Structure-function studies of Group A streptococcal pilus subunits

by

Jonathan Anthony Pointon

Thesis submitted for the degree of Doctor of Philosophy

Institute for Cell and Molecular Biosciences
Newcastle University

September 2010

Abstract

Streptococcus pyogenes (Group A *Streptococcus*, GAS) is an important Gram positive human pathogen that produces a multitude of virulence factors, including cell-surface pili that were first described shortly before the beginning of this project. Pili in M1 GAS strain SF370 mediate adhesion to human tonsil and skin and are composed of a single major protein subunit (Spy0128) forming the pilus shaft and two minor subunits (Spy0125 and Spy0130), whose roles had not been defined. This thesis reports combined molecular microbiology and protein structure studies that provide insights in to the structures and functions of the M1 GAS strain SF370 pilus subunits.

Recombinant Spy0125 (rSpy0125), rSpy0128 and rSpy0130 were purified to apparent homogeneity and low resolution structures were obtained using a combination of biophysical techniques. Novel *intra*-molecular isopeptide bonds discovered in the Spy0128 crystal structure, reported by others during this project, were examined here for their affects on GAS adhesion. This was achieved by combining site-directed and allele replacement mutagenesis to introduce point mutations into the *spy0128* gene in the GAS chromosome. Comparison of parent and mutant strains showed that these bonds were not essential for adhesion of GAS to human keratinocytes *in vitro*, though the pattern of adhesion appeared altered.

Most of the work in this project focused on Spy0125, which was first localized at the tip of the pilus and shown to act as the adhesin. Allele replacement mutagenesis confirmed that the adhesin resided within a stable c. 50 kDa polypeptide corresponding to the C-terminal 2/3 of intact Spy0125. A recombinant version of this region, rSpy0125-CTR, was produced and its high resolution crystal structure determined. In addition to internal *intra*-molecular bonds similar to those recently found by others in Spy0128, the rSpy0125-CTR structure revealed an internal thioester bond between a Cys and a Gln residue that is unprecedented outside of complement and complement-like proteins. Whereas current paradigms of bacterial-host interactions suggest non-covalent forces are involved, the presence of a thioester in Spy0125 reveals for the first time that strong covalent forces may also play a role.

Acknowledgements

There are a number of people who I would like to thank for helping me along on this journey. Firstly Mike, who has been a source of great and constant inspiration, always nudging me in the right direction. Mark for his support over the four years, always available at the end of the telephone to answer my crystallography queries, no matter how small. The members of the Kehoe laboratory, especially Wendy Smith, whose seemingly 'Midas' touch with molecular biology was greatly received and Emily Abbot for helping me get to grips with confocal microscopy. Paul Race for having the patience to put up with all of my questions and his general banter in the lab., along with the rest of the structural biology group for the many crystallography related discussions. Alexandra Solovyova for her help with the biophysical solution work and Joe Gray at Pinnacle for his help with the mass spectrometry data. Also, the Medical Research Council for providing the studentship (and paying the bills!).

Finally, I would also like to thank my close family and friends who are always there for me and of course, Lynsey, for her continued love and support.

Table of Contents

Table of Figures

Table of Tables

Abbreviations and symbols

Amino acids may be abbreviated to either their standard single or three letter code. Atoms and compounds may be referred to by their standard biochemical symbols.

a, b, c, α, β, γ	Real space unit cell dimensions.
Å	Angstrom units (1 x 10^{-10} metre).
ARF	Acute rheumatic fever.
ATP	Adenosine triphosphate.
AUC	Analytical ultracentrifugation.
B-factor	Temperature factor.
BSA	Bovine serum albumin.
CAPS	3-(Cyclohexylamino)-1-propanesulfonic acid.
CCD	Charged-coupled device.
CCP4	Collaborative computational project number 4.
CD	Circular dichroism spectroscopy.
CTR	C-terminal region.
CWSS	Cell wall sorting signal.
DLS	Diamond light source.
DTT	Dithiothreitol.
EDTA	Ethylene diamine tetra-acetic acid.
EMBL	European molecular biology laboratory.
Erm	Erythromycin.
ESRF	European synchrotron radiation facility.
ESU	Estimated standard uncertainties.
$\|F_{obs}\|$	Observed structure factor amplitude.
$\|F_{calc}\|$	Calculated structure factor amplitude.

FCT	Fibronectin binding, collagen binding, T-antigen.
FOM	Figure of merit.
GAS	Group A *Streptococcus*.
GBS	Group B *Streptococcus*.
GSP	General secretion pathway.
h	Hours.
HaCaT	Human adult skin keratinocytes propagated under low Ca^{2+} conditions and elevated temperature.
h, k, l	Reciprocal lattice points.
ID	Internal deletion.
IPTG	Isopropyl β-D-1-thiogalactopyranoside.
kDa	1000 Daltons.
kb	1000 bases.
LB	Luria-Bertani.
LTA	Lipoteichoic acid.
M	Molar.
MAD	Multi-wavelength anomalous diffraction.
MES	2-(N-morpholino)ethanesulfonic acid
min	Minutes.
MIR	Multiple isomorphous replacement.
MME	Monomethyl ether.
NCS	Non-crystallographic symmetry.
NMR	Nuclear magnetic resonance spectroscopy.
dNTP	Deoxynucleotide triphosphate
NTR	N-terminal region.

OD	Optical density.
PBS	Phosphate buffered saline.
PCR	Polymerase chain reaction.
PDB	Protein data bank.
PEG	Polyethylene glycol.
PSAGN	Post streptococcal acute glomerulonephritis.
PVDF	Polyvinylidene fluoride.
R_{work}	Crystallographic R-factor.
R_{free}	Free R-factor.
rmsd	root mean square deviation.
SAD	Single-wavelength anomalous diffraction.
SAXS	Small angle X-ray scattering.
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis.
SeMet	Selenomethionine.
SIC	Streptococcal inhibitor of complement.
SIRAS	Single isomorphous replacement with anomalous scatter.
SLO	Streptolysin O.
SLS	Streptolysin S.
SRP	Signal recognition particle.
SRS	Synchrotron radiation source.
STSS	Streptococcal toxic shock syndrome.
TEMED	N,N,N',N'-Tetramethylethylenediamine
THY	Todd Hewitt broth with yeast extract.
TLS	Translation, libration and screw.

Table 1.2: GAS virulence factors

Location	Virulence factor	Properties	Producing strain
Secreted	Streptolysin S (SLS)	2.8 kDa non-immunogenic polypeptide that inserts into cellular membranes leading to osmotic cell lysis. Also the primary factor causing the characteristic zones of β-haemolysis surrounding GAS colonies on blood agar plates (Nizet, 2002).	All GAS strains
	Streptolysin O (SLO)	57 kDa protein which is able to form extremely large pores of up to 30 nm containing up to 80 monomers of SLO (Palmer *et al.*, 1998; Kehoe *et al.*, 1987). It was revealed that SLO pores were needed for the translocation and possible activation of an effector protein, nicotinamide adenine dinucleotide glycohydrolase (NADase) in a process termed cytolysin mediated translocation (Magassa *et al*., 2010; Madden *et al*., 2001).	All GAS strains
	NADase	Possesses ADP-ribose cyclase activity, cleaving the nicotinamide-ribosyl bond of NAD, producing an intermediate that affects many cellular functions including signal transduction and gene expression eventually leading to cell death (Ghosh *et al.*, 2010).	All GAS strains
	Streptococcal pyrogenic exotoxins; SpeA, SpeC, SpeG-M, SSA, and SMEZ	Encoded by mobile phages except SpeG, SpeJ and SMEZ which are chromosomally located (Fraser and Proft, 2008). These superantigens exert their effect by binding major histocompatability complex class II molecules on the surface of antigen presenting cells where they bind and cross link multiple T-cell receptors at external sites, distinct from those involved in conventional antigen presentation (Sundberg *et al.*, 2002; McCormick *et al.*, 2001). This activates large numbers of T cells leading to a cytokine 'storm' exerting the main pathology of STSS.	Most strains produce one or more. Different combinations in different strains
	SpeB	Secreted cysteine protease that is able to degrade a large number of host serum proteins (Chiang-Ni and Wu, 2008) in addition to a number of GAS proteins. Cleavage of surface associated virulence factors such as C5a peptidase and the M-protein may occupy the host immune system at a distance rather than at the cell surface. SpeB has also been shown to degrade streptococcal superantigens (Nooh *et al.*, 2006).	All GAS strains
	IdeS/Mac and EndoS	Degrade IgG antibodies protecting the organism from opsonisation and phagocytosis (von Pawel-Rammingen and Bjorck, 2003).	Some strains

Secreted	Hyaluronidase	Cleaves the repeating disaccharide of N-acetylglucosamine and D-glucuronic acid that comprises the hyaluronic acid found within human connective tissue and the GAS capsule (Hynes and Walton, 2000; Sandson *et al*., 1968). Proposed to act as a 'spreading factor' for the bacteria (Stevens, 2000) although recent evidence has suggested that hyaluronidase contributes to the spread of protein virulence factors rather than the bacteria (Starr and Engleberg, 2006).	All GAS strains
	Streptokinase	Activates plasminogen that may be free in plasma or bound to surface associated M-protein, GAPDH and α-enolase (see below). Plasminogen is converted to plasmin which acts as a protease on a wide range of substrates, for example, dissolving blood clots which could facilitate bacterial spread throughout human tissues (Svensson *et al.*, 2002).	All GAS strains
	Streptococcal inhibitor of complement (SIC)	Binds soluble C5b-7 membrane attack complex and prevents its insertion into the GAS membrane (Fernie-King *et al.*, 2001; Akesson *et al.*, 1996). Other possible roles for SIC include inhibiting secretory leukocytes, lysozymes, α-defensins and LL-37 anti-microbial peptides (Rooijakkers and van Strijp, 2007; Frick *et al.*, 2003; Fernie-King *et al.*, 2002).	Very few GAS strains
	DNase	GAS are able to circumvent neutrophil extracellular traps (NETs) by the production of four DNases, the most potent of which, DNase1, is crucial for the evasion of neutrophil killing (Buchanan *et al*., 2006; Brinkmann *et al*., 2004).	All GAS strains
Cell wall associated	M-protein	Considered to be the major virulence factor for GAS. Composed of four repeat regions; a highly conserved C-terminal cell wall anchoring region in the C and D repeats, a variable B repeat region and a hypervariable A repeat region which confers strain specificity (Kehoe *et al.*, 1996; Fischetti, 1989). M-protein facilitates adhesion to host cells and contributes to persistence within the host through resistance to phagocytosis and recruitment of host proteins.	All GAS strains at least one
	Enn and Mrp	Functionally related to M-protein. Could play a role in immune evasion.	All strains
	C5a peptidase	A serine protease, highly specific for complement component C5a which is a key chemotactic peptide, cleaving it between His67 and Lys68, residues which are not accessible on inactive C5a, thus reducing phagocyte recruitment to the site of GAS infection (Cleary *et al.*, 1992b).	All GAS strains
	Capsule	Composed of hyaluronic acid and is a key defence mechanism in preventing phagocytosis (Wessels and Bronze, 1994). Strains engineered to produce no capsule are avirulent and easily cleared (Cunningham, 2000). A role for the capsule in GAS adhesion to the CD44 receptor on epithelial cells has been suggested (Stollerman and Dale, 2008; Schrager *et al*., 1998).	Likely by all GAS strains

Cell wall associated	Lipoteichoic acid (LTA)	Believed to facilitate a weak reversible adhesion with epithelial cell membranes, probably through hydrophobic interactions (Sela *et al.*, 2000).	All strains
	Extra cellular matrix binding proteins	Include the fibronectin binding proteins (Fbp); PrtF1 (Hanski and Caparon, 1992), PrtF2 (Kreikemeyer *et al.*, 2004), serum opacity factor (Courtney *et al.*, 2009), Fbp54, FbaA, SfbX and others with ten fibronectin binding proteins described in total, collagen binding proteins (Cpa) (Kreikemeyer *et al.*, 2005) and laminin binding proteins (Lbp) (Terao *et al.*, 2002).	Fbp-Many Cpa-Some Lbp-Many
	Protein G-related α_2M-binding (GRAB)	Bind free α_2-macroglobulin *via* critical arginine residues on the protein surface (Godehardt *et al.*, 2004). The α2-macroglobulins are broad range protease inhibitors which remain active when bound to protein GRAB and could protect GAS surface proteins from host proteases (Godehardt *et al.*, 2004).	Few strains
	GAPDH and α-enolase	Usually found in the cytoplasm and have no N-terminal secretion signal peptide but can be released from the cell by a currently unidentified mechanism (some suspect simply cell lysis, while others suspect a novel active secretion mechanism). Often found tightly associated with the cell wall (Pancholi and Fischetti, 1998; Pancholi and Fischetti, 1992). GAPDH binds fibronectin, lysozyme, myosin and actin as well as weakly to plasmingen (Pancholi and Fischetti, 1992). α-enolase binds plasminogen and could play a role in bacterial spread (Pancholi and Fischetti, 1998).	All GAS strains

Table 2.3 Synthetic oligonucleotide primers used in this study

Primer[a]	Sequence (5' to 3')[b]	Location w.r.t nucleotide 1 of specified ORF[c]	Purpose
JP01 (F)	ctaggatccAAGACTGTTTTTGGTTTAGTAG	M1 Spy0125: 142 to 163	Cloning M1 *spy0125-NTR* into pET28a
JP01 (R)	atgttctcgag**TTA**ACCACCACTCAAAAGATT	790 to 807	
JP02 (F)	ctaggatccAATCAACCTCAAACGACTTCAG	M1 Spy0125: 856 to 877	Cloning M1 *spy0125-CTR* into pET28a
JP02 (R)	atgttctcgag**TTA**TGTAGGAACAACAGGCTC	2152 to 2169	
JP03 (F)	ttacatactgcagCTTTAAAGCAATATTTTCTCA	M1 Spy0125: -122 to -102	Constructing *Δspy0125-NTR* deletion in pG$^+$host9
JP03 (R)	attagtcgacACCAAAAACAGTCTTAGCAC	137 to 156	
JP04 (F)	attagtcgacAATCAACCTCAAACGACTT	856 to 874	
JP04 (R)	atgttctcgagAATCTGTTTTCCATCAATAA	1115 to 1134	
JP05 (F)	CAGTTATTGGCAAATAAGATGTC	M1 Spy0125: -350 to -328	Sequencing primers to confirm *Δspy0125-NTR* deletion in chromosome
JP05 (R)	CAAGCAGGTGATGAAGTGGA	2579 to 2598	

Primer[a]	Sequence (5' to 3')[b]	Location w.r.t nucleotide 1 of specified ORF[c]	Purpose
JP06 (F) JP06 (R) JP07 (F) JP07 (R)	ttacatactgcagGTTATGTATAATGGACATCCACAA attagtcgacGACCCACCAATGCCTCCA attagtcgacCAGCTAACTGACCTTGATTTC atgttctcgagGAAACAGATTCTGAAGGCTAT	M1 Spy0125: 472 to 495 838 to 855 1678 to 1698 2038 to 2058	Constructing *Δspy0125-ID* in pG$^+$host9
JP08 (F) JP08 (R)	attgaattcCGATATGGCTCCTAGTGTAA aagttctcgagGATGTGCATAGATTAATGAG	M1 Spy0128: -411 to -392 1392 to 1411	Constructing *spy0128* + flanking regions in pG$^+$host9
JP09 (F) JP09 (R)	AATAGCCCTTATTCTTATTGG GTTTCAATACTAAATTCTTTGTG	M1 Spy0128: -483 to -463 1477 to 1499	Sequencing primers to confirm wild-type and mutant *spy0128* in chromosome
JP10 (F) JP10 (R)	AAACTAACAGTTACAGCAAACCTTGATTTAGTTAA TTAACTAAATCAAGGTTTGCTGTAACTGTTAGTTT	M1 Spy0128: 91to 125	Constructing K36A mutation in Spy0128

Primer[a]	Sequence (5' to 3')[b]	Location w.r.t nucleotide 1 of specified ORF[c]	Purpose
JP11 (F) JP11 (R)	TATTACAAAGTAACTGCGGAGAAGATAGATAAAGTTC GAACTTTATCTATCTTCTCCGCAGTTACTTTGTAATA	M1 Spy0128: 334 to 370	Constructing E117A mutation in Spy0128
JP12 (F) JP12 (R)	TATAAAGAAGGTAGTGCGGTGCCAATTCAGTTC GAACTGAATTGGCACCGCACTACCTTCTTTATA	M1 Spy0128: 466 to 498	Constructing K161A mutation in Spy0128
JP13 (F) JP13 (R)	ACATTAACGGTGAAGGCAAAAGTTTCAGGTACC GGTACCTGAAACTTTTGCCTTCACCGTTAATGT	M1 Spy0128: 520 to 552	Constructing K179A mutation in Spy0128
JP14 (F) JP14 (R)	GATTATGTTGTCACTGCAGACGATTACAAATCAG CTGATTTGTAATCGTCTGCAGTGACAACATAATC	M1 Spy0128: 757 to 790	Constructing E258A mutation in Spy0128
JP15 (F) JP15 (R)	attgaattcCTGGAAATGTTATGTTTGG aagttctcgagGCTATCATCATGAGTTTGC	M1 Spy0130: -350 to -332 984 to 1002	Constructing *spy0130* + flanking regions in pG^{+}host9

Primer[a]	Sequence (5' to 3')[b]	Location w.r.t nucleotide 1 of specified ORF[c]	Purpose
JP16 (F) JP16 (R)	GTTAAGGGAAATGGGGCAACATCATTTGAACAG CTGTTCAAATGATGTTGCCCCATTTCCCTTAAC	M1 Spy0130: 229 to 261	Constructing K82A mutation in Spy0130
JP17 (F) JP17 (R)	CATCAACTGTTAGGTGCGAATAGTCAATATCATT AATGATATTGACTATTCGCACCTAACAGTTGATG	M1 Spy0130: 301 to 334	Constructing K106A mutation in Spy0130
JP18 (F) JP18 (R)	AACCTAGTTTCTAACGCACTTGGAGAAACCGAA TTCGGTTTCTCCAAGTGCGTTAGAAACTAGGTT	M1 Spy0130: 403 to 435	Constructing K140A mutation in Spy0130
JP19 (F) JP19 (R)	CTTGGAGAAACCGAAGCATCGGAGCTTATTTTTA TAAAAATAAGCTCCGATGCTTCGGTTTCTCCAAG	M1 Spy0130: 421 to 454	Constructing K146A mutation in Spy0130
JP20 (F) JP20 (R)	TCGGAGCTTATTTTTGCACAAGAATATAGTGAA TTCACTATATTCTTGTGCAAAAATAAGCTCCGA	M1 Spy0130: 439 to 471	Constructing K152A mutation in Spy0130

Primer[a]	Sequence (5' to 3')[b]	Location w.r.t nucleotide 1 of specified ORF[c]	Purpose
JP21 (F) JP21 (R)	CAAGAATATAGTGAAGCAACACCGGAACCTCATC GATGAGGTTCCGGTGTTGCTTCACTATATTCTTG	M1 Spy0130: 457 to 490	Constructing K158A mutation in Spy0130
JP22 (F) JP22 (R)	CCAGATACAACTGAGGCGGAAAAACCTCAGAA TTCTGAGGTTTTTCCGCCTCAGTTGTATCTGG	M1 Spy0130: 493 to 524	Constructing K170A mutation in Spy0130
JP23 (F) JP23 (R)	AGTACTCCTTATCAGGAAGTG TGTTCTGTCTAACAATATGGC	M1 Spy0130: -440 to -420 1082 to 1102	Sequencing primers to confirm wild-type and mutant *spy0130* in chromosome

[a] (F) refers to the forward primer and (R) refers to the reverse primer of the corresponding pair.

[b] DNA sequence of primers. Genomic encoded DNA is shown in capitals and extra 'tag' sequences are shown in lower case and contain restriction endonuclease sites which are underlined. Termination codons included in reverse primers to stop potential read through during recombinant proteins production are highlighted in bold.

[c] M1 refers to the M1 GAS strain SF370 used in this study and Spy numbers are the original numbers assigned during genome annotation (Ferretti *et al*., 2001).

```
M1     1 MKKTRFPNKLNTLNTQRVLSKNSKRFTVTLVGVFLMIFALVT--SMVGAKTVFGLVESSTPNAINPDSSSEYRWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLG 118
M28    1 ---MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMLLVLIG-FEGKVRAAHELVEVPVPIFHNPDPQSDYQWYGYEAYTGGYPKYDLFKTYYHDLRVNLHGSKSYQVYCFNVHKHYPRS 116
M12    1 ---MNNKKLQKKQDAPRVSNRKPKQLTVTLVGVFLMFLTLVSSMRGAQSIFGEEKRIEEVSVPKIKSPDDAYPWYGYDSYDSSHPYYERFKVAHDLRVNLNGSKSYQVYCFNINSHYPNR 117
M49    1 -------MQKRDKTNYRSANNKRRQTTIGLLKVFLTFVALIG----IVGFSIRAFGAEEQSVPNRQSSIQDYPWYGYDSYPKGYPDYSPLKTYHNLKVNLEGSKDYQAYCFNLTKHFPSK 109
M3     1 -------MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIG----IVGFSIRAFGAEEQSVPNKQSSVQDYPWYGYDSYSKGYPDYSPLKTYHNLKVNLDGSKEYQAYCFNLTKHFPSK 109
M5     1 -------MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIG-IVGFS---------------------------------------------------------------------  40
M18    1 -------MQKRDKTNYGSANNKRRQTTIGLLKVFLTFVALIG----IVGFSIRAFGAEEQ------------------------------------------------------------  49

M1   119 SDSSVKKWYKKHDGISTKFEDYAMSPRITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPK 238
M28  117 SQSFDRKWYKKLDGTAENFDSLAMEPRVRKEELTKKLRAVMYNAYPNDANGIMKDLEPLNAIKVTQEAVWYYSDSAQIN-PDESFKTEAQSNGINDQQLGLMRKALKELIDPNLGSKYSN 235
M12  118 KNAFSKQWFKRVDGTGDVFTNYAQTPKIRGESLNNKLLSIMYNAYPKNANGYMDKIEPLNAILVTQQAVWYYSDSSYGN-IKTLWASELKDGKIDFEQVKLMREAYSKLISDDLEETSKN 236
M49  110 SDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRILYNGYPNNRNGIMKGIDPLNAILVTQNAIWYYTDSAQIN-PDESFKTEARSNGINDQQLGLMRKALKELIDPNLGSKYSN 228
M3   110 SDSVRSQWYKKLEGTNENFIKLADKPRIEDGQLQQNILRILYNGYPNDRNGIMKGIDPLNAILVTQNAIWYYTDSSYISDTSKAFQQEETDLKLDSQQLQLMRNALKRLINPKEVESLPN 229
M5    40 ----------------------------------------------------------------------------------------------------------------------  40
M18   49 ----------------------------------------------------------------------------------------------------------------------  49

M1   239 QVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPTPGDPPMPPNQPQTT 291
M28  236 KTPSGYRLNVFESHDKT------FQNLLSAEYVPDTPPKPGE--EPPAKTEKT 280
M12  237 KLPQGSKLNIFVPQDKS------VQNLLSAEYVPESPPAPGQSPEPPVQTKKT 283
M49  229 KTPSGYRLNVFESHDKT------FQNLLSAEYVPDTPPKPGE--EPPAKTEKT 273
M3   230 QVPANYQLSIFQSSDKT------FQNLLSAEYVPDTPPKPGE--EPPAKTEKT 274
M5    41 ----------------------------------IRAFGAE--EKSTETKKT  56
M18   50 ---------------------------------------STETKKT  56
```

Figure 4.24: ClustalW sequence alignment of M1 Spy0125-NTR homologues.

Signal peptides are shown in red with mature protein sequence in black. Sequence shown includes residues up to the beginning of the Spy0125-CTR as identified in section 4.4 (NQPQTT) highlighted in yellow. Alignment performed using ClustalW (Larkin et al., 2007).

Figure 4.25: ClustalW sequence alignment of M1 Spy0125-CTR homologues.

Residues denoting the N-terminus of the defined Spy0125-CTR fragment (Asn286-Thr291) are highlighted in yellow. Note proline rich sequence between residues equivalent to Lys274-Asn286 from Spy0125 in homologues that have an N-terminal region. Aligned sequences finish at the 'VVPTG' motif. Alignment performed using ClustalW (Larkin et al., 2007).

Table 6.1: Summary of the major types of Gram negative and Gram positive pili.

Gram type	Pilus type/ Organism	Properties	Target receptor/Ligand
Gram negative	Type I/P pili *E. coli* UPEC[a] *H. influenzae* *P. mirabilis* *Y. pestis*	Assembled by the chaperone/usher pathway. Assembly requires the adhesin, PapG, and three minor subunits, PapE, PapF and PapK, to form the tip fibrillum. The shaft is formed by PapA anchored in the outer membrane by PapH. Subunits are secreted into the periplasm by the GSP and complexed by the chaperone, PapD, which transports them to the usher molecule in the outer membrane, PapC (Wu and Fives-Taylor, 2001). PapA monomers consist of an incomplete Ig-like fold and ~15 residue N-terminal extension. This extension completes the Ig-like domain of the pre-ceding monomer, stabilised by hydrophobic interactions (Sauer *et al*., 2002). The Ig-like C-terminal domain of PapG is anchored at the tip of the pilus, as for PapA, while the ligand binding N-terminal domain adopts a similar β-barrel 'jelly' roll domain (Kline *et al*., 2010).	P pili PapG binds Gal α1-4 Gal containing glycolipids in human kidney. Type I adhesin FimH binds mannosylated receptor on bladder epithelium surface.
Gram negative	Type IV pili *N.gonorrhoeae* *N. meningitidis* *P. aeruginosa* *F. tularensis* *V. cholerae* EPEC[a] ETEC[a]	Subunits translocated into the periplasm by the GSP are retained in the cytoplasmic membrane by an N-terminal signal peptide (subsequently cleaved by PilD). Assembled from the cytoplasmic membrane up, PilG and the PilB ATPase form the base where major subunits (PilA) are recruited one at a time but at three sites simultaneously. The pilus is pushed into the periplasm and through a PilQ pore in the outer membrane by a conformational change in PilB, resulting from hydrolysis of one ATP. This leaves a gap for the next three PilA monomers (Proft and Baker, 2009; Mattick, 2002). Major subunits are composed of a globular β-sheet domain and extended N-terminal α-helix. The C-terminal portion of this helix embeds in the globular domain of the pre-ceding monomer. Some major subunits also harbour a disulphide bond in the globular domain (Craig and Li, 2008).	Role in adhesion less well defined than P pili (Rego *et al*., 2010). Could be facilitated *via* formation of biofilms as Type IV pili interact laterally to form characteristic bundles (Pelicic *et al*., 2008).
Gram negative	Curli *E. coli* *Salmonella spp.*	Formed by the extracellular nucleation/precipitation pathway. Major subunit monomers, CsgA, are secreted into the external milieu by the action of three proteins, CsgE, CsgF and CsgG. Here, they interact with the nucleator protein, CsgB, triggering polymerisation of CsgA monomers at the cell surface (Wu and Fives-Taylor, 2001). Polymerised fibres are β-sheet rich, non-branching and resistant to protease degradation (Proft and Baker, 2009).	Fibronectin (Olsen *et al*., 1989) Plasminogen (Sjobring *et al*., 1994)

Gram negative	CS1 pili ETEC[a]	Assembled by the alternative chaperone/usher pathway. Biogenesis is similar to the formation of P pili although the four proteins required for assembly show no sequence similarity with those of the chaperone/usher pathway (Proft and Baker, 2009).	Unknown.
Gram positive	GAS	Pili in M1 GAS strain SF370 (used in this study) are heterotrimeric, assembled by SrtC and anchored to the cell wall by SrtA, consisting of Spy0130 at the base, Spy0128 forming the shaft, and Spy0125 at the tip. FctB (a Spy0130 homologue) is an all β-domain with an Ig-like fold and polyproline like II helix tail (Linke *et al*., 2010) while Spy0128 is an extended monomer of two (CnaB[b]-like) all β-domains that each accommodate an *intra*-molecular isopeptide bond (Kang *et al*., 2007). Spy0125-CTR was shown in this thesis composed of two β-sandwich domains (each containing an isopeptide bond) and a third domain (with no structural homologues) harbouring an internal thioester bond (Pointon *et al*., 2010).	Unknown. In M1 GAS strain SF370 pili are critical for adhesion to tonsil and cultured keratinocytes (Abbot *et al*., 2007).
Gram positive	GBS	Pili expressed from each of the three pilus islands (PI) form heterotrimeric pili with a single minor subunit at the tip and the base in addition to a major subunit forming the shaft. GBS52 (likely pilus anchor) was revealed to be composed of two CnaB-like (modified Ig) domains with an isopeptide bond in the N2 domain (Krishnan *et al*., 2007). Each PI encodes two sortases, believed to show distinct specificities in incorporating the minor subunits while showing redundancy in polymerising the major subunit (Dramsi *et al*., 2006).	Unknown. Binding to A549 (Krishnan *et al*., 2007) and vhBMEC cells (Maisey *et al*., 2007) *in vitro*.
Gram positive	*S. pneumoniae*	RlrA pili are anchored to the cell wall through RrgC, while polymerised RrgB monomers (containing pilin and E-box motifs) form the pilus shaft. These proteins are suggested to contain two (RrgC) and three (RrgB) *intra*-molecular isopeptide bonds (El Mortaji *et al*., 2010). The adhesin, RrgC, is an elongated monomer of 190 Å on its longest axis, composed of three CnaB-type domains (two of which harbour isopeptide bonds). In addition, a MIDAS motif containing von Willebrand factor domain is positioned at the top of the subunit (Izore *et al*., 2010). These domains have previously been shown to play a role in collagen binding (Emsley *et al*., 2000).	Unknown. The presence of a MIDAS motif in RrgC suggests collagen.
Gram positive	*C. diphtheriae*	The SpaC adhesin (of SpaABC pili) is attached at the pilus tip to the major subunit, SpaA, by SrtA which also polymerises SpaA monomers into pili (Ton-That and Schneewind, 2003). SrtF (the housekeeping sortase) anchors the pilus to the cell wall *via* SpaB (Mandlik *et al*., 2008a). SpaA was revealed to contain a single CnaA[c] (isopeptide harbouring) domain and two CnaB domains, one of which contained an isopeptide and a disulphide bond.	Unknown. SpaABC pili shown to mediate adhesion to D562 pharyngeal

		Lys190 of the SpaA pilin motif was shown to be crucial for polymerisation while the glutamate of the E-box motif was revealed as the catalytic residue for *intra*-molecular isopeptide bond formation (Kang *et al*., 2009b).	cells *in vitro* (Mandlik *et al*., 2007).
Gram positive	*E. faecalis*	Encode endocarditis and biofilm associated pili (Ebp) on the chromosome in addition to biofilm enhancer in enterococci (Bee) pili, on a conjugative plasmid (Nallapareddy *et al*., 2006, Tendolkar *et al*., 2006). The major subunit, EbpC, contains both the pilin and E-box motifs (Nallapareddy *et al*., 2006). Pilus assembly in enterococci has not been extensively studied, although both are expected to form heterotrimeric pili in a similar manner to other Gram positive pili, with the help of pilus associated sortases (Kreikemeyer *et al*., 2010).	Unknown. Pili appear to contribute to adhesion through biofilm formation
Gram positive	*E. faecium*	Genome encodes four distinct pilus gene clusters (PGC 1-4) expected to form heterotrimeric pili catalysed by pilus associated sortases (Kreikemeyer *et al*., 2010; Hendrickx *et al*., 2008). Pilin and E-box motifs are harboured by all four major subunits except *orf904.5* of PGC-4 which does not contain an E-box motif (Hendrickx *et al*., 2009; Hendrickx *et al*., 2008).	Unknown.
Gram positive	*B. cereus*	Pili are heterodimeric composed of the adhesin, BcpB, positioned at the pilus tip and the major subunit, BcpA, polymerised by the pilus associated sortase, SrtD and anchored to the cell wall by SrtA (Budzik *et al*., 2009a; Budzik *et al*., 2008; Budzik *et al*., 2007). The crystal structure of BcpA revealed two CnaB (Cna_2 and Cna_3) domains and an XNA 'jelly' roll fold domain (Budzik *et al*., 2009b). In addition, each domain (including the Cna_1 domain not included in the crystallised construct) was shown to harbour an *intra*-molecular isopeptide bond. (Budzik *et al*., 2009b; Budzik *et al*., 2008).	Unknown.

[a] UPEC-uropathogenic *E. coli*; EPEC-enteropathogenic *E. coli*; ETEC-enterotoxigenic *E. coli*.

[b] In a CnaB-like domain, the isopeptide bond links the first and last β-strands of the domain.

[c] In a CnaA-like domain, the isopeptide bond links the first and penultimate β-strands of the domain.

Chapter 1. General introduction

1.1 Introduction to *Streptococcus pyogenes*

Streptococcus pyogenes (Group A *Streptococcus*, GAS) are Gram positive, non-motile, catalase negative cocci, 0.5-1.0 μM in diameter, which grow in chains. The organism is a facultative anaerobe which is aero-tolerant and a homo-fermenter, producing lactic acid. Optimal growth is achieved at 37^{o}C in nutrient rich media.

Diseases which were most likely to have been caused by GAS were recognised long before streptococci were first described; for example, erysipelas (from the Greek meaning red fire) has been described as long as there have been written records of disease. The term *Streptococcus* was first derived by Billroth (from the Greek: *streptos*, meaning chain, and *kohkos*, meaning berry) in 1874 (Denny, 2000). These organisms were recognised as being responsible for a great number of human infections during the last half of the 19th century (Denny, 2000). These included the greatly feared puerperal fever ('childbed fever'), where outbreaks often had a mortality rate of 50% and scarlet fever with mortality rates of 25-30% (Denny, 2000). Severe post-streptococcal sequelae such as acute rheumatic fever (ARF) and acute glomerulonephritis (PSAGN) were first described over 100 years ago (Martin and Green, 2006). Associations between erysipelas, puerperal fever, scarlet fever and acute rheumatic fever with severe pharyngitis had been noted during the 19th century (Denny, 2000) but the significance of these were not originally understood. Consequently, bacteria isolated from these diseases were frequently named after the condition they caused (*Streptococcus epidemicus, S. erysipelatus*, *S. puerperalis*, *S. scarlatinae*, *S. rheumaticus*, *S. pyogenes*). It was not until detailed classification work had been carried out that it was discovered a single organism was able to cause all of these and a variety of other diseases.

The first steps towards classification of these organisms came when the various kinds of haemolysis produced by different streptococci on blood agar plates were described. These were subsequently defined as; alpha (incomplete or partial), beta (clear zones of 2-4 mm) and gamma (no) haemolysis (Gray, 1998). Streptococci isolated from the diseases described above all displayed β-haemolysis and the collective name *S.*

haemolyticus was used throughout the early 20th century. Around this time definitive studies in typing were carried out. One was 'Lancefield sero-grouping' to distinguish between β-haemolytic streptococci based on immunological differences in their cell wall polysaccharides (Lancefield, 1928b). There are now 15 recognised sero-groups (A-O) based on polysaccharides (groups A, B, C, E, F and G), teichoic acids (groups D and N), or lipoteichoic acids (group H) (Nobbs *et al.*, 2009). These groups do not include all streptococci however, with *Streptococcus pneumoniae* a notable example. Some Lancefield groups contain only a single species (such as Group A) whereas others contain several species (such as Group C). Most β-haemolytic streptococci isolated by Lancefield from serious human infection, belonged to Group A and it was decided to use the name *Streptococcus pyogenes* from the list of names outlined above for this organism.

Lancefield also developed a second typing system to distinguish between Group A streptococcal strains, according to a protease-sensitive 'M-antigen' in the cell wall (Lancefield, 1928a). This method involved boiling the bacteria in hot-HCl to release the 'M-antigen' prior to determining the specific M-antigen serotype (>80 were eventually defined) using standardised typing sera in capillary precipitin tests. At around the same time, another typing system based on antigenic differences in a trypsin resistant antigen, present in the cell wall, called the T antigen was developed (Griffith, 1934). Only 25 serotypes of T antigen have been defined to date and it is common for a streptococcal strain to produce more than one serotype of T-antigen (Johnson and Kaplan, 1993). Although not as discriminating as M-typing, T-typing and various other typing systems (e.g. opacity factor typing) were useful in cases where M-typing could not be performed. For example, some GAS strains could lose expression of M-antigen when grown *in vitro* and a full set of M-typing sera (which were expensive to produce) were rarely available.

Towards the end of the 1990's M serotyping of GAS began to be superseded by a molecular biology approach based on sequencing of 160 bases at the 5' end (the hypervariable region) of the M-protein (*emm*) genes (Beall *et al.*, 1996). Isolates were considered to have the same *emm* type if they shared >95% identity over this region (Facklam *et al.*, 1999). A recent epidemiological study of *emm* type distribution recorded a total of 205 *emm* types, of which the most common was *emm*1

(corresponding to M1) accounting for 18.3% of all isolates in the report (Steer *et al.*, 2009).

1.2 GAS diseases

The skin and mucous membranes of humans are the only known reservoirs for GAS. There have been isolated reports of GAS cultured from mice (Hook *et al.*, 1960) and primates in captivity (Wilson and Salt, 1978) although the bacteria in these cases were probably passed on from humans. GAS can be carried asymptomatically on the skin and in the nasopharynx of up to 70% of the population, without the production of antibodies to GAS antigen (Pichichero and Casey, 2007; Stevens, 2000) but generally, GAS is associated with a wide variety of diseases (Table 1.1). These range from mild self limiting infections of the skin (impetigo) and pharynx (pharyngitis and tonsillitis) to much more severe infections of these and adjacent sites requiring intervention, and occasionally very severe invasive infections such as necrotizing fasciitis and streptococcal toxic shock syndrome (STSS). One of the most common GAS disease manifestations is pharyngitis. GAS are the most common cause of bacterial pharyngitis and the disease primarily affects children aged 5-15 years old. A conservative estimate of the number pharyngitis cases caused by GAS worldwide each year is 616 million (Carapetis *et al.*, 2005b). Fortunately, GAS continues to remain sensitive to penicillin although studies carried out in the last 15 years show that the failure of penicillin to clear GAS infection in the pharynx is in the range of 20-40% (Pichichero and Casey, 2007) in contrast to 4-8% treatment failure reported in the 1950's (Wannamaker *et al.*, 1951; Denny *et al.*, 1950). Many reasons have been suggested for the failure of penicillin treatment including the intracellular localisation of GAS into epithelial and macrophage-like cells in the pharynx (Molinari *et al.*, 2000; Osterlund and Engstrand, 1997; Osterlund *et al.*, 1997; Osterlund and Engstrand, 1995; LaPenta *et al.*, 1994), although the mechanism of GAS intracellular penetration are still unclear and GAS is believed to remain a predominantly extracellular pathogen.

Pyoderma and non-bullous impetigo represent the second most common GAS presentation and are estimated to be responsible for 111 million cases of skin infection worldwide (Carapetis *et al.*, 2005b). These diseases are considered highly infectious and

Table 1.1: Diseases caused by *Streptococcus pyogenes*. Adapted from Efstratiou, 2000.

	Type of disease	Infection
Suppurative	Superficial	Pharyngitis, skin and soft tissue infections, impetigo, erysipelas, vaginitis, post-partum infections
	Deep	Bacteraemia, necrotizing fasciitis, deep soft tissue infections, cellulitis, myositis, peurperal sepsis, pericarditis, meningitis, pneumonia, septic arthritis.
	Toxin-mediated	Scarlet fever, toxic shock-like syndrome.
	Immunologically mediated	Acute rheumatic fever, post-streptococcal acute glomerulonephritis.

secondary transmission of GAS to the pharynx from impetigo lesions has been reported (Bisno and Stevens, 2005). Deeper penetrations of GAS through the epidermis *via* cuts or abrasions can lead to erysipelas affecting superficial layers of the skin, or cellulitis which affects subcutaneous tissues (Bisno and Stevens, 1996).

With the introduction of *emm* typing to distinguish GAS it was revealed that individual serotypes had a specific tissue tropism. That is, specific M types are associated with pharyngitis and others with pyoderma (Bisno, 1980). Five major *emm* patterns, A-E were suggested by the chromosomal arrangement of *emm* subfamily genes (Bessen *et al.*, 1996). *emm* patterns A-C were most likely to cause throat infections while *emm* pattern D were found to be responsible for skin infections, with pattern E *emm* types able to cause infection at both sites (Bessen, 2009; McGregor *et al.*, 2004).

GAS are responsible for a number of post-streptococcal sequelae including PSAGN and ARF. These diseases result from prior infection with specific GAS serotypes. Clinical evidence suggests that PSAGN is caused by a subset of pyoderma associated, 'nephritogenic' M types whereas ARF is caused by a subset of pharyngitis associated 'rheumatogenic' strains (McShan *et al*., 2008). However, a recent study examining the

high rates of ARF in Aboriginal communities in Australia revealed that cases of pharyngitis were extremely rare while pyoderma was the main manifestation of GAS infection (McDonald *et al.*, 2004). This suggested a prior throat infection may not be the sole cause of ARF, although this has yet to be proven. Both PSAGN and ARF are believed to be immunologically mediated, although the exact mechanisms remain to be determined. The mortality rate from PSAGN is reported to be around 1% (Carapetis *et al.*, 2005b). ARF however, is a much more serious disease reportedly responsible for between 233,000 to 492,000 deaths a year. According to the World Health Organisation there are 15.6 to 19.6 million cases of ARF annually worldwide, 95% of which occur in the developing world (Carapetis *et al.*, 2006; Carapetis *et al.*, 2005a; Carapetis *et al.*, 2005b).

The pattern of GAS infection in the last 150 years appears to be changing. Scarlet fever, erysipelas and puerperal sepsis were the most commonly described diseases caused by GAS during the late 19th century but these diseases are now extremely rare. However, their incidence appeared to decrease well before the introduction of sulfonamide and penicillin antibiotics. The improvement in living conditions throughout the 20th century is believed to have contributed significantly to the decreasing incidence of severe GAS disease. The practice of Semmelweis' techniques also played a role in the falling rates of these diseases after a direct link between a lack of hand washing by physicians and puerperal fever was defined (Denny, 2000; Efstratiou, 2000). During the 1930's, GAS were demonstrated to be associated with rheumatic fever by Coburn in the United States and Collis in England (Denny, 2000). Analogous to above, the frequency and mortality rates of rheumatic fever cases also declined over this period. These observations were all made before the introduction of widely available anti-bacterials. By the 1950's GAS infections and rheumatic fever could be kept under control by primary prophylaxis where adequate treatment of pharyngitis with penicillin was shown to all but eliminate episodes of ARF (Denny, 2000). Secondary prophylaxis, the regular administering of penicillin over many years to people with a history of ARF, also greatly reduced the incidence of disease (Carapetis *et al.*, 2006). This combination led to cases of severe GAS disease becoming rare, prompting a relaxation in the way that GAS pharyngitis was managed. During the late 1980's the number of ARF cases began rising along with other extremely severe GAS infections. STSS (GAS infection associated with sudden onset of shock and multi organ failure) was often accompanied by a deep seated skin

and soft tissue infection, often called necrotizing fasciitis (Stevens *et al.*, 1989), causing rapid necrosis of skin and underlying structures. In 60% of cases, GAS entry to the body prior to these diseases was most commonly through the pharynx, skin or vaginal mucosa (Bisno and Stevens, 1996; Chapnick *et al.*, 1992). In sharp contrast to GAS diseases around 100 years ago, these two conditions affected people of all ages who had no pre-disposing conditions and were not immunocompromised (Bisno and Stevens, 2005; Stevens *et al*., 2005).

Reasons for the re-emergence of severe GAS infections in the late 1980's are unclear. It has been proposed that during a period of mostly benign GAS infections throughout most of the 20th century, new subclones of GAS serotypes have evolved, acquiring new genes which improved their ability to infect humans (Aziz and Kotb, 2008), such as the hypervirulent M1T1 clone (Cleary et al., 1992a). The M1T1 strain differs in its genomic content from M1 GAS strain SF370 (Ferretti *et al.*, 2001) used in this study, by up to 70 kb of phage DNA (Cleary *et al.*, 1998).

1.3 GAS virulence factors

GAS is an extremely versatile pathogen capable of causing a wide range of diseases at many different tissue sites within the human body. This is possible due to the multitude of virulence factors produced by GAS, some of which are listed in Table 1.2 and as indicated, a feature of GAS is that some virulence factors are produced by all strains while others are only produced by some. There is much heterogeneity between strains, even between the same virulence factor e.g. M-protein. As humans are the only known reservoir for GAS infection, the organism has evolved many different mechanisms to facilitate initial binding and persistence. This has led to what appears to be a degree of functional redundancy between some virulence factors, as a number of different proteins may exhibit similar effects. The expression of virulence factors is tightly controlled by a network of stand alone response regulators and two component systems which allow expression of specific virulence factors at specific stages of growth. It is also notable that GAS virulence factors are able to work synergistically to subvert the host immune response, such as release of C5a peptidase from the GAS cell wall by SpeB to tackle complement at a distance rather than at the bacterial surface.

1.4 Secretion of proteins in Gram positive bacteria

Up to 25-30% of all bacterial proteins are functionally active in the cytoplasmic membrane or outside of the cell (Driessen and Nouwen, 2008). These include the many virulence factors of GAS which must traverse the cell membrane and diffuse through the thick but largely porous cell wall before they can exert their effect. Unlike Gram negative bacteria where at least six quite distinct protein secretion systems have been defined (Gerlach and Hensel, 2007), in Gram positive bacteria most secreted proteins are translocated across the single cytoplasmic membrane (in an unfolded conformation) by the general secretion pathway (GSP or Sec pathway), which is found in all species. In some Gram positive species, a small minority of (pre-folded) proteins can be secreted by a distinct 'TAT' pathway, but homologues of key *tat* genes have not been found in GAS, so this is not discussed further here (Dilks et al., 2003).

The main features of the GSP are shown in Figure 1.1. The signal recognition particle (SRP) pathway appears to be the major pathway for the secretion of proteins in bacteria and is found in all species. The bacterial SRP is homologous to the eukaryotic system involved in the translocation of proteins across the endoplasmic reticulum, although it is of a lower complexity (Driessen and Nouwen, 2008). The *Bacillus subtilis* SRP consists of a small cytoplasmic RNA (scRNA) molecule associated to fifty four homolog (Ffh) and is homologous to the eukaryotic SRP54 subunit (van Wely *et al.*, 2001). Ffh has been shown to bind to FtsY (step 1, Figure 1.1), a bacterial homolog of the SRP receptor (Bernstein *et al.*, 1989) which can interact directly with SecYEG (see below) (Angelini *et al.*, 2005). In *B. subtilis*, Ffh appears to be a general targeting factor for secretory proteins. If depleted, the levels of extracellular proteins in the external milieu is reduced by up to 80% (Hirose *et al.*, 2000). An alternative to SRP is SecB, extensively studied in *Escherichia coli*, although homologues are not found in GAS and many other species of bacteria. Moreover, it has recently been discovered that some Gram positive species (including *S. pneumoniae*, *Streptococcus gordonii* and *Staphylococcus aureus*) contain two copies of some components of the GSP including SecA and SecY, which may be required for the efficient secretion of distinct proteins (Bensing and Sullam, 2009; Rigel and Braunstein, 2008).

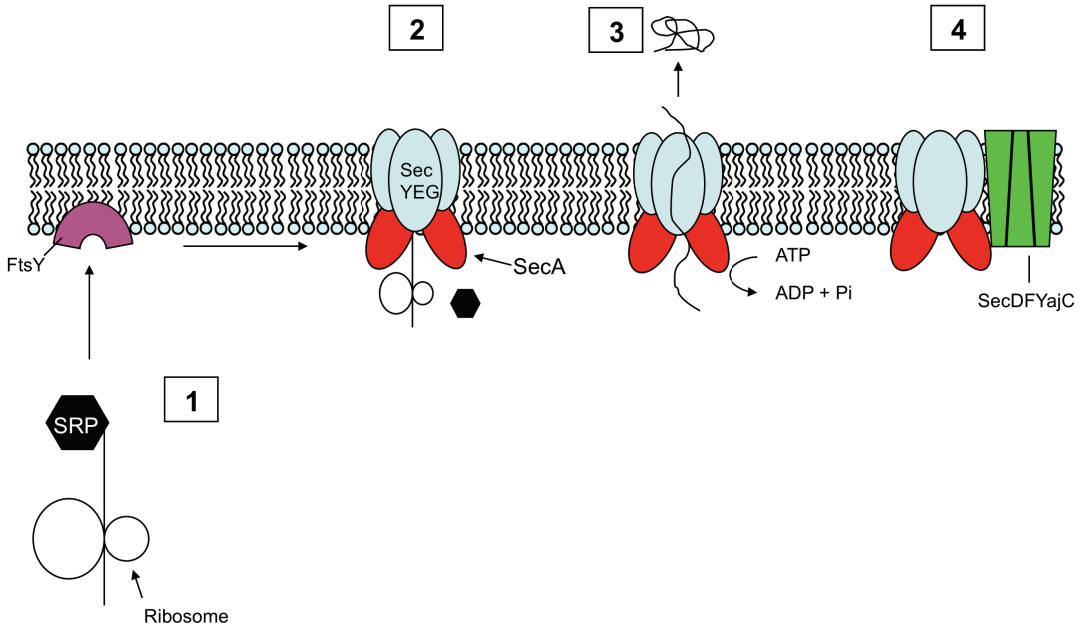

Figure 1.1: The general secretion pathway.

Step 1: Secretory proteins are targeted by the SRP and delivered to FtsY through Ffh. **Step 2**: Three integral membrane proteins, SecY, SecE and SecG come together to form the SecYEG pore which can secrete proteins directly into the external milieu or, laterally through the complex directly into the cytoplasmic membrane (Papanikou *et al.*, 2007). A crystal structure of this complex from *Methanococcus janaschii* has revealed that SecE acts as a 'clamp' to hold both domains of SecY together while SecG localises at the periphery of the complex making only minimal contact with SecY (Van den Berg *et al.*, 2004). **Step 3**: A cytosolic protein, SecA, is believed to be the motor responsible for translocating pre-proteins through SecYEG with continual binding and hydrolysis of ATP. SecA binds pre-proteins through their signal peptide, while the binding of ATP to SecA initiates translocation through the pore. Following hydrolysis, SecA dissociates from the partially inserted pre-protein, de-inserts, and then re-binds triggering an ATP-independent translocation of ~2-2.5 kDa. ATP then rebinds and the process is repeated until the whole protein is translocated through the pore (van der Wolk *et al.*, 1997). The precise mechanism for SecA dependant translocation however remains mostly unresolved (Driessen and Nouwen, 2008). **Step 4**: The SecYEG complex is able to loosely associate with other membrane complexes including the SecDFyajC complex which increases secretion efficiency, possibly by stabilising SecYEG (Duong and Wickner, 1997). Homologues to SecY/E/G, YajC, Ffh and FtsY were all identified in the GAS genome but homologues to SecD/F and SecB were not (Smith, 2004).

Proteins secreted by the GSP are distinguished from cytoplasmic proteins by an amino-terminal secretion signal sequence (from here called signal peptide), which at some stage during secretion is cleaved off by a membrane associated signal peptidase, although it is not clear precisely when this occurs. Signal peptides usually consist of 20-30 residues, although they can be much longer, and share common features even though they may exhibit little sequence homology (von Heijne, 1990). The N-terminal region (N-) contains 1-3 positively charged residues and is followed in sequence by a stretch of 7-15 hydrophobic (H-) residues. The carboxy (C-) terminal region of the signal peptide is usually 3-7 residues in length, is hydrophilic, and contains the cleavage site for signal peptidases (see below) (Tjalsma *et al.*, 2000; von Heijne, 1990). All three of these regions in Gram positive bacteria are longer than their Gram negative counterparts (von Heijne and Abrahmsen, 1989). GSP signal peptides contain a conserved sequence of small residues, commonly Ala-X-Ala (where X is any amino acid) at positions -3 and -1 relating to the cleavage site. Alanine has also been found to be the most common residue at position +1 although all residues (excluding cysteine and proline) have been found at this position (Tjalsma *et al.*, 2000). Once proteins have been translocated through the cytoplasmic membrane, the signal peptide is removed by a specialised serine protease enzyme called a signal peptidase. Type I signal peptidases are membrane associated enzymes located in close proximity to SecYEG translocons and its activity is crucial for cell viability in every species tested to date (Tuteja, 2005).

1.5 Protein secretion in GAS

GAS secrete more then 40 different proteins into the external milieu (Lei *et al.*, 2000) and work on the GSP revealed that secretion of at least some proteins is located at a single microdomain called the ‘ExPortal’ (Rosch and Caparon, 2004). The proposed ‘ExPortal’ is reported to be situated at the site of new wall synthesis at the coccal equator, and is specialised to contain the GSP translocons (Rosch and Caparon, 2004). The group associated with these studies concluded that SecA and HtrA (a surface associated protease which participates in folding reactions) were located at the ‘ExPortal’, suggesting this region functions as an organelle to coordinate protein secretion (Rosch and Caparon, 2005; Rosch and Caparon, 2004). More recently however, Carlsson *et al.* (Carlsson *et al.*, 2006) provided evidence that particular N-

terminal signal peptides direct proteins to different sub-cellular regions of GAS. In contrast to above, their data indicated that SecA was found circumferentially distributed and not localised to a single site. Thus, the exact arrangement of the GSP in GAS is still unclear. It is possible that some proteins can be secreted from any part of the cell but it could be advantageous for wall proteins to be secreted at the site of new cell wall synthesis. The molecular basis for this however remains unknown.

1.6 Gram positive cell walls

Gram positive cell walls consist predominantly of very thick, highly cross linked layers of peptidoglycan with covalently attached anionic polymers (mainly teichoic acid and teichuronic acid), creating a negatively charged wall (Sarvas *et al.*, 2004). Cell wall biosynthesis in Gram positive bacteria has been studied extensively for *S. aureus* although the pathway is almost identical for other Gram positives (Navarre and Schneewind, 1999). This pathway is summarised in Figure 1.2. Of particular interest in the present context is the lipid II precursor (X in Figure 1.2). This molecule includes an amino group that can participate in later transpeptidation reactions. Depending on the species, this can be either the ε-amino group of L-Lys in the penta-peptide moiety, or the equivalent in diaminopimelic acid which is employed in place of L-Lys in some species. Alternatively, when 'cross bridges' are attached at these sites (e.g. penta-glycine in *S. aureus* as shown in Figure 1.2, or di-alanine in GAS), the terminal NH_2 of the 'cross bridge' is available for subsequent transpeptidation reactions. Such transpeptidation reactions play a key role in strengthening the Gram positive cell wall (Nakagawa et al., 1984) and also, more relevant to this project, are the site of attachment for many cell wall proteins.

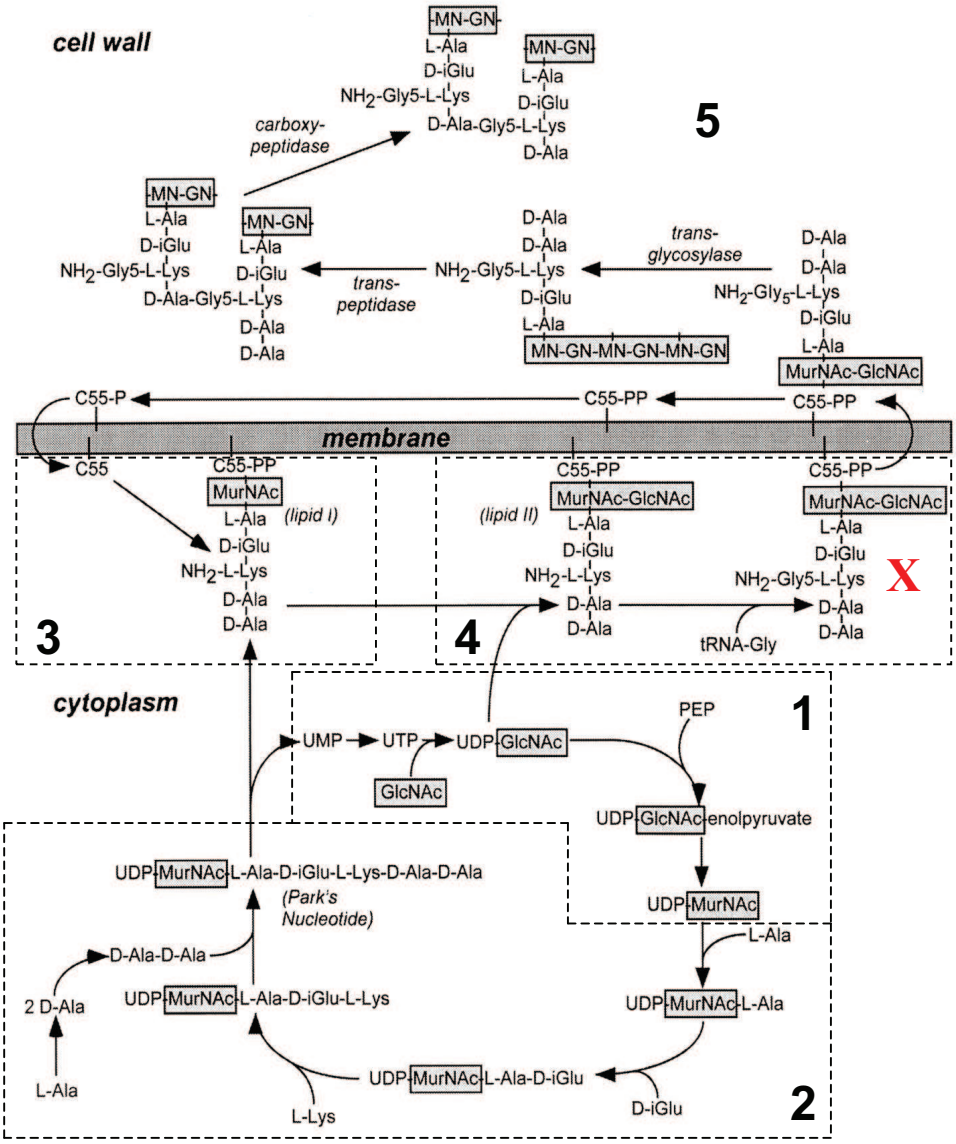

Figure 1.2: Schematic diagram of the cell wall biosynthetic pathway in the Gram positive *Staphylococcus aureus*.

Panel 1: UDP-N-acetyl glucosamine (GlcNAc) together with enolpyruvate forms UDP-N-acetylmuramic acid (MurNAc). **Panel 2**: Five amino acids are then added in sequence, L-Ala, D-isoGlu, L-Lys and D-Ala-D-Ala di-peptide to form Park's nucleotide (Chatterjee and Park, 1964). **Panel 3**: This molecule is then phosphodiester linked to a lipid carrier molecule in the cytoplasmic membrane, displacing UDP to form lipid I (Anderson *et al.*, 1967). **Panel 4**: Linkage to a UDP-GlcNAc molecule creates lipid II and further reactions add the penta-glycine cross bridge to the ε-amino group of L-Lys in the cell wall penta-peptide. **Panel 5**: Lipid II is translocated across the membrane where it becomes the substrate for penicillin binding proteins that catalyse two types of reaction; transpeptidations as mentioned in the text and transglycosylation where glycan strands (GlcNAc-MurNAc) are polymerised, extending between 5-30 subunits in length (Tipper and Strominger, 1965). This pathway is highly similar to cell wall production in GAS, the major difference that GAS has a di-alanine cross bridge in place of penta-glycine attached to lipid II (X). Figure adapted from Navarre and Schneewind, 1999.

1.7 Anchoring of proteins to the Gram positive cell wall

In Gram positive bacteria, once secretory proteins have traversed the cytoplasmic membrane they are free to migrate through the peptidoglycan cell wall and out into the external milieu unless they have a mechanism for attachment. To date, a number of mechanisms for retaining and displaying proteins in the cell wall of Gram positive bacteria have been described.

(i) Less common mechanisms are restricted to individual bacterial species. For example, LytA of *S. pneumoniae* binds to choline residues in teichoic and lipoteichoic acids present in the cell wall, which are anchored in the outer face of the cytoplasmic membrane by their lipid moiety (Holtje and Tomasz, 1975). A second uncommon mechanism for protein retention at the cell surface is exhibited by the InlB protein of *Listeria monocytogenes.* InlB also associates with LTA molecules in the outer face of the cytoplasmic membrane although in a different manner to that described above for LytA. InlB loosely associates with the polyglycerolphosphate backbone of LTA via three 80 amino acid tandem repeats in the C-terminus of the protein, starting with Gly-Trp and called GW modules. Deletion of these residues leads to InlB secretion (Jonquieres *et al.*, 1999; Braun *et al.*, 1997).

(ii) A more common mechanism employed by many bacteria to retain proteins on the outer surface of the cytoplasmic membrane is *via* an N-terminal diacyl-glycerol modification of a cysteine residue in pre-lipoprotein signal peptides. These signal peptides are generally shorter than those found in other proteins and contain a 'lipobox' conserved cleavage sequence of Leu-(Ala/Ser)-(Ala/Gly)-Cys (Tjalsma *et al.*, 2000). During secretion, the cysteine of this motif is modified by a diacylglycerol transferase and following signal peptide cleavage by a distinct 'Type II' signal peptidase, the protein remains anchored in the cell membrane *via* the diacyl-glycerol modification (Tjalsma *et al.*, 2000). Type II signal peptidases show very little sequence homology to other proteases and it has been proposed that they are unusual aspartate proteases with a catalytic triad of Asp-Thr/Ser-Gly (Tjalsma *et al.*, 1999). In contrast to Type I, Type II signal peptidases are not critical for cell viability (Paetzel *et al.*, 2000).

This process was first identified by Hantke and Braun (Hantke and Braun, 1973) in *E. coli* and was later discovered to be applicable to many Gram positive proteins such as BlaZ (Cossart and Jonquieres, 2000; Navarre *et al.*, 1996; Wang and Novick, 1987; Nielsen and Lampen, 1982).

(iii) A third and common mechanism is mediated by a specialised, membrane associated transpeptidase called sortase (as it sorts proteins to the cell wall). Since this is most relevant to this project, sortases and the wall sorting mechanism is described in more detail below.

1.8 Sortases

Many cell wall proteins in GAS and other Gram positive organisms are anchored covalently to the cell wall in a transpeptidation reaction at the outer surface of the membrane, by a membrane associated enzyme called the 'general housekeeping' sortase (usually, but not always, called SrtA) (Marraffini *et al.*, 2006; Navarre and Schneewind, 1999; Kehoe, 1994). All Gram positive bacteria have been found to possess a general sortase but some were found to express additional sortases which tended to be encoded by operons adjacent with the genes whose product the sortase anchors to the cell wall (see later). The SrtA of *S. aureus* is a 206 residue protein with an N-terminal membrane anchor (Mazmanian *et al.*, 1999). The importance of this enzyme was highlighted when *S. aureus* SrtA defective mutants were almost completely cleared from lesions in mice while wild-type bacteria formed kidney abscesses (Mazmanian *et al.*, 2000). This suggested a dramatic decrease in virulence and implicated a role for SrtA in pathogenesis. This enzyme is responsible for anchoring many important surface proteins and virulence factors to the cell wall and has subsequently become an attractive target for anti-microbials (Suree *et al.*, 2009b; Maresso and Schneewind, 2008). Owing to its mechanism of action, SrtA is also being used in a number of biotechnology applications (Proft, 2010; Tsukiji and Nagamune, 2009).

Proteins to be anchored by SrtA are synthesized as precursors with an N-terminal signal peptide and a C-terminal 'stop transfer' signal consisting of a membrane spanning (15-20 residue) hydrophobic region followed by a (2-5 residue) charged tail (Blobel, 1980).

This sequence usually just delays secretion until SrtA can anchor it to the cell wall although this appears to be enough to retain and display the ActA protein of *L. monocytogenes* at the cell surface (Kocks *et al.*, 1992; Blobel, 1980). When examining the C-terminal sequences of surface proteins from Gram positive cocci, a distinct motif located directly upstream of the stop transfer signal was revealed. This sequence consisting of Leucine-Proline-X-Threonine-Glycine (where X is any amino acid) was found to be highly conserved, even at the DNA sequence level (Fischetti *et al.*, 1990). This discovery led to detailed studies of this so-called cell wall sorting signal (CWSS), using protein A from *S. aureus* as a model system. In a series of important experiments it was found that each component of the CWSS; the LPXTG motif, hydrophobic domain and positively charged tail were all required for efficient surface protein attachment to the cell wall (Schneewind *et al.*, 1993; Schneewind *et al.*, 1992). Further experiments showed the LPXTG motif of protein A was cleaved between the threonine and glycine residues by SrtA, and anchored covalently to the penta-glycine cross bridge of the staphylococcal lipid II precursor (Navarre *et al.*, 1998; Ton-That *et al.*, 1997; Schneewind *et al.*, 1995; Navarre and Schneewind, 1994).

1.9 General mechanism of sortase action

Ton-That and colleagues showed that a conserved cysteine (Cys184 in *S. aureus* SrtA) was absolutely required as substitution of this residue to alanine abolished enzyme activity (Ton-That *et al.*, 1999). In addition, the NMR structure of *S. aureus* SrtA revealed a histidine, His120 (Ilangovan *et al.*, 2001) and a later crystal structure revealed an arginine residue, Arg197 (Zong *et al.*, 2004) to be present in the active site of the enzyme close to the conserved cysteine residue. These residues are conserved in all SrtA enzymes and were shown to be required, as alanine substitution severely reduced enzyme activity (Marraffini *et al.*, 2004; Ton-That *et al.*, 2002). Mutation of a non-conserved tryptophan (Trp194 in *S. aureus* SrtA) to alanine did not abrogate attachment but was shown to reduce the kinetics of the reaction (Ton-That *et al.*, 2002). In the two structures of *S. aureus* SrtA discussed above the proposed active site residues (Cys-His-Arg) are not present in conformations consistent with *in vitro* studies. Recent structures of the *S. aureus* SrtA in complex with an LPET peptide, solved by NMR (Suree *et al.*, 2009a), and the GAS SrtA (Race *et al.*, 2009) however, reveal that these

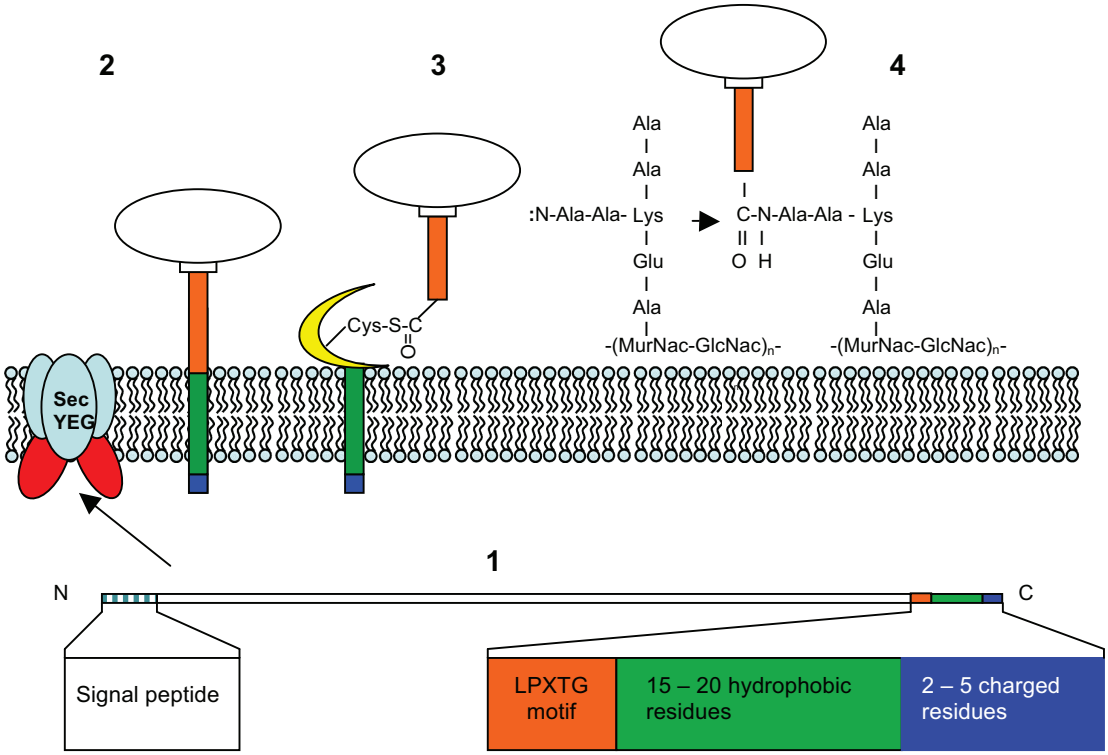

Figure 1.3: Schematic diagram of protein attachment to the cell wall *via* the action of SrtA.

n: represents repeating units of MurNAc-GlcNAc. **Step 1**: Secreted proteins are synthesized as precursors containing all of the features described above. **Step 2**: Following translocation through the GSP the surface protein is held in the cytoplasmic membrane by its stop transfer signal, displaying its LPXTG motif for recognition by SrtA (represented in yellow). **Step 3**: The sulfhydryl group of the conserved cysteine residue in the sortase active site then cleaves the scissile peptide bond between the Thr and Gly of the LPXTG motif creating an acyl-enzyme intermediate with the carboxyl group of the Thr. **Step 4**: This acyl-enzyme intermediate is resolved by a free amino nucleophile provided by the cross bridge of the lipid II peptidoglycan precursor, forming an amide bond with the Thr carboxy terminus of the surface protein (Perry *et al.*, 2002; Ton-That *et al.*, 2000). His120 (*S. aureus* SrtA numbering) mentioned above, has been implicated in activating the conserved Cys (Ton-That *et al.*, 2002), facilitating formation of the acyl-enzyme intermediate while Arg197 could stabilise the reaction (Marraffini *et al.*, 2004).

residues are present in the active site and probably contribute to catalysis of the LPXTG peptide.

The current model describing sortase mediated attachment of surface proteins is outlined in Figure 1.3.

As more types of sortase were discovered, their nomenclature became confusing and this led to attempts to classify these enzymes. Dramsi *et al.* (Dramsi *et al.*, 2005) described four classes of sortase as follows; (i) SrtA, the house-keeping sortase responsible for anchoring LPXTG motif containing surface proteins to the cell wall and has an N-terminal cytoplasmic membrane anchor; (ii) SrtB, contain an N-terminal membrane anchor and recognise motifs specific to the organism and sortase. In the present context this group are responsible for forming long polymeric pili in M1 GAS strain SF370 (section 1.14); (iii) SrtC have been found in corynebacteria, bacilli, clostridial, enterococcal and streptococcal species and have a C-terminal membrane anchor; (iv) SrtD have been found in bacilli, clostridial, actinomycetales and have also been attributed a role in pili formation in *Bacillus cereus* (Budzik *et al.*, 2009a).

1.10 Bacterial pili

For invading pathogens the first step in establishing an infection often involves adhesion to specific tissues within or on the human body. Generally, there are two types of adhesion; those facilitated by non-specific, weak attractive forces, which are sufficient in areas that don't experience strong shear forces; and specific adhesion, which involves binding to specific receptors by non-covalent associations. The net negative charges displayed at the surfaces of both bacterial and mammalian cells tend to repell these surfaces but this repulsive force can be minimised by positioning adhesive protein molecules well away from the bacterial cell surface. Consequently, adhesive molecules are often (though not always) positioned at the tips of long proteinaceous filaments called pili (Latin for 'hair'), or fimbriae (Latin for 'thread' or 'fibre'), which can extend from the bacterial surface for some distance. Pili were first discovered in the Gram negative *E. coli* using electron microscopy in the 1950's (Duguid *et al.*, 1955; Houwink and van Iterson, 1950) while Gram positive pili were first documented in the late 1960's and early 1970's by Yanagawa and colleagues in non-pathogenic strains of corynebacteria (Yanagawa and Honda, 1976; Kumazawa and Yanagawa, 1972; Yanagawa and Otsuki, 1970). Pili are composed of multiple copies of one or more subunits, which are held together by non-covalent interactions in Gram negative bacteria (Proft and Baker, 2009). In direct contrast, the subunits in Gram positive pili are *covalently* attached to each other producing in effect a single large polymer

extending from the bacterial cell surface. Since their discovery, pili in Gram negative bacteria have been studied extensively and a great deal of information is known about their biogenesis and role in pathogenesis, but detailed studies on Gram positive bacterial pili have been comparatively recent.

There are four main groups of Gram negative pili which can be distinguished by their morphology and mechanism of biogenesis; (i) pili such as Type I and P-pili assembled by the chaperone/usher pathway, (ii) Type IV pili, (iii) curli pili assembled by the extracellular nucleation/precipitation pathway, (iv) CS1 pili formed via the 'alternative chaperone-usher pathway'. It is worth noting that these are all quite distinct from pili in Gram positive bacteria and will not be discussed further here. For extensive recent reviews see (Craig, 2009; Proft and Baker, 2009; Pizarro-Cerda and Cossart, 2006; Mattick, 2002; Wu and Fives-Taylor, 2001; Sauer *et al*., 2000).

1.11 Pili in Gram positive bacteria

In the years following their discovery, relatively little progress was made in studying Gram positive pili. Earlier work was limited to oral commensals such as *Actinomyces naeslundii* or other oral streptococci (Wu and Fives-Taylor, 2001). As mentioned above, individual subunits in Gram positive pili are covalently cross linked to each other and the assembled pilus is then anchored to the peptidoglycan cross bridge within the lipid II precursor. These covalent cross links make Gram positive pili extremely robust and the subunits cannot be separated merely by boiling in SDS. The difficulties this presents in trying to purify the individual pilus subunits from Gram positive pili is probably why early work on these structures was so limited after their initial discovery. Significant progress in the field was hampered until genome sequencing data became widely available. This was coupled with the realisation that many important human Gram positive pathogens produced pili, which may be involved in the establishment of an infection and therefore represent novel targets for antimicrobial therapy. Their potential as vaccine candidates has been highlighted in a number of studies which show Gram positive pili proteins elicit a protective immune response in mice (Buccato *et al.*, 2006; Maione *et al.*, 2005; Mora *et al.*, 2005). New information about these intriguing structures was published with increasing frequency throughout the course of this Ph.D project and provided many insights into their biogenesis and function in Gram positive

bacteria. Pili have recently been described in *Enterococcus faecium* (Hendrickx *et al.*, 2008), *Enterococcus faecalis* (Nallapareddy *et al.*, 2006) and *Bacillus cereus* (Budzik *et al.*, 2007) although in this chapter I will only discuss the well characterised pili of *Corynebacterium diphtheriae* and the pathogenic streptococci *Streptococcus agalactiae* (Group B *Streptococcus*, GBS), *S. pneumoniae* and *S. pyogenes*. In each case, these structures consist of a major subunit which forms the shaft of the pilus and either one, but more usually two minor subunits which play a role in host cell adhesion or anchoring the pilus to the cell wall. The genes encoding these subunits are located within the same locus on the chromosome adjacent to genes encoding one or more specific sortases responsible for assembling the pilus (Mandlik *et al.*, 2008b; Oh *et al.*, 2008; Scott and Zahner, 2006). Protein subunits to be incorporated, are synthesized as precursors with a typical N-terminal signal peptide and are secreted *via* the GSP. Detailed studies on these structures have now been performed in important Gram positive pathogens and some of this work is outlined in the following sections.

1.12 Pili in *Corynebacterium diphtheriae*

Pili in Gram positive bacteria were first reported in non-pathogenic corynebacteria although many years later were discovered in *C. diphtheriae*. Studies revealed that *C. diphtheriae* expressed three distinct types of pilus, with the genes involved in the assembly of each located to a particular cluster in the chromosome, called the *spaABC*, *spaDEF* and *spaHIG* (sortase mediated pilus assembly) loci. Each of these clusters encode one or two sortases. Work on the SpaABC pilus revealed that SpaA was located throughout the structure and formed the shaft, with SpaB decorating it and SpaC located at the tip (Ton-That and Schneewind, 2003). The sortase encoded by this gene cluster, confusingly called SrtA but not actually the house keeping sortase (which is termed SrtF in *C. diphtheriae*), was shown to play an important role in pili biogenesis as its deletion abolished SpaA polymerisation into pili (Ton-That and Schneewind, 2003). The discovery of this mechanism meant that a transpeptidation reaction was needed to cross link pili subunits and therefore must involve at least one free amino group. Sequence alignment of the major pilus subunits from a number of Gram positive bacteria identified a so called 'pilin motif' of WxxxVxVYPK, containing a conserved lysine residue at position 190 of SpaA. This lysine was revealed as crucial to pili

biogenesis as substitution for alanine abrogated the polymerisation of SpaA monomers (Ton-That and Schneewind, 2003). The incorporation of SpaB into the pilus was reported to involve the so called 'E-box' motif of SpaA - YxLxETxAPxGY. Mutation of the conserved glutamate within this sequence did not affect polymerisation of SpaA into pili but did seem to affect SpaB incorporation. The precise role that this region plays in pilus assembly was revealed at a later time (see section 1.15). Both the 'pilin motif' and 'E-box' are found in many but not all Gram positive major pilus subunits; for example, neither are found in Spy0128 of M1 GAS strain SF370 (see 1.14.2).

C. diphtheriae ΔspaB mutants released pili from the cell surface after the bacteria were boiled in hot SDS while this was not possible with the wild-type cells (Ton-That *et al.*, 2004). This led to detailed work by Ton-That and colleagues who showed that the housekeeping sortase plays a distinct role in pilus assembly. Deletion of the *srtF* gene from the chromosome led to the secretion of pili into the culture medium. This coupled with the observation that the *ΔspaB* mutant exhibited the same phenotype suggested that SpaB was the final subunit incorporated into the pilus and would therefore be anchored to the cell wall by the house keeping sortase (Mandlik *et al.*, 2008a; Swaminathan *et al.*, 2007). The assembly of SpaABC pili is outlined in Figure 1.4.

The *spaDEF* and *spaHIG* gene clusters each encode a major subunit; SpaD and SpaH respectively, and two minor subunits; SpaF and SpaG likely positioned at the pilus tip and SpaE and SpaI potentially anchoring assembled pili to the cell wall. In contrast to the *spaABC* locus, both harbour genes encoding two sortases. These sortases probably play analogous roles to SrtA and SrtF during SpaA pilus formation described above (Gaspar and Ton-That, 2006; Swierczynski and Ton-That, 2006). Although *C. diphtheriae* produces three distinct pili, only a role for SpaABC pili has been identified so far in pathogenesis. *C. diphtheriae* adhesion to pharyngeal cells (representing the primary site of *C. diphtheriae* infection) was greatly diminished in bacteria which did not express either SpaB or SpaC. Bacterial binding could also be blocked with specific sera to these two proteins while beads coated with SpaB or SpaC both adhered to cultured cells (Mandlik *et al.*, 2007). Deletion of the SpaA protein abolished pilus formation but surprisingly did not reduce bacterial adherence significantly implying that in the absence of SpaA, SpaB and/or SpaC can still mediate adhesion to pharyngeal cells (Mandlik *et al.*, 2007). This suggests that the two remaining pili types are not

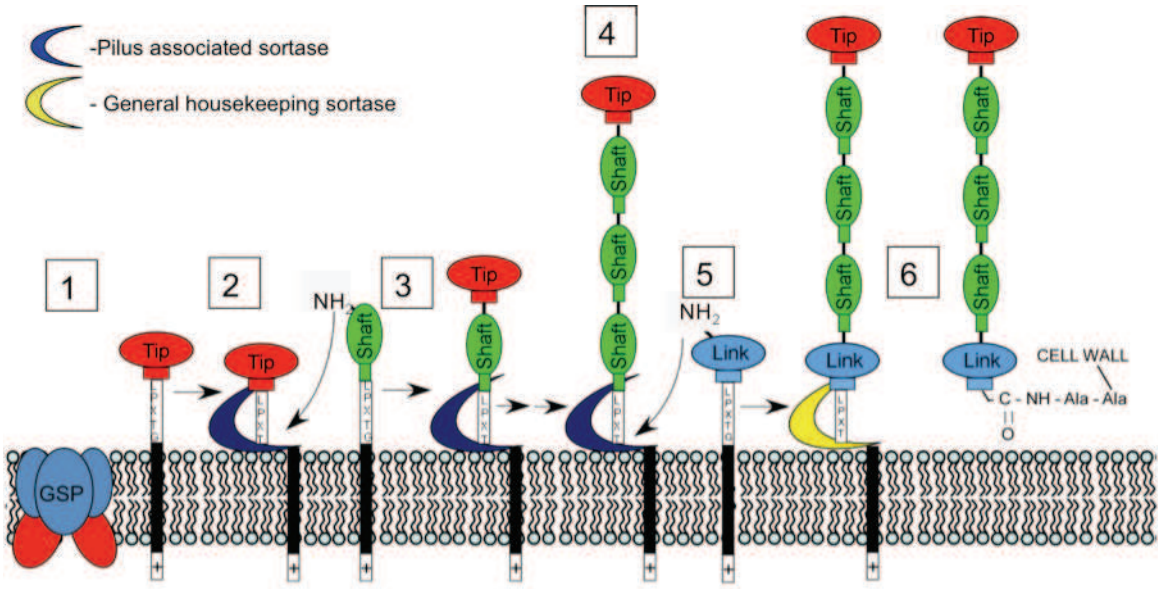

Figure 1.4: General model of heterotrimeric pilus assembly in Gram positive bacteria.

All subunits here are depicted with an 'LPXTG' motif. Depending on the organism in question, subtle variations on this sequence and process are observed. In *C. diphtheriae*, the major shaft subunit SpaA and tip protein SpaC both contain the same C-terminal sorting signal, LPLTG, while SpaB has an LAFTG sorting signal. **Step 1**: Prior to polymerisation into pili, subunits are translocated across the cytoplasmic membrane by the GSP where they are retained in the membrane by their C-terminal stop transfer signal. **Step 2**: Current models of assembly suggest that SpaC is the first subunit to be incorporated and its LPLTG motif is cleaved by the pilus associated sortase between the threonine and glycine creating an acyl-enzyme intermediate as outlined in section 1.9. **Step 3**: This intermediate is presumably resolved by the free ε-amino group from Lys190 of the first SpaA subunit to be incorporated into the pilus. This subunit is then cleaved at its own LPLTG motif by the pilus associated sortase and in turn, this 'new' acyl-enzyme intermediate is resolved by the next SpaA subunit. **Step 4**: This process is repeated many times to produce the polymerised pilus. **Step 5**: The terminal step in pilus assembly most likely involves the incorporation of SpaB ('Link' protein) at the base where a critical lysine at position 139 resolves the final sortase-SpaA acyl-enzyme intermediate (Mandlik *et al.*, 2008a). **Step 6**: The LAFTG motif of SpaB is then cleaved by the housekeeping sortase and anchored to the lipid II precursor in the cell wall (Mandlik *et al.*, 2008a). Recent evidence has shown that all of the components involved in this process localise in the same region of the cytoplasmic membrane leading to a highly specialised centre termed the 'pilusosome' (Guttilla *et al.*, 2009).

involved in adhesion to this cell type and could potentially facilitate adhesion to other, as yet unidentified niches.

1.13 Pili in *Streptococcus agalactiae* and *Streptococcus pneumoniae*

Pili have also been discovered at the surface of important human pathogens from the *Streptococcus* genus. These include *S. agalactiae* (Group B *Streptococcus*, GBS) and *S. pneumoniae*, which show similarities but also distinct differences in pilus assembly to each other and the previously characterised *C. diphtheriae* pili. The presence of pili in GBS was first identified in 2005 (Lauer *et al.*, 2005) and the locus to which they were mapped was termed Pilus Island 1 (PI-1). Two additional pilus loci were subsequently identified in GBS and were considered to be variants of the same island, PI-2a and PI-2b. All three of these loci encode one major and two minor subunits adjacent to two class B sortases (Figure 1.5) (Dramsi *et al.*, 2006; Rosini *et al.*, 2006). Work carried out on pili in GBS revealed that all strains studied to date express at least one type of pilus (Margarit *et al*., 2009). It has been suggested that the two sortases from each locus show a degree of redundancy; while both can polymerise the major subunit into polymeric pili, each shows a distinct specificity for incorporating only one of the minor subunits into the pilus (Dramsi *et al.*, 2006). In an analogous manner to the SpaABC pili of *C. diphtheriae*, GBS *ΔsrtA* mutants released fully polymerised PI-2a pili containing all three subunits into the culture medium (Nobbs *et al.*, 2008). Also, similar to the *C. diphtheriae ΔspaB* mutant, deletion of the *gbs150* minor subunit from PI-2a led to accumulation of pili in the culture medium. This suggests that PI-2a pili are covalently anchored to the cell wall by SrtA *via* GBS150 (Nobbs *et al.*, 2008). GBS is responsible for causing sepsis, pneumonia and meningitis in neonates and attempts to clarify the roles of the minor pilus subunits have been made. Recombinant proteins corresponding to GBS52 and GBS104 (from PI-1, see Figure 1.5) were shown to bind cultured pulmonary epithelial cells although specific antibodies against GBS52 did not block binding (Krishnan *et al.*, 2007). The same study also found that *Δgbs52* mutants displayed dramatically reduced binding and used this as evidence that GBS52 must be the pilus adhesin. However, this reduction in binding was probably due to the release of pili from the cell surface rather than GBS52 acting as the adhesin. Subsequently, GBS104 has been identified as responsible for mediating adhesion to epithelial cells, as

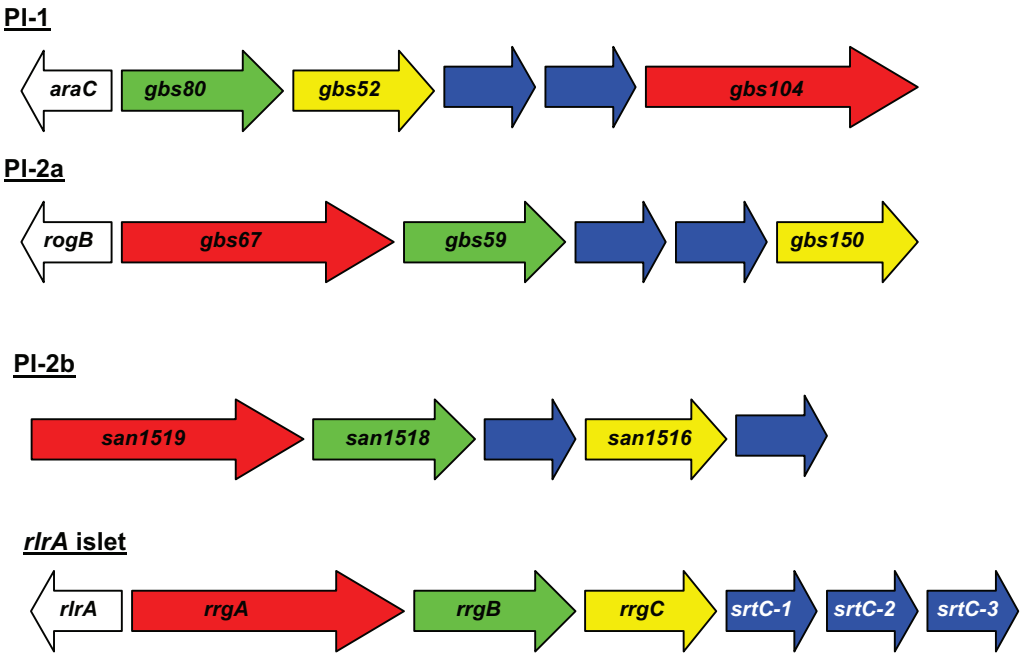

Figure 1.5: Pilus loci of GBS and *S.pneumoniae*.

White arrow: transcriptional regulator. Green arrow: major subunit. Yellow arrow: pilus anchor. Red arrow: adhesin. Blue arrow: sortase. Gene numbers correspond to genome annotations for GBS strain 2603 V/R (PI-1 and PI-2a) and COH1 (PI-2b).

has GBS67 from PI-2a (Pezzicoli *et al.*, 2008). A similar study on a different GBS strain suggested that one subunit, PilA (homologous to GBS67 from PI-2a), was crucial for adhesion while the backbone subunit, PilB, mediated internalisation into brain endothelial cells (Maisey *et al.*, 2007). In contrast, Pezzicoli *et al.* (Pezzicoli *et al.*, 2008) found that deletion of the pilus genes from the chromosome had no effect on the cellular internalisation of GBS, but noted they invaded *via* the paracellular route instead.

Some of the structural pilus proteins encoded by GBS share over 50% sequence homology with proteins encoded by the *rlrA* pilus islet of *S. pneumoniae*. The *rlrA* island is a region of ~11 kb encoding three pilus associated sortase enzymes, *srtC-1*, *srtC-2* and *srtC-3*, an *rlrA* regulator and three wall associated proteins, *rrgA*, *rrgB* and *rrgC*, which have been shown to assemble into polymeric pili (Barocchi *et al.*, 2006). Pili encoded by this islet are composed of repeating RrgB monomers forming the shaft of the pilus (which also contains both the canonical 'E-box' and 'pilin motif') (LeMieux *et al.*, 2006). Early work attempting to elucidate specific roles and positions in the pilus

for RrgA and RrgC suggested that pili formed coiled-coils at the surface of *S. pneumoniae* (Hilleringmann *et al.*, 2008). Detailed EM studies however confirmed that RlrA pili are formed by a single string of RrgB monomers (Hilleringmann *et al.*, 2009) arranged in a head-to-tail conformation with a single monomer of RrgA located at the tip and a single monomer of RrgC at the base (Hilleringmann *et al.*, 2009). This confirmed earlier studies suggesting RrgA was the pilus presented adhesin (Hilleringmann *et al.*, 2008; Nelson *et al.*, 2007; LeMieux *et al.*, 2006). Analysis of the two minor subunits revealed recombinant proteins to both bound to cultured A549 pulmonary epithelial cells although RrgA bound with much greater affinity (Nelson *et al.*, 2007). This arrangement of subunits presents a similar mode of pilus assembly in *S. pneumoniae* to both *C. diphtheriae* and GBS. The recently determined RrgA crystal structure revealed it to be an elongated four domain protein containing a metal ion-dependant adhesion site (MIDAS) motif in its top domain (Izore *et al.*, 2010). These motifs have been shown to mediate protein binding to collagen (Emsley *et al.*, 2000). The RrgA protein has homologues in GBS and M2 GAS pili while the conserved MIDAS motif in these proteins suggests horizontal transfer of the pilus island and a common target for binding. The discovery of three sortase enzymes encoded in the *rlrA* locus led to the hypothesis that there was a one-to-one sortase to surface protein assembly pathway for pili in *S. pneumoniae*. Although this is an attractive model for pilus assembly, recent evidence suggests that this is not strictly the case as the three sortases exhibit (although to a minor degree) redundancy. SrtC-1 was found to be the major sortase involved in polymerisation of RrgB monomers into pili, although in *ΔsrtC-1* strains, 'background' polymerisation of RrgB was observed while SrtC-3 can also form RrgB oligomers *in vitro* (Manzano *et al.*, 2008). Other studies have suggested that while SrtC-1 is responsible for polymerisation of all three subunits into pili, SrtC-3 is required to anchor the pili to a specific location at the cell surface (Falker *et al.*, 2008). Another study by LeMieux *et al.* (LeMieux *et al.*, 2008) reported sortase redundancy in the polymerisation of RrgB but suggested that SrtC-1 incorporated RrgC and SrtC-3 incorporated RrgA into the pilus, while also dismissing the role of the house keeping sortase, as mutants lacking SrtA displayed a wild-type amount of pili at the cell surface. The presence of three sortases in this cluster presents a slightly more complex mechanism of pilus assembly in *S. pneumoniae* than *C. diphtheriae* and GBS although the specific role of each sortase remains unclear. Recent crystal structures of these sortases have provided insights into their mechanism of action, revealing a flexible 'lid'

covering the active site, in contrast to SrtA which instead protects its active site within a hydrophobic pocket (Manzano *et al.*, 2008). The critical Asp and aromatic residues contained within this 'lid' structure, are thought to stabilise the active site residues and upon recognition of a pilus subunit the 'lid' opens and allows entry of the 'LPXTG' type motif (Manzano *et al.*, 2009; Neiers *et al.*, 2009; Manzano *et al.*, 2008). Searches for 'lid' residues in other pilus associated sortases has revealed their presence in a number of Gram positives including GAS. In addition to the RlrA pilus, a second locus has been identified in *S. pneumoniae*, termed pilus island 2 (PI-2) encoding two structural proteins PitB, the major subunit, PitA (whose gene contains an internal stop codon and was therefore not found in mature pili), two sortases (one of which is a pseudogene) and a signal peptidase related product (Bagnoli *et al.*, 2008). Strains containing the polymerised PitB pilus were found to be more adherent to A549 cells than strains where the genes encoding the pilus had been deleted from the chromosome (Bagnoli *et al.*, 2008). Bagnoli *et al*. (Bagnoli *et al.*, 2008) tested their entire strain collection for the presence of the PI-2 locus and found that only 16.4% of their 305 collected isolates harboured the region. This finding was analogous to a similar search for the *rlrA* island in a different strain collection of 465 isolates, which revealed that only 27% contained the region (Aguiar *et al.*, 2008). This study found that not all disease causing strains possessed the *rlrA* pilus island but did show that strains carrying the island were more likely to be resistant to at least one anti-microbial. Interestingly, a number of strains found to be negative for the island did contain remnants of genes contained within it, suggesting horizontal transfer and the presence of this locus in the strain's ancestors. The expression of pili in *S. pneumoniae* did not seem to be crucial for the causation of disease but did contribute to adherence (Aguiar *et al.*, 2008).

1.14 GAS pili

1.14.1 The FCT region of GAS

The genes involved in pili biogenesis in GAS have been mapped to an area of the genome which had previously been designated the FCT region. This was because in some strains this region had been found to contain genes encoding a fibronectin binding protein, a collagen binding protein and a T antigen (Bessen and Kalia, 2002). The FCT region is an 11 to 16 kb locus situated between two highly conserved genes, one encoding a putative heat shock protein and the second of yet undescribed function. This

region has a considerably lower GC% content than the sequences directly up and downstream of it which may be indicative of a pathogenicity island although it lacks other commonly associated features and has been discovered in every GAS genome sequenced to date (Podbielski, 2007). The exact structure of the FCT region and the sequence of genes within it can vary quite extensively between strains and to date nine different FCT structures have been described (see Figure 1.6), although FCT-7 and FCT-9 contain no adhesin (Falugi *et al.*, 2008; Kratovac *et al.*, 2007). Interestingly, the two sortases and minor subunits contained within the FCT-6 region harboured by M2 GAS share 88-93% sequence homology with the same genes from GBS PI-1. This suggests that not only is there recombination of this region between strains of GAS but also between species of streptococci. GBS are able to colonise the vaginal mucosa while M2 GAS are primarily isolated from patients with puerperal sepsis (Falugi *et al.*, 2008). The M-protein has previously been shown to undergo extensive recombination enabling GAS to evade the host immune system (Kehoe *et al.*, 1996) and perhaps this region is subject to similar selective pressures which hastens its evolution.

All FCT regions contain either a *rofA* or *nra* gene and in addition some harbour an *msmR* (multiple sugar metabolism regulator), all of which are regulators of transcription. RofA is important in bacterial-host cell interactions and in response to increased oxygen levels positively regulates itself and the fibronectin binding protein (Kreikemeyer *et al.*, 2002; Fogg and Caparon, 1997; Fogg *et al.*, 1994). The binding of *ΔrofA* mutants to HEp-2 cells was reduced by 40% compared to wild-type bacteria (Beckert *et al.*, 2001). Inactivation of *nra* lead to a decrease in virulence of GAS in a mouse model and it was also noted that Nra appeared to have no effect on *mga* transcription (Luo *et al.*, 2008b). Mga is a stand alone response regulator responsible for the expression of many key virulence factors involved in host cell attachment and immune evasion, including the M-protein and C5a-peptidase (Hondorp and McIver, 2007; Kreikemeyer *et al.*, 2003). Transcription is strongly activated by Mga during the exponential phase of growth (McIver and Scott, 1997) whereas Nra is preferentially expressed during early stationary phase. A recent study showed that transcription of certain virulence factors is regulated by either Mga or Nra but not by both (Luo *et al.*, 2008a). This suggests that there is a highly complex but finely tuned network of genes, cross regulating the expression of virulence factors in GAS as the organism moves from

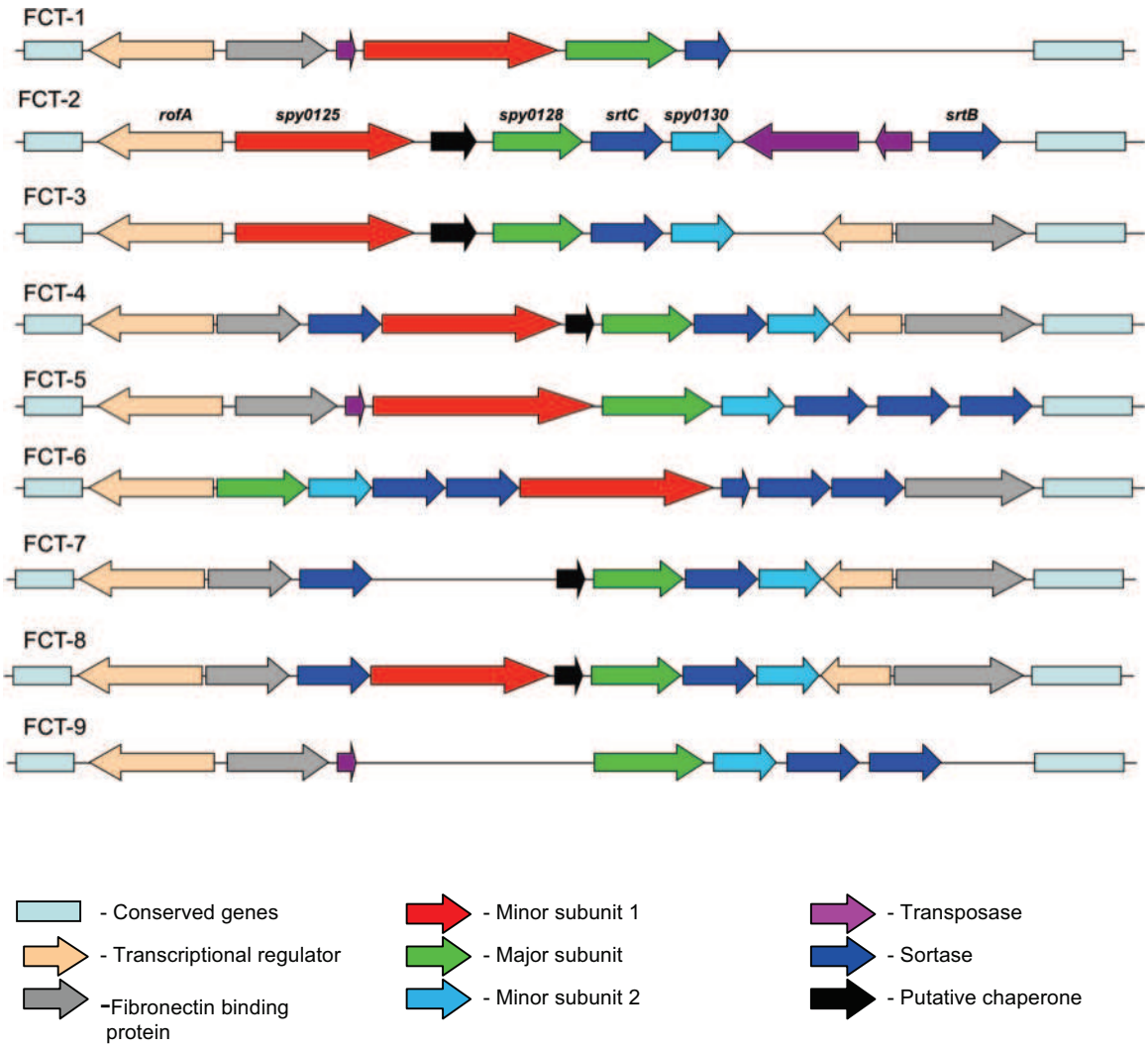

Figure 1.6: The nine FCT regions of GAS.

RofA is the transcriptional regulator encoded in all regions except FCT-3 and FCT-8, which contain Nra. The second regulator of transcription in FCT-3, FCT-4, FCT-7 and FCT-8 is of the AraC/XylS type. FCT genes encoded by FCT-3 and FCT-4 share the greatest degree of similarity and were found to be the most prevalent type in studies by Kratovac *et al*. (Kratovac *et al*., 2007) and Falugi *et al*. (Falugi *et al*., 2008). In contrast, FCT-2 and FCT-6 were found to be rare in the same sample set (Kratovac *et al*., 2007). FCT-7 and FCT-8 are derivatives of FCT-4 (FCT-7 lacks a detectable minor subunit 1 while FCT-8 is identical to FCT-4 apart from *nra* replacing *rofA*). The genes of the M1 GAS strain SF370 FCT-2 region responsible for pilus assembly are highlighted. This figure was compiled from available genome sequence data for FCT-1 to FCT-6 inclusive and from data presented by Kratovac *et al*. (Kratovac *et al*., 2007) and Falugi *et al*. (Falugi *et al*., 2008) for FCT-7 to FCT-9 inclusive.

one stage of growth and infection to the next, potentially conferring on a strain a particular tissue tropism. These regulatory events are likely to be strain specific as Nra from two strains of GAS, both harbouring an FCT-3 region have differing effects; in an M49 GAS strain, Nra negatively regulates pilus expression (Podbielski *et al.*, 1999) while in an M53 GAS strain Nra positively regulates pilus gene expression (Luo *et al.*, 2008b). The gene encoding the MsmR regulator is present in a number of FCT regions and belongs to the AraC/XylS type transcriptional regulators. GAS *ΔmsmR* strains have been shown to exhibit reduced binding to immobilised fibronectin and human pharyngeal cells which was suggested to be as a result of downregulated expression of the major and a single minor pilus subunit in these strains (Nakata *et al.*, 2005). This information taken together suggests that MsmR positively regulates the FCT region genes and consequently pilus expression.

A number of FCT regions, including types 2 and 3, harbour a *sipA* or *lepA* gene (depending on the strain) whose product exhibits homology to previously described signal peptidases. Upon closer examination however, common features such as the catalytically important lysine and serine residues were not present. A role for these proteins in pilus assembly was highlighted when its deletion from the chromosome resulted in unpolymerised major pilus subunit monomers at the cell surface (Nakata *et al.*, 2009). It has been suggested it functions as a chaperone (Zahner and Scott, 2008), although its precise function in pilus biogenesis has yet to be revealed.

1.14.2 Pili in *Streptococcus pyogenes*

Pili in GAS were first visualised extending from the cell surface of serotype M1 GAS strain SF370 (the strain used in this study) by immuno-gold labelling (Mora *et al.*, 2005). These pili are encoded by the FCT-2 locus shown in Figure 1.6. The product of the *spy0128* gene was shown to be the major pilus subunit which forms the shaft and unlike many other major pilus subunits, Spy0128 does not contain an ‘E-box’ or ‘pilin motif’. The products of minor subunits *spy0125* and *spy0130* were also associated with the pilus but were present in far fewer copies (Mora *et al.*, 2005). The minor subunits had been shown to display some homology to previously described fibronectin and collagen binding proteins (Kreikemeyer *et al.*, 2005; Rocha and Fischetti, 1999)

although more recent evidence suggests that neither fibronectin or collagen act as the receptor (Abbot *et al.*, 2007). The FCT region of M1 GAS strain SF370 also encodes a sortase that was designated SrtC, (sometimes referred to as a 'pilin polymerase') but this is actually a type B sortase in the classification system devised by Dramsi (Dramsi *et al*., 2005). A second encoded sortase (designated SrtB), does not contain a signal peptide and is therefore not thought to be functional at the cell surface (Barnett *et al.*, 2004; Barnett and Scott, 2002). Abbot *et al*. (Abbot *et al*., 2007) showed that deletion of this gene had no detectable effects on the expression of pili or other surface proteins examined. In addition, there is a *rofA* transcriptional regulator, *sipA* and two transposase genes highlighting that this region was probably acquired horizontally (Figure 1.6). Mora *et al*. (Mora *et al*., 2005) observed that polymerisation of all three structural subunits into high molecular weight 'ladders' on immunoblots, characteristic of pili, was abolished in a Spy0128 deletion mutant. This phenotype was repeated in a SrtC deletion mutant while they reported a Spy0130 deletion mutant as displaying polymerised pili that appeared to be less frequent at the cell surface. Subsequent work performed by Abbot *et al.* (Abbot *et al.*, 2007) showed that essentially four genes were needed from the FCT region for the formation of functional pili at the cell surface. Deletion of *spy0128* or *srtC* from the chromosome abrogated pilus assembly (confirming the earlier observations of Mora *et al*.) while deletion of *spy0125* or *spy0130* produced polymerised pili that were not functional. This suggested that Spy0125 and/or Spy0130 might act as adhesins but their roles had not been clearly defined at the beginning of this project. This work also showed that GAS pili adhered to specific tissue types, representative of the common sites of GAS infection, including human tonsil, primary keratinocytes and HaCaT's but not others such as A549 and Hep-2 cells that are commonly used in GAS adhesion experiments. This was supported by Manetti *et al*. (Manetti *et al.*, 2007) who showed that recombinant proteins of Spy0125 and Spy0130, both adhered to Detroit human pharyngeal cells. Similar to the Abbot *et al.* study, strains deficient in the major subunit and SrtC were unable to bind to the cultured cell line (Manetti *et al.*, 2007). These data taken together suggest a direct role for pili in the initial adherence of GAS to human cells and therefore a role in the pathogenicity of the organism. Pili in GAS have also been attributed a role in forming biofilms in an *in vitro* model, but the importance and role of biofilms in GAS infection is unclear (Manetti *et al.*, 2007). In contrast to the beneficial adhesive properties of pili, a salivary agglutinin gp340 was found to bind pili from the M1 GAS strain SF370

causing bacterial aggregation, potentially helping with their clearance from the oropharynx (Edwards *et al.*, 2008).

1.15 Isopeptide bonds of Spy0128

During the early stages of this project Kang *et al*. (Kang *et al*., 2007) reported the crystal structure of the major pilus subunit, Spy0128. This was combined with mass spectrometry analysis of intact pili and identified Spy0128 Lys161 as the nucleophile responsible for resolving the SrtC-Spy0128 acyl enzyme intermediate and forming a covalent link to the Thr residue within the 'EVPTG' motif at the C-terminus of the preceeding Spy0128 monomer to be incorporated into the growing pilus. In addition, a particularly novel observation were two *intra*-molecular isopeptide bonds, one in each domain of the protein, linking the first and last β-strands of the domains (Kang *et al.*, 2007). These covalent bonds are self generated between the side chains of lysine and asparagine residues with the help of a crucial glutamate residue. These types of bonds had only been previously reported between polypeptides, rather than within the same molecule and being formed by the action of enzymes such as transglutaminases (Lorand and Graham, 2003), ubiquitin ligases (Pickart, 2001) and sortases (see section 1.9). Viral proteins forming the capsid of the HK97 phage were found to be cross-linked by *inter*-molecular isopeptide bonds that appeared to be self generated by the action of a specific glutamate residue, which when mutated to alanine abolished the cross linking of capsid monomers (Wikoff *et al.*, 2000; Duda, 1998). Mutation of the critical glutamate in Spy0128 also abrogated bond formation (Kang *et al.*, 2007). A plausible mechanism for the formation of *intra*-molecular isopeptide bonds has been proposed (Kang *et al*., 2007). Within the hydrophobic environment of the protein core, the glutamate side chain becomes protonated and is able to polarise the C=O of the asparagine side chain. This event induces a partial positive charge on the asparagine Cγ, which is then subject to nucleophilic attack by the unprotonated lysine ε-amino group, generating the bond (Kang *et al.*, 2009a; Kang *et al.*, 2007). Mutagenic studies carried out to characterise *intra*-molecular isopeptide bonds revealed that their most likely role was to provide extra stability to the protein in a manner analogous to disulphide bonds in eukaryotic proteins. Recombinant Spy0128 harbouring point mutations which abrogated *intra*-molecular bond formation were shown to be much more susceptible to

protease degradation and heating when compared to the native protein (Kang and Baker, 2009).

Following this fascinating discovery crystal structures of pilus subunits from a range of Gram positive pathogens were reported to contain *intra*-molecular isopeptide bonds. The major subunit BcpA responsible for forming pili in *B. cereus* was shown to be a 4 domain protein with each domain harbouring an *intra*-molecular isopeptide bond (Budzik et al., 2009b; Budzik et al., 2008a). The bonds present in the CNA_2 and CNA_3 domains were shown to link the first and last β strands of the domain in a similar way to the bonds present in each domain of Spy0128 (Budzik *et al.*, 2009b). The crystal structure of the SpaA subunit from *C. diphtheriae* revealed the protein to contain an *intra*-molecular isopeptide bond in two of its three domains, in addition to a disulphide bond within its C-terminal domain (Kang *et al.*, 2009b). While the *intra*-molecular isopeptide bond present within the C-terminal domain of SpaA links the first and last β strands of the domain, the bond situated in the middle domain links the first and penultimate β strands (Kang *et al.*, 2009b). The SpaA structure also showed for the first time that the glutamate residue of the so called 'E-box' was in fact the crucial glutamate residue needed for *intra*-molecular isopeptide bond formation (Kang *et al.*, 2009b). As this Ph.D project progressed more information about structural pilus proteins became available and increasingly showed that although these proteins came from a wide variety of organisms they contained similar features such as *intra*-molecular isopeptide bonds. This was further highlighted by recent work done on *S. pneumoniae* pili, showing that all three pilus subunits contained isopeptide bonds. Biochemical characterisation of the RrgB and RrgC proteins revealed that they contain three and two *intra*-molecular bonds respectively (El Mortaji *et al.*, 2010) while the X-ray crystal structure of RrgA highlighted the presence of two isopeptide bonds (Izore *et al.*, 2010). The role of these bonds across the species of bacteria discussed here is probably the same, in that they are required for providing extra stability to these surface exposed structures that are likely to experience significant mechanical forces upon host cell binding.

1.16 Project objectives

The original objectives of this project were to study all three pilus subunits, combining structural and molecular microbiology, with the aim of providing insights into their structure, role in pilus biogenesis and function. At the beginning of this project, Spy0128 had been described as the major pilus subunit that is assembled into the pilus shaft while Spy0125 and Spy0130, had been described as minor subunits but their specific functions were not known (Mora *et al*., 2005). Studies on Spy0128 suggested that the formation of *intra*-molecular isopeptide bonds was not critical to pilus function. From the outset it was anticipated that work on the individual subunits would progress at different rates, depending on the progress made and parallel studies of other laboratories. This proved to be the case, with a large proportion of this project focussing on Spy0125. Experiments by Manetti *et al*. (Manetti *et al*., 2007) suggested that both minor subunits might be pilus adhesins. However, work done in this thesis shows that only one, Spy0125, is responsible for the attachment of M1 GAS strain SF370 pili to host surfaces and is positioned at the pilus tip. Intriguingly, the crystal structure of the Spy0125 binding region presented here reveals a potentially novel mechanism of attachment for bacteria to host cells.

Chapter 2. Materials and methods

2.1 Bacterial strains and plasmids

The *Streptococcus pyogenes* and *Escherichia coli* strains used in this work are described in Tables 2.1 and 2.2. The plasmid vectors used are described in Figures 2.1 and 2.2.

2.2 Chemicals, enzymes and reagents

All chemicals used in this work were purchased from Sigma-Aldrich Company Ltd. (Poole, Dorset, UK) or VWR International (West Chester, Pennsylvania, USA) unless specified otherwise.

Finnzymes Phusion High Fidelity polymerase was supplied by New England Biolabs Ltd (Hertfordshire, UK) and KOD Hot Start DNA Polymerase by Novagen (Merck Chemicals Ltd, Darmstadt, Germany). Restriction endonucleases were purchased from Roche Diagnostics Ltd (Lewes, East Sussex, UK) or Fermentas (York, UK). T4 DNA ligase was purchased from Roche Diagnostics Ltd. All enzymes were supplied with their appropriate reaction buffers and were used as described in the manufacturer's instructions.

Table 2.1. *Streptococcus pyogenes* strains

M Type	Strain	Source / Reference	Disease association
S. pyogenes M1	SF370	ATCC 700294 / (Suvorov and Ferretti, 1996)	Wound infection

Table 2.2. *Escherichia coli* strains

Strain	Genotype and other descriptions	Source / Reference
DH5α	F-, *ø80dlacZΔM15 recA1 endA1 gyrA96 λ- thi-1 hsdR17 (r_K-,m_{K+}) phoA supE44 relA1 deoR Δ(lacZYA-argF)U169*	(Hanahan, 1983)
BL21 (DE3)	F- *ompT $hsdS_B$(r_B-, m_B-) gal dcm* λ(DE3)	Stratagene, Amsterdam, Europe
BL21 (DE3) CodonPlus Type RIL	F-, *ompT hsdS(r_B-,m_B-) dcm* Tet^r *gal* λ(DE3) *endA* Hte [argU ileY leuW Cam^r]	Stratagene, Amsterdam, Europe
B834 (DE3)	F- *ompT gal $hsdS_B$ (r_B-m_B-) met dcm lon* λ(DE3)	Gift from Structural Biology Lab, Newcastle University
TG1-dev	*recA sup hsbΔ5 thi Δ(lac-proAB),* F'[*traD36 proAB+ lacqZΔM15*] c/s:: *repA.* A derivative strain TG1, where the wild-type *repA* gene from the cryptic *Lactococcus lactis* plasmid pWV01 (precursor of pG^+host9) has been inserted into the chromosome. This enables pG+host9 to replicate in *E. coli* at 37°C.	(Law *et al*., 1995) E. Maguin, personal communication

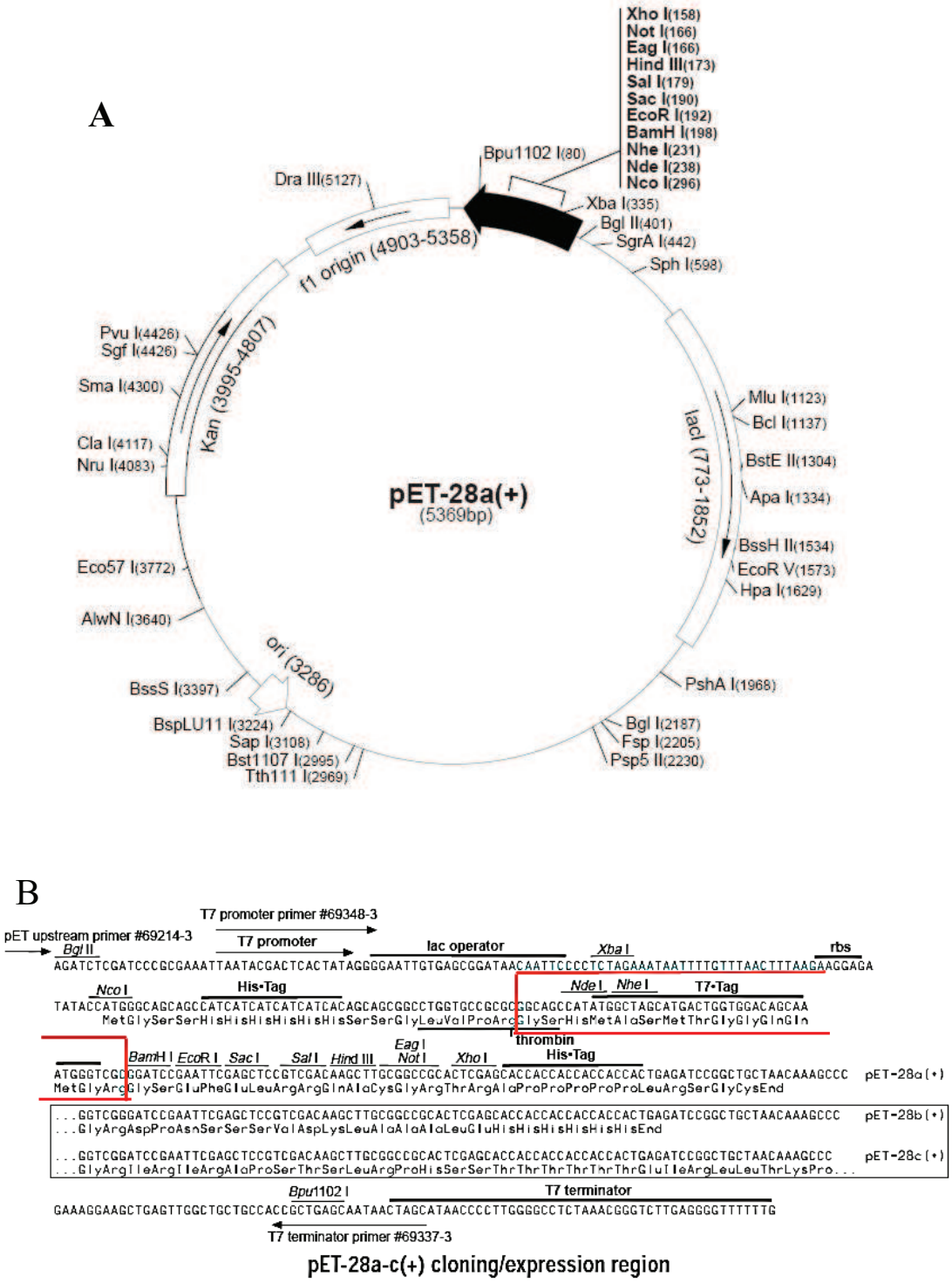

Figure 2.1: pET-28a(+) expression vector and modifications.

Panel A: The pET-28a(+) plasmid is designed to express cloned gene products as translational fusions with an N terminal, or optional C terminal His-tag. The plasmid encodes a thrombin cleavage site and T7 configuration. **Panel B**: The pET-28a(+) cloning region is shown and was modified by the removal of bases 203-248 as indicated by the red box to create pET-28a(+)-D. Source Novagen, Nottingham, England. Figure adapted from www.emdchemicals.com/showBrochure?id=200905.345.

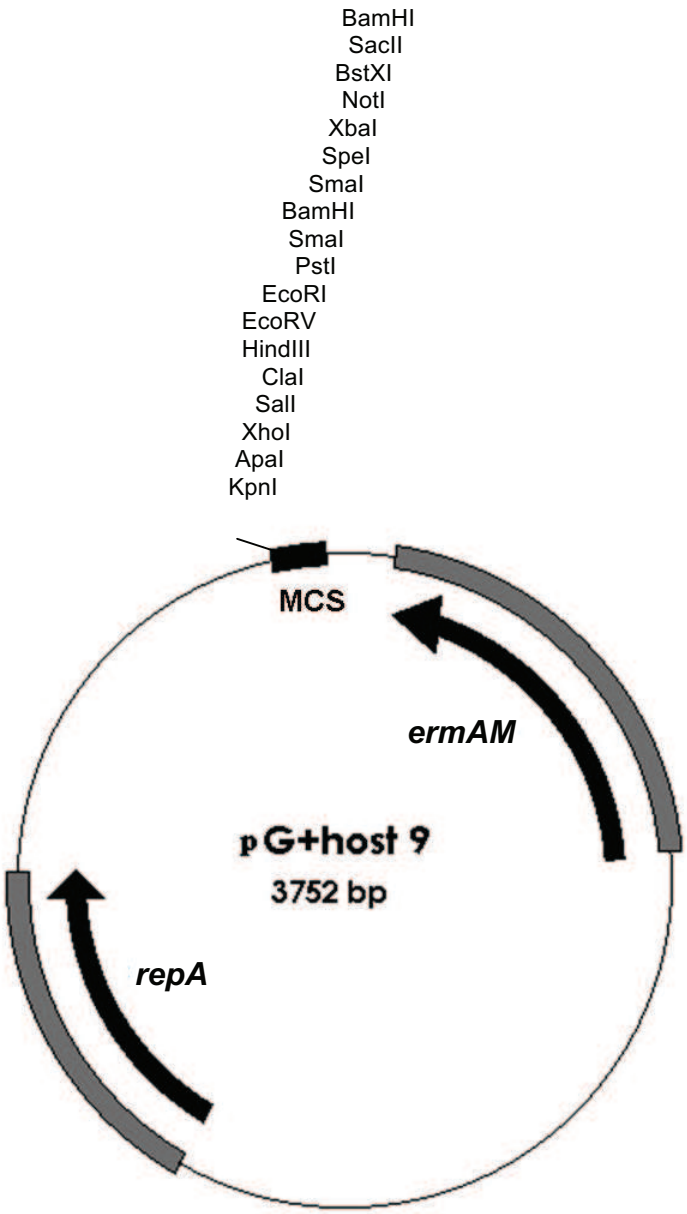

Figure 2.2. pG^{+}host9.

The allele replacement vector pG^{+}host9 encodes *repA*, the temperature sensitive replication gene, and *ermAM* an adenine methylase, which confers resistance to macrolide, lincosamine and streptogrammin (MLS) type antibiotics. The multiple cloning site is derived from pBluescript SK (Stratagene) and the recognition sequences of *NotI*, *BstXI* and *SacI* are overlapping. Figure adapted from Smith, 2004.

2.3 Media, culture conditions and strain storage

All media were made using de-ionised water and sterilised by autoclaving at 121°C, 15 psi for 30 min. Sterilised molten agar was allowed to cool to <45°C before adding heat sensitive supplements after which plates were poured immediately.

S. pyogenes strains were grown in either THY broth (THYB) or on THY blood agar plates in a static incubator at 37°C, 5% CO_2. THYB consisted of Todd Hewitt broth (Difco, Oxford, UK) with 0.5% (w/v) Bacto-yeast extract (Difco). For THY plates, 1.5% (w/v) Bacto-agar (Difco) was added prior to sterilisation. After cooling THY agar to 45°C, 2.5% (v/v) defibrinated horse blood (TCS Biosciences Ltd, Botolph Claydon, Bucks, UK) was added immediately prior to pouring the plates. When selecting for the Erm^r marker encoded by pG^+host9, erythromycin was added to a final concentration of 1 µg/ml. *S. pyogenes* strains were stored at 4°C for up to 4 weeks on THY blood agar plates sealed with Parafilm. For long-term storage glycerol stocks were prepared by inoculating 5 ml THYB with a single colony and cultures were grown for a minimum of 8 h, at which point filter sterilised glycerol was added to a final concentration of 25% (v/v). All stocks were stored at -80°C in sterile cryogenic vials (Nalgene, New York, USA).

E. coli strains were cultured in Luria-Bertani (LB) broth at 37°C (with shaking at 200 rpm) in an orbital incubator (Sanyo, Gallenkamp PLC, Leicester, UK). Individual colonies were grown on LB agar plates in a static incubator at 37°C for 16-24 h. LB consisted of 1% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) sodium chloride, 0.5% (w/v) Bacto-yeast extract. LB agar was prepared by the addition of 1.5% (w/v) Bacto-agar (Difco) prior to sterilisation. When selecting for pET-28a(+), kanamycin was added to media at a final concentration of 50 µg/ml. Selection of the pET30 markers carried by the BL21 (DE3) CodonPlus Type RIL host strain was acheived with the addition of chloramphenicol to the media at a final concentration of 34 µg/ml. When selecting for Erm^r markers encoded by pG^+host9, LB media was supplemented with 300 µg/ml erythromycin. *E. coli* strains were stored long term as glycerol stocks, prepared essentially as described above after being grown for a minimum of 8 h in LB broth at 37°C with shaking at 200 rpm.

Stock solutions of antibiotics were prepared, filter sterilised using 0.2 μm acrodiscs (Pall Life Sciences, Newquay, Cornwall, UK) and stored in sterile microfuge tubes at -20^{o}C prior to use. Stocks of kanamycin (10 mg/ml) were prepared in water. Stocks of chloramphenicol (34 mg/ml) and erythromycin (1 mg/ml and 100 mg/ml) were prepared in ethanol to aid dissolving.

2.4 Bacterial growth rates

For each strain a single well separated colony was inoculated into 10 ml THYB and incubated for 16 h at 37^{o}C with 5% CO_2. Cultures were diluted 1/25 into fresh THYB pre-warmed to 37^{o}C, and then incubated at 37^{o}C with 5% CO_2. Over the next 9-12 h, 1 ml samples were removed from each culture at hourly intervals for OD_{600} measurements.

2.5 General DNA procedures

2.5.1 Isolation of plasmid DNA from *E. coli*

A single well-separated *E. coli* colony transformed with the desired plasmid was used to inoculate 5 ml of LB broth supplemented with appropriate antibiotics, and the cultures were grown for 16 h at 37^{o}C with shaking at 200 rpm. Plasmid DNA was recovered from these cultures using the QIAprep Spin Miniprep Kit (Qiagen Ltd, Crawley, West Sussex, UK). These kits follow the modified alkaline lysis method developed by Birnboim and Doly (Birnboim and Doly, 1979). Cell lysate was neutralised and adjusted to high salt binding conditions before centrifugation at 16,000 x g for 10 min to pellet insoluble cell debris. The supernatant, containing plasmid DNA was then applied to a QIAprep Spin Column to allow binding of plasmid DNA to the column silica gel membrane in high concentrations of chaotropic salts. At this stage contaminating material was removed with several wash steps using the buffers provided in the kit and up to 20 μg of high quality plasmid DNA was eluted from the membrane with 50 μl 10 mM Tris-Cl, pH 8.5, and the DNA was stored at -20^{o}C.

2.5.2 Preparation of GAS chromosomal DNA

Isolation of GAS chromosomal DNA employed cell suspension solution, lytic enzyme, cell lysis solution, protein precipitation solution and DNA hydration solution provided by the Archive Pure DNA Yeast and Gram positive kit (5 PRIME, Flowgen Bioscience, Nottingham, UK).

For preparation of total chromosomal DNA, a single well isolated GAS colony from a fresh THY blood agar plate was used to inoculate 7.5 ml THYB and was grown for 16 h at 37^{o}C with 5% CO_2 in a static incubator. Cells were harvested by centrifugation at 3,600 x g for 15 min and resuspended in 500 µl of cell suspension solution with 25 µl mutanolysin (10 U/µl), 50 µl lysozyme (100 mg/ml) and 60 U lytic enzyme added. The mixture was incubated at 37^{o}C for 3 h to digest cell walls. After incubation, resulting protoplasts were pelleted at 16,000 x g and the supernatant was removed. The cell pellet was resuspended in 500 µl cell lysis solution with gentle pipetting up and down employed to aid cell lysis. Samples were heated to 80^{o}C for 30 min and then allowed to cool to room temperature before 7.5 µl RNase A (4 mg/ml) was added. Samples were then incubated for a further 60 min at 37^{o}C before being cooled on ice and 500 µl of protein precipitation solution was added, followed by vortexing at high speed to aid mixing. Samples were chilled on ice for 5 min before precipitated proteins were pelleted by centrifugation at 16,000 x g. The supernatant containing the chromosomal DNA was collected into a clean microfuge tube with 500 µl of isopropanol added, mixed and centrifuged at 16,000 x g to pellet the DNA. The isopropanol was removed and the DNA was washed with 500 µl of 70% (v/v) ethanol followed by centrifugation at 16,000 x g. The ethanol was poured off and samples were allowed to air dry for 15 min before 100 µl of DNA hydration solution was added. DNA was rehydrated by incubation for 1 hour at 65^{o}C followed by 6 h at room temperature and then stored at 4^{o}C.

2.5.3 Transformation of *E. coli* by modified Hanahan method (Heat Shock)

A slightly modified version of the procedure devised originally by Hanahan (Hanahan, 1983) was used. Competent cells were prepared by inoculating a single well isolated colony from a freshly streaked LB agar plate into 7.5 ml of LB broth, and incubating

this culture for 16 h at 37°C and 200 rpm. Next day, this was used to inoculate 300 ml of LB (1/40) which was incubated at 37°C and 200 rpm until the culture density at OD_{600} reached 0.6. Once the desired density had been achieved, cells were chilled on ice for 20 min and then harvested by centrifugation at 2,355 x g for 20 min at 4°C using a 3-16 K centrifuge (Sigma). Pelleted cells were washed by resuspending in 100 ml of ice cold TFB1 [30 mM potassium acetate, 10 mM $CaCl_2.2H_2O$, 100 mM KCl, 15% (v/v) glycerol, initially dissolved and autoclaved in 900 ml H_2O and then made up to 1 l by the addition of 100 ml autoclaved 500 mM $MnCl_2$]. Cells were harvested as above and resuspended in 7 ml ice cold TFB2 [75 mM $CaCl_2$, 10 mM KCl, 15% (v/v) glycerol initially dissolved in 900 ml H_2O and then made up to 1 l by the addition of 100 ml of autoclaved 100 mM Na-MOPS pH 7.0]. Aliquots of 200 µl were made in pre-cooled sterile microfuge tubes and frozen rapidly in an ethanol-dry ice bath. The frozen competent cells were stored at -80°C.

For transformation, competent cells were thawed on ice and incubated with plasmids or ligation mixtures, immediately upon thawing, for 30 min. Cells were then heat shocked at 42°C for 2 min and returned to ice for 3 min to cool before 1 ml of LB broth was added to each transformation mixture. These were incubated at 37°C and 200 rpm for 120 min to allow expression of antibiotic resistance markers. Samples (e.g. 100 µl) of culture were spread onto LB agar plates containing appropriate antibiotics and incubated for 16 h at 37°C in a static incubator to allow growth of single colonies.

2.5.4 Transformation of *E. coli* by electroporation

Electroporation was used to transform *E.coli* TG1-dev. Freshly competent cells were prepared by inoculating 10 ml of LB broth with a single well-isolated colony from a freshly streaked LB agar plate and the culture was incubated for 16 h at 37°C and 200 rpm. The culture was used to inoculate 500 ml of LB broth (1/100) and was incubated at 37°C and 200 rpm until the OD_{600} reached 0.5-0.6. Once the desired OD had been achieved cells were chilled on ice for 20 min and then harvested by centrifugation at 2,355 x g for 20 min at 4°C. Cells were resuspended in 500 ml of ice cold, sterile H_2O and harvested as above. Cells were washed for a second time in 250 ml of ice cold, sterile H_2O and for a final time in 50 ml of ice cold, 10% (v/v) glycerol (which had been

filter sterilised using a 0.22 μM acrodisc, Millipore) and centrifuged as above. Cells were resuspended in 1 ml of ice-cold 10% (v/v) glycerol (filter sterilised as above) and 60 μl aliquots were dispersed in pre-cooled microfuge tubes, which were frozen rapidly in an ethanol-dry ice bath and stored at -80^{o}C.

For electrotransformation, competent cells were thawed on ice and incubated with 1-10 ng of plasmid DNA for 5 min. The mixture was then transferred to a pre-chilled Gene Pulser cuvette, 0.2 cm electrode gap (Bio-Rad), which was dried and placed in the electroporation chamber. The Gene Pulser II (Bio-Rad) was used to pulse the sample under the following electroporator settings: 2.5 kV voltage, 25 μF capacitance and 200 Ω resistance. The sample was removed from the chamber and 1.5 ml of LB broth was immediately added. Cells were incubated at 37^{o}C for 20 min in a static incubator followed by 40 min at 37^{o}C with shaking at 130 rpm, after which 100 μl of culture was plated out onto LB agar plates containing appropriate antibiotics. The plates were then incubated for 16 h at 37^{o}C.

2.5.5 Transformation of GAS

GAS cells were transformed with plasmid DNA by electroporation using a slightly modified version of the method described by Simon and Ferretti (Simon and Ferretti, 1991). Fresh competent cells were prepared using a single well isolated colony to inoculate 10 ml of THYB which was incubated at 37^{o}C with 5% CO_2 for 16 h. This culture was used to inoculate 200 ml of THYB (1/20) and incubated at 37^{o}C with 5% CO_2 until the OD_{600} of the culture reached 0.22. Once the desired cell density had been achieved the culture was centrifuged for 15 min at 2,355 x g to pellet the cells. The pellet was resuspended in 10 ml of ice-cold filter sterilised 0.5 M sucrose and cells were harvested again at 2,355 x g. This wash step was repeated twice more to remove any trace salts which may have impaired the electrotransformation. The washed cell pellet was finally resuspended in 1 ml of 0.5 M sucrose and 100 μl aliquots were dispersed in pre-cooled sterile microfuge tubes that were frozen rapidly in an ethanol-dry ice bath before storage at -80^{o}C.

For transformation of competent GAS, cells were allowed to thaw before the immediate addition 1-10 ng of plasmid DNA and the mixture was transferred to a pre-chilled Gene Pulser cuvette, 0.2 cm electrode gap (Bio-Rad). The sample was pulsed using a Gene Pulser II (Bio-Rad) under the following electroporator settings: 2.5 kV voltage, 25 μF capacitance and 200 Ω resistance. The sample was removed from the chamber and 1 ml of THYB was immediately added. The cells were incubated at 28^{o}C for 3 h in a static incubator to allow the expression of Erm^{r} markers on the temperature sensitive pG^{+}host9 plasmid. Culture was plated in volumes of 100 μl onto THY blood agar containing the appropriate antibiotics. Plates were incubated at 28^{o}C for 36-48 h to allow growth of single colonies.

2.5.6 Restriction endonuclease digestion and DNA ligation

Restriction endonucleases and T4 ligase together with appropriate buffers were purchased from standard commercial sources (section 2.2). Restriction endonuclease digestions and DNA ligations were carried out according to the manufacturer's protocol using appropriate supplied reaction buffers for the enzyme being used. Restriction endonuclease digests were performed in 40-50 μl reaction volumes and incubated at 37^{o}C for 2-3 h. When required resulting digest products were analysed by agarose gel electrophoresis (section 2.5.11) and fragments were purified using the Qiagen Gel Extraction kit (Qiagen Ltd, section 2.5.12). A reaction volume of 10 μl was used for DNA ligations and samples were incubated at 4^{o}C for 16 h.

2.5.7 Synthesis of oligonucleotide primers

Oligonucleotide primers were synthesised to order by VH Bio (Newcastle upon Tyne, UK). Primers were supplied in a lyophilised form and resuspended in 1 ml of double processed tissue culture water (Sigma-Aldrich Company Ltd.). Primer concentrations were adjusted to 10 pmol/μl, and aliquots were stored at -20^{o}C. The oligonucleotide primers used in this study are described in Table 2.3.

2.5.8 Polymerase chain reaction (PCR)

Standard endpoint PCR was performed essentially as described by Mullis (Mullis, 1990), using Phusion High Fidelity DNA polymerase (section 2.2) and HF buffer supplied by the manufacturer. Reaction mixes consisted of 1 X HF buffer, 0.2 mM of each dNTP, 20 pico-moles of each primer, 1 U Phusion High Fidelity DNA Polymerase and up to 1 µg of template DNA, made up to 50 µl with double processed tissue culture water (Sigma). Thermal cycling was performed using a Biometra T3000 Thermocycler (Biometra, Goettingen, Germany), and amplification of products was achieved using the following conditions:

Initial denaturation	98°C	30 s	
Denaturation	98°C	10 s	30 cycles
Primer annealing	x°C	30 s	
Primer extension	72°C	30 s / kb	
Final extension	72°C	10 min	
Holding step	4°C		

Primer annealing temperature (x) was estimated using the following equation, %GC x 0.41 + 69 – (650/total no. of bases) and was usually the lower melting temperature of the two primers being used. Annealing temperature could then be optimised if needed by changing the Tm of the reaction in 1°C steps. PCR products were visualised by agarose gel electrophoresis and the Tm which gave rise to the best yield of product was used in future reactions.

2.5.9 Site directed mutagenesis

2.5.9.1 Quickchange PCR

A modified quickchange PCR protocol (Stratagene) was used in site directed mutagenesis experiments to construct point mutations. PCR primers encompassing the desired point mutation were used with plasmids purified as described above (section 2.5.1) as template. 10 ng of template in 1 µl was added to a 24 µl reaction mixture containing 1 X buffer for KOD Hot Start DNA polymerase (section 2.2), 1.5 mM

$MgSO_4$, 0.2 mM dNTP's, 0.3 μM each primer, 0.02 U/μl KOD Hot Start DNA polymerase, made up to 25 μl with 15.5 μl double processed tissue culture water (Sigma). Amplification was carried out using the following conditions:

Hot Start	95°C	2 min		
Denaturation	95°C	20 s	}	18 cycles
Annealing	x°C	10 s		
Extension	70°C	25 s / kb		
Holding Step	4°C			

Digestion of the original template plasmid DNA was then achieved in a new reaction in which 10 μl of the above PCR mix was added to a 20 μl reaction mixture containing 16 μl double processed tissue culture water, 1X Tango reaction buffer and 1 μl Dpn-1 restriction enzyme (Fermentas). Dpn-1 specifically digests methylated DNA and consequently digests the original plasmid template while leaving the un-methylated PCR amplified products encompassing the desired point mutation intact. All reactions were incubated for 2 h at 37°C before plasmid DNA harbouring point mutations was purified using a QIAquick PCR purification kit (section 2.5.10, Qiagen Ltd.).

2.5.9.2 Overlap Extension PCR

In addition to quickchange PCR (section 2.5.9.1), overlap extension PCR was employed to construct point mutations in target genes as summarised in Figure 2.3. The two fragments of the target gene harbouring the overlapping sequence needed for the annealing stage were produced in separate forward and reverse 50 μl PCR reactions (panel A, Figure 2.3). Reaction mixes consisted of 1 X HF buffer (section 2.2), 0.2 mM dNTP's, 20 pico-moles of an outside primer with corresponding inside primer containing the mutation and 1 U Phusion High Fidelity DNA polymerase. Up to 1 μg of template GAS chromosomal DNA was added to the reaction mix, which was made up to 50 μl with double processed tissue culture water. Thermal cycling was performed on a Biometra T3000 Thermocycler (Biometra) using the following conditions:

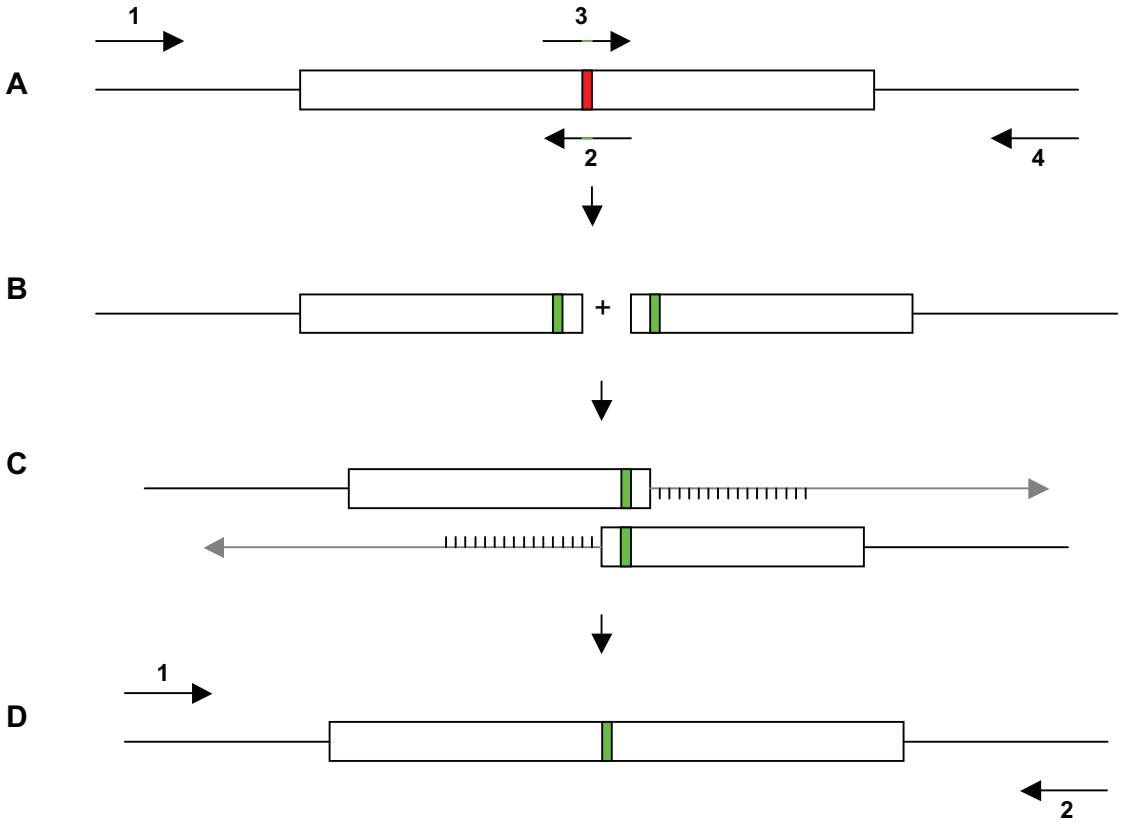

Figure 2.3: Overlap extension PCR.

Panel A: Bases to be mutated are shown in red. Primers depicted by arrows containing green line (2 and 3) represent overlapping mutagenic primers. An initial PCR reaction is set up to include primers labelled 1 and 2, and a second separate PCR reaction is performed using primers labelled 3 and 4. **Panel B**: This creates two fragments with overhangs each containing the mutated residues in green. **Panel C**: These are used in an annealing PCR reaction where all reagents are included in the PCR mix except primers and this allows complementary bases to be 'filled' in by DNA polymerase. **Panel D**: Outside primers used previously (1 and 4) are added and a normal PCR is performed to amplify the desired fragment.

Initial denaturation	98°C	30 s	
Denaturation	98°C	10 s	30 cycles
Primer annealing	x°C	30 s	
Primer extension	72°C	30 s / kb	
Final extension	72°C	10 min	

Amplified fragments (panel B, Figure 2.3) were analysed by agarose gel electrophoresis (section 2.5.11) and purified using a QIAquick PCR purification column (section 2.5.10 Qiagen Ltd.). Fragments to be used in an annealing PCR were diluted 1/10 and 1/50. For the annealing PCR, 1 µl of each purified fragment dilution was added to a reaction mixture containing 1 X HF buffer, 0.2 mM dNTP's, 1 U Phusion High Fidelity DNA polymerase and made up to 40 µl with double processed tissue culture water. Cycling conditions were as described above except an initial 10 cycles were used to anneal the fragments and complete the complementary strands of the template (panel C, Figure 2.3). Upon completion of the annealing PCR, 20 pico-moles of each 'outside' primer (labelled 1 and 4 in Figure 2.3) were added to the reaction and the PCR was continued as above, except with no initial denaturation step, for another 30 cycles to amplify the mutagenised target (panel D, Figure 2.3).

2.5.10 Purification of PCR amplified products

PCR products were purified using the QIAquick PCR purification kit (Qiagen Ltd.). This kit was used for the direct purification of double stranded PCR products (0.1-10 kb) from PCR amplification reactions. The kit was used as per the manufacturer's guidelines. Essentially, the method utilises a silica-gel membrane which absorbs DNA, only in high concentrations of chaotropic salts. After washing to remove contaminants, DNA was eluted from the membrane in 50 µl (or 30 µl to increase DNA concentration) 10 mM Tris, pH 8.5, and stored at -20°C.

2.5.11 Agarose gel electrophoresis

Agarose was purchased from FMC Bioproducts, Rockland, USA. Agarose gels of 0.7% (w/v) or 1.0% (w/v), depending on the size of DNA molecules being examined, were

cast in TAE buffer (40 mM Tris-acetate, 1.0 mM EDTA, pH 8.0) containing 0.5 µg/ml ethidium bromide. Gels were submerged horizontally in TAE running buffer. Prior to loading, a third volume of DNA load buffer [40% (w/v) sucrose, 100 mM EDTA, pH 8.0, 0.01% (w/v) bromophenol blue] was added to DNA samples. Once loaded, gels were run at a constant voltage (usually 100 V) until the bromophenol blue tracker dye had run sufficiently through the gel. DNA was visualised using exposure to long wave ultra violet (UV) light from a transilluminator (Bio-Rad Laboratories Ltd, Hemel Hampstead, Herts., UK) and images of gels were recorded.

2.5.12 Purification of DNA fragments from agarose gels

DNA fragments were subject to agarose gel electrophoresis as described above (section 2.5.11) until separation was sufficient for excision of bands from the gel (using a sterile scalpel blade). Purification of the DNA was carried out using a QIAquick Gel Extraction kit (Qiagen Ltd.) as per the manufacturer's protocol. Briefly, using the supplied solubilisation buffer DNA was released from the gel fragment and bound to the silica-gel membrane in the presence of high concentrations of chaoptropic salts. After washing to remove contaminants, DNA was eluted from the membrane in 50 µl (or 30 µl for increased DNA concentration), 10 mM Tris, pH 8.5 and stored at -20^{o}C.

2.5.13 DNA sequencing

DNA sequencing was performed to order by a commercial company. Between 2-5 µg of purified DNA was sent, with an appropriate sequencing primer (section 2.5.7) to Eurofins Genetic Services Ltd (Eurofins MWG Operon, Ebersberg, Germany) and sequenced using the dideoxy chain termination/cycle sequencing technique on an ABI 3730XL sequencing machine (Applied Biosystems, California, USA).

2.5.14 Allele replacement mutagenesis in GAS

The introduction of mutagenised DNA into the GAS chromosome was done using the allele replacement technique described by Fontaine *et al*. (Fontaine *et al*., 2003). The steps involved are summarised in Figure 2.4. The DNA to be integrated into the GAS

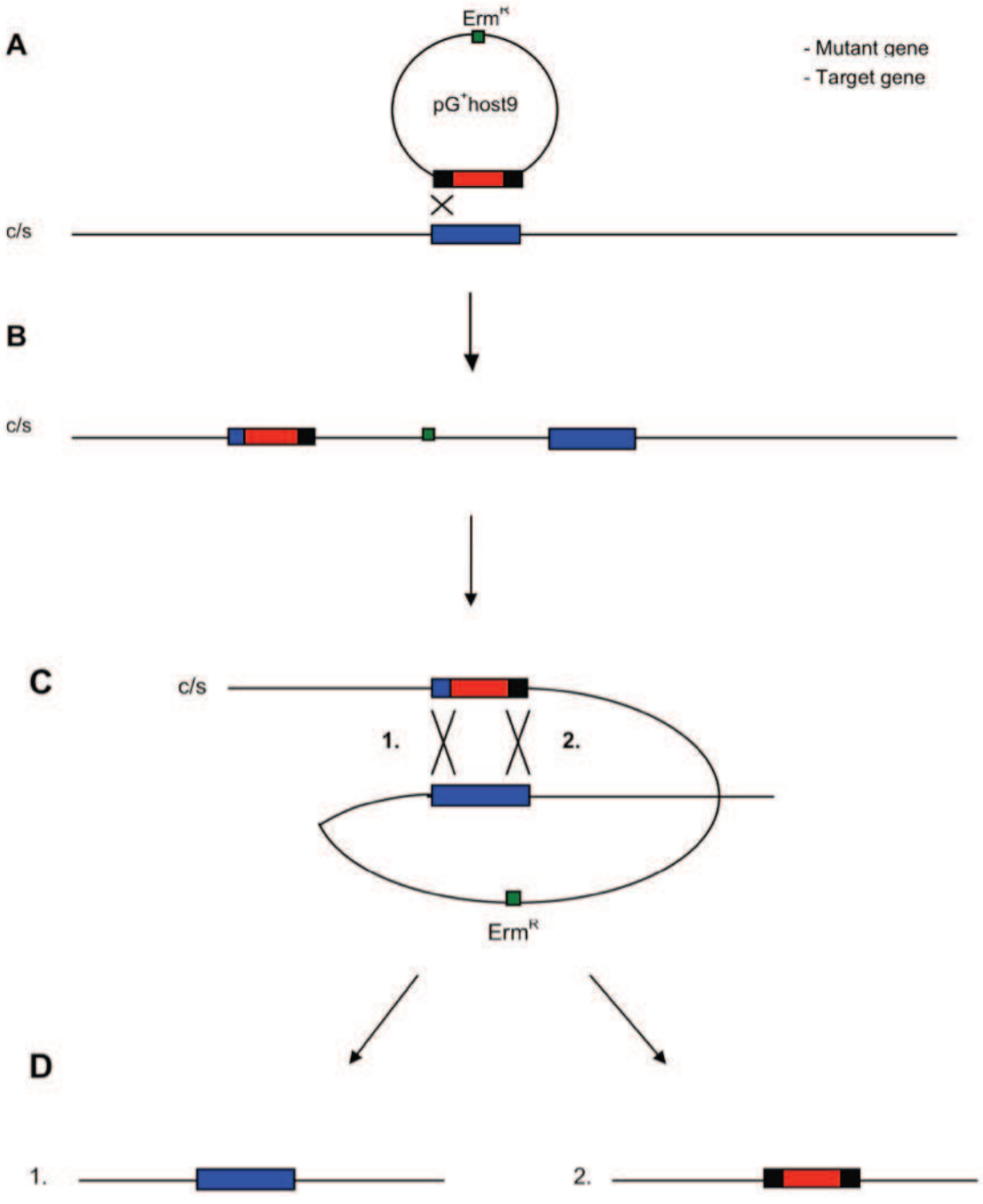

Figure 2.4: Allele replacement

Diagrammatic representation of allele replacement. **Panel A**: integration of the pG^{+}host9 plasmid by a single cross over recombination event at complimentary sequences contained within both the mutant and target gene. **Panel B**: This leads to incorporation of the entire plasmid into the chromosome (c/s). **Panel C**: a second cross over event occurs at complimentary sequences causing excision of the plasmid. **Panel D:** If this event occurs at position 1 the wild-type gene is retained in the c/s but if the second cross over event occurs at position 2, the mutant gene is retained in the c/s.

chromosome was cloned into the temperature sensitive plasmid pG$^+$host9. Competent GAS were transformed with the desired pG$^+$host9 construct by electroporation (section 2.5.5), selecting for single transformant colonies on THY blood agar plates containing 1 μg/ml erythromycin at 28^{o}C (the permissive temperature) for 48 h. A single transformed colony was used to inoculate 1 ml of THYB containing 1 μg/ml erythromycin and was incubated overnight at 28^{o}C. The overnight culture was diluted 1/100 into the same media and incubated at 28^{o}C until the OD_{600} of the culture reached 0.2-0.3, at which point the culture was shifted to the restrictive temperature of 37^{o}C, 5% CO_2 for 6-9 generations (approximately 12 h) to inhibit replication of unincorporated plasmid. Samples were then plated in decreasing 10-fold dilutions from neat to 10^{-4} onto blood agar plates containing 1 μg/ml erythromycin and incubated at 37^{o}C, 5% CO_2 to allow growth of cells in which a single cross over (sco) plasmid integration event had occurred, resulting in the integration of the entire plasmid into the bacterial chromosome.

A single sco colony containing the pG$^+$host9 construct integrated into the chromosome was used to inoculate 1 ml THYB with 1 μg/ml erythromycin added and the culture was incubated overnight at 37^{o}C, 5% CO_2 in a static incubator. The overnight culture was diluted 10^{-5} in drug free THYB and to stimulate the second cross over event the culture was incubated at 28^{o}C for no less than 12 h. This enables replication to be initiated from the integrated plasmid. The pG$^+$host9 plasmid replicates *via* the rolling circle mechanism and resulting single stranded DNA is believed to facilitate second recombination events. To identify Erm^s cells where plasmid excision had occurred, the culture was serially diluted in decreasing 10 fold dilutions to 10^{-6} in PBS and 100 μl of each dilution was plated onto drug free blood agar plates and incubated at 37^{o}C, 5% CO_2 in a static incubator. Sterile wooden tooth picks were used to replica plate single colonies from these plates first onto drug free blood agar and then onto blood agar plates containing 1 μg/ml erythromycin. These plates were incubated overnight at 37^{o}C, 5% CO_2 in a static incubator and colonies that grew on the non-selective plates but did not grow on the erythromycin plates were likely to have undergone a second recombination event deleting the integrated plasmid. Chromosomal DNA preparations (section 2.5.2) were prepared for these colonies and screened by PCR to identify those

containing the desired mutation, which was analysed by sequencing to confirm a double cross over event had occurred.

2.6 Protein procedures

2.6.1 Separation of proteins by discontinuous sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE gels were run essentially as described by Laemmli (Laemmli, 1970) on a Mini Protean II system (Bio-Rad, Hemel Hampstead, Hertfordshire, UK). Resolving gels of appropriate percentage were prepared by mixing either 5.328 ml (8% v/v), 6.66 ml (10% v/v), 8 ml (12% v/v) or 10 ml (15% v/v) of 30% (w/v) acrylamide/bisacrylamide with 5 ml 1.5 M Tris-HCl pH 8.8, 0.2 ml 10% (w/v) SDS made up to 20 ml with water and 170 µl 10% (w/v) ammonium persulphate. Polymerisation was achieved with the addition of 10 µl of N,N,N',N'–Tetramethylethylenediamine (TEMED) immediately prior to pouring. The resolving gel was poured between two glass plates separated by 1 mm until the meniscus reached 2 cm from the top of the smaller plate. Isopropanol was overlaid on top of the resolving gel until it had set to ensure a level surface before being removed by extensive washing with double distilled water. A 5% (v/v) stacking gel was prepared with 830 µl of 30% (w/v) acrylamide/bisacrylamide, 0.63 ml 1M Tris-HCl pH 6.8, 50 µl 10% (w/v) SDS made up to 5 ml with water and 50 µl ammonium persulphate. Polymerisation of the stacking gel was achieved with the addition of 5 µl TEMED immediately prior to pouring on top of the resolving gel. A Teflon comb was inserted into the liquid stacking gel to provide wells. After setting, the gel was placed into the Mini Protean II electrophoresis tank and the comb was removed. The gel was completely covered and both chambers were filled with running buffer [25 mM Tris-HCl, 250 mM glycine and 0.1% (w/v) SDS].

Protein samples were prepared for loading by mixing with SDS load buffer, which consisted of 50 mM Tris-HCl pH 6.8, 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol and 2% (w/v) SDS, before being boiled at 100^{o}C for 5 min to denature proteins. Samples were loaded into wells using gel loading tips (Starlab, Milton Keynes,

UK) and electrophoresis was performed at a constant voltage (usually 250 V) for 40 min or until the bromophenol blue tracker dye had reached the bottom of the resolving gel.

2.6.2 Coomassie staining of SDS-PAGE gels

Protein gels were stained with coomassie blue essentially as described by Laemmli (Laemmli, 1970). Gels were removed from the glass plates and immersed in staining solution which contained 40% (v/v) methanol, 45% (v/v) acetic acid and 0.1% (w/v) Coomassie Brilliant Blue R250 and left for 1 hour with gentle shaking. Gels were rinsed with double distilled water and destained for several h in a solution containing 30% (v/v) methanol, 10% (v/v) acetic acid, with occasional replacement of the destain solution. Destained gels were recorded by scanning onto a computer using a Canon CanoScan 3200F (Canon Inc., Tokyo, Japan).

2.6.3 Production of recombinant proteins

BL21 (DE3) and BL21 (DE3) CodonPlus Type RIL host strains were freshly transformed with the desired recombinant plasmid and trial inductions were carried out on several transformants from each host as soon as possible following transformation. Individual transformants were used to inoculate 1.5 ml of LB broth supplemented with kanamycin for BL21 (DE3) or kanamycin and chloramphenicol for BL21 (DE3) CodonPlus Type RIL as described in section 2.3, and cultures were grown for 16 h at 37^{o}C with shaking at 200 rpm. These cultures were used to inoculate 10 ml of fresh LB medium containing antibiotics (1/20) and cells were grown to OD_{600} 0.6-0.7 at 37^{o}C with shaking at 200 rpm. Protein expression was induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to final concentration 1 mM and continuing incubation at 30^{o}C and 200 rpm. Samples of 1 ml were taken before induction and at hourly intervals post-induction from each culture for up to 4 h. Samples were harvested by centrifugation at 16,000 x g and resuspended in SDS load buffer. Samples were boiled and cell lysate was assessed for protein expression by SDS PAGE (section 2.6.1) to identify optimal conditions for large-scale expression. Once identified, optimised conditions were applied to induction of protein in 1 litre of culture as outlined above. Cells were harvested at 2,355 x g at 4^{o}C for 20 min and stored at -80^{o}C.

2.6.4 Selenomethionine derivatisation of proteins

To prepare protein labelled with selenomethionine, freshly competent B834 (DE3) methionine auxotroph cells were transformed with the appropriate plasmid and a single well separated colony was used to inoculate 10 ml of LB broth containing the appropriate antibiotic. Following incubation at 37^{o}C for 16 h with shaking at 200 rpm, this culture was used to inoculate 100 ml of fresh LB broth, 1/100. This culture was incubated as above until the culture OD_{600} reached 0.2. Cells were harvested by centrifugation at 2,355 x g for 10 min at 4^{o}C. Media to support growth of B834 (DE3) and selenomethionine incorporation (1 l) used 2X M9 media (2 g NH_4Cl, 6 g KH_2PO_4, 12 g Na_2HPO_4), 2 ml of 1 M $MgSO_4$, 25 mg of $FeSO_4$, 10 ml of 40% (w/v) glucose, 1 ml of vitamin mix (1 mg/ml of Niacinamide, Pyroxidine monohydrochloride, Riboflavin and Thiamine), 40 mg of each L-amino acid (excluding methionine and selenomethionine) and the appropriate antibiotic. The cell pellet obtained above was washed in 10 ml of this media and harvested as above to remove traces of LB broth. This pellet was resuspended in a minimal volume and added to 1 l of the above media followed by addition at this stage of 40 mg L(+)-Selenomethionine (Acros Organics, Geel, Belgium). The culture was then incubated at 37^{o}C with shaking at 200 rpm. Cells were induced with IPTG (1 mM final concentration) when the culture density reached between 0.4-0.6 and incubated for 16 h at 20^{o}C with shaking at 200 rpm. Cells were harvested by centrifugation at 2,355 x g and stored at -80^{o}C.

2.6.5 Purification of recombinant proteins

Purification of all proteins was done using an Äktaprime™ system (GE Healthcare, Bucks, UK). All proteins in this study were expressed with an N-terminal hexa-histidine tag to aid purification.

2.6.5.1 His-tag affinity chromatography

Pelleted cells containing recombinant protein were resuspended in 20 ml 'His-load' buffer (20 mM Tris-Cl, 150 mM NaCl, 10 mM Imidazole, pH 7.5) with 0.025% (v/v) β-mercaptoethanol and 100 µl Protease Inhibitor Cocktail Set III (Calbiochem, Darmstadt, Germany) added. Cells were lysed using a Sonoplus HD2070 sonicator (Bandelin,

Berlin, Germany) and the supernatant clarified by centrifugation at 15,000 x g for 40 min. Clear lysate was loaded onto a pre-equilibrated 5 ml Hi-Trap nickel chelating column (pre-equilibrated with nickel, GE healthcare). The column was washed through with 'His-load' buffer until absorbance at 280 nm returned to baseline to remove any non-specifically bound proteins. Bound protein was eluted with an imidazole gradient (10-500 mM) using 'His-elute' buffer (20 mM Tris-Cl, 150 mM NaCl, 1 M imidazole, pH 7.5) over 12 column volumes. Fractions were collected with continuous monitoring of absorbance at 280 nm and peak fractions were analysed by SDS-PAGE (section 2.6.1).

2.6.5.2 Size exclusion gel filtration chromatography

Peak fractions from affinity chromatography containing the protein of interest were pooled and concentrated in an appropriate molecular weight cut off centrifuge cell (Sartorius, GmbH, Goettingen, Germany). Concentrated protein was centrifuged at 15,000 x g at 4^{o}C for 30 min to pellet any insoluble material prior to loading 2 ml volumes onto a Hi-Load 16/60 Superdex 75 or Hi-Load 16/60 Superdex 200 gel filtration column [GE Healthcare, Bucks, UK, pre-equilibrated with gel filtration buffer (20 mM Tris-Cl, 150 mM NaCl, pH 7.5)], depending on the molecular weight of the protein. Gel filtration columns were run at 1 ml/min for 2 column volumes with continuous monitoring of absorbance at 280 nm. Peak fractions were analysed by SDS-PAGE (section 2.6.1) and those containing the protein of interest at >95% purity were pooled and concentrated to 10 mg/ml in an appropriate molecular weight cut off centrifuge cell. Protein concentration was determined by absorbance at 280 nm using the protein's molar extinction coefficient (the amount of light a protein absorbs at a given wavelength) as determined by the ProtParam program (Gasteiger *et al*., 2005) located on the Expert Protein Analysis System proteomics server (http://www.expasy.ch/). For establishing protein concentration, the total absorbance of a sample was divided by the protein of interest's molar extinction coefficient and adjusted accordingly.

2.6.5.3 Purification of selenomethionine labelled proteins

The purification of selenomethionine labelled proteins was as described above but with the addition of 15% (v/v) β-mercaptoethanol to both His-chelation buffers in section 2.6.5 and 5 mM DTT to the gel filtration buffer in 2.6.5.2. This was to ensure that the selenomethionine remained reduced during purification.

2.6.7 Thrombin digest

When necessary, the N-terminal His-tag was cleaved from the expressed proteins using thrombin (GE Healthcare, Bucks, UK). To establish appropriate conditions, test digests were performed by mixing 1 mg of protein with 10 units of thrombin in 20 mM Tris-Cl, 150 mM NaCl, pH 7.5 and incubating at 4^{o}C and 20^{o}C. Samples were taken at 0 min, 30 min, 60 min and then hourly for a further 5 h before a final sample was taken after 16 h. The samples collected at both temperatures were analysed by SDS- PAGE (section 2.6.1) with a decrease of ~2 kDa indicative of tag removal. The optimal temperature and digest time (derived from these tests, usually 16 h at 4^{o}C) were applied to a large scale experiment where 10 mg of protein was digested with 100 units of thrombin in 20 mM Tris, 150 mM NaCl, pH 7.5. After the digest was complete, the sample was applied to a 1 ml His-chelation column (pre-equilibrated with Ni^{2+}, prepared as above), then washed with 3 ml 'His-load' buffer (section 2.6.5.1). The flow-through, which contained cleaved protein, was then applied to a pre-equilibrated 1 ml benzamidine column (GE Healthcare, bucks, UK) to remove protease from the sample, again washing with 3 ml 'His-load' buffer and collecting the flow through. The resulting samples were analysed by SDS-PAGE to ensure pure, cleaved protein had been obtained. The sample concentration was determined by absorbance at 280 nm and concentrated to 10 mg/ml (as before 2.6.5.2).

2.6.8 Extraction of GAS cell wall associated proteins

A single, well-separated GAS colony was used to inoculate 40 ml THYB and cultures were incubated at 37^{o}C with 5% CO_2 for 18 h until they had reached an OD_{600} of 0.6-0.7. Cell density was checked and adjusted if needed so that all cultures were at the same stage of growth. GAS were pelleted by centrifugation at 5,000 x g for 10 min. The supernatant, which could be analysed for any proteins released prior to enzymatic

digestion of the cell wall, was collected and passed through a 0.22 μM pore size filter to remove any remaining bacteria, and the sample was concentrated 40-fold in a 10 kDa cut off centrifuge cell and stored at -80^{o}C.

The pelleted bacterial cells were resuspended and washed in 5 ml of ice-cold TE buffer (50 mM Tris-Cl pH 8.0, 1mM EDTA) and pelleted as above. The wash step was repeated and cells were resuspended in 1 ml TE sucrose buffer (50 mM Tris-Cl pH 8.0, 1 mM EDTA, 20% (w/v) sucrose, 100 mg/ml lysozyme, 100 U mutanolysin and 400 U protease inhibitors) and the mixtures were incubated for 2 h at 37^{o}C to digest cell walls, with constant rotation to aid mixing. After incubation resulting protoplasts were pelleted by centrifugation at 16,000 x g for 5 min and the supernatant, defined as the cell wall extract containing the enzymatically digested cell wall material was harvested into a fresh, sterile 1.5 ml microfuge tube and stored at -80^{o}C. The insoluble cell extract containing the bacterial protoplasts and cell debris, was resuspended in 1 ml H_2O and stored at -80^{o}C.

2.7 Production and purification of rabbit antibodies

2.7.1 Production of specific anti-sera

To produce the inoculum, 700 μl of highly purified recombinant protein at a concentration of 0.4 mg/ml was dialysed into PBS (Sigma). Protein was mixed with Freund's adjuvant (1:1) by passaging through a steel tube held between two sterile glass syringes until a stable white emulsion, extremely resistant to movement had formed. Prior to immunization, 1.4 ml of sterile 2% (v/v) Tween-80 in saline was added and mixed to give a milky suspension. For each antigen, a single, female, New Zealand White rabbit weighing >3 kg was purchased from Charles River UK Ltd. (Margate, Kent, UK). Rabbits were rested for 10-14 days prior to collection of 5 ml pre-immune blood into a sterile glass universal tube *via* a single nick in an ear vein. This sample was incubated at 37^{o}C for 30 min to facilitate clotting and then centrifuged at 2,500 x g for 20 min. Clear serum was removed and stored at -20^{o}C. The above inoculum mix was injected subcutaneously into multiple sites on the rabbits back. After 2 weeks the rabbit was boosted with an inoculum prepared as described above. A 5 ml blood sample was

collected as discussed above, two weeks after boosting and serum was checked for production of specific antibodies by immuno-blotting samples of antigen spotted onto nitrocellulose membrane. Rabbits were boosted on a maximum of 4 occasions at two-weekly intervals. Five days after the last boost, rabbits under terminal anaesthesia were bled by cardiac puncture and serum was collected and stored as above.

2.7.2 Purification of antibodies

A 2.5 ml Protein A affinity column (Sigma) was used to isolate the IgG class of antibodies. The column was prepared by washing with 5 ml of loading buffer (1 M potassium phosphate, pH 9.0). Prior to loading, 1 volume of antiserum was mixed with 3 volumes of loading buffer. Any unbound sample was collected for re-processing before the column was washed with 12 ml of loading buffer. Bound IgG antibodies were eluted with 0.1 M citric acid, pH 3.0 and 1.25 ml fractions were collected in sterile microfuge tubes containing 250 µl of 1.5 M Tris-Cl, pH 8.5 to neutralize the acid. Fractions determined to contain protein at 280 nm were pooled and dialysed into PBS, pH 7.0 and stored at -20^{o}C.

2.8 Blotting procedures

2.8.1 Western blotting

Enzymatically extracted cell wall proteins (section 2.6.8) were run on precast Tris-Cl 4-15% (v/v) gradient gels (Bio-rad). Separated proteins were electroblotted onto nitrocellulose membrane in transfer buffer [200 mM glycine, 25 mM Tris base, 20% (v/v) methanol] for 90 min at 35 mAmps using a Bio-Rad Mini Trans-blot cell. After blotting, all free protein sites on the nitrocellulose membrane were ‘blocked’ by submerging the filter in 10 mM Tris-Cl, 150 mM NaCl, 2 mM EDTA, 5% (w/v) bovine serum albumin, 0.05% (v/v) Triton X-100, pH 7.5 and incubating at 4^{o}C for 16 h. ‘Blocking’ buffer was removed and the membrane was subjected to 3 successive 7 min washes in wash buffer 1 [10 mM Tris-Cl, 2 mM EDTA, 150 mM NaCl, 0.1% (w/v) BSA, 0.01% (v/v) Triton X-100, pH 7.5]. Primary antibodies were then added to the membrane, made up to an appropriate dilution in wash buffer 1, and the membrane was incubated at room temperature for 1 hour with gentle shaking. Any unbound primary

antibodies were washed from the membrane with 3 successive 7 min washes in wash buffer 1, prior to the addition of an appropriate peroxidise conjugated anti-primary antibody at an appropriate dilution in wash buffer 1. The membrane was incubated for 1 hour at room temperature with gentle shaking before any unbound secondary antibody was washed off by 3 successive 7 min washes in wash buffer 2 [10 mM Tris-Cl, 2 mM EDTA, 1 M NaCl, 0.1% (w/v) bovine serum albumin, 0.01% (v/v) Triton X-100, pH 7.5] followed by rinsing in 10 mM Tris-Cl pH 7.5. The peroxidise conjugated antibody was then visualised by submerging the membrane in developing solution. This was comprised of solution A (30 mg 4-chloro-1-naphthol dissolved in 10 ml ice cold methanol, made up 15 min prior to use) mixed with solution B (30 ml ice cold 10 mM Tris-Cl to which 30 µl H_2O_2 was added immediately prior to use) immediately before addition to the membrane. This was incubated at room temperature with gentle shaking. The staining reaction was stopped after 5-30 min by washing the membrane in copious amounts of water.

2.8.2 Electroblotting of proteins for N-terminal sequencing

Immediately prior to use, polyvinylidene difluoride (PVDF) membrane was submerged in 100% (v/v) methanol for 5 min before being transferred to 10 mM CAPS pH 11, 50% (v/v) methanol, 0.02 mg/ml DTT and finally soaked in 10 mM CAPS pH 11, 0.02 mg/ml DTT to prepare membrane for use. Proteins to be investigated by N-terminal sequencing were run on an SDS-PAGE gel which was rinsed in ultra pure water containing 0.02 mg/ml DTT and then soaked in 10 mM CAPS pH 11, 0.02 mg/ml DTT for 10 min with one buffer change after 5 min. Proteins were blotted onto the prepared PVDF membrane at a constant current of 150 mA for 60 min in blotting buffer [10 mM CAPS pH 11, 10% (v/v) methanol]. After blotting the PVDF membrane was rinsed with ultra pure water containing 0.02 mg/ml DTT and stained using 2 mg/ml Coomassie Brilliant Blue R250 in 5:4:1 methanol:water:acetic acid. The membrane was then destained in the same solvent as above until protein bands appeared, at which point the membrane was removed and rinsed in ultra pure water containing 0.02 mg/ml DTT before air drying.

2.9 Mass spectrometry procedures

2.9.1 Fourier Transform-Ion Cyclotron Resonance mass spectrometry (FT-ICR MS)

Protein concentration was determined by absorbance at 280 nm (as above, 2.6.5.2) followed by dilution to 1 mg/ml in standard gel filtration buffer (20 mM Tris-Cl, 150 mM NaCl, pH 7.5). Protein samples were analysed by Pinnacle Laboratory (Newcastle University) using an LTQ-FT Ultra Hybrid Mass Spectrometer (Thermo Scientific, Waltham, MA, USA). Data was collected at a resolution of 400,000 (at m/z = 400). Purified protein, at 10 pmoles/µl in acetonitrile/0.1% (v/v) formic acid (25%/75%), was sprayed directly into the mass spectrometer using an 'Ion Max' ion source (ThermoElectron, Bremen, Germany) at 10 µl/min. The QualBrowser program (ThermoElectron, Bremen, Germany) was used to deconvolute mass spectrum plots to determine the average molecular mass of the protein assuming 'averagine' (0.2678% sulphur) composition.

2.9.2 N-Terminal amino acid sequencing

Proteins of interest were blotted onto PVDF membrane (as described 2.8.2) and submitted for N-terminal protein sequencing at Pinnacle Laboratory (Newcastle University) using the Edman degradation technique on a Beckman Coulter LF 3000 Protein Sequencer (Beckman Coulter, Fullerton, CA, USA).

2.10 Solution biophysical characterisation techniques

2.10.1 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy refers to the differential absorption of the two circularly polarised components of plane polarised light (Kelly *et al*., 2005). One rotates clockwise, the other anti-clockwise and when these interact with a protein, its electronic structure gives rise to characteristic features in specific regions of the CD spectra. For use in CD experiments, the concentration of each stock protein solution was determined by absorbance at 280 nm followed by dilution to 0.5 mg/ml immediately before the spectra was examined. Spectra were recorded in 20 mM sodium phosphate, pH 7.5 (unless other wise stated). Circular dichroism spectra were collected between 180 nm

and 250 nm using a JASCO-810 spectropolarimeter (Jasco, Tokyo, Japan) fitted with a Peltier temperature controller, at 25^{o}C using a 0.2 mm-pathlength cuvette. Five repeat scans were used to generate the final spectra with appropriate protein-free buffer scan subtracted. The data was plotted without smoothing in Microsoft Excel (Microsoft, USA).

2.10.2 Analytical ultracentrifugation (AUC)

The sedimentation velocity technique was used in this study to apply a centrifugal force to samples. This leads to the formation of a concentration boundary as protein molecules are depleted at the meniscus, which moves to the bottom of the cell as a function of time (Lebowitz *et al*., 2002). Sedimentation velocity (SV) experiments were carried out by the Analytical Ultracentrifugation facility (Newcastle University) using a Beckman Coulter ProteomeLab XL-I analytical ultracentrifuge. An experimental temperature of 4^{o}C was used, with samples mounted in an 8-hole AnTi50 rotor with double sector aluminium-epon centrepieces. The centrifuge was operated at a rotation speed of 48,000 rpm. Protein concentrations ranged from 0.15-2 mg/ml, using 7 sample dilutions, in a total volume of 400 µl in standard gel filtration buffer.

2.10.3 Small angle X-ray scattering (SAXS)

SAXS was used to provide information on the size and shape of proteins in solution (Neylon, 2008; Petoukhov and Svergun, 2007). SAXS is carried out on samples in solution containing protein molecules that are randomly orientated throughout the sample. The most common use for SAXS data is to reconstruct the overall 3-dimensional shape of the molecule by *ab initio* shape restoration, which provides a low resolution molecular envelope describing the overall shape of the molecule (Koch *et al*., 2003). AUC is highly complementary to SAXS by providing information about the monodispersity, elongation and aggregation of the sample in solution independent of the SAXS data (Neylon, 2008; Hura *et al*., 2009). Solution X-ray scattering experiments were carried out by Dr. Alexandra Soloyova on beamline 2.1 at the Synchrotron Radiation Source (SRS, Daresbury, UK, at a wavelength of 1.54 Å) or using the X33 camera at the European Molecular Biology Laboratory Hamburg outstation on the storage ring DORIS III of the German Electron Synchrotron (DESY, at a wavelength of

1.5 Å). Protein concentrations used were 5 and 10 mg/ml in gel filtration buffer and measurements were recorded at 4°C. The data for the first and last frames were compared to check for protein aggregation and radiation damage.

2.11 X-ray crystallography

2.11.1 Initial crystallisation trials

Initial protein crystallisation experiments were setup in a 96-well sitting drop format (panel A, Figure 2.5) using the JCSG and Classics sparse matrix screens (as supplied by Qiagen Ltd.). The main well of the crystallisation plate was filled with 100 µl of precipitant solution and a Mosquito nanolitre pipetting robot (TTP LabTech, Royston, UK) was used to setup sitting drops containing 0.5 µl of 10 mg/ml protein mixed with 0.5 µl of precipitant solution in the sub well. After sealing (with adhesive tape), plates were stored at 20°C and checked daily for the growth of protein crystals. In conditions where protein crystals were observed, fresh reagents were used to scale up the crystallisation condition to produce diffraction quality crystals. The sparse matrix condition was varied by altering the pH of the buffer and the concentration of salts/precipitants. Each condition was made up separately in a microfuge tube and centrifuged at 18°C for 10 min at 13,000 x g. Protein samples were centrifuged at 13,000 x g for 30 min at 4°C before use to pellet any precipitate. The hanging drop vapour diffusion technique (panel B, Figure 2.5) was used to set up drops on plastic cover slips containing 2 µl of 10 mg/ml protein and 2 µl of precipitant solution which were sealed over the well with grease.

2.11.2 Data collection

Litholoops (Molecular Dimensions Ltd., UK) were used to transfer crystals into paratone-N oil, which was used as the cryoprotectant in this study, and to subsequently freeze crystals in liquid nitrogen. All X-ray diffraction data were collected using the standard oscillation method on cryo-cooled crystals at synchrotron beamlines, including IO3 of the Diamond Light Source (DLS, Oxford, UK) and BM14 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France).

A. Sitting drop B. Hanging drop

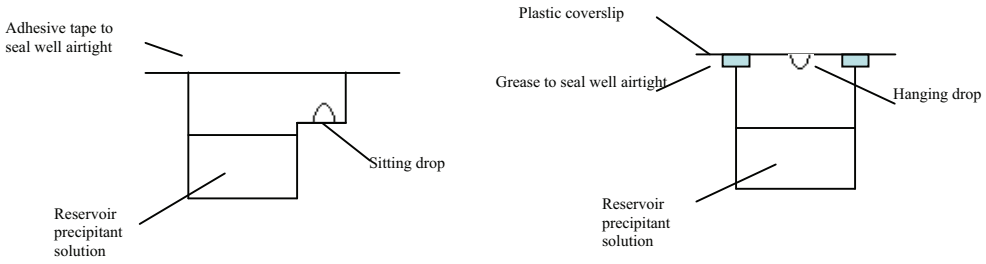

Figure 2.5. Sitting and hanging drop vapour diffusion techniques.

2.11.3 Data processing and reduction

X-ray diffraction images for each dataset were processed using iMOSFLM (Leslie, 2006). This program attempts to automatically determine crystal parameters such as the unit cell dimensions, crystal orientation, point group and can estimate a value for the mosaicity. This process predicts where in reciprocal space conditions satisfying Bragg's Law occur and hence a diffraction spot [Miller index (*hkl*)] is expected. These estimates are then refined to elucidate more accurate parameters enabling the images to be integrated (with further refinement of crystal parameters to best predict *hkl* positions). An experimental measurement of each diffraction spot intensity is obtained (I_{hkl}), along with their standard uncertanties (Leslie, 2006).

Following integration, the data was reduced and scaled using SCALA (Evans, 2006). This program puts all observations on a common scale using the redundancy of multiple measurements of symmetry related reflections (Evans, 2006). An internal control of the quality of crystallographic data is given by *R* factors (reliability factors). Analysis of the data after scaling used R_{merge}, which gives a measure of overall data quality. These statistics can highlight regions of a dataset which agree poorly with the rest of the data (and maybe discarded) and can also reveal whether the crystal has suffered radiation damage during data collection.

SCALA was used as supplied within the Collaborative Computational Project no. 4 suite, 1994.

2.11.4 Obtaining phase information

To produce an interpretable electron density map and ultimately solve a crystal structure two things are needed: amplitudes, which are measured during the diffraction experiment, and phases, which are lost during data collection. The phase problem is a fundamental issue in crystallography and although it is possible to solve small molecule structures by direct methods with extremely high resolution data, this is generally not applicable to most protein structures.

There are three 'umbrella' methodologies used to solve the phase problem in protein crystallography and all of these methods for calculating phases assume some prior knowledge of the electron density or structure (Taylor, 2003). Molecular replacement can be used when there is a homology model of the protein of interest available with typically, at least 25% sequence identity. Initial phases can be generated by positioning the known model correctly in the new unit cell. When there is not a pre-existing model known to be homologous to the protein of interest, so called 'experimental phasing' techniques are employed.

2.11.4.1 Isomorphous replacement

Isomorphous replacement usually involves soaking a native protein crystal in a heavy metal salt solution which can migrate through solvent channels within the crystal and bind to specific sites on the protein. This leads to measurable changes in the intensities collected during a diffraction experiment. As long as the heavy atom derivative crystal remains 'isomorphous' (same unit cell dimensions and orientation of the protein in the unit cell, 'replacement' refers to the fact that the heavy atom is probably replacing a water or other solvent molecule in the native crystal) the positions of the heavy atoms can be solved by Patterson or direct methods and initial phase estimates can be derived for the heavy atoms. These can then be 'bootstrapped' onto all of the atoms in the protein to produce an initial electron density map for the entire crystal. The use of only

one heavy atom derivative creates a twofold phase ambiguity which frequently has to be broken with the use of a second (or more) heavy atom compound (that binds to different sites on the protein), and this technique is known as Multiple Isomorphous Replacement (MIR). The phase ambiguity can also be resolved with the use of anomalous scatter (section 2.11.4.2) in conjunction with isomorphous differences in the Single Isomorphous Replacement with Anomalous Scatter (SIRAS) technique. Screening for derivatised crystals can be a time consuming and largely trial and error process. Native PAGE can be used to search for gel shifts caused by interaction of the protein with a heavy metal compound in solution (Boggon and Shapiro, 2000), suggesting a potential interaction between the heavy atom compound and protein within the crystal.

2.11.4.1.1 Screening heavy metal salt solutions by native PAGE

Protein derivatisation by heavy metal salt compounds in solution was carried out essentially as described by Boggon and Shapiro (Boggon and Shapiro, 2000). Mixtures that contained 2 µl of 2 mg/ml protein and 2 µl of 5 mM heavy metal salt solution were incubated on ice for 10 min. Samples of 1 µl mixed with 6 µl of loading dye (1 M Tris-Cl pH 8.5, 20% (v/v) glycerol and 1% (w/v) bromophenol blue) were loaded onto an 8% (v/v) native PAGE gel that was prepared by mixing 1.33 ml of 30% (w/v) acrylamide/bisacrylamide, 1.66 ml H_2O, 1 ml of 1 M Tris-Cl pH 8.5 and 50 µl of 10% (w/v) ammonium persulphate, polymerised with the addition of 5 µl TEMED immediately prior to pouring. Electrophoresis was performed at a constant voltage of 200 V for 90 min.

2.11.4.1.2 Soaking crystals in heavy atom salt solutions

Crystals were soaked in varying concentrations of heavy atom compounds dissolved in a stabilising solution, which was composed of the crystal mother liquor with a 1-2% increase in major precipitant concentration. Soaks were performed in sitting drop cross bridges (Molecular Dimensions Ltd., UK), placed in a 24 well plate with 300 µl of stabilising solution added to wells to keep soaks hydrated. The well was then sealed with a plastic coverslip.

2.11.4.2 Anomalous diffraction

When X-ray energy (aligned with wavelength) is close to the absorption edge of an atom (e.g. the *K* edge of selenium), some of the radiation is re-emitted with altered phase. This causes a breakdown in Friedel's law (which states that $I_{hkl} = I_{-h-k-l}$), and these anomalous differences can be used in a similar way to isomorphous differences, in a derivative dataset, to determine the positions of the anomalous scatterers and extract phase information (Taylor, 2003). Anomalous differences are much smaller than isomorphous differences and the total anomalous diffraction only generates small differences in intensities (Dodson, 2003). However, due to advances in tunable synchrotron beams, CCD detectors and data collection and phasing algorithms (especially when used in conjunction with cryo-cooled crystals), these differences can be accurately measured to a fraction of a percent of the total scattered intensity (Dauter, 2006).

When X-ray energy is close to the absorption edge of an atom an electron from an inner shell is promoted to a higher shell; when it decays back to its original shell, energy is released as fluorescence. This information can be used to optimise a data collection strategy to maximise the anomalous signal from the crystals by using a fluorescence scan around the relevant absorption edge. This requires a tunable synchrotron beamline, and can result in optimal absorptive (*f''*) and dispersive (*f'*) values.

A textbook excitation spectrum is shown in Figure 2.6 and the program CHOOCH (G. Evans, included as part of the CCP4 suite) is often used to automatically determine the *f'* and *f''* values from the X-ray fluorescence data. It is important to examine the fluorescence spectrum for each derivative crystal directly as experimentally determined values for *f'* and *f''* can vary between crystals and also from the theoretical values.

Two major techniques for deriving phase information using anomalous diffraction are used. The first is single-wavelength anomalous diffraction (SAD), which involves collecting highly redundant data at the high energy side of the absorptive edge (at the peak wavelength, *f''*). This results in a phase ambiguity in a similar way to single

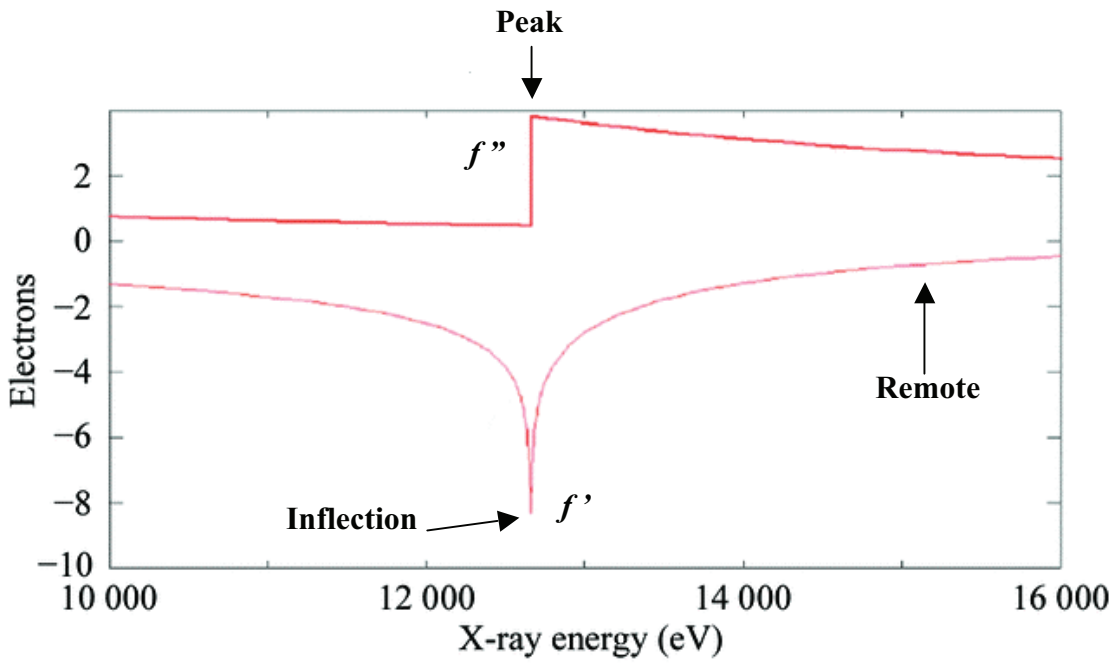

Figure 2.6: Example excitation spectrum at the *K* edge of selenium.

Peak absorption spectrum (*f”*) is shown alongside the dispersive spectrum (*f’*). Using a tunable beamline at a synchrotron a fluorescence scan is performed to determine maximised values for *f”* and *f’* and the three wavelengths chosen to perform a MAD experiment are highlighted (peak, inflection and remote). Figure is adapted from Taylor, 2003.

isomorphous replacement, but this can be resolved with powerful phasing and density modification algorithms (Dodson, 2003). Multi-wavelength anomalous diffraction (MAD) also exploits anomalous diffraction as a phasing tool. During collection of a MAD dataset, data is typically collected at three wavelengths to maximise the anomalous differences (*f”*) and measure the minimum dispersive differences (*f’*). The third wavelength known as the ‘remote dataset’ is collected to provide a measure of the *f’* value away from the edge (highlighted as remote in Figure 2.6). The absorptive and dispersive peaks can be separated by as little as 2 eV which makes it critical to determine these values experimentally. Anomalous and dispersive differences can be used to locate the heavy atom positions in the unit cell, and these positions can be used to initiate phase calculations.

2.11.5 Structure solution (calculating the first interpretable electron density map)

The technological advances in data collection in recent years have been matched by advances in algorithms able to derive maximal information from such data. Many of these algorithms have been incorporated into structure-solution packages that can take X-ray diffraction data, attempt to derive initial phases directly, follow this with density modification (to improve the phases) and ultimately build a protein model into the initial electron density map.

The AutoSol wizard of PHENIX (Adams *et al*., 2002) can attempt to solve protein structures directly from MAD datasets. The AutoSol wizard utilises HYSS (Grosse-Kunstleve and Adams, 2003) to locate heavy atom sites and their substructure positions are subsequently used by SOLVE (Terwilliger and Berendzen, 1999) to determine the phases. Density modification (Terwilliger, 2000) is performed on the best scoring electron density maps from SOLVE (Terwilliger and Berendzen, 1999) by RESOLVE (Terwilliger, 2003) which also attempts autobuilding when supplied with a primary protein sequence. Finally, refinement with phenix.refine produces an initial model.

2.11.6 Refinement/model building and validation

The initial model created by either the automatic building programs or by manual chain tracing in a molecular graphics program will likely obtain a correct fold but will also contain many incorrect features. These include poorly positioned loops between secondary structure elements and other regions where the electron density poorly defines the model; this is because the experimentally derived structure factors [$\left|F_{obs}\right|$ phi(calc)] will be in poor agreement with the calculated structure factors used in the initial model [$\left|F_{calc}\right|$ phi(calc)]. Crystallographic refinement aims to find a set of atomic positions for the protein model that give $\left|F_{calc}\right|$ values that are as close as possible to the $\left|F_{obs}\right|$ values. This is achieved by the adjustment of the atomic positions within the model. REFMAC5 (Murshudov *et al*., 1997) was used for refinement in this Ph.D project and uses the amplitude based maximum likelihood method of refinement. The crystallographic *R* factor is a measure of how well your model matches your data although this is susceptible to manipulation and is therefore not a totally reliable guide to accuracy of the model (Tickle *et al*., 1998). In order to

combat this, the Free *R* factor (R_{free}) was introduced by Brunger (Brunger, 1992) and is the same as the conventional *R* factor but contains between 5-10% of random reflections from the dataset which are excluded from the structure refinement. This has the advantage of measuring the degree to which $|F_{calc}|$ predicts $|F_{obs}|$, without bias from the model, and therefore tells you whether the model has been made better or worse from a paticular manipulation. It is a much more reliable test in assessing the quality of the model. R_{free} will always be higher than the *R* factor and both will drop during the course of a successful refinement. For the majority of refined protein structures the R_{free} value for the final model will be between 20-30%. Final models were produced with iterative cycles of refinement in REFMAC5 and manual rebuilding in COOT (Emsley and Cowtan, 2004). Structures were validated using the validation tools in COOT and finally using the MOLPROBITY server (Davis *et al*., 2007). All crystallography images displayed in this thesis were produced using PyMOL (The PyMOL Molecular Graphics System, Version 0.99rc6, Schrödinger, LLC.).

2.12 Human tissue culture

Unless specified, media and supplements were purchased from Sigma-Aldrich (UK).

HaCaT cells were a gift from C. Todd (Dermatology Sciences, Newcastle University). Cells were handled in a class 2 Microbiological Safety Cabinet (MSC), taking routine precautions to minimise the risk of contamination. All experiments were carried out under containment level II conditions. HaCaT cells are derived from human keratinocytes and have been reported to exhibit most of the characteristics of the cells from which they were derived (Boukamp *et al*., 1988). In previous experiments done by colleagues in the laboratory a number of human tissues and immortalized cell lines had been examined for the binding of M1 GAS strain SF370. It was found that adhesion assays with tonsil explants, primary keratinocytes and HaCaT cells all gave essentially the same results. For this reason, HaCaTs were selected as the model cell line used in this study, as they give relevant physiological results and are easy to maintain in the laboratory.

2.12.1 Passaging HaCaT cells

HaCaT cells were cultured in Dulbecco's Modified Eagle's Medium, supplemented with 10% (v/v) fetal calf serum, 110 μg/ml sodium pyruvate, 100 U/ml penicillin and 100 μg/ml streptomycin (complete HaCaT media). The cells were incubated at 37°C in an atmosphere of 5% CO_2 and grown to confluency in a T25 cm^2 flask. At this point cells were washed 3 times in 10 ml of phosphate buffered saline (PBS). Residual PBS from the final wash was removed prior to detachment of cells from the bottom of the flask, achieved with 1 ml 0.05% (w/v) Trypsin/0.02% (w/v) EDTA incubated at 37% with 5% CO_2 until gaps were seen appearing between cells (usually 10 min). As soon as gaps became apparent, 9 ml of complete HaCaT media was added and a sterile needle and syringe was used to flush attached cells from the bottom of the flask. Cell numbers were assessed by counting cells present in 0.5 ml cell suspension in 20 ml isotone on a Beckman Coulter Z1 dual particle counter (Beckman Coulter, CA, USA). Cells were diluted to 0.5 x 10^6 cells in 10 ml of complete media, added to a new sterile T25 cm^2 flask, and incubated at 37°C with 5% CO_2 until confluency was reached and cells could be passaged again.

2.12.2 Infection of HaCaT cells - adhesion assay

For adhesion assays 1 x 10^5 HaCaT cells were seeded onto 13 mm-diameter glass coverslips in 12 well tissue culture plates and incubated at 37°C with 5% CO_2 until confluent. Cells were washed three times in PBS and 1 ml complete HaCaT media, *minus* antibiotics (assay media) was added to each well, and incubation was continued for 3 h at 37°C with 5% CO_2 prior to infection. In parallel, overnight GAS cultures were used to inoculate 10 ml fresh THY media (1/20) and incubated at 37°C with 5% CO_2 until cultures reached mid log phase of growth (OD_{600}-0.5). For infection, 1 ml of GAS culture was pelleted at 16,000 x g, washed twice in double processed tissue culture water, and resuspended in 1 ml of assay media.

Immediately prior to infection, assay medium was removed from each well of the HaCaT tissue culture plate. Cells were washed once with PBS before 0.5 ml of GAS suspension was added and the plate was then incubated at 37°C with 5% CO_2 for 20 min. Unbound GAS were removed by washing each well three times with PBS before

cells were fixed with 2% (w/v) paraformaldehyde on ice for 15 min. Each well was washed three times with PBS to remove residual fixative before staining.

Fixed cells were permeabilised with the addition of 0.1% (v/v) Trition-X 100 to each well and the plate was incubated at room temperature for 15 min with gentle shaking. Triton solution was removed by washing each well three times with PBS and cells were then 'blocked' for 60 min with 10% (v/v) non-immune rabbit serum in PBS at room temperature. Blocking solution was replaced with 0.5 ml GAS-FITC antibody (5 µg/ml, Acris Antibodies, Herford, Germany) and phalloidin-TRITC (50 µg/ml) in 10% (v/v) rabbit serum in PBS, and incubation was continued in the dark, at room temperature for 60 min with gentle shaking. This staining solution was removed and wells were washed three times with PBS before coverslips were allowed to air dry. On glass slides, dry coverslips were placed face down onto a drop of vectashield, to preserve fluorescence (Vector Laboratories, Peterborough, UK), and covered with foil to minimise exposure to light.

GAS adhesion to HaCaT cells was visualised under a Leica laser scanning confocal microscope (Leica Microsystems (UK) Ltd., Bucks, UK) using a FITC/TRITC glowover and UV 20x 1.4 NA lens, with zoom and pinhole set to 1. Four repeat scans were used to produce a final image with all PMT values recorded for each experiment.

2.12.3 Quantifying GAS adhesion

The quantification of GAS binding to HaCaT cells was achieved using Qwin software (Leica, Bucks, UK) by calculating pixel intensity of top-cell (*x*-*y*) sections. Pixel intensity for the total bacterial component was divided by the calculated pixel intensity for a single bacterium to give the total number of adherent bacteria. Each experiment was repeated on at least three separate occasions, always including the parent wild-type strain as a positive control (100% adhesion). The significance of apparent differences between the wild-type and mutant strains were calculated using the Students t-test, as described in Abbot *et al*., 2007.

Chapter 3. Characterisation of Spy0130 and Spy0128

3.1 Introduction

At the beginning of this project it was known that pili produced by M1 GAS strain SF370 were composed of one major subunit (Spy0128) that forms the shaft, and two minor subunits (Spy0125 and Spy0130) of unknown function. During the first year of this project Manetti *et al*. (Manetti *et al*., 2007) reported that both minor subunits may be acting as pilus adhesins. However, studies by workers in the Kehoe laboratory revealed this was not the case. Briefly, these studies showed that (a) only anti-recombinant (r)Spy0125 sera (and not anti-rSpy0128 or anti-rSpy0130) inhibited adhesion of wild-type M1 GAS strain SF370 to human keratinocytes or tonsil epithelium, which suggests that Spy0128 and Spy0130 do not contribute directly to adhesion; (b) in a *Δspy0125* mutant Spy0128 was assembled into pili that were found mostly in the cell wall (analogous to wild-type) but these pili were non-adhesive; (c) polymerisation of Spy0128 into pili required the action of Spy0129 (the pilin polymerase), and occurred in a *ΔsrtA* mutant but these pili were found mostly in culture supernatants indicating that SrtA was responsible for pilus attachment to the cell wall; (d) in a *Δspy0130* mutant, Spy0128 was assembled into pili but these pili were found mostly in culture supernatants (similar to the *ΔsrtA* mutant) and also appeared longer than their wild-type counterparts suggesting a role for Spy0130 in anchoring the pilus to the cell wall and regulating its length. These data considered together suggest that Spy0130 acts as a 'linker' protein incorporated as the final subunit at the base of the pilus and covalently attached to the lipid II precursor in the cell wall by the housekeeping sortase SrtA (Smith *et al*., 2010). It is now clear that each of the three subunits play distinct but equally essential roles in forming intact, functional pili.

Much of the work presented in this thesis focused on Spy0125 and this is described in detail in chapters 4 and 5. The work in this chapter describes studies on Spy0130 and Spy0128, although these studies were limited for the reasons outlined below.

3.2 Recombinant protein expression and purification

At the beginning of the work described in this chapter, recombinant plasmids consisting of the *E. coli* pET-28a(+) expression vector containing cloned *spy0130* or *spy0128* sequences had been constructed by Dr Wendy Smith and were available in the laboratory. These plasmids, pWS010 and pWS009, had been designed to express the fully processed versions of Spy0130 and Spy0128. This involved omitting sequence for the N-terminal signal peptide, as determined by the SignalP 3.0 server (Emanuelsson *et al*., 2007), and the sequence downstream of the cell wall sorting signal after the glycine of the LPST(G) motif for Spy0130 and the valine of the EVPTG(V) motif for Spy0128 as it was predicted that these sequences would be absent from the mature form of the protein on the GAS cell surface. The final *spy0130* and *spy0128* sequences cloned into pET-28a(+) are shown in panel A of Figures 3.1 and 3.2 with their corresponding amino acid sequences shown in panel B. After examining a range of different strains and induction conditions it was decided to use *E. coli* BL21 (DE3) CodonPlus type RIL for the expression of both proteins, induced at an OD_{600} of 0.6 with 1 mM IPTG for 3 h.

Optimised growth conditions determined during small scale expression tests were applied to large scale inductions of both recombinant proteins. Cells were harvested and lysed by sonication before cell lysate was applied to a His-tag column. Eluted fractions which corresponded to the protein peak on the elution profile at 280 nm were analysed by SDS-PAGE and fractions which were deemed to contain a significant amount of protein were pooled and concentrated. However, this material was observed to contain contaminating host proteins (Figure 3.3). Therefore, concentrated fractions were purified further on a Hi-Load 16/60 Superdex 75 gel filtration column. Panels A and B, Figure 3.4 show recombinant (r)Spy0130 and rSpy0128 gel filtration elution profiles at 280 nm, revealing a single sharp peak for both. Fractions corresponding to this peak were analysed by SDS-PAGE and are shown in Figure 3.4 panels C (rSpy0130) and D (rSpy0128), as highly pure samples. Fractions corresponding to lanes 4-7 for rSpy0130 (panel C, Figure 3.4) and lanes 3-8 for rSpy0128 (panel D, Figure 3.4) were pooled and concentrated to 10 mg/ml. This was determined by absorbance at 280 nm using the molar extinction coefficient of 11920 M^{-1} cm^{-1} for Spy0130 and 26360 M^{-1} cm^{-1} for Spy0128 as determined from the amino acid sequences by the ProtParam tool

A

T7 Promoter → lac operator

TAATACGACTCACTATAGG**GGAATTGTGAGCGGATAACAATTC**CCCTCTAGAAATAATTTTGTTTAACT

rbs His – tag Thrombin site

TTAAG**AAGGAG**ATATACCATGGGCAGCAGC**CATCATCATCATCATCAC**AGCAGCGGC**CTGGTGCCGCGC**

GGCAGCCATATGGAGAACCTCACTGCAAGCATTAATATTGAAGTTATTAATCAAGTTGATGTTGCTACCA
ACAAACAATCCTCTGACATAGATGAAACGTTTATGTTTGTAATTGAAGCGTTAGATAAAGAAAGTCCTTT
ACCTAATTCGGTAACTACTTCTGTTAAGGGAAATGGGAAAACATCATTTGAACAGTTGACTTTTTCAGAA
GTTGGACAATATCATTATAAAATTCATCAACTGTTAGGTAAGAATAGTCAATATCATTATGATGAAACAG
TATATGAAGTTGTTATTTACGTATTGTATAATGAGCAAAGTGGTGCTTTAGAAACTAACCTAGTTTCTAA
CAAACTTGGAGAAACCGAAAAATCGGAGCTTATTTTTAAACAAGAATATAGTGAAAAAACACCGGAACCT
CATCAACCAGATACAACTGAGAAGGAAAAACCTCAGAAAAAACGTAATGGTATTTTGCCATCAACTGGTT
AACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAA

B

GSHMENLTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNGKTSFEQLTFSE
VGQYHYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPEP
HQPDTTEKEKPQKKRNGILPSTG

Figure 3.1: Recombinant plasmid pWS010.

Panel A: The *spy0130* DNA sequences cloned in pWS010 are shown in blue with primer encoded stop codon at the end of the sequence underlined. Vector sequences are shown in black, with important features underlined in bold type and indicated above the sequence. *Nde*I (forward primer) and *Xho*I (reverse primer) restriction endonuclease sites used to clone into the vector are shown in red type adjacent to the cloned DNA. The *spy0130* DNA fragment was cloned in-frame with the vector encoded His-tag creating a fusion protein. **Panel B**: The rSpy0130 amino acid sequence. The Spy0130 sequences are shown in blue, with vector encoded residues remaining at the N-terminus after thrombin cleavage removing the His-tag shown in black.

A

T7 Promoter → lac operator

TAATACGACTCACTATAGG**GGAATTGTGAGCGGATAACAATTC**CCCTCTAGAAATAATTTTGTTTAACT

rbs His-tag Thrombin site

TTAAG**AAGGAG**ATATACCATGGGCAGCAGC**CATCATCATCATCATCAC**AGCAGCGGC**CTGGTGCCGCGC**

GGCAGCCATATGGAGACTGTTGTAAACGGAGCCAAACTAACAGTTACAAAAAACCTTGATTTAGTTAATA
GCAATGCATTAATTCCAAATACAGATTTTACATTTAAAATCGAACCTGATACTACTGTCAACGAAGACGG
AAATAAGTTTAAAGGTGTAGCTTTGAACACACCGATGACTAAAGTCACTTACACCAATTCAGATAAAGGT
GGATCAAATACGAAAACTGCAGAATTTGATTTTTCAGAAGTTACTTTTGAAAAACCAGGTGTTTATTATT
ACAAAGTAACTGAGGAGAAGATAGATAAAGTTCCTGGTGTTTCTTATGATACAACATCTTACACTGTTCA
AGTTCATGTCTTGTGGAATGAAGAGCAACAAAAACCAGTAGCTACTTATATTGTTGGTTATAAAGAAGGT
AGTAAGGTGCCAATTCAGTTCAAAAATAGCTTAGATTCTACTACATTAACGGTGAAGAAAAAAGTTTCAG
GTACCGGTGGAGATCGCTCTAAAGATTTTAATTTTGGTCTGACTTTAAAAGCAAATCAGTATTATAAGGC
GTCAGAAAAAGTCATGATTGAGAAGACAACTAAAGGTGGTCAAGCTCCTGTTCAAACAGAGGCTAGTATA
GATCAACTCTATCATTTTACCTTGAAAGATGGTGAATCAATCAAAGTCACAAATCTTCCAGTAGGTGTGG
ATTATGTTGTCACTGAAGACGATTACAAATCAGAAAAATATACAACCAACGTGGAAGTTAGTCCTCAAGA
TGGAGCTGTAAAAAATATCGCAGGTAATTCAACTGAACAAGAGACATCTACTGATAAAGATATGACCATT
ACTTTTACAAATAAAAAAGACTTTGAAGTGCCAACAGGAGTATAA**CTCGAG**CACCACCACCACCACCACT
GAGATCCGGCTGCTAA

B

GSHMETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTTVNEDGNKFKGVALNTPMTKVTYTNSDKG
GSNTKTAEFDFSEVTFEKPGVYYYKVTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQKPVATYIVGYKEG
SKVPIQFKNSLDSTTLTVKKKVSGTGGDRSKDFNFGLTLKANQYYKASEKVMIEKTTKGGQAPVQTEASI
DQLYHFTLKDGESIKVTNLPVGVDYVVTEDDYKSEKYTTNVEVSPQDGAVKNIAGNSTEQETSTDKDMTI
TFTNKKDFEVPTGV

Figure 3.2: Recombinant plasmid pWS009.

Panel A: The *spy0128* DNA sequences cloned in pWS009 are shown in blue with primer encoded stop codon at the end of the sequence underlined. Vector sequences are shown in black, with important features underlined in bold type and indicated above the sequence. *Nde*I (forward primer) and *Xho*I (reverse primer) restriction endonuclease sites used to clone into the vector are shown in red type adjacent to the cloned DNA. The *spy0128* DNA fragment was cloned in-frame with the vector encoded His-tag creating a fusion protein. **Panel B**: The rSpy0128 amino acid sequence. The Spy0128 sequences are shown in blue, with vector encoded residues remaining at the N-terminus after thrombin cleavage removing the His-tag shown in black.

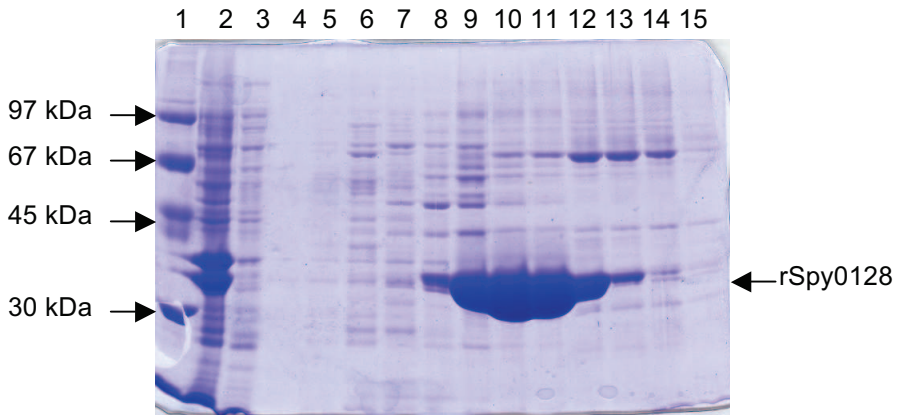

Figure 3.3: Expression and purification of rSpy0128 by affinity chromatography.

Coomassie-blue stained SDS-PAGE gel showing rSpy0128. Lane 1: molecular size standards. Lane 2: sample of induced cell lysate. Lane 3: sample of unbound material from His-tag column. Lanes 4-15: fractions collected during rSpy0128 elution. Fractions corresponding to the absorbance peak of the elution profile, lanes 9-14, were pooled. Fractions eluted during rSpy0130 affinity chromatography were also observed to contain contaminating host proteins (data not shown).

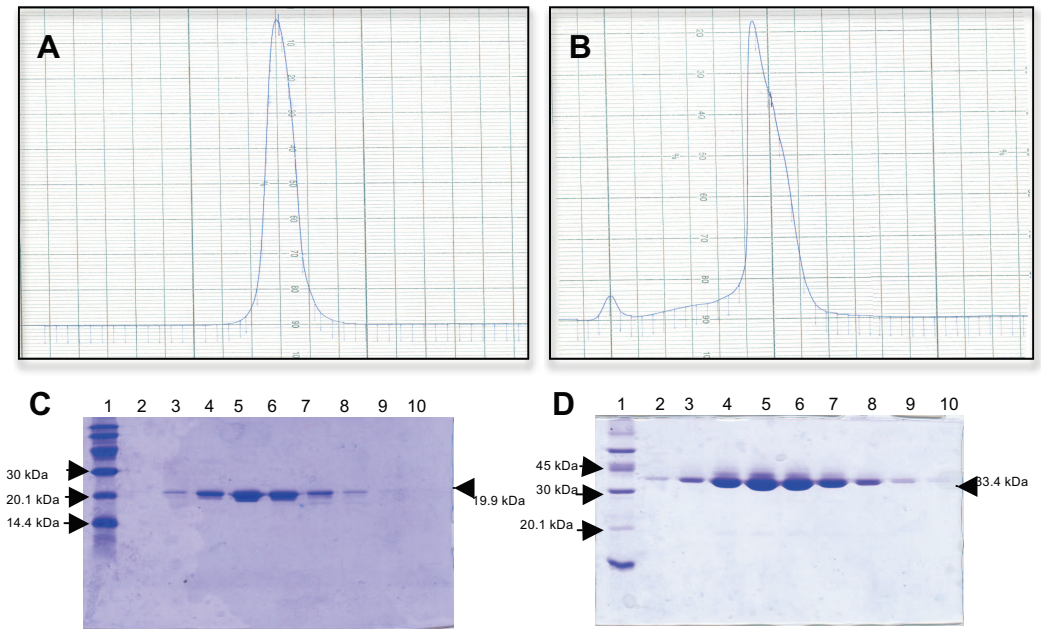

Figure 3.4: Elution profiles for rSpy0130 and rSpy0128 S75 gel filtration and analysis of peak fractions on SDS-PAGE.

Panel A: rSpy0130 gel filtration elution profile at 280 nm. **Panel B**: rSpy0128 gel filtration elution profile at 280 nm. **Panel C**: coomassie blue stained SDS-PAGE gel analysing rSpy0130 peak fractions. **Panel D**: coomassie blue stained SDS-PAGE gel analysing rSpy0128 peak fractions. For both C and D; Lane 1: molecular size standards. Lanes 2-10: samples of fractions corresponding to the major peak from the elution profile.

(Gasteiger *et al*., 2005) on the ExPASy server (www.expasy.ch). When needed, the N-terminal His-tag was removed by digestion with thrombin. After removal of cleaved His-tag (on a His-chelation column) and thrombin (on a benzamidine column) as described in chapter 2, section 2.6.7, rSpy0130 and rSpy0128 were pooled, dialysed into gel filtration buffer and concentrated.

3.3 Biophysical characterisation of rSpy0130 and rSpy0128 in solution

Purified rSpy0130 and rSpy0128 from section 3.2 (with the N-terminal His-tag removed) were used in biophysical characterisation experiments. A number of biophysical techniques were used to probe the physical structure of rSpy0130 and rSpy0128: (i) The secondary structure for both was examined using circular dichroism spectroscopy (CD) (Bulheller *et al*., 2007; Kelly *et al*., 2005); (ii) The heterogeneity and oligomeric state of the proteins in solution were investigated by analytical ultracentrifugation (AUC) using the sedimentation velocity technique (Lebowitz *et al*., 2002); and (iii) Small angle X-ray scattering (SAXS) was used to provide information on the size and shape of the molecules in solution (Neylon, 2008; Petoukhov and Svergun, 2007). As mentioned in the materials and methods, AUC provides information which is highly complementary to SAXS, such as data regarding the monodispersity, elongation and aggregation of the molecule in solution (Hura *et al*., 2009; Neylon, 2008).

Both AUC and SAXS experiments were carried out in collaboration with Dr. Alexandra Solovyova who provides these techniques as a service for our institute.

3.3.1 Secondary structure of rSpy0130

CD was used to probe the rSpy0130 secondary structure. For CD experiments rSpy0130 was diluted to 0.5 mg/ml. Five repeat scans were used to produce the final spectra which were corrected for background “noise” from the sample buffer by subtracting the protein free spectra. The far UV spectra for rSpy0130 is shown as the red graph in Figure 3.5 and reveals it to contain very little identifiable secondary structure. However, features characteristic of internal disorder such as a negative shoulder near 222 nm or minimum at 203 nm are not observed (Receveur-Brechot *et al*., 2006).

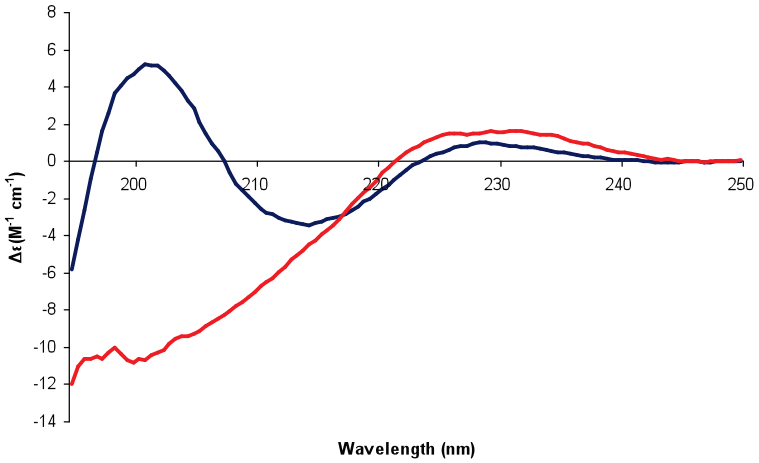

Figure 3.5: Far UV spectra for rSpy0130 and rSpy0128.

rSpy0130 (red line) exhibits little identifiable secondary structure. rSpy0128 (blue line) displays a classical β_I-type sheet spectrum.

Although the spectrum is unusual, a lack of these features combined with the data from the sedimentation velocity and SAXS experiments (see below) suggests that rSpy0130 is not completely unfolded in solution but may exhibit a much elongated structure.

3.3.2 Secondary structure of rSpy0128

rSpy0128 samples were diluted to 0.5 mg/ml for use in CD experiments and the spectra was corrected as described above. The far UV spectrum for rSpy0128 is shown as the blue graph in Figure 3.5 and displays a classical β_I-type sheet spectrum, with a single minimum at 215 nm and a maximum at 202 nm (Solovyova *et al.*, 2010).

3.3.3 Analytical ultracentrifugation of rSpy0130

Concentrations of rSpy0130 used ranged from 0.32-2.04 mg/ml in gel filtration buffer. The continuous c(s) distribution model shown in panel A of Figure 3.6 reveals a single dominant species observed as a single peak at 1.76 S, indicating that rSpy0130 is monodisperse and monomeric in the sample. The calculated sedimentation coefficient for a typically hydrated sphere of equivalent molecular mass to rSpy0130 of 2.32 S is significantly larger than that determined experimentally. This suggests a significant elongation of rSpy0130 in solution. This is supported by a much increased f/f_0^{shape} of

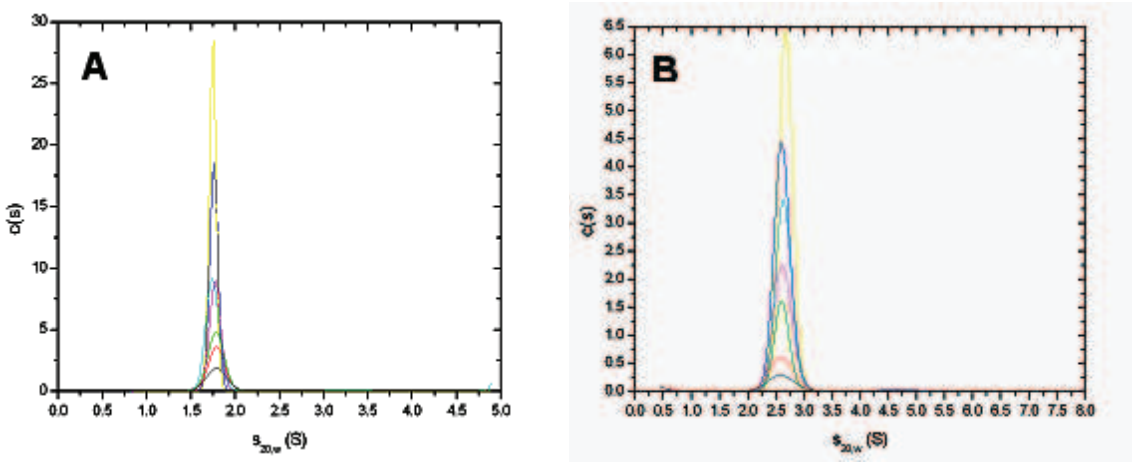

Figure 3.6: Size distribution c(*s*) for rSpy0130 and rSpy0128.

Panel A: A single dominant peak is observed for rSpy0130 at 1.76 S using the general c(*s*) size distribution model. The concentrations of protein examined were: 0.32 mg/ml (black line), 0.61 mg/ml (red line), 0.91 mg/ml (green line), 1.23 mg/ml (magenta line), 1.57 mg/ml (cyan line), 2.04 mg/ml (blue line). **Panel B**: A single dominant peak is observed for rSpy0128 at 2.63 S using the general c(*s*) size distribution model. The concentrations of protein examined were: 0.15 mg/ml (black line), 0.26 mg/ml (red line), 0.43 mg/ml (green line), 0.72 mg/ml (magenta line), 1.01 mg/ml (cyan line), 1.34 mg/ml (blue line) and 1.65 mg/ml (yellow line).

Table 3.1: rSpy0128 and rSpy0130 sedimentation velocity results.

	Calculated parameters		SV experimental data			
	Mass (kDa)	*s* (S)	RMSD[a]	f/f_0^{shape}	$S_{20,w}$	Mass (kDa)
Spy0128	33.9	2.54	0.0109	1.23	2.63 ± 0.019	34.48 ± 1.35
Spy0130	19.9	2.32	0.0121	1.45	1.76 ± 0.003	20.47 ± 0.47

[a] The root mean square deviation (RMSD) of the fit of the data for the highest sample concentration is shown.

1.45 when compared to f/f_0 = 1 for a sphere (Table 3.1). From the sedimentation velocity data it is possible to calculate an accurate mass for the protein present in the sample. The experimentally determined mass of 20.47 kDa reveals that rSpy0130 is monomeric in solution when compared to the calculated mass of 19.9 kDa for the rSpy0130 construct (Solovyova *et al*., 2010).

3.3.4 Analytical ultracentrifugation of rSpy0128

Concentrations of rSpy0128 used in sedimentation velocity experiments ranged from 0.15-1.65 mg/ml in gel filtration buffer. A sharp, single peak for rSpy0128 is revealed by the continuous c(s) distribution model in panel B of Figure 3.6. The derived value of 1.23 for f/f_0^{shape} is significantly larger than 1 (f/f_0 = 1 for a sphere) which implies elongation of the protein (Table 3.1). An experimentally derived mass of 34.5 kDa was obtained for rSpy0128 from the sedimentation velocity data. This is consistent with the expected mass of the expressed Spy0128 construct of 33.9 kDa and reveals that rSpy0128 is monodisperse and monomeric in solution (Solovyova *et al*., 2010).

3.3.5 rSpy0130 solution structure modelled *ab initio*

Solution X-ray scattering experiments were carried out on rSpy0130 that had been confirmed by AUC as monomeric and monodisperse in solution at 5 mg/ml and 10 mg/ml in gel filtration buffer. The SAXS data for rSpy0130 was collected on beamline X33 at the EMBL Hamburg outstation and the low resolution solution structure was determined by Dr Alexandra Solovyova by averaging 20 *ab initio* restored models, constructed as dummy atom models (DAM) using DAMMIN (Svergun, 1999), with a maximum resolution of 25 Å. The DAM for rSpy0130 is shown in panel A Figure 3.7 and was constructed with approximate dimensions of 105 Å x 25 Å x 14 Å, revealing an extremely elongated architecture but suggestive that rSpy0130 is not completely unfolded and displays some globular structure (Solovyova *et al*., 2010).

3.3.6 rSpy0128 solution structure modelled *ab initio*

The SAXS data for rSpy0128 was collected on beamline 2.1 at the SRS (Daresbury, UK). A DAM was constructed for rSpy0128 as described above and is shown in panel B Figure 3.7. An envelope of 66 Å x 37 Å x 24 Å was determined for rSpy0128 and its DAM is suggestive of a two domain structure.

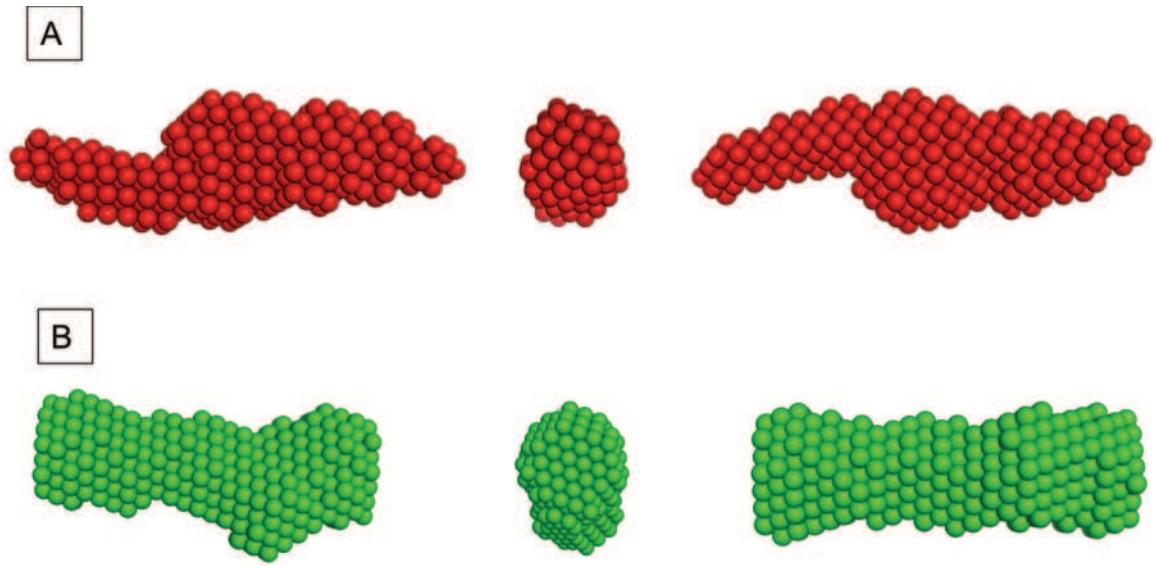

Figure 3.7: SAXS DAMs for rSpy0130 and rSpy0128 in three orientations.

Structures are not to scale. **Panel A**: rSpy0130. Approximate dimensions: 105 Å x 25 Å x 14 Å. **Panel B**: rSpy0128. Approximate dimensions: 66 Å x 37 Å x 24 Å.

3.4 Crystallisation of rSpy0130 and X-ray analysis of SeMet labelled rSpy0128

During the course of this project, attempts to crystallise rSpy0130 proved unsuccessful, potentially for a variety of reasons and these are outlined in the discussion to this chapter. In parallel and in collaboration with Dr Paul Race, who had collected a native rSpy0128 dataset (limited to ~3 Å), this project focussed on producing rSpy0128 selenomethionine (SeMet) crystals. SeMet labelled rSpy0128 was produced essentially as described in chapter 2, section 2.6.4. The rSpy0128 SeMet protein (10 mg/ml) was used to set up initial crystallisation trials using commercially available sparse matrix screens as described in materials and methods (section 2.11.1). Initial crystals were observed in ~70% of all conditions throughout the screens and almost exclusively in conditions harbouring PEG as a precipitant. The most promising crystals grew in 20% (w/v) PEG 6000 and 100 mM tri-sodium citrate pH 5.0. Crystal growth was optimised by screening the concentration of the precipitant against the pH of the buffer using the hanging drop vapour diffusion technique. Crystal growth was observed overnight with maximal dimensions achieved within one week. Large well formed crystals were obtained in 13-14% (w/v) PEG 6000 and 300 mM tri-sodium citrate pH 4.25.

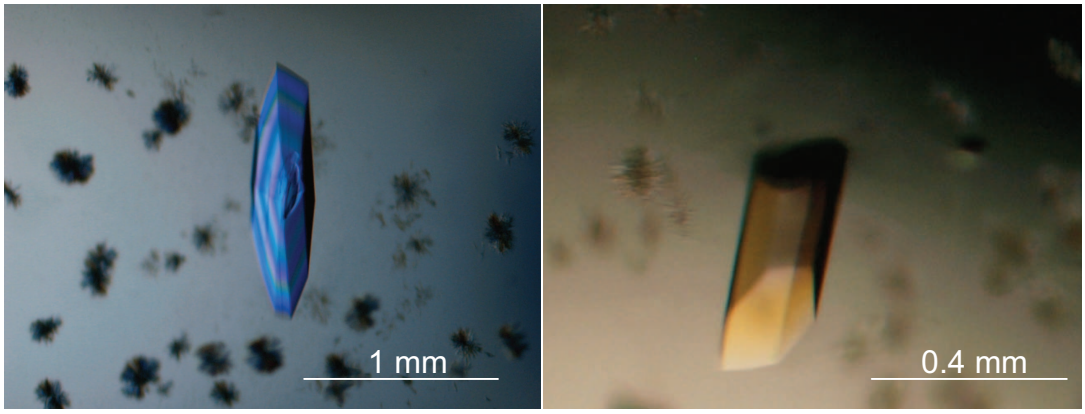

Figure 3.8: Examples of optimised rSpy0128 SeMet crystals.

Examples of these crystals are shown in Figure 3.8. Prior to data collection crystals were harvested in Litholoops, transferred to a cryo-protectant solution of paratone-N and flash frozen in liquid nitrogen. Although these crystals were extremely large and highly bi-refringent, their diffraction of X-rays was highly anisotropic and limited in resolution to ~3 Å. Multiple MAD datasets were collected on rSpy0128 SeMet crystals at the ESRF, Grenoble, France. All were low-resolution (3.5 Å at best), showed evidence of twinning and the structure could not be solved from these crystals. As attempts were being made to solve these problems, we learnt that the structure of Spy0128 had just been solved by a group in New Zealand (subsequently published by Kang *et al*. 2007), and therefore this aspect of the project was not pursued.

3.5 Construction of *spy0128 inter* and *intra* molecular isopeptide bond mutants in the M1 GAS strain SF370 chromosome

Analysis of the rSpy0128 crystal reported by Kang *et al*. (Kang *et al*., 2007), revealed it to be an elongated, two domain protein with individual subunits aligned end on end. This, in combination with mass spectrometry of intact pili identified Lys161 as joining to the Thr of the EVPTG motif (forming the *inter*-molecular bond) on adjacent subunits. In addition, Spy0128 was found to contain two *intra*-molecular isopeptide bonds with one in each domain; between Lys36 and Asn168 in the N-terminal domain and between Lys179 and Asn303 in the C-terminal domain. As explained in chapter 1, section 1.15, formation of these bonds is dependant on an appropriately positioned Glu residue

(Glu117 in the N-terminal domain and Glu258 in the C-terminal domain) (Kang *et al*., 2007). More recently, the importance of the above Glu residues was confirmed with the observation that mutation to alanine abrogates bond formation (Kang and Baker, 2009). These bonds had been confirmed *in vitro* to convey proteolytic and thermal stability upon Spy0128 (Kang and Baker, 2009) but their biological function and role *in vivo* had not been explored. Since this project focused on adhesion, following publication of the Spy0128 structure it was decided to examine the effects of *intra*-molecular bond disruption on GAS adhesion to HaCaT cells. The residues that were targeted for mutagenesis to alanine were; Lys36 and Glu117 from the first *intra*-molecular isopeptide bond, Lys179 and Glu258 from the second, and the *inter*-molecular linker, Lys161.

The mutant alleles were first constructed in the temperature sensitive plasmid pG$^+$host9. The two outer primers (shown as 1 and 2 in Figure 3.9) were used to amplify the entire wild-type *spy0128* including DNA flanking each side of the gene, which would be utilised later during allele replacement mutagenesis (see below). These primers were designed with an *EcoR*I restriction endonuclease site in the forward primer and an *Xho*I site in the reverse primer, to facilitate cloning into the pG$^+$host9 vector which had been cut with the same enzymes. Figure 3.9 shows a schematic diagram of this process. The structure of the resulting plasmid was confirmed by DNA sequencing and it was renamed pJP01. To create each point mutation in *spy0128*, pJP01 was used as template in five separate quickchange PCR mutagenesis reactions (chapter 2, section 2.5.9.1). Each reaction contained a pair of mutagenic primers (Table 2.3) which were designed with up to 18 bases either side of the altered bases to aid annealing to the template DNA and reduce the effects of the base mismatches. The structure of pJP01, highlighting the *spy0128* codons targeted for mutagenesis is shown in Figure 3.10 and each construct is described in Table 3.2. The structures of each of these recombinant plasmids were confirmed by DNA sequencing.

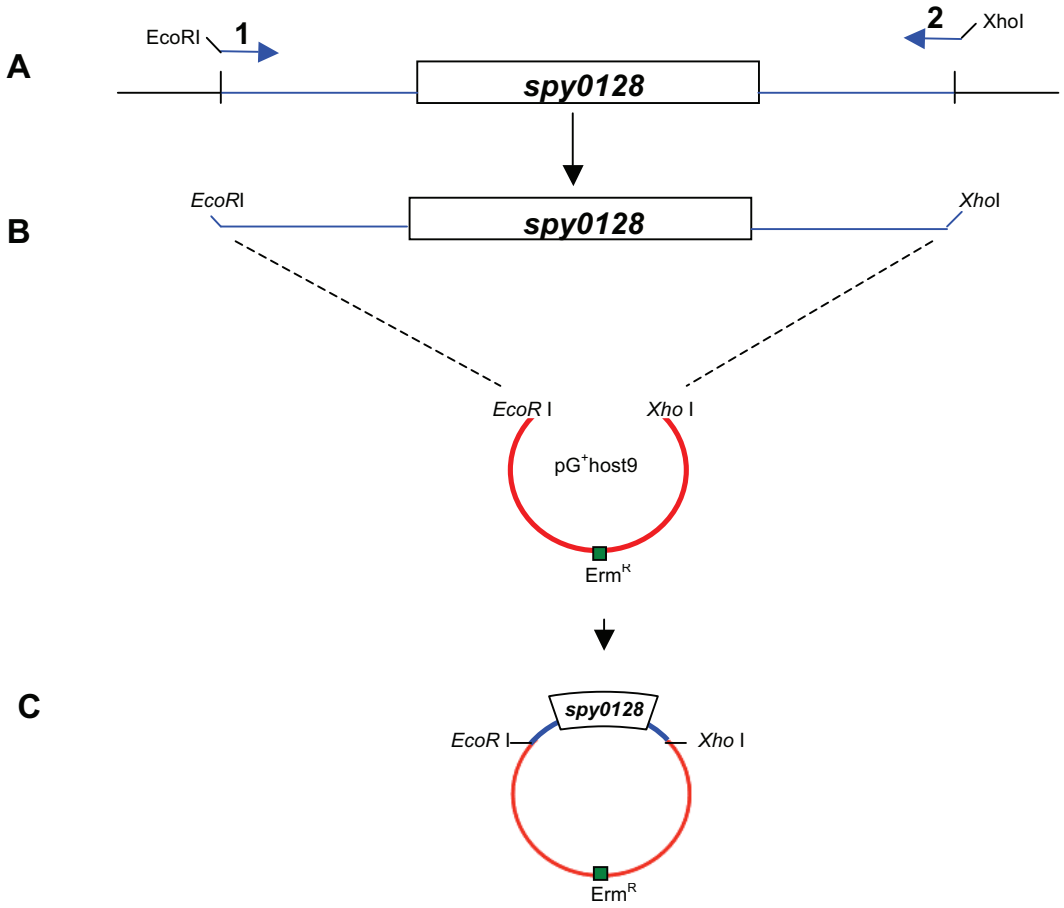

Figure 3.9: Schematic diagram showing construction of the recombinant plasmid pJP01.

Panel A: The primer set 1 + 2 were used to amplify wild-type *spy0128* including flanking regions of DNA directly adjacent to the gene. **Panel B**: The amplified fragment was cleaved with the restriction endonucleases *EcoR*I and *Xho*I. **Panel C**: This fragment was then ligated into the pG^{+}host9 vector, cleaved with the same enzymes to create the pJP01 construct.

Previously in the laboratory, an M1 GAS strain SF370 *spy0128* deletion mutant (*Δspy0128*) had been created, deleting the entire *spy0128* coding region from the chromosome (Abbot *et al*., 2007). Here, *Δspy0128* was used as the target for allele replacement mutagenesis experiments (Figure 2.4) which introduced the wild-type allele carried on plasmid pJP01 and mutant alleles carried on plasmids pJP02, pJP03, pJP04, pJP05 and pJP06 into the M1 GAS strain SF370 chromosome, as summarised in Figure 3.11.

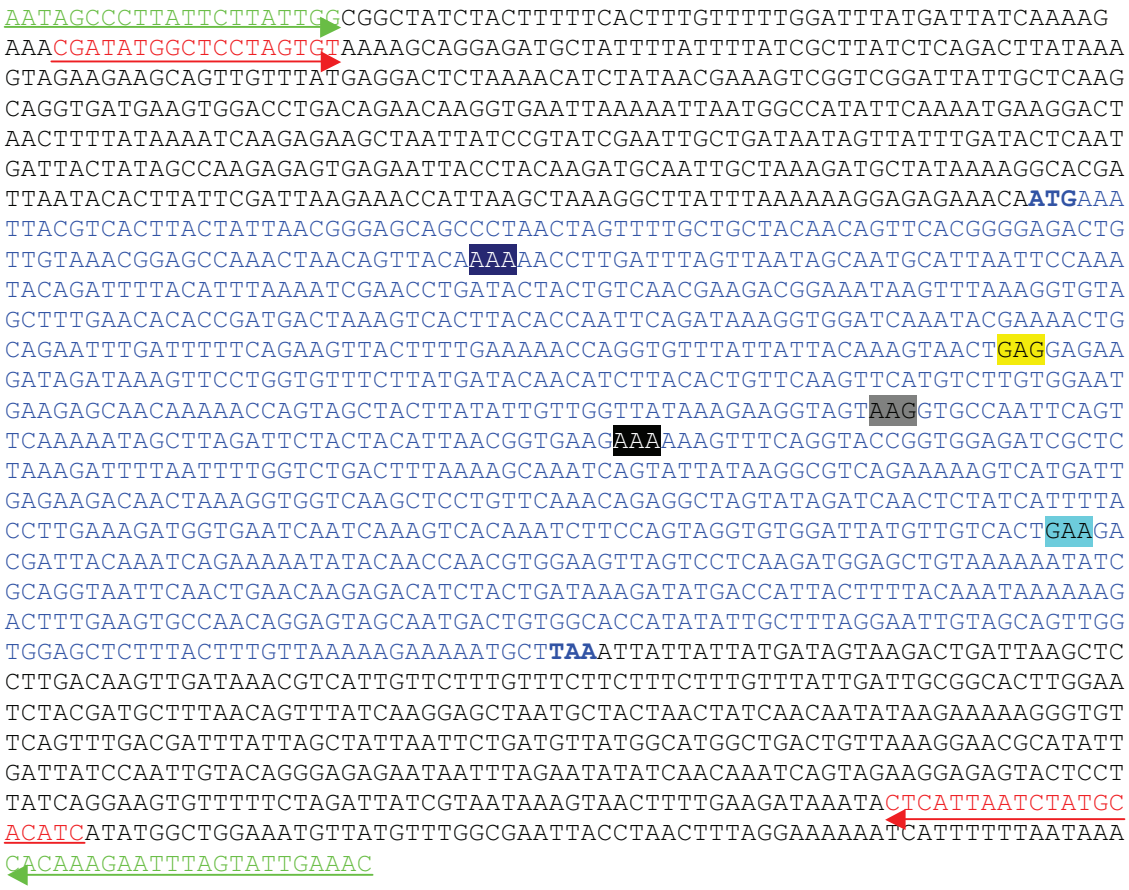

Figure 3.10: DNA sequence for wild-type *spy0128* used to construct the recombinant plasmid pJP01 including bases targeted for mutagenesis.

Coding sequences for *spy0128* are shown in blue with start and stop codons highlighted in bold. Also highlighted within the *spy0128* coding sequence are codons targeted for mutagenesis to alanine. Bases coding for isopeptide bond residues are highlighted as follows; Lys36 in dark blue, Glu117 in yellow, Lys161 in grey, Lys179 in black and Glu258 in sky blue. Adjacent regions of flanking DNA used to replace *spy0128* back into the chromosome are shown in black. Primers used to amplify the gene and clone into pG$^+$host9 are shown in red. Sequencing primers used to confirm the chromosome structure after allele replacement are shown in green.

Table 3.2: Description of recombinant plasmids used in allele replacement to produce Spy0128 isopeptide bond mutants in the M1 GAS strain SF370 chromosome.

Plasmid name	Description of contents within pG$^+$host9
pJP01	Contains wild-type *spy0128* allele
pJP02	Contains *spy0128* allele harbouring K36A point mutation
pJP03	Contains *spy0128* allele harbouring E117A point mutation
pJP04	Contains *spy0128* allele harbouring K161A point mutation
pJP05	Contains *spy0128* allele harbouring K179A point mutation
pJP06	Contains *spy0128* allele harbouring K258A point mutation

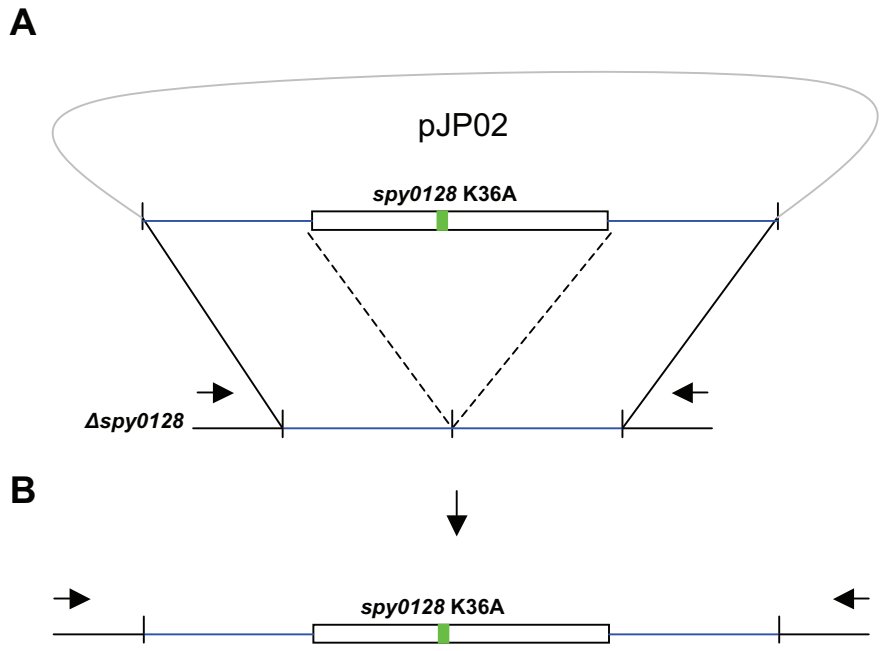

Figure 3.11 Schematic diagram showing construction of the *spy0128* K36A mutant in the M1 GAS strain SF370 chromosome.

Panel A: Previously, a *Δspy0128* strain had been constructed in the laboratory. **Panel B**: the pG^{+}host9 construct containing the *spy0128* K36A mutant allele was transformed into the *Δspy0128* strain and using allele replacement the *spy0128* gene containing the K36A mutation was replaced in the chromosome.

As a control, the wild-type *spy0128* allele (carried on pJP01) was put back into the chromosome of the *Δspy0128* strain. Each of the plasmids in Table 3.2 were introduced into an *Δspy0128* strain by electrotransformation. Allele replacement was used to insert the wild-type and mutagenised *spy0128* alleles into the chromosome as described in chapter 2, section 2.5.14 and shown in the schematic diagram, Figure 3.11. As explained in chapter 2, after the second recombination event involved in the allele replacement procedure, the target DNA is either replaced in the chromosome by the desired allele or the original structure is restored, depending upon which sequences have recombined during excision of the plasmid (chapter 2, Figure 2.3, panel D). To distinguish between the two possible scenarios for each mutant, the chromosomal DNA of 10-12 erythromycin sensitive colonies for each potential mutant was recovered and used as template in a PCR reaction using the sequencing primers highlighted in black in Figure 3.11. The products were examined by agarose gel electrophoresis. The two possible outcomes are readily identifiable, as seen in the example shown in Figure 3.12.

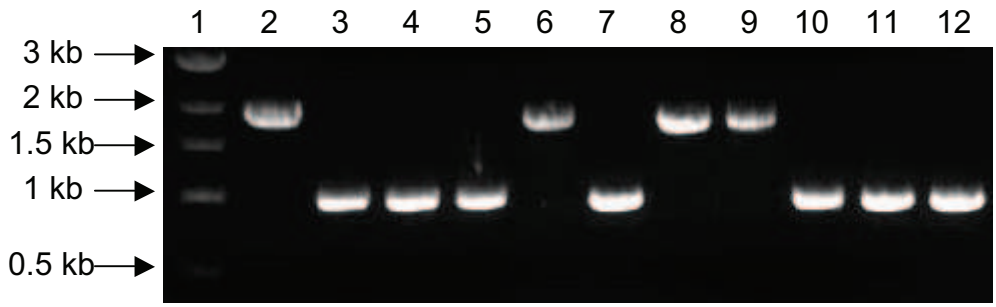

Figure 3.12: Example of PCR screening of Erm^s colonies containing mutagenised *spy0128* alleles.

The result shown is for allele replacement with pJP02 (K36A allele). All other mutants were screened the same way, with similar results. Lane 1: DNA size standards. Lane 2: positive control using wild-type M1 GAS strain SF370 chromosome as template. Lanes 3-12: PCR products from different Erm^s colonies.

The larger products (~1.9 kb) identify clones where the *spy0128* allele has been integrated back into the chromosome of the *Δspy0128* mutant, whereas the smaller products (~0.9 kb) are clones where the *Δspy0128* mutation was restored after the second cross over event. For each mutant, the PCR products of interest were purified and DNA sequenced to confirm the structure of the entire region, including the boundaries between sequences that had been manipulated *in vitro* and un-manipulated flanking regions. Growth rates were examined for all strains on three separate occasions with the same results, which identified no significant differences in growth rates between strains (data not shown).

3.6 Analysis of *inter* and *intra* molecular isopeptide bond mutants

To investigate the effects of isopeptide bond disruption on pili binding to host cells, adhesion to HaCaT cells was examined as outlined in chapter 2, section 2.12.2. Each experiment was repeated on three separate occasions with essentially the same results. An example of a confocal image for each control and mutant is seen in panel A, Figure 3.13. Control strains behaved as expected, with wild-type M1 GAS strain SF370 binding to HaCaT cells, while very few adhering bacteria (probably only remaining because washing steps were too gentle) were observed for the *Δspy0128* strain, consistent with previous observations (showing no binding) (Abbot *et al*., 2007). Panel A, Figure 3.13 also shows that the control strain (*Δspy0128+spy0128*), where wild-type

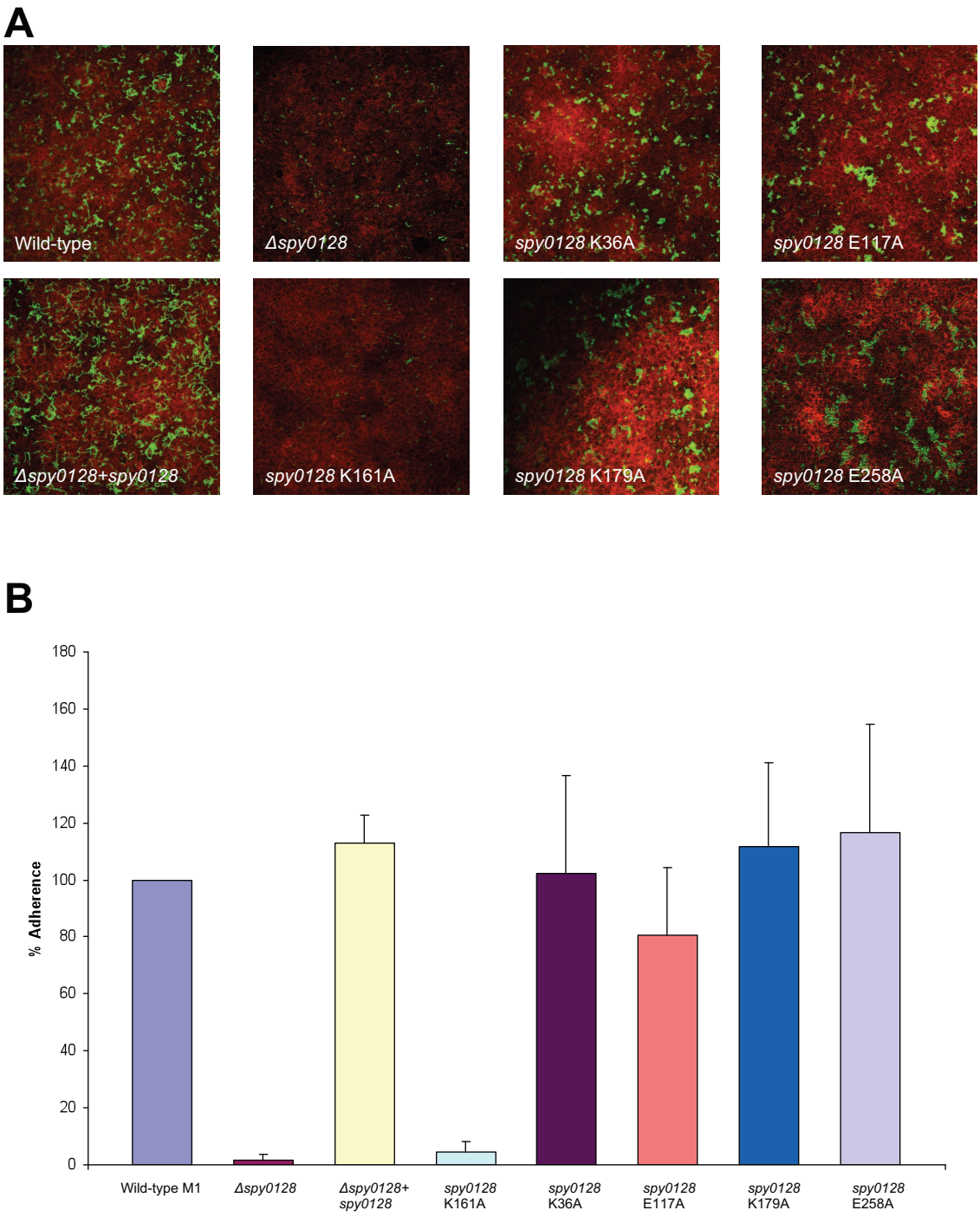

Figure 3.13: Adhesion of wild-type M1 and *spy0128* mutant strains of GAS to HaCaT cells.

Panel A: Example confocal images for wild-type GAS and Spy0128 isopeptide bond mutants adhesion to HaCaT cells at 20x magnification with strain shown indicated. HaCaT cells are labelled red and bacteria green. **Panel B**: Quantification of corresponding GAS strains from panel A. Relative adhesion is shown with wild-type binding considered 100% and error bars indicating variation from three independent experiments. Significant statistical differences compared to wild-type binding were only observed with the *Δspy0128* (P <0.0001) and *spy0128* K161A (P <0.0001) mutant strains.

spy0128 has been put back into the *Δspy0128* chromosome, is essentially the same as the wild-type. This was done in parallel to the *spy0128* mutants to show that wild-type pili function can be restored using allele replacement. Consequently, any effects observed in the mutant strains can be attributed to the point mutation they contain and not because of problems caused inserting *spy0128* back into the chromosome. As predicted, the *inter*-molecular bond mutant, *spy0128* K161A, displayed essentially the same results as *Δspy0128* where very few adherent bacteria were observed. A noticeable feature in all four *intra*-molecular isopeptide bond mutants is the bacteria appear to be 'clumped' together, rather than dispersed across the monolayer in short chains as seen with the wild-type. This was most pronounced in the two glutamate mutants but is observed in all cases.

To assess the contribution of isopeptide bonds in Spy0128 on the ability of M1 GAS strain SF370 to bind to HaCaT cells, the levels of binding for each mutant was quantified as described in chapter 2, section 2.12.3, and is shown in panel B, Figure 3.13. Relative adhesion is shown with wild-type binding considered 100%. The *Δspy0128* (P <0.0001) and *spy0128* K161A (P <0.0001) mutant strains both exhibit P-values which indicate significant differences when compared to the wild-type. It is quite clear from this figure that when the formation of these bonds is abrogated, the bacteria are still able to bind HaCaT cells as effectively as the wild-type. The largest difference in binding was observed between *spy0128* E117A and the wild-type but this was not significant (P <0.2868).

3.7 Effects of Spy0128 isopeptide bonds on pili polymerisation

In conjunction with the adhesion assays described above, western blots were utilised to examine the effects that isopeptide bond disruption might have on polymerisation of pili in these strains. Pili present in cell wall and supernatant fractions of parent and mutant strains (prepared as described in chapter 2, section 2.6.8) were run on a 4-15% gradient SDS-PAGE gel, before being immuno-blotted with anti-rSpy0128 antibody (Figure 3.14). The subunits of Gram positive pili are covalently linked and when M1 GAS strain SF370 pili are blotted with anti-rSpy0128 antibody they appear as characteristic high molecular weight polymer (HMW) 'ladders' on the blot. This characteristic 'laddering' pattern exhibited by pili present in the cell wall extract of wild-type M1

GAS strain SF370 is seen in lane 2 (W), Figure 3.14. The HMW laddering pattern exhibited by the *Δspy0128+spy0128* strain was identical to wild-type (data not shown). This data considered in conjunction with the results from the adhesion assay shows that wild-type function has been restored in the *Δspy0128+spy0128* strain. It is also common to see pili material in the supernatant fraction (S, lane 2, Figure 3.14) as intact pili are released during normal cell wall turnover. Not shown in this figure is the *Δspy0128* mutant as it is not detected by anti-rSpy0128 sera. The expected result can be seen for the *spy0128* K161A mutant in lane 3, Figure 3.14. Only Spy0128 monomer is observed in the cell wall fraction, as the K161A mutation has abrogated pili formation at the cell surface. It is quite clear from this figure that banding patterns exhibited by the two lysine point mutations in lanes 4 and 6 (K36A and K179A) are highly similar to those of the wild-type M1 GAS strain SF370, with most of the 'laddering' pattern observed in the cell wall extract with fewer pili present in the culture supernatants. The 'laddering' patterns observed for the two glutamate point mutations (E117A and E258A) shown in lanes 5 and 7 of Figure 3.14 show fewer HMW polymers in the cell wall fraction. Released into the culture supernatant, especially in the *spy0128* E258A strain, are much lower molecular weight polymers, which do not compare to wild-type at all. However, the ability of these two strains to produce functional pili is not affected when considered in conjunction with the binding data above. Much of the work in this project focussed on Spy0125 and so the reasons for this phenotype were not investigated further.

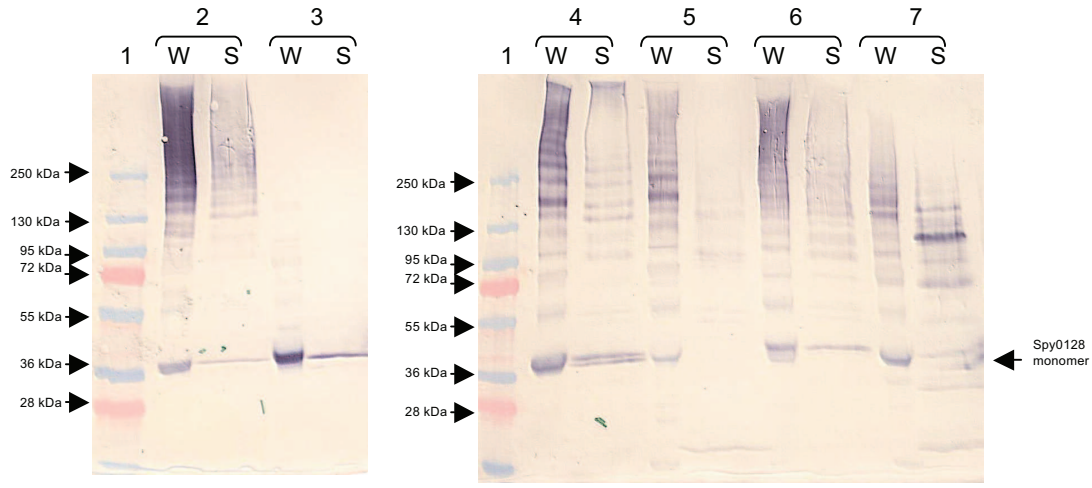

Figure 3.14: Western blot analysis of cell wall extract and supernatant from wild-type M1 GAS and Spy0128 isopeptide bond mutants using anti-rSpy0128 sera.

Lane 1: molecular size standards. **W:** enzymatically digested cell wall extract. **S:** culture supernatants. Lane 2: wild-type M1 GAS strain SF370. Lane 3: *spy0128* K161A. Lane 4: *spy0128* K36A. Lane 5: *spy0128* E117A. Lane 6: *spy0128* K179A. Lane 7: *spy0128* E258A. Also indicated is Spy0128 monomer at ~36 kDa. Only un-polymerised Spy0128 monomer is found in the cell wall fraction of the *spy0128* K161A mutant. Wild-type like 'laddering' patterns (lane 2) are observed for both lysine *intra*-molecular isopeptide bond mutants in lanes 4 and 6. 'Laddering' patterns for the two glutamate point mutations in lanes 5 and 7 reveal pili are much fewer in the cell wall component and much shorter in the supernatant when compared to the wild-type.

3.8 Discussion

Part of the work in this chapter describes the biophysical characterisation of rSpy0130 in solution. The Spy0130 recombinant protein was revealed as an extremely elongated molecule but not completely unfolded. Although the circular dichroism data suggested little evidence of secondary structure, features characteristic of an unfolded protein such as a negative shoulder near 222 nm or minimum at 203 nm were not observed. The data from the AUC and SAXS experiments also showed rSpy0130 to be elongated. Attempts during the project to crystallise rSpy0130 were unsuccessful and we recently learnt that another group had also attempted to crystallise rSpy0130 without success (E.N. Baker, personal communication). However, the Baker group were able to solve the crystal structure of a Spy0130 homologue (called FctB) from a different strain of GAS (Linke *et al*., 2010). Although FctB and Spy0130 share only 33% sequence homology, the dimensions of this recently published (while this thesis was being written) crystal structure are similar to those of the Spy0130 SAXS structure described here suggesting a similar overall shape (Linke *et al*., 2010). Spy0130 has a proline rich tail region and elongated shape which may allow it to tolerate higher tensile forces due to its flexibility (Solovyova *et al*., 2010). These characteristics probably explain why attempts to crystallise rSpy0130 were unsuccessful, as a flexible architecture would prevent intimate packing of molecules during crystal formation.

As mentioned in the introduction to this chapter, Spy0130 was confirmed as the wall 'linker'. It was predicted that as for linkages between other pilus subunits, the C-terminal end of Spy0128 is linked to the side chain of a specific surface exposed lysine residue of Spy0130. Following the unsuccessful crystallisation of rSpy0130, during the latter part of this project attempts were made to identify conserved lysines by aligning all sequenced homologues of Spy0130. Six conserved lysine residues were identified as potential targets. Site directed mutagenesis and allele replacement techniques were used to successfully construct all Lys-to-Ala mutants in the M1 GAS strain SF370 chromosome and preliminary experiments to identify the lysine were initiated. However, around this time we learnt about the FctB structure, and also that this group had identified the critical lysine (K146) in Spy0130 of M1 GAS strain SF370 by mass spectrometry (Linke *et al*., 2010). Because of the failure to crystallise rSpy0130, greater

emphasis was put into characterising Spy0125 (see chapters 4 and 5) and further structural studies on Spy0130 were not pursued.

The remainder of this chapter describes the structure of rSpy0128 in solution and provides an assessment of the roles played by the *inter* and *intra* molecular isopeptide bonds of Spy0128 on the function of pili. The data presented in this chapter concerning the characterisation of rSpy0128 in solution, are entirely consistent with the subsequently published crystal structure of Spy0128 (see below). This highlights the value and place for these low resolution techniques when reliable high resolution crystallographic data are not available. The circular dichroism spectroscopy data reported in this chapter suggested rSpy0128 was composed solely of β-sheet and this matched the high resolution crystal structure of Spy0128 that is composed of two discrete all β domains (Kang *et al*., 2007). Analytical ultracentrifugation experiments determined a sedimentation coefficient of 2.63 S for rSpy0128. This figure is in total agreement with the calculated sedimentation coefficient value for rSpy0128 based on its high resolution crystal structure using HYDROPRO (Garcia De La Torre *et al*., 2000) of 2.54 S (Solovyova *et al*., 2010). The SAXS dummy atom model for rSpy0128 is superimposed with its high resolution crystal structure using SUBCOMB (Kozin and Svergun, 2001) in Figure 3.15. The two have an extremely good fit, confirming that the rSpy0128 crystal structure is a good representation of the protein in solution.

Attempts in this project to determine the high resolution crystal structure of Spy0128 encountered problems and were abandoned when we learnt that a group in New Zealand had solved the structure (Kang *et al*., 2007) using a construct expressing a slightly different version of rSpy0128 protein to that produced in this project. The rSpy0128 produced in this project encompassed Glu24-Val313 (of the 'EVPTGV' sortase recognition motif), and was designed deliberately to exclude the secretion signal peptide and sequence downstream of the CWS motif that would not be present in the protein at the cell surface.

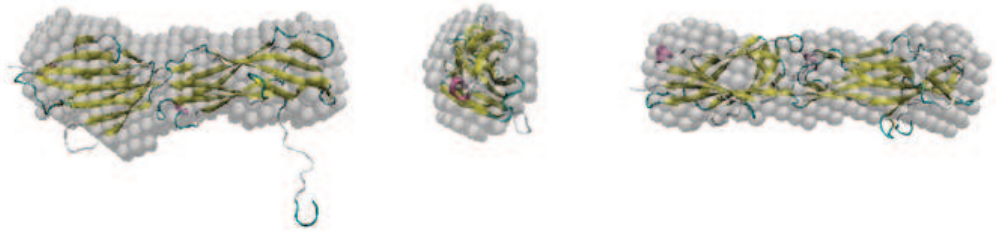

Figure 3.15: rSpy0128 DAM superimposed with the rSpy0128 crystal structure.

In contrast, the rSpy0128 used to solve the structure was comprised of Ala18-Glu308, including part of the secretion signal peptide. It could be argued that the latter was a less well designed product, but it turned out that the extra residues from the signal peptide present in this construct formed a critical crystal contact giving rise to more intimate packing, which in turn, leads to better diffraction. This interaction meant that this group were able to generate crystals that diffracted to 2.2 Å, whereas the best diffraction observed from crystals generated in this work was ~3 Å. The author's remark in their publication that another construct was generated comprising Ala18-Thr311 but did not produce crystals of sufficient quality to rival the Ala18-Glu308 construct. This is probably because the sortase recognition motif is likely to be unstructured and therefore, when not present, tighter packing of molecules within the crystal is possible (Kang *et al*., 2007).

The *intra*-molecular isopeptide bonds discovered within Spy0128 have been shown to confer stability on the protein *in vitro* by making them more resistant to attacks by proteases and giving them a higher thermal stability (Kang and Baker, 2009). However, the influence these bonds have on the function of pili *in vivo*, had not been investigated. This study attempted to address this by substituting the residues involved in *intra*-molecular isopeptide bond formation with alanine in the M1 GAS strain SF370 chromosome. A *Δspy0128* strain of GAS was used as the host in allele replacement. In parallel, wild-type *spy0128* was replaced back into the chromosome of this host to confirm the wild-type phenotype could be restored by this allele replacement approach. All aspects of the resulting strain were similar to wild-type: it bound to wild-type levels in adhesion assays and displayed identical laddering patterns on blots of cell wall and culture supernatant fractions. These data indicated that wild-type function and phenotype had been restored by allele replacement.

Lys161 of Spy0128 was confirmed *in vitro* by mass spectrometry as the residue involved in forming the *inter*-molecular link between Spy0128 molecules in the pilus (Kang *et al*., 2007). Substitution of Lys161 for Ala in the chromosome supported this observation, as it abrogated pilus formation and function. This mutant exhibits a similar phenotype to the *Δspy0128* strain in adhesion assays showing little or no binding when compared to wild-type. Additionally, as predicted, immuno-blots only detected the Spy0128 monomer and not polymerised pili. In contrast, immuno-blots of cell wall preparations from the Lys-Ala *intra*-molecular isopeptide bond mutants with anti-Spy0128 sera, reveal only small effects on the laddering patterns observed, although they appeared slightly less intense than the wild-type. For both bonds, the Glu-Ala substitutions seemed to have a much more pronounced effect on pili. It appears that the Glu mutants produce less HMW polymers at the cell wall and those observed in the culture supernatants are much shorter, with no bands seen above 250 kDa. These effects occur in both Glu mutants but look to be more pronounced in the Glu258Ala substitution (as shown in Figure 3.14). Perhaps the phenotypes observed for these mutants are caused by disrupting local hydrogen bonding networks sufficiently to reduce efficient pilus polymerisation, whilst the effects caused by the loss of the lysine residues are not as great. This is consistent with the findings of Kang *et al*. (Kang and Baker, 2009) who reported that Asn168Ala and Glu117Ala substitutions contained within rSpy0128 protein did not affect the overall structure of the protein (solved by X-ray crystallography) but did cause local disturbances to the protein core. These effects were more pronounced for the Glu117Ala mutation than the Asn168Ala. As crystallisation 'fixes' a protein it in a particular state, this local disturbance could be amplified enough *in vivo* to move Lys161 sufficiently for it to impede efficient attachment to the next Spy0128 monomer to be incorporated into the pilus. Also, the Glu258Ala mutation could cause enough disruption on the C-terminal domain of Spy0128 to compromise the ability of SrtC (which must recognise this domain) to effectively polymerise Spy0128 monomers into pili. Surprisingly, when the ability of these mutants to bind HaCaT cells was examined they all adhered to wild-type levels. The major difference observed between the binding of the mutants and wild-type M1 GAS strain SF370 was that the mutants appeared to bind in 'clumps' rather than spread evenly over the monolayer (see below).

Intra-molecular isopeptide bonds have also been found in the crystal structures of other major Gram positive pilus proteins including SpaA from *Corynebacterium diphtheriae* (Kang *et al*., 2009b) and BcpA from *Bacillus cereus* (Budzik et al., 2008a). Work on BcpA shows that the protein harbours four *intra*-molecular amide bonds. When residues involved in the formation of two of these bonds (in the CNA_2 and XNA domains) are mutated to alanine to abrogate bond formation, results very similar to those obtained in this chapter for Spy0128 are seen. Mutation of Lys174 and Lys273 of these bonds appears to have no effect on pilus phenotype compared to wild-type, as assessed by immuno-blotting and immuno-labelling with gold particles. However, mutation of Glu223 and Asp312 (the catalytic residues required for amide bond formation), results in the much reduced length of pili (Budzik *et al*., 2009b).

The tensile forces experienced by GAS pili upon and after binding have only recently been investigated. Atomic force microscopy (AFM) was used to test the forces that the wild-type Spy0128 protein and isopeptide bond mutant protein could withstand before unfolding to assess their contribution to stability (Alegre-Cebollada *et al*., 2010). While wild-type protein appeared to be almost inextensible, mutations in either of the isopeptide bonds greatly reduced the forces needed to unfold the protein, although these forces were still relatively high (Alegre-Cebollada *et al*., 2010). If such unwinding led to bacterial aggregation, this observation could explain the 'clumping' on the HaCaT monolayer. However, whether forces experienced by intact pili upon binding *in vivo* would be enough to cause unfolding of mutant Spy0128 in the pilus needs further investigation.

AFM has been used in conjunction with optical tweezers to probe the effects tensile forces have on the P pili of *E. coli* (Jass *et al*., 2004). As P pili are stretched, the major subunits which are arranged in a helical structure, unwind causing elongation of the pili. It was suggested that this reveals a 'built in' mechanism to resist shear tensile forces experienced by bacteria attached to epithelial surfaces in the human body (Jass *et al*., 2004). It could be speculated that *intra*-molecular bonds within major subunits of Gram positive bacteria could be playing a similar role.

Chapter 4. Characterisation of Spy0125

4.1 Introduction

As discussed in chapter 1, the FCT region of M1 GAS strain SF370 contains five genes involved in forming cell surface pili. One of these, a specific sortase enzyme (encoded by *spy0129*, originally designated sortase C or SrtC) acts at the outer surface of the bacterial membrane as a transpeptidase, responsible for forming the covalent linkages between pilus subunits following their secretion across the cytoplasmic membrane *via* the general secretion pathway. This action led to SrtC being termed the 'pilus polymerase'. A second gene (*spy0127*) was originally described as encoding a putative signal peptidase based on sequence homologies with other proteins. However, when examined more closely it was discovered that Spy0127 lacked signal peptidase activity and was suggested to play a role as a chaperone to the pilus subunits although this has yet to be confirmed (Zahner and Scott, 2008). The three remaining genes (*spy0125*, *spy0128* and *spy0130*) encode subunits which are incorporated into the pilus structure. As already mentioned in chapter 1, at the beginning of this project, Spy0128 had been described as the major pilus subunit that is assembled into the pilus shaft while the other two, Spy0125 and Spy0130, had been described as minor subunits but their specific roles were not known (Mora *et al*., 2005). All three pilus subunits were to be characterised with the objective of probing their physical structure and role in pilus function and biogenesis. Depending on the progress made with each subunit and parallel work by other laboratories it was anticipated that these studies would progress at different rates. This proved to be the case, with a large proportion of this project focussing on Spy0125.

During the first year of this project, it was reported by Manetti *et al*. (Manetti *et al*., 2007) that recombinant proteins representing the minor subunits (rSpy0125 and rSpy0130) but not the major subunit (rSpy0128) bound to a human pharyngeal cell line. This suggested that both minor subunits could be acting as adhesins and therefore could be positioned at the pilus tip. If this were the case, it would be expected that deletion of one of these genes from the GAS chromosome would only result in a partial loss of adhesion. Studies from our laboratory by Abbot *et al.* (Abbot *et al*., 2007) revealed that this was not the case and that a single deletion of either *spy0125* or *spy0130* resulted in

total loss of adhesion to both human tonsil and cultured keratinocytes. This suggested that both minor subunits play equally essential roles in pilus biogenesis. Later work on Spy0130 revealed it to play a specific role in anchoring the pilus to the cell wall as a 'linker' molecule located at the base of the pilus (see chapter 3). The experiments described in this chapter established that Spy0125 is localised to the tip of the pilus and acts as the pilus presented adhesin.

4.2 Recombinant Spy0125 expression and purification

At the beginning of the work described in this chapter, a recombinant plasmid consisting of the *E. coli* pET-28a(+) expression vector containing cloned *spy0125* sequences had been constructed by Dr Wendy Smith and was available in the laboratory. This plasmid (pWS011) was designed to express recombinant (r)Spy0125 protein lacking the N-terminal signal peptide and the sequences downstream of the C-terminal CWSS after the valine of the VVPTG(V) motif, as these sequences would be absent from the mature form of the protein on the GAS cell surface. The *spy0125* sequence cloned into pET-28a(+) is shown in panel A, Figure 4.1. In this project, prior to large scale rSpy0125 expression, preliminary small scale cultures were setup to identify optimal growth and induction conditions. During these preliminary experiments it was first noted that induced rSpy0125 runs a little slowly on SDS-PAGE, giving an apparent molecular mass of ~94 kDa, whereas the sequence indicates 78.5 kDa. This retarded mobility on SDS-PAGE was subsequently observed consistently throughout this project. Reduced protein mobility on SDS-PAGE is not uncommon and the particular reasons for the reduced mobility of rSpy0125 on SDS-PAGE gels were not explored in this project. It is interesting to note that the RlrA pilus adhesin, RrgA, of *S. pneumoniae* also runs higher than expected on SDS-PAGE (LeMieux *et al*., 2008).

Having examined a variety of strains and induction conditions, it was decided to use *E. coli* BL21 (DE3) CodonPlus type RIL for rSpy0125 expression and that cells were to be induced at a culture density OD_{600} of 0.6 with 1 mM IPTG for 3 h, as this combination produced the greatest quantities of recombinant protein. Growth conditions determined during small scale expression tests were applied to large scale production of rSpy0125. Following induction, cells were harvested and lysed by sonication. The lysate was then subjected to affinity chromatography on a His-tag column and

A

T7 Promoter → lac operator

TAATACGACTCACTATAGG**GGAATTGTGAGCGGATAACAATTC**CCCTCTAGAAATAATTTTGTTTAACT

rbs — His-tag — Thrombin site

TTAAG**AAGGAG**ATATACC**ATG**GGCAGCAGC**CATCATCATCATCATCAC**AGCAGCGGC**CTGGTGCCGCGC**

GGCAGCCATATGAAGACTGTTTTTGGTTTAGTAGAATCCTCGACGCCAAACGCAATAAATCCAGATTCAAGTTCGGAA
TACAGATGGTATGGATATGAATCTTATGTAAGAGGGCATCCATATTATAAACAGTTTAGAGTAGCACACGATTTAAGG
GTTAACTTAGAAGGAAGTAGAAGTTATCAAGTTTATTGCTTTAATTTAAAGAAAGCATTTCCTCTCGGATCAGATAGT
AGTGTTAAAAAGTGGTATAAAAAACATGATGGAATCTCTACAAAATTTGAAGATTATGCGATGAGCCCTAGAATTACG
GGAGATGAGCTAAATCAGAAGTTACGAGCTGTTATGTATAATGGACATCCACAAAATGCCAATGGTATTATGGAAGGC
TTGGAACCCTTGAATGCTATCAGAGTTACACAAGAGGCGGTATGGTACTATTCTGATAATGCTCCTATTTCTAATCCA
GATGAAAGTTTTAAAAGGGAGTCAGAAAGTAACTTGGTTAGTACTTCTCAATTATCTTTGATGCGTCAAGCTTTGAAG
CAACTGATTGATCCGAATTTGGCAACTAAAATGCCAAAACAAGTTCCGGATGATTTTCAGCTAAGTATTTTTGAGTCT
GAGGACAAGGGAGATAAATATAATAAAGGATACCAAAATCTTTTGAGTGGTGGTTTAGTTCCTACTAAACCACCAACT
CCAGGAGACCCACCAATGCCTCCAAATCAACCTCAAACGACTTCAGTACTTATTAGAAAGTATGCTATAGGTGATTAC
TCTAAATTGCTTGAAGGTGCAACATTACAGTTGACAGGGGATAACGTGAATAGTTTTCAAGCGAGAGTGTTTAGCAGT
AATGATATTGGAGAAAGAATTGAACTATCAGATGGAACTTATACTTTAACTGAATTGAATTCTCCAGCTGGTTATAGT
ATCGCAGAGCCAATCACTTTTAAGGTTGAAGCTGGCAAAGTGTATACTATTATTGATGGAAAACAGATTGAAAATCCC
AATAAAGAGATAGTAGAGCCTTACTCAGTAGAAGCATATAATGATTTTGAAGAATTTAGCGTTTTAACTACACAAAAC
TATGCAAAATTTTATTATGCAAAAAATAAAAATGGAAGTTCACAGGTTGTCTATTGCTTTAATGCAGATCTAAAATCT
CCACCAGACTCTGAAGATGGTGGGAAAACAATGACTCCAGACTTTACAACAGGAGAAGTAAAATACACTCATATTGCA
GGTCGTGACCTCTTTAAATATACTGTGAAACCAAGAGATACCGATCCTGACACTTTCTTAAAACATATCAAAAAAGTA
ATTGAGAAGGGTTACAGGGAAAAAGGACAAGCTATTGAGTATAGTGGTCTAACTGAGACACAATTGCGTGCGGCTACT
CAGTTAGCAATATATTATTTCACTGATAGTGCTGAATTAGATAAGGATAAACTAAAAGACTATCATGGTTTTGGAGAC
ATGAATGATAGTACTTTAGCAGTTGCTAAAATCCTTGTAGAATACGCTCAAGATAGTAATCCTCCACAGCTAACTGAC
CTTGATTTCTTTATTCCGAATAACAATAAATATCAATCCCTTATTGGAACTCAGTGGCATCCAGAAGATTTAGTTGAT
ATTATTCGTATGGAAGATAAAAAAGAAGTTATACCTGTAACTCATAATTTAACATTGAGAAAAACGGTGACTGGTTTA
GCTGGTGACAGAACTAAAGATTTCCATTTTGAAATTGAATTAAAAAATAATAAGCAAGAATTGCTTTCTCAAACTGTT
AAAACAGATAAAACAAACCTCGAATTTAAAGATGGTAAAGCAACCATTAATTTAAAACATGGGGAAAGTTTAACACTT
CAAGGTTTACCAGAAGGTTATTCTTACCTTGTCAAAGAAACAGATTCTGAAGGCTATAAGGTTAAAGTTAATAGCCAA
GAAGTAGCAAATGCTACAGTTTCAAAAACAGGAATAACAAGTGATGAGACACTTGCTTTTGAAAATAATAAAGAGCCT
GTTGTTCCTACAGGAGTTTAACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAA

B

GSHMKTVFGLVESSTPNAINPDSSSEYRWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAFPLGSDS
SVKKWYKKHDGISTKFEDYAMSPRITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQEAVWYYSDNAPISNP
DESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKGDKYNKGYQNLLSGGLVPTKPPT
PGDPPMPPNQPQTTSVLIRKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAGYS
IAEPITFKVEAGKVYTIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQVVYCFNADLKS
PPDSEDGGKTMTPDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREKGQAIEYSGLTETQLRAAT
QLAIYYFTDSAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTDLDFFIPNNNKYQSLIGTQWHPEDLVD
IIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNKQELLSQTVKTDKTNLEFKDGKATINLKHGESLTL
QGLPEGYSYLVKETDSEGYKVKVNSQEVANATVSKTGITSDETLAFENNKEPVVPTGV

Figure 4.1: Structure of recombinant plasmid pWS011.

Panel A: The *spy0125* DNA sequences cloned in pWS011 are shown in blue with primer encoded stop codon at the end of the sequence underlined. Vector sequences are shown in black, with important features underlined in bold type and indicated above the sequence. *Nde*I (forward primer) and *Xho*I (reverse primer) restriction endonuclease sites used to clone into the vector are shown in red type adjacent to the cloned DNA. The *spy0125* DNA fragment was cloned in-frame with the vector encoded His-tag creating a fusion protein. **Panel B**: The rSpy0125 amino acid sequence. The Spy0125 sequences are shown in blue, with vector encoded residues remaining at the N-terminus after thrombin cleavage removing the His-tag shown in black.

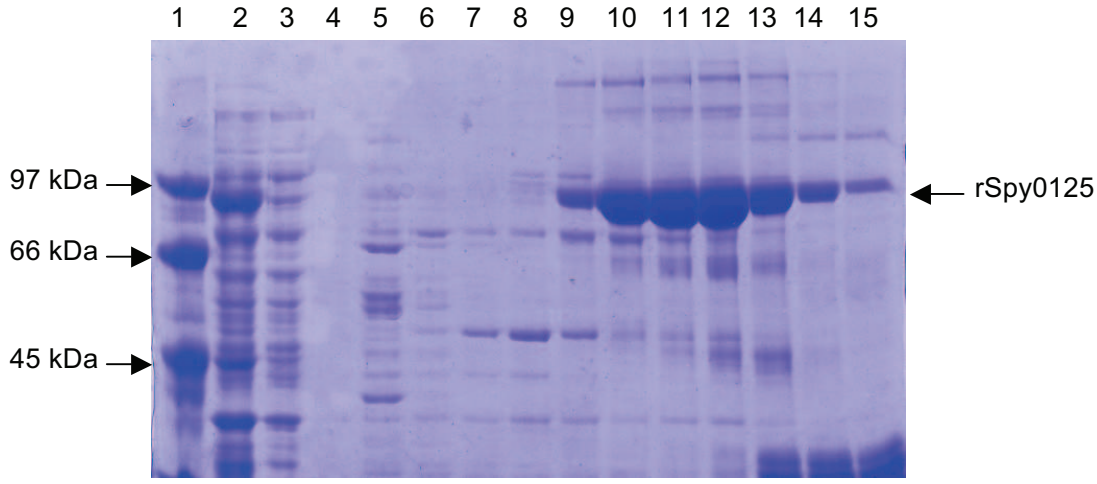

Figure 4.2: Expression and purification of rSpy0125 by affinity chromatography.

Coomassie-blue stained SDS-PAGE gel showing rSpy0125. Lane 1: molecular size standards. Lane 2: sample of induced cell lysate. Lane 3: sample of unbound material from His-tag column. Lanes 4-15: fractions collected during rSpy0125 elution. Fractions corresponding to the absorbance peak of the elution profile, lanes 9-14, were pooled.

eluted fractions that contained large quantities of protein were pooled and concentrated (Figure 4.2). As seen in Figure 4.2 however, these fractions were still contaminated with host cell proteins. Therefore, the pooled and concentrated rSpy0125 was purified further using a Hi-Load 16/60 Superdex 200 gel filtration column. Figure 4.3 panel A shows the rSpy0125 gel filtration elution profile at 280 nm revealing a single predominant peak. Fractions corresponding to this peak were analysed by SDS-PAGE and are shown in Figure 4.3, panel B with rSpy0125 revealed as a single band of highly pure protein, running slightly high at ~94 kDa as noted above. Sample fractions in lanes 5-8 in Figure 4.3 panel B were pooled and concentrated to 10 mg/ml (determined by absorbance at 280 nm using the molar extinction coefficient for Spy0125 of 78745 M^{-1} cm^{-1}) and stored at 4^{o}C. When needed, the N-terminal His-tag was cleaved off by thrombin digestion, followed by removal of the separated His-tag (on a His-chelation column) and thrombin (on a benzamidine column), as described in the materials and methods, section 2.6.7. The resulting rSpy0125 was pooled, dialysed into gel filtration buffer and concentrated. The rSpy0125 amino acid sequence including vector encoded sequences remaining at the N-terminus after removal of the His-tag is shown in panel B, Figure 4.1.

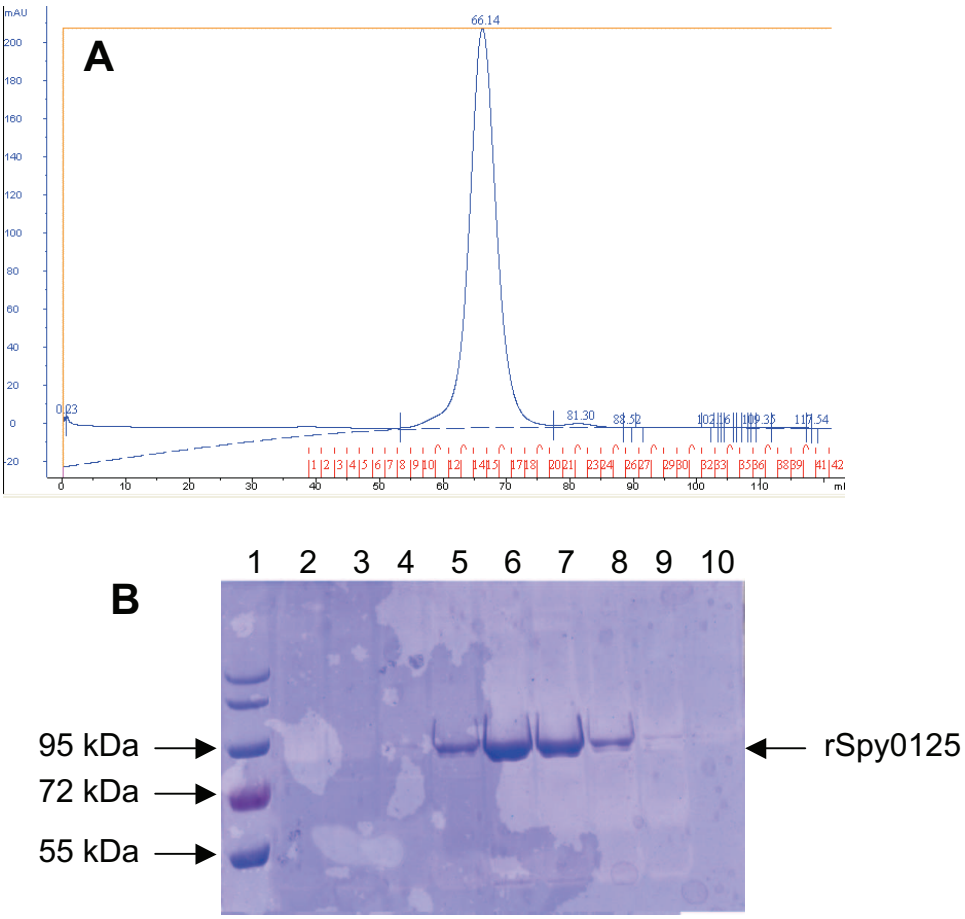

Figure 4.3: Purification of rSpy0125 by gel filtration.

Panel A: rSpy0125 gel filtration elution profile at 280 nm. **Panel B**: coomassie blue stained SDS-PAGE gel analysing peak fractions. Lane 1: molecular size standards. Lanes 2-10: samples of fractions corresponding to the peak from the elution profile. Fractions corresponding to lanes 5-8 were pooled and concentrated.

4.3 Biophysical characterisation of rSpy0125 in solution

Purified rSpy0125 from section 4.2 (with the N-terminal His-tag removed) was used in biophysical characterisation experiments. Circular dichroism spectroscopy (CD), analytical ultracentrifugation (AUC) and small angle X-ray scattering (SAXS) were used to probe the physical structure of rSpy0125 in solution.

Both AUC and SAXS experiments were carried out in collaboration with Dr. Alexandra Solovyova who provides these techniques as a service for our institute.

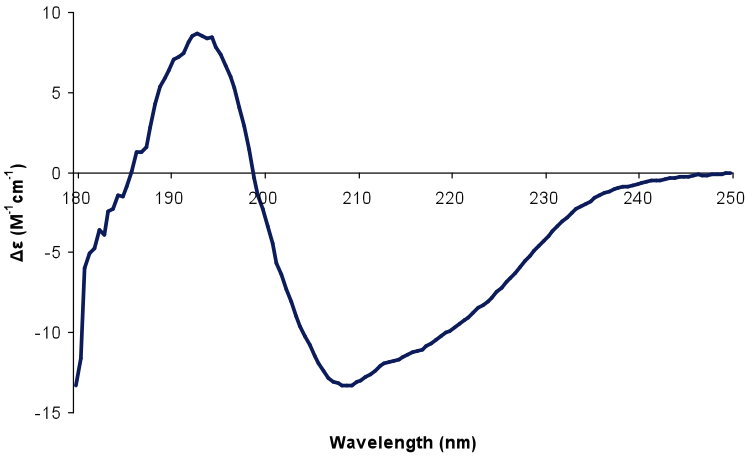

Figure 4.4: Far UV spectrum of rSpy0125.

rSpy0125 displays a non-classical β-$_{II}$ spectrum with a small α-helical component.

4.3.1 Secondary structure of rSpy0125

The secondary structure of rSpy0125 was examined using CD. For use in CD experiments stock rSpy0125 was diluted to 0.5 mg/ml. A final spectrum was determined by averaging 5 repeat scans and subtracting the appropriate protein-free buffer spectrum. Figure 4.4 shows the far UV spectrum for rSpy0125, which displays a non-classical *β*-$_{II}$ spectrum and contains mostly β-sheet with a small proportion of α-helix (Solovyova *et al*., 2010).

4.3.2 Analytical ultracentrifugation of rSpy0125

Concentrations of rSpy0125 in gel filtration buffer (20 mM Tris-Cl, 150 mM NaCl) ranging from 0.15-0.87 mg/ml were used in sedimentation velocity experiments. Using the continuous c(*s*) distribution model, as shown in Figure 4.5, a single dominant species is observed as a single peak at 4.30 S, suggesting that rSpy0125 is monodisperse in the sample. The sedimentation coefficient for rSpy0125 that was determined experimentally, 4.30 S, is significantly lower than would be calculated from a typically hydrated sphere of equivalent molecular mass, 5.65 S, which indicates a significant elongation of rSpy0125 (Solovyova *et al*., 2010).

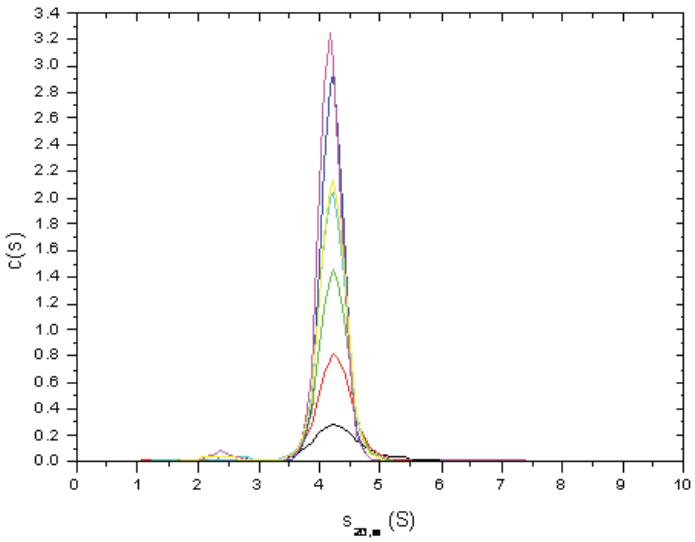

Figure 4.5: Size distribution c(*s*) for rSpy0125.

A single dominant peak is observed at 4.30 S using the general c(*s*) size distribution model. The concentrations of protein examined were: 0.15 mg/ml (black line), 0.31 mg/ml (red line), 0.51 mg/ml (green line), 0.65 mg/ml (cyan line), 0.68 mg/ml (yellow line), 0.76 mg/ml (blue line), 0.87 mg/ml (magenta line).

Table 4.1: rSpy0125 sedimentation velocity results.

	Calculated parameters		SV experimental data			
	Mass (kDa)	*s* (S)	RMSD[a]	f/f_0^{shape}	$S_{20,w}$	Mass (kDa)
Spy0125	76.7	5.65	0.0161	1.41	4.30 ± 0.035	81.81 ± 1.6

[a] The root mean square deviation (RMSD) of the fit of the data for the highest sample concentration is shown.

In addition, the derived value of f/f_0^{shape}, 1.41, was much higher than 1 (f/f_0 = 1 for a sphere) which also implies elongation of the protein (Table 4.1). Using the sedimentation velocity data, an experimentally derived mass of 81.8 kDa was obtained and confirms that rSpy0125 is monomeric in solution when compared to the calculated mass of the expressed Spy0125 construct, 76.7 kDa (Solovyova 2010 *et al*., 2010).

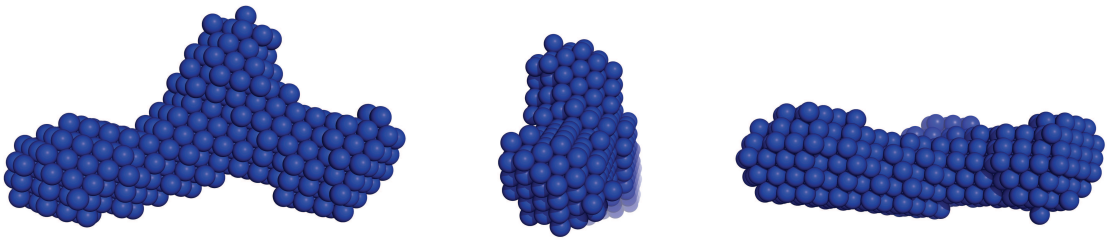

Figure 4.6: SAXS DAM for rSpy0125 in three orientations.

The approximate dimensions for the rSpy0125 monomer were 136 Å x 51 Å x 55 Å.

4.3.3 rSpy0125 solution structure modelled *ab initio*

For use in SAXS experiments stock rSpy0125 sample that had been confirmed as monodisperse and monomeric in solution by AUC was used at 5 mg/ml and 10 mg/ml in gel filtration buffer. The SAXS data for rSpy0125 was collected on beamline 2.1 at the Synchrotron Radiation Source (Daresbury, UK). The low resolution solution structure was determined by Dr Alexandra Solovyova by averaging 20 *ab initio* restored models, using DAMMIN (Svergun, 1999) in the form of dummy atom models (DAM) with a resolution of 25 Å. The constructed DAM for rSpy0125 shown in Figure 4.6 and the approximate dimensions of the protein envelope were: 136 Å x 51 Å x 55 Å (Solovyova *et al*., 2010). No information about atomic positions within the protein molecule can be obtained from SAXS data so it is very difficult to try to predict where potentially interesting features of the protein such as the 'VVPTG' motif and any binding interfaces might actually be located. The DAM for rSpy0125 contains significant branching which is indicative of a multi-domain structure for this protein.

4.4 Characterisation of rSpy0125 breakdown products

It was noted that after purification, rSpy0125 (stored at 4^{o}C) was observed to breakdown over less than a week into two major fragments along with a number of less prominent species (Figure 4.7). In order to explore this in more detail, samples of fresh rSpy0125 (10 mg/ml) were stored at ambient room temperature, 20^{o}C, 4^{o}C, -20^{o}C and -80^{o}C. At regular time intervals over a period of 7 weeks, samples of rSpy0125 stored at each temperature were analysed by SDS-PAGE. As seen in Figure 4.7, it is quite clear

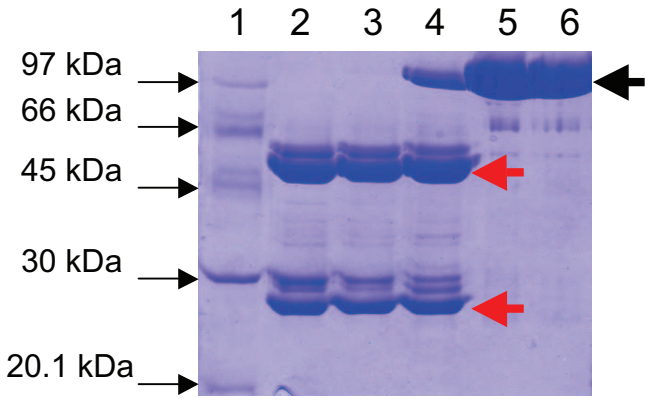

Figure 4.7: rSpy0125 degradation over time.

Coomassie blue stained SDS-PAGE gel of rSpy0125 breakdown after storage at various temperatures for 7 weeks. Lane 1: molecular size standards. Lane 2: ambient room temperature. Lane 3: 20°C. Lane 4: 4°C. Lane 5: -20°C. Lane 6: -80°C. Arrows show full length rSpy0125 (Black) and the two major breakdown products (Red).

at 7 weeks post purification that rSpy0125 stored at ambient room temperature and 20°C has completely degraded to two major fragments of ~25 kDa and ~50 kDa. A similar result is observed for rSpy0125 stored at 4°C although some intact protein remains, whereas rSpy0125 stored at -20°C and -80°C is largely intact with only a hint of breakdown after 7 weeks. It was decided that the two major breakdown fragments observed at ~25 kDa and ~50 kDa were likely to be quite stable because these bands on SDS-PAGE only increased in intensity over time with very little further breakdown. Potentially, these fragments could make up two stable domains of the intact rSpy0125 (~25 kDa + ~50 kDa = ~76.7 kDa) although this does not rule out the possibility that the fragment at ~50 kDa could be degrading into 2 equally sized fragments of ~25 kDa. In order to identify which regions of the intact protein the two peptide fragments represent, the rSpy0125 sample stored at 4°C (lane 4, Figure 4.7) was applied to an S75 gel filtration column to separate the three fragments in solution. Upon isolation, the two breakdown fragments were characterised by mass spectrometry. Both fragments were subject to FT-ICR MS which provides an extremely accurate mass. In the sample eluted at ~25 kDa two distinct masses were identified (data not shown). For analysis, both peptide masses were rounded up to 2 decimal places and entered into the FindPept tool (Gattiker et al., 2002) on the ExPASy server (Gasteiger *et al*., 2005) along with the Spy0125 sequence. Due to the extremely accurate measurements obtained by the FT-

ICR MS this program was able to assign peptide sequence from Spy0125 which had the closest calculated equivalent mass as that achieved experimentally. The first mass of 25674.68 Da, was assigned to the peptide Val54-Gly279, which has a calculated mass of 25674.676 Da and is only 0.004 Da different to the experimentally determined mass. The second mass (25844.94 Da) found within the lower band on SDS-PAGE was assigned to the peptide Gly52-Gly279 which has a calculated mass of 25844.887 Da and is only 0.052 Da different to the experimentally determined mass. The difference between the two experimentally determined masses of 170.26 Da can be attributed to the Gly-Leu (mass 170.2 Da) that is not present on the N-terminus of the peptide with the lower mass.

The sample eluted at ~50 kDa was attributed a mass of 49769.14 Da. This mass alone however would not be enough information to be able to identify the exact peptide which composed the larger degradation product. This is due to the level of accuracy achieved by FT-ICR MS at this mass but more importantly, this fragment was likely to be an internal portion of Spy0125. In order to address this issue and attribute a definitive peptide, the larger breakdown fragment was blotted onto PVDF membrane as described in the materials and methods, section 2.8.2, and submitted for N-terminal amino acid sequencing. During Edman chemistry, amino acids are sequentially removed from the N-terminus and identified by reversed phase high pressure liquid chromatography depending on when they are eluted over time compared to a set of amino acid standards. The area under the amino acid peak is also quantitative which can help identification of the cleaved residue in a mixed sample. Six cycles of Edman chemistry were possible that identified Asn – Gln – Pro – Gln – Thr – Thr as the first 6 amino acids of this degradation product. These residues could be identified precisely from the Spy0125 sequence, and correspond to Asn286-Thr291 (data not shown). This information coupled with the intact mass of the fragment enabled the peptide Asn286-Val725 to be assigned to the larger fragment. The calculated mass (49354.1 Da) of this fragment however was someway short of the intact mass of 49769.14 Da determined by the FT-ICR MS (see discussion). When all the information collected for both degradation products is considered together, the most likely scenario for the breakdown of rSpy0125 in solution is that the protein is cleaving in a proline rich region that separates the two fragments (Figure 4.8). The presence of a proline rich region could indicate that two

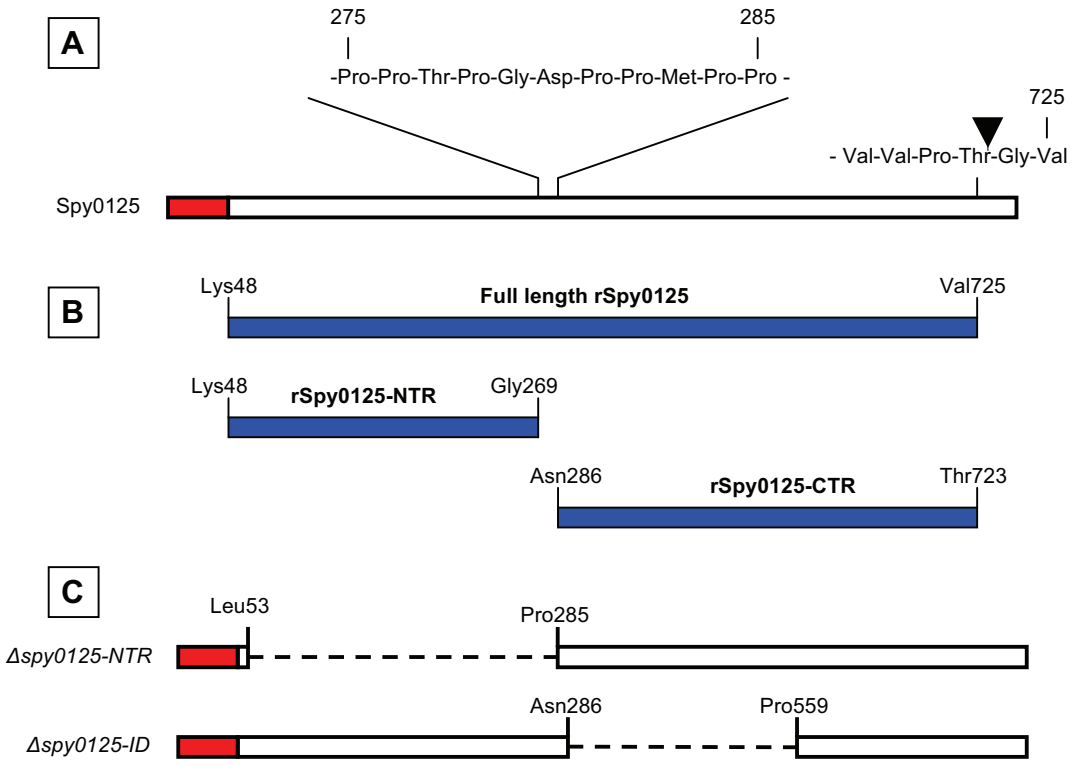

Figure 4.8: Recombinant proteins and chromosomal deletions produced in this chapter.

Panel A: Full length Spy0125 with its N-terminal signal peptide highlighted in red and VVPTGV motif indicated at the C-terminus. Also highlighted is the proline rich region identified as separating the two rSpy0125 breakdown fragments. **Panel B**: Recombinant Spy0125 proteins and where they correspond to full length Spy0125. Lys48 indicates the N-terminus of the mature, secreted protein. **Panel C**: Regions of DNA deleted in *spy0125* with the first and last residues of each deletion indicated (see section 4.6).

domains of rSpy0125 have been isolated and to facilitate further studies, both were recloned into pET-28a(+)-D.

4.5 Cloning, expression and purification of rSpy0125 'breakdown' fragments

The first construct spanned Lys48 (the first residue after the predicted signal peptide is removed)-Gly269 (leaving out part of the proline rich region) and was termed the N-terminal region (Spy0125-NTR), while the second spanned Asn286-Thr723 (the 'T' of the 'VVPTG' motif) to ensure the fully processed version was constructed and was termed the C-terminal region (Spy0125-CTR). The structure of the recombinant fragments and where they correspond to full length Spy0125 is shown in Figure 4.8. Primers were constructed to encompass the regions described above and contained *BamH*I (forward primers) and *Xho*I (reverse primers) restriction endonuclease sites ready for cloning into the pET-28a(+)-D modified expression vector multiple cloning site. Reverse primers also contained a termination codon after the final cloned sequences to stop read through into vector encoded sequences. The final *spy0125-NTR* (pJP07) and *spy0125-CTR* (pJP08) sequences cloned into pET-28a(+)-D are shown in Figures 4.9 and 4.10 and were confirmed by DNA sequencing.

The expression and purification of rSpy0125-NTR and rSpy0125-CTR was essentially as described in section 4.2 for intact rSpy0125. The plamsids pJP07 and pJP08 were transformed into *E.coli* BL21 (DE3) and *E. coli* BL21 (DE3) CodonPlus type RIL strains and small scale expression tests were carried out to identify the optimal induction conditions. As before with intact rSpy0125, the *E. coli* BL21 (DE3) CodonPlus type RIL expression strain in combination with induction at a culture density OD_{600} of 0.6 with 1 mM IPTG for 3 h was observed to give the greatest expression of both proteins. Optimised growth conditions were implemented during large scale production of both proteins and after harvesting, cells were lysed by sonication and soluble His-tagged protein was recovered by Ni^{2+} affinity chromatography. Peak fractions were analysed by SDS-PAGE and those containing large quantities of protein were pooled and concentrated prior to loading on a Hi-Load 16/60 Superdex 75 (rSpy0125-NTR) or 200 (rSpy0125-CTR) gel filtration column to remove impurities carried over from the His-tag purification.

A

T7 Promoter → lac operator

TAATACGACTCACTATAGG**GGAATTGTGAGCGGATAACAATTC**CCCTCTAGAAATAATTTTGTTTAACT

rbs His – tag Thrombin site

TTAAG**AAGGAG**ATATACCATGGGCAGCAGC**CATCATCATCATCATCAC**AGCAGCGGC**CTGGTGCCGCGC**

GGATCCAAGACTGTTTTTGGTTTAGTAGAATCCTCGACGCCAAACGCAATAAATCCAGATTCAAGTTCGG
AATACAGATGGTATGGATATGAATCTTATGTAAGAGGGCATCCATATTATAAACAGTTTAGAGTAGCACA
CGATTTAAGGGTTAACTTAGAAGGAAGTAGAAGTTATCAAGTTTATTGCTTTAATTTAAAGAAAGCATTT
CCTCTCGGATCAGATAGTAGTGTTAAAAAGTGGTATAAAAAACATGATGGAATCTCTACAAAATTTGAAG
ATTATGCGATGAGCCCTAGAATTACGGGAGATGAGCTAAATCAGAAGTTACGAGCTGTTATGTATAATGG
ACATCCACAAAATGCCAATGGTATTATGGAAGGCTTGGAACCCTTGAATGCTATCAGAGTTACACAAGAG
GCGGTATGGTACTATTCTGATAATGCTCCTATTTCTAATCCAGATGAAAGTTTTAAAAGGGAGTCAGAAA
GTAACTTGGTTAGTACTTCTCAATTATCTTTGATGCGTCAAGCTTTGAAGCAACTGATTGATCCGAATTT
GGCAACTAAAATGCCAAAACAAGTTCCGGATGATTTTCAGCTAAGTATTTTTGAGTCTGAGGACAAGGGA
GATAAATATAATAAAGGATACCAAAATCTTTTGAGTGGTGGTTAA**CTCGAG**CACCACCACCACCACCACT
GAGATCCGGCTGCTAA

B

GSKTVFGLVESSTPNAINPDSSSEYRWYGYESYVRGHPYYKQFRVAHDLRVNLEGSRSYQVYCFNLKKAF
PLGSDSSVKKWYKKHDGISTKFEDYAISPRITGDELNQKLRAVMYNGHPQNANGIMEGLEPLNAIRVTQE
AVWYYSDNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDDFQLSIFESEDKG
DKYNKGYQNLLSGG

Figure 4.9: Recombinant plasmid pJP07.

Panel A: Cloned sequences for *spy0125-NTR* are shown in blue, with the primer encoded stop codon at the end of the sequence underlined. The *BamH*I and *Xho*I restriction endonuclease sites are shown in bold red type adjacent to the cloned DNA. Vector encoded sequences are shown in black with important features underlined and indicated above the sequence. The vector encoded start and stop codons are also underlined. **Panel B**: The rSpy0125-NTR amino acid sequence. The rSpy0125-NTR sequences are shown in blue with vector encoded residues remaining at the N-terminus after thrombin cleavage removing the His-tag shown in black.

A

T7 Promoter → lac operator

TAATACGACTCACTATAGG**GGAATTGTGAGCGGATAACAATTC**CCCTCTAGAAATAATTTTGTTTAACT

rbs His-tag Thrombin site

TTAAG**AAGGAG**ATATACCATGGGCAGCAGC**CATCATCATCATCATCAC**AGCAGCGGC**CTGGTGCCGCGC**

GGATCCAATCAACCTCAAACGACTTCAGTACTTATTAGAAAGTATGCTATAGGTGATTACTCTAAATTGC
TTGAAGGTGCAACATTACAGTTGACAGGGGATAACGTGAATAGTTTTCAAGCGAGAGTGTTTAGCAGTAA
TGATATTGGAGAAAGAATTGAACTATCAGATGGAACTTATACTTTAACTGAATTGAATTCTCCAGCTGGT
TATAGTATCGCAGAGCCAATCACTTTTAAGGTTGAAGCTGGCAAAGTGTATACTATTATTGATGGAAAAC
AGATTGAAAATCCCAATAAAGAGATAGTAGAGCCTTACTCAGTAGAAGCATATAATGATTTTGAAGAATT
TAGCGTTTTAACTACACAAAACTATGCAAAATTTTATTATGCAAAAAATAAAAATGGAAGTTCACAGGTT
GTCTATTGCTTTAATGCAGATCTAAAATCTCCACCAGACTCTGAAGATGGTGGGAAAACAATGACTCCAG
ACTTTACAACAGGAGAAGTAAAATACACTCATATTGCAGGTCGTGACCTCTTTAAATATACTGTGAAACC
AAGAGATACCGATCCTGACACTTTCTTAAAACATATCAAAAAAGTAATTGAGAAGGGTTACAGGGAAAAA
GGACAAGCTATTGAGTATAGTGGTCTAACTGAGACACAATTGCGTGCGGCTACTCAGTTAGCAATATATT
ATTTCACTGATAGTGCTGAATTAGATAAGGATAAACTAAAAGACTATCATGGTTTTGGAGACATGAATGA
TAGTACTTTAGCAGTTGCTAAAATCCTTGTAGAATACGCTCAAGATAGTAATCCTCCACAGCTAACTGAC
CTTGATTTCTTTATTCCGAATAACAATAAATATCAATCTCTTATTGGAACTCAGTGGCATCCAGAAGATT
TAGTTGATATTATTCGTATGGAAGATAAAAAAGAAGTTATACCTGTAACTCATAATTTAACATTGAGAAA
AACGGTGACTGGTTTAGCTGGTGACAGAACTAAAGATTTCCATTTTGAAATTGAATTAAAAAATAATAAG
CAAGAATTGCTTTCTCAAACTGTTAAAACAGATAAAACAAACCTCGAATTTAAAGATGGTAAAGCAACCA
TTAATTTAAAACATGGGGAAAGTTTAACACTTCAAGGTTTACCAGAAGGTTATTCTTACCTTGTCAAAGA
AACAGATTCTGAAGGCTATAAGGTTAAAGTTAATAGCCAAGAAGTAGCAAATGCTACAGTTTCAAAAACA
GGAATAACAAGTGATGAGACACTTGCTTTTGAAAATAATAAAGAGCCTGTTGTTCCTACATAA**CTCGAG**C
ACCACCACCACCACCACTGAGATCCGGCTGCTAA

B

GSNQPQTTSVLIRKYAIGDYSKLLEGATLQLTGDNVNSFQARVFSSNDIGERIELSDGTYTLTELNSPAG
YSIAEPITFKVEAGKVYTIIDGKQIENPNKEIVEPYSVEAYNDFEEFSVLTTQNYAKFYYAKNKNGSSQV
VYCFNADLKSPPDSEDGGKTMTPDFTTGEVKYTHIAGRDLFKYTVKPRDTDPDTFLKHIKKVIEKGYREK
GQAIEYSGLTETQLRAATQLAIYYFTDSAELDKDKLKDYHGFGDMNDSTLAVAKILVEYAQDSNPPQLTD
LDFFIPNNNKYQSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNNK
QELLSQTVKTDKTNLEFKDGKATINLKHGESLTLQGLPEGYSYLVKETDSEGYKVKVNSQEVANATVSKT
GITSDETLAFENNKEPVVPT

Figure 4.10: Recombinant plasmid pJP08.

Panel A: Cloned sequences for *spy0125-CTR* are shown in blue, with the primer encoded stop codon at the end of the sequence underlined. The *BamH*I and *Xho*I restriction endonuclease sites are shown in bold red type adjacent to the cloned DNA. Vector encoded sequences are shown in black with important features underlined and indicated above the sequence. The vector encoded start and stop codons are also underlined. **Panel B**: rSpy0125-CTR amino acid sequence is shown in blue with remaining vector encoded residues at the N-terminus after thrombin cleavage of His-tag shown in black.

The elution profiles for both proteins are shown in Figure 4.11 accompanied by samples of the peak fractions assessed by SDS-PAGE. Panels A and B of Figure 4.11 reveal that rSpy0125-NTR purifies in a monomer-dimer equilibrium and fractions which contained monomer (lanes 5-10 Panel B, Figure 4.11) were of sufficient purity to be pooled and concentrated to 10 mg/ml, determined by absorbance at 280 nm using the molar extinction coefficient for Spy0125-NTR of 38850 M^{-1} cm^{-1}. rSpy0125-CTR purifies as a single predominant peak as indicated in Figure 4.11 panel C and is assessed by SDS-PAGE to be homogenous and of high purity (panel D, Figure 4.11). Peak fractions represented in lanes 3-8 were pooled and concentrated to 10 mg/ml, determined by absorbance at 280 nm using the molar extinction coefficient for Spy0125-CTR of 39770 M^{-1} cm $^{-1}$. When needed, the N-terminal His-tag was removed by digestion with thrombin. After removal of cleaved His-tag (on a His-chelation column) and thrombin (on a benzamidine column) as described in chapter 2, section 2.6.7, rSpy0125-NTR and CTR was pooled, dialysed into gel filtration buffer and concentrated.

The secondary structure of rSpy0125-NTR and CTR were examined by CD as described above for rSpy0125 and Figure 4.12 shows example far UV spectra for both proteins. Panel A shows rSpy0125-NTR to be well folded and contain a significant α-helical component. The example spectra for rSpy0125-CTR in panel B shows this protein is also well folded and is predominantly composed of β-strand with a small α-helical component.

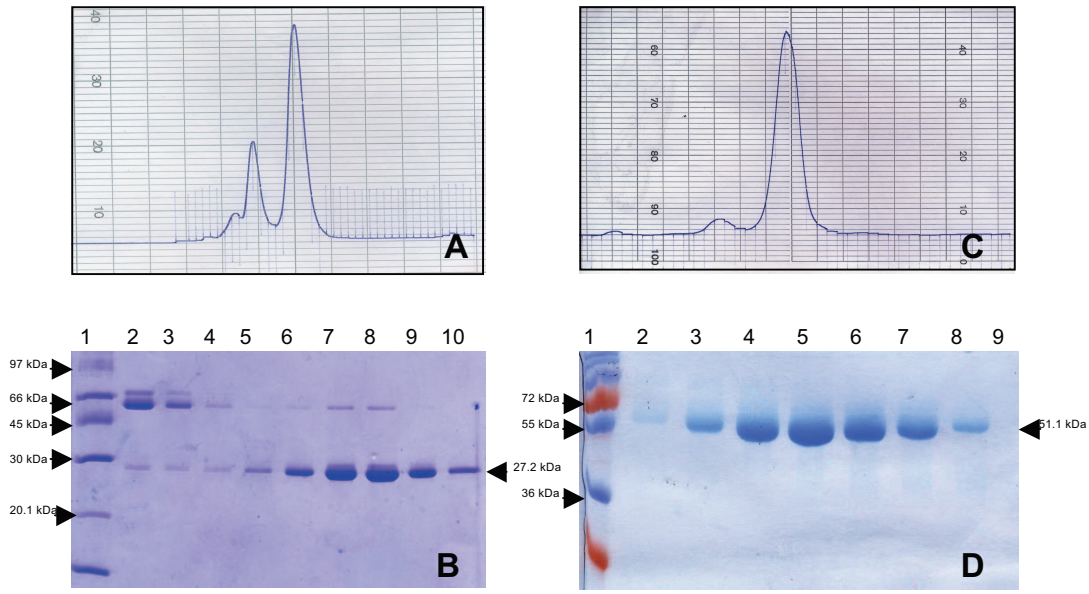

Figure 4.11: Elution profiles and peak fraction analysis of rSpy0125-NTR and CTR purification.

Panel A: rSpy0125-NTR elution profile at 280 nm. **Panel B**: Coomassie blue stained gel of corresponding peak fractions from rSpy0125-NTR gel filtration (lanes 2-10). **Panel C**: rSpy0125-CTR elution profile at 280 nm. **Panel D**: Coomassie blue stained gel of corresponding major peak fractions from rSpy0125-CTR gel filtration (lanes 2-9). Lane 1 in panels C and D: molecular size standards.

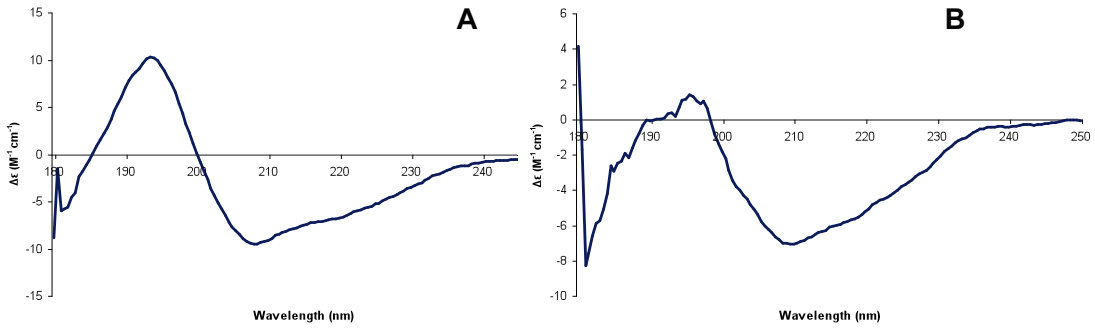

Figure 4.12: Far UV spectra for rSpy0125-NTR and CTR.

Panel A: rSpy0125-NTR spectrum displays a large α-helical component. **Panel B**: rSpy0125-CTR spectrum reveals a largely β-strand protein with a small α-helical component.

4.6 Construction of *Δspy0125-NTR* and *Δspy0125-ID* mutants

In close collaboration, colleagues in the laboratory used the three recombinant proteins produced in this chapter to generate specific anti-sera to each (as described in materials and methods, section 2.7), and the remaining intact subunits (Spy0128 and Spy0130). In experiments carried out by Dr Wendy Smith and Dr Emily Abbot, discussed here for clarity, it was revealed that only anti-sera specific to Spy0125 blocked the adhesion of wild-type M1 GAS strain SF370 to human tonsil to a significant extent. Further experiments were carried out with anti-sera raised against the two Spy0125 fragments identified in section 4.4 which found that only anti-sera specific to the CTR of Spy0125 blocked adhesion of wild-type bacteria. Intriguingly, sequence alignments of all Spy0125 homologues reveals that both the M5 and M18 GAS Spy0125 equivalents are completely missing the NTR as described above (see discussion). The signal peptides are present but the amino acid sequence starts in almost direct alignment with the start of the Spy0125-CTR as identified above in section 4.4. Other than missing this region, there is a high level of sequence homology between the three proteins which suggests either that the NTR is not involved in binding to the target receptor at all or that the M5 and M18 GAS strains utilise other means for initial attachment to host cells. In order to examine this, it was decided to create two deletion mutants within *spy0125* on the M1 GAS strain SF370 chromosome. The first would delete the entire coding region for the Spy0125-NTR described above while the second would be an internal deletion (ID) lacking central sequences.

The mutant alleles were designed and constructed with a colleague in the laboratory (Dr. Wendy Smith), who produced the recombinant plasmids pWS031 and pWS032 carrying the *Δspy0125-NTR* and *Δspy0125-ID* deletions, essentially as outlined in Figure 4.13. PCR primers were designed to delete DNA encoding Leu53-Pro285 inclusive, for the Spy0125-NTR deletion (*Δspy0125-NTR*) and Asn286-Pro559 inclusive, for the Spy0125-ID deletion (*Δspy0125-ID*), whilst being careful not to disrupt the reading frame (panel C, Figure 4.8). In designing primers to construct the *Δspy0125-NTR*, care was taken to maintain the bases encoding the predicted signal peptide and the following five amino acids, to minimise any potential secretion and processing errors. The exact sequences deleted from *spy0125* are shown in Figure 4.14

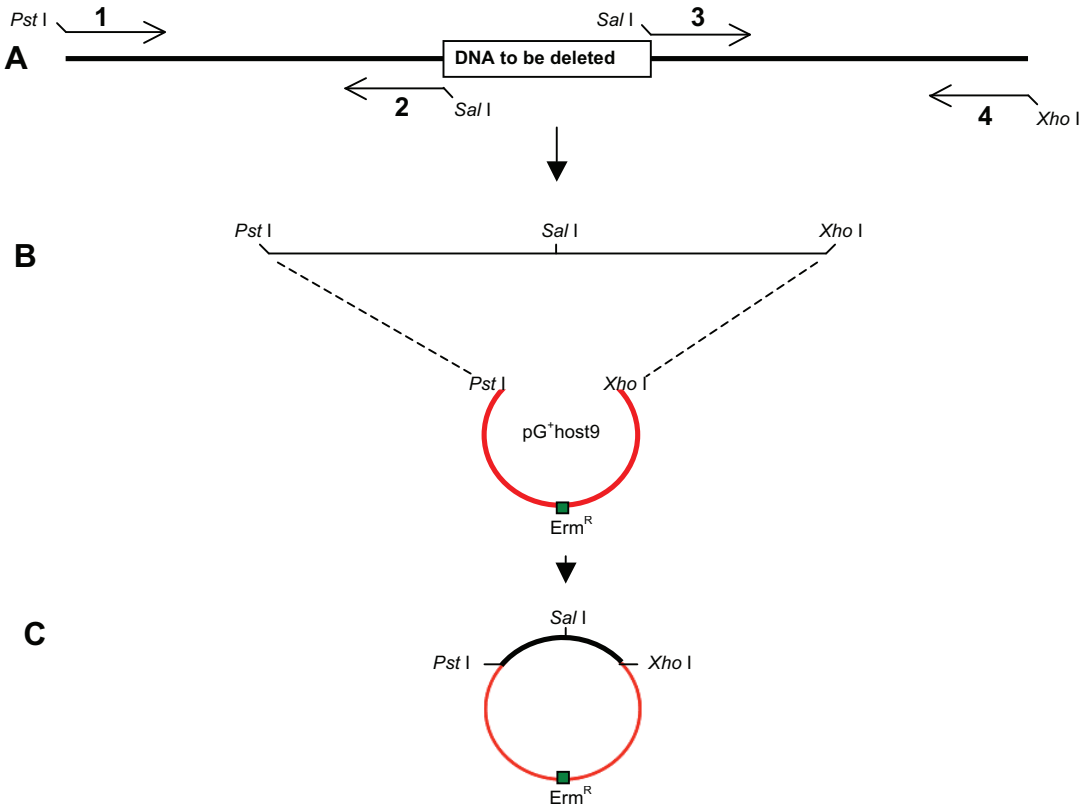

Figure 4.13: Schematic showing construction of recombinant pG$^+$host9 plasmid used in allele replacement to produce targeted deletions in the M1 GAS strain SF370 chromosome.

Panel A: The two sets of primers 1 + 2 and 3 + 4 were used to amplify regions of flanking DNA to the target to be deleted. **Panel B**: The amplified fragments were digested with *Sal*I, whose sites were contained within primers 2 and 3, and ligated to form the full length mutant allele. The purified product of this ligation was used as template in a further PCR reaction before being purified again and cleaved by *Pst*I and *Xho*I whose sites were contained within primers 1 and 4. **Panel C**: The cleaved full length fragment was subsequently ligated into the pG$^+$host9 vector which had been digested with the same two enzymes to construct the recombinant plasmid.

(*Δspy0125-NTR*) and Figure 4.15 (*Δspy0125-ID*) along with flanking regions and the locations of all primers used to create the mutant alleles. Following confirmation by DNA sequencing, these plasmids were then used by me to introduce the mutations into the M1 GAS strain SF370 chromosome by allele replacement mutagenesis.

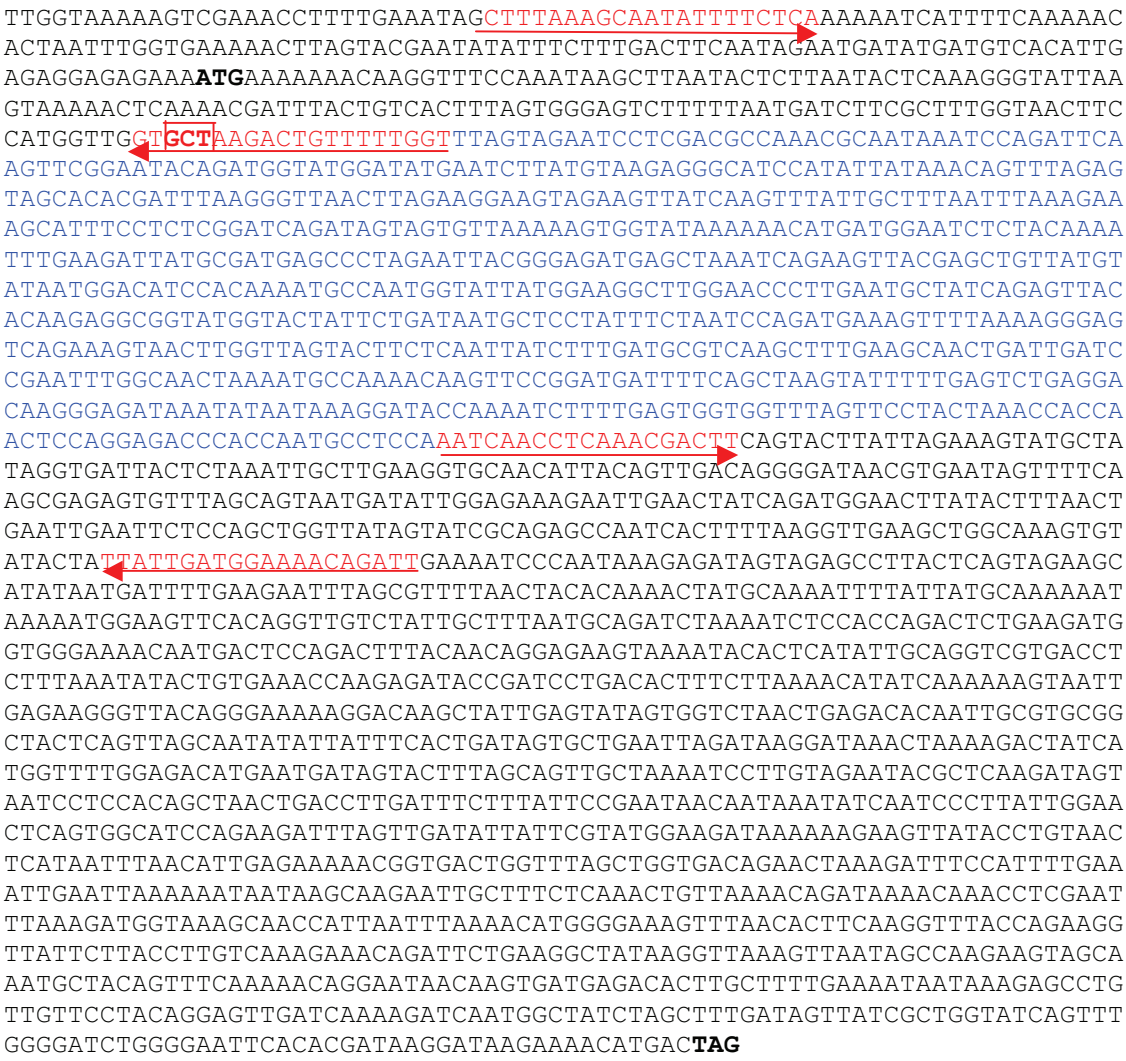

Figure 4.14: Region of DNA sequence deleted illustrating the construction of *Δspy0125-NTR*.

The entire *spy0125* coding region is shown with start and stop codons highlighted in bold. DNA flanking regions retained in the chromosome are shown in black. Primers used to construct the deletion are shown in red with their direction indicated by arrows under the sequence. The Ala codon highlighted and in bold type (**GCT**) codes for the predicted final alanine residue of the signal peptide. It was decided to keep five extra residues after the signal peptide to ensure proper processing of the new protein. Primer sequences were retained in the chromosome. DNA sequence deleted from *spy0125* is shown in blue.

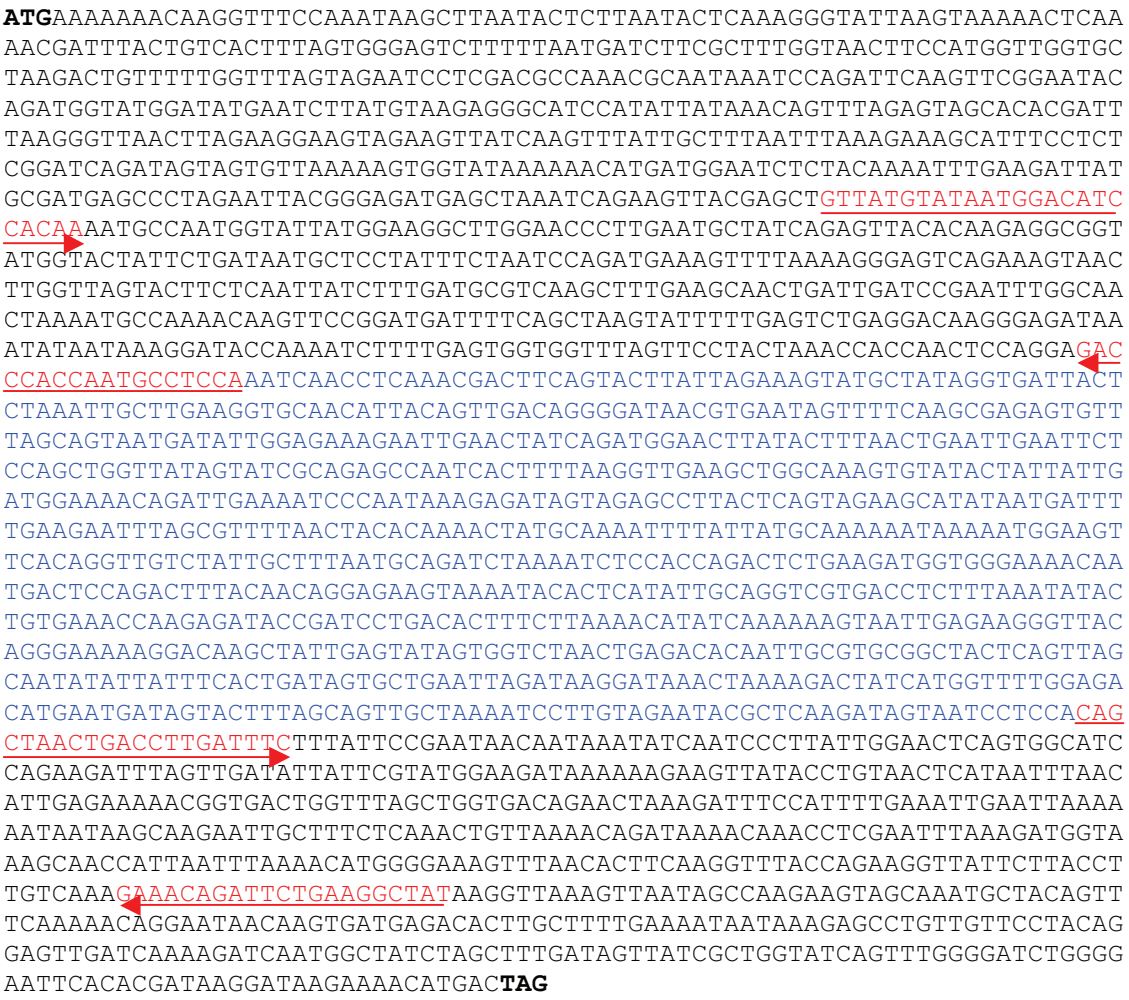

Figure 4.15: Region of DNA sequence deleted illustrating the construction of *Δspy0125-ID*.

The entire *spy0125* coding region is shown with start and stop codons highlighted in bold. DNA flanking regions retained in the chromosome are shown in black. Primers used to construct the deletion are shown in red with their direction indicated by arrows under the sequence. Primer sequences were retained in the chromosome. DNA sequence deleted from *spy0125* is shown in blue.

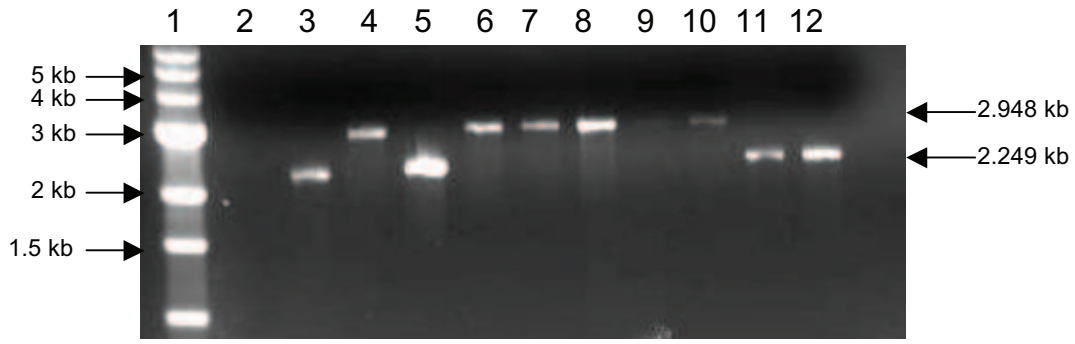

Figure 4.16: Screening of Erm^{s} *Δspy0125-NTR* colonies by PCR.

Erythromycin sensitive colonies of wild-type M1 GAS strain SF370 which had been subject to allele replacement using the pWS031 plasmid were screened by PCR and their products analysed on 1% agarose gel. Lane 1: DNA size standards. Lane 2: negative control containing no template DNA. Lanes 3-12: PCR products from different screened colonies. The larger DNA product at 2.948 kb represents the restored wild-type allele whereas in lanes 3, 5, 11 and 12 the smaller DNA product of 2.249 kb indicates that allele replacement has deleted the *NTR* of *spy0125*. Similar results were obtained for the *Δspy0125-ID*.

Plasmids pWS031 and pWS032 were introduced into wild-type M1 GAS strain SF370 by electrotransformation and allele replacement was used to delete the defined DNA sequences outlined above, as described in chapter 2, section 2.5.14. After the second cross over event either the desired DNA is deleted or the wild-type allele can be restored depending on which sequences have recombined during excision of the plasmid (see chapter 2, Figure 2.4). To distinguish between the two different outcomes, the chromosomal DNA of 10 erythromycin sensitive colonies was recovered and used as template in PCR reactions using primers outside the sequences that had originally been cloned to produce pWS031 and pWS032. The resulting products were examined by agarose gel electrophoresis. The two different outcomes are readily identifiable in Figure 4.16 for *Δspy0125-NTR*. The larger products running at 2.948 kb are where the wild-type allele had been restored while the smaller products running at 2.249 kb are where the *spy0125-NTR* had been deleted. Similar results were obtained while screening for potential *Δspy0125-ID* mutants. The PCR products of interest were purified and confirmed by DNA sequencing. The final DNA sequence remaining in the chromosome after allele replacement for *Δspy0125-NTR* is shown in Figure 4.17 and for *Δspy0125-ID* in Figure 4.18.

TTGGTAAAAAGTCGAAACCTTTTGAAATAGCTTTAAAGCAATATTTTCTCAAAAATCATTTTCAAAAAC
ACTAATTTGGTGAAAAACTTAGTACGAATATATTTCTTTGACTTCAATAGAATGATATGATGTCACATTG
AGAGGAGAGAAA**ATG**AAAAAAACAAGGTTTCCAAATAAGCTTAATACTCTTAATACTCAAAGGGTATTAA
GTAAAAACTCAAAACGATTTACTGTCACTTTAGTGGGAGTCTTTTTAATGATCTTCGCTTTGGTAACTTC
CATGGTTGGT**GCT**AAGACTGTTTTTGGT**GTCGAC**AATCAACCTCAAACGACTTCAGTACTTATTAGAAAG
TATGCTATAGGTGATTACTCTAAATTGCTTGAAGGTGCAACATTACAGTTGACAGGGGATAACGTGAATA
GTTTTCAAGCGAGAGTGTTTAGCAGTAATGATATTGGAGAAAGAATTGAACTATCAGATGGAACTTATAC
TTTAACTGAATTGAATTCTCCAGCTGGTTATAGTATCGCAGAGCCAATCACTTTTAAGGTTGAAGCTGGC
AAAGTGTATACTATTATTGATGGAAAACAGATTGAAAATCCCAATAAAGAGATAGTAGAGCCTTACTCAG
TAGAAGCATATAATGATTTTGAAGAATTTAGCGTTTTAACTACACAAAACTATGCAAAATTTTATTATGC
AAAAAATAAAAATGGAAGTTCACAGGTTGTCTATTGCTTTAATGCAGATCTAAAATCTCCACCAGACTCT
GAAGATGGTGGGAAAACAATGACTCCAGACTTTACAACAGGAGAAGTAAAATACACTCATATTGCAGGTC
GTGACCTCTTTAAATATACTGTGAAACCAAGAGATACCGATCCTGACACTTTCTTAAAACATATCAAAAA
AGTAATTGAGAAGGGTTACAGGGAAAAAGGACAAGCTATTGAGTATAGTGGTCTAACTGAGACACAATTG
CGTGCGGCTACTCAGTTAGCAATATATTATTTCACTGATAGTGCTGAATTAGATAAGGATAAACTAAAAG
ACTATCATGGTTTTGGAGACATGAATGATAGTACTTTAGCAGTTGCTAAAATCCTTGTAGAATACGCTCA
AGATAGTAATCCTCCACAGCTAACTGACCTTGATTTCTTTATTCCGAATAACAATAAATATCAATCCCTT
ATTGGAACTCAGTGGCATCCAGAAGATTTAGTTGATATTATTCGTATGGAAGATAAAAAAGAAGTTATAC
CTGTAACTCATAATTTAACATTGAGAAAAACGGTGACTGGTTTAGCTGGTGACAGAACTAAAGATTTCCA
TTTTGAAATTGAATTAAAAAATAATAAGCAAGAATTGCTTTCTCAAACTGTTAAAACAGATAAAACAAAC
CTCGAATTTAAAGATGGTAAAGCAACCATTAATTTAAAACATGGGGAAAGTTTAACACTTCAAGGTTTAC
CAGAAGGTTATTCTTACCTTGTCAAAGAAACAGATTCTGAAGGCTATAAGGTTAAAGTTAATAGCCAAGA
AGTAGCAAATGCTACAGTTTCAAAAACAGGAATAACAAGTGATGAGACACTTGCTTTTGAAAATAATAAA
GAGCCTGTTGTTCCTACAGGAGTTGATCAAAAGATCAATGGCTATCTAGCTTTGATAGTTATCGCTGGTA
TCAGTTTGGGGATCTGGGGAATTCACACGATAAGGATAAGAAAACATGAC**TAG**

Figure 4.17: Nucleotide sequence of *Δspy0125-NTR* mutant strain of M1 GAS strain SF370.

The nucleotide sequence of the chromosome is shown after the *NTR* of *spy0125* has been deleted. The exact same labelling is used as in Figure 4.14; chromosomal sequences are shown in black including start and stop codons highlighted in bold, primer sequences used to create the mutant are shown in red with the predicted final alanine of the signal peptide is highlighted and in bold type (**GCT**). Located in between the two internal primers highlighted in blue and underlined is the *Sal*I site.

ATGAAAAAAACAAGGTTTCCAAATAAGCTTAATACTCTTAATACTCAAAGGGTATTAAGTAAAAACTCAA
AACGATTTACTGTCACTTTAGTGGGAGTCTTTTTAATGATCTTCGCTTTGGTAACTTCCATGGTTGGTGC
TAAGACTGTTTTTGGTTTAGTAGAATCCTCGACGCCAAACGCAATAAATCCAGATTCAAGTTCGGAATAC
AGATGGTATGGATATGAATCTTATGTAAGAGGGCATCCATATTATAAACAGTTTAGAGTAGCACACGATT
TAAGGGTTAACTTAGAAGGAAGTAGAAGTTATCAAGTTTATTGCTTTAATTTAAAGAAAGCATTTCCTCT
CGGATCAGATAGTAGTGTTAAAAAGTGGTATAAAAAACATGATGGAATCTCTACAAAATTTGAAGATTAT
GCGATGAGCCCTAGAATTACGGGAGATGAGCTAAATCAGAAGTTACGAGCTGTTATGTATAATGGACATC
CACAAAATGCCAATGGTATTATGGAAGGCTTGGAACCCTTGAATGCTATCAGAGTTACACAAGAGGCGGT
ATGGTACTATTCTGATAATGCTCCTATTTCTAATCCAGATGAAAGTTTTAAAAGGGAGTCAGAAAGTAAC
TTGGTTAGTACTTCTCAATTATCTTTGATGCGTCAAGCTTTGAAGCAACTGATTGATCCGAATTTGGCAA
CTAAAATGCCAAAACAAGTTCCGGATGATTTTCAGCTAAGTATTTTTGAGTCTGAGGACAAGGGAGATAA
ATATAATAAAGGATACCAAAATCTTTTGAGTGGTGGTTTAGTTCCTACTAAACCACCAACTCCAGGAGAC
CCACCAATGCCTCCA**GTCGAC**CAGCTAACTGACCTTGATTTCTTTATTCCGAATAACAATAAATATCAAT
CCCTTATTGGAACTCAGTGGCATCCAGAAGATTTAGTTGATATTATTCGTATGGAAGATAAAAAAGAAGT
TATACCTGTAACTCATAATTTAACATTGAGAAAAACGGTGACTGGTTTAGCTGGTGACAGAACTAAAGAT
TTCCATTTTGAAATTGAATTAAAAAATAATAAGCAAGAATTGCTTTCTCAAACTGTTAAAACAGATAAAA
CAAACCTCGAATTTAAAGATGGTAAAGCAACCATTAATTTAAAACATGGGGAAAGTTTAACACTTCAAGG
TTTACCAGAAGGTTATTCTTACCTTGTCAAAGAAACAGATTCTGAAGGCTATAAGGTTAAAGTTAATAGC
CAAGAAGTAGCAAATGCTACAGTTTCAAAAACAGGAATAACAAGTGATGAGACACTTGCTTTTGAAAATA
ATAAAGAGCCTGTTGTTCCTACAGGAGTTGATCAAAAGATCAATGGCTATCTAGCTTTGATAGTTATCGC
TGGTATCAGTTTGGGGATCTGGGGAATTCACACGATAAGGATAAGAAAACATGAC**TAG**

Figure 4.18 Nucleotide sequence of *Δspy0125-ID* mutant strain of M1 GAS strain SF370.

The nucleotide sequence of the chromosome is shown after an internal portion of *spy0125* has been deleted. The exact same labelling is used as in Figure 4.15; chromosomal sequences are shown in black including start and stop codons highlighted in bold. Primer sequences used to create the mutant are shown in red and in between the two internal primers highlighted in blue and underlined is the *Sal*I site.

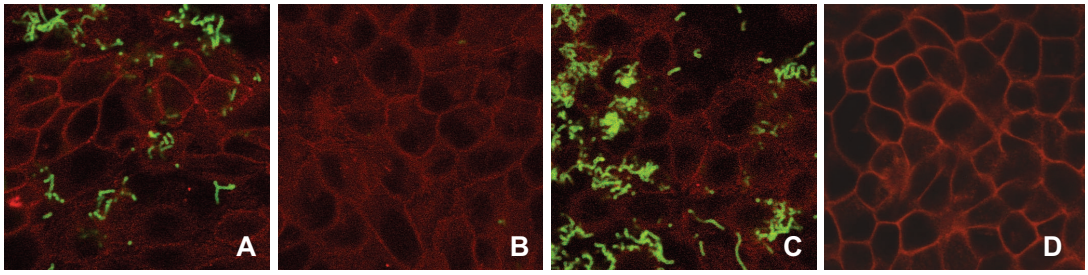

Figure 4.19: Adhesion of wild-type M1 GAS strain SF370 and *spy0125* mutants to HaCaT cells.

Samples were examined using laser scanning confocal microscopy and PMT values for each excitation channel were recorded. HaCaT cells are labelled red and bacteria are stained green. **Panel A**: wild-type M1 GAS strain SF370. **Panel B**: *Δspy0125*. **Panel C**: *Δspy0125-NTR*. **Panel D**: *Δspy0125-ID*.

4.7 Analysis of *Δspy0125-NTR* and *Δspy0125-ID* mutants

The two M1 GAS strain SF370 *spy0125* mutants were constructed to localise the region of Spy0125 responsible for binding to host cells. For adhesion assays the *Δspy0125-NTR* and *Δspy0125-ID* mutants were cultured in parallel with two control strains; wild-type M1 GAS strain SF370 and the *Δspy0125* mutant which had been constructed previously in the laboratory (Abbot *et al*., 2007). The four strains were grown to mid-log phase and added to confluent HaCaT cells on glass coverslips essentially as described in chapter 2, section 2.12.2. Prior to staining, cells were blocked in rabbit serum before the actin cytoskeleton of the HaCaTs was labelled with phalloidin-TRITC (which fluoresces red) and the bacteria were stained using a polyclonal antibody to GAS with a FITC conjugate (which fluoresces green). The adhesion assays were repeated on three separate occasions with essentially the same results. It can quite clearly be seen in panel C, Figure 4.19, that the *Δspy0125-NTR* mutant still binds to HaCaT cells showing that the NTR of Spy0125 is not critical for binding. The *Δspy0125-ID* strain (panel D, Figure 4.19) exhibited a similar phenotype to that shown for the *Δspy0125* mutant (panel B, Figure 4.19), with no binding observed. The binding of the four strains was also analysed by assessing the percentage adherence of each strain compared to wild-type binding, which was considered to be 100%. Leica Qwin software was used to count the fluorescence on each image and the percentage adherence was averaged for the three repeat experiments. Figure 4.20 clearly shows that the *Δspy0125-NTR* mutant clearly still binds to HaCaT cells at wild-type levels while the *Δspy0125-ID* mutant

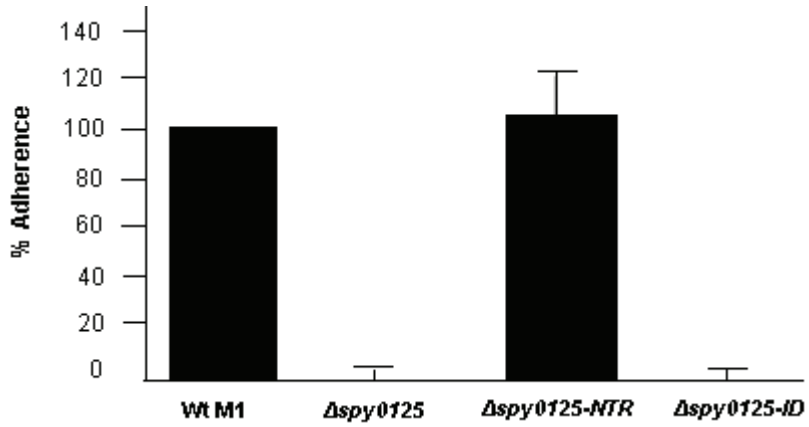

Figure 4.20: Binding of wild-type M1 and *spy0125* mutant strains to HaCaT cells.

Relative adhesion is shown with wild-type binding considered 100% and error bars indicating variation from three independent experiments. Adhesion of strains containing the *spy0125-NTR* deletion was at the same level as the wild-type, even slightly exceeding it. The adhesion of the *Δspy0125-ID* strain however was essentially the same as the *Δspy0125* mutant exhibiting no binding.

exhibits no binding, similar to the *Δspy0125* mutant. These results taken together suggest that the region of Spy0125 responsible for adhesion to host surfaces resides in the CTR of the molecule.

4.8 Localisation of Spy0125 at the pilus tip

As discussed in chapter 1 section 1.15, Kang *et al.* (Kang *et al*., 2007) confirmed that polymerisation of the major subunit, Spy0128, into pili in M1 GAS strain SF370 involves cleavage of its EVPTG motif between Thr311 and Gly312. This is followed by covalent linkage of the carboxyl group of Thr311 to the ε-amino group of Lys161 in the next Spy0128 monomer to be incorporated. This process is then repeated many times to produce the pilus. It was noted that unlike Spy0130 (which has a canonical LPXTG motif), Spy0125 has a similar CWSS (VVPTGV) to Spy0128 (EVPTGV). In addition, there is also a modest degree of homology between the C-terminal regions of Spy0125 and Spy0128 (37% identity). This suggested that Spy0125 could be recognised by SrtC and incorporated into the pilus *via* attachment to Lys161 of Spy0128. To test this prediction site directed mutagenesis and allele replacement mutagenesis were combined to create a Lys161Ala point mutation in *spy0128* in the M1 GAS strain SF370 chromosome (as described in chapter 3, section 3.5). Unlike section 4.6 where allele replacement mutagenesis was used to delete sections of *spy0125*, it was used here to

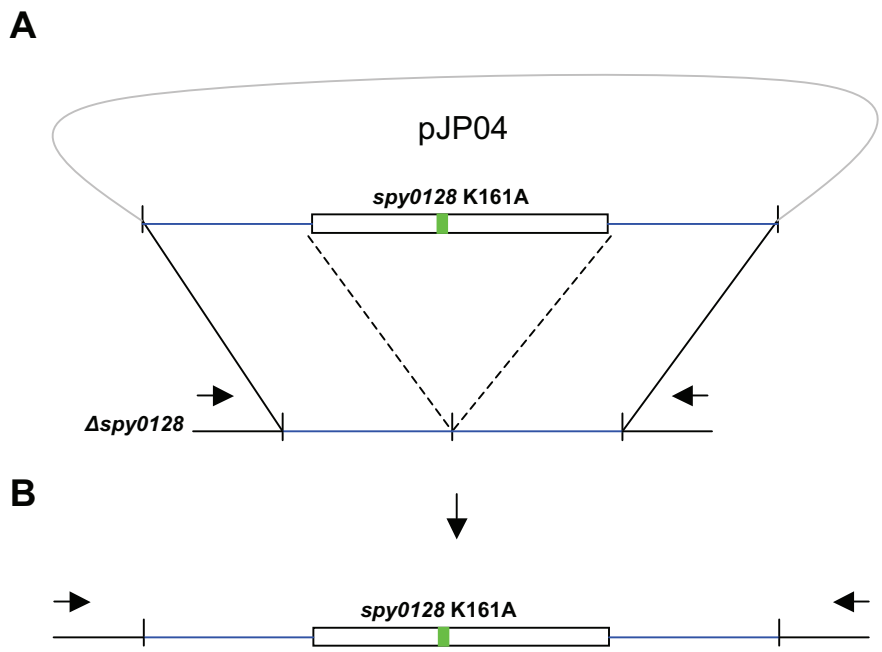

Figure 4.21 Schematic diagram showing construction of the *spy0128* K161A mutant in the M1 GAS strain SF370 chromosome.

Panel A: Previously, a *Δspy0128* strain had been constructed in the laboratory. **Panel B**: The pJP04 construct containing the *spy0128* K161A mutant allele was transformed into the *Δspy0128* strain and using allele replacement the *spy0128* gene containing the K161A mutation was replaced in the chromosome.

introduce the wild-type *spy0128* allele (control) and *spy0128* K161A mutant allele back into the chromosome of an *Δspy0128* strain of GAS which had been created previously in the laboratory. This process is outlined in the schematic diagram in Figure 4.21. The wild-type *spy0128* DNA sequence after allele replacement is shown in Figure 4.22 with the position of the Lys161 codon highlighted in yellow prior to mutagenesis to Ala (GCG).

To examine whether the *spy0128* K161A mutation had any effect on its attachment to Spy0125, cell wall fractions from the wild-type parent and *spy0128* K161A mutant were run on two 4-15% gradient SDS-PAGE gels. One was immuno-blotted with anti-rSpy0128 sera and the second with anti-rSpy0125 antibody. As subunits in Gram positive pili are covalently linked, when blotted with the anti-rSpy0128 antibody the wild-type M1 GAS strain SF370 pili can be observed as characteristic 'ladders' in the cell wall fraction, as seen in panel A, Figure 4.23 (lane 2), while no protein is detected

CGATATGGCTCCTAGTGTAAAAGCAGGAGATGCTATTTTATTTTATCGCTTATCTCAGACTTATAAAGTA
GAAGAAGCAGTTGTTTATGAGGACTCTAAAACATCTATAACGAAAGTCGGTCGGATTATTGCTCAAGCAG
GTGATGAAGTGGACCTGACAGAACAAGGTGAATTAAAAATTAATGGCCATATTCAAAATGAAGGACTAAC
TTTTATAAAATCAAGAGAAGCTAATTATCCGTATCGAATTGCTGATAATAGTTATTTGATACTCAATGAT
TACTATAGCCAAGAGAGTGAGAATTACCTACAAGATGCAATTGCTAAAGATGCTATAAAAGGCACGATTA
ATACACTTATTCGATTAAGAAACCATTAAGCTAAAGGCTTATTTAAAAAAGGAGAGAAACA**ATG**AAATTA
CGTCACTTACTATTAACGGGAGCAGCCCTAACTAGTTTTGCTGCTACAACAGTTCACGGGGAGACTGTTG
TAAACGGAGCCAAACTAACAGTTACAAAAAACCTTGATTTAGTTAATAGCAATGCATTAATTCCAAATAC
AGATTTTACATTTAAAATCGAACCTGATACTACTGTCAACGAAGACGGAAATAAGTTTAAAGGTGTAGCT
TTGAACACACCGATGACTAAAGTCACTTACACCAATTCAGATAAAGGTGGATCAAATACGAAAACTGCAG
AATTTGATTTTTCAGAAGTTACTTTTGAAAAACCAGGTGTTTATTATTACAAAGTAACTGAGGAGAAGAT
AGATAAAGTTCCTGGTGTTTCTTATGATACAACATCTTACACTGTTCAAGTTCATGTCTTGTGGAATGAA
GAGCAACAAAAACCAGTAGCTACTTATATTGTTGGTTATAAAGAAGGTAGTAAGGTGCCAATTCAGTTCA
AAAATAGCTTAGATTCTACTACATTAACGGTGAAGAAAAAAGTTTCAGGTACCGGTGGAGATCGCTCTAA
AGATTTTAATTTTGGTCTGACTTTAAAAGCAAATCAGTATTATAAGGCGTCAGAAAAAGTCATGATTGAG
AAGACAACTAAAGGTGGTCAAGCTCCTGTTCAAACAGAGGCTAGTATAGATCAACTCTATCATTTTACCT
TGAAAGATGGTGAATCAATCAAAGTCACAAATCTTCCAGTAGGTGTGGATTATGTTGTCACTGAAGACGA
TTACAAATCAGAAAAATATACAACCAACGTGGAAGTTAGTCCTCAAGATGGAGCTGTAAAAAATATCGCA
GGTAATTCAACTGAACAAGAGACATCTACTGATAAAGATATGACCATTACTTTTACAAATAAAAAAGACT
TTGAAGTGCCAACAGGAGTAGCAATGACTGTGGCACCATATATTGCTTTAGGAATTGTAGCAGTTGGTGG
AGCTCTTTACTTTGTTAAAAAGAAAAATGCT**TAA**ATTATTATTATGATAGTAAGACTGATTAAGCTCCTT
GACAAGTTGATAAACGTCATTGTTCTTTGTTTCTTCTTTCTTTGTTTATTGATTGCGGCACTTGGAATCT
ACGATGCTTTAACAGTTTATCAAGGAGCTAATGCTACTAACTATCAACAATATAAGAAAAAGGGTGTTCA
GTTTGACGATTTATTAGCTATTAATTCTGATGTTATGGCATGGCTGACTGTTAAAGGAACGCATATTGAT
TATCCAATTGTACAGGGAGAGAATAATTTAGAATATATCAACAAATCAGTAGAAGGAGAGTACTCCTTAT
CAGGAAGTGTTTTTCTAGATTATCGTAATAAAGTAACTTTTGAAGATAAATACTCATTAATCTATGCACA
TC

Figure 4.22: DNA sequence for wild-type *spy0128* used to construct the recombinant plasmid pJP01.

Coding sequences for *spy0128* are shown in blue with start and stop codons highlighted in bold. Highlighted in yellow is the Lys161 codon targeted for site directed mutagenesis to alanine. Adjacent regions of flanking DNA used to replace *spy0128* back into the chromosome are shown in black. Primer sequences used to amplify *spy0128* and clone into pG^{+}host9 are shown in red.

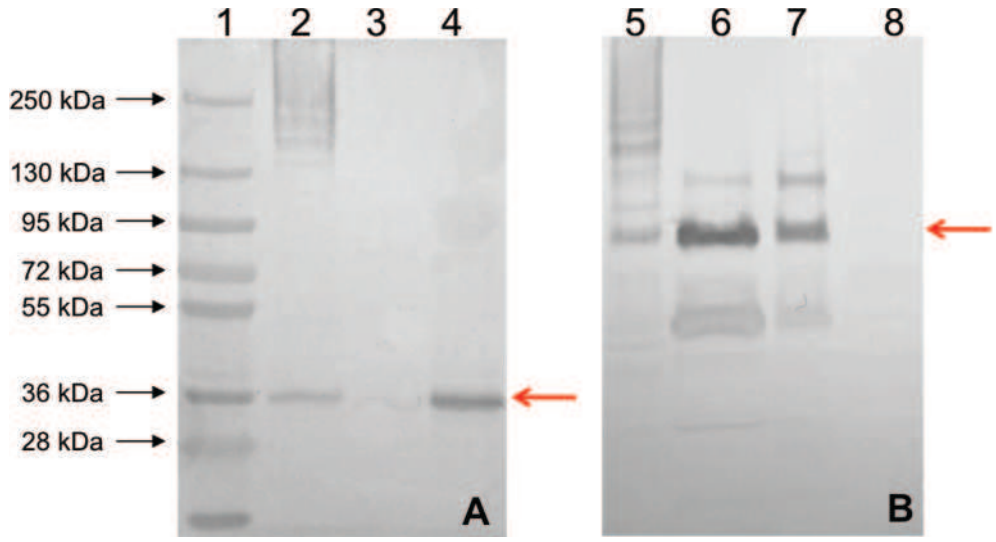

Figure 4.23: Immuno-blots of wild-type M1 and *spy0128* mutant cell wall fractions with anti-rSpy0128 and anti-rSpy0125 sera.

Lys161 of Spy0128 is needed for Spy0125 incorporation at the pilus tip. **Panel A**: Cell wall fractions immuno-blotted with anti-Spy0128 sera. Lane 1: molecular size standards. Lane 2: wild-type M1 GAS strain SF370. Lane 3: *Δspy0128*. Lane 4: *spy0128* K161A. **Panel B**: Cell wall fractions immuno-blotted with anti-Spy0125 sera. Lane 5: wild-type M1 GAS strain SF370. Lane 6: *Δspy0128*. Lane 7: *spy0128* K161A. Lane 8: *Δspy0125*. Red arrows indicate unincorporated Spy0128 (panel A) and Spy0125 (panel B) monomers.

in the *Δspy0128* mutant (lane 3). The *spy0128* K161A mutant in lane 4 is only seen to produce a single band indicating a build up of Spy0128 K161A monomer that is unable to polymerise into pili. When the cell wall fractions from the same three strains are blotted with an anti-rSpy0125 antibody, as seen in panel B of Figure 4.23, a distinctive 'ladder' pattern is observed in the wild-type M1 GAS strain SF370 indicative of polymerised pili with Spy0125 incorporated (lane 5). In cell wall samples for the *Δspy0128* (lane 6) and *spy0128* K161A (lane 7) mutants an intense band is detected at ~94 kDa representing monomeric Spy0125 (as outlined above, this protein runs slowly on SDS-PAGE) along with smaller faint bands that correspond to Spy0125 breakdown products and a single larger band. As the larger band at ~125 kDa is also present in the *Δspy0128* sample it therefore cannot correspond to a heterodimer of Spy0128-Spy0125. This band is also not observed in the blot with anti-rSpy0128. These data would suggest that the incorporation of Spy0125 into pili would be *via* a transpeptidation reaction linking the ε-amino group of Lys161 of Spy0128 to Thr723 of the Spy0125 VVPTG motif, positioning it solely at the pilus tip.

4.9 Discussion

The results presented in this chapter describe the Spy0125 protein, provide evidence that it is located at the tip of the pilus and that this protein is the adhesin, responsible for attachment of the bacteria to host cell surfaces.

Recombinant (r)Spy0125 was observed to breakdown into two stable fragments which were identified as being the N-terminal third of the protein and the remaining two thirds making up the C-terminal region. As mentioned at the end of section 4.4, a proline rich region separates the two isolated fragments and is situated directly upstream of the Asn – Gln – Pro – Gln – Thr – Thr sequence identified as the first six residues of the C-terminal fragment. This region contains the following residues (shown in parenthesis);

- Lys_{274} – [Pro – Pro – Thr – Pro – Gly – Asp – Pro – Pro – Met – Pro – Pro] – Asn_{286} –

Proline rich regions are known for being acid labile, especially Asp-Pro bonds (highlighted above in red) where the Asp side chain can cleave the peptide backbone (Li *et al*., 2009; Piszkiewicz *et al*., 1970). The C-terminal fragment, from Asn286 (determined as the N-terminal residue for this fragment) to Val725 (the final residue of rSpy0125) has a mass of 49354.1 Da while the intact mass obtained by FT-ICR MS was 49769.14 Da. This discrepancy could be accounted for by a variety of reasons. A number of the additional residues from the sequences shown above may have been present on the N-terminus of the larger breakdown fragment when its mass was analysed by FT-ICR MS but under the acidic conditions of the Edman chemistry during N-terminal sequencing these residues may have cleaved off, revealing Asn286 to be the first residue at the N-terminus. The sample of rSpy0125 containing breakdown products used for FT-ICR MS analysis may have harboured several molecules with different N-termini, also containing additional residues from the sequence shown above from the highly labile Pro rich region. This could lead to a determined mass being an average of several unresolved species with slightly varying ends, possibly formed as discussed above, or a post-translational modification, although the mass difference observed suggests that this was not the case. Also, the difference in mass observed here cannot be determined precisely by addition or subtraction of residues from the N or C terminus of this region. The reason for this discrepancy is not clear and was not explored further

however, as this project aimed to focus on protein function and the data was sufficient to show the larger breakdown fragment was composed of the C-terminal two thirds of Spy0125.

It has been shown here that deletion of the DNA encoding the N-terminal region of Spy0125 in the chromosome by allele replacement, does not abrogate binding of the bacteria to HaCaT cells. This would suggest the binding site is contained within the C-terminal 2/3 of Spy0125. This is supported by data from the *Δspy0125-ID* mutant where an internal portion of Spy0125 has been deleted and this strain fails to adhere to HaCaT cells. This result matches the binding phenotype of the *Δspy0125* mutant which also fails to adhere to cultured HaCaT cells or tonsil explants (Abbot *et al*., 2007). However, blots of cell wall fractions for the *Δspy0125-ID* mutant confirmed this variant was still incorporated into polymerised pili indicating that the region responsible for host receptor attachment could be situated within the residues deleted from Spy0125 in this mutant (Smith *et al*., 2010). In contrast, the residues removed in the *Δspy0125-NTR* mutant had no effect on the bacteria's ability to bind HaCaT cells, suggesting that this region is not needed for adhesion. Work done by colleagues in the laboratory, with antibodies specific to the NTR and CTR of Spy0125, supported this finding. When wild-type M1 GAS strain SF370 were pre-incubated with these sera prior to their addition to HaCaT cells, only anti-CTR sera blocked binding (Smith *et al*., 2010). Sequence analysis of Spy0125 homologues reveals that the M5 and M18 strain equivalents do not contain an NTR as discussed in this chapter for Spy0125 (see Figures 4.24 and 4.25). After the predicted signal peptide, these sequences align almost identically to the beginning of the Spy0125-CTR that is described here. This opens up a number of interesting possibilities. Perhaps the NTR contributed nothing to the binding of these proteins to host surfaces and was, therefore, lost during evolution or that pilus mediated adhesion is not the critical method of binding in these strains. It is also possible that the NTR of Spy0125 does play a role of some form in adhesion to targets not present on HaCaT cells and not examined here. The loss of this region for these strains would suggest they would need to utilise different methods of binding for host attachment. Work has been approached by others in the laboratory to determine the role played by pili in M5 GAS strain Manfredo but pili in this strain are quite sparse and anti-sera raised to individual subunits was quite poor. This lead to high background

binding on immuno-blots with these sera making results difficult to interpret. A potentially interesting experiment to test whether the M5 Spy0125 homologue (Spy0104) does indeed bind to host cells such as tonsil explants or HaCaT cells (it could be predicted it *would* mediate adhesion to these targets) would be to replace the *spy0125* gene in the M1 chromosome with the *spy0104* gene from M5 GAS strain Manfredo.

The second crucial result discussed in this chapter is that Spy0125 is shown to be positioned at the tip of the pilus. The major subunit, Spy0128, is polymerised into extended pili by the action of SrtC encoded by *spy0129*. This enzyme is responsible for forming the covalent bond between Thr311 of the CWSS and the free ε-amino group on the side chain of Lys161 of the next Spy0128 monomer to be incorporated into the pilus. The C-terminal regions of Spy0125 and Spy0128 show a reasonable level of sequence homology (37% identity between the two regions), suggestive of a similar domain structure. For this reason, when constructing the *Δspy0125-NTR* and *Δspy0125-ID* mutants, care was taken to include this region and the signal peptide with a few additional residues to ensure each subunit was incorporated into pili. This region may facilitate recognition of both proteins by SrtC and therefore attachment of Spy0125 to Spy0128 by this sortase. The western blots shown in this chapter of cell wall preparations from the wild-type M1 GAS strain SF370 and *spy0128* K161A mutant with anti-rSpy0128 sera show no polymerisation of Spy0128 in the K161A harbouring strain which is as expected. In the repeat blot with anti-Spy0125 sera (panel B, Figure 4.23), three bands are quite clearly seen. The band of smaller mass probably corresponds to the Spy0125-CTR breakdown product. The two bands of larger mass correspond to Spy0125 monomer and an as yet unidentified product (see below). These two bands are also present in the lane corresponding to the cell wall fraction of the *Δspy0128* mutant, suggesting the larger band cannot be a Spy0128-Spy0125 heterodimer. In the *spy0128* K161A mutant there is no polymerisation of Spy0125 subunits which implies that K161 of Spy0128 is needed for the attachment of Spy0125 via Thr723 of its CWSS. Smith *et al.* (Smith *et al*., 2010) confirmed this linkage by mass spectrometry using pili purified from the wild-type M1 bacteria by tryptic digest followed by LC MS/MS analysis.

As mentioned above, the third band with largest mass present on the cell wall blots performed using anti-Spy0125 sera (Figure 4.23) is as yet unidentified. It is tempting to

speculate that in the absence of Spy0128 monomers, or polymerised monomers at the cell surface, Spy0125 is attached covalently to Spy0130 and anchored to the cell wall. This is supported by the observation that on a duplicate immuno-blot containing the same cell wall fractions as above with anti-Spy0130 sera, a band of similar mass is detected at ~125 kDa which would suggest that this band contains Spy0125 joined to Spy0130. However, the anti rSpy0130 antibody used in this blot was quite poor and high background antibody binding was seen. Further experiments would be needed to confirm this potential Spy0130-Spy0125 hetero-dimer linkage, such as using mass spectrometry as outlined in Smith *et al*., 2010. In *S. pneumoniae*, a similar situation has been observed on immuno-blots, where the two minor subunits, RrgC and RrgA, were suggested to form a covalently linked hetero-dimer in the absence of the major subunit, RrgB as suggested here for Spy0130-Spy0125 (LeMieux *et al*., 2008).

Chapter 5. Crystallisation and X-ray analysis of Spy0125

5.1 Introduction

The work described in the previous chapter details the molecular characterisation of Spy0125. This analysis revealed Spy0125 to be extremely labile in solution and prone to degradation at a region defined by a proline rich sequence of residues separating the recombinant protein into roughly one and two thirds. These fragments were identified by mass spectrometry in conjunction with N-terminal sequencing and this allowed both to be re-cloned and expressed. Colleagues in the laboratory produced antibodies to each fragment and subsequent experiments revealed that only anti rSpy0125-CTR sera inhibited binding of wild-type M1 GAS strain SF370 to target cells. It was noted however that rSpy0125-NTR was unstable when stored at 4^{o}C so no further experiments were pursued with this fragment. Deletion of the DNA encoding the Spy0125-NTR led to no decrease in the binding levels of this mutant when compared to wild-type, which strongly suggests that the NTR of Spy0125 is not needed for adhesion. It was also noted that Spy0125 homologues from some other M serotypes of GAS were devoid of the NTR, suggesting that this region has been lost through evolution, possibly because it is not essential. A second mutant, deleting DNA encoding the middle part Spy0125 did not bind to the selected cell line and these data confirm the NTR plays little role in binding and that the crucial binding interface for the initial host-pathogen interaction is contained within the Spy0125-CTR. The major aim for this project was to probe the structures of the M1 GAS strain SF370 pili subunits. Work in this chapter focuses on the determination and analysis of the Spy0125-CTR X-ray crystal structure which reveals a number of exciting features, including evidence suggesting a potentially novel mechanism for bacterial adherence to human cells.

5.2 Crystallisation and X-ray analysis of rSpy0125-CTR

Recombinant (r)Spy0125-CTR purified in chapter 4 was used to set up crystallisation screens as described in chapter 2, section 2.11.1. Initial crystals of rSpy0125-CTR were obtained in 30% (w/v) PEG 5000 MME, 200 mM ammonium sulphate and 100 mM MES pH 6.5. Examples of these crystals are seen in panel A, Figure 5.1. Crystal growth was optimised by screening the concentration of precipitant against the pH of the buffer.

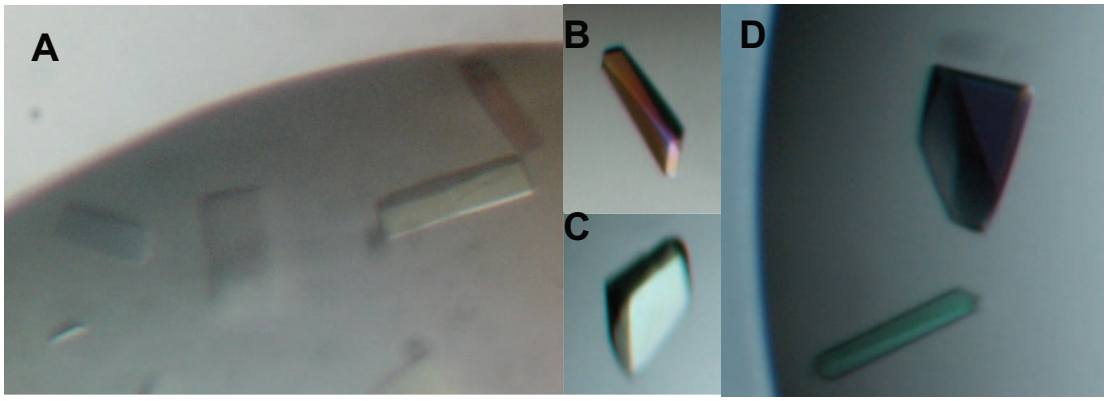

Figure 5.1: Examples of native rSpy0125-CTR crystals.

Panel A: Initial rSpy0125-CTR crystals. **Panels B-D**: Optimised rSpy0125-CTR crystals.

Large well formed crystals were grown in 34-36% (w/v) PEG 5000 MME, 100 mM ammonium sulphate and 100 mM MES pH 6.0-6.5 and examples of these crystals are shown in Figure 5.1, panels B-D. For data collection crystals were harvested in Litholoops and transferred to a cryo-protectant solution of paratone-N, to protect them from radiation damage, and flash frozen in liquid nitrogen. Initial crystals diffracted X-rays extremely poorly and although this was improved by growing larger crystals, the maximum resolution achieved was usually limited to ~3 Å. A complete diffraction dataset was collected from a single cryo-frozen rSpy0125-CTR crystal on beamline IO3 at Diamond Light Source (DLS, Oxford, UK), consisting of 600 images with an oscillation angle of 0.75^{o} on an ADSC Q315 CCD detector, diffracting to a maximum resolution of 2.65 Å. The data was processed using iMOSFLM (Leslie, 2006) then scaled and reduced with SCALA (Evans, 2006). The data collection statistics are shown in Table 5.1. These crystals belonged to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions of a = 45.81 Å, b = 117.50 Å, c = 177.75 Å and $\alpha = \beta = \gamma = 90^{o}$.

5.3 Isomorphous replacement of rSpy0125-CTR

As discussed in chapter 2, section 2.11.4.1, the isomorphous replacement technique can be used to extract phase information. Native PAGE was used to screen for interactions between a large number of compounds and rSpy0125-CTR without having to sacrifice crystals as outlined in chapter 2, section 2.11.4.1.1 (Boggon and Shappiro 2000). rSpy0125-CTR was incubated with many different heavy atom salt solutions on ice for

Table 5.1: Summary of data collection and processing statistics for native rSpy0125-CTR crystals.

	Native rSpy0125-CTR
Wavelength (Å)	0.970
Space Group	$P2_12_12_1$
Resolution range (Å)[a]	55.00 – 2.65 (2.79 – 2.65)
Unit cell parameters (Å)	a = 45.81, b = 117.50, c = 177.75
No. of unique reflections	28853 (4136)
Multiplicity[b]	16.7 (17.3)
I/σ(I)	18.6 (8.2)
Completeness (%)[b]	99.9 (100.0)
R_{merge} (%)[c]	10.7 (41.1)

[a] Values in parenthesis represent figures for the highest resolution shell.

[b] Multiplicity and completeness. Given value is for merged data. Values in parenthesis are for the highest resolution shell.

[c] Reflection statistics are as reported in SCALA. R_{merge} is calculated as described in Evans, 2006.

10 min and mixtures were analysed by native PAGE (data not shown). Six compounds; Cl_4K_2Pt, $KAu(CN)_2$, $K_2Pt(NO_2)_4$, $KAuCl_4$, $C_9H_9HgNaO_2S$ and $HoCl_3.6H_2O$ showed band shifts compared to native rSpy0125-CTR suggestive of an interaction and these compounds were pursued in soaking experiments. Following soaking, crystals were bathed in a cryo-protectant solution of paratone-N and flash frozen in liquid nitrogen as before. Data was collected on a number of different derivatives although no heavy atom sites could be resolved, indicating that the crystals had not been derivatised or that binding sites were disordered.

5.4 Expression and purification of selenomethionine labelled rSpy0125-CTR

Currently the most widely used method of introducing heavy atom scatterers into protein crystals is to incorporate them during protein production. This is routinely achieved using an *E. coli* methionine auxotroph expression strain, grown in media

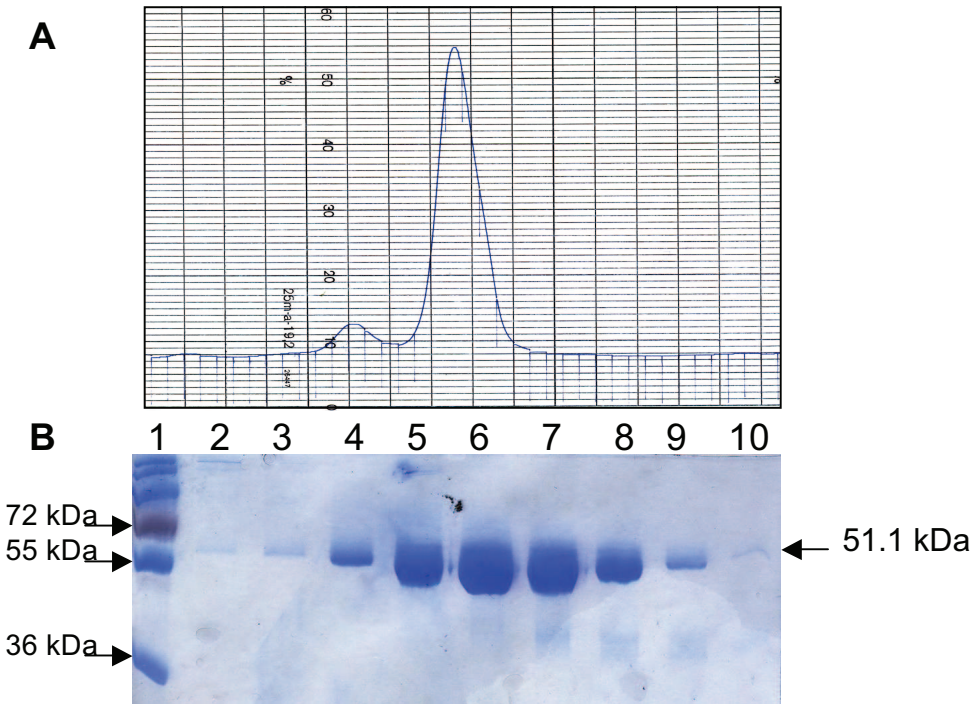

Figure 5.2: Elution profile and peak fraction analysis of rSpy0125-CTR SeMet purification.

Panel A: rSpy0125-CTR SeMet elution profile at 280 nm. **Panel B**: Lane 1 contains molecular size standards. Lanes 2-10: Coomassie blue stained gel of corresponding major peak fractions.

supplemented with all L-amino acids and L-selenomethionine in place of L-methionine. As described in chapter 2, section 2.6.4, B834 (DE3) methionine auxotroph cells were transformed with the pJP08 plasmid and grown in LB broth prior to being washed and resuspended in minimal media. Prior to the addition of L-selenomethionine a sample of culture was harvested and this control was incubated alongside the main culture. During incubation, the cell density of the main culture increased as expected while the control culture did not, confirming growth depended on incorporation of L-selenomethionine. Cells were harvested and lysed by sonication in buffer containing reducing agent to ensure incorporated selenium remained reduced prior to recovery of soluble His-tagged rSpy0125-CTR selenomethionine (SeMet) labelled protein, using Ni^{2+} affinity chromatography. Peak fractions were analysed by SDS-PAGE and those containing large quantities of protein were concentrated and applied to a Hi-Load 16/60 Superdex 200 gel filtration column to remove impurities. The elution profile for rSpy0125-CTR SeMet is shown in panel A, Figure 5.2 revealing a large predominant peak and fractions

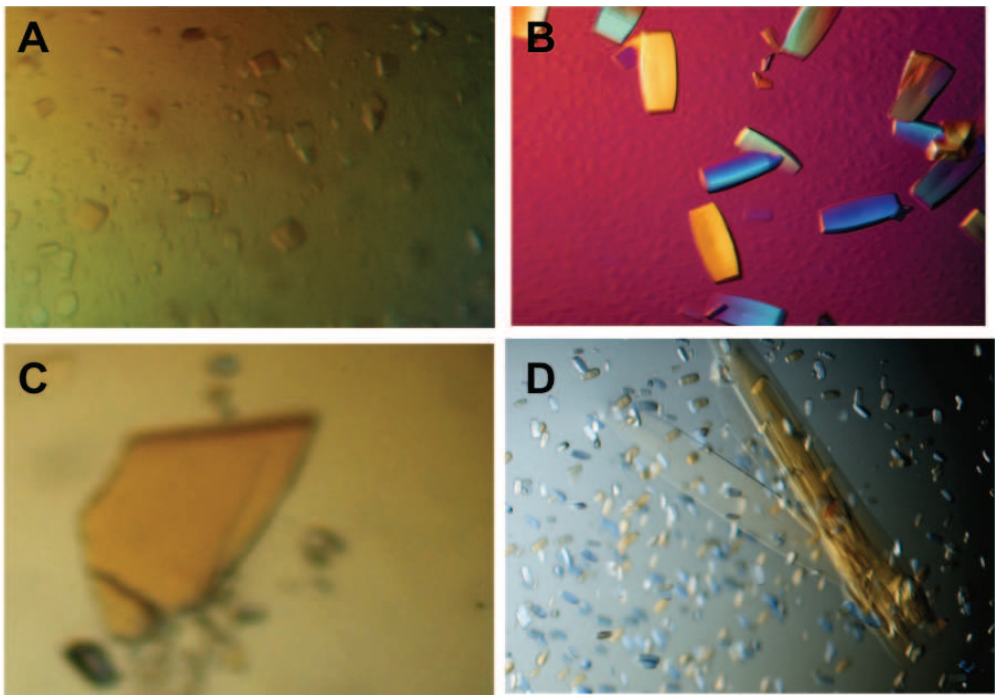

Figure 5.3: Examples of rSpy0125-CTR SeMet crystals.

Panel A: Initial crystals. **Panel B:** Large well formed crystals. **Panel C**: Thin 'plate' like crystals. **Panel D**: Both rSpy0125-CTR SeMet crystal morphologies growing in the same drop.

corresponding to this peak were analysed by SDS-PAGE (panel B, Figure 5.2). Fractions containing protein (lanes 4-9, Figure 5.2) were pooled and concentrated to 10 mg/ml, determined by absorbance at 280 nm using the molar extinction coefficient for rSpy0125-CTR of 39770 M^{-1} cm^{-1}. For use in crystallisation experiments the His-tag was removed by digestion with thrombin, followed by removal of the separated His-tag (on a His-chelation column) and thrombin (on a benzamidine column) as described in the material and methods, section 2.6.7. The digested rSpy0125-CTR SeMet was dialysed into gel filtration buffer prior to concentrating as described above.

5.5 Crystallisation of rSpy0125-CTR SeMet

Initial crystals were obtained for rSpy0125-CTR SeMet as described for native rSpy0125-CTR and were well formed but very small. The most encouraging crystals were obtained in 30% (w/v) PEG 4000, 200 mM ammonium acetate and 100 mM *tri-*

sodium citrate pH 5.6. Examples of these crystals are shown in panel A, Figure 5.3. Using the hanging drop vapour diffusion technique, screening PEG concentration against pH of the buffer, optimised crystals were obtained in 38% (w/v) PEG 4000, 200 mM ammonium acetate and 100 mM *tri*-sodium citrate pH 5.6. This condition routinely gave rise to large, well formed crystals (panel B, Figure 5.3) but in a small number of wells, crystals of different morphology were observed (panel C, Figure 5.3). These 'plate-like' crystals were extremely thin and fragile and the two different crystal forms would occasionally grow in the same drop as seen in panel D, Figure 5.3. For data collection, intact well formed crystals (such as those shown in panel B, Figure 5.3) were harvested in a litho-loop, prior to bathing in a cryo-protectant solution of paratone-N oil and flash frozen in liquid nitrogen. Thin 'plate' like crystals were broken into pieces and a single fragment harvested and frozen in the same manner.

5.6 rSpy0125-CTR SeMet data collection

The data used to solve the rSpy0125-CTR structure was collected on BM14 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France), using the selenomethionine labelled protein crystals described above. Multi-wavelength Anomalous Diffraction (MAD) was chosen for phasing to optimise the anomalous signal as only three methionines are encoded in the primary Spy0125-CTR sequence of 438 residues. Assuming there was full incorporation of SeMet into the protein this would represent a weak anomalous signal (common theory suggests at least one methionine should be present per 100 residues or 10 kDa). Three wavelength MAD data was collected on a single cryo-cooled crystal for both crystal morphologies, after excitation spectrums had been measured for each using a Bruker xFLASH SDD detector and maximised *f''* and *f'* values were determined by CHOOCH (G. Evans, included in the CCP4 suite) to ascertain optimised wavelengths. The experimentally determined excitation spectrum for a thin 'plate' like crystal is shown in Figure 5.4 and a dataset was collected at each wavelength (as determined by CHOOCH) to 2.1 Å, consisting of the same 300 images with an oscillation range of 1^{o} on a MAR 225 CCD detector. Each dataset was processed in iMOSFLM (Leslie, 2006), then scaled and reduced using SCALA (Evans, 2006). The data collection statistics are shown in Table 5.2. An additional redundant dataset was collected on this crystal at the peak wavelength, consisting of 650 images with an oscillation angle of 1^{o}. The data was

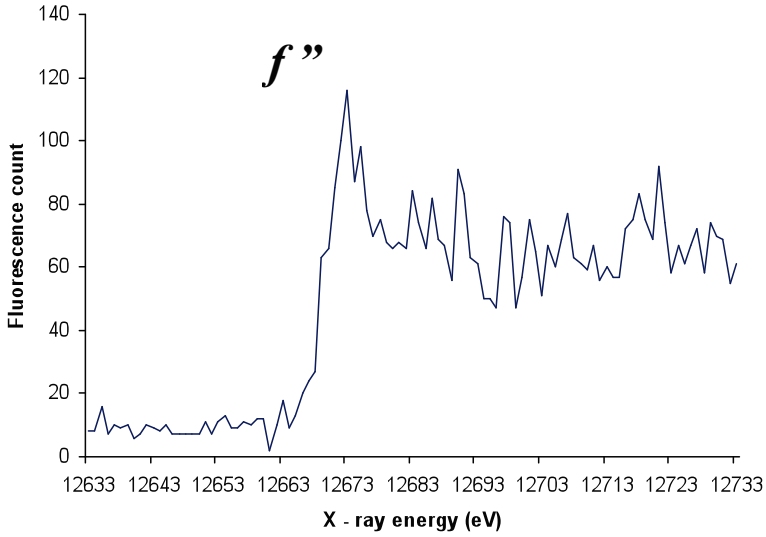

Figure 5.4: Excitation spectrum at the *K* edge of selenium for a thin 'plate' like rSpy0125-CTR SeMet crystal.

processed and scaled as above to a resolution of 1.9 Å and the statistics for this dataset are shown in Table 5.2 (Peak '2'). The protein crystallised in the triclinic space group P1, with unit cell dimensions that remained isomorphous throughout data collection (Table 5.2). Unit cell volume calculations were consistent with 2 molecules being present in the asymmetric unit.

The large well formed crystals (panel B, Figure 5.3) diffracted to a maximum resolution of 2.9 Å and a dataset was collected at each MAD wavelength (as determined by CHOOCH) consisting of the same 180 images with an oscillation range of 1^{o} on a MAR 225 CCD detector. Each dataset was processed and scaled as above and the statistics are listed in Table 5.3. These crystals belonged to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions that remained isomorphous throughout the data collection (shown in Table 5.3) and $\alpha = \beta = \gamma = 90^{o}$. Unit cell volume calculations were consistent with 2 molecules being present in the asymmetric unit.

Following data collection it was clear that the two different crystal morphologies belonged to two distinct space groups. Therefore at this stage, the thin 'plate' like (P1) crystals were renamed A-form crystals and the large well formed ($P2_12_12_1$) crystals as B-form crystals.

Table 5.2: Summary of data collection and processing statistics for rSpy0125-CTR SeMet A-form (P1) crystals.

	Peak	Inflection	Remote	Peak '2'
Wavelength (Å)	0.978	0.979	0.975	0.978
Space Group	P1	P1	P1	P1
Resolution range (Å)[a]	38.47 – 2.10 (2.21 – 2.10)	38.45 – 2.10 (2.21 – 2.10)	38.50 – 2.10 (2.21 – 2.10)	38.00 – 1.90 (2.00 – 1.90)
Unit cell parameters (Å, °)	*a* = 35.43 *b* = 66.28, *c* = 101.75, α = 86.19, β = 90.04, γ = 75.02	*a* = 35.42 *b* = 66.26, *c* = 101.69, α = 86.20, β = 90.06, γ = 75.06	*a* = 35.42 *b* = 66.27, *c* = 101.63, α = 86.36, β = 90.12, γ = 75.09	*a* = 35.48 *b* = 66.46, *c* = 101.80, α = 86.06, β = 90.06, γ = 75.08
No. of unique reflections	50476 (7366)	50433 (7361)	50458 (7354)	66113 (7725)
Multiplicity[b]	3.3 (3.3), 1.7 (1.7)	3.3 (3.3), 1.6 (1.6)	3.3 (3.3), 1.7 (1.7)	7.0 (6.7)
I/σ(I)	12.5 (7.2)	16.4 (7.7)	14.0 (6.9)	15.4 (5.0)
Completeness (%)[b]	97.0 (96.0), 96.0 (95.1)	97.0 (96.0), 96.1 (95.2)	97.0 (96.1), 96.1 (95.1)	93.6 (74.9)
R_{merge} (%)[c]	4.9 (11.5)	3.9 (10.9)	4.2 (12.2)	7.4 (31.8)
No. of sites (SOLVE)		6		
FOM (SOLVE/RESOLVE)		0.33/0.67		

[a] Values in parenthesis represent figures for the highest resolution shell.

[b] Multiplicity and completeness. First given value is for merged data and the second represents separated anomalous pairs. Values in parenthesis are for the highest resolution shell.

[c] Reflection statistics are as reported in SCALA. R_{merge} is calculated as described in Evans, 2006.

Table 5.3: Summary of data collection and processing statistics for rSpy0125-CTR SeMet B-form ($P2_12_12_1$) crystals.

	Peak	Inflection	Remote
Wavelength (Å)	0.978	0.979	0.975
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution range (Å)[a]	56.66 – 2.90 (3.06 – 2.90)	56.71 – 2.90 (3.06 – 2.90)	56.73 – 2.85 (3.06 – 2.85)
Unit cell parameters (Å, °)	a = 43.10, b = 113.32, c = 175.36 $\alpha = \beta = \gamma = 90^o$	a = 43.13, b = 113.42, c = 175.49	a = 43.06, b = 113.46, c = 175.64
Unique reflections	19907 (2830)	19952 (2836)	20994 (2980)
Multiplicity[b]	7.2 (7.3), 3.8 (3.8)	7.2 (7.3), 3.8 (3.8)	7.2 (7.3), 5.8 (3.8)
I/σ(I)	13.2 (5.2)	13.3 (5.0)	12.8 (4.5)
Completeness (%)[b]	100 (100), 100 (100)	100 (100), 100 (100)	100 (100), 100 (100)
R_{merge} (%)[c]	10.6 (35.6)	10.6 (36.5)	11.1 (41.7)
No. of sites (SOLVE)		6	
FOM (SOLVE/RESOLVE)		0.25/0.67	

[a] Values in parenthesis represent figures for the highest resolution shell.

[b] Multiplicity and completeness. First given value is for merged data and the second represents separated anomalous pairs. Values in parenthesis are for the highest resolution shell.

[c] Reflection statistics are as reported in SCALA. R_{merge} is calculated as described in Evans, 2006.

5.7 rSpy0125-CTR structure solution

The AutoSol wizard in PHENIX (Adams *et al*., 2002) was used to solve the two rSpy0125-CTR structures as described in chapter 2, section 2.11.5 (using the MAD data). Unit cell calculations predicted 2 molecules in the asymmetric unit for each crystal form and PHENIX was able to find the 6 Se sites in both. The initial electron density map produced by the AutoSol wizard for the A-form crystals using the density modified MAD phases to 2.1 Å was easily interpretable and the RESOLVE auto build feature was able to produce an initial model comprising 518 residues docked within the density and a further 58 residues docked as poly-Ala or poly-Gly chains. These regions show connectivity in the electron density where the actual protein sequence cannot be traced. The initial electron density map produced by the AutoSol wizard for the B-form crystals using the density modified MAD phases to 2.9 Å was also interpretable but due to the limited resolution of the data, RESOLVE auto build was only able to place 47 residues docked in the sequence with a further 500 residues docked as poly-Ala or poly-Gly chains.

5.8 Refinement and rebuilding

5.8.1 A-form crystals

The initial model produced by the AutoSol wizard in PHENIX is just that and although it had reasonable *R* factors as defined by phenix.refine it could be improved further. REFMAC5 (Murshudov *et al*., 1997) was used to refine the initial model against the A-form "peak 2" dataset to 1.9 Å using the 'SAD data directly' target function as the *f*+ and *f*- differences were already separated in the collected data. Since the 'peak 2' data had the highest resolution and redundancy, it was used for refinement, as it would be the most accurate in describing the final model. Interestingly, the 'peak 2' dataset could be used to solve the rSpy0125-CTR structure by SAD but the maps produced were not of the same quality as those produced by the MAD data. The model was then inspected and rebuilt with COOT (Emsley and Cowtan, 2004) using both the experimental map calculated with the density modified MAD phases and the σ-weighted $2\left|F_{obs}\right|-\left|F_{calc}\right|$ and $\left|F_{obs}\right|-\left|F_{calc}\right|$ maps from REFMAC5. It was obvious from the initial round of refinement that the electron density describing two domains of rSpy0125-CTR was of high quality, but there was limited evidence for an ordered third domain (see section

5.13). Intriguingly, in the initial maps continuous electron density was observed linking the side chains of Lys297-Asp595 and Cys426-Gln575. At this stage water molecules were added using ARP/wARP solvent (Langer *et al*., 2008) which searches for peaks in the $|F_{obs}| - |F_{calc}|$ electron density map greater than 5.4 sigma, and removes water molecules placed in the $2|F_{obs}| - |F_{calc}|$ electron density below 1 sigma. All water molecules were inspected in COOT and those able to form hydrogen bonds with adjacent molecules were retained after refinement. A final model was produced *via* iterative cycles of refinement with REFMAC5 and rebuilding in COOT. During the final stages of refinement, links were created between the side chains of the residues described above and translation, libration and screw-rotation (TLS) restraints were used grouping residues Thr290-Glu387 and Pro388-Glu585 in chain A and residues Gln287-Glu387 and Pro388-Glu585 in chain B. The final model comprised residues Thr291-Thr603 in chain A, excluding residues Gly316-Ala324 which were disordered, and residues Thr291-Asn605 in chain B, excluding residues Asn318-Phe322 which were disordered. Final refinement statistics for the A-form rSpy0125-CTR crystal structure are given in Table 5.4.

5.8.2 B-form crystals

The initial model produced by the AutoSol wizard in PHENIX was used in conjunction with positioning the refined A-form structure into the initial experimental map produced using the density modified MAD phases. This was achieved with MOLREP (Vagin and Teplyakov, 2010) using the 'search for model in the map' performing 'rotation function'. Unlike for the A-form crystals, continuous electron density describing both molecules was present enabling manual building of the missing domain in COOT, using both the experimental map calculated with the density modified MAD phases and the σ-weighted electron density maps ($2|F_{obs}| - |F_{calc}|$ and $|F_{obs}| - |F_{calc}|$) calculated by REFMAC5. Continuous electron density linking the amino acid side chains described above (Lys297-Asp595 and Cys426-Gln575) was also found in the initial B-form maps in addition to a link between the side chains of Lys610-Asn715. A final model that described rSpy0125-CTR was achieved as above with iterative cycles of refinement using REFMAC5 and rebuilding in COOT maintaining non-crystallographic symmetry (NCS) restraints (tight on main chain, medium on side chain) throughout.

Table 5.4: Refinement statistics for rSpy0125-CTR crystal structures.

	rSpy0125-CTR, A-form	rSpy0125-CTR, B-form	rSpy0125-CTR, Native
Dataset used	Peak '2'	Peak	DLS data
R_{work} (%)[a, b]	20.6 (23.7)	23.1 (30.9)	22.8 (34.8)
R_{free} (%)[a, b]	23.2 (29.6)	27.4 (41.0)	26.8 (42.7)
RmsBond (Å)	0.015	0.005	0.015
RmsAngles (°)[b]	1.41	0.91	1.42
No. of non-hydrogen atom	5423	6742	6746
ESU (based on maximum likelihood) (Å)[b]	0.08	0.35	0.30
Ramachandran outliers (%)[c]	0	0	0

[a] Values in parenthesis are those for the highest resolution shell.

[b] Refinement statistics are as reported by REFMAC5 (Murshudov *et al*., 1997). RmsBond and RmsAngle refer to the root-mean-square deviation from 'ideal' geometric values described in the Refmac dictionary. $R_{free} = \sum(| |F_{obs} - |F_{calc} |) / \sum |F_{obs}|$ where $|F_{obs}|$ are observed structure factor amplitudes for a given reflection and $|F_{calc}|$ are corresponding calculated structure factor amplitudes obtained from the refined model. R_{free} uses only random reflections set aside as a 'test set' and not used in refinement. R_{work} is calculated as per R_{free} but using the 'working set' of reflections.

[c] As calculated by MOLPROBITY (Davis *et al*., 2007).

During the final stages of refinement, links were created between the side chains of the above residues and TLS refinement was used grouping residues Thr290-Glu387 in chain A and Gln287-Glu387 in chain B and Pro388-Glu585 and Asn605-Val720 in chains A and B, all forming separate groups. The final model comprised residues Thr290-Pro719 in chain A, excluding residues Asn318-Ser321 and Ile373-Lys376 which were disordered and residues 288-720 in chain B, excluding residues Asp317-Asn320 and Ile373-Lys376 which were disordered. Final refinement statistics for the B-form rSpy0125-CTR crystal structure are given in Table 5.4. The completed B-form rSpy0125-CTR structure is shown in Figure 5.5 with the top domain in green (residues

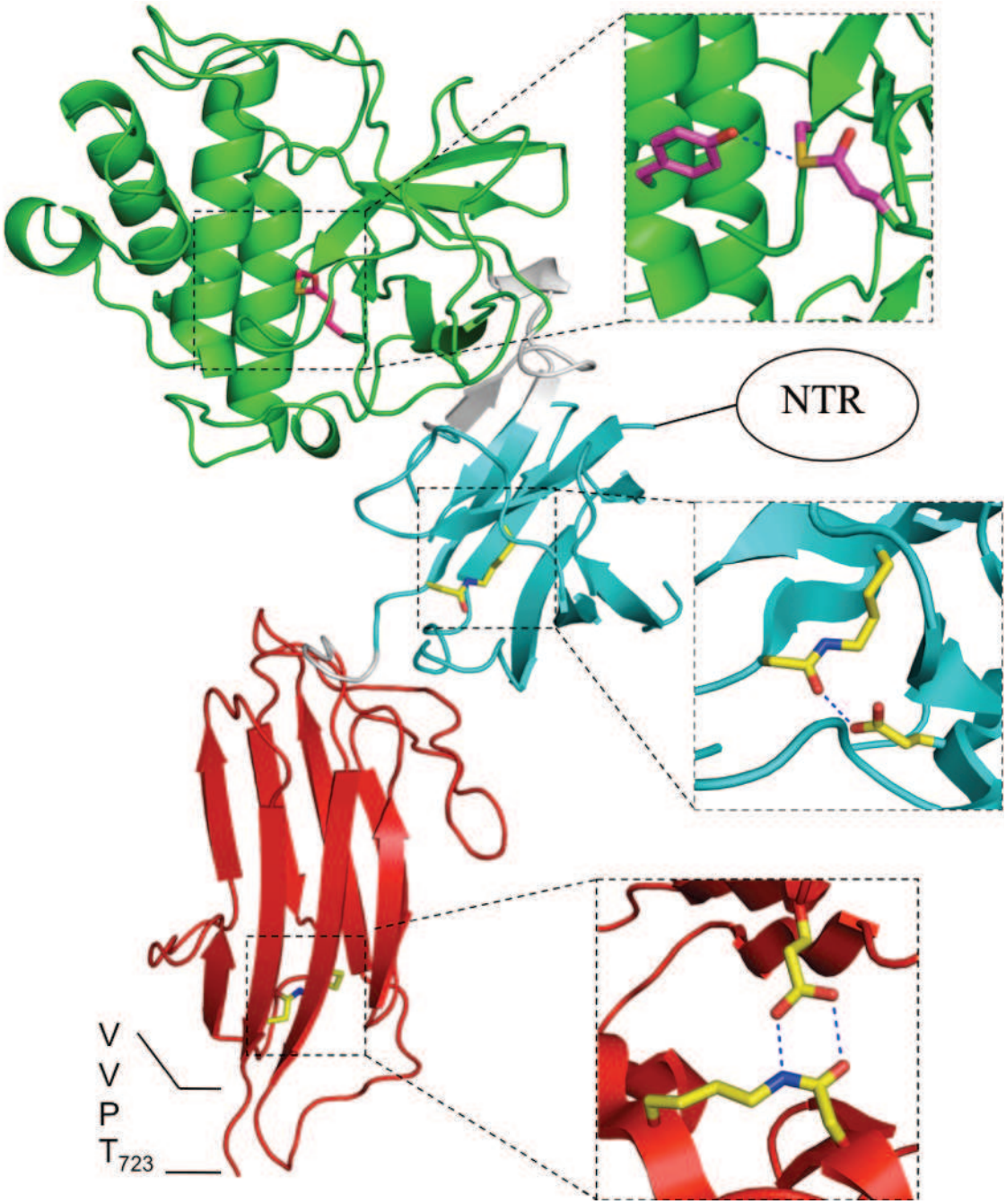

Figure 5.5: Cartoon ribbon representation of the rSpy0125-CTR monomer.

The rSpy0125-CTR monomer is coloured as follows; the top domain (residues Ser390-His583) in green, the middle domain (residues Gln290-Ile372) in sky blue and the bottom domain (residues Thr603-Pro719) in red. Domain linking regions are shown in grey. Indicated at the N-terminus of the monomer is Spy0125-NTR. Highlighted in boxes are the positions of the *intra*-molecular bonds; a Lys-Asn isopeptide bond in the bottom domain (shown with catalytic Glu residue), a Lys-Asp isopeptide bond in the middle domain (shown with catalytic Glu residue) and Cys-Gln thioester bond in the top domain (shown with potentially catalytic Tyr residue). Blue dashed lines indicate hydrogen bonding between atoms. Highlighted at the C-terminus of rSpy0125-CTR is Thr723, which is attached to Lys161 of Spy0128, positioning Spy0125 at the pilus tip.

Ser390-His583), the middle domain in blue (residues Thr290-Ile372 and Ile590-Lys597), the bottom domain in red (residues Thr603-Pro719) with domain linking regions in grey. Also highlighted are the positions of the *intra*-molecular bonds (see section 5.11) and the N-terminus of Spy0125-CTR. Domains were assigned based on their relative position when Spy0125-CTR is modelled at the tip of the pilus. The bottom domain of Spy0125-CTR contains the 'VVPTG' sortase recognition motif and this would form the 'bottom' of the molecule attached to the 'top' of Spy0128 (highlighted in Figure 5.5).

5.8.3 Native B-form crystals

PHASER (McCoy *et al*., 2007) was used to solve the native rSpy0125-CTR structure with the data collected at DLS by molecular replacement to 2.65 Å using the final B-form structure as a homologous search model. A single chain from the final B-form model described above in section 5.8.2 was separated into two different input files, one containing the 'middle and top' domains and the second containing the 'bottom' domain for rotation/translation searches. A final model that described the rSpy0125-CTR was produced as above with iterative cycles of refinement using REFMAC5 (Murshudov *et al*., 1997) maintaining NCS restraints (tight on main chain, medium on side chain) throughout and rebuilding in COOT (Emsley and Cowtan, 2004).

The final model comprised residues Thr290-Pro719 in chain A, excluding residues Val319-Ser321, Ile373-Lys376 and Glu598-His604 which were disordered and residues Pro288-Pro719 in chain B, excluding Ile373-Asp374 which were disordered. Final refinement statistics for the native rSpy0125-CTR crystal structure are given in Table 5.4 and the final electron density map quality for each refined structure is shown in Figure 5.6.

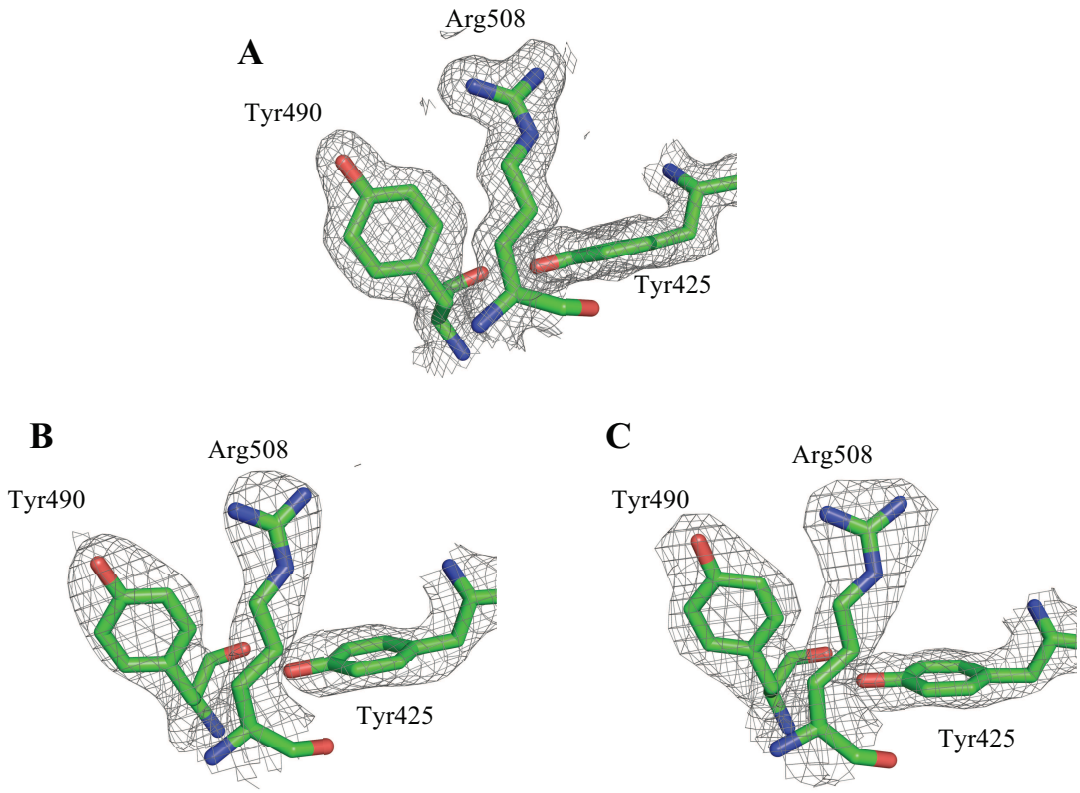

Figure 5.6: Example of the final electron density for rSpy0125-CTR structures.

The 2 $|F_{obs}|$ - $|F_{calc}|$ electron density is contoured at 1.2 σ calculated with phases from the final model. **Panel A**: A-form map refined to 1.9 Å. **Panel B**: B-form map refined to 2.9 Å. **Panel C**: Map derived from the DLS data refined to 2.65 Å. Note the improvement in electron density map quality with increasing resolution.

5.9 Structure validation

All three structures were checked for geometric and structural validity using COOT tools during refinement and MOLPROBITY (Davis *et al.*, 2007) to generate Ramachandran plots for the final structures. These plots are a way of visualising the ϕ against ψ dihedral angles of each of the residues present in the structure. MOLPROBITY also analyses the conformations of each amino acid in protein structures deposited in the PDB and whether the residues present in the structure being examined are in acceptable conformations. Ramachandran plots for the three finished rSpy0125-CTR structures are shown in Figures 5.7, 5.8 and 5.9. All residues are in acceptable conformations and there are no outliers.

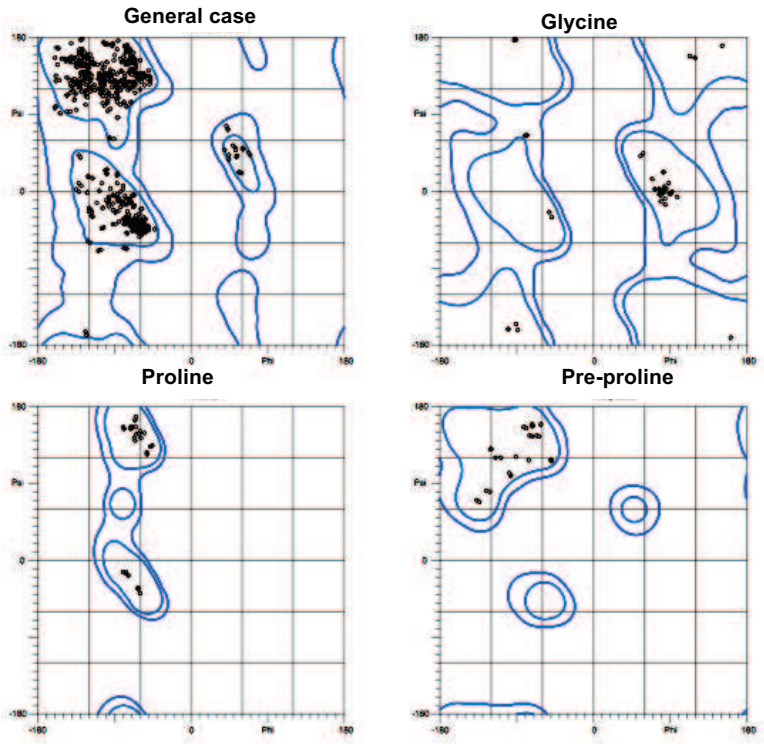

Figure 5.7: Ramachandran plots for the final refined A-form rSpy0125-CTR structure as determined by MOLPROBITY.

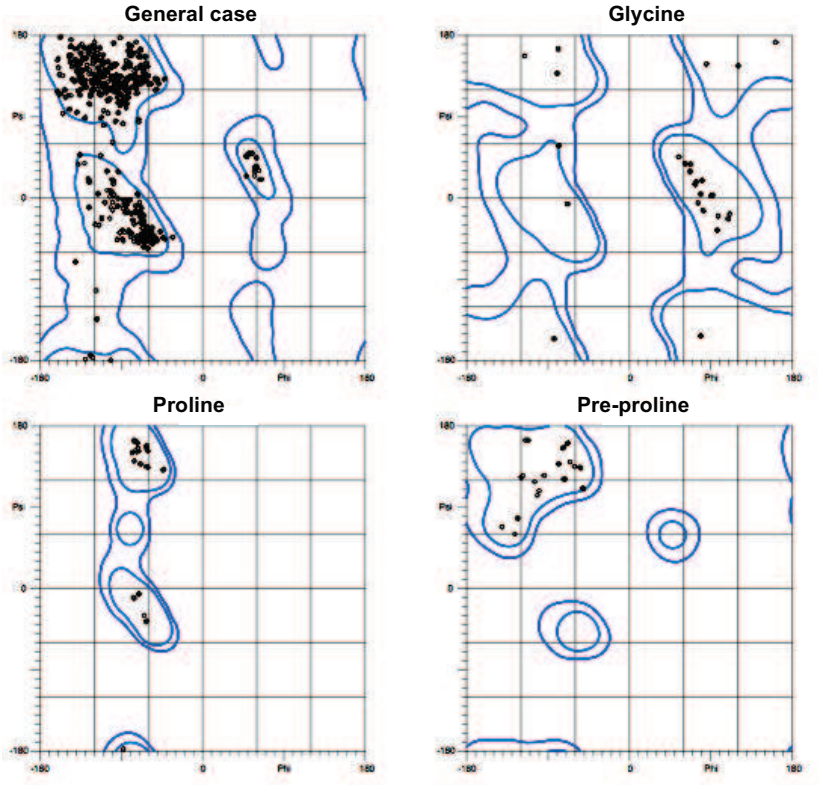

Figure 5.8: Ramachandran plots for the final refined B-form rSpy0125-CTR structure as determined by MOLPROBITY.

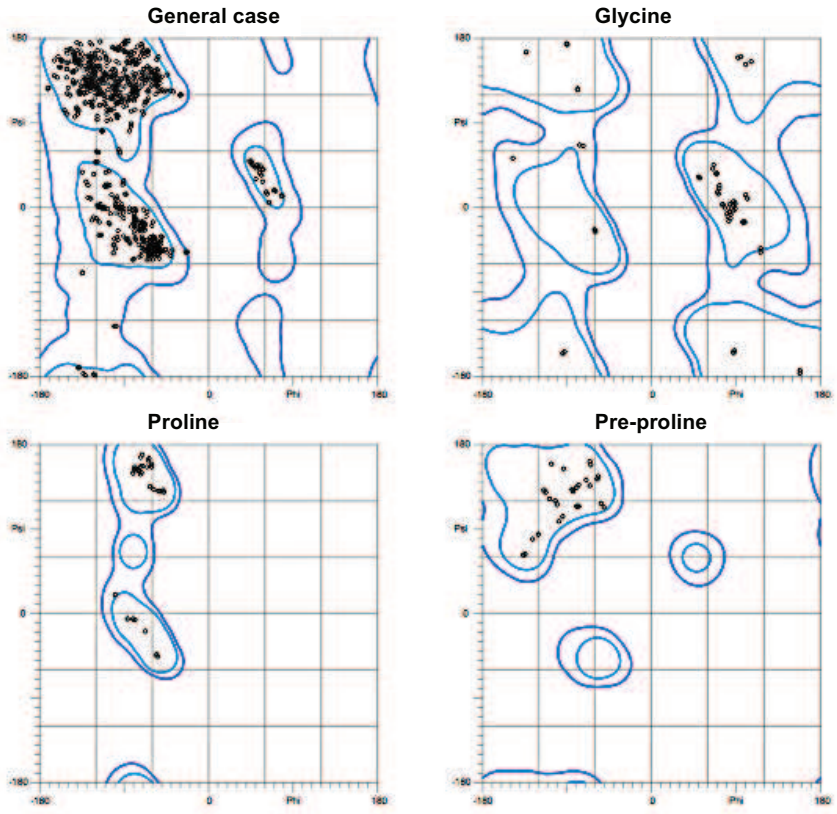

Figure 5.9: Ramachandran plots for the final refined rSpy0125-CTR structure derived from the DLS data as determined by MOLPROBITY.

5.10 Analysis of rSpy0125-CTR structures

LSQMAN (Kleywegt *et al*., 2001) was used to analyse the domains contained within each completed structure and determine their similarity. This is achieved by aligning structures and generating a root mean square deviation (rmsd) which shows the distance of the average displacement of C_α atoms. An rmsd of around ~2.0 Å or lower can be considered to show significant structural similarity while two random structures would generate an rmsds of greater than 12.0 Å. The middle and bottom domain of rSpy0125-CTR both form a β–sandwich arrangement. Overlaying the middle domains from each of the structures generates an rmsd over the range 0.25-0.50 Å (81-85 equivalent C_α positions). Comparison of the bottom domains generates an rmsd of 0.053-0.33 Å (115-117 equivalent C_α positions) while overlaying the top domains from each of the structures generates an rmsd over the range 0.06-0.55 Å (185-194 equivalent C_α positions). Note, the lowest rmsd values generated for each overlay is between domains from the same crystal, i.e. the A-chain and B-chain from the same asymmetric unit (these have already been restrained against each other using NCS restraints during

refinement). The value described here show that domain structure is highly analogous between the different structures. A single region Asp439-Val453 however, adopts a slightly different conformation in the A-form and B-form crystals. This region is well defined in the electron density of both structures however and the most reasonable explanation for this minor rearrangement is the differing crystal contacts throughout this region. Evaluation of the B-form rSpy0125-CTR model derived from the MAD data collected at the ESRF reveals that the electron density covering Pro470-Gln560 in chain B is considerably poorer than that describing the rest of the model in both chains. This region also displays higher than average B factors (which describe the atomic displacement parameters). Despite this, the region refines in the same conformation in each chain from each crystal, including chain A of the same asymmetric unit. If this region is removed and the model refined, the R_{free} value starts to increase indicating that deletion of this region makes for a poorer model describing the electron density. Therefore this region is retained in the final model.

5.11 *Intra*-molecular bonds

The most fascinating aspects of rSpy0125-CTR revealed by the crystal structures are the presence of three internal bonds, one in each domain of the protein. Two are predicted to be covalent isopeptide bonds which provide stability to the protein in an analogous manner to disulphide bonds and the third is a highly unusual thioester bond that might be involved in the binding of Spy0125 to host tissue surfaces (see below).

5.11.1 Lys610-Asn715 *intra*-molecular isopeptide bond

The C-terminal domain of rSpy0125-CTR (bottom domain) harbours an *intra*-molecular isopeptide bond forming a covalent link between the side chains of Lys610-Asn715. It has been shown previously that a conserved glutamate (or aspartate) residue is associated with each bond and is essential, as mutation of this residue abrogates isopeptide bond formation (Kang *et al*., 2007). As seen in panel A, Figure 5.10, Glu680 is in close enough proximity to be able to form hydrogen bonds with both NH on Lys610 and C=O on Asn715, and it is probable that this bond is formed as previously suggested (Kang *et al*., 2007). An OMIT electron density map is shown for this region

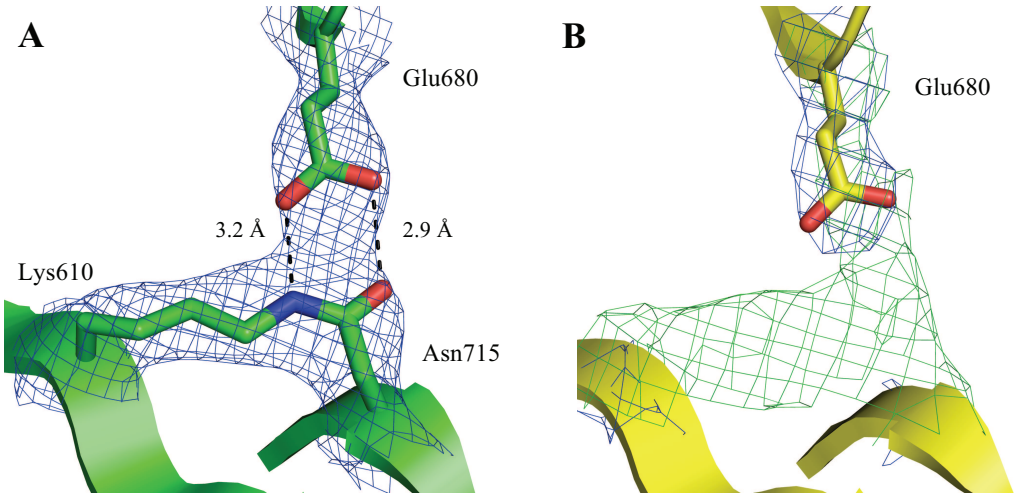

Figure 5.10: The Lys610-Asn715 *intra*-molecular isopeptide bond.

Residues are shown in stick mode with $2|F_{obs}|-|F_{calc}|$ density shown in blue contoured at 1.2 σ and positive $|F_{obs}|-|F_{calc}|$ density in green contoured at 3.5 σ. **Panel A**: Amide linkage formed between the side chains of Lys610 and Asn715. Glu680 is also shown and could play a role in bond formation. **Panel B**: OMIT electron density map of the same residues. Connected green density confirms the presence of the isopeptide bond.

in Panel B, Figure 5.10. In this electron density map, isopeptide bond residues have been removed from the model, which has then been refined. The presence of positive, connected $|F_{obs}|-|F_{calc}|$ electron density (shown in green), indicates that the omitted residues are in the correct conformation in the model. Clusters of hydrophobic residues surrounding these bonds have been implicated in creating a suitable environment for their formation. In the bottom domain of rSpy0125-CTR, residues Phe623, Phe625, Tyr686, Val688 and Phe713 form the hydrophobic cluster. The position of this bond within the bottom domain of rSpy0125-CTR is shown in Figure 5.5.

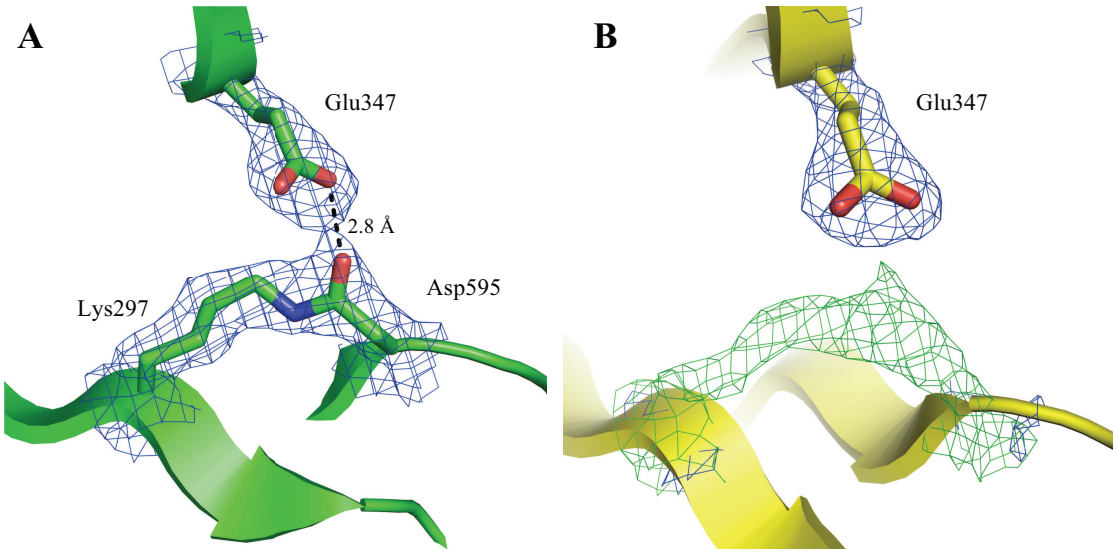

Figure 5.11: The Lys297–Asp595 *intra*-molecular isopeptide bond.

Residues are shown in stick mode with 2 $|F_{obs}| - |F_{calc}|$ density shown in blue contoured at 1.2 σ and positive $|F_{obs}| - |F_{calc}|$ density in green contoured at 3.5 σ. **Panel A**: Amide linkage formed between the side chains of Lys297 and Asp595. Glu347 is also shown and could play a role in bond formation. **Panel B**: OMIT electron density map of the same residues. Connected green density confirms the presence of the isopeptide bond.

5.11.2 Lys297-Asp595 *intra*-molecular isopeptide bond

A second *intra*-molecular isopeptide bond is found within the middle domain of rSpy0125-CTR. Surprisingly, this covalent link is formed between the side chains of Lys297-Asp595 rather than between Lys and Asn side chains as outlined above and for previously described *intra* molecular isopeptide bonds, such as those contained within Spy0128 (Kang *et al*., 2007). This is the first time that an *intra*-molecular isopeptide bond between these two residues has been observed. As seen in panel A, Figure 5.11, Glu347 is probably acting in an analogous manner as it does in the creation of Lys-Asn isopeptide bonds. Glu347 is within a suitable distance to form hydrogen bonds with the O moiety of the Lys-Asp isopeptide bond. An OMIT electron density map was produced for this region as described above (Panel B, Figure 5.11) and the presence of positive $|F_{obs}| - |F_{calc}|$ electron density (shown in green) which is connected, confirms the linkage of these side chains in the model. As described for Lys-Asn isopeptide bonds, a hydrophobic environment is needed for bond formation and the residues involved in forming this cluster are Leu307, Ala310, Leu312, Pro351, Tyr354, Ala 357 and Met593.

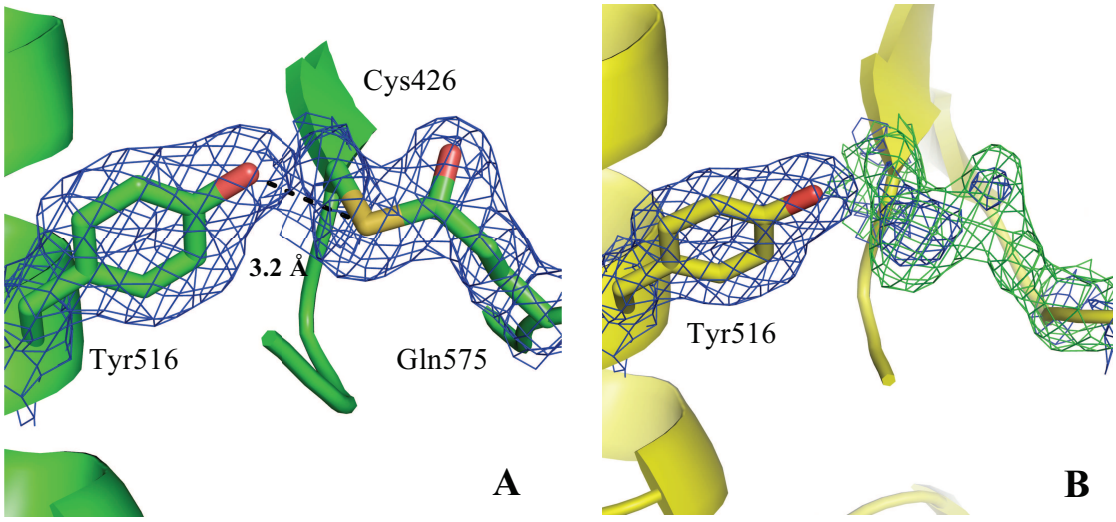

Figure 5.12: The Cys426-Gln575 thioester bond.

Residues are shown in stick mode with 2 $|F_{obs}| - |F_{calc}|$ density shown in blue contoured at 1.2 σ and positive $|F_{obs}| - |F_{calc}|$ density in green contoured at 3.5 σ. **Panel A**: The thioester link formed between the side chains of Cys426 and Gln575, with local Tyr516 also shown, forming a potential hydrogen bond with Cys426 sulphur. **Panel B**: OMIT electron density map of the same residues. Connected green density confirms the thioester linkage.

5.11.3 Cys426-Gln575 internal thioester bond

The most fascinating feature revealed by the rSpy0125-CTR structure is the presence of continuous electron density between the side chains of Cys426-Gln575 within the top domain of the protein. The link between these two residues has been termed an isoglutamyl cysteine thioester bond. These residues are located in a shallow cleft that is partially solvent exposed, at the end of adjacent β-strands. Unlike the two isopeptide bonds described above, this bond could form spontaneously with protein folding providing the energy and environment necessary for bond formation (Pangburn, 1992). Nearby residues however, could also be crucial for thioester bond formation, with Tyr516 able to form a hydrogen bond with the cysteine sulphur. This bond is shown in panel A, Figure 5.12 along with Tyr516. Its corresponding OMIT electron density map created as above is seen in panel B, Figure 5.12 and the presence of positive $|F_{obs}| - |F_{calc}|$ electron density covering these residues suggests the presence of a link between

the side chains of Cys426 and Gln575 in the final model. The position of the thioester in the top domain of rSpy0125-CTR is shown in Figure 5.5.

This particular bond has only been found previously in mammalian proteins such as α_2-macroglobulin and complement proteins where it is required to form covalent bonds with their target ligands. This raises the possibility that they might be responsible here for GAS adhesion (see discussion). The structure of the native complement C3 protein revealed that the thioester was shielded from reacting with water or free nucleophiles by a pocket of hydrophobic residues (Janssen *et al*., 2005) and in rSpy0125-CTR residues Tyr412, Phe427, Ala429, Pro434, Tyr516 and Leu577 form a protective pocket around Cys426-Gln575.

5.12 Initial mass spectrometry of rSpy0125-CTR *intra*-molecular bonds

The Spy0128 structure contained two Lys-Asn isopeptide bonds and mass spectrometry of the intact protein revealed a mass, 34.8 Da lower than that calculated from the amino acid sequence (Kang and Baker, 2009). This accounts for the loss of an NH_3 molecule during the formation of each isopeptide bond. This method was used to investigate the mass of rSpy0125-CTR and whether any differences in the experimentally determined mass and the calculated mass could be accounted for by the presence of the *intra*-molecular bonds described above. A fresh sample of rSpy0125-CTR was prepared for FT-ICR MS analysis, which included removal of the His-tag (chapter 2, section 2.6.7). The intact mass of rSpy0125-CTR using this technique was 49288.5 Da in the neutral state (data not shown). The calculated mass of the Asn286-Thr723, rSpy0125-CTR construct, plus the vector encoded Gly-Ser residues left on the N-terminus after thrombin digestion is 49342.0 Da, as determined by the ProtParam tool on the ExPASy server (Gasteiger *et al*., 2005). The difference between these two values of 53.5 Da, coupled with the experimental error of FT-ICR MS at this mass of 1 Da, could be described by the following; one NH_3 group (17 Da) is lost during the formation of the Lys610-Asn715 isopeptide bond, one H_2O (18 Da) is lost during the formation of the Lys297-Asp595 isopeptide bond and one NH_3 group (17 Da) is lost during the spontaneous formation of the Cys426-Gln575 thioester bond.

5.13 Discussion

The results presented in this chapter describe the high resolution crystal structure of the previously described CTR of Spy0125. In chapter 4, Spy0125 was shown to be the adhesin positioned solely at the pilus tip and the *Δspy0125-NTR* and *Δspy0125-ID* mutant strains revealed critically, that the Spy0125-CTR contains the host cell binding interface.

The A-form structure of rSpy0125-CTR is unusual as it appears to display a non-continuous crystal lattice (panel A, Figure 5.13) which is, of course, a physical impossibility that has occurred due to limitations in modelling dynamic regions in protein structures based on crystallographic data. There is significant electron density (albeit 'smeared') at the modelled C-terminus in each chain suggesting the presence of the bottom domain within the crystal; there is also sufficient space for this domain in the un-modelled region. Further evidence for the flexibility of this domain is highlighted in panel B, Figure 5.13 which shows an overlay (produced using LSQMAN, Kleywegt *et al*., 2001) of both chains in the asymmetric unit of the B-form crystals. It is clear that while the top and middle domains of both chains overlay very tightly, the bottom domains do not. The movement in the bottom domain can be accounted for by a translation of ~14 Å and a slight rotation. Crystal contacts between rSpy0125-CTR monomers in the B-form crystals have caused more rigidity allowing electron density for this region to be seen.

It had been reported previously that the C-terminal domain (CTD) of Spy0125 had a significant degree of sequence similarity with the CTD of Spy0128 (Kang *et al*., 2007). When the structures of these to two domains were aligned (in LSQMAN, using the Spy0125-CTR bottom domain and the Spy0128 crystal structure, PDB code 3B2M) an rmsd of 1.3 Å was calculated for 96 equivalent Cα positions. An overlay of both structures reveals that they are very similar. All β-strands present in rSpy0125-CTR bottom domain are in the same positions in Spy0128 CTD, with the major differences being an extra three short β-strands and large extensions to two loop regions in Spy0128 CTD. It was no surprise that the isopeptide bonds present in these domains are located in exactly the same position within both structures. Also, these domains both contain the

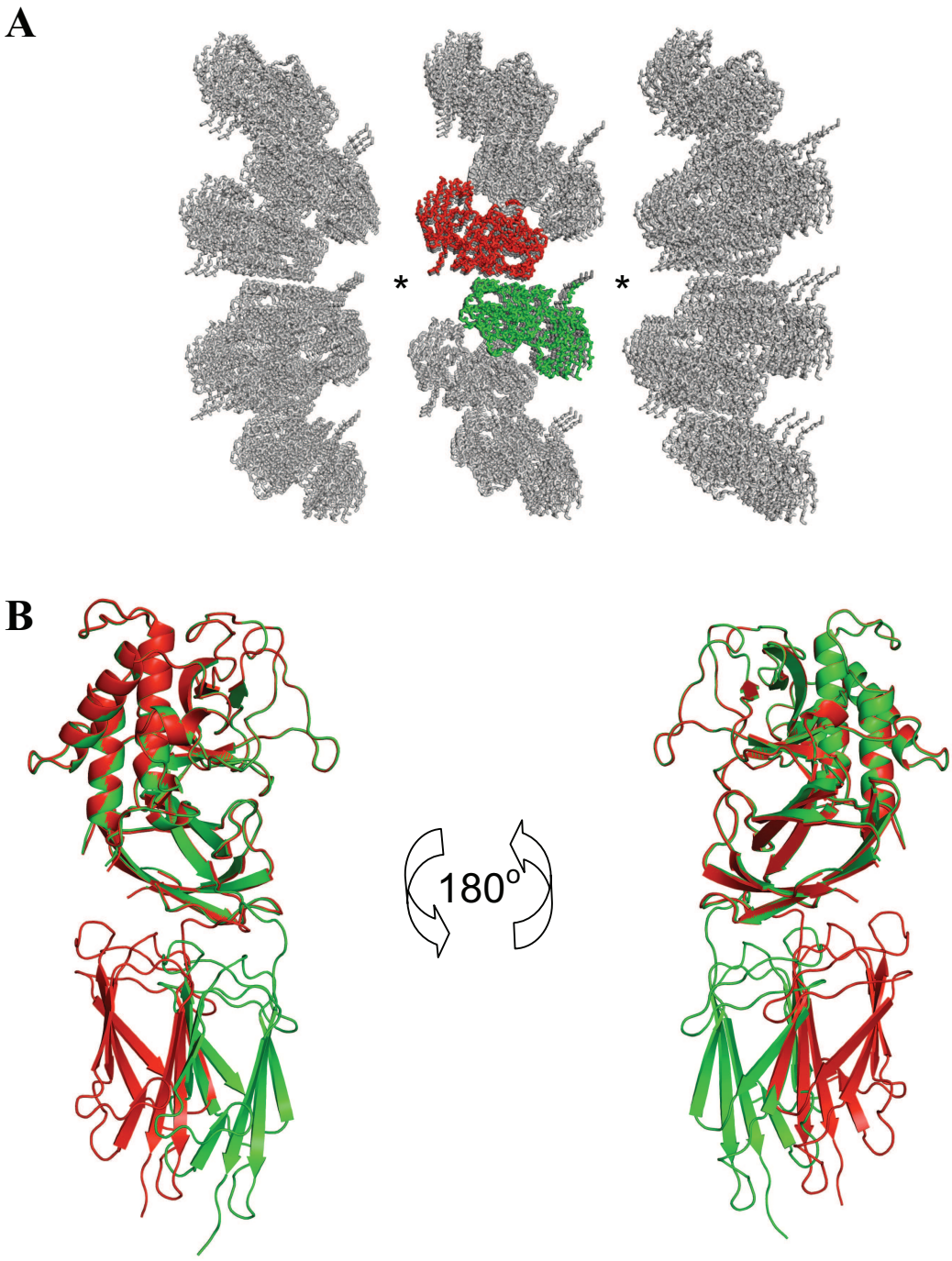

Figure 5.13: Non-continuous lattice in the A-form (P1) crystals.

Panel A: Two clear channels can be seen through the electron density (*) with enough space to suggest that the bottom domain is present in this crystal form. Shown in red and green are the two molecules in the asymmetric unit. **Panel B**: Both molecules from the B-form asymmetric unit are shown, chain A in green and chain B in red. Top and middle domains overlay very tightly but a major translation of 14 Å in the bottom domain is clear.

sortase recognition motifs and are therefore recognised by the pilin polymerase (SrtC) so it follows that these two domains should be structurally similar.

Searches were performed using DALI (Holm and Sander, 1995) to identify any potential structural homologues to the middle and top domains of rSpy0125-CTR. The middle domain of rSpy0125-CTR was found to be structurally similar to the N2 domain of GBS52, a minor pilus subunit from GBS (Krishnan *et al*., 2007). Structure alignments in LSQMAN for these two domains (using the GBS52 crystal structure, PDB entry code 2PZ4) generated an rmsd over 65 equivalent Cα positions of 1.59 Å, suggesting a similar domain architecture for these two structures even though they share little overall sequence similarity. As described above for the bottom domain of rSpy0125-CTR, the isopeptide bond identified here in the rSpy0125-CTR middle domain, occupies a similar position in both GBS52 and rSpy0125-CTR. However, while this bond in GBS52 is formed between the side chains of Lys/Asn residues as described previously for Spy0128 and the bottom domain of rSpy0125-CTR, the corresponding bond in the middle domain of rSpy0125-CTR is between a Lys/Asp pair and has not been found previously. The process of autocatalytic formation of Lys/Asn and Lys/Asp *intra*-molecular isopeptide bonds is shown in Figure 5.14. The *intra*-molecular isopeptide bonds discovered in the middle and bottom domains of the rSpy0125-CTR structure have both been confirmed in the recombinant protein by mass spectrometry and are key features that probably play a stabilising role similar to that reported for other proteins (Pointon *et al*., 2010). Much like rSpy0128, rSpy0125-CTR is stable when stored in solution at 4^{o}C for many months, with no clear degradation visible on SDS-PAGE.

No proteins were found to contain a domain with any structural similarity with the top domain of rSpy0125-CTR, although it is interesting to note the positioning of the top domain within the protein as a whole. In rSpy0125-CTR, the Cα backbone runs (presumably) from the NTR into the middle domain, up into the top domain and then back through the middle domain to the bottom domain (the aspartate residue involved in the middle domain isopeptide bond is located on this short linking region). This may indicate that a gene, encoding the top domain, has been inserted into a loop in the middle domain as parts of amino acid sequence which are distal, yet associated in a structure are evidence for a gene insertion event. Indeed, the reverse could also be true

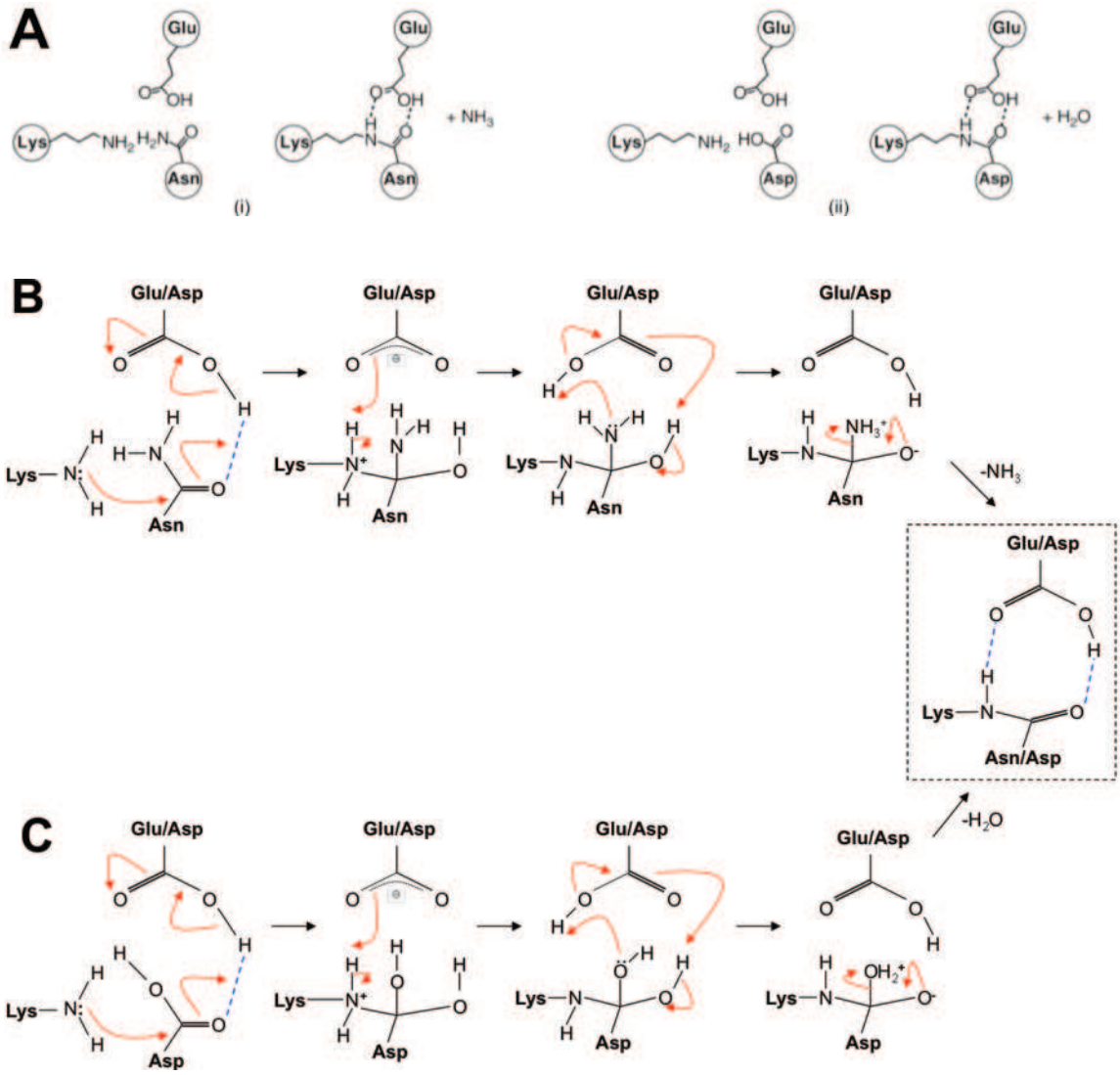

Figure 5.14: Schematic representation of autocatalytic *intra*-molecular isopeptide bond formation.

Panel A: Chemical reaction diagram for Lys-Asn (i) and Lys-Asp (ii) isopeptide bond formation. **Panel B**: Catalysed by a Glu or Asp residue, the potential reaction mechanism for the formation of Lys-Asn isopeptide bonds. **Panel C**: Catalysed by a Glu or Asp residue, the potential reaction mechanism for the formation of Lys-Asp isopeptide bonds. The final stage of both reactions is highlighted, with hydrogen bonds between atoms indicated by blue dashed lines. Figure adapted from Kang and Baker, 2010.

where DNA encoding the top domain of Spy0125-CTR has been deleted from the GBS52 gene. As this thesis was being written the crystal structure of the pilus adhesin, RrgA from *S. pneumoniae* was published which revealed an extremely elongated structure and exhibited evidence of 'indel' events similar to those described above for the possible insertion/deletion of the top domain of Spy0125-CTR. RrgA harbours a MIDAS motif in the top domain of the protein for host cell interaction which is very different to the potential binding interaction reported here for Spy0125-CTR. The position of this motif, in the top domain of the protein is similar to the position of the thioester present in the top domain of rSpy0125-CTR.

The most exciting and intriguing feature revealed by the rSpy0125-CTR structure was the presence of continuous electron density linking the side chains of Cys426 and Gln575, suggestive of an internal thioester within the top domain of the protein. As mentioned in section 5.11.3, the only other proteins shown to contain an internal thioester bond are the α2-macroglobulin protease inhibitors and the complement proteins C3 and C4, which utilise this bond to bind *covalently* to target molecules upon activation (Pangburn, 1992). The thioester contained within the complement proteins C3 and C4 is absolutely critical for their ability to bind target molecules such as bacterial cell walls, and this event is central to the initiation of the complement system and most of its downstream effects. This raises the possibility that the thioester observed within the rSpy0125-CTR structure could be involved in *covalent* binding of M1 GAS strain SF370 to human cells. Such covalent interactions have never before been reported for a bacterial adhesin and its receptor.

To probe any effects the thioester could have on the ability of M1 GAS strain SF370 to bind host cells, a colleague in the laboratory created a *spy0125* Cys426Ala mutation in the bacterial chromosome. The ability of this mutant to bind HaCaT cells was severely compromised compared to wild-type. This suggests that the thioester discovered in the top domain of rSpy0125-CTR plays a critical role in the initial attachment of M1 GAS strain SF370 to host cells (Pointon *et al*., 2010). Adhesion was not totally diminished however and this could be due to the residues surrounding the thioester providing an interface for binding the receptor initially, probably through hydrophobic interactions. This initial binding could trigger a conformational change leading to activation of the thioester, followed by formation of a covalent bond between the ligand and thioester.

This observation would suggest that Spy0125 binding is not indiscriminate but targeted to a specific ligand. The thioester clearly plays an important role in adhesion and although its precise function is not clear, the covalent linkage of M1 GAS strain SF370 to its target receptor is an attractive hypothesis.

The thioester in C3, between Cys988-Gln991, forms a thiolactone ring structure and is protected from reacting with small nucleophiles or water by a hydrophobic pocket around the bond (Janssen *et al*., 2005). Similarly, a hydrophobic pocket surrounds the thioester in rSpy0125-CTR. Whereas the thioester in C3 is formed by residues close in primary sequence, in rSpy0125-CTR the bond is formed between residues distant in primary sequence, that come into close contact following protein folding. Activation of the thioester in C3 occurs with cleavage of a single peptide bond by C3 convertase, resulting in C3a and C3b. This triggers a large conformational change in C3b allowing His1104 to attack the thioester, forming a free thiolate anion and an acyl-imidazole intermediate. The free thiolate then catalyses the transfer of the acyl group of Gln991 to an acceptor hydroxyl and this whole process occurs within a fraction of a second (Janssen *et al*., 2005; Gadjeva *et al*., 1998; Law and Dodds, 1997; Dodds *et al*., 1996). The complement C4 molecule has two isotypes; C4A and C4B. The binding mechanism of C4B is also reliant on a histidine residue (His1106). However, the binding of C4A (containing an Asp at position 1106) and α2-macroglobulin does not go through an intermediate stage and upon their activation, the thioester becomes available to bind amino nucleophiles (Law and Dodds, 1997; Carroll *et al*., 1990).

Mechanisms by which the thioester in Spy0125-CTR may be activated remains a matter of speculation as either of the mechanisms described above may apply. Prior to activation in C3, His1104 is too far away from the thioester (11.7 Å) to have any effect on it but upon cleavage to C3b (Janssen *et al*., 2005), is able to move in and attack the bond. In a similar manner in rSpy0125-CTR, His533 is located 8 Å away from the thioester. It is possible upon activation, a conformational change moves His533 to within a suitable distance of the thioester to enable it to catalyse the formation of the acyl-imidazole intermediate. It would be interesting to test the possible contribution of this residue by site directed mutagenesis. Alternatively, it is possible that free amine nucleophiles could be enough to resolve the thioester bond in Spy0125-CTR although activation probably requires a conformational change to expose the thioester.

Sequence alignments of Spy0125 homologues reveal that the residues involved in thioester bond formation, including those which form the hydrophobic pocket, and His533, are highly conserved. The conservation of these sequences raises the possibility that the thioester bond also plays a role in adhesion in these strains. All Spy0125 homologues belong to FCT types 2, 3 and 4 (chapter 1 section 1.14.1). Other predicted pilus adhesins, from GAS FCT type 1 (M6 Spy0157) and FCT type 5 (M4 Spy0114), show little sequence homology to the CTR of Spy0125 although they do harbour a 'VYCFN' (thioester cysteine containing) motif. Both also contain a 'KDQS' motif which is highly similar to the 'KYQS' motif (containing Gln575) present in Spy0125-CTR. In other Gram positive pilus adhesins described to date, only those from *Corynebacterium diphtheriae* (SpaC, SpaF and SpaG) contain any cysteine residues. Much larger number of cysteines are present compared to Spy0125 (2 Cys); SpaC contains 28 Cys; SpaF contains 6 Cys and SpaG contains 10 Cys but neither of the motifs from above are found other than a 'KGQS' motif in SpaC. However, this does not exclude the possibility that thioester bonds are formed in these adhesins or even disulphide bonds, which have been shown to be present in the *C. diphtheriae* SpaA (major pilus subunit) crystal structure (Kang *et al*., 2009b), although this cannot be known until detailed structural or biochemical analyses of these proteins are carried out. Cysteine residues are not found in the pilus adhesins of Group B *Streptococcus*, *Streptococcus pneumoniae* or *Bacillus cereus*.

The presence of the thioester in rSpy0125-CTR was confirmed by colleagues in the laboratory using UV-Vis spectroscopy (Pointon *et al*., 2010). The next challenge however, and this would no small undertaking, is to confirm its presence in native pili at the cell surface, identify the mechanism by which the thioester is activated and the receptor involved. One possibility for activation could be cleavage of the Spy0125-NTR although this is unlikely as the *Δspy0125-NTR* mutant still binds at wild-type levels (see chapter 4). A more attractive hypothesis is that the surface residues surrounding the thioester associate with the receptor, which may not be a strong attraction but could be sufficient to expose it to adjacent sites, facilitating strong covalent binding. Further experiments would be needed to confirm this however.

Chapter 6. General discussion

Of the many fascinating characteristics displayed by pili in Gram positive bacteria, one of the more intriguing is that although they are expressed by a wide range of different species, they are formed in essentially the same manner, covalently polymerised by the action of sortases with in most cases, the same number of subunits. This contrasts significantly with Gram negative bacteria where pili, that are held together by non-covalent interactions can be grouped into four distinct types based on their assembly pathway. The major types of Gram negative and Gram positive pili are summarised in Table 6.1 on page 171. The pili investigated in this thesis are encoded in the FCT-2 region of the M1 GAS strain SF370 chromosome. To date, nine FCT regions have been described and the pili they encode are believed to form in a similar manner with small differences in the details. These regions resemble pathogenicity islands (PI) suggesting they probably are, or were, mobile and transferred between species and strains within a species. Extensive recombination between islands is likely to help pili avoid selective pressures exerted by the immune system similar to the M and M-like proteins (Kehoe *et al*., 1996). Evidence for PI status is highlighted by the presence of transposase genes in the FCT-2 region of M1 GAS strain SF370 and the very close relationship (over 50% sequence homology) implying horizontal transfer of the adhesin from PI-1 of *S. agalactiae*, PI-1 of *S. pneumoniae* and FCT-6 of M2 GAS. In this instance, pili probably confer a specific tissue tropism as both *S. agalactiae* and M2 GAS strains are frequently isolated from the vaginal mucosa (Falugi *et al*., 2008). Isolation of *S. pneumoniae* from this region of the body has yet to be reported although this probably has more to do with the low carriage rates of the *rlrA* pilus islet and a more specific transmission and tissue tropism that involves other adhesive surface proteins (Aguiar *et al*., 2008). It has been noted however, that the FCT region of GAS, although believed to be a mobile pathogenicity island, lacks features commonly associated with these regions; one has been found in every GAS strain studied to date, situated between two conserved genes and the regions are not flanked by tRNA genes commonly associated with true pathogenicity islands (Podbielski, 2007).

The work described in this thesis began following the discovery that M1 GAS strain SF370 expressed polymeric pili that were attached to the cell wall. Three proteins

composed the pilus, with one forming the shaft (Spy0128) and two ancillary proteins whose functions were unknown at the beginning of this Ph.D project. Work described in this thesis has shown that one of these proteins acts an adhesin and is situated at the tip of the pilus (Spy0125), while parallel studies by colleagues in the laboratory have shown the remaining subunit acts as a 'linker' molecule attaching the pilus structure to the cell wall (Spy0130, Smith *et al*., 2010). Although pili from Gram negative bacteria had been studied extensively for many years, there was relatively little information in the literature about Gram positive pili and certainly no structural data concerning these proteins. The work presented in this thesis attempted to address these issues.

During the early stages of this project Manetti *et al*. (Manetti *et al*., 2007) reported both Spy0130 and Spy0125 bound to Detroit pharyngeal cells suggesting both could be the adhesin. As mentioned in chapter 4, only antibodies to recombinant (r)Spy0125 blocked adhesion of intact M1 GAS strain SF370 to target cells. Subsequent work showed that only Spy0125 is positioned at the tip acting as the pilus presented adhesin. Studies by others in the laboratory indicated that Spy0130 acted as a 'linker' between the last Spy0128 subunit incorporated into the growing pilus and the di-alanine cross bridge of the lipid II precursor. Consistent with this, work in chapter 3 of this thesis confirmed rSpy0130 to be an extremely elongated molecule in solution, probably reflecting its role at the cell surface. As discussed previously, a large degree of flexibility should be expected as Spy0130 and the pilus structure as a whole would be subjected to considerable tensile forces upon binding host surfaces. However, this property was probably the primary reason for unsuccessful attempts to crystallise rSpy0130, preventing intimate packing of molecules within a crystal. Due to time constraints and focussing efforts on Spy0125 crystallisation, attempts to gain insights into the high resolution structure of Spy0130 were not pursued. While this thesis was being written we learnt that another group had tried to crystallise rSpy0130 without success although they were able to solve the X-ray crystal structure of a homologue whose globular structure closely matches the SAXS structure for rSpy0130 presented in chapter 3 (Linke *et al*., 2010).

The major part of the work described in this thesis focused on Spy0125 when it was discovered to be the pilus presented adhesin. rSpy0125 and its breakdown products were characterised in chapter 4 and recombinant protein produced in this work was

given to colleagues in the laboratory who carried out experiments that suggested rSpy0125 N-terminal region (NTR) bound to human tissues but rSpy0125 C-terminal region (CTR) did not. This contrasted antibody inhibition experiments where only anti-sera against rSpy0125-CTR inhibited binding of intact M1 GAS strain SF370. Taken together, these data raise doubts over the use of the recombinant proteins so in order to examine which of the two regions were more likely to be responsible for adhesion, the region of the M1 GAS strain SF370 chromosome encoding Spy0125-NTR was deleted using allele replacement. No reduction in binding was observed in the absence of the Spy0125-NTR, suggesting any contribution it makes to binding is very small and the part of the protein responsible for attachment, resides in the CTR of Spy0125. These results confirm the importance of looking at the function of proteins *in vivo* and not just relying on what is observed in the test tube with recombinant proteins.

Not only are pili in Gram positive bacteria formed in the same manner from a similar number of subunits, the proteins within them also contain many similar features. These include the folding of domains within subunits, even with little homology at the sequence level such as between the middle domain of rSpy0125-CTR presented in this thesis and GBS52 (rmsd-1.59 Å, both are ancillary proteins that probably play different roles in their respective pilus; Spy0125-adhesin, GBS52-cell wall anchor), and the presence of *intra*-molecular isopeptide bonds. These bonds were first observed in the crystal structure of Spy0128 in 2007 (Kang *et al*., 2007) and have since been found in a growing number of Gram positive pili proteins, not just restricted to major subunits. The RrgA adhesin from *S. pneumoniae* RlrA pili contains two *intra*-molecular isopeptide bonds, as does rSpy0125-CTR. In all cases where proteins have been demonstrated to contain *intra*-molecular isopeptide bonds, these proteins are extremely stable in solution. For example, rSpy0128 and rSpy0125-CTR examined in this project exhibited minimal breakdown after storage at 4^{o}C for one year and would even crystallise after this length of time. In contrast, both rSpy0130 and rSpy0125-NTR were labile in solution. Degradation of rSpy0125-NTR was observed to begin during purification and it might be predicted from this that this region does not contain an *intra*-molecular isopeptide bond. Also, the experimentally determined intact masses for the two NTR fragments identified from the breakdown of rSpy0125 were so accurate as to suggest no *intra*-molecular bonds were present (no loss of NH_3 or H_2O was indicated). Work done examining the role of the *intra*-molecular isopeptide bonds in Spy0128 revealed they

contributed to the thermal and structural stability of the protein (Kang and Baker, 2009) and probably play a similar role in Spy0125-CTR. This is advantageous as they are expressed outside of the cell and therefore exposed to potentially harsh environments. A role for these bonds in withstanding strong mechanical forces upon adhesion to a host receptor has also been suggested (Alegre-Cebollada *et al*., 2010). As above, this is an attractive hypothesis because when Spy0128 monomers are positioned in a head-to-tail conformation (like they would be presented in native pili at the cell surface), the *intra*-molecular isopeptide bonds are parallel to the long axis of the pili, running throughout the polymer creating an extremely strong structure (the covalent linkages within and between subunits in Gram positive pili has been described elsewhere as like 'beads on a string'). Prior to this project however, no study had been conducted into their importance for host cell attachment. The work described in chapter 3, where mutations in each of the bonds were engineered into the GAS chromosome, suggests that *intra*-molecular isopeptide bonds are dispensable for M1 GAS strain SF370 attachment to HaCaT cells. The pilus phenotypes suggested that although disruption of the critical Glu residue significantly altered pilus polymerisation, ultimately all strains bound to HaCaT cells at wild-type levels. The extra stability provided by the isopeptide bonds is probably needed to withstand the strong tensile forces experienced throughout the pili upon attachment to a target surface. In these experiments however, only gentle washing forces were applied, highlighted by the presence of background binding in the *Δspy0128* and *spy0128* K161A strains. More significant washing forces would probably be experienced by bacteria in the oropharynx by the action of saliva for example, and perhaps this could be explored in future experiments.

Perhaps the most fascinating result of this Ph.D project was the discovery of continual electron density linking the side chains of Cys426 and Gln575 in the rSpy0125-CTR crystal structure, revealing the presence of an internal thioester bond in the top domain of the protein. This feature has not been reported previously in any other bacterial protein. It is particularly exciting when it is considered that the only proteins identified to date which harbour an internal thioester bond are the complement proteins C3 and C4 and the α-2 macroglobulins. In these instances the thioester is responsible for one function; *covalent* attachment to target molecules *via* free amine or hydroxyl functional groups. Pili in M1 GAS strain SF370 have been shown to be critical for initial binding

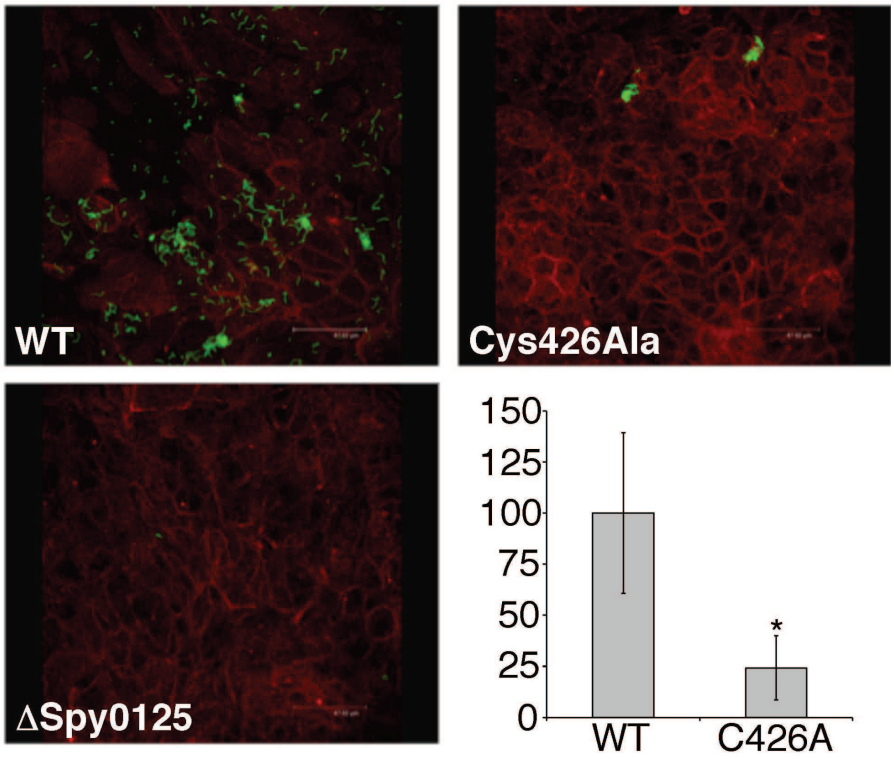

Figure 6.1: Adhesion of M1 GAS strain SF370 Spy0125 Cys426Ala mutant to HaCaT cells, reproduced from Pointon *et al*., 2010.

of the bacteria to human tonsil (Abbot *et al*., 2007) and the presence of a thioester bond in the adhesin raises the intriguing possibility that this strain could bind to human tissues covalently. Binding *via* a thioester bond alone could be indiscriminate, to a number of different receptors that display the correct functional group. This scenario is unlikely however, as GAS displays a specific tissue tropism. Moreover, M1 GAS strain SF370 pili do not bind to Hep-2 or A549 cells (Abbot *et al*., 2007). While this thesis was being written, to test the role of the thioester, colleagues in the laboratory produced a Cys426Ala mutant strain which exhibited dramatically reduced binding to HaCaT cells compared to wild-type GAS (Figure 6.1, reproduced with permission from Pointon *et al*. 2010). Binding is not totally abrogated however. The initial interaction between the specific receptor on human cells and Spy0125 (likely a hydrophobic one) could be mediated by residues surrounding the thioester and this binding could then slightly open the shallow cleft that the bond sits in, exposing it, allowing a more intimate covalent interaction to occur. If this were found to be the case, in effect a single, large, covalently cross linked polymer would link the bacteria to the host cell, strengthened by the

presence of *intra*-molecular isopeptide bonds in both Spy0125-CTR and Spy0128 all facing in the same direction. This would represent an extremely strong and irreversible reaction which could be key during the initial stages of an infection in an environment such as the oropharynx. Further experiments by colleagues in the laboratory have confirmed the presence of the thioester within rSpy0125-CTR using a chemical approach. rSpy0125-CTR was treated with methylamine, a small molecule nucleophile which was used to break open the thioester and methylate Gln575, specifically in combination with iodoacetamide which alkylated the now free thiol, and these modifications were confirmed by mass spectrometry analysis. No modifications were observed when rSpy0125-CTR was incubated with iodoacetamide alone. Now the goal has to be to prove the existence of this bond in native pili at the GAS cell surface. Were this to be reported, it would be an extremely exciting result that would pave the way for future studies attempting to find out exactly what receptor GAS pili adhere to in the initial stages of infection at the skin and pharynx, and potentially provide a target for vaccine development.

The potential for this bond also highlights that although Gram positive pili proteins display many similar characteristics, they also harbour features unique from one another that can only be identified by determining their high resolution structures. Other features as yet identified as unique to individual Gram positive pili proteins include the disulphide bond found in the C-terminal domain of *C. diphtheriae* SpaA (Kang *et al*., 2009b). As mentioned in chapter 5, there are varying numbers of cysteine residues present in pilus adhesins from other Gram positive bacteria ranging from none to 28 in SpaC from *C. diphtheriae*. As more structural information about pilus presented proteins and adhesin proteins in general becomes available, it will be interesting to note whether a thioester bond as revealed here will be present or whether this feature is specific to Spy0125 of GAS and its homologues.

Publications arising from this thesis

Solovyova, A. S., Pointon, J. A., Race, P. R., Smith, W. D., Kehoe, M. A. and Banfield, M. J. (2010) 'Solution structure of the major (Spy0128) and minor (Spy0125 and Spy0130) pili subunits from *Streptococcus pyogenes*', *Eur Biophys J,* 39, (3), pp. 469-80.

Smith, W. D., Pointon, J. A., Abbot, E., Kang, H. J., Baker, E. N., Hirst, B. H., Wilson, J. A., Banfield, M. J. and Kehoe, M. A. (2010) 'Roles of minor pilin subunits Spy0125 and Spy0130 in the M1 *Streptococcus pyogenes* strain SF370', *J Bacteriol*, 192, (18), pp. 4651-9.

Pointon, J. A., Smith, W. D., Saalbach, G., Crow, A., Kehoe, M. A. and Banfield, M. J. (2010) 'A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction', *J Biol Chem*, 285, (44), pp. 33858-66.

References

'Collaborative Computational Project, n. (1994)', *Acta Crystallogr D Biol Crystallogr,* 50, pp. 760-763.

Abbot, E. L., Smith, W. D., Siou, G. P., Chiriboga, C., Smith, R. J., Wilson, J. A., Hirst, B. H. and Kehoe, M. A. (2007) 'Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin', *Cell Microbiol*.

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C. (2002) 'PHENIX: building new software for automated crystallographic structure determination', *Acta Crystallogr D Biol Crystallogr,* 58, (Pt 11), pp. 1948-54.

Aguiar, S. I., Serrano, I., Pinto, F. R., Melo-Cristino, J. and Ramirez, M. (2008) 'The presence of the pilus locus is a clonal property among pneumococcal invasive isolates', *BMC Microbiol,* 8, pp. 41.

Akesson, P., Sjoholm, A. G. and Bjorck, L. (1996) 'Protein SIC, a novel extracellular protein of *Streptococcus pyogenes* interfering with complement function', *J Biol Chem,* 271, (2), pp. 1081-8.

Alegre-Cebollada, J., Badilla, C. L. and Fernandez, J. M. (2010) 'Isopeptide bonds block the mechanical extension of PILI in pathogenic *Streptococcus pyogenes*', *J Biol Chem*.

Anderson, J. S., Matsuhashi, M., Haskin, M. A. and Strominger, J. L. (1967) 'Biosythesis of the peptidoglycan of bacterial cell walls. II. Phospholipid carriers in the reaction sequence', *J Biol Chem,* 242, (13), pp. 3180-90.

Angelini, S., Deitermann, S. and Koch, H. G. (2005) 'FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon', *EMBO Rep,* 6, (5), pp. 476-81.

Aziz, R. K. and Kotb, M. (2008) 'Rise and persistence of global M1T1 clone of *Streptococcus pyogenes*', *Emerg Infect Dis,* 14, (10), pp. 1511-7.

Bagnoli, F., Moschioni, M., Donati, C., Dimitrovska, V., Ferlenghi, I., Facciotti, C., Muzzi, A., Giusti, F., Emolo, C., Sinisi, A., Hilleringmann, M., Pansegrau, W., Censini, S., Rappuoli, R., Covacci, A., Masignani, V. and Barocchi, M. A. (2008) 'A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells', *J Bacteriol,* 190, (15), pp. 5480-92.

Barnett, T. C., Patel, A. R. and Scott, J. R. (2004) 'A novel sortase, SrtC2, from *Streptococcus pyogenes* anchors a surface protein containing a QVPTGV motif to the cell wall', *J Bacteriol,* 186, (17), pp. 5865-75.

Barnett, T. C. and Scott, J. R. (2002) 'Differential recognition of surface proteins in *Streptococcus pyogenes* by two sortase gene homologs', *J Bacteriol,* 184, (8), pp. 2181-91.

Barocchi, M. A., Ries, J., Zogaj, X., Hemsley, C., Albiger, B., Kanth, A., Dahlberg, S., Fernebro, J., Moschioni, M., Masignani, V., Hultenby, K., Taddei, A. R., Beiter, K., Wartha, F., von Euler, A., Covacci, A., Holden, D. W., Normark, S., Rappuoli, R. and Henriques-Normark, B. (2006) 'A pneumococcal pilus influences virulence and host inflammatory responses', *Proc Natl Acad Sci U S A,* 103, (8), pp. 2857-62.

Beall, B., Facklam, R. and Thompson, T. (1996) 'Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci', *J Clin Microbiol,* 34, (4), pp. 953-8.

Beckert, S., Kreikemeyer, B. and Podbielski, A. (2001) 'Group A streptococcal rofA gene is involved in the control of several virulence genes and eukaryotic cell attachment and internalization', *Infect Immun,* 69, (1), pp. 534-7.

Bensing, B. A. and Sullam, P. M. (2009) 'Characterization of *Streptococcus gordonii* SecA2 as a paralogue of SecA', *J Bacteriol,* 191, (11), pp. 3482-91.

Bernstein, H. D., Poritz, M. A., Strub, K., Hoben, P. J., Brenner, S. and Walter, P. (1989) 'Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle', *Nature,* 340, (6233), pp. 482-6.

Bessen, D. E. (2009) 'Population biology of the human restricted pathogen, *Streptococcus pyogenes*', *Infect Genet Evol,* 9, (4), pp. 581-93.

Bessen, D. E. and Kalia, A. (2002) 'Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in *Streptococcus pyogenes*', *Infect Immun,* 70, (3), pp. 1159-67.

Bessen, D. E., Sotir, C. M., Readdy, T. L. and Hollingshead, S. K. (1996) 'Genetic correlates of throat and skin isolates of group A streptococci', *J Infect Dis,* 173, (4), pp. 896-900.

Birnboim, H. C. and Doly, J. (1979) 'A rapid alkaline extraction procedure for screening recombinant plasmid DNA', *Nucleic Acids Res,* 7, (6), pp. 1513-23.

Bisno, A. L. (1980) *The concept of rheumatogenic and non-rheumatogenic group A streptococci. In Streptococcal diseases and the immune response.*: Academic Press Inc., New York, N.Y.

Bisno, A. L. and Stevens, D. L. (1996) 'Streptococcal infections of skin and soft tissues', *N Engl J Med,* 334, (4), pp. 240-5.

Bisno, A. L. and Stevens, D. L. (2005) *Streptococcus pyogenes. In Principles and Practice of Infectious Diseases.*: Churchill Livingstone, Philadelphia.

Blobel, G. (1980) 'Intracellular protein topogenesis', *Proc Natl Acad Sci U S A,* 77, (3), pp. 1496-500.

Boggon, T. J. and Shapiro, L. (2000) 'Screening for phasing atoms in protein crystallography', *Structure,* 8, (7), pp. R143-9.

Boukamp, P., Petrussevska, R. T., Breitkreutz, D., Hornung, J., Markham, A. and Fusenig, N. E. (1988) 'Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line', *J Cell Biol,* 106, (3), pp. 761-71.

Braun, L., Dramsi, S., Dehoux, P., Bierne, H., Lindahl, G. and Cossart, P. (1997) 'InlB: an invasion protein of *Listeria monocytogenes* with a novel type of surface association', *Mol Microbiol,* 25, (2), pp. 285-94.

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D. S., Weinrauch, Y. and Zychlinsky, A. (2004) 'Neutrophil extracellular traps kill bacteria', *Science,* 303, (5663), pp. 1532-5.

Brunger, A. T. (1992) 'Free R value: a novel statistical quantity for assessing the accuracy of crystal structures', *Nature,* 355, (6359), pp. 472-5.

Buccato, S., Maione, D., Rinaudo, C. D., Volpini, G., Taddei, A. R., Rosini, R., Telford, J. L., Grandi, G. and Margarit, I. (2006) 'Use of Lactococcus lactis expressing pili from group B *Streptococcus* as a broad-coverage vaccine against streptococcal disease', *J Infect Dis,* 194, (3), pp. 331-40.

Buchanan, J. T., Simpson, A. J., Aziz, R. K., Liu, G. Y., Kristian, S. A., Kotb, M., Feramisco, J. and Nizet, V. (2006) 'DNase expression allows the pathogen group A *Streptococcus* to escape killing in neutrophil extracellular traps', *Curr Biol,* 16, (4), pp. 396-400.

Budzik, J. M., Marraffini, L. A. and Schneewind, O. (2007) 'Assembly of pili on the surface of *Bacillus cereus* vegetative cells', *Mol Microbiol,* 66, (2), pp. 495-510.

Budzik, J. M., Marraffini, L. A., Souda, P., Whitelegge, J. P., Faull, K. F. and Schneewind, O. (2008a) 'Amide bonds assemble pili on the surface of bacilli', *Proc Natl Acad Sci U S A,* 105, (29), pp. 10215-20.

Budzik, J. M., Oh, S. Y. and Schneewind, O. (2008b) 'Cell wall anchor structure of BcpA pili in *Bacillus anthracis*', *J Biol Chem,* 283, (52), pp. 36676-86.

Budzik, J. M., Oh, S. Y. and Schneewind, O. (2009a) 'Sortase D forms the covalent bond that links BcpB to the tip of *Bacillus cereus* pili', *J Biol Chem,* 284, (19), pp. 12989-97.

Budzik, J. M., Poor, C. B., Faull, K. F., Whitelegge, J. P., He, C. and Schneewind, O. (2009b) 'Intramolecular amide bonds stabilize pili on the surface of bacilli', *Proc Natl Acad Sci U S A,* 106, (47), pp. 19992-7.

Bulheller, B. M., Rodger, A. and Hirst, J. D. (2007) 'Circular and linear dichroism of proteins', *Phys Chem Chem Phys,* 9, (17), pp. 2020-35.

Carapetis, J. R., Mayosi, B. M. and Kaplan, E. L. (2006) 'Controlling rheumatic heart disease in developing countries', *Cardiovasc J S Afr,* 17, (4), pp. 164-5.

Carapetis, J. R., McDonald, M. and Wilson, N. J. (2005a) 'Acute rheumatic fever', *Lancet,* 366, (9480), pp. 155-68.

Carapetis, J. R., Steer, A. C., Mulholland, E. K. and Weber, M. (2005b) 'The global burden of group A streptococcal diseases', *Lancet Infect Dis,* 5, (11), pp. 685-94.

Carlsson, F., Stalhammar-Carlemalm, M., Flardh, K., Sandin, C., Carlemalm, E. and Lindahl, G. (2006) 'Signal sequence directs localized secretion of bacterial surface proteins', *Nature,* 442, (7105), pp. 943-6.

Carroll, M. C., Fathallah, D. M., Bergamaschini, L., Alicot, E. M. and Isenman, D. E. (1990) 'Substitution of a single amino acid (aspartic acid for histidine) converts the functional activity of human complement C4B to C4A', *Proc Natl Acad Sci U S A,* 87, (17), pp. 6868-72.

Chapnick, E. K., Gradon, J. D., Lutwick, L. I., Kim, J., Levi, M., Kim, M. H. and Schlievert, P. M. (1992) 'Streptococcal toxic shock syndrome due to noninvasive pharyngitis', *Clin Infect Dis,* 14, (5), pp. 1074-7.

Chatterjee, A. N. and Park, J. T. (1964) 'Biosynthesis of Cell Wall Mucopeptide by a Particulate Fraction from *Staphylococcus aureus*', *Proc Natl Acad Sci U S A,* 51, pp. 9-16.

Chiang-Ni, C. and Wu, J. J. (2008) 'Effects of streptococcal pyrogenic exotoxin B on pathogenesis of *Streptococcus pyogenes*', *J Formos Med Assoc,* 107, (9), pp. 677-85.

Cleary, P. P., Kaplan, E. L., Handley, J. P., Wlazlo, A., Kim, M. H., Hauser, A. R. and Schlievert, P. M. (1992a) 'Clonal basis for resurgence of serious *Streptococcus pyogenes* disease in the 1980s', *Lancet,* 339, (8792), pp. 518-21.

Cleary, P. P., LaPenta, D., Vessela, R., Lam, H. and Cue, D. (1998) 'A globally disseminated M1 subclone of group A streptococci differs from other subclones by 70 kilobases of prophage DNA and capacity for high-frequency intracellular invasion', *Infect Immun,* 66, (11), pp. 5592-7.

Cleary, P. P., Prahbu, U., Dale, J. B., Wexler, D. E. and Handley, J. (1992b) 'Streptococcal C5a peptidase is a highly specific endopeptidase', *Infect Immun,* 60, (12), pp. 5219-23.

Cossart, P. and Jonquieres, R. (2000) 'Sortase, a universal target for therapeutic agents against gram-positive bacteria?', *Proc Natl Acad Sci U S A,* 97, (10), pp. 5013-5.

Courtney, H. S., Li, Y., Twal, W. O. and Argraves, W. S. (2009) 'Serum opacity factor is a streptococcal receptor for the extracellular matrix protein fibulin-1', *J Biol Chem,* 284, (19), pp. 12966-71.

Craig, L. (2009) 'Type IV Pilus Structure', in Jarrell, K. F.(ed), *Pili and Flagella - current Research and Future trends.* Caister Academic Press, pp. 1-17.

Craig, L. and Li, J. (2008) 'Type IV pili: paradoxes in form and function', *Curr Opin Struct Biol,* 18, (2), pp. 267-77.

Cunningham, M. W. (2000) 'Pathogenesis of group A streptococcal infections', *Clin Microbiol Rev,* 13, (3), pp. 470-511.

Dauter, Z. (2006) 'Estimation of anomalous signal in diffraction data', *Acta Crystallogr D Biol Crystallogr,* 62, (Pt 8), pp. 867-76.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. and Richardson, D. C. (2007) 'MolProbity: all-atom contacts and structure validation for proteins and nucleic acids', *Nucleic Acids Res,* 35, (Web Server issue), pp. W375-83.

Denny, F. W., Wannamaker, L. W., Brink, W. R., Rammelkamp, C. H., Jr. and Custer, E. A. (1950) 'Prevention of rheumatic fever; treatment of the preceding streptococcic infection', *J Am Med Assoc,* 143, (2), pp. 151-3.

Denny, F. W. J. (2000) *History of haemolytic streptococci and associated diseases. In Streptococcal infections* New York: Oxford University Press.

Dilks, K., Rose, R. W., Hartmann, E. and Pohlschroder, M. (2003) 'Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey', *J Bacteriol,* 185, (4), pp. 1478-83.

Dodds, A. W., Ren, X. D., Willis, A. C. and Law, S. K. (1996) 'The reaction mechanism of the internal thioester in the human complement component C4', *Nature,* 379, (6561), pp. 177-9.

Dodson, E. (2003) 'Is it jolly SAD?', *Acta Crystallogr D Biol Crystallogr,* 59, (Pt 11), pp. 1958-65.

Dramsi, S., Caliot, E., Bonne, I., Guadagnini, S., Prevost, M. C., Kojadinovic, M., Lalioui, L., Poyart, C. and Trieu-Cuot, P. (2006) 'Assembly and role of pili in group B streptococci', *Mol Microbiol,* 60, (6), pp. 1401-13.

Dramsi, S., Trieu-Cuot, P. and Bierne, H. (2005) 'Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria', *Res Microbiol,* 156, (3), pp. 289-97.

Driessen, A. J. and Nouwen, N. (2008) 'Protein translocation across the bacterial cytoplasmic membrane', *Annu Rev Biochem,* 77, pp. 643-67.

Duda, R. L. (1998) 'Protein chainmail: catenated protein in viral capsids', *Cell,* 94, (1), pp. 55-60.

Duguid, J. P., Smith, I. W., Dempster, G. and Edmunds, P. N. (1955) 'Non-flagellar filamentous appendages (fimbriae) and haemagglutinating activity in Bacterium coli', *J Pathol Bacteriol,* 70, (2), pp. 335-48.

Duong, F. and Wickner, W. (1997) 'Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme', *EMBO J,* 16, (10), pp. 2756-68.

Edwards, A. M., Manetti, A. G., Falugi, F., Zingaretti, C., Capo, S., Buccato, S., Bensi, G., Telford, J. L., Margarit, I. and Grandi, G. (2008) 'Scavenger receptor gp340 aggregates group A streptococci by binding pili', *Mol Microbiol,* 68, (6), pp. 1378-94.

Efstratiou, A. (2000) 'Group A streptococci in the 1990s', *J Antimicrob Chemother,* 45 Suppl, pp. 3-12.

El Mortaji, L., Terrasse, R., Dessen, A., Vernet, T. and Di Guilmi, A. M. (2010) 'Stability and assembly of pilus subunits of *Streptococcus pneumoniae*', *J Biol Chem*.

Emanuelsson, O., Brunak, S., von Heijne, G. and Nielsen, H. (2007) 'Locating proteins in the cell using TargetP, SignalP and related tools', *Nat Protoc,* 2, (4), pp. 953-71.

Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. and Liddington, R. C. (2000) 'Structural basis of collagen recognition by integrin alpha2beta1', *Cell,* 101, (1), pp. 47-56.

Emsley, P. and Cowtan, K. (2004) 'Coot: model-building tools for molecular graphics', *Acta Crystallogr D Biol Crystallogr,* 60, (Pt 12 Pt 1), pp. 2126-32.

Evans, P. (2006) 'Scaling and assessment of data quality', *Acta Crystallogr D Biol Crystallogr,* 62, (Pt 1), pp. 72-82.

Facklam, R., Beall, B., Efstratiou, A., Fischetti, V., Johnson, D., Kaplan, E., Kriz, P., Lovgren, M., Martin, D., Schwartz, B., Totolian, A., Bessen, D., Hollingshead, S., Rubin, F., Scott, J. and Tyrrell, G. (1999) 'emm typing and validation of provisional M types for group A streptococci', *Emerg Infect Dis,* 5, (2), pp. 247-53.

Falker, S., Nelson, A. L., Morfeldt, E., Jonas, K., Hultenby, K., Ries, J., Melefors, O., Normark, S. and Henriques-Normark, B. (2008) 'Sortase-mediated assembly and surface topology of adhesive pneumococcal pili', *Mol Microbiol,* 70, (3), pp. 595-607.

Falugi, F., Zingaretti, C., Pinto, V., Mariani, M., Amodeo, L., Manetti, A. G., Capo, S., Musser, J. M., Orefici, G., Margarit, I., Telford, J. L., Grandi, G. and Mora, M. (2008) 'Sequence variation in group A *Streptococcus* pili and association of pilus backbone types with lancefield T serotypes', *J Infect Dis,* 198, (12), pp. 1834-41.

Fernie-King, B. A., Seilly, D. J., Davies, A. and Lachmann, P. J. (2002) 'Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme', *Infect Immun,* 70, (9), pp. 4908-16.

Fernie-King, B. A., Seilly, D. J., Willers, C., Wurzner, R., Davies, A. and Lachmann, P. J. (2001) 'Streptococcal inhibitor of complement (SIC) inhibits the membrane attack complex by preventing uptake of C567 onto cell membranes', *Immunology,* 103, (3), pp. 390-8.

Ferretti, J. J., McShan, W. M., Ajdic, D., Savic, D. J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A. N., Kenton, S., Lai, H. S., Lin, S. P., Qian, Y., Jia, H. G., Najar, F. Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S. W., Roe, B. A. and McLaughlin, R. (2001) 'Complete genome sequence of an M1 strain of *Streptococcus pyogenes*', *Proc Natl Acad Sci U S A,* 98, (8), pp. 4658-63.

Fischetti, V. A. (1989) 'Streptococcal M protein: molecular design and biological behavior', *Clin Microbiol Rev,* 2, (3), pp. 285-314.

Fischetti, V. A., Pancholi, V. and Schneewind, O. (1990) 'Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci', *Mol Microbiol,* 4, (9), pp. 1603-5.

Fogg, G. C. and Caparon, M. G. (1997) 'Constitutive expression of fibronectin binding in *Streptococcus pyogenes* as a result of anaerobic activation of rofA', *J Bacteriol,* 179, (19), pp. 6172-80.

Fogg, G. C., Gibson, C. M. and Caparon, M. G. (1994) 'The identification of rofA, a positive-acting regulatory component of prtF expression: use of an m gamma delta-based shuttle mutagenesis strategy in *Streptococcus pyogenes*', *Mol Microbiol,* 11, (4), pp. 671-84.

Fontaine, M. C., Lee, J. J. and Kehoe, M. A. (2003) 'Combined contributions of streptolysin O and streptolysin S to virulence of serotype M5 *Streptococcus pyogenes* strain Manfredo', *Infect Immun,* 71, (7), pp. 3857-65.

Fraser, J. D. and Proft, T. (2008) 'The bacterial superantigen and superantigen-like proteins', *Immunol Rev,* 225, pp. 226-43.

Frick, I. M., Akesson, P., Rasmussen, M., Schmidtchen, A. and Bjorck, L. (2003) 'SIC, a secreted protein of *Streptococcus pyogenes* that inactivates antibacterial peptides', *J Biol Chem,* 278, (19), pp. 16561-6.

Gadjeva, M., Dodds, A. W., Taniguchi-Sidle, A., Willis, A. C., Isenman, D. E. and Law, S. K. (1998) 'The covalent binding reaction of complement component C3', *J Immunol,* 161, (2), pp. 985-90.

Garcia De La Torre, J., Huertas, M. L. and Carrasco, B. (2000) 'Calculation of hydrodynamic properties of globular proteins from their atomic-level structure', *Biophys J,* 78, (2), pp. 719-30.

Gaspar, A. H. and Ton-That, H. (2006) 'Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*', *J Bacteriol,* 188, (4), pp. 1526-33.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) *Protein Identification and Analysis Tools on the ExPASy Server. In The proteomics protocols handbook.*: Humana press, pp 571-607.

Gattiker, A., Bienvenut, W. V., Bairoch, A. and Gasteiger, E. (2002) 'FindPept, a tool to identify unmatched masses in peptide mass fingerprinting protein identification', *Proteomics,* 2, (10), pp. 1435-44.

Gerlach, R. G. and Hensel, M. (2007) 'Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens', *Int J Med Microbiol,* 297, (6), pp. 401-15.

Ghosh, J., Anderson, P. J., Chandrasekaran, S. and Caparon, M. G. (2010) 'Characterization of *Streptococcus pyogenes* beta-NAD+ glycohydrolase: re-evaluation of enzymatic properties associated with pathogenesis', *J Biol Chem,* 285, (8), pp. 5683-94.

Godehardt, A. W., Hammerschmidt, S., Frank, R. and Chhatwal, G. S. (2004) 'Binding of alpha2-macroglobulin to GRAB (Protein G-related alpha2-macroglobulin-binding protein), an important virulence factor of group A streptococci, is mediated by two charged motifs in the DeltaA region', *Biochem J,* 381, (Pt 3), pp. 877-85.

Gray, M. G. (1998) *Streptococcal Infections. In Bacterial Infections of Humans.*: Plenum Publishing Corporation, New York.

Griffith, F. (1934) 'The serological classification of *Streptococcus pyogenes*.', *Journal of Hygene (Cambridge),* 34, pp. 542-584.

Grosse-Kunstleve, R. W. and Adams, P. D. (2003) 'Substructure search procedures for macromolecular structures', *Acta Crystallogr D Biol Crystallogr,* 59, (Pt 11), pp. 1966-73.

Guttilla, I. K., Gaspar, A. H., Swierczynski, A., Swaminathan, A., Dwivedi, P., Das, A. and Ton-That, H. (2009) 'Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria', *J Bacteriol,* 191, (18), pp. 5603-12.

Hanahan, D. (1983) 'Studies on transformation of *Escherichia coli* with plasmids', *J Mol Biol,* 166, (4), pp. 557-80.

Hanski, E. and Caparon, M. (1992) 'Protein F, a fibronectin-binding protein, is an adhesin of the group A *Streptococcus Streptococcus pyogenes*', *Proc Natl Acad Sci U S A,* 89, (13), pp. 6172-6.

Hantke, K. and Braun, V. (1973) 'Covalent binding of lipid to protein. Diglyceride and amide-linked fatty acid at the N-terminal end of the murein-lipoprotein of the *Escherichia coli* outer membrane', *Eur J Biochem,* 34, (2), pp. 284-96.

Hendrickx, A. P., Bonten, M. J., van Luit-Asbroek, M., Schapendonk, C. M., Kragten, A. H. and Willems, R. J. (2008) 'Expression of two distinct types of pili by a hospital-acquired *Enterococcus faecium* isolate', *Microbiology,* 154, (Pt 10), pp. 3212-23.

Hendrickx, A. P., Willems, R. J., Bonten, M. J. and van Schaik, W. (2009) 'LPxTG surface proteins of enterococci', *Trends Microbiol,* 17, (9), pp. 423-30.

Hilleringmann, M., Giusti, F., Baudner, B. C., Masignani, V., Covacci, A., Rappuoli, R., Barocchi, M. A. and Ferlenghi, I. (2008) 'Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A', *PLoS Pathog,* 4, (3), pp. e1000026.

Hilleringmann, M., Ringler, P., Muller, S. A., De Angelis, G., Rappuoli, R., Ferlenghi, I. and Engel, A. (2009) 'Molecular architecture of *Streptococcus pneumoniae* TIGR4 pili', *EMBO J,* 28, (24), pp. 3921-30.

Hirose, I., Sano, K., Shioda, I., Kumano, M., Nakamura, K. and Yamane, K. (2000) 'Proteome analysis of Bacillus subtilis extracellular proteins: a two-dimensional protein electrophoretic study', *Microbiology,* 146 (Pt 1), pp. 65-75.

Holm, L. and Sander, C. (1995) 'Dali: a network tool for protein structure comparison', *Trends Biochem Sci,* 20, (11), pp. 478-80.

Holtje, J. V. and Tomasz, A. (1975) 'Specific recognition of choline residues in the cell wall teichoic acid by the N-acetylmuramyl-L-alanine amidase of Pneumococcus', *J Biol Chem,* 250, (15), pp. 6072-6.

Hondorp, E. R. and McIver, K. S. (2007) 'The Mga virulence regulon: infection where the grass is greener', *Mol Microbiol,* 66, (5), pp. 1056-65.

Hook, E. W., Wagner, R. R. and Lancefield, R. C. (1960) 'An epizootic in Swiss mice caused by a group A *Streptococcus*, newly designated type 50', *Am J Hyg,* 72, pp. 111-9.

Houwink, A. L. and van Iterson, W. (1950) 'Electron microscopical observations on bacterial cytology; a study on flagellation', *Biochim Biophys Acta,* 5, (1), pp. 10-44.

Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S. J., Scott, J. W., Dillard, B. D., Adams, M. W. and Tainer, J. A. (2009) 'Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS)', *Nat Methods,* 6, (8), pp. 606-12.

Hynes, W. L. and Walton, S. L. (2000) 'Hyaluronidases of Gram-positive bacteria', *FEMS Microbiol Lett,* 183, (2), pp. 201-7.

Ilangovan, U., Ton-That, H., Iwahara, J., Schneewind, O. and Clubb, R. T. (2001) 'Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*', *Proc Natl Acad Sci U S A,* 98, (11), pp. 6056-61.

Izore, T., Contreras-Martel, C., El Mortaji, L., Manzano, C., Terrasse, R., Vernet, T., Di Guilmi, A. M. and Dessen, A. (2010) 'Structural Basis of Host Cell Recognition by the Pilus Adhesin from *Streptococcus pneumoniae*', *Structure,* 18, (1), pp. 106-115.

Janssen, B. J., Huizinga, E. G., Raaijmakers, H. C., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B. and Gros, P. (2005) 'Structures of complement component C3 provide insights into the function and evolution of immunity', *Nature,* 437, (7058), pp. 505-11.

Jass, J., Schedin, S., Fallman, E., Ohlsson, J., Nilsson, U. J., Uhlin, B. E. and Axner, O. (2004) 'Physical properties of *Escherichia coli* P pili measured by optical tweezers', *Biophys J,* 87, (6), pp. 4271-83.

Johnson, D. R. and Kaplan, E. L. (1993) 'A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci', *J Med Microbiol,* 38, (5), pp. 311-5.

Jonquieres, R., Bierne, H., Fiedler, F., Gounon, P. and Cossart, P. (1999) 'Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria', *Mol Microbiol,* 34, (5), pp. 902-14.

Kang, H. J. and Baker, E. N. (2009) 'Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*', *J Biol Chem,* 284, (31), pp. 20729-37.

Kang, H. J. and Baker, E. N. (2010) 'Intramolecular isopeptide bonds: protein crosslinks built for stress?', *Trends Biochem Sci.*

Kang, H. J., Coulibaly, F., Clow, F., Proft, T. and Baker, E. N. (2007) 'Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure', *Science,* 318, (5856), pp. 1625-8.

Kang, H. J., Middleditch, M., Proft, T. and Baker, E. N. (2009a) 'Isopeptide bonds in bacterial pili and their characterization by X-ray crystallography and mass spectrometry', *Biopolymers,* 91, (12), pp. 1126-34.

Kang, H. J., Paterson, N. G., Gaspar, A. H., Ton-That, H. and Baker, E. N. (2009b) 'The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds', *Proc Natl Acad Sci U S A,* 106, (40), pp. 16967-71.

Kehoe, M. A. (1994) *Cell wall associated proteins in Gram positive bacteria. In Bacterial Cell Wall, pp 217-259.* Elsevier.

Kehoe, M. A., Kapur, V., Whatmore, A. M. and Musser, J. M. (1996) 'Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology', *Trends Microbiol,* 4, (11), pp. 436-43.

Kehoe, M. A., Miller, L., Walker, J. A. and Boulnois, G. J. (1987) 'Nucleotide sequence of the streptolysin O (SLO) gene: structural homologies between SLO and other membrane-damaging, thiol-activated toxins', *Infect Immun,* 55, (12), pp. 3228-32.

Kelly, S. M., Jess, T. J. and Price, N. C. (2005) 'How to study proteins by circular dichroism', *Biochim Biophys Acta,* 1751, (2), pp. 119-39.

Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. and Jones, T. A. (2001) *International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules.* Kluwer Academic Publishers, Dordrecht.

Kline, K. A., Dodson, K. W., Caparon, M. G. and Hultgren, S. J. (2010) 'A tale of two pili: assembly and function of pili in bacteria', *Trends Microbiol,* 18, (5), pp. 224-32.

Koch, M. H., Vachette, P. and Svergun, D. I. (2003) 'Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution', *Q Rev Biophys,* 36, (2), pp. 147-227.

Kocks, C., Gouin, E., Tabouret, M., Berche, P., Ohayon, H. and Cossart, P. (1992) 'L. monocytogenes-induced actin assembly requires the actA gene product, a surface protein', *Cell,* 68, (3), pp. 521-31.

Kozin, M. B. and Svergun, D. I. (2001) 'Automated matching of high- and low-resolution structural models', *Journal of Applied Crystallography,* 34, pp. 33-41.

Kratovac, Z., Manoharan, A., Luo, F., Lizano, S. and Bessen, D. E. (2007) 'Population genetics and linkage analysis of loci within the FCT region of *Streptococcus pyogenes*', *J Bacteriol,* 189, (4), pp. 1299-310.

Kreikemeyer, B., Beckert, S., Braun-Kiewnick, A. and Podbielski, A. (2002) 'Group A streptococcal RofA-type global regulators exhibit a strain-specific genomic presence and regulation pattern', *Microbiology,* 148, (Pt 5), pp. 1501-11.

Kreikemeyer, B., Gamez, G., Margarit, I., Giard, J. C., Hammerschmidt, S., Hartke, A. and Podbielski, A. (2010) 'Genomic organization, structure, regulation and pathogenic role of pilus constituents in major pathogenic Streptococci and Enterococci', *Int J Med Microbiol*.

Kreikemeyer, B., McIver, K. S. and Podbielski, A. (2003) 'Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions', *Trends Microbiol,* 11, (5), pp. 224-32.

Kreikemeyer, B., Nakata, M., Oehmcke, S., Gschwendtner, C., Normann, J. and Podbielski, A. (2005) '*Streptococcus pyogenes* collagen type I-binding Cpa surface protein. Expression profile, binding characteristics, biological functions, and potential clinical impact', *J Biol Chem,* 280, (39), pp. 33228-39.

Kreikemeyer, B., Oehmcke, S., Nakata, M., Hoffrogge, R. and Podbielski, A. (2004) '*Streptococcus pyogenes* fibronectin-binding protein F2: expression profile, binding characteristics, and impact on eukaryotic cell interactions', *J Biol Chem,* 279, (16), pp. 15850-9.

Krishnan, V., Gaspar, A. H., Ye, N., Mandlik, A., Ton-That, H. and Narayana, S. V. (2007) 'An IgG-like domain in the minor pilin GBS52 of *Streptococcus agalactiae* mediates lung epithelial cell adhesion', *Structure,* 15, (8), pp. 893-903.

Kumazawa, N. and Yanagawa, R. (1972) 'Chemical properties of the pili of *Corynebacterium renale*', *Infect Immun,* 5, (1), pp. 27-30.

Laemmli, U. K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature,* 227, (5259), pp. 680-5.

Lancefield, R. C. (1928a) 'The antigenic complex of *Streptococcus haemolyticus.* III. Chemical and immunological properties of the species specific substance.', *Journal of Experimental Medicine,* 47, pp. 481-489.

Lancefield, R. C. (1928b) 'The antigenic complexes of *Streptococcus haemolyticus.* I. Demonstration of a type specific substance. ', *Journal of Experimental Medicine,* 47, pp. 91-103.

Langer, G., Cohen, S. X., Lamzin, V. S. and Perrakis, A. (2008) 'Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7', *Nat Protoc,* 3, (7), pp. 1171-9.

LaPenta, D., Rubens, C., Chi, E. and Cleary, P. P. (1994) 'Group A streptococci efficiently invade human respiratory epithelial cells', *Proc Natl Acad Sci U S A,* 91, (25), pp. 12115-9.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics,* 23, (21), pp. 2947-8.

Lauer, P., Rinaudo, C. D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A. R., Mora, M., Rappuoli, R., Grandi, G. and Telford, J. L. (2005) 'Genome analysis reveals pili in Group B *Streptococcus*', *Science,* 309, (5731), pp. 105.

Law, J., Buist, G., Haandrikman, A., Kok, J., Venema, G. and Leenhouts, K. (1995) 'A system to generate chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes', *J Bacteriol,* 177, (24), pp. 7011-8.

Law, S. K. and Dodds, A. W. (1997) 'The internal thioester and the covalent binding properties of the complement proteins C3 and C4', *Protein Sci,* 6, (2), pp. 263-74.

Lebowitz, J., Lewis, M. S. and Schuck, P. (2002) 'Modern analytical ultracentrifugation in protein science: a tutorial review', *Protein Sci,* 11, (9), pp. 2067-79.

Lei, B., Mackie, S., Lukomski, S. and Musser, J. M. (2000) 'Identification and immunogenicity of group A *Streptococcus* culture supernatant proteins', *Infect Immun,* 68, (12), pp. 6807-18.

LeMieux, J., Hava, D. L., Basset, A. and Camilli, A. (2006) 'RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae* rlrA pathogenicity islet', *Infect Immun,* 74, (4), pp. 2453-6.

LeMieux, J., Woody, S. and Camilli, A. (2008) 'Roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus', *J Bacteriol,* 190, (17), pp. 6002-13.

Leslie, A. G. (2006) 'The integration of macromolecular diffraction data', *Acta Crystallogr D Biol Crystallogr,* 62, (Pt 1), pp. 48-57.

Li, N., Fort, F., Kessler, K. and Wang, W. (2009) 'Factors affecting cleavage at aspartic residues in model decapeptides', *J Pharm Biomed Anal,* 50, (1), pp. 73-8.

Linke, C., Young, P. G., Kang, H. J., Bunker, R. D., Middleditch, M. J., Caradoc-Davies, T. T., Proft, T. and Baker, E. N. (2010) 'Crystal structure of the minor pilin FctB reveals determinants of Group A streptococcal pilus anchoring', *J Biol Chem,* 285, (26), pp. 20381-9.

Lorand, L. and Graham, R. M. (2003) 'Transglutaminases: crosslinking enzymes with pleiotropic functions', *Nat Rev Mol Cell Biol,* 4, (2), pp. 140-56.

Luo, F., Lizano, S., Banik, S., Zhang, H. and Bessen, D. E. (2008a) 'Role of Mga in group A streptococcal infection at the skin epithelium', *Microb Pathog,* 45, (3), pp. 217-24.

Luo, F., Lizano, S. and Bessen, D. E. (2008b) 'Heterogeneity in the polarity of Nra regulatory effects on streptococcal pilus gene transcription and virulence', *Infect Immun,* 76, (6), pp. 2490-7.

Madden, J. C., Ruiz, N. and Caparon, M. (2001) 'Cytolysin-mediated translocation (CMT): a functional equivalent of type III secretion in gram-positive bacteria', *Cell,* 104, (1), pp. 143-52.

Magassa, N., Chandrasekaran, S. and Caparon, M. G. (2010) '*Streptococcus pyogenes* cytolysin-mediated translocation does not require pore formation by streptolysin O', *EMBO Rep,* 11, (5), pp. 400-5.

Maione, D., Margarit, I., Rinaudo, C. D., Masignani, V., Mora, M., Scarselli, M., Tettelin, H., Brettoni, C., Iacobini, E. T., Rosini, R., D'Agostino, N., Miorin, L., Buccato, S., Mariani, M., Galli, G., Nogarotto, R., Nardi Dei, V., Vegni, F., Fraser, C., Mancuso, G., Teti, G., Madoff, L. C., Paoletti, L. C., Rappuoli, R., Kasper, D. L., Telford, J. L. and Grandi, G. (2005) 'Identification of a universal Group B *Streptococcus* vaccine by multiple genome screen', *Science,* 309, (5731), pp. 148-50.

Maisey, H. C., Hensler, M., Nizet, V. and Doran, K. S. (2007) 'Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells', *J Bacteriol,* 189, (4), pp. 1464-7.

Mandlik, A., Das, A. and Ton-That, H. (2008a) 'The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria', *Proc Natl Acad Sci U S A,* 105, (37), pp. 14147-52.

Mandlik, A., Swierczynski, A., Das, A. and Ton-That, H. (2007) '*Corynebacterium diphtheriae* employs specific minor pilins to target human pharyngeal epithelial cells', *Mol Microbiol,* 64, (1), pp. 111-24.

Mandlik, A., Swierczynski, A., Das, A. and Ton-That, H. (2008b) 'Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development', *Trends Microbiol,* 16, (1), pp. 33-40.

Manetti, A. G., Zingaretti, C., Falugi, F., Capo, S., Bombaci, M., Bagnoli, F., Gambellini, G., Bensi, G., Mora, M., Edwards, A. M., Musser, J. M., Graviss, E. A., Telford, J. L., Grandi, G. and Margarit, I. (2007) '*Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation', *Mol Microbiol,* 64, (4), pp. 968-83.

Manzano, C., Contreras-Martel, C., El Mortaji, L., Izore, T., Fenel, D., Vernet, T., Schoehn, G., Di Guilmi, A. M. and Dessen, A. (2008) 'Sortase-mediated pilus fiber biogenesis in *Streptococcus pneumoniae*', *Structure,* 16, (12), pp. 1838-48.

Manzano, C., Izore, T., Job, V., Di Guilmi, A. M. and Dessen, A. (2009) 'Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens', *Biochemistry,* 48, (44), pp. 10549-57.

Maresso, A. W. and Schneewind, O. (2008) 'Sortase as a target of anti-infective therapy', *Pharmacol Rev,* 60, (1), pp. 128-41.

Margarit, I., Rinaudo, C. D., Galeotti, C. L., Maione, D., Ghezzo, C., Buttazzoni, E., Rosini, R., Runci, Y., Mora, M., Buccato, S., Pagani, M., Tresoldi, E., Berardi, A., Creti, R., Baker, C. J., Telford, J. L. and Grandi, G. (2009) 'Preventing bacterial infections with pilus-based vaccines: the group B *Streptococcus* paradigm', *J Infect Dis,* 199, (1), pp. 108-15.

Marraffini, L. A., Dedent, A. C. and Schneewind, O. (2006) 'Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria', *Microbiol Mol Biol Rev,* 70, (1), pp. 192-221.

Marraffini, L. A., Ton-That, H., Zong, Y., Narayana, S. V. and Schneewind, O. (2004) 'Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. A conserved arginine residue is required for efficient catalysis of sortase A', *J Biol Chem,* 279, (36), pp. 37763-70.

Martin, J. M. and Green, M. (2006) 'Group A *Streptococcus*', *Semin Pediatr Infect Dis,* 17, (3), pp. 140-8.

Mattick, J. S. (2002) 'Type IV pili and twitching motility', *Annu Rev Microbiol,* 56, pp. 289-314.

Mazmanian, S. K., Liu, G., Jensen, E. R., Lenoy, E. and Schneewind, O. (2000) '*Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections', *Proc Natl Acad Sci U S A,* 97, (10), pp. 5510-5.

Mazmanian, S. K., Liu, G., Ton-That, H. and Schneewind, O. (1999) '*Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall', *Science,* 285, (5428), pp. 760-3.

McCormick, J. K., Yarwood, J. M. and Schlievert, P. M. (2001) 'Toxic shock syndrome and bacterial superantigens: an update', *Annu Rev Microbiol,* 55, pp. 77-104.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) 'Phaser crystallographic software', *J Appl Crystallogr,* 40, (Pt 4), pp. 658-674.

McDonald, M., Currie, B. J. and Carapetis, J. R. (2004) 'Acute rheumatic fever: a chink in the chain that links the heart to the throat?', *Lancet Infect Dis,* 4, (4), pp. 240-5.

McGregor, K. F., Spratt, B. G., Kalia, A., Bennett, A., Bilek, N., Beall, B. and Bessen, D. E. (2004) 'Multilocus sequence typing of *Streptococcus pyogenes* representing most known emm types and distinctions among subpopulation genetic structures', *J Bacteriol,* 186, (13), pp. 4285-94.

McIver, K. S. and Scott, J. R. (1997) 'Role of mga in growth phase regulation of virulence genes of the group A *Streptococcus*', *J Bacteriol,* 179, (16), pp. 5178-87.

McShan, W. M., Ferretti, J. J., Karasawa, T., Suvorov, A. N., Lin, S., Qin, B., Jia, H., Kenton, S., Najar, F., Wu, H., Scott, J., Roe, B. A. and Savic, D. J. (2008) 'Genome sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*', *J Bacteriol,* 190, (23), pp. 7773-85.

Molinari, G., Rohde, M., Guzman, C. A. and Chhatwal, G. S. (2000) 'Two distinct pathways for the invasion of *Streptococcus pyogenes* in non-phagocytic cells', *Cell Microbiol,* 2, (2), pp. 145-54.

Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A. G., Maggi, T., Taddei, A. R., Grandi, G. and Telford, J. L. (2005) 'Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens', *Proc Natl Acad Sci U S A,* 102, (43), pp. 15641-6.

Mullis, K. B. (1990) 'Target amplification for DNA analysis by the polymerase chain reaction', *Ann Biol Clin (Paris),* 48, (8), pp. 579-82.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) 'Refinement of macromolecular structures by the maximum-likelihood method', *Acta Crystallogr D Biol Crystallogr,* 53, (Pt 3), pp. 240-55.

Nakagawa, J., Tamaki, S., Tomioka, S. and Matsuhashi, M. (1984) 'Functional biosynthesis of cell wall peptidoglycan by polymorphic bifunctional polypeptides. Penicillin-binding protein 1Bs of *Escherichia coli* with activities of transglycosylase and transpeptidase', *J Biol Chem,* 259, (22), pp. 13937-46.

Nakata, M., Podbielski, A. and Kreikemeyer, B. (2005) 'MsmR, a specific positive regulator of the *Streptococcus pyogenes* FCT pathogenicity region and cytolysin-mediated translocation system genes', *Mol Microbiol,* 57, (3), pp. 786-803.

Nallapareddy, S. R., Singh, K. V., Sillanpaa, J., Garsin, D. A., Hook, M., Erlandsen, S. L. and Murray, B. E. (2006) 'Endocarditis and biofilm-associated pili of *Enterococcus faecalis*', *J Clin Invest,* 116, (10), pp. 2799-807.

Navarre, W. W., Daefler, S. and Schneewind, O. (1996) 'Cell wall sorting of lipoproteins in *Staphylococcus aureus*', *J Bacteriol,* 178, (2), pp. 441-6.

Navarre, W. W. and Schneewind, O. (1994) 'Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria', *Mol Microbiol,* 14, (1), pp. 115-21.

Navarre, W. W. and Schneewind, O. (1999) 'Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope', *Microbiol Mol Biol Rev,* 63, (1), pp. 174-229.

Navarre, W. W., Ton-That, H., Faull, K. F. and Schneewind, O. (1998) 'Anchor structure of staphylococcal surface proteins. II. Cooh-terminal structure of muramidase and amidase-solubilized surface protein', *J Biol Chem,* 273, (44), pp. 29135-42.

Neiers, F., Madhurantakam, C., Falker, S., Manzano, C., Dessen, A., Normark, S., Henriques-Normark, B. and Achour, A. (2009) 'Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity', *J Mol Biol,* 393, (3), pp. 704-16.

Nelson, A. L., Ries, J., Bagnoli, F., Dahlberg, S., Falker, S., Rounioja, S., Tschop, J., Morfeldt, E., Ferlenghi, I., Hilleringmann, M., Holden, D. W., Rappuoli, R., Normark, S., Barocchi, M. A. and Henriques-Normark, B. (2007) 'RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*', *Mol Microbiol,* 66, (2), pp. 329-40.

Neylon, C. (2008) 'Small angle neutron and X-ray scattering in structural biology: recent examples from the literature', *Eur Biophys J,* 37, (5), pp. 531-41.

Nielsen, J. B. and Lampen, J. O. (1982) 'Glyceride-cysteine lipoproteins and secretion by Gram-positive bacteria', *J Bacteriol,* 152, (1), pp. 315-22.

Nizet, V. (2002) 'Streptococcal beta-hemolysins: genetics and role in disease pathogenesis', *Trends Microbiol,* 10, (12), pp. 575-80.

Nobbs, A. H., Lamont, R. J. and Jenkinson, H. F. (2009) '*Streptococcus* adherence and colonization', *Microbiol Mol Biol Rev,* 73, (3), pp. 407-50, Table of Contents.

Nobbs, A. H., Rosini, R., Rinaudo, C. D., Maione, D., Grandi, G. and Telford, J. L. (2008) 'Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in *Streptococcus agalactiae*', *Infect Immun,* 76, (8), pp. 3550-60.

Nooh, M. M., Aziz, R. K., Kotb, M., Eroshkin, A., Chuang, W. J., Proft, T. and Kansal, R. (2006) 'Streptococcal mitogenic exotoxin, SmeZ, is the most susceptible M1T1 streptococcal superantigen to degradation by the streptococcal cysteine protease, SpeB', *J Biol Chem,* 281, (46), pp. 35281-8.

Oh, S. Y., Budzik, J. M. and Schneewind, O. (2008) 'Sortases make pili from three ingredients', *Proc Natl Acad Sci U S A,* 105, (37), pp. 13703-4.

Olsen, A., Jonsson, A. and Normark, S. (1989) 'Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*', *Nature,* 338, (6217), pp. 652-5.

Osterlund, A. and Engstrand, L. (1995) 'Intracellular penetration and survival of *Streptococcus pyogenes* in respiratory epithelial cells in vitro', *Acta Otolaryngol,* 115, (5), pp. 685-8.

Osterlund, A. and Engstrand, L. (1997) 'An intracellular sanctuary for *Streptococcus pyogenes* in human tonsillar epithelium--studies of asymptomatic carriers and in vitro cultured biopsies', *Acta Otolaryngol,* 117, (6), pp. 883-8.

Osterlund, A., Popa, R., Nikkila, T., Scheynius, A. and Engstrand, L. (1997) 'Intracellular reservoir of *Streptococcus pyogenes* in vivo: a possible explanation for recurrent pharyngotonsillitis', *Laryngoscope,* 107, (5), pp. 640-7.

Paetzel, M., Dalbey, R. E. and Strynadka, N. C. (2000) 'The structure and mechanism of bacterial type I signal peptidases. A novel antibiotic target', *Pharmacol Ther,* 87, (1), pp. 27-49.

Palmer, M., Harris, R., Freytag, C., Kehoe, M., Tranum-Jensen, J. and Bhakdi, S. (1998) 'Assembly mechanism of the oligomeric streptolysin O pore: the early membrane lesion is lined by a free edge of the lipid membrane and is extended gradually during oligomerization', *EMBO J,* 17, (6), pp. 1598-605.

Pancholi, V. and Fischetti, V. A. (1992) 'A major surface protein on group A streptococci is a glyceraldehyde-3-phosphate-dehydrogenase with multiple binding activity', *J Exp Med,* 176, (2), pp. 415-26.

Pancholi, V. and Fischetti, V. A. (1998) 'alpha-enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci', *J Biol Chem,* 273, (23), pp. 14503-15.

Pangburn, M. K. (1992) 'Spontaneous thioester bond formation in alpha 2-macroglobulin, C3 and C4', *FEBS Lett,* 308, (3), pp. 280-2.

Papanikou, E., Karamanou, S. and Economou, A. (2007) 'Bacterial protein secretion through the translocase nanomachine', *Nat Rev Microbiol,* 5, (11), pp. 839-51.

Pelicic, V. (2008) 'Type IV pili: e pluribus unum?', *Mol Microbiol,* 68, (4), pp. 827-37.

Perry, A. M., Ton-That, H., Mazmanian, S. K. and Schneewind, O. (2002) 'Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring', *J Biol Chem,* 277, (18), pp. 16241-8.

Petoukhov, M. V. and Svergun, D. I. (2007) 'Analysis of X-ray and neutron scattering from biomacromolecular solutions', *Curr Opin Struct Biol,* 17, (5), pp. 562-71.

Pezzicoli, A., Santi, I., Lauer, P., Rosini, R., Rinaudo, D., Grandi, G., Telford, J. L. and Soriani, M. (2008) 'Pilus backbone contributes to group B *Streptococcus* paracellular translocation through epithelial cells', *J Infect Dis,* 198, (6), pp. 890-8.

Pichichero, M. E. and Casey, J. R. (2007) 'Systematic review of factors contributing to penicillin treatment failure in *Streptococcus pyogenes* pharyngitis', *Otolaryngol Head Neck Surg,* 137, (6), pp. 851-857.

Pickart, C. M. (2001) 'Mechanisms underlying ubiquitination', *Annu Rev Biochem,* 70, pp. 503-33.

Piszkiewicz, D., Landon, M. and Smith, E. L. (1970) 'Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations', *Biochem Biophys Res Commun,* 40, (5), pp. 1173-8.

Pizarro-Cerda, J. and Cossart, P. (2006) 'Bacterial adhesion and entry into host cells', *Cell,* 124, (4), pp. 715-27.

Podbielski, A. (2007) 'Flexible architecture of the *Streptococcus pyogenes* FCT genome region: finally the clue for understanding purulent skin diseases and long-term persistence?', *J Bacteriol,* 189, (4), pp. 1181-4.

Podbielski, A., Woischnik, M., Leonard, B. A. and Schmidt, K. H. (1999) 'Characterization of nra, a global negative regulator gene in group A streptococci', *Mol Microbiol,* 31, (4), pp. 1051-64.

Pointon, J. A., Smith, W. D., Saalbach, G., Crow, A., Kehoe, M. A. and Banfield, M. J. (2010) 'A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction', *J Biol Chem,* 285, (44), pp. 33858-66.

Proft, T. (2010) 'Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation', *Biotechnol Lett,* 32, (1), pp. 1-10.

Proft, T. and Baker, E. N. (2009) 'Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease', *Cell Mol Life Sci,* 66, (4), pp. 613-35.

Race, P. R., Bentley, M. L., Melvin, J. A., Crow, A., Hughes, R. K., Smith, W. D., Sessions, R. B., Kehoe, M. A., McCafferty, D. G. and Banfield, M. J. (2009) 'Crystal structure of *Streptococcus pyogenes* sortase A: implications for sortase mechanism', *J Biol Chem,* 284, (11), pp. 6924-33.

Receveur-Brechot, V., Bourhis, J. M., Uversky, V. N., Canard, B. and Longhi, S. (2006) 'Assessing protein disorder and induced folding', *Proteins,* 62, (1), pp. 24-45.

Rego, A. T., Chandran, V. and Waksman, G. (2010) 'Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis', *Biochem J,* 425, (3), pp. 475-88.

Rigel, N. W. and Braunstein, M. (2008) 'A new twist on an old pathway--accessory Sec [corrected] systems', *Mol Microbiol,* 69, (2), pp. 291-302.

Rocha, C. L. and Fischetti, V. A. (1999) 'Identification and characterization of a novel fibronectin-binding protein on the surface of group A streptococci', *Infect Immun,* 67, (6), pp. 2720-8.

Rooijakkers, S. H. and van Strijp, J. A. (2007) 'Bacterial complement evasion', *Mol Immunol,* 44, (1-3), pp. 23-32.

Rosch, J. and Caparon, M. (2004) 'A microdomain for protein secretion in Gram-positive bacteria', *Science,* 304, (5676), pp. 1513-5.

Rosch, J. W. and Caparon, M. G. (2005) 'The ExPortal: an organelle dedicated to the biogenesis of secreted proteins in *Streptococcus pyogenes*', *Mol Microbiol,* 58, (4), pp. 959-68.

Rosini, R., Rinaudo, C. D., Soriani, M., Lauer, P., Mora, M., Maione, D., Taddei, A., Santi, I., Ghezzo, C., Brettoni, C., Buccato, S., Margarit, I., Grandi, G. and Telford, J. L. (2006) 'Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*', *Mol Microbiol,* 61, (1), pp. 126-41.

Sandson, J., Hamerman, D., Janis, R. and Rojkind, M. (1968) 'Immunologic and chemical similarities between the *Streptococcus* and human connective tissue', *Trans Assoc Am Physicians,* 81, pp. 249-57.

Sarvas, M., Harwood, C. R., Bron, S. and van Dijl, J. M. (2004) 'Post-translocational folding of secretory proteins in Gram-positive bacteria', *Biochim Biophys Acta,* 1694, (1-3), pp. 311-27.

Sauer, F. G., Mulvey, M. A., Schilling, J. D., Martinez, J. J. and Hultgren, S. J. (2000) 'Bacterial pili: molecular mechanisms of pathogenesis', *Curr Opin Microbiol,* 3, (1), pp. 65-72.

Sauer, F. G., Pinkner, J. S., Waksman, G. and Hultgren, S. J. (2002) 'Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation', *Cell,* 111, (4), pp. 543-51.

Schneewind, O., Fowler, A. and Faull, K. F. (1995) 'Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*', *Science,* 268, (5207), pp. 103-6.

Schneewind, O., Mihaylova-Petkov, D. and Model, P. (1993) 'Cell wall sorting signals in surface proteins of gram-positive bacteria', *Embo J,* 12, (12), pp. 4803-11.

Schneewind, O., Model, P. and Fischetti, V. A. (1992) 'Sorting of protein A to the staphylococcal cell wall', *Cell,* 70, (2), pp. 267-81.

Schrager, H. M., Alberti, S., Cywes, C., Dougherty, G. J. and Wessels, M. R. (1998) 'Hyaluronic acid capsule modulates M protein-mediated adherence and acts as a ligand for attachment of group A *Streptococcus* to CD44 on human keratinocytes', *J Clin Invest,* 101, (8), pp. 1708-16.

Scott, J. R. and Zahner, D. (2006) 'Pili with strong attachments: Gram-positive bacteria do it differently', *Mol Microbiol,* 62, (2), pp. 320-30.

Sela, S., Marouni, M. J., Perry, R. and Barzilai, A. (2000) 'Effect of lipoteichoic acid on the uptake of *Streptococcus pyogenes* by HEp-2 cells', *FEMS Microbiol Lett,* 193, (2), pp. 187-93.

Simon, D. and Ferretti, J. J. (1991) 'Electrotransformation of *Streptococcus pyogenes* with plasmid and linear DNA', *FEMS Microbiol Lett,* 66, (2), pp. 219-24.

Sjobring, U., Pohl, G. and Olsen, A. (1994) 'Plasminogen, absorbed by *Escherichia coli* expressing curli or by Salmonella enteritidis expressing thin aggregative fimbriae, can be activated by simultaneously captured tissue-type plasminogen activator (t-PA)', *Mol Microbiol,* 14, (3), pp. 443-52.

Smith, W. D. (2004) *A functional genomic analysis of group A streptococcal virulence factors.* Ph.D Thesis thesis. University of Newcastle upon Tyne.

Smith, W. D., Pointon, J. A., Abbot, E., Kang, H. J., Baker, E. N., Hirst, B. H., Wilson, J. A., Banfield, M. J. and Kehoe, M. A. (2010) 'Roles of minor pilin subunits Spy0125 and Spy0130 in the serotype M1 *Streptococcus pyogenes* strain SF370', *J Bacteriol,* 192, (18), pp. 4651-9.

Solovyova, A. S., Pointon, J. A., Race, P. R., Smith, W. D., Kehoe, M. A. and Banfield, M. J. (2010) 'Solution structure of the major (Spy0128) and minor (Spy0125 and Spy0130) pili subunits from *Streptococcus pyogenes*', *Eur Biophys J,* 39, (3), pp. 469-80.

Starr, C. R. and Engleberg, N. C. (2006) 'Role of hyaluronidase in subcutaneous spread and growth of group A *Streptococcus*', *Infect Immun,* 74, (1), pp. 40-8.

Steer, A. C., Law, I., Matatolu, L., Beall, B. W. and Carapetis, J. R. (2009) 'Global emm type distribution of group A streptococci: systematic review and implications for vaccine development', *Lancet Infect Dis,* 9, (10), pp. 611-6.

Stevens, D. L. (2000) *Group A Beta-Hemolytic Streptococci: Virulence Factors, Pathogenesis, and Spectrum of Clinical Infections. In Streptococcal infections.* New York: Oxford University Press.

Stevens, D. L., Bisno, A. L., Chambers, H. F., Everett, E. D., Dellinger, P., Goldstein, E. J., Gorbach, S. L., Hirschmann, J. V., Kaplan, E. L., Montoya, J. G. and Wade, J. C. (2005) 'Practice guidelines for the diagnosis and management of skin and soft-tissue infections', *Clin Infect Dis,* 41, (10), pp. 1373-406.

Stevens, D. L., Tanner, M. H., Winship, J., Swarts, R., Ries, K. M., Schlievert, P. M. and Kaplan, E. (1989) 'Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A', *N Engl J Med,* 321, (1), pp. 1-7.

Stollerman, G. H. and Dale, J. B. (2008) 'The importance of the group a *Streptococcus* capsule in the pathogenesis of human infections: a historical perspective', *Clin Infect Dis,* 46, (7), pp. 1038-45.

Sundberg, E. J., Li, H., Llera, A. S., McCormick, J. K., Tormo, J., Schlievert, P. M., Karjalainen, K. and Mariuzza, R. A. (2002) 'Structures of two streptococcal superantigens bound to TCR beta chains reveal diversity in the architecture of T cell signaling complexes', *Structure,* 10, (5), pp. 687-99.

Suree, N., Liew, C. K., Villareal, V. A., Thieu, W., Fadeev, E. A., Clemens, J. J., Jung, M. E. and Clubb, R. T. (2009a) 'The structure of the *Staphylococcus aureus* sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized', *J Biol Chem,* 284, (36), pp. 24465-77.

Suree, N., Yi, S. W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M. E. and Clubb, R. T. (2009b) 'Discovery and structure-activity relationship analysis of *Staphylococcus aureus* sortase A inhibitors', *Bioorg Med Chem,* 17, (20), pp. 7174-85.

Suvorov, A. N. and Ferretti, J. J. (1996) 'Physical and genetic chromosomal map of an M type 1 strain of *Streptococcus pyogenes*', *J Bacteriol,* 178, (18), pp. 5546-9.

Svensson, M. D., Sjobring, U., Luo, F. and Bessen, D. E. (2002) 'Roles of the plasminogen activator streptokinase and the plasminogen-associated M protein in an experimental model for streptococcal impetigo', *Microbiology,* 148, (Pt 12), pp. 3933-45.

Svergun, D. I. (1999) 'Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing', *Biophys J,* 76, (6), pp. 2879-86.

Swaminathan, A., Mandlik, A., Swierczynski, A., Gaspar, A., Das, A. and Ton-That, H. (2007) 'Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*', *Mol Microbiol,* 66, (4), pp. 961-74.

Swierczynski, A. and Ton-That, H. (2006) 'Type III pilus of corynebacteria: Pilus length is determined by the level of its major pilin subunit', *J Bacteriol,* 188, (17), pp. 6318-25.

Taylor, G. (2003) 'The phase problem', *Acta Crystallogr D Biol Crystallogr,* 59, (Pt 11), pp. 1881-90.

Tendolkar, P. M., Baghdayan, A. S. and Shankar, N. (2006) 'Putative surface proteins encoded within a novel transferable locus confer a high-biofilm phenotype to *Enterococcus faecalis*', *J Bacteriol,* 188, (6), pp. 2063-72.

Terao, Y., Kawabata, S., Kunitomo, E., Nakagawa, I. and Hamada, S. (2002) 'Novel laminin-binding protein of *Streptococcus pyogenes*, Lbp, is involved in adhesion to epithelial cells', *Infect Immun,* 70, (2), pp. 993-7.

Terwilliger, T. C. (2000) 'Maximum-likelihood density modification', *Acta Crystallogr D Biol Crystallogr,* 56, (Pt 8), pp. 965-72.

Terwilliger, T. C. (2003) 'Automated main-chain model building by template matching and iterative fragment extension', *Acta Crystallogr D Biol Crystallogr,* 59, (Pt 1), pp. 38-44.

Terwilliger, T. C. and Berendzen, J. (1999) 'Automated MAD and MIR structure solution', *Acta Crystallogr D Biol Crystallogr,* 55, (Pt 4), pp. 849-61.

Tickle, I. J., Laskowski, R. A. and Moss, D. S. (1998) 'Rfree and the rfree ratio. I. Derivation of expected values of cross-validation residuals used in macromolecular least-squares refinement', *Acta Crystallogr D Biol Crystallogr,* 54, (Pt 4), pp. 547-57.

Tipper, D. J. and Strominger, J. L. (1965) 'Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine', *Proc Natl Acad Sci U S A,* 54, (4), pp. 1133-41.

Tjalsma, H., Bolhuis, A., Jongbloed, J. D., Bron, S. and van Dijl, J. M. (2000) 'Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome', *Microbiol Mol Biol Rev,* 64, (3), pp. 515-47.

Tjalsma, H., Zanen, G., Venema, G., Bron, S. and van Dijl, J. M. (1999) 'The potential active site of the lipoprotein-specific (type II) signal peptidase of Bacillus subtilis', *J Biol Chem,* 274, (40), pp. 28191-7.

Ton-That, H., Faull, K. F. and Schneewind, O. (1997) 'Anchor structure of staphylococcal surface proteins. A branched peptide that links the carboxyl terminus of proteins to the cell wall', *J Biol Chem,* 272, (35), pp. 22285-92.

Ton-That, H., Liu, G., Mazmanian, S. K., Faull, K. F. and Schneewind, O. (1999) 'Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif', *Proc Natl Acad Sci U S A,* 96, (22), pp. 12424-9.

Ton-That, H., Marraffini, L. A. and Schneewind, O. (2004) 'Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*', *Mol Microbiol,* 53, (1), pp. 251-61.

Ton-That, H., Mazmanian, S. K., Alksne, L. and Schneewind, O. (2002) 'Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis', *J Biol Chem,* 277, (9), pp. 7447-52.

Ton-That, H., Mazmanian, S. K., Faull, K. F. and Schneewind, O. (2000) 'Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates', *J Biol Chem,* 275, (13), pp. 9876-81.

Ton-That, H. and Schneewind, O. (2003) 'Assembly of pili on the surface of *Corynebacterium diphtheriae*', *Mol Microbiol,* 50, (4), pp. 1429-38.

Tsukiji, S. and Nagamune, T. (2009) 'Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering', *Chembiochem,* 10, (5), pp. 787-98.

Tuteja, R. (2005) 'Type I signal peptidase: an overview', *Arch Biochem Biophys,* 441, (2), pp. 107-11.

Vagin, A. and Teplyakov, A. (2010) 'Molecular replacement with MOLREP', *Acta Crystallogr D Biol Crystallogr,* 66, (Pt 1), pp. 22-5.

Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. and Rapoport, T. A. (2004) 'X-ray structure of a protein-conducting channel', *Nature,* 427, (6969), pp. 36-44.

van der Wolk, J. P., de Wit, J. G. and Driessen, A. J. (1997) 'The catalytic cycle of the *Escherichia coli* SecA ATPase comprises two distinct preprotein translocation events', *EMBO J,* 16, (24), pp. 7297-304.

van Wely, K. H., Swaving, J., Freudl, R. and Driessen, A. J. (2001) 'Translocation of proteins across the cell envelope of Gram-positive bacteria', *FEMS Microbiol Rev,* 25, (4), pp. 437-54.

von Heijne, G. (1990) 'The signal peptide', *J Membr Biol,* 115, (3), pp. 195-201.

von Heijne, G. and Abrahmsen, L. (1989) 'Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts', *FEBS Lett,* 244, (2), pp. 439-46.

von Pawel-Rammingen, U. and Bjorck, L. (2003) 'IdeS and SpeB: immunoglobulin-degrading cysteine proteinases of *Streptococcus pyogenes*', *Curr Opin Microbiol,* 6, (1), pp. 50-5.

Wang, P. Z. and Novick, R. P. (1987) 'Nucleotide sequence and expression of the beta-lactamase gene from *Staphylococcus aureus* plasmid pI258 in *Escherichia coli*, Bacillus subtilis, and *Staphylococcus aureus*', *J Bacteriol,* 169, (4), pp. 1763-6.

Wannamaker, L. W., Rammelkamp, C. H., Jr., Denny, F. W., Brink, W. R., Houser, H. B., Hahn, E. O. and Dingle, J. H. (1951) 'Prophylaxis of acute rheumatic fever by treatment of the preceding streptococcal infection with various amounts of depot penicillin', *Am J Med,* 10, (6), pp. 673-95.

Wessels, M. R. and Bronze, M. S. (1994) 'Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice', *Proc Natl Acad Sci U S A,* 91, (25), pp. 12238-42.

Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W. and Johnson, J. E. (2000) 'Topologically linked protein rings in the bacteriophage HK97 capsid', *Science,* 289, (5487), pp. 2129-33.

Wilson, C. D. and Salt, G. F. (1978) 'Streptococci in animal disease', *Soc Appl Bacteriol Symp Ser,* 7, pp. 143-56.

Wu, H. and Fives-Taylor, P. M. (2001) 'Molecular strategies for fimbrial expression and assembly', *Crit Rev Oral Biol Med,* 12, (2), pp. 101-15.

Yanagawa, R. and Honda, E. (1976) 'Presence of pili in species of human and animal parasites and pathogens of the genuscorynebacterium', *Infect Immun,* 13, (4), pp. 1293-5.

Yanagawa, R. and Otsuki, K. (1970) 'Some properties of the pili of *Corynebacterium renale*', *J Bacteriol,* 101, (3), pp. 1063-9.

Zahner, D. and Scott, J. R. (2008) 'SipA is required for pilus formation in *Streptococcus pyogenes* serotype M3', *J Bacteriol,* 190, (2), pp. 527-35.

Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. and Narayana, S. V. (2004) 'Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex', *J Biol Chem,* 279, (30), pp. 31383-9.