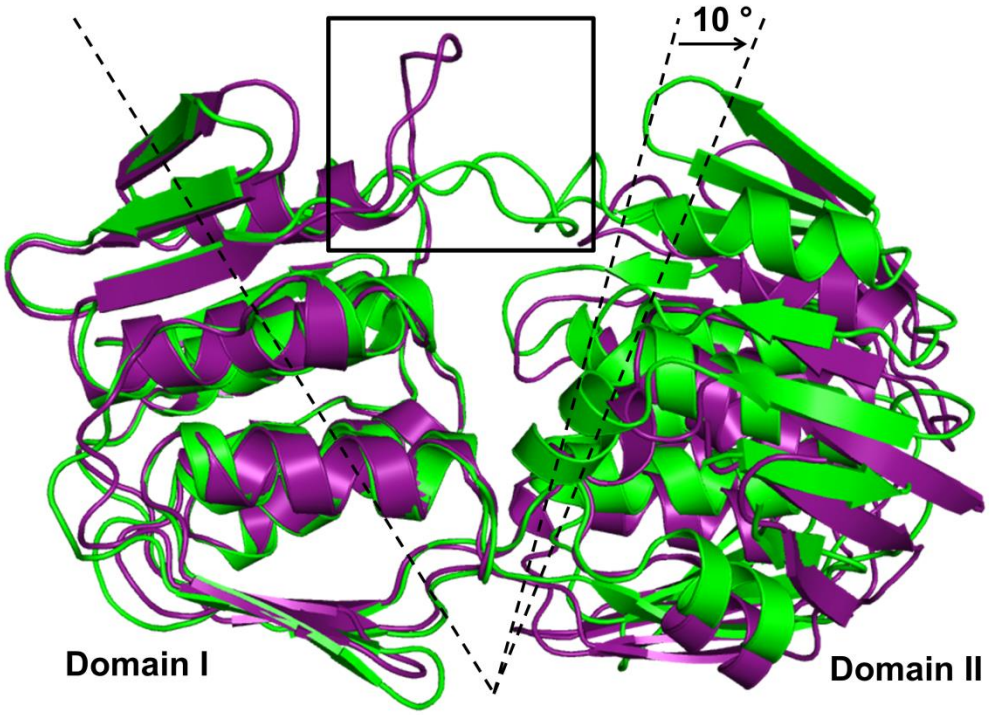


Figure 5.6: 'Lid'-like loop in closed and open conformations
Bacillus MurA LCA structure (green) is in a closed conformation as it is bound to substrate UNAG whereas *E.cloacae* MurA (PDB 1NAW) (purple) has no substrate bound and hence is in an open conformation. The box shows the conformation of the 'lid'-like loop in the open and closed states. Domain I of both structures is overlaid and shows the unbound structure of *E.cloacae* MurA (purple) to be more open than *Bacillus* MurA LCA by 10°.

5.2.1.6 Inhibitor binding in the active site

In spite of sharing mechanistic and structural similarities, contemporary MurA and AroA are exclusively inhibited by fosfomycin and glyphosate respectively. However, it is possible that the ancestral enzymes MurA and AroA are more similar, which may allow inhibition by glyphosate or fosfomycin interchangeably. Eschenburg et al (2003) determined that the reason why glyphosate could not inhibit MurA was due to distance clashes of glyphosate to the residues Arg120, Arg371 and UNAG as well as the lack of a specific residue that could hold glyphosate in place *via* hydrogen bonds. An alignment of contemporary *E. cloacae* C115D MurA liganded with UNAG (PDB 3V4T) with *Bacillus* MurA LCA shows that the orientation of corresponding residues

for Arg120, Arg371 and substrate UNAG remains approximately the same as in the ancestral enzyme (Figure 5.7). This suggests that the structural changes necessary for glyphosate inhibition of MurA are not present in *Bacillus* MurA LCA, in agreement with the IC50 results obtained in section 3.2.4.2.

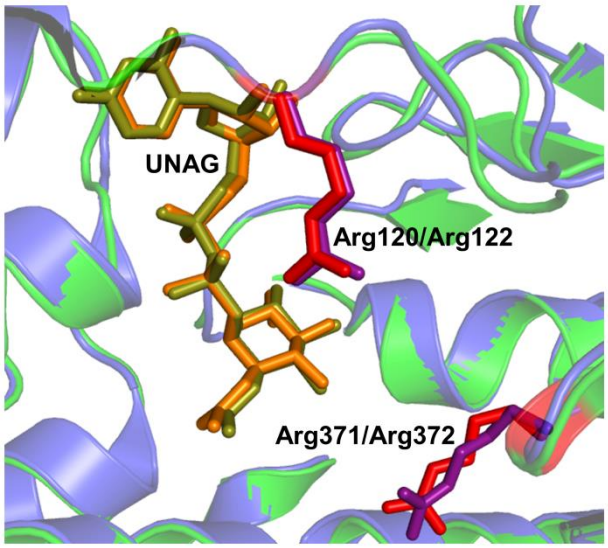


Figure 5.7: Orientation of residues Arg120/Arg122 and Arg371/Arg372 and UNAG in *Enterobacter cloacae* MurA and *Bacillus* MurA LCA respectively
Residues Arg120, Arg371 (purple) and UNAG (olive green) from *Enterobacter cloacae* MurA (PDB 3V4T) (blue) show similar orientation and positioning as residues Arg122, Arg372 (red) and UNAG (orange) from *Bacillus* MurA LCA (green).

5.2.1.7 Comparison to contemporary structures

A structural comparison of *Bacillus* MurA LCA to the entire PDB archive using PDBeFOLD (Krissinel & Henrick 2004) found *B. anthracis* (PDB 3SG1) to be the closest structural homologue with a sequence identity of 82%. Comparing the structures of *Bacillus* MurA LCA and *B. anthracis* MurA would allow us to identify the main differences between contemporary and ancestral MurA enzymes within the same genus. The majority of the remaining structures from the PDBeFOLD analysis share around 46% to 51% sequence identity to *Bacillus* MurA LCA. To properly analyse the differences between ancestral MurA against contemporary MurA, a comparison of the enzymes in their bound form to UNAG is also necessary as the enzyme adopts a different conformation

upon substrate binding. Table 5.3 gives structural alignment parameters for a selection of the structures discussed in this chapter (3SG1 and 1NAW without ligand; 3V4T, 2RL1 and 1UAE with UNAG bound).

Table 5.3: PDBeFOLD structural alignment of PDB structures to *Bacillus* MurA LCA

Structure	RMSD	% seq[1]	Q-score	P-score	Z-score	% sse[2]	PDB code	Reference
B. anthracis	0.60	82	0.96	64.0	24.3	95	3SG1	To be published
E.cloacae	0.89	51	0.89	61.2	23.8	95	3V4T	(Zhu et al. 2012)
E.coli	1.02	50	0.86	64.2	24.2	97	1UAE	(Skarzynski et al. 1996)
H. influenzae	1.06	49	0.86	54.4	22.3	95	2RL1	(Yoon et al. 2008)
E.cloacae	1.77	48	0.67	38.6	18.9	97	1NAW	(Schönbrunn et al. 1996)

[1]%seq: % of sequence identity
[2]%sse: % of secondary structure of *Bacillus* MurA LCA identified in target protein

The PDBeFOLD structural alignment of *Bacillus* MurA LCA to various structures in the PDB archive shows an overall high structural similarity between ancestral MurA structure to various contemporary MurA enzymes.

The superposition of *Bacillus* MurA LCA with unliganded *B. anthracis* (PDB 3SG1) and with liganded *E. cloacae* (PDB 3V4T) and liganded *Haemophilus influenza* (PDB 2RL1) are shown in Figure 5.8. *Bacillus* MurA LCA differs from the three structures in respect to individual α-helix and β-strand length and relative orientation. The length and conformation of loops connecting the secondary structures also differ between the structures. Using the topology diagram of *Bacillus* MurA LCA (Figure 5.3) as comparative tool for

highlighting structural differences, it was observed that *B. anthracis* MurA, *H. influenza* MurA and *E. cloacae* MurA have an additional short β-strand in front of α3 and have one single long α-helix instead of helices α3 and α4. The presence of α13 in subunit VI of *Bacillus* MurA LCA structure forms the motif of $\beta\alpha\beta\alpha^{*}\alpha\beta\beta$. Even though the current literature does not describe this deviation from the commonly identified repetitive motif βαβαββ, the additional short α helix in subunit VI is present in all the contemporary MurA structures. The positioning and orientation of UNAG is also very similar. Most of the noticeable differences of *Bacillus* MurA LCA and the contemporary structures are located on the outer surface of the protein, mostly where the B factors are high, leaving the active site very similar to that of the contemporary enzymes.

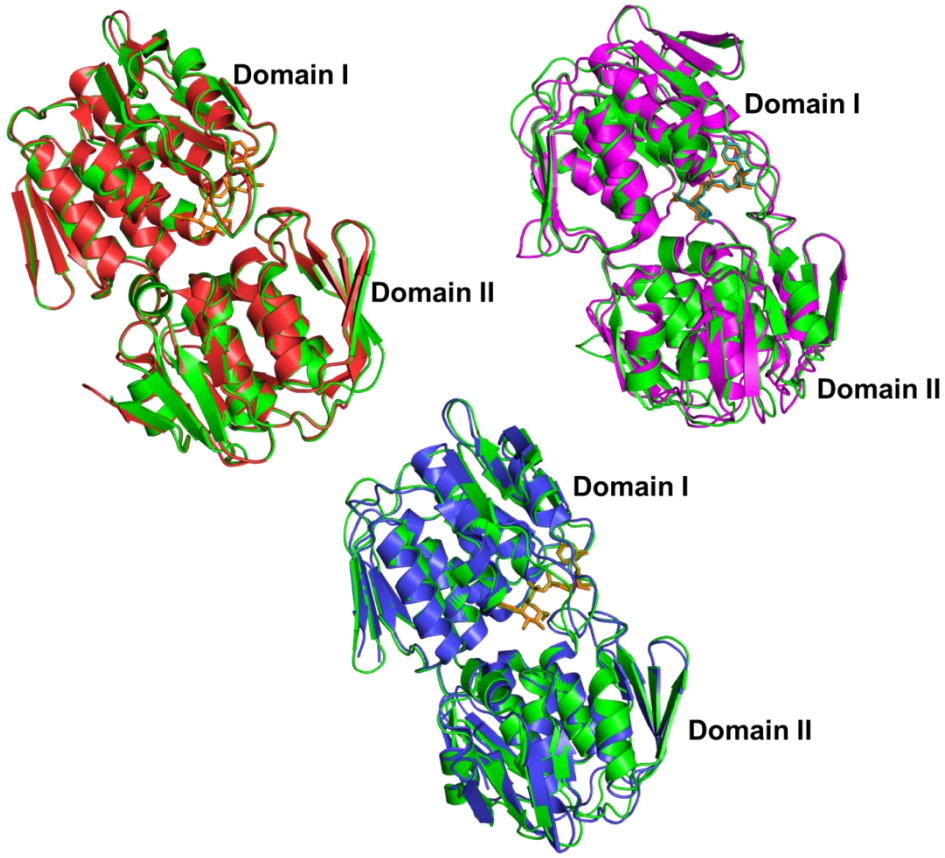


Figure 5.8: Secondary structure matching (SSM) overlay of *Bacillus* MurA LCA bound to UNAG with contemporary structures showing the overall structural similarity
Bacillus MurA LCA (green) bound to UNAG (orange) is superimposed onto the structures of A) *Bacillus anthracis* MurA (red) (PDB 3SG1) B) *Haemophilus influenza* MurA (magenta) (PDB 2RL1) bound to UNAG (teal) and C) *Enterobacter cloacae* MurA (blue) (PDB 3V4T) bound to UNAG (olive green).

The ancestral MurA enzyme was further compared to contemporary *B. anthracis* MurA enzyme by performing a detailed analysis of the amino acid differences between them. All the amino acids that differ between *Bacillus* MurA LCA and *B. anthracis* MurA, which are present in the LCA structure, are shown in Figure 5.9. It was observed that majority of the amino acid residues identified were present on the surface of the protein. The closest residue to the active site was found to be Leu97 with a distance of 4.7 Å from substrate UNAG.

Bacillus MurA LCA differs significantly from *B. anthracis* MurA in its thermostability. There is a large literature that depicts the contribution of protein amino acid composition to thermophily. The most commonly known factors responsible for thermophily include: a higher core hydrophobicity/number of hydrophobic residues, increase in ion pairs on the protein surface/increase in charged residues (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001; Zhou et al. 2008; Taylor & Vaisman 2010), decrease in polar residues (Fukuchi & Nishikawa 2001) and decrease in Ser content (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001). An increase in aromatic residues, higher Pro, Arg and Glu content and lower frequency of Met has also been reported to be associated with increased thermophily (Zhou et al. 2008). The amino acid composition of moderately thermophilic *Bacillus* MurA LCA and mesophilic *B. anthracis* MurA was compared to identify any such trends that could explain the increased thermostability within MurA LCA (Table 5.4). This was achieved with the help of EMBOSS (Rice et al. 2000). It was observed that there is an increase of seven non polar residues and a decrease of five polar residues in MurA LCA in comparison to *B. anthracis* MurA. The decrease of polar residues Ser, Thr, Gln and Asn which is asscociated with thermophily (Fukuchi & Nishikawa 2001; Zhou et al. 2008), is also observed within MurA LCA except for Thr.

The amino acid differences between *Bacillus* MurA LCA and *B. anthracis* MurA are predominantly located on the surface of the protein (Figure 5.9). The amino acid composition on the protein surface adapts to the physico-chemical environment of the protein (Fukuchi & Nishikawa 2001). Therefore, the difference in amino acid composition on the surface of LCA protein in comparison to *B. anthracis* MurA may have resulted from such an adaptation to thermophilic and mesophilic environments respectively.

The increase in charged residues has been reported multiple times as a characteristic of increase in thermophily/thermostability (Zhou et al. 2008; Taylor & Vaisman 2010). However, such an increase in charged residues has

not been observed in *Bacillus* MurA LCA as both thermophilic LCA and mesophilic *B. anthracis* MurA have an equal number of charged residues (Table 5.4). On the other hand, there is a noticeable difference in the Ser content between the ancestral and contemporary MurA. An apparent decrease in Ser content is associated with increase in thermophily (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001) and this trend was observed within the moderately thermophilic *Bacillus* MurA LCA (Table 5.4). With regards to other single amino acid trends associated with thermophily, no particular trend was observed with Met content and a minor increase of 1-4 residues of Pro, Arg and Glu was observed in *Bacillus* MurA LCA in comparison to *B. anthracis* MurA.

Therefore, the increased thermophily observed in *Bacillus* MurA LCA might have resulted from the decreased serine content, increased hydrophobic residues, decreased polar residues and/or due to the amino acid differences on the protein surface. A combination of these factors or any of these factors on its own may have been responsible for the increase in thermophily as there is no specific mechanism to define thermostability and often even subtle redistributions of amino acids in the protein may influence thermostability (Zhou et al. 2008; Hobbs et al. 2012).

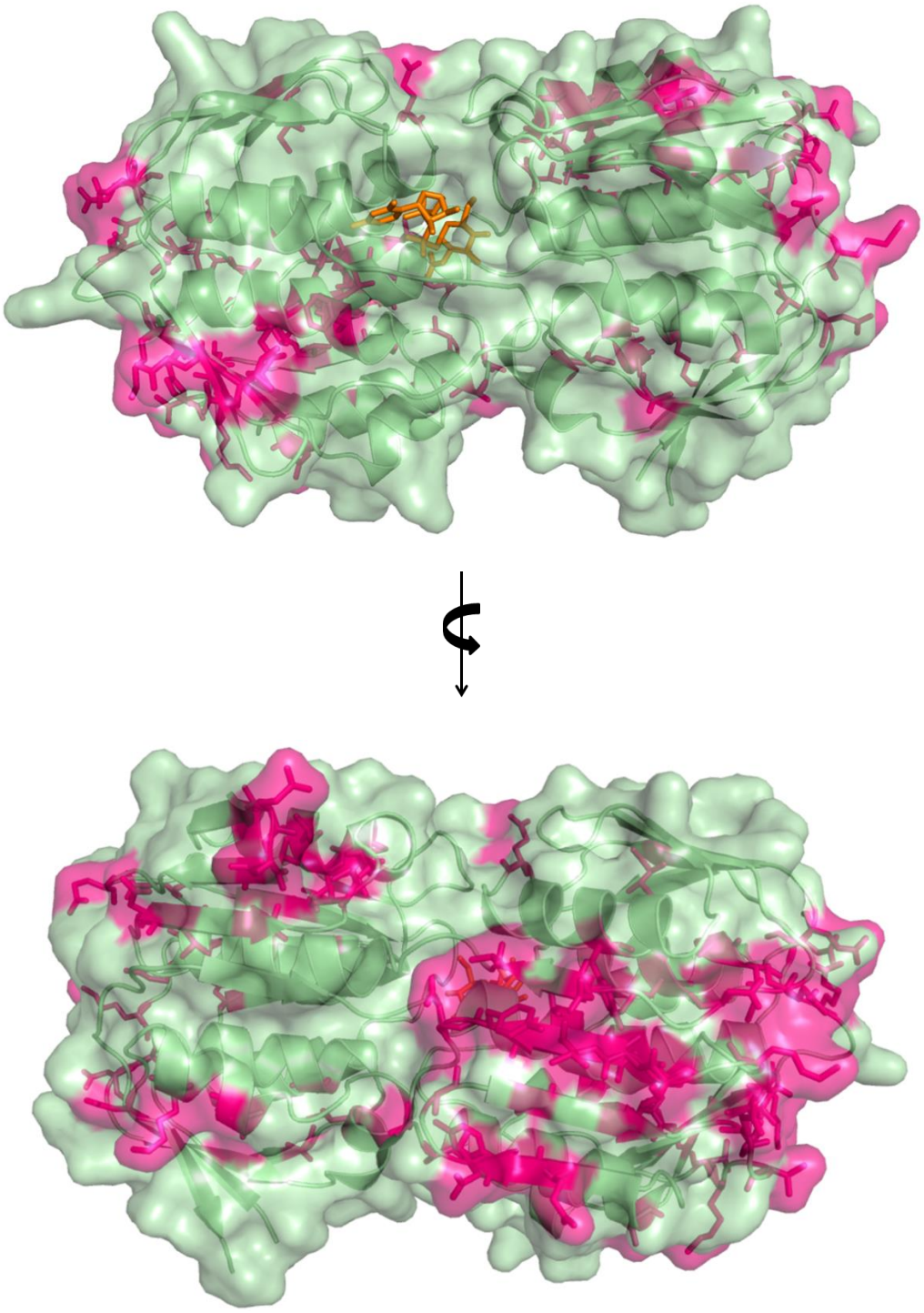


Figure 5.9: Amino acid differences between *Bacillus* MurA LCA and *Bacillus anthracis* MurA
The cartoon representation of *Bacillus* MurA LCA structure (green) with a transparent surface is shown here. The residues that differ between MurA LCA and *B. anthracis* MurA are highlighted on the MurA LCA structure (magenta). The surface exposed residues show magenta colour on the surface. UNAG (orange) is bound to *Bacillus* MurA LCA.

Table 5.4: Amino acid composition of *Bacillus* MurA LCA and *Bacillus anthracis* MurA

Amino acid group[1]	Amino acid	*Bacillus* MurA LCA	*Bacillus anthracis* MurA
Non polar	G	42	40
	A	56	50
	V	36	40
	L	34	36
	I	37	33
	P	17	15
	M	17	17
Total		**239**	**231**
Polar - Uncharged	N	21	22
	Q	7	9
	C	1	1
	S	9	16
	T	25	21
Total		**63**	**69**
Polar – Charged	R	24	23
	K	24	25
	H	7	9
	D	16	18
	E	44	40
Total		**115**	**115**
Aromatic	F (NP)[2]	11	12
	W (NP)[2]	0	0
	Y (P)[3]	8	7
Total		**19**	**19**
		436 residues in total	**434 residues in total**

[1] The amino acid grouping correspond to classification by Zhou et al. (2008)
[2] NP = Non polar residue
[3] P = Polar residue

5.2.2 AroA

5.2.2.1 Crystallisation of Bacillus AroA LCA

Crystallisation screens were laid down as explained in section 2.5.1 with *Bacillus* AroA LCA protein. Out of the 384 conditions tested (Hampton Research), various conditions resulted in the formation of microcrystals and one condition, 'Crystal 14', formed crystal needles. Further crystallisation screen trials (section 2.5.2) were performed on these conditions to improve crystal formation. However, none of the microcrystals improved and fine screens on 'Crystal 14' only resulted in very thin needles originating from a single point. *Bacillus* AroA LCA treated with trypsin (section 2.5.2.3), was used to grow a needle crystal in robot screen condition. 'Crystal 14' containing 0.2 M calcium chloride dihydrate, 0.1 M HEPES sodium pH 7.5 and 28% v/v PEG 400, was taken to the synchrotron and gave diffraction data at 3Å resolution.

Shikimate has been reported to have been used as an alternative substrate for AroA and has also been crystallised in complex with *E.coli* AroA (Priestman et al. 2005b). Therefore, attempts were also made to crystallise *Bacillus* AroA LCA in the presence of shikimate. The drop from which 'Crystal 14' crystal was taken was used as a seed stock for batch seeding (section 2.5.2.4). Trypsin treated protein laid down with 1:100 ratio of the seed stock, in the presence of a final concentration of 10 mM shikimate and mother liquor consisting of 'Crystal 14' condition, formed a crystal which diffracted to 2.03 Å (Figure 5.10).

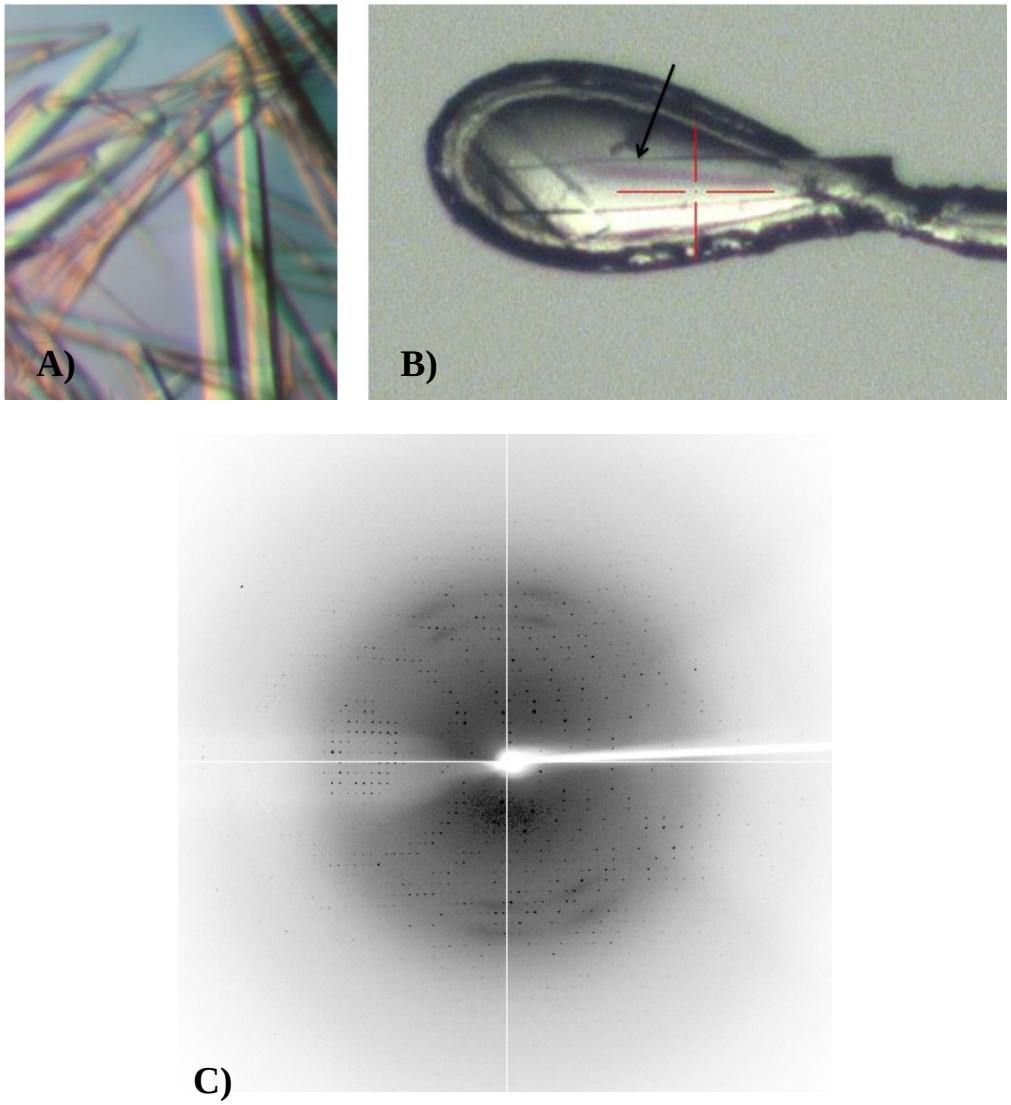


Figure 5.10: *Bacillus* AroA LCA crystals and diffraction pattern
A) Crystals grown with trypsin treated protein using batch seeding with mother liquor of 0.2 M calcium chloride dihydrate, 0.1 M HEPES sodium pH 7.5 and 28% v/v PEG 400 in presence of 10 mM final concentration of shikimate. B) Two crystals in the loop. The arrow indicates the crystal from which diffraction data was collected. C) X-ray diffraction pattern for *Bacillus* AroA LCA.

5.2.2.2 X-ray Data Collection

The MX1 beamline at the Australian synchrotron was used to collect data from *Bacillus* AroA LCA crystals. Two datasets were collected at a resolution of 3.0 Å and 2.03 Å with detector distance of 199 mm, 50% beam attenuation and 1° oscillation for 360° with one second exposure per frame. The 2.03 Å dataset was used for all further analysis.

5.2.2.3 Data processing

The data collected from the synchrotron was processed the same way as the data for *Bacillus* MurA LCA. The collected dataset was autoindexed, scaled and merged at the Australian synchrotron. The space group was found to be $P2_1$ with data at a resolution of 2.03 Å. 'Truncate' (French & Wilson 1978) was used to generate the FreeR for 5% of the data and Matthews coefficient (Matthews 1968) predicted the presence of two molecules in the asymmetric unit. The data collection statistics are given in Table 5.5.

Table 5.5: Data collection statistics for *Bacillus* AroA LCA

Data Statistics	Overall	Outer Shell
Space group	$P2_1$	
Wavelength (Å)	0.95370	
Cell dimensions		
a b c (Å)	62.83 / 82.64 / 81.22	
α β γ (°)	90.00 / 99.68 / 90.00	
Mosaicity	0.29	
Monomers in asymmetric unit	2	
Resolution range (Å)	45.43-2.03	2.09-2.03
Number of observations	399780	28517
Number of unique reflections	52652	3861
R-merge	0.133	0.895
Mean I/σI	11.1	2.0
Completeness %	99.9	99.5
Multiplicity	7.6	7.4

The structure of *Bacillus* AroA LCA was solved by molecular replacement using Phaser (McCoy et al. 2007) in CCP4 (Winn et al. 2011). The structure of *Bacillus halodurans* AroA (PDB 3RMT) with a sequence identity of 73% was used as the model for molecular replacement. Due to the high Rmerge, resolution was further truncated to 2.15 Å for all subsequent steps. The *Bacillus*

AroA LCA sequence was built into the electron density map with automated model building (Terwilliger et al. 2008) in Phenix (Adams et al. 2010). The structure was refined multiple times using Phenix refinement (Afonine et al. 2012) or refmac5 in CCP4 (Murshudov et al. 2011) with manual model building in COOT (Emsley & Cowtan 2004). Amino acid residues with no electron density were removed and the ones in poor electron density regions were deleted and built back manually, the majority of which were in the second half of chain A. Each domain of a *Bacillus* AroA LCA monomer was defined as an individual TLS group for the final refinement. The refinement statistics are given in Table 5.6.

Table 5.6: Refinement and model statistics for *Bacillus* AroA LCA

Refinement	
R-factor (%)	17.7
R_{free} (%)	23.4
Total number of atoms	6355
Total number of protein atoms	6052
Chain A	3019
Chain B	3033
Total number of solvent atoms	303
RMSD	
Bond lengths (Å)	0.0163
Bond angles (°)	1.848
Average B factors (Å^2)	
Chain A	25.846
Chain B	31.244
Ramachandran analysis	
Preferred regions (%)	96.85 % (769 residues)
Allowed regions (%)	2.52 % (20 residues)
Outliers (%)	0.63 % (5 residues)

5.2.2.4 Overall structure

Bacillus AroA LCA consists of 430 amino acids with an N-terminal His-tag. As mentioned above (section 5.2.2.1), *Bacillus* AroA LCA was trypsin treated prior to crystallisation in the presence of its alternative substrate shikimate. However, the crystallised protein does not have shikimate bound to it. There is no electron density observed for the N-terminal His-tag and for the first 11 amino acid residues of *Bacillus* AroA LCA. Amino acid residues from 337 to 347 in chain A and 338 to 346 in chain B lack a properly defined electron density and hence were deleted from the structure. This may be due to disorder in the crystal and/or the flexibility of the loop. Residues corresponding to this region are also absent in *B. halodurans* AroA (PDB 3RMT) and the corresponding region is identified as the most flexible area in structures for *Streptococcus pneumoniae* AroA (PDB 1RF5) and *Agrobacterium* CP4 AroA (PDB 2GG4).

The *Bacillus* AroA LCA monomer, similar to that of *Bacillus* MurA LCA, consists of two globular domains, domain I (residues 23-230) and domain II (residues 12-18 and 235-430) connected together by two strands (residues 19-22 and 231-234) (Figure 5.11). *Bacillus* AroA LCA, in its unliganded form, has an open configuration similar to that of *S. pneumoniae* AroA (Park et al. 2004).

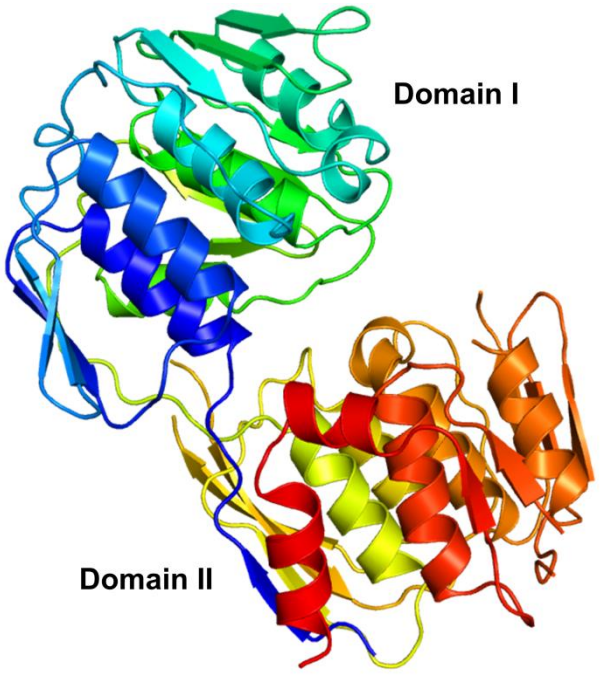


Figure 5.11: Cartoon representation of monomeric *Bacillus* AroA LCA showing unliganded open conformation
The colour spectrum from dark blue to red goes from N-terminal to C-terminal.

The main structural feature shared by AroA and MurA is the six fold repetitive motif of βαβαββ forming an inside-out α/β barrel (Stallings et al. 1991; Schönbrunn et al. 2001; Eschenburg et al. 2003; Park et al. 2004; Oberdorfer et al. 2012). It was observed that the overall inside-out α/β barrel structure is maintained by *Bacillus* AroA LCA (Figure 5.12), however there are minor variations to the previously identified repetitive βαβαββ motif due to chain differences as further explained below and topology differences as explained in section 5.2.2.7 (Figure 5.13).

The *Bacillus* AroA LCA structure consists of two monomers in the asymmetric unit, chain A and chain B, with an RMSD of 1.57 Å indicating significant differences between the two chains (potentially imposed by crystal development). The differences between the two chains lie in their secondary structure elements (their lengths and orientation), which is a result of chain B being more open than chain A by approximately an angle of 8 ° (Figure 5.12). Such differences in orientation between the chains have resulted in the

disruption of the hydrogen bonds responsible in defining a few short α helices and β strands in certain regions of chain B.

The two globular domains of *Bacillus* AroA LCA structure consists of three subdomains each (Figure 5.13). Subunits III and IV form the previously described repetitive βαβαββ motif. Subunit I in chain A also forms a βαβαββ configuration; however a β strand is missing in chain B. All the remaining subunits typically consist of three β strands and two α helices. The residues of the N-terminal region of the protein usually form a β strand in subunit VI. However, density for this region is missing; hence the presence of this β strand is unconfirmed. An additional short α helix was observed in subunits II and VI (consisting of four and six residues respectively), which corresponds to the extra 3_{10}-helices in the structure of unliganded *S. pneumoniae* AroA (Park et al. 2004).

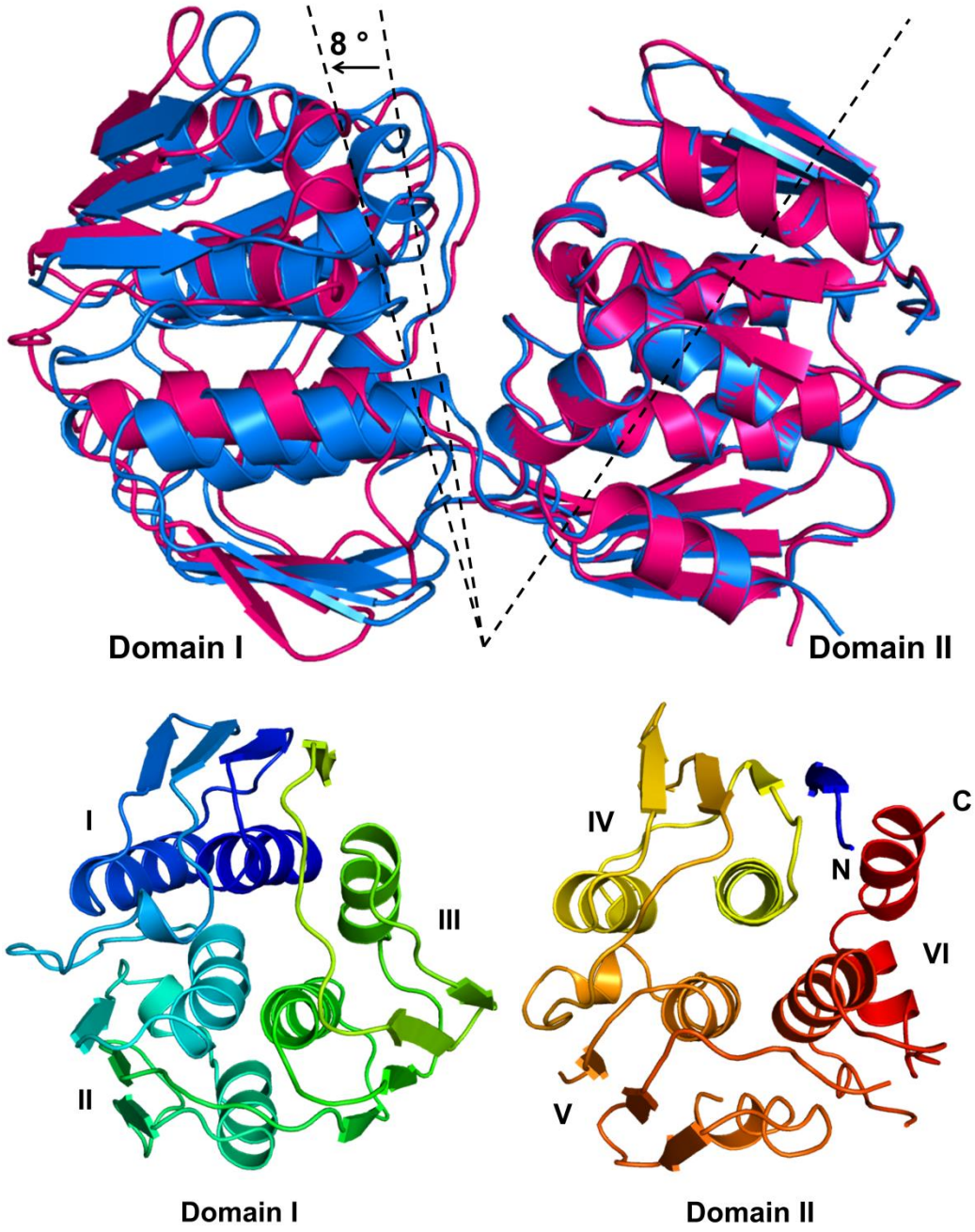


Figure 5.12: Conformational differences between chain A and chain B of *Bacillus* AroA LCA and the two domains forming the inside-out α/β barrel configuration within an individual AroA LCA monomer

A) Overlay of *Bacillus* AroA LCA chain A (blue) and chain B (magenta). Chain B is in a more open state than chain A by an angle of approximately 8 °. B) Cartoon representation of domain I and II of chain B with the six subunits labelled (I-VI). The orientation chosen shows the inside-out α/β barrel configuration of *Bacillus* AroA LCA.

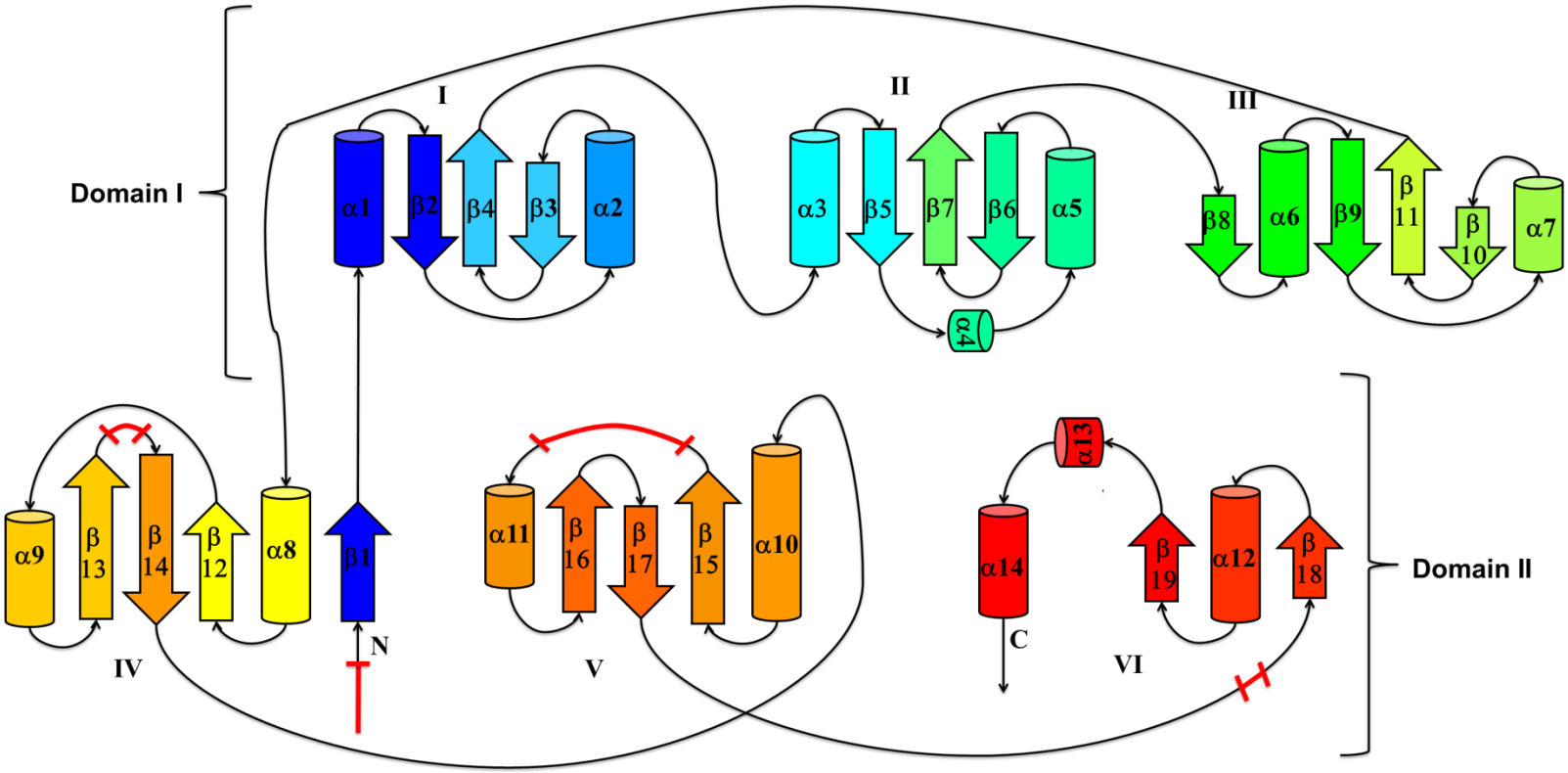


Figure 5.13: Topology sketch of *Bacillus* AroA LCA (Chain B)
The figure displays the two globular domains I and II, each consisting of three subunits (I-VI). Each α helix and β strand is labelled sequentially. The components are coloured corresponding to the colouration used in Figure 5.11. The missing regions in the structure are indicated in red lines.

Bacillus AroA LCA protein was purified using size exclusion chromatography and showed two peaks with the dominant peak corresponding to a dimer. The dimer of *Bacillus* AroA LCA was used for crystallography and the protein crystallised as a dimer (Figure 5.14) as *Bacillus* MurA LCA did. The dimer is formed between the chains from adjacent unit cells sharing a contact area of 725.4 $Å^2$ as calculated by PDBePISA (Krissinel & Henrick 2007). PyMOL was used to generate the symmetry mates for *Bacillus* AroA LCA (Figure 5.14) and this confirmed the formation of dimer.

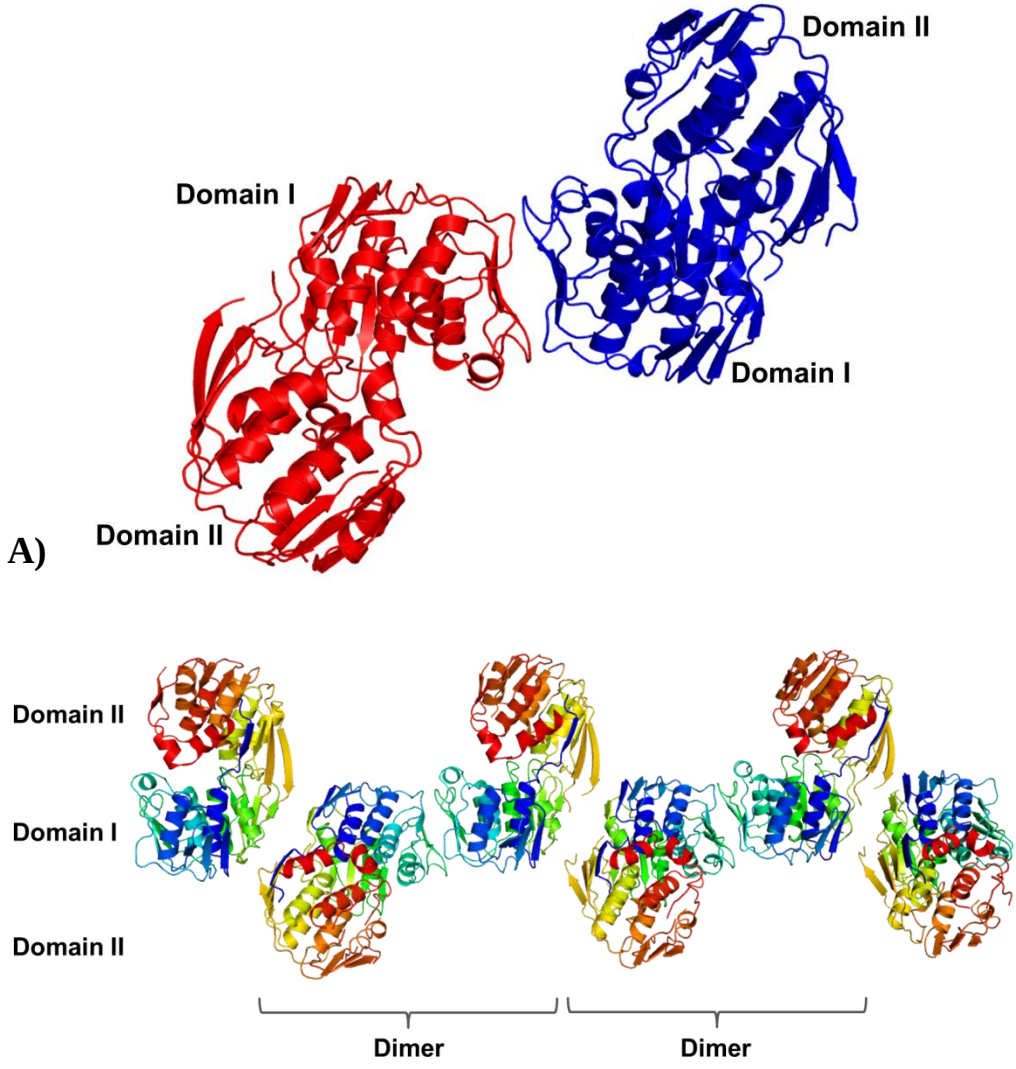


Figure 5.14: *Bacillus* AroA LCA dimer
A) Cartoon representation of *Bacillus* AroA LCA dimer formed by chains from adjacent unit cells. B) Cartoon representation of *Bacillus* AroA LCA dimer formed from symmetry mates generated by PyMOL.

5.2.2.5 Substrate binding site

Similar to that of MurA, the two domains of AroA approach each other in a screw like movement to form the active site in the interdomain cleft (Skarzynski et al. 1996; Schönbrunn et al. 2001). Many residues have been identified in different structures as residues involved in the catalytic activity. However a few residues have been specifically identified as universal active site residues and they are Asp313, Arg386, Arg344, Glu341, Lys411, Arg124 and Lys22 (*E.coli* numbering)(Eschenburg et al. 2003). The residue Asp313 plays an important role in the addition-elimination reaction of the enzyme. In the absence of the aspartate side chain, the reaction will stop after the addition step (Eschenburg et al. 2003). It was proposed by Schönbrunn et al. (2001) that Glu341 acts as a proton donor to the substrate PEP. The remaining residues have been implicated to be taking part in PEP binding (Shuttleworth et al. 1999; Schönbrunn et al. 2001; Eschenburg et al. 2003). It has also been reported that mutagenesis of Lys411 and His385 results in inactive enzymes or drastic decreases in catalytic efficiency (Schönbrunn et al. 2001; Park et al. 2004). The above mentioned catalytically important residues correspond to Asp316, Arg390, Arg347, Glu344, Val416 (instead of Lys), Arg123, Lys23, and His389 respectively in the *Bacillus* AroA LCA structure. These residues are illustrated in Figure 5.15, however residues Arg347 and Glu344, which are part of the highly flexible loop missing in the *Bacillus* AroA LCA structure, are unable to be shown. Residue Lys411 from *E.coli* was found to be involved in the binding of PEP, however, the corresponding residue Thr412 in *S. pneumoniae* AroA does not interact with the substrates (Park et al. 2004). The corresponding residue in *Bacillus* AroA LCA is Val416 which is conserved in *Bacillus* species and may also not interact with PEP. An overlay of *Bacillus* AroA LCA with contemporary *B. halodurans* AroA (PDB 3RMT) (both unliganded), shows the orientation and positioning of the mentioned catalytically important residues (Figure 5.15). The slight differences observed in the orientation of the residues may indicate a possible effect on the catalytic mechanism. However, further comparisons with a structure of *Bacillus* AroA LCA bound to its substrate would need to be performed before any conclusions may be drawn.

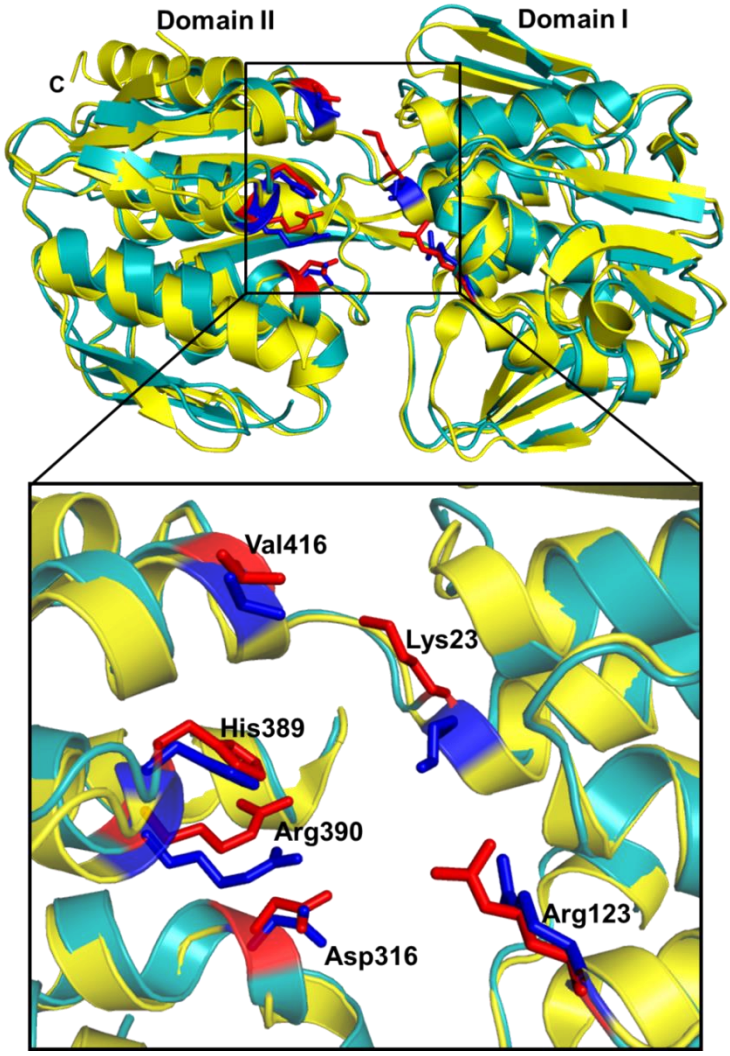


Figure 5.15: Overlay of *Bacillus* AroA LCA and *Bacillus halodurans* AroA showing the active site residues of the enzymes
Bacillus AroA LCA (blue) is overlaid onto *B. halodurans* AroA (yellow) (PDB 3RMT). The figure shows the differences in positioning and orientation of AroA active site residues from *B.* AroA LCA (red) and *B. halodurans* AroA (dark blue). Labels correspond to *Bacillus* AroA LCA numbering only.

5.2.2.6 Inhibitor binding in the active site

Bacillus AroA LCA shows many characteristics similar to class II AroA, yet its sensitivity towards glyphosate in comparison to class II *B. subtilis* AroA has led us to the hypothesis that AroA LCA may be neither from class I nor class II (section 4.2.5.1). For a comprehensive analysis of the AroA LCA structure to verify this hypothesis, it may be necessary to obtain the AroA LCA structure bound to its inhibitor. However, currently this may only be explored using the unliganded AroA LCA structure in comparison to unliganded class II *S. pneumoniae* AroA structure. Comparison to inhibitor bound AroA structures

will not provide the necessary information as unliganded AroA adopts an open conformation whereas liganded AroA adopts a closed conformation (Schönbrunn et al. 2001).

Glyphosate was found to interact with amino acid residues Lys20, Ser21, Arg25, Gly92, Thr93, Arg120, Arg124, Ser166, Ala167, Gln168, Asp312, Lys339, Glu340, Arg343, His384 and Arg385 within class II *S. pneumoniae* AroA structure (Park et al. 2004). These amino acid residues are conserved within the AroA LCA structure with minor changes in the position and orientation of the residues (Figure 5.16) (Numbering of the above residues in *B.* AroA LCA: Lys23, Ser24, Arg28, Gly95, Thr96, Arg123, Arg127, Ser168, Ala169, Gln170, Asp316, Lys343, Glu344, Arg347, His389 and Arg390). The residues Lys343, Glu344 and Arg347 cannot be shown in Figure 5.16 as they are absent from the structure due to lack of electron density. Similar to that of *E. coli* AroA which shows a difference in glyphosate sensitivity due to steric hindrances upon Gly96Ala mutation (Priestman et al. 2005b), the difference in sensitivity between AroA LCA from contemporary class II AroA enzymes may be explained to be the result of mere rearrangements of amino acids within the AroA LCA structure.

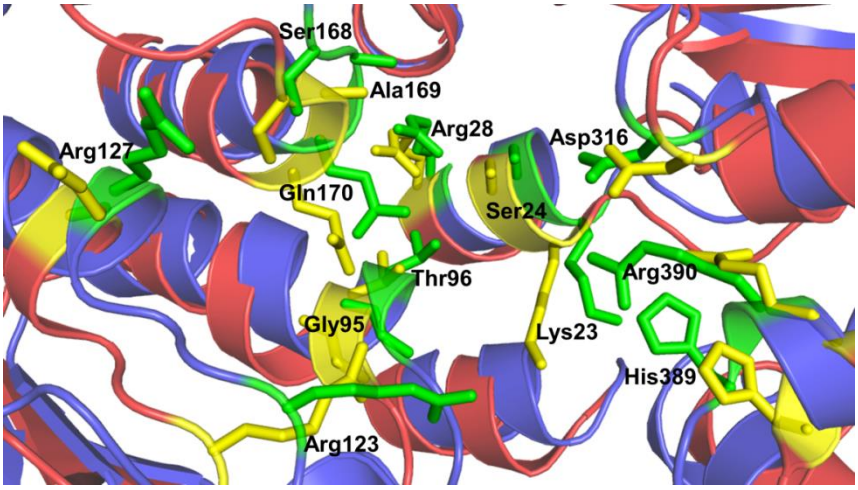


Figure 5.16: Overlay of *Bacillus* AroA LCA with *Streptococcus pneumoniae* AroA showing the residues involved in glyphosate interaction
S. pneumoniae AroA (red) and its residues (yellow); *B.* AroA LCA (blue) and its residues (green). The residue numbering corresponds to *B.* AroA LCA.

5.2.2.7 Comparison to contemporary structures

The structural comparison of *Bacillus* AroA LCA against the entire PDB archive was performed using PDBeFOLD (Krissinel & Henrick 2004) which found the structure of *B. halodurans* AroA (PDB 3RMT) to be the closest structural homologue with a sequence identity of 71%. Table 5.7 gives the structural alignment parameters for a selection of structures discussed in this chapter (3RMT, 1RF5 and 2GG4 are unliganded and 1G6S is bound glyphosate and shikimate-3-phosphate). It is clear from the table that the structure of *Bacillus* AroA LCA has an overall high structural similarity to the different contemporary AroA enzymes as indicated by the %sse. *E.coli* shares the lowest sequence identity of 28% and still identified 79% of secondary structures of *Bacillus* AroA LCA.

Table 5.7: PDBeFOLD structural alignment of PDB structures

Structure	RMSD	% seq[1]	Q-score	P-score	Z-score	% sse[2]	PDB code	Reference
B. halodurans	1.56	71	0.73	44.1	20.1	100	3RMT	---
S. pneumoniae	2.13	52	0.60	22.0	14.9	85	1RF5	(Park et al. 2004)
***Agrobacterium* Cp4**	2.15	43	0.55	24.1	15.4	86	2GG4	(Funke et al. 2006)
E. coli	1.92	28	0.58	17.7	13.4	79	1G6S	(Schönbrunn et al. 2001)

[1] %seq : % of sequence identity
[2]%sse: % of secondary structure of *Bacillus* AroA LCA identified in target protein

The structure overlays of *Bacillus* AroA LCA in its unliganded form with unliganded *B. halodurans* AroA (PDB 3RMT) and *S. pneumoniae* AroA also unliganded (PDB 1RF5) are shown in Figure 5.17. The general fold of the structures are very similar. The topology Figure 5.13 was used to identify the differences between *Bacillus* AroA LCA and contemporary AroA structures from Table 5.7. It was observed that, in comparison to *Bacillus* AroA LCA structure, the structures of contemporary AroA have an extra β strand in subunits I, II, V

and VI positioned beside α1, α3, α10 and β19 respectively, thus generating the repetitive motif of βαβαββ in all the subunits except for subunit VI (*Agrobacterium* CP4 is the only one to form the βαβαββ in subunit VI). It is possible that the additional β strand in subunit VI may be present in *Bacillus* AroA LCA, however this density is missing in the structure. Detailed analysis of the contemporary structures reveals an additional short α helix in subdomains II and VI, corresponding to α4 and α13 in *Bacillus* AroA LCA. These additional α helices are not usually mentioned in the broad topological descriptions in the current literature. Other differences between the ancestral AroA structure and the contemporary AroA structures lie in the lengths and orientations of α helices, β strands and loops.

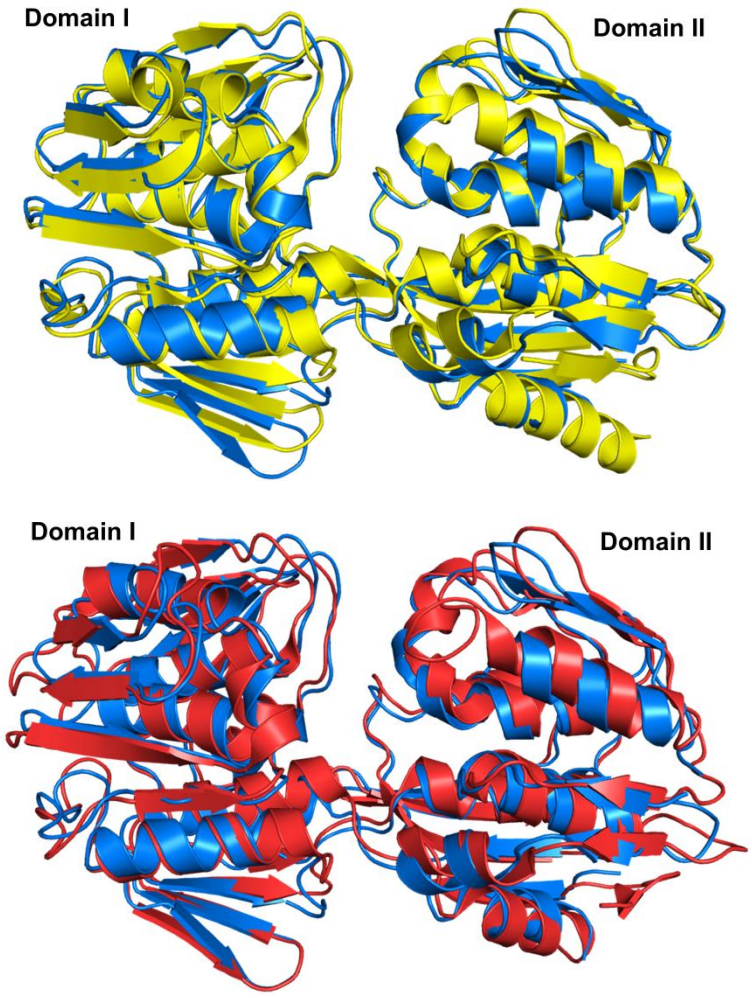


Figure 5.17: Secondary structure matching (SSM) overlay of *Bacillus* AroA LCA with contemporary structures showing the overall structural similarity
Bacillus AroA LCA (blue) overlaid with A) *B. halodurans* AroA (PDB 3RMT) (yellow) and B) *S. pneumoniae* AroA (PDB 1RF5) (red).

The overall structure of *Bacillus* AroA LCA appears to be similar to that of contemporary AroA. The active site residues of the ancestral enzyme are in a similar position to the corresponding residues in its closest structural homologue *B. halodurans* AroA. However, it should be noted that both structures are in their unliganded forms. In order to properly analyse the catalytic site residues and inhibitor interactions, it is necessary to have a *Bacillus* AroA LCA structure in its bound form to its substrates and/or glyphosate.

Further comparison between ancestral AroA enzyme and contemporary *B. halodurans* AroA enzyme was achieved by a detailed analysis of their amino acid compositions (Table 5.8). All the amino acid differences between *Bacillus* AroA LCA and *B. halodurans* AroA that are present on the LCA structure are shown in Figure 5.18. Majority of the amino acid differences are located on the surface of the AroA LCA protein structure with no amino acid closer than 6Å to the substrate binding site.

The thermostability of *Bacillus* AroA LCA was found to differ from that of contemporary *B. halodurans* AroA, in a similar way as *Bacillus* MurA LCA. Therefore, an attempt to find out the factors responsible for thermophily in ancestral AroA was performed as was done for MurA in section 5.2.1.7. The amino acid composition of moderately thermophilic *Bacillus* AroA LCA and mesophilic *B. halodurans* AroA were analysed to identify any trends which could explain the increase in thermostability in ancestral AroA.

The increase in hydrophobic residues and a decrease in polar residues are commonly associated to increased thermophily (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001; Zhou et al. 2008; Taylor & Vaisman 2010). However, it must be noted that in the case of *Bacillus* AroA LCA, there is an apparent decrease in total number of non polar residues (Table 5.8). Furthermore, there is an increase in the number of polar residues in AroA LCA in comparison to *B. halodurans* AroA, which is the opposite of the trend generally observed in thermophilic proteins.

The amino acid differences between *Bacillus* AroA LCA and *B. halodurans* AroA are mostly located on the surface of the protein (Figure 5.18). Therefore, these surface amino acid residues may be responsible for the increased thermophily in

Bacillus AroA LCA as the amino acid composition on the protein surface is known to adapt to the physico-chemical environment of the protein (Fukuchi & Nishikawa 2001).

Another major feature associated with increase in thermophily is the presence of increased ionic interactions on the protein surface/increased charged residues (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001; Zhou et al. 2008; Taylor & Vaisman 2010). In contrary to *Bacillus* MurA LCA, which failed to show any trend with regards to charged residues, *Bacillus* AroA LCA has an apparent increase in number of charged residues in comparison to *B. halodurans* AroA (Table 5.8). This may have contributed to the increase in thermophily in AroA LCA.

Similarly to *Bacillus* MurA LCA, AroA LCA also shows a decrease in Ser content which is also known to be a trend observed in thermophiles (Table 5.8) (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001). Other singular amino acid residue trends observed in thermophily such as lower Met, higher Glu, Arg and Pro content were also analysed (Zhou et al. 2008). Although there is a decrease in Met content in *Bacillus* AroA LCA, it only differs by three residues in comparison to *B. halodurans* AroA. An increase in Glu content is observed, however, this was not observed for Arg and Pro residues.

Therefore, themophily in *Bacillus* AroA LCA might be a consequence of increased number of charged residues or trends observed in individual amino acids such as Ser, Glu or Met. However, it is not possible to find the exact factor responsible for thermophily as a universal mechanism for thermostability has not been identified yet (Yano & Poulos 2003). It must be noted that some of the trends observed in AroA LCA differ from the trends observed in MurA LCA. Therefore, the factors responsible for thermophily may be different in different proteins and/or any of the identified factors singularly or in combination may have influenced thermophily, as even subtle redistributions can affect thermostability (Zhou et al. 2008; Hobbs et al. 2012)

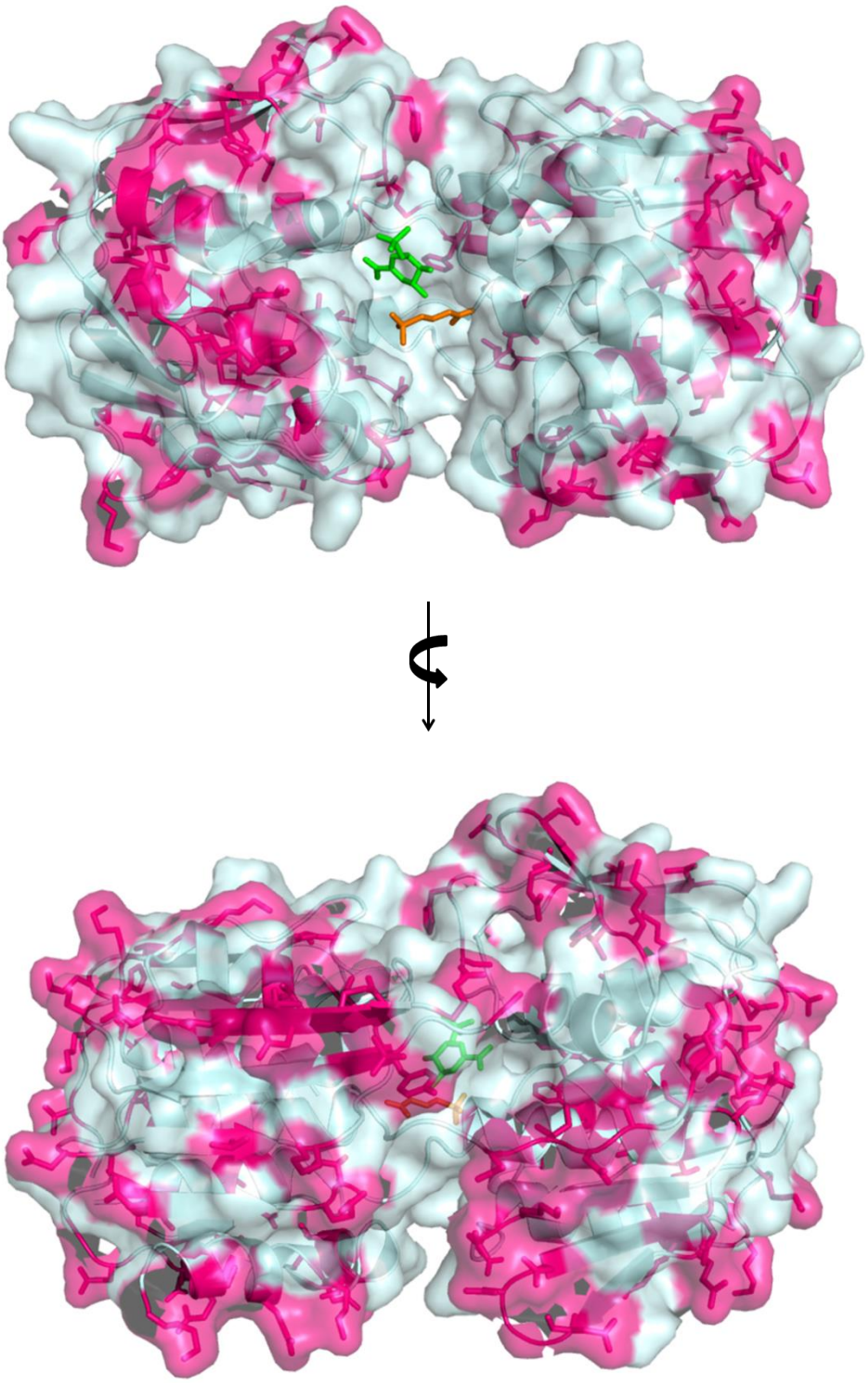


Figure 5.18: Amino acid differences between *Bacillus* AroA LCA and *Bacillus halodurans* AroA

The cartoon representation of Bacillus AroA LCA structure (blue) with a transparent surface is shown here. The residues that differ between AroA LCA and *B. halodurans* AroA are highlighted on the AroA LCA structure (magenta). The surface exposed residues show magenta colour on the surface. The active site is shown by placing S3P (green) and glyphosate (orange) from *S. pneumoniae* AroA structure (1RF6).

Table 5.8: Amino acid composition of *Bacillus* AroA LCA and *Bacillus halodurans* AroA

Amino acid group[1]	Amino acid	*Bacillus* AroA LCA	*Bacillus halodurans* AroA
Non polar	G	47	48
	A	31	35
	V	28	35
	L	36	38
	I	42	37
	P	16	20
	M	10	13
Total		**210**	**226**
Polar - Uncharged	N	15	8
	Q	8	11
	C	3	2
	S	23	29
	T	35	35
Total		**84**	**85**
Polar – Charged	R	18	19
	K	31	25
	H	8	12
	D	28	22
	E	38	30
Total		**123**	**108**
Aromatic	F (NP)[2]	10	10
	W (NP)[2]	0	1
	Y (P)[3]	3	1
Total		**13**	**12**
		430 residues in total	**431 residues in total**

[1] The amino acid grouping correspond to classification by Zhou et al. (2008)
[2] NP = Non polar residue
[3] P = Polar residue

5.2.3 Conclusion

Many studies have identified the structural similarities and differences of MurA and AroA (Schönbrunn et al. 1996; Skarzynski et al. 1996). Therefore, a comparison between the *Bacillus* LCA of MurA and AroA would provide us with insight into the evolution of these enolpyruvyl transferase enzymes. However, an accurate comparison is not possible with the currently obtained structures as *Bacillus* MurA LCA is in its bound form (closed conformation) and *Bacillus* AroA LCA is in its unbound form (open conformation).

The overall enzyme structure consisting of two globular domains built by a 6-fold repetition of βαβαββ motif to give an inside out α/β barrel configuration is the unique protein architecture shared by MurA and AroA enzymes (Eschenburg et al. 2003; Oberdorfer et al. 2012). From the description of *Bacillus* MurA LCA and *Bacillus* AroA LCA in section 5.2.1 and section 5.2.2 respectively, it is clear that both ancestral enzymes share a similar protein fold made of two globular domains consisting of a total of six individual subunits, maintaining the inside out α/β barrel configuration. While some of the subunits within MurA LCA and AroA LCA retained the commonly described βαβαββ motif, some of the subunits showed deviations from this. The structure of *Bacillus* AroA LCA showed more such deviations in comparison to the contemporary AroA enzymes than the deviations observed between *Bacillus* MurA LCA and contemporary MurA enzymes. It has also been noted that there is an additional α helix in subunit VI in *Bacillus* MurA and AroA LCA as well as in all contemporary MurA and AroA enzymes discussed in this chapter. Overall, both MurA and AroA LCA maintain a high structural similarity with contemporary enzymes, with the differences between LCA enzyme and contemporary enzymes predominantly present on the protein surface leaving the active site intact.

The active site of MurA and AroA LCA is situated in the interdomain pocket, similar to that of contemporary MurA and AroA enzymes. Strictly conserved and catalytically important residues of MurA and AroA enzymes are also conserved within the LCA enzymes. However, upon comparison of unliganded AroA LCA with unliganded *B. halodurans* AroA, the orientation and positioning of some of these residues were found to differ. The impact of such changes in the catalytic

properties of AroA could be further investigated after obtaining the structure of AroA bound to its substrate.

The position and orientation of amino acid residues and UNAG in contemporary MurA that are believed to prevent inhibition by the AroA inhibitor glyphosate, remains the same within MurA LCA, suggesting that the LCA also do not support inhibition by glyphosate. The residues involved in the interaction with glyphosate in class II AroA are conserved within AroA LCA with minor variations in position and orientation. However, an inhibitor bound structure of AroA LCA would provide more information on the inhibitory mechanism of AroA LCA.

MurA LCA and AroA LCA are moderately thermophilic and both enzymes show different trends which are attributed to increase in thermophily. While MurA LCA shows an increase in hydrophobic residues, a decrease in polar residues and a decrease in Ser residues, AroA LCA shows an increase in charged residues, decrease in Ser residues and the opposite trend with respect to hydrophobic and polar residues. Therefore, it is difficult to ascertain the exact factors contributing to the increased thermostability in MurA and AroA LCA, in comparison to their mesophilic counterparts. Any of these factors alone or in combination or, any of the differences present on the protein surface, may have been responsible for the increased thermostability.

It would be of further interest to obtain the structure of *Bacillus* AroA LCA bound to its substrate for detailed analysis of its active site. The structures of MurA and AroA LCA bound to their respective inhibitors would also provide a more detailed understanding of inhibitory site of the enzymes.

Chapter 6

Discussion

6.1 General background

The two enolpyruvyl transferase enzymes, AroA and MurA, from the shikimate pathway and peptidoglycan synthesis pathway respectively, have been extensively studied as they are attractive targets for the development of antimicrobial drugs due to the absence of these pathways in mammals, as well as their essentiality to the survival of microorganisms (Eschenburg et al. 2003). Information available in the literature ranges from the enzyme mechanism, enzyme kinetic characteristics, structural properties, inhibitory mechanism and possible new inhibitor designs. The functional and structural similarities of AroA and MurA known to date also leads us to believe that AroA and MurA may have originated from a common ancestor, making the study of their evolution all the more interesting.

Ancestral sequence reconstruction (ASR) was chosen as the tool to study the evolution of AroA and MurA enzymes. The methodology of ancestral sequence reconstruction used in this study has been used for various evolutionary studies, including the study of thermophily within the *Bacillus* genus (Hobbs et al. 2012), reconstruction of ancestral β-lactamases (Risso et al. 2013), and studies on the mechanism of gene duplication (Voordeckers et al. 2012). In this research, ASR was used to resurrect ancestral enzymes of AroA and MurA from the *Bacillus* genus.

AroA and MurA are exclusively inhibited by glyphosate and fosfomycin respectively (Skarzynski et al. 1996; Eschenburg et al. 2003). Hence characterisation of ancestral enzymes with respect to their inhibitor specificity, as well as their inhibitor promiscuity, will provide us with evolutionary information on these enzymes as well as allow us to explore the hypothesis of AroA and MurA originating from a common ancestor.

The aim of this study was to reconstruct ancestral AroA and MurA enzymes from the *Bacillus* genus and compare their kinetic, inhibitory, thermal and structural properties to contemporary *Bacillus* species in order to attain knowledge pertaining to the evolution of AroA and MurA. This study also focussed on

assessing the robustness of the ASR technique in the study of evolution of thermophily.

6.1.1 Functional Evolution of MurA and AroA

Four ancestral MurA enzymes, estimated to be between 550 myr and 950 myr old, were reconstructed to study the evolution of MurA enzymes, and concurrently investigate the evolution of thermophily using these ancestral enzymes. The inferred ancestral sequences have a 76-90% sequence identity to the contemporary sequences used for the inference. The four MurA ancestral enzymes were catalytically active presenting comparable kinetic parameters to those of contemporary *Bacillus subtilis* and *Bacillus coagulans* MurA enzymes at their optimum temperature of activity (T_{opt}). MurA ANC1 and MurA LCA have very similar kinetic properties only differing in their T_{opt} by 4 °C. Out of all the characterised ancestral MurA and contemporary MurA enzymes, MurA ANC2 shows the highest catalytic efficiency. Inhibition of MurA LCA by the MurA inhibitor fosfomycin does not differ substantially from the inhibition exhibited by *B. subtilis* MurA. It is also clear MurA LCA cannot be specifically inhibited by the AroA inhibitor glyphosate. The thermoactivity (T_{opt}) and thermostability (T_m) of *B. subtilis* and *B. coagulans* MurA enzymes confirmed the thermal adaptation of their host organisms, thus confirming the correlation between organism growth temperature and enzyme T_{opt} and T_m. All four MurA ancestors exhibited similar T_{opt} and T_m profiles, all being moderately thermophilic.

In a similar way to MurA, AroA from the LCA of *Bacillus* genus was reconstructed using ASR. The sequence of the resurrected AroA ancestor, aged 957 myr old, has a 64-82% sequence identity to the contemporary *Bacillus* species used for the inference. Shikimate-3-phosphate (S3P), substrate of AroA enzyme reaction, was synthesised and purified to 12.7% purity for the characterisation of the AroA enzyme. The catalytically active AroA LCA showed comparable kinetic parameters to *B. subtilis* AroA, with AroA LCA showing more than twice the catalytic efficiency (k_{cat}/K_M) of *B. subtilis* AroA at the enzyme's T_{opt}. Contemporary *B. subtilis* AroA, classified as a class II AroA, concurred with the class II AroA characteristics of having very high tolerance to glyphosate as well as high affinity towards PEP. AroA LCA, on the other hand, is considerably more sensitive towards glyphosate than *B. subtilis* AroA by approximately 12 fold, yet exhibits high affinity towards PEP. As expected, similar to that of MurA, AroA

LCA does not show sensitivity toward the MurA inhibitor fosfomycin. Thermal characterisation of AroA LCA resulted in moderately thermophilic properties. Ancestral enzymes reconstructed *via* ASR may also be used to study the relationship between its biochemical properties and the organismal fitness (Hobbs et al. 2015). Unlike *B. subtilis* AroA, AroA LCA failed to rescue the Δ*aroA E. coli* cells in spite of having favourable biochemical properties such as thermostability, increased affinity towards its substrates, as well as greater catalytic efficiency than its contemporary counterpart.

6.1.2 Structural evolution of MurA and AroA

The structure of *Bacillus* MurA LCA was solved at a resolution of 2.11Å at spacegroup $P2_12_12_1$. MurA LCA crystallised bound to its substrate UNAG and a phosphate molecule. The structure of *Bacillus* AroA LCA, in the apo-form, was solved to a resolution of 2.15Å with $P2_1$ space group.

The overall structure of *Bacillus* MurA LCA and *Bacillus* AroA LCA consists of two globular domains connected *via* a double stranded linker. The entire structure consists of six very similar motifs, many of them comprising typically of four β-strands and two α-helices, giving an overall inside out α/β barrel structure, which is a characteristic feature of enolpyruvyl transferase enzymes.

MurA shows variation from the characteristic βαβαββ repetitive motif in its subunits II and VI (Figure 5.3). However, subunit VI in MurA LCA is consistent with that of contemporary MurA structures, as it was observed that the additional short α helix in subunit VI constituting $\beta\alpha\beta\alpha^{*}\alpha\beta\beta$ motif is a common feature among contemporary MurA structures, inspite of not being specifically mentioned in the literature. UNAG was situated in the catalytic site of MurA, in the interdomain cleft, with the enzyme in its 'closed state' with the loop from Pro114-Pro123 forming a lid over the catalytic site. The most commonly recognized catalytically important and conserved residues in MurA enzymes were also identified in MurA LCA. The AroA inhibitor glyphosate does not specifically inhibit MurA LCA. The structure of MurA LCA confirmed that the distance clashes associated with the inability of glyphosate to inhibit MurA enzymes (Eschenburg et al. 2003) were also present within MurA LCA, hence confirming that no structural changes have occurred to allow inhibition by glyphosate. The structure of MurA LCA shows an overall high structural similarity to

contemporary MurA enzymes, the closest being to *Bacillus anthracis* MurA (PDB 3SG1), differing mainly with respect to the relative orientation and length of individual loops, β strands and α helices. Such structural differences were observed on the protein surface, leaving the active site very similar to that of contemporary MurA enzymes.

Bacillus AroA LCA was crystallised with two chains in the asymmetric unit, with one chain more open than the other by approximately 8 °. This resulted in topological differences between both chains caused by disruptions in hydrogen bonds defining α helices and β strands. AroA LCA shows deviations from the repetitive βαβαββ motif in all subunits except for III and IV (Figure 5.13). The additional short α helix in subdomains II and VI is consistent with contemporary AroA enzymes. AroA LCA, crystallised in its unliganded form, displayed the 'open' state conformation with its active site in the interdomain cleft. Conserved and universal active site residues are also conserved within the LCA structure, however the orientation and positioning of these residues differ slightly from that of the contemporary *Bacillus halodurans*, the closest contemporary structure (Figure 5.15). However, the effect of these differences on the catalytic mechanism can only be determined by comparing AroA LCA and contemporary AroA structures bound to its substrates.

The thermal stability of the ancestral enzymes in comparison to the contemporary enzymes was investigated *via* comparison of the enzyme's amino acid composition and structure. The literature provides a variety of factors attributed to thermostability (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001; Zhou et al. 2008; Taylor & Vaisman 2010), for which *Bacillus* MurA LCA and AroA LCA, both moderately thermophilic, exhibit different trends. *Bacillus* MurA LCA in comparison to *B. anthracis* MurA showed a decrease in Serine content, increase in hydrophobic residues and decrease in polar residues. On the other hand, *Bacillus* AroA LCA shows an apparent decrease in total number of nonpolar residues and an increase in the number of polar residues in comparison to *B. halodurans* AroA, which is opposite to the trend generally observed in thermophilic proteins. Additionally, *Bacillus* AroA LCA has an apparent increase in number of charged residues in comparison to *B. halodurans* AroA as well as a decrease in Serine content which are known to be trends observed in thermophiles (Szilágyi & Závodszky 2000; Fukuchi & Nishikawa 2001; Zhou et al. 2008;

Taylor & Vaisman 2010). The majority of the amino acid differences between MurA LCA and *B. anthracis* MurA, and between *Bacillus* AroA LCA and *B. halodurans* AroA, were present on the surface of the protein structure. The amino acid composition on a protein surface adapts to its physico-chemical environment, thus displaying the differences in amino acid composition between proteins of thermophilic bacteria and mesophilic bacteria to be predominantly on the protein surface (Fukuchi & Nishikawa 2001).

Despite all the changes observed on the ancestral MurA and AroA enzymes, it is suspected that 1-3 critical mutations may have been responsible for the exhibited thermal properties. However, the exact mutations responsible for this are unknown. Moreover, subtle global redistributions of amino acids in the protein structure may also have influenced thermostability (Hobbs et al. 2012).

6.1.3 Stability of ancestral enzymes and ASR as a tool to study evolution of thermophily

Ancestral enzymes resurrected using ASR were reported to exhibit high thermostability, kinetic stability, and high catalytic activity in comparison to modern enzymes (Perez-Jimenez et al. 2011; Hobbs et al. 2012; Akanuma et al. 2013; Risso et al. 2013; Groussin et al. 2015; Hobbs et al. 2015). However, there are also reconstructed ancestral enzymes that do not agree with this, such as LeuB ANC2 (Hobbs et al. 2012). In regards to AroA and MurA enzymes, due to their lack of tryptophan residues, kinetic stability which is measured as the free energy for unfolding $\Delta G^{\ddagger}$, could not be determined.

Considering the catalytic properties of AroA and MurA ancestral enzymes, not all the reconstructed enzymes showed higher catalytic activity/catalytic efficiency than modern enzymes. MurA ANC2, ANC3 and AroA LCA showed higher catalytic efficiency than contemporary enzymes, while MurA ANC1 and MurA LCA only showed higher catalytic efficiency to *B. subtilis* MurA (but similar k_{cat}), but lower catalytic efficiency than *B. coagulans* MurA. Therefore, no specific pattern or trend of increasing catalytic activity/efficiency with evolutionary time was observed.

While a broad trend of increasing thermophily has been reported with evolutionary time (Gaucher et al. 2008; Perez-Jimenez et al. 2011; Akanuma et al. 2013), Hobbs et al. (2012), albeit at different timescales, reported fluctuations

within this broad thermal trend, with thermophily evolving twice *via* independent evolutionary paths. *Bacillus* AroA LCA and all *Bacillus* MurA ancestral enzymes are moderately thermophilic with a constant temporal trend of thermophily (Figure 6.1). The temporal trend in thermophily obtained from *Bacillus* MurA ancestors cannot be compared with the trend showed by LeuB due to the difference in timescale of the reconstructed ancestors. The only enzymes that are possible to compare are AroA and MurA LCA, aged 957 and 950 myr old, which is the same age as the LeuB LCA LeuB4 (950 myr). However the thermal properties of the *Bacillus* LCA shown by AroA, MurA and LeuB enzymes differ significantly from each other, with T_{opt} of 51 °C, 58 °C and 70 °C respectively, and T_m of 62 °C, 59.7 °C and 65.3 °C respectively. This implies that the LCA of the *Bacillus* genus, aged ~950 myr old exhibits different thermal properties at the same point of time, which is not realistic.

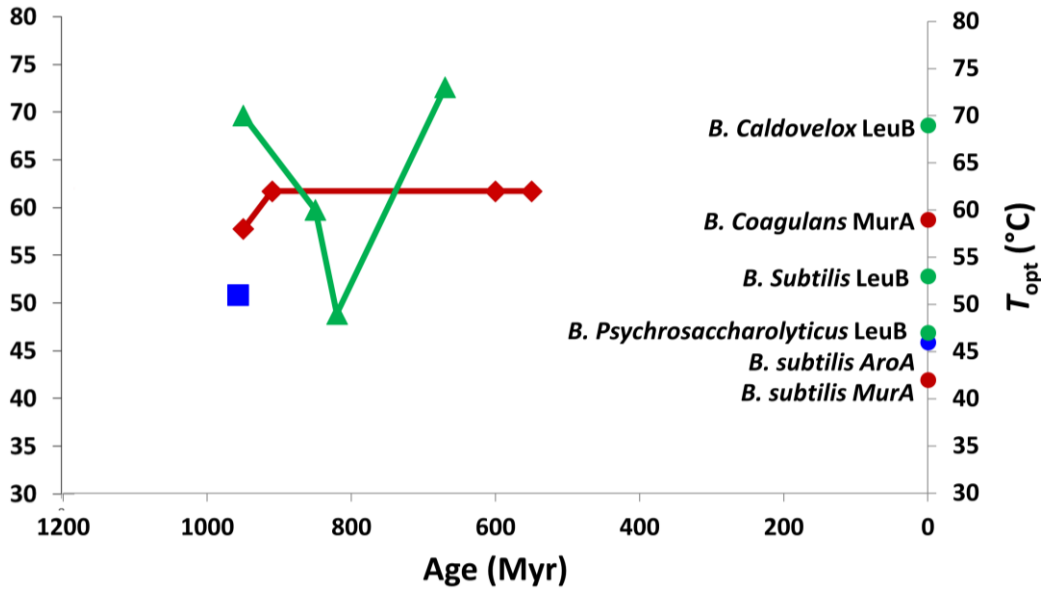


Figure 6.1: Temperature profile displayed by ancestral AroA, MurA and LeuB enzymes
The data shows the age estimates for ancestral nodes against their T_{opt} values for each ancestral enzyme. Data for LeuB is taken from Hobbs et al. (2012). The T_{opt} values for contemporary enzymes are shown at age 0 Myr. Colour and symbol key: LeuB enzymes (green), LeuB ancestors (triangle), MurA enzymes (red), MurA ancestors (diamond), AroA enzymes (blue), AroA ancestor (square) and all contemporary enzymes (circle). The figure also shows that the T_{opt} values obtained for *B. subtilis* enzymes differs beween enzymes.

The occurrence of different temperature profiles by the *Bacillus* LCA from three different enzymes inferred using the same ASR methodology, (considering the absolute thermal values measured here), may be explained three ways: 1) The ancestral inference is protein dependent and depends on the species chosen for the

phylogenetic tree. AroA, MurA and LeuB differ in their inference primarily in the fact that both have different contemporary species and tree topologies; 2) The absolute T_{opt} or T_m values are protein dependant. This is suggested by the difference in T_{opt} or T_m values obtained for *B. subtilis* AroA, MurA and LeuB enzymes (Figure 6.1); 3) As proposed by Woese et al. (1998), the universal ancestor is not a discrete entity but a community. Hence the concept that a group of contemporary species evolved from a single common ancestor may be an oversimplification of the true evolutionary events. Therefore the ancestral enzymes reconstructed here may in fact represent a consortium of species that existed during a particular time point. The present day *Bacillus* species inhabit multiple environments. Similarly, the data obtained here for the reconstructed ancestor may represent a single candidate from the consortium, ie., each reconstructed ancestor only represents a snapshot of the evolving host.

ASR remains useful in the study of thermophily, however, the biological conclusions and thermal trends inferred from resurrected ancestors needs to be treated with caution as, even after surpassing biases associated with ancestral inference, the reconstructed ancestral protein may still only be a representative from a consortium of species present at that point of time and the ancestor with the respective biochemical and biophysical properties may not even have existed.

6.1.4 Evolution of Enzyme inhibition

The specific inhibition of AroA and MurA LCA, by their inhibitors glyphosate and fosfomycin respectively, was explored to evaluate the effect of evolution on the enzyme's inhibitory properties. While MurA LCA showed very similar inhibitor specificity to fosfomycin in comparison to its contemporary enzyme, the inhibition of AroA LCA in comparison to *B. subtilis* AroA by glyphosate showed interesting characteristics. AroA LCA showed approximately 12 fold more sensitivity towards glyphosate in comparison to the naturally tolerant class II *B. subtilis* AroA. Yet AroA LCA maintains high affinity to PEP, similar to *B. subtilis* AroA, a characteristic of class II enzymes. AroA LCA also shares a sequence identity of 76-77% with class II *Bacillus cereus* AroA and *B. subtilis* AroA. The residues involved in glyphosate binding in class II *Streptococcus pneumoniae* AroA, were also identified within the AroA LCA structure, with minor variations in their position and orientation. This evidence indicates AroA LCA to be intermediate between class I and class II AroA and the present day

class II *Bacillus* species evolved from a glyphosate sensitive enzyme. It also agrees with the hypothesis that reconstructed ancestors only represent the properties of a candidate from a consortium of species present at a particular time point. Therefore it is possible that the LCA is only a representative candidate with different properties to the typical class I and class II enzymes.

A phylogenetic analysis demonstrates that class I and class II AroA are distinctly grouped together (Figure 6.2) (previous studies have also shown this phylogenetic distinction between class I and class II (Tian et al. 2012; Zhang et al. 2014)). This sets the question as to which among these two classes would have been the LCA of both classes. The preliminary data obtained from AroA LCA suggests that the ancestor would have been class I AroA as AroA LCA is observed to become more sensitive as it goes back in time.

An alternative reasoning for inhibitory features exhibited by AroA LCA is that the characteristics defining a class I and II AroA may be easily obtainable subjected to appropriate selection pressures. This is exemplified by *Pseudomonas* genus within which some species are tolerant to inhibition by glyphosate while some are sensitive to inhibition by glyphosate (Schulz et al. 1985).

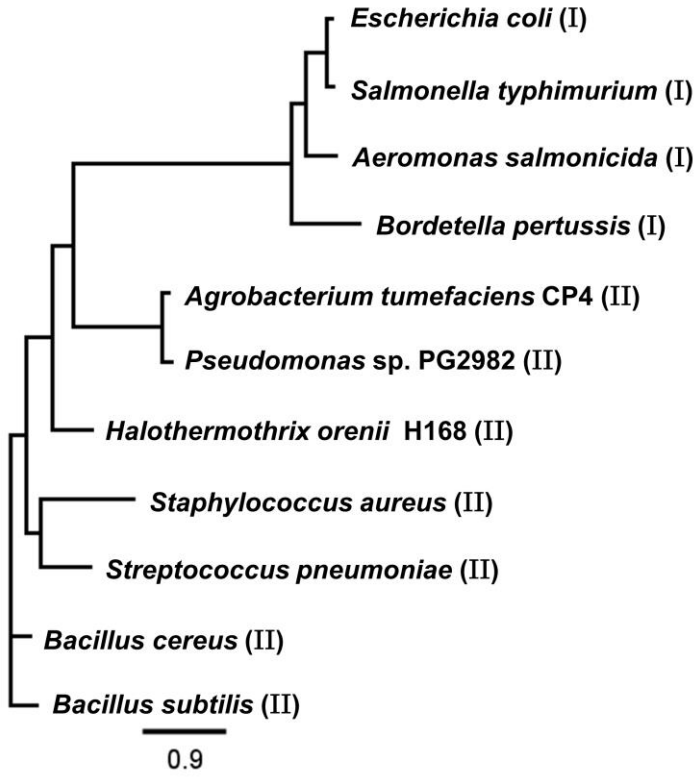


Figure 6.2: Phylogenetic tree analysis of selected class I and class II AroA species
The unrooted phylogenetic tree was generated using GARLI 1.0. The scale for substitution rate per site per unit time is indicated at the bottom of the tree. I and II next to the species indicate the class of each species. The accession numbers for the species used in the phylogenetic tree are given in Table A.3 in Appendix A.

The promiscuity of the inhibitors was also explored to ascertain any evolutionary changes that may allow specific inhibition of AroA and MurA LCA by fosfomycin and glyphosate respectively. If we assume that AroA and MurA enzymes originated from a common ancestor, this implies that going back in evolutionary time, at a certain point both enzymes would show more similarities structurally and functionally to each other. Therefore it is possible that at a particular evolutionary time, both AroA and MurA had biochemical and biophysical characteristics allowing their inhibition by fosfomycin and glyphosate interchangeably. However, upon exploring such inhibitor promiscuity, it was observed that AroA LCA did not show specific inhibition with the MurA inhibitor fosfomycin and vice versa. Hence, it is clear that the LCA of AroA and MurA only possessed structural and functional characteristics to allow exclusive inhibition by glyphosate and fosfomycin respectively at this point of time.

6.1.5 *In vivo* fitness of reconstructed AroA ancestor

Reconstructed ancestral proteins provide a unique opportunity to study the relationship between the protein's biochemical and biophysical properties and organismal fitness *in vivo* (Hobbs et al. 2015). Such a test was possible for AroA. AroA LCA has twice the catalytic efficiency and higher affinity to PEP and S3P than *B. subtilis* AroA, and lower k_{cat} than *B. subtilis* AroA both at 37 °C and at their respective T_{opt}. Upon studying the fitness of these enzymes using Δ*aroA E. coli* cells, AroA LCA failed to rescue the cells at 37 °C inspite of the efficient enzyme properties, in comparison to the contemporary enzyme. However, it has been shown that improved properties such as protein stability and higher catalytic efficiency does not correlate to organismal fitness (Hobbs et al. 2015). There are many possible reasons for the fitness cost observed here. For example, AroA LCA has a three times lower k_{cat} than *B. subtilis* AroA at 37 °C which may also have affected the fitness *in vivo.* Another reasoning for the fitness cost may be associated with an increase in uncoupling between ancestral AroA LCA enzyme and the other enzymes of the shikimate pathway (Hobbs et al. 2015). The regulation of the shikimate pathway in microorganisms is by feedback inhibition and by repression of the first enzyme (Fischer et al. 1986; Herrmann & Weaver 1999). The insertion of ancestral *Bacillus* AroA LCA enzyme in modern *E. coli* cells, may have caused improper regulation of the other enzymes present in the pathway. This in turn may result in the accumulation of different pathway intermediates (Hobbs et al. 2015). The improper regulation of enzymes of the shikimate pathway may lead to an inability to synthesise the aromatic amino acids, or a decrease in end product molecules of the shikimate pathway, resulting in failure of feedback inhibition and repression of the first enzyme DAHP synthase (Fischer et al. 1986). There is also the possibility for toxicity of accumulated intermediate metabolites, especially DAHP or S3P (Fischer et al. 1986). Furthermore, organismal fitness also depends on inter- and intra-molecular interactions such as interaction of the protein with individual components within the cellular network (Kaçar & Gaucher 2013). Differences between interactions exhibited by the contemporary AroA protein within the cellular network in comparison to the ancestral AroA may have also been responsible for the fitness cost imposed by AroA LCA. However, the exact molecular basis is unknown.

6.1.6 Robustness and implications of ASR in study of evolution

The robustness of ancestral inference is fundamental to this study. All the data presented here is in context of this ASR method. Therefore, a discussion of its strengths and weaknesses is important.

The LCA of *Bacillus* genus using AroA and MurA were reconstructed using the Maximum Likelihood criterion (ML). At a given position and at a given internal node, ML inference assigns the ancestral state with the highest posterior probability (Yang 2006). The ML method tends to choose the most frequently observed amino acids at a given site, eliminating the less frequent amino acids that may even be slightly detrimental to structural stability. This in turn leads to the well-known inference bias, overestimation of thermal stability (Williams et al. 2006). An alternative and possible solution to such a bias is the Bayesian inference method which chooses the amino acid residues from the posterior probability distribution instead of choosing the one with the maximum probability (Williams et al. 2006). However, Hobbs et al. (2012) observed a greater number of errors in the ancestral sequence inferred using Bayesian method over ML (Hobbs et al. 2012). Another bias associated with the ancestral inference is the predominance of hydrophobic amino acids in the ancestral sequence (Gaucher et al. 2008).

The data from reconstructed *Bacillus* AroA and MurA ancestral enzymes does not support bias in their reconstruction. Evidence for unbiased reconstruction is detailed below:

(i) Phylogenetic robustness

The phylogenetic trees for the ancestral inference of AroA, MurA and LeuB enzymes all employed extant protein sequences from the *Bacillus* genus. While AroA and MurA employed the same set of 20 *Bacillus* sequences, the LeuB phylogenetic tree differed in that it included psychrophilic (20 °C) and thermophilic (60–80 °C) sequences (Hobbs et al. 2012). The three phylogenetically conserved clades identified by Alcaraz et al (2010), the pathogenic clade, the soil clade, and halophile clade are conserved within the three phylogenetic trees. The relationship between the species described by their grouping within the phylogenetic tree is consistent with each other, and with previous literature. For instance, the pathogenic clade, also referred to as *Bacillus cereus* group, comprises of primarily six species, *B. pseudomycoides, B. mycoides,*

B. weihenstephanensis, B. cereus, B. thuringiensis and *B. anthracis;* to which *B. cytotoxicus* is closely related (Guinebretière et al. 2013). Similarly, the grouping of species within the soil and halophile clades are consistent with the literature (Alcaraz et al. 2010; Hobbs et al. 2012; Guinebretière et al. 2013), suggesting confidence in the phylogenetic relationship between species. The robustness of the phylogenies of AroA and MurA are also validated by the high bootstrap percentage, including a 100% bootstrap value for the LCA of the *Bacillus* genus. However, it was observed that these three phylogenetic trees differ from each other in their overall topology. The pathogenic and soil clades are grouped together within MurA (with MurA ANC2 as common ancestor) and LeuB (with LeuB3 as common ancestor), with the halophile clade branching from them, while the soil and halophile clades are grouped together within AroA phylogeny.

(ii) Sequence identity and posterior probability

There is high sequence identity between the species used for the inference. The sequence alignment of *Bacillus* species for MurA has 50.2% identical sites while AroA alignment has 35% identical sites. The fact that there is more difference between the AroA sequences than the MurA sequences may be due to greater rate of evolution as observed from the phylograms (data not shown). This may also explain some of the minor differences in topology between AroA, MurA and LeuB phylogenies.

The reconstructed ancestral sequences show high similarity towards the contemporary sequences, (MurA LCA: 76-90% and AroA LCA 64-82%). However the ancestral enzymes do not show a significantly higher identity to one contemporary sequence than the others, suggesting that there is no bias posed on the ancestral nodes by extant sequences. The high sequence identity between the *Bacillus* sequences implies a higher accuracy for the phylogeny inferred from them (Kumar & Filipski 2007) and will result in a biologically relevant phylogenetic tree (Brown 2002).

All the reconstructed ancestral nodes had high posterior probability given the model of evolution used and low number ambiguous sites.

(iii) Functionality of reconstructed enzymes

The ancestral enzymes show comparable kinetic characteristics to the contemporary enzymes, and the structure of LCA is highly similar to the

contemporary structure. This suggests that the enzymes have been accurately reconstructed, as any error in ancestral inference would have resulted in inactive or biologically unrealistic enzymes.

(iv) Thermal properties

Williams et al (2006) suggested that the ancestral enzymes reconstructed by the ML method tend to overestimate protein thermostability. However, considering the thermal properties of MurA ancestors, although all ancestors were consistently moderately thermophilic, there is a gradual decrease in the thermal properties of the ancestors with evolutionary time, therefore not supporting a bias towards thermostability. The ML method assigns the ancestral sequence with the amino acid residues present in the majority of sequences known to be stabilizing, thus conferring the predicted ancestor a thermostability higher than contemporary orgainsms. However, the AroA and MurA ancestors do not show considerably much higher thermostability than the contemporary sequences such as *Bacillus coagulans*, thereby suggesting no overestimation of thermostability. At the same time, it must be noted that the reconstructed ancestors are not an average of the extant species thermostabilities as shown by LeuB ancestral enzymes (Hobbs et al. 2012). The sequences of the inferred ancestors for AroA and MurA show highest pairwise identity to sequences from different thermal environments: MurA ANC1 to *Bacillus megaterium* (25-30 °C), MurA ANC2 to *Bacillus cytotoxicus* (45-50 °C), MurA ANC3 to *Bacillus cellulosyliticus* (37 °C), MurA LCA to *Bacillus cellulosyliticus* (37 °C), and AroA LCA to *Bacillus cytotoxicus* (45-50 °C) then to *Bacillus pumilus* (25-30 °C) (Table 3.1 and Table 4.1). Therefore there is no specific similarity to contemporary sequences from a specific thermal habitat, suggesting that contemporary sequences do not bias the thermal properties of ancestral sequences and hence they are more likely to exhibit their true thermal properties.

(v) Hydrophobicity

An increase of seven hydrophobic residues was observed within MurA LCA, which accounts for only a 1% increase in hydrophobic residues in comparison to contemporary mesophilic *B. anthracis* MurA (Table 5.4), while on the contrary, *Bacillus* AroA LCA has a decrease of ~4% in hydrophobic residues compared to its contemporary mesophilic *B. halodurans* AroA (Table 5.8), and yet AroA LCA

shows a higher T_m than MurA LCA, hence ruling out bias associated to hydrophobic residues.

Although we cannot completely rule out the possibility of bias or errors in the ancestral inference, this evidence provides a strong confidence in the validity of the resurrected ancestors, hence there is no reason to doubt the validity of the data measured for the reconstructed enzymes.

It is however acknowledged that a variety of assumptions are associated with this study. We assume that the resurrected ancestral enzyme is a representation of the true ancestral state. However, ancestral sequence inference identifies the best approximation of the ancestral state and is not the true ancestor (Thornton 2004). The ancestral reconstruction by ML uses a set model of evolution, alignment and tree topology for inference. We assume confidence in these parameters upon which the ancestral inference depends. It is also assumed that the properties exhibited by the ancestral enzymes were not affected by its expression and characterisation within extant cell systems (Thornton 2004). The absolute values for the thermal properties as well as ancestral evolutionary ages were taken for this study.

In spite of these, this data provides ample information on the structural and functional aspects of AroA and MurA enzyme evolution. There is no specific trend observed with respect to thermal stability and catalytic activity/efficiency with evolutionary time. Such an exhibit of distinct properties by the ancestors suggests that the inference does not show any bias. Furthermore, the mechanistic and structural similarities between AroA and MurA led us to hypothesis that AroA and MurA originated from a common ancestor. Upon analysing the chronogram of prokaryotic evolution generated by Battistuzzi et al. (2004), it is observed that the evolution of MurA may have succeeded AroA, as indicated by the presence of AroA enzyme within archaea, while peptidoglycan is absent from archaea cell wall and hence MurA enzyme. This suggests that the LCA of AroA and MurA may possibly be the LUCA. Therefore, it is only natural that the reconstruction of AroA and MurA ancestral enzymes that date back to only 950 myr old would not show characteristic features differing from contemporary AroA and MurA. Therefore, the results obtained in this study concur with the expected evolutionary behavior.

6.1.7 Implications of ASR

The use of this ASR technique, most certainly has a significant implication in the study of functional and structural protein evolution. However, certain aspects of this ASR technique, such as possible bias due to taxon sampling, and error in inferences due to external factors such as horizontal gene transfer, insertions and deletions, may be addressed with the use of a reconciled tree (Groussin et al. 2015). Reconciled trees (S-aware trees) encompass a phylogeny based on protein sequences (gene tree) as well as a species tree. Incorporating a species tree along with a gene tree, would allow factoring in genetic events such as duplications, deletions and lateral gene transfers, which may otherwise affect the phylogeny based on protein sequences alone. The use of a reconciled tree was shown to improve the accuracy of ancestral inference *in silico* as well as *in vitro* (Groussin et al. 2015). Accurate phylogeny translates into accurate ancestral inference and resurrection of more evolutionarily realistic ancestors (Groussin et al. 2015). Therefore using a reconciled tree for *Bacillus* species with a species tree involving only species that are fully genome sequenced, would account for any genetic events and would standardise the contemporary species used for ancestral inference. This is more likely to result in more consistent phylogenies between proteins. This would also provide further confidence in the node age estimations. The use of a reconciled tree may or may not cause uniformity in the thermal trends exhibited by the ancestors from the same time point. However, this must be confirmed by experimentally testing AroA, MurA and LeuB LCA of the *Bacillus* genus reconstructed using reconciled trees. The current reconstruction method utilises a site-homogenous LG model (Le & Gascuel 2008) which uses only a single substitution matrix for a given phylogenetic tree. On the other hand a site-heterogeneous model utilises several such matrices and helps decrease bias in the estimation of model parameters (Groussin et al. 2013). However, results obtained from trees reconstructed by both site-homogenous model and site-heterogeneous were very similar (Groussin et al. 2015).

Therefore, incorporating the use of a reconciled tree to the current ASR methodology could improve the overall robustness of ancestral inference and hence the biological conclusions drawn from them.

6.2 Future Work

As previously discussed, the use of a reconciled tree for the ancestral inference procedure would result in more accurate phylogenies which in turn would resurrect more evolutionarily accurate ancestral enzymes. Therefore, AroA and MurA ancestral enzymes from the *Bacillus* genus could be reconstructed using a reconciled tree and the enzymes characterised, to see whether the measured properties differ from those observed in this study.

It would be of further interest to obtain the structure of *Bacillus* AroA LCA bound to its substrate for detailed analysis of its active site. This may also help in identifying the structural factors responsible for 3-6 fold higher affinity of *Bacillus* AroA LCA in comparison to *B. subtilis* AroA. The structure of *Bacillus* AroA LCA bound to its inhibitor glyphosate would be necessary to structurally evaluate the reasons for its increased sensitivity towards glyphosate in comparison to *B. subtilis* AroA. The structural comparison between AroA LCA and MurA LCA could not be performed as MurA LCA bound to UNAG was in its 'closed' state and unliganded AroA LCA was in its 'open' state. Therefore, the structure of AroA LCA bound to its substrate or unliganded MurA LCA could be used for such a comparative study.

The *in vivo* fitness cost posed by AroA LCA needs to be investigated further. To achieve this, the following experiments could be carried out (Hobbs et al. 2015): 1) To have the ancestral AroA on the chromosome of contemporary *B. subtilis,* or cloning the genes encoding for *B. subtilis* AroA and *Bacillus* AroA LCA such that they come under the control of the same promoter; 2) growth experiments of Δ*aroA* cells complemented with AroA LCA and *B. subtilis* AroA in M9 media and sampling periodically to determine the growth rates; 3) competition assays may also be performed by growing the two strains in the same culture and estimating the relative fitness of AroA LCA in comparison to the contemporary AroA; and 4) placing the AroA LCA onto the *Bacillus* chromosome and forward evolving the strain until fitness improves. The resulting mutations unique to AroA could then be studied *in vitro* (Hobbs et al. 2015).

As explained in this study, the exact factors involved in the characterisation of class I and class II AroA are unknown. However, it would be of further interest to reconstruct ancestral enzymes that go further back to the LCA of class I and class II AroA and thus gain more insight into the evolution of these two classes of

AroA. Taking the reconstruction process further back in evolutionary time would also allow us further study the evolution of AroA and MurA enzymes. To further explore the hypothesis that AroA and MurA originated from a common ancestor, reconstructing older ancestors may allow us to observe convergence between AroA and MurA enzymes. With the help of new ancestral inference procedures with reconciled trees, there is greater potential for successful reconstruction of ancestral enzymes that are much further back in evolutionary time.

References

Abascal F, Zardoya R, Posada D 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21(9): 2104-2105.

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66(2): 213-221.

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD 2012. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallographica Section D: Biological Crystallography 68(4): 352-367.

Akanuma S, Nakajima Y, Yokobori S-I, Kimura M, Nemoto N, Mase T, Miyazono K-I, Tanokura M, Yamagishi A 2013. Experimental evidence for the thermophilicity of ancestral life. Proceedings of the National Academy of Sciences 110(27): 11067-11072.

Akiyama H, Okunuki H, Tsuzuki S, Arami S, Miura H, Sakushima J, Teshima R, Goda Y, Hino A, Sawada J 1999. Construction of an expression system of shikimate kinase II and preparation of shikimic acid 3-phosphate. Journal of the Food Hygienic Society of Japan:

Alcaraz LD, Moreno-Hagelsieb G, Eguiarte LE, Souza V, Herrera-Estrella L, Olmedo G 2010. Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. BMC Genomics 11(1): 332.

Amer FA, El-Behedy EM, Mohtady HA 2008. New targets for antibacterial agents. Biotechnology and Molecular Biology Reviews 3(3): 46-57.

An M, Maitra U, Neidlein U, Bartlett PA 2003. 5-enolpyruvylshikimate 3-phosphate synthase: Chemical synthesis of the tetrahedral intermediate and assignment of the stereochemical course of the enzymatic reaction. Journal of the American Chemical Society 125(42): 12759-12767.

Anderson KS, Sikorski JA, Benesi AJ, Johnson KA 1988a. Isolation and structural elucidation of the tetrahedral intermediate in the EPSP synthase enzymic pathway. Journal of the American Chemical Society 110(19): 6577-6579.

Anderson KS, Sikorski JA, Johnson KA 1988b. Evaluation of 5-enolpyruvoylshikimate-3-phosphate synthase substrate and inhibitor binding by stopped-flow and equilibrium fluorescence measurements. Biochemistry 27(5): 1604-1610.

Arai K-I, Kawakita M, Kaziro Y 1972. Studies on Polypeptide Elongation Factors from *Escherichia coli* II. Purification of factors Tu-Guanosine

diphosphate, Ts, and Tu-Ts, and crystallization of Tu-Guanosine diphosphate and Tu-Ts. Journal of Biological Chemistry 247(21): 7029-7037.

Arca P, Reguera G, Hardisson C 1997. Plasmid-encoded fosfomycin resistance in bacteria isolated from the urinary tract in a multicentre survey. Journal of Antimicrobial Chemotherapy 40(3): 393-399.

Arenas M, Posada D 2010. The effect of recombination on the reconstruction of ancestral sequences. Genetics 184(4): 1133-1139.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. Molecular Systems Biology 2(1):

Barrett J 2011. Biochemistry of survival. In: Perry RN, Wharton DA eds. Molecular and Physiological Basis of Nematode Survival. Wallingford, U.K., CABI. Pp. 282-310.

Barry GF, Kishore GM, Padgette SR 1992. Patent No. WO 9,204,449. DNA encoding class II 5-enolpyruvyl shikimate-3-phosphate synthase for producing plants and bacteria tolerant to glyphosate herbicides.

Battistuzzi FU, Feijao A, Hedges SB 2004. A genomic timescale of prokaryote evolution: Insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evolutionary Biology 4: 44.

Baum EZ, Montenegro DA, Licata L, Turchi I, Webb GC, Foleno BD, Bush K 2001. Identification and characterization of new inhibitors of the *Escherichia coli* MurA enzyme. Antimicrobial Agents and Chemotherapy 45(11): 3182-3188.

Baykov AA, Evtushenko OA, Avaeva SM 1988. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. Analytical Biochemistry 171(2): 266-270.

Beavis AD, Powers M 2004. Temperature dependence of the mitochondrial inner membrane anion channel. The relationship between temperature and inhibiton by magnesium. Journal of Biological Chemistry 279(6): 4045-4050.

Berti PJ, Chindemi P 2009. Catalytic residues and an electrostatic sandwich that promote enolpyruvyl shikimate 3-phosphate synthase (AroA) catalysis. Biochemistry 48(17): 3699-3707.

Biery JJ 2007. Probing the antibiotic target MurA from *Staphylococcus aureus* and *Bacillus subtilis*. Unpublished MSc thesis, University of Kansas.

Blake KL, O'Neill AJ, Mengin-Lecreulx D, Henderson PJF, Bostock JM, Dunsmore CJ, Simmons KJ, Fishwick CWG, Leeds JA, Chopra I 2009. The nature of *Staphylococcus aureus* MurA and MurZ and approaches for

detection of peptidoglycan biosynthesis inhibitors. Molecular Microbiology 72(2): 335–343.

Blum JS, Bindi AB, Buzzelli J, Stolz JF, Oremland RS 1998. *Bacillus arsenicoselenatis*, sp. nov., and *Bacillus selenitireducens*, sp. nov.: Two haloalkaliphiles from Mono Lake, California that respire oxyanions of selenium and arsenic. Archives of Microbiology 171(1): 19-30.

Brown ED, Marquardt JL, Lee JP, Walsh CT, Anderson KS 1994. Detection and characterization of a phospholactoyl-enzyme adduct in the reaction catalyzed by UDP-*N*-acetylglucosamine enolpyruvoyl transferase, MurZ. Biochemistry 33(35): 10638-10645.

Brown TA 2002. Genomes (Chapter 16: Molecular phylogenetics). 2nd ed. New York, Wiley-Liss, from http://www.ncbi.nlm.nih.gov/books/NBK21122/.

Cai W, Pei J, Grishin NV 2004. Reconstruction of ancestral protein sequences and its applications. BMC Evolutionary Biology 4(1): 33.

Cassidy PJ, Kahan FM 1973. A stable enzyme-phosphoenolpyruvate intermediate in the synthesis of uridine-5'-diphospho-*N*-acetyl-2-amino-2-deoxyglucose 3-O-enolpyruvyl ether. Biochemistry 12(7): 1364-74.

Castañeda-García A, Blázquez J, Rodríguez-Rojas A 2013. Molecular mechanisms and clinical impact of acquired and intrinsic fosfomycin resistance. Antibiotics 2(2): 217-236.

Cerritos R, Vinuesa P, Eguiarte LE, Herrera-Estrella L, Alcaraz-Peraza LD, Arvizu-Gómez JL, Olmedo G, Ramirez E, Siefert JL, Souza V 2008. *Bacillus coahuilensis* sp. nov., a moderately halophilic species from a desiccation lagoon in the Cuatro Cienegas Valley in Coahuila, Mexico. International Journal of Systematic and Evolutionary Microbiology 58(4): 919-923.

Chang C-M, Chern J, Chen M-Y, Huang K-F, Chen C-H, Yang Y-L, Wu S-H 2014. Avenaciolides: Potential MurA-targeted inhibitors against peptidoglycan biosynthesis in methicillin-resistant *Staphylococcus aureus* (MRSA). Journal of the American Chemical Society:

Christensen AM, Schaefer J 1993. Solid-state NMR determination of intra- and intermolecular ^{31}P-^{13}C distances for shikimate 3-phosphate and [1-^{13}C] glyphosate bound to enolpyruvylshikimate-3-phosphate synthase. Biochemistry 32(11): 2868-2873.

Christensen BG, Leanza WJ, Beattie TR, Patchett AA, Arison BH, Ormond RE, Kuehl FA, Albers-Schonberg G, Jardetzky O 1969. Phosphonomycin: structure and synthesis. Science 166(3901): 123-125.

Collins TM, Wimberger PH, Naylor GJP 1994. Compositional bias, character-state bias, and character-state reconstruction using parsimony. Systematic Biology: 482-496.

Costantini S, Colonna G, Facchiano AM 2008. ESBRI: A web server for evaluating salt bridges in proteins. Bioinformation 3(3): 137-8.

Couñago R, Chen S, Shamoo Y 2006. *In vivo* molecular evolution reveals biophysical origins of organismal fitness. Molecular Cell 22(4): 441-449.

Cowtan K 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallographica Section D: Biological Crystallography 62(9): 1002-1011.

Csermely P, Agoston V, Pongor S 2005. The efficiency of multi-target drugs: The network approach might help drug design. Trends in Pharmacological Sciences 26(4): 178-182.

Dai HJ, Parker CN, Bao JJ 2002. Characterization and inhibition study of MurA enzyme by capillary electrophoresis. Journal of Chromatography B 766(1): 123-132.

De Smet KAL, Kempsell KE, Gallagher A, Duncan K, Young DB 1999. Alteration of a single amino acid residue reverses fosfomycin resistance of recombinant MurA from *Mycobacterium tuberculosis*. Microbiology 145(11): 3177-3184.

Denizci AA, Kazan D, Abeln ECA, Erarslan A 2004. Newly isolated *Bacillus clausii* GMBAE 42: An alkaline protease producer capable to grow under higly alkaline conditions. Journal of Applied Microbiology 96(2): 320-327.

DePristo MA, Weinreich DM, Hartl DL 2005. Missense meanderings in sequence space: A biophysical view of protein evolution. Nature Reviews Genetics 6(9): 678-687.

Dib EG, Dib SA, Korkmaz DA, Mobarakai NK, Glaser JB 2003. Nonhemolytic, nonmotile Gram-positive rods indicative of *Bacillus anthracis*. Emerging Infectious Diseases 9(8): 1013-1015.

Du W, Brown JR, Sylvester DR, Huang J, Chalker AF, So CY, Holmes DJ, Payne DJ, Wallis NG 2000. Two active forms of UDP-*N*-acetylglucosamine enolpyruvyl transferase in Gram-positive bacteria. Journal of Bacteriology 182(15): 4146-4152.

Easter AD 2010. Decoupling enzyme catalysis from thermal denaturation. Unpublished MSc thesis, University of Waikato, Hamilton, New Zealand.

El Zoeiby A, Sanschagrin F, Levesque RC 2003. Structure and function of the Mur enzymes: Development of novel inhibitors. Molecular Microbiology 47(1): 1-12.

Emsley P, Cowtan K 2004. Coot: Model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography 60(12): 2126-2132.

Eschenburg S, Healy ML, Priestman MA, Lushington GH, Schönbrunn E 2002. How the mutation glycine96 to alanine confers glyphosate insensitivity to

5-enolpyruvyl shikimate-3-phosphate synthase from *Escherichia coli*. Planta 216(1): 129-135.

Eschenburg S, Kabsch W, Healy ML, Schönbrunn E 2003. A new view of the mechanisms of UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. Journal of Biological Chemistry 278(49): 49215-49222.

Eschenburg S, Priestman M, Schönbrunn E 2005a. Evidence that the fosfomycin target Cys115 in UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) is essential for product release. Journal of Biological Chemistry 280(5): 3757-3763.

Eschenburg S, Priestman MA, Abdul-Latif FA, Delachaume C, Fassy F, Schönbrunn E 2005b. A novel inhibitor that suspends the induced fit mechanism of UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA). Journal of Biological Chemistry 280(14): 14070-14075.

Eschenburg S, Schönbrunn E 2000. Comparative X-ray analysis of the un-liganded fosfomycin-target MurA. Proteins: Structure, Function, and Bioinformatics 40(2): 290-298.

Evans PR, Murshudov GN 2013. How good are my data and what is the resolution? Acta Crystallographica Section D: Biological Crystallography 69(7): 1204-1214.

Fadda G, Cattani P, Rossi A, Sechi L, Ungheri D, Zanetti S 2006. An Atlas of Photographs Illustrating the Effects of Fosfomycin Trometamol on Bacterial Adhesiveness. Retrieved January, 2015, from http://www.gimmoc.it/editoria/quaderni/quaderni_2006/q3-2006.pdf.

Fauci AS 2001. Infectious diseases: Considerations for the 21st century. Clinical Infectious Diseases 32(5): 675-685.

Fischer RS, Berry A, Gaines CG, Jensen RA 1986. Comparative action of glyphosate as a trigger of energy drain in eubacteria. Journal of Bacteriology 168(3): 1147-1154.

Fitch WM 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Systematic Biology 20(4): 406-416.

Forterre P 1995. Thermoreduction, a hypothesis for the origin of prokaryotes. Comptes Rendus de l'Academie des Sciences-Serie III-Sciences de la Vie 318(4): 415-422.

French S, Wilson K 1978. On the treatment of negative intensity observations. Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography 34(4): 517-525.

Fritze D, Pukall R 2001. Reclassification of bioindicator strains *Bacillus subtilis* DSM 675 and *Bacillus subtilis* DSM 2277 as *Bacillus atrophaeus*.

International Journal of Systematic and Evolutionary Microbiology 51(1): 35-37.

Fukuchi S, Nishikawa K 2001. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. Journal of Molecular Biology 309(4): 835-843.

Funke T, Han H, Healy-Fried ML, Fischer M, Schönbrunn E 2006. Molecular basis for the herbicide resistance of Roundup Ready crops. Proceedings of the National Academy of Sciences 103(35): 13010-13015.

Funke T, Healy-Fried ML, Han H, Alberg DG, Bartlett PA, Schönbrunn E 2007. Differential inhibition of class I and class II 5-enolpyruvylshikimate-3-phosphate synthases by tetrahedral reaction intermediate analogues. Biochemistry 46(46): 13344-13351.

Funke T, Yang Y, Han H, Healy-Fried M, Olesen S, Becker A, Schönbrunn E 2009. Structural basis of glyphosate resistance resulting from the double mutation Thr97→ Ile and Pro101→ Ser in 5-enolpyruvylshikimate-3-phosphate synthase from *Escherichia coli*. Journal of Biological Chemistry 284(15): 9854-9860.

Funke TA 2008. Targeting 5-enolpyruvylshikimate-3-phosphate synthase and the shikimate pathway. Unpublished PhD thesis, University of Kansas.

Gasser CS, Winter JA, Hironaka CM, Shah DM 1988. Structure, expression, and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. Journal of Biological Chemistry 263(9): 4280-4287.

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A 2005. Protein identification and analysis tools on the ExPASy server. In: Walker JM ed. The Proteomics Protocols Handbook. Totowa, NJ, Humana Press. Pp. 571-607.

Gaucher EA, Govindarajan S, Ganesh OK 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451(7179): 704-707.

Gaucher EA, Thomson JM, Burgan MF, Benner SA 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425(6955): 285-288.

Gautam A, Rishi P, Tewari R 2011. UDP-*N*-acetylglucosamine enolpyruvyl transferase as a potential target for antibacterial chemotherapy: Recent developments. Applied Microbiology and Biotechnology 92(2): 211-225.

Ge X, d'Avignon DA, Ackerman JJH, Sammons RD 2010. Rapid vacuolar sequestration: The horseweed glyphosate resistance mechanism. Pest Management Science 66(4): 345-348.

Georlette D, Blaise V, Collins T, D'Amico S, Gratia E, Hoyoux A, Marx J-C, Sonan G, Feller G, Gerday C 2004. Some like it cold: Biocatalysis at low temperatures. FEMS Microbiology Reviews 28(1): 25-42.

Gordon RE 1977. The genus *Bacillus*. In: Laskin AI, Lechevalier H eds. CRC Handbook of Microbiology. Vol. 1. Boca Raton, FL, CRC Press. Pp. 319-336.

Griffith SM, Brewer TG, Steiner JJ 2001. Thermal dependence of the apparent K_M of glutathione reductase from three wetland grasses and maize. Annals of Botany 87(5): 599-603.

Gromiha MM, Oobatake M, Sarai A 1999. Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. Biophysical Chemistry 82(1): 51-67.

Groussin M, Boussau B, Gouy M 2013. A branch-heterogeneous model of protein evolution for efficient inference of ancestral sequences. Systematic Biology 62(4): 523-538.

Groussin M, Hobbs JK, Szöllősi GJ, Gribaldo S, Arcus VL, Gouy M 2015. Toward more accurate ancestral protein genotype–phenotype reconstructions with the use of species tree-aware gene trees. Molecular Biology and Evolution 32(1): 13-22.

Gruys KJ, Walker MC, Sikorski JA 1992. Substrate synergism and the steady-state kinetic reaction mechanism for EPSP synthase from *Escherichia coli*. Biochemistry 31(24): 5534-5544.

Guinebretière M-H, Auger S, Galleron N, Contzen M, De Sarrau B, De Buyser M-L, Lamberet G, Fagerlund A, Granum PE, Lereclus D, De Vos P, Nguyen-The C, Sorokin A 2013. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. International Journal of Systematic and Evolutionary Microbiology 63(Pt 1): 31-40.

Hall BG 2006. Simple and accurate estimation of ancestral protein sequences. Proceedings of the National Academy of Sciences 103(14): 5431-5436.

Hanson-Smith V, Kolaczkowski B, Thornton JW 2010. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. Molecular Biology and Evolution 27(9): 1988-1999.

Harms MJ, Thornton JW 2010. Analyzing protein structure and function using ancestral gene reconstruction. Current Opinion in Structural Biology 20: 360-366.

Harms MJ, Thornton JW 2013. Evolutionary biochemistry: Revealing the historical and physical causes of protein properties. Nature Reviews Genetics 14(8): 559-571.

Hendlin D, Stapley EO, Jackson M, Wallick H, Miller AK, Wolf FJ, Miller TW, Chaiet L, Kahan FM, Foltz EL 1969. Phosphonomycin, a new antibiotic produced by strains of *Streptomyces*. Science 166(3901): 122-123.

Herrmann KM 1995a. The shikimate pathway as an entry to aromatic secondary metabolism. Plant Physiology 107(1): 7.

Herrmann KM 1995b. The shikimate pathway: Early steps in the biosynthesis of aromatic compounds. The Plant Cell 7(7): 907.

Herrmann KM, Weaver LM 1999. The shikimate pathway. Annual Review of Plant Biology 50(1): 473-503.

Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ 1992. Experimental phylogenetics: Generation of a known phylogeny. Science 255(5044): 589-592.

Hobbs JK, Prentice EJ, Groussin M, Arcus VL 2015 Reconstructed ancestral enzymes impose a fitness cost upon modern bacteria despite exhibiting favourable biochemical properties. Journal of Molecular Evolution

Hobbs JK, Shepherd C, Saul DJ, Demetras NJ, Haaning S, Monk CR, Daniel RM, Arcus VL 2012. On the origin and evolution of thermophily: Reconstruction of functional Precambrian enzymes from ancestors of *Bacillus*. Molecular Biology and Evolution 29(2): 825-835.

Huelsenbeck JP, Bollback JP 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Systematic Biology 50(3): 351-366.

Jakeman DL, Mitchell DJ, Shuttleworth WA, Evans JNS 1998. Effects of sample preparation conditions on biomolecular solid-state NMR lineshapes. Journal of Biomolecular NMR 12(3): 417-421.

Jin B-S, Han S-G, Lee W-K, Ryoo SW, Lee SJ, Suh S-W, Yu YG 2009. Inhibitory mechanism of novel inhibitors of UDP-N-acetylglucosamine enolpyruvyl transferase from *Haemophilus influenzae*. Journal of Microbiology and Biotechnology 19: 1582-1589.

Jin D, Lu W, Ping S, Zhang W, Chen J, Dun B, Ma R, Zhao Z, Sha J, Li L 2007. Identification of a new gene encoding EPSPS with high glyphosate resistance from the metagenomic library. Current microbiology 55(4): 350-355.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL 2008. NCBI BLAST: A better web interface. Nucleic Acids Research 36(suppl 2): W5-W9.

Jones DT, Taylor WR, Thornton JM 1992. The rapid generation of mutation data matrices from protein sequences. Computer Applications in the Biosciences: CABIOS 8(3): 275-282.

Ju J, Xu S, Wen J, Li G, Ohnishi K, Xue Y, Ma Y 2009. Characterization of endogenous pyridoxal 5′ -phosphate-dependent alanine racemase from *Bacillus pseudofirmus* OF4. Journal of Bioscience and Bioengineering 107(3): 225-229.

Kaçar B, Gaucher EA 2013. Experimental evolution of protein-protein interaction networks. Biochemical Journal 453(3): 311-319.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647-1649.

Keating GM 2013. Fosfomycin trometamol: A review of its use as a single-dose oral treatment for patients with acute lower urinary tract infections and pregnant women with asymptomatic bacteriuria. Drugs 73(17): 1951-1966.

Kedar GC, Brown-Driver V, Reyes DR, Hilgers MT, Stidham MA, Shaw KJ, Finn J, Haselbeck RJ 2008. Comparison of the essential cellular functions of the two *murA* genes of *Bacillus anthracis*. Antimicrobial Agents and Chemotherapy 52(6): 2009-2013.

Kim DH, Lees WJ, Haley TM, Walsh CT 1995a. Kinetic characterization of the inactivation of UDP-GlcNAc enolpyruvyl transferase by (*Z*)-3-fluorophosphoenolpyruvate: Evidence for two oxocarbenium ion intermediates in enolpyruvyl transfer catalysis. Journal of the American Chemical Society 117(5): 1494-1502.

Kim DH, Lees WJ, Kempsell KE, Lane WS, Duncan K, Walsh CT 1996a. Characterization of a Cys115 to Asp substitution in the *Escherichia coli* cell wall biosynthetic enzyme UDP-GlcNAc enolpyruvyl transferase (MurA) that confers resistance to inactivation by the antibiotic fosfomycin. Biochemistry 35(15): 4923-4928.

Kim DH, Lees WJ, Walsh CT 1994. Formation of two enzyme-bound reaction intermediate analogs during inactivation of UDP-GlcNAc enolpyruvyl transferase by (*E*)-or (*Z*)-3-fluorophosphoenolpyruvate. Journal of the American Chemical Society 116(14): 6478-6479.

Kim DH, Lees WJ, Walsh CT 1995b. Stereochemical analysis of the tetrahedral adduct formed at the active site of UDP-GlcNAc enolpyruvyl transferase from the pseudosubstrates,(*E*)-and (*Z*)-3-fluorophosphoenolpyruvate, in D_2O. Journal of the American Chemical Society 117(23): 6380-6381.

Kim DH, Tucker-Kellogg GW, Lees WJ, Walsh CT 1996b. Analysis of fluoromethyl group chirality establishes a common stereochemical course for the enolpyruvyl transfers catalyzed by EPSP synthase and UDP-GlcNAc enolpyruvyl transferase. Biochemistry 35(17): 5435-5440.

Kock H, Gerth U, Hecker M 2004. MurAA, catalysing the first committed step in peptidoglycan biosynthesis, is a target of Clp-dependent proteolysis in *Bacillus subtilis*. Molecular Microbiology 51(4): 1087-1102.

Krekel F, Oecking C, Amrhein N, Macheroux P 1999. Substrate and inhibitor-induced conformational changes in the structurally related enzymes UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS). Biochemistry 38(28): 8864-8878.

Krishnan NM, Seligmann H, Stewart C-B, de Koning APJ, Pollock DD 2004. Ancestral sequence reconstruction in primate mitochondrial DNA: Compositional bias and effect on functional inference. Molecular Biology and Evolution 21(10): 1871-1883.

Krissinel E, Henrick K 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallographica Section D: Biological Crystallography 60(12): 2256-2268.

Krissinel E, Henrick K 2007. Inference of macromolecular assemblies from crystalline state. Journal of Molecular Biology 372(3): 774-797.

Kumar S, Filipski A 2007. Multiple sequence alignment: In pursuit of homologous DNA positions. Genome Research 17(2): 127-135.

Lanzetta PA, Alvarez LJ, Reinach PS, Candia OA 1979. An improved assay for nanomole amounts of inorganic phosphate. Analytical Biochemistry 100(1): 95-97.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23(21): 2947-2948.

Laskowski RA 2001. PDBsum: Summaries and analyses of PDB structures. Nucleic Acids Research 29(1): 221-222.

Lauer B, Süssmuth R, Kaiser D, Jung G, Bormann C 2000. A putative enolpyruvyl transferase gene involved in nikkomycin biosynthesis. The Journal of Antibiotics 53(4): 385-392.

Le SQ, Gascuel O 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution 25(7): 1307-1320.

Leslie AGW, Powell HR 2007. Processing diffraction data with MOSFLM. In: Evolving Methods for Macromolecular Crystallography. Dordrecht, Netherlands, Springer. Pp. 41-51.

Levin JG, Sprinson DB 1964. The enzymatic formation and isolation of 3-enolpyruvylshikimate 5-phosphate. Journal of Biological Chemistry 239(4): 1142-1150.

Lewendon A, Coggins JR 1983. Purification of 5-enolpyruvylshikimate 3-phosphate synthase from *Escherichia coli*. Biochemical Journal 213: 187-191.

Li G, Steel M, Zhang L 2008. More taxa are not necessarily better for the reconstruction of ancestral character states. Systematic Biology 57(4): 647-653.

Livermore DM 2004. The need for new antibiotics. Clinical Microbiology and Infection 10(s4): 1-9.

Maeda H, Dudareva N 2012. The shikimate pathway and aromatic amino acid biosynthesis in plants. Annual Review of Plant Biology 63: 73-105.

Majumder K, Selvapandiyan A, Fattah FA, Arora N, Ahmad S, Bhatnagar RK 1995. 5-enolpyruvylshikimate-3-phosphate synthase of *Bacillus subtilis* is an allosteric enzyme. European Journal of Biochemistry 229(1): 99-106.

Marquardt JL, Brown ED, Lane WS, Haley TM, Ichikawa Y, Wong C-H, Walsh CT 1994. Kinetics, stoichiometry, and identification of the reactive thiolate in the inactivation of UDP-GlcNAc enolpyruvoyl transferase by the antibiotic fosfomycin. Biochemistry 33(35): 10646-10651.

Marquardt JL, Brown ED, Walsh CT, Anderson KS 1993. Isolation and structural elucidation of a tetrahedral intermediate in the UDP-*N*-acetylglucosamine enolpyruvoyl transferase enzymic pathway. Journal of the American Chemical Society 115(22): 10398-10399.

Marquardt JL, Siegele DA, Kolter R, Walsh CT 1992. Cloning and sequencing of *Escherichia coli murZ* and purification of its product, a UDP-*N*-acetylglucosamine enolpyruvyl transferase. Journal of Bacteriology 174(17): 5748-5752.

Mascaros PA, Bazzicalupi C, Bianchi A, Giorgi C, Valero MDG, Garzón RL, Salido MLG, Valtancoli B 2011. Molecular recognition of ADP over ATP in aqueous solution by a polyammonium receptor containing a pyrimidine residue. Chemical Communications 47(10): 2814-2816.

Matthews BW 1968. Solvent content of protein crystals. Journal of Molecular Biology 33(2): 491-497.

McConkey GA 1999. Targeting the shikimate pathway in the malaria parasite *Plasmodium falciparum*. Antimicrobial Agents and Chemotherapy 43(1): 175-177.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ 2007. Phaser crystallographic software. Journal of Applied Crystallography 40(4): 658-674.

McDevitt D, Payne DJ, Holmes DJ, Rosenberg M 2002. Novel targets for the future development of antibacterial agents. Journal of Applied Microbiology 92(S1): 28S-34S.

Millar G, Lewendon A, Hunter MG, Coggins JR 1986. The cloning and expression of the *aroL* gene from *Escherichia coli* K12. Purification and complete amino acid sequence of shikimate kinase II, the *aroL*-gene product. Biochemical Journal 237: 427-437.

Mizyed S, Wright JEI, Byczynski B, Berti PJ 2003. Identification of the catalytic residues of AroA (enol pyruvylshikimate 3-phosphate synthase) using partitioning analysis. Biochemistry 42(23): 6986-6995.

Molina-López J, Sanschagrin F, Levesque RC 2006. A peptide inhibitor of MurA UDP-*N*-acetylglucosamine enolpyruvyl transferase: The first committed step in peptidoglycan biosynthesis. Peptides 27(12): 3115-3121.

Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography 67(4): 355-367.

Nakamura LK 1998. *Bacillus pseudomycoides* sp. nov. International Journal of Systematic Bacteriology 48(3): 1031-1035.

Nakamura LK, Jackson MA 1995. Clarification of the taxonomy of *Bacillus mycoides*. International Journal of Systematic Bacteriology 45(1): 46-49.

National Institutes of Health 2007. Understanding emerging and re-emerging infectious diseases. In: Biological Sciences Curriculum Study. Bethesda, MD. Retrieved December, 2014, from http://www.ncbi.nlm.nih.gov/books/NBK20370/.

Nicholas HB, Arst HN, Caddick MX 2001. Evaluating low level sequence identities. European Journal of Biochemistry 268(2): 414-419.

Nielsen P, Fritze D, Priest FG 1995. Phenetic diversity of alkaliphilic *Bacillus* strains: Proposal for nine new species. Microbiology 141(7): 1745-1761.

Nobeli I, Favia AD, Thornton JM 2009. Protein promiscuity and its implications for biotechnology. Nature Biotechnology 27(2): 157-167.

Nock S, Grillenbeck N, Ahmadian MR, Ribeiro S, Kreutzer R, Sprinzl M 1995. Properties of isolated domains of the elongation factor Tu from *Thermus thermophilus* HB8. European Journal of Biochemistry 234(1): 132-139.

Nogi Y, Takami H, Horikoshi K 2005. Characterization of alkaliphilic *Bacillus* strains used in industry: Proposal of five novel species. International Journal of Systematic and Evolutionary Microbiology 55(6): 2309-2315.

Oakley TH, Cunningham CW 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. Evolution 54(2): 397-405.

Oberdorfer G, Binter A, Ginj C, Macheroux P, Gruber K 2012. Structural and functional characterization of NikO, an enolpyruvyl transferase essential in nikkomycin biosynthesis. Journal of Biological Chemistry 287(37): 31427-31436.

Okunuki H, Akiyama H, Teshima R, Hino A, Goda Y, Sawada J, Toyoda M, Maitani T 2003. Determination of enzymatic activity of 5-enolpyruvylshikimate-3-phosphate synthase by LC/MS. Journal of the Food Hygienic Society of Japan 44(2): 77-82.

Oliveira JS, Mendes MA, Palma MS, Basso LA, Santos DS 2003. One-step purification of 5-enolpyruvylshikimate-3-phosphate synthase enzyme from

Mycobacterium tuberculosis. Protein Expression and Purification 28(2): 287-292.

Oliveira JS, Pinto CA, Basso LA, Santos DS 2001. Cloning and overexpression in soluble form of functional shikimate kinase and 5-enolpyruvylshikimate 3-phosphate synthase enzymes from *Mycobacterium tuberculosis*. Protein Expression and Purification 22(3): 430-435.

Pace NR 1991. Origin of life-facing up to the physical setting. Cell 65(4): 531-533.

Padgette SR, Huynh QK, Borgmeyer J, Shah DM, Brand LA, Re DB, Bishop BF, Rogers SG, Fraley RT, Kishore GM 1987. Bacterial expression and isolation of *Petunia hybrida* 5-enol-pyruvylshikimate-3-phosphate synthase. Archives of Biochemistry and Biophysics 258(2): 564-573.

Padgette SR, Re DB, Gasser CS, Eichholtz DA, Frazier RB, Hironaka CM, Levine EB, Shah DM, Fraley RT, Kishore GM 1991. Site-directed mutagenesis of a conserved region of the 5-enolpyruvylshikimate-3-phosphate synthase active site. Journal of Biological Chemistry 266(33): 22364-22369.

Pagel M, Meade A, Barker D 2004. Bayesian estimation of ancestral character states on phylogenies. Systematic Biology 53(5): 673-684.

Parish T, Stoker NG 2002. The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. Microbiology 148(10): 3069-3077.

Park H, Hilsenbeck JL, Kim HJ, Shuttleworth WA, Park YH, Evans JN, Kang C 2004. Structural studies of *Streptococcus pneumoniae* EPSP synthase in unliganded state, tetrahedral intermediate-bound state and S3P-GLP-bound state. Molecular Microbiology 51(4): 963-971.

Pauling L, Zuckerkandl E 1963. Chemical paleogenetics. Molecular "Restoration Studies" of extinct forms of life. Acta Chemica Scandinavica 17: S9-S16.

Pedroso MM, Ely F, Basso LA, Santos DS n.d. The Enzymatic Synthesis of Shikimate-3-phosphate. http://www.pucrs.br/research/salao/2006-VIISalaoIC/Arquivos2006/CienciasBiologicas/37105%20-%20MARCELO%20MONTEIRO%20PEDROSO.pdf (accessed June 2014).

Perez-Jimenez R, Inglés-Prieto A, Zhao Z-M, Sanchez-Romero I, Alegre-Cebollada J, Kosuri P, Garcia-Manyes S, Kappock TJ, Tanokura M, Holmgren A 2011. Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nature Structural & Molecular Biology 18(5): 592-596.

Posada D 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution 25(7): 1253-1256.

Priestman MA, Funke T, Singh IM, Crupper SS, Schönbrunn E 2005a. 5-enolpyruvylshikimate-3-phosphate synthase from *Staphylococcus aureus* is insensitive to glyphosate. FEBS Letters 579(3): 728-732.

Priestman MA, Healy ML, Funke T, Becker A, Schönbrunn E 2005b. Molecular basis for the glyphosate-insensitivity of the reaction of 5-enolpyruvylshikimate 3-phosphate synthase with shikimate. FEBS Letters 579(25): 5773-5780.

Ramasamy D, Lagier J-C, Gorlas A, Raoult D, Fournier P-E 2013. Non contiguous-finished genome sequence and description of *Bacillus massiliosenegalensis* sp. nov. Standards in Genomic Sciences 8(2): 264.

Rambaut A 2009. Molecular Evolution, Phylogenetics and Epidemiology. http://tree.bio.ed.ac.uk/software/figtree/ (accessed June 2015).

Rannala B, Yang Z 1996. Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. Journal of Molecular Evolution 43(3): 304-311.

Rice P, Longden I, Bleasby A 2000. EMBOSS: The European molecular biology open software suite. Trends in Genetics 16(6): 276-277.

Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW 2006. Evolutionary origins of the eukaryotic shikimate pathway: Gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryotic Cell 5(9): 1517-1531.

Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian β-lactamases. Journal of the American Chemical Society 135: 2899-2902.

Saitou N 1989. A theoretical study of the underestimation of branch lengths by the maximum parsimony principle. Systematic Biology 38(1): 1-6.

Salemi M, Vandamme A-M eds. 2003. The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. Cambridge, U.K., Cambridge University Press.

Salisbury BA, Kim J 2001. Ancestral state estimation and taxon sampling density. Systematic Biology 50(4): 557-564.

Samland AK, Etezady-Esfarjani T, Amrhein N, Macheroux P 2001. Asparagine 23 and aspartate 305 are essential residues in the active site of UDP-*N*-acetylglucosamine enolpyruvyl transferase from *Enterobacter cloacae*. Biochemistry 40(6): 1550-1559.

Sanangelantoni AM, Cammarano P, Tiboni O 1996. Manipulation of the *tuf* gene provides clues to the localization of sequence element (s) involved in the thermal stability of *Thermotoga maritima* elongation factor Tu. Microbiology 142(9): 2525-2532.

Sanderson MJ 2003. r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19(2): 301-302.

Sanson GFO, Kawashita SY, Brunstein A, Briones MRS 2002. Experimental phylogeny of neutrally evolving DNA sequences generated by a bifurcate series of nested polymerase chain reactions. Molecular Biology and Evolution 19(2): 170-178.

Schönbrunn E, Eschenburg S, Krekel F, Luger K, Amrhein N 2000. Role of the loop containing residue 115 in the induced-fit mechanism of the bacterial cell wall biosynthetic enzyme MurA. Biochemistry 39(9): 2164-2173.

Schönbrunn E, Eschenburg S, Shuttleworth WA, Schloss JV, Amrhein N, Evans JNS, Kabsch W 2001. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proceedings of the National Academy of Sciences 98(4): 1376-1380.

Schönbrunn E, Sack S, Eschenburg S, Perrakis A, Krekel F, Amrhein N, Mandelkow E 1996. Crystal structure of UDP-*N*-acetylglucosamine enolpyruvyltransferase, the target of the antibiotic fosfomycin. Structure 4(9): 1065-1075.

Schönbrunn E, Svergun DI, Amrhein N, Koch MHJ 1998. Studies on the conformational changes in the bacterial cell wall biosynthetic enzyme UDP-*N*-acetylglucosamine enolpyruvyltransferase (MurA). European Journal of Biochemistry 253(2): 406-412.

Schrödinger 2010. The PyMOL Molecular Graphics System [Computer software].

Schultz TR, Churchill GA 1999. The role of subjectivity in reconstructing ancestral character states: A Bayesian approach to unknown rates, states, and transformation asymmetries. Systematic Biology: 651-664.

Schulz A, Krüper A, Amrhein N 1985. Differential sensitivity of bacterial 5-enolpyruvylshikimate-3-phosphate synthases to the herbicide glyphosate. FEMS Microbiology Letters 28(3): 297-301.

Shahab M, Verma M, Pathak M, Mitra K, Misra-Bhattacharya S 2014. Cloning, expression and characterization of UDP-*N*-acetylglucosamine enolpyruvyl transferase (MurA) from *Wolbachia* endosymbiont of human lymphatic filarial parasite *Brugia malayi*. PLOS ONE 9(6): e99884.

Shuttleworth WA, Evans JNS 1994. Site-directed mutagenesis and NMR studies of histidine 385 mutants of 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase. Biochemistry 33(23): 7062-7068.

Shuttleworth WA, Pohl ME, Helms GL, Jakeman DL, Evans JNS 1999. Site-directed mutagenesis of putative active site residues of 5-enolpyruvylshikimate-3-phosphate synthase. Biochemistry 38(1): 296-302.

Silver LL 2007. Multi-targeting by monotherapeutic antibacterials. Nature Reviews Drug Discovery 6(1): 41-55.

Skarzynski T, Mistry A, Wonacott A, Hutchinson SE, Kelly VA, Duncan K 1996. Structure of UDP-*N*-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-*N*-acetylglucosamine and the drug fosfomycin. Structure 4(12): 1465-1474.

Stallings WC, Abdel-Meguid SS, Lim LW, Shieh H-S, Dayringer HE, Leimgruber NK, Stegeman RA, Anderson KS, Sikorski JA, Padgette SR, Kishore GM 1991. Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase: A distinctive protein fold. Proceedings of the National Academy of Sciences of the United States of America 88: 5046-5050.

Stauffer ME, Young JK, Evans JNS 2001. Shikimate-3-phosphate binds to the isolated N-terminal domain of 5-enolpyruvylshikimate-3-phosphate synthase. Biochemistry 40(13): 3951-3957.

Steinrücken HC, Amrhein N 1984a. 5-enolpyruvylshikimate-3-phosphate synthase of *Klebsiella pneumoniae*. European Journal of Biochemistry 143(2): 341-349.

Steinrücken HC, Amrhein N 1984b. 5-enolpyruvylshikimate-3-phosphate synthase of *Klebsiella pneumoniae.* 2. Inhibition by glyphosate [N-(phosphonomethyl)glycine]. European Journal of Biochemistry 143(2): 351-7.

Stetter KO 1996. Hyperthermophiles in the history of life. Proceedings of the Ciba Foundation Symposium 202-Evolution of Hydrothermal Ecosystems on Earth (And Mars?), Wiley Online Library. Pp. 1-23.

Sun Y-C, Chen Y-C, Tian Z-X, Li F-M, Wang X-Y, Zhang J, Xiao Z-L, Lin M, Gilmartin N, Dowling DN 2005. Novel AroA with high tolerance to glyphosate, encoded by a gene of *Pseudomonas putida* 4G-1 isolated from an extremely polluted environment in China. Applied and Environmental Microbiology 71(8): 4771-4776.

Szilágyi A, Závodszky P 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. Structure 8(5): 493-504.

Takahata S, Ida T, Hiraishi T, Sakakibara S, Maebashi K, Terada S, Muratani T, Matsumoto T, Nakahama C, Tomono K 2010. Molecular mechanisms of fosfomycin resistance in clinical isolates of *Escherichia coli*. International Journal of Antimicrobial Agents 35(4): 333-337.

Tavaré S 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences 17: 57-86.

Taylor TJ, Vaisman II 2010. Discrimination of thermophilic and mesophilic proteins. BMC Structural Biology 10(Suppl 1): S5.

Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, Read RJ, Adams PD 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallographica Section D: Biological Crystallography 64(1): 61-69.

Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV 2000. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. European Journal of Biochemistry 267(17): 5313-5329.

Thornton JW 2004. Resurrecting ancient genes: Experimental analysis of extinct molecules. Nature Reviews Genetics 5: 366-375.

Tian Y-S, Xu J, Han J, Zhao W, Fu X-Y, Peng R-H, Yao Q-H 2013. Complementary screening, identification and application of a novel class II 5-enopyruvylshikimate-3-phosphate synthase from *Bacillus cereus*. World Journal of Microbiology and Biotechnology 29(3): 549-557.

Tian Y-S, Xu J, Xiong A-S, Zhao W, Gao F, Fu X-Y, Peng R-H, Yao Q-H 2012. Functional characterization of class II 5-enopyruvylshikimate-3-phosphate synthase from *Halothermothrix orenii* H168 in *Escherichia coli* and transgenic *Arabidopsis*. Applied Microbiology and Biotechnology 93(1): 241-250.

Tzin V, Galili G 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Molecular Plant 3(6): 956-972.

Vitkup D, Kharchenko P, Wagner A 2006. Influence of metabolic network structure and function on enzyme evolution. Genome Biology 7(5): R39.

Vogt G, Woell S, Argos P 1997. Protein thermal stability, hydrogen bonds, and ion pairs. Journal of Molecular Biology 269(4): 631-643.

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. PLOS Biology 10: e1001446.

Walsh CT, Benson TE, Kim DH, Lees WJ 1996. The versatility of phosphoenolpyruvate and its vinyl ether products in biosynthesis. Chemistry & Biology 3(2): 83-91.

Wanke C, Amrhein N 1993. Evidence that the reaction of the UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase proceeds through the O-phosphothioketal of pyruvic acid bound to Cys115 of the enzyme. European Journal of Biochemistry 218(3): 861-870.

Wanke C, Falchetto R, Amrhein N 1992. The UDP-*N*-acetylglucosamine 1-carboxyvinyl-transferase of *Enterobacter cloacae* Molecular cloning, sequencing of the gene and overexpression of the enzyme. FEBS Letters 301(3): 271-276.

Williams PD, Pollock DD, Blackburne BP, Goldstein RA 2006. Assessing the accuracy of ancestral protein reconstruction methods. PLOS Computational Biology 2(6): 598-605.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A 2011. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D: Biological Crystallography 67(4): 235-242.

Woese C 1998. The universal ancestor. Proceedings of the National Academy of Sciences 95(12): 6854-6859.

Woese CR 1987. Bacterial evolution. Microbiological Reviews 51(2): 221.

Yang Z 2006. Computational Molecular Evolution. Oxford, U.K., Oxford University Press.

Yang Z 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24(8): 1586-1591.

Yang Z, Kumar S, Nei M 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141(4): 1641-1650.

Yang Z, Rannala B 1997. Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. Molecular Biology and Evolution 14(7): 717-724.

Yano JK, Poulos TL 2003. New understandings of thermostable and peizostable enzymes. Current Opinion in Biotechnology 14(4): 360-365.

Yoon H-J, Lee SJ, Mikami B, Park H-J, Yoo J, Suh SW 2008. Crystal structure of UDP-*N*-acetylglucosamine enolpyruvyl transferase from *Haemophilus influenzae* in complex with UDP-*N*-acetylglucosamine and fosfomycin. Proteins: Structure, Function, and Bioinformatics 71(2): 1032-1037.

Yu L, Fink BD, Herlein JA, Sivitz WI 2013. Mitochondrial function in diabetes: Novel methodology and new insight. Diabetes 62(6): 1833-1842.

Zhang F, Berti PJ 2006. Phosphate analogues as probes of the catalytic mechanisms of MurA and AroA, two carboxyvinyl transferases. Biochemistry 45(19): 6027-6037.

Zhang J, Nei M 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. Journal of Molecular Evolution 44(1): S139-S146.

Zhang Y, Yi L, Lin Y, Zhang L, Shao Z, Liu Z 2014. Characterization and site-directed mutagenesis of a novel class II 5-enopyruvylshikimate-3-phosphate (EPSP) synthase from the deep-sea bacterium *Alcanivorax* sp. L27. Enzyme and Microbial Technology:

Zhi X-Y, Yao J-C, Li H-W, Huang Y, Li W-J 2014. Genome-wide identification, domain architectures and phylogenetic analysis provide new insights into

the early evolution of shikimate pathway in prokaryotes. Molecular Phylogenetics and Evolution 75: 154-164.

Zhou X-X, Wang Y-B, Pan Y-J, Li W-F 2008. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. Amino Acids 34(1): 25-33.

Zhu J-Y, Yang Y, Han H, Betzi S, Olesen SH, Marsilio F, Schönbrunn E 2012. Functional consequence of covalent reaction of phosphoenolpyruvate with UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase (MurA). Journal of Biological Chemistry 287(16): 12657-12667.

Zwickl DJ 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Unpublished PhD thesis, University of Texas, Austin, TX.

Appendices

Appendix A: Accession Numbers and species details

Table A.1: Bacterial strains and accession numbers of MurA used in ancestral inference

Organism	Strain/serovar	Accession number
Bacillus amyloliquefaciens	DSM 7	NC_014551
Bacillus anthracis	Sterne	NC_005945
Bacillus atrophaeus	1942	NC_014639
Bacillus cellulosilyticus	DSM 2522	NC_014829
Bacillus cereus	ATCC 14579	AE016877
Bacillus clausii	KSM-K16	NC_006582
Bacillus coagulans	36D1	NC_016023
Bacillus coahuilensis	m4-4	NZ_ABFU01000064
Bacillus cytotoxicus	NVH 391-98	NC_009674
Bacillus halodurans	C-125	NC_002570
Bacillus licheniformis	ATCC 14580	NC_006270
Bacillus megaterium	DSM 319	CP001982
Bacillus mycoides	Rock3-17	NZ_ACMW01000102
Bacillus pseudofirmus	OF4	NC_013791
Bacillus pseudomycoides	DSM 12442	NZ_ACMX01000083
Bacillus pumilus	ATCC 7061	NZ_ABRX01000004
Bacillus selenitireducens	MLS10	NC_014219
Bacillus subtilis	168	This study
Bacillus thuringiensis	97-27	NC_005957
Bacillus weihenstephanensis	KBAB4	NC_010184
Clostridium acetobutylicum	ATCC 824	NC_003030
Clostridium butyricum	5521	NZ_ABDT01000114

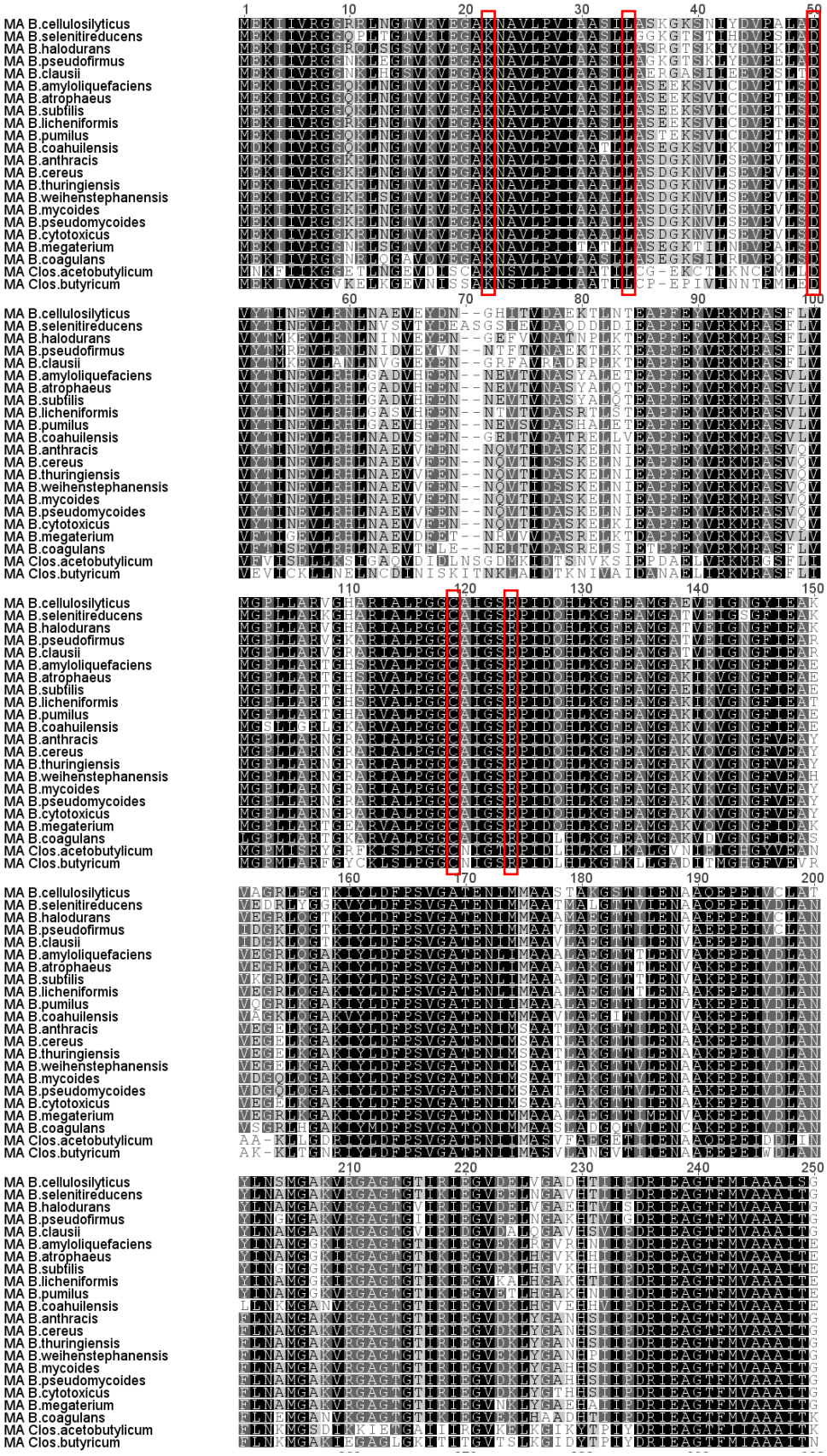
Table A.2: Bacterial strains and accession numbers of AroA used in ancestral inference

Organism	Strain/serovar	Accession number
Bacillus amyloliquefaciens	DSM7	NC_014551
Bacillus anthracis	Ames	NC_003997
Bacillus atrophaeus	1942	NC_014639
Bacillus cellulosilyticus	DSM 2522	NC_014829
Bacillus cereus	ATCC 14579	NC_004722
Bacillus clausii	KSM-K16	NC_006582
Bacillus coagulans	36D1	NC_016023
Bacillus coahuilensis	m4-4	NZ_ABFU01000037
Bacillus cytotoxicus	NVH 391-98	NC_009674
Bacillus halodurans	C-125	NC_002570
Bacillus licheniformis	ATCC 14580	NC_006322
Bacillus megaterium	DSM319	NC_014103
Bacillus mycoides	Rock3-17	NZ_ACMW01000069
Bacillus pseudofirmus	OF4	NC_013791
Bacillus pseudomycoides	DSM 12442	NZ_ACMX01000055
Bacillus pumilus	ATCC 7061	NZ_ABRX01000005
Bacillus selenitireducens	MLS10	NC_014219
Bacillus subtilis	168	This study
Bacillus thuringiensis	Al Hakam	NC_008600
Bacillus weihenstephanensis	KBAB4	NC_010184
Clostridium acetobutylicum	ATCC 824	NC_003030
Clostridium butyricum	5521	NZ_ABDT01000092

Table A.3: Bacterial species and accession numbers of class I and class II AroA

Class	Organism	Accession number
I	*Aeromonas salmonicida*	Q03321
II	*Agrobacterium tumefaciens. CP4*	Q9R4E4
II	*Bacillus cereus*	NP_832685
II	*Bacillus subtilis*	P20691
I	*Bordetella pertussis*	P12421
I	*Escherichia coli*	P0A6D3
II	*Halothermothrix orenii. H168*	ACL69793
II	*Pseudomonas sp. PG2982*	P0A2Y4
I	*Salmonella typhimurium*	CAA71382
II	*Staphylococcus aureus*	Q05615
II	*Streptococcus pneumoniae*	Q9S400

Appendix B: Protein sequence alignment of Bacillus MurA and Bacillus AroA



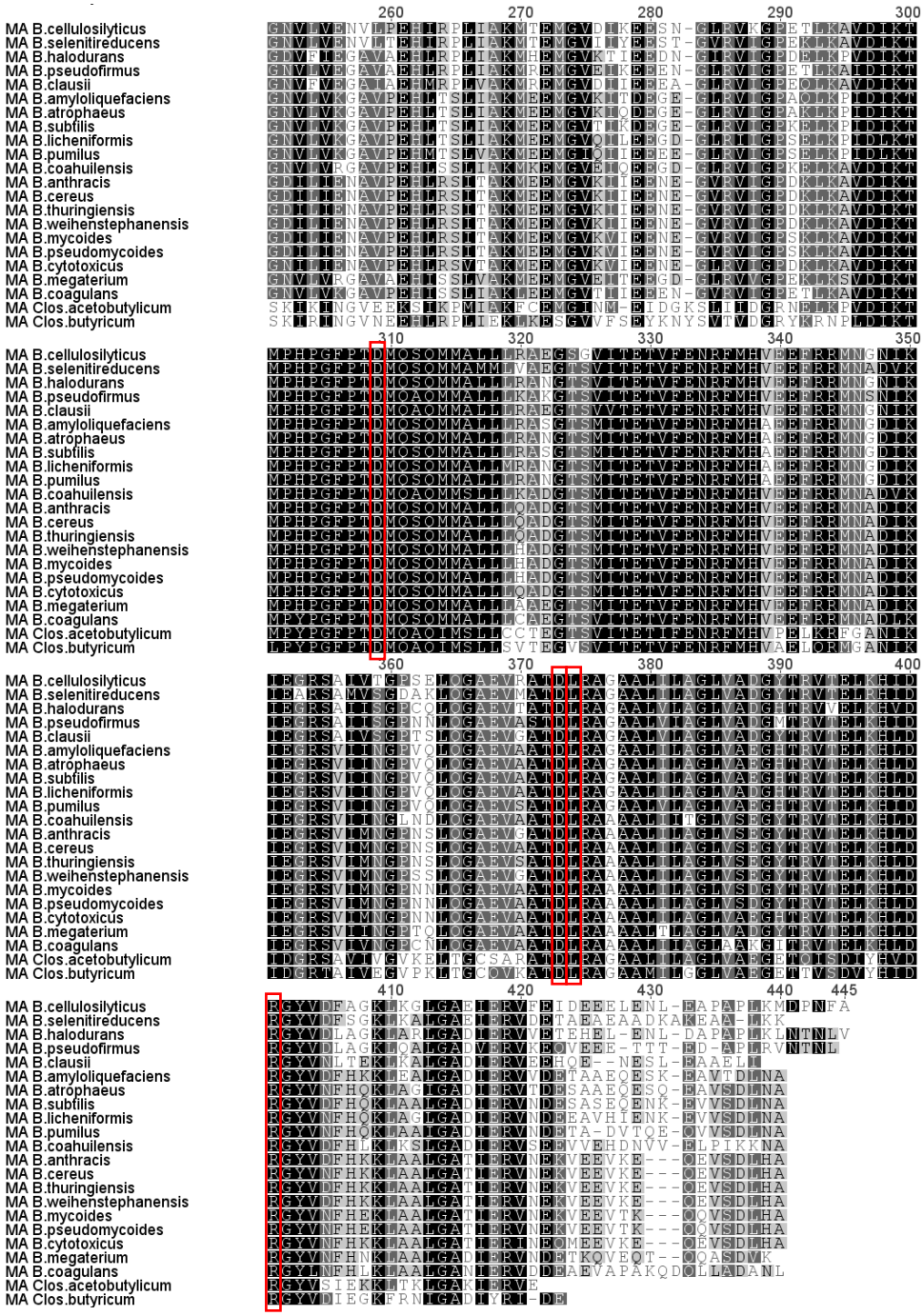
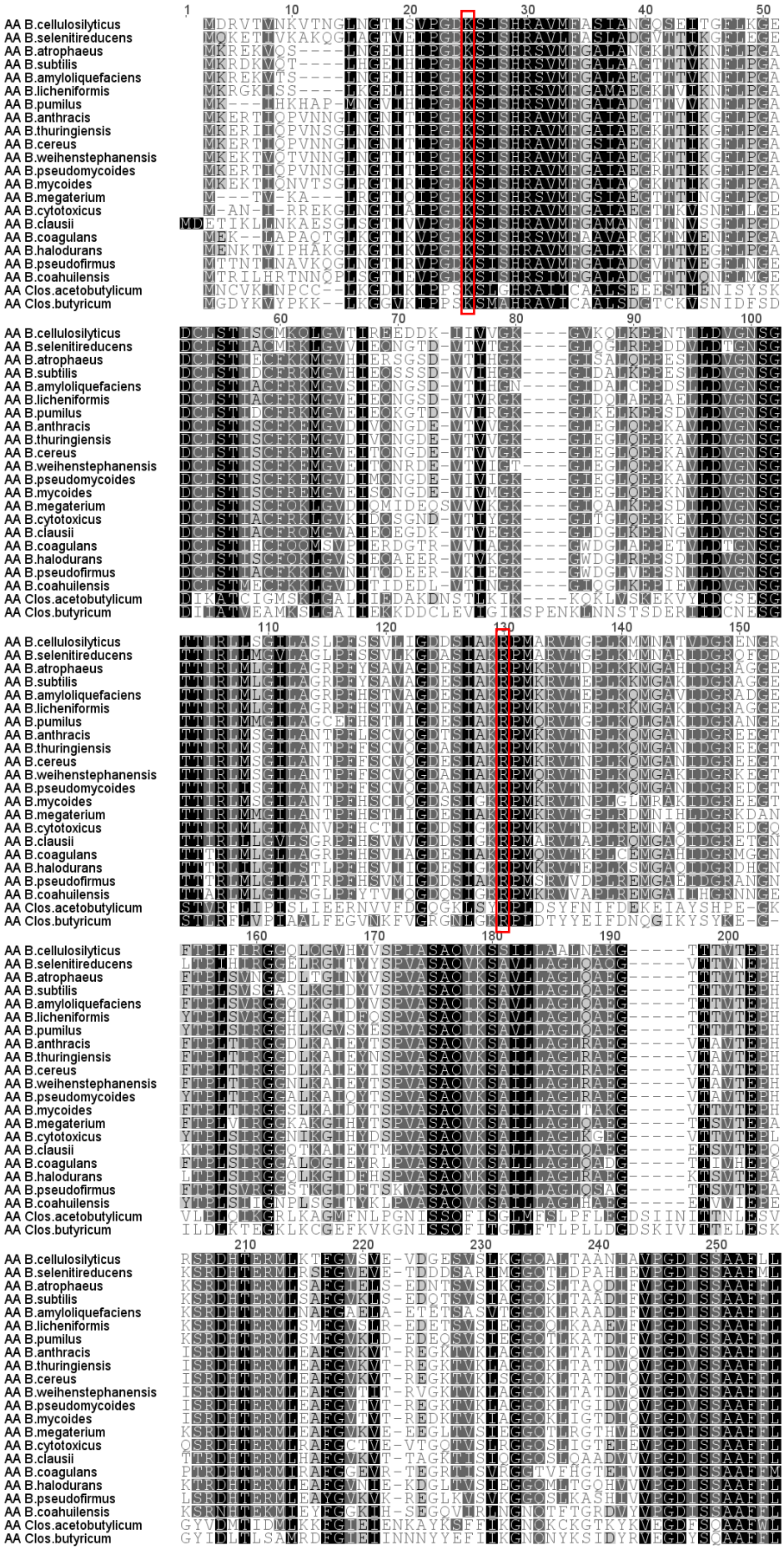


Figure A.1: *Bacillus* MurA protein sequence alignment
The different shades indicate the percentage similarity at each position in the alignment. Black = 100% similar, dark grey = 80 to 100% similar, light grey = 60 to 80% similar and white = less than 60% similar. Alignment was generated with ClustalW {Larkin, 2007 #222} in Geneious {Kearse, 2012 #221}. Conserved residues mentioned in section 3.2.1.1 are shown in red boxes.


AA B.cellulosilyticus
AA B.selenitireducens
AA B.atrophaeus
AA B.subtilis
AA B.amyloliquefaciens
AA B.licheniformis
AA B.pumilus
AA B.anthracis
AA B.thuringiensis
AA B.cereus
AA B.weihenstephanensis
AA B.pseudomycoides
AA B.mycoides
AA B.megaterium
AA B.cytotoxicus
AA B.clausii
AA B.coagulans
AA B.halodurans
AA B.pseudofirmus
AA B.coahuilensis
AA Clos.acetobutylicum
AA Clos.butyricum

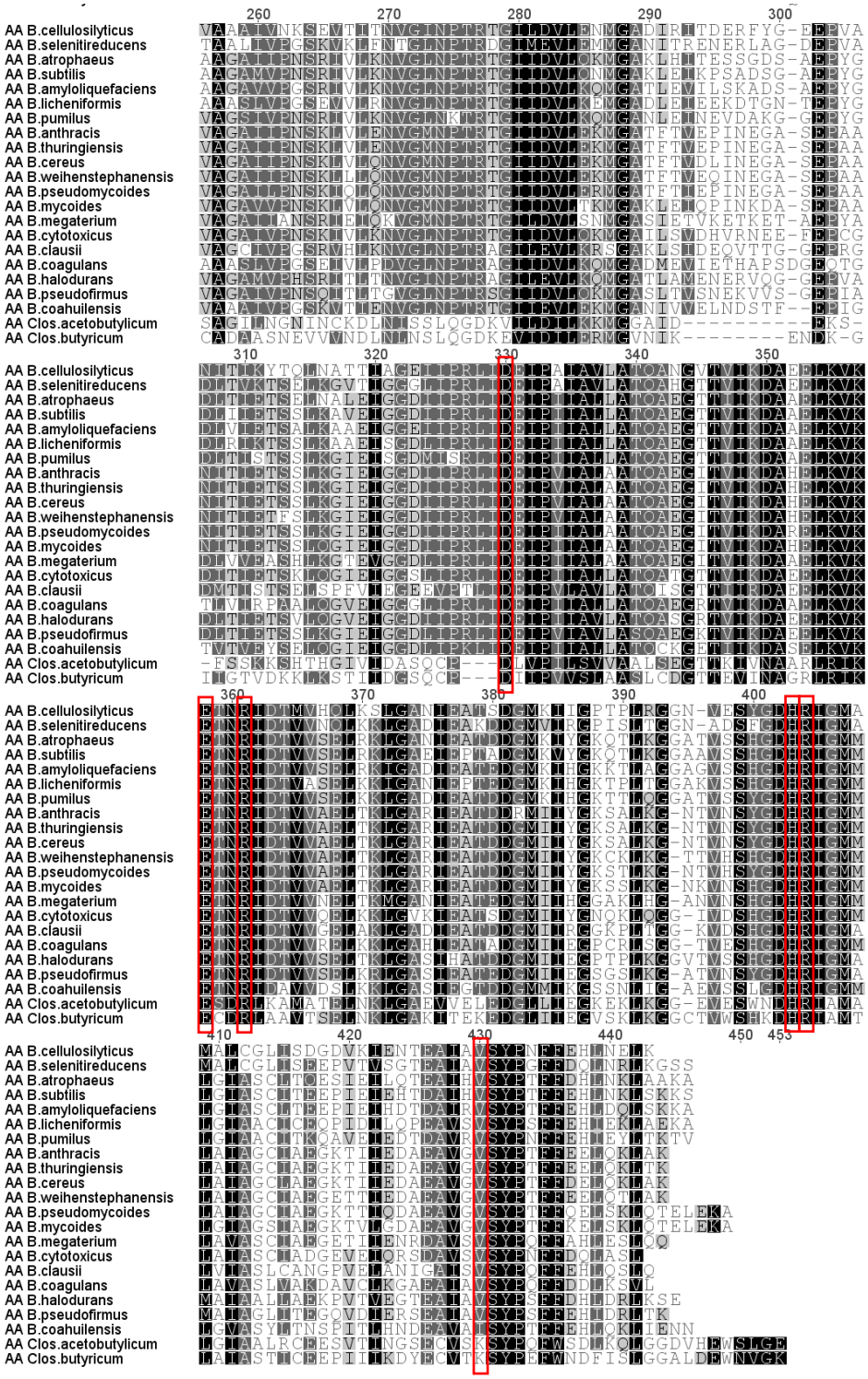


Figure A.2: *Bacillus* AroA protein sequence alignment
The different shades indicate the percentage similarity at each position in the alignment. Black = 100% similar, dark grey = 80 to 100% similar, light grey = 60 to 80% similar and white = less than 60% similar. Alignment was generated with ClustalW {Larkin, 2007 #222} in Geneious {Kearse, 2012 #221}. Conserved residues mentioned in section 4.2.1.1 are shown in red boxes.

Appendix C: Nucleotide and amino acid sequence information

Bacillus subtilis murA (1311 bp)

TTGGAAAAAATCATCGTCCGCGGCGGTCAGAAGTTAAACGGCACAGTCAAAGTTG
AAGGCGCTAAAAATGCCGTTTTACCTGTTATCGCTGCATCTTTATTAGCAAGTGA
AGAAAAAAGCGTAATTTGTGATGTACCTACGCTCTCCGATGTATATACAATTAAC
GAAGTGTTGCGTCATTTAGGAGCAGATGTGCATTTTGAAAATAATGAAGTGACTG
TAAATGCTTCATACGCTTTGCAAACTGAAGCACCTTTTGAATATGTTCGTAAAAT
GCGTGCGTCTGTGCTTGTCATGGGGCCGCTTCTTGCGCGTACAGGTCATGCAAGA
GTTGCACTTCCGGGCGGATGCGCAATTGGTTCCAGACCGATTGATCAGCATTTAA
AAGGTTTTGAAGCAATGGGCGCAGAAATCAAAGTCGGTAATGGCTTCATTGAAGC
TGAAGTAAAAGGCCGACTGCAAGGCGCAAAAATTTATCTGGACTTCCCAAGTGTA
GGAGCTACAGAGAACCTGATTATGGCAGCCGCTCTAGCTGAAGGAACAACAACGC
TGGAAAACGTGGCAAAAGAACCCGAAATCGTTGATTTAGCAAACTATATCAACGG
CATGGGCGGAAAAATCCGCGGAGCTGGCACCGGCACCATCAAAATTGAAGGAGTC
GAAAAGCTTCACGGCGTAAAACACCATATTATTCCTGACCGTATTGAAGCGGGCA
CATTTATGGTTGCTGCTGCAATCACTGAAGGAAACGTATTAGTAAAAGGAGCGGT
TCCTGAGCACCTCACCTCTTTAATTGCAAAAATGGAAGAGATGGGTGTAACAATT
AAGGATGAAGGTGAAGGTCTGCGTGTCATCGGCCCGAAAGAGCTTAAACCGATTG
ACATCAAAACAATGCCTCACCCGGGCTTCCCGACTGATATGCAGTCACAAATGAT
GGCGCTTCTGCTTCGTGCAAGCGGCACAAGCATGATTACAGAAACCGTTTTTGAA
AACCGTTTTATGCATGCGGAAGAATTCCGCCGTATGAATGGTGATATCAAGATTG
AAGGACGTTCTGTCATCATTAACGGTCCTGTACAGCTTCAGGGAGCTGAAGTTGC
AGCGACTGATTTGCGTGCAGGTGCAGCGCTGATTCTTGCGGGGTTAGTGGCTGAA
GGTCACACACGTGTTACTGAATTGAAGCACTTAGACCGCGGTTACGTTGATTTCC
ATCAGAAGCTTGCCGCTCTGGGCGCAGACATCGAACGTGTAAATGATGAGTCTGC
TTCTGAGCAAGAGAATAAAGAAGTCGTTTCTGACTTAAATGCATAA

Bacillus subtilis MurA

MEKIIVRGGQKLNGTVKVEGAKNAVLPVIAASLLASEEKSVICDVPTLSDVYTIN
EVLRHLGADVHFENNEVTVNASYALQTEAPFEYVRKMRASVLVMGPLLARTGHAR
VALPGGCAIGSRPIDQHLKGFEAMGAEIKVGNGFIEAEVKGRLQGAKIYLDFPSV
GATENLIMAAALAEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGV
EKLHGVKHHIIPDRIEAGTFMVAAAITEGNVLVKGAVPEHLTSLIAKMEEMGVTI
KDEGEGLRVIGPKELKPIDIKTMPHPGFPTDMQSQMMALLLRASGTSMITETVFE
NRFMHAEEFRRMNGDIKIEGRSVIINGPVQLQGAEVAATDLRAGAALILAGLVAE
GHTRVTELKHLDRGYVDFHQKLAALGADIERVNDESASEQENKEVVSDLNA

436 amino acids
Predicted molecular weight: 46701 Da
Theoretical pI: 5.55

***Bacillus coagulans murA* (1314 bp)**

ATGGAAAAAATCATTGTGCGTGGTGGTAATCGTCTGCAGGGTGCAGTTCAGGTTG
AAGGTGCAAAAAATGCAGTTCTGCCGGTTATTGCAGCAAGCCTGCTGGCAAGCGA
AGGTAAAAGCATTATTCGTGATGTTCCGCAGCTGAGTGATGTTTTTACCATTAGC
GAAGTTCTGCGTCATCTGAATGCAGAAGTTACCTTTCTGGAAAACGAAATCACCG
TTGATGCAAGCCGTGAACTGAGCATTGAAACCCCGTTTGAATATGTTCGTAAAAT
GCGTGCAAGCTTTCTGGTTCTGGGTCCGCTGCTGGCACGTACCGGTAAAGCACGT
GTTGCACTGCCTGGTGGTTGTGCAATTGGTAGCCGTCCGATTGATCTGCATCTGA
AAGGTTTTGAAGCAATGGGTGCAAAAGTTGATGTGGGCAATGGTTTTATTGAAGC
CAGCGTTAGCGGTCGCCTGCATGGTGCCAAAATTTACATGGATTTTCCGAGCGTT
GGTGCCACCCAGAATATTATGATGGCAGCCAGCCTGGCAGATGGTCAGACCGTTA
TTGAAAATTGTGCAAAAGAACCGGAAATTGTGGATCTGGCCAATTTTCTGAATGA
AATGGGTGCCAATGTTAAAGGTGCAGGCACCGGCACCATTAAAATCGAAGGTGTT
GAAAAACTGCATGCCGCAGATCATACCATTATTCCGGATCGTATTGAAGCAGGCA
CCTTTATGGTTGCAGCAGCAATTACCAAAGGTAATGTTCTGGTGAAAGGTGCCGT
GCCGGAACATATTAGCAGCCTGATTGCAAAACTGGAAGAAATGGGCGTGACCATT
ATTGAAGAAGAAAACGGCGTTCGCGTTATTGGTCCGGAAACCCTGAAAGCAGTGG
ATATCAAAACCATGCCGTATCCGGGTTTTCCGACCGATATGCAGAGCCAGATGAT
GGCACTGCTGCTGTGTGCAGAAGGCACCAGCGTTATTACCGAAACCGTGTTTGAA
AATCGCTTTATGCACGTGGAAGAATTTCGTCGTATGAATGCCGATCTGAAAATTG
AAGGTCGTAGCGTTATTGTTAATGGTCCGTGTAATCTGCAAGGTGCCGAAGTTGC
AGCAACCGATCTGCGTGCAGCAGCCGCACTGATTATTGCAGGTCTGGCAGCAAAA
GGTATTACCCGTGTTACCGAACTGAAACATCTGGATCGTGGTTATCTGAACTTTC
ATCTGAAACTGGCAGCACTGGGTGCGAATATTGAACGTGTTGATGATGAAGCGGA
AGTTGCACCGGCAAAACAGGATCAGCTGCTGGCCGATGCAAATCTGTAA

***Bacillus coagulans* MurA**

MEKIIVRGGNRLQGAVQVEGAKNAVLPVIAASLLASEGKSIIRDVPQLSDVFTIS
EVLRHLNAEVTFLENEITVDASRELSIETPFEYVRKMRASFLVLGPLLARTGKAR
VALPGGCAIGSRPIDLHLKGFEAMGAKVDVGNGFIEASVSGRLHGAKIYMDFPSV
GATQNIMMAASLADGQTVIENCAKEPEIVDLANFLNEMGANVKGAGTGTIKIEGV
EKLHAADHTIIPDRIEAGTFMVAAAITKGNVLVKGAVPEHISSLIAKLEEMGVTI
IEEENGVRVIGPETLKAVDIKTMPYPGFPTDMQSQMMALLLCAEGTSVITETVFE
NRFMHVEEFRRMNADLKIEGRSVIVNGPCNLQGAEVAATDLRAAAALIIAGLAAK
GITRVTELKHLDRGYLNFHLKLAALGANIERVDDEAEVAPAKQDQLLADANL

437 amino acids
Predicted molecular weight: 46649 Da
Theoretical pI: 5.30

***Bacillus murA ANC1* (1311 bp)**

ATGGAAAAAATCATTGTGCGTGGTGGTAAACGTCTGAATGGCACCGTTAAAGTTG
AAGGTGCAAAAAATGCAGTTCTGCCGGTTATTGCAGCAAGCCTGCTGGCAAGCGA
AGGTAAAAGCGTTATTTATGATGTTCCGGCACTGAGTGATGTGTATACCATTAAT
GAAGTTCTGCGTCATCTGAATGCCGAAGTGGAATTTGAAAATAATCAGGTTACCG

TTGATGCAAGCCGTGAACTGAAAACCGAAGCACCGTTTGAATATGTTCGTAAAAT
GCGTGCAAGCGTTCTGGTTATGGGTCCGCTGCTGGCACGTACCGGTCATGCACGT
GTTGCACTGCCTGGTGGTTGTGCAATTGGTAGCCGTCCGATTGATCAGCATCTGA
AAGGTTTTGAAGCAATGGGTGCCAAAGTTAAAGTGGGCAATGGTTTTATTGAAGC
CAAAGTGGAAGGTCGTCTGCAGGGTGCGAAAATCTATCTGGATTTTCCGAGCGTT
GGTGCAACCGAAAACATTATGATGGCAGCAGCACTGGCAGAAGGCACCACCATTA
TTGAAAATGTTGCAAAAGAACCGGAAATCGTGGATCTGGCAAATTTTCTGAACGC
CATGGGAGCAAAAGTTCGTGGTGCAGGCACCGGCACCATTCGTATTGAAGGTGTT
GATAAACTGTATGGTGCCGAACATACCATTATTCCGGATCGCATTGAAGCAGGCA
CCTTTATGGTTGCAGCAGCAATTACCGGTGGTAATGTTCTGGTGAAAGGTGCAGT
TCCTGAACATCTGAGCAGCCTGATTGCAAAAATGGAAGAAATGGGCGTGGAAATT
ATTGAAGAGGGTGATGGTCTGCGTGTTATTGGTCCGGAAAAACTGAAAGCAGTGG
ATATCAAAACCATGCCGCATCCGGGTTTTCCGACCGATATGCAGAGCCAGATGAT
GGCACTGCTGCTGCGTGCCGAAGGTACAAGCATGATTACCGAAACCGTTTTTGAA
AACCGCTTTATGCACGTGGAAGAATTTCGTCGTATGAACGCAGACATTAAAATCG
AAGGTCGTAGCGTGATTATTAACGGTCCGAATAAACTGCAAGGTGCAGAAGTTGC
AGCAACCGATCTGCGTGCAGCAGCCGCACTGATTCTGGCAGGTCTGGTTGCGGAA
GGTTATACCCGTGTTACCGAACTGAAACATCTGGATCGTGGTTATGTGAACTTCC
ACAAAAAACTGGCAGCCCTGGGTGCAGATATTGAACGTGTTAATGATGAAGCCGA
AGAAGAAGAGGAAAACAAACAAGAGGTTAGCGATCTGAACGCGTAA

***Bacillus* MurA ANC1**

MEKIIVRGGKRLNGTVKVEGAKNAVLPVIAASLLASEGKSVIYDVPALSDVYTIN
EVLRHLNAEVEFENNQVTVDASRELKTEAPFEYVRKMRASVLVMGPLLARTGHAR
VALPGGCAIGSRPIDQHLKGFEAMGAKVKVGNGFIEAKVEGRLQGAKIYLDFPSV
GATENIMMAAALAEGTTIIENVAKEPEIVDLANFLNAMGAKVRGAGTGTIRIEGV
DKLYGAEHTIIPDRIEAGTFMVAAAITGGNVLVKGAVPEHLSSLIAKMEEMGVEI
IEEGDGLRVIGPEKLKAVDIKTMPHPGFPTDMQSQMMALLLRAEGTSMITETVFE
NRFMHVEEFRRMNADIKIEGRSVIINGPNKLQGAEVAATDLRAAAALILAGLVAE
GYTRVTELKHLDRGYVNFHKKLAALGADIERVNDEAEEEEENKQEVSDLNA

436 amino acids
Predicted molecular weight: 46930 Da
Theoretical pI: 5.36

***Bacillus murA ANC2* (1311 bp)**

ATGGAAAAAATCATTGTGCGTGGTGGTAAACGTCTGAATGGCACCGTTCGTGTTG
AAGGTGCAAAAAATGCAGTTCTGCCGGTTATTGCAGCAAGCCTGCTGGCAAGCGA
AGGTAAAAGCGTTATTTATGATGTTCCGGCACTGAGTGATGTGTATACCATTAAT
GAAGTTCTGCGTCATCTGAATGCCGAAGTGGAATTTGAAAATAATCAGGTTACCG
TGGATGCCAGCAAAGAACTGAAAACCGAAGCACCGTTTGAATATGTTCGTAAAAT
GCGTGCAAGCGTTCTGGTTATGGGTCCGCTGCTGGCACGTACCGGTCATGCACGT
ATTGCACTGCCTGGTGGTTGTGCAATTGGTAGCCGTCCGATTGATCAGCATCTGA
AAGGTTTTGAAGCAATGGGTGCCAAAGTTAAAGTGGGCAATGGTTTTATTGAAGC
CAAAGTTGAAGGTCGTCTGCAGGGTGCGAAAATCTATCTGGATTTTCCGAGCGTT
GGTGCAACCGAAAACATTATGATGGCAGCAGCACTGGCAGAAGGCACCACCATTA

TTGAAAATGCAGCAAAAGAACCGGAAATTGTGGATCTGGCAAATTTTCTGAACGC
CATGGGAGCAAAAGTTCGTGGTGCAGGCACCGGCACCATTCGCATTGAAGGCGTT
GATAAACTGTATGGTGCAGAACATACCATTATTCCGGATCGTATTGAAGCAGGCA
CCTTTATGGTTGCAGCAGCAATTACCGGTGGTAATGTTCTGGTGGAAAATGCCGT
TCCTGAACATATCCGTAGCCTGATTGCAAAAATGGAAGAAATGGGCGTGGAAATT
ATTGAAGAGGGTGATGGTCTGCGTGTTATTGGTCCGGAAAAACTGAAAGCAGTGG
ATATCAAAACCATGCCGCATCCGGGTTTTCCGACCGATATGCAGAGCCAGATGAT
GGCACTGCTGCTGCGTGCCGAAGGTACAAGCATGATTACCGAAACCGTTTTTGAA
AACCGCTTTATGCACGTGGAAGAATTTCGTCGTATGAACGCAGACATTAAAATCG
AAGGTCGTAGCGTGATTATTAACGGTCCGAATAAACTGCAAGGTGCCGAAGTTGC
AGCAACCGATCTGCGTGCAGCAGCCGCACTGATTCTGGCAGGTCTGGTTGCGGAA
GGTTATACCCGTGTTACCGAACTGAAACATCTGGATCGTGGTTATGTGAACTTCC
ACAAAAAACTGGCAGCCCTGGGTGCAGATATTGAACGTGTTAATGAAGAAGCCGA
AGAAGAAGAGGAAAACAAACAAGAGGTTAGCGATCTGAACGCGTAA

***Bacillus* MurA ANC2**

MEKIIVRGGKRLNGTVRVEGAKNAVLPVIAASLLASEGKSVIYDVPALSDVYTIN
EVLRHLNAEVEFENNQVTVDASKELKTEAPFEYVRKMRASVLVMGPLLARTGHAR
IALPGGCAIGSRPIDQHLKGFEAMGAKVKVGNGFIEAKVEGRLQGAKIYLDFPSV
GATENIMMAAALAEGTTIIENAAKEPEIVDLANFLNAMGAKVRGAGTGTIRIEGV
DKLYGAEHTIIPDRIEAGTFMVAAAITGGNVLVENAVPEHIRSLIAKMEEMGVEI
IEEGDGLRVIGPEKLKAVDIKTMPHPGFPTDMQSQMMALLLRAEGTSMITETVFE
NRFMHVEEFRRMNADIKIEGRSVIINGPNKLQGAEVAATDLRAAAALILAGLVAE
GYTRVTELKHLDRGYVNFHKKLAALGADIERVNEEAEEEEENKQEVSDLNA

436 amino acids
Predicted molecular weight: 47057 Da
Theoretical pI: 5.31

***Bacillus murA ANC3* (1311 bp)**

ATGGAAAAAATCATTGTGCGTGGTGGTAAACGTCTGAATGGCACCGTTCGTGTTG
AAGGTGCAAAAAATGCAGTTCTGCCGGTTATTGCAGCAAGCCTGCTGGCAAGCAA
AGGCACCAGCGTTATTTATGATGTTCCGGCACTGGCAGATGTGTATACCATTAAT
GAAGTTCTGCGTAATCTGAACGCCGAGGTGGAATATGATAATGGTCAGATTACCG
TTGATGCCGAGAAAGAACTGAAAACCGAAGCACCGTTTGAATATGTTCGTAAAAT
GCGTGCAAGCTTTCTGGTTATGGGTCCGCTGCTGGCACGTATTGGTCATGCCCGT
ATTGCACTGCCTGGTGGTTGTGCAATTGGTAGCCGTCCGATTGATCAGCATCTGA
AAGGTTTTGAAGCAATGGGTGCCAAAGTGGAAATTGGCAATGGTTTTATTGAAGC
GAAAGTTGAAGGTCGTCTGCAGGGTGCGAAAATCTATCTGGATTTTCCGAGCGTT
GGTGCAACCGAAAACATTATGATGGCAGCAGCACTGGCCGAAGGCACCACCATTA
TTGAAAATGCAGCACAAGAACCGGAAATTGTTGATCTGGCAAATTATCTGAATGC
CATGGGAGCAAAAGTTCGTGGTGCAGGCACCGGCACCATTCGCATTGAAGGCGTT
GATGAACTGAATGGTGCAGAACATACCATTATTCCGGATCGTATTGAAGCAGGCA
CCTTTATGGTTGCAGCAGCAATTACCGGTGGTAATGTTCTGGTTGAAAATGTTGT
GCCTGAACATATCCGTCCGCTGATTGCAAAAATGACCGAAATGGGTGTGAAAATC
ATCGAAGAAGAAAACGGTCTGCGCGTTATTGGTCCGGAAAAACTGAAAGCAGTGG

ATATCAAAACCATGCCGCATCCGGGTTTTCCGACCGATATGCAGAGCCAGATGAT
GGCACTGCTGCTGCGTGCAGAAGGTACAAGCGTTATTACCGAAACCGTTTTTGAA
AACCGCTTTATGCACGTGGAAGAATTTCGTCGTATGAATGCCGACATTAAAATCG
AAGGTCGTAGCGCAATTGTTAACGGTCCGAATAAACTGCAAGGTGCCGAAGTTGC
AGCAACCGATCTGCGTGCGGGTGCAGCACTGATTCTGGCAGGTCTGGTTGCAGAT
GGTTATACCCGTGTTACCGAACTGAAACATATTGATCGTGGCTATGTTGACTTTG
CCGGTAAACTGAAAGCCCTGGGTGCAGATATTGAACGTGTTAATGAAAAAGCCGA
AGAAGAGGAAGAGAACAAAGAAGAAGCCGCACCGCTGAAAGCATAA

***Bacillus* MurA ANC3**

MEKIIVRGGKRLNGTVRVEGAKNAVLPVIAASLLASKGTSVIYDVPALADVYTIN
EVLRNLNAEVEYDNGQITVDAEKELKTEAPFEYVRKMRASFLVMGPLLARIGHAR
IALPGGCAIGSRPIDQHLKGFEAMGAKVEIGNGFIEAKVEGRLQGAKIYLDFPSV
GATENIMMAAALAEGTTIIENAAQEPEIVDLANYLNAMGAKVRGAGTGTIRIEGV
DELNGAEHTIIPDRIEAGTFMVAAAITGGNVLVENVVPEHIRPLIAKMTEMGVKI
IEEENGLRVIGPEKLKAVDIKTMPHPGFPTDMQSQMMALLLRAEGTSVITETVFE
NRFMHVEEFRRMNADIKIEGRSAIVNGPNKLQGAEVAATDLRAGAALILAGLVAD
GYTRVTELKHIDRGYVDFAGKLKALGADIERVNEKAEEEEENKEEAAPLKA

436 amino acids
Predicted molecular weight: 46885 Da
Theoretical pI: 5.20

***Bacillus murA LCA* (1311 bp)**

ATGGAAAAAATCATTGTGCGTGGTGGTAAACGTCTGAATGGCACCGTTCGTGTTG
AAGGTGCAAAAAATGCAGTTCTGCCGGTTATTGCAGCAAGCATTCTGGCAAGCAA
AGGCACCAGCACCATTTATGATGTTCCGGCACTGGCAGATGTGTATACCATTAAT
GAAGTTCTGCGTAATCTGAACGCCGAGGTGGAATATGATAATGGTCAGATTACCG
TTGATGCCGAGAAAACCCTGAAAACCGAAGCACCGTTTGAATATGTTCGTAAAAT
GCGTGCAAGCTTTCTGGTTATGGGTCCGCTGCTGGCACGTATTGGTCATGCCCGT
ATTGCACTGCCTGGTGGTTGTGCAATTGGTAGCCGTCCGATTGATCAGCATCTGA
AAGGTTTTGAAGCAATGGGTGCCAAAGTGGAAATTGGCAATGGTTTTATTGAAGC
GAAAGTTGAAGGTCGTCTGCAGGGTGCGAAAATCTATCTGGATTTTCCGAGCGTT
GGTGCAACCGAAAACATTATGATGGCAGCAGCAATGGCAGAAGGCACCACCATTA
TTGAAAATGCAGCACAAGAACCGGAAATTGTTGATCTGGCAAATTATCTGAATGC
CATGGGAGCAAAAGTTCGTGGTGCAGGTACAGGCACCATTCGCATTGAAGGCGTT
GATGAACTGAATGGTGCAGAACATACCATTATTCCGGATCGTATTGAAGCAGGCA
CCTTTATGGTTGCAGCAGCCATTACCGGTGGTAATGTTCTGGTTGAAAATGTTGT
GCCTGAACATATCCGTCCGCTGATTGCAAAAATGACCGAAATGGGTGTGAAAATC
ATCGAAGAAGAAAACGGTCTGCGCGTTATTGGTCCGGAAAAACTGAAAGCAGTGG
ATATCAAAACCATGCCGCATCCGGGTTTTCCGACCGATATGCAGAGCCAGATGAT
GGCACTGCTGCTGCGTGCAGAAGGTACAAGCGTTATTACCGAAACCGTTTTTGAA
AACCGCTTTATGCACGTGGAAGAATTTCGTCGTATGAACGCCAACATTAAAATCG
AAGGTCGTAGCGCAATTGTTACCGGTCCGAATAAACTGCAAGGTGCCGAAGTTGC
AGCAACCGATCTGCGTGCCGGTGCAGCACTGATTCTGGCAGGTCTGGTTGCAGAT
GGTTATACCCGTGTTACCGAACTGAAACATATTGATCGTGGCTATGTTGACTTTG

CCGGTAAACTGAAAGCCCTGGGTGCAGATATTGAACGTGTTAATGAAAAAGCCGA
AGAAGAGGAAGAAAATCTGGAAGCAGCAGCACCGCTGAAAGTTTAA

Bacillus MurA LCA

MEKIIVRGGKRLNGTVRVEGAKNAVLPVIAASILASKGTSTIYDVPALADVYTIN
EVLRNLNAEVEYDNGQITVDAEKTLKTEAPFEYVRKMRASFLVMGPLLARIGHAR
IALPGGCAIGSRPIDQHLKGFEAMGAKVEIGNGFIEAKVEGRLQGAKIYLDFPSV
GATENIMMAAAMAEGTTIIENAAQEPEIVDLANYLNAMGAKVRGAGTGTIRIEGV
DELNGAEHTIIPDRIEAGTFMVAAAITGGNVLVENVVPEHIRPLIAKMTEMGVKI
IEEENGLRVIGPEKLKAVDIKTMPHPGFPTDMQSQMMALLLRAEGTSVITETVFE
NRFMHVEEFRRMNANIKIEGRSAIVTGPNKLQGAEVAATDLRAGAALILAGLVAD
GYTRVTELKHIDRGYVDFAGKLKALGADIERVNEKAEEEEENLEAAAPLKV

436 amino acids
Predicted molecular weight: 46818 Da
Theoretical pI: 5.31

Bacillus subtilis aroA (1287 bp)

ATGAAACGAGATAAGGTGCAGACCTTACATGGAGAAATACATATTCCCGGTGATA
AATCCATTTCTCACCGCTCTGTTATGTTTGGCGCGCTAGCGGCAGGCACAACAAC
AGTTAAAAACTTTCTGCCGGGAGCAGATTGTCTGAGCACGATCGATTGCTTTAGA
AAAATGGGTGTTCACATTGAGCAAAGCAGCAGCGATGTCGTGATTCACGGAAAAG
GAATCGATGCCCTGAAAGAGCCAGAAAGCCTTTTAGATGTCGGAAATTCAGGTAC
AACGATTCGCCTGATGCTCGGAATATTGGCGGGCCGTCCTTTTTACAGCGCGGTA
GCCGGAGATGAGAGCATTGCGAAACGCCCAATGAAGCGTGTGACTGAGCCTTTGA
AAAAAATGGGGGCTAAAATCGACGGCAGAGCCGGCGGAGAGTTTACACCGCTGTC
AGTGAGCGGCGCTTCATTAAAAGGAATTGATTATGTATCACCTGTTGCAAGCGCG
CAAATTAAATCTGCTGTTTTGCTGGCCGGATTACAGGCTGAGGGCACAACAACTG
TAACAGAGCCCCATAAATCTCGGGACCACACTGAGCGGATGCTTTCTGCTTTTGG
CGTTAAGCTTTCTGAAGATCAAACGAGTGTTTCCATTGCTGGTGGCCAGAAACTG
ACAGCTGCTGATATTTTTGTTCCTGGAGACATTTCTTCAGCCGCGTTTTTCCTTG
CTGCTGGCGCGATGGTTCCAAACAGCAGAATTGTATTGAAAAACGTAGGTTTAAA
TCCGACTCGGACAGGTATTATTGATGTCCTTCAAAACATGGGGGCAAAACTTGAA
ATCAAACCATCTGCTGATAGCGGTGCAGAGCCTTATGGAGATTTGATTATAGAAA
CGTCATCTCTAAAGGCAGTTGAAATCGGAGGAGATATCATTCCGCGTTTAATTGA
TGAGATCCCTATCATCGCGCTTCTTGCGACTCAGGCGGAAGGAACCACCGTTATT
AAGGACGCGGCAGAGCTAAAAGTGAAAGAAACAAACCGTATTGATACTGTTGTTT
CTGAGCTTCGCAAGCTGGGTGCTGAAATTGAACCGACAGCAGATGGAATGAAGGT
TTATGGCAAACAAACGTTGAAAGGCGGCGCTGCAGTGTCCAGCCACGGAGATCAT
CGAATCGGAATGATGCTTGGTATTGCTTCCTGTATAACGGAGGAGCCGATTGAAA
TCGAGCACACGGATGCCATTCACGTTTCTTATCCAACCTTCTTCGAGCATTTAAA
TAAGCTTTCGAAAAAATCCTGA

***Bacillus subtilis* AroA**

MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAGTTTVKNFLPGADCLSTIDCFR
KMGVHIEQSSSDVVIHGKGIDALKEPESLLDVGNSGTTIRLMLGILAGRPFYSAV
AGDESIAKRPMKRVTEPLKKMGAKIDGRAGGEFTPLSVSGASLKGIDYVSPVASA
QIKSAVLLAGLQAEGTTTVTEPHKSRDHTERMLSAFGVKLSEDQTSVSIAGGQKL
TAADIFVPGDISSAAFFLAAGAMVPNSRIVLKNVGLNPTRTGIIDVLQNMGAKLE
IKPSADSGAEPYGDLIIETSSLKAVEIGGDIIPRLIDEIPIIALLATQAEGTTVI
KDAAELKVKETNRIDTVVSELRKLGAEIEPTADGMKVYGKQTLKGGAAVSSHGDH
RIGMMLGIASCITEEPIEIEHTDAIHVSYPTFFEHLNKLSKKS

428 amino acids
Predicted molecular weight: 45240 Da
Theoretical pI: 6.33

***Bacillus aroA LCA* (1293 bp)**

ATGAAAGAGAAAACCATTAACACCGCAGAAAAAGGTCTGAATGGCACCATTCGTA
TTCCGGGTGATAAAAGCATTAGCCATCGTGCAGTTATGTTTGGTGCAATTGCCGA
AGGCACCACCACCGTTAAAAACTTTCTGCCTGGTGAAGATTGTCTGAGCACCATT
GCATGTTTTCGTAAACTGGGTGTTGAAATCGAACAGAATGGTGATGATGTTACCA
TTAACGGTAAAGGCCTGGATGGTCTGAAAGAACCGAAAGATGTTCTGGATGTTGG
TAATAGCGGTACAACCATTCGTCTGATGCTGGGTATTCTGGCAAATCGTCCGTTT
CATAGCACCATTATTGGTGATGAAAGTATTGCAAAACGTCCGATGAAACGTGTTA
CCGATCCGCTGCGTGAAATGGGTGCACAGATTGATGGTCGTGAAGATGGCAAATT
TACACCGCTGAGCATTCGTGGTGGTAATCTGAAAGGTATCGATTATAACAGTCCG
GTTGCAAGCGCACAGGTTAAAAGCGCAATTCTGCTGGCAGGTCTGCAGGCAGAAG
GTACAACAACCGTTACCGAACCGCATAAAAGCCGTGATCATACCGAACGTATGCT
GGAAGCATTTGGTGTTAAAGTTGAAGAAGAAGGTCAGACCGTTAGCATTGAAGGT
GGTCAGACCCTGAAAGGCACCGATATTGAAGTTCCAGGTGATATTAGCAGCGCAG
CATTTTTTCTGGTTGCCGGTGCCATTGTTCCGAATAGCCGTATTGTTCTGAAAAA
TGTTGGTCTGAATCCGACCCGTACCGGTATTATTGATGTGCTGAAAAAAATGGGA
GCCAACCTGGAAATTGATCAGGTTCGTGAAGAGGGTGCAGAACCGTATGGTGATC
TGACCATTGAAACCAGCAGTCTGAAAGGGATTGAAATTGGTGGCGATCTGATTCC
GCGTCTGATTGATGAAATTCCGATTATTGCACTGCTGGCAACCCAGGCGGAAGGC
ACAACCGTGATTAAAGATGCAGAAGAACTGAAAGTTAAAGAAACCAACCGTATTG
ATACCGTTGTGAGCGAACTGAAAAAACTGGGTGCAAATATTGAAGCAACCGATGA
TGGCATGATTATCCATGGTAAAACCACACTGAAAGGTGGTGCAACCGTTGATAGC
CATGGTGATCATCGTATTGGTATGATGCTGGCCATTGCAAGCTGTATTGCAGAAG
GTGAAGTTGAAATTGAAGATACCGATGCAGTTAGCGTTAGCTATCCGAAATTTTT
CGAACATCTGGAAAGCCTGAAAAAATGA

***Bacillus* AroA LCA**

MKEKTINTAEKGLNGTIRIPGDKSISHRAVMFGAIAEGTTTVKNFLPGEDCLSTI
ACFRKLGVEIEQNGDDVTINGKGLDGLKEPKDVLDVGNSGTTIRLMLGILANRPF
HSTIIGDESIAKRPMKRVTDPLREMGAQIDGREDGKFTPLSIRGGNLKGIDYNSP
VASAQVKSAILLAGLQAEGTTTVTEPHKSRDHTERMLEAFGVKVEEEGQTVSIEG

GQTLKGTDIEVPGDISSAAFFLVAGAIVPNSRIVLKNVGLNPTRTGIIDVLKKMG
ANLEIDQVREEGAEPYGDLTIETSSLKGIEIGGDLIPRLIDEIPIIALLATQAEG
TTVIKDAEELKVKETNRIDTVVSELKKLGANIEATDDGMIIHGKTTLKGGATVDS
HGDHRIGMMLAIASCIAEGEVEIEDTDAVSVSYPKFFEHLESLKK

430 amino acids
Predicted molecular weight: 45930 Da
Theoretical pI: 5.07

***Escherichia coli shikimate kinase (aroL)* (525 bp)**

ATGACACAACCTCTTTTTCTGATCGGGCCTCGGGGCTGTGGTAAAACAACGGTCG
GAATGGCCCTTGCCGATTCGCTTAACCGTCGGTTTGTCGATACCGATCAGTGGTT
GCAATCACAGCTCAATATGACGGTCGCGGAGATCGTCGAAAGGGAAGAGTGGGCG
GGATTTCGCGCCAGAGAAACGGCGGCGCTGGAAGCGGTAACTGCGCCATCCACCG
TTATCGCTACAGGCGGCGGCATTATTCTGACGGAATTTAATCGTCACTTCATGCA
AAATAACGGGATCGTGGTTTATTTGTGTGCGCCAGTATCAGTCCTGGTTAACCGA
CTGCAAGCTGCACCGGAAGAAGATTTACGGCCAACCTTAACGGGAAAACCGCTGA
GCGAAGAAGTTCAGGAAGTGCTGGAAGAACGCGATGCGCTATATCGCGAAGTTGC
GCATATTATCATCGACGCAACAAACGAACCCAGCCAGGTGATTTCTGAAATTCGC
AGCGCCCTGGCACAGACGATCAATTGTTGA

***Escherichia coli* shikimate kinase**

MTQPLFLIGPRGCGKTTVGMALADSLNRRFVDTDQWLQSQLNMTVAEIVEREEWA
GFRARETAALEAVTAPSTVIATGGGIILTEFNRHFMQNNGIVVYLCAPVSVLVNR
LQAAPEEDLRPTLTGKPLSEEVQEVLEERDALYREVAHIIIDATNEPSQVISEIR
SALAQTINC

174 amino acids
Predicted molecular weight: 19151 Da
Theoretical pI: 4.72

Appendix D: Bacterial strains and Plasmids

D1: Plasmids and cell strains

Table A.4: Plasmids used in this study

Plasmid	Description	His Tag used
pPROEX HTb	*E. coli* expression vector. Trc promoter. Encodes an N-terminal 6 x histidine tag with an rTEV protease cleavage site. Ampicillin resistance	N-terminal
pET28b-PstI	Bacterial expression vector with T7 promoter; Modified pET28b where the *Nco*I sites have been replaced with *Pst*I sites; C- or N-terminal His-tags; Kanamycin resistance	C-terminal

Table A.5: Bacterial strains used in the study

Cell strain	Description
DH5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17
BL21 (DE3)	F- ompT hsdSB (rB-mB-) gal dcm (DE3)
Keio collection – parent strain: *E. coli* BW25141	*rrnB3* DE*lacZ4787* DE*phoBR580hsdR514* DE(*araBAD*)*567DE(rhaBAD)568galU95 DEendA9::FRT DEuidA3::pir(wt) recA1 rph-1*
***E. coli* BW25141 – *aroA* KO**	*rrnB3* DE*lacZ4787* DE*phoBR580hsdR514* DE(*araBAD*)*567DE(rhaBAD)568galU95 DEendA9::FRT DEuidA3::pir(wt) recA1 rph-1* Δ*aroA*

D2: Genetically modified organisms (GMOs) used in this study

Table A.6: Transformants generated in this study

Cell strain	Plasmid
***E. coli* DH5α**	pPROEX HtB- *Bacillus* AroA LCA
	pPROEX HtB- *Bacillus* MurA LCA
	pPROEX HtB- *Bacillus* MurA ANC1
	pPROEX HtB- *Bacillus* MurA ANC2
	pPROEX HtB- *Bacillus* MurA ANC3
	pPROEX HtB- *Bacillus subtilis* MurA
	pPROEX HtB- *Bacillus coagulans* MurA
***E.. coli* BL21**	pPROEX HtB- *Bacillus subtilis* AroA
	pET28b-pstI- Shikimate kinase
***E. coli* BW25141**	pPROEX HtB
***E. coli* BW25141 – *aroA* KO**	pPROEX HtB
	pPROEX HtB- *Bacillus subtilis* AroA
	pPROEX HtB- *Bacillus* AroA LCA

Appendix E: Reagents, Buffers, Growth Media and Gels

E1: Reagents and buffers

Coomassie stain	0.05 % w/v coomassie blue R-250, 25 % v/v isopropanol, 10% v/v acetic acid
Destain solution	10% v/v acetic acid
10x DNA loading dye	0.4 % w/v bromophenol blue, 0.4 % w/v xylene, 50 % v/v glycerol
4x SDS loading dye	200 mM Tris-HCl pH 6.8, 8% (w/v) SDS, 40% (v/v) glycerol, 0.4% (w/v) bromophenol blue, 400 mM β-mercaptoethanol
1x SDS-PAGE running buffer	25 mM Tris-HCl pH 8.3, 250 mM glycine, 0.1% (w/v) SDS
1x TAE buffer	40 mM Tris-acetate, 1 mM EDTA

E2: Buffers used in lysis buffer screen and Protein Melt Analysis

pH 4.0 :	50 mM sodium acetate + 300 mM NaCl
pH 5.0 :	50 mM sodium acetate + 300 mM NaCl
pH 6.0 :	50 mM citrate + 300 mM NaCl
pH 7.0 :	50 mM sodium phosphate + 300 mM NaCl
pH 7.5 :	50 mM Tris + 300 mM NaCl
pH 8.0 :	50 mM sodium phosphate + 300 mM NaCl
pH 9.0 :	50 mM glycine + 300 mM NaCl
pH 10.0 :	50 mM glycine + 300 mM NaCl

E3: Growth media

LB broth	1 % (w/v) bactotryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl pH 8	
LB agar	1 % (w/v) bactotryptone, 0.5 % (w/v) yeast extract, 1 % (w/v) NaCl, 15 g/L agar pH 8	
M9 minimal media	5 x M9 salts:	64 g/L $Na_2HPO_4.7H_20$
		15 g/L KH_2PO_4
		2.5 g/L NaCl
		5.0 g/L NH_4Cl
	M9 media:	200 ml/L 5x M9 salts
		2 ml/L 1 M $MgSO_4$
		20 ml/L carbon source (0.2 μm filter sterilised)
		0.1 ml/L $CaCl_2$
		780 ml/L H_2O
M9 agar	200 ml/L 5x M9 salts 2 ml/L 1 M $MgSO_4$ 20 ml/L carbon source (0.2 μm filter sterilised) 0.1 ml/L $CaCl_2$ 780 ml/L H_2O with 15 g agar	
Terrific Broth	Solution A	12 g tryptone
		24 g yeast extract
		4 ml glycerol
		Made up to 900 ml with H_2O
	Solution B	2.31 g KH_2PO_4
		12.54 g K_2HPO_4
		Made up to 100 ml with H_20

E4: Gels

Agarose gel	1 % w/v agarose in TAE buffer and 2-5 μL SYBR safe DNA gel stain (Invitrogen) per gel

15% SDS-PAGE gel recipe

SDS-PAGE gels were made up of a resolving gel layer overlaid with a stacking gel layer

Resolving gel layer:	15% acrylamide 375 mM Tris (pH 8.8) 0.1% (w/v) SDS 0.05% (w/v) ammonium persulphate 0.05 % (v/v) TEMED
Stacking gel layer:	5% acrylamide 128 mM Tris (pH 6.8) 0.1% (w/v) SDS 0.05% (w/v) ammonium persulphate 0.05 % (v/v) TEMED