Imperial College
London

Proteomic and Bioinformatic Analysis of Surface Proteins in *Clostridium difficile*

Anne Elizabeth Wright

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy and the Diploma of Imperial College London

Centre for Molecular Microbiology and Infection
Department of Biological Sciences
Imperial College London

November 2006

BIBL LONDON Univ.

Declaration

I hereby declare that my thesis entitled "Proteomic and bioinformatic analysis of surface proteins in *Clostridium difficile*" is not the same as any that I have submitted for a degree or diploma or other qualification at any other university.

All data presented in this thesis are exclusively the result of my work, with the following exceptions: Mass spectrometry was carried out by Dr. Robin Wait at the Kennedy Institute, Imperial College, London and Dr. Jerry Thomas at the University of York; electron microscopy was carried out by Dave Goulding of Imperial College.

Signed..... Anne Wright .....

Date..... 21st December 2016 .....

Abstract

Clostridium difficile causes serious nosocomial infections that can lead to antibiotic associated diarrhoea and pseudomembranous colitis. Treatment of *C. difficile* infections can be problematic, with the disease recurring in many patients, and there is currently no vaccine available. The major effectors of tissue damage are the two toxins produced, TcdA and TcdB. However, toxigenic strains can show a range of virulence, and relatively little is known about how the bacterium colonises a host. To address this lack of information, a polyphasic approach was taken to survey the proteins present on the surface of *C. difficile* cells. The recently sequenced genome of strain 630 was searched using bioinformatic tools for genes with traits found in other virulence factors, resulting in the identification of 364 putative virulence genes. The complement of proteins present on the surface of *C. difficile* was mapped using proteomics, and changes in the proteins present under different conditions were compared. Mapping the surface proteome has identified several surface proteins for the first time, including eighteen proteins common to the bioinformatics search. Additionally, 45 proteins were found to be differentially expressed under different conditions. Extracts of *C. difficile* surface proteins were probed first using rabbit antiserum, and then using human antisera taken from patients suffering from *C. difficile* infection after an outbreak in a Dublin hospital. The immunoreactive proteins were identified, confirming their immunogenicity and *in vivo* expression. Compiling the results of these analyses shows that nine proteins are common to all the searches, and these proteins represent putative virulence factors and vaccine candidates. These proteins include the surface layer proteins and six surface layer protein paralogues, components of the flagella, ABC transporters and two proteins with unknown function. One of these proteins, Slp paralogue 4, was investigated further, although its putative function as a cysteine protease could not be confirmed.

Acknowledgements

First I would like to thank my supervisors, Dr. Neil Fairweather and Dr. Katy Brown, for giving me the opportunity to work on this project. You have given me endless advice, ideas, patience and encouragement, and also allowed me to broaden my horizons, try new things and earn the title of "Conference Queen". Thank-you!

I would also like to thank the numerous collaborators that have helped me: Robin Wait and Jerry Thomas for mass spec analysis at the eleventh hour; Ben Crossett and Judit Nagy for many discussions and tips about every aspect of proteomics; Nick Thomson and Mo Sebaihia for helping with the bioinformatics searches; Thomas Bertrand and Ben Moore for lots of patient advice and help with SAXS; Riccardo Bennett-Lovsey for assistance with protein structure modelling; Ben Moore (again!) for help and many explanations about CD; and Denise Drudy for providing the Dublin outbreak strain and the patient antisera.

My time at Imperial has been made all the more enjoyable by the members of the CMMI and the Fairweather, Brown, Frankel and Dougan lab members, past and present. In particular, I would like to thank Omar, Ema, Jenny, Robert, Ben C, Ben M, Karen, Alexandria, Emma, Gemma, Thomas, Kathy, Nandi, Liz, Anthony, Simon, Steve, Anne, and Del, for advice, good fun, lab banter and celebrations/commiserations in the Holland Club.

Finally, and most importantly, I would like to thank my friends and family, especially Jamie, and my Mum and Dad for all their love and support.

Contents

Declaration....2

Abstract....3

Acknowledgements....4

Contents....5

List of Figures....11

List of Tables....14

Abbreviations....16

Chapter 1 - Introduction....18

1.1. Clostridium difficile as a pathogen....19

1.1.1. Description and history....19

1.1.2. Clinical and pathological findings....20

1.1.3. Epidemiology and typing....22

1.1.4. Pathogenesis....24

1.1.4.1 C. difficile toxins....25

1.1.4.2 Further virulence factors....28

1.1.5 *C. difficile* treatments....29

1.2. The Gram-positive bacterial cell wall....31

1.2.1. The Gram-positive cell wall....31

1.2.1.1 Peptidoglycan....31

1.2.1.2 Secondary cell wall polymers....31

1.2.1.3 Proteins....33

1.2.1.4 Capsule....33

1.2.1.5 Flagella....33

1.2.1.6 Pili....34

1.2.2. Protein secretion and attachment to the cell wall....34

1.2.2.1 Signal peptide....34

1.2.2.2 Sec secretion pathway....36

1.2.2.3 Twin arginine translocation....39

1.2.2.4 ATP binding cassette transporters....39

1.2.2.5 Flagella export apparatus....40

1.2.2.6 Sortase....42

1.2.2.7 Lipoproteins....44

1.2.2.8 Frequency of use of secretion routes....44

1.2.3. The S-layer....45

1.2.3.1. Biochemistry of the S-layer....47

1.2.3.2 Genetics of the S-layer....50

1.2.3.3 Attachment 51
1.2.3.4 Evolution and functional roles 53
1.3. The cell wall of *C. difficile* 55
1.4. Proteomics 58
1.5 Project aims 62
Chapter 2 - Materials and Methods 63
2.1 Materials 64
2.1.1 Bacterial strains and vectors 64
2.1.2 Chemical reagents 67
2.1.3 Media 67
2.1.4 Buffers 67
2.2 General procedures for DNA manipulation 70
2.2.1 Plasmid purification 70
2.2.2 Polymerase Chain Reaction 70
2.2.3 Agarose gel electrophoresis 70
2.2.4 Purification of DNA from agarose gels 72
2.2.5 Restriction endonuclease digestion of DNA 72
2.2.6 Ligation 72
2.2.7 DNA sequencing 72
2.2.8 Preparation of electrocompetent *E. coli* 72
2.2.9 Electroporation of *E. coli* 73
2.2.10 TOPO cloning reaction and chemical transformation 73
2.3 Expression and purification of recombinant proteins 73
2.3.1 Growth and induction 73
2.3.2 Native 6 x His-tag protein purification 74
2.3.3 Denaturing 6 x His-tag protein purification 74
2.3.4 Gel filtration chromatography 75
2.4 Protein characterisation 75
2.4.1 Polyacrylamide electrophoresis (PAGE) and Coomassie staining 75
2.4.2 Bicinchoninic acid (BCA) protein concentration assay 75
2.4.3 Western blots 76
2.4.4 Anti-*C. difficile* surface protein antibody production 76
2.4.5 Circular dichroism analysis 77
2.4.6 Small angle X-ray scattering analysis 77
2.5 Cysteine protease activity assays 78
2.5.1 Zymograms 78
2.5.2 Milk degradation 79
2.5.3 Azocoll degradation 79

2.5.4 Gelatinase activity 79
2.5.5 *N*-Benzoyl-DL-arginine 4-nitroanilide hypochloride (BAPNA) degradation ... 79
2.6 Extraction of *C. difficile* SLPs 80
2.6.1 Glycine extraction of the SLPs 80
2.6.2 Extraction of the cell wall component of *C. difficile* cells 80
2.6.3 Total protein extraction from *C. difficile cells* 81
2.6.4 Total protein concentration assay 81
2.6.5 Assay for cell lysis: isocitrate dehydrogenase activity 81
2.6.6 Assay for cell lysis: glyceraldehyde 3-phosphate dehydrogenase activity ... 81
2.7 2D SDS-PAGE 82
2.7.1 First dimension: isoelectric focusing 82
2.7.2 Second dimension 82
2.7.3 Colloidal Coomassie staining 83
2.7.4 Silver staining 83
2.7.5 Ruthenium II tris bathophenanthroline disulphonate (RuBPS) staining 83
2.7.6 Computational analysis of results using Phoretix 84
2.8 Mass spectrometry 85
2.8.1 In-gel trypsin digest 85
2.8.2 MALDI-TOF analysis of trypsin digested proteins 86
2.8.3 Tandem electrospray mass spectrometry 86
2.9 Bioinformatic analyses 87
Chapter 3 - Bioinformatic analysis of putative virulence factors in the Clostridium difficile genome 90
3.1 Introduction 91
3.1.1 Bioinformatics as a tool for the analysis of genome data 91
3.1.2 The *C. difficile* genome 93
3.2 Results: Bioinformatic analysis of the *C. difficile* genome 94
3.2.1 Selection of virulence CDSs 96
3.2.2 Filtering criteria: housekeeping functions 97
3.2.2.1 ABC transporters/permeases 97
3.2.2.2 Two component regulatory systems and signalling proteins 98
3.2.2.3 Phage/transposons 101
3.2.2.4 Sporulation 101
3.2.3 Filtering criteria: unknown proteins 102
3.2.3.1 ORFans 102
3.2.3.2 Hypothetical proteins 102
3.2.3.3 Putatively assigned proteins 103
3.2.4 Possible virulence factors with putatively assigned functions 103

3.2.4.1 Autolysins 103
3.2.4.2 Flagella and chemotaxis 112
3.2.4.3 Pili 112
3.2.4.4 Drug resistance 112
3.2.4.5 Lantibiotics 118
3.2.4.6 Oligopeptide transporters 119
3.2.4.7 Proteases 119
3.2.4.8 Toxins 122
3.2.4.9 Adhesins 122
3.2.4.10 Sortase and sortase substrates 122
3.2.4.11 Iron scavenging 123
3.2.4.12 Quorum sensing 123
3.2.4.13 Miscellaneous CDSs 125
3.3 Discussion 126
Chapter 4 - The surface proteome of *Clostridium difficile* and structural studies of the surface layer proteins 129
4.1. Introduction 130
4.2 Results:The surface proteome of *C. difficile* 132
4.2.1 Methods for extracting SLPs from *C. difficile* 132
4.2.2 Cell lysis assays 138
4.2.2.1 Isocitrate dehydrogensae activity 140
4.2.2.2 Glyceraldehyde 3-phosphate dehydrogenase activity 140
4.2.3 Mapping the surface proteome 140
4.3 Structural studies of the SLPs 161
4.3.1 Expression and purification of SLPs 162
4.3.2 Analysis of native SLPs removed by glycine treatment 165
4.3.3 *In silico* prediction of SLP secondary structure. 165
4.3.4 Circular dichroism (CD) 169
4.3.5 Small Angle X-ray Scattering analysis 169
4.3.3 Crystallisation trials of the LMW SLP 173
4.4 Discussion 172
Chapter 5 - Proteomic analysis of the differential expression of *Clostridium difficile* surface proteins grown under a variety of conditions 178
5.1 Introduction 179
5.1.1 Growth phase induced changes in the proteome 179
5.1.2 Stress-induced changes in the proteome 180
5.2 Results: Proteins differentially expressed on the surface of *C. difficile* at 3 growth phases 181

5.2.1 Sample preparation and SDS-PAGE 181
5.2.2 Analysis of 2D gels 183
5.2.3 Differential expression seen within glycine extracts 183
5.2.4 Differential expression seen within lysozyme extracts 187
5.2.4.1 Analysis of lysozyme extracts using zoom gels 190
5.3. Effect of antibiotics on differentially expressed proteins in *C. difficile* 203
5.3.1 Determining the minimal inhibition concentration of the antibiotics 203
5.3.2 Sample preparation and SDS-PAGE 203
5.3.3 Analysis of 2D gels 204
5.3.4 Differential expression seen within glycine extracts 204
5.3.5 Differential expression seen within lysozyme extracts 209
5.4. Results: Effect of heat-shock on differentially expressed proteins 214
5.4.1 Sample preparation and SDS-PAGE 214
5.4.2 Analysis of 2D gels 216
5.4.3 Differential expression seen within glycine extracts 216
5.4.4 Differential expression seen with in lysozyme extracts 222
5.5 Discussion 229
Chapter 6 - The immunoproteome of *Clostridium difficile* 234
6.1 Introduction 235
6.2 Results: The immunoproteome of *C. difficile* recognised by rabbit antiserum .. 236
6.2.1 Generation of antibodies to *C. difficile* surface proteins 236
6.2.2 1D SDS-PAGE and Western blots 236
6.2.3 2D SDS-PAGE and Western blots 239
6.3 The immunoproteome of *C. difficile* recognised by human patient sera 248
6.3.1 Background 248
6.3.2 1D SDS-PAGE and Western blot 248
6.3.3 2D SDS-PAGE and Western blots 251
6.4 Discussion 260
Chapter 7 - Studies of Slp paralogue 4 (Cwp84) 263
7.1 Introduction 264
7.1.1 Slp paralogue 4 264
7.1.2 Cysteine proteases 265
7.2 Sequence analysis of Slp paralogue 4 268
7.2.1 Sequence similarities to other cysteine proteases 268
7.2.2 Modelling the Slp paralogue 4 structure 270
7.3 Cloning, expression and purification of Slp paralogue 4 270
7.3.1 Construct pEC29(7) 270
7.3.2 Cloning and expression of full length Slp paralogue 4 270

7.4 Proteolytic activity of Orf4CHis1 274
7.4.1 Zymograms 274
7.4.2 Casein degradation assay 274
7.4.3 Gelatine degradation activity 277
7.4.4 Azocoll degradation 277
7.4.5 BAPNA assay 281
7.4.6 Conincubation assay with SlpA 281
7.5 Discussion 285
Chapter 8 - Discussion 288
8.1 Introduction 289
8.2 Complementary approaches to identify *C. difficile* virulence factors and vaccine candidates 290
8.3 Responses of *C. difficile* to stress 293
8.4 Functional studies of Slp paralogue 4 294
References 295

List of Figures

Figure 1.1 Gross pathology and histology of normal colon and colon with pseudomembranous enterocolitis.....21

Figure 1.2 A. The pathogenicity locus of *C. difficile*
B. Molecular structure of C. *difficile* toxins.....26

Figure 1.3 Generalised structure of Gram-positive peptidoglycan.....32

Figure 1.4 Model of signal peptide insertion into the cytoplasmic membrane and cleavage by SPase.....35

Figure 1.5 Sec factors and the translocation processes.....37

Figure 1.6 A. Ribbon diagram of a typical ABC transporter, BtuCD
B. Schematic mechanism of transport.....41

Figure 1.7 Sortase - mediated protein anchoring in *S. aureus*.....43

Figure 1.8 Electron micrographs of S-layers.....46

Figure 1.9 Schematic of different S-layer lattice types.....48

Figure 1.10 Electron micrograph and schematic of prokaryotic S-layers.....52

Figure 1.11 Arrangement of ORFs in the region of *slpA* of *C.difficile*.....56

Figure 3.1 Strategy for selection of virulence factors in the *C. difficile* genome.....95

Figure 3.2 Artemis view showing cluster of genes thought to be involved in pilus biogenesis.....113

Figure 3.3 Artemis views showing CDS involved in drug resistance.....117

Figure 3.4 Artemis views showing CDS involved in lantibiotic resistance.....120

Figure 3.5 Mechanism of quorum sensing by the accessory gene regulator (Agr) system.....124

Figure 4.1 Extraction of S-layer and associated proteins from *C. difficile* 630 and analysis by SDS-PAGE.....133

Figure 4.2 Schematic of cell fractionation procedure.....135

Figure 4.3 Structure of a typical Gram-positive peptidoglycan with cleavage sites for degradative enzymes indicated.....136

Figure 4.4 SDS-PAGE of *C. difficile* 630 after fractionation.....137

Figure 4.5 Strains of *C. difficile* after treatment with digestion buffer viewed under the microscope.....139

Figure 4.6 2-DE of SLPs and associated proteins extracted from *C. difficile* by glycine treatment.....143

Figure 4.7 2-DE of proteins extracted by lysozyme treatment.....144

Figure 4.8 2-DE and identification of protein spots extracted from *C. difficile* by glycine treatment.....145

Figure 4.9 Examples of MS spectra and protein coverage of the LMW SLP and HMW SLPs.....147
Figure 4.10 Position and peptide coverage of SLP paralogue 15 on 2-DE gel of proteins extracted from *C. diffcile* by glycine treatment.....155
Figure 4.11 Expression and purification of LMW SLP.....163
Figure 4.12 Expression and purification of HMW SLP.....164
Figure 4.13 Size exclusion chromatography of SLPs extracted from *C. difficile* by glycine treatment.....166
Figure 4.14 PAGE analysis of glycine extracted SLPs.....167
Figure 4.15 Secondary structure prediction of the SLPs.....168
Figure 4.16 Circular dichroism analysis of the recombinant SLPs.....170
Figure 4.17 SAXS - based shape predictions for the SLPs.....172
Figure 5.1. SDS-PAGE analysis of surface proteins extracted from a time-course culture of *C. difficile* 630 by glycine and lysozyme treatment.....182
Figure 5.2 The construction of a Phoretix experiment.....184
Figure 5.3 Average gels displaying proteins extracted from the surface of a growing culture of *C. difficile* 630 by glycine treatment.....186
Figure 5.4 Histograms of identified protein spots changing over time.....189
Figure 5.5 2D PAGE analysis of proteins extracted from the surface of a growing culture *C. difficile* 630 by lysozyme treatment.....191
Figure 5.6 Histogram of identified protein spots changing over time.....195
Figure 5.7 Narrow range 2D PAGE analysis of proteins extracted from the surface of a growing culture *C. difficile* 630 by lysozyme treatment.....196
Figure 5.8 Histograms of spot volumes of identified proteins changing in expression levels during growth.....201
Figure 5.9 SDS-PAGE analysis of surface proteins extracted by glycine and lysozyme treatment from cultures of *C. difficile* 630 treated with antibiotics.....205
Figure 5.10 2D PAGE analyses of proteins extracted by glycine treatment from the surface of a culture *C. difficile* 630 treated with antibiotics.....206
Figure 5.11 2D PAGE analyses of proteins extracted by lysozyme treatment from the surface of a culture *C. difficile* 630 treated with antibiotics.....210
Figure 5.12 Histograms showing the change in spot volumes of identified proteins from cultures treated with antibiotics.....213
Figure 5.13 SDS-PAGE analysis of surface proteins extracted by glycine and lysozyme treatment from cultures of *C. difficile* 630 that have undergone heat-shock treatment.....215

Figure 5.14 2D PAGE analyses of proteins extracted by glycine treatment from the surface of a culture *C. difficile* 630 before and after heat-shock.....217
Figure 5.15 Histogram of the change in spot volume for proteins differentially expressed after heat-shock.....221
Figure 5.16 2D PAGE analyses of proteins extracted by lysozyme treatment from the surface of a culture *C. difficile* 630 before and after heat-shock.....223
Figure 5.17 Histograms of spot volumes of identified proteins changing in expression levels in heat-shocked bacteria.....227
Figure 6.1 EM images of immuno-gold labelling of antiserum raised against *C. difficile* 630.....237
Figure 6.2 SDS-PAGE and Western blot of analysis of *C. difficile* proteins using rabbit antibodies.....238
Figure 6.3 2-DE and Western blot of glycine extracted *C. difficile* surface proteins.....240
Figure 6.4 2-DE and Western blot of lysozyme extracted *C. difficile* surface proteins.....242
Figure 6.5 2-DE and Western blot of total proteins from *C. difficile*.....244
Figure 6.6 SDS-PAGE and Western blot of *C. difficile* strain M9 using human patient antibodies.....249
Figure 6.7 2D Western blots of lysozyme cell surface extracts from *C. difficile* strain M9 using human patient antisera.....252
Figure 6.8 2-D SDS-PAGE of lysozyme extracted cell surface proteins from *C. difficile* M9 with identified seroreactive spots.....254
Figure 7.1 Mechanism of action of cysteine proteases.....267
Figure 7.2 Sequence alignment of the protease region of Slp paralogue 4 with other cysteine proteases.....269
Figure 7.3 Views of the predicted structure of Slp paralogue 4.....271
Figure 7.4 SDS-PAGE (A) and Western blot (B) of *E. coli* cell lysates induced to express Orf4CHis1.....273
Figure 7.5 Native (A) and denaturing (B) purification of Orf4CHis1 on a batch Nickel-NTA column.....275
Figure 7.6 Typical zymograms testing for proteolytic activity of Orf4CHis1.....276
Figure 7.7 Caseinase activity determined by growth on agar plates supplemented with skimmed milk.....278
Figure 7.8 Gelatinase activity.....279
Figure 7.9 Coincubation assay with SlpA and Orf4CHis1.....283

List of Tables

Table 1.1 Characteristic chemical properties of S-layers.....49

Table 2.1 *E. coli* strains used in this study and their genotypes.....65

Table 2.2 Plasmid vectors used in this study and their properties.....66

Table 2.3 Sequence of the oligonucleotides used.....71

Table 3.1 Virulence related ABC transporter components.....99

Table 3.2 Virulence related two component signal transduction proteins.....101

Table 3.3 Putative virulence factors with unknown function.....104

Table 3.4 Autolysins in the *C. difficile* 630 genome.....109

Table 3.5 Putative virulence factors involved in drug resistance.....115

Table 3.6 Possible virulence factors with protease domains.....121

Table 4.1 Assays for detection of cell lysis during surface protein extraction from *C. difficile* cells.....141

Table 4.2 Summary of results of MALDI-TOF protein identifications from proteins extracted from *C. difficile* by glycine treatment.....146

Table 4.3 Proteins identified by MS/MS from surface proteins extracted from *C. difficile* 630 by glycine treatment.....150

Table 4.4 Proteins identified by MS/MS from surface proteins extracted from *C. difficile* 630 by lysozyme digestion treatment.....156

Table 4.5 Secondary structure analysis from CD data for the recombinant SLPs.....171

Table 4.6 SAXS data collection.....171

Table 5.1 Proteins identified from 2D gels of glycine extracted proteins from cultures at different stages of growth.....188

Table 5.2 Proteins identified from 2D gels of lysozyme extracted proteins from cultures at different stages of growth.....193

Table 5.3 Proteins identified from Zoom 2D gels of lysozyme extracted proteins from cultures at different stages of growth.....199

Table 5.4 Proteins identified from 2D gels of glycine extracted proteins from cultures treated with antibiotics.....208

Table 5.5 Proteins identified from 2D gels of lysozyme extracted proteins from cultures treated with antibiotics.....212

Table 5.6 Proteins identified from 2D gels of glycine extracted proteins from heat-shocked cultures.....219

Table 5.7 Proteins identified from 2D gels of lysozyme extracted proteins from heat-shocked cultures.....225

Table 6.1 Proteins identified from 2D gels of glycine extracted proteins recognised by rabbit antiserum to surface proteins.....245

Table 6.2 Proteins identified from 2D gels of lysozyme extracted proteins recognised by rabbit antiserum to surface proteins.....246
Table 6.3 Proteins identified from 2D gels of total protein of *C. difficile* recognised by rabbit antiserum to surface proteins.....247
Table 6.4 Proteins identified from 2D gels of total protein of *C. difficile* M9 recognised by rabbit antiserum to surface proteins.....256
Table 7.1 Results of the azocoll degradation experiment.....280
Table 7.2 BAPNA assay.....282

Abbreviations

1-D.....One dimensional

2-D.....Two dimensional

3-D.....Three dimensional

A_{600}.....Absorbance at 600 nm

AAD.....Antibiotic associated diarrhoea

ABC.....ATP binding cassette

ACN.....Acetonitrile

ANOVA.....Analysis of variance

ATP.....Adenosine triphosphate

BAPNA.....*N*-Benzoyl-DL-arginine 4-nitroanilide hypochloride

BLAST...Basic local alignment search tool

BSA.....Bovine serum albumin

CD.....Circular dichroism

CDAD.....*Clostridium difficile* associated diarrhoea

CDS.....Coding sequence

CHAPS.....3-Cholamidopropyl dimethylammonio-propanesulfonic acid

CROP.Clostridial repetitive oligopeptide

CTns.....Conjugative transposons

DIGE..Differential in gel electrophoresis

DNA.....Deoxyribonucleic acid

dNTP.....Deoxyribonucleotide triphosphate

DTT.....Dithiothreitol

ECL.....Enhanced chemiluminescence

ECM.....Extracellular matrix

EDTA...Ethylenediaminetetraacetic acid

ELISA.....Enzyme-linked immunosorbent assay

ESI.....Electrospray ionisation

FACS.....Fluorescence Activated Cell Sorting

FPLC.....Fast protein liquid chromatography

GST.....Glutathione *S*-transferase

GTP.....Guanosine triphosphate

h.....Hours

HEPES...4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

HMW.....High molecular weight

HSP.....High scoring pairs

IPG.....Immobilised pH gradient

IPTG.....Isopropyl-beta-D-thiogalactopyranoside

LB.....Luria bertani broth

LD_{50}.....Lethal dose 50%

LMW.....Low molecular weight

LPS.....Lipopolysaccharide

MALDI-TOF.....Matrix assisted laser desorption/ionisation time of flight

MBP.....Maltose binding protein

MIC.....Minimal inhibition concentration

MS.....Mass spectrometry

MSD.....Membrane spanning domain

m/z.....Mass to charge ratio

NAD.Nicotinamide adenine dinucleotide

ORF.....Open reading frame

PAGE.....Polyacrylamide gel electrophoresis

PBS.....Phosphate buffered saline

PCR.....Polymerase chain reaction

pI.....Isoelectric point

PMC.....Pseudomembranous colitis

RFLP.....Restriction fragment length polymorphism

RNA.....Ribonucleic acid

RT-PCR.....Reverse transcriptase polymerise chain reaction

RuBPS.....Ruthenium II tris bathophenanthroline disulphonate

s.....Seconds

SCWP.....Secondary cell wall polymer

SDS.....Sodium dodecyl sulphate

S-layer.....Surface layer

SLH.....S-layer homology

SLP.....Surface layer protein

SPase.....Signal peptidase

SPPase.....Signal peptide peptidase

SRP.....Signal recognition particle

TAE.....Tris acetate EDTA buffer

TAT.....Twin arginine translocation

TEMED.....Tetramethylethylenediamine

TFA.....Trifluoroacetic acid

T_m.....Melting temperature

UDP.....Uridine triphosphate

UK.....United Kingdom

UV.....Ultra violet light

v/v.....Volume per volume

w/v.....Weight per volume

Chapter 1

Introduction

1.1. Clostridium difficile as a pathogen

Clostridia (from the Latin *clostridium* for spindle) are anaerobic, spore-forming bacilli that include a number of species pathogenic to humans. Best known amongst these are the agents for botulism (*Clostridium botulinum*), tetanus (*Clostridium tetani*) and gas gangrene (several species, mostly *Clostridium perfringens*). *Clostridium difficile* is another member of this genus, and is linked with antibiotic associated diarrhoea (AAD) and pseudomembranous colitis (PMC), whose incidence has increased markedly in recent years.

1.1.1. Description and history

C. difficile occurs in the shape of straight rods, 0.5-1.9 µm in width and 3.0-16.9 µm in length. It is motile with peritrichous flagella and typically forms oval, sub-terminal spores. Although young cultures are uniformly Gram-positive, 24-48 hour cultures may appear Gram-negative. *C. difficile* was originally discovered in 1935 in the intestinal flora of neonates and was given its name because it was difficult to isolate and slow growing in culture (Hall & O'Toole, 1935). It was initially believed to be a non-pathogenic organism, since it appeared to be a commensal in infants, despite cell-free culture supernatants being highly lethal to experimental animals. *C. difficile* is widespread in the environment, and is found in the stools of a number of animal species including horses, cats and dogs (Riley *et al.*, 1991; Baverud *et al.*, 2003).

It was not until the 1970s that *C. difficile* was associated with a human disease, pseudomembranous colitis (PMC). This condition had been described nearly a century before as a fatal complication in a young woman that had undergone major abdominal surgery, and given its name from the observation at autopsy of a 'diphtheric membrane' in the small bowel (Finney, 1893). In the 1950s, shortly after the introduction of widespread use of antibiotics in clinical practice, PMC was recognised as a relatively common complication of treatments with broad-spectrum antibiotics. In particular, a high rate of PMC (10%) was reported in patients treated with clindamycin, a protein synthesis inhibitor of the lincomycin group commonly used in the treatment of anaerobic infections (Tedesco, 1974). The link between PMC and *C. difficile* was proven through parallel studies of human patients and a rodent model that showed faecal filtrates from patients or sick animals were cytotoxic on tissue culture cells (Green, 1974; Larson *et al.*, 1978), and that this effect could be neutralised by a polyvalent antiserum against *Clostridium sordellii* toxin (Rifkin, 1977; Chang *et al.*, 1978). However, *C. difficile* rather than *C. sordellii* was found to be the dominant species in the intestinal contents of both PMC patients (Bartlett, 1978a; George, 1978;

Larson *et al.*, 1978) and antibiotic-treated animals (Bartlett, 1977, 1978b) and was responsible for the production of the toxin cytopathic to tissue culture cells. The neutralising effect of *C. sordellii* anti-toxin was shown to be due to cross-reaction (Bartlett *et al.*, 1978).

Following *C. difficile*'s original identification as the causative agent of PMC, it soon became clear that the organism is a frequent cause of antibiotic-associated diarrhoea (AAD), defined as otherwise unexplained diarrhoea that occurs with the administration of antibiotics. It is the most common cause of nosocomial diarrhoea and is responsible for 10-20% of all cases of AAD, including the majority of such cases with colitis, and for 95-100% of cases of PMC (Bartlett, 2002).

1.1.2. Clinical and pathological findings

The two toxins produced by *C. difficile* lead to a number of histological changes in the colon. Patchy epithelial necrosis occurs with an exudate of neutrophils, which sometimes erupt into a 'volcanic' lesion. More prominent histological changes are characterised by diffuse epithelial necrosis and ulceration with development of an overlying pseudomembrane containing cellular debris, leukocytes, fibrin and mucin (Figure 1.1) (Medline *et al.*, 1976).

Patients suffering from *C. difficile* infection present with a wide range of clinical manifestations, ranging from asymptomatic carriage to life-threatening PMC (Mylonakis *et al.*, 2001). Typically voluminous and watery diarrhoea, accompanied by cramps and tenderness begin 4-10 days after initiation of antibiotic therapy, although antimicrobials taken 1-2 months previously can still pre-dispose to infection. Diarrhoea has a characteristic foul odour and is rarely bloody. Systemic symptoms such as fever, nausea and malaise are normally present (Kelly *et al.*, 1994). Fulminant colitis develops in approximately 1-3% of patients resulting in ileus, toxic megacolon (dilation due to loss of muscular tone), colonic perforation and a mortality rate of up to 30% (Kelly & LaMont, 1998).

1.1.3. Diagnosis

Currently, there is no single laboratory test that is specific and sensitive for diagnosing *C. difficile*-associated disease (CDAD). In the UK, toxin testing is the most common method, using kits that detect toxin A and/or toxin B. Cell cytotoxicity assays and culture on selective media are also in widespread use (Wilcox & Smyth, 1998).

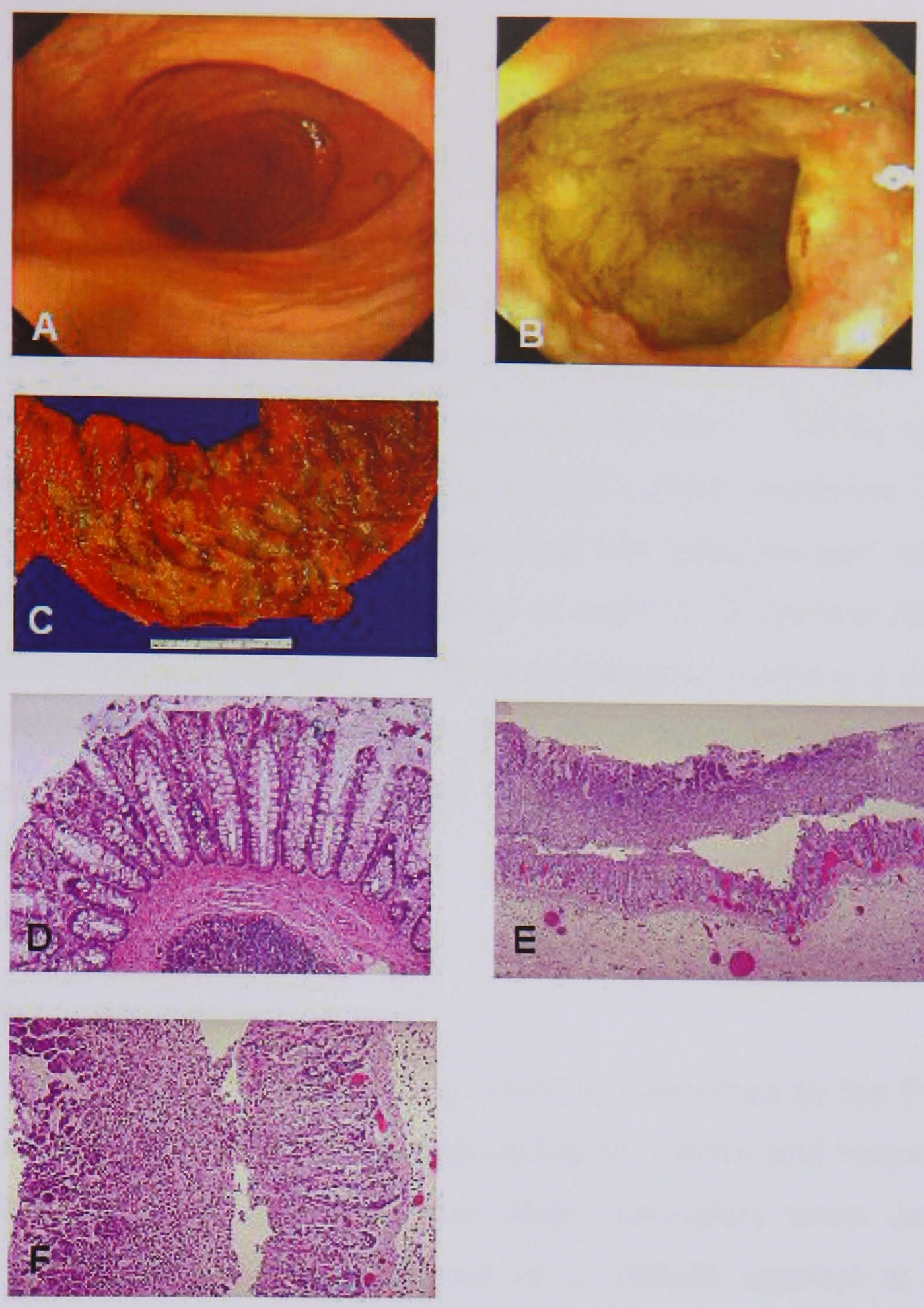


Figure 1.1 Gross pathology and histology of normal colon and colon with pseudomembranous enterocolitis. **A**. Normal colonic view with endoscopy. **B**. Endoscopic view of pseudomembranous colitis, showing typical pseudomembranes adherent to the colonic mucosa. **C**. Cross section of a colon with pseudomembranous enterocolitis. Yellow-green plaques partially cover a hyperaemic mucosa, which itself is not eroded. **D**. Normal colonic mucosa viewed with medium power microscopy. **E**. Low- and **F**. Medium-power microscopy of colitis. The pseudomembrane is composed of inflammatory cells (mainly neutrophils), necrotic epithelium, and mucus in which the overgrowth of microorganisms takes place. The underlying mucosa shows congested vessels, but is still intact. Images are from The Internet Pathology Laboratory for Medical Education, Florida State University College of Medicine at http://www-medlib.med.utah.edu/WebPath/ORGAN.html.

Despite small country to country variations, these tests have a similar popularity across Europe (Barbut *et al.*, 2003).

The tissue culture cytotoxicity assay is highly sensitive (94-100%) and specific (99%), but is expensive and lengthy (2-3 days) (Kelly & LaMont, 1998). Cell monolayers are exposed to stool filtrates where toxins, primarily toxin B, exert a cytopathic effect characterised by cell rounding. The specificity of the test is established by neutralising toxin B with the cross-reacting *C. sordellii* antitoxin (Rifkin, 1977). More rapid, same-day results are obtained with enzyme immunoassays for the *C. difficile* toxins (Viscidi *et al.*, 1983; Whittier *et al.*, 1993). These assays have a lower sensitivity (70-90%) but good specificity (99%). However, it is important that kits detecting both toxin A and B are used, as there is an increasing emergence of toxin A^- B^+ strains (Barbut *et al.*, 2002; Yucesoy *et al.*, 2002). Culturing *C. difficile* on selective media is a very sensitive method of detection, detecting low numbers of bacteria, but when used alone without toxin testing may lead to low specificity and misdiagnosis of CDAD due to asymptomatic carriage (Delmee, 2001).

1.1.3. Epidemiology and typing

The surveillance of *C. difficile* in England and Wales is performed by the Public Health Laboratory Service Communicable Disease Surveillance Centre and hospital reporting of cases in patients over the age of 65 has been mandatory since January 2004 (Department of Health, 2006). The incidence of *C. difficile* appears to be growing throughout the developed world, with the annual number of positive cases in England and Wales increasing from 1,576 in 1993 to 44,488 in 2004 (Wilcox & Smyth, 1998; Department of Health, 2006). Europe wide, this incidence represents 1.1 cases per 1000 patient admissions (Barbut *et al.*, 2003) and is leading to an increasing number of hospital ward closures (Wilcox & Smyth, 1998). Each *C. difficile* infected patient in the UK costs an average of £4000 in extra health care costs due to the extended time in hospital, drugs and tests required (Wilcox, 1996). In the US, there are estimated to be 3 million cases of PMC and AAD caused by *C. difficile* per year (Johnson *et al.*, 1990) costing the US $1.1 billion per year (Kyne *et al.*, 2002a).

In 2002, hospitals in Quebec, Canada experienced an outbreak of hypervirulent *C. difficile*, with a high incidence and mortality rate (Eggertson, 2005a, 2005b). This strain was found to express binary toxin and has an 18 base pair deletion in *tcdC*, the negative regulator of the toxins (Warny *et al.*, 2005) (see section 1.1.4.1). The hypervirulent strain has been associated with not only the use of cephalosporins, but

also fluoroquinolones, to which it is resistant (Loo *et al.*, 2005). The cumulative attributable mortality of the outbreak strain is reported to be as high as 16.7% (Pepin *et al.*, 2005). This strain, or very closely related strains, now appear to be causing more geographically dispersed outbreaks, including in the UK (McDonald *et al.*, 2005).

Asymptomatic faecal colonisation by *C. difficile* commonly occurs in neonates (47% in one study (Bolton *et al.*, 1984)), despite a large portion of these having titres of cytopathic faecal toxin. There is no evidence of acquisition from the mother during delivery and colonisation appears to be unrelated to the means of delivery, infant sex, means of feeding, duration of hospital stay or antibiotic treatment (Bolton *et al.*, 1984). For unknown reasons, the rate of carriage diminishes during the first year of life and toxinogenic *C. difficile* is isolated from stool specimens in only 0-3% of healthy adults (Knoop *et al.*, 1993). However, antibodies against toxin A are present in 60% of the population (Kelly *et al.*, 1992). During hospitalisation, colonisation frequently occurs (Johnson *et al.*, 1990) although in approximately a third of cases the infection is asymptomatic. The balance between asymptomatic carriage and CDAD is very fine, and dependant as much on host factors as bacterial ones, and it is not known whether the carriage of *C. difficile* is stable or transient in these cases. Presumably, these asymptomatic carriers act as a reservoir of infection, with patient-patient contact spreading the disease. The clonality of the causative strains of hospital outbreaks supports this assumption (Meakins *et al.*, 2003). *C. difficile* has also been shown to be prevalent in hospitals and facilities providing long-term care (Al-Eidan *et al.*, 2000). The organism can be cultured from the hands of personnel and equipment such as toilets, bedpans, bedding, floors, mops, furniture and telephones. The spores are highly resistant and therefore difficult to eliminate from these sites (Kim *et al.*, 1981).

Strain typing is important for the investigation of outbreaks of *C. difficile* and for descriptive epidemiology. Methods need to be easy, quick, cheap, reproducible and suitably discriminatory (Streulens, 1998). A number of methods have been developed and are in use (reviewed in Brazier, 2001).

Phenotypic methods for typing include serological and electrophoretic analysis of bacterial proteins. Serological analysis with rabbit antisera is performed by agglutination on slides, and has led to the recognition of 10 serotypes (Delmee *et al.*, 1985). Electrophoretic analysis is performed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of whole cell proteins. Each protein pattern is characterised by 1-3 densely staining bands and is highly reproducible (Costas *et al.*, 1994).

Genotypic methods use restriction endonucleases to digest genomic DNA, and the characteristic patterns generated are viewed by electrophoresis (Kuijper *et al.*, 1987). Digestion of genomic DNA with a restriction enzyme that recognises six-base sequences such as *Hin*dIII, followed by electrophoresis was originally used (Wren & Tabaqchali, 1987), and this method has been shown to correlate well to serological analysis (Clabots *et al.*, 1993). However, the method is labour intensive, complex, difficult to compare and the DNA can be susceptible to self-nuclease digestion.

A number of polymerase chain reaction (PCR) based methods have also been developed for typing purposes. PCR ribotyping amplifies the spacer region between the 16S and 23S rRNA regions. Like many other bacteria, *C. difficile* has been shown to possess multiple copies of the rRNA genes, which vary characteristically in number and length on the genome. Ribotyping is reproducible and shows good correlation with serotypes (Collier *et al.*, 1996). This method has become routine at the UK Anaerobic Reference Unit in Cardiff, which since 1995 has ribotyped over 2000 strains and identified 116 ribotypes (Stubbs *et al.*, 1999). Other PCR based methods for typing strains include amplification of the toxin and flagellin genes in combination with sequencing or restriction fragment length polymorphism (RFLP) (Rupnik *et al.*, 1998; Tasteyre *et al.*, 2000).

PCR Ribotype 001 is the predominant strain in England and Wales (UK Anaerobe Reference Unit) and is endemic in virtually all hospitals. Other ribotypes are prevalent in other countries, for example the hypervirulent strain currently causing problems in Quebec is ribotype 027. Although most clusters of CDAD cases are due to a single strain, occasionally unrelated strains cause outbreaks, probably due to positive selection by local environmental pressures in hospitals (Brazier, 2001).

1.1.4. Pathogenesis

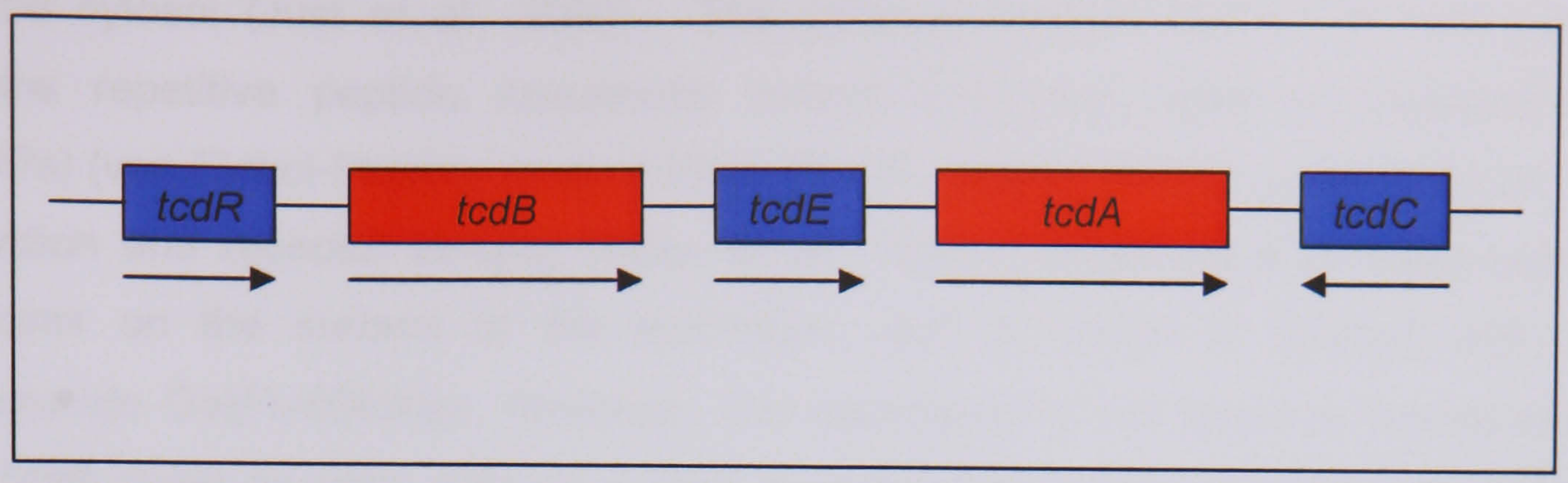
C. difficile infection is most commonly seen after treatment with antimicrobial drugs. Normally *C. difficile* does not cause disease, as it is out competed in the gut by the established microflora. This protective effect is known as 'colonisation resistance'. Disruption of the normal gut microbiota allows *C. difficile* to become established and cause disease. Colonisation resistance has been demonstrated *in vitro* with faecal emulsions that inhibited the growth and toxin production of *C. difficile* (Borriello & Barclay, 1986). The complexities of colonisation resistance remain to be fully elucidated, but various genera of faecal bacteria (*Bacteroides* sp., bifidobacteria, lactobacilli, pseudomonads, staphylococci and streptococci) appear to inhibit growth of

C. difficile in vitro (Rolfe *et al.*, 1981). A number of broad-spectrum antibiotics have been implicated in facilitating *C. difficile* infection, particularly clindamycin, second- and third-generation cephalosporins, ampicillin and amoxycillin. Restricting the use of the pre-disposing agent has been shown to control and reduce the number of outbreaks (McNulty *et al.*, 1997). The strong association between CDAD and antibiotic exposure distinguishes this disease from that caused by any other enteropathogenic bacteria. Increased age, the severity of underlying disease and lengthy periods in hospital are other important risk factors that predispose humans to the development of CDAD (McFarland *et al.*, 1990; Kyne *et al.*, 2002b).

1.1.4.1 C. difficile *toxins*

The primary effectors of tissue damage and inflammation in CDAD are two exotoxins, TcdA (toxin A) and TcdB (toxin B). Both toxins are lethal upon systemic administration to animals and are cytopathic to tissue culture cells. TcdB is approximately 1000 times more potent as a cytotoxin than TcdA, and this is thought to be due to differences in toxin binding, enzymatic activity and substrate specificity (Chaves-Olarte *et al.*, 1997). TcdA has been shown to possess enterotoxic activity (causing fluid secretion, mucosal damage and inflammation) when injected into rabbit ileum (Triadafilopoulos *et al.*, 1987). The two toxins are encoded on a pathogenicity locus along with a number of small open reading frames (ORFs), thought to be regulators (Figure 1.2A). TcdR has been shown to be an alternative sigma factor and is suspected to be a positive regulator of the toxin genes, although it does not bind directly to the toxin promoters (Mani & Dupuy, 2001). *TcdR* is transcribed during stationary phase and is responsive to environmental conditions (Hundsberger *et al.*, 1997). The *tcdE* gene product has structural similarity to bacteriophage holin proteins that lyse host cells before the progeny bacteriophage are released (Tan *et al.*, 2001). TcdE is speculated to facilitate the release of the two toxins through permeabilisation of the cell wall. *TcdC* is transcribed during exponential phase and is suspected to be a negative regulator of the toxin genes (Hammond, 1997). The toxins are large single chain polypeptides with molecular masses of 308 kDa for TcdA and 270 kDa for TcdB (von Eichel-Streiber *et al.*, 1992; Knoop *et al.*, 1993). Toxins A and B are homologous, sharing 50% identity at the protein sequence level. Each toxin contains three functional domains (Figure 1.2B). Residues 1-958 in TcdA and 1-956 in TcdB specify monoglucosyltransferase activity which catalyses the incorporation of glucose into small GTP-binding proteins (Rho, Rac and Cdc42) (Just *et al.*, 1995a, 1995b). These GTP-binding proteins regulate the actin cytoskeleton, forming stress fibres and focal adhesions. A putative translocation domain exists between amino acid residues 958-1849 in TcdA and 956-

A



B

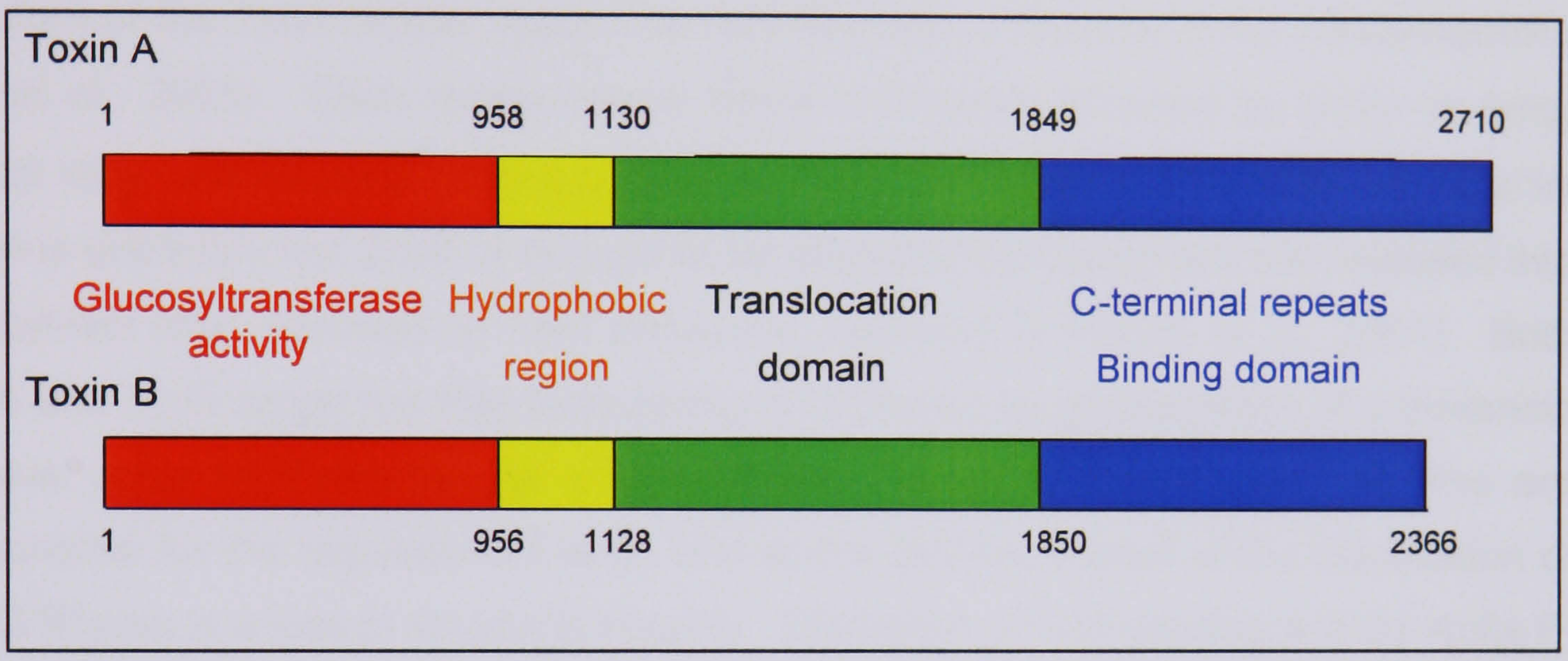


Figure 1.2. A. The pathogenicity locus of *C. difficile* (not to scale). Red boxes indicate the toxin structural genes. Blue boxes indicate small open reading frames, believed to be regulators of expression (*tcdR* and *tcdC*) and possibly components of a secretion apparatus (*tcdE*). Arrows indicate the direction of transcription. Modified from von Eichel-Streiber *et al.* (1992). **B. Molecular structure of *C. difficile* toxins.** The two genes are homologous, each containing a three-domain structure. The N-terminal ~960 amino acids encode the enzymatic domain that glucosylates threonine-37 in RhoA and threonine-35 in Rac and Cdc42. The central domain contains a putative translocation domain, including a hydrophobic region, thought to cross the membrane and which binds ATP. The C-terminal domain encodes the receptor-binding domain made up of a number of repeat sequences. Modified from Pothoulakis and Lamont (2001).

1850 in TcdB, each of which contains an amino-terminal hydrophobic region. The function of this region is unknown, but is thought to mediate translocation of the toxin into the cytosol (Just *et al.*, 2000). The carboxyl-terminal domain of both toxins contains repetitive peptide sequences termed combined repetitive oligopeptides (CROPs) (von Eichel-Streiber *et al.*, 1992). CROPs appear to play a role in target cell interaction and receptor binding (Wren *et al.*, 1991). TcdA binds to carbohydrate structures on the surface of the epithelium, and is known to interact with the disaccharide Galβ1-4GlcNac, (however, this saccharide is not found in human cells) and TcdB binds to cells without a thick carbohydrate matrix, although its specific receptor is unknown. The toxins then enter the cells by receptor-mediated endocytosis through an acidic intracellular component (Thelestam & Chaves-Olarte, 2000). The structure of the TcdA repeat region has recently been solved by X-ray crystallography (Ho *et al.*, 2005). Each repeat region forms a β-hairpin followed by short- or long-repeat loop, and together form a β-solenoid fold. The route of entry into the cells for TcdA is unknown, but TcdB is thought to be delivered by lysosomes and released into the cytosol after cleavage by host proteases; reviewed in Poxton *et al.*, 2001). Both TcdA and TcdB target the Ras superfamily of GTPases by glucosylation of a threonine residue using UDP-glucose as a cosubstrate. Small GTPases such as Rho are responsible for the regulation of actin, and so the primary impact of the inactivation of the GTPases is a loss of structural integrity. Disruption of GTP-binding activity shifts F-actin to G-actin resulting in alterations to the epithelial microvilli and surface projections (Chang *et al.*, 1979) and characteristic cell rounding (Pothoulakis & Lamont, 2001). In addition, Rho, Rac and Cdc42 are involved in a number of other cellular events, for example cell cycle regulation and signalling through mitogen-activated protein kinases. TcdA can induce changes in the integrity of the epithelial layer, allowing further toxin (or bacterial) access to the basolateral epithelium (Canny *et al.*, 2005). Here, inflammation is induced, including disruption of tight junctions, neutrophil infiltration, substance P production, chemokine production, reactive oxygen intermediate production and activation of the apoptotic cell death pathway (reviewed in Voth & Ballard, 2005). Indeed, many of the pathologies of the disease can be attributed to inflammatory events modulated by the toxins, rather than to the direct effects on the cytoskeleton.

A series of experiments in a hamster model of infection showed a correlation of virulence with the amount of TcdA produced, and therefore TcdA was assumed to play the crucial pathogenic role (Borriello *et al.*, 1987). However, the isolation of $TcdA^-$ $TcdB^+$ strains from severe cases of diarrhoea and PMC has challenged this assumption (Alfa *et al.*, 2000; Barbut *et al.*, 2002). Two types of TcdA mutant have

been isolated: the first contains a ~5.9 kb deletion at the 3' terminus of the TcdA gene (Soehn *et al.*, 1998); the second, more common mutant contains a ~1.8 kb deletion in the 3' CROP region of the TcdA gene (Rupnik *et al.*, 2003). The incidence of $TcdA^-$ mutants has been described as ranging from less than 1% (Sambol *et al.*, 2000) to 56.5% (Samra *et al.*, 2002). Therefore, TcdB may be more important in the manifestation of disease than previously thought.

In addition to TcdA and TcdB, a number of strains have been found to produce a binary toxin with actin-specific ADP-ribosyltransferase activity (Popoff *et al.*, 1988; Perelle *et al.*, 1997). Of the strains submitted to the UK Anaerobe Reference Unit, 6.4% contained the *cdtA* (enzymatic domain) and *cdtB* (binding) components of the binary toxin (Stubbs *et al.*, 2000). Indeed, a recent study linked CDAD severity to the presence of the binary toxin genes, rather than the strain or severity of underlying conditions (McEllistrem *et al.*, 2005). Although the binary toxin has been shown to have cytotoxic activity (Perelle *et al.*, 1997), its role in human disease has not yet been fully defined.

1.1.4.2 Further virulence factors

A role for *C. difficile* virulence factors in addition to TcdA and TcdB was suggested by the occurrence of asymptomatic carriage of toxin-positive strains and the finding that toxigenic strains display a range of virulence in the hamster model (Borriello *et al.*, 1987). Relatively little is known about the conditions necessary for colonisation of the gut by *C. difficile* or the identity of bacterial factors necessary for the initial stages of pathogenesis.

Possible virulence factors that may be involved in pathogenesis of *C. difficile* are those that mediate survival in the host, adhesion to host cells and interaction with the host immune system. The full mechanisms behind these processes are not known, although several putative virulence factors have been described. Adherence has been shown to be modified by a number of environmental factors such as increased sodium, calcium, an acidic pH and iron starvation (Waligora *et al.*, 1999), but the specific mechanisms by which this occur are not clear. Components of the blood or extracellular matrix may be utilised by *C. difficile* to adhere to intestinal cells (Eveillard *et al.*, 1993; Hennequin *et al.*, 2003). Flagella, in particular FliC (flagellin) and FliD (flagellar cap protein) are believed to have a role in pathogenicity. Flagellated strains adhere 10 times more strongly than non-flagellated strains to mouse cecum tissue. FliC, FliD and crude flagella themselves adhere to axenic mouse cecal mucus *in vitro*

(Tasteyre *et al.*, 2001). GroEL is a molecular chaperone that is extra-cellularly released by *C. difficile* after heat-shock and can be associated with the surface of the bacterium. Although not previously associated with virulence, GroEL specific antibodies and an excess of the purified protein were found to partially block adherence of *C. difficile* to Vero cells (Hennequin *et al.*, 2001b). GroEL may therefore be one of a number of adhesins of *C. difficile*. A number of other heat shock proteins have been implicated in virulence, but these remain to be identified and their role determined (Eveillard *et al.*, 1993; Karjalainen *et al.*, 1994). Studies using ELISA and FACS analysis on the high molecular weight surface layer protein (HMW SLP) and its antibodies have shown that the HMW SLP is able to mediate binding to HeLa cells and to purified extracellular matrix components collagen I, thrombospondin and vitronectin (Calabi *et al.*, 2002). A paralogue of the surface layer proteins (SLPs), Cwp66, has also been shown to have adhesive properties (Waligora, 2001). Immunolabelling with specific antibodies showed the C-terminal domain of Cwp66 was cell surface exposed. Purified fragments of Cwp66 and antibodies against it were found to partially inhibit adherence to Vero cells. The putative roles of the SLPs in pathogenesis are described further in section 1.3. No correlation has been made between presence of fimbriae or capsule and the virulence status of the organism (Borriello *et al.*, 1988; Davies & Borriello, 1990).

Taken together, these results suggest that there are multiple adhesins and putative virulence factors expressed by *C. difficile*, only a few of which have been discovered and their roles elucidated. The recent completion of the *C. difficile* genome sequence (Sebaihia *et al.*, 2006) will facilitate the discovery of further virulence factors, allowing a greater understanding of the mechanisms by which this organism causes disease.

1.1.5 *C. difficile* treatments

Treatment of *C. difficile* infections begins by the withdrawal of the inciting antibiotic, usually followed by therapy with metronidazole or vancomycin. However in many cases the infection recurs as soon as the antibiotic treatment is withdrawn, sometimes due to the germination of spores in the colon not killed by the antibiotic treatment and sometimes due to reinfection. A number of alternative therapies have been explored, with mixed results, but most lack randomised placebo-controlled trials and so their success is not fully proven. In the case of recurrent infections, the use of probiotics given alongside antibiotic therapy may be beneficial. *Saccharomyces boulardii* is the most studied probiotic, but lactobacilli and faecal enemas with healthy stools also show some levels of success (reviewed in McFarland, 2005). Alternatively, administration of

immune products may be therapeutic, particularly as many patients with recurrent infections fail to produce a high anti-TcdA antibody response (Wilcox, 2004). A low level of mucosal IgA is associated with CDAD, and if severely reduced, may predispose to recurrent infection (Johal *et al.*, 2004). Another key area for hospitals to target to reduce the incidence of CDAD is infection control and antibiotic management. Hand washing, gloves and ward-decontamination have been shown to be effective at reducing CDAD (Gerding *et al.*, 1995). Hospital policies to restrict the use of broad-spectrum antibiotics, particularly clindamycin, have proved successful in limiting CDAD in two hospitals (Brown *et al.*, 1990; Pear *et al.*, 1994).

Knowledge of the mechanisms and factors that contribute to *C. difficile* disease is now facilitating the rational design of treatments and vaccines. A serum antibody response to TcdA has been shown to be protective against *C. difficile* infection (Kim *et al.*, 1987a; Aboudola *et al.*, 2003), and so the focus of a number of studies has been TcdA immunity. Inactivated toxoids given intra-muscularly were protective against *C. difficile* challenge in a hamster model, where serum antibodies showed neutralising activity against TcdA and B (Giannasca *et al.*, 1999). In order to define which portion of the TcdA and B polypeptides could elicit protective antibodies, Kink and Williams (1998) created a number of recombinant peptides that spanned the whole length of the toxins. Hens were immunised with these peptides, and IgY was extracted from their eggs, which was administered to hamsters to assess their level of passive protection. The most effective domain to elicit toxin neutralising antibodies was the CROP region. Indeed, antibodies directed against TcdA were sufficient to prevent disease, but antibodies to both toxins were able to offer complete therapeutic protection from death and diarrhoea. Hamsters treated with anti-toxin antibodies did not develop any relapsing disease months later. Toxoid-based vaccinations have now been tried in healthy volunteers and a small number of CDAD patients, where they were well tolerated and resolved a number of recurrent infections (Sougioultzis *et al.*, 2005). In an alternative approach, Ward and co-workers made a number of constructs containing the CROP repeats in frame with the fragment C region of tetanus toxin, which they expressed in an attenuated strain of *Salmonella typhimurium* (Ward *et al.*, 1999a, 1999b). Good immune responses were observed in vaccinated mice, both at mucosal surfaces and in the serum where it was capable of neutralising the cytotoxic activity of the toxin. Another vaccine target are the SLPs, where anti-SLP antibodies aided survival in a hamster model of disease (O'Brien *et al.*, 2005). Anti-SLP antibodies are believed to aid in the elimination of *C. difficile* by promoting phagocytosis. A further understanding of the virulence factors of *C. difficile* will help in the design of new vaccines and treatments.

1.2. The Gram-positive bacterial cell wall

1.2.1. The Gram-positive cell wall

The primary role of the cell wall is to provide a rigid structure to enable the cell to withstand the turgor pressure from the cytoplasm and also to provide shape and strength whilst allowing the free diffusion of substances through its structure. The Gram-positive cell wall is host to a variety of molecules serving a wide range of functions, and is described here in general terms, not specific to *C. difficile*.

1.2.1.1 Peptidoglycan

The primary component of the Gram-positive cell wall is peptidoglycan (Figure 1.3). This molecule has three components; the glycan backbone, short wall peptides and peptide cross-bridges. The glycan backbone is composed of 2 sugar derivatives, *N*-acetylglucosamine and *N*-acetylmuramic acid, connected by a β-1,4 linkage. The length of the backbone can vary between 5-30 subunits, depending on bacterial species (Glauner *et al.*, 1988). The glycan backbones of the peptidoglycan are linked by short wall peptides, usually in the sequence L-Ala – D-Glu – DAA – D-Ala, where DAA is a species-specific diamino acid, often L-Lysine, an isomer of diaminopimelic acid or ornithine. The L-Ala amino acid is attached to the lactyl group of muramic acid residues and the D-Ala amino acid is cross-linked through its side chain carboxyl to the diamino acid where the type and number of cross-linking amino acids varies between species. In its simplest form, the cross-link can be a direct peptide bond, often found in *Bacillus* species. Alternatively, the cross-links can be up to 5 amino acids long, and be composed of a single type of amino acid, e.g. the pentaglycine bridge in *Staphylococcus aureus*, or a more complex combination of residues. In Gram-positive bacteria, up to 90% of the cell wall is composed of peptidoglycan, and it can be several layers thick, to a depth of 20-50 nm thick (Hancock, 1988).

1.2.1.2 Secondary cell wall polymers

Gram-positive bacteria commonly synthesise secondary cell wall polymers (SCWPs) such as teichoic acids. Teichoic acids are acidic polysaccharides, which include all wall, membrane or capsular polymers containing glycerol-, ribitol-, or glucosyl-phosphate residues. When the repeat unit contains uronic acid, it is described as a teichuronic acid. The SCWPs are commonly decorated with sugars or esterified with D-alanine. Teichoic acids are either covalently attached to peptidoglycan via phosphate esters, or are bound to membrane lipids where they are known as lipoteichoic acids. Because the SCWPs are anionic, the surface of the cell has an

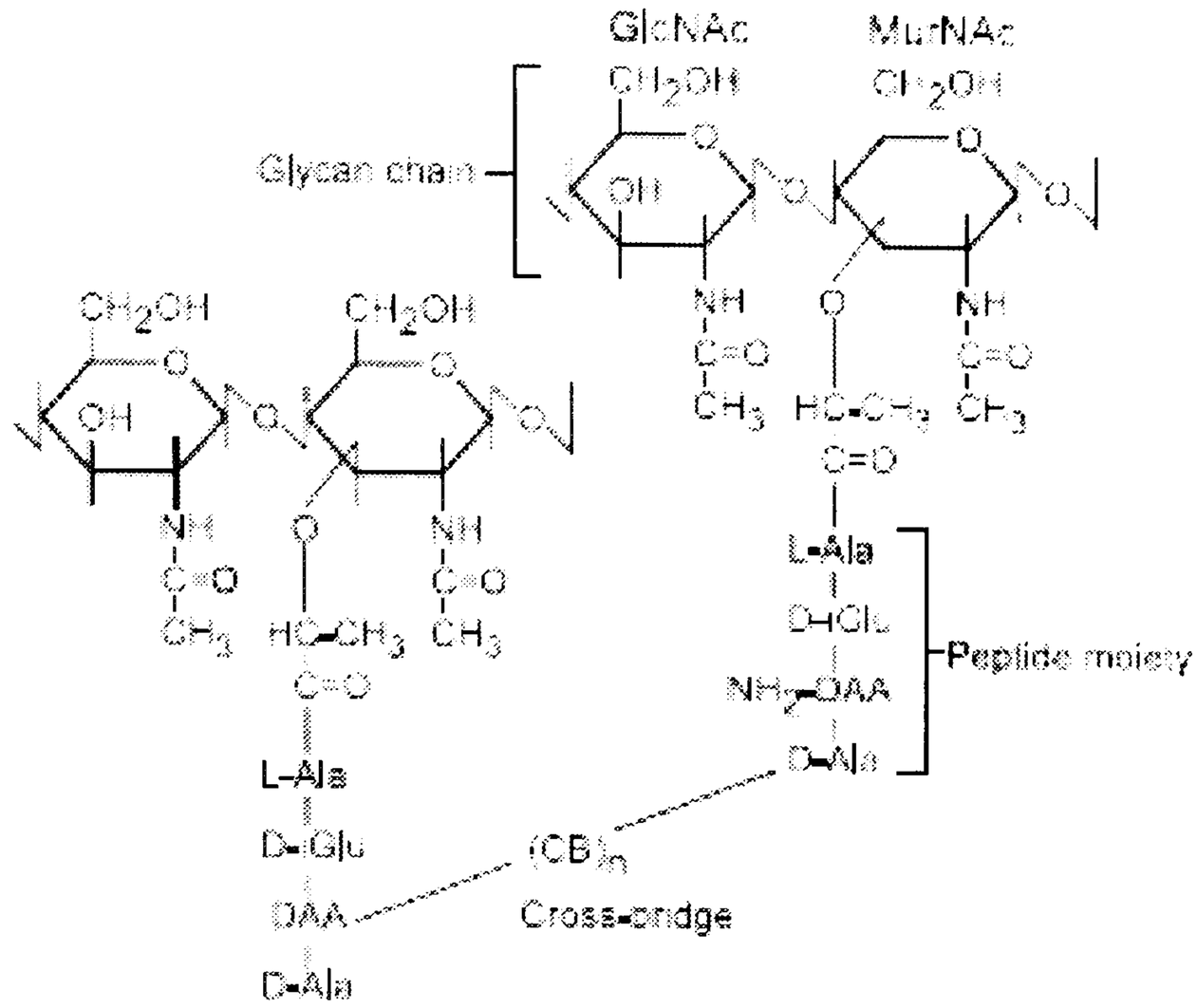


Figure 1.3 Generalised structure of Gram-positive peptidoglycan. The structure is composed of glycan chains, made up of a repeating disaccharide GlcNAc-MurNAc, short wall peptides and cross-bridges (CB). The glycan chain is linked to tetrapeptides of the sequence L-Ala – D-Glu – DAA – D-Ala (where DAA represents a diaminoacid) via a lactyl linkage of MurNAc to alanine of the tetrapeptide. The peptidoglycan chains are linked by cross-bridges, whose structure varies between bacterial species. From Ton-That *et al.* (2004).

overall negative charge, which may allude to the SCWP's as yet undetermined function (Hancock, 1988). Several putative roles have been put forward, such as aiding the capture and passage of ions through the cell wall or protection against noxious hydrophobic compounds e.g. bile salts (Fischer & Tomasz, 1984). They may also provide binding sites for enzymes that sculpt peptidoglycan, e.g. the LytA amidase of *Streptococcus pneumoniae* binds to the choline moiety of cell wall teichoic acids (Holtje & Tomasz, 1975). Although their role is not yet fully understood, the SCWPs are highly antigenic molecules that are found in great variety on the surface of all Gram-positive bacteria.

1.2.1.3 Proteins

Proteins are commonly found associated with the cell wall, either directly attached (discussed in section 1.2.2), or more loosely associated. In some cases, the attached protein has a specific receptor function, e.g. staphylococcal protein A has a high affinity for the Fc region of immunoglobulin, causing precipitation of the antibodies (Sjodahl, 1977; Wright *et al.*, 1977). However, the function of many surface associated proteins is still unknown. Specific details of proteins associated with the cell wall of *C. difficile* are described below (section 1.3).

1.2.1.4 Capsule

A number of microorganisms produce a capsule external to their cell wall. This capsule-layer is of amorphous structure, and is often composed of high-molecular weight polysaccharides. Expression of a capsule is conditional, and depends upon the growth conditions (Fouet & Mesnage, 2002). It is believed that the capsule aids in virulence and the evasion of the host immune system, with its slipperiness resisting phagocytosis (Taylor & Roberts, 2005).

1.2.1.5 Flagella

Anchored to the cell surface of some bacteria are flagella, the structure of movement. A flagellum is a helical filament driven by a rotating motor at its base, which propels the cell through the medium (Macnab, 2004). A flagellum is composed of 5 parts: an outermost filament with its distal cap that functions like a propeller; a hook which acts as a joint; a basal body which receives torque from the motor and transmits it to the hook and filament; a motor which operates and controls the rotation; and the export apparatus for the translocation of components prior to assembly (discussed further in section 1.2.2.5). The filament is composed of units of flagellin, a self-assembling

protein unit that extends 5-10 μm into the medium. The hook is short curved structure, which forms a constant length arrangement larger in diameter than the filament. The basal body is a small flat structure embedded in the cell surface. It is composed of 15 or more proteins that aggregate to form a rod with two rings attached (in Gram-positive bacteria), one embedded in the cell membrane, and one in the teichoic acid component of wall. The motor uses energy supplied by the proton motive force for turning the flagellum. The motor also contains a number of receptors that allows it to respond to chemical stimuli, thereby resulting in chemotaxis (Kojima & Blair, 2004).

1.2.1.6 Pili

Another type of surface appendage are pili. Pili are composed of pilin and originate from the cell membrane, extending up to 2 μm into the medium. They commonly mediate attachment to surfaces, with the tip-protein exhibiting a high degree of substrate specificity (Tasteyre, 2001).

1.2.2. Protein secretion and attachment to the cell wall

Any protein which has been translocated across the cytoplasmic membrane of a bacterial cell can be termed extracellular, a definition which includes the periplasmic and outer membrane proteins of Gram-negative bacteria. Translocation occurs by a number of dedicated systems, as it must proceed without compromising the permeability of the membrane. Exported proteins have a wide range of functions, e.g. enzymes for hydrolysis of natural polymers or cell density signalling. Alternatively, exported proteins may be retained at the cell wall for tasks such as cell wall turnover, substrate binding, or protein secretion (Briggs, 1986).

1.2.2.1 Signal peptide

The majority of proteins destined for secretion contain an amino-terminal signal sequence (Blobel, 1980) that is recognised by molecular chaperones as it is transcribed and transported to the cytoplasmic membrane. A schematic diagram of the proposed model for signal peptide mediated translocation is shown in Figure 1.4. Generally signal peptides are composed of three distinct domains, N, H and C. The N-domain is positively charged and contains a number of arginine or lysine residues (Silhavy & Beckwith, 1983). The N-domain is thought to interact with the negatively charged phospholipids in the lipid bilayer of the membrane and may also interact with translocation machinery (Akita *et al.*, 1990). The H-domain is formed by 15-20 hydrophobic residues that adopt an α-helical conformation in the membrane

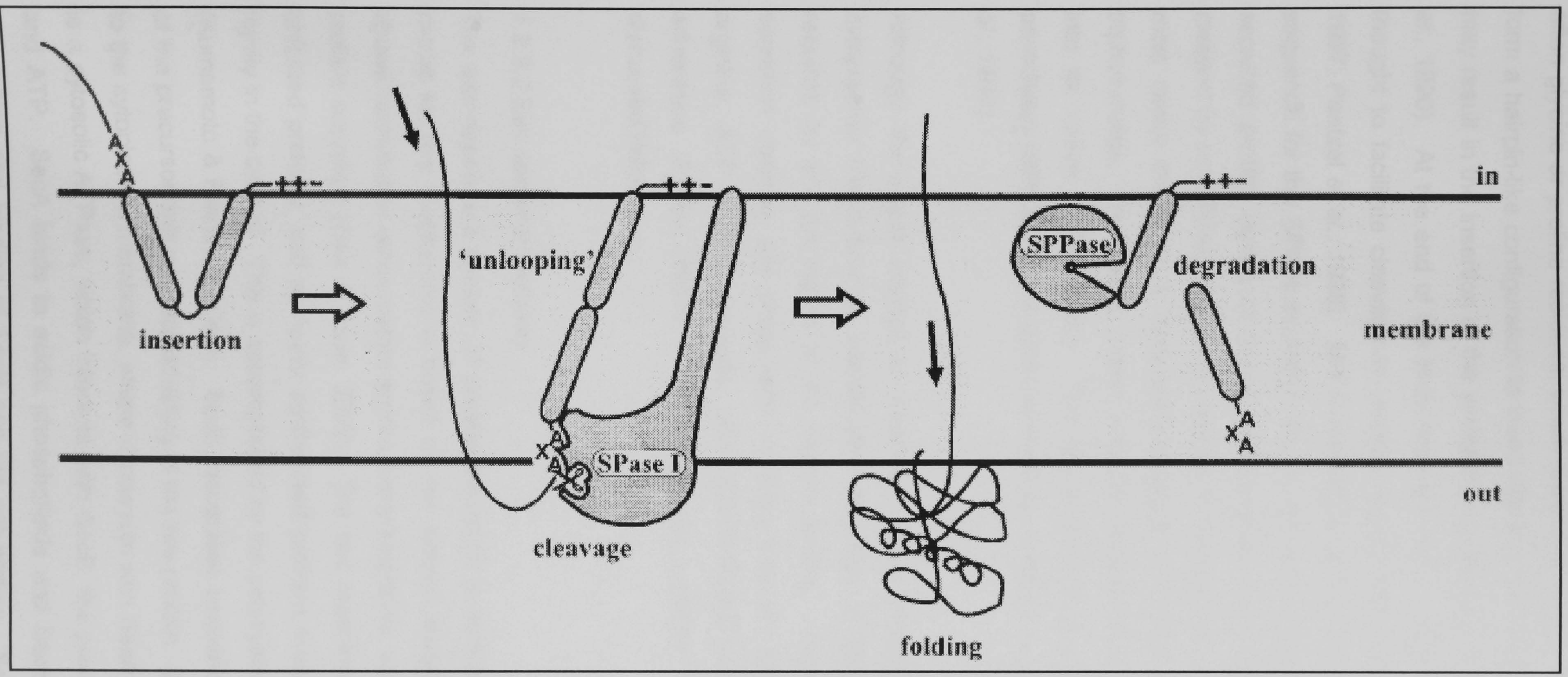


Figure 1.4 Model of signal peptide insertion into the cytoplasmic membrane and cleavage by SPase. The positively charged N-domain interacts with the negatively charged phospholipids in the membrane. Next, the H-domain integrates into the membrane in a hairpin-like loop structure. As it unloops, the first part of the mature protein is pulled through the membrane. After translocation by specific apparatus (not shown here), the signal peptide is cleaved by SPase and subsequently degraded by SPPase. After translocation, the mature protein folds into its native conformation. From Tjalsma *et al.* (2000).

(Emr & Silhavy, 1982; Briggs & Gierasch, 1986). In the middle of the H-domain are often glycine or proline residues, which break the α-helix structure and may allow it to form a hairpin-like configuration to insert into the membrane. Unlooping of the hairpin may result in the insertion of the whole signal peptide into the membrane (de Vrije *et al.*, 1990). At the end of the H-domain are more α-helix breaking residues that are thought to facilitate cleavage by specific signal peptidases (SPases) (Dalbey *et al.*, 1997; Paetzel *et al.*, 1998). The C-domain is hydrophilic and contains the consensus sequence for the SPase cleavage site to remove the signal peptide from the mature secreted protein during or shortly after translocation (Dalbey *et al.*, 2000). Once cleaved by an SPase, the mature protein is transported across the cell membrane, in most cases driven by a translocation motor that binds and hydrolyses nucleoside triphosphates. The mature protein is finally released from the membrane and can fold into its native conformation. The signal peptide is degraded by signal peptide peptidases (SPPases) and removed from the membrane (Dev & Ray, 1990; Dalbey *et al.*, 1997).

Although the signal peptide is necessary for a pre-protein to be directed to the cytoplasmic membrane for translocation to begin, a dedicated secretion system is required for the full export of the mature protein. Currently, seven major protein secretion systems are recognised in Gram-positive bacteria, sec-dependent, twin arginine, ATP binding cassette, pilin, ESAT-6/WXG100, flagella export and tight adherence (Pallen, 2003). The four most common translocation systems are discussed below.

1.2.2.2 Sec secretion pathway

The sec-dependent system of protein secretion is summarised in Figure 1.5. The model for this mechanism is based on the system elucidated in *E. coli*, however it shows remarkable conservation across many kingdoms, confirming its essential role in protein secretion (Cao & Saier, 2003). The sec machinery is only able to transport unfolded proteins, and so newly synthesised proteins must be prevented from folding tightly in the cytosol. This is accomplished by the sec-system specific chaperone SecB (Kumamoto & Francetic, 1993). SecB recognises sequence motifs in the mature part of the precursor and can retard folding of the pre-protein. SecB targets the pre-protein to the cytoplasmic membrane, where it interacts with SecA (step 2, Figure 1.5). SecA is a cytosolic ATPase, which interacts with SecB, the pre-protein, the cell membrane and ATP. SecA binds to acidic phospholipids and SecY, where it is activated for recognition and binding of SecB and the pre-protein. SecY is a component of the

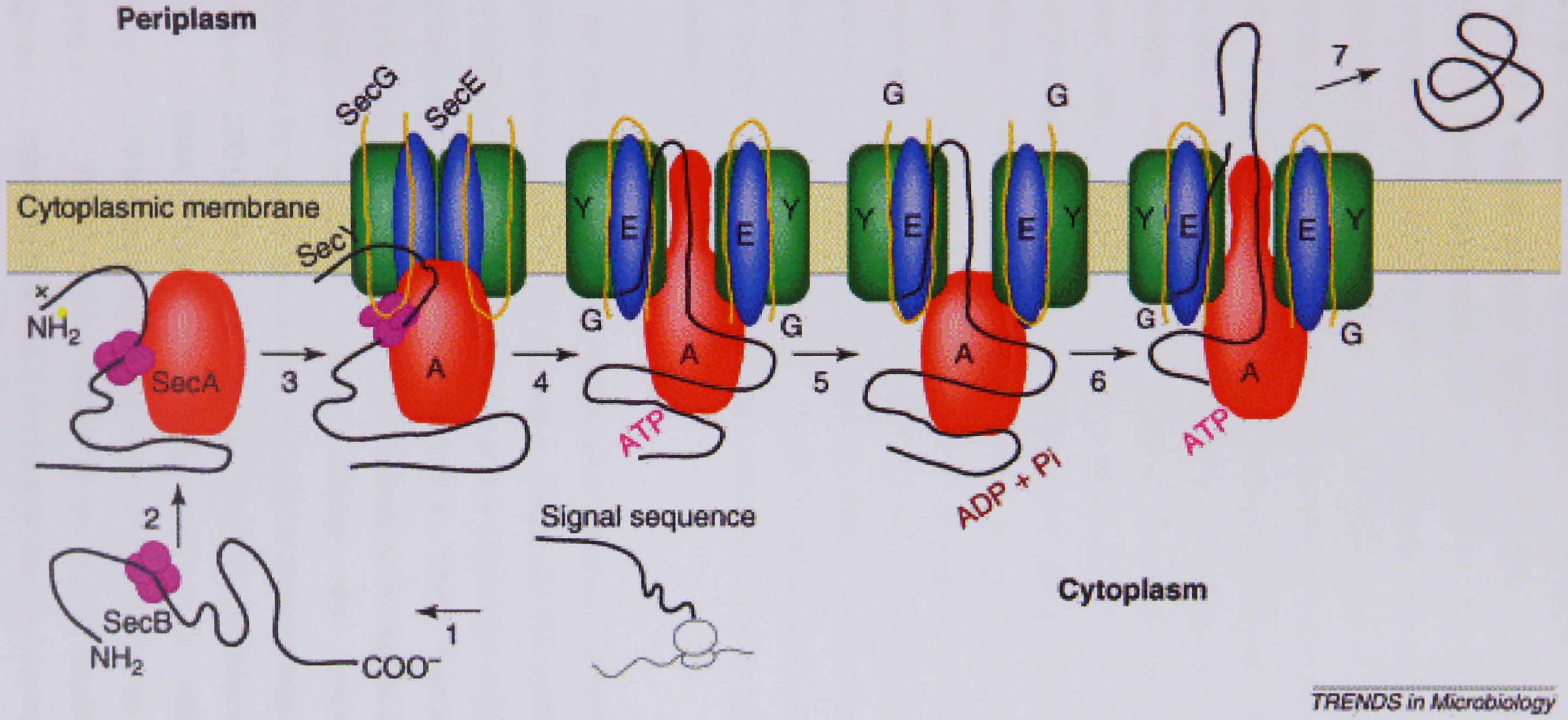


Figure 1.5 Sec factors and the translocation processes. The protein is represented as a black line. Steps 1-3 show targeting of the pre-protein to the cytoplasmic membrane. The signal sequence is recognised SecB, the sec-system-specific chaperone, which targets the pre-protein towards the export apparatus, specifically SecA. SecA, bearing the pre-protein, then binds to the membrane. The heterotrimeric complex SecYEG forms the channel for polypeptide translocation. In steps 4 and 5 initiation of translocation begins with the binding of ATP. Translocation (step 6) occurs with cycles of ATP binding and hydrolysis and the proton motive force, each cycle pushing 20-30 residues more of the pre-protein through the SecYEG complex. At step 7, completion, the mature protein is released. The model for sec-secretion is *E. coli*, on which this figure is based. Only small differences exist in Gram-positive bacteria. From Mori and Ito (2001).

heterotrimeric SecYEG complex, which forms a protein-conducting channel through the cytoplasmic membrane. Negative staining has revealed that SecYEG complexes with several other SecYEG units to form ring-like structures with a central hole of 20-30 Å (Manting *et al.*, 2000). SecE is thought to function as a surrogate signal peptide when the SecY channel is in its closed form in the absence of translocating protein. Upon arrival of a signal peptide, SecE is displaced thus opening the SecY channel for transport. SecG may assist in the stabilising of the inserted state of SecA. When ATP binds to SecA (step 4, Figure 1.5), it changes the conformation of SecA, leading to the release of SecB and insertion of the C-terminus of SecA into the membrane. This conformational rearrangement drives movement of the protein through the membrane. Upon hydrolysis of ATP, SecA is induced to release the pre-protein and return to its previous configuration. Repeated cycles of these movements are thought to drive the stepwise movement of the pre-protein into and across membrane (Ramamurthy & Oliver, 1997). Three membrane proteins, SecD, SecF and YajC associate with the SecYEG pore and appear to regulate SecA dependent translocation activity (Duong & Wickner, 1997). Once translocation is completed (step 7, Figure 1.5), the mature protein is released and can fold into its native configuration. The mechanism of completion is not yet fully understood, but it is thought that SecD, SecF and SecY are involved in these late step reactions (Matsuyama *et al.*, 1993). The sec-secretion pathway is monitored by SecM. SecM modulates the translation of SecA by monitoring the protein export and integration activities of the cell (Nakatogawa & Ito, 2001; Murakami *et al.*, 2004).

To be exported by the sec-dependent pathway, proteins must be maintained in an export competent, i.e. unfolded state. There are two pathways by which this can occur; the SecB route described above, or the alternative signal recognition particle (SRP) pathway (de Gier *et al.*, 1997; Valent *et al.*, 1998). Rather than maintain the translated protein in an unfolded state, the SRP binds to the signal peptide and temporally arrests translation of the protein (Walter & Blobel, 1982, 1983a, 1983b). The SRP has a preference for signal sequences with higher hydrophobicity compared to SecB (Lee & Bernstein, 2001). In *E. coli*, the SRP is a ribonucleoprotein consisting of Ffh protein and 4.5S RNA (Poritz *et al.*, 1988; Ribes *et al.*, 1990). The SRP binds to the hydrophobic N-terminal signal sequence of the protein as it emerges from the ribosome (Bunai *et al.*, 1996). Once bound to the signal peptide, the SRP migrates to the cytoplasmic membrane where it docks onto the membrane receptor FtsY (Luirink *et al.*, 1994; de Leeuw *et al.*, 1997). Once the SRP has delivered the ribosome and nascent chain, the complex dissociates under the control of GTPases in both the SRP particle and FtsY receptor. Translation resumes at this point, and the pre-protein is delivered

to the sec translocation channel (Valent *et al.*, 1998). A major benefit of this cotranslational membrane insertion system is that it may provide a mechanism for the expression of hydrophobic membrane proteins that may otherwise aggregate before integration into the membrane.

Signal peptides that direct translocation via the SRP pathway rather than SecB tend to be longer and more hydrophobic (Zanen *et al.*, 2005). In *B. subtilis*, most signal peptides are characteristic of the SRP route, and as there is no SecB homologue in this organism, this is the likely route of targeting to the Sec translocon (Tjalsma *et al.*, 2000). In the *C. difficile* 630 genome, there are two copies of the *secA* gene, and also genes encoding for SecY, SecE and SecG. There is no gene homologous to *secB* in the genome.

1.2.2.3 Twin arginine translocation

This alternative mechanism by which proteins can be translocated through the cytoplasmic membrane was first discovered in chloroplasts involved in a pH dependent protein import into the thylakoid lumen, but was later found to be widespread among eubacteria (Chaddock *et al.*, 1995). The twin argine translocation (TAT)-pathway differs from the sec-pathway because it allows the passage of fully folded and even oligomeric proteins (Hynds *et al.*, 1998). This is thought to be especially useful when exporting proteins that use a cofactor. The key components of the TAT-translocon are three integral membrane proteins, TatA TatB and TatC (Bogsch *et al.*, 1998; Sargent *et al.*, 1998). Together, TatABC forms large multimeric complexes of 500-600 kDa. The TAT export pathway is extremely unusual because it has no requirement for ATP (Cline *et al.*, 1992), and instead is thought to rely on the proton motive force (Alder & Theg, 2003; Finazzi *et al.*, 2003). Although the mechanism by which the TAT-export system operates has not been fully elucidated, the following model has been proposed by Robinson & Bolhuis (2004). They believe the substrate signal peptide may adopt a partially helical structure in vicinity of membrane. This aids in its binding to the TatABC complex, specifically to the TatBC site where the first interaction is known to occur (Alami *et al.*, 2003). Binding is thought to trigger the recruitment of homooligomeric TatA complexes, resulting in the generation of an active super complex. Translocation then occurs, possibly driven by the proton motive force.

1.2.2.4 ATP binding cassette transporters

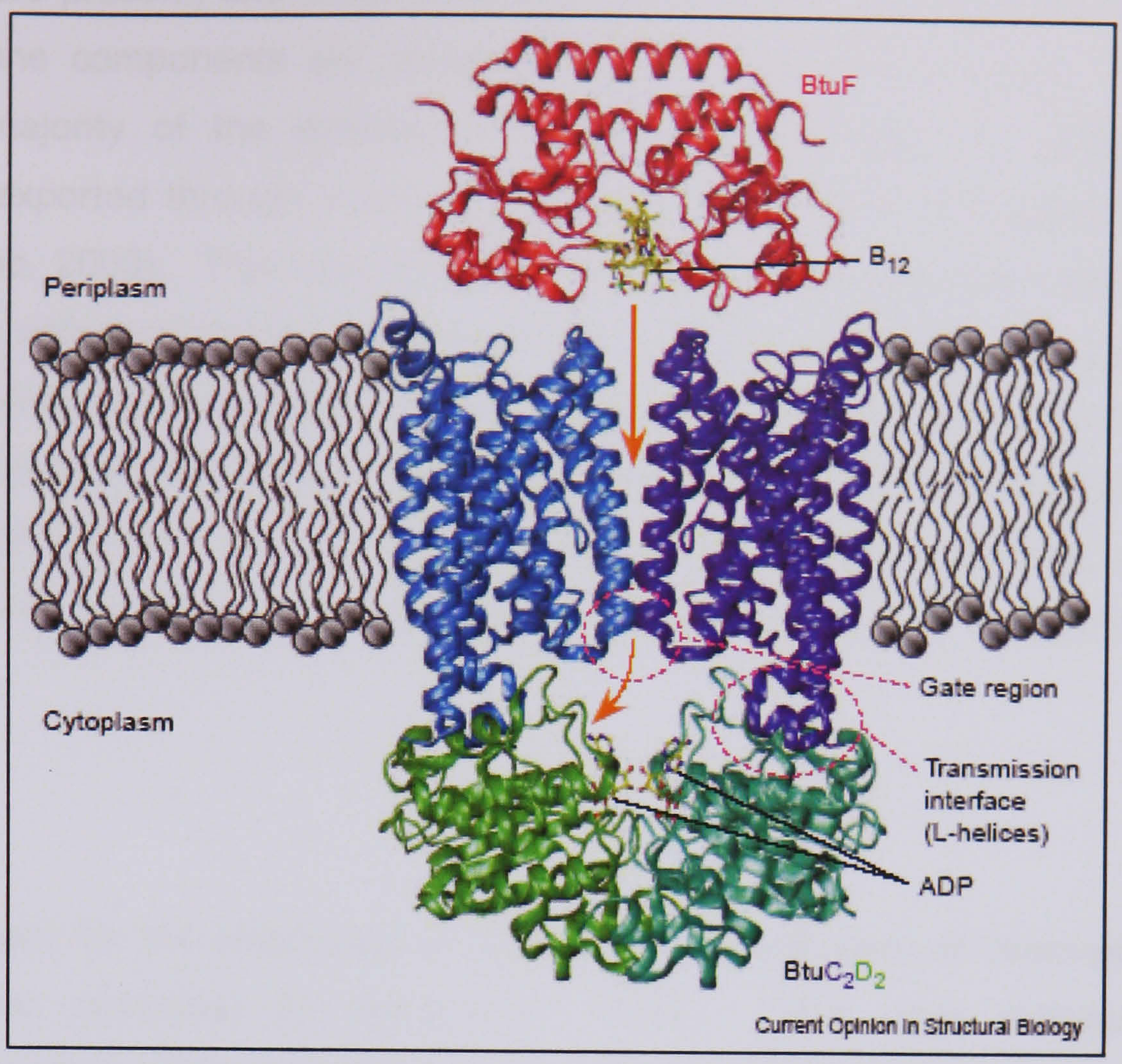
ATP-binding cassette (ABC) transporters facilitate the transport of chemically diverse substrates across a cell membrane in a single direction (reviewed in Locher & Borths,

2004). Even though the transporters must be diverse to translocate a wide variety of substrates, they do appear to share a common architecture and basic structure; a common engine attached to a specialised translocation pathway (Holland & Blight, 1999). Two ABC domains in the system bind and hydrolyse ATP to energise translocation, and the translocation pathway is provided by two more variable membrane-spanning domains (MSDs) (Schneider & Hunke, 1998). The crystal structure of the ABC importer of vitamin B_{12} was determined by Locher (2002), and is shown in Figure 1.6A. Importers are often composed of separate subunits, whereas exporters often contain the ABC and MSD domains fused into a single polypeptide. The chemical to be imported is delivered by an extra-cytoplasmic specific binding protein (Nikaido & Hall, 1998). The binding proteins exhibit high affinity and selectivity to their substrates, which is thought to ensure specificity of transport (Quiocho & Ledvina, 1996). The mechanism by which translocation occurs is still under debate. One proposed scheme is interface modulation (Figure 1.6B). Initially the central cavity of the importer is open to the periplasm or medium and closed to the cytoplasm by interacting loops between the dimers. As the binding protein and substrate docks, a signal is sent to the ABC domain. The ABC domains are thought to move by several angstroms, pinning two ATP molecules into the partner ABC domains (Locher *et al.*, 2002), and hydrolysis occurs. The conformational change caused by the tight engagement of the ATP molecules forces the cytoplasmic gate open, allowing the substrate to escape into the cytoplasm. The binding partner protein may serve as an additional gate to prevent the substrate from being released to the wrong side of the membrane. After ATP hydrolysis, the conformational changes force the binding protein to dissociate from the complex. This energy dependent, conformational change mechanism could control the proper sequencing of gate opening and closing. Finally, the ABC-MSD complex returns to its resting state. A major benefit of this proposed tilting mechanism is that access is alternated between the two sides of the membrane, but never both together, maintaining membrane integrity. Although this mechanism has been proposed based on the structure of the vitamin B_{12} translocon, the arrangement of ABC and MSD domains appears to be conserved, and so a similar mechanism may apply to other ABC-transporters.

1.2.2.5 Flagella export apparatus

The bacterial flagellum is both a motor structure and a protein export/assembly apparatus. The flagellum extends from the cytoplasm to the cell exterior, and the substructures of the flagellum (section 1.2.1.5) are built up in a sequential fashion, proceeding from more proximal structures to more distal ones, i.e. from integral and

A



B

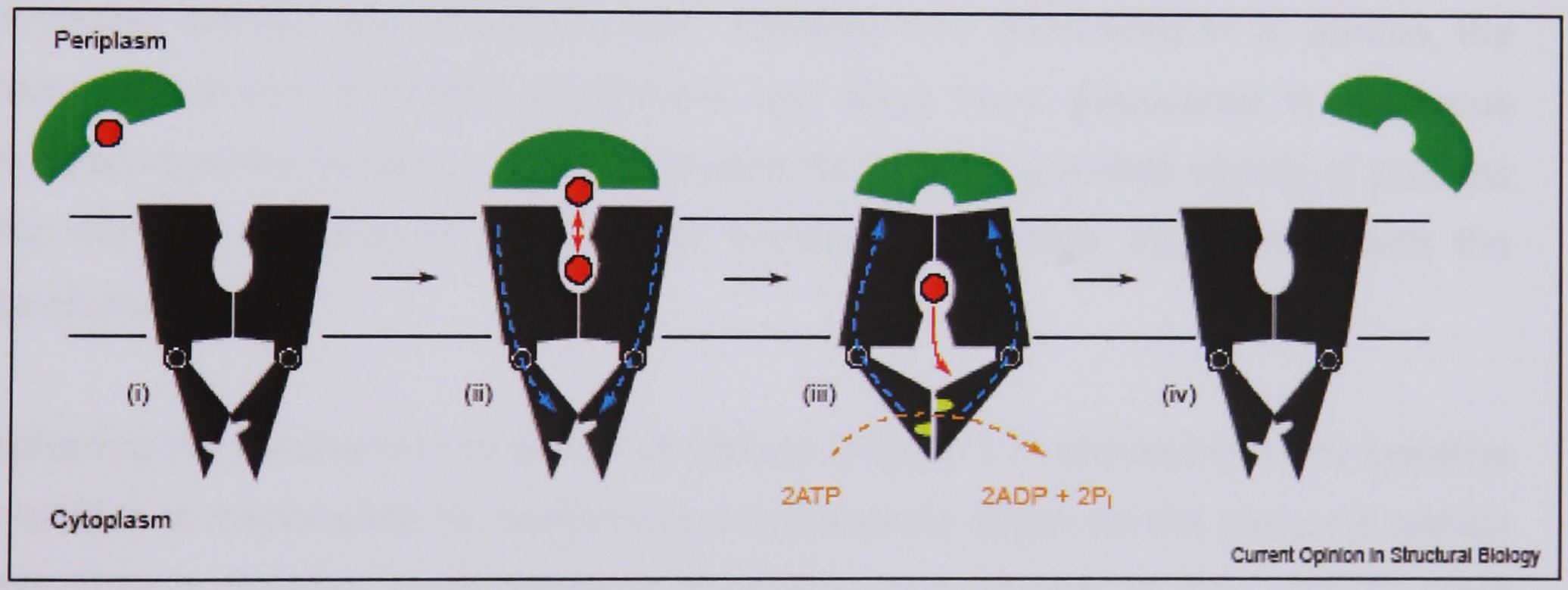


Figure 1.6 A. Ribbon diagram of a typical ABC transporter, BtuCD. BtuF captures the substrate, vitamin B_{12}, and places it over the translocation pathway at the interface of the MSDs. When ATP is not bound, the extracytoplasmic side of the gateway is in an open configuration. From Locher *et al.* (2002). **B. Schematic mechanism of transport.** The red circle represents vitamin B_{12} and the green shape represents BtuF, the binding protein. Conformational changes caused by ATP hydrolysis may allow substrate to enter the cytoplasm by an open/closed gating system. From Locher (2004).

peripheral membrane components (to periplasmic components and outer membrane components where present) and finally to components that lie in the cell exterior. The integral membrane components are secreted by the sec-pathway (section 1.2.2.2). However, the majority of the proteins that are located beyond the cytoplasmic membrane are exported through a specialized pathway, the type III flagellar export pathway (Macnab, 2003). This type of export pathway characteristically has a well-defined structure with a continuous central channel in the nascent structure. Subunits arrive in the channel via apparatus in the MS ring at the base of the structure in the cytoplasmic membrane, and diffuse in single-file along the length of the flagellum to be added at the distal end (Suzuki *et al.*, 1998). Proteins exported by this pathway contain no cleavable signal peptide, and do not appear to have a conserved recognition site (Macnab, 2003).

1.2.2.6 Sortase

A novel mechanism for the attachment of proteins to the cell walls of Gram-positive bacteria was first described by Schneewind (1992). This work showed that transpeptidase sortase enzymes catalysed the attachment of proteins containing a specific motif to the peptidoglycan cell wall. Protein substrates for sortase-mediated attachment must contain a conserved LPXTG motif near the C-terminus, a C-terminal hydrophobic domain and a charged tail. Although first discovered in *S. aureus*, the sortase mechanism of protein attachment has since been discovered in numerous other Gram-positive bacteria. It is responsible for anchoring a wide variety of proteins to the surface of bacterial cells, which function to facilitate interactions with the extracellular milieu.

Deciphering the mechanism of action of sortase (Figure 1.7) showed that the cysteine 184 residue is responsible for performing a nucleophilic attack on the carbonyl carbon at the scissile peptide bond between threonine and glycine of the LPXTG motif (Schneewind *et al.*, 1995; Ton-That *et al.*, 1997). This results in the formation of a thioester with the carboxyl-group of threonine, while the amino-group of the cleaved LPXTG motif is released (Ton-That & Schneewind, 1999). Sortase then anchors its substrate to the tetrapeptide murein subunits or free unsubstituted pentapeptide subunits of the peptidoglycan. This mechanism of attachment of proteins to the cell wall was shown to occur in many Gram-positive bacteria, even though the precise details of attachment can vary (Schneewind *et al.*, 1993; Navarre & Schneewind, 1999; Dhar *et al.*, 2000). Typically, Gram-positive bacteria encode two sortase genes (Pallen *et al.*, 2001). However, the sequenced strain of *C. difficile* (630) has only one intact

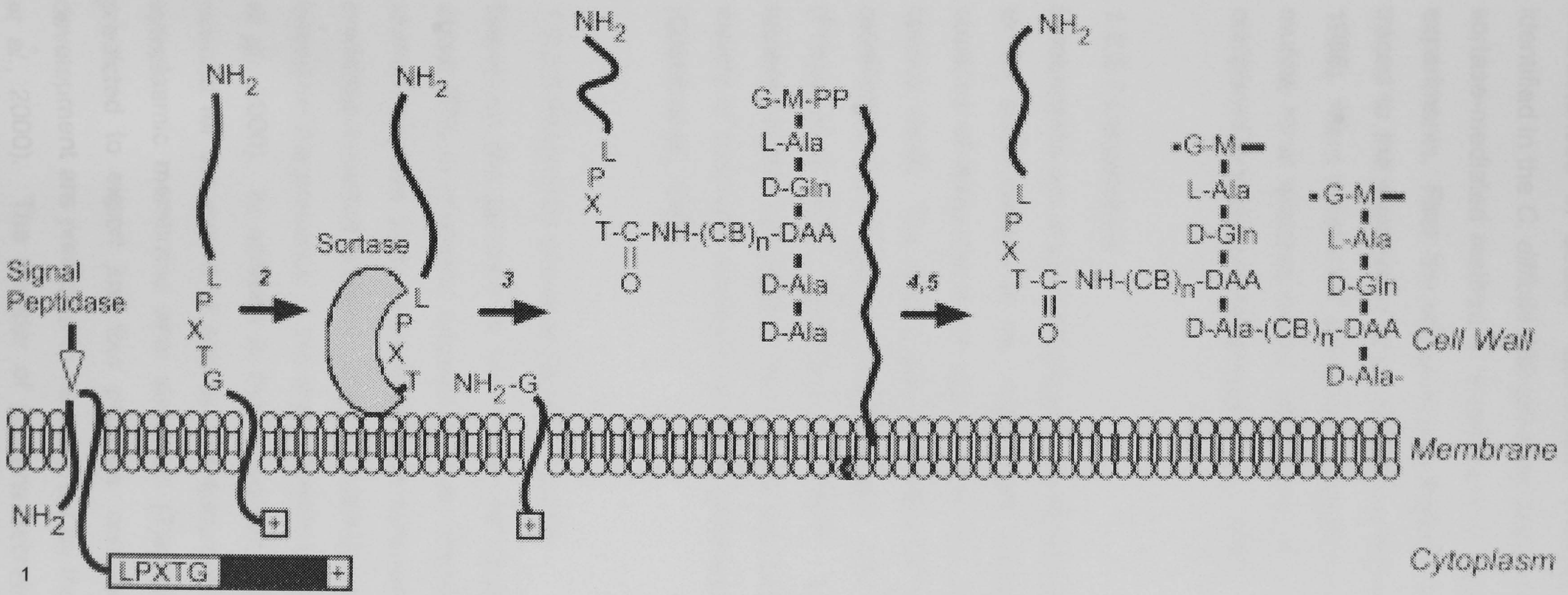


Figure 1.7 Sortase - mediated protein anchoring in *S. aureus*. (1) **Export.** Pre-proteins with the appropriate signal peptide are targeted to the Sec-pathway. (2) **Retention.** The C-terminal sorting signal retains the polypeptide within the Sec-translocon. (3) **Cleavage.** Sortase cleaves the LPXTG motif between threonine and glycine forming a thioester intermediate. (4) **Linkage.** Nucleophilic attack of the free amino group of lipid II at the thioester bond resolves the acyl-enzyme intermediate, synthesising the amide bond between surface proteins and the pentaglycine cross-bridge and regenerating the active site sulphydryl. (5) **Incorporation**. The transglycosylation reaction incorporates the lipid-linked surface protein into the cell wall. The murein pentapeptide subunit with attached surface protein is then cross-linked to other cell wall peptides, generating the mature murein tetrapeptide. From Ton-That *et al.* (2004).

sortase gene; the other gene contains a frame-shift mutation. A number of genes for sortases have been recognised in other clostridial species, however, none have been characterised in detail. Seventeen putative sortase substrates have also been identified in the *C. difficile* 630 genome, discussed in Chapter 3. The importance of the sortase-mediated method of attachment in virulence has been shown in a number of experiments. First, the adhesive and serum binding properties of *S. aureus* have been traced to the expression and correct localisation of surface proteins (Foster & Hook, 1998). More substantial evidence comes from the analysis of sortase mutants (*srt*) in a murine renal abscess model, where the LD_{50} of *srt* mutants is reduced by 1.5 logs compared to wild type (Mazmanian *et al.*, 2000).

1.2.2.7 Lipoproteins

Lipoproteins are a specific group of secreted proteins that remain covalently attached to the outer surface of the cytoplasmic membrane after translocation. The signal peptides of lipoproteins are shorter than sec-pathway signal peptides and contain a lipobox motif. The final cysteine residue of the lipobox motif is the target for lipid modification by diacylglyceryl transferase, and anchors the protein to the membrane (Pragai *et al.*, 1997; Tjalsma *et al.*, 1999). Whilst the N-termini of lipoproteins are tethered to the cytoplasmic membrane, the C-terminal domains are thought to fulfill a variety of functions similar to those of periplasmic proteins in Gram-negative bacteria (Gilson *et al.*, 1988).

1.2.2.8 Frequency of use of secretion routes

Based on the genome of the archetypal Gram-positive organism, *Bacillus subtilis,* about 70% of predicted proteins lack an amino-terminal signal peptide or membrane anchor, and are likely to be located in cytoplasm (Tjalsma *et al.*, 2000). Of the 180 predicted exported proteins, about 160 are thought to be exported by the sec-pathway, based on the presence of the characteristic tripartite structured signal peptide (Tjalsma *et al.*, 2000). In addition to these secreted proteins, the sec-dependent pathway is needed for the export of approx 114 lipoproteins, which will remain anchored to the cytoplasmic membrane after secretion (Tjalsma *et al.*, 1999). The ABC-system is predicted to export just four proteins, and four proteins required for competence development are predicted to be exported by the pseudopilin export pathway (Tjalsma *et al.*, 2000). The number of proteins secreted by the TAT-pathway is difficult to estimate on the basis of signal peptide prediction. A small proportion of exported proteins are thought to be specifically attached to wall (~ 0.5%), the remainder of

exported proteins (~ 4%) have the potential to pass through the cell wall and be secreted into the environment (Tjalsma *et al.*, 2000).

Occasionally other methods of attachment have been found, sometimes involving anchoring to SCWPs. For example, the InlB invasion protein of *L. monocytogenes* is anchored to lipoteichoic acids via its cell surface anchor domain containing GW tandem repeats.

1.2.3. The S-layer

Surface layers (or S-layers) represent the outermost boundary layer in many prokaryotes, external to, and in close apposition with the cell envelope (Sleytr, 1978). They form a paracrystalline two-dimensional array, covering the surface of the cell, and are usually composed of a single (glyco)protein species. S-layers are found in all types of prokaryotes, the archaea (where the S-layer occasionally forms the exclusive cell wall component), as well as Gram-positive and Gram-negative bacteria (Sleytr & Sara, 1997). Apart from their localisation and organisation, S-layers are extremely diverse, showing a range of structures and functions.

Structural studies have been undertaken on a number of different SLPs. Circular dichroism was used to show the composition of *Azotobacter vinelandii* SLP consists mainly of β-sheet (Bingle *et al.*, 1987a). The SLPs of *Methanothermus fervidus* and the Gram-positive bacterium *Geobacillus stearothermophilus* have recently been crystallised, however the 3-D crystal structure of these proteins has not yet been reported (Pavkov *et al.*, 2003). Only one 3-D crystal structure of an SLP has so far been reported. Of the three domains of the *Methanosarcina* SLP, two have been crystallised by Jing *et al.*, (2002). One domain contains a unique, highly symmetrical structure with a seven-bladed β-propeller fold. The other crystallised domain contains a polycystic kidney disease superfamily fold, while the third domain is predicted to adopt a β-helical fold. Despite these advances in determining the structures of some SLPs, very little can be conferred as to the structure of SLPs from other organisms due to the heterogeneity of this group of proteins.

S-layers are detected and visualised by electron microscopy in combination with negative staining or freeze-etching (Figure 1.8). Two- and three-dimensional image reconstruction has revealed structural information to 1 nm (Lupas *et al.*, 1994; Sleytr & Sara, 1997). Atomic force microscopy has been used to view S-layer surface topology, and shows the opening and closing of channels within S-layer subunits

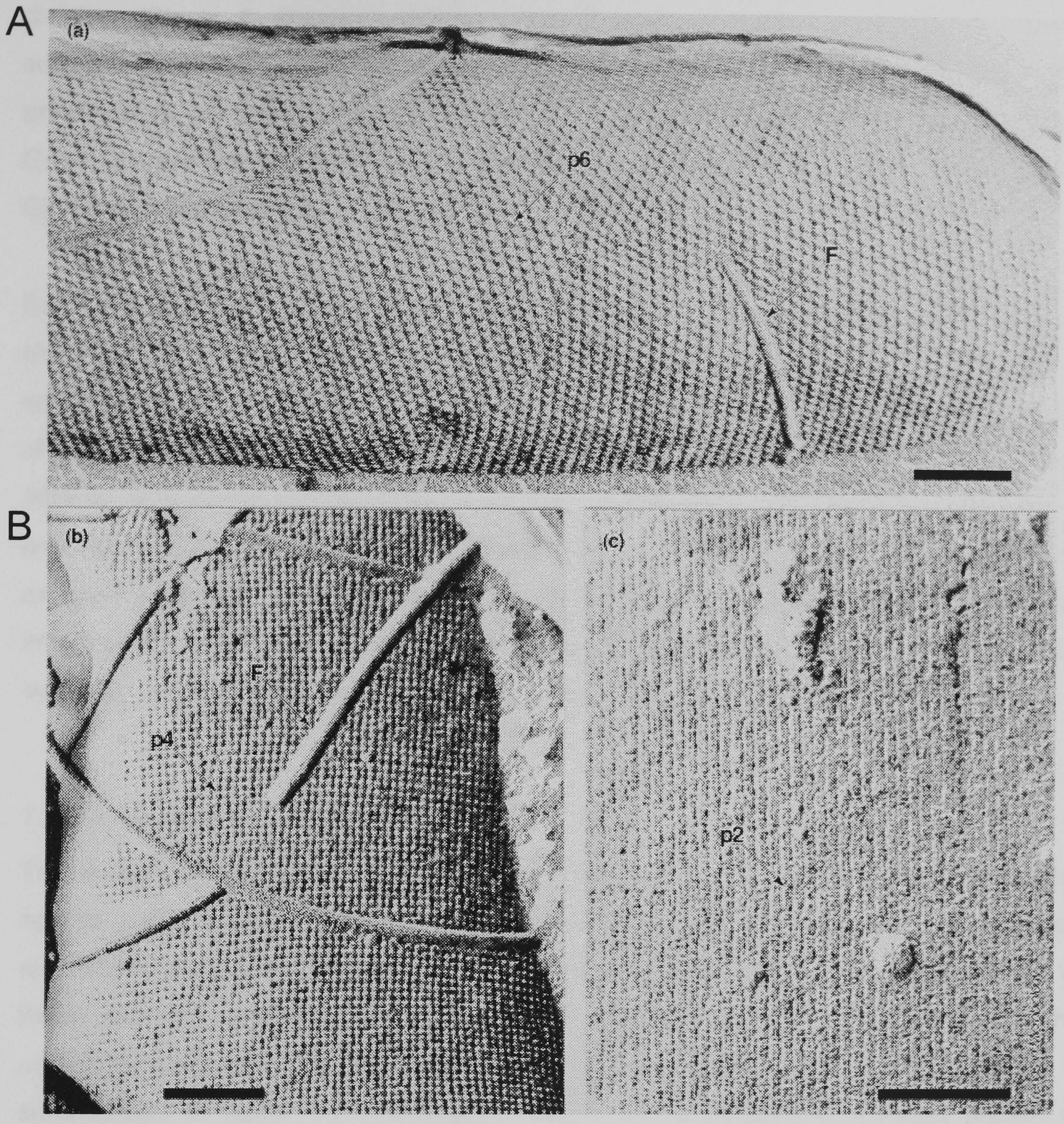


Figure 1.8 Electron micrographs of S-layers. Electron micrographs of freeze-etched preparations of intact cells of: (a) *Thermoanaerobacter thermohydrosulfuricus* with hexagonal (p6) surface lattice; (b) *Desulfotomaculum nigrificans* with square (p4) lattice; (c) *Lactobacillus helveticus* with an oblique (p2) lattice. 'F' are flagellae overlying the S-layer. Scale bars = 100 nm. From Sleytr and Beveridge (1999).

(Muller *et al.*, 1996). These methods have shown that S-layers have a lattice structure, 5-25 nm thick, that completely covers the surface of the cell at all times of growth and division (Sleytr & Glauert, 1976). The S-layer is composed of one or more superimposed layers, and has a smooth outer and corrugated inner surface. The presence of an S-layer creates an additional surface compartment in archaea and Gram-positive bacteria that may be functionally equivalent to the periplasmic space in Gram-negative bacteria.

Each morphological unit of the S-layer is composed of 1-6 identical subunits with oblique (p1, p2), square (p4) or hexagonal (p3, p6) symmetry, and centre-to-centre spacing of 2-35 nm (Figure 1.9). Lattice pores in the crystalline protein meshwork are of identical size and morphology, and are 2-8 nm in size. The pores can occupy from 30% to 70% of the surface area of the cell. It takes 5×10^5 monomers to cover an average rod shaped cell, and so transcription and translation of the genes must be extremely efficient. The morphological units are able to spontaneously self-assemble into multimers or even whole two-dimensional sheets, either in suspension or on surfaces, forming very low free energy structures (Bingle *et al.*, 1987a, 1987b).

1.2.3.1. Biochemistry of the S-layer

The S-layer is normally isolated from purified cell wall fragments using chaotrophic agents such as urea or guanidine hydrochloride, detergents at low pH or cation substitution (Li^+ or Na^+ replacing Ca^{2+}) (Koval & Murray, 1984). The effectiveness of these methods suggests that non-covalent interactions anchor the S-layer to the underlying cell wall. The composition of S-layer proteins from several species is summarised in Table 1.1. Normally, S-layers are composed of a single protein or glycoprotein species. A number of secondary modifications to SLPs have been reported, including cleavage of the amino- or carboxyl-termini, phosphorylation and glycosylation (Bahl *et al.*, 1997; Schaffer & Messner, 2001). In a few cases two superimposed S-layers each composed of a different protein species have been found. Two SLPs have been found in *C. difficile* (Calabi *et al.*, 2001), *Campylobacter rectus* (Kobayashi *et al.*, 1993), *B. anthracis* (Engelhardt & Peters, 1998) and Aquaspirillum (Kist & Murray, 1984). In all cases, antibodies against each SLP do not cross-react with the other, but it is unclear whether two S-layers are formed by all these species at all times.

In Gram-positive bacteria, the S-layer assembles by the insertion of multiple, helically arranged bands on cylindrical part of cell, whereas in Gram-negative bacteria,

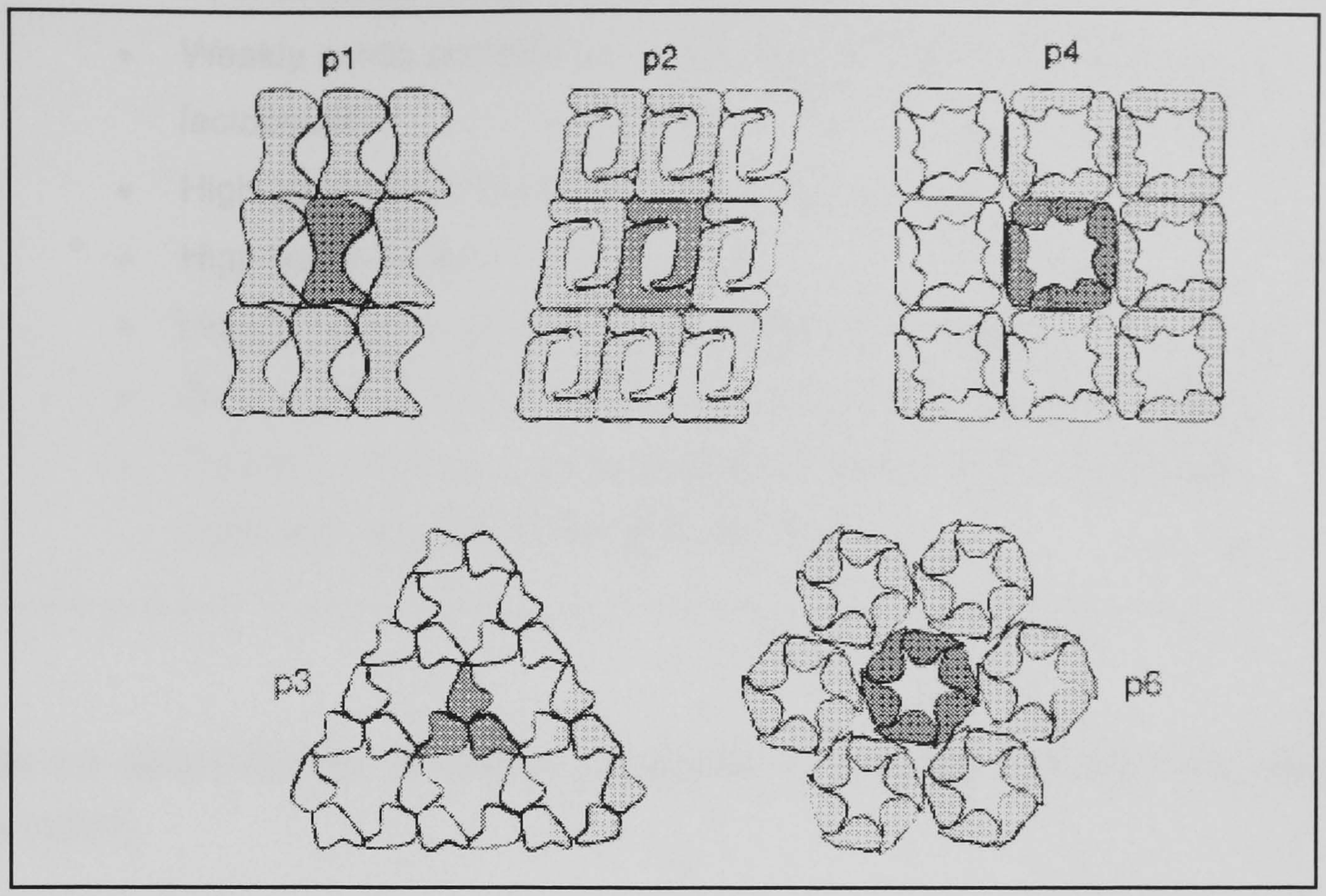


Figure 1.9 Schematic of different S-layer lattice types. The regular arrays exhibit oblique (p1, p2), square (p4) or hexagonal lattices (p3, p6). The morphological units are composed of one, two, three, four or six identical subunits. From Sleytr (1999).

- M_r of constituent subunits in the range of 40,000-170,000
- Weakly acidic proteins (pI = 3-6) (except *Methanothermus* and lactobacilli)
- High glutamic and aspartic acid content (~ 15%)
- High lysine content (~ 10%)
- High in hydrophobic amino acid residues (~ 40-60%)
- Zero or low content of sulphur-containing amino acid residues
- Circular dichroism shows approximately 40% of protein organised as β-sheet and 20% as α-helix.

Table 1.1 Characteristic chemical properties of S-layers. Modified from Sleytr & Sara (1997).

assembly occurs at random (Gruber & Sleytr, 1988; Sleytr & Beveridge, 1999). The detailed mechanism by which this occurs is unknown.

1.2.3.2 Genetics of the S-layer

The SLP genes have now been cloned and sequenced from over 40 different organisms covering a wide range of taxa (reviewed in Sleytr (1999)). No similarity exists between the SLP sequence in archaea and eubacteria, and there is only some similarity between closely related species. For example, *Lactobacillus acidophilus* and *Lactobacillus helveticus* SLPs are 80% similar, but have barely any similarity to the SLP of *Lactobacillus brevis* (Sleytr, 1999). There are some local similarities between Gram-positive and Gram-negative SLPs at the amino-termini because of binding specificities to components of supporting cell wall layers. An example of this is the surface layer homology (SLH) domain, discussed in 1.2.3.3.

SLPs are the predominant protein of cells with an S-layer, forming up to 15% of the total protein content. With a bacterial generation time of 20 minutes, up to 500 protein molecules must be synthesised per second to maintain coverage of the cell. Therefore, regulation of SLP production is critical and is thought to occur at the levels of transcription, translation and secretion. Strong promoters are required to drive transcription, and in fact, the SLP promoter in *L. acidophilus* is twice as strong as the *lac* promoter of *Escherichia coli*, one of the strongest promoters known (Boot *et al.*, 1996a). Investigations into the synthesis of *L. brevis* SLPs found that in this organism at least, SLP production very tightly regulated. Transcription is directed by two promoters, with the level of P2 transcripts being ten times higher than P1 transcripts at all stages of the cell growth cycle. Additionally, the transcript is made up of stable mRNA with a long 5' untranslated region, which has the potential to fold into stable secondary structures, with a half-life of up to 14 min, allowing translation for longer periods of time (Boot & Pouwels, 1996; Kahala *et al.*, 1997).

A number of organisms are able to switch their SLP type. Of the eight SLP variants that exist in *Campylobacter fetus,* only one is expressed at a given time. The SLP variant expressed can switch in a RecA-dependent manner (Dworkin *et al.*, 1995; Dworkin *et al.*, 1997). SLP expression uses a single promoter on an invertible DNA segment, thus allowing antigenic variation and avoidance of the host's immune response (Dworkin & Blaser, 1996). Other species that can switch SLPs include *L. acidophilus* (Boot *et al.*, 1996a; Boot *et al.*, 1996b), *Bacillus sphaericus* (Bowditch *et*

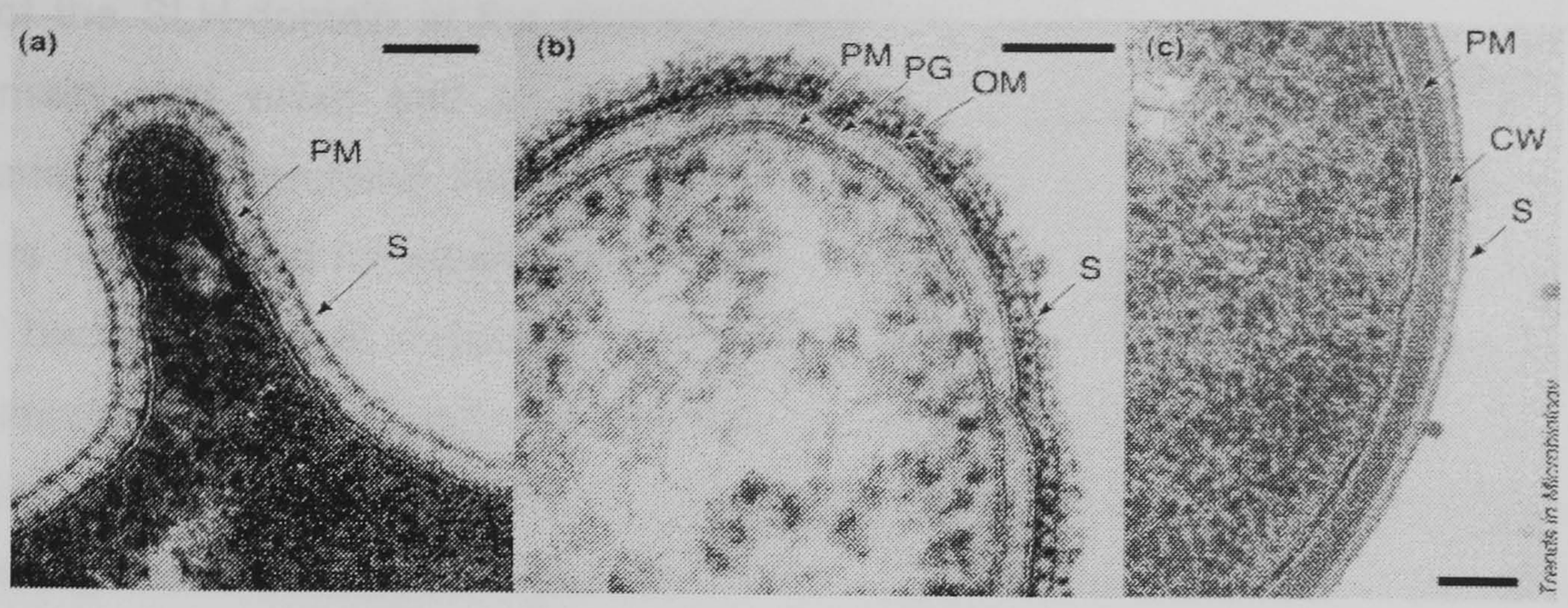
al., 1989), *Rickettsia prowazekii* (Ching *et al.*, 1992), *Bacillus stearothermophilus* (Bahl *et al.*, 1997) and *L. brevis* (Jakava-Viljanen *et al.*, 2002).

The SLP gene sequences from different strains of *C. glutamicum* were found to show a mosaic structure composed of conserved and variable domains. Two domains were found to be especially conserved, those at the amino-terminal signal peptide and the carboxyl-terminal anchor region (Hansmeier *et al.*, 2004). Studies in *B. sphaericus* suggest that S-layers may have evolved and spread via horizontal DNA transfer followed by gene rearrangements (Pollmann *et al.*, 2005).

1.2.3.3 Attachment

Several modes for the attachment of the S-layer to the underlying cell wall components have been described (Figure 1.10). In eubacteria, the SLPs are most commonly attached to cell wall polysaccharides. In Gram-positive bacteria, the SLPs interact with peptidoglycan or secondary cell wall polymers (discussed in section 1.2.1.2) (Masuda & Kawata, 1980; Mesnage *et al.*, 2000; Schaffer *et al.*, 2000). Gram-negative bacteria SLPs bind to the O-polysaccharide side chains of lipopolysaccharide (LPS) (Engelhardt & Peters, 1998; Weygand *et al.*, 1999). One common method of anchoring SLPs to the underlying peptidoglycan is through an SLH domain (first described by Lupas *et al.* (1994)). This is a domain spanning 50-60 amino acid residues, which is predicted to adopt a distinctive secondary structure with two α-helices separated by short β-strand and loop (H_I – S – L – H_{II}). Despite structural conservation between SLH domains from different species, the DNA sequence encoding the SLH domain varies, and can be difficult to detect. The SLH-domain is sufficient for targeting to the cell wall. However, at least in *B. anthracis*, the SLH domain does not attach directly to the cell wall but instead to secondary cell wall polymers (Egelseer *et al.*, 1998; Mesnage *et al.*, 1999a, 1999b; Brechtel & Bahl, 1999). Mesnage and co-workers found that the SLH-domain in the two SLPs produced by *B. anthracis,* EA1 and Sap, interact with pyruvylated polysaccharides on the peptidoglycan. The gene required for pyruvylation, *csaB* is located immediately upstream from the SLP genes. Other species with SLH domains within their SLPs also contain genes homologous to *csaB*, so the authors postulate that pyruvylation of polysaccharides is a trademark of the anchoring of SLPs via the SLH-domain (Mesnage *et al.*, 2000). A number of proteins other than SLPs also employ SLH-domains for attachment, and therefore SLH-mediated attachment has been classified into groups. Group I consist of SLPs; group II are cell bound exoenzymes or other proteins involved in polysaccharide degradation; and group III are outer membrane proteins such as porins (Engelhardt & Peters, 1998). In SLPs there are 2-3

A



B

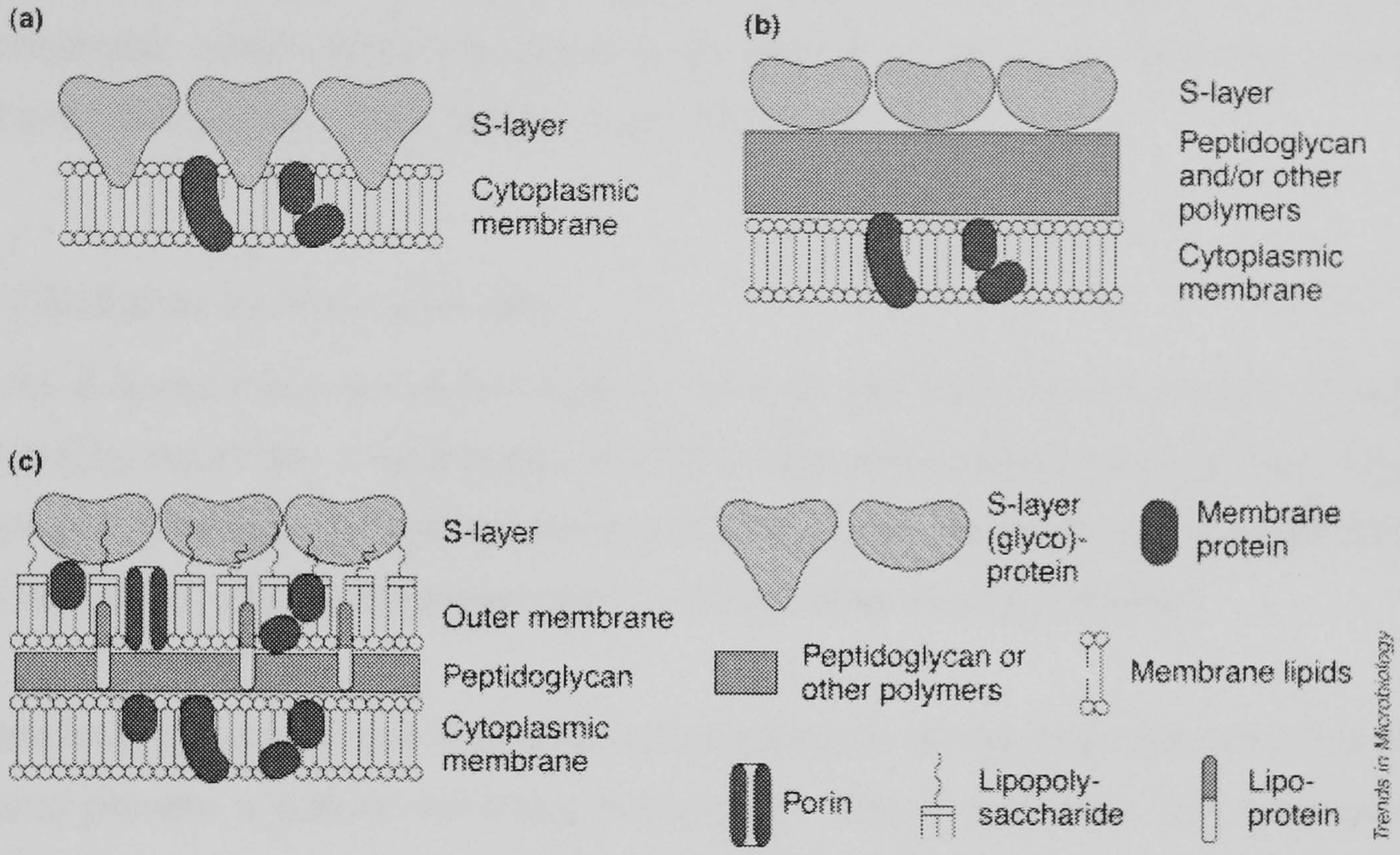


Figure 1.10 Electron micrograph and schematic of prokaryotic S-layers. A. Electron micrographs of thin sections of **(a)** an archeon *Sulfolobus acidocaldarius*, **(b)** a Gram-negative bacterium *Aeromonas salmonicida* and **(c)** a Gram-positive bacterium *Bacillus thuringiensis.* Abbreviations: CW, Gram-positive cell wall; OM, outer membrane; PG, peptidoglycan; PM, plasma membrane; S, S-layer. Scale bar = 50 nm. **B.** Schematic illustration of the supramolecular architecture of the three major classes of prokaryotic cell envelopes containing crystalline bacterial cell surface layers. **(a)** Cell envelope structure of Gram-negative Archaea with S-layer as only component external to the plasma membrane. **(b)** Gram-positive Archaea and Bacteria. In Bacteria, the rigid cell wall component is peptidoglycan; in Archaea it is composed of other polymers e.g. pseudomurein. **(c)** Gram-negative Bacteria with thin peptidoglycan and outer membrane. If present, the S-layer is associated with the lipopolysaccharide of the outer membrane. From Sleytr and Beveridge (1999).

copies of the SLH-domain at the amino-terminus; in group II, multiple copies of the SLH-domain also exist; and in group III, there is only a single copy, but the oligermerisation of proteins into trimers or tetramers occurs. However, the SLH-domain is not required for attachment of S-layer proteins to the cell wall in all Gram-positive bacteria. An SLH-domain is absent from the SLPs of *Lactobacteriaceae*, which recognise neutral polysaccharides (Masuda & Kawata, 1985). In *B. stearothermophilus* the SLPs bind to teichuronic acids as a means of attachment to the cell wall (Egelseer *et al.*, 1998; Schaffer *et al.*, 1999). In archaea and some Gram-positive and Gram-negative bacteria (e.g. *Corynebacterium glutamicum* and *R. prowazekii*), the SLPs are anchored simply to the underlying membrane by a sequence of hydrophobic amino acids predicted to consist of a membrane spanning α-helical coiled coil (Ching *et al.*, 1992; Chami *et al.*, 1997).

1.2.3.4 Evolution and functional roles

Bacterial S-layers have an ancient history, as they can be found on organisms widely separated by evolution. It is probable that structural and functional divergence followed in response to widely separate environmental pressures, although it is possible that S-layers may have originated independently several times during evolution.

S-layers must be subject to a strong selective pressure, as the large amounts of protein produced present a substantial biosynthetic cost. This is consistent with findings that certain bacteria lose their S-layers under laboratory conditions, and outgrow their wild-type counterparts. As yet, an all-encompassing natural function has not been found for the S-layer, but a number of specific functions have been investigated (reviewed in Beveridge *et al.*, (1997)).

The S-layer is a porous barrier, and can therefore act as a molecular sieve, regulating the exchange of molecules between the cell and the environment (Sara & Sleytr, 1987). The S-layer lattices of some mesophilic and thermophilic bacilli have sharp exclusion limits between 15-45 kDa, and permeability studies show the pores have excellent anti-fouling properties, essential for the exchange of metabolites and nutrients (Sara & Sleytr, 1987). Thus an outer cellular compartment is created, external to the plasma membrane, where macromolecules that are important for the processing/transport of nutrients or folding/export of secreted or membrane proteins can be retained (Beveridge & Graham, 1991). This molecular sieve function may also afford protection from noxious agents such as biocides, lytic enzymes or prokaryotic predators. For example, the S-layer in *C. fetus* and *Cauobacter crescentus* protects

against attack from the carnivorous bacteria *Bdellovibrilo bacteriovorous* (Koval & Hynes, 1991). The S-layer lattice of *Synechococcus* enables it to grow in lakes with exceptionally high gypsum and calcite concentrations. Its hexagonal lattice acts as template for fine-grain mineralisation and is continuously shed from the cell surface of the organism, preventing the clogging of further cell envelope components (Schultze-Lam *et al.*, 1992). The S-layer can also act as an attachment site for exo-enzymes. For example, the SLP of *B. stearothermophilus* has been shown to act as an adhesion site for amylase (Egelseer *et al.*, 1995). The S-layer has also been shown to have a role in cell shape and the initiation of fission in two archaea (Messner *et al.*, 1986; Pum *et al.*, 1991).

Several pieces of evidence indicate that S-layers can be virulence factors, i.e. determinants of disease, in human and animal pathology. The S-layer of *C. fetus* enables the organism to resist phagocytosis and the bactericidal activity of serum because the C3b component of the complement cascade is unable to bind to it (Blaser & Pei, 1993; Dworkin & Blaser, 1997). Mutations preventing S-layer synthesis or chemical removal of S-layer make the organism sensitive to serum and phagocytosis-mediated killing. It is hypothesised that the S-layer masks other immunogenic structures e.g. LPS, also aiding avoidance of the immune system (Dworkin & Blaser, 1997). Loss of the S-layer is associated with a decreased adherence of *C. rectus* cells to human gingival fibroblasts, and loss of the S-layer upon subculturing *in vitro* correlated with reduced virulence in a mouse model (Kesavalu *et al.*, 1991). Furthermore, the ability of *C. fetus* to undergo SLP variation may enhance its virulence. Phase variation of SLPs in *C. fetus* has been associated with relapsing infections. This was found to occur by three mechanisms, by upregulating virulence, by antigenic variation and by allowing the bacterium to become latent (Tu *et al.*, 2005). In *Bacillus cereus*, a Gram-positive spore forming facultative anaerobe and an agent of both gastrointestinal and non-gastrointestinal disease, the S-layer has been shown to mediate tissue adherence. Strains with an intact S-layer show increased binding to extracellular matrix components such as type I collagen, laminin, fibronectin and fibrinogen compared to cultures that have lost the S-layer through *in vitro* subculturing (Kotiranta *et al.*, 2000). The HMW SLP of *C. difficile* has also been shown to have a role in adherence, discussed in section 1.3.

The SLPs of *Rickettsia typhi* and *R. prowazekii* elicit protective immune responses in animals. Vaccination of guinea pigs with purified SLP prevented disease caused by challenge with live *Rickettsiae* by stimulation of both the humoral and cell-mediated arms of the immune system (Carl & Dasch, 1989).

1.3. The cell wall of *C. difficile*

The S-layer of *C. difficile* is very unusual in that it has two SLPs derived from a common precursor, SlpA (Calabi *et al.*, 2001). Processing of SlpA involves removal of a signal peptide and a second internal cleavage at position 355 in strain 630 to release the two mature SLPs (Figure 1.11A). The high molecular weight (HMW) SLP is highly conserved between strains and shows homology to *N*-acetyl muramoyl-L-alanine amidase from *B. subtilis*. The HMW SLPs range in size from 48-56 kDa and antibodies against them do cross-react with those from other strains. The HMW SLP shows peptidoglycan hydrolase activity, as determined by zymogram on *Micrococcus luteus* cell walls. The low molecular weight (LMW) SLP shows much greater sequence diversity between strains (Calabi & Fairweather, 2002). The LMW SLPs range from 37-45 kDa in size and show no antigenic cross-reaction between strains (McCoubrey & Poxton, 2001). The variability of the LMW SLP could make it useful in typing as an alternative to serotyping (Karjalainen *et al.*, 2002). Freeze etch electron microscopy on the cell surface of *C. difficile* revealed two S-layers. The outer S-layer was arranged in a square symmetry while the inner S-layer had hexagonal symmetry (Cerquetti *et al.*, 2000). This suggests, but does not prove, that each S-layer in this organism is composed of different protein species, comprised of units of the HMW- and LMW SLPs.

A family of 28 gene products related to the HMW SLP are found in the genome of *C. difficile* strain 630. Each of these gene products, or paralogues, contains a domain with homology to an amidase from *B. subtilis*. Eleven of the genes encoding Slp paralogues are clustered around *slpA* (Figure 1.11B), and almost all of the gene products contain a second, variable domain. These variable domains are not related to each other by sequence, but do show similarity between strains where studied, and are thought in some cases to give the protein its function (Calabi & Fairweather, 2002). Six of these genes were shown by reverse transcriptase-polymerase chain reaction (RT-PCR) to be transcribed (Calabi *et al.*, 2001). Most paralogues also contain a signal sequence, indicating the possibility of surface localisation. Indeed, the gene downstream from s*lpA* is one of two copies of *secA* in the genome, whose gene product is an essential component of the general secretory pathway (section 1.2.2.2).

The HMW SLP has been found to mediate binding of *C. difficile* to gastrointestinal tissues (Calabi *et al.*, 2002). ELISA and FACS analysis show that S-layer proteins extracted from *C. difficile* and recombinant HMW SLP bind to HeLa cells, and to human and mouse gastrointestinal tissues. Additionally, anti-serum against the HMW SLP partially blocks adherence. The HMW SLP binds to epithelial and connective tissue of

A

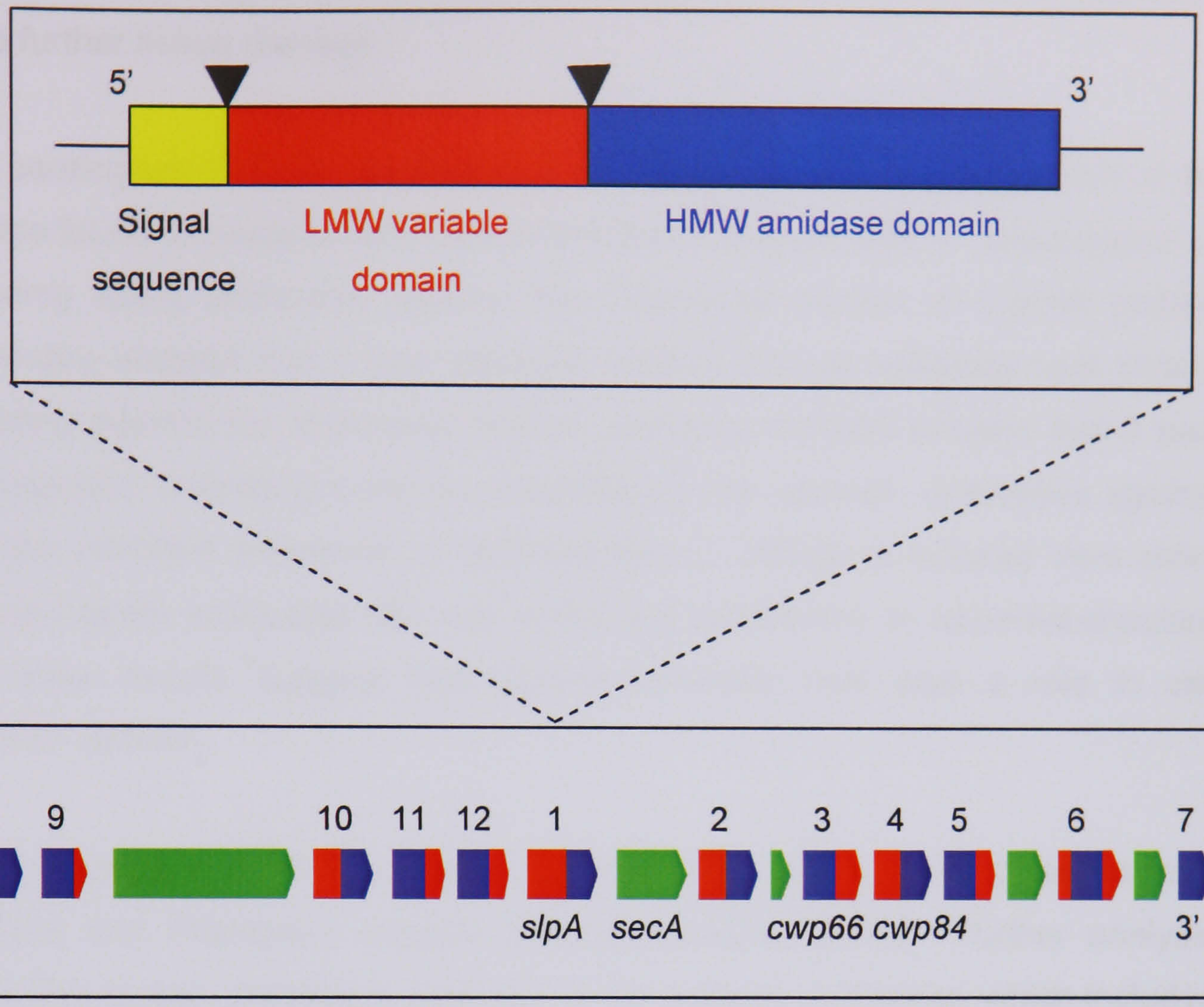


Figure 1.11. Arrangement of ORFs in the region of *slpA* of *C.difficile*. **A.** Domains of SlpA. Arrows indicate post-translational cleavage sites. **B.** Analysis of a region of the genome sequence of *C. difficile* strain 630 (www.sanger.ac.uk/Projects/C_difficile) revealed a series of ORFs that show amino acid sequence homology to the C-terminal domain (the HMW-subunit) of SlpA, and to the *N*-acetyl muramyl-L-alanine amidase from *B. subtilis.* These ORFs are all potentially secreted as they contain signal sequences. Blue shading indicates regions of homology to amidase, and the non-homologous region is shown in red. Yellow indicates unrelated ORFs. Not to scale.

mice and humans, mostly on the brush border, corresponding to the most highly differentiated enterocytes. SLP-mediated binding of *C. difficile* to enterocytes could allow the targeted delivery of toxins, without dilution or degradation, and may itself contribute to further tissue damage.

One of the paralogues of SlpA, paralogue 3 or Cwp66 (clostridial wall protein of 66 kDa), was also found to possess adhesive properties (Waligora, 2001). Visualisation of immunolabelling using antibodies against the C-terminal portion of Cwp66 protein (variable domain) showed that it was exposed outside the cell following heat shock. Immunolabelling against the N-terminal portion (amidase domain) showed that it was not surface exposed, indicating it may be embedded in the cell wall. Antibodies against Cwp66 partially inhibited adherence of heat-shocked *C. difficile* to cultured Vero cells. However, anti-Cwp66 antibodies did not inhibit the adherence of non-heat-shocked bacteria. These results suggest that several adhesins may play a role in cell adherence of *C. difficile.*

Another SlpA paralogue, ORF 4 or Cwp84 has homology to a cysteine protease from *Canis familiaris* and *Phaseolus vulgaris* (Panagiotopoulou, 2001). Further analysis suggests that this protein belongs to Clan CA of the cysteine proteases, which includes cathepsins and papain. The secreted proteases of *C. difficile* have been found to have generally low activity with a 'suspect' role in pathogenicity (Poilane *et al.*, 1998). As *cwp84* is located downstream of *slpA* in the genome, an attractive hypothesis is that Cwp84 is involved in the cleavage of SlpA or other members of this family of proteins. When expressed as a recombinant protein in *E. coli*, Cwp84 displayed weak, non-specific proteolytic activity on gelatine, skimmed milk and azocoll, but attempts to purify the protein failed (Savariau-Lacomme *et al.*, 2003). Expression of this gene in *C. difficile* could only be detected by Northern blotting and RT-PCR during early exponential phase. The amino acid sequence was found to be highly conserved between the 28 strains studied, particularly around the predicted active site (Savariau-Lacomme *et al.*, 2003), suggesting that Cwp84 has an important physiological function for *C. difficile*.

Recently, a number of studies have looked at the immunological response to surface components from *C. difficile*, with the aim of identifying vaccine candidates. The variablility and immunogenicity of a number of surface components from *C. difficile* have been analysed (Pechine *et al.*, 2005b). FliC, FliD, Cwp66 and Cwp84 were all found to be expressed during the course of an infection. Patient blood serum showed antibodies had been generated to FliC, FliD, Cwp84 and the carboxyl-terminal

(variable) domain of Cwp66. Analysis of the heterogeneity of these proteins in different strains by RFLP showed that FliC, FliD, Cwp84 and the amino-terminal amidase domain of Cwp66 were conserved. This led the authors to conclude that FliC, FliD and Cwp84 should be considered as vaccine candidates. Further studies comparing serum antibody levels to each of these proteins in *C. difficile* patients and asymptomatic but infected controls showed that all of the healthy individuals had higher levels of antibodies to Cwp66 amino-terminal domain, FliC, FliD and the fibronectin binding protein, Fbp68, as opposed to the anti-toxin antibodies which showed no difference between patient and control (Pechine *et al.*, 2005a). The levels of protection afforded by passively administered antibodies against the SLPs in a lethal hamster challenge model have been investigated (O'Brien *et al.*, 2005). Hamsters that received antibodies tended to survive longer than their counterparts, but all animals in the study eventually died of the infection. Analysis of the hamster sera revealed prolonged survival was due to the enhancement of phagocytosis. Drudy *et al.* (2004) investigated the humoral immune response to *C. difficile* SLPs. Although there was no significant difference in the levels of serum antibodies (measured by ELISA) between patients, carriers and controls, this study did find that patients that suffered multiple-relapsing infections have a lower IgM level than single-episode patients.

1.4. Proteomics

The proteome is a term first coined to describe the set of proteins encoded by the genome (Wasinger *et al.*, 1995; Wilkins *et al.*, 1996), but can now be defined more precisely as the overall protein composition of a cell or tissue under exactly defined conditions. The proteome is dynamic, as opposed to the relatively static genome. The growing availability of genome sequence data, coupled with recent advances in technology are fuelling the growth of proteomics, allowing a greater understanding of biological processes at the molecular level.

Traditionally, two dimensional gel electrophoresis (2-DE) is used to separate proteins in a complex mixture to generate a global picture of samples under defined conditions. First developed by O'Farrell in 1975, 2-DE is currently the only technique available for the parallel quantitative expression profiling of large sets of complex protein mixtures e.g. whole cell lysates (O'Farrell, 1975). 2-DE delivers a map of the proteins present in a sample, comparison of which reveals changes in protein expression level, isoforms or post-translational modifications. Depending on the size of the gel and the pH gradient used, 2-DE can resolve more than 5000 proteins simultaneously (~2000 proteins routinely) and detect and quantify <1 ng protein per spot (Gorg *et al.*, 2004).

The first stage in proteomics is sample preparation. The protein sample must be denatured, disaggregated, reduced and solubilised in order to disrupt molecular interactions and to try to ensure each spot represents an individual polypeptide. The three fundamental steps of sample preparation are cell disruption, inactivation or removal of interfering substances and solubilisation of the proteins. Protein samples frequently have a high dynamic range and diversity, so it is often preferable to carry out a pre-fractionation step to reduce the complexity of the sample or to enrich for certain proteins e.g. low copy number or membrane. This can be achieved a number of techniques such as isolation of specific cell types, isolation of cell compartments and/or organelles, selective precipitation or sequential extraction of certain protein classes (Gorg *et al.*, 2004).

Once prepared, protein samples are first separated according to charge by IEF and then by molecular mass with SDS-PAGE. This procedure allows the resolution of several thousand proteins, but suffers from a bias towards abundant and soluble proteins. After 2-DE, the separated proteins have to be visualised. Protein stains must be compatible with mass spectrometry (MS), be highly sensitive and exhibit a broad linear range. Traditionally, Coomassie blue or silver stain are used, but recently a number of fluorescent 'Sypro' stains have been developed, which detect protein with a sensitivity approaching silver, a linear dynamic range greater than Coomassie, and are compatible with MS.

2-DE is routinely coupled with MS to identify proteins. Typically, protein spots on a gel are excised, digested with an enzyme such as trypsin that cleaves at defined protein sequences and the mixture is subjected to MS. One example of MS, a time-of-flight mass spectrometer (MS-TOF) consists of an ion source, a mass analyser that measures the mass-to-charge ratio (*m/z)* of ionised analytes and a detector that registers the number of ions at each *m/z*. As well as the invention of MS-TOF, another key development in proteomics was the discovery of non-destructive methods to convert proteins into volatile ions, recognised by the 2002 Nobel Prize for Chemistry (Fenn, Tanaka and Wüthrich). There are two ways to do this: Matrix Assisted Laser Desorption Ionisation (MALDI) and Electrospray Ionisation (ESI). For MALDI-TOF MS, samples are solidified in an acidified matrix, which absorbs energy in a specific UV range, and dissipates this energy thermally. The rapidly transferred energy generates a vaporised plume of matrix and simultaneously ejects the analytes into the gas phase where they acquire a charge. Strong electrical fields between the MALDI plate and the entrance to the MS tube forces the charged analytes to reach the entrance at different speeds based on their *m/z*. The deduced masses of the peptides can be compared to

those calculated for an *in silico* digest of the predicted proteins from the organism's genome or a protein database, ideally allowing the rapid, unambiguous identification of individual spots (Aebersold & Mann, 2003; Zhu *et al.*, 2003). This method is known as 'peptide mass fingerprinting' (Henzel *et al.*, 1993).

There is a second approach available for mass spectrometric protein identification. In a two-step procedure, MALDI fingerprinting is the first step (Schevchenko *et al.*, 1996b). The second part of the method relies on the fragmentation of individual peptides in the mixture to gain sequence information. Peptides are ionised by electrospray ionisation directly from the liquid phase, from a very fine needle (internal diameter of 1 μm). The peptide ions are sprayed into a 'tandem mass spectrometer' which has the ability to resolve peptides in a mixture, isolate one species at a time and dissociate it into amino- or carboxyl-terminal containing fragments, designated 'b' or 'y' ions respectively. Fragmentation analysis of the mass of these peptide fragments reveals sequence information, and using information derived from several peptides, is much more specific for the identification of a protein. As a result of its simplicity, excellent mass accuracy, high resolution and sensitivity and unlimited mass range capabilities, MALDI-TOF is commonly used to identify proteins by peptide mass fingerprinting (Mann *et al.*, 2001). Tandem mass spectrometry identification of a protein requires more expertise, equipment and is harder to scale-up quantities, but identifications are more clear-cut because in addition to peptide mass, it also provides information about peptide sequence.

The first published proteome-map appeared in 1995 for *Mycoplasma genitalium* (Wasinger *et al.*, 1995). Since then, the field of proteomics has grown rapidly. The proteomes of many pathogenic bacteria have been mapped, including *Haemophilus influenzae*, *Helicobacter pylori*, *S. aureus*, *L. monocytogenes*, and *Mycobacterium tuberculosis* as well as model bacteria such as *E. coli* and *B. subtilis*. Proteomics has been used to examine changes in an organism that can lead to its pathogenicity. For example, the pathogenicity mechanisms of *Candida albicans* have been elucidated (Niimi *et al.*, 1999) and immunoreactive proteins identified using human sera (Pitarch *et al.*, 1999). Proteomics has been used to compare different strains of bacteria e.g. *M. tuberculosis* and *Mycobacterium bovis* BCG strain, to reveal insights into the differences in pathogenicity of these closely related strains (Jungblut *et al.*, 1999). A goal of these applications of proteomics is to identify proteins involved in the disease process, hopefully leading to the development of medical interventions.

Proteomics has been employed as a complement to bioinformatics and functional genomics, as illustrated by Sinchaikul's study of the *Bacillus* species in a systems biology approach (Sinchaikul *et al.*, 2002). This integrated approach is hoped to give a fuller understanding of a microbial system, including aspects such as metabolism, response to stimuli, compartmentalism etc, and their inter relations.

MS-based proteomics has only been made possible by the availability of sequence databases, particularly entire genomes. Three hundred and forty microbial and other genomes have now been published (Genomes On Line Database, (Liolios *et al.*, 2006)), and many more are currently being sequenced and annotated. This represents a key achievement in the development of proteomics as a technique, and an invaluable tool for data interpretation. Proteomics technology has had to be supported by the concurrent development of bioinformatic tools to analyse, retrieve and interpret proteomic data. With the recent completion of the *C. difficile* 630 genome sequence, proteomics offers a valuable chance to gain an insight into the lifestyle of this recalcitrant pathogen.

1.5 Project aims

The evidence reviewed above outlines the substantial and increasing medical importance of infections by *C. difficile*. However, the mechanisms of pathogenicity remain to be fully elucidated, and further studies are required to provide effective methods for the treatment and prevention of infection. In this context, this project is focused on the identification and characterisation of proteins involved in the pathogenicity of *C. difficile*. A global approach utilising the complementary techniques of bioinformatics and proteomics was undertaken for the identification of surface exposed proteins and putative virulence factors. The surface proteome of the bacterium was mapped, and then extended to look for differentially regulated proteins under a range of conditions. Conditions designed to mimic those encountered in an infection and induce stress were studied, namely growth, antibiotic stress and heat shock. Next, the surface proteome was probed with antibodies from immunised rabbits and human CDAD patients to identify immunoreactive proteins. The culmination of these approaches was the identification of nine proteins common to all techniques that are surface exposed and immunogenic, and therefore represent putative virulence factors and potential vaccine candidates. Finally, one of the proteins identified by proteomics as surface exposed and immunogenic, SLP paralogue 4 (or Cwp84) was investigated by biochemical methods to examine whether it fulfilled its predicted function as a cysteine protease.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Bacterial strains and vectors

Clostridium difficile 630 (PCR ribotype 012, toxin A^+/B^+) (Wust & Hardegger, 1983) was provided by Dr. Peter Mullany, Eastman Dental Institute, London. *C. difficile* strain M9 (PCR ribotype 017, toxin A^+/B^-) and its corresponding patient sera were provided by Dr. Denise Drudy, University College, Dublin.

Strains were stored in Robertson's double cooked meat medium, and grown streaked on blood-agar base II plates supplemented with 7.5% defibrinated horse blood at 37°C in an anaerobic container (Don Whitley Scientific) in an atmosphere of 10% CO_2 / 10% H_2 / 80% N_2. Liquid cultures were grown in brain heart infusion (BHI) medium generally for 24 h under the same conditions starting from plates not older than 2 days. Strains were periodically re-isolated by streaking on blood-agar base II plates supplemented with 7.5% defibrinated horse blood and *C. difficile* selective supplement for all strains, and additionally 10 μg/ml tetracycline and erythromycin for strain 630.

Escherichia coli K-12 strains TOP10, JM109 and XL1-Blue and strain BL21(λ*DE3*) were stored frozen at –80 °C in microbank tubes. Working stocks were streaked on LB plates and grown at 37 °C. Liquid cultures were grown in LB medium at 37 °C with shaking at 220 rpm. The genotypes of the *E. coli* strains used in this study can be found in Table 2.1. Plasmids used were based on pET28a (Novagen), pGEX 4T-1 (Amersham Bioscience), pMAL p2E (New England Biolabs) and pCR2.1-TOPO (Invitrogen). Plasmid-encoded antibiotic resistance was selected with 100 μg/ml ampicillin or 50 μg/ml kanamycin, as appropriate. To induce expression of cloned genes under the control of the lactose repressor, isopropyl-β-D-thiogalactoside (IPTG) was added to late log-phase cultures (A_{600} 0.5-0.8) to a final concentration of 1 mM unless otherwise stated. Table 2.2 shows the plasmid vectors used in this study.

Lactococcus lactis strain NZ9000 (Kuipers *et al.*, 1998) was grown in M17 medium supplemented with 0.5% glucose at 30 °C without shaking. Chloramphenicol was added to 5 μg/ml (liquid cultures) or 10 μg/ml (plates) to select for plasmids based on pNZ8020. Frozen stocks were made in the same medium containing 23% glycerol and stored at –80 °C. To induce expression of SlpA from pNZ-1e4H (Table 2.2), nisin (a gift from Microscience) was added to late log cultures (A_{600} = 0.5) to a final concentration of 0.3 ng/ml.

Strain	Genotype	Source
E. coli BL21(λ*DE3*)	F⁻ *ompT*, *hsdS*$_B$($r_B^- m_B^-$), *gal*, *dcm* (λ*DE3*)	Stratagene Ltd.
E. coli K12 TOP10	F⁻ *mcr*A Δ(*mrr- hsd*RMS- *mcr*BC) Φ80*lac* ZΔM15 Δ*lacX74 rec*A1, *deo*R, *ara*D139 Δ(*ara-leu*)7697 *gal*U, *gal*K, *rps*L (Str^R) *end*A1, *nup*G	Invitrogen Ltd.
E. coli K12 XL-1 Blue	F', *pro*AB *lac*I^q ZΔM15 Tn*10 end*A1, *rec*A1, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, *lac*	Stratagene Ltd.
E. coli K12 JM109	F' *traD36 proAB lacI^q ZΔM15 e14-(McrA-) recA1, endA1, gyrA96, thi-1, hsdR17*(rK-mK+) *supE44, relA1, D(lac-proAB)*	Stratagene Ltd.

Table 2.1 *E. coli* strains used in this study and their genotypes.

Plasmid name	Plasmid based on	Antibiotic resistance	Further notes
pCR2.1-TOPO		Ampicillin and kanamycin	Used for TA cloning (Invitrogen)
pET28a		Kanamycin	Expression plasmid, hexa-histidine tag (Novagen)
pGEX 4T-1		Ampicillin	Expression plasmid, GST[1]-tag (Amersham)
pMAL p2E		Ampicillin	Expression plasmid, MBP[2] tag exported to the periplasm (New England Biolabs)
pET-H	pET28a	Kanamycin	Used to express HMW SLP[3] with C-terminal hexa-histidine tag (Calabi, 2001)
pET-L	pET28a	Kanamycin	Used to express LMW SLP[4] with N-terminal hexa-histidine tag (Calabi, 2001)
pEC29(7)	pET28a	Kanamycin	Used to express N-terminal putative protease domain of SLP paralogue 4 with C-terminal hexa-histidine tag
pNZ-1e4H	pNZ8020	Chloramphenicol	Used to express full length SlpA with a C-terminal hexa-histidine tag in *L. lactis* (Calabi, 2001)

Table 2.2 Plasmid vectors used in this study and their properties.

---

[1] GST – glutathione S-transferase

[2] MBP – maltose binding protein

[3] HMW SLP – high molecular weight surface layer protein

[4] LMW SLP – low molecular weight surface layer protein

2.1.2 Chemical reagents

All chemicals used in this study were from Sigma-Aldrich Company Ltd unless otherwise stated.

2.1.3 Media

Blood agar base II and brain-heart infusion (BHI) broth and *C. difficile* selective supplement (CCFA) were obtained from Oxoid. Luria Bertani (LB) broth and agar were from Merck, Robertson's double cooked meat medium was from the Southern Group Laboratory and M17 broth was from VWR International. Horse blood was from TCS-Biosciences.

2.1.4 Buffers

Agarose gel loading buffer

Bromophenol blue	0.25% (w/v)
Xylene cyanol	0.25% (w/v)
Glycerol	25% (v/v)
1 x TAE	74.5% (v/v)

Colloidal Coomassie stain

Ammonium sulphate	50 g
Phosphoric acid	1% (v/v)
Coomassie Blue G250	0.1% (w/v)
Water	to 500 ml

Colloidal Coomassie fixative

Acetic acid	10% (v/v)
Methanol	40% (v/v)
Water	to 500 ml

Coomassie Blue stain

Coomassie Brilliant Blue R250	0.2% (w/v)
Acetic acid	10% (v/v)
Methanol	50% (v/v)
Water	to 500 ml

Coomassie Blue destain

Acetic acid	10% (v/v)
Methanol	50% (v/v)
Water	to 500 ml

Matrix solution

α-cyano-4-hydroxycinnanic acid	10 mg
Acetonitrile	495 µl
Ethanol	495 µl
Trifluoroacetic acid	0.001% (v/v)

Rehydration Solution

Urea	8 M
CHAPS	2% (w/v)
IPG buffer	0.5% (v/v)
Bromophenol Blue	0.05% (w/v)
MilliQ water	to 10 ml

Add 4.2 mg DTT per 1.5 ml immediately prior to use

SDS equilibration buffer

Tris-HCl pH 8.8	50 mM
Urea	6 M
Glycerol	30% (v/v)
SDS	2% (w/v)
MilliQ water	to 50 ml

SDS-PAGE sample buffer (2 x)

Tris-HCl pH 6.8	125 mM
β-mercaptoethanol	10% (v/v)
Glycerol	10% (v/v)
SDS	2% (w/v)
Bromophenol blue	0.1% (w/v)

SDS-PAGE resolving gel (12%)	
Acrylamide/*bis*-acrylamide ratio 30:0.8	4.0 ml
1.5 M Tris-HCl pH 8.8 (4 x)	2.5 ml
SDS	1% (w/v)
Ammonium persulphate	0.1% (w/v)
TEMED	6 µl
Water	to 10 ml

SDS-PAGE running buffer (10 x)	
Tris-HCl pH 8.3	30.8 g
Glycine	144.14 g
SDS	10 g
Water	to 1 l

SDS-PAGE stacking gel	
Acrylamide/*bis*-acrylamide ratio 30:0.8	1.2 ml
0.5 M Tris pH 6.8 (4 x)	1.5 ml
SDS	1% (w/v)
Ammonium persulphate	0.1% (w/v)
TEMED	6 µl
Water	to 6 ml

SOC medium	
Bacto-tryptone	20 g
Yeast extract	5 g
NaCl	0.58 g
KCl	2.5 mM
$MgCl_2$	10 mM
Glucose	20 mM
Water	to 1 l
Adjusted to pH 7.0	

TAE (50 x)	
Tris	242 g
Acetic acid	57.1 ml
Na_2EDTA	37.2 g
Water to	1 l

Transfer Buffer	
CAPS (3[cyclohexyamino]-1-propanesulphonic acid), pH 11	10 mM
Methanol	10% (v/v)
Water	to 1 l

2.2 General procedures for DNA manipulation

2.2.1 Plasmid purification

Plasmids were purified using QIAprep Spin kits (Qiagen), following the protocol recommended by the manufacturer. Briefly, a 10 ml overnight culture was centrifuged at 3,500 *g* for 5 min and the cell pellet resuspended in 250 µl of buffer P1. An equal volume of buffer P2 was added and mixed gently, followed, after 5 min incubation, by addition of 350 µl of buffer N3. The sample was again gently mixed and then centrifuged at 10,000 *g* for 5 min. The supernatant was applied to a QIAprep column and centrifuged at 10,000 *g* for 1 min. Residual buffer was removed by centrifuging a second time at 10,000 *g* for 1 min. DNA was eluted by adding 30-50 µl of 10 mM Tris-HCl, pH 8.5, incubating for 1 min at room temperature and centrifuging at 10,000 *g* for 1 min.

2.2.2 Polymerase Chain Reaction

Polymerase Chain Reactions (PCRs) were carried out in a 30 µl reaction volume, containing 1 x supplied PCR buffer, 2.5 mM $MgCl_2$, 0.1 µM each forward and reverse primers (Thermoelectron, Table 2.3), 0.2 mM dNTPs, 0.2 U *Taq* DNA polymerase (Roche) and 1 µg genomic template DNA. Reactions were carried out using a Thermo Hybaid thermal cycler. Initial denaturation was at 95 °C for 5 min followed by 30 cycles of denaturation at 95 °C for 45 s, annealing at 5 °C below primers T_m for 1 min and extension at 72 °C for 30 s per 500 nucleotides of product length. A final extension at 72 °C for 5 min was included to ensure the completion of all synthesis.

2.2.3 Agarose gel electrophoresis

Agarose gels were prepared by melting 1% (w/v) agarose powder in 1 x TAE in a microwave oven. When sufficiently cool, ethidium bromide was added to 1 µg/ml and the gel cast on a horizontal perspex plate. When fully set, gels were placed in a tank containing 1 x TAE. DNA samples were mixed with 0.25 volumes loading buffer and then loaded into the gel wells. Electrophoresis was performed at approximately 10 V/cm. DNA was visualised on a UV transilluminator at 312 nm, and images recorded

Primer	Sense	Sequence 5' to 3'	RE sites	Notes and purpose
pGEXCwp84EcoR1	F	GGGTAGAATTCAGAAAGTATAAATCA	EcoR1	Clone full length *cwp84* into pGEX 6P-1
pGEXCwp84Xho1	R	TCTCTCGAGTCACTATTTTCCTAAAAG	Xho1	Clone full length *cwp84* into pGEX 6P-1
pMALCwp84EcoR1	F	CCGGAATTCAGAAAGTATAAATCA	EcoR1	Clone *cwp84* into pMAL p2E
pMALCwp84Sal1	R	GCGTCGACTCACTATTTTCCTAAAAG	Sal1	Clone *cwp84* into pMAL p2E
pGEX6P1forseq	F	ACGTATTGAAGCTATCCCAC	-	Sequence pGEX 6P-1 insert
pGEX6P1revseq	R	ATACACTCCGCTATCGCTAC	-	Sequence pGEX 6P-1 insert
pMALP2Xforseq	F	GGTCGTCAGACTGTCGATGAAGCC	-	Sequence pMAL p2-E insert
pMALP2Xrevseq	R	CGCCAGGGTTTTCCCAGTCACGAC	-	Sequence pMAL p2-E insert

Table 2.3 Sequence of the oligonucleotides used. Sense: F = forward, R = reverse. Restriction enzyme (RE) sites are underlined.

using a UVIPro system.

2.2.4 Purification of DNA from agarose gels

DNA fragments were gel-purified using a PCR purification kit (Qiagen), following the manufacturers instructions. Briefly, a gel slice containing the fragment of interest was dissolved in 3 volumes of buffer QG by incubating at 50 °C with occasional vortexing. One volume of isopropanol was added and the solution applied to a QIAquick column. This was processed as the QIAquick column used for plasmid purification (see section 2.2.1). The DNA was eluted in 30-50 µl of 10 mM Tris-HCl pH 8.5.

2.2.5 Restriction endonuclease digestion of DNA

Restriction digests were carried out in the supplied buffer with 20 units of each endonuclease (New England Biolabs). Reactions were at 37 °C for at least 3 h, followed by inactivation at 60 °C for 20 min.

2.2.6 Ligation

Ligations were carried out with 1 µl T4 DNA ligase (Promega) in 10 µl volumes with a range of vector: insert ratios in the supplied ligase reaction buffer. Reactions were left overnight at 14 °C.

2.2.7 DNA sequencing

DNA sequences were determined by the Advanced Biotechnology Centre (Imperial College, London) using sequencing primers detailed in Table 2.3.

2.2.8 Preparation of electrocompetent *E. coli*

A 10ml overnight bacterial culture (*E. coli* JM109, XL1-Blue or BL21(λ*DE3*)) was diluted in 500ml fresh medium and grown until mid-log phase (A_{600} = 0.4). The culture was chilled on ice for 20 min prior to being pelleted by centrifugation at 1,900 *g* for 20 min. The cell pellet was resuspended and centrifuged in decreasing volumes of chilled dH_2O (500 and 250 ml) also at 1,900 *g* for 20 min. The resulting cell pellet was resuspended in 50 ml chilled 10% glycerol, centrifuged as before and resuspended for a final time in 1 ml chilled 10% glycerol. The 1 ml of competent cells was divided into 60 µl aliquots, flash frozen in dry ice/ethanol and stored at –80 °C.

2.2.9 Electroporation of *E. coli*

Competent cells were defrosted on ice. Between 10 and 60 ng of DNA were added and incubated on ice for a further 1 min. The suspension was then transferred to a 0.7 cm electroporation cuvette and electroporated (Gene Pulser II, BioRad) at the following settings: voltage = 1.75 kV; capacitance = 25 μF; resistance = 800 Ω. Immediately after electroporation, 1.5 ml of pre-warmed SOC medium were added to the cuvette and the mixture transferred to a 15 ml Falcon tube, which was incubated at 37 °C with shaking at 220 rpm for 90 min. Dilutions of the culture were made and 100 μl spread onto LB agar with the appropriate antibiotic, and incubated overnight at 37 °C. Transformants were picked and streaked onto LB-agar plates with the appropriate antibiotic.

2.2.10 TOPO cloning reaction and chemical transformation

PCR products were cloned into pCR2.1-TOPO vector. Briefly, 4 μl PCR product were mixed with 1 μl supplied salt solution and 1 μl TOPO vector and incubated for 5 min at room temperature. From this reaction, 2 μl were gently mixed with a vial of "One Shot" chemically competent *E. coli* and incubated on ice for 5 min. The cells were subjected to heat shock at 42 °C for 30 s and then placed on ice. Pre-warmed (37°C) SOC medium was added, and the tubes shaken (220 rpm) at 37 °C for 1 hour. Aliquots were then spread onto pre-warmed selective plates and incubated overnight at 37 °C.

2.3 Expression and purification of recombinant proteins

2.3.1 Growth and induction

Overnight cultures of *E. coli* BL21(λ*DE3*) harbouring plasmids based on the pET vector system were used to inoculate 250 ml LB (with kanamycin at 50 μg/ml). Cultures were incubated at the appropriate temperature (typically 37 °C) in a shaking incubator until the A_{600} was between 0.8-1.0. A 1 ml sample was taken, pelleted and stored at –20 °C. Expression of the cloned genes was chemically induced by alleviation of the repression from the *lac* repressor by using IPTG. Cultures were induced by the addition of IPTG at the appropriate concentration, typically 1 mM. The culture was incubated for a further 3-4 h, and the bacterial cells pelleted by centrifugation at 16,700 *g* for 15 min. Bacterial cell pellets were stored at –20 °C.

2.3.2 Native 6 x His-tag protein purification

E. coli cell pellets, which were kept on ice at all times, were resuspended in lysis buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 10 mM imidazole) with lysozyme added to 1 mg/ml. Cells were sonicated (Sonoplus sonicator, Bandelin) on ice for 2.5 min in pulses of 10 s on, 10 s off. Lysed cultures were centrifuged at 15,000 *g* for 20 min at 4 °C and the supernatant containing the soluble protein removed and filtered through a 0.45 µm filter.

Induced *L. lactis* cells harbouring pNZ plasmids were processed as *E. coli* up to, but not including, the point of sonication. Solid urea was added to a concentration of 6 M, and once completely dissolved, the insoluble material was removed by centrifugation at 10,000 *g* for 5 min. The soluble protein was then processed as for *E. coli*, described below.

1 ml of resuspended nickel-NTA agarose (Qiagen) was added to a 20 ml universal tube and washed with 5 ml lysis buffer. Tubes were centrifuged at 2,700 *g* for 5 min and the buffer removed. The lysate was applied to the washed nickel-NTA agarose and rotated at 4 °C overnight. The resin was then poured into a disposable chromatography column (BioRad). Flow-through from the column was reapplied twice more to the same column. After washing with 10-20 bed volumes of wash buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 20 mM imidazole), bound material was eluted with 4 bed volumes of elution buffer (50 mM NaH_2PO_4 pH 8.0, 300 mM NaCl, 250 mM imidazole).

2.3.3 Denaturing 6 x His-tag protein purification

Bacterial cell pellets were resuspended in denaturing lysis buffer (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, adjusted to pH 8.0) and gently rocked for 1 hour at room temperature. Lysed cultures were centrifuged at 16,700 *g* for 20 min at 4 °C and the supernatants removed, filtered, and applied to the nickel resin in the same way as section 2.3.2. Flow-through was reapplied to the column twice. The column was washed with 4 ml denaturing wash buffer (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, adjusted to pH 6.3). Bound material was eluted with 4 bed volumes of denaturing elution buffer 1 (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, adjusted to pH 5.9), followed by 4 bed volumes of denaturing elution buffer 2 (100 mM NaH_2PO_4, 10 mM Tris-HCl, 8 M urea, adjusted to pH 4.5). Two different elution buffers are used because protein monomers generally elute in buffer 1, while multimers and aggregates generally elute in buffer 2.

2.3.4 Gel filtration chromatography

A number of proteins were further purified using gel filtration chromatography. Two columns were used: an Amersham Biosciences HiLoad 16/60 superdex 200 prep grade column (column bed volume 120 ml); and an Amersham Biosciences HiPrep 16/60 Sephacryl S-100 HR column (column bed volume 120 ml). The columns were attached to an AKTA FPLC (Amersham Biosciences) and equilibrated with 360 ml 20 mM HEPES pH 7.4 running buffer at a flow rate of 0.5 ml/min (6 cm/hour). Protein preparations were applied to the column with a flow rate of 0.5 ml/min (6 cm/hour). Proteins were eluted in 1.5 column volumes of running buffer and 2 ml fractions collected and assessed by SDS-PAGE (section 2.4.1). Fractions were pooled based on quantity and purity of the desired protein present. Purified protein was concentrated using a Centriprep YM-10 filter unit (Millipore), with a 10-kDa cut-off limit, at 2,500 *g* and 4 °C.

2.4 Protein characterisation

2.4.1 Polyacrylamide electrophoresis (PAGE) and Coomassie staining

Protein samples were analysed by electrophoresis using polyacrylamide gels in the presence of SDS using a Mini Protean II system (BioRad). Protein samples were mixed with an equal volume of SDS-PAGE sample buffer and boiled for 3-5 min before loading onto the gel. Gels (1 mm thick) were prepared using an acrylamide:*bis*-acrylamide ratio of 30:0.8 using solutions detailed in section 2.1.4, and electrophoresed in a vertical minigel apparatus (Mini Protean II, Bio-Rad) at constant voltage (120 V), until the bromophenol blue indicator reached the bottom of the gel. Proteins were stained with Coomassie Blue R-250 (at least 1 hour at room temperature), followed by destaining overnight. Gels were scanned using a UMAX scanner.

2.4.2 Bicinchoninic acid (BCA) protein concentration assay

The Pierce BCA assay was used and the manufacturers instructions followed. A series of standards containing known concentrations of bovine serum albumin (BSA) were made up from a stock solution of 2 mg/ml. 20 µl of each standard solution were added to a 96 well micro-titre plate in triplicate. Test proteins were then diluted 1:5, 1:10, 1:15 or 1:20. 20 µl of each sample were added to the plate in duplicate. Pierce working reagent was made up in a 50:1 ratio of reagent A: reagent B and 180 µl of this mixture were added to each sample well and the plate incubated for 30 min covered in clingfilm. Absorbencies were measured on an ELISA plate reader (BioRad) at 595 nm

and the standards used to plot a protein concentration standard curve, allowing the unknown concentrations to be calculated.

2.4.3 Western blots

For Western blotting, samples were separated by PAGE. Following SDS-PAGE, proteins were electroblotted using a mini Trans-Blot cell (Bio-Rad) onto a nitrocellulose membrane (BioRad) in transfer buffer. Transfers were for 1 hour at 70 V and 4 °C. Membranes were blocked with 10% skimmed milk for 1-dimensional SDS-PAGE or 3% BSA for 2-dimensional SDS-PAGE. Triplicate washes after each incubation step were in phosphate buffered saline (PBS) with 0.1% Tween 20 for 5 min. Commercial antibodies (anti-His tag) were used at the manufacturers recommended concentration in 1% skimmed milk powder in PBS-Tween. Other antibodies (anti-SLP (gift from E. Calabi) and anti-*C. difficile* surface proteins (this study)) were used at optimal concentrations determined by ELISA or Western blot. The optimal dilution for the anti-HMW SLP and anti-LMW SLP antibodies was 1:100,000 and the optimal dilution for the anti-*C. difficile* antibodies was 1:200,000. Antibodies were diluted in 1% skimmed milk powder in PBS-Tween, or 0.3% BSA in PBS-Tween. Bound primary antibodies were detected with horse radish peroxidase (HRP)-conjugated rabbit anti-mouse IgG, goat anti-rabbit IgG (Dako), or goat anti-human IgG, IgA and IgM (Sigma) at 1:2,000 in the same buffer as the primary antibody. Each incubation step was for a minimum of 1 hour at room temperature. Blots were developed by enhanced chemiluminescence (ECL) (Amersham Biosciences), following the manufacturer's instructions. Briefly, the supplied developing reagents were mixed at a ratio of 1:1, and the solution applied to the nitrocellulose membrane to completely wet the surface. After 5 min incubation, any excess developing reagents were blotted dry using tissues and the nitrocellulose membrane was wrapped in clingfilm. The blot was visualised using a FUJI LAS-3000 Imager, using the 'wizard' feature to automatically calculate the optimal development time.

2.4.4 Anti - *C. difficile* surface protein antibody production

Rabbit anti-*C. difficile* 630 surface protein antisera were prepared by subcutaneously injecting two New Zealand White rabbits with approx 10^9 whole, heat killed *C. difficile*, prepared according to the method of Poxton (Poxton, 1979). Using this protocol, antibodies should be generated against any surfaced exposed antigens, but not intra-cellular antigens.

C. difficile colonies were scraped from 36-hour blood agar plates, suspended in 1 ml PBS and washed twice, each time with 1 ml PBS. The cell suspension was heated twice at 60 °C for 30 min each time, followed by the addition of 0.5% formaldehyde. The viability of the bacterial suspension was checked by spotting samples onto blood agar plates and incubating overnight. The suspension was mixed 1:1 with Freund's complete adjuvant. Rabbits were inoculated subcutaneously with 10^9 bacteria (corresponding to approximately 500 μl) and were boosted after 2, 4, 6 and 11 weeks using the same dose of antigen but in incomplete Freund's adjuvant. Sample bleeds were collected to assay antibody titres, and for the experiments described, sera collected after the third inoculation were used. The Central Biomedical Service (Imperial College) carried out immunisations and bleeds. Appropriate Home Office licences were obtained prior to all animal experimentation.

2.4.5 Circular dichroism analysis

Circular dichroism (CD) measurements were performed on an AVIV 62DS spectropolarimeter. The instrument was calibrated with sulphonic acid for optical rotation and benzene vapour for wavelength. Data were recorded at 0.2 nm intervals, and at 25 °C, in a temperature-controlled chamber. Samples or baselines consisting of buffer only (20 mM HEPES, pH 7.4) were examined in 0.05 cm pathlength Suprasil cuvettes. Three scans between 193 nm and 300 nm were collected for each sample. Individual scans were smoothed with a Savitsky-Golay filter (Savitsky & Golay, 1964) and the averaged baseline spectrum was subtracted from the averaged sample spectrum. Mean residue molecular weights of 35,000 for the recombinant LMW SLP and 40,300 for the HMW SLP were used in the calculations of mean residue elipticity. Spectra were analysed using the Dichroweb server (Lobley *et al.*, 2002).

2.4.6 Small angle X-ray scattering analysis

Data were collected at the EMBL outstation at DESY (Deutsches Elektronen-Synchrotron), Hamburg, Germany. Scattering of 1.54 Å radiation emitted from the DORIS III storage ring was detected using the EMBL X33 camera at the D1/2 beamline with linear or quadrant multiwire proportional chambers with delayed readout. For low-angle measurements, protein sample concentrations were typically 1-2 mg/ml to avoid interparticle interferences whereas for higher angles the samples were concentrated to 10-15 mg/ml. Approximately 120 μl of sample were injected into the sample cuvette. Buffer scattering was measured after every sample and BSA was used as a control. The scattering of the buffer was subtracted from the sample.

The main parameters that can be extracted from the scattering curve are the molecular weight and the radius of gyration (R_g). It was established by Guinier (1955) that for small values of the scattering vector, s + $4\pi\sin\theta/\lambda$, I(s) can be expressed as follows

$$I(s) = I(0) \exp(-1/3R_g^2 s^2)$$

If log[I(s)] is plotted against s^2 (Guinier plot) and then extrapolated to s = 0 this gives I(0), the forward scattering, which is related to the number of electrons in the particle and hence molecular weight. By comparing I(0) of the sample with I(0) for a standard of known molecular weight and concentration, e.g. BSA, the molecular weight of the sample can be estimated. The gradient of the plot is proportional to the square of the radius of gyration of the scattering density. Two modelling programs were used. The program Dammin (Svergun, 1999) evaluates different models consisting of densely packed spherical 'dummy atoms' that are assigned by configuration vectors to either the protein, hydration shell or solvent. The program uses simulated annealing to find a configuration that best fits the experimental data, whilst also allowing connectivity and compactness constraints to be imposed. The program Gasbor (Svergun *et al.*, 2001) has fewer free parameters. Proteins are represented as an assembly of dummy residues and simulated annealing is used to find a solution that is compatible with the polypeptide chain.

2.5 Cysteine protease activity assays

2.5.1 Zymograms

Proteins were assayed for non-specific protease activity using zymograms (Lantz & Ciborowski, 1994). Samples were suspended in SDS sample buffer minus β-mercaptoethanol and incubated for 30 min at room temperature. Samples were then run on denaturing SDS-PAGE, using a resolving gel containing 0.1 mg/ml gelatine or skimmed milk powder. After electrophoresis at 4 °C and 100 V, the gel was washed twice in 2.5% Triton X-100; twice in 2.5% Triton X-100, 50 mM Tris-HCl pH 7.4; and twice in 50 mM Tris-HCl pH 7.4. Each wash was for 10 min at room temperature. Gels were then incubated at 37 °C for 1.5 h in 50 mM Tris-HCl pH 7.4, 1 mM DTT. Gels were stained by immersion in Coomassie Blue R-250 and lytic bands were visualised against the blue background of the gels after destaining.

2.5.2 Milk degradation

The following method was carried out as described in Savariau-Lacomme *et al.* (2003). Recombinant *E. coli* were streaked onto LB agar containing the appropriate antibiotics to maintain the plasmid, and incubated at 37 °C overnight. A sterile skimmed milk solution was added to LB agar to a final concentration of 1%, and poured into petri dishes to set. Half of the milk-agar plates also contained 1 mM IPTG, and the appropriate antibiotics were also added. The overnight plates were then replica plated onto the milk plates with and without IPTG using velvets. After overnight incubation at 37 °C, the zone of lysis around the colonies was recorded using a digital camera (Canon). The positive control was *Streptococcus pyogenes* H372, and the negative controls were *E. coli* BL21(λ*DE3*) and *E. coli* BL21(λ*DE3*) containing either pET28a or pGEX 4T-1.

2.5.3 Azocoll degradation

The following method was carried out as described in Savariau-Lacomme *et al.* (2003). Azocoll was made up to 10 mg/ml in 50 mM Tris-HCl pH 7.4, 1 mM $CaCl_2$. A 1 ml suspension was vortexed and mixed with an equal volume of an overnight culture of *E. coli*, induced to express recombinant protein. Cultures were incubated under aerobic conditions at 37 °C for 6 h with the azocoll. After centrifugation at 16,000 *g* for 3 min, the A_{520} of the solution was measured and background absorbance was subtracted. A positive control of 0.1 mg of clostripain per 1 ml azocoll was used. Negative controls were LB, *E. coli* BL21(λ*DE3*) and *E. coli* BL21(λ*DE3*) containing either pET28a or pGEX 4T-1.

2.5.4 Gelatinase activity

The following methods were carried out as described in Savariau-Lacomme *et al.* (2003). Photographic film was coated with a sterile suspension of 10% gelatine, 1% activated charcoal. Strips of film were placed in cultures of *E. coli* induced to express recombinant protein and incubated in a static 37 °C incubator overnight. The extent of charcoal release was recorded using a digital camera. A positive control of *S. pyogenes* was used. Negative controls were LB alone, BL21(λ*DE3*), and the vector only strain (pET 28a, pGEX 4T-1).

2.5.5 *N*-Benzoyl-DL-arginine 4-nitroanilide hypochloride (BAPNA) degradation

The following methods were carried out as described in Savariau-Lacomme *et al.* (2003). Samples of purified protein and cell lysates were incubated with 1.2 mM

BAPNA in 50 mM Tris-HCl pH 8.8, 20 mM cysteine hydrochloride for 30 min at 37 °C. The A_{412} was recorded using a Thermo-Spectronic Helios Epsilon spectrophotometer, and the background absorbance subtracted. Positive controls of trypsin, papain and clostripain were used, and negative controls were LB alone, BL21 (λ*DE3*) and the vector only strain (pET28a, pGEX 4T-1), as above.

2.6 Extraction of *C. difficile* SLPs

2.6.1 Glycine extraction of the SLPs

Standard SLP extraction was essentially as described in Dubreuil *et al.* (1988). *C. difficile* from a 24 hour culture in BHI (20-200 ml) were harvested by centrifugation (15 min at 3,500 *g*), washed once in PBS and re-suspended in 0.04 volumes of 0.2 M glycine pH 2.2. After rotating incubation for 30 min at room temperature, the mixture was centrifuged to remove bacteria (5 min at 10,000 *g*) and the supernatant neutralised to pH 7.0 by titration with 2 M Tris.

Alternative extraction methods replaced glycine with one of the following solutions:

- 8 M urea, 50 mM Tris-HCl pH 8.0, 1 mM phenyl-methyl-sulphonide-fluoride (optional: followed by ultra centrifugation at 100,000 *g* for 30 min at 4 °C);
- 10 or 70 mM Na_2EDTA in PBS pH 7.4 (incubated 45 °C for 30 min);
- 1 or 5 M LiCl.

2.6.2 Extraction of the cell wall component of *C. difficile* cells

This extraction procedure was developed from methods described by Schneewind *et al.* (1992) and Jonquieres *et al.* (1999). Cultures (24 hour, 20 ml) of *C. difficile* were pelleted by centrifugation at 31,600 *g* for 3 min. Bacterial pellets were washed once in PBS and once in Tris-sucrose buffer (TS, 10 mM Tris-HCl pH 6.9, 10 mM $MgCl_2$, 0.5 M sucrose). Cell pellets were resuspended in 2 ml digestion buffer (TS with 60 µg/ml mutanolysin, 1 mg/ml lysozyme, 50 µg/ml lysostaphin, 250 µg/ml RNAseA, 2 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF)) and incubated for 2 h at 37 °C with gentle rotating agitation using a Hybaid hybridisation oven. Cells were checked for lysis under the light microscope using x 40 objective lens and phase contrast. Cells were pelleted by centrifugation at 16,700 *g* for 5 min. Supernatant fluids containing cell wall associated proteins were collected, filtered and stored at 4 °C.

2.6.3 Total protein extraction from *C. difficile cells*

The total protein content of *C. difficile* cells was extracted in a manner compatible with 2-dimensional gel electrophoresis (2-DE) (section 2.7) using the Bio-Rad ReadyPrep protein extraction kit, according to the manufacturers protocol. Briefly, the supplied sample buffer was rehydrated, and 10 µl/ml tributylphosphine and 0.2% (w/v) ampholyte were added. Frozen cell pellets from 20 ml culture were resuspended in 1 ml sample buffer and sonicated on ice for 5 min, in 30 s bursts. The lysed cells were warmed to room temperature and then centrifuged at 16,000 *g* for 20 min to pellet cell debris.

2.6.4 Total protein concentration assay

A 2-D Quant Kit (Amersham Biosciences) was used to measure the total protein concentration in samples prepared for 2-DE. The method according to the manufacturers instructions was followed. Briefly, the protein in the sample was precipitated by the addition of 500 µl of precipitant and 500 µl co-precipitant and centrifuged at 10,000 *g* for 5 min. The supernatant was removed and the precipitated protein resuspended in 100 µl of copper reagent and 400 µl of water. Working colour reagent (1ml) was added and the colour left to develop at room temperature for 15 min. The absorbance of each sample was measured at 480 nm using a Thermo-Spectronic Helios Epsilon spectrophotometer using water as a reference.

2.6.5 Assay for cell lysis: isocitrate dehydrogenase activity

The method followed was as described by Harwood *et al.* (1990), and was used to check for lysis of the bacteria under the conditions described in section 2.6.2. The reaction mixture was composed of 50 mM potassium phosphate buffer pH 8.0, 5 mM $MgCl_2$, 0.4 mM NADP and 420 µl cell extract at suitable dilution. The reaction was initiated by the addition of 1 mM isocitrate pH 7.0, and the increase in absorbance at 340 nm at 25 °C was monitored using a Shimadzu UV-1601 spectrophotometer (Shimadzu Corporation, Japan). Ten units of commercially available isocitrate dehydrogenase were used as a positive control.

2.6.6 Assay for cell lysis: glyceraldehyde 3-phosphate dehydrogenase activity

This protocol was based on the method of Ferdinand (1964), and modified by the laboratory of Dr. Peter Moody, University of Leicester. Cell wall extract was added to 0.25 mM NAD, 50 mM Na_2HPO_4, 50 mM CHAPS. The reaction was started by the

addition of 0.2 mM glyceraldehyde 3-phosphate and monitored at 340 nm using a Shimadzu UV-1601 spectrophotometer (Shimadzu Corporation, Japan). For a positive control, 10 units of commercially available glyceraldehyde 3-phosphate dehydrogenase were used in place of the cell wall extract.

2.7 2D SDS-PAGE

2.7.1 First dimension: isoelectric focusing

For a 7 cm immobilised pH gradient (IPG) strip (pH 3-10) (Invitrogen), 1-20 µg of protein sample were mixed with rehydration buffer to a final volume of 125 µl. For 11 cm IPG strips (BioRad), 200 µl rehydration buffer were used to rehydrate the strips and the protein sample loaded using a cup during isoelectric focusing. Rehydration buffer was pipetted along the bottom of a rehydration tray for a length corresponding to the length of the strip being used. The protective cover was removed from the IPG strip and placed face down in the rehydration buffer, ensuring the buffer was evenly distributed and no air bubbles were present. Strips were overlaid with 2 ml of PlusOne cover fluid (Amersham Bioscience) covered, levelled and left overnight. Strips were transferred to a universal strip holder and paper wicks and electrodes carefully placed on each end of the strip. Strip holders were placed on the IPGPhor (Pharmacia) and covered with 4 ml of PlusOne cover fluid. Samples applied to 7 cm IPG strips were run on the following programme: step 1; 500 V for 1 min, step 2; 4000 V for 1.5 h, step 3; 5000 V for 45 min or a minimum of 7.1 kVh. Samples applied to 11 cm IPG strips were run on the following programme: step 1; 500 V for 1 min, step 2; 4000 V for 1.5 h, step 3; 8000 V for 1.5 h or a minimum of 14 kVh.

2.7.2 Second dimension

The IPG strips were equilibrated in 10 ml SDS equilibration buffer with 100 mg DTT added immediately prior to use and incubated for 10 min with constant agitation. This was followed by a second wash in SDS equilibration buffer with 250 mg iodoacetamide, for 10 min with agitation. The strip was then dipped in 1 X MES running buffer (50 mM Tris, 50 mM 3-(N-morpholino)propanesulphonic acid, 1 mM EDTA, 0.01% SDS, pH 7.3) (Invitrogen or BioRad) and inserted into the IPG well of the gel, with the anodic end towards the molecular weight markers well. The IPG strip was overlaid with MES-agarose (1 x MES SDS running buffer, 0.5% agarose (w/v), 0.05% bromophenol blue (w/v)). The 7 cm 4-12% gradient Bis-Tris gels (Invitrogen) were placed in a Novex tank (Invitrogen). The upper chamber of the Novex tank was filled with 1 x MES SDS running buffer plus antioxidant (500 µl per 200 ml) and the lower chamber with 1 x MES

SDS running buffer. The 11 cm 4-12% gradient Bis-Tris gels (BioRad) were placed in Criterion tanks (BioRad) and both chambers filed with 1 x MES buffer. Electrophoresis was carried out at 140 V for 40 min, or until the dye front reached the end of the gel. Gels were stained with Colloidal Coomassie Blue, silver or ruthenium II tris bathophenanthroline disulphonate (see below).

2.7.3 Colloidal Coomassie staining

Gels were placed in fixative solution for at least 1 hour, and then washed in distilled water 3 times for 10 min each on an orbital shaker. The staining solution was prepared by mixing 4 volumes of stock staining solution (section 2.1.4) with 1 volume of methanol. Gels were then submerged in staining solution overnight with gentle agitation. To remove residual stain, gels were washed in distilled water at 50 °C.

2.7.4 Silver staining

A modified method of silver staining, compatible with mass spectrometry, was used, based on the method of Shevchenko *et al.* (1996). Proteins were fixed for 20 min in 50% methanol, 5% acetic acid, followed by 1 wash in 50% methanol for 10 min and an overnight wash in MilliQ water. The MilliQ water was discarded and sufficient sensitiser (0.02% sodium thiosulphate) to submerge the gel was applied for 1 min, and then washed twice for 1 min each in MilliQ water. Gels were incubated in cold silver reagent (0.1% silver nitrate) for 20 min at 4 °C, and then washed twice more in MilliQ water for 1 min each time. Developer (0.04% formalin in 2% sodium carbonate) was applied for between 5 and 10 min, with frequent changes as the developer solution discoloured. Staining was stopped using 5% acetic acid for 5 min, and gels were washed again, and stored in MilliQ water.

2.7.5 Ruthenium II tris bathophenanthroline disulphonate (RuBPS) staining

Proteins were stained according to a modified version of Rabilloud and coworkers (Rabilloud *et al.*, 2001), based on an alternative to the commercial Sypro Ruby dye (Molecular Probes).

A 20 mM stock solution of RuBPS was prepared according to a modified version of (Rabilloud *et al.*, 2001). Approximately 50 mg of potassium pentachloroaquo ruthenate ($K_2Cl_5Ru.H_2O$) (Alfa Aesar) was suspended as a slurry in water at 10 mg/ml. 1 ml of slurry was transferred to a conical heat-resistant glass centrifuge tube siliconised with RepelSilane (Amersham Bioscience) and capped. The tube was inserted into a

heating block preheated to 150 °C and filled with mineral oil to create a seal. The slurry was boiled, and therefore refluxed for 5 min in a fume hood. When cool, the slurry was transferred to a 7 ml plastic sterilin bottle containing 45 mg anhydrous bathophenanthroline disulphonate disodium salt ($C_{24}H_{14}N_2O_6S_2Na_2$) (Alfa Aesar), and vortexed. The suspension was transferred back to the glass centrifuge tube and heating continued at 150 °C for 20 min. A colour change to deep green/brown was observed. 0.25 ml of 500 mM sodium ascorbate (Alfa Aesar) were added and heating continued at 150 °C for 20 min, causing a further colour change to deep ruby red. The solution was transferred to an eppendorf tube and stored in the dark at 4 °C.

Electrophoresed protein gels were fixed overnight in 40% ethanol, 10% acetic acid. The gels were washed 4 times for 30 min each time in MilliQ water, and then proteins were stained in 1 μM RuBPS for 6 h with rotating agitation. Destaining was carried out overnight in the fixative solution. Protein was visualised on a FUJI FLA-5000 scanner using the blue laser (λ = 473 nm) and the 510 nm long pass filter (Y510), at 400 V sensitivity and 50 μm resolution.

2.7.6 Computational analysis of results using Phoretix

Phoretix 2D Evolution (Non-Linear Dynamics) is a computer software package specifically designed for the analysis of 2-D PAGE gels. Protein patterns and spot intensities can be accurately compared, adjusting for gel-to-gel variations due to staining and protein loading.

The software was used in manual editing mode. First, an experiment was created and the images of the gels to be compared were loaded into the experiment. Typically, 3 gels of each condition tested (sub-gels) were used to create an "average gel" of that condition. Although three gels were prepared for all the conditions compared, occasionally one of the gels would not run properly, and if the pattern of one gel was obviously not consistent with the others in its group, then it was excluded from the analysis. The average gel with the most spots was used to create a "reference gel", to which all spots were matched against, and allowed comparison between spots, gels and conditions.

First, protein spots were manually detected on all of the sub-gels using the editing tools. Then gels were viewed in matching mode, where the gels were warped to the average gel for each condition. This makes the manual spot matching much easier, and far more accurate to perform. Protein spots were manually matched, and checked

using the montage window to check that the spots in each sub-gel were matched to their counterparts on the average gel. Once gels were edited and matched the "Add to Reference gel" icon was used to add all unmatched spots to the average gel, ensuring that the spots present on all gels were ultimately represented on the reference gel. The background was subtracted from each sub-gel using the mode of non-spot and the default setting of 45. Spot volumes were normalised using the total spot volume multiplied by 100. This procedure of gel warping, spot matching, and adding to the reference was repeated to compare each average gel to the reference gel. Finally, all the spot numbers were synchronised.

Following analysis, the data generated for each spot on each sub-gel was viewed in the Comparisons window. The reference spot numbers are listed vertically and each row contains normalised spot volumes for that spot as it occurs in each of the sub-gels. This spreadsheet of data was used for subsequent statistical analyses comparing expression levels between sample sets. Data were statistically analysed to look for consistency between matched spots, and differences between conditions, using the analysis of variance (ANOVA) function within the Phoretix software. This statistic is similar to a multi-directional T-test. Spots with a 95% confidence rating were retained for further investigation: all other spots were filtered out of the analysis. The remaining spots were manually checked again to make sure they were correctly matched, and where possible duplicate spots from different conditions chosen for mass spectrometry analysis.

2.8 Mass spectrometry

2.8.1 In-gel trypsin digest

Spots were excised from gels using a sterile scalpel. The resulting gel plugs were destained in 10 µl 50 mM ammonium bicarbonate, 50% acetonitrile (ACN) at room temperature for 20 min or until the blue colour had disappeared. After removing destain, the plugs were vacuum dried for 5 min using a Savant Speed Vac, which helps the subsequent solutions to penetrate the gel plug. Proteins in the gel plug were reduced by the addition of 15 µl 10 mM DTT in 100 mM ammonium bicarbonate for 30 min at room temperature. Proteins were then alkylated by adding 15 µl 55 mM iodoacetamide in 100 mM ammonium bicarbonate for 20 min at room temperature and samples were washed with 20 µl 100 mM ammonium bicarbonate for 5 min at room temperature. The gel plugs were dehydrated by the addition of 30 µl ACN and vacuum dried for 5 min. Proteins in the gel plug were digested with 60 ng trypsin in 50 mM ammonium bicarbonate overnight at 37 °C. The supernatants were removed and

retained, and 15 µl 0.1% trifluoroacetic acid (TFA) added to the gel plug to stop the digestion (incubation at 37 °C for 30 min). The TFA was removed and pooled with the previous fraction. Peptides remaining in the gel plug were extracted by the addition of 20 µl of 6:4 v/v mix of ACN: 0.1% TFA, and incubated at room temperature for 30 min. The supernatants were removed and added to the previous fractions. The combined supernatants from these stages contain the majority of peptides from the gel plug, and these were concentrated to a volume of approximately 10 µl using a speed vacuum. The peptides were stored at 4 °C until spotting onto the MALDI-TOF target plate just before the start of the MALDI session.

2.8.2 MALDI-TOF analysis of trypsin digested proteins

A tripeptide calibration mix was prepared by mixing 10 pmol/µl each of angiotensin I, rennin substrate and adrenocorticotropic hormone (18-39 fragment) in 0.1% TFA in a ratio of 3:1:3. Sample or calibration mixture (2 µl) were mixed with an equal volume of matrix solution and spotted onto a target plate in duplicate using 1 µl per well. Lockmass (1 µl of 5 pmol/µl in matrix solution) was spotted into the centre of each cluster of wells. The plate was allowed to dry at room temperature and then loaded into the MALDI-TOF machine (M@LDI-TOF-R, Micromass). Sample lists were created and data processed using MassLynx software and peptides searched against the non-redundant protein database and *C. difficile* genome database using ProteinLynx software.

2.8.3 Tandem electrospray mass spectrometry

Due to the technical expertise required and availability of equipment, tandem electrospray mass spectrometry (MS/MS) was carried out by Dr Robin Wait (Kennedy Institute of Rheumatology, Imperial College, London) and Dr Jerry Thomas (Proteomics Facility, University of York). The principal stages of the process are outlined here.

Tandem electrospray mass spectra were recorded at the Kennedy Institute using a Q-TOF hybrid quadrupole/orthogonal acceleration TOF spectrometer (Micromass), interfaced to a Micromass CapLC capillary chromatograph. Samples were dissolved in 0.1% aqueous formic acid (with the stream select valve diverting the column effluent to waste). The flow rate was then reduced to 1 µl/min, the stream select valve was switched to the data acquisition position and the peptides were eluted into the mass spectrometer with an ACN/0.1% formic acid gradient (5-70% ACN over 20 min). The capillary voltage was set to 3500 V and data dependent MS/MS acquisitions were performed on precursors with charge state of 2, 3 or 4 over a survey mass range of

540-1000. Trypsin autolysis products and keratin derived precursor ions were automatically excluded. The collision voltage was varied between 18 and 45 V depending on the charge and mass of the precursor. Product ion spectra were charge-state de-encrypted and deisotoped with a maximum entropy algorithm (MaxEnt 3, Micromass).

MALDI-MS and MS/MS were carried out at the University of York using an Applied Biosystems 4700 Proteomics Analyzer. Gel plugs in PCR tubes were posted at room temperature in protective packaging.

2.9 Bioinformatic analyses

The *C. difficile* genome sequence was available at http://www.sanger.ac.uk/projects/C_difficile/. The chromosome is 4.29 Mb, including a 7 kb plasmid, with a total of 3775 predicted coding sequences.

Artemis (release 6) was used to visualise ORFs within the *C. difficile* genome. Artemis is a Java-based DNA sequence visualisation and annotation tool that can be used for the analysis of Bacteria, Archaea and lower eukaryotic genomes (Rutherford *et al.*, 2000). Genomes are viewed in context of DNA sequence and its six frame translations. Sequences are viewed in two windows that can be used to view the same sequence at different zoom levels simultaneously. The results of external programmes (e.g., BLAST) can be imported into Artemis and viewed as features overlaid onto the sequence. Artemis is available under the GNU General Public License from http://www.sanger.ac.uk/Software/Artemis (Rutherford *et al.*, 2000).

Nucleotide sequence similarities to entries in the SwissProt database were determined using BLAST-N and protein sequence similarities determined using BLAST-P, both available at http://www.ncbi.nlm.nih.gov/BLAST. Alignments with no gaps are referred to as high scoring pairs (HSPs). The number of random HSP scores equal or greater than S is described by the Poisson distribution. This is the P value associated with the score S. Highly significant scores have P values close to zero. For gapped alignments, the significance of a given alignment with score S is represented by the E (expected) value, the expected number of chance alignments with a score of S or better. This can be evaluated by looking at alignment scores generated by mock databases of random sequences of comparable length and composition. The E value decreases exponentially as the Score (S) that is assigned to a match between two

sequences increases. The E value reflects the size of the database and the scoring system in use (Altschul *et al.*, 1990; Altschul and Gish, 1996).

Pairwise alignment of protein sequences was conducted using the general purpose multiple sequence alignment programme ClustalW at http://www.ebi.ac.uk/clustalw/. ClustalW calculates the best match for the selected sequences and lines them up so that the identities, similarities and differences can be seen. The basic multiple alignment algorithm consists of three main stages: 1) all pairs of sequences are aligned separately in order to calculate a distance matrix giving the divergence of each pair of sequences; 2) a guide tree is calculated from the distance matrix; 3) the sequences are progressively aligned according to the branching order in the guide tree (Higgins, 1994).

The PROSITE database of protein families and domains available at http://ca.expasy.org/prosite/ was used to determine which known protein family (if any) a new sequence belongs. Protein sequence families have some regions that are better conserved than others during evolution. These regions are generally important for the function of a protein and for the maintenance of its 3D structure. By analysing the constant and variable properties of such groups of similar sequences, it is possible to derive a signature for a protein family or domain, which distinguishes its members from all other unrelated proteins. PROSITE is particularly useful when the structure and function of an unknown protein, which is too distantly related to any protein of known structure to detect its affinity by overall sequence alignment, may be identified by its possession of a particular cluster of residue types classified as a motif (Hulo *et al.*, 2004).

Protein sequences were searched against the Pfam database (http://www.sanger.ac.uk/Software/Pfam/) to identify functional domains within the sequence. Pfam contains over 6190 protein families and 75% of proteins contain at least one match to Pfam, making it a valuable tool for assessing the putative role of a protein (Mulder *et al.*, 2003).

Three-dimensional structures of protein sequences were predicted using three-dimensional position-specific scoring matrix, 3D-PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm/index2.html). This method recognises remote protein sequence homologues, using structural alignments of homologous proteins to obtain a structural equivalence of residues (Kelley *et al.*, 2000). Alternatively, a local dynamic programme alignment (Smith & Waterman, 1981) was used to align residue

equivalencies between a given protein and the closest available model structure (identified by BLAST analysis). This maps the respective equivalencies onto the pdb structure and can then be viewed in the programme Rasmol.

SignalP, available at http://www.cbs.dtu.dk/services/SignalP/, uses a combined neural network approach to the recognition of signal peptides and their cleavage sites, using one network to recognise the cleavage site and another network to distinguish between signal peptides and non-signal peptides (Bendtsen *et al.*, 2004).

PSORT, available at http://psort.nibb.ac.jp/, predicts protein localisation sites in cells when provided with the amino acid sequence of the protein and its origin i.e. Gram-positive bacteria, yeast etc. PSORT first predicts the presence of a signal peptide, then transmembrane regions and finally lipoproteins, and combines this data to predict the cellular location of any given protein (Nakai & Horton, 1999).

Protein molecular masses and pI values were calculated using the compute pI/MW tool at http://us.expasy.org/tools/pi_tool.html. Protein molecular mass is calculated by the addition of average isotopic masses of amino acids in the protein and the average isotopic mass of one water molecule. Protein pI is calculated using pK values of the amino acids (Boeckmann *et al.*, 2003).

Chapter 3

Bioinformatic analysis of putative virulence factors in the *Clostridium difficile* genome.

3.1 Introduction

3.1.1 Bioinformatics as a tool for the analysis of genome data

With the advent of genome sequencing, a new resource became available for the study and understanding of biological systems. Over 340 genomes have now been published (GOLD (Liolios *et al.*, 2006)), and many more sequencing projects are underway. However, this raft of information needs to be handled, viewed and interpreted, and this has necessitated the development of computer programmes and computer-based analysis tools. A large number of programmes are now available in a web-based format and can provide information about a gene product's likely processing, location and function.

When analysing the genome of a pathogen, a natural feature to investigate are genes encoding virulence factors. Predicted coding sequences (CDS) can be compared to protein sequences in databases using programmes such as Basic Local Alignment Sequence Tool (BLAST), which identifies matches to known genes (Altschul *et al.*, 1990). BLAST can also indicate the presence of motifs from families of genes with the same function, thereby conferring a function to the query CDS. However, BLAST is unable to identify the function of novel genes. In most genomes sequenced to date, approximately 20% of CDS have no significant similarity to anything else in the database, and a further 15-20% have homology with genes of unknown function. An alternative way to identify new virulence factors is to search for genes that are co-regulated with known virulence factors, or to look for particular motifs that are indicative of virulence factors. For example, tandem repeats at the 5' end of the gene are common in virulence factors as recombination within the repeats can turn expression on and off with high frequency (phase variation) (Hood *et al.*, 1996). This can contribute to an attenuated or virulent phenotype of the bacterium. Secreted or surface exposed proteins can sometimes be virulence factors that are important for adhesion and colonisation of a host. These proteins can be identified using programmes that recognise such motifs, e.g. transmembrane helices (PHD programme) or secretory leader sequences (SignalP).

A recent application of the *in silico* analysis of genomes is in vaccine discovery. Pizza *et al.* (2000) carried out the first published example of such a study on *Neisseria meningitidis* serogroup B, a leading cause of meningococcal meningitis and sepsis. Previous efforts to design a vaccine against this microorganism were hampered by the similarity of the capsular polysaccharide to human tissues and sequence variation of surface exposed proteins (Finne *et al.*, 1987; Hayrinen *et al.*, 1995; Poolman *et al.*,

1995). In the Pizza study, the whole genome was analysed for novel surface exposed or exported proteins, resulting in the selection of 570 ORFs. High throughput cloning and expression produced 350 candidate antigens, which were used to immunise mice. The antisera produced were tested for bactericidal response, antigen surface exposure and antigen conservation. The result of this work was the identification of seven vaccine candidates that are now being investigated further (Jacobsson *et al.*, 2006). This was the first study to prove the potential of the bioinformatic approach and has since been repeated on other pathogens including *Helicobacter pylori* (L. Lissolo, unpublished), *Streptococcus pneumoniae* (Wizemann *et al.*, 2001), *Chlamydia pneumoniae* (Montigiani *et al.*, 2002), *Staphylococcus aureus* (Etz *et al.*, 2002; Vytvytska *et al.*, 2002), *Porphyromas gingivalis* (Ross *et al.*, 2001), *Edwardsiella tarda* (Srinivasan *et al.*, 2003) and *Mycobacterium tuberculosis* (Betts, 2002). This approach has been termed 'reverse vaccinology' (Rappuoli, 2000) and has identified more potential vaccine candidates in the last eight years than have been identified over the past 40 years by traditional means (Scarselli *et al.*, 2005). The antigens identified through a genomic approach are usually different to those found using a conventional approach. For example, many of the newly discovered antigens include surface exposed proteins with a globular structure and without membrane spanning domains, and often have a low abundance on the cell surface. These vaccine candidates are sometimes able to induce wide strain coverage; indeed, the MenB vaccine candidates are similar to proteins found in other *N. meningitidus* serogroups and even *Neisseria gonorrhoeae*, and so are likely to elicit cross-protection (Pizza *et al.*, 2000).

The prediction of vaccine candidates by genome analysis can be strengthened by a combination of other approaches focusing on protein expression, surface localisation and immunoreactivity. A study of this type has been carried out on *Bacillus anthracis*, the causative agent of the potentially fatal disease anthrax, and possible bioweapon (Ariel *et al.*, 2002; Ariel *et al.*, 2003; Chitlaru *et al.*, 2004). As well as using bioinformatics to search for virulence factors, the surface proteome of *B. anthracis* was mapped using two-dimensional gel electrophoresis and mass spectrometry, and then probed with sera from *B. anthracis* immunised animals. The benefit of this approach is that surface localisation and expression of the proteins is proven, complementing the bioinformatic analysis. The antigens identified serve as candidates for the development of an improved anthrax vaccine (Chitlaru *et al.*, 2004).

The increasing number of sequenced genomes available has allowed for the development of comparative genomics. This can be used, for example, to compare genetic variability between pathogens and closely related non-pathogenic bacteria.

Comparative genomics can be used to isolate virulence genes, provide a rational basis for the attenuation of vaccine candidates, or monitor gene content variability within a species. Recently the sequenced pathogenic clostridial genomes have been compared, with particular focus on extra-chromosomal DNA (Bruggemann, 2005). In comparison to enteric bacteria, most clostridia have relatively static genomes. However, several toxin loci and virulence regions show signatures of genetic mobilisation. Some toxins have been located to extra-chromosomal elements such as phage and plasmids; indeed, plasmids have been found in all sequenced clostridia to date. Two examples of plasmid-located virulence factors are the gene encoding tetanus toxin and ABC exporters conferring drug resistance in *C. botulinum*. Bruggemann (2005) suggests that based on comparative genomics, mobilisable genetic elements may be a source for horizontally acquired genes which confer fitness in the clostridia. Comparative genomics can be used to investigate inter-strain variability. For example, comparative genome analysis of *Helicobacter pylori* strains revealed a limited number of strain specific genes and a similar genome organisation. This was unexpected in a pathogen previously believed to be subject to extensive genomic and allelic diversity (Alm & Trust, 1999). In contrast, the comparison of two *M. tuberculosis* strains revealed more extensive variability than had been expected (Fleischmann *et al.*, 2002). These findings have significant implications for pathogenesis and immunity against microorganisms. However, as it is unlikely that a single strain can be paradigmatic for the biology and pathogenic potential of a species, it is imperative to validate results based on bioinformatics. In the post-genomic era, this is possible through functional genomics, including DNA microarrays and proteomics, which in turn utilise different aspects of bioinformatics themselves.

3.1.2 The *C. difficile* genome

The sequencing and annotation of the *C. difficile* genome has recently been completed (Sebaihia *et al.*, 2006). The genome sequence was that of *C. difficile* strain 630, a virulent and multi-drug resistant strain isolated from an outbreak in a Zürich hospital in 1982 (Wust *et al.*, 1982). The genome contains a circular chromosome of 4,290,252 bp and a plasmid of 7,881 bp. The chromosome is predicted to contain 3,776 CDS, and the plasmid 11 CDS. Comparison of the *C. difficile* genome with the other sequenced clostridia illustrates the diversity of this genus. 50% of *C. difficile* genes are unique to the species whereas only 15% are common to other clostridial genomes. Unlike the other sequenced clostridia, the genome of *C. difficile* is highly mobile, with 11% of the genome made up of mobile genetic elements (Bruggemann, 2005). The *C. difficile* mobile genetic elements are comprised of seven conjugative transposons

(CTns), a likely mobilisable transposon (Tn*5398*) (Farrow *et al.*, 2001), two prophages, a skin element (Haraldsen & Sonenshein, 2003), IStrons (Burrus *et al.*, 2002) and several other unknown mobile elements. Each conjugative transposon contains different accessory genes that may be involved in antimicrobial resistance and transport, e.g. CTn*3* mediates tetracycline resistance (Mullany *et al.*, 1990). Analysis of the *C. difficile* genome has also revealed some of the metabolic traits that the bacterium utilises to aid adaptation and survival in the gastrointestinal tract. For example, a large number of CDSs are predicted to be involved in carbohydrate transport and metabolism, suggesting that this is the preferred nutrient source. *C. difficile* appears to have other metabolic adaptations that aid survival in the gut, such as genes predicted to be involved in bile acid conversion and tolerance. Another important event in the pathogenic lifestyle of *C. difficile* is sporulation and germination. Like the other sequenced clostridia, *C. difficile* appears to differ in its sporulation process compared to the bacilli. Several genes involved in the phosphorelay process that trigger sporulation are absent. In addition the mobile skin element must be excised from the sporulation sigma factor gene it interrupts for efficient sporulation to occur (Haraldsen & Sonenshein, 2003). However, the mechanism by which sporulation is triggered in *C. difficile* remains unknown. For example, the GerA family of receptors that triggers sporulation in other *Bacillus* and *Clostridium* species is absent in *C. difficile* (Feavers *et al.*, 1990). Elucidation and analysis of the genome sequence has given new insight into the *C. difficile* lifestyle, revealing a more mobile and adaptable content than had been expected. The genome sequence also provides a starting point for further analyses of the *C. difficile* lifestyle.

In this study, the *C. difficile* genome has been examined in detail, through a series of bioinformatic searches and manual curation. This approach has identified CDSs with a potential role in the virulence of the bacterium. In addition those CDSs with a surface localisation may be considered vaccine candidates. This bioinformatic analysis is complementary to other approaches investigating bacterial pathogenesis, such as proteomics.

3.2 Results: Bioinformatic analysis of the *C. difficile* genome

The *C. difficile* genome contains 3,776 CDSs. The strategy by which these CDSs were searched for potential virulence factors is shown in Figure 3.1. A search was made for a series of specified motifs, based on those likely to be present in bacterial protein virulence factors. In this way, it was hoped to positively search for proteins containing

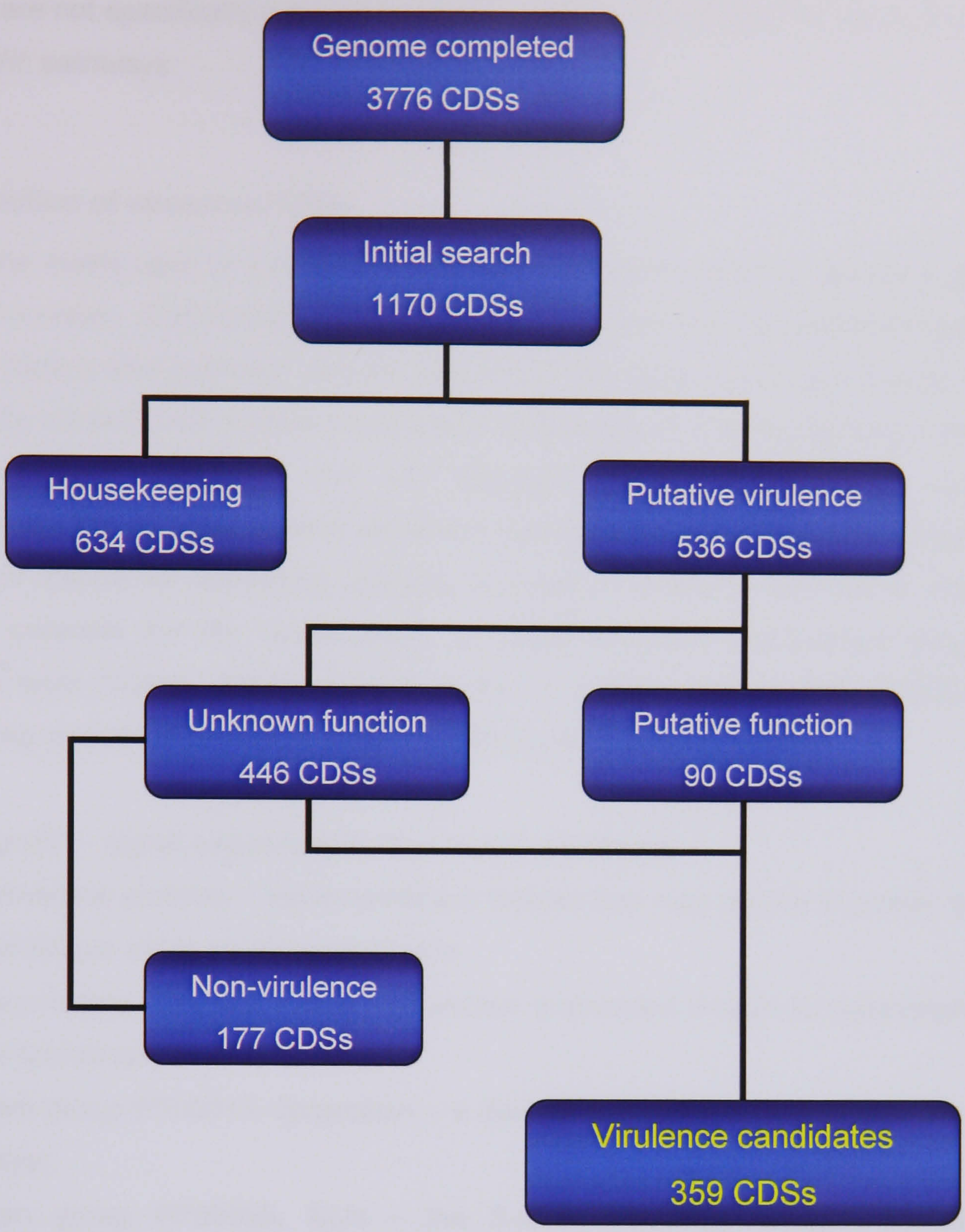


Figure 3.1 Strategy for selection of virulence factors in the *C. difficile* genome. Flowchart of the computational analysis and the filtering steps. The initial search selected ORFs with features of virulence factors, and following filtering steps removed unlikely virulence ORFs from this list. The resulting number of ORF products at each step is noted in each shaded box.

these motifs. Then, a series of filtering criteria were applied to discard proteins whose functions are not specifically required for virulence, e.g. "housekeeping" genes involved in metabolic pathways.

3.2.1 Selection of virulence CDSs

Many of the motifs used in the bioinformatic searches either direct or anchor a protein to the cell surface. Such proteins are of particular interest when searching for putative virulence factors and potential vaccine candidates, because during the course of an infection the bacterial cell surface components and secreted proteins form the interface for bacterial-host interaction and are exposed to the host immune system. Consequently, the bacterial surface proteome and secretome represent the subcellular fractions of choice for identifying proteins involved in virulence and which possess antigenic potential for the development of novel vaccines, antibacterial drugs or diagnostic tools (Grandi, 2001). Therefore, the *C. difficile* genome was searched for the following classes of protein or those containing the following motifs:

- SignalP – signal sequences for the export of proteins
- Membrane proteins – transmembrane helices that may anchor a protein to the lipid bilayer of the plasma membrane
- Lipoproteins – certain motifs that anchor a secreted protein to components of the lipid bilayer
- Pfam group PF00013, lipoprotein – a domain mediating attachment to the lipid bilayer
- Pfam group PF00395, SLH – the S-layer homology domain, sometimes responsible for attachment of S-layer and other proteins to the peptidoglycan
- Pfam group PF01473, cell wall binding 1 – domain for anchoring proteins to the cell wall
- Pfam group PF01476, LysM – this domain is found in enzymes involved in cell wall degradation or may have a peptidoglycan binding function
- Pfam group PF01510, amidase 2 – a family of enzymes that can bind to and sculpt the cell wall
- Pfam group PF01520, amidase 3 – a family of enzymes that can bind to and sculpt the cell wall
- Pfam group PF04122, cell wall binding 2 – anchors to the cell wall
- Pfam group PF04203, sortase – a transpeptidase that attaches proteins to the cell wall by the threonine of an LPXTG motif

- Sortase substrates – proteins with the LPXTG motif (or similar) predicted to be attached to the cell wall by sortase enzymes
- Twin-arginine pathway substrates – proteins with signals for secretion by the twin-arginine pathway
- Flagella – components of the flagella have been found to be adhesins and therefore virulence factors in *C. difficile* (Tasteyre, 2001)
- SlpA – the high molecular weight surface layer protein of *C. difficile* (HMW SLP) and a paralogue of this protein, Cwp66, have been shown to be adhesins and therefore virulence factors (Waligora, 2001; Calabi *et al.*, 2002). Other paralogues to these proteins may have a similar role
- Nucleotide repeat sequences – a common feature in virulence factors

The results of these searches were compiled into a single data set containing 1170 CDSs. The selection of potential virulence CDSs was designed to be as inclusive as possible to maximise the chance of identifying virulence factors. However, this large group of CDSs needs to be inspected for false positive virulence factors.

3.2.2 Filtering criteria: housekeeping functions

Using the programme Artemis to view the genome, all CDSs highlighted by the search above were manually inspected, and those shown by BLAST analysis to have homology to housekeeping genes were removed. A housekeeping gene is judged to be involved in cellular functions such as energy metabolism, information transfer, central or intermediary metabolism or regulators. In addition, CDSs containing disruptions such as frame-shift mutations were removed from the list. There were a number of exceptions to the exclusions listed above, and these are discussed further below.

3.2.2.1 ABC transporters/permeases

ABC transporters facilitate the unidirectional transport of small molecules across the cell membrane by the hydrolysis of ATP. In total, 152 components of ABC transporters were found in the *C. difficile* genome, including permeases, substrate-binding proteins, solute-binding proteins, ATP-binding proteins, carrier proteins, membrane and lipoproteins. On the whole, components of ABC transporters were not considered to be directly involved in virulence, as they are primarily involved in the transport of nutrients. However, a small number of ABC transporters may be of benefit to *C. difficile* during the course of an infection, and genes encoding these components were

retained in the list of putative virulence factors. This group includes ABC transporters involved in antibiotic resistance, lantibiotic transport, iron scavenging and oligopeptide transport (Table 3.1). Antibiotic resistance is a feature of *C. difficile* infection, and so the mechanisms that facilitate this are key to virulence. Lantibiotics are lanthionine-containing peptide antibiotics produced by Gram-positive bacteria that cause cell death of other Gram-positive bacteria. Resistance to lantibiotics provides *C. difficile* with a competitive advantage over other gut microflora, therefore aiding colonisation and virulence. Iron availability is highly restricted in a host, and limits the growth of colonising bacteria; hence an ability to scavenge iron is considered to be a virulence factor (Brown & Holden, 2002). Oligopeptide transport allows the utilisation of peptides as a nutrient source, which may be important for growth and pathogenesis.

As well as ABC-transporter components, nine putative transporter proteins and four components of the sec-dependent secretion system were found. These proteins were judged not to be directly involved in the pathogenesis of *C. difficile* (rather, it is sometimes secretion system substrates that are the virulence factors) and were therefore removed from the list of putative virulence factors.

3.2.2.2 Two component regulatory systems and signalling proteins

Two component regulatory systems are thought to have evolved from a primitive signalling device encoded by a single pair of genes. These genes are thought to have been duplicated many times. These duplications have enabled the gene products to diverge and adapt to serve analogous signal transduction functions for a range of processes (Neidhardt *et al.*, 1990). This conserved signalling pathway, regulated by phosphorylation, controls diverse pathways such as chemotaxis, gene expression, virulence and developmental pathways. Sixty-six proteins belonging to this group were found in the *C. difficile* genome, comprising sensor histidine kinases, response regulators, signalling and sensory box proteins. On the whole, members of this group were not considered virulence determinants, as they do not perform responses themselves, but rather signal other proteins to respond. However, three proteins are potentially important for the pathogenicity of *C. difficile*, all of which signal responses to antibiotics and lantibiotics (Table 3.2).

In addition to the 66 proteins above, two serine/threonine protein kinases were found in the *C. difficile* genome. These were proteins were not regarded as possible virulence factors.

Table 3.1 Virulence related ABC transporter components. Genes are listed in sequential CDS number. Yellow boxes indicate similarity to ABC transporters involved in antibiotic resistance, pink indicates lantibiotic resistance components, green indicates oligopeptide transporters and blue indicates iron transport. The species to which the CDS has highest homology are listed.

CDS	Function	Homologous to protein from:
CD0161	Putative antibiotic ABC transporter, ATP-binding protein	*Bacillus subtilis / Streptococcus pyogenes / Lactococcus lactis*
CD0363	Putative antibiotic ABC transporter, membrane protein	*Clostridium perfringens / B. subtilis*
CD0364	Putative antibiotic ABC transporter, ATP-binding protein	*B. subtilis / Staphylococcus epidermidis*
CD0365	Putative antibiotic ABC transporter, permease protein	*Enterococcus faecalis / Bacillus licheniformis*
CD0479	Putative lantibiotic ABC transporter, permease protein	*B. subtilis / Clostridium tetani*
CD0480	Putative lantibiotic ABC transporter, permease protein	*B. subtilis / Streptococcus mutans*
CD0823	Putative lantibiotic resistance ABC transporter, permease protein	*Bacillus sp. / B. licheniformis / Ruminococcus gnavus*
CD0824	Putative lantibiotic resistance ABC transporter, permease	*B. licheniformis / S. pyogenes / R. gnavus*
CD0853	Oligopeptide ABC transporter, permease protein	*B. subtilis / L. lactis*
CD0854	Oligopeptide ABC transporter, permease protein	*B. subtilis*
CD0855	Oligopeptide ABC transporter, substrate-binding lipoprotein	*B. subtilis*
CD1095	Putative lantibiotic ABC transporter, permease protein	*S. pyogenes / C. difficile / B. licheniformis*
CD1096	Putative lantibiotic ABC transporter, permease protein	*L. lactis / B. licheniformis*
CD1350	Putative antibiotic-related ABC transport system, permease protein	*S. mutans / B. subtilis*
CD1351	Putative antibiotic-related ABC transport system, permease protein	*S. mutans / B. subtilis / S. pyogenes*
CD1647	Ferric siderophore ABC transporter, permease protein	*Campylobacter coli / Vibrio anguillarum / B. subtilis*
CD1648	Ferric siderophore ABC transporter, permease protein	*C. coli / V. anguillarum / B. subtilis*

Table 3.1 continued

CD1650	Ferric siderophore ABC transporter, substrate-binding protein	*C. coli / V. anguillarum*
CD1873	ABC transporter, permease protein (vancomycin resistance)	*Streptococcus pneumoniae*
CD1875	ABC transporter, permease protein (VexP3)	*S. pneumoniae / Enterococcus faecalis*
CD2672	Oligopeptide ABC transporter, substrate-binding protein	*B. subtilis / Listeria innocua*
CD2674	Oligopeptide ABC transporter, permease protein	*B. subtilis*
CD2876	Putative ferrichrome ABC transporter, permease protein	*B. subtilis*
CD2877	Putative ferrichrome ABC transporter, permease protein	*B. subtilis*
CD2878	Putative ferrichrome ABC transporter, substrate-binding protein	*B. subtilis*
CD2953	ABC transporter, substrate-binding protein (possible iron binding)	*Bacillus thuringiensis*
CD2998	Putative iron ABC transporter, permease protein	*Corynebacterium diphtheriae / Bordetella avium*
CD2999	Putative iron ABC transporter, substrate-binding protein	*C. tetani*
CD3525	Putative iron ABC transporter, solute-binding protein	*Bacillus cereus*
CD3526	Putative iron ABC transporter, permease protein	*V. vulnificus*
CD3528	Putative iron ABC transporter, solute-binding lipoprotein	*Bacillus anthracis*
CD3529	Putative iron ABC transporter, permease protein	*V. vulnificus / Actinobacillus pleuropneumoniae*

Table 3.2 Virulence related two component signal transduction proteins

CDS	Protein	Function	Homology to
CD0482	SpaK	Putative lantibiotic resistance two-component sensor kinase	*B. subtilis*
CD1625	VanS	Two-component system, sensor kinase	*E. faecalis*
CD1877	VncS	Two-component system, sensor histidine kinase	*S. pneumoniae/ E. faecalis*

3.2.2.3 Phage/transposons

The *C. difficile* 630 chromosome contains a large number of phage and transposons, with mobile elements constituting 11% of the genome. Seven conjugative transposons are present in the genome, including Tn*5397* and three further highly related transposons, which confer tetracycline resistance. Two transposons are similar to Tn*1549*, which confers VanB-type glycopeptide resistance in *E. faecalis* (Garnier *et al.*, 2000). A further transposon contains lantibiotic resistance genes and carries a gene (CD1113) whose product is thought to be involved in conjugative transfer. A mobilisable, non-conjugative transposon, Tn*5398*, is also present and encodes erythromycin resistance. Furthermore, two prophage and one putative prophage are found within the genome. In addition, several remnants of phage and transposons are distributed throughout the genome.

Most of the 56 phage and transposon proteins identified by the bioinformatics searches have been classified as non-pathogenicity related, because they may be variable between strains and therefore would not make good vaccine candidates. In most cases, the cargo genes carried within the phage/transposons were not highlighted by the search criteria. However, in one case, CD0386, a putative collagen binding surface protein was identified, located within transposon 1, which may be an important aid to the bacterium in adhesion to the host.

3.2.2.4 Sporulation

A number of genes associated with sporulation were identified by the genome searches. Although sporulation is key for the bacterium to survive adverse conditions, these genes have not been classified as putative virulence determinants as they are unlikely to be responsible for the manifestation of disease in patients. Twenty-four genes thought to be involved in sporulation were found, and eliminated from the list of putative virulence factors.

From the original list of 1170 putative virulence CDSs, 634 CDSs were designated as non-virulent by the above criteria and were removed from the list.

3.2.3 Filtering criteria: unknown proteins

After removal of housekeeping genes above, 536 CDSs remain on the list of possible virulence factors. These CDSs were subdivided and screened further (Figure 3.1).

3.2.3.1 ORFans

Of the 541 CDSs remaining, 113 have no similarity to any other known protein, and have been classed as "ORFans". Considering the difficulty in distinguishing short non-coding ORF products from real genes, and given the bias towards short proteins in datasets of ORF products having no matches to proteins in databases (Skovgaard *et al.*, 2001), a further reduction of putative virulence CDSs based on ORF length was carried out. Thirty-three gene products predicted to be smaller than 150 residues were removed from the list of putative virulence factors, leaving 80 CDSs still considered as putative virulence factors. Although these remaining CDSs are not homologous to other proteins, some contain motifs that may indicate their localisation and/or function. For example, 17 CDSs are predicted to be exported from the cell and 51 are likely to be associated with the cell membrane. Four CDSs contained motifs that may confer a function to the protein. CD0839 contains a domain for a membrane-associated CAAX amino-terminal protease. CD2408 has a motif typical of a multi-drug efflux pump, and CD2831 and CD3246 are predicted to be surface-anchored proteins.

3.2.3.2 Hypothetical proteins

Hypothetical proteins have similarity to other predicted proteins but have not been studied experimentally, and so are uncharacterised in terms of expression and putative function. The genome searches identified 55 hypothetical proteins. 22 CDSs were predicted to encode proteins smaller than 150 residues and were removed from the list of putative virulence factors, for the reasons described in 3.2.3.1 (Skovgaard *et al.*, 2001). After this step, 33 CDSs remain on the list of putative virulence factors. To reduce the number of candidate virulence factors further, the BLAST analysis of CDSs were compared. CDSs for which similarity to a protein from a non-pathogenic organism (e.g., *C. acetobutylicum* and *B. subtilis*) was more significant than the hits to other taxa (expectation values at least 100 times lower than that of the nearest competitor) were removed from the list of virulence candidates. The rationale behind this filtering criterion is that proteins that are conserved between related pathogenic

and non-pathogenic bacteria are unlikely to be responsible for the unique attributes of *C. difficile* virulence. However, ORF products with similarity to pathogenic bacteria from unrelated taxa or from eukaryotic organisms were retained. After this, 29 hypothetical proteins were still considered possible virulence factors, of which 20 are conserved, five are predicted membrane proteins, one is a putative lipoprotein and one is a conserved vancomycin resistance-related hypothetical membrane protein.

3.2.3.3 Putatively assigned proteins

A large number of putatively assigned proteins have been identified by the bioinformatic searches. Putative proteins have been recognised by specific domains found within the polypeptide sequence, e.g., transmembrane domains, signal peptides, lipoprotein anchoring or similarity to known proteins. Putatively assigned proteins were analysed in the same way as the ORFans and hypothetical proteins, with the removal of 116 CDSs encoding proteins predicted to be smaller than 150 residues or similar to non-pathogenic bacteria. By these filtering criteria, 281 CDSs were reduced to 165 putative virulence factors, comprising 120 membrane proteins, 31 exported proteins and 14 lipoproteins. There are a number of CDSs of particular interest within this group. CD0183 is a putative exported protein with similarity over its carboxyl-terminal domain to *N. meningitidis* outer membrane proteins and *Listeria* species invasion associated protein p60. The gene for another putatively exported protein, CD2797, is located within a cluster of genes encoding surface proteins and contains Pfam motifs for fibronectin type III and dockerin repeats. Two highly similar lipoproteins, CD0569 and CD0570 contain a basic membrane protein motif, Pfam group PF02608. This family of proteins are usually found in the outer membrane of Gram-negative bacteria and are antigenic in nature when expressed by pathogens.

This bioinformatic analysis has identified a total of 196 CDSs of unknown function as potential virulence factors (Table 3.3).

3.2.4 Possible virulence factors with putatively assigned functions

A total of 90 CDSs with a predicted function remain on the list of putative virulence factors. These will be divided into groups and discussed.

3.2.4.1 Autolysins

The autolysin family of proteins in *C. difficile* contains two important groups, the autolysins themselves and the SLPs (Table 3.4). Autolysins are members of a widely

Table 3.3 Putative virulence factors with unknown function

CDS	Identified by search criteria	Similarity
CD0044	Membrane	*Clostridium perfringens*
CD0111	Membrane/SigP	*C. perfringens/Streptococcus agalactiae*
CD0183	SigP	*C. perfringens/Listeria innocua* p60 protein
CD0190	Membrane/SigP	*Enterococcus faecalis/Lactobacillus plantarum*
CD0272	Membrane	*Clostridium tetani*
CD0281	Membrane/SigP	No similarities
CD0314	Membrane	*C. perfringens/Staphylococcus aureus*
CD0347	Membrane/SigP	No similarities
CD0354	Membrane	*Pseudomonas aeruginosa*
CD0393	Membrane/SigP	*C. tetani/Streptococcus gordonii*
CD0415	Membrane	*E. faecalis/S. agalactiae*
CD0419	Membrane	*E. faecalis*
CD0420	Lipoprotein/membrane/SigP	No similarities
CD0429	Membrane/SigP	*S.agalactiae/Streptococcus pneumoniae*
CD0430	Lipoprotein/membrane/ PS00013/SigP	*Treponema denticola/E. faecalis*
CD0438	Membrane/SigP	*Bacillus cereus*
CD0457	Membrane	No similarities
CD0517	PS00013/membrane	*Clostridium acetobutylicum/C. tetani*
CD0518	Membrane	*C. perfringens/C. tetani*
CD0519	Membrane	*Porphyromonas gingivalis*
CD0529	Membrane/SigP	*C. tetani/Fusobacterium nucleatum subsp. Vincentii*
CD0530	Membrane/SigP	*C. tetani/F. nucleatum subsp. Vincentii*
CD0545	Lipoprotein/PS00013/ membrane/SigP	*Salmonella typhi*
CD0546	Membrane	*C. perfringens/Bacillus megaterium*
CD0550	Membrane/SigP	*P. gingivalis*
CD0628	Membrane	No similarities
CD0632	Repeat	No similarities
CD0651	Membrane/SigP	*E. faecalis/L. plantarum*
CD0676	Membrane	*C. tetani*
CD0702	Membrane/SigP	*Listeria monocytogenes/Streptococcus mutans*
CD0710	Membrane	No similarities
CD0738	Membrane/SigP	*C. tetani/Treponema denticola*
CD0739	Membrane/SigP	No similarities
CD0747	Lipoprotein/PS00013/SigP	*E. faecalis*
CD0777	Membrane	*C. tetani/F. nucleatum*
CD0786	Membrane/SigP	*C. tetani*

Table 3.3 continued

CDS	Identified by search criteria	Similarity
CD0798	Membrane	*Streptomyces hygroscopicus/ Campylobacter jejuni*
CD0881	Membrane/SigP	*C. difficile/Escherichia coli*
CD0981	Membrane	No similarities
CD0984	Membrane	*C. tetani*
CD1012	Membrane/SigP	No similarities
CD1032	Membrane/SigP	*C. tetani/C. acetobutylicum*
CD1045	Membrane	*C. tetani*
CD1052	Membrane/SigP	No similarities
CD1080	Lipoprotein/PS00013/SigP	No similarities
CD1085	Membrane/SigP	*E. faecalis/L. plantarum*
CD1088	Membrane	No similarities
CD1112	Membrane	*E. faecalis/S. agalactiae*
CD1135	SigP	*L. monocytogenes/L. innocua*
CD1136	Membrane	No similarities
CD1155	Membrane	No similarities
CD1156	Lipoprotein/membrane/SigP	No similarities
CD1210	Membrane/SigP	No similarities
CD1221	Membrane/SigP	*C. tetani/C. acetobutylicum*
CD1243	Membrane	No similarities
CD1245	Membrane	No similarities
CD1301	Membrane	*Bacillus subtilis/Bacillus anthracis*
CD1346	Membrane	No similarities
CD1347	Membrane/SigP	No similarities
CD1348	Lipoprotein/PS00013/SigP	No similarities
CD1354	Membrane	*C. perfringens*
CD1377	Membrane	No similarities
CD1393	Membrane/SigP	*C. tetani/S. aureus*
CD1406	Membrane/SigP	No similarities
CD1415	Membrane	No similarities
CD1435	Membrane	No similarities
CD1438	Membrane	*C. tetani/Rattus norvegicus*
CD1440	Membrane	*C. perfringens/ Methanosarcina acetivorans*
CD1441	Membrane/SigP	*C. acetobutylicum/C. perfringens*
CD1443	Membrane/SigP	*S. aureus/Staphylococcus epidermidis*
CD1486	Lipoprotein/PS00013/SigP	No similarities
CD1496	SigP	*C. tetani/B. subtilis*
CD1505	Membrane/SigP	No similarities
CD1509	Lipoprotein/membrane/SigP	No similarities
CD1515	PS00013	*P. aeruginosa/B. cereus*
CD1520	Membrane/SigP	No similarities
CD1523	Membrane	*F. nucleatum/C. tetani*

Table 3.3 continued

CD1575	Membrane/SigP	*C. perfringens/C. tetani*
CD1590	Membrane	*C. perfringens/B. anthracis*
CD1609	Membrane/SigP	No similarities
CD1611	Membrane/SigP	No similarities
CD1616	Membrane	*B. cereus/B. anthracis*
CD1619	Membrane	*Streptococcus pyogenes/S. mutans*
CD1622	Lipoprotein/SigP	*L. innocua/L. plantarum*
CD1630	Membrane	*C. perfringens/Bacillus halodurans*
CD1645	Membrane	*Xanthomonas campestris/B. halodurans*
CD1656	Membrane	*S. mutans/P. gingivalis*
CD1668	Membrane/SigP	No similarities
CD1673	Membrane	*Bacteroides nodosus* (possible fimbrial assembly protein)/ *Lactococcus lactis*
CD1677	Membrane	*C. perfringens/S. pyogenes*
CD1678	Membrane	*C. perfringens/T. tencongensis*
CD1683	Membrane/SigP	*C. perfringens/L. plantarum*
CD1687	Lipoprotein/membrane/SigP	No similarities
CD1766	Lipoprotein/PS00013	No similarities
CD1768	Membrane	*P. gingivalis*
CD1781	Lipoprotein/PS00013/SigP	No similarities
CD1789	Membrane	*C. tetani/C. acetobutylicum*
CD1815	SigP	*C. perfringens/C. tetani*

CDS	Identified by search criteria	Similarity
CD1819	Membrane	*Vibrio paraheamolyticus/Vibrio vulnificus*
CD1842	Membrane	*Staphylococcus gallinarum/C. perfringens*
CD1888	Membrane	No similarities
CD1890	Membrane/SigP	*Streptococcus hominis/S. algalactiae*
CD1928	Membrane	No similarities
CD1929	Membrane/SigP	*Xanthomonas campestris*
CD1930	SigP	*C. acetobutylicum/C. tetani*
CD1933	Membrane	*B. cereus*
CD1940	Lipoprotein/membrane	No similarities
CD1968	Membrane	*L. monocytogenes*
CD1971	SigP	No similarities
CD1994	Membrane	*Streptococcus iniae/S. mutans*
CD1995	Membrane/SigP	*C. tetani*
CD2015	Membrane	*Paenibacillus popilliae/B. anthracis*
CD2029	Lipoprotein/PS00013/SigP	*B. cereus*
CD2044	Membrane/SigP	No similarities
CD2051	Membrane	*B. subtilis/C. perfringens*
CD2052	Lipoprotein/PS00013/SigP	No similarities
CD2056	Membrane	*C. tetani*
CD2058	Membrane/SigP	No similarities

Table 3.3 continued

CDS	Identified by search criteria	Similarity
CD2062	Membrane	*B. anthracis/B. subtilis*
CD2063	SigP	*L. monocytogenes (antiterminator bvrA)*
CD2067	Membrane/SigP	No similarities
CD2097	Membrane/SigP	*C. perfringens*
CD2127	Membrane/SigP	No similarities
CD2142	Membrane	*V. vulnificus*
CD2159	Membrane	*L. monocytogenes/C. acetobutylicum*
CD2205	Membrane	*C. perfringens*
CD2228	Membrane	*L. innocua*
CD2229	Membrane/SigP	*C. perfringens*
CD2243	Membrane	No similarities
CD2250	Membrane	*C. tetani*
CD2255	Membrane	No similarities
CD2265	PS00013/membrane	*C. tetani*
CD2273	Membrane	*C. tetani/T. tencongensis*
CD2295	Membrane	No similarities
CD2299	Membrane	No similarities
CD2315	PS00013/membrane/SigP	*C. difficile*
CD2337	Membrane/SigP	No similarities
CD2346	PS00013/membrane/SigP	No similarities
CD2366	Membrane/SigP	No similarities
CD2376	Membrane	*B. anthracis/L. monocytogenes*
CD2395	Membrane	No similarities
CD2434	Membrane	*C. perfringens*
CD2479	Membrane	*C. tetani/F. nucleatum*
CD2520	Membrane	*C. acetobutylicum/Paenibacillus popilliae*
CD2617	Membrane	*C. tetani*
CD2624	Membrane/SigP	No similarities
CD2625	PS00013/membrane/SigP	*C. perfringens/S. pyogenes*
CD2628	Membrane/SigP/Sortase substrate	No similarities
CD2649	Membrane/SigP	*C. perfringens/T. tencongensis*
CD2689	Lipoprotein/PS00013/SigP	*C. tetani/C. perfringens*
CD2699	Membrane	*E. coli*
CD2718	Membrane	*C. perfringens*
CD2720	PS00013/membrane	*B. anthracis*
CD2727	PS00013/membrane	*C. perfringens*
CD2728	PS00013/membrane/SigP	*C. perfringens*
CD2763	PS00013/membrane/SigP	No similarities
CD2785	PS00013/membrane	*Leptospira interrogans*
CD2797	SigP	No similarities

Table 3.3 continued

CDS	Identified by search criteria	Similarity
CD2803	Membrane	No similarities
CD2808	Membrane	No similarities
CD2810	Membrane	*L. monocytogenes/C. acetobutylicum*
CD2820	Membrane	*E. coli/V. parahaemolyticus*
CD2821	Membrane	*E. coli/Vibrio cholera*
CD2830	Membrane/SigP	No similarities
CD2856	Membrane	*C. perfringens/E. faecalis/S. agalactiae*
CD2881	PS00013	*B. anthracis/L. innocua*
CD2977	Repeat	*Archaeoglobus fulgidus*
CD2980	SigP	*C. tetani/F. nucleatum*
CD3042	Membrane/SigP	*C. tetani*
CD3043	Membrane/SigP	No similarities
CD3051	SigP	*B. cereus/Gloeobacter violaceus*
CD3059	Membrane	No similarities
CD3063	Membrane/SigP	*C. tetani*
CD3073	PS00013/membrane	*E. coli/L. plantarum*
CD3078	Membrane/SigP	*C. perfringens/B. subtilis*
CD3108	Membrane/SigP	*F. nucleatum/C. tetani*
CD3114	PS00013/membrane	*E. faecalis/F. nucleatum*
CD3163	PS00013	*F. nucleatum/L. interrogans*
CD3218	Membrane/SigP	*C. difficile*
CD3228	Membrane	*S. pneumoniae*
CD3236	Membrane	*E. faecalis/L. plantarum*
CD3264	Membrane	No similarities
CD3291	Membrane/SigP	*Borrelia hermsii*
CD3347	Membrane	No similarities
CD3366	Membrane/SigP	*C. tetani*
CD3369	Membrane	No similarities
CD3464	Lipoprotein/membrane/SigP	*C. tetani/C. acetobutylicum*
CD3506	Membrane/SigP	No similarities
CD3510	Membrane	No similarities
CD3519	Membrane/SigP	No similarities
CD3556	Membrane/SigP	*C. perfringens/L. plantarum*
CD3617	PS00013	*B. cereus/S. aureus*
CD3633	Membrane/SigP	*B. anthracis/O. iheyensis*
CD3636	Membrane/SigP	No similarities
CD3660	Membrane/SigP	*B. anthracis/O. iheyensis*
CD3669	Lipoprotein/PS00013/SigP	*C. tetani/Methanosarcina acetivorans*

Table 3.4 Autolysins in the *C. difficile* 630 genome.

CDS	Protein	Function	Identified by search	Similarity
CD0106	CwlD	Germination-specific *N*-acetylmuramoyl-L-alanine amidase	Membrane/PF01520	*B. subtilis*
CD0125	-	Putative cell wall endopeptidase	Membrane/SigP	*C. tetani/T. tencongensis*
CD0440	SLP paralogue 27	Cell surface protein	Lipoprotein/ PF04122/SigP	*C. difficile/B. subtilis*
CD0514	SLP paralogue 15	Putative cell surface hemagglutinin/adhesin	PF04122/SigP	*C. difficile/B. subtilis/Streptococcus gordonii*
CD0784	-	Putative *N*-acetylmuramoyl-L-alanine amidase	PF01520/Membrane/SigP	*B. cereus/C. perfringens*
CD0844	SLP paralogue 25	Cell surface protein	PF04122/SigP	*C. difficile*
CD1034	-	Putative mannosyl-glycoprotein endo-beta-*N*-acetylglucosamidase	Membrane/SigP	*C. tetani*
CD1035	SLP paralogue 16	Putative cell surface *N*-acetylmuramoyl-L-alanine amidase	PF01506/ PF04122/SigP	*C. difficile/B. subtilis*
CD1036	SLP paralogue 17	Putative cell surface *N*-acetylmuramoyl-L-alanine amidase	PF01506/ PF04122/SigP	*C. difficile/B. subtilis*
CD1047	SLP paralogue 18	Putative cell surface protein	PF01422/SigP	*C. difficile/B. subtilis*
CD1233	SLP paralogue 29	Putative cell surface protein	Lipoprotein/ PF04122/SigP	*C. difficile/C. tetani*
CD1304	-	Putative endo-beta-*N*-acetylglucosamidase	SigP	*S. caprae (fibronectin binding autolysin)/ B. cereus (enterotoxin)/ S aureus (bifunctional autolysin precursor)*
CD1469	SLP paralogue 20	Putative cell surface protein	PF04122/SigP	*B. cereus (penicillin binding protein)/C. difficile*
CD1751	SLP paralogue 13	Putative cell surface protein similar to Cwp84	PF04122/ membrane/SigP	*C. difficile*
CD1803	SLP paralogue 23	Putative cell surface protein	PF04122/SigP	*C. difficile*
CD1987	SLP paralogue 28	Putative cell surface protein	PF04122/SigP	*C. difficile/B. subtilis*
CD2193	SLP paralogue 24	Putative cell surface protein	PF04122/membrane/SigP	*C. difficile/C. tetani*
CD2402	-	Putative cell wall hydrolase	SigP	*B. cereus (enterotoxin)/ L. monocytogenes (invasion associated protein p60)*

Table 3.4 continued

CDS	Protein	Function	Identified by search	Similarity
CD2518	SLP paralogue 26	Cell surface protein	PF04122/SigP	*C. difficile/B. subtilis*
CD2713	SLP paralogue 22	Cell surface protein	Lipoprotein/PF1473/PF04122/SigP	*C. difficile*
CD2735	SLP paralogue 14	Putative cell surface protein	PF04122/SigP	*C. difficile*
CD2761	-	Putative *N*-acetylmuramoyl-L-alanine amidase	PF01520/membrane	*Paenibacillus polymyxa/B. subtilis*
CD2767	SLP paralogue 19	Putative cell surface protein	PF04122/SigP	*C. difficile*
CD2768	-	Putative cell-wall hydrolase	Membrane/SigP	*Listeria welshimeri/T. tencongensis*
CD2782	SLP paralogue 7	Putative cell surface protein	PF04122/ membrane/ SigP	*C. difficile*
CD2784	SLP paralogue 6	Putative cell surface protein	PF01520/ PF04122/SigP	*C. difficile*
CD2786	SLP paralogue 5	Putative cell surface protein	PF04122/ membrane/ SigP	*C. difficile*
CD2787	Cwp84	Cell surface-associated cysteine protease	PF04122/ membrane/ SigP	*C. difficile*
CD2789	Cwp66	Cell surface protein	PF04122/SigP	*C. difficile*
CD2791	SLP paralogue 2	Putative S-layer protein precursor	PF04122/SigP	*C. difficile*
CD2793	SlpA	S-layer precursor protein	PF04122/membrane/ SigP	*C. difficile*
CD2794	SLP paralogue 12	Putative surface protein	PF04122/ membrane/ SigP	*C. tetani/B. licheniformis*
CD2795	SLP paralogue 11	Putative cell surface protein	PF04122/SigP	*C. tetani/B. subtilis*
CD2796	SLP paralogue 10	Putative cell surface protein	PF04122/ membrane/ SigP	*C. tetani/B. subtilis/C. difficile*
CD2798	SLP paralogue 9	Putative cell surface protein	PF04122/ membrane/ SigP	*B. subtilis/C. difficile*
CD2799	SLP paralogue 8	Putative cell surface protein	PF04122/SigP	*C. difficile*
CD3192	SLP paralogue 21	Putative cell surface protein	PF04122/SigP	*C. difficile/C. tetani*

distributed group of enzymes that naturally digest the peptidoglycan backbone of the cell wall, and play a role in a variety of cellular functions, including cell division and sporulation. According to their hydrolytic bond-specificity, autolysins are classified as muraminidases, glucosaminidases, *N*-acetylmuramoyl-L-alanine amidases, amidases and endopeptidases. In the *C. difficile* 630 genome, seven autolysins were found by the bioinfomatic searches. CD0106 encodes CwlD, an *N*-acetylmuramoyl-L-alanine amidase that has a specific role in sporulation. This makes it unlikely to be a virulence determinant, and it will not be considered further. There are 2 further putative *N*-acetylmuramoyl-L-alanine amidases (CD0784, CD2761) in the genome, along with 2 glucosamidases (CD1034, CD1304). Interestingly, one of the glucosamidases, CD1304, also shows similarity to the *Staphylococcus caprae* fibronectin-binding autolysin, and the *B. cereus* enterotoxin. Two putative cell wall hydrolases (CD2402, CD2768) have also been found, both containing the *E. coli* NlpC/Listeria P60 motif, PF00877. This motif is located near the carboxyl-terminus of a number of different bacterial and viral proteins. In bacteria, the proteins that contain this motif include putative tail components of prophage, invasin, invasion associated proteins, putative lipoproteins, cell wall hydrolases, and putative endopeptidases. Proteins containing the NlpC/P60 motif sometimes shows cysteine peptidase activity, e.g. dipeptidyl-peptidase VI from *Bacillus sphaericus*. However, the motif has also been found in non-peptidase homologues that have either been found experimentally to be without peptidase activity, or lack amino acid residues essential for catalytic activity. One of the hydrolases, CD2402, shows homology to *B. cereus* (enterotoxin) and *L. monocytogenes* (invasion associated protein p60).

The largest group of proteins in the autolysin family is a set of paralogues of the surface layer protein, SlpA. This protein is post-translationally cleaved into two proteins; one of which is highly variable between *C. difficile* strains (LMW SLP) and the other has homology to CwlB of *B. subtilis*, with *N*-acetylmuramoyl-L-alanine amidase activity (HMW SLP). There are 28 paralogues of SlpA within the genome of *C. difficile*, each with a common *N*-acetylmuramoyl-L-alanine amidase domain and a second domain that varies between paralogues. Each paralogue also contains a signal peptide, indicating the potential to be exported from the cell, but some also contain transmembrane domains, and so may be anchored to the cell envelope. None of the paralogues were found to contain an SLH-domain, which is important for attachment of the SLPs to the cell wall in other bacteria e.g. *B. anthracis* and *Thermonanaerobacterium* (Mesnage *et al.*, 1999b; Mesnage *et al.*, 2000; Cava *et al.*, 2004). The majority of Slp paralogues do not have a known function. Exceptions to this are SlpA (surface layer proteins), Cwp66 (an adhesin) and Cwp84 (a cysteine

protease). The variable domains of some other paralogues contain motifs that may confer a function to the protein. For example, CD0514, SLP paralogue 15 contains repeat motifs for haemagglutinin binding and CD1751 is very similar to *cwp84* and contains domains for cysteine protease activity. Not highlighted by the bioinformatic searches, but also of interest is CD0843, which encodes a putative glycosyl transferase. This may act to glycosylate the adjacent Slp paralogue 25.

3.2.4.2 Flagella and chemotaxis

The *C. difficile* 630 genome encodes the full machinery necessary to produce functional flagella (even though not all strains express them (Tasteyre *et al.*, 2000)), and also six proteins involved in chemotaxis, totalling 35 motility CDSs. Flagella, and in particular flagellin (FliC) and the flagellar cap protein (FliD) are associated with adherence and gut colonisation in *C. difficile* (Tasteyre, 2001). Flagella are also involved in adherence and colonisation in many other species of bacteria, e.g. *Aeromonas,* enteropathogenic *E. coli, Burkholderia cenocepacia* and *Salmonella enterica* (Van Asten *et al.*, 2000; Giron *et al.*, 2002; Tomich *et al.*, 2002; Gavin *et al.*, 2003). Furthermore, the ability of bacteria to move in response to chemical stimuli has also been associated with virulence in a number of bacteria, for example *Vibrio parahaemolyticus* (reviewed in Ottemann & Miller (1997)).

3.2.4.3 Pili

There is one published report of pili in *C. difficile* in the literature (Borriello *et al.*, 1988), and consistent with this, the bioinformatic search of the genome identified a cluster of CDSs that are predicted to encode the machinery for type IV pilus biogenesis (Figure 3.2). Predicted gene products in this cluster include type IV pilin (CD3507, CD3508, CD3513), assembly proteins (CD3509, CD3511, CD3512), leader peptidases (CD3503, CD3504) and a pilus retraction protein that may cause twitching mobility (CD3505). Three further pilus-related proteins were found in separate locations in the genome. These were CD3294 that encodes a putative type IV pilin; CD3295 that encodes a putative type IV pilus assembly protein; and CD2305 that encodes a putative pilin/general secretion pathway protein.

3.2.4.4 Drug resistance

C. difficile infection is linked with the use of antibiotics. Therefore resistance to antibiotics is particularly relevant in *C. difficile* infections. The bioinformatics searches



Figure 3.2 Artemis screen shot showing cluster of genes thought to be involved in pilus biogenesis. These ORFs encode a number of components of a type IV pilus. ORFs are predicted to encode proteins as follows: CD3503 prepilin leader peptidase; CD3504 prepilin leader peptidase; CD3505 pilus retraction protein; CD3506 putative exported protein; CD3507 type IV pilin; CD3508 type IV pilin; CD3509 type IV pilus assembly protein; CD3510 putative membrane protein; CD3511 type IV pilus assembly protein; CD3512 type IV pilus assembly protein; and CD3513 pilin.

identified 29 CDSs involved in drug resistance, detailed in Table 3.5. All genes located within mobilisable genetic elements (e.g. tetracycline resistance on CTn*5397* or skin element encoded teicoplanin resistance) were not included in the analysis, as their presence may vary between strains.

A cluster of genes involved in vancomycin resistance was identified by the bioinformatic searches (Figure 3.3A); however, these genes do not form the full operon needed for drug resistance. Within this cluster are encoded a two-component signal transduction system composed of a response regulator VanR, and a histidine kinase VanS. CD1626 encodes VanG, a putative D-alanine:D-alanine ligase, similar to vancomycin/teicoplanin A-type resistance proteins from *Enterococcus faecalis*. CD1627 is a putative D-alanyl-D-alanine carboxylpeptidase and CD1628 encoding VanTG is a serine/alanine racemase; both have homologues in *E. faecalis*. In addition to the cluster of genes, isolated CDSs with homology to vancomycin resistance proteins were identified. CD1436 is a conserved vancomycin resistance-related hypothetical membrane protein with amino acid similarity to similar proteins in *Thermoanaerobacter tencongensis* and *Clostridium tetani*. CD3599 is a two-component sensor histidine kinase similar to VanS in *E. faecium* and is located adjacent to *luxS*, encoding S-ribosylhomocysteinase, a putative quorum-sensing molecule (Carter *et al.*, 2005). Despite the presence of these genes, *C. difficile* 630 is sensitive to vancomycin at concentrations greater than 1 μg/ml (Chapter 5).

Seven CDSs with a putative role in β-lactam resistance were found. This includes a possible operon (Figure 3.3B) containing a putative β-lactamase, the machinery for an ABC-transport system, a two-component signal transduction system and a TetR transcriptional regulator (discussed below). In addition, two penicillin-binding protein repressors, CD2765 and CD2766, are present at the start of a cell surface complex cluster of genes.

A number of genes were found that contain a MatE motif (PF01554). Two CDSs, CD1506 and CD3361 contain two motifs conserved amongst efflux proteins of the DinF branch of the multi-antimicrobial and toxic compound extrusion (MATE) family. Although not annotated as such, these genes may encode CdeA, described by Dridi *et al.* (2004), which mediates *C. difficile* resistance to ethidium bromide, acriflavin and fluoroquinolone antibiotics. Other characterised members of the MATE family have been found to function as drug/sodium antiporters mediating resistance to a wide range of cationic dyes, fluroquinolones, aminoglycosides and other structurally diverse antibodies and drugs. MATE proteins are found in bacteria,

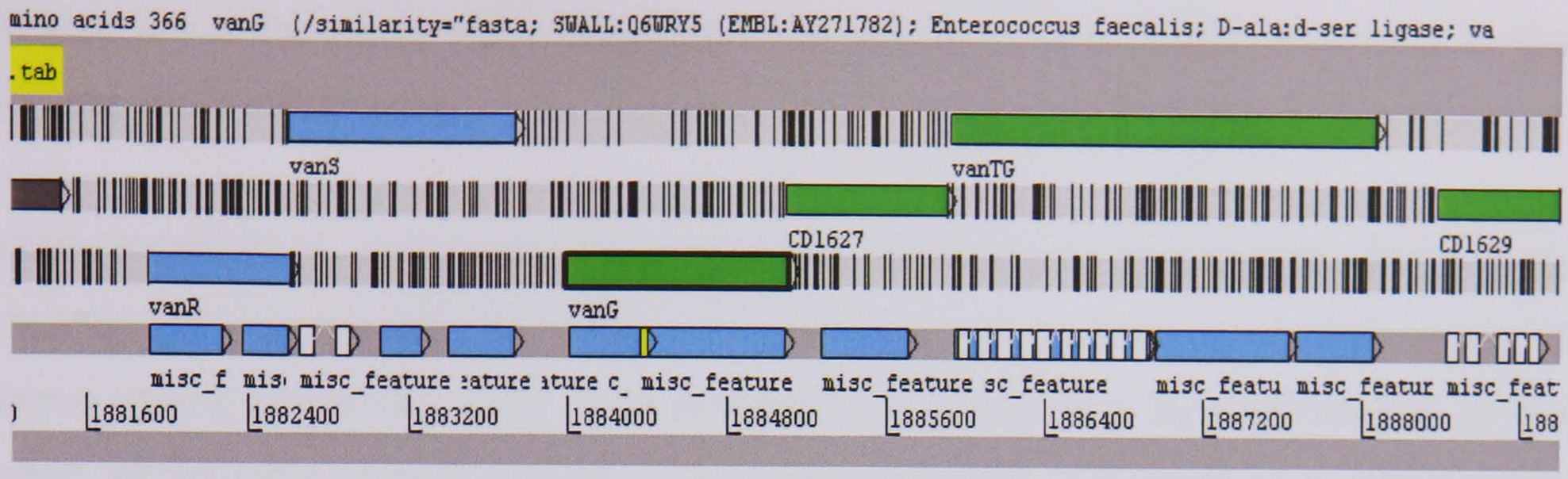
Table 3.5 Putative virulence factors involved in drug resistance

CDS	Function	Identified by search	Similarity
CD0169	Putative drug/sodium antiporter (MatE)	Membrane	*C. perfringens*
CD0458	Putative β-lactamase	Membrane/SigP	*S. typhimurium/E. coli*
CD0470	Putative β-lactamase-inducing penicillin-binding protein, BlaR	Membrane/SigP	*B. licheniformis/S. aureus*
CD0488	Quaternary ammonium compound-resistance protein, SugE	Membrane	*E. coli*
CD0578	Putative efflux protein	Membrane	*Enterococcus spp./L. lactis*
CD0630	Putative drug/sodium antiporter (MatE)	Membrane	*Porphyromonas gingivalis/C. tetani*
CD0895	Metallo-β-lactamase superfamily exported protein	Membrane/SigP	*C. acetobutylicum/B. cereus*
CD0985	Putative drug/sodium antiporter (MatE)	PS00013/membrane	*C. perfringens/ C. tetani/ Bacteroides thetaiotaomicron*
CD1380	Putative transport-related membrane protein	Membrane/SigP	*Methanobacterium thermoautotrophicum/ C. acetobutylicum*
CD1418	Putative multidrug efflux protein (MatE)	Membrane	*F. nucleatum/C. difficile*
CD1436	Conserved vancomycin resistance-related hypothetical membrane protein	Membrane/SigP	*T. tencongensis/C. tetani*
CD1506	EA protein Multi Antimicrobial Extrusion (MATE) family	Membrane	*C. difficile/F. nucleatum*
CD1541	Putative drug/sodium antiporter (MatE)	Membrane	*C. perfringens*
CD1625	Two-component system, sensor kinase, VanS	Membrane	*E. faecalis*
CD1628	Amino acid racemase, VanT/G	Membrane	*E. faecalis/ Enterococcus gallinarum*
CD1681	TetR-family family transcriptional regulator	SigP	*Methanosarcina acetivorans/B. subtilis*
CD2108	Putative drug/sodium antiporter (MatE)	Membrane	*L. monocytogenes*
CD2389	Putative β-lactamase inducer	Membrane	*L. monocytogenes/S. aureus*

Table 3.5 continued

CDS	Function	Identified by search	Similarity
CD2407	Putative multidrug efflux pump, membrane protein	Membrane	*O. iheyensis*
CD2408	Putative multidrug efflux pump, outer membrane protein	SigP	No similarities
CD2478	Metallo β-lactamase superfamily lipoprotein	Lipoprotein/PS00013	*C. tetani/T. tencongensis*
CD2765	Putative penicillin-binding protein repressor	Membrane/SigP	*E. faecalis/Enterococcus hirae/ C. perfringens*
CD2766	Putative penicillin-binding protein repressor	Membrane/SigP	*E. faecalis/E. hirae*
CD2874	Putative drug/sodium antiporter (MatE)	Membrane	*C. perfringens*
CD3198	Multidrug resistance protein, Cme	Membrane	*C. difficile/Streptococcus spp.*
CD3207	Putative Na^+ driven multidrug efflux pump (MatE)	Membrane/SigP	*F. nucleatum*
CD3299	Probable transporter (tetracenomycin c resistance and export)	PS00013/membrane/SigP	*C. difficile/Streptomyces glaucescens*
CD3361	Putative transmembrane MATE family transporter protein	Membrane	*C. difficile/S. aureus*
CD3599	Two component sensor histidine kinase, VanS	Membrane	*E. faecium*

A



B

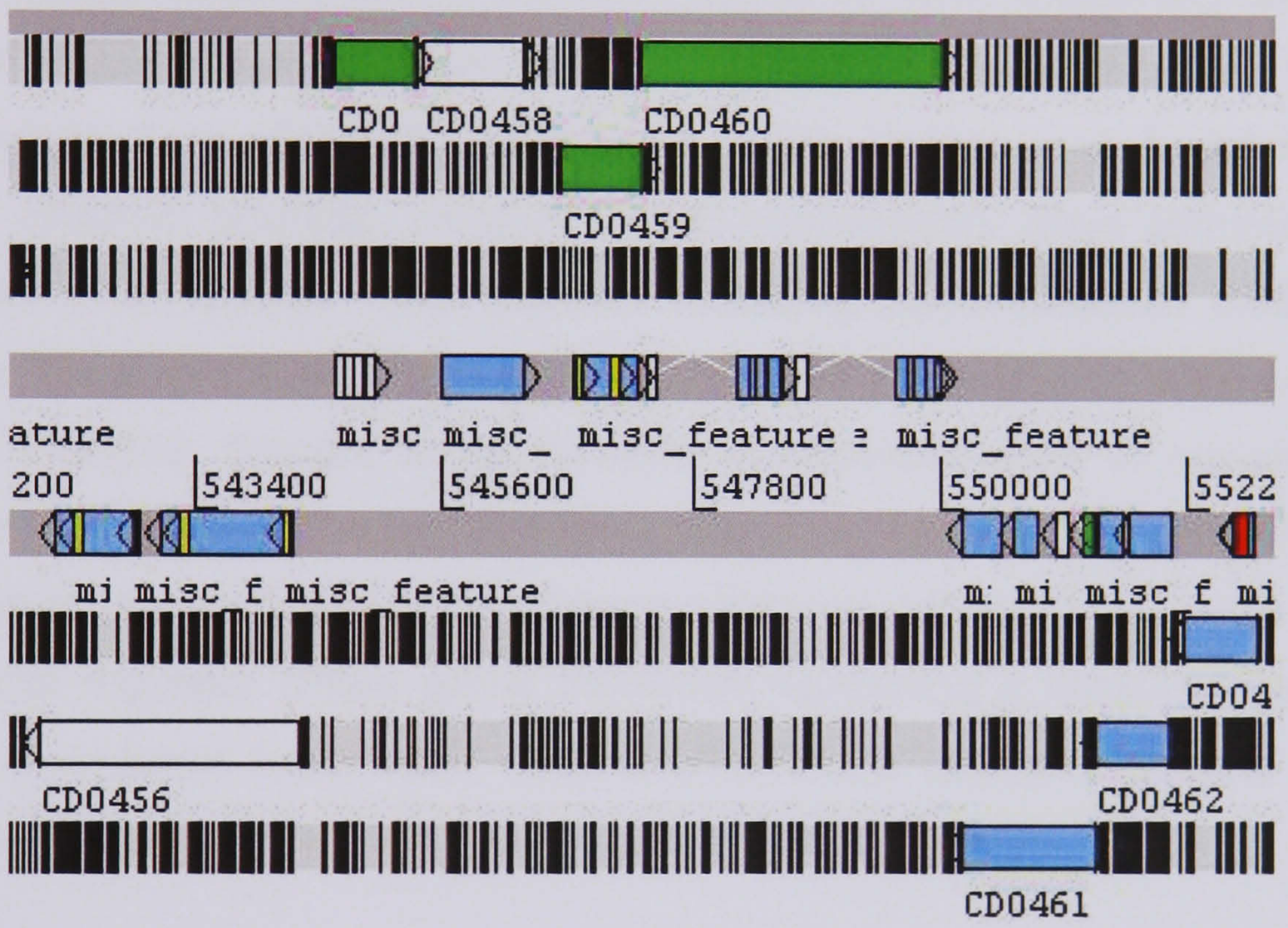


Figure 3.3 Artemis views showing CDS involved in drug resistance. **A**. View from Artemis showing a cluster of genes putatively involved in vancomycin resistance. **B**. View from Artemis showing cluster of genes putatively involved in β-lactam resistance. CD0458 encodes a putative β-lactamase. CD0456 encodes a putative antibiotic resistance ABC transporter ATP-binding protein, and CD0459 and CD0460 encode other components of an ABC-transporter, an ATP-binding protein, and permease protein respectively. CD0461 encodes a two-component sensor histidine kinase, CD0462 encodes a two-component response regulator and CD0463 encodes a TetR-family transcriptional regulator.

archaea and eukaryotes. The MatE Pfam group (PF01554) was found in 10 other proteins highlighted by the bioinformatic search, and a further search of the genome revealed 3 additional CDSs containing this motif. In some cases, the drug/sodium antiporters are adjacent to ABC transporter components, e.g. CD2874 is next to the ferrichrome ABC transporter.

A protein involved in resistance to quaternary ammonium compounds, SugE was detected by the bioinformatic searches. A quaternary ammonium compound is any of numerous strong bases and their salts derived from ammonium by replacement of the hydrogen atoms with organic radicals. They are especially important as surface-active agents, disinfectants, and drugs. SugE was first identified as a suppressor of *groEL* mutations but is also a member of the small multidrug resistance family. High-level expression of *sugE* leads to resistance to a subset of toxic quaternary ammonium compounds in *E. coli* (Chung & Saier, 2002). Additionally, one copy of the multidrug resistance protein Cme (CD3198) was found. This protein was found to confer resistance to ethidium bromide, safranin O and erythromycin when expressed in *E. faecalis* (Lebel *et al.*, 2004). Two adjacent multidrug pump proteins, CD2407 and CD2408 were found by the bioinformatic searches, and two transport related proteins, CD1380 and CD3299, thought to be involved in tetracenomycin C resistance and export were also detected. The bioinformatics searches identified one copy of the *tetR* gene, although further searches of the genome reveal an additional 13 copies. TetR is a transcriptional regulator (Pfam group PF00440) containing a helix-turn-helix motif. Although not a virulence factor itself, TetR is often found upstream of other potential virulence factors, such as the β-lactamase cluster described above.

3.2.4.5 Lantibiotics

Lantibiotics are lanthionine-containing peptide antibiotics produced by Gram-positive bacteria that cause cell death of other Gram-positive bacteria. Eight gene products were found to be involved in lantibiotic export and resistance, excluding those found on mobilisable genetic elements. Many of these genes are located within clusters on the genome. Although *C. difficile* has many genes for lantibiotic resistance, it does not appear to synthesise any lantibiotics itself (excluding CD1107 on CTn*4*, which has similarity to bacteriocin-like proteins from *E. faecalis*).

Most of the lantibiotic resistance genes encoded on the *C. difficile* 630 genome form part of ABC transporters. Most are found in pairs of permeases. CD0823 and CD0824 are both predicted to encode lantibiotic ABC transporter permeases, and are situated in

the genome close to genes for other ABC components, a two-component signal transduction system and some oxidative-stress response proteins. A cluster of genes encoding components similar to subtilin transport in *B. subtilis* were found and are shown in Figure 3.4. This cluster encodes an ABC transporter and a two-component sensory system.

3.2.4.6 Oligopeptide transporters

Utilisation of oligopeptides may confer *C. difficile* with a competitive advantage in the colon. It is possible that when the microbiota is disturbed, for example by antibiotics, peptide concentrations rise in the colon, forming a potential nutrient source for *C. difficile*. The bioinformatic searches of the genome identified 10 CDSs with a putative role in oligopeptide transport, mainly as components of ABC transporters.

3.2.4.7 Proteases

Proteases are widely distributed throughout the genomes of pathogens, and aid the progression of disease by degrading host and competitor proteins, providing nutrients and acting as a defence mechanism. The bioinformatic searches of the *C. difficile* 630 genome identified 12 putative proteases (Table 3.6), although the substrates of many cannot be predicted. CD0125 and CD2787 are also members of the autolysin family, and have been discussed in 3.2.4.1. Two putative membrane associated CAAX amino-terminal proteases were found, CD0156 and CD0839. This family of proteins are believed to be proteases and contain a highly conserved Glu-Glu motif at the amino-terminus and two histidine residues that may be involved in zinc binding (Pei & Grishin, 2001). CD0181 encodes a putative membrane associated metalloprotease. This ORF contains a Pfam PF05569 motif, indicative of a BlaR1 peptidase, and indeed this ORF shows similarity to the amino-terminal half of the *Bacillus licheniformis* BlaR regulatory protein. Interestingly, integral membrane homologues to this protein act as a signal-transducers, which together with transcriptional repressors regulate production of β-lactamase and penicillin-binding protein 2a, which mediates staphylococcal resistance to β-lactam antibiotics (Archer & Bosilevac, 2000). The signal transducer is a fusion protein with penicillin-binding and zinc metalloprotease domains. The signal for protein expression is transmitted by site-specific proteolytic cleavage of both the transducer, which auto-activates, and the repressor, which is inactivated, unblocking gene transcription (Zhang *et al.*, 2001). It is unclear as to whether a similar function could be performed by the product of CD0181.

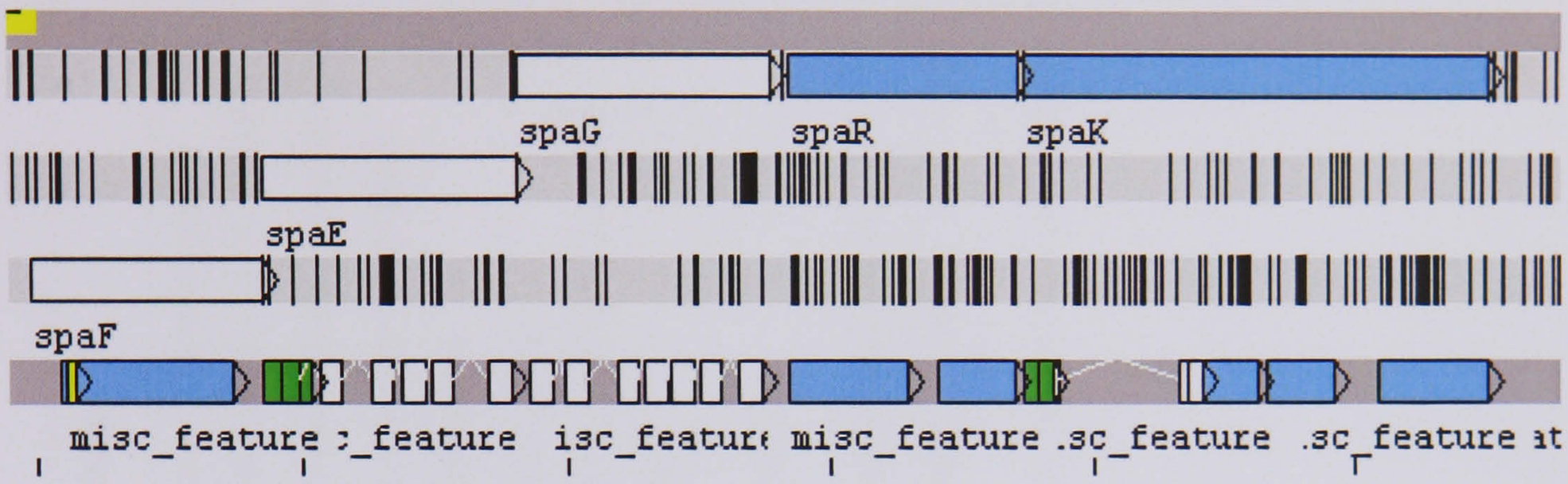


Figure 3.4 Artemis views showing CDS involved in lantibiotic resistance. Screen shot from Artemis showing *spa* cluster of genes involved in lantibiotic resistance. SpaF, SpaE and SpaG encode components of an ABC transporter. SpaR and SpaK encode a two-component response regulator and histidine kinase, respectively.

Table 3.6 Possible virulence factors with protease domains.

CDS	Function	Identified by search	Similarity
CD0125	Putative cell wall endopeptidase	Membrane/SigP	*C. tetani/T. tencongensis*
CD0156	Putative membrane-associated CAAX amino terminal protease	Membrane	*C. acetobutylicum/ B. subtilis*
CD0181	Putative membrane-associated metalloprotease	Membrane/SigP	*B. licheniformis/L. monocytogenes/S. aureus incl MRSA*
CD0695	Putative exported peptidase	Membrane/SigP	*Fusobacterium nucleatum*
CD0839	Putative membrane-associated CAAX amino terminal protease	Membrane	No similarities
CD1294	Metalloprotease	Membrane/SigP	*T. tencongensis*
CD2129	Putative membrane-associated protease	Membrane	*T. tencongensis/E. faecalis*
CD2582	Putative membrane-associated neutral zinc metallopeptidase	Membrane	*C. tetani*
CD2787	Putative cell surface-associated cysteine proteinase, Cwp84	Cwp66/PF04122/ membrane/SigP	*C. difficile*
CD3284	Probable protease	Membrane/SigP	*C. acetobutylicum/E. coli*
CD3455	Putative exported carboxyl-terminal processing protease	Membrane/SigP	*Bartonella bacilliformis/B. subtilis*
CD3652	Putative peptidase	Lipoprotein/PS00013 /membrane/SigP	*T. tencongensis/Streptomyces coelicolor*

3.2.4.8 Toxins

The bioinformatic searches identified the toxin locus (*tcdA, B, C, D* and *E*) in the *C. difficile* 630 genome. In addition, genes *cdtA* and *cdtB* encoding the binary toxin were found, but both have frame shift mutations in this strain of *C. difficile*. These genes and their corresponding proteins have been well characterized in other strains and are discussed in section 1.1.5.1.

3.2.4.9 Adhesins

Several genes were found that contain extracellular matrix binding domains. These include CD2831 (collagen binding, discussed in 3.2.4.10), CD2592 and CD2797 (fibronectin binding). CD2797 also contains a dockerin repeat, which in *C. thermocellum* binds to cohesin in the cellulosome complex (Doi & Kosugi, 2004). However a corresponding cohesin domain could not be detected in the *C. difficile* genome. CD3145 encodes a putative surface anchored protein, which contains six thrombospondin domains and may mediate binding to extracellular matrix components.

3.2.4.10 Sortase and sortase substrates

The sortase enzyme can catalyse the attachment of secreted proteins containing an LPXTG motif to the peptidoglycan of the cell wall (discussed in 1.2.2.9). Most genomes usually contain two genes encoding sortase (Pallen *et al.*, 2001), but in *C. difficile* 630, one of these (CD3146) is disrupted by an in-frame stop codon, so only one sortase enzyme (CD2718) has the potential to function.

Proteins attached by sortase to the cell wall may be considered virulence candidates, as this location is important for mediating interactions between the microbe and host. Seventeen potential sortase substrates were identified, four of which are carried on conjugative transposons and so will not be considered further here. Eight of these sortase substrates do not have any known motifs or homology that could assist in defining their putative function. Interestingly one of these ORFs is a paralogue to the HMW SLP (CD2782, Slp paralogue 7). CD1380 is a putative permease and contains domains for sugar transport and CD2833 is predicted to encode a transport related ATPase. CD2498 encodes D-alanyl-D-alanine carboxylpeptidase and is a putative sporulation specific penicillin-binding protein. CD3145 is of particular interest as a potential virulence factor as it has homology to a serine-aspartate-rich surface anchored fibrinogen-binding protein. This property is conferred by six PF02412 thrombospondin domains, which can bind to the extracellular matrix components

fibrinogen, fibronectin, laminin and type V collagen. CD2831 is predicted to be a putative surface anchored membrane protein and contains two interesting Pfam groups, which may confer collagen-binding properties. Studies in *S. aureus* have shown that the PF05738 motif folds into a β-sandwich structure forming a stalk that extends away from the bacterial cell surface and presents the PF05737 ligand-binding domain to the extra cellular milieu. Interestingly, three of these sortase substrate genes, CD2831 (collagen-binding protein), CD3145 (fibrinogen-binding protein) and CD3246 contain polymeric tracts of DNA immediately downstream from the 5' signal sequence. The region encoded by the polymeric-tract is exposed on the cell surface, and so it is possible that these proteins could undergo phase variation through slipped-strand mispairing of the gene, thereby evading the host immune response.

3.2.4.11 Iron scavenging

Iron is a limiting factor for bacterial growth within a host, so an ability to scavenge iron efficiently is advantageous to pathogens. The bioinformatic analysis identified 17 proteins involved in iron acquisition. Many form parts of ABC transporters, for example CD2874-CD2878 form a ferrichrome ABC transporter and CD1647-CD1649 are form a ferric siderophore ABC transporters.

3.2.4.12 Quorum sensing

It is not clear as to whether *C. difficile* uses a quorum sensing system. Toxin production only begins *in vitro* when a high cell density is achieved (Dupuy, 1998), which could be due to a quorum sensing system, or alternatively be part of a global stress response. Carter *et al.* (2005) have investigated the potential of the *luxS* gene product to act as an autoinducer cell-signalling molecule, and while it did induce bioluminescence in a *luxS* reporter strain, it did not have any discernable effect on toxin production. An alternative quorum sensing mechanism is the accessory gene regulator (Agr) system. For this pathway to function, an operon of four genes is required, expression of which directs the induction of further operons, including virulons (Figure 3.4). The bioinformatic search of the genome identified two genes from the Agr operon, CD2749A encoding AgrD, the autoinducer peptide and CD2750 encoding AgrB, the accessory gene regulator. Further searches of the genome identified CDSs with similarity to AgrA, a two-component response regulator (CD2603) and AgrC, a two-component histidine kinase (CD0338). However, it is unclear as to whether these gene products could co-ordinately function as the location of these genes on the genome are far apart.

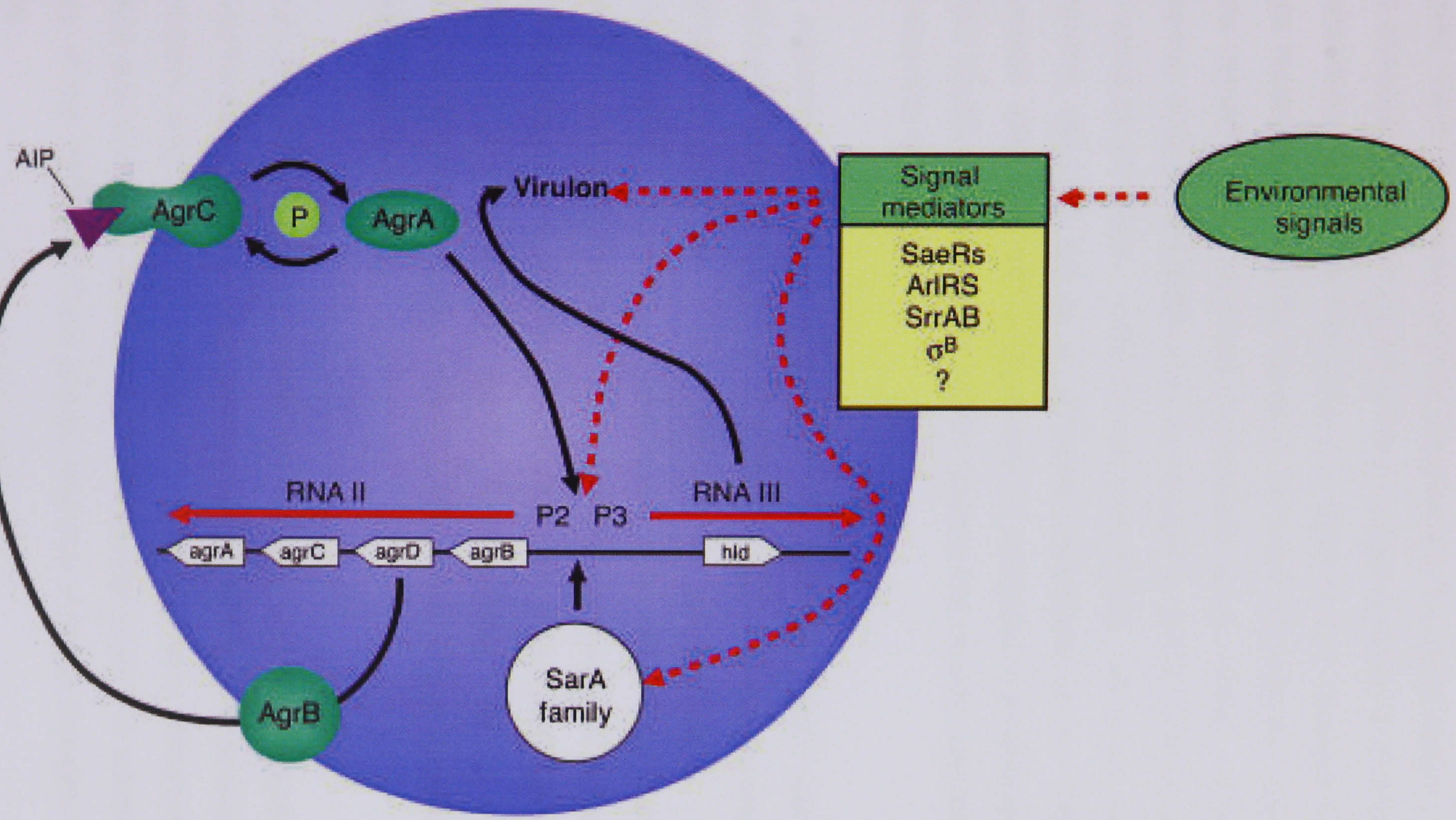


Figure 3.5 Mechanism of quorum sensing by the accessory gene regulator (Agr) system. The P2 operon encodes the signalling mechanism and the P3 operon acts as the effector in *S. aureus*. All P2 components are present in the *C. difficile* genome, but not organised as an operon. Black lines indicate interactions and dashed red lines indicate potential regulatory pathways from environmental signals. From Yarwood and Schlievert (2003).

3.2.4.13 Miscellaneous CDSs

Many CDSs identified by the bioinformatic analysis do not form part of a group of similar genes, and so will be discussed individually below.

Situated within the cluster of SLP paralogues is a gene called *mviN*. The protein this gene encodes has been shown to be a virulence factor in *Salmonella enterica* serovar Typhimurium. Deletion of this gene in *S. typhimurium* mutants reduces virulence in the mouse model of typhoid-like disease, although the specific mechanism by which this occurs remains to be elucidated (Carsiotis *et al.*, 1989). Another copy of a gene containing a MviN PF03023 motif is present in the *C. difficile* genome, however it is interrupted by an IStron. CD0194 encodes GroEL, which as well as functioning as a chaperone, can also act as an adhesin (Hennequin *et al.*, 2001b). GroEL is found associated with the cell surface and is released extracellularly after heat shock. Purified GroEL and antibodies raised against this protein are able to inhibit *C. difficile* attachment to Vero cells.

Two adjacent and highly similar (52.5% identical) CDSs were highlighted by the bioinfomatic search, CD0569 and CD0570. Both proteins exhibit similarity to several immunogenic proteins; TcsA (CD4+ T cell-stimulating antigen) of *Listeria monocytogenes*, BmpA (the immunogenic P39 antigen) of *Borrelia burgdorferi*, and TmpC of *Treponema pallidum*. The proteins are predicted to be attached to the lipid bilayer and contain Pfam group PF02608, which is highly antigenic when found in other proteins. The CD0569 gene is preceded by a polymeric tract of $(GTAAATAGGATGTAAAA)_{15}$, which could allow phase variation of the protein by slipped-strand mispairing.

A gene, CD1546, thought to encode a haemolysin-like membrane protein was found by the bioinformatic analysis. This gene is similar to the *C. acetobutylicum* and *B. cereus* haemolysin and contains motifs for secretion, lipid attachment, transmembrane domains and a mitochondrial energy transfer signature. In addition, CD1546 contains a copy of the PF03006 Pfam group, and this motif is thought to confer haemolysin activity.

A putative surface protein was identified with interesting homology. CD2717 is 24.12% identical to a *Plasmodium falciparum* mature parasite-infected erythrocyte surface antigen. This parasite protein is exported from the parasite to the erythrocyte membrane, where it binds to membrane skeletal proteins, and is thought to play a

major role in intraerythrocytic parasite viability (Magowan *et al.*, 1995). CD2717 contains one transmembrane repeat but no other motifs, and its function is unclear.

3.3 Discussion

To date, relatively little has been discovered about the arsenal of virulence factors employed by *C. difficile* to infect and colonise a host. In addition, the development of a vaccine against this pathogen does not appear to be imminent. The availability of the *C. difficile* chromosomal draft sequence and annotation has allowed the *in silico* identification of putative virulence factors and vaccine candidates selected from the complete protein repertoire.

This study represents the first genome-scale analysis of *C. difficile* putative virulence factors and vaccine candidates. This analysis has identified putative virulence factors, some of which may be considered vaccine candidates. Those proteins representing vaccine candidates are likely to be surface exposed and/or contain motifs characteristic of virulence factors or immunogens and/or are similar to documented virulence related proteins. From the whole genome of 3776 ORFs, 1170 were identified by a series of searches as possibly being involved in pathogenesis. The searches were designed to include as many possible putative virulence factors. Next, this large set of CDSs was subjected to a reductive strategy, in order to remove false positive hits. Careful manual curation of the CDSs resulted in 69% of the ORFs identified by the bioinformatic searches being categorised as non-virulent and these were removed from the list of putative virulence factors. Finally, this analysis has resulted in identification of 359 putative virulence/immunogenic genes, 90 of which have a putative function and 269 are unknown/hypothetical proteins.

The number of possible virulence factors and/or vaccine candidates identified from the *C. difficile* 630 genome is comparable to other bioinformatic screens of other pathogen genomes. The first published bioinformatic selection of vaccine candidates identified 570 ORFs from the genome of *N. meningitidis* (Pizza *et al.*, 2000), of which 350 were tested as protective antigens. Subsequent studies in *S. pneumoniae* and *C. pneumoniae* identified 130 and 133 putative vaccine candidates respectively (Wizemann *et al.*, 2001; Montigiani *et al.*, 2002). The selection criteria of these analyses were slightly different to this study, as the emphasis was on finding suitable vaccine candidates that could undergo high throughput cloning, expression and immunisation trials. Therefore these searches concentrated on surface components with few transmembrane domains, to maximise the chances of later expressing a

soluble antigen. In all of these studies, proteins have been identified that can induce a protective immune response in an animal model, and many are under further investigation as vaccine candidates, validating the usefulness of a bioinformatic approach.

The virulence genes identified from the *C. difficile* 630 genome with a putatively assigned function were grouped according to their function. The categories of virulence gene products are autolysins, flagella and chemotaxis, pili, drug resistance, lantibiotics, oligopeptide transporters, proteases, toxins, adhesins, sortase and sortase substrates, iron scavenging, quorum sensing and 'others'. However, assignment to a group is not unequivocal, as large proteins are often comprised of more than a single functional domain. For example some autolysins also contain protease domains and some sortase substrates also can also act as adhesins. Although representative proteins from most groups have been already been studied to some degree, many of the putative virulence factors described here were previously unknown in *C. difficile*. This analysis also complements and extends the existing knowledge in some areas of *C. difficile* pathogenesis, for example, there are many reports of flagellated *C. difficile* cells, but nothing has been reported on chemotaxis proteins. This analysis has confirmed the full repertoire of flagella-related proteins, and identified six chemotaxis related proteins. Furthermore, this study confirms some reports in the *C. difficile* literature, e.g. pili have only once been described in *C. difficile* cells (Borriello *et al.*, 1988), and this analysis has found a putative operon encoding type IV pili. The CDSs identified by these searches that do not have a known function may encode new virulence systems, and are still valid as potential vaccine candidates. This group of proteins are particularly worthy of further study.

Bioinformatics alone is unable to prove the expression and function of a protein. The limitations of bioinformatics are particularly acute in the cases where partial operons appear to be present, or genes for a particular function are scattered throughout the genome. An example of this is lantibiotic production and resistance. The *C. difficile* genome contains genes encoding one putative lantibiotic producing ORF (on a transposon), many lantibiotic ABC transporting systems and a few lantibiotic two-component signal-transducing systems. From this assortment of genes, it is unclear which lantibiotics *C. difficile* can sense and excrete. Another example is the Agr quorum sensing genes. The *C. difficile* genome does contain genes with similarity to each of the four genes required for this system to function. However these genes are widely dispersed over the chromosome and are not located within an operon, so it is unclear as to whether they can function in concert.

Although bioinformatics has been shown to be a powerful tool for the identification of new virulence factors and vaccine candidates, it clearly has limitations. The bioinformatic approach is able to circumvent some limitations of laboratory experiments (such as scale), resulting in the identification of gene candidates representing the complete gene repertoire of organism. For the judicious selection of virulence factors, it is desirable to combine the bioinformatic approach with functional genomics methods, such as microarray or proteomics, to confirm the expression of proteins in the expected cellular location, and to assess whether these proteins are conserved between strains. Bioinformatics also provides a valuable starting point for further investigations into the putative functions of potentially interesting CDSs.

The 364 gene products identified by this analysis as potential virulence factors and vaccine candidates will be of particular interest when compared to data in the following proteomics chapters. Confirmation of expression will lend credence to their status as virulence factors.

Chapter 4

The surface proteome of *Clostridium difficile* and structural studies of the surface layer proteins

4.1. Introduction

The repertoire of proteins a bacterium displays on its surface is critical for its ability to interact with its environment, and in the case of a pathogen, for its ability to cause infection. The cell surface has multiple roles including protection against the host's immune system, environmental sensing, nutrient uptake and attachment, all of which are important factors for survival. As such, the study of cell surface associated proteins is key to understanding the early stages of infection as the pathogen colonises the host.

Extracellular proteins can be defined as having been translocated across the cytoplasmic membrane. Upon exiting the cytoplasmic membrane, protein molecules encounter the highly cross-linked, semi-porous charged cell wall. Smaller protein molecules may be able to diffuse through, but the passage of larger proteins or those with many polar residues at their surface are likely to be affected by the molecular sieving or ion-exchanging properties of the cell wall. There are a number of mechanisms by which secreted proteins may be immobilised to the cell surface (section 1.2.2). The primary anchor-site for surface associated molecules, particularly in Gram-positive bacteria, is the cell wall. The cell wall is composed of peptidoglycan and anionic polymers (e.g. teichoic acid), which form a highly cross-linked and semi-porous structure, the composition and structure of which may vary with culture conditions. Functional proteins may be covalently attached to the peptidoglycan, or be more loosely associated with either the peptidoglycan or charged secondary wall polymers (discussed further in section 1.2).

External to the cell wall, *Clostridium difficile* has a surface layer (S-layer). This paracrystalline layer is composed of two proteins encoded by a single gene, *slpA*, (described section 1.3) (Calabi *et al.*, 2001). Briefly, *slpA* is post-translationally cleaved into two surface layer proteins (SLPs); a high molecular weight (HMW) protein and a low molecular weight (LMW) protein. Both SLPs lack any common cell wall binding domains, and it is currently unclear how, if at all, the S-layer is anchored to the cell wall. From freeze-etch electron microscopy it appears that there are two superimposed S-layers on surface of *C. difficile*; the outer S-layer is arranged in a square symmetry while the inner S-layer has a hexagonal symmetry (Cerquetti *et al.*, 2000). This suggests, but does not prove, the possibility that each S-layer is composed of a single protein species, either the HMW SLP or LMW SLP. Few structural data exist relating to the SLPs. Only one structure has been published to date (Ho *et al.*, 2005), but the heterogeneity between SLPs means that little can be inferred as to the structure of other SLPs, including those of *C. difficile*. The HMW SLP is highly conserved between

C. difficile strains and shows homology to *N*-acetyl muramoyl-L-alanine amidase from *B. subtilis*. The HMW SLPs can range in size between strains from 48–56 kDa and antibodies against them do cross-react with those from other strains. The HMW SLP has peptidoglycan hydrolase activity, shown by zones of degradation in a zymogram using *Micrococcus luteus* cell walls. The LMW SLP shows much greater sequence diversity between strains (Calabi *et al.*, 2001). The LMW SLPs range from 37–45 kDa depending on strain and show no antigenic cross-reaction between strains (McCoubrey & Poxton, 2001). As discussed in section 1.3, a family of 28 paralogues related to the HMW SLP are found in the genome of *C. difficile* strain 630, 11 of which are clustered around *slpA* (Figure 1.13). Each of these paralogues contains a domain with homology to the *N*-acetyl muramoyl-L-alanine amidase from *B. subtilis*, and most contain a second domain that varies between paralogues. These variable domains are not related to each other by sequence, but do show similarity between strains where studied, and are thought in some cases to give the protein its unique function (Calabi & Fairweather, 2002). Most paralogues also contain a signal sequence, and are thought to be exported from the cell, but like the SLPs, no common anchoring motif is present and it is unclear if or how they may be attached to the cell surface. Furthermore, the gene downstream from s*lpA* is one of two copies of *secA*, whose gene product is an essential component of the general secretory pathway.

To date, a number of cell-surface associated virulence factors involved in the adherence of *C. difficile* to the gut wall have been identified and characterised. The HMW SLP, SlpA paralogue 6 (also known as Cwp66), FliC (flagellin) and FliD (the flagellar cap protein) are all believed to be involved in the adhesion of *C. difficile* to the host gut epithelium, and are discussed further in section 1.3. These findings suggest that several adhesins may play a role in the adherence of *C. difficile* to the gut during an infection. However, it is difficult to prove that any of these proposed adhesins are essential for attachment and colonisation due to the lack of genetic tools available for use in *C. difficile*. Because of this, and because the cell surface has not been rationally investigated, it is feasible that many more adhesins exist on the cell surface that have not yet been discovered.

A valuable tool to dissect the complex array of proteins that are associated with the cell wall is proteomics: separation of the proteins using two-dimensional electrophoresis (2-DE) followed by mass spectrometry (MS) identification. This approach has been used in a number of studies recently, and the cell wall associated subproteomes mapped for *Escherichia coli* K-12 (Molloy *et al.*, 2000), *Staphylococcus aureus (*Nandakumar *et al.*, 2005), *Candida albicans* (Pitarch *et al.*, 2002), *Chlamydia pneumoniae* (Montigiani *et*

al., 2002) and the spore coat of *B. subtilis* and *Bacillus anthracis* (Lai *et al.*, 2003). Each of these studies revealed the presence of proteins on the cell surface that had not previously been demonstrated or predicted before, showing that proteomics gives a novel insight into the composition of the cell surface. Proteomics is providing information on not only the composition of sub-sections of the proteome, but also the relevance of the proteins found there. For example, Hughes and co-workers (2002) have mapped the major surface proteins of *Streptococcus agalactiae* and found that antibodies generated against two of these proteins were partially protective against *S. agalactiae* infection. In addition, the cell wall subproteome of *Listeria monocytogenes* has been mapped (Schaumburg *et al.*, 2004), revealing the expression of a number of genes with no orthologue in the non-pathogenic *Listeria innocua* genome, which therefore may represent virulence candidates.

The recent completion of the *C. difficile* 630 genome (Sebaihia *et al.*, 2006) has enabled the use of proteomics to investigate of the composition of the cell wall-associated subproteome in this bacterium. The aim of this study was to identify proteins present on the surface of the cell that may have a role in its virulence. The successful analysis of cell wall proteins is contingent on the development of robust surface protein extraction methods, and so the first aim was to optimise this procedure. Once optimised, two methods with complementary compositions were chosen for further analysis by proteomics. The process of separating the proteins using 2-DE was optimised, and many of the proteins were identified using MS, thus mapping the cell wall-associated subproteome. Establishing procedures to study the cell wall sub-proteome is a necessary precursor for further studies on the changing expression of surface proteins under different conditions that may be of relevance to the virulence status of *C. difficile*. As well as examining the composition of proteins associated with the cell surface, the structure of the predominant proteins in this location, the SLPs, was also investigated. To accomplish this, recombinant SLPs were expressed, purified and analysed using biochemical and biophysical methods.

4.2 Results: The surface proteome of *C. difficile*

4.2.1 Methods for extracting SLPs from *C. difficile*

There are a number of methods in the literature for the removal of S-layer and associated proteins from the surface of a bacterial cell (Sleytr, 1999; Cerquetti *et al.*, 2000; Calabi *et al.*, 2001). A number of different methods have been compared, and were found to remove one or both of the SLPs, with the concomitant release of varying numbers of associated proteins. Figure 4.1 shows 1D SDS-PAGE analysis of the

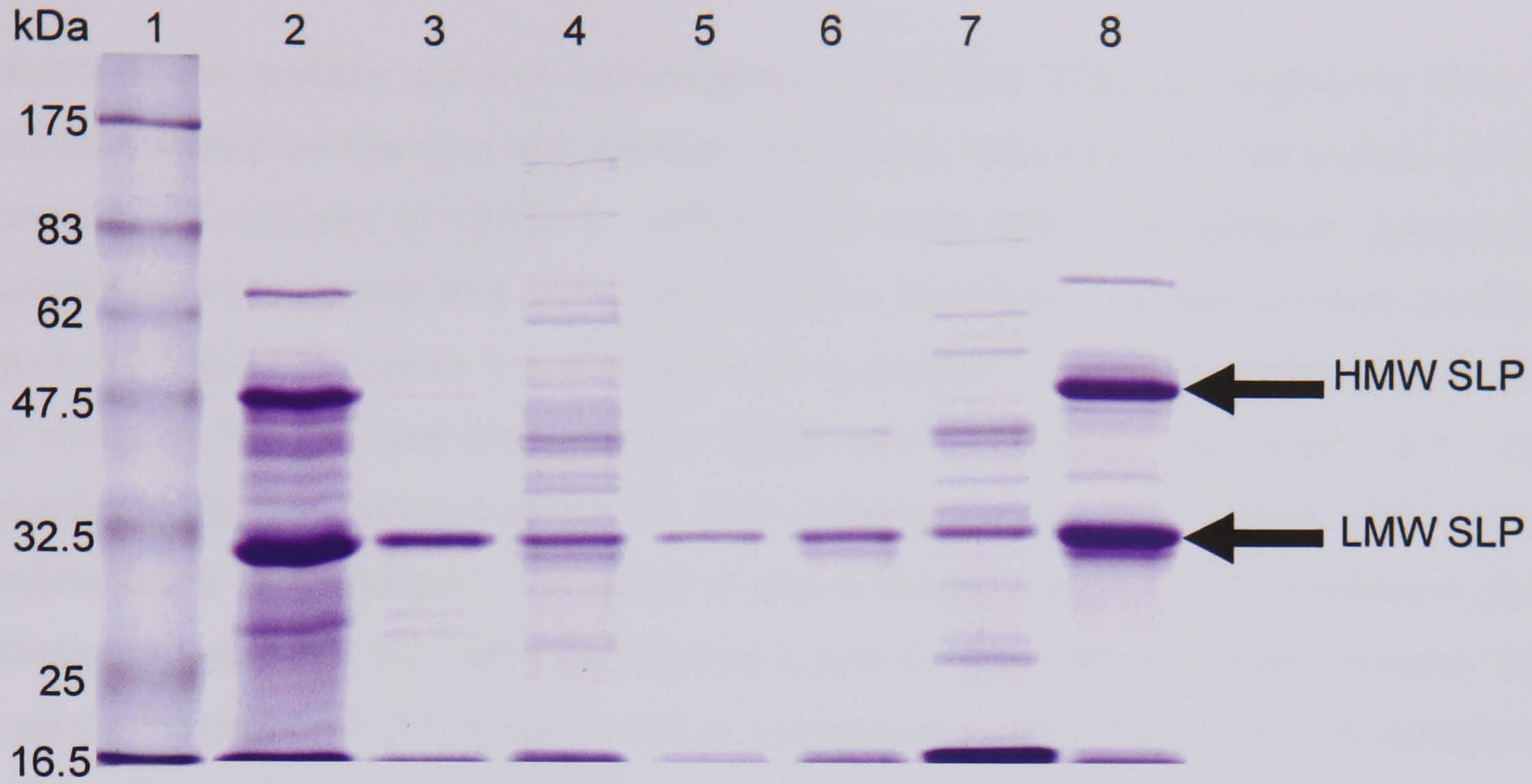


Figure 4.1 Extraction of S-layer and associated proteins from *C. difficile* 630 and analysis by SDS-PAGE. S-layers were extracted from an overnight culture of *C. difficile* 630 using: lane 2, 8 M urea; lane 3, 5 M LiCl; lane 4, 1 M LiCl; lane 5, 70 mM EDTA; lane 6, 10 mM EDTA; lane 7, lysozyme treatment and lane 8, 0.2 M glycine pH 2.2. Molecular masses in kDa are shown in lane 1. Arrows indicate the high (HMW) and low (LMW) molecular weight SLPs.

proteins removed by treatment of the whole bacteria with urea, LiCl, EDTA, lysozyme or low pH glycine as detailed in section 2.6. Most of these treatments act on the cell wall, but with different mechanisms of action, resulting in slightly different compositions for each extract.

Analysis of the surface extracts immediately shows that different treatments have a significant impact on the type and number of proteins released from the surface of the bacterium. Treatment of whole *C. difficile* 630 cells with urea releases apparently equimolar amounts of the two SLPs, along with a significant number of other proteins (Figure 4.1, lane 2). Lanes 3 and 4 show the extraction of surface proteins with LiCl. LiCl releases higher quantities of the LMW SLP relative to the HMW SLP, but interestingly higher concentrations of LiCl release much less protein than lower concentrations (5 M versus 1 M, lanes 3 and 4 respectively). EDTA treatment also preferentially releases the LMW SLP (lanes 5 and 6), and very few other proteins, but the concentration of EDTA appears not to dramatically affect the number of additional proteins released compared to LiCl. Lane 8 shows the effect of glycine treatment of whole cells. The two SLPs are released, along with a small number of additional proteins.

Because of the variations in the efficiency of these extraction methods, a new technique for removing all proteins from the surface of the cell was developed. The method builds on techniques optimised for the Gram-positive organisms *L. monocytogenes*, *Streptococcus pyogenes* and *S. aureus* (Schneewind *et al.*, 1992; Piard *et al.*, 1997; Jonquieres *et al.*, 1999). A basic schematic of this process is shown in Figure 4.2. The growth medium was separated from the bacterial cells by centrifugation and the secreted proteins were concentrated by trichloroacetic acid (TCA) precipitation. The cell pellet was washed and incubated in a digestion buffer containing enzymes that degrade the peptidoglycan wall. The cell wall was treated with three enzymes, lysozyme, mutanolysin and lysostaphin, each of which digests a specific bond within the peptidoglycan layer (Figure 4.3). The hypothesis was that this method should give a more defined cell wall extract than the procedures discussed above. Removal of the cell wall from an intact bacterium should result in the formation of protoplasts. These can be sub-fractionated further into membrane and cytosolic fractions by a simple freeze-thaw lysis followed by centrifugation. Samples of each fraction were run on SDS-PAGE, an example of which is Figure 4.4. The culture supernatant (Figure 4.4, lane 3) was found to contain a few faint protein bands, including one protein consistent in size with the LMW SLP. These proteins were not present in the control consisting of TCA precipitated brain heart infusion (BHI) medium

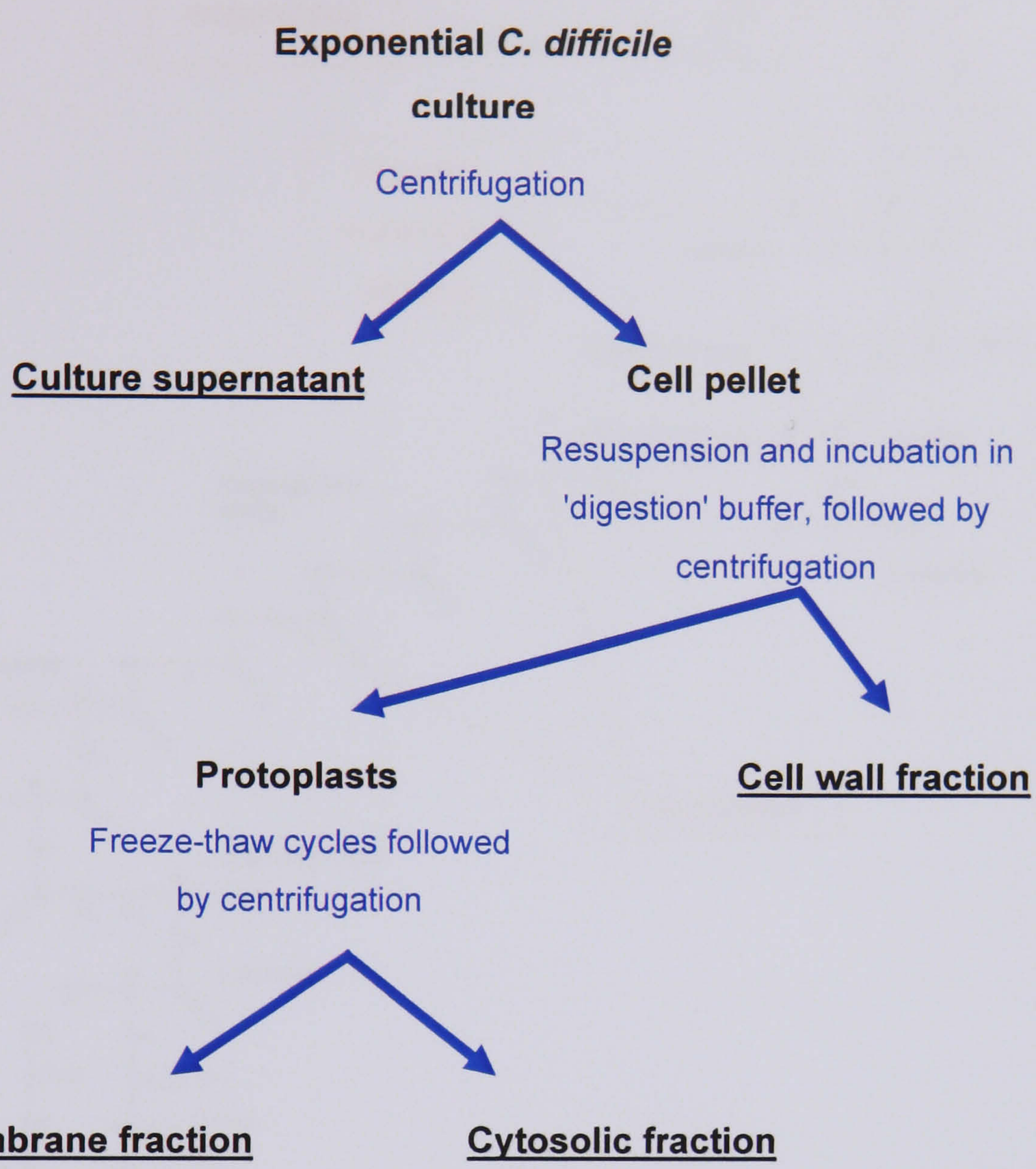


Figure 4.2 Schematic of cell fractionation procedure. The *C. difficile* culture is fractionated into separate components. Each underlined fraction is analysed by SDS-PAGE. Procedures are highlighted in blue, full details of which are described in section 2.6.

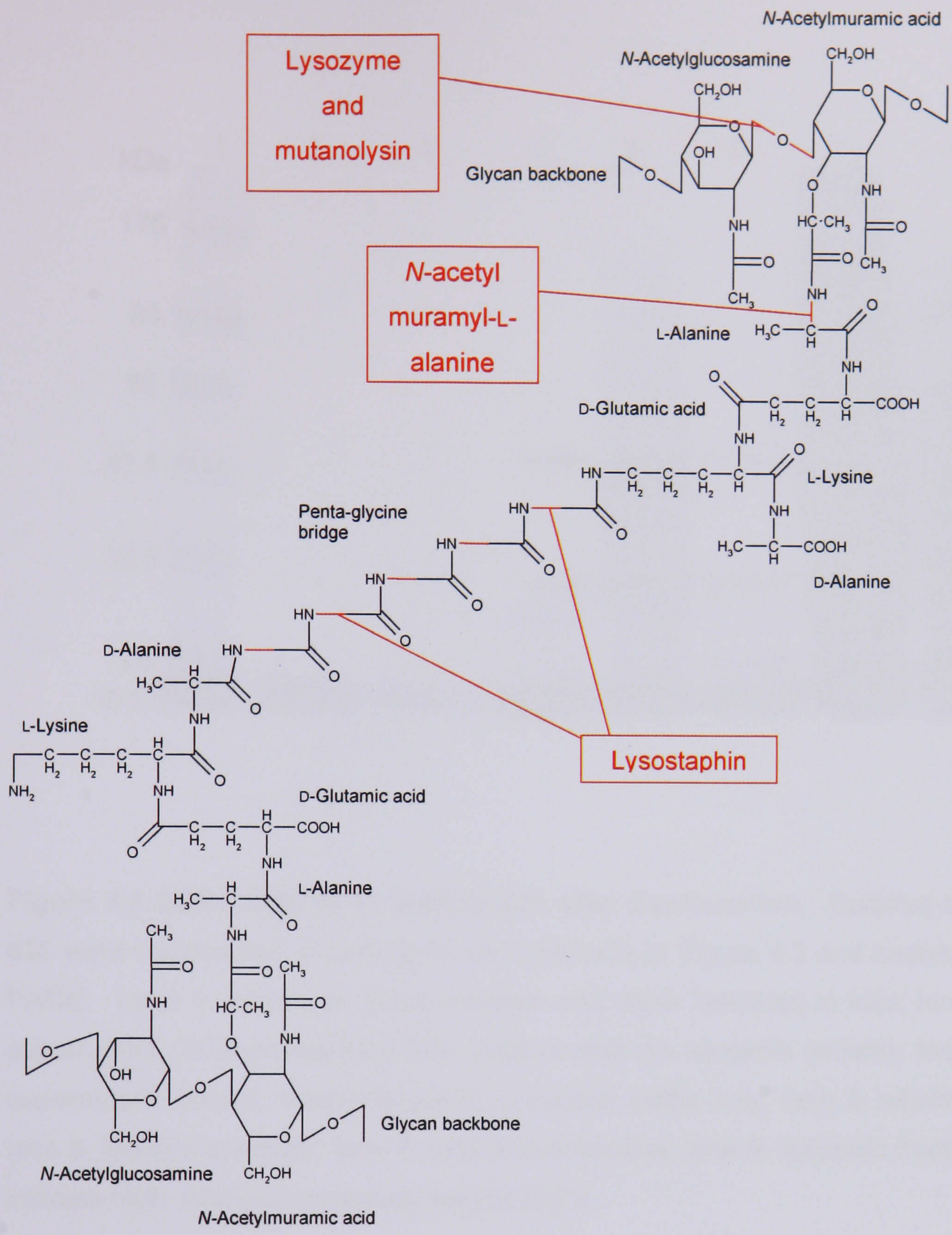


Figure 4.3 Structure of a typical Gram-positive peptidoglycan with cleavage sites for degradative enzymes indicated. Lysozyme and mutanolysin are muramidases that hydrolase the β-1,4 glycosidic bond between *N*-Ac-Mur and *N*-Ac-Glc. Mutanolysin cleaves lysozyme resistant bonds. Lysostaphin is a peptidase and hydrolyses the –Gly-|-Gly- bond in the pentaglycine interpeptide link (Hancock & Poxton, 1988). The HMW SLP shows homology to *N*-acetyl muramoyl-L-alanine amidase and shows peptidoglycan hydrolase activity by zymogram (Calabi *et al.*, 2001).

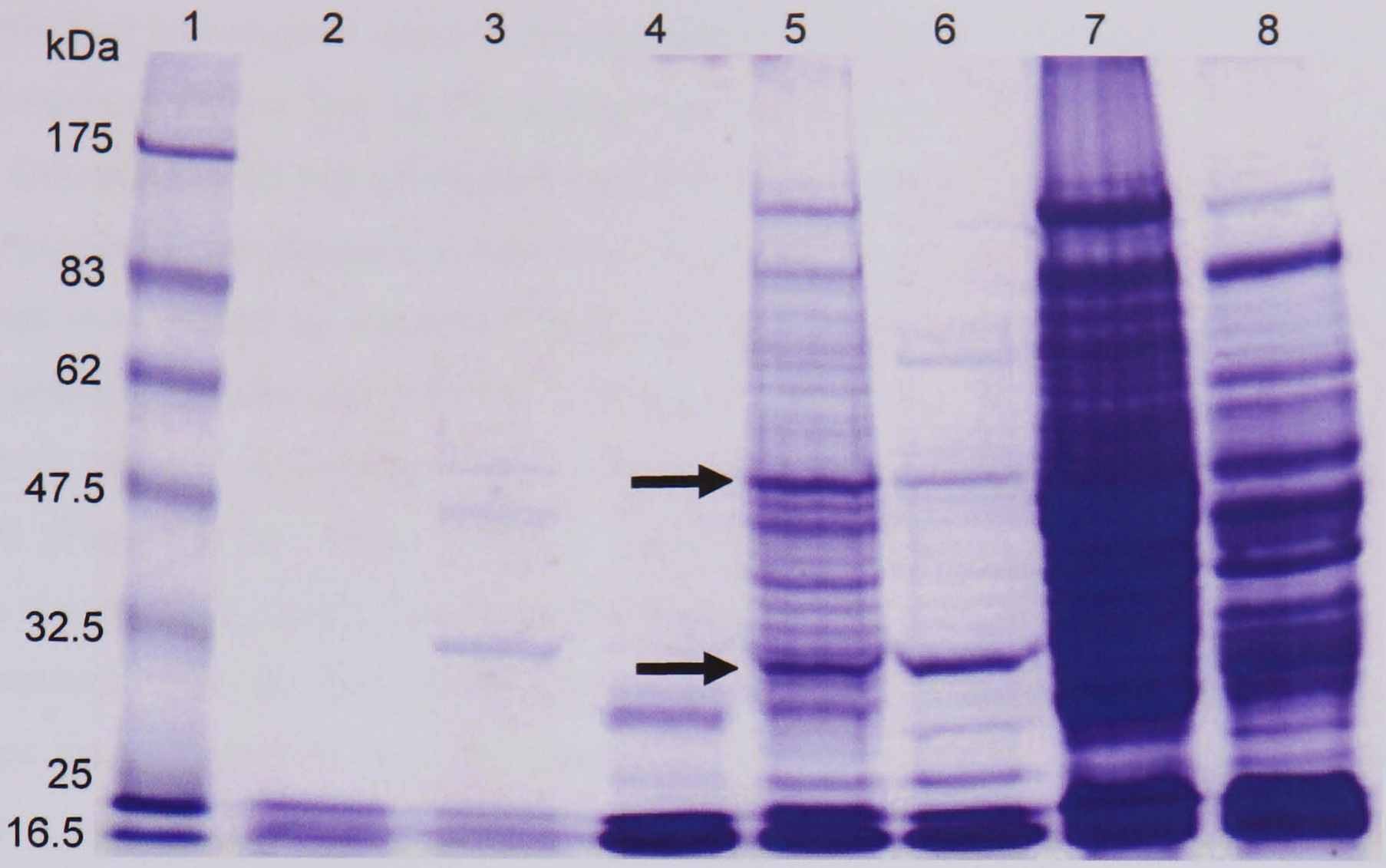


Figure 4.4 SDS-PAGE of *C. difficile* 630 after fractionation. Cultures of *C. difficile* 630 were fractionated according to the schematic in Figure 4.2 and analysed by SDS-PAGE. Lane 1, molecular mass markers with sizes indicated in kDa; lane 2, control supernatant (TCA precipitated BHI medium with no bacterial growth); lane 3, culture supernatant; lane 4, lysozyme-based extraction buffer only; lane 5, whole cell lysate; lane 6, lysozyme extract; lane 7, membrane fraction; lane 8, cytosolic fraction. Arrows indicate high- and low- molecular weight SLPs.

(Figure 4.4, lane 2). The lysozyme-based cell wall digestion buffer was found to contain a small number of low molecular weight proteins (Figure 4.4, lane 4), and bands of a similar size also appeared in the lysozyme extracted proteins lane (Figure 4.4, lane 6). These proteins presumably correspond to the lysozyme, RNAse, mutanolysin and lysostaphin used in the extraction procedure. The lysozyme extract is mainly comprised of the two SLPs, with a few higher molecular weight proteins also present. Comparison to the whole cell lysate in lane 5 shows how dominant the SLPs are in the bacterial complement of proteins. The cell membrane fractions contain many more bands than would be expected relative to the cytosolic fractions, suggesting that the lysis procedure was not optimal. This was confirmed in a number of different strains of *C. difficile* by microscopy, ensuring the phenomenon was not peculiar to strain 630 (Figure 4.5). Most bacterial cells appeared to have an intact cell wall, indicating that the enzymatic digestion of the cell wall by lysozyme was incomplete. If true protoplasts had not been made, then it is unlikely that the freeze-thaw cell lysis step would be sufficient to lyse the bacteria. Variations on the lysozyme digestion buffer were tested, employing using higher concentrations of enzymes and longer incubation times, but protoplasts were never generated. Anecdotal evidence suggests that it may not be possible to generate protoplasts of *C. difficile* (Peter Borriello, personal communication), and so the analysis of all fractions of the *C. difficile* proteome was not possible. Consequently only the cell wall lysozyme extract was investigated further. The lysozyme fraction (Figure 4.4, lane 6) contains both SLPs but greater quantities of the LMW SLP, as well as a number of further bands that putatively represent other cell wall associated proteins. Figure 4.1 compares this type of extract (lane 7) with the other extraction methods by SDS-PAGE. The lysozyme extraction method releases a considerable amount of LMW SLP, but relatively little HMW SLP. A significant number of protein bands are also present in the lysozyme extract, many of which are not seen in other extracts.

4.2.2 Cell lysis assays

Despite the evidence that the cell wall was not fully removed during the lysozyme cell wall digestion step, some of the additional bands present in the lysozyme extract may be due to cell lysis. This was investigated by a number of assays for intracellular enzymes, and comparing the activity of *C. difficile* sonicated lysates, lysozyme extracts and intact cells.

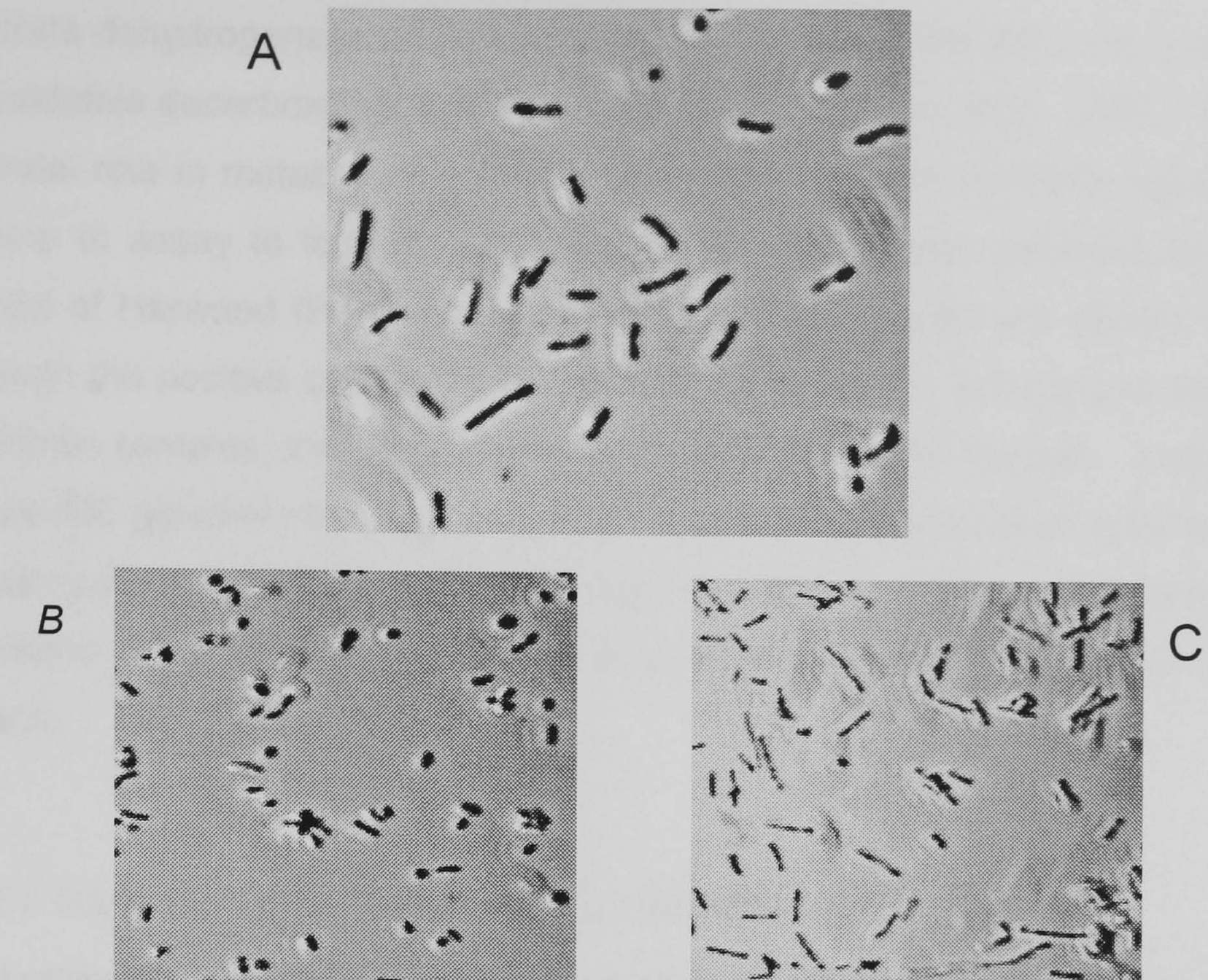


Figure 4.5 Strains of *C. difficile* after treatment with digestion buffer viewed under the microscope. **A**. Strain 630; **B**. Strain 1; **C**. Strain 17; **D**. Strain DE2; **E**. Strain Y. All bacterial cells appear to be intact after 2 hours incubation in lysozyme digestion buffer at 37 °C with rotating incubation. Magnification x 1000, phase contrast.

4.2.2.1 Isocitrate dehydrogensae activity

Isocitrate dehydrogenase (ICDH) is a key component of the citric acid cycle, catalysing the oxidative decarboxylation of isocitrate (Stryer, Biochemistry, 1995). ICDH plays as essential role in metabolism and is strictly cytosolic, and therefore represents a good enzyme to assay to look for cell lysis. This enzyme was assayed according to the method of Harwood (Harwood *et al.*, 1990) and the results are shown in Table 4.1A. Although the positive control worked well in this assay, no activity was detected for any *C. difficile* samples, including freshly sonicated *C. difficile* lysates. Analysis of the *C. difficile* 630 genome revealed the presence of a gene predicted to encode for isocitrate dehydrogenase, CD0834. Despite this, no activity could be detected under the conditions used, and so this method was not able to quantify cell lysis in the surface extracts.

4.2.2.2 Glyceraldehyde 3-phosphate dehydrogenase activity

Glyceraldehyde 3-phosphate dehydrogenase (G3PDH) is an important component of glycolysis, catalysing the conversion of glyceraldehyde 3-phosphate to 1, 3-bisphosphoglycerate exclusively in the cytoplasm. G3PDH was assayed by a method based on that of Ferdinand (1964), and modified by the laboratory of Dr. Peter Moody (University of Leicester). Activity of G3PDH was monitoring through the rate of production of nicotinamide adenine dinucleotide (NADH) at A_{340}. The results are shown in Table 4.1B. Over the 2 min reaction time, the positive control of commercially acquired G3PDH increased the absorbance of the reaction mixture by 0.16. All *C. difficile* fractions tested showed no change in absorbance under the same conditions. Analysis of the *C. difficile* 630 genome shows that two copies of G3PDH gene are present, CD1767 and CD3174. It remains unclear as to why this assay was not able to detect G3PDH activity in these *C. difficile* samples.

Although these assays have not quantified the level of cell lysis after extracting the surface proteins, the evidence from the microscopy and SDS-PAGE analysis of the fractionation procedure suggests that the levels of lysis are relatively low. Therefore it was decided to proceed with further analysis of the lysozyme extracted proteins.

4.2.3 Mapping the surface proteome

Two SLP extracts were chosen for further study by 2-DE. The low pH glycine extract was chosen as it removes both SLPs with just a small number of additional proteins. The lysozyme digestion method was also chosen as this should remove all proteins

A. Isocitrate dehydrogenase activity

	10 U Isocitrate dehydrogenase	Whole bacteria	Bacterial lysate	Lysozyme extract
t = 0	0.0	0.0	0.0	0.0
t = 10 min	0.5	0.0	0.0	0.0

B. Glyceraldehyde 3-phosphate dehydrogenase activity

	5 U Glyceraldehyde 3-phosphate dehydrogenase	Whole bacteria	Bacterial lysate	Lysozyme extract
t = 0	0.00	0.00	0.00	0.00
t = 2 min	0.16	0.00	0.00	0.00

Table 4.1 Assays for detection of cell lysis during surface protein extraction from *C. difficile* cells. *C. difficile* cells were grown to log phase and lysed using either sonication or lysozyme. Enzyme activities were measured in triplicate for each fraction using a published method (see Chapter 2). The change in absorbance over time was recorded in each assay. Isocitrate dehydrogenase activity was measured at A_{340}, and glyceraldehyde 3-phosphate dehydrogenase activity was measured at A_{340}. Absorbency readings were taken at the time points (t) indicated.

exterior to the cell wall in a methodical way and give a complete picture of the cell wall subproteome.

The proteins extracted from *C. difficile* 630 using glycine at pH 2.2 were separated by 2-DE, in the first dimension through an 11 cm non-linear gradient of pH 3-10, and in the second dimension by a 4-12% gradient SDS-polyacrylamide gel. Although these extracts typically contain predominantly the two SLPs and relatively few additional bands on 1-D SDS-PAGE, when separated by 2-DE, a number of further proteins co-extracted with the SLPs can be seen (Figure 4.6). Silver staining revealed over 50 protein spots in this extract. There was a concentration of proteins in the low pI range (approximately 4-5) with a dense region of proteins around the two SLPs, and few spots in the higher pI range.

The lysozyme cell wall extract was found to be composed of over 100 unique spots when silver stained (Figure 4.7). This treatment released approximately twice as many proteins as the glycine treatment, reflecting the variety of proteins associated with the cell wall. Again, the majority of spots were concentrated in the low pI range, but with a number of small basic proteins, identified mainly as ribonuclease and lysozyme which were used as part of the extraction method

Initially, Coomassie-stained spots from glycine extracts run on 7 cm mini-format 2-DE gels were analysed by MALDI-TOF as described in 2.8. Based on abundance, major spots on these gels were individually excised, digested with trypsin and subjected to MALDI-TOF MS. Comparison of the peptide profiles against the non-redundant database and *in silico* digests of the partially finished genome coupled with BLAST analysis (www.ncbi.nlm.nih.gov/BLAST/) (section 2.9) allowed the identification of protein spots. The spots identified by this method are shown in Figure 4.8 and full details of the protein matches are given in Table 4.2. The peptide fragments from spot 1 matched those from theoretical digests of SlpA in the non-redundant and *C. difficile* genome databases. On further study, these peptides matched specifically to the carboxyl-terminal portion of the protein, which is the HMW SLP after post-translational processing. The peptides from spot 2 mapped to the amino-terminus of SlpA, corresponding to the LMW SLP after post-translational cleavage. Unsurprisingly, these are the dominant proteins of the extract. Figure 4.9 shows the mass spectra for the HMW SLP and LMW SLP and the coverage of the peptides over the protein sequence. Spot 4 also contained peptides from SlpA, but just to one portion of the LMW SLP, suggesting that this is a degradation product. Spot 3 peptides matched a putative phospholipase. Phospholipases have two principal roles, as digestive enzymes or in

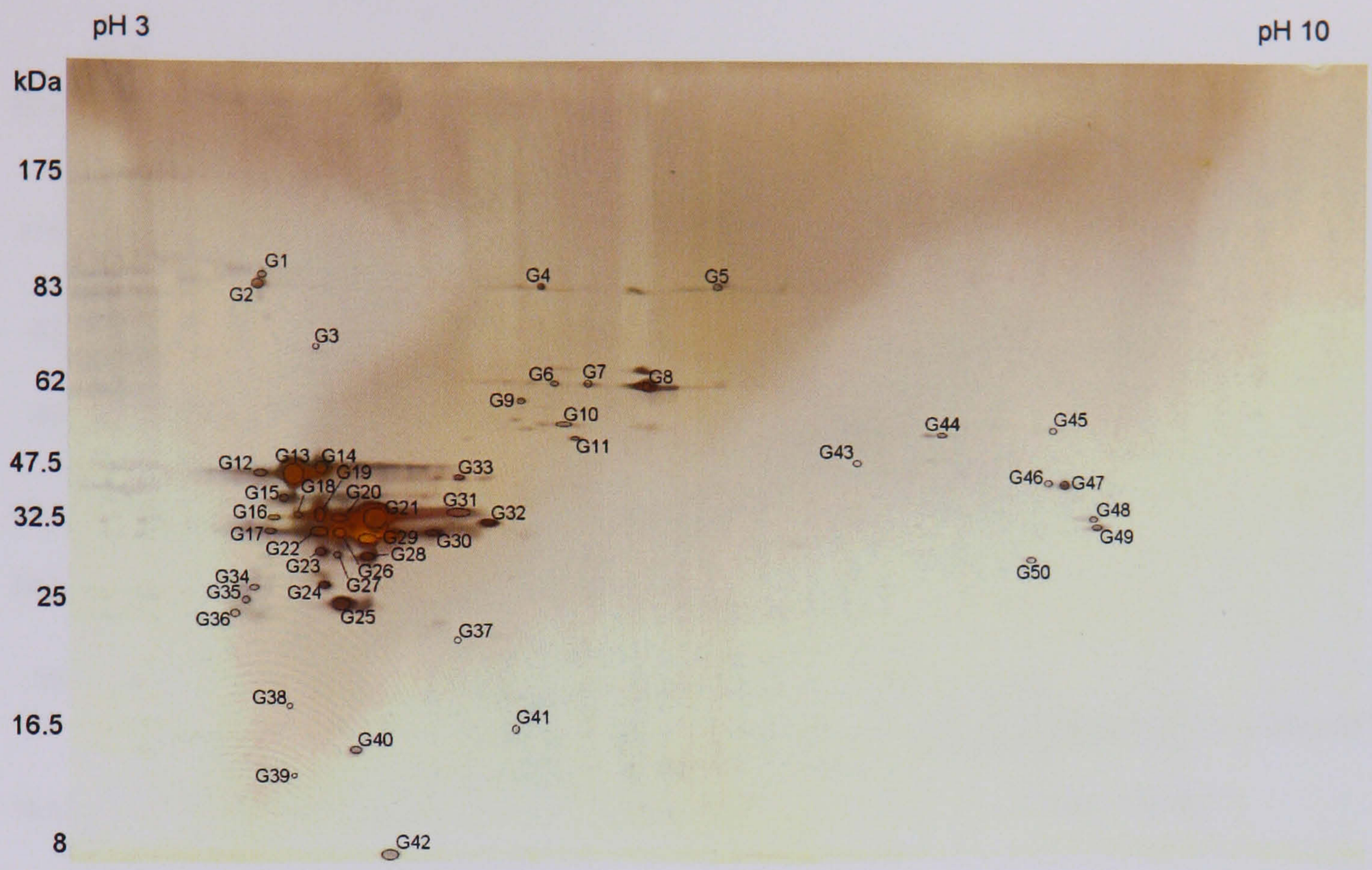


Figure 4.6 2-DE of SLPs and associated proteins extracted from *C. difficile* by glycine treatment. Proteins were extracted from *C. difficile* 630, grown at 37°C for 24 h in BHI medium, by treatment with glycine pH 2.2. 50 μg protein extract was resolved on 2-DE and visualised by silver staining. Fifty spots were identified by MS and are numbered G1-G50 (see Table 4.3).



Figure 4.7 2-DE of proteins extracted by lysozyme treatment. Proteins were extracted from *C. difficile* 630, grown at 37 °C for 24 h in BHI medium, by treatment with lysozyme, mutanolysin, lysostaphin and RNAseA. 50 μg protein extract was resolved on 2-DE and visualised by silver staining. Sixty-nine spots were identified by MS and are numbered L1-L69 (see Table 4.4).

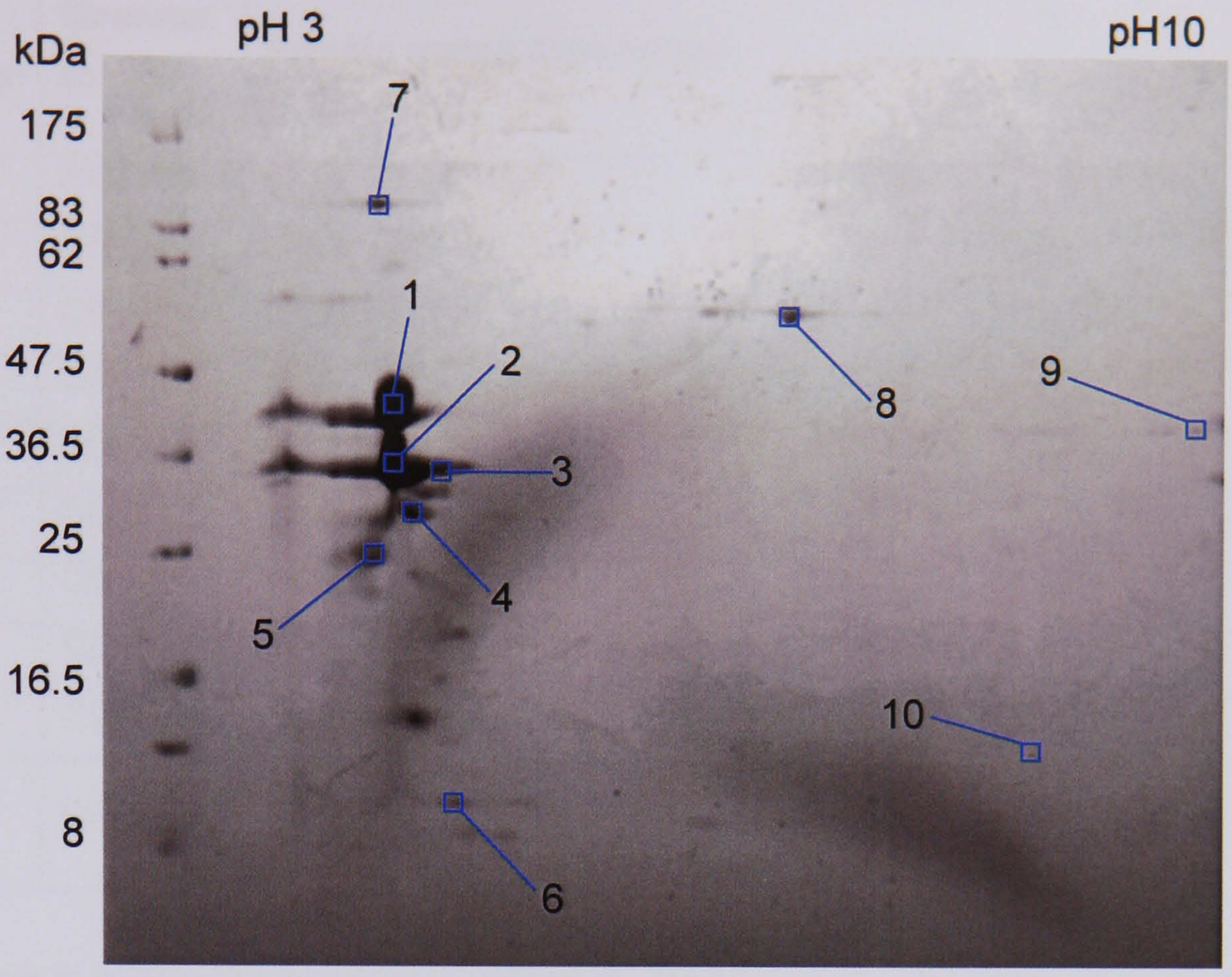


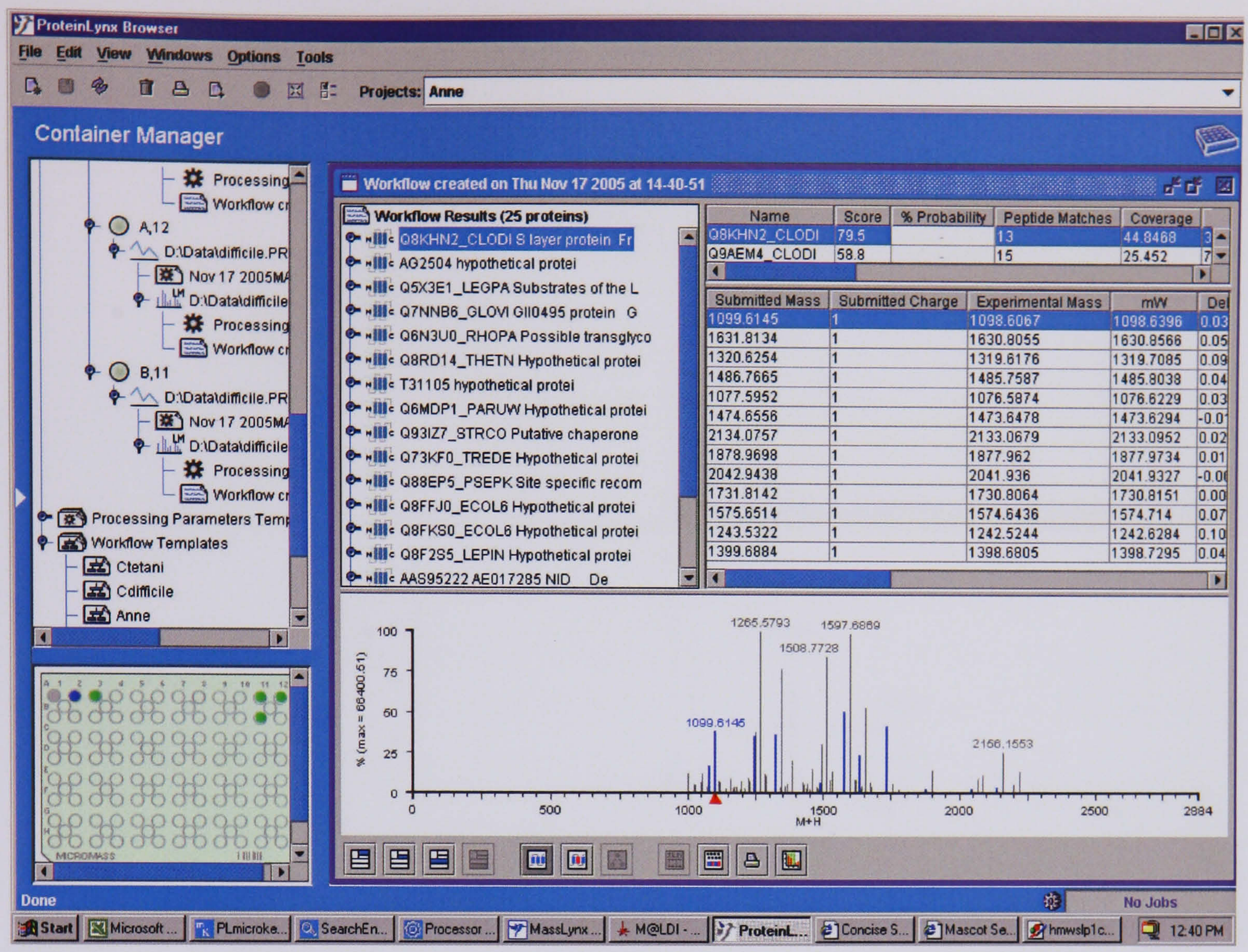
Figure 4.8 2-DE and identification of protein spots extracted from *C. difficile* by glycine treatment. Proteins were extracted from *C. difficile* 630, grown at 37 °C for 24 h in BHI medium, by treatment with 0.2 M glycine pH 2.2. Approximately 25 μg was resolved by 2-DE and visualised by staining with Coomassie blue. Molecular masses in kDa are indicated on the left.

Six spots were identified by MALDI-TOF as: spot 1, C-terminal SlpA; spot 2, N-terminal SlpA; spot 3, putative phospholipase; spot 4, SlpA; spot 5, FliC (flagellum subunit); spot 6, conserved hypothetical protein; spot 7, putative exported protein; spot 8, putative signalling protein; spot 9, Slp paralogue ORF 8; spot 10, ABC transporter, ATP binding protein. Details of protein matches are shown in Table 4.2.

Spot number	Genome ID	Protein match description	Protein probe score	% coverage	Number of peptide matches
1	CD2793	SlpA, HMW SLP	8.2	21.3	11
2	CD2793	SlpA, LMW SLP	9.5	44.8	13
3	CD2799	Putative phospholipase	7.6	29.3	13
4	CD2793	SlpA, LMW SLP	8.1	17.8	9
5	CD0239	FliC, flagellin	7.1	34.8	11
6	CD3617	Conserved hypothetical	4.1	17.2	7
7	CD1135	Putative exported protein	4.6	13.8	8
8	CD1015	Putative signalling protein	4.0	7.4	4
9	CD2799	SLP paralogue 8	5.1	18.1	9
10	CD0785	ABC transporter ATP-binding protein	5.8	6.1	3

Table 4.2 Summary of results of MALDI-TOF protein identifications from proteins extracted from *C. difficile* by glycine treatment. Spot numbers correspond to Figure 4.8. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Protein Probe score determined by Protein Lynx Global Server (Micromass).

A

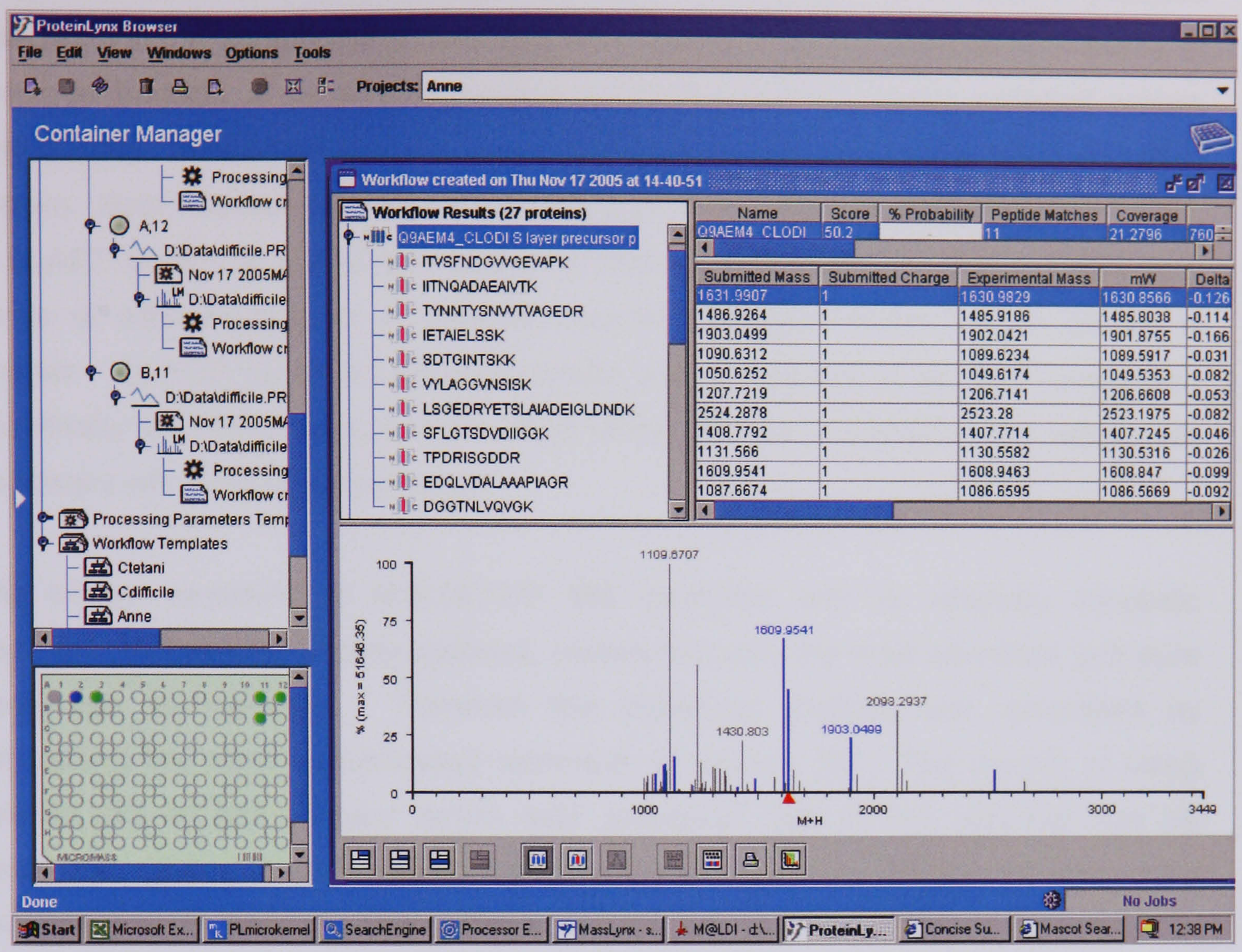


B



Figure 4.9 Examples of MS spectra and protein coverage of the LMW SLP and HMW SLPs. Screen shots from Protein Lynx Global Server showing mass spectra and protein coverage for the SLPs. **A.** Screen shot showing the spectrum obtained from spot 1 (Figure 4.8) and the peptides that match the LMW SLP. **B.** Coverage of peptides over the LMW SLP protein sequence. A large number of peptides match and cover a significant proportion of the protein sequence indicating a high degree of confidence in the match. Continued overleaf

C



D

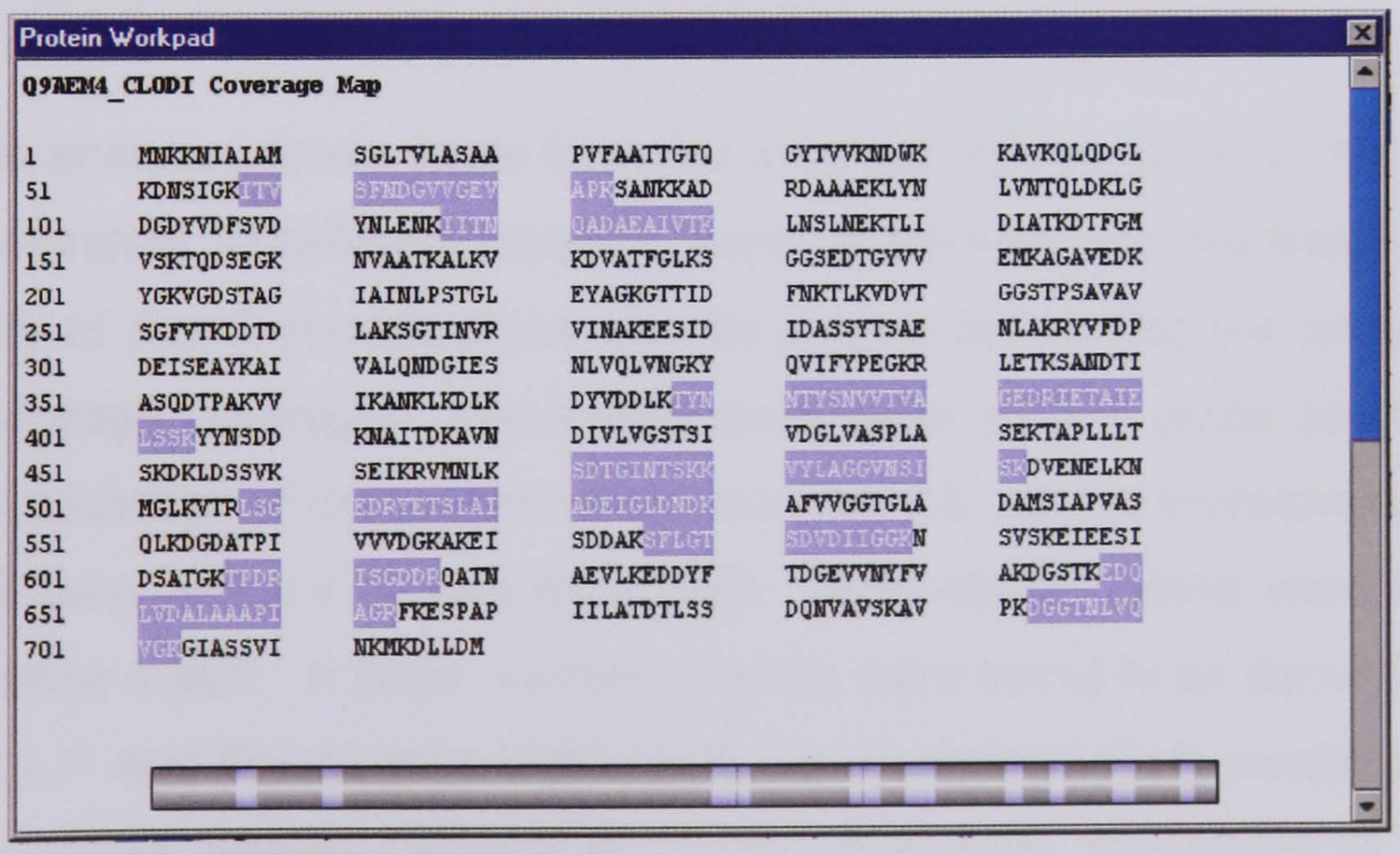


Figure 4.9 continued. **C.** Screen shot showing the spectrum from spot 2 (Figure 4.8) and the peptides that match to the HMW SLP. **D.** Coverage of peptides over the HMW SLP sequence.

the generation of signal molecules. They preferentially hydrolyse substrates located in bilayer membranes, micelles or lipoprotein particles (Stryer, 1997). Spot 5 peptides matched FliC, a component of the flagellum. Spots 6 and 7 matched to proteins of unknown function, a conserved hypothetical protein and a putative exported protein respectively. Spot 8 had a peptide mass fingerprint consistent with a putative signalling protein. Spot 9 showed highest similarity to the SlpA paralogue ORF 8 (Figure 1.13B). A BLAST search of the unique domain of this paralogue showed low similarity to a range of different families of proteins, and the putative function of this protein is unclear. Spot 10 contained peptides similar to a component of an ABC transporter, specifically an ATP-binding protein. Most of the proteins identified in this extract are consistent with a surface localisation.

The limited sensitivity of MALDI-TOF MS combined with its relatively simplistic approach of matching peptide patterns, means that only the most abundant and pure spots can be identified. Therefore this proteomic analysis was developed by introducing the more sophisticated technique of tandem MS. The benefit of using tandem MS is that primary amino acid sequence data of the peptides can be generated, giving a very reliable identification of proteins. MS/MS is also more sensitive and multiple proteins within a single spot can be confidently identified. In collaboration with Dr. Robin Wait (Kennedy Institute of Rheumatology, Imperial College London), protein spots were analysed by tandem MS, using a Q-Tof hybrid quadrupole/orthogonal acceleration TOF spectrometer.

Using MS/MS to analyse protein spots from the resolved glycine extracts, the identities of 50 spots, comprising 12 different proteins, were determined and are shown in Table 4.3. The quality of these identifications can be judged by viewing the protein probe score and percentage coverage. The lowest acceptable protein probe score was 18 and percentage coverage of the protein sequence was 3.5, as this corresponded to the protein with the least amount of sequence data. Most identifications were based on more than 1 peptide match. A large number of spots were found to be derived from the 45 kDa HMW SLP and the 36 kDa LMW SLP. A number of SlpA paralogues were detected in this extract, namely ORFs 2, 6, 12, 15, 24 and 25. In contrast to SlpA, the paralogues 2, 6, 12, 24, and 25 were all visualised as spots of molecular mass close to that predicted for the mature protein after removal of the signal peptide, indicating that a second internal cleavage had not taken place. Interestingly, paralogue 15, which has a predicted molecular mass of 167 kDa was visualised as a series of spots of molecular mass between 47 kDa and 85 kDa. Paralogue 15 contains a cell wall binding domain linked to a series of 9 peptide repeats each of 120 residues. This

Table 4.3 Proteins identified by MS/MS from surface proteins extracted from *C. difficile* 630 by glycine treatment. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Protein Probe score determined by Protein Lynx Global Server (Micromass). Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length protein and mass of mature protein in parentheses. Signal sequences predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Protein Probe score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
G1	CD2793	LMW SLP	50.81	21.16	3 (73)	76115 (34228)	4.9	√
G2	CD2793	HMW SLP	85.85	24.33	4 (82)	76115 (39504)	4.6	√
	CD0514	SLP paralogue 15	92.66	13.68	4 (69)	167580	4.7	√
G3	CD2793	LMW SLP	39.26	13.62	2 (47)	76115 (34228)	4.9	√
G4	CD0514	SLP paralogue 15	53.60	11.59	2 (35)	167580	4.7	√
G5	CD0514	SLP paralogue 15	111.39	13.44	4 (57)	167580	4.7	√
	CD2793	HMW SLP	40.52	20.32	3 (76)	76115 (39504)	4.6	√
G6	CD2791	SLP paralogue 2	140.79	13.97	6 (87)	66427	6.4	√
G7	CD2791	SLP paralogue 2	59.64	7.06	3 (44)	66427	6.4	√
G8	CD2791	SLP paralogue 2	199.60	20.39	8 (126)	66427	6.4	√
G9	CD2784	SLP paralogue 6	43.28	6.22	2 (42)	73024	7.2	√
G10	CD2784	SLP paralogue 6	18.56	4.29	2 (29)	73024	7.2	√
	CD2793	HMW SLP	33.62	12.83	2 (48)	76115 (39504)	4.6	√
G11	CD2793	HMW SLP	66.78	16.84	3 (63)	76115 (39504)	4.6	√
G12	CD2793	HMW SLP	244.52	51.60	10 (193)	76115 (39504)	4.6	√

Table 4.3 continued

Spot	Genome ID	Description	Protein Probe score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
G13	CD2793	HMW SLP	325.02	55.88	14 (212)	76115 (39504)	4.6	√
G14	CD2793	HMW SLP	230.32	48.13	10 (180)	76115 (39504)	4.6	√
G15	CD2793	HMW SLP	309.36	41.71	11 (150)	76115 (39504)	4.6	√
G16	CD2793	HMW SLP	142.81	24.33	6 (91)	76115 (39504)	4.6	√
G17	CD2793	LMW SLP	906	65.11	11 (209)	76115 (34228)	4.9	√
G18	CD2793	SlpA	378.93	35.33	15 (255)	76115	4.8	√
G19	CD2793	LMW SLP	157.79	29.86	8 (103)	76115 (34228)	4.9	√
G20	CD2793	LMW SLP	2983.63	52.75	12 (182)	76115 (34228)	4.9	√
G21	CD0239	FliC, flagellin	138.29	13.10	4 (38)	30755	8.6	×
	CD2793	LMW SLP	424.16	62.03	14 (214)	76115 (34228)	4.9	√
G22	CD0239	FliC, flagellin	76.16	18.62	3 (54)	30755	8.6	×
	CD2793	LMW SLP	232.43	28.70	9 (99)	76115 (34228)	4.9	√
G23	CD2793	LMW SLP	68.13	17.97	3 (62)	76115 (34228)	4.9	√
	CD0239	FliC, flagellin	59.39	22.76	3 (66)	30755	8.6	×
G24	CD0239	FliC, flagellin	132.38	15.77	3 (49)	30755	8.6	×
	CD2793	HMW SLP	240.14	32.62	10 (121)	76115 (39504)	4.6	√
G25	CD2793	HMW SLP	352.01	34.22	11 (116)	76115 (39504)	4.6	√
G26	CD2793	LMW SLP	89.95	15.36	4 (53)	76115 (34228)	4.9	√
	CD0239	FliC, flagellin	193.92	33.79	7 (98)	30755	8.6	×
G27	CD0844	SLP paralogue 25	40.61	14.38	3 (45)	33776	5.2	√

Table 4.3 continued

Spot	Genome ID	Description	Protein Probe score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
G28	CD0239	FliC, flagellin	799.40	28.62	5 (83)	30755	8.6	×
	CD2793	LMW SLP	61.00	21.45	4 (74)	76115 (34228)	4.9	√
	CD0844	SLP paralogue 25	70.06	14.38	3 (45)	33776	5.2	√
G29	CD2793	LMW SLP	3571.71	55.65	12 (192)	76115 (34228)	4.9	√
	CD0239	FliC, flagellin	105.74	8.48	5 (61)	30755	8.6	×
G30	CD0239	FliC, flagellin	86.52	26.55	4 (52)	30755	8.6	×
G31	CD2793	LMW SLP	232.78	34.49	8 (119)	76115 (34228)	4.9	√
	CD0239	FliC, flagellin	117.35	23.10	4 (67)	30755	8.6	×
G32	CD0255	FlgE, flagellar hook protein	62.17	10.09	2 (33)	34571	5.2	×
G33	CD2793	HMW SLP	225.00	38.23	9 (143)	76115 (39504)	4.6	√
G34	CD2793	LMW SLP	75.80	17.39	3 (60)	76115 (34228)	4.9	√
G35	CD2793	HMW SLP	232.72	33.16	8 (124)	76115 (39504)	4.6	√
G36	CD2793	HMW SLP	75.80	8.34	3 (60)	76115 (39504)	4.6	√
G37	CD2793	HMW SLP	107.24	25.13	6 (94)	76115 (39504)	4.6	√
G38	CD1280	NifU, nitrogen fixation /Fe-S cluster	33.81	8.90	1 (13)	15853	4.5	×
G39	CD2793	HMW SLP	109.02	19.25	4 (72)	76115 (39504)	4.6	√
G40	CD2793	HMW SLP	53.70	6.95	2 (26)	76115 (39504)	4.6	√
G41	CD2793	HMW SLP	66.76	12.83	3 (48)	76115 (39504)	4.6	√
G42	CD2793	HMW SLP	35.09	12.30	2 (46)	76115 (39504)	4.6	√
G43	CD2793	LMW SLP	870.45	21.74	3 (75)	76115 (34228)	4.9	√
G44	CD2794	SLP paralogue 12	59.26	12.10	2 (64)	57586	8.8	√

Table 4.3 continued

Spot	Genome ID	Description	Protein Probe score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
G45	CD0514	SLP paralogue 15	21.97	10.80	1 (22)	167580	4.7	√
G46	CD2793	LMW SLP	177.27	7.54	1 (26)	76115 (34228)	4.9	√
G47	CD0514	SLP paralogue 15	129.52	21.18	5 (68)	167580	4.7	√
G58	CD2793	LMW SLP	914.43	27.54	4 (95)	76115 (34228)	4.9	√
G49	CD1156	Putative exported protein	55.01	5.0	2 (18)	39277	9.4	√
G50	CD2193	SLP paralogue 24	33.34	4.41	1 (13)	31501	7.5	√

repeat region shows homology to haemolysins from bacterial species such as *Shewanella amazonensis* (E = $9e^{-32}$), *Ralstonia solanacearum* (E = $6e^{-31}$) and *Burkholderia cenocepacia* (E = $1e^{-99}$) when analysed by BLAST. The 47 kDa, high pI fragment contained peptides corresponding to the cell wall binding domain, whereas the remaining higher molecular mass spots contain peptides from the peptide repeat region and not the cell wall binding domain (Figure 4.10). The variable domains of several other proteins detected in this extract show no significant homology to other known proteins (by BLAST analysis), and so their putative functions are unknown. This group includes SLP paralogues 2, 6, and 25. The variable domain of Slp paralogue 12 contains a COG2340 domain. This domain belongs to a family of uncharacterised proteins containing SCP/PR1 motifs, and the function of this group of proteins is unknown. BLAST analysis of the Slp paralogue 24 variable domain shows that it has homology to proteins with a cell wall hydrolase function, for example from *Clostridium beijerincki* (E = $5.3e^{-37}$). In addition to the Slp paralogues, two components of the flagellum were identified in this extract, FliC, the flagellin subunit and FlgE, the flagellar hook protein. Two further proteins were also found; a putative exported protein and NifU, a molecular scaffold for Fe-S clusters.

The lysozyme cell wall extract was found to be comprised of 100 spots when resolved by 2-DE and visualised with silver staining (Figure 4.7). This treatment released approximately twice as many proteins as the glycine treatment. Again, the majority of spots were concentrated in the low pI range, but with a number of small basic proteins, mainly ribonuclease and lysozyme, which were used in the extraction method. A total of 69 protein spots, comprising 45 different proteins, were identified by comparison with the translated genome, and are shown in Table 4.4. The two SLPs are predominant in this extract, although the HMW protein is less abundant than the LMW protein. A number of paralogues were also found, namely ORFs 2, 4, 15 and 25. Once again, Slp paralogue 15 appears in several distinct spots (Figure 4.7, spots L1, L2 and L3). Slp paralogue 4 was detected in this extract (spot L66), although with a much smaller molecular mass and a more basic pI than would be predicted for the full length sequence. Only 1 peptide from this protein underwent a second round of MS, resulting in sequence data, and this peptide was from the cell wall binding domain of the protein. The variable domain of ORF 4 has high sequence identity (30%) to cysteine proteases, and is discussed further in chapter 7. In addition, a putative cold shock protein (CD1335) was found, and the heat shock protein DnaK. Several ABC transporter proteins were detected, including AppA, an oligopeptide substrate binding protein, OppA, a second oligopeptide-binding transport protein (CD0855) and a substrate-binding lipoprotein (CD0873). Most of the proteins identified are predicted to contain a

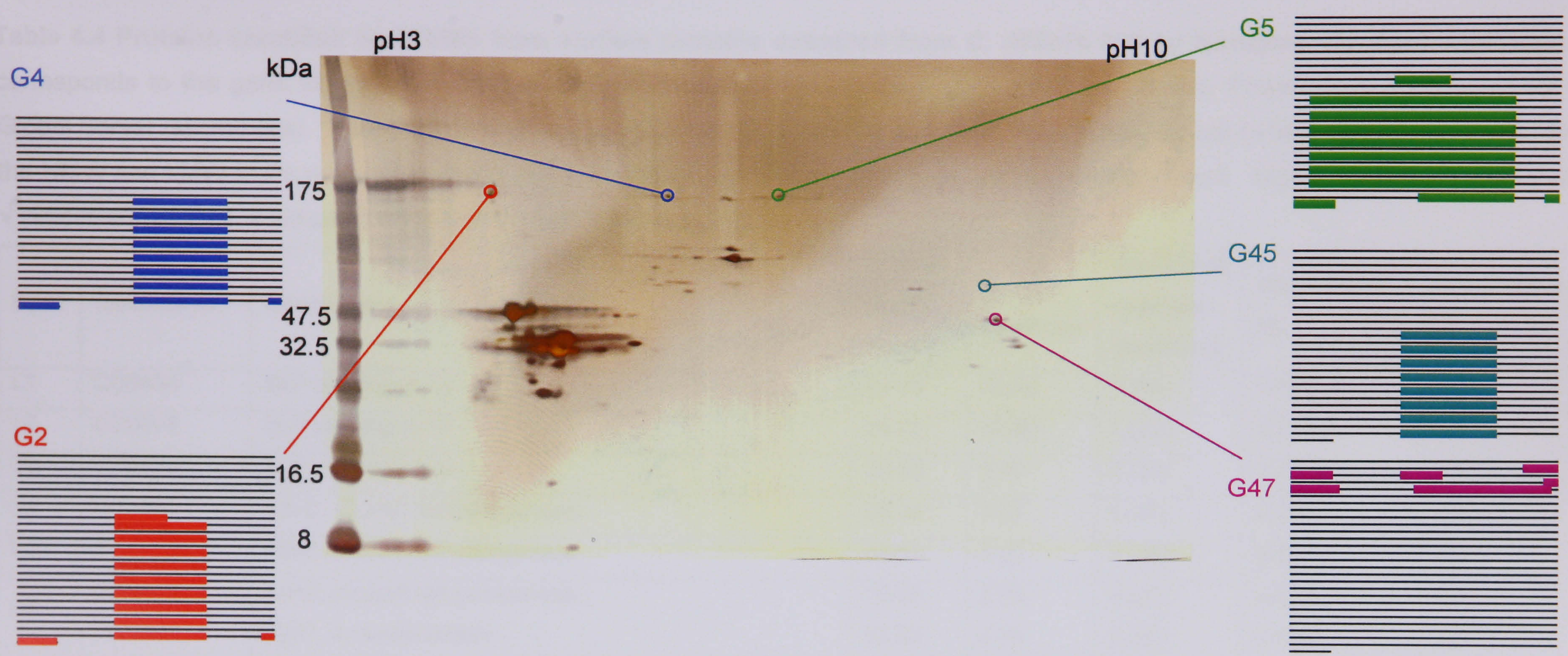


Figure 4.10 Position and peptide coverage of SLP paralogue 15 on 2-DE gel of proteins extracted from *C. diffcile* by glycine treatment. The main panel shows a 2D gel of proteins extracted from *C. difficile* by glycine treatment. Spots identified by MS/MS as SLP paralogue 15 are circled and numbered corresponding to Figure 4.9 and Table 4.3. The five smaller figures give diagrammatical representations of the protein sequence of SLP paralogue 15, with the NH_2 terminus at the top left. Highlighted regions show peptide coverage. Due to the repetitive nature of this protein sequence, a sequenced peptide could come from any (or several) of the repeat sequences and repeating highlighted bars indicate this. Molecular mass markers are indicated on the left in kDa and pH is indicated above the gel.

Table 4.4 Proteins identified by MS/MS from surface proteins extracted from *C. difficile* 630 by lysozyme digestion treatment. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Protein Probe score determined by Protein Lynx Global Server (Micromass). Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Protein probe score	% coverage	Number of peptides (residues)	Theoretical M_r	Theoretical pI	Signal sequence
L1	CD0514	SLP paralogue 15	71.12	18.96	3 (50)	167580	4.7	√
L2	CD0514	SLP paralogue 15	23.72	10.80	1 (22)	167580	4.7	√
L3	CD0514	SLP paralogue 15	56.88	9.45	2 (32)	167580	4.7	√
L4	CD2966	AdhE, alcohol dehydrogenase	96.36	9.05	5 (80)	96674	5.8	×
L5	CD2966	AdhE, alcohol dehydrogenase	45.89	3.96	2 (35)	96674	5.8	×
L6	CD2966	AdhE, alcohol dehydrogenase	75.25	9.28	4 (80)	96674	5.8	×
	CD3394	PykF, pyruvate kinase	52.26	4.78	2 (28)	63092	5.0	×
L7	CD2791	SLP paralogue 2	141.98	14.12	6 (87)	66427	6.4	√
	CD3091	TreA/C, trehalose-6-phosphate hydrolase	146.72	11.71	6 (65)	66104	6.3	×
L8	CD3394	PykF, pyruvate kinase	65.31	4.95	2 (29)	63092	5.0	×
L9	CD2755	PtsI, phosphoenolpyruvate phosphotransferase	52.02	5.26	2 (30)	63132	4.8	×
L10	CD2672	AppA, ABC transporter oligopeptide binding protein	173.01	15.82	8 (84)	59706	5.6	√
L11	CD2672	AppA, ABC transporter oligopeptide binding protein	118.12	14.12	7 (75)	59706	5.6	√
	CD3394	PykF, pyruvate kinase	59.31	4.44	2 (26)	63092	5.0	×

Table 4.4 continued

Spot	Genome ID	Description	Protein probe score	% coverage	Number of peptides (residues)	Theoretical M_r	Theoretical pI	Signal sequence
L12	CD2793	LMW SLP	156.54	14.76	3 (53)	76115 (34228)	4.9	√
	CD0395	FldA, Cinnamoyl-CoA:R-phenyllactate CoA transferase	92.48	18.80	5 (74)	44216	5.2	×
	CD3173	Pgk, Phosphoglycerate kinase	80.10	13.50	3 (54)	43048	5.1	×
	CD1059	ThlA1, Acetyl coA acetyl transferase	61.38	15.37	3 (63)	43171	5.4	×
L13	CD2672	AppA, ABC transporter oligopeptide binding protein	65.95	8.10	3 (43)	59706	5.6	√
	CD0855	Putative transport system peptide-binding protein	19.51	10.73	2 (56)	58496	5.4	√
L14	CD3285	Pgi, glucose 6-phosphate isomerase	53.74	14.92	4 (57)	50519	5.6	×
	CD3170	Eno, enolase	44.23	4.24	2 (19)	48149	4.6	×
L15	CD3173	Pgk, phosphoglycerate kinase	47.05	10.00	2 (40)	43048	5.1	×
L16	CD0395	FldA, Cinnamoyl-CoA:R-phenyllactate CoA transferase	302.60	49.12	13 (196)	44216	5.2	×
L17	CD3285	Pgi, glucose 6-phosphate isomerase	84.25	17.31	4 (78)	50519	5.6	×
L18	CD0395	FldA, Cinnamoyl-CoA:R-phenyllactate CoA transferase	85.69	14.29	3 (57)	44216	5.2	×
	CD1279	NifS cysteine desulfurase	31.22	11.44	2 (45)	44387	6.3	×
L19	CD3174	GapB, Glyceraldehyde 3-phosphate dehydrogenase	167.65	21.19	6 (71)	36015	5.9	×
L20	CD3174	GapB, Glyceraldehyde 3-phosphate dehydrogenase	282.07	29.85	9 (100)	36015	5.9	×
	CD1594	CysM, putative O-acetylserine sulfhydrylase	237.30	37.09	8 (112)	32647	5.4	×
L21	CD0672	Conserved hypothetical	22.65	8.30	2 (24)	33099	5.4	×
L22	CD0672	Conserved hypothetical	79.80	16.96	4 (49)	33099	5.4	×
L23	CD1594	CysM, putative O-acetylserine sulfhydrylase	178.19	36.75	8 (105)	32647	5.4	×

Table 4.4 continued

Spot	Genome ID	Description	Protein probe score	% coverage	Number of peptides (residues)	Theoretical M_r	Theoretical pI	Signal sequence
L24	CD0401	EtfA1, electron transport flavoprotein	114.95	18.26	5 (63)	37129	4.6	×
	CD2139	Tsf, Elongation factor	72.92	18.15	3 (55)	33122	5.2	×
L25	CD3664	Putative aminotransferase	110.37	20.10	5 (80)	44779	5.2	×
	CD0395	FldA, Cinnamoyl-CoA:R-phenyllactate CoA transferase	169.62	17.29	4 (69)	44216	5.2	×
L26	CD3173	Pgk, phosphoglycerate kinase	105.57	17.75	4 (71)	43048	5.1	×
	CD2793	LMW SLP	40.06	4.87	2 (35)	76115 (34228)	4.9	√
	CD3664	Putative aminotransferase	74.23	9.04	3 (36)	44779	5.2	×
L27	CD2793	LMW SLP	240.26	20.31	9 (153)	76115 (34228)	4.9	√
L28	CD3173	Pgk, phosphoglycerate kinase	120.70	19.50	5 (78)	43048	5.1	×
	CD2793	LMW SLP	228.73	18.78	10 (135)	76115 (34228)	4.9	√
L29	CD0394	Ldh, Lactate dehydrogenase	196.29	27.41	7 (90)	36449	5.1	×
	CD2793	LMW SLP	56.35	3.62	2 (23)	76115 (34228)	4.9	√
L30	CD2793	LMW SLP	313.44	27.54	13 (199)	76115 (34228)	4.9	√
L31	CD2793	LMW SLP	64.29	3.62	2 (23)	76115 (34228)	4.9	√
L32	CD2793	LMW SLP	382.00	26.15	13 (187)	76115 (34228)	4.9	√
L33	CD0239	FliC, Flagellin	130.07	37.24	6 (31)	30755	8.6	×
	CD0873	ABC transporter substrate binding lipoprotein	48.21	11.18	2 (38)	36049	5.4	√
L34	CD0844	SLP paralogue 25	55.70	13.10	3 (41)	33776	5.2	√
	CD2793	LMW SLP	122.80	6.81	4 (49)	76115 (34228)	4.9	√

Table 4.4 continued

Spot	Genome ID	Description	Protein probe score	% coverage	Number of peptides (residues)	Theoretical M_r	Theoretical pI	Signal sequence
L35	CD2755	PtsI, phosphoenolpyruvate phosphotransferase	73.57	10.53	3 (60)	63132	4.8	×
	CD3225	Dihydrodipicolinate synthetase	56.36	11.19	3 (32)	31577	4.8	×
L36	CD0795	Conserved hypothetical	49.99	8.5	2 (21)	26930	4.6	×
L37	CD3174	GapB, Glyceraldehyde 3-phosphate dehydrogenase	56.16	8.06	2 (27)	36015	5.9	×
L38	CD0239	FliC, flagellin	71.07	14.14	2 (41)	30755	8.6	×
	CD3172	TpiA, triospehosphate isomerase	68.27	18.62	3 (46)	27174	5.0	×
L39	CD2755	PtsI, phosphoenolpyruvate phosphotransferase	73.31	9.65	3 (55)	63132	4.8	×
L40	CD2240	NanA, N-acetylneuraminate lyase	58.81	16.56	3 (50)	34291	5.3	×
L41	CD2461	DnaK, chaperone protein	129.52	21.18	5 (68)	66449	4.7	×
L42	CD3172	TpiA, triosephosphate isomerase	193.09	36.84	7 (80)	27174	5.0	×
L43	CD0401	EtfA1, electron transport flavoprotein	102.39	12.46	4 (43)	37129	5.1	×
L44	CD0401	EtfA1, electron transport flavoprotein	76.23	6.67	3 (23)	37129	5.1	×
L45	CD0239	FliC, flagellin	267.22	9.31	2 (28)	30755	8.6	×
L46	CD1125	Nitroreductase	92.28	21.15	4 (62)	32711	5.3	×
L47	CD3205	Nitroreductase	35.55	9.65	2 (19)	22344	7.4	×
L48	CD0058	TufA Elongation factor	71.20	5.54	2 (22)	44008	4.6	×
L49	CD2793	HMW SLP	90.85	20.32	4 (76)	76115 (39504)	4.8	√
L50	CD0401	EtfA1, electron transport flavoprotein	68.21	15.65	4 (44)	37129	5.1	×
L51	CD2461	DnaK, chaperone protein	33.69	5.69	2 (35)	66449	4.7	×

Table 4.4 continued

Spot	Genome ID	Description	Protein probe score	% coverage	Number of peptides (residues)	Theoretical M_r	Theoretical pI	Signal sequence
L52	CD1634	TerD, telurium resistance protein	56.20	19.79	3 (38)	20742	4.5	×
L53	CD1280	NifU, nitrogen fixation/Fe-S cluster formation	87.80	26.03	3 (38)	15853	4.5	×
L54	CD2513	Putative translation inhibitor endoribonuclease	55.61	11.12	2(14)	13582	4.9	×
L55	CD1355	CspB, putative cold shock protein	22.70	25.76	1 (17)	7105	4.7	×
L56	CD2793	HMW SLP	70.82	13.90	3 (52)	76115 (39504)	4.6	√
L57	CD2793	HMW SLP	66.78	16.84	3 (63)	76115 (39504)	4.6	√
L58	CD1355	CspB, putative cold shock protein	62.90	48.48	3 (32)	7105	4.7	×
L59	CD1690	TrxA1, thioredoxin	31.01	20.95	2 (22)	11708	5.1	×
L60	CD2756	PtsH, PTS system, phosphocarrier protein	113.88	60.46	4 (51)	9046	5.2	×
L61	CD3663	RpsF, 30S ribosomal protein	125.17	66.30	6 (61)	10502	7.3	×
L62	CD2137	Rrf, ribosome recycling factor	91.69	32.97	5 (61)	21090	7.5	×
L63	Ribonuclease	From sample preparation	488.97	24.82	2 (35)	16442	9.0	-
L64	Ribonuclease	From sample preparation	1469	56.67	5 (85)	16442	9.0	-
L65	Lysozyme	From sample preparation	100.04	12.92	2 (19)	16220	10.7	-
L66	CD2787	SLP paralogue 4	27.82	4.66	1 (17)	38577	6.7	√
	Lysozyme	From sample preparation	635.50	36.73	5 (54)	16220	10.7	-
L67	Lysozyme	From sample preparation	606.92	31.97	4 (47)	16220	9.5	-
L68	Ribonuclease	From sample preparation	2196.24	87.10	6 (101)	16442	9.0	-
L69	Lysozyme	From sample preparation	148.00	30.61	3 (45)	16220	10.7	-

signal peptide and therefore are expected to be located in either the cell wall or associated with the cytoplasmic membrane. However a number of proteins detected do not contain a known signal sequence. For example, a number of proteins involved in metabolic functions were found including CD0394 lactate dehydrogenase, CD0395 E-cinnamoyl CoA:R-phenyllactate CoA transferase large subunit, CD0401 electron transport flavoprotein, CD1059 acetyl coA acetyl transferase, CD1279 CoA:R-phenyllactate CoA transferase large subunit, CD1594 cysteine synthase, CD1690 thioredoxin, CD2240 *N*-acetylneuraminate lyase, CD2755 phosphoenolpyruvate phosphotransferase, CD2756 PTS system, phosphocarrier protein, CD2966 alcohol dehydrogenase, CD3091 trealose-6-phosphate hydrolase, CD3170 enolase, CD3172 triosephosphate isomerase, CD3173 phosphoglycerate kinase, CD3174 glyceraldehyde-3-phosphate dehydrogenase, CD3225 dihydropicolinate synthetase, CD3285 glucose 6-phosphate isomerase, CD3394 pyruvate kinase, and CD3664 putative aminotransferase. Some proteins involved in information transfer were found, CD0058 and CD2139 elongation factors, and CD2513 translation initiation factor. Unexpectedly, a protein thought to be involved in nitrogen fixation was found, CD1280 NifU nitrogen fixation/Fe-S cluster formation, and this protein was also found in the glycine extract. CD1125 and CD3205 both encode nitroreductases. Conserved hypothetical proteins CD0672 and CD0795 were found, as well as a tellurium-resistance protein CD1634.

Comparison of the glycine and lysozyme extracts shows that low pH glycine treatment releases specifically the S-layer proteins and just a few other proteins from the cell surface, whereas digestion of the peptidoglycan cell wall with lysozyme releases many more proteins of greater diversity. The differences between the extracts are illustrated by the detection of different SLP paralogues in each extract. For example, paralogue 4 is only seen in the lysozyme extract whereas paralogues 6, 12 and 24 are only seen in the lysozyme extract. Paralogues 2, 15 and 25 are common to both extracts. A number of proteins found in the lysozyme extract did not have signal peptides, in contrast to the glycine extract that contains predominantly proteins with signal peptides or those exported from the cell by a dedicated mechanism.

4.3 Structural studies of the SLPs

Very little is known about the structure of SLPs in general and the SLPs of *C. difficile* in particular. It is not clear how the SLPs assemble into a paracrystalline array, how and where they interact or how they are attached to the underlying cell wall. With this in mind, and to put into context the proteomics data on the other proteins associated with the cell surface, the structure of the SLPs was investigated. Recombinant SLPs were

expressed, purified, and a number of structural techniques were used for characterisation.

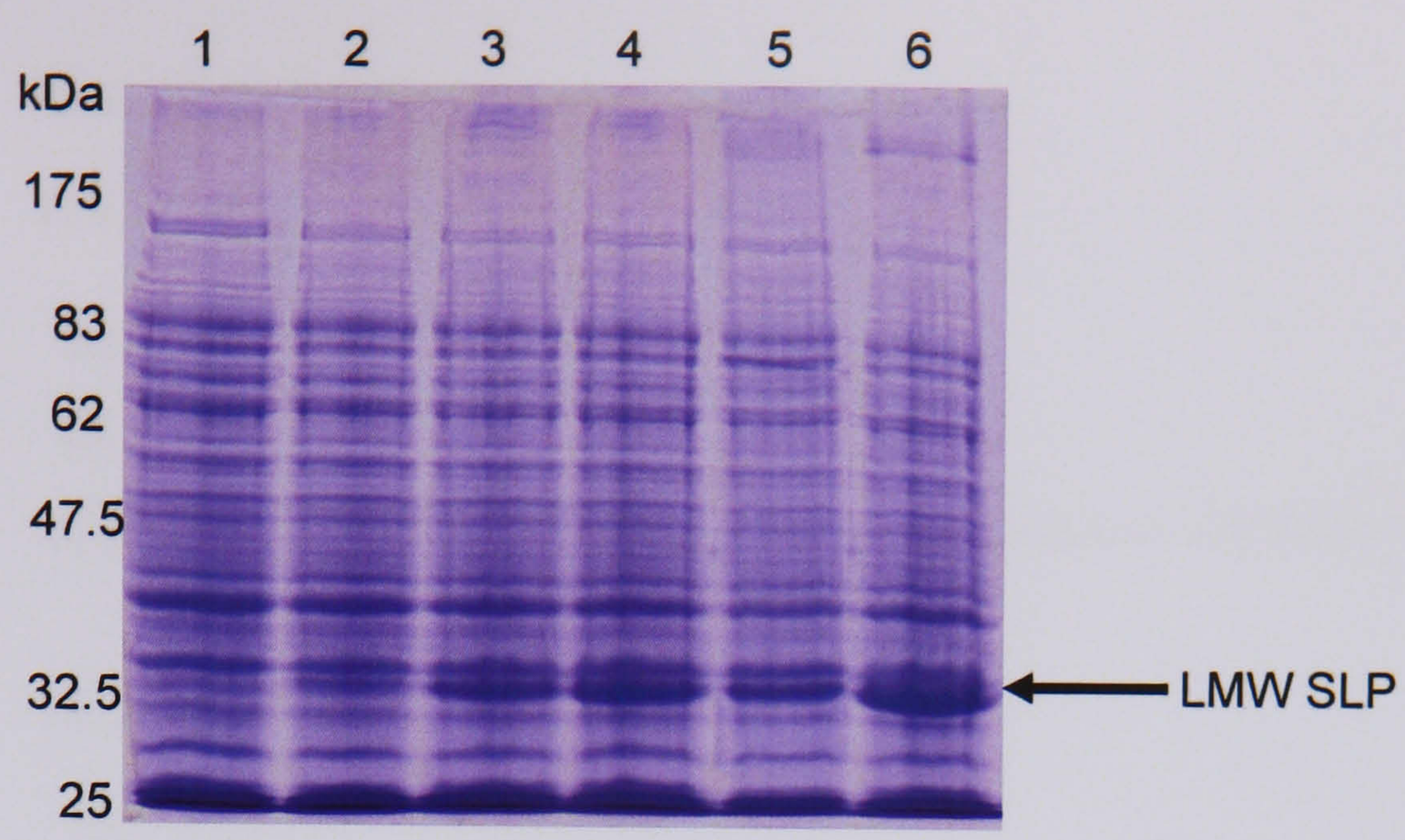
4.3.1 Expression and purification of SLPs

Plasmids pETLow and pETHigh were a gift from Emanuela Calabi. The plasmid pETLow contains DNA encoding the amino-terminal variable domain of *slpA*, without the signal sequence and this corresponds to the LMW SLP. The plasmid pETHigh contains DNA from the carboxyl-terminal cell wall binding domain, and this corresponds to the HMW SLP. Both polypeptides were expressed with carboxyl-terminal hexa-histidine tags in *E. coli* BL21(λ*DE3*) cells. Bacterial cultures were induced for the expression of recombinant protein under optimal conditions of 37 °C with 1 mM IPTG. Samples of the cultures were taken before induction with IPTG and after induction for 30 min, 1, 2, 3 and 24 hours. After IPTG induction, BL21(λ*DE3*) containing pETLow produced an additional protein of approximately 36 kDa (Figure 4.11A). This protein reacted to anti-hexa-histidine tag antibodies when analysed by Western blot, and therefore corresponds to the LMW SLP (data not shown). Similarly, after IPTG induction, BL21(λ*DE3*) containing pETHigh produced an additional protein of approximately 45 kDa (Figure 4.12A). This protein reacted to anti-hexa-histidine tag antibodies when analysed by Western blot, and is therefore believed to correspond to the HMW SLP (data not shown).

The LMW- and HMW SLPs were partially purified from the total soluble protein of bacterial cultures induced for 3 hours using nickel-agarose affinity chromatography (Figures 4.11B and 4.12B). High yields of both proteins were achieved, with a small number of additional bands. Typically, 5 mg LMW SLP and 3 mg HMW SLP were produced per litre of culture. Elution fractions of each protein were pooled and concentrated to a volume of less than 1 ml.

Because the additional proteins in the recombinant SLP preparations were of a range of sizes but dissimilar to the proteins of interest, a second stage of purification was employed using size exclusion chromatography. The pooled and concentrated partially pure LMW- and HMW SLP preparations were loaded onto a S100 column attached to an FPLC and eluted over two column volumes in 20 mM HEPES pH 7.4. Fractions collected at the peaks of absorbance were analysed by SDS-PAGE (data not shown). The fractions of LMW SLP were pure and relatively highly concentrated. The fractions of HMW SLP were less concentrated and still contained a few contaminating proteins. Fractions of each protein were pooled and concentrated for subsequent analysis.

A



B

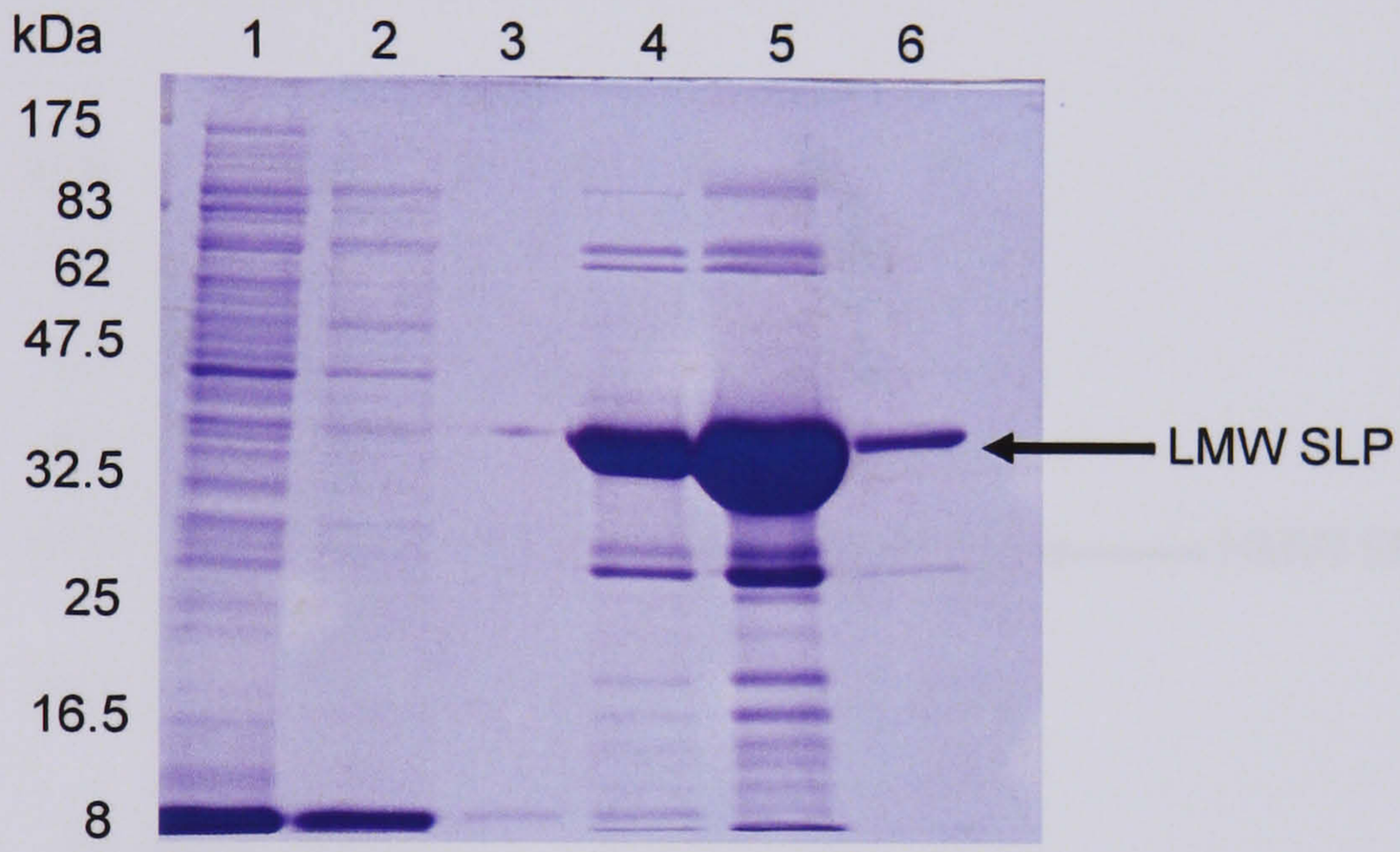
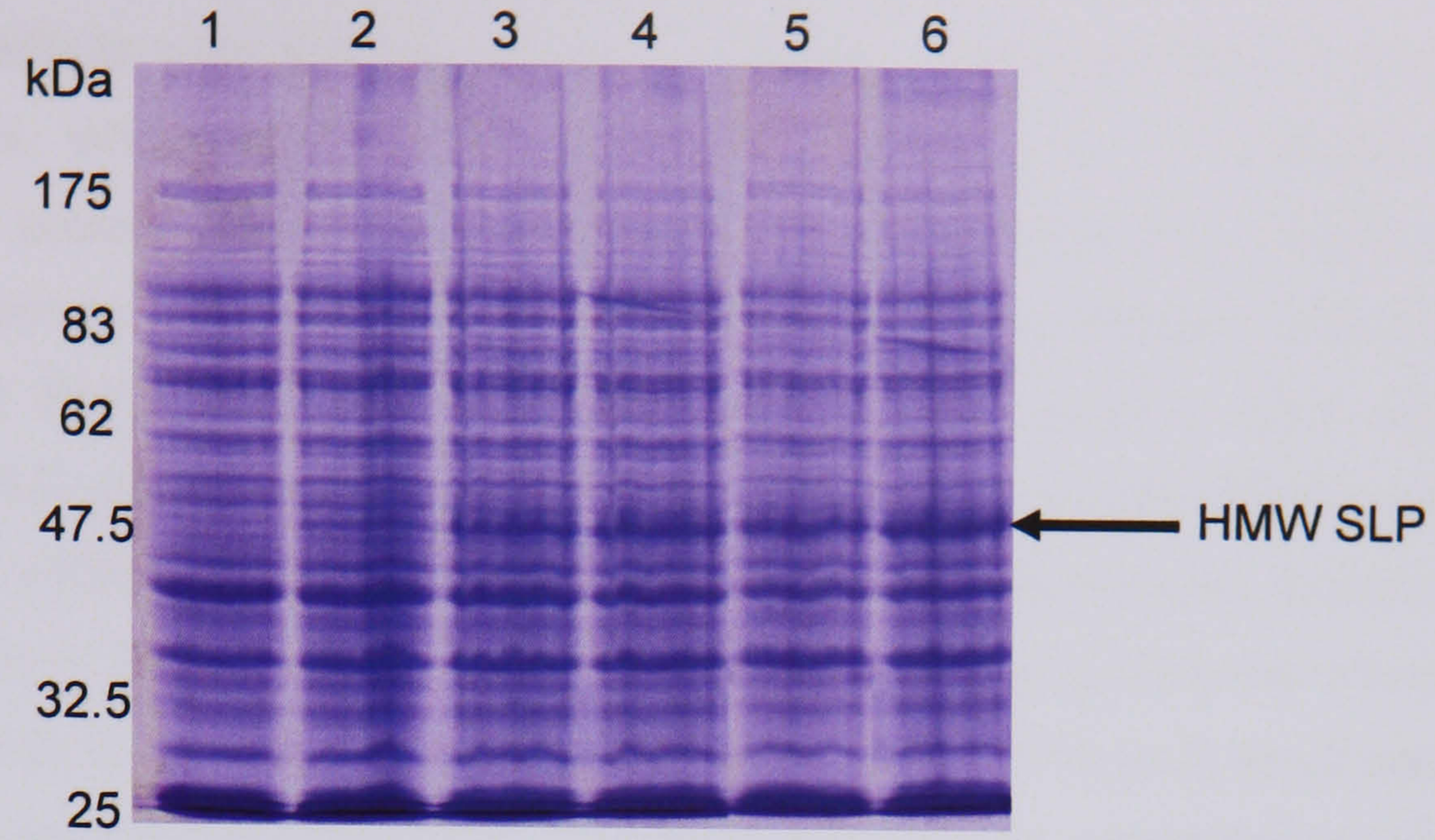


Figure 4.11 Expression and purification of LMW SLP. **A** Total soluble protein from cultures of *E. coli* BL21(λ*DE3*) containing pETLow separated by SDS-PAGE. Lane 1, before induction of LMW SLP expression with IPTG; lane 2, 30 min after addition of IPTG; lane 3, 1 hour induction; lane 4, 2 hours induction; lane 5, 3 hours induction; lane 6 overnight induction. **B** Purification of LMW SLP using nickel-agarose, analysed by SDS-PAGE. Lane 1, column flow through; lane 2, wash 1; lane 3, wash 2; lane 4, elution 1; lane 5, elution 2; lane 6, elution 3. The arrow indicates LMW SLP. Molecular masses in kDa are shown.

A



B

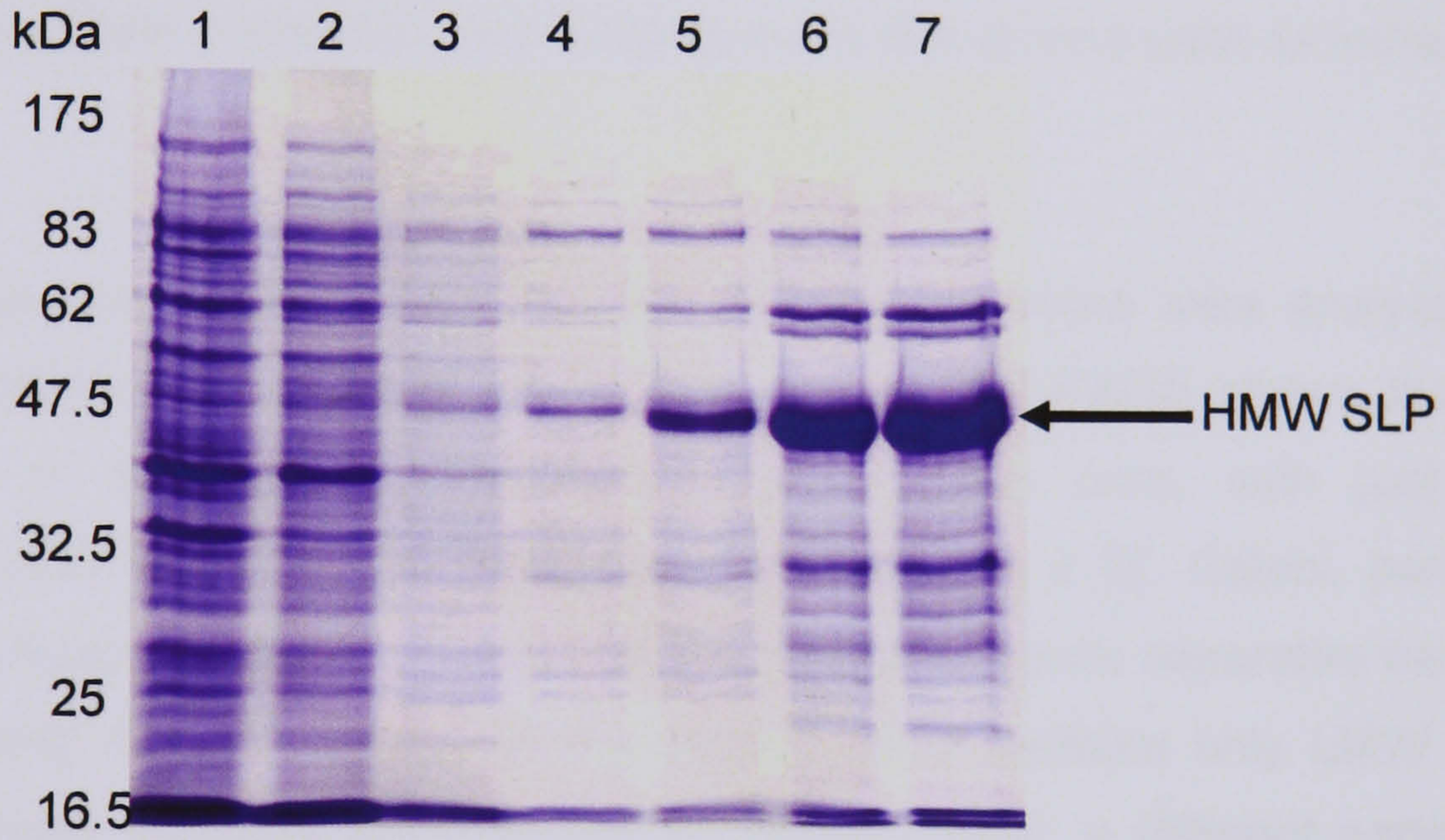


Figure 4.12 Expression and purification of HMW SLP. **A** Total soluble protein from cultures of *E. coli* BL21(λ*DE3*) (pETHigh) separated by SDS-PAGE. Lane 1, before induction of HMW SLP expression with IPTG; lane 2, 30 min after addition of IPTG; lane 3, 1 hour induction; lane 4, 2 hours induction; lane 5, 3 hours induction; lane 6 overnight induction. **B** Purification of HMW SLP using nickel-agarose, analysed by SDS-PAGE. Lane 1, column flow through; lane 2, wash 1; lane 3, wash 2; lane 4, elution 1; lane 5, elution 2; lane 6, elution 3; lane 7, elution 4. The arrow indicates HMW SLP. Molecular masses in kDa are shown.

4.3.2 Analysis of native SLPs removed by glycine treatment

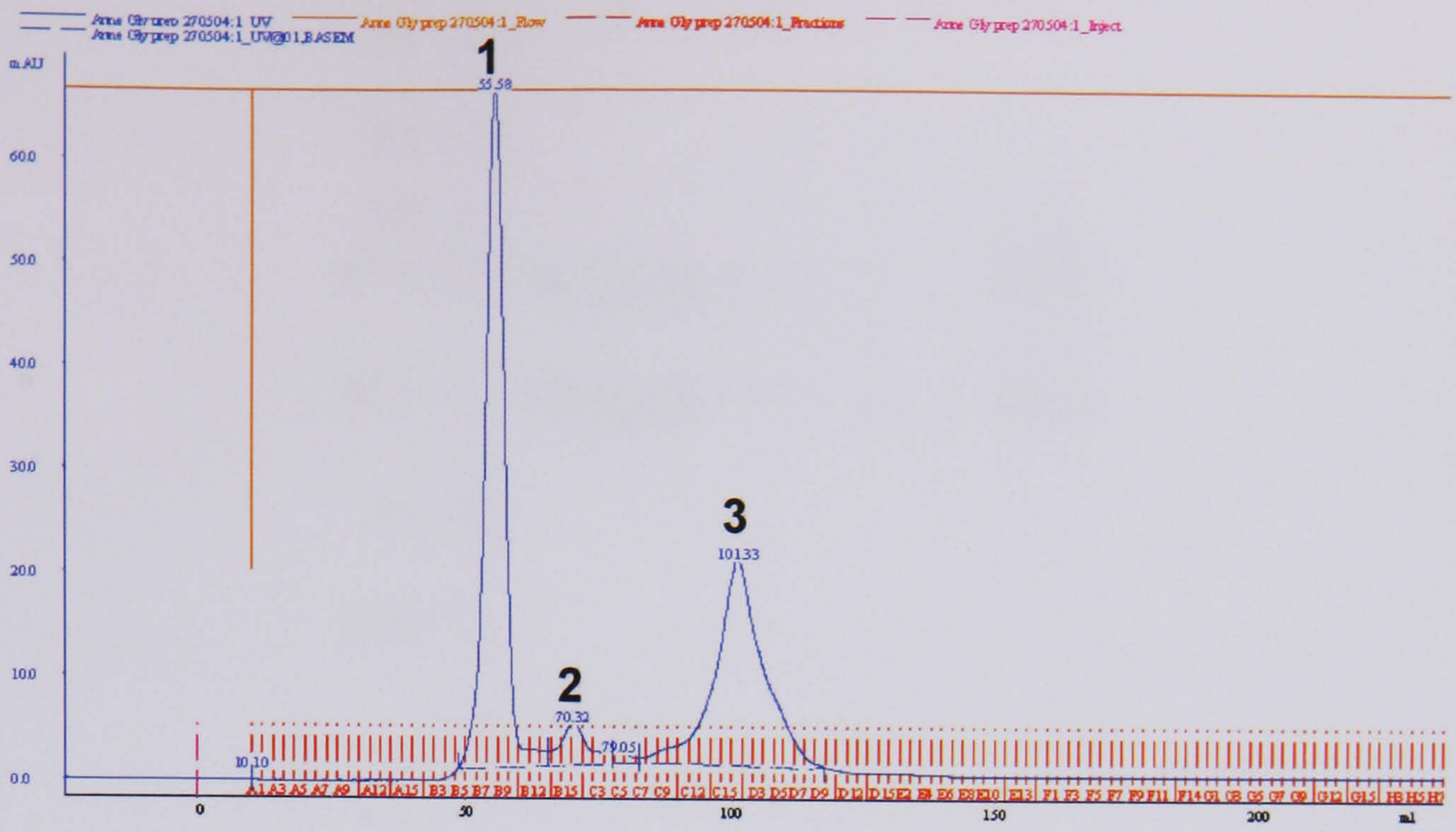
Treatment of *C.* difficile cells with low pH glycine results in the extraction of relatively pure native SLPs, which were further purified and studied as an alternative to recombinant SLP protein. SLPs were extracted from the surface of *C. difficile* cells using glycine treatment and analysed by a combination of gel filtration and PAGE. Glycine extracted SLPs were separated according to size using a S200 column attached to an FPLC and eluted over two column volumes into 20 mM HEPES pH 7.4 (Figure 4.13A). Fractions corresponding to the peaks of absorbance were analysed by SDS-PAGE (Figure 4.13B). This showed that the first peak eluting from the column at 55.6 ml corresponds to a mixture of LMW- and HMW SLPs. The next small peak at 70.3 ml is composed only of LMW SLP. The final peak at 101 ml yielded no visible protein on SDS-PAGE. From standards previously calibrated to the column the first peak was sized as 363 kDa. The second peak was sized as 109 kDa and the third peak as 9 kDa. The final peak that had no protein visible after SDS-PAGE analysis and contains the smallest molecules may correspond to the glycine used to extract the SLPs.

The glycine extract and fractions from the first and second peaks were analysed by denaturing and native PAGE (Figure 4.14). Denaturing SDS-PAGE shows that the samples applied to the gel filtration column were quite pure, with just one 'contaminating' protein identified previously as Slp paralogue 2 (E. Calabi, personal communication). Again, protein from the first size exclusion peak separates into two bands corresponding to both SLPs, and the second peak contains only LMW SLP. However, when these samples are analysed by native PAGE, a different pattern of separation occurs. Only a single band of a large molecular weight (above 150 kDa) is seen for both glycine extracts and the co-eluting mix of SLPs from first size exclusion peak. The fraction containing only LMW SLP from the second size exclusion peak could not be visualised on this gel.

4.3.3 *In silico* prediction of SLP secondary structure.

The secondary structure of a protein sequence can be predicted by *in silico* methods. The programme 3D PSSM (www.sgb.bio.ic.ac.uk) was used to predict the secondary structure of the LMW- and HMW SLPs (Figure 4.15). This programme assigns each residue a predicted secondary structure, α-helix, β-sheet or coil and gives a score of its confidence in this assignment. Using this programme, the secondary structure of the LMW SLP is predicted to contain 3 α-helices, all towards the amino-terminus and making up 18% of the protein. Also predicted are 3 large regions of β-sheet comprising

A



B

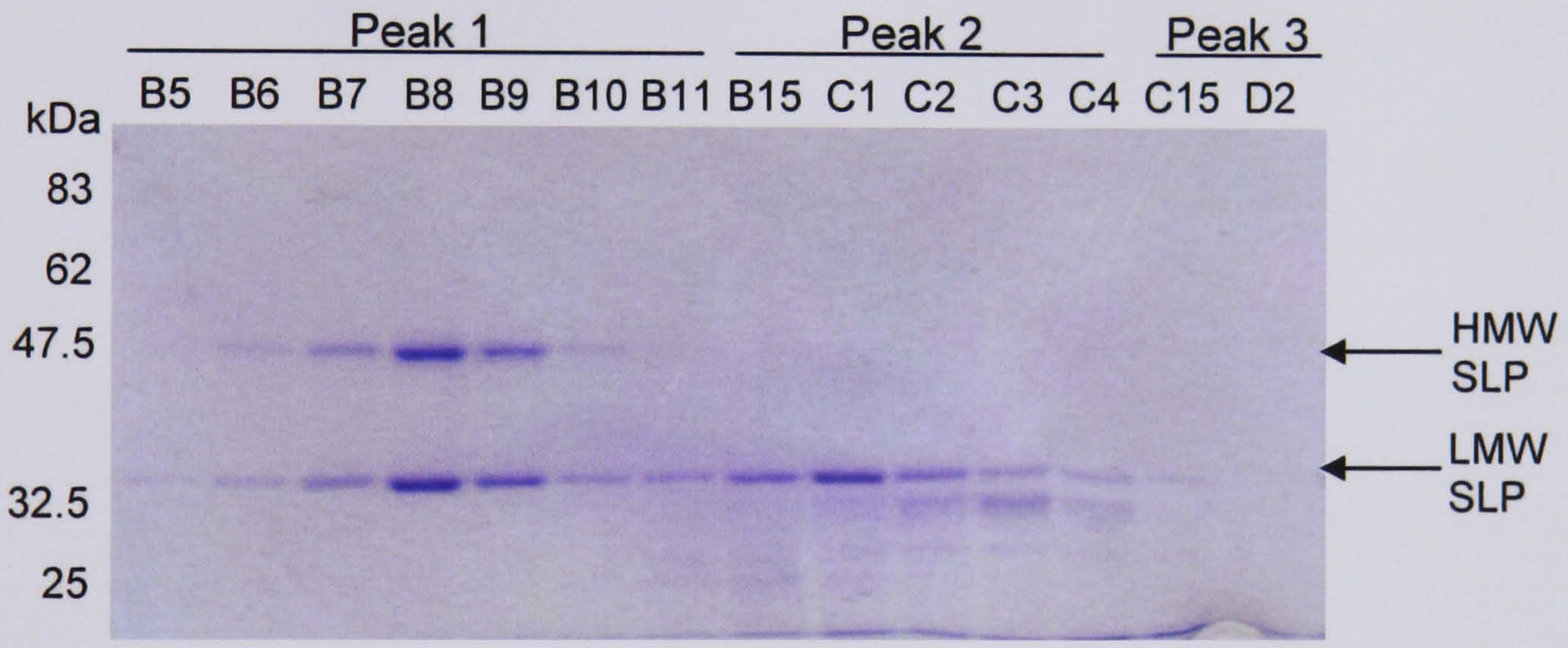


Figure 4.13 Size exclusion chromatography of SLPs extracted from *C. difficile* by glycine treatment. **A** FPLC trace of protein elution. The blue line represents UV absorption, which is indicative of protein. Red bars show the fractions collected. **B** SDS-PAGE analysis of fractions collected around the peaks of absorbance. Fractions are indicated at the top of each lane. Arrows point to the HMW and LMW proteins.

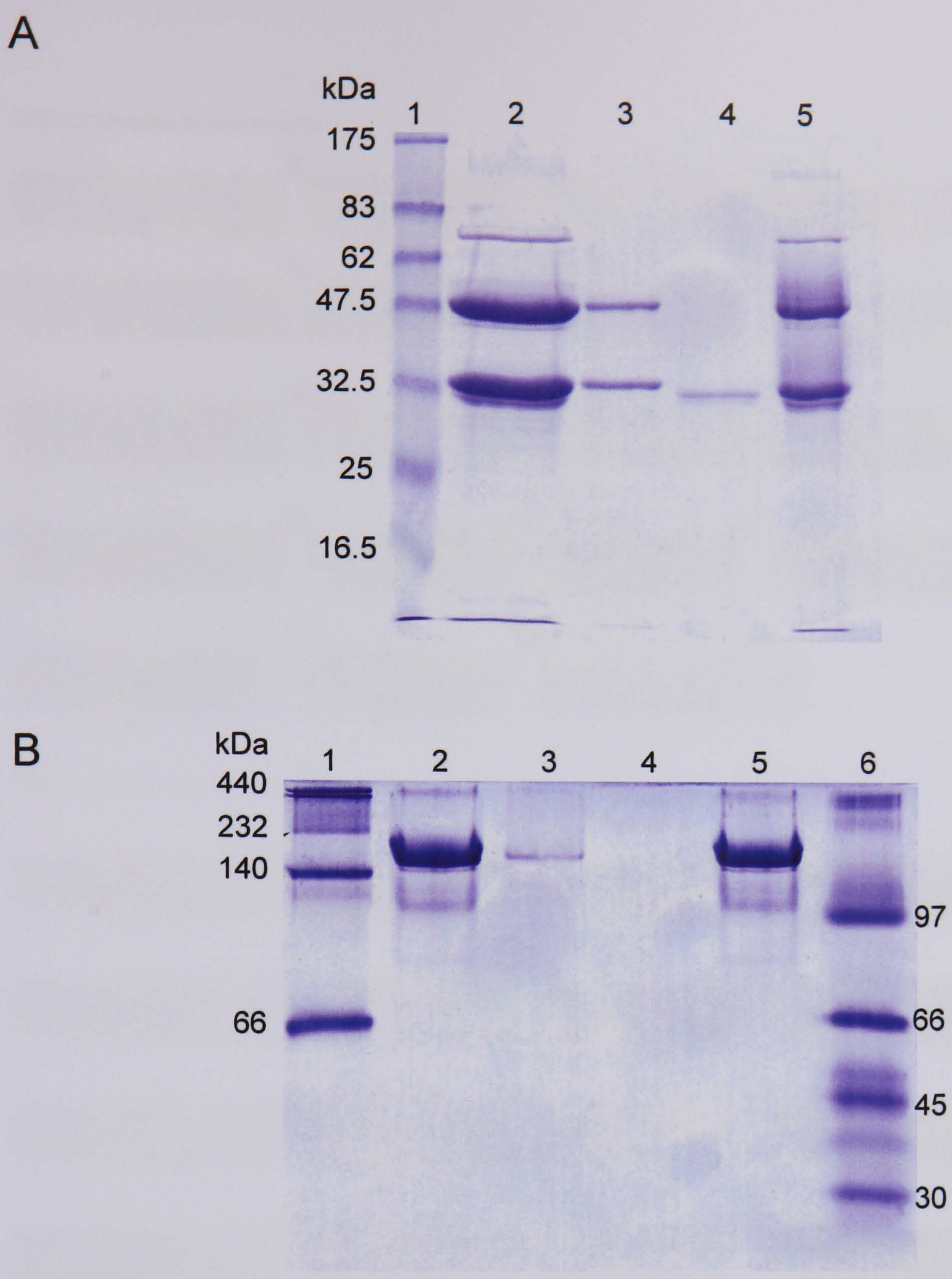
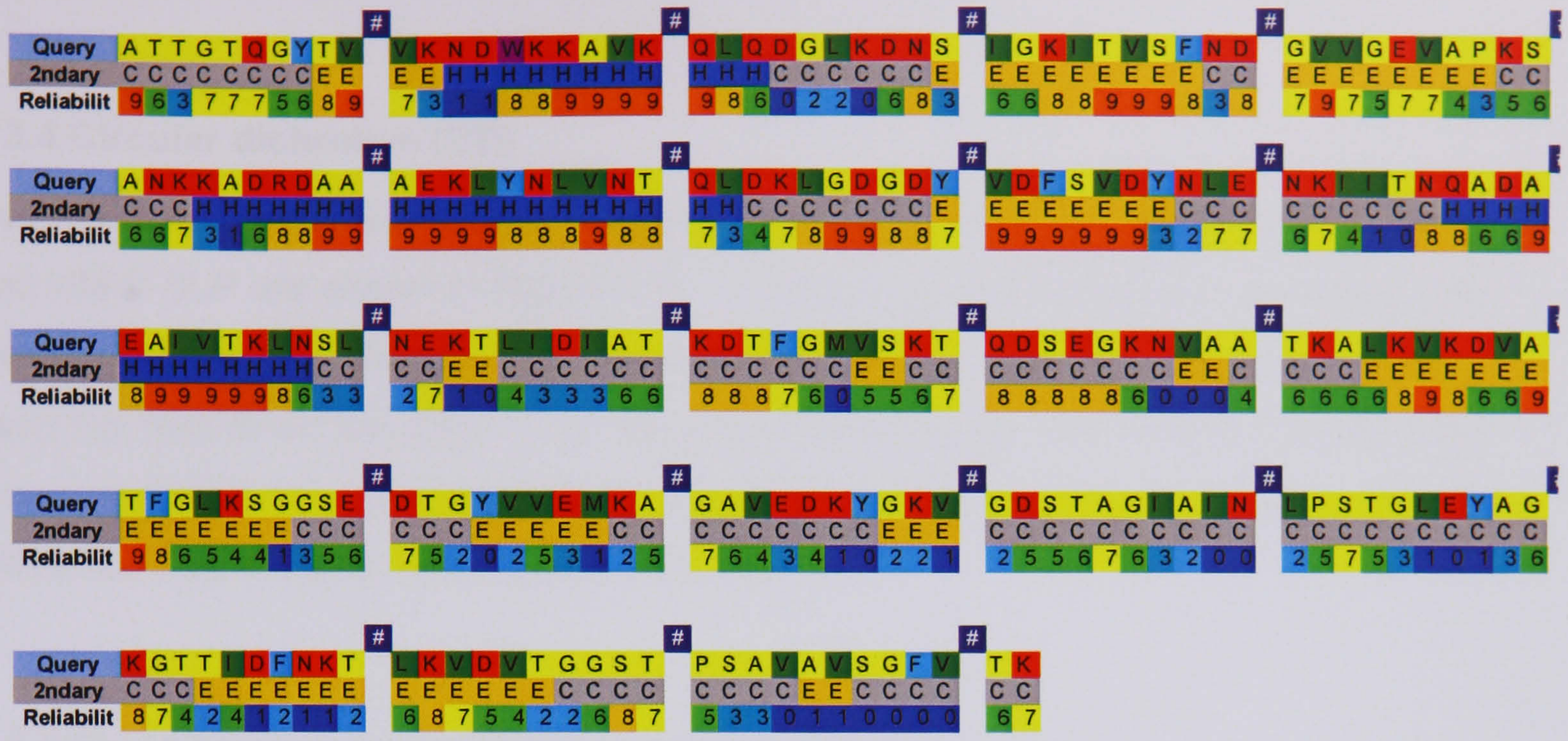


Figure 4.14 PAGE analysis of glycine extracted SLPs. PAGE analysis of SLPs extracted from *C. difficile* by glycine treatment and separated by size exclusion chromatography. **A** Denaturing SDS-PAGE. Lane 1, molecular mass markers in kDa; lane 2, 10 μg glycine extracted SLPs; lane 3, fraction B9 (from size exclusion chromatography; Figure 4.13); lane 4, fraction C1 (Figure 4.13); lane 5, 5 μg glycine extract. **B** Native PAGE. Lane 1, molecular mass markers in kDa; lane 2, 10 μg glycine extracted SLPs; lane 3, fraction B9 (Figure 4.13); lane 4, fraction C1 (Figure 4.13); lane 5, 5 μg glycine extract; lane 6, molecular mass markers in kDa.

LMW SLP Secondary Structure Prediction.



HMW SLP Secondary Structure Prediction.

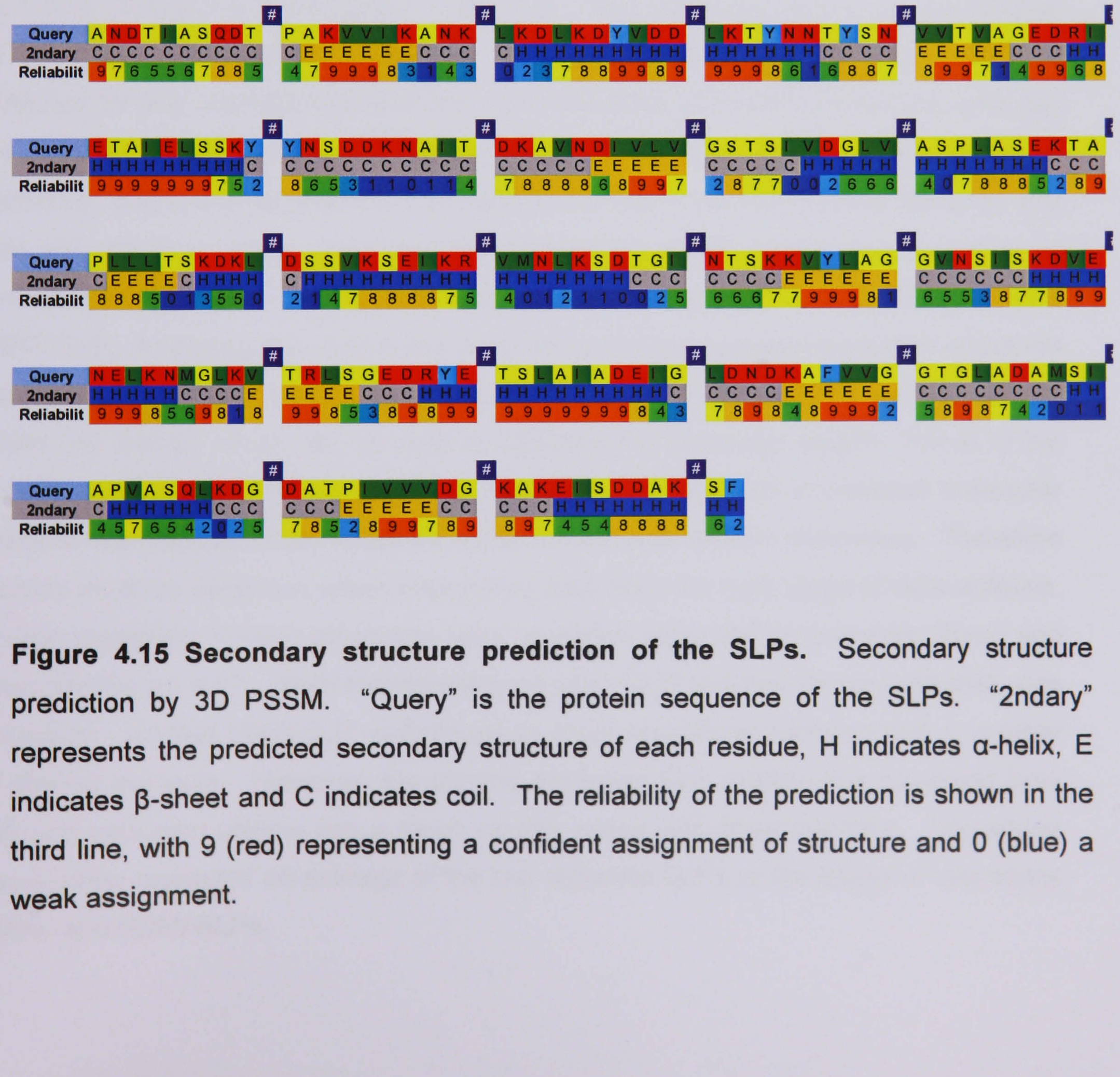


Figure 4.15 Secondary structure prediction of the SLPs. Secondary structure prediction by 3D PSSM. "Query" is the protein sequence of the SLPs. "2ndary" represents the predicted secondary structure of each residue, H indicates α-helix, E indicates β-sheet and C indicates coil. The reliability of the prediction is shown in the third line, with 9 (red) representing a confident assignment of structure and 0 (blue) a weak assignment.

31% of the protein and numerous coiled regions. The HMW SLP is predicted to contain 7 large regions of α-helix, comprising 42% of the protein, 18% of the protein is predicted to form β-sheet and the remainder is coil.

4.3.4 Circular dichroism (CD)

Purified recombinant SLPs were analysed using CD. The CD spectra for the LMW- and HMW SLP are shown in Figure 4.16. Analysis using the programme Selcon, with reference set three, was used to calculate percentages of α-helix and β-sheet which gave the best fit for the data. The calculated percentages of secondary structure are shown in Table 4.5. Both SLPs have a high proportion of β-sheet and turns, but the LMW SLP has a slightly less ordered structure.

4.3.5 Small Angle X-ray Scattering analysis (SAXS)

Purified, recombinant SLPs and glycine extracted size separated SLPs were analysed by SAXS (Svergun & Koch, 2002) (Table 4.6). This technique allows the visualisation of the macromolecular structure of a protein in solution. Ideally data should be obtained for low, intermediate and high concentrations of protein, so that the data can be merged to give information about different aspects of the protein structure. However, due to the large amount of aggregation that occurred in many samples, this was not possible. Data could only be obtained for single protein concentrations of recombinant LMW SLP, the glycine extracted co-eluting SLPs and glycine extracted LMW SLP. Analysing this data in the programme Primus reveals information about the size of the protein in solution (radius of gyration) and the distribution of electron density within the protein, which can be used to calculate the molecular weight. For all of the proteins analysed, interpretation of the data using Primus gave a predicted molecular weights much smaller than would be expected for even protein monomers. Therefore caution must be exercised when interpreting data from the next stage of data analysis, protein modelling. Protein structures were modelled using the programmes Gnom and Dammin (Figure 4.17). Both the recombinant LMW SLP and the glycine extracted, size exclusion - purified LMW SLP were found to adopt complicated shapes with a number of distinct domains. However, the glycine extracted SLP mixture that co-eluted from the size exclusion column has a much simpler shape with three domains. This shape could either represent an average of the two separate SLPs or the shape of interacting LMW- and HMW SLPs.

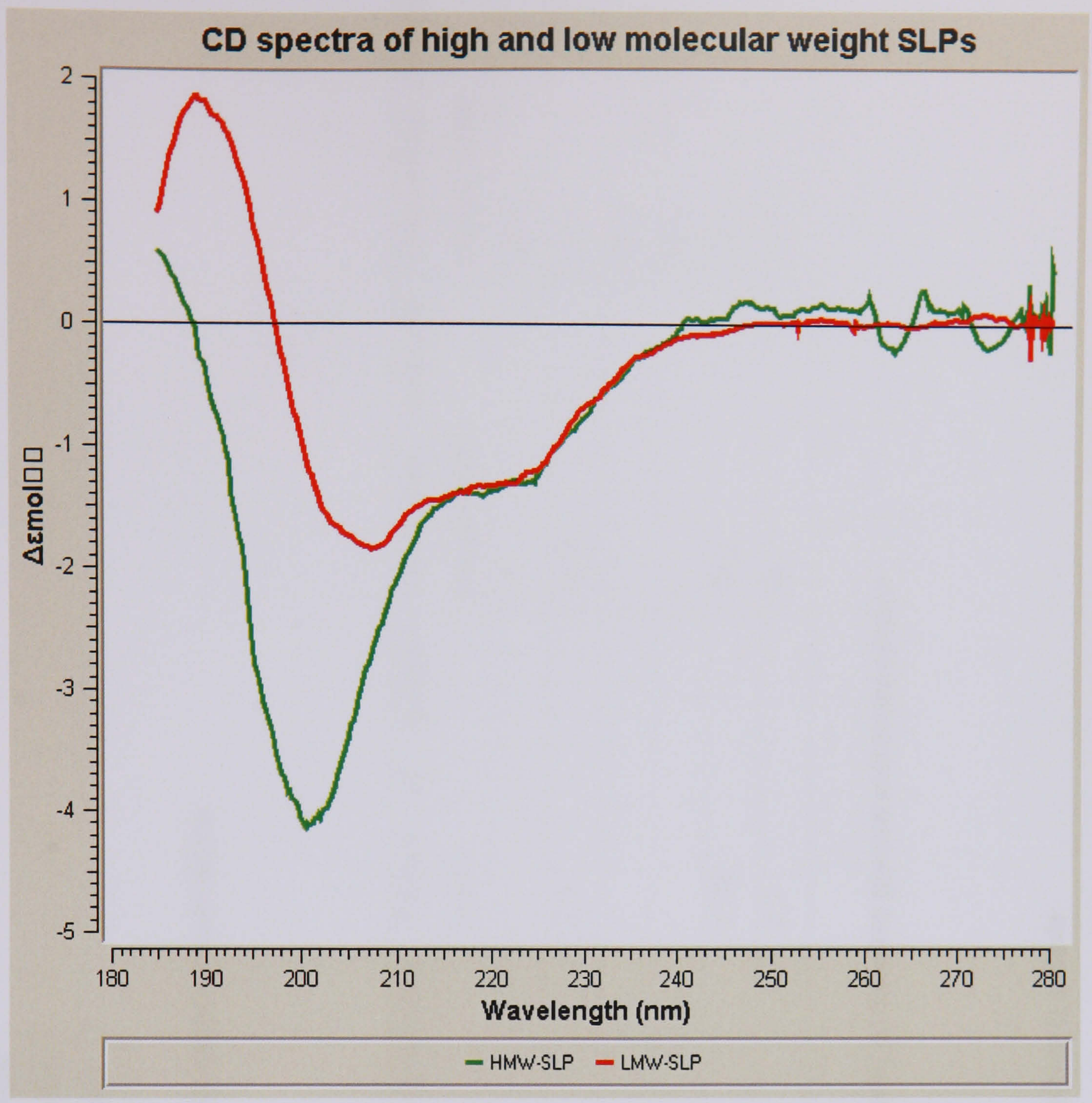


Figure 4.16 Circular dichroism analysis of the recombinant SLPs. CD spectra of LMW SLP (red) and HMW SLP (green).

	% Secondary structure				
Protein	**Alpha helix**	**Beta sheet**	**Beta turn**	**Unordered**	**Total**
LMW SLP	9.9	34.2	18.8	26.5	89.4
HMW SLP	14.6	27.7	21.7	35.4	99.5

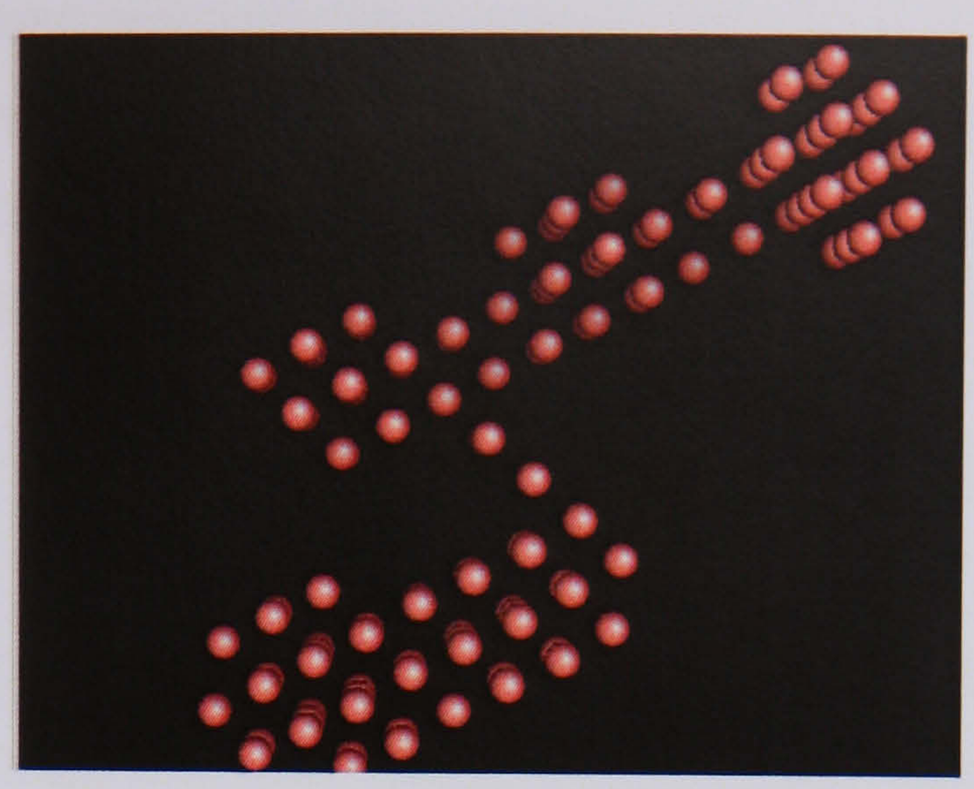
Table 4.5 Secondary structure analysis from CD data for the recombinant SLPs.

Protein sample[a]	Solution state determined by SAXS	Protein concentration (mg/ml)	Radius of gyration (nm)	Calculated molecular weight (kDa)	Theoretical molecular weight (kDa)
rLMW SLP 1	No diffraction	1.0	-	-	38.3
rLMW SLP 2	In solution	3.0	2.5	12.9	38.3
rLMW SLP 3	Aggregated	13.7	-	-	38.3
rHMW SLP	Aggregated	2.8	-	-	39.5
gLMW and HMW SLP	In solution	3.2	3.78	33.4	76.1
gLMW SLP	In solution	2.6	2.7	18	38.3

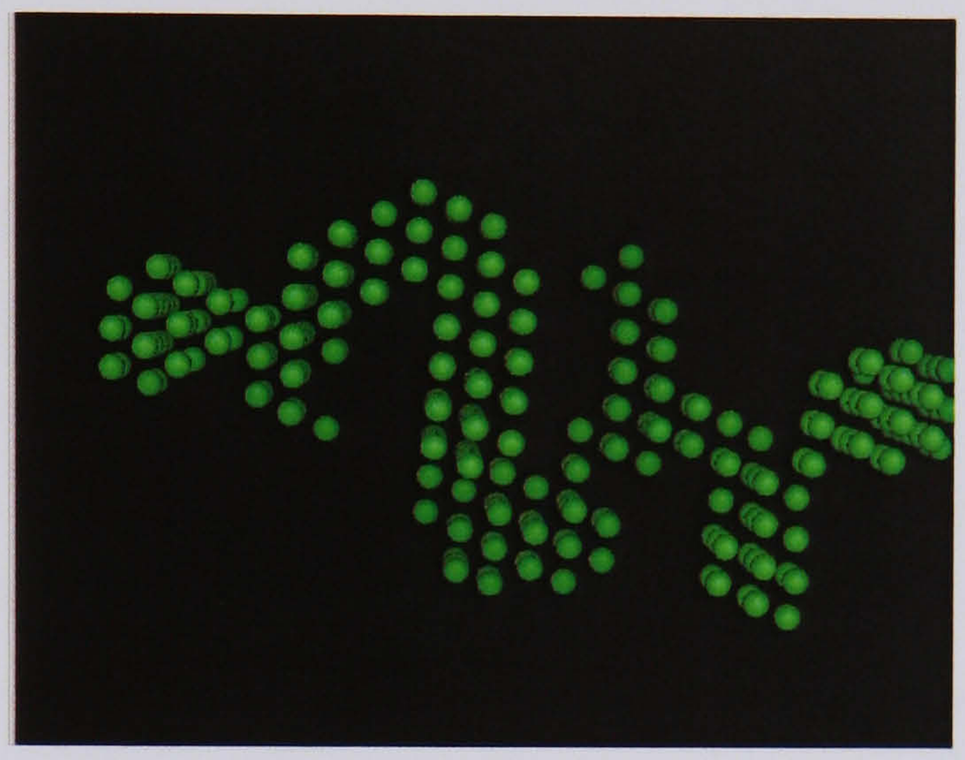
Table 4.6 SAXS data collection. r indicates recombinant protein, g indicates glycine extracted protein

[a] r indicates recombinant protein; g indicates glycine extracted protein from *C. difficile*

A. rLMW SLP



B. gLMW SLP



C. gLMW and HMW SLPs

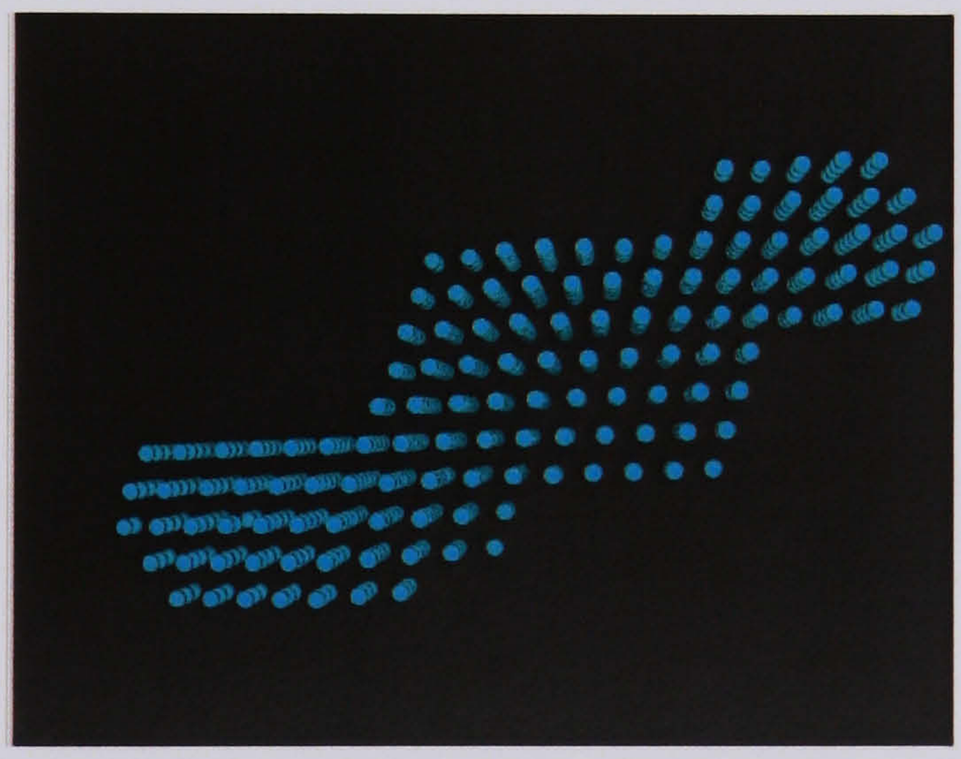


Figure 4.17 SAXS - based shape predictions for the SLPs. Data was processed using Gnom and Dammin. **A.** Recombinant LMW SLP. **B.** LMW SLP extracted from *C. difficile* by glycine treatment and separated by size exclusion chromatography. **C.** LMW- and HMW SLPs extracted from *C. difficile* by glycine treatment and co-eluting from the size exclusion column.

4.3.3 Crystallisation trials of the LMW SLP

Crystallisation trials were performed only with recombinant LMW SLP, due to the large quantities and high concentrations of protein required. Crystallisation trays were set up using concentrations of up to 15 mg/ml protein, however no crystals were obtained during the time frame of the study.

4.4 Discussion

Even though the method of attachment of the S-layer to the underlying cell wall remains to be elucidated, it is possible to draw some conclusions on the nature of the interaction based on the data presented here. The cell wall has previously been shown to swell and shrink in response to changes in pH, ionic strength and osmolarity (Marquis, 1968; Ou & Marquis, 1970). For example, in low salt concentrations similarly charged groups such as the carboxyl and phosphate of ionic polymers will repel, and separate as much as possible, making the cell wall swell. Under conditions of high ionic strength the converse will happen, and the cell wall will be compacted. In addition, the charge density of the wall can be affected by other factors such as the level of D-alanyl ester substitution of anionic polymers. Under physiological conditions, the amino group of D-alanine is protonated, but when substituted, the opposite charge on adjacent carboxyl or phosphate moiety is neutralised, thereby reducing the ability of the cell wall to bind cations (Heptinstall *et al.*, 1970). Growth conditions have been shown to affect the amount of alanine ester in the cell wall of *B. subtilis* (Ellwood & Tempest, 1972). In addition, the charge density of the cell wall can be affected by the level of amidation of carboxyl groups on the tetrapeptide chain. The compactness of the cell wall is affected by these potential variations in charge density, as tighter packing is possible at lower charge densities (Harwood *et al.*, 1990).

Urea treatment removed a large number of proteins from the surface of the bacterium, including large quantities of both SLPs. Urea acts as a denaturant, disrupting hydrogen bonds within the cell wall, and possibly also within the S-layer, and so presumably hydrogen bonding is important for the attachment of the SLP proteins to the bacterial surface. LiCl released varying numbers of proteins according to its concentration. LiCl may act by displacing ions involved in interactions anchoring proteins to the cell wall and/or S-layer, or as a chaotrophic agent. However, the LiCl extracts looked very different to the urea extracts on SDS-PAGE, and contained far fewer proteins. It is also possible that the LiCl salt could affect the porosity of the cell wall, i.e. high concentrations cause the cell wall to contract, trapping proteins within it, while lower concentrations result in a swelling of the cell wall, aiding the release of

associated proteins. In all cases, the LMW SLP was released preferentially to the HMW SLP, which may be an indicator of the relative positions of the proteins to the cell wall. Under all concentrations of LiCl used, only small quantities of the HMW SLP were removed, which may indicate that it is more tightly associated with the underlying cell wall than the easily shed LMW SLP. Indeed, the LMW SLP is shed under typical growth conditions, and is found in culture supernatants (Mukherjee *et al.*, 2002). EDTA removes surface proteins by chelating metal ions that may be essential in maintaining protein interactions. The concentration of EDTA used does not have such a profound effect on the extent of protein removed from the cell surface as the LiCl treatment, which is consistent with the theory of osmotic shock acting on the cell wall. Divalent cations are known to be important for the self-assembly of S-layers in *C. difficile* and other organisms (Takumi *et al.*, 1992; Sleytr & Beveridge, 1999). Removal of Ca^{2+} in particular can cause S-layers to shed from the surface of the bacterium. After S-layers have been isolated from the cell surface of several bacterial species, removal of the disrupting agent and reintroduction of Ca^{2+} can allow the spontaneous formation of the 2D lattice (Takumi *et al.*, 1992; Sleytr & Beveridge, 1999). Glycine is able to disrupt even very tight ionic interactions, and releases predominantly the two SLPs (in approximate equimolar concentrations), along with very few other proteins. Finally, when the cell wall was digested with lysozyme and other peptidoglycan degrading enzymes, many proteins were released that are presumably attached directly or indirectly to the peptidoglycan cell wall or to other cell wall polymers. Surprisingly only small quantities of the HMW SLP were present in this extract, which may indicate that it is not anchored directly to the cell wall, but instead to other polymers associated with the cell wall such as teichoic or techuronic acids, that are unaffected by this digestion treatment. These polymers are usually anionic, and so attachment of the HMW SLP via an ionic interaction would explain the greater efficiency of glycine and urea in removing this protein from the cell surface. Furthermore, it is possible that the LMW SLP (and possibly the other Slp paralogues too) are actually anchored to the cell wall via attachment to the HMW SLP, thus explaining the preferential release of the LMW SLP under all cell wall treatments.

Two methods of extracting proteins from the cell surface were used to map the surface proteome of *C. difficile*. The proteome profiles obtained were highly dependent upon the extraction method used. Glycine extraction was used because this removed both SLPs and other proteins that may be tightly associated with it. Lysozyme was used, producing a "cell wall extract" as this should methodically remove any proteins that are attached to the underlying cell wall. The most dominant proteins in these extracts were the two SLPs, but the next most abundant protein found was flagellin, the main

component of the body of a flagellum. This result is consistent with *C. difficile* being a motile bacterium, and the extraction of surface components from the cells. Flagella are commonly associated with some strains of *C. difficile*, and both flagellin and the flagellar cap protein (not detected in this study) have been linked to the virulence of the organism by aiding attachment (Tasteyre, 2001). This analysis has shown 8 out of 30 paralogues of the SLP proteins to be present on the cell surface under these conditions. Many of the paralogues were found in both types of extract, but several (Slp paralogues 4, 6, 12 and 24) were only found in one type of extract, which may reflect differences in the way these paralogues are attached to the underlying cell wall. Slp paralogue 4 or Cwp84, has been shown to have cysteine protease activity (Savariau-Lacomme *et al.*, 2003), (see also Chapter 7), and although the function of Slp paralogue 2 is unknown, large quantities of it have been detected in culture supernatants (Mukherjee *et al.*, 2002). These data show for the first time in many cases that SLP paralogues are expressed on the surface of the bacterium, and may therefore have important and independent functions. Interestingly, the Slp paralogue and virulence factor Cwp66 was not found in these extracts. This adhesin was discovered in heat-shocked bacteria (Waligora *et al.*, 2001), so it is possible that this study has missed the detection of more virulence factors that are induced by a stress response, and may not be expressed under the growth conditions used here.

The proteomic analysis showed that the cell wall extract in particular contains many proteins without a known signal sequence or dedicated export system. This may indicate that some of the bacteria have lysed under these conditions, despite the difficulties in removing the complete cell wall and the osmotically balanced buffer in which this procedure was carried out. One such protein was DnaK, a chaperone protein. Even though they do not contain any known signal sequence, chaperones are commonly found in many extracellular extracts (Schaumburg *et al.*, 2004), and it has been speculated that they may have a vital function at the cell surface, aiding the folding of proteins as they are exported through the cell wall. Interestingly, a recent study in *Listeria monocytogenes* detected a number of proteins expected to be located in the cytoplasm in cell wall extracts, despite no cell lysis being detected. A number of these "moon-lighting" proteins were able to bind components of the extracellular matrix (ECM), thereby contributing to the virulence of the bacterium (Schaumburg *et al.*, 2004). Enolase was detected in the lysozyme extract of *C. difficile*, where it is possible a similar dual "moon-lighting" role may take place. However, this extract contains many proteins that would be expected to function inside the cell, and even though there are far less bands than a total lysate, it is possible that a small amount of lysis

occurred. Unfortunately, this hypothesis could not be proved by any of the assays for cell lysis used in this study.

The expression and purification of large quantities of recombinant SLPs from *E. coli* has facilitated studies into the structure and interactions between these proteins. The LMW SLP is predicted by bioinformatics to contain a substantial proportion (31%) of β-sheet in its structure. This was confirmed by CD analysis, which showed 34% of the protein was β-sheet. In addition, 18% of the structure was predicted to be α-helical, compared to 10% found experimentally by CD analysis. However, bioinformatic and CD analyses were not consistent in assessing the secondary structure of the HMW SLP. Bioinformatics estimates 42% of the structure to be composed of α-helix, compared to 14.6% found by CD. CD analysis estimates a higher proportion of β-sheet structure, with 27.7%, as opposed to bioinformatics, which estimates 18%. The SAXS analysis performed on these proteins did not give any reliable information regarding the shape of the SLPs, and crystallisation trials were not successful.

Prior to this study, little was known about the interactions of the SLPs, both with the cell wall, and with each other. An interesting observation was that when trying to separate glycine extracted SLPs from *C. difficile* using gel filtration, a large heterologous multimer composed of both SLPs eluted first, followed by homologous LMW SLP multimer. Gel filtration estimated the molecular weight of the LMW SLP-HMW SLP complex to be approximately 360 kDa, and native PAGE confirmed this, sizing the complex as greater than 150 kDa. When visualising this protein mixture by SDS-PAGE, both LMW- and HMW SLP appear to be present in roughly equimolar concentrations. Assuming the SLPs aggregate with a 1:1 ratio, these sizes could represent a heterotrimer (243 kDa) or heterotetramer (324 kDa). The smaller complex composed of only LMW SLP was sized by gel filtration as approximately 109 kDa, which is consistent in size with a trimer. These complexes were observed when the SLPs had been eluted from the gel filtration column into a non-ionic HEPES buffer. The S-layer structure is known to be composed of morphological units, composed of several identical protein subunits, that arrange to form the two dimensional array (Figure 1.11). Therefore it is possible that the interacting proteins make up the morphological units that then assemble to form the lattice. Hence, the two superimposed S-layers with different symmetries found in *C. difficile* (Cerquetti *et al.*, 2000) could be composed of one layer of interacting HMW- and LMW SLPs and a second layer consisting of only LMW SLP that interacts with itself. The properties of the SLPs that allow them to interact in this duplex fashion are unknown, and can probably only be determined by their crystallisation.

This work represents the first description of the surface sub-proteome in *C. difficile*. A number of methods for the removal of proteins from the cell surface have been evaluated, leading to the development of a hypothesis for the mechanism of attachment of the S-layer to the cell wall. In addition, an efficient and reproducible procedure for the expression and purification of recombinant SLPs has been developed. Analysis of these proteins has revealed clues regarding the assembly and structure of the S-layer. Further studies utilising these procedures may yield more conclusive results. The establishment here of a robust procedure for the removal of surface layer proteins from *C. difficile* will facilitate further studies of the changing proteome. The range of proteins released by the glycine and lysozyme treatments was very different, but complementary, and so both of these methods will be utilised in further experiments. For the first time, the expression of a number of *C. difficile* proteins has been confirmed, in particular localising a number of Slp paralogues to the cell surface.

Although the surface sub-proteome for stationary phase cells has been mapped, this represents a 'snap-shot' view of *C. difficile*. To further understand how the bacterium may behave during the course of an infection, it is necessary to study this dynamic sub-proteome further. The changing expression of proteins under different conditions will give an insight as to how the bacterium is able to adapt to challenging environments and ultimately, cause disease. This is explored further in Chapter 5. The understanding of the proteins a bacterial cell presents on its surface may also allow for the rationale design of treatments and vaccines that target these critical entities.

Chapter 5

Proteomic analysis of the differential expression of *Clostridium difficile* surface proteins grown under a variety of conditions

5.1 Introduction

Compared to the relatively static genome, the proteome is a highly dynamic system. Changes in the expression of genes leading to the production of functional proteins allow an organism to adapt and survive in a constantly changing environment. Proteomics is a suitable tool for measuring the biological changes in an organism's cellular structure and has been used to investigate differences between bacterial strains and the effects of metabolic changes induced by nutrient limitation, growth phase changes, temperature or the presence of inhibitory compounds such as antibiotics.

5.1.1 Growth phase induced changes in the proteome

The proteome has been described for several organisms at different stages of their growth. The changes in cytoplasmic proteins during growth of *S. aureus* cells have been thoroughly mapped (Kohler *et al.*, 2005), resulting in the identification of 1123 proteins, representing two-thirds of the expected cytoplasmic proteome. Comparison of the cytoplasmic proteome between exponentially growing and stationary phase cells revealed differences in the metabolic enzymes expressed at these times. Bacteria in the log phase of growth express much higher levels of enzymes involved in protein synthesis, transcription and glycolysis, whereas stationary phase cells express greater quantities of citric acid cycle and gluconeogenesis enzymes. A proteomic study of *Rhodopirellula baltica* at different stages of growth showed the opposing regulation of the citric acid cycle and oxidative pentose phosphate pathway (Gade *et al.*, 2005). Furthermore, during stationary phase, the enzymes involved in amino acid biosynthesis were down regulated while the alternative sigma factor σ^H and the gene products expressed under the control of this sigma factor were up regulated.

The growth phase of a cell also has an impact on the cell wall proteome, which has been studied in the probiotic organism *Lactobacillus salivarius* (Kelly *et al.*, 2005). In this study, three proteins were identified whose expression increased as the cells progressed from lag phase into log phase: the chaperone DnaK, the elongation factor Ef-Ts and pyruvate kinase. Although these proteins would be expected to be cytoplasmic, they were also identified in surface extracts from *Listeria monocytogenes* (Schaumburg *et al.*, 2004). As these studies used different extraction methods, and because the proteins were only detected under certain growth conditions, it is unlikely that they are cytoplasmic contaminants. However, the system by which they could be exported, and their likely function at the cell surface, remains unclear.

5.1.2 Stress-induced changes in the proteome

Bacillus subtilis has the best-studied proteome of the Gram-positive bacteria (reviewed in Sinchaikul *et al.*, (2002)). The cytosolic proteins have been comprehensively mapped (Buttner *et al.*, 2001) and furthermore, the proteins linked to the stress response have been further characterised by proteomics (Bernhardt *et al.*, 1997). This study found that upon exposure to various stresses, some stress-response proteins were induced generally, and others were specific to the type of stress. The expression of 42 general stress proteins required the alternative sigma factor σ^B, whereas seven were SigB independent. The authors also showed that under stressed conditions, the expression of stress proteins constituted up to 40% of the total protein synthesis of the cells, which underlines the importance of the stress regulon.

Further studies on *B. subtilis* examined the effect of antibiotics on exponentially growing cells (Bandow *et al.*, 2003). Each antibiotic treatment caused an individual effect on the differential expression profile, and overlaps in expression profiles were found to reflect similarities in molecular drug mechanisms. This finding was exploited to determine the mechanism of action of a novel antibacterial compound (a peptidyl transferase inhibitor, BAY50-2369), and this technique of proteomic expression profiling could be applied to new antibiotics to elucidate their mechanism of action.

The heat, oxidative and acid stress responses were characterised by proteomics for the intracellular pathogen *Brucella melitensis* (Teixeira-Gomes *et al.*, 2000). These conditions are particularly relevant for the survival of the bacteria within phagocytic cells, and therefore also for virulence. Depending on the type of stress the cells were subjected to, 6.4-12% of the proteome was differentially expressed. Each stress induced the expression of a specific group of proteins, with only one protein common to all stresses. Nineteen of the changing spots were identified, including BvrR (two-component signal tranducing pathway), DnaK and GroEL (chaperones) and superoxide dismutase.

Although the proteome of the surface proteins has been mapped for exponential phase *Clostridium difficile* under normal growing conditions (Chapter 4), the changing expression of proteins can be compared to decipher how *C. difficile* adapts to changing environments and causes disease. This can be achieved by building upon the techniques established in Chapter 4, using the two complementary methods to extract proteins from the surface of *C. difficile*, and separating the individual components by 2-DE. Comparison of protein gels will show differences between expression levels under different conditions, and these changing proteins can be identified by MS. Both the

glycine and lysozyme extraction procedures were used, as the data in Chapter 4 show that they produce different but complementary views of the cell surface. In this study, the effect on the surface proteome of growth phase, heat-shock and exposure to antibiotics to *C. difficile* cells has been examined. These conditions are particularly pertinent to the virulence status of *C. difficile*. The expression of TcdA and TcdB are induced as the bacteria enter the stationary phase of growth (Mukherjee *et al.*, 2002), and this expression may be accompanied by the induction of other virulence genes. Temperature also affects the expression of TcdA and TcdB (Karlsson *et al.*, 2003) and the expression of adhesins such as Cwp66 and GroEL (Eveillard *et al.*, 1993; Hennequin *et al.*, 2001a; Waligora, 2001). The response of *C. difficile* to the application of antibiotics may help to explain how the organism thrives in the presence of certain antibiotics, and why it is so recalcitrant to antibiotic treatment in some cases.

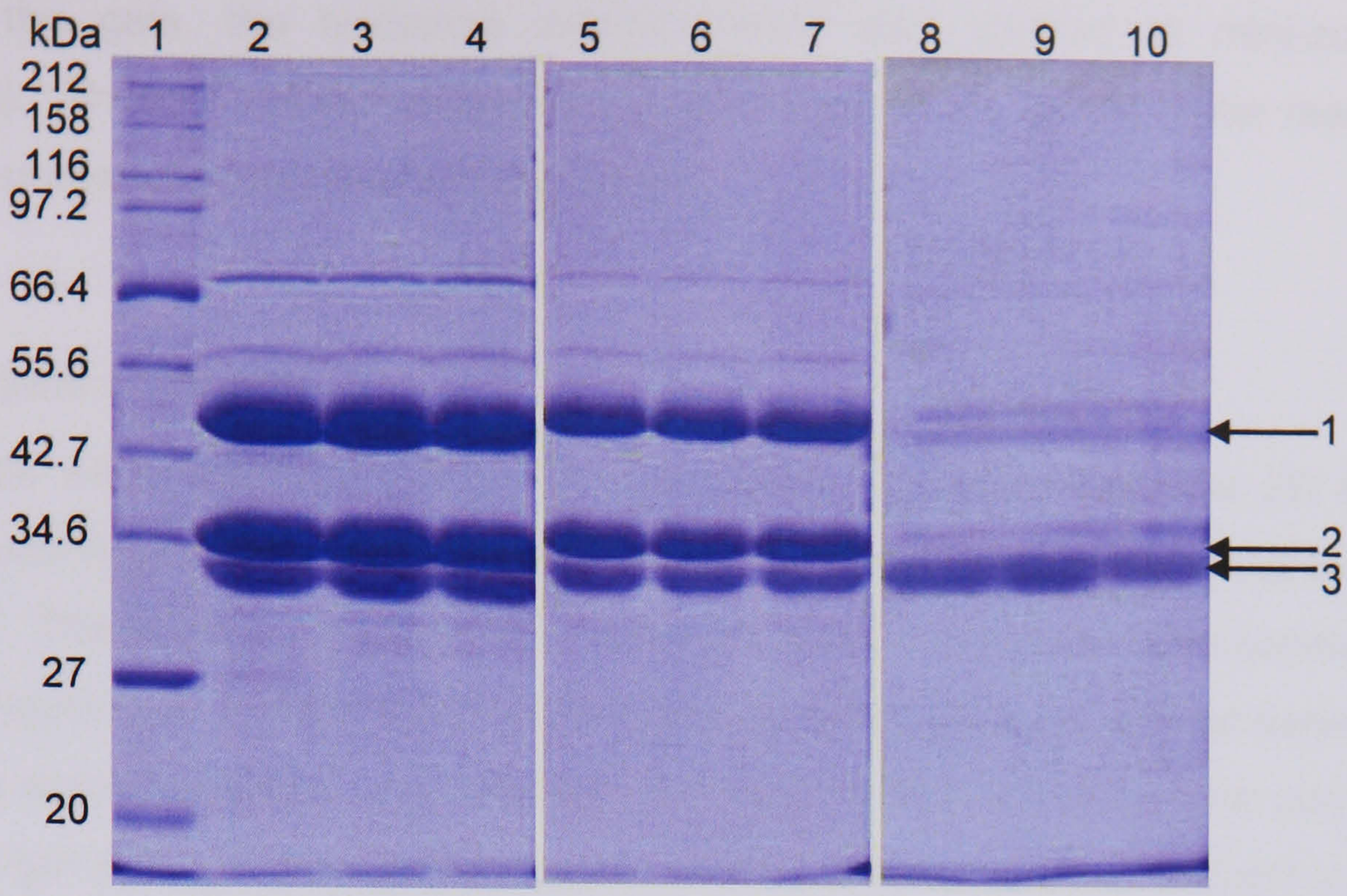
5.2 Results: Proteins differentially expressed on the surface of *C. difficile* at three growth phases

5.2.1 Sample preparation and SDS-PAGE

Three aliquots from a single culture were harvested after: 3 hours (representing early exponential phase); 6 hours (late exponential); and 24 hours (stationary phase). Glycine and lysozyme extractions (developed in Chapter 4) were carried out on each of the three time point samples. Extracts were first analysed by 1D SDS-PAGE (Figure 5.1). In all cases, the surface extracts were reproducible, consistently producing the same pattern of proteins. A stark difference between the time points, particularly noticeable in the glycine extracts, is that three dominant protein species are present during exponential growth but in stationary phase, only one is observed. These three protein bands, indicated by arrows on Figure 5.1A, were subjected to MALDI-TOF MS analysis. The largest molecular weight protein, of approximately 46 kDa and labelled as arrow 1, was identified as the HMW SLP, and the protein labelled as arrow 2 was identified as the LMW SLP. Neither of these protein bands can be seen in the extracts from the culture at stationary phase. The lowest dominant molecular weight protein, labelled as arrow 3, was identified as flagellin, and this protein appears to be unaffected by the growth phase. Other differences between the time points are also apparent, and therefore these extracts were analysed further, by 2D PAGE and MS.

Each of the 18 protein samples (50 µg) were subjected to IEF and 2D PAGE. Glycine extracts separated by 2D PAGE were stained with ruthenium II tris bathophenanthroline disulphonate (RuBPS) and visualised by fluorescent scanning. Proteins extracted by lysozyme treatment and separated by2-DE were stained with

A



B

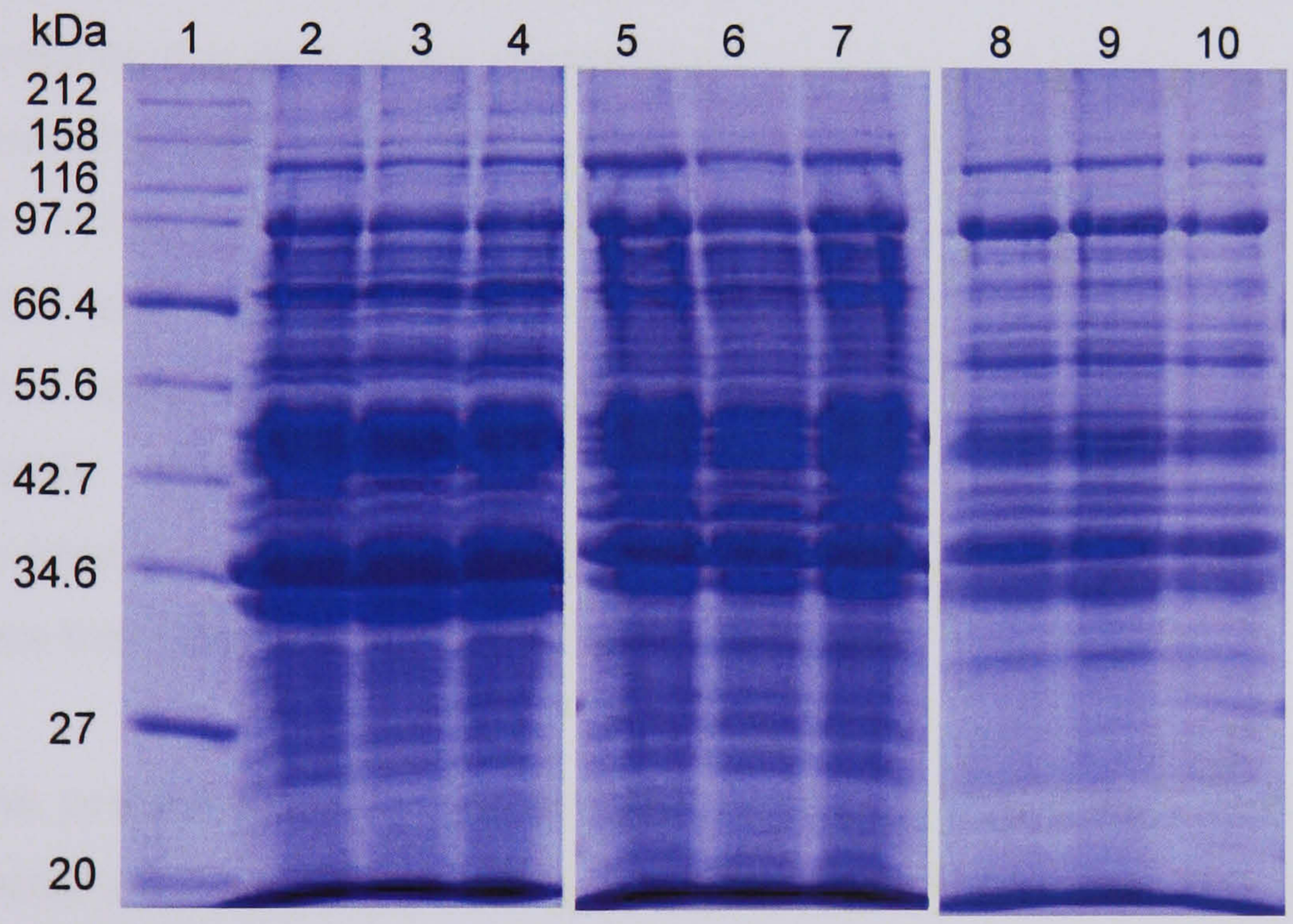


Figure 5.1. 1D SDS-PAGE analysis of surface proteins extracted from a time-course culture of *C. difficile* 630 by glycine and lysozyme treatment. Triplicate extractions for each time point were made from cells taken from a single, large culture. **A.** Glycine extracts; lane 1, molecular mass markers in kDa; lanes 2-4, 3 h culture; lanes 5-7, 6 h culture; lanes 8-10, 24 h culture. Arrows show the position of: 1, HMW SLP; 2, LMW SLP; and 3, FliC. **B.** Lysozyme extracts; lane 1, molecular mass markers in kDa; lanes 2-4, 3 h culture; lanes 5-7, 6 h culture; lanes 8-10, 24 h culture.

colloidal Coomassie blue. Due to the very heavy regions of staining at the lower pH range of the gels, the lysozyme extracts were also applied to mini-zoom gels (Invitrogen). These 7 cm IEF strips have a pH range of 4-7, giving better resolution of the large number of proteins within this range.

5.2.2 Analysis of 2D gels

Protein gels were analysed using the Non Linear Dynamics Phoretix 2D Evolution program, with manual processing. The glycine and lysozyme extracts were analysed separately. The strategy of analysis for each "Phoretix" experiment was constructed as shown in Figure 5.2. Protein spots were manually detected and matched between sub-gels of the same condition. An "average gel" was created for each time point, based on the sub-gel with the most protein spots. Protein spots had to be present on at least two of the three sub-gels to be marked onto the average gel. Each average spot contained statistical data from equivalent spots on all three sub-gels. The creation of average gels improves the statistical significance of the spots, as they must be present in the majority of sub-gels to be represented on the average gel. The three average gels representing each time point were used to construct a reference gel for the experiment. In the same way, the average gel with the most spots was used to visually represent the reference gel, although it contains data from each of the nine sub-gels. All spots from each average gel were marked onto the reference gel, and matched where possible. Creating a reference gel allows a straightforward comparison of all spots in an experiment, as it is implicit that if two spots in different gels are matched to the reference then they are also matched to each other.

Once all the protein spots on all the gels had been identified and matched, certain filtering criteria were applied to select for spots significantly changing in volume. The Phoretix programme contains a number of statistical tools by which spot data can be analysed. As the all the experiments in this study were multi-directional comparisons, an Analysis Of Variance (ANOVA) was applied. This statistical filter ensures that all matched spots from sub-gels are similar and do not vary significantly between replicate samples and gels. In this way, variations between the different conditions being tested must be due to the condition itself, rather than experimental variation. Spots with an ANOVA score of greater than 0.05 were excluded and not considered further.

5.2.3 Differential expression seen within glycine extracts

Representative gels for each time point are shown in Figure 5.3. Spots shown by the Phoretix analysis to be significantly changing in volume are circled on each gel. A total

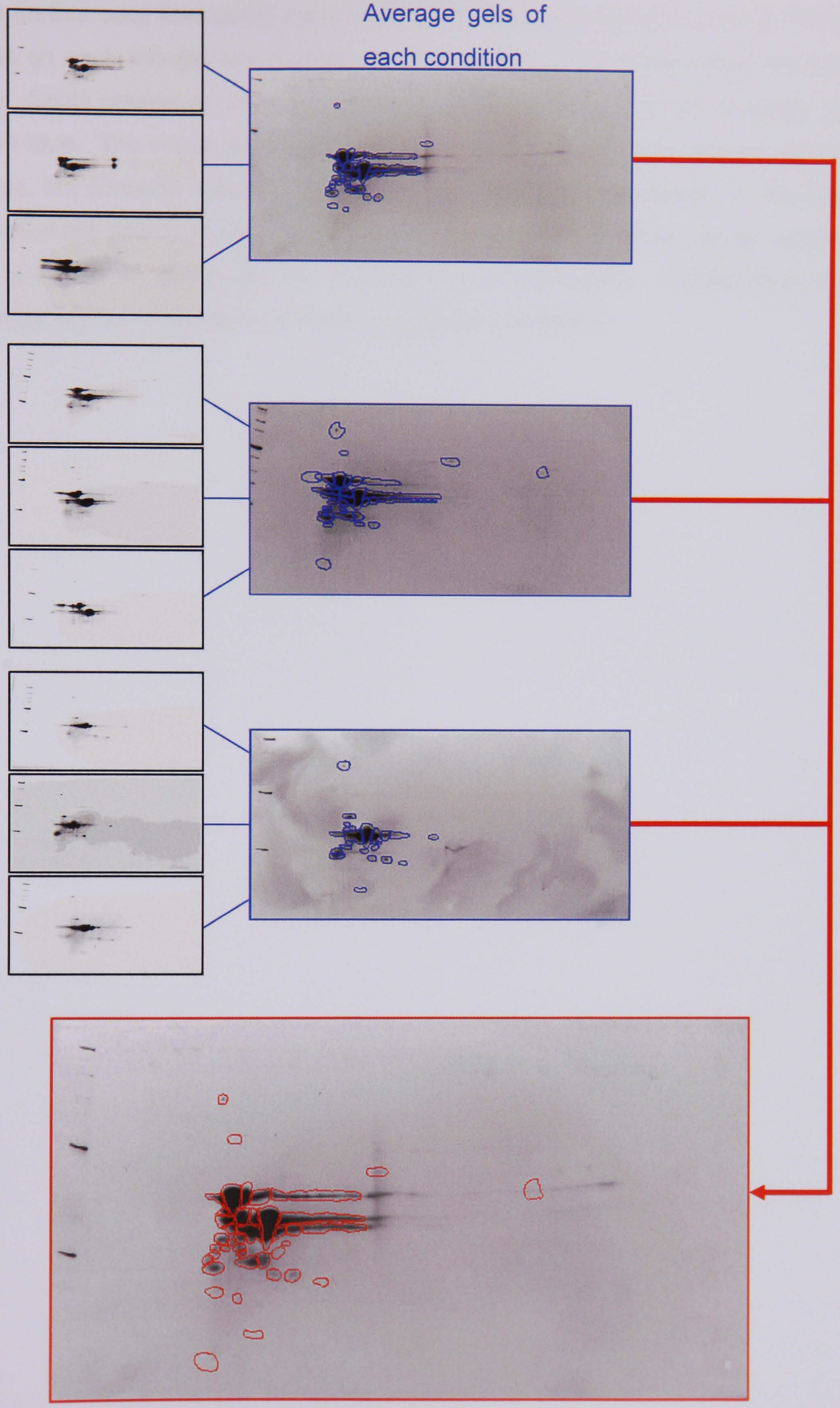


Figure 5.2 The construction of a Phoretix experiment. Legend continued overleaf.

Figure 5.2 The construction of a Phoretix experiment. The 2D gels from a glycine extracted time-course experiment are shown as an example. Glycine extracts of each condition (in this case time point) were made in triplicate and separated by 2D PAGE. The spots on each sub-gel are marked and then matched with others from the same condition. Spots present on at least 2 of the 3 gels are displayed on an "average" gel, outlined in blue. The visual display of an average gel is based on the sub-gel with the most spots, but contains statistical data from each sub-gel it represents. In the same way, the average gels from each condition are compiled into a reference gel (outlined in red), to which all spots can be compared and referenced. Construction of a reference gel allows comparison of spots over all the conditions.

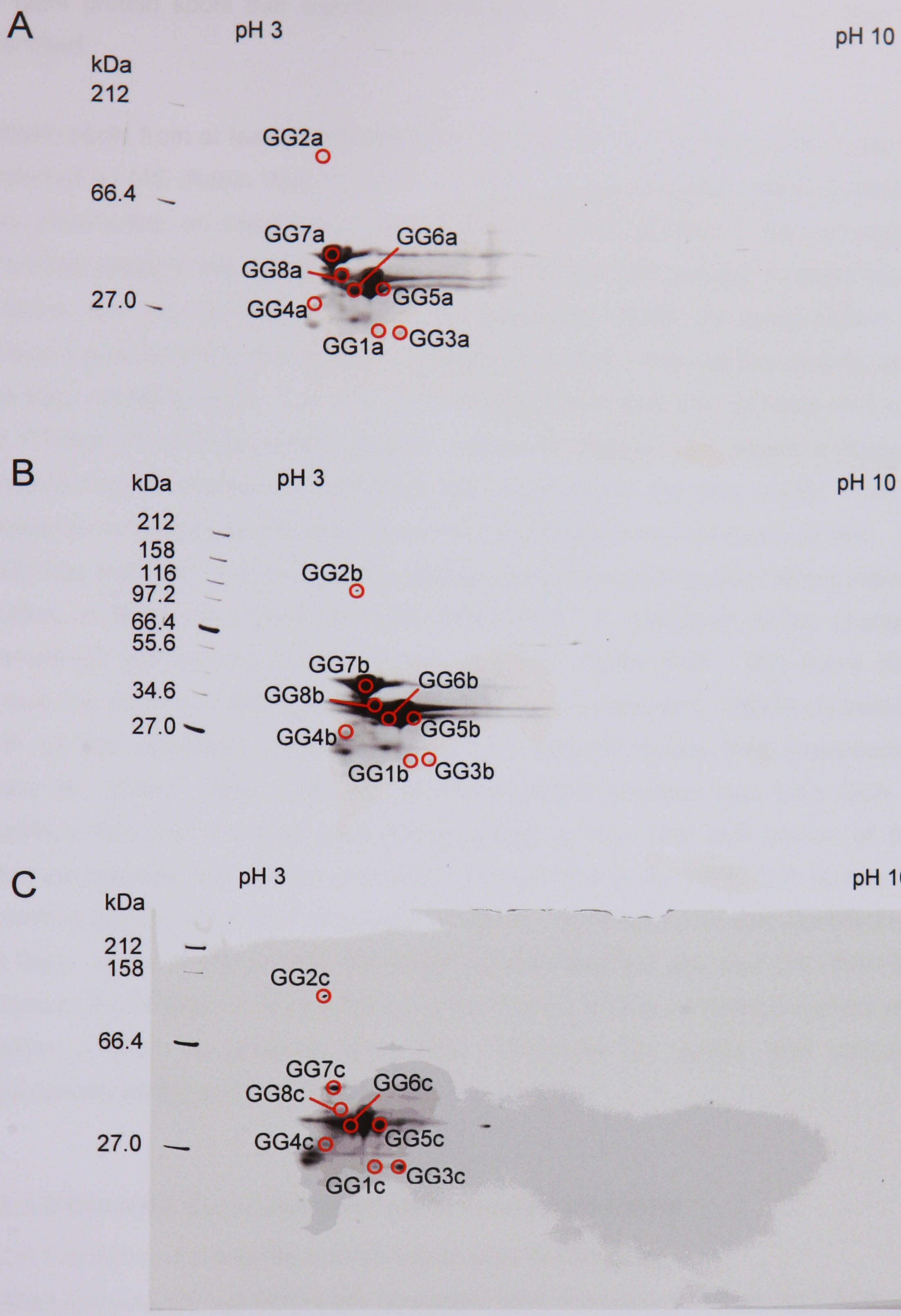


Figure 5.3 Average gels displaying proteins extracted from the surface of a growing culture of *C. difficile* 630 by glycine treatment. **A** Proteins extracted from a 3 h culture. **B** Proteins extracted from a 6 h culture. **C** Proteins extracted from a 24 h culture. The pH is indicated at the top of the protein gels and the molecular masses are shown in kDa on the left. Protein spots identified by MS are circled in red and labelled GG to indicate glycine growth experiment. The suffix 'a' is to show the spot is from the 3 h time point; 'b' is the 6 h time point and 'c' the 24 h time point.

of eight protein spots that significantly changed in expression level over time were identified.

Protein spots from at least two of the three time points were digested with trypsin and analysed by MS (Robin Wait, Imperial College). Of the 20 protein spots analysed in this experiment, an identification was obtained for 13 of them. The successfully identified proteins are shown in Table 5.1. Because the change in expression of proteins over time is being studied, it is important to identify the same protein from different gels to confirm that the spot matching is correct. This was the case for four of the eight changing spots. Spot GG3 was identified from gels from all three time points as CD3642, a putative selenocysteine synthase. Figure 5.4A shows a histogram representing the change in normalised spot volume over the time points. The spot increases in volume as the cells move from exponential into stationary phase. Spot GG7 was identified from gels representing all three time points as the carboxyl-terminal portion of SlpA, corresponding to the HMW SLP. A histogram of the change in normalised spot volume of spot GG7 is shown in Figure 5.4B. This figure clearly shows the amount of HMW SLP extracted from the surface of *C. difficile* by treatment with glycine decreases as the growth of the bacteria moves from exponential to stationary phase. Spot GG8 was found to contain peptides also from SlpA, but peptides from the 3 h time point corresponded to the LMW SLP portion of SlpA, whereas peptides from the 6 h time point corresponded to the HMW SLP portion. It is therefore unclear which SLP this spot represents. Spot GG4 was also identified twice as SlpA. In the 3 h time point, the single peptide identified was from the HMW SLP, however, the MS/MS on the protein spot from the 6 h time point failed to identify which portion of SlpA the peptides came from. Therefore this protein spot cannot be conclusively identified either SLP.

5.2.4 Differential expression seen within lysozyme extracts

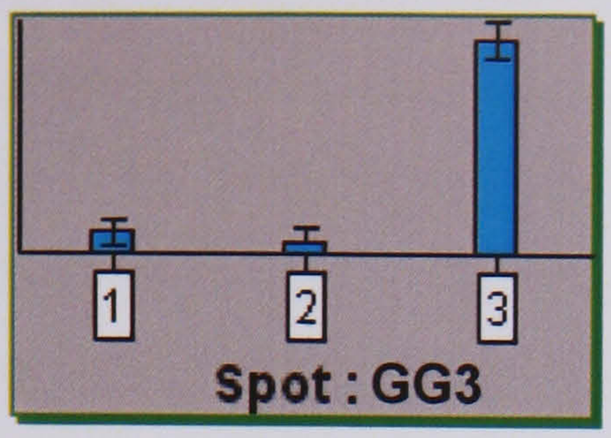
5.2.4.1 Analysis of lysozyme extracts using large format gels

Protein samples (50 μg) from each time point were analysed on 11 cm 2D PAGE gels. The protein was visualised by RuBPS staining, however a very heavy staining region obscured the individual protein spots. To counter this, the gels were restained with the less sensitive colloidal Coomassie blue, which improved the resolution of the spots. The average gels for each time point are shown in Figure 5.5. Spots shown by the Phoretix analysis to be significantly changing in volume are circled on each gel. A total of eleven protein spots that significantly changed in expression level over time were identified.

Spot	Genome ID	Description	Protein Probe Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
GG2b	CD0514	SLP paralogue 15	77.53	2.21	3(36)	167580	4.7	√
GG3a	CD3642	Putative selenocysteine synthase	11.89	3.52	1	40516	5.6	x
GG3b	CD3642	Putative selenocysteine synthase	3.52	20.09	1(13)	40516	5.6	x
GG3c	CD3642	Putative selenocysteine synthase	3.52	17.95	1(13)	40516	5.6	x
GG4a	CD2793	SlpA (LMW SLP)	24.03	1.95	1(14)	76115 (34228)	4.9	√
GG4b	CD2793	SlpA	3.75	8.90	1	76115	4.6	√
GG5a	CD3642	Putative selenocysteine synthase	19.57	3.52	1(13)	40516	5.6	x
	CD0239	FliC Flagellin	14.94	8.62	1(25)	30755	8.6	x
GG5b	CD2793	SlpA (HMW SLP)	1(27)	12.42	3.75	76115 (39504)	4.6	√
GG7a	CD2793	SlpA (HMW SLP)	75.11	5.56	3(40)	76115 (39504)	4.6	√
	CD3525	Putative iron ABC transporter, solute-binding protein	12.81	3.94	1(14)	39154	5.4	√
GG7b	CD2793	SlpA (HMW SLP)	202.99	21.84	11(180)	76115 (39504)	4.6	√
	CD3525	Putative iron ABC transporter, solute-binding protein	13.19	3.94	1(14)	39154	5.4	√
GG7c	CD2793	SlpA (HMW SLP)	1(14)	36.55	1.95	76115 (39504)	4.6	√
GG8a	CD2793	SlpA (HMW SLP)	22.70	1.11	1(8)	76115 (39504)	4.6	√
GG8b	CD2793	SlpA (LMW SLP)	42.96	9.74	3(70)	76115 (34228)	4.9	√

Table 5.1 Proteins identified from 2D gels of glycine extracted proteins from cultures at different stages of growth. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Protein Probe score determined by Protein Lynx Global Server (Micromass). Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

A



B

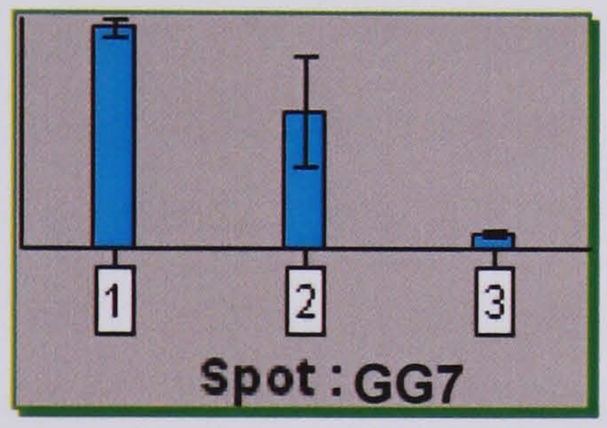


Figure 5.4 Histograms of identified protein spots changing over time. **A**. Spot GG3 identified as CD3642, a putative selenocysteine synthase. **B**. Spot GG7, identified as CD2793, the HMW SLP. The histograms show the normalised spot volumes. Bar 1 represents the 3 h time point, bar 2 represents the 6 h time point and bar 3 represents the 24 h time point. Error bars show the variance between the spot volumes over the three sub-gels

Protein spots from at least two of the three time points were digested with trypsin and analysed by MS (Jerry Thomas, University of York). Of the 22 spots analysed from this experiment, an identification was obtained for 21 of them. The successful identifications are shown in Table 5.2. Only two spots were identified as the same protein from separate gels, spots GL3 and GL8. Spot GL3 was identified from a 3 h and 6 h time point gel as the HMW SLP. A histogram depicting the change in spot volume is shown in Figure 5.6A. The relative abundance of this HMW SLP spot is decreasing as the age of the culture increases. The histogram of spot GL8 is shown in Figure 5.6B. This spot was identified from two separate gels of the 3 h time point sample, as gels from the other time points contained insufficient protein to extract. Spot GL8 was identified as Slp paralogue 2. This protein also appears to be decreasing in volume as the age of the culture increases.

5.2.4.1 Analysis of lysozyme extracts using zoom gels

Due to the difficulties in distinguishing proteins on the 11cm format gels, 20 μg of the same protein samples were applied to 7 cm 2-DE gels, using IEF strips with a narrower pH range, from 4 to 7, known as a zoom gel. The pH range of 4-7 contains the majority of proteins in the sample, and the increase in area over which the proteins are separated increases resolution of protein spots. The separated proteins were visualised using RuBPS in conjunction with a fluorescent scanner and the spot patterns analysed using Phoretix 2D Evolution, with the same experimental design as described previously. Representative gels from each condition are shown in Figure 5.7. While fewer spots are visible on these gels compared to the 11 cm gels, the separation of protein spots was improved. Sixteen spots were confidently found to be changing in volume.

Protein spots from at least two of the three time points for the majority of spots were digested with trypsin and analysed by MS (Jerry Thomas, University of York). All samples that underwent MS/MS analysis (30) were identified. The identities of each spot are displayed in Table 5.3, with details of the quality of each match indicated by an increasing total ion score. Of the sixteen spots changing in volume, twelve were identified as the same protein from more than one gel. The first five spots were all identified as SlpA, with peptides corresponding to the HMW SLP. These five spots have a pI close to 4, and range in size from approximately 10 to 30 kDa. The spots are best seen on the 6 h time point gels where they are strongly reproducible (data not shown). The spots are distinct, but also similar in size and pI, and could represent degradative or post-translationally modified forms of the HMW SLP. The histograms

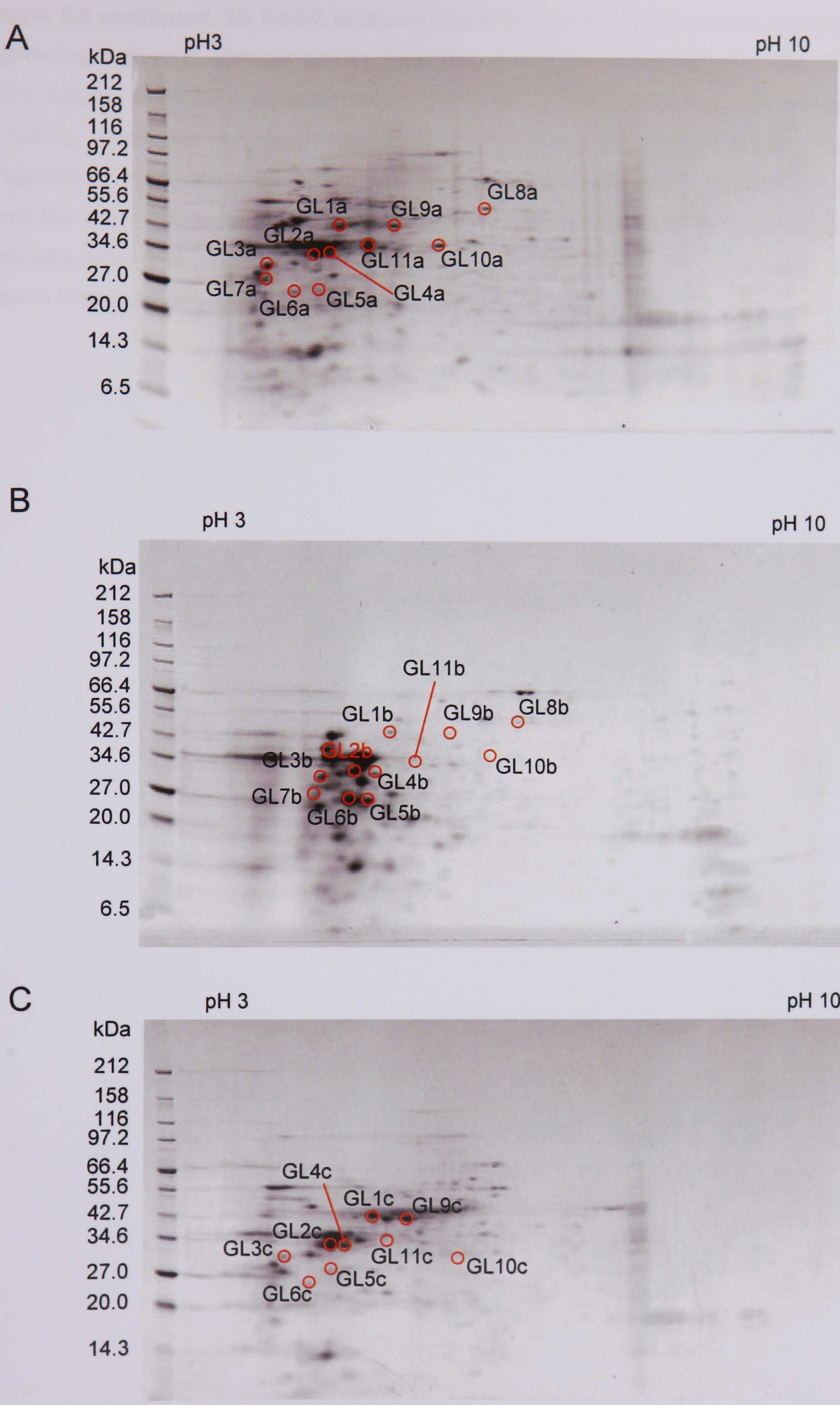


Figure 5.5 2D PAGE analysis of proteins extracted from the surface of a growing culture *C. difficile* 630 by lysozyme treatment. Legend continued overleaf.

Figure 5.5 continued. 2D PAGE analysis of proteins extracted from the surface of a growing culture *C. difficile* 630 by lysozyme treatment. **A** Proteins extracted from a 3 h culture. **B** Proteins extracted from a 6 h culture. **C** Proteins extracted from a 24 h culture. Molecular masses in kDa are indicated on the left and pH above. Spots judged to be differentially expressed are circled in red. The spots are numbered and have the prefix GL to show they are from the growth experiment and extracted by lysozyme treatment. The suffix 'a' is to show the spot is from the 3 h time point; 'b' is the 6 h time point and 'c' the 24 h time point.

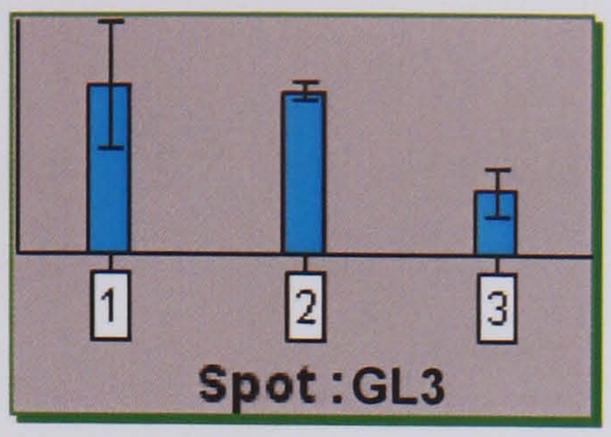
Table 5.2 Proteins identified from 2D gels of lysozyme extracted proteins from cultures at different stages of growth. Spots are labelled according to Figure 5.5, with suffix 'a' indicating 3 h time point, 'b' indicating 6 h time point and 'c' indicating 24 h time point. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). The total ion score was determined by Applied Biosystems. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
GL1b	CD2672	AppA, oligopeptide ABC transporter	337	4.75	6 (74)	58164	5.6	√
GL1c	CD3664	Putative amino acid transferase	371	9.53	9 (114)	44798	5.2	x
GL2a	CD0239	FliC, flagellin	487	10.55	5 (92)	30774	8.6	x
GL2b	CD2793	SlpA, LMW SLP	970	40.84	9 (145)	76115 (34228)	4.9	√
GL3a	CD2793	SlpA, HMW SLP	637	5.38	6 (97)	76115 (39504)	4.6	√
GL3b	CD2793	SlpA, HMW SLP	681	5.99	7 (108)	76115 (39504)	4.6	√
GL4b	CD2793	SlpA, LMW SLP	111	4.23	1 (15)	76115 (34228)	4.9	√
GL4c	CD0239	FliC, flagellin	463	11.70	6 (102)	30774	8.6	x
GL5b	CD2793	SlpA, LMW SLP	787	34.08	9 (121)	76115 (34228)	4.9	√
GL5c	CD1055	EtfB2, electron transport flavoprotein	101	2.89	2 (23)	28909	4.8	x
GL6b	CD2793	SlpA, HMW SLP	365	5.60	5 (101)	76115 (39504)	4.6	√
GL6c	CD0400	EtfB1, electron transport flavoprotein	168	2.80	2 (22)	28671	4.9	x
GL7a	CD0239	FliC, flagellin	409	8.26	4 (72)	30774	8.6	x

Table 5.2 continued

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
GL7b	CD2793	SlpA, HMW SLP	331	3.44	4 (62)	76115 (39504)	4.6	√
GL8a	CD2791	Slp paralogue 2	303	5.45	8 (102)	66446	6.4	√
GL8a	CD2791	Slp paralogue 2	45	0.86	1 (16)	66446	6.4	√
GL9a	CD1175	AckA acetate kinase	234	5.18	5 (62)	43321	5.4	x
GL9c	CD1059	ThlA1 acetyl coA acetyl transfersae	291	5.02	6 (59)	40861	5.4	√
GL10a	CD3174	Glyceraldehyde 3-phosphate dehydrogenase	478	10.13	8 (102)	36034	5.9	x
GL10b	CD2672	AppA, oligopeptide ABC transporter	295	4.33	6 (69)	58164	5.6	√
GL11a	CD3174	Glyceraldehyde 3-phosphate dehydrogenase	539	10.13	8 (102)	36034	5.9	x

A



B

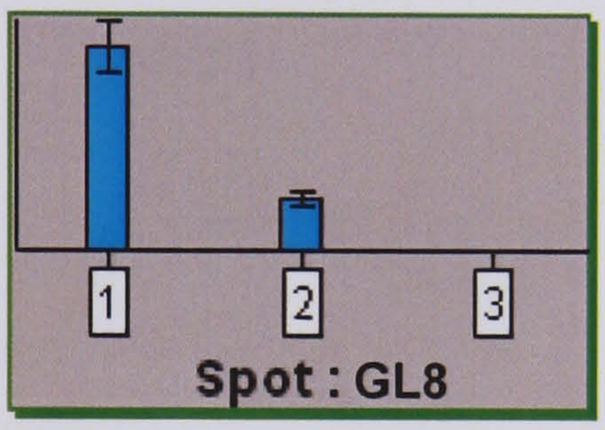


Figure 5.6 Histogram of identified protein spots changing over time. A. Spot GL3 identified as CD2793, SlpA, corresponding to the HMW SLP. **B.** Spot GL8, identified as CD2791, Slp paralogue 2. The histograms show the normalised spot volumes. Bar 1 represents the 3 h time point, bar 2 represents the 6 h time point and bar 3 represents the 24 h time point. Error bars indicate variation over the sub-gels.

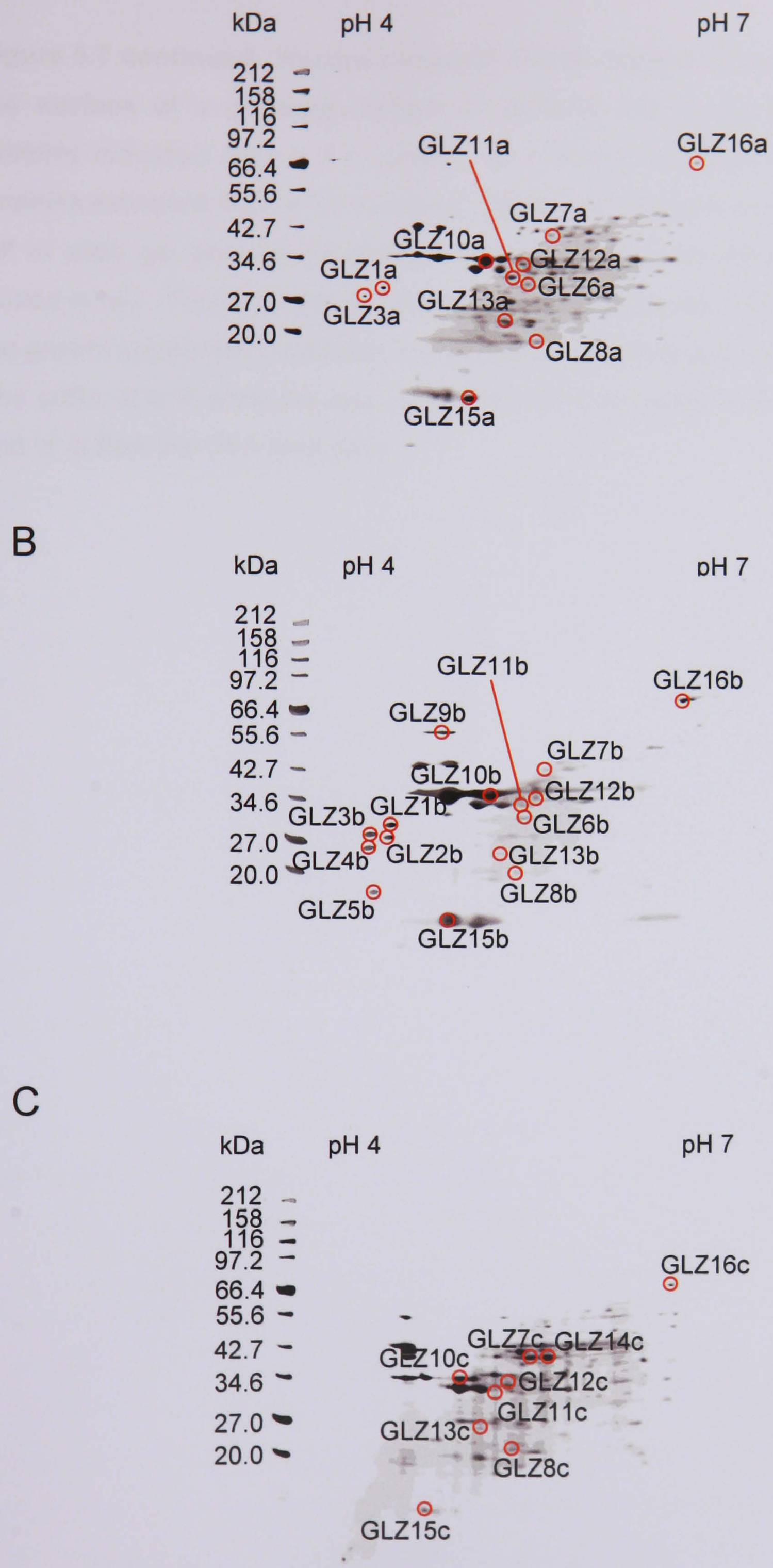


Figure 5.7 Narrow range 2D PAGE analysis of proteins extracted from the surface of a growing culture *C. difficile* 630 by lysozyme treatment. Legend continued overleaf.

Figure 5.7 continued. Narrow range 2D PAGE analysis of proteins extracted from the surface of a growing culture *C. difficile* 630 by lysozyme treatment. A. Proteins extracted from a 3 h culture. **B.** Proteins extracted from a 6 h culture. **C.** Proteins extracted from a 24 h culture. Molecular masses (kDa) are indicated on the left of each gel and the pH above. Spots judged to be differentially expressed are circled in red. The spots are numbered and have the prefix GLZ to show they are from the growth experiment, extracted by lysozyme treatment and visualised on a zoom gel. The suffix 'a' is to show the spot is from the 3 h time point, 'b' is from the 6 h time point and 'c' is from the 24 h time point.

showing the change in volume of spots GLZ1-5 over the growth curve are displayed in Figure 5.8A-E. All five spots have a maximal volume after 6 h growth. Spots GLZ1 and GLZ3 are present in lower quantities after 3 h growth, but all spots decrease in volume after 24 h growth. This finding is consistent with the glycine extracted samples and the 11 cm format gels, all of which show spots containing the HMW SLP decreasing in volume over time.

Spot GLZ15 was identified from the 3 h and 6 h extracts as SlpA. Rather curiously, the peptides sequenced from both of these spots originated from both the LMW and HMW SLPs, specifically the region around the predicted SlpA cleavage site at 355 residues. Spot GLZ15 has a molecular weight of less than 20 kDa, and so could represent a degradation product of SlpA. As the mechanism by which SlpA is post-translationally cleaved remains to be elucidated, it is unclear how this modified protein could arise. Figure 5.8I shows a histogram of the spot volume. The spot appears to increase in volume over exponential phase, and then decreases as the cells enter stationary phase.

Spot GLZ16 was identified as Slp paralogue 2 (Figure 5.8J). This spot is also present at all stages of growth, but reaches a maximum at the end of exponential stage. This confirms the same finding on the large format gels of the same extracts.

Spot GLZ6 was identified twice (albeit weakly each time) as CD2365, a putative sulphonate ABC transporter solute binding lipoprotein. This protein is predicted to contain a signal sequence and its putative function would require a cell surface location. The histogram of the spot volume at the different time points is shown in Figure 5.8F. Greater quantities of CD2365 are expressed during exponential phase, decreasing after 24 h growth. Spot GLZ8 was identified as CD3305, encoding ClpP1, an ATP-dependent Clp protease proteolytic subunit. CD3305 was identified from gels from the 3 h and 24 h time points, and appears to be present at all stages of growth (Figure 5.8G). Spot GLZ9 can only be seen on gels of the 6 h extract, and it was identified on replicate gels of this condition as CD2672, an oligopeptide ABC transporter substrate binding protein. This protein was also identified on the large format gels, but not in duplicate for the same spot. Spot GLZ11 was identified on replicate gels from the 3 h extract as CD2321, encoding Tkt. This protein is similar to the carboxyl-terminal domain of eukaryotic transketolase, a protein that is normally present as two separate proteins in prokaryotes. Spot GLZ11 can be seen at all stages of growth (Figure 5.8H), but with the greatest volume during early exponential growth.

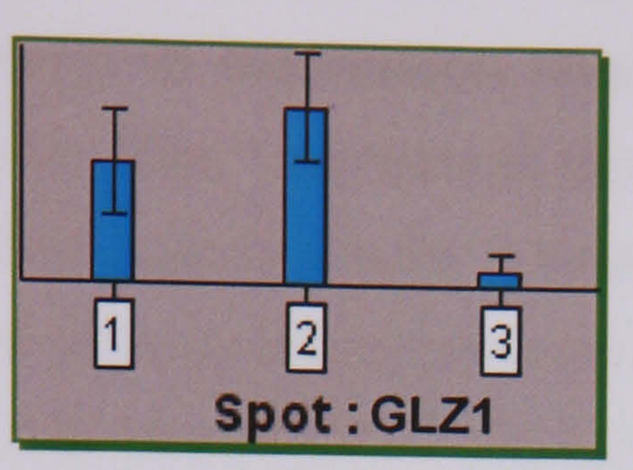
Table 5.3 Proteins identified from Zoom 2D gels of lysozyme extracted proteins from cultures at different stages of growth. Spots are labelled according to Figure 5.7, with suffix 'a' indicating 3 h time point, 'b' indicating 6 h time point and 'c' indicating 24 h time point. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). The total ion score was determined by Applied Biosystems. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
GLZ1a	CD2793	SlpA, HMW SLP	830	6.93	8 (125)	76115 (39504)	4.6	√
GLZ1b	CD2793	SlpA, HMW SLP	508	4.27	5 (77)	76115 (39504)	4.6	√
GLZ2b	CD2793	SlpA, HMW SLP	498	4.27	5 (77)	76115 (39504)	4.6	√
GLZ2b	CD2793	SlpA, HMW SLP	393	3.05	4 (55)	76115 (39504)	4.6	√
GLZ3a	CD2793	SlpA, HMW SLP	574	5.76	6 (104)	76115 (39504)	4.6	√
GLZ3b	CD2793	SlpA, HMW SLP	982	7.48	8 (135)	76115 (39504)	4.6	√
GLZ4b	CD2793	SlpA, HMW SLP	708	5.99	7 (108)	76115 (39504)	4.6	√
GLZ4b	CD2793	SlpA, HMW SLP	739	6.26	7 (113)	76115 (39504)	4.6	√
GLZ5b	CD2793	SlpA, HMW SLP	611	4.71	6 (85)	76115 (39504)	4.6	√
GLZ5b	CD2793	SlpA, HMW SLP	550	4.05	5 (73)	76115 (39504)	4.6	√
GLZ6a	CD2365	Putative sulphonate ABC transporter solute binding lipoprotein	48	0.92	1 (10)	39896	5.8	√
GLZ6b	CD2365	Putative sulphonate ABC transporter solute binding lipoprotein	49	0.09	1 (11)	39896	5.8	√

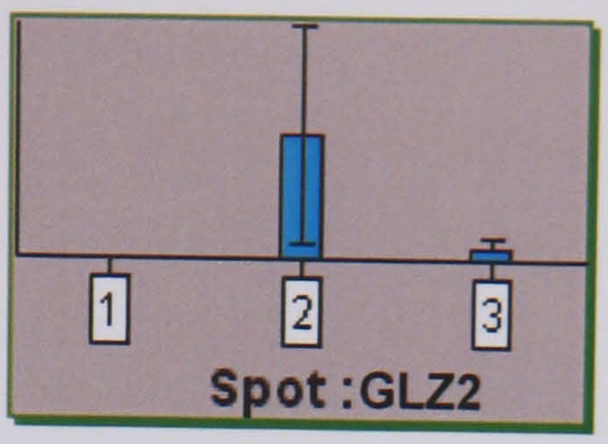
Table 5.3 continued

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
GLZ8a	CD3305	ClpP1, ATP dependent Clp protease proteolytic subunit	148	3.77	2 (22)	21295	5.1	x
GLZ8c	CD3305	ClpP1, ATP dependent Clp protease proteolytic subunit	66	1.88	1 (11)	21295	5.1	x
GLZ9b	CD2672	AppA, oligopeptide ABC transporter	208	2.95	4 (47)	58164	5.6	√
GLZ9b	CD2672	AppA, oligopeptide ABC transporter	311	3.57	5 (57)	58164	5.6	√
GLZ10c	CD0873 CD2793	ABC transporting lipoprotein, substrate binding protein SlpA, LMW SLP	417 143	10.57 10.14	7 (108) 3 (36)	36067 76115 (34228)	5.4 4.9	√ √
GLZ11a	CD2321	Tkt', similar to eukaryotic transketolase	48	1.09	1 (10)	32342	5.1	x
GLZ11a	CD2321	Tkt', similar to eukaryotic transketolase	197	4.46	3 (41)	32342	5.1	x
GLZ12a	CD0394	LdhA, lactate dehydrogenase	46	1.00	1 (10)	36518	5.0	x
GLZ12b	CD0239	FliC, flagellin	296	5.50	3 (48)	30755	8.6	x
GLZ13a	CD0071/ CD0058 (identical)	TufB, elongation factor Tu	79	1.17	1(14)	44027	4.6	x
GLZ13a	CD0070	FusA, translation elongation factor G	279	2.81	6 (58)	75897	4.9	x
GLZ14c	CD1059	ThlA1, acetyl-CoA acetyl transferase	534	6.72	7 (79)	43171	5.4	√
GLZ14c	CD1059	ThlA1, acetyl-CoA acetyl transferase	413	5.11	6 (60)	43171	5.4	√
GLZ15a	CD2793	SlpA, both SLPs	555	4.45	6 (96)	76115	4.8	√
GLZ15b	CD2793	SlpA, both SLPs	336	1.99	4 (43)	76115	4.8	√
GLZ16a	CD2791	Slp paralogue 2	167	2.35	3 (44)	66446	6.4	√
GLZ16b	CD2791	Slp paralogue 2	132	1.60	2 (30)	66446	6.4	√

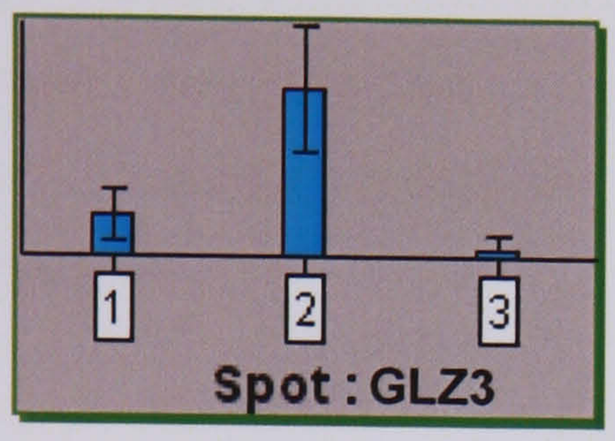
A



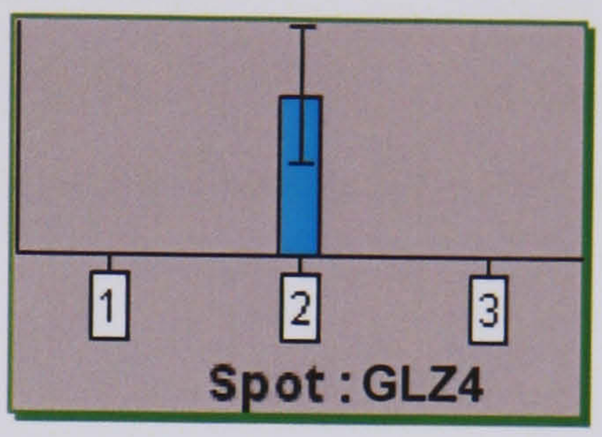
B



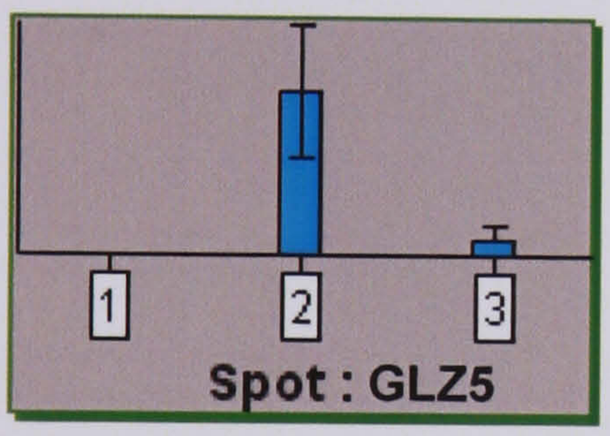
C



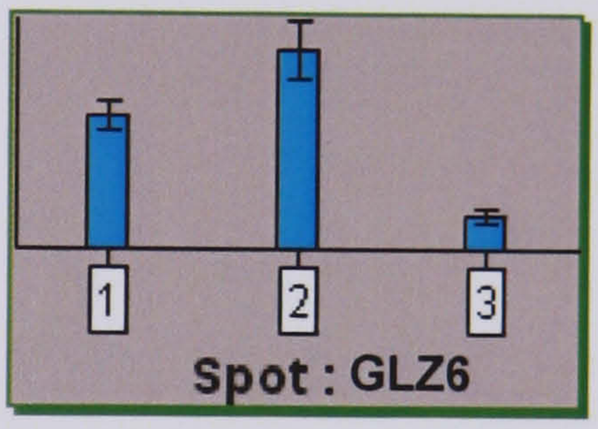
D



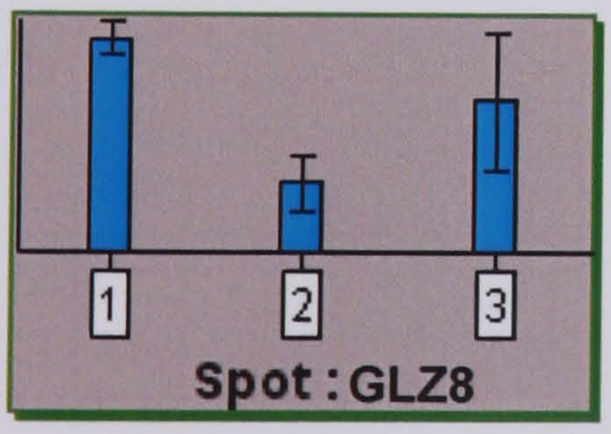
E



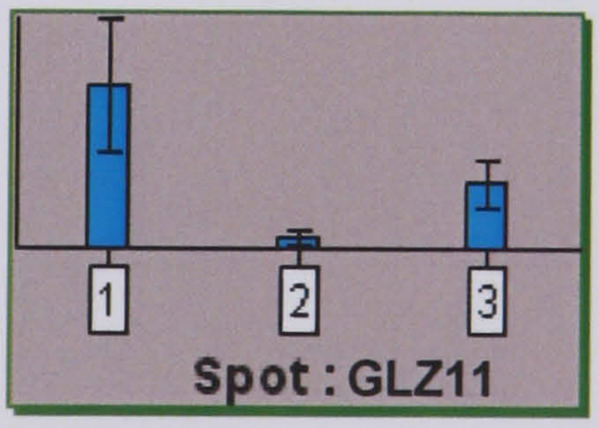
F



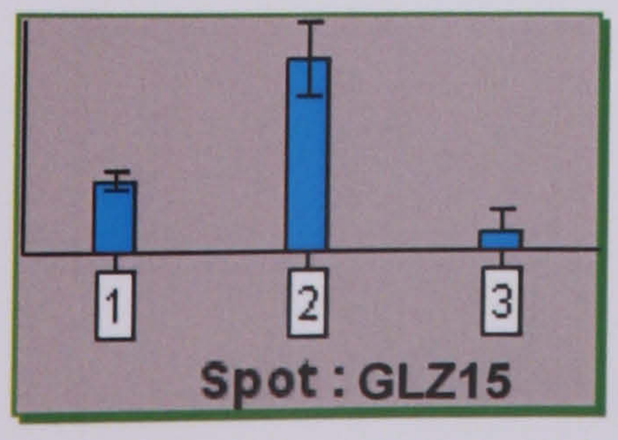
G



H



I



J

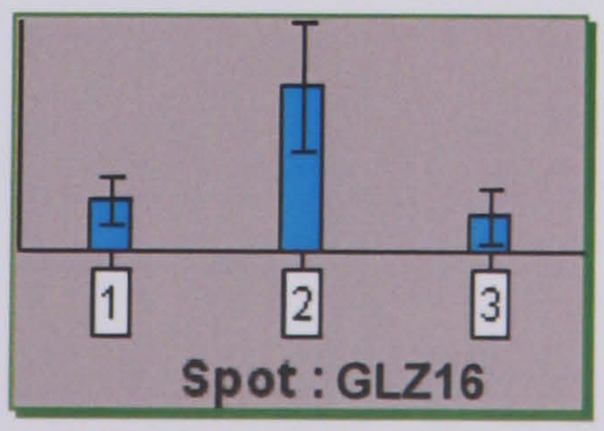


Figure 5.8 Histograms of spot volumes of identified proteins changing in expression levels during growth. Legend continued overleaf.

Figure 5.8 continued. Histograms of spot volumes of identified proteins changing in expression levels during growth. Histograms show normalised spot volumes. Bar 1 represents the 3 h time point, bar 2 represents the 6 h time point and bar 3 represents the 24 h time point. Proteins were extracted from the surface of *C. difficile* cells by lysozyme treatment and separated on zoom 2D gels with a pH range of 4-7. **A-E** Spots GLZ1-5 all identified as CD2793, SlpA, and specifically the HMW SLP. **F** Spot GLZ6 identified as CD2365, a putative sulphonate ABC transporter. **G** Spot GLZ8 identified as ClpP1, an ATP-dependent Clp protease proteolytic subunit. **H** Spot GLZ11, identified as CD2321, encoding Tkt, which is similar to the carboxyl-terminus of eukaryotic transketolase. **I** Spot GLZ15 identified as CD1059, encoding ThlA1, an acetyl-CoA acetyl transferase. **J** Spot GLZ16 was identified as CD2791 encoding Slp paralogue 2.

The protein is not predicted to contain a signal sequence, and it is unclear what the function of this protein may be in the cell surface extract. Spot GLZ14 was only detected in the 24 h extracts, and was identified in duplicate as CD1059, which encodes ThlA1, an acetyl-CoA acetyl transferase. This protein is predicted to contain a signal sequence and is therefore probably exported from the cell. Spot GLZ13 was identified from replicate 3 h extract gels as CD0070 and CD0071. Although the same identification was not obtained in duplicate, both of these adjacent genes encode translation elongation factors.

5.3. Effect of antibiotics on differentially expressed proteins in *C. difficile*

5.3.1 Determining the minimal inhibition concentration of the antibiotics

The effect of vancomycin, ampicillin and clindamycin on the surface proteome of *C. difficile* 630 during early exponential phase was investigated. Ampicillin was chosen as a typical β-lactam that can pre-dispose to *C. difficile* infection. The effect of clindamycin was evaluated as it is the antibiotic that most commonly predisposes patients to *C. difficile* infection. The effect of vancomycin on the surface proteome of *C. difficile* was studied as it is commonly used to treat *C. difficile* infections. The minimal inhibition concentration (MIC) was determined for *C. difficile* 630 for each antibiotic in duplicate, by broth macrodilution using the method of Drummond (2003) (Section 2.6.7). The MICs determined here were in close agreement with those reported by other workers (Drummond *et al.*). The MIC of ampicillin was determined as 4 μg/ml, clindamycin as 256 μg/ml and vancomycin was inhibitory at 1 μg/ml.

5.3.2 Sample preparation and SDS-PAGE

A large vessel containing 1.2 l of BHI broth was inoculated with a 1:20 dilution of overnight *C. difficile* culture and incubated under anaerobic conditions. After 3 h growth and when the A_{600} was 0.4, indicating the bacteria were in early exponential phase, the culture was divided into four 300 ml aliquots in sterile vessels. Each culture was treated with an antibiotic, either vancomycin, ampicillin or clindamycin at concentrations of 50% MIC. One culture received no antibiotic, as a control. The cultures were left to grow under anaerobic conditions for a further two hours. The bacteria were harvested by centrifugation, and glycine and lysozyme extracts were prepared on each sample in triplicate (a total of six extracts per treatment). Extracts were first analysed by 1D SDS-PAGE to ensure the extraction procedure was successful (Figure 5.9). This confirmed that the extraction procedure was reproducible, and 50 μg of each protein sample was subjected to IEF and 2-D PAGE. The glycine

and lysozyme extracts were separated by 2-D PAGE, stained with RuBPS and visualised by fluorescent scanning using a FUJI FLA-5000 laser scanner.

5.3.3 Analysis of 2D gels

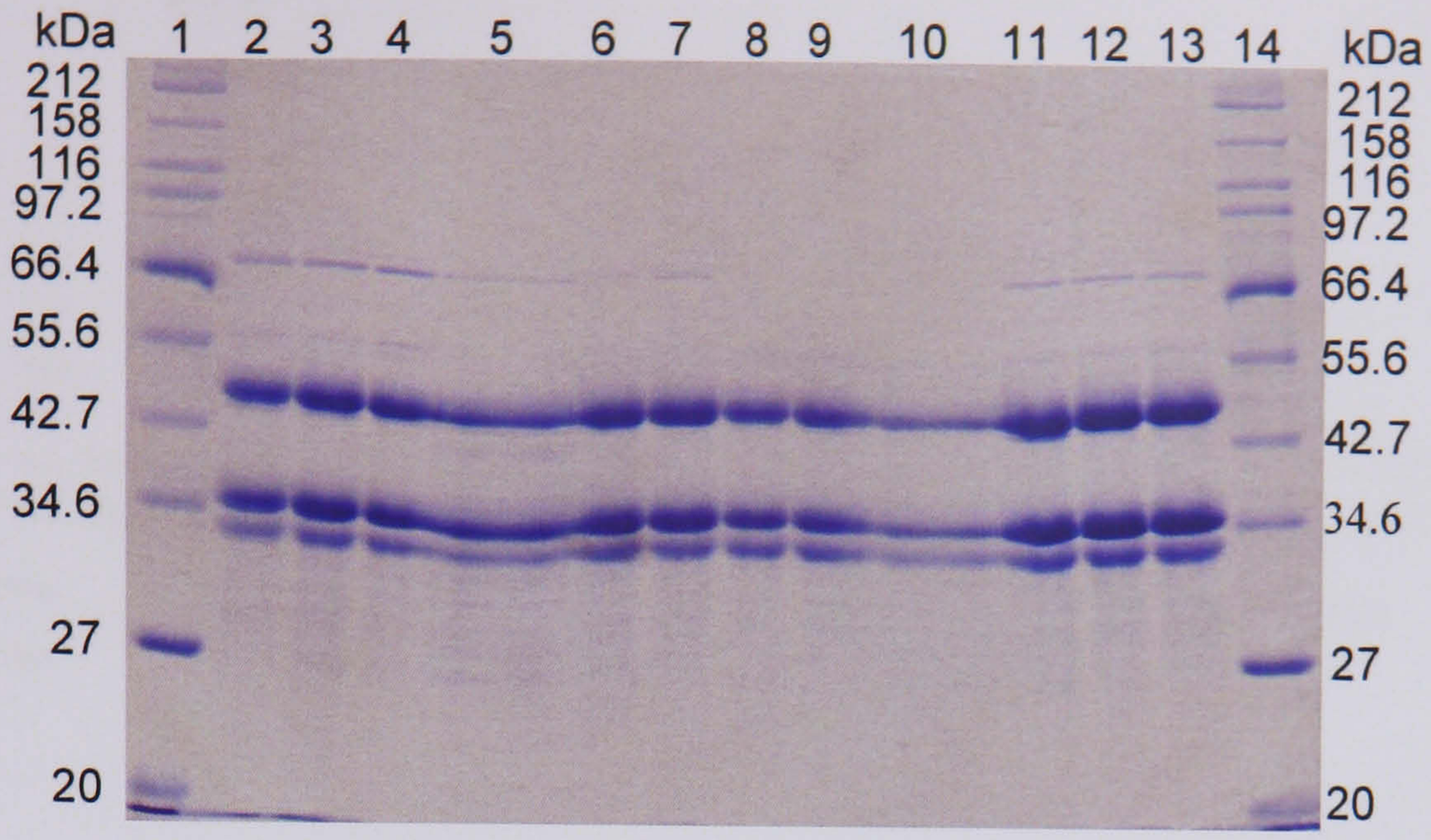
Protein gels were analysed in the Non Linear Dynamics Phoretix 2D Evolution program, in a similar way to that described in section 5.2.2. Briefly, the glycine and lysozyme extracts were analysed in separate experiments, with manual processing at all stages. The spots identified and matched in the triplicate gels of each treatment were compiled into an average gel. The average gels were used to create a reference gel. The ANOVA test was applied to the spot list and those with a $p>0.05$ were discarded.

5.3.4 Differential expression seen within glycine extracts

Glycine extracts (50 μg) from cultures of *C. difficile* 630 treated with antibiotics were analysed by 2-DE. Representative gels are shown in Figure 5.10. Overall, the spot pattern on all twelve gels was broadly similar, with approximately ten additional spots in the extract from the culture without any antibiotic treatment. Spots identified by the Phoretix programme as significantly changing are circled in red. Twelve spots were found to be differentially regulated ($p<0.05$) and are indicated in red on Figure 5.10. The expression of at least ten of these spots was highest in the extracts without antibiotic treatment, suggesting that the effect of all three antibiotics studied here was to reduce the diversity of proteins displayed on the surface of the bacterium.

Fourteen differentially expressed protein spots were analysed by MS/MS (Robin Wait, Imperial College). Five of these spots were identified (Table 5.4). Spot AG6 was identified from three gels (one from a culture without antibiotic treatment, two from vancomycin treated cultures) as SlpA, specifically the HMW SLP. This spot is situated just below a larger spot (Figure 5.10A), identified in Chapter 4 as the HMW SLP. Spot AG6 has the greatest volume on the gels of the vancomycin treated culture extracts, but is present on gels of all the extracts. The other two identified spots were flagellin (spot AG7, no antibiotics) and a putative exported protein (spot AG9, vancomycin treatment), however these spots were not identified in duplicate and so cannot be confirmed as differentially expressed.

A



B

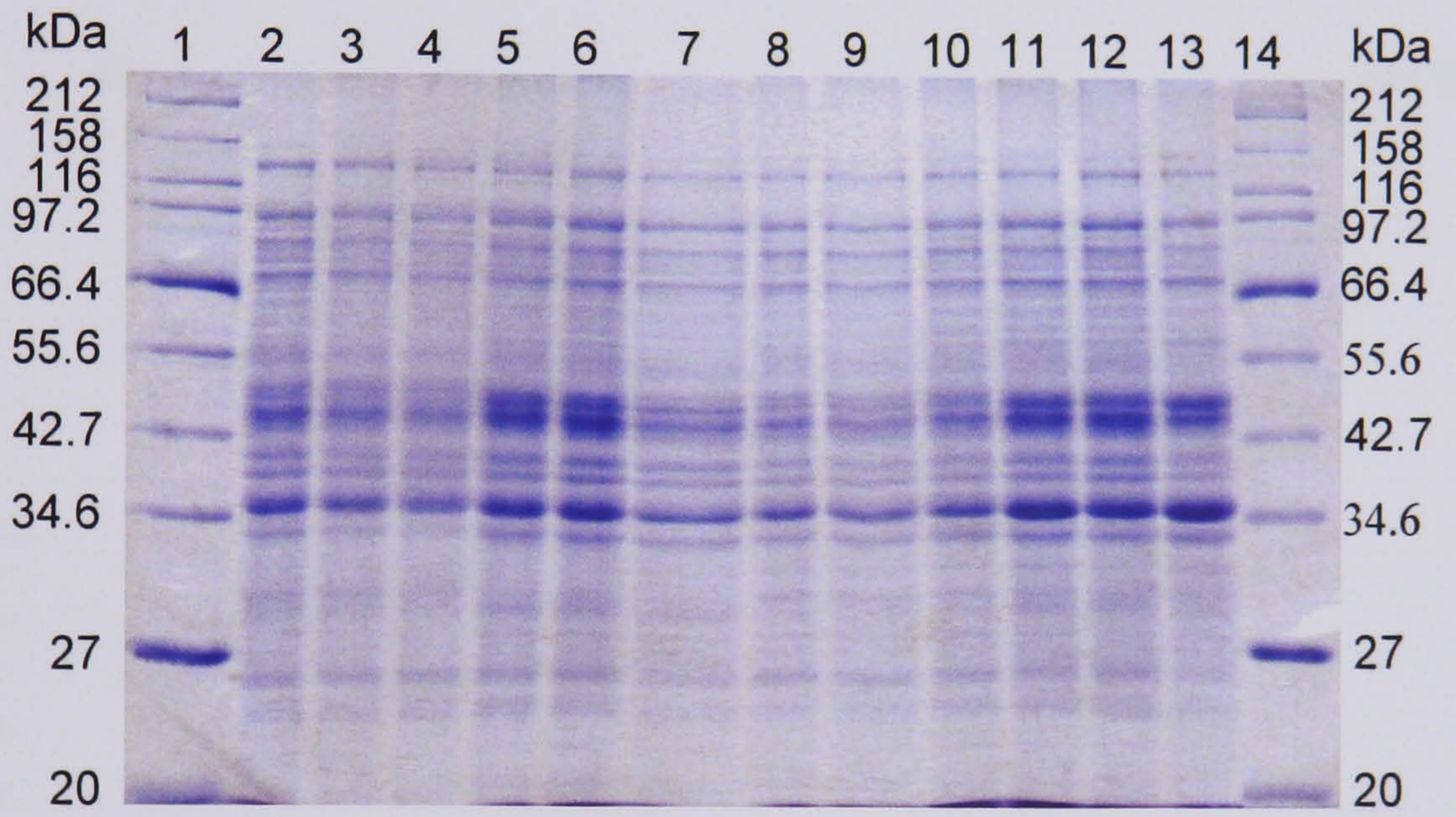


Figure 5.9 1D SDS-PAGE analysis of surface proteins extracted by glycine and lysozyme treatment from cultures of *C. difficile* 630 treated with antibiotics. Triplicate extractions of each condition were made from cells taken from a single, culture. **A.** Glycine extracts; lane 1, molecular mass markers in kDa; lanes 2-4, vancomycin treated culture; lanes 5-7, ampicillin treated culture; lanes 8-10, clindamycin treated culture, lanes 11-13, untreated culture and lane 14, molecular mass markers. **B.** Lysozyme extracts; lane 1, molecular mass markers in kDa; lanes 2-4, vancomycin treated culture; lanes 5-7, ampicillin treated culture; lanes 8-10, clindamycin treated culture, lanes 11-13, untreated culture and lane 14, molecular mass markers.

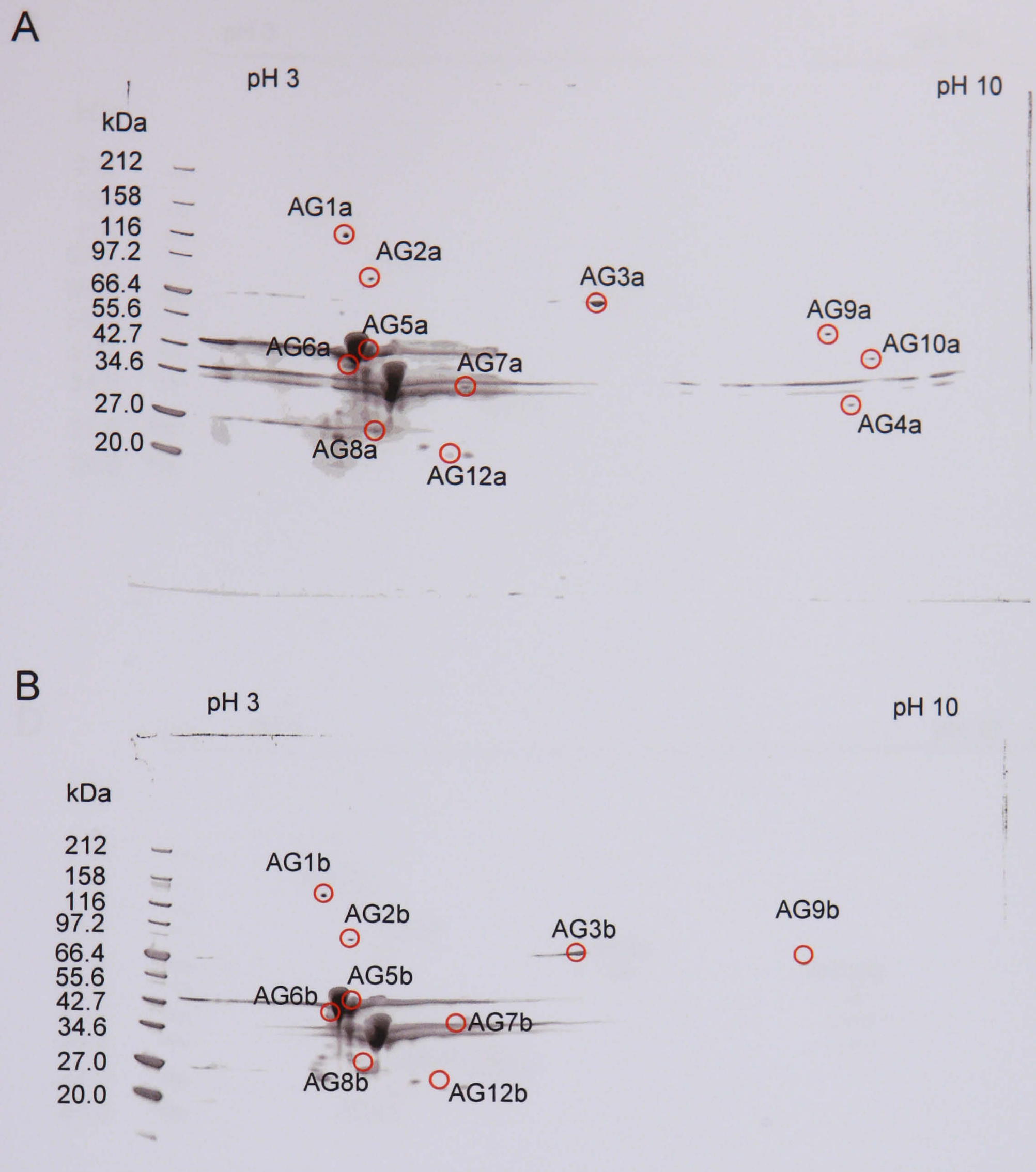


Figure 5.10 2D PAGE analyses of proteins extracted by glycine treatment from the surface of a culture *C. difficile* 630 treated with antibiotics. **A.** Culture grown without antibiotics. **B.** Culture treated with ampicillin. **C.** Culture treated with clindamycin. **D.** Culture treated with vancomycin. Molecular masses (kDa) are indicated on the left and pH above. Spots judged to be differentially expressed are circled in red. The spots are numbered and have the prefix AG to show they are from the antibiotic experiment and extracted by glycine treatment. The suffix 'a' is to show the spot is from the culture without antibiotics. The suffix 'b' is to show the spot is from the culture treated with ampicillin. The suffix 'c' is to show the spot is from the culture treated with clindamycin. The suffix 'd' is to show the spot is from the culture treated with vancomycin. Continued overleaf.

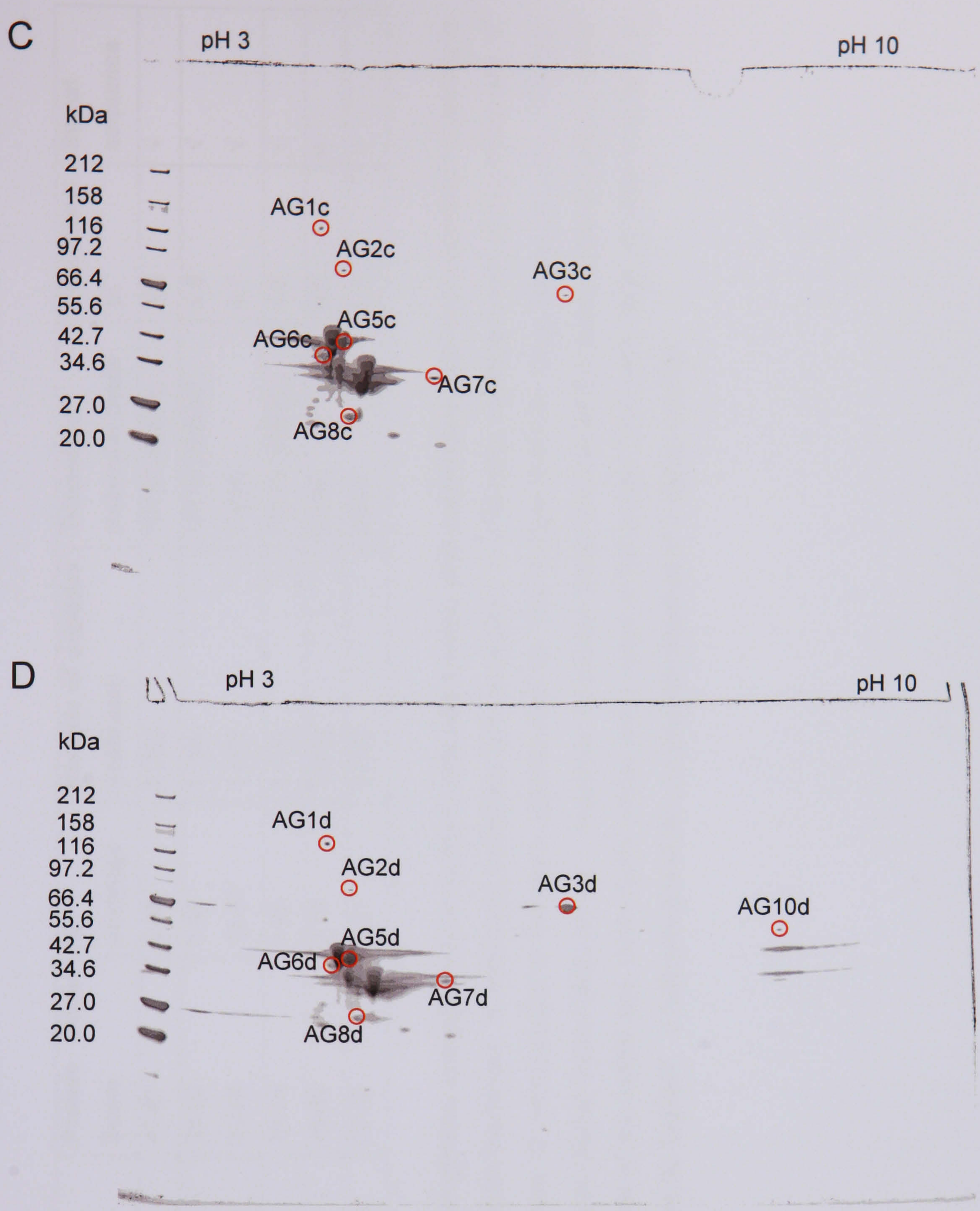


Figure 5.10 continued. 2D PAGE analyses of proteins extracted by glycine treatment from the surface of a culture *C. difficile* 630 treated with antibiotics. C. Culture treated with clindamycin. **D.** Culture treated with vancomycin. Molecular masses (kDa) are indicated on the left and pH above.

Spot	Genome ID	Description	Protein Probe Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
AG6a	CD2793	SlpA, HMW SLP	47.67	5.42	2 (39)	76115 (39504)	4.6	√
AG6d	CD2793	SlpA, HMW SLP	29.08	1.95	1 (14)	76115 (39504)	4.6	√
	CD2251	Putative exported protein	23.08	10.17	1 (12)	13286	4.7	√
AG6d	CD2793	SlpA, HMW SLP	34.14	1.95	1 (14)	76115 (39504)	4.6	√
AG7a	CD0239	FliC, flagellin	48.86	3.79	1 (11)	30755	8.6	x
AG9d	CD1575	Putative exported protein	13.32	4.91	1 (13)	29607	9.8	√

Table 5.4 Proteins identified from 2D gels of glycine extracted proteins from cultures treated with antibiotics. Spots are labelled according to Figure 5.10, with suffix 'a' indicating no antibiotic treatment, 'b' indicating treatment with ampicillin, 'c' indicating treatment with clindamycin and 'd' indicating treatment with vancomycin. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Protein probe score determined by Protein Lynx Global Server (Micromass). Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

PAGE

MISSING

IN

ORIGINAL

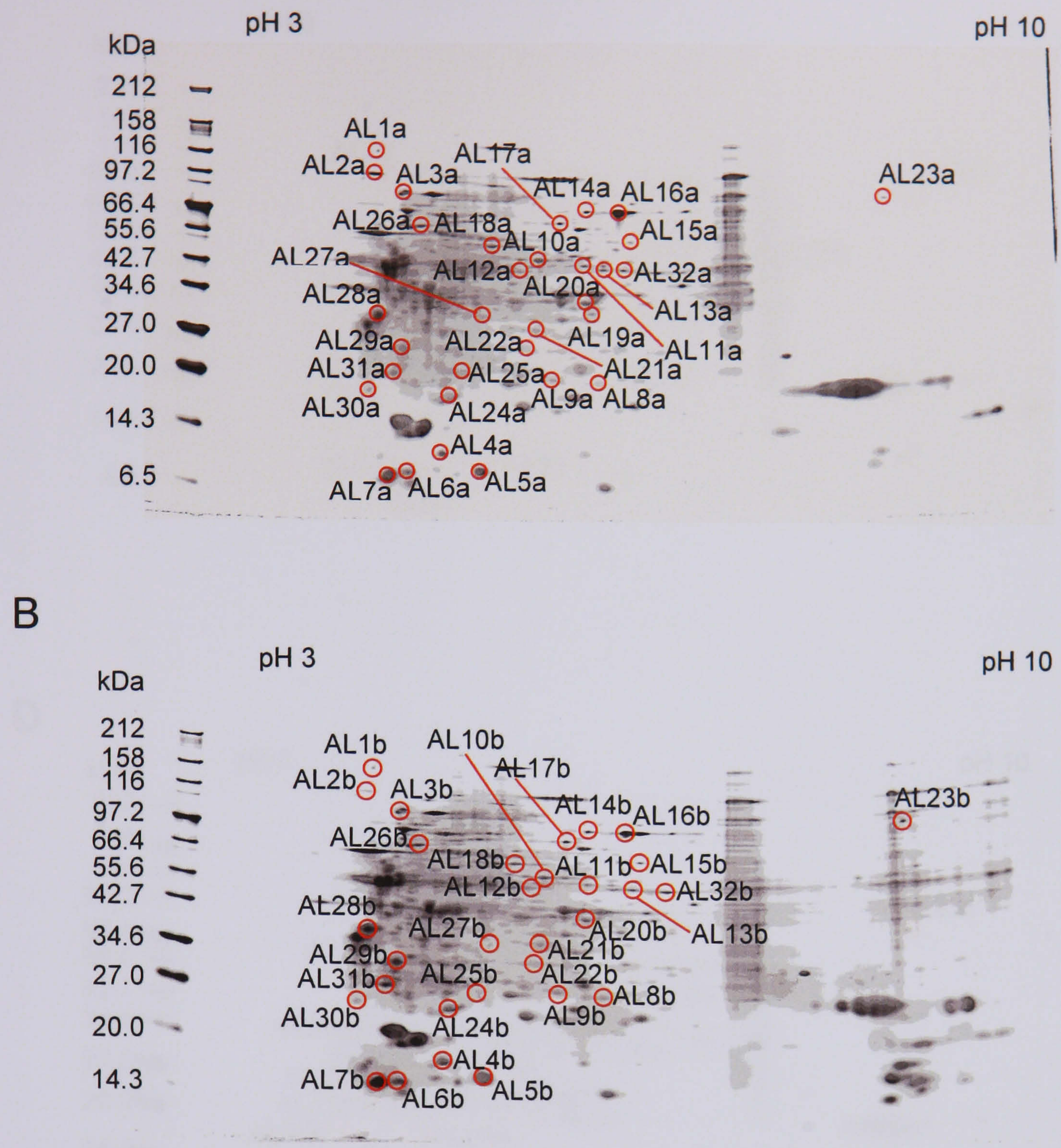
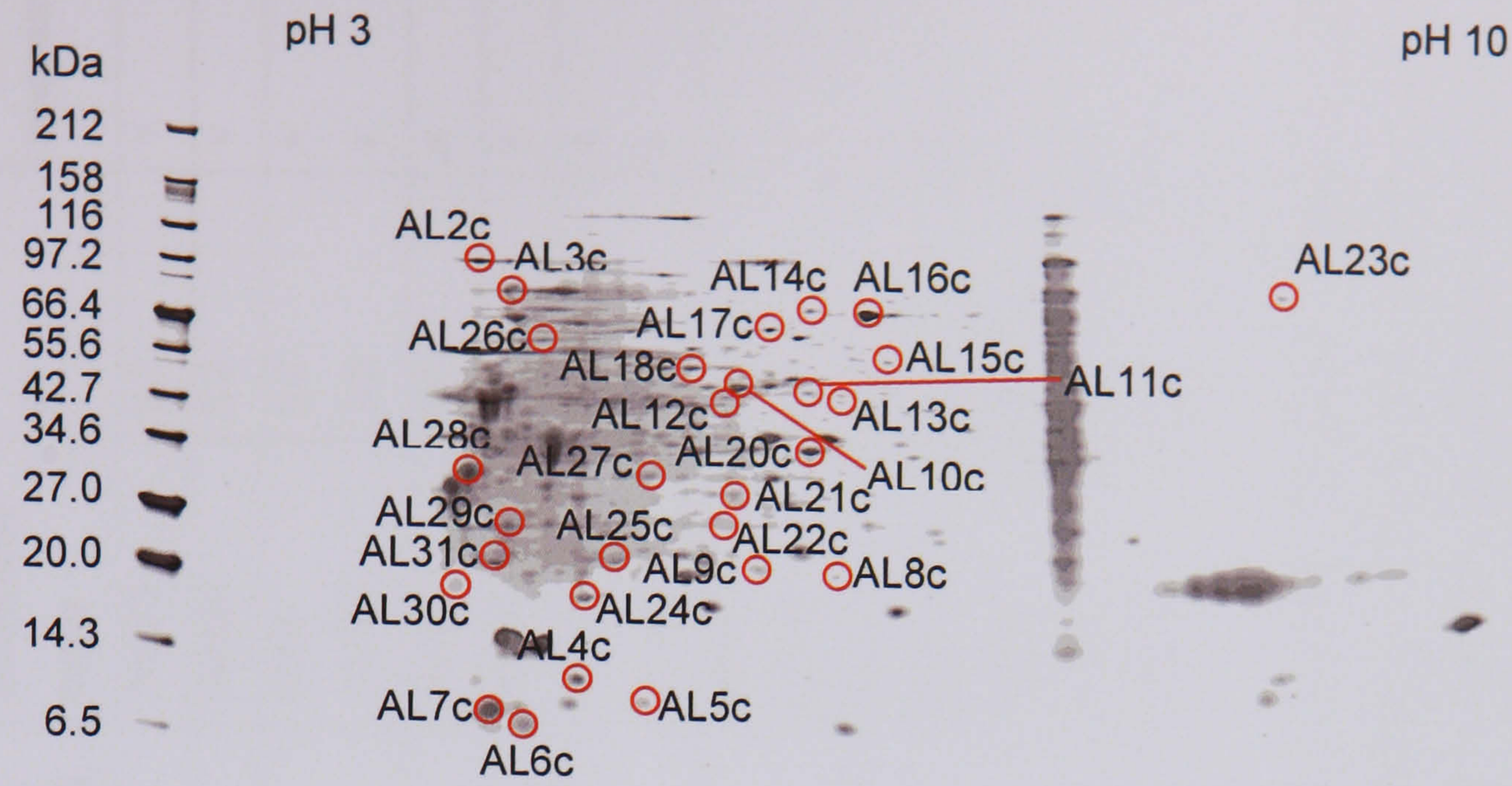


Figure 5.11 2D PAGE analyses of proteins extracted by lysozyme treatment from the surface of a culture *C. difficile* 630 treated with antibiotics. **A.** Culture untreated with antibiotics. **B.** Culture treated with 2 μg/ml ampicillin. **C.** Culture treated with 128 μg/ml clindamycin. **D.** Culture treated with 0.5 μg/ml vancomycin. Molecular masses (kDa) are indicated on the left and pH above. Spots judged to be differentially expressed are circled in red. The spots are numbered and have the prefix AL to show they are from the antibiotic experiment and extracted by lysozyme treatment. The suffix 'a' is to show the spot is from the culture not treated with antibiotics. The suffix 'b' is to show the spot is from the culture treated with ampicillin. The suffix 'c' is to show the spot is from the culture treated with clindamycin. The suffix 'd' is to show the spot is from the culture treated with vancomycin.

C



D

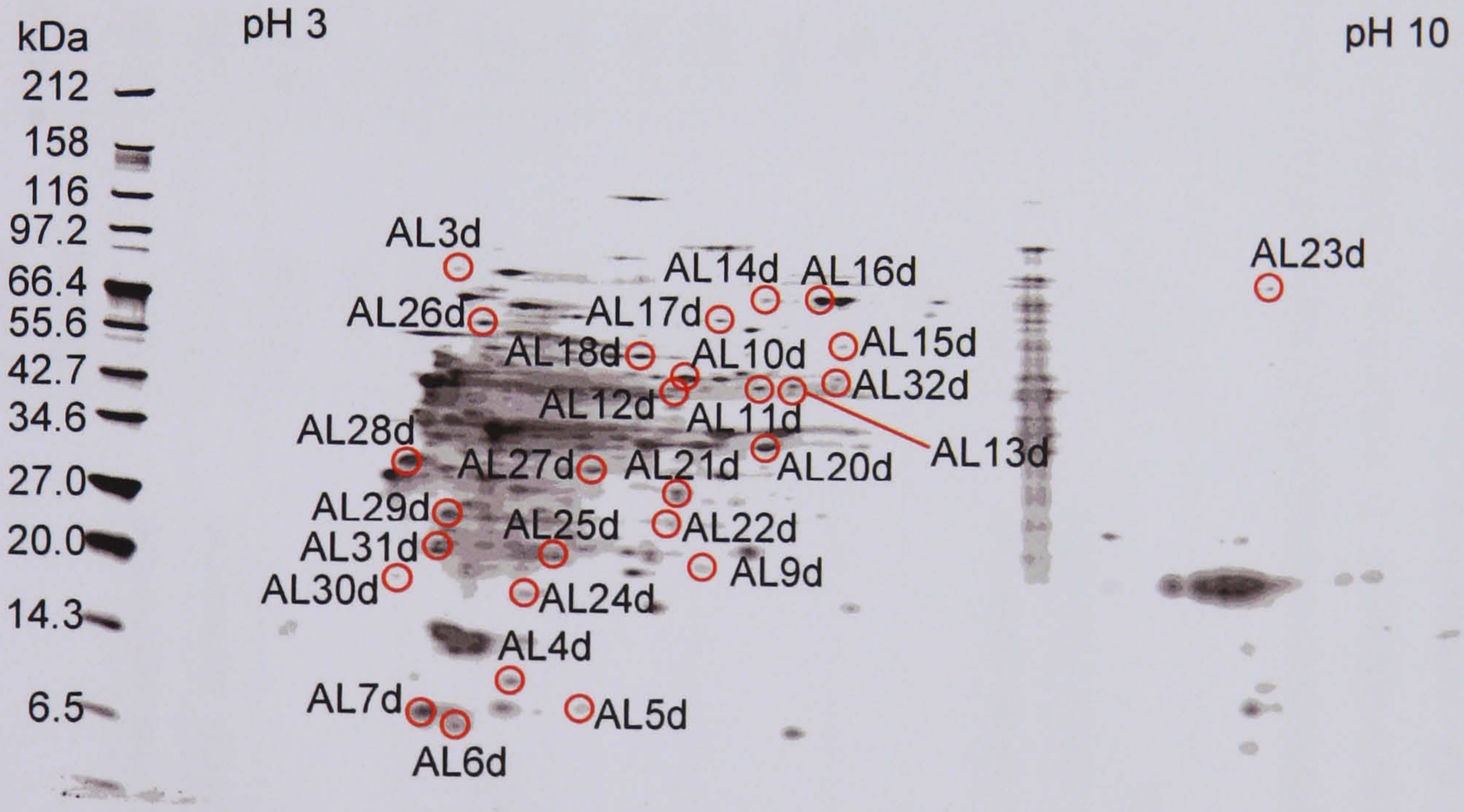


Figure 5.11 continued. 2D PAGE analyses of proteins extracted by lysozyme treatment from the surface of a culture *C. difficile* 630 treated with antibiotics. C. Culture treated with 128 μg/ml clindamycin. **D.** Culture treated with 0.5 μg/ml vancomycin. Molecular masses (kDa) are indicated on the left and pH above.

Spot	Genome ID	Description	Protein Probe Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical p/	Signal sequence
AL2a	CD1350	Putative lantibiotic ABC transporter permease	8.54	3.36	1	26823	9.2	√
AL2a	CD1350	Putative lantibiotic ABC transporter permease	11.10	3.36	1	26823	9.2	√
AL4a	CD2756	PtsH, phosphocarrier protein	20.00	9.30	1 (8)	9046	5.2	x
	CD0236	FliS2, flagellar protein	14.48	11.11	1 (13)	13733	5.0	x
AL5a	CD2756	PtsH, phosphocarrier protein	25.30	11.63	1 (10)	9046	5.2	x
AL8a	CD3317	FdhF, formate dehydrogenase H	10.69	1.97	1 (11)	62653	5.2	x
AL15c	CD3181	Putative aminohydrolase	12.28	1.58	1	49277	5.9	x
	CD3641	Conserved hypothetical	5.91	3.98	1	27164	5.9	x
AL15d	CD3181	Putative aminohydrolase	8.39	1.58	1	49277	5.9	x
	CD3641	Conserved hypothetical	8.18	3.98	1	27164	5.9	x
AL16a	CD2791	Slp paralogue 2	61.72	3.69	2 (23)	66427	6.4	√
AL16c	CD2791	Slp paralogue 2	38.90	4.81	2 (30)	66427	6.4	√
AL22b	CD1157	NorV, putative anaerobic nitric oxide reductase	8.92	3.02	1	44818	5.4	x
AL22c	CD1157	NorV, putative anaerobic nitric oxide reductase	11.38	3.02	1	44818	5.4	x
AL28d	CD0795	Conserved hypothetical protein	10.51	2.83	1 (7)	26948	4.6	x

Table 5.5 Proteins identified from 2D gels of lysozyme extracted proteins from cultures treated with antibiotics. Spots are labelled according to Figure 5.11, with suffix 'a' indicating no antibiotic treatment, 'b' indicating treatment with ampicillin, 'c' indicating treatment with clindamycin and 'd' indicating treatment with vancomycin. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Protein probe score determined by Protein Lynx Global Server (Micromass). Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable). Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

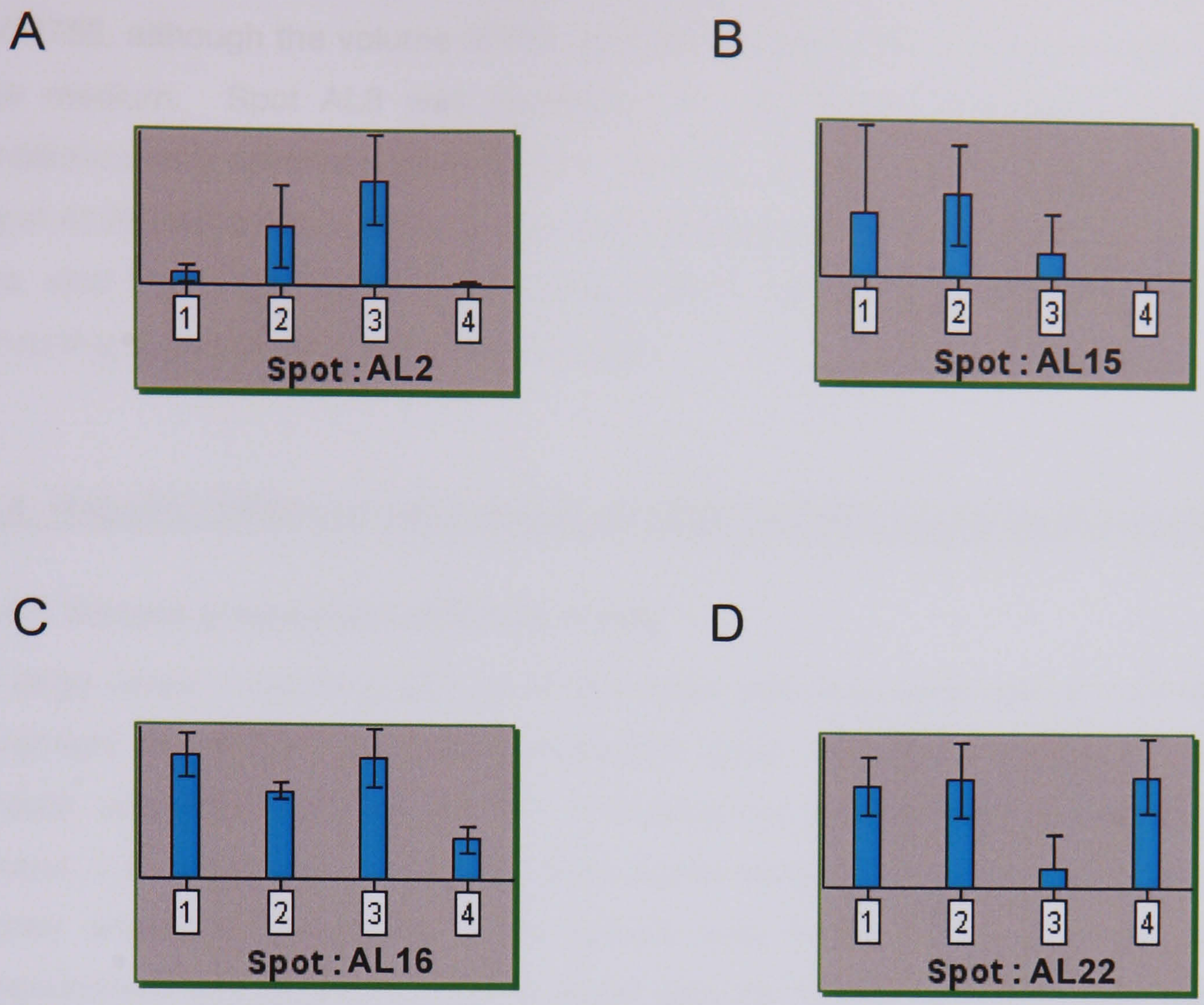


Figure 5.12 Histograms showing the change in spot volumes of identified proteins from cultures treated with antibiotics. The histograms show the normalised spot volume for the extract from the culture treated with: 1, ampicillin; 2, clindamycin; 3, no antibiotic and 4, vancomycin. **A** Spot AL2 identified as CD1350, encoding a component of a lantibiotic ABC transporter complex. **B** Spot AL15, identified as two proteins, CD3181 encoding a putative aminohydrolase, and CD3641 encoding a conserved hypothetical protein. **C** Spot AL16, identified as CD2791, encoding Slp paralogue 2. **D** Spot AL22 identified as CD1157, encoding NorV, a putative anaerobic nitric oxide reductase.

had the greatest volume when antibiotics were present in the medium, particularly vancomycin and clindamycin. Spot AL5 was also found to contain peptides from CD2756, although the volume of this spot was greatest when ampicillin was present in the medium. Spot AL8 was identified from an extract from cells grown without antibiotics, and contained peptides from CD3317, encoding formate dehydrogenase H. Spot AL28 had an increased volume when antibiotics were present in the medium, and the spot from the vancomycin treated culture extract was identified as CD0795, encoding a conserved hypothetical protein.

5.4. Results: Effect of heat-shock on differentially expressed proteins

5.4.1 Sample preparation and SDS-PAGE

A large vessel containing 500 ml of BHI broth was inoculated with a 1:20 dilution of overnight *C. difficile* culture and incubated under anaerobic conditions. After 6 h growth and when the A_{600} was 0.8, indicating the bacteria were in late exponential phase, the culture was divided into three sterile vessels. One culture was left at 37 °C under anaerobic conditions. One culture was sealed to maintain an anaerobic environment and placed in a 60 °C water bath for 10 min, and the other culture was placed in the 60 °C water bath for 30 min. Immediately after the heat treatment, the bacterial cells were harvested by centrifugation. Proteins were extracted by glycine and lysozyme treatment in triplicate for each sample (a total of six extracts per treatment). Extracts were first analysed by 1D SDS-PAGE to ensure the extraction procedure was successful (Figure 5.13). Volumes corresponding to 0.5 A_{600} units of bacteria were loaded onto each lane. This figure shows that increasing durations of heat shock cause a reduction in the amount and diversity of proteins that can be removed from the surface of the bacteria. This is particularly clear on the glycine extracts (Figure 5.13A), which show remarkable similarity to the glycine extracts from the time course experiment (Figure 6.1). Proteins extracted from the surface of *C difficile* after 24 h culture did not contain the two SLPs that were present during exponential growth. Similarly, proteins extracted from a culture that had been subjected to heat shock at 60 °C for 30 min also lacked the two SLPs, whereas the SLPs were the dominant proteins in the extracts from cultures without, and with a brief heat-shock. As with the time course experiment, heat shock did not appear to have any effect on the presence of flagellin in the surface extracts. The heat-shock surface protein extracts were subjected to IEF and 2-D PAGE, stained with RuBPS and visualised by fluorescent scanning.

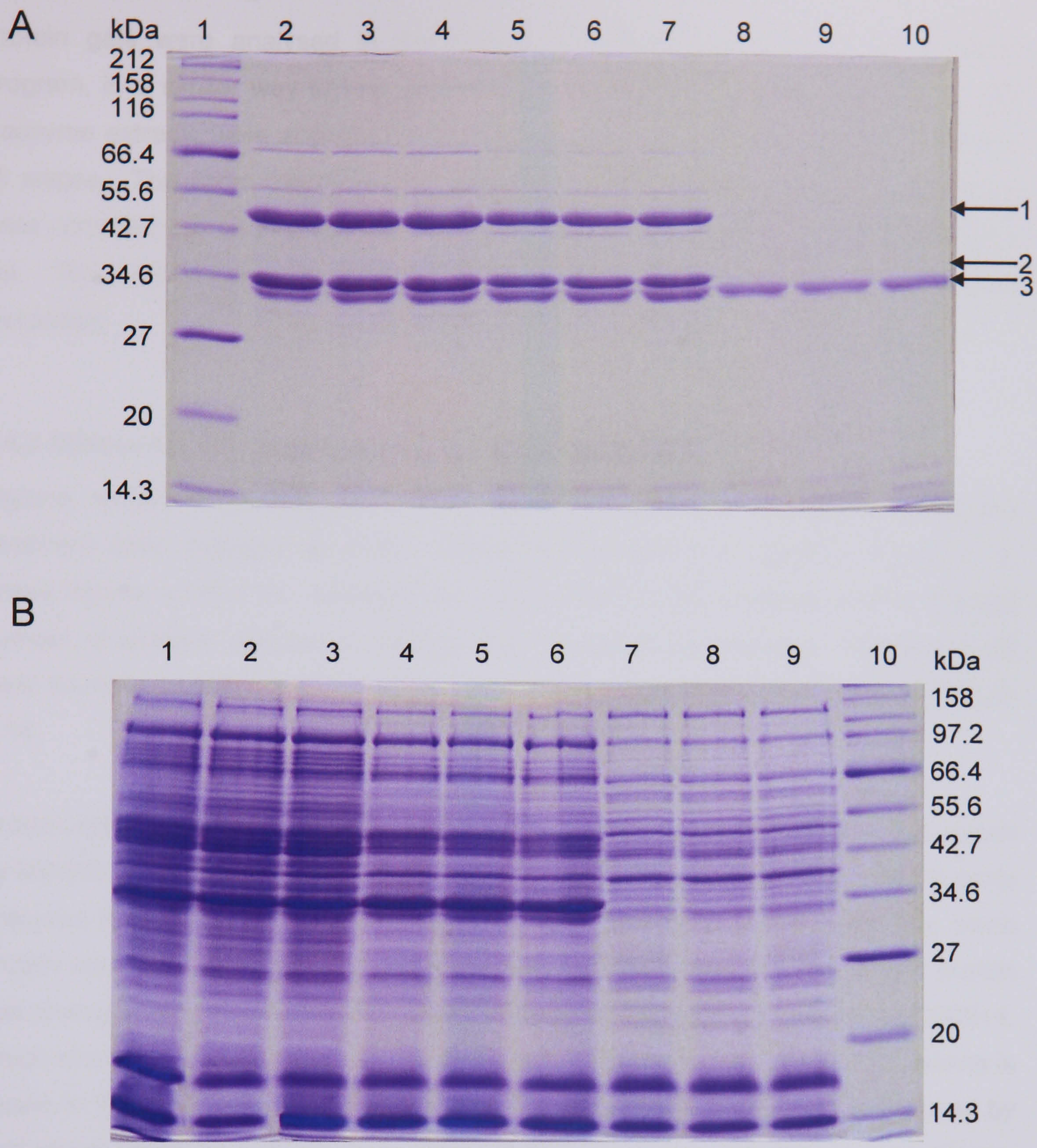


Figure 5.13. 1D SDS-PAGE analysis of surface proteins extracted by glycine and lysozyme treatment from cultures of *C. difficile* 630 that have undergone heat-shock treatment. Triplicate extractions of each condition were made from cells taken from a large, single culture. **A.** Glycine extracts; lane 1, molecular mass markers, in kDa; lanes 2-4, culture remaining at 37 °C; lanes 5-7, culture heat-shocked for 10 min at 60 °C; lanes 8-10, culture heat-shocked for 30 min at 60 °C. Arrows show the position of: 1, HMW SLP; 2, LMW SLP; and 3, FliC. **B.** Lyzozyme extracts; lanes 1-3, culture remaining at 37 °C; lanes 4-6, culture heat-shocked for 10 min at 60 °C; lanes 7-9, culture heat-shocked for 30 min at 60 °C; lane 10, molecular mass markers in kDa.

5.4.2 Analysis of 2D gels

Protein gels were analysed in the Non Linear Dynamics Phoretix 2D Evolution program, in a similar way to that described in section 5.2.2. Briefly, the glycine and lysozyme extracts were analysed in separate experiments, with manual processing at all stages. The spots identified and matched in the triplicate gels of each treatment were compiled into an average gel. The average gels were used to create a reference gel. The ANOVA test was applied to the spot list and those with a $p>0.05$ were discarded.

5.4.3 Differential expression seen within glycine extracts

Glycine extracts from cultures of *C. difficile* 630 that had undergone heat-shock treatment were analysed by 2-DE. Representative gels are shown in Figure 5.14. These figures confirm the findings from the 1D SDS-PAGE analysis, with a reduced number of proteins present in extracts from heat-shocked bacteria. Thirteen spots were found to be differentially expressed and these spots are circled in red on Figure 5.14.

Protein spots identified as differentially expressed were trypsin digested and analysed by MS/MS at least twice for each spot (Jerry Thomas, University of York). All 20 spots analysed by MS/MS were identified (Table 5.6). For many of the spots, the same protein was identified from more than one gel, indicating that the level of the protein was changing between conditions. Spot HG1 was identified from two gels as CD0514, which encodes Slp paralogue 15. The histogram showing the change in spot volume is shown in Figure 5.15A. This graph shows that the volume of this spot is reduced by heat-shock at 60 °C, with a longer shock causing a greater reduction in spot volume. Spot HG2 is composed of Slp paralogue 2, and the histogram of the change in spot volume (Figure 5.15B) shows that this protein is also reduced in volume by the effect of heat-shock. Spots HG3, HG4 and HG7 were all identified as CD0239, encoding flagellin, and the histograms showing the change in spot volume are shown in Figure 5.15C, D and G. Spots HG3 and HG4 are located very close together on the 2D gels, and both of these spots increase in volume after increasing durations of heat-shock at 60 °C. These protein spots are approximately 30 kDa in mass, with a pI of approximately 6. This is quite close to the predicted size of 30.8 kDa and pI of 8.6. Spot HG7 is slightly lower molecular weight, at approximately 27 kDa and has a more acidic pI of approximately 4. Therefore this spot may represent a degraded or modified form of flagellin. Contrary to the trend of increasing spot volume after heat-shock seen with spots HG3 and HG4, spot HG7 decreases in volume after heat-shock. Three

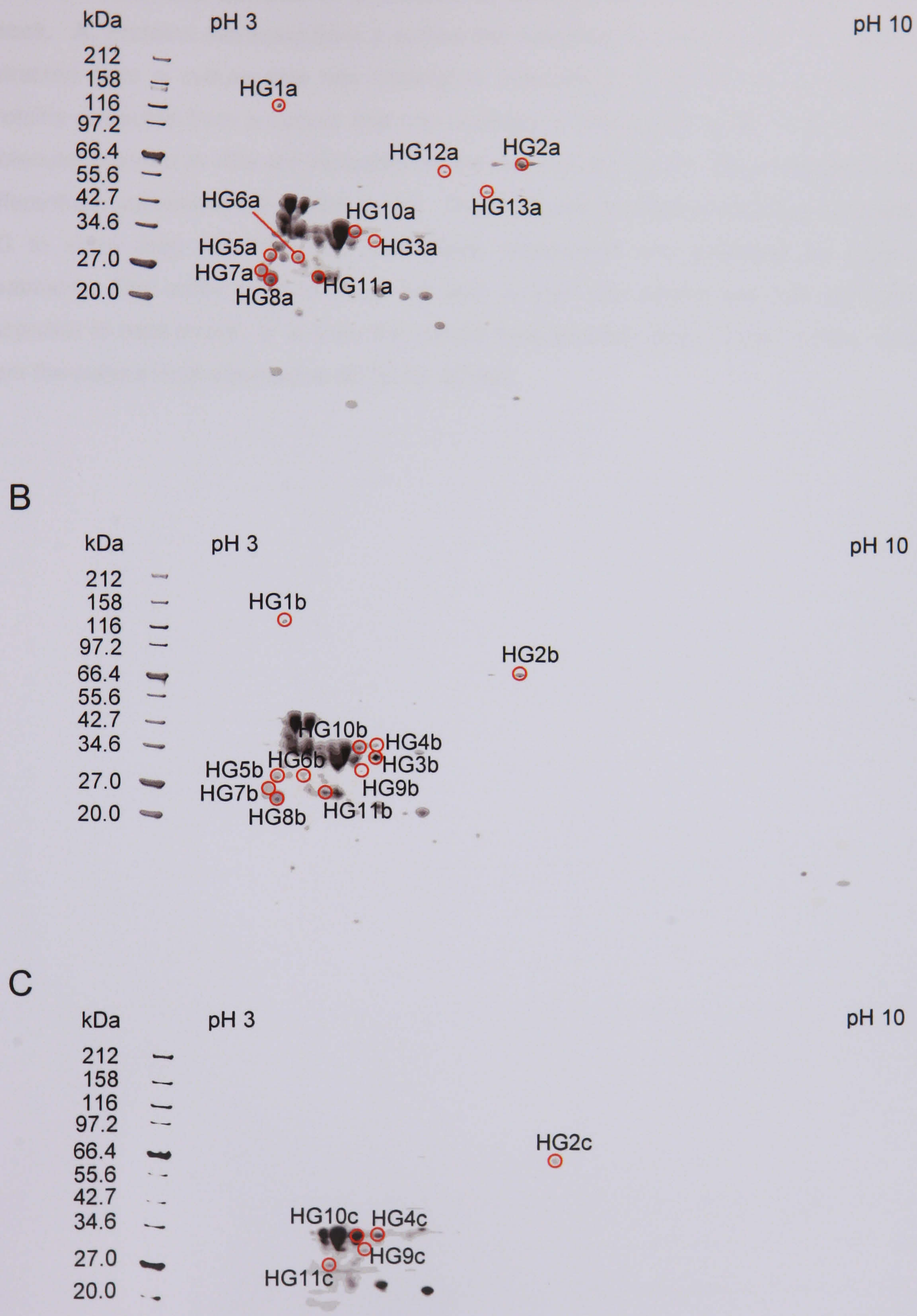


Figure 5.14 2D PAGE analyses of proteins extracted by glycine treatment from the surface of a culture *C. difficile* 630 before and after heat-shock. Legend continued overleaf.

Figure 5.14 continued. 2D PAGE analyses of proteins extracted by glycine treatment from the surface of a culture *C. difficile* 630 before and after heat-shock. A. Proteins extracted from a culture not subjected to heat-shock. **B.** Proteins extracted from a culture that has undergone heat-shock at 60 °C for 10 min. **C.** Proteins extracted from a culture that has undergone heat-shock at 60 °C for 30 min. Molecular masses in kDa are indicated on the left and pH above. Spots judged to be differentially expressed are circled in red. The spots are numbered and have the prefix HG to show they are from the heat-shock experiment and extracted by glycine treatment. The suffix 'a' is to show the spot is from the culture that has not been subjected to heat-shock, 'b' is from the culture heat-shocked at 60 °C for 10 min, 'c' is from the culture heat-shocked at 60 °C for 30 min.

Table 5.6 Proteins identified from 2D gels of glycine extracted proteins from heat-shocked cultures. Spots are labelled according to Figure 5.14, with suffix 'a' indicating no heat-shock, 'b' indicating heat-shock at 60 °C for 10 min and 'c' indicating heat-shock at 60 °C for 30 min. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). The total ion score was determined by Applied Bioscience. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
HG1a	CD0514	Slp paralogue 15	466	1.97	6 (96)	167580	4.7	√
HG1b	CD0514	Slp paralogue 15	389	1.97	6 (96)	167580	4.7	√
HG2a	CD2791	Slp paralogue 2	625	7.22	9 (135)	66427	6.4	√
HG2b	CD2791	Slp paralogue 2	365	5.08	7 (95)	66427	6.4	√
HG3a	CD0239	FliC, flagellin	320	8.26	4 (72)	30755	8.6	x
HG3b	CD0239	FliC, flagellin	440	10.89	6 (95)	30755	8.6	x
HG4c	CD0239	FliC, flagellin	487	11.58	6 (101)	30755	8.6	x
HG4c	CD0239	FliC, flagellin	531	11.93	7 (104)	30755	8.6	x
HG5a	CD2793	SlpA, HMW SLP	509	1.99	6 (97)	76115 (39504)	4.6	√
HG5b	CD2793	SlpA, HMW SLP	594	6.82	6 (123)	76115 (39504)	4.6	√
HG6a	CD2793	SlpA, HMW SLP	508	6.10	7 (110)	76115 (39504)	4.6	√
HG6b	CD2793	SlpA, HMW SLP	114	0.89	1 (16)	76115 (39504)	4.6	√
HG7a	CD0239	FliC, flagellin	326	6.42	4 (56)	30755	8.6	x
HG7a	CD0239	FliC, flagellin	290	6.54	4 (57)	30755	8.6	x

Table 5.6 continued

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
HG8a	CD2793	SlpA, HMW SLP	490	5.04	6 (91)	76115 (39504)	4.6	√
HG8b	CD2793	SlpA, HMW SLP	418	4.27	5 (77)	76115 (39504)	4.6	√
HG10a	CD2793	SlpA, LMW SLP	924	39.44	10 (140)	76115 (34228)	4.9	√
HG10b	CD2793	SlpA, LMW SLP	949	43.38	10 (154)	76115 (34228)	4.9	√
HG11a	CD2793	SlpA, HMW SLP	573	6.49	8 (117)	76115 (39504)	4.6	√
HG11b	CD0514	Slp paralogue 15	283	1.64	5 (80)	167580	4.7	√

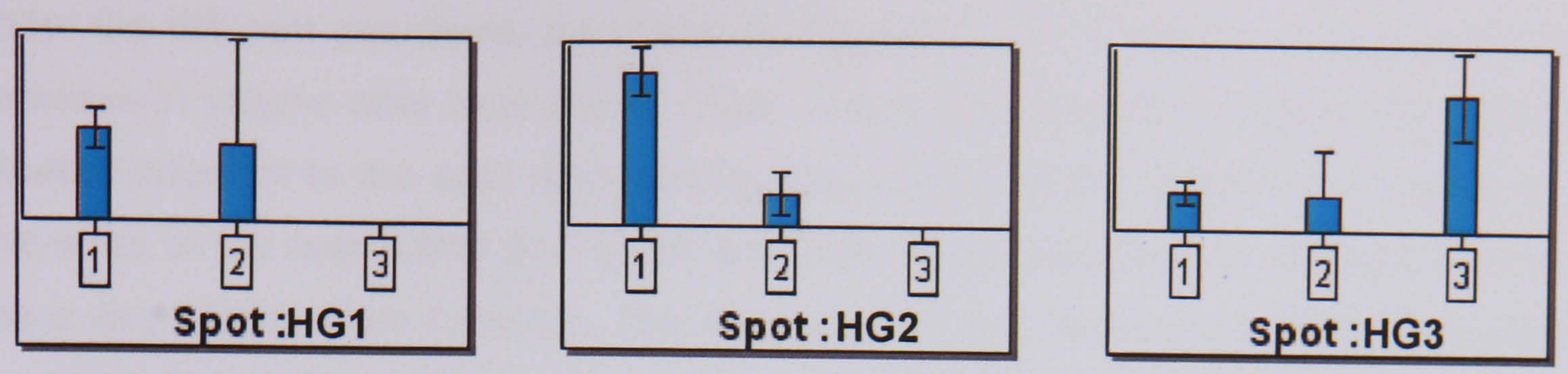


Figure 5.15 Histogram of the change in spot volume for proteins differentially expressed after heat-shock. The histograms show the normalised spot volume for the extract from the culture treated with: 1, no heat-shock; 2, heat-shock at 60 °C for 10 min; 3, heat-shock at 60 °C for 30 min. **A** Spot HG1 identified as CD0514, encoding Slp paralogue 15. **B** Spot HG2, identified as CD2791, encoding Slp paralogue 2. **C** Spot HG3, identified as CD0239, flagellin. **D** Spot HG4 identified as CD0239, flagellin. **E** Spot HG5 identified as CD2793 encoding the HMW SLP. **F** Spot HG6 identified as CD2793 encoding the HMW SLP. **G** Spot HG7 identified as CD0239, flagellin. **H** Spot HG8 identified as CD2793 encoding the HMW SLP. **I** Spot HG10 identified as CD2793 encoding the LMW SLP.

spots, namely HG5, HG6 and HG8, were identified by MS/MS as the HMW SLP. These spots are located close together on the 2D gels, and are much smaller in size than the main HMW SLP spot. The histograms showing the volumes of the spots under the different conditions are shown in Figures 5.15E, F and H. These spots all decrease in volume after heat-shock. Spot 10 was identified as the LMW SLP, and is situated adjacent to the spot representing the majority of the LMW SLP (Chapter 4). The sizes of the main LMW SLP spots and spot 10 are very similar, although spot 10 has a very slightly more basic pI. Therefore spot 10 may represent a modified form of the LMW SLP, rather than a degraded form. After 30 min heat shock at 60 °C, the volume of spot 10 increases markedly (Figure 5.15I), and this also corresponds to a decrease in the volume of the main LMW SLP spot. Therefore it is possible that heat-shock induces a post-translational modification of the LMW SLP, although the mechanism by which this may occur is unknown.

5.4.4 Differential expression seen with in lysozyme extracts

Lysozyme extracts (50 μg) of the surface proteins of *C. difficile* 630 subjected to heat-shock were analysed by 2-DE. Figure 5.16 shows representative 2D gels of extracts from bacteria kept at 37 °C, heat-shocked at 60 °C for 10 min, and heat-shocked at 60 °C for 30 min respectively. The gels were stained with RuBPS and visualised using a FUJI FLA-5000 laser scanner. Analysis of the 2D gels using the Phoretix programme resulted in the identification of 27 differentially expressed spots. These spots are circled in red and labelled on Figure 5.16.

All of the 35 differentially expressed spots analysed by MS/MS (Jerry Thomas, University of York) were identified (Table 5.7). Thirteen spots were identified as the same protein on more than one gel, indicating that these proteins are indeed differentially expressed. Spot HL2 was identified from two extracts from bacteria not subjected to heat-shock as CD2461, encoding DnaK, a chaperone protein. A histogram of the normalised volume of spot HL2 is shown in Figure 5.17A, and shows unexpectedly that this spot is only found in extracts from bacteria that have not been heat-shocked. Although the sequence for DnaK is not predicted to contain a signal sequence, it is commonly found in surface proteomes of other bacteria (Schaumburg *et al.*, 2004). Spot HL4 was identified as enolase. This spot has a large increase in volume after heat-shock at 60 °C for 30 min (Figure 5.17B). Spots HL5 and HL6 were both identified as SlpA, localised to the LMW SLP. The normalised volumes of spots HL5 and HL6 both reach a peak after heat-shock for 10 min, and markedly decrease when the heat-shock is extended to 30 min (Figure 5.17C and D). The

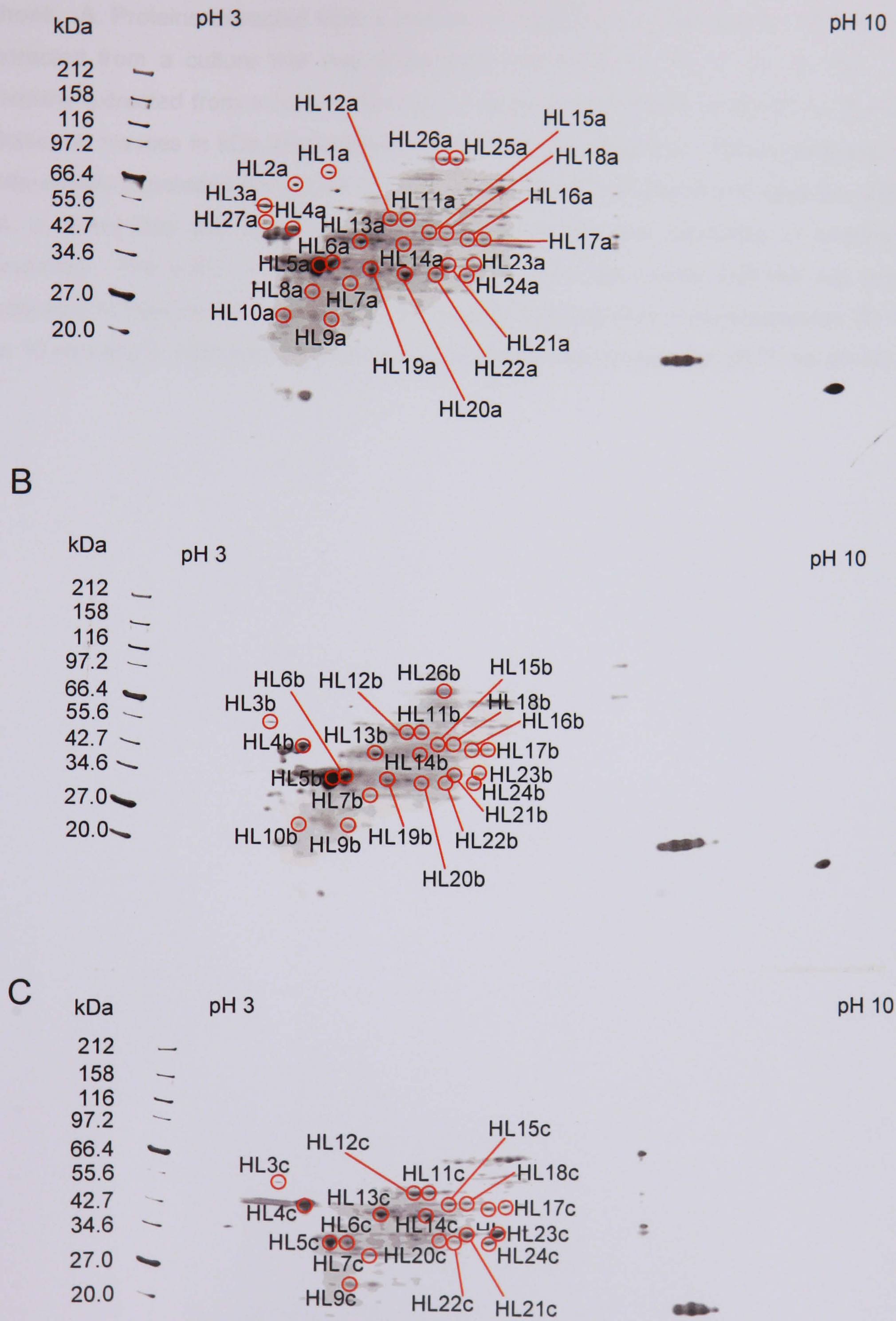


Figure 5.16 2D PAGE analyses of proteins extracted by lysozyme treatment from the surface of a culture *C. difficile* 630 before and after heat-shock. Legend continued overleaf.

Figure 5.16 continued. 2D PAGE analyses of proteins extracted by lysozyme treatment from the surface of a culture *C. difficile* 630 before and after heat-shock. **A**. Proteins extracted from a culture not subjected to heat-shock. **B**. Proteins extracted from a culture that has undergone heat-shock at 60 °C for 10 min. **C**. Proteins extracted from a culture that has undergone heat-shock at 60 °C for 30 min. Molecular masses in kDa are indicated on the left and pH above. Spots judged to be differentially expressed are circled in red. The spots are numbered and have the prefix HL to show they are from the heat-shock experiment and extracted by lysozyme treatment. The suffix 'a' is to show the spot is from the culture that has not been subjected to heat-shock, 'b' indicates the culture that has been heat-shocked at 60 °C for 10 min and 'c' indicates the culture that has been heat-shocked at 60 °C for 30 min.

Table 5.7 Proteins identified from 2D gels of lysozyme extracted proteins from heat-shocked cultures. Spots are labelled according to Figure 5.16, with suffix 'a' indicating no heat-shock, 'b' indicating heat-shock at 60 °C for 10 min and 'c' indicating heat-shock at 60 °C for 30 min. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Total ion score determined by Applied Bioscience. Predicted molecular mass (Da) based on sequences containing signal peptide except for the HMW and LMW SLPs which show mass of full length SlpA and mature protein in parentheses. Signal sequences were predicted by SignalP, √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
HL1a	CD1056	EtfA2, acetyl CoA acetyl transferase	341	5.35	5 (54)	36073	5.1	x
HL2a	CD2461	DnaK, chaperone protein	428	5.96	8 (110)	66449	5.3	x
HL2a	CD2461	DnaK, chaperone protein	107	1.95	4 (36)	66449	5.3	x
HL3a	CD2672	AppA, oligopeptide ABC transporter	640	8.28	10 (132)	59706	5.6	√
HL3c	CD0194	GroEL, 60 kDa chaperonin	168	2.33	4 (38)	57772	4.7	x
HL4a	CD3170	Eno, enolase	499	7.82	7 (101)	48149	4.6	x
HL4b	CD3170	Eno, enolase	526	8.36	7 (108)	48149	4.6	x
HL5a	CD2793	SlpA, LMW SLP	1045	41.41	10 (147)	76115 (34228)	4.9	√
HL5b	CD2793	SlpA, LMW SLP	934	38.03	9 (135)	76115 (34228)	4.9	√
HL6a	CD2793	SlpA, LMW SLP	669	32.39	8 (115)	76115 (34228)	4.9	√
HL6b	CD2793	SlpA, LMW SLP	1022	41.41	10 (147)	76115 (34228)	4.9	√
HL9a	CD0058/ CD0071	TufA/TufB, elongation factor Tu	191	3.94	4 (47)	44008	4.6	x
HL9b	CD0058/ CD0071	TufA/TufB, elongation factor Tu	111	4.86	4 (58)	44008	4.6	x
HL12a	CD1175	AckA, acetate kinase	179	6.27	6 (75)	43321	5.4	x

Table 5.7 continued

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
HL12b	CD2335	GuaB, inosine monophosphate dehydrogenase	120	1.87	3 (28)	54778	5.3	x
HL13a	CD3664	Putative amino acid aminotransferase	270	6.27	6 (75)	44798	5.1	x
HL13b	CD3664	Putative amino acid aminotransferase	278	7.02	7 (84)	44798	5.1	x
HL14a	CD1175	AckA, acetate kinase	215	7.78	7 (93)	43321	5.4	x
HL14c	CD1175	AckA, acetate kinase	174	5.18	5 (62)	43321	5.4	x
HL15a	CD3285	Pgi, glucose-6-phosphate isomerase	611	8.75	9 (118)	50538	5.6	x
HL15c	CD3285	Pgi, glucose-6-phosphate isomerase	242	3.48	4 (47)	50538	5.6	x
HL16a	CD2791	Slp paralogue 2	141	2.78	3 (52)	66427	6.4	√
HL16b	CD2726	GlyA, putative serine hydroxymethyltransferase	133	2.09	2 (26)	45796	5.7	x
HL16c	CD0410	Putative DNA replication protein (CTn2)	34	0.12	1 (8)	33273	6.3	x
HL17a	CD2726	GlyA, putative serine hydroxymethyltransferase	108	2.89	3 (36)	45796	5.7	x
HL17b	CD2726	GlyA, putative serine hydroxymethyltransferase	122	2.09	2 (26)	45796	5.7	x
HL18a	CD0084	RplE, 50S ribosomal protein L5	13	1.48	1 (8)	20437	9.5	x
HL18c	CD0084	RplE, 50S ribosomal protein L5	16	1.48	1 (8)	20437	9.5	x
HL21a	CD3174	Glyceraldehyde-3-phosphate dehydrogenase	132	3.08	3 (31)	36015	5.9	x
HL21b	CD3174	Glyceraldehyde-3-phosphate dehydrogenase	334	6.26	6 (63)	36015	5.9	x
HL24a	CD2270	FruK, putative 1-phosphofructokinase	32	1.96	2 (18)	33832	5.6	x
HL24b	CD2270	FruK, putative 1-phosphofructokinase	91	2.84	3 (26)	33832	5.6	x
HL25a	CD2966	AdhE, aldehyde-alcohol dehydrogenase	134	1.32	3 (35)	96674	5.8	x
HL27a	CD3306	Tig, trigger factor	196	3.50	5 (47)	48187	4.6	x
HL27a	CD3306	Tig, trigger factor	13	0.52	1 (7)	48187	4.6	x

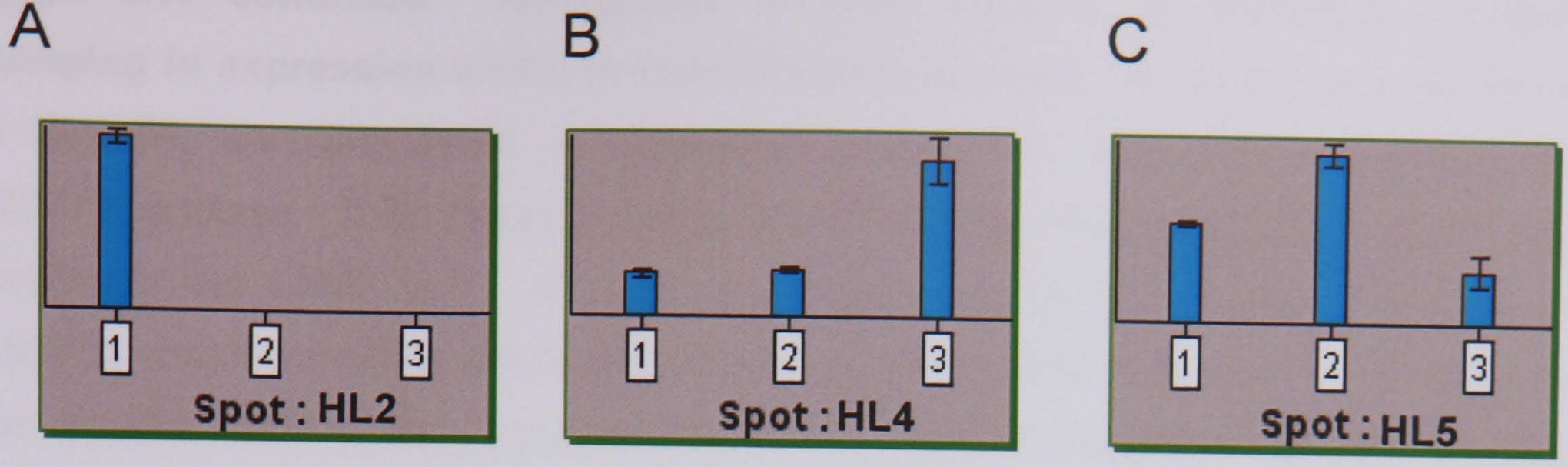


Figure 5.17 Histograms of spot volumes of identified proteins changing in expression levels in heat-shocked bacteria. Legend continued overleaf.

Figure 5.17 continued. Histograms of spot volumes of identified proteins changing in expression levels in heat-shocked bacteria. A. Spot 2 was identified as CD2461, encoding DnaK, a chaperone protein. **B.** Spot 4 was identified as CD3170, enolase. **C-D.** Spots 5 and 6 were both identified as CD2793, SlpA, and specifically the LMW SLP. **E.** Spot 9 contains peptides from either CD0058 or CD0071, which are 100% identical and encode the elongation factor Tu. **F.** Spot 13 was identified as CD3664, a putative amino acid aminotransferase. **G.** Spot 14 was identified as CD1175, which encodes acetate kinase. **H.** Spot 15 was identified as CD3285, glucose 6-phosphate isomerase. **I.** Spot 17 was identified as CD2726, encoding a putative serine hydroxymethyltransferase. **J.** Spot 18 was identified as CD0084, the ribosomal protein RplE. **K.** Spot 21 was identified as CD3174, glyceraldehyde 3-phosphate dehydrogenase. **L.** Spot 24was identified as CD2270, which is thought to encode phosphofructokinase. **M.** Spot 27 was identified as CD3306, which encodes a trigger factor. Histograms show normalised spot volumes. Bar 1 represents the culture not subjected to heat-shock, bar 2 represents the culture heat-shocked at 60 °C for 10 min and bar 3 represents the culture heat-shocked at 60 °C for 30 min. Error bars indicate variation within a condition.

peptides from spot HL9 could correspond to one of two identical CD numbers, CD0058 or CD0071, which are 100% identical and encode the elongation factor Tu. The volume of this spot appears to increase after increasing durations of temperature stress (Figure 5.17E). Spot HL13 is CD3664, a putative amino acid aminotransferase. This spot increases in volume after increasing durations of heat-shock (Figure 5.17F). Spot HL14 contains peptides from CD1175, which encodes acetate kinase. This spot increases in volume after heat-shock for 30 min at 60 °C (Figure 5.17G). Spot HL15 is CD3285, glucose 6-phosphate isomerase. This spot increases in volume after heat-shock for 30 min (Figure 5.17H). Spot HL17 was identified from two extracts as CD2726, encoding a putative serine hydroxymethyltransferase. This spot is upregulated with increasing durations of heat shock (Figure 5.17I). Spot HL18 contains the ribosomal protein RplE. Normally this protein is present at very low levels in the surface extracts, however after a 10 min heat-shock at 60 °C its levels are elevated, and this effect is even more pronounced after 30 min heat-shock (Figure 5.17J). Spot HL21 was identified from two separate extracts as glyceraldehyde 3-phosphate dehydrogenase, and this spot increases in volume after heat-shock (Figure 5.17K). Spot HL24 was identified as CD2270, which is thought to encode phosphofructokinase. This spot increases in volume after the bacteria are exposed to heat-shock (Figure 5.17L). Spot HL27 has been identified from replicate gels of the surface extract from bacteria that had not been heat-shocked as CD3306, which encodes a trigger factor. This spot is not present in extracts from heat-shocked bacteria. This gene is located immediately upstream of two genes encoding the Clp protease.

5.5 Discussion

The differential expression of *C. difficile* surface proteins has been studied under a variety of conditions. A total of 202 protein spots from surface extracts were analysed by MS/MS after separation by 2-DE, resulting in identifications for 125 protein spots. Of these proteins, 45 protein spots were identified in duplicate and confirmed as differentially expressed.

The transition from exponential to stationary phase is generally characterised by a wide range of stress conditions such as nutrient depletion (Siegele & Kolter, 1992; Ishihama, 1997). In *C. difficile*, this transition is particularly relevant, as the two key virulence factors (the toxins) are differentially regulated according to the growth phase (Dupuy & Sonenshein, 1998). Transcription of the toxin genes is repressed during growth but substantially increases as the cells enter stationary phase. Furthermore, the toxin genes are subject to catabolite repression, as the stationary phase induction can be

alleviated by the addition of glucose to the medium. Because of the importance of growth phase in the regulation of the toxins, the differential expression of surface proteins during growth was investigated, to see if other surface associated virulence factors (e.g. adhesins) were regulated in the same way.

Proteins were extracted in triplicate from the surface of *C. difficile* 630 at early-exponential, mid-exponential and stationary phases of growth by glycine and lysozyme treatments in parallel. These duplicate extraction methods have been shown to produce different but complementary views of the cell surface (Chapter 4). A total of sixteen protein spots were identified in duplicate and can be confirmed as differentially expressed during the growth of *C. difficile*. Two proteins increased in volume as the age of the culture increased, namely a putative selenocysteine synthase and an acetyl CoA acetyl transferase. Five different proteins decreased in intensity as the bacteria entered stationary phase. These proteins are the HMW SLP, Slp paralogue 2, a component of a sulphonate ABC transporter, transketolase and ClpP1, a protease. Four different proteins were maximally expressed during late exponential phase. This includes five spots that were all identified as the HMW SLP, and all of these spots had a minimum intensity after 24 h culture. Interestingly, duplicate spots were identified from the 6 h zoom gels displaying the lysozyme extract as containing peptides from the central region of SlpA. These protein spots have a low molecular weight (less than 20 kDa), and so cannot correspond to the full-length protein. Additionally, spots corresponding to Slp paralogue 2 and AppA, a component of an oligopeptide transporter were identified from the zoom gels of the lysozyme extract, again with a maximum intensity at 6 h growth. With so few confirmed changes, it is difficult to see any broad trends for changing classes of proteins. In addition, it may not be possible to detect some proteins that are down regulated if they have a long half-life in the cell, and so this is a conservative representation of the proteins that are differentially expressed during growth. Other recent studies have also detected a wide range of differentially expressed proteins during growth, including proteins involved in cellular metabolism, RNA and protein biosynthesis and stress adaptation (Folio *et al.*, 2004; Gade *et al.*, 2005). Generally, proteins involved in RNA and protein synthesis are down regulated as cells enter stationary phase, whereas proteins involved in cellular metabolism are up regulated. In *Rhodopirellula baltica*, citric acid cycle activity is reduced and the pentose phosphate pathway plays a more important role as cells enter stationary phase (Gade *et al.*, 2005). However, the differentially regulated proteins detected in *C. difficile* do not necessarily conform to these trends, with an increase in a biosynthetic protein (selenocysteine synthase) and a concomitant increase and decrease in two enzyme involved in metabolism (acetyl CoA acetyl transferase

increases, transketolase decreases). The HMW SLP spot decreases significantly in intensity as the age of the culture increases. This may be because less protein is produced, or the protein is shed, or because it is less easily removed from the cell surface. Due to the heavy metabolic demand of producing an S-layer, and the decrease in nutrient availability as the culture enters stationary phase, it is possible that lower quantities of the SLPs are produced. The intensity of the Slp paralogue 2 protein spot also decreases in stationary phase, but as the function of this protein is unknown, the significance of this cannot be determined. It is known that Slp paralogue 2 can be detected in the culture medium, particularly when glucose is present in the medium, maintaining cells in the exponential phase of growth (Mukherjee *et al.*, 2002). Therefore during stationary phase, when the toxins are expressed, Slp paralogue 2 is found at lower levels in both the culture supernatant and at the cell surface.

The action of antibiotics on cultures of *C. difficile* is especially relevant to CDAD infections, as certain antibiotics predispose to the condition. Therefore, the effect of two pre-disposing antibiotics, clindamycin and ampicillin, on the surface proteome was investigated, along with the effect of one of the main treatments for CDAD, vancomycin. From the 79 protein spots identified from these extracts, a total of five differentially expressed protein spots were identified in duplicate from cultures containing sub-inhibitory levels of antibiotics. First, the expression of two proteins was suppressed in the presence of antibiotics, Slp paralogue 2 and CD1350 encoding a permease from a lantibiotic ABC transporter. Three protein spots had higher intensities in the presence of antibiotics: a presumed degradative product of the HMW SLP; NorV anaerobic nitric oxide reductase; and a spot containing peptides from a putative aminohydrolase and a conserved hypothetical protein. Again, the small number of differentially expressed proteins identified and their diversity makes it difficult to see any broad trends. But, in common with the experiment studying the growth phases, the HMW SLP and Slp paralogue 2 are differentially regulated under these conditions of stress. The upregulation of a nitric oxide reductase in the presence of antibiotics may also be part of a general stress response. Nitric oxide reductases are produced by many pathogens, and are believed to protect against nitric oxide attack from host macrophage (Stevanin *et al.*, 2005).

The response to an increase in temperature, or heat-shock, has been widely studied in a number of organisms (Hecker *et al.*, 1996; Bernhardt *et al.*, 1997). Typically, this results in the induction of a number of proteins through a complex response system, which includes a group of heat shock proteins. The study of the response to heat-shock is particularly interesting in *C. difficile* as some adhesins are produced as part of

this response (Karjalainen *et al.*, 1994; Waligora, 2001). The presence of the adhesin Cwp66 and other possible co-regulated virulence factors on the surface of *C. difficile* cells after heat-shock was analysed using proteomics.

Despite MS/MS analysis of 55 differentially regulated protein spots, Cwp66 was not identified in the surface extracts. However, fifteen different proteins were found to be differentially expressed following increasing durations of heat-shock at 60 °C. Nine proteins were up regulated by heat-shock including proteins involved in metabolism, (enolase, glucose 6-phosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, phosphofructokinase and acetate kinase), biosynthesis (elongation factor Tu, amino acid transferase and serine hydroxymethyltransferase) and the ribosomal protein RplE. All of these proteins would be expected to be located in the cytoplasm. These results appears to indicate that heat-shock induces an increase in glycolysis, with four enzymes from this pathway increasing in expression. This may be necessary to fuel an increase in biosynthetic activity, perhaps preparing the bacteria to undertake sporulation. RplE is differentially regulated in *B. subtilis* exposed to cold shock (Beckering *et al.*, 2002), and so could be part of a general stress response. Six proteins were down regulated by heat-shock, including DnaK, a trigger factor, Slp paralogues 2 and 15 and the LMW- and HMW SLPs. Finding DnaK is reduced in expression after heat-shock is unexpected, as it is known to be a heat-shock protein in other organisms (Volker *et al.*, 1994; Hecker *et al.*, 1996). Another common heat-shock protein of the same class as DnaK, GroEL, has been proven to be upregulated in *C. difficile* in response to heat, and is exported to the cell surface and culture medium (Hennequin *et al.*, 2001a). Furthermore, GroEL can also aid in the adhesion of *C. difficile* to the host tissue (Hennequin *et al.*, 2001b). The down-regulation of the HMW SLP and Slp paralogue 2 in response to heat-stress is consistent with the findings from the growth and antibiotic experiments. The function of Slp paralogue 2 remains unknown but its high levels of expression and sensitivity to stress indicates that it probably has an important role for the bacterium. The reasons for the sensitivity of the HMW SLP to stress are also unclear. Further proteins may also be down regulated after heat shock, but cannot be seen in this study due to long half-lives, or an inability to be resolved by 2-DE.

Finally, although the differential expression of some of the identified proteins could not be confirmed, 24 different proteins were identified in these extracts that were not identified in the mapping of the surface proteome in Chapter 4. Therefore this study has also established the expression of these proteins in the *C. difficile* surface sub-proteome. The identities of these proteins are: CD0070 FusA translation elongation

factor, CD0071 TufB elongation factor Tu, CD0084 RplE ribosomal protein, CD0194 GroEL 60 kDa chaperonin, CD0236 FliS2 flagellar protein, CD0400 EtfB1 electron transport flavoprotein, CD0410 (CTn2) putative DNA replication protein, CD1055 EtfB2 electron transport flavoprotein, CD1056 EtfA2 acetyl CoA acetyl transferase, CD1175 acetate kinase, CD1350 permease from lantibiotic ABC transporter, CD2270 phosphofructokinase, CD2321 Tkt transketolase, CD2335 GuaB inosine-5'-monophosphate dehydrogenase, CD2365 sulphonate ABC transporter solute binding lipoprotein, CD2726 serine hydroxymethyltransferase, CD2966 AdhE, aldehyde-alcohol dehydrogenase, CD3181 putative aminohydrolase, CD3305 ClpP1 protease, CD3306 Tig trigger factor, CD3317 formate dehydrogenase H, CD3525 iron ABC transporter component, CD3641 conserved hypothetical protein and CD3642 selenocysteine synthase. The expression and localisation of some of these proteins have not previously been reported.

In conclusion, a wide variety of proteins respond to specific types of stress in *C. difficile*. Two proteins were down regulated in response to all three types of stress studied here, the HMW SLP and Slp paralogue 2. The reasons behind the sensitivity of these two proteins to general stress is not apparent from this analysis, and merits further study in the future. A possible method to investigate this would be to use metabolic labelling and construct a pulse-chase experiment to monitor the degradation of certain proteins over time or under different conditions. Alternatively, the regulation of these proteins could be investigated at the level of transcription, using RT-PCR. In addition further proteomics-based experiments could be used to characterise the role these proteins play during the course of a *C. difficile* infection.

Chapter 6

The immunoproteome of *Clostridium difficile*

6.1 Introduction

The immunoproteome can be defined as the set of proteins produced by a pathogen during the course of an infection that is recognised by the immune system of the host. The immunoproteome can be studied using the antiserum produced by the infected host. Certain antigens expressed by the pathogen will stimulate the production of antibodies that could eventually lead to the clearing of the infection by the host immune system: such antigens are termed 'protective' and can form the basis of vaccines that protect against disease.

Studies of the immunoproteome of pathogens are facilitating the rapid development of vaccines. Classically, some anti-bacterial vaccines have been based on killed (e.g. cholera, pertussis) or attenuated (e.g. tuberculosis) bacteria and although this complex mixture of antigens can induce a broad protective response, there is a greater risk of adverse effects. Therefore modern vaccines tend to be based on subunit vaccines, which can be composed of one or a mixture of a few defined protective antigens. The selection of protective antigens in a vaccine is critical. For example, only 2.5% of antigens identified in *Helicobacter pylori* were protective against disease (Ferrero & Labigne, 2001). Although poorly defined, there are a number of relevant antigen parameters that can preselect promising vaccine candidates. In enteric bacteria, antigens that are surface exposed are more likely to illicit a protective immune response, as their localisation makes them more likely to be exposed to the stomach mucosa and antigen presenting cells (Mai *et al.*, 1992). Highly abundant antigens are likely to be more immunogenic than low-abundance proteins since immune responses are generally dose dependent (Bumann *et al.*, 2004). And antigens that are seroreactive make promising vaccine candidates, as they must be expressed *in vivo* and be accessible to the host immune system.

Proteomics provides one solution to assess large numbers of antigens for vaccine candidate parameters and can enhance the hit rate for protective antigen selection. For example, the combined subproteomes of *H. pylori*, the causative agent of gastric ulcers and cancer (Genta, 1997), have been analysed (Bumann *et al.*, 2004). At least 8 of the 15 vaccine candidates predicted by the proteomic analysis were protective, compared to 10 out of 400 found by more traditional methods (Ferrero & Labigne, 2001), a substantially superior hit rate.

The aim of this study was to map the immunoproteome of *C. difficile*. Initially the antigens recognised by the serum of immunised rabbits were determined. Then the

study was extended to look at the antigens recognised by human patient sera, and the response of patients experiencing single episode and recurrent infections were compared.

6.2 Results: The immunoproteome of *C. difficile* recognised by rabbit antiserum

6.2.1 Generation of antibodies to *C. difficile* surface proteins

Antibodies against the surface proteins of *C. difficile* 630 were generated as described in section 2.4.4 following the method of Poxton (Poxton, 1979). New Zealand white rabbits were inoculated four times with 10^9 whole, heat-killed, formalin fixed *C. difficile* cells, in order to generate antibodies against surface exposed antigens. Bacteria were fixed with formalin to prevent cell lysis, and so an antibody response should be raised primarily against surface antigens. However, the specificity of antibodies to the cell surface may not be complete due to the possibility of macrophage ingesting and displaying cytoplasmic antigens. Therefore, the specificity of the raised antibodies to the surface of the bacterium was assessed qualitatively by immuno-gold electron microscopy (EM), carried out by Dave Goulding (Imperial College, London). Briefly, sections of *C. difficile* were incubated with the rabbit antiserum, and the antibodies were detected using Protein A-gold particles. The gold particles were detected by transmission electron microscopy (Figure 6.1), and are primarily localised at the cell surface, indicating the specificity of the antibodies to the cell surface. This antiserum will be referred to as rabbit anti-surface *C. difficile* 630.

6.2.2 SDS-PAGE and Western blots

Using the methods developed in Chapter 4, the surface proteins were prepared from *C. difficile* 630 cells by glycine and lysozyme treatment. Protein extracted by these treatments (10 μg), and protein from bacterial cells lysed by boiling for 10 min (10 μg) were loaded in duplicate onto SDS-PAGE gels. The proteins on one gel were stained using Coomassie blue while the proteins on the duplicate gel were transferred to nitrocellulose, and probed by Western blot using the rabbit anti-*C. difficile* surface proteins antibodies generated above (section 6.2.1). A secondary anti-rabbit antibody conjugated to HRP in combination with ECL was used to detect seroreactive proteins (Figure 6.2). Pre-immune serum did not recognise any *C. difficile* proteins (data not shown). A large number of proteins were recognised by the antiserum in all extracts, confirming that the sera contained antibodies reacting with surface associated proteins. Western blots were performed on extracts from cells during early-exponential growth,

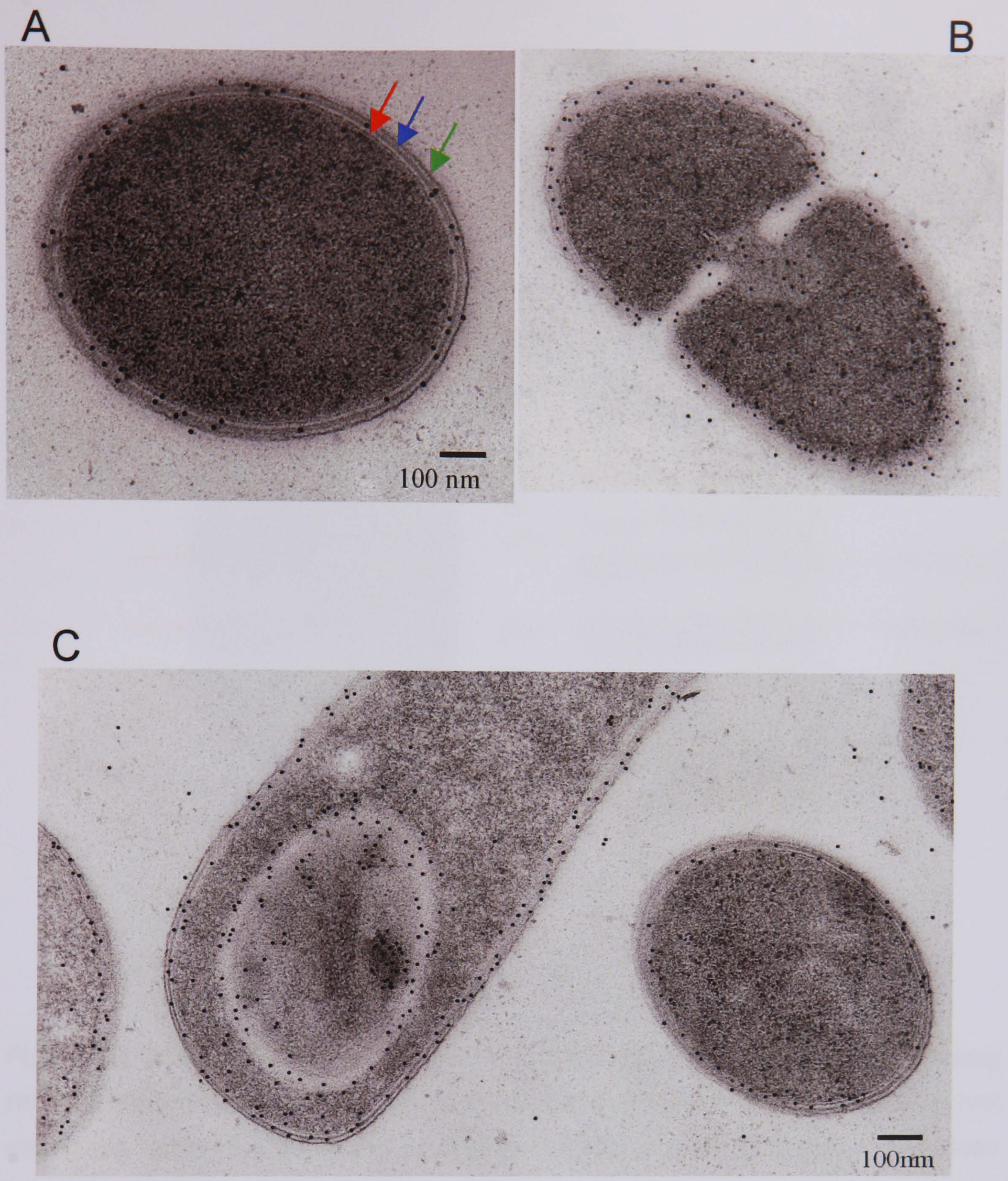


Figure 6.1 EM images of immuno-gold labelling of antiserum raised against *C. difficile* 630. A whole cell (**A**), a dividing cell (**B**) and a sporulating cell (**C**) are shown. The red arrow indicates the cell membrane, the blue arrow the cell wall and the green arrow the S-layer. Gold particles can be seen as defined circular black spots and have preferred specificity to the surface of all the cells, and the endospore.

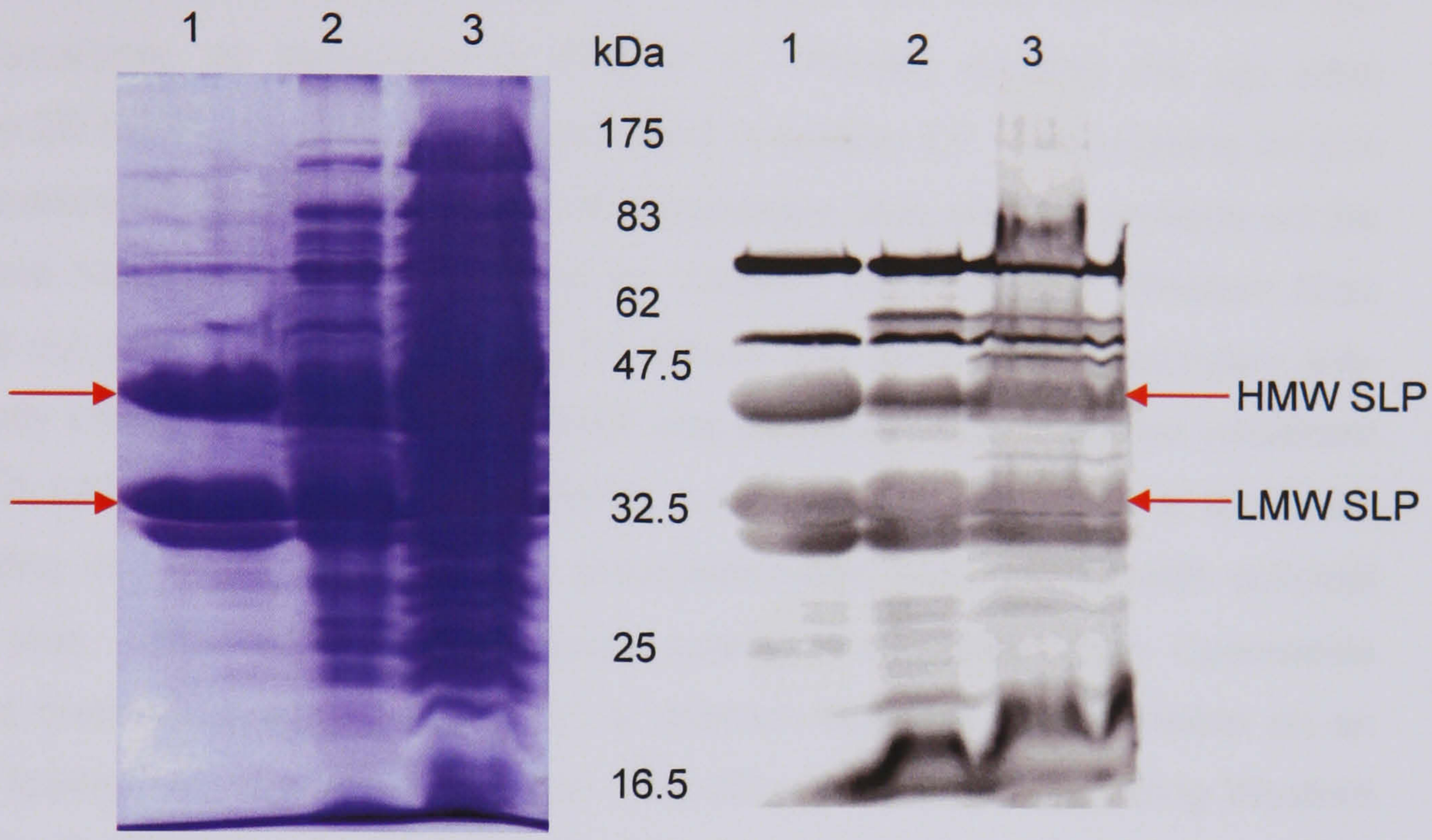


Figure 6.2 SDS-PAGE and Western blot of analysis of *C. difficile* proteins using rabbit antibodies. **A.** SDS-PAGE stained with Coomassie blue. **B.** Western blot using rabbit antibodies generated against the surface antigens of *C. difficile* visualised by HRP and ECL. Lane 1, glycine extracted surface proteins; lane 2, lysozyme extracted surface proteins; lane 3, total lysate. Molecular mass markers are indicated in kDa and the SLPs are indicated by arrows.

late exponential growth and during stationary phase. No significant difference in seroreactivity was seen between these time points (data not shown), and so all future experiments were carried out on extracts from cultures during late-exponential growth (6 hours culture time).

6.2.3 2D SDS-PAGE and Western blots

Proteins were extracted from the surface of *C. difficile* 630 cells by treatment with glycine or lysozyme, as developed in Chapter 4. Protein extracts (50 μg) were separated by 2D PAGE, in duplicate, as described in section 2.7. The proteins on one gel were visualised by staining with colloidal Coomassie blue and the proteins on the other gel were transferred to nitrocellulose for Western blot analysis. Western blots were carried out using rabbit anti-surface *C. difficile* 630 sera and a secondary anti-rabbit antibody conjugated to HRP. The HRP was detected using ECL and visualised using a FUJI LAS-3000 imager. In addition, a second gel containing a ten times greater loading of protein (500 μg) was electrophoresed and stained with colloidal Coomassie blue. ECL detects much lower quantities of protein than Coomassie staining, and therefore a protein visible on a Western blot may not be visible on an equivalently loaded Coomassie stained gel. The 2D gels and corresponding Western blots were compared (Figures 6.3 (glycine extracts), 6.4 (lysozyme extracts) and 6.5 (total bacterial protein)).

The 2D gels and Western blots were overlaid, and the seroreactive protein spots excised from the stained gel and analysed by MS (Jerry Thomas, University of York). The identities of the protein spots circled in red in Figures 6.3, 6.4 and 6.5 were determined and are detailed in Tables 6.1, 6.2 and 6.3, respectively. The fourteen protein spots identified from the glycine extracts were comprised of just five different proteins. The rabbit anti-surface *C. difficile* 630 antibodies recognised both surface layer proteins (SLPs), Slp paralogues 2 and 15 and flagellin. All these proteins are predicted to have a cell surface localisation and have been detected previously in surface extracts (Chapter 4). Analysis of fifteen protein spots from the lysozyme extracts led to the identification of eight different proteins. In common with the glycine extracts, both SLPs and Slp paralogue 2 were recognised by the rabbit antiserum. In addition, a further five proteins were recognised by the rabbit cell surface antiserum: AppA oligopeptide ABC transporter, enolase, amino acid transferase, selenocysteine synthase and threonine dehydrogenase. The gene for the AppA component of the oligopeptide transporter encodes a signal sequence, and therefore the AppA protein would be expected to be located at the cell surface. Although the other four proteins

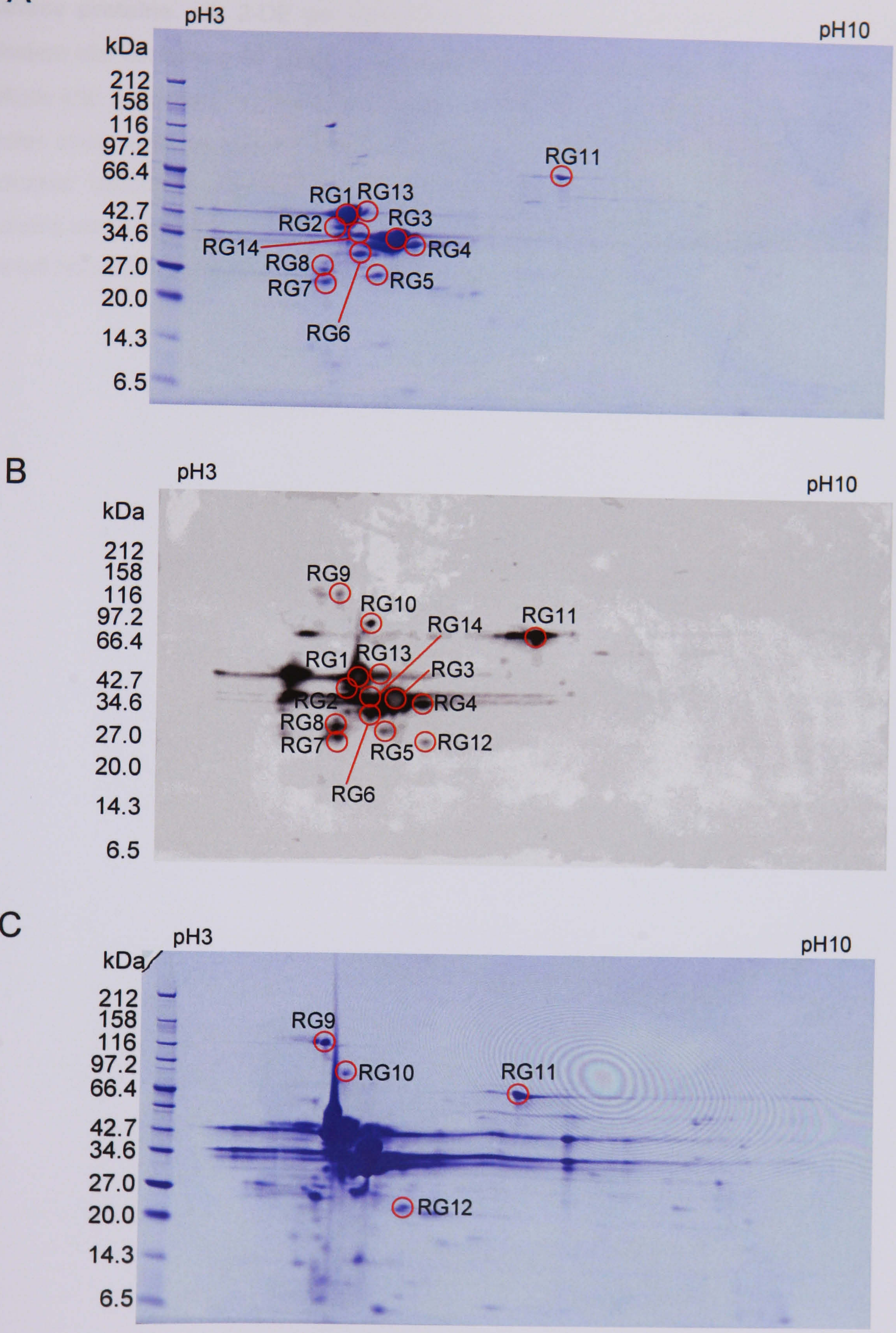


Figure 6.3 2-DE and Western blot of glycine extracted *C. difficile* surface proteins. Legend continued overleaf.

Figure 6.3 continued. 2-DE and Western blot of glycine extracted *C. difficile* surface proteins. **A.** 2-DE gel containing 50 μg glycine extracted protein. **B.** 2D Western blot containing 50 μg glycine extracted protein and using rabbit anti-surface *C. difficile* 630 antibodies. **C.** 2-DE gel containing 500 μg glycine extracted protein. Red circles show spots picked for MS/MS analysis. Proteins are labelled RGx, where R indicates detection with rabbit antibodies, G indicates glycine extracted surface proteins and x is the protein spot number. Molecular mass markers (kDa) are shown on the left hand side of each gel and the pH indicated above the gel.

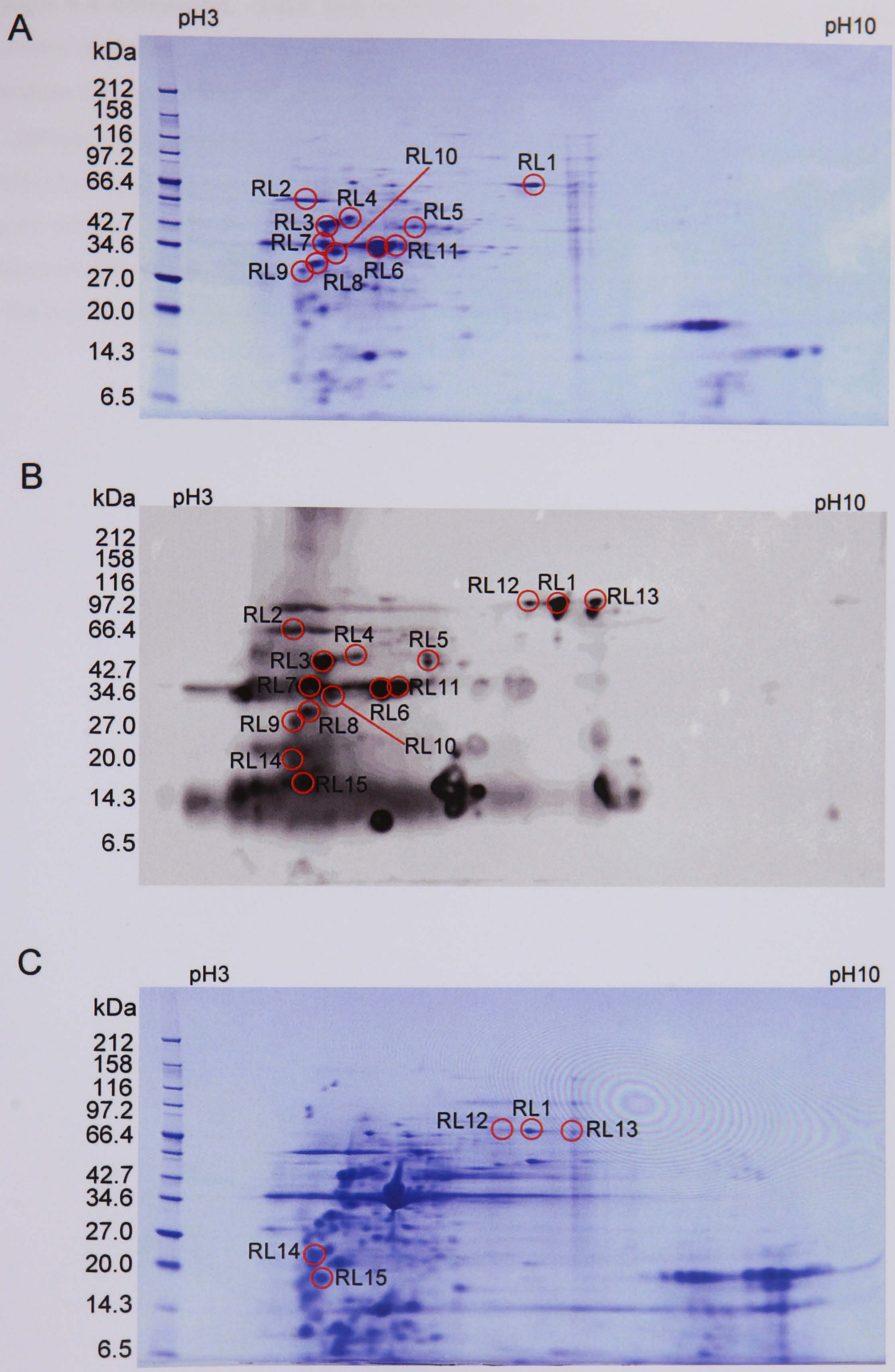


Figure 6.4 2-DE and Western blot of lysozyme extracted *C. difficile* surface proteins. Legend continued overleaf.

Figure 6.4 continued. 2-DE and Western blot of lysozyme extracted *C. difficile* surface proteins. **A.** 2-DE gel containing 50 μg lysozyme extracted protein. **B.** 2D Western blot containing 50 μg lysozyme extracted protein and using rabbit anti-surface *C. difficile* 630 antibodies. **C.** 2-DE gel containing 500 μg lysozyme extracted protein. Molecular mass markers are indicated in kDa and pH above the gel. Red circles show spots picked for MS/MS analysis. Proteins are labelled RLx, where R indicates detection with rabbit antibodies, L indicates lysozyme extracted surface proteins and x is the protein spot number.

A



B

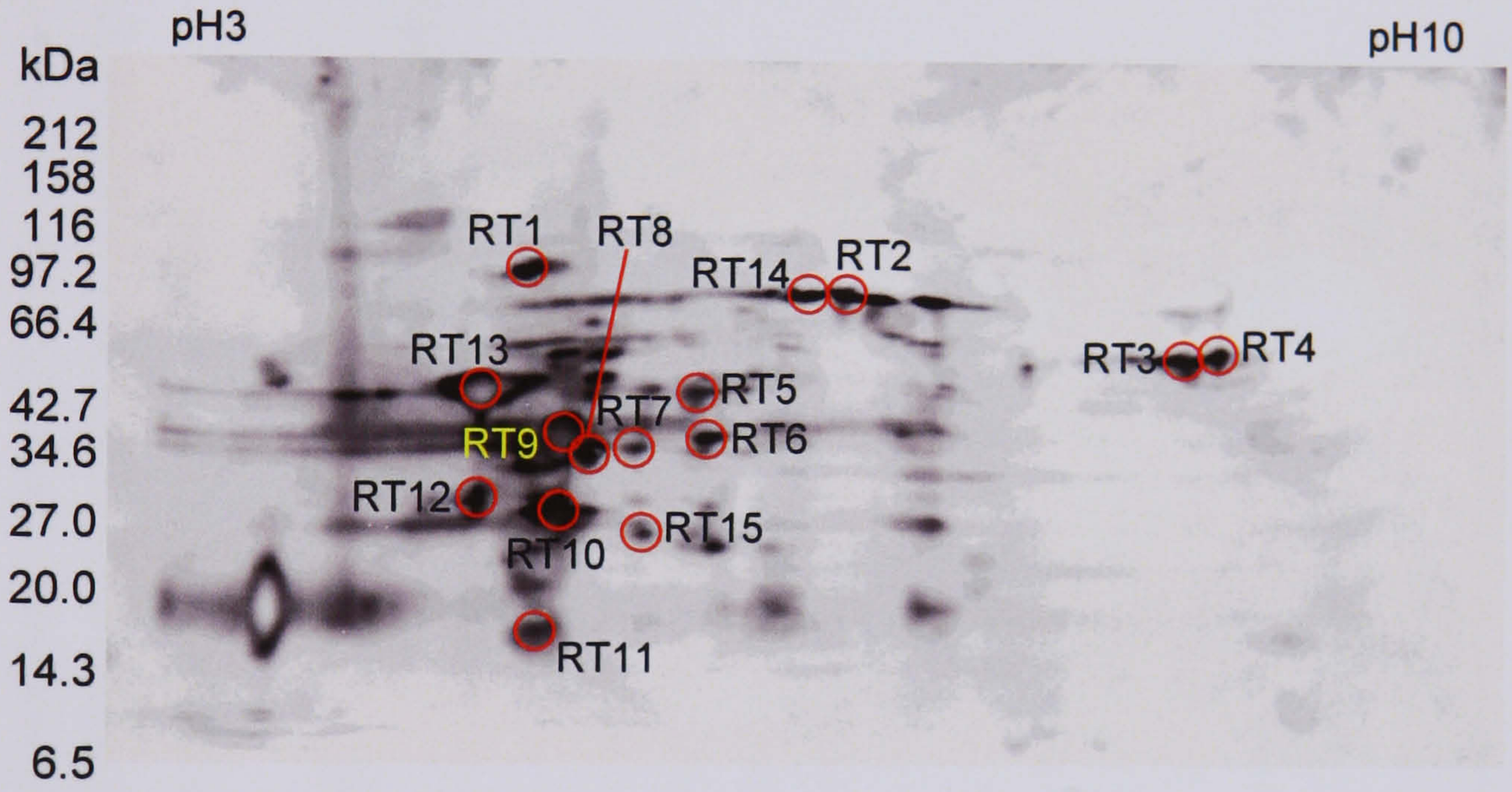


Figure 6.5 2-DE and Western blot of total proteins from *C. difficile*. **A** 2-DE gel containing 50 μg total protein. **B** 2D Western blot containing 50 μg total protein and using rabbit anti-surface *C. difficile* 630 antibodies. Molecular weight markers are indicated in kDa and the pH above the gel. Red circles show spots picked for MS/MS analysis. Proteins are labelled RTx, where R indicates detection with rabbit antibodies, T indicates total protein extract and x is the protein spot number.

Table 6.1 Proteins identified from 2D gels of glycine extracted proteins recognised by rabbit antiserum to surface proteins. Spots are labelled as detailed in Figure 6.3. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Total ion score was determined by Applied Biosystems. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
RG1	CD2793	HMW SLP	990	39.3	9 (147)	76115 (39504)	4.6	√
RG2	CD2793	HMW SLP	1035	57.5	9 (215)	76115 (39504)	4.6	√
RG3	CD2793	LMW SLP	914	35.8	8 (115)	76115 (34228)	4.9	√
RG4	CD0239	FliC	522	11.6	6 (101)	30755	8.6	x
RG5	CD2793	HMW SLP	405	14.7	4 (55)	76115 (39504)	4.6	√
RG6	CD0239	FliC	576	11.8	7 (103)	30755	8.6	x
RG7	CD2793	HMW SLP	214	11.8	3 (44)	76115 (39504)	4.6	√
RG8	CD0239	FliC	340	10.2	6 (89)	30755	8.6	x
RG9	CD0514	Slp paralogue 15	590	2.7	8 (133)	167580	4.7	√
RG10	CD2793	HMW SLP	155	8.3	2 (31)	76115 (39504)	4.6	√
RG11	CD2791	Slp paralogue 2	608	8.2	10 (154)	66427	6.4	√
RG12	CD0239	FliC	347	7.7	4 (67)	30755	8.6	x
RG13	CD2793	HMW SLP	80	7.9	3 (57)	76115 (39504)	4.6	√
RG14	CD2793	HMW SLP	10	3.6	1 (26)	76115 (39504)	4.6	√

Table 6.2 Proteins identified from 2D gels of lysozyme extracted proteins recognised by rabbit antiserum to surface proteins. Spots are labelled as detailed in Figure 6.4. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Total ion score was determined by Applied Biosystems. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
RL1	CD2791	Slp paralogue 2	177	4.2	5 (78)	66427	6.4	√
RL2	CD2672	AppA, oligopeptide transporter	393	7.1	9 (111)	59706	5.6	√
RL3	CD2793	HMW SLP	943	45.7	9 (171)	76115 (39504)	4.6	√
RL4	CD3170	Enolase	716	10.4	9 (134)	48149	4.6	x
RL5	CD3664	Putative amino acid transferase	274	5.6	6 (67)	44779	5.2	x
RL6	CD2793	LMW SLP	944	41.7	9 (134)	76115 (34228)	4.9	√
RL7	CD2793	HMW SLP	683	33.4	8 (125)	76115 (39504)	4.6	√
RL8	CD2793	HMW SLP	732	43.6	9 (163)	76115 (39504)	4.6	√
RL9	CD2793	HMW SLP	781	34.8	8 (130)	76115 (39504)	4.6	√
RL10	CD3642	Putative selenocysteine synthase	19	3.52	1 (13)	40516	5.6	x
RL11	CD1792	Putative L-threonine dehydrogenase	23	5.4	1 (17)	35616	5.2	x
RL12	CD2791	Slp paralogue 2	57	1.2	1 (8)	66427	6.4	√
RL13	CD2791	Slp paralogue 2	19	3.5	1 (22)	66427	6.4	√
RL14	CD2793	HMW SLP	69	3.3	2 (24)	76115 (39504)	4.6	√
RL15	CD2793	HMW SLP	61	3.3	2 (24)	76115 (39504)	4.6	√

Table 6.3 Proteins identified from 2D gels of total protein of *C. difficile* recognised by rabbit antiserum to surface proteins. Spots are labelled as detailed in Figure 6.5. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Total ion score was determined by Applied Biosystems. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
RT1	CD2461	DnaK chaperone	513	8.8	10 (162)	66449	4.7	x
RT2	CD2791	Slp paralogue 2	429	5.7	8 (106)	66427	6.4	√
RT3	CD2784	Slp paralogue 6	342	3.8	7 (77)	73024	7.2	√
RT4	CD1035	Slp paralogue 16	222	3.5	6 (72)	74678	8.7	√
RT5	CD1059	ThlA1, acetyl-CoA acetyltransferase	260	4.1	5 (50)	43171	5.4	√
RT6	CD1594	CysM, putative O-acetylserine sulfhydrylase	348	7.2	5 (65)	32647	5.4	x
RT7	CD2321	Tkt, transketolase	358	7.8	6 (77)	32342	5.1	x
RT8	CD0273	HtpG chaperone	442	5.0	8 (96)	74924	4.9	x
RT9	CD2793	HMW SLP	234	12.6	3 (47)	76115 (39504)	4.6	√
RT10	CD0400	EtfB1, electron transfer flavoprotein	621	13.1	8 (103)	28671	4.9	x
RT11	CD0064	RplL, 50S ribosomal protein	490	20.3	6 (74)	12620	4.8	x
RT12	CD0400	EtfB1, electron transfer flavoprotein	589	14.3	9 (112)	28671	4.9	x
RT13	CD2793	HMW SLP	793	35.6	7 (133)	76115 (39504)	4.6	√
RT14	CD2791	Slp paralogue 2	50	6.4	3 (40)	66427	6.4	√
RT15	CD0091	Adk, adenylate kinase	46	4.6	1 (10)	23700	5.3	x

are not predicted to be exported from the cell, three have been found previously in surface extracts (Chapters 4 and 5), namely enolase, amino acid transferase and selenocysteine synthase. Perhaps unsurprisingly, the greatest diversity of seroreactive proteins was found in the total protein extracts. Of the fifteen protein spots identified, eleven different proteins were detected. Two of these proteins are common to the surface extracts, the HMW SLP and Slp paralogue 2. Two further Slp paralogues (6 and 16) also reacted with the rabbit antibodies. The other seroreactive proteins have a wide variety of functions, including two chaperones, a ribosomal protein, metabolic and biosynthetic enzymes. A number of seroreactive proteins remain unidentified from this extract, including some dominant proteins of low MW and acidic pI, as it was occasionaly difficult to determine which Coomassie stained protein corresponded to a protein on the Western blot.

6.3 The immunoproteome of *C. difficile* recognised by human patient sera

6.3.1 Background

C. difficile strain M9 (PCR ribotype 017, toxin A^+/B^-) caused an outbreak of CDAD in a Dublin hospital. The strain and the blood serum of six patients infected by this strain were provided by Dr. Denise Drudy (University College, Dublin). Four of the patients had a single episode of CDAD, whereas the two other patients experienced recurrent infections. For all of the patients, blood serum was obtained on the first day CDAD was confirmed, and again 12 days later.

6.3.2 1D SDS-PAGE and Western blot

Proteins extracted from the surface of *C. difficile* M9 cells by glycine or lysozyme treatments, and total cell lysates were separated by 1D SDS-PAGE. Replicate protein gels were transferred to nitrocellulose and Western blots performed with the twelve samples of patient serum (diluted 1:500). The primary antibodies were detected by anti-human IgG, IgA and IgM antibodies conjugated to HRP, in combination with ECL (Figure 6.6). All sera reacted with multiple proteins from all of the cell extracts, confirming that *C. difficile* proteins were recognised by components of the immune system of all 6 patients. In general, the sera taken twelve days after CDAD had been confirmed recognised more proteins than the sera taken on the day CDAD was confirmed. This suggests that all six patients were mounting an immune response against their infections. The strong antibody response seen in all patients on day 1 of the infection may be a result of a previous infection with *C. difficile*, or it may be that the patients were suffering from the infection for a number of days before it was confirmed.

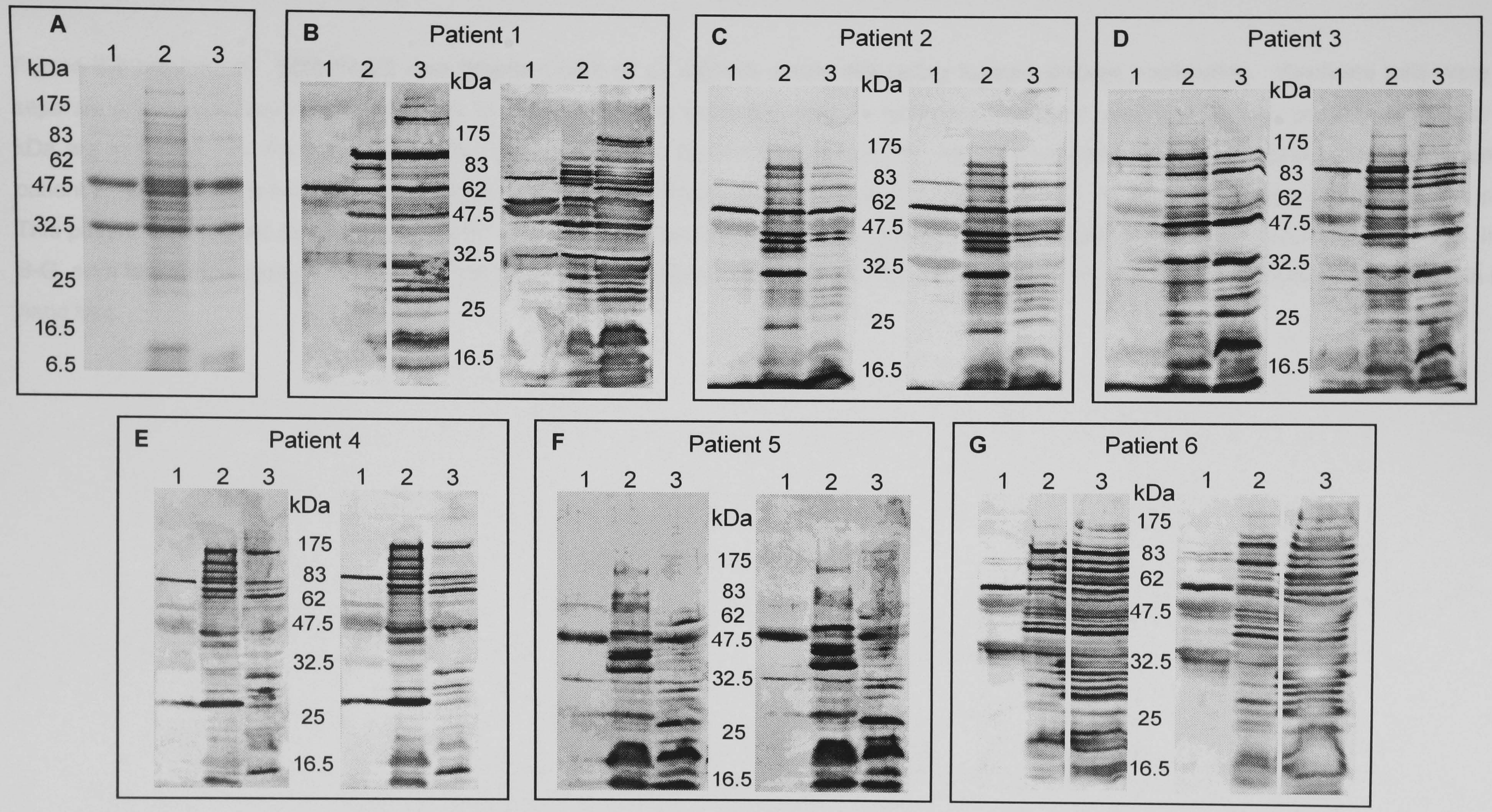


Figure 6.6 1D SDS-PAGE and Western blot of *C. difficile* strain M9 using human patient antibodies. Legend continued overleaf.

Figure 6.6 continued. SDS-PAGE and Western blot of *C. difficile* strain M9 using human patient antibodies. Replicate gels were used to separate proteins extracted from *C. difficile* by: lane 1 - glycine treatment, lane 2 - lysozyme treatment, and lane 3 - total protein. Molecular mass in kDa are indicated. **A.** Proteins stained by Coomassie blue. **B.** Proteins probed with serum from patient 1. **C.** Proteins probed with serum from patient 2. **D.** Proteins probed with serum from patient 3. **E.** Proteins probed with serum from patient 4. **F.** Proteins probed with serum from patient 5. This patient experienced recurrent infections. **G.** Proteins probed with serum from patient 6. This patient experienced recurrent infections. In panels B-G, sera taken from pateints on day 1 of the infection are shown in the left hand blot and sera taken on day 12 of the infection are shown in the right hand blot.

There is no discernable qualitative difference between the proteins recognised by the sera of the patients that suffered a single episode of infection (patients 1-4) compared to those that had relapsing infections (patients 5-6).

6.3.3 2D SDS-PAGE and Western blots

In order to determine which *C. difficile* proteins the patient antibodies recognised, the surface proteins were separated by 2-DE and probed again with the patient sera (Figure 6.7 and 6.8). Replicate gels containing 50 μg surface proteins extracted by treatment with lysozyme were electrophoresed. The proteins from six gels were transferred to nitrocellulose membranes and probed with the patient antisera taken on day twelve of the infection. The proteins from one gel were stained with colloidal Coomassie blue and a further gel containing 500 μg protein was also stained with Coomassie. Proteins that reacted with the patient antibodies were identified by overlaying images of the Western blots and Coomassie stained gels. Immunoreactive proteins were subjected to MS/MS to determine their identity (Jerry Thomas, University of York) (Table 6.4).

Each of the six patients recognised a discrete pattern of *C. difficile* proteins. The patient antibodies recognised between six and 31 protein spots on the Western blots. Some of these proteins were common between patients but others were unique to a single patient. The patterns of immunoreactive proteins recognised by the antibodies from the patients that suffered a single episode of CDAD (patients1-4, Figure 6.7A-D) were not significantly different to those from the two patients that suffered relapsing infections (patients 5 and 6, Figure 6.7E-F). The range of immunoreactive proteins identified was diverse (42 different proteins), although some proteins were common to a number of patients. Five of the patient sera contained antibodies against the LMW and HMW SLPs, and to Slp paralogue 2. These were by far the most common antigens, but other antigens recognised by the serum of more than one patient include Slp paralogue 4, the chaperones DnaK and HtpG, enolase, AppA component of an oligopeptide ABC transporter, a lipoprotein of an ABC transporter, the translation elongation factors TufA/B, CD0119 phosphomannomutase, CD0395 E-cinnamoyl-CoA:R-phenyllactate CoA transferase, CD0399 acyl coA dehydrogenase, CD1594 acetyl serine sulfhydrolase, CD2150 a putative endonuclease and CD3394 pyruvate kinase. Not all of these proteins are predicted to be exported from the bacterium, and they have a wide range of functions. Most immunoreactive proteins were only antigenic in individual patients (55%), although in some instances the same protein

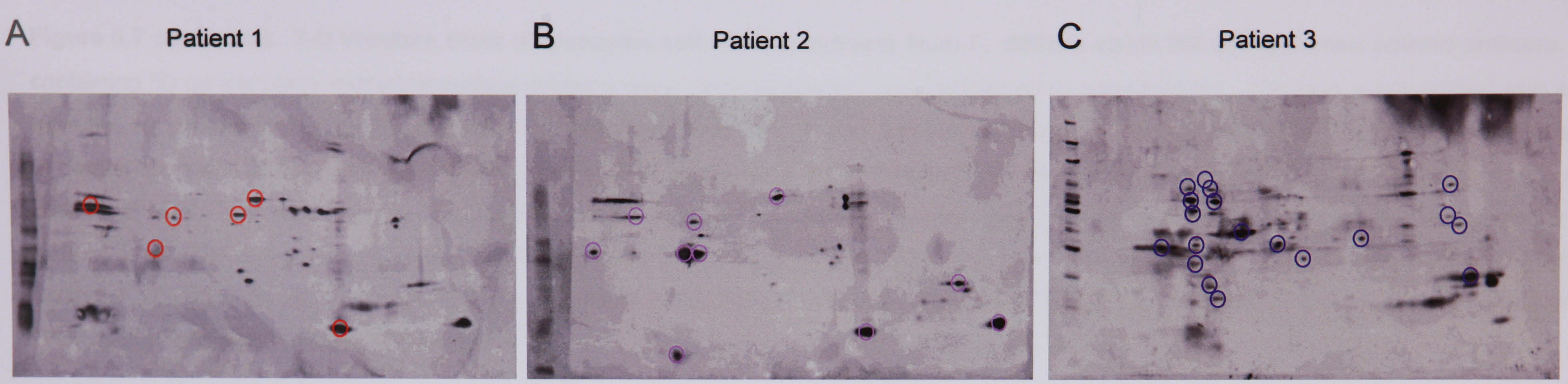


Figure 6.7 2-D Western blots of lysozyme cell surface extracts from *C. difficile* strain M9 using human patient antisera. Legend continued overleaf.

Figure 6.7 continued. 2-D Western blots of lysozyme cell surface extracts from *C. difficile* strain M9 using human patient antisera. 2D gels containing 50 μg lysozyme extracted surface proteins were electrophoresed, and transferred to nitrocellulose. The blots were probed with antiserum from the six patients, taken on day twelve of the infection. A mixture of goat anti-human IgG, IgA and IgM antibodies, conjugated to HRP, was used in combination with ECL to detect the primary antibodes. These blots were compared with replicate gels (Figure 6.8) stained by colloidal Coomassie blue, and where possible, immunoreactive proteins were picked for MS/MS analysis (indicated by circles). **A.** Western blot using antiserum from patient 1. **B.** Western blot using antiserum from patient 2. **C.** Western blot using antiserum from patient 3. **D.** Western blot using antiserum from patient 4. **E.** Western blot using antiserum from patient 5, this patient experienced recurrent infections. **F.** Western blot using antiserum from patient 6, this patient experienced recurrent infections.

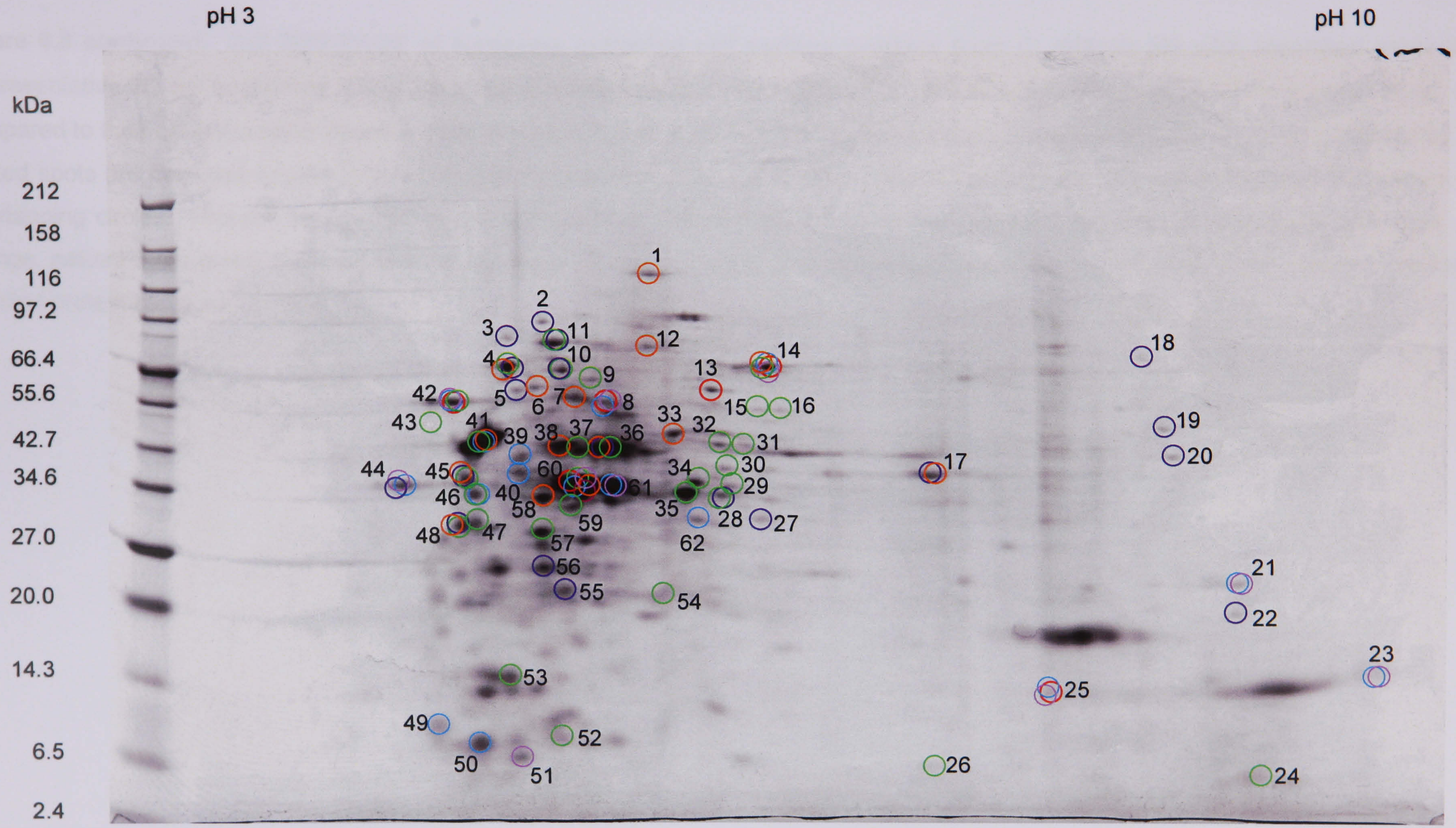


Figure 6.8 2-D SDS-PAGE of lysozyme extracted cell surface proteins from *C. difficile* M9 with identified seroreactive spots. Legend continued overleaf.

Figure 6.8 continued. 2-D SDS-PAGE of lysozyme extracted cell surface proteins from *C. difficile* M9 with identified seroreactive spots. Representative 2D gel containing 50 μg lysozyme extracted surface proteins from *C. difficile*, stained with colloidal Coomassie blue. 2D gels were compared to their corresponding Western blots (Figure 6.7) and immunoreactive proteins were picked for MS/MS analysis (indicated by circles). The picked spots are coloured according to which patient antiserum recognised them, and in many cases this was more than one patient, indicated by overlapping circles. Proteins recognised by the antiserum of patient 1 are coloured red, patient 2 are lilac, patient 3 are dark blue, patient 4 are orange, patient 5 are green and patient 6 are light blue. Molecular masses in kDa are indicated on the left and the pH above. The identities of the labelled protein spots are in Table 6.4.

Table 6.4 Proteins identified from 2D gels of total protein of *C. difficile* M9 recognised by human antiserum to surface proteins. Spots are labelled as described in Figure 6.8. Genome ID corresponds to the gene identification by The Sanger Institute (www.sanger.ac.uk/C_difficile). Total ion score was determined by Applied Biosystems. Predicted molecular mass (Da) based on sequences containing signal peptide (where applicable) except for the case of the HMW and LMW SLPs which show mass of full length SlpA and mass of mature protein in parentheses. Signal sequences were predicted by SignalP. √ indicates presence, x indicates absence of a signal peptide.

Spot	Patient	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
1	4	CD2682	NifJ, pyruvate-flavodoxin oxidoreductase	426	7.3	8 (86)	128884	5.4	x
2	3	CD0070	FusA, translation elongation factor G	584	14.8	10 (102)	75897	4.9	x
3	3	CD2793	SlpA	334	7.9	4 (57)	76115	4.8	√
4	3, 4, 5	CD2461	DnaK, chaperone	816	22.9	10 (141)	66449	4.7	x
5	3	CD2755	PtsI, phosphoenolpyruvate phosphotransferase	729	17.0	8 (97)	63132	4.8	x
6	4	CD3171	Pgm, phosphoglycerate mutase	257	12.0	6 (61)	56375	5.0	x
7	4	CD2335	GuaB, inosine-5'-monophosphate dehydrogenase	409	13.4	7 (67)	54778	5.3	x
8	1, 2, 6	CD2787	Cwp84, Slp paralogue 4	281	8.5	6 (68)	38577	6.7	√
9	5	CD0070	FusA, translation elongation factor G	353	12.4	9 (85)	75897	4.9	x
10	3, 5	CD0273	HtpG, chaperone	401	14.7	8 (95)	74924	4.9	x
11	3, 5	CD3394	PykF, pyruvate kinase	493	18.8	8 (110)	63092	5.0	x
12	4	CD0759	PlfB, formate acetyltransferase	508	15.1	8 (112)	83564	5.4	x
13	1	CD2786	Slp paralogue 5	23	1.5	1 (8)	57214	6.7	√
14	1, 2, 4, 5, 6	CD2791	Slp paralogue 2	1006	22.2	10 (138)	66427	6.4	√

Table 6.4 continued

Spot	Patient	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
15	5	CD2341	AbfD, gamma-aminobutyrate metabolism dehydratase/isomerase	33	1.4	1 (7)	55102	5.8	x
16	5	CD2341	AbfD, gamma-aminobutyrate metabolism dehydratase/isomerase	96	5.6	3 (28)	55102	5.8	x
17	3, 4	CD0399	Bcd1, acyl-CoA dehydrogenase	182	14.3	4 (54)	40864	6.4	x
18	3	CD0084	RplE, 50S ribosomal protein	33	4.4	1 (8)	20437	9.5	x
19	3	CD2767	Slp paralogue 19	124	4.7	4 (33)	77830	8.6	√
20	3	CD0084	RplE, 50S ribosomal protein	41	4.4	1 (8)	20437	9.5	x
21	2, 6	CD0119	Phosphoglucomutase/ phosphomannomutase mutase	27	1.3	1 (6)	51126	5.2	x
22	3	CD0084	RplE, 50S ribosomal protein	28	4.4	1 (8)	20437	9.5	x
23	2, 6	CD2150	Putative endonuclease/ exonuclease/phosphatase	16	10.0	2 (21)	24083	8.9	x
24	5	CD0085	RpsH, 30S ribosomal protein	195	34.1	5 (45)	14680	9.2	x
25	1, 2, 6	CD2150	Putative endonuclease/ exonuclease/ phosphatase	12	5.3	1 (11)	24083	8.9	x
26	5	CD1467	ABC transporter, permease protein	15	1.3	1 (11)	96241	9.4	√
27	3	CD1058	Hbd, 3-hydroxybutyryl-CoA dehydrogenase	235	18.1	5 (51)	30679	5.8	x
28	3, 5	CD1594	CysM, putative O-acetylserine sulfhydrylase	236	22.2	5 (67)	32647	5.4	x
29	5	CD3174	Glyceraldehyde-3-phosphate dehydrogenase	182	11.0	4 (37)	36015	5.9	x

Table 6.4 continued

Spot	Patient	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
30	5	CD0994	Putative aminotransferase	293	86.1	5 (62)	40045	5.6	x
31	5	CD2726	GlyA, putative serine hydroxymethyltransferase	135	11.1	4 (46)	45796	5.7	x
32	5	CD2726	GlyA, putative serine hydroxymethyltransferase	190	15.7	6 (65)	45796	5.7	x
33	4	CD3285	Pgi, glucose-6-phosphate isomerase	505	20.1	7 (94)	50538	5.6	x
34	5	CD3174	GapB, glyceraldehyde-3-phosphate dehydrogenase	673	28.1	9 (94)	36015	5.9	x
35	5	CD1594	CysM, putative O-acetylserine sulfhydrylase	236	20.2	4 (61)	32647	5.4	x
36	3, 4, 5	CD0395	FldA E-cinnamoyl-CoA:R-phenyllactate CoA transferase	437	18.8	6 (75)	44216	5.2	x
37	5	CD3094	Putative sigma-54-dependent transcriptional regulator	22	2.2	2 (15)	76607	7.1	x
38	4	CD0058/ CD0071	TufA, elongation factor Tu	507	26.4	8 (105)	44027	4.9	x
39	6	CD2793	HMW SLP	645	24.9	6 (93)	76115 (39504)	4.6	√
40	6	CD2793	LMW SLP	166	5.9	2 (19)	76115 (34228)	4.9	√
41	4, 5	CD3170	Eno, enolase	692	24.2	7 (104)	48149	4.6	x
42	1, 2, 5, 6	CD2672	AppA, oligopeptide ABC transporter	538	19.1	9 (99)	58164	5.6	√
43	5	CD3306	Tig, trigger factor	255	11.7	5 (50)	48187	4.6	x
44	2, 3, 6	CD0873	ABC transporter, substrate-binding lipoprotein	515	25.3	6 (92)	36067	5.4	√
45	3, 4, 5	CD2793	HMW SLP	627	29.7	8 (111)	76115 (39504)	4.6	√

Table 6.4 continued

Spot	Patient	Genome ID	Description	Total Ion Score	% coverage	Number of peptides (residues)	Theoretical molecular mass	Theoretical pI	Signal sequence
46	5	CD2793	HMW SLP	549	21.9	5 (82)	76115 (39504)	4.6	√
47	5	CD2793	HMW SLP	966	35.0	8 (131)	76115 (39504)	4.6	√
48	3, 4, 5	CD2793	HMW SLP	560	26.5	6 (99)	76115 (39504)	4.6	√
49	6	CD2793	HMW SLP	567	23.0	6 (86)	76115 (39504)	4.6	√
50	6	CD2793	HMW SLP	490	17.1	7 (64)	76115 (39504)	4.6	√
51	2	CD2793	HMW SLP	191	5.9	2 (22)	76115 (39504)	4.6	√
52	5	CD2857	Putative hydrolase	20	3.1	1 (7)	25776	5.6	x
53	5	CD0064	RplL, 50S ribosomal protein	462	53.7	6 (65)	12621	4.8	x
54	5	CD2139	Tsf, elongation factor Ts	156	13.5	4 (41)	33122	5.2	
55	3	CD0058/ CD0071	TufA, elongation factor Tu	398	20.4	6 (81)	44027	4.9	x
56	3	CD0058/ CD0071	TufA, elongation factor Tu	472	20.4	6 (81)	44027	4.9	x
57	5	CD3172	TpiA, triosephosphate isomerase	549	35.6	7 (88)	27174	5.0	
58	4	CD0239	FliC, flagellin	159	12.1	2 (35)	30755	8.6	x
59	5	CD2321	Tkt', transketolase	395	33.3	7 (102)	32342	5.1	x
60	1, 2, 4, 5, 6	CD2793	LMW SLP	167	5.9	2 (19)	76115 (34228)	4.9	√
61	2, 6	CD2793	LMW SLP	156	5.9	2 (19)	76115 (34228)	4.9	√
62	6	CD1047	Slp paralogue 18	335	20.0	6 (70)	37403	8.9	√

was detected in multiple spots on the same gel (e.g. patient 5 spots 31 and 32, or spots 28 and 35).

6.4 Discussion

Preliminary experiments using rabbits injected intramuscularly with heat killed *C. difficile* revealed antibodies against at least 20 different proteins, many of which had been previously identified in cell surface extracts from *C. difficile*. These seroreactive proteins have a wide range of predicted functions, although not all would be expected to be located at the cell surface based on their DNA sequences. This experiment confirms the expression of these proteins, and shows their capacity to induce an antibody response. These proteins are also likely to have a cell surface localisation.

The immune systems of human patients was found to recognise and produce an antibody response against a large number and wide diversity of antigens during infection with *C. difficile.* Between six and 31 immunoreactive proteins were identified in the immunoproteomes from each of the six CDAD patients. From the six patients, a total of 42 different immunoreactive proteins were identified. Seventeen of these proteins were common antigens in multiple patients (detailed in section 6.3.3), and these proteins are particularly worthy of further study. Interestingly, eight of the antigens recognised by more than one patient were also recognised in the rabbit immunoproteome. These proteins are the two SLPs, Slp paralogue 2, the chaperones DnaK and HtpG, AppA oligopeptide ABC transporter, enolase and CysM acetyl serine sulfhydrolase. These eight proteins are particularly promising vaccine candidates and merit further investigation.

This proteomic approach has identified several proteins that fulfil the parameters discussed in section 6.1 for promising candidate antigens. The parameters that should be addressed in a strategy of rational vaccine design include abundance, localisation and seroreactivity of potential antigens. These parameters will be considered to further evaluate the eight promising vaccine candidates defined above. All eight proteins were identified in this study on the basis of their seroreactivity, and therefore all eight antigens conform to this parameter. In addition, all eight antigens were recognised from lysozyme extracted *C. difficile* proteins, and therefore fulfil the requirement for surface localisation. Indeed, the two SLPs, Slp paralogue 2 and AppA oligopeptide transporter are all predicted to encode signal peptides, and would be expected to have a surface localisation. Limitations with the computer algorithms used to predict the cellular localisation of proteins can mean that proteins predicted *in silico* to be

cytoplasmic actually have been shown by functional studies to associate with the cell surface (Chakravarti *et al.*, 2001). For example, heat shock proteins and enolase have been detected in studies on other bacteria at the cell surface (Ferrero & Labigne, 2001; Hughes *et al.*, 2002; Schaumburg *et al.*, 2004), and CysM acetyl serine sulfhydrolase was shown in Chapter 4 to be present on the cell surface of *C. difficile* 630. This means that it is possible that these antigens, although predicted to be cytoplasmic, are exported to the cell surface of *C. difficile*. In terms of protein abundance, the extracts analysed in this study show that three of the proteins (LMW- and HMW SLPs and Slp paralogue 2) are present at high levels in *C. difficile* cells. Although the remaining five proteins are not expressed at such high levels, they are all strongly stained by colloidal Coomassie blue, which detects proteins at levels greater than 10 ng. Therefore all eight antigens fulfil the requirements for a vaccine candidate. Even though only 2.5% of antigens are protective against disease (Ferrero & Labigne, 2001), selecting antigens in such a way resulted in a greater than 50% success rate at identifying protective antigens in *H. pylori* (Bumann *et al.*, 2004).

A number of studies have been published on the immunoproteome of other pathogenic microorganisms. Some of these antigens are common to this study in *C. difficile.* For example, seven seroreactive proteins were identified in *Bacillus anthracis* using guinea pig and rabbit antiserum (Chitlaru *et al.*, 2004). These antigens include the two *B. anthracis* SLPs (EA1 and Sap), two unknown proteins that contain surface-layer homology (SLH) domains, two heat shock proteins including DnaK and an oligopeptide transporter named OppA. Seroreactivity for DnaK has also been reported in *Francisella tularensis* (Havlasova *et al.*, 2002), *H. pylori* (Haas *et al.*, 2002), and *Streptococcus pneumoniae* (Hamel *et al.*, 1997). In addition, enolase was shown to be immunoreactive in *Candida albicans* (Fernández-Arenas *et al.*, 2004).

To date, attempts to produce a vaccine for CDAD have focussed primarily on the toxins TcdA and TcdB, and to a lesser extent the SLPs, to prevent disease. The many publications that explore the use of inactivated toxins to protect against disease have established that immunity against both toxins is necessary for complete protection in a hamster model of infection (Fernie *et al.*, 1983; Kim *et al.*, 1987b; Libby and Wilkins, 1982) and the essential effector molecules that afford this protection are antibodies (Giannasca *et al.*, 1999). In particular, the magnitude and kinetics of the IgG response to toxin A are critical in determining the clinical outcome of a *C. difficile* infection, with a strong response minimising the chances of developing diarrhoea and recurrent infections (Aronsson *et al.*, 1985; Leung *et al.*, 1991; Warny *et al.*, 1994; Johnson *et al.*, 1995; Kyne *et al.*, 2000, 2001). Due to the importance of generating an antibody

response against the toxins, the active and passive immunisation against *C. difficile* diarrhoea and colitis using toxoid preparations has been attempted, with varying levels of success (Torres *et al.*, 1995; Giannasca *et al.*, 1999). Other attempts to develop a vaccine to CDAD have focussed on the SLPs, which are known to elicit a high IgG response (Pantosti *et al.*, 1989), which is associated with a reduced risk of recurrent infections (Drudy *et al.*, 2004). Passive immunisation of hamsters using antibodies to the SLPs significantly prolongs post challenge survival in the hamster model, but does not prevent death (O'Brien *et al.*, 2005). Despite these advances in developing a vaccine to *C. difficile*, so far a fully protective method is yet to be found. To elicit a protective response from a broad range of patients, it is likely that a vaccine will have to be composed of several subunits. In light of these investigations, this study has the identified the two SLPs as strong vaccine candidates. The two large toxins were not identified in this study, presumably as their large size prevents them from entering 2D gels and it is unlikely they would be located in a cell wall fraction of the bacteria. However, this study has also identified five further vaccine candidates, which merit further investigation.

The first steps in designing a vaccine to protect against *C. difficile* infection include the identification of antigens that are produced by preferably all, or at least the majority, of strains that are problematic in the hospital environment. Another important point to consider is strain-to-strain primary amino acid sequence variability of the antigens, which should be minimal for good vaccine candidates. This should be studied by sequence analysis of the common strains in the clinical settings. Based on these premises, the eight proteins identified above should be considered for further investigation. The next stage in this project would be to test these antigens in an animal model of disease. Currently the best model is the Syrian hamster, which shows severe symptoms of diarrhoea upon oral feeding of *C. difficile* spores (Sambol *et al.*, 2001). It is also possible to monitor colonisation of the hamster by *C. difficile* using non-toxic strains. Therefore by immunising animals with one or more of these antigens, the ability to induce an immune response that slowed down or prevented colonisation by non-toxic strains could be studied. As it is likely that any future vaccine would include components of one or both of the two toxins TcdA and TcdB, any antigen identified could be used together with recombinant, non-toxic derivatives of the toxins, e.g. the cell binding domains, together with the antigens that showed the highest efficacy in prevention of colonisation. Functional studies of the antigens identified here could also help in the rational design of vaccine candidates, as the neutralisation of factors known to play an important role in the virulence of *C. difficile* will limit the development of disease.

Chapter 7

Studies of Slp paralogue 4 (Cwp84)

7.1 Introduction

7.1.1 Slp paralogue 4

The Slp paralogue 4 gene is the fifth open reading frame (ORF) downstream from *slpA*, the gene encoding both surface layer proteins (SLPs) (Figure 1.13). Like the other SlpA paralogues, it is comprised of a secretion signal, a conserved amidase-like domain and a third, variable domain that is different to the other paralogues. The variable, amino-terminal domain in Slp paralogue 4 displays amino acid homology with several cysteine proteases (Panagiotopoulou, 2001). Recently, a number of studies have looked at the expression and activity of this protein, which has been renamed Cwp84, signifying a clostridial wall protein of 84 kDa. The sequence variation of the *cwp84* gene has been analysed in 28 strains, and was shown to be very highly conserved (Savariau-Lacomme *et al.*, 2003). Furthermore, a Northern blot using a probe specific to *cwp84* showed that this gene is transcribed during early exponential growth of the bacteria, but not in late exponential or stationary phase (Savariau-Lacomme *et al.*, 2003). Savariau-Lacomme *et al* (2003) also expressed recombinant Cwp84 and studied its activity on a range of substrates. The full length Cwp84 protein expressed poorly in *Escherichia coli* as a fusion protein with glutathione *S*-transferase (GST), and could not be purified due to extensive proteolytic degradation. Instead, the protease activity of Cwp84 was investigated in *E. coli* cells containing the plasmid pGEXΩ*cwp84*. When the expression of Cwp84 was induced, a light degradation of gelatine, skimmed milk and azocoll was observed (Savariau-Lacomme *et al.*, 2003). In a subsequent study by the same group, the catalytic amino-terminal domain of *cwp84* was cloned into the expression system pGEX-6P-1. The truncated Cwp84 protein was expressed and purified as an in-frame fusion-protein tagged with GST (Janoir *et al.*, 2004). This time the truncated Cwp84 protein showed proteolytic activity on gelatine and *N*-Benzoyl-DL-arginine 4-nitroanilide hypochloride (BAPNA, a non-specific colorimetric substrate), but not on azocoll, in contrast to the first study. Inhibition experiments on the purified protein confirmed that Cwp84 belongs to the cysteine protease family.

The role that Cwp84 and a number of other *Clostridium difficile* surface proteins play during the course of an infection has recently been investigated (Pechine *et al.*, 2005b). The blood serum of several of *C. difficile* patients was assayed for the development of antibodies to a number of surface proteins, and 15 of the 17 patients tested produced antibodies specific to Cwp84. This confirms Cwp84 is expressed during the course of an infection and as it is a highly conserved protein, the authors postulate that it would make a good vaccine candidate.

Two studies have been published on the general proteolytic activity of *C. difficile* strains. Both studies detected varying types of proteolytic activity in all strains tested (Seddon & Borriello, 1992; Poilane *et al.*, 1998). In the study by Poilane and co-workers, two proteases were identified in the culture supernatant that degraded gelatine, and *C. difficile* isolates were also able to degrade collagen and casein. However, the levels of proteolytic activity detected were very low (Poilane *et al.*, 1998). In the study by Seddon and Borriello, an association was found between proteolytic activity and increased levels of virulence in a hamster model, which was not correlated to the toxigenic status of the isolates. The source of at least part of this proteolytic activity was found to be a cell-associated cysteine protease, which was isolated and characterised, and found to be most closely related to clostripain from *Clostridium histolyticum* (Seddon & Borriello, 1992). These studies confirm that *C. difficile* does have proteolytic activity, which may be a contributory factor in the virulence of the organism. Further studies utilising the *C. difficile* genome information (Sebaihia *et al.*, 2006) may allow a greater understanding of the protease activity of this bacterium.

7.1.2 Cysteine proteases

There are at least seven evolutionary lines of cysteine proteases, known as clans, and within each clan are one or more families of closely related cysteine proteases (Barrett and Rawlings, 2001). The largest clan (clan CA) contains about half the total families, and the archetypal cysteine protease, the plant peptidase papain.

The crystal structure of papain (Drenth *et al.*, 1968; Drenth *et al.*, 1971; Kamphuis *et al.*, 1984) and a number of other cysteine proteases (Kamphuis *et al.*, 1985) have been solved. These structures have revealed that the papain-like enzymes share a common fold, with two distinct domains. The mature enzyme is reminiscent of a closed book, with the left (L) domain comprising the amino-terminal half of the molecule and the right (R) domain comprising the carboxyl-terminal half, separated by a V-shaped active site cleft. In the middle of the cleft are the two catalytic residues, Cys25 and His159 (papain numbering) from the L- and R-domains respectively. The L-domain is predominantly α-helical in structure, and the R-domain forms a β-barrel, which also includes a shorter α-helix. Both the amino and carboxyl terminal ends of the polypeptide cross over to the other domain, creating a clamp holding the two domains closely together.

The mechanism of action of cysteine proteases has been investigated in considerable detail (reviewed in Grzonka *et al.* (2001)) (Figure 7.1). The enzymatic activity of

cysteine proteases is dependent upon a catalytic diad formed by a cysteine and histidine residue, which in the pH interval 3.5-8.0 exists as the ion pair thiolate and protonated imidazole. Nucleophilic attack of the thiolate group of the cysteine residue on the carbonyl group of the hydrolysed bond releases the carboxyl-terminal fragment of the cleaved product and an intermediate S-acyl-enzyme. This reacts with a water molecule, releasing the amino-terminal fragment and the regenerated free cysteine protease molecule. Uniquely to the papain-family of cysteine proteases, a third residue is also important in the catalytic mechanism. An asparagine residue (Asn175) is in close proximity to the active site histidine, and forms a hydrogen bond between the amide O of the asparagine side chain and the N of the histidine. This interaction is thought to orient the side chain of the histidine favourably for catalysis.

Cysteine proteases are often synthesized as pro-proteins. The pro-segments are typically around 100 residues and contain the ExxxRxxxFxxNxxxIxxxN (where x=any amino acid) sequence motif (known hereafter as the ERFNIN motif). The crystal structures of two cysteine protease pro-enzymes have been reported (Coulombe *et al.*, 1996; Groves *et al.*, 1996), and despite a lack of strong sequence homology, the pro-regions of these enzymes were structurally similar. The pro-regions consist of a common fold containing a helix-turn-strand hairpin super-secondary structure followed by an α-helix. These pro-regions are wedged into the substrate-binding site of the enzymes. The orientation of the pro-region is opposite to that of the peptidyl substrate, ensuring that the catalytic residues do not have the right spatial arrangement and cleavage does not occur. The pro-region also blocks access by the substrate, inhibiting enzymatic activity (reviewed in Groves *et al.* (1998)). The pro-regions of papain-like cysteine proteases sometimes fulfil further functions, such as protein folding, stabilisation and targeting of the enzyme (McIntyre & Erickson, 1993; Mach *et al.*, 1994; McIntyre *et al.*, 1994; Tao *et al.*, 1994; Bryan *et al.*, 1995). Activation of the enzyme occurs by removal of the pro-region. *In vitro*, this is an autocatalytic process, which could occur by an inter- or intra-molecular mechanism that is yet to be elucidated.

Cysteine proteases have been shown to play a role in bacterial virulence. An example of this is *Streptococcus pyogenes*, which produces an exotoxic cysteine protease known as SpeB. SpeB is able to degrade host proteins such as human fibronectin and vitronectin (Kapur, 1993). It was confirmed as a virulence factor when *speB* knockout mutants were evaluated in a mouse model and had lost their lethality (Lukomski *et al.*, 1997).

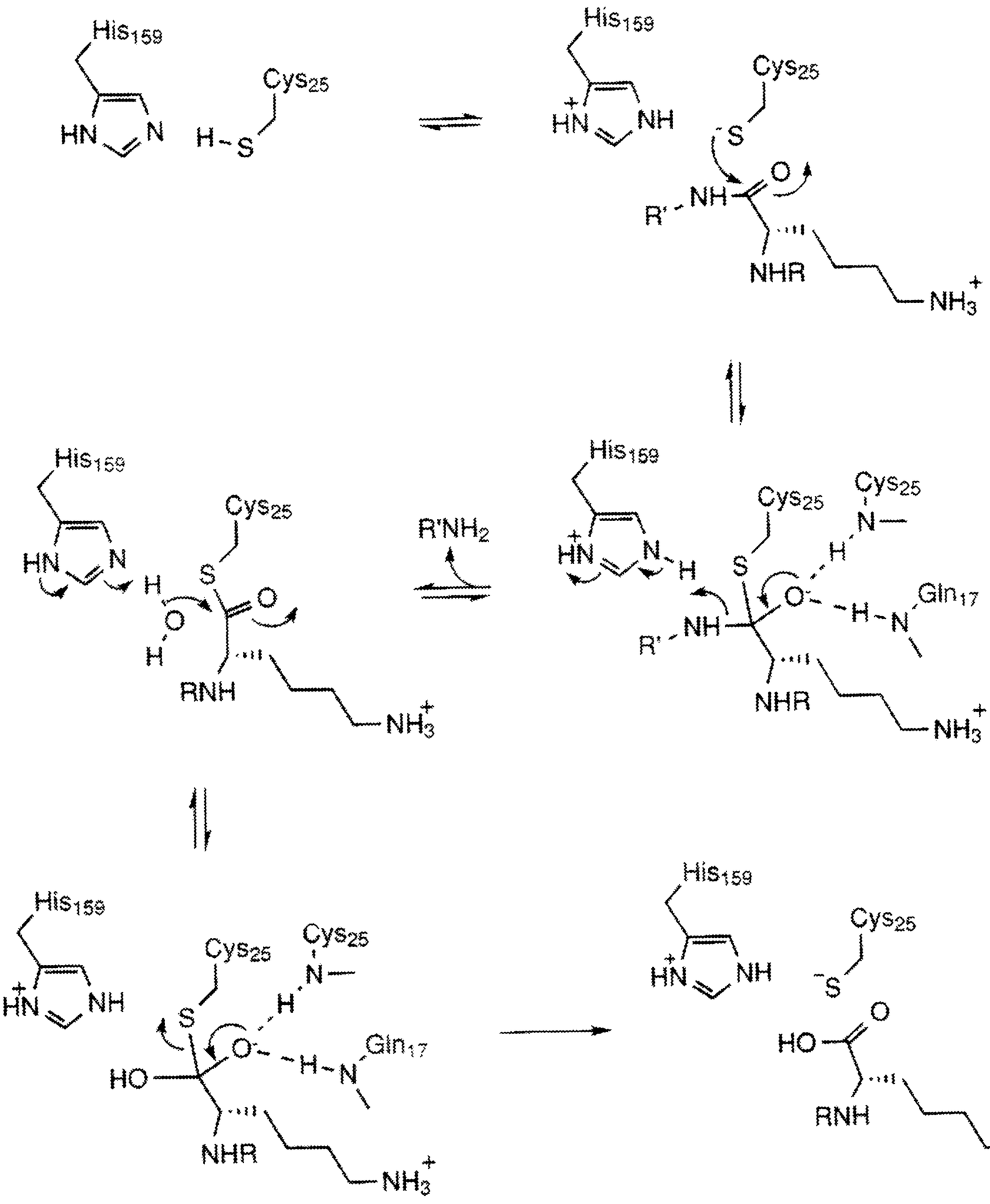


Figure 7.1 Mechanism of action of cysteine proteases. Schematic representation of the various steps and putative intermediates and transition states in the reaction pathway for the hydrolysis of an amide substrate by a cysteine protease. Details are given in the text. From Storer & Menard (1994).

In this study, the role of Slp paralogue 4 as a cysteine protease has been investigated. This protein has been detected in surface extracts of *C. difficile* cells (Chapter 4) and is recognised by patient antibodies (Chapter 6, Pechine *et al.* (2005b)). An attractive hypothesis, and one which other groups have suggested (Savariau-Lacomme *et al.*, 2003), is that Slp paralogue 4 may have proteoytic activity against other proteins encoded by the same gene cluster, and in particular, may play a role in the maturation of the SLPs by cleavage of SlpA. This hypothesis has been tested by evaluating the activity of paralogue 4 on full length SlpA, and the non-specific proteolytic activity of paralogue 4 has been assessed by a number of methods.

7.2 Sequence analysis of Slp paralogue 4

7.2.1 Sequence similarities to other cysteine proteases

A search for sequences related to the cysteine protease domain of Slp paralogue 4 was carried out against the non-redundant database (including non-redundant CDS translations, PDB, SwissProt, PIR and PRF) using the BLASTP search tool available at http://www.ncbi.nlm.nih.gov/BLAST/. Slp paralogue 4 was found to be homologous to a number of cysteine proteases from the papain C1A family, with the most significant match to a cysteine protease from *Syntrophobacter fumaroxidans* (E = 1e-33, identity 29%). The Slp paralogue 4 amino acid sequence was aligned with the sequences of other proteases from the C1A family using ClustalW, available at http://www.ebi.ac.uk/clustalw/ (Figure 7.2). The Slp paralogue 4 sequence was aligned to these sequences because they are from the papain family and have structural data available, which will aid the structural modelling of Slp paralogue 4. Slp paralogue 4 has strong homology over a large proportion of its sequence. In particular, the regions around the catalytic cysteine and histidine residues are very well conserved, and closely match the consensus sequences for a C1A cysteine protease (indicated in red in Figure 7.2). In the Slp paralogue 4 sequence, the asparagine residue that is thought to stabilise the thiolate-imidazolium ion pair has been substituted for a serine residue, although the rest of the motif is well conserved. This asparagine to serine substitution is unlikely to disrupt the function of the residue, as it will retain a polar nature and the correct spatial orientation. One motif that is present in the other cysteine protease sequences, but absent from Slp paralogue 4, is the ERFNIN sequence, which represents the proregion of the enzyme. Instead of the ERFNIN sequence, a signal sequence is located immediately before the protease domain (not shown on Figure 7.2). The absence of a proregion in Slp paralogue 4 may affect its folding and activation.

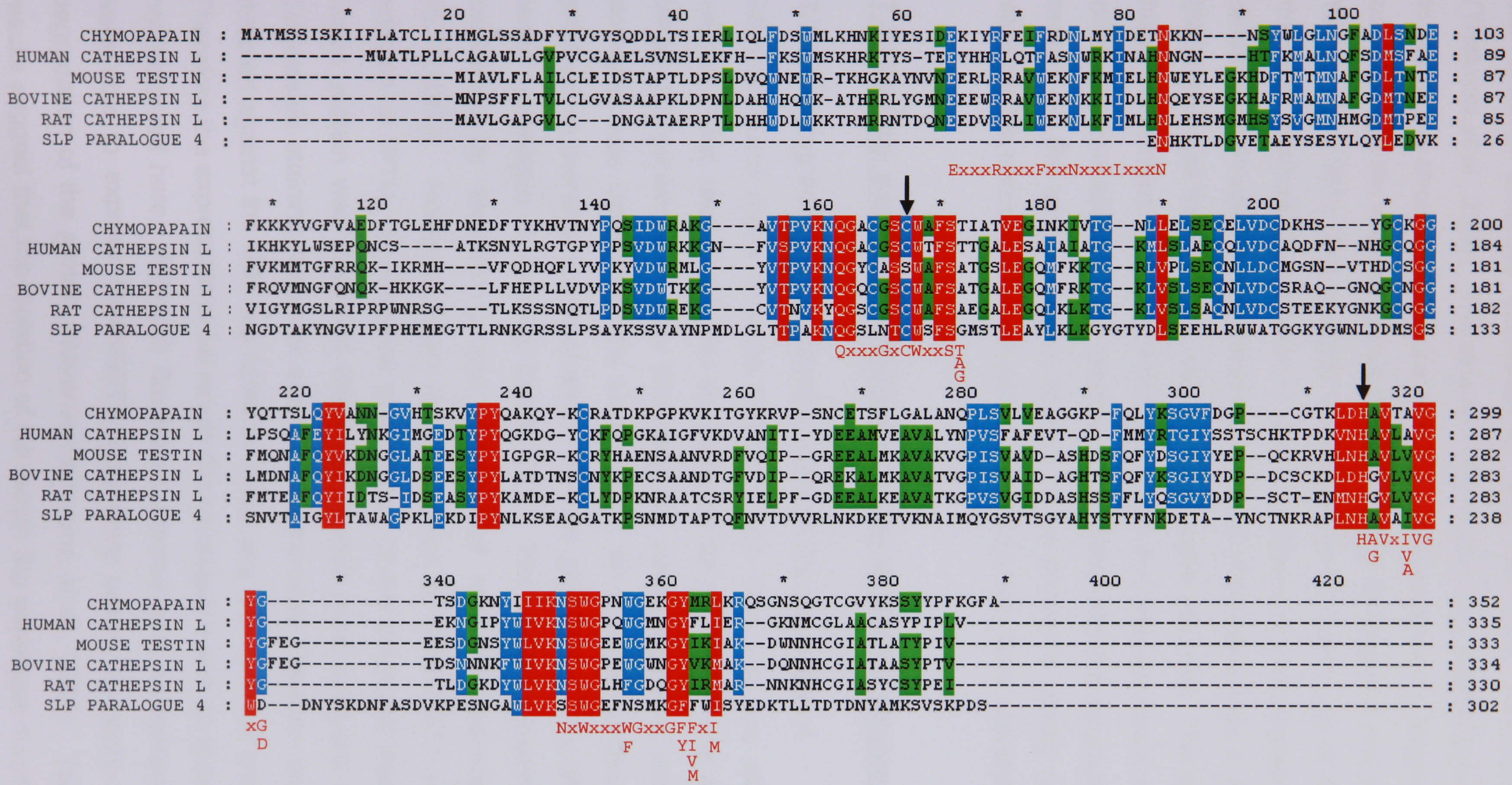


Figure 7.2 Sequence alignment of the protease region of Slp paralogue 4 with other cysteine proteases. Slp paralogue 4 is at the bottom of the alignment, and is shown aligned to the sequences for chymopapain, human cathepsin L mouse testin, bovine cathepsin L and rat cathepsin L. Conserved residues are highlighted in red, 80% similarity is indicated in green and 60% in blue. The consensus sequence for four domains of a cysteine protease is shown in red text. Black arrows indicate the catalytic cysteine and histidine residues.

7.2.2 Modelling the Slp paralogue 4 structure

The amino acid sequence of the protease domain of Slp paralogue 4 was searched against the PDB database using BLASTP. The best match was to a 1.9 Å structure of the cysteine protease ervatamin C, from the plant *Tabernaemontana divaricata* (Thakurta *et al.*, 2004). The structure the ervatamin C homologue was used as a basis to model the structure of Slp paralogue 4. A local dynamic programme alignment was used to align the residue equivalencies between paralogue 4 and ervatamin C (Smith & Waterman, 1981). This mapped the respective equivalencies of Slp paralogue 4 onto the PDB structure of ervatamin C, and was viewed in the programme Rasmol. Three-dimensional images of the predicted structure of Slp paralogue 4 (Figure 7.3) show that the protein has two domains with a central active site cleft. The catalytic cysteine and histidine residues are in the centre of the cleft and also situated in close proximity is the putatively interacting serine residue. Consistent with the structures of other cysteine proteases, the amino-terminal domain of the protein is mostly composed of α-helix, whereas the carboxyl-terminal part has a β-sheet structure with short region of α-helix.

7.3 Cloning, expression and purification of Slp paralogue 4

7.3.1 Cloning and expression of full length Slp paralogue 4

Attempts were made to clone the full length ORF into two alternative expression vectors. First, taking the same approach as Savariau-Lacomme *et al.* (2003), the full-length Slp paralogue 4 gene (including the signal sequence, protease and amidase domains) was cloned into the vector pGEX 6P-1. This construct was designed to produce an in-frame fusion of amino-terminal glutathione *S*-transferase (GST) to full-length paralogue 4, termed Orf4NGST1. The resulting plasmid constructs were confirmed by sequencing and transformed into the expression strain *E. coli* BL21(λ*DE3*). However, when the expression of Orf4NGST1 was induced by the addition of IPTG, no evidence of a protein of the predicted size of the GST-ORF4 fusion protein was observed when cell lysates were analysed by SDS-PAGE and Coomassie staining (data not shown). Furthermore, Western blots using antibodies directed against the GST tag failed to detect any GST-labelled proteins in the lysates. Therefore the expression of full-length Slp paralogue 4 with an N-terminal GST tag was not achieved here. Although Savariau-Lacomme and co-workers reported that they were able to express Orf4NGST1, they were unable to purify it due to the rapid degradation of the protein (Savariau-Lacomme *et al.*, 2003). Taken together, these results suggest that the isolation of full length Slp paralogue 4 with an amino-terminal GST tag may be highly problematic.

A



B

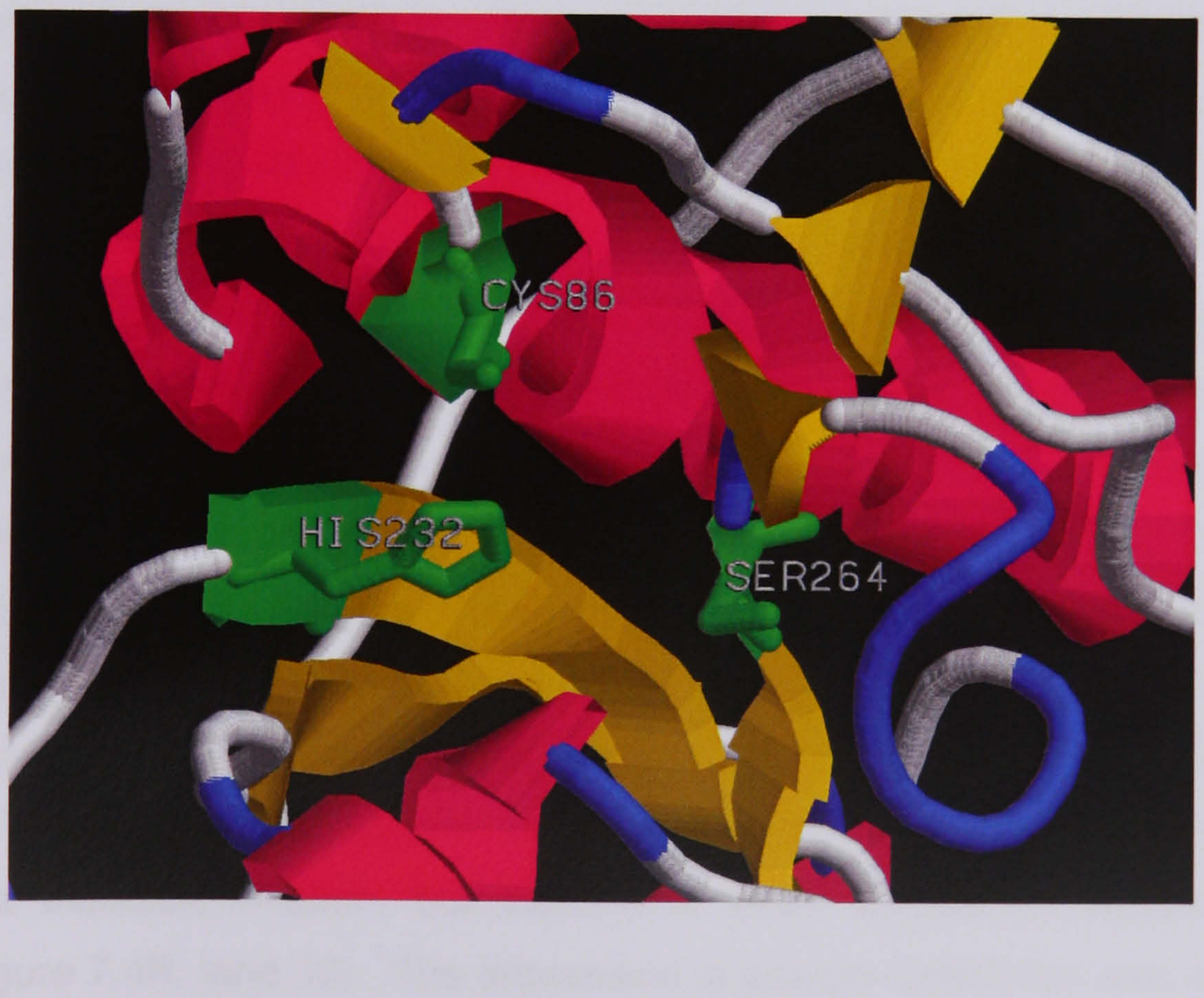


Figure 7.3 Views of the predicted structure of Slp paralogue 4. **A.** View of the predicted tertiary structure, with N-terminus at the top of the image, and the C-terminus at the bottom. **B.** Zoom image of the predicted active site cleft. The active site residues are highlighted in green and labelled. Pink indicates α-helix and yellow indicates β-sheet secondary structure.

The second approach taken to clone the full-length Slp paralogue 4 gene was to use the expression vector pMAL p2-X. This vector was chosen to add an amino-terminal maltose binding-protein (MBP) tag to Slp paralogue 4, and direct the folding of the protein to the periplasm. This has the benefit that proteins that are usually insoluble in the cytoplasm can fold more slowly and in conditions that may be similar to their normal expression, resulting in higher yields of soluble protein. However, despite the successful PCR amplification of Slp paralogue 4 using these primers (data not shown), a construct containing the correct insertion sequence could not be obtained. Therefore the properties of an MBP-tagged Slp paralogue 4 fusion protein could not be investigated any further.

7.3.2 Construct pEC29(7)

Construct pEC29(7) was a gift from Dr. Emanuela Calabi. This plasmid contains the sequence for the protease domain of Slp paralogue 4 cloned into the vector pET28a with a carboxyl-terminal hexa-histidine tag (Orf4CHis1). Due to the difficulties in cloning the full-length Slp paralogue 4 sequence, it was decided to proceed using this recombinant fragment of the protein. The expression host *E. coli* BL21(λ*DE3*) was transformed with construct pEC29(7). The growth of *E. coli* containing pEC29(7) was slightly impaired compared to *E. coli* alone at 37 °C, both before and after expression of Orf4CHis1 was induced by the addition of 1 mM IPTG (data not shown). Expression of a protein of the predicted molecular weight, 42 kDa, can be seen when *E. coli* lysates were analysed by SDS-PAGE (Figure 7.4A, lanes 3, 6 and 9). However, when samples of the *E. coli* induced to express Orf4CHis1 were lysed by sonication and the soluble and insoluble fractions viewed by SDS-PAGE, the vast majority of the recombinant protein was in the insoluble fraction of the cells (Figure 7.4A). To ensure the over-expressed protein was the protein of interest, a Western blot with antibodies against the hexa-histidine tag was performed (Figure 7.4B). As predicted, the histidine-tagged protein can be seen at approximately 40 kDa, consistent with this being Orf4CHis1. The Western blot reveals that the majority of the protein of interest is present in the insoluble fraction, but there is a small amount of soluble protein expressed (Figure 7.4B, lane 10). The expression of soluble Orf4CHis1 was optimised by growing the bacteria under a range of temperatures and inducing with a range of concentrations of IPTG. The conditions that maximised soluble protein expression were found to be induction with 0.1 mM IPTG for 4 hours at 20 °C (data not shown).

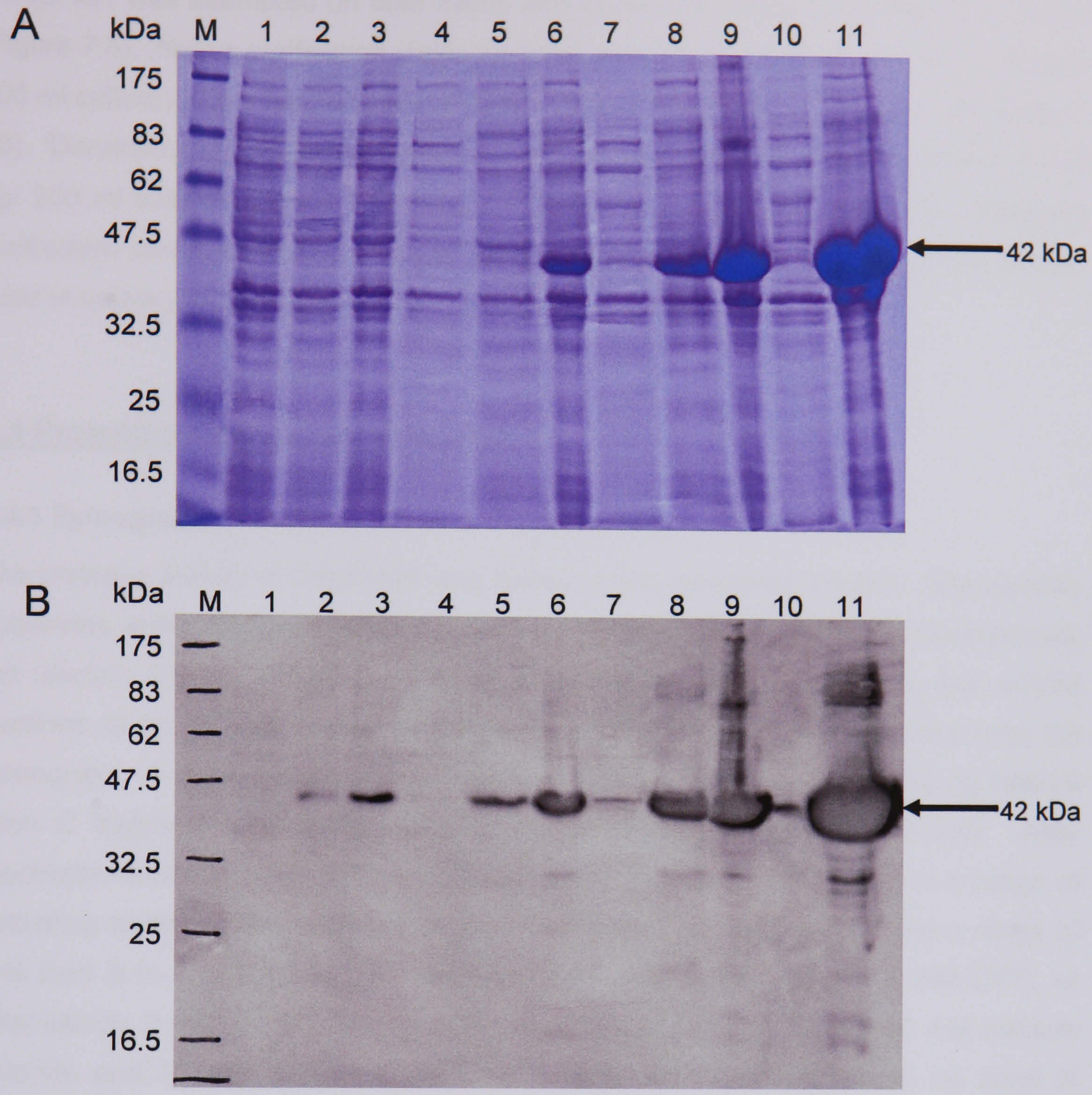


Figure 7.4 SDS-PAGE (A) and Western blot (B) of *E. coli* cell lysates induced to express Orf4CHis1. **A.** SDS-PAGE separated proteins visualised by staining with Coomassie blue. **B.** Western blot of duplicate gel A using antibodies directed against the hexa-histidine tag. Growth was at 37 °C and induction with 1 mM IPTG. Cells were lysed by boiling in SDS sample buffer for 5 min prior to loading onto gels. Molecular mass markers are indicated by M. Lane 1, total lysate of BL21(λ*DE3*); lane 2, total lysate of BL21(λ*DE3*) (pEC29) -uninduced; lane 3, total lysate of BL21(λ*DE3*) (pEC29) induced for 1 h; lane 4, soluble fraction; lane 5, insoluble fraction; lane 6, total lysate of BL21(λ*DE3*) (pEC29) induced for 4 h; lane 7, soluble fraction; lane 8, insoluble fraction; lane 9, total lysate of BL21(λ*DE3*) (pEC29) induced overnight; lane 10, soluble fraction; lane 11, insoluble fraction.

A series of experiments were conducted to purify the Orf4CHis1 protein. Purification of Orf4CHis1 was attempted on both native and denaturing nickel-NTA agarose columns (Figure 7.5). Native purification yielded a small amount of soluble Orf4CHis1 (~100 µg/ 200 ml culture) with a few low-abundance contaminating proteins (Figure 7.5A, lanes 7-10). Denaturing purification resulted in a much higher yield of very pure protein (~160 µg/ 200 ml culture) (Figure 7.5B, lanes 6 and 7). As the protein from the denaturing purification was in a solution containing 6 M urea and would need refolding, the smaller yield of natively folded protein was used in further experiments.

7.4 Proteolytic activity of Orf4CHis1

7.4.1 Zymograms

The protease activity of Orf4CHis1 was tested by the zymogram method. Non-specific substrates (milk powder or gelatine) were incorporated into a 12% SDS-polyacrylamide gel (section 2.5.1). Protein purified on a native nickel-agarose column and soluble fractions of an *E. coli* lysate induced to express Orf4CHis1 were loaded onto the zymogram. A positive control of 1 ng papain and a negative control of 5 ng tetanus toxin C fragment were also loaded onto the zymogram gels (Figure 7.6). After electrophoresis, the SDS was soaked out of the gels before incubation in a range of activating buffers. The activation buffers tested were 20 mM Tris-HCl (at a range of pHs from 6 to 9 at 0.5 intervals), with and without a reducing agent (1 mM DTT), or alternatively 2 mM EDTA, 10 mM cysteine, 100 mM glycine pH 8.3, 10 mM calcium chloride and 10 mM sodium phosphate. Under all conditions tested, no zone of clearing around the Orf4CHis1 protein was visible (representative example in Figure 7.6). A range of incubation times were tested, from 2 h to overnight, with 2 h considered optimal as longer incubation times led to diffusion of the protein bands. Zymograms containing both SDS and native PAGE gels were tested. No proteolytic activity on milk or gelatine for Orf4CHis1 could be determined by the zymogram method.

7.4.2 Casein degradation assay

Following the method of Savariau-Lacomme (Savariau-Lacomme *et al.*, 2003), the ability of *E. coli* BL21(λ*DE3*) expressing recombinant Orf4CHis1 to degrade skimmed milk in the growth medium was tested. Cultures were streaked out onto LB agar plates with the appropriate antibiotic. After overnight incubation at 37 °C, the cultures were replica plated onto LB plates supplemented with 1% milk and the appropriate antibiotic, with or without 1 mM IPTG. After overnight growth, the plates were compared and

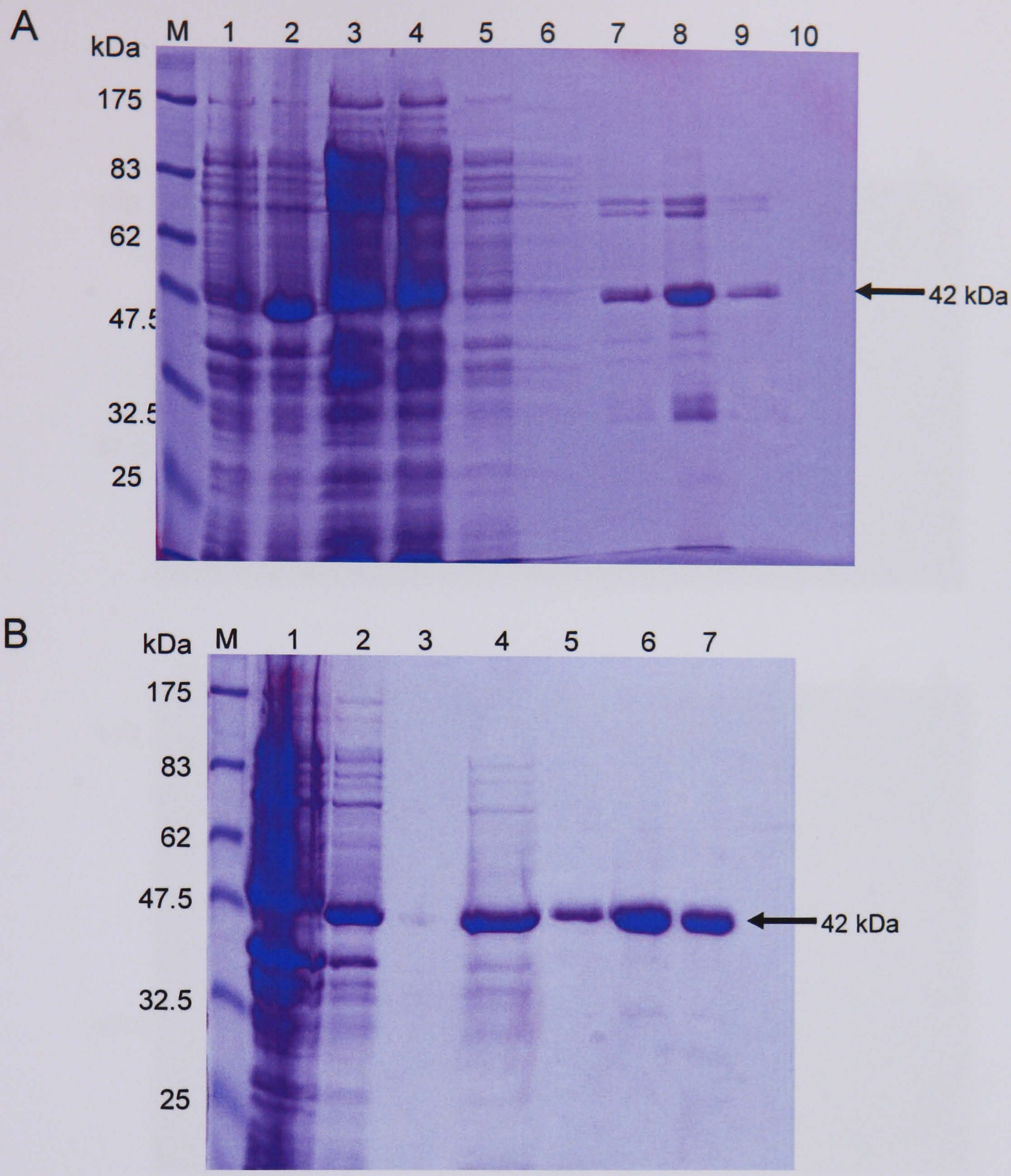


Figure 7.5 Native (A) and denaturing (B) purification of Orf4CHis1 on a batch Nickel-NTA column. Orf4CHis1 was expressed using an optimised procedure (growth at 20 °C, induction with 0.1 mM IPTG). **A.** Native purification of Orf4CHis1 on a nickel-column. Molecular mass markers are indicated by M. Lane 1, total lysate of BL21(λ*DE3*) (pEC29) before induction; 2, total lysate after 4 h induction for the expression of Orf4CHis1; 3, soluble fraction after 4 h induction for the expression of Orf4CHis1; 4, flow through from column; 5, wash 1; 6, wash 2; 7, elution 1; 8, elution 2; 9, elution 3; 10, elution 4. **B.** Denaturing purification of Orf4CHis1 on a nickel column. Lane 1, total lysate of BL21(λ*DE3*) (pEC29) after 4 h induction for the expression of Orf4CHis1; 2, flow through from column; 3, empty lane; 4, wash 1; 5, wash 2; 6, elution D; 7, elution E. Proteins were visualised with Coomassie Blue. Arrow indicates Orf4CHis1.

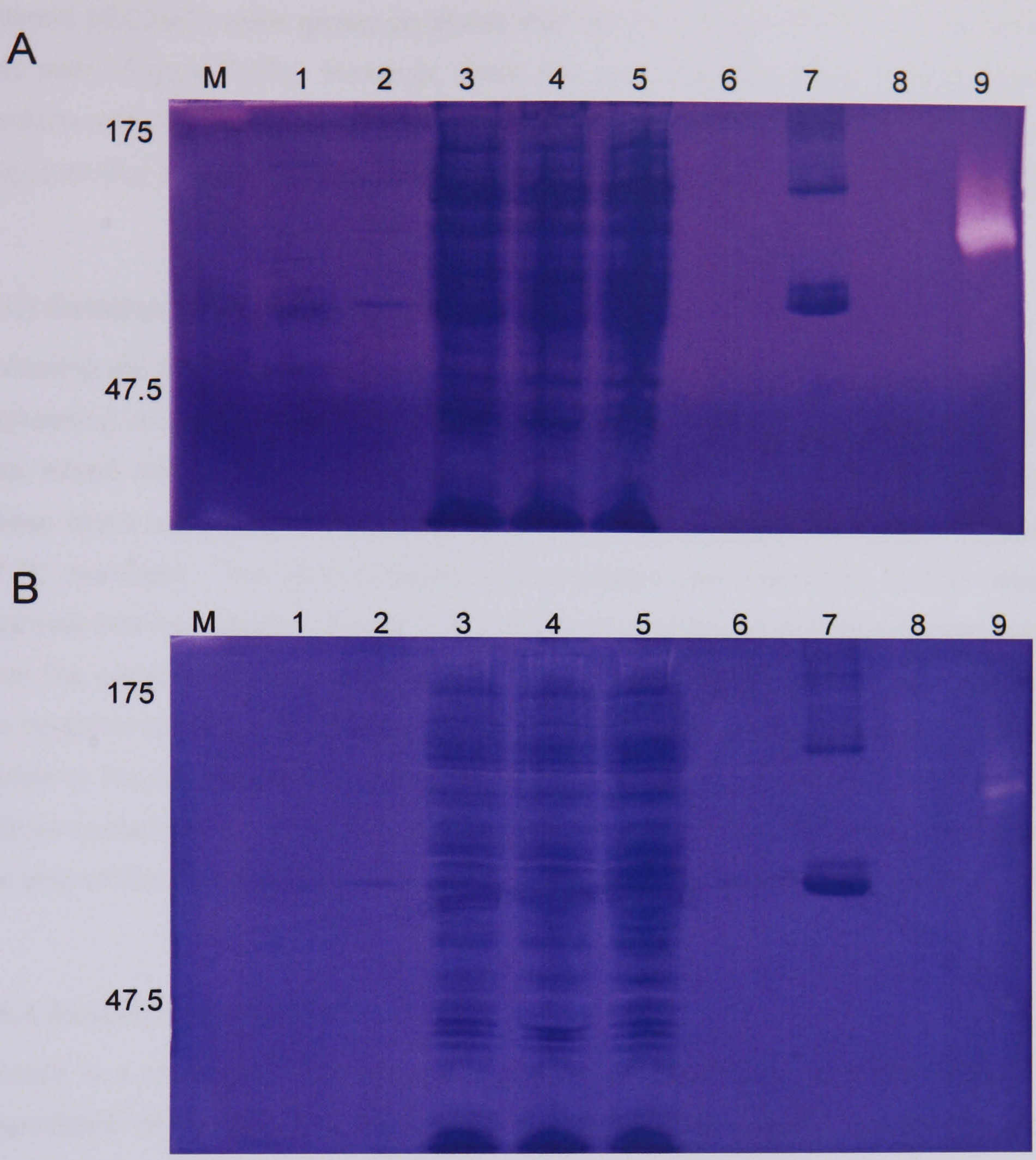


Figure 7.6 Typical zymograms testing for proteolytic activity of Orf4CHis1. A. Zymogram containing gelatine as a substrate. **B.** Zymogram containing skimmed milk powder as a substrate. Molecular mass marker are indicated by M and shown in kDa. Lane 1, *C. difficile* lysate; lane 2, purified Orf4CHis1; lane 3, *E. coli* lysate induced to express Orf4CHis1; lane 4, *E. coli* lysate induced to express Orf4NGST1; lane 5, *E. coli* BL21(λ*DE3*) lysate (no vector); lane 6, empty lane; lane 7, 5 ng tetanus toxin C fragment; lane 8, empty lane; lane 9, 1 ng papain. A zone of clearing can only be seen in lane 9 (papain).

cultures that produced a zone of lysis were deemed to have caseinase activity. A negative control of BL21(λ*DE3*) and a positive control of *Streptococcus pyogenes* (Nomizu *et al.*, 2001) were used (Figure 7.7A and B). When BL21(λ*DE3*) containing plasmid pEC29(7) were grown on plates that did not contain IPTG, no lysis of the milk was seen (Figure 7.7C). However, when the same colonies were replica plated onto medium with IPTG, a faint zone of lysis in the milk could be seen (Figure 7.7D). This suggests that induction of Orf4CHis1 does impart caseinase activity onto BL21(λ*DE3*).

7.4.3 Gelatine degradation activity

Following the methods of Savariau-Lacomme (2003), the activity of *E. coli* BL21(λ*DE3*) expressing recombinant Orf4CHis1 on gelatine was investigated. Activated charcoal was mixed into a 10% gelatine solution and coated onto strips of gel-backing film. These strips were placed into bacterial cultures and incubated in a static incubator at 37 °C overnight. The level of gelatine degradation was monitored by the release of charcoal into the medium (Figure 7.8). Although high levels of charcoal were released from the gelatine by the positive control, *S. pyogenes*, charcoal was also released by the negative control of *E. coli* BL21(λ*DE3*). Furthermore, a charcoal sediment was also visible in the tubes containing only LB medium or water. Therefore, even though the culture containing *E. coli* BL21(λ*DE3*) expressing Orf4CHis1 did release charcoal from the strip of film, it cannot be confirmed that this was due to proteolytic activity.

7.4.4 Azocoll degradation

Azocoll is a chemical that comprises collagen attached to an azo dye group. Upon degradation of the collagen, the azo dye is released into solution, resulting in a colour change that can be measured through the optical density at 520 nm. Clostripain was used as a positive control, and consistently released the azo dye (Table 7.1). However, the negative controls of BL21(λ*DE3*) and BL21(λ*DE3*) containing plasmids pET28a and pGEX 6P-1 also released significant amounts of azo dye, and in similar quantities to cultures expressing recombinant Slp paralogue 4. Therefore this assay did not conclusively measure the degradation of azocoll by Orf4CHis1.

A



B



C



D



Figure 7.7 Caseinase activity determined by growth on agar plates supplemented with skimmed milk. **A.** Negative control of BL21(λ*DE3*) shows no zone of lysis in the milk. **B.** Positive control *S. pyogenes* shows caseinase activity indicated by zone of lysis around colonies. **C.** BL21(λ*DE3*) containing pEC29(7) grown without IPTG. **D.** BL21(λ*DE3*) containing pEC29(7) grown with IPTG and therefore expressing Orf4CHis1. Arrows indicate zone of lysis. These plates are representative of three replicate experiments.

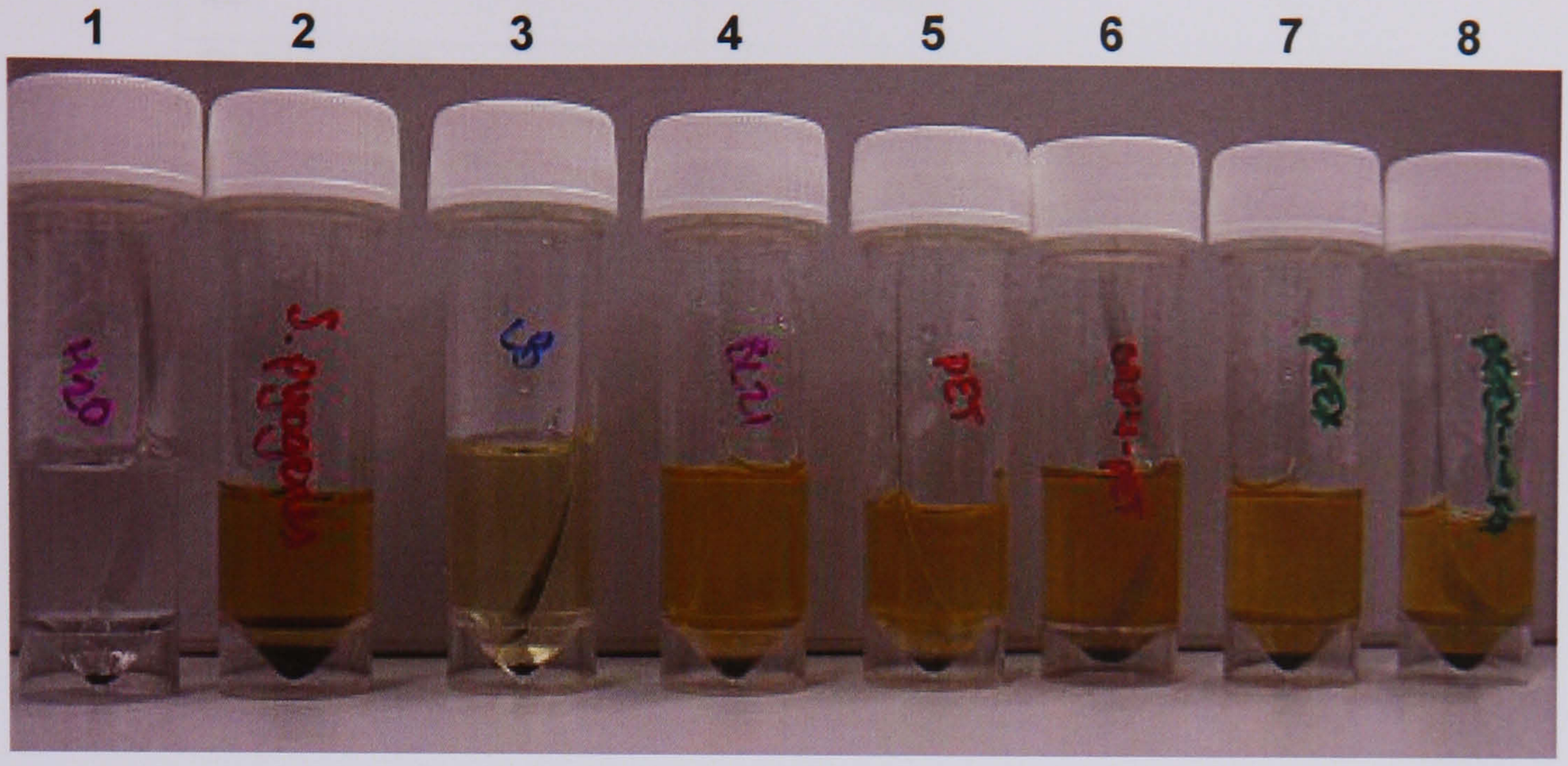


Figure 7.8 Gelatinase activity. Strips of film coated in a suspension of charcoal in 10% gelatine were placed in bacterial cultures incubated overnight. Gelatinase activity is indicated by charcoal sediment in the base of the universal tubes. Tube 1, negative control of water; tube 2, positive control of *S. pyogenes*; tube 3, negative control of LB medium only; tube 4, negative control of *E. coli* BL21(λ*DE3*); tube 5, negative control of *E. coli* BL21(λ*DE3*) containing plasmid pET28a; tube 6, *E. coli* BL21(λ*DE3*) containing plasmid pEC29(7) and 1 mM IPTG to induce expression of Orf4CHis1; tube 7, negative control of *E. coli* BL21(λ*DE3*) containing plasmid pGEX 4T-1; tube 8, *E. coli* BL21(λ*DE3*) containing plasmid pGEXΩ*orf4* and 1 mM IPTG to induce expression of Orf4NGST1. This result is representative of a duplicate experiment. All tubes contain a charcoal sediment and therefore this assay is unsuitable for detecting gelatinase activity.

Sample	Mean OD_{520}	Standard deviation
Water	0.30	0.050
LB	0.34	0.151
BL21(λ*DE3*)	0.49	0.055
BL21(λ*DE3*) pET28a	0.64	0.179
BL21(λ*DE3*) pEC29(7)	0.58	0.129
BL21(λ*DE3*) pGEX 6P-1	0.51	0.150
BL21(λ*DE3*) pGEXΩ*Orf4*	0.46	0.070
Clostripain	2.03	0.313

Table 7.1 Results of the azocoll degradation experiment. Azocoll (10 mg.ml) was mixed 1:1 with overnight cultures of *E. coli* BL21(λ*DE3*), induced to express recombinant protein. Cultures were incubated under aerobic conditions at 37 °C for 6 h with the azocoll. Degradation of collagen releases an azo dye into solution, which is measured by the optical density at 520 nm. These results are the average of three experiments.

7.4.5 BAPNA assay

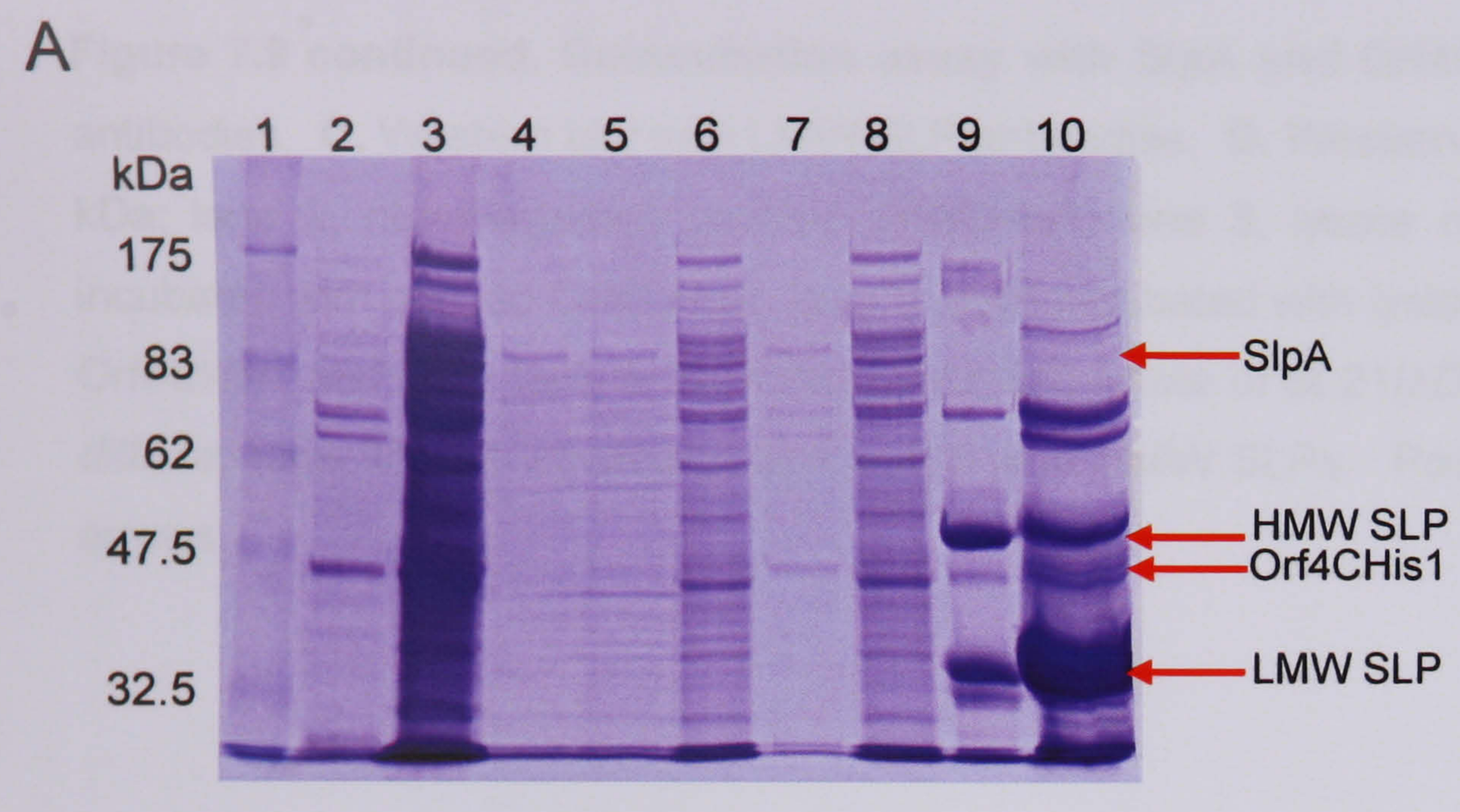
The method employed was that described by Janoir for the measurement of proteolytic activity by the protease domain of Slp paralogue 4 (Janoir *et al.*, 2004). BAPNA is a non-specific colorimetric substrate, and its degradation by proteases can be followed by an increasing optical absorbance at 412 nm. Table 7.2 shows the results of this assay. Surprisingly, papain shows very little proteolytic activity by this method, with a low level of absorbance similar to that of the negative control, tetanus toxin fragment C. The other positive control, clostripain, caused significant degradation of BAPNA, and increased the A_{412} of the solution to 0.247. The nickel-agarose-purified Orf4CHis1 increased the A_{412} of the solution to 0.082, higher than that of the negative control but not to the same extent as clostripain. However, when lysates of BL21(λ*DE3*) that do not express Orf4CHis1 are compared, there is no significant difference between the A_{412}. Therefore this assay is not conclusive in establishing the proteolytic activity of Orf4CHis1.

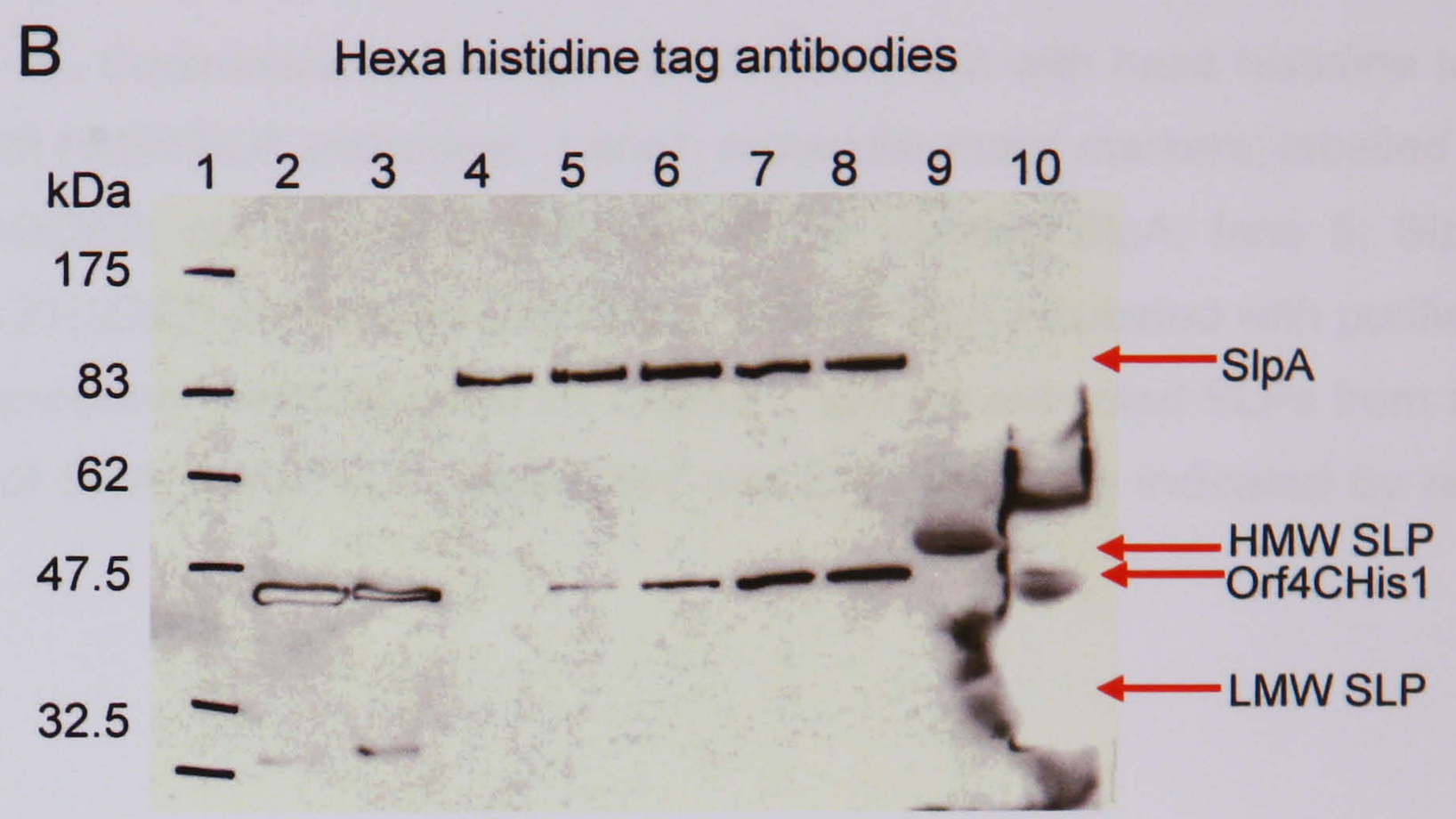
7.4.6 Co-incubation assay with SlpA

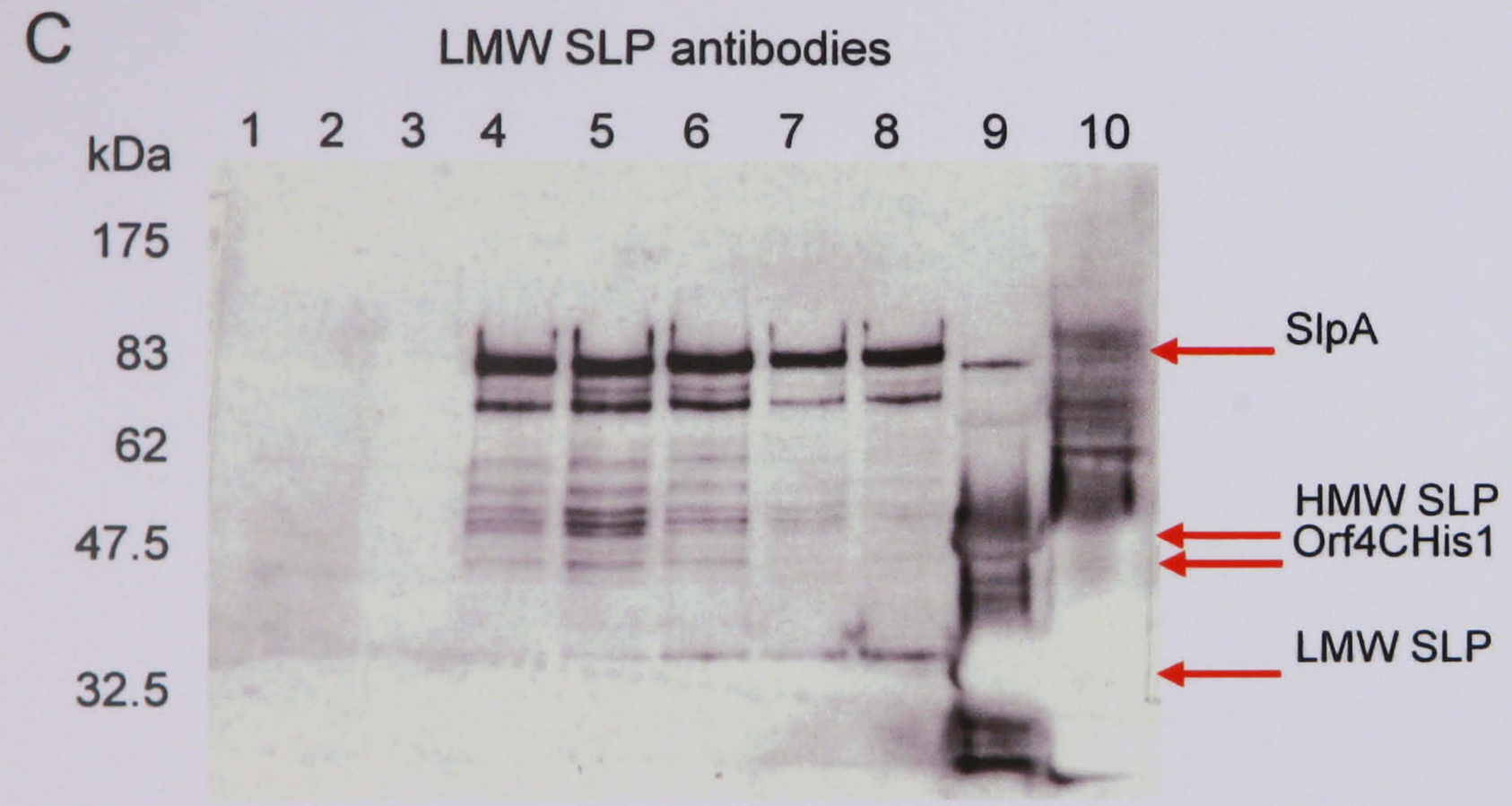
To determine whether SlpA is a substrate of Slp paralogue 4, a coincubation assay was set up. Equimolar concentrations of full length SlpA (produced in *L. lactis*, section 2.3.2) and Orf4CHis1 (nickel-agarose purified and as part of a lysate of BL21(λ*DE3*) induced to express Orf4CHis1) were incubated together for four hours at 37 °C. The incubation was carried out in duplicate, with one of the reactions also containing 1 mM DTT to ensure the putative catalytic cysteine residue was in a reduced state, and therefore able to carry out the reaction. Each component was also incubated alone, so that degradation of the proteins could be monitored. After the incubation was complete, the samples were applied to four 12% SDS-PAGE gels (Figure 7.9). Also applied to the gels were a glycine extract of the native SLPs from *C. difficile* 630, and a mixture of recombinantly expressed and partially purified SLPs (Figure 7.9, lanes 9 and 10 respectively). One gel was stained with Coomassie blue (Figure 7.9A). The remaining three gels were transferred to a nitrocellulose membrane, and Western blots performed. Figure 7.9B shows a Western blot probed with antibodies against the hexa-histidine tag that the recombinant proteins are labelled with. SlpA is visible as a band of approximately 80 kDa (lanes 4-8), and this band does not change in size under any of the conditions tested. Orf4CHis1 can be seen as a smaller band at approximately 42 kDa (lanes 1-7). Due to the large quantities of recombinantly expressed SLPs loaded onto the gel, the Western blot is overexposed and the individual SLPs cannot be clearly seen (lane 10). However, it is apparent from this gel that SlpA has not been

Sample	**A_{412}**
Water	0.007
Papain	0.020
Clostripain	0.247
Tetanus toxin fragment C	0.024
Purified Orf4CHis1	0.082
C. difficile lysate	0.136
E. coli BL21(λ*DE3*)	0.337
BL21(λ*DE3*) pEC29(7) lysate	0.247
BL21(λ*DE3*) pGEXΩ*orf4*	0.260

Table 7.2 BAPNA assay. Purified proteins and cell lysates were incubated with 1.2 mM BAPNA for 30 min at 37 °C. Cell lysates were made by culturing *E. coli* BL21(λ*DE3*) to early exponential phase at 37 °C with shaking at 220 rpm and the expression of recombinant protein was induced by the addition of 1 mM IPTG for 3 hours under the same conditions. Cultures were lysed by sonication. Degradation of BAPNA is measured by an increasing absorbance at 412 nm.







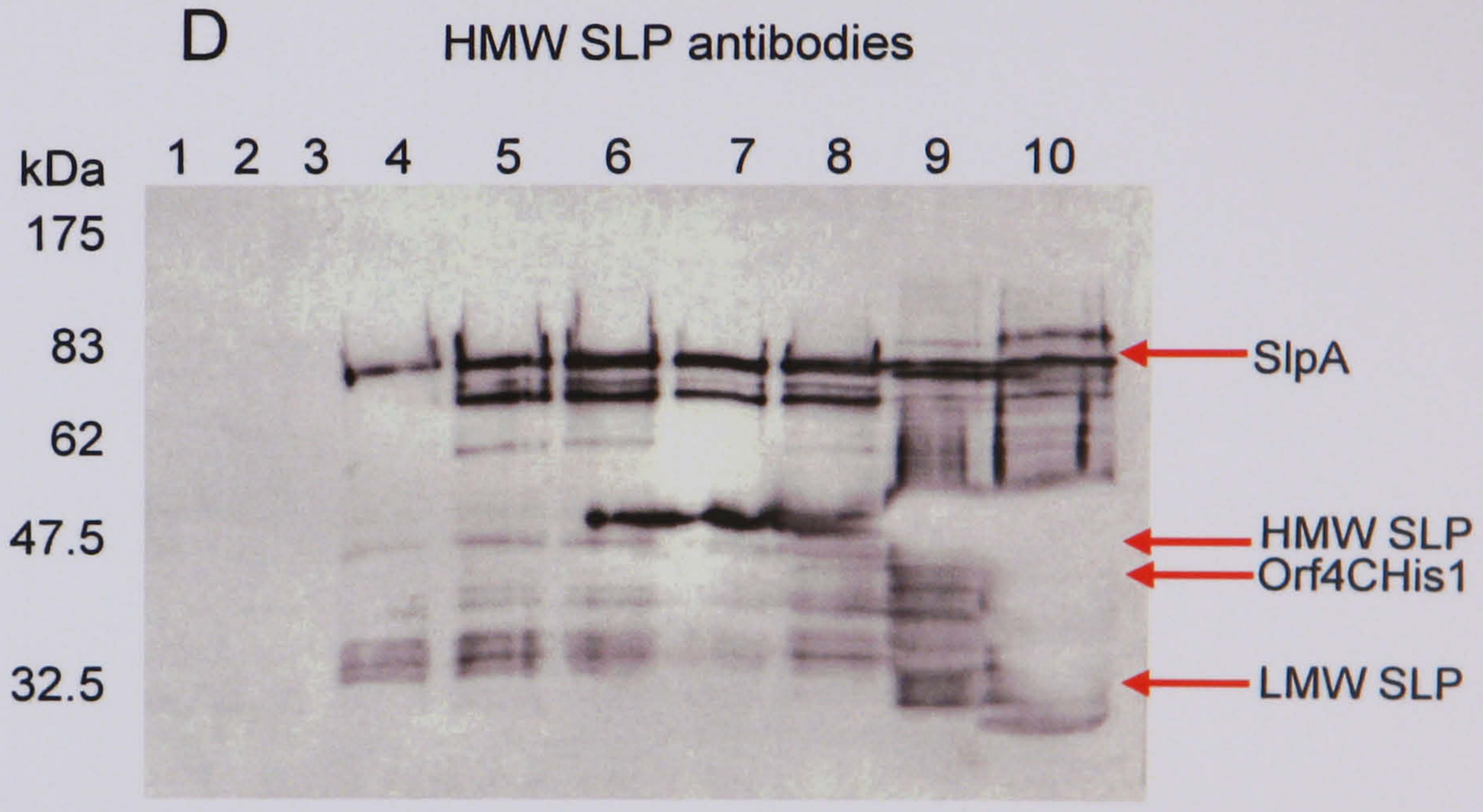


Figure 7.9 Coincubation assay with SlpA and Orf4CHis1. Legend continued overleaf.

Figure 7.9 continued. Coincubation assay with SlpA and Orf4CHis1. **A.** Coomassie stained gel. **B.** Western blot with hexa histidine tag antibodies. **C.** Western blot with LMW SLP antibodies. **D.** Western blot with HMW SLP antibodies. Lane1, molecular mass markers, labelled in kDa; lane 2, nickel-agarose purified Orf4CHis1; lane 3, lysate of BL21(λ*DE3*) expressing Orf4CHis1; lane 4, purified SlpA; lane 5, SlpA incubated with purified Orf4CHis1; lane 6, SlpA incubated with lystae of BL21(λ*DE3*) expressing Orf4CHis1; lane 7, SlpA incubated with purified Orf4CHis1 and DTT; lane 8, SlpA incubated with lysate of BL21(λ*DE3*) expressing Orf4CHis1 and DTT; lane 9, glycine extracted SLPs from *C. difficile*; lane 10, mix of recombinant LMW- and HMW SLPs. Positions of SlpA, HMW SLP, Orf4CHis1 and LMW SLP are indicated by red arrows.

degraded by Orf4CHis1. The Western blots probed with antibodies against the LMW SLP and HMW SLP (Figure 7.9C and D) also show that there is no change in size of the SlpA protein after incubation with Orf4CHis1. Under these conditions, this assay was unable to detect cleavage of SlpA by the putative paralogue 4 protease.

7.5 Discussion

Slp paralogue 4 has been detected on the surface of *C. difficile* by proteomics (Chapter 4). A peptide from the amidase domain of the protein was identified from a lysozyme extract of the surface proteins, confirming the expression of this protein on the surface of the bacterium. In addition, Slp paralogue 4 was recognised by the antisera of three patients (Chapter 6). This confirms that Slp paralogue 4 is expressed during the course of an infection, and is accessible to the host's immune system. Therefore this study aimed to investigate the function of Slp paralogue 4.

Slp paralogue 4 has significant homology to the papain family of cysteine proteases (29% identical), and the amino acid residue sequence of the protease domain shows that it does contain the motifs typical of this family of enzymes. The model predicting the structure of the protease domain of Slp paralogue 4 shows that the protein is likely to contain two domains with a central cleft representing the active site. The predicted catalytic residues of cysteine and histidine, and serine in place of the usual asparagine, all face the centre of the cleft, in a similar fashion to that described for other cysteine proteases (Drenth *et al.*, 1968; Kamphuis *et al.*, 1985).

Slp paralogue 4 proved a very difficult gene to clone. Only one construct, pEC29(7), encoding the protease domain of paralogue 4, resulted in the expression of protein. It is possible that the predicted proteolytic function of the gene product causes it to be toxic to *E. coli.* The damaging nature of Slp paralogue 4 is also consistent with the low levels of soluble protein produced by pEC29(7); most of the protein produced was insoluble and so less likely to be lethal to the *E. coli* cells. Also, maximum amounts of soluble protein were produced when the *E. coli* cells were induced with a reduced amount of IPTG and at low temperature (0.1 mM and 20 °C respectively). Orf4CHis1 was purified under native and denaturing conditions on a nickel-agarose column. Native purification yielded a small amount of soluble protein (~500 μg/l culture), with a small number of contaminating proteins. Denaturing purification resulted in a much higher yield of very pure protein (~800 μg/l culture). A study of Slp paralogue 4 (also known as Cwp84) has recently been published (Savariau-Lacomme *et al.*, 2003). The full-length Slp paralogue 4 gene was cloned into pGEX 6P-1 with an amino-terminal

GST tag, but could not be reproduced in this study. Savariau-Lacomme *et al.* (2003) successfully expressed full length Slp paralogue 4 in *E. coli*, and in common with this study, found that high concentrations of IPTG used to induce expression caused the protein to be insoluble, resulting in the formation of inclusion bodies. In addition, Savariau-Lacomme *et al.* were unable to purify protein Orf4NGST1, finding a number of degradative products even in the presence of cysteine protease inhibitors. This was hypothesised to be due to an autocatalytic activity of the protease. This is a plausible explanation, as cysteine proteases are typically synthesized as pro-proteins (section 7.1.2). The pro-region folds into the active site to prevent the protein from causing cellular damage as it is transported to its site of activity. Once sectreted, the protease auto-cleaves the pro-region to become active. Thus, autocatalytic activity, particularly at the termini, is common in this group of proteins. The difficulties in purifying extra-cellular cysteine proteases due to autocatalysis has been described elsewhere, for example the SpeB protein of *Streptococcus pyogenes* (Nomizu *et al.*, 2001).

The specificity of cysteine proteases varies considerably between individual proteins (Rawlings & Barrett, 1994). Because the specificity and the substrate of Slp paralogue 4 are unknown, a number of general assays using a non-specific substrate were investigated. First, the zymogram technique was used. This technique has successfully shown the activity of many cysteine proteases (Lantz & Ciborowski, 1994; Poppelmann *et al.*, 2002). The zymogram technique was used to try and detect non-specific protease activity of Orf4CHis1, which could be extended to more specific substrates if preliminary evidence of activity was found. However, no protease activity could be found for Orf4CHis1 under the conditions described. It is possible that the substrates used were not suitable to show activity of Orf4CHis1, or the reaction buffers were not appropriate, or that the protein failed to refold correctly after electrophoresis. Therefore a number of further assays designed to detect non-specific proteolytic activity were used.

The studies by Savariau-Lacomme (2003) and Janoir (2004) found that some non-specific protease activity could be found in the *E. coli* clone induced to express the full-length and the protease domain of the Slp paralogue 4 protein. Weak degradative activity on azocoll (positive in the Savariau-Lacomme study, negative in the Janoir study) gelatine, skimmed milk and BAPNA was found. These assays were repeated here, using protein Orf4CHis1, and *E. coli* clones induced to express Orf4CHis1. The gelatine assay did not detect proteolytic activity under the conditions used here, as all samples had a positive result including the negative controls. The assays detecting activity on skimmed milk and azocoll did not show any significant activity specific to the

presence of Orf4CHis1, and the BAPNA degradation assay showed weak, borderline activity for the purified protein only, and not for clones induced to express the protein. In conclusion, these assays did not show clear evidence that Orf4CHis1 has non-specific proteolytic activity.

Due to the negative results obtained for the detection of non-specific proteolytic activity of Slp paralogue 4, an assay was developed using the hypothesised substrate, SlpA. Both SlpA and Slp paralogue 4 are encoded on the same gene cluster, and both are predicted to be exported from the cell as they contain a signal sequence. It is known that SlpA is post-translationally cleaved to create the mature LMW- and HMW SLPs (Calabi, 2001), but the method by which this occurs has not been elucidated. Therefore, due to its similarity to cysteine proteases, Slp paralogue 4 was considered a possible candidate to carry out this cleavage reaction. To study this, purified recombinant SlpA from *Lactococcus lactis* was used as a substrate in a simple coincubation assay with partially purified Orf4CHis1 and a lysate of *E. coli* BL21(λ*DE3*) induced to express Orf4CHis1. The assay was carried out at 37 °C for 4 hours with and without the presence of a reducing agent to optimise the reaction conditions. A sensitive and unambiguous method to detect cleavage of SlpA by Orf4CHis1 was through a Western blot using antibodies directed against the hexa-histidine tag that both SlpA and Orf4CHis1 were labelled with. This assay did not show any degradation of SlpA by Orf4CHis1 under the conditions used here, and has not shown any proteolytic activity of Slp paralogue 4.

To summarise, this study has shown that Slp paralogue 4 contains many of the structural features of a cysteine protease, but has not shown significant proteolytic activity with the numerous assays attempted here. The proteomic analysis of the cell surface of *C. difficile* (Chapter 4) has shown that Slp paralogue 4 is expressed on the surface of the bacterium, and it is immunoreactive (Chapter 6), but its function in this location remains unclear.

Ideally, the function of Slp paralogue 4 as a protease could be investigated in *C. difficile*. The methods to accomplish the construction of targeted knockout mutations are at a very early stage of development, and have been hampered by a host restriction/modification barrier, a lack of suitable plasmids and difficulties with electroporating *C. difficile*. Until genetic tools become freely available to study *C. difficile*, the role of numerous potential virulence factors, including the Slp paralogues and paralogue 4 in particular may not be fully understood.

Chapter 8

Discussion

8.1 Introduction

Clostridium difficile is a pathogen that is becoming increasingly problematical in the nosocomial setting. The incidence of *C. difficile* infections is greatly increasing throughout the developed world (Department of Health, 2006), and the recent emergence of a hyper-virulent strain in Canada is increasing incidence and mortality rates (Eggertson, 2005a). *C. difficile* infections are usually treated by withdrawing the inciting antibiotics, followed by administration of metronidazole or vancomycin, but relapsing infections are common. To date, there is not a vaccine available to prevent *C. diffiicle* disease.

Currently, most attention on the mechanisms of virulence of *C. difficile* has focused on the two toxins, TcdA and TcdB, which are the main effectors of tissue damage. However, toxigenic strains can be carried asymptomatically and display a range of virulence (Borriello *et al.*, 1987), so other factors may also be important. Relatively little is known about how the bacterium colonises its host and initiates infection, although it is almost certainly a multifactorial process. Several adhesins have been described, including the HMW SLP (Calabi *et al.*, 2002), the Slp paralogue Cwp66 (Waligora, 2001) and the chaperone GroEL (Hennequin *et al.*, 2001b). However the lack of genetic tools available for *C. difficile* makes it difficult to establish the relative importance of these virulence factors, and more may exist.

The sequencing and annotation of the *C. difficile* genome has recently been completed (Sebaihia *et al.*, 2006). The availability of genome sequence information will revolutionise future studies of *C. difficile*. It will enable a global approach to the identification of virulence factors, antimicrobial and vaccine candidates, both *in silico* and also using functional genomics, microarray and proteomic technologies.

The availability of genomic information for other pathogens has been utilised in vaccine development. For example, the term "reverse vaccinology" has been coined to describe the *in silico* selection of vaccine candidates using bioinfomatics. This technique has been applied to a multitude of pathogens, including *Neisseria meningitidis* (Pizza *et al.*, 2000) and *Streptococcus pneumoniae* (Wizemann *et al.*, 2001), and some of the antigens identified by these studies are now being investigated in clinical trials as vaccines. Nevertheless, it is often necessary to validate the results of *in silico* selection strategies through functional genomic studies, to ensure that candidate proteins are expressed *in vivo*, and that an immune response against them can be protective against disease. This has been achieved by a proteomics-based approach for the pathogens *Bacillus anthracis* (Ariel *et al.*, 2003; Chitlaru *et al.*, 2004)

and *Helicobacter pylori* (Bumann *et al.*, 2004). These studies have mapped the surface proteomes, and then probed the bacterial extracts with antibodies generated in animals against the pathogens. This strategy proved very successful in *H. pylori*, identifying 15 vaccine candidates, at least eight of which have been shown to be protective (Bumann *et al.*, 2004).

8.2 Complementary approaches to identify *C. difficile* virulence factors and vaccine candidates

In this study, a number of complementary approaches have been taken to survey the range of proteins presented on the surface of *C. difficile*, with the aim of identifying potential vaccine candidates and virulence factors.

First, a bioinformatics approach was taken to analyse the *C. difficile* genome sequence to identify genes that contained traits typical of virulence factors (Chapter 3). After selecting gene products that were predicted to be sorted to the cell surface or have similarity to known virulence factors, a reductive strategy was employed to remove genes that were unlikely to have a role in virulence. Genes predicted to encode proteins with "housekeeping" functions, or with strong similarity to other genes in non-pathogenic organisms were discounted, as well as very short coding sequences (CDS). This analysis led to the identification of 364 CDSs with a possible role in virulence. Of these, 196 have an unknown function. The remaining 168 CDSs have putatively assigned functions, which were grouped as autolysins, flagella and chemotaxis proteins, pili, drug resistance, lantibiotics, oligopeptide transporters, proteases, toxins, sortase and sortase substrates, iron scavenging, quorum sensing and miscellaneous. This is the first time that reverse vaccinology has been applied to a clostridial species, but this information now needs to be validated by further analyses, to confirm the expression, localisation and immunogenicity of the identified CDSs.

The repertoire of proteins expressed on the surface of *C. difficile* was mapped using proteomics (Chapters 4 and 5, (Wright *et al.*, 2005)). First, a number of methods were developed to remove proteins from the surface of the cells, and these were assessed by SDS-PAGE. Two methods were chosen for further analysis by proteomics: glycine treatment, which released large quantities of the surface layer proteins (SLPs) and lysozyme treatment, which released a greater range of proteins, although the level of cll lysis caused by these methods could not be determined. The surface extracts were separated by 2D PAGE and protein spots identified by MS/MS. Seventy-one different proteins were detected in the surface extracts, and the expression of some of these

proteins has not previously been described. Many of these proteins were predicted to be exported from the cell by motifs present in their DNA, including proteins such as the SLPs and certain Slp paralogues, components of ABC transporters and components of flagella. However, many of the proteins detected were not predicted to have a cell surface location, including some metabolic and biosynthetic enzymes, and their function in this location is unclear. Studies of the surfaces of other bacterial pathogens have reported similar findings, and have even detected some of the same proteins as this study, for example enolase, DnaK and elongation factor Tu on the surface of *Listeria monocytogenes* (Schaumburg *et al.*, 2004) and glyceraldehyde 3-phosphate dehydrogenase on the surface of *Streptococcus agalactiae* (Hughes *et al.*, 2002). Interestingly, 18 of the proteins identified in cell surface extracts by proteomics were also highlighted by the bioinformatic analysis of the genome. These proteins are flagellin, the flagellar hook protein, the two SLPs, Slp paralogues 2, 4, 6, 12, 24 and 25, OppA and AppA oligopeptide transporter substrate binding proteins, a putative lantibiotic transporter permease, a putative iron ABC transporter solute binding protein and two putative exported proteins with no predicted function (CD1156 and CD1575).

The immunoproteome of *C. difficile* has been mapped (Chapter 6). Proteins extracted from the surface of *C. difficile* were separated by 2D PAGE and probed with antisera. Antiserum was produced in rabbits inoculated with whole, killed *C. difficile* cells. Nineteen immunoreactive proteins were identified from the 2D gels using this serum. Lysozyme extracted surface proteins from *C. difficile* strain M9 were probed using antisera from six patients that had been infected with this strain. Each patient antiserum recognised a characteristic set of proteins. There did not appear to be any significant difference between the antigens recognised by patients that recovered from disease and patients that suffered several relapsing infections. A total of 42 immunoreactive proteins were recognised by the patient antisera, including seventeen antigens that were recognised by the antiserum of more than one patient.

Comparison of the bioinformatic, proteomic and immunoproteome results show that nine proteins are common to all of the analyses. These are flagellin, AppA oligopeptide ABC transporter, the HMW- and LMW SLPs and Slp paralogues 2, 4, 5, 6 and 15. These proteins are predicted by reverse vaccinology to be good vaccine candidates, and proteomics has confirmed their expression, surface localisation and immunogenicity. By using a polyphasic approach, the short-list of vaccine candidates identified here is strengthened. Following the parameters set by Bumann (Bumann *et al.*, 2004), these proteins fulfil all the requirements for good vaccine candidates. In addition, vaccine candidates need to be conserved between different strains of a

species. Several of these proteins have already been studied in different *C. difficile* strains, so some of this information is available. Flagellin shows little phenotypic or genotypic variation between strains (Tasteyre *et al.*, 2000) and is commonly recognised by the antisera of CDAD patients (Pechine *et al.*, 2005b). The LMW SLP shows significant protein sequence divergence between *C. difficile* strains (Calabi & Fairweather, 2002), and therefore may not be suitable for vaccination studies. However the HMW SLP is a highly conserved protein (Calabi & Fairweather, 2002), and in addition this conserved amidase-like region is replicated in the other Slp paralogues. Other than the conserved amidase domain, the sequence conservation between the other paralogues in different strains is largely unknown. An exception is Slp paralogue 2, which is absent in strain 167 (Calabi & Fairweather, 2002), but is 98% identical in the sequenced strain 630 and the strain causing the outbreak in Canada, and also similar to the strains currently causing outbreaks in the UK and USA (Dr. Robert Fagan, personal communication). Slp paralogue 4 is also well conserved between strains at its amino-terminal (Pechine *et al.*, 2005b). Further investigations are required to fully determine the conservation of the remaining Slp paralogues and AppA in different strains of *C. difficile* before vaccination studies begin. In addition, a wider survey of the antibody responses of patients would improve the significance of this data, and separately probing for IgG, IgA and IgM antibodies would help to discern the type of antibody response mounted in patients.

A logical progression of this work would be to clone and express these nine proteins in *E. coli* and determine if they are protective against infection in the hamster model of *C. difficile* disease. A similar experiment has already attempted the passive immunisation of hamsters using antibodies to both SLPs. Although post-challenge survival was extended in the treated hamsters, this passive immunisation did not prevent death (O'Brien *et al.*, 2005). As vaccination using *C. difficile* toxins is also not always protective against disease (Giannasca & Warny, 2004), it is likely that a vaccine against *C. difficile* will need to contain several antigens to fully protect against disease.

Ideally, the importance of these eight proteins in the virulence of *C. difficile* should also be investigated. Until suitable genetic tools become widely available for making mutants of *C. difficile*, this is difficult to determine. An alternative approach could be to clone and express recombinant proteins, and use them to raise antisera. The recombinant proteins and antisera could then be used to block attachment of *C. difficile* cells to human epithelial cells grown in tissue culture, or sections of human gut tissue. This type of method has already been tested on flagellin and the SLPs (Calabi *et al.*, 2002; Tasteyre, 2001), and has shown that flagellin and the HMW SLP are adhesins.

8.3 Responses of *C. difficile* to stress

The effects of growth, antibiotics and heat-shock on the cell surface sub-proteome of *C. difficile* 630 were analysed. Several differentially expressed proteins were identified for each condition, but due to the relatively small number of proteins identified and their diversity, it is difficult to see any broad trends, either for individual conditions or in the experiment as a whole. Two proteins, the HMW SLP and Slp paralogue 2 were detected in lower quantities after the bacteria had been stressed in some way. The three conditions that caused this response were progression into stationary phase growth, the presence of antibiotics in the medium and heat shock at 60 °C. It is not clear why the expression of these two proteins is stress sensitive. Possible reasons could be that they are shed as the bacteria prepare to sporulate, or their expression is down regulated to conserve energy, or that shedding the S-layer may aid in the release of the toxins. The expression of toxins is known to be induced by catabolite repression (Dupuy & Sonenshein, 1998), and begins as the bacteria enter stationary phase, but the mechanism by which these large protein toxins are excreted is unknown. In addition, it is not known whether toxin expression is part of a stress response and could also be triggered by other stressful conditions such as the presence of antibiotics or extremes of temperature. Under none of the conditions tested was the adhesin Cwp66 detected, even though the expression of this protein is reported to be induced by heat-shock (Waligora, 2001).

These results could be improved by repeating this proteomic analysis, but with several modifications. Analysing the cell surface extracts on several 2D zoom gels with overlapping pH ranges, and using larger format gels, would help to improve the resolution of the protein spots, thus making the Phoretix analysis more accurate and the excision of protein spots easier. A major problem in proteomics is reproducibility, and this problem was exacerbated in these experiments due to the difficulties in growing cultures of *C. difficile*. Protein extractions were performed in triplicate, to minimise sample-to-sample variation in the experiment, and to ensure that differences between conditions were due to biological variation and not sample variation. However, it is notoriously difficult to reproduce 2D electrophoresis, and identical samples can appear different on identical gels run in the same conditions. To eliminate these gel-to-gel variations, a differential in gel electrophoresis (DIGE) procedure could be undertaken. This involves using three Cy dyes to stain combinations of each protein extract, before separating the proteins simultaneously on the same gel. This technique avoids the problem of gel-to-gel variation and makes it very easy to visualise changes in protein spots.

8.4 Functional studies of Slp paralogue 4

The proteomics and bioinfomatics analyses performed here have identified a number of proteins that may be important for *C. difficile* to cause disease. However, many of these proteins have not previously been characterised, and their functions are uncertain. In order to fully understand the implications of a proteomic analysis, it is necessary to further investigate each protein identified.

Slp paralogue 4 was identified by proteomics in extracts of the cell surface made by lysozyme treatment (Chapter 4). Slp paralogue 4 was also recognised by the antisera of three *C. difficile* patients, confirming its immunogenicity and expression *in vivo* (Chapter 7). Therefore the properties of Slp paralogue 4 were investigated.

Like the other Slp paralogues, Slp paralogue 4 is composed of an amidase-like domain and a second variable domain. This variable domain shows homology to cysteine proteases, and analysis of its sequence showed that the two residues essential for catalytic activity are conserved. The cloning, expression and purification of Slp paralogue 4 proved to be problematical, and only small quantities of the cysteine protease domain with a hexa-histidine tag could be obtained. Several experiments designed to show proteolytic activity of the recombinant protein on non-specific substrates were attempted, but no significant proteolytic activity was found. Furthermore, no proteolytic activity could be detected using the hypothesised substrate of Slp paralogue 4, full length SlpA. Other studies of this protein have found that it has weak proteolytic properties on non-specific substrates (Janoir *et al.*, 2004; Savariau-Lacomme *et al.*, 2003), but a precise function remains to be elucidated.

8.5 Conclusions

This study represents one of the first post-genomic analyses of *C. difficile*. The complementary analyses of bioinformatics and proteomics have comprehensively mapped the range of proteins present at the *C. difficile* cell surface. The immunoproteome has been defined using rabbit and human patient antisera, and this has confirmed the *in vivo* expression of many proteins for the first time. Nine proteins were common to all the analyses performed, and represent a strong starting-point for further investigations into the mechanisms of virulence, and development of a vaccine against *C. difficile*.

References

Aboudola, S., Kotloff, K.L., Kyne, L., Warny, M., Kelly, E.C., Sougioultzis, S., Giannasca, P.J., Monath, T.P., and Kelly, C.P. (2003) *Clostridium difficile* vaccine and serum immunoglobulin G antibody response to toxin A. *Infect Immun* **71**: 1608-1610.

Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**: 198-207.

Akita, M., Sasaki, S., Matsuyama, S., and Mizushima, S. (1990) SecA interacts with secretory proteins by recognizing the positive charge at the amino terminus of the signal peptide in *Escherichia coli*. *J Biol Chem* **265**: 8164-8169.

Alami, M., Luke, I., Deitermann, S., Eisner, G., Koch, H.G., Brunner, J., and Muller, M. (2003) Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. *Mol Cell* **12**: 937-946.

Alder, N.N., and Theg, S.M. (2003) Energetics of protein transport across biological membranes. A study of the thylakoid DeltapH-dependent/cpTat pathway. *Cell* **112**: 231-242.

Al-Eidan, McElnay, Scott, and Kearney (2000) *Clostridium difficile*-associated diarrhoea in hospitalised patients. *J Clin Pharm Ther* **25**: 101-109.

Alfa, M.J., Kabani, A., Lyerly, D.M., Moncrief, S., Neville, L.M., Al-Barrak, A., Harding, G.K.H., Dyck, B., Olekson, K., and Embil, J.M. (2000) Characterization of a toxin A-negative, toxin B-positive strain of *Clostridium difficile* responsible for a nosocomial outbreak of *Clostridium difficile*-associated diarrhea. *Journal of Clinical Microbiology* **38**: 2706-2714.

Alm, R.A., and Trust, T.J. (1999) Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J Mol Med* **77**: 834-846.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Altschul, S.F., and Gish, W. (1996) Local alignment statistics. *Methods Enzymol* **266**: 460-480.

Archer, G.L., and Bosilevac, J.M. (2000) Signaling antibiotic resistance in Staphylococci. *Science* **291**: 1915-1916.

Ariel, N., Zvi, A., Grosfeld, H., Gat, O., Inbar, Y., Velan, B., Cohen, S., and Shafferman, A. (2002) Search for potential vaccine candidate open reading frames in the *Bacillus anthracis* virulence plasmid pXO1: in silico and in vitro screening. *Infect Immun* **70**: 6817-6827.

Ariel, N., Zvi, A., Makarova, K.S., Chitlaru, T., Elhanany, E., Velan, B., Cohen, S., Friedlander, A.M., and Shafferman, A. (2003) Genome-Based Bioinformatic Selection of Chromosomal *Bacillus anthracis* Putative Vaccine Candidates Coupled with Proteomic Identification of Surface-Associated Antigens. *Infect. Immun.* **71**: 4563-4579.

Aronsson, B., Granstrom, M., Mollby, R., and Nord, C.E. (1985) Serum antibody response to *Clostridium difficile* toxins in patients with *Clostridium difficile* diarrhoea. *Infection* **13**: 97-101.

Bahl, H., Scholz, H., Bayan, N., Chami, M., Leblon, G., Gulik-Krzywicki, T., Shechter, E., Fouet, A., Mesnage, S., and Tosi-Couture, E. (1997) Molecular biology of S-layers. *FEMS Microbiology Reviews* **20**: 47-98.

Bandow, J.E., Brotz, H., Leichert, L.I., Labischinski, H., and Hecker, M. (2003) Proteomic approach to understanding antibiotic action. *Antimicrob Agents Chemother* **47**: 948-955.

Barbut, F., Lalande, V., Burghoffer, B., Thien, H.V., Grimprel, E., and Petit, J.C. (2002) Prevalence and genetic characterization of toxin A variant strains of *Clostridium difficile* among adults and children with diarrhea in France. *J Clin Microbiol* **40**: 2079-2083.

Barbut, F., Delmee, M., Brazier, J.S., Petit, J.C., Poxton, I.R., Rupnik, M., Lalande, V., Schneider, C., Mastrantonio, P., Alonso, R., Kuijper, E., and Tvede, M. (2003) A European survey of diagnostic methods and testing protocols for *Clostridium difficile*. *Clin Microbiol Infect* **9**: 989-996.

Barrett, A.J., and Rawlings, N.D. (2001) Evolutionary lines of cysteine peptidases. *Biological Chemistry* **382**: 727-733.

Bartlett, J.G., Moon, N., Chang, T.W., Taylor, N., and Onderdonk, A.B. (1978) Role of *Clostridium difficile* in antibiotic-associated pseudomembranous colitis. *Gastroenterology* **75**: 778-782.

Bartlett, J.G. (2002) *Clostridium difficile*-associated Enteric Disease. *Curr Infect Dis Rep* **4**: 477-483.

Bartlett, J.G., Chang, T.-W., Gurwith, M., Gorbach, S.L. and Onderdonk, A.B. (1978a) Antibiotic-induced enterocolitis in hamsters: studies with eleven agents and evidence to support the pathogenic role of toxin producing *Clostridia*. *Am. J. Vet. Res.* **39**: 1525-1530.

Bartlett, J.G., Chang, T.-W., Gurwith, M., Gorbach, S.L. and Onderdonk, A.B. (1978b) Antibiotic-associated pseudomembranous colitis due to toxin-producing *Clostridia*. *N. Eng. J. Med.* **298**: 531-534.

Bartlett, J.G., Onderdonk, A.B., Cisneros, R.L. and Kasper, D.L. (1977) Clindamycin-associated colitis due to a toxin-producing species of *Clostridium* in hamsters. *J. Infect. Dis.* **136**: 701-705.

Baverud, V., Gustafsson, A., Franklin, A., Aspan, A., and Gunnarsson, A. (2003) *Clostridium difficile*: prevalence in horses and environment, and antimicrobial susceptibility. *Equine Vet J* **35**: 465-471.

Beckering, C.L., Steil, L., Weber, M.H., Volker, U., and Marahiel, M.A. (2002) Genomewide transcriptional analysis of the cold shock response in *Bacillus subtilis*. *J Bacteriol* **184**: 6395-6402.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**: 783-795.

Bernhardt, J., Volker, U., Volker, A., Antelmann, H., Schmid, R., Mach, H., and Hecker, M. (1997) Specific and general stress proteins in *Bacillus subtilis*--a two-dimensional protein electrophoresis study. *Microbiology* **143**: 999-1017.

Betts, J.C. (2002) Transcriptomics and proteomics: tools for the identification of novel drug targets and vaccine candidates for tuberculosis. *IUBMB Life* **53**: 239-242.

Beveridge, T.J., and Graham, L.L. (1991) Surface layers of bacteria. *Microbiol Rev* **55**: 684-705.

Beveridge, T.J., Pouwels, P.H., Sara, M., Kotiranta, A., Lounatmaa, K., Kari, K., Kerosuo, E., Haapasalo, M., Egelseer, E.M., Schocher, I., Sleytr, U.B., Morelli, L., Callegari, M.L., Nomellini, J.F., Bingle, W.H., Smit, J., Leibovitz, E., Lemaire, M., Miras, I., Salamitou, S., Beguin, P., Ohayon, H., Gounon, P., Matuschek, M., and Koval, S.F. (1997) Functions of S-layers. *FEMS Microbiol Rev* **20**: 99-149.

Bingle, W.H., Engelhardt, H., Page, W.J., and Baumeister, W. (1987a) Three-dimensional structure of the regular tetragonal surface layer of *Azotobacter vinelandii*. *J Bacteriol* **169**: 5008-5015.

Bingle, W.H., Whippey, P.W., Doran, J.L., Murray, R.G., and Page, W.J. (1987b) Structure of the *Azotobacter vinelandii* surface layer. *J Bacteriol* **169**: 802-810.

Blaser, M.J., and Pei, Z. (1993) Pathogenesis of *Campylobacter fetus* infections: critical role of high-molecular-weight S-layer proteins in virulence. *Journal of Infectious Diseases* **167**: 372-377.

Blobel, G. (1980) Regulation of intracellular protein traffic. *Harvey Lect* **76**: 125-147.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**: 365-370.

Bogsch, E.G., Sargent, F., Stanley, N.R., Berks, B.C., Robinson, C., and Palmer, T. (1998) An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. *J Biol Chem* **273**: 18003-18006.

Bolton, R.P., Tait, S.K., Dear, P.R., and Losowsky, M.S. (1984) Asymptomatic neonatal colonisation by *Clostridium difficile*. *Arch Dis Child* **59**: 466-472.

Boot, H.J., Kolen, C.P., Andreadaki, F.J., Leer, R.J., and Pouwels, P.H. (1996a) The *Lactobacillus acidophilus* S-layer protein gene expression site comprises two consensus promoter sequences, one of which directs transcription of stable mRNA. *J Bacteriol* **178**: 5388-5394.

Boot, H.J., Kolen, C.P., and Pouwels, P.H. (1996b) Interchange of the active and silent S-layer protein genes of *Lactobacillus acidophilus* by inversion of the chromosomal slp segment. *Mol Microbiol* **21**: 799-809.

Boot, H.J., and Pouwels, P.H. (1996) Expression, secretion and antigenic variation of bacterial S-layer proteins. *Mol Microbiol* **21**: 1117-1123.

Borriello, S.P., and Barclay, F.E. (1986) An *in-vitro* model of colonisation resistance to Clostridium difficile infection. *J Med Microbiol* **21**: 299-309.

Borriello, S.P., Ketley, J.M., Mitchell, T.J., Barclay, F.E., Welch, A.R., Price, A.B., and Stephen, J. (1987) *Clostridium difficile*--a spectrum of virulence and analysis of putative virulence determinants in the hamster model of antibiotic-associated colitis. *J Med Microbiol* **24**: 53-64.

Borriello, S.P., Davies, H.A., and Barclay, F.E. (1988) Detection of fimbriae amongst strains of *Clostridium difficile*. *FEMS Microbiology Letters* **49**: 65-67.

Bowditch, R.D., Baumann, P., and Yousten, A.A. (1989) Cloning and sequencing of the gene encoding a 125-kilodalton surface-layer protein from *Bacillus sphaericus* 2362 and of a related cryptic gene. *J Bacteriol* **171**: 4178-4188.
Brazier, J.S. (2001) Typing of *Clostridium difficile*. *Clin Microbiol Infect* **7**: 428-431.
Brechtel, E., and Bahl, H. (1999) In *Thermoanaerobacterium thermosulfurigenes* EM1 S-Layer Homology Domains Do Not Attach to Peptidoglycan. *J. Bacteriol.* **181**: 5017-5023.
Briggs, M.S., and Gierasch, L.M. (1986) Molecular mechanisms of protein secretion: the role of the signal sequence. *Adv Protein Chem* **38**: 109-180.
Brown, E.B., Talbot, G.H., Axelrod, P., Provencher, M., and Hoegg, C. (1990) Risk factors for *Clostridium difficile* toxin-associated diarrhea. *Infect Control Hosp Epidemiol* **11**: 283-290.
Brown, J.S., and Holden, D.W. (2002) Iron acquisition by Gram-positive bacterial pathogens. *Microbes and Infection* **4**: 1149-1156.
Bruggemann, H. (2005) Genomics of clostridial pathogens: implication of extrachromosomal elements in pathogenicity. *Curr Opin Microbiol* **8**: 601-605.
Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G., and Gallagher, T. (1995) Catalysis of a protein folding reaction: mechanistic implications of the 2.0 A structure of the subtilisin-prodomain complex. *Biochemistry* **34**: 10310-10318.
Bumann, D., Jungblut, P.R., and Meyer, T.F. (2004) *Helicobacter pylori* vaccine development based on combined subproteome analysis. *Proteomics* **4**: 2843-2848.
Bunai, K., Takamatsu, H., Horinaka, T., Oguro, A., Nakamura, K., and Yamane, K. (1996) *Bacillus subtilis Ffh,* a homologue of mammalian SRP54, can intrinsically bind to the precursors of secretory proteins. *Biochem Biophys Res Commun* **227**: 762-767.
Burrus, V., Pavlovic, G., Decaris, B., and Guedon, G. (2002) The ICESt1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**: 77-97.
Buttner, K., Bernhardt, J., Scharf, C., Schmid, R., Mader, U., Eymann, C., Antelmann, H., Volker, A., Volker, U., Hecker, M. (2001) A comprehensive two-dimensional map of cytosolic proteins of *Bacillus subtilis*. *Electrophoresis* **22**: 2908-35
Calabi, E., Calabi, F., Phillips, A.D., and Fairweather, N.F. (2002) Binding of *Clostridium difficile* Surface Layer Proteins to Gastrointestinal Tissues. *Infect. Immun.* **70**: 5770-5778.
Calabi, E., and Fairweather, N. (2002) Patterns of Sequence Conservation in the S-Layer Proteins and Related Sequences in *Clostridium difficile*. *J. Bacteriol.* **184**: 3886-3897.
Calabi, E., Ward, S., Wren, B., Paxton, T., Panico, M., Morris, H., Dell, A., Dougan, G., Fairweather, N. (2001) Molecular characterization of the surface layer proteins from *Clostridium difficile*. *Mol Microbiol* **40**: 1187-1199.
Canny, G., Drudy, D., Macmathuna, P., O'Farrelly, C., and Baird, A.W. (2005) Toxigenic *C. difficile* induced inflammatory marker expression by human intestinal epithelial cells is asymmetrical. *Life Sci.*
Cao, T.B., and Saier, M.H., Jr. (2003) The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biochim Biophys Acta* **1609**: 115-125.
Carl, M., and Dasch, G.A. (1989) The importance of the crystalline surface layer protein antigens of rickettsiae in T-cell immunity. *J Autoimmun* **2 Suppl**: 81-91.
Carsiotis, M., Stocker, B.A., Weinstein, D.L., and O'Brien, A.D. (1989) A *Salmonella typhimurium* virulence gene linked to *flg*. *Infect Immun* **57**: 3276-3280.
Carter, G.P., Purdy, D., Williams, P., and Minton, N.P. (2005) Quorum sensing in *Clostridium difficile*: analysis of a *luxS*-type signalling system. *J Med Microbiol* **54**: 119-127.
Cava, F., de Pedro, M.A., Schwartz, H., Henne, A., and Berenguer, A. (2004) Binding to pyruvylated compounds as an ancestral mecnanism to anchor the outer envelope in primative bateria. *Mol Microbiol* **52**: 677-690.
Cerquetti, M., Molinari, A., Sebastianelli, A., Diociaiuti, M., Petruzzelli, R., Capo, C., and Mastrantonio, P. (2000) Characterization of surface layer proteins from different *Clostridium difficile* clinical isolates. *Microbial Pathogenesis* **28**: 363-372.
Chaddock, A.M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R.G., Klosgen, R.B., and Robinson, C. (1995) A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. *Embo J* **14**: 2715-2722.
Chakravarti, D.N., Fiske, M.J., Fletcher, L.D., and Zakursky, R.J. (2001) Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* **19**: 601-612.

Chami, M., Bayan, N., Peyret, J.L., Gulik-Krzywicki, T., Leblon, G., and Shechter, E. (1997) The S-layer protein of *Corynebacterium glutamicum* is anchored to the cell wall by its C-terminal hydrophobic domain. *Mol Microbiol* **23**: 483-492.

Chang, T.W., Gorbach, S.L., and Bartlett, J.B. (1978) Neutralization of *Clostridium difficile* toxin by *Clostridium sordellii* antitoxins. *Infect Immun* **22**: 418-422.

Chang, T.W., Lin, P.S., Gorbach, S.L., and Bartlett, J.G. (1979) Ultrastructural changes of cultured human amnion cells by *Clostridium difficile* toxin. *Infect Immun* **23**: 795-798.

Chaves-Olarte, E., Weidmann, M., Eichel-Streiber, C., and Thelestam, M. (1997) Toxins A and B from *Clostridium difficile* differ with respect to enzymatic potencies, cellular substrate specificities, and surface binding to cultured cells. *J Clin Invest* **100**: 1734-1741.

Ching, W.M., Carl, M., and Dasch, G.A. (1992) Mapping of monoclonal antibody binding sites on CNBr fragments of the S-layer protein antigens of *Rickettsia typhi* and *Rickettsia prowazekii*. *Mol Immunol* **29**: 95-105.

Chitlaru, T., Ariel, N., Zvi, A., Lion, M., Velan, B., Shafferman, A., and Elhanany, E. (2004) Identification of chromosomally encoded membranal polypeptides of *Bacillus anthracis* by a proteomic analysis: prevalence of proteins containing S-layer homology domains. *Proteomics* **4**: 677-691.

Chung, Y.J., and Saier, M.H., Jr. (2002) Overexpression of the *Escherichia coli sugE* gene confers resistance to a narrow range of quaternary ammonium compounds. *J Bacteriol* **184**: 2543-2545.

Clabots, C.R., Johnson, S., Bettin, K.M., Mathie, P.A., Mulligan, M.E., Schaberg, D.R., Peterson, L.R., and Gerding, D.N. (1993) Development of a rapid and efficient restriction endonuclease analysis typing system for *Clostridium difficile* and correlation with other typing systems. *J Clin Microbiol* **31**: 1870-1875.

Cline, K., Ettinger, W.F., and Theg, S.M. (1992) Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two lumenal proteins are transported in the absence of ATP. *J Biol Chem* **267**: 2688-2696.

Collier, M.C., Stock, F., DeGirolami, P.C., Samore, M.H., and Cartwright, C.P. (1996) Comparison of PCR-based approaches to molecular epidemiologic analysis of *Clostridium difficile*. *J Clin Microbiol* **34**: 1153-1157.

Costas, M., Holmes, B., On, S.L., Ganner, M., Kelly, M.C., and Nath, S.K. (1994) Investigation of an outbreak of *Clostridium difficile* infection in a general hospital by numerical analysis of protein patterns by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *J Clin Microbiol* **32**: 759-765.

Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J.S., and Cygler, M. (1996) Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *Embo J* **15**: 5492-5503.

Dalbey, R.E., Lively, M.O., Bron, S., and van Dijl, J.M. (1997) The chemistry and enzymology of the type I signal peptidases. *Protein Sci* **6**: 1129-1138.

Dalbey, R.E., Chen, M., Jiang, F., and Samuelson, J.C. (2000) Understanding the insertion of transporters and other membrane proteins. *Curr Opin Cell Biol* **12**: 435-442.

Davies, H.A., and Borriello, S.P. (1990) Detection of capsule in strains of *Clostridium difficile* of varying virulence and toxigenicity. *Microb Pathog* **9**: 141-146.

De Gier, J.W., Valent, Q.A., Von Heijne, G., and Luirink, J. (1997) The *E. coli* SRP: preferences of a targeting factor. *FEBS Lett* **408**: 1-4.

de Leeuw, E., Poland, D., Mol, O., Sinning, I., ten Hagen-Jongman, C.M., Oudega, B., and Luirink, J. (1997) Membrane association of FtsY, the E. coli SRP receptor. *FEBS Lett* **416**: 225-229.

de Vrije, G.J., Batenburg, A.M., Killian, J.A., and de Kruijff, B. (1990) Lipid involvement in protein translocation in *Escherichia coli*. *Mol Microbiol* **4**: 143-150.

Delmee, M., Homel, M., and Wauters, G. (1985) Serogrouping of *Clostridium difficile* strains by slide agglutination. *J Clin Microbiol* **21**: 323-327.

Delmee, M. (2001) Laboratory diagnosis of *Clostridium difficile* disease. *Clin Microbiol Infect* **7**: 411-416.

Department of Health (2006) www.dh.gov.uk.

Dev, I.K., and Ray, P.H. (1990) Signal peptidases and signal peptide hydrolases. *J Bioenerg Biomembr* **22**: 271-290.

Dhar, G., Faull, K.F., and Schneewind, O. (2000) Anchor structure of cell wall surface proteins in *Listeria monocytogenes*. *Biochemistry* **39**: 3725-3733.

Doi, R.H., and Kosugi, A. (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat Rev Microbiol* **2**: 541-551.

Drenth, J., Jansonius, J.N., Koekoek, R., Swen, H.M., and Wolthers, B.G. (1968) Structure of papain. *Nature* **218**: 929-932.

Drenth, J., Jansonius, J.N., Koekoek, R., and Wolthers, B.G. (1971) The structure of papain. *Adv Protein Chem* **25**: 79-115.
Dridi, L., Tankovic, J., and Petit, J.C. (2004) CdeA of *Clostridium difficile*, a new multidrug efflux transporter of the MATE family. *Microb Drug Resist* **10**: 191-196.
Drudy, D., Calabi, E., Kyne, L., Sougioultzis, S., Kelly, E., Fairweather, N., and Kelly, C.P. (2004) Human antibody response to surface layer proteins in *Clostridium difficile* infection. *FEMS Immunol Med Microbiol* **41**: 237-242.
Drummond, L.J., Smith, D.G., and Poxton, I.R. (2003) Effects of sub-MIC concentrations of antibiotics on growth of and toxin production by *Clostridium difficile*. *J Med Microbiol* **52**: 1033-1038.
Dubreuil, J.D., Logan, S.M., Cubbage, S., Eidhin, D.N., McCubbin, W.D., Kay, C.M., Beveridge, T.J., Ferris, F.G., and Trust, T.J. (1988) Structural and biochemical analyses of a surface array protein of *Campylobacter fetus*. *J Bacteriol* **170**: 4165-4173.
Duong, F., and Wickner, W. (1997) The SecDFYajC domain of preprotein translocase controls preprotein movement by regulating SecA membrane cycling. *Embo J* **16**: 4871-4879.
Dupuy, B., and Sonenshein, A.L. (1998) Regulated transcription of *Clostridium difficile* toxin genes. *Mol Microbiol* **27**: 107-120.
Dworkin, J., Tummuru, M.K., and Blaser, M.J. (1995) A lipopolysaccharide-binding domain of the *Campylobacter fetus* S-layer protein resides within the conserved N terminus of a family of silent and divergent homologs. *J Bacteriol* **177**: 1734-1741.
Dworkin, J., and Blaser, M.J. (1996) Generation of *Campylobacter fetus* S-layer protein diversity utilizes a single promoter on an invertible DNA segment. *Mol Microbiol* **19**: 1241-1253.
Dworkin, J., and Blaser, M.J. (1997) Molecular mechanisms of *Campylobacter fetus* surface layer protein expression. *Mol Microbiol* **26**: 433-440.
Dworkin, J., Shedd, O.L., and Blaser, M.J. (1997) Nested DNA inversion of *Campylobacter fetus* S-layer genes is *recA* dependent. *J Bacteriol* **179**: 7523-7529.
Egelseer, E., Schocher, I., Sara, M., and Sleytr, U. (1995) The S-layer from *Bacillus stearothermophilus* DSM 2358 functions as an adhesion site for a high-molecular-weight amylase. *J. Bacteriol.* **177**: 1444-1451.
Egelseer, E.M., Leitner, K., Jarosch, M., Hotzy, C., Zayni, S., Sleytr, U.B., and Sara, M. (1998) The S-layer proteins of two *Bacillus stearothermophilus* wild-type strains are bound via their N-terminal region to a secondary cell wall polymer of identical chemical composition. *J Bacteriol* **180**: 1488-1495.
Eggertson, L. (2005a) *C. difficile strain* 20 times more virulent. *CMAJ* **172**: 1279-.
Eggertson, L. (2005b) Quebec reports *C. difficile* mortality statistics. *CMAJ* **173**: 139-.
Ellwood, D.C., and Tempest, D.W. (1972) Influence of culture pH on the content and composition of teichoic acids in the walls of *Bacillus subtilis*. *J Gen Microbiol* **73**: 395-402.
Emr, S.D., and Silhavy, T.J. (1982) Molecular components of the signal sequence that function in the initiation of protein export. *J Cell Biol* **95**: 689-696.
Engelhardt, H., and Peters, J. (1998) Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions. *J Struct Biol* **124**: 276-302.
Etz, H., Minh, D.B., Henics, T., Dryla, A., Winkler, B., Triska, C., Boyd, A.P., Sollner, J., Schmidt, W., von Ahsen, U., Buschle, M., Gill, S.R., Kolonay, J., Khalak, H., Fraser, C.M., von Gabain, A., Nagy, E., and Meinke, A. (2002) Identification of in vivo expressed vaccine candidate antigens from *Staphylococcus aureus*. *Proc Natl Acad Sci U S A* **99**: 6573-6578.
Eveillard, M., Fourel, V., Barc, M.C., Kerneis, S., Coconnier, M.H., Karjalainen, T., Bourlioux, P., and Servin, A.L. (1993) Identification and characterization of adhesive factors of *Clostridium difficile* involved in adhesion to human colonic enterocyte-like Caco-2 and mucus-secreting HT29 cells in culture. *Mol Microbiol* **7**: 371-381.
Farrow, K.A., Lyras, D., and Rood, J.I. (2001) Genomic analysis of the erythromycin resistance element Tn5398 from *Clostridium difficile*. *Microbiology* **147**: 2717-2728.
Feavers, I.M., Foulkes, J., Setlow, B., Sun, D., Nicholson, W., Setlow, P., and Moir, A. (1990) The regulation of transcription of the *gerA* spore germination operon of *Bacillus subtilis*. *Mol Microbiol* **4**: 275-282.
Ferdinand, W. (1964) The isolation and specific activity of rabbit-muscle glyceraldehyde phosphate dehydrogenase. *Biochem J* **92**: 578-585.
Fernández-Arenas, E., Molero, G., Nombela, C., Diez-Orejas, R., and Gil, C. (2004) Low virulent strains of *Candida albicans*: Unravelling the antigens for a future vaccine. *Proteomics* **4**: 3007–3020.

Fernie, D.S., Thomson, R.O., Batty, I., and Walker, P.D. (1983) Active and passive immunization to protect against antibiotic associated caecitis in hamsters. *Dev Biol Stand* **53**: 325-332.
Ferrero, R.L., and Labigne, A. (2001) *Helicobacter pylori* vaccine development in the post-genomic era: can *in silico* translate to *in vivo*. *Scand J Immunol* **53**: 443-448.
Finazzi, G., Chasen, C., Wollman, F.A., and de Vitry, C. (2003) Thylakoid targeting of Tat passenger proteins shows no delta pH dependence *in vivo*. *Embo J* **22**: 807-815.
Finne, J., Bitter-Suermann, D., Goridis, C., and Finne, U. (1987) An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. *J Immunol* **138**: 4402-4407.
Finney, J.M.T. (1893) Gastro-entrostomy for cicatrizing ulcer of the pylorus. *Bull. John Hopkins Hosp.* **4**: 53-55.
Fischer, H., and Tomasz, A. (1984) Production and release of peptidoglycan and wall teichoic acid polymers in pneumococci treated with beta-lactam antibiotics. *J Bacteriol* **157**: 507-513.
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs Jr, W.R., Jr., Venter, J.C., and Fraser, C.M. (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* **184**: 5479-5490.
Folio, P., Chavant, P., Chafsey, I., Belkorchia, A., Chambon, C., and Hebraud, M. (2004) Two-dimensional electrophoresis database of *Listeria monocytogenes* EGDe proteome and proteomic analysis of mid-log and stationary growth phase cells. *Proteomics* **4**: 3187-3201.
Foster, T.J., and Hook, M. (1998) Surface protein adhesins of *Staphylococcus aureus*. *Trends Microbiol* **6**: 484-488.
Fouet, A., and Mesnage, S. (2002) *Bacillus anthracis* cell envelope components. *Curr Top Microbiol Immunol* **271**: 87-113.
Gade, D., Stuhrmann, T., Reinhardt, R., and Rabus, R. (2005) Growth phase dependent regulation of protein composition in *Rhodopirellula baltica*. *Environ Microbiol* **7**: 1074-1084.
Garnier, F., Taourit, S., Glaser, P., Courvalin, P., and Galimand, M. (2000) Characterization of transposon Tn1549, conferring VanB-type resistance in *Enterococcus* spp. *Microbiology* **146 (Pt 6)**: 1481-1489.
Gavin, R., Merino, S., Altarriba, M., Canals, R., Shaw, J.G., and Tomas, J.M. (2003) Lateral flagella are required for increased cell adherence, invasion and biofilm formation by *Aeromonas* spp. *FEMS Microbiol Lett* **224**: 77-83.
Genta, R.M. (1997) The immunobiology of *Helicobacter pylori* gastritis. *Semin Gastrointest Dis* **8**: 2-11.
George, R.H., Symonds, J.M., Dimock, F., Brown, J.D., Arabi, Y., Shinagawa, N., Keighley, M.R., Alexander Williams, J. and Buron, D.W. (1978) Identification of *Clostridium difficile* as a cause of pseudomembraneous colitis. *Br. Med. J.* **1**: 695.
Gerding, D.N., Johnson, S., Peterson, L.R., Mulligan, M.E., and Silva, J., Jr. (1995) *Clostridium difficile*-associated diarrhea and colitis. *Infect Control Hosp Epidemiol* **16**: 459-477.
Giannasca, P.J., Zhang, Z.X., Lei, W., Boden, J.A., Giel, M.A., Monath, T.P., and Thomas, W.D.J. (1999) Serum antitoxin antibodies mediate systemic and mucosal protection from *Clostridium difficile* disease in hamsters. *Infect Immun* **67**: 527-538.
Giannasca, P.J., and Warny, M. (2004) Active and passive immunization against *Clostridium difficile* diarrhea and colitis. *Vaccine* **22**: 848-856.
Gilson, E., Alloing, G., Schmidt, T., Claverys, J.P., Dudler, R., and Hofnung, M. (1988) Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in Mycoplasma. *Embo J* **7**: 3971-3974.
Giron, J.A., Torres, A.G., Freer, E., and Kaper, J.B. (2002) The flagella of enteropathogenic *Escherichia coli* mediate adherence to epithelial cells. *Mol Microbiol* **44**: 361-379.
Glauner, B., Holtje, J.V., and Schwarz, U. (1988) The composition of the murein of *Escherichia coli*. *J Biol Chem* **263**: 10088-10095.
Gorg, A., Weiss, W., and Dunn, M.J. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**: 3665-3685.
Grandi, G. (2001) Antibacterial vaccine design using genomics and proteomics. *Trends Biotechnol* **19**: 181-188.
Green, R.H. (1974) The association of viral activation with penicillin toxicity in guinea pigs and hamsters. *Yale J. Biol. Med* **47**: 166-181.

Groves, M.R., Taylor, M.A., Scott, M., Cummings, N.J., Pickersgill, R.W., and Jenkins, J.A. (1996) The prosequence of procaricain forms an alpha-helical domain that prevents access to the substrate-binding cleft. *Structure* **4**: 1193-1203.

Groves, M.R., Coulombe, R., Jenkins, J., and Cygler, M. (1998) Structural basis for specificity of papain-like cysteine protease proregions toward their cognate enzymes. *Proteins: Structure, Function and Genetics* **32**: 504-514.

Gruber, K., and Sleytr, U.B. (1988) Localized insertion of new S-layer during growth of *Bacillus stearothermophilus* strains. *Arch Microbiol* **149**: 485-491.

Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Lamkiewicz, L., Wiczk, W., Wieczerzak, E., Ciarkowski, J., Drabik, P., Janowski, R., Kozak, M., Jaskolski, M., and Grubb, A. (2001) Structural studies of cysteine proteases and their inhibitors. *Acta Biochimica Polonica* **48**: 1-20.

Haas, G., Karaali, G., Ebermayer, K., Metzger, W.G., Lamer, S., Zimny-Arndt, U., Diescher, S., Goebel, U.B., Vogt, K., Roznowski, A.B., Wiedenmann, B.J., Meyer, T.F., Aebischer, T., and Jungblut, P.R. (2002) Immunoproteomics of *Helicobacter pylori* infection and relation to gastric disease. *Proteomics* **2**: 313-324.

Hall, I.C., and O'Toole, E. (1935) Intestinal flora in new born infents with a description of a new pathogenic anaerobe, *Bacillus difficilis*. *Am. J. Dis. Child* **49**: 390-402.

Hamel, J., Martin, D., and Brodeur, B.B. (1997) Heat shock response of *Streptococcus pneumoniae*: identification of immunoreactive stress proteins. *Microb Pathog* **23**: 11-21.

Hammond, G.A., Lyerly, D.M. and Johnson, J.L. (1997) Transcriptional analysis of the toxigenic element of *Clostridium difficile*. *Microbial Pathogenesis* **22**: 143-154.

Hancock, I.C., and Poxton, I.R. (1988) *Bacterial Cell Surface Techniques*: John Wiley & Sons.

Hansmeier, N., Bartels, F. W., Ros, R., Anselmetti, D., Tauch, A., Puhler, A., Kalinowski, J. (2004) Classification of hyper-variable *Corynebacterium glutamicum* surface-layer proteins by sequence analyses and atomic force microscopy. *J Biotechnol.* **112**: 177-93

Haraldsen, J.D., and Sonenshein, A.L. (2003) Efficient sporulation in *Clostridium difficile* requires disruption of the *sigma K* gene. *Mol Microbiol* **48**: 811-821.

Harwood, C.R., Coxon, R.D., and Hancock, I.C. Chapter 8. In *Molecular Biology Methods for Bacillus*. Harwood, C.R. and Cutting, S.M. (eds).

Harwood, C.R., Coxon, R.D., and Hancock, I.C. (1990) The *Bacillus* cell envelope and secretion. In *Molecular biological methods for Bacillus*. Harwood, C.R. and Cutting, S.M. (eds). Chichester: Wiley, pp. 378.

Havlasova, J., Hernychova, L., Halada, P., Pellantova, V., Krejsek, J., Stulik, J., Macela, A., Jungblut, P.R., Larsson, P., and Forsman, M. (2002) Mapping of immunoreactive antigens of *Francisella tularensis* live vaccine strain. *Proteomics* **2**: 857-867.

Hayrinen, J., Jennings, H., Raff, H.V., Rougon, G., Hanai, N., Gerardy-Schahn, R., and Finne, J. (1995) Antibodies to polysialic acid and its N-propyl derivative: binding properties and interaction with human embryonal brain glycopeptides. *J Infect Dis* **171**: 1481-1490.

Hecker, M., Schumann, W., and Volker, U. (1996) Heat-shock and general stress response in *Bacillus subtilis*. *Mol Microbiol* **19**: 417-428.

Hennequin, C., Collignon, A., and Karjalainen, T. (2001a) Analysis of expression of GroEL (Hsp60) of *Clostridium difficile* in response to stress. *Microb Pathog* **31**: 255-260.

Hennequin, C., Porcheray, F., Waligora-Dupriet, A., Collignon, A., Barc, M., Bourlioux, P., and Karjalainen, T. (2001b) GroEL (Hsp60) of *Clostridium* difficile is involved in cell adherence. *Microbiology* **147**: 87-96.

Hennequin, C., Janoir, C., Barc, M.C., Collignon, A., and Karjalainen, T. (2003) Identification and characterization of a fibronectin-binding protein from *Clostridium difficile*. *Microbiology* **149**: 2779-2787.

Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases *Proc Natl Acad Sci* **90**: 5011-5

Heptinstall, S., Archibald, A.R., and Baddiley, J. (1970) Teichoic acids and membrane function in bacteria. *Nature* **225**: 519-521.

Higgins, D.G. (1994) CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol Biol* **25**: 307-318.

Ho, J. G., Greco, A., Rupnik, M., Ng, K. K. (1999) Crystal structure of receptor-binding C-terminal repeats from *Clostridium difficile* toxin A. *Proc Natl Acad Sci* **102**: 18373-8.

Holland, I.B., and Blight, M.A. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. *J Mol Biol* **293**: 381-399.

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**: 6072-6076.

Hood, D.W., Deadman, M.E., Jennings, M.P., Bisercic, M., Fleischmann, R.D., Venter, J.C., and Moxon, E.R. (1996) DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *Proc Natl Acad Sci U S A* **93**: 11121-11125.

Huber, C., Ilk, N., Runzler, D., Egelseer, E.M., Weigert, S., Sleytr, U.B., and Sara, M. (2005) The three S-layer-like homology motifs of the S-layer protein SbpA of *Bacillus sphaericus* CCM 2177 are not sufficient for binding to the pyruvylated secondary cell wall polymer. *Mol Microbiol* **55**: 197-205.

Hughes, M.J., Moore, J.C., Lane, J.D., Wilson, R., Pribul, P.K., Younes, Z.N., Dobson, R.J., Everest, P., Reason, A.J., Redfern, J.M., Greer, F.M., Paxton, T., Panico, M., Morris, H.R., Feldman, R.G., and Santangelo, J.D. (2002) Identification of major outer surface proteins of *Streptococcus agalactiae*. *Infect Immun* **70**: 1254-1259.

Hulo, N., Sigrist, C.J., Le Saux, V., Langendijk-Genevaux, P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004) Recent improvements to the PROSITE database. *Nucleic Acids Res* **32**: D134-137.

Hundsberger, T., Braun, V., Weidmann, M., Leukel, P., Sauerborn, M., and von Eichel-Streiber, C. (1997) Transcription analysis of the genes tcdA-E of the pathogenicity locus of *Clostridium difficile*. *Eur J Biochem* **244**: 735-742.

Hynds, P.J., Robinson, D., and Robinson, C. (1998) The sec-independent twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane. *J Biol Chem* **273**: 34868-34874.

Ilk, N., Kosma, P., Puchberger, M., Egelseer, E.M., Mayer, H.F., Sleytr, U.B., and Sara, M. (1999) Structural and functional analyses of the secondary cell wall polymer of *Bacillus sphaericus* CCM 2177 that serves as an S-layer-specific anchor. *J Bacteriol* **181**: 7643-7646.

Ishihama, A. (1997) Adaptation of gene expression in stationary phase bacteria. *Curr Opin Genet Dev* **7**: 582-588.

Jacobsson, S., Thulin, S., Molling, P., Unemo, M., Comanducci, M., Rappuoli, R., and Olcen, P. (2006) Sequence constancies and variations in genes encoding three new meningococcal vaccine candidate antigens. *Vaccine* **24**: 2161-2168.

Jakava-Viljanen, M., Avall-Jaaskelainen, S., Messner, P., Sleytr, U.B., and Palva, A. (2002) Isolation of three new surface layer protein genes (*slp*) from *Lactobacillus brevis* ATCC 14869 and characterization of the change in their expression under aerated and anaerobic conditions. *J Bacteriol* **184**: 6786-6795.

Janoir, C., Grenery, J., Savariau-Lacomme, M.P., and Collignon, A. (2004) [Characterization of an extracellular protease from *Clostridium difficile*]. *Pathol Biol (Paris)* **52**: 444-449.

Jing, H., Takagi, J., Liu, J.H., Lindgren, S., Zhang, R.G., Joachimiak, A., Wang, J.H., and Springer, T.A. (2002) Archaeal surface layer proteins contain beta propeller, PKD, and beta helix domains and are related to metazoan cell surface proteins. *Structure (Camb)* **10**: 1453-1464.

Johal, S.S., Lambert, C.P., Hammond, J., James, P.D., Borriello, S.P., and Mahida, Y.R. (2004) Colonic IgA producing cells and macrophages are reduced in recurrent and non-recurrent *Clostridium difficile* associated diarrhoea. *J Clin Pathol* **57**: 973-979.

Johnson, S., Clabots, C.R., Linn, F.V., Olson, M.M., Peterson, L.R., and Gerding, D.N. (1990) Nosocomial *Clostridium difficile* colonisation and disease. *Lancet* **336**: 97-100.

Johnson, S., Sypura, W.D., Gerding, D.N., Ewing, S.L., and Janoff, E.N. (1995) Selective neutralization of a bacterial enterotoxin by serum immunoglobulin A in response to mucosal disease. *Infect Immun* **63**: 3166-3173.

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**: 902-914.

Jungblut, P.R., Schaible, U.E., Mollenkopf, H.J., Zimny-Arndt, U., Raupach, B., Mattow, J., Halada, P., Lamer, S., Hagens, K., and Kaufmann, S.H. (1999) Comparative proteome analysis of *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG strains: towards functional genomics of microbial pathogens. *Mol Microbiol* **33**: 1103-1117.

Just, I., Selzer, J., Wilm, M., von Eichel-Streiber, C., Mann, M., and Aktories, K. (1995a) Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature* **375**: 500-503.

Just, I., Wilm, M., Selzer, J., Rex, G., von Eichel-Streiber, C., Mann, M., and Aktories, K. (1995b) The enterotoxin from *Clostridium difficile* (ToxA) monoglucosylates the Rho proteins. *J Biol Chem* **270**: 13932-13936.

Just, I., Hofmann, F., and Aktories, K. (2000) Molecular mode of action of the large clostridial cytotoxins. *Curr Top Microbiol Immunol* **250**: 55-83.
Kahala, M., Savijoki, K., and Palva, A. (1997) In vivo expression of the *Lactobacillus brevis* S-layer gene. *J. Bacteriol.* **179**: 284-286.
Kamphuis, I.G., Kalk, K.H., Swarte, M.B., and Drenth, J. (1984) Structure of papain refined at 1.65 A resolution. *J Mol Biol* **179**: 233-256.
Kamphuis, I.G., Drenth, J., and Baker, E.N. (1985) Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *J Mol Biol* **182**: 317-329.
Karjalainen, T., Barc, M.C., Collignon, A., Trolle, S., Boureau, H., Cotte-Laffitte, J., and Bourlioux, P. (1994) Cloning of a genetic determinant from *Clostridium difficile* involved in adherence to tissue culture cells and mucus. *Infect Immun* **62**: 4347-4355.
Karjalainen, T., Saumier, N., Barc, M.C., Delmee, M., and Collignon, A. (2002) *Clostridium difficile* genotyping based on *slpA* variable region in S-layer gene sequence: an alternative to serotyping. *J Clin Microbiol* **40**: 2452-2458.
Karlsson, S., Dupuy, B., Mukherjee, K., Norin, E., Burman, L.G., and Akerlund, T. (2003) Expression of *Clostridium difficile* toxins A and B and their sigma factor TcdD is controlled by temperature. *Infect Immun* **71**: 1784-1793.
Kelley, L.A., MacCallum, R.M., and Sternberg, M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J Mol Biol* **299**: 499-520.
Kelly, C., Pothoulakis, C., Orellana, J., and LaMont, J. (1992) Human colonic aspirates containing immunoglobulin A antibody to *Clostridium difficile* toxin A inhibit toxin A-receptor binding. *Gastroenterology* **102**: 35-40.
Kelly, C.P., Pothoulakis, C., and LaMont, J.T. (1994) C*lostridium difficile* colitis. *N Engl J Med* **330**: 257-262.
Kelly, C.P., and LaMont, J.T. (1998) Clostridium difficile infection. *Annu Rev Med* **49**: 375-390.
Kelly, P., Maguire, P.B., Bennett, M., Fitzgerald, D.J., Edwards, R.J., Thiede, B., Treumann, A., Collins, J.K., O'Sullivan, G.C., Shanahan, F., and Dunne, C. (2005) Correlation of probiotic *Lactobacillus salivarius* growth phase with its cell wall-associated proteome. *FEMS Microbiol Lett* **252**: 153-159.
Kesavalu, L., Holt, S.C., Crawley, R.R., Borinski, R., and Ebersole, J.L. (1991) Virulence of *Wolinella recta* in a murine abscess model. *Infect Immun* **59**: 2806-2817.
Kim, K.H., Fekety, R., Batts, D.H., Brown, D., Cudmore, M., Silva, J., Jr., and Waters, D. (1981) Isolation of *Clostridium difficile* from the environment and contacts of patients with antibiotic-associated colitis. *J Infect Dis* **143**: 42-50.
Kim, P.H., Iaconis, J.P., and Rolfe, R.D. (1987a) Immunization od adult hamsters against *Clostridium difficile*-assocaited ileocecitis and transfer of protection to infant hamsters. *Infect Immun* **55**: 2984-2992.
Kim, P.H., Iaconis, J.P., and Rolfe, R.D. (1987b) Immunization of adult hamsters against *Clostridium difficile*-associated ileocecitis and transfer of protection to infant hamsters. *Infect Immun* **55**: 2984-2992.
Kink, J.A., and Williams, J.A. (1998) Antibodies to recombinant *Clostridium difficile* toxins A and B are an effective treatment and prevent relapse of *C. difficile*-associated disease in a hamster model of infection. *Infect Immun* **66**: 2018-2025.
Kist, M. L., Murray, R. G. (1984) Components of the regular surface array of *Aquaspirillum serpens* MW5 and their assembly in vitro. *J Bacteriol* **157**: 599-606
Knoop, F.C., Owens, M., and Crocker, I.C. (1993) *Clostridium difficile*: clinical disease and diagnosis. *Clin Microbiol Rev* **6**: 251-265.
Kobayashi, Y., Ohta, H., Kokeguchi, S., Murayama, Y., Kato, K., Kurihara, H., and Fukui, K. (1993) Antigenic properties of *Campylobacter rectus (Wolinella recta)* major S-layer proteins. *FEMS Microbiol Lett* **108**: 275-280.
Kohler, C., Wolff, S., Albrecht, D., Fuchs, S., Becher, D., Buttner, K., Engelmann, S., and Hecker, M. (2005) Proteome analyses of *Staphylococcus aureus* in growing and non-growing cells: A physiological approach. *Int J Med Microbiol* **295**: 547-565.
Kojima, S., and Blair, D.F. (2004) The bacterial flagellar motor: structure and function of a complex molecular machine. *Int Rev Cytol* **233**: 93-134.
Kotiranta, A., Lounatmaa, K., and Haapasalo, M. (2000) Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes Infect* **2**: 189-198.
Koval, S.F., and Murray, R.G. (1984) The isolation of surface array proteins from bacteria. *Can J Biochem Cell Biol* **62**: 1181-1189.
Koval, S.F., and Hynes, S.H. (1991) Effect of paracrystalline protein surface layers on predation by *Bdellovibrio bacteriovorus*. *J Bacteriol* **173**: 2244-2249.

Kuijper, E.J., Oudbier, J.H., Stuifbergen, W.N., Jansz, A., and Zanen, H.C. (1987) Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. *J Clin Microbiol* **25**: 751-753.

Kuipers, O.P., de Ruyter, P., Kleerebezem, M., and de Vos, W.M. (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *J Biotechnol* **64**: 15-21.

Kumamoto, C.A., and Francetic, O. (1993) Highly selective binding of nascent polypeptides by an *Escherichia coli* chaperone protein in vivo. *J Bacteriol* **175**: 2184-2188.

Kyne, L., Warny, M., Qamar, A., and Kelly, C.P. (2000) Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. *N Engl J Med* **342**: 390-397.

Kyne, L., Warny, M., Qamar, A., and Kelly, C.P. (2001) Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. *Lancet* **357**: 189-193.

Kyne, L., Hamel, M.B., Polavaram, R., and Kelly, C.P. (2002a) Health care costs and mortality associated with nosocomial diarrhea due to *Clostridium difficile*. *Clin Infect Dis* **34**: 346-353.

Kyne, L., Sougioultzis, S., McFarland, L.V., and Kelly, C.P. (2002b) Underlying disease severity as a major risk factor for nosocomial *Clostridium difficile* diarrhea. *Infect Control Hosp Epidemiol* **23**: 653-659.

Lai, E.M., Phadke, N.D., Kachman, M.T., Giorno, R., Vazquez, S., Vazquez, J.A., Maddock, J.R., and Driks, A. (2003) Proteomic analysis of the spore coats of *Bacillus subtilis* and *Bacillus anthracis*. *J Bacteriol* **185**: 1443-1454.

Lantz, M.S., and Ciborowski, P. (1994) Zymographic techniques for detection and characterization of microbial proteases. *Methods Enzymol* **235**: 563-594.

Larson, H.E., Price, A.B., Honour, P., and Borriello, S.P. (1978) *Clostridium difficile* and the aetiology of pseudomembranous colitis. *Lancet* **1**: 1063-1066.

Lebel, S., Bouttier, S., and Lambert, T. (2004) The *cme* gene of *Clostridium difficile* confers multidrug resistance in *Enterococcus faecalis*. *FEMS Microbiol Lett* **238**: 93-100.

Lee, H.C., and Bernstein, H.D. (2001) The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci U S A* **98**: 3471-3476.

Leung, D.Y., Kelly, C.P., Boguniewicz, M., Pothoulakis, C., LaMont, J.T., and Flores, A. (1991) Treatment with intravenously administered gamma globulin of chronic relapsing colitis induced by *Clostridium difficile* toxin. *J Pediatr* **118**: 633-637.

Libby, J.M., and Wilkins, T.D. (1982) Production of antitoxins to two toxins of *Clostridium difficile* and immunological comparison of the toxins by cross-neutralization studies. *Infect Immun* **35**: 374-376.

Liolios, K., Tavernarakis, N., Hugenholtz, P., and Kyrpides, N.C. (2006) The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res* **34**: D332-334.

Lobley, A., Whitmore, L., and Wallace, B.A. (2002) DICHROWEB: an interactive website for the anlaysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **18**: 211-212.

Locher, K.P., Lee, A.T., and Rees, D.C. (2002) The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**: 1091-1098.

Locher, K.P. (2004) Structure and mechanism of ABC transporters. *Curr Opin Struct Biol* **14**: 426-431.

Locher, K.P., and Borths, E. (2004) ABC transporter architecture and mechanism: implications from the crystal structures of BtuCD and BtuF. *FEBS Lett* **564**: 264-268.

Loo, V.G., Poirier, L., Miller, M.A., Oughton, M., Libman, M.D., Michaud, S., Bourgault, A.M., Nguyen, T., Frenette, C., Kelly, M., Vibien, A., Brassard, P., Fenn, S., Dewar, K., Hudson, T.J., Horn, R., Rene, P., Monczak, Y., and Dascal, A. (2005) A Predominantly Clonal Multi-Institutional Outbreak of *Clostridium difficile*-Associated Diarrhea with High Morbidity and Mortality. *N Engl J Med.*

Luirink, J., ten Hagen-Jongman, C.M., van der Weijden, C.C., Oudega, B., High, S., Dobberstein, B., and Kusters, R. (1994) An alternative protein targeting pathway in *Escherichia coli:* studies on the role of FtsY. *Embo J* **13**: 2289-2296.

Lukomski, S., Sreevatsan, S., Amberg, C., Reichardt, W., Woischnik, M., and Podbielski, A. (1997) Inactivation of *Streptococcus pyogenes* extracellular cysteine protease significantly decreases mouse lethality of serotype M3 and M49 strains. *J Clin Invest* **99**: 2574-2580.

Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S., and Baumeister, W. (1994) Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. *J Bacteriol* **176**: 1224-1233.

Mach, L., Mort, J.S., and Glossl, J. (1994) Maturation of human procathepsin B. Proenzyme activation and proteolytic processing of the precursor to the mature proteinase, in vitro, are primarily unimolecular processes. *J Biol Chem* **269**: 13030-13035.

Macnab, R.M. (2003) How bacteria assemble flagella. *Annu Rev Microbiol* **57**: 77-100.

Macnab, R.M. (2004) Type III flagellar protein export and flagellar assembly. *Biochim Biophys Acta* **1694**: 207-217.

Magowan, C., Coppel, R.L., Lao, A.U., Moronne, M.M., Tchernia, G., and Mohandas, N. (1995) Role of the *Plasmodium falciparum* mature-parasite-infected erythrocyte surface antigen (MESA/PfEMP-2) in malarial infection of erythrocytes. *Blood* **86**: 3196-3204.

Mai, U.E., Perez-Perez, G.I., Allen, J.B., Wahl, S.M., Blaser, M.J., and Smith, P.D. (1992) Surface proteins from *Helicobacter pylori* exhibit chemotactic activity for human leukocytes and are present in gastric mucosa. *J Exp Med* **175**: 517-525.

Mani, N., and Dupuy, B. (2001) Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *Proc Natl Acad Sci U S A* **98**: 5844-5849.

Mann, M., Hendrickson, R.C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annual Review of Biochemistry* **70**: 437-473.

Manting, E.H., van Der Does, C., Remigy, H., Engel, A., and Driessen, A.J. (2000) SecYEG assembles into a tetramer to form the active protein translocation channel. *Embo J* **19**: 852-861.

Marquis, R.E. (1968) Salt-induced contraction of bacterial cell walls. *J Bacteriol* **95**: 775-781.

Masuda, K., and Kawata, T. (1980) Reassembly of the regularly arranged subunits in the cell wall of *Lactobacillus brevis* and their reattachment to cell walls. *Microbiol Immunol* **24**: 299-308.

Masuda, K., and Kawata, T. (1985) Reassembly of a regularly arranged protein in the cell wall of *Lactobacillus buchneri* and its reattachment to cell walls: chemical modification studies. *Microbiol Immunol* **29**: 927-938.

Matsuyama, S., Fujita, Y., and Mizushima, S. (1993) SecD is involved in the release of translocated secretory proteins from the cytoplasmic membrane of *Escherichia coli*. *Embo J* **12**: 265-270.

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**: 5510-5515.

McCoubrey, J., and Poxton, I.R. (2001) Variation in the surface layer proteins of *Clostridium difficile*. *FEMS Immunology and Medical Microbiology* **31**: 131-135.

McDonald, L.C., Killgore, G.E., Thompson, A., Owens, R.C., Jr., Kazakova, S.V., Sambol, S.P., Johnson, S., and Gerding, D.N. (2005) An Epidemic, Toxin Gene-Variant Strain of *Clostridium difficile*. *N Engl J Med*.

McEllistrem, M.C., Carman, R.J., Gerding, D.N., Genheimer, C.W., and Zheng, L. (2005) A hospital outbreak of *Clostridium difficile* disease associated with isolates carrying binary toxin genes. *Clin Infect Dis* **40**: 265-272.

McFarland, L.V., Surawicz, C.M., and Stamm, W.E. (1990) Risk factors for *Clostridium difficile* carriage and *C. difficile*-associated diarrhea in a cohort of hospitalized patients. *J Infect Dis* **162**: 678-684.

McFarland, L.V. (2005) Alternative treatments for *Clostridium difficile* disease: what really works? *J Med Microbiol* **54**: 101-111.

McIntyre, G.F., and Erickson, A.H. (1993) The lysosomal proenzyme receptor that binds procathepsin L to microsomal membranes at pH 5 is a 43-kDa integral membrane protein. *Proc Natl Acad Sci U S A* **90**: 10588-10592.

McIntyre, G.F., Godbold, G.D., and Erickson, A.H. (1994) The pH-dependent membrane association of procathepsin L is mediated by a 9-residue sequence within the propeptide. *J Biol Chem* **269**: 567-572.

McNulty, C., Logan, M., Donald, I.P., Ennis, D., Taylor, D., Baldwin, R.N., Bannerjee, M., and Cartwright, K.A. (1997) Successful control of *Clostridium difficile* infection in an elderly care unit through use of a restrictive antibiotic policy. *J Antimicrob Chemother* **40**: 707-711.

Meakins, S.M., Adak, G.K., Lopman, B.A., and O'Brien, S.J. (2003) General outbreaks of infectious intestinal disease (IID) in hospitals, England and Wales, 1992-2000. *Journal of Hospital Infection* **53**: 1-5.

Medline, A., Shin, D.H., and Medline, N.M. (1976) Pseudomembranous colitis associated with antibiotics. *Human Pathology* **7**: 693-703.

Mesnage, S., Tosi-Couture, E., and Fouet, A. (1999a) Production and cell surface anchoring of functional fusions between the SLH motifs of the *Bacillus anthracis* S-layer proteins and the *Bacillus subtilis* levansucrase. *Mol Microbiol* **31**: 927-936.

Mesnage, S., Tosi-Couture, E., Mock, M., and Fouet, A. (1999b) The S-layer homology domain as a means for anchoring heterologous proteins on the cell surface of *Bacillus anthracis*. *J Appl Microbiol* **87**: 256-260.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *EMBO J.* **19**: 4473-4484.

Messner, P., Pum, D., Sara, M., Stetter, K.O., and Sleytr, U.B. (1986) Ultrastructure of the cell envelope of the archaebacteria *Thermoproteus tenax* and *Thermoproteus neutrophilus*. *J Bacteriol* **166**: 1046-1054.

Molloy, M.P., Herbert, B.R., Slade, M.B., Rabilloud, T., Nouwens, A.S., Williams, K.L., and Gooley, A.A. (2000) Proteomic analysis of the *Escherichia coli* outer membrane. *Eur J Biochem* **267**: 2871-2881.

Montigiani, S., Falugi, F., Scarselli, M., Finco, O., Petracca, R., Galli, G., Mariani, M., Manetti, R., Agnusdei, M., Cevenini, R., Donati, M., Nogarotto, R., Norais, N., Garaguso, I., Nuti, S., Saletti, G., Rosa, D., Ratti, G., and Grandi, G. (2002) Genomic approach for analysis of surface proteins in *Chlamydia pneumoniae*. *Infect Immun* **70**: 368-379.

Mori, H., and Ito, K. (2001) The Sec protein-translocation pathway. *Trends Microbiol* **9**: 494-500.

Mukherjee, K., Karlsson, S., Burman, L.G., and Akerlund, T. (2002) Proteins released during high toxin production in Clostridium difficile. *Microbiology* **148**: 2245-2253.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R.R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S.E., Pagni, M., Peyruc, D., Ponting, C.P., Selengut, J.D., Servant, F., Sigrist, C.J., Vaughan, R., and Zdobnov, E.M. (2003) The InterPro Database, 2003 brings increased coverage and new features. *Nucleic Acids Res* **31**: 315-318.

Mullany, P., Wilks, M., Lamb, I., Clayton, C., Wren, B., and Tabaqchali, S. (1990) Genetic analysis of a tetracycline resistance element from *Clostridium difficile* and its conjugal transfer to and from *Bacillus subtilis*. *J Gen Microbiol* **136**: 1343-1349.

Muller, D., Baumeister, W., and Engel, A. (1996) Conformational change of the hexagonally packed intermediate layer of *Deinococcus radiodurans* monitored by atomic force microscopy. *J. Bacteriol.* **178**: 3025-3030.

Murakami, A., Nakatogawa, H., and Ito, K. (2004) Translation arrest of SecM is essential for the basal and regulated expression of SecA. *PNAS* **101**: 12330-12335.

Mylonakis, E., Ryan, E.T., and Calderwood, S.B. (2001) *Clostridium difficile*--Associated diarrhea: A review. *Arch Intern Med* **161**: 525-533.

Nakai, K., and Horton, P. (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem Sci* **24**: 34-36.

Nakatogawa, H., and Ito, K. (2001) Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol Cell* **7**: 185-192.

Nandakumar, R., Nandakumar, M. P., Marten, M. R., Ross, J. M. (2005) Proteome analysis of membrane and cell wall associated proteins from *Staphylococcus aureus*. *J Proteome Res* **4**: 250-7

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**: 174-229.

Neidhardt, F.C., Ingraham, J.L., and Schaechter, M. (1990) *Physiology of the Bacterial Cell*: Sinauer Associates Inc.

Niimi, M., Cannon, R.D., and Monk, B.C. (1999) *Candida albicans* pathogenicity: a proteomic perspective. *Electrophoresis* **20**: 2299-2308.

Nikaido, H., and Hall, J.A. (1998) Overview of bacterial ABC transporters. *Methods Enzymol* **292**: 3-20.

Nomizu, M., Pietrzynski, G., Kato, T., Lachance, P., Menard, R., and Ziomek, E. (2001) Substrate specificity of the streptococcal cysteine protease. *J Biol Chem* **276**: 44551-44556.

O'Brien, J.B., McCabe, M.S., Athie-Morales, V., McDonald, G.S., Ni Eidhin, D.B., and Kelleher, D.P. (2005) Passive immunisation of hamsters against *Clostridium difficile* infection using antibodies to surface layer proteins. *FEMS Microbiol Lett* **246**: 199-205.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**: 4007-4021.

Ottemann, K.M., and Miller, J.F. (1997) Roles for motility in bacterial-host interactions. *Mol Microbiol* **24**: 1109-1117.

Ou, L.T., and Marquis, R.E. (1970) Electromechanical interactions in cell walls of gram-positive cocci. *J Bacteriol* **101**: 92-101.
Paetzel, M., Dalbey, R.E., and Strynadka, N.C. (1998) Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature* **396**: 186-190.
Pallen, M.J., Lam, A.C., Antonio, M., and Dunbar, K. (2001) An embarrassment of sortases - a richness of substrates? *Trends Microbiol* **9**: 97-102.
Panagiotopoulou, I. (2001) Bioinformatic analysis of surface proteins from *Clostridium difficile*. *Msc Thesis*.
Pantosti, A., Cerquetti, M., Viti, F., Ortisi, G., and Mastrantonio, P. (1989) Immunoblot analysis of serum immunoglobulin G response to surface proteins of *Clostridium difficile* in patients with antibiotic-associated diarrhea. *J. Clin. Microbiol.* **27**: 2594-2597.
Pavkov, T., Oberer, M., Egelseer, E.M., Sara, M., Sleytr, U.B., and Keller, W. (2003) Crystallization and preliminary structure determination of the C-terminal truncated domain of the S-layer protein SbsC. *Acta Crystallogr D Biol Crystallogr* **59**: 1466-1468.
Pear, S.M., Williamson, T.H., Bettin, K.M., Gerding, D.N., and Galgiani, J.N. (1994) Decrease in nosocomial *Clostidium difficile*-associated diarrhea by restricting clindamycin use. *Ann Intern Med* **120**: 272-277.
Pechine, S., Gleizes, A., Janoir, C., Gorges-Kergot, R., Barc, M.C., Delmee, M., and Collignon, A. (2005a) Immunological properties of surface proteins of *Clostridium difficile*. *J Med Microbiol* **54**: 193-196.
Pechine, S., Janoir, C., and Collignon, A. (2005b) Variability of *Clostridium difficile* surface proteins and specific serum antibody response in patients with *Clostridium difficile*-associated disease. *J Clin Microbiol* **43**: 5018-5025.
Pei, J., and Grishin, N.V. (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**: 700-712.
Pepin, J., Valiquette, L., and Cossette, B. (2005) Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *Cmaj* **173**: 1037-1042.
Perelle, S., Gibert, M., Bourlioux, P., Corthier, G., and Popoff, M. (1997) Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by *Clostridium difficile* CD196. *Infect Immun* **65**: 1402-1407.
Piard, J.C., Hautefort, I., Fischetti, V.A., Ehrlich, S.D., Fons, M., and Gruss, A. (1997) Cell wall anchoring of the *Streptococcus pyogenes* M6 protein in various lactic acid bacteria. *J Bacteriol* **179**: 3068-3072.
Pitarch, A., Pardo, M., Jimenez, A., Pla, J., Gil, C., Sanchez, M., and Nombela, C. (1999) Two-dimensional gel electrophoresis as analytical tool for identifying *Candida albicans* immunogenic proteins. *Electrophoresis* **20**: 1001-1010.
Pitarch, A., Sanchez, M., Nombela, C., Gil, C. (2002) Sequential fractionation and two-dimensional gel analysis unravels the complexity of the dimorphic fungus *Candida albicans* cell wall proteome. *Mol Cell Proteomics* **1**: 967-82
Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., Jennings, G.T., Baldi, L., Bartolini, E., Capecchi, B., Galeotti, C.L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., Scarselli, M., Storni, E., Zuo, P., Broeker, M., Hundt, E., Knapp, B., Blair, E., Mason, T., Tettelin, H., Hood, D.W., Jeffries, A.C., Saunders, N.J., Granoff, D.M., Venter, J.C., Moxon, E.R., Grandi, G., and Rappuoli, R. (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* **287**: 1816-1820.
Poilane, I., Karjalainen, T., Barc, M.C., Bourlioux, P., and Collignon, A. (1998) Protease activity of *Clostridium difficile* strains. *Can J Microbiol* **44**: 157-161.
Pollmann, K., Raff, J., Schnorpfeil, M., Radeva, G., and Selenska-Pobell, S. (2005) Novel surface layer protein genes in *Bacillus sphaericus* associated with unusual insertion elements. *Microbiology* **151**: 2961-2973.
Poolman, J.T., Kriz-Kuzemenska, P., Ashton, F., Bibb, W., Dankert, J., Demina, A., Froholm, L.O., Hassan-King, M., Jones, D.M., Lind, I., and et al. (1995) Serotypes and subtypes of *Neisseria meningitidis*: results of an international study comparing sensitivities and specificities of monoclonal antibodies. *Clin Diagn Lab Immunol* **2**: 69-72.
Popoff, M.R., Rubin, E.J., Gill, D.M., and Boquet, P. (1988) Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect Immun* **56**: 2299-2306.
Poppelmann, M., Becker, W.M., and Petersen, A. (2002) Combination of zymography and immunodetection to analyze proteins in complex culture supernatants. *Electrophoresis* **23**: 993-997.

Poritz, M.A., Strub, K., and Walter, P. (1988) Human SRP RNA and E. coli 4.5S RNA contain a highly homologous structural domain. *Cell* **55**: 4-6.
Pothoulakis, C., and Lamont, J.T. (2001) Microbes and Microbial Toxins: Paradigms for Microbial- Mucosal Interactions II. The integrated response of the intestine to Clostridium difficile toxins. *Am J Physiol Gastrointest Liver Physiol* **280**: G178-183.
Poxton, I.R. (1979) Serological identification of Bacteroides species by an enzyme-linked immunosorbent assay. *J Clin Pathol* **32**: 294-298.
Poxton, I.R., McCoubrey, J., and Blair, G. (2001) The pathogenicity of *Clostridium difficile*. *Clin Microbiol Infect* **7**: 421-427.
Pragai, Z., Tjalsma, H., Bolhuis, A., van Dijl, J.M., Venema, G., and Bron, S. (1997) The signal peptidase II (*lsp*) gene of *Bacillus subtilis*. *Microbiology* **143 (Pt 4)**: 1327-1333.
Pum, D., Messner, P., and Sleytr, U.B. (1991) Role of the S layer in morphogenesis and cell division of the archaebacterium *Methanocorpusculum sinense*. *J Bacteriol* **173**: 6865-6873.
Purdy, D., O'Keeffe, T.A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N.P. (2002) Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol* **46**: 439-452.
Quiocho, F.A., and Ledvina, P.S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol* **20**: 17-25.
Rabilloud, T., Strub, J.M., Luche, S., van Dorsselaer, A., and Lunardi, J. (2001) A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels. *Proteomics* **1**: 699-704.
Ramamurthy, V., and Oliver, D. (1997) Topology of the integral membrane form of *Escherichia coli* SecA protein reveals multiple periplasmically exposed regions and modulation by ATP binding. *J Biol Chem* **272**: 23239-23246.
Rappuoli, R. (2000) Reverse vaccinology. *Curr Opin Microbiol* **3**: 445-450.
Rawlings, N.D., and Barrett, A.J. (1994) Families of cysteine peptidases. *Methods Enzymol* **244**: 461-486.
Ribes, V., Romisch, K., Giner, A., Dobberstein, B., and Tollervey, D. (1990) E. coli 4.5S RNA is part of a ribonucleoprotein particle that has properties related to signal recognition particle. *Cell* **63**: 591-600.
Rifkin, G.D., Fekety, F.R. and Silva, J.J. (1977) Antibiotic induced colitis: Implications of a toxin neutralised by *Clostridium sordellii* antitoxin. *Lancet* **26**: 1103-1106.
Riley, T.V., Adams, J.E., O'Neill, G.L., and Bowman, R.A. (1991) Gastrointestinal carriage of *Clostridium difficile* in cats and dogs attending veterinary clinics. *Epidemiol Infect* **107**: 659-665.
Robinson, C., and Bolhuis, A. (2004) Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim Biophys Acta* **1694**: 135-147.
Rolfe, R.D., Helebian, S., and Finegold, S.M. (1981) Bacterial interference between *Clostridium difficile* and normal fecal flora. *J Infect Dis* **143**: 470-475.
Ross, B.C., Czajkowski, L., Hocking, D., Margetts, M., Webb, E., Rothel, L., Patterson, M., Agius, C., Camuglia, S., Reynolds, E., Littlejohn, T., Gaeta, B., Ng, A., Kuczek, E.S., Mattick, J.S., Gearing, D., and Barr, I.G. (2001) Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. *Vaccine* **19**: 4135-4142.
Rupnik, M., Avesani, V., Janc, M., von Eichel-Streiber, C., Delmee, M. (1998) A novel toxinotyping scheme and correlation of toxinotypes with serogroups of *Clostridium difficile* isolates. *J Clin Microbiol* **36**: 2240-7
Rupnik, M., Kato, N., Grabnar, M., and Kato, H. (2003) New types of toxin A-negative, toxin B-positive strains among *Clostridium difficile* isolates from Asia. *J Clin Microbiol* **41**: 1118-1125.
Rupnik, M., Pabst, S., von Eichel-Streiber, C., Urlaub, H., and Soling, H.D. (2005) Characterization of the cleavage site and function of resulting cleavage fragments after limited proteolysis of *Clostridium difficile* toxin B (TcdB) by host cells. *Microbiology* **151**: 199-208.
Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944-945.
Sambol, S.P., Merrigan, M.M., Lyerly, D.M., Gerding, D.N., and Johnson, S. (2000) Toxin analysis of a variant strain of *Clostridium difficile* that causes human clinical disease. *Infect Immun* **68**: 5480-5487.

Sambol, S.P., Tang, J.K., Merrigan, M.M., Johnson, S., and Gerding, D.N. (2001) Infection of hamsters with epidemiologically important strains of *Clostridium difficile*. *J Infect Dis* **183**: 1760-1766.
Samra, Z., Talmor, S., and Bahar, J. (2002) High prevalence of toxin A-negative toxin B-positive *Clostridium difficile* in hospitalized patients with gastrointestinal disease. *Diagn. Microbiol. Infect. Dis.* **43**: 189-192.
Sara, M., and Sleytr, U.B. (1987) Molecular sieving through S layers of *Bacillus stearothermophilus* strains. *J Bacteriol* **169**: 4092-4098.
Sargent, F., Bogsch, E.G., Stanley, N.R., Wexler, M., Robinson, C., Berks, B.C., and Palmer, T. (1998) Overlapping functions of components of a bacterial Sec-independent protein export pathway. *Embo J* **17**: 3640-3650.
Savariau-Lacomme, M.P., Lebarbier, C., Karjalainen, T., Collignon, A., and Janoir, C. (2003) Transcription and analysis of polymorphism in a cluster of genes encoding surface-associated proteins of *Clostridium difficile*. *J Bacteriol* **185**: 4461-4470.
Savitsky, A., and Golay, M.J.E. (1964) Smoothing and differentiation of data by simplified least squares procedure. *Analytical Chemistry* **36**: 1627-1639.
Scarselli, M., Giuliani, M.M., Adu-Bobie, J., Pizza, M., and Rappuoli, R. (2005) The impact of genomics on vaccine design. *Trends Biotechnol* **23**: 84-91.
Schaffer, C., Muller, N., Christian, R., Graninger, M., Wugeditsch, T., Scheberl, A., and Messner, P. (1999) Complete glycan structure of the S-layer glycoprotein of *Aneurinibacillus thermoaerophilus* GS4-97. *Glycobiology* **9**: 407-414.
Schaffer, C., Dietrich, K., Unger, B., Scheberl, A., Rainey, F.A., Kahlig, H., and Messner, P. (2000) A novel type of carbohydrate-protein linkage region in the tyrosine-bound S-layer glycan of *Thermoanaerobacterium thermosaccharolyticum* D120-70. *Eur J Biochem* **267**: 5482-5492.
Schaffer, C., and Messner, P. (2001) Glycobiology of surface layer proteins. *Biochimie* **83**: 591-599.
Schaffer, C., and Messner, P. (2005) The structure of secondary cell wall polymers: how Gram-positive bacteria stick their cell walls together. *Microbiology* **151**: 643-651.
Schaumburg, J., Diekmann, O., Hagendorff, P., Bergmann, S., Rohde, M., Hammerschmidt, S., Jansch, L., Wehland, J., and Karst, U. (2004) The cell wall subproteome of *Listeria monocytogenes*. *Proteomics* **4**: 2991-3006.
Schneewind, O., Model, P., and Fischetti, V.A. (1992) Sorting of protein A to the staphylococcal cell wall. *Cell* **70**: 267-281.
Schneewind, O., Mihaylova-Petkov, D., and Model, P. (1993) Cell wall sorting signals in surface proteins of gram-positive bacteria. *Embo J* **12**: 4803-4811.
Schneewind, O., Fowler, A., and Faull, K.F. (1995) Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. *Science* **268**: 103-106.
Schneider, E., and Hunke, S. (1998) ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol Rev* **22**: 1-20.
Schultze-Lam, S., Harauz, G., and Beveridge, T.J. (1992) Participation of a cyanobacterial S layer in fine-grain mineral formation. *J Bacteriol* **174**: 7971-7981.
Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeño-Tárraga, A.M., Wang, H., Holden, M.T.G., Wright, A., Churcher, C., Quail, M., Baker, S., Bason, N., Brooks, K., Chillingworth, T., Cronin, A., Davis, P., Dowd, L., Fraser, A., Feltwell, T., Hance, Z., Holroyd, S., Jagels, K., Moule, S., Mungall, K., Price, C., Rabbinowitsch, E., Sharp, S., Simmonds, M., Stevens, K., Unwin, L., Whitehead, S., Dupuy, B., Dougan, G., Barrell, B., and Parkhill, J. (2006) The multidrug resistant human pathogen *Clostridium difficile* has a highlymobile, mosaic genome. *Nature Genet.* **38**: 779-786
Seddon, S.V., and Borriello, S.P. (1992) Proteolytic activity of *Clostridium difficile*. *J Med Microbiol* **36**: 307-311.
Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996a) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**: 850-858.
Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H., Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acad Sci* **93**: 14440-5
Siegele, D.A., and Kolter, R. (1992) Life after log. *J Bacteriol* **174**: 345-348.
Silhavy, T.J., and Beckwith, J. (1983) Isolation and characterization of mutants of *Escherichia coli* K12 affected in protein localization. *Methods Enzymol* **97**: 11-40.

Sinchaikul, S., Sookkheo, B., Topanuruk, S., Juan, H.F., Phutrakul, S., and Chen, S.T. (2002) Bioinformatics, functional genomics, and proteomics study of *Bacillus* sp. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**: 261-287.

Sjodahl, J. (1977) Repetitive sequences in protein A from *Staphylococcus aureus*. Arrangement of five regions within the protein, four being highly homologous and Fc-binding. *Eur J Biochem* **73**: 343-351.

Skovgaard, M., Jensen, L.J., Brunak, S., Ussery, D., and Krogh, A. (2001) On the total number of genes and their length distribution in complete microbial genomes. *Trends Genet* **17**: 425-428.

Sleytr, U.B., and Glauert, A.M. (1976) Ultrastructure of the cell walls of two closely related clostridia that possess different regular arrays of surface subunits. *J Bacteriol* **126**: 869-882.

Sleytr, U.B. (1978) Regular arrays of macromolecules on bacterial cell walls: structure, chemistry, assembly, and function. *Int Rev Cytol* **53**: 1-62.

Sleytr, U.B., and Sara, M. (1997) Bacterial and archaeal S-layer proteins: structure-function relationships and their biotechnological applications. *Trends Biotechnol* **15**: 20-26.

Sleytr, U.B., and Beveridge, T.J. (1999) Bacterial S-layers. *Trends Microbiol* **7**: 253-260.

Sleytr, U.B., Messner, P., Pum, D. and Sara, M. (1999) Crystalline bacterial cell surface layers (S-layers): From supramolecular cell structures to biomimetics and nanotechnology. *Angew. Chem. Int. Edit.* **38**: 1035-1054.

Smith, T.F., and Waterman, M.S. (1981) Identification of common molecular subsequences. *J Mol Biol* **147**: 195-197.

Soehn, F., Wagenknecht-Wiesner, A., Leukel, P., Kohl, M., Weidmann, M., von Eichel-Streiber, C., and Braun, V. (1998) Genetic rearrangements in the pathogenicity locus of *Clostridium difficile* strain 8864 - implications for transcription, expression and enzymatic acivity of toxins A and B. *Mol Gen Genet* **258**: 222-232.

Sougioultzis, S., Kyne, L., Drudy, D., Keates, S., Maroo, S., Pothoulakis, C., Giannasca, P.J., Lee, C.K., Warny, M., Monath, T.P., and Kelly, C.P. (2005) *Clostridium difficile* toxoid vaccine in recurrent *C. difficile*-associated diarrhea. *Gastroenterology* **128**: 764-770.

Srinivasan, V., Schnitzlein, W.M., and Tripathy, D.N. (2003) A consideration of previously uncharacterized fowl poxvirus unidirectional and bidirectional late promoters for inclusion in homologous recombinant vaccines. *Avian Dis* **47**: 286-295.

Stevanin, T.M., Moir, J.W., and Read, R.C. (2005) Nitric oxide detoxification systems enhance survival of *Neisseria meningitidis* in human macrophages and in nasopharyngeal mucosa. *Infect Immun* **73**: 3322-3329.

Storer, A.C., and Menard, R. (1994) Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol* **244**: 486-500.

Streulens, M.J. (1998) Molecular epidemiologic typing systems of bacterial pathogens:current issues and perspectives. *Memoranda Institut Oswaldo Cruz* **93**: 581-585.

Stryer, L. (1997) *Biochemistry*: W.H.Freeman and Company.

Stubbs, S., Rupnik, M., Gibert, M., Brazier, J., Duerden, B., and Popoff, M. (2000) Production of actin-specific ADP-ribosyltransferase (binary toxin) by strains of *Clostridium difficile*. *FEMS Microbiol Lett* **186**: 307-312.

Stubbs, S.L.J., Brazier, J.S., O'Neill, G.L., and Duerden, B.I. (1999) PCR Targeted to the 16S-23S rRNA Gene Intergenic Spacer Region of *Clostridium difficile* and Construction of a Library Consisting of 116 Different PCR Ribotypes. *J. Clin. Microbiol.* **37**: 461-463.

Suzuki, H., Yonekura, K., Murata, K., Hirai, T., Oosawa, K., and Namba, K. (1998) A structural feature in the central channel of the bacterial flagellar FliF ring complex is implicated in type III protein export. *J Struct Biol* **124**: 104-114.

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

Svergun, D. I., Petoukhov, M. V., Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80**: 2946-53

Svergun, D.I., and Koch, M.H. (2002) Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol* **12**: 654-660.

Takumi, K., Endo, Y., Koga, T., Oka, T., and Natori, Y. (1992) In vitro self-assembly of the S layer subunits from *Clostridium difficile* GAI 0714 into tetragonal arrays. *Tokushima J Exp Med* **39**: 95-100.

Tan, K.S., Wee, B.Y., and Song, K.P. (2001) Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. *J Med Microbiol* **50**: 613-619.

Tao, K., Stearns, N.A., Dong, J., Wu, Q.L., and Sahagian, G.G. (1994) The proregion of cathepsin L is required for proper folding, stability, and ER exit. *Arch Biochem Biophys* **311**: 19-27.

Tasteyre, A., Karjalainen, T., Avesani, V., Delmee, M., Collignon, A., Bourlioux, P., and Barc, M.C. (2000) Phenotypic and genotypic diversity of the flagellin gene (*fliC*) among *Clostridium difficile* isolates from different serogroups. *J Clin Microbiol* **38**: 3179-3186.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H., Karjalainen, T. (2001) Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun* **69**: 7937-7940.

Taylor, C.M., and Roberts, I.S. (2005) Capsular polysaccharides and their role in virulence. *Contrib Microbiol* **12**: 55-66.

Tedesco, F.J., Barton, R.W. and Alpers, D.H. (1974) Clindamycin-associated colitis. *Ann. Intern. Med.* **81**: 429-433.

Teixeira-Gomes, A. P., Cloeckaert, A., Zygmunt, M. S. (2000) Characterization of heat, oxidative, and acid stress responses in *Brucella melitensis*. *Infect Immun* **68**: 2954-61

Thakurta, P.G., Biswas, S., Chakrabarti, C., Sundd, M., Jagannadham, M.V., and Dattagupta, J.K. (2004) Structural basis of the unusual stability and substrate specificity of ervatamin C, a plant cysteine protease from *Ervatamia coronaria*. *Biochemistry* **43**: 1532-1540.

Thelestam, M., and Chaves-Olarte, E. (2000) Cytotoxic effects of the *Clostridium difficile* toxins. *Curr Top Microbiol Immunol* **250**: 85-96.

Tjalsma, H., Kontinen, V.P., Pragai, Z., Wu, H., Meima, R., Venema, G., Bron, S., Sarvas, M., and van Dijl, J.M. (1999) The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of alpha-amylase, a non-lipoprotein. *J Biol Chem* **274**: 1698-1707.

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**: 515-547.

Tomich, M., Herfst, C.A., Golden, J.W., and Mohr, C.D. (2002) Role of flagella in host cell invasion by *Burkholderia cepacia*. *Infect Immun* **70**: 1799-1806.

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**: 22285-22292.

Ton-That, H., and Schneewind, O. (1999) Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. *J Biol Chem* **274**: 24316-24320.

Ton-That, H., Marraffini, L.A., and Schneewind, O. (2004) Protein sorting to the cell wall envelope of Gram-positive bacteria. *Biochim Biophys Acta* **1694**: 269-278.

Torres, J.F., Lyerly, D.M., Hill, J.E., and Monath, T.P. (1995) Evaluation of formalin-inactivated *Clostridium difficile* vaccine administered by parenteral and mucosal routes of immunisation in hamsters. *Infect Immun* **63**: 4619-4627.

Triadafilopoulos, G., Pothoulakis, C., O'Brien, M.J., and LaMont, J.T. (1987) Differential effects of Clostridium difficile toxins A and B on rabbit ileum. *Gastroenterology* **93**: 273-279.

Tu, Z.C., Gaudreau, C., and Blaser, M.J. (2005) Mechanisms underlying *Campylobacter fetus* pathogenesis in humans: surface-layer protein variation in relapsing infections. *J Infect Dis* **191**: 2082-2089.

Valent, Q.A., Scotti, P.A., High, S., de Gier, J.W., von Heijne, G., Lentzen, G., Wintermeyer, W., Oudega, B., and Luirink, J. (1998) The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon. *Embo J* **17**: 2504-2512.

Van Asten, F.J., Hendriks, H.G., Koninkx, J.F., Van der Zeijst, B.A., and Gaastra, W. (2000) Inactivation of the flagellin gene of *Salmonella enterica* serotype enteritidis strongly reduces invasion into differentiated Caco-2 cells. *FEMS Microbiol Lett* **185**: 175-179.

Viscidi, R.P., Yolken, R.H., Laughon, B.E., and Bartlett, J.G. (1983) Enzyme immunoassay for detection of antibody to toxins A and B of *Clostridium difficile*. *J Clin Microbiol* **18**: 242-247.

Volker, U., Engelmann, S., Maul, B., Riethdorf, S., Volker, A., Schmid, R., Mach, H., and Hecker, M. (1994) Analysis of the induction of general stress proteins of *Bacillus subtilis*. *Microbiology* **140 (Pt 4)**: 741-752.

von Eichel-Streiber, C., Laufenberg-Feldmann, R., Sartingen, S., Schulze, J., and Sauerborn, M. (1992) Comparative sequence analysis of the *Clostridium difficile* toxins A and B. *Mol Gen Genet* **233**: 260-268.

Voth, D.E., and Ballard, J.D. (2005) *Clostridium difficile* toxins: mechanism of action and role in disease. *Clin Microbiol Rev* **18**: 247-263.

Vytvytska, O., Nagy, E., Bluggel, M., Meyer, H.E., Kurzbauer, R., Huber, L.A., and Klade, C.S. (2002) Identification of vaccine candidate antigens of *Staphylococcus aureus* by serological proteome analysis. *Proteomics* **2**: 580-590.

Waligora, A.J., Barc, M.C., Bourlioux, P., Collignon, A., and Karjalainen, T. (1999) *Clostridium difficile* cell attachment is modified by environmental factors. *Appl Environ Microbiol* **65**: 4234-4238.

Waligora, A.J., Hennequin, C., Mullany, P., Bourlioux, P., Collignon, A., Karjalainen, T. (2001) Characterization of a cell surface protein of *Clostridium difficile* with adhesive properties. *Infect Immun* **69**: 2144-2153.

Walter, P., and Blobel, G. (1982) Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. *Nature* **299**: 691-698.

Walter, P., and Blobel, G. (1983a) Disassembly and reconstitution of signal recognition particle. *Cell* **34**: 525-533.

Walter, P., and Blobel, G. (1983b) Signal recognition particle: a ribonucleoprotein required for cotranslational translocation of proteins, isolation and properties. *Methods Enzymol* **96**: 682-691.

Ward, S.J., Douce, G., Dougan, G., and Wren, B.W. (1999a) Local and Systemic Neutralizing Antibody Responses Induced by Intranasal Immunization with the Nontoxic Binding Domain of Toxin A from *Clostridium difficile. Infect. Immun.* **67**: 5124-5132.

Ward, S.J., Douce, G., Figueiredo, D., Dougan, G., and Wren, B.W. (1999b) Immunogenicity of a *Salmonella typhimurium aroA aroD* vaccine expressing a nontoxic domain of *Clostridium difficile* toxin A. *Infect Immun* **67**: 2145-2152.

Warny, M., Vaerman, J.P., Avesani, V., and Delmee, M. (1994) Human antibody response to *Clostridium difficile* toxin A in relation to clinical course of infection. *Infect Immun* **62**: 384-389.

Warny, M., Pepin, J., Fang, A., Killgore, G., Thompson, A., Brazier, J., Frost, E., and McDonald, L.C. (2005) Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* **366**: 1079-1084.

Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humphery-Smith, I. (1995) Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium. Electrophoresis* **16**: 1090-1094.

Weygand, M., Wetzer, B., Pum, D., Sleytr, U.B., Cuvillier, N., Kjaer, K., Howes, P.B., and Losche, M. (1999) Bacterial S-layer protein coupling to lipids: x-ray reflectivity and grazing incidence diffraction studies. *Biophys J* **76**: 458-468.

Whittier, S., Shapiro, D.S., Kelly, W.F., Walden, T.P., Wait, K.J., McMillon, L.T., and Gilligan, P.H. (1993) Evaluation of four commercially available enzyme immunoassays for laboratory diagnosis of *Clostridium difficile*-associated diseases. *J Clin Microbiol* **31**: 2861-2865.

Wilcox, M.H., and Smyth, E.T. (1998) Incidence and impact of *Clostridium difficile* infection in the UK, 1993-1996. *J Hosp Infect* **39**: 181-187.

Wilcox, M.H. (2004) Descriptive study of intraveneous immunoglobulin for the treatment of recurrent *Clostridium difficile* diarrhoea. *Journal of Antimicrobial Chemotherapy*.

Wilcox, M.H., Cunniffe, J.G., Trundle, C., Redpath, C. (1996) Financial burden of hospital-acquired Clostridium difficile infection. *J Hosp Infect* **34**: 23-30.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., Williams, K.L., and Hochstrasser, D.F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**: 61-65.

Wizemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E., Gayle, A., Brewah, Y.A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S., and Koenig, S. (2001) Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect Immun* **69**: 1593-1598.

Wren, B.W., and Tabaqchali, S. (1987) Restriction endonuclease DNA analysis of *Clostridium difficile. J Clin Microbiol* **25**: 2402-2404.

Wren, B.W., Russell, R.R., and Tabaqchali, S. (1991) Antigenic cross-reactivity and functional inhibition by antibodies to *Clostridium difficile* toxin A, *Streptococcus mutans* glucan-binding protein, and a synthetic peptide. *Infect Immun* **59**: 3151-3155.

Wright, A., Wait, R., Begum, S., Crossett, B., Nagy, J., Brown, K., and Fairweather, N. (2005) Proteomic analysis of cell surface proteins from *Clostridium difficile. Proteomics* **5**: 2443-2452.

Wright, C., Willan, K.J., Sjodahl, J., Burton, D.R., and Dwek, R.A. (1977) The interaction of protein A and Fc fragment of rabbit immunoglobulin G as probed by complement-fixation and nuclear-magnetic-resonance studies. *Biochem J* **167**: 661-668.

Wust, J., Sullivan, N.M., Hardegger, U., and Wilkins, T.D. (1982) Investigation of an outbreak of antibiotic-associated colitis by various typing methods. *J Clin Microbiol* **16**: 1096-1101.
Wust, J., and Hardegger, U. (1983) Transferable resistance to clindamycin, erythromycin and tetracycline in *Clostridium difficile*. *Antimicrob Agents Chemother* **23**: 784-786.
Yarwood, J.M., and Schlievert, P.M. (2003) Quorum sensing in Staphylococcus infections. *J. Clin. Invest.* **112**: 1620-1625.
Yucesoy, M., McCoubrey, J., Brown, R., and Poxton, I.R. (2002) Detection of toxin production in *Clostridium difficile* strains by three different methods. *Clin Microbiol Infect* **8**: 413-418.
Zanen, G., Houben, E. N., Meima, R., Tjalsma, H., Jongbloed, J. D., Westers, H., Oudega, B., Luirink, J., van Dijl, J. M., Quax, W. J. (2005) Signal peptide hydrophobicity is critical for early stages in protein export by *Bacillus subtilis*. *Febs J* **272**: 4617-30
Zhang, H.Z., Hackbarth, C.J., Chansky, K.M., and Chambers, H.F. (2001) A proteolytic transmembrane signalling pathway and resistance to beta-lactams in staphylococci. *Science* **291**: 1962-1965.
Zhu, H., Bilgin, M., and Snyder, M. (2003) Proteomics. *Annu Rev Biochem* **72**: 783-812.