Structural and Biochemical Characterisation of Ribitol-5- phosphate Transferases

Thomas Evan Ramsey

PhD

University of York

Department of Chemistry

February 2025

Abstract

Wall teichoic acids (WTAs) play a role in cell growth, division and morphology as well as colonisation and adhesion in Gram-positive bacterial species such as *Staphylococcus aureus* and *Bacillus subtilis.* In *S. aureus* and *B. subtilis* W23*,* WTAs largely consists of a ribitol-5-phosphate (RboP) polymer, which is synthesised by TarK and TarL enzymes using CDP-ribitol (CDP-Rbo). This donor substrate, CDP-Rbo, is synthesised from RboP by TarI. Deletion of the enzymes along the WTA pathway results in restored antibiotic susceptibility in methicillin-resistant *S. aureus*, demonstrating the potential of WTA biosynthesis as a therapeutic target. Therefore the characterisation of the Tar enzymes could aid in the development of novel antimicrobial agents. This project endeavoured to facilitate the generation of the difficult-to-synthesise donor substrate, CDP-Rbo, via TarI and to characterise the TarK and TarL enzymes, thereby providing insight into the protein activity and 3D structures.

Firstly, CDP-Rbo synthesis was achieved using a chemoenzymatic approach utilising TarI with RboP produced by members of the group via chemical synthesis. Efforts were then made to recombinantly express a range of TarK/L homologues which highlighted that large set of expression trials were required to produce soluble protein. Recombinant TarK and TarL from *B. subtilis* W23 were successfully solubilised by using Rosetta2 plysS *E. coli* and, for TarL, changing the His6 tag location to the C-terminal end. Buffers containing CHAPS were required, therefore aiding in solubility throughout the multi-step chromatography protocols. Once the target proteins and variants of interest were purified, stability was assessed using circular dichroism and nanoDSF with/without ligands. TarK activity was measured by liquid chromatography–mass spectrometry and could utilise CDP-Rbo. Crystallisation was not successful, but mass photometry results identified multimeric forms of TarL ideal for cryo-EM. 2D classes then yielded a possible tetrameric form of TarL facilitating further structural studies using this technique.

Contents

Acknowledgements

I would like to thank my supervisors Dr Lianne Willems and Professor Gideon Davies for giving me the opportunity to start this PhD and for your support and guidance throughout. Additionally, I would like to thank my TAP member Professor Tony Wilkinson for giving me further advice on this research during our meetings.

Thank you to all technical staff for their assistance in this project. I would specifically like to thank Dr Julia Walton for her help throughout especially in all protein purification troubleshooting and with keeping the lab running. Thank you also to all staff in the technology facility, namely Dr Andrew Leech and Dr Katy Cornish for their experimental insights on all things nanoDSF and mass photometry and thank you Dr Thijs Voskuilen for providing various ligands for this project.

I would like to thank Dr Jamie Blaza and Sam Hart for introducing me to and helping me with cryo-EM. Thanks also to Dr Wendy Offen for her supervision and advice regarding all the crystallisation trials that were performed.

A big thanks to Dr Saeed Akkad, Dr Tessa Keenan and Abby Walklett for your help, support and wisdom throughout my PhD both inside and outside the lab. Also, thanks to all members of B-block and the Willems group, both past and present, for making my PhD manageable. Thanks to Josh and Connor for their wild conversations and discussions that definitely made the weeks fun, thanks to Morgan, Matt and Ella for the various climbing trips and thanks to all the members of the Organics football team for giving me a place to show off my limited football prowess (Campeones!).

To my family and closest friends, thank you for your unwavering love and support during the past three and a half years. To my sister Emily and Dad, thank you for the various visits to York to encourage me, listening to my rants and for giving me the push I needed to carry on and to Auntie Claire, thank you for checking up on me every now and again from all the way down south.

Finally, thank you to Mum, whom I dedicate this thesis to. Thank you for loving and supporting me for as long as you could and for being so proud. I hope you still are watching from wherever you may be.

Author’s Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Chapter 1: Introduction

1.1 Wall Teichoic Acids

Teichoic acids were first discovered in 1958 when a study attempted to uncover the function of cytidine-5'-diphosphate-glycerol (CDP-Gro) and cytidine-5'-diphosphate-ribitol (CDP-Rbo) in various Gram-positive bacteria (Armstrong *et al.*, 1958). They consist of phosphodiester-linked polyol repeat units and are a major component of Gram-positive bacterial cell walls (Ward, 1981). Teichoic acids are long anionic polymers which are categorised into two types: the membrane embedded lipoteichoic acids (LTAs) and wall teichoic acids (WTAs) which are covalently bound to the peptidoglycan (PG) layers and extend far from the cell wall (Aasjord and Grov, 1980) (**Figure 1.1**). In *Bacillus subtilis and Staphylococcus aureus* WTAs comprise of about 20-40 polyol-phosphate repeats (Kojima, Araki and Ito, 1985; Swoboda *et al.*, 2010). The main chain of WTAs can be separated into two units: the main chain polymer that consists of polyol-phosphate repeat units linked through phosphodiester bonds, and the linkage unit which anchors the polymer to the peptidoglycan layer. The linkage unit is conserved across bacterial species and is a disaccharide of *N*-acetylglucosamine 1-phosphate (P-GlcNAc) and *N*-acetylmannosamine (ManNAc) with 1 or 2 glycerol 3-phosphate units (GroP) attached to the C4 oxygen position of ManNAc (Araki and Ito, 1989; Brown *et al.*, 2010; Brown, Santa

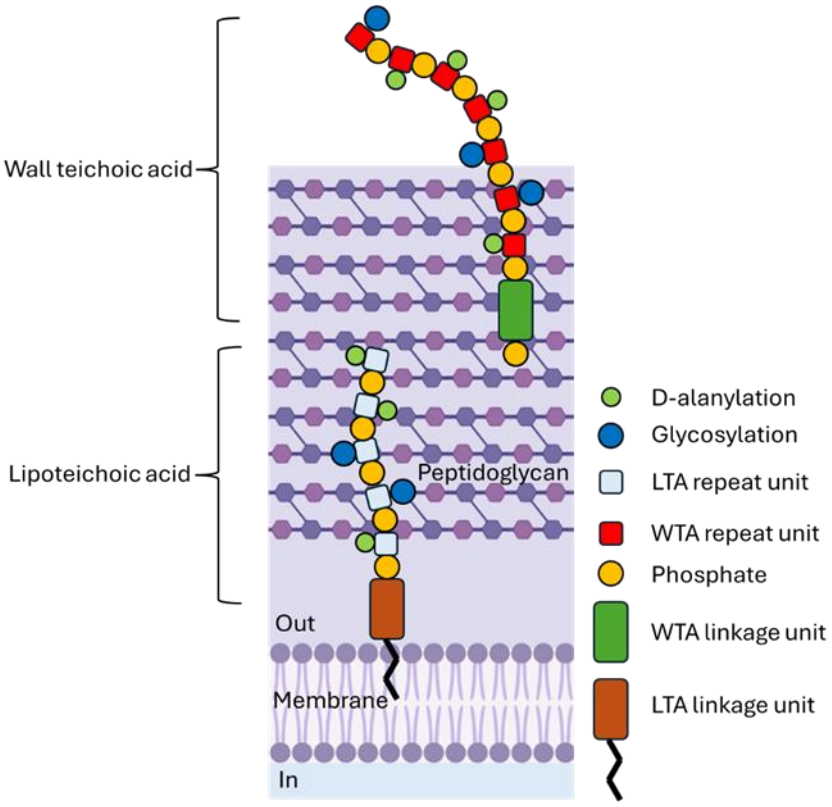


Figure 1.1: Schematic representation of teichoic acids located within and beyond the cell wall of Gram-positive bacteria. WTA = wall teichoic acid, LTA = lipoteichoic acid.

Maria and Walker, 2013a) (**Figure 1.2**). Several variations of the repeat unit can be observed in different bacterial species (**Figure 1.2**); however, the focus of this work is on the most characterised type 1 repeat units which are found in *B. subtilis* and *S. aureus* and consist of repeating GroP or ribitol-5-phosphate (RboP) units. Among the same bacterial species, the use of either GroP or RboP is not conserved. Subspecies of *B. subtilis,* specifically *B. subtilis* W23 and *B. subtilis* 168, display either RboP or GroP polymers, respectively (Brown, Santa Maria and Walker, 2013a). Despite WTA structural differences, they all contain a backbone which is largely negatively charged, and the functions remain the same as outlined below.

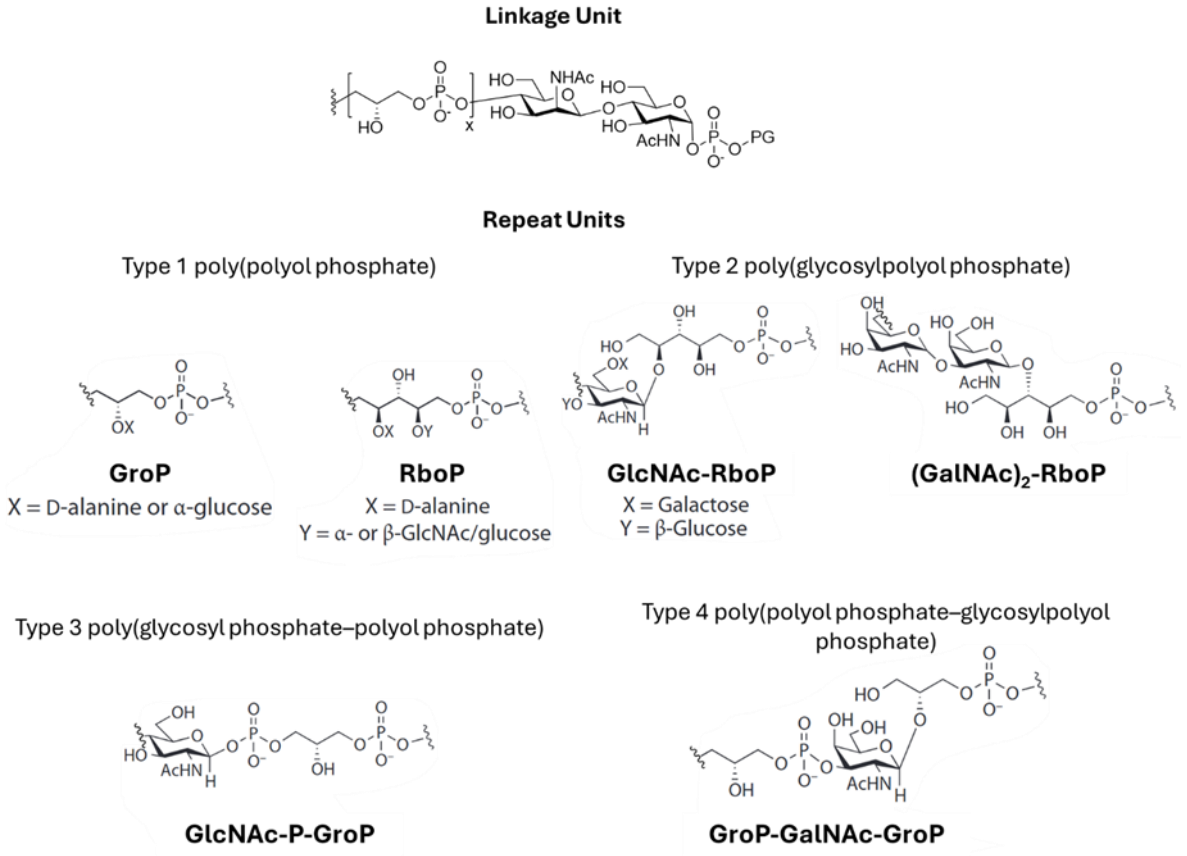


Figure 1.2: The linkage unit and the repeat units found in WTA. The most common linkage unit, GroP-ManNAc-GlcNAc-P-PG is conserved but the repeat unit is diverse. PG = peptidoglycan. Figure adapted from Brown, Santa Maria and Walker, 2013.

The main backbone of WTAs is further altered with the addition of cationic D-alanine or saccharides such as glucose and/or GlcNAc (**Figure 1.2**). In species utilising RboP repeat units, D-alanine moieties are bonded to position two of RboP, while sugar moieties are bonded to position four (Neuhaus and Baddiley, 2003; Vinogradov *et al.*, 2006) or in some cases position three (Gerlach *et al.*, 2018). D-alanine modification also has been found to fluctuate with differences in pH, salt concentration and temperature in the cell's external environment, whereas saccharide modifications do not appear to change (Jenni and Berger-Bächi, 1998; Neuhaus and Baddiley, 2003). Further differences in WTA lie between bacterial species; *S.*

aureus RboP repeats are almost solely decorated with GlcNAc modifications whereas in *B. subtilis* it was observed that some strands are fully glycosylated while others are not (Chin, Burger and Glaser, 1966; Jenni and Berger-Bächi, 1998). The anomeric configuration among strains of the same species also differs. In *S. aureus* the glycosidic bonds can be found only in α or β configuration or an assortment of both (Chin, Burger and Glaser, 1966; Nathenson *et al.*, 1966; Endl *et al.*, 1983, 1984; Brown, Santa Maria and Walker, 2013a).

1.2 Wall Teichoic Acid Biosynthesis

The location of WTA synthesis was first found using electron microscopy. Synthesis was observed to initiate at the wall-membrane interface, with the WTA polymer later appearing outside the cell attached to the cell wall (Anderson *et al.*, 1978; Ward, 1981). Most of the initial work on identifying the steps to WTA synthesis characterised the nature of the enzymatic steps but was not expansive enough to identify individual enzymes (Ward, 1981). This is in part due to most of the biosynthetic genes being essential, therefore rendering most gene knockout studies inadequate and unusable (Bhavsar, Beveridge and Brown, 2001; Kobayashi *et al.*, 2003; Bhavsar *et al.*, 2004; D'Elia, Millar, *et al.*, 2006; D'Elia, Pereira, *et al.*, 2006). However, genetic analysis of temperature sensitive and phage-resistant mutants allowed the GroP and RboP gene clusters to be elucidated in *B. subtilis* 168 and *B. subtilis* W23, respectively (Karamata, Pooley and Monod, 1987; Young *et al.*, 1989). Further biochemical characterisation experiments involving the identified proteins were impeded by the scarcity of the undecaprenyl substrate onto which WTAs are assembled (**Figure 1.3**). This lipid substrate is difficult to handle due to the long carbon chain and is of low abundance in bacterial cells, therefore isolation of enough of this substrate for biochemical and enzyme activity experiments is challenging (Brown, Santa Maria and Walker, 2013a). However, isolation of the WTA from *S. aureus* has been achieved allowing activity assays of TarL to be performed (Li *et al.*, 2024). Chemoenzymatic methods have also been established therefore allowing the synthesis of various WTA substrates with truncated lipid chains allowing easier access to, and handling of these compounds (Ginsberg *et al.*, 2006; Zhang *et al.*, 2006; Brown, Zhang and Walker, 2008; Brown *et al.*, 2010; Li *et al.*, 2024). These methods have allowed researchers to discover the function of specific WTA enzymes using activity assays and mechanistic studies on the recombinant proteins to uncover the broader process of WTA synthesis, modification, export and attachment (**Figure 1.3**).

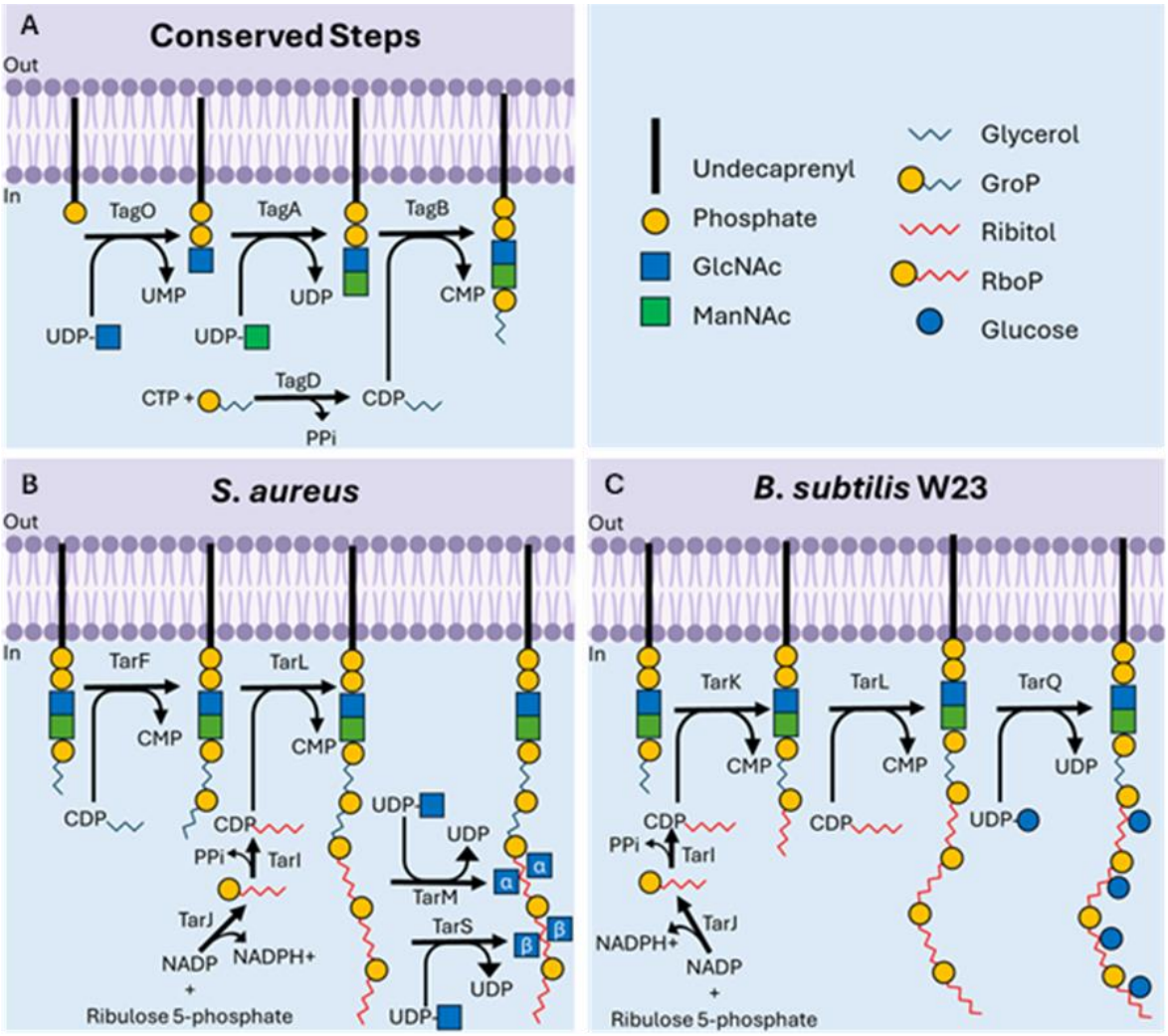


Figure 1.3: Schematic diagram of the conserved steps (A) and diverging steps of the downstream synthesis of WTA in S. aureus (B) and B. subtilis W23 (C) within the cell before export and attachment to the PG wall.

1.2.1 TagO

The initial step in the synthesis of the conserved unit in WTAs involves the addition of GlcNAc to undecaprenyl-phosphate embedded in the cytosolic side of the cell membrane by TagO (**Figure 1.3**). TagO was found to have sequence homology to other Uridine diphosphate (UDP)-N-acetylglucosamine: undecaprenyl-P N-acetylglucosaminyl 1-P transferases (Soldo, Lazarevic and Karamata, 2002). Further investigations made use of the *B subtilis* 168 strain in which native TagO was truncated. Inducible expression of full-length TagO was achieved by the isopropyl β-D-1-thiogalactopyranoside (IPTG)-mediated P_{spac} promoter. When the bacteria were grown in the presence of IPTG, typical cell morphology and phosphate, GlcNAc and glycerol cell wall content was observed. When transferred into IPTG-free media, the cells displayed a coccoid cell morphology which is typically found in mutants that cannot produce WTA. Reductions to phosphate, GlcNAc and glycerol content within the cell wall were also

observed (Soldo, Lazarevic and Karamata, 2002). This suggests that TagO likely has a role very early on in the synthesis of WTA.

1.2.2 TagA

TagA acts on the TagO product to install ManNAc to form the disaccharide ManNAc(β 1→4)GlcNAc (**Figure 1.3**). This is the first committed step of WTA biosynthesis (Zhang *et al.*, 2006). Murazumi *et al.*, 1988, managed to partially purify TagA using ammonium sulphate precipitation and various chromatography techniques. Enzyme kinetics experiments concluded that the purified enzyme has a high affinity for UDP-ManNAc (K_m = 4.4 μM) (Murazumi *et al.*, 1988). Other studies used purified TagA to replicate lipid disaccharide production *in vitro*. The TagA product was identified, using liquid chromatography-mass spectrometry (LC-MS), as the speculated ManNAc-GlcNAc-pp-lipid product, therefore supporting TagA's role in WTA synthesis (Ginsberg *et al.*, 2006; Zhang *et al.*, 2006).

1.2.3 TagB

TagB catalyses the transfer of the first GroP residue onto the C4 position of the ManNAc of the TagA product (**Figure 1.3**). TagB uses an activated form of the donor substrate, CDP-glycerol, synthesised from glycerol phosphate and cytidine triphosphate (CTP) by the enzyme TagD (Park *et al.*, 1993). Similar *in vitro* assays to the work on TagA were used to observe the synthesis of the GroP-disaccharide-pp-lipid, however in one study, radioactive [β-^{32}P]CDP-glycerol was used to track synthesis and not LC-MS (Bhavsar, Truant and Brown, 2005). Native TagB with a green fluorescent protein (GFP) tag was observed to interface at the membrane using fluorescent microscope *in vivo* which suggests that this protein is membrane bound (Bhavsar, Truant and Brown, 2005; Ginsberg *et al.*, 2006). TagB also shares homology with TagF, a glycerol phosphate polymerase present in species in which GroP is the main polymer in WTA (*B. subtilis* 168 and *S. epidermidis,* for example). However, TagB does not exhibit polymerisation activity but is required for 'priming' the chain with the first residue of GroP for TagF to act upon, thus leading to the name glycerol phosphate primase (Bhavsar, Truant and Brown, 2005).

After the TagB step, the biosynthetic path of WTA diverges into either GroP or RboP polymerisation. It is worth noting that the names of the previously mentioned enzymes change depending on the main constituent of the WTA polymer. If the polymer contains mainly GroP (e.g. *B. subtilis* 168 and *S. epidermidis)* then the enzymes are labelled TagX (Teichoic acid glycerol) and if they mainly contain RboP (e.g. *B. subtilis* W23 and *S. aureus)* then they are labelled TarX (Teichoic acid ribitol). Nevertheless, the enzymes retain the same biological functions.

The following sections focus on the RboP polymerisation steps present in *B. subtilis* W23 and *S. aureus* and explores the differences between the biosynthetic pathways in these species.

1.2.4 TarI and TarJ

To install RboP residues on the growing WTA chain, the activated form of RboP, CDP-ribitol (CDP-Rbo), is synthesised (**Figure 1.3**). Both *S. aureus* and *B. subtilis* W23 utilise a TarI and TarJ heterodimer reminiscent of a bifunctional enzyme found in *H. influenzae* Bcs1 (Zolli, Kobric and Brown, 2001). In early publications the co-overexpression of both TarI and TarJ was necessary to produce these enzymes in any meaningful quantity, first suggesting a hetero-dimeric protein complex (Pereira and Brown, 2004). The same study demonstrated that TarJ first reduces ribulose-5-phosphate to RboP which is then converted into CDP-Rbo by TarI with the use of CTP (Pereira and Brown, 2004). Despite the complex formation between these enzymes in separate species suggesting a functional association, rapid mixing experiments on TarIJ and Bcs1 highlighted that the RboP product from TarJ is not channelled directly into TarI (Pereira and Brown, 2004). This observation was later confirmed by the elucidation of the structure of TarIJ using X-ray protein crystallography (Li *et al.*, 2021a). A TarIJ hetero-tetramer has also been speculated along with the suggestion that TarIJ may be a subunit of a larger enzyme complex, possibly involving TarL, as it has been observed associating to the membrane (Formstone *et al.*, 2008; Li *et al.*, 2021a).

1.2.5 *S. aureus* TarF, TarK and TarL

Initially, TarF in *S. aureus* was proposed to install two GroP residues on the TagB product followed by the 'priming' of the chain with the first RboP residue by TarK (Lazarevic *et al.*, 2002; D'Elia, Pereira, *et al.*, 2006; Qian *et al.*, 2006). However, after *in vitro* reconstruction of the WTA synthesis pathway using overexpressed *S. aureus* proteins, a revised pathway was designed (Brown, Zhang and Walker, 2008). In these experiments, each product of the pathway from TagB to TarL was analysed and confirmed by a mixture of HPLC and LC-MS analysis after each enzymatic reaction was completed (Brown, Zhang and Walker, 2008). Conclusions were then drawn regarding how many and what type of residues were added after each enzyme treatment. This approach quantified the exact number of GroP residues attached by TarF which was reduced to only one residue (Brown, Zhang and Walker, 2008) (**Figure 1.3B**). Furthermore, TarK was found to not be required to polymerise RboP as the TarL enzyme could act on the TarF product directly, acting as both a 'primase' and polymerase (Brown, Zhang and Walker, 2008; Brown *et al.*, 2010) (**Figure 1.3B**). TarL was found to extend the RboP polymer to about 40 RboP residues long (Ward, 1981; Yokoyama *et al.*, 1986). The function of TarK in *S. aureus* is still unknown. Additionally, *S. aureus* contains another set of *tarIJL* genes labelled as *tarI'J'K* (Qian *et al.*, 2006; Meredith, Swoboda and Walker, 2008).

However, the function of this duplication is still unknown and may be redundant (Qian *et al.*, 2006; Meredith, Swoboda and Walker, 2008; Pereira, D'Elia, *et al.*, 2008).

1.2.6 *B. subtilis* W23 TarK and TarL

RboP WTA synthesis in *B. subtilis* W23 was originally suspected to be identical to *S. aureus* (Lazarevic *et al.*, 2002). However, despite being essential in *S. aureus,* studies exploring deletions of the *tarF* gene in *B. subtilis* W23 found growth rate and WTA content did not change, suggesting that synthesis of the TarF product is not required for RboP polymerisation (Brown *et al.*, 2010). In the same study, reverse transcriptase-polymerase chain reactions (RT-PCR) were also carried out during different stages to observe any changes in *tarF* expression levels. No changes in transcription were observed suggesting that, despite TarF being functional in other species, in *B. subtilis* W23 the enzyme is not require for poly RboP synthesis (Brown *et al.*, 2010).

Deletions of *tarK* could not be made. Therefore, unlike in *S. aureus,* TarK is essential in *B. subtilis* for WTA synthesis (Brown *et al.*, 2010). This was further supported by RT-PCR data showing the abundance of transcription of TarK during growth (Brown *et al.*, 2010). Furthermore, a hybrid strain of *S. aureus* in which the native *tarF* gene was deleted and the *tarK* and *tarL* genes from *B. subtilis* were incorporated maintained WTA synthesis of native length, suggesting that TarK assumes the role of TarF in *B. subtilis* (Brown *et al.*, 2010). Reactions that mimicked WTA synthesis *in vitro* were also performed in the same publication. Using His-tagged *B. subtilis* TarK and TarL (*Bs*TarK and *Bs*TarL) enzymes overexpressed in *E. coli* and purified using immobilized metal affinity chromatography (IMAC), reactions were performed on chemically synthesised substrates. After reaction, the samples were analysed using HPLC and MS. Results showed that *Bs*TarK does not utilise CDP-glycerol as a substrate and instead uses CDP-Rbo to install one RboP residue on the TagB product, making the *Bs*TarK enzyme a RboP 'primase' (Brown *et al.*, 2010). Reactions with intermediates with a differing number of GroP residues confirmed that *Bs*TarK has a strong preference for substrates with a single GroP unit i.e. the TagB product. Tandem reactions of *Bs*TarK and *Bs*TarL using radio labelled substrates also showed that no RboP polymer is formed on substrates with 2 GroP units, further confirming that *Bs*TarK can only act on substrates with one GroP unit. Furthermore, it was shown that *Bs*TarL cannot polymerise RboP without the first RboP residue installed by *Bs*TarK (Brown *et al.*, 2010). Thus, *Bs*TarK and *Bs*TarL act together as a primase-polymerase pair acting sequentially on the TagB product, only (Figure 1.3**C**).

1.3 End-stage Polymer Modifications

1.3.1 TarQ

Post-polymerisation modification of WTA has been suggested to also occur inside the cell prior to WTA export (Allison *et al.*, 2011; Brown *et al.*, 2012a). In *B. subtilis* W23 the poly RboP is modified via β-glycosylation using the enzyme TarQ (**Figure 1.3C**). This was found through deletions of the *tarQ* gene in *B. subtilis* and analysis of the extracted WTA. MS of the wild type (WT) WTA and the WTA from the strain with the *tarQ* deletion found that the mass of a single WTA unit decreased by about ~162 Da in the mutant strain (Brown *et al.*, 2012a). This change in mass is the equivalent to the addition of glucose therefore suggesting that TarQ is responsible for glucose modification (Brown *et al.*, 2012a). Overexpression of the *tarQ* gene was also used to supplement a mutant strain of *S. aureus* which did not express the suspected WTA modifying enzymes TarM and TarS. This strain did not modify WTAs through glycosylation but, with the addition of the *tarQ* gene, glucose modification was restored (Brown *et al.*, 2012a).

1.3.2 TarM, TarS and TarP

In *S. aureus* TarM and TarS have been found to modify the RboP polymer with α-GlcNAc and β-GlcNAc, respectively (Xia *et al.*, 2010; Brown *et al.*, 2012a) (**Figure 1.3B**). Initially *tarM* had an unknown function, however, when a *tarM* deletion strain of *S. aureus* was produced, the WTA produced did not contain any α-GlcNAc modification. This was elucidated using proton NMR to compare WT WTA to the knockout mutant strain's WTA, whereby the presence of the peak characteristic of the anomeric proton in α-GlcNAc was not identified in the spectra of the mutant WTA (Xia *et al.*, 2010). This was restored with the addition of a plasmid encoding *tarM* (Xia *et al.*, 2010). Further investigation using overexpressed TarM, unmodified WTA and UDP-GlcNAc found that TarM modifies the WTA with α-GlcNAc *in vitro* confirming its role as an α-GlcNAc-transferase (Xia *et al.*, 2010). However, β-GlcNAc modification was still observed suggesting that another enzyme exists to perform this reaction.

TarS was identified in the same operon as TarM (NCTC8325) and therefore, it was suspected to function as the unknown β-GlcNAc transferase (Brown *et al.*, 2012a). Using a similar approach to the elucidation of TarM's function, TarS was overexpressed and purified, and reaction mixtures were prepared with radiolabelled UDP-GlcNAc and poly RboP WTA synthesised using a combined chemical and enzymatic approach. Completed reactions were analysed using poly acrylamide gel electrophoresis (PAGE) and phosphor-imaging which found that the starting material depleted and products with a similar mobility to the WTA

polymers appeared (Brown *et al.*, 2012a). To discover the stereochemistry of this modification, multiple knockout mutant strains of *S. aureus* were designed which disabled either *tarS*, *tarM* or both and their WTAs were isolated, hydrolysed and the monomers were analysed by NMR similar to the TarM study. This highlighted that the characteristic peak for the anomeric proton of β-*O*-GlcNAc was not present in the NMR spectra of WTAs of the TarS and TarS + TarM knockout mutants but was present in the TarM knockout mutant (Brown *et al.*, 2012a). Much like in the TarM study, when the double knockout mutant was supplemented with a plasmid encoding the *tarS* gene, β-*O*-GlcNAc modification was restored (Brown *et al.*, 2012a). Thus, TarS is a β-GlcNAc-transferase.

Another enzyme, named TarP, was also found to be responsible for β-GlcNAc modification of the RboP polymer in WTA. Further investigation using proton NMR found that TarP selectively glycosylates the C3 position of the RboP residue in Methicillin-Resistant *S. aureus* (MRSA) WTAs which differs from the C4 modification facilitated by TarS predominantly in many other *S. aureus* strains (Gerlach *et al.*, 2018). Structure elucidation of TarP with several RboP units also confirmed that the geometry of the catalytic residues in the active site appear close to the C3 hydroxyl group similar to the location of the catalytic residues around the RboP C4 hydroxyl group in the TarS structure (Sobhanifar *et al.*, 2016; Gerlach *et al.*, 2018). To assess the effect glycosylation of the different hydroxyl groups had on the immune response, human Immunoglobulin G (IgG) was applied to isolated WTA from different strains containing different *tarS/tarP* gene knockouts. WTAs from dual knockouts of *TarS and TarP* had the lowest deposition of IgG which indicates that glycosylated WTA is a prominent antigen in humans. The *tarP* gene knockout strain had the highest deposition of IgG suggesting that β-GlcNAc on the C4 position of RboP is an important epitope for human anti-*S. aureus* antibodies (Gerlach *et al.*, 2018). In the strain expressing TarP with and without TarS expression there was only a slight increase in in IgG binding compared to the mutant without any glycosylation(Gerlach *et al.*, 2018). This would suggest that a small percentage of *S. aureus*-specific antibodies can bind WTA with β-GlcNAc at the C3 position of RboP, and *tarP*-expressing *S. aureus* are less likely to be detected and eliminated by human phagocytes. Thus, TarP is important in MRSA immune evasion and further research could improve MRSA treatment.

1.4 Polymer Export, Attachment and D-alanylation

1.4.1 TagGH/TarGH

After the fully assembled and glycosylated polymer is completed, the chain is translocated through the membrane, most likely by a two-component ATP binding cassette (ABC) transporter. This aspect of WTA synthesis is not as well understood as the above steps,

however some work has been done in identifying the *tagGH* operon as the ABC transporter in *B. subtilis* 168 (Lazarevic and Karamata, 1995). TagH contains an ATPase domain, thereby providing the energy required to alter the structure of TagG to facilitate WTA export (Lazarevic and Karamata, 1995; Schirner, Stone and Walker, 2011; Brown, Santa Maria and Walker, 2013a). It is worth again noting that these Tag proteins refer to species which produce GroP WTA, however, it is likely that the Tar equivalents, which synthesise RboP WTA, in *S. aureus* and *B. subtilis* W23 (TarGH) retain the same function s (Schirner, Stone and Walker, 2011). Structural elucidation of the N terminal domain of TagH has provided further evidence of the role of TagH as an ATPase as the dimeric structure is very common to other ATPases and ABC transporters like RecA (Yang *et al.*, 2021). The TagH dimer structure also revealed a positively charged pocket on the inner side of the membrane possibly used to facilitate TagG binding (Yang *et al.*, 2021).

How WTAs are pumped across the membrane is unknown. There have been two models proposed. In the first model, the polymer is recognised by TagH at the non-reducing end and pumped through TagG, followed by the lipid linkage unit (Lazarevic and Karamata, 1995). In the second model the lipid-PP-GlcNAc-ManNAc linkage unit, common to all species with WTA, is first recognised by TagH and then flipped across the membrane with the polymer following behind (**Figure 1.4**) (Schirner, Stone and Walker, 2011). Some evidence has been found which suggests that the latter model is more likely. By substituting the *tagGH* genes in *B. subtilis* 168, which would transport WTAs with mainly GroP units, with the *tarGH* genes from *S. aureus,* which would transport WTAs with mainly RboP units, it was found that the *B. subtilis* 168 phenotype did not change and the WTA polymer was identical to the WT strain (Schirner, Stone and Walker, 2011). This suggests that recognition and translocation of WTA is independent of the main chain polymer. Therefore, recognition of the common linkage unit (lipid-PP-GlcNAc-ManNAc) may only be needed which implies the linkage unit is transported first.

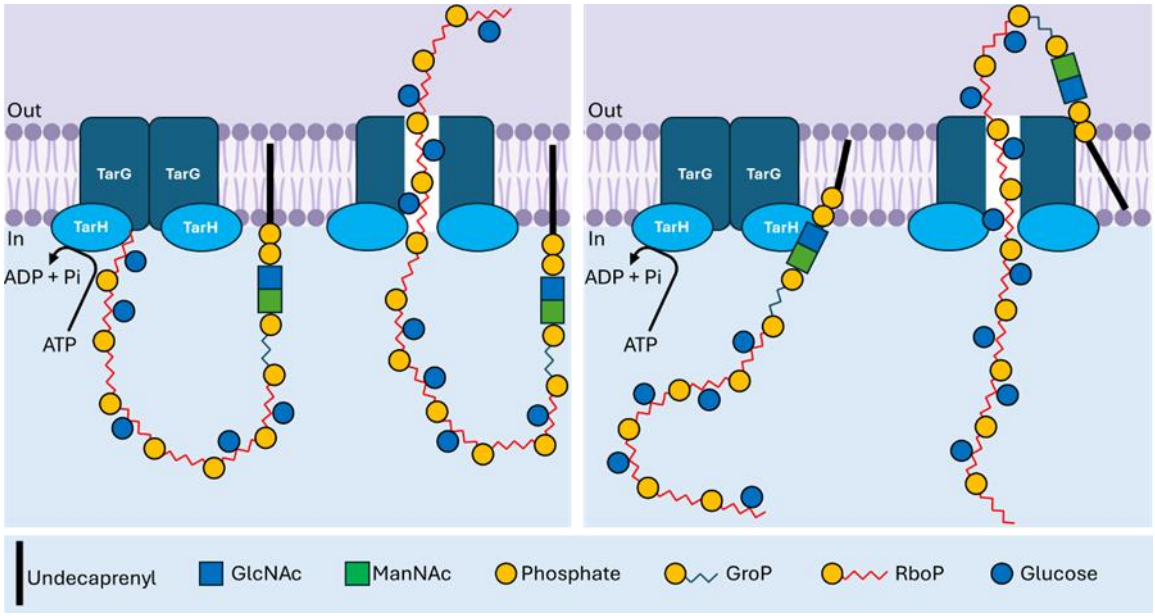


Figure 1.4: Schematic diagram of the two proposed methods of WTA translocation. Modelled here is the transport of the polymer synthesised in B. subtilis W23 using the equivalent enzymes to TagGH, TarGH. A = TarH first recognises the non-reducing end of the polymer and then proceeds with translocation, using energy from ATP hydrolysis, with the linkage unit following. B = TarH first recognises the linkage unit which is then flipped outside the cell via ATP hydrolysis, after which the polymer follows.

1.4.2 TarTUV/TagTUV

After WTA translocation, the polymer is covalently bonded to the PG via a phosphodiester linkage to the C6 hydroxyl of a MurNAc unit. TagTUV has been proposed to carry out this reaction based on genetic evidence of other uncharacterised LytR-CpsA-Psr (LCP) proteins (**Figure 1.5**) (Kawai *et al.*, 2011; Dengler *et al.*, 2012). Gene knockout studies of each of these three genes in *B. subtilis* 168 found that single and double gene knockouts do not drastically change the morphology of the cell under normal growth conditions, however, when all three were disrupted, the cells were not viable (Kawai *et al.*, 2011). Isolation of cell wall material from these mutants also identified a significant reduction in phosphate content in the triple deletion strain when compared to the WT strain, similar to results observed in TagO deletions (Kawai *et al.*, 2011). Three members of the LCP protein family have also been identified in *S. aureus* which may have a homologous function (Over *et al.*, 2011).

Whether or not WTAs are synthesised and attached onto pre-existing mature or nascent PG is still debated. In cryogenic electron microscopy (cryo-EM) experiments, a heavier stained region at the septa of *B. subtilis* and *S. aureus* was identified as WTA which suggests that WTAs are attached to new PG during cell division (Matias and Beveridge, 2005, 2007). When WTA synthesis was inhibited, the stained area receded over time, further suggesting that WTA synthesis proceeds with PG synthesis (Campbell *et al.*, 2011). Other studies utilised fluorescent protein tagging to make protein fusions of various WTA biosynthetic enzymes which showed that the proteins again localise at the septum in *B. subtilis* and *S. aureus* during

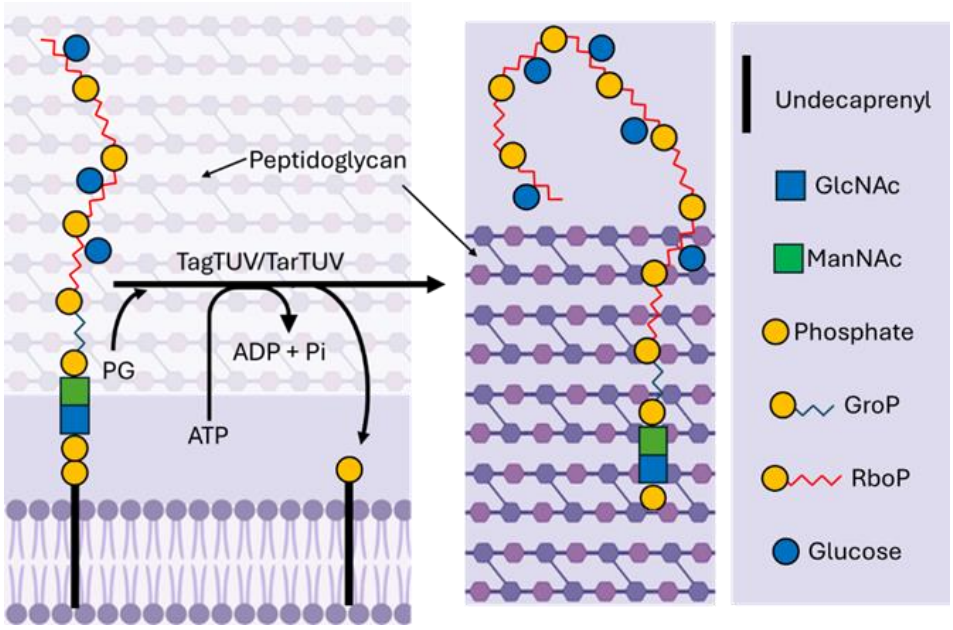


Figure 1.5: Schematic diagram of the supposed enzymes involved in WTA attachment to the peptidoglycan (PG). Linkage is made through a phosphodiester bond to the C6 hydroxyl of the MurNAc unit of PG.

cell division (Bhavsar, Truant and Brown, 2005; Formstone *et al.*, 2008; Atilano *et al.*, 2010). In contrast, other studies have also observed that WTAs are attached to mature PG. Andre *et al.*, 2011, used fluorescently tagged lectins specific to glucose to image WTA localisation in *Lactobacillus plantarum*. They found that no fluorescence could be observed at the septum region of the dividing cells as well as at the newly formed poles of the daughter cells (Andre *et al.*, 2011). Ultimately, other methods of elucidating WTA localisation are required.

1.4.3 DltA, B, C and D

The final step in WTA biosynthesis is the decoration of the WTA polymer with D-alanine. This occurs after WTA translocation, differing from other modifications. Four enzymes have been linked to D-alanylation in LTA and WTA synthesis in many species: DltA through to DltD (Neuhaus and Baddiley, 2003; Koprivnjak *et al.*, 2006). It is thought that in *S. aureus,* direct attachment of D-alanine to the WTA polymer occurs at very low levels. Instead, LTAs are first charged with D-alanine and the alanylated LTAs then become donors to WTA through the action of an enzyme named FmtA (Reichmann, Cassona and Gründling, 2013; Rahman *et al.*, 2016).

DltA is a D-alanine-D-alanyl carrier protein ligase which catalyses the charging of D-alanine using ATP and the transfer of the activated amino acid to the carrier protein DltC (Heaton and Neuhaus, 1994; Debabov *et al.*, 1996). The functions of DltB and D are unclear. DltB has been proposed to facilitate the transfer of D-alanine from DltC to C_{55}-P, to create a similar lipid intermediate seen in the early steps of WTA synthesis, and also transfer this substrate across

the membrane (Perego *et al.*, 1995). However, the speculated lipid intermediate has not yet been identified. Additionally, the structure of DltB in complex with DltC has been found, highlighting a 'funnel' extending into the middle of the membrane, possibly for allowing access for the charged DltC to interface with the lipid inside the membrane for synthesis of the intermediate, or for the charged DltC protein to directly interact with LTAs and/or WTAs (Ma *et al.*, 2018). DltD function is debated. DltD is generally accepted to be a membrane bound protein, however two models have been proposed for its function: one in which DltD acts outside the cell to transfer D-alanine from the lipid intermediate onto either WTA or more likely LTA (Perego *et al.*, 1995), and the other suggesting that the protein acts inside the cell to increase the efficiency of DltA-mediated ligation of D-alanine to DltC (Debabov, Kiriukhin and Neuhaus, 2000; Neuhaus and Baddiley, 2003).

1.5 Functions of Wall Teichoic Acids

1.5.1 Cell Morphology and Division

Many studies report that bacterial strains which lack WTAs have a decreased rate of growth and have various morphological abnormalities (Boylan *et al.*, 1972; Bhavsar, Beveridge and Brown, 2001; Soldo, Lazarevic and Karamata, 2002; Bhavsar, Truant and Brown, 2005). Rod shaped bacteria such as *B. subtilis* also lose their shape without WTAs and are therefore displaying the coccoidal phenotype (Boylan *et al.*, 1972; Bhavsar, Beveridge and Brown, 2001; Soldo, Lazarevic and Karamata, 2002). These data suggest that WTAs are likely required for the correct assembly and localisation of PG biosynthetic enzymes during cell division. Furthermore, many of the enzymes in PG and WTA synthesis seem to colocalise, therefore supporting the possible function of WTAs in regulating nascent PG synthesis (Formstone *et al.*, 2008; Kawai *et al.*, 2011). Additionally, there is evidence that WTA synthesis may be attached to new PG at the septum, further supporting a possible relationship with PG synthesis and therefore cell shape and division (Bhavsar, Truant and Brown, 2005; Matias and Beveridge, 2007; Formstone *et al.*, 2008; Campbell *et al.*, 2011).

In *S. aureus,* the WTA enzymes have been found to interact with the major autolysin Atl responsible for hydrolysis of PG during cell division (Schlag *et al.*, 2010). Without the WTA polymer, Atl is distributed across the cell surface, thereby increasing vulnerability to autolysis. This causes defects in cell separation of the daughter cells, observable by electron microscopy (Campbell *et al.*, 2011). Additionally, PG crosslinking enzymes PBP4 and FmtA lose septal localisation in strains without WTAs (Atilano *et al.*, 2010; Qamar and Golemi-Kotra, 2012). In *B. subtilis,* the same decrease in PG crosslinking is observed in *tarO* deletion strains (Schlag *et al.*, 2010). Furthermore, TarO inhibitors were found to facilitate susceptibility to β-lactams

in MRSA, implying a connection between the biosynthetic pathways of WTAs and highly crosslinked PG which is resistant to antibiotics (Campbell *et al.*, 2011; Farha *et al.*, 2013).

1.5.2 Ion Regulation

Due to the highly charged anionic nature of the WTA polymer, it is thought that WTAs can bind metal cations in the surrounding environment beyond the PG wall. (Wickham *et al.*, 2009; Kern *et al.*, 2010). WTA biosynthesis enzymes are also upregulated in decreased metal ion conditions, suggesting that WTAs act as a scavenging system utilised during colonisation (Endl *et al.*, 1983). Additionally, WTAs bind protons and strains lacking WTA were shown to have a measurable decrease in proton uptake. Because of this, WTAs have been suggested to create local changes in pH, thus possibly altering the activity of certain enzyme present, like autolysins (Biswas *et al.*, 2012). Ion binding may also decrease the repulsive effect of neighbouring charged phosphate groups which can affect the structure of the WTA polymer, and consequently cell wall integrity (Wickham *et al.*, 2009).

1.5.3 Host Defence Evasion and Antibiotic Resistance

As WTA are highly decorated long polymers extending far beyond the cell wall, they contribute to the overall charge and hydrophobicity of the surface of the bacterial cell. This, therefore, will alter the binding of various wanted and unwanted molecules. By altering the modification of the WTA, bacteria can protect themselves from threats and undesirable conditions. Decreased WTA content has been observed to increase bacterial susceptibility to high temperatures and high salt environments, implying that WTAs have a role in regulating temperature tolerance and osmotic stress (Atilano *et al.*, 2010; Holland, Conlon and O'Gara, 2011). In addition, cells without WTAs have an increased susceptibility to antimicrobial fatty acids present in humans likely due to easy access to the less hydrophilic cell wall, therefore facilitating binding and causing cell membrane damage (Kohler, Weidenmaier and Peschel, 2009). Furthermore lysozyme and lysostaphin, which cleave PG, are blocked by the presence of WTAs (Bera *et al.*, 2007; Wu *et al.*, 2019).The same is true for serum amyloid P component which binds to PG and promotes leukocyte-mediated phagocytosis (An *et al.*, 2013).

Resistance to daptomycin-calcium complexes is observed when WTA production and D-alanylation is upregulated, possibly due to the introduction of electrostatic repulsion which limits antibiotic access (Bertsche *et al.*, 2013). This phenomenon is also observed with certain other cationic antimicrobial peptides (CAMPs) (Peschel *et al.*, 1999, 2000; Collins *et al.*, 2002; Walter *et al.*, 2007). The mechanism behind this resistance is unknown, however it has been speculated that the absence of D-alanine increases the overall negative charge of the bacterial surface, therefore promoting CAMP binding and action (Vadyvaloo *et al.*, 2004). In line with

this theory, both *B. subtilis* and *S. aureus* strains which lack D-alanine modification bind positively charge cytochrome *c* more than the WT strains (Peschel *et al.*, 1999).

As mentioned previously, WTAs seem to infer β-lactam resistance in *B. subtilis* and MRSA (Campbell *et al.*, 2011; Brown *et al.*, 2012a; Farha *et al.*, 2013). Further research found that removal of WTA β-GlcNAc modifications restored antibiotic susceptibility, feasibly through acting as a scaffold for proteins involved in resistance (Brown *et al.*, 2012a). Specifically, a transpeptidase found in MRSA named PBP2A was found to selectively interact with WTAs (Qamar and Golemi-Kotra, 2012). As sugar moieties have been identified in phage receptor binding, it possible that other proteins with carbohydrate binding sites may interact with decorated WTA, suggesting further uses as a protein scaffold (Baptista, Santos and São-José, 2008; Xia *et al.*, 2011).

1.5.4 Adhesion and Colonisation

The WTA polymer can interact with a variety of surfaces vital for colonisation such as host epithelial cells and artificial materials. For example, glycosylated WTA in *S. aureus* can interact with the scavenger receptor SREC-I to promote nasal colonisation (Baur *et al.*, 2014; Winstel *et al.*, 2015). Furthermore, WTAs have also been associated with intestinal epithelial cell and umbilical vein endothelial cell colonisation in mouse models, whereby *tagO* deletions in *S. aureus* resulted in unsuccessful infection (Misawa *et al.*, 2015).

Bacteria with WTAs promote biofilm formation, an important part of adhesion and colonisation. Without WTAs or D-alanine WTA modification, biofilm formation is lessened thereby decreasing adhesion to plastic and glass surfaces (M. Gross *et al.*, 2001; Holland, Conlon and O'Gara, 2011). It is thought that the neutralisation of the negative surface charge of WTAs by D-alanylation allows greater adhesion to uncharged surfaces like glass and plastic through hydrophobic interaction (M. Gross *et al.*, 2001). For this reason, D-alanine esters on WTAs are considered virulence factors as their absence does not cause major cell defects but does decrease pathogenicity (Andre *et al.*, 2011; Campbell *et al.*, 2011). Additionally, WTAs have also been classified as virulence factors, even though the absence of WTAs causes defects *in vitro,* as mutants lacking WTAs have a drastic decrease in survivability inside host cells due to all the above functions (Weidenmaier *et al.*, 2004; Walter *et al.*, 2007).

1.6 Antivirulence and Inhibition of WTA Biosynthesis

The various stages of WTA synthesis have been suggested as targets for new antibiotics and antivirulence treatments for MRSA and other pathogenic *Staphylococci*. Strains in which early synthetic enzymes are knocked out, like *tarO*, result in bacteria unable to colonise and infect animal models; therefore, inhibitors of these enzymes may yield novel antibiotics (Misawa *et al.*, 2015). Furthermore, the enzymes responsible for D-alanylation, DltA, B, C and D, may also be targets for inhibition as their deletion attenuates pathogenicity, colonisation and biofilm formation (M. Gross *et al.*, 2001). This is likely caused by the increased negative charge of the WTA polymer when D-alanylation is halted which increases repulsive forces between the cell surface and the stratum (M. Gross *et al.*, 2001). However, it is argued that these types of targets will not successfully treat established infections but instead could be used to decrease virulence of antibiotic resistance strains.

Inhibitors of TarO have been identified, for example tunicamycin (Hancock, Wiseman and Baddiley, 1976; Campbell *et al.*, 2011) which has been used to observe the effects of *in vitro* inhibition of WTA biosynthesis. Unfortunately, tunicamycin also inhibits GlcNAc phosphotransferase (GPT) which is vital for *N*-linked glycan biosynthesis in animals and can therefore not be used as an antibiotic (Price and Tsvetanova, 2007). More nontoxic inhibitors are required which specifically target the synthesis of WTA or the D-alanine modification steps.

One highly druggable target which has been identified is TarG. As a transmembrane transporter, it is much more accessible to inhibitors than the cytoplasmic enzymes involved in WTA biosynthesis (Wang *et al.*, 2013). Studies have reported that the only inhibitors identified as disrupting late-stage WTA synthesis are all compounds which block the presumed substrate channel formed by heterodimeric TarGH, therefore disallowing translocation across the membrane of the complete WTA polymer (Swoboda *et al.*, 2009; Wang *et al.*, 2013).

Another possible strategy of combating antibiotic resistant strains is through using WTA inhibitors in combination with β-lactams. As mentioned previously, resistant strains which either lack WTAs or lack glycosylated WTAs become susceptible to β-lactams (Campbell *et al.*, 2011; Brown *et al.*, 2012a; Farha *et al.*, 2013; Lee *et al.*, 2016). Therefore, if inhibitors for mid to late-stage WTA synthesis were identified and supplemented with β-lactam antibiotics the combination could revert antibiotic resistance (Lee *et al.*, 2016). When tested *in vitro*, the synergy of TarG inhibitors with Imipenem (a newer class of β-lactam antibiotics) was unfortunately lower than predicted (Wang *et al.*, 2013). This could be because β-lactams are only effective against actively growing bacteria and that WTA inhibitors have a bacteriostatic effect, thus acting somewhat antagonistically with β-lactam antibiotic action (Campbell *et al.*,

2012). Contrary to this, the combined treatments, in arguably high dosages, have favourable effects in animal models of MRSA infection by decreasing the bacterial burden to levels associated with Linezolid and Ciprofloxacin treatments (Wang *et al.*, 2013). Another set of inhibitors for TarO, tarocin A and tarocin B, was also found to restore broad-spectrum β-lactam antibiotic susceptibility in MRSA in mouse models (Lee *et al.*, 2016). Additionally, TarS could be another target for inhibitors which supplement β-lactam antibiotics. Without sugar moieties present on the WTA polymer, like in TarS knockout mutants, PBP2A (the active transpeptidase in MRSA) may not selectively interact with WTAs which could remove some resistance towards β-lactams (Brown *et al.*, 2012a). However, despite the structure of TarS having been elucidated, no inhibitor screens have been performed (Guo *et al.*, 2021).

All WTA biosynthetic enzymes from TarB to TarH are essential and therefore are targets for novel antibiotics. As mentioned, TarG inhibitors have had the most promise likely due to its accessible position as a transmembrane transporter protein. However, the mid-stages of WTA synthesis, specifically the enzymes responsible for RboP priming and polymerisation, TarK and TarL, have not been well characterised and inhibitors have not yet been discovered. Recently, the structure of TarL from *S. aureus* (*Sa*TarL) has been solved but potential inhibitors remain unknown (Li *et al.*, 2024). Further work is required to uncover the mechanisms of catalysis and solve the atomic structures of TarK and TarL homologues from other bacterial species, such as the model organism *B. subtilis* W23, to facilitate the development of novel chemical probes and inhibitors as tools for further research.

1.7 Aims

The aim of the research in this thesis is to structurally characterise the enzymes responsible for RboP incorporation in WTA biosynthesis (TarK and TarL) and develop assays for the analysis of their activity *in vitro*. Furthermore, inhibitors will be tested to attempt to quantify promising compounds with the later aim to elucidate the 3D protein structure with theses inhibitors bound. If successful, this will act as a basis for structure-guided design of antibacterial agents for therapeutically relevant bacteria. Mutants will also be designed and tested to allow insights into the catalytic action of these enzymes.

Chapter 2 describes the development of a combined chemical and biological approach to the synthesis of CDP-ribitol, the donor required for TarK and TarL activity. TarI will be recombinantly overexpressed, purified and used for the synthesis of CDP-ribitol from RboP by other members of the group. Purified CDP-ribitol will then allow the development of activity assays for TarK and TarL to proceed.

Chapter 3 describes the identification and assessment of various TarK and TarL homologs. Sequence and structural analysis will be performed to highlight conserved residues of interest among well-characterised TagF-like enzymes for later mutagenesis studies. In addition, small scale expression and purification of various homologues will establish suitable methods to produce the target proteins, BsTarK and BsTarL, in soluble form.

Chapter 4 describes the methods used to optimise WT and mutant *Bs*TarK expression and purification. Various tagged constructs will be cloned, and various expression conditions will be tested. After the optimised plasmid construct is cloned and expression and purification protocols are established, the solubilised protein will take part in X-ray protein crystallography trials and characterisation of the protein through circular dichroism (CD) and nano differential scanning fluorimetry (nanoDSF). CDP-Rbo inhibitor analogues from our collaborators will also be tested using nanoDSF to observe melting temperature shifts suggestive of binding. Early LC-MS activity assays will also be performed with the purified enzyme to confirm enzyme activity.

Chapter 5 describes methods used to optimise WT and mutant *Bs*TarL expression and purification. Various tagged constructs will be cloned, and various expression conditions will be tested. After the optimised plasmid construct is cloned and expression and purification protocols are established, the solubilised protein will take part in X-ray protein crystallography trials to solve the atomic structure. Characterisation of the protein by CD and nanoDSF will also be performed. To test for larger oligomeric *Bs*TarL proteins which have been observed in the literature in homologous proteins, mass photometry will be used. If larger oligomeric

structures are detected, this would allow to cryogenic electron microscopy to be a viable method of elucidating the protein structure.

Chapter 2: TarI Expression and Purification

2.1 Introduction

TarI is a member of a large family of cytidylyltransferases which span numerous eukaryotic and prokaryotic species (Zolli, Kobric and Brown, 2001). Specifically, TarI is responsible for the catalysis of CDP-ribitol (CDP-Rbo) formation using cytidine 5′-triphosphate (CTP) and ribitol-5-phosphate (RboP) in some species of gram positive bacteria including *Bacillus subtilis* W23 and *Staphylococcus aureus* (**Figure 2.1**) (Pereira and Brown, 2004). The *tarI* gene in *B. subtilis* was first proposed as a CDP-Rbo synthase by sequence alignment with the *H. influenza* gene, *acs1* (Lazarevic *et al.*, 2002). Further studies on the TarI gene found in *S. aureus* then confirmed TarI activity as a RboP cytidylyltransferase (Pereira, D'Elia, *et al.*, 2008). TarI has been proposed to act in tandem with TarJ, a medium-chain alcohol dehydrogenase/reductase which reduces ribulose-5-P (RibP) to RboP using NADPH (Pereira and Brown, 2004). Structural data has shown that the TarI/J pair form a heterodimer and computational experiments have proposed that the proteins associate further to form a heterotetramer, however there is no channelling of the RibP substrate from TarJ to TarI (Pereira and Brown, 2004; Li *et al.*, 2021b).

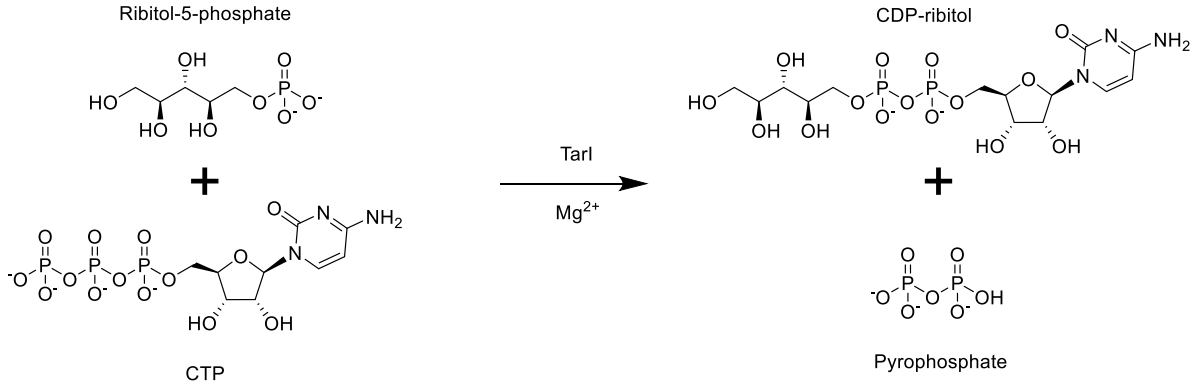


Figure 2.1: CDP-ribitol synthesis performed by TarI. Ribitol-5-phosphate is produced by TarJ and charged with CTP to form the donor substrate for WTA synthesis, CDP-ribitol.

The product of the TarI reaction, CDP-Rbo, is required for the synthesis of the RboP polymers present in wall teichoic acids (WTA) which are produced by ribitol phosphate transferase enzymes (TarK and TarL) in *B. subtilis* W23 and *S. aureus.* WTAs are involved in many functions such as cell shape determination, cell division, biofilm formation, cell adhesion and colonisation and antibiotic resistance, therefore, inhibition of the WTA biosynthetic pathway, like the synthesis of CDP-Rbo or the RboP polymer, could lead to novel antimicrobials and antivirulence compounds (Matthias Gross *et al.*, 2001; Soldo, Lazarevic and Karamata, 2002; Weidenmaier *et al.*, 2004; Brown *et al.*, 2010; Farha *et al.*, 2013; Wu *et al.*, 2021). This is

supported by evidence that TarI knockout mutants in *S. aureus* result in cell death (Meredith, Swoboda and Walker, 2008). Moreover, allosteric inhibitors of other cytidylyltransferases, namely 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferases (IspD) in *Arabidopsis thaliana* and *Plasmodium vivax,* have already been discovered, but the product produced, CDP-4-diphosphocytidyl-2*C*-methyl-D-erythritol (CDP-ME), is unrelated to WTA synthesis and the compounds may not translate well to TarI inhibition studies (Witschel *et al.*, 2011; Kunfermann *et al.*, 2014; Schwab *et al.*, 2017).

Further research is required to increase our understanding of the biosynthetic pathways involved in WTA production to enable the discovery of inhibitory compounds and combat the increasing antibiotic resistance threat.

2.1.1 Aims

To support characterisation experiments and structural studies involving TarK and TarL as described in Chapters 3-5, it is paramount that the CDP-Rbo donor is available. While producing CDP-Rbo through chemical synthesis is challenging, the chemical production of the TarI substrate RboP is straightforward through the ring opening and reduction of the much cheaper and commercially available sugar, ribose 5-phosphate. Therefore, the chemoenzymatic synthesis of CDP-Rbo can be achieved using chemically synthesised RboP, CTP and TarI (Pereira and Brown, 2004; Brown, Zhang and Walker, 2008; Baur *et al.*, 2009; Brown *et al.*, 2010). Hence, this Chapter pursues the overexpression and protein purification of TarI required for CDP-Rbo synthesis.

Initially, the recombinant gene products need to be synthesised, and expression performed to express soluble TarI. If successful, larger scale cultures will be used to produce the target protein which will then be isolated using multistep column protein purification techniques.

2.2 Results and Discussion

2.2.1 Protein Expression and Purification

Using the expression and purification methods outlined in section **2.4.1, Table 2-1,** immobilized metal affinity chromatography (IMAC) was first used to isolate the target protein from sonicated resuspended cell paste lysate (see **2.4.1** for expression conditions and lysis methods). Following purification on a Nickel (HisTrap) affinity column, samples of the soluble (cleared lysates) and insoluble (cell debris) fractions prior to purification, along with flow through, wash and elution fractions of interest were prepared for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis (see **7.2.8**) (**Figure 2.2**).

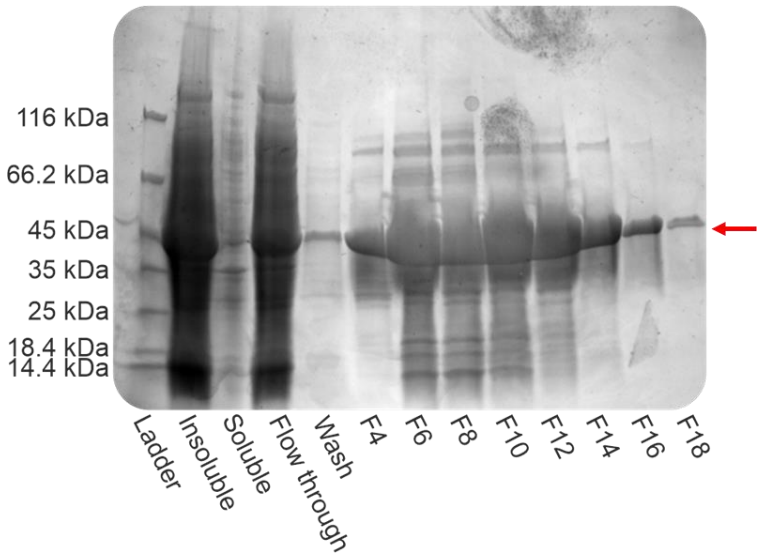


Figure 2.2: First IMAC purification of His-SUMO-TarI from E. coli lysate. Red arrows indicate the presence of target protein in the eluted fractions (expected mass ~40 kDa). Resolved on a 4-15% polyacrylamide gel.

From the gel, intense bands at around the expected mass of His-SUMO-TarI (~40 kDa) can be observed in fractions 4-18. This indicates successful capture and elution of the His6 tagged protein from the cell lysate. However, within the fractions which contain the target protein, there are impurities observed at around 14, 18, 35 and 66 kDa, and therefore further purification steps are required. Also, in the flow through and wash fractions a band is observed around 40 kDa. This suggests that some of the target protein did not bind strongly to the column. This may be due to the very high soluble expression levels of this protein which may result in the HisTrap column reaching its binding capacity. To combat this issue, a smaller amount of cell pellet could be used for lysis, or several columns could be attached in series to improve binding capacity. Additionally, the insoluble fraction also appears to contain the target protein represented by a band at ~40 kDa. This may again be caused by a high level of protein expression which may cause protein aggregation. To limit this, a lower isopropyl β-D-1-

thiogalactopyranoside (IPTG) could be used to induce protein expression or the induction time of the expression could be decreased. Regardless, a large amount of soluble protein has been captured.

Following the first successful capture step, the fractions containing His-SUMO-TarI were combined, and the SUMO tag was cleaved off using the SUMO protease protocol (Invitrogen). To separate the His-SUMO tag from the now native TarI, a second IMAC step was performed (**2.4.1**). As the His-SUMO retains the His6 tag, the target protein should be observed in the flowthrough, while just the His-SUMO tag should be observed in the elution fractions. The completed reaction mixture was loaded onto the column and purification proceeded. After the purification was completed, samples of each fraction from the IMAC purification were prepared for SDS-PAGE analysis (see **7.2.8**) (**Figure 2.3**).

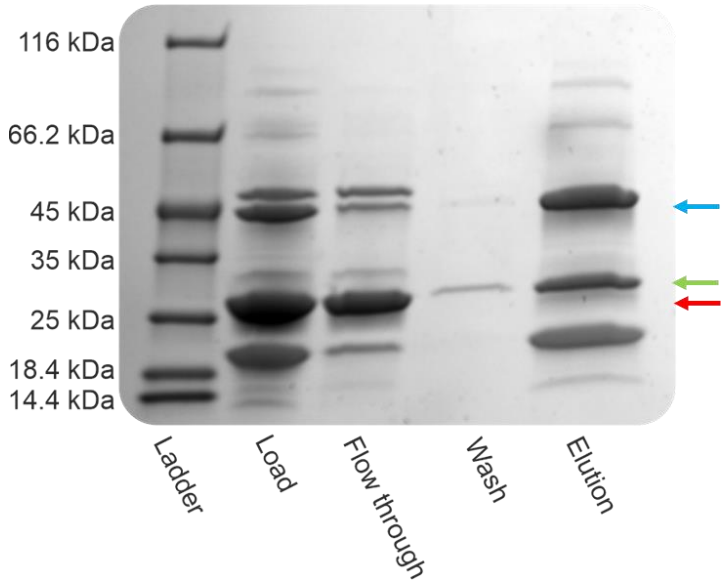


Figure 2.3: Second IMAC purification of the SUMO protease treated His-SUMO-TarI from the previous purification. The red arrow indicates the presence of the digested target protein, native TarI, in the flow through fraction (expected mass ~27 kDa). The blue arrow represents undigested protein (~40 kDa). The green arrow may represent digested target protein or more likely the SUMO protease (~28 kDa). Resolved on a 4-15% polyacrylamide gel.

An intense band is observed in the flow through lane which is located at the correct mass expected for the cleaved target protein (~27 kDa). However, the same band can also be observed in the elution lane. One reason for this could be that the cleaved TarI protein has some affinity for the column, however a more likely explanation is that this band represents the His6 tagged SUMO protease enzyme. The protease, coincidentally, has a similar mass to the native TarI protein (~28 kDa) and carries a His6 tag, and therefore this band likely contains the protease. Another band of note in the elution lane is the intense band at ~45 kDa. This band probably consists of undigested His-SUMO-TarI protein as it is present in the elution

lane, and therefore likely retains a His6 tag, and it has a similar mass to the fusion protein (~40 kDa). The reason to why some of the His-SUMO-TarI protein was not cleaved may be caused by decreased activity of the SUMO protease enzyme. This decrease in activity could be caused by the temperature used for the incubation step or by the age of the protease stock. To prevent this in the future, fresh protease should be produced, or a larger excess of protease could be added to the protein solution to encourage full cleavage of the target fusion protein. Nonetheless, a large amount of native TarI has been separated from the undigested His-SUMO-TarI. The flowthrough sample still contains many impurities which are represented by bands observed at ~20, ~30 and ~45 kDa, indicating that further purification is required.

The flow through of the previous IMAC purification was subsequently purified by size exclusion chromatography (SEC) (see **2.4.1**). After the purification was completed, peaks in the 280 nm absorption spectrum were identified and eluted fractions under these peaks were combined and sampled. The fraction samples and a sample taken from the loaded mixture (before concentrating to 2 mL) were prepared for SDS-PAGE analysis (**7.2.8**) (**Figure 2.4**).

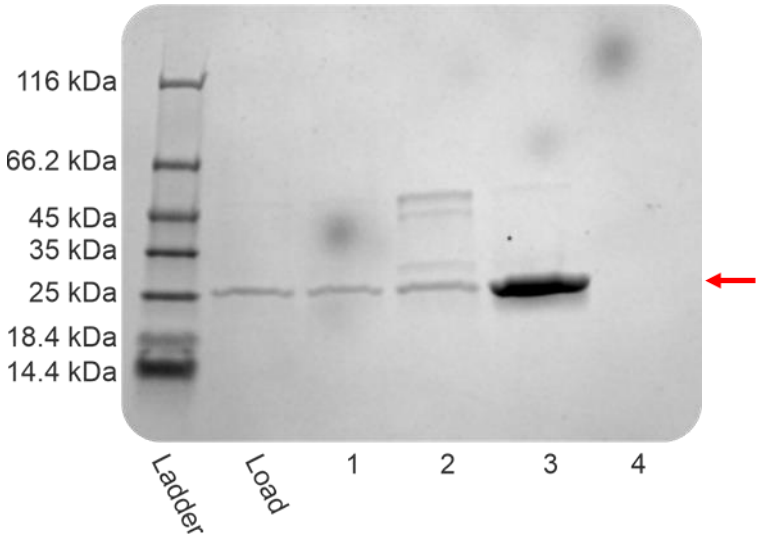


Figure 2.4: SEC of TarI from the previous purification. The red arrow indicates the presence of the digested target protein, native TarI (expected mass ~27 kDa). Resolved on a 4-15% polyacrylamide gel.

A single intense band at the expected mass of TarI (~27 kDa) is observed in the third pooled SEC fraction. This suggests that SEC purification was successful in separating TarI from the other impurities observed in the previous purification step. However, this band is also observed in the other two combined samples suggesting that the TarI protein also eluted in these fractions. Conversely, these bands are much less intense than the single band in sample three; thus, the majority of the TarI protein is present in sample three. The purified protein was concentrated by spin concentration (see **7.2.7**) to ~1.4 mg/mL and stored at -70 °C in 100 µL aliquots (6mL total) for use in the chemoenzymatic synthesis of CDP-Rbo.

With TarI successfully expressed and purified, the protein was used to produce CDP-Rbo by postdoctoral research associate Saeed Akkad. For this purpose, synthetic RboP was reacted with purified TarI in the presence of CTP overnight at 37ºC. Liquid chromatography-mass spectrometry (LC-MS) monitoring revealed correct product formation (m/z 537.08), after which the product was purified by treatment with alkaline phosphatase (to remove unreacted RboP), anion exchange chromatography and purification on a Teledyne ACCQ Prep high-performance liquid chromatography (HPLC) system. The purified material was used as a donor for TarK and TarL reactions described in Chapters 3-5.

2.3 Conclusion and Future Work

The expression, and purification of soluble TarI described in this chapter was essential in synthesising CDP-Rbo which is required for various downstream experiments, described in the later chapters, involving the TarK and TarL enzymes. The overexpression of the enzyme was straightforward, and the 3-step purification procedure proved highly effective in isolating the target protein. However, optimisations could be performed. First, the SUMO cleavage step did not cleave all the fusion protein. To increase the yield of this step, fresh SUMO protease should be used and, if the yield remains low, a screen involving the evaluation of different buffers, reaction temperatures and incubation times could be performed to elucidate the ideal conditions for this cleavage reaction. Secondly, the SEC purification could be refined by altering the flow rate and column type to encourage all the target protein to elute at the same retention volume for increased resolution. Nevertheless, the purification methods shown here yielded a suitable amount of pure protein which allowed other members of the group to trial various conditions for the successful synthesis of CDP-Rbo using TarI, CTP and RboP.

2.4 Methods

2.4.1 Protein Expression and Purification

The pSUMO-SaTarI plasmid (see **Table 7-1** entry 22), previously engineered by postdoctoral research associate Saeed Akkad, was transformed into BL21(DE3) *E. coli* (see **7.2.5**) and colonies from the selection plate were used to inoculate a starter culture of 100 mL of kanamycin LB. This culture was incubated overnight at 37 °C, 180 RPM. 6 mL of this culture was used to inoculate larger 600 mL kanamycin LB cultures which were incubated at 37 °C, 180 RPM until the OD_{600} reached ~0.6. Uninduced samples were taken, and the cultures were induced with the addition of IPTG at a working concentration of 0.5 mM. Protein expression then proceeded at 18 °C, 180 RPM overnight. The cultures were then centrifuged at 6,000 xg to produce cell pellets and the supernatant was discarded. These pellets were then stored at -70 °C for future purification.

The three-step purification method used for the purification of TarI is outlined in **Table 2-1**. The buffers for each step are listed in **Table 2-2**.

Table 2-1: Outline of the method used for purifying TarI

Method no.	Protein	Step 1		Step 2		Step 3	
		Column method	Buffers used	Column method	Buffers used	Column method	Buffers used
1	His-SUMO-TarI	IMAC (Ni^{2+})	His A/B	IMAC (Ni^{2+})	His A/B	SEC	SEC

Table 2-2: List of buffers used for purifying TarI

Buffer no.	Name	Buffer component conc	Salt conc	pH	Imidazole conc	Additives
1	His A	50 mM Tris	500 mM NaCl	6.5	10 mM	N/A
2	His B	50 mM Tris	500 mM NaCl	6.5	500 mM	N/A
3	SUMO	50 mM Tris	150 mM NaCl	6.5	N/A	5 mM DTT
4	SEC	50 mM Tris	150 mM NaCl	6.5	N/A	5 mM DTT

For lysis, the cell pellet was resuspended in His A buffer (**Table 2-2**) supplemented with protease inhibitors, benzonase and lysozyme, and was lysed using sonication (Fisherbrand™ 505 Sonicator) at 40% amplitude for 2 mins with 30 secs on and 30 secs off. Insoluble cell debris was pelleted by centrifugation (38,000 xg for 30 mins) and the soluble fraction was collected. Cell lysate was subsequently loaded onto a 5 mL HisTrap FF Crude column (Cytiva),

equilibrated with His A buffer, using an AKTA Go for IMAC purification. Flow through was collected and the column was washed with 15 column volumes (CV) of His A buffer. Proteins were then eluted off the column with a linear gradient of 0-100% of His B buffer (**Table 2-2**) over 20 CV. 3 mL fractions were collected. The soluble and insoluble fractions, flow through, wash and elution fractions of interest were sampled and prepared for SDS-PAGE analysis (**7.2.8**).

To remove the SUMO tag, the fractions containing TarI were combined and buffer exchanged into SUMO buffer (**Table 2-2**). The solution was then subjected to SUMO protease cleavage by adding 10 µg of SUMO protease per 1 mg of protein (**7.2.6**). The protease reaction proceeded at 20 °C at 200 RPM overnight. The completed reaction mixture was then loaded onto a 5 mL HisTrap FF Crude column (Cytiva), equilibrated with His A buffer, and purified using an AKTA Go. The flow though was collected, and the column was washed with 5 CV of His A buffer. Bound proteins were subsequently eluted off the column with 5 CV of 100% His B buffer. Samples of the flow through, wash and elution were taken and were prepared for SDS-PAGE analysis (**7.2.8**).

The fractions which contained the target protein were combined and buffer exchanged into SEC buffer (**Table 2-2**) (see **7.2.7**). The sample was concentrated to ~2 mL and applied onto a Superdex 75 pg 16/600 column which was previously equilibrated with SEC buffer. The sample was centrifuged to remove any precipitate and subsequently injected into the system. The purification proceeded at 1 mL/min with SEC buffer until 130 mL was washed through. 2 mL fractions were collected, and samples were taken from fractions of interest defined by the UV trace. These fractions were prepared for SDS-PAGE analysis (see **7.2.8**). The fractions containing pure TarI were combined, concentrated to ~1.4 mg/mL (see **7.2.6**) and stored as 100 µL flash frozen aliquots at -70 °C.

Chapter 3: Investigating TarK and TarL Homologues

3.1 Introduction

TarK and TarL are responsible for the addition of the poly-ribitol phosphate structure present in the wall teichoic acids (WTA) of *Bacillus subtilis* W23 and *Staphylococcus aureus* synthesised within the cytoplasm of bacterial cells (Brown *et al.*, 2010). These enzymes belong to a of monotopic membrane proteins known as the TagF-like family. This family contains other enzymes which are responsible for the previous steps of WTA synthesis such as TagB and TarF which both install glycerol-phosphate (GroP) successively on the substrate lipid-P-P-GlcNAc-ManNAc (Brown *et al.*, 2010). TagF, from which the family is named, polymerises GroP on the same substrate in bacterial species in which glycerol-phosphate is the primary constituent of WTAs like *B. subtilis* 168 and *S. epidermidis* (Pereira, Schertzer, *et al.*, 2008; Lovering *et al.*, 2010). In contrast, TarK and TarL install D-ribitol-5-phosphate (RboP) sequentially on the acceptor substrate, lipid-P-P-GlcNAc-ManNAc-GroP$_x$ (x = 1 or 2 for *B. subtilis* W23 and *S. aureus,* respectively), using the donor substrate CDP-ribitol (CDP-Rbo) (Brown *et al.*, 2010). This process diverges between Gram-positive bacterial species, with both *B. subtilis* W23 and *S. aureus* TagB priming the lipid disaccharide substrate with the first GroP residue but subsequent addition of pentitol-phosphate units being catalysed by different enzymes. TarK primes the mono-GroP acceptor with the first RboP in *B. subtilis* W23 (*Bs*TarK), while TarF adds a second GroP residue in *S. aureus* using the substrate CDP-glycerol. Subsequently, TarL polymerises RboP in both species, with *S. aureus* TarL (*Sa*TarL, as opposed to *Bs*TarL in *B. subtilis*) acting without TarK to prime the first RboP residue and subsequently polymerise the chain (Brown *et al.*, 2010) (**Figure 3.1**). In both species, this poly-RboP chain is of about 20-40 residues long, *in vivo* (Brown *et al.*, 2010; Swoboda *et al.*, 2010). After the assembly of the WTA polymer, which is still attached to the lipid carrier, it is transported through the membrane using an ATP binding cassette transport complex called TarGH by a process that is not yet known (Lazarevic and Karamata, 1995; Schirner, Stone and Walker, 2011). After transportation, the mature polymer is attached to the peptidoglycan

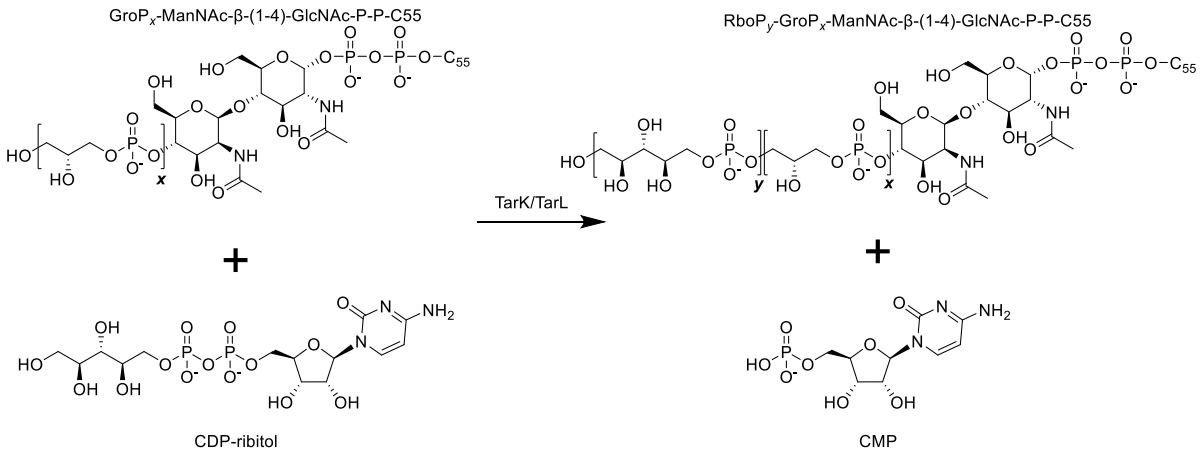


Figure 3.1: Ribitol-phosphate transferase reactions catalysed by B. subtilis W23 and S. aureus TarK and TarL. x = 1 for B. subtilis W23, with GroP addition being catalysed by TagB; x = 2 for S. aureus, in which GroP addition is catalysed by TagB and then TarF. y = 20-40: in B. subtilis W23, TarK catalyses addition of one RboP after which TarL adds additional RboP units. In S. aureus, TarL both primes and polymerises the RboP chain.

(PG) where it extends beyond the cell wall into the environment and becomes decorated with D-alanine, while the lipid carrier is recycled back to the cytoplasmic side of the cell membrane (Kawai *et al.*, 2011).

WTAs have many functions, mainly in cell growth, division and morphology as well as colonisation and adhesion in pathogenic bacteria (Swoboda *et al.*, 2010; Brown, Santa Maria and Walker, 2013b). Additionally, WTAs impart protection against host defences that function to cleave the cell wall and/or disrupt the cellular membrane (Bera *et al.*, 2007; Wu *et al.*, 2019). Multiple studies have shown that inhibition of enzymes involved in the biosynthetic and export pathways results in no observable WTAs and other studies have also observed that disrupting the biosynthesis of WTAs in mutant methicillin-resistant *Staphylococcus aureus* (MRSA) strains restores β-lactam sensitivity (Hancock, Wiseman and Baddiley, 1976; Swoboda *et al.*, 2009; Brown *et al.*, 2012b). A combined approach utilising broad spectrum antibiotics and inhibitors of WTA production together has been shown to have favourable effects in animal *in vivo* models of MRSA infection and colonisation (Farha *et al.*, 2013; Lee *et al.*, 2016).

Further research is required to broaden our understanding of WTA synthesis and allow the groundwork to be laid for further development of WTA-targeting therapeutic strategies to combat the growing resistance to antibiotics. Therefore, it is important to explore the functions of the various biosynthetic enzymes. More specifically, TarK and TarL are of importance as TarK and TarL knockout strains in both *B. subtilis* W23 and *S. aureus* are non-viable, hence inhibition of these enzymes may act as a novel antibiotic strategy (Brown *et al.*, 2010). This vulnerability may be caused by the accumulation of toxic intermediates or by the depletion of the lipid phosphate precursor which is vital for peptidoglycan synthesis (Bouhss *et al.*, 2008). Currently, inhibitors of TarK and TarL have not yet been discovered despite recent advances in ribitol phosphate transferase structural and biochemical characterisation. Thus, a combined

approach of chemical probe/inhibitor synthesis and structural and enzymological biology is required to further understand this process and allow for structure-based drug design.

3.1.1 Aims

To enable structural and enzymological studies of TarK and TarL, it is necessary to establish workflows for the production and purification of the recombinant proteins. In this chapter, expression trials will be performed to find suitable conditions for the isolation of soluble TarK and TarL.

First, bioinformatical techniques will be used to search for homologous gene sequences encoding putative TarK and TarL enzymes and align and identify conserved residues using protein sequence alignment tools. The 3D structures of proteins that do not have a published experimental structure, such as *Bs*TarK and *Bs*TarL, will be predicted by AlphaFold2. The homologues identified will then be superimposed onto the experimentally found TarL structure to highlight structurally conserved residues and other domains of interest (Li *et al.*, 2024). Following homologue identification, the initial targets, *Bs*TarK/L and *Sa*TarL, and the enzymes with the highest sequence identity, will be expressed in *Escherichia coli* with His6 tags for protein expression trials. Promising soluble proteins will then be selected and preliminary immobilised metal affinity column purification will be performed to isolate the proteins of interest.

3.2 Results and Discussion

3.2.1 Homologue Selection

The expression of *B. subtilis* W23 TarK and TarL has been reported in the literature, however, preliminary experiments in this work revealed that the protein expression levels were very low (Brown *et al.*, 2010). Hence, other homologues were explored with the aim to find TarK/L proteins with better expression levels.

The Basic Local Alignment Search Tool (BLAST) database was first used to elucidate close homologs to TarK and TarL for preliminary expression trials (**Table 3-1**). Six homologues were chosen, three with high percent identities (the percentage of the protein sequence that is identical over its entire length) to *Bs*TarL of 65% (*Bacillus pumilus*, BpuTarL), 79% (*Bacillus velezensis*, *Bv*TarL) and 87% (B*acillus paralicheniformis*, *Bpa*TarL) and three with high percent identities to *Bs*TarK of 51% (*Bacillus sp.* TH007, *Bs*TH007TarK), 63% (*Bacillus gobiensis*, *Bg*TarK) and 79% (*Bacillus velezensis*, *Bv*TarK) (**Table 3-1**). The protein sequences of these proteins were compiled from GenBank (NIH) (see **3.4.1**).

Table 3-1: A list of the top three homologs of BsTarK and BsTarL. All sequences were acquired from GenBank, and their protein names and accession numbers are shown. Theoretical pI was calculated using the Compute pI/Mw tool from the Expasy server (Gasteiger et al., 2005).

Species	Protein Name	% identity	GenBank Accession number	Bp	MW (kDa)	MW (kDa) including His tag	Theoretical pI
***Bacillus subtilis* (W23) TarL**	polyribitolphosphotransferase	N/A		1860	72.60	74.89	7.72
Bacillus paralicheniformis	Teichoic acid poly(ribitol-phosphate) polymerase	87%	TWM36451.1	1302	50.98	53.27	6.26
Bacillus velezensis	polyribitolphosphotransferase	79%	AVB11903.1	1842	72.35	74.64	9.15
Bacillus pumilus	polyribitolphosphotransferase	65%	AOC58576.1	1836	72.60	74.88	6.39
***Bacillus subtilis* (W23) TarK**	ribitolphosphotransferase	N/A		1170	45.00	47.27	6.54
Bacillus velezensis	CDP-glycerol glycerophosphotransferase	79%	WP_238994089.1	3030	118.40	120.67	8.97
Bacillus gobiensis	ribitolphosphotransferase	63%	ALC84247.1	1170	45.17	47.46	8.14
Bacillus sp. TH007	ribitolphosphotransferase	51%	KRV44443.1	1164	45.80	48.09	8.64

All the top homologues found appear to be of similar molecular weight between TarK homologs and TarL homologs, except for *Bv*TarK which has a mass of ~118 kDa. Coincidently, the sum of the masses of *Bs*TarK and *Bs*TarL equals ~117 kDa, possibly indicating that *Bv*TarK contains both TarK and TarL enzymes. This is further supported by the domain

organisation of *Bv*TarK which highlights that the TarK GT-B domain is upstream from the domains identical to the whole TarL protein (**Figure 3.2**). Therefore, *Bv*TarK could act as both a RboP primase and polymerase similar to *Sa*TarL (Brown *et al.*, 2010). However, *Sa*TarL achieves both functions without this supposed gene fusion and has a similar mass to other TarL homologs. Activity assays of *Bv*TarK are required to determine if the functions are indeed similar *Sa*TarL.

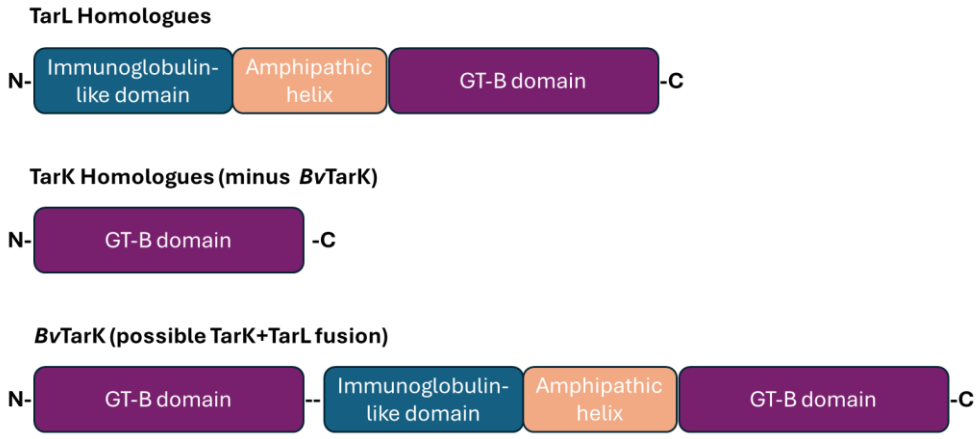


Figure 3.2: Illustration of the different domain organisation of the TarL and TarK homologues and the possible BvTarK fusion protein.

3.2.2 Alignments and Structural Similarities

To highlight areas of sequence interest, such as conserved catalytic residues and motifs which would provide evidence of a similar function to TarK and L, TarK and TarL sequences were aligned using Clustal Omega (EMBL-EBI) (**Figure 3.3** and **Figure 3.4**) (see **3.4.2**). There are large areas of homology across the selected sequences. Specifically, in TarK and TarL protein sequence alignments, the conserved WH, HP and T(S in some TarL homologs)DYSSV(L in *Sa*TarL)XXD/E motifs are also conserved in the glycerol-phosphate polymerase enzyme TagF from *Staphylococcus epidermidis* (*Se*TagF) (Lovering *et al.*, 2010) (**Figure 3.3** and **Figure 3.4**). The donor substrate in *Se*TagF is CDP-glycerol which is similar in structure to CDP-ribitol, the natural donor substrate of TarK/L enzymes. Structural analysis and mutagenesis studies of *Se*TagF highlighted two conserved His residues within the WH and HP motifs as catalytic residues (Lovering *et al.*, 2010). Therefore, it is likely that these residues in TarK/L are also catalytic and could be mutated in later mutagenesis studies to confirm their catalytic function. Furthermore, the conserved TDYSSVXXD/E motif in *Se*TagF typically coordinates the nucleotide in the sugar donor in Glycosyltransferase B (GT-B) members, suggesting that in

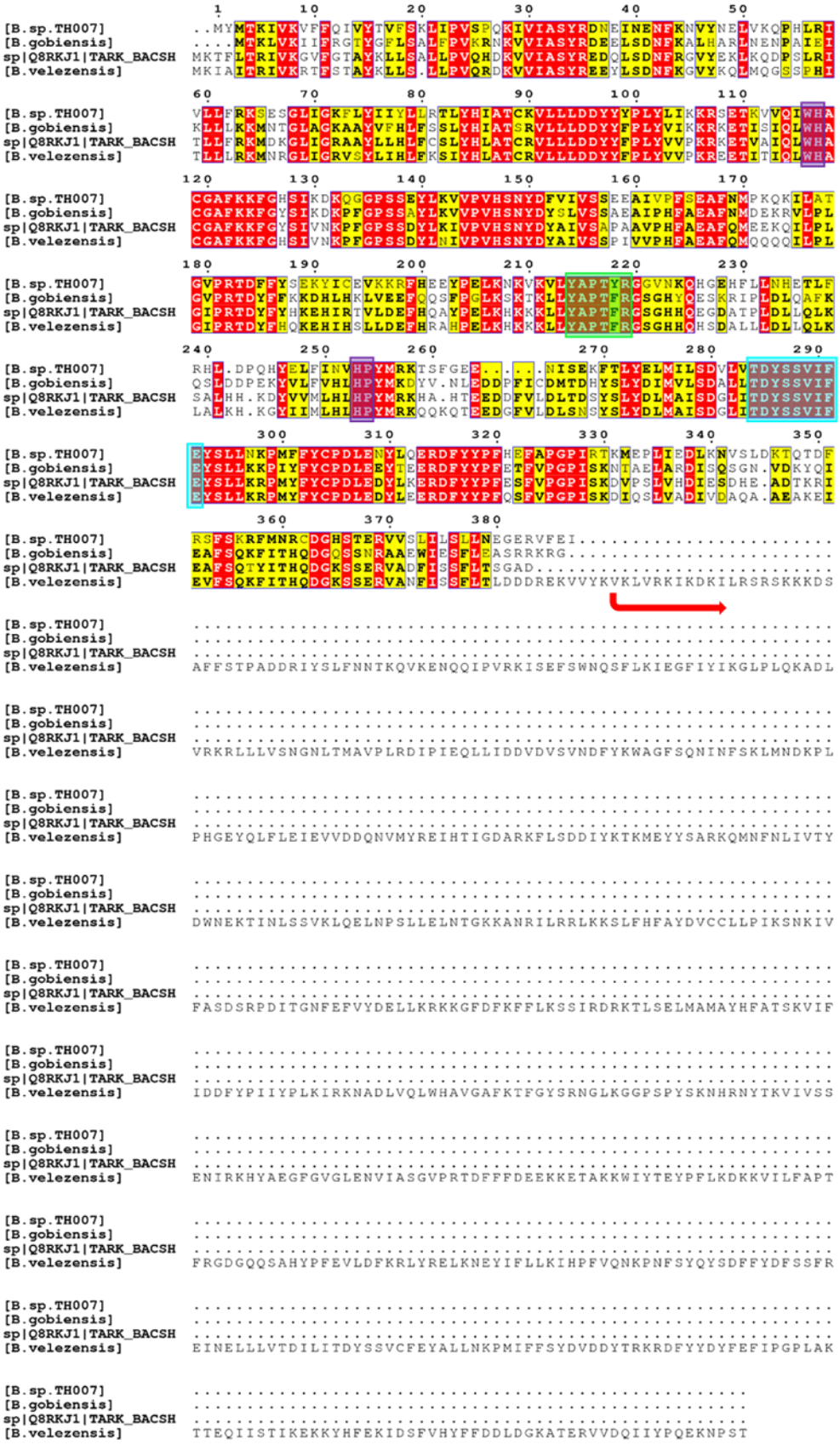


Figure 3.3: Protein sequence alignment of BsTH007TarK, BgTarK, BsTarK and BvTarK. Motifs which are conserved across TarK homologs and SeTagF are in coloured boxes. Purple = Conserved catalytic residues, green = suspected glycerol-phosphate recognition motif, cyan = CDP binding motif. The red arrow indicates the downstream sequence of BvTarK which is identical to BvTarL.

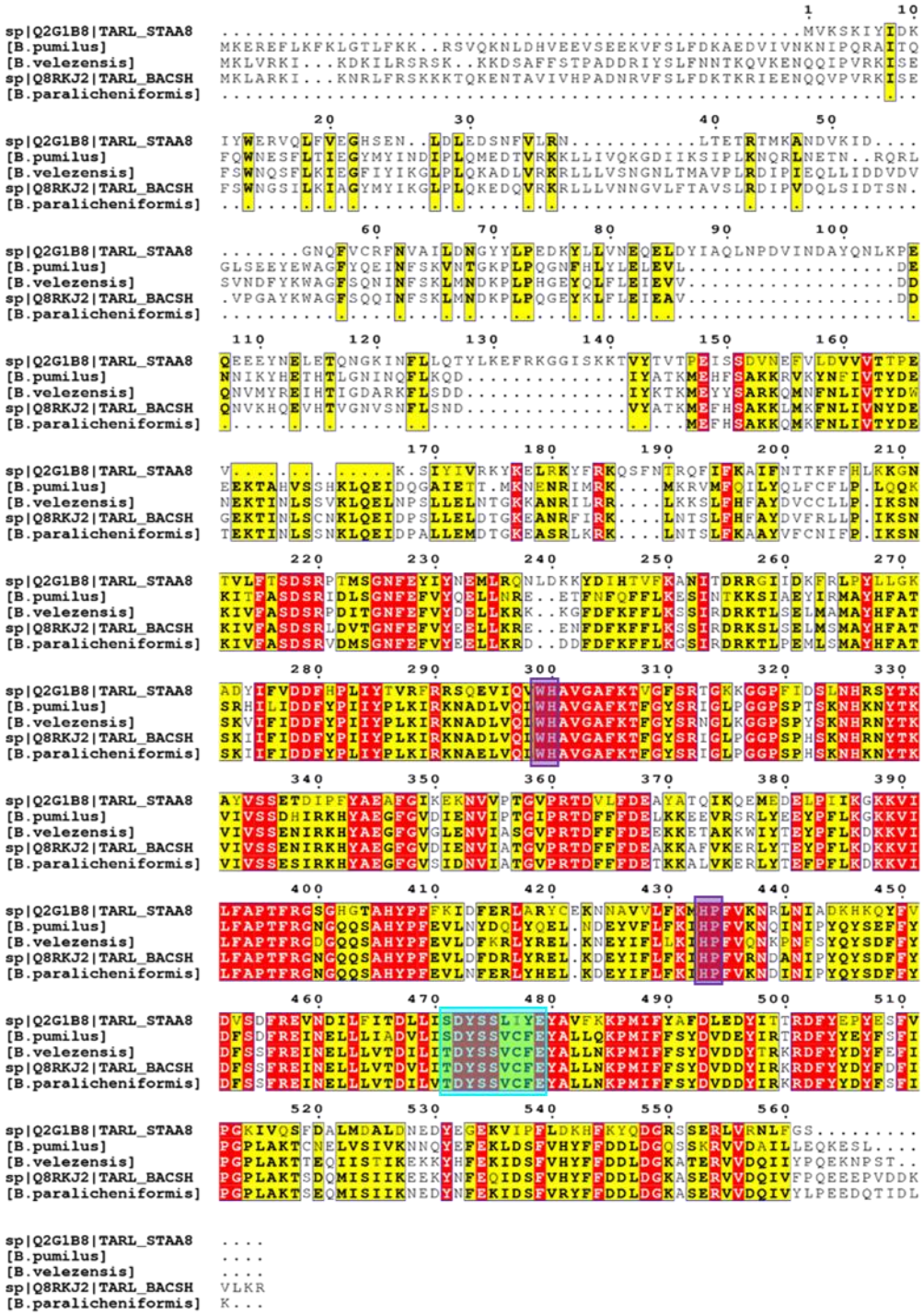


Figure 3.4: Protein sequence alignment of SaTarL, BpuTarL, BvTarL, BsTarL and BpaTarL. Motifs which are conserved across TarL homologs and SeTagF are in coloured boxes. Purple = Conserved catalytic residues, cyan = CDP binding motif.

TarK/L this motif has the same function (Lovering *et al.*, 2010). One other motif that is conserved in TarK and *Se*TagF but not in TarL is the YAPTW(Y/F in TarK homologs)R motif. It was suggested that this motif may be used in acceptor glycerol recognition in *Se*TagF (Lovering *et al.*, 2010), which is intriguing as TarK acts on the same acceptor substrate (glycerol-phosphate) as TagF during WTA synthesis. This may explain the lack of this motif in TarL as this enzyme acts on the pre-existing ribitol-phosphate residues. Finally, *Bv*TarK is

significantly larger than the other TarK homologs (~120kDa compared to ~45kDa). It appears that the entire protein sequence downstream of V397 (the valine mRNA codon, GUG, can be used as a start codon in prokaryotes (Blattner *et al.*, 1997)) is identical to *Bv*TarL, suggesting that this enzyme is a gene fusion of both enzymes as mentioned above. The resulting higher molecular weight makes *Bv*TarK a possible cryogenic electron microscopy (cryo-EM) target.

Another protein sequence alignment was performed between *Bs*TarK/L, Bcs3 (a ribitol phosphate transferase in *Haemophilus influenzae,* (Cifuente *et al.*, 2023)) and *Se*TagF using the same method to examine the conserved residues among these proteins (**Figure 3.5**). The same WH motif is observed in Bcs3, providing further evidence of this conserved catalytic histidine, as well as the DYSSV(L in *Sa*TarL)XXD/E motifs mentioned previously. Furthermore, in Cifuente *et al.*, 2023, it was found that the G99 residue which is conserved in ribitol phosphate transferases but not glycerol phosphate transferases appears to allow the longer chain RboP moiety of CDP-Rbo to bind. In glycerol phosphate transferases, the equivalent residue is a larger proline residue which possibly inhibits CDP-Rbo binding whist allowing CDP-Gro to be accommodated in the active site (Cifuente *et al.*, 2023).

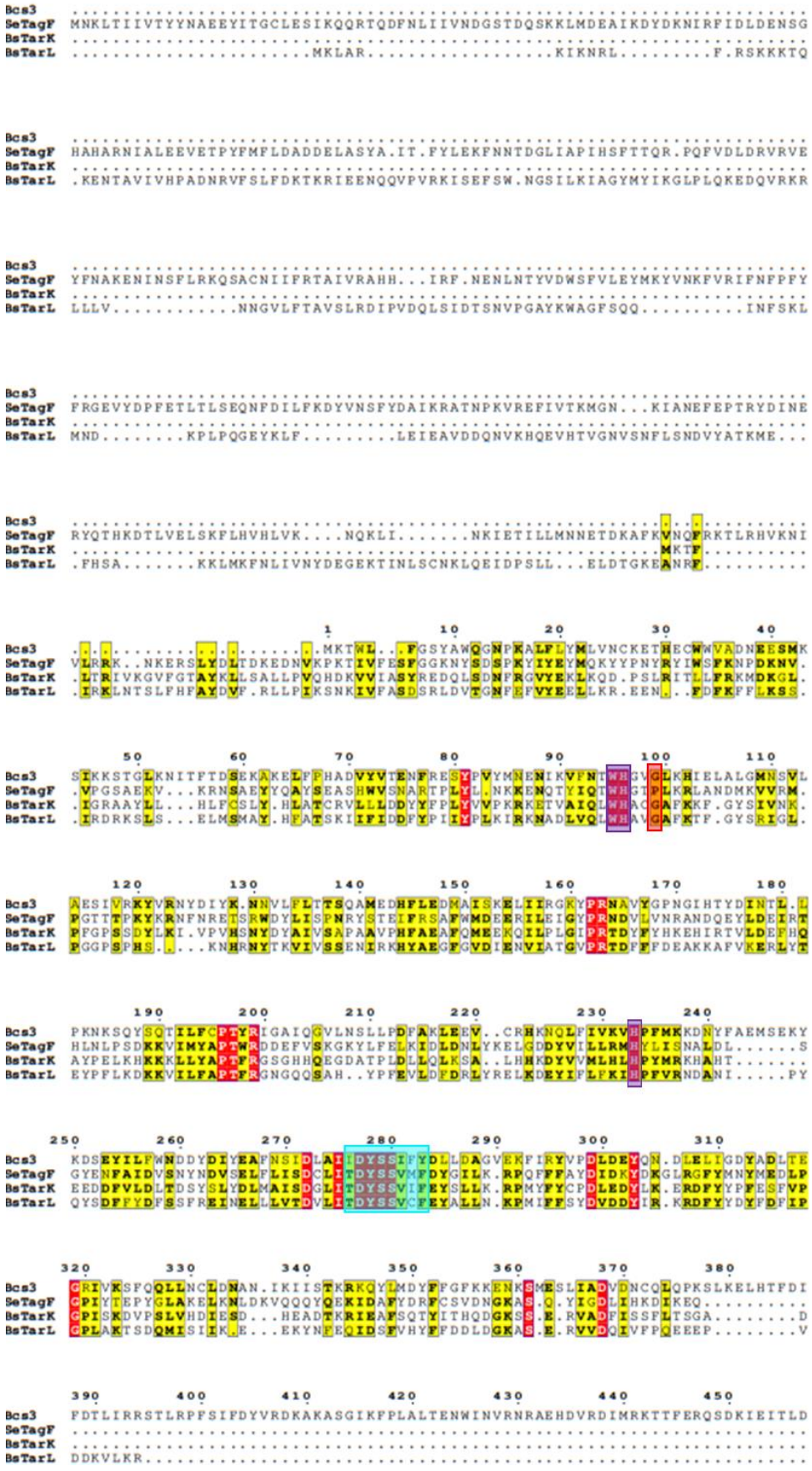


Figure 3.5: Protein sequence alignment of Bcs3 (1-457), SeTagF, BsTarK and BsTarL. Motifs which are conserved are in coloured boxes. Purple = Conserved catalytic residues, red = G/P active site residue depending on enzyme function, cyan = CDP binding motif.

To compare the structural similarities and the arrangement of the conserved catalytic residues in 3D space between polyol-phosphate transferases, AlphaFold2 predicted protein structures of *Bs*TarK and *Bs*TarL were generated (Jumper *et al.*, 2021) (see **3.4.2**). These structures were then compared to published protein structures of other polyol-phosphate transferases such as *Sa*TarL (Li *et al.*, 2024) and *Se*TagF (Lovering *et al.*, 2010) using UCSF Chimera (Pettersen *et al.*, 2004) (**Figure 3.6**).

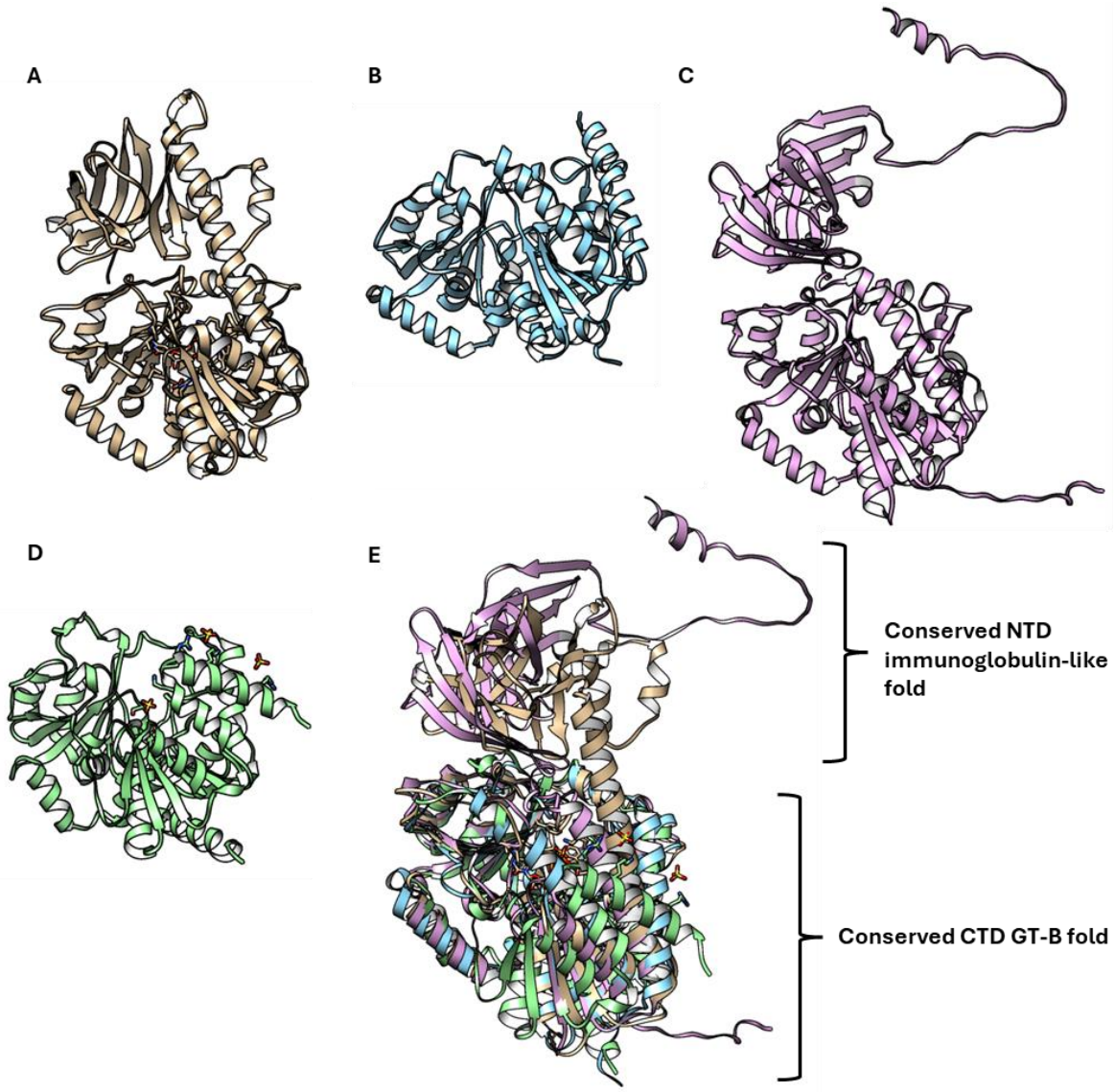


Figure 3.6: 3D representations of relevant polyol-phosphate transferases. A = SaTarL H300N, PDB ID 8VA1, B = BsTarK AlphaFold2 prediction, C = BsTarL AlphaFold2 prediction, D = SeTagF (313-723) PDB ID 3L7I, E = superposition of all 4 structures with the conserved domains labelled. All structures visualised in UCSF Chimera.

All the structures contain a similar C-terminal glycosyltransferase-B (GT-B) domain, which is predicted to be membrane binding (Li *et al.*, 2024). The polymerases (*Bs*TarL and *Se*TagF, for example, not *Bs*TarK) contain an additional N-terminal domain (NTD) which does not have a consistent fold (the *Se*TagF structure is truncated to remove the NTD (1-312) which is why it is not pictured in **Figure 3.6D**). The function of this NTD is not well characterised but is

thought to allow recognition of the growing polyol-phosphate polymer. For the AlphaFold generated *Bs*TarL structure, this NTD is seemingly disordered at the very start of the protein, which may be due to an incorrect prediction or could indicate it is present in the native structure of the enzyme. The function of this domain, assuming the correct prediction, could provide an anchor point into the membrane or act as a dimer arm to facilitate dimerisation as observed in the structure of the TarIJ heterodimer (Li *et al.*, 2021b). The structure of *Sa*TarL was found to be a tetramer of TarL subunits using cryo-EM, which also suggested that the NTD is central for tetramerisation. Therefore, it is highly possible that *Bs*TarL also forms a higher order protein oligomer through the same interactions observed between the NTD and CTD of the adjacent monomer as observed in *Sa*TarL. The possibility of *Bs*TarL oligomerisation should be explored as any higher order oligomer is suitable for single particle cryo-EM as the mass would easily exceed the lower limit of the current EM equipment of ~100 kDa (a *Bs*TarL dimer would be ~146 kDa, for example).

The experimentally determined structures of *Se*TagF 313-723 and *Sa*TarL H300N (Lovering *et al.*, 2010; Li *et al.*, 2024) were then superimposed with the predicted structures of *Bs*TarK and *Bs*TarL in UCSF Chimera (Pettersen *et al.*, 2004) (**Figure 3.7**).

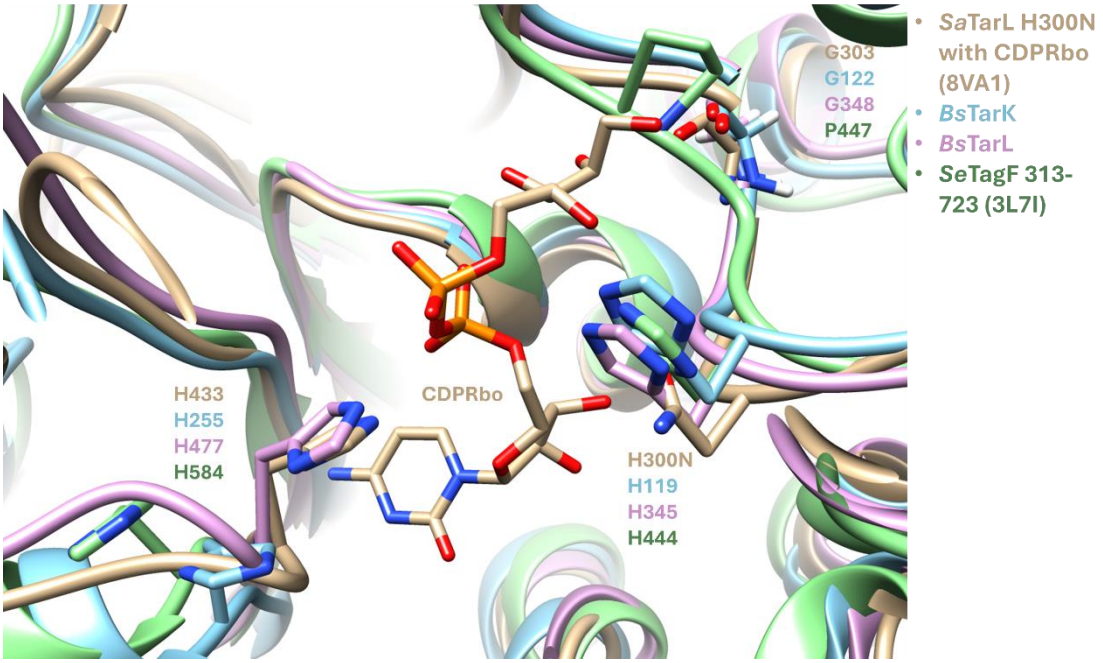


Figure 3.7: Superposition of SaTarL H300N, BsTarK, BsTarL and SeTagF. CDP-Rbo is bound to the active site of SaTarL H300N. The conserved catalytic residues are shown and labelled. The G/P residue for GroP and RboP transferases, respectively, are also shown and labelled. All structures visualised in UCSF Chimera.

This analysis shows that the identified conserved catalytic residues found in the sequence alignments are also somewhat spatially conserved. Both the catalytic histidine from *Sa*TarL H433 and *Bs*TarL H477, which is responsible for binding the alpha phosphate moiety, seem to superimpose well but the corresponding histidines from *Bs*TarK H477 and *Se*TagF H584

do not. This is likely due to the structure of *Bs*TarK being a predicted structure without any ligand bound and therefore the histidine might be out of position compared to the donor-bound conformation. The same could be said for *Se*TagF as this structure also is not ligand bound and therefore it is possible the histidine only swings into position with the presence of a ligand. The histidine residues at this location in both *Bs*TarK and *Bs*TarL could be mutated to alanine residues to inactivate the enzyme without changing the structure for activity and structural studies. The mutagenesis of this residue to alanine was completed in studies of both SeTagF and SaTarL and showed that the enzyme becomes inactivated, supporting the role of this residue in catalysis (Lovering *et al.*, 2010; Li *et al.*, 2024).

The other catalytic residues appear to superimpose very well. It is worth noting that in the *Sa*TarL structure, H300 is mutated to an asparagine to inactivate the enzyme and trap the CDP-Rbo substrate (Li *et al.*, 2024). The equivalent mutation was also performed in studies involving *Se*TagF and it was found to also inactivate the enzyme and trap the CDP-Gro substrate (Lovering *et al.*, 2010). Therefore, mutating the equivalent histidine to asparagine in *Bs*TarK and *Bs*TarL could yield the same effect and could be useful in activity assays, to confirm the proposed role of this histidine residue, and in structure studies to generate donor-bound complexes.

The final superposition explored was the conserved glycine residue among RboP transferases and the equivalent conserved proline residue in GroP transferases. The glycine residue within the structures of *Sa*TarL, *Bs*TarK and *Bs*TarL is located away from the terminal hydroxyl of the ribitol chain when superimposing the CDP-Rbo bound structure of *Sa*TarL onto *Bs*TarK and L, whereas the proline observed in *Se*TagF appears to sterically impede its binding. This fits with the hypothesis that the conserved glycine in RboP transferases allows the longer CDP-Rbo donor to bind, whereas in *Se*TagF the presence of a bulkier proline inhibits CDP-Rbo binding whilst allowing CDP-Gro binding (Cifuente *et al.*, 2023). Therefore, the mutation of this glycine to proline in *Bs*TarK and *Bs*TarL (G122 and G348, respectively) could inhibit the activity of these enzymes and could be useful in activity studies. Furthermore, this mutation may turn RboP transferases into GroP transferases and perhaps vice versa in *Se*TagF mutants.

Overall, three mutations of conserved residues present in other RboP transferases have been identified for the generation of inactive enzyme variants. In *Bs*TarK and *Bs*TarL these mutations are: H119N and H345N (to trap CDPRbo and inactivate catalysis), G122P and G348P (to inhibit CDPRbo binding), and H255A and H477A (to inactivate catalysis), respectively. These mutations can be made through single point mutagenesis, and the

expression of these protein variants should be tested for later activity assays and structural studies.

3.2.3 Expression Trials

The coding sequences of the six homologues and the three initial TarK/L enzymes (*Bs*TarK, *Bs*TarL and *Sa*TarL) were ordered in pET-28a (+) plasmids from GenScript with a N-terminal His6 tag and transformed into *E. coli* cells. The gene sequences were codon optimised for expression in *E. coli.* 1 L scale cultures containing each plasmid were grown to the desired OD_{600} in 2x Yeast Extract Tryptone (2xYT) growth media supplemented with kanamycin and protein expression was induced with IPTG (see **3.4.3**). Uninduced and induced samples were taken and lysed using sonication in Tris buffer. The clarified lysates (soluble protein fraction) and pellets (insoluble protein fraction) were then analysed by Western blot (**Figure 3.8**).

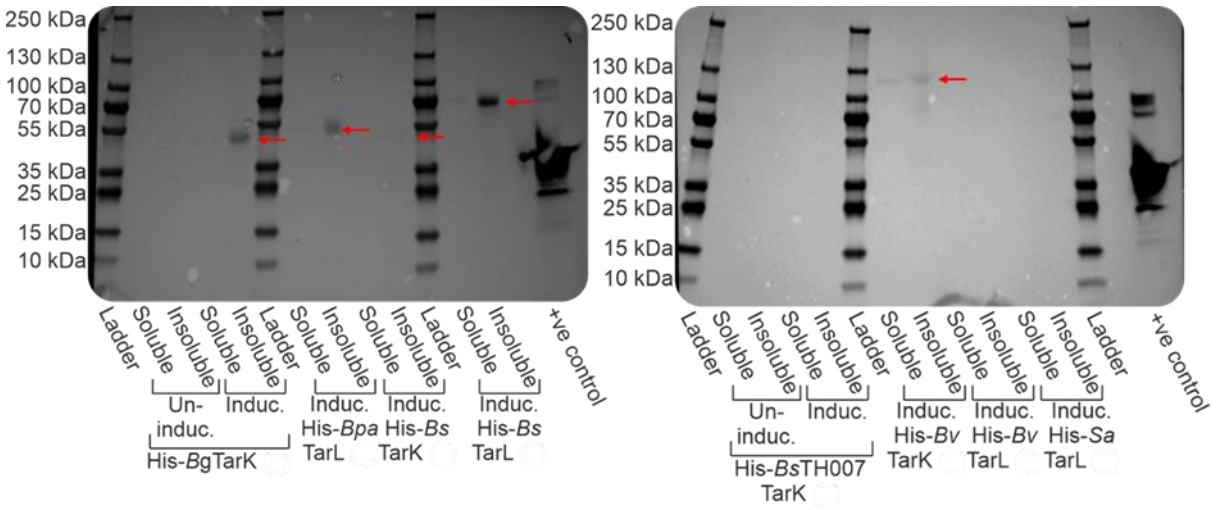


Figure 3.8: Anti His6 tag Western blots of soluble and insoluble proteins fractions from E. coli transformed with various plasmids coding for His-tagged TarK/L enzymes. Uninduced samples are included as negative controls. The positive (+ve) control represents a confirmed His-tagged protein standard. Red arrows indicate positive bands at the expected molecular weights for the proteins. Resolved on a 4-15% polyacrylamide gel.

Due to cell cultures transformed with His-*Bpu*TarL growing too slowly to the desired OD_{600} compared to the other constructs, protein overexpression was not initiated hence there is no Western blot result for this protein. Cultures transformed with His-*Bs*TH007TarK, His-*Bv*TarL and His-*Sa*TarL did not present any bands in their induced lanes suggesting that protein expression has not worked or is too low to detect. These proteins were omitted from later trials. Both His-*Bs*TarK and His-*Bv*TarK appear to have low levels of expression as indicated by faint bands in the insoluble fraction at ~47 kDa and ~121 kDa, respectively. Additionally, His-*Bv*TarK appears to be more soluble than the other constructs as a band of about equal intensity is observed in the soluble and insoluble fractions. His-*Bg*TarK, His-*Bpa*TarL and His-*Bs*TarL all show good levels of expression due to bands present at ~47, ~52 and ~75 kDa,

respectively, with His-*Bs*TarL expressing at the highest levels based on the most intense band. However, most protein from these three constructs is found in the insoluble fraction.

To attempt to increase the amount of protein found in the soluble fraction, buffers containing specific single additives were prepared. Two of the buffers used were variants of a 50 mM Tris, 100 mM NaCl buffer with either 0.6% w/v CHAPS or 20% v/v glycerol added as these additives have been used in published *Bs*TarK and *Bs*TarL studies (Brown *et al.*, 2010). The other buffer contained no additives. Pellets from cells expressing His-*Bg*TarK, His-*Bpa*TarL, His-*Bs*TarK, His-*Bs*TarL or His-*Bv*TarK were prepared like the previous trial and lysed using sonication after being resuspended in either of the three aforementioned buffers (see **3.4.3**). The samples were then prepared for Western blot analysis (**Figure 3.9**).

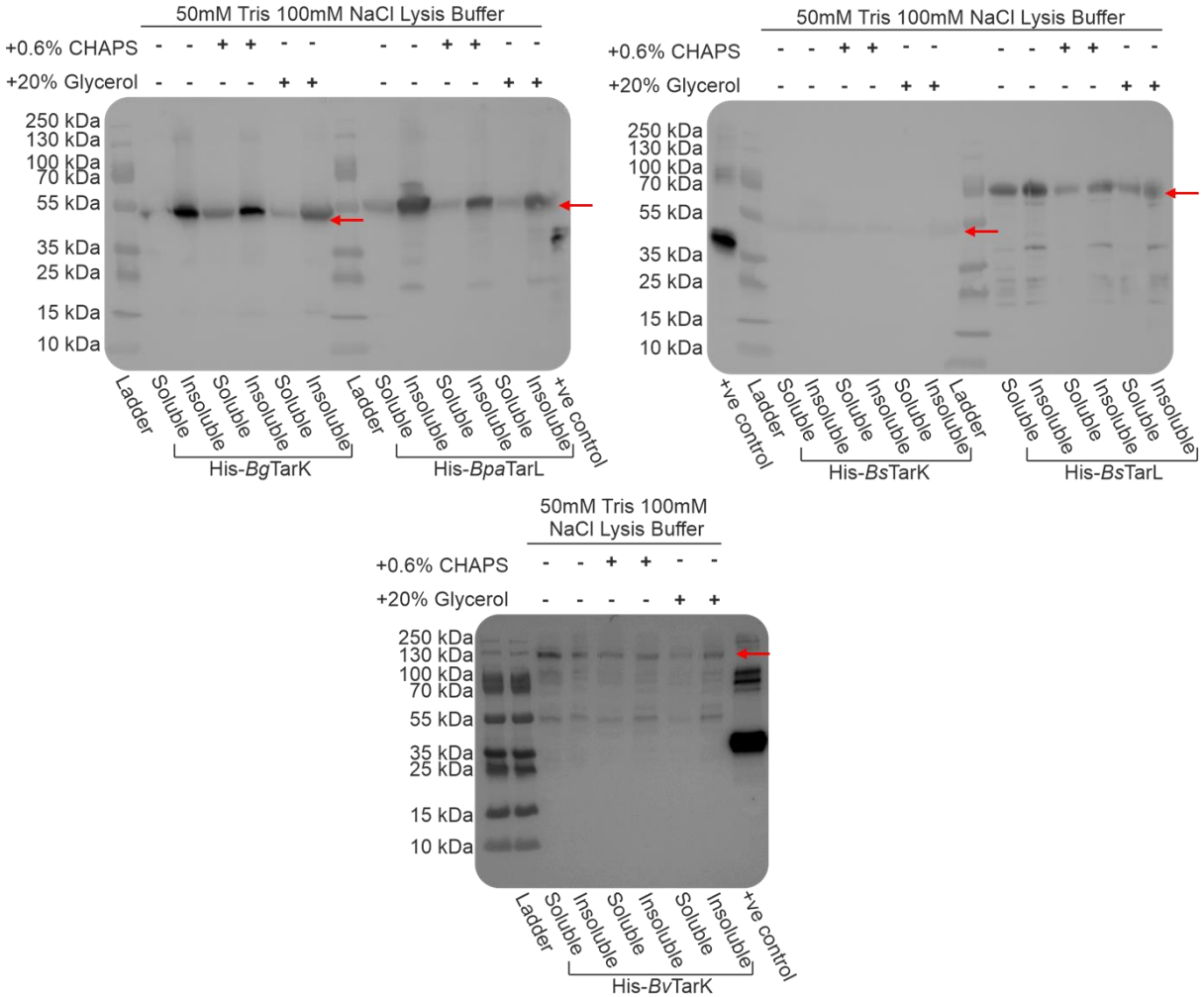


Figure 3.9: Anti His6 tag Western blots of various TarK/L enzymes after sonication in Tris salt buffer with different additives. The positive (+ve) control represents a confirmed His-tagged protein standard. Red arrows indicate positive bands at the correct molecular weight for each target protein. Resolved on a 4-15% polyacrylamide gel.

From the Western blot it appears that His-*Bs*TarK expression does not yield a large amount of protein. This is indicated by faint bands at ~47 kDa present in all the buffer conditions assessed, which is consistent with the previous Western blot result. For His-*Bv*TarK, faint bands are also observed at ~120 kDa across different buffer conditions suggesting that protein

expression is low. However, using just the Tris buffer without glycerol or CHAPS appears to solubilise most of the protein expressed as the band appears to be more intense in the soluble fraction than in the insoluble counterpart. Furthermore, in the +20% v/v glycerol buffer, a more intense band in the insoluble fraction compared to the soluble fraction is observed, suggesting that the addition of glycerol decreases solubility. His-*Bg*TarK expression is comparatively high but mostly results in insoluble protein as indicated by the intense band at ~47 kDa in all buffer conditions. Similarly, His-*Bpa*TarL expression is almost entirely within the insoluble fraction across all buffer conditions as intense bands are observed at ~52 kDa compared to the very faint bands seen in the soluble fractions. The Western blot analysis of His-*Bs*TarL after lysis using the Tris, salt buffer condition without additives shows bands of almost equal intensity at ~75 kDa, therefore suggesting that the protein is equally soluble and insoluble. There is no obvious increase in soluble band intensity compared to the insoluble bands when using the buffers with additives, thus indicating that these additives do not increase protein solubility. In some of the proteins evaluated, namely His-*Bpa*TarL, His-*Bs*TarL and His-*Bv*TarK, several other His tagged protein bands are observed at lower molecular weights. This is likely a result of proteases cleaving the protein as protease inhibitors were not added to the lysis buffers. Addition of protease inhibitors to the lysis buffer should remedy this issue when attempting protein purification. From the five proteins analysed, His-*Bg*TarK, His-*Bs*TarL and His-*Bv*TarK were selected for affinity purification trials using the 50 mM Tris, 100 mM NaCl lysis buffer (see **3.4.4**). These proteins were selected based on their relative soluble expression levels in the results above.

3.2.4 Preliminary Immobilised Metal Affinity Column Purification

1 L of *E. coli* culture transformed with either of the three plasmids was used to express the target proteins using the conditions mentioned in the previous section. Each pellet was resuspended in the abovementioned Tris lysis buffer supplemented with 30 mM imidazole, protease inhibitors, benzonase and lysozyme and lysed by sonication. The cell lysate was then purified by nickel affinity chromatography with the AKTA start purification system (Cytiva) (see **3.4.4**).

Following the purification attempt of His-*Bg*TarK, the UV chromatogram did not display an obvious peak representative of the target protein. Some fractions were pooled (F6-F18) and

the insoluble, flow through and wash fractions, were sampled analysed by SDS-PAGE analysis (**Figure 3.10**) (see **3.4.4)**.

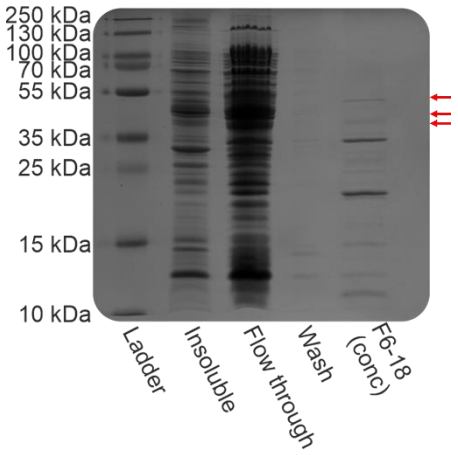


Figure 3.10 Coomassie stained SDS-PAGE gel of the fractions from affinity purification of His-BgTarK. The fractions which may contain the target protein have been combined and concentrated. Red arrows indicate the possible presence of His-BgTarK. Resolved on a 4-15% polyacrylamide gel.

Unfortunately, there is not an intense band observed around ~47 kDa in the pooled purification fractions which would indicate successful purification of His-*Bg*TarK. However, there are two faint bands between 35 and 55 kDa which could consist of the target protein. Because bands can be observed in both the insoluble and flow through within the same mass range, it is also possible that the protein was expressed but mostly insoluble or was not successfully purified. Furthermore, two prominent impurities are present in the 'purified' fractions indicated by two bands, one at ~35 kDa and one just below ~25 kDa. The proteins in these bands could represent degraded His-*Bg*TarK or contaminating, histidine rich *E. coli* proteins. Other impurities can be observed also which could be removed with a different purification method such as size exclusion chromatography (SEC).

Following the purification attempt of His-*Bs*TarL, the UV chromatogram did not display an obvious peak representative of the target protein. Therefore, every second fraction along with the soluble, insoluble, flow through and wash fractions, were sampled and prepared for SDS-PAGE analysis (see **3.4.4**) (**Figure 3.11**).

The bands present above 66.2 kDa in the soluble, insoluble, flow through and the wash fractions are likely to consist of His-*Bs*TarL (~75 kDa). Unfortunately, the most intense band observed is in the insoluble fraction followed by the band at the same molecular weight in the flow through. This suggests that the protein is mostly insoluble, and any solubilised protein did not bind strongly to the column, therefore eluting mostly in the flow though. The insignificant

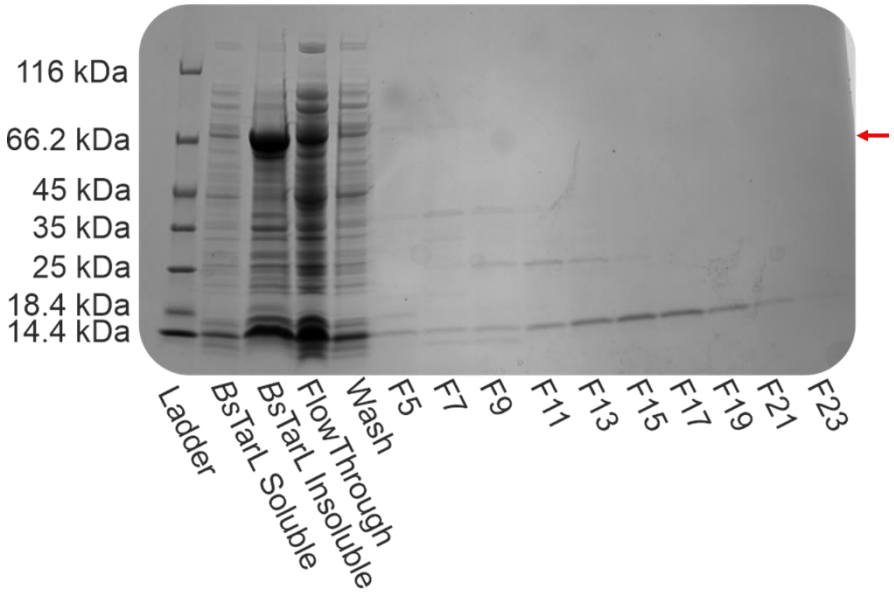


Figure 3.11: Coomassie stained SDS-PAGE gel of the fractions from affinity purification of His-BsTarL. The possible existence of the target protein is indicated by red arrows. The F numbers represent the numbered fractions. Resolved on a 4-15% polyacrylamide gel.

amount of protein in the eluted fractions represented by the very faint bands observed in F5 and F7 also confirms this. It was likely that the imidazole concentration in the wash buffer was too high. To remedy this, the flow through, wash fraction and elution fractions 1-7 were combined and buffer exchanged into imidazole-free Tris buffer using dialysis, and a wash buffer was made with a 3-fold decrease in imidazole concentration (from 30 mM down to 10 mM). The affinity purification was then repeated and to confirm the presence of the protein, samples from the input, flow through and wash were prepared for SDS-PAGE analysis along with every other fraction starting from 3 to 23 (see **3.2.4**) (**Figure 3.12**).

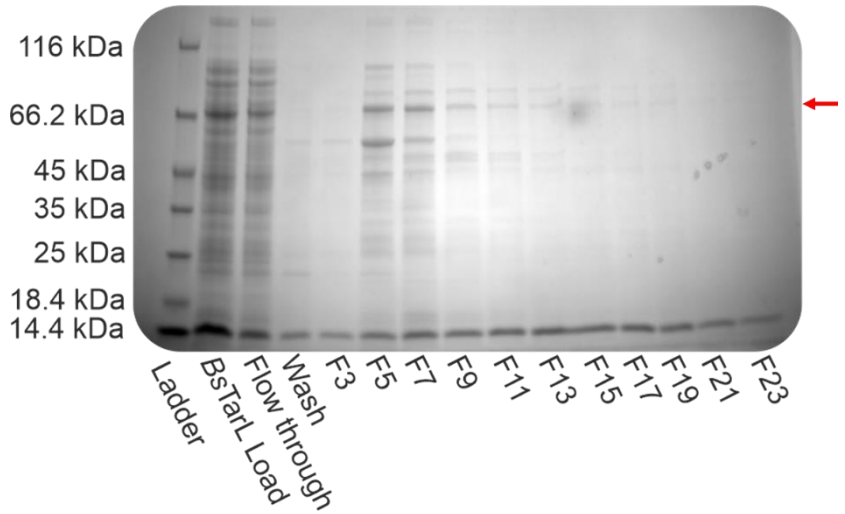


Figure 3.12: Coomassie stained SDS-PAGE gel of the fractions from the second attempt at affinity purification. The possible existence of the target protein is indicated by red arrows. The F numbers represent the numbered fractions. Resolved on a 4-15% polyacrylamide gel.

SDS-PAGE analysis yielded bands across F5-15 which are at the expected mass for His-*Bs*TarL (~75 kDa). The same band can still be observed in the flow through, which suggests that the decrease in imidazole concentration did increase the target proteins affinity to the column but did not result in full binding. The concentration of imidazole could be lowered further in subsequent protein purification trials to attempt to increase affinity to the column, therefore increasing yield.

Following the purification attempt of His-*Bv*TarK, the UV chromatogram did not display an obvious peak representative of the target protein. Therefore, to confirm the existence of the target protein, every second fraction along with the soluble, insoluble, flow through and wash fractions, were sampled and prepared for SDS-PAGE analysis (see **3.2.4)** (**Figure 3.13**)

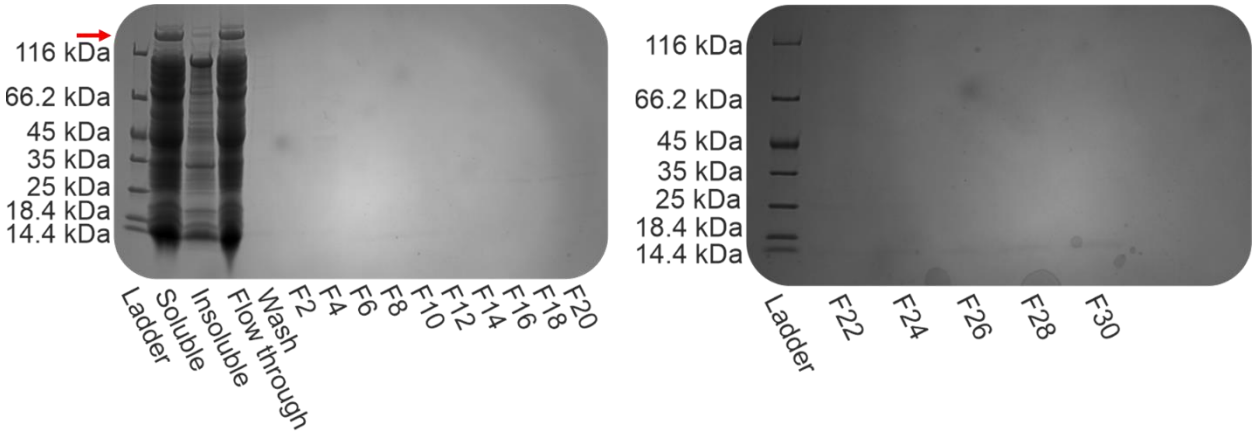


Figure 3.13: Coomassie stained SDS-PAGE gel of the fractions from Ni-NTA affinity purification of His-BvTarK. The possible existence of the target protein in the soluble lane and flow through is indicated by red arrows. The F numbers represent the numbered fractions. Resolved on a 4-15% polyacrylamide gel.

Within the eluted fractions, there are no bands observed which would correspond to the molecular weight of His-*Bv*TarK (~121 kDa). However, there are two intense bands found within the soluble and flow through lanes at the expected molecular weight of the target protein, suggesting that the protein did express and is soluble but did not bind to the affinity column. This could have been due to the imidazole concentration of the lysis buffer being too high (30 mM). To remedy this problem, the imidazole concentration should be decreased to 10 mM for future purifications to ensure the target protein binds to the column.

3.3 Conclusions and Future Work

To conclude, a homology search was performed to identify putative ribitol-transferases TarK and TarL in various strains of Gram-positive bacteria and a small number of these were selected for sequence analysis, structural analysis and expression trials. One of the selected proteins, His-*Bs*TarL, was successfully purified from expression trials using affinity chromatography. Both His-*Bg*TarK and His-*Bv*TarK appeared to have been expressed at appreciable levels but were not successfully purified using the initial affinity purification procedure. This could be remedied with decreasing the imidazole concentration of the lysis buffer for future purifications, increasing the length of the His6 tag to 8 or 10 residues or using a different type of affinity tag. The majority of *Bg*TarK and *Bs*TarL seems to remain in the insoluble fraction, and therefore further expression trials are required to solubilise these proteins. The expression of soluble *Bs*TarK and *Bs*TarL, as well as the identified polyol-phosphate transferases mentioned earlier in this chapter (*Sa*TarL (Li *et al.*, 2024) and *Se*TagF (Lovering *et al.*, 2010)) have been reported (Brown, Zhang and Walker, 2008; Brown *et al.*, 2010). The methods by which each paper produces soluble polyol-phosphate transferases have some similarities. For example, Brown *et al.*, 2010, Lovering *et al.*, 2010 and Li *et al.*, 2024 all use the detergent CHAPS at concentrations between 0.5 to 40 mM in the lysis buffers, which suggests that CHAPS concentration should be a factor in future expression trials in combination with other variables. Brown *et al.*, 2010 also uses a different *E. coli* cell to the cell line used in this chapter, Rosetta2(DE3) pLysS. This could imply that the overexpression of *Bs*TarK and *Bs*TarL is toxic as the pLysS *E. coli* strain carries the gene encoding T7 lysozyme, which lowers the background expression level of target genes under the control of the T7 promoter. Why Rosetta2 *E. coli* is used is unclear. This strain supplies tRNAs for 7 rare codons typically used for eukaryotic protein expression; therefore, this strain may also supply the same rare codons found in Gram-positive gene sequences for expression in Gram-negative systems (like *E. coli*). However, the original genes ordered from GeneScript were codon optimised for expression in *E. coli* so this should not be an issue. Nevertheless, changing the cell line is another variable that could be tested in future expression trials.

Another change from the expression trials described here when compared to the literature is the location of the His6 tag. N-terminal tags were used here, but both Brown *et al.*, 2010, and Li *et al.*, 2024, use C-terminal tags. In the case of *Sa*TarL expression, this tag is also cleavable (Li *et al.*, 2024). A N-terminal tag for *Bs*TarL could interfere with tetramerisation which was seen in *Sa*TarL, therefore causing the protein to lose the ability to undergo stabilising multimer interactions and no longer be soluble. Hence the creation of C-terminal His6 tagged gene constructs should be performed and the subsequent protein overexpression tested.

The purification methods also differ among polyol-phosphate transferases. Brown *et al.*, 2010, uses a rocking bed of His-Bind resin instead of fast protein liquid chromatography (FPLC) for purification of *Sa*TarK and *Sa*TarL. However, they only use the purified enzymes for activity assays and not for structural determination, so it is possible the purity may be lower. Notably, no data regarding the expression or purification of these proteins are provided by the authors. After the initial IMAC step, Li *et al.*, 2024 used a heparin affinity column which they state, "separated unfolded from folded protein, and highly pure samples from this procedure were critical in obtaining diffracting crystals". Therefore, trying different purification methods could increase the yield of soluble protein.

3.4 Methods

3.4.1 Homologue selection using BLAST

To search for close homologues of *Bs*TarK and *Bs*TarL, their protein sequences were submitted for a BLAST nr_clustered protein search. Six homologues were chosen originating from four distinct species and ranging from 51-87% identity to *Bs*TarK and *Bs*TarL. All TarK and TarL DNA and protein sequences were compiled from GenBank/GenPept (NIH) and aligned using Clustal Omega (Madeira *et al.*, 2022). Each protein's theoretical pI was calculated using the ExPASy Server (Gasteiger *et al.*, 2005). All Tar enzyme protein sequence alignments were annotated using ESPript (Robert and Gouet, 2014).

3.4.2 AlphaFold2 Protein Structure Generation and Alignments

The AlphaFold2 Protein Structure Prediction package was used on the Viking computing cluster to generate and rank five protein structures each of *Bs*TarK and *Bs*TarL (Jumper *et al.*, 2021). The structures of *Sa*TarL and *Se*TagF were obtained from the PDB (Berman *et al.*, 2000). All structural visualisations were generated using UCSF Chimera and alignments were performed using the MatchMaker tool (Pettersen *et al.*, 2004).

3.4.3 Expression Trials

Flasks containing 10 mL of kanamycin 2xYT (see **7.1.4**) were inoculated with BL21(DE3) *E. coli* cells transformed with pET-28a (+) containing either *Bg*TarK, *Bpa*TarL, *Bpu*TarL, *Bs*TarK, *Bs*TarL, *Bs*TH007TarK, *Bv*TarK, *Bv*TarL or *Sa*TarL (all N-terminally His6 tagged) from previously made glycerol stocks (plasmids named p"protein-ofinterest"28a, see **7.1.2** for plasmid list and **7.2.5** for transformation protocols). The flasks were incubated at 37 °C in a shaking incubator at 200RPM overnight. Flasks containing 1 L of kanamycin 2xYT were then inoculated with these starter cultures and were incubated at 37 °C, 200 RPM until the cultures

reached an OD_{600} value of 0.5-0.8. The cultures were then induced with 1 mM IPTG to initiate expression of the various His6 tagged Tar proteins. 1 mL samples were taken before induction. After overnight expression at 16 °C, 200 RPM, the cell cultures were pelleted by centrifugation (6000 xg, 30 mins, 4 °C) and stored at -70 °C after determining each pellet's mass. Samples of each pellet were taken and resuspended in either Tris A (50 mM Tris pH 7.25, for *Bg*TarK, *Bpa*TarL, *Bs*TarK, *Bs*TarL and *Sa*TarL) or Tris B buffer (50 mM Tris pH 8.15, for *Bs*TH007TarK, *Bv*TarK and *Bv*TarL) and sonicated on ice at 22 µm amplitude for 20-30 cycles of 30 secs pulses with 30 secs rest. In the case of the additive trials, these buffers were also supplemented with either 100 mM NaCl and 0.6 % (w/v) CHAPS or 100 mM NaCl and 20 % (v/v) glycerol. Insoluble cell debris was pelleted using centrifugation (38,000 xg, 30 mins, 4 °C) and the soluble fraction was collected. Cell debris was then resuspended as much as possible in an equal volume of lysis buffer. The samples were subsequently prepared for SDS-PAGE to observe expression (see **7.2.8**). Samples were also prepared for Western blot analysis (see **7.2.9**).

3.4.4 Preliminary Immobilised Metal Affinity Column Purification

Larger scale cultures (~1 L of kanamycin 2xYT) were produced by inoculating the 1 L flasks with 10 mL of starter culture containing p*Bg*TarK28a, p*Bs*TarL28a or p*Bv*TarK28a. The flasks were incubated at 37 °C, 200 RPM until 0.5-0.8 OD_{600} was reached. Protein expression was induced with the addition of 1 mM IPTG, and expression proceeded at 16 °C, 200 RPM overnight. The cells were subsequently pelleted by centrifugation (6000 xg, 30 mins, 4 °C) and the supernatant was removed. Pellets were resuspended with Tris A lysis buffer (50 mM Tris, 100 mM NaCl, 30 mM imidazole, pH 7.25), supplemented with benzonase, lysozyme and protease inhibitors, using a ratio of 1 mL of buffer per 0.1 g of cell pellet. The solutions were then sonicated on ice at 22 µm amplitude for 30 secs 20 times with a 30 secs rest between pulses. Insoluble cell debris was pelleted using centrifugation (38,000 xg, 30 mins, 4 °C) and the soluble fraction was collected. Cell lysates were subsequently loaded onto a 5 mL HisTrap FF Crude column (Cytiva) using an AKTA start (Cytiva) for affinity purification. The flow through was collected and the column was washed with 15 CV of Tris A lysis buffer. Proteins were then eluted with a linear gradient of 0-100% of Tris A elution buffer (50 mM Tris, 100 mM NaCl, 500 mM imidazole, pH 7.25) over 15 CV. 3 mL fractions were collected. The soluble and insoluble fractions, flow through, wash and fractions of interest were sampled and prepared for SDS-PAGE analysis (see **7.2.8**). The flowthrough fraction of *Bs*TarL contained most of the target protein. This fraction, along with the wash and the elution fractions up to and including F7, were combined and dialysed overnight into a Tris A lysis buffer with a

reduced concentration of imidazole (50 mM Tris, 100 mM NaCl, 10 mM imidazole, pH 7.25). The affinity purification was then repeated using the same protocol as above.

Chapter 4: *Bs*TarK Wild Type and Mutant Protein Expression and Purification

4.1 Introduction

TarK belongs to a family of monotopic membrane proteins responsible for the synthesis of the poly-ribitol phosphate chain present in the wall teichoic acid (WTA) structure which is synthesised inside the cytoplasm of Gram-positive bacterial cells including *Bacillus subtilis* W23 and *Staphylococcus aureus.* Specifically, TarK from *Bacillus subtilis* W23 (*Bs*TarK) is known as a 'primase' which acts on the TagB product lipid-P-P-GlcNAc-ManNAc-GroP to install a single D-ribitol-5-phosphate (RboP) residue on the C1 hydroxyl position of GroP using the donor substrate CDP-Ribitol (CDP-Rbo) (Brown *et al.*, 2010) (**Figure 4.1**). *Bs*TarL then continues to polymerise RboP residues until the polymer is about 20-40 residues long (Brown *et al.*, 2010; Swoboda *et al.*, 2010). In *S. aureus,* it has been proposed that TarK does not have a function and instead both the priming and polymerisation is achieved with a dual function TarL enzyme (Brown *et al.*, 2010). However, some studies indicate that TarK acts like and replaces the dual fuctional TarL in *tarL* deletion strains (Meredith, Swoboda and Walker, 2008). This polymer is then exported and covalently attached to the peptidoglycan (PG) and the polymer extends far beyond the PG layers (Neuhaus and Baddiley, 2003).

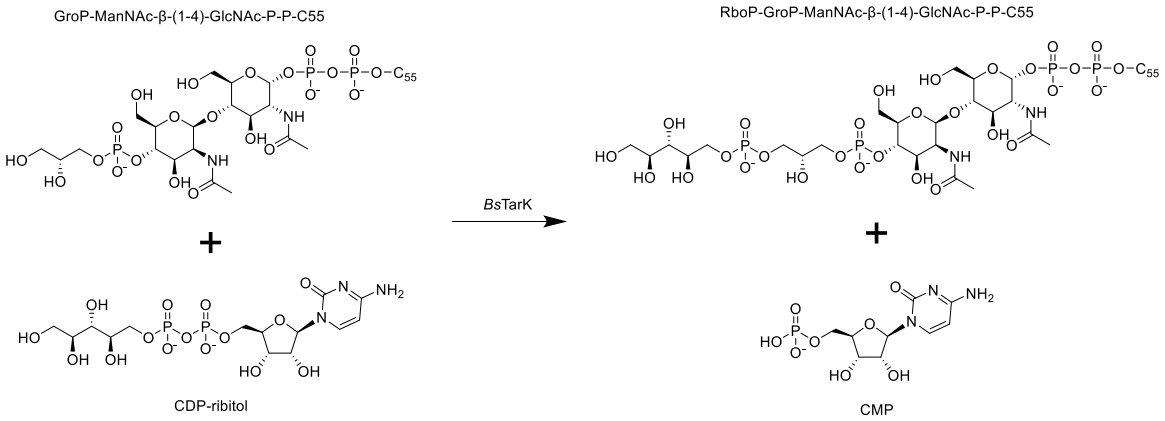


Figure 4.1: B. subtilis ribitol-phosphate priming reaction facilitated by TarK. BsTarK acts only on the TagB product, C_{55}-P-P-GlcNAc-ManNAc-GroP.

WTAs are known to be involved in many functions, namely cell shape determination, cell division, biofilm formation, cell adhesion and colonisation and antibiotic resistance (M. Gross *et al.*, 2001; Soldo, Lazarevic and Karamata, 2002; Weidenmaier *et al.*, 2004; Swoboda *et al.*, 2010; Farha *et al.*, 2013; Wu *et al.*, 2021). Hence, enzymes along the biosynthetic pathway have been proposed as antivirulence targets. Several studies have found that mutant Methicillin-resistant *Staphylococcus aureus* (MRSA) strains lacking WTAs regain sensitivity to

β-lactam antibiotics (Brown *et al.*, 2012). Additionally, research indicates that inhibiting enzymes involved in the biosynthesis and export of WTAs leads to their complete absence (Hancock, Wiseman and Baddiley, 1976; Swoboda *et al.*, 2010). Furthermore, a combined strategy using broad-spectrum antibiotics along with inhibitors of WTA production has shown positive outcomes in animal models of MRSA infection and colonization (Farha *et al.*, 2013; Lee *et al.*, 2016).

Further research is needed to deepen our understanding of WTA synthesis and establish a foundation for new drug studies aimed at combating the increasing resistance to antibiotics. It is crucial to investigate the roles of these biosynthetic enzymes and any related homologs. In particular, TarK is of significance, as knockout strains of *B. subtilis* W23 without this enzyme are non-viable (Brown *et al.*, 2010). The vulnerability of TarK knockout strains may arise from the buildup of toxic WTA intermediates or from the depletion of lipid phosphate precursors which are also essential for PG synthesis (Bouhss *et al.*, 2008). Currently, no inhibitors for TarK have been identified, despite recent progress in the study of ribitol phosphate transferases. Therefore, a combined approach involving the synthesis of chemical probes or inhibitors, along with structural and enzymological studies, is necessary to further understand this process and facilitate structure-based drug design.

4.1.1 Aims

In order to enable structural and enzymological studies of TarK, it is necessary to establish workflows for the production and purification of the recombinant protein and any active site variants. In this chapter, the previous expression trials of *Bs*TarK described in chapter 3 will be expanded upon to find suitable conditions for the isolation of soluble TarK therefore facilitating downstream characterisation of the enzyme.

First, recombinant gene products with different tags varying in location will be synthesised for further, more expansive expression trials. The best conditions for bacterial growth and protein expression and the tag type and location that gives highest yields of soluble protein will then be used for larger scale protein purification using various optimised chromatography methods. Upon the production of highly pure *Bs*TarK, various methods of enzyme characterisation will be used such as circular dichroism to check if the protein is folded and nano differential scanning fluorimetry to assess the protein melting temperature and to screen possible TarK inhibitors synthesised by collaborators. The purified enzyme will also undergo protein crystallisation trials with ligands to attempt to solve the protein structure by X-ray crystal diffraction. Finally, a liquid chromatography–mass spectrometry (LC-MS)-based assay will be developed to screen *Bs*TarK activity by utilising donor substrates synthesised in-house and synthetic lipid-linked acceptors.

4.2 Results and Discussion

4.2.1 Molecular Cloning

In chapter 3, it was shown that *Bs*TarK can be expressed as an N-terminal His fusion protein in BL21(DE3) cells, although levels of soluble protein appeared to be relatively low under the conditions tested. To express and purify soluble *Bs*TarK, a few different gene constructs were designed and synthesised using the In-Fusion cloning method. First, the *Bs*TarK gene was inserted into Champion™ pET SUMO (shortened to pSUMO) which has an N-terminal His6 tag followed by a Small Ubiquitin-like Modifier (SUMO) solubility tag for the initial expression trials. This tag is intended to increase the solubility of proteins and also contains a protease cleavage site immediately downstream of the SUMO tag, leaving no extra amino acids at the N-terminus of the target protein after removal of the tag (Butt *et al.*, 2005).

The Champion™ pET SUMO vector was first linearised at the desired insertion site using designed forward and reverse primers and commercial PCR reagents (see **4.4.1**). Analysis of correct linearisation was performed using agarose gel electrophoresis (see **7.2.2**) (**Figure 4.2**). Successful linearisation of the vector is evident from the intense signal for all samples at ~5.5 kb. Each band was extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C.

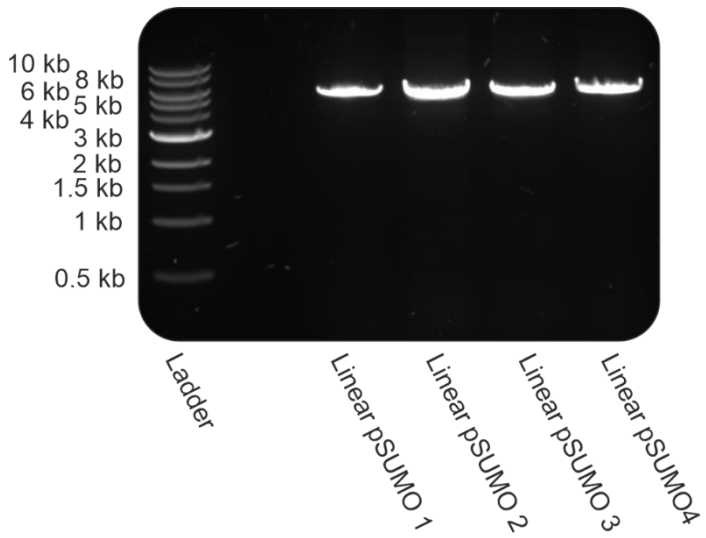


Figure 4.2: Linearised pSUMO resolved on a 0.8% agarose gel. Expected length of the linearised plasmid is 5.643 Kb.

The corresponding insert with homologous overlaps with the linearised pSUMO plasmid was synthesised by cloning the *Bs*TarK gene from the ordered plasmid pUC57*Bs*TarK using forward and reverse primers and commercial PCR reagents (see **4.4.1**). Completed PCR reactions were analysed using DNA gel electrophoresis (see **7.2.2**) (**Figure 4.3**). Successful synthesis of the *Bs*TarK insert is represented by the intense band observed at ~1.2 kb

reminiscent of the correct insert length. The faint bands observed at ~0.5 kb are likely primer dimers. The inserts located at the correct length were gel extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C.

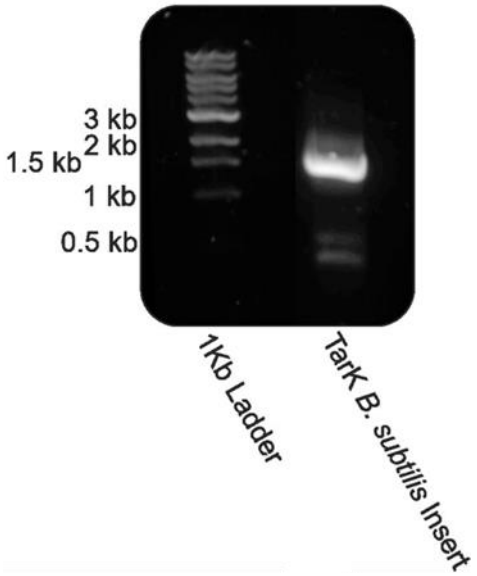


Figure 4.3: BsTarK insert for linear pSUMO resolved on a 0.8% agarose gel. Expected band is 1.2 kb

To complete ligation, the In-Fusion kit (Takara Bio) was used and completed reactions were transformed into *E. coli* cells. Antibiotic resistant colonies were isolated, and colony PCR was used to confirm the presence of the ligated plasmid with the *Bs*TarK insert (see **7.2.3**) (**Figure 4.4**). The single intense band at ~1.7 kb across all the colonies tested is indicative of the presence of the correctly engineered plasmid (1.766 kb). The same colonies were cultured and miniprepped and the isolated pSUMO*Bs*TarK plasmids were stored at -20 °C.

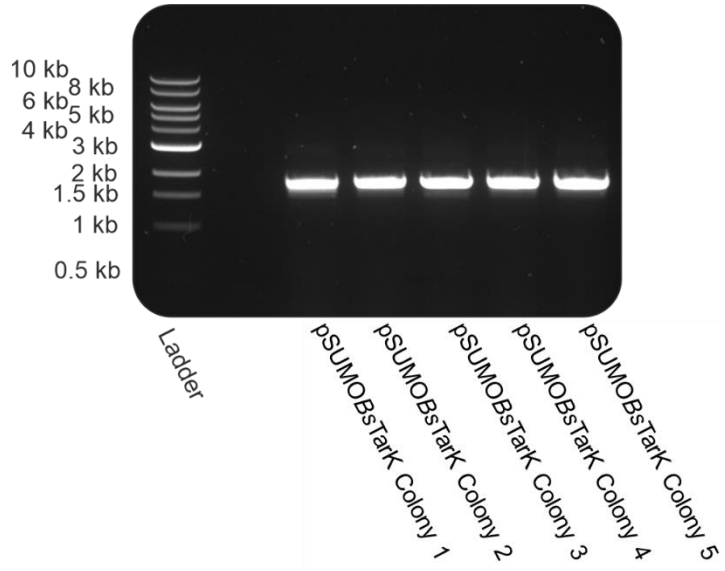


Figure 4.4: Colony PCR products of pSUMOBsTarK resolved on a 0.8% agarose gel. Correct length expected to be 1.766 kb.

C-terminally His6 tagged BsTarK constructs were also designed for In-Fusion cloning. First, the pET-28a(+) vector was linearised at the desired insertion site, to allow C-terminal His6 tagging, using commercial PCR reagents and designed plasmids (see **4.4.1**). Analysis of the correct linearised vector was performed using DNA electrophoresis (**Figure 4.5**).

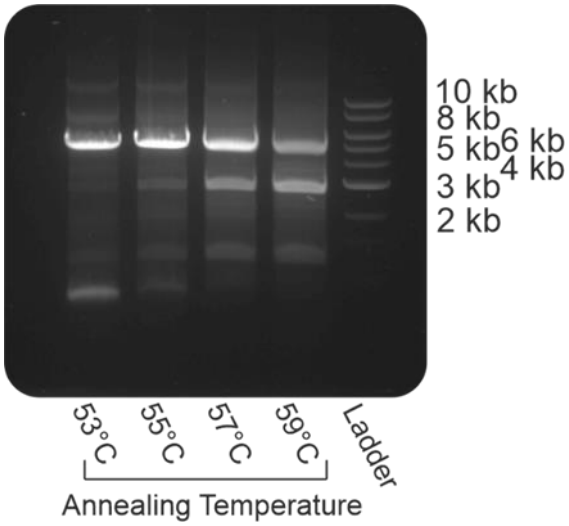


Figure 4.5: Linearised pET-28a(+) for C-terminal His6 tag genes resolved on a 0.8% agarose gel. Expected length of the linearised plasmid is 5.2 kb.

Correct linearisation is represented by the intense bands located at ~5.2 kb. The faint bands observed at around 3 kb are likely to be incorrectly linearised plasmids resulting in truncated amplifications due to differences in annealing temperature. The other faint bands at length less than 2 kb are likely primer dimers. Nevertheless, separation is clear, therefore the linearised plasmid was subsequently gel extracted (see **7.2.1**) and stored at -20 °C.

The corresponding *Bs*TarK insert was synthesised by PCR using the ordered plasmid pHis-*Bs*TarK28a designed primers complementary to the insertion site of the linearised vector (see **4.4.1**). Correct synthesis of the *Bs*TarK insert was analysed by DNA gel electrophoresis (**Figure 4.6**). The presence of an intense band at around 1.2 kb indicates correct synthesis of the *Bs*TarK insert at every annealing temperature tested. The inserts located at the correct length were gel extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C.

To complete ligation, the In-Fusion kit (Takara Bio) was used and completed reactions were transformed into *E. coli* cells. Successful colonies were cultured and miniprepped and isolated plasmid DNA was sent for Sanger DNA sequencing (Eurofins Genomics) which confirmed correct ligation. Recombinant plasmids were then stored at -20 °C.

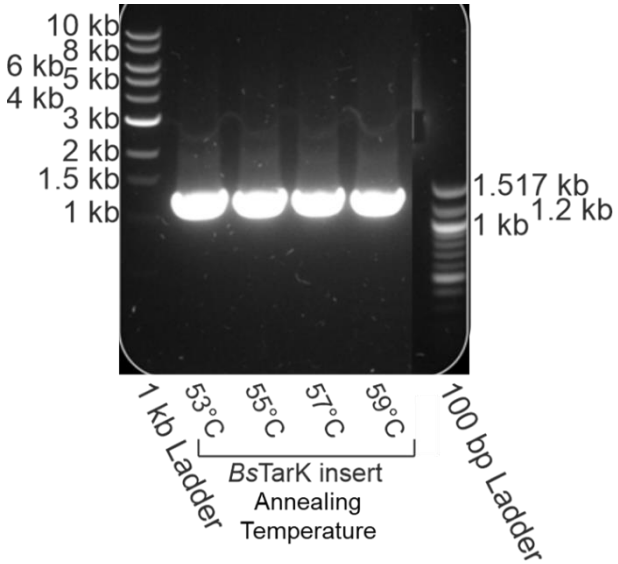


Figure 4.6: BsTarK insert for linear pET-28a(+) resolved on a 0.8% agarose gel. Expected band is 1.2 kb.

4.2.2 Expression Trials

The expression of the His-SUMO-*Bs*TarK protein was first assessed. Initially, the pSUMO*Bs*TarK plasmid was transformed into BL21 (DE3) *E. coli* and successful colonies were cultured for protein overexpression with the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (see **4.4.2**). To assess protein expression, samples of the uninduced and induced pelleted cells were taken, lysed using BugBuster reagent, and prepared for SDS-PAGE and anti-His Western Blot analysis (see **7.2.8** and **7.2.9**) (**Figure 4.7**). Clarified lysates (soluble fractions) were compared to cell debris (insoluble fractions). The presence of a band between 45 and 66.2kDa in the induced lane (**Figure 4.7**) may correspond to the expected mass of the His-SUMO-BsTarK construct (58.378kDa). However, this band is only observed in the insoluble fraction. This observation is further supported by the presence of an intense band at the expected mass in the induced, insoluble lane in the anti-His6 tag Western blot (**Figure 4.7B**). A faint band is also observed in the induced, soluble lane of the Western blot indicating a small amount of soluble protein expression. Unfortunately, the presence of a SUMO tag seems to have not completely solubilised the protein, thus further expression trials are required.

A few variables for expression of *Bs*TarK have been changed in the literature. For example, the concentration of IPTG has been halved to 0.5 mM in the procedure by Brown, Zhang and

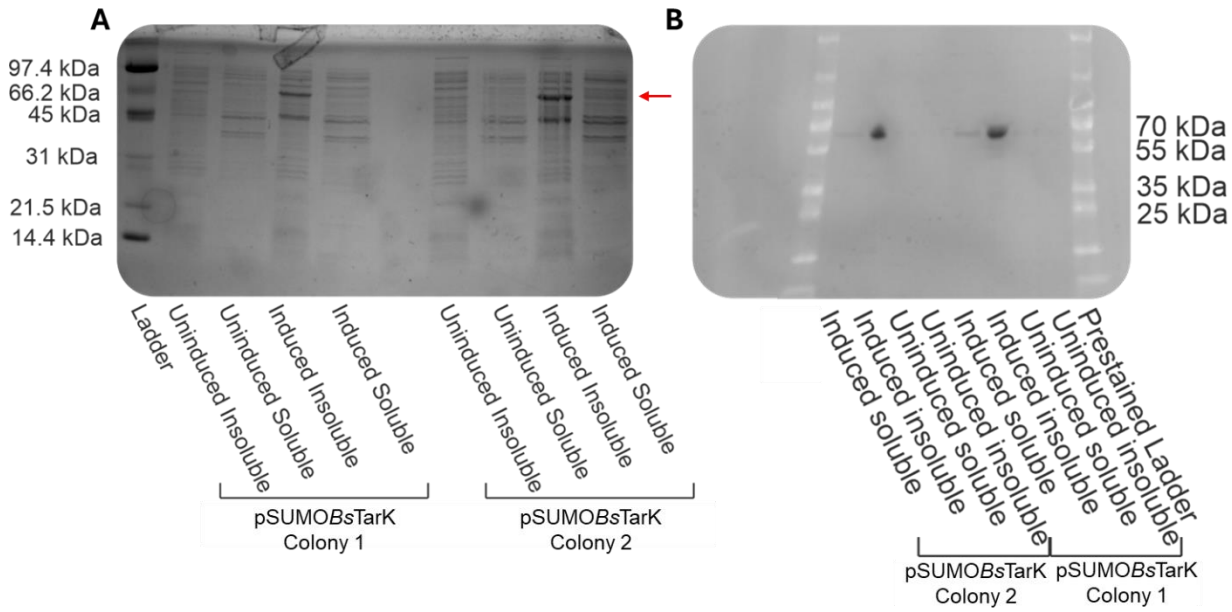


Figure 4.7: SDS-PAGE and Western blot analysis the initial expression trial of His-SUMO-BsTarK. A = Coomassie stained SDS-PAGE gel. Bands containing the target protein are labelled with a red arrow. B = An anti-His6 tag western blot of the same expression trial showing a His6 tagged construct was expressed. Two colonies from the transformation were assessed. Resolved on a 4-15% polyacrylamide gel.

Walker, 2008. Additionally, the buffer used for cell lysis in Brown *et al.*, 2010 contained additives such as 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) and glycerol. Other TagF-like enzymes have also been solubilised for structural studies using CHAPS at varying concentrations (Brown, Zhang and Walker, 2008; Brown *et al.*, 2010; Lovering *et al.*, 2010; Li *et al.*, 2024). Therefore, differing IPTG concentrations for protein induction were subsequently trialled along with a lysis buffer which contained CHAPS and glycerol. Four different IPTG concentrations were tested: 0.1, 0.25, 0.5 and 1.0 mM (see **4.4.2**). Resuspended uninduced and induced cell pellets were lysed using sonication in the buffer identical to the buffer used by Brown *et al.*, 2010, which contains 0.6% (w/v) CHAPS and 20% (v/v) glycerol. Samples were prepared for SDS-PAGE analysis (see **7.2.8**) to assess protein expression (**Figure 4.8**).

The results indicate that different IPTG concentrations have trivial effect on the expression of His-SUMO-*Bs*TarK. More importantly however, cell lysis with the new lysis buffer appears to have solubilised many of the bacterial proteins previously observed in the insoluble fraction in **Figure 4.7**. This includes the target protein, His-SUMO-BsTarK, as shown by the band observed in the soluble, overnight (O/N) fraction approximately at the expected mass (58.378 kDa) which is not present in the uninduced lane.

The next expression trial involved testing different expression media. Terrific broth (TB), superior broth (SB) and the rich growth medium 2xYT were chosen and expression conditions were altered to only induce protein expression with 1 mM IPTG (see **4.4.2**). Uninduced and

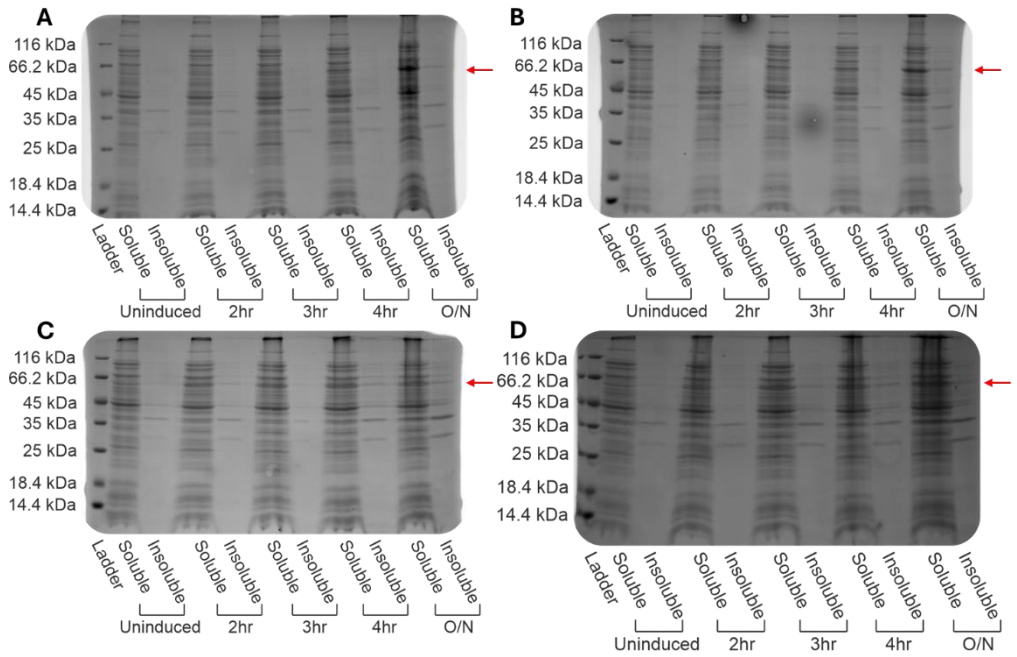


Figure 4.8: Coomassie stained SDS-PAGE gels from His-SUMO-BsTarK expression trial with samples taken at the indicated time intervals after induction. A = induction with 0.1 mM IPTG. B = induction with 0.25 mM IPTG. C = induction with 0.5 mM IPTG. D = induction with 1 mM IPTG. Bands which are likely the target protein are labelled with red arrows at the expected mass of 58.378 kDa. O/N = overnight. Resolved on a 4-15% polyacrylamide gel.

induced samples were taken, and pellets were lysed using the same method as the previous expression trial (see **4.4.2**). Samples were then prepared for SDS-PAGE analysis (see **7.2.8**) (**Figure 4.9**).

The Coomassie gels indicate that 2xYT growth medium appears to increase expression of the target protein due to the intense band at the expected mass. It was also observed that 2xYT media increased the growth rate of the cultures as they reached the target OD_{600} sooner than the other types of media. The corresponding band in the results from TB and SB media is harder to observe. This could be due to the protein expressing at lower levels. The very intense band at ~14.4kDa on both gels is suspected to be the lysozyme added to the lysis buffer.

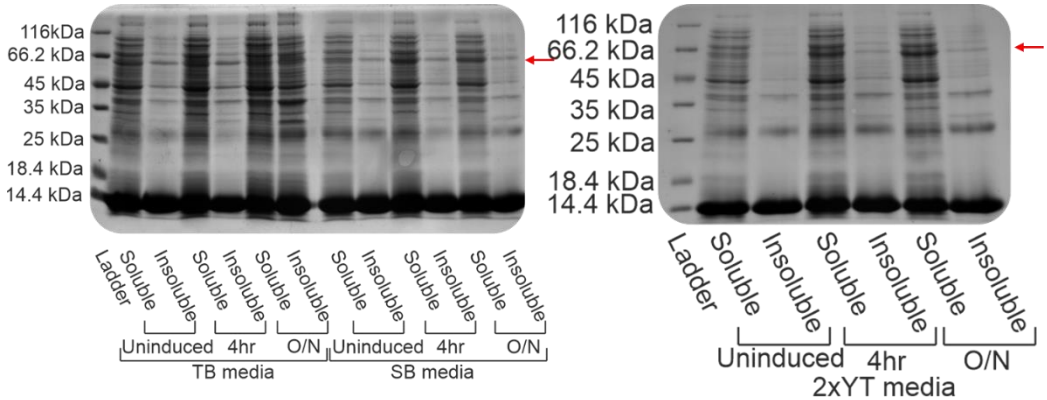


Figure 4.9: Coomassie stained SDS-PAGE gels from His-SUMO-BsTarK expression trial involving different media. Left = Terrific Broth and Superior Broth. Right = 2xYT media Samples of soluble (lysates) and insoluble (pellet) protein fractions are shown before and after induction. The intense band identified by the red arrows corresponds to the approximate expected mass of His-SUMO-BsTarK (58.378kDa). O/N = overnight. Resolved on a 4-15% polyacrylamide gel.

With a selection of growth and expression conditions assessed, the optimised conditions were chosen for the next set of expression trials. These conditions were: 2xYT media, cells cultured until 0.5-0.8 OD_{600}, then induction of protein expression with the addition of 1mM IPTG and growth O/N at 16°C. As the N-terminal SUMO tag does not seem to cause increased solubility, it was decided to return to the non-SUMO-tagged, N-terminally His6-tagged protein trialled previously (**Chapter 3**) and also test the newly synthesised C-terminal His6-tagged protein as expression levels remained low. In the literature, *Bs*TarK and other TagF-like proteins are C-terminally tagged (Brown, Zhang and Walker, 2008; Brown *et al.*, 2010; Lovering *et al.*, 2010). Additionally, the CHAPS concentration in the lysis buffer differs greatly. For *Bs*TarK expression, Brown *et al.*, 2010 uses 0.6% w/v (~9.8 mM) CHAPS in their lysis buffer which is around the critical micelle concentration (CMC) of this detergent (the same buffer as used above), while *SaTarL* expression and lysis described by Li *et al.*, 2024, uses a CHAPS concentration of 16 mM. The highest concentration of CHAPS used in a lysis buffer is seen in the expression and purification of TagF reported by Lovering *et al.*, 2010, where the concentration is 4-fold higher than the CMC of CHAPS at 40 mM. Increasing the concentration of CHAPS in the lysis buffer may yield a higher amount of solubilised *Bs*TarK. Finally, another observation from the literature is that *Bs*TarK is expressed in the Rosetta2(DE3) pLysS strain of *E. coli* (Brown *et al.*, 2010). Therefore, it is worth evaluating the expression of *Bs*TarK in other strains than BL21(DE3). The reason for the use of Rosetta2(DE3) pLysS strain of *E. coli* is not known. Rosetta2 strains contain more rare codons which are used for eukaryotic protein expression, but *Bs*TarK is a prokaryotic protein. It is possible the use of this expression strain helps with Gram-positive protein expression in Gram-negative systems as there are codons which are rarer in *E. coli.* However, the gene sequences used in this work have been codon optimised for expression in *E. coli.* The pLysS aspect of this strain encodes and expresses T7 lysozyme, which suppresses basal expression of T7 RNA polymerase prior to induction.

Both the N- and C-terminally His6 tagged *Bs*TarK constructs were transformed into Rosetta2(DE3) pLysS and expression of *Bs*TarK proceeded using the optimised conditions described above (see **4.4.2**). Uninduced and induced samples were taken and lysed using sonication in the 'high' (40 mM) CHAPS buffer from Lovering *et al.*, 2010. The soluble and insoluble fractions were prepared for SDS-PAGE and Western blotting analysis (see **7.2.8** and **7.2.9**) (**Figure 4.10**).

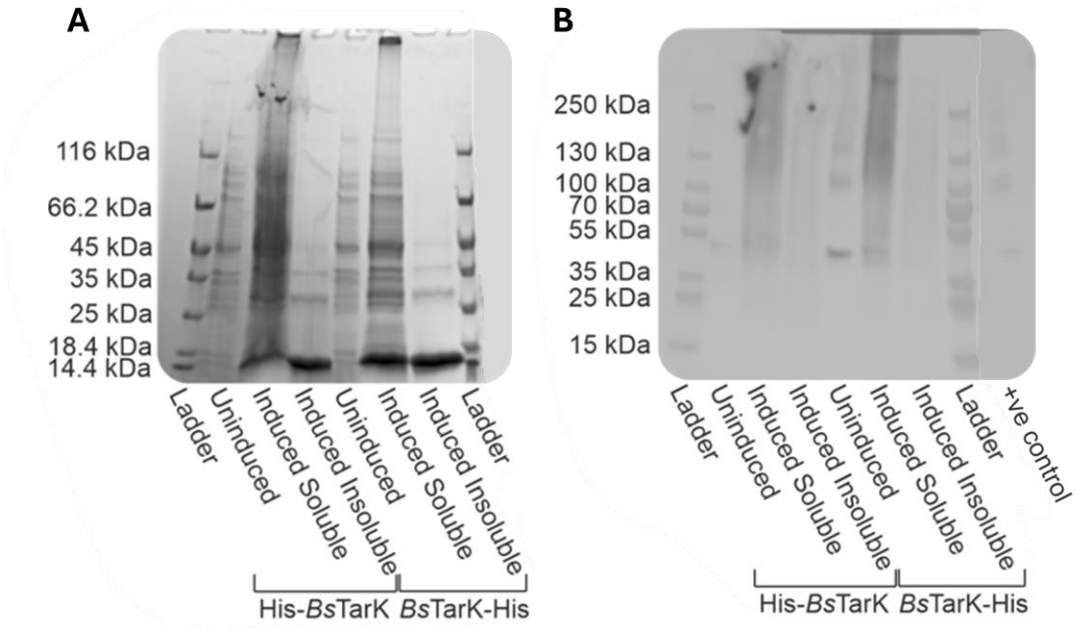


Figure 4.10: SDS-PAGE and Western blot analysis of the expression of N- and C-terminally His6 tagged BsTarK. A = Coomassie stained SDS-PAGE gel of BsTarK expression. B = An anti-His6 tag western blot of the same expression trial. The uninduced lane represents the total cell lysate (soluble and insoluble protein). Resolved on a 4-15% polyacrylamide gel.

The expression of both tagged proteins is unclear in both the Coomassie stained gel and the Western blot. There may be *Bs*TarK expression observed in the Coomassie gel as a band is present at around 45 kDa (the expected mass is ~47 kDa for both the N- and C- terminally tagged constructs), however this band is present in both the uninduced and induced soluble fraction. The same observation is present in the Western blot suggesting leaky expression. There is also substantial smearing in the Westing blot which is likely due to non-specific binding of the anti-His6 tag antibody. The reason for this may be due to the high amount of detergent present in the loaded samples. Additionally, there are bands present at the very top of the Coomassie gel in the induced soluble fractions. This could indicate that the presence of CHAPS at a concentration far exceeding the CMC is causing the soluble protein to aggregate and inhibiting entry into the gel. This can also explain the smears present in the induced soluble fractions of both images which may consist of a continuous range of aggregated protein of varying molecular weight.

From the results of the previous trial, it was decided to revert to using the 'low' CHAPS buffer for subsequent trials. In this trial, the N-terminally His6 tagged *Bs*TarK protein was expressed in two different strains: Rosetta2(DE3) pLysS (Novagen) and SHuffle lysY (New England Biolabs) (see **4.4.2**). Protein expression conditions remained the same and uninduced and induced samples were taken. Samples were subsequently lysed using sonication in the 'low'

(10 mM) CHAPS buffer (see **4.4.2**) and prepared for SDS-PAGE and Western Blot analysis (see **7.2.8** and **7.2.9**) (**Figure 4.11**).

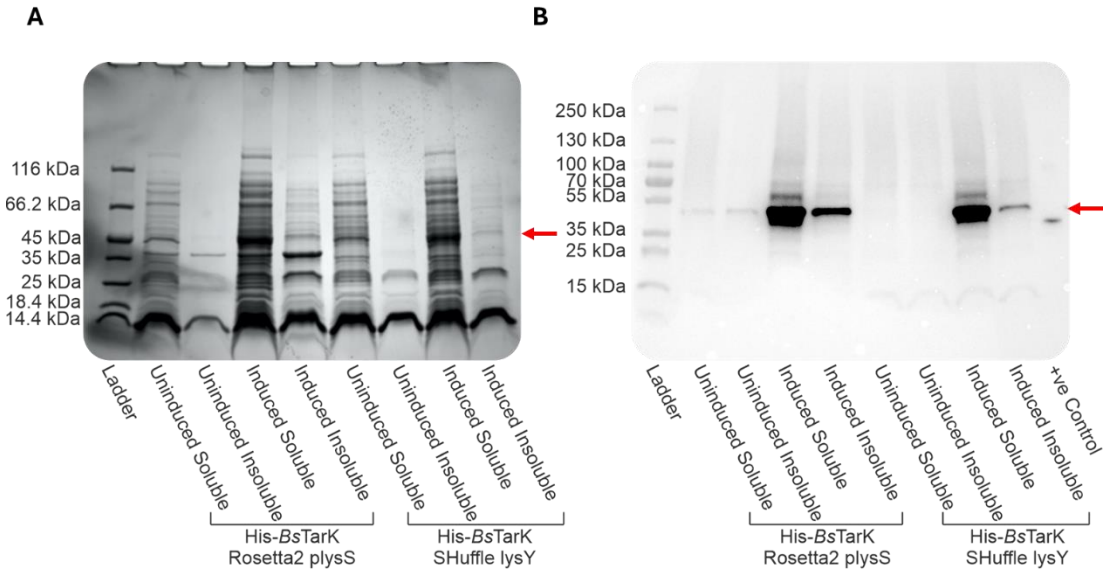


Figure 4.11: SDS-PAGE and Western blot analysis of the expression of N-terminally His6 tagged BsTarK in different strains. A = Coomassie stained SDS-PAGE gel of N-terminally His6 tagged BsTarK expression. Bands containing the target protein are labelled with a red arrow. B = An anti-His6 tag western blot of the same expression trial. Resolved on a 4-15% polyacrylamide gel.

Soluble expression of His-*Bs*TarK is observed in the Coomassie stained SDS-PAGE gel as an intense band is present in the induced soluble lane, but not in the uninduced lanes, for both strains of bacteria at around 45 kDa (expected mass ~47 kDa) (**Figure 4.11A**). This is also supported by the clear and highly intense bands observed in the Western blot in the same lanes located between the 35 kDa and 55 kDa marks (**Figure 4.11B**). The Western blot also displays faint bands in the induced insoluble lanes of both strains implying that not all the protein is soluble. Furthermore, in the Rosetta2 strain faint bands are observed in the uninduced fractions which may indicate leaky expression. Nevertheless, using these strains in combination with the 'low' CHAPS buffer increases soluble expression considerably, with the Rossetta2 strain yielding slightly more intense bands when observed in the Western blot than the SHuffle strain.

Next, the same trial was performed but the C-terminally His6 tagged *Bs*TarK was expressed instead (see **4.4.2**). All conditions remained the same. SDS-PAGE and Western blotting were used to assess expression (see **7.2.8** and **7.2.9**) (**Figure 4.12**).

The Coomassie-stained SDS-PAGE gel reveals the soluble expression of *Bs*TarK-His, evident as a prominent band in the induced soluble lane for both bacterial strains, which is absent in the uninduced lanes. This band appears at approximately 45 kDa, close to the expected mass of ~47 kDa (**Figure 4.12A**). The observation is further supported by the Western blot, which

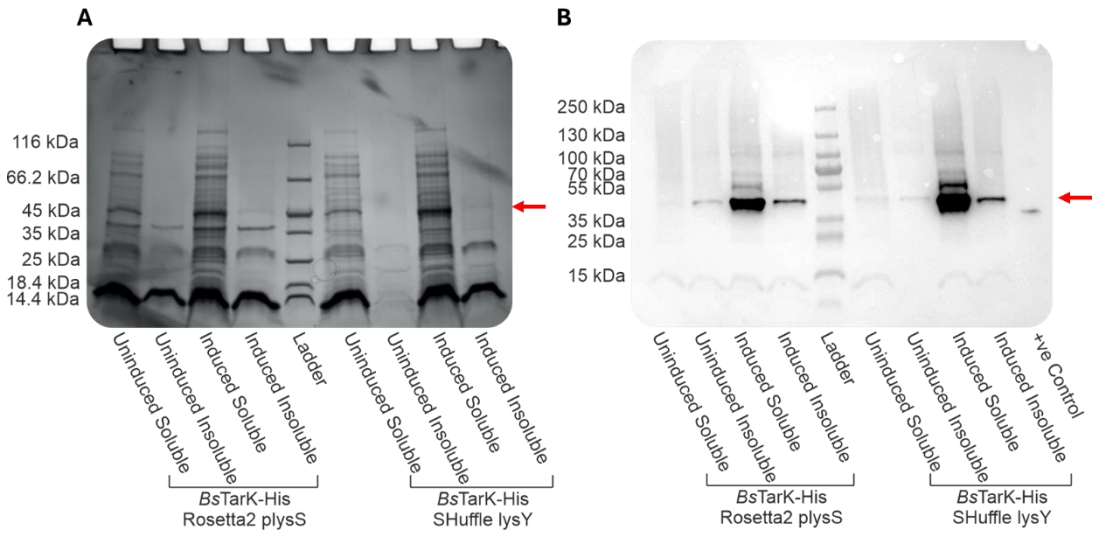


Figure 4.12: SDS-PAGE and Western blot analysis of the expression of C-terminally His6 tagged BsTarK in different strains. A = Coomassie stained SDS-PAGE gel of C-terminally His6 tagged BsTarK expression. Bands containing the target protein are labelled with a red arrow. B = An anti-His6 tag western blot of the same expression trial. Resolved on a 4-15% polyacrylamide gel.

shows clear and highly intense bands in the induced soluble lanes between the 35 kDa and 55 kDa markers **Figure 4.12B**. In both strains however, a faint band can be observed in the induced insoluble lanes of both strains implying that not all the protein is soluble. Furthermore, faint bands can also be observed in the uninduced soluble and insoluble fractions of both strains tested. This may again imply leaky expression. The SHuffle strain appears to express *Bs*TarK-His slightly more than the Rosetta2 strain as the bands present in the Western blot seem more intense.

Ultimately, using the 'low' CHAPS buffer (Brown *et al.*, 2010) appears to solubilise *Bs*TarK incredibly well compared to the 'high' CHAPS buffer (Lovering *et al.*, 2010) when expressed in the Rosetta2(DE3) pLysS (Novagen) and SHuffle lysY (New England Biolabs) strains and also decreases the smears observed in the Western blots. Between the two strains, *Bs*TarK expression levels are at comparable levels. Furthermore, the location of the His6 tag seems to not affect the expression of *Bs*TarK, however, it was decided to switch to the C-terminally His6 tagged construct for later protein purification work in line with the literature (Brown, Zhang and Walker, 2008; Brown *et al.*, 2010; Lovering *et al.*, 2010).

4.2.3 Protein Purification

Early attempts to purify *Bs*TarK used the engineered pSUMO*Bs*TarK plasmid and the expression conditions outlined in **4.4.2, Table 4-1**, entry 3 using the 2xYT growth medium. Using the methods outlined in section **4.4.4, Table 4-2**, entry 1, immobilised metal affinity chromatography (IMAC) was first used to isolate the target protein from the lysate. Samples of the soluble, insoluble, flow through and wash fractions were taken, and after purification finished, samples from the eluted fractions, which may have contained the target protein by observation of 280 nm absorption peaks, were taken. All the samples were prepared for SDS-PAGE analysis (see **7.2.8**) (**Figure 4.13**).

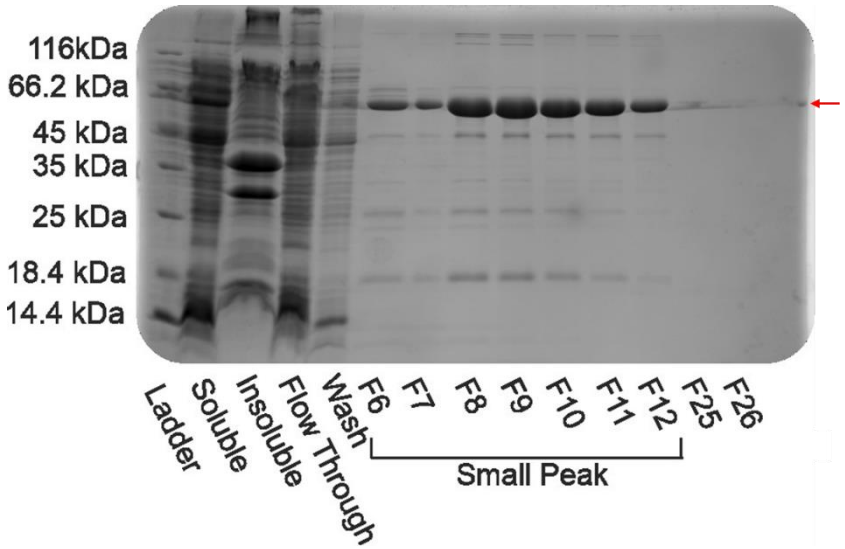


Figure 4.13: First IMAC purification of His-SUMO-BsTarK from E. coli lysate. Red arrows indicate the presence of target protein in the eluted fractions (expected mass ~58 kDa). Resolved on a 4-15% polyacrylamide gel.

From the gel, intense bands at around the expected mass of His-SUMO-*Bs*TarK (~58 kDa) can be observed in fractions 6-12. This indicates successful capture and elution of the His6 tagged protein from the cell lysate. However, within the fractions which contain the target protein, there are impurities at around 18, 25 and 45 kDa, therefore further purification steps are required.

Following the first successful capture step, the fractions containing His-SUMO-*Bs*TarK were combined, and the SUMO tag was cleaved off using the SUMO protease protocol (Invitrogen). To separate the His-SUMO tag from the now native *Bs*TarK, a second IMAC step was performed (**4.4.4**). As the His-SUMO retains the His6 tag, the target protein should be observed in the flowthrough while just the His-SUMO tag should be observed in the elution. The completed reaction mixture was loaded onto the column and purification proceeded. After the purification was completed, samples of the completed SUMO protease reaction and each fraction from the IMAC purification were prepared for SDS-PAGE analysis (see **7.2.8**) (**Figure** 4.14).

Intense bands are observed at the expected mass of native *Bs*TarK (~45 kDa) in the post SUMO reaction and the flow through lanes (**Figure 4.14**), therefore indicating successful cleavage of the His-SUMO tag from *Bs*TarK. There is a faint band observed at ~58 kDa in both the post SUMO and flow through fractions suggesting that the cleavage of His-SUMO from His-SUMO-*Bs*TarK was not 100% completed. Nevertheless, the amount of undigested target protein is small, and the major component is the native *Bs*TarK protein. Some impurities remain; therefore, a third purification step is required.

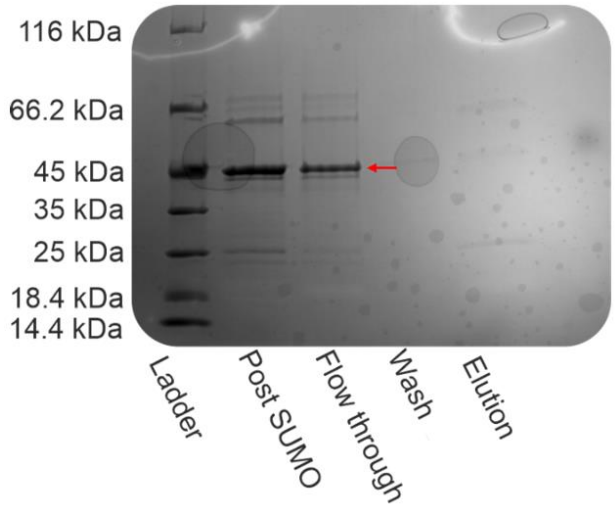


Figure 4.14: Second IMAC purification of the SUMO protease treated His-SUMO-BsTarK from the previous purification. The red arrow indicates the presence of the digested target protein, native BsTarK, in the flow through fraction (expected mass ~45 kDa). Resolved on a 4-15% polyacrylamide gel.

The flow through of the previous IMAC purification was subsequently purified by size exclusion chromatography (SEC) (**4.4.4**). After the purification was completed, peaks in the 280 nm absorption spectrum were identified and eluted fractions under these peaks were sampled. The sampled fractions and a sample taken from the loaded mixture were prepared for SDS-PAGE analysis (see **7.2.8**) (**Figure 4.15**).

The very intense band in the loaded sample suggests that *Bs*TarK is the majority species before SEC proceeded. This lane also highlights the numerous impurities present at ~18-35 kDa and ~66 kDa. The band at ~66 kDa is likely the His-SUMO-*Bs*TarK which did not get cleaved by SUMO protease. The bands observed across fractions 1D10-1F10 are at the expected mass of *Bs*TarK (~45 kDa) indicating that *Bs*TarK has been purified. However, in fractions 1E10-1F6 there is a faint band observed at ~66 kDa, therefore indicating that SEC has failed to separate this impurity. Nevertheless, this impurity is only a fraction of the overall

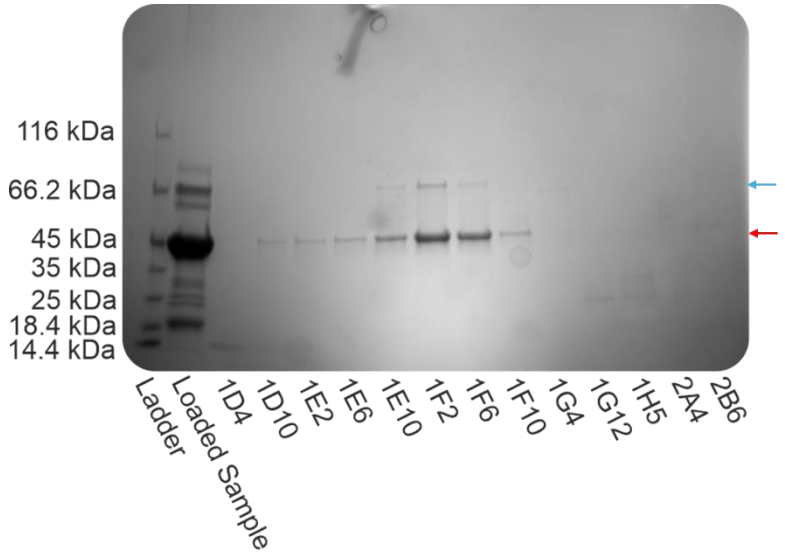


Figure 4.15: SEC of BsTarK from the previous purification. Red arrows indicate the presence of the digested target protein, native BsTarK (expected mass ~45 kDa). Blue arrow indicates the presence of an impurity at ~66 kDa. Resolved on a 4-15% polyacrylamide gel.

protein content. When combined with the other fractions which only contain *Bs*TarK (1D10-1E6 and 1F10), the impurities may be negligible for *Bs*TarK crystallography and enzymatic characterisation.

Later attempts to purify *Bs*TarK involved using the same expression conditions as the previous method except the cell line used was Rosetta2(DE3) pLysS and the plasmid used contained the C-terminally His6 tagged *Bs*TarK (p*Bs*TarK-His28a, **Table 7-1,** entry 4). The second purification method (outlined in section **4.2.3**, **Table 4-2** entry 2) uses a heparin affinity column purification step which was used in the published procedure for TagF purification and allowed the protein structure to be elucidated by X-ray crystallography (Lovering *et al.*, 2010). This column contains immobilised heparin which is nature's most negatively charged biomolecule. The negatively charged ligand could mimic the negatively charged TagB produced which the active site of *Bs*TarK interacts with therefore *Bs*TarK could bind the heparin strongly. Furthermore, the paper mentions that the 'heparin purification of TagF separated unfolded from folded protein, and highly pure samples from this procedure were critical in obtaining diffracting crystals.' As TarK is part of the same enzyme family, the heparin purification method could serve the same function in producing highly pure *Bs*TarK for crystallography.

The first step of this second purification method uses IMAC similarly to the first step of the method used for His-SUMO-BsTarK (see **4.2.3**). Samples of the total lysate were taken and after purification completed samples of the flow through, wash and the fractions which contained protein, according to 280 nm absorption peaks observed, were taken. All samples were prepared for SDS-PAGE analysis (see **7.2.8**) (**Figure 4.16**).

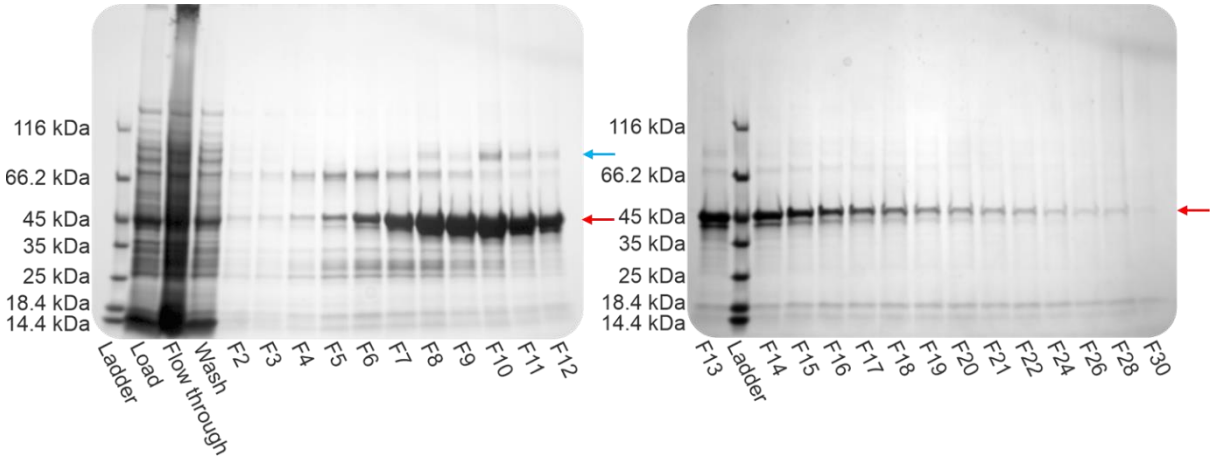


Figure 4.16: IMAC purification of BsTarK-His from E. coli lysate. Red arrows indicate the presence of target protein in the eluted fractions (expected mass ~45 kDa). Blue arrow indicates possible dimer (~90 kDa). Resolved on a 4-15% polyacrylamide gel.

Intense bands at around the expected mass of *Bs*TarK-His (~46 kDa) can be observed in fractions 4-28. This indicates successful capture and elution of the His6 tagged protein from the cell lysate. However, within some fractions which contain the target protein, there are impurities. Fractions 4-13 appear to contain faint bands at around 14 kDa, 18 kDa, three bands between 25 and 35 kDa, and 66 kDa; the proteins at these masses are largely unknown, however the bands at the lower molecular mass (~14 kDa) could be lysozyme which is added to the lysis buffer. There is also a band observed above 66 kDa which could consist of dimeric *Bs*TarK-His (~90 kDa). The dimerisation may be useful for cryo-EM studies unless the dimerisation is caused by protein aggregation.

To purify *Bs*TarK-His further, a heparin affinity purification column was used (see **4.4.4**). The fractions which contained the target protein were combined, buffer exchanged (see **7.2.7**) into the required buffer (**Table 4-3**) and a sample was taken before starting the purification method. After purification completed, samples were taken of the flow through, wash and the fractions which contained protein according to detected absorption at 280 nm. All samples were prepared for SDS-PAGE analysis (**7.2.8**) (**Figure 4.17**).

From the gel, bands at around the expected mass of *Bs*TarK-His (~46 kDa) can be observed in all fractions sampled. The most intense bands representing successful purification of *Bs*TarK-His can be observed in fractions A9-B12. The faint bands present at the expected mass of *Bs*TarK-His in the flow through and wash fractions indicate that not all the target protein bound to the column or bound with insufficient affinity. This could be due to either the column becoming oversaturated and therefore unable to bind more of the target protein, or

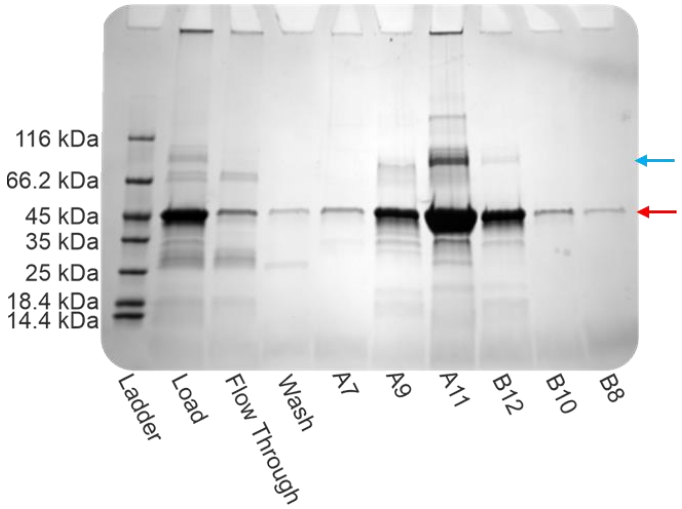


Figure 4.17: Second purification step using a heparin affinity column of BsTarK-His from the previously eluted fractions. Red arrows indicate the presence of target protein in the eluted fractions (expected mass ~45 kDa). Blue arrow indicates possible dimer (~90 kDa). Resolved on a 4-15% polyacrylamide gel.

the protein could be partially unfolded and therefore unable to bind/bind tightly (Lovering *et al.*, 2010). In fraction A11, another band can be observed above 66 kDa. This may be the hypothesised *Bs*TarK-His dimer observed previously. Furthermore, there are faint bands observed in the fractions which contain the highest amount of target protein between 18 kDa and 35 kDa. These bands could represent a small number of impurities or degradation products. Nevertheless, the impurities observed are a small fraction of the total protein and when fractions A9-B12 are combined, the impurities may be negligible for *Bs*TarK crystallography and enzymatic characterisation.

4.2.4 Circular Dichroism

After overexpression and purification of BsTarK from the SUMO construct, the purified protein was buffer exchanged into a suitable buffer for variable temperature circular dichroism (CD) (see **7.2.10**). This CD buffer contained 10 mM sodium phosphate (Na-PB), 10% v/v glycerol, and 5 mM DTT at pH 7.5. CD is used to elucidate the folded regions of the protein and provide a crude melting temperature curve. The CD data were plotted using GraphPad Prism (**Figure 4.18**).

The curve generated by the variable temperature CD experiment identifies shallow minima at 208 and 222 nm, characteristic of the spectra of α-helices, suggesting that the protein is folded. As the temperature increases these peaks approach 0 mdeg indicating that the α-helices in the protein are unfolding. Unfortunately, due to increasing noise to signal ratio, the positive peaks at 193 and 195 nm, indicative of α-helices and β-sheets, were not recorded. This is likely due to the addition of DTT in the CD buffer which absorbs around the wavelength range

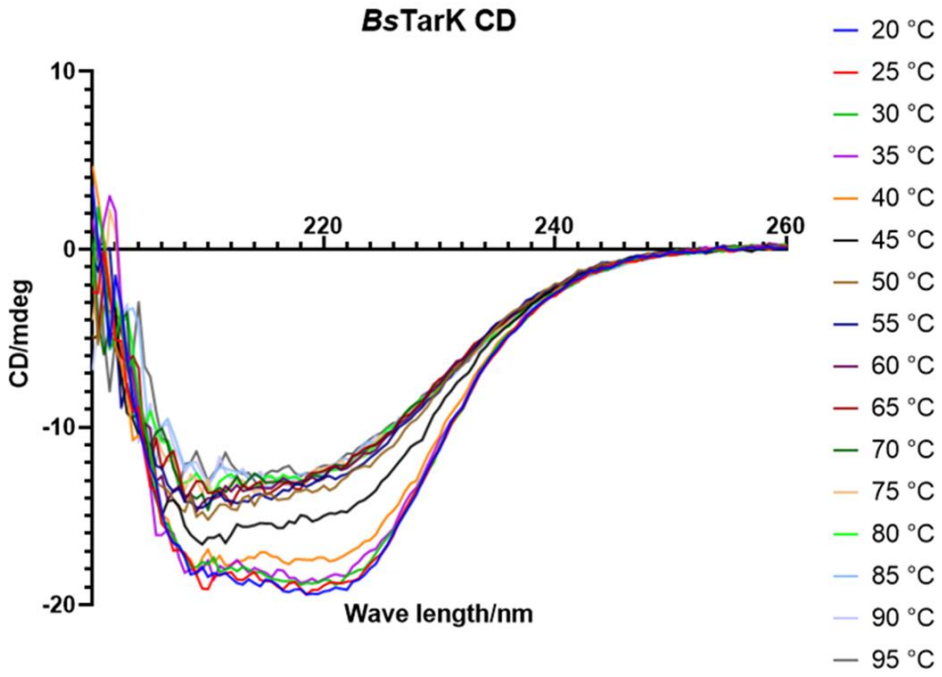


Figure 4.18: CD spectra of BsTarK at various temperatures.

used in CD experiments (Seo *et al.*, 2013). To improve the signal to noise ratio, another reducing agent should be used which does not absorb at these wavelengths such as TCEP or beta-mercaptoethanol (β-ME). Furthermore, the buffer used contains sodium phosphate and glycerol which are both able to structurally mimic parts of the enzyme's native ligands. These components could be interacting with the active site, therefore stabilising the enzyme, and leading to a misrepresentation of apo enzyme stability. For example, phosphate ions are similar to the phosphate groups present on both the acceptor ligand (glycerol phosphate) and donor ligand (CDP-ribitol). Glycerol is also present in the acceptor ligand and is structurally analogous ribitol.

Using the CD data collected at 217 and 222 nm, a melting temperature plot was derived (**Figure 4.19**). A typical sigmoidal function is produced which suggests that the protein is folded and unfolds expectedly when temperature is increased and allows the melting temperature to be calculated. The melting temperature of the protein was found to be ~44 °C. Further CD studies of *Bs*TarK in different buffers and with different ligands should be used in the future to determine the ideal conditions in which the protein is most stable in order to increase the likelihood of protein crystallisation.

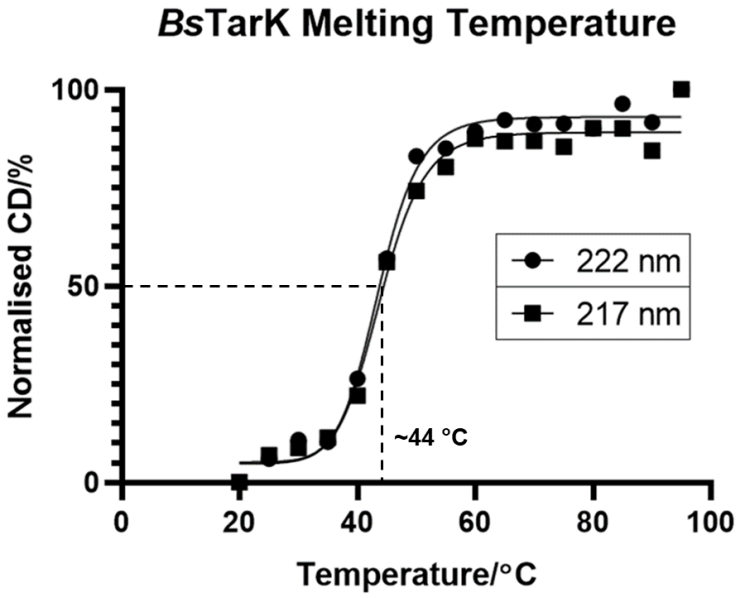


Figure 4.19: Thermal melting temperature curves of BsTarK using the CD results at 217 nm and 222 nm. Melting temperature was calculated to be ~44 °C for both the 217 nm and 222 nm curves.

4.2.5 Crystallography

Purified *Bs*TarK, from the same protein preparation and in the same buffer used for the CD experiments, was concentrated to 15mg/ml for protein crystallisation trials. Three trial plates were used: Morpheus (A) (Molecular Dimensions), Crystal Screen HT (B) (Hampton Research) and PACT (C) (Molecular Dimensions) (see **4.4.5**). Both wells of the 96 well plates were used with one containing only the protein (top well) and the other containing the protein with a ~8x molar excess of CDP and MgCl2 (bottom well). Two conditions from two different trials yielded multiple crystals in the wells containing CDP and $MgCl_2$: plate A, well C12 = AC12 (**Figure 4.20A**) and plate B, well G11 = BG11 (**Figure 4.20B**). The conditions of AC12 were 0.01 M $ZnCl_2$, 0.01 M HEPES, pH 7, 20% (w/v) polyethylene glycol 6000 and the conditions of BG11 were 0.1M HEPES, pH7.5, 70% (v/v) 2-methyl-2,4-pentanediol. These crystals were sent to the Diamond Light Source for X-ray diffraction experiments (**Figure 4.20**).

Unfortunately, the crystals from this trial did not yield obvious and characteristic protein diffraction patterns and Diamond's auto processing software failed. This suggests that these crystals are not protein. Furthermore, the diffraction patterns from some of the crystals

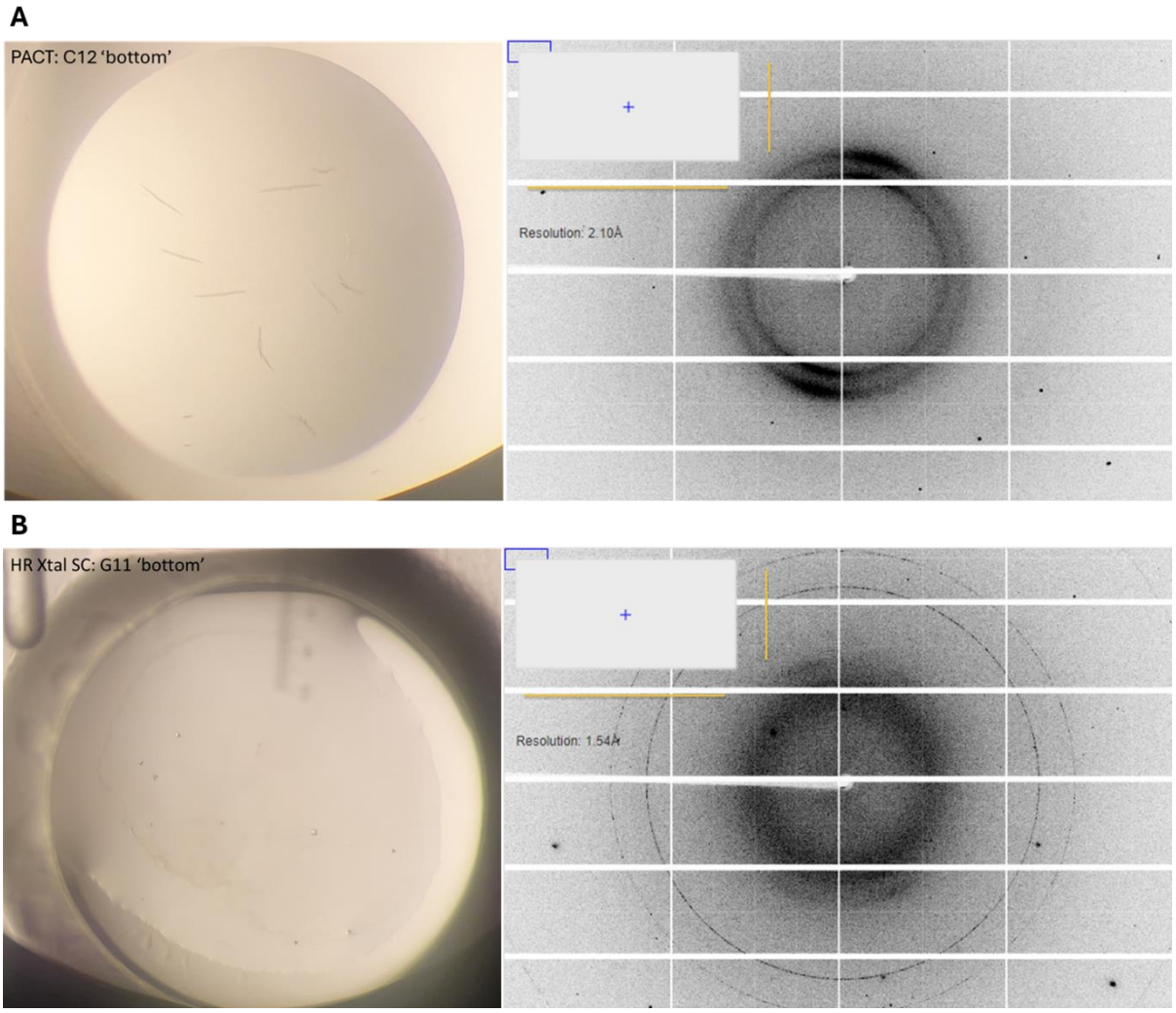


Figure 4.20: Microscopic photographs and X-ray diffraction images from BsTarK crystallisation trials. A = photograph of the crystals of the 'PACT' screen well C12 'bottom' (left) and the resulting X-ray diffraction (right). B = photograph of the crystals of the 'crystal screen HT' screen well G11 'bottom' (left) and the resulting X-ray diffraction (right).

screened did generate large spots which are spaced far apart. This is reminiscent of salt crystal diffraction. Therefore, more crystal trials were performed.

Further crystallisation trials of *Bs*TarK used the C-terminal His6 tagged protein construct (*Bs*TarK-His) from the second purification method (**4.4.4, Table 4-2**, entry 2). The protein was buffer exchanged into SEC buffer with 20% (v/v) glycerol and concentrated to 14 mg/ml and crystal screens were performed (see methods). Crystals were only observed in two wells of the PDB screen: A7 (0.2 M $Mg(OAc)_2$, 20% (w/v) polyethylene glycol 8000, 0.1 M NaCAC (sodium cacodylate), pH 6.5) and D1 (25% (w/v) polyethylene glycol 3350, 0.1 M HEPES, pH 7.5 (**Figure 4.21**). Crystals from both these wells were fished and were assessed in-house for X-ray diffraction (**Figure 4.22**).

Unfortunately, the resulting diffraction patterns from the in-house X-ray source are not reminiscent of protein diffraction but instead salt crystal diffraction. This is because the

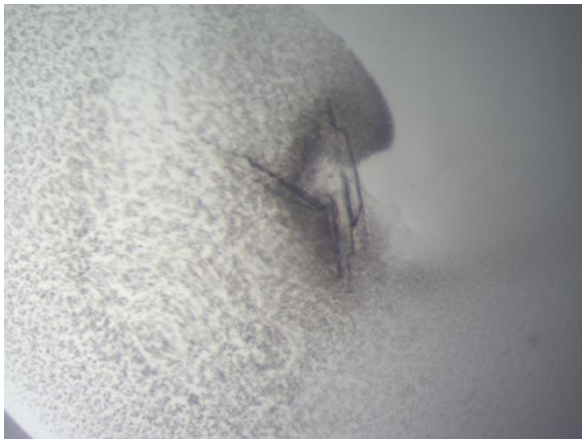


Figure 4.21: Microscopic photograph of possible BsTarK-His crystals in the PBD trial, well D1.

patterns consist of large spots spaced far apart. Some reasons behind the difficulties in *Bs*TarK crystallisation may be explained by the possible membrane binding domain present in the protein which is also in all TagF-like enzymes (Formstone *et al.*, 2008; Lovering *et al.*, 2010; Li *et al.*, 2024). Membrane associated proteins are infamously hard to crystallise and the addition of detergents can result in protein-free micelles which can hamper crystallisation (Kermani, 2021). However, removal of the detergents decreases *Bs*TarK solubility. If the protein forms a dimer naturally, possibly seen in **Figure 4.17**, then cryo-EM may be a valid technique to elucidate protein structure. This method does not require a large amount of protein, is not inhibited by the presence of detergents and has been used before to resolve the structure of another TagF-like enzyme (Li *et al.*, 2024). However, the dimer mass is only ~90 kDa which is slightly under the recommended mass for the cryo-EM equipment.

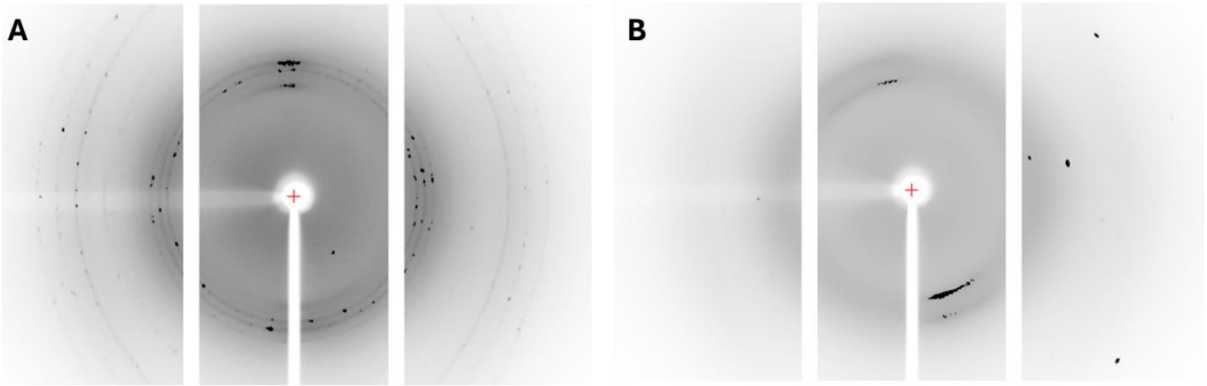


Figure 4.22: X-ray diffraction images of the possible BsTarK crystals fished from this trial. A = Diffraction pattern from a crystal fished from PDB trial, well D1. B = Diffraction pattern from a crystal fished from PDB trial, well A7.

4.2.6 Nano Differential Scanning Fluorimetry and Inhibitor Screening

Nano differential scanning fluorimetry (nanoDSF) was first used to generate melting temperature curves of *Bs*TarK-His in two different buffers at pH 6.0 and 7.5, both supplemented with CHAPS and glycerol (see **4.4.4, Table 4-3**, buffer 1 without imidazole and buffer 4) to assess stability (**4.4.6**) (**Figure 4.23**).

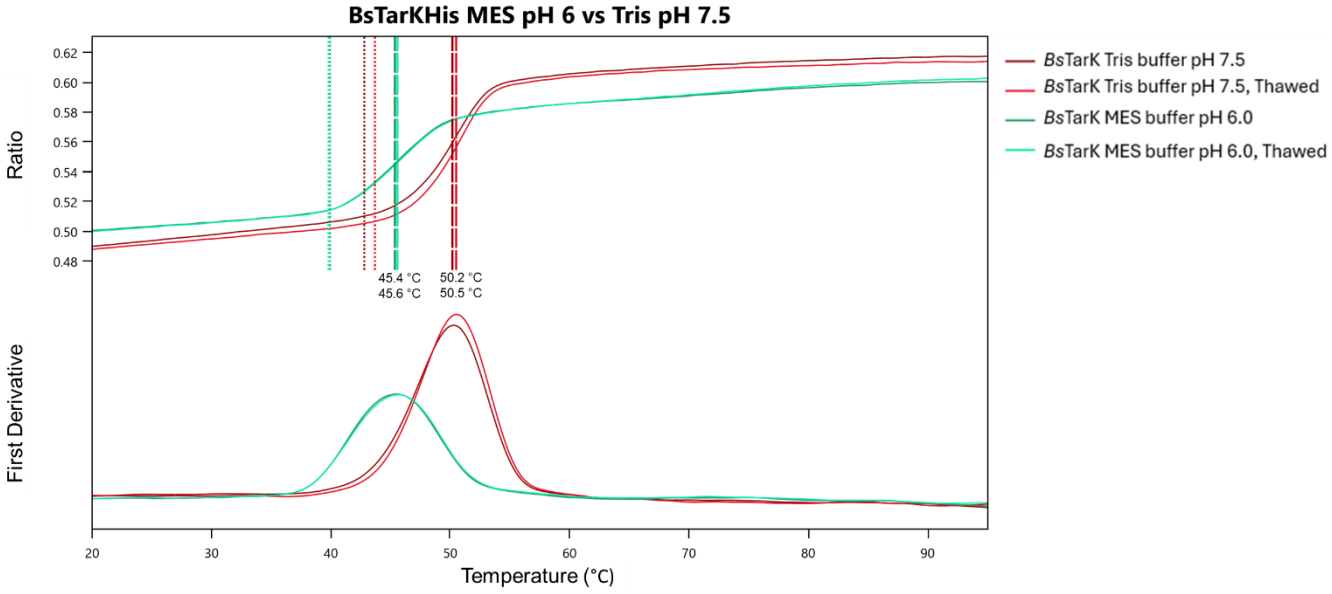


Figure 4.23: NanoDSF protein melting temperature experiment of fresh or freeze-thawed BsTarK at two different buffer pHs. The coloured dotted lines identify the onset of unfolding. The coloured dashed lines identify the inflection points and therefore the melting temperatures of BsTarK. The 350 nm:330 nm ratio of tryptophan fluorescence and the first derivative are plotted against temperature.

The curves generated by the experiment display the characteristic sigmoidal function of a protein melting temperature where the inflection point represents the melting temperature (Tm) of the protein. The protein in the pH 7.5 buffer has a higher Tm (~50 °C) than in the pH 6.0 buffer (~45 °C) suggesting that the protein is more stable at pH 7.5. The pH 6.0 buffer is used in the heparin purification step; therefore, it is important to buffer exchange the protein into the pH 7.5 buffer immediately after the purification is completed. Furthermore, the Tm of the flash frozen and thawed protein in each buffer seems to not change greatly from their freshly prepared counterparts, which suggests that freezing and subsequently thawing *Bs*TarK-His does not affect protein stability.

Next, nanoDSF was used to observe if the Tm of *Bs*TarK-His shifts with the addition of the donor substrate, CDP-Rbo. Purified protein was first exchanged into a reaction buffer used in other *Bs*TarK studies (100 mM NaCl, 20 mM Tris pH 7.5, and 30 mM $MgCl_2$, Brown *et al.*, 2010) (**4.4.6**). Three runs of *Bs*TarK-His with and without CDP-Rbo were performed in parallel resulting in a plot of six curves (**Figure 4.24**). The Tm of the protein was then calculated from the first derivative of the curves.

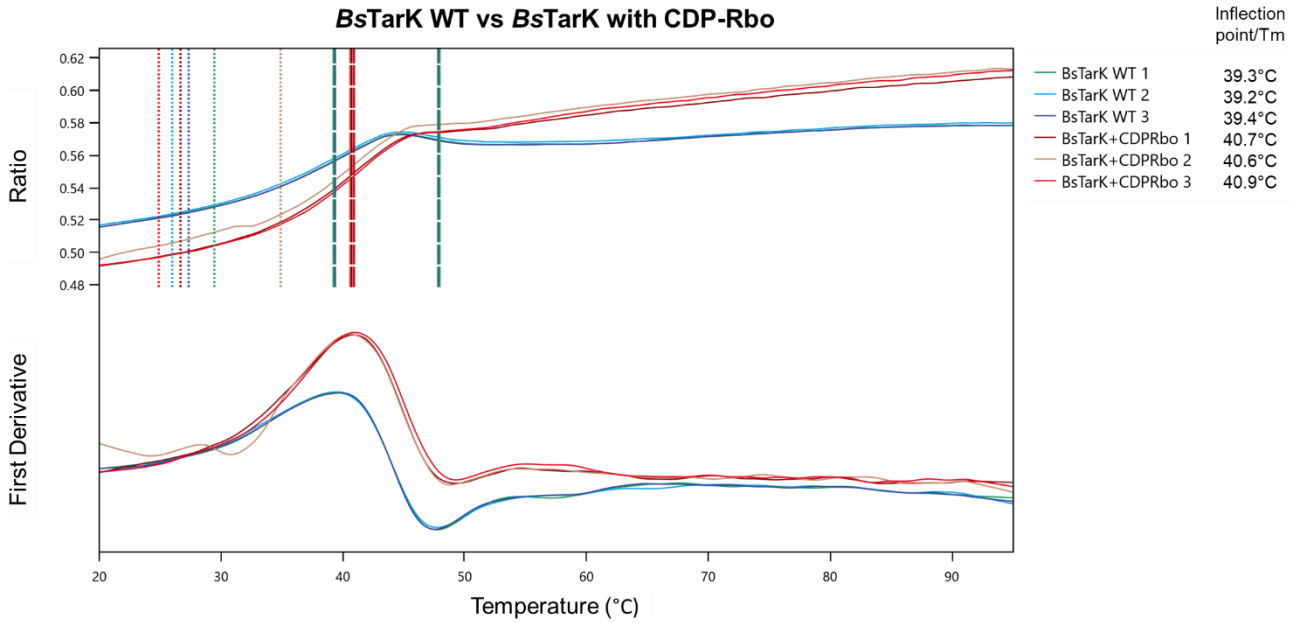


Figure 4.24: NanoDSF protein melting temperature experiment of BsTarK with and without CDP-Rbo in triplicate. The coloured dotted lines identify the onset of unfolding. The coloured dashed lines identify the inflection points and therefore the melting temperatures of BsTarK. The 350 nm:330 nm ratio of tryptophan fluorescence and the first derivative are plotted against temperature.

The curves generated by the experiment display the characteristic sigmoidal function of a protein melting temperature. The Tm of *Bs*TarK-His without donor substrate was on average 39.3 °C and the Tm of *Bs*TarK-His supplemented with CDP-Rbo was on average 40.7 °C. This difference of 1.4 °C was deemed statistically significant via un-paired T-test analysis and therefore shows that the addition of CDP-Rbo increases protein stability. However, the melting temperatures elucidated are several degrees lower than the previous experiment in the Tris buffer (**Figure 4.23**). This suggests that without the presence of CHAPS, glycerol or both additives the protein loses stability. Furthermore, the NaCl concentrations for this experiment is five times lower than in the previous one which may also impact protein solubility and stability. Nevertheless, this experiment showed that the Tm of *Bs*TarK can shift with the addition of a ligand such as CDP-Rbo, therefore, this method could be used to screen several possible inhibitors to uncover the ones which bind and stabilise *Bs*TarK with great effect.

The same experiment was subsequently performed along with six inhibitor analogues of CDP-Rbo (see **4.4.6**). Each inhibitor was evaluated in triplicate. The curves were plotted (**Figure 4.25**) and the inflection points were identified for each condition allowing Tm values to be calculated (**Figure 4.26**).

All the samples appear to display protein melting curves much like the results above, however the calculated Tm values of all the samples which contain a ligand are lower than apo *Bs*TarK. This observation incudes the samples supplemented with CDP-Rbo which are identical to the

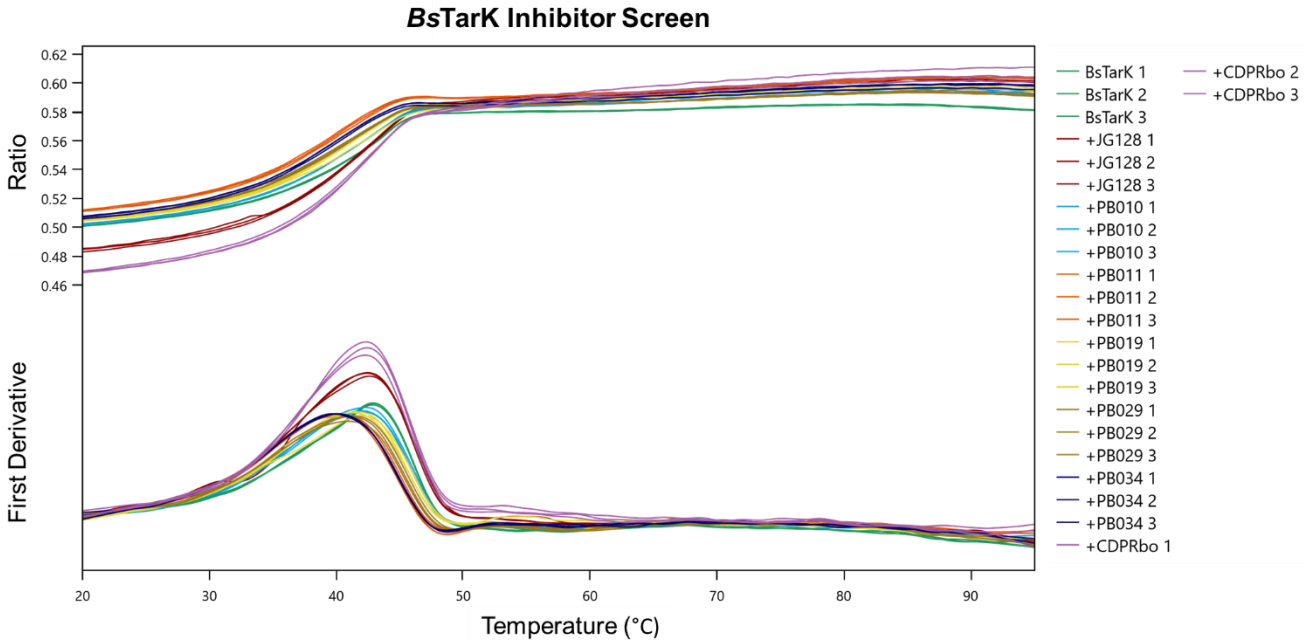


Figure 4.25: NanoDSF protein melting temperature experiment of BsTarK with and without ligands in triplicate. The 350 nm:330 nm ratio of tryptophan fluorescence and the first derivative are plotted against temperature.

previous experiment. These results indicate that that none of the inhibitors stabilise *Bs*TarK and they also call into question if CDP-Rbo has any stabilising effect.

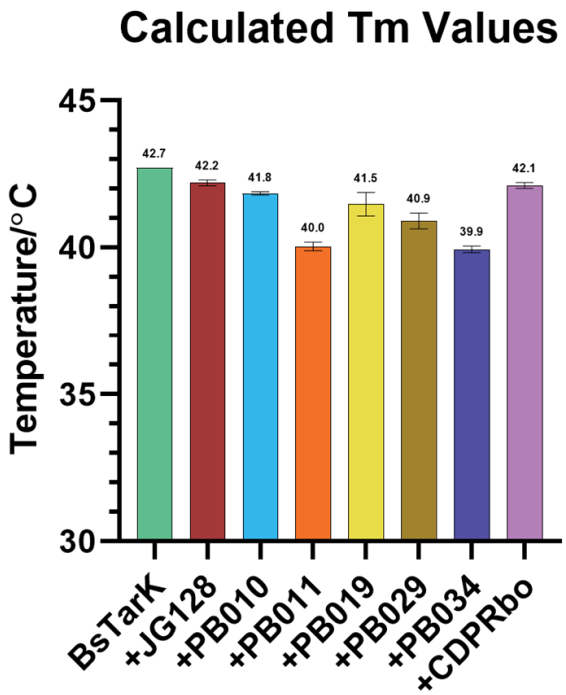


Figure 4.26: Calculated average Tm values of BsTarK with and without ligands. Specific average Tm values of each sample are displayed above their respective bars.

Another interesting observation involved the initial fluorescence readings before starting the temperature ramp experiment. The fluorescence values drop significantly with the addition of any of the ligands evaluated when compared to the apo *Bs*TarK sample. This could be indicative of substrate binding as the fluorescence signal may be quenched due to the

presence of a ligand. This is a common technique used to derive K_d values for various ligands and inhibitors (Pollard, 2010). For this reason, an experiment was performed involving 2-fold serial dilutions of the acceptor substrate CDP-Rbo and a fixed concentration of *Bs*TarK with controls (no enzyme and just buffer). Each condition was repeated in duplicate, and 330 nm fluorescence was recorded (see **4.4.6**). The data was then transformed, normalised, and plotted (**Figure 4.27**).

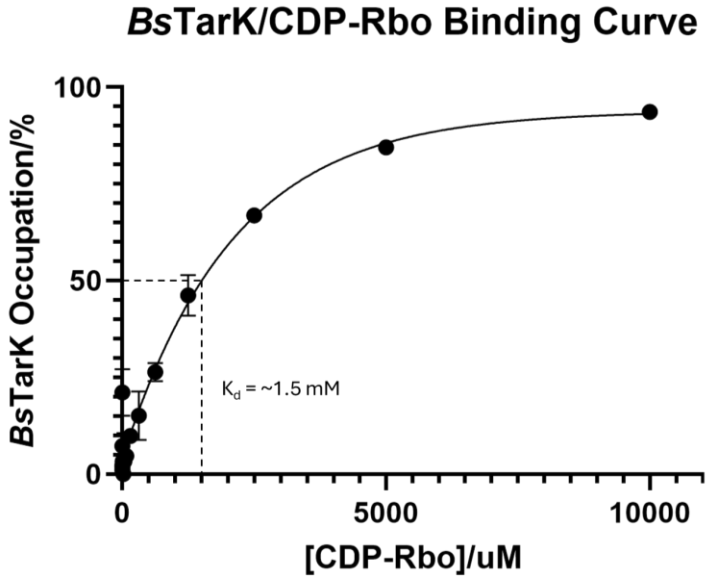


Figure 4.27: Binding curve of CDP-Rbo and BsTarK. The K_d of CDP-Rbo was estimated to be ~1.5 mM.

The graph and curve generated appear to be reminiscent of a ligand binding curve with the K_d of CDP-Rbo estimated to be ~1.5 mM. However, more data should be gathered at higher concentrations to define the curve in more detail. Also, there seems to be an outlier or two that do not fit the curve; therefore, this experiment may require a higher number of repeats. Another detail to note which was only realised after this experiment is that for these experiments the sample is excited with radiation of a wavelength of 280 nm which tryptophan residues absorb and then emit at a wavelength of 330/350 nm (depending on solvent exposure). The 280 nm wavelength may also be absorbed by the CDP-containing ligands as the cytosine ring absorbs in the 250-290 nm wavelength region (maximum ~270 nm, (Fischer, 1995)). This may mean that the decrease in fluorescence is not due to the quenching of this wavelength by the ligand but is instead caused by the ligand itself absorbing the 280 nm wavelength radiation and therefore decreasing the overall exposure and excitation of the tryptophan residues and hence decreasing the fluorescence signal when ligand concentration is increased. This could mean that the curve in **Figure 4.27**Figure 4.27: Binding curve of CDP-Rbo and BsTarK. The Kd of CDP-Rbo was estimated to be ~1.5 mM. represents absorption by CDP-Rbo rather than enzyme binding and, similarly, that any changes in the observed melting temperature curves in **Figure 4.26** are obscured by CDP-Rbo absorption.

To assess if the decrease in fluorescence is caused by the presence of the ligand not by the binding, the same experiment was performed but a different protein was used. This protein was carbonic anhydrase and is not known to bind CDP-Rbo. The 330 nm fluorescence signal was recorded, and the data was then transformed, normalised, and plotted (**Figure 4.28**).

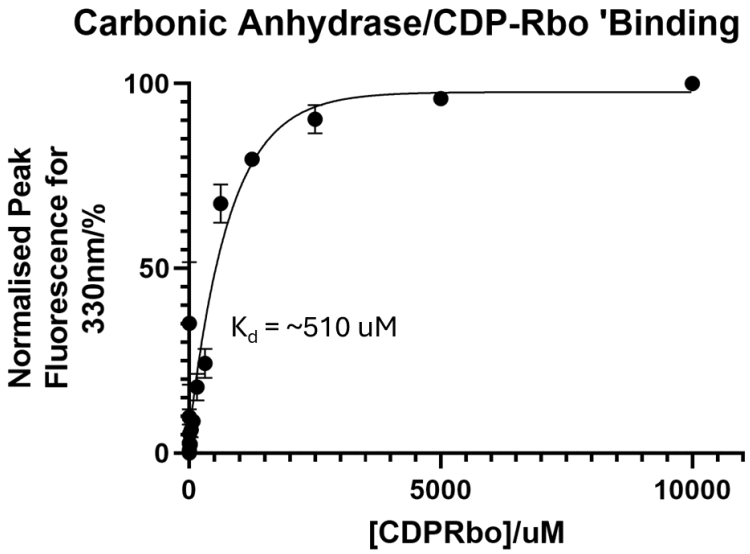


Figure 4.28: Supposed CDP-Rbo 'binding curve' of carbonic anhydrase.

The curve observed from the graph suggests that carbonic anhydrase also binds CDP-Rbo which cannot be the case. This hints that the trend observed in both experiments could be due to the cytosine ring of CDP-Rbo absorbing the 280 nm wavelength as mentioned previously. Therefore, the observed decrease in fluorescence cannot be reliably used to determine the K_d of any of the cytosine-containing ligands and inhibitors. Also, this explains the difficulty in obtaining reliable Tm values in the presence of CDP-Rbo as the ligand is likely obscuring any shifts in the fluorescence signal. However, the curves are not identical so there could be another component in the *Bs*TarK curve which may involve ligand binding and therefore effects the rate of fluorescence decay. Given more time, a more reliable assay could be developed or other methods of determining the K_d of the inhibitors could be attempted. One such method that could be used is isothermal titration calorimetry (ITC). Unfortunately, the amount of inhibitor required is quite large and exceeds the amount available at the time.

Two active site mutants of *Bs*TarK-His (H119N and H255A) were also assessed using nanoDSF to observe if the mutations influenced Tm. These mutants were selected as they are the corresponding residues which align to the catalytic residues identified in *Se*TagF, *Sa*TarL and Bcs3 (Lovering *et al.*, 2010; Cifuente *et al.*, 2023; Li *et al.*, 2024). The H119N mutation should render the enzyme non-functional which is useful in activity assays and the H255A mutant should make the enzyme non-functional and trap CDP-Rbo inside the active site which is useful for structural studies seen previously in other TagF-like protein structures

(Lovering *et al.*, 2010; Cifuente *et al.*, 2023; Li *et al.*, 2024). The mutants were synthesised using single point mutagenesis (see **4.4.3**) and the C-terminally HIS-tagged proteins were expressed and purified by using the heparin affinity column procedure outlined above (**4.4.4**). The buffers used for nanoDSF were the SEC buffer (**Table 4-3**, buffer 3) with or without 20% v/v glycerol. Unfortunately, the H119N mutant did not produce a sigmoidal curve and is not shown. Data for the WT protein and H255A variant are shown in **Figure 4.29**.

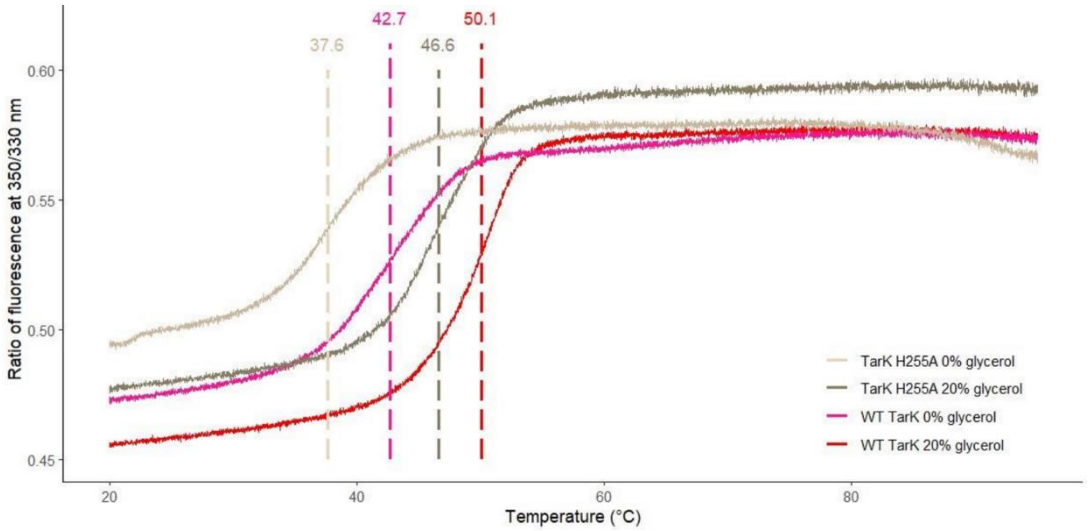


Figure 4.29: Intrinsic tryptophan fluorescence ratio at 350/330 nm of WT TarK and the TarK H255A variant. Dashed lines show the melting points for each sample, at which 50% is unfolded.

Results show that the TarK H255A variant is less stable relative to the WT regardless of the glycerol concentration: there is a 3.5 °C and 5.1 °C decrease in the variant melting temperature at 20% and 0% glycerol, respectively. This suggests that the H255A mutation in TarK decreases overall stability. It can also be observed that the presence of glycerol increases the Tm by 9 °C in the mutant, compared to the 7.4 °C increase in the WT. It is unknown why the other mutant (H119N) did not produce a standard melting temperature curve, but this may be due to inherent instability from the mutation which denatured the protein prior to this experiment.

4.2.7 Activity Assays

To assess *Bs*TarK-His activity, several reactions were performed and analysed by liquid chromatography–mass spectrometry (LC–MS) (see **4.4.7** for the reaction conditions and content). After the reaction was incubated for 30 mins at 21 °C then quenched with the addition of 30 µL of methanol, the sample was loaded onto the LC-MS machine and the MS chromatograms were used to identify starting material and/or product formation (**Figure 4.30**).

The main peak observed in mass chromatogram of the initial reaction shows no obvious product formation (MW of *Bs*TarK product = 971.27). Only the accepter substrate used (TV1055, MW = 757.24) and the donor substrate (CDP-Rbo, MW = 537.08) were detected.

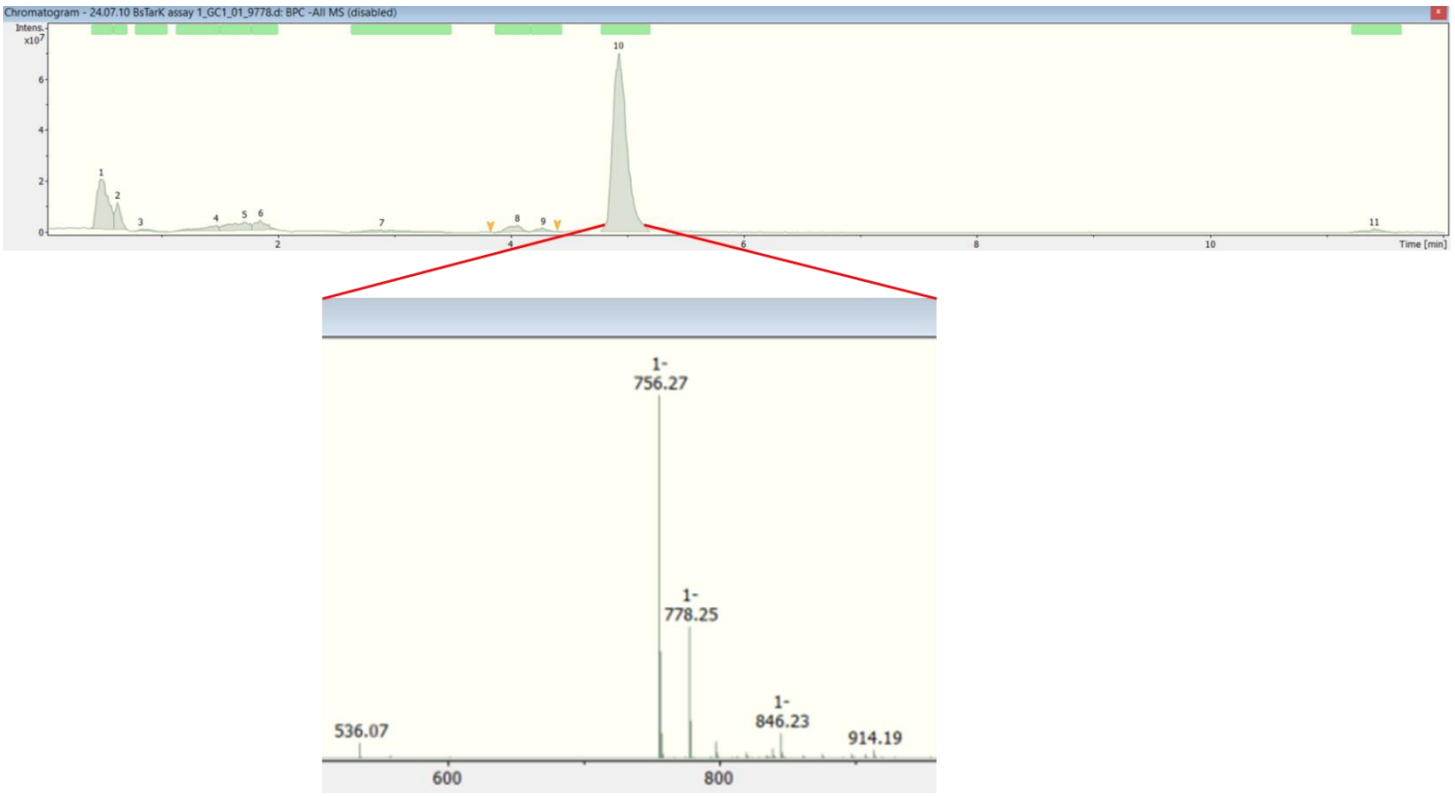


Figure 4.30: Mass chromatogram and mass spectrum of the main peak observed after the initial BsTarK reaction with the synthetic acceptor substrate TV055 and CDP-ribitol as the donor.

To confirm if any product had been synthesised, the extracted ion (EI) chromatogram tool was used (**Figure 4.31**).

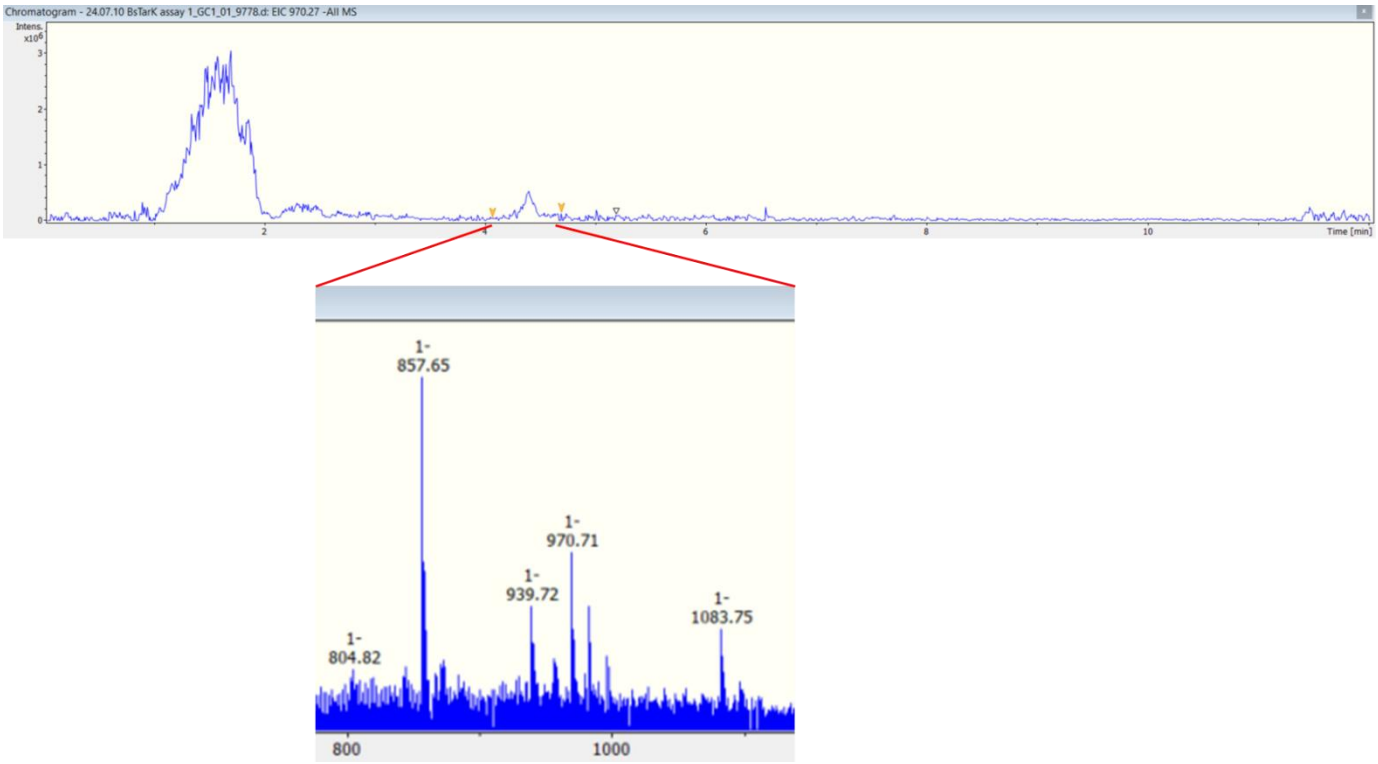


Figure 4.31: Extracted ion chromatogram and mass spectra of the peaks which contain ions with the product mass of ~971.27 from the initial BsTarK reaction. The product mass found in the spectra is shown with a red arrow.

The EI chromatogram tool identified two peaks, the first at about 1-2 mins and the second smaller peak at around 4-4.5 mins. The MS spectrum of the smaller peak showed a low intensity product signal (MW = 971.27) which suggests that the enzyme is active and synthesised the expected *Bs*TarK product. To check if this conclusion is correct, another experiment was performed involving heat treated *Bs*TarK as a negative control (**Figure 4.32**).

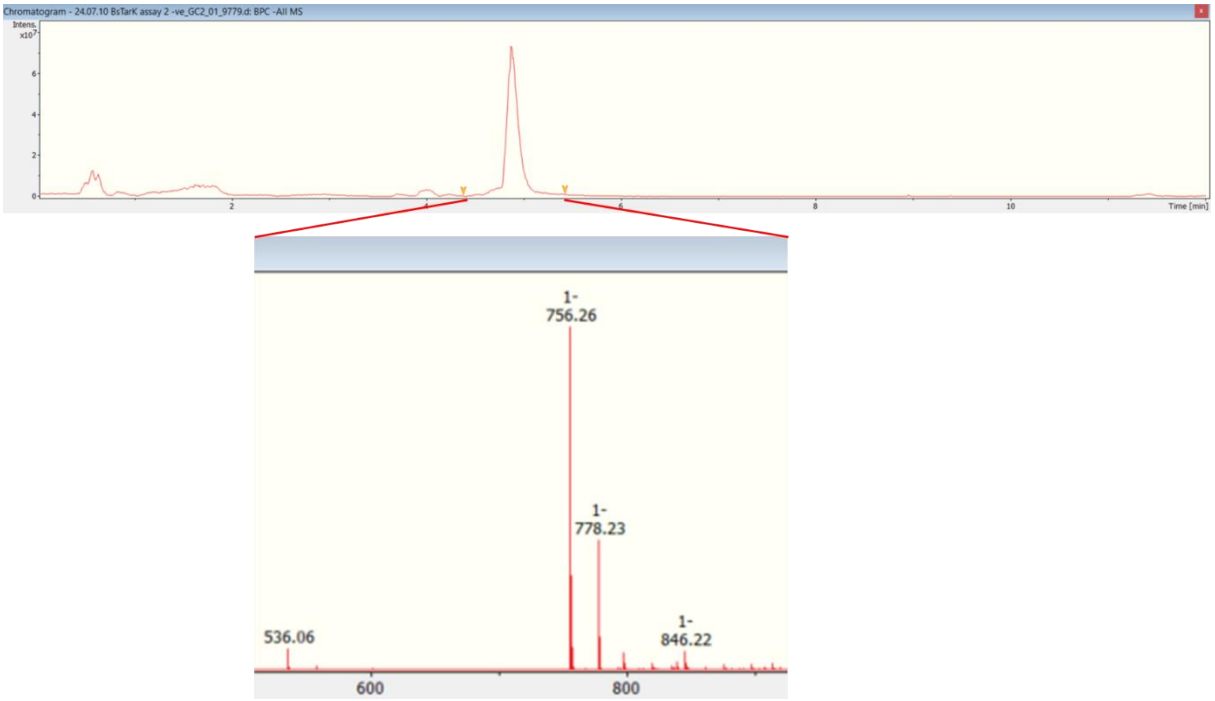


Figure 4.32: Chromatogram and mass spectra of the main peak of heat treated BsTarK reactions.

The main peak and mass spectra observed in this control reaction are very similar to the previous reaction (**Figure 4.30**). The intensities of the peaks in the MS spectra of the starting materials (CDP-Rbo MW = 537.08 and TV1041 MW = 757.24) are also comparable to the previous mass spectra. This shows that in the previous reaction the amount of *Bs*TarK product synthesised is likely quite small. The EI chromatogram tool was then used to attempt to find any masses that represent the *Bs*TarK product (**Figure 4.33**). No product peak was observed in this negative control reaction, which supports the conclusion that *Bs*TarK appears to be active as the previously observed product peak is not caused by background ions. Although the MS chromatogram cannot give comparable and quantitative results, it does appear that a large amount of the starting substrates remained after synthesising the *Bs*TarK product. This suggests that the activity of *Bs*TarK is very low. To estimate turnover rate, another assay could be developed either by using an assay such as the UMP/CMP-Glo™ Glycosyltransferase Assay (Promega) or using fluorescently tagged acceptor substrates for easier detection by UV absorbance. Furthermore, other conditions could be altered, like temperature, to find the optimal conditions for *Bs*TarK activity.

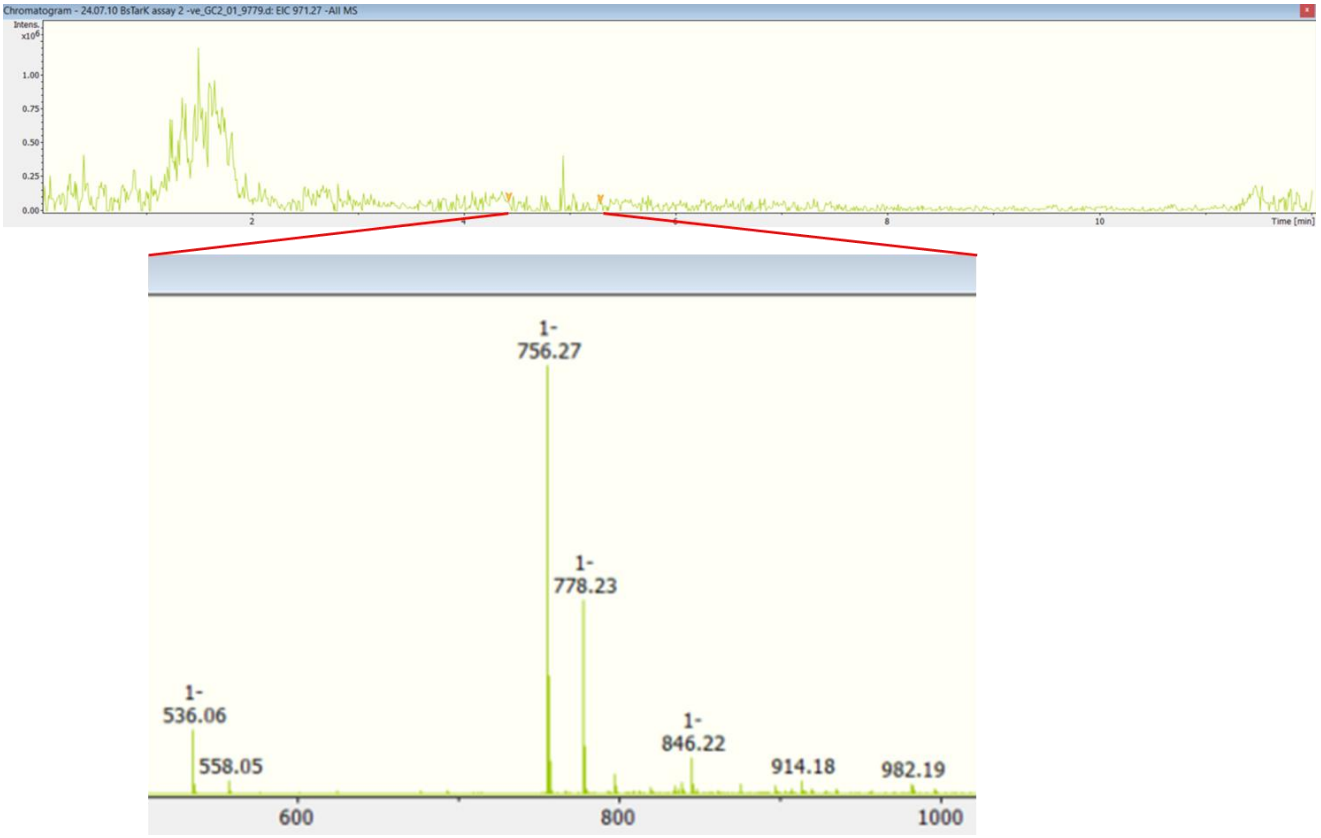


Figure 4.33: Extracted ion chromatogram and mass spectra of the peaks which may contain ions with the product mass of ~971.27 from the heat treated BsTarK reaction. Extracted ion chromatogram and mass spectra of the peaks which may contain ions with the product mass of ~971.27 from the heat treated BsTarK reaction.

4.3 Conclusions and Future Work

The cloning of various tagged *Bs*TarK constructs described in this chapter were vital in elucidating the conditions required for the soluble expression of the target protein through extensive expression trials. With the conditions finalised, purification of the target protein using a multistep protocol, with inspiration from the literature, allowed highly pure and soluble *Bs*TarK to be produced. Using the soluble protein produced, early characterisation of *Bs*TarK and the active site mutant has been performed, with mixed results, and the activity of the WT enzyme has been tested. Furthermore, preliminary protein crystallisation trials have been performed, however no protein crystals have been generated.

Future work on *Bs*TarK should focus on further biochemical characterisation. The first step should be testing different methods to assess the binding of the various ligands. This should use other equipment which do not interact with the moieties of the ligands to detect melting temperature shifts with the addition of ligands, like fluorescence dye based thermal shift assay. Other methods include ITC to measure the binding energies of the ligands. However, a larger amount of inhibitor is required and therefore may not be time efficient.

Later work on *Bs*TarK should involve solving the protein structure with further crystallisation trials, possibly in complex with ligands which bind tightly to the enzyme and using the mutants generated in this work to trap native ligands like CDP-Rbo. Cryo-EM could be an attractive alternative for structure elucidation if *Bs*TarK forms a larger oligomeric complex, however more work is needed to explore multimerisation.

Finally, the amount of starting substrate remaining and the small amount of the *Bs*TarK product formed from the LC-MS assays of *Bs*TarK suggests that it is not very active. To estimate the turnover rate, an alternative assay should be developed, either by utilizing an assay like the UMP/CMP-Glo™ Glycosyltransferase Assay (Promega) or by employing fluorescently tagged acceptor substrates for more straightforward detection via UV absorbance. Additionally, other parameters, such as temperature, should be modified to identify the optimal conditions for *Bs*TarK activity.

4.4 Methods

4.4.1 Molecular Cloning

Linearisation of the Champion™ pET SUMO vector (Invitrogen) and synthesis of the *Bs*TarK insert was performed using Q5 High-Fidelity DNA Polymerase (New England Biolabs) PCR and primers designed from the primer design tool (Takara Bio) (see **7.2.1**) For linearisation of the vector the primers used were 'linear pSUMO F' and 'linear pSUMO R' (see **Table 7-2,** entries 7 and 8). For *Bs*TarK insert synthesis the pUC57*Bs*TarK plasmid (GenScript) was used with the primers 'SUMOTarKBs F' and 'SUMOTarKBs R' (see **Table 7-2**, entries 1 and 2). Correct insert and linearised vector formation were confirmed by agarose gel electrophoresis. Subsequently, DNA isolation, ligation and transformation of *E. coli* was performed as described in section **7.2.1**.

The cloning of the p*Bs*TarK-His28a plasmid proceeded using the methods outlined in **7.2.1**. Linearisation of the vector was achieved using the 'p28aCtermHis lin F' and 'p28aCtermHis lin R' primers (see **Table 7-2**, entries 19 and 20). Synthesis of the insert was achieved using the 'pBsTarK-His-28a F' and 'pBsTarK-His-28a R' primers (see **Table 7-2**, entries 21 and 22). Correct insert and linearised vector were confirmed, isolated, ligated and transformed using the methods outlined in **7.2.1**.

4.4.2 Expression Trials

The first trial proceeded using the pSUMO*Bs*TarK plasmid and the conditions in **Table 4-1**, entry 1. Overnight cultures were first grown using 5-10 mL of kanamycin LB and inoculated with a colony from the transformation plate of BL21 (DE3) containing the pSUMO*Bs*TarK construct. Flasks containing 50 mL of kanamycin LB were subsequently inoculated with the overnight cultures and grown at 37 °C, 200 RPM until the desired OD_{600} was reached. The cultures were then induced with 1 mM IPTG and incubated overnight at 16 °C at 200 RPM. 1 mL samples were taken before induction and after overnight incubation. The samples were pelleted by centrifugation at 8,000 xg for 5 mins and pellets were lysed using 20 µL of BugBuster reagent (Novagen). Insoluble cell debris was pelleted by centrifugation (16,000xg for 20mins) and the soluble fraction was collected. Cell debris was then resuspended as much as possible in 20 µL of BugBuster reagent. Samples were then taken for SDS-PAGE and Western blot analysis (see general methods SDS-PAGE gels and Western Blotting).

Table 4-1: List of the BsTarK expression trials and the variables used.

Trial no.	Plasmid	OD_{600} range	IPTG conc (mM)	Expression temp and max time	*E. coli* Strain	Media
1	pSUMO*Bs*TarK	0.5-0.8	1.0	16 °C, O/N	BL21(DE3)	LB
2	pSUMO*Bs*TarK	0.5-0.8	0.1, 0.25, 0.5 or 1.0	16 °C, O/N	BL21(DE3)	LB
3	pSUMO*Bs*TarK	0.5-0.8	1.0	16 °C, O/N	BL21(DE3)	TB, SB or 2xYT
4	pHis-*Bs*TarK28a or p*Bs*TarK-His28a	0.5-0.8	1.0	16 °C, O/N	Rosetta2(DE3) pLysS	2xYT
5	pHis-*Bs*TarK28a	0.5-0.8	1.0	16 °C, O/N	Rosetta2(DE3) pLysS or SHuffle lysY	2xYT
6	p*Bs*TarK-His28a	0.5-0.8	1.0	16 °C, O/N	Rosetta2(DE3) pLysS or SHuffle lysY	2xYT

The second trial proceeded using the conditions outlined in **Table 4-1,** entry 2. Overnight cultures were first grown using 5-10 mL of kanamycin LB and a colony looped from the transformation plate of BL21 (DE3) containing the pSUMO*Bs*TarK construct or from glycerol stocks stored at -70 °C. Flasks containing 50 mL of kanamycin LB were subsequently inoculated with the overnight cultures and grown at 37°C, 200RPM until the desired OD_{600} was reached. The cultures were then induced with IPTG at the concentrations listed (**Table 4-1** entry 2) and incubated for the necessary time (2 hrs to overnight) at 16C at 200RPM. 1 mL samples were taken before induction, at hourly intervals and after overnight incubation. The samples were pelleted by centrifugation at 8000xg for 5 mins and pellets were resuspended in 1 mL of lysis buffer (100 mM Tris, 500 mM NaCl, 0.6% (w/v) CHAPS and 20% (v/v) glycerol, pH7.5) supplemented with protease inhibitors, benzonase and lysozyme and lysed by sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (16,000 xg for 20 mins) and the soluble fraction was collected. Cell debris was then resuspended as much as possible using the 1 mL of lysis buffer). All the samples were then prepared for SDS-PAGE analysis (see **7.2.8**).

The third trial used the conditions outlined in **Table 4-1,** entry 3. Overnight cultures were first grown using 5-10 mL of kanamycin LB and a colony looped from the transformation plate of BL21 (DE3) containing the pSUMO*Bs*TarK construct, or from glycerol stocks stored at -70 °C. Flasks containing 50 mL of kanamycin TB, SB or 2xYT were subsequently inoculated with the overnight cultures and grown at 37 °C, 200 RPM until the desired OD_{600} was reached. The cultures were then induced with 1 mM IPTG and incubated for the necessary time and

temperature at 200 RPM. 1 mL samples were taken before induction, at the 4-hr mark and after overnight incubation. The samples were pelleted by centrifugation at 8,000 xg for 5 mins and pellets were resuspended in 1mL of lysis buffer supplemented with protease inhibitors, benzonase and lysozyme (100 mM Tris, 500 mM NaCl, 0.6% (w/v) CHAPS and 20% (v/v) glycerol, pH7.5) and lysed by sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (16,000xg for 20 mins) and the soluble fraction was collected. Cell debris was then resuspended as much as possible using the 1 mL of lysis buffer). All the samples were then prepared for SDS-PAGE analysis (see **7.2.8**).

The fourth trial used the conditions outlined in **Table 4-1,** entry 4. This trial used the N-terminal His6 tagged *Bs*TarK construct (pHis-BsTarK28a, ordered from GenScript) and the synthesised C-terminal His6 tagged *Bs*TarK construct (pBsTarK-His28a). Overnight cultures were first grown using 5-10 mL of kanamycin LB and a colony looped from the transformation plate of Rosetta2(DE3) pLysS containing either the N- or C-terminal tagged constructs, or from previously made glycerol stocks stored at -70 °C. Flasks containing 50 mL of kanamycin 2xYT were subsequently inoculated with the overnight cultures and grown at 37 °C, 200 RPM until the desired OD_{600} was reached. The cultures were then induced with 1 mM IPTG and incubated overnight at 16 °C and at 200 RPM. 1 mL samples were taken before induction and after overnight incubation. The samples were pelleted by centrifugation at 8,000 xg for 5 mins and pellets were resuspended in 1 mL of 'high' CHAPS lysis buffer (30 mM Na-PB, 500 mM NaCl, 5% (v/v) glycerol, 2 mM βME, 40 mM CHAPS, pH 7.5) supplemented with protease inhibitors, benzonase and lysozyme and lysed by sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (16,000 xg for 20 mins) and the soluble fraction was collected. Cell debris was then resuspended as much as possible using the 1 mL of lysis buffer). All the samples were then prepared for SDS-PAGE and Western blot analysis (see **7.2.8** and **7.2.9**).

The fifth expression trial used the conditions outlined in **Table 4-1,** entry 5. Overnight cultures were first grown using 5-10 mL of kanamycin LB and a colony looped from the transformation plate of either Rosetta2(DE3) pLysS or SHuffle lysY containing the N-terminal His6 tagged *Bs*TarK construct (pHis-BsTarK28a). Flasks containing 50 mL of kanamycin 2xYT were subsequently inoculated with the overnight cultures and grown at 37 °C, 200 RPM until the desired OD_{600} was reached. The cultures were then induced with 1 mM IPTG and incubated overnight at 16 °C and at 200 RPM. 1 mL samples were taken before induction and after overnight incubation. 1 mL samples were taken before induction and after overnight incubation. The samples were pelleted by centrifugation at 8,000 xg for 5 mins and pellets

were resuspended in 1ml of lysis buffer (100 mM Tris, 500 mM NaCl, 0.6% (w/v) CHAPS and 20% (v/v) glycerol, pH7.5) supplemented with protease inhibitors, benzonase and lysozyme and lysed by sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (16,000 xg for 20 mins) and the soluble fraction was collected. Cell debris was then resuspended as much as possible using the 1 mL of lysis buffer. All the samples were then prepared for SDS-PAGE and Western blot analysis (see **7.2.8** and **7.2.9**).

The sixth and final expression used conditions outlined in **Table 4-1,** entry 6. This method of this trial is identical to the previous trial, however the plasmid used was the C-terminal His6 tagged *Bs*Tark construct (pBsTarK-His28a).

4.4.3 Single Point Mutagenesis

The standard method of mutagenesis is described in section **7.2.4**. For the *Bs*TarK H119N mutant, the primers 'BsTarK H119N F' and 'BsTarK H119N R' (**Table 7-2**, entries 29 and 30) were used for the PCR reactions. For the *Bs*TarK H255A mutant, the primers 'BsTarK H255A F' and 'BsTarK H255A R' (**Table 7-2,** entries 31 and 32) were used for the PCR reactions. Both mutants were made using the pBsTarK-His28a plasmid as the template.

4.4.4 Protein Purification

Two methods of purifying *Bs*TarK fusion proteins were developed (**Table 4-2**). The buffers for each step of the two methods are listed in **Table 4-3**.

Table 4-2: Outline of the two methods used for purifying BsTarK.

Method no.	Protein	Step 1		Step 2		Step 3	
		Column method	Buffers used	Column method	Buffers used	Column method	Buffers used
1	His-SUMO-*Bs*TarK	IMAC (Ni^{2+})	His A/B	IMAC (Ni^{2+})	His A/B	SEC	SEC
2	*Bs*TarK-His	IMAC (Ni^{2+})	His A/B	AC (Heparin)	Hep A/B	N/A	N/A

Table 4-3: List of buffers used for purifying BsTarK.

Buffer no.	Name	Buffer component conc	Salt conc	CHAPS conc	Other additives concs	pH	Imidazole conc
1	His A	100 mM Tris	500 mM NaCl	0.6% (w/v)	2 mM β-ME, 20% (v/v) glycerol	7.5	30 mM
2	His B	100 mM Tris	500 mM NaCl	0.6% (w/v)	2 mM β-ME, 20% (v/v) glycerol	7.5	500 mM

3	SEC	100 mM Tris	500 mM NaCl	0.6% (w/v)	2 mM β-ME, 10% (v/v) glycerol	7.5	N/A
4	Hep A	100 mM MES	500 mM NaCl	0.6% (w/v)	2 mM β-ME, 20% (v/v) glycerol	6.0	N/A
5	Hep B	100 mM MES	2 M NaCl	0.6% (w/v)	2 mM β-ME 20% (v/v) glycerol	6.0	N/A

The first purification method used the engineered pSUMO*Bs*TarK plasmid which was expressed using the conditions in **Table 4-1**, entry 3 using the 2xYT media. Several larger scale cultures of about 1 L of 2xYT media were inoculated with overnight cultures of BL21(DE3) transformed with pSUMO*Bs*TarK. The cultures were incubated at 37 °C, 200 RPM until the desired OD_{600} was reached. Protein expression was induced with 1 mM IPTG and incubated overnight at 16 °C and at 200 RPM. The cultures were pelleted at 8,000 xg for 30 mins at 4 °C using ultracentrifugation and the cell paste was stored at -70 °C. The pellet was then resuspended in 50 mL of the His A buffer, supplemented with protease inhibitors, benzonase and lysozyme, and was lysed using sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (38,000 xg for 30 mins) and the soluble fraction was collected. Cell lysate was subsequently loaded onto a 5 mL HisTrap FF Crude column (Cytiva), equilibrated with His A buffer, using an AKTA Go for IMAC purification. Flow through was collected and the column was washed with 15 column volumes (CV) of His A buffer. Proteins were then eluted off the column with a linear gradient of 0-100% of His B buffer over 15 column volumes. 2 mL fractions were collected. The soluble and insoluble fractions, flow through, wash and fractions of interest were sampled and prepared for SDS-PAGE analysis (see **7.2.8**).

The fractions which contained the target protein were combined and concentrated and buffer exchanged into His A buffer (see **7.2.7**). The solution was then prepared for SUMO protease cleavage by adding 10 µg of SUMO protease per 1 mg of protein (see **7.2.6**). The protease reaction proceeded at 20 °C at 200 RPM overnight and samples were taken before and after the reaction. The completed reaction mixture was then loaded onto a 5 mL HisTrap FF Crude column (Cytiva), equilibrated with His A buffer, using an AKTA Go. The flow though was collected, and the column was washed with 5 CV of His A buffer. Bound proteins were subsequently eluted off the column with 5 CV of 100% His B buffer. Samples of the flow through, wash and elution were taken and, along with the pre and post SUMO reaction samples, were prepared for SDS-PAGE analysis (see **7.2.8**).

The fractions which contained the target protein were combined and buffer exchanged into SEC buffer (see **7.2.7**). The sample was concentrated to ~500 µL and applied onto a Superdex 200 Increase 10/300 GL column which was previously equilibrated with SEC buffer. The sample was centrifuged to remove any precipitate and subsequently injected into the system. The purification proceeded at 0.35 mL/min with SEC buffer until 30 mL was washed through. 0.5 mL fractions were collected, and samples were taken from fractions of interest defined by the UV trace. These fractions were prepared for SDS-PAGE (see **7.2.8**). The fractions containing pure *Bs*TarK were combined and prepared for downstream applications.

The second purification method used the engineered plasmid pBsTarK-His28a which was expressed using the conditions in **Table 4-1,** entry 6 using the Rosetta2(DE3) pLysS strain. Several larger scale cultures of about 1 L of 2xYT media were inoculated with overnight cultures of Rosetta2(DE3) transformed with pBsTarK-His28. The cultures were incubated at 37 °C, 200 RPM until the desired OD_{600} was reached. Protein expression was induced with 1 mM IPTG and incubated overnight at 16 °C and at 200 RPM. The cultures were pelleted at 8000 xg for 30 mins using ultracentrifugation and the cell paste was stored at -70 °C. The pellets were resuspended in His A buffer supplemented with protease inhibitors, benzonase and lysozyme, and was lysed using sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (38,000 xg for 30 mins) and the soluble fraction was collected. Cell lysate was subsequently loaded onto a 5 mL HisTrap FF Crude column (Cytiva), equilibrated with His A buffer, using an AKTA Go for IMAC purification. Flow through was collected and the column was washed with 15 column volumes (CV) of His A buffer. Proteins were then eluted off the column with a linear gradient of 0-100% of His B buffer over 20 CV. 2 mL fractions were collected. The soluble and insoluble fractions, flow through, wash and fractions of interest were sampled and prepared for SDS-PAGE analysis (see **7.2.8**).

The fractions of interest containing the target protein were combined and buffer exchanged into Hep A buffer (see **7.2.7**). The sample was subsequently loaded onto a 5 mL HiTrap Heparin HP affinity column, equilibrated in Hep A buffer, using an ATKA Go for AC purification. Proteins were then eluted off the column with a linear gradient of 0-100% Hep B buffer over 10CV. 2 mL fractions were collected. Samples from the loaded volume, flow through, wash and fractions of interest were prepared for SDS-PAGE analysis (see **7.2.8**).

The mutant variants of *Bs*TarK-His, H119N and H255A, were purified using the second set of purification methods described above.

4.4.5 Crystallography

Initial crystallography trials used pure *Bs*TarK from the first set of purification methods. First, the protein was buffer exchanged into CD buffer (10 mM Na-PB, 10% (v/v) glycerol, 5 mM DTT, pH 7.5) and concentrated to ~15 mg/mL (see **7.2.7**). A sample was taken and supplemented with CDP and $MgCl_2$ at ~8x molar excess. Both the apo and CDP ligand protein solutions were then dispensed in a ratio of 150 nl:150 nL of protein to reservoir solution into three 96 well plate crystallisation trials, Morpheus (Molecular Dimensions), Crystal Screen HT (Hampton Research) and PACT (Molecular Dimensions), using the Mosquito robot (SPT Labtech). Apo protein was placed in the top well and CDP ligand plus protein was placed in the bottom well for sitting drop vapour diffusion crystallization. The plates were left for ~3 weeks and crystals were fished, flash frozen in liquid nitrogen and sent to the Diamond Light Source for X-ray diffraction.

The second set of crystallography trials used pure *Bs*TarK-His from the second set of purification methods. The purified protein was buffer exchanged into a buffer containing 100 mM Tris, 500 mM NaCl, 0.6% (w/v) CHAPS, 20% (v/v) glycerol and 2 mM BME at pH 7.5. The protein was concentrated to ~14 mg/mL (see **7.2.7**). Protein solutions were then dispensed in a ratio of 150 nL:150 nL of protein to reservoir solution into six 96 well plate crystallisation trials; Crystal Screen HT (Hampton Research), JCSG+ (Molecular Dimensions), Morpheus (Molecular Dimensions), PEG/Ion (Hampton Research), PBD (University of York) and PGA (Molecular Dimensions), using the Mosquito robot (SPT Labtech). The sitting drop vapour diffusion method was used. The plates were left for ~3 weeks and crystals were fished, flash frozen in liquid nitrogen and evaluated for X-ray diffraction using the in-house X-ray source (Rigaku Oxford Diffraction XtaLAB Synergy-R (Cu) Generator with HyPix ARC 150° photon counting detector).

4.4.6 NanoDSF and Inhibitor Screening

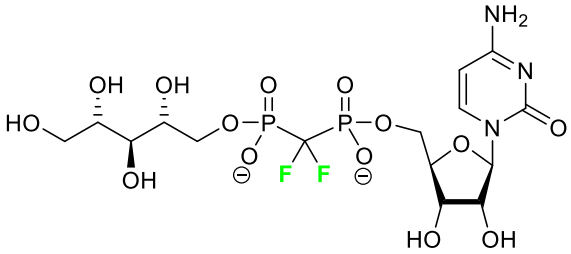
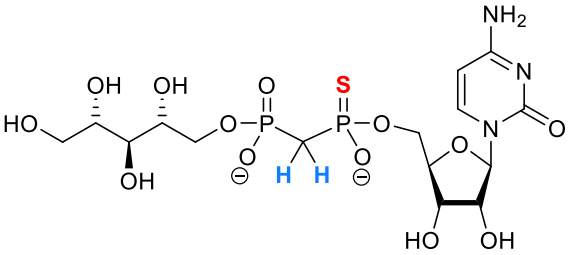
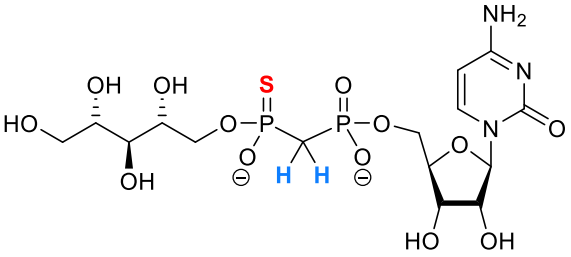
All analyses were performed on the Prometheus nanoDSF instrument (NanoTemper Technologies) and used the same method described in section **7.2.11**.

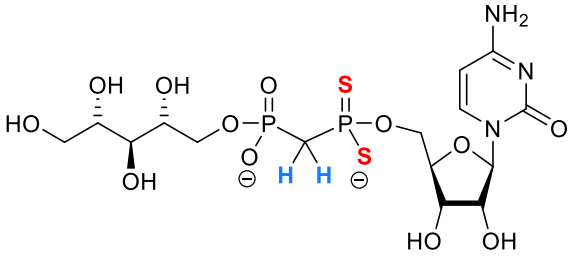
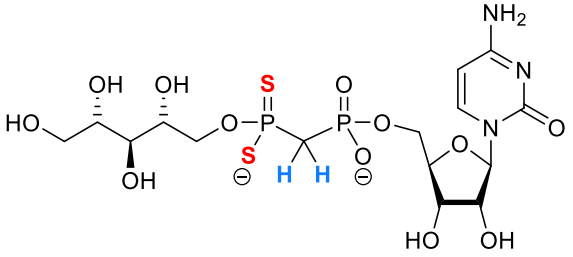
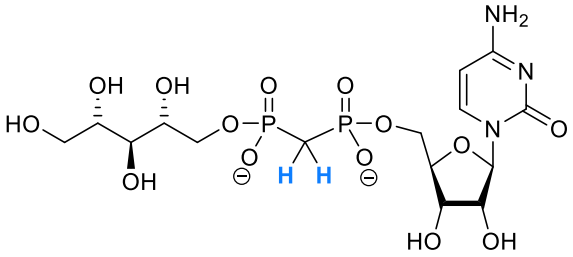
For the initial melting temperature curve of *Bs*TarK, pure protein from the second set of purification methods was used. *Bs*TarK-His was buffer exchanged into either His A buffer without imidazole or the Hep A buffer (**Table 4-3**, entries 1 and 4) and concentrated to ~1 mg/mL (see **7.2.7**). Samples from each solution were flash frozen in liquid nitrogen also. The melting temperature experiment proceeded as described in section **7.2.11**. The two mutants of *Bs*TarK-His, H119N and H255A, were also evaluated using this method in the same buffer His A buffer without imidazole and with or without 20% (v/v) glycerol.

The experiments assessing the effect of CDP-Rbo on the melting temperature of *Bs*TarK-His were again performed using the same method as described in **7.2.11**. Purified protein was first exchanged into a reaction buffer used in other *Bs*TarK studies (100 mM NaCl, 20 mM Tris pH 7.5, and 30 mM $MgCl_2$) (Brown *et al.*, 2010) and concentrated to ~1 mg/mL (see **7.2.7**). Three samples were taken and supplemented with CDP-Rbo at a working concentration of 1.5 mM. Three samples of apo protein were also taken. The melting temperature experiment proceeded as described in section **7.2.11**.

For the inhibitor screen, the same methods were used as described above. Each inhibitor (listed in **Table 4-4**) was added to the protein samples to a working concentration of 1.5 mM. The experiment was completed in triplicate. The melting temperature experiment proceeded as described in section **7.2.11**. The derived melting temperatures for each condition were plotted on a graph using GraphPad Prism.

Table 4-4: List of possible inhibitors for BsTarK/L.

Code	Structure
JG128	 Molecular Weight: 569.30
PB010	 Molecular Weight: 549.38
PB011	 Molecular Weight: 549.38

PB019	 Molecular Weight: 565.44
PB029	 Molecular Weight: 565.44
PB034	 Molecular Weight: 533.32

For the CDP-Rbo binding curve, serial 2-fold dilutions of a stock of CDP-Rbo into the reaction buffer (100 mM NaCl, 20 mM Tris pH 7.5, and 30 mM $MgCl_2$) were performed. In duplicate, each different concentration of CDP-Rbo was added to samples of pure *Bs*TarK-His at ~1 mg/ml. The working concentrations of CDP-Rbo ranged from 0-10 mM. The peak 330 nm fluorescence values from the discovery scan were normalised and transformed then plotted as a function of CDP-Rbo concentration using GraphPad Prism. This experiment was repeated using carbonic anhydrase as a negative control.

4.4.7 Activity Assays

30 µL reactions were prepared using the reaction buffer (100 mM NaCl, 20 mM Tris pH 7.5, and 30 mM $MgCl_2$ (Brown *et al.*, 2010)) which contained 20 µM of pure *Bs*TarK-His, 200 µM of CDP-Rbo and 200 µM of the acceptor substrate TV1041 from collaborators working in Leiden University (**Figure 4.34**). Heat treated *Bs*TarK-His was used as a negative control. The

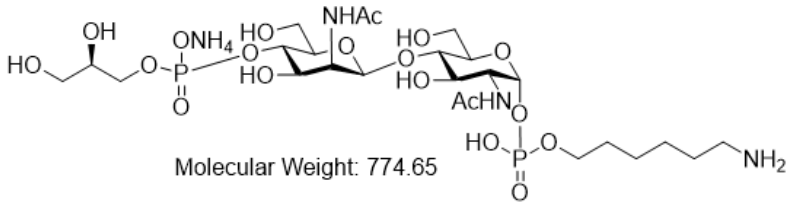


Figure 4.34: The acceptor substrate TV1041.

reactions were incubated for 30 mins at 21 °C then quenched with the addition of 30 µL of methanol. Denatured protein was pelleted by centrifugation at 10,000 xg for 10 mins and the supernatant was collected and freeze dried. The samples were then redissolved in 45 µL of 50% v/v acetonitrile and LC-MS grade water with 1% v/v formic acid. The samples were applied using an injection volume of 30 µL to a Accucore HILIC 2.1x50 mm, 2.6 µm silica particle size HPLC column and bound compounds were eluted using a gradient from 95% v/v to 5% v/v acetonitrile in water over 12 mins at a flow rate of 300 µL/min using the Thermo Scientific UltiMate 3000 HPLC System. The CT ultra ETD II ion trap was then used to identify eluted compounds of interest using MS.

Chapter 5: *Bs*TarL Wild Type and Mutant Protein Expression and Purification

5.1 Introduction

TarL is part of the TagF-like family of monotopic membrane proteins and is responsible for the synthesis of the poly-ribitol phosphate chain present in the wall teichoic acid (WTA) structure which is synthesised inside the cytoplasm of Gram-positive bacterial cells including *Bacillus subtilis* W23 and *Staphylococcus aureus.* TarL from *Bacillus subtilis* W23 (*Bs*TarL) is specifically known as a D-ribitol-5-phosphate (RboP) polymerase and acts on the TarK product lipid-P-P-GlcNAc-ManNAc-GroP-RboP to install many successive D-ribitol-5-phosphate (RboP) residues. Each of these is attached via its phosphate group onto the C1 hydroxyl position of the latest installed RboP using the donor substrate CDP-Ribitol (CDP-Rbo) (Brown *et al.*, 2010) (**Figure 5.1**). In *S. aureus* it has been proposed that TarL (*Sa*TarL) performs both the activities of TarK and TarL, therefore acting as a dual function enzyme that both primes and polymerises the RboP chain (Brown *et al.*, 2010). After the RboP polymer is synthesised (which consists of 20-40 RboP residues), the polymer is exported outside the cell membrane and is covalently attached to the peptidoglycan (PG) where the polymer extends far beyond the layers of the cell wall (Neuhaus and Baddiley, 2003).

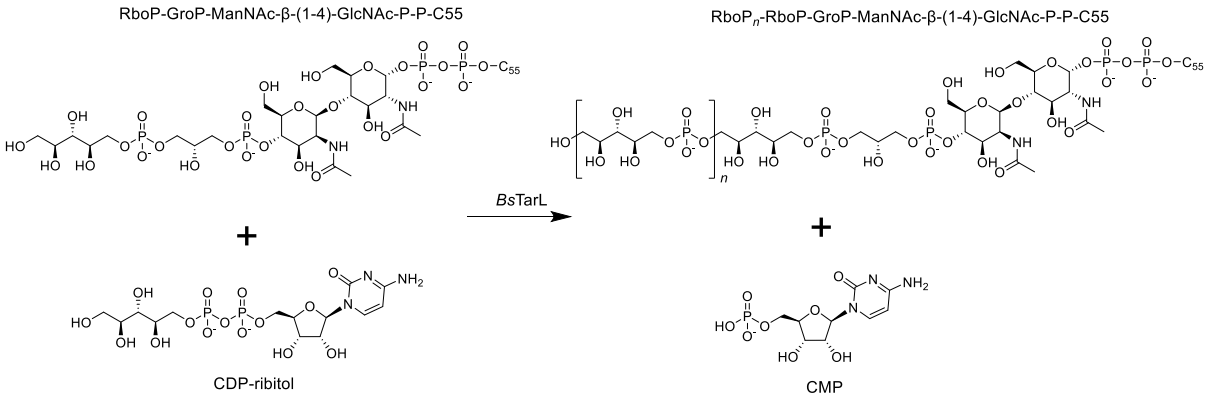


Figure 5.1: B. subtilis ribitol-phosphate polymerisation reaction facilitated by TarL. BsTarL acts on the TarK product, C_{55}-P-P-GlcNAc-ManNAc-GroP-RboP, to install RboP moieties successively until the polymer (n) is 20-40 residues long.

WTAs are known to be involved in many functions, namely cell shape determination, cell division, biofilm formation, cell adhesion and colonisation and antibiotic resistance (Matthias Gross *et al.*, 2001; Soldo, Lazarevic and Karamata, 2002; Weidenmaier *et al.*, 2004; Swoboda *et al.*, 2010; Farha *et al.*, 2013; Wu *et al.*, 2021). Hence, enzymes along the biosynthetic pathway of WTAs have been proposed as antivirulence targets. Some studies have found that mutant Methicillin-resistant *Staphylococcus aureus* (MRSA) strains lacking WTAs regain

sensitivity to β-lactam antibiotics (Brown *et al.*, 2012b). Additionally, research indicates that inhibiting enzymes involved in the biosynthesis and export of WTAs leads to the complete absence of WTAs (Hancock, Wiseman and Baddiley, 1976; Swoboda *et al.*, 2010; Farha *et al.*, 2013; Lee *et al.*, 2016). Furthermore, a combined strategy using broad-spectrum antibiotics along with inhibitors of WTA production has shown positive outcomes in animal models of MRSA infection and colonisation (Farha *et al.*, 2013; Lee *et al.*, 2016).

WTA polymerisation enzymes are targets for novel inhibitory compounds as knockout mutant strains of *B. subtilis* W23 in which the *tarK* and *tarL* genes are removed result in cell death (Brown *et al.*, 2010). It is therefore important that further research into these biosynthetic enzymes is performed to deepen our understanding of WTA synthesis. TarL is of particular interest as the dual function enzyme in *S. aureus* could be an effective target for new drug studies aimed at combating the increasing resistance to antibiotics in pathogens such as MRSA. No inhibitors of TarL have been identified. Thus, a comprehensive approach that includes the development of chemical probes or inhibitors, along with structural and enzymological studies, is essential for gaining a deeper understanding of the process catalysed by this enzyme and advancing structure-based drug design, and could enable a new frontier in fighting other resistant pathogenic species.

5.1.1 Aims

To facilitate downstream structural and enzymological studies of TarL, protocols need to be developed which allow the production and purification of the recombinant protein and any desired active site mutants. Despite previous papers describing the successful expression of TarL, these papers do not show the expression and purification results (Brown *et al.*, 2010). Because of this, the purity of the proteins produced using the described methods is unknown. To ensure high purity of TarL for structural studies, muti-step fast protein liquid chromatography (FPLC) methods will be used instead of the nickel bead settle bed method presented in previous research (Brown *et al.*, 2010).

In this chapter, the previous expression trials of *Bs*TarL will be expanded upon, with the aid of the successful expression conditions identified for *Bs*TarK production in chapter 4, to elucidate the optimal conditions to produce pure and soluble TarL. This would then permit downstream structural and characterisation experiments of the enzyme.

Initially, recombinant gene products with different tags varying in terminal location will be synthesised for more expansive protein expression trials. The ideal conditions and tag type and location which permit the greatest yield of solubilised protein will be identified, therefore allowing larger scale protein expression and purification using immobilized metal affinity

chromatography (IMAC) and size exclusion chromatography (SEC) techniques. Once highly pure *Bs*TarL is produced, various methods of enzyme characterisation will be used, including nanoDSF and circular dichroism, to assess protein stability and correct folding. Active site mutants will also be synthesised and assessed in a similar manner. Mass photometry will be utilised to examine the possibility of *Bs*TarL multimerisation. The purified enzyme will also undergo protein crystallisation trials and/or, if the multimeric protein forms, cryogenic electron microscopy (cryo-EM) to solve the protein structure in the presence or absence of the natural ligands or possible inhibitors.

5.2 Results and Discussion

5.2.1 Molecular Cloning

To express and purify soluble *Bs*TarL, a few different gene constructs were designed and synthesised using the In-Fusion cloning method. The *Bs*TarL gene was first inserted into Champion™ pET SUMO (shortened to pSUMO) which has an N-terminal His6 tag followed by a Small Ubiquitin-like Modifier (SUMO) solubility tag. This tag is intended to increase the solubility of proteins and also contains a cleavage site immediately downstream of the SUMO tag, leaving no extra amino acids at the N-terminal of the target protein upon cleavage (Butt *et al.*, 2005).

The Champion™ pET SUMO vector was first linearised at the desired insertion site using designed forward and reverse primers (**5.2.1**) and commercial PCR reagents (**7.2.1**). The results from this linearisation are show in section **4.2.1, Figure 4.2**.

The corresponding insert with homologous overlaps with the linearised pSUMO plasmid was synthesised by cloning the *Bs*TarL gene from the ordered plasmid pUC57*Bs*TarL using forward and reverse primers and commercial PCR reagents (see **5.2.1** and **7.2.1**). Completed PCR reactions were analysed using DNA gel electrophoresis (**7.2.2**) (**Figure 5.2**).

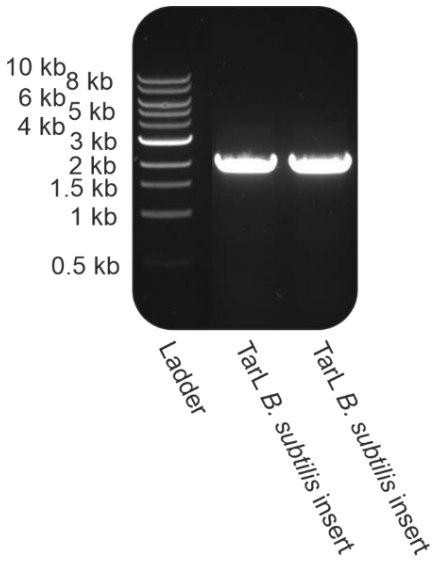


Figure 5.2: BsTarL insert for linear pSUMO resolved on a 0.8% agarose gel. Expected band is 1.89 kb.

Successful synthesis of the *Bs*TarL insert is represented by the single intense band observed at ~1.9 kb reminiscent of the correct insert length. The inserts located at the correct length were gel extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C.

To complete ligation, the In-Fusion kit (Takara Bio) was used and completed reactions were transformed into BL21 *E. coli* cells. Successful colonies were isolated, and colony PCR was used to confirm the presence of the ligated plasmid with the *Bs*TarL insert (**7.2.3**) (**Figure 5.3**).

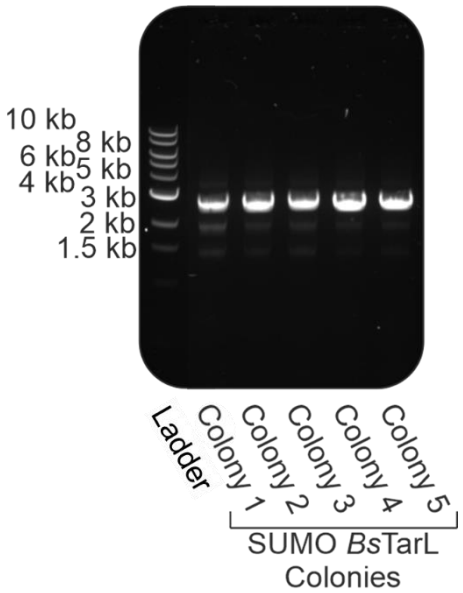


Figure 5.3: Colony PCR products of pSUMOBsTarL resolved on a 0.8% agarose gel. Correct length expected to be 2.456kb.

The single intense band at ~2.5 kb across all the colonies tested is indicative of the presence of the correctly engineered plasmid (2.456 kb). However, there are two other faint bands present in all the colonies tested at ~1.5 kb and ~2 kb. These bands may be primer dimers or degradation products of the engineered plasmid. The same colonies were cultured and purified using a Miniprep DNA purification kit and the pSUMO*Bs*TarL plasmids were stored at -20 °C.

C-terminally His6 tagged *Bs*TarL constructs were also designed for In-Fusion cloning. First, the pET-28a(+) vector was linearised at the desired insertion site, to allow C-terminal His6 tagging, using commercial PCR reagents and designed primers (**5.2.1**). Analysis of the correct linearised vector was performed and the results are shown in section **4.2.1**, **Figure 4.5.**

The corresponding *Bs*TarL insert was synthesised by PCR using the ordered plasmid pHis-*Bs*TarL28a and designed primers complementary to the insertion site of the linearised vector **(5.4.1**). Correct synthesis of the *Bs*TarL insert was analysed by DNA gel electrophoresis (**Figure 5.4**).

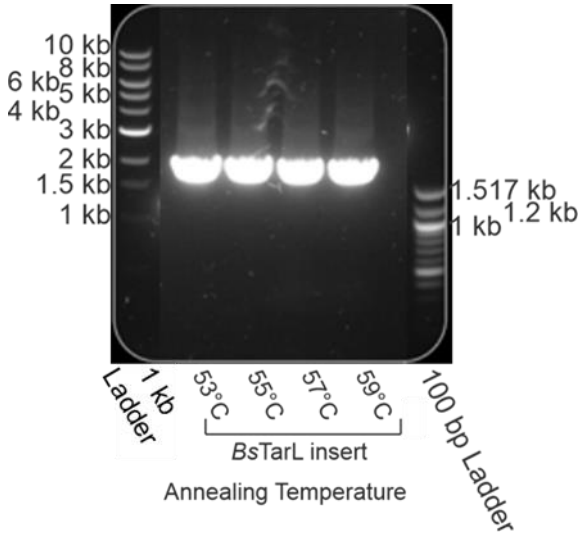


Figure 5.4: BsTarL insert for linear pET-28a(+) resolved on a 0.8% agarose gel. Expected band is ~1.9 kb.

The presence of an intense band at around 1.9 kb indicates correct synthesis of the *Bs*TarL insert at every annealing temperature tested. The inserts located at the correct length were gel extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C.

To complete ligation, the In-Fusion kit (Takara Bio) was used and completed reactions were transformed into *E. coli* cells. Successful colonies were cultured and miniprepped and isolated plasmid DNA was sent for Sanger DNA sequencing (Eurofins Genomics) which confirmed correct ligation. Recombinant plasmids were then stored at -20 °C.

5.2.2 Expression Trials

The expression of the engineered N-terminally tagged His-SUMO-*Bs*TarL protein was first assessed. The pSUMO*Bs*TarL plasmid was transformed into BL21 (DE3) *E. coli* and successful colonies were cultured for expression. Protein overexpression proceeded with the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (see **5.4.2**). The conditions used for the first trial were similar to the optimised His-SUMO-*Bs*TarK expression conditions (**4.4.2**) because of the successful production of soluble protein observed in that case. Resuspended uninduced and induced cell pellets were lysed using sonication in the lysis buffer reported by Brown *et al.*, 2010, which contained CHAPS and glycerol to aid in solubilisation of the target protein (**5.4.2**). Samples were prepared for SDS-PAGE analysis (**7.2.8**) to assess protein expression (**Figure 5.5**).

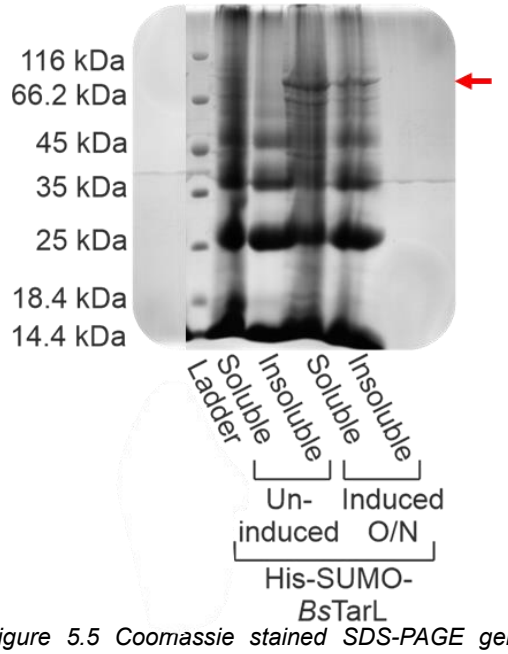


Figure 5.5 Coomassie stained SDS-PAGE gels from His-SUMO-BsTarL expression trial with samples taken after overnight (O/N) expression. Bands which are likely the target protein are labelled with red arrows at the expected mass of 86.0 kDa. Resolved on a 4-15% polyacrylamide gel.

The faint bands located in both induced lanes above the 66.2 kDa mark may represent successful expression of the target protein (His-SUMO-*Bs*TarK has an expected mass of ~86 kDa) as these bands are not seen in the uninduced lanes. However, these bands are considerably less intense than the other bands observed in all lanes, presumably *E. coli* proteins, at ~25, ~35 and ~45 kDa suggesting that the overexpression of the target protein is not substantial compared to native *E. coli* proteins.

As the SUMO tagged constructs seemed to exhibit relatively low levels of expression, the location of the tag and the cell line used for protein expression were changed with the aim to yield higher expression levels. As mentioned in previous chapters, TarL, TarK and other TagF-like proteins have been C-terminally tagged in the literature (Brown, Zhang and Walker, 2008; Brown *et al.*, 2010; Lovering *et al.*, 2010; Li *et al.*, 2024) which suggests that the location of the tag is vital to soluble expression. Additionally, the CHAPS concentration in the lysis buffer differs greatly between different studies. For *Bs*TarL expression, Brown *et al.*, 2010 uses 0.6% w/v (~9.8 mM) CHAPS in their lysis buffer which is around the critical micelle concentration (CMC) of this detergent, while *Sa*TarL expression and lysis by Li *et al.*, 2024, uses a higher CHAPS concentration of 16 mM. The highest concentration of CHAPS used in a lysis buffer was reported for the expression and purification of TagF by Lovering *et al.*, 2010, in which case the concentration is 4-fold higher than the CMC of CHAPS at 40 mM. Increasing the concentration of CHAPS in the lysis buffer may yield a higher amount of solubilised *Bs*TarL but it may also result in smeary gels as seen previously (**Figure 4.10**). Finally, another

observation from the literature is that *Bs*TarL is expressed in the Rosetta2(DE3) pLysS strain of *E. coli* (Brown *et al.*, 2010). Therefore, it is worth evaluating the expression of *Bs*TarL in strains other than BL21(DE3). The reason for the use of Rosetta2(DE3) pLysS strain of *E. coli* is not known. Rosetta2 strains contain more rare codons which are used for eukaryotic protein expression, but *Bs*TarL is a prokaryotic protein. Additionally, they may better facilitate Gram-positive protein expression in Gram-negative systems as there are codons which are rarer in *E. coli,* however, the gene sequences used in this work have been codon optimised for expression in *E. coli.* The pLysS aspect of the Rosetta2(DE3) pLysS strain encodes and expresses T7 lysozyme, which suppresses basal expression of T7 RNA polymerase prior to induction.

It was decided to return to the N-terminally His6-tagged protein trialled previously (**3.2.3**) alongside the newly synthesised C-terminally His6-tagged protein. The conditions used for bacterial growth and expression remained the same. These conditions were: 2xYT media, culture until 0.5-0.8 OD_{600}, then induce protein expression with the addition of 1mM IPTG and leave overnight (O/N) at 16°C. Both the N- and C-terminally His6 tagged *Bs*TarL constructs were transformed into Rosetta2(DE3) pLysS and expression of *Bs*TarL proceeded using the previous conditions (**5.4.2**). Uninduced and induced samples were taken and lysed using sonication in the 'high' CHAPS buffer from Lovering *et al.*, 2010 (see **5.4.2**). Samples of soluble and insoluble fractions were prepared for SDS-PAGE and Western blotting analysis (**Figure 5.6**).

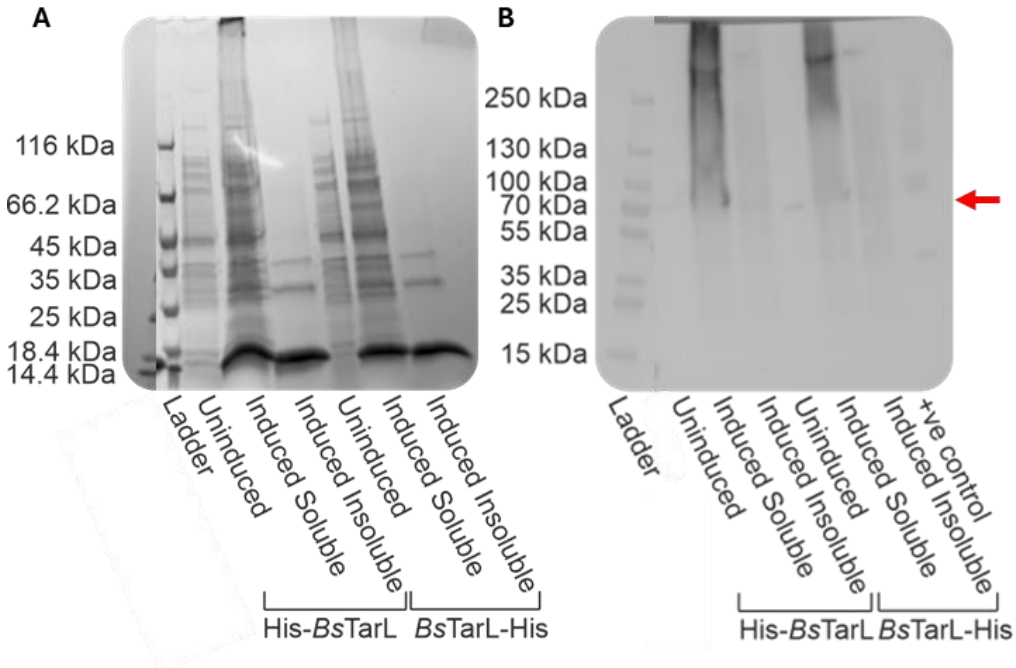


Figure 5.6: SDS-PAGE and Western blot analysis of the expression of N- and C-terminally His6 tagged BsTarL. A = Coomassie stained SDS-PAGE gel of BsTarL expression. B = An anti-His6 tag western blot of the same expression trial. Red arrows indicate possible presence of the target protein (~74 kDa). Resolved on a 4-15% polyacrylamide gel.

The expression of both tagged proteins is unclear in both the coomassie stained gel and the Western blot. In the coomassie stained gel there are multiple bands around the expected molecular weight (expected mass of ~74 kDa) in both the uninduced and induced soluble lanes suggesting that the target protein has not been expressed at detectable levels. In the Western blot faint bands can be observed in the induced soluble lanes at around ~74 kDa suggesting that the protein has been expressed. Furthermore, other bands are observed above 250 kDa which suggests that the target protein may be aggregating. There is also substantial smearing in the Western blot which is likely due to non-specific binding of the anti-His6 tag antibody. The reason for this may be due to the high amount of detergent present in the loaded samples or due to a high antibody concentration used. The higher molecular weight bands observed along with the smears could be explained by the presence of CHAPS at a concentration far exceeding the CMC which causes the soluble protein to aggregate.

Based on the results of the previous trial (**Figure 5.6**), the 'low' CHAPS buffer (**Figure 5.5**Figure 5.5 Coomassie stained SDS-PAGE gels from His-SUMO-BsTarL expression trial with samples taken after overnight (O/N) expression. Bands which are likely the target protein are labelled with red arrows at the expected mass of 86.0 kDa.) was selected for further expression trials. In the next experiment, the N-terminally His6-tagged *Bs*TarL protein was expressed in two different *E. coli* strains: Rosetta2(DE3) pLysS (Novagen) and SHuffle lysY (New England Biolabs) (**5.4.2**). Protein expression conditions were kept constant, with samples collected both before and after induction. The samples were then lysed by sonication in the 'low' CHAPS buffer (**5.4.2**) and prepared for SDS-PAGE and Western Blot analysis (**7.2.8** and **7.2.9**) (**Figure 5.7**).

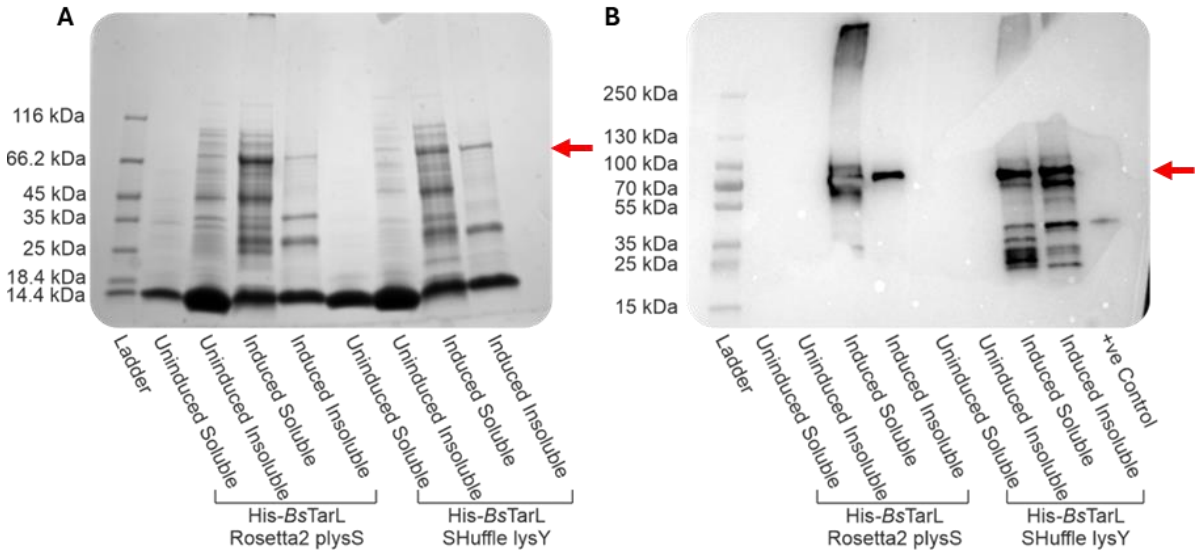


Figure 5.7: SDS-PAGE and Western blot analysis of the expression of N-terminally His6 tagged BsTarL in different strains. A = Coomassie stained SDS-PAGE gel of N-terminally His6 tagged BsTarL expression. Bands containing the target protein are labelled with red arrows. B = An anti-His6 tag western blot of the same expression trial. Resolved on a 4-15% polyacrylamide gel.

Soluble expression of His-*Bs*TarL is evident on the coomassie stained SDS-PAGE gel, with a strong band present in the induced soluble lane, but not present in the uninduced lanes, at a position above the 66.2 kDa marker (expected mass ~74 kDa) for both bacterial strains. Successful target protein expression is further confirmed by distinct, highly intense bands on the anti-His Western blot in the same lanes, positioned between the 70 kDa and 100 kDa markers. The Western blot also shows bands of similar intensity in the induced insoluble lanes of both strains, suggesting that about half of the protein is soluble. Conversely, the band in the induced soluble lane of the Rosetta2 strain seems to be obscured by a 'burn' mark which can be caused by a signal too intense. This would indicate that the soluble expression of the target protein is higher in the soluble fraction of Rosetta2 plysS strain. In the Rosetta2 plysS coomassie gel, this soluble band also appears more intense than the insoluble further supporting this conclusion. Other bands are also observed in the Western blot that are of a lower molecular mass than the target protein only in the induced lanes of SHuffle cells. This suggests that while His-*Bs*TarL is present, some of the protein produced is degraded in SHuffle *E. coli*. Additionally, in the induced soluble lane of Rosetta2 cells, there is an intense band observed above the 250 kDa marker which suggests that some of the protein produced is aggregating. Nonetheless, combining these strains with the 'low' CHAPS buffer significantly enhances soluble target protein expression, with the Rosetta2 strain appearing superior in terms of the level of expressed protein and lack of degradation.

Next, the C-terminally His6 tagged *Bs*TarL protein was expressed under the same conditions (see **5.4.2**). SDS-PAGE and Western blot analysis were used to evaluate expression (**Figure 5.8**).

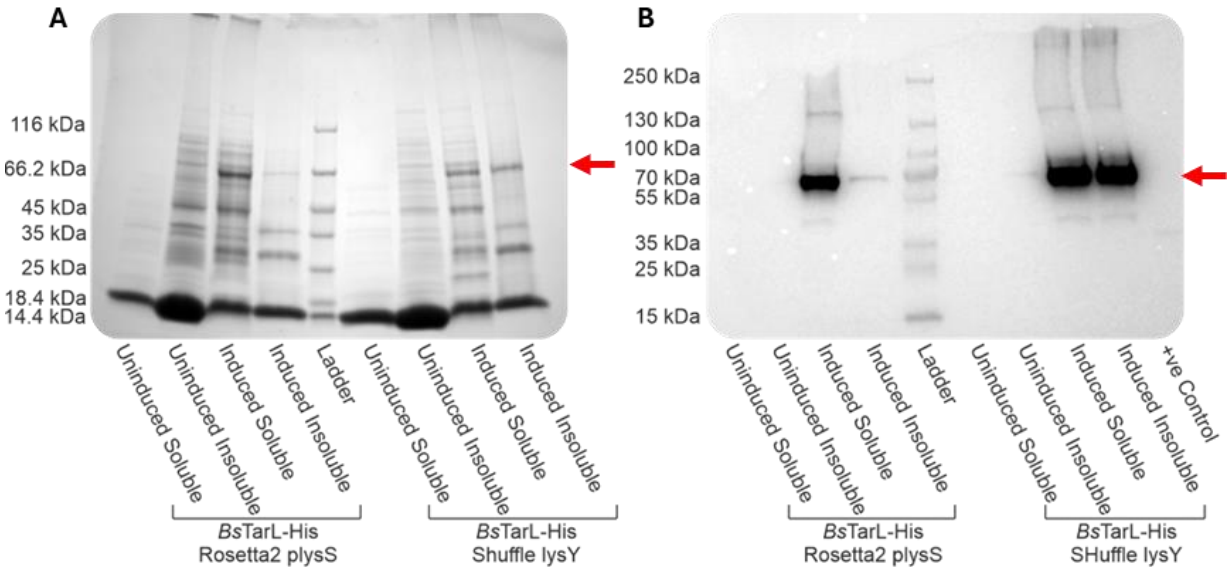


Figure 5.8: SDS-PAGE and Western blot analysis of the expression of C-terminally His6 tagged BsTarL in different strains. A = Coomassie stained SDS-PAGE gel of C-terminally His6 tagged BsTarL expression. Bands containing the target protein are labelled with red arrows. B = An anti-His6 tag western blot of the same expression trial. Resolved on a 4-15% polyacrylamide gel.

Soluble expression of *Bs*TarL-His is observed in the coomassie stained SDS-PAGE gel as an intense band is present in the induced soluble lane, but not in the uninduced lanes, for both strains of bacteria at around 70 kDa (expected mass ~74 kDa). This is also supported by the clear and highly intense bands observed in the Western blot in the induced soluble lanes located at the 70 kDa mark. In Shuffle cells, a prominent band appears in the insoluble protein fraction following induction. This band has the same intensity as the corresponding band in the soluble protein fraction on Western blot, indicating that a substantial portion of the target protein is insoluble. In contrast, in the Rosetta strain, the amount of insoluble protein is significantly lower than that of the soluble protein. This suggests that the majority of *Bs*TarL-His is soluble in the Rosetta2 plysS strain, while in SHuffle lysY the soluble to insoluble protein ratio is about 50:50. Additionally, the Western blot reveals another band of low intensity just above the 130 kDa mark. This could correspond to a dimer of *Bs*TarL-His with an expected mass of ~148 kDa. Multimeric TarL, specifically tetrameric and octameric configurations, have been reported in the literature, most notably in the cryo-EM structure of *Sa*TarL (Li *et al.*, 2024). Therefore, investigations into the multimeric state of *Bs*TarL should be undertaken to assess the possibility of cryo-EM structure elucidation.

To conclude, using the 'low' CHAPS buffer (Brown *et al.*, 2010) appears to improve the solubility of *Bs*TarL significantly compared to the 'high' CHAPS buffer (Lovering *et al.*, 2010) when expressed in the Rosetta2(DE3) pLysS (Novagen) and SHuffle lysY (New England Biolabs) strains. This buffer also reduces the smearing observed in the Western blots. Between the two strains and the two tag configurations used, it appears that *Bs*TarL-His in the Rosetta2 pLysS strain yields the highest amount of soluble protein with very low levels of insoluble protein. Therefore, the C-terminally tagged *Bs*TarL construct expressed in Rosetta2(DE3) pLysS was used for larger scale expression and protein purification.

5.2.3 Protein Purification

Larger scale bacterial cultures were used to synthesise *Bs*TarL-His using the optimised conditions mentioned previously (**5.4.2**). The first step of *Bs*TarL purification, outlined in **Table 5-2**, utilised immobilized metal affinity chromatography (IMAC) to isolate the target protein from the lysate. Samples of the total lysate, flow through and wash fractions were taken, and after purification finished, samples from the eluted fractions, which may have contained the target protein by observation of 280 nm absorption peaks, were taken. All the samples were prepared for SDS-PAGE analysis (**Figure 5.9**).

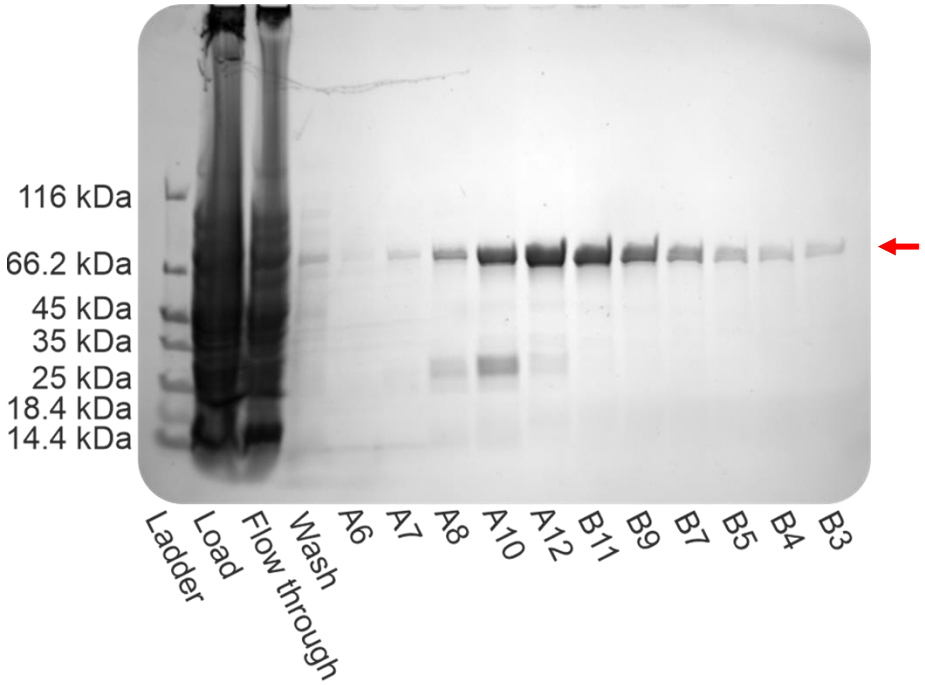


Figure 5.9: IMAC purification of BsTarL-His from E. coli lysate. Target protein has an expected mass of ~74 kDa. Resolved on a 4-15% polyacrylamide gel.

Intense bands at around the expected mass of *Bs*TarL-His (~74 kDa) can be observed in fractions A7-B3. This indicates successful capture and elution of the His6 tagged protein from the cell lysate. However, within some of the fractions which contain the target protein, namely A8-A12, an impurity of ~25 kDa can be observed, indicating that further purification steps may be required. Fractions B11-B3 appear to be very pure as only the target band can be observed. These fractions could be isolated and used in structural and characterisation studies (**5.2.4** and **5.2.5**).

In the literature, the second step in the purification of *Sa*TarL uses the Superose 6 Increase size exclusion chromatography (SEC) column (Cytiva) (Li *et al.*, 2024). The authors found using cryo-EM that TarL forms tetramers which are embedded in the top and bottom of a CHAPS micelle disc. *Bs*TarL has a sequence identity to *Sa*TarL of 39.08%. Therefore, the same oligomeric configuration may be observed in *Bs*TarL. If this 2x tetrameric arrangement is present, the mass of the multimeric *Bs*TarL protein would total ~592 kDa plus the mass of the CHAPS micelle disc. This suggests that the Superose 6 Increase column, which has an upper limit of 5 MDa, should be able to separate the various oligomeric states which may be formed. This oligomeric state is also of sufficient molecular weight to allow structural determination by cryo-EM (the smallest protein structures resolved at <4 Å (atomic resolution) rarely are sub-100 kDa (Merk *et al.*, 2016; Herzik, Wu and Lander, 2019)).

`The fractions containing the target protein from the previous purification step were combined and buffer exchanged into size exclusion chromatography (SEC) buffer for SEC using the Superose 6 Increase column (Cytiva) (see **5.4.4**). After the purification was completed, four large peaks were identified in the 280 nm absorption spectrum (named P1: A8-B7, P2: B6-B4, P3: B3-B1 and P4: C1-C4) and eluted fractions under these peaks were sampled. The sampled fractions and a sample taken from the loaded mixture were prepared for SDS-PAGE analysis (**Figure 5.10**).

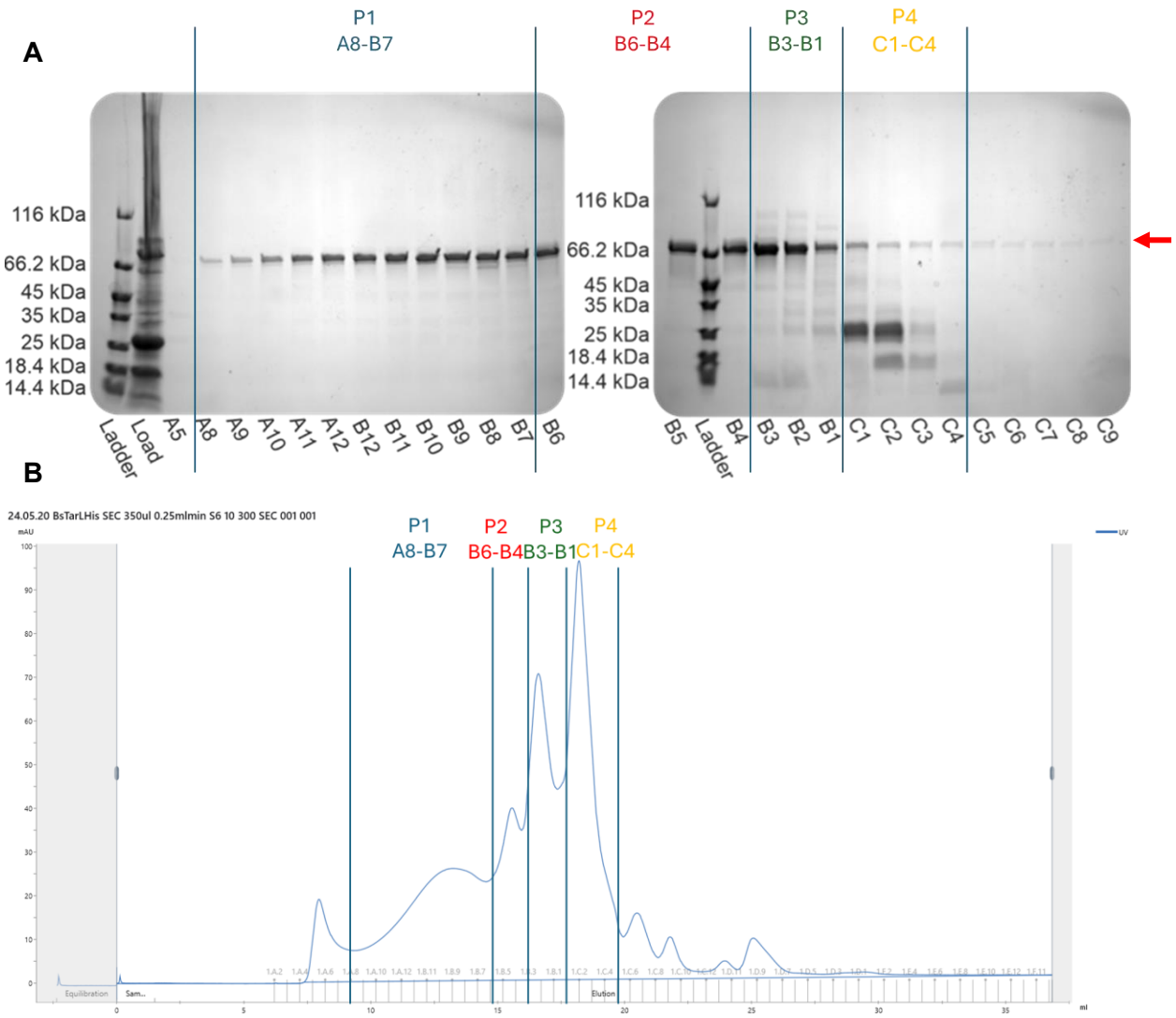


Figure 5.10: SEC purification of BsTarL-His. Target protein with an expected mass of ~74 kDa is seen across many fractions in the SDS-PAGE gel (A). The fractions that are observed under the four prominent peaks on the UV chromatogram (B) are shown and labelled as P1, P2, P3 and P4. Resolved on a 4-15% polyacrylamide gel.

From the gels, it is clear that the band observed across fractions A8-C4 at ~74 kDa represents the target protein *Bs*TarL-His. P1 appears to contain very pure protein as only a single intense band is observed across the labelled fractions; this is also true for P2. P3 contains some impurities represented by faint bands at ~14 kDa, ~25 kDa, 45 kDa and above ~66 kDa, however the main constituent of these fractions is the target protein shown by the intense bands present at ~74 kDa. P4 also appears to contain the target protein as faint bands are

observed at ~74 kDa, however there are much more intense bands observed at ~25 kDa and 18 kDa suggesting that the majority of this peak contains impurities. These results suggest that *Bs*TarL-His is forming a range of multimers across fractions A8-B1 as multiple peaks which contain the target protein are observed at different retention volumes. For example, P3 may contain monomeric protein, P2 may contain Trimeric or tetrameric protein and P1 may contain octameric protein. As mentioned previously, tetrameric *Sa*TarL has been observed in the literature (Li *et al.*, 2024). Here, the tetramer was found to be embedded into a CHAPS micelle disc with an additional tetramer on top of the disc with D4 symmetry, resulting in a octameric protein (Li *et al.*, 2024). It seems likely that P1 contains octameric *Bs*TarL with P2 and P3 containing the broken-down products of the multimer (trimers, dimers, monomers). To explore the multimeric nature of *Bs*TarL, mass photometry or size exclusion chromatography with multi-angle static light scattering (SEC-MALS) should be attempted.

5.2.4 Circular Dichroism

To provide evidence that the protein is folded with tertiary structures which unfolds expectedly with an increase in temperature, circular dichroism (CD) was performed. Pure *Bs*TarL-His from the IMAC purification step was prepared in SEC buffer with and without 10% v/v glycerol (**7.2.7**). The experiment then proceeded as outlined in section **7.2.10**. The CD data were plotted using GraphPad Prism (**Figure 5.11**).

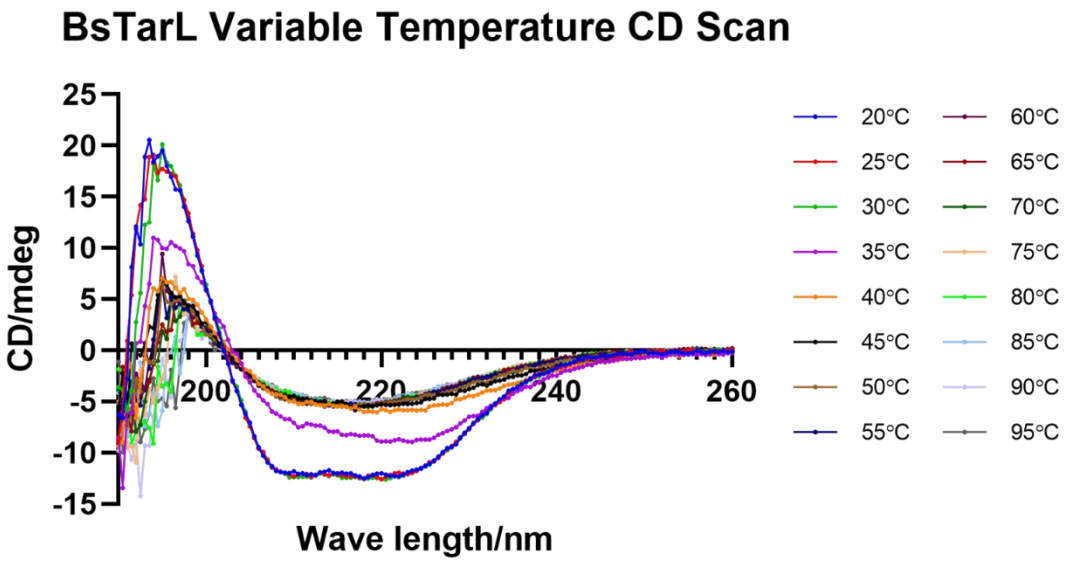


Figure 5.11: CD spectra of BsTarL at various temperatures.

The curve generated by the variable temperature CD experiment showed negative values at 208 and 222 nm, characteristic of the spectra of α-helices, and positive values at 193 and 195 nm characteristic of α-helices and β-sheets, respectively. This suggests that the protein is

folded. As the temperature increases these peaks approach 0 mdeg indicating that the secondary structures in the protein are unfolding.

Using the CD data collected at 218 and 222 nm for both the glycerol and glycerol-free buffers, a melting temperature plot was derived (**Figure 5.12**) A typical sigmoidal function is produced in both runs which suggests that the protein is folded and unfolds expectedly when temperature is increased. The resulting curve allows the melting temperature to be calculated. The melting temperatures of the protein were found to be on average 35 °C and 41°C for the glycerol-free and 10% v/v glycerol buffers, respectively. This suggests that the addition of glycerol increases the stability of the target protein similar to what is reported in the literature with *Bs*TarL and other homologues (Brown *et al.*, 2010; Lovering *et al.*, 2010). Further CD studies of *Bs*TarL in different buffers and with different ligands should be used in the future to determine the ideal conditions in which the protein is most stable in order to increase the likelihood of protein crystallisation.

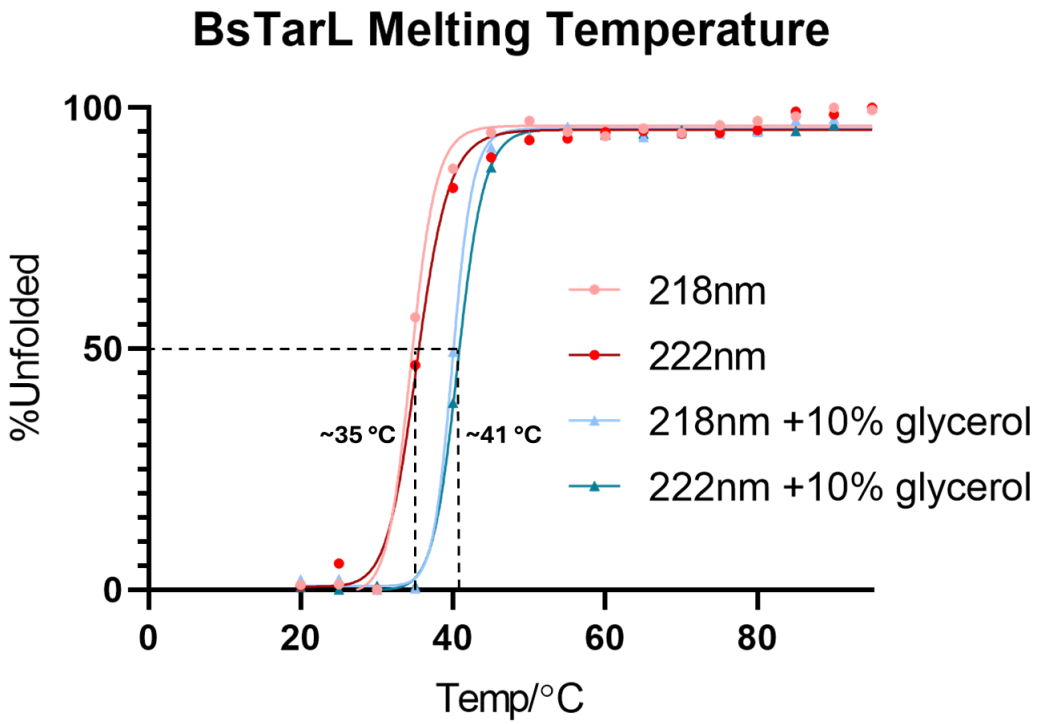


Figure 5.12: Thermal melting temperature curves of BsTarL using the CD results at 218 nm and 222 nm. Melting temperature was calculated to be 35 °C and 41°C for the glycerol-free and 10% v/v glycerol buffers, respectively.

5.2.5 Crystallography

Purified *Bs*TarL, from the same protein preparation and in the same buffer used for the CD experiments, with 10% v/v glycerol, was concentrated to 12 mg/mL for protein crystallisation trials. Initial crystal hits were identified in the Crystal Screen HT (well F10) and PEG Ion HT screens (wells F1-2, 7-8, 10 and A3). The conditions were further optimised as described in **5.4.5** and the protein concentration was increased to 15 mg/ml. Crystals were observed in

three wells: A3 (**Figure 5.13**), B3 and C3 which contain a 0.1M MES buffer, pH 6.5 with either 6, 8 or 10% v/v PEG 20000, respectively. These crystals were sent to the Diamond Light Source for X-ray diffraction experiments.

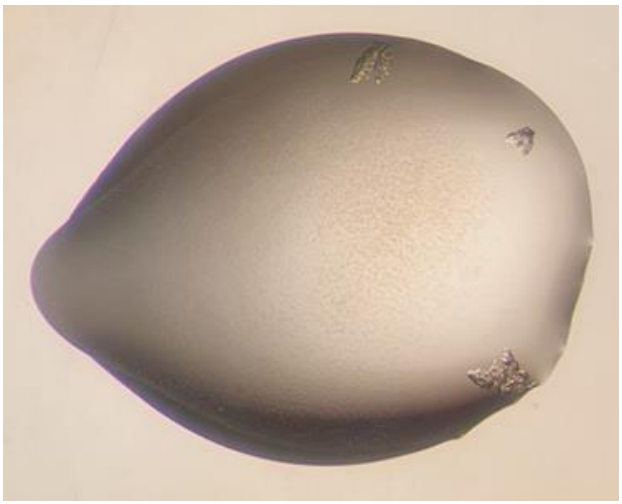


Figure 5.13: Microscopic photograph of possible BsTarL-His crystals in the Crystal Screen HT screen optimisation trial, well A3.

Unfortunately, the crystals from this trial did not yield obvious and characteristic protein diffraction patterns and Diamond's auto processing software failed. This suggests that these crystals are not protein. As mentioned in the previous chapter, the reason why *Bs*TarL is difficult to crystallise may be explained by the membrane binding domain present in the protein and other TagF-like proteins. Membrane associated proteins, like these enzymes, are infamously hard to crystallise. However, evidence from the literature and the SEC purification results suggest that *Bs*TarL may form a multimer like *Sa*TarL (Li *et al.*, 2024). If this is the case, then cryo-EM is a feasible technique to elucidate protein structure as even the formation of a dimer (~148 kDa) exceeds the lower limit (100 kDa) of the equipment, with the possibility of an even larger octamer (~592 kDa) to form.

5.2.6 Nano Differential Scanning Fluorimetry

To assess the thermal stability of *Bs*TarL and mutated variants, nano differential scanning fluorimetry (nanoDSF) was used instead of CD as the different proteins can be screened in parallel to save time and the buffer composition is not a concern for the wavelengths used (for example Cl^- can absorb at the wavelengths used for CD measurements therefore Cl^- concentration has to remain low). *Bs*TarL and two active site mutants, H345N and H477A, were purified. The mutants were synthesised using single point mutagenesis using an approach similar to that used in the previous chapter (**5.4.3**). These mutants were selected as they are the residues which align to the catalytic residues identified in *Se*TagF, *Sa*TarL and Bcs3 (Lovering *et al.*, 2010; Cifuente *et al.*, 2023; Li *et al.*, 2024). The H345N mutation, which

natively (H345) may have a role in activating the nucleophile, should render the enzyme non-functional which is useful in activity assays (Lovering *et al.*, 2010; Li *et al.*, 2024). The H477A mutant, which natively (H477) may have a role in protonating the leaving group, should make the enzyme non-functional and also has the ability to trap CDP-Rbo inside the active site which is useful for structural studies, as seen previously in other TagF-like protein structures (Lovering *et al.*, 2010; Cifuente *et al.*, 2023; Li *et al.*, 2024). The proteins were prepared for nano differential scanning fluorimetry (nanoDSF) experiments outlined in section **5.4.6** to assess the effects the mutations had on the melting temperature (Tm), and therefore stability, of the protein (**Figure 5.14**).

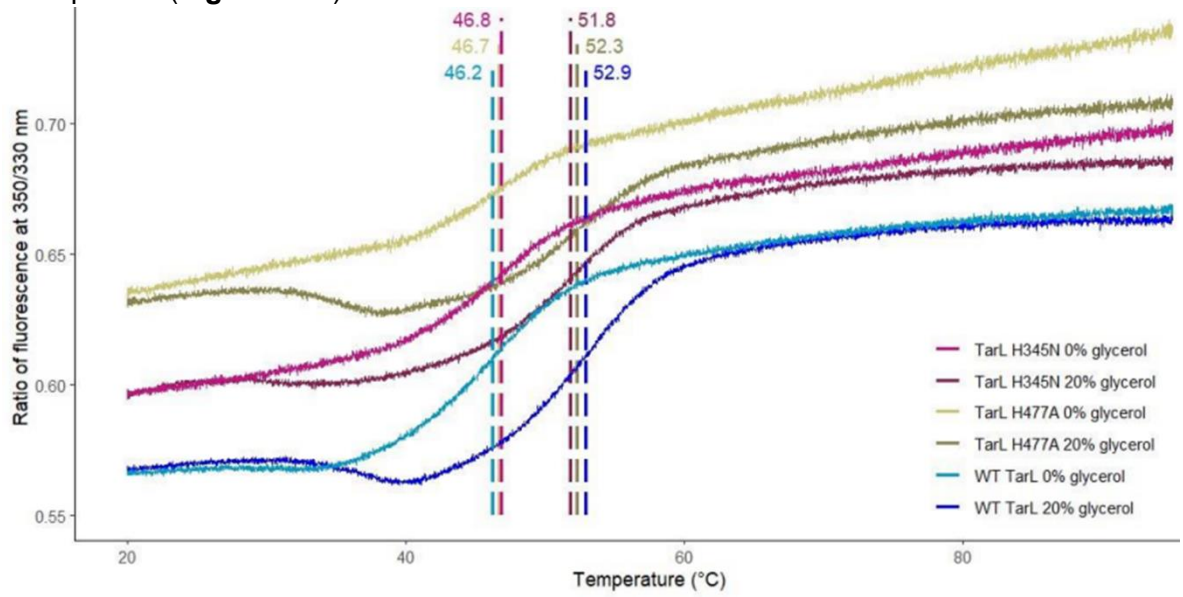


Figure 5.14: Intrinsic tryptophan fluorescence ratio at 350/330 nm of WT TarL and the TarL H345N and TarL H477A variants. Dashed lines show the melting points for each sample, at which 50% is unfolded.

All proteins tested exhibit the typical sigmoidal curve expected from the experiment and therefore Tm values can be calculated. The Tm values of the mutants in the 0% v/v glycerol buffer (46.2 °C and 46.8 °C for H345N and H477A, respectively) are very similar to the WT enzyme (46.7 °C) which suggests that the mutations do not affect protein stability greatly. The same observation can be made for the experiments involving the addition of 20% v/v glycerol. Here the Tm values were 52.9 °C, 51.8 °C and 52.3 °C for the WT, H345N and H477A enzymes, respectively. There is however a large shift in Tm values of all the enzymes between the glycerol-free and the 20% v/v glycerol buffers, with the TarL WT Tm value increasing by 6.7 °C with the addition of glycerol. This indicates that glycerol does have a stabilising effect on WT TarL and the mutant enzymes. The melting temperature values of the WT enzymes are much higher than the previous CD values measured. This is likely due to the buffer differences between the two experiments. As Cl^- ions absorb at some of the wavelengths used in CD, the buffer did not contain NaCl. This likely decreases solubility of the protein and decreases the melting temperature value. With nanoDSF, NaCl concentration is not an issue,

therefore this buffer contained 500 mM NaCl which likely aided in increasing the stability and solubility of the target protein.

5.2.7 Mass Photometry

The fractions from the separated peaks of the SEC purification (P1, P2 and P3) were combined and prepared for mass photometry experiments as described in **5.4.7**. The initial measurements were taken at a working enzyme concentration of 25 nM. Histograms were plotted using the Refeyn DiscoverMP software (**Figure 5.15**).

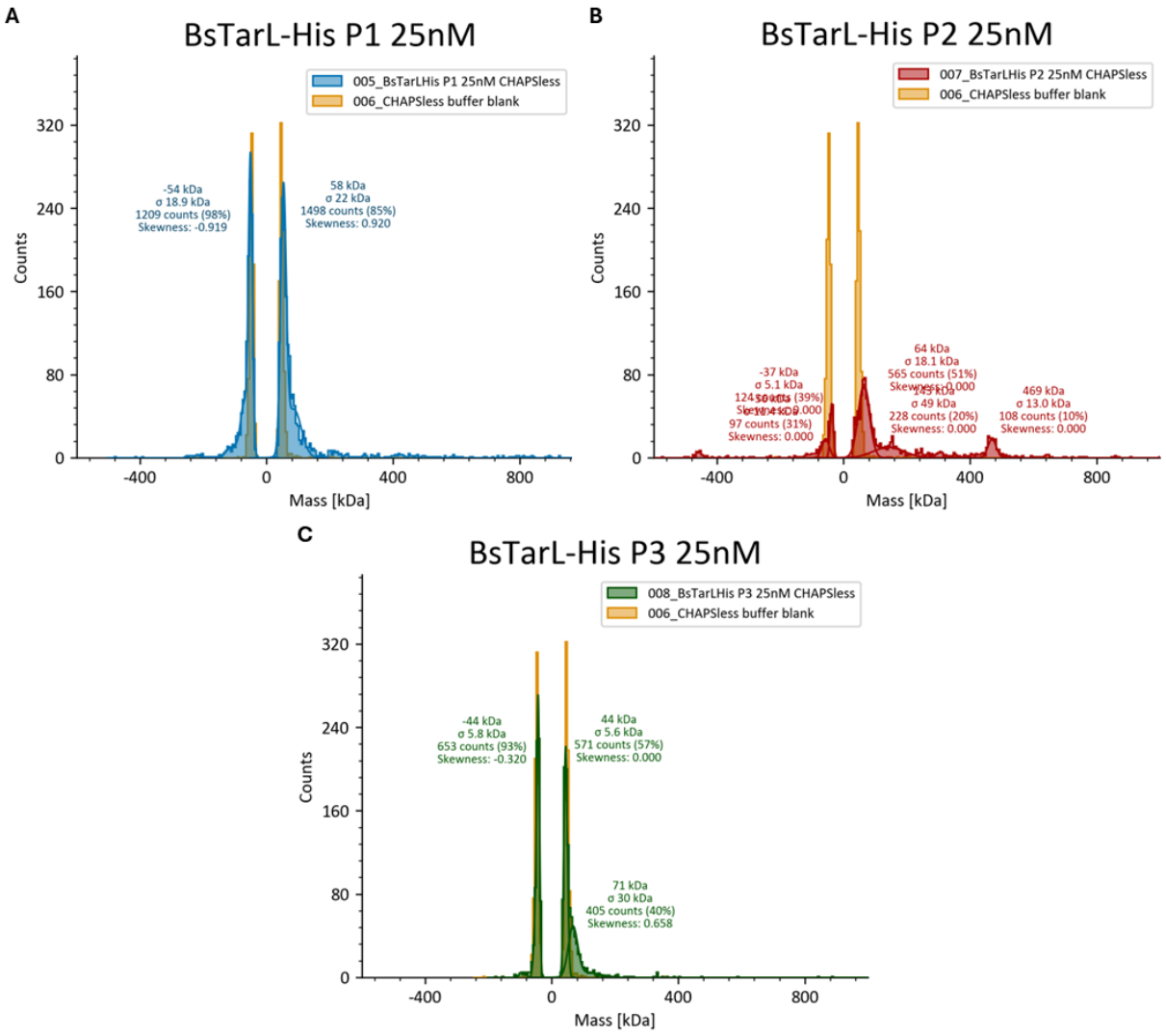


Figure 5.15: Histograms of the BsTarL mass photometry results for the three different SEC peaks at a working concentration of 25 nM. The buffer blank run is overlaid in yellow. Gaussian distribution was used to identify peaks. Large negative mass values are representative of high background noise. A = P1 sample: a single peak with an average mass of ~58 kDa (σ 22 kDa) is observed. B = P2 sample: multiple peaks are observed. The average masses identified are ~64 kDa (σ 18.1 kDa), 149 kDa (σ 49 kDa) and 469 kDa (σ 13 kDa). C = P3 sample: two peaks are observed. The average masses identified are ~44 kDa (σ 5.6 kDa) and 71 kDa (σ 30 kDa).

The combined fractions of P1 (**Figure 5.15A**) show a single peak representing a protein mass of ~58 kDa which would suggest that the *Bs*TarL protein is in a monomeric form despite this peak eluting the earliest in the SEC step. However, the presence of a negative mass peak and

the identical peaks present in the buffer blank suggest that there is contamination in the buffer used to dilute the protein and questions that this peak represents monomeric protein. The background noise can occur even in detergent-free buffers which have been vacuum filtered as seen in the literature (Niebling *et al.*, 2022). It may be necessary to include an extra step during sample preparation such as spin or disc filtering, as it would appear centrifugation of the sample and only one step of filtration of the buffer to remove particulates and impurities is insufficient. Also, negative mass values corresponding to the protein peaks of the same positive mass are not common but can occur at a lower frequency. This would explain some of the other small negative mass peaks observed in P2 (~-470 kDa, for example) and later experiments. This is often caused by the protein molecules interacting with the glass surface rather than moving away from it.

The combined fractions of P2 (**Figure 5.15B**) show multiple peaks of protein masses of ~64 kDa, ~149 kDa and ~469 kDa. This would suggest that P2 contains three multimeric states of the *Bs*TarL protein: the monomer, dimer and hexamer (~74, ~148 and ~444 kDa, respectively). However, the detected species at 469 kDa is of slightly higher molecular weight than the expected hexameric mass (444 kDa) even accounting for the standard deviation, therefore this peak appears to be either skewed by errors in the experiment or contains an impurity. Interestingly, this sample does not contain large negative mass readings like P1. This would suggest that centrifugation of the sample prior to the experiment did remove impurities.

The reason for fraction P2 containing a mixture of multimeric states is unclear. It may suggest that formation of multimers is dependent on the presence of detergent and that large dilutions of the sample into CHAPS-free buffer disrupts the stability of these structures. This would result in the breakdown of multimeric species in P2 into its constituent parts. Mass photometry experiments have been completed using detergents at and above their respective critical micelle concentration values (CMC (CHAPS CMC = 6-10 mM (Qin *et al.*, 2010)), however this increases the lowest detectable protein mass (Olerinyova *et al.*, 2021). Nonetheless, for a CHAPS concentration of 1% of the CMC as used in this experiment, masses of >70 kDa should still be detectable on the TwoMP instrument (Refeyn application notes).

The combined fractions of P3 (**Figure 5.15C**) show two peaks of protein masses ~44 kDa and ~71 kDa. The first peak is likely caused by impurities in the buffer as this peak has a negative mass reading similar to P1. The second peak, however, likely represents the monomeric form of *Bs*TarL as the mass is similar to the mass of the protein (~74 kDa). This is expected as this is the last separate peak that eluted of the SEC column and therefore should consist of the smallest unit of *Bs*TarL. Conversely, this peak is not resolved well due to the presence of the buffer impurity peak and therefore this experiment should be repeated using more stringent

buffer and sample filtering techniques mentioned previously to remove the presence of the impurity peak.

To make the mass photometry peaks more prominent, the working concentration of *Bs*TarL was increased to 50 nM and the experiment was repeated. In order to limit the potential dissociation of multimeric protein complexes in the CHAPS-free buffer used for these measurements, each sample was diluted to the working concentration one at a time and immediately applied to the instrument instead of preparing all samples at the same time. Histograms of this experiment were plotted using the Refeyn DiscoverMP software (**Figure 5.16**).

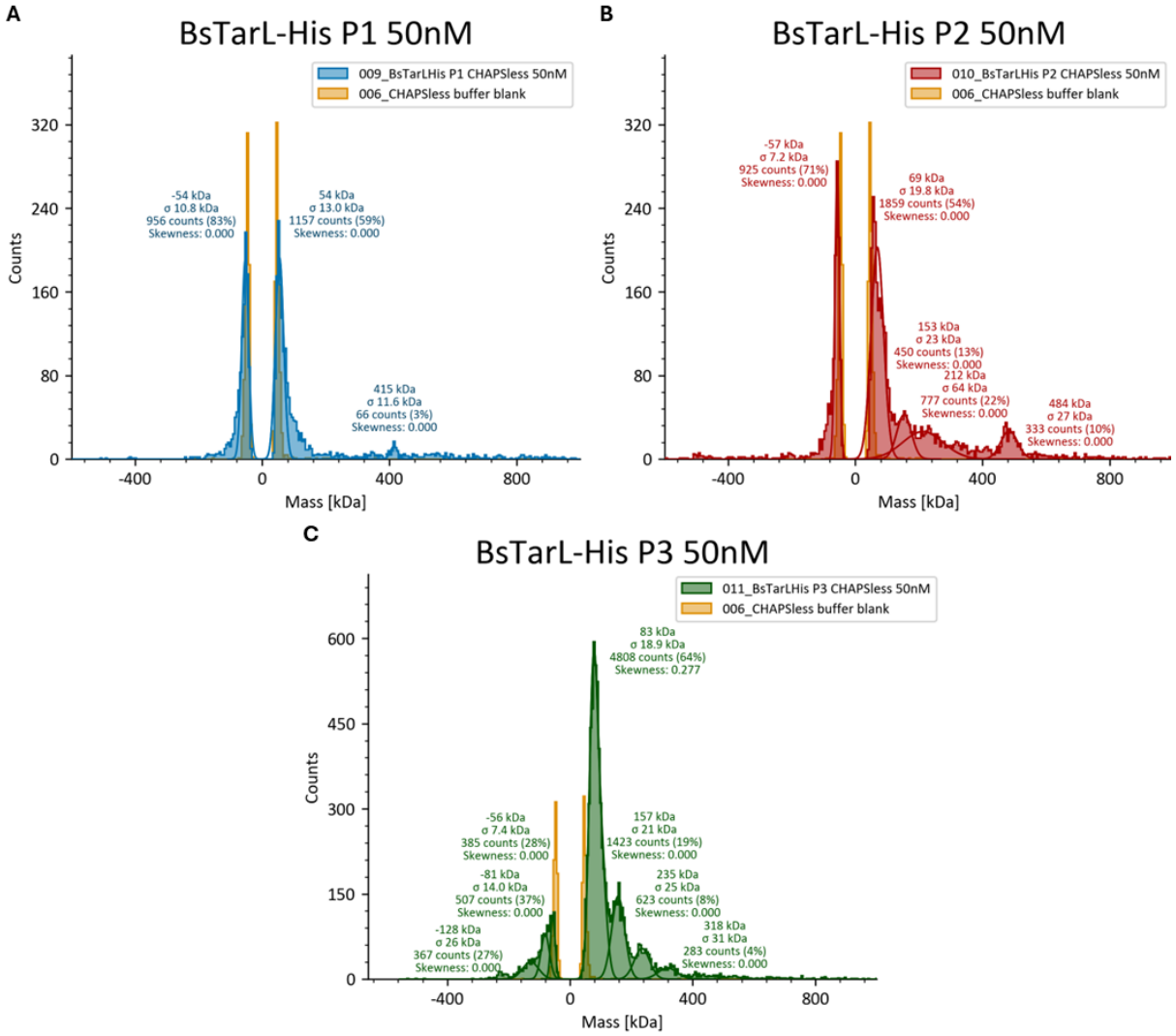


Figure 5.16: Histograms of the BsTarL mass photometry results for the three different SEC peaks at a working concentration of 50 nM. The buffer blank run is overlaid in yellow. Gaussian distribution was used to identify peaks. Large negative mass values are representative of high background noise. A = P1 sample: two peaks are observed. The average masses identified are 54 kDa (σ 13 kDa) and 415 kDa (σ 11.6 kDa). B = P2 sample: multiple peaks are observed. The average masses identified are ~69 kDa (σ 19.8 kDa), 153 kDa (σ 23 kDa), 212 kDa (σ 64 kDa) and 484 kDa (σ 27 kDa). C = P3 sample: multiple peaks are observed. The average masses identified are 83 kDa (σ 18.9 kDa), 157 kDa (σ 21 kDa), 235 kDa (σ 25 kDa) and 318 kDa (σ 31 kDa).

The combined fractions of P1 (**Figure 5.16A**) show two peaks representing a protein mass of ~54 kDa and ~415 kDa. The first peak appears to again be caused by buffer impurities as the buffer blank is also present in this range and has a corresponding negative mass peak. The second peak is more convincing of a multimeric species, however the mass shown is not equal to the expected pentameric or hexameric mass (370 and 444 kDa, respectively) even accounting for the standard deviation. Therefore, this peak is either skewed by errors in the experiment or contains an impurity. Furthermore, this second peak was not observed in the previous experiment which may suggest that the decreased time interval between diluting the sample in the CHAPS-free buffer and taking the measurement may partly preserve the presence of a higher order oligomer/aggregate, but again this mass is not equal to the expected mass of the closest multimer (444 kDa).

The combined fractions of P2 (**Figure 5.16B**) show four peaks representing protein masses of ~69, ~153, ~212 and ~484 kDa. The first may represent the monomeric species, however the impurity peak seen throughout these experiments may be interfering with the resolution of this peak. The second and third peaks (~153 and 212 kDa) are also not well resolved. These two peaks could represent the dimeric and trimeric multimers as the expected masses (~148 kDa and ~222 kDa, respectively) are within the range of each peak's standard deviation. The dimeric species was also seen in the previous experiment, but the trimeric species was not. This suggests that the decreased time interval between diluting the sample in the CHAPS-free buffer and taking the measurement may partly preserve the trimeric protein. The fourth peak has a similar mass to the previous measurement of the P2 sample and does not quite represent the expected hexameric mass (~444 kDa). A possible explanation is that a CHAPS micelle might be forming which the protein is embedded into and is increasing the perceived mass.

The combined fractions of P3 (**Figure 5.16C**) show four peaks representing protein masses of ~83, ~157, ~235 and ~318 kDa. The first and largest peak likely represents the monomeric form of the protein which is expected as this sample eluted last off the SEC column. The next peak likely represents the dimeric form of the protein as the expected mass (~148 kDa) is close to the average mass observed. The next peak likely represents the trimeric form of the protein (expected mass of ~222 kDa), however this peak is less distinct compared to the previous peaks. The final peak may represent the tetrameric form of BsTarL similar to *Sa*TarL in the literature (Li *et al.*, 2024) and which has an expected mass of ~296 kDa. Interestingly, the mass identified is again larger than the expected tetrameric mass like the previous P2 experiment. This may again suggest that a CHAPS micelle is present with the protein therefore increasing the perceived mass.

The fact that P3 contains a range of oligomers is also interesting as this is not seen in the previous P3 experiment and these fractions eluted last in the SEC purification. This could be because oligomers of *Bs*TarL form spontaneously in the presence of CHAPS micelles (SEC buffer CHAPS concentration is within the CMC range) and are in equilibrium with the monomeric form of the protein. When separating the monomeric form from the established equilibrium through SEC, the monomers begin forming oligomers to return to this equilibrium. Then, when diluting the samples with CHAPS-free buffer, the CHAPS concentration falls far below the CMC and the multimer dissociates. Because the time the samples spent in the CHAPS-free buffer was decreased in this experiment, the multimer is briefly observed in higher abundances before degrading unlike the previous set of experiments. This idea is supported by the range of oligomers observed in the P2 and P3 50 nM experiments (**Figure 5.16B** and **C**) where, despite eluting off the column at separate times, the monomeric form is the largest component of each sample, suggesting multimer degradation when CHAPS concentration is decreased.

The main issue with these experiments is the background noise observed. All experiments should be repeated with a buffer which has been through several filtration steps to remove impurities. Furthermore, the samples should be centrifuged for longer to ensure the removal of any particulates present. Also, diluting the samples in CHAPS-free buffer may not be required as detergents can be used in mass photometry measurements as long as the noise peaks do not obscure the protein mass peaks. A separate run should be performed with the SEC buffer to observe the mass peaks CHAPS may cause. If the false masses are of lower intensity? than the possible *Bs*TarL dimer masses, then this buffer could be used to observe higher order oligomer mass peaks (Refeyn application notes). This would preserve the CHAPS micelles and possibly the larger multimers of the protein.

Ultimately, there is some evidence that *Bs*TarL forms at the least dimers (**Figure 5.16**). Therefore, cryo-EM could be used to determine the protein structure. An added benefit of cryo-EM is that detergents are not an issue, which means that the proposed stable multimeric form of *Bs*TarL in a CHAPS micelle could be seen in initial 2D classes similar to *Sa*TarL (Li *et al.*, 2024).

5.2.8 Cryogenic Electron Microscopy

Each of the isolated peaks from the SEC purification (P1, P2 and P3) were prepared for cryo-EM grids as described in section **5.4.8**. A representative micrograph is shown in **Figure 5.17 A** with a comparison with the *Sa*TarL micrograph from Franco K. K. Li *et al*, 2024 (**Figure 5.17B**). The RELION pipeline was used until the initial 2D classification stage. **Figure 5.18** shows some of the 2D classes identified from the initial screens and data collection of the P2 and P3 *Bs*TarL samples.

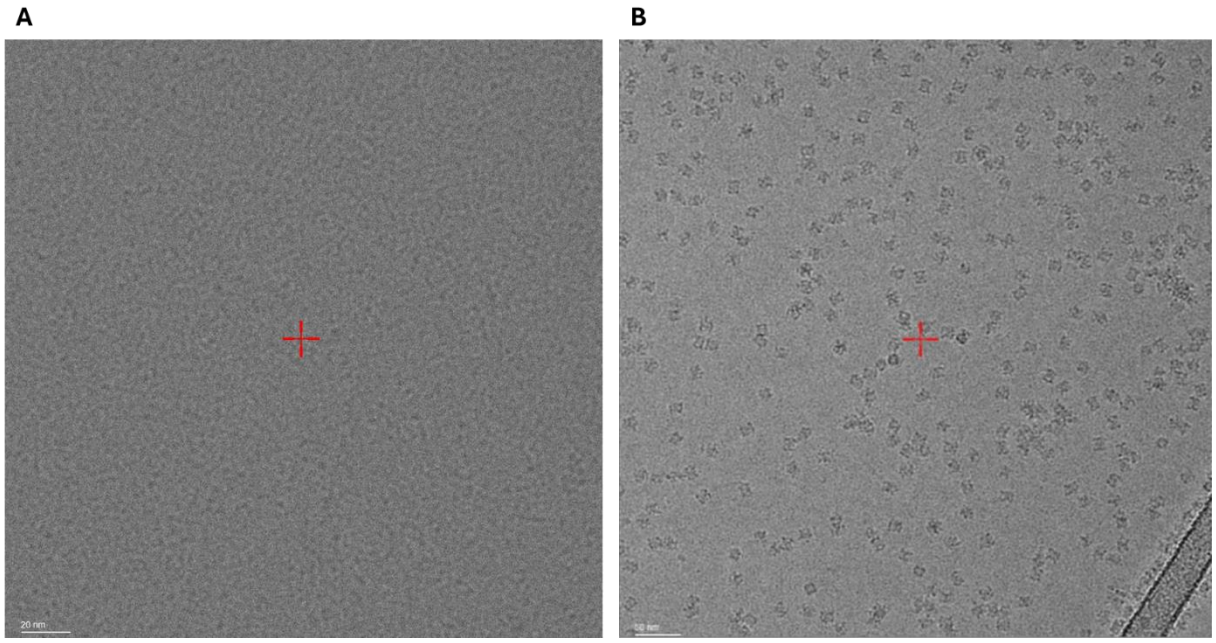


Figure 5.17: Raw micrographs from this work (A) and Franco K. K. Li et al*, 2024 (B). The scale bar represents 20 nm and 50 nm for A and B, respectively.*

Comparing the representative micrographs from this work and Franco K. K. Li *et al*, 2024 (**Figure 5.17**) shows a clear difference in the particles present. In this work, the particles, if any, are not clearly observed unlike the micrograph from Franco K. K. Li *et al*, 2024 where tetrameric *Sa*TarL embedded into a CHAPS micelle can clearly be observed. This suggests that similar tetrameric particles of *Bs*TarL are not present or are obscured by noise. This could be due unoptimised grid preparation or, that simply *Bs*TarL does not form stable tetramers.

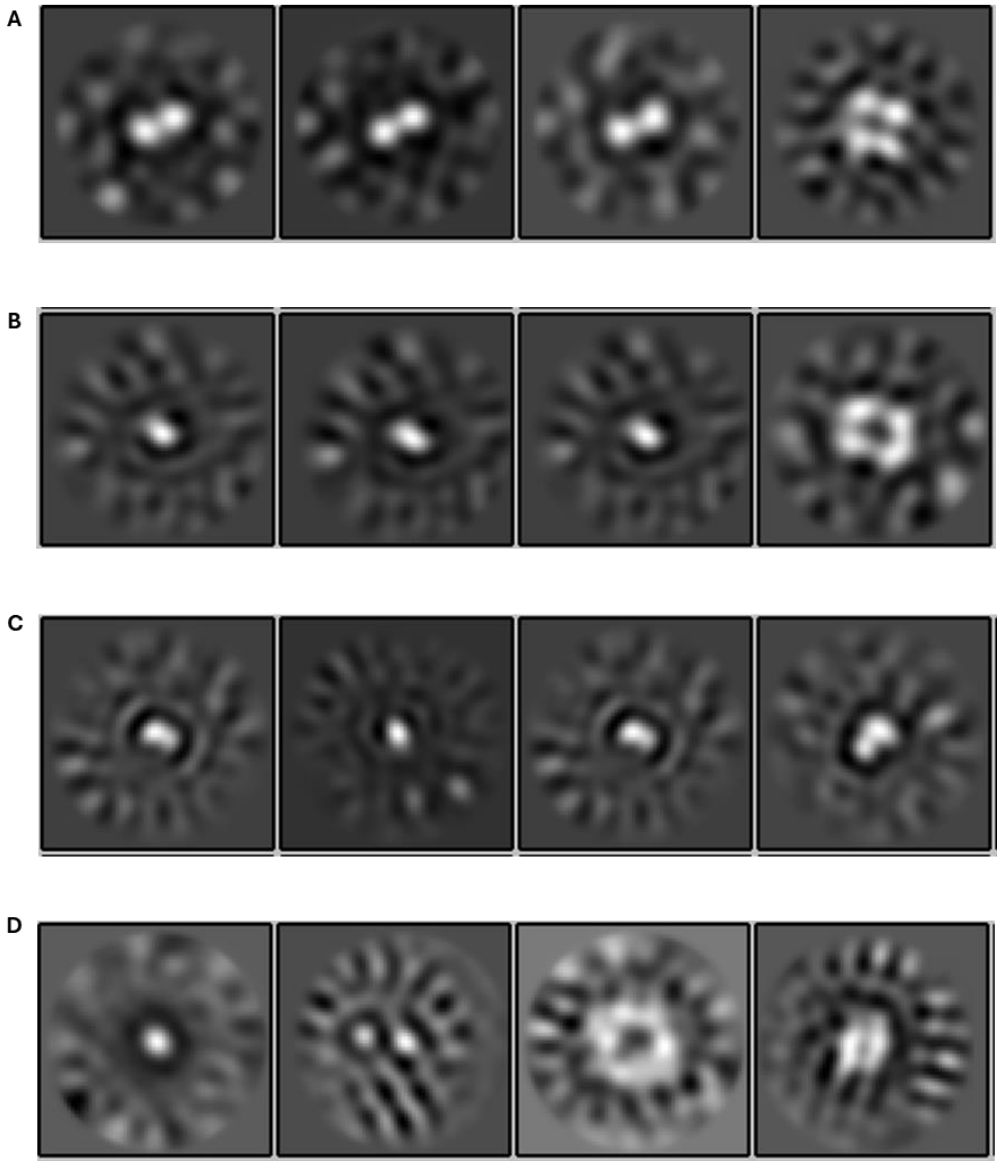


Figure 5.18: Examples of the top 2D classes of BsTarL from P2 (peak fractions from SEC, B6-B4) and P3 (peak fractions from SEC, B3-B1). A = P2 sample at 1 mg/mL blotted with 0 force for 5 secs. B = P2 sample at 1 mg/mL blotted with -10 force for 5 secs. C = P3 sample at 1 mg/mL blotted with -10 force for 5 secs. D = P2 sample at 8 mg/mL blotted with 5 force for 5 secs.

In **Figure 5.18A** there are 2D classes that are reminiscent of either a dimer or a tetramer rotated 90° perpendicular to the symmetry axis. This provides evidence that *Bs*TarL does seem to readily form a multimer. The final class in this figure also suggests a tetrameric form is present in this grid with 4-fold symmetry like the tetrameric structures of *Sa*TarL and *Se*TagF (Lovering *et al.*, 2010; Li *et al.*, 2024). However, this class is most likely a common *E. coli* contaminant protein. The same observation can be made in the last class in **Figure 5.18B**, however the majority of 2D classes in this preparation appear to be representative of the monomeric form of the protein or a contaminate.

Figure 5.18C shows 2D classes which resemble the monomeric form of the protein. This is expected as the P3 sample eluted last in the SEC purification step and is consistent with the mass photometry results where the monomeric form was the most abundant.

To summarise, the initial screening of grid preparation conditions and subsequent initial 2D classification may identify the formation of a *Bs*TarL homotetramer. However, due to time restraints, an initial 3D model was not generated. Furthermore, there seems to be some heterogeneity in the samples. The SEC purification may require optimisation to fully separate the different *Bs*TarL species, or the pipeline from purification to grid preparation could be sped up to decrease the possibility of the large multimers disassembling. Also, further optimisation of the grid preparation parameters as well as ideal protein concentrations should be performed to ensure grids are of the best quality for larger micrograph data collections.

5.3 Conclusions and Future Work

The use of different tagged versions of the *Bs*TarL protein described in this chapter was vital in uncovering the conditions required for the soluble expression of the target enzyme. Extensive expression trials involving the use of several different *E. coli* cell lines and lysis condition also aided in finalising ideal expression variables. The finalised conditions are growth in 2xYT media until an OD_{600} of 0.5-0.8, induction with 1 mM IPTG at 16 °C overnight using Rosetta2 plysS *E. coli* cells transformed with the C-terminal His6 tag version of *Bs*TarL. With the soluble expression protocol of *Bs*TarL developed, purification of the protein proceeded using a multi-step method, with inspiration from the literature, which involves the initial capture of the target protein using IMAC followed by a clean-up step using SEC to isolate various multimeric versions of *Bs*TarL. This allowed highly pure and soluble *Bs*TarL to be produced for characterisation experiments using CD and nanoDSF. The active site mutants of *Bs*TarL have also been produced and had their stability assessed using nanoDSF. These experiments showed that the proteins are folded and are stabilised in the presence of glycerol. Furthermore, protein crystallisation trials have been performed, however no protein crystals have been produced. Nonetheless, interesting results from the SEC purification uncovered the presence of oligomeric versions of *Bs*TarL which were explored further using mass photometry. The range of multimeric species identified in the mass photometry experiments then opened the door to early cryo-EM experiments and single particle analysis which resulted in 2D classes that indicate the presence of a homotetramer.

Future work on *Bs*TarL should focus on further biochemical characterisation with the assessment of the enzyme's activity using either liquid chromatography–mass spectrometry (LC–MS) or the UMP/CMP-Glo™ Glycosyltransferase Assay (Promega). This, however, would require the synthesis of the acceptor substrate either by *Bs*TarK or chemical synthesis, or isolation of native WTAs. Also, screening of the inhibitors used in the *Bs*TarK chapter should also be performed using methods such as fluorescence dye based thermal shift assays or isothermal titration calorimetry to highlight the inhibitors with the greatest binding affinity.

The preliminary work of solving the structure of *Bs*TarL by cryo-EM looks promising. However, issues of homogeneity and ice thickness and quality in the grid preparation need to be solved by further optimisation. This would allow higher quality particles to be observed and therefore finer, more detailed 2D classes to be generated. The current data set could also be worked on further using RELION to generate low-resolution 3D classes to further confirm the existence of the homotetrameric species and allow comparisons to be made with the *Sa*TarL homotetramer. Once the optimisations are finalised, substrate bound structures of the WT and

mutant enzymes using CDP-Rbo can then be solved along with any promising inhibitors to aid in further inhibitor design.

5.4 Methods

5.4.1 Molecular Cloning

The linearisation of the Champion™ pET SUMO vector (Invitrogen) is outlined in the methods section of the previous chapter (**4.4.1**). For *Bs*TarL insert synthesis the pUC57*Bs*TarL plasmid (GenScript) was used with the primers 'SUMOTarLBs F' and 'SUMOTarLBs R' (**Table 7-2**, entries 3 and 4). Correct insert and linearised vector formation were confirmed by agarose gel electrophoresis. Subsequently, DNA isolation, ligation and transformation of *E. coli* was performed as described in section **7.2.1**.

The cloning of the p*Bs*TarL-His28a plasmid was carried out following the In-Fusion cloning protocol outlined in section **7.2.1**. To linearise the vector, the 'p28aCtermHis lin F' and 'p28aCtermHis lin R' primers were used ((**Table 7-2**, entries 19 and 20). The insert was synthesised using the 'pBsTarL-His-28a F' and 'pBsTarL-His-28a R' primers (**Table 7-2,** entries 23 and 24). Verification of the correct insert and linearised vector, followed by isolation, ligation, and transformation, was performed as described in the In-Fusion cloning section of the general methods (**7.2.1**).

5.4.2 Expression Trials

The first trial of *Bs*TarL proceeded using the pSUMO*Bs*TarL plasmid and the conditions in **Table 5-1**, entry 1. Overnight cultures were first grown using 5-10 mL of kanamycin 2xYT and inoculated with a colony from the transformation plate of BL21 (DE3) containing the pSUMO*Bs*TarL construct. Flasks containing 50 mL of kanamycin 2xYT were subsequently inoculated with the overnight cultures and grown at 37 °C, 200 RPM until the desired OD_{600} was reached. The cultures were then induced with 1 mM IPTG and incubated overnight at 16 °C at 200 RPM. 1 mL samples were taken before induction and after overnight incubation. The samples were pelleted by centrifugation at 8,000 xg for 5 mins and pellets were lysed using lysis buffer (100 mM Tris, 500 mM NaCl, 0.6% (w/v) CHAPS and 20% (v/v) glycerol, pH 7.5) supplemented with protease inhibitors, benzonase and lysozyme and lysed by sonication on ice at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (16,000 xg for 20 mins) and the soluble fraction was collected. Cell debris was then resuspended as much as possible using the 1 mL of lysis buffer). All the samples were then prepared for SDS-PAGE analysis (**7.2.8**).

The subsequent trials listed in **Table 5-1** numbered 2, 3 and 4 are analogous to the trials listed in **4.4.2, Table 4-1** numbered 4, 5 and 6, respectively. All conditions are identical with the exception of the plasmids used and thus the proteins expressed.

Table 5-1: List of the BsTarL expression trials and the variables used

Trial no.	Plasmid	OD_{600} range	IPTG conc (mM)	Expression temp and max time	*E. coli* Strain	Media
1	pSUMO*Bs*TarL	0.5-0.8	1.0	16 °C, O/N	BL21(DE3)	2xYT
2	pHis-*Bs*TarL28a or p*Bs*TarL-His28a	0.5-0.8	1.0	16 °C, O/N	Rosetta2(DE3) pLysS	2xYT
3	pHis-*Bs*TarL28a	0.5-0.8	1.0	16 °C, O/N	Rosetta2(DE3) pLysS or SHuffle lysY	2xYT
4	p*Bs*TarL-His28a	0.5-0.8	1.0	16 °C, O/N	Rosetta2(DE3) pLysS or SHuffle lysY	2xYT

5.4.3 Single Point Mutagenesis

The standard method of mutagenesis is described in section **7.2.4**. For the *Bs*TarL H345N mutant, the primers 'BsTarL H345N F' and 'BsTarK H119N R' (**Table 7-2**, entries 33 and 34) were used for the PCR reactions. For the *Bs*TarL H477A mutant, the primers 'BsTarL H477A F' and 'BsTarL H477A R' (**Table 7-2,** entries 35 and 36) were used for the PCR reactions. Both mutants were made using the pBsTarL-His28a plasmid as the template.

5.4.4 Protein Purification

The outline of the two-step purification method used for the purification of *Bs*TarL-His is outlined in **Table 5-2**. The buffers for each step are listed in **Table 5-3**.

Table 5-2: Outline of the method used for purifying BsTarL.

Method no.	Protein	Step 1		Step 2	
		Column method	Buffers used	Column method	Buffers used
1	*Bs*TarL-His	IMAC (Ni^{2+})	His A/B	SEC	SEC

Table 5-3: List of buffers used for purifying BsTarL-His.

Buffer no.	Name	Buffer component conc	Salt conc	CHAPS conc	pH	Imidazole conc
1	Lysis	20 mM HEPES	500 mM NaCl	16 mM	8.0	25 mM

2	His A	20 mM HEPES	500 mM NaCl	6.5 mM	8.0	25 mM
3	His B	20 mM HEPES	500 mM NaCl	6.5 mM	8.0	25 mM
4	SEC	20 mM HEPES	500 mM NaCl	6.5 mM	8.0	N/A

The expression of *Bs*TarL-His proceeded using the pBsTarL-His28a plasmid which was expressed using the conditions in **Table 5-1**, entry 4 using the Rosetta2(DE3) pLysS strain. This method of protein expression is also identical to the conditions used in the second purification method of *Bs*TarK-His (**4.4.4**). The cell paste was resuspended in 50 mL of lysis buffer (**Table 5-3**, entry 1) supplemented with protease inhibitors, benzonase and lysozyme, and was lysed using sonication at 40% amplitude for 2 mins with 10 secs on and 5 secs off. Insoluble cell debris was pelleted by centrifugation (38,000 xg for 30 mins at 4 °C) and the soluble fraction was collected. Cell lysate was subsequently loaded onto a 5 mL HisTrap FF Crude column (Cytiva), equilibrated with His A buffer (**Table 5-3**, entry 2), using an AKTA Go for IMAC purification. Flow through was collected and the column was washed with 15 column volumes (CV) of His A buffer. Proteins were then eluted off the column with a linear gradient of 0-100% of His B buffer (**Table 5-3**, entry 3) over 20 CV. 2 mL fractions were collected. The soluble and insoluble fractions, flow through, wash and fractions of interest were sampled and prepared for SDS-PAGE analysis (**7.2.8**).

The fractions of interest containing the target protein were combined and buffer exchanged into SEC buffer (**Table 5-3**, entry 4) (see **7.2.7**). The sample was concentrated using spin concentration to <500 µL (see **7.2.7**) and applied onto a Superose™ 6 Increase 10/300 column (Cytiva) which was previously equilibrated with SEC buffer. The sample was centrifuged to remove any precipitate and subsequently injected into the system. The purification proceeded at 0.25 mL/min with SEC buffer until 30 mL had washed through and the proteins eluted. 0.5 mL fractions were collected, and samples were taken from fractions of interest defined by the UV trace. These fractions were prepared for SDS-PAGE (**7.2.8**). The combined fractions from different peaks containing pure *Bs*TarL-His were then prepared for downstream applications.

For nanoDSF experiments the wild type and mutant variants of *Bs*TarL-His, H345N and H477A, were purified using the second set of purification methods from section **4.4.4**, **Table 4-2**, entry 2.

5.4.5 Crystallography

Pure *Bs*TarL-His from the IMAC purification step was buffer exchanged into His A buffer without imidazole and concentrated to 12 mg/mL. For sitting drop vapour diffusion crystallisation, the protein solution was dispensed in a ratio of 150nL:150nL of protein to reservoir solution into five 96 well plate crystallisation screens (Crystal Screen HT (Hampton

Research), JCSG+ (Molecular Dimensions), PEG Ion HT (Hampton Research), MPD Screen (Qiagen) and Ammonium Sulphate (Qiagen)) using the Mosquito robot (SPT Labtech). Following crystal observations in the Crystal Screen HT (well F10) and PEG Ion HT screens (wells F1-2, 7-8, 10 and A3), two optimisation plates were designed and set up using fresh *Bs*TarL at a concentration of 15mg/mL and the Mosquito robot. After a week of incubation, three wells (A3, B3 and C3: 0.1M MES, pH 6.5 with either 6, 8 or 10% v/v PEG 20000, respectively) yielded crystals. The crystals were fished, flash frozen in liquid nitrogen and sent to the Diamond Light Source for X-ray diffraction.

5.4.6 NanoDSF

All analyses were performed on the Prometheus nanoDSF instrument (NanoTemper Technologies) and used the same method described in section **7.2.11**.

Purified *Bs*TarL (WT and the two variants, H345N and H477A) was buffer exchanged into 100 mM Tris, 500 mM NaCl, 0.6% (w/v) CHAPS and 2 mM β-ME, pH 7.5 with and without 20% v/v glycerol and concentrated to 1.5 mg/mL using spin concentrators (**7.2.7**). The melting temperature experiment proceeded as described in section **7.2.11.**

5.4.7 Mass Photometry

Separate peaks from the SEC purification which may represent different multimeric states of *Bs*TarL, named P1, P2 and P3, were isolated and diluted to a concentration of 0.1 mg/mL (~1350nM) using SEC buffer. The samples were then diluted using CHAPS-free SEC buffer to an initial working concentration of 25 nM and centrifuged at 1000 xg for 1 min to remove particulates. Protein solutions were then applied to the Refeyn TwoMP mass photometer (Refeyn). The sample was recorded for 1 min in 'normal' mode and the data was acquired using the Refeyn AcquireMP 2024.1.0 software. The data generated was assessed using data from a previous run with protein standards of known molecular weight to calibrate the machine and determine the molecular weights observed in the sample. The data was plotted as a histogram using the DiscoverMP software and peaks were identified using a gaussian function.

This method was repeated for all samples using a working concentration of 50 nM.

5.4.8 Cryogenic Electron Microscopy

Separate peaks from the SEC purification which may represent different multimeric states of *Bs*TarL, named P1, P2 and P3, were isolated and concentrated to a concentration of ether 1 or 8 mg/mL. 3 µL of each sample was applied to glow discharged Quantifoil™ 300 copper mesh R1.2/1.3 holey carbon film grids. The grids were blotted with ether -10 or 0 force using

the Thermo Fisher (FEI) Vitrobot Mark IV at 90% humidity for 1.5 s and vitrified using liquid ethane. The grids were then stored in liquid nitrogen.

To screen for grid quality, each grid was loaded into the Thermo Fisher Glacios Cryo-Transmission Electron Microscopy (TEM) equipped with a Falcon 4 direct electron detector set to a magnification of 240,000x allowing for a pixel size of 0.574 Å. The defocus list was also set to -2.0, -1.8, -1.6, -1.4. Grid squares of sufficient quality and ice thinness were selected, and micrographs were automatically acquired overnight yielding ~500 images. All micrograph images were imported into RELION 5.0 (Scheres, 2012). Particle picking was performed using the Topaz auto picking pipeline (Bepler *et al.*, 2019). To generate initial 2D classes, each set of particles underwent 140 iterations of the 2D classification job. All computational analysis was performed using the Viking computing cluster.

Chapter 6: Summary and Future Prospects

Wall teichoic acids (WTAs) are known to play a role in cell growth, division and morphology as well as colonisation and adhesion in Gram-positive bacterial species such as *Staphylococcus aureus* and *Bacillus subtilis* (Matthias Gross *et al.*, 2001; Soldo, Lazarevic and Karamata, 2002; Weidenmaier *et al.*, 2004; Swoboda *et al.*, 2010; Wu *et al.*, 2021). Specifically in *S. aureus* and *B. subtilis* W23, WTAs largely consists of a ribitol-5-phosphate (RboP) polymer, which is synthesised by TarK and TarL enzymes using the donor CDP-ribitol (CDP-Rbo). This donor substrate, CDP-Rbo, is synthesised from RboP by TarI (Pereira and Brown, 2004; Meredith, Swoboda and Walker, 2008; Brown *et al.*, 2010). Some studies have shown that deletion of the enzymes along the WTA pathway results in restored antibiotic susceptibility in methicillin-resistant *S. aureus* (MRSA), demonstrating the potential of WTA biosynthesis as a therapeutic target. Therefore the characterisation of the Tar enzymes could aid in the development of novel antimicrobial and antivirulence agents (Brown *et al.*, 2012a; Farha *et al.*, 2013; Lee *et al.*, 2016). This project endeavoured to facilitate the generation of the difficult-to-synthesise donor substrate, CDP-Rbo, via TarI and to characterise the TarK and TarL enzymes, thereby providing insight into the protein activity and 3D structures.

Firstly, CDP-Rbo synthesis was achieved using a chemoenzymatic approach. This used the TarI protein successfully expressed and purified as described in Chapter 2 along with RboP produced by members of the group via chemical synthesis. The generation of CDP-Rbo was vital in testing the activity of TarK as described in Chapter 4 and will be required in the future for ligand binding studies and testing TarL activity.

The initial screen of the various homologues of TarK and TarL from different bacterial species described in Chapter 3 yielded mixed results. *Bacillus subtilis* (*Bs*)TarL was the only soluble protein purified from the initial selection of homologues. However, this Chapter provided useful insights into the direction to take for the subsequent expression trials in (Chapters 4 and 5). This involved testing various conditions for the expression of these proteins such as the cell line, expression medium and the position of the affinity tag. Other alterations involved using different lysis buffers with additives such as CHAPS and/or glycerol to solubilise the target proteins. Chapter 3 also identified active site residues that are potentially involved in catalytic activity, which could be mutated for both TarK and TarL. These residues of interest were found by using protein sequence alignment with other WTA enzymes such as *Se*TagF and *Sa*TarL mentioned in the literature (Lovering *et al.*, 2010; Cifuente *et al.*, 2023; Li *et al.*, 2024).

After initial difficulties in expressing *Bs*TarK, Chapter 4 describes the various expression trials performed to find the optimised expression conditions for the expression of soluble affinity

tagged protein. The expression involved using either Rosetta2(DE3) pLysS (Novagen) or SHuffle lysY (New England Biolabs) *E. coli* strains grown in 2x yeast tryptone media to an OD_{600} of 0.5-0.8. Expression of *Bs*TarK-His was then induced using 1 mM IPTG and protein expression proceeded at 16 °C overnight. The slower expression due to the low temperature likely aided in the solubilising the protein while the tag location likely did not matter. This then allowed highly pure *Bs*TarK to be produced using an optimised multi-step column purification protocol involving immobilised metal affinity chromatography (IMAC) followed by heparin affinity chromatography. These methods were also used in successfully purifying the single point mutants of the target residues identified in the previous chapter which were generated using site-directed mutagenesis. Pure *Bs*TarK was then tested with CD to observe the secondary structures present in the folded protein. Protein crystallography trials were performed but yielding protein crystals proved challenging. Melting temperature studies using nano differential scanning fluorimetry (nanoDSF) were also accomplished. The melting temperature values were found to be ~43 °C for the WT enzyme without 20% v/v glycerol which increased to 50 °C with 20% v/v glycerol. The *Bs*TarK H255A mutant was also tested, and the melting temperatures were ~38 °C and ~47 °C without and with 20% v/v glycerol, respectively. The wild type (WT) enzyme was also tested with various ligands which identified that the addition of the native donor (CDP-Rbo) did not increase the melting temperature of the protein. However, this was also true for all the inhibitors listed in **Table 4-4**. However, this assay was ultimately found to be unsuitable for using CDP-containing ligands since the absorbance of the cytidine moiety interferes with the assay readout, meaning that the obtained results may not represent the true melting temperatures of the protein with the ligands present as described in Chapter 4. The activity of *Bs*TarK also was tested using LC-MS assays which showed that the enzyme was active, however only a small amount of product formed.

Similar to *Bs*TarK in Chapter 4, *Bs*TarL expression was further optimised successfully to produce large amounts of soluble protein in Chapter 5. This was achieved by testing different *E. coli* cell lines and tag locations and buffers containing CHAPS detergent. The expression protocol involved using Rosetta2(DE3) pLysS (Novagen) *E. coli* strains grown in 2x yeast tryptone media to an OD_{600} of 0.5-0.8. Expression of *Bs*TarL-His was then induced using 1 mM IPTG and protein expression proceeded at 16 °C overnight. Unlike *Bs*TarK, the tag location seems to matter, with the C-terminal His6 tag increasing *Bs*TarL solubility. In *Sa*TarL, the N-terminal domain is required for tetramerisation, therefore, if the protein was N-terminally tagged this could affect the formation of a possibly more stable and soluble oligomeric protein (Li *et al.*, 2024). This domain is not present in TarK which would explain that the tag location does not matter as much in TarK than TarL. Then a multi-step protein column purification method was developed to produce highly pure protein. This method used IMAC followed by

size exclusion chromatography. *Bs*TarL mutants were also generated and purified in a similar manner to WT *Bs*TarK and the variants. The melting temperatures of all *Bs*TarL variants were tested by nanoDSF to assess protein stability. The melting temperature values were ~46 °C for all variants which increased to ~52-53 °C with the addition of 20% v/v glycerol. This showed that the melting temperature values did not change much between mutants and WT and the addition of 20% v/v glycerol had a stabilising effect on all variants. The positive effect of glycerol on protein melting temperature is somewhat expected as it increases the hydrophobic interactions between nonpolar amino acids within the protein, effectively stabilising its native structure and preventing unfolding. Initial crystallography trials of *Bs*TarL were not successful, however, size exclusion chromatography results identified for the first time that *Bs*TarL may form a multimeric protein like *Sa*TarL and *Se*TagF (Lovering *et al.*, 2010; Li *et al.*, 2024). This observation was explored further using mass photometry which highlighted that *Bs*TarL exists as a range of multimeric species. This opened the door to cryogenic electron microscopy (cryo-EM) as a technique to solve the structure of this protein as any multimer is large enough for this method to be effective. Initial 2D classification of the multimeric *Bs*TarL resolved a rudimentary image of a *Bs*TarL tetrameric class, however due to time restraints 3D classes were not developed.

The obvious next step for the project would be finish the RELION pipeline to generate an initial 3D model of *Bs*TarL, therefore yielding a novel protein structure. Using this pipeline, the active site mutants could also have their structure solved. Specifically, for *Bs*TarL, the H345N mutant may be able to trap the CDP-Rbo substrate. In *Se*TagF and *Sa*TarL the equivalent H345 residues (H444 and H300, respectively) function as an active-site base but when mutated to asparagine, the enzyme activity is reduced for both enzymes and the carbonyl group of asparagine forms hydrogen bond interactions with the hydroxyl group of the CDP-polyol substrate (Lovering *et al.*, 2010; Li *et al.*, 2024). If this could be replicated for *Bs*TarL this would provide insights to the mechanistic action of the enzyme. This knowledge could then be used to observe binding of non-hydrolysable ligand derivatives which may act as inhibitors, to fuel structure-based inhibitor/probe design. This should also be pursued for *Bs*TarK through further crystal trials or cryo-EM studies if multimeric studies of *Bs*TarK yield a protein large enough for this method (>100 kDa).

Other future work should focus on developing a reliable method to assay *Bs*TarK and *Bs*TarL activity. Firstly, the CDP-Rbo donor substrate can now be synthesised (Chapter 2), however the acceptor substrate of *Bs*TarK used in this Chapter (**Figure 4.34**: The acceptor substrate TV1041.) is not native. One solution would be to grow and harvest *B. subtilis* W23 WTAs and digest them to produce the TagO product similar to a procedure used to study the activity of *Sa*TarL (Meredith, Swoboda and Walker, 2008). Alternatively, to quantitatively test the activity

of these enzymes, an acceptor substrate with a fluorescent tag could be used along with the liquid chromatography–mass spectrometry (LC-MS) method described in Chapter 4. This would allow the observation of product formation by comparing fluorescence signals between the 'no enzyme' control and the experimental reaction therefore allowing the quantification of product formed and therefore reaction velocity. After the RboP primase reaction is optimised, the same method could then be used to measure the activity of *Bs*TarL by using the *Bs*TarK product. With the activity assays developed, possible inhibitors could then be added at known concentration to determine the IC_{50} therefore leading into further antimicrobial studies. These studies could involve testing the growth rate of *S. aureus* and/or *B. subtilis* when these possible inhibitory compounds are added to the growth medium.

For other ligand binding studies isothermal titration calorimetry (ITC) could be used. This method requires a large amount of material; however, this approach would allow the deduction of the equilibrium constants and binding affinities of possible inhibitors and ligands in much more detail than the protein melting temperature experiments performed in Chapters 4 and 5. This technique is also not susceptible to the absorption properties of the CDP-containing ligands as the ligand affinities are measured using the change in heat energy. The results of these experiments would then act as a rough guide as to what ligands should be tested for the aforementioned enzyme inhibition studies.

Finally, as a structure exists of *Sa*TarL, it would be advisable to focus on working with this enzyme for future ligand binding studies if carrying on with this work appears difficult (Li *et al.*, 2024). The authors have reported the methods required to synthesise the target protein at large quantities and at sufficient purity, therefore a future student or researcher could attempt to replicate their work or collaborate with the group to aid further ribitol phosphate transferase research. Combined with future drug/inhibitor screening and the cryo-EM methods described in Li *et al.*, 2024, novel ligand bound structures of *Sa*TarL could be resolved therefore furthering our understanding of the catalytic mechanisms of these enzymes and facilitating downstream optimisation of synthetic ligand design for antimicrobial studies.

Chapter 7: Materials and General Methods

7.1 Materials

7.1.1 Bacterial Strains

Stellar™ Competent Cells (Takara Bio)

Genotype: F-, ara, Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, dcm

XL10-Gold Ultracompetent Cells (Agilent)

Genotype: TetrD(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZDM15 Tn10 (Tetr) Amy Camr]

SHuffle T7 Competent Cells (New England Biolabs)

Genotype: F´ lac, pro, lacIq / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB rpsL150(StrR) Δgor Δ(malF)3

BL21(DE3) Competent Cells (New England Biolabs)

Genotype: F– *ompT* $hsdS_B$ (r_B–, m_B–) *gal dcm* (DE3)

Rosetta™ 2(DE3) pLysS Competent Cells (Novagen)

F^- *ompT* $hsdS_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE2 (Cam^R)

7.1.2 Plasmids

Table 7-1: List of vectors and plasmids used in this work.

Construct	Tag Type/ Location	Gene	Vector	Description	Antibiotic Resistance	Source
Champion™ pET SUMO	N/A	N/A	N/A	Plasmid vector	Kanamycin	Invitrogen
pBsTarKH119N-His28a	Hexa-histidine tag/C-terminus	*Bs*TarK H119N mutant	pET-28a(+)	For mutant protein overexpression and cloning	Kanamycin	This work
pBsTarKH225A-His28a	Hexa-histidine tag/C-terminus	*Bs*TarK H225A mutant	pET-28a(+)	For mutant protein overexpression and cloning	Kanamycin	This work

pBsTarK-His28a	Hexa-histidine tag/C-terminus	*Bs*TarK	pET-28a(+)	For protein overexpression and cloning	Kanamycin	This work
pBsTarLH345N-His28a	Hexa-histidine tag/C-terminus	*Bs*TarL H345N mutant	pET-28a(+)	For mutant protein overexpression and cloning	Kanamycin	This work
pBsTarLH477A-His28a	Hexa-histidine tag/C-terminus	*Bs*TarL H477A mutant	pET-28a(+)	For mutant protein overexpression and cloning	Kanamycin	This work
pBsTarL-His28a	Hexa-histidine tag/C-terminus	*Bs*TarL	pET-28a(+)	For protein overexpression and cloning	Kanamycin	This work
pBvTarK-His28a	Hexa-histidine tag/C-terminus	*Bv*TarK	pET-28a(+)	For protein overexpression and cloning	Kanamycin	This work
pHis-BgTarK28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bg*TarK	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pHis-BpaTarL28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bpa*TarL	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pHis-BpuTarL28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bpu*TarL	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pHis-BsTarK28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bs*TarK	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pHis-BsTarL28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bs*TarL	pET-28a(+)	For protein overexpression and co-overexpression with pHis-BsTarK15b and cloning	Kanamycin	GenScript
pHis-BsTH007TarK28a	Thrombin cleavable hexa-	*Bs*TH007 TarK	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript

	histidine tag/N-terminus					
pHis-BvTarK28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bv*TarK	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pHis-BvTarL28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Bv*TarL	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pHis-SaTarL28a	Thrombin cleavable hexa-histidine tag/N-terminus	*Sa*TarL	pET-28a(+)	For protein overexpression and cloning	Kanamycin	GenScript
pL238-BsTarL-His28a	Hexa-histidine tag/C-terminus	*Bs*TarL NTD truncation	pET-28a(+)	For mutant protein overexpression and cloning	Kanamycin	This work
pSUMOBsTarK	Hexa-histidine-SUMO tag/N-terminus	*Bs*TarK	Champion ™ pET SUMO	For protein overexpression and cloning	Kanamycin	This work
pSUMOBsTarL	Hexa-histidine-SUMO tag/N-terminus	*Bs*TarL	Champion ™ pET SUMO	For protein overexpression and cloning	Kanamycin	This work
pSUMOSaTarL	Hexa-histidine-SUMO tag/N-terminus	*Sa*TarL	Champion ™ pET SUMO	For protein overexpression and cloning	Kanamycin	This work
pSUMOTarI	Hexa-histidine-SUMO tag/N-terminus	TarI	Champion ™ pET SUMO	For protein overexpression and cloning	Kanamycin	This work
pUC57BsTarK	N/A	*Bs*TarK	pUC57	For cloning	Ampicillin	GenScript
pUC57BsTarL	N/A	*Bs*TarL	pUC57	For cloning	Ampicillin	GenScript
pUC57SaTarK	N/A	*Sa*TarL	pUC57	For cloning	Ampicillin	GenScript

7.1.3 Primers

Table 7-2: List of all primers used in this work.

Number	Name	Sequence	Description
1	SUMOTarKBs F	GAACAGATTGGTGGTATGAAGACATTTCTAACTAGGATAG	Forward primer for cloning out TarKBs from pUC57 for insertion into linear pSUMO using in-fusion cloning
2	SUMOTarKBs R	TACCTAAGCTTGTCTTTAATCCGCGCCGC	Reverse primer for cloning out TarKBs from pUC57 for insertion into linear pSUMO using In-Fusion cloning
3	SUMOTarLBs F	GAACAGATTGGTGGTATGAAACTAGCTAGGAAGATAAAAA	Forward primer for cloning out TarLBs from pUC57 for insertion into linear pSUMO using In-Fusion cloning
4	SUMOTarLBs R	TACCTAAGCTTGTCTTTAGCGCTTCAGAACTTTGTCA	Reverse primer for cloning out TarLBs from pUC57 for insertion into linear pSUMO using In-Fusion cloning
5	SUMOTarLSa F	GAACAGATTGGTGGTATGGTAAAATCAAAAATATATATTG	Forward primer for cloning out TarLSa from pUC57 for insertion into linear pSUMO using In-Fusion cloning
6	SUMOTarLSa R	TACCTAAGCTTGTCTTTAAGAGCCGAACAAGTT	Reverse primer for cloning out TarLSa from pUC57 for insertion into linear pSUMO using In-Fusion cloning
7	Linear pSUMO F	TAAAGACAAGCTTAGGTATTTATTCGG	Forward primer used for linearising pSUMO
8	Linear pSUMO R	ACCACCAATCTGTTCTCTG	Reverse primer used for linearising pSUMO
9	Linear pET28a F	ATGGCTAGCATGACTGGTGGAC	Forward primer used for linearising pET28a with a N-terminal 6x His tag
10	Linear pET28a R	ATGGCTGCCGCGCGGCAC	Reverse primer used for linearising pET28a with a N-terminal 6x His tag
11	pET28a TarKBs F	CCGCGCGGCAGCCATATGAAGACATTTCTAACTAGGATAG	Forward primer for cloning out TarKBs from pUC57 for insertion into linear pET28a (N-term tag) using In-Fusion cloning
12	pET28a TarKBs R	AGTCATGCTAGCCATTTAATCCGCGCCGCTGGT	Reverse primer for cloning out TarKBs from pUC57 for insertion into linear pET28a (N-term tag) using In-Fusion cloning
13	pET28a TarLSa F	CCGCGCGGCAGCCATATGGTAAAATCAAAAATATATTATTG	Forward primer for cloning out TarLSa from pUC57 for insertion into linear pET28a (N-term tag) using In-Fusion cloning
14	pET28a TarLSa R	AGTCATGCTAGCCATTTAAGAGCCGAACAAGTT	Reverse primer for cloning out TarLSa from pUC57 for insertion into linear pET28a (N-term tag) using In-Fusion cloning
15	pET28a TarLBs F	CCGCGCGGCAGCCATATGAAACT	Forward primer for cloning out TarLBs from pUC57 for insertion into linear pET28a (N-term tag) using In-Fusion cloning

		AGCTAGGAAGAT AAAAA	
16	pET28a TarLBs R	AGTCATGCTAGC CATTTAGCGCTT CAGAACTTTGTC A	Reverse primer for cloning out TarLBs from pUC57 for insertion into linear pET28a (N-term tag) using In-Fusion cloning
17	pET Upstream	ATGCGTCCGGC GTAGA	Forward primer for sequencing pDuet S1
18	T7 Terminator	GCTAGTTATTGC TCAGCGG	Reverse primer for sequencing pDuet S2
19	p28aCterm His lin F	CTCGAGCACCA CCACCAC	Forward primer for linearising pHis-BsTarL28a. Removes His-BsTarL and allows C-term His Tag
20	p28aCterm His lin R	TGGTATATCTCC TTCTTAAAG	Reverse primer for linearising pHis-BsTarL28a. Removes His-BsTarL and allows C-term His Tag
21	pBsTarK-His-28a F	GAAGGAGATATA CCAATGAAGACA TTTCTAACTAGG ATAG	Forward primer for cloning out BsTarK from pHis-BsTarK28a into p28aCtermHis
22	pBsTarK-His-28a R	GTGGTGGTGCT CGAGATCCGCG CCGCTGGTCAG	Reverse primer for cloning out BsTarK from pHis-BsTarK28a into p28aCtermHis
23	pBsTarL-His-28a F	GAAGGAGATATA CCAATGAAACTA GCTAGGAAGATA AAAA	Forward primer for cloning out BsTarL from pHis-BsTarL28a into p28aCtermHis
24	pBsTarL-His-28a R	GTGGTGGTGCT CGAGGCGCTTC AGAACTTTGTCA	Reverse primer for cloning out BsTarL from pHis-BsTarL28a into p28aCtermHis
25	pBvTarK-His-28a F	GAAGGAGATATA CCAATGAAGATT GCTATAACAAGG ATAG	Forward primer for cloning out BvTarK from pHis-BvTarK28a into p28aCtermHis
26	pBvTarK-His-28a R	GTGGTGGTGCT CGAGGGTACTC GGGTTCTTTTCC TG	Reverse primer for cloning out BvTarK from pHis-BvTarK28a into p28aCtermHis
27	BvTarK Mid F	GGTCCGATCTC GAAAGACAT	Forward primer for mid BvTarK gene sequencing
28	BvTarK Mid R	CTTTGGACGTCG CAAAGTGG	Reverse primer for mid BvTarK gene sequencing
29	BsTarK H119N F	CTATCCAGCTGT GGAATGCGTGC GGTGCC	Forward primer for mutagenesis of BsTarK H119 to N
30	BsTarK H119N R	GGCACCGCACG CATTCCACAGCT GGATAG	Reverse primer for mutagenesis of BsTarK H119 to N
31	BsTarK H255A F	ATTATGTGGTCA TGCTGCACCTAG CCCCGTATATGC GTAA	Forward primer for mutagenesis of BsTarK H255 to A
32	BsTarK H255A R	TTACGCATATAC GGGGCTAGGTG	Reverse primer for mutagenesis of BsTarK H255 to A

		CAGCATGACCAC ATAAT	
33	BsTarL H345N F	CTGGTGCAGCT GTGGAATGCAGT TGGTGCAT	Forward primer for mutagenesis of BsTarL H345 to N
34	BsTarL H345N R	ATGCACCAACTG CATTCCACAGCT GCACCAG	Reverse primer for mutagenesis of BsTarL H345 to N
35	BsTarL H477A F	ACGAATACATTT TCCTGTTTAAAA TCGCTCCGTTTG TTCGTAACGACG	Forward primer for mutagenesis of BsTarL H478 to A
36	BsTarL H477A R	CGTCGTTACGAA CAAACGGAGCG ATTTTAAACAGG AAAATGTATTCG T	Reverse primer for mutagenesis of BsTarL H478 to A
37	BsTarL 238-His F	TACCAATGCTTA ATACCAGTCTGT TCCACTTTG	Forward primer to remove the N-terminal domain of BsTarL
38	BsTarL 238-His R	TATTAAGCATTG GTATATCTCCTT CTTAAAGT	Reverse primer to remove the N-terminal domain of BsTarL

7.1.4 Bacterial Growth Media

All bacterial growth media was made up in Milli-Q water and autoclaved before use. Antibiotic supplemented media and agar contained 50 µg/mL kanamycin or 100 µg/mL ampicillin depending on the resistance gene present in the plasmid used.

LB Media

Tryptone (1% w/v), yeast extract (0.5% w/v), NaCl (1% w/v)

LB-agar

Tryptone (1% w/v), yeast extract (0.5% w/v), NaCl (1% w/v), agar (1.5% w/v)

2xYT media

Tryptone (1.6% w/v), yeast extract (1% w/v), NaCl (0.5% w/v)

Super Broth

Tryptone (3.5% w/v), yeast extract (2% w/v), NaCl (0.5% w/v), NaOH (5 mM)

Terrific Broth

Tryptone (1.2% w/v), yeast extract (2.4% w/v), Glycerol (0.4% v/v), K_2HPO_4 (1.3% w/v), KH_2PO_4 (0.2% w/v)

7.2 General Methods

7.2.1 In-Fusion Cloning

DNA primers were designed using the sequences of both the linearised vector and insert and the primer design tool (Takara Bio). Plasmid vectors were linearised by PCR using the manufacturer's protocol (Q5 High-Fidelity DNA Polymerase, New England Biolabs)) and the relevant forward and reverse primers (indicated in **Table 7-2**, entries 7-10, 19 and 20). Successful linearisation was confirmed by DNA agarose gel electrophoresis (see **7.2.2**). Bands of the correct length, indicative of the linearised vector, were excised from the gel and DNA was extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C. DNA concentrations were calculated from the absorbance at 260 nm measured using the DS-11 Series Spectrophotometer/Fluorometer (DeNovix).

Inserts were cloned out of plasmids, which were either made previously or ordered from GenScript, by following the manufacturer's PCR protocol and the primers proposed by the design tool (**Table 7-2**, entries 1-6, 11-16, 21-26, 37 and 38). Successful PCR reactions were confirmed by DNA agarose gel electrophoresis (see **7.2.2**). Bands of the correct length, indicative of the desired gene insert, were excised from the gel and DNA was extracted using a QIAquick gel extraction kit as per the manufacturer's protocol (QIAGEN) and stored at -20 °C. DNA concentrations were calculated from the absorbance at 260 nm measured using the DS-11 Series Spectrophotometer/Fluorometer (DeNovix).

Complementary inserts and linearised vectors were prepared with the In-Fusion HD Enzyme Premix and incubated according to the manufacturer's protocol (Takara Bio). Completed reactions were then transformed into Stellar Competent *E. coli* cells (see **7.2.5**). Colonies were selected and cultured overnight at 37 °C in LB media for DNA extraction using a QIAprep miniprep kit (QIAGEN). Isolated plasmid DNA concentrations were calculated from the absorbance at 260 nm measured using DS-11 Series Spectrophotometer/Fluorometer (DeNovix). Successful incorporation of the various inserts into the various vectors was confirmed using colony PCR and Sanger sequencing (see **7.2.3**). The correct plasmid constructs were stored at -20 °C.

7.2.2 DNA Agarose Gel Electrophoresis

Completed PCR reactions were prepared for DNA agarose gel electrophoresis by adding the correct volume of 6x Gel Loading Dye, Purple (New England Biolabs) to the samples. The samples were mixed by pipetting and then loaded onto a 0.8% w/v agarose gel submerged in

TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA). Electrophoresis then proceeded at 100 V for ~40 mins or until the dye front reached the end of the gel. Completed gels were then imaged for DNA bands using the Gbox Chemi XX6 (Syngene).

7.2.3 Colony PCR

Colonies found on LB-agar plates from cloning attempts and subsequent transformations were numbered and looped to make numbered patch LB-agar plates with the relevant antibiotic. The patch plates were incubated at 37 °C overnight. The remainder of the identical looped colonies were then resuspended in 100 µL of Milli-Q water each and boiled at 95 °C for 5 mins. 25 µL PCR reactions were then set up for each colony according to the relevant PCR reagent's protocol, with 1 µL of the boiled colony used as the template DNA. The primers used for this protocol were specific to each vector and insert which allowed amplification over the cloning site (**Table 7-2**, entries 1-6, 11-18 and 21-26). To confirm correct gene insertion, the completed reactions were loaded into a 0.8% w/v agarose gel for gel electrophoretic analysis. Colonies which indicated correct gene insertion were then looped from the numbered patch plate and used to inoculate 5 or 10 mL of LB with the relevant antibiotic for overnight growths at 37 °C, 200 RPM. A sample of each culture was stored as glycerol stock and the remainder was used for plasmid isolation using the QIAprep Spin Miniprep kit. Plasmid concentrations were determined using the DS-11 Series Spectrophotometer/Fluorometer (DeNovix) and correct formation of DNA constructs was further confirmed using Sanger DNA sequencing (Eurofins Genomics).

7.2.4 Site-Directed Mutagenesis

QuikChange site-directed mutagenesis kits (Agilent) were used as per the manufacture's protocol, with primers designed using the QuikChange Primer Design Program (**Table 7-2**, entries 29-36). Transformations of each DpnI-treated construct into XL-10 Gold *E. coli* were performed (see **7.2.5**). Colonies were cultured and plasmid was isolated as described above and stored at -20 °C. Correct mutations were confirmed by Sanger sequencing (Eurofins Genomics).

7.2.5 Transformations

Competent *E. coli* cells were aliquoted (50 µL) into 1.5 mL Eppendorf tubes with the addition of 1 µL of ~1 ng/mL of the desired plasmid and incubated on ice for 30 mins. Cells were then heat shocked at 42 °C for exactly 30 seconds. Cells were incubated on ice for 5 mins after which 950 µL of Super Optimal broth with Catabolite repression (SOC) media was added. Cultures were incubated at 37 °C with shaking for 1 hr before plating on LB-Agar plates containing the relevant antibiotic. 50 µL of the cultures were first used to plate, described as

the “dilute” plate, and the rest of the cultures were pelleted at 6,000 xg for 5 mins. The supernatants were discarded, and the cell pellets were resuspended in 50 µL of media and plated. These plates were described as the “concentrated” plate. All plates were incubated at 37 °C for 16-18 hrs.

7.2.6 Protein Concentration Determinations

Concentrations of purified proteins were estimated using the protein’s predicted extinction coefficient (AAT Bioquest) and the absorption at 280 nm. Protein 280 nm absorption readings were taken using the DS-11 Series Spectrophotometer/Fluorometer (DeNovix), using the microvolume setting, after blanking with the relevant buffer.

7.2.7 Buffer Exchange and Concentrations

Buffer exchange was achieved by either dialysis, spin concentration or via a HiPrep 26/10 desalting column using AKTA FPLC systems.

Dialysis tubing (10K molecular weight cut off (MWCO), Thermo scientific) was cut to the appropriate length and soaked in the chosen buffer for dialysis for 5 mins. One end was tied and clamped, and the sample was transferred to the tubing. The other end was tied and clamped, and the sample was transferred to a container filled with a 20x greater volume of chosen buffer. The solution was stirred at 4 °C. Dialysis proceeded for 2 hrs before transferring the sample into fresh buffer which was stirred overnight (16-18 hrs) at 4 °C. After dialysis, the sample was transferred into a clean vessel, and protein concentration measurements were taken (see **7.2.6**)

Spin concentrators (30 kDa MWCO, Vivaspin) were equilibrated with the chosen buffer by centrifugation at 4,700 xg, 4 °C for 2 mins. Protein samples were loaded and concentrated using centrifugation at 4,700 xg, 4 °C until a small volume remained. To exchange into a different buffer, the chosen buffer was added to the concentrated sample until the tube was full and concentrated again. This was repeated twice. The final buffer exchanged samples were either concentrated or diluted depending on the desired protein concentration (see **7.2.6**) and transferred into a clean vessel.

A HiPrep 26/10 desalting column (Cytiva) was equilibrated using 2 CV (106 mL) of the desired buffer using an AKTA FPLC system at a maximum flow rate of 10 mL/min. A maximum of 15 mL of the protein solution was loaded onto the column. Elution of the loaded sample proceeded for 1.5 CV (79.5 mL), at a maximum flow rate of 10 mL/min, and 2 mL fractions were collected. Fractions which contained the protein of interest were identified using a clear peak on the UV absorption trace (280 nm) and these fractions were combined into a clean

vessel. If protein samples exceeded a volume of 15 mL this method was repeated. The combined fractions were concentrated using spin concentrators or diluted to the required protein concentration (see **7.2.6**).

7.2.8 SDS-PAGE Gels

5x reducing sample buffer (10% SDS, 10 mM β-mercaptoethanol, 20% v/v glycerol, 200 mM Tris HCl, 0.05% w/v bromophenol blue, pH 6.8) was added to samples in a 1:5 (v/v) ratio. The samples were then mixed and subsequently boiled (~95 °C) for 10 mins. Prepared samples were then loaded into either self-made 12% polyacrylamide gels or gradient precast 4-15% polyacrylamide gels (Bio-Rad). Gels were submerged in running buffer (20 mM Tris HCl, 30 mM glycine, 1% w/v SDS). SDS-PAGE then proceeded at 200 V for ~55 mins or ~30 mins (or until the bands reached the end of the gel) for the self-made or precast gels, respectively. The completed gels were stained using Coomassie dye (0.1% w/v Coomassie Brilliant Blue R-250, 50% v/v ethanol, 10% v/v glacial acetic acid) and subsequently destained using destaining solution (40% v/v ethanol, 10% v/v glacial acetic acid) until protein bands were visible and could clearly be imaged using the Gbox Chemi XX6 (Syngene).

7.2.9 Western Blotting

Protein gels were run as described before (see **7.2.8**) without staining. Completed gels were rinsed with MilliQ water and placed into pre-soaked polyvinylidene fluoride (PVDF) membrane transfer stacks (Invitrogen). Fully assembled stacks were then transferred into the Power Blotter XL System (Invitrogen) for semi dry protein transfer. Transfer proceeded at 1.3 A for 7 mins. The membranes were then blocked with 5% w/v milk in TBST (Tris-buffered saline with 0.1% v/v Tween 20) on a rocker for 1hr at room temperature or overnight at 4 °C. Membranes were then drained and probed for His-tagged proteins using a solution containing HRP-anti-His antibody (Monoclonal antibody (Proteintech)) diluted 2000-fold in TBST with 1% w/v milk. Antibody binding then proceeded on a rocker for 1hr at room temperature or overnight at 4 °C. Excess antibody solution was drained off and the membranes were washed three times in TBST for 5 mins on a rocker. Amersham ECL Detection Reagent (Cytiva) was applied to the membranes and incubated for 5 mins. The membranes were immediately imaged for chemiluminescence to detect His-tagged protein bands using the Gbox Chemi XX6 (Syngene).

7.2.10 Circular Dichroism

Purified protein samples were prepared at a concentration of approximately 0.2-0.4 mg/mL in a buffer suitable for CD. A 400 µL aliquot was transferred into a quartz 1mm cuvette which was placed into a Jasco J-1500 CD Spectrophotometer (JASCO). A temperature/wavelength

scan was performed on each sample which took a full CD spectrum from 190 to 260 nm every 5 °C from 20 to 95 °C. Each scan was repeated five times. The temperature was then plotted as a function of the normalised CD spectra (mdeg) of two different wavelengths (218 nm and 222 nm) to determine the melting temperatures of the proteins.

7.2.11 Nano Differential Scanning Fluorimetry

Purified protein samples were prepared to a minimum concentration of 1 mg/mL in various buffers either before or after flash freezing and thawing. The samples were then transferred into separate Prometheus NT.48 capillaries (NanoTemper Technologies) and placed into the Prometheus nanoDSF instrument (NanoTemper Technologies). An initial scan was performed at 10% excitation power and excitation power was increased if tryptophan fluorescence signal was too weak. Fluorescence was detected at 330 nm and 350 nm along with the level of backscattering at regular temperature intervals from 20-95 °C. The fluorescence ratio of 350 nm to 330 nm was calculated and plotted as a function of temperature to resolve a protein melting temperature curve. The level of light backscattering was also plotted as a function of temperature. The melting temperature of the protein samples corresponded to the point of inflection of the resulting sigmoidal curve. The onset of aggregation temperature was recorded at the point in which backscattering started to increase.

Abbreviations

2D - Two-Dimensional

2xYT – 2× Yeast Extract Tryptone medium

3D - Three-Dimensional

ABC - ATP Binding Cassette

ATP - Adenosine Triphosphate

B. subtilis - *Bacillus subtilis*

***Bg*TarK** – *Bacillus gobiensis* TarK

BLAST – Basic Local Alignment Search Tool

***Bpa*TarL** – *Bacillus paralicheniformis* TarL

***Bpu*TarL** – *Bacillus pumilus* TarL

***Bs*TarK** - *Bacillus subtilis* W23 TarK

***Bs*TarL** - *Bacillus subtilis* W23 TarL

***Bs*TH007TarK** – *Bacillus sp. TH007* TarK

***Bv*TarK** – *Bacillus velezensis* TarK

***Bv*TarL** – *Bacillus velezensis* TarL

CAC – Cacodylate

CAMP - Cationic Antimicrobial Peptides

CD - Circular Dichroism

CDP – Cytidine diphosphate

CDP-Gro - Cytidine-5'-diphosphate-glycerol

CDP-ME - CDP-4-diphosphocytidyl-2C-methyl-D-erythritol

CDP-Rbo - Cytidine-5'-diphosphate-ribitol

CHAPS - 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

CMP - Cytidine monophosphate

Cryo-EM - Cryogenic Electron Microscopy

CTD – C-Terminal Domain

CTP - Cytidine triphosphate

CV - Column Volume

DTT – Dithiothreitol

E. coli - *Escherichia coli*

EI - Extracted Ion

FPLC - Fast Protein Liquid Chromatography

GFP - Green Fluorescent Protein

GlcNAc - N-acetylglucosamine

GPT - GlcNAc phosphotransferase

GroP - Glycerol 3-phosphate

GT-B – Glycosyltransferase B

H. influenzae - *Haemophilus influenzae*

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His6 – Hex-histidine tag

HPLC - High-Performance Liquid Chromatography

Hrs – Hours

IC_{50} - Half-maximal Inhibitory Concentration

IgG - Immunoglobulin G

IMAC - Immobilised Metal Affinity Chromatography

IPTG - Isopropyl β-D-1-thiogalactopyranoside

IspD - 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase

ITC - Isothermal Titration Calorimetry

LB - Luria-Bertani

LC-MS - Liquid Chromatography-Mass Spectrometry

LCP - LytR-CpsA-Psr

LTA - Lipoteichoic Acid

m/z - Mass-to-charge Ratio

ManNAc - N-acetylmannosamine

MES - 2-(N-morpholino)ethanesulfonic acid

Mins – Minutes

MRSA - Methicillin-Resistant *Staphylococcus aureus*

MS - Mass Spectrometry

MWCO - Molecular Weight Cut Off

nanoDSF - Nano Differential Scanning Fluorimetry

NMR - Nuclear Magnetic Resonance

NTD – N-Terminal Domain

O/N – Overnight

OD_{600} - Optical density at 600 nm

PAGE - Polyacrylamide Gel Electrophoresis

PCR - Polymerase Chain Reaction

PDB - Protein Data Bank

PEG - Polyethylene Glycol

PG - Peptidoglycan

P-GlcNAc - N-acetylglucosamine 1-phosphate

PP - Pyrophosphate

pp-lipid - Pyrophosphate-lipid

PVDF - Polyvinylidene Fluoride

RboP - Ribitol-5-phosphate

RibP - Ribulose-5-phosphate

RPM - Revolutions Per Minute

RT-PCR - Reverse Transcriptase-Polymerase Chain Reaction

S. aureus - *Staphylococcus aureus*

***Sa*TarL** - *Staphylococcus aureus* TarL

SB - Superior Broth

SDS-PAGE – Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

SEC - Size Exclusion Chromatography

SEC-MALS - Size Exclusion Chromatography with Multi-Angle Light Scattering

Secs – Seconds

***Se*TagF** – *Staphylococcus epidermidis* TagF

SOC - Super Optimal broth with Catabolite repression media

SUMO - Small Ubiquitin-like Modifier

TB - Terrific Broth

TBST - Tris-Buffered Saline with 0.1% v/v Tween 20

Tm - Melting Temperature

Tris - Tris(hydroxymethyl)aminomethane

UDP - Uridine diphosphate

UDP-GlcNAc - Uridine Diphosphate N-acetylglucosamine

UMP - Uridine monophosphate

UV – Ultraviolet

WT - Wild Type

WTA - Wall Teichoic Acid

xg - Centrifugal force

β-GlcNAc - β-linked *N*-acetylglucosamine

References

Aasjord, P. and Grov, A. (1980) 'Immunoperoxidase and electron microscopy studies of staphylococcal lipoteichoic acid', *Acta Pathologica Et Microbiologica Scandinavica. Section B, Microbiology*, 88(1), pp. 47–52. Available at: https://doi.org/10.1111/j.1699-0463.1980.tb02602.x.

Allison, S.E. *et al.* (2011) 'Studies of the genetics, function, and kinetic mechanism of TagE, the wall teichoic acid glycosyltransferase in Bacillus subtilis 168', *The Journal of Biological Chemistry*, 286(27), pp. 23708–23716. Available at: https://doi.org/10.1074/jbc.M111.241265.

An, J.-H. *et al.* (2013) 'Human SAP is a novel peptidoglycan recognition protein that induces complement- independent phagocytosis of Staphylococcus aureus', *Journal of immunology (Baltimore, Md. : 1950)*, 191(6), pp. 3319–3327. Available at: https://doi.org/10.4049/jimmunol.1300940.

Anderson, A.J. *et al.* (1978) 'Cell wall assembly in Bacillus subtilis: location of wall material incorporated during pulsed release of phosphate limitation, its accessibility to bacteriophages and concanavalin A, and its susceptibility to turnover', *Journal of Bacteriology*, 136(3), pp. 886–899. Available at: https://doi.org/10.1128/jb.136.3.886-899.1978.

Andre, G. *et al.* (2011) 'Fluorescence and atomic force microscopy imaging of wall teichoic acids in Lactobacillus plantarum', *ACS chemical biology*, 6(4), pp. 366–376. Available at: https://doi.org/10.1021/cb1003509.

Araki, Y. and Ito, E. (1989) 'Linkage Units in Cell Walls of Gram-Positive Bacteria', *Critical Reviews in Microbiology*, 17(2), pp. 121–135. Available at: https://doi.org/10.3109/10408418909105745.

Armstrong, J.J. *et al.* (1958) 'Isolation and structure of ribitol phosphate derivatives (teichoic acids) from bacterial cell walls', *Journal of the Chemical Society (Resumed)*, (0), pp. 4344–4354. Available at: https://doi.org/10.1039/JR9580004344.

Atilano, M.L. *et al.* (2010) 'Teichoic acids are temporal and spatial regulators of peptidoglycan cross-linking in Staphylococcus aureus', *Proceedings of the National Academy of Sciences of the United States of America*, 107(44), pp. 18991–18996. Available at: https://doi.org/10.1073/pnas.1004304107.

Baptista, C., Santos, M.A. and São-José, C. (2008) 'Phage SPP1 reversible adsorption to Bacillus subtilis cell wall teichoic acids accelerates virus recognition of membrane receptor YueB', *Journal of Bacteriology*, 190(14), pp. 4989–4996. Available at: https://doi.org/10.1128/JB.00349-08.

Baur, S. *et al.* (2009) 'Synthesis of CDP-activated ribitol for teichoic acid precursors in Streptococcus pneumoniae', *Journal of Bacteriology*, 191(4), pp. 1200–1210. Available at: https://doi.org/10.1128/JB.01120-08.

Baur, S. *et al.* (2014) 'A Nasal Epithelial Receptor for Staphylococcus aureus WTA Governs Adhesion to Epithelial Cells and Modulates Nasal Colonization', *PLOS Pathogens*, 10(5), p. e1004089. Available at: https://doi.org/10.1371/journal.ppat.1004089.

Bepler, T. *et al.* (2019) 'Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs', *Nature Methods*, 16(11), pp. 1153–1160. Available at: https://doi.org/10.1038/s41592-019-0575-8.

Bera, A. *et al.* (2007) 'Influence of Wall Teichoic Acid on Lysozyme Resistance in Staphylococcus aureus', *Journal of Bacteriology*, 189(1), pp. 280–283. Available at: https://doi.org/10.1128/JB.01221-06.

Berman, H.M. *et al.* (2000) 'The Protein Data Bank', *Nucleic Acids Research*, 28(1), pp. 235–242. Available at: https://doi.org/10.1093/nar/28.1.235.

Bertsche, U. *et al.* (2013) 'Increased cell wall teichoic acid production and D-alanylation are common phenotypes among daptomycin-resistant methicillin-resistant Staphylococcus aureus (MRSA) clinical isolates', *PloS One*, 8(6), p. e67398. Available at: https://doi.org/10.1371/journal.pone.0067398.

Bhavsar, A.P. *et al.* (2004) 'Teichoic acid is an essential polymer in Bacillus subtilis that is functionally distinct from teichuronic acid', *Journal of Bacteriology*, 186(23), pp. 7865–7873. Available at: https://doi.org/10.1128/JB.186.23.7865-7873.2004.

Bhavsar, A.P., Beveridge, T.J. and Brown, E.D. (2001) 'Precise Deletion of tagD and Controlled Depletion of Its Product, Glycerol 3-Phosphate Cytidylyltransferase, Leads to Irregular Morphology and Lysis of Bacillus subtilis Grown at Physiological Temperature', *Journal of Bacteriology*, 183(22), pp. 6688–6693. Available at: https://doi.org/10.1128/JB.183.22.6688-6693.2001.

Bhavsar, A.P., Truant, R. and Brown, E.D. (2005) 'The TagB protein in Bacillus subtilis 168 is an intracellular peripheral membrane protein that can incorporate glycerol phosphate onto

a membrane-bound acceptor in vitro', *The Journal of Biological Chemistry*, 280(44), pp. 36691–36700. Available at: https://doi.org/10.1074/jbc.M507154200.

Biswas, R. *et al.* (2012) 'Proton-Binding Capacity of Staphylococcus aureus Wall Teichoic Acid and Its Role in Controlling Autolysin Activity', *PLoS ONE*, 7(7), p. e41415. Available at: https://doi.org/10.1371/journal.pone.0041415.

Blattner, F.R. *et al.* (1997) 'The Complete Genome Sequence of Escherichia coli K-12', *Science*, 277(5331), pp. 1453–1462. Available at: https://doi.org/10.1126/science.277.5331.1453.

Bouhss, A. *et al.* (2008) 'The biosynthesis of peptidoglycan lipid-linked intermediates', *FEMS Microbiology Reviews*, 32(2), pp. 208–233. Available at: https://doi.org/10.1111/j.1574-6976.2007.00089.x.

Boylan, R.J. *et al.* (1972) 'Regulation of the bacterial cell wall: analysis of a mutant of Bacillus subtilis defective in biosynthesis of teichoic acid', *Journal of Bacteriology*, 110(1), pp. 281–290. Available at: https://doi.org/10.1128/jb.110.1.281-290.1972.

Brown, S. *et al.* (2010) 'Staphylococcus aureus and Bacillus subtilis W23 Make Polyribitol Wall Teichoic Acids Using Different Enzymatic Pathways', *Chemistry & Biology*, 17(10), pp. 1101–1110. Available at: https://doi.org/10.1016/j.chembiol.2010.07.017.

Brown, S. *et al.* (2012a) 'Methicillin resistance in Staphylococcus aureus requires glycosylated wall teichoic acids', *Proceedings of the National Academy of Sciences of the United States of America*, 109(46), pp. 18909–18914. Available at: https://doi.org/10.1073/pnas.1209126109.

Brown, S., Santa Maria, J.P. and Walker, S. (2013a) 'Wall Teichoic Acids of Gram-Positive Bacteria', *Annual Review of Microbiology*, 67(1), pp. 313–336. Available at: https://doi.org/10.1146/annurev-micro-092412-155620.

Brown, S., Zhang, Y.-H. and Walker, S. (2008) 'A revised pathway proposed for Staphylococcus aureus wall teichoic acid biosynthesis based on in vitro reconstitution of the intracellular steps', *Chemistry & Biology*, 15(1), pp. 12–21. Available at: https://doi.org/10.1016/j.chembiol.2007.11.011.

Butt, T.R. *et al.* (2005) 'SUMO fusion technology for difficult-to-express proteins', *Protein Expression and Purification*, 43(1), pp. 1–9. Available at: https://doi.org/10.1016/j.pep.2005.03.016.

Campbell, J. *et al.* (2011) 'Synthetic lethal compound combinations reveal a fundamental connection between wall teichoic acid and peptidoglycan biosyntheses in Staphylococcus aureus', *ACS chemical biology*, 6(1), pp. 106–116. Available at: https://doi.org/10.1021/cb100269f.

Campbell, J. *et al.* (2012) 'An antibiotic that inhibits a late step in wall teichoic acid biosynthesis induces the cell wall stress stimulon in Staphylococcus aureus', *Antimicrobial Agents and Chemotherapy*, 56(4), pp. 1810–1820. Available at: https://doi.org/10.1128/AAC.05938-11.

Chin, T., Burger, M.M. and Glaser, L. (1966) 'Synthesis of teichoic acids. VI. The formation of multiple wall polymers in Bacillus subtilis W-23', *Archives of Biochemistry and Biophysics*, 116(1), pp. 358–367. Available at: https://doi.org/10.1016/0003-9861(66)90042-7.

Cifuente, J.O. *et al.* (2023) 'A multi-enzyme machine polymerizes the Haemophilus influenzae type b capsule', *Nature Chemical Biology*, 19(7), pp. 865–877. Available at: https://doi.org/10.1038/s41589-023-01324-3.

Collins, L.V. *et al.* (2002) 'Staphylococcus aureus strains lacking D-alanine modifications of teichoic acids are highly susceptible to human neutrophil killing and are virulence attenuated in mice', *The Journal of Infectious Diseases*, 186(2), pp. 214–219. Available at: https://doi.org/10.1086/341454.

Debabov, D.V. *et al.* (1996) 'The D-Alanyl carrier protein in Lactobacillus casei: cloning, sequencing, and expression of dltC', *Journal of Bacteriology*, 178(13), pp. 3869–3876. Available at: https://doi.org/10.1128/jb.178.13.3869-3876.1996.

Debabov, D.V., Kiriukhin, M.Y. and Neuhaus, F.C. (2000) 'Biosynthesis of lipoteichoic acid in Lactobacillus rhamnosus: role of DltD in D-alanylation', *Journal of Bacteriology*, 182(10), pp. 2855–2864. Available at: https://doi.org/10.1128/JB.182.10.2855-2864.2000.

D'Elia, M.A., Pereira, M.P., *et al.* (2006) 'Lesions in teichoic acid biosynthesis in Staphylococcus aureus lead to a lethal gain of function in the otherwise dispensable pathway', *Journal of Bacteriology*, 188(12), pp. 4183–4189. Available at: https://doi.org/10.1128/JB.00197-06.

D'Elia, M.A., Millar, K.E., *et al.* (2006) 'Wall teichoic acid polymers are dispensable for cell viability in Bacillus subtilis', *Journal of Bacteriology*, 188(23), pp. 8313–8316. Available at: https://doi.org/10.1128/JB.01336-06.

Dengler, V. *et al.* (2012) 'Deletion of hypothetical wall teichoic acid ligases in Staphylococcus aureus activates the cell wall stress response', *FEMS microbiology letters*, 333(2), pp. 109–120. Available at: https://doi.org/10.1111/j.1574-6968.2012.02603.x.

Endl, J. *et al.* (1983) 'Chemical composition and structure of cell wall teichoic acids of staphylococci', *Archives of Microbiology*, 135(3), pp. 215–223. Available at: https://doi.org/10.1007/BF00414483.

Endl, J. *et al.* (1984) 'Determination of cell wall teichoic acid structure of staphylococci by rapid chemical and serological screening methods', *Archives of Microbiology*, 137(3), pp. 272–280. Available at: https://doi.org/10.1007/BF00414557.

Farha, M.A. *et al.* (2013) 'Inhibition of WTA synthesis blocks the cooperative action of PBPs and sensitizes MRSA to β-lactams', *ACS chemical biology*, 8(1), pp. 226–233. Available at: https://doi.org/10.1021/cb300413m.

Fischer, J. (1995) 'Specific detection of nucleotides, creatine phosphate, and their derivatives from tissue samples in a simple, isocratic, recycling, low-volume system', *LC-GC International*, 8, pp. 254–264.

Formstone, A. *et al.* (2008) 'Localization and Interactions of Teichoic Acid Synthetic Enzymes in Bacillus subtilis', *Journal of Bacteriology*, 190(5), pp. 1812–1821. Available at: https://doi.org/10.1128/jb.01394-07.

Gasteiger, E. *et al.* (2005) 'Protein Identification and Analysis Tools on the ExPASy Server', in J.M. Walker (ed.) *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607. Available at: https://doi.org/10.1385/1-59259-890-0:571.

Gerlach, D. *et al.* (2018) 'Methicillin-resistant Staphylococcus aureus alters cell wall glycosylation to evade immunity', *Nature*, 563(7733), pp. 705–709. Available at: https://doi.org/10.1038/s41586-018-0730-x.

Ginsberg, C. *et al.* (2006) 'In Vitro Reconstitution of Two Essential Steps in Wall Teichoic Acid Biosynthesis', *ACS Chemical Biology*, 1(1), pp. 25–28. Available at: https://doi.org/10.1021/cb0500041.

Gross, M. *et al.* (2001) 'Key role of teichoic acid net charge in Staphylococcus aureus colonization of artificial surfaces', *Infection and Immunity*, 69(5), pp. 3423–3426. Available at: https://doi.org/10.1128/IAI.69.5.3423-3426.2001.

Guo, Y. *et al.* (2021) 'Cell wall glycosylation in Staphylococcus aureus: targeting the tar glycosyltransferases', *Current Opinion in Structural Biology*, 68, pp. 166–174. Available at: https://doi.org/10.1016/j.sbi.2021.01.003.

Hancock, I.C., Wiseman, G. and Baddiley, J. (1976) 'Biosynthesis of the unit that links teichoic acid to the bacterial wall: inhibition by tunicamycin', *FEBS letters*, 69(1), pp. 75–80. Available at: https://doi.org/10.1016/0014-5793(76)80657-6.

Heaton, M.P. and Neuhaus, F.C. (1994) 'Role of the D-alanyl carrier protein in the biosynthesis of D-alanyl-lipoteichoic acid', *Journal of Bacteriology*, 176(3), pp. 681–690. Available at: https://doi.org/10.1128/jb.176.3.681-690.1994.

Herzik, M.A., Wu, M. and Lander, G.C. (2019) 'High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM', *Nature Communications*, 10, p. 1032. Available at: https://doi.org/10.1038/s41467-019-08991-8.

Holland, L.M., Conlon, B. and O'Gara, J.P. (2011) 'Mutation of tagO reveals an essential role for wall teichoic acids in Staphylococcus epidermidis biofilm development', *Microbiology (Reading, England)*, 157(Pt 2), pp. 408–418. Available at: https://doi.org/10.1099/mic.0.042234-0.

Jenni, R. and Berger-Bächi, B. (1998) 'Teichoic acid content in different lineages of Staphylococcus aureus NCTC8325', *Archives of Microbiology*, 170(3), pp. 171–178. Available at: https://doi.org/10.1007/s002030050630.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Karamata, D., Pooley, H.M. and Monod, M. (1987) 'Expression of heterologous genes for wall teichoic acid in Bacillus subtilis 168', *Molecular and General Genetics MGG*, 207(1), pp. 73–81. Available at: https://doi.org/10.1007/BF00331493.

Kawai, Y. *et al.* (2011) 'A widespread family of bacterial cell wall assembly proteins', *The EMBO journal*, 30(24), pp. 4931–4941. Available at: https://doi.org/10.1038/emboj.2011.358.

Kermani, A.A. (2021) 'A guide to membrane protein X-ray crystallography', *The FEBS Journal*, 288(20), pp. 5788–5804. Available at: https://doi.org/10.1111/febs.15676.

Kern, T. *et al.* (2010) 'Dynamics Characterization of Fully Hydrated Bacterial Cell Walls by Solid-State NMR: Evidence for Cooperative Binding of Metal Ions', *Journal of the American Chemical Society*, 132(31), pp. 10911–10919. Available at: https://doi.org/10.1021/ja104533w.

Kobayashi, K. *et al.* (2003) 'Essential Bacillus subtilis genes', *Proceedings of the National Academy of Sciences of the United States of America*, 100(8), pp. 4678–4683. Available at: https://doi.org/10.1073/pnas.0730515100.

Kohler, T., Weidenmaier, C. and Peschel, A. (2009) 'Wall teichoic acid protects Staphylococcus aureus against antimicrobial fatty acids from human skin', *Journal of Bacteriology*, 191(13), pp. 4482–4484. Available at: https://doi.org/10.1128/JB.00221-09.

Kojima, N., Araki, Y. and Ito, E. (1985) 'Structure of the linkage units between ribitol teichoic acids and peptidoglycan.', *Journal of Bacteriology*, 161(1), pp. 299–306.

Koprivnjak, T. *et al.* (2006) 'Cation-Induced Transcriptional Regulation of the dlt Operon of Staphylococcus aureus', *Journal of Bacteriology*, 188(10), pp. 3622–3630. Available at: https://doi.org/10.1128/jb.188.10.3622-3630.2006.

Kunfermann, A. *et al.* (2014) 'Pseudilins: Halogenated, Allosteric Inhibitors of the Non-Mevalonate Pathway Enzyme IspD', *Angewandte Chemie International Edition*, 53(8), pp. 2235–2239. Available at: https://doi.org/10.1002/anie.201309557.

Lazarevic, V. *et al.* (2002) 'Comparison of ribitol and glycerol teichoic acid genes in Bacillus subtilis W23 and 168: identical function, similar divergent organization, but different regulationThe EMBL accession numbers for the nucleotide sequences reported in this paper are AJ313428, AJ318465, AJ318466, AJ318467, AJ318468, AJ318469 and AJ318470.', *Microbiology*, 148(3), pp. 815–824. Available at: https://doi.org/10.1099/00221287-148-3-815.

Lazarevic, V. and Karamata, D. (1995) 'The tagGH operon of Bacillus subtilis 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids', *Molecular Microbiology*, 16(2), pp. 345–355. Available at: https://doi.org/10.1111/j.1365-2958.1995.tb02306.x.

Lee, S.H. *et al.* (2016) 'TarO-specific inhibitors of wall teichoic acid biosynthesis restore β-lactam efficacy against methicillin-resistant staphylococci', *Science Translational Medicine*, 8(329), p. 329ra32. Available at: https://doi.org/10.1126/scitranslmed.aad7364.

Li, F.K.K. *et al.* (2021a) 'Crystallographic analysis of TarI and TarJ, a cytidylyltransferase and reductase pair for CDP-ribitol synthesis in *Staphylococcus aureus* wall teichoic acid biogenesis', *Journal of Structural Biology*, 213(2), p. 107733. Available at: https://doi.org/10.1016/j.jsb.2021.107733.

Li, F.K.K. *et al.* (2024) 'Cryo-EM analysis of S. aureus TarL, a polymerase in wall teichoic acid biogenesis central to virulence and antibiotic resistance', *Science Advances*, 10(9), p. eadj3864. Available at: https://doi.org/10.1126/sciadv.adj3864.

Lovering, A.L. *et al.* (2010) 'Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis', *Nature Structural & Molecular Biology*, 17(5), pp. 582–589. Available at: https://doi.org/10.1038/nsmb.1819.

Ma, D. *et al.* (2018) 'Crystal structure of a membrane-bound O-acyltransferase', *Nature*, 562(7726), pp. 286–290. Available at: https://doi.org/10.1038/s41586-018-0568-2.

Madeira, F. *et al.* (2022) 'Search and sequence analysis tools services from EMBL-EBI in 2022', *Nucleic acids research*, p. gkac240. Available at: https://doi.org/10.1093/nar/gkac240.

Matias, V.R.F. and Beveridge, T.J. (2005) 'Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space', *Molecular Microbiology*, 56(1), pp. 240–251. Available at: https://doi.org/10.1111/j.1365-2958.2005.04535.x.

Matias, V.R.F. and Beveridge, T.J. (2007) 'Cryo-electron microscopy of cell division in Staphylococcus aureus reveals a mid-zone between nascent cross walls', *Molecular Microbiology*, 64(1), pp. 195–206. Available at: https://doi.org/10.1111/j.1365-2958.2007.05634.x.

Meredith, T.C., Swoboda, J.G. and Walker, S. (2008) 'Late-Stage Polyribitol Phosphate Wall Teichoic Acid Biosynthesis in Staphylococcus aureus', *Journal of Bacteriology*, 190(8), p. 3046. Available at: https://doi.org/10.1128/JB.01880-07.

Merk, A. *et al.* (2016) 'Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery', *Cell*, 165(7), pp. 1698–1707. Available at: https://doi.org/10.1016/j.cell.2016.05.040.

Misawa, Y. *et al.* (2015) 'Staphylococcus aureus Colonization of the Mouse Gastrointestinal Tract Is Modulated by Wall Teichoic Acid, Capsule, and Surface Proteins', *PLoS pathogens*, 11(7), p. e1005061. Available at: https://doi.org/10.1371/journal.ppat.1005061.

Murazumi, N. *et al.* (1988) 'Partial purification and properties of UDP-N-acetylmannosamine:N-acetylglucosaminyl pyrophosphorylundecaprenol N-acetylmannosaminyltransferase from Bacillus subtilis', *Journal of Biochemistry*, 104(6), pp. 980–984. Available at: https://doi.org/10.1093/oxfordjournals.jbchem.a122594.

Nathenson, S.G. *et al.* (1966) 'Enzymatic synthesis and immunochemistry of alpha- and beta-N-acetylglucosaminylribitol linkages in teichoic acids from several strains of Staphylococcus aureus', *The Journal of Biological Chemistry*, 241(3), pp. 651–658.

Neuhaus, F.C. and Baddiley, J. (2003) 'A Continuum of Anionic Charge: Structures and Functions of d-Alanyl-Teichoic Acids in Gram-Positive Bacteria', *Microbiology and Molecular Biology Reviews : MMBR*, 67(4), pp. 686–723. Available at: https://doi.org/10.1128/MMBR.67.4.686-723.2003.

Niebling, S. *et al.* (2022) 'Biophysical Screening Pipeline for Cryo-EM Grid Preparation of Membrane Proteins', *Frontiers in Molecular Biosciences*, 9, p. 882288. Available at: https://doi.org/10.3389/fmolb.2022.882288.

Olerinyova, A. *et al.* (2021) 'Mass Photometry of Membrane Proteins', *Chem*, 7(1), pp. 224–236. Available at: https://doi.org/10.1016/j.chempr.2020.11.011.

Over, B. *et al.* (2011) 'LytR-CpsA-Psr proteins in Staphylococcus aureus display partial functional redundancy and the deletion of all three severely impairs septum placement and cell separation', *FEMS microbiology letters*, 320(2), pp. 142–151. Available at: https://doi.org/10.1111/j.1574-6968.2011.02303.x.

Park, Y.S. *et al.* (1993) 'Expression, purification, and characterization of CTP:glycerol-3-phosphate cytidylyltransferase from Bacillus subtilis', *The Journal of Biological Chemistry*, 268(22), pp. 16648–16654.

Perego, M. *et al.* (1995) 'Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in Bacillus subtilis. Identification of genes and regulation', *The Journal of Biological Chemistry*, 270(26), pp. 15598–15606. Available at: https://doi.org/10.1074/jbc.270.26.15598.

Pereira, M.P., D'Elia, M.A., *et al.* (2008) 'Duplication of Teichoic Acid Biosynthetic Genes in Staphylococcus aureus Leads to Functionally Redundant Poly(Ribitol Phosphate) Polymerases', *Journal of Bacteriology*, 190(16), pp. 5642–5649. Available at: https://doi.org/10.1128/JB.00526-08.

Pereira, M.P., Schertzer, J.W., *et al.* (2008) 'The Wall Teichoic Acid Polymerase TagF Efficiently Synthesizes Poly(glycerol phosphate) on the TagB Product Lipid III', *ChemBioChem*, 9(9), pp. 1385–1390. Available at: https://doi.org/10.1002/cbic.200800026.

Pereira, M.P. and Brown, E.D. (2004) 'Bifunctional Catalysis by CDP-ribitol Synthase: Convergent Recruitment of Reductase and Cytidylyltransferase Activities in Haemophilus

influenzae and Staphylococcus aureus', *Biochemistry*, 43(37), pp. 11802–11812. Available at: https://doi.org/10.1021/bi048866v.

Peschel, A. *et al.* (1999) 'Inactivation of the dlt operon in Staphylococcus aureus confers sensitivity to defensins, protegrins, and other antimicrobial peptides', *The Journal of Biological Chemistry*, 274(13), pp. 8405–8410. Available at: https://doi.org/10.1074/jbc.274.13.8405.

Peschel, A. *et al.* (2000) 'The D-alanine residues of Staphylococcus aureus teichoic acids alter the susceptibility to vancomycin and the activity of autolytic enzymes', *Antimicrobial Agents and Chemotherapy*, 44(10), pp. 2845–2847. Available at: https://doi.org/10.1128/AAC.44.10.2845-2847.2000.

Pettersen, E.F. *et al.* (2004) 'UCSF Chimera—A visualization system for exploratory research and analysis', *Journal of Computational Chemistry*, 25(13), pp. 1605–1612. Available at: https://doi.org/10.1002/jcc.20084.

Pollard, T.D. (2010) 'A Guide to Simple and Informative Binding Assays', *Molecular Biology of the Cell*, 21(23), pp. 4061–4067. Available at: https://doi.org/10.1091/mbc.e10-08-0683.

Price, N.P.J. and Tsvetanova, B. (2007) 'Biosynthesis of the tunicamycins: a review', *The Journal of Antibiotics*, 60(8), pp. 485–491. Available at: https://doi.org/10.1038/ja.2007.62.

Qamar, A. and Golemi-Kotra, D. (2012) 'Dual roles of FmtA in Staphylococcus aureus cell wall biosynthesis and autolysis', *Antimicrobial Agents and Chemotherapy*, 56(7), pp. 3797–3805. Available at: https://doi.org/10.1128/AAC.00187-12.

Qian, Z. *et al.* (2006) 'Genomic characterization of ribitol teichoic acid synthesis in Staphylococcus aureus: genes, genomic organization and gene duplication', *BMC Genomics*, 7(1), p. 74. Available at: https://doi.org/10.1186/1471-2164-7-74.

Qin, X. *et al.* (2010) 'Concentration-Dependent Aggregation of CHAPS Investigated by NMR Spectroscopy', *The Journal of Physical Chemistry B*, 114(11), pp. 3863–3868. Available at: https://doi.org/10.1021/jp911720w.

Rahman, M.M. *et al.* (2016) 'The Staphylococcus aureus Methicillin Resistance Factor FmtA Is a d-Amino Esterase That Acts on Teichoic Acids', *mBio*, 7(1), pp. e02070-15. Available at: https://doi.org/10.1128/mBio.02070-15.

Reichmann, N.T., Cassona, C.P. and Gründling, A. (2013) 'Revised mechanism of d-alanine incorporation into cell wall polymers in Gram-positive bacteria', *Microbiology*, 159(Pt 9), pp. 1868–1877. Available at: https://doi.org/10.1099/mic.0.069898-0.

Robert, X. and Gouet, P. (2014) 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42(W1), pp. W320–W324. Available at: https://doi.org/10.1093/nar/gku316.

Scheres, S.H.W. (2012) 'RELION: Implementation of a Bayesian approach to cryo-EM structure determination', *Journal of Structural Biology*, 180(3), pp. 519–530. Available at: https://doi.org/10.1016/j.jsb.2012.09.006.

Schirner, K., Stone, L.K. and Walker, S. (2011) 'ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers', *ACS chemical biology*, 6(5), pp. 407–412. Available at: https://doi.org/10.1021/cb100390w.

Schlag, M. *et al.* (2010) 'Role of staphylococcal wall teichoic acid in targeting the major autolysin Atl', *Molecular Microbiology*, 75(4), pp. 864–873. Available at: https://doi.org/10.1111/j.1365-2958.2009.07007.x.

Schwab, A. *et al.* (2017) 'Mechanism of Allosteric Inhibition of the Enzyme IspD by Three Different Classes of Ligands', *ACS Chemical Biology*, 12(8), pp. 2132–2138. Available at: https://doi.org/10.1021/acschembio.7b00004.

Seo, A. *et al.* (2013) 'Using UV-absorbance of intrinsic dithiothreitol (DTT) during RP-HPLC as a measure of experimental redox potential in vitro', *Analytical and bioanalytical chemistry*, 405(19), pp. 6379–6384. Available at: https://doi.org/10.1007/s00216-013-7063-2.

Sobhanifar, S. *et al.* (2016) 'Structure and Mechanism of Staphylococcus aureus TarS, the Wall Teichoic Acid β-glycosyltransferase Involved in Methicillin Resistance', *PLoS Pathogens*, 12(12), p. e1006067. Available at: https://doi.org/10.1371/journal.ppat.1006067.

Soldo, B., Lazarevic, V. and Karamata, D. (2002) 'tagO is involved in the synthesis of all anionic cell-wall polymers in Bacillus subtilis 168', *Microbiology (Reading, England)*, 148(Pt 7), pp. 2079–2087. Available at: https://doi.org/10.1099/00221287-148-7-2079.

Swoboda, J.G. *et al.* (2009) 'Discovery of a small molecule that blocks wall teichoic acid biosynthesis in Staphylococcus aureus', *ACS chemical biology*, 4(10), pp. 875–883. Available at: https://doi.org/10.1021/cb900151k.

Swoboda, J.G. *et al.* (2010) 'Wall Teichoic Acid Function, Biosynthesis, and Inhibition', *Chembiochem : a European journal of chemical biology*, 11(1), pp. 35–45. Available at: https://doi.org/10.1002/cbic.200900557.

Vadyvaloo, V. *et al.* (2004) 'Cell-surface alterations in class IIa bacteriocin-resistant Listeria monocytogenes strains', *Microbiology (Reading, England)*, 150(Pt 9), pp. 3025–3033. Available at: https://doi.org/10.1099/mic.0.27059-0.

Vinogradov, E. *et al.* (2006) 'Structural elucidation of the extracellular and cell-wall teichoic acids of Staphylococcus aureus MN8m, a biofilm forming strain', *Carbohydrate Research*, 341(6), pp. 738–743. Available at: https://doi.org/10.1016/j.carres.2006.01.012.

Walter, J. *et al.* (2007) 'D-alanyl ester depletion of teichoic acids in Lactobacillus reuteri 100-23 results in impaired colonization of the mouse gastrointestinal tract', *Environmental Microbiology*, 9(7), pp. 1750–1760. Available at: https://doi.org/10.1111/j.1462-2920.2007.01292.x.

Wang, H. *et al.* (2013) 'Discovery of Novel Wall Teichoic Acid Inhibitors as Effective anti-MRSA β-lactam Combination Agents', *Chemistry & biology*, 20(2), pp. 272–284. Available at: https://doi.org/10.1016/j.chembiol.2012.11.013.

Ward, J.B. (1981) 'Teichoic and teichuronic acids: biosynthesis, assembly, and location', *Microbiological Reviews*, 45(2), pp. 211–243. Available at: https://doi.org/10.1128/mr.45.2.211-243.1981.

Weidenmaier, C. *et al.* (2004) 'Role of teichoic acids in Staphylococcus aureus nasal colonization, a major risk factor in nosocomial infections', *Nature Medicine*, 10(3), pp. 243–245. Available at: https://doi.org/10.1038/nm991.

Wickham, J.R. *et al.* (2009) 'Revisiting magnesium chelation by teichoic acid with phosphorus solid-state NMR and theoretical calculations', *The Journal of Physical Chemistry. B*, 113(7), pp. 2177–2183. Available at: https://doi.org/10.1021/jp809313j.

Winstel, V. *et al.* (2015) 'Wall Teichoic Acid Glycosylation Governs Staphylococcus aureus Nasal Colonization', *mBio*, 6(4), p. 10.1128/mbio.00632-15. Available at: https://doi.org/10.1128/mbio.00632-15.

Witschel, M.C. *et al.* (2011) 'Inhibitors of the Herbicidal Target IspD: Allosteric Site Binding', *Angewandte Chemie International Edition*, 50(34), pp. 7931–7935. Available at: https://doi.org/10.1002/anie.201102281.

Wu, X. *et al.* (2019) 'Reducing Staphylococcus aureus resistance to lysostaphin using CRISPR-dCas9', *Biotechnology and Bioengineering*, 116(12), pp. 3149–3159. Available at: https://doi.org/10.1002/bit.27143.

Wu, X. *et al.* (2021) 'Wall teichoic acids: physiology and applications', *FEMS Microbiology Reviews*, 45(4), p. fuaa064. Available at: https://doi.org/10.1093/femsre/fuaa064.

Xia, G. *et al.* (2010) 'Glycosylation of Wall Teichoic Acid in Staphylococcus aureus by TarM', *The Journal of Biological Chemistry*, 285(18), pp. 13405–13415. Available at: https://doi.org/10.1074/jbc.M109.096172.

Xia, G. *et al.* (2011) 'Wall teichoic Acid-dependent adsorption of staphylococcal siphovirus and myovirus', *Journal of Bacteriology*, 193(15), pp. 4006–4009. Available at: https://doi.org/10.1128/JB.01412-10.

Yang, C.-S. *et al.* (2021) 'Crystal structure of the N-terminal domain of TagH reveals a potential drug targeting site', *Biochemical and Biophysical Research Communications*, 536, pp. 1–6. Available at: https://doi.org/10.1016/j.bbrc.2020.12.028.

Yokoyama, K. *et al.* (1986) 'Structure and functions of linkage unit intermediates in the biosynthesis of ribitol teichoic acids in Staphylococcus aureus H and Bacillus subtilis W23', *European Journal of Biochemistry*, 161(2), pp. 479–489. Available at: https://doi.org/10.1111/j.1432-1033.1986.tb10469.x.

Young, M. *et al.* (1989) 'Pseudo-allelic relationship between non-homologous genes concerned with biosynthesis of polyglycerol phosphate and polyribitol phosphate teichoic acids in Bacillus subtilis strains 168 and W23', *Molecular Microbiology*, 3(12), pp. 1805–1812. Available at: https://doi.org/10.1111/j.1365-2958.1989.tb00166.x.

Zhang, Y.-H. *et al.* (2006) 'Acceptor substrate selectivity and kinetic mechanism of Bacillus subtilis TagA', *Biochemistry*, 45(36), pp. 10895–10904. Available at: https://doi.org/10.1021/bi060872z.

Zolli, M., Kobric, D.J. and Brown, E.D. (2001) 'Reduction precedes cytidylyl transfer without substrate channeling in distinct active sites of the bifunctional CDP-ribitol synthase from Haemophilus influenzae', *Biochemistry*, 40(16), pp. 5041–5048. Available at: https://doi.org/10.1021/bi002745n.