Structure and Function of the Accessory Secretion System in Gut Symbionts

Ryan Griffiths - 100086871

A thesis submitted for the degree of Doctor of Philosophy (PhD) To the University of East Anglia.

Abstract

The accessory secretion (aSec) system is a protein export pathway uniquely present in gram-positive bacteria and dedicated to the secretion of large, glycosylated cell-wall anchored adhesins called serine rich repeat proteins (SRRPs). This system has been primarily characterised in pathogens in the context of biofilm formation and virulence. The recent discovery of aSec and SRRPs in *Limosilactobacillus reuteri* strains ATCC 53608 and 100-23 provided new insights into the structure, strain specific glycosylation and function of SRRPs in symbiotic gut bacteria but the aSec machinery has not been investigated. In *L. reuteri*, aSec consists of the translocation machinery (SecA2, SecY2, Asp4), chaperones (Asp1, Asp2, Asp3) and a variable number of glycosyltransferases (GTs) that *O*-glycosylate the secretory target SRRP. Here, a combination of biochemical, biophysical, and structural approaches was used to investigate the structure and function of aSec components in *L. reuteri* ATCC 53608 and 100-23 strains. To this aim, recombinant proteins of the *L. reuteri* aSec pathway were produced in *E. coli* and purified individually or in complex.

The crystal structure of LrGtfC$_{100-23}$, one of the GTs involved in strain-specific glycosylation of SRRP, was determined by X-ray crystallography showing a classical GT-B fold. Site-directed mutagenesis of LrGtfC$_{100-23}$ revealed the importance of Ser238 in conferring UDP-Glc specificity as shown using thermal shift assays while LrGtfC C240W$_{53608}$ mutation introduced promiscuity with positive thermal shifts for both UDP-Glc and UDP-GlcNAc ligands.

Small-angle X-ray scattering (SAXS) and Alphafold2 approaches were used to construct structural models of individual aSec components and complexes. The LrSecA2-SecY2-Asp4$_{53608}$ complex consisted of a predicted dimerised motor-ATPase SecA2, membrane protein SecY2 and Asp4. LrAsp1-Asp2-Asp3$_{53608}$ formed a 1:1:1 complex, where LrAsp2$_{53608}$ exhibited acetylesterase activity. Native SRRPs purified from both *L. reuteri* ATCC 53608 and 100-23 strains showed high binding avidity to both LrAsp1-Asp2-Asp3$_{53608}$ and the secretion complex LrSecA2-SecY2-Asp4$_{53608}$ with cross-strain recognition.

Together these data suggest a highly regulated and conserved secretion process that progresses the SRRP secretion cargo through *O*-glycosylation, *O*-acetylation, and transport in a sequential manner. This work opens new avenues of research for further biophysical characterisation by cryo-EM and potential for glycoengineering applications.

Access Condition and Agreement

Acknowledgements

Contents

Figures and Tables:

Abbreviations:

Abbreviations	Full name
AFLP	Amplified fragment length polymorphism
agWGA	Agarose-bound wheat germ agglutinin
AMP	Antimicrobial peptide
aSec	Accessory secretion
AST	Alanine-Serine-Threonine
ADP	Adenosine di-phosphate
ATP	Adenosine tri-phosphate
BR	Binding region
BSA	Bovine serum albumin
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
DNA	Deoxyribonucleic acid
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
EPS	Exopolysaccharide
FAP	Fimbriae associated protein
g	gram
GALT	Gut associated lymphoid tissue
GlcNAc	N-acetylglucososamine
GSP	General secretion pathway
GspB	Gordonii surface protein B
gSRR1	Glycosylated serine-rich repeat region 1
GT	Glycosyltransferase
GTP	Guanidine tri-phosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hr	Hour
HRP	Horseradish peroxidase
Ig	Immunoglobulin
IRA	Intramolecular regulator
Kd	Dissociation constant
kDa	Kilo Dalton
K_M	Michaelis constant
L	Litre
LDMII	*Lactobacillus* defined medium type II
LSP	Large surface protein
M	Molar
mg	milligram
Min	minute
mL	millilitre
MOPS	3-(N-morpholino)propane sulfonic acid
MRS	de Man, Rogosa and Sharpe
MUB	Mucin binding protein
MWCO	Molecular weight cut off
NBD	Nucleotide binding domain
ng	nanogram
NiNTA	Nickel nitriloacetic acid
OD	Optical density
OMV	Outer membrane vesicle

OST	Oligosaccharyltransferase
PAGE	Polyacrylamide gel electrophoresis
PBD	Protein binding domain
PBS	Phosphate buffered saline
PDB	Protein data bank
PEG	Polyethyleneglycol
PMSF	Phenylmethylsulfonyl fluoride
pNP	p-nitrophenyl
P(r)	pair distance distribution function
PSI	Pounds per square inch
rcf	Relative centrifugal force
Rpm	Rotations per minute
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulfate
SEC	Size-exclusion chromatography
Sec	secretion
sec	seconds
SEGF	Size-exclusion gel filtration
SOC	Super optimal broth with catabolite repression
SP	Signal peptide
SraP	Serine-rich adhesion protein
SPRIP	Single-primer reaction in parallel
SRR	Serine-rich region
SRRP	Serine-rich region protein
TAT	Twin-arginine translocation
TMB	Tetramethylbenzidine
TSA	Thermal shift assay
Tris	Tris(hydroxymethyl)aminomethane
UDP	Uridine diphosphate
Vmax	Maximum velocity
Vol	volume
WGA	Wheat germ agglutinin
w/v	Weight /volume
Å	Angstrom
°C	Degree centigrade

Chapter 1 Introduction

1.1 The vertebrate gastrointestinal tract

1.1.1 Anatomy and physiology

The vertebrate gastrointestinal (GI) tract and its accessory organs including salivary glands, liver, gallbladder, and exocrine pancreas are involved in the digestion of food, absorption of nutrients, waste elimination, as well as immune surveillance through gut associated lymphoid tissue (GALT). The composition of the GI tract can vary between vertebrates reflecting evolutionary history driven by differing features in food chemistry, which is a major driver in diversification of gut morphology, physiology, and biochemistry (pH, enzymes, bile salts, and mucus) (1, 2). Carnivores have a relatively simple GI anatomy with a shorter colon reflecting the digestion of high-protein dietary intake, whereas folivorous species may have an enlarged colon and caecum by comparison and frugivorous species developing intermediary structures (3). So, while mice, pigs, and humans share many of the major key anatomical organs of the digestive tract (oesophagus, stomach, small intestine, and large intestine), their sizes, functions and physiological properties are fundamentally different. For example, rodent stomachs are divided into glandular (thick walled with columnar epithelia) for secretion of mucus, stomach acid, enzymes, and non-glandular (thin walled with squamous epithelia) for food storage as well as digestion (4). Conversely the human and pig stomach is glandular type, with the pig stomach being 2-3 times larger with different distributions of cell types (cardiac mucosa) (Figure 1.1) (5).
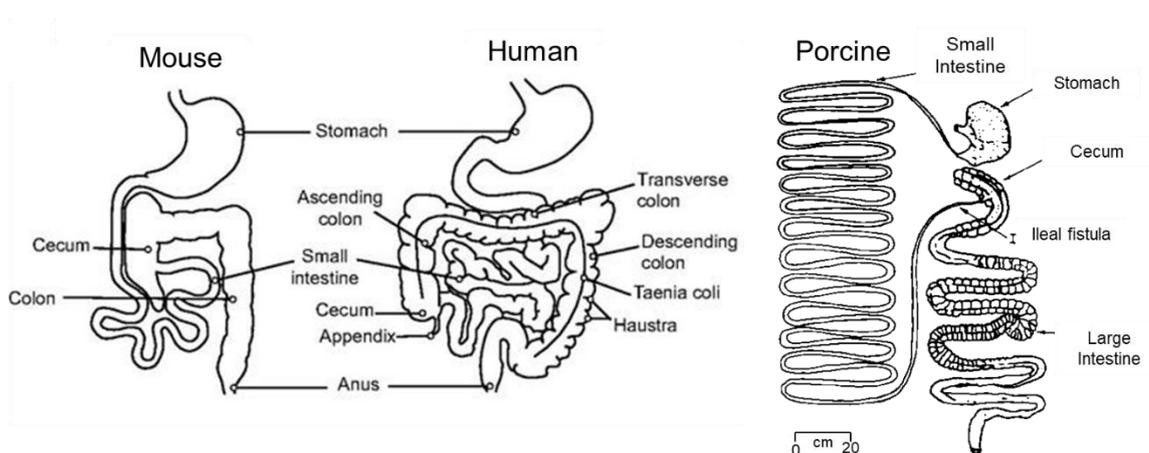
Figure 1.1 Schematic representation of the vertebrate gut anatomy. Mouse and human gut schematic annotated and taken from (6). Porcine gut schematic annotated and taken from (7).

The GI tract is covered by mucus layer whereby the composition, organisation, glycosylation, and thickness vary along the GI tract and between species (8, 9). The stomach and large intestine have a two-layered mucus system, composed of an inner, attached mucus and an outer, unattached, loose mucus layer, while the small intestine has a loose and penetrable mucus layer, consistent with the physiology of these organs (10). The average mucus thickness in the stomach is around 52 μm in rodents, 144 μm in humans, and 190 μm in pigs while in the large intestine, the inner mucus thickness estimates are ≈50 μm in mice and ≈200 μm in humans (11-13). Secreted gel-forming MUC2 (Muc2 in rodents) are the main structural components of mucus in the small and large intestine while gel-forming MUC5AC and MUC6 are the main mucins in the stomach (10). Mucin glycosylation profiles also show regio-specificity along the GI tract. The terminal epitopes show considerable variation with a decreasing gradient of fucose and ABH blood group expression and an increasing gradient of blood group Sd^a/Cad-related epitopes and sialic acid from the ileum to the colon in humans and reverse gradients in mice where the small intestine is dominated by sialylated structures and the colon with those terminating in fucose (Figure 1.2) (14, 15).
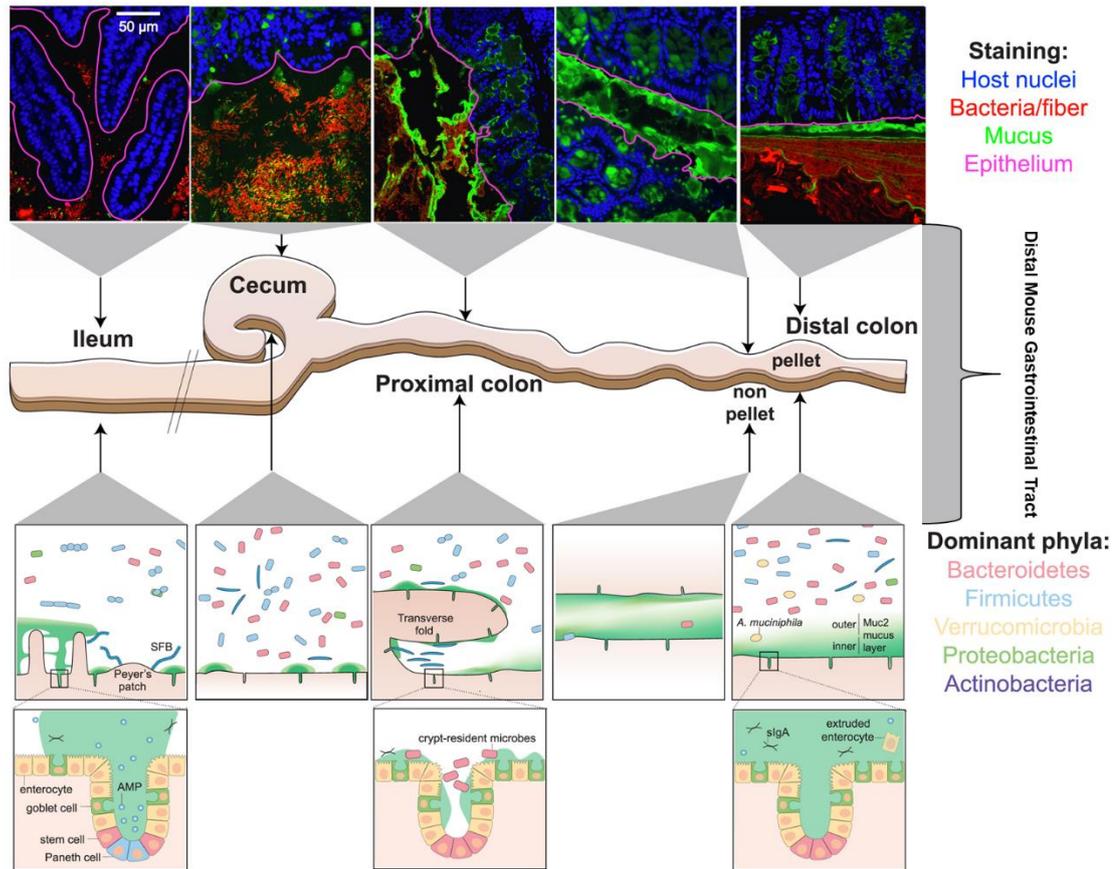
Figure 1.2 Biogeography of the mouse gastrointestinal microbiota. From top to bottom: confocal micrographs of intestinal sections stained with ulex europaeus-1 lectin (green) recognising fucose and 4′,6-diamidino-2-phenylindole (blue in epithelium, red in lumen). Fucose is most abundant in the mouse distal colon and less so in the caecum and proximal colon. The epithelial boundary is overlaid (magenta). A schematic of the distal mouse GI tract with following smaller schematics of characteristics for each intestinal location showing structure and bacterial localisation are heterogeneous along the longitudinal and transverse axes of the murine GI tract. The MUC2-dependent layer becomes increasingly dense and impenetrable by microbes towards the colon. The density and diversity of bacteria increase along the longitudinal axis, with the small intestine favouring facultative anaerobic, proteolytic bacteria, and the colon favouring anaerobic, saccharolytic bacteria. Along the transverse axis, most bacteria are spatially segregated from the host tissue by immunological and physical barriers (16, 17), with a few notable exceptions (18-20). Mucus structure in live animals is reviewed in (21). AMP = antimicrobial peptide. sIgA = secretory IgA. SFB = segmented filamentous bacteria. Figure from (22).

Together, these differences have an impact on the fitness of the organism to respond to its dietary challenges but may also affect the localisation as well as colonisation strategies of the trillions of microbes inhabiting the GI tract.

15

1.1.2 The gut microbiota

The gut microbiota is comprised of bacteria, viruses, fungi, archaea, and some eukaryotes whose populations vary across the intestinal biogeography (23). The gut microbiota is involved in metabolism, physiology, nutrition acquisition, immune function and is therefore considered another 'organ' of the GI tract due to its global impacts on the host organism (24, 25). This is particularly apparent whereby disruptions of the microbiota (or conditions of dysbiosis) have been associated with an increasing number of diseases such as obesity, diabetes, chronic inflammatory diseases but also neurological or brain disorders (26).

Most bacterial species inhabiting the vertebrate gut microbiota are members of four phyla: Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes (27, 28). The gut microbiota composition varies along and across the GI tract, due to the chemical, nutritional, and immunological gradients across the gut topography (Figure 1.2) (22). These microbial communities may differ between individuals within a species, however the notion of a 'core microbiome' has been suggested which describes microbial features common within a species (29, 30). This may include microbial populations present at certain times throughout the organism's lifespan, microbes involved in essential biological functional processes for the host, and microbes that may have co-evolved with the host to improve ecological fitness.

The evolution of complex life has always occurred in association with bacteria, so as a result there is an increasing research focus on the role microbes play in adaptation and natural selection. Microbiomes (the collective genome of the microbiota) greatly extend the number of genomes in one organism and can consequently have a substantial phenotypic effect on their host, thus potentially providing a stronger adaptive potential (31). The hologenome theory, albeit controversial, purports that the combined genomes of host and associated microbiota should be assigned as one evolutionary unit and together contribute to the overall ecological fitness (32, 33). For bacterial-mammal associations, phylosymbiosis ("microbial community relationships that recapitulate the phylogeny of their host") is apparent where there is higher compositional similarity between gut bacteria colonising related hosts versus distantly related hosts (34).

Conversely inheritance of relative microbial abundance between generations is not always vertically transmittable and can be influenced by a variety of factors. This diversity of inheritance and modularity of microbiomes complicates the matter of modelling microbial impact on host evolution.

The gut microbiota composition is subject to various changes due to selective pressures originating from various intrinsic and extrinsic factors that can occur over the entirety of an organism's lifespan (Figure 1.3) (35, 36). For example, in humans, starting from birth, the method of infant delivery and breast feeding is a critical stage in the vertical transmission of commensal bacteria from the maternal host to the offspring (37). Early colonisation by gut symbiotic *Bifidobacterium* and *Lactobacilli* strains has a huge impact on the development of the infant gut microbiota, metabolism, and health (37-39). The commensal microbial diversity is modified throughout the entire aging process and favours different compositions at various milestone developmental periods e.g., infancy, adolescence, adulthood (Figure 1.3 (35). Host genetics may also create dispositions towards particular commensals, as shown in twin studies, whereby gut microbiota composition is more similar between monozygotic twins than between dizygotic twins (40, 41). Medical interventions to diseases, such as treatments with broad-spectrum antibiotics have a major impact on the gut microbiota composition (42). During homeostasis, the gut microbiota is also greatly affected by the diet, which can change the metabolic landscape through nutrient availability for the various microbial members, thus enforcing selective pressures (43, 44). Finally, the gut can also be transiently colonised through the ingestion of probiotics ("live microorganisms that, when administered in adequate amounts, confer a health benefit on the host"), or the consumption of fermented foods, such as live yogurts, which contain bacteria able to populate or modulate the gut microbiota (45-47).
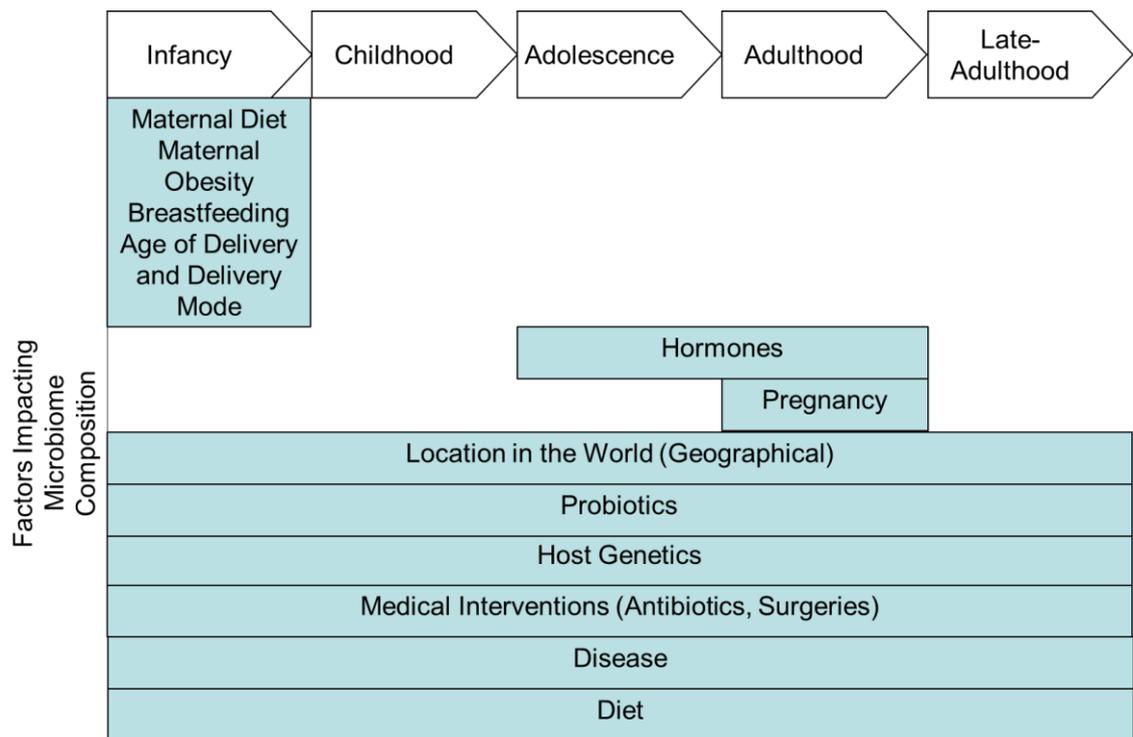
Figure 1.3. Factors which may influence the gut microbiota composition in humans during lifespan milestones. Information assimilated from (24, 37, 39, 42-45, 48).

1.2 Limosilactobacillus reuteri: a model gut symbiont
1.2.1 Host specialisation of *L. reuteri* strains

L. reuteri is present in the gut of a diverse range of vertebrates and has been the subject of great interest in respect to the co-evolutionary relationship it displays with its hosts (49). *L. reuteri* belongs to the taxonomic classification of Bacteria, Firmicutes/Bacillota, Bacilli, Lactobacillales, Lactobacillaceae and has recently been reclassified from *Lactobacillus* to *Limosilactobacillus* based on core genome phylogeny, (conserved) pairwise average amino acid identity, clade-specific signature genes, physiological criteria and the ecology of the organism (50).

In vertebrates such as rodents, pigs and chickens, *L. reuteri* is considered a dominant species of the GI tract (51, 52). In humans, the symbiotic relationship with *L. reuteri* appears to be ambiguous since its discovery in the mid-20th century as an autochthonous (indigenous) species. A recent taxonomic study reported that only 4% of individuals harboured *L. reuteri,* in line with the decline in human microbiome diversity predicted to be a result of medical interventions, diets and lifestyles (53-56). However, other studies showed that *L. reuteri* could be repeatedly isolated

18

from human gastric biopsies over a period of several months from 4 out of 12 patients (57, 58). In addition, *L. reuteri* has also been isolated from human breast milk, highlighting the capacity for vertical transmission from mother to infant and a possible role in early colonisation of the infant gut (59, 60).

In rodents, the murine forestomach, in pigs the pars esophagea (stomach), and in chickens, the crop, all have a stratified squamous epithelium composed of various columnar cell types, which for mouse have been observed coated with *L. reuteri* biofilms, as shown by electron microscopy (Figure 1.4AB) (61-63). Conversely, humans lack stratified squamous epithelia and colonisation by *L. reuteri* is confined to the mucus as shown using human small intestinal biopsies (Figure 1.4C) (64). *L. reuteri* has been shown to transmit horizontally through animal behaviours, for example, rodents exhibit coprophagy, as well as vertically to subsequent generations through breast milk and vaginal mucus contact in pigs, thus being transferred maternally to the offspring (65).
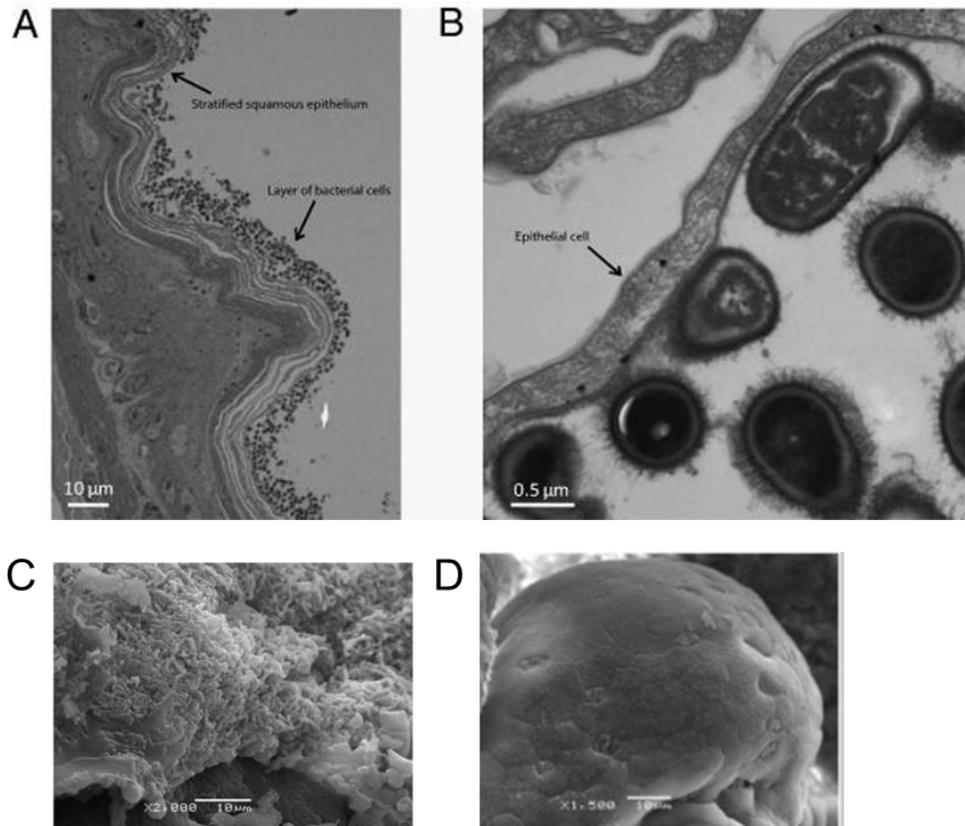
Figure 1.4. Host specific associations of *L. reuteri* strains. (A) *L. reuteri* 100-23 biofilm were visible on the stratified squamous epithelium present in the forestomach of an ex-*Lactobacillus*-free mouse 7 days post-inoculation. (B) *L. reuteri* cells were shown attached directly to stratified cells. These micrographs were produced by transmission electron microscopy and taken from (61) (C) Scanning electron microscopy of duodenal biopsies incubated with *L. reuteri* ATCC 6475 for 6 hr (D) *L. reuteri* appeared restricted to the outer mucus layer and did not contact the duodenum epithelium surface. Taken from (64).

L. reuteri host-specific adaptation is apparent at the genomic level. Analysis of the population structure of a large collection of *L. reuteri* strains from multiple vertebrate hosts highlighted the host-constrained diversification of *L. reuteri* populations. Amplified-fragment length polymorphism (AFLP) of 16S rRNA genes revealed distinct clusters that are host specific for rodent (B, D, E), pig (A, Cii) and human (Ciii) (Figure 1.5) (66).
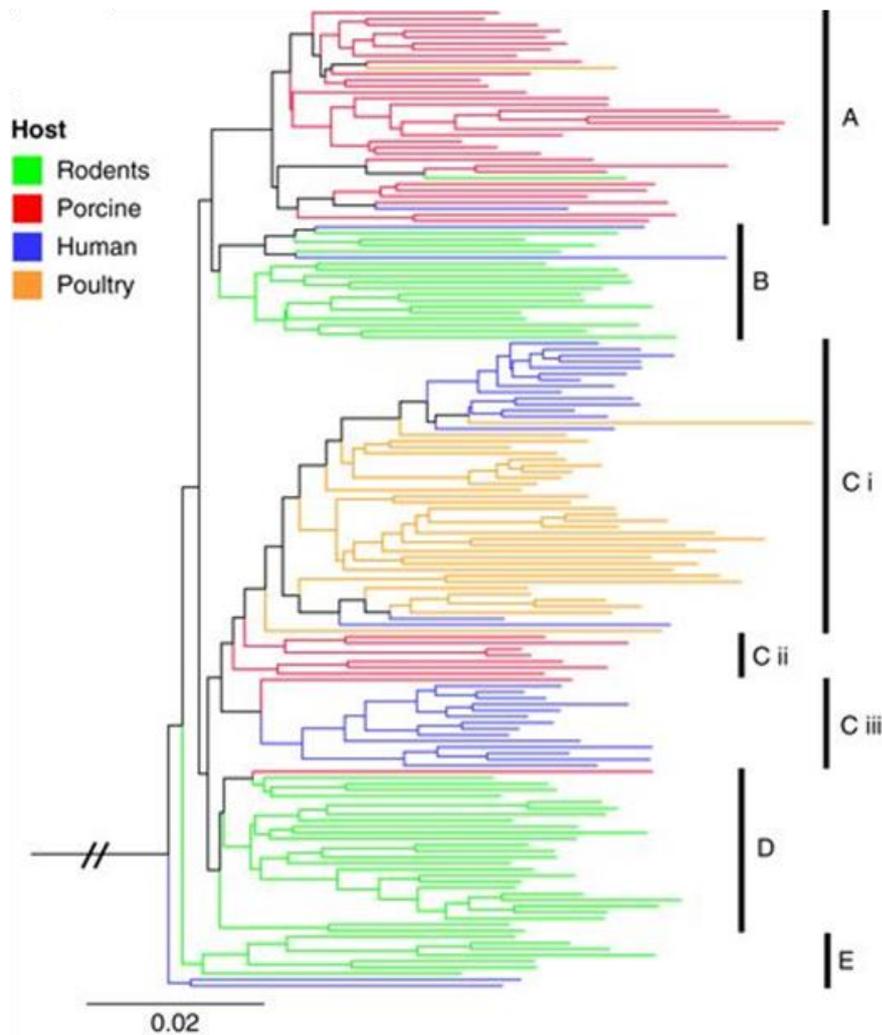
Figure 1.5. Amplified-fragment length polymorphism (AFLP) analysis of *L. reuteri* strains. Searches were performed using strains from humans, rodents (mice and rats), pigs and poultry (chicken and turkey). A neighbour-joining phylogeny of 165 strains based on 439 markers obtained from two AFLP primer pairs was generated, and branches were coloured according to host origin. Taken from (66).

Some of these host clades were interrupted with *L. reuteri* strains mostly of human, bovine or chicken origin, perhaps as a result of recombination. Further multi-locus-based analysis leading to the construction of a more detailed genealogy tree for three conserved loci also revealed dominant host-specific clades (66). This analysis confirmed that although recombination was prominent in *L. reuteri* strains, it did not confound the ancestry of individual lineages. The ecological fitness of *L. reuteri* strains within preferential natural host niches has been demonstrated *in vivo* using mouse models (67). It was found that only *L. reuteri* strains isolated

from rodents could effectively colonise germ-free (GF) mice and outcompete *L. reuteri* strains from pig, chicken, or human origin, resulting in biofilm formation (Figure 1.6) (67). Similar strain specific colonisation in GF mice by *L. reuteri* strains has been observed consistently (68-70). Analysis of faecal samples of vertebrates administered *L. reuteri* strains from humans, pigs, chicken, and rodent reported a strong host adaptation of *L. reuteri* to rodents and chickens, indicating a joint evolution of this bacterial species with several vertebrate hosts (71). Additionally, it is important to note that cohabitation of biofilms in rodents can result in ecological facilitation between two geographically distinct vertebrate gut symbionts (*L. reuteri* and *Lactobacillus taiwanensis*), highlighting that interspecies interactions in the gut microbiota can also be a factor in driving host specific adaptation (72). Taken together, these data suggest that the specific environment provided by distinct vertebrate hosts is a key driver of diversification of the ancestral *L. reuteri* population.
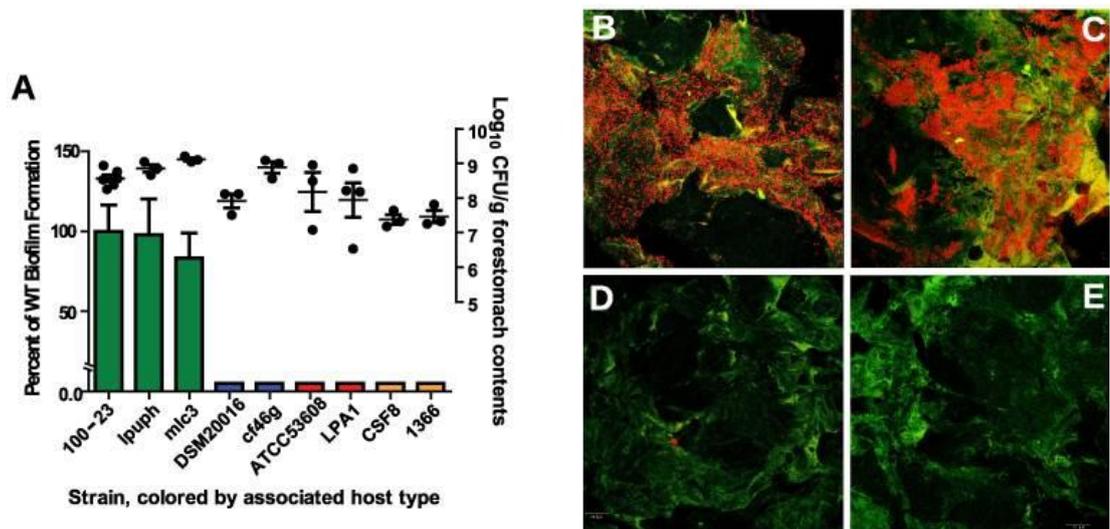
Figure 1.6. Biofilm formation of *L. reuteri* strains is associated with host type. (A) Quantification of biofilm density (relative to biofilm of strain 100-23) by confocal microscopy and cell counts in germ-free mouse forestomach contents of *L. reuteri* strains two days after gavage with a single dose of $\sim 10^7$ cells. Bars are colour coded according to host origin (green, rodent; blue, human; red, pig, and orange, chicken). Confocal micrographs showing density and pattern of bacteria (red) by strains (B) Lpuph (mouse), (C) Mlc3 (mouse), (D) DSM20016T (human), and (E) ATCC 53608 (pig). From (67).

1.2.2 *L. reuteri* colonisation factors

As a gram-positive bacterium, *L. reuteri* has a peptidoglycan cell wall layer surrounding the cell membrane (36). Briefly, peptidoglycans are synthesised from nucleotide precursors, N-acetylmuramic acid (MurNAc), D- or L-amino acids, which are then covalently linked to lipid carriers to form lipid I, or further modified with UDP-GlcNAc to form lipid II (37, 38). These are then polymerised by dedicated transglycosylases and transpeptidases to form cross-linked peptidoglycans forming the cell wall (39). The cell wall is then further decorated with exopolysaccharides (EPS) and various cell wall embedded proteins. Finally, *Lactobacilli* are usually surrounded by glycosylated, proteinaceous S-layers (Figure 1.7) (40).
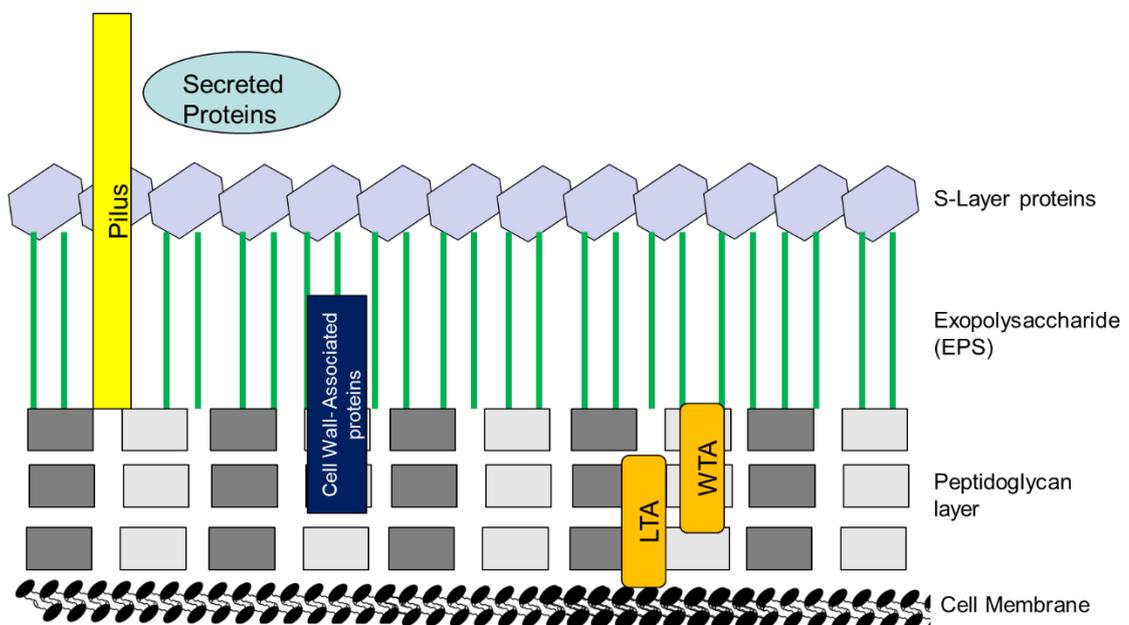
Figure 1.7. Schematic representation of the cell surface architecture in *Lactobacilli*. The cell membrane of *Lactobacilli* is coated with a multi-layer consisting of peptide-glycan, lipotechoic acids (LTA), wall teichoic acids (WTA), various cell wall associated proteins. This is followed by an exopolysaccharide (EPS) coating and S-layer proteins. Adapted from (73).

Components present on the cell surface are a primary point of interaction for bacteria with their surroundings and also contribute to the various host colonisation strategies to adapt to physiological differences in the GI tracts of vertebrates. Several surface proteins such as mucus binding protein (MUB, CmbA), large surface protein (Lsp) and serine-rich repeat protein (SRRP)

have all been shown to mediate interaction of *L. reuteri* with the host tissues in a strain-specific manner *in vitro* (74-77). The binding mechanisms of MUB, through its long and linear multi-repeat structure with mucin glycans, are predicted to potentiate the retention of bacteria within the outer mucus layer (78) while SRRP has been shown to be essential for biofilm formation of rodent *L. reuteri* strains in gnotobiotic mice (67). These proteins contain LPXTG cell wall binding motifs, are large, have binding regions and contain multiple repeated motifs. The LPXTG motif is integral to cell-wall integration which is carried out by sortases present in the periplasm between the cell membrane and cell wall that carry out transpeptidation in a fairly conserved manner for a wide range of substrates (79, 80). Briefly the LPXTG motif is cleaved and a Thr residue of the cleaved product is subsequently crosslinked to an acceptor active amino group of the gram-positive peptidoglycan cell wall stem peptide (80).

Additional factors involved in host-strain colonisation of *L. reuteri* in the GI tract include biologically active prophages advantageous for killing competitor strains, or the production of metabolites often associated with health promoting properties, such as reuterin (a mixture of different forms of 3-hydroxypropionaldehyde) (81, 82). Reuterin is a potent antimicrobial derived from glycerol metabolism, which has been shown to inhibit the growth of bacteria, but also yeasts, fungi or viruses (83). Other *L. reuteri* antimicrobial metabolites include lactic acid, acetic acid, ethanol or reutericyclin, and conversely cross-feeding relationships with other bacteria, such as *Bifidobacterium bifidum* in mouse, where trophic interactions based on 1,2-propandiol also increased ecological fitness (84, 85). All these competitive advantages are predicted to be involved in potential beneficial health outcomes whereby for example, *L. reuteri* was shown to protect against enteropathogenic *E. coli* (EPEC) infection, a major cause of diarrheal infant death in developing countries, by competing for binding sites through strain-specific adhesins (86). Diarrhoeal diseases are also a major burden in animal husbandry, where incidence can be lowered in piglets fed with *L. reuteri* I5007 (a pig strain) due to inhibition of enterotoxigenic *E. coli* (87, 88). Research on identifying probiotic *L. reuteri* strains for humans and animal husbandry is

ongoing and the European Food Safety Authority (EFSA) has recognised the beneficial effects of *L. reuteri* CECT 8700 as a zootechnical additive and gut flora stabiliser (89).

In summary, *L. reuteri* is a good model organism to study the evolutionary and biochemical mechanisms underpinning bacteria-host interactions.

1.3 Bacterial protein glycosylation

Glycosylation is a common polypeptide medication in nature whereby carbohydrates can be covalently bound to proteins through glycosidic linkages catalysed by glycosyltransferases (GTs) to the amide nitrogen of Asn residues (*N*-glycosylation), and the hydroxyl oxygen of Ser/Thr residues (*O*-glycosylation). Bacterial protein glycosylation can occur in the periplasmic space where a glycan attached to a lipid carrier (lipid-linked oligosaccharide) is transported from cytoplasm to periplasm via a flippase allowing for *en bloc* protein glycosylation by an oligosaccharyltransferase (OST) (90-92). Protein glycosylation in bacteria can also occur in a stepwise sequential addition by GTs in the cytoplasm.

N-glycosylation systems have been mainly discovered in gram-negative bacteria with some exceptions, such as *Mycoplasma sp.*, with *Campylobacter jejuni* as the model organism (93-95). In *C. jejuni*, nucleotide activated sugars are assembled onto a undecaprenylphosphate carrier and are flipped across the inner-membrane into the periplasm, where subsequently the OST PglB can transfer the broad range of structurally diverse glycan donors on to a specific protein acceptor sequence: Asn/Asp-X$_{-1}$-Asn-X$_{+1}$-Ser/Thr (X$_{-1,+1}$ ≠ Pro) (96, 97). Homologues of PglB are present in most δ- and ε- proteobacteria, however the organisation of gene clusters encoding GTs and number of OSTs can vary greatly between species and even strains (98). In contrast to the *en bloc* glycosylation pathway observed in *Campylobacter*, β- and γ-proteobacteria use a sequential addition pathway for cell surface adhesins, such as the HMW1 from *Haemophilus influenza* (99).

O-glycosylation is widespread for bacteria and is particularly abundant for large-cell surface structures (100). For example, in several gram-negative, the major subunit of type IV pili PilE has been shown to be glycosylated with trisaccharides in *Neisseria meningitidis*, which is carried

out *en bloc* by the OST Pgl (101). Similar pathways have also been identified in *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* for pilin glycosylation (102, 103). An *en bloc* protein *O*-mannosylation pathway has been identified in some gram-positive Actinomycetes including *Mycobacterium*, *Corynebacterium* and *Streptomyces* (104-106). Stepwise cytoplasmic *O*-glycosylation has been well-studied in flagellins, where glycosylation is essential for flagellar motility and can vary in complexity of length and number of sugar moieties (107). Stepwise *O*-glycosylation also occurs along with dedicated secretion systems for the glycosylated cargo, as is the case for the accessory secretion (aSec) system present in some gram-positive bacteria including *L. reuteri* (108).

1.4 Bacterial secretion systems

All domains of life secrete circa one-third of their entire proteome for it to be across or inserted within biological membranes. As a result, dedicated membrane associated systems have evolved to facilitate this export and summarised in Figure 1.8. There are several ways to categorise and subdivide these by shared characteristics. One way is to organise secretion systems dedicated to crossing 'self'-cell membranes (Sec, Tat, Type I, Type II, Type V, Type X secretion systems) or other recipient cell membranes (Type III, IV, VI) (109, 110). There are also secretion systems dedicated to specific processes such as type VIII for curli protein (amyloid fibre) biogenesis in gram-negative bacteria, or secretion systems only present in specific groups of organisms, for example type IX in *Bacteroidetes,* type VIIA in *Mycobacteria sp* and VIIB in *Firmicutes* (*Staphylococcus aureus*) (111, 112). The different membrane and cell-wall compositions between gram-positive, gram-negative, or species-specific structures (mycobacterial envelopes) also require different secretion pathways. Briefly, gram-negative bacteria are surrounded by a thin peptidoglycan cell wall, which itself is surrounded by an inner membrane and outer membrane containing lipopolysaccharide (LPS), whereas gram-positive bacteria lack an outer membrane, but are surrounded by thicker layers of peptidoglycan as described above (113) (Figure 1.7). The space between layers is generally referred to as the periplasmic space or periplasm. The number of secretion systems is constantly being updated, but this section aims to provide a brief overview

about the main gram-positive and gram-negative bacterial "type" secretion systems determined so far (for a summary, see Figure 1.8).
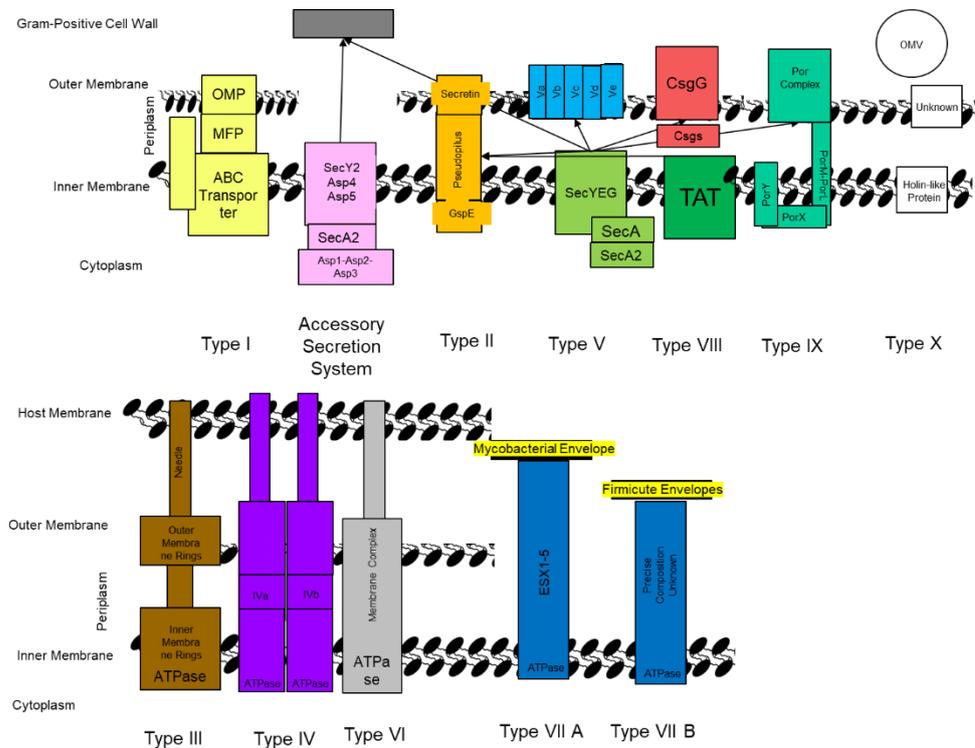
Figure 1.8. Summary schematic of secretion systems in bacteria. Type I (yellow), II (orange), V (light blue), VIII (red), IX (turquoise), X (white) have multiple steps when crossing between cell membrane to the outer membrane. SecYEG (or Sec or General Secretion Pathway) (light green) and TAT (dark green) facilitate transport across the inner membrane for a majority of secretion systems. The aSec system (pink) shares similarities with SecYEG but is dedicated for transport for gram-positive bacteria. Type III (brown), Type IV (purple), Type VI (grey) are dedicated for transport to host membranes. Finally Type VII (dark blue) secretion systems are typically found in gram-positive bacteria, but also appear to export through specific envelope conditions as in the mycobacterial envelope. This information was assimilated from (109, 110, 114-117). MFP = Membrane Fusion protein. OMP = Outer Membrane Protein. OMV = Outer Membrane Vesicle.

1.4.1 The general secretion pathway (GSP) or Sec

Secretion through the GSP (or Sec) is a multi-step process that can be divided into three main stages (Figure 1.9) (116): (1) Protein sorting and targeting (2) Translocation (3) Release and maturation. Proteins targeted for secretion are considered pre-proteins and typically carry a cleavable N-terminal signal peptide (SP), which is typically between 20-30 residues long and

contains a positively charged N-terminus, a hydrophobic core and a polar C-terminal end containing a peptidase recognition site. Lipoprotein SPs secreted via the GSP contain a modified lipobox motif with a LxxC consensus sequence whereby the cysteine residue becomes a target for lipidation and membrane anchorage (118). Targeting to the membrane SecYEG translocon occurs co-translationally, or post-translationally where a major determining factor is the hydrophobicity of the SP and destination of the protein. Piloting factors such as the ribonucleoprotein signal-recognition particle (SRP) can target inner membrane proteins for secretion co-translationally along with the SecYEG-bound SRP receptor FtsY. SRP can also target YidC, which can insert membrane proteins in a cooperative manner with the translocon (119, 120). SecA-dependent targeting occurs co-translationally when determined by a highly hydrophobic SP, and post-translationally when further supported by chaperones such as the ribosomal trigger factor for less hydrophobic SPs, or SecB (121, 122). The SecB chaperone is not commonly present in gram-positive bacteria and is only associated with gram-negative bacteria or *Mycobacterium tuberculosis* (123). Other analogous proteins have been shown to fulfil similar roles to SecB, for example CsaA in *Bacillus subtilis* (124). Binding by cytosolic factors to the pre-protein can occur while the polypeptide chain is being elongated but is not necessarily mechanistically involved in protein synthesis (125). SecB bound pre-proteins are brought directly to the SecA subunit of the translocase machinery (119, 126) (Figure 1.9).
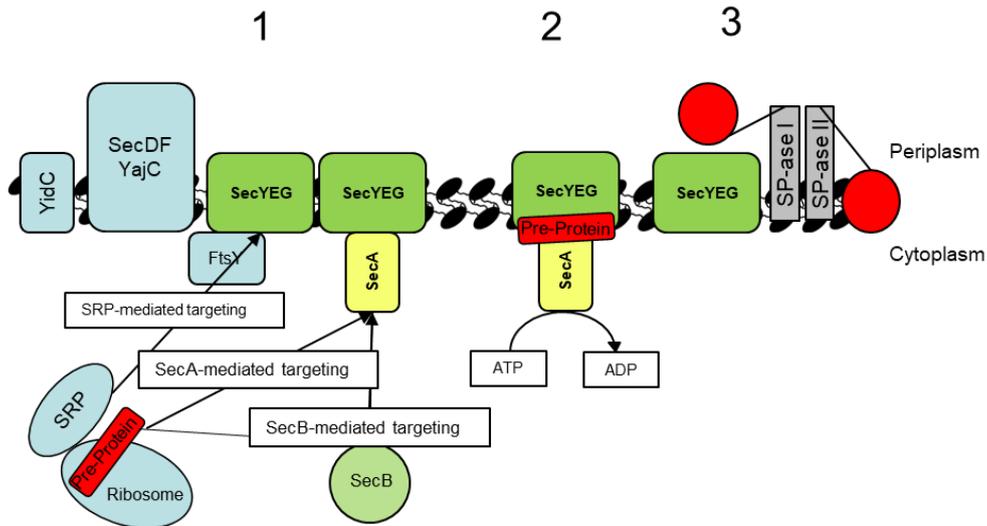
Figure 1.9. Schematic representation of the general secretion pathway (GSP) in bacteria. The bacterial pre-protein translocase subunits consist of the SecYEG pre-protein-conducting channel (green) and the ATPase motor SecA (yellow). SecYEG can associate with the auxiliary proteins SecDFYajC and YidC (blue). (The secretion GSP process follows the sequential events: (1) binding of SP target pre-protein (red) by cytosolic chaperones SecB or FtsY and association with translocase machinery, (2) ATP-driven transport of target pre-protein through translocase machinery, and (3) translocation of the pre-protein into the periplasmic space or laterally embedded into the cell membrane and cleavage of the SP by signal-peptidases I or II allowing for proper folding. Adapted from (116).

The translocase describes a trans-membrane protein channel built of SecY, SecE and SecG forming a heterotrimer as well as a mobile SecA ATPase, which drives translocation using ATP hydrolysis (127). The SecDF-YajC also interact with the translocon channel and have been implicated with release of the preprotein on the periplasmic side, stabilising SecY as well as forming a holo-translocon with SecYEG and YidC which becomes more dependent on proton-motive force (128, 129). The structure and biophysical properties of SecA have been extensively studied (126, 127, 130-133). The domain organisation of SecA consists of a nucleotide binding domain (NBD) and two intramolecular regulator of the ATPase (IRA) domains which contain a conserved DEAD motor, as well as a pre-protein binding domain (PBD) (Figure 1.10). The DEAD motor is highly flexible to regulate ATPase activity (131). ADP binding stabilizes the domain, however; it still acts allosterically controlling the PBD. Protein binding of SecA is two-fold

whereby the PBD can bind the SP on the mature pre-protein domains and the tip of the C-terminus has also been shown to interact with translocating polypeptides (134). The oligomeric state of the SecA component appears difficult to capture in its most physiological relevant state. Despite being usually isolated as both parallel and anti-parallel dimers from a broad range of bacteria such as *T. thermophilus, E. coli,* or *S. aureus* strains, the crystal structure of SecA from *Bacillus subtilis* 168 favoured a monomeric state (135).
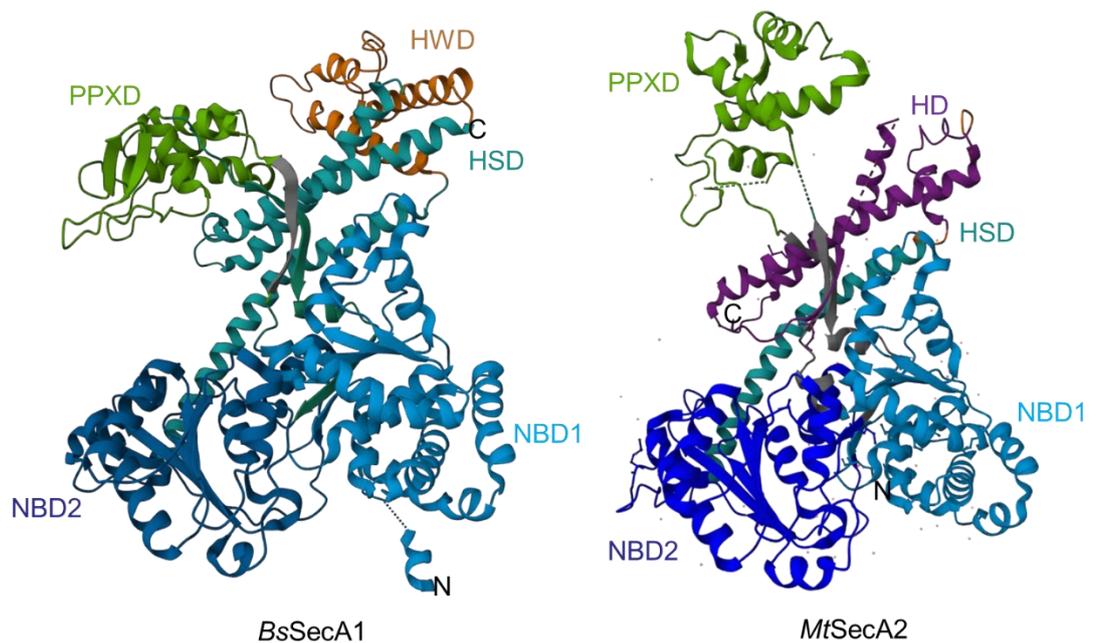
Figure 1.10 Structure of monomeric *B. subtilis* SecA1 and *M. tuberculosis* SecA2. Monomeric *Bs*SecA$_{168}$ (PDB: 1TF5) and *Mt*SecA2$_{H37Rv}$ presented as a cartoon with annotated and coloured folds. Nucleotide binding domain 1 (NBD1) (light blue), NBD2 (dark blue), protein-protein cross-linking domain (PPXD) (light green), helical scaffold domain (HSD) (cyan), and helical wing domain (HWD) (orange), helical domain (HD) (purple). Adapted from (136, 137).

The structures of SecY from *Thermus thermophilus* HB8 and *E. coli* K12 have been solved by X-ray crystallography and cryo-EM as part of the SecYEG complex. The SecYEG trimer may be sufficient for translocation, although there could be dimeric SecYEG pores or even higher-order assemblies with ribosomes (Figure 1.11) (127). The pore is formed between the halves of SecY each containing five transmembrane helices. The inactive channel harbours a α-helical plug and

a ring of six hydrophobic residues, which prevent ion leakages or dissipation in the proton motive force. A consensus has yet to emerge on the dynamics of protein translocation through this system and several models have been proposed including i) a processive power-stroke movement of the substrate for every ATP-molecule hydrolysed or ii) passive diffusive motions given directionality at the expense of ATP through a Brownian Ratchet movement (133). Other auxiliary proteins have been defined (such as SecD or SecF), which are required for a functional channel, but their roles have not been completely elucidated (138).
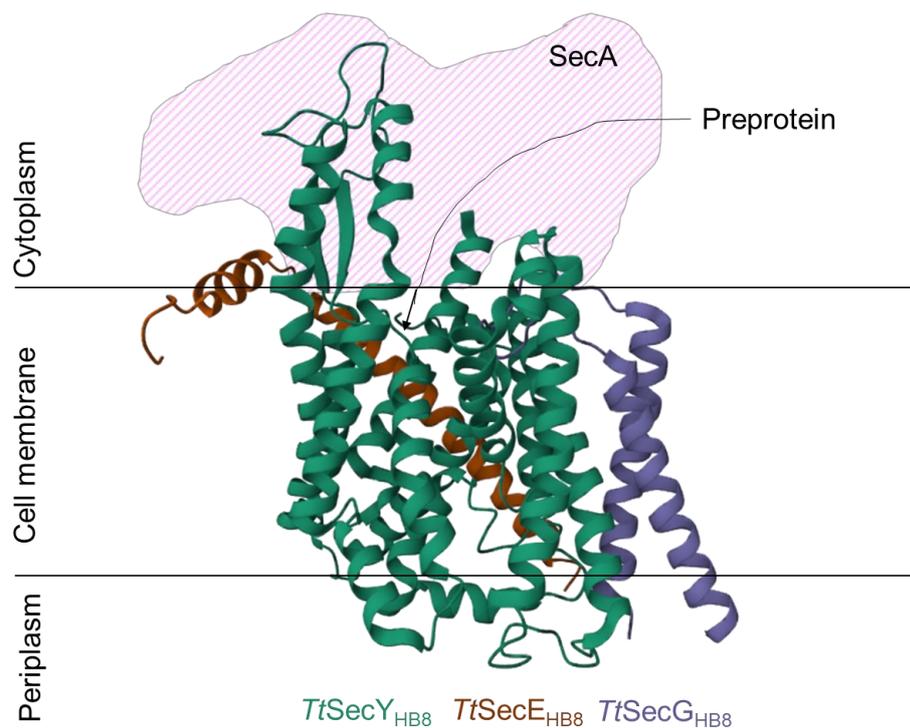
Figure 1.11. Structures of the SecYEG complex from *T. thermophilus* HB8. The *Tt*SecYEG$_{HB8}$ complex in a cartoon ribbon presentation where SecY is coloured green, SecE is orange, and SecG is purple (PDB: 5CH4). SecA (shaded pink) and preprotein (line) bind cytoplasmic regions of SecY. Adapted from (139).

1.4.2 Tat secretion pathway

Twin-arginine translocation (Tat) secretion also allows for initial translocation across the cell membrane and is also involved in some of the other "Type" secretion systems. It was first discovered in both prokaryotes and chloroplasts and can be essential for virulence in pathogenic bacteria such as *Pseudomonas aeruginosa* and *Yersinia pseudotuberculosis*, as well as viability

of some halophilic archaea (109, 140). In contrast to the Sec system, it is dedicated to fully folded proteins, uses a proton motive force, and responds to a different SP sequence containing an N-terminal arginine dipeptide providing its namesake (Arg-Arg-Polar amino acid-Phe-Leu-Lys) (141-143). The Tat pathway of protein secretion consists of 2–3 components: TatA, TatB, and TatC. In a representative *E. coli* K12 model, *TatABCD* are part of one operon whereby *TatE* is located on another chromosome locus (144). TatA is an L-shaped small protein composed of one transmembrane helix and one amphipathic helix which forms the translocation channel (145). TatB is involved in SP recognition and NMR revealed a structural fold in 4-α helices that have flexible regions suitable for interactions with a diverse range of substrates (146). TatC is a transmembrane protein with 6 transmembrane helices as shown by the crystal structure of TatC from *Aquifex aeolicus*VF5 (PDB: 4B4A) (147).

In summary, the protein translocation process begins with essential signal peptide recognition and substrate binding to a TatBC complex. Upon formation of the TatBC–substrate protein complex, the TatA subunits are recruited. Together, perturbation of the cell membrane by the transmembrane domains may facilitate transport (142). However, different versions of Tat systems exist, for example minimalistic pathways in *Staphylococcus aureus* containing only TatA, TatC, and Tat systems containing multiple TatAC translocation channels as reported in *B. subtilis* 168 (143, 148, 149).

1.4.3 Type I secretion system

The type I secretion system is exclusive to gram-negative bacteria and consists of a series of complexes that translocate primarily toxins from the cytoplasm directly across the cell envelope thus bypassing the periplasm completely (150, 151). The translocation follows a particular sequence of events which have been primarily characterised in the *E. coli* model for the secretion of hemolysin in a virulence context: (1) an accessory protein that spans from the periplasm to the cytoplasm is critical for the stabilisation of an inner membrane ABC-transporter with the outer membrane translocation tunnel (TolC), resulting in (2) a complex formation between the membrane transporter and TolC allowing for recognition of a C-terminal recognition sequence in

the secretory target, (3) the energy for translocation is then facilitated by ATP hydrolysis, which also opens the translocation tunnel, and (4) the substrate is secreted and the trimolecular complex between the ABC-transporter, TolC and the secretory target dissociates.

1.4.4 Type II secretion system

The type II secretion system is exclusive to gram-negative bacteria and utilises the aforementioned Tat and Sec pathways for initial secretion across the inner membrane. A combination of X-ray crystallography and more recently cryo-EM investigations have provided new insights into the pilus-like biomolecular machinery of *E. coli* K12, allowing secretion of a wide variety of virulence factors, metalloproteases, and digestive enzymes without the need for an identifying SP (152). The architecture of the translocation machinery is extremely complex with 40-70 proteins hypothesised to span from the cytoplasm and extending beyond the outer membrane with ATP as an energy donor for translocation (115, 153). Folded substrates are recruited in the periplasm activating the ATPase portion in the cytoplasm, allowing for transport through the translocation channel in the outer membrane portion called the secretin (153).

1.4.5 Type III secretion system

The type III secretion system is well described and forms a characteristic "nano-syringe", allowing gram-negative pathogens (e.g. *Salmonella typhimurium* strains) to inject toxins into host cells in an ATP-dependent manner (154, 155). Briefly, this complex is a multicomponent structure that contains the following: (A) an inner membrane anchor, (B) an export machinery with substrate recognition and ATP-hydrolysis function, (C) a needle-like filament that serves as a translocation channel for proteins and effectors and (D) the translocation pore, which inserts into the eukaryotic host cell membrane. The activity of this secretion system is closely linked with the virulence and infection process of the pathogen.

1.4.6 Type IV secretion system

The type IV secretion system is dedicated to DNA-nucleoprotein molecules and facilitates conjugation between gram-negative/positive bacteria and has been associated with horizontal

gene-transfer between bacterial species (and even interkingdom transfer) (114). Type IV secretion systems have been subdivided into IVA, IVB and "other". Type IVA are systems structurally and functionally resembling the well-described *Agrobacterium tumefaciens* VirB/D4 which has been applied for the genetic modification of plants (114). Type IVB are systems resembling *Legionella pneumophila* Dot/Icm (156). "Other" refers to those that have no ancestral relatedness to the IVA, IVB pathways or have not been well characterised. Mechanistically, there is a lot of variation in Type IV systems, but all recruit multiple ATPases for secretion (156).

1.4.7 Type V secretion system

The type V secretion system allows for transport across gram-negative outer membranes, and thus relies on the Sec pathway for the primary translocation across the inner membrane. Type Vs are subdivided from Va to Ve (157). Type V secretion is considered unusual compared to other secretion systems, as it has a high compositional diversity between bacteria. Briefly, a complex is formed between the passenger domain containing a cleavable SP sequence (the functional protein to be secreted), a linking region, and a membrane translocating b-barrel domain. In type Va, the membrane translocating domain forms the initial pore allowing the passenger substrate to bypass the membrane before autocatalytically cleaving and releasing it into the extracellular space (158). In type Vb, the passenger domain and membrane translocating domain are separate, and thus no cleavage step is necessary for transport into the extracellular space (159). The Type Vc are obligate trimers and mainly encompass trimeric autotransporter adhesins (TAAs) (160). Type Vd are similar to Va but contain an additional periplasmic domain structurally similar to the membrane secretion domain (160). Type Ve are similar auto-transporters to Va, but the order of passenger and membrane translocation domain are reversed (161). Transport is not driven by ATP, but rather through free energy from folding the passenger protein through the membrane. Notably, the type V system is aided by other outer-membrane systems, such as, BamA, which facilitate the initial pore formation (162).

1.4.8 Type VI secretion system

The type VI secretion system is a large, contractile ATP dependent puncturing structure that is utilised by gram-negative bacteria for a myriad of effector protein secretion but has also been heavily implicated with bacterial communication in a contact-dependent manner (163). Type VI secretion systems are anchored to the cell envelope by a 1700 kDa membrane core complex comprised of 10 heterotrimeric complexes containing the proteins TssJ, TssL and TssM. This acts as a docking site for a cytoplasmic baseplate (TssEFGK) which is assembled around a spike (VgrG-PAAR) (164, 165). TssBC polymerises into a contractile sheath and is coordinated by TssA, which most commonly ends up forming a distal cap at the end of the sheath and tube (166). This sheath surrounds an inner tube composed of hexameric rings of hemolysin-coregulated protein (Hcp). Cargo can be Hcp-dependent which is then stored in the inner tube lumen, or cargo can be associated with the outside of the spike by associating with VgrG, or PAAR. Typically type VI secretion systems contain multiple homologues of Hcp, VgrG and/or PAAR proteins to accommodate specificity towards their respective effectors (167). The sheath can contract rapidly which allows the Hcp-VgrG-PAAR to propel out of the secreting cell and into the target cell thus delivering cargo effectors. Overall, the ancestral function of type VI secretion targets other strains and species of bacteria, however, systems targeting eukaryotic cells have also been identified in the context of pathogenesis (168).

1.4.9 Type VII secretion system

The type VII secretion system is exclusive to some gram-positive bacteria often with lipidated cell walls that require specialised systems to circumvent it (such as the mycomembrane of *M. tuberculosis* strains). However, type VII systems have also been identified in other gram-positive acteria, leading to further classification into lettered systems VIIa and VIIb. The mycobacterial Type VIIA secretion system contains 5 conserved membrane components (EccB, EccC, EccD, EccE and MycP with a 6:6:12:6:3 stoichiometry) which form the secretion system across the inner membrane, for which now two high resolution CryoEM structures exist for *M. tuberculosis* and *Mycobacterium xenopi* (111, 169-171). The overall membrane assembly can be described as a

trimer of dimers, whereby each dimer contains one MycP and two protomers with each containing one copy of EccB, EccC, EccE and two copies of EccD. Chaperones like EspG prevent aggregation of secretion substrates. Both EccC and EccA contain nucleotide-binding domains (NBDs) consistent with ATPase activity (172). Particularly EccC performs an active role adopting an extended and contracted form on the cytoplasmic side which may provide a basis for substrate selection, recognition, and transport across the membrane (171). No structural data so far provides a foundation to hypothesise how the transport across the mycomembrane is facilitated (117). VIIb has been best described in *S. aureus* strains although bioinformatic work has identified candidates present in many other gram-positives such as *Streptococci, B. subtilis* strains (112, 173). The main characteristics shared between Type VIIa and VIIb secretion systems are a conserved ATPase and the presence of a specific C-terminal SP on the substrate (174).

1.4.10 Type VIII secretion system

The type VIII secretion system is exclusively involved in curli biogenesis. Curli are amyloid fibres and a major proteinaceous component of a complex extracellular matrix produced by many *Enterobacteriaceae* and are involved in adhesion, auto-aggregation, and biofilm formation (175, 176). Their amyloidogenic and tangled nature makes them toxic due to aggregative potential. Transport begins with secretion of the curli subunits (CsgA and CsgB) via Sec across the inner membrane. In the periplasm, CsgC prevents premature folding and polymerisation of the CsgA fibre and CsgE chaperones it towards the CsgG transmembrane channel. Recruitment and (partial) enclosure of CsgA in the secretion complex is proposed to create an entropy gradient over the channel that favours CsgA's outward diffusion as an unfolded, soluble polypeptide (177).

1.4.11 Type IX secretion system

Type IX secretion system, also known as the Por secretion system (PorSS) or PerioGate, is exclusively present in the *Bacteroidetes* phylum (178). First, the cargo proteins are guided by a classical SP to the Sec machinery in the inner membrane. During translocation, the SP is cleaved off by type I signal peptidase, and the cargo is released into the periplasm. In the periplasm,

transported proteins fold into a stable conformation, as indicated from the accumulation of their soluble forms in the periplasm of T9SS secretory mutants. Whether the cargo proteins require chaperone(s) to assist in folding and/or guiding them to the outer membrane translocon is still unknown (179). Substrates for the Type IX secretion system have a conserved C-terminal domain which is important for recognition by the outer membrane porin, although there is a large variety of functional attributions to the proteins secreted by this system including their involvement in nutrition, virulence and motility (180).

1.4.12 Type X secretion system

The Type X secretion system in gram-negative bacteria describes a two-step pathway for transport of substrates using holins (inner membrane) and cell wall editing peptidoglycan hydrolases (muramidases) located in the periplasm to make space for transport across the outer membrane via mechanisms like outer membrane vesicle (OMVs) formation (110). In the *S. marcescens* DB10 model, peptidoglycans are held to the outer membrane through the presence of lipoproteins (110). Here, ChiX is transported into the periplasmic space, where it cleaves cross-bridges. Localisation of this activity allows for subsequent crowding of substrates which can then be transported by the formation of OMVs, as shown in *Salmonella enterica* serovar Typhi (181). Holin/muramidase clusters associated with enterotoxins were also discovered in *E. coli* and *Cronobacter sakazakii* strains (110).

1.5 The accessory secretion system

1.5.1 The accessory secretion system in pathogens

The accessory secretion (aSec) system, also known as "non-canonical secretion system" or "SecA2/Y2" is restricted to gram-positive bacteria and often considered as non-essential for survival (182). This system appears to be specialised for the transport of post-translationally modified, glycosylated cell-wall anchored proteins, as opposed to pre-proteins. However, two versions aSec pathways have been observed so far. Bacteria such as *Mycobacteria*, *Clostridium difficile* or *Listeria sp.*, only possess SecA2, i.e., the paralog of SecA, where it has been

characterised in the context of virulence for targeting proteins that have typical Sec SP, no SP and even Tat SP (183, 184). On the other hand, many *Streptococci (S. parasanguinis, S. gordonii)*, and *Lactobacilli* including *L. reuteri* have a dedicated gene cluster containing SecA2, SecY2 and other accessory secretion proteins. The relationship between these two secretion systems is illustrated in Figure 1.12 (185).
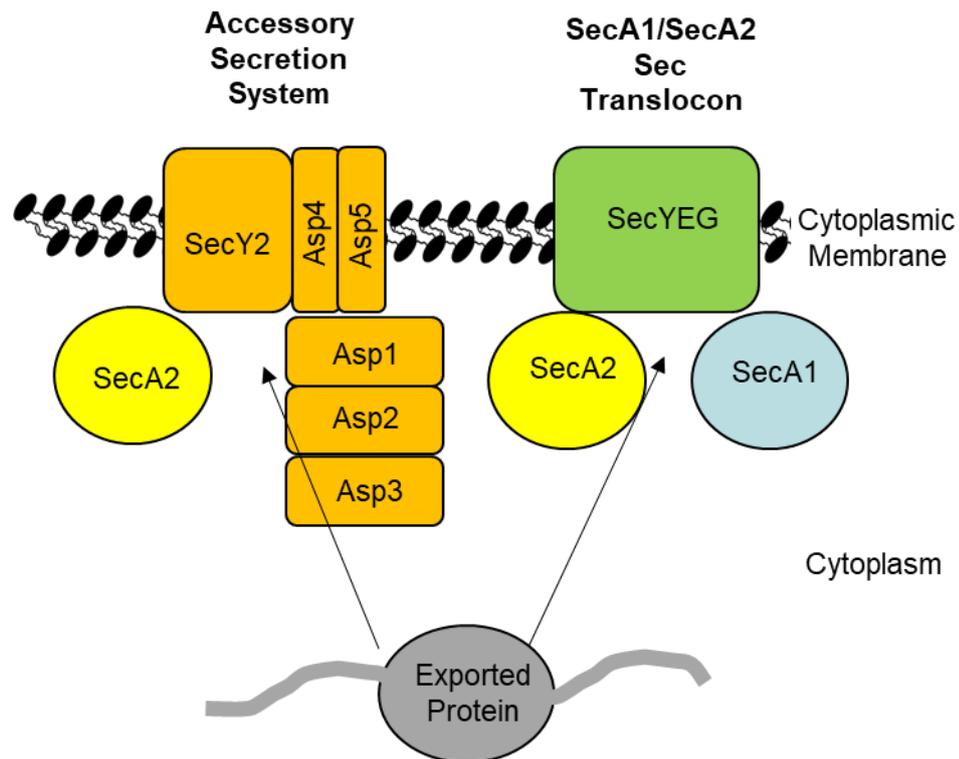
Figure 1.12. **Models for SecA2-dependent export.** In aSec systems, a SecA2-exported protein (shown in grey) is predicted to be exported either through an accessory secretion system translocon or the canonical SecA1/SecYEG translocon with the assistance of SecA2. The example of a translocon is modelled on the SecA2/SecY2 system of *S. gordonii* M99. Adapted from (185).

SecA2 is homologous to SecA from the GSP, whereby sequence alignments and structural modelling predict many conserved domains (185, 186). This was further validated by the resolution of SecA2 crystal structures from *M. tuberculosis* H37Rv (PDB: 4UAQ) and *C. difficile* 630 (PDB: 6SXH) (187, 188). These include a DEAD-like motor domain containing nucleotide binding domains (NBD) and an intramolecular regulator of ATPase (IRA). The ability to hydrolyse ATP has been demonstrated *in vitro* (130, 189). Generally, SecA2 appears to be smaller than SecA due to the truncated C-terminal linker domains and helical wing domains, which are

located near the pre-protein binding cleft in SecA, implicated in substrate recognition. The significance of these truncations is pending further research, however the generation of SecA-SecA2 chimeric proteins has been shown to result in non-functional proteins in *S. gordonii* M99 and DL1 strains (189). SecA2 has also been shown to interact with SP components of the secretion cargo, so these structural differences could be related to substrate recognition specificity (190). This suggests that the aSec system with only SecA2 interacting with the SecYEG channel of the GSP and the aSec system containing the full pathway cluster may be fundamentally different not only in relation to their secretion cargo, but also in their oligomeric interactions.

SecY2 forms the main transmembrane pore and is a paralog of SecY from the GSP. No crystal structure of SecY2 is currently available. Genetic SecY2 knockout strains did not seem to significantly affect the export of glycosylated secretome in *S. aureus* H37rv strain, but this also supports findings that the export of glycosylated products is not entirely dependent on the aSec pathway (191).

Asp4 and Asp5 are accessory secretion proteins that are reminiscent of SecE and SecG, however, they appear to be non-essential and are present variably across aSec gene clusters of gram-positive bacteria, so their roles remain to be unravelled (192).

The secretion-cargo transported by the SecA2-SecY2-Asp4-Asp5 translocation machinery has so far been exclusively identified as large, cell-wall anchored glycosylated adhesins referred to as serine-rich repeat proteins (SRRPs). These proteins are characterised by large serine-rich repeat regions (SRR) flanking a binding region (BR) and the presence of an elongated SP sequence distinguishing them from other secretion systems (108, 193) (Figure 1.13). The role of the extended SP is not fully understood; however, the presence of a polybasic motif and hydrophobic core have been shown to facilitate protein-lipid interactions and thus may aid in the trafficking process (194). The glycosylation profile of SRRs varies depending on the nature and number of glycosyltransferases (GTs) encoded within their respective aSec clusters (195). To date, several SRRPs associated with the aSec system across gram-positive bacteria, mainly in *Streptococcus*

and *Staphylococcus* pathogenic bacteria, have been functionally characterised, revealing a range of binding partners involved in adhesion and biofilm formation (Table 1.1) (108, 196-199).

Figure 1.13 General architecture of SRRPs in gram-positive bacteria. SRRPs transported by the aSec system contain a long signal-peptide (SP) (blue), a short alanine-serine-threonine domain (AST), the first serine rich region (SRR1) (yellow), a binding region (BR) (green), a larger second serine rich region (SRR2) and an LPXTG cell wall anchoring motif.

Table 1.1. Functionally characterised SRRPs from gram-positive bacteria.

SRRP name	Bacterial origin	Binding target
Fap1	*Streptococcus parasanguinis* FW213	Saliva-coated hydroxylapatite, other *S. parasanguinis* FW213 cells (200, 201)
	Streptococcus oralis IE12	Salivary mucin (Mucin 7- MUC7), sialic acid containing glycan epitopes (platelets, oral epithelial cells)(202)
FapA	*Streptococcus oralis* F0392	Inter/intraspecies(203)
FapB		
FapC		Sialic acid containing substrates in saliva (203)
GspB	*Streptococcus gordonii* M99	Human salivary proteins, sialyl-T antigen, glycoprotein Ibα (204, 205)
Hsa	*Streptococcus gordonii* DL1	Sialyl-T antigen, α(2,3) sialyllactosamine, sialic acid-containing MUC7, human salivary proteins, glycoprotein Ibα(204, 205)
SRRP	*Streptococcus sanguinis* SK1	Sialyl-T antigen, sialyl-Lewis X, α(2,3) sialyllactosamine(206)
	Streptococcus sanguinis NCTC10712	Sialyl-T antigen, sialyl-Lewis X, α(2,3) sialyllactosamine(206)
	Streptococcus. sanguinis SK678	Sialyl-Lewis X, α(2,3) sialyllactosamine(206)
	Streptococcus mitis SF100	Sialyl-T antigen(206)
	Limosilacotobacillus. reuteri 100-23	Mouse forestomach stratified squamous epithelium(207)
	Limosilacotobacillus. reuteri ATCC 53608	Porcine gastric mucin, DNA, intestinal epithelium, polygalacturonic acid, rhamnogalacturonan I and chondroitin sulfate A(208)
PsrP	*Streptococcus pneumoniae* TIGR4	Keratin-10, PrsP, extracellular DNA, syalilated ligands on platelets (196, 209)
SraP	*Staphylococcus aureus* COL	Sialylated ligands on platelets (198)
	Staphylococcus aureus ISP479C	
	Staphylococcus aureus PS750	
SRR-1	*Streptococcus agalactiae* CNCTC10/84	Fibrinogen, keratin-4 (210, 211)

SRR-2	*Streptococcus agalactiae* COH1	Fibrinogen (212)
SssP1	*Streptococcus suis* CZ130302	HEp-2, human microvessel endothelial cells, sialylated structures (213)
SrpA	*Streptococcus cristatus* CR311	*Corynebacterium matruchotii* ATCC 11426 (214)
		Fusobacterium nucleatum 364 (214)
	Streptococcus salivarius JIM8777	Inter/Intraspecies (215)
SrpB	*Streptococcus salivarius* JIM8777	Epithelial cells (HT-29, A549, FaDu, HEp-2) (215)
SrpC	*Streptococcus salivarius* JIM8777	Extracellular matrix components (-Mucin2 MUC2, fibronectin, MaxGel ECM, laminin, elastin) (215)

The Asp1-Asp2-Asp3 complex composed of Asp1, Asp2 and Asp3 (or Gap1, Gap2, Gap3 in *S. parasanguinis*FW213) are defined as chaperones for the fully glycosylated substrates and associate with the cell membrane and SecA2, and thus are predicted to aid in the translocation process (216). The crystal structures of Asp1, Asp1 in complex with Asp3, and Asp1 in complex with Asp2 and Asp3 have been solved by X-ray crystallography from *S. gordonii* M99 and *S. pneumoniae*TIGR4 (see Table 1.2) (217). These data suggest that Asp1 and Asp3 are structurally related to carbohydrate-binding proteins. Asp1 appears to be a catalytically inactive member of the GT-B family of glycosyltransferases, and Asp3 is similar to the carbohydrate binding module (CBM) domain of carbohydrate-active enzymes (classified in the CAZy database, http://www.cazy.org/Carbohydrate-Binding-Modules.html) (218). The presence of carbohydrate-binding motifs in Asp1 and Asp3 is in agreement with the prediction that substrates of aSec are exclusively glycosylated proteins. These binding interactions were highlighted to be fundamental for the transport of GspB, the SRRP from *S. gordonii* M99, with site-directed mutagenesis studies revealing Y131 in *S. gordonii* M99 Asp3 (*Sg*Asp3$_{M99}$) as a critical residue for secretion (218). Additionally, the Asp1-Asp2-Asp3 complex from *S. gordonii* M99 (*Sg*Asp1-Asp2-Asp3$_{M99}$) complex has been shown to interact with phospholipids when incubated with liposomes (217). On the other hand, pulldown assays and size-exclusion gel filtration between Asp1 and Asp3 with secretion substrates have failed to show strong interactions so far (219). Although a direct binding of Asps to carbohydrates of secretion cargo has not yet been shown, the low-binding affinity can

be characteristic of CBMs (220). The Asp1-Asp2-Asp3 complex from *S. pneumoniae* TIGR4 (*Sp*Asp1-Asp2-Asp3$_{TIGR4}$) complex has been shown to increase ATPase activity of *Sp*SecA2$_{TIGR4}$ in *in vitro* assays (216).

The Asp2 protein of *S. gordonii* M99 strain (*Sg*Asp2$_{M99}$) has been determined to be an acetyltransferase that modifies *O*-GlcNAc moieties on SRR domains of GspB (221). Asp2 has a C-terminal serine-based catalytic triad reminiscent of a diverse group of enzymes, including esterases, proteases, lipases, and glycoside hydrolases (150). Whilst targeted mutations of the *Sg*Asp2$_{M99}$ catalytic domain had no effect on transportation, it abolished acetylation and consequent transportation was rerouted via the canonical system instead of the aSec pathway (221). *S. gordonii* M99 lacking *O*-acetylated GspB had significantly reduced virulence and platelet binding capability and transportation was rerouted to the other secretion pathways instead of the aSec system (221). Further evidence to suggest that Asp2 acts as an acetyltransferase is that the SRR protein cargo of *Streptococcus salivarius* JIM8777 aSec system has acetylated *O*-GlcNAc moieties (221). In summary, Asp2 appears to be critical to the aSec transport of secretion targets and the biological function of the substrate, but the precise catalytic steps mediated through the enzymatic activity of Asp2 remain to be elucidated including presence of cofactors, or its protein complex partners Asp1 and Asp3.

Taken together, the pathogen models show that SRRPs relevant for virulence and attachment to host components are transported by a dedicated aSec system. These adhesins are modified by glycosyltransferases and Asp1-Asp2-Asp3 before reaching the membrane-localised SecA2-SecY2-Asp4-Asp5 translocation complex. Disruption of the secretion system has adverse consequences on post-translational modifications (glycosylation, acetylation) of the secretion cargo and subsequent pathogenicity (193, 222). Table 1.2 summarises some of the existing structural information available for aSec systems in gram-positive bacteria. While the secretion target SRRPs and their binding regions (BRs) have been characterised in *L. reuteri* ATCC 53608 and 100-23 strains, the structural components of the aSec system, SecA2-Y2 and Asps have only been studied in pathogenic bacteria.

Table 1.2. Summary of aSec proteins discovered in gram-positive bacteria. BR refers to binding region of the secretion cargo SRRPs. References are provided to studies reporting the discovery of aSec proteins (from SecA2Y2 and SecA1SecA2 systems) in bacteria and to those which have been structurally characterised.

Bacterial Family	aSec Protein	PDB
Aerococcaceae including: *Ignavigranum sp.* (223)		
Bacillacea including: *Bacillus sp.* (224) *Lycinibacillus sp.*(225) *Parageobacillus sp.* (226)		
Clostridiaceae Clostridium sp. (186)	CdSecA2$_{630}$ (188)	6SXH, 6T4H
Corynebacteriaceae Corynebacterium sp.(227)		
Enterobacteriaceae Enterococcus sp. (228)		
Lactobacillaceae Including: *Leuconostoc sp.,* (229) *Lactobacillus sp.,*(195) *Limosilactobacillus sp.*(195)	LrSRRP-BR$_{53608}$ (208)	5NXK
	LrSRRP-BR$_{100-23}$ (208)	5NY0
Listeriaceae Including: *Listeria sp.*(230)		
Mycobacteriaceae Including: *Mycobacterium sp.* (187)	MtSecA2$_{H37rv}$ (137)	4AUQ
Staphylococcaceae including: *Staphylococcus sp.* (231, 232) *Macrococcus sp.* (232)	SaSraP-BR$_{NCTC8325}$ (161)	4M00, 4M02
Streptococcaceae including: *Streptococcus sp.* (197, 222)	SpPsrP-BR$_{TIGR4}$.(141)	3ZGH
	SpAsp1-Asp2- Asp3$_{TIGR4}$ (216).	6LNW
	SgAsp1-Asp3$_{M99}$ (217)	5VAE
	SgAsp1$_{M99}$ (217)	5VAF

1.5.2 The accessory secretion system in *L. reuteri*

L. reuteri aSec pathway clusters have been identified in the genomes of strains isolated from murine and porcine hosts, however no aSec pathway appears to be present in human isolates. *L. reuteri* clusters in these strains contain the *SRRP, SecA2, SecY2, Asp1, Asp2, Asp3*, a varying number of glycosyltransferases (**gtfs**), but all contain *GtfA, GtfB and GtfC* (Figure 1.14) (233).
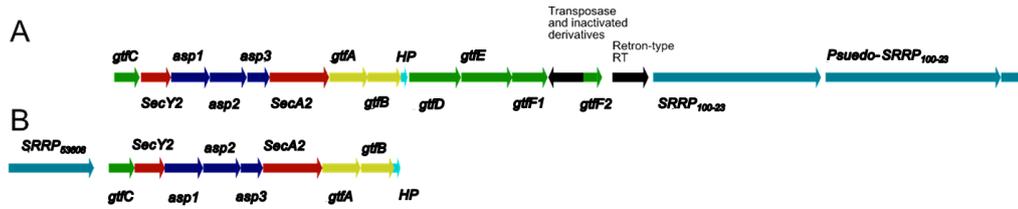
Figure 1.14. Organisation of aSec clusters from *L. reuteri* strains. A) *L. reuteri* 100-23 aSec cluster, (B) *L. reuteri* ATCC 53608 a Sec cluster. The genes encoding the SecA2/Y2 translocation machinery are shown in red, the accessory secretion proteins Asp1-5 in blue and the priming GTs, GtfA and GtfB. Genes encoding additional GTs are shown in green and the genes encoding SRRPs are in teal. Black arrows represent genes that are not part of the aSec machinery. HP: hypothetical protein. Adapted from (233).

The SecA2/Y2 cluster and SRRP in the murine isolate *L. reuteri* 100-23 have been shown to be critical for adhesion of the bacteria to the forestomach epithelium of the murine GI tract, as shown by colonisation experiments in mice with *L. reuteri* 100-23C (a plasmid-free derivative strain of 100-23) wild-type and mutants (Figure 1.15) (207). Mutants lacking the *srrp* gene (annotated as Lr_70902) showed the most pronounced reduction in colonisation, compared to other targeted adhesin genes tested. Mutants of the s*ecA2* gene showed defective adhesion that was not completely eliminated, suggesting that SRRPs may still be secreted through other means (207). This is in agreement with aSec-alternative secretion of GspB in *S. gordonii* M99 strain, albeit with less efficiency or to fitness detriment (234).
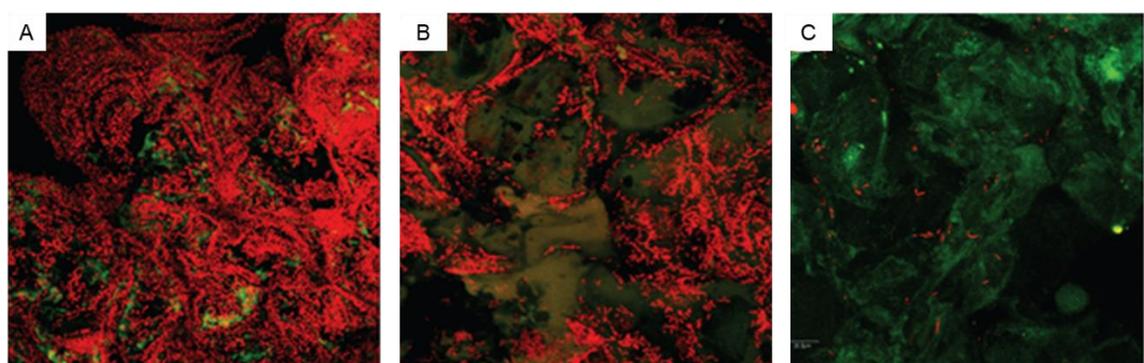
Figure 1.15. Characterization of *in vivo* biofilms of *L. reuteri* 100-23C variants. Confocal micrographs of forestomach tissue from mice colonized with wild type and mutant *L. reuteri* 100-23C two days after gavage with 10^7 cells (A) wild type, (B) *secA2* mutant, and (C*) lr70902 (SRRP)* mutant. Taken from (207).

SRRPs from *L. reuteri* ATCC 53608 strain (isolated from pig) and 100-23 strain (isolated from mouse) have recently been characterised in terms of structure, glycosylation, and role in adhesion/ host specificity (195). SRRPs are typically composed of several discrete subdomains: a long N-terminal SP of around 90 amino acids, followed by an alanine-serine-threonine (AST) domain, a short serine-rich repeat region (SRR-1), a binding region (BR), a longer second SRR (SRR-2) and a C-terminal cell wall anchoring motif (LPTXG) (Figure 1.16) (208). The crystal structures of BRs from SRRPs of *L reuteri* ATCC 53608 and 100-23 strains have been determined. They display a β-solenoid fold formed of repetitive right-handed helix of parallel β-sheets (Figure 1.17). In addition, SRRP-BR$_{53608}$ was shown to bind to the epithelium and to dietary components in a pH-dependent manner (208). Phylogenetic analyses of *L. reuteri* SRRP-BRs based on amino acid alignments showed one group consisting mainly of SRRPs from porcine isolates that included strain ATCC 53608, but also rodent isolate 100–23, a second group mainly of pseudo-SRRPs from rodent and sourdough strains and a third group of pseudo-SRRPs of porcine origin (108).
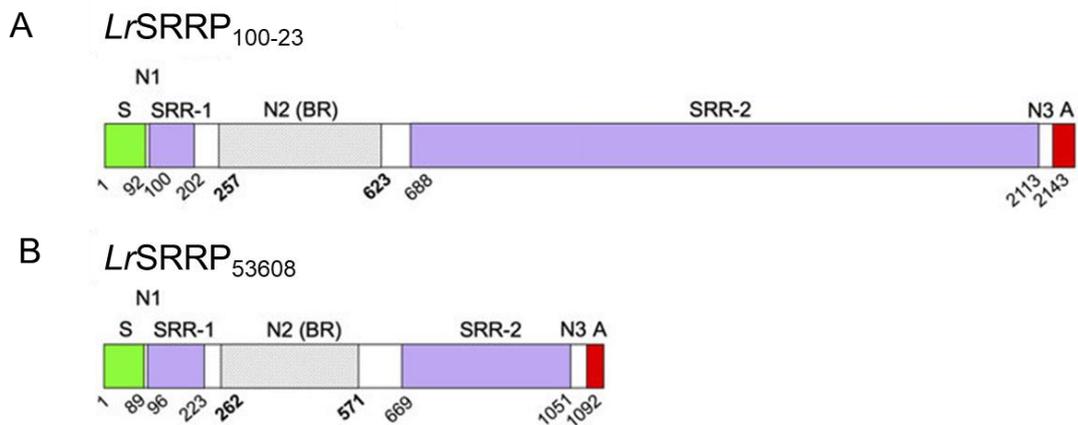
Figure 1.16. Schematic domain organisation of *L. reuteri* SRRPs. (A) *L. reuteri* 100-23 SRRP (*Lr*SRRP$_{100-23}$), (B) *L. reuteri* ATCC 53608 SRRP (*Lr*SRRP$_{53608}$). Domains are labelled as follows: A, cell wall anchor including LPXTG motif; N1, nonrepeat region 1; N2 (BR), nonrepeat region 2 (putative binding region); N3, nonrepeat region 3; S, secretion signal sequence; SRR-1, serine-rich region 1; SRR-2, serine-rich region 2. The beginning amino acid position is indicated below each domain. Regions of the BR that were resolved by crystallography are shaded grey and span amino acids 257-623 for *Lr*SRRP$_{100-23}$ and amino acids 262-571 for *Lr*SRRP$_{53608}$. Taken from (208).
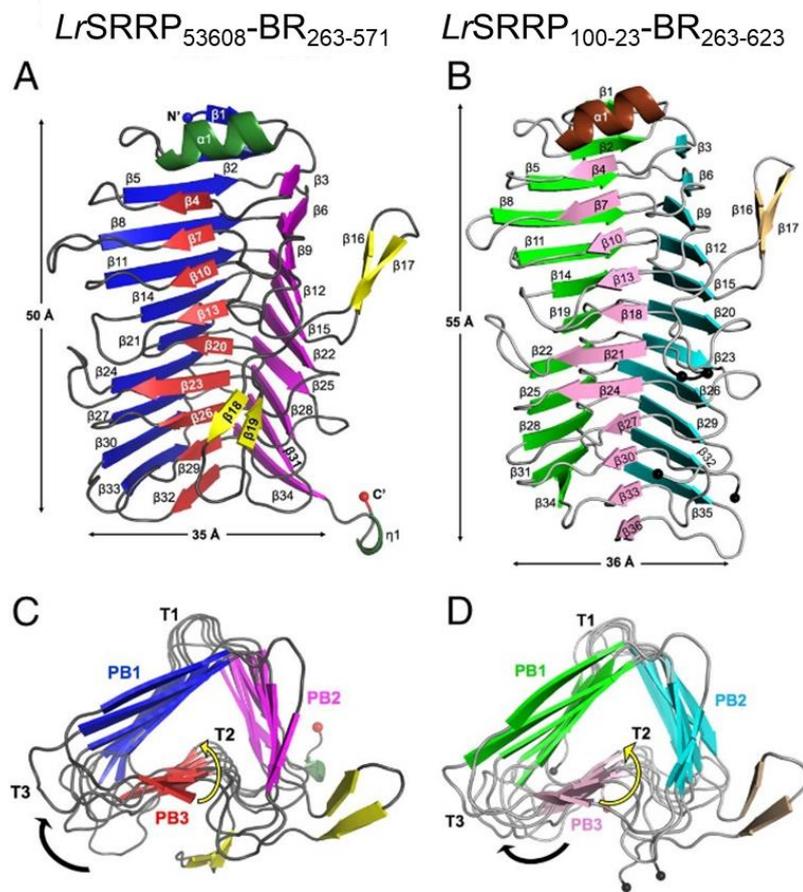
Figure 1.17. Crystal structures of *Lr*SRRP$_{53608}$-BR$_{262–571}$ and *Lr*SRRP$_{100-23}$-BR$_{257–623}$. (A) Cartoon representation of *Lr*SRRP$_{53608}$-BR$_{262–571}$. The β-strands in putative binding regions PB1, PB2, and PB3 are shown in deep blue, magenta, and red, respectively; α-helices are dark green, and β-strands of the loop are yellow. (B) Cartoon representation of *Lr*SRRP$_{100-23}$-BR$_{257–623}$. PB1, PB2, and PB3 β-strands are in light green, cyan, and pink, respectively. The α-helix is in brown, and loop β-strands are in beige. The black spheres indicate the gaps in the model between amino acids 413–421 and 568–583. (C) Cartoon cross-section of *Lr*SRRP$_{53608}$-BR$_{262–571}$. (D) Cartoon cross-section of *Lr*SRRP$_{100-23}$-BR$_{257–623}$ showing β-solenoid superhelices along the helical axis from the N to the C terminal. In both, β1 and α2 are omitted for clarity. The black arrow is the direction in which the polypeptide chains fold around the helical axis, showing a right-handed superhelix. The helical twist down each β-sheet is indicated by yellow arrows. Along the helical axis, the β-strands in each parallel β-sheet increasingly twist toward the left with respect to each other. Taken from (208).

SRRPs from *L. reuteri* 100-23 and ATCC 53608 strains have been shown to be differentially glycosylated (235). *O*-glycosylation is initiated by the two glycosyltransferases, *Lr*GtfA and *Lr*GftB, which form a *Lr*GtfAB complex mediating the addition of *N*-acetylglucosamine (GlcNAc) to Ser/Thr residues in the SRR domains. In *Lr*SRRP$_{100-23}$, this is followed by addition

of Glc by $LrGtfC_{100-23}$ whereas $SRRP_{53608}$ is extended by another GlcNAc by $LrGtfC_{53608}$, reflecting the ligand preference of $LrGtfC_{100-23}$ and $LrGtfC_{53608}$ for UDP-Glc and UDP-GlcNAc, respectively (235). *L. reuteri* 100-23 aSec system also has additional glycosyltransferases (*Lr*GtfE/GtfD) which may lead to further elongation of the *O*-glycan chain (233, 235). Furthermore, *in vivo* glycoengineering in *E. coli* led to glycosylation of $LrSRRP_{53608}$ variants with α-GlcNAc and GlcNAcβ (1→6) GlcNAcα moieties (235). Together these data indicate that $LrSRRP_{100-23}$ and $LrSRRP_{53608}$ are modified with Hex-Glc-GlcNAc and di-GlcNAc moieties, respectively (Figure 1.18). It is then expected that *Lr*Asp1-Asp2-Asp3 chaperones modify SRRPs bringing them to the secretion complex formed by a *Lr*SecA2-SecY2-Asp4 complex for export across the cell membrane as proposed in Figure 1.18, although the mechanisms of transport and characterisation of the protein complexes mediating these steps remain to be demonstrated. It was hypothesised that the strain-specific glycosylation of SRRP adhesins could contribute to *L. reuteri* host-adaptation strategy.
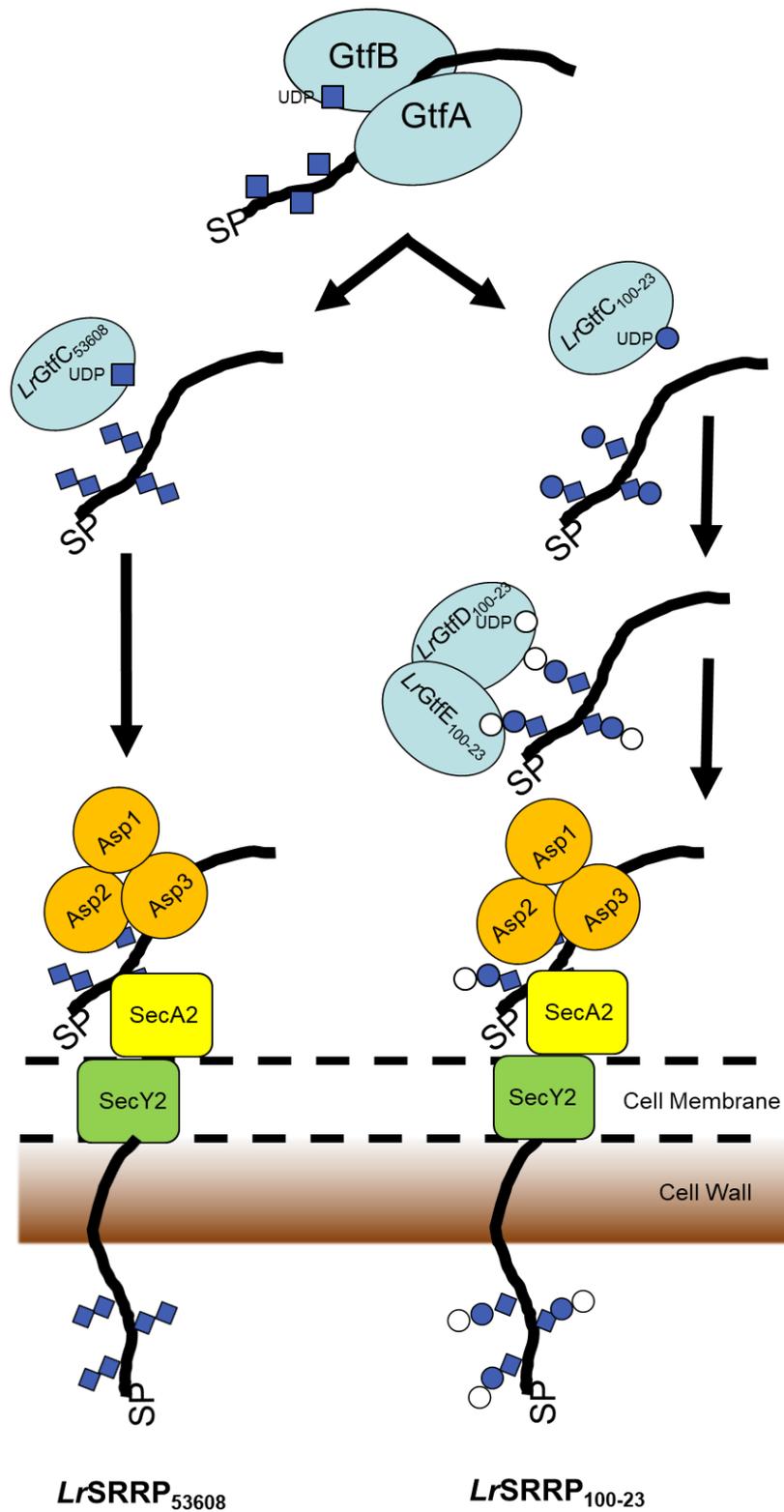
Figure 1.18. Glycosylation mechanisms of SRRPs from *L. reuteri* strains ATCC 53608 and 100-23. The *Lr*GtfA/B complex initiates the glycosylation of the *L. reuteri* SRRPs with GlcNAc residues, while *Lr*GtfCs extends the glycans with either GlcNAc (*Lr*SRRP$_{53608}$) or Glc (*Lr*SRRP$_{100-23}$). The glycosylated *Lr*SRRP$_{53608}$ is then secreted through the aSec system, whereas the LrSRRP$_{100-23}$ is further glycosylated by *Lr*GtfD/E before secretion. Adapted from (233).

Aim and objectives

The overall aim of this PhD project is to characterise the accessory secretion (aSec) system in the gut symbiont *L. reuteri*. Beyond being a model vertebrate symbiont to study host-strain adaptation, *L. reuteri* also has beneficial applications across species. The aSec system dedicated to the translocation of large-cell wall anchored adhesins called SRPPs, has primarily been investigated in human pathogenic bacteria in the context of virulence. Thus far, work related to *L. reuteri* aSec system has focused on the characterisation of the secretion cargo SRRP, while the cytoplasmic sequence of glycan modification, translocation and the molecular mechanisms underpinning these events are yet to be elucidated.

To address this knowledge gap, the aSec components of *L. reuteri* ATCC 53608 and 100-123 strains from porcine and rodent hosts, respectively, have been heterologously expressed and purified individually or in complex for (1) biochemical characterisation using enzymatic activity and binding assays and (2) structural analysis through X-ray crystallography, Alphafold2 and small-angle x-ray scattering (SAXS). Together these approaches provided new structural insights into the function of aSec systems in gut symbionts.

Chapter 2 Materials and Methods

2.1 Bacterial strains and culture conditions

2.1.1 *Limosilactobacillus reuteri*

The *L. reuteri* strains used in this work, ATCC 53608 and 100-23, were grown aerobically at 37°C statically and light protected using de Man, Rogosa and Sharpe (MRS) or *Lactobacillus* defined medium type II (LDMII) broth (Table 2.1). MRS was composed of 10 g proteose peptone no.3, 10 g beef extract, 5 g yeast extract, 20 g dextrose, 1 g polysorbate 80, 2 g ammonium citrate, 5 g sodium acetate, 0.1 g magnesium sulphate, 0.05 g manganese sulphate, 2 g dipotassium phosphate, dissolved in 1 L of water and sterilised by autoclave at 121°C for 30 min at 15 pounds per square inch (PSI). LDMII was composed of 1.5 g K_2HPO_4, 1.5 g KH_2PO_4, 15 g sodium acetate, 0.22 g sodium citrate, 50 mg tryptophan, 50 mg, asparagine, 50 mg cysteine, 50 mg glycine, 50 mg serine, 50 mg alanine, 50 mg phenylalanine, 50 mg histidine, 50 mg isoleucine, 50 mg methionine, 50 mg proline, 50 mg threonine, 50 mg valine, 50 mg valine, 50 mg tyrosine, 50 mg leucine, 50 mg glutamine, 50 mg aspartic acid, 50 mg glutamic acid, 0.2 mg thiamine-HCl, 0.04 mg para-aminobenzoic acid, 0.4 mg calcium pantothenic acid, 1.0 mg niacin, 0.5 mg pyridoxine-HCl, 0.05 mg biotin, 0.1 mg folic acid, 0.4 mg riboflavin, 10 mg adenine sulphate, 20 mg uracil, 10 mg guanine-HCl, 50 mg cytidine, 1.6 µg thymidine, 1.0 mL Tween-80, 0.164 g $MgSO_4.H_2O$, 23.4 mg $MnSO_4.7H_2O$, 13 mg $FeSO_4.7H_2O$, and 30 g sucrose was dissolved in 1 L of water and sterilised by filtration through a 0.22 µM filter (236).

Table 2.1. *Limosilactobacillus reuteri* **strains and culture conditions.**

L. reuteri strain	origin	starter culture conditions	growth culture conditions
ATCC 53608	porcine intestine, www.ATCC.org	de Man, Rogosa and Sharpe (MRS) broth	*Lactobacillus* defined medium type II (LDMII) (236)
100-23	mouse forestomach (207)		

2.1.2 *Escherichia coli*

The *E. coli* strains used in this work were DH5α (New-England Biolabs) and STELLAR (TakaraBio) for plasmid propagation and cloning, as well as BL21(DE3) (New-England Biolabs) for heterologous expression of target aSec proteins (Table 2.2). *E. coli* strains and derivatives

were grown in Luria-Bertani (LB) (Miller) broth composed of 10 g peptone, 5 g yeast extract and 10 g sodium chloride dissolved in 1 L water and sterilised by autoclave at 121°C for 30 min at 15 PSI. Auto-induction medium (AIM) used for the heterologous expression of recombinant aSec proteins was AIM-LB Broth Base including Trace elements (Formedium) composed of 10 g tryptone, 5 g yeast extract, 3.3 g ammonium sulphate, 6.8 g potassium dihydrogen phosphate, 7.1 g disodium hydrogen phosphate, 0.5 g glucose, 2.0 g lactose, 0.15 g magnesium sulphate and 0.03 g trace elements, dissolved in 1 L of water and sterilised by autoclave at 121°C for 30 min at 15 PSI. Antibiotics were used depending on antibiotic resistance genes introduced by plasmid vectors with carbenicillin used at 100 µg/mL, kanamycin at 50 µg/mL and chloramphenicol at 34 µg/mL final concentrations (as listed in Table 2.4).

Table 2.2. *Escherichia coli* **strains and culture conditions.**

E. coli strain	origin	starter Culture Conditions	growth culture conditions
DH5α	NEB (international.NEB.com)	LB ± antibiotics	LB ± antibiotics
STELLAR	TakaraBio (www.takarabio.com)	LB± antibiotics	LB ± antibiotics
BL21(DE3)	NEB (international.NEB.com)	LB± antibiotics	AIM (Formedium) ± antibiotics

2.2 Cloning, site-directed mutagenesis, and heterologous expression

2.2.1 Genomic DNA purification

Genomic DNA (gDNA) was extracted from *L. reuteri* ATCC 53608 and 100-23 strains grown overnight using the Monarch® Genomic DNA Purification Kit (New-England Biolabs) as per manufacturer's instructions for gram-positive bacteria. Briefly, *L. reuteri* cells were harvested by centrifugation at 12,000 rpm for 1 min. The supernatant was discarded, and the pellet resuspended in 10 mM Tris-HCl pH 8.0 by vortexing. A lysozyme (Merck) solution of 25 mg/mL was prepared and 20 µl was added to the *L. reuteri* cell suspension along with 100 µl of Monarch® tissue lysis buffer. Following incubation at 37°C until the suspension became clear. 10 µL Proteinase K (New-England Biolabs) was added and incubated at 56°C for 45 min in a thermal mixer

(Thermofisher) at 1000 rotations per min (rpm). Then 3 μL of RNase A (New-England Biolabs) was added and incubated at 56°C for 10 min in a thermal mixer (Thermofisher) at 1000 rpm. The resulting lysate was then mixed with Monarch® gDNA Binding Buffer and purified with dedicated Monarch® gDNA purification columns. *L. reuteri* gDNA was eluted from the column using TE buffer (10 mM Tris-HCl pH 8.0 with 1 mM EDTA) and quality and quantity analysed by Nanodrop™ 2000 (Thermofisher).

2.2.2 Cloning and heterologous expression

Primers were designed for the amplification of DNA encoding aSec protein targets from *L. reuteri* ATCC 53608 and 100-23 genomic DNA by polymerase chain reactions (PCR) for the purpose of cloning into multiple-cloning sites of expression vector plasmids (Table 2.3). Primer synthesis and quality assurance was carried out by Eurofins Genomics (https://eurofinsgenomics.eu/).

Table 2.3. Primers used in cloning of aSec components.

Primer (5'→3')	Template	Annealing temperature °C	Target	Restriction site
			Cloning vector	
F': ccgcgcggcagccatgattatttagttcctgcttgg		63	Asp1	NdeI
R': gtgcggccgcaagcttcacttcgttaaaactttttcc			pET28b	HindIII
F': ccgcgcggcagccataatgtcttacagttaggc		62	Asp2	NdeI
R': gtgcggccgcaagctttattctcctcgtccaaaagc			pET28b	HindII
F': ccgcgcggcagccatatgcagcagttttattaatcaagg		61	Asp3	NdeI
R': gtgcggccgcaagcttaaattcaaaagagggataa			pET28b	HindIII
F': gcgcggcagccataatgattatttagttcctgcttgg		60	Asp1-Asp2-Asp3	NdeI
R': taaattcaaaagagggataaaattcaaaagagggataa			pET15b	EcoRI
F': aggagatataccatgcattgaaaatcg	*L. reuteri* ATCC 53608 gDNA	58	Asp4	EcoRI
R': taaattcaaaagaggattcggaggctgaacta			pET15b	BamHI
F': cgcgcggcagccatatgccaaatttctc		59	SP	NdeI
R': taaattcaaaagaggattctatataattctctatcgagtat			pET15b	HindIII
F': ccgcgcggcagccatgctgacatttaatctaatatgg		62	Asp2	NcoI
R'gtgcggccgcaagctattctaaattactacttgtgagc			pET28b	AvrII
F': accacagccaggatcatgctgaaactctatactct		58	SecY2	BamHI
R': tgcggccgcaagctattctatatgtgattacattta			prSFduet-1 MCS 1	HindIII
F': accaggacatacatgctgtgcaaatatttacgc		63	SecA2	NdeI
R'gtagcgtgccgtaatatattctgtctatttacgcggacgg			pRSFduet-1 MCS 2	AvrII
F': ccgcgcggcagccatacgtaccccaattttaggttac		62	SecA2	NdeI
R': gtgcggccgcaagctttaattaaaatgaacaattaa			pET28b	HindIII
F': ccgcgcggcagccatatgctatattaaggtctctatagg		59	Asp2	NdeI

R' gtgcggccgcaagctattctgcaactgacagtctcc	*L. reuteri*		pET28b	HindIII
F': ccgcgcgggcagccatatgctagtgcttactctagg	100-23	60	SecA2	EcoRI
R': gtgcggccgcaagctattctttggacatctgacc	gDNA		pET28b	HindIII

PCR reactions (25 µL) contained 12.5 µL Q5 Hi-Fi 2X Master Mix (New England Biolabs), 1.25 µL 10 µM forward primer, 1.25 µL 10 µM reverse primer, 1 µL genomic template DNA and nuclease-free water. PCR was performed in a T100™ Thermal Cycler (Bio-Rad) with the following steps: initial denaturation at 98°C for 30 sec; 30 amplification cycles with 10 sec denaturation at 98°C, annealing for 20 sec at temperatures predicted using New England Biolabs web-tool (http://tmcalculator.neb.com/#!/main) and extension time of 20 sec/kb at 72°C; followed by a final extension for 2 min also at 72°C.

The PCR products encoding the aSec components from *L. reuteri* ATCC 53608 and *L. reuteri* 100-23 were then cloned into pET15b, pRSFduet-1, pET28b, pET28a, pETcoco-1 and pOPINF vectors (Table 2.3). A full list of plasmids, inserts, purification fusion-tags, protease cleavage sites and antibiotic-resistance cassettes used for this work is provided in Table 2.4. These plasmids allow the fusion of the insert with an affinity purification tag to facilitate further purification, and expression of the recombinant protein is under control of a lactose (lac) inducible promoter. Briefly, the plasmids were linearised using restriction enzymes from New England Biolabs according to manufacturer's instructions at restriction sites listed in Table 2.3. The digestion products were confirmed by electrophoresis using agarose gels and then purified using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs).

The In-fusion HD cloning plus (Takara) was used to generate the recombinant plasmids according to the manufacturer's instructions. Briefly, 2 µL of the In-fusion HD enzyme premix (Takara) was added to the linearised plasmid (at 50-200 ng) and PCR-amplified insert (10-200 ng) in a final volume of 10 µL. DNA concentrations were estimated using NanoDrop™ 2000 (Thermofisher). The suspension was then used to transform competent *E. coli* Stellar (TakaraBio) cells. Briefly, approximately 5 ng of DNA was added to the thawed Stellar cells and incubated on ice for 30 min. The cells were then heat-shocked for 45 sec at 42°C. A volume of 900 µL SOC (Super Optimal Broth with Catabolite Repression), was added to the transformation reaction to a

final volume of 1 mL and incubated with shaking at 160 rpm for 1 hr at 37°C. 1 L of SOC contains 20 g tryptone, 5 g yeast extract, 4.8 g MgSO$_4$, 3.6 g dextrose, 0.5 g NaCl and 0.18 g KCl. An aliquot of 100 µL was plated on selective medium with appropriate antibiotics (Merck) depending on the resistance cassette of the plasmid (see Table 2.4). Plates were incubated at 37°C overnight.

Recombinant colonies were grown in LB liquid culture at 37°C shaking for 4-5 hr. Plasmids were extracted using the Monarch® Plasmid Miniprep Kit (New England Biolabs). The presence of the gene was first determined by analytical digest followed by agarose gel electrophoresis and the integrity of the insert confirmed by sequencing using the Mix2Seq kit (Eurofins) with T7 promoter specific primers "F': TAATACGACTCACTATAGGG".

Once validated for sequence integrity, recombinant plasmid DNA (3 ng/µL for single-plasmid and 1 ng/µL for co-transformation) was used to transform *E. coli* Bl21(DE3) (Novagen). Briefly, the suspension was incubated on ice for 30 min before heat-shocking at 42°C for 10 sec. The transformation reaction was completed to 1 mL with SOC medium and incubated at 37°C shaking for 1 hr before plating on selective media.

Table 2.4. Summary of recombinant plasmids encoding *Lr*aSec components.

Plasmid(s)	Insert(s)	Tags and Resistance Casettes	Plasmid Map
pOPINF-*Lr*GtfC$_{53608}$*	*Lr*GtfC$_{53608}$	H6-tag, Amp^R, C3-cleavage site	Supplementary Figure S1.1
pOPINF-*Lr*GtfC F174L$_{53608}$	*Lr*GtfC F174L$_{53608}$		
pOPINF-*Lr*GtfC$_{53608}$ P238S	*Lr*GtfC$_{53608}$ P238S		
pOPINF-*Lr*GtfC C240W$_{53608}$	*Lr*GtfC C240W$_{53608}$		
pOPINF-*Lr*GtfCD101A$_{53608}$	*Lr*GtfCD101A$_{53608}$		
pET28a-*Lr*GtfC$_{100-23}$*	*Lr*GtfC$_{100-23}$	H6-tag, Kan^R, thrombin cleavage site	Supplementary Figure S1.2
pET28a-*Lr*GtfC L174F$_{100-23}$	*Lr*GtfC L174F$_{100-23}$		
pET28a-*Lr*GtfC S238P$_{100-23}$	*Lr*GtfC S238P$_{100-23}$		
pET28a-*Lr*GtfC W240C$_{100-23}$	*Lr*GtfC W240C$_{100-23}$		
pET28a-*Lr*GtfCD101A$_{100-23}$	*Lr*GtfC D101A$_{100-23}$		
pET28b-*Lr*Asp1$_{53608}$	*Lr*Asp1$_{53608}$		Supplementary Figure S1.3
pET28b- *Lr*Asp2$_{53608}$	*Lr*Asp2$_{53608}$		Supplementary Figure S1.4
pET28b-*Lr*Asp2S349A$_{53608}$	*Lr*Asp2S349A$_{53608}$		

pET28b- $LrAsp2_{100-23}$	$LrAsp2_{100-23}$		Supplementary Figure S1.5
pET28b-$LrAsp3_{53608}$	$LrAsp3_{53608}$		Supplementary Figure S1.6
pET15b- LrAsp1-Asp2-Asp3$_{53608}$	LrAsp1-Asp2-Asp3$_{53608}$	H6-tag, Amp^R	Supplementary Figure S1.7
pET28b-LrSecA2$_{53608}$	LrSecA2$_{53608}$	H6-tag, Kan^R, thrombin cleavage site	Supplementary Figure S1.8
pET28b-LrSecA2K114A$_{53608}$	LrSecA2K114A$_{53608}$		
pET28b-LrSecA2$_{100-23}$	LrSecA2$_{100-23}$		Supplementary Figure S1.9
pET28b-LrSecY2$_{53608}$	LrSecY2$_{53608}$		Supplementary Figure S1.10
pET15b- LrAsp4$_{53608}$	LrAsp4$_{53608}$	Amp^R	Supplementary Figure S1.11
pET15b- $LrSP_{53608}$	$LrSP_{53608}$	H6-tag, Amp^R, thrombin cleavage site	Supplementary Figure S1.12
pET15b-LrSRR1$_{100-23}$*	LrSRR1$_{100-23}$	H6-tag, Amp^R,	Supplementary Figure S1.13
pETcoco-1-LrGtfA-GtfB-GtfC$_{53608}$*	LrGtfA-GtfB-GtfC$_{53608}$	Chl^R	Supplementary Figure S1.14
pRSFDUET-1-LrSecA2-SecY2$_{53608}$	LrSecA2-SecY2$_{53608}$	H6-tag, Kan^R	Supplementary Figure S1.15
pRSFDUET-1-LrGtfA-GtfB$_{53608}$*	LrGtfA-GtfB$_{53608}$	Kan^R	Supplementary Figure S1.16

*Plasmids were generated by Dr Dimitris Latousakis in (235).

2.2.3 Site-directed mutagenesis

Site-directed mutagenesis for *LrGtfCs*, *LrSecA2*$_{53608}$ and *LrAsp2*$_{53608}$ was carried out by PCR using a method adapted from Edelheit *et al.* (2009) (Figure 2.1) (237) . Briefly, forward, and reverse primers listed in Table 2.5 for mutagenesis targets were used in separate single-primer reactions in parallel using high-fidelity polymerase Platinum SuperFI II (Thermofisher), which using an annealing temperature of 60°C resulted in PCR products containing parental plasmid DNA, the 5'-3' and 3'-5' amplicons were then combined and mixed by pipette aspiration. The mixture was then denatured at 95°C for 5 min, and slowly cooled down in a sequential thermal ramp from 90°C to 25°C (90°C for 2 min, 80°C for 2 min, 70°C for 2 min, 60°C for 2 min, 50°C for 2 min, 40°C for 2 min, 30°C for 2 min and 25°C for 2 min), promoting random annealing of plasmid strands. The products were digested with DpnI removing methylated parental template strands leaving only the annealed, mutated plasmids. The digestion product was used to transform *E. coli* as described above (see section 2.2.2). All reagents and restriction enzymes were from New England Biolabs.

Table 2.5. Primers for site-directed mutagenesis of *Lr*aSec proteins.

Primer	Template	Annealing Temperature °C	Site-Directed Mutagenesis Target
F:ggtaattaatcttgctggtaacc	*Lr*GtfC$_{53608}$-pOPINF	60	*Lr*GtfC F174L$_{53608}$
R'ccaatggtcgttctaattaatgg			
F':gtcaaatgagagttattgtaaaa			*Lr*GtfC P238S $_{53608}$
R'ttttacaataactctcatttgac			
F':tgagccatattggaaaaactacatg			*Lr*GtfC C240W$_{53608}$
R' catgtagttttttccaatatggctca			
F:gtttatgggcgcgcaatcagatcc			*Lr*GtfC D101A$_{53608}$
R'ggatctgattgcgcgcccataaac			
F':ggtaattaattttgctggtaacc	*Lr*GtfC$_{100-23}$-pET28a		*Lr*GtfC L174F$_{100-23}$
R': ggttaccagcaaaattaattacc			
F': tggtcaaatgagccatattggaaaaac			*Lr*GtfC S238P$_{100-23}$
R': gttttccaatatggctcatttgacca			
F': tggtttagtatgttcaaatgagtc			*Lr*GtfC W240C$_{100-23}$
R': gactcatttgaacatactaaacca			
F': agtttatgggctggcaatcagatcc			*Lr*GtfC D101A$_{100-23}$
R': ggatctgattgccagcccataaact			
F': ttcatcgctagcctgcgcttatcat	*Lr*Asp2$_{53608}$-pET28b		*Lr*Asp2 S349A$_{53608}$
R': atgataagcgcaggctagcgatgaa			
F': aacatgcgtgcgccatctgactgc	*Lr*SecA2$_{53608}$-pET28b		*Lr*SecA2 K114A$_{53608}$
R': gcagtcagatggcgcacgcatgtt			
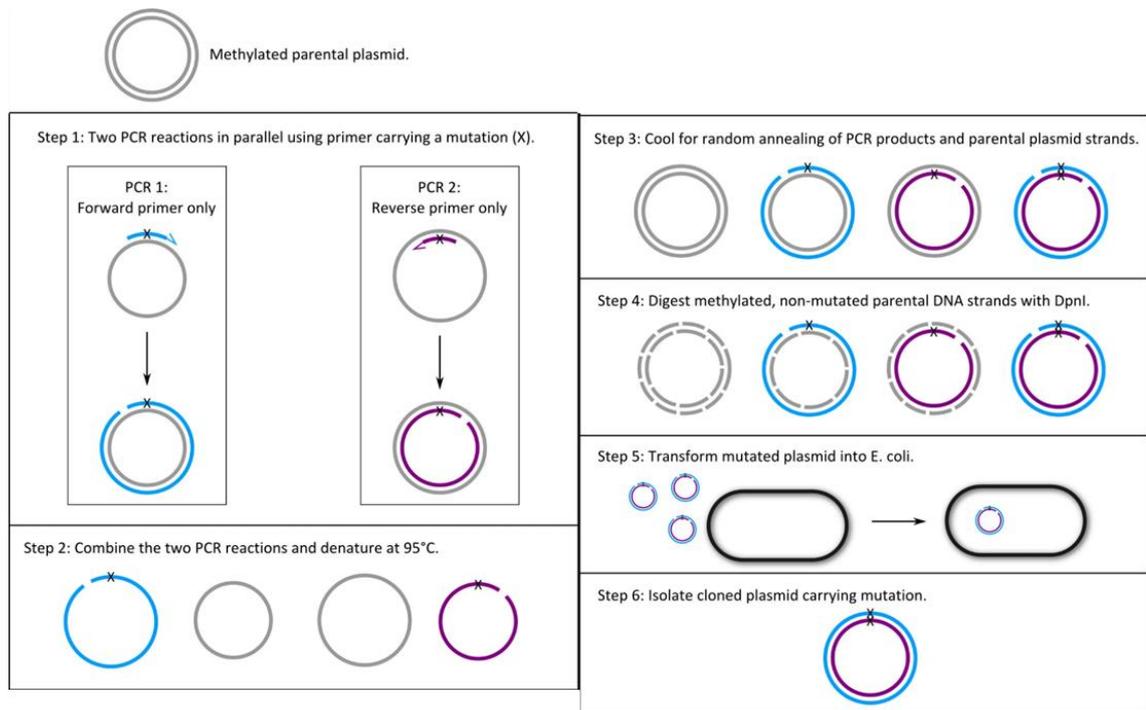
Figure 2.1. Flow-chart of the single-primer site-directed mutagenesis method. The parental plasmid is shown in grey colour and the two PCR synthesized strands are shown in blue and purple. The letter "X" marks the position of the mutation. Taken from (237).

2.3 Protein purification

2.3.1 Production and purification of recombinant aSec proteins

E. coli Bl21(DE3) pre-cultures harbouring the recombinant plasmids were grown overnight and used to inoculate LB AIM (Formedium) supplemented with the appropriate antibiotics in bevelled conical flasks (see Table 2.4). The cultures were grown aerobically under 200 rpm shaking for 3 hr at 37°C and then for 60 hrs at 16°C. Cells were pelleted by centrifugation at 4°C for 30 min at 5000 rpm.

Cells were then lysed, and proteins extracted using the BPER Complete Reagent (Thermofisher) in 150 mM NaCl and 20 mM Tris pH 7.9 following manufacturer's instructions and then sonicated at 50% amplitude with a 3.1 mm probe on ice 5 times in 30 sec intervals with 30 sec pauses. The suspension was then centrifuged at 4°C at 10,000 rpm for 30 min to separate soluble and insoluble protein fractions.

For LrSecA2-SecY2-Asp4$_{53608}$, the lysis step was modified as follows. The cell pellet was resuspended in phosphate buffered saline (PBS) pH 7.5 with 3 mM CHAPS and protease inhibitors (0.1 mM PMSF and 1 mM benzamidine) and sonicated at 50% amplitude with a 3.1 mm probe on ice 5 times in 30 sec intervals with 30 sec pauses. This mixture was then incubated on a rotary shaker at 4°C for 7 hr. The solution was then centrifuged at 10,000 rpm for 30 min at 4°C to separate soluble and insoluble protein fractions.

Purification of his-tagged proteins was carried out using nickel-nitrilotriacetic acid (NiNTA). Briefly, clarified extracts of soluble proteins were loaded onto a 2.5 mL gravity flow His-Bind® Column (Novagen) according to the manufacturer's instructions. Bound proteins were eluted in the column strip buffer (150 mM NaCl, 20 mM Tris and 100 mM EDTA) with the aim to release all nickel bound proteins, as high concentrations of imidazole would impact downstream purification processes.

When needed, his-tag removal by thrombin or C3-protease cleavage was performed at 4°C overnight using 1 unit of bovine thrombin (Merck) or 1-unit C3-protease (provided by collaborators at University of East Anglia) and 10 mL of up to 3 mg/mL protein in column strip buffer containing 10 mM CaCl$_2$. The thrombin and cleavage product were separated using the 1 mL HItrap Benzamidine Fast Flow column on the ÄKTA Pure (GE Healthcare). C-3 cleavage products were cleaned-up using size-exclusion chromatography.

Purification of recombinant proteins with O-GlcNAc, such as LrgSRR1$_{53608}$, was carried out by gravity flow using agarose-bound wheat germ agglutinin (agWGA) (VectorLabs) as previously described (235). Briefly, clarified soluble protein extracts were bound and washed with 10 column vol of HEPES buffer (20 mM HEPES pH 7.5, 200 mM NaCl), and the bound proteins were eluted with six column vol of HEPES buffer containing 500 mM GlcNAc. The proteins were extensively dialysed in H$_2$O to remove any free GlcNAc.

Further purification of recombinant proteins was carried out by size-exclusion gel filtration (SEGF) chromatography. Briefly, proteins were concentrated to 2.5 mL by centrifugation using

a Vivaspin® 10,000 MWCO spin-filter at 4°C and 8000 rpm. The samples were then loaded onto a gravity flow desalting PD-10 column (Merck) and buffer-exchanged into 150 mM NaCl, 20 mM Tris pH 7.9. This solution was further spin-filter concentrated to 1 mL and loaded onto a HiLoad® 16/600 Superdex® 75 pg column (GE Healthcare), or a HiLoad® 16/600 Superdex® 200 pg column (GE Healthcare). SEGF purification was conducted on the ÄKTA Pure FPLC (GE Healthcare) using an initial loading volume of 2 mL, at a flow rate of 0.75 mL/min and monitored by UV-detection to determine fraction collection. Proteins were eluted in 150 mM NaCl, 20 mM HEPES at pH 7.5 in 1 mL or 1.8 mL fractions. Protein complexes were further concentrated using Vivaspin® 100,000 MWCO spin-filter by centrifugation at 4°C and 5000 rpm.

2.3.2 Purification of native *Lr*SRRP

Purification of *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ was conducted as previously described (235). Briefly, *L. reuteri* 100-23 and ATCC 53608 strains were grown in LDM-II for 24 hr at 37°C aerobically until cell density were in stationary phase (OD$_{600nm}$ >2.5). The bacteria were discarded following centrifugation at 10,000 rpm for 10 mins. Ammonium sulphate was added to the spent medium at a final concentration of 60% (w/v) to precipitate the proteins. The suspension was stirred overnight at 4°C. The precipitated proteins were recovered by centrifugation at 10,000 rpm for 30 min. The proteins were resuspended in HEPES buffer (HEPES 20 mM, NaCl 150 mM, pH 7.5) and *Lr*SRRP purified by gravity flow affinity chromatography, using agarose-bound wheat germ agglutinin (agWGA), as described above (see section 2.3). Loosely bound proteins were removed with 10 column volumes of HEPES buffer, and the bound proteins were eluted with six column volumes of HEPES buffer containing 500 mM GlcNAc. The proteins were extensively dialysed in H$_2$O to remove any free GlcNAc.

2.4 Protein analysis

2.4.1 Protein gel electrophoresis and staining

Proteins were separated under denaturing conditions on sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE). Briefly, NuPAGE™ LDS Sample Buffer (Invitrogen) was

added to protein samples and boiled at 80°C for 5 min. Samples were loaded onto NuPAGE™ 4-12% Bis-Tris gels (Invitrogen) and run at 200V for 50 min in 1 X NuPAGE™ MOPS buffer.

For native-PAGE analysis, the NuPAGE Tris-Acetate Gels 3-8% (Invitrogen) were used according to manufacturer's instructions with some modification of voltage and runtime to facilitate separation of native proteins. In native-PAGE, proteins are separated according to the net charge, size, and shape of their native structure. Briefly, purified protein samples were combined with Novex™ Tris-Glycine Native Sample Buffer (Invitrogen) and run at 150V for 65 min in 1 X Novex™ Tris-Glycine Native Running Buffer (Invitrogen).

Gels were stained with InstantBlue™ (Expedeon) for 10 min at room temperature and imaged using a fluorchemE (ProteinSimple). Gels destined for western-blotting were not stained at this stage and were transferred to Tris-Glycine Transfer Buffer (Invitrogen) for subsequent processing.

2.4.2 Western blot analysis

SDS-PAGE or native gels were blotted onto a polyvinylidene difluoride (PVDF) membrane at 30V for 2 hrs in NuPAGE transfer buffer using an X-cell II blot module (ThermoFisher Scientific). The blots were then incubated in a Protein-free (PBS) blocking solution (ThermoFisher Scientific) for 1 hr. Protein with his-tag fusions could be detected by incubation with anti-his HRP monoclonal antibody (Invitrogen) at 1:1500 dilution in PBS for 30 mins and visualised using the Clarity™ Western ECL Substrate (BioRad) according to manufacturer's instructions. Glycoproteins with *O*-GlcNAc moieties were detected by incubation with fluorescein-WGA (VectorLabs) at 5 µg/mL in PBS for 1 hr. After staining, blots were washed with H_2O and scanned on a GS-800 calibrated densitometer (Bio-Rad, UK) with default settings for colorimetric and fluorescein analysis respectively.

2.5 Protein structural characterisation

2.5.1 Protein crystallisation

Crystallisation trials were dispensed in 96 well crystallisation trays (Molecular Dimensions) for sitting drop vapour diffusion, using the Gryphon crystallisation robot (Art Robbins Instruments). Crystallisation trials were conducted for the following proteins and at the concentrations listed in Table 2.6.

Table 2.6. Crystallisation trial conditions for recombinant *Lr*aSec proteins. Recombinant aSec proteins were exposed to different buffer for crystallisation.

aSec protein	Concentrations (mg/mL)	Protein buffers	Crystallisation conditions (Molecular Dimensions)
*Lr*GtfC$_{53608}$	10	20 mM TrisHCl pH 8.0, 150 mM NaCl	PGA-Screen^{Tm}
			PACT-Premier^{Tm}
	12	20 mM HEPES pH 7.5, 150 mM NaCl	Wizard polyethylene glycol-Ion^{Tm}
			3D-Structure Screen^{Tm}
	15	20 mM TrisHCl pH 8.0, 150 mM NaCl, 5 mM UDP	The BCA Screen^{Tm}
			Clear Strategy Screen^{Tm}
		20 mM TrisHCl pH 8.0, 150 mM NaCl, 5 mM UDP-GlcNAc	Morpheus I and II®
			JGSC Plus^{Tm}
*Lr*GtfC$_{100-23}$	9	20 mM TrisHCl pH 8.0, 150 mM NaCl	PGA-Screen^{Tm}
			PACT-Premier^{Tm}
	18	20 mM TrisHCl pH 8.0, 150 mM NaCl, 5 mM UDP-Glc	Clear Strategy Screen^{Tm}
		20 mM TrisHCl pH 8.0, 150 mM NaCl, 1 mM UDP	JGSC Plus^{Tm}
		20 mM TrisHCl pH 8.0, 150 mM NaCl, 1 mM UDP-GlcNAc	3D-Structure Screen^{Tm}
*Lr*Asp1$_{53608}$	5	20 mM HEPES pH 7.5, 100 mM NaCl	PGA-Screen^{Tm}
			PACT-Premier^{Tm}
	10	20 mM MES pH 6.5, 100 mM NaCl	Wizard polyethylene glycol-Ion^{Tm}
			3D-Structure Screen^{Tm}
	20	20 mM HEPES pH 8, 200 mM NaCl	The BCA Screen^{Tm}
			Clear Strategy Screen^{Tm}
*Lr*Asp2$_{53608}$	7.5	20 mM HEPES pH 8, 200 mM NaCl	3D-Structure Screen^{Tm}
	12.5	20 mM HEPES pH 8, 200 mM NaCl, 1 mM pNP-acetate	Clear Strategy Screen^{Tm}
			The BCA Screen^{Tm}

	15	20 mM HEPES pH 8, 200 mM NaCl, 1 mM Acetylcholine	MIDAS Screen[Tm]
			PACT-Premier[Tm]
*Lr*Asp2100-23	7.5	20 mM HEPES pH 8, 200 mM NaCl	3D-Structure Screen[Tm]
	12.5	20 mM HEPES pH 8, 200 mM NaCl, 1 mM pNP-acetate	Clear Strategy Screen[Tm]
			The BCA Screen[Tm]
	15	20 mM HEPES pH 8, 200 mM NaCl, 1 mM Acetylcholine	MIDAS Screen[Tm]
			PACT-Premier[Tm]
*Lr*SecA253608	8	20 mM HEPES pH 7.8, 100 mM NaCl	PGA-Screen[Tm]
			PACT-Premier[Tm]
	12.5	20 mM HEPES pH 7.8, 100 mM NaCl, 1 mM ATP	Clear Strategy Screen[Tm]
			JGSC Plus[Tm]
	16	20 mM HEPES pH 7.8, 100 mM NaCl, 1 mM GTP	3D-Structure Screen[Tm]
			MIDAS Screen[Tm]
*Lr*SecA2100-23	10	20 mM HEPES pH 7.8, 100 mM NaCl	PGA-Screen[Tm]
			PACT-Premier[Tm]
	15	20 mM HEPES pH 7.8, 100 mM NaCl, 1 mM ATP	Clear Strategy Screen[Tm]
			JGSC Plus[Tm]
*Lr*SecY253608	5	20 mM HEPES pH 7.4, 100 mM NaCl	Clear Strategy Screen[Tm]
	10	20 mM BTP pH 7.0, 100 mM NaCl	JGSC Plus[Tm]
	15	20 mM MES pH 6.5, 100 mM KCl	MemStart-MemSys Screen[Tm]
*Lr*Asp1-Asp2-Asp353608	10	20 mM HEPES pH 7.4, 100 mM NaCl, 1% glycerol	Clear Strategy Screen[Tm]
			JGSC Plus[Tm]
			3D-Structure Screen[Tm]
			MIDAS Screen[Tm]
	15		PGA-Screen[Tm]
			PACT-Premier[Tm]
			MemStart-MemSys Screen[Tm]
*Lr*SecA2-SecY2-Asp453608	3	20 mM HEPES pH 7.8, 100 mM NaCl	Clear Strategy Screen[Tm]
	5		JGSC Plus[Tm]
	10	20 mM HEPES pH 7.8, 100 mM NaCl, 1 mM ATP	3D-Structure Screen[Tm]
			MIDAS Screen[Tm]
	12.5		PGA-Screen[Tm]
			PACT-Premier[Tm]
	15	20 mM Tris pH 8.1, 50 mM NaCl	MemStart-MemSys Screen[Tm]

2.5.2 X-ray crystallography

Crystallising conditions for 10 mg/mL LrGtfC$_{100\text{-}23}$ wereP: 0.2 M potassium thiocyanate, 0.1M bis-Tris propane pH 7.5, 20% w/v polyethylene glycol for the apo form and co-crystallised with 1 mM UDP, or in 0.2 M ammonium citrate dibasic, 20% (w/v) PEG 3,350 with 1 mM UDP-GlcNAc. Crystals were sent to Diamond Light Source (Oxford, UK) for diffraction analysis. Diffracting datasets obtained at a wavelength of 0.9762Å were solved by molecular replacement using *Streptococcus parasanguinis* Gtf3 glycosyltransferase (PDB 3QKW) as a model by Dr Gareth Ashworth, Hans Pfalzgraf and Dr Andrew Hemmings at the University of East Anglia (UEA) [238].

2.5.3 Size-exclusion chromatography coupled small angle x-ray scattering (SEC-SAXS)

SEC-SAXS was performed at the Diamond Light Source (Oxford, UK) operating using a HiLoad® 16/600 Superdex for size-exclusion and a scattering vector (q) range from 0.0032 to 0.38 Å^{-1} [239]. Briefly, LrAsp2$_{53608}$, LrAsp1-Asp2-Asp3$_{53608}$ and LrSecA2-SecY2-Asp4$_{53608}$ were prepared at 10 mg/mL in HEPES pH 7.5, 200 mM NaCl and flash-frozen. Data were analysed in BioXtas RAW using ATSAS and GNOM plug-ins to derive pair distance distribution function (P(r)), scattering angle at 0 (I(0)) as well as other constants needed for structural interpretation and model construction [240-242]. *Ab initio* electron density determination was conducted from solution scattering data using DENSS [243]. Molecular weight estimations were conducted by volume of correlation, Porod volume and Bayesian inference [244-246].

2.5.4 Protein homology modelling

Homology modelling of LrGtfC$_{53608}$ was conducted using SwissModel using the LrGtfC$_{100\text{-}23}$ structure as a template [247]. The model was evaluated based on QMean statistics (a predictor of quality of local similarities between target and template) for quality assessment [248].

2.5.5 Alphafold2 structural modelling

Alphafold2 was used to construct models of LrGtfAB$_{53608}$, LrGtfAB$_{100-23}$, LrAsp1$_{53608}$, LrAsp1$_{100-23}$, LrAsp2$_{53608}$, LrAsp2$_{100-23}$, LrAsp3$_{53608}$, LrAsp3$_{100-23}$, LrAsp4$_{53608}$, LrAsp1-Asp2-Asp3$_{53608}$, LrSecA2$_{53608}$, LrSecA2$_{100-23}$, LrAspY2$_{53608}$ and LrSecY2$_{100-23}$ (249). This was done through submission of aSec protein amino-acid sequences derived from publicly available sequence data and executed using the Google Collaboratory Alphafold2 trained with structures available in the protein data bank (PDB), UniProtv2021_04, UniRef90: v2022_01, MGnify v2022_05, and BFD (250).

2.5.6 PDB-Image processing and rendering

Mol* (/'molstar/) was used to produce and render graphical representations of protein 3D-structural models and is a web-based open-source toolkit for visualisation and analysis of large-scale molecular data (251). Superimpositions and structural homology analysis were also performed in Mol* and compared with Gesamt superimpositions in CCP4mg for validation (252).

2.5.7 Protein domain movement analysis

DynDom was used to determine domains, hinge axes and hinge bending residues in proteins where two or more conformations were available, such as LrGtfC$_{100-23}$ in the presence or absence of UDP or UDP-GlcNAc (253).

2.6 Enzymatic assays for characterisation of aSec proteins

2.6.1 Malachite green ATP/GTPase assays

The Malachite Green Phosphate Assay Kit (MAK307, Merck) was used to measure the release of free orthophosphate from ATP or GTP hydrolysis by LrSecA2$_{53608}$, LrSecA2K114A$_{53608}$, LrSecA2$_{100-23,}$ the LrSecA2-SecY2-Asp4$_{53608}$ complex and the LrAsp1-Asp2-Asp3$_{53608}$ complex according to the manufacturer's instructions (Figure 2.2). A free phosphate standard was provided in the kit to be used for quantification. ATP (Merck) or GTP (Merck) was used as a substrate at 50 µM, 100 µM, 150 µM, 250 µM, 500 µM and 750 µM concentration in a final volume of 100

μL. The final concentration of proteins used was 180 nM (molarity derived from protein theoretical molecular weight and mg/mL by absorbance at 280 nm and predicted protein extinction coefficient) in a reaction buffer containing HEPES pH 7.5, 200 mM NaCl and 1 mM $MgCl_2$. The reaction product was monitored at 610 nm using a Fluostar Optima 96-well plate reader (BMG-Labtech) in clear low-binding 96-well plates (BMG-Labtech) at 25°C. Kinetic analysis and non-linear regression analysis were applied using Microsoft Excel.
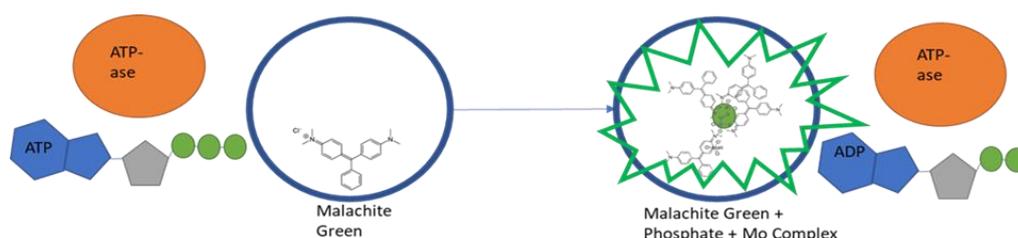
Figure 2.2 Schematic of Malachite Green ATP-assay. ATP-hydrolysis by ATPases releasing ADP and inorganic phosphate in the presence of malachite green forms a green molybdate (Mo) complex detectable at 610 nM.

2.6.2 Acetylesterase assays

2.6.2.1 *p*-nitrophenyl (*p*NP)-acetate assay

Acetylesterase activity was determined by measuring the hydrolysis of the *p*-nitrophenyl (*p*NP)-acetate (Merck) following adaptation and combination of several methods (216, 219, 221) (Figure 2.3). The reaction was performed in 96-well microtiter plates (BMG-Labtech) with a total volume of 200 μL. The reaction buffer was the same as the enzyme storage buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1% glycerol) and *p*NP-acetate (1 M stock dissolved in 100% ethanol) was used at 0.025, 0.5, 5, 7.5 mM final concentration. Purified $Lr\text{Asp2}_{53608,}$ $Lr\text{Asp2S349A}_{53608}$, $Lr\text{Asp2}_{100-23}$ and $Lr\text{Asp1-Asp2-Asp3}_{53608}$ were added to the mixture at a final concentration of 600 nM for monomeric proteins and 300 nM for protein complexes (molarity derived from protein theoretical molecular weight and mg/mL by absorbance at 280 nm and predicted protein extinction coefficient). The reaction was carried out for 30 min at 30°C. The release of *p*NP was measured by monitoring change in absorbance at 405 nm using a Fluostar Optima 96-well plate

reader (BMG-Labtech). This experiment was performed with 10 replicates and kinetic analysis and non-linear regression were applied using Microsoft Excel.
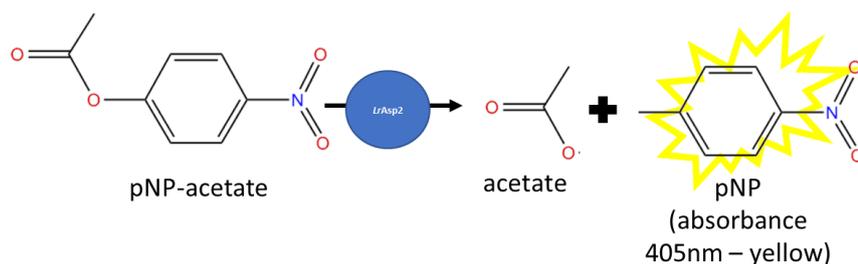
Figure 2.3. Schematic of acetylesterase assay using *p*NP-acetate. The substrate *p*NP-acetate can be hydrolysed by acetylesterases into acetate and *p*NP which can be detected at 405 nM

2.6.2.2 Indoxylacetate enzymatic assay

Acetylesterase activity was also measured by the hydrolysis of indoxylacetate (Merck) into indigogenic indoxyl and acetate (Figure 2.4). The reaction was performed in 96-well microtiter plates (BMG-labtech) in a total volume of 200 µL. The reaction buffer was the same as the enzyme storage buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 1% glycerol) and indoxylacetate (stock of 100 mM dissolved in 100% DMSO) was used at 50, 100, 150, 200 and 250 µM final concentration. Purified *Lr*Asp2$_{53608}$, *Lr*Asp2S349A$_{53608}$ and *Lr*Asp2$_{100\text{-}23}$ were added to the reaction mixture at 150 nM final concentration (molarity derived from protein theoretical molecular weight and mg/mL by absorbance at 280 nm). The reaction was carried out for 120 min at 30°C. The release of indoxyl was measured by monitoring change in absorbance at 375 nm using a Fluostar Optima 96-well plate reader (BMG-Labtech). This experiment was performed with 3 replicates and kinetic analysis and non-linear regression were applied using Microsoft Excel.

Inhibition of indoxylacetate hydrolysis by *Lr*Asp2 proteins was tested using paraoxonethyl (solubilised to working concentrations in reaction buffer) following similar reaction conditions (enzyme storage and reaction buffer were 20 mM HEPES pH 7.5, 200 mM NaCl, 1% glycerol). Inhibitor kinetics were determined by addition of paraoxonethyl at 25, 50, 100, 150, 200, 250, 300 µM final concentration at the beginning of the reaction. From % rate inhibition, inhibitor

kinetic parameters, such as the half-maximal inhibitory concentration (IC50), were derived using data processing packages like Microsoft Excel.
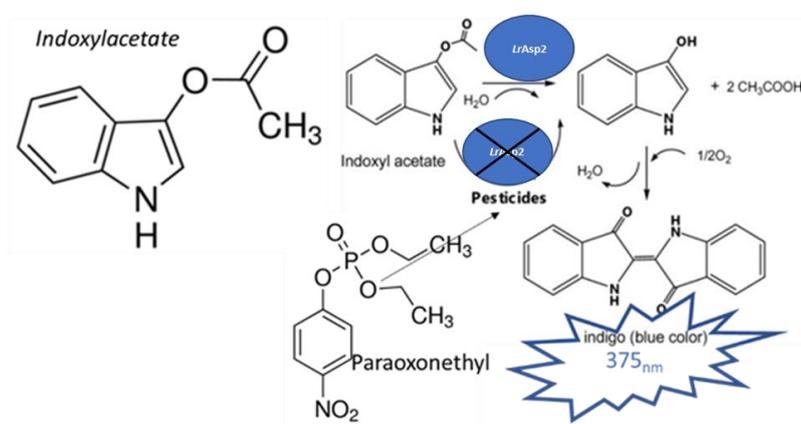
Figure 2.4 Schematic of acetylesterase assay using indoxylacetate. Indoxylacetate can be hydrolysed into an indoxyl complex which can be detected at 375 nm. Chemicals such as pesticides like paraoxonethyl can act as inhibitors to acetylesterases catalysing this reaction.

2.7 *In vitro* binding assays

2.7.1 ELISA-based binding assays

ELISA-based binding assays were performed using the Nunc MaxiSorp™ high protein-binding capacity 96 well ELISA plates (Thermofisher). First, recombinant aSec proteins were immobilised overnight at 4°C. Following 3 washing steps with washing buffer (PBS containing 0.05% Tween 20), the plate was incubated for 1 hr with blocking buffer (PBS containing 0.05% Tween 20 and 0.1% BSA). The plate was then washed three times with washing buffer, followed by incubation with purified his-tagged aSec protein (analyte) for 1 hr. Then the plate was washed 3 times with blocking buffer and incubated with 1:1000 6x-His Tag monoclonal antibody (HIS.H8), HRP (Invitrogen) for 1 hr. After antibody incubation, the plate was washed three times with washing buffer and then incubated with 100 µL 3,3′,5,5′-tetramethylbenzidine (TMB) substrate (Thermofisher) for 5 min to allow for colour development. The reaction was quenched with 100 µL 2N H_2SO_4 and an end-point OD measurement was made at 450 nm and 570 nm

(reference wavelength). BSA and PBS in place of an immobilised protein served as negative controls. Two different protein preparations were used each in triplicate for protein complexes (n=6) and three different protein preparations were used for monomeric proteins (n=3), except for LrGtfCs (n=4). The immobilised aSec proteins and binding partners (analytes) are listed in Table 2.7.

Table 2.7. Summary of recombinant aSec proteins used in ELISA-based binding assays.

Immobilised protein (coated to well)	Analyte protein	Analyte concentration (µg/mL)
LrSRRP$_{53608}$	LrGtfC$_{53608}$	20
	LrAsp1$_{53608}$	15
	LrAsp2$_{53608}$	15
	LrAsp1-Asp2-Asp3$_{53608}$	5 and 10
	LrSecA2$_{53608}$	10
	LrSecA2$_{100-23}$	10
	LrSecA2-SecY2-Asp4$_{53608}$	5 and 10
LrSRRP$_{100-23}$	LrGtfC$_{100-23}$	20
	LrAsp1$_{53608}$	15
	LrAsp2$_{53608}$	15
	LrSecA2$_{53608}$	10
	LrSecA2$_{100-23}$	10
	LrSecA2-SecY2-Asp4$_{53608}$	5 and 10
LrgSRR1$_{100-23}$	LrGtfC$_{53608}$	20
	LrGtfC$_{100-23}$	20
	LrAsp1$_{53608}$	15
LrSP$_{53608}$	LrGtfC$_{53608}$	20
	LrGtfC$_{100-23}$	20
	LrAsp1$_{53608}$	15
	LrSecA2-SecY2-Asp4$_{53608}$	5 and 10
LrSecA2$_{53608}$	LrAsp1$_{53608}$	5
	LrAsp2$_{53608}$	5
	LrAsp1-Asp2-Asp3$_{53608}$	5

2.7.2 Biolayer interferometry (BLI)

BLI was performed using the Octet RED96 system (ForteBio). Assays were conducted in black 96-well plates (Nunc™ F96 MicroWell™ plate, Thermo Scientific) at 30 °C, under shaking (1000 rpm) using pre-hydrated ARG2 Octet® Amine Reactive 2nd-Generation Biosensors (ForteBio). Assays were conducted as follows: (1) biosensor equilibration in H_2O for 600 sec (2) amine-coupling activation with EDC/NHS (10:20 mM) for 300 sec (3) ligand loading in 10 mM sodium acetate buffer pH 5.8 for 400 sec (4) biosensor quenching with 1 M ethanolamine pH 9.0 for 200

sec (5) baseline in reaction buffer containing PBS pH 7.4 and 0.001% Triton X100 for 60 sec (6) association with analyte in reaction buffer for 900 sec, and (7) dissociation in reaction buffer for 900 sec. The immobilised proteins (ligands) and the association partner proteins (analytes) used in this work are listed in Table 2.8. Binding kinetics were treated for global fit using a 1:1 binding model with incomplete dissociation analysed in the ForteBio Octet data analysis version 11.1 software package using triplicate data (n=3).

Table 2.8. Summary of recombinant aSec proteins used in BLI binding kinetics.

Immobilised protein	Analyte protein	Analyte concentration (nM)
LrSRRP$_{53608}$	LrAsp1-Asp2-Asp3$_{53608}$	75, 100, 150, 200, 250
	LrSecA2-SecY2-Asp4$_{53608}$	30, 60, 90, 120, 150
LrSRRP$_{100-23}$	LrAsp1-Asp2-Asp3$_{53608}$	75, 100, 150, 200, 250
	LrSecA2-SecY2-Asp4$_{53608}$	30, 60, 90, 120, 150
Anti-His6 HRP monoclonal antibody (Invitrogen)	LrSecA2-SecY2-Asp4$_{53608}$	30, 60, 90, 120, 150
LrSecY2$_{53608}$	LrSecA2$_{53608}$	10,20

2.7.3 Thermal shift assay

The thermal shift assays (TSA) or thermofluor were conducted in the StepOnePlus Real Time PCR system (Thermofisher). Reactions were performed in a final volume of 25 µL using 20 µM protein and SYPRO Orange (Merck) dye at 10X final concentration as previously described (254). For LrGtfC WT and mutant, the binding reactions were conducted in 50 mM Tris pH 7.5, 100 mM NaCl in the absence or presence of 3 mM UDP-Glc (Carbosynth) or 3 mM UDP-GlcNAc (Carbosynth). A thermal ramp from 20 – 90°C was used at 1.5% stepwise increases. Samples (n=4) were randomised in a MicroAmp^{Tm} 96-well plate (Thermofisher) and resulting melt-curve data analysed in Microsoft Excel.

2.8 Bioinformatic tools

Genetic and protein sequence visualisation and alignments were performed in Mega11 (255). Protein structure bioinformatics was conducted using PROTPARAM and PRO-SITE from Expasy Swiss Institute of Bioinformatics (256, 257).

2.9 Statistics and data processing tools

Data organisation and manipulation (such as non-linear regression analysis) was performed using Microsoft Excel. For statistical work, G*Power was used for determination of sample-sizes and appropriate statistical tests (249). The statistical tests were performed in Microsoft Excel, or equipment associated software, for example, Data Analysis v11.1 software for BLI binding kinetics using the Octet (see section 2.7.2).

Chapter 3 Structure and function of glycosyltransferases from the *L. reuteri* aSec system

3.1 Introduction

The biosynthesis of glycans and glycosylation of various molecules is mediated by glycosyltransferases (GT); enzymes that transfer sugar moieties from activated donor molecules to specific acceptors thus forming glycosidic bonds. GTs belong to a large subclass of enzymes categorised in 115 families (as of April 2022) based on sequence identity in the CAZy database (www.cazy.org). In bacteria, protein glycosylation has mainly been characterised in pathogens, whereby complex oligosaccharide structures have been shown to play key role in pathogenicity (258-260). Both *N*-and *O*-glycosylation pathways have been found in prokaryotes showing similarities with eukaryotic and archaeal kingdoms (99, 260). *N*-glycosylation refers to glycans linked with the amide nitrogen atom of Asn residues of the protein acceptor. For gram-negative bacteria, such as *Campylobacter jejuni*, the presence of *N*-glycosylation has been demonstrated as essential for viability (261). No *N*-glycosylation pathways have been determined in gram-positive bacteria thus far. *O*-glycosylation occurs at the oxygen atom of Ser, Thr, or Tyr residues as is the case of *L. reuteri* SRRPs (195, 235). Recent studies reported *O*-glycosylation of flagellin of *E. coli* and *Salmonella enterica* strains(262, 263). GTs involved in either *N*- or *O*-glycosylation show specificity for the type of glycosidic bond (α or β) via conserved mechanisms.

There are two widely accepted catalytic mechanisms in GTs (Figure 3.1) (264). The inverting mechanism follows a single displacement by a nucleophilic attack of the acceptor on the C-1 of the sugar donor inverting the anomeric stereochemistry, which may also require divalent cations (e.g., Mg^{2+}, Mn^{2+}) (Figure 3.1A). The retaining mechanism proposes a double displacement whereby the intermediate state is a covalently bound glycosyl-enzyme complex (Figure 3.1B).
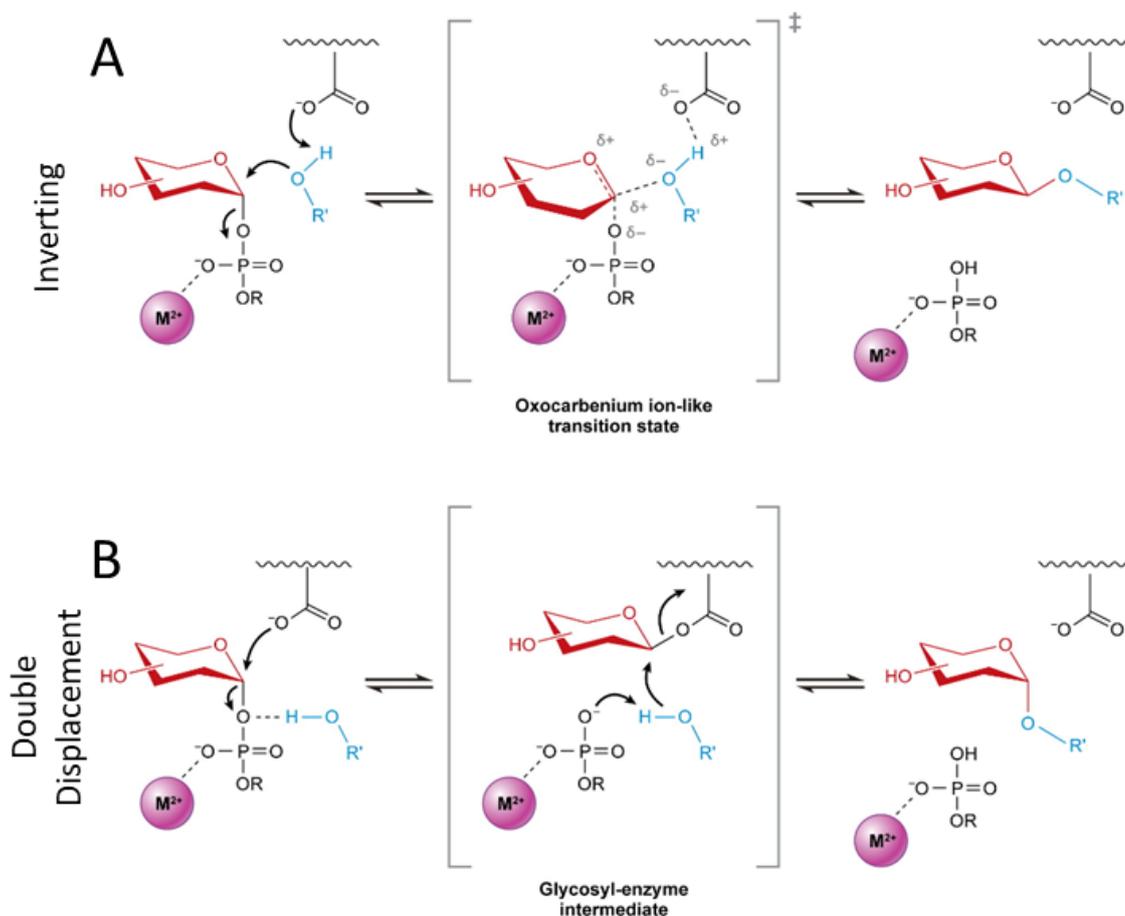
Figure 3.1. Proposed catalytic mechanisms of GTs (A) Inverting GTs utilise a direct-displacement that results in an inverted anomeric configuration via a single oxocarbenium ion-like transition state. (B) Retaining GTs may use a double-displacement mechanism involving the formation of a covalently bound glycosyl-enzyme intermediate. R = a nucleoside, a nucleoside monophosphate, a lipid phosphate, or phosphate (phosphorylases classified as GTs); and R'OH, an acceptor group (e.g., another sugar, a protein, or an antibiotic). Taken from (264).

Cytosolic GTs in eukaryotes and prokaryotes can display GT-A type or GT-B type folds (265). The GT-A fold is primarily characterised by an open, twisted β-sheet surrounded by α-helices on both sides generating two closely flanking Rossman-like folds (a series of alternating β and α helical segments wherein the β-strands are hydrogen bonded forming a β-sheet) (264, 266). Modified GT-A type enzymes exist where the order of β-strands vary (e.g., in sialyltransferases) (267). The GT-B fold also forms two β-α-β Rossman folds, but these are less tightly associated than in GT-A and, by facing each other, can form an active-site cleft (264). Transmembrane GTs contain 8-13 transmembrane helices and are categorised as GT-C type folds (268). GT structures

have also been described as flexible with several mobile loops, that can accommodate diverse and sometimes large acceptor substrates (269).

In general, GTs show specificity for donor-sugar, acceptor, and type of linkage (one enzyme = one linkage hypothesis) (270). In GT-A fold enzymes, the N-terminal domain is the site of sugar-nucleotide recognition whilst the C-terminal domain is responsible for binding the acceptor molecule (271-273). The roles of the C- and N-terminal domains are reversed in GT-B fold enzymes. In comparison to GT-A fold enzymes, GT-B C-terminal domains share higher sequence similarity between themselves reflecting similar donor-sugars, whereas the greater variety in potential acceptors leads to a more diverse N-terminal domains (272, 273). There are many examples where GTs have been engineered to alter specificity for the sugar donor, sugar acceptor, or enzymatic rate of the reaction. For example, a triple mutant of the glucosyltransferase OleD from *Streptomyces antibioticus* ATCC 8663 was reported to have a 60-fold higher catalytic efficiency with UDP-Glc and additionally displayed an improved promiscuity towards using other glycosyl donors *in vitro* (274). Chimeric GTs have also been generated taking advantage of highly homologous sequence identity to 'switch' specificity towards respective sugar nucleotides (275).

O-glycosylation of Ser/Thr residues is an important post-translational modification occurring on a range of acceptors including protein cargos secreted through the aSec pathway such as cell-surface adhesins SRRPs (195). The number and types of GTs encoded in the *L. reuteri* aSec secretion pathway are strain-dependent (Figure 3.2) (195). The *L. reuteri* ATCC 53608 aSec operon encodes *gtfA*, *gtfB* and *gtfC*, whereas the *L. reuteri* 100-23 aSec operon encodes *gtfA, gtfB, gtfC, gtfD, gtfE* and *gtfF* (Figure 3.2) (235). Previous work showed that the conserved primary *O*-GlcNAc moiety is deposited by a *Lr*GtfA-GtfB (*Lr*GtfAB) complex in *L. reuteri* SRRPs (235). This step is conserved in pathogenic bacteria like *S. gordonii* M99 and *S. parasanguinis* FW213 strains (238, 276). Mutation of *gtfB* in *L. reuteri* 100-23 inhibited transport of SRRP, highlighting the importance of the initial GlcNAc deposition for secretion (235). Subsequent glycosylation by *Lr*GtfC introduces the strain-specific glycosylation patterns observed in both native *Lr*SRRPs (di-GlcNAc in *Lr*SRRP$_{53608}$ and Hex-Glc-GlcNAc in *Lr*SRRP$_{100-23}$) and glycoengineered SRR

regions (*Lr*gSRR) in *E. coli* (235). The sugar nucleotide preferences of *Lr*GtfC were determined by saturation transfer difference NMR and thermofluor or differential scanning fluorimetry (DSF) with *Lr*GtfC$_{53608}$ specific for UDP-GlcNAc and *Lr*GtfC$_{100-23}$ UDP-Glc (235).

The aim of the work reported in this chapter was to characterise biochemically and structurally GTs present in *L. reuteri* 53608 and 100-23 with a focus on *Lr*GtfCs as the main mediator of strain-specific glycosylation of *Lr*SRRPs.
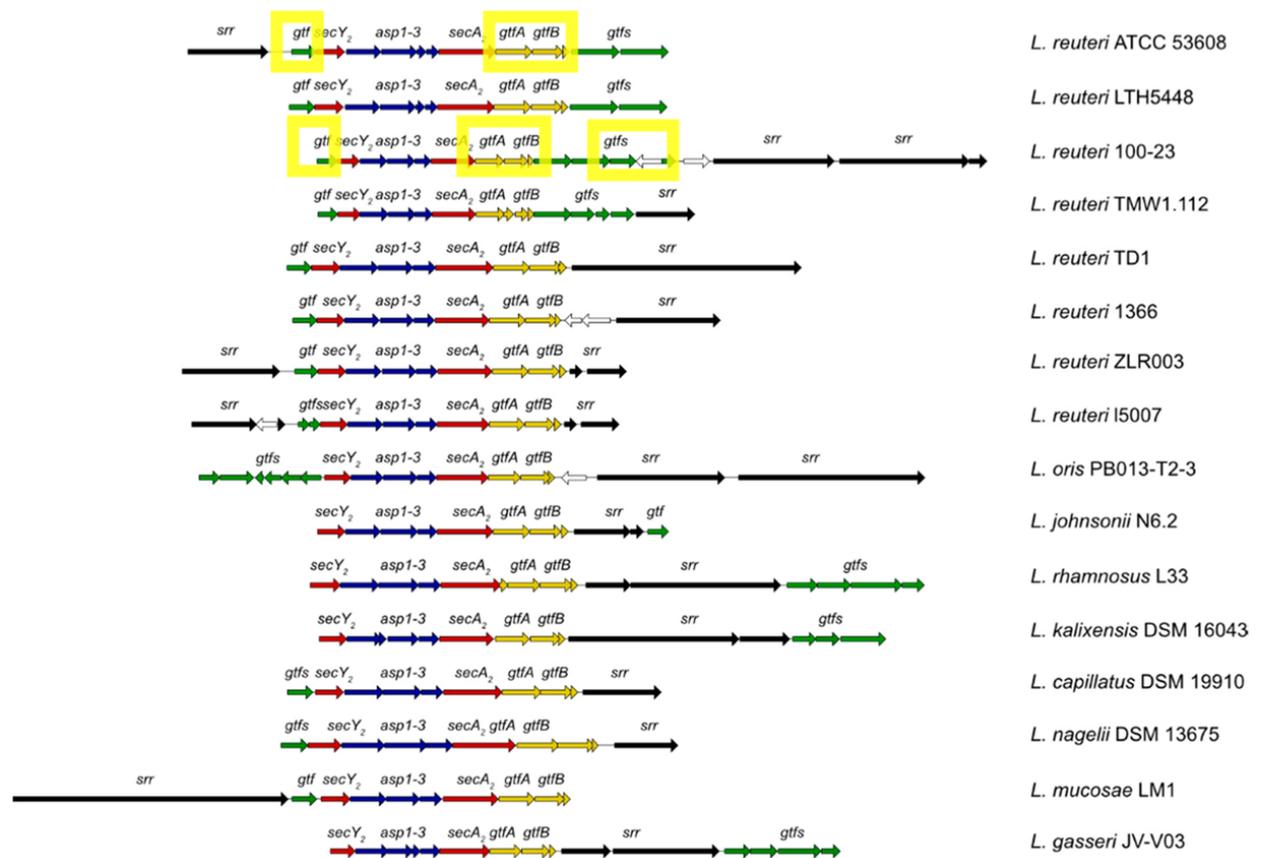
Figure 3.2. Organisation of the aSec clusters identified in *Limosilactobacillus* genomes. The genes encoding the translocases SecA2 and SecY2 are shown in red, the accessory secretion proteins Asp1–Asp2-Asp3 in blue and the priming GTs, GtfA and GtfB, in yellow. Genes encoding additional GTs are shown in green and the genes encoding serine-rich repeat proteins are illustrated in black. White arrows represent genes that are not part of the SecA2 machinery. The GTs of *L. reuteri* ATCC 53608 and 100-23 aSec systems are shown in yellow boxes. Adapted from (195).

3.2 Results
3.2.1 Production of recombinant LrGtfCs

For the biochemical and structural characterisation of LrGtfCs encoded by the *L. reuteri* aSec operon, recombinant LrGtfC$_{53608}$ and LrGtfC$_{100-23}$ were produced heterologously in *E. coli* as his-tagged proteins using the recombinant plasmids containing LrGtfC$_{100-23}$ in pET28a and LrGtfC$_{53608}$ in pOPIN-F (235). In addition, site-directed mutagenesis was used to produce catalytic and binding site mutants of LrGtfC$_{53608}$ and LrGtfC$_{100-23}$ (see section 3.2.3) following the Single-Primer Reactions in Parallel (SPRIP) protocol. This approach differed from traditional double-primer PCR mutagenesis methods by synthesising two mutant DNA strands from the template plasmid separately using single primers for subsequent annealing (237) (for experimental details see section 2.2.3).

Recombinant LrGtfC$_{100-23}$ and LrGtfC$_{53608}$ WT and mutants were purified by nickel affinity chromatography (NiNTA) showing expected sizes of 38 kDa on SDS-PAGE (Figure 3.3A). Bands that were apparent around 72 kDa may be contaminants or a dimerised form of LrGtfCs (76 kDa). For protein crystallisation, his-tags were removed by 3C-protease and thrombin cleavage for LrGtfC$_{100-23}$ and LrGtfC$_{53608}$, respectively. Further purification using size-exclusion gel filtration (SEGF) for LrGtfC$_{53608}$ retained higher molecular weight oligomers of LrGtfC$_{53608}$ or contaminants that co-eluted in all relevant fractions (Figure 3.3B). The yield of recombinant LrGtfCs WT and mutant was around 10 mg/L of *E. coli* culture after purification by NiNTA and SEGF (Appendix 3, supplementary Table S3.1).
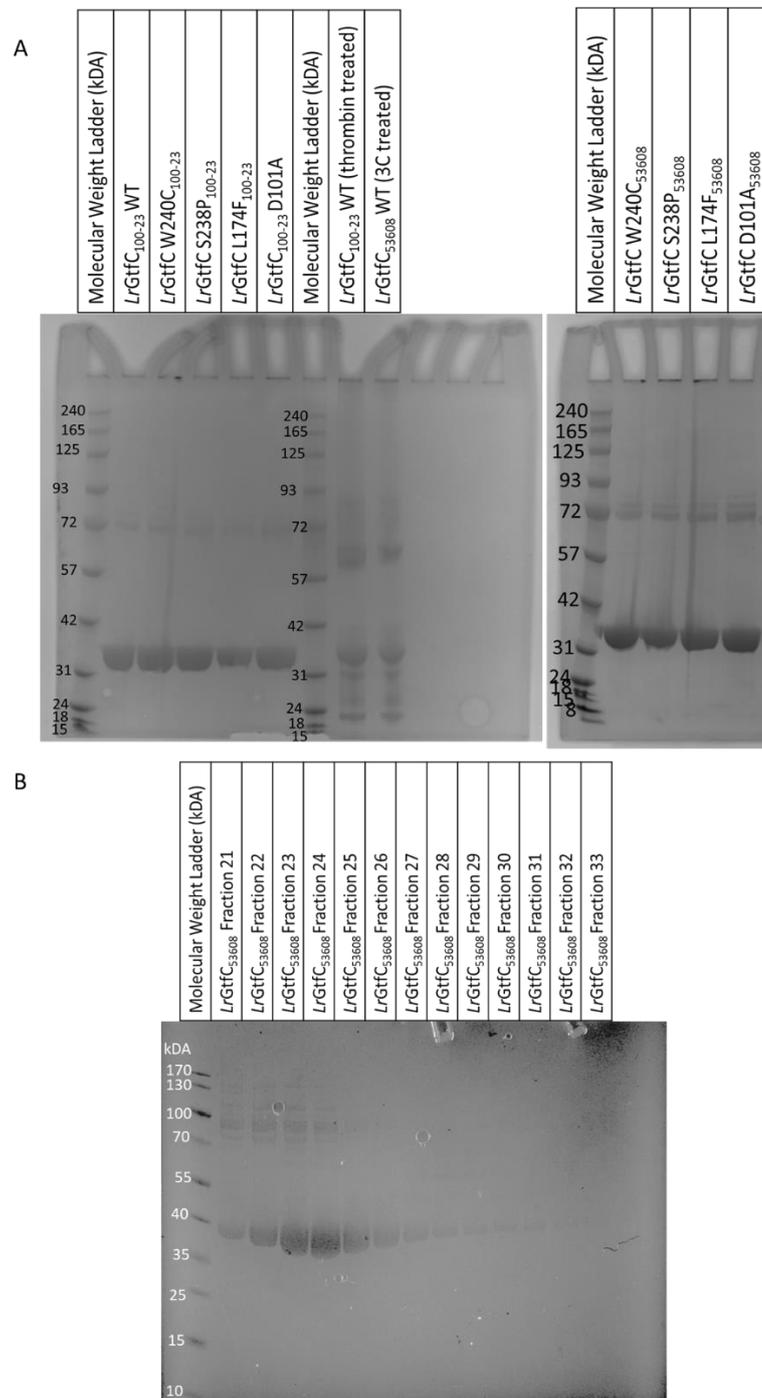
Figure 3.3 SDS-PAGE of purified recombinant GtfCs variants from *L. reuteri* ATCC 53608 and 100-23. (A) Recombinant *Lr*GtfC$_{100-23}$ (expressed from pET28a) and *Lr*GtfC$_{53608}$ (expressed from pOPIN-F) WT and mutant proteins were produced in *E. coli* BL21(DE3) and purified by NiNTA. His-tag removal was performed for the WT proteins using thrombin or 3C-protease for *Lr*GtfC$_{100-23}$ and *Lr*GtfC$_{53608}$, respectively. (B) Protein fractions of recombinant *Lr*GtfC$_{53608}$ purified by NiNTA and SEGF. Proteins were analysed by SDS-PAGE under denaturing conditions on 10% agarose and a molecular weight ladder was also run for size estimation. The gels were stained with instant-blue (Coomassie).

3.2.2 Structural Characterisation of *Lr*GtfCs

Purified recombinant *Lr*GtfC$_{100-23}$ and *Lr*GtfC$_{53608}$ were used in crystallisation trials in the presence or absence of ligands (Appendix 2, supplementary Figure S2.1, supplementary Figure S2.2). *Lr*GtfC$_{100-23}$ was successfully crystallised in the absence of a ligand as well as with UDP or UDP-GlcNAc under the following conditions: 10 mg/ml *Lr*GtfC$_{100-23}$ apo form and with UDP, 0.2 M potassium thiocyanate, 0.1 M bis-Tris propane pH 7.5, 20% w/v polyethylene glycol and in 0.2 M ammonium citrate dibasic, 20% (w/v) PEG 3,350 with UDP-GlcNAc. X-ray crystallography of crystals used in subsequent analysis had a P 21 21 21 orthorhombic space group. No crystals suitable for X-ray diffraction were observed for *Lr*GtfC$_{53608}$. The 3D structures were then solved by Dr Gareth Ashworth and Hans Pfalzgraf as part of the Hemming's group at the University of East Anglia (UEA, Norwich Research Park).

The crystal structure of *Lr*GtfC$_{100-23}$ obtained at 3.18 Å resolution was consistent with a tetramer for the apo enzyme, as well as in a complex with UDP or UDP-GlcNAc (Figure 3.4A). Co-crystallisation attempts with UDP-Glc substrate failed to identify the ligand in the active-site, possibly due to *Lr*GtfC$_{100-23}$ enzymatic activity. The monomers displayed a classic GT-B fold consisting of two similar Rossman fold subdomains containing a conserved central 6 stranded β sheet (Figure 3.4B). A flexible loop was observed close to the sugar-nucleotide binding site. This is an important structural determinant in GTs as demonstrated in glutamyltransferases from *B. subtilis* 168 and may act as a lid to the active site during catalysis (277).

Homology modelling of *Lr*GtfC$_{53608}$ was carried out using the crystal structure of *Lr*GtfC$_{100-23}$ as the template using Swissmodel (278). This allowed for superimposition of the active sites with UDP and UDP-GlcNAc bound structures, which helped identify key residues at different conformational stages which may be involved with strain-specific differences in sugar specificity of *Lr*GtfC$_{100-23}$ and *Lr*GtfC$_{53608}$ (Figure 3.4C). In *Lr*GtfC$_{100-23,}$ the presence of UDP-GlcNAc led to a transition of Trp240 and Tyr239 away from the active site cavity, whilst His250 and Asn249 rotated inwards. Additionally, domain movement tracing was carried out using DynDom revealing a 4.2° rotation of the C-terminal domain in red relative to the N-terminal domain in blue

along a hinge region in green (Figure 3.4D) (253). In both *Lr*GtfC 3D-models, a catalytic Asp101 was conserved along with a His250, which due to its spatial proximity may interact with UDP. The superimposition of *Lr*GtfC$_{100-23}$ active site onto *Lr*GtfC$_{53608}$identified residues which may play a role in *Lr*GtfC$_{100-23}$ and *Lr*GtfC$_{53608}$ reported sugar specificity (UDP-Glc for *Lr*GtfC$_{100-23}$ and UDP-GlcNAc for *Lr*GtfC$_{53608}$). The following residues were selected for site-directed mutagenesis: Leu174, Ser238 and Trp240 in *Lr*GtfC$_{100-23}$ corresponding to Phe174, Pro238 and Cys240 in *Lr*GtfC$_{53608}$ (see section 3.2.3).
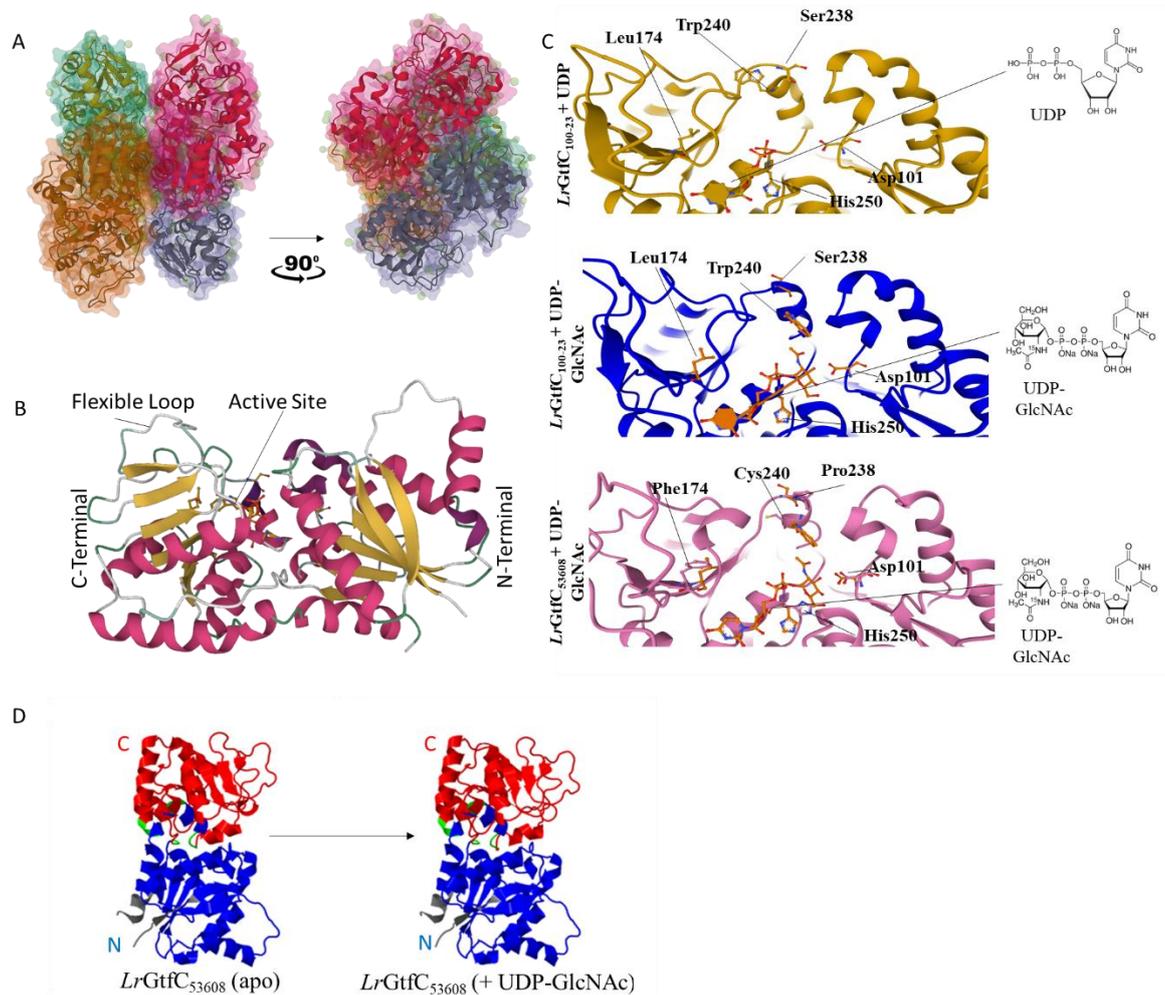
Figure 3.4. Structural characterisation of *Lr*GtfCs. (A) Tetrameric *Lr*GtfC$_{100-23}$ X-ray crystal structures in the absence of ligand superimposed with UDP-bound and UDP-GlcNAc-bound states as cartoon models and transparent molecular surface representations coloured by chain. (B) Single monomer of *Lr*GtfC$_{100-23}$ in complex with UDP in cartoon representation coloured by secondary-structures. (C) Active site of *Lr*GtfC$_{100-23}$ bound to UDP (orange), UDP-GlcNAc (blue) and superimposition of *Lr*GtfC$_{53608}$ homology model to *Lr*GtfC$_{100-23}$ bound to UDP-GlcNAc active site (pink) (D) Representation of DynDom N-(blue) and C-terminal (red) domain rotation of *Lr*GtfC$_{100-23}$ apoenzyme upon binding to UDP-GlcNAc with mobile residues highlighted (green).

3.2.3 Targeted engineering of *Lr*GtfC donor specificity

Alignment analysis of the amino acid sequences of 19 publicly available *Lr*GtfC sequences were used to determine if the binding site residues identified in the structural analysis (see section 3.2.2) followed host-specific patterns (Table 3.1, Figure S4.1). For residues at position 240, this appeared to be the case whereby most *L. reuteri* strains from porcine hosts (pink) shared the Cys240 representative of *Lr*GtfC$_{53608}$, whereas *L. reuteri* strains from rodent (grey) and chicken (red) shared exclusively the Trp240 representative of *Lr*GtfC$_{100-23}$. A similar pattern was shown for residue 174, however with 3/11 porcine *L. reuteri* strains sharing the Leu174 with *Lr*GtfC$_{100-23}$. In pig isolates, only *L. reuteri* pg3b strain had both Trp240 and Leu174 found in *Lr*GtfC$_{100-23}$ as also shown for rodent/chicken strains. Uniquely, *Lr*GtfC$_{100-23}$ was the only sequence containing Ser238 while all other sequences had Pro in this position. Taken together, *Lr*GtfC residues corresponding to position 174 and 240 appeared to be host-specific and were therefore selected for site-directed mutagenesis.

Table 3.1. Occurrence of *Lr*GtfC binding site residues across *L. reuteri* strains isolated from different hosts. Amino acid alignment using MUSCLE was utilised to classify *L. reuteri Lr*GtfC sequences relative to the nature of residues found at position 174, 238 and 240 in $Lr\text{GtfC}_{100\text{-}23}$ and $Lr\text{GtfC}_{53608}$. *L. reuteri* strains from rodents are shaded grey, from pigs are shaded pink, and from chicken are shaded orange.

Residue	*Lr*GtfC	Residue	*Lr*GtfC
Trp240	100-23, mlc3, Td1, lpuph2, byun-re, 484-32	Cys240	An417, ZLR003, 19-E-6, 167-67, 3C6, I5007, 121, ATCC 53608
	KLR1002, pg3b		
	SKKU-OGDONS-01, 1366		
Ser238	100-23	Pro238	mlc3, Td1, lpuph2, byun-re, 484-32
			KLR1002, pg3b, An417, ZLR003, 19-E-6, 167-67, 3C6, I5007, 121, ATCC 53608
			SKKU-OGDONS-01, 1366
Leu174	100-23, mlc3, Td1, lpuph2, byun-re, 484-32	Phe174	An417, ZLR003, 19-E-6, 167-67, 3C6, KLR1002, 121, ATCC 53608
	Pg3b, I5007, 3c6		
	SKKU-OGDONS-01, 1366		

Guided by the sequence alignment and structural analysis of the $Lr\text{GtfC}_{100\text{-}23}$ crystal structure and $Lr\text{GtfC}_{53608}$ homology model, site-directed mutagenesis was used to generate catalytic inactive mutants targeting the Asp101 residue in *Lr*GtfC D101A_{53608} and *Lr*GtfC $\text{D101A}_{100\text{-}23}$, or residues potentially implicated in sugar specificity: *Lr*GtfCP238S$_{53608}$, *Lr*GtfC C240W$_{53608}$, *Lr*GtfC

F174L$_{53608}$ and *Lr*GtfC S238P$_{100-23}$, *Lr*GtfC W240C$_{100-23}$, *Lr*GtfC L174F$_{100-23}$. Mutants of *Lr*GtfCs were expressed recombinantly in *E. coli* and purified as described in section 3.2.1 (Figure 3.3).

In order to investigate the sugar specificity of the mutants, a thermofluor approach was used to monitor potential shifts in thermal stability upon binding UDP-sugars. SYPRO Orange fluorescent dye binds hydrophobic surfaces but is quenched by water. Upon thermal denaturation of proteins, the hydrophobic core is exposed leading to higher fluorescence signal which can be measured to report differences in the melting temperature (Tm °C) in the presence or absence of potential ligands. Here *Lr*GtfC$_{53608}$ and *Lr*GtfC$_{100-23}$ at 20 μg/mL were subjected to an increasing +2% thermal gradient from 25°C - 90 °C in the absence or presence of 3 mM UDP-Glc or 3 mM UDP-GlcNAc (Figure 3.5). The *Lr*GtfC$_{53608}$ WT showed a positive shift in Tm °C for UDP-GlcNAc compared to the conditions with no ligand, or in the presence of UDP-Glc (Figure 3.5A). *Lr*GtfC P238S$_{53608}$ showed a positive thermal shift upon addition of UDP-Glc, but not with UDP-GlcNAc, whereas the other three mutants (*Lr*GtfC C240W$_{53608}$, *Lr*GtfC F174L$_{53608}$, *Lr*GtfC D101A$_{53608}$) showed a positive shift to UDP-GlcNAc as observed in *Lr*GtfC$_{53608}$ WT (Figure 3.5B-E).

For *Lr*GtfC$_{100-23}$, an expected positive thermal shift was observed upon addition of UDP-Glc compared to UDP-GlcNAc and no ligand conditions (Figure 3.5F). A similar effect was observed for *Lr*GtfC S238P$_{100-23}$, *Lr*GtfC L174$_{100-23}$ and *Lr*GtfC D101A$_{100-23}$ mutants (Figure 3.5G, I, J). In contrast, *Lr*GtfC W240C$_{100-23}$ showed a positive thermal shift with both UDP-GlcNAc and UDP-Glc. This was the only residue that introduced some degree of donor promiscuity effect to *Lr*GtfC$_{100-23}$ (Figure 3.5H).

The thermal shifts of *Lr*GtfCs determined from the mean derivative data in Figure 3.5 are summarised in Table 3.2. Taken together, these results showed that site-directed mutagenesis of residues in the active site of *Lr*GtfCs could induce changes in UDP-sugar specificity as was the case for *Lr*GtfC P238S$_{53608}$ to UDP-Glc or *Lr*GtfC$_{100-23}$ W240C$_{100-23}$ for both UDP-GlcNAc and UDP-Glc. Additionally, mutation of the catalytic Asp (D101) to Ala did not impact the thermal stability of both *Lr*GtfC$_{53608}$ and *Lr*GtfC$_{100-23}$, suggesting the catalytic mutants retained the

capacity to bind the ligand. The *Lr*GtfCD101A catalytic mutants remain to be tested for enzymatic activity in vitro and in ligand-complex for X-ray crystallography.
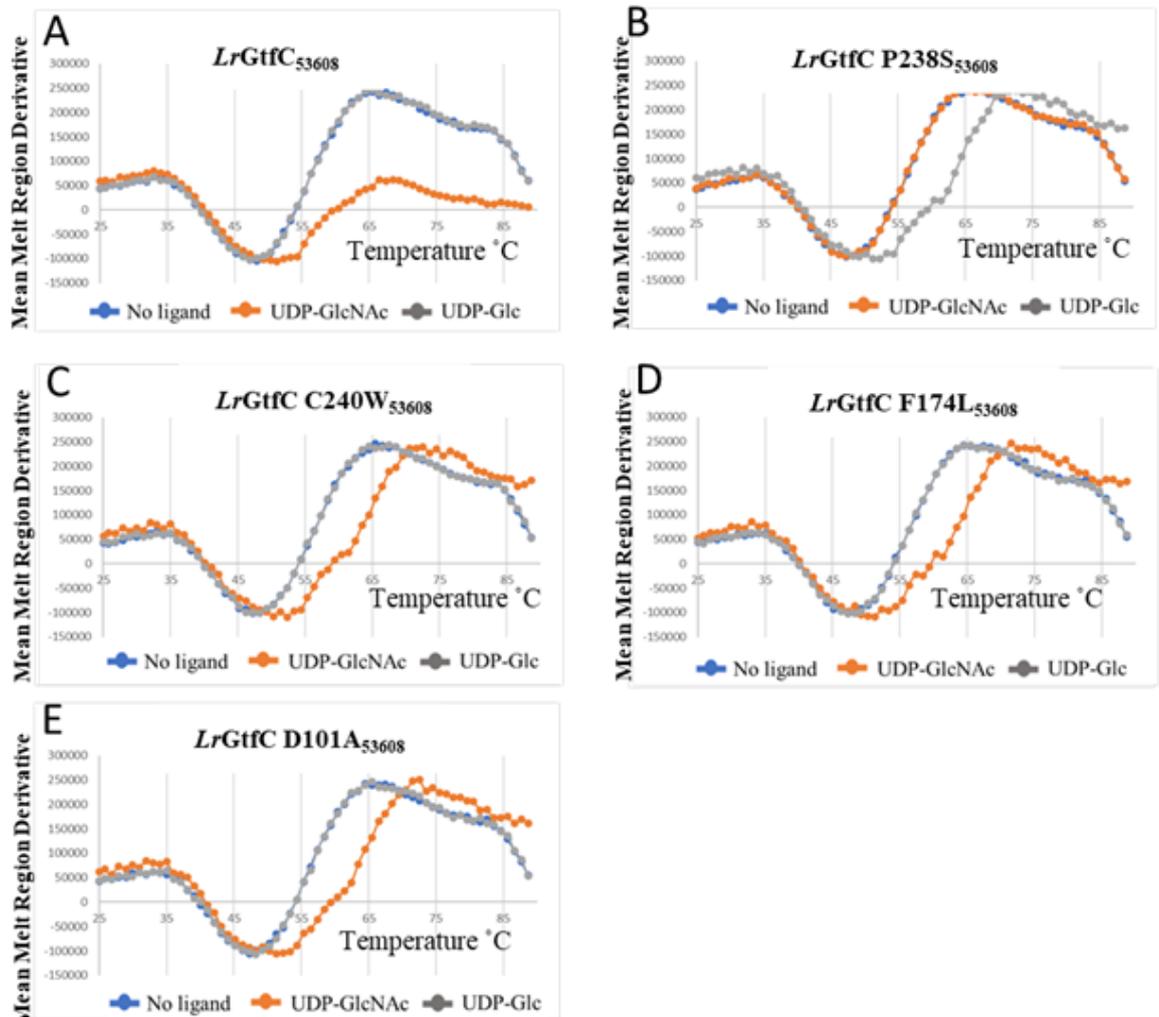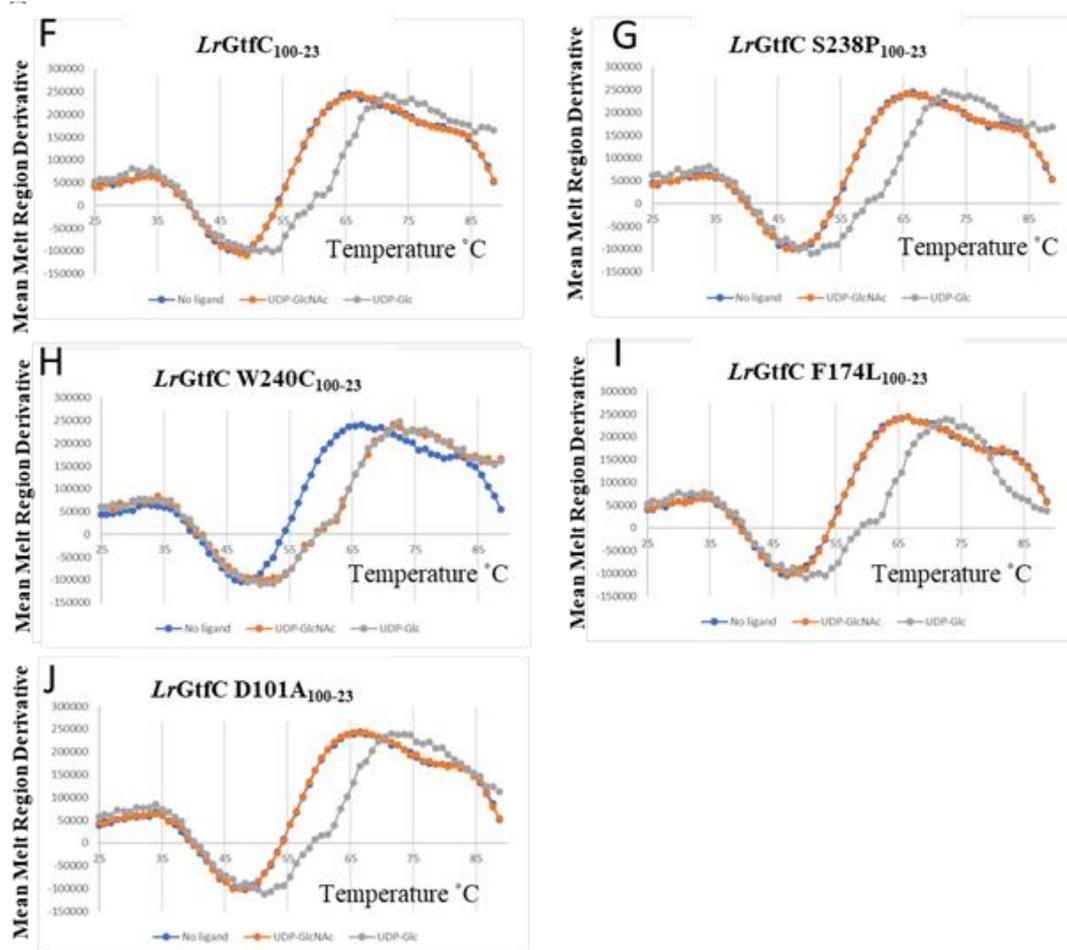
Figure 3.5. Thermal shift analysis of *Lr*GtfC variants. (A) *Lr*GtfC$_{53608}$ WT (B) *Lr*GtfC P238S$_{53608}$ (C) *Lr*GtfC C240W$_{53608}$ (D) *Lr*GtfC F174L$_{53608}$ (E) *Lr*GtfC D101A$_{53608}$ (F) *Lr*GtfC$_{100-23}$WT (G) *Lr*GtfC S238P$_{100-23}$ (H) *Lr*GtfC W240C$_{100-23}$ (I) *Lr*GtfC L174F$_{100-23}$ (J) *Lr*GtfC D101A$_{100-23}$. Recombinant *Lr*GtfC$_{53608}$ and *Lr*GtfC$_{100-23}$ WT and mutant proteins in 50 mM Tris pH 7.5 were analysed by Thermofluor in the absence or presence of 3mM UDP-Glc or 3 mM UDP-GlcNAc. n=4, data points expressed as sample means.

Table 3.2. Summary of Thermofluor binding assays of *Lr*GtfC variants.

	*Lr*GtfC[a]	Tm°C[a]		
		No ligand	**3 mM UDP Glc**	**3 mM UDP GlcNAc**
*Lr*GtfC$_{53608}$	WT	48.27 ± 0.3	48.7 ± 1.1	51.32 ± 0.4
	P238S	48.11 ± 1.2	51.41 ± 0.84	49.14 ± 0.4
	C240W	48.15 ± 0.3	49.12 ± 1.1	52.30 ± 1.3
	F174L	48.26 ± 0.4	48.24 ± 1.2	50.63 ± 1.0
	D101A	48.27 ± 0.3	48.5 ± 0.1	50.99 ± 0.4
*Lr*GtfC$_{100-23}$	WT	48.32 ± 0.2	53.34 ± 0.2	48.35 ± 0.3
	S238P	47.38 ± 0.1	50.11 ± 0.7	48.29 ± 0.3
	W240C	48.51 ± 0.4	51.14 ± 0.5	51.23 ± 0.2
	L174F	48.42 ± 0.3	51.32 ± 0.4	48.33 ± 0.1
	D101A	48.13 ± 0.2	51.33 ± 0.4	48.35 0.4

[a]Recombinant *Lr*GtfC$_{53608}$ and *Lr*GtfC$_{100-23}$ WT and mutant proteins in 50 mM Tris pH 7.5 analysed by Thermofluor in the absence or presence of 3 mM UDP-Glc or 3 mM UDP-GlcNAc. (n=4).

3.2.4 *Lr*GtfC acceptor recognition specificity

The main predicted acceptor for *Lr*GtfCs of the aSec system is the *O*-GlcNAc attached to SRRs of the SRRP secretion cargo. The extended SP of SRRPs is unique to the accessory secretion system and may play a role in the presentation of the SRRPs for their glycosylation by the aSec system. To gain insight into *L. reuteri* GtfC acceptor specificity, an ELISA based binding assay was conducted with the native fully glycosylated SRRP purified from *L. reuteri* ATCC 53608 or 100-23, recombinantly produced *Lr*gSRR1$_{100-23}$ partially glycosylated by *Lr*GtfAB$_{100-23}$ and the *Lr*SRRP$_{53608}$ signal-peptide (SP$_{53608}$).

Native SRRPs were purified from *L. reuteri* spent medium as previously described using agWGA-affinity chromatography and detected by fluorescein-WGA on western blot at a size over 240 kDa (Figure 3.6A) (235).

The *Lr*SRR1$_{100-23}$ region (spanning amino acid residues 100 to 205) cloned in pET15b and co-expressed with *Lr*GtfA-GtfB-GtfC$_{100-23}$ (*Lr*GtfABC$_{100-23}$) cloned in a pETcoco-1resulted in the expression of a glycosylated protein *Lr*gSRR1ABC$_{100-23}$ detected by fluorescein-WGA on western blot at an apparent molecular weight between 45-66 kDa (Figure 3.6B). This recombinant form of *Lr*gSRR1ABC$_{100-23}$ was previously confirmed to be predominantly glycosylated with GlcNAc

(by *Lr*GtfAB$_{100-23}$) and Glc (by *Lr*GtfC$_{100-23}$) (235). Here, a partially glycosylated gSRR1$_{100-23}$ without a his-tag was generated as the presence of his-tag on the immobilised binding partner would interfere with the design of the binding assay (outlined in section 2.7.2). Briefly, the *Lr*SRR1$_{100-23}$ region (spanning amino acid residues 100 to 205) was cloned into *E. coli* pET15b and co-expressed with *Lr*GtfAB$_{100-23}$ cloned in pETduet-1. The recombinant *Lr*gSRR1AB$_{100-23}$ was purified using agWGA chromatography and detected on western blot by fluorescein-WGA (Figure 3.6B). This protein showed a lower molecular weight, between 35-45 kDa as compared to 45-66 kDa for *Lr*gSRR1ABC$_{100-23}$, in line with the absence of additional glycosylation beyond *O*-GlcNAc.

The *Lr*SP$_{53608}$, corresponding to residues 1-95 of *Lr*SRRP$_{53608}$ was cloned into pET28b with his-tag and thrombin cleavage site which allowed for on-column his-tag removal and subsequent removal of thrombin by benzamidine affinity chromatography (see section 2.3.2 for experimental details). The recombinant *Lr*SP$_{53608}$ proteins eluted from the NiNTA and benzamidine affinity chromatography steps were analysed by SDS-PAGE, showing a band at the predicted 11 kDa (Figure 3.6C).

Time constraints and technical challenges in molecular cloning prevented production of *Lr*gSRR1$_{53608}$ and *Lr*SP$_{100-23}$ constructs. In addition, attempts to produce unglycosylated forms of *Lr*SRR or *Lr*SRRP led to misfolding and aggregation of the recombinant proteins, further confirming the importance of glycosylation in folding and secretion, as reported earlier for *Lr*SRRP$_{53608}$ (208, 235).
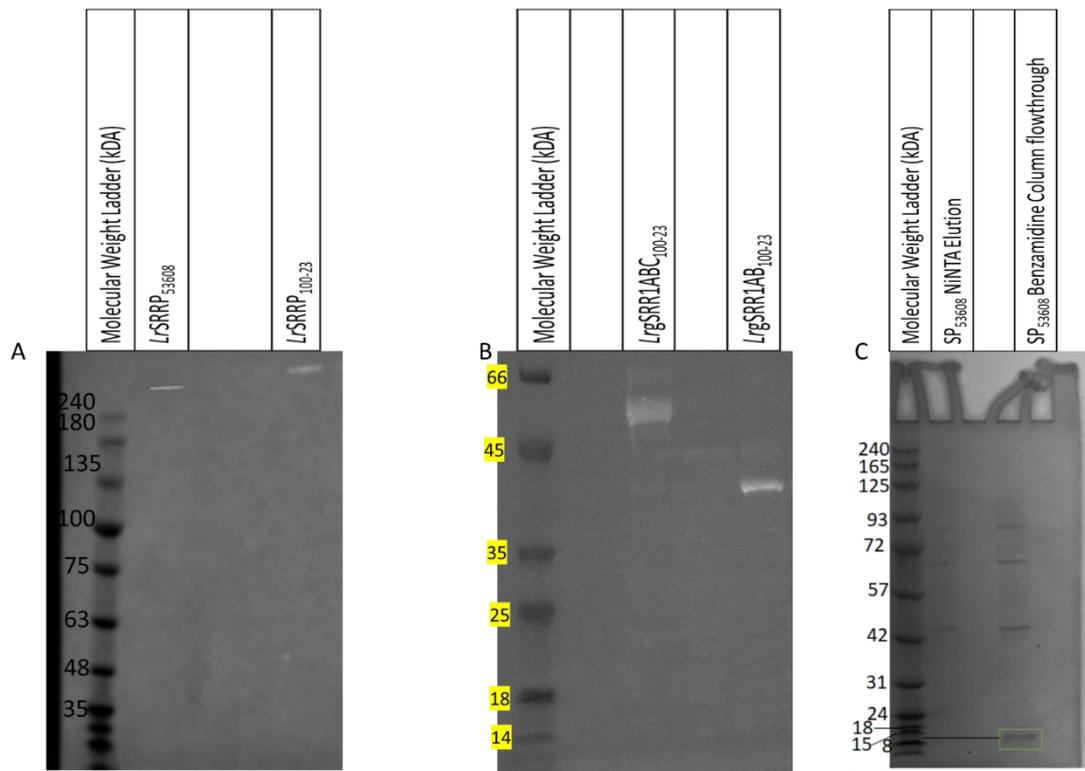
Figure 3.6. Purification of *Lr*SRRP-derived protein variants. (A) Western-blot of native *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ extracted from spent medium by ammonium precipitation and purified by WGA affinity chromatography, stained with fluorescein-WGA. (B) Western-blot of recombinant *Lr*gSRR1$_{100-23}$ glycosylated with *Lr*GtfABC$_{100-23}$ (*Lr*gSRR1ABC$_{100-23}$) or *Lr*GtfAB$_{100-23}$ (*Lr*gSRR1AB$_{100-23}$) purified by WGA affinity chromatography stained with fluorescein-WGA. (C) SDS-PAGE of SP$_{53608}$ purified by NiNTA and benzamidine affinity chromatography, stained with Coomassie instant blue stain (green box).

Binding of *Lr*GtfC$_{53608}$ and *Lr*GtfC$_{100-23}$ to immobilised *Lr*SRRPs, *Lr*gSRR1AB$_{100-23}$ partially glycosylated with *Lr*GtfAB$_{100-23}$, *Lr*SP$_{53608}$ and BSA (negative control) was determined by ELISA using an anti-his HRP fusion monoclonal antibody. Statistical inference was used to determine differences between test conditions by two-tailed student t-tests.

Both *Lr*GtfC$_{53068}$ and *Lr*GtfC$_{100-23}$ showed significant binding to their respective native *Lr*SRRPs as compared to BSA (p=0.002, degrees of freedom (df) =4) (Figure 3.7). Neither of the *Lr*GtfCs showed association with *Lr*SP$_{53608}$. Both *Lr*GtfCs significantly bound to the recombinant *Lr*gSRR1AB$_{100-23}$, with *Lr*GtfC$_{100-23}$ showing higher binding than *Lr*GtfC$_{53608}$ (p=0.005, df=4).

Taken together, these data suggest that the partially glycosylated regions of the *Lr*SRRP are the main putative binding partners of *Lr*GtfCs, irrespective of strain origin of the protein acceptor.
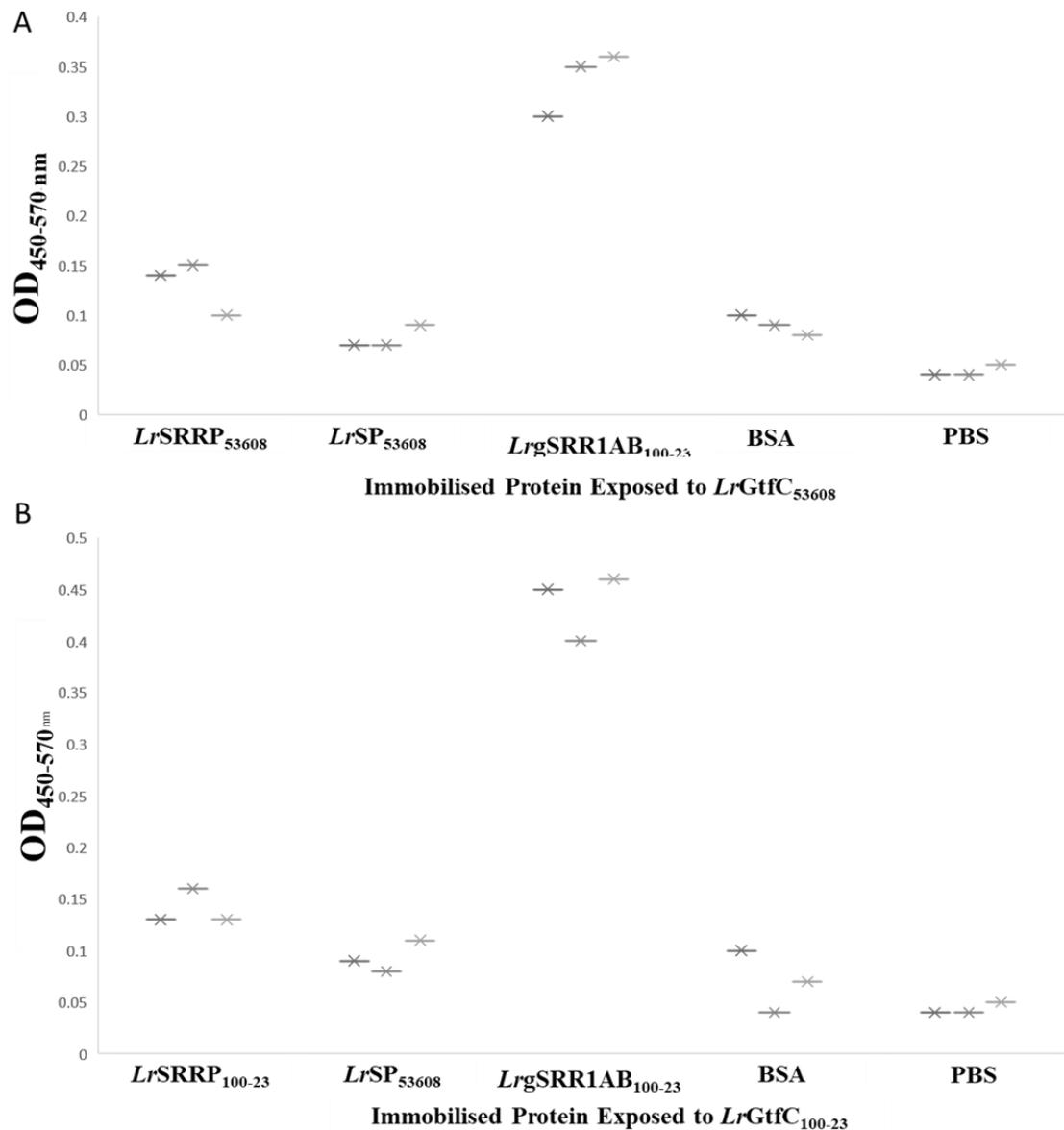
Figure 3.7. ELISA-based binding assay between *Lr*GtfCs and native *Lr*SRRPs or recombinant SRRP domains. (A) *Lr*GtfC$_{53608}$ at 20 μg/mL final concentration was added to wells coated with *Lr*SRRP$_{53608}$, *Lr*SP$_{53608}$ *Lr*gSRR1AB$_{100\text{-}23}$, BSA and PBS. (B) *Lr*GtfC$_{100\text{-}23}$ at 20 μg/mL final concentration was added to wells coated with *Lr*SRRP$_{100\text{-}23}$, *Lr*SP$_{53608}$ *Lr*gSRR1AB$_{100\text{-}23}$ BSA and PBS. Experiments were performed in triplicate and binding by monitoring OD 450-570 nm following TMB hydrolysis by anti-his-HRP monoclonal antibodies.

3.2.5 *Lr*GtfAB protein structures

The Alphafold2 models revealed GT-B folds for both *Lr*GtfA$_{53608}$ (green) and *Lr*GtfB$_{53608}$ (blue) (Figure 3.8A). The typical Rossman folds forming the canonical 'U' shape was apparent for both enzymes. These models were superimposed with the existing crystal structure of the GtfAB tetramer from *S. gordonii* M99 strain (*Sg*GtfAB$_{M99}$) co-crystallised with UDP-GlcNAc (PDB: 59EU) (276). On inspection of the active site of *Lr*GtfAB$_{53608}$, the residues needed for catalysis (Glu413) or UDP binding (Lys342 and Arg337) were found to be conserved in *Lr*GtfA$_{53608}$ (Figure 3.8B). *Lr*GtfB$_{53608}$ lacked these residues and the cleft itself was smaller (12Å) when compared to *Lr*GtfA$_{53608}$ (27.9 Å), as depicted in the molecular surface model representation (Figure 3.8C).

*Lr*GtfA$_{53608}$ and *Lr*GtfB$_{53608}$ share 74% and 75% amino acid sequence identity with *Lr*GtfA$_{100-23}$ and *Lr*GtfB$_{100-23}$, respectively. The structural model of *Lr*GtfAB$_{100-23}$ indicated a high structural homology to *Lr*GtfAB$_{53608}$, with conserved active site residues in *Lr*GtfA$_{100-23}$ (Glu418, Lys345, Arg342) (Figure 3.8D). In contrast, residues involved in catalysis, or UDP-sugar binding were absent in *Lr*GtfB$_{100-23}$. The cleft size of *Lr*GtfB$_{100-23}$ was smaller than *Lr*GtfA$_{100-23}$ at 15.4 and 24.6 Å, respectively, whereas for the *Lr*GtfAB$_{53608}$ complex this was 12 and 27.9 Å for *Lr*GtfB$_{53608}$ and *Lr*GtfA$_{53608}$, respectively.
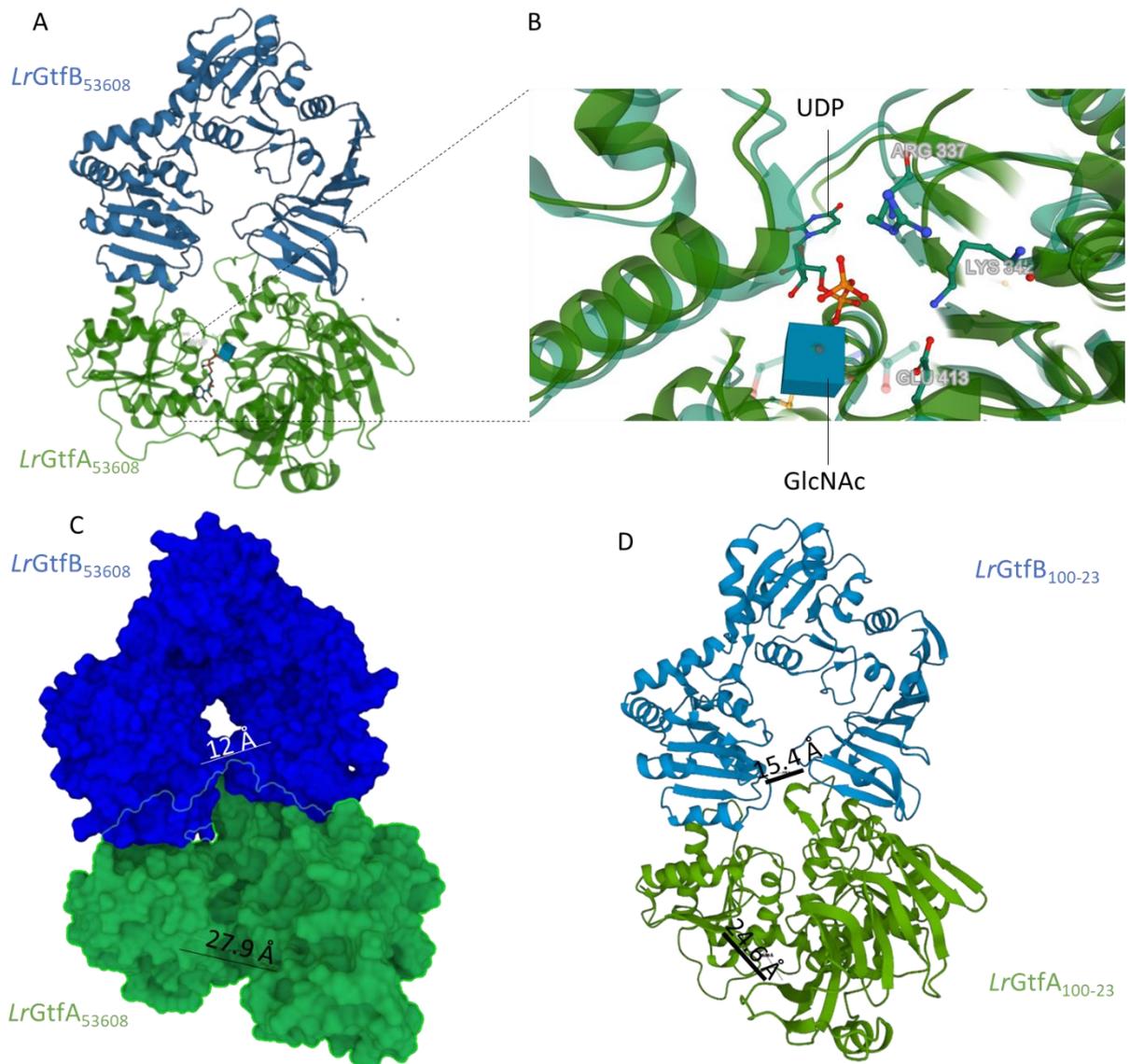
Figure 3.8 Alphafold2 structural model of the *Lr*GtfAB complex. (A) *Lr*GtfA$_{53608}$ (green) and *Lr*GtfB$_{53608}$ (blue) cartoon structure representation generated using Alphafold2 superimposed with *Sg*GTFAB$_{M99}$ crystal structure (59EU) in front view (276) (B) Close-up view of *Lr*GtfA$_{53608}$ active site superimposed with *Sg*GtfA$_{M99}$ with key active residues labelled (Arg337, Lys342, Glu413) and UDP GlcNAc ligand. (C) Front view of *Lr*GtfAB$_{53608}$ as a molecular surface model with width of cleft (Å) highlighted in this conformation. (D) Front view of *Lr*GtfAB$_{100-23}$ with width of cleft (Å) highlighted in this conformation.

3.3 Discussion

Glycosylation is a critical step in preserving the structure and function of the accessory secretion cargo SRRP facilitated by the sequential addition of sugar moieties by GTs (207). Here, a

combination of X-ray crystallography, machine learning modelling (Alphafold2) and homology modelling approaches was used to structurally characterise *L. reuteri* GTs involved in the aSec system and subsequently guide biochemical determination of the mechanism underpinning their sugar-donor/acceptor substrate specificity.

The first step in SRRP glycosylation is carried out by the GtfAB complex which catalyses *O*-glycosylation of SRR regions by covalent attachment of the core GlcNAc moiety with the hydroxyl group of Ser/Thr residues (193). This was demonstrated in gram-positive pathogens such as *S. gordonii*, *S. parasanguinis* and *S. agalactiae* strains and more recently in *L. reuteri* ATCC 53608 and 100-23 strains where *Lr*GtfAB was shown to add GlcNAc (222, 235, 279, 280). Genetic disruption of either *gtfA* or *gtfB* in *S. gordonii* M99 prevented platelet binding by the GspB SRRP adhesin required for pathogenesis due to a complete loss of glycosylation (281, 282).

Here, structural modelling predicted that *Lr*GtfAB is comprised of two GT-B type GTs with distinct substrate catalytic and acceptor recognition roles. Superimposition of *Lr*GtfAB model to SgGtfAB$_{M99}$ complex bound to UDP-GlcNAc confirmed the presence of putative catalytic nucleophile (Glu) and UDP-sugar binding residues (Lys, Arg) in LrGtfA$_{53608}$ and LrGtfA$_{100-23}$ of *Lr*GtfAB (276). These residues were shown to be required for catalysis and ligand binding in SgGtfA$_{M99}$ by *in vitro* glycosylation assays (276). Glu has also been identified as a nucleophile in other characterised GTs from both prokaryotes (e.g., *S. enterica* serovar Typhi) and eukaryotes (*Anemarrhena asphodeloides*, or plants) (283, 284). Lys and Arg residues are also important components in UDP binding in a range of glyco-active enzymes, such as epimerases (285-287). In contrast to GtfAs, GtfBs from *L. reuteri* and *S. gordonii* strains lacked all of these key residues, which supports the auxiliary role of these proteins as acceptor recognition.

Another interesting facet of the *Lr*GtfAB complex relates to the predicted size and conformation of the cleft formed between the respective C- and N-terminal domains of the GTs. The size of binding clefts of both LrGtfB$_{53608}$ and LrGtfB$_{100-23}$, of 12 Å, and 15.4 Å, respectively in the structural predictions was larger than that reported for SgGtfB$_{M99}$ at approximately 8 Å (276). The GtfA clefts were more structurally conserved between species as overall size and distances across

the clefts measured within 3 Å of each other (276). The reason for GtfB predicted cleft size discrepancies is unclear but may be due to structural differences in topography of Ser/Thr acceptor sites in SRRPs. In summary, *L. reuteri* GtfAB structural models provided further evidence that main residues are shared with pathogens in line with the conserved capacity of *Lr*GtfAB to deposit the primary *O*-GlcNAc.

The next step in *L. reuteri* SRRP glycosylation is carried out by *Lr*GtfC, which introduces the strain-specific glycosylation pattern, whereby *Lr*GtfC$_{53068}$ shows a preference for UDP-GlcNAc and *Lr*GtfC$_{100-23}$ for UDP-Glc, respectively (235). Here, the crystal structure of *Lr*GtfC$_{100-23}$ was solved in an apoenzyme and in UDP-bound or UDP-GlcNAc bound state. The protein crystallised as a tetramer and occupied P 21 21 21 space group, which is consistent with other GtfC structures available in the protein data bank i.e., Gtf3 of *S. parasanguinis* (*Sp*Gtf3$_{FW213}$) (3QKW) and GtfC of *S. agalactiae* COH1 (*Sa*GtfC$_{COH1}$) (4W6Q) (238, 288). The presence of UDP bound to the active site was consistent with previous work showing binding of *Lr*GtfC$_{100-23}$ to UDP, UDP-Gal, UDP-GlcNAc and most favourable binding with UDP-Glc using Thermofluor and STD NMR (235). As UDP-GlcNAc was not the preferred substrate of *Lr*GtfC$_{100-23}$, co-crystallisation may have been facilitated as it could be retained in the active site without being hydrolysed. Simulating the transition from ligand bound and unbound states demonstrated flexibility of the GT-B hinge region which is consistent with the reported mobility of GTs (264, 268). Additionally, several amino acid residues specific of *Lr*GtfC$_{53608}$ or *Lr*GtfC$_{100-23}$ in the active site were identified through superimposition with the homology model of *Lr*GtfC$_{53608}$ (generated based on the *Lr*GtfC$_{100-23}$ crystal structure). Residues differing in the active site of *Lr*GtfC$_{100-23}$ and *Lr*GtfC$_{53608}$ were as follows: Leu174, Ser238 and Trp240 in *Lr*GtfC$_{100-23}$ corresponding to Phe174, Pro238 and Cys240 in GtfC$_{53608}$. Some of these are shared with *Sp*Gtf3$_{FW213}$ (a glucosyltransferase) which has Phe172, Ser237 and Cys242 at the corresponding positions showing conservation across other GtfCs (289).

Site-directed mutagenesis was used to validate the role of these residues in terms of the UDP-sugar binding specificity of *Lr*GtfC$_{53608}$ and *Lr*GtfC$_{100-23}$. The change of Pro238 to Ser238 in

LrGtfC$_{53608}$ induced a thermal shift in favour of UDP-Glc, as reported for LrGtfC$_{100-23}$, indicating that Ser238 may facilitate UDP-Glc specificity (235). Conversely, mutation of Ser238 to Pro238 in LrGtfC$_{100-23}$ retained preference for UDP-Glc, suggesting that multiple residues facilitate UDP-GlcNAc specificity. Following sequence analysis across *L. reuteri* strains analysed, only LrGtfC$_{100-23}$ was found to possess Ser238, however, SpGtf3$_{FW213}$ (which shares a preference for UDP-Glc) has a Ser at the same position as well as the other potential residues involved in ligand specificity in LrGtfC P238S$_{53608}$ (Leu174, S238, C240). The only other mutation that induced a differential thermal shift compared to the WT enzymes, was W240C in LrGtfC$_{100-23}$ which seemed to favour binding to both UDP-Glc and UDP-GlcNAc. Introduction of promiscuity has been achieved in various glycoengineering attempts in other GT systems (290, 291). Taken together, this approach showed that it may be possible to switch the substrate specificity of aSec GtfCs by the modification of a single amino acid, but multiple mutations may be required to induce further changes in UDP-Glc and UDP-GlcNAc specificities.

The *O*-GlcNAc site of LrSRR regions is the proposed acceptor for LrGtfCs, however, it has been shown that glycan elongating GTs, like GtfC, can also recognise underlying polypeptide sequences of their acceptor (292). Based on ELISA binding assays, we showed that binding of LrGtfC$_{53608}$ and LrGtfC$_{100-23}$ was highest to the recombinant, partially glycosylated gSRR1$_{100-23}$ (i.e., glycosylated by LrGtfAB$_{100-23}$) as compared to the fully glycosylated native LrSRRP$_{53608}$ or LrSRRP$_{100-23}$, or recombinant LrSP$_{53608}$. The interaction between LrGtfC$_{100-23}$ and LrgSRR1AB$_{100-23}$ was stronger than that of LrGtfC$_{53608}$ with LrgSRR1AB$_{100-23}$, suggesting that while the primary *O*-GlcNAc was the main acceptor irrespective of LrGtfC strain origin, there may be underlying interactions with the polypeptide backbone driven by host specificity influencing acceptor specificity. The weaker binding observed with native LrSRRPs may be due to steric hindrance caused by further post translational modifications, i.e. further glycosylation having already occurred on the LrSRRP$_{100-23}$ or *O*-acetylation, which can prevent over-glycosylation (217).

O-glycosylation systems are increasingly being discovered in a range of prokaryotes with glycoproteins diversifying from the traditional examples of S-layer proteins, flagellins or pilins

(293). The enormous variability present in the glycan structures from bacteria (often different from mammalian ones) and in the underlying GTs is accompanied with technical difficulties in isolating and analysing glycoproteins (294). This complexity is illustrated for example in the oral pathogen *Porphyromonas gingivalis* DPG3 by the glycosylation of fimbriae with fucose, xylose, mannose, Gal, Glc, N-acetylgalactosamine and GlcNAc residues as revealed by gas chromatography mass spectrometry (GC-MS) (295). Protein glycosylation is often linked with an organism's ecological fitness, so GTs are likely to be under severe evolutionary pressure for the conservation of residues dictating donor-specificity, whereby single-nucleotide polymorphisms could have a major impact on recognition, as reported here for *L. reuteri* GtfCs.

Future work should expand the range of *Lr*GtfC mutants to identify the residues required for UDP-GlcNAc or UDP-Glc specificity. Whilst modification of W240C seems to have introduced promiscuity for both UDP-Glc and UDP-GlcNAc, the underlying reason for this is unclear. Further binding assays such as isothermal titration calorimetry (ITC) or biolayer interferometry (BLI) will help derive binding kinetics to validate these findings. Additionally, it would be of interest to determine the glycosylation profile of recombinant gSRR1s using *Lr*GtfC mutants generated in this work. Finally, solving the crystal structure of ligand-bound *Lr*GtfC mutants, such as *Lr*GtfC P238S$_{53608}$ in complex with UDP-Glc, would help identify the structural basis of ligand specificity.

Chapter 4 Structure and Function of Accessory Secretion Proteins (Asps) of the *L. reuteri* aSec System

4.1 Introduction

Chaperones play a crucial role in many secretion systems in both prokaryotes and eukaryotes to mediate sequential processing and hierarchy within a secretion pathway (296, 297). Post-translational modifications such as glycosylation, ubiquitination, prenylation, phosphorylation and acetylation are all processes carried out by bacteria, which greatly expand the diversity of the secreted proteome (298-300). In addition to the presence of GTs for glycosylation of SRRPs, the aSec system contains chaperones called accessory secretion proteins (Asps).

Asp1, Asp2 and Asp3 have been extensively investigated in the pathogenic organisms *S. gordonii* M99 and *S. pneumoniae* TIGR4 where they are predicted to not only help traffic the SRRP to the secretion machinery, but also facilitate *O*-acetylation through the formation of an Asp1-Asp2-Asp3 complex (216, 219). The crystal structures of $SgAsp1_{M99}$, $SgAsp1\text{-}Asp3_{M99}$ and $SpAsp1\text{-}Asp2\text{-}Asp3_{TIGR4}$ with partial degradation of C-terminal regions of Asp2 and Asp3 have been determined. (216, 217). These revealed an overall inactive GT-B type fold for $SgAsp1_{M99}$ and predicted carbohydrate binding module (CBM) for $SgAsp3_{M99}$, in line with their putative functions for the secretion of SRRP (GspB). $SgAsp2_{M99}$ was proposed to be an *O*-acetyltransferase which will prevent potential over-glycosylation of the secretion cargo (218, 219). Additionally, it was reported that $SgAsp1\text{-}Asp2\text{-}Asp3_{M99}$ and $SpAsp1\text{-}Asp2\text{-}Asp3_{TIGR4}$ respectively form a complex and that $SgAsp1_{M99}$ and $SgAsp3_{M99}$ can interact with $SgSecA2_{M99}$, suggesting a chaperoning function. Liposome co-flotation assays have also shown association of the $SgAsp1\text{-}Asp2\text{-}Asp3_{M99}$ complex with lipids, so to help modify and bring the SRRP to the secretion machinery at the cell membrane (217).

This chapter reports the biochemical and structural analysis of Asp1, Asp2 and Asp3 from *L. reuteri* ATCC 53608 and 100-23 strains as individual components and in complex.

4.2 Results

4.2.1 Production and analysis of recombinant *Lr*Asps

For the biochemical and structural characterisation of Asps encoded by the *L. reuteri* aSec operon, recombinant Asps from ATCC 53608 and 100-23 strains were produced heterologously in *E. coli* using the plasmids containing N-terminal his-tags for $LrAsp1_{53608}$, $LrAsp2_{53608}$, $LrAsp3_{53608}$, $LrAsp2_{100-23}$ in pET28b, whereas $LrAsp1\text{-}Asp2\text{-}Asp3_{53608}$ (including intergenic regions) was cloned into pET15b. In addition, site-directed mutagenesis was performed to generate $LrAsp2$ $S349A_{53608}$ using the SPRIP protocol (237) (see section 2.3 for experimental details).

Recombinant Asp WT and mutants were purified by nickel affinity chromatography (NiNTA) as his-tagged proteins. The apparent size of recombinant *Lr*Asps proteins was shown by SDS-PAGE to be around 60 kDa for $LrAsp1_{53608}$, 58 kDa for $LrAsp2_{53608,}$ $LrAsp2S349A_{53608}$, $LrAsp2_{100-23}$ and 34 kDa for $LrAsp3_{53608}$, respectively (Figure 4.1AB). The recombinant protein consisting of $LrAsp1\text{-}Asp2\text{-}Asp3_{53608}$ was purified by NiNTA followed by size-exclusion gel filtration (SEGF) chromatography and analysed by native PAGE, suggesting the formation of a 1:1:1 $LrAsp1_{53608}$: $LrAsp2_{53608}$: $LrAsp3_{53608}$ complex at approx. 150 kDa (Figure 4.1CD).
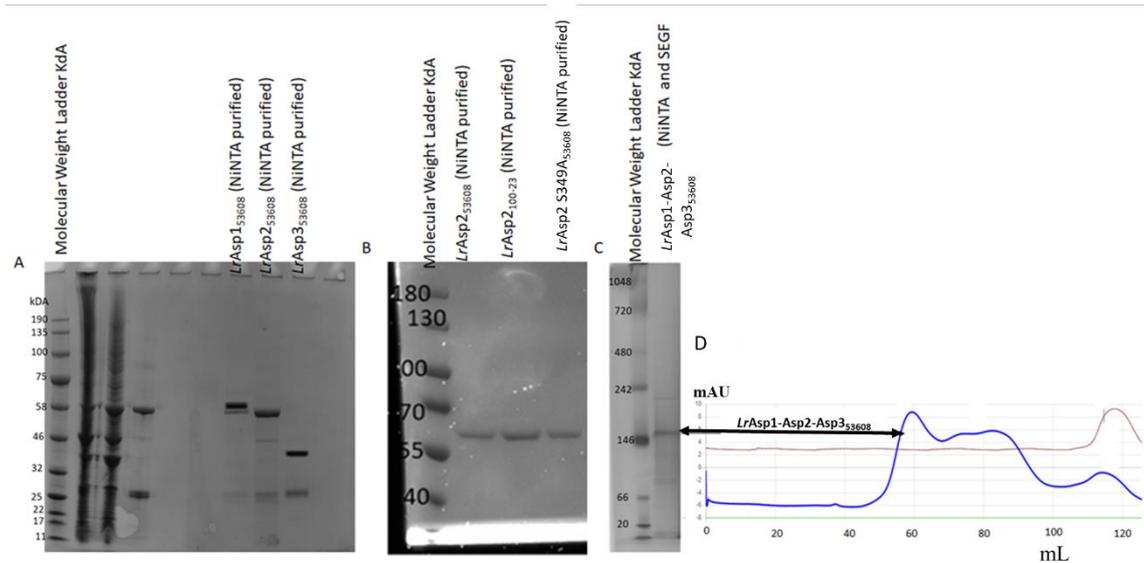
Figure 4.1. Gel electrophoresis of purified recombinant *Lr*Asp variants from *L. reuteri* ATCC 53608 and 100-23 strains. (A) Recombinant $LrAsp1_{53608}$, $LrAsp2_{53608}$ and $LrAsp3_{53608}$ expressed from pET28b purified by NiNTA and analysed by SDS-PAGE (B) Recombinant $LrAsp2_{53608}$, $LrAsp2S349A_{53608}$ and $LrAsp2_{100-23}$ purified NiNTA and analysed by SDS-PAGE. (C) Recombinant $LrAsp1$-$Asp2$-$Asp3_{53608}$ purified by NiNTA and SEGF analysed by native PAGE. The gels were stained with coomassie (D) Recombinant $LrAsp1$-$Asp2$-$Asp3_{53608}$ purified by SEGF showing UV-absorption of eluting proteins.

The yield of the purified recombinant *Lr*Asp proteins varied between 1 mg to 6 mg with $LrAsp1$-$Asp2$-$Asp3_{53608}$ complex showing the lowest yield (Appendix 3, supplementary Table S3.1).

4.2.2 Structure and function of *Lr*Asp1

Asp1 has been attributed the role of chaperone in pathogenic aSec models, such as for *S. gordonii* M99 and *S. pneumoniae* TIGR4, due to structural similarities to carbohydrate binding proteins, lipid association and binding of GspB to $SgAsp1_{M99}$ (217). To test the possible chaperoning function of *Lr*Asp1, the binding of recombinant $LrAsp1_{53608}$ was tested against native *Lr*SRRPs, the recombinant $LrgSRR1AB_{100-23}$ and $LrSP_{53508}$ (generated in section 3.2.4) by ELISA using anti-his HRP fusion monoclonal antibody for detection (see 2.7.1 for experimental details). Statistical inference by two-tailed student t-test was used to determine differences between conditions.

$LrAsp1_{53608}$ bound significantly to both native $LrSRRP_{53608}$ and $LrSRRP_{100-23}$ when compared to BSA (p<0.0001, df=6, t=13.14 and p<0.0001, df=6, t=15.2, respectively) (Figure 4.2). However,

no statistically significant difference in binding was observed between LrAsp1$_{53608}$ and LrSRRP$_{53608}$ and LrSRRP$_{100-23}$ (p=0.09, df=6, t=2.2024), indicating that, *in vitro*, binding can occur irrespective of the strain origin of the secretion cargo. No binding could be determined between LrAsp1$_{53608}$ and the LrSP$_{53608}$ or LrgSRR1AB$_{100-23}$. The absence of binding between LrAsp1$_{53608}$ and partially glycosylated LrgSRR1AB$_{100-23}$ suggests that interaction between LrAsp1$_{53608}$ and the SRRP may occur after *O*-glycosylation to facilitate the sequential processing in the aSec pathway, i.e., premature SRRPs are not bound, or chaperoned to the secretion complex. However, it cannot be excluded that the absence of binding was due to differences in glycosylation present on LrgSRR1AB$_{100-23}$ compared to LrSRRP$_{53608}$.
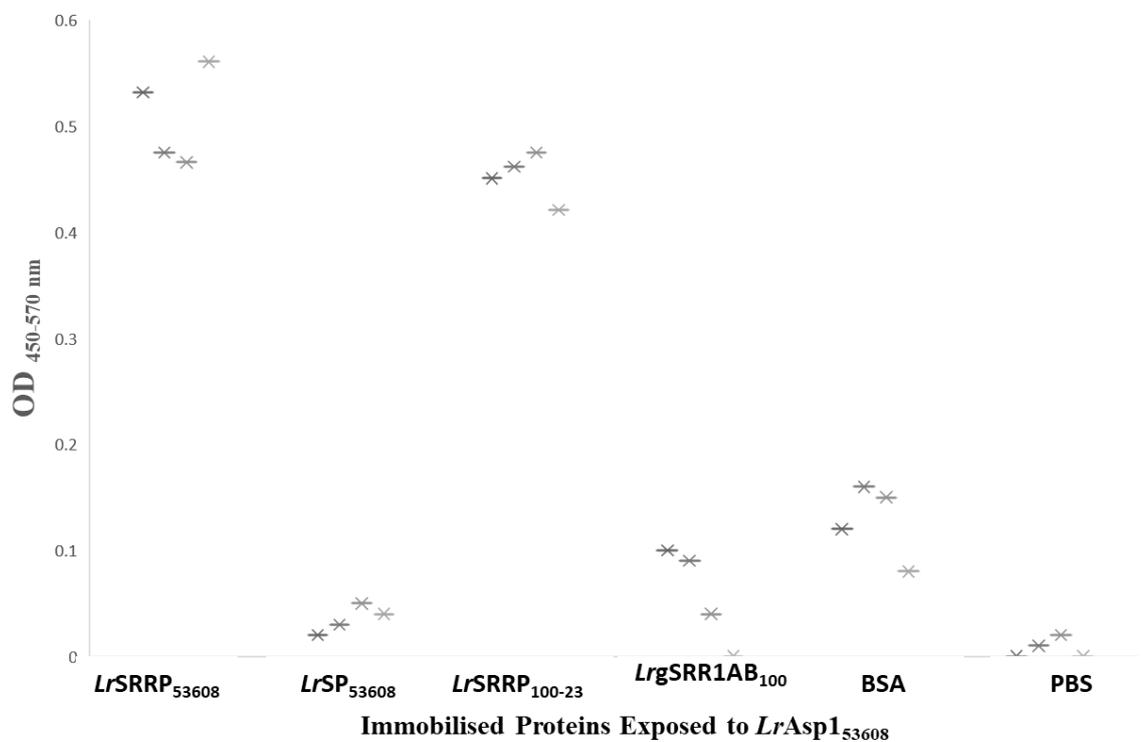
Figure 4.2. ELISA-based binding assay between LrAsp1$_{53608}$ and LrSRRP components. LrAsp1$_{53608}$ at 15 µg/mL final concentration was added to wells coated with LrSRRP$_{53608}$, LrSP$_{53608}$, LrSRRP$_{100-23}$, LrgSRR1$_{100-23}$, BSA or PBS. Binding was determined by monitoring OD$_{450-570nm}$ following hydrolysis of TMB substrate by anti-his-HRP monoclonal antibodies (n=4).

To elucidate the structural basis for LrAsp1 binding to LrSRRP, crystallisation of recombinant LrAsp1$_{53608}$ was attempted, but protein crystals could not be obtained in any of the conditions

tested which included screens of various precipitants, salts, and pH (see Table 2.6). As a consequence, LrAsp1$_{53608}$ and LrAsp1$_{100-23}$ structural models were constructed from sequence data using Alphafold2 (250, 301). LrAsp1$_{53608}$ and LrAsp1$_{100-23}$ share 75% amino acid sequence identity, and both structures appeared highly similar, displaying the two typical Rossman folds consistent with GT-B glycosyltransferases (Figure 4.3AB). Both structures showed the presence of a flexible loop in the C-terminal domain (residues ~290-320 in LrAsp1), which has been highlighted as more disordered during model construction and does have some amino acid differences between LrAsp1$_{53608}$ and LrAsp1$_{100-23,}$ suggesting strain-specific determinants. However, no pattern or conservation of these particular residues within this loop could be determined when compared with LrAsp1 from other $L. reuteri$ strains (not shown).

Structural homology was overall conserved between LrAsp1 presented here and LrAsp1 X-ray crystal structures from pathogens (SgAsp1$_{M99}$ and SpAsp1$_{TIGR4}$) (Figure 4.3C) (216, 302), indicating that the structure-function relationship of LrAsp1 may be conserved across gram-positive species.

A chaperoning role for Asp1 was further evidenced by the structural homology of LrAsp1$_{100-23}$ and LrAsp1$_{53608}$ with LrGtfB of the aSec system, as shown upon superimposition of LrAsp1$_{100-23}$ and LrAsp1$_{53608}$ with the surface model of the LrGtfB$_{53608}$ 3D-model constructed previously (section 3.2.5) (Figure 4.3D). As shown for LrGtfB$_{53608}$, no catalytic residues characteristic of glycosyltransferases could be identified, suggesting that LrAsp1$_{53608}$ and LrAsp1$_{100-23}$ may be involved in substrate recognition, i.e., binding of SRRPs through carbohydrate or protein interactions, rather than an enzymatic function. However, LrGtfB$_{53608}$ structure models did not share the flexible C-terminal loop extension present in LrAsp1, and the predicted cleft opening highlighted above was wider in LrAsp1$_{53608}$ and LrAsp1$_{100-23}$ (16 Å and 20 Å, respectively) compared to 12 Å in the LrGtfB$_{53608}$ structural prediction, which may help accommodate the different sizes of O-glycan chains decorating SRRPs.
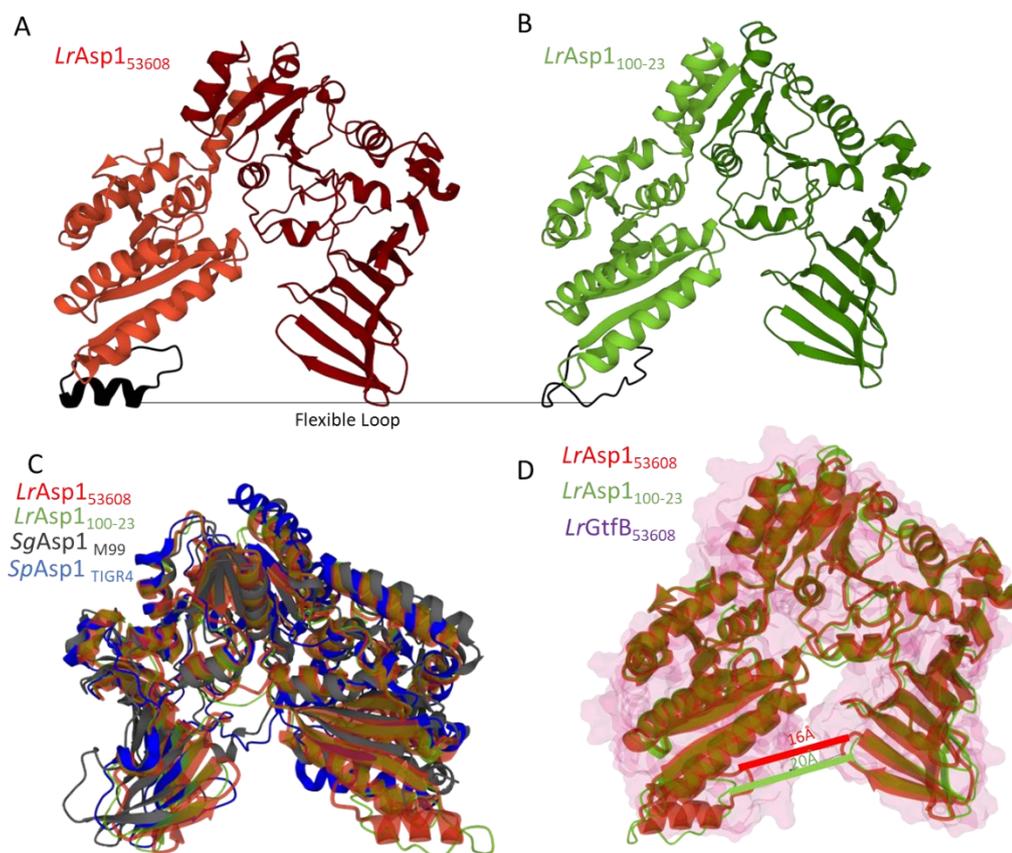
Figure 4.3. Alphafold2 structural models of *Lr*Asp1₅₃₆₀₈ and *Lr*Asp1₁₀₀₋₂₃. (A) Front-view of cartoon *Lr*Asp1₅₃₆₀₈ structural model (N-terminal domain in dark red, C-terminal domain in light red, C-terminal flexible loop in black) (B) Front-view of cartoon Asp1₁₀₀₋₂₈ structural model (N-terminal domain in dark green, C-terminal domain in light green, C-terminal flexible loop in black) (C) Superimposition of cartoon *Lr*Asp1₅₃₆₀₈, *Lr*Asp1₁₀₀₋₂₃ with surface-model representation of *Lr*GtfB₅₃₆₀₈ (purple). (D) Superimposition of *Lr*Asp1₅₃₆₀₈, *Lr*Asp1₁₀₀₋₂₃, *Sg*Asp1ₘ₉₉ (5VAF, black) and *Sp*Asp1ₜᵢGR4 (6LNW, blue) (216, 302).

4.2.3 Structure and function of *Lr*Asp2

Recombinant *Lr*Asp2₅₃₆₀₈ and *Lr*Asp2₁₀₀₋₂₃ were first tested for their acetylesterase activity *in vitro*. This also included a putative catalytic mutant *Lr*Asp2S349A₅₃₆₀₈, which was generated based on structural homology with a *Staphylococcus aureus O*-acetyltransferase OatA catalytic domain crystal structure (6VJP) (303). Acetylesterase activity refers to the hydrolysis of an acetic-ester using water releasing acetate and an alcohol, therefore it can be determined colorimetrically at 405 nm by monitoring *p*NP released from *p*NP-acetate cleavage quantified by using a *p*NP standard curve (Figure 4.4A). A velocity plot was derived as a function of substrate concentration

to determine kinetic parameters (Figure 4B, Table 4.1). Both recombinant LrAsp2$_{53608}$ and LrAsp2$_{100-23}$ exhibited similar kinetic parameters with a k$_{cat}$ of 1.26 ± 0.0009 and 1.3 ± 0.0007, respectively. Furthermore, the catalytic mutant LrAsp2 S349A$_{53608}$ showed loss of enzymatic activity, confirming S349 as the putative nucleophile for acetate-ester hydrolysis.
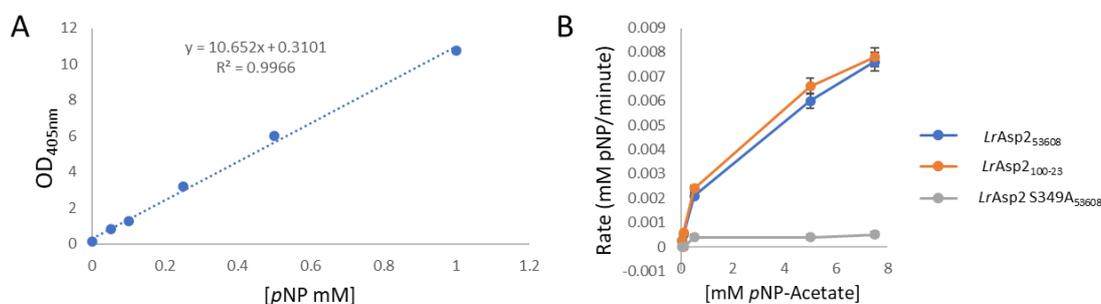
Figure 4.4 LrAsp2 acetylesterase activity with pNP-acetate. (A) Standard curve of pNP absorbance at OD$_{405nm}$ used for quantification of pNP released during pNP-substrate hydrolysis. (B) Rate (mM/min pNP produced) of pNP-acetate hydrolysis by LrAsp2$_{53608}$ (in blue), LrAsp2$_{100-23}$ (in orange) and LrAsp2S349A$_{53608}$ (in grey) with increasing pNP-acetate substrate concentrations (from 0.5 - 7.5 mM). Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl. Mean rates were derived with 10 replicates and SD error bars are displayed. Blank reactions containing no LrAsp2$_{53608,}$ or LrAsp2$_{100-23}$ were used as negative controls.

Table 4.1. Kinetic parameters of LrAsp2$_{53608}$ and LrAsp$_{100-23}$ acetylesterase activity with pNP-acetate.

Enzymes	Kinetic parameters		
	V_{max} (µmol pNP/min)	K_m (µm)	k$_{cat}$ (µmol/min/µmol)
LrAsp2$_{53608}$	7.7 ± 0.04	162 ± 15	1.26 ± 0.0009
LrAsp2$_{100-23}$	7.8 ± 0.04	160 ± 11	1.3 ± 0.0007

To test other potential acetyl-donors and inhibitors for LrAsp2, an indoxylacetate hydrolysis assay was established. Here, the activity of LrAsp2$_{53608}$, LrAsp2$_{100-23}$ and LrAsp2S349A$_{53608}$ was determined by monitoring the release of the indoxyl molecule which can be detected at OD$_{375nm}$ following the enzymatic hydrolysis of indoxylacetate. The concentration of indoxyl was calculated using the Beer-Lambert law (ε = 2.54 mM^{-1} cm^{-1}). The reaction showed release of the indoxyl molecule for LrAsp2$_{53608}$ and LrAsp2$_{100-23}$ (Figure 4.5A). LrAsp2$_{53608}$ Ser349A showed no acetylesterase activity. A velocity plot was derived as a function of substrate concentration to determine the kinetic parameters (Figure 4.5B, Table 4.2). Both LrAsp2$_{53608}$ and LrAsp2$_{100-23}$

exhibited similar kinetic parameters with this substrate and a kcat of 3.13 ± 0.0019 and 3.25 ± 0.0004 μmol/min/μmol, respectively. This suggests that the turn-over of indoxylacetate is faster than pNP-acetate (1.26 ± 0.0009 and 1.3 ± 0.0007 μmol/min/μmol), suggesting a preference for certain acetyl-donor. It should be noted that while buffer salt and pH were accounted for and kept the same during kinetics, indoxylacetate was solubilised in DMSO, whereas pNP-acetate was solubilised in ethanol which may have an influence on the enzymatic activity.

Paraoxonethyl is a common inhibitor of acetylcholinesterases and irreversibly disrupts the hydrolytic activity by blocking the catalytic residue Ser of these enzymes (304). Addition of paraoxonethyl to the enzymatic reaction at 20 μM final concentration abolished acetylesterase activity of LrAsp2$_{53608}$ and LrAsp2$_{100\text{-}23}$ (Figure 4.5A). To determine the potency of the inhibitor, LrAsp2$_{53608}$ kinetic parameters were measured on indoxylacetate with concentrations of paraoxonethyl ranging from 2.5 μM to 35 μM and the % of inhibition was derived as a function of the enzymatic rate without inhibitor (Figure 4.5C). The half-maximal inhibitory concentration (IC50) for paraoxonethyl acting on LrAsp2$_{53608}$ was calculated to be 11 ± 0.4 μM.
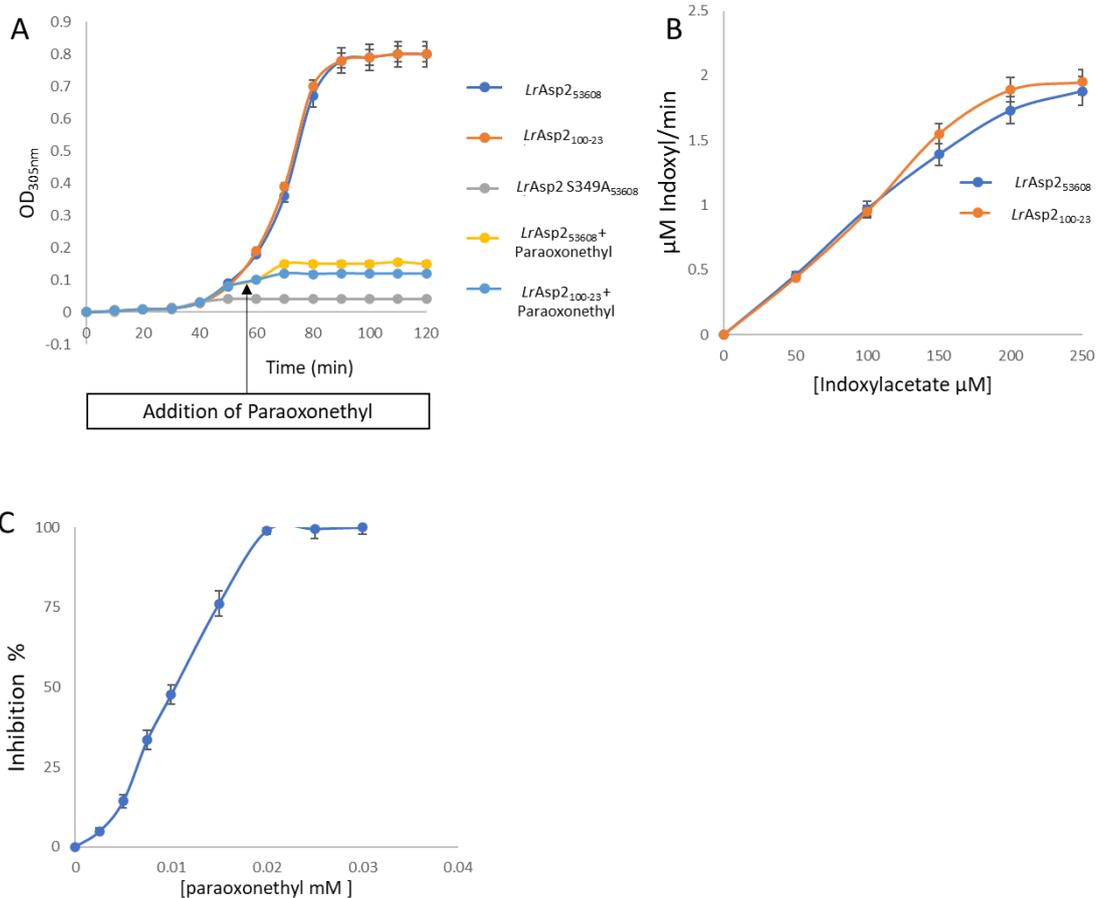
Figure 4.5. Rate of *Lr*Asp2 acetylesterase activity with indoxylacetate. (A) Indoxylacetate hydrolysis *Lr*Asp2$_{53608}$ (in blue), *Lr*Asp2$_{100\text{-}23}$ (in orange) and *Lr*Asp2S349A$_{53608}$ (in grey), as well as with 20 μM paraoxonethyl with *Lr*Asp2$_{53608}$ (in yellow) and *Lr*Asp2$_{100\text{-}23}$ (light blue) monitored at OD$_{305nm}$. (B) Rate (μM indoxyl produced per min) of indoxylacetate hydrolysis by *Lr*Asp2$_{53608}$ (in blue), *Lr*Asp2$_{100\text{-}23}$ (in orange) with increasing μM indoxylacetate substrate concentrations. Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl and 25 μM final concentration of *Lr*Asp2. Mean rates were calculated from 3 replicates and SD error bars are displayed. (C) IC50 inhibitor kinetics with *Lr*Asp2$_{53608}$ and paraoxonethyl.

Table 4.2. Kinetic parameters of *Lr*Asp2$_{53608}$ and *Lr*Asp2$_{100\text{-}23}$ acetylesterase activity with indoxylacetate.

Enzyme	V_{max} (μmol indoxyl/min)	K_m (μm)	k_{cat} (μmol/min/μmol)
*Lr*Asp2$_{53608}$	1.88 ± 0.07	120 ± 9	3.13 ± 0.0019
*Lr*Asp2$_{100\text{-}23}$	1.98 ± 0.08	121 ± 2	3.25 ± 0.0004

To investigate the capacity of *Lr*Asp2$_{53608}$ and *Lr*Asp2$_{100\text{-}23}$ to recognise *Lr*SRRPs, binding of *Lr*Asp2$_{53608}$ and *Lr*Asp2$_{100\text{-}23}$ to immobilised targets was determined by ELISA following detection with an anti-his HRP fusion monoclonal antibody. Statistical inference by two-tailed

student t-test was used to determine differences between conditions. No binding could be determined between LrAsp2$_{53608}$ or LrAsp2$_{100-23}$ with native LrSRRPs from either strain (Figure 4.6). This suggested that LrAsp2 O-acetylation may require interactions with other aSec proteins, or that the fully glycosylated and secreted LrSRRP was not suitable as a binding partner due to all available sites being already acetylated.
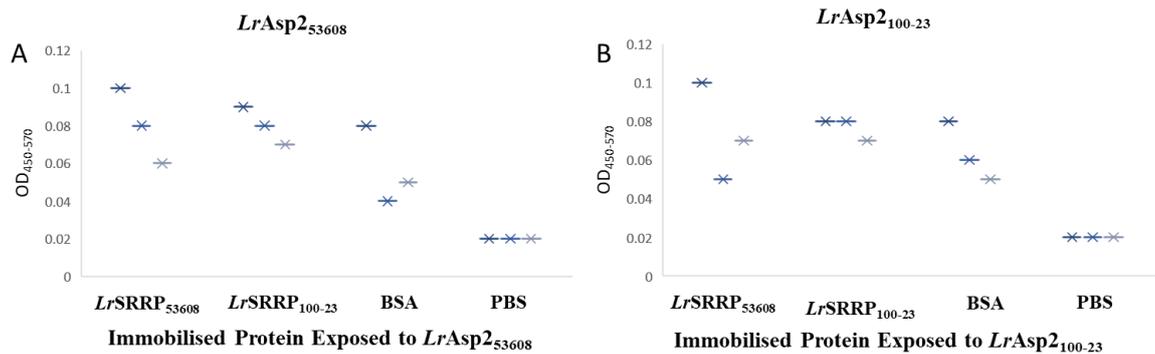
Figure 4.6. ELISA-based binding assay between LrAsp2$_{53608}$ and LrAsp2$_{100-23}$ and LrSRRPs. (A) LrAsp2$_{53608}$ at 15 μg/mL final concentration was added to well coated with LrSRRP$_{53608}$, LrSRRP$_{100-23}$, BSA and PBS. n=3. (B) LrAsp2$_{100-23}$ at 15 μg/mL final concentration was added to well coated with LrSRRP$_{53608}$, LrSRRP$_{100-23}$, BSA or PBS. Binding was monitored at OD$_{450-470\ nm}$ following TMB hydrolysis by anti-his-HRP monoclonal antibodies (n=3).

Attempts were made to crystallise LrAsp2$_{53608}$ and LrAsp2$_{100-23}$ using recombinant *L. reuteri* Asp2 proteins (Figure 4.1B), but whilst crystals were found for LrAsp2$_{53608}$ and LrAsp2$_{100-23}$, the quality of diffraction data was poor or resembled salt (Appendix 2, supplementary Figures S2.3 and S2.4). Instead, size exclusion small-angle scattering of X-rays (SEC SAXS) was utilised to determine the overall shape, size as well as electron density map of LrAsp2$_{53608}$ and help validate LrAsp2$_{53608}$ and LrAsp2$_{100-23}$ structural models constructed from sequence data using Alphafold2, which could then be fitted over a SAXS bead-model.

SEC-SAXS combines size-exclusion gel filtration with X-ray scattering of macromolecules, whereby the scattered beam as a function of the scattering angle can provide information of overall structure, shape, and quaternary structure *ab initio* (305). The size-exclusion step enables separation of contaminants or larger aggregates which could otherwise interfere with the downstream analysis. The size-exclusion trace of LrAsp2$_{53608}$ showed the presence of some

additional peaks which may have occurred due to X-ray exposure or following transportation and storage at the beamline (Figure 4.7A). The buffer-baseline was identified for subtraction and a peak between 360-372 frames was identified for downstream analysis, which also agreed with the expected run-time for the size of LrAsp2$_{53608}$ (58 kDa). The scattering profile and Guinier fit analysis of the peak area indicated a compact, folded globular cytosolic protein (Figure 4.7BC). The Kratky-Plot which can qualitatively assess the degree of unfolding in the sample indicated strong likelihood of a folded, soluble LrAsp2$_{53608}$, as shown by the gaussian bell shaped distribution (Figure 4.7D). The P(r) function provided information on the shape and size of LrAsp2$_{53608}$ by providing an estimation of the maximum dimension (Dmax) from which molecular weight estimates were derived (Table 4.3). A slight tail end was observed when approaching Dmax, suggesting that some aggregation, albeit minor, may have occurred (Figure 4.7E). Multiple size and shape calculations were taken into consideration with molecular weight from volume correlation providing an estimate of 58.9 kDa, in good agreement with the predicted molecular weight calculated from the LrAsp2$_{53608}$ amino acid sequence of 58 kDa (Table 4.3). A bead model of LrAsp2$_{53608}$ was also constructed using SAXS data and successfully fitted using an Alphafold2 generated model of LrAsp2$_{53608}$ (Figure 4.7F). The SAXS model confirmed that the recombinant LrAsp2$_{53608}$ was stable as a monomer and validated the Alphafold2 structural prediction.

An Alphafold2 structural model of LrAsp2$_{100-23}$ was also generated and superimposed onto the LrAsp2$_{53608}$ model (Figure 4.8A). Overall, there was a high structural homology between the two models with the proteins sharing 65% amino acid sequence identity. Both LrAsp2$_{53608}$ and LrAsp2$_{100-23}$ models displayed two Swiss-roll folds containing β-sheets stacked in alternating orientations. On initial inspection of the LrAsp2, the catalytic residues governing acetylesterase activity in SgAsp2$_{M99}$ could not be identified (Ser362, Glu452, and His482) (221). However, guided by a *Staphylococcus aureus* O-acetyltransferase OatA (*Sa*OatA) catalytic domain crystal structure (6VJP), an alternate set of catalytic residues was identified (Figure 4.8B) (303). These were Ser349, His470 and Asp472, which were found to be conserved between LrAsp2$_{53608}$ and

*Lr*Asp$_{100-23}$. The nucleophilic Ser349 pointed towards the deprotonating His470, which is consistent with the capacity of *Lr*Asp2 to act as an *O*-acetyltransferase and with the loss of acetylesterase activity of the *Lr*Asp2 S349A$_{53608}$ mutant (Figure 4.4, Figure 4.5).
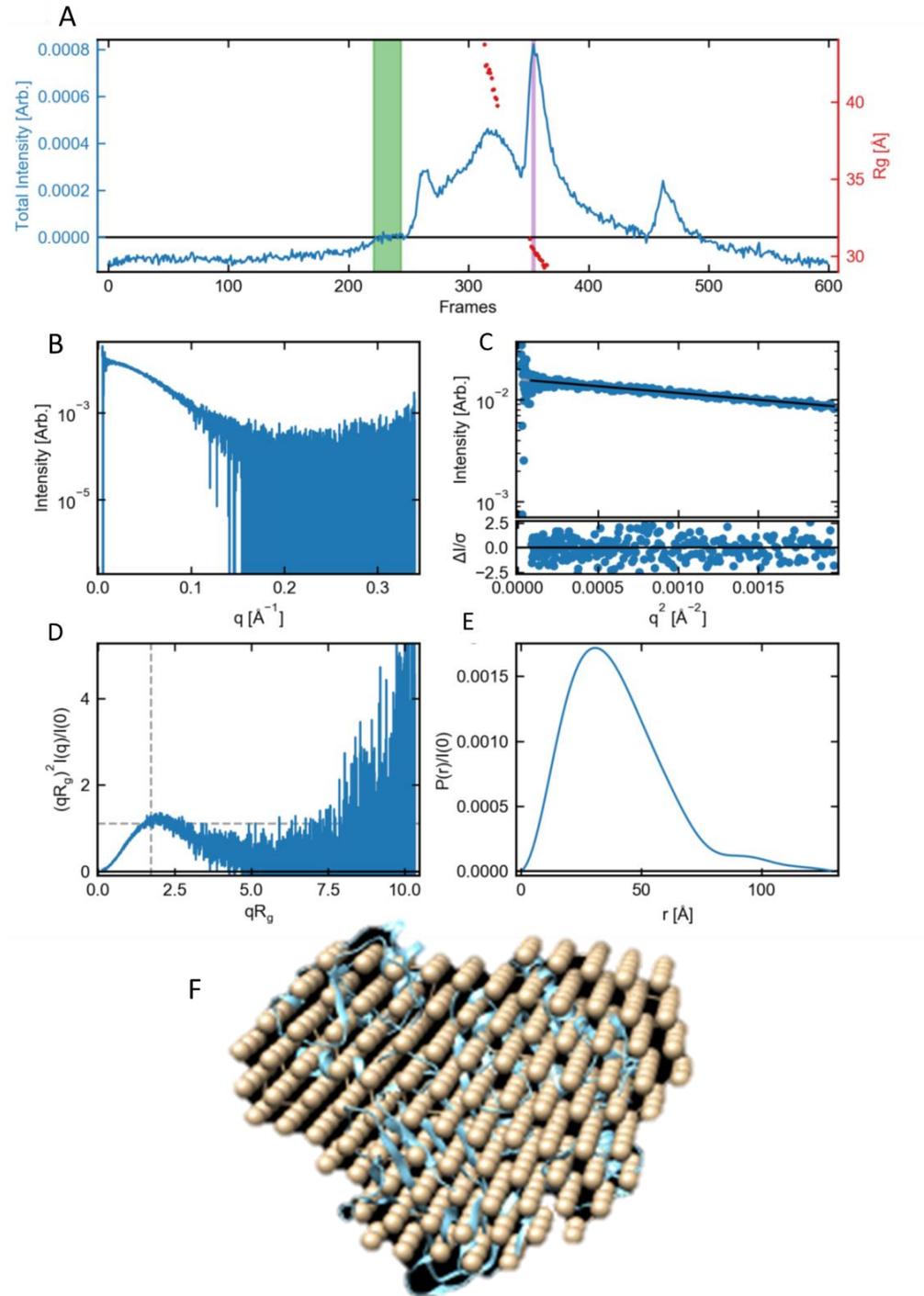
Figure 4.7. SAXS data summary for *Lr*Asp2$_{53608}$. (A) Series intensity (blue, left axis) vs. frame, and Rg vs. frame (red, right axis). Green shaded regions are buffer regions, purple shaded regions are sample regions. (B) Scattering profile(s) on a log-lin scale. (C) Guinier fit(s) (top) and fit residuals (bottom). (D) Normalized Kratky plot. Dashed lines show where a globular system would peak. (E) P(r) function(s), normalized by I(0). (F) Fitted bead-model of *Lr*Asp2$_{53608}$ using GNOM and Alphafold2 structure prediction of *Lr*Asp2$_{53608}$.

Table 4.3. SAXS-derived properties of *Lr*Asp2$_{53608}$.

Molecular weight and shape calculation	Molecular weight and shape estimate
Molecular weight (Vp)	71.4 kDa
Porod Volume	$8.61e^4\ A^3$
Molecular weight (Vc)	58.9 kDa
Molecular weight (S&S)	64.3 kDa
Shape (S&S)	Compact
Dmax (S&S)	110.1
Molecular weight. (Bayes)	65.5 kDa
Molecular weight (Bayes-Confidence Interval)	61.6 to 67.9 kDa
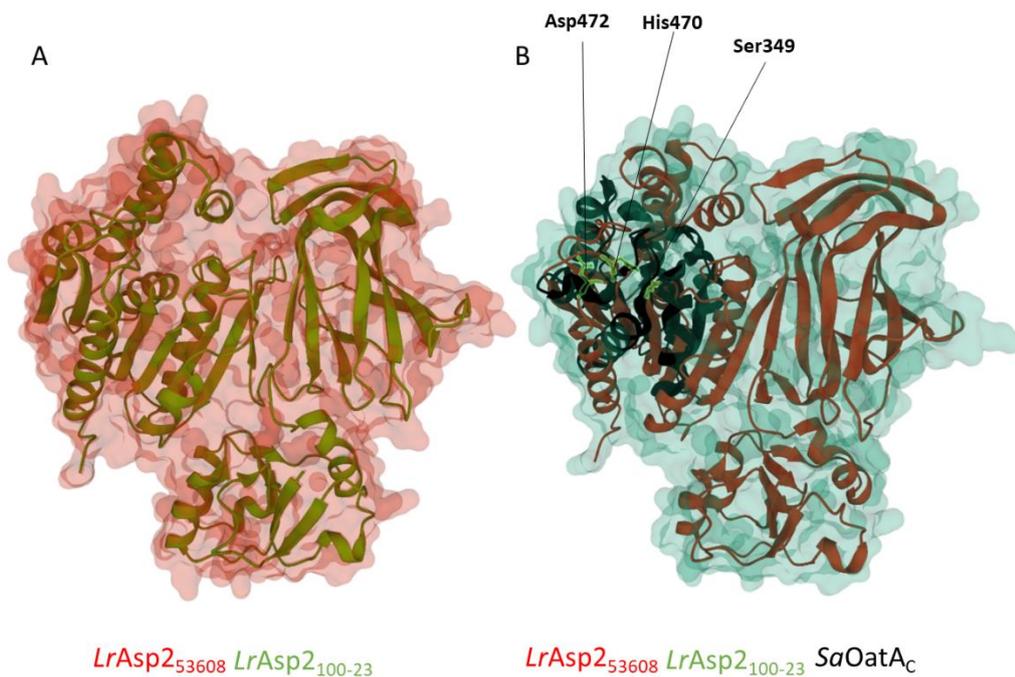
Figure 4.8. Alphafold2 structural models of *Lr*Asp2$_{53608}$ and *Lr*Asp2$_{100-23}$. (A) *Lr*Asp2$_{53608}$ (red) surface model and cartoon *Lr*Asp2$_{100-23}$ (green) superimposition. (B) Conserved catalytic triad (Asp472, His470, Ser349) in superimposed *Lr*Asp2$_{53608}$, *Lr*Asp2$_{100-23}$ and *Sa*OatA catalytic domain models (303).

4.2.4 Structure and function of *Lr*Asp3

When comparing the Alphafold2 generated model of *Lr*Asp3 with the partial crystal structures available for *Sg*Asp3$_{M99}$ (5VAE, black) and *Sp*Asp3$_{TIGR4}$(6LNW, blue), high structural homology of the β-sheet in the N-terminal domains was also apparent (216, 217) (Figure 4.9C). Tools to analyse protein topography and conserved motifs (TOPS, PROSITE) reported no structural analogues for comparison, nor identification of possible structure-function relationships (306,

307). However, the β-sandwich arrangement of *Lr*Asp3 is a common feature found in carbohydrate-binding modules (CBM) classified in the CAZy database ([www.cazy.org).](www.cazy.org) Particularly Type C exo-CBMs for binding termini of glycans on shorter sugar ligands of 1-3 monosaccharide units (e.g., reported in CBM9, CBM13, CBM32 families) best reflected the overall shape and features of *Lr*Asp3, as opposed to planar or endo-CBMs accommodating longer sugar chains (Type A and Type B) (308, 309). However, none of the available structural models for those families could be used to accurately identify the family or class *Lr*Asp3 belong to using superimposition or sequence analysis. Taken together, these data suggest that Asp3 is unique to aSec pathways with no structural analogues present in other secretion pathways or otherwise. It may also be that Asp3 functions by binding SRRP through protein-carbohydrate interactions only once part of the *Lr*Asp1-Asp2-Asp3 complex.
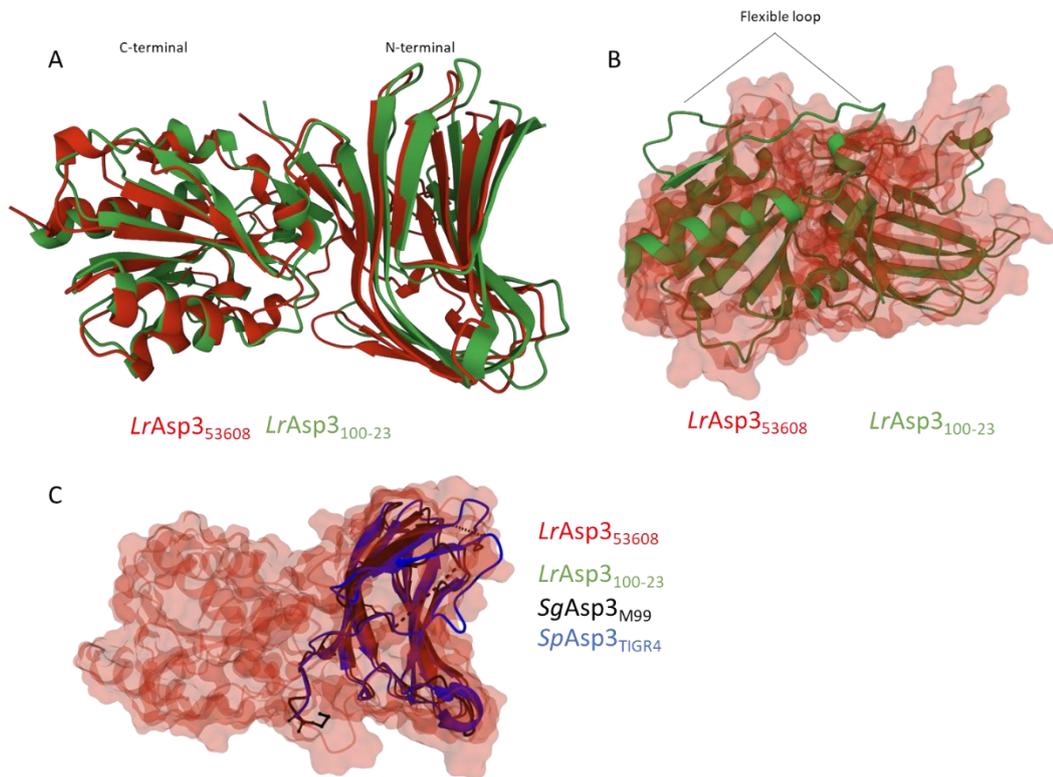
Figure 4.9. Alphafold2 structural models of LrAsp3$_{53608}$ and LrAsp3$_{100-23}$. (A) Front-view of cartoon LrAsp3$_{53608}$ superimposed with LrAsp3$_{100-23}$ structural model (B) Front-view of cartoon LrAsp3$_{100-23}$ structural model with LrAsp3$_{53608}$ surface model (C) Superimposition of surface model LrAsp3$_{53608}$ and LrAsp1$_{100-23}$ with partial SgAsp3$_{M99}$ (5VAE, black) and SpAsp3$_{TIGR4}$ (6LNW, blue) crystal structures (216, 219)

4.2.5 Structure and function of LrAsp1-Asp2-Asp3$_{53608}$

LrAsp1, LrAsp2 and LrAsp3 from *L. reuteri* ATCC 53608 were co-expressed in *E. coli* and purified as a 1:1:1 complex (Figure 4.1), as shown previously for SgAsp1-Asp2-Asp3$_{M99}$ and SpAsp1-Asp2-Asp3$_{TIGR4}$ (216, 217). The main predicted target of the LrAsp1-Asp2-Asp3$_{53608}$ complex is the native SRRP secretion cargo, so to test its possible chaperoning function, the LrAsp1-Asp2-Asp3$_{53608}$ complex was purified and assessed as a binding partner for native LrSRRPs *in vitro* using biolayer interferometry (BLI) and ELISA assays.

For the ELISA-binding assays, results showed a dose-dependent interaction between LrAsp1-Asp2-Asp3$_{53608}$ and the native LrSRRP$_{53608}$ with a two-fold increase in binding when LrAsp1-Asp2-Asp3$_{53608}$ was used at 10 µg/mL as compared to 5 µg/mL (Figure 4.10). The binding appeared to be specific to LrSRRP, as no signal was detected against BSA, or when PBS was

used as a negative control. A higher response was observed for LrAsp1-Asp2-Asp3$_{53608}$ as compared to LrAsp1$_{53608}$, suggesting that the formation of the complex increased affinity of LrAsp1$_{53608}$ or that multiple interactions with other components, such as LrAsp3$_{53608}$, may result in increased binding.

BLI was then used to calculate binding kinetic data and avidity constants. Amine-reactive biosensors (ARG2) allowed for amine-coupling of the native LrSRRP$_{53608}$ and LrSRRP$_{100-23}$ to the biosensor, while LrAsp1-Asp2-Asp3$_{53608}$ was used as the analyte. A baseline without analyte was used as a negative control and as a reference sensor to subtract shifts resulting in buffer response. There was an increased response following increasing LrAsp1-Asp2-Asp3$_{53608}$ concentrations which allowed for fitting the association and dissociation kinetics with LrSRRP$_{53608}$ (Figure. 4.11A-E). The same was repeated with LrSRRP$_{100-23}$ immobilised to the biosensor (Figure 4.11F). The binding responses followed a similar pattern for LrSRRP$_{100-23}$ and for LrSRRP$_{53608}$. Steady-state kinetic data showed a high-avidity interaction between LrAsp1-Asp2-Asp3$_{53608}$ and both LrSRRP$_{53608}$ and LrSRPP$_{100-23}$ with a final Kd of $2.3\ x\ 10^{-7}$ and $2.2\ x\ 10^{-7}$ M, respectively (Table 4.4).
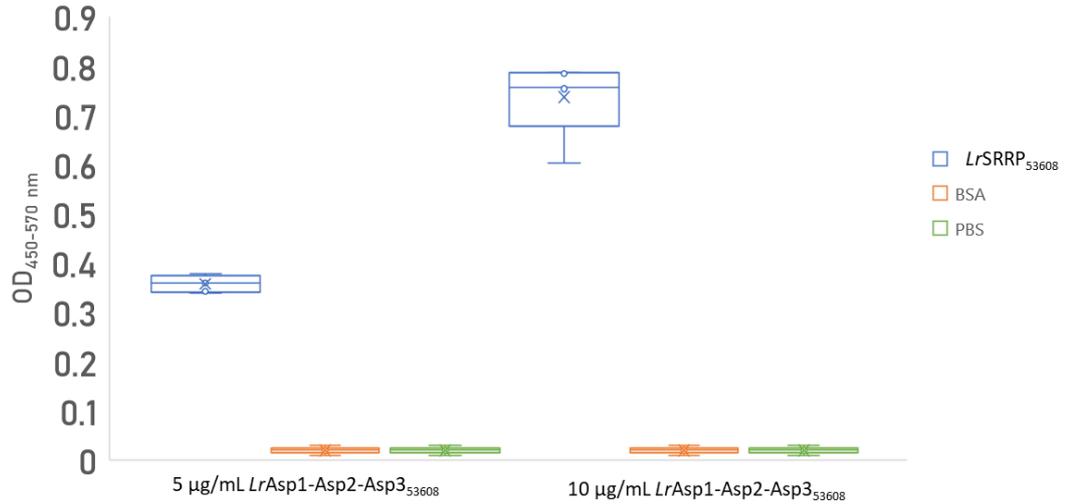
Figure 4.10. ELISA-based binding assay between *Lr*Asp1-Asp2-Asp3$_{53608}$ and *Lr*SRRP$_{53608}$. *Lr*Asp1-Asp2-Asp3$_{53608}$ at 5 and 10 μg/mL final concentration was added to well coated with *Lr*SRRP$_{53608}$, BSA and PBS. Binding was determined following incubation of anti-his-HRP monoclonal antibodies and TMB substrate by measurement at OD$_{450\text{-}570nm}$ (n=8).
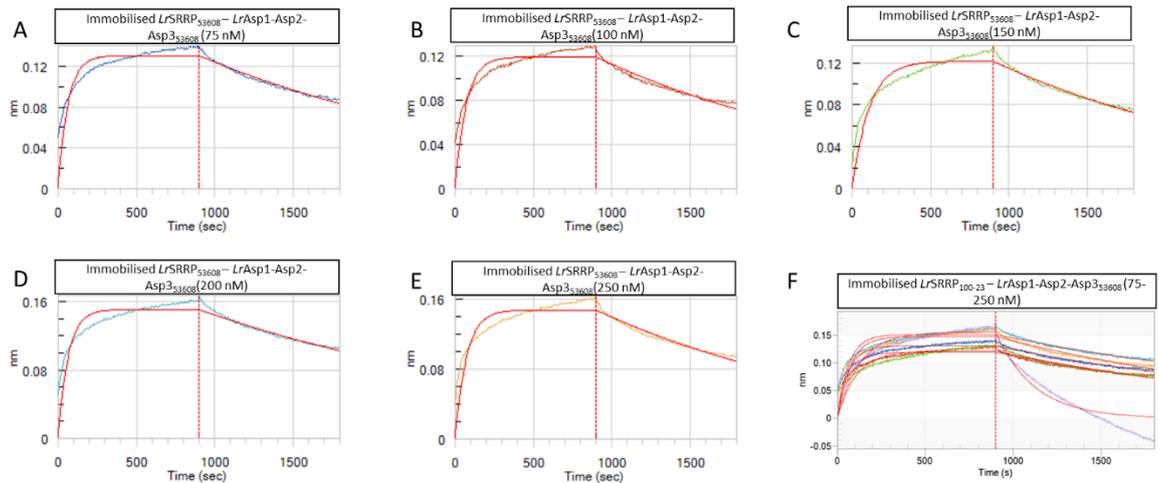
Figure 4.11. Biolayer interferometry (BLI) binding between *Lr*Asp1-Asp2-Asp3$_{53608}$ complex and native *Lr*SRRPs. (A) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*Asp1-Asp2-Asp3$_{53608}$ (final concentration 75 nM) association and dissociation fitting. (B) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*Asp1-Asp2-Asp3$_{53608}$ (final concentration 100 nM) association and dissociation fitting. (C) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*Asp1-Asp2-Asp3$_{53608}$ (final concentration 150 nM) association and dissociation fitting. (D) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*Asp1-Asp2-Asp3$_{53608}$ (final concentration 200 nM) association and dissociation fitting. (E) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*Asp1-Asp2-Asp3$_{53608}$ (final concentration 250 nM) association and dissociation fitting. (F) Immobilised *Lr*SRRP$_{100\text{-}23}$ exposed to *Lr*Asp1-Asp2-Asp3$_{53608}$ (final concentration gradient 75 – 250 nM) association and dissociation fitting. (n=3)

Table 4.4 Kinetic parameters of *Lr*Asp1-Asp2-Asp3$_{53608}$ and *Lr*SRRP binding by BLI.

Immobilised protein	Analyte protein	Kinetic parameters		
		R^2	Rmax	Kd (M)
*Lr*SRRP$_{53608}$		0.84	0.174 ± 0.01	$2.3 \ x \ 10^{-7} \pm$ $9.1 \ x \ 10^{-9}$
*Lr*SRRP$_{100\text{-}23}$	*Lr*Asp1-Asp2-Asp3$_{53608}$	0.79	0.172 ± 0.03	$2.2 \ x \ 10^{-7} \pm$ $1.1 \ x \ 10^{-8}$

To investigate the potential effect of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex on *Lr*Asp2$_{53608}$ acetylesterase activity, kinetic parameters were determined using *p*NP-acetate as a substrate. A velocity plot was derived as a function of substrate concentration allowing for calculation of kinetic parameters (Figure 4.12, Table 4.5). When accounting for the final stoichiometry of *Lr*Asp2$_{53608}$ that is part of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex, the specific activity was found to be 10-fold higher in complex than as a monomer (12.7 ± 0.008 to 1.26 ± 0.0009 respectively). *Lr*Asp1$_{53608}$ and *Lr*Asp3$_{53608}$ were also tested for activity, but no *p*NP-acetate hydrolysis could be detected (not shown). Taken together these data suggest that the formation of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex increases *Lr*Asp2$_{53608}$ acetylesterase activity, which may be due to increased stability or increased binding to the substrate.
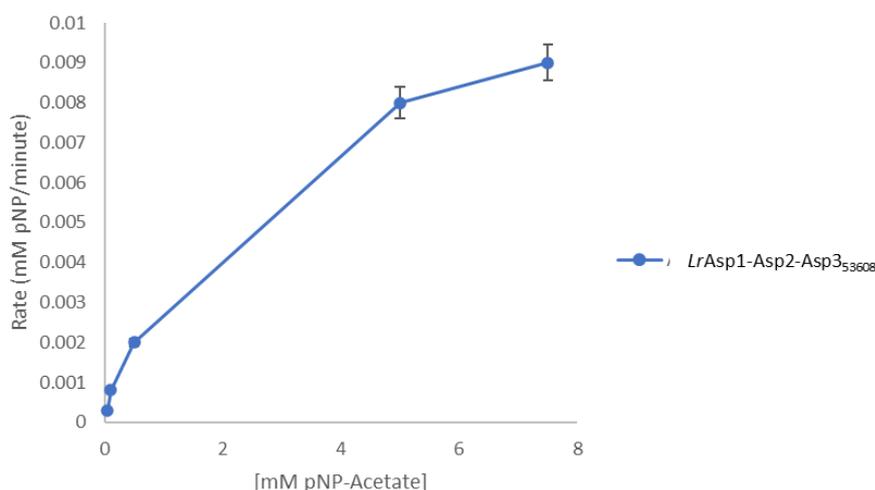
Figure 4.12. *Lr*Asp1-Asp2-Asp3$_{53608}$ acetylesterase activity with *p*NP-acetate. Rate (mM *p*NP produced per min) of *p*NP-acetate hydrolysis by *Lr*Asp1-Asp2-Asp3$_{53608}$ (in blue), with increasing concentration of *p*NP-acetate substrate. Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl. Mean rates were derived from 5 replicates and SD error bars are displayed.

Table 4.5. Kinetic parameters of *Lr*Asp1-Asp2-Asp3 acetylesterase activity with *p*NP-acetate as a substrate.

Protein (s)	Kinetic parameters					
	*p*NP-acetate			indoxylacetate		
	V_{max} (µmol *p*NP/min)	K_m (µm)	k_{cat} (µmol/min/µmol)	V_{max} (µmol *p*NP/min)	K_m (µm)	k_{cat} (µmol/min/µmol)
*Lr*Asp2$_{53608}$	7.7 ± 0.04	162 ± 15	1.26 ± 0.0009	1.88 ± 0.07	120 ± 9	3.13 ± 0.0019
*Lr*Asp2$_{100-23}$	7.8 ± 0.04	160 ± 11	1.3 ± 0.0007	1.98 ± 0.08	121 ± 2	3.25 ± 0.0004
*Lr*Asp1-Asp2-Asp3$_{53608}$	9.12 ± 0.09	362 ± 35	12.7 ± 0.008	ND	ND	ND
*Sp*Asp1-Asp2-Asp3$_{TIGR4}$	4.79 (216)	957.61 (216)	ND	ND	ND	ND

Despite several attempts, crystals that formed of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex did not diffract (Appendix 2, supplementary Figure S2.5). SEC-SAXS was performed, revealing a size-exclusion trace with a split peak, suggesting some degradation of *Lr*Asp1-Asp2-Asp3$_{53608}$ over exposure, during transportation and storage to the beamline, or the presence of different *Lr*Asp1-Asp2-Asp3$_{53608}$ oligomerisations (Figure 4.13A). Minor adjustments were made to the buffer baseline and the peak between 320-330 frames was identified for downstream analysis, also corresponding to the expected run-time for size of *Lr*Asp1-Asp2-Asp3$_{53608}$ (150 kDa). Scattering profile and Guinier fit analysis of the peak area were indicative of a compact, folded globular cytosolic protein (Figure 4.13BC). The Kratky-Plot which can qualitatively assess the degree of unfolding in the sample reported the strong likelihood of a folded, soluble *Lr*Asp1-Asp2-Asp3$_{53608}$, as shown by the gaussian bell shaped distribution (Figure 4.13D). The P(r) function provided information about the shape and size of *Lr*Asp1-Asp2-Asp3$_{53608}$, estimating the maximum dimension (Dmax) from which molecular weight estimates were determined. A slight tail end was observed when approaching Dmax, suggesting that some aggregation, albeit minor, may have occurred (Figure 4.13E). Multiple size and shape calculations were taken into consideration with the Bayes estimation predicting the *Lr*Asp1-Asp2-Asp3$_{53608}$ to be between

127.5 to 151.4 kDa (Table 4.6), in agreement with the calculated 152 kDa based on the amino acid composition. A bead model of LrAsp1-Asp2-Asp3$_{53608}$ was also constructed using SAXS data, showing electron-density maps likely corresponding to one LrAsp1$_{53608}$ monomer, one LrAsp2$_{53608}$ monomer and one LrAsp3$_{53608}$ monomer. This provided guidance to build the LrAsp1-Asp2-Asp3$_{53608}$ complex using the Alphafold2 generated models of the individual component proteins (Figure 4.13F).

The LrAsp1-Asp2-Asp3$_{53608}$ model was further analysed to investigate how the different subunits may interact with each other (Figure 4.14A). It was apparent that all three Asps have a possible interface with one another. Of particular interest was that the LrAsp2$_{53608}$ catalytic triad (Ser349, His470, Asp472) was within 14.5 Å of a flexible loop in LrAsp1$_{53608}$, suggesting that LrAsp1$_{53608}$ may facilitate the binding and orientation of the glycosylated SRRP for O-acetylation by LrAsp2$_{53608}$ (Figure 4.14B).
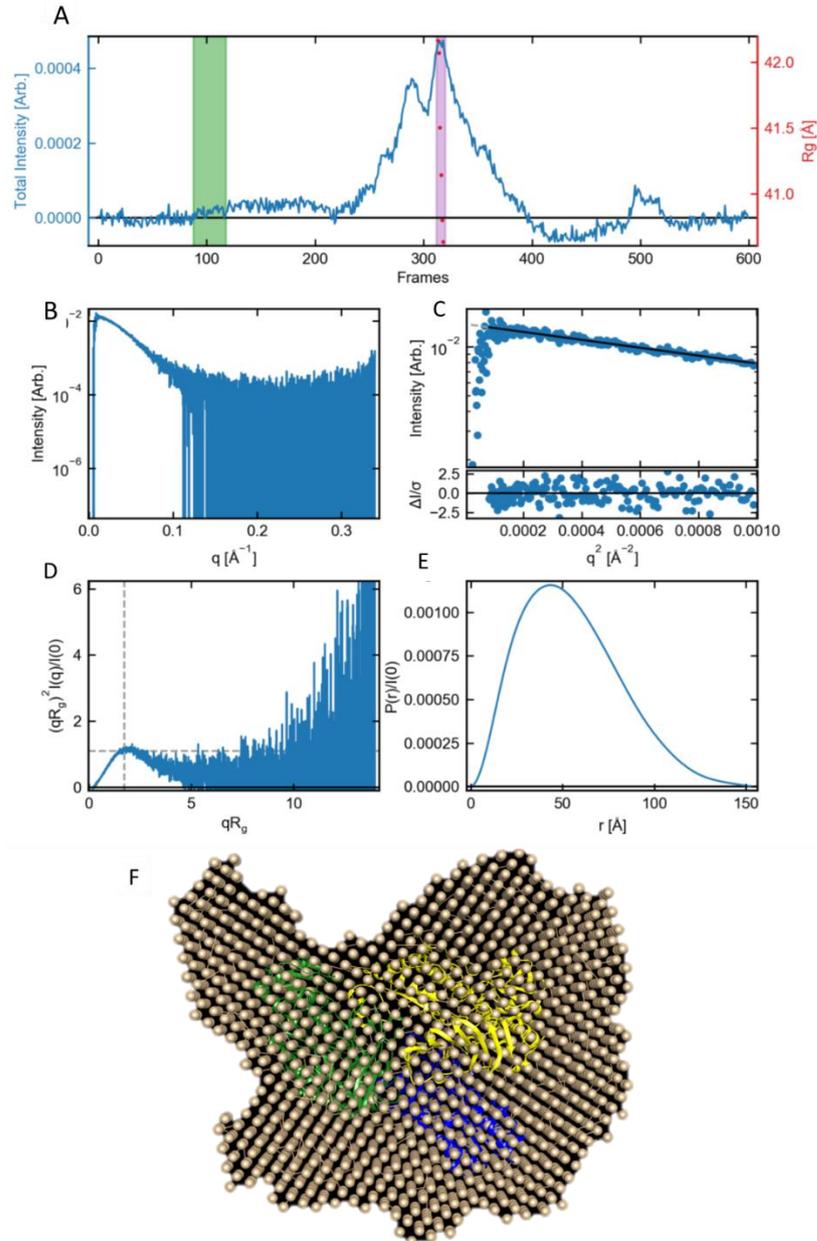
Figure 4.13. SAXS data summary for *Lr*Asp1-Asp2-Asp3$_{53608}$ (A) Series intensity (blue, left axis) vs. frame, and Rg vs. frame (red, right axis). Green shaded regions are buffer regions, purple shaded regions are sample regions. (B) Scattering profile(s) on a log-lin scale. (C) Guinier fit(s) (top) and fit residuals (bottom). (D) Normalized Kratky plot. Dashed lines show where a globular system would peak. (E) P(r) function(s), normalized by I(0). (F) Fitted bead-model of Asp2$_{53608}$ using GNOM and Alphafold2 structure prediction of *Lr*Asp1-Asp2-Asp3$_{53608}$ (*Lr*Asp1$_{53608}$ in yellow, *Lr*Asp2$_{53608}$ in green, *Lr*Asp3$_{53608}$ in blue)

Table 4.6. SAXS-derived properties of *Lr*Asp1-Asp2-Asp3$_{53608}$.

Molecular weight and shape calculation	Molecular weight and shape estimate
Molecular weight (Vp)	158.4 kDa
Porod Volume [A^3]	$1.91e^5 A^3$
Molecular weight (Vc)	131.5 kDa
Molecular weight (S&S)	134 kDa
Shape (S&S)	Compact
Dmax (S&S)	148.8
Molecular weight (Bayes)	138.2 kDa
Molecular weight (Bayes-Confidence Interval)	127.5 to 151.4 kDa
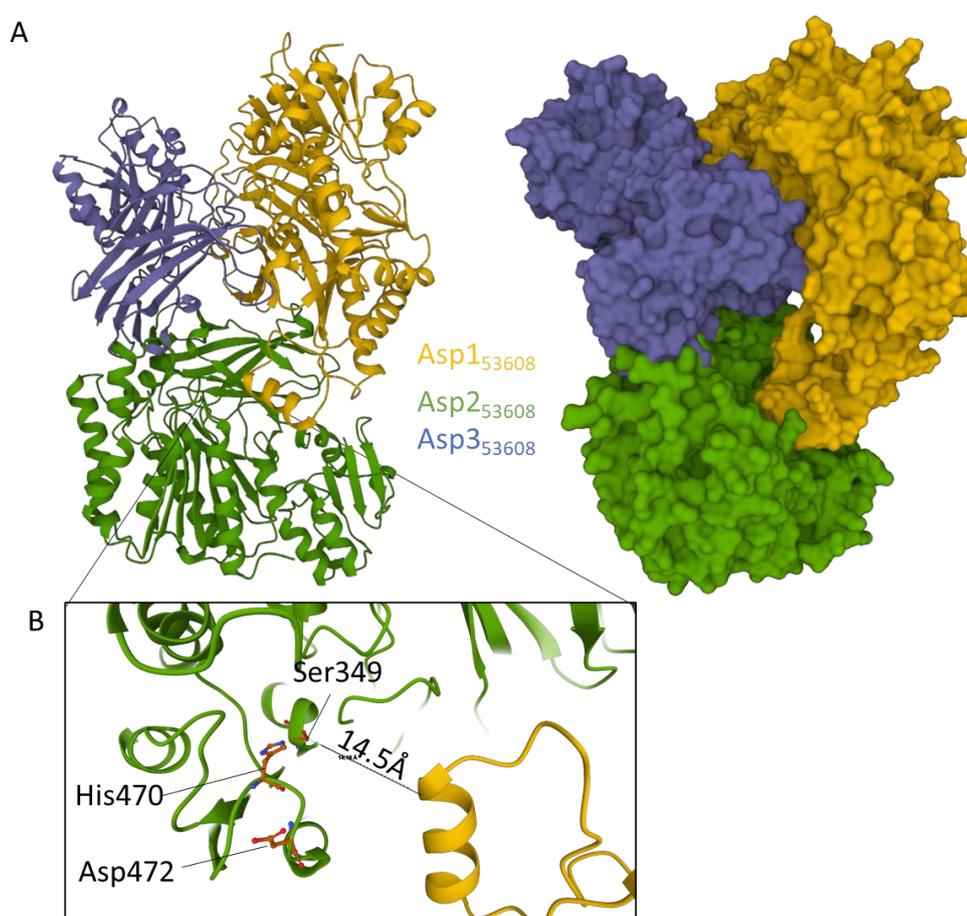
Figure 4.14. Alphafold2 structural model of *Lr*Asp1-Asp2-Asp3$_{53608}$ complex. (A) Cartoon and molecular surface representation of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex (*Lr*Asp1$_{53608}$ in yellow, *Lr*Asp2$_{53608}$ in green and *Lr*Asp3$_{53608}$ in blue) (B) Distance between *Lr*Asp2$_{53608}$ catalytic residues and *Lr*Asp1$_{53608}$ flexible C-terminal loop.

4.3 Discussion

Cytosolic processing of the secretion cargo by chaperoning and post-translational modifications are common features of secretion pathways including the aSec system (185, 263, 300). Here, biochemical, and biophysical approaches, including machine learning modelling (Alphafold2) and SAXS, were used to investigate the structure and function of Asps involved in the *L. reuteri* aSec pathway.

The oligomeric state of LrAsp1$_{53608}$, LrAsp2$_{53608}$ and LrAsp3$_{53608}$ was shown to be a 1:1:1 LrAsp1:Asp2:Asp3$_{53608}$ complex by native-PAGE and SEC-SAXS. SEC-SAXS provides three-dimensional low resolution structures *ab initio* and is a useful tool when crystallisation of proteins is not possible due to protein flexibility, stability or unusual oligomerisations (305). Alphafold2 has shown unprecedented levels of accuracy in modelling single chain protein structures and is able to compute large multimeric protein complexes (310-312). As a result, it offers the most reliable methodology in terms of *in silico* protein structural prediction despite bias towards protein conformations that are prone to crystallisation (as a result of relying on PDB as a training database) (313). Here, Alphafold2 was used to generate models of LrAsp1, LrAsp2 and LrAsp3 from *L. reuteri* ATCC53608 and 100-23 strains. Due to increasing complexity of multimers and reduced representation in structural databases compared to monomers, experimental approaches like SAXS can help validate protein complex structural predictions and provide statistical confirmation of likelihood of fit. Here, the resulting bead model enabled virtual assembly of the LrAsp1-Asp2-Asp3$_{53608}$ complex using Alphafold2 generated models of the individual components. As the structural homology between Asps of *L. reuteri* ATCC 53608 and 100-23 strains are very high, a similar oligomerisation is anticipated for LrAsp1-Asp2-Asp3$_{100-23}$ strains. Despite incomplete resolution of Asp2 and Asp3 in recombinant Asp1-Asp2-Asp3 crystal structures from *S. gordonii* M99 and *S. pneumoniae* TIGR4 strains, a similar oligomerisation and orientation of the individual Asps was apparent (216, 217).

Here, the structural modelling of LrAsp2 and LrAsp3 from *L. reuteri* ATCC 53608 and 100-23 strains helped piece together missing structural data on C-terminal regions of Asp2 and Asp3 in

gram-positive bacteria. For example, we predicted that the C-terminal catalytic triad of *L. reuteri* Asp2 utilises Ser, His, Asp which was shared with *S. aureus* acetyltransferase OatA (303). Thus, a similar reaction mechanism can be proposed whereby the carboxyl group of Asp472 forms a salt bridge with a nitrogen atom in the imidazole ring of His470. In turn, this would enable His470 to deprotonate Ser349, creating a nucleophilic attack on the carbonyl carbon of the acetyl donor molecule. In this scenario, local residues at the active site would form an oxyanion hole that stabilises the transition state, which will then collapse into a covalently bound acetyl-enzyme intermediate. In SgAsp2$_{M99}$, sequence alignment suggests the presence of a Glu instead of Asp in LrAsp2, but which may play similar roles. The significance of these differences in catalytic residue composition across gram-positive species may reflect specificities towards the target SRRP as opposed to acetyl-donor substrates, since the diversity of potential acetyl donors has been reported to be quite low with acetyl phosphate being suggested as a high-energy donor across many bacterial systems (314, 315).

The acetylesterase activity of LrAsp2 was characterised as a monomer and as part of the LrAsp1-Asp2-Asp3$_{53608}$ complex. Both pNP-acetate and indoxylacetate were found to be acetyl-donor substrates for LrAsp2$_{53608}$ *in vitro*. LrAsp2$_{53608}$ showed highest enzymatic rate when part of the LrAsp1-Asp2-Asp3$_{53608}$ complex using pNP-acetate as a substrate (Table 4.5). This 10-fold difference in activity between LrAsp2$_{53608}$ and LrAsp1-Asp2-Asp3$_{53608}$ suggested that formation of the complex is important for the enzyme activity. Comparable constants have been derived for SpAsp1- Asp2- Asp3$_{TIGR4}$, but no kinetic parameters are available for SgAsp1-Asp2-Asp 3$_{M99}$ (216, 217).

Paraoxonethyl was identified as an inhibitor of LrAsp2$_{53608}$ acetylesterase activity. Paraoxonethyl is commonly used as one of the most potent pesticides available and acts as a acetylcholinesterase inhibitor, thus being a nerve agent (304). It targets the active site where it irreversibly binds to the nucleophilic Ser needed for hydrolysis (304). While it is not suitable for drug use in the context of inhibiting pathogenic Asp2s due to its high toxicity, the indoxylacetate based assay reported

here for *Lr*Asp2 may be used for screening less toxic analogues of paraoxonethyl that may absorb at the same wavelength as *p*NP.

In contrast to *Lr*Asp2, *Lr*Asp1 and *Lr*Asp3 did not show enzymatic activity on these substrates in the conditions tested. This was reflected by the predicted structures of *Lr*Asp1 and *Lr*Asp3 from ATCC 53608 and 100-23 strains, whereby *Lr*Asp1 formed the U-shaped canonical form of GT-B type glycosyltransferases but lacked any of the catalytic residues involved in the glycosylation reaction, as also reported for *Lr*GtfB (276). A recognition function could be confirmed for *Lr*Asp1$_{53608}$ as it bound the native glycosylated *Lr*SRRPs of both *L. reuteri* ATCC 53608 and 100-23 strains equally in binding assays in contrast to monomeric *Lr*Asp2$_{53608}$ and *Lr*Asp2$_{100\text{-}23}$ which were unable to bind *Lr*SRRPs *in vitro*. Recombinant *Lr*Asp3$_{53608}$ appeared stable only when part of *Lr*Asp1-Asp 2-Asp 3$_{53608}$, as also reported for recombinant *Sg*Asp3$_{53608,}$ and thus could only be biochemically characterised as part of the complex (217).

For *Sg*Asp1$_{M99}$, *Sg*Asp2 $_{M99}$ and *Sg*Asp3 $_{M99}$, site-directed mutagenesis of key residues (such as catalytic Ser362 of *Sg*Asp2$_{M99}$) and gene insertion deletions have been shown to drastically affect *Sg*GspB$_{M99}$ secretion and glycosylation, which is consistent with the capacity of Asps to interact with SRRP, but no binding kinetics of aSec components are available to date across gram-positive bacteria (217). Here, the binding kinetics for *Lr*Asp1-Asp2-Asp3$_{53608}$ were determined, showing nanomolar-avidity towards *Lr*SRRPs irrespective of the strain origin. Avidity was preferred to affinity, which is specific to a single binding site, because it is unclear which parts of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex bound *Lr*SRRP and multiple interactions are likely to occur. Protein-carbohydrate interactions may be mediated by GT-like *Lr*Asp1$_{53608}$ or by the potential CBM in *Lr*Asp3$_{53608}$. However, it is possible that protein-protein interactions may also occur to facilitate high avidity interactions, considering cross-strain binding of *Lr*SRRPs *in vitro* in spite of strain-specific glycosylation (235, 316). In type III secretion systems, the chaperone FlgN in *S. typhimurum* MM9001 has been shown to have nanomolar affinity for substrates FlgK and FlgL to prevent aggregation and proteolysis of the secretion cargo (317). However, *Lr*Asp1-Asp2-Asp3$_{53608}$ also performs *O*-acetylation in addition to a chaperoning role, so it is possible that

conformational changes upon acetylation allow for decoupling of the LrAsp1-Asp2-Asp3 complex from the LrSRRP to facilitate further post-translational modification of the SRRs.

In summary, our biochemical and structural analysis of the recombinant LrAsps revealed acetylesterase activity of LrAsp2 as a monomer and as part of the LrAsp1-Asp2-Asp3$_{53608}$ complex and binding interactions of LrAsp1 as a monomer and as part of the LrAsp1-Asp2-Asp3$_{53608}$ complex with LrSRRPs. Formation of the LrAsp1-Asp2-Asp3$_{53608}$ complex demonstrated the best functional capacity *in vitro*, suggesting its probable native oligomeric state. Insights into the Alphafold2 structure of LrAsp1-Asp2-Asp3$_{53608}$ constructed using SAXS model confirmed a 1:1:1 oligomerisation and provided novel information on LrAsp2$_{53608}$ and LrAsp3$_{53608}$ C-terminal domains, as well as proximity of the LrAsp2$_{53608}$ catalytic site and LrAsp1$_{53608}$ mobile loop structure close to the binding cleft. A new substrate and inhibitor have been tested for LrAsp2$_{53608}$. Furthermore, avidity binding kinetics derived for the secretion complex with the native substrate irrespective of strain-specific glycosylation of SRRPs.

Future work on LrAsp1-Asp2-Asp3 is required to determine specific sites of interaction between aSec components and LrSRRPs. Nuclear magnetic resonance (NMR) based techniques can be used to monitor the transfer of the acetyl group from donor to O-glycan and to determine binding partners or epitopes in complement of the thermal shift and enzymatic assays used in this work. NMR could also help explore structurally which regions of LrAsp1 and LrAsp3 are binding LrSRRPs. Such information is needed to advance our understanding of the cytosolic steps occurring to O-acetylate as well as chaperone LrSRRP to the cell membrane. Additionally, due to the conserved nature of the Asp1-Asp2-Asp3 complex across gram-positive bacteria, this research could provide targets of pharmacological interest for therapeutics targeting aSec systems of pathogenic bacteria.

Chapter 5 Structure and function of aSec translocation machinery proteins of *L. reuteri* strains.

5.1 Introduction

The aSec translocation machinery at the cell membrane is predicted to be comprised of SecA2, SecY2, Asp4, Asp5, due to sequence homology with its Sec pathway paralogues SecA, SecY, SecE and SecG (182) (see section 1.3.1) . Here, SecA2 is the motor-ATPase providing energy for active transport and SecY2 plays the role of a transmembrane protein forming the pore for the SRRP to traverse the cell membrane (185, 318). The roles of Asp4 and Asp5 are poorly understood but, based on their similarity to SecE and SecG, they are likely membrane proteins with a high percentage of exposed hydrophobic residues involved in assembly of the secretion complex as well as insertion into the cell membrane.

Out of the components of the aSec translocation machinery in pathogenic gram-positive bacteria, SecA2 has been characterised the most. It is a cytoplasmic enzyme but has also been shown to interact with membrane lipids in *S. gordonii* M99 in order to localise with the other aSec proteins (SecY2, Asp4) (189). At a biochemical level, the ATPase activity of recombinant SecA2 proteins from *S. gordonii* M99 (SgSecA2$_{M99}$), *S. gordonii* DL1 (SgSecA2$_{DL1}$), *C. difficile* 630 (CdSecA2$_{630}$) and *M. tuberculosis* H37Rv (MtSecA2$_{H37Rv}$) have been reported (188, 189, 318, 319). Monomeric 3D-crystal structure of CdSecA2$_{630}$ and MtSecA2$_{H37Rv}$ have been solved by X-ray crystallography, however the native structural form of aSec SecA2 remains controversial. In the Sec system, both monomeric and dimeric forms of SecA have been reported from various bacterial species including *E. coli* K12 (EcSecA$_{K12}$) and *T. thermophilus* HB8 (TtSecA$_{HB8}$) (320, 321). In other secretion systems, structural data for SecA2 are limited to SecA1/SecA2-dependent systems, which have different secretory targets and utilise the SecY of the Sec system, as opposed to SRRP and SecY2 of the aSec system (137, 188). SgSecA2$_{M99}$ has also been shown to interact with cytosolic proteins of the aSec system such as Asp2 and Asp3 (322). It has also been reported that SecA2 is important for aSec mediated transport of SRRPs, whereby deletion mutations of SecA2 ked to loss of biofilm formation in pathogens (*S. pneumoniae* TIGR4, *L. monocytogenes* EGD-e) and gut symbiont *L. reuteri* 100-23C alike (207, 323, 324).

SecY2, Asp4 and Asp5 have been expressed recombinantly in *S. gordonii* DL1 where they have been shown to form a complex (318). So far, all structural properties of the SecY2-Asp4-Asp5 complex have been derived from homology modelling and bioinformatics analysis based on existing crystal structures of the SecYEG complex (127). At a biochemical level, it been shown that the ATPase activity of recombinant SgSecA2$_{DL1}$ increases four-fold when part of SgSecY2-Asp4-Asp5$_{DL1}$ complex (318). Interestingly, many *L. reuteri* strains including ATCC 53608 and 100-23 do not encode Asp5, suggesting a functional redundancy in these auxiliary proteins, or differences in the *L. reuteri* membrane translocation machinery assembly.

This chapter reports the biochemical and structural analysis of SecA2, SecY2 and Asp4 from *L. reuteri* ATCC 53608 and 100-23 strains as individual components and in complex.

5.2 Results
5.2.1 Production and analysis of recombinant translocation proteins from *L. reuteri* aSec system

For the biochemical and structural characterisation of SecA2, SecY2 and Asp4 encoded by the *L. reuteri* aSec operon, recombinant proteins from ATCC 53608 and 100-23 strains were produced heterologously in *E. coli* using the plasmids containing N-terminal his-tag fusions for LrSecA2$_{53608}$, LrSecY2$_{53608}$, LrSecA2$_{100-23}$ in pET28b, whereas the LrSecA2-SecY2-Asp4$_{53608}$ complex was encoded by two plasmids with LrSecA2-SecY2$_{53608}$ cloned into pETDUET-1 and LrAsp4$_{53608}$ cloned in pET15b for co-expression. In addition, site-directed mutagenesis was performed to generate LrSecA2K114A$_{53608}$ using the SPRIP protocol (see 2.3 for experimental details) (237).

Recombinant aSec secretion machinery proteins were purified by nickel affinity chromatography (NiNTA) as his-tagged proteins. In the case of WT and mutant LrSecA2 proteins, the his-tags were also removed following thrombin cleavage for X-ray crystallography. Removal of the purification tag could not be performed with LrSecY2$_{53608}$ due to incompatible cleavage sites. The apparent size of the purified recombinant proteins was shown by SDS-PAGE to be approx. 90

kDa for LrSecY2$_{53608}$, LrSecA2$_{53608}$ and LrSecA2K114A$_{53608}$, and 43 kDa for LrSecY2$_{53608}$, respectively (Figure 5.1AB).
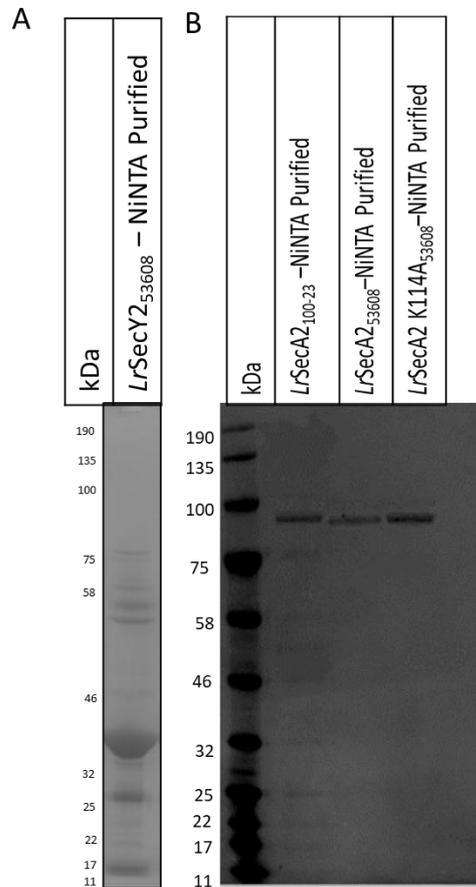
Figure 5.1 Gel electrophoresis of purified recombinant WT and mutant aSec translocation machinery proteins from *L. reuteri* ATCC 53608 and 100-23. (A) Recombinant LrSecY2$_{53608}$ purified by NiNTA and analysed by SDS-PAGE. (B) Recombinant LrSecA2$_{53608}$, LrSecA2$_{100-23,}$ LrSecA2 K114A$_{53608}$ purified by NiNTA followed by his-tag removal by thrombin cleavage and analysed by SDS-PAGE. Gels were stained with Coomassie.

To purify the recombinant LrSecA2-SecY2-Asp4$_{53608}$ complex, the use of 3 mM CHAPS detergent during the lysis step was essential to yield a soluble complex. This may be due to the zwitterionic detergent solubilising the exposed hydrophobic surfaces of LrSecY2$_{53608}$ and LrAsp4$_{53608}$, allowing for the whole complex to remain in solution for further purification steps. The recombinant complex was purified by NiNTA as a his-tagged protein followed by size-exclusion gel filtration (SEGF) (Figure 5.2AB). The elution at 59 mL using the Superdex200 column was consistent with a protein of apparent size between 158 and 440 kDa (Figure 5.2A).

This fraction was further analysed by SDS-PAGE, showing bands corresponding to the expected sizes for LrSecA2$_{53608}$, (90 kDa) and LrSecY2$_{53608}$ (45 kDa), based on their amino acid sequences (Figure 5.2B). Some aberrant gel migration or degradation was apparent, highlighted by the presence of two close bands for each predicted protein component. LrAsp4$_{53608}$ had an expected size of 10 kDa, from the amino acid sequence, so it may be that the band just above the 15 kDa marker represents a dimerised form of LrAsp4$_{53608}$. The fraction was also analysed by native-PAGE which showed a band just above the 242 kDa marker, which could suggest the following composition based on the theoretical mass of the individual proteins: dimeric LrSecA2$_{53608}$ (180 kDa), monomeric LrSecY2$_{53608}$ (45 kDa) and dimeric LrAsp4$_{53608}$ (20 kDa), resulting in a 245kDa complex (Figure 5.2C).

The yield of recombinant LrSecA2 proteins was around 6 mg/L culture while membrane associated LrSecY2$_{53608}$ was around 3 mg/L and the LrSecA2-SecY2-Asp4$_{53608}$ was further reduced to approx.0.8 mg/L (Appendix 3, supplementary Table S3.1).
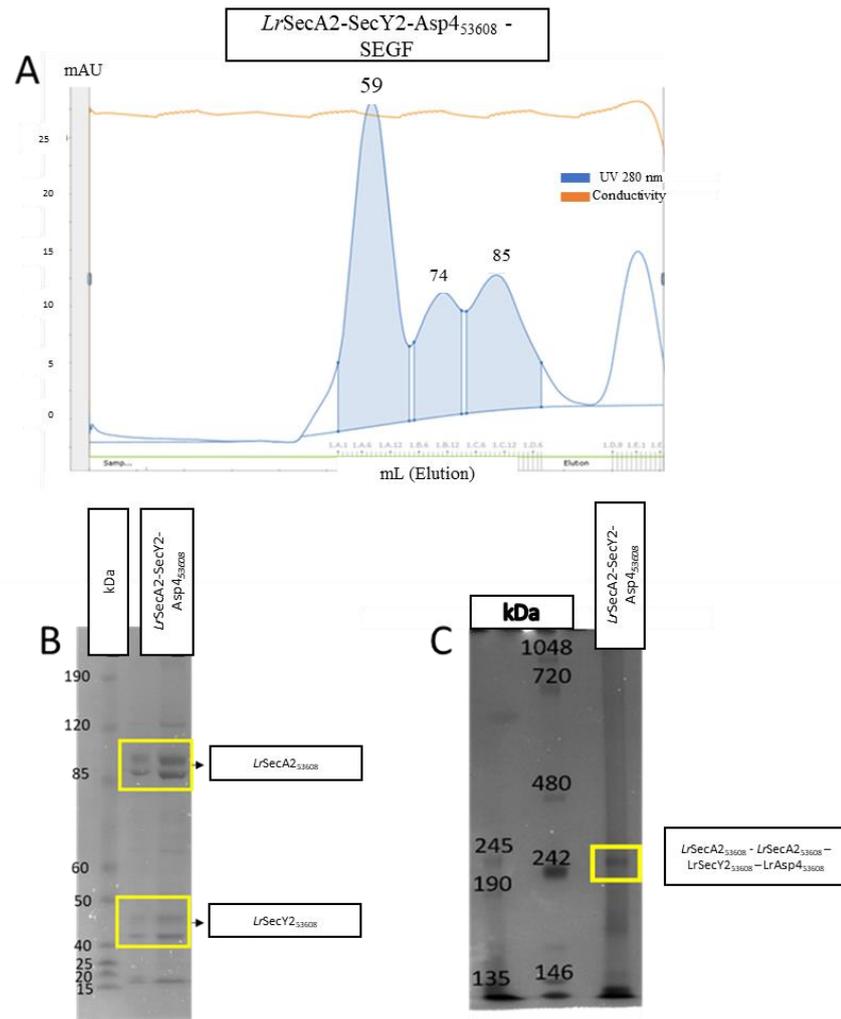
Figure 5.2. Purification of the recombinant *Lr*SecA2-SecY2-Asp4$_{53608}$. Recombinant *Lr*SecA2-SecY2-Asp4$_{53608}$ purified by NiNTA and SEGF (A) SEGF trace of *Lr*SecA2-SecY2-Asp4$_{53608}$ with main peak eluted at 59 mL (B) *Lr*SecA2-SecY2-Asp4$_{53608}$ analysed by SDS-PAGE (C) *Lr*SecA2-SecY2-Asp4$_{53608}$ analysed by native-PAGE. Gels were stained with Coomassie instant-blue. Yellow boxes highlight bands corresponding to expected sizes for the aSec translocon components, or complex.

5.2.2 Structure and function of *Lr*SecA2

SecA2 has been identified as the motor-ATPase of the aSec system, facilitating transport of the SRRP across the cell membrane via SecY2. The energy for transport has been identified to derive from the hydrolysis of high energy phosphoanhydride bonds like ATP to ADP + P$_i$. Here, a malachite green assay was used to colorimetrically detect free P$_i$ released from hydrolysis of ATP and GTP by recombinant *Lr*SecA2 variants. This also included a putative catalytic mutant

*Lr*SecA2 K114A$_{53608}$, which was targeted for site-directed mutagenesis based on structural and sequence homology with *Ec*SecA$_{K12}$ (6GOX), where the corresponding catalytic Lys residue of the Walker A motif conserved in ATPases was identified and validated by site-direct mutagenesis and enzymatic characterisation (127).

Phosphate was quantified using an external standard curve (Figure 5.3A). A velocity plot was derived as a function of ATP/GTP substrate concentration to determine kinetic parameters (Figure 5.3BC, Table 5.1). *Lr*SecA2$_{100-23}$ hydrolysed ATP at a faster rate than *Lr*SecA2$_{53608}$ with a k$_{cat}$ of 5.6 ± 0.2 and 5.3 ± 0.1 nmol/min/nmol, which was also statistically significantly different (t=3.00, df=8, p=0.0171). Both *Lr*SecA2$_{100-23}$ and *Lr*SecA2$_{53608}$ hydrolysed GTP at the same rate with a k$_{cat}$ of 4.1 nmol/min/nmol. The differences in Km and rates between using ATP and GTP as a substrate are summarised in Table 5.1. Overall *Lr*SecA2s showed a preference for ATP although both ATP and GTP could be utilised *in vitro*. Furthermore, the *Lr*SecA2 K114A$_{53608}$ mutant showed a loss of enzymatic activity, confirming K114 as the putative catalytic residue in the *Lr*SecA2 Walker A motif.
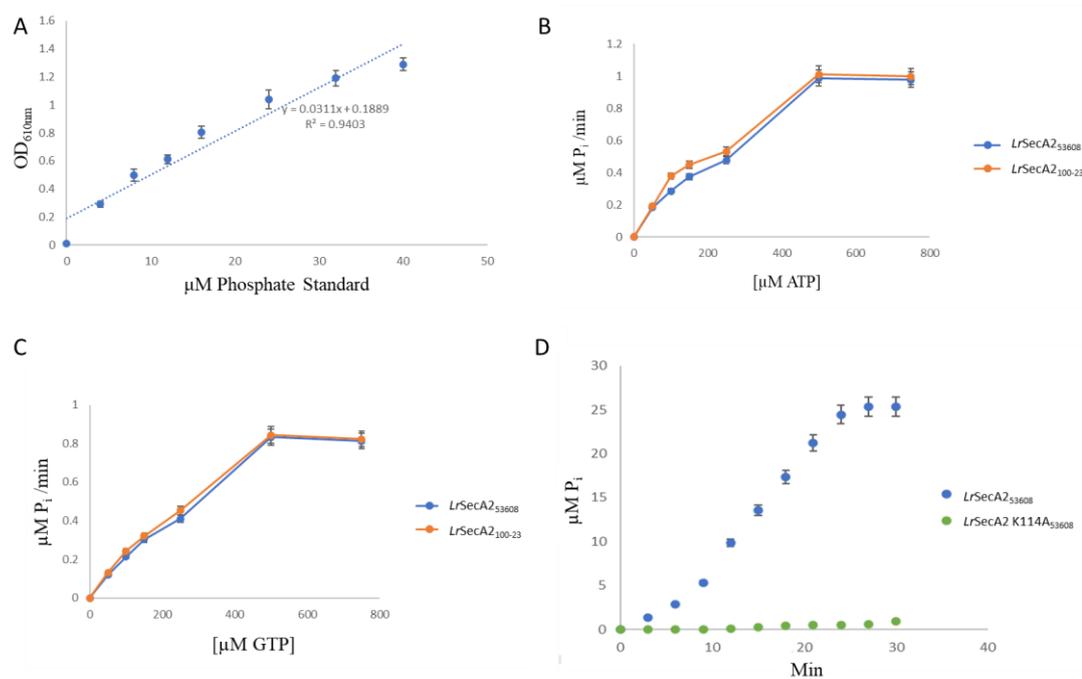
Figure 5.3 *Lr*SecA2 ATP/GTPase enzymatic activity using the malachite green assay. (A) Phosphate standard curve for quantification of free phosphates detected at OD_{610nm}. (B) Rate (μM P_i produced per min) of ATP hydrolysis by *Lr*SecA2$_{53608}$ (blue) and *Lr*SecA2$_{100-23}$ (orange) with increasing ATP concentrations (0-750 μM). (C) Rate (μM P_i produced per min) of GTP hydrolysis by *Lr*SecA2$_{53608}$ (blue) and *Lr*SecA2$_{100-23}$ (orange) with increasing GTP concentrations (0-750 μM). (D) P_i (μM) released per min of ATP hydrolysis by *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{53608}$K144A. Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl, 1 mM MgCl$_2$. Mean rates were derived with 5 replicates and SD error bars are displayed. Blank reactions containing no *Lr*SecA2$_{53608}$, or *Lr*SecA2$_{100-23}$ were used as negative controls.

Table 5.1 Kinetic parameters of *Lr*SecA2 ATP/GTPase activity.

Protein	Kinetic parameters					
	ATP			GTP		
	V_{max} (μmol *p*NP/min)	K_m (μm)	k_{cat} (nmol/min/nmol)	V_{max} (μmol *p*NP/min)	K_m (μm)	k_{cat} (μmol/min/μmol)
*Lr*SecA2$_{53608}$	1.06 ± 0.4	187.8 ± 9.3	5.3 ± 0.1	0.813 ± 0.1	232 ± 6.2	4.1 ± 0.08
*Lr*SecA2$_{100-23}$	1.12 ± 0.3	175.3 ± 11	5.6 ± 0.2	0.818 ± 0.1	229 ± 5.1	4.1 ± 0.09

To test possible associations between *Lr*SecA2s and aSec components, the binding of *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{100-23}$ was tested against native *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ as well as recombinant aSec proteins produced previously, *Lr*Asp1$_{53608}$, *Lr*Asp2$_{53608}$ and *Lr*Asp1-Asp2-Asp3$_{53608}$ (see section 4.2.1) and *Lr*GtfC$_{53608}$ (see section 3.2.1). Binding was detected by ELISA

using an anti-his HRP fusion monoclonal antibody. Statistical inferences were made by two-tailed student t-tests to determine any significant differences between conditions.

LrSecA2$_{53608}$ was shown to bind to LrSRRP$_{53608}$, LrSRRP$_{100-23}$, LrAsp1$_{53608}$, LrAsp2$_{53608}$, and LrAsp1-Asp2-Ap3$_{53608}$ (Figure 5.4A). Differences in the binding of LrSecA2$_{53608}$ to LrSRRP proteins were not statistically significant (p=0.95, t=0.07, df=10), indicating that interactions between the aSec secretion motor ATPase and the secretion cargo was highly conserved. LrSecA2$_{53608}$ did not bind LrGtfC$_{53608.}$ LrSecA2$_{100-23}$ was shown to bind both LrSRRP$_{53608}$ and LrSRRP$_{100-23}$, but the binding responses were not statistically different from one another (p=0.92, t=0.109, df=10). Taken together, these data suggest that *in vitro* associations between LrSecA2 can occur with LrSRRPs irrespective of the strain origin of the secretion cargo. The interactions demonstrated here between LrSecA2$_{53608}$ and LrAsp proteins as monomers and as part of the LrAsp1-Asp2-Asp3$_{53068}$ complex support the hypothesis that cytosolic LrAsp proteins have a chaperone function orientated towards the secretion machinery, and that these cytosolic components of the aSec system may even form larger complexes to facilitate the post-translational modification as well as transport of the SRRP.
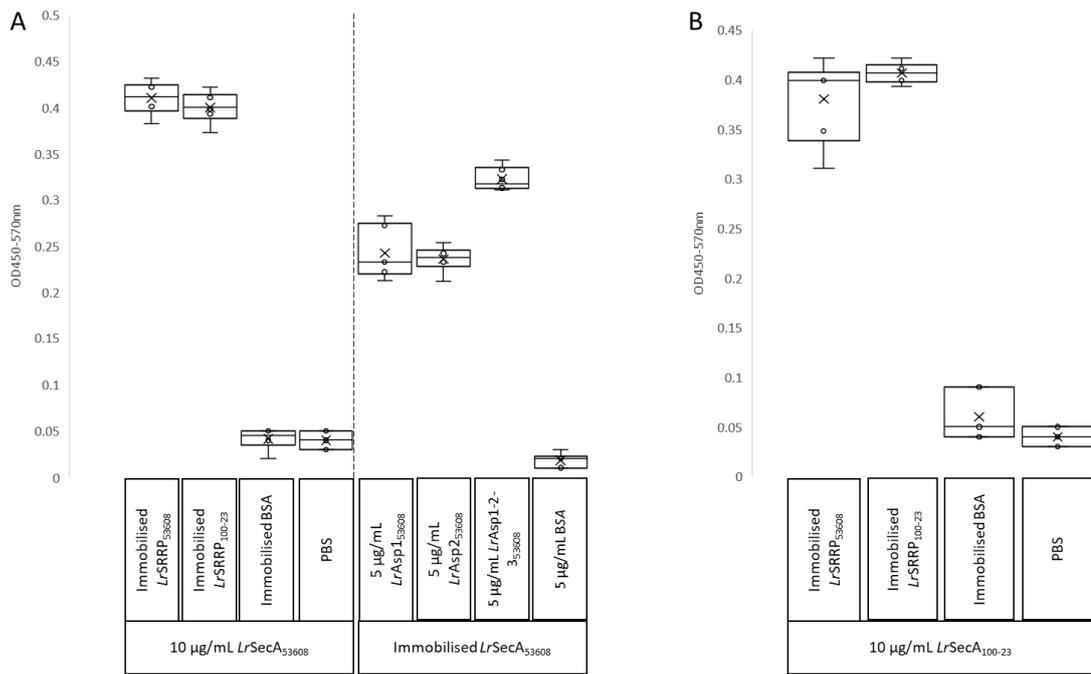
Figure 5.4 ELISA-based binding assay between *Lr*SecA2s and *Lr*aSec components. (A) *Lr*SecA2$_{53608}$ at 10 µg/mL final concentration was added to wells coated with *Lr*SRRP$_{53608}$, *Lr*SRRP$_{100-23}$, BSA or PBS. *Lr*SecA2$_{53608}$ without his-tag was coated to wells and then 5 µg/mL final concentration recombinant *Lr*Asp1$_{53608}$, *Lr*Asp2$_{53608}$, or *Lr*Asp1-Asp2-Asp3$_{53608}$ (B) *Lr*SecA2$_{100-23}$ at 10 µg/mL final concentration was added to wells coated with *Lr*SRRP$_{53608}$, *Lr*SRRP$_{100-23}$, BSA or PBS. Binding was monitored at OD$_{450-570nm}$ following hydrolysis of TMB by HRP-conjugated anti-his monoclonal antibodies. (n=6).

It was attempted to crystallise recombinant forms of *Lr*SecA2 as apoprotein and in ligand complexes, but these trials were unsuccessful (Appendix 2, supplementary Figure S2.6, and Figure S2.7). Due to the lack of crystals suitable for *Lr*SecA2s for X-ray crystallography, Alphafold2 structural models of monomeric *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{100-23}$ were generated and analysed (Figure 5.5AB). *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{100-23}$ share 80% amino acid sequence identity and thus also have similar domain architectures consisting of two nucleotide binding domains (NBDs) separated by a protein-protein cross-linking (PPXD) domain. The Walker A motif and catalytic residue K114 in *Lr*SecA2are located in the NBD1 region, however both NBDs contribute residues associated with nucleotide-binding. The PPXD is attached to NBD1 with a pair of antiparallel β-strands forming a sheet with five hydrogen bonds. Above it, there is a central α-helix surrounded three 3 α-helices, a helix-turn, and a pair of antiparallel β-strands on each side.

Upon structural comparison to other SecA2 structures from SecA1/SecA2 systems as well as SecA from the Sec pathway, some differences between SecA2 proteins from different organisms were noted (Figure 5.5C). CdSecA2$_{630}$ and EcSecA$_{K12}$ had an extended helical-wing domain, but these were missing/truncated in LrSecA2s and SaSecA2$_{H37Rv}$. Helical fingers in the C-terminus of CdSecA2$_{630}$, SaSecA2$_{H37Rv,}$ and EcSecA$_{K12}$ associated with insertion into SecY were not present in LrSecA2, which may be consistent with specific interaction with LrSecY2. LrSecA2 proteins (as well as other SecA2s) also lack the C-terminal tail present in EcSecA$_{K12}$ which is known to facilitate ribosomal interaction for secretion of unfolded proteins, possibly due to the canonical secretion cargo of SecA2 proteins having folded components (such as the SRRP BR) or glycosylation.
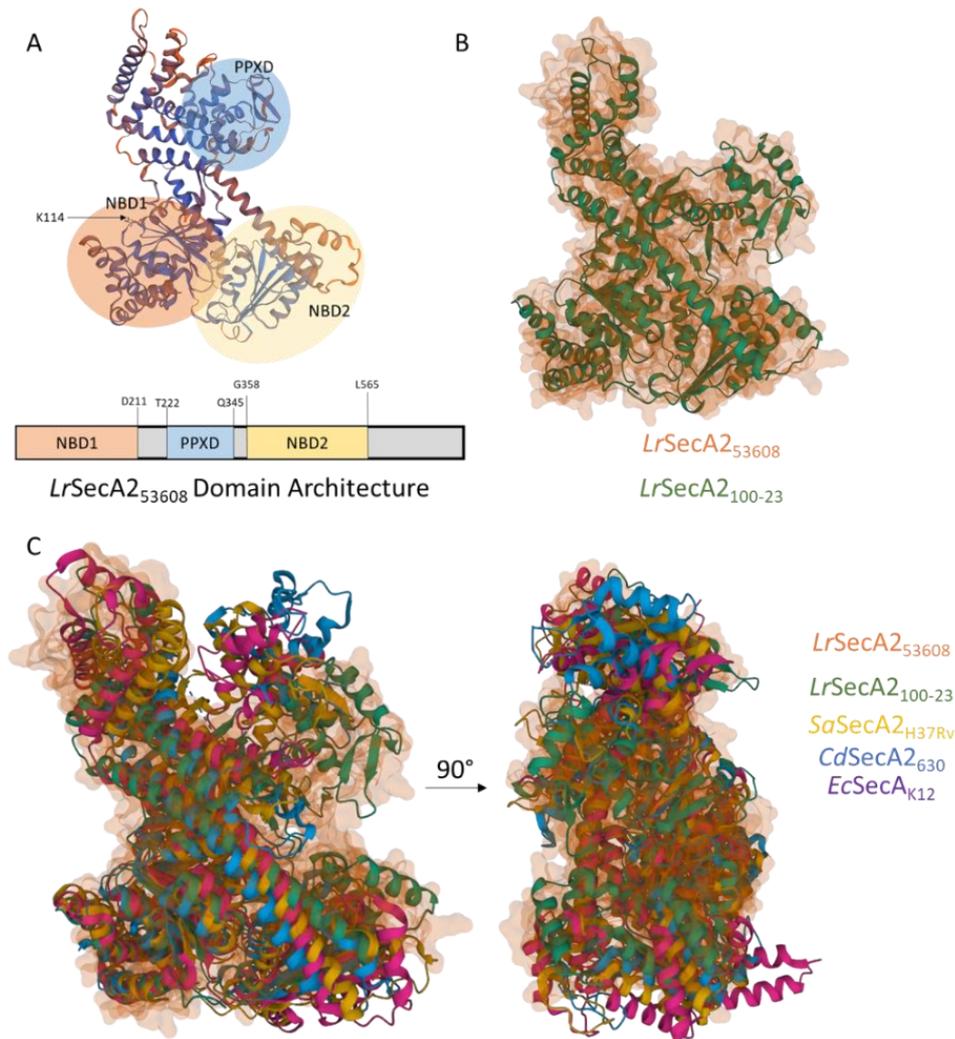
Figure 5.5 Alphafold2 structural models of *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{100-23}$. (A) *Lr*SecA2$_{53608}$ cartoon model with domain organisation and catalytic K114 in the NBD1 walker A motif. NBD = Nucleotide Binding Domain. PPXD = Protein-Protein Cross-linking Domain. (B) *Lr*SecA2$_{53608}$ molecular surface (orange) superimposed on *Lr*SecA2$_{100-23}$ (green) cartoon model. (C) *Lr*SecA2$_{53608}$ molecular surface (orange) superimposed on *Lr*SecA2$_{100-23}$ (green), *Sa*SecA2$_{H37Rv}$ (gold) (4AUQ), *Cd*SecA2$_{630}$ (blue) (6SXH) and *Ec*SecA$_{K12}$ (purple) (6GOX) cartoon models.

5.2.3 Structure and function of *Lr*SecY2

SecY2 is a membrane protein and forms the translocon channel of the aSec system. Therefore, it is predicted to interact favourably with SecA2, the enzyme that provides the energy for transport though hydrolysis of high-energy phosphoanhydride bonds like ATP. To test whether *Lr*SecA2 and *Lr*SecY2 interact *in vitro*, biolayer interferometry (BLI) was used to confirm any associations via amine-reactive biosensors (ARG2) allowing for amine-coupling of recombinant *Lr*SecY2$_{53608}$ as the immobilised ligand and *Lr*SecA2$_{53608}$ as the analyte. A baseline without analyte was used

as a negative control and as a reference sensor to substrate shifts resulting from the buffer response. LrSecA2$_{53608}$ showed a dose-dependent response at 10 nM and 20 nM, indicating that LrSecA2$_{53608}$ and LrSecY2$_{53608}$ bind one another, likely forming a protein complex (Figure 5.6).
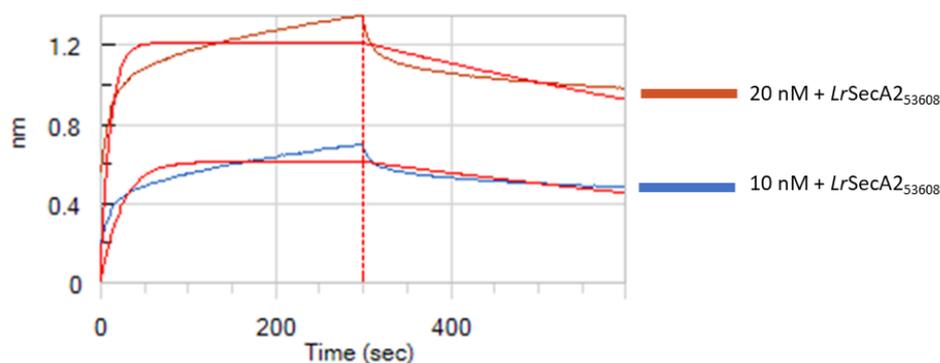
Figure 5.6 Biolayer interferometry binding between LrSecY2$_{53608}$ and LrSecA2$_{53608}$. LrSecY2$_{53608}$ was immobilised to amine-reactive biosensors (ARG2) and exposed to LrSecA2$_{53608}$ as the associating analyte at 10 nM (blue) and 20 nM (red) final concentration.

Since LrSecY2$_{53608}$ and LrSecA2$_{53608}$ show a stable binding response *in vitro*, it is possible that the interaction between these two proteins would have an impact on ATPase activity of LrSecA2$_{53608}$. To test this hypothesis, a malachite green assay was used to colorimetrically detect free P_i released from hydrolysis of ATP by recombinant LrSecA2$_{53608}$, as reported previously (Figure 5.3). A velocity plot was derived as a function of ATP substrate concentration to determine kinetic parameters upon addition of LrSecY2$_{53608}$ (Figure 5.7, Table 5.2). It was confirmed that LrSecY2$_{53608}$ did not hydrolyse ATP. Then addition of LrSecY2$_{53608}$ to LrSecA2$_{53608}$ at equimolar concentration showed a higher rate of hydrolysis with a k_{cat} of 6.1 ± 0.2 nmol/min/nmol compared to LrSecA2$_{53608}$ on its own with 5.3 ± 0.1 nmol/min/nmol. The substrate affinity was also increased with addition of LrSecY2$_{53608}$ (Km = 137.8 ± 8.1 µm) as compared to LrSecA2$_{53608}$ on its own (Km =187.8 ± 9.3 µm). Taken together, these results suggest that LrSecY2$_{53608}$ and LrSecA2$_{53608}$ may form a complex required for optimal enzymatic performance for secretion.
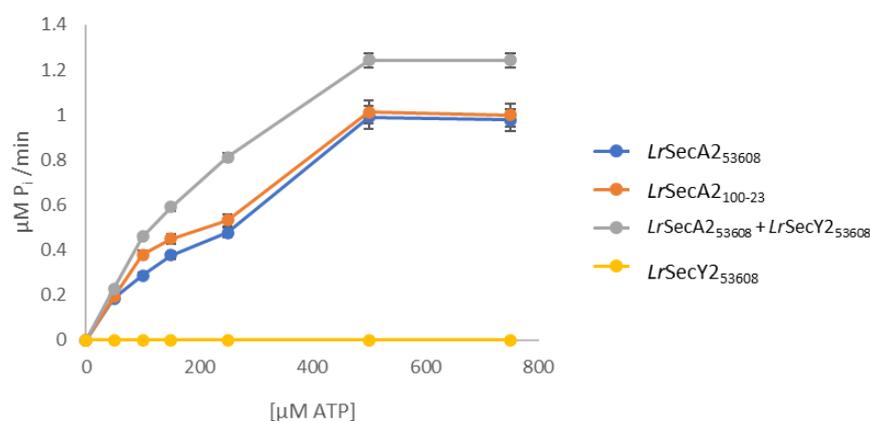
Figure 5.7 _Lr_SecA2 and _Lr_SecY2$_{53608}$ ATPase activity using the malachite green assay. Rate (μM P$_i$ produced per minute) of ATP hydrolysis by _Lr_SecA2$_{53608}$ (blue) and _Lr_SecA2$_{100-23}$ (orange), _Lr_SecA2$_{53608}$ + _Lr_SecY2$_{53608}$ (grey) and _Lr_SecY2$_{53608}$ (yellow) with increasing ATP concentrations (0-750 μM). Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl, 1 mM MgCl$_2$. Mean rates were derived with 5 replicates and SD error bars are displayed. Blank reactions containing no _Lr_SecA2$_{53608}$, or _Lr_SecA2$_{100-23}$ were used as negative controls.

Table 5.2. Kinetic parameters of _Lr_SecA2$_{53608}$ with _Lr_SecY2$_{53608}$ ATPase activity.

Proteins	Kinetic parameters		
	ATP		
	V$_{max}$ (μmol _p_NP/min)	K$_m$ (μm)	k$_{cat}$ (nmol/min/nmol)
_Lr_SecA2$_{53608}$ + _Lr_SecY2$_{53608}$	1.22 ± 0.1	137.8 ± 8.1	6.1 ± 0.2

Some crystallisation trials were conducted for recombinant _Lr_SecY2$_{53608}$ but crystals that grew did not diffract under X-ray conditions (Appendix 2, supplementary Figure S2.8). Alphafold2 models for _Lr_SecY2$_{53608}$ and _Lr_SecY2$_{100-23}$ were generated in an unrelaxed form (Figure 5.8A). Both proteins have 10 predicted transmembrane helices forming a 'plugged' central pore, which likely can form an open conformation during translocation. The distribution of hydrophilic and hydrophobic residues in relation to the pore suggested an orientation where N- and C-terminal ends are exposed at the cytosolic side of the membrane (Figure 5.8B). Whilst SecY2 and SecY share overall structural homology, the C-terminal loop differs between the proteins, as shown when _Lr_SecY2$_{53608}$ was superimposed onto _Tm_SecY$_{MSB8}$ (3DIN) (Figure 5.8C) (186, 325). This

loop is conserved across SecYs and interacts with SecA as part of Sec system transport, as shown by X-ray crystallography and protein cross-linking experiments (325). An additional difference to SecY is the presence of bulkier amino acids such as Met or Leu residues in LrSecY2 around the pore on the cytosolic side, but it is unclear what functionality this may serve in the context of specificity towards the large, glycosylated secretion cargo.
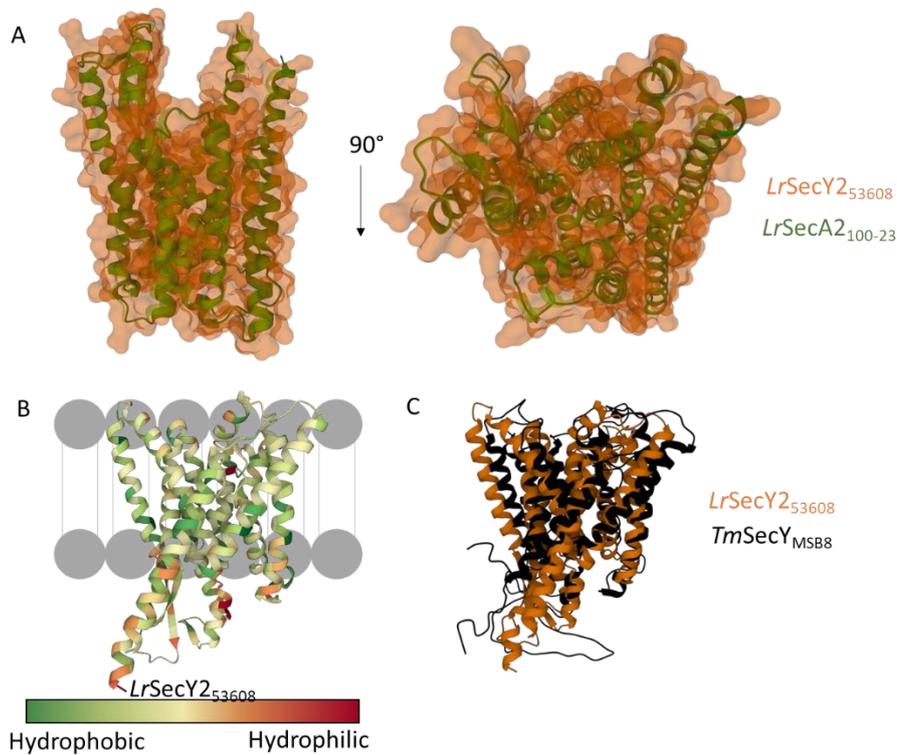
Figure 5.8 Alphafold2 structural models of LrSecY2$_{53608}$ and LrSecY2$_{100-23}$. (A) LrSecY2$_{53608}$ surface model (orange) superimposed on a LrSecY2$_{53608}$ cartoon model (B) LrSecY2$_{53608}$ cartoon model coloured by hydrophobicity and suspected membrane/cytosol orientation. (C) LrSecY2$_{53608}$ cartoon model (orange) superimposed on TmSecY$_{MSB8}$ (black) (3DIN) cartoon model.

5.2.4 Structure and function of the LrSecA2-SecY2-Asp4 complex

LrSecA2$_{53608}$, LrSecY2$_{53608}$ and LrAsp4$_{53608}$ were co-expressed in $E.$ $coli$, but the precise oligomerisation state of the resulting complex is unknown (mainly due to the small size of LrAsp4$_{53608}$ at a theoretical 10 kDa relative to the complex based on amino acid sequence). The precise function of LrAsp4$_{53608}$ remains undetermined and it could not be purified in a soluble

form for characterisation (data not shown). However, since the main target of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex is expected to be the native SRRP secretion cargo, the binding of purified recombinant *Lr*SecA2-SecY2-Asp4$_{53608}$ complex to native *Lr*SRRPs was tested test *in vitro* by ELISA-based assays and BLI.

For the ELISA-based binding assays, results showed a dose-dependent interaction between *Lr*SecA2-SecY2-Asp4$_{53608}$ and native *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ with a two-fold increase in binding when 10 µg *Lr*SecA2-SecY2-Asp4$_{53608}$ was used as compared to 5 µg (Figure 5.9). There was no binding to recombinant *Lr*SP$_{53608}$ or the BSA negative control. The association of *Lr*SecA2-SecY2-Asp4$_{53608}$ to *Lr*SRRPs was irrespective of *L. reuteri* strain origin.

BLI was then used to determine binding kinetic data and avidity constants. Amine-reactive biosensors (ARG2) allowed for amine-coupling of the native *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ to the biosensor, while *Lr*SecA2-SecY2-Asp4$_{53608}$ was used as the analyte. A baseline without analyte was used as a negative control and as a reference sensor to subtract shifts resulting in buffer response. There was an increased response following increasing *Lr*Asp1-Asp2-Asp3$_{53608}$ concentrations which allowed for fitting the association and dissociation kinetics with *Lr*SRRP$_{53608}$ (Figure. 5.10A-E). The same was repeated with *Lr*SRRP$_{100-23}$ immobilised to the biosensor (Figure 5.10F). The binding responses followed a similar pattern for *Lr*SRRP$_{100-23}$ and for *Lr*SRRP$_{53608}$. A positive control based on the anti-his monoclonal antibody was also used to provide a benchmark of a specific binding interaction and test whether the presence of detergent with *Lr*SecA2-SecY2-Asp4$_{53608}$ may interfere with the assay (Figure 5.10G). Interaction with the anti-his monoclonal antibody to the his-tagged *Lr*SecY2$_{53608}$ within the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex showed a Kd of 5.1 x 10^{-9}. Steady-state kinetic data also showed a high-avidity interaction between *Lr*SecA2-SecY2-Asp4$_{53608}$ and both *Lr*SRRP$_{53608}$ and *Lr*SRPP$_{100-23}$ with a Kd of 2.1 x 10^{-8} M confirming strong binding (Table 5.3). This is approximately 10-fold higher than the interaction of both *Lr*SRRP$_{53608}$ and *Lr*SRPP$_{100-23}$ with *Lr*Asp1-Asp2-Asp3$_{53608}$, reported previously (Table 5.4).
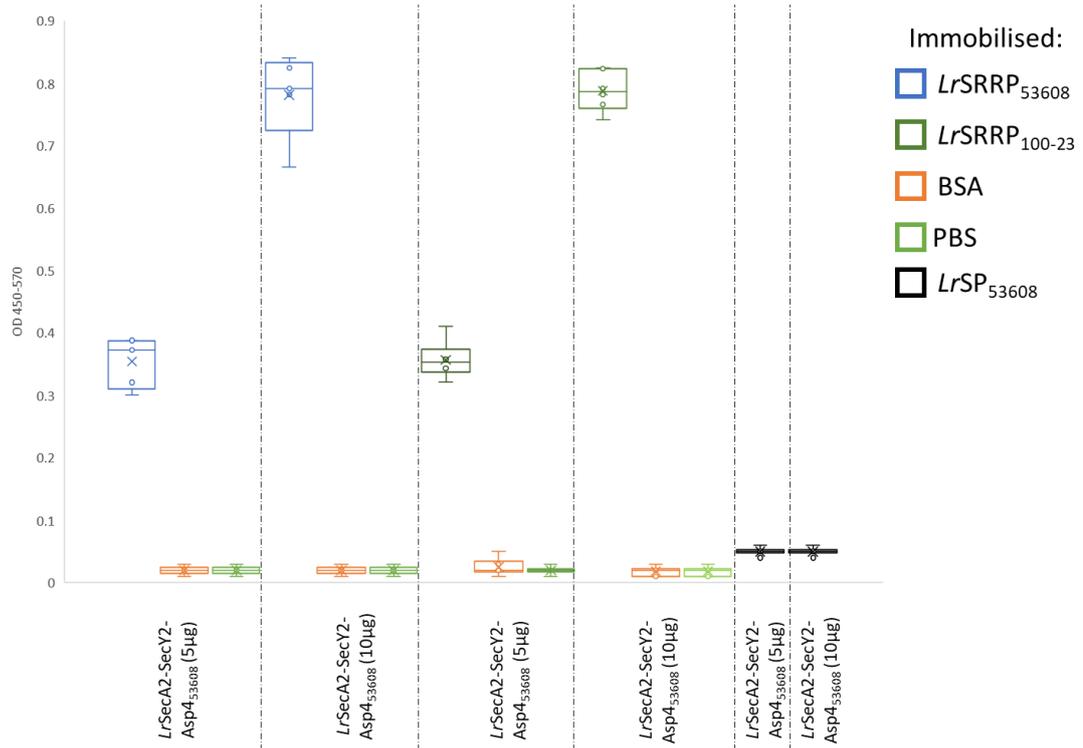
Figure 5.9 ELISA-based binding assay between *Lr*SecA2-SecY2-Asp4$_{53608}$ and *Lr*SRRPs. *Lr*SecA2-SecY2-Asp4$_{53608}$ at 5 and 10 µg was added to wells coated with *Lr*SRRP$_{53608}$, *Lr*SRRP$_{100-23}$, *Lr*SP$_{53608}$, BSA or PBS. Binding was determined by monitoring OD$_{450-570nm}$ following hydrolysis of TMB by HRP-conjugated anti-his monoclonal antibodies. (n=6).
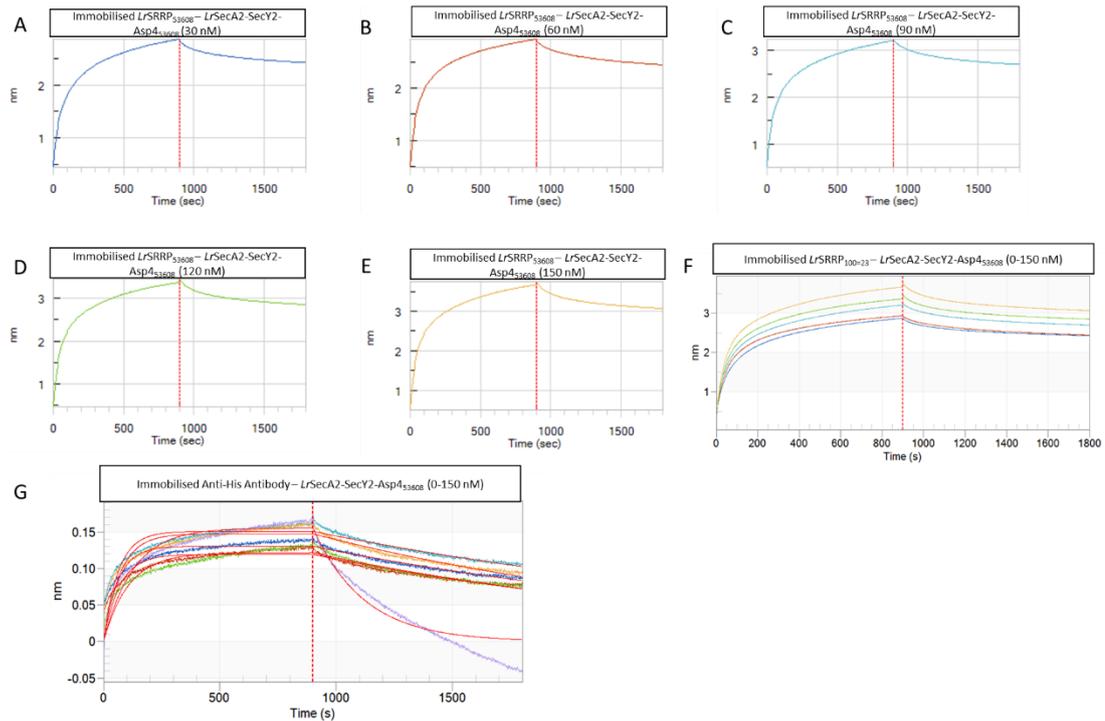
Figure 5.10 Biolayer interferometry binding between *Lr*SecA2-SecY2-Asp4$_{53608}$ complex and *Lr*SRRPs. (A) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration 30 nM) association and dissociation fitting. (B) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration 60 nM) association and dissociation fitting. (C) Immobilised SRRP$_{53608}$ exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration 90 nM) association and dissociation fitting. (D) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration 120 nM) association and dissociation fitting. (E) Immobilised *Lr*SRRP$_{53608}$ exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration 150 nM) association and dissociation fitting. (F) Immobilised *Lr*SRRP$_{100-23}$ exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration gradient 0 – 150 nM) association and dissociation fitting. (G) Immobilised anti-his monoclonal antibody exposed to *Lr*SecA2-SecY2-Asp4$_{53608}$ (final concentration gradient 0 – 150 nM) association and dissociation fitting (positive-control). (n=3)

Table 5.3 *Lr*SecA2-SecY2-Asp4$_{53608}$ BLI binding steady-state kinetic parameters.

Ligand (immobilised to biosensor	Analyte (association)	Kinetic parameters		
		R^2	Rmax	Kd (M)
*Lr*SRRP$_{53608}$		0.84	3.67 ± 0.15	$2.1 \times 10^{-8} \pm 1.0 \times 10^{-9}$
*Lr*SRRP$_{100-23}$	*Lr*SecA2-SecY2-Asp4$_{53608}$	0.82	3.64 ± 0.13	$2.1 \times 10^{-8} \pm 1.1 \times 10^{-9}$
Anti-His antibody		0.72	0.188 ± 0.18	$5.1 \times 10^{-9} \pm 2.9 \times 10^{-10}$

As *Lr*SecA2 was shown to interact with *Lr*SecY2 in the context of ATPase activity (see section 5.2.3) and to *Lr*Asps in the context of binding interactions (see section 5.2.2), the effect of

interactions between LrSecA2-SecY2-Asp4$_{53608}$ and LrAsp1-Asp2-Asp3$_{53608}$ on enzymatic activity was also explored.

The ATPase activity of the LrSecA2-SecY2-Asp4$_{53608}$ complex was tested in the presence of LrAsp1-Asp2-Asp3$_{53608}$ at an equimolar concentration using the malachite green assay (Figure 5.11A). The LrSecA2-SecY2-Asp4$_{53608}$ complex exhibited ATP hydrolysis and addition of LrAsp1-Asp2-Asp3$_{53608}$ greatly increased the enzymatic rate resulting in an almost doubling of the k_{cat} when compared to LrSecA2-SecY2-Asp4$_{53608}$ on its own (11.03 and 6.79 nmol/min/nmol, respectively) (Table 5.4). This was also higher than LrSecA2$_{53608}$ by itself or when LrSecA2$_{53608}$ was assayed in the presence of LrSecY2$_{53608}$ (Figure 5.6). LrAsp1-Asp2-Asp3$_{53608}$ on its own exhibited no ATPase activity.

The acetylesterase activity of LrAsp1-Asp2-Asp3$_{53608}$ was also tested in combination with LrSecA2-SecY2-Asp4$_{53608}$ at an equimolar concentration using pNP-acetate as a substrate as reported previously for LrAsp1-Asp2-Asp3$_{53608}$ (see section 4.2.4, Figure 4.12, Table 4.6). LrSecA2-SecY2-Asp4$_{53608}$ showed no activity on pNP-acetate (Figure 5.11B). However, the presence of LrSecA2-SecY2-Asp4$_{53608}$ positively affected the enzymatic rate of LrAsp1-Asp2-Asp3$_{53608}$, with a k_{cat} of 14.25 µmol/min/µmol compared to 12.7 µmol/min/µmol for LrAsp1-Asp2-Asp3$_{53608}$ alone (Table 5.5).

Taken together, the positive effects reported for the respective enzymatic activity of LrSecA2-SecY2-Asp4$_{53608}$ and LrAsp1-Asp2-Asp3$_{53608}$ suggest that a larger LraSec complex may form at the cell membrane facilitating post-translational modification of the LrSRRP and subsequent transport.
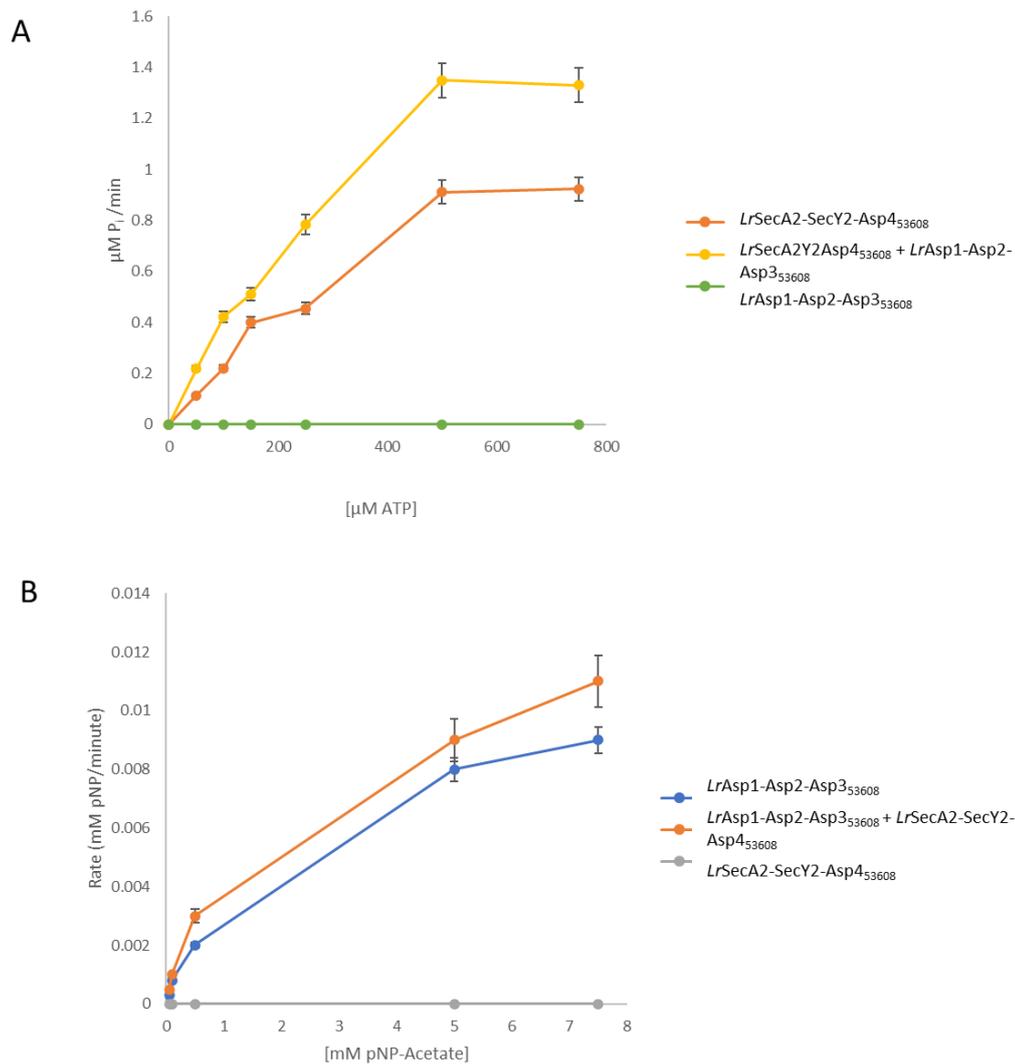
Figure 5.11 Effect of *Lr*aSec complex interactions on ATPase and acetylesterase activities. (A) Malachite green ATPase assay showing rate (μM Pi produced per minute) of ATP hydrolysis by *Lr*SecA2-SecY2-Asp4$_{53608}$ (orange), *Lr*SecA2-SecY2-Asp4$_{53608}$ + *Lr*Asp1-Asp2-Asp3$_{53608}$ (yellow) and *Lr*Asp1-Asp2-Asp3$_{53608}$ (green)with increasing ATP concentrations (0-750 μM). Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl, 1 mM MgCl2. Mean rates were derived and SD error bars are displayed (n=5). (B) Rate (mM *p*NP produced per minute) of *p*NP-acetate hydrolysis by *Lr*Asp1-Asp2-Asp3$_{53608}$ (in blue), *Lr*Asp1-Asp2-Asp3$_{53608}$ + *Lr*SecA2-SecY2-Asp4$_{53608}$ (orange), and *Lr*SecA2-SecY2-Asp4$_{53608}$ (grey) with increasing concentration of *p*NP-acetate substrate. Reactions were conducted in 96-well plates and incubated at 25°C in HEPES pH 7.5, 200 mM NaCl. Mean rates were derived and SD error bars are displayed (n=5).

Table 5.4. Effect of *Lr*aSec complex interactions on ATPase kinetic parameters.

Proteins	Kinetic parameters		
	ATP		
	V_{max} (µmol P_i/min)	K_m (µm)	k_{cat} (nmol/min/nmol)
*Lr*SecA2-SecY2-Asp4$_{53608}$	0.813 ± 0.4	176 ± 4.7	6.78 ± 0.2
*Lr*SecA2-SecY2-Asp4$_{53608}$ + *Lr*Asp1-Asp2-Asp3$_{53608}$	1.323 ± 0.9	163 ± 5.3	11.03 ± 0.8
*Lr*Asp1-Asp2-Asp3$_{53608}$	0	0	0

Table 5.5. Effect of *Lr*aSec complex interactions on acetylesterase activity kinetic parameters.

Proteins	Kinetic parameters		
	ATP		
	V_{max} (µmol pNP/min)	K_m (µm)	k_{cat} (µmol/min/µmol)
*Lr*Asp1-Asp2-Asp3$_{53608}$	9.12 ± 0.09	362 ± 35	12.7 ± 0.008
*Lr*Asp1-Asp2-Asp3$_{53608}$ + *Lr*SecA2-SecY2-Asp4$_{53608}$	10.12 ± 0.08	347 ± 38	14.25 ± 0.01
*Lr*SecA2-SecY2-Asp4$_{53608}$	0	0	0

Despite several attempts, the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex crystals diffracted as salt (Appendix 2, supplementary Figure S2.9). Instead, SEC-SAXS was performed, revealing a size-exclusion trace with two peaks, suggesting some degradation or aggregation of *Lr*SecA2-SecY2-Asp4$_{53608}$ over exposure, during transportation and storage to the beamline, or the presence of different oligomerisations of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex (Figure 5.12A). Ultimately, the larger peak was selected for further analysis between 268-274 frames and was used for downstream analysis, since it corresponded to the expected run-time for the theoretical size of *Lr*SecA2-SecY2-Asp4$_{53608}$ (245 kDa). The scattering profile and Guinier fit analysis of the peak area were indicative of a soluble, globular protein, albeit with some abnormally distributed residuals (Figure 5.12 BC). This was supported by the Kratky-Plot which assessed the degree of unfolding in the sample whereby the peak of the gaussian bell shaped distribution was slightly skewed to the right. The size and shape were estimated using the P(r) function allowing for estimation of size and shape as well as the maximum dimension (Dmax) needed for further molecular weight estimates. A slight tail end was observed when approaching Dmax suggesting

that some aggregation, albeit minor, may have occurred (Figure 5.12E). Multiple size and shape calculations were taken into consideration with the Bayes estimation predicting the apparent size of LrSecA2-SecY2-Asp4$_{53608}$ to be between 147.5 to 233.4 kDa (Table 5.6). A bead model of LrSecA2-SecY2-Asp4$_{53608}$ was constructed showing a cylindrical shape (5.12F). Unfortunately, it was not possible to fit existing Alphafold2 models into the bead model of LrSecA2-SecY2-Asp4$_{53608}$ reliably with a high-enough degree of statistical confidence that would allow for the proteins to be orientated and oligomerised correctly. However, the size and protein shape estimation seem consistent with that expected for LrSecA2-SecY2-Asp4$_{53608}$ as a large protein complex of 245 kDa.
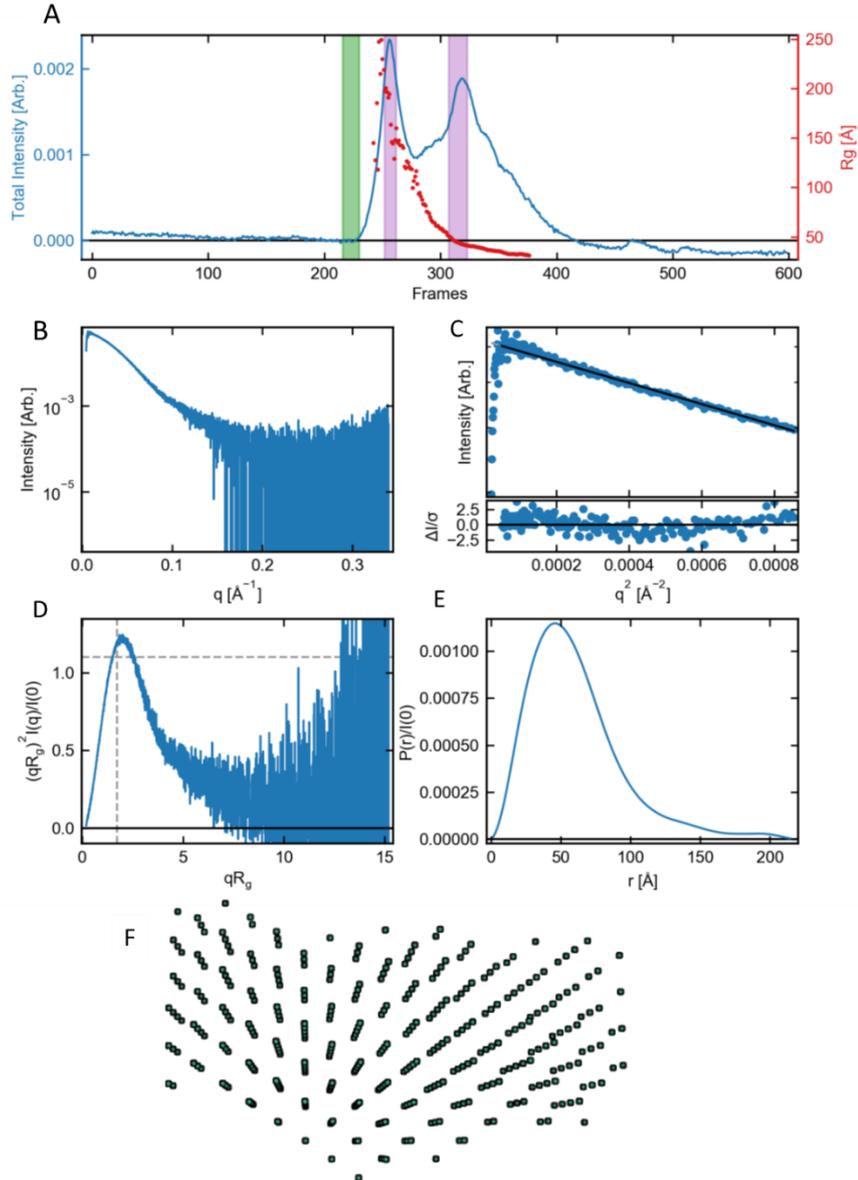
Figure 5.12 SAXS data summary for _Lr_SecA2-SecY2-Asp4$_{53608}$ (A) Series intensity (blue, left axis) vs. frame, and Rg vs. frame (red, right axis). Green shaded regions are buffer regions, purple shaded regions are sample regions. (B) Scattering profile(s) on a log-lin scale. (C) Guinier fit(s) (top) and fit residuals (bottom). (D) Normalized Kratky plot. Dashed lines show where a globular system would peak. (E) P(r) function(s), normalized by I(0). (F) Fitted bead-model of _Lr_SecA2-SecY2-Asp4$_{53608}$.

Table 5.6. SAXS-derived properties of *Lr*SecA2-SecY2-Asp4$_{53608}$

Molecular weight and shape calculation	Molecular weight and shape estimate
Molecular weight (Vp)	218.4 kDa
Porod Volume [A^3]	$2.01e^5 A^3$
Molecular weight (Vc)	194.5 kDa
Molecular weight (S&S)	212.2 kDa
Shape (S&S)	Compact
Dmax (S&S)	148.8
Molecular weight (Bayes)	221.2 kDa
Molecular weight (Bayes-Confidence Interval)	147.5 to 233.4 kDa

Despite these limitations, a structural model of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex was constructed using the individually generated Alphafold2 structures and superimposing with the existing SecAYEG structure from *Geobacillus thermodenitrificans* NG80-2 (5EUL) (Figure 5.13) (320). The model depicts *Lr*SecA2-SecY2-Asp4$_{53608}$ with a monomeric *Lr*SecA2$_{53608}$, as the structural form a *Lr*SecA2$_{53608}$ dimer is unknown. The *Lr*SecA2-SecY2-Asp4$_{53608}$ structure did however highlight how *Lr*SecY2$_{53608}$ could directly associate with *Lr*SecA2$_{53608}$ at the PPXD domain through amino acid residues 305-318, as well as at the C-terminal NBD1 through amino acid residues 390-398. In addition to the undetermined oligomeric state of *Lr*SecA2$_{53608}$, another limitation of the model was the construction of *Lr*Asp4$_{53608}$ structural model (both by Alphafold2 or by homology modelling). Due to its flexible nature, it was difficult to superimpose, despite sharing 30% amino acid sequence identity with SecE from *Thermus thermophilus* HB8 used here as a template (Figure 5.13B).
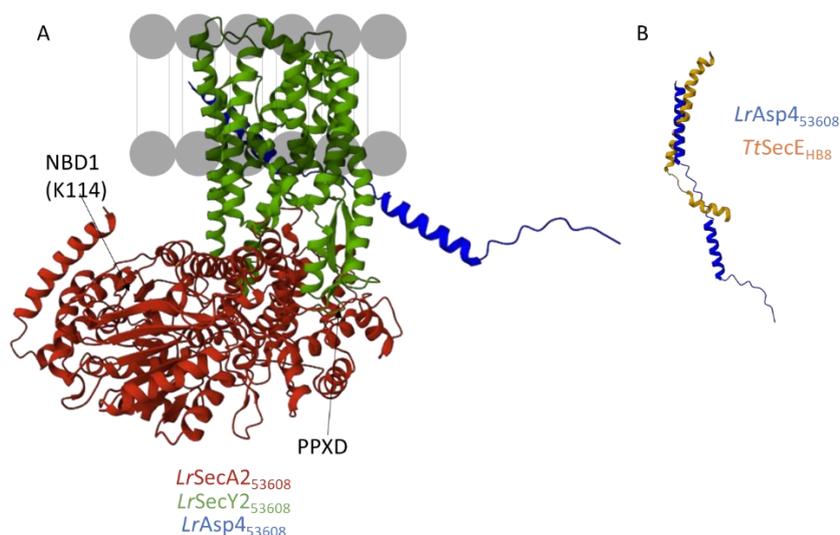
Figure 5.13 Alphafold2 structural model of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex. (A) Cartoon model of *Lr*SecA2-SecY2-Asp4$_{53608}$ consisting of *Lr*SecA2$_{53608}$ (red), *Lr*SecY2$_{53608}$ (green) and *Lr*Asp4$_{53608}$ (blue). (B) Cartoon model of *Lr*Asp4$_{53608}$ superimposed on *Tt*SecE$_{HB8}$ (5AWW).

5.3 Discussion

The aSec translocation machinery is dedicated to the transport of large, glycosylated SRRPs across the cell membrane in some gram-positive bacteria in an ATP-dependent manner. Here, biochemical, and biophysical approaches, including machine learning modelling (Alphafold2) and SAXS, activity and binding assays were applied to investigate the structure and function of SecA2, SecY2 and Asp4 involved in the *L. reuteri* aSec pathway.

Alphafold2 was used to generate structural models of *Lr*SecA2, *Lr*SecY2 and *Lr*Asp4. To date, no structural data exist for the translocation machinery of the aSec pathway, except for SecA2 proteins from SecA1/SecA2 based systems (*Mt*SecA2$_{H37Rv}$ (4AUQ), *Cd*SecA2$_{630}$) which are structurally distinct from aSec *Lr*SecA2 systems, with helical-wing domains and helical fingers associated with integration to SecY of the Sec system (130, 188). The SecA2 proteins from the Sec-dependent pathway have also only been crystallised as monomers, with only dimers being crystallised for SecA in both parallel and anti-parallel conformations (132, 321). So, whilst there is some agreement that transport in the Sec system is a dimer-driven translocation, and by

comparison the aSec system as well, the structural form of the dimeric SecA and SecA2 remains controversial (127, 130, 189). Here, the *Lr*SecA2 structural models showed functional residues for ATPase activity present in the NBD1 catalytic region as part of a Walker A motif and the putative catalytic residue was identified as Lys114 in *Lr*SecA2$_{53608}$ and confirmed to be essential for ATPase hydrolysis by site-directed mutagenesis, as previously reported for Lys156 in *Cd*SecA2$_{630}$ (188).

The *Lr*SecY2 structural models confirmed their structural similarities to SecY. However, no discrete regions attributed to ribosomal and SecA binding required for transport of unfolded pre-proteins were apparent. This was consistent with differences to the canonical secretion system and specificity of SecY2 translocation for the large, glycosylated SRRP as well as complex formation with SecA2 (318). Binding interaction between *Lr*SecY2$_{53608}$ and *Lr*SecA2$_{53608}$ was confirmed *in vitro* by BLI. Increased ATPase activity was reported for *Lr*SecA2$_{53608}$ in the presence of *Lr*SecY2$_{53608}$, suggesting that is likely that these proteins may form a complex.

Indeed, structural analysis of the *Lr*SecA2-SecY2-Asp4$_{53608}$ by SAXS, as well as analysis of the complex by SDS-PAGE and native-PAGE, indicated that co-expression of all recombinant proteins predicted to be part of the aSec translocation machinery proteins may result in the formation of a 245 kDa complex. This complex is likely comprised of a dimerised SecA2, monomeric SecY2 and either a monomeric or dimeric Asp4. Unlike pathogenic aSec clusters (e.g., *S. gordonii* M99, *S. gordonii* DL1, or *S. pneumoniae* TIGR4), *L. reuteri* strains ATCC 53608 and 100-23 lack *Asp5* in their genome (195, 318). Not much is known of the function of Asp4 or Asp5, but due to the presence of hydrophobic residues, they are predicted to aid with assembly of the transmembrane complex (318). Due to the highly disordered and unusual structure of the proteins, protein modelling tools such as Swissmodel and Alphafold2 were not able to generate a structure of *Lr*Asp4 that could be accommodated in the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex and used to fit SAXS bead model structures. A structural model of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex was built based on superimposition on the *Gt*SecAYEG$_{NG80-2}$ crystal structure (5EUL) (320). However, due to aforementioned lack of clarity on structural form of both the dimerised

*Lr*SecA2 and *Lr*Asp4, it is unlikely that an *in-silico* method can be utilised to gain further insights for the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex, therefore obtaining high-resolution structural data will be necessary to address remaining gaps in the structure of these complexes.

The ATPase activities of recombinant *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{100-23}$ were determined as monomers and for *Lr*SecA2$_{53608}$, in the presence of *Lr*SecY2$_{53608}$ as well as part of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex in the absence or presence of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex. The k$_{cat}$ of 4.1 µmol/min/µmol for *Lr*SecA2$_{53608}$ and *Lr*SecA2$_{100-23}$, consistent with reports for recombinant *Mt*SecA2$_{H37Rv}$ (4.73 µmol/min/µmol) (130). We also showed that *Lr*SecA2 proteins could utilise GTP as a substrate, which had not been tested previously for SecA or SecA2 proteins but has been shown to be a substrate of multi-drug ABC transporters PatA/PatB from *S. pneumoniae* R6 (326). *Lr*SecA2$_{100-23}$ showed a marginally higher rate of ATP hydrolysis than *Lr*SecA2$_{53608}$, which could be due to the size of *Lr*SRRP$_{100-23}$ being larger than *Lr*SRRP$_{53608}$, therefore driving selection for a more energetically efficient enzyme. However, the main differences in ATPase activity were observed when *Lr*SecA2$_{53608}$ was part of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex and in the presence of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex, showing a 3-fold increased k$_{cat}$ as compared to monomeric *Lr*SecA2$_{53608}$.

The *Lr*SecA2-SecY2-Asp4$_{53608}$ complex also had a positive impact on the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex acetylesterase activity *in vitro*. The enzymatic activity of these recombinant aSec protein complexes had not been investigated before. Here, the findings suggest that the most favourable state for the best activity of the enzymatic components of the aSec protein complexes is when in direct association with one another. It further supports the hypothesis that the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex acts as a chaperone for the SRRP cargo bringing it to the membrane-localised translocation machinery and that additionally *O*-acetylation may therefore occur as the SRRP is processed through the membrane.

The binding kinetics for *Lr*SecA2-SecY2-Asp4$_{53608}$ showed nanomolar-avidity towards *Lr*SRRPs irrespective of their strain origin. Avidity was preferred to affinity, which is specific to a single binding site, because it is unclear which parts of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex bound

*Lr*SRRP and multiple interactions are likely to occur. This binding avidity was 10-fold stronger than for *Lr*Asp1-Asp2-Asp3$_{53608}$, which could be related to the sequence of cytosolic events leading the SRRP from cytoplasmic complexes (*Lr*GtfAB, *Lr*Asp1-Asp2-Asp3) to the membrane complex (*Lr*SecA2-SecY2-Asp4) with an increasing gradient of molecular binding avidities. However, unlike *Lr*Asp1-Asp2-Asp3, which showed carbohydrate-binding motifs and structural features alluding to SRRP binding, such features were not observed in the structural model of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex (or as individual proteins).

In summary, our structural and biochemical characterisation of the recombinant *L. reuteri* aSec translocation machinery revealed ATP/GTPase activity of *Lr*SecA2 as a monomer and as part of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex whereby the highest activity and ATP-affinity was observed in the presence of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex. It was also shown that binding interactions occurred between monomeric *Lr*SecA2 and the native *Lr*SRRP proteins and recombinant *Lr*Asp1-Asp2-Asp3$_{53068}$. These interactions were also reported for the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex. Formation of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex demonstrated the best functional capacity *in vitro*, but its precise oligomeric state is still unclear. This is compounded by challenges in solving the structural form of *Lr*Asp4$_{53608}$ as well as the unclear dimer orientation of *Lr*SecA2. Insights into the *Lr*SecA2-SecY2-Asp4$_{53608}$ provided by Alphafold2 structures and SAXS did however suggest the presence of a dimerised *Lr*SecA2, monomeric *Lr*SecY2 and a monomeric or dimeric *Lr*Asp4 in the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex.

Future work on the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex should aim at getting high resolution structural data using single particle cryo electron microscopy (cryoEM), as some of the challenges crystallising membrane protein complexes may be averted using this technique. Since no overt motifs for specificity towards SRRPs could be found in structural models of the *Lr*SecA2-SecY2-Asp4$_{53608}$ complex so far, cryoEM may also lend itself well to structurally characterising the *Lr*SecA2-SecY2-Asp4$_{53608}$ in complex with *Lr*SRRPs as we have established that high avidity interactions take place and cryoEM is suitable for glycoproteins (as opposed to X-ray

crystallography) (327, 328). Preliminary attempts at solving the 3D structure by cryoEM of the native LrSRRP$_{53608}$ and LrSRRP$_{100-23}$ are in progress.

Chapter 6 Conclusions and Future Work

The characterisation of the aSec system has mainly been a research focus in pathogenic gram-positive bacteria (182, 192). The strain-specific glycosylation of *Lr*SRRP from ATCC 53608 (from pig) and 100-23 (from rodent) and the crystal structure of *Lr*SRRP-BR provided first insights into the structure and function of aSec systems in gut symbionts (195, 235). Here, recombinant aSec proteins from *L. reuteri* ATCC 53608 and 100-23 strains were characterised biochemically and structurally, uncovering interactions occurring between individual aSec proteins or aSec protein complexes and their respective ligands. Our results are consistent with a stepwise process where *Lr*SRRP are first modified through *O*-glycosylation by dedicated *Lr*GTs as well as *O*-acetylation by *Lr*Asp2 in the cytosol, and then being transported to the cell membrane by the chaperoning action of the *Lr*Asp1-Asp2-Asp3 complex and translocation using the *Lr*SecA2-SecY2-Asp4 complex (Figure 6.1).
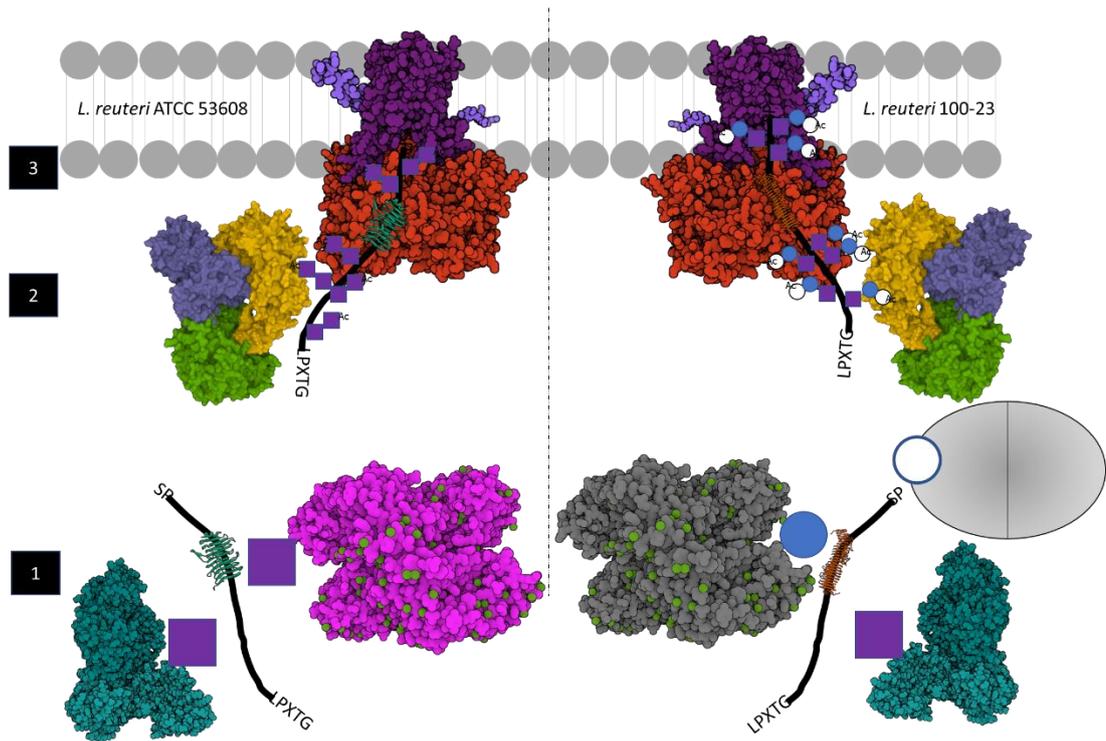
Figure 6.1 Model of aSec system in *L. reuteri* ATCC 53608 and 100-23 strains. The aSec secretion pathway is governed by three steps: (1) Glycosylation of *Lr*SRRP (2) *O*-acetylation and chaperoning of the *Lr*SRRP and (3) Translocation of *Lr*SRRP across the cell membrane by active transport. Stylized protein structural models of aSec proteins and protein complexes: *Lr*GtfAB (cyan), *Lr*GtfC$_{53608}$ (magenta), *Lr*GtfC$_{100-23}$ (grey), *Lr*GtfEF$_{100-23}$ (gray), *Lr*Asp1 (yellow)-Asp2 (green)-Asp3 (blue), *Lr*SecA2 (red)-SecY2 (purple)-Asp4 (light blue) *Lr*SRRP (black) with ribbon model of respective BR. GlcNAc, Glc and hexose are represented in standard glycan symbols.

Step one involves the glycosylation of SRRPs at SRR regions by a diverse set of glycosyltransferases that vary between bacteria on a strain level. The deposition of a core *O*-GlcNAc by the GtfAB complex is highly conserved and present in all SRRPs characterised thus far (195, 222, 235, 276, 280). The Alphafold2 structure of *Lr*GtfAB$_{53608}$ and *Lr*GtfAB$_{100-23}$ generated here showed a high structural homology to known crystal structures of the *Sg*GtfAB$_{M99}$ complex (276). Strain-specific glycosylation in the aSec pathway is introduced in *L. reuteri* by *Lr*GtfC (235). Here, X-ray crystal structures showed a tetrameric *Lr*GtfC$_{100-23}$ in the apo form or in complex with UDP or UDP-GlcNAc, supporting a structurally guided site-directed mutagenesis strategy. We showed that substitutions of single amino acid residues in the *Lr*GtfC

active-site cleft could affect UDP-sugar specificity, whereby LrGtfC P238S$_{53608}$ conferred *in vitro*

UDP-Glc specificity characteristic of LrGtfC$_{100-23}$ and LrGtfC W240C$_{100-23}$ introduced *in vitro*

promiscuity to both UDP-GlcNAc and UDP-Glc.

However, further *in vitro* biochemical data are warranted to validate the LrGtfC residue-mediated

specificity reported here. Whilst ligand binding to a target enzyme can stabilise its native state,

therefore increasing the bound protein's melting temperature which was exploited here in thermal

shift assays, it remains to be demonstrated that LrGtfC mutants will transfer UDP-sugars to the

target acceptors (329). Therefore, additional *in vitro* assays need to be performed using

recombinant LrGtfC WT and mutants generated in this work, with UDP-sugar ligands and a

recombinant LrgSRR1 acceptor to monitor alterations in sugar specificity. The characterisation

of the glycosylation profile can then be performed by western-blot analysis with lectin staining,

or by glycomics using MALDI-TOF MS/MS and 2D NMR, as was done previously for LrSRRPs

(235). Further strain-specific glycosylation differences between *L. reuteri* SRRPs are due to other

GTs present in LraSec operons, leading to further glycosylation of LrSRRP by hexose residues

by GtfE/F in *L. reuteri* 100-23 strain (235). These GTs remain to be structurally investigated.

There is an increasing appreciation of the importance and extent of protein glycosylation in

bacteria influencing cell surface adhesion, colonisation and biofilm formation leading to virulence

in the context of pathogenesis or symbiosis in the case of commensals/symbionts (330, 331). As

gut symbionts and their microbial products become increasingly attractive tools for human health

and animal husbandry, the *L. reuteri* aSec system may be exploited in the future for

biotechnological applications in the context of glycoengineering or for the development of novel

recombinant glycoprotein secretion strategies.

The second step of the aSec system pathway in *L. reuteri* involves the bi-functional LrAsp1-Asp2-

Asp3 complex which both modifies the SRRP through *O*-acetylation, but also chaperones it to

the translocation machinery. This step is highly conserved across gram-positive species with the

aSec system and involves a 1:1:1 heterotrimeric complex in *L. reuteri*, as also reported for

SgAsp1-Asp2-Asp3$_{M99}$ and SpAsp1-Asp2-Asp3$_{TIGR4}$ from pathogenic bacteria (216, 217). Here,

*Lr*Asp2 has been shown to be an acetyltransferase able to use both *p*NP-acetate and indoxylacetate as a donor substrate and this acetylesterase activity could be irreversibly inhibited by paraoxonethyl. The chaperoning capacity of the Asp1-Asp2-Asp3 complex has been reported previously through protein cross-linking experiments with *Sg*GspB$_{M99}$ (217, 218, 221). Here, we used ELISA-based binding assays and BLI, to determine binding kinetic parameters between the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex and the SRRP cargo, revealing high affinity interactions in the nanomolar range. Although no binding kinetics are available for other Asps characterised to date, these results are consistent with kinetic data available for chaperones of the Type III secretion systems in *Salmonella* sp. (317). *Lr*Asp1-Asp2-Asp3$_{53608}$ shared similar binding constants when tested with *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ *in vitro*, supporting the hypothesis that this step of the aSec pathway is conserved in *L. reuteri*.

The third step of the *L. reuteri* aSec pathway encompasses the processing of SRRP by the translocation machinery comprised of *Lr*SecA2, *Lr*SecY2 and *Lr*Asp4 at the cell membrane. The energy from active transport is mediated by the hydrolysis of ATP by *Lr*SecA2, as highlighted in other aSec systems (186, 189, 230). Here, we showed that GTP could also be hydrolysed by *Lr*SecA2 *in vitro* which had not been reported previously, albeit to a lesser kinetic rate, and the biological relevance of this finding remains to be investigated. Interestingly, the ATPase activity of the *Lr*SecA2 component of the *Lr*SecA2-SecY2-Asp4$_{53068}$ complex was enhanced in the presence of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex. Similarly, the acetylesterase activity of the *Lr*Asp2 component of the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex was enhanced in the presence of *Lr*SecA2-SecY2-Asp4$_{53068}$. Additionally, the binding avidity of the *Lr*SecA2-SecY2-Asp4$_{53068}$ to *Lr*SRRP$_{53608}$ and *Lr*SRRP$_{100-23}$ in the M^{-8} Kd range was stronger than the *Lr*Asp1-Asp2-Asp3$_{53608}$ complex for *Lr*SRRPs. Therefore, steps 2 and Step 3 may occur on the cytosolic side of the cell membrane, consistent with evidence of membrane association for aSec components in *S. gordonii* M99 (217). Our work also suggests that a higher level of interactions between the *Lr*SecA2-SecY2-Asp4 and *Lr*Asp1-Asp2-Asp3-Asp4$_{53068}$ complex may be facilitating post-translational modification as well as transport of *Lr*SRRP across the *L. reuteri* cell membrane.

However, despite successful purification and biophysical characterisation of *Lr*SecA2-SecY2-Asp4$_{53608}$ by SAXS analysis, the structural characteristics of the complex remain to be determined. The use of Alphafold2 for generating models of the monomers in this complex shows the limitations of *in silico* protein structure generation in capturing only a single state of the protein. Uncertainties related to structural forms for *Lr*Asp4 or dimerised *Lr*SecA2 were observed, affecting its fit within the SAXS-bead model. It is also difficult for artificial intelligence (AI) to predict other important aspects of protein structures such as metal ions, cofactors and ligands, although recent developments are focusing on molecular docking (Alphafill) (332). Furthermore, the predominant structural form of a dimerised SecA2 proteins remains elusive, as both parallel and anti-parallel conformations have been observed for SecA homologues. Here we predict that the structural assembly of the *L. reuteri* aSec translocation machinery is similar to the canonical SecA-SecY-SecE-SecG complex (182). The lack of Asp5 in *L. reuteri* ATCC 53608 and 100-23 strains, otherwise present in aSec systems from pathogenic gram-positive organisms such as *S. gordonii* M99, may suggest different oligomeric forms of the translocation machinery. These differences may account for the diversity observed in the overall structure or glycosylation of SRRPs across gram-positive bacteria, and/or changes in the composition of the aSec operon over its evolutionary history.

To gain further structural and functional insights into *L. reuteri* aSec system, high-resolution interactions between recombinant aSec proteins produced in this work including (1) *Lr*SRRP and *Lr*SecA2-SecY2-Asp4 (2) *Lr*SRRP and *Lr*Asp1-Asp2-Asp3 (3) *Lr*SRRs at different stages of glycosylation by *Lr*GTs, will be necessary to fully understand the molecular dynamics involved in transporting such a large glycoprotein through the aSec system. The binding kinetics reported here suggest that interactions between aSec components are strong enough to facilitate analysis through biophysical means such as cryoEM and NMR, which remain in liquid-phase and thus avoid some of the challenges in protein crystallisation observed throughout this work and the current limitation of *in silico* techniques like Alphafold2 to predict glycan structures. Preliminary work on cryoEM of *Lr*SRRPs is underway. In the future, structural data added into the protein

data bank will feed into modern machine learning programs to predict similar structures, and also identify sites that could be relevant to the field of pharmacological drug discovery for development of therapeutics against pathogenic gram-positive bacteria which use the aSec system as part of their virulence mechanisms.

Citations

1.   Gillois K, Lévêque M, Théodorou V, Robert H, Mercier-Bonin M. Mucus: An Underestimated Gut Target for Environmental Pollutants and Food Additives. Microorganisms. 2018;6(2):53.
2.   Karasov WH, Douglas AE. Comparative digestive physiology. Compr Physiol. 2013;3(2):741-83.
3.   Chivers DJ, Hladik CM. Morphology of the gastrointestinal tract in primates: comparisons with other mammals in relation to diet. J Morphol. 1980;166(3):337-86.
4.   Kararli TT. Comparison of the gastrointestinal anatomy, physiology, and biochemistry of humans and commonly used laboratory animals. Biopharm Drug Dispos. 1995;16(5):351-80.
5.   Hatton GB, Yadav V, Basit AW, Merchant HA. Animal Farm: Considerations in Animal Gastrointestinal Physiology and Relevance to Drug Delivery in Humans. J Pharml Sci. 2015;104(9):2747-76.
6.   Nguyen TLA, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? Dis Model Mech. 2015;8(1):1-16.
7.   Varel VH, Yen JT, Kreikemeier KK. Addition of cellulolytic clostridia to the bovine rumen and pig intestinal tract. Appl Environ Microbiol. 1995;61(3):1116-9.
8.   Johansson ME, Ambort D, Pelaseyed T, Schütte A, Gustafsson JK, Ermund A, *et al*. Composition and functional role of the mucus layers in the intestine. Cell Mol Life Sci. 2011;68(22):3635-41.
9.   Ermund A, Schütte A, Johansson ME, Gustafsson JK, Hansson GC. Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches. Am J Physiol Gastrointest Liver Physiol. 2013;305(5):G341-7.
10.   Johansson ME, Sjövall H, Hansson GC. The gastrointestinal mucus system in health and disease. Nat Rev Gastroenterol Hepatol. 2013;10(6):352-61.
11.   Varum FJ, McConnell EL, Sousa JJ, Veiga F, Basit AW. Mucoadhesion and the gastrointestinal tract. Crit Rev Ther Drug Carrier Syst. 2008;25(3):207-58.
12.   Varum FJ, Veiga F, Sousa JS, Basit AW. Mucus thickness in the gastrointestinal tract of laboratory animals. J Pharm Pharmacol. 2012;64(2):218-27.
13.   Atuma C, Strugala V, Allen A, Holm L. The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. Am J Physiol Gastrointest Liver Physiol. 2001;280(5):G922-9.
14.   Robbe C, Capon C, Coddeville B, Michalski JC. Structural diversity and specific distribution of *O*-glycans in normal human mucins along the intestinal tract. Biochem J. 2004;384(Pt 2):307-16.
15.   Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, *et al.* Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. J Biol Chem. 2003;278(47):46337-48.
16.   Johansson MEV, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc Natl Acad Sci. 2008;105(39):15064-9.
17.   Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, et al. The Antibacterial Lectin RegIIIγ; Promotes the Spatial Segregation of Microbiota and Host in the Intestine. Science. 2011;334(6053):255-8.
18.   Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, *et al*. Induction of Intestinal Th17 Cells by Segmented Filamentous Bacteria. Cell. 2009;139(3):485-98.
19.   Lee SM, Donaldson GP, Mikulski Z, Boyajian S, Ley K, Mazmanian SK. Bacterial colonization factors control specificity and stability of the gut microbiota. Nature. 2013;501(7467):426-9.

20.   Sonnenberg GF, Monticelli LA, Alenghat T, Fung TC, Hutnick NA, Kunisawa J, *et al.* Innate Lymphoid Cells Promote Anatomical Containment of Lymphoid-Resident Commensal Bacteria. Science. 2012;336(6086):1321-5.

21.   Pelaseyed T, Bergström JH, Gustafsson JK, Ermund A, Birchenough GMH, Schütte A, et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. Immunol Rev. 2014;260(1):8-20.

22.   Tropini C, Earle KA, Huang KC, Sonnenburg JL. The Gut Microbiome: Connecting Spatial Organization to Function. Cell Host & Microbe. 2017;21(4):433-42.

23.   Shetty SA, Hugenholtz F, Lahti L, Smidt H, de Vos WM. Intestinal microbiome landscaping: insight in community assemblage and implications for microbial modulation strategies. FEMS Microbiol Rev. 2017;41(2):182-99.

24.   Cani PD. Human gut microbiome: hopes, threats and promises. Gut. 2018;67(9):1716-25.

25.   Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. Nat Rev Microbiol. 2021;19(1):55-71.

26.   Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. Microb Ecol Health Dis. 2015;26:26191.

27.   Hugon P, Dufour JC, Colson P, Fournier PE, Sallah K, Raoult D. A comprehensive repertoire of prokaryotic species identified in human beings. Lancet Infect Dis. 2015;15(10):1211-9.

28.   Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J. 2017;474(11):1823-36.

29.   Neu AT, Allen EE, Roy K. Defining and quantifying the core microbiome: Challenges and prospects. Proc Natl Acad Sci. 2021;118(51):e2104429118.

30.   Risely A. Applying the core microbiome to understand host–microbe systems. J Anim Ecol. 2020;89(7):1549-58.

31.   Henry LP, Bruijning M, Forsberg SKG, Ayroles JF. The microbiome extends host evolutionary potential. Nat Commun. 2021;12(1):5141.

32.   Rosenberg E, Zilber-Rosenberg I. The hologenome concept of evolution after 10 years. Microbiome. 2018;6(1):78.

33.   Morris JJ. What is the hologenome concept of evolution? F1000Res. 2018;7.

34.   Lim SJ, Bordenstein SR. An introduction to phylosymbiosis. Proc. Bioll Sci. 2020;287(1922):20192900.

35.   Bajinka O, Tan Y, Abdelhalim KA, Özdemir G, Qiu X. Extrinsic factors influencing gut microbes, the immediate consequences and restoring eubiosis. AMB Express. 2020;10(1):130.

36.   Moran-Ramos S, Lopez-Contreras BE, Villarruel-Vazquez R, Ocampo-Medina E, Macias-Kauffer L, Martinez-Medina JN, *et al.* Environmental and intrinsic factors shaping gut microbiota composition and diversity and its relation to metabolic health in children and early adolescents: A population-based study. Gut Microbes. 2020;11(4):900-17.

37.   Asnicar F, Manara S, Zolfo M, Truong DT, Scholz M, Armanini F, *et al.* Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. mSystems. 2017;2(1):e00164-16.

38.   Xue C, Xie Q, Zhang C, Hu Y, Song X, Jia Y, *et al.* Vertical transmission of the gut microbiota influences glucose metabolism in offspring of mice with hyperglycaemia in pregnancy. Microbiome. 2022;10(1):122.

39.   Granger CL, Embleton ND, Palmer JM, Lamb CA, Berrington JE, Stewart CJ. Maternal breastmilk, infant gut microbiome and the impact on preterm infant health. Acta Paediatrica. 2021;110(2):450-7.

40.   Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, *et al.* Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe. 2016;19(5):731-43.

41.   Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, *et al.* A core gut microbiome in obese and lean twins. Nature. 2009;457(7228):480-4.

42.   Konstantinidis T, Tsigalou C, Karvelas A, Stavropoulou E, Voidarou C, Bezirtzoglou E. Effects of Antibiotics upon the Gut Microbiome: A Review of the Literature. Biomedicines. 2020;8(11).

43.   Leeming ER, Johnson AJ, Spector TD, Le Roy CI. Effect of Diet on the Gut Microbiota: Rethinking Intervention Duration. Nutrients. 2019;11(12).

44.   Singh RK, Chang HW, Yan D, Lee KM, Ucmak D, Wong K, *et al. I*nfluence of diet on the gut microbiome and implications for human health. J Transl Med. 2017;15(1):73.

45.   Leeuwendaal NK, Stanton C, O'Toole PW, Beresford TP. Fermented Foods, Health and the Gut Microbiome. Nutrients. 2022;14(7).

46.   Hemarajata P, Versalovic J. Effects of probiotics on gut microbiota: mechanisms of intestinal immunomodulation and neuromodulation. Therap Adv Gastroenterol. 2013;6(1):39-51.

47.   Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, et al. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nature Reviews Gastroenterology & Hepatology. 2014;11(8):506-14.

48.   Gupta VK, Paul S, Dutta C. Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity. Frontiers in Microbiology. 2017;8.

49.   Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, et al. Lifestyles in transition: evolution and natural history of the genus Lactobacillus. FEMS Microbiol Rev. 2017;41(Supp_1):S27-s48.

50.   Zheng J, Wittouck S, Salvetti E, Franz C, Harris HMB, Mattarelli P, *et al.* A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. Int J Syst Evol Microbiol. 2020;70(4):2782-858.

51.   Brooks SPJ, McAllister M, Sandoz M, Kalmokoff ML. Culture-independent phylogenetic analysis of the faecal flora of the rat. Can J Microbiol. 2003;49(10):589-601.

52.   Leser TD, Amenuvor JZ, Jensen TK, Lindecrona RH, Boye M, Moller K. Culture-independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. Appl Environ Microbiol. 2002;68(2):673-90.

53.   Molin G, Jeppsson B, Johansson ML, Ahrné S, Nobaek S, Ståhl M, *et al*. Numerical taxonomy of *Lactobacillus spp*. associated with healthy and diseased mucosa of the human intestines. J Appl Bacteriol. 1993;74(3):314-23.

54.   Walter J. Ecological role of lactobacilli in the gastrointestinal tract: implications for fundamental and biomedical research. Appl Environ Microbiol. 2008;74(16):4985-96.

55.   Blaser MJ, Falkow S. What are the consequences of the disappearing human microbiota? Nat Rev Microbiol. 2009;7:887.

56.   Rackaityte E, Lynch SV. The human microbiome in the 21st century. Nat Commun. 2020;11(1):5256.

57.   Ryan KA, Jayaraman T, Daly P, Canchaya C, Curran S, Fang F, *et al*. Isolation of *lactobacilli* with probiotic properties from the human stomach. Lett Appl Microbiol. 2008;47(4):269-74.

58.   Walter J, Britton RA, Roos S. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proc Natl Acad Sci U S A. 2011;108 Suppl 1(Suppl 1):4645-52.

59.   Sinkiewicz G, Ljunggren L. Occurrence of *Lactobacillus reuteri* in human breast milk. Microb Ecol Health Dis. 2008;20(3):122-6.

60.   Sinkiewicz G, Nordström EA. 353 Occurrence of *Lactobacillus reuteri, Lactobacilli* and *Bifidobacteria* in Human Breast Milk. Pediatr Res. 2005;58(2):415-.

61.   Frese SA, Benson AK, Tannock GW, Loach DM, Kim J, Zhang M, *et al*. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. PLoS Genet. 2011;7(2):e1001314.

62.   Fuller R, Brooker BE. *Lactobacilli* which attach to the crop epithelium of the fowl. The Am J Clin Nutr. 1974;27(11):1305-12.

63.   Fuller R, Barrow PA, Brooker BE. Bacteria associated with the gastric epithelium of neonatal pigs. Appl Environ Microbiol. 1978;35(3):582-91.
64.   Walsham AD, MacKenzie DA, Cook V, Wemyss-Holden S, Hews CL, Juge N, et al. Lactobacillus reuteri Inhibition of Enteropathogenic *Escherichia coli* Adherence to Human Intestinal Epithelium. Front Microbiol. 2016;7:244.
65.   Walter J, Britton RA, Roos S. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. Proceedings of the National Academy of Sciences. 2011;108(supplement_1):4645-52.
66.   Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, *et al*. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. Isme j. 2010;4(3):377-87.
67.   Frese S, MacKenzie D, Peterson D, Schmaltz R, Fangman T, Zhou Y, *et al*. Molecular Characterization of Host-Specific Biofilm Formation in a Vertebrate Gut Symbiont. PLoS genet. 2013;9:e1004057.
68.   Oh PL, Benson AK, Peterson DA, Patil PB, Moriyama EN, Roos S, et al. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. Isme J. 2009;4:377.
69.   Carbajal N, Casas I, Dobrogosz W. Effect of host-specific *Lactobacillus reuteri* on ileal tissue development in gnotobiotic BALB/c mice. Microbial Ecol Health Dis. 1999;11:184.
70.   Molin G, Andersson R, Ahrne S, Lonner C, Marklinder I, Johansson ML, *et al*. Effect of fermented oatmeal soup on the cholesterol level and the *Lactobacillus* colonization of rat intestinal mucosa. Antonie van Leeuwenhoek. 1992;61(3):167-73.
71.   Duar RM, Frese SA, Lin XB, Fernando SC, Burkey TE, Tasseva G, *et al*. Experimental Evaluation of Host Adaptation of *Lactobacillus reuteri* to Different Vertebrate Species. Appl Environ Microbiol. 2017;83(12).
72.   Lin XB, Wang T, Stothard P, Corander J, Wang J, Baines JF, *et al*. The evolution of ecological facilitation within mixed-species biofilms in the mouse gastrointestinal tract. The ISME Journal. 2018;12(11):2770-84.
73.   Teame T, Wang A, Xie M, Zhang Z, Yang Y, Ding Q, *et al*. Paraprobiotics and Postbiotics of Probiotic Lactobacilli, Their Positive Effects on the Host and Action Mechanisms: A Review. Front Nutr. 2020;7.
74.   MacKenzie DA, Tailford LE, Hemmings AM, Juge N. Crystal structure of a mucus-binding protein repeat reveals an unexpected functional immunoglobulin binding activity. J Biol Chem. 2009;284(47):32444-53.
75.   Jensen H, Roos S, Jonsson H, Rud I, Grimmer S, van Pijkeren JP, *et al*. Role of Lactobacillus reuteri cell and mucus-binding protein A (CmbA) in adhesion to intestinal epithelial cells and mucus *in vitro*. Microbiology (Reading). 2014;160(Pt 4):671-81.
76.   Etzold S, MacKenzie DA, Jeffers F, Walshaw J, Roos S, Hemmings AM, *et al*. Structural and molecular insights into novel surface-exposed mucus adhesins from *Lactobacillus reuteri* human strains. Mol Microbiol. 2014;92(3):543-56.
77.   MacKenzie DA, Jeffers F, Parker ML, Vibert-Vallet A, Bongaerts RJ, Roos S, et al. Strain-specific diversity of mucus-binding proteins in the adhesion and aggregation properties of *Lactobacillus reuteri*. Microbiology (Reading). 2010;156(Pt 11):3368-78.
78.   Gunning AP, Kavanaugh D, Thursby E, Etzold S, MacKenzie DA, Juge N. Use of Atomic Force Microscopy to Study the Multi-Modular Interaction of Bacterial Adhesins to Mucins. Int J Mol Sci. 2016;17(11).
79.   Ton-That H, Marraffini LA, Schneewind O. Protein sorting to the cell wall envelope of Gram-positive bacteria. Biochim Biophys Acta. 2004;1694(1-3):269-78.
80.   Siegel SD, Reardon ME, Ton-That H. Anchoring of LPXTG-Like Proteins to the Gram-Positive Cell Wall Envelope. Curr Top Microbiol Immunol. 2017;404:159-75.
81.   Doleyres Y, Beck P, Vollenweider S, Lacroix C. Production of 3-hydroxypropionaldehyde using a two-step process with Lactobacillus reuteri. Appl Microbiol Biotechnol. 2005;68(4):467-74.

82.   Oh J-H, Lin XB, Zhang S, Tollenaar SL, Özçam M, Dunphy C, *et al*. Prophages in *Lactobacillus reuteri* Are Associated with Fitness Trade-Offs but Can Increase Competitiveness in the Gut Ecosystem. Appl EnvMicrobiol. 2019;86(1):e01922-19.

83.   Cleusix V, Lacroix C, Vollenweider S, Duboux M, Le Blay G. Inhibitory activity spectrum of reuterin produced by *Lactobacillus reuteri* against intestinal bacteria. BMC Microbiology. 2007;7(1):101.

84.   Spinler JK, Taweechotipatr M, Rognerud CL, Ou CN, Tumwasorn S, Versalovic J. Human-derived probiotic *Lactobacillus reuteri* demonstrate antimicrobial activities targeting diverse enteric bacterial pathogens. Anaerobe. 2008;14(3):166-71.

85.   Cheng CC, Duar RM, Lin X, Perez-Munoz ME, Tollenaar S, Oh J-H, *et al*. Ecological Importance of Cross-Feeding of the Intermediate Metabolite 1,2-Propanediol between Bacterial Gut Symbionts. Applied and Environmental Microbiology. 2020;86(11):e00190-20.

86.   McGrath CJ, Laveckis E, Bell A, Crost E, Juge N, Schüller S. Development of a novel human intestinal model to elucidate the effect of anaerobic commensals on *Escherichia coli* infection. Dis Model Mech. 2022;15(4).

87.   Fairbrother JM, Nadeau É, Gyles CL. *Escherichia coli* in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. Anim Health Res Rev. 2005;6(1):17-39.

88.   Francis DH. Enterotoxigenic *Escherichia coli* infection in pigs and its diagnosis. J Swine Health Prod. 2002;10(4):171-5.

89.   Hazards EPanel oB, Koutsoumanis K, Allende A, Alvarez-Ordóñez A, Bolton D, Bover-Cid S, *et al.* Update of the list of QPS-recommended biological agents intentionally added to food or feed as notified to EFSA 11: suitability of taxonomic units notified to EFSA until September 2019. EFSA J. 2020;18(2):e05965.

90.   Szymanski CM. Bacterial glycosylation, it's complicated. Front Mol Biosci. 2022;9.

91.   Schäffer C, Messner P. Emerging facets of prokaryotic glycosylation. FEMS MicrobiolRev. 2017;41(1):49-91.

92.   Li H, Debowski AW, Liao T, Tang H, Nilsson HO, Marshall BJ, *et al*. Understanding protein glycosylation pathways in bacteria. Future Microbiol. 2017;12:59-72.

93.   Nothaft H, Scott NE, Vinogradov E, Liu X, Hu R, Beadle B, et al. Diversity in the Protein *N*-Glycosylation Pathways Within the Campylobacter Genus*. Mol Cell Proteomics. 2012;11(11):1203-19.

94.   Daubenspeck JM, Jordan DS, Simmons W, Renfrow MB, Dybvig K. General *N*-and *O*-Linked Glycosylation of Lipoproteins in Mycoplasmas and Role of Exogenous Oligosaccharide. PLoS One. 2015;10(11):e0143362.

95.   Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, *et al*. *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science. 2002;298(5599):1790-3.

96.   Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, *et al*. Definition of the bacterial *N*-glycosylation site consensus sequence. EMBO J. 2006;25(9):1957-66.

97.   Weerapana E, Glover KJ, Chen MM, Imperiali B. Investigating bacterial *N*-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. J Am Chem Soc. 2005;127(40):13766-7.

98.   Ollis AA, Chai Y, Natarajan A, Perregaux E, Jaroentomeechai T, Guarino C, *et al*. Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique *N*-glycosylation site preferences. Sci Rep. 2015;5(1):15237.

99.   Nothaft H, Szymanski CM. Bacterial protein *N*-glycosylation: new perspectives and applications. J Biol Chem. 2013;288(10):6912-20.

100.   Schäffer C, Messner P. Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. Glycobiology. 2004;14(8):31R-42R.

101.   Gault J, Ferber M, Machata S, Imhaus A-F, Malosse C, Charles-Orszag A, *et al*. *Neisseria meningitidis* Type IV Pili Composed of Sequence Invariable Pilins Are Masked by Multisite Glycosylation. PLOS Pathog. 2015;11(9):e1005162.

102.   Aas FE, Vik A, Vedde J, Koomey M, Egge-Jacobsen W. *Neisseria gonorrhoeae O-*linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. Mol Microbiol. 2007;65(3):607-24.

103.   Smedley JG, 3rd, Jewell E, Roguskie J, Horzempa J, Syboldt A, Stolz DB, *et al.* Influence of pilin glycosylation on *Pseudomonas aeruginosa* 1244 pilus function. Infect Immun. 2005;73(12):7922-31.

104.   Howlett R, Read N, Varghese A, Kershaw C, Hancock Y, Smith MCM. *Streptomyces coelicolor* strains lacking polyprenol phosphate mannose synthase and protein *O*-mannosyl transferase are hyper-susceptible to multiple antibiotics. Microbiology (Reading). 2018;164(3):369-82.

105.   Liu CF, Tonini L, Malaga W, Beau M, Stella A, Bouyssié D, *et al.* Bacterial protein-*O*-mannosylating enzyme is crucial for virulence of *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A. 2013;110(16):6560-5.

106.   Mahne M, Tauch A, Pühler A, Kalinowski J. The *Corynebacterium glutamicum* gene pmt encoding a glycosyltransferase related to eukaryotic protein-*O*-mannosyltransferases is essential for glycosylation of the resuscitation promoting factor (Rpf2) and other secreted proteins. FEMS Microbiol Lett. 2006;259(2):226-33.

107.   Iwashkiw JA, Vozza NF, Kinsella RL, Feldman MF. Pour some sugar on it: the expanding world of bacterial protein *O*-linked glycosylation. Mol Microbiol. 2013;89(1):14-28.

108.   Latousakis D, MacKenzie DA, Telatin A, Juge N. Serine-rich repeat proteins from gut microbes. Gut Microbes. 2020;11(1):102-17.

109.   Green ER, Mecsas J. Bacterial Secretion Systems: An Overview. Microbiol Spectr. 2016;4(1).

110.   Palmer T, Finney AJ, Saha CK, Atkinson GC, Sargent F. A holin/peptidoglycan hydrolase-dependent protein secretion system. Mol Microbiol. 2021;115(3):345-55.

111.   Rivera-Calzada A, Famelis N, Llorca O, Geibel S. Type VII secretion systems: structure, functions and transport models. Nat Rev Microbiol. 2021;19(9):567-84.

112.   Spencer BL, Tak U, Mendonça JC, Nagao PE, Niederweis M, Doran KS. A type VII secretion system in Group B *Streptococcus* mediates cytotoxicity and virulence. PLoS Pathog. 2021;17(12):e1010121.

113.   Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect Biol. 2010;2(5):a000414.

114.   Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E. Biogenesis, Architecture, and Function of Bacterial Type IV Secretion Systems. Annu Rev of Microbiol. 2005;59(1):451-85.

115.   Naskar S, Hohl M, Tassinari M, Low HH. The structure and mechanism of the bacterial type II secretion system. Mol Microbiol. 2021;115(3):412-24.

116.   Papanikou E, Karamanou S, Economou A. Bacterial protein secretion through the translocase nanomachine. Nat Rev Microbiol. 2007;5(11):839-51.

117.   Palmer T. Structural biology: Mycobacterial ESX secrets revealed. Nat Microbiol. 2017;2(6):17074.

118.   von Heijne G. The structure of signal peptides from bacterial lipoproteins. Protein Eng. 1989;2(7):531-4.

119.   Luirink J, Sinning I. SRP-mediated protein targeting: structure and function revisited. Biochim Biophys Acta. 2004;1694(1-3):17-35.

120.   Randall L, Hardy S. SecB, one small chaperone in the complex milieu of the cell. Cell Mol Life Sci. 2002;59(10):1617-23.

121.   Oswald J, Njenga R, Natriashvili A, Sarmah P, Koch HG. The Dynamic SecYEG Translocon. Front Mol Biosci. 2021;8:664241.

122.   Zhu Z, Wang S, Shan SO. Ribosome profiling reveals multiple roles of SecA in cotranslational protein export. Nat Commun. 2022;13(1):3393.

123.   Sala A, Bordes P, Genevaux P. Multitasking SecB chaperones in bacteria. Front Microbiol. 2014;5.

124.   Linde D, Volkmer-Engert R, Schreiber S, Müller JP. Interaction of the *Bacillus subtilis* chaperone CsaA with the secretory protein YvaY. FEMS Microbiol Lett. 2003;226(1):93-100.

125. Randall LL. Translocation of domains of nascent periplasmic proteins across the cytoplasmic membrane is independent of elongation. Cell. 1983;33(1):231-40.
126. Hartl F-U, Lecker S, Schiebel E, Hendrick JP, Wickner W. The binding cascade of SecB to SecA to SecYE mediates preprotein targeting to the *E. coli* plasma membrane. Cell. 1990;63(2):269-79.
127. Lycklama ANJA, Driessen AJ. The bacterial Sec-translocase: structure and mechanism. Philos Trans R Soc Lond B Biol Sci. 2012;367(1592):1016-28.
128. Schulze RJ, Komar J, Botte M, Allen WJ, Whitehouse S, Gold VAM, *et al*. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG–SecDF–YajC–YidC. Proc Natl Acad Sci U S A. 2014;111(13):4844-9.
129. Nouwen N, Piwowarek M, Berrelkamp G, Driessen AJM. The Large First Periplasmic Loop of SecD and SecF Plays an Important Role in SecDF Functioning. J Bacteriol. 2005;187(16):5857-60.
130. Hou JM, D'Lima NG, Rigel NW, Gibbons HS, McCann JR, Braunstein M, *et al.* ATPase activity of *Mycobacterium tuberculosis* SecA1 and SecA2 proteins and its importance for SecA2 function in macrophages. J Bacteriol. 2008;190(14):4880-7.
131. Karamanou S, Gouridis G, Papanikou E, Sianidis G, Gelis I, Keramisanou D, *et al*. Preprotein-controlled catalysis in the helicase motor of SecA. EMBO J. 2007;26(12):2904-14.
132. Banerjee T, Lindenthal C, Oliver D. SecA functions *in vivo* as a discrete anti-parallel dimer to promote protein transport. Mol Microbiol. 2017;103(3):439-51.
133. Allen WJ, Corey RA, Oatley P, Sessions RB, Baldwin SA, Radford SE, *et al.* Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. eLife. 2016;5:e15598.
134. Gelis I, Bonvin AMJJ, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, *et al.* Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. Cell. 2007;131(4):756-69.
135. Osborne AR, Clemons WM, Rapoport TA. A large conformational change of the translocation ATPase SecA. Proc Natl Acad Sci U S A. 2004;101(30):10937-42.
136. Osborne AR, Clemons WM, Jr., Rapoport TA. A large conformational change of the translocation ATPase SecA. Proc Natl Acad Sci U S A. 2004;101(30):10937-42.
137. Swanson S, Ioerger TR, Rigel NW, Miller BK, Braunstein M, Sacchettini JC. Structural Similarities and Differences between Two Functionally Distinct SecA Proteins, *Mycobacterium tuberculosis* SecA1 and SecA2. J Bacteriol. 2016;198(4):720-30.
138. Duong F, Wickner W. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. EMBO J. 1997;16(10):2756-68.
139. Tanaka Y, Sugano Y, Takemoto M, Mori T, Furukawa A, Kusakizako T, *et al*. Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. Cell Rep. 2015;13(8):1561-8.
140. Thomas JR, Bolhuis A. The *tatC* gene cluster is essential for viability in halophilic archaea. FEMS Microbiol Lett. 2006;256(1):44-9.
141. Alcock F, Stansfeld PJ, Basit H, Habersetzer J, Baker MA, Palmer T, *et al.* Assembling the Tat protein translocase. Elife. 2016;5.
142. Lee PA, Tullman-Ercek D, Georgiou G. The bacterial twin-arginine translocation pathway. Annu Rev Microbiol. 2006;60:373-95.
143. Palmer T, Berks BC. The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol. 2012;10(7):483-96.
144. Frain KM, Robinson C, van Dijl JM. Transport of Folded Proteins by the Tat System. Protein J. 2019;38(4):377-88.
145. Gohlke U, Pullan L, McDevitt CA, Porcelli I, de Leeuw E, Palmer T, *et al*. The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. Proc Natl Acad Sci U S A. 2005;102(30):10482-6.
146. Zhang Y, Wang L, Hu Y, Jin C. Solution structure of the TatB component of the twin-arginine translocation system. Biochim Biophys Acta. 2014;1838(7):1881-8.
147. Rollauer SE, Tarry MJ, Graham JE, Jääskeläinen M, Jäger F, Johnson S, *et al.* Structure of the TatC core of the twin-arginine protein transport system. Nature. 2012;492(7428):210-4.

148.   Jongbloed JD, Martin U, Antelmann H, Hecker M, Tjalsma H, Venema G, *et al.* TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway. J Biol Chem. 2000;275(52):41350-7.

149.   Biswas L, Biswas R, Nerz C, Ohlsen K, Schlag M, Schäfer T, *et al.* Role of the twin-arginine translocation pathway in *Staphylococcus*. J Bacteriol. 2009;191(19):5921-9.

150.   Delepelaire P. Type I secretion in gram-negative bacteria. Biochim Biophys Acta. 2004;1694(1-3):149-61.

151.   Thomas S, Holland IB, Schmitt L. The Type 1 secretion pathway - the hemolysin system and beyond. Biochim Biophys Acta. 2014;1843(8):1629-41.

152.   Yan Z, Yin M, Xu D, Zhu Y, Li X. Structural insights into the secretin translocation channel in the type II secretion system. Nat Struct Mol Biol. 2017;24(2):177-83.

153.   Korotkov KV, Sandkvist M, Hol WGJ. The type II secretion system: biogenesis, molecular architecture and mechanism. Nat Rev Microbiol. 2012;10(5):336-51.

154.   Coburn B, Sekirov I, Finlay BB. Type III secretion systems and disease. Clin Microbiol Rev. 2007;20(4):535-49.

155.   Lombardi C, Tolchard J, Bouillot S, Signor L, Gebus C, Liebl D, *et al*. Structural and Functional Characterization of the Type Three Secretion System (T3SS) Needle of *Pseudomonas aeruginosa*. Front Microbiol. 2019;10:573.

156.   Meir A, Macé K, Lukoyanova N, Chetrit D, Hospenthal MK, Redzej A, *et al.* Mechanism of effector capture and delivery by the type IV secretion system from *Legionella pneumophila.* Nat Commun. 2020;11(1):2864.

157.   Fan E, Chauhan N, Udatha DBRKG, Leo JC, Linke D. Type V Secretion Systems in Bacteria. Microbiol Spect. 2016;4(1):4.1.10.

158.   Meuskens I, Saragliadis A, Leo JC, Linke D. Type V Secretion Systems: An Overview of Passenger Domain Functions. Front Microbiol. 2019;10:1163.

159.   Kang'ethe W, Bernstein HD. Charge-dependent secretion of an intrinsically disordered protein via the autotransporter pathway. Proc Natl Acad Sci. 2013;110(45):E4246-E55.

160.   Hoiczyk E, Roggenkamp A, Reichenbecher M, Lupas A, Heesemann J. Structure and sequence analysis of Yersinia YadA and Moraxella UspAs reveal a novel class of adhesins. The EMBO J. 2000;19(22):5989-99.

161.   Leo Jack C, Lyskowski A, Hattula K, Hartmann Marcus D, Schwarz H, Butcher Sarah J, *et al*. The Structure of *E. coli* IgG-Binding Protein D Suggests a General Model for Bending and Binding in Trimeric Autotransporter Adhesins. Structure. 2011;19(7):1021-30.

162.   Leo JC, Linke D. A unified model for BAM function that takes into account type Vc secretion and species differences in BAM composition. AIMS Microbiol. 2018;4(3):455-68.

163.   Russell AB, Hood RD, Bui NK, LeRoux M, Vollmer W, Mougous JD. Type VI secretion delivers bacteriolytic effectors to target cells. Nature. 2011;475(7356):343-7.

164.   Santin YG, Cascales E. Domestication of a housekeeping transglycosylase for assembly of a Type VI secretion system. EMBO rep. 2017;18(1):138-49.

165.   Zoued A, Durand E, Brunet YR, Spinelli S, Douzi B, Guzzo M, *et al.* Priming and polymerization of a bacterial contractile tail structure. Nature. 2016;531(7592):59-63.

166.   Schneider JP, Nazarov S, Adaixo R, Liuzzo M, Ringel PD, Stahlberg H, *et al.* Diverse roles of TssA-like proteins in the assembly of bacterial type VI secretion systems. EMBO J. 2019;38(18):e100825.

167.   Cianfanelli FR, Monlezun L, Coulthurst SJ. Aim, load, fire: the type VI secretion system, a bacterial nanoweapon. Trends Microbiol. 2016;24(1):51-62.

168.   Gallique M, Bouteiller M, Merieau A. The Type VI Secretion System: A Dynamic System for Bacterial Communication? Front Microbiol. 2017;8.

169.   Beckham KSH, Ciccarelli L, Bunduc CM, Mertens HDT, Ummels R, Lugmayr W, *et al.* Structure of the mycobacterial ESX-5 type VII secretion system membrane complex by single-particle analysis. Nat Microbiol. 2017;2(6):17047.

170.   Beckham KSH, Ritter C, Chojnowski G, Ziemianowicz DS, Mullapudi E, Rettel M, *et al*. Structure of the mycobacterial ESX-5 type VII secretion system pore complex. Sci Adv. 2021;7(26).

171.   Bunduc CM, Fahrenkamp D, Wald J, Ummels R, Bitter W, Houben ENG, *et al.* Structure and dynamics of a mycobacterial type VII secretion system. Nature. 2021;593(7859):445-8.

172.   Crosskey TD, Beckham KSH, Wilmanns M. The ATPases of the mycobacterial type VII secretion system: Structural and mechanistic insights into secretion. Prog Biophys Mol Biol. 2020;152:25-34.

173.   Bowman L, Palmer T. The Type VII Secretion System of *Staphylococcus*. Annu Rev Microbiol. 2021;75:471-94.

174.   Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, et al. Type VII secretion--mycobacteria show the way. Nat Rev Microbiol. 2007;5(11):883-91.

175.   Barnhart MM, Chapman MR. Curli biogenesis and function. Annu Rev Microbiol. 2006;60:131-47.

176.   Bhoite S, van Gerven N, Chapman MR, Remaut H. Curli Biogenesis: Bacterial Amyloid Assembly by the Type VIII Secretion Pathway. EcoSal Plus. 2019;8(2).

177.   Van Gerven N, Klein RD, Hultgren SJ, Remaut H. Bacterial amyloid formation: structural insights into curli biogensis. Trends Microbiol. 2015;23(11):693-706.

178.   Sato K, Sakai E, Veith PD, Shoji M, Kikuchi Y, Yukitake H, *et al.* Identification of a new membrane-associated protein that influences transport/maturation of gingipains and adhesins of *Porphyromonas gingivalis*. J Biol Chem. 2005;280(10):8668-77.

179.   Lasica AM, Ksiazek M, Madej M, Potempa J. The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. Front Cell Infect Microbiol. 2017;7:215.

180.   Gorasia Dhana G, Lunar Silva I, Butler Catherine A, Chabalier M, Doan T, Cascales E, *et al.* Protein Interactome Analysis of the Type IX Secretion System Identifies PorW as the Missing Link between the PorK/N Ring Complex and the Sov Translocon. Microbiol Spectr. 2022;10(1):e01602-21.

181.   Geiger T, Pazos M, Lara-Tejero M, Vollmer W, Galán JE. Peptidoglycan editing by a specific LD-transpeptidase controls the muramidase-dependent secretion of typhoid toxin. Nat Microbiol. 2018;3(11):1243-54.

182.   Rigel NW, Braunstein M. A new twist on an old pathway – accessory secretion systems. Mol Microbiol. 2008;69(2):291-302.

183.   Feltcher ME, Gibbons HS, Ligon LS, Braunstein M. Protein export by the mycobacterial SecA2 system is determined by the preprotein mature domain. J Bacteriol. 2013;195(4):672-81.

184.   Braunstein M, Brown AM, Kurtz S, Jacobs WR, Jr. Two nonredundant SecA homologues function in mycobacteria. J Bacteriol. 2001;183(24):6979-90.

185.   Braunstein M, Bensing BA, Sullam PM. The Two Distinct Types of SecA2-Dependent Export Systems. Microbiol Spectr. 2019;7(3).

186.   Bensing BA, Seepersaud R, Yen YT, Sullam PM. Selective transport by SecA2: An expanding family of customized motor proteins. Biochim Biophys Acta. 2014;1843(8):1674-86.

187.   Kurtz S, McKinnon KP, Runge MS, Ting JP, Braunstein M. The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. Infect Immun. 2006;74(12):6855-64.

188.   Lindič N, Loboda J, Usenik A, Vidmar R, Turk D. The Structure of *Clostridioides difficile* SecA2 ATPase Exposes Regions Responsible for Differential Target Recognition of the SecA1 and SecA2-Dependent Systems. Int J Mol Sci. 2020;21(17).

189.   Bensing BA, Sullam PM. Characterization of *Streptococcus gordonii* SecA2 as a paralogue of SecA. J Bacteriol. 2009;191(11):3482-91.

190.   Bensing BA, Yen YT, Seepersaud R, Sullam PM. A Specific Interaction between SecA2 and a Region of the Preprotein Adjacent to the Signal Peptide Occurs during Transport via the Accessory Sec System*. J Biol Chem. 2012;287(29):24438-47.

191.   Sibbald MJ, Winter T, van der Kooi-Pol MM, Buist G, Tsompanidou E, Bosma T, *et al.* Synthetic effects of *secG* and *secY2* mutations on exoproteome biogenesis in *Staphylococcus aureus*. J Bacteriol. 2010;192(14):3788-800.

192.   Bandara M, Corey RA, Martin R, Skehel JM, Blocker AJ, Jenkinson HF, *et al*. Composition and Activity of the Non-canonical Gram-positive SecY2 Complex. J Biol Chem. 2016;291(41):21474-84.

193.   Chan JM, Gori A, Nobbs AH, Heyderman RS. *Streptococcal* Serine-Rich Repeat Proteins in Colonization and Disease. Front Microbiol. 2020;11.

194.   Spencer C, Bensing BA, Mishra NN, Sullam PM. Membrane trafficking of the bacterial adhesin GspB and the accessory Sec transport machinery. J Biol Chem. 2019;294(5):1502-15.

195.   Latousakis D, Juge N. How Sweet Are Our Gut Beneficial Bacteria? A Focus on Protein Glycosylation in *Lactobacillus*. Int J Mol Sci. 2018;19(1):136.

196.   Sanchez CJ, Shivshankar P, Stol K, Trakhtenbroit S, Sullam PM, Sauer K, *et al*. The *Pneumococcal* Serine-Rich Repeat Protein Is an Intra-Species Bacterial Adhesin That Promotes Bacterial Aggregation *In Vivo* and in Biofilms. PLOS Pathog. 2010;6(8):e1001044.

197.   Bensing BA, Sullam PM. An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. Mol Microbiol. 2002;44(4):1081-94.

198.   Siboo IR, Chambers HF, Sullam PM. Role of SraP, a Serine-Rich Surface Protein of *Staphylococcus aureus,* in binding to human platelets. Infect Immun. 2005;73(4):2273-80.

199.   Garnett JA, Simpson PJ, Taylor J, Benjamin SV, Tagliaferri C, Cota E, *et al*. Structural insight into the role of *Streptococcus parasanguinis* Fap1 within oral biofilm formation. Biochem Biophys Res Commun. 2012;417(1):421-6.

200.   Wu H, Mintz KP, Ladha M, Fives-Taylor PM. Isolation and characterization of Fap1, a fimbriae-associated adhesin of *Streptococcus parasanguis* FW213. MoMicrobiol. 1998;28(3):487-500.

201.   Froeliger EH, Fives-Taylor P. *Streptococcus parasanguis* Fimbria-Associated Adhesin Fap1 Is Required for Biofilm Formation. Infect Immun. 2001;69(4):2512-9.

202.   Chahal G, Quintana-Hayashi MP, Gaytán MO, Benktander J, Padra M, King SJ, *et al*. *Streptococcus oralis* Employs Multiple Mechanisms of Salivary Mucin Binding That Differ Between Strains. Front Cell Infect Microbiol. 2022;12.

203.   Ronis A, Brockman K, Singh AK, Gaytán MO, Wong A, McGrath S, *et al*. *Streptococcus oralis* subsp. *dentisani* Produces Monolateral Serine-Rich Repeat Protein Fibrils, One of Which Contributes to Saliva Binding via Sialic Acid. Infect Immun. 2019;87(10).

204.   Takamatsu D, Bensing BA, Prakobphol A, Fisher SJ, Sullam PM. Binding of the streptococcal surface glycoproteins GspB and Hsa to human salivary proteins. Infect Immun. 2006;74(3):1933-40.

205.   Takamatsu D, Bensing BA, Cheng H, Jarvis GA, Siboo IR, López JA, *et al*. Binding of the *Streptococcus gordonii* surface glycoproteins GspB and Hsa to specific carbohydrate structures on platelet membrane glycoprotein Ibalpha. Mol Microbiol. 2005;58(2):380-92.

206.   Bensing BA, Khedri Z, Deng L, Yu H, Prakobphol A, Fisher SJ, *et al*. Novel aspects of sialoglycan recognition by the Siglec-like domains of *streptococcal* SRR glycoproteins. Glycobiol. 2016;26(11):1222-34.

207.   Frese SA, MacKenzie DA, Peterson DA, Schmaltz R, Fangman T, Zhou Y, *et al*. Molecular Characterization of Host-Specific Biofilm Formation in a Vertebrate Gut Symbiont. PLOS Genet. 2013;9(12):e1004057.

208.   Sequeira S, Kavanaugh D, MacKenzie DA, Šuligoj T, Walpole S, Leclaire C, *et al*. Structural basis for the role of serine-rich repeat proteins from *Lactobacillus reuteri* in gut microbe-host interactions. Proc Natl Acad Sci U S A. 2018;115(12):E2706-e15.

209.   Shivshankar P, Sanchez C, Rose LF, Orihuela CJ. The *Streptococcus pneumoniae* adhesin PsrP binds to Keratin 10 on lung cells. Mol Microbiol. 2009;73(4):663-79.

210.   Seo HS, Mu R, Kim BJ, Doran KS, Sullam PM. Binding of glycoprotein Srr1 of *Streptococcus agalactiae* to fibrinogen promotes attachment to brain endothelium and the development of meningitis. PLoS Pathog. 2012;8(10):e1002947.

211.   Wang NY, Patras KA, Seo HS, Cavaco CK, Rösler B, Neely MN, et al. Group B streptococcal serine-rich repeat proteins promote interaction with fibrinogen and vaginal colonization. J Infect Dis. 2014;210(6):982-91.

212.   Six A, Bellais S, Bouaboud A, Fouet A, Gabriel C, Tazi A, et al. Srr2, a multifaceted adhesin expressed by ST-17 hypervirulent Group B *Streptococcus* involved in binding to both fibrinogen and plasminogen. Mol Microbiol. 2015;97(6):1209-22.

213.   Zhang Y, Lu P, Pan Z, Zhu Y, Ma J, Zhong X, et al. SssP1, a *Streptococcus suis* Fimbria-Like Protein Transported by the SecY2/A2 System, Contributes to Bacterial Virulence. Appl Environ Microbiol. 2018;84(18).

214.   Lancy P, Jr., Appelbaum B, Holt SC, Rosan B. Quantitative *in vitro* assay for "corncob" formation. Infect Immun. 1980;29(2):663-70.

215.   Couvigny B, Lapaque N, Rigottier-Gois L, Guillot A, Chat S, Meylheuc T, *et al.* Three glycosylated serine-rich repeat proteins play a pivotal role in adhesion and colonization of the pioneer commensal bacterium, *Streptococcus salivarius*. Environ Microbiol. 2017;19(9):3579-94.

216.   Guo C, Feng Z, Zuo G, Jiang Y-L, Zhou C-Z, Chen Y, *et al.* Structural and functional insights into the Asp1/2/3 complex mediated secretion of pneumococcal serine-rich repeat protein PsrP. Biochem Biophys Res Comm. 2020;524(3):784-90.

217.   Chen Y, Bensing BA, Seepersaud R, Mi W, Liao M, Jeffrey PD, *et al*. Unraveling the sequence of cytosolic reactions in the export of GspB adhesin from *Streptococcus gordonii*. Journal of Biological Chemistry. 2018;293(14):5360-73.

218.   Seepersaud R, Bensing BA, Yen YT, Sullam PM. Asp3 mediates multiple protein–protein interactions within the accessory Sec system of *Streptococcus gordonii*. Mol Microbiol. 2010;78(2):490-505.

219.   Chen Y, Bensing BA, Seepersaud R, Mi W, Liao M, Jeffrey PD, *et al.* Unraveling the sequence of cytosolic reactions in the export of GspB adhesin from *Streptococcus gordonii*. J Biol Chem. 2018;293(14):5360-73.

220.   Nagae M, Yamaguchi Y. Three-dimensional structural aspects of protein–polysaccharide interactions. International journal of molecular sciences. 2014;15(3):3768-83.

221.   Seepersaud R, Sychantha D, Bensing BA, Clarke AJ, Sullam PM. *O*-acetylation of the serine-rich repeat glycoprotein GspB is coordinated with accessory Sec transport. PLOS Pathog. 2017;13(8):e1006558.

222.   Mistou MY, Dramsi S, Brega S, Poyart C, Trieu-Cuot P. Molecular dissection of the secA2 locus of group B Streptococcus reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. J Bacteriol. 2009;191(13):4195-206.

223.   De Paulis AN, Bertona E, Gutiérrez MA, Ramírez MS, Vay CA, Predari SC. *Ignavigranum ruoffiae*, a rare pathogen that caused a skin abscess. JMM Case Rep. 2018;5(2):e005137.

224.   Clair G, Roussi S, Armengaud J, Duport C. Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions. Mol Cell Proteomics. 2010;9(7):1486-98.

225.   Nguyen-Mau SM, Oh SY, Kern VJ, Missiakas DM, Schneewind O. Secretion genes as determinants of *Bacillus anthracis* chain length. J Bacteriol. 2012;194(15):3841-50.

226.   Lebre PH, Aliyu H, De Maayer P, Cowan DA. In silico characterization of the global *Geobacillus* and *Parageobacillus* secretome. Microb Cell Fact. 2018;17(1):156.

227.   Caspers M, Freudl R. *Corynebacterium glutamicum* possesses two *secA* homologous genes that are essential for viability. Arch Microbiol. 2008;189(6):605-10.

228.   Sharma K, Sultana T, Liao M, Dahms TES, Dillon J-AR. EF1025, a Hypothetical Protein From *Enterococcus faecalis*, Interacts With DivIVA and Affects Cell Length and Cell Shape. Front Microbiol. 2020;11.

229.   Andreevskaya M, Hultman J, Johansson P, Laine P, Paulin L, Auvinen P, *et al.* Complete genome sequence of *Leuconostoc gelidum* subsp. *gasicomitatum* KG16-1, isolated from vacuum-packaged vegetable sausages. Stand Genomic Sci. 2016;11:40.

230.   Lenz LL, Mohammadi S, Geissler A, Portnoy DA. SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. Proc Natl Acad Sci U S A. 2003;100(21):12432-7.

231.   Siboo IR, Chaffin DO, Rubens CE, Sullam PM. Characterization of the accessory Sec system of *Staphylococcus aureus*. J Bacteriol. 2008;190(18):6188-96.

232. Rodriguez BV, Kuehn MJ. *Staphylococcus aureus* secretes immunomodulatory RNA and DNA via membrane vesicles. Scientific Reports. 2020;10(1):18293.

233. Latousakis D, MacKenzie D, Telatin A, Juge N. Serine-rich repeat proteins from gut microbes. Gut Microbes. 2020;11(1):102-117. doi: 10.1080/19490976.2019.1602428.

234. Bensing BA, Gibson BW, Sullam PM. The *Streptococcus gordonii* platelet binding protein GspB undergoes glycosylation independently of export. J Bacteriol. 2004;186(3):638-45.

235. Latousakis D, Nepravishta R, Rejzek M, Wegmann U, Le Gall G, Kavanaugh D, *et al*. Serine-rich repeat protein adhesins from *Lactobacillus reuteri* display strain specific glycosylation profiles. Glycobiology. 2019;29(1):45-58.

236. Kotarski SF, Savage DC. Models for study of the specificity by which indigenous *lactobacilli* adhere to murine gastric epithelia. Infect Immun. 1979;26(3):966-75.

237. Edelheit O, Hanukoglu A, Hanukoglu I. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. BMC Biotech. 2009;9(1):61.

238. Zhu F, Erlandsen H, Ding L, Li J, Huang Y, Zhou M, *et al*. Structural and Functional Analysis of a New Subfamily of Glycosyltransferases Required for Glycosylation of Serine-rich *Streptococcal* Adhesins*. J of Biol Chem. 2011;286(30):27048-57.

239. Cowieson NP, Edwards-Gayle CJC, Inoue K, Khunti NS, Doutch J, Williams E, *et al*. Beamline B21: high-throughput small-angle X-ray scattering at Diamond Light Source. J Synchrotron Radiat. 2020;27(5):1438-46.

240. Manalastas-Cantos K, Konarev PV, Hajizadeh NR, Kikhney AG, Petoukhov MV, Molodenskiy DS, *et al*. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. J Appl Crystallogr. 2021;54(1):343-55.

241. Hopkins JB, Gillilan RE, Skou S. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. J Appl Crystallogr. 2017;50(5):1545-53.

242. Svergun D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992;25(4):495-503.

243. Grant TD. *Ab initio* electron density determination directly from solution scattering data. Nat Methods. 2018;15(3):191-3.

244. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. Nature. 2013;496(7446):477-81.

245. Piiadov V, Ares de Araújo E, Oliveira Neto M, Craievich AF, Polikarpov I. SAXSMoW 2.0: Online calculator of the molecular weight of proteins in dilute solution from experimental SAXS data measured on a relative scale. Prot Sci. 2019;28(2):454-63.

246. Hajizadeh NR, Franke D, Jeffries CM, Svergun DI. Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. Scientific Reports. 2018;8(1):7204.

247. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-W303.

248. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2011;27(3):343-50.

249. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39:175-91.

250. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Žídek A, *et al.* Highly accurate protein structure prediction for the human proteome. Nature. 2021;596(7873):590-6.

251. Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, *et al*. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res. 2021;49(W1):W431-W7.

252. McNicholas S, Potterton E, Wilson KS, Noble MEM. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr. 2011;67(4):386-94.

253. Veevers R, Hayward S. Methodological improvements for the analysis of domain movements in large biomolecular complexes. Biophys and Physicobiol. 2019;16:328-36.

254.   Reinhard L, Mayerhofer H, Geerlof A, Mueller-Dieckmann J, Weiss MS. Optimization of protein buffer cocktails using Thermofluor. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2013;69(Pt 2):209-14.

255.   Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38(7):3022-7.

256.   Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, *et al*. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform. 2002;3(3):265-74.

257.   Garg VK, Avashthi H, Tiwari A, Jain PA, Ramkete PW, Kayastha AM, et al. MFPPI - Multi FASTA ProtParam Interface. Bioinformation. 2016;12(2):74-7.

258.   Posch G, Pabst M, Brecker L, Altmann F, Messner P, Schäffer C. Characterization and scope of S-layer protein *O*-glycosylation in *Tannerella forsythia*. J Biol Chem. 2011;286(44):38714-24.

259.   Pabst M, Grouzdev DS, Lawson CE, Kleikamp HBC, de Ram C, Louwen R, *et al.* A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. ISME J. 2022;16(2):346-57.

260.   Nothaft H, Szymanski CM. Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol. 2010;8(11):765-78.

261.   Jervis Adrian J, Langdon R, Hitchen P, Lawson Andrew J, Wood A, Fothergill Joanne L, *et al*. Characterization of *N*-Linked Protein Glycosylation in *Helicobacter pullorum*. J Bacteriol. 2010;192(19):5228-36.

262.   Logan SM. Flagellar glycosylation - a new component of the motility repertoire? Microbiology (Reading). 2006;152(Pt 5):1249-62.

263.   Ohtsubo K, Marth JD. Glycosylation in cellular mechanisms of health and disease. Cell. 2006;126(5):855-67.

264.   Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: Structures, Functions, and Mechanisms. Annu Rev Biochem. 2008;77(1):521-55.

265.   Coutinho PM, Deleury E, Davies GJ, Henrissat B. An Evolving Hierarchical Family Classification for Glycosyltransferases. J Mol Biol. 2003;328(2):307-17.

266.   Hanukoglu I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. Biochem Mol Biol Educ. 2015;43(3):206-9.

267.   Chiu CPC, Watts AG, Lairson LL, Gilbert M, Lim D, Wakarchuk WW, *et al*. Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. Nat Struct Mol Biol. 2004;11(2):163-70.

268.   Liu J, Mushegian A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. Protein Sci. 2003;12(7):1418-31.

269.   Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A. Structures and mechanisms of glycosyltransferases. Glycobiology. 2006;16(2):29R-37R.

270.   Simoni RD. Saul Roseman: his many contributions to biochemistry over eight decades. Proc Natl Acad Sci U S A. 2011;108(45):18219-20.

271.   Davies GJ, Gloster TM, Henrissat B. Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr Opin Struct Biol. 2005;15(6):637-45.

272.   Chang A, Singh S, Phillips GN, Jr., Thorson JS. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. Curr Opin Biotechnol. 2011;22(6):800-8.

273.   Erb A, Weiß H, Härle J, Bechthold A. A bacterial glycosyltransferase gene toolbox: Generation and applications. Phytochemistry. 2009;70(15):1812-21.

274.   Williams GJ, Zhang C, Thorson JS. Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. Nat Chem Biol. 2007;3(10):657-62.

275.   Krauth C, Fedoryshyn M, Schleberger C, Luzhetskyy A, Bechthold A. Engineering a Function into a Glycosyltransferase. Chemistry & Biology. 2009;16(1):28-35.

276.   Chen Y, Seepersaud R, Bensing BA, Sullam PM, Rapoport TA. Mechanism of a cytosolic *O*-glycosyltransferase essential for the synthesis of a bacterial adhesion protein. Proc Natl Acad Sci U S A. 2016;113(9):E1190-9.

277. Calvio C, Romagnuolo F, Vulcano F, Speranza G, Morelli CF. Evidences on the role of the lid loop of γ-glutamyltransferases (GGT) in substrate selection. Enzyme and Microb Technol. 2018;114:55-62.

278. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-w303.

279. Takamatsu D, Bensing BA, Sullam PM. Four Proteins Encoded in the *gspB-secY2A2* Operon of *Streptococcus gordonii* Mediate the Intracellular Glycosylation of the Platelet-Binding Protein GspB. J of Bacteriol. 2004;186(21):7100-11.

280. Lizcano A, Akula Suresh Babu R, Shenoy AT, Saville AM, Kumar N, D'Mello A, *et al*. Transcriptional organization of pneumococcal *psrP-secY2A2* and impact of *GtfA* and *GtfB* deletion on PsrP-associated virulence properties. Microbes Infect. 2017;19(6):323-33.

281. Takamatsu D, Bensing BA, Sullam PM. Genes in the accessory sec locus of *Streptococcus gordonii* have three functionally distinct effects on the expression of the platelet-binding protein GspB. Mol Microbiol. 2004;52(1):189-203.

282. Takamatsu D, Bensing BA, Sullam PM. Four proteins encoded in the gspB-secY2A2 operon of Streptococcus gordonii mediate the intracellular glycosylation of the platelet-binding protein GspB. J Bacteriol. 2004;186(21):7100-11.

283. Esposito D, Günster RA, Martino L, El Omari K, Wagner A, Thurston TLM, *et al*. Structural basis for the glycosyltransferase activity of the *Salmonella* effector SseK3. J Biol Chem. 2018;293(14):5064-78.

284. Huang J, She Y, Yue J, Chen Y, Li Y, Li J, *et al*. Exploring the catalytic function and active sites of a novel *C*-glycosyltransferase from *Anemarrhena asphodeloides*. Synthetic and Systems Biotechnology. 2022;7(1):621-30.

285. Swanson BA, Frey PA. Identification of lysine 153 as a functionally important residue in UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry. 1993;32(48):13231-6.

286. Samland AK, Amrhein N, Macheroux P. Lysine 22 in UDP-N-acetylglucosamine enolpyruvyl transferase from *Enterobacter cloacae* is crucial for enzymatic activity and the formation of covalent adducts with the substrate phosphoenolpyruvate and the antibiotic fosfomycin. Biochemistry. 1999;38(40):13162-9.

287. Bhattacharyya U, Dhar G, Bhaduri A. An arginine residue is essential for stretching and binding of the substrate on UDP-glucose-4-epimerase from *Escherichia coli*. Use of a stacked and quenched uridine nucleotide fluorophore as probe. J Biol Chem. 1999;274(21):14573-8.

288. Ito K, Ito S, Shimamura T, Weyand S, Kawarasaki Y, Misaka T, *et al.* Crystal structure of glucansucrase from the dental caries pathogen *Streptococcus mutans.* J Mol Biol. 2011;408(2):177-86.

289. Zhu F, Erlandsen H, Ding L, Li J, Huang Y, Zhou M, *et al.* Structural and Functional Analysis of a New Subfamily of Glycosyltransferases Required for Glycosylation of Serine-rich *Streptococcal* Adhesins. Journal of Biological Chemistry. 2011;286(30):27048-57.

290. Tytgat HLP, Lin C-w, Levasseur MD, Tomek MB, Rutschmann C, Mock J, *et al.* Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. Nat Commun. 2019;10(1):5403.

291. Biswas A, Thattai M. Promiscuity and specificity of eukaryotic glycosyltransferases. Biochem Soc Trans. 2020;48(3):891-900.

292. Rini JM, Moremen KW, Davis BG, Esko JD. Glycosyltransferases and Glycan-Processing Enzymes. 4th ed: Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 2022 2022.

293. Dell A, Galadari A, Sastre F, Hitchen P. Similarities and Differences in the Glycosylation Mechanisms in Prokaryotes and Eukaryotes. Int J Microbiol. 2010;2010:148178.

294. Schäffer C, Messner P. Emerging facets of prokaryotic glycosylation. FEMS Microbiol Rev. 2017;41(1):49-91.

295. Zeituni AE, McCaig W, Scisci E, Thanassi DG, Cutler CW. The Native 67-Kilodalton Minor Fimbria of *Porphyromonas gingivalis* Is a Novel Glycoprotein with DC-SIGN-Targeting Motifs. J Bacteriol. 2010;192(16):4103-10.

296.   Thomas NA, Deng W, Baker N, Puente J, Finlay BB. Hierarchical delivery of an essential host colonization factor in enteropathogenic *Escherichia coli*. J Biol Chem. 2007;282(40):29634-45.

297.   Manera K, Kamal F, Burkinshaw B, Dong TG. Essential functions of chaperones and adaptors of protein secretion systems in Gram-negative bacteria. FEBS J. 2022;289(16):4704-17.

298.   Schroeder GN, Petty NK, Mousnier A, Harding CR, Vogrin AJ, Wee B, *et al.* *Legionella pneumophila* Strain 130b Possesses a Unique Combination of Type IV Secretion Systems and Novel Dot/Icm Secretion System Effector Proteins. J Bacteriol. 2010;192(22):6001-16.

299.   Walsh CT, Garneau-Tsodikova S, Gatto Jr GJ. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. Angew Chem Int Ed Engl. 2005;44(45):7342-72.

300.   Ma Q, Zhang Q, Chen Y, Yu S, Huang J, Liu Y, *et al.* Post-translational Modifications in Oral Bacteria and Their Functional Impact. Front Microbiol. 2021;12:784923.

301.   Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, *et al.* Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

302.   Chen Y, Bensing BA, Seepersaud R, Mi W, Liao M, Jeffrey PD, *et al.* Unraveling the sequence of cytosolic reactions in the export of GspB adhesin from *Streptococcus gordoni*i. J Biol Chem. 2018;293(14):5360-73.

303.   Jones CS, Sychantha D, Howell PL, Clarke AJ. Structural basis for the *O*-acetyltransferase function of the extracytoplasmic domain of OatA from *Staphylococcus aureus*. Journal of Biological Chemistry. 2020;295(24):8204-13.

304.   Poirier L, Plener L, Daudé D, Chabrière E. Enzymatic decontamination of paraoxon-ethyl limits long-term effects in planarians. Sci Rep. 2020;10(1):3843.

305.   Bernadó P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol Biosyst. 2012;8(1):151-67.

306.   Michalopoulos I, Torrance GM, Gilbert DR, Westhead DR. TOPS: an enhanced database of protein structural topology. Nucleic Acids Res. 2004;32(Database issue):D251-4.

307.   de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 2006;34(Web Server issue):W362-5.

308.   Boraston Alisdair B, Bolam David N, Gilbert Harry J, Davies Gideon J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J. 2004;382(3):769-81.

309.   Armenta S, Moreno-Mendieta S, Sánchez-Cuapio Z, Sánchez S, Rodríguez-Sanoja R. Advances in molecular engineering of carbohydrate-binding modules. Proteins. 2017;85(9):1602-17.

310.   Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun. 2022;13(1):1265.

311.   Bryant P, Pozzati G, Zhu W, Shenoy A, Kundrotas P, Elofsson A. Predicting the structure of large protein complexes using AlphaFold and Monte Carlo tree search. Nature Commun. 2022;13(1):6028.

312.   Cummins MC, Jacobs TM, Teets FD, DiMaio F, Tripathy A, Kuhlman B. AlphaFold accurately predicts distinct conformations based on the oligomeric state of a *de novo* designed protein. Protein Sci. 2022;31(7):e4368.

313.   Flower TG, Hurley JH. Crystallographic molecular replacement using an *in silico*-generated search model of SARS-CoV-2 ORF8. Protein Science. 2021;30(4):728-34.

314.   Christensen DG, Xie X, Basisty N, Byrnes J, McSweeney S, Schilling B, *et al.* Post-translational Protein Acetylation: An Elegant Mechanism for Bacteria to Dynamically Regulate Metabolic Functions. Front Microbiol. 2019;10.

315.   Tobin Kåhrström C. Acetyl phosphate revealed as a global acetyl donor. Nat Rev Microbiol. 2013;11(9):597-.

316.   Liyanage SH, Yan M. Quantification of binding affinity of glyconanomaterials with lectins. Chem Commun (Camb). 2020;56(88):13491-505.

317.   Kinoshita M, Nakanishi Y, Furukawa Y, Namba K, Imada K, Minamino T. Rearrangements of α-helical structures of FlgN chaperone control the binding affinity for its cognate substrates during flagellar type III export. Mol Microbiol. 2016;101(4):656-70.
318.   Bandara M, Corey RA, Martin R, Skehel JM, Blocker AJ, Jenkinson HF, *et al.* Composition and Activity of the Non-canonical Gram-positive SecY2 Complex. J Biol Chem. 2016;291(41):21474-84.
319.   Swanson S, Ioerger TR, Rigel NW, Miller BK, Braunstein M, Sacchettini JC. Structural Similarities and Differences between Two Functionally Distinct SecA Proteins, *Mycobacterium tuberculosis* SecA1 and SecA2. J Bacteriol. 2015;198(4):720-30.
320.   Li L, Park E, Ling J, Ingram J, Ploegh H, Rapoport TA. Crystal structure of a substrate-engaged SecY protein-translocation channel. Nature. 2016;531(7594):395-9.
321.   Vassylyev DG, Mori H, Vassylyeva MN, Tsukazaki T, Kimura Y, Tahirov TH, *et al*. Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. J Mol Biol. 2006;364(3):248-58.
322.   Bensing BA, Yen YT, Seepersaud R, Sullam PM. A Specific interaction between SecA2 and a region of the preprotein adjacent to the signal peptide occurs during transport via the accessory Sec system. J Biol Chem. 2012;287(29):24438-47.
323.   Renier S, Chagnot C, Deschamps J, Caccia N, Szlavik J, Joyce SA, *et al*. Inactivation of the SecA2 protein export pathway in *Listeria monocytogenes* promotes cell aggregation, impacts biofilm architecture and induces biofilm formation in environmental condition. Environ Microbiol. 2014;16(4):1176-92.
324.   Bandara M, Skehel JM, Kadioglu A, Collinson I, Nobbs AH, Blocker AJ, *et al.* The accessory Sec system (SecY2A2) in *Streptococcus pneumoniae* is involved in export of pneumolysin toxin, adhesion and biofilm formation. Microbes and Infection. 2017;19(7):402-12.
325.   Zimmer J, Nam Y, Rapoport TA. Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature. 2008;455(7215):936-43.
326.   Orelle C, Durmort C, Mathieu K, Duchêne B, Aros S, Fenaille F, *et al.* A multidrug ABC transporter with a taste for GTP. Scientific Reports. 2018;8(1):2309.
327.   Yuan Y, Cao D, Zhang Y, Ma J, Qi J, Wang Q, et al. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. Nat Commun. 2017;8(1):15092.
328.   Frenz B, Rämisch S, Borst AJ, Walls AC, Adolf-Bryfogle J, Schief WR, *et al.* Automatically Fixing Errors in Glycoprotein Structures with Rosetta. Structure. 2019;27(1):134-9.e3.
329.   Kranz JK, Clemente JC. Binding techniques to study the allosteric energy cycle. Methods Mol Biol. 2012;796:3-17.
330.   Fulton KM, Smith JC, Twine SM. Clinical applications of bacterial glycoproteins. Expert Rev Proteomics. 2016;13(4):345-53.
331.   Brandis G, Cao S, Hughes D. Operon Concatenation Is an Ancient Feature That Restricts the Potential to Rearrange Bacterial Chromosomes. Mol Biol Evol. 2019;36(9):1990-2000.
332.   Hekkelman ML, de Vries I, Joosten RP, Perrakis A. AlphaFill: enriching AlphaFold models with ligands and cofactors. Nat Methods. 2022.

Appendix 1

The maps of the plasmids used for recombinant aSec protein expression (section 2.2.3, Table 2.4)

are presented here. Plasmids used for site-directed mutagenesis display the corresponding mutated

residues annotated onto the map.
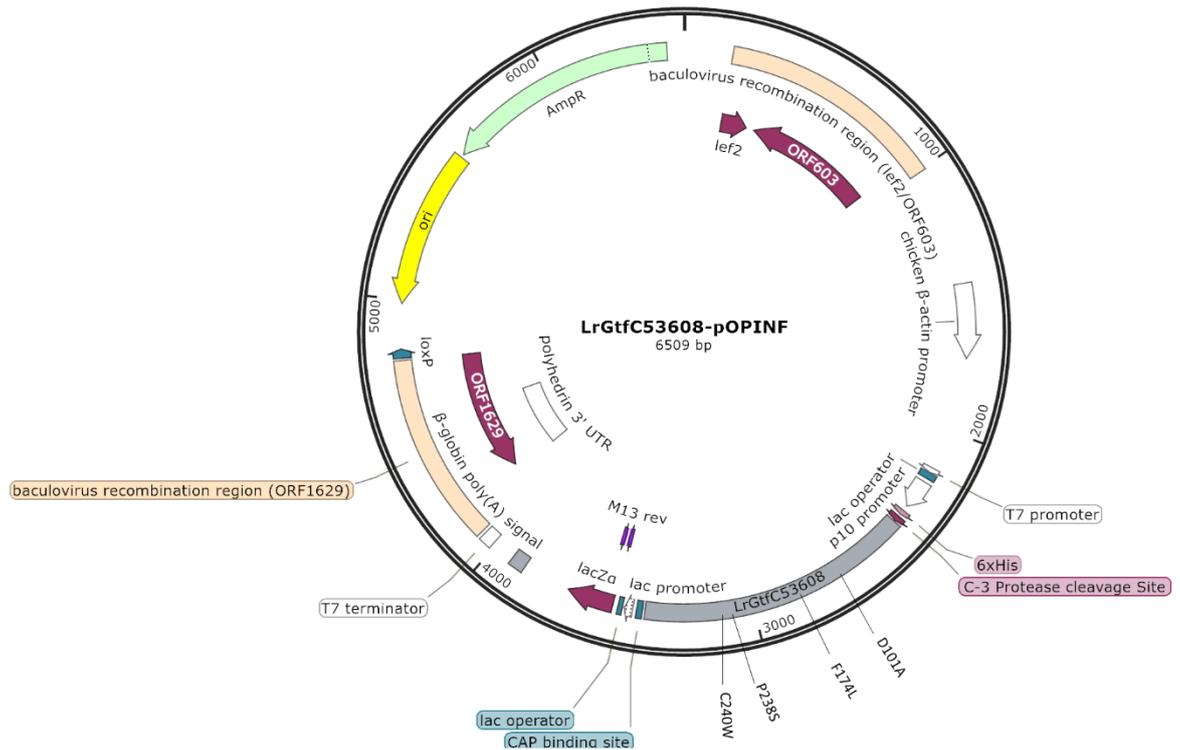
Supplementary Figure S1.1 Map of recombinant *Lr*GtfC$_{53608}$-pOPINF plasmid. Site-directed mutagenesis carried out in this plasmid generated *Lr*GtfC F174L$_{53608}$, *Lr*GtfC P238S$_{53608}$, *Lr*GtfC C240W$_{53608}$ and *Lr*GtfC D101A$_{53608}$ mutant proteins. Site-directed amino acid targets have been annotated at their respective positions.
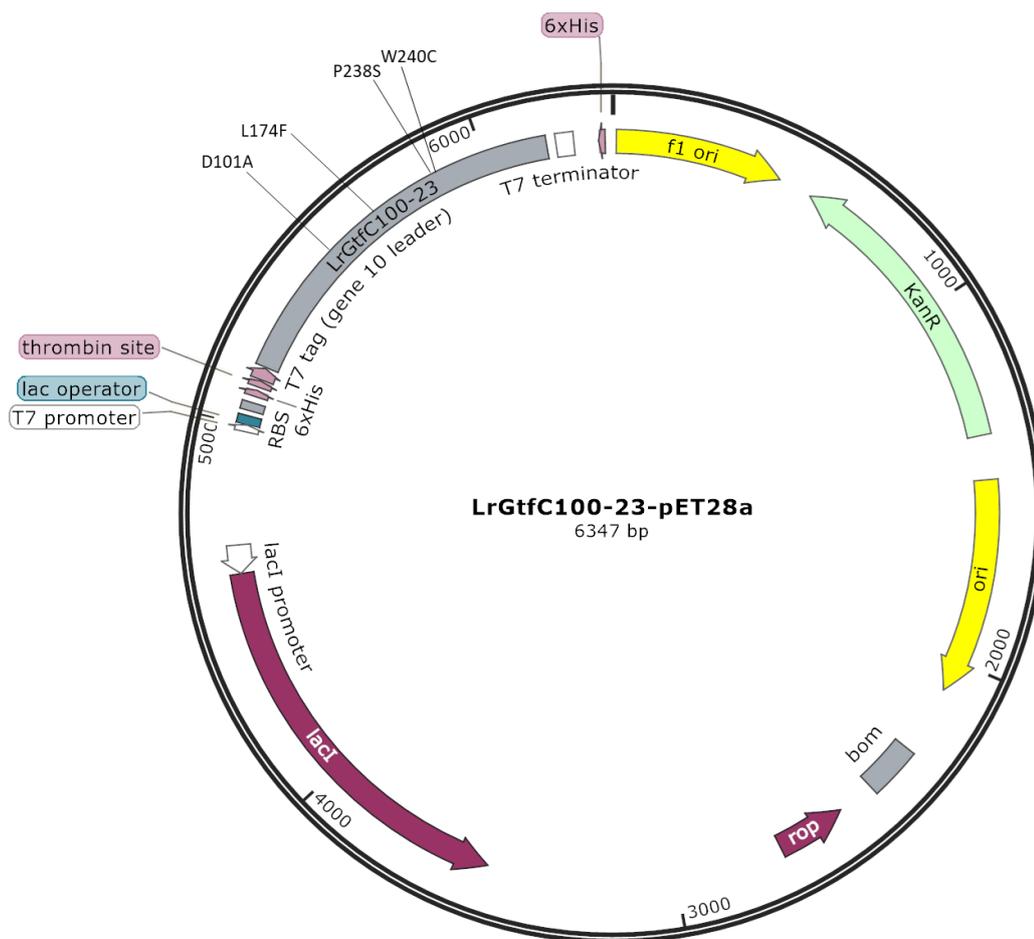
Supplementary Figure S1.2 Map of recombinant *Lr*GtfC$_{100-23}$-pET28a plasmid. Site-directed mutagenesis carried out in this plasmid generated *Lr*GtfC L174F$_{100-23}$, *Lr*GtfC S238P$_{100-23}$, *Lr*GtfC$_{100-23}$ and *Lr*GtfC D101A$_{100-23}$ mutant proteins. Site-directed amino acid targets have been annotated at their respective positions.
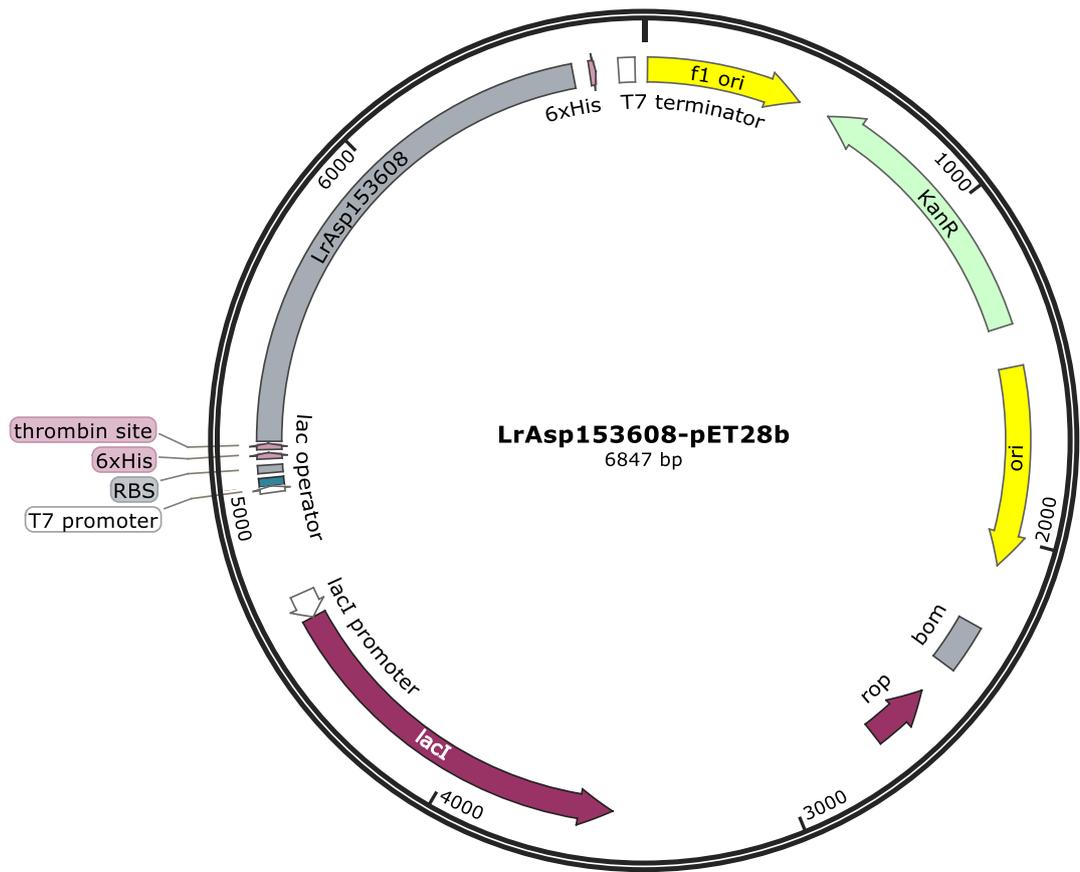
Supplementary Figure S1.3 Map of *Lr*Asp1$_{53608}$-pET28b recombinant plasmid.
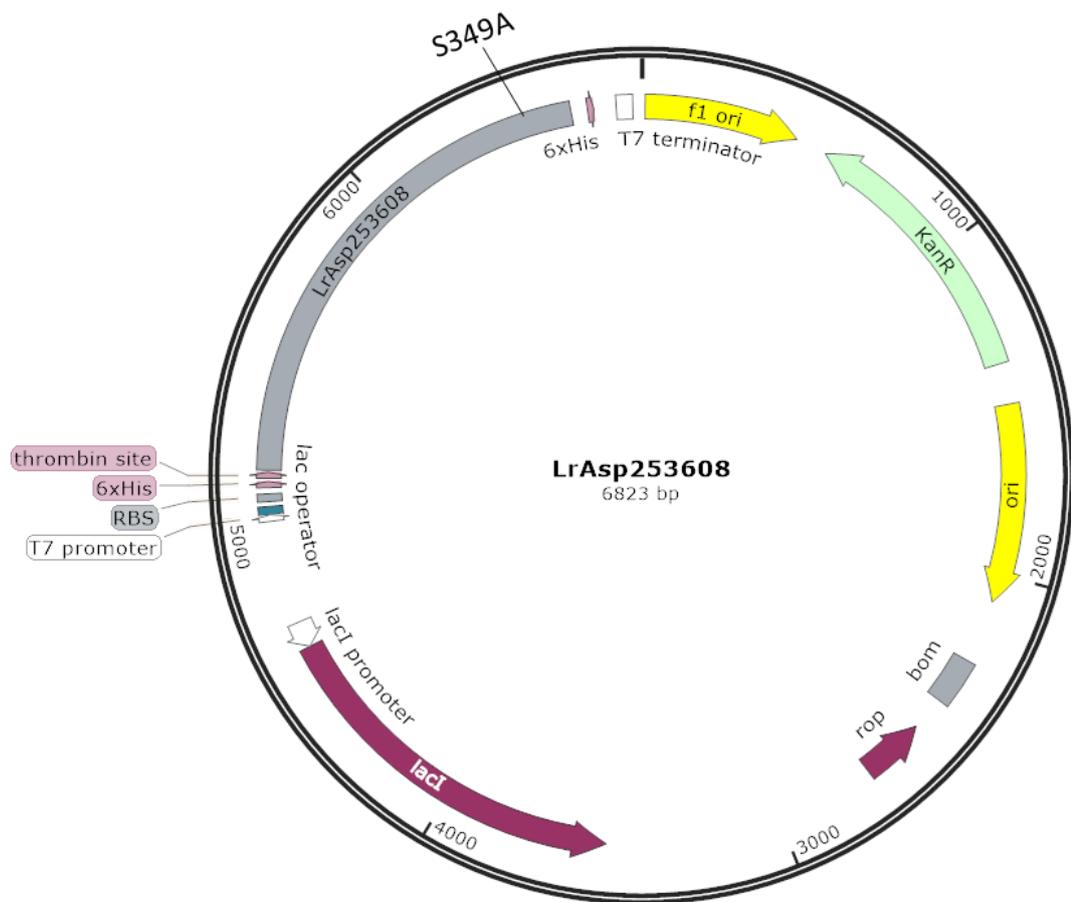
Supplementary Figure S1.4 Map of *Lr*Asp2$_{53608}$-pET28b recombinant plasmid. Site-directed mutagenesis target generating *Lr*Asp2S349A$_{53608}$ has been annotated at its position.
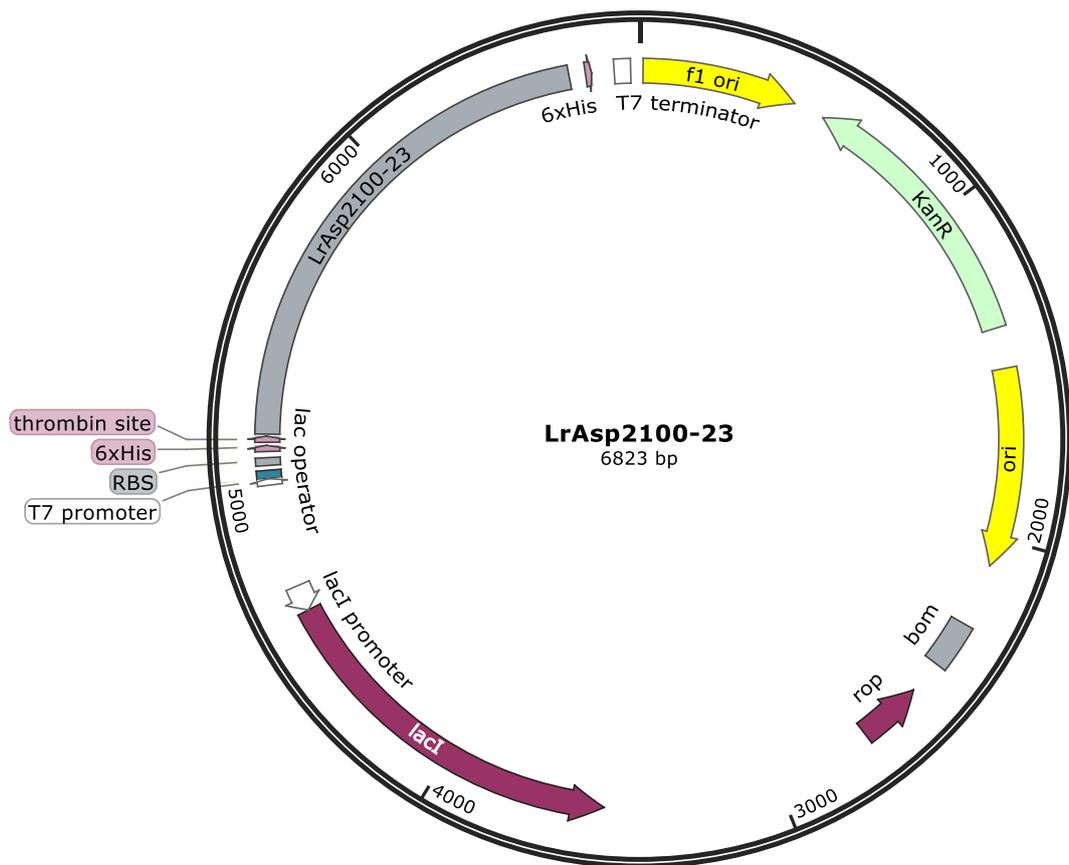
Supplementary Figure S1.5 Map of *Lr*Asp2$_{100\text{-}23}$-pET28b recombinant plasmid.
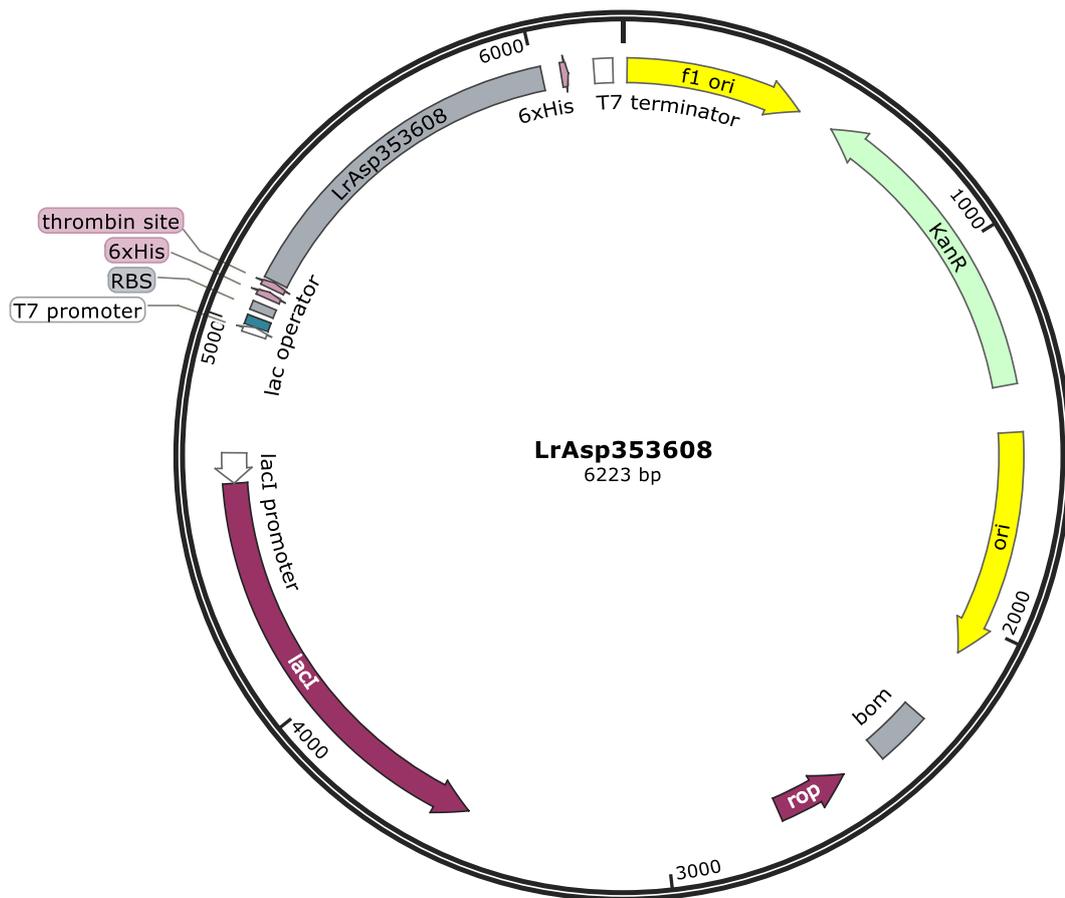
Supplementary Figure S1.6 Map of *Lr*Asp3₅₃₆₀₈-pET28b recombinant plasmid
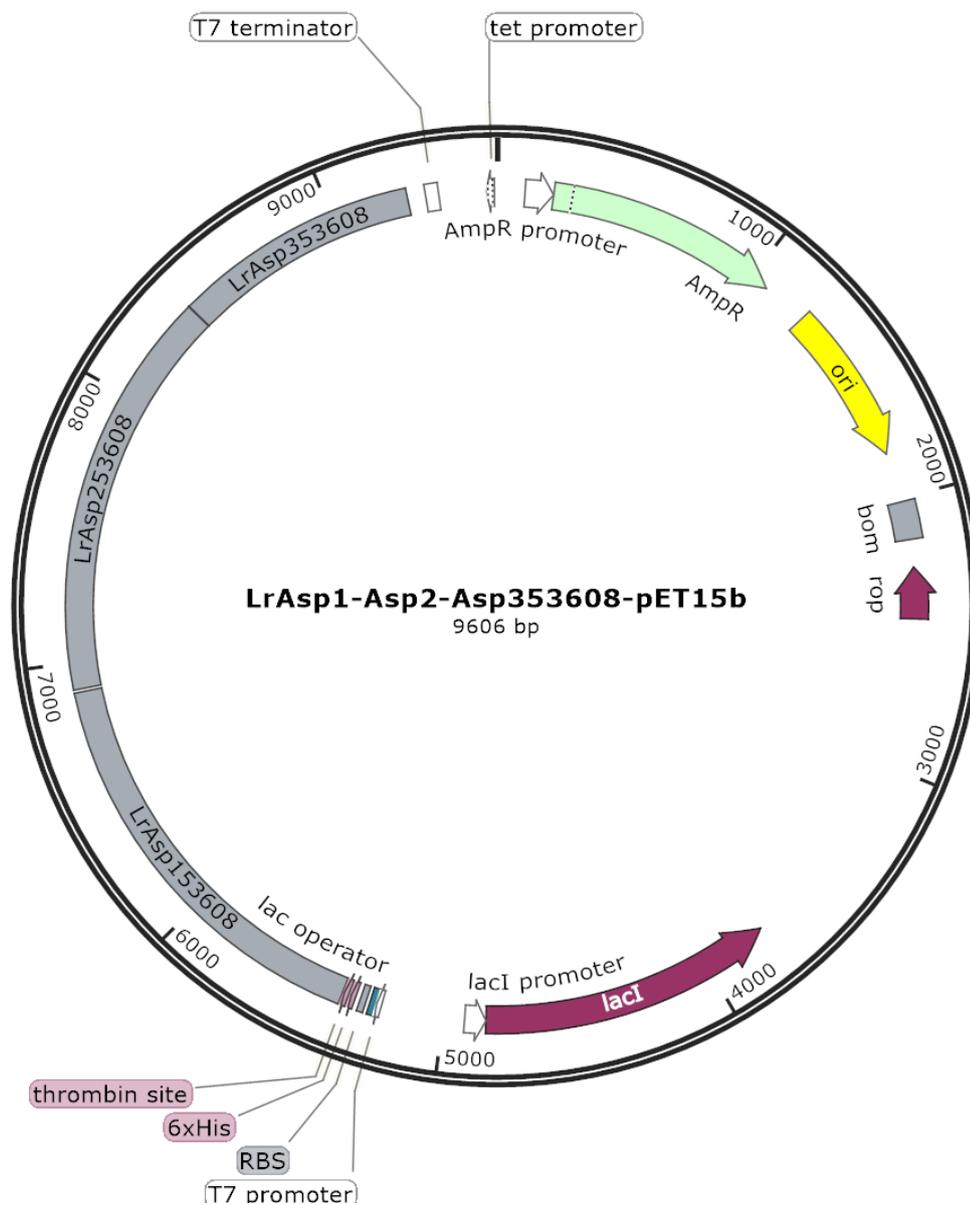
Supplementary Figure S1.7 Map of *Lr*Asp1-Asp2-Asp3$_{53608}$-pET28b recombinant plasmid.
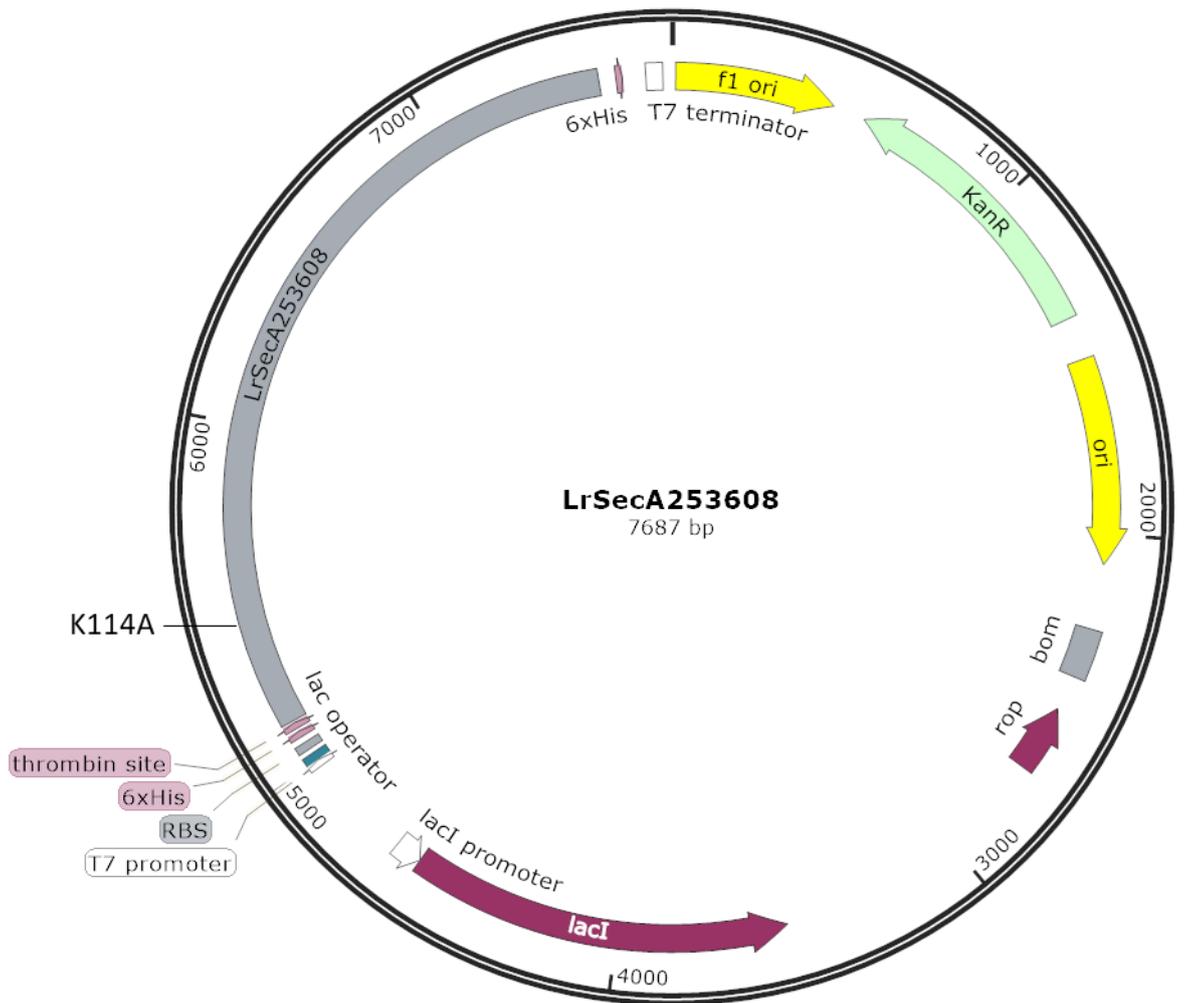
Supplementary Figure S1.8 Map of *Lr*SecA2$_{53608}$-pET28b recombinant plasmid.
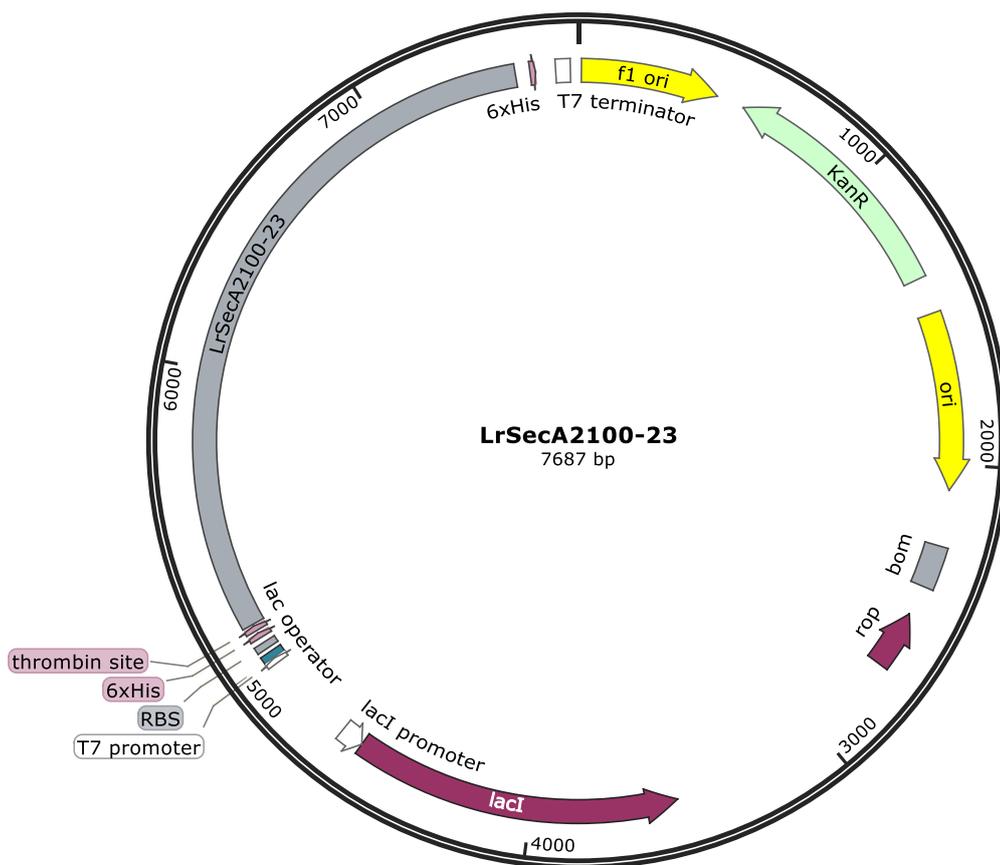
Supplementary Figure S1.9 Map of *Lr*SecA2$_{100\text{-}23}$-pET28b recombinant plasmid.
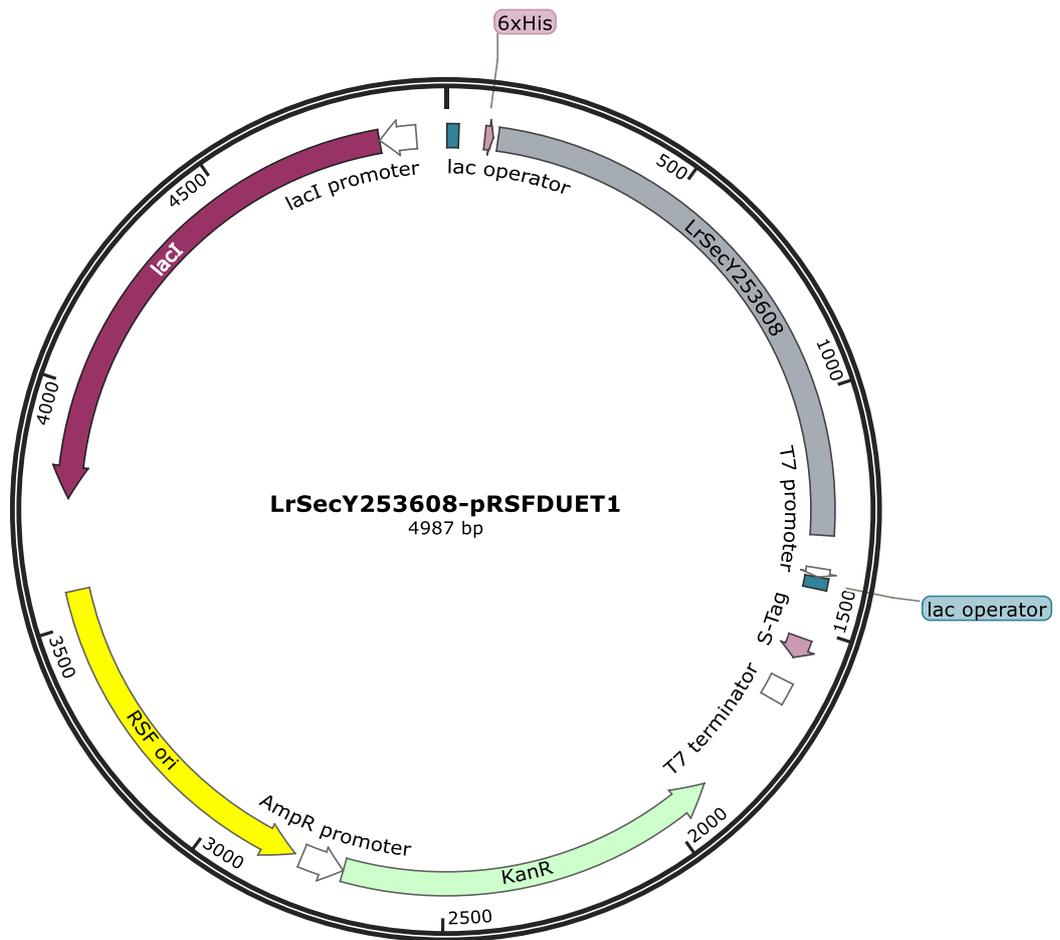
Supplementary Figure S1.10 Map of recombinant *Lr*SecY2$_{53608}$-pRSFDUET-1 plasmid.
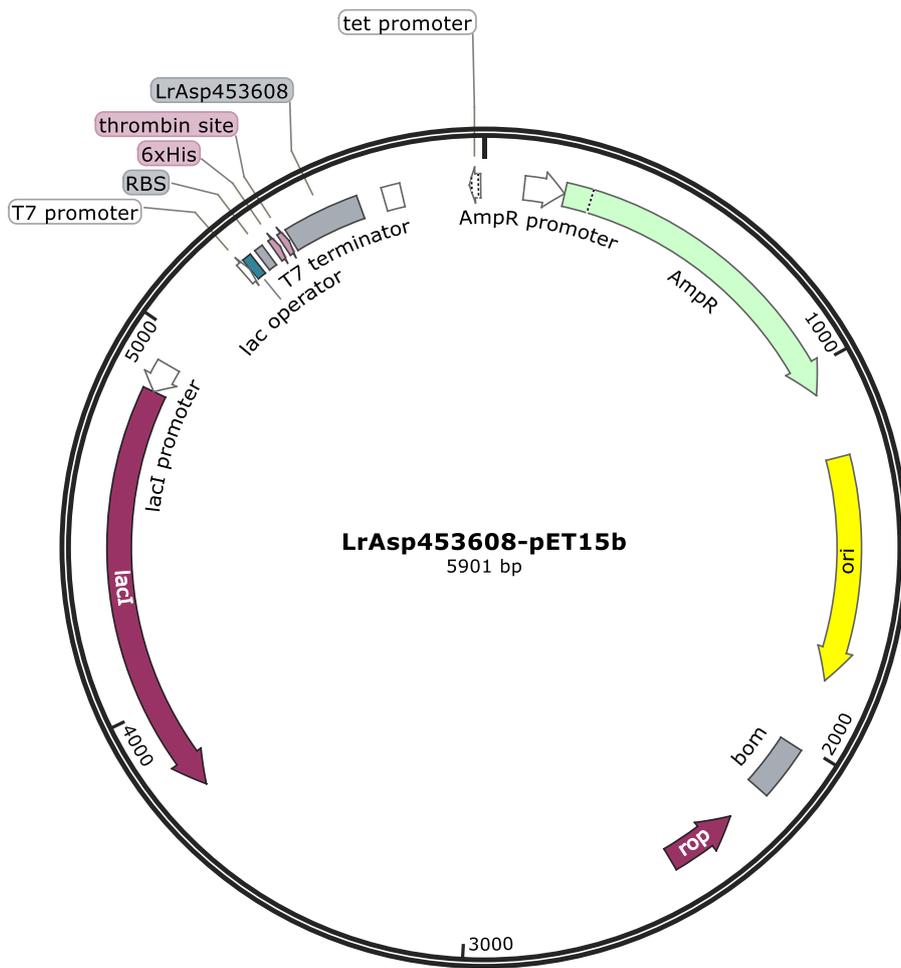
Supplementary Figure S1.11 Maps of recombinant *Lr*Asp4$_{53608}$-pET15b plasmid.
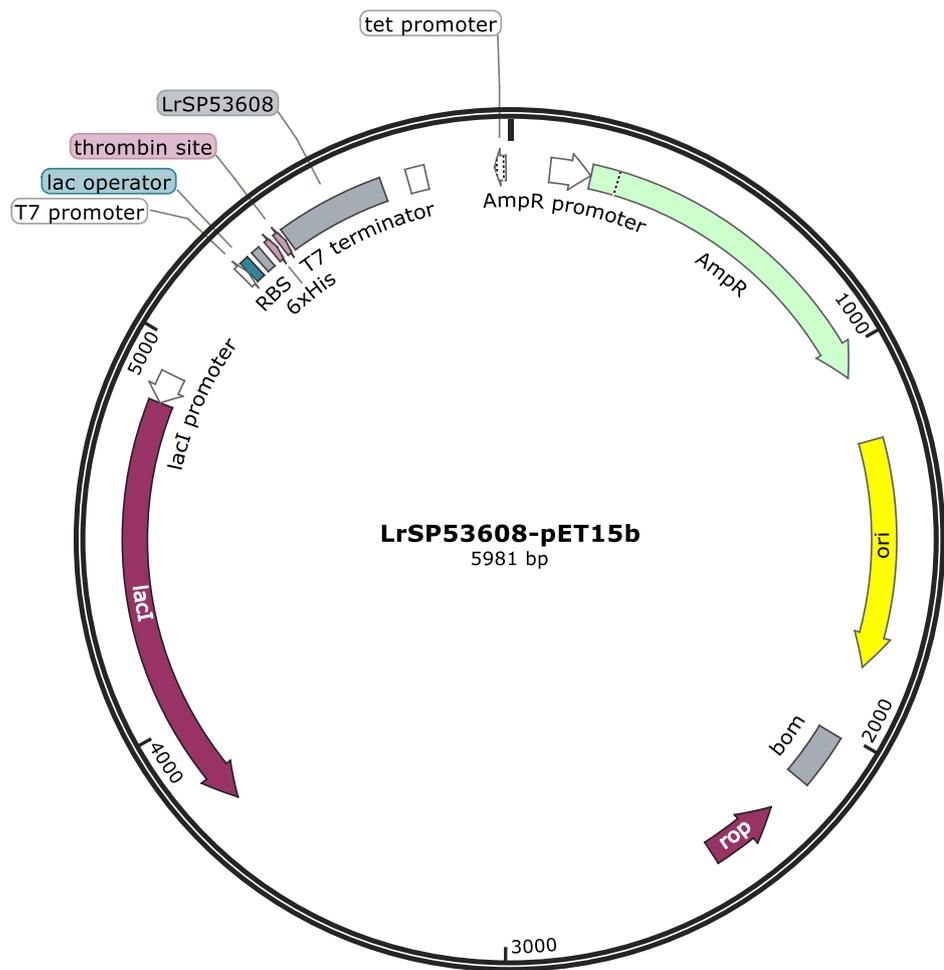
Supplementary Figure S1.12 Map of *Lr*SP₅₃₆₀₈-pET15b recombinant plasmid.
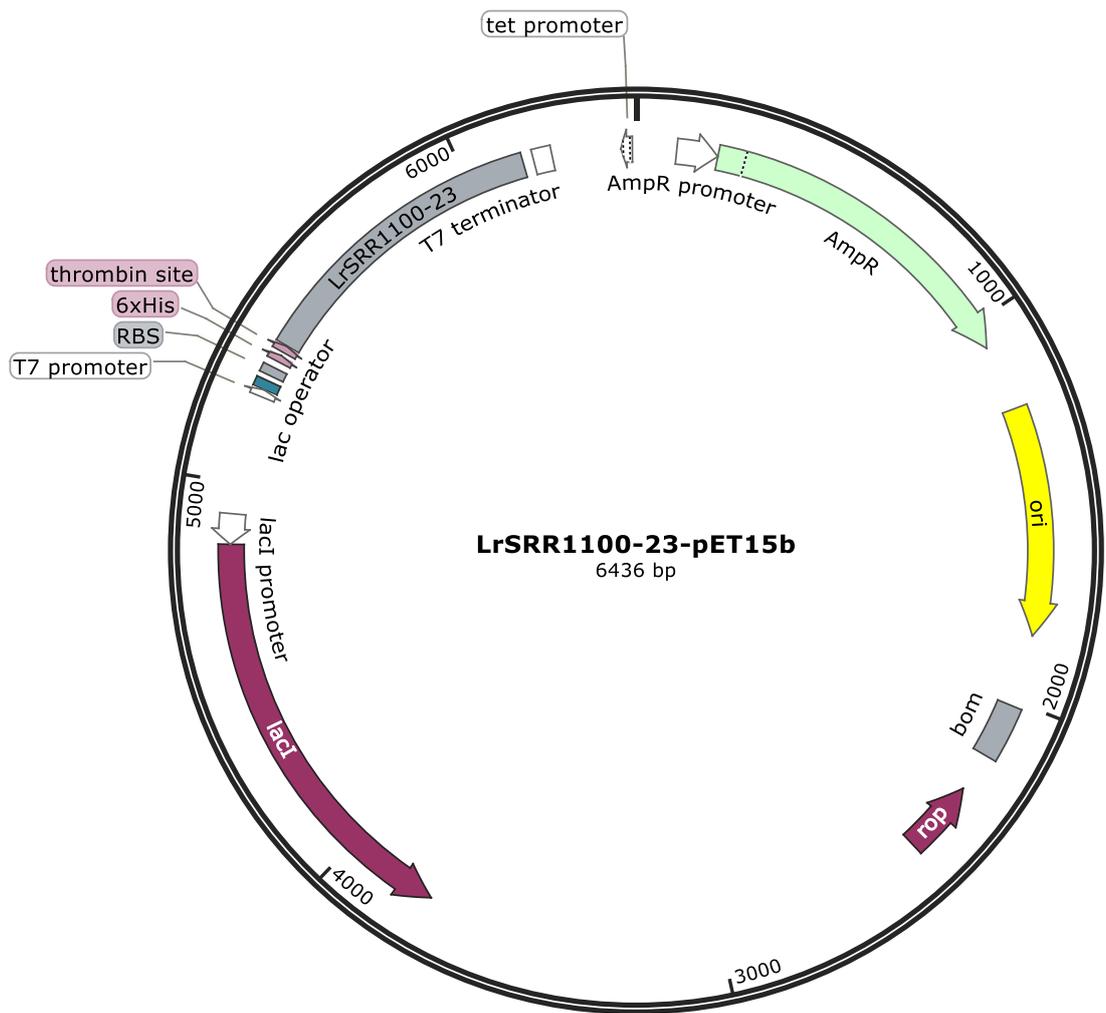
Supplementary Figure S1.13 Map of recombinant *Lr*SRR1₁₀₀₋₂₃-pET15b.
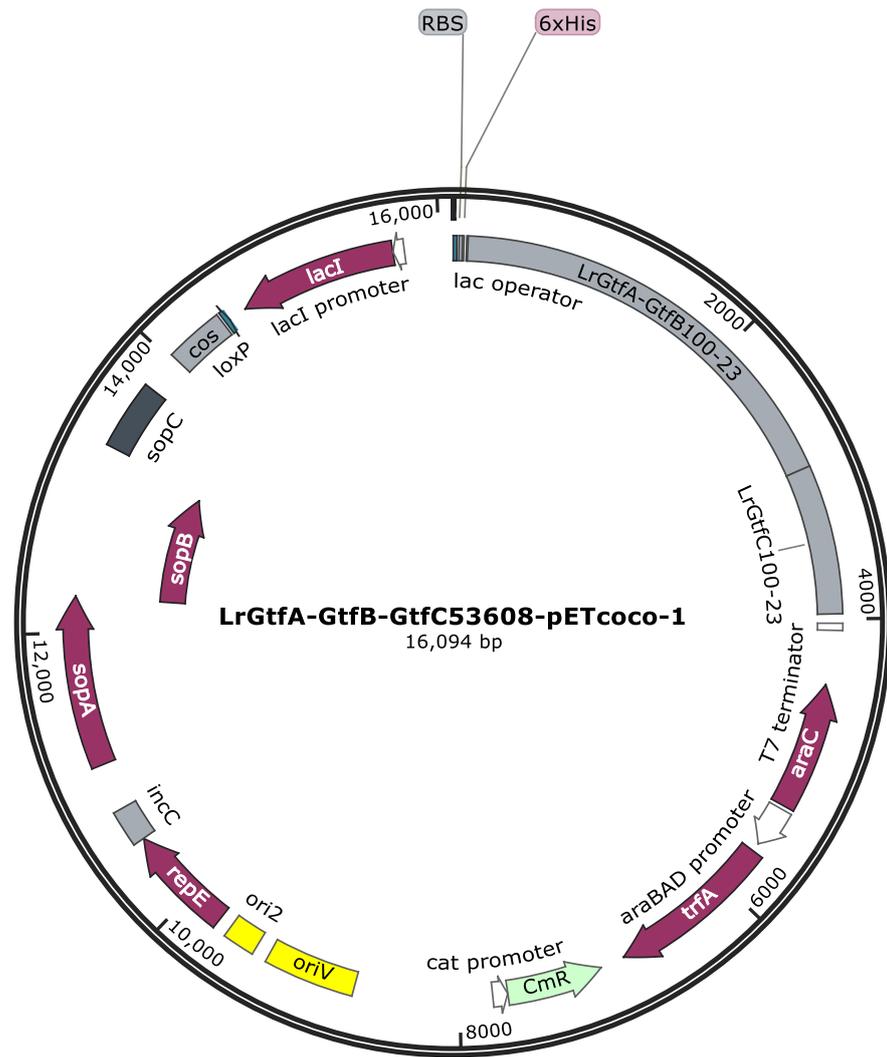
Supplementary Figure S1.14 Map of recombinant of *Lr*GtfA-GtfB-GtfC$_{100\text{-}23}$-pETcoco-1 plasmid.
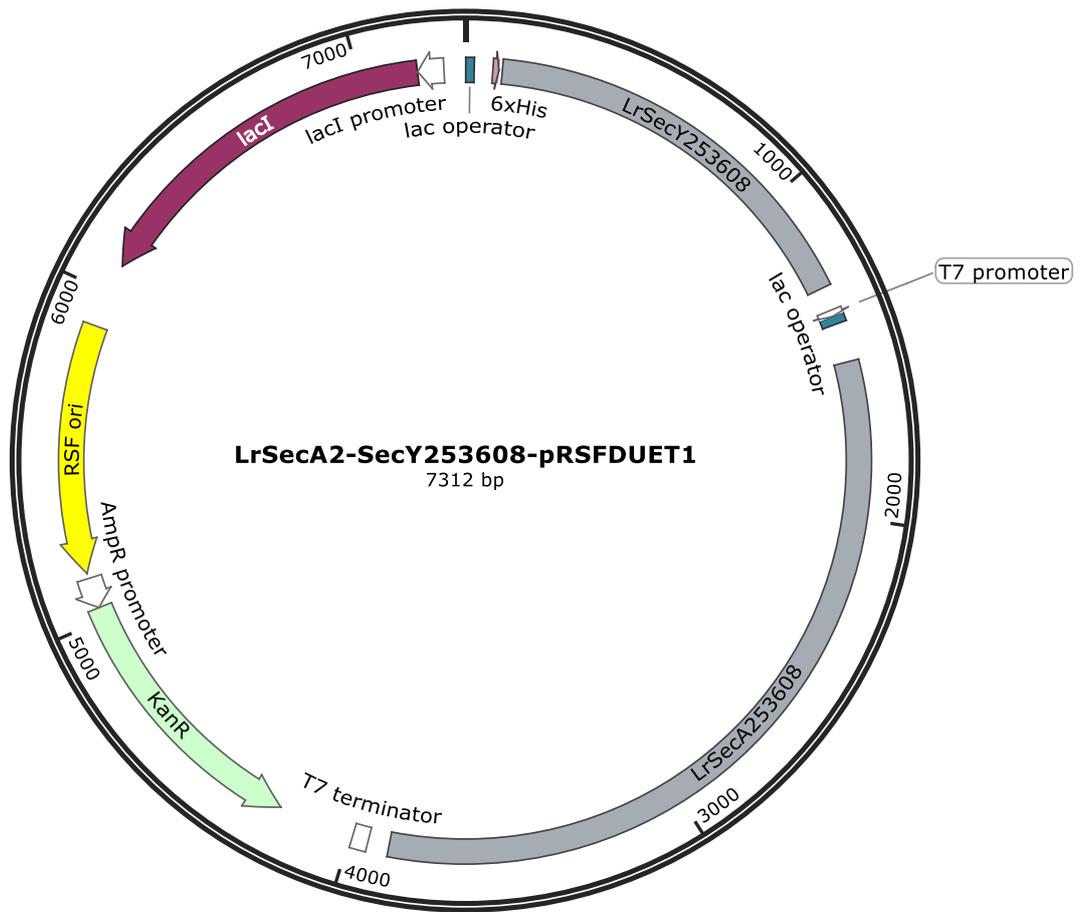
Supplementary Figure S1.15 Map of recombinant *Lr*SecA2-SecY2$_{53608}$-pRSFDUET-1 plasmid.
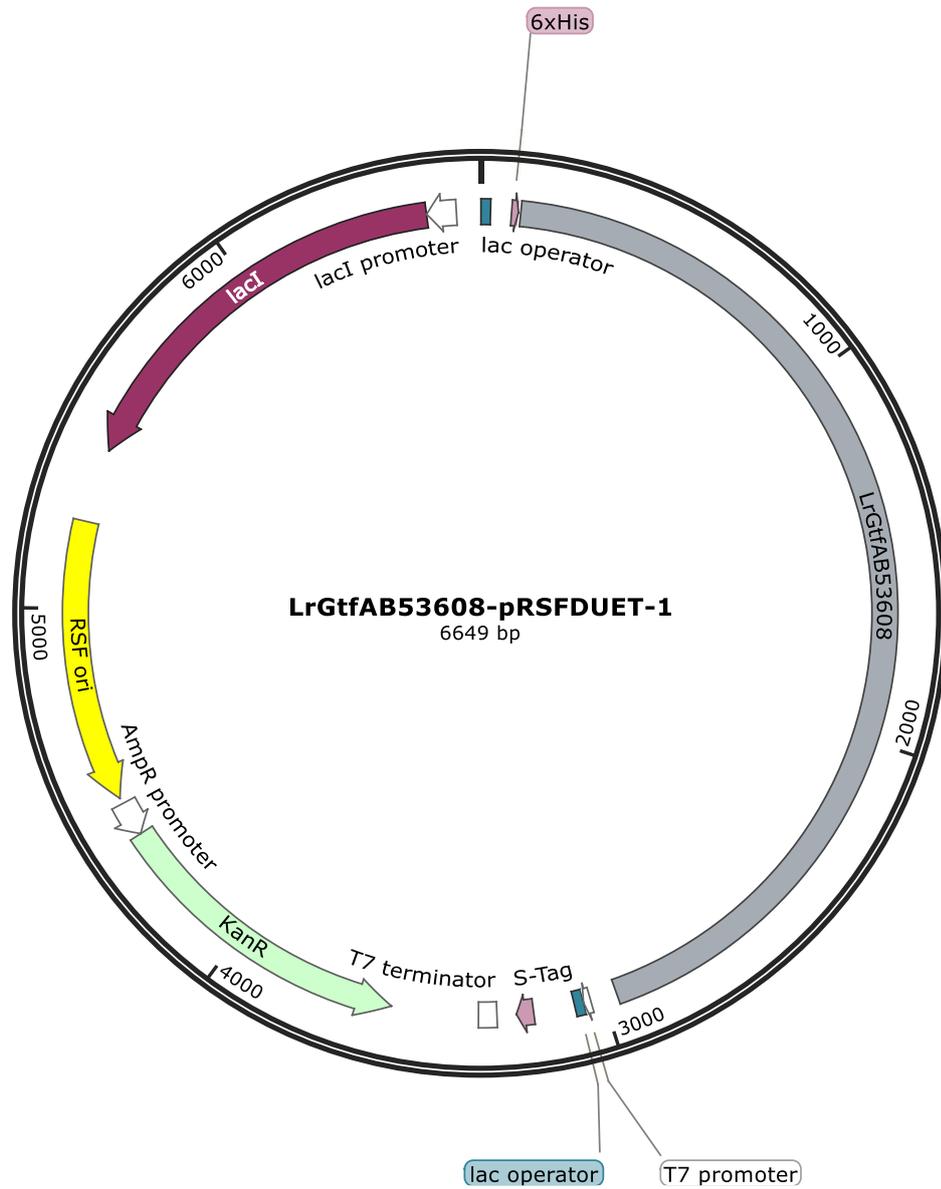
Supplementary Figure S1.16 Map of recombinant *Lr*GtfA-GtfB$_{53608}$-pRSFDUET-1 plasmid.

Appendix 2

Here the crystals formed from recombinant aSec proteins using conditions described in section 2.5.1 are reported.
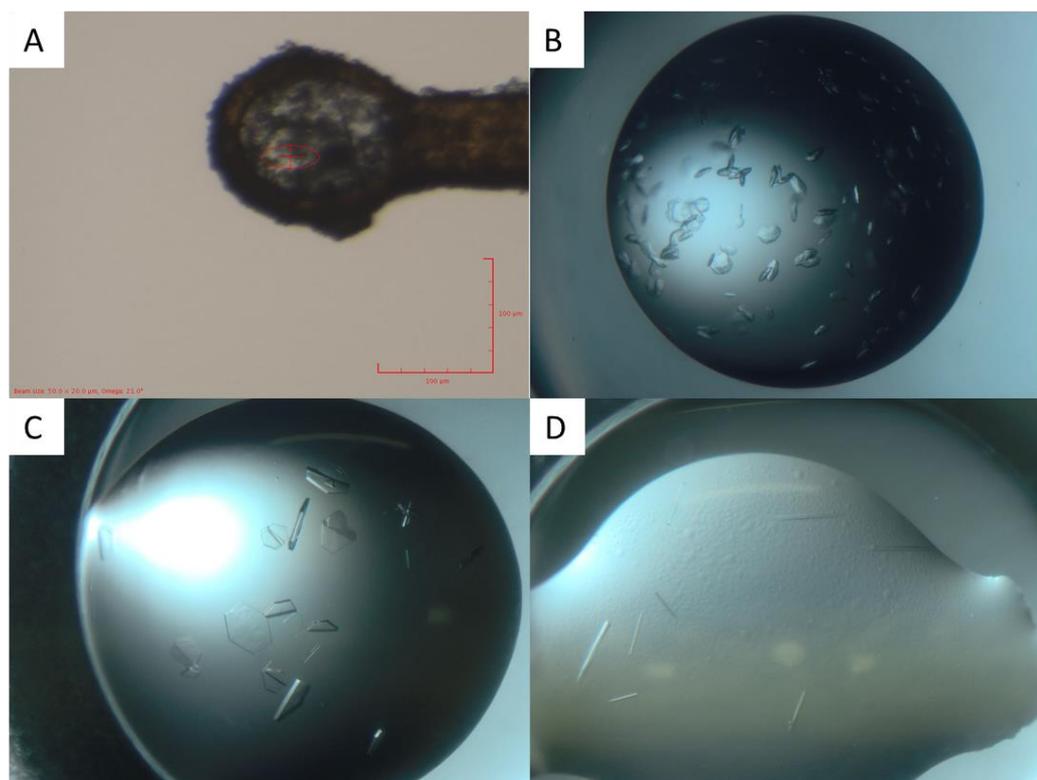
Supplementary Figure S2.1 Crystallisation of recombinant *Lr*GtfC$_{100\text{-}23}$. (A) 10 mg/mL *Lr*GtfC$_{100\text{-}23}$ in 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 7.5, 20% w/v Polyethylene Glycol 8000 (B) 10 mg/mL *Lr*GtfC$_{100\text{-}23}$ in 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 7.5, 20% w/v polyethylene glycol 8000, 1 mM UDP (C) 10 mg/mL *Lr*GtfC$_{100\text{-}23}$ in 0.2 M Potassium Thiocyanate, 0.1 M Bis-Tris propane pH 7.5, 20% w/v polyethylene glycol 8000, 1 mM UDP-GlcNAc (D) 10 mg/mL *Lr*GtfC$_{100\text{-}23}$ in 0.2 M potassium thiocyanate, 0.1 M Bis-Tris propane pH 7.5, 20% w/v polyethylene glycol 8000, 1 mM UDP-GlcNAc. Protein crystallisation trials were held at 25°C with a ratio of 0.3 µL protein solution : 0.5 µL crystallisation solution.

Supplementary Figure S2.2 Crystallisation of recombinant *Lr*GtfC$_{53608}$. 10 mg/mL *Lr*GtfC$_{53608}$ in 0.1 M potassium thiocyanate, 0.1 M bis-Tris propane pH 7.0, 40% polyethylene glycol 8000. Protein crystallisation trials were held at 25°C with a ratio of 0.25 µL protein solution : 0.3 µL crystallisation solution.
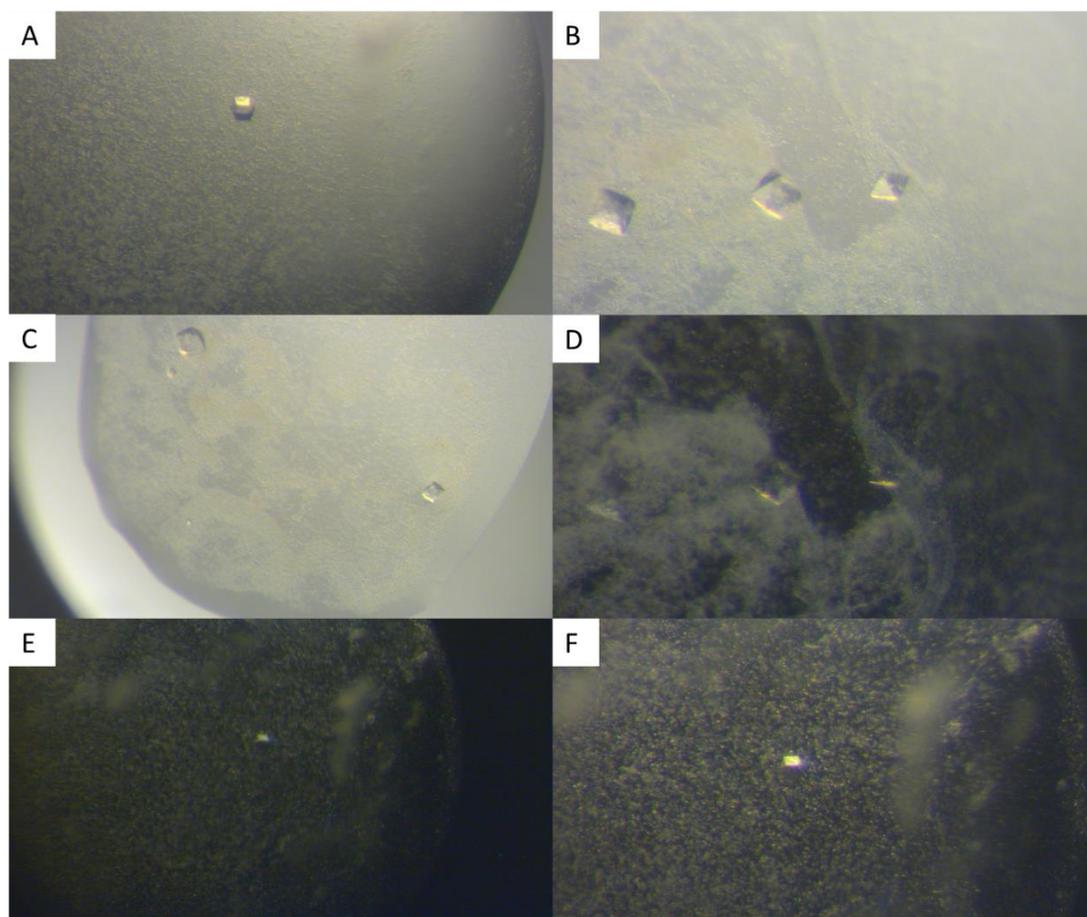
Supplementary Figure S2.3 Crystallisation of recombinant *Lr*Asp2$_{100\text{-}23}$. (A) 12.5 mg/mL *Lr*Asp2$_{100\text{-}23}$ in 0.1 M magnesium chloride, 30% polyethylene glycol 6000 **(B)** 12.5 mg/mL *Lr*Asp2$_{100\text{-}23}$ in 0.1 M Tris pH 8.0, 0.3 M NaCl, 20% polyethylene glycol 3000**(C)** 12.5 mg/mL *Lr*Asp2$_{100\text{-}23}$ in 0.1 M sodium formate, 0.1 M sodium cacodylate pH 6.5, 8% γ-PGA **(D)** 15 mg/mL *Lr*Asp2$_{100\text{-}23}$ in 0.2 M magnesium chloride, 0.1 M sodium cacodylate pH 6.5, 8% γ-PGA **(E)** 15 mg/mL *Lr*Asp2$_{100\text{-}23}$ in 0.2 M lithium sulphate, 20% polyethylene glycol 4000 **(F)** 15 mg/mL *Lr*Asp2$_{100\text{-}23}$ in 0.1 M calcium chloride, 0.1 M bis-Tris propane pH 5.5, 25% polyethylene glycol 3350. Protein crystallisation trials were held at 25°C with a ratio of 0.3 μL protein solution : 0.5 μL crystallisation solution.
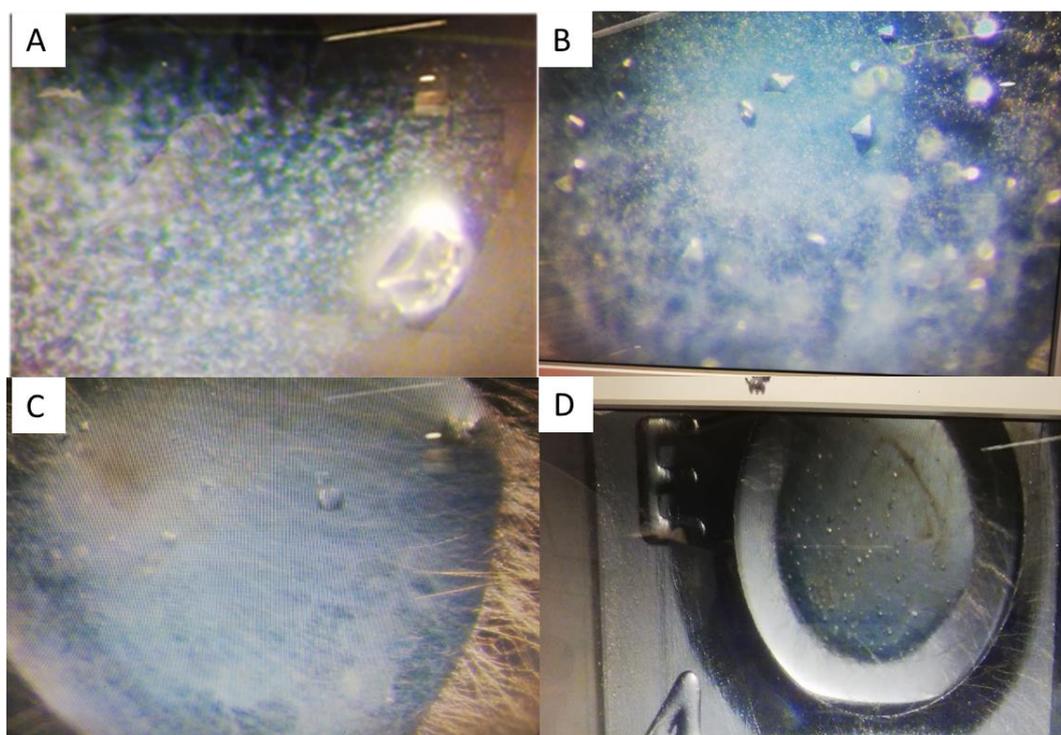
Supplementary Figure S2.4 Crystallisation of recombinant *Lr*Asp2$_{53608}$. (A) 12.5 mg/mL *Lr*Asp2$_{53608}$ in 0.1 M HEPES pH 7.5, 1 M sodium chloride, 5% polyethylene glycol 3350 (B) 12.5 mg/mL *Lr*Asp2$_{53608}$ in 0.1 M sodium formate, 0.1 M sodium cacodylate pH 6.5, 8% γ-PGA (C) 12.5 mg/mL *Lr*Asp2$_{53608}$ in 0.01 M magnesium chloride, 0.1 M HEPES pH 7.5, 25% polyethylene glycol 3350 (D) 15 mg/mL *Lr*Asp2$_{53608}$ in 1.26 M ammonium sulphate, 0.1 M Tris pH 8.6. Protein crystallisation trials were held at 25°C with a ratio of 0.3 µL protein solution : 0.5 µL crystallisation solution.
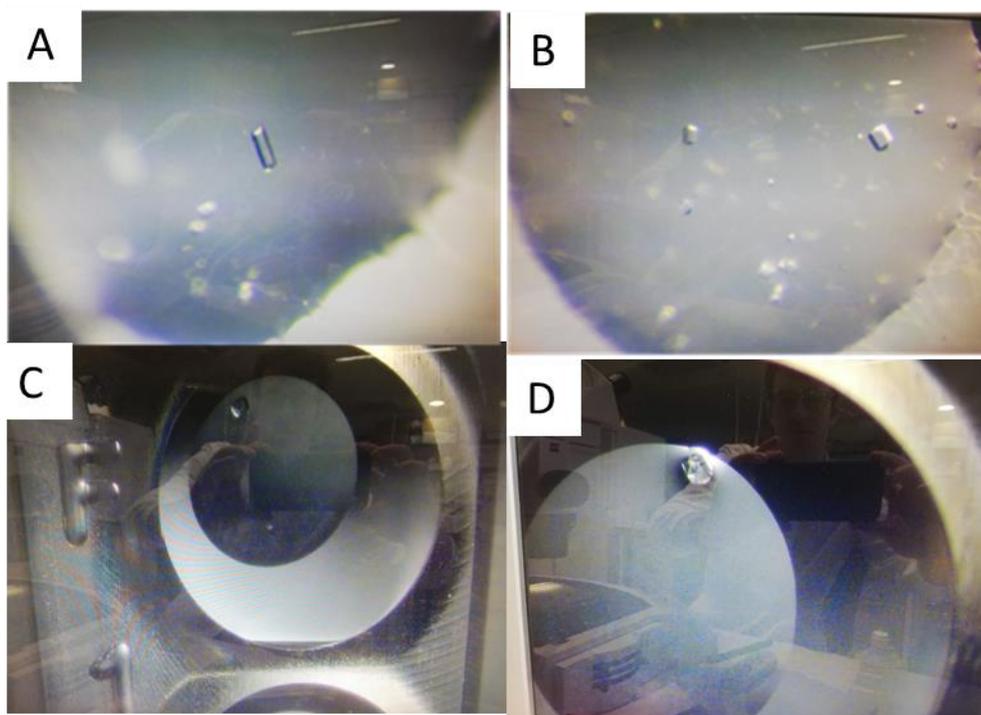
Supplementary Figure S2.5 Crystallisation of recombinant _Lr_Asp1-Asp2-Asp3$_{53608}$. (A) 10 mg/mL _Lr_Asp1-Asp2-Asp3$_{53608}$ in 0.1 M MES, pH 6.0, 0.2 M calcium chloride, 20% polyethylene glycol 6000 (B) 10 mg/mL _Lr_Asp1-Asp2-Asp3$_{53608}$ in 0.1 M MES, pH 6.0, 0.2 M calcium chloride, 20% polyethylene glycol 6000 (C) 10 mg/mL _Lr_Asp1-Asp2-Asp3$_{53608}$ in 0.1 M bis-Tris propane pH 7.0, 0.2 M sodium fluoride, 20% polyethylene glycol 3350 (D) 10 mg/mL _Lr_Asp1-Asp2-Asp3$_{53608}$ in 0.1 M bis-Tris propane pH 7.5, 0.2 M sodium fluoride, 20% polyethylene glycol 3350 . Protein crystallisation trials were held at 25°C with a ratio of 0.3 µL protein solution : 0.5 µL crystallisation solution.
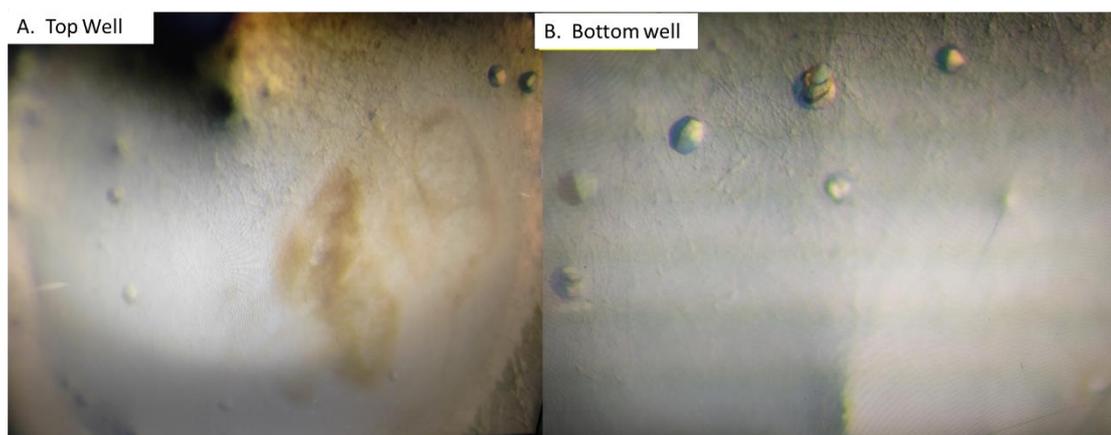
Supplementary Figure S2.6 Crystallisation of recombinant _Lr_SecA2$_{100-23}$. (A) 10 mg/mL _Lr_SecA2$_{100-23}$ in 0.1 M NaCl, 0.1 M MOPS pH 6.1, 30% polyethylene glycol 400. (B) 10 mg/mL _Lr_SecA2$_{100-23}$ in 0.1 M NaCl, 0.1 M MOPS pH 6.1, 30% polyethylene glycol 400. Protein crystallisation trials were held at 25°C with a ratio of 0.25 µL protein solution : 0.25 µL crystallisation solution for the top well and 0.25 µL protein solution : 0.25 µL crystallisation solution for the bottom well.
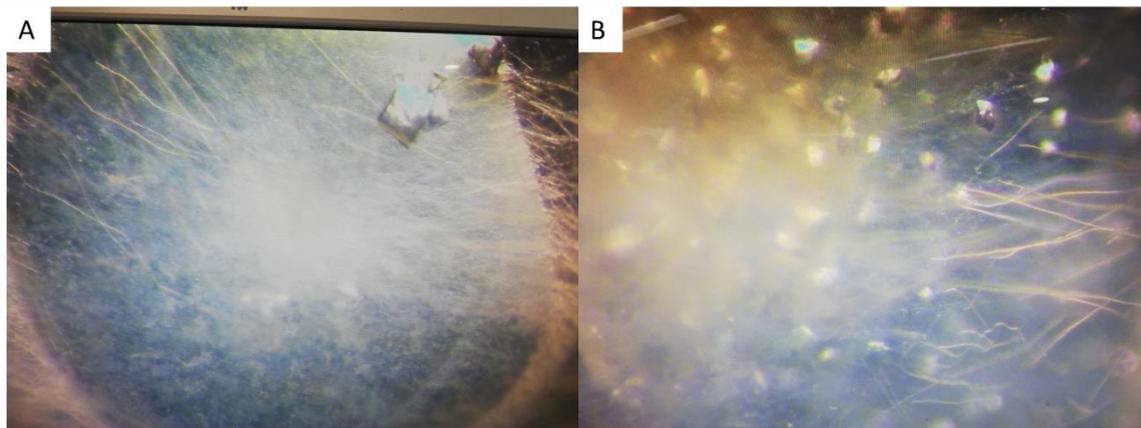
Supplementary Figure S2.7 Crystallisation of recombinant *Lr*SecA2₅₃₆₀₈ (A) 10 mg/mL *Lr*SecA2₅₃₆₀₈ in 0.2 M calcium chloride, 0.1 M Bicine pH 9.0, 25% polyethylene glycol 4000. (B) 10 mg/mL *Lr*SecA2₅₃₆₀₈in 0.1 M sodium chloride, 0.1 M MOPS pH 6.1, 30% polyethylene glycol 400. Protein crystallisation trials were held at 25°C with a ratio of 0.3 µL protein solution : 0.5 µL crystallisation solution.
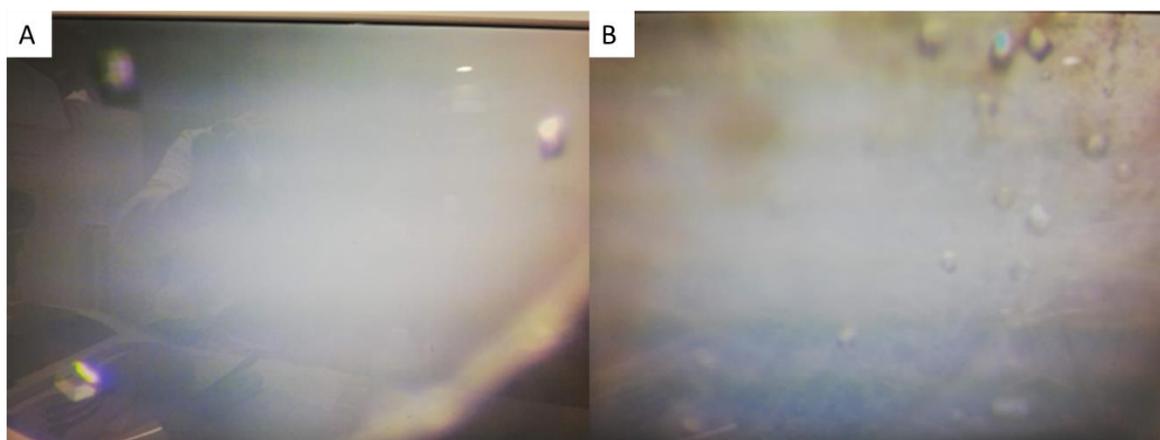
Supplementary Figure S2.8 Crystallisation of recombinant *Lr*SecY2₅₃₆₀₈ (A) 5 mg/mL *Lr*SecY2₅₃₆₀₈ in 0.5 M ammonium sulphate, 10% polyethylene glycol 8000. (B) 10 mg/mL *Lr*SecY2₅₃₆₀₈in 0.2 M magnesium chloride, 0.1 M Tris pH 8.5, 12% polyethylene glycol 4000. Protein crystallisation trials were held at 25°C with a ratio of 0.3 µL protein solution : 0.5 µL crystallisation solution.
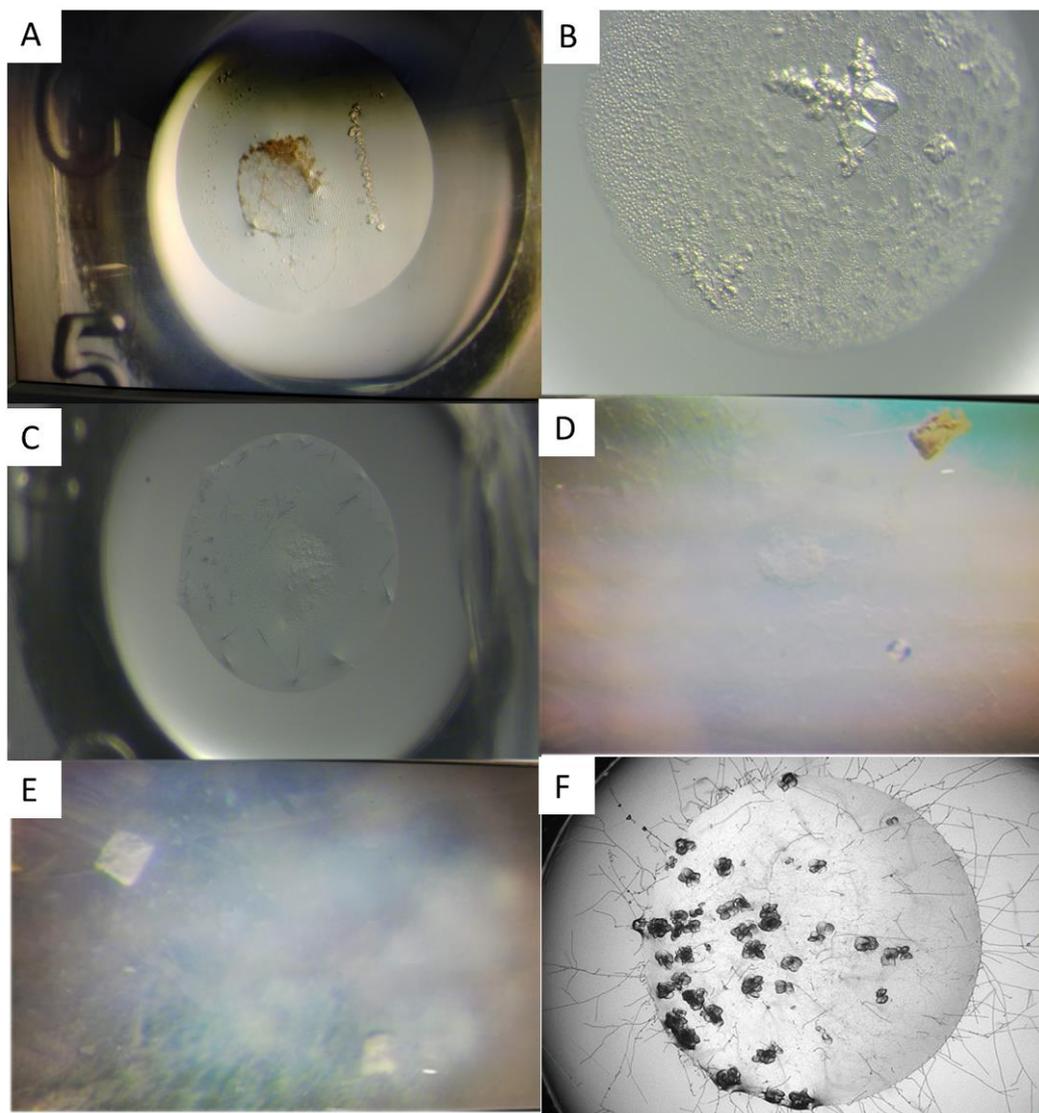
Supplementary Figure S2.9 Crystallisation of recombinant *Lr*SecA2-SecY2-Asp4$_{53608}$. (A) 5 mg/ *Lr*SecA2-SecY2-Asp4$_{53608}$ in 0.1 M Tris, pH 8.0, 0.2 M calcium chloride, 20% polyethylene glycol 6000 (B) 5 mg/mL *Lr*SecA2-SecY2-Asp4$_{53608}$ in 0.1 M MES, pH 6.0, 0.2 M calcium chloride, 20% polyethylene glycol 6000 (C) 10 mg/mL *Lr*SecA2-SecY2-Asp4$_{53608}$ in 0.1 M bis-Tris propane pH 7.0, 0.2 M sodium fluoride, 20% polyethylene glycol 3350 (D) 10 mg/mL *Lr*SecA2-SecY2-Asp4$_{53608}$ in 1.0 M potassium sodium tartrate tetrahydrate, 0.1 M HEPES pH 7.5 (E) 15 mg/mL *Lr*SecA2-SecY2-Asp4$_{53608}$ in 0.2 M calcium chloride, 20% polyethylene glycol 6000. (F) 10 mg/mL *Lr*SecA2-SecY2-Asp4$_{53608}$ in 0.1 M magnesium chloride, 0.1 M sodium acetate pH 5.5, 12% polyethylene glycol 6000.   Protein crystallisation trials were held at 25°C with a ratio of 0.25 µL protein solution : 0.3 µL crystallisation solution.

Appendix 3

Here, the purification process of the recombinant *Lr*aSec proteins produced in this work is detailed providing information on purification steps, presence/absence of his-tags, yield in mg of purified protein per litre (L) of *E. coli* culture (Supplementary Table S3.1). Depending on application, for an instance for certain ELISA based binding assays or for X-ray crystallography, his-tags may have been retained or removed as mentioned in the text.

Supplementary Table S3.1 Summary of recombinant *Lr*aSec proteins produced and purified in this work.

Proteins	Purification steps	His-tag	Purification yields (mg/L)
*Lr*GtfC$_{53608}$	NiNTA and SEGF	N-terminal	10
		3C cleaved	8.5
*Lr*GtfC F174L$_{53608}$ *Lr*GtfC P238S$_{53608}$ *Lr*GtfC C240W$_{53608}$ *Lr*GtfC D101A$_{53608}$	NiNTA	N-terminal	10
*Lr*GtfC$_{100-23}$	NiNTA and SEGF	N-terminal	10
		Thrombin cleaved	8.5
*Lr*GtfC L174F$_{100-23}$ *Lr*GtfC S238P$_{100-23}$ *Lr*GtfC W240C$_{100-23}$ *Lr*GtfC D101A$_{100-23}$	NiNTA	N-terminal	10
*Lr*Asp1$_{53608}$	NiNTA	N-terminal	6
*Lr*Asp2$_{53608}$	NiNTA and SEGF	N-terminal	5
		Thrombin cleaved	4
*Lr*Asp2S349A$_{53608}$	NiNTA and SEGF	N-terminal	5
		Thrombin cleaved	4
*Lr*Asp2$_{100-23}$	NiNTA and SEGF	N-terminal	5
		Thrombin cleaved	4
*Lr*Asp3$_{53608}$	NiNTA	N-terminal	3
*Lr*Asp1-Asp2-Asp3$_{53608}$	NiNTA and SEGF	N-terminal (*Lr*Asp1$_{53608}$)	1
*Lr*SecA2$_{53608}$	NiNTA and benzamidine	N-terminal	6
		Thrombin cleaved	5.5
*Lr*SecA2K114A$_{53608}$	NiNTA and benzamidine	N-terminal	6
		Thrombin cleaved	5.5
*Lr*SecA2$_{100-23}$	NiNTA and benzamidine	N-terminal	6
		Thrombin cleaved	5.5
*Lr*SecY2$_{53608}$	NiNTA	N-terminal	3
*Lr*SecA2-SecY2-Asp4$_{53608}$	NiNTA and SEGF	N-terminal (*Lr*SecY2$_{53608}$)	0.8
*Lr*SP$_{53608}$	NiNTA and benzamidine	Thrombin cleaved	13
*Lr*gSRR1ABC$_{100-23}$ (glycosylated by *Lr*GtfA-GtfB-GtfC$_{100-23}$)	agWGA	N-terminal	3
*Lr*gSRR1AB$_{100-23}$ (glycosylated by *Lr*GtfA-GtfB$_{100-23}$)	agWGA	None	4

Appendix 4

Amino acid alignment using MUSCLE was utilised to classify *L. reuteri* LrGtfC sequences relative to the nature of residues found at position 174, 238 and 240 (Figure S4.1).

Figure S4.1. MUSCLE-alignment of *Lr*GtfC amino acid sequences. *Lr*GtfC amino acid sequences from 18 *L. reuteri* strains were aligned using MUSCLE. A range between residues 172 and 315 are shown with arrows pointing to amino acid at position 174, 238 and 240.