A Proteomics Approach to Investigate Uropathogenic

*Escherichia coli*

by

Nizar Hamed E. Saeedi

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Biomedical Sciences

April 2019
This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author’s prior consent.
Acknowledgements

No one walks alone on the journey of life. Just where do you start to thank those that joined you, walked beside you, and helped you along the way.

I would like to show my greatest appreciation to Professor Mathew Upton for his guidance and constant supervision. I can't say thank you enough for his tremendous support and help. Without his encouragement and guidance, this project would not have been completed.

My thanks and appreciations also go to Dr Vikram Sharma for his great help with the proteomics assays and data analysis; and Dr Majed Alghoribi for his help with the genomic data.

To Professor Simon Jackson, Upton’s lab members, and the laboratory staff at the University of Plymouth, who’s help and advice was indispensable throughout my studies; I would like to extend my sincerest thanks. I am highly indebted to my sponsor, University of Tabuk, for sponsoring my study.

I would like to express my heartfelt thanks to my beloved parents, family and friends for their blessings and wishes for the successful completion of my studies.

Finally, yet importantly, I am forever indebted to my wife Zainab for her understanding, endless patience and encouragement when it was most needed. I am also grateful to my kids Ghena and Majed for their support.
Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship from University of Tabuk.

Presentations at conferences:

- **Title of poster:** Proteomic Analysis of ST127 Uropathogenic *E. coli* Reveals Distinct Secretomes.
  Title of conference attended: Microbiology Society Annual Conference 2017.
  Date attended / Location: 03 - 06 April 2017 / Edinburgh, UK.

- **Title of poster:** Proteomic Analysis of Uropathogenic *Escherichia coli* Sequence Type 127.
  Title of conference attended: 27th European Congress of Clinical Microbiology and Infectious Diseases.
  Date attended / Location: 22 - 25 April 2017 / Vienna, Austria.

- **Title of oral presentation:** Metabolic Adaptation of *E. coli* Causing Bacteraemia that Originates in Urinary Tract Infection.
  Title of conference attended: ASM Microbe 2017.
  Date attended / Location: 01 - 05 June 2017 / New Orleans, USA.

- **Title of poster:** Metabolic Adaptation in *Escherichia coli* Isolates During Transition from UTI to Bloodstream Infection.
  Title of conference attended: Microbiology Society Annual Conference 2018.
  Date attended / Location: 10 - 13 April 2018 / Birmingham, UK.
• Title of poster: Metabolic Adaptation of Uropathogenic *Escherichia coli* Strains in Causing Bacteraemia.
  Title of conference attended: 28th European Congress of Clinical Microbiology and Infectious Diseases.
  Date attended / Location: 21 - 24 April 2018 / Madrid, Spain.

• Title of poster: Understanding the Pathogenic Process of Uropathogenic *Escherichia coli* ST127 using Proteomics on Uroepithelial Co-culture Samples.
  Title of conference attended: Microbiology Society Annual Conference 2019.
  Date attended / Location: 08 - 11 April 2019 / Belfast, UK.

Word count of main body of thesis: 35,719

Signed … Nizar Hamed E. Saeedi …

Date ……08/04/2019…………
Abstract

Nizar Hamed E. Saeedi

A Proteomics Approach to Investigate Uropathogenic Escherichia coli

Uropathogenic Escherichia coli (UPEC) is the most commonly known cause of urinary tract infection (UTI). The possession of many virulence factors and global dissemination of multi-drug resistant UPEC strains is posing a critical risk to treatment worldwide. The most virulent UPEC strains have descended from major lineages such as sequence type (ST) 131, ST38 and ST127. Strains of ST127 exhibit the highest virulence potential of most UPEC lineages, but little is known about their pathogenicity. Previous studies have reported the important role of metabolic adaptations in UPEC virulence. This study investigated the metabolic adaptations of UPEC isolates, with a focus on ST127, using a quantitative proteomics approach.

We employed a classical shotgun proteomics approach to analyse proteins extracted from multiple strains of UPEC following growth in various environments (Lysogeny broth, artificial urine medium and co-culture with uroepithelial cells). The digested proteins and peptides from all fractions were separated on a Dionex Ultimate 3000 RSLC nano flow system and analysed in an Orbitrap Velos Pro FTMS. Data were processed using Perseus software.

Expression and Gene Ontology Enrichment analysis revealed different proteomic profiles of UPEC ST127 strains cultured in LB and AUM. This study showed that environmental changes have an effect on the metabolic pathways expressed by a specific strain. These conditions can be engineered to simulate the variety of environments UPEC isolates may need to survive in. This methodology was then employed to enhance our understanding of the pathogenesis of UPEC.

Building on previous work with UPEC paired isolates (urine and blood), proteomics analysis revealed that members of some STs have a higher metabolic potential that could provide virulence advantages. The study also revealed minimal proteomic variations among the paired isolates for some STs.

The methods were then applied during co-culture experiments with UPEC and HT1197 bladder epithelial cells. UPEC strain SA189 caused exfoliation effects in HT1197 not seen with other UPEC strains. Analysis of the UPEC SA189 secretome revealed highly abundant bacterial proteins, including hemolysin A.

Proteomic analysis is helping to understand the pathogenic potential of UPEC and may facilitate identification of novel diagnostic or therapeutic targets to reduce UTI and potential subsequent bacteraemia.
Chapter One: General Introduction and Literature Review

1.1 *Escherichia coli*: the microorganism

1.1.1 General and cultural features

1.1.2 Serological types of *Escherichia coli*

1.1.3 Phylogenetic organisation of *E. coli*

1.1.4 Pathotypes of *Escherichia coli*

1.2 Urinary tract infection

1.2.1 Epidemiology of UTI

1.2.2 Pathogenicity of UTI

1.2.2.1 The ascending pathway for development of UTI

1.2.2.2 Haematogenous pathway of developing UTI

1.2.2.3 The lymphatic pathway

1.3 Uropathogenic *Escherichia coli*

1.3.1 Molecular epidemiology of UPEC

1.3.2 Virulence factors of UPEC

1.3.3 UPEC survival in the urinary tract

1.3.3.1 Metabolic adaptations of UPEC

1.3.4 Antimicrobial resistance in UPEC

1.4 Proteomics: an investigative methodology

1.4.1 Technologies for investigating the proteome

1.4.1.1 Mass spectrometry proteomics

1.4.2 Proteomic approaches to investigate UPEC pathogenesis
1.5 Aims of this study ........................................................................................................ 59

Chapter Two: Materials and Methods ................................................................. 61
2.1 Bacterial strains and culture conditions ......................................................... 62
  2.1.1 Artificial urine medium .................................................................................. 62
2.2 Growth kinetics .................................................................................................... 65
2.3 Lipopolysaccharide extraction and silver staining ........................................ 66
2.4 Tissue culture ....................................................................................................... 67
  2.4.1 Maintaining epithelial cell culture ................................................................. 67
  2.4.2 Co-culture conditions ..................................................................................... 67
  2.4.3 Cells viability and MTT assay ........................................................................ 69
2.5 Protein extraction ................................................................................................. 69
  2.5.1 Soluble protein extraction ............................................................................. 70
  2.5.2 Extracting insoluble proteins from the cell pellet ......................................... 70
  2.5.3 Epithelial cells protein extraction ................................................................. 71
  2.5.4 Protein precipitation for secretome analysis ............................................... 71
  2.5.5 Protein estimation ......................................................................................... 72
  2.5.6 SDS gel electrophoresis and Coomassie Brilliant blue stain ...................... 72
2.6 Proteomics ............................................................................................................ 73
  2.6.1 Sample preparation ....................................................................................... 73
  2.6.2 Mass spectrometry analysis ........................................................................... 75
  2.6.3 Proteomics analytical approach ..................................................................... 78
    2.6.3.1 Perseus software ...................................................................................... 78
    2.6.3.2 Microsoft Excel software ......................................................................... 81
    2.6.3.3 Venn diagrams ....................................................................................... 82
    2.6.3.4 GraphPad Prism ...................................................................................... 82
    2.6.3.5 GProX software ...................................................................................... 83
    2.6.3.6 STRING database ................................................................................. 83
    2.6.3.7 DAVID tool ............................................................................................ 84
    2.6.3.8 Proteomics data analysis pipeline ........................................................... 85

8
Chapter Three: Investigating the Effect of Growth in Different Environments on Phenotypic and Proteomics Profiles of UPEC

3.1 Introduction

3.1.1 Aim and objectives

3.2 Experimental design

3.3 Results

3.3.1 Growth kinetics

3.3.2 LPS detection and analysis

3.3.3 Protein extractions

3.3.3.1 Soluble proteins

3.3.3.2 Insoluble proteins

3.3.4 Mass spectrometry and data analysis

3.3.4.1 Expression fold change analysis

3.3.4.2 Enrichment analysis

3.3.5 Genomic analysis of UPEC ST127 strains

3.4 Discussion

3.4.1 Growth conditions

3.4.2 Investigating LPS structure

3.4.3 Protein extraction

3.4.4 Analysis and interpretation of proteomics data

3.4.5 Conclusion

Chapter Four: Metabolic Adaptation in Escherichia coli Isolates During Transition from UTI to Bloodstream Infection

4.1 Introduction

4.1.1 Aim and objectives

4.2 Experimental design

4.3 Results
4.3.1 Protein extractions .............................................................. 130
4.3.2 Overview analysis of the proteome profiles ................................. 130
4.3.3 UPEC STs strain analysis ...................................................... 135
  4.3.3.1 Perseus analysis of all UPEC strains .................................... 135
  4.3.3.2 Protein-Protein interactions network of all UPEC strains .......... 138
4.3.4 UPEC ST127 strains analysis ............................................... 145
  4.3.4.1 Perseus analysis of UPEC ST127 strains ............................... 145
  4.3.4.2 Protein-Protein interactions network of UPEC ST127 strains ....... 148
4.3.5 Genomic analysis of paired UPEC isolates .................................. 159

4.4 Discussion .................................................................................. 166
  4.4.1 Growth conditions and protein extractions ................................. 166
  4.4.2 Overview of the proteomic profile analysis .................................. 167
  4.4.3 Paired UPEC metabolic pathways .......................................... 168
  4.4.4 ST127 paired isolates showed exclusive proteome profile .......... 171
  4.4.5 Variations in the genomes of paired isolates ............................... 174
  4.4.6 Conclusion ........................................................................... 176

Chapter Five: Proteomic Analysis of Uropathogenic Escherichia coli
ST127 During Co-Culture with Uroepithelial Cells .................. 178

5.1 Introduction ................................................................................... 179
  5.1.1 Aim and objectives .................................................................. 180

5.2 Experimental design ..................................................................... 181

5.3 Results .......................................................................................... 183
  5.3.1 Co-culture experiment .............................................................. 183
    5.3.1.1 Uropathogenic E. coli SA189 effect on HT1197 cells ............... 183
  5.3.2 Protein extractions ................................................................. 186
  5.3.3 Analysis of the bacterial proteome profiles during co-culture with HT1197 cells 189
  5.3.4 Analysis of the secretome profiles from UPEC and HT1197 cells during co-culture experiments .......................................................... 191
5.3.5 Analysis of the epithelial proteome profiles during co-culture experiments with UPEC and HT1197 cells ................................................................. 199

5.4 Discussion ............................................................................................................. 204

5.4.1 Co-culture conditions and proteins extractions .............................................. 204

5.4.2 Overview of the proteomic profiles' analysis ................................................. 206

5.4.2.1 Bacterial proteomic changes during the co-culture experiments .......... 207

5.4.2.2 Secreted proteins observed during the co-culture experiments with UPEC and K12 *E. coli* with HT1197 cells ................................................................. 211

5.4.2.3 Epithelial proteome determination during the co-culture experiments with *E coli* and HT1197 cells ........................................................................... 213

5.4.3 Conclusion ........................................................................................................ 215

Chapter Six: Final Discussion and Concluding Remarks .................. 217

6.1 Is shotgun proteomics an effective tool for investigating the bacterial proteome? ........................................................................................................... 218

6.2 Are there proteomic variations between paired isolates of UPEC from urosepsis and if so, what are they? ................................................................. 219

6.3 What metabolic adaptations of UPEC took place during infection of the epithelial cells? ......................................................................................... 220

6.4 Future directions .................................................................................................. 221

Appendices ............................................................................................................... 224

References ............................................................................................................... 237
List of Figures

Figure 1.1: Generalised structure of a Gram-negative rod shaped bacterium.. 25
Figure 1.2: Development by the ascending pathway of urinary tract infection caused by uropathogens ................................................................. 33
Figure 1.3: *E. coli* CFT073 complete genome presenting the seven housekeeping gene targets, according to the Achtman MLST protocol .... 36
Figure 1.4: Summary of the main metabolic adaptations of UPEC during urinary tract infection (UTI) .................................................................................. 45
Figure 1.5: An example of a modern mass spectrometry proteomics workflow.. ......................................................................................... 53
Figure 1.6: Peptide sequencing using the tandem mass spectrometry (MS/MS) ......................................................................................... 55
Figure 2.1: General work flow chart of the methodology followed for proteomics studies and mass spectrometry analysis .............................................. 77
Figure 2.2: The Perseus data analysis platform ........................................ 79
Figure 2.3: Outline summary of the Perseus software workflow ............. 81
Figure 2.4: General pipeline chart of the methodology followed for proteomics data analysis ................................................................. 85
Figure 3.1: Experimental design of the differential proteomics study .......... 91
Figure 3.2: Growth kinetics of *E. coli* EC18, EC41 and SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM) ....................................................... 93
Figure 3.3: Colony forming unit (log10) of *E. coli* EC18 and the control strains *E. coli* EC41 and *E. coli* SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM) ........................................................................ 94
Figure 3.4: Silver stained SDS-PAGE analysis of LPS extractions for *E. coli* EC18, *E. coli* EC41, and *E. coli* SA189 .................................................. 95
Figure 3.5: BCA estimation for soluble protein extracted from *E. coli* EC41 in artificial urine medium (AUM) ................................................................. 97
Figure 3.6: BCA estimation for protein yield extracted from *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM).................................................................................................................. 97

Figure 3.7: Protein extractions for *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 loaded on SDS PAGE gels ................................................................................................................................................... 98

Figure 3.8: Correlation matrix of label free quantification (LFQ) of biological replicates of UPEC ST127 isolates (EC18, EC41, and SA189) grown in LB and AUM.............................................................................................................................................. 99

Figure 3.9: Overlaps between soluble and insoluble proteins in UPEC ST127 strains ........................................................................................................................................................................................................ 101

Figure 3.10: Hierarchical clustering (HCL) of z normalized label-free quantification (LFQ) Intensities for EC18, EC41, and SA189 ........................................................................................................ 102

Figure 3.11: Row clusters (n=6) of HCL of label-free protein quantification (LFQ) intensities for ST127 strains (EC18, EC41, and SA189) in LB and AUM... ........................................................................................................................................ 103

Figure 3.12: Overlaps between differentially altered proteins in three strains of ST127 ........................................................................................................................................................................................................ 105

Figure 3.13: Pathways enriched in proteins up-regulated by ≥2 fold during growth in Lysogeny Broth compared to artificial urine medium ............... 107

Figure 3.14: Pathways enriched in proteins down-regulated by ≥2 fold during growth in Lysogeny Broth compared to artificial urine medium ............... 109

Figure 3.15: BLAST ring image of UPEC isolates of ST127 strains EC18, EC41, and SA189 ................................................................................................................................................ 112

Figure 3.16: BLASTn comparison of O-antigen gene cluster (Green) and colanic acid gene cluster (Red) of UPEC EC18, EC41, SA189 and UPEC ST127 reference strain 536 .................................................................................................................... 113

Figure 3.17: General chemical structure of LPS from Gram-negative enterobacteria ...................................................................................................................................................... 116

Figure 3.18: Glycolysis / Gluconeogenesis pathway of *E. coli* O18:K1:H7 UTI89 (UPEC)........................................................................................................................................................................ 121

Figure 4.1: Experimental design of the paired isolates study .................. 129
Figure 4.2: BCA estimation for protein yield extracted from paired isolates (Urine and Blood) of UPEC strains cultured in in Lysogeny Broth (LB)...

Figure 4.3: Samples of protein extractions for *E. coli* EC18 (Blood and Urine), and *E. coli* EC41 (Blood) loaded on SDS PAGE gels.

Figure 4.4: Overlaps between differentially altered proteins in four STs showing the up-regulated and down-regulated proteins in blood vs. urine isolates.

Figure 4.5: Hierarchical clustering (HCL) of relative expression data of LFQ ratios for paired isolates of ST131 (EC1 and EC10), ST127 (EC18 and EC41), ST73 (EC22 and EC73), and ST38 (EC47 and EC86).

Figure 4.6: Row clusters (n=10) of HCL of relative expression data of paired isolates of ST131 (EC1 and EC10), ST127 (EC18 and EC41), ST73 (EC22 and EC73) and ST38 (EC47 and EC86).

Figure 4.7: Interactions network enrichment analysis of cluster no. 1152 (all UPEC entries).

Figure 4.8: Interactions network enrichment analysis of cluster no. 1148 (all UPEC entries).

Figure 4.9: Interactions network enrichment analysis of cluster no. 1146 (all UPEC entries).

Figure 4.10: Extended interactions network enrichment analysis of cluster no. 1132 (all UPEC entries).

Figure 4.11: Interactions network enrichment analysis of cluster no. 1120 (all UPEC entries).

Figure 4.12: Interactions network enrichment analysis of cluster no. 995 (all UPEC entries).

Figure 4.13: Hierarchical clustering (HCL) of relative expression data of label-free protein quantification (LFQ) ratios for paired isolates of ST127 (EC18 and EC41).

Figure 4.14: Row clusters (n=10) of HCL of relative expression data of paired isolates of ST127 (EC18 and EC41).

Figure 4.15: Interactions network enrichment analysis of cluster no. 1153.
Figure 4.16: Interactions network enrichment analysis of cluster no. 1152 .... 151
Figure 4.17: Interactions network enrichment analysis of cluster no. 1151 .... 152
Figure 4.18: Interactions network enrichment analysis of cluster no. 1137 .... 152
Figure 4.19: Interactions network enrichment analysis of cluster no. 1150 .... 153
Figure 4.20: Interactions network enrichment analysis of cluster no. 1148 .... 154
Figure 4.21: Interactions network enrichment analysis of cluster no. 1147 .... 155
Figure 4.22: Interactions network enrichment analysis of cluster no. 1146 .... 156
Figure 4.23: Interactions network enrichment analysis of cluster no. 1139 .... 157
Figure 4.24: Interactions network enrichment analysis of cluster no. 1113 .... 158
Figure 4.25: BLAST ring image of paired isolates (blood and urine) of ST131 strains (EC1 and EC10)......................................................................................... 160
Figure 4.26: BLAST ring image of paired isolates (blood and urine) of ST127 strains (EC18 and EC41)......................................................................................... 161
Figure 4.27: BLAST ring image of paired isolates (blood and urine) of ST73 strains (EC22 and EC73)......................................................................................... 162
Figure 4.28: BLAST ring image of paired isolates (blood and urine) of ST38 strains (EC47 and EC86)......................................................................................... 163
Figure 4.29: Schematic diagram illustrating the specific cascade of Galactose metabolism.............................................................................................................. 169
Figure 5.1: Experimental design of the co-culture analysis of growth of ST127 UPEC strains and E. coli K12 with HT1197 uroepithelial cells ............. 181
Figure 5.2: General work flow chart of the methodology followed for the co-culture proteomics studies to isolate bacterial proteins, epithelial proteins, and secretome. ........................................................................................................ 182
Figure 5.3: HT1197 uroepithelial cells co-cultured with UPEC ST127 strains and E. coli K12 ........................................................................................................ 184
Figure 5.4: MTT assay for the co-cultured UPEC SA189 with HT1197. ....... 185
Figure 5.5: BCA estimation for bacterial protein yield extracted from *E. coli* K12, *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 in co-culture conditions and MEM only control ........................................................................................................... 187

Figure 5.6: BCA estimation for protein yield extracted from HT1197 uroepithelial cells co-cultured with *E. coli* K12, *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 .................................................................................................................. 187

Figure 5.7: Protein extractions for HT1197 cells co-cultured with *E. coli* EC18, *E. coli* EC41, *E. coli* SA189 and *E. coli* K12 loaded on SDS PAGE gels........................................................................................................ 188

Figure 5.8: Overlaps between identified secretomes in all strains of UPEC ST127 and *E. coli* K12 in co-culture and control conditions ............ 193

Figure 5.9: Overlaps between identified unique secretomes in co-culture conditions of all strains of UPEC ST127 and *E. coli* K12 .................. 194

Figure 5.10: Pathways enriched in collections of proteins that were identified as part of the secretome under various co-culture conditions with UPEC ST127 strains and HT1197 cells, but were not present in bacterial secretome databases ......................................................................................... 198

Figure 5.11: Hierarchical clustering (HCL) following label-free quantification (LFQ) of proteins from HT1197 cells (CLL) co-cultured with *E. coli* K12, EC18, EC41, SA189, compared to a control of culture (CON) in MEM alone .................................................................................................................. 200

Figure 5.12: Overlaps between quantified epithelial proteins of HT1197 cells co-cultured with UPEC ST127 strains and *E. coli* K12 ............ 201

Figure 5.13: LFQ intensity of paxillin protein extracted from uroepithelial HT1197 cells co-cultured with *E. coli* K12, *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 .................................................................................................................. 203

Figure 5.14: Analysis of HlyA protein using the PSORTb tool .................. 208

Figure 5.15: Acquisition of amino acids and roles of gluconeogenesis and the TCA cycle in fitness of UPEC .................................................................................................................. 210

Figure 5.16: Probability analysis of HlyA protein on SignalP 5.0 tool ........ 212
List of Tables

Table 1.1: Most common virulence factors of *E. coli* involved in urinary tract infection.................................................................................................................. 40

Table 1.2: Summary of metabolic functions in uropathogenic *E. coli* (UPEC) that represent potential targets for novel antimicrobial approaches ............ 47

Table 2.1: *Escherichia coli* strains used in this study........................................ 63

Table 2.2: Artificial urine medium composition.................................................... 64

Table 2.3: Sample preparation for SDS PAGE gel electrophoresis...................... 73

Table 3.1: Row clusters (n=6) of HCL of label-free protein quantification (LFQ) intensities for ST127 strains (EC18, EC41, and SA189) in LB and AUM. .................................................................................................................. 103

Table 3.2: Up-regulated and down-regulated proteins (>2 fold) in all strains of ST127 ............................................................................................................ 104

Table 3.3: Pathways enriched in up-regulated proteins (≥2 fold with *p*<0.05) in LB vs. AUM ........................................................................................................... 108

Table 3.4: Pathways enriched in down-regulated proteins (≥2 fold with *p*<0.05) in LB vs. AUM ........................................................................................................... 110

Table 4.1: Metabolic activity profiles of paired blood and urine *E. coli* isolates.......................................................................................................................... 126

Table 4.2: Antibiotics susceptibility profiles of paired blood and urine *E. coli* isolates................................................................................................................. 127

Table 4.3: Up-regulated and down-regulated proteins (>1.5 fold) in all four STs ....................................................................................................................... 133

Table 4.4: Row clusters (n=10) of HCL of relative expression data of label-free protein quantification (LFQ) ratios for paired isolates of ST131 (EC1 and EC10), ST127 (EC18 and EC41), ST73 (EC22 and EC73), and ST38 (EC47 and EC86) ........................................................................................................... 136

Table 4.5: Row clusters (n=10) of HCL of relative expression data of label-free protein quantification (LFQ) ratios for paired uro-sepsis isolates of ST127 (EC18 and EC41) ........................................................................................................... 147
Table 4.6: MLST, Serotyping and Fim-typing of each strain examined from the four UPEC lineages ................................................................. 164

Table 4.7: Summarised genomic analysis comparing the urine and blood isolates of each strain of UPEC STs............................................................... 165

Table 5.1: Total proteome profiles of all strains of UPEC ST127 and E. coli K12......................................................................................... 189

Table 5.2: Pathways up-regulated in the bacterial proteome of all strains of UPEC ST127 and E. coli K12 ................................................................. 190

Table 5.3: Pathways down-regulated in the bacterial proteome of all strains of UPEC ST127 and E. coli K12 ................................................................. 191

Table 5.4: Total secretome profiles of HTT1197 cells and all strains of UPEC ST127 and E. coli K12 in co-culture and control conditions .......... 192

Table 5.5: Bacterial and epithelial proteins identified in the secretomes during co-culture of all strains of UPEC ST127 with HT1197 cells. .............. 196

Table 5.6: Details of bacterial secreted proteins from various co-culture conditions of all strains of UPEC ST127 with HT1197 cells ................. 197

Table 5.7: Total proteome profiles of all epithelial cells co-cultured with UPEC ST127 strains and E. coli K12 ................................................................. 199

Table 5.8: Differentially regulated epithelial proteins of HT1197 co-cultured with UPEC ST127 strains and E. coli K12. ..................................................... 202
List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>Amino acid</td>
</tr>
<tr>
<td>ABC</td>
<td>Ammonium bicarbonate</td>
</tr>
<tr>
<td>AC</td>
<td>Accession number</td>
</tr>
<tr>
<td>AMR</td>
<td>Antimicrobial resistance</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATCC</td>
<td>American Type Culture Collection</td>
</tr>
<tr>
<td>AUM</td>
<td>Artificial urine media</td>
</tr>
<tr>
<td>B-PER</td>
<td>Bacterial Protein Extraction Reagent</td>
</tr>
<tr>
<td>BCA</td>
<td>Bicinchoninic acid</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BRIG</td>
<td>BLAST Ring Image Generator</td>
</tr>
<tr>
<td>CAA</td>
<td>Chloroacetamide</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony-forming unit</td>
</tr>
<tr>
<td>DF</td>
<td>Dilution factor</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide, C$_2$H$_6$OS</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPBS</td>
<td>Dulbecco's Phosphate Buffered Saline</td>
</tr>
<tr>
<td>DTT</td>
<td>1,4-Dithiothreitol</td>
</tr>
<tr>
<td>EC</td>
<td><em>E. coli</em></td>
</tr>
<tr>
<td>ECACC</td>
<td>European Collection of Authenticated Cell Cultures</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FASP</td>
<td>Filter aided sample preparation method</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal bovine serum</td>
</tr>
<tr>
<td>GML</td>
<td><em>Galleria mellonella</em> larvae</td>
</tr>
<tr>
<td>HCL</td>
<td>Hierarchical clustering</td>
</tr>
<tr>
<td>IS</td>
<td>Insertion sequence</td>
</tr>
<tr>
<td>LB</td>
<td>Lysogeny broth</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid chromatography–mass spectrometry</td>
</tr>
<tr>
<td>LFQ</td>
<td>Label free quantification</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharide</td>
</tr>
<tr>
<td>Lys-c</td>
<td>Endoproteinase Lys-c</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MEM</td>
<td>Minimum Essential Medium</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>MRM</td>
<td>Multiple reaction monitoring</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MTT</td>
<td>Thiazolyl Blue Tetrazolium Blue</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NEAA</td>
<td>Non-essential Amino Acid</td>
</tr>
<tr>
<td>OD</td>
<td>Optical density</td>
</tr>
<tr>
<td>p-</td>
<td>Probability value</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PTMs</td>
<td>Post-translational modifications</td>
</tr>
<tr>
<td>PTS</td>
<td>phosphotransferase system</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>SA</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>SCV</td>
<td>Small colony variant</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SDS PAGE</td>
<td>Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>SDT</td>
<td>SDS dithiothreitol Tris/HCl</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard error of the mean</td>
</tr>
<tr>
<td>ST</td>
<td>Sequence type</td>
</tr>
<tr>
<td>STRING</td>
<td>Search Tool for the Retrieval of Interacting Genes/Proteins</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic acid</td>
</tr>
<tr>
<td>TFA</td>
<td>Trifluoroacetic acid</td>
</tr>
<tr>
<td>UA</td>
<td>Uric acid</td>
</tr>
<tr>
<td>UPEC</td>
<td>Uropathogenic <em>Escherichia coli</em></td>
</tr>
<tr>
<td>UTI</td>
<td>Urinary tract infection</td>
</tr>
<tr>
<td>xg</td>
<td>Gravitational force</td>
</tr>
</tbody>
</table>
Chapter One: General Introduction and Literature Review
1.1 *Escherichia coli*: the microorganism

*Escherichia coli* is currently one of the most well characterized bacteria. A vast number of studies have been conducted to characterise this microorganism following episodes of infection outbreaks that have placed it as one of the major causes of increasing human mortality, given its burden on the health care system (DuPont *et al*., 1971; Russo & Johnson, 2003; Staley *et al*., 1972).

However, for most of the last half century, *E. coli* was considered to be just a part of the normal flora of the mammalian intestine (including that of humans) and exhibiting low levels of pathogenicity (Ferguson & June, 1952). *E. coli* was named in 1885 after it’s discoverer Dr. Theodor Escherich, a German paediatrician (Shulman *et al*., 2007), who made his initial description of the microorganism and named it *Bacterium coli commune*, which was reprinted in English in 1989 (Escherich, 1989). Over a decade later in 1897, Kiyoshi, a Japanese scientist, discovered another bacterium named *Bacillus dysentericus* (*Shigella dysenteriae*), which has been shown to be closely related to *E. coli* (Trofa *et al*., 1999). Later on, the names of *Escherichia* and *Shigella* were assigned to these two organisms, respectively, following their motility, metabolic profiles and distinguishing clinical symptoms (Castellani & Chalmers, 1919).

Nowadays, *E. coli* is considered to be a global pathogen that is frequently causing bacterial infections in humans. Four main types of infections can be linked to *E. coli* based on the clinical symptoms and the virulence factors that mediate them. These infections are: urinary tract infections (UTI); bacteraemia (bacterial blood stream infection); neonatal meningitis; and gastroenteritis (Katouli, 2010; Nataro & Kaper, 1998; Weintraub, 2007).
The emergence of antibiotic resistant *E. coli* strains presents a significant global health concern. The two main issues of reduced susceptibility toward antibiotics and the limited understanding of virulence mechanisms have to be addressed to contain the threat of this pathogen.

Moreover, different environments can place different challenges that the pathogens have to overcome. These situations result in the bacteria utilising different virulence factors, of which many are carried on plasmids, to establish infections. For instance, *E. coli* evolution is mainly centred around the recombination process (Petty *et al.*, 2014), as antimicrobial resistance genes can be widely acquired on plasmids and other mobile elements via horizontal genes transfer (Diard *et al.*, 2007). In addition, some diseases are caused by multi antibiotic resistance pathogens, which possess the ability for genetic exchange (Petty *et al.*, 2014; Rogers *et al.*, 2011). In the last decade or so, reported cases regarding outbreaks of antibiotic resistant *E. coli* strains have increased worldwide (Croxall *et al.*, 2011; Lavollay *et al.*, 2006).

Therefore, it essential to have a deeper understanding of *E. coli* virulence and transmission mechanisms, and this extends to the metabolic activity employed by bacteria to survive and establish various infections (Hussain *et al.*, 2017; Luthje & Brauner, 2014).
1.1.1 General and cultural features

*E. coli* bacterium is a Gram-negative bacillus that grows in facultative-anaerobe conditions, so has the ability to undergo fermentation. It belongs to *Enterobacteriaceae* family. *E. coli* is a motile, non sporing and versatile organism that can grow naturally in a wide variety of environments. *E. coli* can grow optimally between 7°C and 50°C and exhibits acid tolerance down to pH 4.4. It can be cultured in the laboratory using simple culture media when incubated aerobically or anaerobically at 37°C (Salyers & Whitt, 2002). *E. coli* uses these durability characteristics to survive in numerous conditions including water, soil and on surfaces; and to establish its presence in different forms as a commensal strain or as a pathogen in humans or animals (van Elsas *et al.*, 2011).

The morphological structure of *E. coli* on solid medium is observed in formation of non-pigmented colonies with ‘rough’ or ‘smooth’ forms. The rough colony form can be described as flat and irregular in shape, while the smooth colonies are raised and circular in shape (Hasman *et al.*, 2000).

*E. coli* cell structure (Figure 1.1) consists mainly of the cell wall made of protein and carbohydrates, which prevent the cell from lysing; a plasma membrane, made of lipids and proteins with a key function of controlling the movement of molecules in and out of the cell; ribosomes for protein production; fimbriae and flagella for the capability of adherence and motility, respectively; and a nucleoid, a membrane-less structure, which contains DNA (Bayer & Anderson, 1965; Gumbart *et al.*, 2014; Reshes *et al.*, 2008).
1.1.2 Serological types of *Escherichia coli*

Some key features of the outer membrane in *E. coli* are the components that are the focus of serotyping methods. Serotyping is an effective epidemiological tool has been used to differentiate lineages and population composition in numerous bacterial species. *E. coli* strains can be serologically classified based on three criteria: the O antigens (lipopolysaccharide); the K antigens (capsule); and the H antigens (flagella). Kauffmann initially introduced this classification scheme (Kauffmann, 1964) to be later developed to include antibodies for detection of 173 O antigens, 80 K antigens, and 56 H antigens, creating over 100,000 possible patterns of identification (Orskov & Orskov, 1992).
Solid connections between different serotypes and various pathotypes have been shown in many studies. In addition, UTIs have been linked with some serotypes (Johnson et al., 1997), which are also associated with specific subgroups of *E. coli* (e.g. UPEC). For example, O antigen groups O11, O17, O73, and O77 are shared in UPEC clonal group A strains (*E. coli* phylogenetic group D and ST69) (Lloyd et al., 2007; Prats et al., 2000; Skjot-Rasmussen et al., 2013; Tartof et al., 2005).

### 1.1.3 Phylogenetic organisation of *E. coli*

The multi-locus enzyme electrophoresis (MLEE) analysis of 88 reference strains of *E. coli* (72 human isolates and 16 other mammalians isolates) defined six major phylogenetic groups, known as A, B1, B2, C, D, and E (Ochman & Selander, 1984).

These groups have great ecological indications. For instance, groups A and B2 are the most common in humans with 40.5% and 25.5%, respectively, when compared to groups B1 and D (17%). On the other hand, B1 strains are the predominant groups in animals (41%), followed by A (22%), B2 (21%) and D (16%) (Chaudhuri & Henderson, 2012).

Remarkably, many strains of group B2 and some of group D strains are ExPEC, while strains of groups A and B1 are usually lacking extraintestinal virulence factors. This is signifying a connection between phylogeny and extraintestinal virulence (Bingen et al., 1998; Chaudhuri & Henderson, 2012).
1.1.4 Pathotypes of *Escherichia coli*

*E. coli* are known to be present in different forms, either being non-pathogenic commensal strains normally found in the gastrointestinal tract of humans and animals, or as a pathogen in humans or animals, where they can cause serious infections.

*E. coli* strains can be divided into two main categories based on the nature of their pathogenesis. These two types are enteric *E. coli* (EC) and extra-intestinal pathogenic *E. coli* (ExPEC). EC strains are essentially limited causing infections of the mucous lining of the intestines, while ExPEC strains are capable of spread further to various parts of the body (e.g. urinary tract and the bloodstream) and can cause several serious infections (Kuhnert *et al.*, 2000; Marrs *et al.*, 2005; Xie *et al.*, 2006).

Furthermore, EC strains can also be divided further based on their pathogenic features into six subtypes: enteropathogenic *E. coli* (EPEC); enterohaemorrhagic *E. coli* (EHEC); enterotoxigenic *E. coli* (ETEC); enteroinvasive *E. coli* (EIEC); enteroaggregative *E. coli* (EAEC); and diffusely adherent *E. coli* (DAEC) (Kaper *et al.*, 2004; Weintraub, 2007). Members of the ExPEC group are more commonly known for causing UTI, neonatal sepsis and bloodstream infections (Manges & Johnson, 2012; Miajlovic *et al.*, 2016; Pitout, 2012).

1.2 Urinary tract infection

UTI is one of the critical causes of morbidity and mortality around the world, affecting all age groups. It has one of the highest costs of treatment in
the United States, costing ~US$3.5 billion per year, including health care costs and loss of work hours from the workforce (Flores-Mireles et al., 2015). In 2011, 14% of antibiotic prescriptions for community-acquired infections in England were for UTIs (Sadler et al., 2017). Recently, for October 2018, the recommended first-line UTI treatment in England (Nitrofurantoin) has been prescribed over 350 thousand times, with a total cost of over £3 million (OpenPrescribing.net, 2017).

Anatomically, the urinary tract is partitioned into an upper part defined by the kidneys, renal pelvis, and ureters; and a lower part made up of the urinary bladder and urethra. UTI is an inflammatory reaction of the urothelium to bacterial infection that is typically connected with bacteriuria (the presence of bacteria in the urine) and pyuria (the presence of pus in the urine, typically during a bacterial infection). UTI may include just the lower urinary tract or both the upper and lower tracts (Flores-Mireles et al., 2015).

In addition, UTI can be symptomatic or asymptomatic infection. In the latter case, the bacteria are present in patient’s urine but no UTI signs or symptoms are present. Asymptomatic infection occurs most commonly in patients of a higher age, with an incidence of 40-50% in those above 80 years of age (Givler & Givler, 2018). However, most of these cases will not develop into symptomatic UTIs nor require antibiotic treatments, except for certain cases such as pregnant women (Ditkoff et al., 2018; Wullt et al., 2019).

1.2.1 Epidemiology of UTI

UTI can be divided clinically into two types: uncomplicated UTI that classically occurs in individuals who are otherwise healthy and have no
structural or neurological urinary tract anomalies (Hooton, 2012); and complicated UTI, which is a disease associated with abnormal aspects affecting the urinary tract or even the host defence system (Lichtenberger & Hooton, 2008).

Uncomplicated UTIs falls into two categories: lower UTIs (cystitis) and upper UTIs (pyelonephritis) (Hannan et al., 2012). Furthermore, Some of the risk factors contributing to occurrence of cystitis are: being a female gender, a past UTI, sexual intercourse, diabetes, vaginal infection, being overweight and being genetically susceptible (Foxman, 2014; Hannan et al., 2012).

In contrast, complicated UTI risk factors include urinary obstruction due to the presence of foreign objects (indwelling catheters and calculi), neurological disease causing retention of urine, kidney failure or transplantation, pregnancy, and immunosuppression in hospitalised patients (Levison & Kaye, 2013). The majority of complicated UTIs are Catheter-associated, where it accounts for ~80% of the UTI cases in the United States (Assadi, 2018). In addition, CAUTI is one of the contributing factors of increased rates of morbidity, mortality, and establishing a secondary infection of bacteraemia (Chenoweth et al., 2014).

Recurrence of UTI could be due to relapses or re-infection. A recurrence of bacteriuria is due to reappearance of the same microorganism that was displayed before treatment (Nickel, 2005). In E. coli, The bacterial serine-threonine protein kinase HipA is responsible for catalysing hibernation, resulting in temporarily tolerance to antibiotics as it stops growing and become dormant bacteria (Semanjski et al., 2018). Re-infection is a repeat of bacteriuria with a different microorganism from the original infection or could also be with the original microorganism. Chronic UTI indicates perseverance of the same
microorganism for a long time with relapses after treatment (Kodner & Thomas Gupton, 2010). In addition, many of these infections are associated with antimicrobial resistant microorganisms (Pallett & Hand, 2010).

Both Gram-positive and Gram-negative bacteria can cause UTIs. In addition, some fungi such as Candida albicans can infect the bladder and kidneys leading to Funguria (Behzadi et al., 2015). However, uropathogenic E. coli (UPEC) is the most common bacterial strain that causes uncomplicated and complicated UTIs (Flores-Mireles et al., 2015). E. coli represents around 50% of hospital acquired UTIs and up to 90% of community acquired infections (Toval et al., 2014). Within the E. coli species, various subgroups (serogroups O1, O2, O4, O6, O7, O8, O18, O25, O68 and O75) are often detected in patients with UTI (Ananias & Yano, 2008; Delannoy et al., 2017; Li et al., 2010).

The causative agents for the rest of community acquired UTI vary between Gram-negative bacteria, for example, Klebsiella and Proteus; and Gram-positive Enterococcus faecalis and Staphylococcus saprophyticus (Flores-Mireles et al., 2015). The rest of clinically presenting UTIs typically happen after colonisation with Enterobacter, Citrobacter, Serratia, Pseudomonas aeruginosa, Providencia, E. faecalis, or Staphylococcus epidermidis (Tabibian et al., 2008). Remarkably, the patient's age may impact the kind of infective microorganism present, where Staphylococcus saprophyticus now represents 10-15% of UTIs in young females in contrast with under 1% in elderly female patients (Kline et al., 2010). In fact, UTI is considerably more frequently associated with females than males. This indicates the significance of the ascending pathway of UTI as the female urethra is shorter and closer to the perianal area than the male urethra, thus, more susceptible to contamination.
1.2.2 Pathogenicity of UTI

UTI happens as a result of the interaction between the virulence mechanisms of the bacteria and host defence systems. There are three possible courses whereby microbes can attack and spread inside the urinary tract; the ascending pathway, the haematogenous (descending) pathway, and lymphatic route.

In many studies, different models were used to study and visualise the different modes of developing UTI (Hannan & Hunstad, 2016; Hung et al., 2009). For example, a well-established murine model was used for a study on antibiotic treatments for long term ascending UTI (Hvidberg et al., 2000). These models enable us to expand our understanding of UTI pathogenicity, thus identifying new therapeutic and preventative measures for acute, chronic, and recurrent UTI.

1.2.2.1 The ascending pathway for development of UTI

In this route, infection originates mostly from bowel microorganisms, which ascend through the urethra to the bladder and then possibly the upper urinary tract (Figure 1.2). It has been demonstrated that organisms that cause UTI in females colonise the vaginal and periurethral areas before the occurrence of UTI (Gilbert et al., 2017). Inside the bladder, microorganisms may multiply and after that spread to the ureter, particularly if vesicoureteral reflux is present, then to the renal pelvis and parenchyma (Niveditha et al., 2012).

Sexual acts can increase transfer of microbes into the female bladder. Using spermicides can facilitate colonisation of the vagina with uropathogens. In spite of the fact that the predominant Lactobacillus vaginal flora is more
susceptible to nonoxynol-9 (an organic compound widely used in contraceptives for its spermicidal properties) than *E. coli* (Ojha *et al.*, 2003), this compound also enhances adherence of *E. coli* to vaginal epithelial cells (Steiner & Cates, 1997). Lack of oestrogen is presently perceived as a risk factor for repetitive UTI in post-menopausal females, in light of the resulting vaginal flora changes with loss of defensive *Lactobacillus*, which are replaced by different uropathogens (Kodner & Thomas Gupton, 2010). Using various virulence factors (Table 1.1), especially hemolysin A, UPEC strains are able to invade the uroepithelial cells layers (Dhakal & Mulvey, 2012).

In a previous study, it was shown that UPEC (except avian pathogenic *E. coli* strains or *fimH* deficient mutant UPEC) could interfere with contractility in a rat ureter model (Table 1.1) (Floyd *et al.*, 2012). In addition, the study also revealed that UPEC strains specifically inhibit contraction of the ureters disrupting the flow of urine, which may facilitate ascending infection.

The ascending route can also be progressed further with uropathogens invading the blood stream via the kidneys to cause septicemia (Bahagon *et al.*, 2007). In fact, *E. coli* bacteremic urinary tract infections increased in England by 15.6% in the year 2014 (Goto & Al-Hasan, 2013; Lishman *et al.*, 2018).
1.2.2.2 Haematogenous pathway of developing UTI

Kidney infection in healthy individuals via the haematogenous route is uncommon but may lead to UTI. Infrequently, the renal parenchyma of a hospitalised patient may be infected during *Staphylococcus aureus* bacteraemia (Muder *et al.*, 2006) or *Candida* fungemia that originates from oral sources in immunosuppressed patients (Fisher *et al.*, 2011). However, infections of the kidney with Gram-negative bacilli rarely occur by the haematogenous route (Agarwal *et al.*, 2012).
1.2.2.3 The lymphatic pathway

This pathway of UTI occurrence is very uncommon. On rare events, bacteria can spread and invade the urinary tract from the closest organs via the lymphatic system. Retroperitoneal abscesses, necrotizing enterocolitis, and acute abdominal infections are the most common clinical cases connected to the lymphatic route of UTI (Winter et al., 2016). In females, the urogenital tract can be infected by pathogens travelling through preuterine lymphatics, while in males, the prostate bladder could possibly become infected via the rectal and colonic lymphatic (Meares, 1984).

1.3 Uropathogenic *Escherichia coli*

Uropathogenic *E. coli* (UPEC) are considered one of the most common causes of UTI, especially in females (Flores-Mireles et al., 2015). Generally, bacteriuria is indicated when a midstream urine CFU count is $>10^5$ CFU/ml. Yet, some cases of UTI in women appears with $<10^5$ CFU/ml (Lloyd et al., 2007).

UPEC strains frequently deliver hemolysins that can result in lysis of host cells. These strains have the unique ability to colonise and cause infection in the urinary tract (bladder) transitioning from the intestinal tract (Dhakal & Mulvey, 2012; Mann et al., 2017).

1.3.1 Molecular epidemiology of UPEC

Within UPEC, some strains have been seen to have an increased prevalence or virulence. The current strategy used to investigate the genetic diversity and relatedness between isolates of pathogenic *E. coli*, including
UPEC, is multilocus sequence typing (MLST) (Wirth et al., 2006). Improvements and progress made to date in DNA sequencing has helped MLST to become an essential technique for typing of epidemiologically significant pathogens (Lau et al., 2008a). This technique has been useful in local or sporadic outbreak investigation and widely utilized to successfully characterise UPEC strains (Croxall et al., 2011; Lau et al., 2008b; Tartof et al., 2005).

MLST detects the differences in sequence of normally seven specific genes called housekeeping genes (Figure 1.3). In the most common MLST scheme for *E. coli*, these genes are: *adk* (Adenylate kinase), *fumC* (Fumarate hydratase class II), *gyrB* (DNA gyrase subunit B), *icd* (Isocitrate dehydrogenase [NADP]), *mdh* (Malate dehydrogenase), *purA* (Adenylosuccinate synthetase), and *recA* (Protein RecA). Any selective pressure such as antibiotic exposure does not affect these genes, nor do they accumulate neutral genomic variations at a high rate. MLST involves DNA sequence analysis of fragments of 450-500 bp for each gene target, which enables accurate detection of the genome fragment with single pair of primers. Each gene has different sequences that, in turn, are assigned to different allele numbers. Then, these allele numbers of the housekeeping genes provide an allelic profile representing a specific sequence type (ST) (Enright & Spratt, 1999; Spratt, 1999; Urwin & Maiden, 2003).

Application of MLST to global collections of UPEC have revealed the importance of a relatively small number of strains/clones including ST10, ST69, ST73, ST95, ST127 and ST131 (Petty et al., 2014; Price et al., 2013). ST131 isolates are well established in the wider UPEC population and are acknowledged to be a key strain of UPEC (Johnson et al., 2010).
Figure 1.3: *E. coli* CFT073 complete genome presenting the seven housekeeping gene targets, according to the Achtman MLST protocol (Wirth *et al.*, 2006). Adapted from (Gibreel, 2011).

In contrast to the global significance of ST131 isolates, ST127 has been described as a recently emerged virulent clone that has not been widely studied (Gibreel *et al.*, 2012b). ST127 isolates possessed the common uropathogenic O6 serotype (Johnson, 1991). However, isolates of UPEC ST127 are not often reported in prevalence studies, although they have been reported to be significant members of prospective strain collections (Alghoribi *et al.*, 2015; Banerjee *et al.*, 2013). One explanation for the low prevalence of ST127 strains in some studies could be linked to the fact that UPEC ST127 isolates often exhibit full susceptibility to antibiotics that are widely used with UTIs (Croxall *et al.*, 2011; Hertz *et al.*, 2016). In addition, knowing that most epidemiological
studies are based on clinical laboratory isolates from patients with antibiotic resistant strains or recurrent infection, it is more likely for ST127 isolates to be underestimated in these surveys.

In a recent study on community-acquired UTIs, UPEC ST127 isolates were the second most prevalent strains, following ST95 strains (Yamaji et al., 2018). The study also showed that the rate of ST127 isolate detection within a Northern California college community had increased from 11% in 1999-2000 to 16% in 2016-2017. Furthermore, another study revealed that ST127 isolates have higher lethality in the in vivo model of Galleria mellonella larvae (GML) infection and a higher virulence profile compared to isolates of other STs, based on PCR detection of known UPEC related virulence factors (Alghoribi et al., 2014). In the same study, one isolate of ST127 (UPEC EC18) was found to be avirulent toward the wax moth larvae. Genomic analysis revealed that that strain has an insertion sequence (IS1) mediated deletion in the O-antigen cluster, which could explain the cause of its behavior in the Galleria model (Alghoribi et al., 2014).

Although using MLST has been proven useful to identify different STs, it is a costly and time-consuming tool (O'Farrell et al., 2012). Thus, there is a need to find more effective and direct approach to identify specific ST. Development of ST-specific PCR showed great potentials in surveillance or in culture collections screening for a variety of STs (e.g. STs 69, 73, 95, and 131) but not for a ST127 clone (Clermont et al., 2009; Doumith et al., 2015; Li et al., 2010). However, a recent study by O'Hara and colleagues (O'Hara et al., 2019) focused on finding ST127-specific PCR, was able to identify a three-gene multiplex PCR, including an extraction/PCR control (gyrB) and two ST127 specific genes (upaG and fliC). The assay is highly specific to UPEC ST127 and
it can be applied in clinical settings for urgent detection and monitoring of this high-virulent clone.

### 1.3.2 Virulence factors of UPEC

Many virulence factors that contribute to the ability of UPEC to cause UTI are well described and the most important ones are listed in Table 1.1. UPEC strains produce type 1 pili, which attach to mannose-containing glycoproteins (uroplakins) present on the luminal surfaces of bladder epithelial cells through the tip protein adhesin (FimH), mediating cellular invasion. Moreover, P pili bind particularly to a globobiose (Gal-1,4-Gal) disaccharide connected to a ceramide lipid (the P blood group antigens) through the tip protein adhesin (PapG) (Mulholland et al., 1984). Ceramide lipid was found particularly on the surfaces of host erythrocytes and uroepithelial cells (Hannan et al., 2012). Thus, the presence of the P pili receptor in the human population dictates an individual’s susceptibility to UTIs caused by UPEC (Bien et al., 2012). Furthermore, a study has shown that both function and expression of type 1 pili of UPEC strain UTI89 was affected negatively following growth in filtered human urine (Greene et al., 2015).

Despite the fact that the physical flow of urine and many host factors can limit the contact of UPEC with epithelial cells of the bladder, once contact is made, adhesion and invasion can be the result. This will prompt replication inside the cellular vacuoles and lead to lysis and sloughing of the cells (Wu et al., 2017). On the other hand, a percentage of the microorganisms can persist within the underlying epithelial cells, where they cannot be recognised by the immune system (Nagamatsu et al., 2015).
It was suggested in previous studies that multiple expression of virulence factors of UPEC isolates from the upper urinary tract is higher than pathogens of the lower urinary tract (Johnson, 1991; Yamamoto, 2007). This suggests that upper UTI is likely to develop further and overcome the host innate immune system to cause subsequent infections (Table 1.1).

By shifting expressions of their virulence proteins, UPEC strains can drastically change their virulence and pathological profiles (Terlizzi et al., 2017). A group of known phenotypes have resulted from horizontal gene transfer, through plasmid conjugation and phage mediated gene exchange. The majority of UPEC strains have pathogenicity islands (PAIs) (Schmidt & Hensel, 2004), which can facilitate gene exchange between different UPEC strains or other Enterobacteriaceae species (Messerer et al., 2017; Putze et al., 2009; Schubert et al., 2009). For example, UPEC P. fimbriae island contains virulence factors such as haemolysin, pili, cytotoxic necrosing factor, and uropathogenic specific protein (USP) (Nakano et al., 2001).

The existence of these strains illustrates the ease with which bacteria can change over time, starting with one virotype (organisms of the same species with the same pathogenicity) then evolving into different or related types. Obviously, the potential for new virotypes to develop is high, and E. coli subgroups that vary in their ability to cause these infections without a doubt will emerge in the future (Warnes et al., 2012).
Table 1.1: Most common virulence factors of *E. coli* involved in urinary tract infection*.

<table>
<thead>
<tr>
<th>Virulence factors</th>
<th>Location</th>
<th>Function</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha hemolysin</td>
<td>Exported</td>
<td>Cytotoxicity, haemolysis, and leukocyte lysis.</td>
<td>Alpha hemolysin is one of the common main virulence factors in ExPEC isolated from pyelonephritis, cystitis, asymptomatic bacteriuria, and faecal isolates.</td>
<td>(Dhakal &amp; Mulvey, 2012; Emody <em>et al.</em>, 2003; Johnson, 1991)</td>
</tr>
<tr>
<td>Cytotoxic necrotising factor 1</td>
<td>Exported</td>
<td>Intrusion with phagocytosis and apoptosis process.</td>
<td>CNF1 frequently associated with UTI isolates and significantly related to prostatitis cases.</td>
<td></td>
</tr>
<tr>
<td>Secreted autotransporter toxin</td>
<td>Exported</td>
<td>Cytotoxicity</td>
<td>Sat is more frequently associated with isolates causing pyelonephritis.</td>
<td></td>
</tr>
<tr>
<td>Enterobactin</td>
<td></td>
<td>Facilitate growth under iron restriction conditions.</td>
<td>These factors are mainly associated with invasive UTI. They play complex roles to aid iron acquisition to disrupt of phagocytosis and apply direct toxicity toward host cells.</td>
<td></td>
</tr>
<tr>
<td>Type 1 fimbriae (<em>fimA</em> gene)</td>
<td>Bacterial surface</td>
<td>Adhesion to mucosal epithelium and tissue matrix, invasion, and biofilm formation.</td>
<td>Present in isolates from pyelonephritis (73%), cystitis (49%), bacteraemia (58%), asymptomatic bacteriuria (41%) and faecal (38%). Enforce <em>E. coli</em> to cause serious and complicated UTI, as it helps bacterial growth in iron-limited conditions.</td>
<td>(Emody <em>et al.</em>, 2003; Johnson, 1991)</td>
</tr>
<tr>
<td>P fimbriae (<em>papC</em> gene) (<em>papGIII</em> gene)</td>
<td>Bacterial surface</td>
<td>Adhesion to mucosal epithelium, and cytokine induction.</td>
<td>Present in all <em>E. coli</em> isolates, specifically more detected in <em>E. coli</em> phylogroups A and D. UPEC had the ability to form intracellular bacterial communities (IBCs) due to the high expression of the <em>fimA</em> gene, coding for type 1 fimbriae (T1F), which are responsible for adherence to and invasion of luminal facet cells of the bladder.</td>
<td>(Berry <em>et al.</em>, 2009; Emody <em>et al.</em>, 2003; Johnson, 1991; Piatti <em>et al.</em>, 2008)</td>
</tr>
</tbody>
</table>

* Adapted from (Alghoribi, 2015)
Table 1.1: Most common virulence factors of *E. coli* involved in urinary tract infection*.

<table>
<thead>
<tr>
<th>Virulence factors</th>
<th>Location</th>
<th>Function</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1C fimbriae</td>
<td>Bacterial surface</td>
<td>Adhesion to endothelial cells and mucosal epithelium.</td>
<td>14% of F1C fimbriae <em>E. coli</em> strains are associated with UTI, as the fimbriae bind to the epithelial and endothelial cells of human kidney and bladder.</td>
<td>(Khan et al., 2000)</td>
</tr>
<tr>
<td>S fimbriae</td>
<td>Bacterial surface</td>
<td>Adhesion to endothelial cells, mucosal epithelium, and tissue matrix.</td>
<td>S fimbriae <em>E. coli</em> strains are associated with newborn meningitis and/or UTI, as it was reported that the fimbriae bind to the epithelial cells of the proximal and distal tubules, and renal vascular endothelium.</td>
<td>(Korhonen et al., 1986; Marre et al., 1990)</td>
</tr>
<tr>
<td>Lipopolysaccharide</td>
<td>Bacterial surface</td>
<td>Endotoxicity, ‘O’-antigen, cytokine induction, and serum resistance.</td>
<td>Lipopolysaccharide is a major factor contributing to the pathogenesis of <em>E. coli</em> infection and responsible for the initial immune response. Induces cytokine expression through toll-like receptor 4 (TLR4).</td>
<td>(Kaper et al., 2004; Morrison &amp; Ryan, 1987)</td>
</tr>
<tr>
<td>K antigen</td>
<td></td>
<td>Antiphagocytic, anticomplement effect, serum resistance, and evasion of innate immune system recognition.</td>
<td>Certain types of K antigen such as K1, K5, K12, K13, and K20 are detected in UTI isolates. Moreover, the major K antigens (63%) detected in isolates from pyelonephritis in females are K1 and K5. While O antigens associated with UTI isolates are O1, O2, O4, O6, O16, O18, O22, O25 and O75. These antigens serve as an important tool to differentiate UPEC from other <em>E. coli</em> strains.</td>
<td>(Emody et al., 2003; Johnson, 1991; Wiles et al., 2008)</td>
</tr>
<tr>
<td>O antigen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H antigen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Adapted from (Alghoribi, 2015)
1.3.3 UPEC survival in the urinary tract

UPEC are a pathotype of ExPEC and originally reside in the intestine, where they do not normally cause any complications (Wiles et al., 2008). However, following deployment of the virulence factors described earlier, UPEC can disseminate and colonize other host environments such as the urinary tract and bloodstream (Flores-Mireles et al., 2015). In addition, the ability of UPEC isolates to exploit nutrition in diverse environments starting from the intestines and the urinary tract to the bloodstream indicates the significance of metabolic adaptations in pathogenesis (Eisenreich et al., 2010; Mann et al., 2017). These metabolic pathways are highly regulated and any nutritional changes in the host can affect their expression (Brown et al., 2008). Expression of the pili enables UPEC to attach to the bladder epithelial cells (Table 1.1), leading to invasion and intracellular replication to form intracellular bacterial communities (IBCs) (Rosen et al., 2007). In IBCs, microorganisms are hidden from antibiotic reach and can rapidly multiply and disseminate to invade other cells. Nevertheless, IBCs have not been observed within the bladder wall of humans’ cases and only seen in relation to mouse models (De Nisco et al., 2019). Transiting to a different environment applies pressure on UPEC to metabolically adapt, either in intracellular and extracellular microenvironments (Mann et al., 2017).

For successful transitions to different surfaces and sub-surfaces, UPEC must survive and establish growth, initially this being in urine as they transition from the gastrointestinal tract. Urine is the major nutrient source for UPEC in the urinary tract, especially during the extracellular phase. It is essential to understand the mechanism behind the utilization of urine components by UPEC strains; however, urine has a complex composition and variability between
different individuals presents challenges to researchers in this field (Brooks & Keevil, 1997). Still, the Urine Metabolome Database has all the metabolites identified in human urine to date (Bouatra et al., 2013).

Nutritional sources present at high concentration in human urine are urea, creatinine, organic acids, D-serine amino acids, inorganic ions (ammonia), sodium, potassium, purines, pyrimidines, and water-soluble toxins, which can be utilized by UPEC (Bouatra et al., 2013; Vejborg et al., 2012). Nevertheless, urine is considered nutritionally a limited growth medium due to the limited concentration of arginine, methionine, valine, uracil, adenine, isoleucine, and iron (Stamey & Mihara, 1980; Vejborg et al., 2012). In order to meet the basic requirements for growth, UPEC will depend highly on the ability for synthesis (via catabolism) of these low-concentration metabolites, such as glucose, oxytocin, angiotensin, and melatonin (Bouatra et al., 2013; Mann et al., 2017). Therefore, UPEC strains are well developed metabolically and have the ability to survive the harsh conditions of the urinary tract.

1.3.3.1 Metabolic adaptations of UPEC

The metabolic effect of different environments on UPEC growth is yet to be fully understood (Mann et al., 2017). However, UPEC can apply adaptive strategies by activating genes and metabolic pathways in order to survive the urine conditions and cause UTIs (Figure 1.4). The main adaptations happen in central carbon metabolism, amino catabolism and other pathways.

Central carbon metabolism pathways are essential for intracellular colonization of the bladder and kidneys (Alteri et al., 2009). For example, deletion of the \textit{sdhB} gene, coding for succinate dehydrogenase involved in
conversion of fumarate to succinate during the TCA cycle, showed attenuation in pathogenesis in the bladder and kidneys of mouse models (Alteri et al., 2009). The same study also showed that UPEC fitness in UTI models is reduced when the gluconeogenesis process is disrupted by pckA deletion. The pckA gene encodes phosphoenolpyruvate carboxykinase that converts oxaloacetate to phosphoenolpyruvate (Alteri et al., 2009).

The amino acid catabolism pathway is significant for UPEC growth in human urine and arginine, serine and histidine transporters have been shown to be up-regulated in comparison to growth in LB medium (Conover et al., 2016). In addition, dppA and oppA deletion mutants showed significantly attenuated growth in urine, indicating the importance of small peptide uptake (Alteri et al., 2009).

Purines and pyrimidine pathways are important for metabolic adaptation in the bloodstream. A study has showed that growth of purA, purE, pyrC, and pyrE mutant strains of E. coli was impaired in LB conditions (Samant et al., 2008). In addition, deletion of purF, a purine biosynthesis-associated gene, stopped the synthesis process in UPEC UTI89 but did not affect its adhesion to the epithelial cells. However, UPEC intracellular survival was constrained and cells were unable to disseminate further into the intracellular bacterial communities (Shaffer et al., 2017).
Figure 1.4: Summary of the main metabolic adaptions of UPEC during urinary tract infection (UTI). Genes in blue color denote genes that play a role in UPEC fitness in the urinary tract, or whose deletion results in attenuation in a mouse or in vitro model of invasion and intracellular bacterial community formation. The genes in red denote those specifically identified to play a role in kidney infection. Transporters are shown in green boxes over the cell membrane. *sdhB*, succinate dehydrogenase; *pckA*, phosphoenolpyruvate carboxykinase; *tpiA*, triosephosphate isomerase; *srlA*, sorbitol transporter; *lacZ*, β-galactosidase; *gaiK*, galactokinase; *dsdX*, D-serine specific transporter; *dsdA*, D-serine deaminase; *dppA*, periplasmic dipeptide transport protein; *oppA*, periplasmic oligopeptide-binding protein; *tonB*, ferric iron uptake mediator; *pur*, genes involved in purine synthesis. Adapted from (Mann et al., 2017).

Iron uptake and transport systems such as siderophores play an important role during the intracellular lifecycle of UPEC infection. A deletion of *tonB*, the ferric iron uptake mediator, resulted in reduced colonization of UPEC in the mouse kidney model of UTI after 48 hours post-infection (Torres et al., 2001). Moreover, many genes involved in siderophore synthesis, such as salmochelin receptor *iroN* and the high pathogenicity island encoded *ybtSETU* for synthesis
and uptake of yersiniabactin, were shown to be up-regulated during in vivo mouse UTI (Conover et al., 2016; Henderson et al., 2009; Reigstad et al., 2007). These findings indicate the requirement of iron uptake and transport systems in UPEC pathogenesis in iron-starved conditions. Other known metabolic regulators of UPEC, such as nutritional sensing, are also crucial for UPEC metabolic adaptation and survival. Examples of these are: RyhB (iron-acquisition) (Porcheron et al., 2014); BarA-UvrY (gluconeogenesis and glycolysis) (Palaniyandi et al., 2012); and Fnr (oxygen related genes) (Barbieri et al., 2014).

Furthermore, genes identified to play a role in UPEC adaptation strategies for specific nutrients, have been considered as potential candidates for antimicrobial targets for UTIs (Table 1.2).
Table 1.2: Summary of metabolic functions in uropathogenic *E. coli* (UPEC) that represent potential targets for novel antimicrobial approaches.

<table>
<thead>
<tr>
<th>Targets</th>
<th>Target purpose</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>tpiA, opgH, pyk, dsdA, pur</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transporters</td>
<td>Drug development</td>
<td>No drug is required to enter the cell</td>
<td>Functional redundancy by other iron transporters</td>
<td></td>
</tr>
<tr>
<td><em>(tonB, oppA, dppA)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron receptors</td>
<td>Vaccine development</td>
<td>Impacts the initial stages of UPEC colonization</td>
<td>Less effective, as not all UPECs have all receptors</td>
<td>(Brumbaugh <em>et al.</em>, 2013; Garcia <em>et al.</em>, 2011; Miethke &amp; Marahiel, 2007; Mike <em>et al.</em>, 2016; Wieser <em>et al.</em>, 2010)</td>
</tr>
<tr>
<td><em>(fyuA, iutA, ireA)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siderophore synthesis</td>
<td>Vaccine development</td>
<td>Potential for a universal vaccine candidate</td>
<td>Functional redundancy in iron acquisition systems</td>
<td></td>
</tr>
<tr>
<td><em>(Aerobactin and salmochelin synthesis)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trimethylamine, acetic acid, creatinine,</td>
<td>Diagnosis</td>
<td>Inexpensive and quick diagnosis via easy incorporation on to</td>
<td>Changeable outcomes due to varying levels of</td>
<td>(Ariza-Heredia <em>et al.</em>, 2014; Karlsen &amp; Dong, 2015; Su <em>et al.</em>, 2016)</td>
</tr>
<tr>
<td>guanine, hypoxanthine</td>
<td></td>
<td>strips.</td>
<td>metabolites</td>
<td></td>
</tr>
</tbody>
</table>
1.3.4 Antimicrobial resistance in UPEC

Antimicrobial resistance is a major global concern and this threat requires urgent attention to control the growth and expansion of antibiotic resistant strains. The Review on Antimicrobial Resistance stated that antibiotic resistant bacteria kill around 50,000 people a year in the United States and Europe. Currently, it is estimated to kill more than 700,000 people a year around the world, with an increase to 10 millions deaths a year in 2050 (AMR, 2014).

*E. coli* has been the focus of numerous studies regarding surveillance of antimicrobial resistance. Many studies have suggested that there is no correlation or that there is a negative correlation between the virulence factors and the antibiotic resistance profile of UPEC strains (Alghoribi *et al*., 2014; Gibreel *et al*., 2012b; Johnson *et al*., 2002; Velasco *et al*., 2001; Vila *et al*., 2002). One study highlighted that UPEC isolates might develop resistance, regardless of their virulence profiles, when associated with immune-compromised patients, who are undergoing antibiotic treatment (Johnson, 1991). In addition, in Japan similar outcomes were demonstrated, indicating the existence of a complex association of virulence potential, antimicrobial resistance profile and phylogenetic background of UPEC (Kawamura-Sato *et al*., 2010).

Reduced susceptibility to antibiotics could possibly lead to the emergence of opportunistic pathogens already present as commensal strains, which cause UTI and bloodstream infections that are then more difficult to treat (Russo & Johnson, 2003). In fact, pathogenic *E. coli* does not have exclusivity of antibiotic resistance, as commensal strains could help pathogenic strains by being a major reservoir of resistant genes (van den Bogaard & Stobberingh, 2000).
The most commonly used antimicrobial agents for UTI treatment are β-lactam antibiotics. Uncomplicated UTIs are generally treated with nitrofurantoin and trimethoprim/sulfamethoxazole (Bartoletti et al., 2016; Watson et al., 2018). Third generation cephalosporin (e.g. ceftriaxone) and cefazolin are used for complicated UTIs treatment (Abe et al., 2016; Jorgensen et al., 2018).

During the last two decades, there has been a significant increase in antibiotic resistance in UPEC strains isolated from clinical cases (Olorunmola et al., 2013). Some of the key threats are in strains resistant to the main therapeutic options for UTI, including fluoroquinolones, the β-lactam antibiotics (including the cephalosporins), aminoglycosides and trimethoprim (Phan et al., 2015; Pitout, 2012; Tang et al., 2014).

As mentioned previously in this chapter, ST131 is one of the leading clones of UPEC currently causing human infections. Members of ST131 are often seen to carry resistance to multiple classes of antibiotic, including all front line therapies (Alghoribi et al., 2015; Sarkar et al., 2018). Interestingly, and in stark contrast, ST127 isolates are comparatively more susceptible to the action of a wide range of antibiotics than members of other important clones, like ST131 (Alghoribi et al., 2015; Gibreel et al., 2012b).

In a recent study focused on specimens taken from patients at the point of care, within a Californian university community, ST127 found to be the second most prevalent lineage increasing from 11% to 16% over a period of 17 years (Yamaji et al., 2018). In addition, ST127 UPEC represented 2.3% of the isolates (out of 44 isolates) from community acquired UTI study in Switzerland, characterized for clonal diversity, virulence potential, and antimicrobial resistance (Nuesch-Inderbinen et al., 2017). It was also found that ST127
isolates have the highest virulence factor score among other STs but exhibited no resistance to any of the antibiotics used studied. Given their high virulence potential and recent evidence of emerging resistance to antibiotics, including the cephalosporins (Ferjani et al., 2017; Gomi et al., 2017), clinical cases associated with ST127 strains could substantially increase, resulting in a significant threat to human health.

1.4 Proteomics: an investigative methodology

Proteins have a fundamental role in cell biology, regulating various functions of different cells, controlling activities that range from movement to mitosis. Proteins are complex molecules made of one or more chains of amino acids ordered specifically according to the sequence of nucleotides in the encoding DNA. Following the translation process, more complexity is added via post-translational modifications (PTMs). These modifications include phosphorylation, sulphation, hydroxylation, methylation, and glycosylation (Han & Martinage, 1992). The complete protein content that is analysed with regard to their interactions, kinetics, localization, and PTMs is called the proteome (Wilkins et al., 1996). In other words, proteomics is not only limited to the identification and quantification of proteins (Fields, 2001).

The proteome varies from cell to cell and is continually changing as a result of biochemical interaction with the genome and the environment. Any one microorganism has very diverse protein expression in various parts of a host body, in various phases of its life cycle and in responding to external stimuli it may experience in various environmental conditions. The ability to view biological processes occurring and, with understanding that proteins are
representatives of mRNA levels, to investigate host translational control and regulation, give proteomics an advantage over genomics in understanding the cell lifecycle and functions. It could be said that this makes the proteome the most significant data set that can be used to understand a biological system (Cox & Mann, 2007).

Although proteomics plays an important role nowadays in developing and validating targets for diagnostics and therapy, eukaryotic proteomic studies are quite complex due to PTMs, which occur at many sites via various routes (Krishna & Wold, 1993). Moreover, prokaryotic proteomic studies are also challenging. The difficulties arise from the vast variation in characteristics such as quantity dynamic range, molecular size, and hydrophilicity and hydrophobicity features (Pandey & Mann, 2000).

1.4.1 Technologies for investigating the proteome

Proteomics in one of the “omics” technologies that are now widely used in biomedical research to understand cell biology. These technologies include genomics, transcriptomics, proteomics, and metabolomics. Proteomic methods allow investigation of proteins in relation to the underlying genomic architecture in an organism, and are especially useful for understanding the structure and function of proteins.

There are many approaches used in the proteomics field such as conventional, advanced, quantitative, high throughput, and bioinformatic analysis techniques. Conventional techniques for protein purification are chromatography-based techniques, enzyme-linked immunosorbent assay (ELISA), and Western blotting (Hage et al., 2012). The advanced techniques
include protein microarray, Edman sequencing, mass spectrometry (MS), and gel-based approaches. Moreover, stable isotopic labelling with amino acids in cell culture (SILAC) (Wang et al., 2018), isobaric tagging for relative and absolute quantitation (iTRAQ) (Wiese et al., 2007), and Isotope-coded affinity tag (ICAT) labelling (Shiio & Aebersold, 2006) are examples of the quantitative techniques. In addition, high throughput techniques include both X-ray crystallography and NMR-spectroscopy (Aslam et al., 2017).

1.4.1.1 Mass spectrometry proteomics

MS-based proteomics has become an essential tool for analysis of protein expression on a global scale. For mass spectrometry proteomics experiments (Figure 1.5), the protein sample is firstly extracted from a biological source (e.g. bacteria or cell culture) followed by protein quantification and purification. The purification step is usually done on SDS-PAGE gels in order to allow removal of a specific band or to check the integrity of the total extracted protein. The proteins can be then in-gel digested or total proteome analysis can be conducted by digesting the proteins in-solution to yield a set of peptides using a sequence-specific protease, such as Trypsin and Lys-C. Both of these cleave peptides on the C-terminal side of lysine and arginine amino acid residues (Chen et al., 2009; Rajmakers et al., 2010). In fact, Trypsin/Lys-C protease combination has enhanced protein digestion and, thus, MS analysis (Saveliev et al., 2013).
Figure 1.5: An example of a modern mass spectrometry proteomics workflow.

In MS, peptides’ size is measured and compared to reference databases, although MS can measure whole intact proteins. Proteins are difficult to work with due to solubility variations. Moreover, MS is more sensitive for peptide analysis as whole proteins make determining the masses of the resulting isoforms impossible due to the wide range of modification possible. Furthermore, MS is efficient at detecting sequences ~20 residues long, thus peptides are better for obtaining sequence information for protein identification (Mørtz et al., 1996; Steen & Mann, 2004; Sze et al., 2002).

Filter aided sample preparation (FASP) is a common method to obtain peptides for LC-MS analysis. The advantage of this method is that it allows analysis of detergent-lysed cells and tissues, even at high concentrations of detergent. Thus, it is ideal for total proteome studies and fractions containing biological membranes. Since the method utilises a filtration device, we can obtain a protein digest at high concentrations with no need for precipitation and the analysed samples will be free of cell debris (Wisniewski et al., 2009).
Following protein digestion and obtaining peptides, the sample can be analysed on MS firstly via separation using a high-performance liquid chromatography (HPLC) system that is connected directly “in-line” to the MS. This is followed by injecting the peptides after ionization with soft ionization techniques, such as electrospray ionization (Fenn et al., 1989) and matrix-assisted laser desorption/ionization (MALDI) (Karas & Hillenkamp, 1988).

It is essential for the peptides to be electrically charged for detection in the MS as it passes through voltage gates eliminating any neutrally charged peptides. In a tandem MS (MS/MS) (Figure 1.6), there are two consecutive stages of MS; the total mass is firstly measured in MS followed by peptide collision with an inert gas (e.g. Helium atoms) resulting in fragmentation of the peptides (Steen & Mann, 2004). This process is called collision-induced dissociation (CID) fragmentation, which helps confirm the structure of the detected peptide and can allow de novo peptide sequence determination (Demarque et al., 2016). The actual mass of the molecules is not measured by mass spectrometry, rather it determines the mass-to-charge ratio (m/z). Electrospray ionization creates differently charged ions. To calculate the molecular weight of a peptide, the detected m/z value has to be multiplied by z and corrected for the number of attached protons (which equals z) (Steen & Mann, 2004).
Figure 1.6: Peptide sequencing using the tandem mass spectrometry (MS/MS). This technique uses collision-induced dissociation (CID) fragmentation. m/z stands for mass-to-charge ratio.

At the end of the analysis pathway, the sequence data for fragmented peptides are used in searches against a range of protein databases. There are many software tools available for analysis of proteomic data, but MaxQuant, a complementary computational framework for processing and statistical assessment of MS/MS data, facilitates a sensitive six-frame database search and is well suited for detailed proteomic analyses. It is worth noting that these approaches can also be used in global genome re-annotation applications (Krug et al., 2013).
Proteomic data can be searched against many databases such as NIH genetic sequence database (GenBank), NCBI reference sequence database (RefSeq), the universal protein resource (UniProt), and the protein information resource (PIR) (Aslam et al., 2017). Other databases can be used for understanding the role of proteins in certain pathways including metabolism, signalling and network or cellular interactions. Examples of these databases are BioCarta, kyoto encyclopedia of genes and genomes (KEGG) (Kanehisa et al., 2017), and Reactome (Croft et al., 2011).

In general, proteins act in complex ways with each other. It is essential to use protein-protein interaction (PPI) databases to investigate and understand biological systems at the protein level to help understand the complex connections. BioGRID, IntAct, MINT and HRPD are examples of the most widely used PPI databases. In addition, one of the most common PPI databases applications is the ‘search tool for the retrieval of interacting genes/proteins’ (STRING), which is connected to other databases for more confirmatory resources (Szklarczyk et al., 2015).

1.4.2 Proteomic approaches to investigate UPEC pathogenesis

During the last decade, significant changes in the field of proteomics have taken place, especially in investigating the bacterial proteome (Cox et al., 2014; VerBerkmoes et al., 2004). Proteomics complements genomic analysis as it can indicate, contrary to mRNA, the sub-cellular location of proteins, PTMs, macromolecular interactions, structure and molecular characterization (Cox & Mann, 2007). However, proteomic studies have much greater analytical challenges. Proteins can be expressed in different concentrations depending on
the protein function and location, while the genome usually has a set number of copies of every gene. In addition, an amplification technique is not available for proteomics studies in contrast to genomics studies, where polymerase chain reaction (PCR) can increase sensitivity (VerBerkmoes et al., 2004).

A number of studies have been conducted on the bacterial proteome (Liu et al., 2015; Osbak et al., 2016; Schmidt et al., 2016). For example, a study on proteome quantification in *E. coli*, demonstrated the use of multi-enzyme digestion FASP and the total protein approach (Wisniewski & Rakus, 2014). The study also supported the proposed use of Trypsin/Lys-C combinations for better protein digestion (Saveliev et al., 2013) and demonstrated that the total protein approach estimates of protein abundances are relative to the values of well characterised protein standards. This demonstrates the simplicity and applicability of using such a methodology to investigate large-scale proteomics (Wisniewski & Rakus, 2014). Another study included a comprehensive analysis on the closely related strains of *E. coli* B and *E. coli* K-12, investigating the intracellular, outer membrane, and extracellular proteomes (Yoon et al., 2012). The analysis revealed that *E. coli* B is favourable for recombinant proteins production due to a higher amino acid biosynthesis, lack of flagella, and fewer proteases. In addition, *E. coli* B has more secreted proteins due to possession of additional type II secretion system and a different outer membrane composition.

Currently, there are only a limited number of studies investigating the virulence of UPEC using proteomic approaches (Lo et al., 2017). Smith and colleagues investigated the cellular proteome of *E. coli* isolates from UTIs and random faecal samples (Smith et al., 2011). The analysis revealed that individual patients present a similar *E. coli* population, but different to other
patients with the metabolic variations are the highlighted difference. In addition, the surface proteins of UPEC were the focus of a previous shotgun proteomics study. The analysis showed that using protease to trim the surface-exposed peptides off the bacterial cell is not applicable for Gram-negative bacteria as it caused contamination to the sample with cytoplasmic proteins (Walters & Mobley, 2009).

More recently, differences in protein expression profiles were used to analyse UPEC biofilms and distinguish the different populations within the biofilms using matrix-assisted laser desorption/ionization time-of-flight imaging mass spectrometry (MALDI-TOF IMS) (Floyd et al., 2015). The study also showed that oxygen tension regulates type 1 pili expression, as it was only observed in the top-most layer of the biofilm. Moreover, Wurpel and colleagues analysed 54 UPEC isolates for their outer membrane proteome (Wurpel et al., 2015). This study provided an inventory of surface proteins and a comparative proteomic analysis of artificially induced outer membrane vesicles using tandem mass spectrometry. In a subsequent study, the same group identified the surface proteomic profiles of UPEC during growth in pooled human urine and identified new factor, UCA-like (UCL) fimbriae, that may contribute to biofilm formation in UTI (Wurpel et al., 2016). However, more studies are required for a deeper understanding of UPEC biology and pathogenicity, in particular for revealing metabolic adaptations. Such proteomic studies may allow identification of new therapeutic targets.
1.5 Aims of this study

Considering the reported significance of UPEC infection and its effect on the population health and global economics, more detailed studies are required into the pathogenesis of UPEC. There has been a substantial focus on leading drug-resistant clones on UPEC in many recent studies. One lineage that has received less attention is ST127, members of which are often widely susceptible to antibiotics, but have a high virulence potential, based on virulence gene content and ability to kill *G. mellonella* larvae. Acquisition of antibiotic resistance by ST127 strains could result in highly pathogenic and resistant UPEC, so the clone requires more detailed investigation.

By understanding the biology of UPEC in more detail, it may be possible to design new therapies that do not result in further increases in antibiotic resistance. One aspect of the pathogenesis of UPEC strains that has not received much attention is investigation of their metabolic adaptations during infection. Previous work in the group has indicated that there may be some variation in the phenotype of UPEC recovered from urine and blood, even though the blood isolates were assumed to be derived from the urine colonising strains. This suggests metabolic adaptation occurs during transition from urine to blood. Proteomic studies can offer a better means to understand the pathogenic potential and metabolic adaptation of UPEC, especially when employed alongside genomic analysis.

With the above in mind, the work conducted during this project had the overall purpose of using proteomic analysis to explore the biology of UPEC clones, in a number of settings, with a focus on members of ST127. The work
required development of relevant methods and their application in a range of in vitro conditions. The following Aims and Objectives were set:

Firstly, to establish and optimize the use of proteomics for investigating pathogenesis of UPEC strains, the proteome profiles of genetically different background strains of UPEC will be determined and compared following growth in different laboratory media. This will establish the methodology that will enable further investigation of UPEC STs in different settings.

Secondly, to increase our understanding of the metabolic adaptations and underlying mechanisms of UPEC adaptation to different clinical settings (e.g. blood versus urine) and to examine the ability of strains to disseminate into and grow in blood, a quantitative proteomics and genomic analysis will be conducted to investigate and compare the proteome profiles of paired UPEC isolates (from blood and urine of the same patient). To reveal changes that might be specific to UPEC of certain genetic backgrounds, investigation will need to be conducted for strains of different STs.

Finally, as there have been no comprehensive investigations into the underlying mechanisms of ST127 UPEC pathogenesis, it will be informative to conduct co-culture experiments with uroepithelial cells and different strains of ST127. By using quantitative proteomic approaches to investigate the interaction of UPEC with uroepithelial cells, it may be possible to reveal potential targets for novel therapeutic or vaccine-based means to control or prevent UTI.
Chapter Two: Materials and Methods
2.1 Bacterial strains and culture conditions

Bacterial strains used in this study are listed in Table 2.1. These strains were previously collected as a part of other studies (Alghoribi et al., 2014; Alghoribi et al., 2015; Gibreel et al., 2012b). Frozen stocks of the used strains were stored at -80°C in the Microbank™ preservation system (Pro-Lab Diagnostics, UK). E. coli growth cultures were prepared with Lysogeny broth (LB) (Fisher BioReagents, UK) and artificial urine medium (section 2.1.1), then incubated aerobically at 37°C overnight with constant shaking (~130 rpm).

For agar cultures, 1.2% LB agar was used, and plates were incubated aerobically at 37°C overnight.

2.1.1 Artificial urine medium

Artificial urine medium (AUM) was also used for growth of E. coli strains. AUM was prepared using a method adopted from Brooks and Keevil (1997), as shown in Table 2.2. The pH for AUM was made up to 6.5 and then the medium was sterilised via filtration with 0.2 µm nylon membrane filters, as autoclaving AUM will lead to precipitation (Brooks & Keevil, 1997). Sterilised AUM was stored at 4°C.
Table 2.1: *Escherichia coli* strains used in this study.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sample</th>
<th>Hospital / Department</th>
<th>Sequence Type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>K12</td>
<td>Urine</td>
<td>Manchester Royal Infirmary</td>
<td>ST131</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC1U</td>
<td>Blood</td>
<td>Manchester Royal Infirmary</td>
<td>ST131</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC1B</td>
<td>Urine</td>
<td>Mid Yorkshire NHS Trust</td>
<td>ST131</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC10U</td>
<td>Blood</td>
<td>Mid Yorkshire NHS Trust</td>
<td>ST131</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC18U</td>
<td>Urine</td>
<td>Mid Yorkshire NHS Trust</td>
<td>ST127</td>
<td>(Alghoribi et al., 2014)</td>
</tr>
<tr>
<td>EC18B</td>
<td>Blood</td>
<td>Mid Yorkshire NHS Trust</td>
<td>ST127</td>
<td>(Alghoribi et al., 2014)</td>
</tr>
<tr>
<td>EC41U</td>
<td>Urine</td>
<td>Manchester Royal Infirmary</td>
<td>ST127</td>
<td>(Alghoribi et al., 2014)</td>
</tr>
<tr>
<td>EC41B</td>
<td>Blood</td>
<td>Manchester Royal Infirmary</td>
<td>ST127</td>
<td>(Alghoribi et al., 2014)</td>
</tr>
<tr>
<td>SA189</td>
<td>Urine</td>
<td>Riyadh, Saudi Arabia</td>
<td>ST127</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC22U</td>
<td>Urine</td>
<td>Manchester Royal Infirmary</td>
<td>ST73</td>
<td>(Gibreel et al., 2012b)</td>
</tr>
<tr>
<td>EC22B</td>
<td>Blood</td>
<td>Manchester Royal Infirmary</td>
<td>ST73</td>
<td>(Gibreel et al., 2012b)</td>
</tr>
<tr>
<td>EC73U</td>
<td>Urine</td>
<td>Manchester Royal Infirmary</td>
<td>ST73</td>
<td>(Gibreel et al., 2012b)</td>
</tr>
<tr>
<td>EC73B</td>
<td>Blood</td>
<td>Manchester Royal Infirmary</td>
<td>ST73</td>
<td>(Gibreel et al., 2012b)</td>
</tr>
<tr>
<td>EC47U</td>
<td>Urine</td>
<td>Mid Yorkshire NHS Trust</td>
<td>ST38</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC47B</td>
<td>Blood</td>
<td>Mid Yorkshire NHS Trust</td>
<td>ST38</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC86U</td>
<td>Urine</td>
<td>Manchester Royal Infirmary</td>
<td>ST38</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
<tr>
<td>EC86B</td>
<td>Blood</td>
<td>Manchester Royal Infirmary</td>
<td>ST38</td>
<td>(Alghoribi et al., 2015)</td>
</tr>
</tbody>
</table>
Table 2.2: Artificial urine medium composition*.

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity (g)</th>
<th>Concentration (mmol L⁻¹)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptone</td>
<td>1</td>
<td>N/A</td>
<td>Lab M</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>0.005</td>
<td>N/A</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Lactic acid</td>
<td>0.1</td>
<td>1.1</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Citric acid</td>
<td>0.4</td>
<td>2</td>
<td>VWR</td>
</tr>
<tr>
<td>Sodium Bicarbonate</td>
<td>2.1</td>
<td>25</td>
<td>Fisher Scientific</td>
</tr>
<tr>
<td>Urea</td>
<td>10</td>
<td>170</td>
<td>Fisher Scientific</td>
</tr>
<tr>
<td>Uric acid</td>
<td>0.07</td>
<td>0.4</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Creatinine</td>
<td>0.8</td>
<td>7</td>
<td>Acros Organics</td>
</tr>
<tr>
<td>Calcium chloride.2H₂O</td>
<td>0.37</td>
<td>2.5</td>
<td>Fisher Scientific</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>5.2</td>
<td>90</td>
<td>Oxoid</td>
</tr>
<tr>
<td>Iron II sulphate.7H₂O</td>
<td>0.0012</td>
<td>0.005</td>
<td>VWR</td>
</tr>
<tr>
<td>Magnesium sulphate.7H₂O</td>
<td>0.49</td>
<td>2</td>
<td>VWR</td>
</tr>
<tr>
<td>Sodium sulphate.10H₂O</td>
<td>3.2</td>
<td>10</td>
<td>Sigma-Aldrich</td>
</tr>
<tr>
<td>Potassium dihydrogen phosphate</td>
<td>0.95</td>
<td>7</td>
<td>VWR</td>
</tr>
<tr>
<td>Di-potassium Hydrogen phosphate</td>
<td>1.2</td>
<td>7</td>
<td>VWR</td>
</tr>
<tr>
<td>Ammonium chloride</td>
<td>1.3</td>
<td>25</td>
<td>VWR</td>
</tr>
<tr>
<td>Distilled water</td>
<td>To 1 L</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Adapted and updated from (Brooks & Keevil, 1997)
2.2 Growth kinetics

Overnight growth cultures (10ml) in LB were prepared. Then, 1 ml of each strain was inoculated into 100 ml flasks of LB and AUM. The flasks were incubated aerobically at 37°C with continuous shaking (~130 rpm). The optical density was measured at 600nm at the following time points: Zero, 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours, and 18 hours. One ml of the culture was placed in a cuvette and measured at OD<sub>600</sub> with a photometer cuvette reader (Eppendorf, UK) to obtain a growth curve. The last two readings were diluted by 50% in phosphate buffered saline (PBS) in order to get accurate readings of the growth curves.

In addition, a serial dilution was performed at these time points to allow determination of colony forming unit (CFU) counts using the Miles and Misra method (Miles et al., 1938). All dilutions were drop plated with 3 drops of 10 µl each on LB agar plates, which were incubated at 37°C overnight, counting the colonies formed the following day. Three independent experiments were performed for growth kinetic studies. The CFU/ml was calculated using the following formula:

\[
\text{CFU per ml} = \frac{\text{no. of colonies}^* \times \text{DF}}{\text{volume of culture plate}}
\]

Where DF means the dilution factor used.

*average of the number of the inoculation dose then the average of the replicates done
2.3 Lipopolysaccharide extraction and silver staining

An overnight culture (OD$_{600}$ of 0.8-1.2) of 5 ml LB broth, incubated at 37°C, was used to extract LPS using an LPS Extraction Kit (iNtRON Biotechnology). First, bacterial pellets were obtained by centrifugation at 16,000 xg for 10 minutes and all traces of supernatant were removed. Second, cells were lysed by adding 1 ml of lysis buffer, with vigorous vortexing. Then, 200 µl of chloroform was added and samples were vortexed vigorously for 10-20 seconds, and incubated at RT for 5 minutes. Chloroform was added to separate the phenol layer from the aqueous layer. Next, after centrifugation at 15,115 xg for 10 minutes at 4°C, two layers were formed with a separating white precipitate containing proteins and genomic DNA. The upper layer (400 µl) was transferred to a new 1.5 ml tube. To purify LPS from other components, 800 µl of purification buffer was added and samples were mixed well with incubation for 10 minutes at -20 °C, followed by centrifugation at 15,115 xg for 15 minutes at 4°C. The LPS pellet was washed with 1ml of 70% EtOH and centrifuged at 15,115 xg for 15 minutes. Finally, 70 µl of 10 mM Tris-HCl buffer (pH 8.0) was added to the LPS pellet, which was dissolved at 100°C in a heat block for 3 minutes. LPS extract was stored at -80°C.

LPS extract was analysed, along with positive control of LPS extract from E. coli (Sigma-Aldrich, UK), by electrophoresis on a Bolt™ SDS gel kit, with 4-12% Bis-Tris Plus Gels (Invitrogen™) using the conditions described in section 2.5.6. Following LPS extraction and SDS PAGE, the gel was stained by using the ProteoSilver™ Silver stain kit (Sigma-Aldrich, UK), following the manufacturer’s protocol.
2.4 Tissue culture

2.4.1 Maintaining epithelial cell culture

Epithelial cell line HT1197 (ECACC 87032403) (Public Health England-Culture collections, UK) was used in this study. The cells were maintained using MEM with glutamine, 1% non-essential amino acids (NEAA), and 10% Foetal Bovine Serum (FBS) (Thermo Fisher Scientific, UK). MEM needs to be supplemented with FBS as it lacks proteins and growth factors. Moreover, due to the use of sodium bicarbonate buffer system (2.2 g/L) in MEM, it requires a 5% CO₂ environment to maintain the medium pH. Cell culture flasks were incubated at 37°C until the cells reached 70-80% confluence (~3.5 days) when the cells were split using Trypsin/EDTA and seeded at 2x10⁴ to 5x10⁴ cells/cm² into new flasks.

2.4.2 Co-culture conditions

In this study we co-cultured the epithelial cells HT1197 with different UPEC strains. A single colony of each strain was inoculated in a 10 ml of LB broth and incubated overnight at 37°C. Then, 10 ml of AUM was inoculated with a 1:100 volume of overnight cultures for each strain and incubated at 37°C for 3 hours (mid log phase).

Cell culture flasks (75 cm²) were used to carry out the experiment, where cell estimations for each flask were ~2.5x10⁶ cells. All flasks were washed twice with Dulbecco's Phosphate Buffered Saline (DPBS) before adding E. coli. For multiplicity of infection (MOI) of 10 we need 2.5x10⁷ CFU of each strain. After calculating the volume required for each strain, it was inoculated into 13 ml of
MEM medium only (excluding FBS and NEAA). Then added to the flask and incubated at 37°C in CO₂ incubator for 5 hours (Wurpel et al., 2016).

During co-culture, each experiment was monitored every hour to observe any cytopathic effects. Following 5 hours incubation, media were collected into falcon tubes to be centrifuged at 5000 xg for 10 minutes. Both pellet (containing bacterial cells) and supernatant (containing secreted proteins – the ‘secretome’) were stored at -80°C for downstream analysis. Attached epithelial cells were washed twice with DPBS, as this is recommended to minimize the recovery of bacterial cells. Cells were then detached using Trypsin/EDTA and collected in 5 ml of MEM medium only then centrifuged at 134 xg for 10 minutes. Supernatants were discarded and pellets were stored at -80°C for downstream analysis.

In some cases, collecting the media at the beginning could result in collecting both the bacteria and epithelial cells. To separate them, the pellet was re-suspended in 2 ml MEM media only and transferred into an Ultrafree-CL centrifugal filter unit (Merck Millipore, UK) with a pore size of 5 µm. Following centrifugation at 3000 xg for 3 minutes, flow-through (containing bacterial cells) was collected into another tube and centrifuged at 5000 xg for 10 minutes. Pellets were stored at -80°C for downstream analysis.

Following collection of bacteria in the flow through, epithelial cells were lysed on the filter using 500 µl of 2% SDT lysis buffer [2%(w/v) SDS, 100mM Tris/HCl pH 7.6] and incubated with shaking for 15 minutes. Next, samples were centrifuged at 4000 xg for 40 minutes. To eliminate sample viscosity, 20 µl of Benzonase Nuclease (≥250 units/µl) (Sigma-Aldrich, UK) was added. Moreover,
3 µl/sample of Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, UK) were added and samples were stored at -80°C for downstream analysis.

2.4.3 Cells viability and MTT assay

In order to check for cell viability after being co-cultured with UPEC strains, the MTT assay was performed. Thiazolyl Blue Tetrazolium Blue (MTT) (Sigma-Aldrich, UK) was diluted in PBS with a final concentration of 5 mg/ml, then filtered with a 0.2 µm syringe filter and stored at 4°C in the dark, as it is light sensitive.

MTT creates a yellowish colour solution that is transformed to a dark blue colour. For co-culture experiments in 24 well plates, we added 45 µl/well of MTT at 3hrs, 4hrs, and 5hrs incubation periods. After incubation, excess volumes were discarded and 450 µl/well of DMSO was added. The absorbance for each plate was then measured at 560 nm and 670 nm with a PHERAsar FS plate reader (BMG LABTECH).

2.5 Protein extraction

For the media comparison study (LB and AUM) and the paired isolates experiment (LB only), growth culture media were used to grow UPEC strains for performing protein extraction, which was carried out at mid log phase for all strains (~3 hours). An overnight culture of 10 ml LB broth was used to inoculate 100 ml of LB and AUM broth, which were then incubated at 37°C for 3 hours with continuous shaking (~130 rpm). For all the samples, we extracted both
fractions of soluble and insoluble proteins. An overview of this process is provided in Figure 2.1.

2.5.1 Soluble protein extraction

Proteins were extracted using the B-PER complete bacterial protein extraction reagent (Thermo Fisher Scientific, UK). During the procedure, the bacterial suspension was centrifuged at 5000 \( xg \) for 10 minutes followed by addition of equilibrated reagents to 5 ml/g of biomass. Next, 2\( \mu \)L of 0.5M EDTA per mL of B-PER Complete Reagent and 10\( \mu \)L of Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, UK) were added per ml of sample to produce a 1X final concentration. After adding all the reagents, the suspension was mixed well until homogeneous and incubated for 15 minutes at room temperature with vortex every 5 minutes for 30 seconds each. Finally, the lysate was centrifuged at 16,000 \( xg \) for 20 minutes to separate soluble proteins from the insoluble proteins (pellet) and then soluble proteins were transferred into separate a tube and stored at -80°C.

2.5.2 Extracting insoluble proteins from the cell pellet

Following B-PER protein extraction, SDT lysis buffer [4\%(w/v) SDS, 100mM Tris/HCl pH 7.6, 0.1M DTT] was added to the insoluble pellets to extract all the proteins that were insoluble in B-PER. DTT was added to the sample after performing protein estimation, as it will interfere with the OD readings. SDT lysis buffer was added to samples in a ratio of 1:10. Next, samples were sonicated in a water bath at 4°C for 3-5 minutes. After pellets were dissolved,
they were centrifuged at 16,000 xg for 10 minutes and protein was transferred into separate tubes for storage at -80°C.

2.5.3 Epithelial cells protein extraction

Following the collection of the co-cultured HT1197 epithelial cells, we lysed them using 500 µl of 2% SDT lysis buffer [2%(w/v) SDS, 100mM Tris/HCl pH 7.6] and incubated, after gentle vortexing, for 15 minutes. Next, samples were centrifuged at 16000 xg for 20 minutes. Then, 20 µl of Benzonase Nuclease (≥250 units/µl) (Sigma-Aldrich, UK) was added to eliminate sample viscosity followed by 3 µl/sample of Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific, UK) were added and then stored at -80°C for downstream analysis.

2.5.4 Protein precipitation for secretome analysis

Trichloroacetic acid (TCA) was used to precipitate proteins collected in the supernatant of the co-culture experiment. TCA (110.132 g) was dissolved in 50 ml of dH2O and kept at RT. Then, TCA was added to a volume of 20% of the sample and kept on ice for 1.5 hours (incubation was also continued overnight where necessary). Next, samples were centrifuged at 3500 xg for 30 minutes at 4°C. Supernatants were discarded and pellets were transferred into another tube and centrifuged again at 16000 xg for 30 minutes. Following that, pellets were washed twice with 75% Acetone to remove the TCA. Samples were then dried in a speed-vac at 30°C for 30 minutes and re-suspended well in 100 µl of 8M urea buffer. Samples stored at -80°C for downstream analysis.
2.5.5 Protein estimation

Following protein extraction, protein estimation was carried out using the Pierce™ BCA (Bicinchoninic Acid Assay) (Thermo Fisher Scientific, UK). The microplate procedure was followed, as per manufacturer’s instructions, with a standard prepared using the B-PER reagent and SDT lysis buffer as diluent for soluble and insoluble protein extractions, respectively. Absorbance was measured at 562 nm with a PHERAnstar FS plate reader (BMG LABTECH). Using MARS data analysis software (BMG LABTECH) analysis of readings was conducted and standard curves generated.

2.5.6 SDS gel electrophoresis and Coomassie Brilliant blue stain

Extracted proteins were separated using a Bolt™ SDS gel kit, with 4-12% Bis-Tris Plus Gels (Invitrogen™). Samples were prepared as in Table 2.3 with 20 to 50 µg of sample. After preparation, samples were heated at 70°C for 10 minutes. Then, they were loaded onto the gels and 10 µl of SeeBlue® Plus2 Pre-Stained Protein Standard (Life Technologies) was also used. Run conditions for the gel were as follows; 160V for 35 minutes.

For gel staining, one gram of Brilliant blue stain was added to prepare 1 L of the stain solution. Coomassie de-stain solution was prepared as follows: 50% (v/v) Methanol, 10% (v/v) Acetic Acid, and 40% (v/v) dH₂O. The gel was stained with Coomassie stain for one hour followed with 3-4 de-stain steps for 30 minutes each. Lastly, gels were washed with 50% de-stain solution for 30 minutes and then kept in 10% Acetic acid for storage at 4°C. In some experiments, InstantBlue stain (Expedeon Ltd, UK) was used for rapid detection.
of protein bands. The gel was stained for one hour then washed and maintained in dH$_2$O.

Table 2.3: Sample preparation for SDS PAGE gel electrophoresis*.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Reduced Sample</th>
<th>Non-Reduced Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>X µl</td>
<td>X µl</td>
</tr>
<tr>
<td>Bolt™ LDS Sample Buffer (4X)</td>
<td>10 µl</td>
<td>10 µl</td>
</tr>
<tr>
<td>Bolt™ Reducing Agent (10X)</td>
<td>4 µl</td>
<td>____</td>
</tr>
<tr>
<td>Deionized Water</td>
<td>Up to 26 µl</td>
<td>Up to 30 µl</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td>40 µl</td>
<td>40 µl</td>
</tr>
</tbody>
</table>

* Adapted from Bolt™ 4-12% Bis-Tris Plus Gels Kit.

2.6 Proteomics

2.6.1 Sample preparation

A filter aided sample preparation (FASP) method was used to prepare the samples for MS (Wisniewski et al., 2009). A total of 20-150 µg of estimated protein was diluted with UA buffer (8 M urea in 0.1 M Tris/HCl pH 8.5) to a minimum volume of 300µl and passed through a filter unit [Amicon Ultra-0.5ml (10kDa cut off), Merck Millipore] by centrifugation at 16,160 xg for 15 minutes. This step was repeated with longer centrifugation times (~20 minutes) to make sure the detergent concentration was minimal. Second, 4 µl of 1 M DTT (Dithiothreitol, GE Healthcare) was added and the volume made up to 400 µl with UA buffer and samples were incubated at RT for 30 minutes then centrifuged at 16,160 xg for another 30 minutes. Next, 40 µl of 50 mM CAA
(Chloroacetamide) was added and the volume made up to 400 µl with UA buffer and incubated at RT in the dark for 20 minutes, then centrifuged at 16,160 xg for another 30 minutes. After that, the filter was washed twice with UA buffer and centrifuged at 16,160 xg. Later, 1:50 of Lys-c enzyme (Lysyl Endopeptidase, Wako Bioproducts, Neuss, Germany) was added and samples were incubated at RT overnight. The next day, 50 mM ABC buffer (Ammonium Bicarbonate) was added to dilute the samples to ≤ 2 M of urea. Next, 1:100 of Trypsin enzyme (Promega) was added and samples were mixed on thermo mixer at 6500 rpm for 1 minute and incubated at RT overnight. The following day, collection tubes for the filter units were changed and samples centrifuged at 16,160 xg for 30 minutes. Next, 50 µl of 0.5 M NaCl was added and samples were centrifuged at 16,160 xg for 20 minutes. After that, 50 µl of 2% trifluoroacetic acid (TFA) was added to acidify the filtrate and stop the digestion reaction. Sample acidity was checked with pH indicator (pH ≤ 3) before moving to the next steps.

Small amounts of peptide digest for LC-MS analysis can be desalted using StageTips (Rappsilber et al., 2007). Two plugs (1mm diameter) punched out of C18 discs (EMPORE ™ - 3M) were packed to load 10 µg of digested proteins. Firstly, each StageTip was conditioned with 100 µl of 100% methanol. For each step in the centrifuge, 1475 xg was used for 5 minutes. If after centrifugation a small volume still remains in the tip, centrifugation can be repeated by increasing the time and optimising according to experimental condition (this was applied for all steps). Next, 100 µl of buffer B (80% acetonitrile in 0.5% acetic acid) was added and StageTips were centrifuged. Then 100 µl of buffer A (0.5% Acetic Acid) was added and StageTips were centrifuged; this step was repeated twice. After washing the membrane, samples were added and the assembly
centrifuged as described above. Then, the samples were eluted with 25 µl of buffer B (twice) into new collection tubes. Lastly, samples were dried in a speed-vac at 30°C for 25 minutes. Concentrated samples were transferred to glass micro-vials for storage 4°C.

2.6.2 Mass spectrometry analysis

This work was carried out by the staff in the University of Plymouth Systems Biology Centre. Peptides were separated on a Dionex Ultimate 3000 RSLC nano flow system (Dionex, Camberly UK). An aliquot (3 µl) of sample was loaded in TFA/acetonitrile (2% acetonitrile in 0.1% TFA) onto an Acclaim Pep Map 100 µm × 2 cm, 3 µm C18 nano trap column, at a flow rate of 5 µl/min, bypassing the analytical column. Elution of bound peptides was performed with the trap column inline with an Acclaim PepMap C18 nano column 75 µm × 25 cm, 3 µm, 100 Å (Analytical Column) with a linear gradient of 96% buffer A and 4% buffer B to 60% buffer A and 40% buffer B, at a constant flow rate of 300nl/min over 240 minutes. The sample was ionized in positive ion mode using a Proxeon nano spray ESI source (Thermo Fisher Scientific, UK) and analysed in an Orbitrap Velos Pro FTMS (Thermo Finnigan, Bremen, Germany). The Orbitrap Velos Pro instrument, running Xcalibur2.1 software, was operated in the data dependent mode to automatically switch between MS and MS/MS acquisition. MS spectra of intact peptides (m/z 350-1600) with an automated gain control accumulation target value of 1000000 ions were acquired with a resolution of 60000. The ten most intense ions were sequentially isolated and fragmented in the linear ion trap by collision induced dissociation (CID) at a target value of 10,000 or maximum ion time of 200 ms. A dynamic exclusion of
ions previously sequenced within 45” was applied. All the singly charged and unassigned charge state ions were excluded from sequencing. Typical mass spectrometric conditions were: spray voltage, 2.3 kV; no sheath and auxiliary gas flow; heated capillary temperature, 275°C; normalized CID collision energy 30% for MS2 in LTQ. The ion selection threshold was 10000 counts for MS2. An activation q = 0.25 and activation time of 30 ms were used.

Peptides and proteins were identified by Andromeda via automated database searching of all tandem mass spectra against a curated target/decoy database, forward and reversed version of the *E. coli* K12, UPEC 536, and UTI89 protein sequence database (http://www.uniprot.org/ UniProt. Release July 2015) containing all *E. coli* protein entries from Swiss-Prot and TrEMBL. Spectra were initially searched with a mass tolerance of 6 ppm in MS mode and 0.5 Da in MS/MS mode and strict trypsin specificity and allowing up to 2 missed cleavage sites. The resulting Andromeda peak list-output files were automatically loaded into inbuilt MaxQuant software modules for further processing, label free quantification (LFQ) and a maximum false discovery rate of 1% were fixed for the result output files (Cox et al., 2014).
Figure 2.1: General work flow chart of the methodology followed for proteomics studies and mass spectrometry analysis.
2.6.3 Proteomics analytical approach

Raw files obtained after mass spectrometry analysis were provided by the Systems Biology Centre and were processed using MaxQuant. All proteomics data of this thesis are attached in Appendix VII. Following MaxQuant, we processed the data through multiple softwares and online databases for data analysis including; Perseus software (Tyanova et al., 2016), GProX software (Rigbolt et al., 2011), Microsoft Excel, STRING database (Szklarczyk et al., 2017), and DAVID tool (Huang da et al., 2009; Huang et al., 2008a). An overview of these tools is given below.

2.6.3.1 Perseus software

Perseus has a comprehensive collection of statistical tools for high-dimensional omics data analysis including value normalization, pattern recognition, cross-omics comparisons, and multiple-hypothesis testing (Figure 2.2). In this study, we used Perseus software v1.6.1.3 (http://www.perseus-framework.org) as the main tool for the data analysis (Tyanova et al., 2016).

The software has five main sections: Loading the data; Processing; Analysis; Multi-processing; and data Export (Figure 2.3). At the start, we uploaded the output file of MaxQuant as a tab-delimited .txt file, which has a header row with the names of all columns. The type of data is specified during file loading into three main columns categories; Main columns [contain expression columns, e.g. LFQ intensity (Cox et al., 2014), iBAQ (Schwanhausser et al., 2011), or ratios of samples], Numerical columns (all additional numerical data), and Text columns (identifiers such as protein IDs) (Tyanova & Cox, 2018).
Figure 2.2: The Perseus data analysis platform. The core data structure of Perseus is the data matrix, containing samples and expression values (e.g., protein, mRNA). Perseus incorporates data filtration and normalization, and multiple methods for examining analysis such as histogram charts, intensity curves and scatter plots. Statistical tools including t-tests, PCA, correlation analysis and enrichment analysis support classical expression “omics” data analysis. Adapted from (Tyanova et al., 2016).
Following data loading comes data processing, where firstly all expression columns were $\log_2$ transformed. Then, all rows were categorically annotated by annotating all the samples belonging to one condition with the same annotation in one group. Following that, data filtration is performed by filtering rows based on valid values, where are minimum number of valid values in each group are two. This step will reduce the data matrix to only valid groups. Then, we performed data imputation by replacing missing values from normal distribution. ANOVA test was done for matrix data using permutation-based FDR (FDR 0.05) for truncation. Although data normalization is not always required, different types of normalization can be performed on the data to make it comparable (Tyanova & Cox, 2018).

Regarding the analysis step, multi scatter plots were generated to validate the replicate samples for data correlations. Moreover, after performing further data treatment in Excel (section 2.6.3.2), we visualized the data using Hierarchical clustering, with regard to using Euclidean distance measure and complete linkage measure to calculate the squared Euclidean distance (Yilm & Ramdeen, 2015). It is recommended to preserve the order of the columns tree (samples) but not the rows (proteins) in order to have clear clusters of the data.

Finally, the matrix data were exported and saved as a tab-delimited .txt file, where it can be read using any spreadsheet software (e.g. Microsoft Excel).
2.6.3.2 Microsoft Excel software

Microsoft Excel is a common spreadsheet program used to create grids of text, numbers and formulas specifying calculations. It is extremely useful for reading the output text file of Perseus and MaxQuant.

In this study, we used Excel 2011 to organise and highlight the up-regulated and down-regulated proteins in colour codes in addition to perform a number of statistical calculations. Originally, all protein identifiers come as Fasta headers from the MaxQuant output as in the following format:

```
>db|UniqueIdentifier|EntryName  ProteinName OS=OrganismName
OX=OrganismIdentifier  [GN=GeneName]PE=ProteinExistence
SV=SequenceVersion
```
Where db is either ‘sp’ for UniProtKB/Swiss-Prot or ‘tr’ for UniProtKB/TrEMBL and UniquelIdentifier is the primary accession number of the UniProtKB entry.

In this study, we focused on using the Unique Identifier (accession number; AC) and the Gene Name for the downstream data analysis. The AC is allocated to each sequence upon insertion into UniProtKB. Accession numbers are constant throughout different version releases (The UniProt Consortium, 2017). However, using only AC as a protein Identifier in some databases (e.g. STRING) is not suitable, as it lacks an updated list from the Uniprot database. Thus, using gene names was more appropriate, where the majority of the protein entries in the list were recognised.

Examples of the uses of Excel in this study include generating a ratios column of LFQ values, inverse of the function of log₂, and finding the relative expression values. Relativity in expression data can be found out by dividing the average LFQ value of a replicate by the highest value among the samples.

### 2.6.3.3 Venn diagrams

A Venn diagram is a diagram that displays all possible logical relations between a limited collection of different data sets. Venny 2.1 was the primary tool used for the overlap analysis in this study (Oliveros, 2015).

### 2.6.3.4 GraphPad Prism

The main statistical software used for data analysis was GraphPad Prism v6 ([https://www.graphpad.com/scientific-software/prism/](https://www.graphpad.com/scientific-software/prism/)). It was also used to generate all the graph figures presented in this study.
2.6.3.5 GProX software

The Graphical Proteomics Data Explorer (GProX) is a freely available complete software platform for comprehensive and combined analysis and visualization of large proteomics datasets. The program provides a range of data distributions visualization tools such as heatmaps, scatter plots, and histograms (Rigbolt et al., 2011).

GProX v1.1.16 (http://gprox.sourceforge.net/) was used in this study. Mainly it functions similarly to Perseus but generates higher quality output figures. Data are uploaded as tab-delimited .txt files, where first row must contain column headers (non-redundant). The AC must be present for all protein ids (Rigbolt et al., 2011).

2.6.3.6 STRING database

STRING is a database of identified and predicted protein-protein interactions (von Mering et al., 2003). The interactions include physical (direct) and functional (indirect) associations; they originate from computational prediction, from information transfer between organisms, and from interactions accumulated from other (primary) databases (Szklarczyk et al., 2015).

Data sources of interactions in STRING are obtained from five main sources; genomic context predictions, high-throughput lab experiments, conserved co-expression, automated textmining, and prior information in databases (Szklarczyk et al., 2017).

We used STRING database v10.5 (https://string-db.org/) to analyse the clusters generated by Perseus software. Using gene names, clusters of proteins
were loaded into STRING. *E. coli* CFT073 (NCBI:txid199310) data was used as a reference strain due to it being a UPEC strain and because a high number of proteins were matched with the samples (distinct protein-coding genes: 5369). Following loading of the protein list, an interaction map is generated. The interaction network edges are based on the confidence of data (line thickness indicates the strength of data support). Network analysis can show functional enrichment in process such as biological process, molecular functions, and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Kanehisa et al., 2010). All data can be exported as tab-delimited .txt file format.

### 2.6.3.7 DAVID tool

The Database for Annotation, Visualization and Integrated Discovery (DAVID) is an online tool that incorporates functional genomic annotations with graphical summaries (Dennis et al., 2003; Huang da et al., 2009; Huang et al., 2008a). DAVID can interpret genome-scale datasets by simplifying the conversion from data assembly to biological meaning.

In this study, we used DAVID v6.7 ([https://david-d.ncifcrf.gov](https://david-d.ncifcrf.gov)) for our functional annotation analysis. AC for proteins of interest were used to upload the data. Moreover, options such as Fold Enrichment, Bonferroni, Benjamini, and FDR were selected in the analysis. Data can be downloaded as tab-delimited .txt file format.
2.6.3.8 Proteomics data analysis pipeline

To summarise, the raw data files were obtained from LC-MS/MS to be processed in MaxQuant. The output files underwent further processing in Perseus for quality control. Following confirmation of the data integrity, expression data were averaged on Microsoft Excel to examine the fold change (Chapters 3 and 5) or the relative expression (Chapter 4). The data were then uploaded again into Perseus to generate hierarchical clusters, which were later displayed using Venn diagrams. Next, protein groups of interest were investigated using Functional Annotation Enrichment with DAVID tool (Chapters 3 and 5) or through Protein-Protein Interaction (PPI) network analysis on STRING (Chapter 4) (Figure 2.4).

Figure 2.4: General pipeline chart of the methodology followed for proteomics data analysis.
2.7 Genomic analysis

2.7.1 Sample preparation and DNA sequence analysis

Genome sequencing was provided by MicrobesNG (Birmingham, UK) ([http://www.microbesng.uk](http://www.microbesng.uk)), which is supported by the BBSRC (grant number BB/L024209/1). A single colony of the strain to be sequenced was mixed in 100 µl sterile PBS then cultured on LB agar. All bacterial biomass was collected off the plate and mixed into a bead tube (Pro-Lab Diagnostics, UK) provided by MicrobesNG. Leaving the cryopreservative solution in, tubes were sealed and send at room temperature to MicrobesNG. The turnaround time for generating raw genome sequence data was 10-12 weeks.

At MicrobesNG, DNA was extracted directly from an aliquot taken from the tube. Illumina MiSeq and HiSeq 2500 platforms were used to perform all sequencing. Samples were sequenced using 2x250bp paired-end reads. The output of the analysis was provided in fasta, gbk, gff, and fastq file formats.

MicrobesNG processed the data through a standard analysis-sequencing pipeline. This started with the Kraken programme ([https://ccb.jhu.edu/software/kraken/](https://ccb.jhu.edu/software/kraken/)) (Wood & Salzberg, 2014) to identify the nearest associated genome, followed by mapping the reads to a reference genome in order to evaluate the data quality using BWA mem 0.7.15 ([http://bio-bwa.sourceforge.net/](http://bio-bwa.sourceforge.net/)) (Li & Durbin, 2010). Next, SPAdes 3.8.1 ([http://bioinf.spbau.ru/spades](http://bioinf.spbau.ru/spades)) was used complete the de novo assembly of the reads (Nurk et al., 2013). Lastly, the contigs were ordered and annotated using MUMmer 3.0 ([http://mummer.sourceforge.net/](http://mummer.sourceforge.net/)) (Kurtz et al., 2004) and Prokka 1.1 ([http://www.vicbioinformatics.com/software.prokka.shtml](http://www.vicbioinformatics.com/software.prokka.shtml)) (Seemann, 2014), respectively.
2.7.2 Bioinformatics analysis

Contigs were ordered according to the complete genome of UPEC ST127 reference strain 536 [Accession ref NC_008253; (Hochhut et al., 2006)] and annotated using Prokka v1.12 (Prokka: Prokaryotic Genome Annotation System - http://www.vicbioinformatics.com/software.prokka.shtml). This work was done in collaboration with Dr Majed Alghoribi (King Abdullah International Medical Research Center, Saudi Arabia).

Multiple software tools including BLAST (Altschul et al., 1990), Artemis comparison tool (ACT) (Carver et al., 2005), BLAST Ring Image Generator (BRIG 0.95-dev0004) (Alikhan et al., 2011), Easyfig (Sullivan et al., 2011) and Tablet (Milne et al., 2013) were used to compare and visualize UPEC genomes.
Chapter Three: Investigating the Effect of Growth in Different Environments on Phenotypic and Proteomics Profiles of UPEC
3.1 Introduction

Uropathogenic *E. coli* is the most commonly known cause of UTI, being responsible for up to 85% of community-acquired infections and 40% of hospital-acquired cases (Toval *et al.*, 2014). UPEC strains possess many virulence factors that add to their capacity to cause infection (Bien *et al.*, 2012). In addition, the global spread of multi-drug resistant UPEC strains is posing a critical risk to treatment (Warnes *et al.*, 2012).

The most virulent UPEC strains are descended from major lineages, such as ST38, ST127 and ST131, some of which have become globally disseminated (Gibreel *et al.*, 2012b). This project is investigating different isolates of the ST127 lineage. One strain in particular (strain EC18) has a deletion in the O-antigen cluster, which results in attenuation of virulence toward the *Galleria mellonella* insect larvae infection model (Alghoribi *et al.*, 2014). The O antigen is a major component of LPS (Wang & Quinn, 2010), which means there could be a complete or partial absence of LPS in *E. coli* EC18. It is not clear how this may affect the pathogenicity of strain EC18. The study is also investigating *E. coli* EC41 and SA189, a UPEC ST127 isolates of clinical UTI with intact O-antigen gene clusters, as comparison strains to EC18.

Proteomics allows investigation of changes in relative abundance of proteins in different conditions. Bacteria may exhibit very diverse protein expression profiles in various parts of the human body, in various phases of the life cycle and under various environmental conditions (Cox *et al.*, 2014; Krug *et al.*, 2013). Deciphering these changes can inform our understanding of bacterial pathogenesis and proteomic methods have been applied here to investigate gene expression profiles in isolates of the ST127 lineage. This study will
demonstrate the effect of environmental changes on the metabolic pathways expressed by a specific strain. Proteomic based analysis of UPEC strains will significantly enhance our understanding of the host-pathogen interaction of UPEC isolates in different environments, from a nutrient-limited environment (Urine) to a richer environment (Blood); and may facilitate development of novel therapies for UTI.

3.1.1 Aim and objectives

The overall Aim of this chapter is to establish the use of proteomic approach for investigating the pathogenesis of UPEC strains. To achieve this aim, the following objectives were set:

- Establish methods for reproducible growth of several strains of ST127 in a range of laboratory media.
- Extract LPS from strains *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 for comparison of LPS structure.
- Ensure protein extraction yields were reproducible from different strains under a range of conditions.
- Prepare extracted proteins for analysis on the OrbiTrap MS platform.
- Carry out bioinformatics analysis of proteomics data.
3.2 Experimental design

In this chapter we have three UPEC strains of ST127 (EC18, EC41, and SA189). Each strain will be grown in two different conditions (LB and AUM). Each sample will have two protein fractions (soluble and insoluble). Two individual biological replicates were done for each sample in this project, which will bring the total number of samples to 24 samples (Figure 3.1).

Figure 3.1: Experimental design of the differential proteomics study. The figure indicates the sub category samples for each strain.
3.3 Results

3.3.1 Growth kinetics

To investigate if the O-antigen deficient strain *E. coli* EC18 is affected in its ability to grow in a rich medium or not compared to control strains *E. coli* EC41 and *E. coli* SA189 (with intact O-antigen gene clusters), growth assays were performed. All strains were grown in LB and AUM without any antibiotic selection. Over 18 hours incubation, the OD measurement was recorded for growth in LB and AUM (Figure 3.2).

Comparing the growth curves within a single medium, using 2-way ANOVA analysis, there was a statistically significant difference \((p \leq 0.0001)\) between SA189 and other strains in LB media for the last three time points. However, all strains in AUM were at a similar density after 18 hours growth. On the other hand, it is clear from the graph that there is a significant difference \((p \leq 0.0001)\) between growth in LB and AUM for all three strains, starting from the 2 hours’ time point. In addition, analysing the growth curves, it is apparent that the mid-log phase for all strains grown in either LB or AUM is at \(\sim 3\) hours of incubation.

Colony forming units were also calculated for all three UPEC strains in both media. In LB media, EC18 showed a statistically significant difference \((p \leq 0.01)\) to EC41 in the last three time points and to SA189 at the 18 hours’ time point (Figure 3.3 LB). Moreover, for the last three time points, EC18 and EC41 CFU variations in AUM were highly significant \((p \leq 0.001)\). In addition, EC18 and SA189 were only significantly different \((p \leq 0.01)\) at the 18 hours’ time point (Figure 3.3 AUM).
Figure 3.2: Growth kinetics of *E. coli* EC18, EC41 and SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM). Cultures of all strains were grown in LB broth and AUM without any selection. OD600 measurements were taken at zero minutes, 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours, and 18 hours. Data are plotted as the mean of three biological replicates and error bars represent standard error of the mean (SEM). *p*-Value represented as (**** = *p*≤0.0001, *** = *p*≤0.001, ** = *p*≤0.01, * = *p*≤0.05) for *E. coli* SA189 compared to *E. coli* EC18 and EC41 in LB.
Figure 3.3: Colony forming unit (log10) of *E. coli* EC18 and the control strains *E. coli* EC41 and *E. coli* SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM). Duplicate cultures of all strains were grown in LB broth and AUM, then incubated on LB agar plates without any selection. Colonies counted at zero minutes, 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours, and 18 hours. Data are plotted as the mean and error bars represent standard error of the mean (SEM).
3.3.2 LPS detection and analysis

To investigate the LPS structure, LPS was extracted from all strains and compared on SDS PAGE gels (Figure 3.4). The results confirmed there is a partial absence of LPS structure (O antigen substituted LPS) in *E. coli* EC18 compared to *E. coli* EC41 and SA189. The missing O-antigen repeats for EC18 are highlighted in the red box (Figure 3.4). Nevertheless, lipid A core oligosaccharide was shown in all strains (Alexander & Rietschel, 2001).

![Figure 3.4: Silver stained SDS-PAGE analysis of LPS extractions for E. coli EC18, E. coli EC41, and E. coli SA189. LPS extract from E. coli (Sigma-Aldrich, UK) was used as a positive control. Arrows show LPS core and O-antigen repeats. Different dilutions were used for each sample.](image)
3.3.3 Protein extractions

3.3.3.1 Soluble proteins

All samples were cultured from a single colony in 10 ml LB broths and incubated at 37°C overnight. Extraction was performed at mid log phase (~ 3 hours) on 100 ml of media per sample, as extracting proteins at further time points will result in a lower concentration of proteins extracted (Figure 3.5). Soluble proteins were extracted using the B-PER complete bacterial protein extraction reagent. BCA protein estimation indicated higher protein extraction yield in both media for EC18 than the other strains (Figure 3.6). Moreover, EC18 was statistically significant in unpaired t test analysis compared to SA189 in LB ($p = 0.0061$) and to EC41 in AUM ($p = 0.0224$). On the gel, it is clear that EC18 protein extractions in both media had more consistent results with well-resolved protein bands visible (Figure 3.7A).

3.3.3.2 Insoluble proteins

As results from the soluble protein extractions indicated that a large proportion of the protein might have been retained in the pellet, we sought to extract these proteins using SDT lysis buffer. Following extraction from pellets, BCA estimation showed higher insoluble protein extraction yields compared to the soluble ones, with statistically significant differences in the AUM samples (Figure 3.6), which was observed on SDS PAGE analysis (Figure 3.7B). There were no statistically significant variations between insoluble proteins for the same medium.
Figure 3.5: BCA estimation for soluble protein extracted from *E. coli* EC41 in artificial urine medium (AUM). Soluble proteins extracted by B-PER complete bacterial protein extraction reagent at different time points. Data are plotted as the mean of two biological replicates and error bars represent standard error of the mean (SEM).

Figure 3.6: BCA estimation for protein yield extracted from *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM). Data are plotted as the mean of two biological replicates and error bars represent standard error of the mean (SEM). *p*-Value represented as (**** = $p \leq 0.0001$, *** = $p \leq 0.001$, ** = $p \leq 0.01$, * = $p \leq 0.05$, n.s. = not significant) for the soluble and insoluble proteins fractions comparison of each sample.
Figure 3.7: Protein extractions for *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 loaded on SDS PAGE gels. (A) Soluble proteins extracted by B-PER complete bacterial protein extraction reagent; 30 µg of protein was loaded for all samples except for EC41 LB and SA189 LB (20 µg). (B) Insoluble proteins extracted by SDT lysis buffer. 50 µg of protein was loaded for all samples. InstantBlue staining was used.

### 3.3.4 Mass spectrometry and data analysis

As mentioned above, all the samples were subdivided into two fractions: Soluble and Insoluble and further processed for mass spectrometry analysis. The output data were processed using MaxQuant (1.5.3.1) and the Perseus software package. The data indicated a good correlation between replicates of ≥ 0.9 (Figure 3.8). Raw files from soluble and insoluble fractions were merged to have the overall proteomic profile of the strain in each specific environment. Figure 3.9 showed the overlapped protein of both extraction methods (soluble and insoluble) in each culture condition of each strain.
Figure 3.8: Correlation matrix of label free quantification (LFQ) of biological replicates of UPEC ST127 isolates (EC18, EC41, and SA189) grown in LB and AUM. Significant correlations of same condition replicates are represented in red squares.
Hierarchical clustering (HCL) in Perseus software, based on the level of protein detection, showed different label-free quantification (LFQ) clustering between LB and AUM samples (Figure 3.10). In addition, strain EC18 clustered separately from the O-antigen positive strains in both media.

A total of 1702 proteins were identified through interrogation of the *E. coli* K12 and UPEC strains (UTI89 and 536) database. However, only 747 proteins were quantified across all the conditions due to existence in one condition and absence in the other, which means that a ratio could not be obtained. The remaining proteins (747) were quantified and provide this study with many insights into the proteome profile of these strains of UPEC.

Examining the HCL of label-free protein quantification (LFQ) intensities for ST127 strains (EC18, EC41, and SA189), we can see that 747 proteins were clustered into six clusters with a distance threshold of 3.86 (Figure 3.10). These clusters are different in size, ranging from 313 proteins to 63 proteins (Table 3.1).
Figure 3.9: Overlaps between soluble and insoluble proteins in UPEC ST127 strains. The indicated proteins are the soluble proteins (blue) and insoluble proteins (yellow) of UPEC EC18, EC41, and SA189 in Lysogeny Broth (LB) and artificial urine medium (AUM). Venny 2.1 tool was used to produce the figures (Oliveros, 2015).
3.3.4.1 Expression fold change analysis

With a closer look at the row clusters (Figure 3.11), we can see what we describe in the heat map in addition to more detailed information. Cluster 740 showed spike in protein expression for EC41 in LB, while clusters 730 and 733 has spiked in protein expression for EC18 in AUM and EC18 in LB, respectively. Moreover, protein expression is elevated for isolates grown in LB,
in cluster 739. However, cluster 737 showed raised expression in proteins of isolates in the AUM condition.

Table 3.1: Row clusters (n=6) of HCL of label-free protein quantification (LFQ) intensities for ST127 strains (EC18, EC41, and SA189) in LB and AUM. Size indicates the number of proteins in the cluster.

<table>
<thead>
<tr>
<th>Cluster No.</th>
<th>740</th>
<th>739</th>
<th>738</th>
<th>737</th>
<th>733</th>
<th>730</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>93</td>
<td>63</td>
<td>313</td>
<td>99</td>
<td>98</td>
<td>81</td>
</tr>
</tbody>
</table>

Figure 3.11: Row clusters (n=6) of HCL of label-free protein quantification (LFQ) intensities for ST127 strains (EC18, EC41, and SA189) in LB and AUM. Colour scale ranges from 0.0 (red) to >3.5 (light green) indicating distance from centre.
EC18 has significantly less up-regulated proteins than the other strains, while it has the most down-regulated proteins among the ST127 strains examined (Table 3.2). On the other hand, examining the up-regulated and down-regulated proteins numbers using Venny 2.1 tool (Oliveros, 2015) give us insight into the proteome profiles of the strains. In general, EC41 and SA189 have more than 50% of their proteome profiles differentially regulated. In fact, EC41 and SA189 have more shared regulated proteins compared to EC18. EC41 and SA189 share 179 (23.9%) up-regulated proteins and 110 (14.7%) down-regulated proteins (Figure 3.12). A list of intersecting proteins of Figure 3.12 are listed in Appendix I.

Table 3.2: Up-regulated and down-regulated proteins (>2 fold) in all strains of ST127. The numbers indicates the differentially regulated proteins in LB compared to AUM.

<table>
<thead>
<tr>
<th></th>
<th>EC18</th>
<th>EC41</th>
<th>SA189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated proteins</td>
<td>123</td>
<td>258</td>
<td>253</td>
</tr>
<tr>
<td>Down-regulated proteins</td>
<td>195</td>
<td>158</td>
<td>168</td>
</tr>
<tr>
<td>Total</td>
<td>318</td>
<td>416</td>
<td>421</td>
</tr>
<tr>
<td>Percentage</td>
<td>42.6%</td>
<td>55.7%</td>
<td>56.4%</td>
</tr>
</tbody>
</table>
Figure 3.12: Overlaps between differentially altered proteins in three strains of ST127. The indicated proteins are the up-regulated (A) and down-regulated (B) proteins in Lysogeny Broth (LB) compared to artificial urine medium (AUM). Venny 2.1 tool was used to produce the figures (Oliveros, 2015).
3.3.4.2 Enrichment analysis

Pathway Enrichment Analysis was performed on up-regulated (Figure 3.12A) and down-regulated (Figure 3.12B) proteins using the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool (Huang da et al., 2009; Huang et al., 2008b).

Figure 3.13 shows pathways that were enriched by growth in LB, compared to AUM. A cut off of up-regulation by ≥2 fold with \( p < 0.05 \) was used. From the pathways that were up-regulated in LB, an overlap analysis was carried out using the Venny tool (Oliveros, 2015), which shows that there are six commonly shared pathways in all strains (Table 3.3). However, no pathways were exclusively up-regulated in EC18 nor shared exclusively with another strain. On the other hand, EC41 and SA189 showed exclusive pathway enrichment for three and two pathways, respectively. Moreover, they shared exclusively five pathways (Table 3.3).

Pathways enriched by down-regulation during growth in LB (≥2 fold with \( p < 0.05 \)) are illustrated in Figure 3.14. The overlap analysis of the pathways using Venny tool shows the three common shared pathways. EC18 shared exclusively three down-regulated pathways with SA189, and only one with EC41 (Table 3.4). In addition, five pathways were exclusively down-regulated in EC18. On the other hand, only two pathways were exclusively down-regulated in EC41, in addition to the down-regulation of the two-component system pathway in SA189 (Table 3.4).
Figure 3.13: Pathways enriched in proteins up-regulated by ≥2 fold during growth in Lysogeny Broth compared to artificial urine medium. Also shown is the overlap in up-regulated pathways for each of the three UPEC strains.
Table 3.3: Pathways enriched in up-regulated proteins (≥2 fold with \(p<0.05\)) in LB vs. AUM.

<table>
<thead>
<tr>
<th>Pathways</th>
<th>EC18</th>
<th>EC41</th>
<th>SA189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Aminoacyl-tRNA biosynthesis</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>RNA polymerase</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Ribosome</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Glycolysis / Gluconeogenesis</td>
<td></td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Trinitrotoluene degradation</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Fatty acid biosynthesis</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Propanoate metabolism</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial secretion system</td>
<td></td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>Protein export</td>
<td></td>
<td>↑</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.14: Pathways enriched in proteins down-regulated by ≥2 fold during growth in Lysogeny Broth compared to artificial urine medium. Also shown is the overlap in down-regulated pathways for each of the three UPEC strains.
Table 3.4: Pathways enriched in down-regulated proteins (≥2 fold with $p<0.05$) in LB vs. AUM.

<table>
<thead>
<tr>
<th>Pathways</th>
<th>EC18</th>
<th>EC41</th>
<th>SA189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Glycine, serine and threonine metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Alanine, aspartate and glutamate metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>ABC transporters</td>
<td></td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Cysteine and methionine metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Glycolysis / Gluconeogenesis</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Glutathione metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Glyoxylate and dicarboxylate metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Methane metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Protein export</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>RNA degradation</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Two-component system</td>
<td></td>
<td></td>
<td>↓</td>
</tr>
</tbody>
</table>
3.3.5 Genomic analysis of UPEC ST127 strains

Genomic analysis was carried between UPEC strains EC18, EC41, SA189, and UPEC ST127 reference strain 536 (Hochhut et al., 2006).

The analysis indicated an insertion-sequence (IS1) mediated deletion (from galE to wcaA) within the O-antigen and colanic acid gene clusters in EC18. Many studies have indicated that the O-antigen gene cluster is encoded in the region between the essential genes galF and gnd (DebRoy et al., 2011; Reyes et al., 2012).

As a result of this insertion, the majority of the O-antigen gene cluster was deleted. This has resulted in a break in the sequence assembly and showed evidence of an insertion sequence (IS1) at that position (Figure 3.15)(Figure 3.16).

In comparison, the O-antigen cluster was completely assembled into a single contig in the EC41 and SA189 assemblies (Figure 3.16). Furthermore, a whole genome comparison between all the UPEC strains and the complete genome of UPEC 536 was performed. This was done in order to make sure that the deleted EC18 O-antigen genes were not present elsewhere in the genome.

In addition, the hlyA gene was not identified in the genome sequence of UPEC EC18 nor EC41. Nevertheless, UPEC SA189 genome possessed this gene. The significance of this finding will be discussed later in Chapter 5.
Figure 3.15: BLAST ring image of UPEC isolates of ST127 strains EC18, EC41, and SA189. UPEC 536 was used as a reference strain. O antigen is indicated in red, demonstrating it is absent in EC18.
Figure 3.16: BLASTn comparison of O-antigen gene cluster (Green) and colanic acid gene cluster (Red) of UPEC EC18, EC41, SA189 and UPEC ST127 reference strain 536. The grey shading indicates high nucleotide identity between the sequences (99–100%). In EC18, there is an insertion-sequence (IS1) mediated deletion of most of the O-antigen gene cluster and the colanic acid gene cluster compared to other UPEC strains (vertical dotted line denotes contig boundaries in EC18). Easyfig was used to prepare this figure (Sullivan et al., 2011).
3.4 Discussion

3.4.1 Growth conditions

Bacterial growth is significant in the overall pathogenesis of an organism. It was shown that the rapid growth of *E. coli*, compared to other less common urinary pathogens, might explain why it is the most commonly known cause of UTI (Anderson *et al.*, 1979).

In this study, the O-antigen deficient strain *E. coli* EC18 was analyzed to see if the absence of virulence towards the *Galleria mellonella* model was due to growth impairment as the rapid growth of *E. coli* is essential in the pathogenesis during UTI (Alteri & Mobley, 2007; Anderson *et al.*, 1979). It was clearly shown (Figure 3.2) there were statistically significant differences between SA189 and other strains in LB media, while all strains in AUM were at a similar density after 18 hours growth. This does not explain the lack of virulence in the wax moth model. In addition, all three strains grew differently in LB compared to AUM. Complex rich media, such as LB, contain a good carbon and energy source (e.g., glucose), which has resulted in rapid growth of the UPEC strains. While in a minimal medium, such as AUM, nutrients were limited (Tao *et al.*, 1999). This variation of medium nutrients could explain the statistically significant difference in growth curves of LB and AUM.

Using AUM instead of filtered normal human urine provides many advantages, where different chemical composition of human urine is an issue of standardization of the study methods (Bouatra *et al.*, 2013). In addition, urine composition is dynamic, which varies according to diet, age, gender and health status (Ipe *et al.*, 2016). The AUM medium used in the current study has been shown to have all parameters of normal human urine within or close to their
normal physiological ranges (Chutipongtanate & Thongboonkerd, 2010). In the same study, which compared seven different media simulating human urine, AUM data were consistent with respect to cell compatibility and cytotoxicity profiles, where it has the least cytotoxic effects. AUM has been proposed to be an alternative for renal physiological study using *in vitro* cell culture models (Chutipongtanate & Thongboonkerd, 2010).

AUM was used to simulate *in vivo* conditions for the isolated strains with its limited source of nutrients. The enriched LB media has been widely used for cultivation of bacterial pathogens and other fastidious bacteria and was used here as a control medium. The structure of *Proteus mirabilis* biofilms has been studied while growing in AUM and LB media. Two different biofilm structures are formed, where AUM biofilm structure was observed as a flat layer lacking nutrient channels and inducing swarmer cells to protrude out into the bulk fluid where nutrients are abundant (Jones *et al.*, 2007).

In this study, proteins were extracted at mid-log phase, which was at ~3 hours of incubation. The significance of mid-log phase extraction has been shown in various studies (Makrides, 1996; Sorensen & Mortensen, 2005; Stevens, 2000), as this is representative of the phenotype of rapidly growing *E. coli* cells in the bladder during UTI. At the early log phase, *E. coli* cells are healthy and are in the prime state to produce proteins. Extracting proteins at a later stage will result in low protein yield (Figure 3.5) and high protein degradation as was observed in our optimization steps (result not shown).
3.4.2 Investigating LPS structure

O antigen is a major component of LPS, as it can promote resistance to activity of the innate immune system, such as serum killing and phagocytosis and its presence enhances the pathogenicity of *E. coli* (Reyes *et al.*, 2012). This point is particularly interesting as strain EC18 was recovered from a patient suffering urosepsis (Alghoribi *et al.*, 2014) indicating that the lack of an O antigen did not limit the ability of the strain to cause a serious invasive infection.

LPS structure consists of three main parts as follows: lipid A, the membrane base and the endotoxic part of LPS; the core saccharide of a branched chain of hexose and heptose sugars; and the O-antigen side chain, which has a repeating unit of sugars that extends into the extracellular space (Figure 3.17). The core region is composed of and inner core and outer core. The inner core is linked to lipid A and contains of 2-keto-3-deoxy-octanoate (Kdo) and L-glycero-D-mannoheptose. The outer core is bound to the O antigen and consist of hexoses and hexosamines (Silhavy *et al.*, 2010).

Figure has been removed due to Copyright restrictions.

Figure 3.17: General chemical structure of LPS from Gram-negative enterobacteria (Alexander & Rietschel, 2001).
There are three types of LPS structure, where absence of a specific part of the LPS structure determines its type. The three types are smooth, semi-rough, and rough. Production of LPS with the full structure results in smooth LPS, while absence of the O-specific chain results in the semi-rough type and rough LPS is present if the outer core is missing as well (Reyes et al., 2012). Absence of the O-specific chain on the SDS gels for strain EC18 indicates that it carries a semi-rough type structure of LPS. Many studies have indicated that bacteria with rough or semi-rough LPS have a higher LPS extraction yields than smooth LPS bacteria (Darveau & Hancock, 1983; Helander et al., 1992). Moreover, LPS was shown to play important role in motility as a wetting agent; thus its mutation could indicate limitation in swarming motility (Toguchi et al., 2000).

3.4.3 Protein extraction

In this study, most bacterial proteins were extracted using the two methods described earlier. Using SDS to solubilize and extract the insoluble proteins was appropriate for this study as the aim was not to isolate functional proteins. Although both soluble and insoluble fractions of proteins were processed separately, data files from both fractions were merged in final analysis to have the full proteome profile of growth comparison in AUM and LB, as the focus of this chapter was method development and optimisation. However, data for each extraction method could be examined later in further studies to give insight into the difference in proteome profiles between the soluble and insoluble proteins of UPEC. Why strain EC18 provided a greater proportion of soluble proteins than strains EC41 and SA189 can be linked to the difference in LPS (Darveau & Hancock, 1983; Helander et al., 1992).
3.4.4 Analysis and interpretation of proteomics data.

The LFQ method in mass spectrometry aims to determine the relative abundance of proteins in a biological sample without using stable isotope-containing compounds or other chemical labelling methods (Wasinger et al., 2013). Using LFQ proteomics approaches requires less preparation steps, thus it is a simpler and cost-effective method than stable isotope-based labelling methods. In addition, the applicability of LFQ to any type of samples with no limitation on the number of compared samples makes it more preferable to use (Cox et al., 2014). The data also indicates that the methodology of this study was robust with a correlation between replicates of ≥ 0.9.

The unique clustering of EC18 in HCL for both conditions (LB and AUM) compared to EC41 and SA189 correlates with the previous findings of genomics analysis and *Galleria mellonella* infection model data for strain EC41 (Alghoribi et al., 2014) and SA189 (Klazar & Upton, Unpublished), and suggests that the lack of an O-antigen has significantly affected the proteins being expressed by the strain. Fold enrichment (the ratio of two proportions) was used to illustrate the up-regulated and down-regulated pathways (Figure 3.13)(Figure 3.14). Some pathways (n=5) included both up-regulated and down-regulated proteins and that is due to the different regulation of related proteins within the pathways.

The pathway for amino sugar and nucleotide sugar carbohydrate metabolism was commonly up-regulated in all strains in LB. This rational regulation is most likely to be due to the rich medium content of LB. The UPEC strains also showed up-regulation in genetic information processing pathways, including transcription (RNA polymerase) and translation (Aminoacyl-tRNA
biosynthesis and Ribosome), which correlate with the growth of organisms in LB medium. The increased rate of transcription/translation is a result to adapt to this rich and complex environment, where nutrients (e.g. tryptone and yeast exact) are necessary to grow bacteria to high cell densities (Christensen et al., 2017). In addition, both nucleotide metabolism pathways for Purine and Pyrimidine were up-regulated in all strains.

Purine has a key role in cellular energy systems and cellular signaling. Purine and pyrimidine are both contributors in RNA and DNA production. This correlate to the increased growth rate in LB condition compared to AUM where more synthesis of cellular proteins is required (Klumpp & Hwa, 2008). They may be produced de novo or reprocessed by a salvage pathway from regular catabolism. Uric acid is the final product of purine catabolism (Xi et al., 2000), while citric acid cycle intermediates are the outcome of pyrimidine metabolism (Turnbough & Switzer, 2008).

UPEC EC41 and SA189 showed exclusive up-regulation in the pentose phosphate pathway and in glycolysis / gluconeogenesis (Table 3.3). These two pathways could be an important route of carbon flux through the central metabolic pathways during growth in a rich medium, such as LB (Sprenger, 1995). The activation of an alternative pathway does not necessarily require any expressional changes, but instead it may depend on the accumulation of an intermediate metabolite (Nakahigashi et al., 2009). This mechanism can mediate the strength of these metabolic networks. In addition, gluconeogenesis pathway is utilized by E. coli for glucose-6-phosphate, which will be eventually used in cell envelope synthesis including the O-specific chain of LPS structure (Figure 3.17) (Silhavy et al., 2010).
Furthermore, starch and sucrose metabolism and pyruvate metabolism pathways are connected to the glycolysis / gluconeogenesis pathway (Figure 3.18). The glycolysis process has multiple reactions that metabolize glucose to two molecules of pyruvate, yielding the associated pair of molecules of ATP. Gluconeogenesis is the process of glucose synthesis from non-carbohydrate origins such as pyruvate and lactic acid (Berg et al., 2002). This does not necessary mean they work as in opposition to each other.

On the other hand, all three UPEC strains showed down-regulation of the TCA cycle pathway and the amino acid metabolism of Glycine, Serine, and Threonine in LB and thus up-regulation in AUM (Table 3.4), which simulates the conditions found in UTI. In urine and AUM, peptides and amino acids are limited and are the key carbon source for UPEC during UTI; so central metabolic pathways are essential for UPEC fitness in vivo (Alteri et al., 2009).

Furthermore, the glycolysis / gluconeogenesis pathway was down-regulated in EC18 and SA189 in AUM. As we describe above, some essential steps of glycolysis are not essential in the gluconeogenesis process.
Figure 3.18: Glycolysis / Gluconeogenesis pathway of *E. coli* O18:K1:H7 UTI89 (UPEC). Green boxes indicating genes presence in the genome and the completeness of the pathway. Map00010 of Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2017).
Membrane transport systems include ABC transporters, the phosphotransferase system (PTS) and bacterial secretion systems. These are required for processing environmental information. ATP-binding cassettes (ABC) are essential membrane protein transporters that dynamically carry molecules across cell membranes, implying the energy originated from the hydrolysis of ATP to ADP (Moussatova et al., 2008). The pathway was down-regulated in both EC18 and SA189. The low energy source in AUM could explain the reason behind up-regulation to compensate for the nutrient limitation.

As shown in Table 3.4, EC18 showed more down-regulated pathways than EC41 and SA189 in contrast to the up-regulated pathways in Table 3.3. Some of these pathways were exclusive to EC18 including carbohydrate metabolism pathways such as Glyoxylate and Dicarboxylate metabolism and pyruvate metabolism. This could be due to the lack of O-antigen (Figure 3.16) and thus the repeating polysaccharides of O antigen in EC18 strain (Figure 3.17) meaning the strain will not require carbohydrate building blocks at the levels needed in other strains.

3.4.5 Conclusion

The G. mellonella model showed its usefulness in many studies. However, it did not detect the pathogenicity of EC18 strain, as it was isolated from clinical case of UTI. This indicates clearly that it is not a good surrogate model for UTI isolates.
From these findings we can conclude that using proteomics approaches to investigate the bacterial pathogenesis of UPEC is appropriate. The proteomics findings of different strains in different conditions were reproducible with a high correlation. In addition, the proteome profile analysis reflected variations in O-antigen genomic data for the isolates being examined.

In this study, one aim was to represent the UPEC ST127 proteome profile. However, EC18 could not be considered as a typical representative of that sequence type. On the other hand, EC41 and SA189 were similar regarding their proteome profiles. Nevertheless, due to their differences in geographical isolation, adding more isolates of ST127 in addition to reference strains of ST127 such as *E. coli* 536 (O6:K15:H31) will support this representation of the ST127 proteomic profile.

This study has shown that environmental changes have an effect on the metabolic pathways expressed by a specific strain. These conditions can be engineered to simulate the environments UPEC isolates may encounter in the transition from a nutrient-limited environment (Urine) to a richer environment (Blood).

The metabolic pathways highlighted above are essential for the growth of UPEC strains. However, the redundancy of some of these pathways for glucose metabolism, or the biosynthesis of amino acids, purines and pyrimidines can be interpreted as being a result of the necessity for UPEC isolates to be able to survive in a variety of environments. These tactics will enable the bacteria to exploit different environmental resources and cope with the lack, or limitation, of the essential small molecules required for metabolic functions (Kim & Copley, 2007).
Chapter Four: Metabolic Adaption in *Escherichia coli* Isolates During Transition from UTI to Bloodstream Infection
4.1 Introduction

Generally, *E. coli* inhabits the lower gastrointestinal tract (Moller et al., 2003). However, for some strains such as Uropathogenic *E. coli* (UPEC), it is also adapted to the relatively harsh environment of the urinary tract. Urinary tract infections (UTIs) are mostly initiated as a result of UPEC strains contaminating and colonising the urethra and then migrating to the bladder lumen (Rosen *et al.*, 2007). Several studies suggested that some UPEC strains can invade the epithelial cells of the bladder and initiate an intracellular infection cycle (Justice *et al.*, 2004; Martinez *et al.*, 2000).

UPEC is the most common cause of UTI. UPEC strains can colonize the bladder and kidney, which may sometimes lead to possibly a fatal case of bacteraemia (Kaper *et al.*, 2004; Smith *et al.*, 2010). Many studies showed that *E. coli* is the most common pathogen responsible for Gram-negative bacteraemia (Al-Hasan *et al.*, 2010; Dale *et al.*, 2018; Lee *et al.*, 2016; Shaw *et al.*, 2015).

In addition, previous work with paired urine and bloodstream UPEC strains revealed identical results for molecular typing and virulence gene content (Alghoribi, 2015), supporting the notion that strains progress from the urine to the bloodstream. However, the study showed different metabolic activity (Table 4.1) and antibiotic susceptibility profiles (Table 4.2) in some pairs of isolates indicating phenotypic changes, potentially due to single nucleotide polymorphisms (SNPs) or epigenetic events. These changes were stable during regular maintenance in culture in laboratory media, such as LB and it is possible that the metabolic changes reflect the different requirements for growth in different clinical sites.
Table 4.1: Metabolic activity profiles of paired blood and urine *E. coli* isolates. Adapted from (Alghoribi, 2015).

<table>
<thead>
<tr>
<th>NO.</th>
<th>Sample</th>
<th>ST</th>
<th>Phylogroup</th>
<th>Biochemical Tests</th>
<th>Bio-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GGT</td>
<td>ProA</td>
</tr>
<tr>
<td>1</td>
<td>Blood</td>
<td>131</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>Urine</td>
<td>131</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Blood</td>
<td>131</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Urine</td>
<td>131</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>Blood</td>
<td>127</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>Urine</td>
<td>127</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>Blood</td>
<td>127</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>41</td>
<td>Urine</td>
<td>127</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>Blood</td>
<td>73</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>Urine</td>
<td>73</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>Blood</td>
<td>73</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>73</td>
<td>Urine</td>
<td>73</td>
<td>B2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>Blood</td>
<td>38</td>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>47</td>
<td>Urine</td>
<td>38</td>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>86</td>
<td>Blood</td>
<td>38</td>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>86</td>
<td>Urine</td>
<td>38</td>
<td>D</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

ST=Sequence Type; 1= Positive; 0= Negative; GGT= Gamma-Glutamyl-Transferase; ProA= L-Proline Arylamidase; TyrA= Tyrosine Arylamidase; SAC= Saccharose/Sucrose; 5KG= 5-Keto-D-Gluconate; ILATk= L-Lactate alkalisation; SUCT= Succinate alkalisation; GlyA= Glycine Arylamidase; BGUR= Beta-Glucoronidase; O129R= O/129 Resistance; IMLTa= L-Malate assimilation; ELLM= Ellman.
Table 4.2: Antibiotics susceptibility profiles of paired blood and urine *E. coli* isolates. Adapted from (Alghoribi, 2015).

<table>
<thead>
<tr>
<th>NO.</th>
<th>Sample</th>
<th>ST</th>
<th>AMO</th>
<th>AMP</th>
<th>AMC</th>
<th>PIP/TAZ</th>
<th>CF</th>
<th>CXM</th>
<th>CXM A.</th>
<th>FOX</th>
<th>CTX</th>
<th>CAZ</th>
<th>FEP</th>
<th>ATM</th>
<th>ETP</th>
<th>MEM</th>
<th>AMK</th>
<th>GM</th>
<th>TOB</th>
<th>Cip</th>
<th>OFX</th>
<th>TGC</th>
<th>TRI</th>
<th>Anti-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Blood</td>
<td>131</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>1</td>
<td>Urine</td>
<td>131</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>10</td>
<td>Blood</td>
<td>131</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>I</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
</tr>
<tr>
<td>10</td>
<td>Urine</td>
<td>131</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>I</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
</tr>
<tr>
<td>18</td>
<td>Blood</td>
<td>127</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>18</td>
<td>Urine</td>
<td>127</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>41</td>
<td>Blood</td>
<td>127</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
</tr>
<tr>
<td>41</td>
<td>Urine</td>
<td>127</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
</tr>
<tr>
<td>22</td>
<td>Blood</td>
<td>73</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>22</td>
<td>Urine</td>
<td>73</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>73</td>
<td>Blood</td>
<td>73</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>73</td>
<td>Urine</td>
<td>73</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>47</td>
<td>Blood</td>
<td>38</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>47</td>
<td>Urine</td>
<td>38</td>
<td>R</td>
<td>R</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>86</td>
<td>Blood</td>
<td>38</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>86</td>
<td>Urine</td>
<td>38</td>
<td>R</td>
<td>S</td>
<td>S</td>
<td>I</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

R= Resistant; S= Susceptible; I= Intermediate; AMO= amoxicillin; AMP= ampicillin; AMC= amoxicillin/clavulanic acid; PIP/TAZ= piperacillin/tazobactam; CF= cefotaxim; CXM= cefuroxime; CXM A.= cefuroxime axetil; FOX= cefoxitin; CTX= cefotaxime; CAZ= cefazidime; FEP= ceftipime; ATM= aztreonam; ETP= ertapenem; MEM= meropenem; AMK= Amikacin; GM= Gentamicin; TOB= Tobramycin; Cip= Ciprofloxacin; OFX= ofloxacin; TGC= tigecycline; TRI= trimethoprim.
4.1.1 Aim and objectives

We aim to reveal the mechanisms underlying metabolic adaptation of UPEC to allow growth in blood, using a label free quantitative proteomics and genomic approach. In this chapter we also aim to investigate how ST127 isolates behave compared to other UPEC STs isolates. To achieve these Aims, the following objectives were set:

- Extract proteins (soluble and insoluble) from all UPEC strains including both isolates of blood and urine paired cultures.
- Ensure protein extraction yields were reproducible from different strains during preparation of extracted proteins for analysis on the OrbiTrap MS platform.
- Carry out detailed analysis of MS data with different means of analysis and presentation of bioinformatics analysis of proteomics data.
- Validate the proteomics findings with genomics analysis.

4.2 Experimental design

In this chapter we have four STs of UPEC isolates (ST38, ST73, ST127, and ST131). Each ST has two strains with paired isolates from blood and urine. Each sample will have two protein fractions (soluble and insoluble). Three individual biological replicates were done for each sample in this project, which will bring the total number of the project samples to 96 samples (Figure 4.1).
Figure 4.1: Experimental design of the paired isolates study. The image indicates the sub category samples for each ST, with three biological replicates each.
4.3 Results

4.3.1 Protein extractions

All samples were cultured from a single colony in 10 ml LB broths and incubated at 37°C overnight. Extraction was performed at mid log phase (~ 3 hours) on 100 ml of LB media per sample. Soluble proteins were extracted using the B-PER complete bacterial protein extraction reagent, while insoluble proteins were extracted using SDT lysis buffer. BCA protein estimation indicated higher protein extraction yield in insoluble fragments than the soluble one. Moreover, in unpaired t test analysis, insoluble proteins were statistically more significant than soluble fraction proteins in all paired isolates except ST127 strains (EC18 and EC41) and *E. coli* EC86 U (Figure 4.2).

On the gel, it is clear that biological replicates of protein extractions on SDS PAGE analysis in blood isolates of EC18 and EC41 are consistent and with well resolved protein bands visible (Figure 4.3).

4.3.2 Overview analysis of the proteome profiles

Following analysing all the paired isolate samples, including the soluble and insoluble fractions in addition to all the replicates, we followed a specific analytical approach described in (2.6.3) to interpret and understand the biological meaning of the data set.

The initial analysis was done using Perseus, Excel, and Venny tool, an interactive tool for comparing lists with Venn’s diagrams (Oliveros, 2015) (http://bioinfogp.cnb.csic.es/tools/venny/), to have an overview of the proteome profiles.
Figure 4.2: BCA estimation for protein yield extracted from paired isolates (Urine and Blood) of UPEC strains cultured in Lysogeny Broth (LB). Soluble proteins extracted by B-PER complete bacterial protein extraction reagent. Insoluble proteins extracted by SDT lysis buffer. Data are plotted as the mean of three biological replicates and error bars represent standard error of the mean (SEM). $p$-Value represented as (**** = $p \leq 0.0001$, *** = $p \leq 0.001$, ** = $p \leq 0.01$, * = $p \leq 0.05$, n.s. = not significant) for the soluble and insoluble proteins fractions comparison of each sample.
Figure 4.3: Samples of protein extractions for *E. coli* EC18 (Blood and Urine), and *E. coli* EC41 (Blood) loaded on SDS PAGE gels. (A) Soluble proteins extracted by B-PER complete bacterial protein extraction reagent; 30 µg of protein was loaded for all samples. (B) Insoluble proteins extracted by SDT lysis buffer. Blood isolates are referred to as B1 and B2. Urine isolates are referred to as U. 50 µg of protein was loaded for all samples. InstantBlue staining was used.

A total of 2025 proteins were identified across all strains. The data indicated a good correlation between replicates of ≥ 0.9 (Appendix II). 1164 of these proteins were quantified in all strains. Moreover, 381 (32.73%) of these proteins were up-regulated (>1.5 fold) in blood isolates compared to urine isolates, and 194 (16.67%) proteins were down-regulated (>2 fold) across all strains in blood isolates compared those from urine (Figure 4.4). In addition, slightly up-regulated (1.3 - 1.5 fold) and down-regulated (1.5 - 2 fold) proteins with statistical significance were included in the study. The ANOVA test indicates that 943 (81.01%) proteins were statistically significant.
In general, ST38 and ST73 have more shared regulated proteins compared to other STs. ST38 and ST73 share 69 (27.5%) up-regulated proteins and 41 (34.7%) down-regulated proteins. On the other hand, ST127 share considerably less differentially expressed proteins with ST131. In fact, ST131 has less differentially regulated proteins than was observed for other STs (Table 4.3). This table shows a summary of differentially regulated proteins in all STs and their percentages regarding to the total quantified proteins. Remarkably, all STs isolates shared just one up-regulated protein (UPF0267 protein YqfB), while no down-regulated protein was shared between all STs (Figure 4.4). A list of intersecting proteins of Figure 4.4 are listed in Appendix III.

Table 4.3: Up-regulated and down-regulated proteins (>1.5 fold) in all four STs. The numbers indicate the proteins common to all strains of an ST.

<table>
<thead>
<tr>
<th></th>
<th>ST38</th>
<th>ST73</th>
<th>ST127</th>
<th>ST131</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up-regulated proteins</td>
<td>177</td>
<td>143</td>
<td>136</td>
<td>37</td>
</tr>
<tr>
<td>Down-regulated proteins</td>
<td>79</td>
<td>80</td>
<td>85</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>256</td>
<td>223</td>
<td>221</td>
<td>54</td>
</tr>
<tr>
<td>Percentage</td>
<td>22%</td>
<td>19.2%</td>
<td>19%</td>
<td>4.6%</td>
</tr>
</tbody>
</table>
Figure 4.4: Overlaps between differentially altered proteins in four STs showing the up-regulated and down-regulated proteins in blood vs. urine isolates. Venny 2.1 tool was used to produce the figures (Oliveros, 2015).
4.3.3 UPEC STs strain analysis

Following the initial analysis, we continue examining the data using a method that examines relative expression. Relative expression data can be found out by dividing the average LFQ value of a replicate on the highest value among the samples. The first step was done on Excel, and then the *.txt file was uploaded into Perseus as described above (2.6.3.1).

4.3.3.1 Perseus analysis of all UPEC strains

Examining the relative expression data, we can see that 1164 proteins were clustered into ten clusters with a distance threshold of 2.028 (Figure 4.5). These clusters are different in size ranging from 426 proteins to just 7 proteins (Table 4.4).

In Figure 4.5, we can see the preserved order of sample starting with urine isolates then the blood isolates of each strain of an ST. The heat map indicates the relativity to the highest LFQ value in each protein. Remarkably, ST127 strains (EC18 and EC41) in both isolates (urine and blood) displayed similar proteins expression profiles to each other. On the other hand, the blood isolates of ST131 (EC1 and EC10) showed similar profiles to the urine isolates of UPEC, especially at cluster 1152 (Figure 4.6). With a closer look at the row clusters, we can see what we describe in the heat map in addition to more detailed information. 1148 and 1132 clusters showed spikes in protein expression for both EC18 isolates, while it drops in cluster 1146. Moreover, proteins expression spiked for both isolates of EC73 and EC47 in clusters number 1120 and 995, respectively.
Figure 4.5: Hierarchical clustering (HCL) of relative expression data of LFQ ratios for paired isolates of ST131 (EC1 and EC10), ST127 (EC18 and EC41), ST73 (EC22 and EC73), and ST38 (EC47 and EC86). Row clusters ID numbers are shown. Scale range from 0.0 (light green) to 1 (light red).

Table 4.4: Row clusters (n=10) of HCL of relative expression data of label-free protein quantification (LFQ) ratios for paired isolates of ST131 (EC1 and EC10), ST127 (EC18 and EC41), ST73 (EC22 and EC73), and ST38 (EC47 and EC86). Size indicates the number of proteins in the cluster.

<table>
<thead>
<tr>
<th>Cluster No.</th>
<th>1153</th>
<th>1152</th>
<th>1151</th>
<th>1150</th>
<th>1148</th>
<th>1146</th>
<th>1145</th>
<th>1132</th>
<th>1120</th>
<th>995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>98</td>
<td>307</td>
<td>426</td>
<td>38</td>
<td>117</td>
<td>106</td>
<td>32</td>
<td>23</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Matched proteins in STRING</td>
<td>94</td>
<td>290</td>
<td>400</td>
<td>34</td>
<td>109</td>
<td>105</td>
<td>26</td>
<td>21</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Coverage %</td>
<td>95.92</td>
<td>94.46</td>
<td>93.90</td>
<td>89.47</td>
<td>93.16</td>
<td>99.06</td>
<td>81.25</td>
<td>91.30</td>
<td>80.00</td>
<td>85.71</td>
</tr>
</tbody>
</table>
Figure 4.6: Row clusters (n=10) of HCL of relative expression data of paired isolates of ST131 (EC1 and EC10), ST127 (EC18 and EC41), ST73 (EC22 and EC73) and ST38 (EC47 and EC86). Colour scale ranges from 0.0 (red) to >3.5 (light green) indicating distance from centre.
4.3.3.2 Protein-Protein interactions network of all UPEC strains

To examine the clusters in Figure 4.5, we used STRING tool to look at the protein-protein interactions network. This step will enable us to visualise the enriched metabolic pathways and how significant they are in each cluster. 

*E. coli* CFT073 (NCBI:txid199310) data was used as a reference due to being UPEC strain and the high number of proteins from the samples that matched it (distinct protein-coding genes: 5369).

Following analysis on STRING tool, the average coverage of matched proteins in all UPEC strain clusters (n=10) was 90.42% (Table 4.4). The analysis was focused on comparing between different STs, in particular ST127 compared to other UPEC STs. Some clusters showed statistically significant enrichments detected in functional subsystems (n=7), while other showed no significant enrichment (n=3).

Furthermore, clusters 1152 (Figure 4.7) and 1148 were enriched in metabolic pathways, especially in riboflavin, pyrimidine and carbon metabolism for cluster 1148 (Figure 4.8). Cluster 1146 was also enriched in metabolic pathways (e.g. histidine metabolism) in addition to biosynthesis of secondary metabolites (Figure 4.9).

Although cluster 1132 showed highest fold change for strain EC18 isolates, it did not show any enriched function (data not shown). However, in STRING, we can expand the examined list of proteins to include the closely relative proteins to the original list for PPI and functional annotation enrichment analysis. Thus, examining the extended network, we found out functional enrichment for galactose, and pyruvate metabolism (Figure 4.10). On the other hand, the expressed proteins in cluster 995 indicate enrichment
in Pentose and glucuronate interconversions, and metabolic pathways for EC47 isolates (Figure 4.12). Moreover, cluster 1120 showed enrichment of metabolism of alanine, aspartate, glutamate and pyrimidine for EC73 isolates (Figure 4.11).
Figure 4.7: Interactions network enrichment analysis of cluster no. 1152 (all UPEC entries). Combined screenshots from the STRING website, showing results obtained upon entering a set of 307 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.8: Interactions network enrichment analysis of cluster no. 1148 (all UPEC entries). Combined screenshots from the STRING website, showing results obtained upon entering a set of 117 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.9: Interactions network enrichment analysis of cluster no. 1146 (all UPEC entries). Combined screenshots from the STRING website, showing results obtained upon entering a set of 106 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.10: Extended interactions network enrichment analysis of cluster no. 1132 (all UPEC entries). Combined screenshots from the STRING website, showing results obtained upon entering a set of 23 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.11: Interactions network enrichment analysis of cluster no. 1120 (all UPEC entries). Combined screenshots from the STRING website, showing results obtained upon entering a set of 10 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.

Figure 4.12: Interactions network enrichment analysis of cluster no. 995 (all UPEC entries). Combined screenshots from the STRING website, showing results obtained upon entering a set of 7 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
4.3.4 UPEC ST127 strains analysis

4.3.4.1 Perseus analysis of UPEC ST127 strains

To have a closer look at ST127 strains, we performed a separate HCL using relative expression dataset from ST127 strains. We defined ten row clusters of the 1164 proteins with a distance threshold of 1.05 (Figure 4.13). These clusters are different in size ranging from 273 proteins to just 11 proteins (Table 4.5). The following analysis of expression clusters are not to be compared with the earlier analysis in this chapter. The cluster numbers are generated randomly in Perseus and the data are not related to the previous clusters.

In Figure 4.13, the order of samples has been specifically preserved, starting with urine isolates then the blood isolates of EC18 and EC41. The heat map indicates the relativity to the highest LFQ value in each protein.

Remarkably, EC18 isolates (urine and blood) displayed similar protein expression profiles to each other, which was also the case for EC41 isolates. This clearly can be seen in Figure 4.13 and in clusters numbered 1151, 1152, and 1146 (Figure 4.14).

Moreover, protein expression for urine isolates of EC18 and EC41 were higher than the blood isolates in cluster number 1139. However, cluster number 1137 showed more elevated expression in blood isolates compared the urine isolates ST127 strains.
Figure 4.13: Hierarchical clustering (HCL) of relative expression data of label-free protein quantification (LFQ) ratios for paired isolates of ST127 (EC18 and EC41). Row clusters ID numbers are shown. Scale range from 0.0 (light green) to 1 (light red).
Table 4.5: Row clusters (n=10) of HCL of relative expression data of label-free protein quantification (LFQ) ratios for paired uro-sepsis isolates of ST127 (EC18 and EC41). Size indicates the number of proteins in the cluster.

<table>
<thead>
<tr>
<th>Cluster No.</th>
<th>1153</th>
<th>1152</th>
<th>1151</th>
<th>1150</th>
<th>1148</th>
<th>1147</th>
<th>1146</th>
<th>1139</th>
<th>1137</th>
<th>1113</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>273</td>
<td>117</td>
<td>11</td>
<td>245</td>
<td>163</td>
<td>132</td>
<td>76</td>
<td>61</td>
<td>29</td>
<td>57</td>
</tr>
<tr>
<td>Matched proteins in STRING</td>
<td>262</td>
<td>108</td>
<td>10</td>
<td>228</td>
<td>153</td>
<td>127</td>
<td>72</td>
<td>56</td>
<td>26</td>
<td>51</td>
</tr>
<tr>
<td>Coverage %</td>
<td>95.97</td>
<td>92.31</td>
<td>90.91</td>
<td>93.06</td>
<td>93.87</td>
<td>96.21</td>
<td>94.74</td>
<td>91.80</td>
<td>89.66</td>
<td>89.47</td>
</tr>
</tbody>
</table>

Figure 4.14: Row clusters (n=10) of HCL of relative expression data of paired isolates of ST127 (EC18 and EC41). Colour scale ranges from 0.0 (red) to >3.5 (light green) indicating distance from centre.
4.3.4.2 Protein-Protein interactions network of UPEC ST127 strains

Following further analysis on STRING tool for ST127 isolates, the average coverage of matched proteins in all clusters (n=10) was found to be 92.80% for ST127 strains (Table 4.5). All clusters showed significant enrichment detected in functional subsystems.

Cluster 1153 showed highly expressed proteins in all isolates of ST127 with enrichment of ribosome-based pathways (Figure 4.15). Both clusters of 1152 and 1151 showed elevated protein expressions for EC18 isolates. Cluster 1152 showed functional enrichment in metabolic pathways including citrate cycle (TCA cycle) and carbon metabolism (Figure 4.16). Moreover, sphingolipid metabolism was the only enriched pathway in cluster 1151 (Figure 4.17). On the other hand, cluster 1146 showed higher protein expression in EC41 isolates. Metabolic pathways, in particular glutathione metabolism, were enriched in this cluster (Figure 4.22).

Interestingly, cluster 1139 showed higher enrichment in only P-loop containing nucleoside triphosphate hydrolase (INTERPRO Protein Domains and Features) in the urine isolates of ST127 UPEC strains, but not the corresponding blood stream isolates (Figure 4.23). However, cluster 1137 showed an opposite protein expression profile indicating functional enrichment in the blood isolates of ST127 for only methane metabolism (Figure 4.18).

Furthermore, a group of clusters (1113, 1147, 1148, 1150, and 1153) showed similar protein expression profiles (either high, intermediate, or low) for all isolates of ST127. Cluster with high expression profiles (1150 and 1153) were enriched in metabolic pathways including pyrimidine, purine, pyruvate
and carbon metabolism. In addition, the pentose phosphate pathway and aminoacyl-tRNA biosynthesis were also enriched (Figure 4.19)(Figure 4.15).

Regarding intermediate expression profile clusters, 1147 and 1148 showed functional enrichment in the metabolic pathways of glycerophospholipid, pyruvate, and beta-Alanine. Biosynthesis of secondary metabolite pathway and the phosphotransferase system (PTS) were also enriched (Figure 4.21)(Figure 4.20).

Lastly, cluster 1113 was the only cluster with a low expression profile across all ST127 isolates (Figure 4.24). Functional here enrichment included metabolic pathways (e.g. histidine metabolism), pentose and glucuronate interconversions and alanine, aspartate and glutamate metabolism.
Figure 4.15: Interactions network enrichment analysis of cluster no. 1153. Combined screenshots from the STRING website, showing results obtained upon entering a set of 273 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.16: Interactions network enrichment analysis of cluster no. 1152. Combined screenshots from the STRING website, showing results obtained upon entering a set of 117 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.17: Interactions network enrichment analysis of cluster no. 1151. Combined screenshots from the STRING website, showing results obtained upon entering a set of 11 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.

Figure 4.18: Interactions network enrichment analysis of cluster no. 1137. Combined screenshots from the STRING website, showing results obtained upon entering a set of 29 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.19: Interactions network enrichment analysis of cluster no. 1150. Combined screenshots from the STRING website, showing results obtained upon entering a set of 245 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.20: Interactions network enrichment analysis of cluster no. 1148. Combined screenshots from the STRING website, showing results obtained upon entering a set of 163 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.21: Interactions network enrichment analysis of cluster no. 1147. Combined screenshots from the STRING website, showing results obtained upon entering a set of 132 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.22: Interactions network enrichment analysis of cluster no. 1146. Combined screenshots from the STRING website, showing results obtained upon entering a set of 76 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.23: Interactions network enrichment analysis of cluster no. 1139. Combined screenshots from the STRING website, showing results obtained upon entering a set of 61 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
Figure 4.24: Interactions network enrichment analysis of cluster no. 1113. Combined screenshots from the STRING website, showing results obtained upon entering a set of 57 proteins. The inset is showing the significant enrichments detected in functional subsystems, where enriched functions have been selected and the corresponding protein nodes in the network are highlighted in colour. Line thickness indicates the strength of data support.
4.3.5 Genomic analysis of paired UPEC isolates

A whole genome comparison between all the UPEC paired isolates was performed using a reference complete genome for each UPEC ST. The BLAST Ring Image Generator (BRIG 0.95-dev0004) (Alikhan et al., 2011) was used to compare and visualize UPEC genomes against the references. Additional data on assembly statistics are provided in Appendix IV.

Genomic analysis was carried out for all UPEC paired isolates including; EC1, EC10, against UPEC ST131 reference strain EC958 (Forde et al., 2014)(Figure 4.25); EC18, EC41, with UPEC ST127 reference strain 536 (Hochhut et al., 2006)(Figure 4.26); EC22, EC73, and UPEC ST73 reference strain CFT073 (Welch et al., 2002)(Figure 4.27); and finally EC47, EC86, and UPEC ST38 reference strain MRSN 346647 (Snesrud et al., 2017)(Figure 4.28).

On the whole, BRIG images indicated a low level of variation between paired isolates. However, EC1 isolates showed notable variation on BRIG analysis (Figure 4.25) and, thus, a further analysis was carried out with findings shown in Table 4.6. The Table shows a summary of the in silico analysis of MLST, serotype and Fim-type for all the UPEC isolates. This reveals the different serotype and Fim-type for blood and urine isolates of strain EC1, which suggests that there was more than one strain causing the UTI that progressed to uro-sepsis. Similar findings were reported by McNally and colleagues (McNally et al., 2013).

The deletion of the EC18 O-antigen genes in both blood and urine isolates was confirmed here and was discussed in the previous chapter. Table 4.7 indicates the numbers of features missing from each strain compared to its
partner and gives the total length of the combined missing sequences. In addition, the number of completely and partially absent genes are shown. With the exception of strain EC1, which clearly has some substantial differences between blood and urine isolates, pairs of strains show limited genetic differences.

Figure 4.25: BLAST ring image of paired isolates (blood and urine) of ST131 strains (EC1 and EC10). UPEC EC958 was used as reference strain. Blood isolates are indicated in red and the urine isolates in yellow.
Figure 4.26: BLAST ring image of paired isolates (blood and urine) of ST127 strains (EC18 and EC41). UPEC 536 was used as reference strain. Blood isolates are indicated in red and the urine isolates in yellow.
Figure 4.27: BLAST ring image of paired isolates (blood and urine) of ST73 strains (EC22 and EC73). UPEC CFT073 was used as reference strain. Blood isolates are indicated in red and the urine isolates in yellow.
Figure 4.28: BLAST ring image of paired isolates (blood and urine) of ST38 strains (EC47 and EC86). UPEC MRSN 346647 was used as reference strain. Blood isolates are indicated in red and the urine isolates in yellow.
## Table 4.6: MLST, Serotyping and Fim-typing of each strain examined from the four UPEC lineages

<table>
<thead>
<tr>
<th>Organism ID</th>
<th>Organism</th>
<th>MLST</th>
<th>SeroType</th>
<th>FimType</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC1U</td>
<td><em>Escherichia coli</em></td>
<td>ST131</td>
<td>O16:H5</td>
<td>fimH41</td>
</tr>
<tr>
<td>EC1B</td>
<td><em>Escherichia coli</em></td>
<td>ST131</td>
<td>O25:H4</td>
<td>fimH30</td>
</tr>
<tr>
<td>EC10U</td>
<td><em>Escherichia coli</em></td>
<td>ST131</td>
<td>O25:H4</td>
<td>fimH30</td>
</tr>
<tr>
<td>EC10B</td>
<td><em>Escherichia coli</em></td>
<td>ST131</td>
<td>O25:H4</td>
<td>fimH30</td>
</tr>
<tr>
<td>EC18U</td>
<td><em>Escherichia coli</em></td>
<td>ST127</td>
<td>H31</td>
<td>fimH2</td>
</tr>
<tr>
<td>EC18B</td>
<td><em>Escherichia coli</em></td>
<td>ST127</td>
<td>H31</td>
<td>fimH2</td>
</tr>
<tr>
<td>EC41U</td>
<td><em>Escherichia coli</em></td>
<td>ST127</td>
<td>O6:H31</td>
<td>fimH2</td>
</tr>
<tr>
<td>EC41B</td>
<td><em>Escherichia coli</em></td>
<td>ST127</td>
<td>O6:H31</td>
<td>fimH2</td>
</tr>
<tr>
<td>EC22U</td>
<td><em>Escherichia coli</em></td>
<td>ST73</td>
<td>H1</td>
<td>fimH10</td>
</tr>
<tr>
<td>EC22B</td>
<td><em>Escherichia coli</em></td>
<td>ST73</td>
<td>H1</td>
<td>fimH10</td>
</tr>
<tr>
<td>EC73U</td>
<td><em>Escherichia coli</em></td>
<td>ST73</td>
<td>O50/O2:H1</td>
<td>fimH10</td>
</tr>
<tr>
<td>EC73B</td>
<td><em>Escherichia coli</em></td>
<td>ST73</td>
<td>O50/O2:H1</td>
<td>fimH10</td>
</tr>
<tr>
<td>EC47U</td>
<td><em>Escherichia coli</em></td>
<td>ST38</td>
<td>O86:H18</td>
<td>fimH5</td>
</tr>
<tr>
<td>EC47B</td>
<td><em>Escherichia coli</em></td>
<td>ST38</td>
<td>O86:H18</td>
<td>fimH5</td>
</tr>
<tr>
<td>EC86U</td>
<td><em>Escherichia coli</em></td>
<td>ST5150</td>
<td>O1:H15</td>
<td>fimH65</td>
</tr>
<tr>
<td>EC86B</td>
<td><em>Escherichia coli</em></td>
<td>ST5150</td>
<td>O1:H15</td>
<td>fimH65</td>
</tr>
</tbody>
</table>
Table 4.7: Summarised genomic analysis comparing the urine and blood isolates of each strain of UPEC STs.

<table>
<thead>
<tr>
<th>Sequence type</th>
<th>ST131</th>
<th>ST127</th>
<th>ST73</th>
<th>ST38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain</td>
<td>EC1</td>
<td>EC10</td>
<td>EC18</td>
<td>EC41</td>
</tr>
<tr>
<td>Missing features</td>
<td>116</td>
<td>133</td>
<td>107</td>
<td>86</td>
</tr>
<tr>
<td>Total length (bp)</td>
<td>364012</td>
<td>269083</td>
<td>111143</td>
<td>295517</td>
</tr>
<tr>
<td>Deleted genes</td>
<td>405</td>
<td>287</td>
<td>90</td>
<td>290</td>
</tr>
<tr>
<td>Partially deleted genes</td>
<td>63</td>
<td>101</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>
4.4 Discussion

4.4.1 Growth conditions and protein extractions

As described earlier, one of main factors in bacterial pathogenesis is rapid growth (Anderson et al., 1979). In laboratories, enriched LB media has been widely used for culturing of diverse range of bacteria including fastidious ones. Complex nutrient rich LB medium contains a good carbon (catabolizable amino acids) and energy source (e.g., glucose) (Sezonov et al., 2007).

Using a single type of media for our experiment with the paired isolates is crucial. To explain this, when two strains of one pair are grown in LB and are then analysed, the differences in their proteomes will be due the stain itself. This means it is not the result of an environmental effect, as described in the previous chapter. In addition, we extracted all fractions of protein at mid-log phase, which was at ~3 hours of incubation. The importance of this time point has been described in the previous chapter, as well as in other studies (Makrides, 1996; Sorensen & Mortensen, 2005; Stevens, 2000). Moreover, as described earlier, bacterial proteins were extracted based on their solubility. Both fractions (soluble and insoluble) of proteins were processed separately then the peaklists/spectra files were combined for each sample.

In this study, extraction of insoluble protein gave a higher yield than the soluble extraction in all paired isolates. However, strangely, ST127 paired isolates (EC18 and EC41) and UPEC EC86 U showed no statistical significance (average $p = 0.642$) in their differences. No explanation could be offered for this finding. The protein extraction results were highly reproducible.
4.4.2 Overview of the proteomic profile analysis

A total of 1164 proteins were quantified in this study across all STs. Almost half of these (528/1164) were differentially regulated (both up-regulation and down-regulation) when comparing isolates from blood with their paired isolate from urine. ST131 strains had the lowest number of differentially regulated proteins among other STs, suggesting conservation of phenotype in isolates from this ST during transition from urine to blood. On the other hand, ST38 isolates possessed the highest number of differentially regulated proteins. The fact that, ST131 isolates shared the lowest number of differentially regulated proteins with ST127 strains emphasises the unique profile of the latter clonal group.

Of all these differentially regulated proteins, only one protein was seen to be up-regulated in blood compared to urine among pairs from all STs. UPF0267 protein YqfB is PUA-like domain-containing protein according to the *E. coli* K-12 sub-strain MG1655 database, however it was identified as hypothetical protein according to *E. coli* CFT073 database. This protein consists of 103 aa/312 bp. The molecular weight of the polypeptide is 11.905 kD and it is located in the cytosol (Shen et al., 2005). Following intensive searches in relevant databases for YqfB, the protein function remains uncharacterized.

A relative expression method was used as the main approach to tackle this big dataset, trying to find a connection between the proteome profiles of all strains examined. The LFQ based analysis has been preferred in many studies (Cox et al., 2014). The LFQ method represents the protein intensities using standardisation to eliminate some outlier results to best represent the ratio of changes between various samples. Since the MaxQuant program can present
different categorical data, including intensity and LFQ, LFQ can be used to compare between samples and intensity for some complex cases (e.g. some fractions were conditioned or treated differently) (Schwanhausser et al., 2011). Intensities are more tidily arranged than LFQ and, significantly, normalized across all samples.

### 4.4.3 Paired UPEC metabolic pathways

Following analysis on STRING tool, the coverage and identification of the gene name entries was at a high rate (90.42%) (Table 4.4). However, functional enrichment was not detected for some clusters (clusters 1150, 1145, and 1132), due to the low number of proteins listed in those clusters.

Interestingly, cluster 1132 showed highly expressed proteins almost exclusively in the EC18 samples, but this cluster did not show any enriched functions. Nevertheless, looking deeper into the extended network we detected functional enrichment in pathways such as galactose (galT) and pyruvate metabolism. The importance of galactose lies in the synthesis of exopolysaccharides (e.g. colanic acid) in E. coli (Whitfield & Paiment, 2003). In this study, yeast extract (a component of LB medium used for paired isolates growth) was the source of the carbohydrates (Grant & Pramer, 1962), which were subsequently metabolized into galactose (Medina et al., 2011).

A deletion of galE can result in inhibition of isomerization of UDP-galactose to UDP-glucose, and vice versa (Figure 4.29). A galE mutant E. coli was unable to grow in the galactose supplemented medium due to accumulation of the intermediate products of galactose-1-phosphate and UDP-galactose (Barkan et al., 2011; Csiszovszki et al., 2011; Sundararajan, 1963).
Nevertheless, galactose was observed efficiently incorporated into the LPS structure of galE mutant *E. coli* via the UDP-galactose-LPS galactosyl-transferase system, where galactose is transferred from the sugar nucleotide into the incomplete polysaccharide moiety of the endogenous lipopolysaccharide (Csizovszki *et al.*, 2011; Schnaitman & Austin, 1990).

![Figure 4.29: Schematic diagram illustrating the specific cascade of Galactose metabolism.](image)

UDP-glucose is essential for the LPS synthesis, cell wall sugar moieties, capsular polysaccharides and membrane-derived oligosaccharides, as well as exopolysaccharides, either directly or via UDP-galactose (Berbis *et al.*, 2015). Knowing that there is an insertion-sequence (IS1) mediated deletion (from galE to wcaA) within the O-antigen and colanic acid gene clusters in EC18 paired isolates (Figure 3.16) that resulted in the semi-rough type of LPS (Figure 3.4)
could explain why proteins associated with metabolism of galactose are up-regulated in this strain, compared to others.

Proteins in EC47 isolates were also exclusively up regulated compared to all UPEC paired isolates in cluster 995 (Figure 4.6). Their up-regulation indicates enrichment in the interconversions of the monosaccharide pentose and glucuronate, the salts or esters of glucuronic acid (carbohydrate metabolism) (Figure 4.12). Additionally, exclusive up-regulation was also noted for EC73 isolates protein expression in cluster 1120 (Figure 4.6). The enriched pathways were metabolism of alanine, aspartate, glutamate (amino acid metabolism); and pyrimidine metabolism (nucleotide metabolism) (Figure 4.11). The changes in these pathways have been shown to affect the bacterial small colony variant (SCV) phenotype. SCV is a slow-growing aberrant form of the wild type bacteria, which has been implicated in cases of persistent, relapsing and infections caused by bacteria that have a reduced susceptibility to antibiotics (Cui et al., 2012; Proctor et al., 2006). In fact, many studies have reported that chronic UTIs are caused by SCV E. coli (Dutt et al., 2014; Hooton, 2001; Trülzsch et al., 2003). It is still unclear how relevant these findings to the strains in question, but it may explain their ability to cause urosepsis.

In cluster 1152 and 1148, some pathways showed elevated expression profiles in the blood isolates from all STs (except those of ST127, which exhibited elevated expression in blood and urine) (Figure 4.6). These pathways range from genetic information translation (e.g. ribosome and aminoacyl-tRNA biosynthesis) (Figure 4.7) to metabolic pathways (e.g. riboflavin, pyrimidine and carbon metabolism) (Figure 4.8). This potentially indicates an overall increase in metabolic activity in blood isolates. Although in the current work there was no analysis performed on the growth curves between the blood and urine isolates
of each strain, this finding could indicate that blood isolates have a higher growth rate than the urine isolates, with exception to ST127 isolates.

Finally, histidine metabolism, which was enriched in cluster 1146 (Figure 4.9) showed mostly intermediate expression except for the drop in ST127 paired isolates (discussed below).

### 4.4.4 ST127 paired isolates showed exclusive proteome profile

HCL expression datasets show clear clustering for both isolates of ST127 strains. Gene ontology analysis of differentially expressed proteins between blood and urine isolates indicated an increase in pathways linked to aerobic respiration, peptidoglycan-based cell wall synthesis, cofactor binding and magnesium ion binding. A decrease in carbohydrate catabolic processes and aminobutyraldehyde dehydrogenase activity was indicated by down-regulation of a number of genes in blood isolates compared to those from urine. Biological processes related to organelle proteins, including the two lipid bilayers, were also altered.

EC18 paired isolates showed contrasting proteomic expression profiles to EC41 isolates. These were more elevated in EC18 than EC41 isolates, primarily in proteins involved in the TCA cycle pathway and galactose metabolism. Central metabolic pathways such as the TCA cycle are essential for UPEC fitness in vivo (Alteri et al., 2009). Increased expression in EC18 strains could be related to the previously discussed missing O-antigen, being a mechanism by which the isolates compensate and persist in both conditions (blood and urine).
In contrast, EC41 isolates exhibited a higher level of expression for proteins related to glutathione metabolism. This pathway is responsible for synthesis of glutathione during the exponential growth phase, which is then secreted into the medium (Bachhawat & Yadav, 2018). Many functions are known for this pathway, including transportation of amino acids (Bachhawat & Kaur, 2017), maturation of cytosolic iron-sulphur proteins (Sipos et al., 2002), iron metabolism (Kumar et al., 2011) and scavenging of glutathione from other lysing cells (Hanigan, 2014). With respect to metabolism involving iron, which is scarce in urine and blood, hydrolysis of L-glutathione can generate a more reactive thiol cysteinylglycine, which leads to ferric iron reduction to ferrous iron and initiates a iron redox cycling (Paolicchi et al., 2002). Therefore, glutathione metabolism is essential as it has key roles in cell viability. This also correlates with the finding that EC41 isolate proteins involved in the TCA cycle pathway were down-regulated compared to EC18 isolates. A study found that disrupting the degradation pathways of L-cysteine, serine and L-glutathione leading to TCA cycle flux reduction, and blocking glycogen formation, significantly improved the glutathione production (Zhang et al., 2016).

The urine isolates of UPEC ST127 strains showed up-regulation in P-loop containing nucleoside triphosphate hydrolase (Figure 4.23). It encodes a protein with a similar function to YqeH GTPase, which is required for ribosome assembly and essential for bacterial growth (Guo et al., 2013; Loh et al., 2007). This could indicate that the ST127 urine isolates are more resilient and active considering the environment they were isolated from, especially when there was not much genomic variation between each of the paired isolates. Additionally, the ability to utilize human urine as a substrate for growth advances the incidence of bacteriuria. Surviving the urine by maintaining the growth of
colonizing organisms requires more ribosomal synthesis (Lpe et al., 2016; Roos et al., 2006; Wullt et al., 1998).

Methane metabolism was the only pathway that was up regulated in blood isolates of UPEC ST127 strains when compared to the urine isolates (Figure 4.18). This pathway is responsible for processing methane as the only carbon and energy source. In *Methylococcus capsulatus*, a Gram-negative coccoid bacterium, metabolism of methane produces pyruvate as the end result in oxygen-limited environments (Kalyuzhnaya et al., 2013). Whilst it is not immediately clear why this pathway would be upregulated in UPEC from blood, this is an environment with reduced oxygen tension, which may drive alternative pathways for pyruvate metabolism (Ettwig et al., 2010; Yu & Liao, 2018). This could explain the elevated expression observed in pyruvate metabolism pathways in different clusters (1148, 1150, and 1153).

In addition, glycerophospholipid metabolism was also enriched across all paired isolates of UPEC ST127. Glycerophospholipids are essential to the Gram-negative bacterial membrane (Dalebroux, 2017). Bacteria need to cope and adapt to changes in the surrounding environment in order to survive. To do so, alterations to some proteins, specifically phospholipids, can reform cellular shape and add to the integrity of the cellular envelope. These adaptations will enable the bacteria to overcome environmental stresses, such as transitioning from the urinary tract into the blood stream (Rowlett et al., 2017).

Furthermore, proteins of the phosphotransferase system (PTS) showed an intermediate elevation in expression profile across all UPEC isolates. Membrane transport systems such as the PTS are required for interpreting environmental information. PTS is a major carbohydrate transport system in
bacteria which catalyses the phosphorylation of received sugar substrates associated with their crossing of the bacterial membrane (Deutscher et al., 2006). In fact, PTS proteins are necessary for the full deployment of pathogenic effects by many bacterial species (Kok et al., 2003).

Some pathways, such as that involved in histidine metabolism, were observed to have a low level of expression in ST127 isolates (Figure 4.24). The histidine metabolism pathway is highly conserved across many bacteria (Itoh et al., 2007) and its main function is to act as a catabolic pathway allowing utilisation of histidine as a carbon and energy source (Bender, 2012). Moreover, proteins of the histidine metabolism pathway (HisA, HisB and HisG) (See Appendix V) had the highest expression in the urine isolates of ST73. This may result in an advantage over isolates from the other STs, regarding growth in urine and establishing UTI.

4.4.5 Variations in the genomes of paired isolates

No detailed genomic analysis was carried out on these data as it was out of the scope of this study. Nevertheless, a whole genome alignment was performed to check for gross differences, for example missing genes in urine isolate compared to blood isolate and vice versa in addition to generating BRIG figures for the isolates of each ST.

Isolates of ST73 for both strains (EC22 and EC73) were largely identical with minor sequence changes. For EC22 isolates, the analysis could not find any hits matching the O type genes, which could indicate a deletion of the O-antigen, but no further analysis was conducted.
A similar description can also be used for EC47 isolates of ST38. EC86 isolates were stated as ST38 in a previous study (Alghoribi, 2015). However, the MLST analysis showed that the isolates are ST5150, which is within the ST38 clonal complex. However, both isolates were similar regarding the other aspects of the genomic analysis.

The genomic analysis of EC1 isolates (ST131) showed major differences among the “pair” regarding serotyping and fim-typing, which could indicate that the clinical case was associated with two different UPEC ST131 as it has been reported in a previous study (McNally et al., 2013). Thus, the discrepancy in number of deleted genes can be justified.

For EC10 blood isolate, there was an issue with sequencing the sample and results indicated sample contamination; it was discovered that the 30% of the sequence data were derived from an environmental organism identified as Stenotrophomonas maltophilia. Data for the ST131 strain in that sample were extracted and compared the urine isolate. Unfortunately, the differences between the assembled genomes were too substantial to allow strong conclusions to be drawn regarding the genomic variations between the EC10 isolates.

In a number of instances there were genes identified in blood isolates that were not present in the corresponding urine isolates. This may be seen as going against the hypothesis of a single strain from the urine ascending to the bloodstream, as gaining “new” genes would not be expected in this situation. One explanation for this might be that the single colonies originally recovered from the clinical samples are taken to represent what is probably a population of clones derived from a single strain that originally colonised the bladder. The
members of the derived clones may vary slightly in genetic content following many generations of replication.

Genomic analysis showed high similarity between isolates of ST127 from urine and blood. There were ~28 SNPs in the EC18 blood isolate compared to the one from urine and no SNPs between EC41 isolates. Plasmid analysis also showed no difference in blood/urine isolates of both strains of ST127. With this being said, our analysis showed no clear correlation between these SNPs and proteomic analysis for these strains.

### 4.4.6 Conclusion

In this chapter, using relative expression values proved to be more useful than just LFQ values, especially when dealing with many different stains. This methodology aims to facilitate a more approachable visualisation of the data by standardisation of the LFQ value to the maximum value observed for any particular protein. Although all paired isolates were maintained in LB for several sub-cultures, the changes in metabolic activities seem to be stable as noticed from our proteomic data.

From the previous analysis, we can see that not all UPEC isolates generated the same profiles for expression of metabolic pathways, although they have all been grown in the same conditions. Some STs have a higher metabolic potential that could provide more virulence advantages. For example, ST131 isolates demonstrated in a previous study a high metabolic potential and this may facilitate its wider dissemination to various environments (Gibreel et al., 2012a).
Furthermore, in many examples we have seen that all isolates of the same ST exhibit similar proteomic profiles. One of the main observations regarding the expression profile of ST127 strains is that there was minimal variation among the paired strains (blood and urine). Similarly, ST131 paired isolates presented limited levels of variation in proteomic profiles between blood and urine isolates.

However, in some STs, one isolate in a pair displayed a different proteomic profile to the other member of the pair. Nevertheless, this only was noted on a small number of proteins. We hoped the variation might give us insights into the different phenotypes seen for some paired isolates considering the proteomic approach we followed. Although we generated highly valid data sets, it did not fully meet our expectations.
Chapter Five: Proteomic Analysis of Uropathogenic *Escherichia coli* ST127 During Co-Culture with Uroepithelial Cells
5.1 Introduction

Having established the use of proteomic approaches to investigate UPEC pathogenicity, we sought to continue to study what variable changes can occur during co-culture with epithelial cells. As the previous chapter demonstrated, the variations in the proteomic profiles of the urine isolates and the blood isolates were not substantial. The two clinical sources of the isolates (urine and blood) represent somehow the status of before and after the transition from UTI to bloodstream infection, respectively. One question that remains to be answered is what happens during the key stage of a developing UTI when bacterial cells are exposed to the bladder epithelium.

Thus, in order to have a more complete picture of the proteome of UPEC ST127 in UTI, we investigated the proteome profile during growth in the presence of uroepithelial cell line (HT1197) (Robino et al., 2014; Tolg et al., 2011). Adherence to and invasion of the bladder epithelial cells is an important step toward establishing uropathogenicity (Martinez et al., 2000). UPEC isolates can then replicate within the intracellular space, either free in the cytosol or in enclosed envelope. The bacteria can survive longer in former condition (Kim et al., 2018). In enclosed vesicles, bacteria can evade and resist the physical flush of urine flow, the immune system and antibiotic treatment (Blango & Mulvey, 2010; Duncan et al., 2004).

Studies have shown that some UPEC isolates are opportunistic invaders of the intracellular space and not exclusively extracellular pathogens (Dhakal et al., 2008). This means that not all UPEC isolates are invasive pathogens and, in fact, isolates of UPEC ST127 showed no invasion of uroepithelial cells in many studies (Alkeskas et al., 2015; O'Hara, 2018).
Proteomic study of this situation of co-culture of UPEC ST127 isolates and HT1197 uroepithelial cells could shed some light into the essential bacterial proteins required for initiating UTI and possibly further invasion into the bloodstream. In addition, investigating the epithelial cell proteome profile could provide us with a base line of the reaction of the epithelial cells during UTI or in some cases persistent recurrent infection.

5.1.1 Aim and objectives

In this chapter we aim to uncover the mechanisms underlying metabolic adaptation of ST127 UPEC strains during growth in co-culture with the uroepithelial cell line HT1197, using a quantitative proteomics analysis. To achieve this Aim, the following objectives were set:

- Global proteomics profiling of ST127 UPEC strains in control (cells only) and co-culture conditions.
- Secretomics approach to identify novel factors secreted by ST127 UPEC strains when co-cultured with HT1197 cells.
- Bioinformatic analysis of proteomics and secretomics datasets.
5.2 Experimental design

In this chapter we have three UPEC strains of ST127 (EC18, EC41, and SA189) and *E. coli* K12 as a control strain. Each strain was co-cultured with uroepithelial cells HT1197 in MEM media. Each sample had three parts to analyse: bacterial cell pellets (soluble and insoluble); supernatants (the secretome); and the epithelial cell pellets. Three individual biological replicates were performed for each sample, bringing the total number of samples to 93 (Figure 5.1)(Figure 5.2).

Figure 5.1: Experimental design of the co-culture analysis of growth of ST127 UPEC strains and *E. coli* K12 with HT1197 uroepithelial cells. The figure indicates the sub category samples for each set of co-culture experiments with three biological replicates being conducted for each.
Figure 5.2: General work flow chart of the methodology followed for the co-culture proteomics studies to isolate bacterial proteins, epithelial proteins, and secretome.
5.3 Results

5.3.1 Co-culture experiment

As described earlier, ST127 UPEC strains (EC18, EC41, and SA189) in addition to non-pathogenic control strain *E. coli* K12 were co-cultured with HT1197 uroepithelial cells for 5 hours. Compared to previous chapters, sample volumes are relatively smaller than previous experiments.

With the exception of strain SA189, all strains showed a minimal effect on the cells and any effect could be suggested to be due to deprivation of the media. HT1197 uroepithelial cells co-cultured with SA189 showed significantly different morphological shape compared to the other *E. coli* strains (Figure 5.3).

5.3.1.1 Uropathogenic *E. coli* SA189 effect on HT1197 cells

Co-culture of SA189 with HT1197 resulted in a unique effect on the cells. Cells appeared to be larger and developed an oval shape after the 3rd hour of incubation, compared to the other *E. coli* strains (Figure 5.3).

The typical shape of uroepithelial cells is an elongated shape. This morphology was lost and replaced with cell rounding (Figure 5.3) in HT1197 cells exposed to SA189. In addition, the nucleus seemed to be more condensed. However, performing an MTT assay on the cell co-cultured with SA189 at different time points (3, 4, and 5 hours) indicated that the cells were still viable (Figure 5.4). Moreover, SA189 also caused the uroepithelial cells to detach from the flask surface, resulting in their flotation in the medium at the end of the five hours incubation.
Figure 5.3: HT1197 uroepithelial cells co-cultured with UPEC ST127 strains and E. coli K12. Photos were taken at 40x magnification using an inverted microscope.
Following that morphological finding, we sought to investigate the cause by going systematically from testing a range of possibilities, including nutrient deprivation to actual invasion of the epithelial cells. This part was done in collaboration with a student (G. Guilas), but the thesis author lead the work and analysed the data. First, HT1197 cells were treated with either an SA189 supernatant from 3 hours incubation (mid log phase) or an overnight culture. The supernatants were filtered twice and treated with gentamycin to minimize the bacterial presence. Both cultures indicated the absence of the effect in question, which could indicate that there is no effect of the bacterial secretome or nutrients deprivation, respectively.
The other possible reason for the effect could be due to the bacterial surface and its proteins/receptors following direct cell-cell contact. Heat killed (70°C and 90°C) UPEC SA189 in addition to gentamycin-killed cells were co-cultured with HT1197 cells. The unique effect was also absent in these cases which possibly indicates that the effect in question only occurs HT1197 cells are when co-cultured with live and untreated UPEC SA189. Further investigation of the cause of this effect was not performed.

5.3.2 Protein extractions

All samples were cultured from a single colony in 10 ml AUM media and incubated at 37°C for 3 hours (mid log phase). Protein extraction was performed after 5 hours of co-culture in 75 cm² cell culture flasks. For bacterial pellets, soluble proteins were extracted using the B-PER complete bacterial protein extraction reagent, while insoluble proteins were extracted using SDT lysis buffer.

BCA protein estimation indicated a higher protein extraction yield in insoluble fragments than the soluble one. Moreover, in unpaired t test analysis, insoluble proteins were statistically more significant in yield than those of the soluble fraction in all control sets and the EC18 co-culture experiment (Figure 5.5). While for epithelial cell protein extractions, we found no statistical significance in co-culture experiments compared to the control set, except for SA189 co-culture (Figure 5.6).
Figure 5.5: BCA estimation for bacterial protein yield extracted from *E. coli* K12, *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189 in co-culture conditions and MEM only control. Data are plotted as the mean of three biological replicates and error bars represent standard error of the mean (SEM). *p*-Value represented as (** *** = *p*≤0.0001, ** = *p*≤0.001, * = *p*≤0.01, n.s. = not significant) for the soluble vs insoluble protein fractions on comparison within each sample.

Figure 5.6: BCA estimation for protein yield extracted from HT1197 uroepithelial cells co-cultured with *E. coli* K12, *E. coli* EC18, *E. coli* EC41 and *E. coli* SA189. Data are plotted as the mean of three biological replicates and error bars represent standard error of the mean (SEM). *p*-Value represented as (** *** = *p*≤0.0001, ** = *p*≤0.001, * = *p*≤0.01, n.s. = not significant) for the proteins comparison to the control sample.
On SDS PAGE analysis, it was clear that HT1197 protein extractions from different co-culture sets and the control have a similar profile with well-resolved protein bands visible. However, due to exclusive effect of SA189 on HT1197, extracted proteins were very low yield, as shown on the gel (Figure 5.7).

Figure 5.7: Protein extractions for HT1197 cells co-cultured with *E. coli* EC18, *E. coli* EC41, *E. coli* SA189 and *E. coli* K12 loaded on SDS PAGE gels. Protein extracted by 2% SDS in 100 mM Tris-Hcl; 30 µg of protein was loaded for all samples. InstantBlue staining was used.
5.3.3 Analysis of the bacterial proteome profiles during co-culture with HT1197 cells

Following analysis of the results using MaxQuant and Perseus to interrogate the *E. coli* K12 and UPEC strains (UTI89 and 536) databases, 1301 proteins were identified in the bacterial pellets (soluble and insoluble) across all co-culture experiments including the control samples of *E. coli* strains in MEM. As there were three biological replicate experiments for each condition, proteins were classified as being present when observed in at least one replicate in either the experiment set or the control set. On the other hand, quantified proteins were those that were identified in a minimum of two replicates in each experimental group (co-culture experiment and control) (Table 5.1).

Table 5.1: Total proteome profiles of all strains of UPEC ST127 and *E. coli* K12. The differentially regulated proteins indicate an experiment to control ratio which indicated proteins up-regulated or down-regulated by >1.5 fold.

<table>
<thead>
<tr>
<th></th>
<th><em>E. coli</em> K12</th>
<th>UPEC EC18</th>
<th>UPEC EC41</th>
<th>UPEC SA189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total proteins identified in all strains</td>
<td>1301</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proteins identified per strain</td>
<td>362</td>
<td>391</td>
<td>398</td>
<td>1282</td>
</tr>
<tr>
<td>Proteins quantified for each strain</td>
<td>34</td>
<td>46</td>
<td>63</td>
<td>1086</td>
</tr>
<tr>
<td>Differentially regulated</td>
<td>15</td>
<td>30</td>
<td>33</td>
<td>288</td>
</tr>
</tbody>
</table>

Pathway enrichment analysis on subset of differentially expressed proteins was performed using DAVID webtool (Dennis *et al.*, 2003) (Table 5.2).
Table 5.2: Pathways up-regulated in the bacterial proteome of all strains of UPEC ST127 and E. coli K12. Pathways are those enriched in up-regulated proteins (≥1.5 fold) in co-culture vs. control.

<table>
<thead>
<tr>
<th>Pathways</th>
<th>K12</th>
<th>EC18</th>
<th>EC41</th>
<th>SA189</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA degradation</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycolysis / Gluconeogenesis</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Butanoate metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Ribosome production</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Aminoacyl-tRNA biosynthesis</td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyrosine metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alanine, aspartate and glutamate metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolism of xenobiotics by cytochrome P450</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutathione metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propanoate metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cysteine and methionine metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One carbon pool by folate biosynthesis pathway</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valine, leucine and isoleucine biosynthesis</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-Chloroacrylic acid degradation</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lipopolysaccharide biosynthesis</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1- and 2-Methylnaphthalene degradation</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novobiocin biosynthesis</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taurine and hypotaurine metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fructose and mannose metabolism</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
UPEC SA189 has more affected pathways when considering the number of isolated proteins (Table 5.1). Three pathways were shared between UPEC isolates in their up-regulation: glycolysis/gluconeogenesis; pyruvate metabolism; and ribosome production pathways.

Down-regulated pathways (n=5) were considerably fewer in number compared to the up-regulated pathways (n=27). Enriched down-regulated pathways were only detected for UPEC EC18 and SA189, with no pathways shared between them (Table 5.3).

Table 5.3: Pathways down-regulated in the bacterial proteome of all strains of UPEC ST127 and E. coli K12. Pathways are those enriched in down-regulated proteins (≥1.5 fold) in co-culture vs. control.

<table>
<thead>
<tr>
<th>Pathways</th>
<th>K12</th>
<th>EC18</th>
<th>EC41</th>
<th>SA189</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td></td>
<td></td>
<td></td>
<td>↓</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td></td>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td></td>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>ABC transporters</td>
<td></td>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td></td>
<td></td>
<td></td>
<td>↓</td>
</tr>
</tbody>
</table>

5.3.4 Analysis of the secretome profiles from UPEC and HT1197 cells during co-culture experiments

We have identified 1878 proteins in secretome across all co-culture experiments including the control samples of E. coli strains in MEM through interrogation of the E. coli K12 and UPEC strain UTI89 and 536 databases. We identified proteins present in at least one replicate of the three biological replicates (Table 5.4).
Table 5.4: Total secretome profiles of HTT1197 cells and all strains of UPEC ST127 and *E. coli* K12 in co-culture and control conditions. The identified proteins were based on the presence of at least one peptide in all replicates.

<table>
<thead>
<tr>
<th>Culture</th>
<th>Identified proteins</th>
<th>Culture</th>
<th>Identified proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEM medium</td>
<td>2</td>
<td>Control cells</td>
<td>854</td>
</tr>
<tr>
<td>K12</td>
<td>2</td>
<td>K12+cells</td>
<td>900</td>
</tr>
<tr>
<td>EC18</td>
<td>13</td>
<td>EC18+cells</td>
<td>969</td>
</tr>
<tr>
<td>EC41</td>
<td>250</td>
<td>EC41+cells</td>
<td>1146</td>
</tr>
<tr>
<td>SA189</td>
<td>634</td>
<td>SA189+cells</td>
<td>1375</td>
</tr>
</tbody>
</table>

For the secretome analysis, we checked for overlapping proteins using Venn diagrams to focus our analysis on specific categories. For instance, for each *E. coli* strain, four samples (e.g. UPEC EC18, control HT1197 cells, MEM medium and the co-culture of EC18 and HT1197) were used in one diagram. The analysis focused on the uniquely identified proteins in the co-culture condition after cross elimination with the control samples (MEM medium, *E. coli* in MEM, and HT1197 cells only). Thus, we have identified 412 proteins uniquely seen in EC18 co-culture, 391 proteins in EC41 co-culture, 305 proteins in SA189 co-culture, and 292 proteins in the K12 co-culture condition (Figure 5.8).

Also, for a closer look into these secretome profiles, all uniquely identified secretomes in the co-culture conditions were compared (Figure 5.9). This step was done to show what proteins are shared between all UPEC ST127 co-cultured conditions (n=17) and what are exclusive in EC18 (n=219), EC41 (n=187) and SA189 (n=235). Interestingly, EC18 and EC41 under co-culture conditions shared more proteins (n=155) than either did with the SA189 co-culture secretome.
Figure 5.8: Overlaps between identified secretomes in all strains of UPEC ST127 and E. coli K12 in co-culture and control conditions. Uniquely identified secretomes in co-culture conditions are highlighted in red. Venny 2.1 tool was used to produce the figures.
Figure 5.9: Overlaps between identified unique secretomes in co-culture conditions of all strains of UPEC ST127 and *E. coli* K12. These Venn diagrams showed all UPEC ST127 strains (EC18, EC41, and SA189) with (left) and without (right) K12. Venny 2.1 tool was used to produce the figures.
Further detail of the proteins shown in Figure 5.9 is given in Table 5.5, which displays the bacterial and epithelial proteins observed in various conditions. The Table shows that the bacterial proteins were the majority of the secreted proteins observed in co-culture conditions, except for the SA189 experiment. Moreover, although more bacterial proteins were shared between EC18 and EC41 than SA189 (as stated above), epithelial proteins were predominant regarding the proteins shared by SA189 and the other UPEC during co-culture conditions.

The bacterial secreted proteins were checked for secretome prediction using SignalP 5.0 tool (Petersen et al., 2011). Out of the final 540 bacterial secretome, only 55 proteins (10.19%) were predicted to be secreted proteins using E. coli UTI89 database (Table 5.5).

The epithelial secreted proteins were also analysed and confirmed using the Vertebrate Secretome Database (VerSeDa) (Cortazar et al., 2017). In general, the database showed that almost half (49.69%) of the epithelial proteins identified during co-culture were predicted to be secreted proteins using the Homo sapiens (GRCh37) dataset.

Analysis on DAVID tool showed that most of the secretory proteins are enriched for ABC transporter pathways. Furthermore, cross-examining the list of the bacterial secretome with the predicted secretory proteins in SignalP and the bacterial proteins extracted from the pellets, we found a list consisting of 192 proteins (35.56%) that were present in the secretome, but were not in the secretory database or in the bacterial extracted proteins (Table 5.6).
Table 5.5: Bacterial and epithelial proteins identified in the secretomes during co-culture of all strains of UPEC ST127 with HT1197 cells.

<table>
<thead>
<tr>
<th>Venn diagram areas</th>
<th>Total</th>
<th>Bacterial proteins</th>
<th>Bacterial secretome database</th>
<th>Epithelial proteins</th>
<th>Vertibrate secretome database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusive to &quot;EC18+cells&quot;</td>
<td>219</td>
<td>203</td>
<td>25</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Exclusive to &quot;EC41+cells&quot;</td>
<td>187</td>
<td>175</td>
<td>17</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Exclusive to &quot;SA189+cells&quot;</td>
<td>235</td>
<td>3</td>
<td>2</td>
<td>232</td>
<td>119</td>
</tr>
<tr>
<td>&quot;EC18+cells&quot; and &quot;EC41+cells&quot;</td>
<td>155</td>
<td>144</td>
<td>9</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>&quot;EC41+cells&quot; and &quot;SA189+cells&quot;</td>
<td>32</td>
<td>9</td>
<td>1</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>&quot;EC18+cells&quot; and &quot;SA189+cells&quot;</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>&quot;EC18+cells&quot;, &quot;EC41+cells&quot; and &quot;SA189+cells&quot;</td>
<td>17</td>
<td>5</td>
<td>1</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>866</td>
<td>540</td>
<td>55 (10.19%)</td>
<td>326</td>
<td>162 (49.69%)</td>
</tr>
</tbody>
</table>

These proteins were analysed on DAVID tool to investigate the enriched pathways. Proteins highlighted in red (Table 5.6) are the only ones that showed pathway enrichments. Figure 5.10 shows some of the enriched pathways, where carbohydrate metabolism pathways such as glycolysis/gluconeogenesis and pyruvate metabolism are common and shared in these conditions.
Table 5.6: Details of bacterial secreted proteins from various co-culture conditions of all strains of UPEC ST127 with HT1197 cells.

<table>
<thead>
<tr>
<th>Venn diagram areas</th>
<th>Secreted bacterial proteins that are not in the bacterial secretome database*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusive in &quot;EC18+cells&quot;</td>
<td>66</td>
</tr>
<tr>
<td>Exclusive in &quot;EC41+cells&quot;</td>
<td>67</td>
</tr>
<tr>
<td>Exclusive in &quot;SA189+cells&quot;</td>
<td>1</td>
</tr>
<tr>
<td>&quot;EC18+cells&quot; and &quot;EC41+cells&quot;</td>
<td>52</td>
</tr>
<tr>
<td>&quot;EC41+cells&quot; and &quot;SA189+cells&quot;</td>
<td>4</td>
</tr>
<tr>
<td>&quot;EC18+cells&quot; and &quot;SA189+cells&quot;</td>
<td>1</td>
</tr>
<tr>
<td>&quot;EC18+cells&quot;, &quot;EC41+cells&quot; and &quot;SA189+cells&quot;</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>192 / 540 (35.56%)</strong></td>
</tr>
</tbody>
</table>

* Red figures indicate conditions under which pathways were observed to have enriched proteins.
Figure 5.10: Pathways enriched in collections of proteins that were identified as part of the secretome under various co-culture conditions with UPEC ST127 strains and HT1197 cells, but were not present in bacterial secretome databases. Pathway enrichment of proteins seen exclusively in EC18, EC41, or both co-cultures conditions is shown.
5.3.5 Analysis of the epithelial proteome profiles during co-culture experiments with UPEC and HT1197 cells

Examining the results of MaxQuant analysis against the human proteome database revealed that 1658 proteins were identified in the combined data for epithelial cells across all experiments, including the control sample (Table 5.7).

These proteins were identified when present in at least one biological replicate in either the experiment set or the control set. The only proteins actually quantified in this work were those identified in a minimum of 2 replicates in each experimental group (co-culture experiment and control). The data indicated a good correlation between replicates of ≥ 0.9 (Appendix VI).

Table 5.7: Total proteome profiles of all epithelial cells co-cultured with UPEC ST127 strains and E. coli K12. The differentially regulated proteins indicate the experiment to control ratio where up-regulated and down-regulated proteins differed by >1.5 fold changes in expression.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Identified in all conditions</th>
<th>Identified (Per condition)</th>
<th>Quantified (Per condition)</th>
<th>Differentially regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT1197 + K12</td>
<td>1658</td>
<td>1393</td>
<td>992</td>
<td>17</td>
</tr>
<tr>
<td>HT1197 + EC18</td>
<td></td>
<td>1173</td>
<td>987</td>
<td>45</td>
</tr>
<tr>
<td>HT1197 + EC41</td>
<td></td>
<td>1202</td>
<td>980</td>
<td>55</td>
</tr>
<tr>
<td>HT1197 + SA189</td>
<td></td>
<td>1308</td>
<td>486</td>
<td>276</td>
</tr>
</tbody>
</table>
HCL clustering showed the unique profiling of the co-culture experiment of HT1197 epithelial cells with UPEC SA189 (Figure 5.11). On the other hand, the control condition and the rest of the UPEC (EC18 and EC41) co-culture conditions clustered together and apart from *E. coli* K12. This finding complements the finding of morphological analysis of the epithelial cells, as described above.

Figure 5.11: Hierarchical clustering (HCL) following label-free quantification (LFQ) of proteins from HT1197 cells (CLL) co-cultured with *E. coli* K12, EC18, EC41, SA189, compared to a control of culture (CON) in MEM alone. LFQ intensities are the output of the MaxLFQ algorithm and are based on the raw intensities and normalized on multiple levels.
The overlap in quantified proteins in all co-culture conditions was examined to identify the similarity in proteome profiles between these conditions (Figure 5.12). All conditions had a similar number of proteins, apart from the SA189 co-culture condition (Table 5.7). Interestingly, the majority of the proteins fell into two sections: proteins that are shared between all co-culture conditions (47.9%), and proteins shared with the exclusion of SA189 co-culture condition (47.4%).

Figure 5.12: Overlaps between quantified epithelial proteins of HT1197 cells co-cultured with UPEC ST127 strains and *E. coli* K12. Venny 2.1 tool was used to produce the figure.
Furthermore, the differentially regulated proteins were also overlapped to detect the percentage of shared proteins between the up-regulated and the down-regulated proteins. With only two proteins (0.6%) shared between the different regulations, it emphasises the similarity of the data obtained across all co-culture conditions.

The two identified proteins are ST13 and AKR1B1. ST13 protein is a fragment of Heat shock cognate 70 (Hsc70) interacting protein, which has a key role in protein folding and protein degradation. This protein was up-regulated in the EC18 condition and down-regulated in co-culture with SA189. The other protein (AKR1B1) was up-regulated in co-culture with EC41 and down-regulated in the SA189 condition. This protein plays a role in folate biosynthesis (Werner et al., 2011).

Regarding the differential regulation of proteins shared between all conditions, most of these proteins were exclusive to the SA189 co-culture experiments. For example, 147 of 186 up-regulated proteins and all 100 down-regulated proteins were exclusively observed in UPEC SA189 co-culture with HT1197 cells. In another words, the ‘All conditions’ proteins section is dominated by proteins seen in SA189 co-culture experiments (Table 5.8).

Table 5.8: Differentially regulated epithelial proteins of HT1197 co-cultured with UPEC ST127 strains and *E. coli* K12.

<table>
<thead>
<tr>
<th>Quantified proteins</th>
<th>No. of proteins</th>
<th>Up-regulated</th>
<th>Down-regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT1197 + K12, EC18, and EC41</td>
<td>478</td>
<td>38</td>
<td>3</td>
</tr>
<tr>
<td>All conditions</td>
<td>483</td>
<td>186</td>
<td>100</td>
</tr>
</tbody>
</table>
Pathways enriched in up-regulated proteins in the UPEC SA189 condition were associated with proteins linked to Parkinson's disease and Alzheimer's disease. However, other pathways were down-regulated, such as some linked to cell cycle, regulation of the actin cytoskeleton and focal adhesion.

On the other hand, the ‘pathogenic Escherichia coli infection pathway’ was the only pathway enriched in the co-culture experiments with other E. coli strains and the ‘ribosome production pathway’ was down-regulated.

The experiment that showed the unique effect of exfoliation (SA189) had undetectable levels of paxillin protein (Figure 5.13). Thus, this condition has a significant decrease ($p=0.0332$) in cytoskeletal paxillin protein compared to the control.

![Figure 5.13: LFQ intensity of paxillin protein extracted from uroepithelial HT1197 cells co-cultured with E. coli K12, E. coli EC18, E. coli EC41 and E. coli SA189. Data are plotted as the mean of three biological replicates and error bars represent standard error of the mean (SEM). $p$-Value represented as (* = $p\leq0.05$, n.s. = not significant) for the protein comparisons to the control sample.](image-url)
5.4 Discussion

5.4.1 Co-culture conditions and proteins extractions

In this experimental set up, we had three sets of samples to analyse: bacterial proteins; epithelial cells proteins; and the secretome (including both bacterial and epithelial proteins). These three samples types were technically challenging with regards to the volumes obtained, since the co-culture experiments could only be done in 75cm$^2$ cell culture flasks, resulting in a maximum of 15 ml yield (supernatant). In addition to low volume of supernatant obtained, this experimental set up was technically challenging due to low number of cells. This is in contrast to previous proteomics experiments where bacteria are normally grown in 100 ml of media. We expected low protein yield as consequence and hence could only run samples in three biological replicates without any technical replicates.

Furthermore, the UPEC SA189 co-culture condition presented a new challenge with the unique effect on HT1197 uroepithelial cells. The rounding of the cells and detachment from the flask surface made the sample separation more difficult between the bacterial and epithelial parts. However, we were able to separate them using a 5 µm filter tube to collect the E. coli cells (2 µm). To investigate this effect further, we performed an MTT assay. Interestingly, SA189 co-cultured epithelial cells (at different time points) showed no statistically significant difference to the control samples, indicating that there was no negative impact on cell viability. In a previous study, using HT1197 cells demonstrated an elevated level of nuclear de novo methyltransferase (DNMT1), known as the “maintenance DNMT”, when inoculated with UPEC UT189 (Tolg et al., 2011). It also revealed epigenetic modification and down-regulation of
growth-suppressor genes, such as cyclin-dependent kinase inhibitor 2A \((CDKN2A)\) and O6-methylguanine-DNA-methyltransferase \((MGMT)\), which suggests that infection initiates pathogen-specific modifications of the host cell epigenome.

Regarding protein extractions, bacterial protein extractions were performed similarly to what have been described in the previous chapters including isolation of soluble and insoluble proteins. The insoluble proteins represented the significant proportion of proteins, compared to soluble ones. The insoluble proteins were also more abundant than soluble ones in the previous chapters, which could say more about the extraction methods of using SDT lysis buffer being more effective for future studies. Epithelial proteins were extracted using SDT lysis buffer. The amounts of proteins were similar to the control, except for the SA189 co-culture experiment. This can be explained due to loss some of the epithelial proteins during the retrieval process using filter tubes. While trying to lyse the epithelial cells on the filter and eluting out the proteins, it is expected that a lower amount of proteins than anticipated will be recovered.

Secretomes form an essential class of proteins that control and regulate many biological and physiological processes, which make it clinically important regarding identifying new biomarkers and therapeutic targets (Dwivedi \textit{et al.}, 2016; Gagic \textit{et al.}, 2016; Monteiro \textit{et al.}, 2016). In this chapter, TCA precipitation assay enable us to isolate the secreted proteins in all co-culture and control conditions. The precipitation technique was found to be more successful with prolonged incubation (overnight) of the samples at 4°C.
5.4.2 Overview of the proteomic profiles’ analysis

Due to the presence of proteins from different organisms in this study, we used different databases to identify the extracted material. For the bacterial proteome, we used a collection of UPEC databases (n=2) in addition to the *E. coli* K12 database in order to cover diverse aspects of the bacterial proteome. Epithelial proteome data was analysed using the Human protein sequence database (forward and reversed version) (http://www.uniprot.org/, UniProt. Release May 2017). The reason not to include all of the databases together as one shared database, is to reduce and eliminate the expected background contamination of the epithelial proteins with the bacterial proteins and vice versa. However, we used a combined database for the secretome data analysis as it contains both bacterial and epithelial secreted proteins.

Pathogenicity islands within UPEC strains encode many bacterial toxins such as cytotoxic necrotising factor 1 (CNF1), vacuolating auto-transporter toxin (Vat), secreted autotransporter toxin (Sat), and Hemolysin A (Bien *et al.*, 2012; Terlizzi *et al.*, 2017; Wiles *et al.*, 2008). It has been reported that 30-50% of UPEC strains express Hemolysin A, which is associated with the infection severity (Mobley *et al.*, 1990).

Hemolysin A consists of 1024 aa/ 3075 bp. The molecular weight of the polypeptide is 110.25 kDa (predicted from the nucleotide sequence) according to the *E. coli* CFT073 database (Welch, 1991; Welch *et al.*, 2002). This protein is encoded in ST127 reference strain 536 by the gene *hlyA* (protein accession no. Q1R2T5) (Dobrindt *et al.*, 2002). In addition, it has been shown that HlyA toxin can target and degrade the paxillin protein, which plays a role in the protection of mammalian cell physical integrity (Dobrindt *et al.*, 2002).
Remarkably, the HlyA protein was only produced and detected during the SA189 isolate co-culture experiment in both the bacterial proteins and the secretome data (data not shown). The paxillin protein was not detected in the epithelial cells co-cultured with SA189, but it was detected in every other sample. This finding could explain the exfoliation effect on the epithelial cells observed in this study. HlyA toxin triggers proteolysis of host proteins, which can also target proteins involved in the cytoskeleton, inflammation and apoptosis of the cells. In addition, the proteolysis process requires stable interactions between HlyA toxin and host membranes (Dhakal & Mulvey, 2012). Further discussion on this topic is presented in section 5.4.2.3.

Furthermore, alpha hemolysin is encoded within the hlyCABD operon (Cavalieri et al., 1984; Linhartova et al., 2010). However, as mentioned in Chapter 3, the hlyA gene was not identified in the genome sequence of UPEC strains EC18 nor EC41. The UPEC SA189 genome possessed this gene as well as expressing the protein.

5.4.2.1 Bacterial proteomic changes during the co-culture experiments

Regarding the analysis of the bacterial and epithelial proteins, they were identified based on presence in at least one of the three biological replicates. This means that there are some proteins present in one condition, but not the other. Therefore, it is not statistically accurate to use this list of proteins to quantify all the differentially regulated proteins. Thus, we set the parameter to have a minimum of two replicates in each experimental group (co-culture experiment and control). Following this methodology, we can obtain average LFQ intensity readings of proteins for each experimental set. Sequentially, the
LFQ intensity ratios of the experiment to the control set can be used to highlight the proteins differentially regulated by 1.5 fold in both ways (up and down). Nevertheless, proteins expressed solely in co-culture experiments were examined (minimum of two replicates) for each strain (data not shown). The analysis generates a short list of proteins, which gave no extra information than what found in the following analysis regarding pathways enrichments.

For alpha hemolysin protein, the database showed that this protein is an extracellular protein (Figure 5.14). Analysis on the PSORTb tool v 3.0.2 (Yu et al., 2010) showed that the localization site is extracellular with the same prediction result.

For the bacterial proteins, enriched up-regulated pathways are listed in Table 5.2. The majority of the pathways fall under the metabolism pathways except for RNA degradation, ribosome and aminoacyl-tRNA biosynthesis pathways, as these come under genetic information processing (Kanehisa et
al., 2017). Carbohydrate metabolism pathways took the lead in this analysis as they were in the majority. For example, glycolysis/gluconeogenesis, citrate cycle (TCA cycle), butanoate metabolism and pyruvate metabolism pathways were among the top shared pathways among UPEC isolates.

The importance of gluconeogenesis and TCA cycle pathways have been described in a previous study indicating that mutant strains in these central metabolic pathways were unable to colonize the bladder nor the kidneys of in vivo mouse model (Alteri et al., 2009). Both pathways are connected and working together to form intermediates of a metabolic pathway, where the TCA cycle provides substrates for gluconeogenesis when cells use amino acids as a carbon source. This also connects the pyruvate metabolism pathway, when amino acid are degraded to oxaloacetate or to pyruvate, which can be converted to acetyl-CoA and enter the TCA cycle (Figure 5.15) (Gottschalk, 1986; Takahashi et al., 2000). The butanoate/butyrate pathway is also associated with the previous pathways via fermentation of acetyl-CoA to butyrate (Hadjifrangiskou et al., 2011; Monk et al., 2013).

Other carbohydrate metabolism pathways were also enriched for SA189 exclusively. In addition, amino acid metabolism pathways (e.g. tyrosine metabolism and alanine, aspartate and glutamate metabolism pathways) came second in up-regulated enrichment for SA189 strain.

From our results, we can see that E. coli K12 has a minimal effect on epithelial cell protein expression compared to the UPEC isolates. The majority of metabolic variations lay with UPEC SA189, which could indicate a high potential for virulence as observation in the co-culture experiments (Gibreel et al., 2012a).
In contrast, down-regulated pathways were all metabolism pathways except ABC transporter pathways (environmental information processing). More specifically, \(dppB\), \(dppD\), and \(dppF\) are the genes involved in down-regulation of Dipeptide transport systems in the SA189 co-culture experiment. These proteins are part of the binding-protein-dependent transport system for dipeptides, which are responsible for the translocation of substrate across the membrane (Haugen et al., 2007). This finding in particular is interesting as it contrasts previous studies that indicated the importance of \(dpp\) operon and reported its up-regulation during UTI (Mehta et al., 2015; Subashchandrabose et al., 2014). These contrasting results may be explained by strain level differences.

There was only a low number of down-regulated pathways and they were only enriched in UPEC EC18 and SA189 experiments.
5.4.2.2 Secreted proteins observed during the co-culture experiments with UPEC and K12 E. coli with HT1197 cells

The analysis regarding the secretome data was conducted using a new approach. The secretome contained both bacterial, epithelial and some cell medium proteins. We were interested to identify which proteins are secreted exclusively during the co-culture conditions. Thus, overlapping proteins with control HT1197 cells and UPEC isolates were excluded from the study.

The uniquely identified proteins were compared to see which proteins are exclusive to experiments with each UPEC isolate and what is shared across all experiments. The majority of these proteins were bacterial proteins. In addition, co-culture with SA189 led to a lower count of secreted bacterial proteins compared to with the other UPEC.

Interestingly, a high number of proteins that were exclusive in either EC18 or EC41 were not predicted as secretory proteins using SignalP tool. Applying SignalP as a prediction tool for signal peptides is currently common practice. Proteins with signal peptides can be targeted to the secretory pathway, but are not necessarily secreted (Nielsen, 2017). A signal peptide is the N-terminal part of the protein that is targeted to the secretory pathway in both prokaryotes and eukaryotes (von Heijne, 1990). SignalP has an accuracy rate of 78.1% for predicting signal sequences (Zhang & Henzel, 2004), yet it remains the leading tool and the most accurate over other predication tools (Choo et al., 2009).

Looking at the non-secretory proteins, ‘ribosome production’ was the major pathway enriched with these proteins. These proteins are generally not secreted but could be present in the secretome following cell lysis during the extraction process. Furthermore, although hemolysin A protein was isolated in
the secretome of UPEC SA189, analysis of its amino acid sequence on SignalP predicts the probability that the sequence does not have any kind of signal peptide (Figure 5.16). However, other studies indicated that type I secretion system is used to translocate hemolysin A (Gentschev et al., 2002; Thomas et al., 2014).

![SignalP-5.0 prediction](image)

**Figure 5.16:** Probability analysis of HlyA protein on SignalP 5.0 tool. The screenshot indicates the final prediction of hemolysin A protein as ‘other’ non signal peptide protein.

<table>
<thead>
<tr>
<th>Protein type</th>
<th>Signal peptide (Sec/SPI)</th>
<th>TAT signal peptide (Tat/SPI)</th>
<th>Lipoprotein signal peptide (Sec/SPI)</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood</td>
<td>0.029</td>
<td>0.0164</td>
<td>0.0061</td>
<td>0.9486</td>
</tr>
</tbody>
</table>

The main differences between Type I secretion system and the other three secretion systems are: (1) the formation of periplasmic intermediates is not required for the passenger proteins, as it secreted directly into the extracellular medium; (2) the targeting signals are located at the carboxyl terminus of the secreted proteins; (3) generally, no cleavage of the secretion signals is
necessary during or after secretion (Gentschev et al., 2002; Holland et al., 1990; Wagner et al., 1983).

Around 50-78% of UPEC isolates from pyelonephritis cases express hemolysin A (Bien et al., 2012; Ristow & Welch, 2016). In many studies focusing on understanding the effects of hemolysin both in vitro and in vivo infection models, it is hypothesized that hemolysin A is secreted at sublytic and lytic doses in vivo (Garcia et al., 2013; Grimminger et al., 1991; Mansson et al., 2007). Expression variability of hemolysin A in different clinical isolates limits our understanding of its regulation process (Murthy et al., 2019; Ristow & Welch, 2016). Yet, it makes a strong vaccine candidate as it not only modulates epithelial cell functions, but also disables macrophages and suppresses inflammatory responses (Dhakal & Mulvey, 2012; Smith et al., 2015).

5.4.2.3 Epithelial proteome determination during the co-culture experiments with E coli and HT1197 cells

The epithelial proteomes were processed similarly to the bacterial proteins, by comparing the co-culture conditions to the control sample of HT1197 uroepithelial cells in MEM medium alone. Noticeably, the epithelial proteins isolated in SA189 experiments were higher in numbers, which may not be surprising considering the unique effect of exfoliation on HT1197 cells. On the other hand, K12, EC18 and EC41 showed a shared and minimal effect on the epithelial proteome.

Overall, host cell exfoliation involves degradation of cell-cell and cell-matrix links, which seemingly requires break down of focal adhesions and other functionally related structures (Nakagawa et al., 2006). Actinin and filamin have
been shown to work cooperatively to enhance the cytoskeletal structures rigidity of the actin filament networks (Esue et al., 2009). In the current study, this was observed in the form of enriched pathways for down-regulated proteins (talin, actinin, rac, filamin, and vinculin) in the UPEC SA189 co-culture experiment, which revealed that the actin filament networks are probably weakening.

Lytic doses of HlyA can induce the exfoliation process. However, sublytic doses can activate some host proteases, including mesotrypsin, which will ultimately lead to the degradation of paxillin (Dhakal & Mulvey, 2012; Ristow & Welch, 2016). Paxillin protein is a 68 kDa multi domain support protein that can control the interaction forces between cytoskeletal changes and host cell adhesion via synchronized interactions with numerous essential and regulatory proteins (Deakin & Turner, 2008). In a previous study, human bladder epithelial cells (BECs) were infected with a reference UPEC isolate (UTI89) to test the stability of paxillin protein. The infected cells showed a decreasing level of the protein over the course of infection, with totally undetectable level after 4 hours of co-culture incubation time (Dhakal & Mulvey, 2012). These findings are consistent with those in the current study, as revealed by extraction of the whole proteome profile after 5 hours of uroepithelial cell co-cultured with different UPEC isolates.

HlyA initiates the degradation of paxillin. For example, targeted deletion of hlyA gene in UPEC strain UTI89 resulted in an elimination of paxillin degradation in BECs. However, complementation of the hlyA mutant restored the phenotype (Dhakal & Mulvey, 2012). HlyA has the possibility to induce host cell apoptosis and tissue turnover rates within BECs (Smith et al., 2008). In fact, possession of the hlyA gene and ability to express it enhances UPEC strain
tissue damage within the urinary tract, consequently causing more severe clinical UTI (Johnson, 1991; Marrs et al., 2005).

The fact that all UPEC strains in this study were clinical isolates that exhibited notable differences in their effect on the epithelial cells adds to the fact that pathogenesis in UPEC is still a complex subject to understand. It varies from one strain to another in the same ST, not to mention the host variations in clinical cases and differences in susceptibility from one case to another.

5.4.3 Conclusion

In this study, different UPEC isolates from clinical samples were examined and observed to vary in their effect on uroepithelial cells. Knowing that all isolates belong to ST127, the data suggest that not all isolates of the same ST necessarily mediate UTI using the same mode of action. Although not surprising, the extent of the difference between the effect of strain SA189 compared to other strains was notable.

The findings of the study also suggest that the observed effect of epithelial cells peeling off the cell culture plate is a type of defence mechanism towards infection with UPEC strain SA189. In addition, changes in cells shape were likely to be due the depletion of paxillin protein, which plays a role in maintaining the cytoskeletal structure of epithelial cells.

Furthermore, from the data, it is reasonable to conclude that the exfoliation effect on HT1197 cells is a result of the presence of the viable UPEC SA189, which may indicate that UPEC SA189 mediates the exfoliation effect via adhesion, potentially followed by invasion. This indication is yet to be confirmed.
This work has demonstrated the potential of whole proteome profiling for investigating host-pathogen interactions. The approach can be applied to investigate both bacterial (soluble and insoluble) and epithelial proteins, which enables detection of variations in enriched pathways using numerous data analytical tools such as DAVID. It should be acknowledged that extracting the proteins following five hours incubation places some challenges on the protein recovery process and cells lysis may result in protein degradation.
Chapter Six: Final Discussion and Concluding Remarks
6.1 Is shotgun proteomics an effective tool for investigating the bacterial proteome?

Uropathogeneic *E. coli* is the most common cause of UTI. These cases have been growing rapidly in the last decade with more AMR strains surfacing causing a hefty burden on the health services due to increased hospital visits, hospitalised patient, and prescribed antibiotics. Currently, there is a need to take a different approach to investigate these pathogens.

This study was able to determine the differences in proteomes of different isolates of UPEC ST127 and the data correlated well with the background data from the genome analysis, in some specific cases. In addition, proteomics data were highly reproducible for each strain as a result of standardisation of the incubation conditions and protein extraction methodology. The ability to have a quantitative analysis of the proteins gives us a more accurate understanding, in contrast to just using the genomic data or traditional protein analysis techniques (Western blot).

These conditions can be engineered to simulate the variety of environments UPEC isolates may need to survive in. The methods that have been developed will support future studies that could significantly enhance our understanding of the host-pathogen interaction.
6.2 Are there proteomic variations between paired isolates of UPEC from urosepsis and if so, what are they?

UPEC strains can invade the epithelial cells of the bladder and initiate an intracellular infection cycle. They can colonize the bladder and kidney, which may sometimes lead to a possibly fatal case of bacteraemia. In many studies it has been shown that \textit{E. coli} is the most common pathogen responsible for Gram-negative bacteraemia. Previous work with UPEC paired urine and bloodstream isolates revealed identical results for molecular typing and virulence gene content. However, different antibiotic susceptibility and metabolic activity profiles indicated phenotypic changes, potentially due to SNPs or epigenetic events.

Functional annotation enrichment analysis of differentially expressed proteins indicated up-regulation in pathways linked to aerobic respiration, peptidoglycan-based cell wall structure, cofactor binding, and magnesium ion binding. Down-regulated genes in a number of blood isolates were involved in carbohydrate catabolic process and aminobutyraldehyde dehydrogenase activity. Biological processes related to organelle proteins, including the two lipid bilayers, were also altered.

The current study revealed that strains of some STs have a higher metabolic potential that could provide more virulence advantages. The study also indicted that one isolate in a pair displayed a different proteomic profile to the other member of the pair, for some STs. However, in other STs, minimal proteomic variations between the paired strains were shown.
6.3 What metabolic adaptations of UPEC took place during infection of the epithelial cells?

The point at which UPEC invade the epithelial cells in the bladder and kidney is significant, as it determines the intracellular colonization and possibly further dissemination into the bloodstream, respectively.

While investigating UPEC ST127 pathogenicity during infection, proteomics revealed upregulation in the pentose phosphate and glycolysis/gluconeogenesis pathways in UPEC EC18 (an O-antigen deficient mutant). These two pathways could be important routes of carbon flux through the central metabolic pathways during growth in urine. Co-culture of UPEC SA189 with uroepithelial cells HT1197 cells leads to exfoliation effects in HT1197 not seen with other UPEC strains. Analysis of the UPEC SA189 secretome revealed highly abundant bacterial proteins, including hemolysin A. These data suggest that not all isolates of the same ST necessarily mediate UTI using the same mode of action.

This work has demonstrated the potential of whole proteome profiling for investigating host-pathogen interactions. The approach can be applied to investigate future co-cultured sets involving UPEC form different STs and primary epithelial cells from healthy individuals compared to patients suffering recurrent UTI.
### 6.4 Future directions

Following this work, to understand in more detail the different virulence of EC18 in GML, mutational analysis will be the recommended direction to follow in order to investigate some of these essential genes. A high-density transposon mutagenesis method (e.g. TraDIS) could be used (Goh et al., 2017; Hadjifrangiskou et al., 2012). This step will identify genes that are essential for growth and virulence in GML and these may represent targets for development of novel vaccines or therapies to prevent and treat UTI. Moreover, generating a recombinant strain of EC18 and attempting to restore the function of the O antigen could help us to understand its function in pathogenesis of this strain.

In addition, biofilm studies could be a useful confirmatory assay to distinguish between the three strains of UPEC ST127 (EC18, EC41, and SA189). As described above, O antigen mutation should have no effect on biofilm formation (Demirdjian et al., 2017), but it may affect the motility. In addition, study of biofilm formation in LB versus AUM should give insight into the capability of these strains to cause catheter related UTI.

Moreover, another approach can be followed regarding the data analysis in this study. As mentioned earlier, we processed both fractions of proteins (soluble and insoluble) separately. However, we pooled data for both fractions as one sample for full analysis. Analysing the data sets of both fractions individually may give a deeper insight into the proteome profile of UPEC ST127. Another approach that can be followed is to perform multiple reaction monitoring (MRM), a targeted proteomic analysis, to confirm the changes in expression of some key proteins (Banaei-Esfahani et al., 2017).
In order to have a more global representation of UPEC STs, we could include addition of a representative strain for each ST into the proteomic profiling. For example, UPEC ST131 reference strain EC958 (Forde et al., 2014); UPEC ST127 reference strain 536 (Hochhut et al., 2006); UPEC ST73 reference strain CFT073 (Welch et al., 2002); UPEC ST38 reference strain MRSN 346647 (Snesrud et al., 2017). This approach would enable similarities between the isolates (e.g. paired isolates) and their representative reference strain with respect to proteomic profile, and therefore would enable a more complete representation of the major lineages of UPEC causing UTI and bacteraemia.

Another direction that could give us a complete comparison of the major lineages of UPEC is to compare the urine isolates growing in AUM (not pooled human urine for standardization purposes) and the blood isolates growing in blood (Berendsen et al., 2017; Dubourg et al., 2019). This approach could give information about how members of each ST react in its original (isolation) environment.

Secretome analysis of the UPEC ST127 strains during infection in the in vivo G. mellonella model could be used to explore further the mechanism by which the bacteria can establish infection in this model. This may help explain why strain EC18 does not kill the larvae, when most other ST127 strains do. How relevant this model is to UTI, however, remains to be seen as strain EC18 was recovered from a case of urosepsis.

The observed link between HlyA and paxillin in this work is notable and exploring the paxillin protein cascade pathway and its related proteins could explain the mechanism by which UPEC SA189 was able to induce the cells to
lyse, in contrast to the other UPEC isolates. Seemingly important, the identification and classification of host receptors and signalling factors, which encourage adherence and invasion of host tissue cells by UPEC isolates, could explain the susceptibility of some people to acquire UTIs. For example, the novel mannoside FimH inhibitors were used to interfere with cell attachment and prevent the occurrence of UTI (Totsika et al., 2013). These findings could, hypothetically, lead to identification of novel targets for therapeutic applications to prevent or treat UTI.

Furthermore, additional validation assays such as an invasion assay and TEM are suggested for further confirmatory work regarding UPEC SA189 mode of action toward the uroepithelial cells.

Nevertheless, the work generated during this project has highlighted the significance of using proteomic approaches in bacterial pathogenesis studies. Proteomic tools are not just used as a confirmatory assay for specific proteins of interest. On the contrary, they can be used as an investigative and screening tool for studying the whole bacterial proteome in different scenarios, starting from bacterial metabolic adaptation in various environments through to host-pathogen interactions. The scientific interest that this project has received, indicate the promising future of this field for bacterial studies in the future.
Appendices
Appendix I: Intersecting proteins from up-regulated and down-regulated proteins in figure 3.12.

<table>
<thead>
<tr>
<th>Venn diagram clusters</th>
<th>Number of proteins</th>
<th>Gene names</th>
</tr>
</thead>
<tbody>
<tr>
<td>Down-regulated proteins</td>
<td>EC18, EC41 and SA189</td>
<td>68</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td>EC18 and EC41</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>EC18 and SA189</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>EC41 and SA189</td>
<td>42</td>
</tr>
</tbody>
</table>
Appendix II: Data correlation of label free quantification (LFQ) of three biological replicates of paired UPEC isolates grown in LB. All correlations are significant (≥ 0.9).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Correlation of biological replicates</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 and 2</td>
<td>1 and 3</td>
<td>2 and 3</td>
</tr>
<tr>
<td>EC1 U</td>
<td>0.951</td>
<td>0.961</td>
<td>0.965</td>
</tr>
<tr>
<td>EC10 U</td>
<td>0.962</td>
<td>0.969</td>
<td>0.953</td>
</tr>
<tr>
<td>EC18 U</td>
<td>0.979</td>
<td>0.945</td>
<td>0.929</td>
</tr>
<tr>
<td>EC41 U</td>
<td>0.959</td>
<td>0.927</td>
<td>0.927</td>
</tr>
<tr>
<td>EC22 U</td>
<td>0.954</td>
<td>0.968</td>
<td>0.966</td>
</tr>
<tr>
<td>EC73 U</td>
<td>0.956</td>
<td>0.970</td>
<td>0.965</td>
</tr>
<tr>
<td>EC47 U</td>
<td>0.955</td>
<td>0.966</td>
<td>0.956</td>
</tr>
<tr>
<td>EC86 U</td>
<td>0.961</td>
<td>0.974</td>
<td>0.957</td>
</tr>
<tr>
<td>EC1 B</td>
<td>0.973</td>
<td>0.968</td>
<td>0.964</td>
</tr>
<tr>
<td>EC10 B</td>
<td>0.965</td>
<td>0.969</td>
<td>0.965</td>
</tr>
<tr>
<td>EC18 B</td>
<td>0.976</td>
<td>0.939</td>
<td>0.954</td>
</tr>
<tr>
<td>EC41 B</td>
<td>0.976</td>
<td>0.951</td>
<td>0.953</td>
</tr>
<tr>
<td>EC22 B</td>
<td>0.924</td>
<td>0.942</td>
<td>0.933</td>
</tr>
<tr>
<td>EC73 B</td>
<td>0.962</td>
<td>0.970</td>
<td>0.953</td>
</tr>
<tr>
<td>EC47 B</td>
<td>0.960</td>
<td>0.966</td>
<td>0.973</td>
</tr>
<tr>
<td>EC86 B</td>
<td>0.964</td>
<td>0.972</td>
<td>0.956</td>
</tr>
</tbody>
</table>
**Appendix III:** Intersecting proteins from up-regulated and down-regulated proteins in figure 4.4.

<table>
<thead>
<tr>
<th>Venn diagram clusters</th>
<th>Number of proteins</th>
<th>Gene names</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST131, ST127, and ST73</td>
<td>3</td>
<td>phoU, nupG, slyB</td>
</tr>
<tr>
<td>ST127, ST73, and ST38</td>
<td>2</td>
<td>exbB, sodC</td>
</tr>
<tr>
<td>ST131, ST127, ST73, and ST38</td>
<td>1</td>
<td>yqfB</td>
</tr>
<tr>
<td><strong>Down-regulated proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST127 and ST73</td>
<td>5</td>
<td>slyD, folP, gsk, baer, UTI89_C1522</td>
</tr>
<tr>
<td>ST127, ST73, and ST38</td>
<td>5</td>
<td>coaA, ispF, ydhD, murD, narP</td>
</tr>
</tbody>
</table>
Appendix IV: Genomes assembly statistics for UPEC paired isolates.

<table>
<thead>
<tr>
<th>Sample id</th>
<th># contigs (&gt;= 0 bp)</th>
<th># contigs (&gt;= 1000 bp)</th>
<th>Total length (&gt;= 0 bp)</th>
<th>Total length (&gt;= 1000 bp)</th>
<th># contigs</th>
<th>Largest contig</th>
<th>Total length</th>
<th>GC (%)</th>
<th>N50</th>
<th>N75</th>
<th>N100</th>
<th>L50</th>
<th>L75</th>
<th># N's per 100 kbp</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC1U</td>
<td>238</td>
<td>125</td>
<td>5259561</td>
<td>5205625</td>
<td>125</td>
<td>32152</td>
<td>5243127</td>
<td>50.79</td>
<td>191087</td>
<td>109319</td>
<td>11</td>
<td>21</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC1BC</td>
<td>172</td>
<td>93</td>
<td>5319155</td>
<td>5273012</td>
<td>145</td>
<td>710931</td>
<td>5311228</td>
<td>50.79</td>
<td>263041</td>
<td>116284</td>
<td>7</td>
<td>14</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC10U</td>
<td>128</td>
<td>71</td>
<td>5314278</td>
<td>5288069</td>
<td>88</td>
<td>611843</td>
<td>5300795</td>
<td>50.65</td>
<td>204841</td>
<td>151505</td>
<td>9</td>
<td>17</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC10BC</td>
<td>390</td>
<td>325</td>
<td>10109725</td>
<td>10076909</td>
<td>352</td>
<td>655261</td>
<td>10097256</td>
<td>57.89</td>
<td>83546</td>
<td>27796</td>
<td>25</td>
<td>81</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC18U</td>
<td>194</td>
<td>67</td>
<td>4938441</td>
<td>4875836</td>
<td>110</td>
<td>557129</td>
<td>4904929</td>
<td>50.65</td>
<td>203407</td>
<td>77181</td>
<td>9</td>
<td>18</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC18BC</td>
<td>216</td>
<td>71</td>
<td>4963854</td>
<td>4893504</td>
<td>115</td>
<td>557129</td>
<td>4923453</td>
<td>50.64</td>
<td>190721</td>
<td>77181</td>
<td>9</td>
<td>19</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC41U</td>
<td>170</td>
<td>80</td>
<td>5304098</td>
<td>5259454</td>
<td>114</td>
<td>786064</td>
<td>5282179</td>
<td>50.53</td>
<td>170883</td>
<td>102403</td>
<td>9</td>
<td>19</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC41BC</td>
<td>157</td>
<td>65</td>
<td>5266795</td>
<td>5222096</td>
<td>99</td>
<td>786064</td>
<td>5244259</td>
<td>50.52</td>
<td>185544</td>
<td>110925</td>
<td>8</td>
<td>17</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC22U</td>
<td>128</td>
<td>80</td>
<td>5135902</td>
<td>5110130</td>
<td>103</td>
<td>468496</td>
<td>5127472</td>
<td>50.53</td>
<td>202690</td>
<td>107883</td>
<td>9</td>
<td>17</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC22BC</td>
<td>165</td>
<td>77</td>
<td>5146678</td>
<td>5114116</td>
<td>98</td>
<td>468602</td>
<td>5129491</td>
<td>50.54</td>
<td>234838</td>
<td>113556</td>
<td>9</td>
<td>16</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC73U</td>
<td>181</td>
<td>80</td>
<td>5138186</td>
<td>5102693</td>
<td>104</td>
<td>806262</td>
<td>5120183</td>
<td>50.51</td>
<td>287488</td>
<td>120801</td>
<td>6</td>
<td>13</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC73BC</td>
<td>2875</td>
<td>587</td>
<td>5988688</td>
<td>5400431</td>
<td>781</td>
<td>104403</td>
<td>5537324</td>
<td>50.67</td>
<td>23196</td>
<td>10625</td>
<td>72</td>
<td>157</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>EC47U</td>
<td>174</td>
<td>99</td>
<td>5250645</td>
<td>5216218</td>
<td>121</td>
<td>413278</td>
<td>5231387</td>
<td>50.55</td>
<td>148279</td>
<td>86740</td>
<td>11</td>
<td>22</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC47BC</td>
<td>167</td>
<td>89</td>
<td>5308622</td>
<td>5270159</td>
<td>115</td>
<td>676219</td>
<td>5289073</td>
<td>50.52</td>
<td>149666</td>
<td>87439</td>
<td>10</td>
<td>21</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC86U</td>
<td>215</td>
<td>100</td>
<td>5196477</td>
<td>5158203</td>
<td>120</td>
<td>452732</td>
<td>5173971</td>
<td>50.56</td>
<td>134514</td>
<td>65691</td>
<td>13</td>
<td>27</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>EC86BC</td>
<td>163</td>
<td>98</td>
<td>5235202</td>
<td>5200661</td>
<td>128</td>
<td>541598</td>
<td>5223729</td>
<td>50.55</td>
<td>149015</td>
<td>65691</td>
<td>10</td>
<td>22</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>
### Appendix V: Pathway enrichment of UPEC ST127 clusters.

<table>
<thead>
<tr>
<th>Cluster No.</th>
<th>Pathway ID</th>
<th>Pathway description</th>
<th>Gene count</th>
<th>FDR</th>
<th>Matching proteins in network (labels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>340</td>
<td>1113</td>
<td>Histidine metabolism</td>
<td>3</td>
<td>0.00236</td>
<td>hisA, hisB, hisG</td>
</tr>
<tr>
<td>1137</td>
<td>680</td>
<td>Methane metabolism</td>
<td>3</td>
<td>0.0447</td>
<td>fdoG, gpmA, gpmB</td>
</tr>
<tr>
<td>1139</td>
<td>IPR027417</td>
<td>P-loop containing nucleoside triphosphate hydrolase</td>
<td>10</td>
<td>0.0287</td>
<td>clpX, ffh,ftsE,gmk, hda, loliD, oblB, udk, ybiT</td>
</tr>
<tr>
<td>1100</td>
<td>Metabolic pathways</td>
<td>22</td>
<td>0.00426</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1110</td>
<td>Biosynthesis of secondary metabolites</td>
<td>20</td>
<td>0.000845</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>Log2FoldChange</td>
<td>q-value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>7</td>
<td>0.00715</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine, serine and threonine metabolism</td>
<td>6</td>
<td>0.0168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycerophospholipid metabolism</td>
<td>5</td>
<td>0.0296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microbial metabolism in diverse environments</td>
<td>16</td>
<td>0.0296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphotransferase system (PTS)</td>
<td>7</td>
<td>0.0296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fructose and mannose metabolism</td>
<td>6</td>
<td>0.0452</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>6</td>
<td>0.0452</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beta-Alanine metabolism</td>
<td>3</td>
<td>0.0496</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolic pathways</td>
<td>62</td>
<td>9.18E-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribosome</td>
<td>15</td>
<td>9.18E-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminoacyl-tRNA biosynthesis</td>
<td>9</td>
<td>1.03E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microbial metabolism in diverse environments</td>
<td>28</td>
<td>3.67E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of secondary metabolites</td>
<td>30</td>
<td>1.18E-04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>17</td>
<td>1.16E-04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathway</td>
<td>ID</td>
<td>E-value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-----</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>9</td>
<td>0.000631</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvate metabolism</td>
<td>10</td>
<td>0.000695</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>8</td>
<td>0.000782</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of amino acids</td>
<td>14</td>
<td>0.00892</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>11</td>
<td>0.00915</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>9</td>
<td>0.011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pentose phosphate pathway</td>
<td>8</td>
<td>0.000782</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of amino acids</td>
<td>14</td>
<td>0.00892</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>11</td>
<td>0.00915</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>9</td>
<td>0.011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycolysis / Gluconeogenesis</td>
<td>7</td>
<td>0.0175</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphingolipid metabolism</td>
<td>2</td>
<td>0.00101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolic pathways</td>
<td>40</td>
<td>8.27E-09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citrate cycle (TCA cycle)</td>
<td>8</td>
<td>1.29E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carbon metabolism</td>
<td>14</td>
<td>2.79E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of secondary metabolites</td>
<td>21</td>
<td>7.26E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microbial metabolism in diverse environments</td>
<td>16</td>
<td>0.000607</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fatty acid biosynthesis</td>
<td>4</td>
<td>0.00321</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactose metabolism</td>
<td>5</td>
<td>0.0152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>5</td>
<td>0.0152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1152</td>
<td>1212</td>
<td><strong>Fatty acid metabolism</strong></td>
<td>4</td>
<td>0.0152</td>
<td>c1187,fabD,fabH,fabl</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>---------------------------</td>
<td>---</td>
<td>------</td>
<td>---------------------</td>
</tr>
</tbody>
</table>
| 1100 | **Metabolic pathways** | 83 | 7.72E-15 | accA,accB,apt,aroK,aspA,atpD,c0775,c3528,c4155,c5350,carB,cdyA,cysE, 
| | | | | pepD,pepN,pgi,proA,proB,proC,proA,proB,proD,pykA,pyrG,pyrH,ribH,rpoA, 
| 1110 | **Biosynthesis of secondary metabolites** | 33 | 8.54E-05 | accA,accB,aroK,c0775,c3528,c4155,c5350,dapA,eno,fbp,fdhF,frdA,frdC,frmA, 
<p>| | | | | suhB,talB,tniA,ublE,yleB |
| 3020 | <strong>RNA polymerase</strong> | 4 | 8.65E-05 | rpoA,rpoB,rpoC,rpoZ |
| 1200 | <strong>Carbon metabolism</strong> | 17 | 0.000737 | accA,accB,cysE,eda,eno,fbp,fdhF,frdA,frdC,frmA,fumA,kdgK,pckA,pgi,pykA,talB,tniA |
| 10 | <strong>Glycolysis / Gluconeogenesis</strong> | 9 | 0.00258 | agp,c0775,eno,fbp,frmA,pckA,pgi,pykA,tniA |
| 61 | <strong>Fatty acid biosynthesis</strong> | 5 | 0.00524 | accA,accB,fabA,fabB,fabF |
| 550 | <strong>Peptidoglycan biosynthesis</strong> | 6 | 0.00524 | ddlA,ddlB,mrcA,murA,murC,murF |
| 3018 | <strong>RNA degradation</strong> | 5 | 0.00672 | dnaK,eno,groL,pnp,me |
| 480 | <strong>Glutathione metabolism</strong> | 5 | 0.0104 | gshB,pepB,pepD,pepN,speE |</p>
<table>
<thead>
<tr>
<th></th>
<th>Metabolism</th>
<th>ID</th>
<th>P-value</th>
<th>Signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>3060</td>
<td>Protein export</td>
<td>5</td>
<td>0.0104</td>
<td>ftsY, secA, secD, secY, yidC</td>
</tr>
<tr>
<td>520</td>
<td>Amino sugar and nucleotide sugar metabolism</td>
<td>8</td>
<td>0.0134</td>
<td>c0775, galU, glmM, glmU, murA, nanA1, nanE, pgi</td>
</tr>
<tr>
<td>620</td>
<td>Pyruvate metabolism</td>
<td>8</td>
<td>0.0213</td>
<td>accA, accB, frdA, frdC, fumA, gloA, pckA, pykA</td>
</tr>
<tr>
<td>1212</td>
<td>Fatty acid metabolism</td>
<td>5</td>
<td>0.026</td>
<td>accA, accB, fabA, fabB, fabF</td>
</tr>
<tr>
<td>30</td>
<td>Pentose phosphate pathway</td>
<td>6</td>
<td>0.0307</td>
<td>c0775, eda, fbp, kdgK, pgi, talB</td>
</tr>
<tr>
<td>562</td>
<td>Inositol phosphate metabolism</td>
<td>2</td>
<td>0.0372</td>
<td>suhB, tpiA</td>
</tr>
</tbody>
</table>
Appendix VI: Data correlation of label free quantification (LFQ) of three biological replicates of uroepithelial cells co-cultured with *E. coli* strains. All correlations are significant (≥ 0.9).

<table>
<thead>
<tr>
<th>Sample</th>
<th>Correlation of biological replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 and 2</td>
</tr>
<tr>
<td>HT1197 control</td>
<td>0.972</td>
</tr>
<tr>
<td>K12 co-culture</td>
<td>0.983</td>
</tr>
<tr>
<td>EC18 co-culture</td>
<td>0.967</td>
</tr>
<tr>
<td>EC41 co-culture</td>
<td>0.960</td>
</tr>
<tr>
<td>SA189 co-culture</td>
<td>0.918</td>
</tr>
</tbody>
</table>
Appendix VII: CD containing the proteomics data of all chapters in the thesis in the following files:

- Chapter 3 proteomics Data.
- Chapter 4 proteomics Data.
- Chapter 5.1 Bacterial proteomics Data.
- Chapter 5.2 Secretome Data.
- Chapter 5.3 Epithelial proteomics Data.
References


medical and biological research = Revista brasileira de pesquisas medicas e biologicas 41, 877-883.


resistant Enterobacteriaceae are associated with severe sepsis at admission: implication for empirical therapy. *Medicine* 95, e3694.


Petty, N. K., Ben Zakour, N. L., Stanton-Cook, M., Skippington, E., Totsika, M., Forde, B. M., Phan, M. D., Gomes Moriel, D., Peters, K. M., Davies, M.,


Microbiology and molecular biology reviews: MMBR 72, 266-300, table of contents.


Wullt, B., Sundén, F. & Grabe, M. (2019). Asymptomatic bacteriuria is harmless and even protective: Don't treat if you don't have a very specific reason. European Urology Focus 5, 15-16.


