

**Genomic variability in the control of
the signaling molecule cyclic-di-GMP
and its role in pathogenicity and
biofilm formation of commensal and
pathogenic *Escherichia coli***

Inaugural-Dissertation

To obtain the academic degree of

Doctor rerum naturalium (Dr. rer. nat.)

**Submitted to the Department of Biology, Chemistry and Pharmacy
of the Freie Universität, Berlin**

by

Tatyana Leonidovna Povolotsky

Donets'k, USSR

September 2014

This work was accomplished between January 2011 and August 2014 at the Freie Universität, Berlin under the supervision of Prof. Dr. Regine Hengge.

1st Reviewer: Prof. Dr. Regine Hengge

2nd Reviewer: Prof. Dr. Rupert Mutzel

Date of Defence: December 17, 2014

List of publications stemming from this study

1. **Povolotsky, T.L.** and Hengge, R. (2012) 'Life-style' control networks in *Escherichia coli*: Signaling by the second messenger c-di-GMP. *J of Biotechnology* 160(1-2): 10-16.
2. Richter*, A.M.; **Povolotsky***, T.L.; Wieler, L.H.; Hengge, R. (2014) Cyclic-di-GMP signaling and biofilm-related properties of the Shiga toxin-producing 2011 German outbreak *Escherichia coli* O104:H4. *EMBO Molecular Medicine* 6, 1622-1637
(*These authors contributed equally to this work)
3. **Povolotsky, T.L.**; Klauck, G.; Hengge, R. (2014) Genomic variability in the control of the signaling molecule cyclic-di-GMP in commensal and pathogenic *Escherichia coli*.
(In preparation)

Acknowledgements

My sincere thanks to Prof. Dr. Regine Hengge for seeing my potential and taking me on as her student to mentor, guide, encourage, enabling me to work on a stimulating topic as well as providing the opportunity to meet and interact with other scientists in the global community via conferences and meetings, also for having the patience and finding the time to discuss progress and future direction of the project with helpful feedback.

I would like to especially thank Gisela Klauck for her patience, guidance, support, making the time for discussions, lending of expertise and mentoring me in general in the wet lab as well as practical assistance in lab work.

My gratitude extends to Alexandra Poßling for inoculating cultures early in the morning and running the radioactive portions of some of my experiments, as well as preparing figures 4.33 and 4.41 included in this work. I would also like to thank Franziska Skopp for all her help with protein purification protocols, her patience and cheery disposition. Also, a thank you to Anja Richter for her collaboration in working in Prof. Dr. Wieler's lab with pathogenic *E. coli*, for allowing the use of her figures 4.44 and 4.45 to be included in this work and help with the French Press.

I would like to thank my former and present colleagues for the stimulating discussions and their help and expertise in the lab: Eberhard Klauck, Nicole Sommerfeldt, Franziska Mika, Diego Serra, Tina Jaenicke, Martin Lorkowski, Vanessa Pfiffer, Susanne Herbst, Anja Richter, Alexandra Poßling, Franziska Skopp, Gisela Klauck, Natalia Tschowri, Olga Sarenko and Sandra Lindenberg.

Thanks to Angelika Strack and Raúl Trepel for keeping the lab running smoothly by preparing the media, stock solutions, cleaning of glassware and sterilization. Their contribution cannot be understated as it enabled me to focus directly on the project.

My thanks to our former and present secretaries for making the lab environment and bureaucratic issues run smoothly; Mary Wurm and Felicitas Reindl.

I would like to thank my GRK1673 colleagues, especially Annesha Lahiri, for the great atmosphere and support. With special acknowledgement to the program coordinator Esther-Maria Antão and Elisabeth Otto for helping with all of the bureaucracy of the program, for filling all of the orders and being there to answer all the questions, and Prof. Dr. Wieler for heading the program, for his advice and constructive criticism of the project as well as allowing us to use his lab for working with pathogenic *E. coli*.

I would like to express my unyielding gratitude to my parents, Vera and Leonid, my sister, Olga, and my grandmother, Bronislava, for their continued and unwavering support no matter in what corner of the world I found myself in and what adventure in life I was determined to pursue.

My thanks and esteem to my family and friends, especially Sarra Wolynskaja, Elena Volkova, Dorian Weber and Eitan Gershenson, who were there with me through all my trials and tribulations, who were there to share with me my joys and sorrows and who helped shape me into the person I am today.

Table of Contents

| | |
|---|------|
| Figures | VIII |
| Tables..... | IX |
| Abbreviations | X |
| Summary..... | 1 |
| 1 Introduction | 5 |
| 1.1 Commensal and pathogenic <i>E. coli</i> | 5 |
| 1.2 Bacterial environmental adaptation and biofilm formation..... | 10 |
| 1.3 C-di-GMP as a ubiquitous second messenger in bacteria..... | 11 |
| 1.3.1 Biosynthesis and degradation of c-di-GMP by GGDEF, EAL and HD-GYP domain- containing proteins..... | 11 |
| 1.3.2 Multiple GGDEF/EAL domain-containing proteins in single organisms..... | 13 |
| 1.3.3 Enzymatically inactive degenerate GGDEF/EAL domain-containing proteins in <i>E. coli</i> .. | 14 |
| 1.3.4 The 29 GGDEF/EAL domain-containing proteins <i>E. coli</i> K-12 W3110..... | 16 |
| 1.3.5 C-di-GMP binding effectors and their targets | 17 |
| 1.3.6 C-di-GMP inversely controls motility and curli fiber formation in <i>E. Coli</i> | 18 |
| 1.3.7 Biofilm matrix components | 21 |
| 1.3.8 C-di-GMP's role in virulence regulation..... | 24 |
| 2 Aims | 26 |
| 3 Materials and Methods | 27 |
| 3.1 Chemicals, materials and technical equipment..... | 27 |
| 3.2 Media and Additives | 29 |
| 3.2.1 Media..... | 29 |
| 3.2.2 Additives | 30 |
| 3.3 Bacterial strains, bacteriophages, plasmids and oligonucleotide primers used in this work | 30 |
| 3.3.1 Bacterial strains..... | 30 |
| 3.3.2 Bacteriophages | 32 |
| 3.3.3 Plasmids | 33 |
| 3.3.4 Primers..... | 34 |
| 3.4 Microbiological methods..... | 37 |
| 3.4.1 Sterilization | 37 |
| 3.4.2 Long-term storage of bacterial strains..... | 37 |
| 3.4.3 Growth conditions | 37 |
| 3.4.4 Bacterial motility assay | 37 |
| 3.4.5 Determination of the cell density of liquid bacterial cultures | 37 |
| 3.4.6 Transformation..... | 37 |

| | | |
|--------|--|-----|
| 3.4.7 | P1 transduction | 38 |
| 3.5 | Molecular biological and biochemical methods..... | 39 |
| 3.5.1 | Determination of DNA concentrations and sample purity | 39 |
| 3.5.2 | Polymerase chain reaction (PCR)..... | 39 |
| 3.5.3 | Preparation of chromosomal DNA and plasmid DNA | 39 |
| 3.5.4 | Agarose gel electrophoresis | 40 |
| 3.5.5 | Plasmid construction | 40 |
| 3.5.6 | Construction of chromosomal <i>lacZ</i> fusions | 41 |
| 3.5.7 | Construction of deletion/ insertion mutants | 41 |
| 3.5.8 | Flipping out resistance cassettes..... | 41 |
| 3.5.9 | Complementation experiments | 42 |
| 3.5.10 | Determination of β -galactosidase activity | 42 |
| 3.5.11 | RNA preparation for primer extension | 42 |
| 3.5.12 | Primer extension analysis | 43 |
| 3.5.13 | Extraction of proteins from bacterial culture samples | 44 |
| 3.5.14 | SDS-polyacrylamide gel electrophoresis (SDS-PAGE)..... | 44 |
| 3.5.15 | PageBlue Protein Staining | 45 |
| 3.5.16 | Immunoblot analysis (Western blot) | 45 |
| 3.5.17 | Protein over-expression and purification | 46 |
| 3.5.18 | Determination of protein concentrations | 47 |
| 3.5.19 | <i>In vivo</i> protein-protein interaction analysis..... | 47 |
| 3.6 | Databases and bioinformatic analyses..... | 48 |
| 3.6.1 | Strains and genome sequences..... | 49 |
| 3.6.2 | Phylogenetic, hydropathy and sequence analyses | 50 |
| 3.6.3 | Motif analysis | 51 |
| 4 | Results | 52 |
| 4.1 | Bioinformatic Analysis – variations in GGDEF/EAL genes in 61 <i>E. coli</i> genomes..... | 52 |
| 4.1.1 | Genes encoding GGDEF domain-containing proteins | 56 |
| 4.1.2 | EAL domain-containing proteins—core and pathotype-specific proteins..... | 71 |
| 4.1.3 | Degenerate GGDEF/EAL domain proteins..... | 86 |
| 4.1.4 | Biofilm matrix components and transcriptional regulators..... | 88 |
| 4.2 | Regulation of newly identified DGCs and PDEs..... | 94 |
| 4.2.1 | Expression of <i>dgcX</i> | 95 |
| 4.2.2 | Expression of <i>yneF</i> | 99 |
| 4.2.3 | Expression of <i>pdeT</i> | 101 |
| 4.3 | Functional characterization of newly identified DGCs and PDEs | 104 |
| 4.3.1 | Protein-protein interaction: PdeT (VmpA) does not interact with YcdT..... | 104 |
| 4.3.2 | Diguanylate cyclase activity of DgcX and YneF | 106 |

| | | |
|-------|--|-----|
| 4.3.3 | Can DgcX complement a <i>yegE</i> knockout?..... | 107 |
| 4.3.4 | Morphology of macrocolony biofilms in K-12 strains with <i>dgcX</i> integrated into the chromosome..... | 111 |
| 4.4 | Biofilm producing properties of O104:H4 outbreak strain (LB226692)..... | 113 |
| 4.4.1 | Morphology of macrocolony biofilms of O104:H4 outbreak strain (LB226692) | 114 |
| 4.4.2 | Expression of biofilm regulator CsgD in O104:H4 outbreak strain (LB226692)..... | 115 |
| 5 | Discussion | 117 |
| 5.1 | Acquisition of a novel DGC, DgcX, and changes in other DGC/PDE genes in EAECs including outbreak strain O104:H4 LB226692 | 117 |
| 5.2 | Acquisition of a novel PDE, PdeT (VmpA) and changes in other DGC/PDE genes in EHECs of the O157:H7 serotype..... | 119 |
| 5.3 | The uncoupling of the PDE YcgG's sensory domain from the EAL domain | 120 |
| 5.4 | Acquisition of a novel PDE, PdeY and changes in other DGC/PDE genes in UPECs | 121 |
| 5.5 | DGC/PDE complement reduced in K-12 strains W3110, MG1655, MDS42, DH10B and BW2952..... | 122 |
| 5.6 | Loss of curli through prophage insertion in the <i>mlrA</i> gene | 123 |
| 5.7 | Cellulose and PGA, biofilm matrix components least maintained among <i>E. coli</i> strains . | 124 |
| 5.8 | Conservation of a core eight c-di-GMP signaling housekeeping genes in <i>E. coli</i> | 126 |
| 5.9 | O104:H4 outbreak strain's (LB226692) unique ensemble of biofilm forming properties may contribute to its high virulence..... | 128 |
| 6 | References | 130 |

Figures

| | |
|---|-----|
| FIGURE 1.1: SCHEMATIC REPRESENTATION OF <i>ESCHERICHIA COLI</i> GROUPS AND PATHOTYPES. | 6 |
| FIGURE 1.2: PATHOGENIC SCHEMA OF DIARRHEAGENIC <i>E. COLI</i> | 9 |
| FIGURE 1.3: C-DI-GMP TURNOVER. | 13 |
| FIGURE 1.4: THE STANDARD COMPLEMENT OF 29 GGDEF/EAL DOMAIN-CONTAINING PROTEINS IN <i>E. COLI</i> K-12 W3110. ... | 17 |
| FIGURE 1.5: THE REGULATORY NETWORK OF MOTILITY, CURLI FIBERS AND CELLULOSE SYNTHESIS. | 21 |
| FIGURE 1.6: GENOMIC LAYOUTS OF A) <i>CSG</i> GENES, B) <i>BCS</i> GENES AND C) <i>PGA</i> GENES IN THE K-12 W3110. | 22 |
| FIGURE 4.1: FREQUENCY OF PROTEIN CONSERVATION OF THE PUTATIVE DIGUANYLATE CYCLASES | 56 |
| FIGURE 4.2: ALIGNMENT OF GENOMIC REGIONS OF THE <i>YNEF</i> GENE COMPARED IN 5 STRAINS OF <i>E. COLI</i> | 59 |
| FIGURE 4.3: ALIGNMENT OF GENOMIC REGIONS OF <i>PGAA-YCDT</i> GENES COMPARED IN 5 STRAINS OF <i>E. COLI</i> | 61 |
| FIGURE 4.4: GENOMIC ALIGNMENT OF THE <i>YDDV</i> GENE IN UPEC 536 WITH THE MODIFIED REFERENCE STRAIN W3110. | 62 |
| FIGURE 4.5: PREDICTED TOPOLOGY OF THE <i>DGCX</i> PROTEIN | 63 |
| FIGURE 4.6: ALIGNMENT OF GENOMIC REGIONS OF THE <i>DGCX</i> GENE COMPARED IN 5 STRAINS OF <i>E. COLI</i> | 64 |
| FIGURE 4.7: GENOMIC LAYOUT OF THE <i>DGCX</i> GENE. | 65 |
| FIGURE 4.8: SEQUENCE ALIGNMENT OF THE NOVEL <i>DgcX</i> AND ITS CLOSEST TWO HOMOLOGS <i>YeaI</i> AND <i>YcdT</i> | 66 |
| FIGURE 4.9: HYDROPATHY PLOT OF THE <i>DgcY</i> PROTEIN. | 67 |
| FIGURE 4.10: GENOMIC LAYOUT OF REGIONS OF <i>E. COLI</i> SMS-3-5. | 68 |
| FIGURE 4.11: GGDEF DOMAIN SEQUENCE ALIGNMENT OF THE NOVEL <i>DgcY</i> AND ITS CLOSEST HOMOLOG <i>YdeH</i> | 68 |
| FIGURE 4.12: CONSENSUS SEQUENCE OF THE MOTIF FOUND IN <i>DgcX</i> , <i>YcdT</i> AND <i>YeaI</i> PROTEINS. | 69 |
| FIGURE 4.13: PHYLOGENETIC TREE OF THE GGDEF DOMAIN-CONTAINING PROTEINS. | 71 |
| FIGURE 4.14: FREQUENCY OF PROTEIN CONSERVATION OF THE PUTATIVE PHOSPHODIESTERASES. | 72 |
| FIGURE 4.15: ALIGNMENT OF GENOMIC REGIONS OF THE <i>YAHA</i> GENE COMPARED BETWEEN 5 STRAINS OF <i>E. COLI</i> | 74 |
| FIGURE 4.16: <i>YAHA</i> (PDE) SEQUENCE ALIGNMENT FROM K-12 W3110 STRAIN AND FROM 55989 (EAEC) STRAIN. | 74 |
| FIGURE 4.17: SEQUENCE ALIGNMENT OF THE HTH MOTIF IN THE <i>YAHA</i> PROTEIN BETWEEN K-12 W3110 AND 55989. | 74 |
| FIGURE 4.18: OPERON LAYOUT OF THE <i>YCDT</i> GENE. | 78 |
| FIGURE 4.19: SEQUENCE ALIGNMENT OF THE NOVEL <i>PdeT</i> (PUTATIVE PDE) AND ITS CLOSEST HOMOLOG <i>Rtn</i> (PDE). | 79 |
| FIGURE 4.20: GENOMIC LAYOUT OF <i>PdEX</i> IN <i>E. COLI</i> 536 (UPEC). | 80 |
| FIGURE 4.21: SEQUENCE ALIGNMENT OF THE NOVEL <i>PdEX</i> (PUTATIVE PDE) AND ITS CLOSEST HOMOLOG <i>Ylie</i> (PDE). | 81 |
| FIGURE 4.22: GENOMIC LAYOUT OF THE <i>PDEY</i> GENE. | 82 |
| FIGURE 4.23: SEQUENCE ALIGNMENT OF THE ENTIRE NOVEL <i>PdeY</i> (PUTATIVE PDE) AND ENTIRE <i>Yhjh</i> (PDE). | 82 |
| FIGURE 4.24: GENOMIC LAYOUT OF <i>PDEZ</i> IN <i>E. COLI</i> E24377A (ETEC). | 83 |
| FIGURE 4.25: EAL-DOMAIN SEQUENCE ALIGNMENT OF THE NOVEL <i>PdeZ</i> (PUTATIVE PDE) AND THE EAL DOMAINS OF ITS CLOSEST HOMOLOGS <i>CsrD</i> (DEGENERATE GGDEF/EAL DOMAIN PROTEIN) AND <i>YcIR</i> (PDE). | 84 |
| FIGURE 4.26: PHYLOGENETIC TREE OF THE EAL DOMAIN-CONTAINING PROTEINS (20). | 85 |
| FIGURE 4.27: FREQUENCY OF PROTEIN CONSERVATION OF DEGENERATE GGDEF/EAL DOMAIN-CONTAINING PROTEINS. | 86 |
| FIGURE 4.28: FREQUENCY OF PROTEIN CONSERVATION INVOLVED IN CURLI SYNTHESIS. | 88 |
| FIGURE 4.29: ALIGNMENT OF GENOMIC REGIONS OF THE <i>MLRA</i> GENE COMPARED IN 5 STRAINS OF <i>E. COLI</i> | 90 |
| FIGURE 4.30: FREQUENCY OF PROTEIN CONSERVATION INVOLVED IN CELLULOSE BIOSYNTHESIS. | 91 |
| FIGURE 4.31: FREQUENCY OF PROTEIN CONSERVATION INVOLVED IN <i>PGA</i> BIOSYNTHESIS. | 93 |
| FIGURE 4.32: PROMOTER REGION OF THE <i>DGCX</i> (TOP) AND <i>YNEF</i> GENE (BOTTOM). | 96 |
| FIGURE 4.33: DETERMINATION OF TRANSCRIPTIONAL START SITES BY PRIMER EXTENSION FOR <i>YNEF</i> (LEFT) AND <i>DGCX</i> (RIGHT) 5' END MRNA. | 97 |
| FIGURE 4.34: EXPRESSION OF <i>LACZ</i> REPORTER FUSIONS TO <i>DGCX</i> AND OTHER <i>E. COLI</i> DGC GENES. | 98 |
| FIGURE 4.35: <i>DGCX</i> STRONGLY EXPRESSED AND UNDER PARTIAL CONTROL OF <i>RPOs</i> IN K-12. | 99 |
| FIGURE 4.36: <i>YNEF</i> WEAKLY EXPRESSED AND UNDER <i>RPOs</i> CONTROL IN K-12. | 100 |
| FIGURE 4.37: <i>YNEF</i> WEAKLY EXPRESSED BUT UNDER <i>RPOs</i> REGULATION AND PARTIAL <i>CsgD</i> CONTROL IN K-12. | 101 |
| FIGURE 4.38: <i>PdeT</i> IN JOINT OPERON WITH <i>YcdT</i> SHOWS ALMOST NO EXPRESSION IN K-12. | 103 |
| FIGURE 4.39: <i>PdeT</i> AS ITS OWN TRANSCRIPTIONAL UNIT SHOWS WEAK EXPRESSION ONLY AT 28°C IN K-12. | 104 |
| FIGURE 4.40: <i>PdeT</i> (<i>VmpA</i>) DOES NOT DIRECTLY INTERACT WITH <i>YcdT</i> | 106 |
| FIGURE 4.41: DGC ASSAY SHOWS NO C-DI-GMP SYNTHESIS BY <i>DgcX</i> OR <i>YneF</i> | 107 |
| FIGURE 4.42: <i>DgcX</i> UNABLE TO COMPLEMENT $\Delta YEGE$ | 110 |
| FIGURE 4.43: MACROCOLONIES ON CR PLATES AT 28°C & 37°C AFTER 6 DAYS. | 113 |
| FIGURE 4.44: MACROCOLONIES ON A) LB NO SALT AT 28°C AND B) CR PLATES AT 28°C & 37°C AFTER 7 DAYS. | 115 |
| FIGURE 4.45: CELLULAR <i>CsgD</i> LEVELS. | 116 |
| FIGURE 5.1: THE POTENTIAL NEW START CODON FOR THE TRUNCATED VERSION OF <i>YCGG</i> | 121 |

Tables

| | |
|---|----|
| TABLE 3.1: SUPPLIERS OF CHEMICALS USED IN THIS WORK. | 27 |
| TABLE 3.2: SUPPLIERS OF TECHNICAL EQUIPMENT USED IN THIS WORK. | 28 |
| TABLE 3.3: ADDITIVES AND THEIR FINAL CONCENTRATIONS IN MEDIA. | 30 |
| TABLE 3.4: BACTERIAL STRAINS USED IN THIS WORK. | 30 |
| TABLE 3.5: BACTERIOPHAGES USED IN THIS WORK. | 32 |
| TABLE 3.6: PLASMIDS USED IN THIS WORK. | 33 |
| TABLE 3.7: OLIGONUCLEOTIDE PRIMERS USED IN THIS WORK. | 34 |
| TABLE 3.8: PROGRAMS AND DATABASES USED FOR THE BIOINFORMATIC ANALYSIS | 48 |
| TABLE 4.1: THE 35 GGDEF/EAL DOMAIN PROTEINS FOUND WITHIN 61 STRAINS OF <i>E. COLI</i> | 54 |
| TABLE 4.2: 61 <i>E. COLI</i> STRAINS ANALYZED IN THIS WORK AND THEIR COMPLEMENTS OF DGC, PDE AND BIOFILM MATRIX COMPONENT PROTEINS. | 55 |
| TABLE 4.3: THE NOVEL GGDEF DOMAIN-CONTAINING PROTEINS AND THEIR NATIVE IN STRAIN DESIGNATION. | 63 |
| TABLE 4.4: THE NOVEL EAL DOMAIN-CONTAINING PROTEINS AND THEIR NATIVE IN STRAIN DESIGNATION. | 77 |

Abbreviations

| | |
|--------------------|---|
| A | adenine |
| AAF | aggregative adherence fimbriae |
| ABU | asymptomatic bacteriuria <i>E. coli</i> |
| AIEC | adherent and invasive <i>E. coli</i> |
| AMP | adenosine monophosphate |
| AmpR | ampicillin resistance |
| APEC | avian pathogenic <i>E. coli</i> |
| APS | ammonium persulfate |
| A-site | active site |
| ATP | adenosine triphosphate |
| β -gal. act. | β -galactosidase activity |
| BCIP | 5-bromo-4-chloro-3-indolyl phosphate |
| C | cytosine |
| cAMP | cyclic adenosine monophosphate |
| cat | chloramphenicol resistance gene |
| c-di-GMP | cyclic di-guanosine monophosphate |
| CmR | chloramphenicol resistance |
| C-terminal | carboxy-terminal |
| DAEC | diffusely adherent <i>E. coli</i> |
| DGC | diguanylate cyclase |
| DMF | dimethylformamide |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| dNTP | deoxynucleoside triphosphate |
| DTT | 1,4,-dithiothreitol |
| EAEC | enteroaggregative <i>E. coli</i> |
| EDTA | ethylenediaminetetraacetate |
| EHEC | enterohemorrhagic <i>E. coli</i> |
| EIEC | enteroinvasive <i>E. coli</i> |
| EPEC | enteropathogenic <i>E. coli</i> |
| ETEC | enterotoxigenic <i>E. coli</i> |
| E σ | RNA polymerase holoenzyme |
| ExPEC | extraintestinal pathogenic <i>E. coli</i> |
| Fig. | figure |
| G | guanine |
| GDP | guanosine diphosphate |
| GMP | guanosine monophosphate |
| GTP | guanosine triphosphate |

| | |
|--------------|--|
| HUS | hemolytic uremic syndrome |
| IPTG | isopropyl- β -D-1-thiogalactopyranoside |
| I-site | inhibitory site |
| kan | kanamycin resistance gene |
| KanR | kanamycin resistance |
| LB | Luria-Bertani |
| MNEC | meningitis-associated <i>E. coli</i> |
| mRNA | messenger RNA |
| NBT | nitroblue tetrazolium chloride |
| NTD | amino-terminal domain |
| N-terminal | amino-terminal |
| OD | optical density |
| ONPG | ortho-nitrophenyl- β -D-galactopyranoside |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| PDE | (c-di-GMP-specific) phosphodiesterase |
| poly[d(I-C)] | poly-deoxy-inosinic-deoxy-cytidylic acid |
| PPi | pyrophosphate |
| (p)ppGpp | guanosine penta- and tetraphosphate |
| RNA | ribonucleic acid |
| RNAP | RNA polymerase |
| SDS | sodium dodecyl sulfate |
| spec. | specific |
| STEC | shiga toxin producing <i>E. coli</i> |
| Stx | Shiga toxin |
| T | thymine |
| TEMED | N,N,N',N'-Tetramethylethylenediamine |
| TMS | transmembrane segment |
| TetR | tetracycline resistance |
| Tn | transposon |
| Tris | trishydroxyaminomethane |
| U | uracil |
| UPEC | uropathogenic <i>E. coli</i> |
| UTI | urinary tract infections |
| wt | wild-type |
| X-Gal | 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside |

Zusammenfassung

Zyklisches di-Guanosinemonophosphat (c-di-GMP) ist ein sekundärer Botenstoff, der im gesamten Bakterienreich anzutreffen ist. Er ist die treibende Kraft hinter der Regulation verschiedener Prozesse, einschließlich Biofilmbildung, bakterieller Adhärenz, Zell-Zell-Kommunikation, Differenzierung, Motilität und Virulenz. Der zelluläre c-di-GMP Spiegel wird durch zwei Sätze von Proteinen moduliert: Diguanylate Cyclasen (DGCs) synthetisieren und Phosphodiesterasen (PDEs) degradieren c-di-GMP. DGCs werden durch die Anwesenheit der GGDEF- und PDEs durch die der EAL-Domäne identifiziert. In den K-12 Stämmen von *Escherichia coli* W3110 und MG1655 wurden 29 verschiedene Gene für GGDEF/EAL-Domänen enthaltende Proteine annotiert, allerdings ist nur wenig darüber bekannt wie sich diese Liste zwischen pathogenen und kommensalen *E. coli* Stämmen unterscheidet. Diese Studie präsentiert einen systematischen Vergleich zwischen den prognostizierten Genen in insgesamt 61 Stämmen von *E. Coli*, einschließlich enterohämorrhagischen (EHEC), uropathogenen (UPEC), enteroaggregative (EAG), enteropathogenen (EPEC) und enterotoxigene (ETEC) sowie kommensalen Stämmen, die für DGC und PDE Proteine sowie für Proteine codieren, die mit der Biofilmbildung assoziiert werden. Es wurde gezeigt, dass einige Gruppen von pathogenen *E. coli* potentiell neuartige DGC und PDE-Gene besitzen, während andere DGCs und PDEs tatsächlich fehlen. So wurden sechs neue mutmaßliche Gene entdeckt, vier für PDEs (PdeT, PdeX, PdeY und PdeZ) und zwei für DGCs (DgcX und DgcY). Dagegen enthalten eine große Anzahl an Stämmen, einschließlich der *E. coli* K-12 Stämme, eine lange 5' Deletion im DGC Gen *yneF*, was impliziert, dass dieses Gen in diesen Stämmen nicht exprimiert wird. Die allgemeine „Grundausstattung“ umfasst 8 Gene, die für GGDEF- und/oder EAL-Domänenproteine codieren und unter allen hier untersuchten 61 *E. coli* Stämmen konserviert sind. Desweiteren wurden jeweils zwei Varianten der PDE Gene *yahA* und *ycgG* gefunden. Bei ersterem gibt es aufgrund einer stromaufwärts vorliegenden Insertion eines Gens für ein AidA-I-Adhäsion-ähnliches Protein eine Veränderung der 5'-Sequenz, welche für das LuxR-ähnliche N-terminale DNA-Bindemotif in YahA codiert. Im Falle von *ycgG* findet man eine Vollversion inklusive der für die Transmembrandomäne mit CSS-Motif codierenden Sequenzen sowie eine verkürzte Version, die lediglich für die cytoplasmatische EAL Domäne codiert. Die *csg* Gene sind fast universell in allen für diese Studie herangezogenen Stämmen konserviert und codieren für Proteine, die an der Produktion und Export der amyloiden Curlifasern beteiligt sind, darunter auch der Biofilmregulator CsgD, der ebenfalls für die Produktion der Biofilmmatrixkomponente Cellulose essentiell ist.

Im Jahr 2011 wurden fast 4000 Personen in Deutschland von einem Shiga-Toxin (Stx) produzierenden *Escherichia coli* O104:H4 infiziert, wobei mehr als 20% der Patienten hämolytisch-urämisches Syndrom (HUS) entwickelten. Der Ausbruchsstamm ist einem EAEC

genetisch am ähnlichsten, hat aber einen das Stx Gen tragenden Phagen aus EHEC erworben (Mellmann *et al.*, 2011). Er enthält außerdem das für die neuartige Diguanylatcyclase DgcX codierende Gen, welches in allen sechs untersuchten EAEC O104:H4 konserviert sowie in zwei ETECs, E24377A und ETEC H10407, und dem kommensalen Stamm SE11 zu finden ist. Aufgrund dieses Ausbruchs wurde das DgcX Protein zur weiteren Charakterisierung der Regulation und Funktion ausgewählt sowie der Fokus auf den Ausbruchsstamm und seine besonderen Eigenschaften bezüglich DGC/PDE-Gene und mit Biofilm assoziierten Genen gesetzt. DgcX wurde als die am höchsten exprimierte DGC von allen anderen bisher in *E. coli* untersuchten identifiziert. Es wird sowohl bei 28°C als auch 37°C exprimiert, und dies während des gesamten Wachstumszyklus in *E. coli*. Der Ausbruchsstamm produziert besonders starke Biofilme (Al Safadi *et al.*, 2012) und diese Arbeit zeigt, dass er mit zwei neuartigen DGCs ausgestattet, jedoch Cellulose-negativ ist. Der Ausbruchsstamm exprimiert den Biofilmregulator CsgD und amyloide Curlifasern bei 37°C. Die hohe Inzidenz von HUS und Adhärenz des Ausbruchstammes könnte auf seine hohen c-di-GMP- und starken Curlifasernsynthese bei gleichzeitiger Defizienz in der Zelluloseproduktion zurückzuführen sein. Curli Fasern bewirken eine starke proinflammatorische Reaktion (Tükel *et al.*, 2005, 2009), während Zellulose dieser entgegenwirkt. So ist die starke proinflammatorische Reaktion ausgelöst durch eine Infektion mit dem Ausbruchsstamm vermutlich das Ergebnis der hohen Curli Produktion ohne Zellulose und könnte die systemische Absorption des Shiga-Toxins vom Körper und den Transport durch den Blutstrom zu den Nieren ermöglichen, was schließlich zu hämorrhagischen Durchfall und HUS führt. Diese Studie trägt zu der Aufklärung des komplexen Regulationsnetzwerks von c-di-GMP bei und schafft einen Einblick, welche DGCs und PDEs für c-di-GMP Regulation in *E. coli* unverzichtbar sein könnten. Es lenkt die Aufmerksamkeit auf den Ausbruchsstamm und sein einzigartiges Ensemble von Eigenschaften, die vermutlich in Zusammenhang mit seiner erhöhten Virulenz stehen.

Summary

Cyclic-di-GMP is a second messenger molecule found ubiquitously throughout the bacterial kingdom. It is the driving force behind the regulation of processes including biofilm formation, bacterial adherence, cell-cell signaling, differentiation, motility, and virulence. Its cellular level is modulated by two sets of proteins: diguanylate cyclases (DGCs) synthesize and phosphodiesterases (PDEs) degrade c-di-GMP. DGCs are characterized by the presence of the GGDEF domain and PDEs are identified by the EAL domain. In the K-12 strains *Escherichia coli* W3110 and MG1655, 29 different GGDEF/EAL domain-containing proteins have been identified. However, little is known about how that list may differ between strains of *E. coli*, particularly comparing pathogenic and commensal strains. This study presents a systematic comparison of the complement of putative DGC and PDE proteins as well as genes associated with biofilm formation, among 61 strains of *E. coli* including enterohemorrhagic (EHEC), uropathogenic (UPEC), enteroaggregative (EAEC), enteropathogenic (EPEC) and enterotoxigenic (ETEC) as well as commensal strains. It has been found that some groups of pathogenic *E. coli* possess potentially novel DGC and PDE genes, whereas other common DGCs and PDEs are absent. 8 GGDEF/EAL domain-containing genes were found to be universally conserved among the 61 *E. coli* strains analyzed. Four novel putative PDEs (PdeT, PdeX, PdeY and PdeZ) and two novel DGCs (DgcX and DgcY) were discovered. A large number of strains, including *E. coli* K-12 strains, contain a large 5' deletion in the DGC gene *yneF*, implying that this gene is not expressed in these strains. Two variants of the PDE gene *yahA* have been detected, one of which contains an upstream insertion of an *aidA-I adhesin-like* gene resulting in the alteration of the 5' end encoding the LuxR-like N-terminal domain in the DNA-binding motif. The *ycgG* gene was also found in two variants, one full version encoding its transmembrane domain containing a CSS motif and a shortened version encoding only the EAL domain uncoupled from the transmembrane sensory domain. *csg* genes were almost universally conserved among the analyzed strains and encode proteins for amyloid curli fiber production and export and the biofilm regulator CsgD, essential for the production of two biofilm matrix components: curli fibers and cellulose. The novel *dgcX* gene has been found in a total of nine strains and was conserved among all EAECs of the O104:H4 serotype: 55989, HUSEC041, LB226692, 2011C-3493, 2009EL-2050 and 2009EL-2071. The three additional strains containing the *dgcX* gene were two ETECs: E24377A and ETEC H10407, and one commensal strain SE11.

In 2011 nearly 4000 persons in Germany were infected by a Shiga toxin (Stx)-producing *Escherichia coli* O104:H4, with more than 20% of patients developing hemolytic uremic syndrome (HUS). The outbreak strain is genetically most similar to an EAEC but has acquired a Stx carrying phage from EHEC (Mellmann *et al.*, 2011) and contains the novel *dgcX* gene. This outbreak led to the DgcX protein being chosen for further characterization of regulation and

function as well as focus on the outbreak strain and its special characteristics with respect to DGC/PDE genes and genes associated with biofilm formation. DgcX has been found to be the most highly expressed DGC of all of the others studied so far in *E. coli*. It is expressed at both 28°C and 37°C and throughout the *E. coli*'s growth cycle. The outbreak strain was shown to produce thick biofilms (Al Safadi *et al.*, 2012) and this work shows that it contains two novel DGCs (one of them highly active and atypical of many *E. coli*), expressing the biofilm regulator CsgD and amyloid curli fibers at 37°C but is cellulose-negative. The outbreak strains high incidence of HUS and adherence may be due to its high production of c-di-GMP and curli fiber, while at the same time not being able to produce cellulose. Curli fibers cause a strong proinflammatory response (Tükel *et al.*, 2005, 2009) while cellulose has been shown to counteract this. Thus the strong proinflammatory response triggered by an infection by the outbreak strain may be the result of high curli production without cellulose and may facilitate the systemic absorption of the shiga-toxin by the body and transport through the bloodstream to the kidneys, which can lead to hemorrhagic diarrhea and HUS. This study will contribute to elucidating the complex regulatory network of c-di-GMP as well as shed light upon which DGCs and PDEs may be indispensable for c-di-GMP regulation in *E. coli* and which may be linked to virulence. It brings attention to the outbreak strain and its unique ensemble of properties that may be linked to its increased virulence.

1 Introduction

1.1 Commensal and pathogenic *E. coli*

Escherichia coli is a stunningly versatile microorganism that typically inhabits the mucosal layer of the gastrointestinal tract of mammals, where it is the most abundant facultative anaerobe (Sweeney *et al.*, 1996; Kaper *et al.*, 2004). This process begins within a few hours of birth, where the mother's intestinal microflora is passed onto the infant during the birthing process (Bettelheim *et al.*, 1974). For the most part, commensal *E. coli* can co-exist with its host in a symbiotic relationship, unless the host is immunocompromised or the *E. coli* manages to travel to another part of the body or there is a breach in the gastrointestinal lining (Kaper *et al.*, 2004). Commensal *E. coli* play an important role as part of the intestinal microbiota, helping in digestion, the production of vitamin K and by simply taking up space so that other enteric pathogens have no place to settle (Schierack *et al.*, 2009). The bacterium also plays a pivotal role in modern biology via its use as a cloning host in recombinant DNA technology. But *E. coli* also has a darker side, because of some strains acquisitions of virulence factors that have transformed it from a benign inhabitant to harbinger of disease and even death. These pathogenic *E. coli* are thought to have emerged from commensal *E. coli* as the result of horizontal gene transfer of specific virulence factors encoded in the genetic elements from other bacterial pathogens to commensal *E. coli* strains (Kaper *et al.*, 2004). *E. coli* can pick up these genetic elements from a number of different bacteria and the result may become a completely novel combination, the most successful of which persisting to the next generation and potentially becoming locked into the *E. coli* genome. These different arrangements of genetic elements are what separates *E. coli* strains from each other and are referred to as the pathotype (Kaper *et al.*, 2004).

In contrast to commensal strains, pathogenic *E. coli* are capable of causing disease in healthy individuals. The symptoms associated with these infections are generally divided into three categories: enteric/diarrheal disease, urinary tract infections (UTIs) and sepsis/meningitis (Kaper *et al.*, 2004). UTI is caused predominately by uropathogenic *E. coli* (UPEC) and is the most common of the extraintestinal infections, occurring in the urinary tract (Foxman, 2002). Meningitis and sepsis constitute another major class of extraintestinal infections and are caused by meningitis-associated *E. coli* (MNEC) (Dawson *et al.*, 1999). *E. coli* pathotypes associated with extraintestinal infections are termed as ExPECs (Russo *et al.*, 2000) and according to this broad definition both UPECs and MNECs would also fall into this category. The avian pathogenic *E. coli* (APEC) causes extraintestinal infections in poultry, specifically respiratory infections (Johnson & Russo, 2002; Kaper *et al.*, 2004). The major intestinal pathogens can be further divided into six groups: enterohaemorrhagic *E. coli* (EHEC), enteroaggregative *E. coli* (EAEC), enteropathogenic *E. coli* (EPEC) (Bray, 1945; Kaper, 1996), enterotoxigenic *E. coli* (ETEC),

Introduction

enteroinvasive *E. coli* (EIEC) (Kaper *et al.*, 2004), diffusely adherent *E. coli* (DAEC) (Nataro *et al.*, 1998; Scaletsky *et al.*, 2002) and adherent and invasive *E. coli* (AIEC) (Nash *et al.*, 2010).

EHECs have been of especial interest due to their association with many foodborne outbreaks, most recently the May 2011 outbreak in Germany, and their ability to cause hemorrhagic diarrhea or hemolytic uremic syndrome (HUS). Based on genome sequence the recent May 2011 outbreak strain (LB226692) is an EAEC derivative, but because it has acquired virulence factors and the shiga toxin from classical EHEC strains it has the ability to also cause HUS (Mellmann *et al.*, 2011; Frank *et al.*, 2011), leading to it being categorized as an EHEC for a time (Mellmann *et al.*, 2011). EHECs can technically also be classified as STECs for shiga toxin-producing *E. coli*, but not all STECs can be classified as EHECs because unlike EHECs, STECs don't have to contain the Locus of Enterocyte Effacement (LEE) (McDaniel *et al.*, 1995) pathogenicity island (Kaper *et al.*, 2004). Because it lacks the LEE pathogenicity island, the STEC categorization is more appropriate than EHEC. However, because of the 2011 outbreak strains high genomic similarity to EAECs, in this study I will be referring to it as an EAEC. A schematic of how the various *E. coli* strains are organized relative to their pathotypes can be seen in figure 1.1.

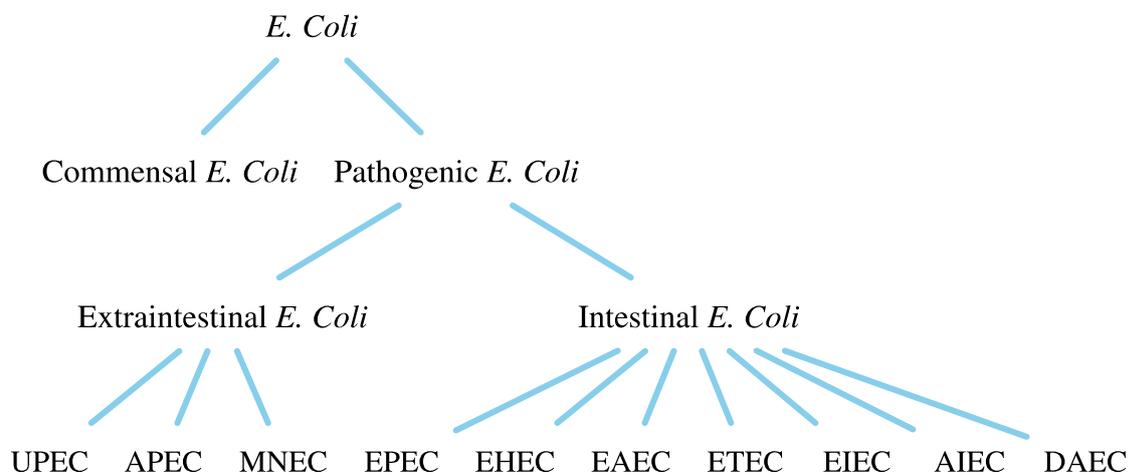


Figure 1.1: Schematic representation of *Escherichia coli* groups and pathotypes. UPEC, uropathogenic *E. coli*; APEC, avian pathogenic *E. coli*; MNEC, meningitis-associated *E. coli*; EPEC, enteropathogenic *E. coli*; EHEC, enterohaemorrhagic *E. coli*; EAEC, enteroaggregative *E. coli*; ETEC, enterotoxigenic *E. coli*; EIEC, enteroinvasive *E. coli*; AIEC, adherent and invasive *E. coli*; DAEC, diffusely adherent *E. coli*.

Pathogenic *E. coli* strains are named based on their serotypes which are characterized by shared lipopolysaccharides (LPS), designated as 'O' in the strain's name, and flagellar antigens, designated as 'H' in the strains name. Sometimes they are referred to by their serogroups which are characterized by the 'O' antigen alone (Whittam *et al.*, 1996; Nataro *et al.*, 1998). *E. coli* strains can also be classified using the Clermont *E. coli* phylo-typing method where the strains are sorted into eight groups (A, B1, B2, C, D, E, F and I) using the method of multiplex PCR,

Introduction

detecting specifically chosen genes and the presence or absence of these genes would then enable the *E. coli* strain to be categorized phylogenetically (Clermont *et al.*, 2012). Most commensal strains of *E. coli* fall into the phylogenetic group A or B1 as they are typically lacking virulence factors of pathogenic strains (Picard *et al.*, 1999). ExPECs have been shown to fall mostly in the D and B2 groups (Smith *et al.*, 2007), ETECs fall in the B1 and A groups (Sahl *et al.*, 2011), while EPECs, EAECs and EHECs fall primarily in groups A, B1, D and E (Smith *et al.*, 2007).

In general, the mechanism of pathogenesis involves the colonization of the host's mucosal site, evasion of the host immune system, multiplication and eventual host damage (Kaper *et al.*, 2004). The vast majority of *E. coli* strains are extracellular in their colonization process, with the notable exception of EIEC which actually invades the cell and then replicates from within the epithelial cell or macrophage (Darfeuille-Michaud, 2002). Different groups of *E. coli* employ different methods of colonization and the majority are able to do this by producing specific adherence factors. Often these adherence factors are in the form of either fimbriae/pili or fibrillae. These structures are distinct from flagella; fimbriae are rod-like ranging from 5-10nm in diameter, while fibrillae are either curly and flexible or long and wiry ranging from 2-4nm in diameter (Cassels *et al.*, 1995).

EAECs are an example of *E. coli* that utilize fimbriae in their colonization strategy. They produce thick biofilms using aggregative adherence fimbriae (AAF) to form a stacked-brick configuration in the epithelia of the intestinal mucosa (Nataro *et al.*, 1992, 1994; Czczulin *et al.*, 1997). Other adhesion strategies include the use of outer-membrane proteins such as Intimin produced by EHECs and UPECs (Kaper *et al.*, 2004). For example, EHECs attach to the colon's enterocytes by inducing cytoskeletal rearrangement via the protein Intimin, thereby destroying the native microvillar architecture and causing the formation of pedestal-like structures to which the bacteria then attaches to (also known as attaching and effacing lesion; the genes responsible for this process are encoded on the LEE pathogenicity island) (Donnenberg *et al.*, 1993; Kaper *et al.*, 2004; Croxen & Finlay, 2010). Like EPECs, EHECs induce the attaching and effacing lesion but instead adhere to the small bowel enterocytes (Elliott *et al.*, 1998; Kaper *et al.*, 2004). Unlike EPECs, EHECs also produce shiga toxin (also known as verocytotoxin), the systemic absorption of which can lead to potentially life threatening complications (Karmali *et al.*, 1983; Kaper 1996). The toxin is produced in the colon, where it causes apoptosis of intestinal epithelial cells and is then transported through the bloodstream to the kidneys, where it wreaks havoc on the endothelial cells, obstructs the microvasculature and leads to renal inflammation which can ultimately result in HUS (Andreoli *et al.*, 2002, Kaper *et al.*, 2004). EHECs are commonly found in the mammalian digestive tract of cattle where it is actually considered a commensal strain because ruminants lack vascular Gb3 receptors for the Shiga toxin to bind to (Pruimboom-Brees *et al.*,

Introduction

2000). Since the toxin cannot bind to receptors in cattle, it cannot be endocytosed and transported to other parts of the body to cause vascular damage.

ETECs adhere to the small bowel mucosa using various fibrillar colonization factors (CFs) (Gaastra & Svennerholm, 1996) or proteinaceous fimbrial, where they express two groups of enterotoxins, heat-labile and heat-stable (So *et al.*, 1976, 1978), which lead to intestinal secretion and watery diarrhea (Wolf *et al.*, 1997; Kaper *et al.*, 2004). AIECs are associated with Crohn's disease and are found in the ileal lesions of Crohn's disease patients, where they adhere to the epithelial cells found in the intestine (Darfeuille-Michaud, 2002). AIECs are capable of invading and surviving within the epithelial cells and macrophages (Nash *et al.*, 2010). DAECs colonize the small bowel enterocytes and elicit a signal transduction cascade that leads to the growth of long tentacle-like cellular projections that then wrap around the bacteria (Bernet-Carnard *et al.*, 1996; Kaper *et al.*, 2004). A schematic of colonization strategies of intestinal *E. coli* infections is provided in figure 1.2.

Introduction

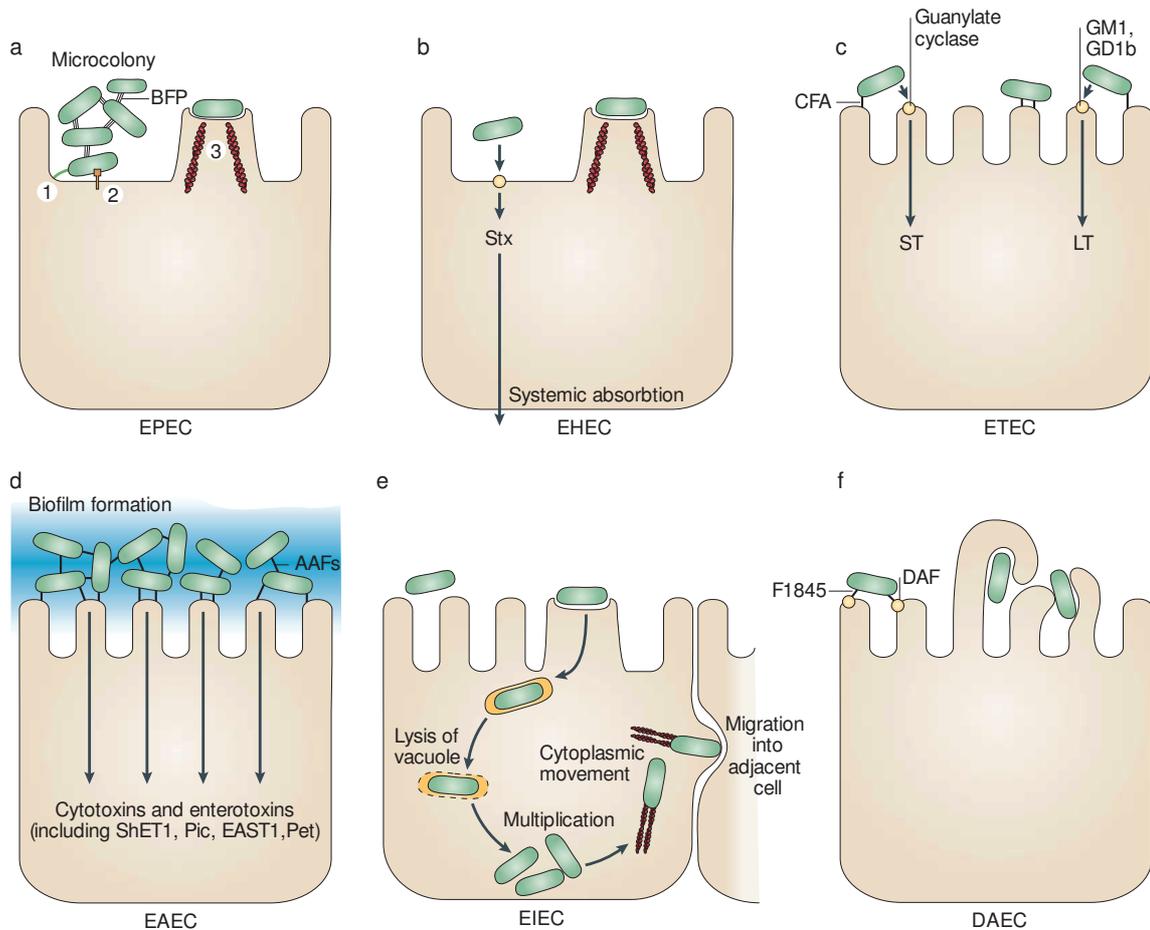


Figure 1.2: Pathogenic schema of diarrheagenic *E. coli*. Here the six recognized categories of diarrheagenic *E. coli* are schematically depicted in their colonization process of eukaryotic cells. These descriptions are the result of mostly *in vitro* studies and may not comprehensively reflect *in vivo* processes. a) EPECs induce the attaching and effacing lesion and adhere to the small bowel enterocytes. Diarrhea and the inflammatory response accompanies the cytoskeletal derangements. 1. Initial adhesion using bundle-forming pili (BFP), 2. Protein translocation by type III secretion, 3. Pedestal formation. b) EHECs induce the attaching and effacing lesion and adhere to the colon's enterocytes. EHECs produce shiga toxin (Stx) and its systemic absorption by the body and transport through the bloodstream and to the kidneys can lead to hemorrhagic diarrhea or hemolytic uremic syndrome (HUS) and death. c) ETECs adhere to the small bowel mucosa using various fibrillar colonization factors (CFs) or proteinaceous fimbriae, where they express two groups of enterotoxins [heat-labile (LT) and heat-stable (ST)] which lead to intestinal secretion and watery diarrhea. d) EAECs produce thick biofilms using aggregative adherence fimbriae (AAF) to form a stacked-brick configuration in the epithelia of the intestinal mucosa and secrete a variety of enterotoxins and cytotoxins including Shigella enterotoxin 1 (ShET1), Pic, Pet and enteroaggregative *E. coli* ST1 (EAST1). e) EIECs invade colonic epithelial cells and then replicate from within the epithelial cell. They are capable of moving laterally through the epithelium by directly spreading from cell to cell or they can exit and re-enter the baso-lateral plasma membrane. f) DAECs colonize the small bowel enterocytes and elicit a signal transduction cascade via release of decay-accelerating factors (DAFs) that leads to the growth of long tentacle-like cellular projections that then wrap around the bacteria. This figure is a reproduction from Kaper *et al.*, (2004).

UPECs use a variety of factors usually encoded in UPEC-specific pathogenicity islands to colonize and ascend the urinary tract to the bladder and potentially the kidneys. In the bladder UPECs use type 1 fimbriae to attach to the mannose moieties of the uroplakin receptors that coat the transitional epithelial cells. They then invade the epithelial cells and form pod-like bulging biofilms of bacteria that are shrouded by a matrix of exopolysaccharide components, most notably

poly-GlcNAc (PGA) (Mulvey *et al.*, 1998; Kaper *et al.*, 2004). These biofilms may then serve as a reservoir for recurrent infection and help UPECs evade the host immune system and resist antibiotic treatments (Mulvey *et al.*, 1998; Anderson *et al.*, 2003; Kaper *et al.*, 2004). MNECs infect the central nervous system and form K1 capsules to protect the bacteria. The capsules contain sialic acid that not only help the bacteria survive the penetration of the blood-brain-barrier but also helps evade the human innate immune system as sialic acid is widely used by the human body and thus does not trigger an immune response (Hoffman *et al.*, 1999).

Pathogenic *E. coli* in contrast to commensal *E. coli* have acquired specific virulence attributes that have enabled them to adapt to new niches, outcompete other bacteria, cause disease and be transmitted from host to host. Many have acquired specific pathogenicity islands and plasmids that enable them to colonize, evade the host immune system, multiply and eventually cause damage and disease. It is interesting to note that among all of the various *E. coli* strains the core shared genome by all strains is about 20%, thus making *E. coli* one of the most diverse groups of bacteria (Rasko *et al.*, 2008).

1.2 Bacterial environmental adaptation and biofilm formation

Bacteria in a given environment are subject to a bombardment of conditions, including temperature swings, nutrient limitation, host immune system, to shifts in osmolarity. In order to ensure survival, bacteria have evolved various mechanisms to respond to the environmental stresses. Chief among them is the ability to switch ‘life styles’ from a planktonic motile single-cellular state to an adhesive sedentary multicellular state known as a biofilm (Ross *et al.*, 1991; Simm *et al.*, 2004; Kader *et al.*, 2006; Weber *et al.*, 2006; Hengge 2009). By forming biofilms, some bacteria have been able to evade the host immune system and resist the effects of antibiotics (Stewart & Costerton 2001; Hoiby *et al.*, 2002). These advantages come at a price of a limited nutrient supply to all of the cells as well as increased stress from the concentrated exposure to fermentation-generated acid that the cells are forced to produce because of their dense population limiting their access to oxygen (Hengge 2011).

In a colony there is no smooth transition between the growth phases and bacteria in different growth phases can coexist within the same colony in different regions of that colony (Serra *et al.*, 2013a; 2013b). The transition between growth phases can most acutely be seen in batch cultures where the majority of the bacterial cells are homogeneous with respect to their physiological state. Under these conditions, the highly motile state is linked to the post-exponential growth phase [starting at approximately 1.5 optical density at wavelength 578nm (OD₅₇₈) in LB]. In this phase nutrients, though diminishing, are still readily available but not equally distributed as in the exponential growth phase (Adler & Templeton 1967; Amsler *et al.*, 1993; Arnqvist *et al.*, 1994; Weber *et al.*, 2006; Zhao *et al.*, 2007; Hengge 2011). This heterogeneity of nutrient distribution

forces the *E. coli* to ‘forage’ for food by expressing flagella and undergoing chemotaxis. If the resources decrease further, *E. coli* enters the stationary phase (starting at OD₅₇₈ of approximately 3 in LB) in which motility is down-regulated and adhesins start to be formed such as curli fibers as well as exopolysaccharides like PGA, cellulose and colonic acid (Adler & Templeton 1967; Amsler *et al.*, 1993; Arnvist *et al.*, 1994; Weber *et al.*, 2006; Zhao *et al.*, 2007; Beloin *et al.* 2008; Hengge 2011).

In *E. coli*, biofilm formation is intimately entwined with the general stress response regulator RpoS (σ^S), which controls over 500 genes and enables the transition into the stationary phase (Weber *et al.*, 2005; Hengge 2011). It counteracts the deleterious stress effects that triggered the transition and bolsters general resistance to stresses not yet encountered. σ^S is able to usher in the general stress response because it is able to outcompete σ^{70} (the housekeeping sigma factor) with the help of accessory proteins when recruiting RNA polymerase (RNAP) holoenzyme, even though there is always a higher σ^{70} concentration in the cell (Grigorova *et al.*, 2006) and σ^{70} has a higher holoenzyme affinity (Maeda *et al.*, 2000). Through this method of the RNAP utilizing alternative sigma factors, *E. coli* and bacteria in general are able to reprogram gene expression and achieve sweeping changes in lifestyle, morphology or differentiation events (like sporulation in sporulating bacteria) (Weber *et al.*, 2005). One of the main factors in the transition between ‘lifestyles’ in bacteria is the level of the second messenger molecule bis-(3’-5’)-cyclic dimeric guanosine monophosphate (c-di-GMP) (Henge 2009).

1.3 C-di-GMP as a ubiquitous second messenger in bacteria

1.3.1 Biosynthesis and degradation of c-di-GMP by GGDEF, EAL and HD-GYP domain-containing proteins

For a long time the importance of c-di-GMP, a second messenger molecule found ubiquitously in bacteria, has been completely underestimated. The molecule has been obscured by other intercellular signaling molecules like cyclic AMP and was not discovered until 1987 by Benziman and co-workers (1987). Yet it remained in the periphery of the field until the genomics era revealed the startling infiltration of this molecule throughout the bacterial kingdom. The second messenger functions in a gamut of molecular processes from transcription to post-transcriptional control, as well as the regulation of enzyme activity and large cellular structures (Hengge 2009). When this regulation is examined on the physiological scale, it amounts to biofilm formation, cell-cell signaling, differentiation, motility, and importantly, virulence (Paul *et al.*, 2004; Méndez-Ortiz *et al.*, 2006; Jonas *et al.*, 2008; Kulasakara *et al.*, 2006; Hammer & Bassler, 2009). In *E. coli*, c-di-GMP is required for the transition between a motile-planktonic single-cell state and the biofilm state (Simm *et al.*, 2004; Weber *et al.*, 2006; Hengge, 2011).

Introduction

The level of the second messenger, c-di-GMP, is regulated by two sets of proteins: diguanylate cyclases (DGCs) and phosphodiesterases (PDEs). DGCs synthesize and PDEs degrade c-di-GMP. Often these two proteins form antagonistic partnerships by which the c-di-GMP level is regulated and with the assistance of effectors they act upon targets. Together all of these components form control modules (Cotter & Stibitz, 2007; Hengge 2009).

DGCs are characterized by the presence of the GGDEF domain, named after the domain's active site (A-site) amino acid (AA) sequence 'GGDEF' where the D can also be an E to be functionally active (Tal *et al.*, 1998; Chan *et al.*, 2004; Malone *et al.*, 2007; Ryjenkov *et al.*, 2005). The GGDEF domain may also contain an inhibiting site (I-site) five residues upstream of the A-site, characterized by the RxxD residues (Chan *et al.*, 2004; Christen *et al.*, 2006). The I-site acts as an allosteric inhibitor of DGC activity, thereby preventing excessive GTP consumption and establishing an upper boundary for the cellular concentration of c-di-GMP (Chan *et al.*, 2004; Christen *et al.*, 2006). The GGDEF domain binds GTP and then dimerizes with another GTP-bound GGDEF domain protein, forming two phosphodiester bonds and producing a c-di-GMP molecule (Schirmer & Jenal, 2009).

PDEs are characterized by the EAL or HD-GYP domain, though *E. coli* strains only have the EAL domain (Chang *et al.*, 2001; Christen *et al.*, 2005; Ryan *et al.*, 2006; Schmidt *et al.*, 2005; Tamayo *et al.*, 2005). An active EAL domain has an Mg²⁺- or Mn²⁺-chelating site and requires it for activity (Rao *et al.*, 2008). Ca²⁺ or Zn²⁺ inhibit the activity of the EAL domain (Christen *et al.*, 2005; Schmidt *et al.*, 2005; Tamayo *et al.*, 2005). The EAL domain is thought to be intact and active with the conservation of eight non-contiguous amino acids which make up the Mg²⁺- or Mn²⁺-chelating site (Rao *et al.*, 2008). Its proteins hydrolyze c-di-GMP to linear 5' pGpG and the product is further hydrolyzed into two GMP molecules by other hydrolases (Christen *et al.*, 2005; Schmidt *et al.*, 2005). The HD-GYP domain-containing protein directly hydrolyzes c-di-GMP into two GMP molecules. The HD-GYP domain contains the HHExxDGxxGYP motif and is unrelated to the EAL domain (Dow *et al.*, 2006). The significance of having two unrelated c-di-GMP hydrolyzing domains within the same genome is still remains to be elucidated. An overview of the synthesis and degradation of c-di-GMP may be seen in figure 1.3.

There is also a subset of proteins that may contain either a GGDEF and/or an EAL domain but where the active site's amino acid sequence is not conserved and is thus enzymatically inactive. These proteins are collectively termed degenerate GGDEF/EAL domain proteins and often act as effectors or targets in the c-di-GMP regulation cascade (Jenal *et al.*, 2006; Hengge, 2009; Povolotsky & Hengge, 2012). A given protein may contain either a GGDEF or an EAL domain, a combination of the two together or one of the domains in combination with another sensory domain located in the N-terminus of the protein (Galperin *et al.*, 2001; Galperin 2004; Hengge 2009). These sensory domains can control the activities of the GGDEF and EAL domains

which are present in the C-terminus of a protein. In *E. coli*, there are some proteins that contain both an EAL and a GGDEF domain; this may seem conflicting at first, but in these cases one of the domain's active site is degenerated and the protein lacks that domain's enzymatic activity.

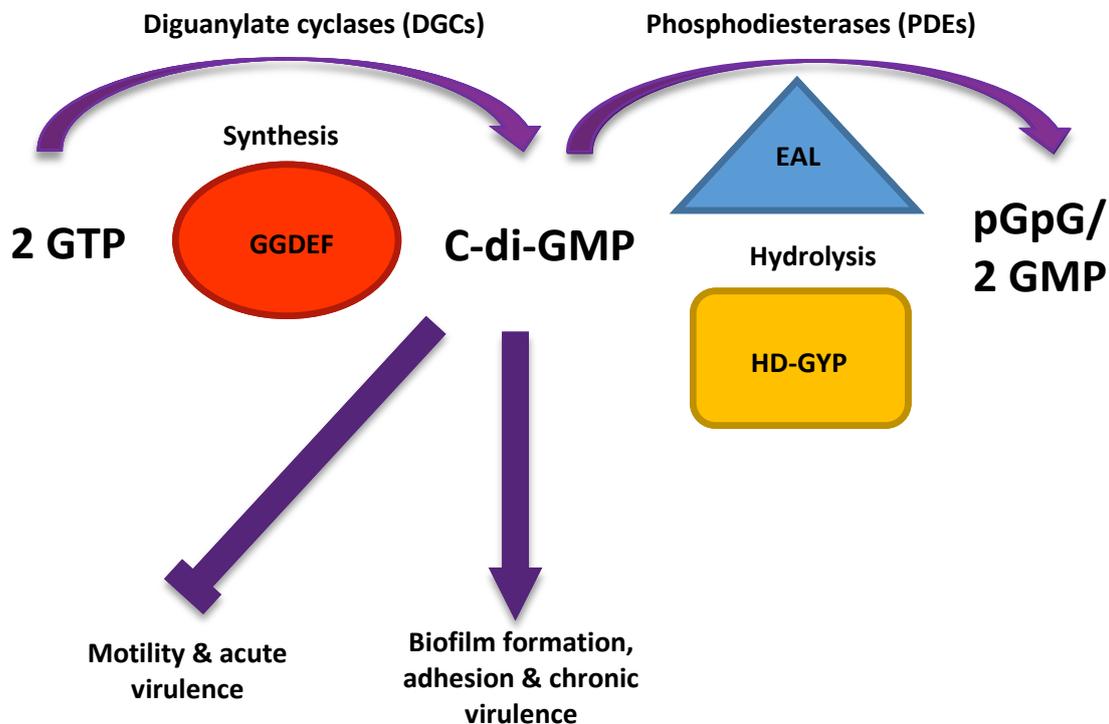


Figure 1.3: C-di-GMP turnover. The level of c-di-GMP is regulated by two sets of proteins: diguanylate cyclases (DGCs) and phosphodiesterases (PDEs). DGCs synthesize and are characterized by the GGDEF domain which binds GTP and then dimerizes with another GTP bound GGDEF domain protein, forming two phosphodiester bonds and producing a c-di-GMP molecule. PDEs degrade c-di-GMP and are characterized by the EAL or HD-GYP domain. The EAL domain proteins hydrolyze c-di-GMP to linear 5' pGpG and the product is further hydrolyzed into two GMP molecules by other hydrolases. The HD-GYP domain protein directly hydrolyzes c-di-GMP into two GMP molecules. High cellular levels of c-di-GMP promote biofilm formation, adhesion and chronic virulence, while low cellular levels of c-di-GMP promote motility and acute virulence.

1.3.2 Multiple GGDEF/EAL domain-containing proteins in single organisms

It is interesting to note the large number of GGDEF and EAL domain proteins found within a single genome. In the commensal *E. coli* K-12 W3110 strain genes exist for 29 GGDEF/EAL domain proteins. *Salmonella enterica* serovar Typhimurium (hereafter referred to as simply *Salmonella*) encodes 19 GGDEF/EAL domain proteins. *Pseudomonas aeruginosa* has 41 and *Vibrio cholera* encodes 72 GGDEF/EAL/HD-GYP domain proteins (Galperin *et al.*, 2001). This multiplicity raises the question of the necessity of such a large collection of seemingly redundantly functioning proteins. A possible explanation is that many of the GGDEF/EAL domain proteins are coupled with an accessory sensory input domain in the N-terminus, and the GGDEF/EAL domain proteins present in a single genome have a unique sensory input domain (Galperin *et al.*, 2001; Galperin 2004).

Introduction

The commonly occurring sensory input domains include GAF, PAS, REC, HTH, sensor globin, BLUF, LuxR-like, MASE1, MASE2, CSS, CZB and HAMP, to name a few (Galperin, 2005; this study). Though the exact mechanism of the varying sensory domains for most of the proteins is yet to be elucidated several proteins have been well studied. One example is the Dos protein of *E. coli* which is regulated by sensing oxygen, to which it binds, resulting in the shift of the heme molecule which in turn leads to a conformational change of the PAS domain (Takahashi & Shimizu, 2006; Wan *et al.*, 2009). In *Acetobacter xylinum*, the PDE A1 protein has a heme containing PAS domain and by reversibly binding to O₂ regulates the phosphodiesterase activity of the protein in much the same way as in Dos (Chang *et al.*, 2001; Wan *et al.*, 2009). The GAF domain has been shown to bind small molecules like cGMP, cAMP and various chromophores (Ho *et al.*, 2000). The BLUF domain is a blue light sensor that upon irradiation allows the *E. coli* protein BluF to bind the YcgE repressor which releases the latter from its operator DNA resulting in de-repression of specific target genes (Hasegawa *et al.*, 2006; Tschowri *et al.*, 2009).

Intriguingly there exist proteins with both a GGDEF and an EAL domain present. This raises the question of their function, as they seem to have conflicting effects. For most of these cases it has been demonstrated *in vitro* that these proteins perform only one of the domain-associated activities (Christen *et al.*, 2005; Schmidt *et al.*, 2005; Takahashi & Shimizu, 2006). There exists a GGDEF-EAL compound protein in *Rhodobacter sphaeroides*, that *in vitro* shows PDE activity and upon removal of the EAL domain also shows DGC activity (Tarutina *et al.*, 2006). Whether this translates to *in vivo* bifunctional activity remains to be demonstrated.

Furthermore, there is clear evidence of spatial and temporal regulation of DGCs and EALs, so that specific DGCs and PDEs may only be expressed in response to changing environmental conditions and stresses such as nutrient limitation. Thus the entire set of GGDEF/EAL/HD-GYP encoding genes of a single genome is never expressed and present at the same time, contributing to a single converging pool of C-di-GMP. Instead sequestration occurs of DGCs and EALs and even entire c-di-GMP control modules (Jenal & Malone 2006; Kader *et al.*, 2006; Ryan *et al.*, 2006; Hengge 2009; Christen *et al.*, 2010; Ryan *et al.*, 2010). Thus the activity of DGCs and PDEs is tightly modulated within the cell at the transcriptional level, at the translational level via protein stability, allosterically through phosphorylation of accessory domains and by spatial localizations/sequestration within the cell.

1.3.3 Enzymatically inactive degenerate GGDEF/EAL domain-containing proteins in *E. coli*

Whenever a compound protein emerges, it is conceivable that one of the two domain's functions may mutate and evolve to serve a new function, independent of its origins. In *E. coli* there exist four examples of GGDEF/EAL domain proteins that have lost their DGC/PDE activity and have

Introduction

taken on new roles. These proteins are referred to as ‘degenerate’ GGDEF or EAL domain proteins. The BluF (YcgF) protein is one such example since it contains a BLUF and a degenerate EAL domain (see Table 4.1 for details on which amino acids in EAL domain are mutated). The BLUF domain is a well characterized blue light sensor (Hasegawa *et al.*, 2006) and is located in the N-terminus of BluF. BluF senses blue light, in the presence of which its affinity to complex with the MerR-like repressor protein, YcgE, increases. Upon YcgE-BluF binding, YcgE is de-repressed and becomes unbound from its operator DNA. In this manner gene expression of a regulon (*ycgZ-ymgAB*) comprised of eight small proteins is induced. These proteins then stimulate the activity of the Rcs phosphorelay system which in turn activate colonic acid production, acid resistance genes and *biofilm production linked (bdm)* gene expression and the down regulation of adhesive curli fibers (Tschowri *et al.*, 2009). Furthermore, this pathway is strongly activated at 16°C as opposed to 37°C, leading to the idea that it is an evolved adaptation in *E. coli* to survive outside of the host (Tschowri *et al.*, 2009).

CsrD (YhdA) is a degenerate composite GGDEF-EAL domain protein, but lacks the critical amino acid residues in both of the domains to have either DGC or PDE activity (see Table 4.1). Instead it is involved in indirect post-transcriptional regulation of the flagellar master regulator, FlhDC, PGA inhibition, and CsgD expression. CsrD targets CsrB and CsrC, two small non-coding RNAs for degradation by RNaseE (Suzuki *et al.*, 2006). The exact mechanism of this process is not yet fully worked out, but binding of YhdA to CsrB and CsrC has been confirmed and in $\Delta yhdA$, CsrB and CsrC are both stable. When CsrB and CsrC are degraded, CsrA activity is increased. CsrA is an RNA binding protein that controls carbon metabolism, biofilm formation and motility. Stable CsrB and CsrC bind CsrA and sequester it, thereby inhibiting CsrA function. CsrA binds PGA mRNA and prevents its translation, effectively inhibiting PGA expression. CsrA regulates target mRNAs at the post-transcriptional level of two GGDEF domain proteins, YcdT and YdeH, which in turn control flagella-mediated swimming motility. In *E. coli*, CsrA binding to the mRNAs of YcdT and YdeH lead to a strong down regulation in transcript levels. CsrA stabilizes the mRNA of *flhDC* by binding to the *flhDC* leader which is required for flagellum biosynthesis (Babitzke & Romeo, 2007).

The YdiV protein has a degenerate EAL domain and is an anti-FlhD(4)C(2) factor that is best expressed in nutrient poor conditions in *Salmonella* (Wada *et al.*, 2011). Furthermore, it has been shown to inhibit expression of class 2 and class 3 genes through its interaction with FlhDC, thereby preventing FlhDC from interacting with DNA (Chilcott & Hughes, 2000; Chevance & Hughes, 2008; Wada *et al.*, 2011). YdiV directly binds FlhDC and delivers it to the protease ClpXP for proteolytic degradation (Takaya *et al.*, 2012). In *Salmonella*, the work of Simm *et al.*, (2009) shows that in $\Delta ydiV$, also known as STM1344 in *Salmonella*, CsgD is down regulated compared to the wild-type strain. CsgD is the major regulator of the expression of the ‘rdar’ (red,

dry and rough) morphotype (Römling *et al.*, 1998b) which is also down regulated in $\Delta ydiV$, as well as cellulose and curli fibers which are components of the extracellular matrix. $\Delta ydiV$ mutants show slightly more motility, measured in swimming and swarming ability, as compared to the wild-type strain. When a $\Delta ydiV$ is complemented with a *ydiV* containing vector and an empty vector, only the complementation with the *ydiV*-vector, leads to a slight rdar morphotype enhancement and a motility down regulation at 28°C (Simm *et al.*, 2009). Studies have shown that YdiV is also capable of inactivating FlhDC in *E. coli*, but that *in vivo* the protein is not expressed at high enough concentrations to inhibit flagellar synthesis, as the gene is transcribed but not translated (Wada *et al.*, 2012).

YeaI is a GGDEF domain protein but with a degenerate A-site and an intact I-site. As of yet the function of this protein remains a unknown but it might serve as a c-di-GMP effector protein. As mentioned earlier, degenerate GGDEF/EAL domains may also be used as sensory domains, such an example would be PopA (c-di-GMP effector) protein of *Caulobacter crescentus* where the A-site of the GGDEF domain is no longer functional but the I-site is fully intact and c-di-GMP is still bound (Duerig *et al.*, 2009).

1.3.4 The 29 GGDEF/EAL domain-containing proteins *E. coli* K-12 W3110

Of the 29 GGDEF/EAL domain genes featured in *E. coli* K-12 substr. W3110, 12 encode GGDEF domain proteins, 10 encode EAL domain proteins and seven encode GGDEF-EAL domain composite proteins (Hengge 2009). Based on biochemical evidence and amino acid sequence conservation (see Table 4.1), 12 of these proteins are putative DGCs, 13 are putative PDEs and four are degenerate GGDEF/EAL domain-containing proteins (see 1.3.3) (Fig. 1.4). A number of these proteins are crucial in the inverse regulation of motility and biofilm formation (see 1.3.6). For about half of these proteins, their respective functions remain to be discerned.

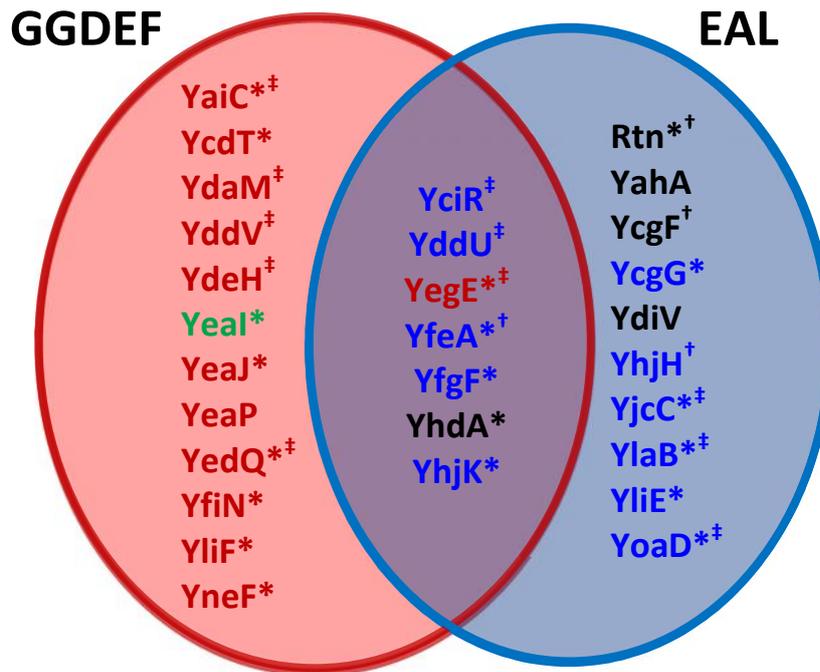


Figure 1.4: The standard complement of 29 GGDEF/EAL domain-containing proteins in *E. coli* K-12 W3110. In red are 12 predicted DGC proteins, in blue are 13 predicted PDE proteins, in green is an effector with a GGDEF A1⁺ domain and in black are degenerate GGDEF/EAL proteins. * designates the 19 membrane bound protein, ‡ designates proteins which are under positive control of stationary phase sigma factor σ^S , and † designates proteins that are under negative control by σ^S . 22 of these 29 proteins are expressed in LB.

1.3.5 C-di-GMP binding effectors and their targets

Besides DGCs and PDEs which control the production and degradation of c-di-GMP, its regulation also requires effectors and effector targets. All of these components are part of the c-di-GMP control module. Effectors bind directly to c-di-GMP and thus ‘sense’ its presence, leading to a conformational change that then activates them to seek out target molecules. The interaction between the target molecule and the effector induces a change in the output production of the target molecule. Effectors are known to exist in the form of proteins, transcription factors and RNAs. Examples include PilZ domain proteins named after the PilZ protein in *P. aeruginosa*, where it was first discovered, the transcription factor FleQ, the PelD protein (Amikam & Galperin, 2006; Lee *et al.*, 2007; Hickman & Harwood, 2008) and a class of riboswitches characterized by the GEMM motif (Genes for the Environment, for Membranes and Motility) (Lee *et al.*, 2010; Sudarsan *et al.*, 2008). The riboswitches undergo structural changes upon binding to c-di-GMP which can affect transcriptional termination, intron splicing and/or translation of coding regions located further downstream (Lee *et al.*, 2010; Sudarsan *et al.*, 2008). Suitable c-di-GMP control module ‘target’ candidates can be promoters, RNAs, enzymes or molecular structures such as the flagellar basal body (Hengge, 2010).

The best studied effector of c-di-GMP in *E. coli* is the PilZ domain family of proteins. Like the GGDEF and EAL domains, the PilZ domain is widely distributed throughout the bacterial

kingdom but is absent from the eukaryotic and archaeal kingdoms. The PilZ domain can exist on its own or coupled to a GGDEF/EAL/HD-GYP domain (Amikam & Galperin, 2006; Merighi *et al.*, 2007). In *E. coli* YcgR and BcsA are PilZ domain proteins. YcgR is involved in the control of motility via the flagellum (Boehm *et al.*, 2010; Fang & Gomelsky, 2010; Paul *et al.*, 2010) and BcsA is involved in the control of cellulose production (see section 1.3.6) (Ross *et al.*, 1987). YcgR also controls motility in *Salmonella*. $\Delta yhjH$ mutants in both *E. coli* and *Salmonella* have greatly reduced swimming ability, but when coupled with *ycgR* inactivation swarming and swimming is restored to 80% of the levels seen in wild-types. In *C. crescentus*, the PilZ homolog, DgrA, is dependent on c-di-GMP for mediating motility of the cell by binding the flagellar protein FliL and in turn blocking motor function (Christen *et al.*, 2007).

FleQ of *P. aeruginosa* is another example of a c-di-GMP effector. FleQ not only activates the expression of flagella synthesizing genes but also represses transcription of the *pel* operon which is involved in the production of extracellular polysaccharides (EPS). When FleQ is bound to c-di-GMP at high c-di-GMP levels the *pel* operon is de-repressed and Pel EPS are biosynthesized. At low levels of cellular c-di-GMP FleQ is unbound and acts as a transcriptional regulator that binds directly to the *pel* promoter DNA and thereby represses its transcription (Hickman & Harwood, 2008). When the *pel* operon is expressed, it produces another effector of c-di-GMP, the PelD protein. PelD expression and binding to c-di-GMP are required for Pel EPS synthesis (Lee *et al.*, 2007).

The PopA protein in *C. crescentus* is a 'degenerate' DGC with a degenerate A-site and an intact I-site which has evolved from a DGC to a c-di-GMP effector. When bound to c-di-GMP, PopA is sequestered to the old cell pole of a differentiating cell during the transition from the G1 to the S phase. There it helps recruit CtrA and RcdA as well as the machinery that degrades CtrA. CtrA is a cell cycle regulator, whose degradation at the appropriate time is necessary in order for chromosomal DNA replication initiation. In this way c-di-GMP is a spatial and temporal regulator (Duerig *et al.*, 2009).

When taking a closer look at various the known c-di-GMP effectors, it becomes evident that c-di-GMP controls widely diverse activities that include flagellar motor function, transcriptional regulation, enzymatic activity, EPS formation and the cell cycle via localized proteolysis.

1.3.6 C-di-GMP inversely controls motility and curli fiber formation in *E. Coli*

When transitioning into the stationary phase, *E. coli* switch from the motile planktonic single-cell state to a sessile state where they exhibit adhesive surface appendices that enable them to adhere to each other and to varying surfaces. This state is known as biofilm formation. For *E. coli* K-12 W3110 strain this transition results in curli fiber expression and occurs in LB medium below 30°C at an OD₅₇₈ of ~3. This change of states or 'life-styles' is mediated by the stationary phase

Introduction

sigma factor RpoS (σ^S) which is the general stress response regulator that controls over 500 genes and enables the transition into the stationary phase (Weber *et al.*, 2005; Hengge 2011), counteracts the deleterious stress effects that triggered the transition and bolsters general resistance to stresses not yet encountered. RpoS is able to reprogram gene expression to achieve these sweeping changes in the cell (like morphology and lifestyle) by outcompeting the vegetative sigma factor RpoD (σ^{70}) and other sigma factors with the help of accessory proteins when recruiting the RNA polymerase holoenzyme (RNAP) (Weber *et al.*, 2005).

When faced with nutrient limitation and other stresses, motility and chemotaxis enables bacteria to move away from the negative stimuli of a microenvironment to seek out better environments. *E. coli* achieves motility in the post-exponential phase through the synthesis of flagella and associated components. This is an intricate process that is activated by the master regulator FlhDC which together with RpoD controls the expression of over 50 genes encoded in 14 operons. The gene expression is divided into a hierarchical structure that consists of three tiers (Chilcott & Hughes, 2000; Aldridge & Hughes, 2002; Soutourina & Bertin, 2003; Chevance & Hughes, 2008). The master operon, *flhDC*, is at the top of this hierarchy and encodes the subunits that make up FlhD(4)C(2), a DNA-binding protein (Wang *et al.*, 2006). FlhDC directly activates the expression of the class 2 genes, the second tier of the hierarchical structure, which encode among other proteins; the flagellar sigma factor FliA (σ^{FliA}), the anti-sigma factor FlgM, hook basal body (HBB) proteins, the hook of the flagellum and the FliZ protein. The FliZ protein is a regulatory factor that interferes with the expression of σ^S dependent genes, in this way giving precedence to the genes expressed by σ^{FliA} and to motility over the general stress response condition (Pesavento *et al.*, 2008; Chevance & Hughes, 2008). FlgM inhibits the activity of FliA during HBB construction, but once the HBB is completed it acts like a type IV secretion system and exports FlgM out of the cell (Minamino *et al.*, 2006). Once FliA is released from the FlgM interaction, it activates the expression of the class 3 genes, the third and final tier (Hughes *et al.*, 1993; Kutsukake 1994; Karlinsey *et al.*, 2000; Pesavento *et al.*, 2008), which control chemotaxis, motor activity, and the assembly of the flagellar filament.

Among the class 3 genes under σ^{FliA} control are the genes *yhjH* and *ycgR* (Girgis *et al.*, 2007). *yhjH* codes for a highly active PDE (Pesavento *et al.*, 2008) and *ycgR* codes for a PilZ-like c-di-GMP effector (Amikam & Galperin, 2006; Ryjenkov *et al.*, 2006). YhjH promotes motility in the early post-exponential phase by hydrolyzing c-di-GMP. The low cellular level of c-di-GMP ensures that YcgR stays unbound and thereby inactive. In the late post-exponential phase FlhDC expression stops and the remaining pool of FlhDC and σ^{FliA} becomes depleted through ClpXP and Lon-mediated degradation (Tomoyasu *et al.*, 2003; Pesavento *et al.*, 2008). This results in the termination of the production of FliZ, YhjH and YcgR, whose levels are slowly depleted by continued slow growth (Pesavento *et al.*, 2008). In parallel, RpoS accumulates and priority is now

Introduction

given to σ^S -dependent gene expression (Hengge-Aronis, 2002). Among these genes are DGCs that outweigh YhjH's PDE activity, so that c-di-GMP is produced faster than it is degraded, which in turn allows YcgR to take effect and affect the cell's motility by altering the pattern of rotation of the flagellum and swimming speed (Boehm *et al.*, 2010; Fang & Gomelsky, 2010; Paul *et al.*, 2010). Which DGCs are expressed at this point is also dependent on the temperature. YedQ is expressed at both 37°C and 28°C but plays a minor role compared to its two counterparts, YeaJ and YegE. At 37°C YeaJ is expressed while at 28°C YegE is expressed (Pesavento *et al.*, 2008).

At 28°C YegE and YedQ inhibit motility and activate *csgD* expression by generating more c-di-GMP than YhjH can degrade, thereby allowing YcgR to become activated by binding c-di-GMP and interfering with the flagellar motor (Pesavento *et al.*, 2008). The higher levels of c-di-GMP also relieve the repression of the DGC YdaM by the PDE YciR, as YciR has higher binding affinity to c-di-GMP. The freed YdaM is then able to increase cellular c-di-GMP levels further and activate the MerR-like transcription factor MlrA to induce transcription of the *csgDEFG* operon which is critical for the expression biofilm matrix components like amyloid curli fibers and cellulose (Lindenberg *et al.*, 2013).

Introduction

et al., 2001; Römling 2005; Weber *et al.*, 2006; Itoh *et al.*, 2008). Curli are adhesive structures that enable cell-cell and cell-surface adherence and in *E. coli* and *Salmonella* are vital components of biofilm. The transcriptional regulator, CsgD, mediates the expression of both curli and cellulose by inducing the *csgBAC* operon and the DGC YaiC, respectively (Olsen *et al.*, 1993; Arnqvist *et al.*, 1994; Hammar *et al.*, 1995; Römling *et al.*, 1998a; Brown *et al.*, 2001).

Curli expression is regulated via a signaling cascade that involves several DGCs, PDEs, the stationary phase RNA polymerase σ^S , the transcription factor MlrA (see section 1.3.5) and the expression of both of the *csgDEFG* and *csgBAC* operons. These two operons are divergently transcribed and are separated by an intergenic region of 752 bps (Fig. 1.6A). The *csgBAC* operon encodes the two subunits of curli fiber. CsgA, also known as curlin, is the major structural protein of curli fiber while CsgB is required for nucleating CsgA polymerization into an amyloid fiber on the cell's surface, thus CsgB adopts a confirmation that templates CsgA aggregation (Hammar *et al.*, 1995; 1996). The third protein encoded by the *csgBAC* operon is CsgC and its exact role remains unclear. Mutated strains lacking *csgC* altogether still were able to assemble curli fibers but exhibited defects in auto-aggregation and binding affinity to both soluble fibronectin and Congo red (colonies paler than wild-type) (Hammar *et al.*, 1995). It is interesting to note that despite many bacterial species known to be able to form amyloid fibers only bacteria within the *Enterobacteriaceae* group have the accessory CsgC protein. Based on the crystal structure of CsgC, the work of Taylor *et al.* (2011) suggests redox activity and with CsgG being the substrate. The *csgDEFG* operon is required for the expression and translocation of the curli fiber subunits. CsgG is an oligomeric, outer-membrane translocator protein (a pore) that secretes CsgA and CsgB onto the cell's surface. CsgE and CsgF are two accessory proteins. CsgE is essential for subunit stability and CsgF is necessary for the cell-association of the fibers (Nenninger *et al.*, 2009, Robinson *et al.*, 2006). CsgD is a transcriptional regulator and has a LuxR-like helix-turn-helix DNA binding motif and is required for the expression of the *csgBAC* operon. The expression of the *csgDEFG* operon is in turn under the control of the transcriptional regulator MlrA.

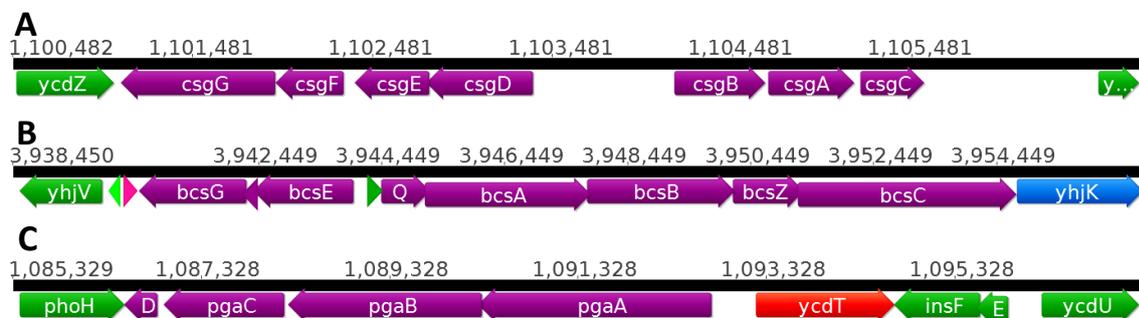


Figure 1.6: Genomic layouts of A) *csg* genes, B) *bcs* genes and C) *pga* genes in the K-12 W3110.

Cellulose synthesis occurs below 30°C and is under the control of the YaiC/YoaD system. YaiC is a putative DGC (fully intact GGDEF domain but DGC activity has not been

Introduction

experimentally substantiated) and YoaD is its antagonistic putative PDE partner (Brombacher *et al.*, 2003; Brombacher *et al.*, 2006). YaiC is expressed in the stationary phase and is under the control of the σ^S /YdaM/MlrA/CsgD cascade (see section 1.3.5). Its antagonistic counterpart is controlled solely by σ^S (Weber *et al.*, 2006) and is down regulated by the overproduction of CsgD (Brombacher *et al.*, 2006). YaiC initiates cellulose production via direct synthesis of c-di-GMP. High levels of c-di-GMP activate cellulose biosynthesis using *bcs* genes (Römling *et al.*, 2000; Simm *et al.*, 2004). In *Salmonella* (Garcia *et al.*, 2004) and *E. coli* (Da Re & Ghigo 2006) the c-di-GMP producing function of YaiC is replaced by the DGC YedQ.

The *bcsQABZC* and the *bcsEFG* operons are divergently transcribed (Fig. 1.6B) where *bcsA* and *bcsB* code for the cellulose synthase. The *bcsA* gene encodes the catalytic subunit of cellulose synthase, BcsA, that forms the transmembrane pore across the cellular inner membrane and binds UDP-glucose. The *bcsB* gene encodes the periplasmic protein, BcsB, which is anchored in the inner membrane and guides the cellulose polymer across the periplasm. BcsC is a putative oxidase that is predicted to be involved in complex assembly and required for cellulose synthesis *in vivo* (Standel *et al.*, 1994). BcsZ is a periplasmic cellulase that appears to enhance cellulose production *in vivo*, though the exact mechanism is still to be elucidated (Standel *et al.*, 1994; Mazur *et al.*, 2011). The *bcsABZC* operon has been shown to be essential for cellulose formation in both *E. coli* and *Salmonella* (Zogaj *et al.*, 2001). BcsQ is also essential for cellulose formation and localizes at the cell pole. It has been suggested to be involved in subcellular localization of the cellulose biosynthesis apparatus (Le Quéré *et al.*, 2009). The *bcsEFG* operon has been shown to be essential for cellulose production in *Salmonella* (Solano *et al.*, 2002) and in *E. coli* (Richter, Povolotsky *et al.*, submitted) but their exact functions remain to be uncovered.

The *pgaABCD* operon is required for the synthesis and excretion of the PGA, an essential polysaccharide adhesion that enables biofilm formation. It is an intercellular adhesion that also attaches to abiotic surfaces (Wang *et al.*, 2004). This exopolysaccharide is used as a biofilm matrix component by some *E. coli*, notably UPECs (Itoh *et al.*, 2008). The *pgaABCD* locus encodes four proteins PgaA, PgaB, PgaC and PgaD (Fig 1.6C). PgaA forms a pore within the cellular membrane and together with PgaB, a N-acetyl-glucosamine deacetylase, exports the PGA polymer across the membrane (Itoh *et al.*, 2008). PgaC and PgaD are inner membrane proteins (Daley *et al.*, 2005). PgaC is a glycosyltransferase which catalyzes the formation of PGA from its UDP-GlcNAc precursor. PgaD is also essential for PGA biosynthesis (Wang *et al.*, 2004; Itoh *et al.*, 2008), though its exact role has not yet been clarified. The *pgaABCD* locus is regulated by the YddV protein (Tagliabue *et al.*, 2010). CsrA inhibits PGA by blocking translation of the *pgaA* gene and degrading the *pgaABCD* mRNA (Wang *et al.*, 2004). CsrA also down-regulates YcdT (DGC) and YdeH by binding to the mRNAs of these GGDEF domain proteins and thus preventing their translation. YdeH (DGC) up-regulates PGA and PgaD through c-di-GMP, but the

exact mechanism is yet to be elucidated. PGA is expressed under various stress conditions including high salt concentration (though the cation-responsive regulatory protein HhaR is necessary for that) (Goller *et al.*, 2006) or the presence of ethanol or glucose (these two being independent of CsrA) (Cerca *et al.*, 2008).

1.3.8 C-di-GMP's role in virulence regulation

Biofilm forming bacteria are difficult to treat as biofilms enable bacteria to evade the immune system and provide antibiotic resistance (Anderl *et al.*, 2000; Stewart & Costerton 2001; Hoiby *et al.*, 2002; Harrison *et al.*, 2007, 2009). UPEC strains are the leading cause of UTIs, where biofilms are involved in urinary catheter cystitis. Using biofilms they are able to anchor themselves to the superficial umbrella cells which they invade and as a result are more resistant to the host immune system and antibiotic treatments (Nickel *et al.*, 1994; Hatt & Rather, 2008). The work of Cerca *et al.*, (2007) found that in UPEC strains a major part of the clinical isolates expressed PGA in the host. Furthermore, it has been found that the *pgaABCD* locus (*E. coli*) is a close homolog of the *icaADBC* locus (*Staphylococci*) and PGA is biochemically indistinguishable from PNAG (the product of *icaADBC*). PNAG is also a polymer which promotes biofilm formation and has been shown to be a virulence factor in *Staphylococci* (Kropec *et al.*, 2005; Rupp *et al.*, 1999). Thusly it seems to also be a virulence factor in UPECs (Agladze *et al.*, 2005). UPECs rely on PGA for intracellular adhesion which is expressed at 37°C (Wang *et al.*, 2004; Cerca *et al.*, 2008). The *pgaABCD* operon is regulated by the DGC YddV (see section 1.3.6 for details) (Tagliabue *et al.*, 2010). PGA expression is dependent on the ratio of ppGpp to c-di-GMP (Boehm *et al.*, 2009) and by extension on the ratio between DGCs and PDEs, the proteins that make and break c-di-GMP, respectively. Basal ppGpp levels have been shown to be a necessity for optimal virulence in several species of bacteria (Braeken *et al.*, 2006). Besides being a virulence factor, PGA is an exopolysaccharide that UPECs use along with other bacteria as a biofilm matrix component. One example is *Yersinia pestis*, the causative agent of the Bubonic Plague or the Black Death, which ravaged Europe throughout the Middle Ages and uses PGA to colonize the flea as part of its infection cycle (Hinnebusch & Erickson, 2008).

Xanthomonas campestris is a plant pathogen that is responsible for black rot of crucifers in which low levels of c-di-GMP are associated with virulence. Clp is a transcriptional regulator that controls more than 300 virulence genes and is repressed when bound to c-di-GMP (He *et al.*, 2007). At high levels of c-di-GMP Clp is bound and virulence is abolished. The PDEs RpfG and RavR hydrolyze c-di-GMP and together create a low c-di-GMP level cellular environment where Clp is in its active form. RpfG and RavR are activated through phosphorylation by RpfC and RavS, which in turn are activated by hypoxia and diffusible signaling factor (DSF). Finally, DSF synthesis is controlled by RpfF. The PDE RpfG is a response regulator that not only indirectly

Introduction

regulates virulence factor secretion but also suppression of biofilm formation and stimulates type IV pilus-mediated motility (Andrade *et al.*, 2006; Römling *et al.*, 2013).

Biofilm formation has been linked with virulence extensively in *Vibrio cholerae* and *P. aeruginosa*. *V. cholerae* uses *vibrio* exopolysaccharides (VPS) to form biofilms. VPS biosynthesizing enzymes are encoded in two operons that are controlled by the transcriptional activators VpsR and VpsT (Yildiz *et al.*, 2001; Casper-Lindley *et al.*, 2004; Krasteva *et al.*, 2010). By maintaining low levels of cellular c-di-GMP the PDE VieA inhibits biofilm formation by repressing the transcription of the *vps exopolysaccharide* operon and indirectly induces acute virulence gene expression through positive transcriptional regulation of *ctxAB* [cholera-toxin (CT)] and the transcription factor *toxT* (Tischler & Camilli, 2005). *V. cholerae* uses biofilms as part of its strategy to survive in the environment and in the host switches lifestyles to the planktonic, motile state. This transition is regulated by c-di-GMP and VieA (Tischler & Camilli, 2005). *V. cholerae* El Tor uses the quorum sensing regulator HapR to indirectly inhibit biofilm formation by reducing the overall intercellular c-di-GMP concentration (Waters *et al.*, 2008). *P. aeruginosa* has a reversed relationship between c-di-GMP and acute virulence, where c-di-GMP promotes chronic virulence. *P. aeruginosa* from chronically infected respiratory tract specimens of cystic fibrosis (CF) patients has been shown to exist in a genetically diverse bacterial population. Some of the bacteria form 'small colony variants' (SCV) which show increased antibiotic resistance to a broad range of antimicrobial agents (Govan *et al.*, 1996). This SCV phenotype is due to the expression of fimbriae encoded by the chaperone usher pathway (*cupA*) gene cluster in *P. aeruginosa*. CupA expression is modulated by the intracellular level of c-di-GMP via proteins containing EAL and GGDEF domains (Friedman & Kolter, 2004; Jackson *et al.*, 2004). The EPS alginate is one biofilm matrix component of mucoid *P. aeruginosa*, a virulence factor that is often recovered from bacteria inhabiting the infected lung of cystic fibrosis patients (Govan *et al.*, 1996; Friedman & Kolter, 2004; Jackson *et al.*, 2004). Alg44 is a c-di-GMP binding protein (via its PilZ domain) that is required for the polymerization and transport of alginate. Nonmucoid *P. aeruginosa* uses the EPSs, Pel and Psl, as the major components in the extracellular matrix of its biofilm. Pel and Psl are positively regulated by c-di-GMP on the transcriptional level, and biofilm formation increases when the DGC WspR is activated (Hickman *et al.*, 2005). The study of Kulasakara *et al.*, (2006) found a strong correlation between the formation of hyperbiofilms and high levels of intracellular c-di-GMP (Kulasakara *et al.*, 2006).

To summarize, in certain pathogens chronic infection is linked with high cellular levels of c-di-GMP and biofilm formation, while in others low cellular levels of c-di-GMP and the motile state are associated with acute virulence.

2 Aims

The Hengge group has been working on understanding the underlying mechanisms of stress response and biofilm formation in *E. coli*. The second messenger molecule, c-di-GMP, is a driving force behind the regulation of these processes. Its activity is modulated by two sets of proteins: diguanylate cyclases (DGCs) synthesize and phosphodiesterases (PDEs) degrade c-di-GMP. DGCs are characterized by the presence of the GGDEF domain and PDEs are identified by the EAL domain. In the *Escherichia coli* K-12 strains W3110 and MG1655, 29 different GGDEF/EAL domain-containing proteins have been identified. However, little is known about how that list may differ between strains of *E. coli*, particularly comparing pathogenic and commensal strains. The aim of this work was to systematically compare the complement of putative DGC and PDE proteins as well as genes associated with biofilm formation, among 61 strains of *E. coli* including enterohemorrhagic (EHEC), uropathogenic (UPEC), enteroaggregative (EAEC), enteropathogenic (EPEC) and enterotoxigenic (ETEC) as well as commensal strains. The overarching goal was to uncover potential trends, anomalies or novelties, to determine whether pathogens employ the same complement or whether they have evolved novel adaptations, and to study specific new DGCs/PDEs and their regulation and function experimentally – ideally in relation to the pathotype/virulence properties of the respective strain.

The 2011 outbreak in Germany of the Shiga toxin (Stx)-producing *Escherichia coli* O104:H4 had a high incidence of HUS (more than 20% of patients). The outbreak strain is genetically most similar to an EAEC but has acquired a Stx carrying phage from EHEC (Mellmann *et al.*, 2011) and contains a novel gene *dgcX*. This prompted the further investigation of biofilm forming properties of the outbreak strain and four other strains (the classical EAEC 55989, the closely related EAEC HUSEC041, the classical EHEC EDL933 and the laboratory K-12 strain W3110), as well as exploration of the function of *dgcX* and its relationship to virulence in EAECs of the O104:H4 serotype.

3 Materials and Methods

3.1 Chemicals, materials and technical equipment

In Tables 3.1 and 3.2 is the comprehensive list of all chemicals and equipment used in this study, and their corresponding suppliers.

Table 3.1: Suppliers of chemicals used in this work.

| Product | Supplier |
|---|---------------------|
| Agarose | Biozym |
| Alkaline phosphate: Calp Intestinal Phosphate (CIP) | New England Biolabs |
| Antibiotics | Roth, Sigma-Aldrich |
| Anti-Rabbit IgG-Alkaline Phosphate antibody | Sigma-Aldrich |
| Anti-HIS antibody | USBiological |
| APS (Ammonium persulfate) | Roth |
| BCIP (5-Bromo-4-chloro-3-indolyl phosphate) | Roth |
| Blotting membrane for proteins: Roti-PVDF, pore size 0.45 μ M | Roth |
| Blotting paper | Whatman |
| Bradford solution for protein determination | AppliChem |
| Bromphenol blue | Roth |
| Chitin-Beads | New England Biolabs |
| Coomassie Brilliant Blue R-250 | AppliChem |
| DNA preparation kit: QIAamp DNA Mi | Analytik Jena |
| DNaseI for protein purification buffer (powder) | Roche |
| DNA size markers: 100 bp DNA Ladder, λ DNA-BstEII Digest | New England Biolabs |
| DNA polymerase: Opti-Taq | Roboklon |
| dNTPs Mix | MP Biomedicals |
| Gel extraction kit | Qiagen |
| IPTG (isopropyl β -D-1-thiogalactopy | Roth |
| Milk powder | Roth |

Materials and Methods

| Product | Supplier |
|--|-----------------------|
| NBT (Nitroblue tetrazolium chloride) | AppliChem |
| Oligonucleotide primers | Metabion |
| ONPG (ortho-nitrophenyl- β -D-galactopyranoside) | Roth, Serva |
| PCR purification kit | Qiagen |
| Plasmid mini preparation kit: QIAprep Spin Miniprep Kit, InnuPREP Plasmid Mini Kit | Qiagen, Analytik Jena |
| Polyvinylidene fluoride (PVDF) membrane | Sigma-Aldrich |
| Protein size marker: Prestained Protein Marker | New England Biolabs |
| Restriction enzymes | New England Biolabs |
| T4 DNA ligase | New England Biolabs |
| TEMED (N,N,N',N'-Tetramethylethylenediamine) | Roth |
| X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) | AppliChem |
| Xylene Cyanole | Roth |

Table 3.2: Suppliers of technical equipment used in this work.

| Equipment | Supplier |
|--|---|
| Äkta system | GE Healthcare |
| Centrifuges | Eppendorf, Heraeus, Thermo Scientific |
| Electroporator | BioRad |
| ELISA-Reader | BioRad |
| French Press | Thermo Electron Corporation/Amincon SLM Instruments |
| Gel documentation | Alpha Innotech/ Biozym |
| Gel electrophoresis | BioRad, Biozym |
| Gel filtration column: Superdex 75 (16/60) | GE Healthcare |
| Incubators | Infors HT |
| Incubator, shaking (Thermo Mixer) | Eppendorf |
| NanoDrop | Peqlab |
| PCR-Thermocycler | Thermo Scientific |

Materials and Methods

| Equipment | Supplier |
|----------------------------|------------|
| PhosphorImager (FLA-2000G) | Fujifilm |
| Photo developing machine | Protec |
| Photometer | Healthcare |
| Semi-dry Electro Blotter | Peqlab |

3.2 Media and Additives

3.2.1 Media

- Bacteria grown in nutrient-rich medium followed the Silhavy *et al.*, (1984) protocol for Luria-Bertani (LB) medium and contained 5g of yeast extract, 10g of Bacto-tryptone and 5g of NaCl per liter.
- Congo red (CR) agar plates contained LB agar without salt and 20mL CR solution per liter of medium. The LB agar without salt was sterilized via autoclaving while the CR solution was sterilized through filtration.
- Congo red (CR) solution consisted of 2mg/mL of Congo red and 1mg/mL Coomassie Brilliant Blue G mixed with 70% Ethanol. It was sterilized through filtration.
- 6X DNA loading dye buffer consisted of 0.25 % Bromphenol Blue, 0.25 % Xylencyanol and 30 % Glycerin mixed in 1X TAE buffer.
- LB/agar plates contained 15g of Bacto-Agar, 5g of yeast extract, 10g of Bacto-tryptone and 5g of NaCl per liter of LB medium.
- LB/agar plates without salt contained 15g of Bacto-Agar, 5g of yeast extract and 10g of Bacto-tryptone per liter of LB medium.
- 10X M9 minimal buffer contained 75.2g of $\text{Na}_2\text{HPO}_4 \times \text{H}_2\text{O}$, 10g of KH_2PO_4 , 10g of $\text{NH}_4 \text{Cl}$, 5g of NaCl per liter of total volume of DI water. The buffer was sterilized through autoclaving.
- Minimal media (M9) plates contained 15g of Bacto-agar in 900mL of DI water sterilized through autoclaving, to which 100mL of sterile 10X M9 minimal buffer was added.
- Motility (swim) plates contained 5g of Bacto-tryptone, 5g of NaCl and 3g of Bacto-agar per liter of total volume of DI water.
- ONPG (ortho-Nitrophenyl- β -galactoside) consisted of 0.4g of ONPG diluted in total volume of 50mL water. The container was shielded from light by wrapping aluminum foil around it. It was prepared fresh before each use.
- 50X TAE buffer consisted of 242g of Tris-Base, 57.1mL acetic acid and 100mL 0.5M EDTA in 1L a total volume of water.

Materials and Methods

- TOP-Agar used for creating *lacZ* fusions contained 7g of Bacto-Agar per 1L of total volume LB-medium and was autoclaving.
- 2X TSS (transformation and storage solution) (Chung *et al.*, 1989) contained 90mL of LB medium, 20g of PEG-600, 10mL of 100% DMSO (dimethyl sulfoxide), and 10mL of 1M MgSO₄ or MgCl₂, at a final pH of 6.5. Sterilized through autoclaving and stored at -4°C.
- Z-buffer (β-Galaktosidase enzyme buffer) contained 39mM of NaH₂PO₄, 61mM of Na₂HPO₄, 10mM KCL and 10mM of MgSO₄ x 7 H₂O in 1L total volume of water.

3.2.2 Additives

Concentrated stock solutions were prepared of the media listed in Table 3.3:

Table 3.3: Additives and their final concentrations in media.

| Additive | Dissolvent | Final concentration (If not otherwise specified) |
|-----------------|-------------------------|--|
| Ampicillin | water | 100 µg/mL |
| Chloramphenicol | 70% ethanol | 15 or 30 µg/mL |
| Kanamycin | water | 50 µg/mL |
| Sodium citrate | water | 20 mM |
| Tetracylin | 70% ethanol | 5 µg/mL |
| X-Gal | DMF (dimethylformamide) | 30 µg/mL |

3.3 Bacterial strains, bacteriophages, plasmids and oligonucleotide primers used in this work

3.3.1 Bacterial strains

Table 3.4: Bacterial strains used in this work. For strains with *lacZ* fusions, the end points of the fused promoter/gene fragments are noted with numbers indicating the position relative to the translational start site of the respective gene. *ΔrpoS::kan* was P1 transduced from W3110 *rpoS::kan* (F. Mika, unpublished). *ΔcsgD::kan* was P1 transduced from W3110 *csgD::kan* (described in Sommerfeldt *et al.*, 2009). *ΔyegE::scar* was P1 transduced from W3110 *yegE::kan* (described in Weber *et al.*, 2006) and the KAN cassette was then flipped out using Datsenko and Wanner method (2000).

| <i>Wild-type strains</i> | | |
|--------------------------|--|-------------------------------|
| Strain | Genotype or description | Reference |
| W3110 | E.coli K12 thyA36 deoC2 IN(rrnD-rrnE)I | Hayashi <i>et al.</i> , 2006 |
| LB226692 | Sequenced 2011 German outbreak strain isolate (EAEC) | Mellmann <i>et al.</i> , 2011 |
| RKI II-2027 | Official 2011 German outbreak strain isolate (EAEC) | |

Materials and Methods

| HUSEC041 | A historic O104:H4 HUS isolate from 2001 (01-09591) (EAEC) | Mellmann <i>et al.</i> , 2011 |
|--|---|---|
| 55989 | 'classical' EAEC, 1999 isolate from HIV patient from Central African Republic | Mossoro <i>et al.</i> , 2002 |
| EDL933 | 'classical' O157:H7 EHEC, 1982 isolate from outbreak in USA | Riley <i>et al.</i> , 1983 |
| <i>Derivatives constructed by other members of the Hengge group</i> | | |
| Strain | Genotype | Reference |
| GB1000 | E.coli K12 thyA36 deoC2 IN(rrnD-rrnE)I, Dlac(I-A) | Gisela Klauck & Regine Hengge, unpublished data |
| GB1100 | GB1000 csgB(-190, +43)::lacZ | Gisela Klauck & Regine Hengge, unpublished data |
| NS252 | GB1000 <i>yaiC</i> (-213, +79)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS255 | GB1000 <i>ycdT</i> (-582, +58)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS307 | GB1000 <i>ydaM</i> (-307, +43)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS303 | GB1000 <i>yddV</i> (-391, +49)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS298 | GB1000 <i>ydeH</i> (-321, +43)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS313 | GB1000 <i>yeaJ</i> ::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS296 | GB1000 <i>yeaP</i> (-309, +40)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS304 | GB1000 <i>yedQ</i> (-196, +58)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS305 | GB1000 <i>yegE</i> (-360, +22)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS253 | GB1000 <i>yfiN</i> (-321, +55)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| NS301 | GB1000 <i>yliF</i> (-321, +40)::lacZ (<i>hybr.</i>) | Sommerfeldt <i>et al.</i> , 2009 |
| <i>Derivatives constructed in this work</i> | | |
| TLP1 | GB1000 <i>pdeT</i> (-356, +21)::lacZ | This work |
| TLP2 | GB1000 <i>yneF</i> (-373, +14)::lacZ | This work |
| TLP3 | GB1000 <i>dgcX</i> (-412, +18)::lacZ | This work |
| TLP4 | GB1000 <i>ycdTpdeT</i> (-2053, +21)::lacZ | This work |
| TLP7 | GB1000 <i>dgcX</i> (-412, +18)::lacZ Δ <i>rpoS</i> ::kan | This work |
| TLP8 | GB1000 <i>dgcX</i> (-412, +18)::lacZ Δ <i>csgD</i> ::kan | This work |
| TLP9 | GB1000 <i>yneF</i> (-373, +14)::lacZ Δ <i>rpoS</i> ::kan | This work |

| | | |
|-------|--|-----------|
| TLP10 | GB1000 <i>yneF</i> (-373, +14):: <i>lacZ</i> Δ <i>csgD</i> :: <i>kan</i> | This work |
| TLP11 | GB1000 <i>pdeT</i> (-356, +21):: <i>lacZ</i> Δ <i>rpoS</i> :: <i>kan</i> | This work |
| TLP12 | GB1000 <i>pdeT</i> (-356, +21):: <i>lacZ</i> Δ <i>csgD</i> :: <i>kan</i> | This work |
| TLP13 | GB1000 <i>ycdTpdeT</i> (-2053, +21):: <i>lacZ</i> Δ <i>rpoS</i> :: <i>kan</i> | This work |
| TLP14 | GB1000 <i>ycdTpdeT</i> (-2053, +21):: <i>lacZ</i> Δ <i>csgD</i> :: <i>kan</i> | This work |
| TLP15 | GB1000 <i>cell</i> ⁺ :: <i>scar</i> * <i>dgcX</i> :: <i>lacZ</i> | This work |
| TLP16 | GB1000 <i>cell</i> ⁺ :: <i>scar</i> * <i>dgcX</i> :: <i>lacZ</i> Δ <i>rpoS</i> :: <i>kan</i> | This work |
| TLP71 | GB1000 <i>cell</i> ⁺ :: <i>scar</i> * <i>dgcX</i> :: <i>lacZ</i> Δ <i>csgD</i> :: <i>kan</i> | This work |
| TLP17 | GB1000 with <i>dgcX</i> integrated into the chromosome at lambda <i>attB</i> site | This work |
| TLP18 | TLP 17 <i>cell</i> ⁺ :: <i>scar</i> * | This work |
| TLP30 | GB1000 <i>cell</i> ⁺ :: <i>scar</i> * | This work |
| TLP33 | TLP 17 <i>cell</i> ⁺ :: <i>scar</i> * Δ <i>yegE</i> :: <i>scar</i> | This work |
| TLP34 | TLP 17 Δ <i>yegE</i> :: <i>scar</i> | This work |
| TLP29 | GB1000 Δ <i>yegE</i> :: <i>scar</i> | This work |
| TLP41 | GB1000 <i>cell</i> ⁺ :: <i>scar</i> * Δ <i>yegE</i> :: <i>scar</i> | This work |

**cell*⁺::*scar* refers to cellulose forming W3110 derivative where the 6th codon of *bcsQ* gene was changed from a TAG (stop) to TTG (leucine) in the chromosomal DNA of the W3110 strain. The mutation was achieved using a one-step replacement procedure described by Datsenko & Wanner (2000) with the primers described in Serra *et al.*, (2013a), where the final strain had a KAN marker at the end of this mutation. A P1 lysate was made from this strain and the *bcsQ* mutation was then transferred into the GB1000 strain by P1 transduction and then the antibiotic resistance cassette was flipped out using the Datsenko & Wanner (2000) protocol creating a *scar*.

3.3.2 Bacteriophages

Table 3.5: Bacteriophages used in this work.

| Lysat | Reference |
|----------------|---------------------------------------|
| λ RS45 | (Simons <i>et al.</i> , 1987) |
| λ RS74 | (Simons <i>et al.</i> , 1987) |
| P1vir | Laboratory collection of Hengge group |

3.3.3 Plasmids

Table 3.6: Plasmids used in this work. For pJL28 derivatives with *lacZ* fusions, the end points of the fused promoter/gene fragments are noted with numbers indicating the position relative to the translational start site of the respective gene. Antibiotic resistance encoded by the respective plasmid is noted as follows: AmpR: ampicillin resistance, KanR: kanamycin resistance, CmR: chloramphenicol resistance, TetR: tetracylin resistance

| Cloning vectors and plasmids generated elsewhere | | |
|---|---|-----------------------------------|
| Plasmid | Description | Reference |
| pBT | BacterioMatch II Two-Hybrid System bait vector, encodes the full-length bacteriophage λ repressor, CmR | Agilent Technologies |
| pBT- <i>rpoS</i> | pBT-derivative expressing RpoS fused to the full-length bacteriophage λ repressor | Klauck & Hengge, unpublished data |
| pBT- <i>rpoS</i> _K173E | pBT-derivative with a mutated RpoS fused to the full-length bacteriophage λ repressor | Klauck & Hengge, unpublished data |
| pCAB18 | IPTG-inducible low copy number vector carrying the <i>p_{tac}</i> promoter, AmpR | Barembuch & Hengge 2007 |
| pCP20 | helper plasmid that enables the elimination of the antibiotic resistance genes after one-step inactivation; encoding FLP recombinase, synthesis of FLP is temperature-inducible; temperature-sensitive replication; AmpR, CmR | Datsenko & Wanner 2000 |
| pJL28 | vector for the generation of <i>lacZ</i> fusions, AmpR | Lucht <i>et al.</i> , 1994 |
| pKD4 | Template plasmid for one-step inactivation, carrying the <i>kan</i> -cassette; AmpR, KanR | Datsenko & Wanner 2000 |
| pKD13 | Template plasmid for one-step inactivation, carrying the <i>cat</i> -cassette; AmpR, CmR | Datsenko & Wanner 2000 |
| pKD46 | helper plasmid for one-step inactivation; encodes the λ Red recombinase; temperature-sensitive replication; AmpR | Datsenko & Wanner 2000 |
| pTRG | BacterioMatch II Two-Hybrid System target vector; encodes the RNAP α NTD, TetR | Agilent Technologies |
| pTRG- <i>rssB</i> | pTRG-derivative expressing RssB fused to the RNAP α NTD | Klauck & Hengge, unpublished data |
| pQE30 | Protein expression vector for expression of proteins with a N-terminal 6X HIS-tag | Qiagen |
| pQE60 | Protein expression vector for expression of proteins with a C-terminal 6X HIS-tag | Qiagen |
| Constructs generated in this work | | |
| Plasmid | Description | Reference |

Materials and Methods

| | | |
|------------------------------------|---|-----------|
| pCAB18- <i>dgcX</i> | pCAB18-derivative expressing DgcX | This work |
| pCAB18- <i>yaiC</i> | pCAB18-derivative expressing YaiC | This work |
| pBT- <i>ycdT</i> | pBT-derivative expressing YcdT fused to the full-length bacteriophage λ repressor pBT-derivative expressing RpoS fused to the full-length bacteriophage λ repressor | This work |
| pBT- <i>pdeT</i> | pBT-derivative expressing PdeT fused to the full-length bacteriophage λ repressor | This work |
| pTRG- <i>ycdT</i> | pTRG-derivative expressing YcdT fused to the RNAP α NTD | This work |
| pTRG- <i>pdeT</i> | pTRG-derivative expressing PdeT fused to the RNAP α NTD | This work |
| pJL28- <i>yneF</i> | pJL28-derivative with YneF promoter fused to lacZ | This work |
| pJL28- <i>pdeT</i> | pJL28-derivative with PdeT promoter fused to lacZ | This work |
| pJL28- <i>ycdTpdeT</i> | pJL28-derivative with YcdTPdeT promoter fused to lacZ | This work |
| pJL28- <i>dgcX</i> | pJL28-derivative with DgcX promoter fused to lacZ | This work |
| pJL28- <i>dgcX</i> _{full} | pJL28-derivative with whole DgcX used to integrate <i>dgcX</i> into chromosomal DNA of GB1000 at <i>attB</i> site | This work |
| pQE30- <i>dgcX</i> | pQE30-derivative expressing DgcX (GGDEF domain only) | This work |
| pQE30- <i>yneF</i> | pQE30-derivative expressing YneF (GGDEF domain only) | This work |
| pQE60- <i>dgcX1</i> | pQE60-derivative expressing DgcX 1 st construct | This work |
| pQE60- <i>dgcX2</i> | pQE60-derivative expressing DgcX 2 nd construct | This work |
| pQE60- <i>yaiC</i> | pQE60-derivative expressing YaiC | This work |

3.3.4 Primers

Table 3.7: Oligonucleotide primers used in this work. Bold italic lettering denotes restriction sites and lowercase lettering indicates point mutations in primers used for recombinant PCR.

| I. Primers used for generating <i>lacZ</i> fusions | | |
|---|------------------------------|---|
| Primers for cloning gene fusion fragments into the pJL28 vector | | |
| lacZ fusion | Primer name | Sequence |
| <i>dgcX</i> (-412, +18):: <i>lacZ</i> | <i>dgcX</i> -d(-412) (EcoRI) | 5'-GCAA <i>GAaTTc</i> TGTGCCTCAGTTTTGTC-3' |
| | <i>dgcX</i> -u-18 (HindIII) | 5'-ATAG <i>AaGcTTG</i> GTTGATAATCATGTACGC ACC-3' |
| <i>yneF</i> (-373, +14):: <i>lacZ</i> | <i>yneF</i> -d(-373) (EcoRI) | 5'-GGCG <i>GAaTTc</i> GTCGCTAAACGTC-3' |
| | <i>yneF</i> -u-14 (HindIII) | 5'-GGG <i>TaagCtTGC</i> ATTAAATTATCTGTCCATA C-3' |

Materials and Methods

| | | |
|--|---|---|
| <i>pdeT</i> (-356, +21):: <i>lacZ</i> | <i>pdeT</i> -d(-356)(EcoRI) <i>pdeT</i> -u-21 (BamHI) | 5'-GGCG GAaTTc GTCGCTAAACGTC -3' 5'-AGACG GAaTTC GATACTGAACGTGCG-3' |
| <i>ycdTpdeT</i> (-2053, +21):: <i>lacZ</i> | <i>pdeT</i> -d(-2053) (EcoRI) <i>pdeT</i> -u-21 (BamHI) | 5'- GACG GAaTTC GATACTGAACGTGCG -3' 5'-AGACG GAaTTC GATACTGAACGTGCG-3' |
| <i>dgcX</i> (-412, +1318):: <i>lacZ</i> | <i>dgcX</i> -d-412 (EcoRI) <i>dgcX</i> -u(-1318) (HindIII) | 5'-GCAAA GAaTTc TGTGCCTCAGTTTTGTC-3' 5'-GGGCT GaAgcTT AAAAAGGTGTTTATTTT AGTGAG -3' |
| Primers for testing single lysogeny of chromosomal <i>lacZ</i> fusions (Powell <i>et al.</i> , 1994) | | |
| Primer name | | Sequence |
| P1 | | 5'-GAGGTACCAGCGCGGTTTGATC-3' |
| P2 | | 5'-TTTAATATATTGATATTATATCATTTTACGTTTCTCGTTC-3' |
| P3 | | 5'-ACTCGTCGCGAACCGCMC-3' |
| Primers for extracting gene fusions and re-sequencing, (<i>lacZ</i> -u-110 constructed by Athanasios Typas) | | |
| lacZ fusion | Primer name | Sequence |
| <i>dgcX</i> (-412, +18):: <i>lacZ</i> | <i>dgcX</i> -d(-412) (EcoRI) <i>lacZ</i> -u-110 | 5'-GCAAA GAaTTc TGTGCCTCAGTTTTGTC-3' 5'-CGCCAGCTGGCGAAAGGG-3' |
| <i>yneF</i> (-373, +14):: <i>lacZ</i> | <i>yneF</i> -d(-373) (EcoRI) <i>lacZ</i> -u-110 | 5'-GGCG GAaTTc GTCGCTAAACGTC-3' 5'-CGCCAGCTGGCGAAAGGG-3' |
| <i>pdeT</i> (-356, +21):: <i>lacZ</i> | <i>pdeT</i> -d(-356)(EcoRI) <i>lacZ</i> -u-110 | 5'-GGCG GAaTTc GTCGCTAAACGTC -3' 5'-CGCCAGCTGGCGAAAGGG-3' |
| <i>ycdTpdeT</i> (-2053, +21):: <i>lacZ</i> | <i>pdeT</i> -d(-2053) (EcoRI) <i>lacZ</i> -u-110 | 5'- GACG GAaTTC GATACTGAACGTGCG -3' 5'-CGCCAGCTGGCGAAAGGG-3' |
| <i>dgcX</i> (-412, +1318):: <i>lacZ</i> | <i>dgcX</i> -d-412 (EcoRI) <i>lacZ</i> -u-110 | 5'-GCAAA GAaTTc TGTGCCTCAGTTTTGTC-3' 5'-CGCCAGCTGGCGAAAGGG-3' |
| II. Primers for cloning <i>dgcX</i> and <i>yaiC</i> into pCAB18 | | |
| Primer name | | Sequence |
| <i>dgcX</i> -d(-62) (EcoRI) _{fwd} | | 5'- GTAAG gAATTc TCGTTATTTGTAGCGATACATATTA -3' |
| <i>dgcX</i> -u-1360 (HindIII) _{rev} | | 5'- GAACT GaAGct TATTATATGGAATAATATTCTGCAGT -3' |
| <i>yaiC</i> -d(-8) (PciI/AflIII) _{fwd} | | 5'- GTTCT acATGT TCCCAAAAATAATGAATGATGAAAAC -3' |

| | | |
|--|---|---|
| yaiC-u-1094 (BglII) _{rev} | 5'- CGCCGG AgaTct GGCCGCCACTTCGGTGC -3' | |
| III. Primers for cloning <i>dgxX</i> and <i>yaiC</i> into pQE60 | | |
| Primer name | Sequence | |
| dgxX-d-(-62) (EcoRI) _{fwd} 1 st construct | 5'- GTAAG GgAATTc tCGTTATTTGTAGCGATACATATTA-3' | |
| dgxX-u-1310 (BglII) _{rev} 1 st construct | 5'- AAAAG GgAgaTct TTTTGAGTGAGTTATCACC -3' | |
| dgxX-d-(-8) (NcoI) _{fwd} 2 nd construct | 5'- GTGCGT cCATGg TTATCAACCAGGTAC-3' | |
| dgxX-u-1310 (BglII) _{rev} 2 nd construct | 5'- AAAAG GgAgaTct TTTTGAGTGAGTTATCACC -3' | |
| yaiC-d-(-8) (PciI/AflIII) _{fwd} | 5'- GTTTCT acATGT TCCCAAAAATAATGAATGATGAAAAC -3' | |
| yaiC-u-1094 (BglII) _{rev} | 5'- CGCCGG AgaTct GGCCGCCACTTCGGTGC -3' | |
| IV. Primers for cloning <i>dgxX</i> and <i>yneF</i> into pQE30 for DGC Assay | | |
| Primer name | Sequence | |
| dgxX-d-819 (BamHI) _{fwd} | 5'- TTCAG GggATcca ATCATGCTAATCATATGGC -3' | |
| dgxX-u-1318 (HindIII) _{rev} | 5'- GGGCT GaAgcTT AAAAAGGTGTTTATTTGAGTGAG -3' | |
| yneF-d-858 (BamHI) _{fwd} | 5'- TTTAG CggatCcg CGATCAATTCGCTAATGAAG -3' | |
| yneF-u-1400 (HindIII) _{rev} | 5'- GAATA AAAGctt TCAGACAACCTCCTCACCGTAACG -3' | |
| V. Primers for cloning <i>pdeT</i> and <i>ycdT</i> into plasmids of the bacterial two-hybrid system | | |
| Construct | Primer name | Sequence |
| pBT-pdeT | pdeT-d- (EcoRI) | 5'-ACTT gaattc gCAGCGCTGGATATCTCGG-3' |
| | pdeT-u- (BamHI) | 5'-GGTT ggatcc ttaTCTGTTTACATACCATTGATAAAAC-3' |
| pBT-ycdT | ycdT-d- (EcoRI) | 5'-TTCAG gaattc AcgaGTAACAAAGAACATTGCACATC-3' |
| | ycdT-u- (BamHI) | 5'-AAAT ggatcc ttaTGGTGACTCACAAAATTACACC-3' |
| pTRG- pdeT | pdeT-d- (EcoRI) | 5'-GGT gaattc TttaTCTGTTTACATACCATTGATAAAAC-3' |
| | pdeT-u- (BamHI) | 5'-CTTC ggatcc CAGCGCTGGATATCTCGGAA-3' |
| pTRG-ycdT | ycdT-d- (EcoRI) | 5'-AAA gaattc TttaTGGTGACTCACAAAATTACACC -3' |
| | ycdT-u- (BamHI) | 5'-CAGT ggatcc cgagTAACAAAGAACATTGCACAT -3' |

3.4 Microbiological methods

3.4.1 Sterilization

All media and chemical solutions, unless stated otherwise, were wet-autoclaved for 20 minutes at 120°C and 1 bar of pressure. Glassware was dry-heat-sterilized at 180°C for 6 hours and heat-sensitive solutions were sterilized by filtration.

3.4.2 Long-term storage of bacterial strains

Bacterial strains were stored long-term at -80°C by mixing 930µL of overnight culture with 70µL of DMSO (dimethyl sulfoxide) (final concentration of 7%).

3.4.3 Growth conditions

Bacterial liquid cultures were grown either at 28°C or 37°C in glass conical flasks or glass test tubes and were filled 20-25% of the total volume. Conical flasks were either incubated in water baths with shaking set to 300 rpm to allow for aeration or in air shakers with shaking set to 130 rpm. Cultures in glass test tubes were aerated by rolling. Bacteria grown on solid medium were incubated at either 28°C or 37°C. For standard experiments, bacterial cultures grown in LB medium were inoculated to a start OD_{578nm} of 0.05.

3.4.4 Bacterial motility assay

Motility assays were performed by inoculating 3µL of overnight cultures adjusted to an OD_{578nm} of 2 or 3 into motility plates (see 3.2.1 for recipe) and allowed to grow for 4-6 hours at 28°C or 3-4 hours at 37°C. The radius of the swimming halo for each inoculated sample was measured and compared to the wild type.

3.4.5 Determination of the cell density of liquid bacterial cultures

The cell density of liquid bacterial cultures was determined using a spectrometer measuring the optical density at a wavelength of 578nm. The sterile medium in which the bacteria were grown was used as a blanking reference. To ensure accurate measurements, probes whose optical density (OD) of 578nm was higher than 0.4, were diluted first.

3.4.6 Transformation

The process of transformation involves the direct uptake and incorporation exogenous DNA from the bacterial surroundings through the cell membrane. The two methods used in this study to transfer plasmid DNA are TSS (transformation and storage solution) transformation and electroporation. To transfer purified DNA fragments, electroporation was used exclusively. Electroporation as described by Calvin and Hanawalt (1988) has the highest transformation efficiency of the two methods and was used in this project to transfer plasmid DNA generated by

ligation and purified DNA fragments generated by PCR. The DNA fragments generated by PCR were purified by running on electrophoresis gel and extracted using the QIAquick gel extraction kit from Qiagen. Electrocompetent cells were created following the protocol of Sambrook et al., (1989). TSS transformations were used to transform purified plasmids and were carried out by first inoculating sterile glass test tubes containing 3-5mL of LB with a single colony of the strain that was to be transformed. The culture was then incubated at 37°C on rollers until an OD₅₇₈ of approximately 0.8 was reached. 200µL of culture, 200µL of 2X TSS (20g PEG-600, 10mL 1M MgSO₄, 90mL LB medium, and 10mL of 100% DMSO) and 1-3µL of plasmid were aliquoted into 1.5mL Eppendorf reaction tubes and stored on ice for 30 minutes and then for 1 hour at either 37°C or 30°C (depending on the plasmid) with shaking of 650rpm. A control TSS transformation contained no plasmid in the mixture. Finally the culture was plated on selection plates and allowed to grow overnight at either 37°C or 28°C, depending on the plasmid.

3.4.7 P1 transduction

Transduction is the transfer of bacterial DNA via a bacteriophage. P1 transduction utilizes the P1 virion which is capable of transferring up to 15 kb of DNA and unlike Lambda, is not dependent on site-specific mechanisms for recombination. P1 transduction is a form of general transduction where the transducing particles are composed entirely of host DNA and completely lack any DNA originating from the viral vector. P1 lysogens are not readily induced, though there have been claims that induction may be possible with exposure to UV light (Nicole Sommerfeldt, unpublished data).

In this study, P1 transduction was used to transfer specific gene mutations or deletions generated by using the Datsenko and Wanner (2000) method. P1 lysates were generated from these strains by first inoculating sterile glass test tubes containing 3-5mL of LB with single colony of the strain that was to donate its DNA to be transferred to another strain. The culture was then incubated at 37°C on rollers until an OD₅₇₈ of approximately 0.2-0.3 was reached. 1-2 drops of 1M CaCl₂ were added along with 1-3 drops of the wild-type P1 vir lysate. To ensure sufficient cell lysis, the mixture was then incubated at 37°C on rollers for 3-12 hours. 5-10 drops of chloroform were then added in order to kill off residual cells. The mixture was allowed to incubate again on rollers at 37°C for 15-30 minutes before centrifugation was performed for 10 minutes at 6000rpm. The supernatant which contained the P1 phages with fragments of the donor strains' DNA, encompassing the entire chromosomal DNA packed into its heads, was then transferred to a sterile glass vial for storage at 4°C and 2 more drops of chloroform were added to the created lysate.

The P1 transduction was achieved by first inoculating glass test tubes containing 2-3mL of LB with a single colony of the strain that was to be the recipient of the donor DNA from the

generated P1 lysate. The culture was then incubated at 37°C on rollers until an OD₅₇₈ of approximately 0.5 was reached. 1-2 drops of 1M CaCl₂ were added and 500µL of the mixture was transferred to a sterile glass test tube. 3-5 drops of the generated P1 lysate were then added, the mixture vortexed and allowed to react for 15mins at room temperature. The reaction was terminated with the addition of 50µL sodium citrate; this binds the Ca⁺ ions that were previously mediating P1's adherence to the cell membrane and thus facilitated transduction. The mixture was incubated for 1 hour on rollers at 37°C before being plated on selective media plates containing sodium citrate. Thus, only the bacterial cells that received the mutated gene of interest along with the gene encoding antibiotic resistance was able to grow on the selective media plates that contained the given antibiotic.

3.5 Molecular biological and biochemical methods

3.5.1 Determination of DNA concentrations and sample purity

The spectrophotometer Nanodrop was used to measure DNA concentrations. The DNA sample to be measured was exposed to ultraviolet light at a wavelength of 260nm. Some of that light would be absorbed by the DNA molecule, resulting in less light that would reach the photodetector and a higher optical density. Using the intensities of the incident light and the transmitted light can be used to calculate the optical density = $\log\left(\frac{\text{Intensity of Incident Light}}{\text{Intensity of Transmitted Light}}\right)$. This value can be used in Beer-Lambert's law to calculate the concentration of the measured DNA sample. 1 A₂₆₀ unit dsDNA = 50µg/mL, 1 A₂₆₀ unit ssDNA = 37µg, and 1 A₂₆₀ unit ssRNA = 40µg.

To assess the purity of a sample, the samples optical density was measured at two different wavelengths and their ratio compared. A value of A₂₆₀/A₂₈₀ = 1.8 confirmed DNA purity while a value of A₂₆₀/A₂₈₀ = 2.0 confirmed RNA purity.

3.5.2 Polymerase chain reaction (PCR)

PCR was performed according to standard protocols (Sambrook *et al.*, 1989) using the DNA polymerase Opti-Taq (Roboklon) and the primers listed in Table 3.6. Primers were synthesized by Metabion. PCR products were analyzed on 1% agarose gels. PCR-fragments used for cloning into plasmids, one step inactivation (Wanner & Detsenko method) and DNA sequence analysis were purified from agarose gels using the QIAquick gel-extraction kit (Qiagen).

3.5.3 Preparation of chromosomal DNA and plasmid DNA

Chromosomal DNA was extracted using the QIAamp DNA Mini Kit from Qiagen. Plasmid DNA was extracted from bacterial cells using the InnuPREP Plasmid Mini Kit from Analytik Jena.

3.5.4 Agarose gel electrophoresis

In order to separate, purify, isolate or analyze DNA agarose gel, electrophoresis was employed. Samples of DNA were ran on 1% agarose gels (in 1X TAE buffer (40mM Tris (trishydroxyaminomethane), 1mM EDTA, 20mM acetic acid). The DNA samples were treated with 6X DNA sample buffer (0.25% bromphenol blue, 0.25% xylene cyanole, 30% glycerol) and loaded along with DNA size markers (100 bp extension DNA Ladder, λ DNA-BstEII Digest, New England Biolabs). Gels were run at 95-120 V and visualized by staining with either ethidium bromide (0.5 μ g/mL) or gelRed (0.75 μ g per 100mL of 1% agarose gel mixed directly into the gel).

3.5.5 Plasmid construction

All of the constructed plasmids used in this study are listed in Table 3.6.

3.5.5.1 Restriction digest and ligation

In order to construct the various plasmids used in this project, the plasmids and PCR amplified DNA fragments were first digested with restriction enzymes then purified and finally re-ligated. Digestion protocols recommended by the manufacturer (New England Biolabs) were adhered to. In some cases the digested plasmids were additionally incubated with calf intestinal phosphatase (CIP) (New England Biolabs) for 30min at 37°C, in order to prevent re-ligation of the plasmid. Following the instructions of the QIAquick PCR purification kit from Qiagen, the digested products were purified. Ligation of the digested DNA fragments and plasmid was induced by the addition of T4 DNA ligase (New England Biolabs) along with T4 buffer at a specific cycle of temperatures achieved using the PCR machine.

3.5.5.2 Electroporation and purification

Electroporation was used to transfer the created ligated plasmids into bacterial cells. Between 1-3 μ L of the ligated plasmids were mixed with 40-50 μ L of electrocompetent cells in an electroporation cuvette and an electro pulse of 50eV was applied to the sample. The mixture was then added to 500 μ L of medium (usually LB) and allowed to incubate for 1hr at 37°C while shaken for aeration before being plated on selective media and grown overnight at either 37°C or 28°C depending on the plasmid. The overnight colonies were then purified by single colony streaking out onto fresh selective media plates and allowed to grow again overnight at either 37°C or 28°C.

3.5.5.3 Sequence confirmation

In order to confirm correct plasmid construction, FLAG-tag construction, one-step inactivation, SCAR constructs, two-step integrations or the sequence anomalies found in genome databases the entire plasmid or PCR amplified DNA fragments were sent to GATC (Konstanz) for sequencing.

3.5.6 Construction of chromosomal *lacZ* fusions

Construction of *lacZ* fusions consisted first of plasmid construction (see section 3.5.5) of the promoter region of the gene of interest into the pJL28 plasmid (Lucht *et al.*, 1994). This resulted in a translational fusion that allowed for more accurate expression more indicative of the physiological expression of the protein. After sequence confirmation of the generated plasmid, the protocols established in Simons *et al.*, (1987) paper were followed in order to transfer the plasmid *lacZ* fusion into the bacterial chromosome. The phages λ RS45 or λ RS74 were employed to accomplish the *in vitro* recombination between the *bla* and *lacZ* genes encoded in the pJL28 plasmid and *bla'* and *lacZ'* genes encoded by the phage. The phage then integrated the genes at the lambda *attB* site in the bacterial chromosome. To insure a single chromosomal insertion, all of the generated fusions were tested by PCR using the procedure described by Powell *et al.*'s 1994 publication.

3.5.7 Construction of deletion/ insertion mutants

Non-polar in-frame mutations as well as deletions or mutations were generated using the Datsenko and Wanner method (Datsenko & Wanner, 2000), using the primers and plasmids listed in Tables 3.6 and 3.5, respectively, or via P1 transduction of mutated region of interest along with antibiotic resistance marker. Then SCARs were created by flipping out the antibiotic resistance marker as described in section 3.5.8.

3.5.8 Flipping out resistance cassettes

The pCP20 plasmid was TSS transformed into the *E. coli* strain in order to eliminate resistance marker from the strain. The transformed strain was plated on AMP plates and grown overnight (ON) at 28°C. Transformants were purified by first streaking them out on AMP plates and grown ON at 28°C. Then the transformants were streaked out on LBXGAL plates and grown ON at 43°C in order for them to lose the pCP20 plasmid. The transformants were once again streaked out on plates and grown ON at 37°C. For the last step three different plates were used in parallel: LBXGAL plates, LBXGAL AMP and LBXGAL plates containing an antibiotic that was attempted to be flipped out. The last step was performed in order to confirm that the antibiotic resistance cassette of interest was flipped out and that the pCP20 plasmid was also gone. The colonies that grew on the LBXGAL plates but not on either of the plates that contained the antibiotic were grown in LB broth ON and collected as detailed in section 3.4.2.

3.5.9 Complementation experiments

Complementation experiments can show whether a given protein can complement the activity of another protein or same protein when that protein is knocked out. ON cultures were made from each of the six strains and 70mL of sterile LB medium containing 300mL flasks were inoculated with the ON culture to an initial OD_{578nm} of 0.05 and grown in a 28°C water bath for 24 hours. The OD_{578nm} was recorded and 50µL samples of the growing bacterial cultures were taken starting 3 hours after inoculation, with an hourly frequency until 12 hours after inoculation and a last sample after 24 hours after the inoculation (ON). The samples were directly frozen at -20°C for use on the following day in determining β-galactosidase activity (see Material & Methods, section 3.4.9). 7 hours after inoculation the bacterial cultures were split equally into two 100mL flasks and one of the flasks was treated with IPTG in order to induce expression of the plasmid. Samples of the cultures with IPTG were taken starting 9 hours after inoculation at an OD_{578nm} of approximately 3.

This complementation protocol was followed using two different plasmids: pCAB18 (low-medium copy) and pQE60, a high copy plasmid which has a HIS-tag in its C-terminus.

3.5.10 Determination of β-galactosidase activity

The β-galactosidase activity of a *lacZ* fusion was assayed using *o*-nitrophenyl-β-D-galactopyranoside (ONPG) as a substrate according to Miller's (1972) method. The specific β-gal activity was calculated as µmol of *o*-nitrophenol per min per mg of cellular protein. Experiments showing the expression of *lacZ* fusions were performed at least in three replicas and displayed with standard deviation (SD) values from that dataset. When samples were collected along a growth curve then they were performed at least twice in a given condition though in many cases they were repeated more times.

3.5.11 RNA preparation for primer extension

RNA was extracted using the hot phenol/chloroform extraction method. 5mL of bacterial culture containing the plasmid of interest was harvested, mixed with 630µL of ice-cold RNA-stop solution (5% phenol in ethanol) and incubated on ice for 5 minutes. The mixture was centrifuged on max speed at 4°C, the supernatant was discarded and the pellet was stored at -80°C. The method described by Tani *et al.*, (2002) was used for cell lysis and nucleic acid extraction. 10µL of RNase-free DNaseI was used to digest the DNA. The procedure was performed at 37°C for 20 minutes. RNA was extracted in succession in 1 volume of Aqua-PCI (phenol:chloroform:isoamyl alcohol, 25:24:1, pH 4.5-5) and 1 volume of CI (chloroform:isoamyl alcohol, 24:1). This was followed by adding 0.1 volumes of 3M sodium acetate (pH 5.2), 3 volumes of ice-cold 100% ethanol and allowing precipitation to occur at -80°C for at least 30 minutes. The RNA was then

pelleted at maximum speed by centrifugation at 4°C and the pellet was washed with 80% ice-cold ethanol, pelleted again and air-dried at room temperature. The RNA was then re-suspended in 50µL of RNase-free water. RNA purity and concentration was calculated using the methods described in section 3.5.1.

3.5.12 Primer extension analysis

Primer extension analysis was performed by Alexandra Poßling. The total RNA was prepared as outlined in section 3.5.11 from W3110 derivative carrying the pJL28 plasmid grown overnight at 37°C in LB medium.

Primers listed in table 3.7 were labelled with γ -³²P-ATP by mixing together 2µL of the primer (10pmol/µL), 1µL of T4 polynucleotide kinase (PNK, 10u/µL, Fermentas), 1.5µL of T4 PNK-buffer (Fermentas), 2µL of γ -³²P-ATP (Hartmann Analytic) and 8.5µL of water. The mixture was allowed to incubate at 37°C for 60 minutes. Then the T4 PNK was inactivated by heat-shocking the mixture for 3 minutes at 90°C. The labelled primer was purified out according to the manufacturer's instructions by employing a Sephadex G-25 column (GE Healthcare).

2µL of the labelled primers were incubated at 65°C for 5 minutes along with 10µg of RNA in 11µL of water and 1µL of RNase-free dNTPs mix (10mM, Invitrogen). This mixture was then incubated on ice for 5 minutes, after which 1µL of DDT (0.1M, provided with Superscript III reverse transcriptase) was added to the test tube and the whole mixture was incubated for an additional hour at 45°C. The enzyme was inactivated by heat shock for 15 minutes at 70°C.

Using the CycleReader DNA Sequencing Kit from Fermentas, a DNA sequence ladder was generated to be used as a marker. 1.5µL of the previously generated labelled primers and 100fmol of a template DNA fragment were employed according to the manufacturer's instructions to make the sequencing reaction.

The solution consisting of 2µL of the primer extension reaction, 5µL of Stop-Solution from the CycleReader DNA Sequencing Kit and 3µL of each sequencing reaction mixed with 4µL of 1:4-diluted Stop-Solution, was heated for 3 minutes at 90°C before being loaded onto a denaturing 7M urea, 6% polyacrylamide sequencing gel (42g urea, 10mL 10x TBE, 15mL Rotiphorese Gel40, 50µL TEMED, 500µL 10% APS in a total volume of 100mL DI water). The gel was pre-warmed by being previously run at 60W for approximately 60 minutes. The loaded gel was then run for 2 hours at 60W and then was vacuum dried for 90 minutes before being autoradiographed using the FLA-2000G Imager from Fujifilm.

3.5.13 Extraction of proteins from bacterial culture samples

Extraction of proteins from bacterial culture samples was carried out by collecting the calculated amount of culture and centrifuging it on high for 10 minutes at approximately 13000rpm. The supernatant was sucked out and the pellet was re-suspended in 60 μ L of 1X SDS sample buffer (0.06M Tris pH 6.8, 2% SDS, 10% glycerol, 3% β -mercaptoethanol, 0.005% bromphenol blue). The samples were cooked for 15 minutes at 100 $^{\circ}$ C and if the proteins contained transmembrane segments, they were cooked first for 10 minutes at 70 $^{\circ}$ C. The amount of culture collected was calculated by measuring the OD_{578nm} and using the relationship discovered by Miller (Miller, 1972) that 1mL of a cell suspension at an OD_{578nm} of 1 contains approximately 107 μ g of protein. The amount collected was always 60 μ g of protein to which 60 μ L of 1X SDS buffer was added. The volume collected was calculated using the relationship

$$\text{Volume} = \frac{60\mu\text{g} \cdot 1\frac{\mu\text{g}}{\mu\text{L}}}{107\mu\text{g} \cdot \text{OD}_{578\text{nm}}}$$

3.5.14 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used in order to separate proteins according to their electrophoretic mobility. In order to linearize the protein samples, the anionic detergent sodium dodecyl sulfate (SDS) was applied. For the purposes of the experiments carried out in this project, it was important to linearize the protein molecules so that an accurate comparison could be made between proteins where their mobility along the polyacrylamide gel would be solely dependent on the length of the protein molecule and its mass-to-charge ratio. Otherwise it would be difficult to positively identify a protein according to its running size as compared to a standard marker. The SDS also binds to the protein's polypeptide chain imbuing it with an even distribution of charge per unit mass, which in turn allows the protein to move through the gel while a charge is applied.

The standard SDS protocols as first outlined in Laemmli's publication (1970) and then later modified in Sambrook *et al.*'s publication (1989) were adhered to. The polyacrylamide gel was comprised of two levels, an upper and a lower level. The lower level was 12% polyacrylamide gel and contained 2.5mL of LT buffer (36.34g Tris, 0.8g SDS in 200mL water; pH 8.8), 3.45mL DI water, 4mL Rotiphorese Gel 30 from Roth, 5 μ L TEMED and 50 μ L of 10% APS. The upper level was 4% polyacrylamide gel and consisted of 1.25mL UT buffer (6.06g Tris pH6.8, 0.8g SDS in 100mL water), 3.07mL water, 0.65mL Rotiphorese Gel 30 from Roth, 5 μ L TEMED and 50 μ L of 10% APS. The gels were run at 25-30mA per gel in 1X SDS-PAGE buffer (25mM Tris, 0.19M glycerol, 0.1% SDS) for 45-60mins. To visualize, the gels were stained with PageBlue protein staining solution (Pierce).

3.5.15 PageBlue Protein Staining

PageBlue protein staining solution (Pierce) uses the Coomassie G-250 dye for staining on polyacrylamide gels. The solution is sensitive to a protein range from 5-500ng and is ~10 times more sensitive than Coomassie Brilliant Blue R-250 based dyes.

The gel was first washed three times with DI water with gentle agitation for 5 minutes after being microwaved on high power for 1 minute. The gel was then bathed in the PageBlue staining solution, just enough to cover the gel, again with gentle agitation for 20 minutes. The gel and solution were first microwaved at full power for 20-30 sec so that the reagents did not boil. The staining solution could be reused at least 3 times. Finally the gel was rinsed with DI water and then washed with DI water for 5 minutes with gentle agitation, at which point the protein bands on the gel became clearly visible.

3.5.16 Immunoblot analysis (Western blot)

Western blotting was used to detect specific proteins. These proteins were tagged by specific markers, either HIS or FLAG, to which specific antibodies would bind to. The antibodies specific for a given marker were generated in either mouse or rabbit and a secondary antibody specific to either mouse or rabbit (the host in which the primary antibody was generated in) was applied. Western blotting first utilized the SDS-PAGE method for protein separation. The gel was then transferred to a polyvinylidene fluoride (PVDF) membrane (Sigma-Aldrich) via a semi-dry blotting transfer. The semi-dry blotter from Peqlab was used and the manufacturer's instructions were adhered to. The transfer was performed for 60 minutes at a constant conduction current of 150mA for a small gel of 100cm² surface area. The blotted PVDF membrane was then blocked with TBSTM [TBST buffer (20mM Tris at pH 7.5, 150mM NaCl, 0.05% Tween-20) with 5% milk powder] for a minimum of 30 minutes.

The primary antibody was added to fresh TBSTM and the blocking was usually conducted overnight. 2 μ L of antibody in 20mL TBSTM for the anti-HIS antibody (produced in rabbits) incubated with the membrane for 2 hours with gentle agitation. The washing procedure was performed 3 times for 10 minutes each with TBST buffer and gentle agitation. The secondary antibody was added to fresh TBSTM and the membrane was allowed to soak in the mixture for 1 hour with gentle agitation. 3 μ L of secondary antibody (goat-anti-rabbit-alkaline phosphatase conjugate, Sigma-Aldrich) was used in 20mL TBSTM. A final washing step with TBST was performed for 10 minutes with gentle agitation followed by an equilibration step with alkaline phosphatase (AP) buffer (100mM Tris pH 9.5, 100mM NaCl, 5mM MgCl₂) incubated with gentle agitation for 10 minutes. Finally, the proteins were visualized by the application of 33 μ L BCIP (50mg/mL in dimethylformamide) and 66 μ L NBT (50mg/mL in 70% dimethylformamide) in 10mL AP buffer.

The visualization process of the proteins works on the principle that the secondary antibody, an alkaline phosphatase, converts a soluble chromogenic substrate, in our case BCIP and NBT, to a colored, insoluble product that precipitates onto the membrane. Since the alkaline phosphatase is bound to the primary antibody, which in turn is bound to the tag on our protein of interest, the precipitate only forms at the location of the protein and the final product is colored bands. The intensity of these bands correlates to the amount of protein present. The development was then stopped by washing away the soluble substrate. The developed membrane was then scanned for storage and in some cases, where protein quantification was required, it was analyzed with ImageGauge V3.45. The unknown quantity of the studied protein was compared to specific markers and its quantity calculated.

3.5.17 Protein over-expression and purification

Constructed plasmids (AMP resistant) containing the protein of interest were TSS transformed into the GB1000 strain. ON cultures were produced by inoculating transformed colonies in 5mL of LB and 7.5 μ L of 100 μ g/mL AMP. 5mL of the ON culture was then inoculated into a 1.5L flask containing 300mL of LB medium and 300 μ L of 100 μ g/mL AMP. The mixture was grown at 37°C for approximately 90 minutes. At an approximate OD₅₇₈ of 0.8 the plasmids were induced using 300 μ L of 100 μ g/mL sterilized IPTG and grown overnight at 16°C. Protein samples of the over-expression culture were taken before the induction step and after the ON step. The thawed cells were handled on ice or in a 4°C room during the entire extraction process.

The over-expression culture was then centrifuged for 12 minutes at 6000rpm at 4°C. The supernatant was discarded and the pelleted cells were stored at -80°C for later purification. The previously collected protein samples were analyzed on SDS-PAGE gel to confirm over-expression and only in cases where the difference in protein-band intensity was not distinct enough or the protein's running size did not coincide with the expected value was a western blot performed to confirm over-expression. After the confirmation of protein over-expression, the stored harvested cells were treated with Ni-NTA Wash buffer (50mL 1M Tris pH 7.5, 5mL 2M MgCl₂, 30mL 5M NaCl, 10mL 2M Imidazole, 905mL VE water) mixed with protease inhibitor tablets (1 tablet/10mL [add 2 tablets for 15mL]). The mixture was thawed on ice for 60 minutes. The cells were then re-suspended and lysed with a French Press. Centrifuging the lysed cells for 40mins at 17000rpm enabled the removal of cell debris and the retention of the over-expressed proteins located within the supernatant. In parallel to the centrifugation step, 4mL of chitin beads were calibrated in Lysis buffer (50mL 1M Tris pH 7.5, 5mL 2M MgCl₂, 60mL 5M NaCl, 5mL 2M Imidazole, 880mL VE water). The calibration/lysis buffer was then discarded and the chitin beads were incubated with the supernatant containing the over-expressed proteins on rollers at 4°C from 2 to 10 hours.

The supernatant of overexpressed proteins along with the chitin beads was washed with 200-300mL of Ni-NTA Wash buffer. After washing, the optical density at 280nm of the eluted wash was measured. If the OD was less than 0.03, the washing was stopped; otherwise it was continued until the wash elutant reached an OD₂₈₀ of 0.03 or less. The proteins were eluted by treating the column with 3-5mL of 1X Elution buffer (5mL 1M Tris pH 7.5, 500μL 2M MgCl₂, 6mL 5M NaCl, 15mL 2M Imidazole, 73.5mL VE water) and letting the mixture incubate for 30 minutes. The elution was then collected in 2mL Eppendorf reaction tubes. An SDS-PAGE was performed to confirm the purity of the eluted target proteins. The samples were collected along the protocol at the following steps: before induction, ON after induction, centrifuged supernatant of lysed cells, pellet, wash and each of the elution fractions from the 2mL collection Eppendorf reaction tubes. The elution fraction containing only the purified protein and no debris was then dialyzed ON into storage buffer (500mM NaCl, 20 mM Tris (pH 8), 0.2mM DTT). The samples were then collected in 1.5mL Eppendorf reaction tubes and stored at -80°C.

3.5.18 Determination of protein concentrations

The protein concentration of the purified protein was measured with the NanoDrop spectrophotometer. The ultraviolet wavelength of 280nm was used to measure the absorbance of the purified protein. Setting 1Abs = 1 mg/mL along with the proteins' size in kDa and specific extinction coefficient were used by the NanoDrop software to calculate the concentration using Beer-Lambert's law $A = \epsilon \cdot b \cdot c$, where A is the absorbance value, ϵ is the wavelength-dependent molar absorptivity coefficient or extinction coefficient in L/mol · cm, b is the path length in cm, and c is the analyte concentration in mol/L. The buffer in which the purified protein was stored was used as a blank.

3.5.19 *In vivo* protein-protein interaction analysis

The BacterioMatch II two-hybrid system from Agilent Technologies was used to detect *in vivo* protein-protein interactions. The system detects protein-protein interactions via the transcriptional activation of the *HIS3* reporter gene. One of the proteins to be tested (the bait) is cloned into the pBT vector to construct a fusion to the full-length bacteriophage λ repressor. The other protein to be tested (target protein) is cloned into the pTRG vector to construct a fusion to the N-terminal domain of the A-subunit of RNA polymerase. The bait is attached to the λ operator sequence upstream of the reporter promoter through the DNA-binding domain of λ cI. If the two tested proteins interact with each other, they enable the binding of the RNA polymerase at the promoter of the *HIS3* reporter gene, thereby enabling its subsequent expression. The expression of the His3 enzyme allows for the reporter strain to grow on minimal media. The reporter strain has a mutation in the *hisB* gene which prevents its own synthesis of histidine and thus is unable to grow on minimal media, but the expression of *HIS3* successfully complements this mutation. In the

absence of transcriptional activation by the RNA polymerase, the *HIS3* gene is still produced in minimal levels from the reporter gene cassette. In order to circumvent this, the 3-amino-1,2,4-triazole (3-AT) is added to minimal media lacking histidine. 3-AT is a competitive inhibitor of the His3 enzyme. On this media, only the constructs with protein-protein interaction and subsequent transcriptional activation of *HIS3* will be able to grow.

This system offers several advantages to the traditional yeast based system. *E. coli* has both a higher transformation efficiency and faster growth rate as compared to yeast. The overall system's sensitivity can be fine-tuned by modifying the amount of 3-AT added to the media.

The bacterial two-hybrid assay was performed five times for each tested condition according to the instructions of the manufacturer with the following adjustments: the final concentration of the selective screening media was 3.5mN 3-AT, 2-3 μ L of the respective plasmid was co-transformed with 50 μ L aliquots of competent cells of the reporter strain and the media and chemicals indicated in BacterioMatch II's protocol were scaled down to match the aliquot size, except for the amount of M9⁺ His-dropout incubation medium remained 1mL.

Special plating conditions were carried out, depending on which conditions were to be tested. When co-transforming with an empty pBT or pTRG vector, 400 μ L of 1:100 dilution in M9⁺ His-dropout medium of the co-transformant mixture was plated on non-selective media plates. On the selective media plates 400 μ L of undiluted mixture was plated. When co-transforming with a vector containing a gene of interest, 200 μ L of undiluted co-transformation mixture was plated on screening plates.

The co-transformants were grown overnight at 37°C on selective and non-selective media plates. In order to have a direct comparison of the bacterial growth of the different vector combinations a patching step was introduced, where all of the different combinations tested were patched onto a single plate. 5 of the grown co-transformants of each condition were then patched onto selective media followed by non-selective media large plates which were incubated overnight at 37°C or for 3-4 days at 28°C.

3.6 Databases and bioinformatic analyses

Table 3.8: Programs and databases used for the bioinformatic analysis and addresses of websites where the programs or databases can be accessed or downloaded from.

| Program/Database Name | Website address |
|---|---|
| BLAST (Basic Local Alignment Search Tool) | http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi |
| Clustal X | http://www.clustal.org/clustal2/ |

| Program/Database Name | Website address |
|---|---|
| Ecocyc | http://www.ecocyc.org |
| Geneious (purchased) | http://www.geneious.com |
| HMMTOP | http://www.enzim.hu/hmmtop/ |
| The MEME Suite & MAST | http://meme.nbcr.net/meme/ |
| NCBI (National Center for Biotechnology Information) | http://www.ncbi.nlm.nih.gov |
| PATRIC (The Pathosystems Resource Integration Center) | http://www.patricbrc.org |
| SMART EMBL | http://smart.embl-heidelberg.de |
| TMHMM | http://www.cbs.dtu.dk/services/TMHMM/ |
| TreeView | http://www.treeview.net/tv/download.asp |

3.6.1 Strains and genome sequences

This study started with a systematic analysis of all of the 39 completed *E. coli* genomes available on National Center for Biotechnology Information (NCBI) database was performed in May 2011. In May 2012 an updated search was performed and incorporated an additional 9 genomes bringing the total to 48 genomes analyzed. In June 2014 an additional 13 genomes were selectively added as there were too many new completed *E. coli* genomes added to the NCBI database to add to this study. The 13 selected genomes were added to complement the previous strains, to fill out the different pathogenic groups and if possible to provide additional strains to groups which previously only had one representative. The grand total of strains included in this study was 61. The strains analyzed included pathogens from the following groups: EAEC, ETEC, EHEC, EPEC, ExPEC, UPEC, APEC, AIEC, MNEC; as well as commensal strains. The *E. coli* K-12 W3110 was used as a reference strain in this study, as it is one of the most popular experimental models for c-di-GMP signaling/regulation in *E. coli*. All of the strains analyzed in this study can be found in Table 4.2.

The Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1996) searched for proteins within a single selected genome of *E. coli*, at a time, for GGDEF and EAL domains. For the GGDEF domain search, the 169 AA long GGDEF domain of the YdaM protein (as identified by the SMART EMBL database) was used as the query sequence. For the EAL domain search, the 245 AA long EAL domain of the YjcC protein (as identified by the SMART EMBL database) was used as the query sequence. Using this approach previously unidentified GGDEF and EAL domain proteins could be located within a given genome. If a protein was absent from the genome

of a strain a second BLAST search was performed using the missing proteins analog from the W3110 strain as the query sequence. This approach yielded discrepancies in the length of the proteins. To verify whether these differences were genuine reflections of differing topologies or arbitrary differences in the annotation of the strain a follow-up tBLAST search was performed, where the amino acid sequence was searched directly within the nucleotide database by the tBLAST algorithm translating the nucleotide sequence into an amino acid sequence, again using proteins from the W3110 strain as the query sequence. In this way further proteins that were annotated to be 'missing' were found and others were confirmed to be truly absent from the genome. Discrepancies in the topology of the various proteins between strains were further addressed by looking at the nucleotide sequence, at the annotated start codon and the upstream region of the Shine-Dalgarno sequence. It proved often to be the case that a protein's nucleotide sequence was nearly identical and that a different start codon was annotated, where the differences in sequence did not result in early termination of the protein. Deleterious disruptions of a protein, such as insertion, deletions or point mutations were also checked by looking at the nucleotide sequence and this sequence was then compared to a wild-type nucleotide sequence of the specific gene. When a gene was missing altogether from the genome, it was further confirmed by searching in directly the nucleotide database using nucleotide blast function and the nucleotide sequence from the wild-type version of the gene as the query sequence. Table 4.2 presents the cumulative results with an 'X' marking the presence of a protein within the strain.

If novel proteins were identified the nomenclature of naming them was based on their potential functions determined by which domain they contained, either a GGDEF or an EAL. So that a protein with an intact GGDEF domain (based on the conservation of key amino acids in the domain that predict domain activity) was named Dgc for its putative diguanylate cyclase activity and a protein with an intact EAL domain (based on the conservation of key amino acids in the domain that predict domain activity) would be named Pde for its putative phosphodiesterases activity, plus some additional letter that was not already taken by a previously identified DGC or PDE.

3.6.2 Phylogenetic, hydropathy and sequence analyses

The alignments of two protein or DNA sequences were performed by the BLAST tool located on the NCBI website. Multiple alignments (greater than two sequences compared at a time) were performed separately (using default parameters) on all of the GGDEF domain-containing proteins and all of the EAL domain-containing proteins found within the *E. coli* strains and phylogenetic trees were generated using the ClustalX program (Thompson *et al.*, 1997). The phylogenetic trees were visualized using TreeView (Zhai & Saier 2002; Zhai *et al.*, 2002). Topological analyses of

the individual proteins were performed by the TMHMM (Krogh *et al.*, 2001) and HMMTOP (Tusnady *et al.*, 2001) programs.

3.6.3 Motif analysis

All of the GGDEF domain-containing proteins and all of the EAL domain-containing proteins were analyzed for potential uncharacterized motifs using the MEME (Bailey & Elkan, 1995) program. Default settings were used, except that the condition “any number of repetitions” was selected for the prediction of how single motifs were distributed among the sequences. The locations of the motifs were determined for individual proteins relative to the locations of the TMSs using the hydropathy plots generated by the TMHMM program.

4 Results

4.1 Bioinformatic Analysis – variations in GGDEF/EAL genes in 61 *E. coli* genomes

Bacteria possess a complement of GGDEF/EAL domain-containing proteins and in the commonly used K-12 laboratory strain *E. coli* W3110, the full complement is 29 of these proteins. 12 encode GGDEF domain proteins, 10 encode EAL domain proteins and 7 encode GGDEF-EAL domain composite proteins (Hengge, 2009). Based on biochemical evidence and amino acid sequence conservation (see Table 4.1), 12 of these proteins are putative DGCs, 13 are putative PDEs and four are degenerate GGDEF/EAL domain-containing proteins. However, little is known about how that list may differ between strains of *E. coli*, particularly comparing pathogenic and commensal strains. Which genes are most conserved and which ones have the greatest fluctuation? To answer these questions a systematic analysis was performed on all complete genomes available on the NCBI database (Materials & Methods 3.5). The initial analysis in May 2011 included 39 strains, an updated analysis was performed in May 2012 and added an additional 9 genomes. Finally, in June 2014 a final update was performed and this time 13 additional strains were selectively added as there were too many new completed *E. coli* genomes added to the NCBI database to add to this study. The 13 selected genomes were added to complement the previous strains, to fill out the different pathogenic groups and if possible to provide additional strains to groups which previously only had one representative. Among the 61 genomes both commensal and pathogenic strains were included and pathogens contained the following groups: enteroaggregative *E. coli* (EAEC), enterotoxigenic *E. coli* (ETEC), enterohemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), extraintestinal pathogenic *E. coli* (ExPEC), enterotoxigenic *E. coli* (ETEC), uropathogenic *E. coli* (UPEC), avian pathogenic *E. coli* (APEC), adherent and invasive *E. coli* (AIEC) and meningitis-associated *E. coli* (MNEC) (Table 4.2). By meticulously analyzing the genomes of 61 *E. coli* strains, trends were uncovered that are discussed in detail in the following sections.

The 2011 outbreak in Germany of the Shiga toxin (Stx)-producing *Escherichia coli* O104:H4 had a high incidence of HUS (more than 20% of patients). The outbreak strain is genetically most similar to an EAEC but has acquired a Stx carrying phage from EHEC (Mellmann *et al.*, 2011) and contains a novel gene *dgcX*. This prompted the further investigation of biofilm forming properties of the outbreak strain and four other strains (55989, HUSEC041, EDL933 and K-12 strain W3110), as well as exploring the function of *dgcX* and its relationship to virulence in EAECs of the O104:H4 serotype. The classical EAEC 55989 was chosen to be included in the laboratory experiments because the outbreak strain shares 98% of its genome with it (Frank *et al.*, 2011). HUSEC041 (01-09591) was included because it is a historic isolate from 2001 closely related to the outbreak strain LB226692. The classical EHEC strain EDL933 was included

Results

because both the outbreak strain (LB226692) and HUSEC041 harbor a Stx2 containing prophage inserted into the *wrbA* gene, the same integration site where the phage is found in EDL933. The EAEC 55989 strain has an intact *wrbA* gene not occupied by a prophage. The K-12 laboratory strain W3110 was included as a reference commensal strain that is not a pathogen.

Results

Table 4.1: The 35 GGDEF/EAL domain proteins found within 61 strains of *E. coli*. In pink are the putative diguanylate cyclases (DGCs), in blue are the putative phosphodiesterases (PDEs) and in green are the degenerate GGDEF/EAL domain-containing proteins. Highly conserved residues that are required for enzymatic function of the given domain are printed bold.

| | Function | Number of TMSs | Sensory Domains | GGDEF Domain | | EAL Domain | | | | |
|-----------------------------|--|----------------|---|------------------------|---------------------|----------------|------------------|------------------------|--------------------------|--|
| | | | | I-Site RxxD | A-Site GG[D/E]EF | EALxR | C-di-GMP binding | Catalysis | Mg ²⁺ binding | |
| Putative DGC | | | | | | | | | | |
| YaiC (AdrA) | Cellulose synthesis | 6 | MASE2 | RGSDVIGRFGGDEF | | | | | | |
| YcdT | Unknown | 8 | | RPDDLARVGGEEF | | | | | | |
| YdaM | Activates CsgD transcription | 0 | PAS, sensory box AtoS, PAS ₂ | RKGDVLFVRWGGEF | | | | | | |
| YddV (DosC) | PGA regulation | 0 | Sensor globin | RSSDYVFRYGGDEF | | | | | | |
| YdeH | Unknown | 0 | CZB | RDYETVYRYGGEEF | | | | | | |
| YeaJ | Downregulates motility | 2 | | RKSDYAIRLGGDEF | | | | | | |
| YeaP | Indirect biofilm regulation | 0 | GAF | QNGEVIGRLGGDEF | | | | | | |
| YedQ | Cellulose synthesis | 2 | | RAQDVAGRVGGEEF | | | | | | |
| YegE | Downregulates motility, activates CsgD transcription | 11 | MASE1, PAS, PAC-PAS, | RSSDVLRARLGGDEF | EARNL | -/-/- | -/- | | | |
| YfiN | Unknown | 2 | | GLRHKAYRLGGDEF | | | | | | |
| YliF | Unknown | 2 | | VDKGKVYRFGGDEF | | | | | | |
| YneF | Unknown | 8 | | GDKGLVARMGGEEF | | | | | | |
| DgcX | Unknown | 8 | | RKEDILGRLGGEEF | | | | | | |
| DgcY | Unknown | 6 | | QSEDVVVRYGGEEF | | | | | | |
| Putative PDE | | | | | | | | | | |
| Rtn | Unknown | 2 | CSS motif | | EVLLR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YahA | Unknown | 0 | LuxR-C-like DNA binding | | EALVR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YcgG | Unknown | 2 | CSS motif | | EVLAR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YciR | Inhibits YdaM | 0 | PAS, RNase II stability modulator | EHDQVLARPGGDEF | EALVR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YddU (DosP) | | 0 | PAS, PAS | KPDQYLCRIEGTQF | EALAR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YfeA (YfdA) | Unknown | 8 | MASE1, | QENEKLYQLPGSEL | EILAR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YfgF | Unknown | 9 | MASE1, | EPGEDVYQLSGNDL | EILLR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YhjH | Motility | 0 | | | ELLTV | Q/V/D/A | L/E | E/N/P/R/D/K/E/Q | | |
| YhjK | Unknown | 2 | HAMP | SPRMILAQISGYDF | EVLLR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YjcC | Unknown | 2 | CSS motif | | EALLR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YlaB | Unknown | 2 | CSS motif | | EALAR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YliE | Unknown | 2 | | | EALCR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| YoaD (AdrB) | Unknown | 2 | CSS motif | | EILLR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| PdeT (VmpA) | Motility | 2 | CSS motif | | EALMR | Q/R/D/D | T/S | E/N/E/E/D/K/E/Q | | |
| PdeX | Unknown | 0 | | | ELLVR | Q/R/D/D | T/E | E/N/E/E/D/K/E/Q | | |
| PdeY | Unknown | 0 | | | EMLSD | E/D/D/S | S/E | E/N/E/E/D/K/E/Q | | |
| PdeZ | Unknown | 0 | | | ELLSR | E/R/D/D | S/E | E/L/E/E/D/K/E/Q | | |
| Degenerate GGDEF/EAL | | | | | | | | | | |
| BluF (YcgF) | Anti-repressor | 0 | BluF | | EAIVQ | N/Q/H/S | T/A | E/N/E/E/D/K/M/Q | | |
| YeaI | Unknown | 8 | | RPDDLARLEGEVF | | | | | | |
| CsrD (YhdA) | Turnover of RNAs CsrB/CsrC | 2 | | YPGALLARYHRSDF | ELMCR | Q/R/Q/H | A/R | E/E/E/E/N/K/E/Q | | |
| YdiV | Proteolytic targeting of FlhDC, but not well expressed in <i>E. coli</i> | 0 | | | EVLR | F/H/N/D | N/D | E/-/L/E/G/M/G/Q | | |

Results

Table 4.2: 61 *E. coli* strains analyzed in this work and their complements of DGC, PDE and biofilm matrix component proteins. "X" marks the presence of a gene. "2 parts" denotes genes that have a point mutation and then have an immediate in-frame start codon and thus may be expressed in two parts. "Pt. mut." denotes disruption of a gene as a result of a point mutation that prematurely introduced a Stop. "ins." denotes a gene disruption as a result of an insertion of 1 or more nucleotides. "Δ" denotes a gene disruption as a result of a deletion of 1 or more nucleotides. "IS" denotes a gene disruption as a result of an insertion of an insertion element. "Δ5'" denotes a large deletion in the 5' region of the gene and "Δ3'" denotes a large deletion in the 3' region of the gene. "frag." denotes that only a fragment of the gene can be found and is clarified in section 4.1.1 (*yddV*). "abn." denotes the presence of the upstream of *yahA* insertion of *aidA-like adhesin protein* gene. A blank slot represents the gene is missing in its entirety from the genome. *E. coli* are grouped according to pathogenic groups. Genes involved in biofilm formation are first, followed by genes encoding putative DGCs, followed by genes encoding putative PDEs, finally coming to the genes encoding degenerate GGDEF/EAL domain proteins. Thick lines denote separation in the gene section.

| Strain | Phylogroup | <i>csgB</i> | <i>csgA</i> | <i>csgC</i> | <i>csgD</i> | <i>csgE</i> | <i>csgF</i> | <i>csgG</i> | <i>mlrA</i> | <i>bcsE</i> | <i>bcsF</i> | <i>bcsG</i> |
|---|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| O42 (EAEC) | D | X | X | X | X | X | X | X | X | X | X | X |
| 55989 (EAEC) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| O104:H4 str. 01-09591 (HUSEC041) (EAEC) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| O104:H4 str. LB226692 (EAEC) | B1 | X | X | X | X | X | X | X | X | ins. | X | X |
| O104:H4 str. 2009EL-2050 (EAEC) | B1 | X | X | X | X | X | X | X | X | ins. | X | X |
| O104:H4 str. 2009EL-2071 (EAEC) | B1 | X | X | X | X | X | X | X | X | ins. | X | X |
| O104:H4 str. 2011C-3493 (EAEC) | B1 | X | X | X | X | X | X | X | X | ins. | X | X |
| E24377A (ETEC) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| ETEC H10407 (ETEC) | A | X | X | X | X | X | X | X | X | X | X | X |
| UMNF18 (porcine ETEC) | A | X | X | X | X | X | X | X | X | X | X | X |
| UMNK88 (porcine ETEC) | A | X | X | X | X | X | X | X | X | X | X | X |
| O103:H2 str. 12009 (EHEC) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| O111:H- str. 11128 (EHEC) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| O157:H7 str. EC4115 (EHEC) | E | X | X | X | X | X | X | X | phage | X | X | X |
| O157:H7 str. EDL933 (EHEC) | E | X | X | X | X | X | X | X | phage | X | X | X |
| O157:H7 str. Sakai (EHEC) | E | X | X | X | X | X | X | X | phage | X | X | X |
| O157:H7 str. TW14359 (EHEC) | E | X | X | X | X | X | X | X | phage | X | X | X |
| Xuzhou21 (EHEC) (O157:H7) | | X | X | X | X | X | X | X | phage | X | X | X |
| O26:H11 str. 11368 (EHEC) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| O145:H28 str. RM13516 (STEC) | | X | X | ins. | X | X | X | X | phage | $\Delta 5'$ | X | X |
| O145:H28 str. RM13514 (STEC) | | X | X | X | X | X | X | X | phage | $\Delta 5'$ | X | X |
| O145:H28 str. RM12761 (STEC) | | X | X | ins. | X | X | X | X | phage | $\Delta 5'$ | X | X |
| O145:H28 str. RM12581 (STEC) | | X | X | X | X | X | X | X | phage | $\Delta 5'$ | X | X |
| JJ1886 (Subclone of ST131) | | X | X | X | X | X | X | X | X | X | X | X |
| O127:H6 str. E2348/69 (EPEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| O55:H7 str. CB9615 (EPEC) | E | X | X | X | X | X | X | X | X | X | X | X |
| O55:H7 str. RM12579 (EPEC) | E | X | X | X | X | X | X | X | phage | X | X | X |
| O7:K1 str. CE10 (ExPEC, NMEC) | D | X | X | X | X | X | X | X | phage | X | X | IS |
| IHE3034 (ExPEC, MNEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| IAI39 (ExPEC) (O7:K1) | F | X | X | X | X | X | X | X | phage | X | X | IS |
| PMV-1 (ExPEC) | | X | X | X | X | X | X | X | X | X | X | X |
| S88 (ExPEC) (O45:K1:H7) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| UMN026 (ExPEC) (AR) (O7:K1) | D | X | X | X | X | X | X | X | X | X | X | X |
| 536 (UPEC) (O6:K15:H31) | B2 | X | X | X | X | X | X | ins | X | X | X | X |
| CFT073 (UPEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| UM146 (UPEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| UT189 (UPEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| LF82 (AIEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| O83:H1 str. NRG 857C (AIEC) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| APEC O1 (APEC) (O1:K1:H7) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| APEC O78 (APEC) | | X | X | X | X | X | X | X | X | X | X | X |
| SMS-3-5 (Environmental isolate) | F | X | X | X | X | X | X | X | X | X | X | X |
| ABU 83972 (ABU) (C) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| B str. REL606 (C) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| BL21(DE3) (C) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| 'BL21-Gold(DE3)pLysS AG' (NP) | A | X | X | X | X | X | X | X | X | X | X | X |
| BW2952 (C) (K-12 derivative) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| DH1 (C) (K-12 derivative) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| ED1a (C) (O81) | B2 | X | X | X | X | X | X | X | phage | X | X | X |
| HS (C) (O9) | A | X | X | X | X | X | X | X | ins. | | | $\Delta 5'$ |
| IAI1 (C) (O8) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| KO11FL (C) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| Nissle 1917 (C) | | X | X | X | X | X | X | X | X | X | X | X |
| SE11 (C) (O152:H28) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| SE15 (C) (O150:H5) | B2 | X | X | X | X | X | X | X | X | X | X | X |
| W (C) (Lab) | B1 | X | X | X | X | X | X | X | X | X | X | X |
| str. K-12 substr. DH10B (C) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| str. K-12 substr. MDS42 (C) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| str. K-12 substr. MG1655 (C) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |
| str. K-12 substr. W3110 (C) (Lab) | A | X | X | X | X | X | X | X | X | X | X | X |

csgB *csgA* *csgC* *csgD* *csgE* *csgF* *csgG* *mlrA* *bcsE* *bcsF* *bcsG*

| <i>yeaP</i> | <i>ycdT</i> | <i>yddV</i> | <i>yedQ</i> | <i>yegE</i> | <i>yneF</i> | <i>novel DGC</i> | <i>rtn</i> | <i>yhjH</i> | <i>yfgF</i> | <i>yhjK</i> | <i>ylaB</i> | <i>yciR</i> | <i>yfeA/yfdA</i> | <i>yliE</i> |
|-------------|-------------|-------------|-------------|-------------|-------------|------------------|------------|-------------|-------------|-------------|-------------|-------------|------------------|-------------|
| X | | X | X | ins. | X | | X | X | X | X | X | X | X | X |
| X | ins. | X | 2 parts | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | X | X | 2 parts | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | X | X | 2 parts | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | X | X | 2 parts | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | X | X | 2 parts | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | X | X | X | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | Pt. mut. | X | X | X | Δ | X dgcX | X | X | X | X | X | X | X | X |
| X | X | X | ins. | X | Δ5' | | X | X | X | X | X | X | X | X |
| X | | X | X | X | X | | X | X | IS | X | X | X | X | X |
| X | X | | X | X | X | | X | X | X | X | X | X | X | X |
| X | | X | X | X | X | | X | X | X | X | X | Pt. mut. | X | X |
| X | X | X | X | ins. | X | | X | X | X | X | X | Δ | X | X |
| Pt. mut. | X | X | X | ins. | X | | X | X | X | X | X | X | X | X |
| Pt. mut. | X | X | X | ins. | X | | X | X | X | X | X | X | X | X |
| X | X | X | X | ins. | X | | X | X | X | X | X | Δ | X | X |
| Pt. mut. | X | X | X | ins. | X | | X | X | X | X | X | X | X | X |
| X | | X | X | Δ | X | | X | X | X | X | X | X | X | X |
| X | | X | X | X | Δ | | X | X | X | X | X | X | X | X |
| X | | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | | frag. | Δ | IS | X | | X | X | X | X | X | X | X | X |
| X | Δ | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | Δ | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | X | X dgcY | X | X | Pt. mut. | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | X | | X | X | Pt. mut. | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | Pt. mut. |
| X | X | X | X | X | Δ | | X | X | X | X | X | X | X | X |
| X | | X | X | X | X | X dgcY | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | X | | X | X | | X | X | X | X | X | X | X | Pt. mut. |
| X | X | X | | X | X | | X | X | X | X | X | X | X | Pt. mut. |
| X | X | X | X | X | X | | X | X | X | X | X | X | X | Pt. mut. |
| X | X | X | X | X | Δ5' | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | Δ5' | | X | X | X | X | X | X | X | X |
| X | | frag. | X | X | | | X | X | X | X | X | X | IS | X |
| X | X | X | Pt. mut. | X | Δ5' | | X | X | Pt. mut. | X | X | X | X | X |
| X | | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | X | frag. | ins. | X | X | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | X | X dgcX | X | X | X | X | X | X | X | X |
| X | X | frag. | X | X | X | | X | X | X | X | X | X | X | Pt. mut. |
| X | X | X | X | X | X | | X | X | X | X | X | X | X | X |
| X | IS | X | X | X | Δ5' | | X | X | X | X | X | X | X | X |
| X | | X | X | X | Δ5' | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | Δ5' | | X | X | X | X | X | X | X | X |
| X | X | X | X | X | Δ5' | | X | X | X | X | X | X | X | X |
| <i>yeaP</i> | <i>ycdT</i> | <i>yddV</i> | <i>yedQ</i> | <i>yegE</i> | <i>yneF</i> | <i>novel DGC</i> | <i>rtn</i> | <i>yhjH</i> | <i>yfgF</i> | <i>yhjK</i> | <i>ylaB</i> | <i>yciR</i> | <i>yfeA/yfdA</i> | <i>yliE</i> |

4.1.1 Genes encoding GGDEF domain-containing proteins

Genes encoding GGDEF domain-containing proteins were conserved with varying frequency. Figure 4.1 displays the frequency of protein conservation among the 61 analyzed strains. Only proteins for which the gene was free from deleterious corruptions were included in the figure. All corruptions are described thoroughly in the results and apart from two exceptions were compared in reference to the wild-type variant from the K-12 W3110 strain. The two exceptions were the *yneF* gene where the wild-type variant used for gene comparison came from the EAEC 55989 strain, and the *yddV* gene where the wild-type variant used for gene comparison came from the K-12 MG1655 strain.

There were two DGC protein encoding genes that were conserved within all 61 *E. coli* strains analyzed in this study. These were *yaiC* (*adrA*) and *yliF*. *YaiC* has been shown to be CsgD dependent and expressed in late stationary phase (Brombacher *et al.*, 2003; Sommerfeldt *et al.*, 2009), as well as required for cellulose specific activity (Simm *et al.*, 2004). The function of *YliF* remains to be elucidated.

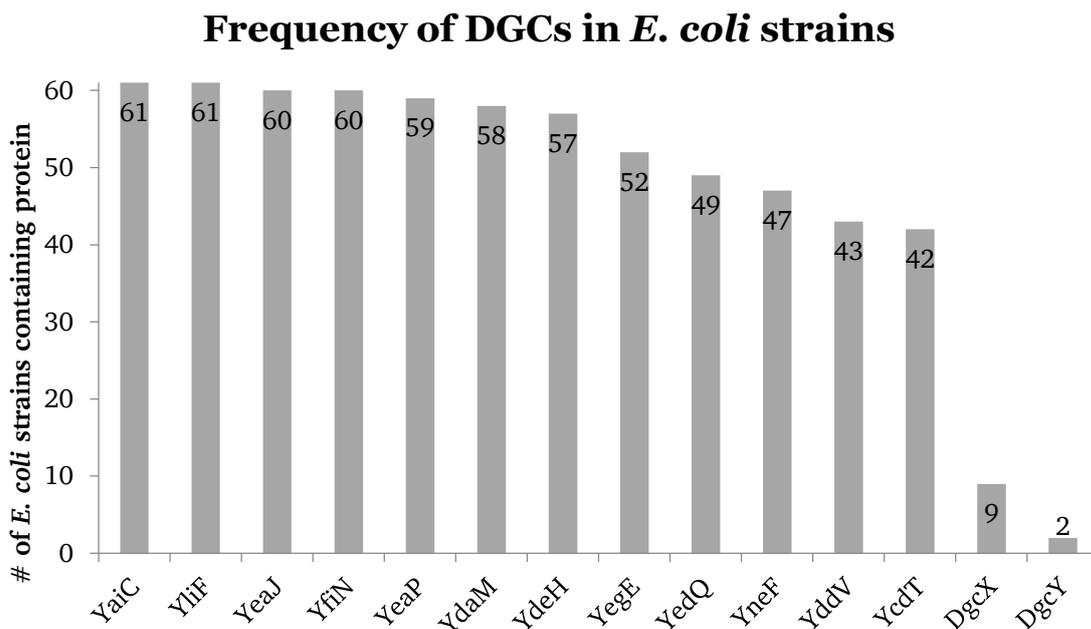


Figure 4.1: Frequency of protein conservation of the putative diguanylate cyclases in the 61 analyzed *E. coli* strains. Displayed are the DGCs and how often their genes are found fully intact.

The gene *yeaJ* was conserved in 60 of the analyzed strains and was only corrupted in *E. coli* str. K-12 substr. MDS42. The first 1425 nucleotides of the gene corresponding to the wild-type variant were deleted altogether from the genome and only the last 66 nucleotides of the gene remained intact. The deletion was up to the intergenic region between *yeaJ* and *yeaI* and left *yeaI* completely intact.

Results

The gene *yfiN* was only corrupted in a single case and conserved in all of the other 60 strains. In the EPEC O55:H7 str. RM12579 there was a deletion of a G in the 371st codon (after 1114 nucleotides) which led to a frameshift and early termination after 376 AAs whereas the wild-type version had 408 AAs. The GGDEF motif was located within the portion predicted to be expressed but it is unclear how missing the end portion of the protein affects the protein's stability.

The *yeaP* gene was disrupted in three of the EHEC O157:H7 strains: EDL933, Sakai and Xuzhou21. In the three strains the point mutation occurred at the same position after the 847th nucleotide in the 283rd codon where a G (in the wild type variant) was mutated into a T which resulted in the TAA stop codon. The point mutation occurred after the region encoding the GGDEF motif though the last 59 AAs would not be expressed.

The *ydaM* gene was missing altogether from the genome of *E. coli* str. K-12 substr. MDS42 and was corrupted in *E. coli* 0127:H6 E2348/69 (EPEC) and *E. coli* O7:K1 str. CE10 strain (ExPEC, NMEC). In E2348/69 the corruption consisted of the deletion of 88 nucleotides starting in the 187th codon. The corresponding deleted region in the wild-type strain was flanked on both sides by the sequence 'CGTCATG' and in E2348/69 this sequence marked where the deletion occurred. This deletion was not in-frame and caused early termination after encoding for 228 AAs before the GGDEF motif was reached. Within the *E. coli* O7:K1 str. CE10 strain (ExPEC, NMEC) the *ydaM* gene had an in-frame deletion of 12 nucleotides after the 934th nucleotide of the gene. The missing 12 nucleotides encoded the residues 'ALAR' in the C-terminus and were located in the corresponding codons 313-316 in the wild-type variant. It is unclear how this deletion influences the protein's activity or stability; the A-site, I-site and the claimed proteolysis target sequence (LAA; Flynn *et al.*, 2003) are all intact.

The *ydeH* gene was corrupted in four strains: *E. coli* ETEC H10407, *E. coli* O26:H11 str. 11368 (EHEC), *E. coli* UMNK88 (ExPEC) and *E. coli* ABU 83972. In *E. coli* ETEC H10407, the 5' end consisting of 119 bps was deleted and the remainder of the gene starting with the 35th codon relative to the wild-type was intact. The deleted region was replaced by a divergently transcribed IS1 insertion element which consisted of *insA* and *insAB* genes. This insertion spanned the genome up to the *ETEC_1608* gene which encoded a dipeptidyl carboxypeptidase II. In the *E. coli* UMNK88 (ExPEC) strain the corruption consisted of a point mutation occurring after the 775th nucleotide in the 259th codon where 'GAA' was mutated into the stop codon 'TAA.' The stop codon occurs after the GGEEF motif. The *ydeH* gene in the *E. coli* O26:H11 str. 11368 (EHEC) strain also had an IS1 element insertion after the 603rd nucleotide into its 205th codon. The IS1 element was 776 nucleotides long and contained *insA* and *insAB* genes. The insertion element was flanked by the sequence 'GTTTATCG.' The two IS1 insertion elements reported in this paragraph share a 90% identity with each other. Finally, the *E. coli* ABU 83972

Results

strain had a 70 nucleotide deletion in the *ydeH* gene. The deletion started after the 118th codon and resulted in a frameshift and early termination after 133 AAs before the GGEEF motif was reached.

The *yegE* gene was affected in eight strains, with all of the EHECs of the O157:H7 serotype (EC4115, EDL933, Sakai, TW14359 and Xuzhou21) sharing the same disruption. After the 457th nucleotide in the 153rd codon there was an insertion of an A which led to a frameshift and early termination after 177 AAs and before the GGEEF domain was reached. The *yegE* gene in the EAEC 042 strain was also corrupted by an insertion. An extra G in a string of 6 G's in the 138th codon after the 413th nucleotide led to a frameshift and early termination after 177 AAs before the GGEEF domain was reached. In the EPEC O127:H6 str. E2348/69, the *yegE* gene was disrupted by a large insertion of 1551 nucleotides of an ISEC13 transposase. The disruption occurred after the 418th nucleotide in the 140th codon and the transposase was flanked by the sequence 'GTTGGGGG.' In the EHEC O26:H11 str. 11368 strain, a deletion of 7 nucleotides after the 1467th nucleotide in the 490th codon corrupted the *yegE* gene and led to frameshift and early termination after 506 AAs before the GGDEF domain was reached. Finally, in the commensal *E. coli* B str. REL606 strain, the *yegE* gene had an in-frame deletion of 12 nucleotides encoding the residues 'GTQL.' In the wild-type variant these residues are encoded in the codons 40-43.

The *yedQ* gene was completely absent in two commensal *E. coli* strains: B str. REL606 and BL21(DE3). In the porcine ETEC UMN18 the *yedQ* gene had an insertion of four nucleotides after the 155th nucleotide in the 53rd codon that caused a frameshift and early termination after 58 AAs and before the GGDEF domain was reached. In the EPEC O127:H6 E2348/69 the *yedQ* gene was disrupted by a single nucleotide deletion of an A after the 1282nd nucleotide in the 427th codon which led to a frameshift and early termination after 461 AAs and before the GGEEF motif. A point mutation disrupted the *yedQ* gene in the commensal *E. coli* strain HS. The point mutation occurred after the 341st nucleotide in the 114th codon and changed the codon from a 'TGG' to a 'TAG.' In the commensal *E. coli* Nissle 1917 strain the 3' end of the *yedQ* gene was deleted, including the GGEEF motif (after 411th codon) and instead DNA encoding flagellar export pore protein was found in that region. In six the strains EAECs: 55989, HUSEC041, LB226692, 2011C-3493, 2009EL-2071 and 2009EL-2050 the same interesting anomaly was observed. The *yedQ* gene had a point mutation after the 934th nucleotide in the 312th codon from a 'CAG' in the wild-type variant to the stop codon 'TAG.' The stop codon was followed by a start codon (314th codon relative to the wild-type variant) within the same reading frame.

The *yneF* gene was affected in 14 *E. coli* strains and it was completely absent from the genome of the commensal *E. coli* ED1a. In four strains a single nucleotide deletion caused a reading frame shift and subsequent early termination. In both STEC strains O145:H28 str. RM13516 and O145:H28 str. RM12761, which are clonal-isolates from the 2007 STEC outbreak

Results

in Belgium collected from ice-cream (Cooper *et al.*, 2014), the deletion of a T in the 32nd codon after the 95th nucleotide led to early termination after only 49 AAs. In the APEC O78, a deletion of a single C in the 31st codon after the 92nd nucleotide also led to early termination after 49 AAs. In the ETEC H10407 a deletion in the *yneF* gene occurred after the 348th nucleotide in the 117th codon. The deletion was of an A in a string of four A's that led to a shift in the reading frame and early termination after 122 AAs and before reaching the GGDEF domain.

In nine strains the *yneF* gene had the same anomaly of having a large 5' end deletion. The deletion spanned 433 nucleotides and the DNA upstream of the deletion was identified as the *yneG* gene, encoding a glutaminase. In these six cases the protein was annotated as consisting of 315 AA residues instead of 472, but based on the large deletion it is unlikely that *yneF* would be expressed. The 5' end deletion is shown in figure 4.2 where *yneF* is displayed in five strains and the intact version is compared to the version with the large 5' deletion. The strains affected by this deletion were the porcine ETEC UMNF18 and the commensals: ATCC 8739, BW2952, DH1, HS, K12 substr. DH10B, K-12 substr. MG1655, K-12 substr. W3110, K-12 substr. MDS42. All of the K-12 strains had the 5' deletion. The *yneF* gene was also found in *Shigella flexneri*, *Klebsiella pneumoniae*, *Escherichia albertii*, *Escherichia fergusonii*, *Citrobacter amalonaticus* and *Citrobacter rodentium*. The encoded protein YneF was chosen for further investigation into its regulation and possible function reported later in this project.



Figure 4.2: Alignment of genomic regions of the *yneF* gene compared in 5 strains of *E. coli*. At the top is the K-12 W3110 strain used as a reference, followed by the EAECs 55989, HUSEC041, the outbreak strain LB226692 and the EHEC EDL933. In red are genes of predicted GGDEF domain proteins. *yneF* is shown as incomplete in the K-12 W3110, the same 5' deletion is also found in a number of other strains discussed in this study.

The *ycdT* gene was completely absent from 14 genomes and corrupted in 5. The following strains had it deleted in its entirety: 042 (EAEC), UMNK88 (porcine ETEC), O111:H- str. 11128 (EHEC), O26:H11 str. 11368 (EHEC), O145:H28 str. RM13516 (STEC), O145:H28 str. RM13514 (STEC), O145:H28 str. RM12761 (STEC), O145:H28 str. RM12581 (STEC), O127:H6 str. E2348/69 (EPEC), UMN026 (ExPEC), SMS-3-5 (unknown pathogen), ED1a (commensal), IA11 (commensal) and K-12 MDS42 (commensal). The ETEC H10407 had a point

Results

mutation in the 450th codon (1348th nucleotide) that caused the ‘GAG’ codon according to wild-type variant to be transformed into the stop codon ‘TAG.’ This premature stop omitted the last three AAs of the YcdT protein and the predicted protein was 449 AA residues long compared to the wild-type 452 AA. Both EPEC strains O55:H7 str. CB9615 and O55:H7 str. RM12579 possessed a single nucleotide deletion of a C after the 483rd nucleotide in the 162nd codon led to a frameshift and early termination after 162 AAs before the GGEEF motif was reached. In the commensal K-12 strain DH10B an IS2 insertion element disrupted the *ycdT* gene. The element consisted of 1336 nucleotides, *insA* and *insAB* genes. The insertion occurred after 612 nucleotides of the *ycdT* gene in the 205th codon and resulted in a frameshift. 219 AAs were predicted to be expressed before a stop codon would be reached. The IS2 insertion element was flanked on both ends by the sequence ‘GTGGTTA.’

Lastly, in the EAEC 55989 the *ycdT* gene was disrupted by a 7 nucleotide insertion that was a repetition of the sequence immediately following the insertion ‘TTTGTTT.’ This insert occurred after the 461st nucleotide in the 153rd codon and was experimentally confirmed in this study by re-sequencing. The insertion resulted in a shift in the reading frame and early termination so that only 171 AAs were predicted to be expressed and the GGDEF domain would not be reached. The LB226692 and HUSEC041 sequences that are available on the NCBI database also claimed a 7 nucleotide insertion in the *ycdT* gene in these strains but re-sequencing of this DNA region obtained by PCR had shown that in fact the *ycdT* gene was intact and the entire region (4363bps) including the *pgaA* gene and the *ycdT* gene was 99% identical (4350/4363 identities) to the analogous region in the K-12 W3110 strain (Fig. 4.3), and the intergenic region between the *pgaA-ycdT* was entirely identical (588/588 identities). These results were confirmed by the latest sequenced strain added to this study (June 2014), *E. coli* O104:H4 str. 2011C-3493. The strain is the clonal isolate, isolated from an American HUS patient who had traveled to Germany and had contracted the infection (Ahmed *et al.*, 2012). Figure 4.3 depicts the genomic region around the *ycdT* gene in five strains including the disrupted version in *E. coli* 55989 which also has an insertion element, IS1 between *pgaA* and *ycdT*.

Results

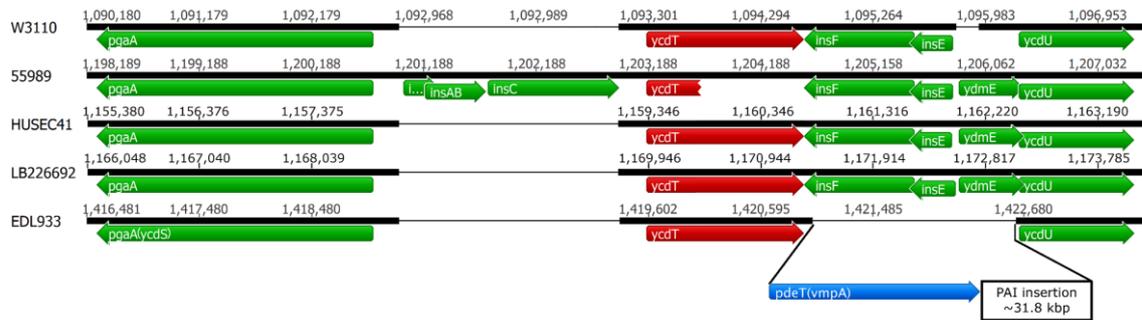


Figure 4.3: Alignment of genomic regions of *pgaA-ycdT* genes compared in 5 strains of *E. coli*. At the top is the K-12 W3110 strain used as a reference, followed by the EAECs 55989, HUSEC041, the outbreak strain LB226692 and the EHEC EDL933. In red are genes of predicted GGDEF domain proteins, in blue are genes of predicted EAL domain proteins. 7 nucleotide insertion in the 55989 strain resulted in a frameshift and early termination of *ycdT* and the IS1 insertion elements may be providing an additional promoter for *pgaA*. The EHEC EDL933 shows the location of the novel *pdeT* (*vmpA*) gene, in the same operon as *ycdT*. This arrangement is also found in all other EHECs of the O157:H7 serotype.

It is also worth noting that YcdT was annotated as having three different lengths among the *E. coli* genomes of either 452 AAs, 476 AA or 496 AA, further analysis of the DNA sequence revealed that two different start codons were responsible for the difference in size. The 452 AA annotation used an ‘ATG’ as the start codon while the 476 AA and 496 AA variant used a ‘TTG,’ but at different locations. As ‘ATG’ is more common, it is probable that the 452 AA variant is the actual size of the protein. Additionally, the 41 AA difference in size between the 496 AA and 472 AA annotations also predicts an additional transmembrane segment (TMS) meaning that the YcdT would have 9 TMSs instead of 8 TMSs in the longer variant with the N-terminus oriented toward the outside and the C-terminus oriented toward the inside of the cell. The 8 TMS version is predicted to have both the N- and C-termini oriented inside the cell.

The *yddV* (*dosC*) gene was affected in 17 instances. In the EHEC O103:H2 str. 12009 it was absent altogether while in the other 16 cases the same type of deletion events took place. The 5’ end of the *yddV* gene consisting of 644 nucleotides was deleted and the deletion went further to correspond with the region -38bps upstream of the ATG start codon found in the wild-type variant. This deletion included the native Shine-Dalgarno sequence of the *yddV* gene. A brief fragment (44bps) of the *yddV* gene was found corresponding to the codons 215-230 in the wild-type variant. This sequence was followed by the end portion of the *yddV* gene corresponding to the codons 427-461 in the wild-type variant, implying another internal deletion of 530bps. The 17 strains affected by this phenomenon are: O127:H6 str. E2348/69 (EPEC), JJ1886 (STEC), IHE3034 (ExPEC, MNEC), S88 (ExPEC), PMV-1 (ExPEC), 536 (UPEC), CFT073 (UPEC), UTI89 (UPEC), LF82 (AIEC), UM146 (AIEC), O83:H1 str. NRG 857c (AIEC), APEC O1, ABU 83972 (commensal), ED1a (commensal), SE15 (commensal) and Nissle 1917 (commensal). Table 4.2 denotes the deletion in the *yddV* gene as ‘frag.’ Strains SE15 and JJ1886 also had an additional deletion of a C in the 445th codon, resulting in a frameshift. Figure 4.4 illustrates this

Results

deletion in the representative UPEC 536 strain compared to the modified W3110 strain where the annotation was manually adjusted to show the correct *yddV* length.

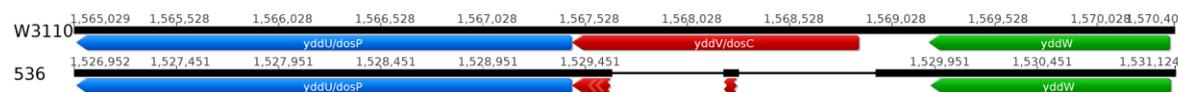


Figure 4.4: Genomic alignment of the *yddV* gene in UPEC 536 with the modified reference strain W3110 (the annotation of the *yddV* gene was adjusted to reflect the true size of the gene). The UPEC 536 strain is a representative for the 16 other strains affected with this deletion (Table 4.2). The first 644bps of the 5' end of *yddV* is deleted along with a secondary internal region (530bps).

YddV was annotated to be 347 AA long in the W3110 strain as opposed to the 460 AA variant in the K-12 MG1655 strain. This proved to be an annotation error within the W3110 genome as the nucleotide sequence is completely intact, corresponding to the 460 AA version. Immune blot experiments detecting the expression levels of the 3X flag tagged protein encoded by its chromosomal copy showed that the size was ~56 kDa, as was expected for a 460 AA length variant plus flag tag as opposed to ~43 kDa, as was expected for a 347 AA plus flag-tag variant (Gisela Klauck, unpublished data).

4.1.1.1 Two novel GGDEF domain proteins in pathogenic *E. coli*: DgcX and DgcY

A novel GGDEF domain protein was found to be encoded in nine of the 61 strains analyzed. These strains were the pathogens 55989 (EAEC), the May 2011 outbreak strain LB226692 (EAEC), its clonal isolate 2011C-3493 (EAEC), its close relatives: HUSEC041 (01-09591) (EAEC), the 2009 Georgian outbreak HUS O104:H4 strain 2009EL-2071 (EAEC), a clonal isolate of the 2009 Georgian outbreak HUS O104:H4 strain 2009EL-2050 (EAEC), the ETECs: E24377A and H10407 and a commensal, SE11. Table 4.3 lists the designations and GI numbers of the novel GGDEF domain-containing proteins within the native strains. The protein had no official name and is designated as DgcX, for diguanylate cyclase X. DgcX is 443 AA long and shares a 98.87% identity (438/443) between the nine strains. DgcX was predicted by the NCBI database to have 8 TMSs (Fig. 4.5). This was also confirmed by the TMHMM software. Based on the amino acid sequence the protein had both an intact A-site (GGEEF) and I-site (RKED) and therefore most likely has DGC activity. The protein had a tail consisting of 170 AA in the C-terminus where the GGDEF domain resides, and was oriented inside the cell membrane (GGDEF is in the cytoplasm). A search with NCBI yielded no other known domains to be associated with this protein and the tail region was large enough to only accommodate the GGDEF domain.

Results

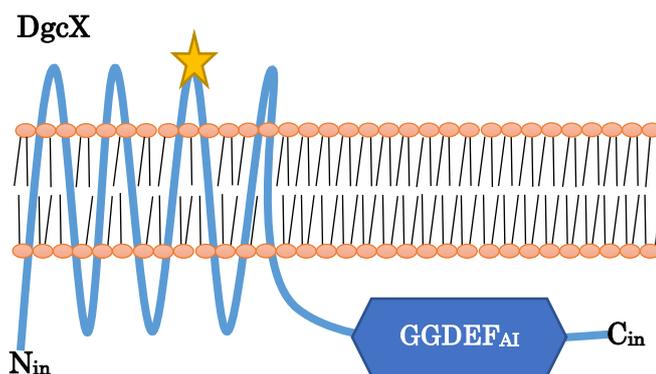


Figure 4.5: Predicted topology of the DgcX protein, including the location of the known domains within the protein. 443 AA residues long, 8 TMSs, with “in” denoting the cytoplasmic side of the membrane, “A” denoting an intact A-site and “I” denoting an intact I-site. The star in yellow denotes the location of motif A discussed in section 4.1.1.3.

Table 4.3: The novel GGDEF domain-containing proteins and their native in strain designation.

| Designation in this study | Strain | Designation in strain | GI number |
|---------------------------|-----------------|-----------------------|-----------|
| DgcX | LB226692 | HUSEC_04298 | 340741263 |
| | 2011C-3493 | O3K_17495 | 407483024 |
| | HUSEC041 | HUSEC41_04052 | 340735556 |
| | 55989 | EC55989_0813 | 218694243 |
| | 2009EL-2071 | O3O_07785 | 407468242 |
| | 2009EL-2050 | O3M_17475 | 410483577 |
| | EPEC H10407 | Not annotated | |
| | E24377A | EcE24377A_0835 | 157155149 |
| | SE11 | ECSE_1457 | 209918648 |
| DgcY | SMS-3-5 | EcSMS35_1716 | 170517710 |
| | O7:K1 str. CE10 | CE10_1648 | 386624007 |

Upon performing operon analysis of the *dgcX* gene within the nine *E. coli* strains it was found in all of the strains that the *dgcX* gene was located as a single transcription unit, not as part of an operon (Fig. 4.6). In eight of the strains the gene was in the same location flanked by the *ybhB* and *bioA* genes upstream and a large prophage insertion downstream. YbhB is a putative kinase inhibitor protein (Serre *et al.*, 2001). BioA is part of the biotin biosynthesis pathway and is

Results

a 7, 8-diaminopelargonic acid synthase enzyme (Stoner *et al.*, 1975; Eliot *et al.*, 2002). In all of the EAECs of the O104:H4 serotype: *E. coli* 55989, LB226692, 2011C-3493, HUSEC041, 2009EL-2071 and 2009EL-2050 the prophage insertion downstream of *dgcX* was from the prophage DLP12 and was approximately 43.8 kbp. Upon closer inspection the prophage used the lambda *attB* attachment site to insert itself between the *ybhC* and *ybhB* genes. The genomic context of *dgcX* in H10407 (ETEC) was similar to the other pathogenic strains, except that the prophage insertion upstream differed and was approximately 51.9 kbp in length. The *E. coli* E24377A (ETEC) had a different prophage insertion upstream of *dgcX* and was approximately 21.9 kbp in length.

In the commensal *E. coli* SE11 the *dgcX* gene was situated in a different location. There was also a prophage insertion (~39.3 kbp) upstream of *dgcX* (though different from the prophages found in the other strains), but downstream were genes encoding the universal stress protein F (UspF) and an outer membrane protein porin, OmpN (Fig. 4.7). The location of *dgcX* in SE11 was 636,161 bps upstream of *ybhB* with no other insertions at the lambda *attB* attachment site. DgcX may have evolved from a YeaI-like protein, as it is its closest homolog, with YcdT being the second closest (Fig. 4.13, Fig. 4.8). However, YeaI is a GGDEF domain-containing protein with a degenerate A-site and an intact I-site. It is still unclear what the proteins' function is but due to its intact I-site it might function as a c-di-GMP-binding effector protein. Two of the strains containing the *dgcX* gene had a full complement of the standard DGC proteins [*E. coli* E24377A (ETEC) and *E. coli* SE11 (commensal)] therefore it is unlikely to have evolved as a compensatory protein to act only as backup for a missing DGC. Due to its conservation among EAECs of the H104:O7 serotype (including the 2011 outbreak strain), this protein was chosen for further investigation into its regulation and possible function. The results are discussed below.

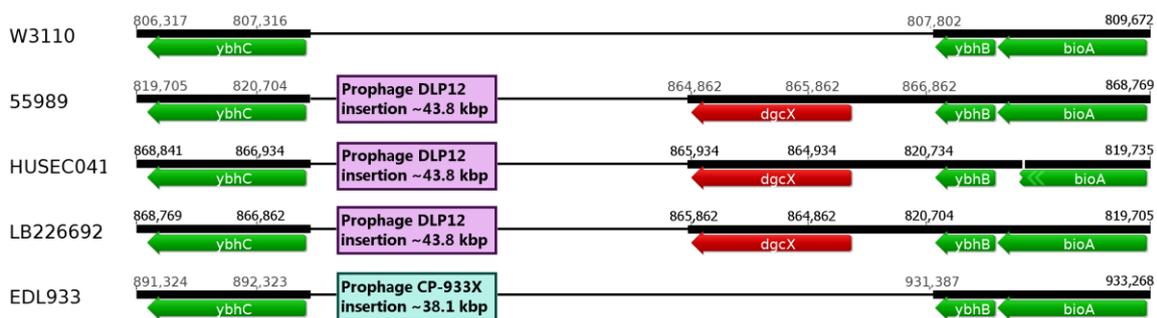


Figure 4.6: Alignment of genomic regions of the *dgcX* gene compared in 5 strains of *E. coli*. At the top is the K-12 W3110 strain used as a reference, followed by the EAECs 55989, HUSEC041, the outbreak strain LB226692 and the EHEC EDL933. In red are genes of predicted GGDEF domain proteins. All of the EAECs of the O104:H4 serotype had the novel *dgcX* gene in the same genomic layout. The different prophages are color coded according to type of prophage so that prophages marked by the same color are identical.

Results

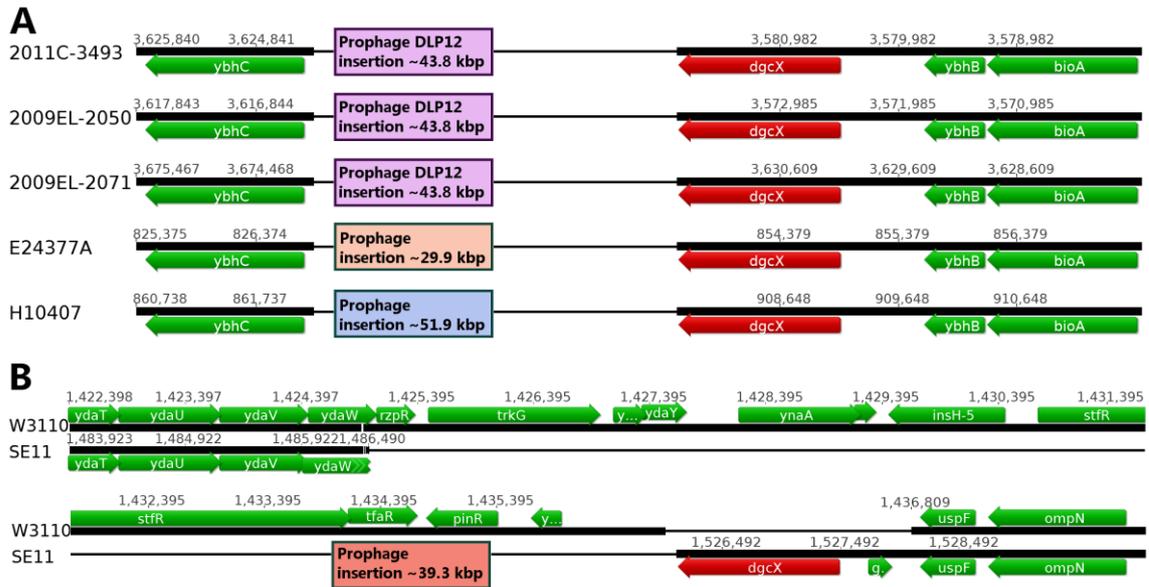


Figure 4.7: Genomic layout of the *dgxX* gene represented by the red arrow in each of the three operon layouts. In all of the layouts the upstream region of *dgxX* is of different prophages and each type of prophage is represented by a different color. A) Depicts the layout found in the 2011 outbreak strain (*E. coli* 2011C-3493), the 2009 Georgian outbreak clonal isolate strains (2009EL-2050 and 2009EL-2071), the ETEC E24377A and the ETEC H10407. In A and B the phages use the lambda *attB* attachment site but are different. B) Depicts the layout found in the commensal *E. coli* SE11. SE11 is the only strain to have this layout with a different position compared to the other two layouts depicted in A.

Results

predicted as a single one. This hypothesis is further supported by the fact that cyclic-di-GMP is known to be an intracellular molecule, but not for being able to cross the cell membrane, so that DgcY is unlikely to be oriented as suggested by the 5 TMS topology. Furthermore, applying the positive-inside rule (von Heijne & Gavel, 1998) yields two regions with the greatest number of positively charged residues (arginine and lysine). If the 5 TMS topology is assumed, one of the two regions would have to be oriented outside of the cellular membrane. According to the TMHMM program, the highest probabilities of cell orientation lie within the N-terminus and the regions between loops 1-3, with a high probability of 0.92539 (Fig. 4.9). When a 6 TMS topology is assumed, both of the regions with the greatest number of positively charged residues are oriented inside the cellular membrane, in accordance with the positive-inside rule. DgcY conserved all of the integral residues for DGC activity in the A-site with the GGEEF sequence but had a degenerate I-site sequence, RQSED (Table 4.2).

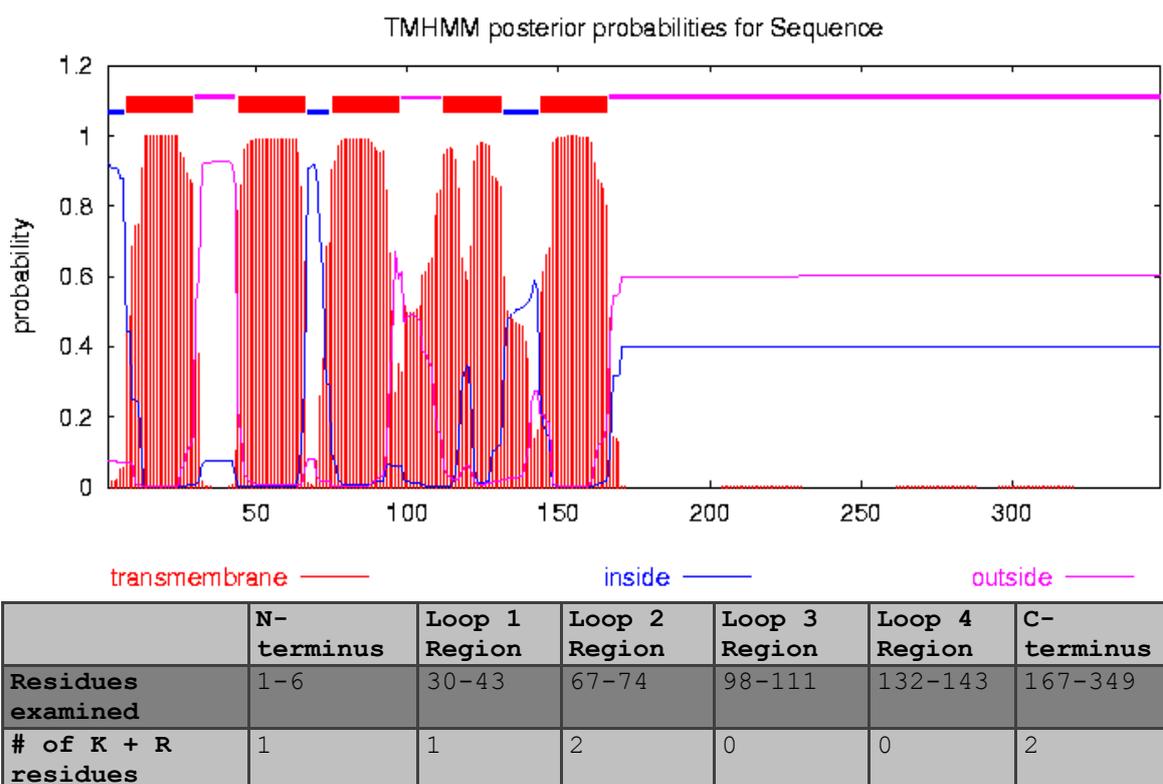


Figure 4.9: Hydropathy plot of the DgcY protein generated by TMHMM, where the highest probability lies in the N-terminus region and the regions of loops 1-3. With a 0.92539 probability that the N-terminus is oriented inside the cell membrane, the false merge is speculated to occur on the 4th TMS where two compressed peaks are clearly visible. Located below is the table describing the DgcY protein which was examined manually for probable orientation within the membrane. Twenty amino acyl residues adjacent to the first or last TMS were examined in both termini, except if TMHMM predicted that fewer residues would be found. In the case of DgcY seven residues were examined in the N-terminus as a result.

DgcY seemed to be located in the same operon as genes coding for two other proteins: metallo-beta-lactamase family protein (EcSMS35_1714) and a small hypothetical protein (EcSMS35_1715) not found in any of the other *E. coli* strains analyzed in this study (Fig. 4.10).

Results

DgcY's closest homologue outside of the *E. coli* species was the CKO_01497 protein found in *Citrobacter koseri* ATCC BAA-895. The two proteins shared a 64% identity throughout their full alignment (data not shown). Based on the multiple alignment of the entire GGDEF domain-containing proteins, DgcY clustered most closely with the YdeH protein followed by YedQ, as visualized in the phylogenetic tree (Fig. 4.13). The alignment of the DgcY and YdeH amino acid sequences can be seen in Figure 4.11.

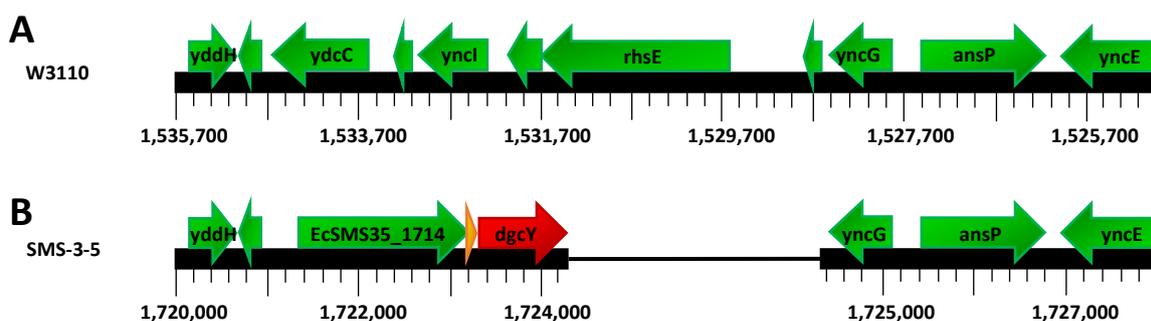


Figure 4.10: Genomic alignment of the *dgcY* gene in *E. coli* SMS-3-5, an antibiotic resistant environmental isolate, with reference strain W3110. The location of the novel gene *dgcY* (*EcSMS35_1716*) is indicated by the red arrow. This layout is also representative for the *E. coli* O7:K1 str. CE10 strain, where *dgcY* is termed *CE10_1648*.

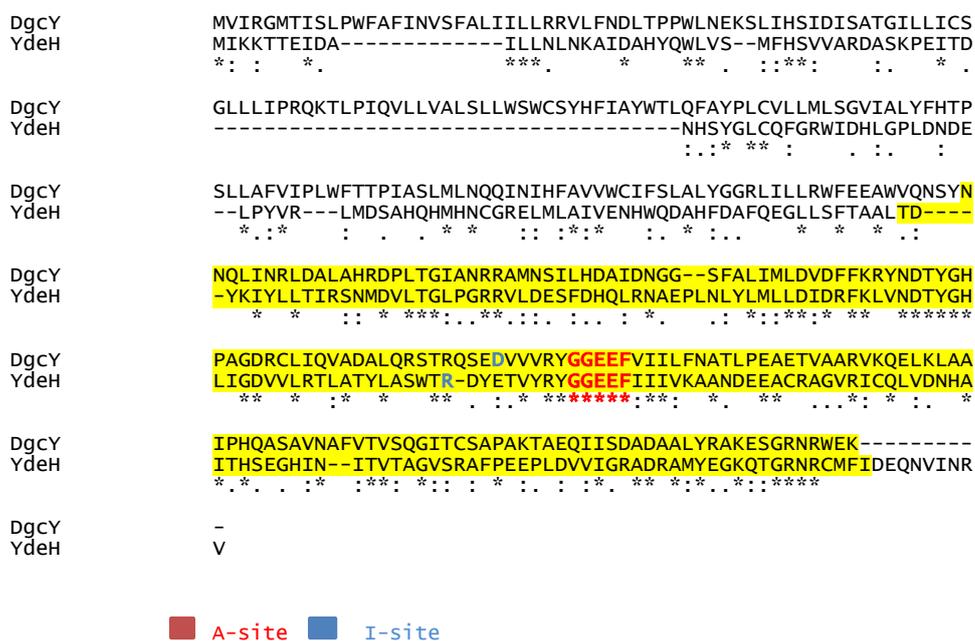


Figure 4.11: GGDEF domain sequence alignment of the novel DgcY and its closest homolog YdeH, using entire protein sequences. Highlighted in yellow is the GGDEF domain region, blue lettering denotes the I-site and red lettering marks the location of the AAs important for the activity of the active site. The alignment was generated using CLUSTAL X (Thompson *et al.*, 1997) and then manually edited to highlight key amino acid positions associated with DGC activity.

4.1.1.2 Novel conserved sensory domain in DgcX, YcdT and Yeal

A MEME (Bailey & Elkan, 1995) analysis was performed on all of the GGDEF domain-containing proteins in their entirety. The analysis further elucidated how the GGDEF domain

proteins relate to each other and why they cluster according to the layout seen in Figure 4.13. Besides sharing the GGDEF domain, some of the proteins shared additional sequence similarities. Proteins YeaI, YcdT and DgcX shared a motif that spanned 13 AAs and was made up of the putative AxxLSSxxxxYxL consensus sequence, as shown in Figure 4.12. The motif was located in the hydrophilic region between TMSs 5 and 6 of the DgcX, YeaI and YcdT protein topology. The 13 AA motif ran along residues 176-188 of DgcX, along residues 221-233 of YeaI and along residues 175-187 of YcdT. All of these proteins possessed 8 TMSs. Trying to match the motif to established protein domains yielded inconclusive results, as various databases and programs including MAST, SMART and ProDom could not match it to any known domain. The motif may be some kind of sensor domain.

Motif analyses were performed with two sets of proteins. The first set contained the GGDEF domain-containing proteins, mostly from the K-12 W3110 strain, with novel or incomplete proteins obtained from their native strains. In this set each type of protein was represented by a single member. This initial analysis identified the existence of the described motif. In order to refine the motif, a second analysis was performed using a second set of proteins, consisting of nine DgcX sequences, seven YcdT sequences and ten YeaI sequences. The DgcX protein sequence was obtained from the nine strains that carry it. These strains were also used to obtain the sequences of YcdT and YeaI. As a reference, YcdT and YeaI were also included from the W3110 strain. The refined motif that was detected by these 2 sets of analyses is presented in figure 4.12.



Figure 4.12: Consensus sequence of the motif found in DgcX, YcdT and YeaI proteins. The motif spans 13 AAs and is located in the hydrophilic region between TMSs 5 and 6 of the DgcX, YcdT and YeaI protein topology. The X-axis is numbered according to the amino acid position within the consensus sequence. The 13 AA motif ran along residues 176-188 of the DgcX protein, residues 221-233 of the YeaI and residues 175-187 of the YcdT protein.

4.1.1.3 Phylogenetic distribution of GGDEF domain-containing proteins

Figure 4.13 shows the phylogenetic relationship based on the entire protein amino acid sequence of all of the GGDEF domain-containing proteins. The difference in the N-terminal sensory

Results

domains mainly dictates the phylogenetic relationship between the proteins (see below). The 16 included proteins are located in 6 clusters, numbered 1-6 in figure 4.13.

The first cluster consists of the proteins YeaI, YcdT and DgcX. All three of them contain eight TMSs and also possess a conserved motif (section 4.1.1.2). The second cluster groups the proteins DgcY, YdeH, YddV and YliF, with DgcY and YdeH, and YddV and YliF clustering closer together. DgcY has six TMSs, YliF and YdeH have two TMSs and YddV has none. YddV has a sensor globin domain while the other three proteins have no identified sensory domains. In the case of these four proteins, it may be the strong similarity in their GGDEF domains that cause them to cluster together rather than any similarity in their N-terminal sensory domains. The alignment of YdeH and DgcY in figure 4.11 supports this claim in that the alignment in the C-terminal is much stronger than that in the N-terminal. In the third cluster, the proteins YneF, YaiC, CsrD and YeaJ cluster together. All four are transmembrane proteins. CsrD and YeaJ have two TMSs, YaiC has six TMSs and YneF contains eight TMSs. All of them lack the potential binding motif found in the proteins of cluster 1 which may be the reason why the two groups cluster separately. YedQ did not closely align with any of the other proteins and is in its own cluster, cluster 4. The proteins found in cluster 5, YdaM and YegE, have PAS domains. Finally the proteins forming cluster 6 are YeaP and YfiN. It is unclear why these two proteins cluster closely together, YeaP contains a GAF sensory domain, while YfiN has no identified sensory domains.

Results

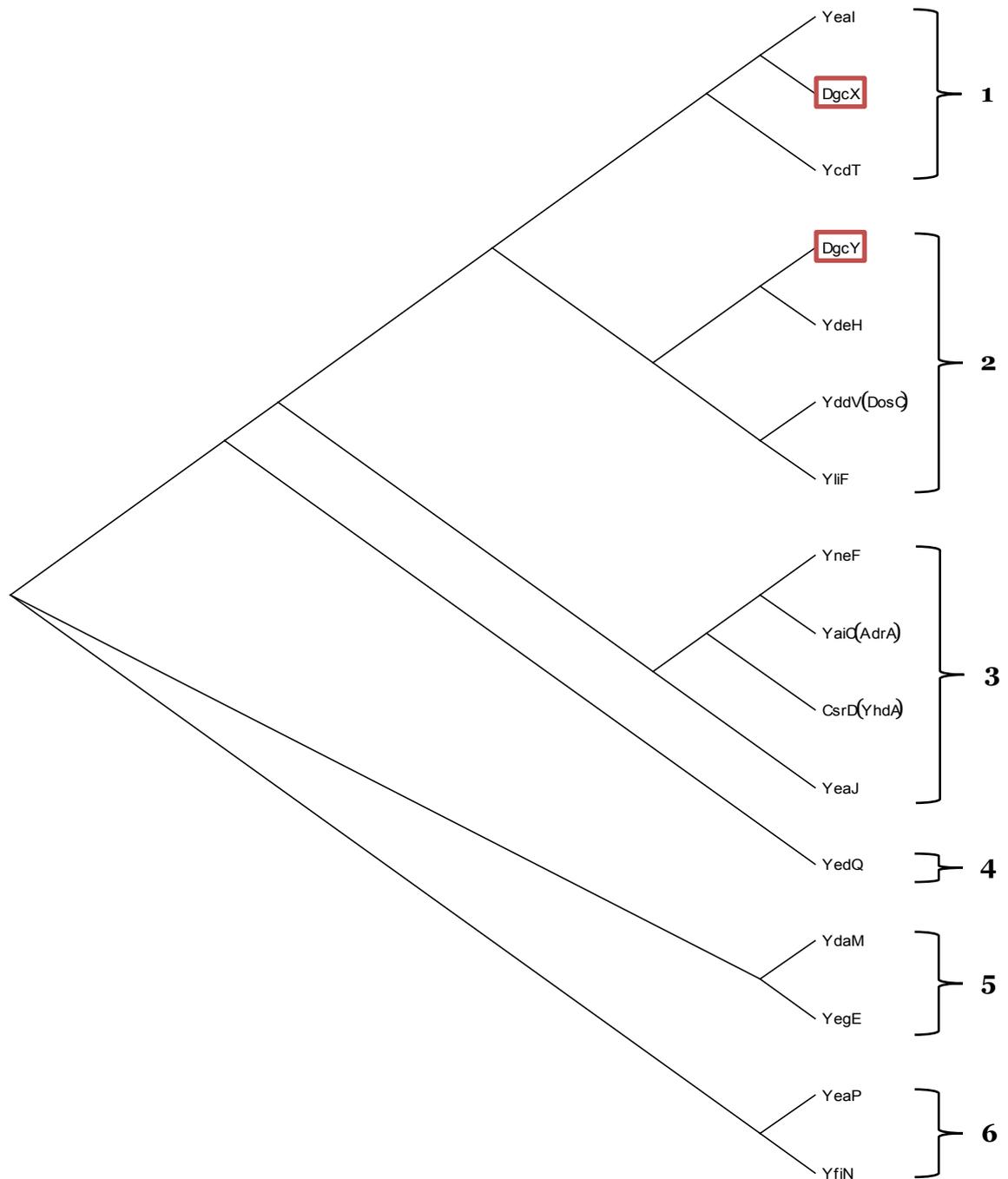


Figure 4.13: Phylogenetic tree of the GGDEF domain-containing proteins. The tree is based on the ClustalX multiple alignment of GGDEF domain-containing proteins in their full entirety and drawn with the TreeViewX program. The novel proteins are boxed in red. The 16 proteins group in 6 clusters, numbered 1-6.

4.1.2 EAL domain-containing proteins—core and pathotype-specific proteins

Genes encoding EAL domain-containing proteins were conserved with varying frequency, as displayed in figure 4.14. Only proteins for which the gene is free from deleterious corruptions were included in the figure. All corruptions are described thoroughly in the results and were compared in reference to the wild-type variant from the K-12 W3110 strain.

Frequency of PDEs in *E. coli* strains

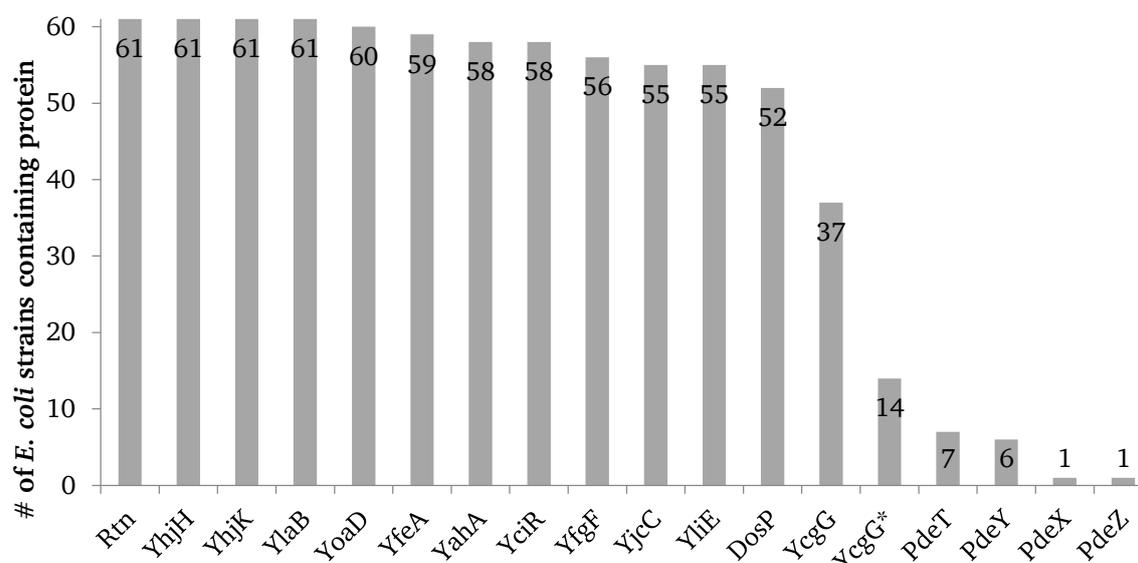


Figure 4.14: Frequency of protein conservation of the putative phosphodiesterases in the 61 analyzed *E. coli* strains. Displayed are PDEs in relation to how often their genes are found fully intact. The * next to the YcgG denotes the truncated version containing only the EAL domain.

There were four PDE protein encoding genes that were conserved within all 61 *E. coli* strains analyzed in this study. These were *rtn*, *yhjH*, *yhjK* and *ylaB*. The *yoaD* gene was corrupted in only a single strain (the ETEC *E. coli* E24377A) and was conserved in all of the other strains. In E24377A, the *yoaD* gene had a deletion of a C in a string of C's after the 354th nucleotide in the 118th codon. This deletion led to a shift in the reading frame and early termination (135 AAs compared to 532 AAs in the wild-type) before the EAL domain was reached.

The *yfeA* gene was completely missing from the genome of *E. coli* K-12 substr. MDS42 and was disrupted in the commensal *E. coli* ED1a strain. The disruption stemmed from an insertion of the IS600 insertion element. The insertion element spanned 1267bps and contained transposase ORF A and transposase ORF B. It started after the 405th nucleotide of the *yfeA* gene in the 135th codon and was flanked by the sequence 'GGC.' The YfeA protein was predicted to end on the 136th codon, brought in by the insertion element which is a TGA. The rest of the *yfeA* gene remained intact following the insertion.

The *yahA* gene was conserved in 58 strains of *E. coli*. It was missing altogether in two commensal strains (*E. coli* BW2952 and *E. coli* ED1a) and corrupted in the APEC O78 strain. A point mutation of the 886th nucleotide from a G in the wild-type to a T caused the 296th codon to become a stop codon (TAA). The point mutation led to early termination and predicted the protein length to be 295 AAs compared to 362 AAs in the wild-type.

The remaining 58 strains all had the *yahA* gene but the gene had two distinct variations. The difference arose in the promoter and the 5' region encoding the N-terminus of the YahA protein.

Results

In one form there was an immediate upstream insertion of an *aidA-I adhesin-like protein* coding gene making it possible that the genes are now within a single operon for strains with this insertion (Fig. 4.15), as no distinct hair pin between the two genes could be found. In table 4.2, the strains carrying the *yahA* gene with upstream insertion are identified and the *yahA* gene is denoted as ‘abn.’

When the two protein variations of YahA were aligned, it became visible that there was a strong discrepancy in the protein’s N-terminus (Fig. 4.16), specifically in the LuxR-C-like helix-turn-helix (HTH) domain. The HTH motif has been shown to be responsible for DNA binding in many proteins like LuxR in *Vibrio fischeri* (Egland *et al.*, 1999), NarL in *E. coli* (Baikalov *et al.*, 1996) and RcsB in *Erwinia amylovora* (Pristovsek *et al.*, 2003) and is likely to also be involved in DNA binding in the YahA protein. According to the work of Pristovsek *et al.*, (2003) the effector domain of the HTH is composed of two helices. The first one is the supporting (stabilizing) helix followed by the looping region (scaffold), and the second one is where the DNA recognition takes place. An alignment of the RcsB LuxR domain with the YahA LuxR-C-like domain showed which residues within the YahA protein would fit into which helices in the HTH motif (data omitted). Between the two YahA variants (Fig. 4.16) were a total of three differences within the HTH motif (Fig. 4.17). Two of them fell into the region potentially forming the helices. The first difference was within the stabilizing helix, where a Q (glutamine) in one variant (W3110) was an E (glutamic acid) in the other (55989). The second difference fell into the DNA recognition helix, where an F (phenylalanine) in one variant (W3110) was a Y (tyrosine) in the other (55989) (Fig. 4.17). These mutations are conservative amino acid substitutions and it could be argued that they have no substantial effect on the DNA binding activity of the HTH but the effect of the other surrounding mutations on the proteins activity is unclear.

The *adhesin-like protein* coding gene was inserted between the *yahA* and the *betT* genes and comparing the genomic sequences within this region, the only difference between the two variants in the strains was the presence of the *adhesin-like protein* gene. It is difficult to draw any conclusions regarding this change being related to any specific pathogenic strain, as a number of commensals had this insertion and the subsequent mutations as well. While the K-12 strains did not have this insertion, all EHEC strains involved in this study as well as all *E. coli* strains of the O104:H7 serotype including the May 2011 outbreak strain (LB226692) had it upstream of the *yahA* gene.

Results

the wild-type) was mutated into an A. This caused the 432nd codon to become a stop (TAG). In the EHECs O157:H7 str. EC4115 and O157:H7 str. TW14359, the disruption took place at the same position. A deletion of 5 nucleotides after the 1573rd nucleotide in the *yciR* gene sequence led to a frameshift and premature termination (525 AAs) relative to the wild-type version (661 AAs).

The *yfgF* gene was conserved in 56 strains. In the porcine ETEC UMNK88 the gene was disrupted by an insertion of an IS1 element encoding transposase proteins InsA and InsAB. The insertion consisted of 1779 nucleotides, occurred after the 776th nucleotide in the 594th codon and was flanked by the sequence 'CGTTTTTC.' It led to a frameshift and early termination after 602 AAs as compared to 747 AAs in the wild-type variant. In the strains O7:K1 str. CE10 (ExPEC, NMEC) and IAI39 (ExPEC) the point mutation was observed at the same position. The 1335th nucleotide was mutated from a C (wild-type variant) to an A. This resulted in a stop codon (TAA) in the 445th codon, truncating the protein to 445 AAs. A different point mutation affected the *yciR* gene in the commensal *E. coli* HS strain. It took place in the 616th nucleotide and changed a C (in the wild-type variant) to a T. The 206th codon became a stop codon (TAG) and the predicted protein was truncated to 205 AAs. The final affected strain is the APEC O1 which had a single nucleotide deletion (G) after the 2121st nucleotide in the 708th codon. The deletion resulted in a shift in the reading frame and early termination after 728 AAs.

The *yjcC* gene was conserved in 55 strains. Four of the affected strains [O127:H6 str. E2348/69 (EPEC), CFT073 (UPEC), ABU 83972 (commensal) and Nissle 1917 (commensal)] experienced a point mutation at the same position. The point mutation occurred in the 1552nd nucleotide and mutated a C in the wild-type variant to a T. The result was the early termination of the *yjcC* gene after the 518th codon (TAA). This is slightly shorter than the wild-type variant of 528 AAs and is unclear how this truncation affects protein stability or activity. A different point mutation was observed in the ETEC H10407 strain, where the 1240th nucleotide was mutated from a C (wild-type variant) to a T. This led to the 414th codon becoming a stop codon (TAG) and early termination of the predicted protein. Finally, the disruption of the *yjcC* gene in the commensal REL606 strain was also due to a point mutation, this time in the 934th nucleotide. There a C in the wild-type variant is replaced by a T, which turned the 312th codon into a stop (TAG).

The *yliE* gene was also conserved in 55 strains. In the EHEC O111:H- str. 11128 the *yliE* gene was augmented by an in-frame 9 nucleotide insertion that extended the predicted proteins length to 785 AAs vs 782 AAs in the wild-type variant. The insertion took place after the 1210th nucleotide in the 404th codon and was a repetition of the sequence immediately following it downstream (TGGAAGCCG). In this region the protein sequence in the wildtype variant was 'EAVEAV' and in the O111:H- str. 11128 it was 'EAVEAVEAV.' The *yliE* gene in *E. coli* strain

Results

O83:H1 str. NRG 857c (AIEC) was corrupted by a point mutation after the 1684th nucleotide, where a G (wild-type variant) got changed to a T. This turned the 562nd codon into a stop codon (TAG). In the strains: REL606 (commensal), BL21 (DE3) (commensal) and 'BL21-Gold (DE3) pLysS AG' (commensal) the point mutation occurred at the same position. The 151st nucleotide was changed from a C (wild-type variant) to a T and the 51st codon became a stop (TAG). The final affected strain was SE15 (commensal) where a different point mutation occurred in the 1337th nucleotide. The T in the wild-type variant was mutated to a G and caused a stop in the 446th codon (TGA).

The *dosP* (yddU) gene was missing as a whole in EHEC O103:H2 str. 12009. The five EHECs of the O157:H7 serotype suffered from the same deletion event, where 10 nucleotides were deleted after the 1214th nucleotide in the 405th codon. This led to a frameshift and early termination after 413 AAs (compared to 799 AAs in wild-type variant). A different deletion was observed in the O7:K1 str. CE10 (ExPEC, NMEC) and IAI39 (ExPEC). Here the deletion of an A took place after the 1662nd nucleotide in the 555th codon and resulted in a frameshift and early termination after 556 AAs. In the UPEC CFT073 strain two insertion events were observed. The first one was of a T and took place after the 1312th nucleotide in the 438th codon, the second one was of a G and took place after the 2311th nucleotide in the 771st codon relative to the wild-type variant. The first insertion was enough to cause a frameshift and early termination after 438 AAs.

The *ycgG* gene was affected in 24 strains and was the least conserved gene out of all of the DGC/PDE encoding genes examined. The gene was missing in eight strains: all five EHECs of the O157:H7 serotype, UM146 (AIEC), UTI89 (UPEC) and K-12 substr. MDS42. In the *E. coli* UMN026 (ExPEC) strain a point mutation after the 81st nucleotide disrupted the *ycgG* gene. The mutation turned a G (wild-type variant) into an A, which generated a stop codon in the 27th codon. A deletion of a single nucleotide (G) in the UMNK88 (porcine, ETEC) strain led to frameshift and early termination after 307 AAs. The deletion took place after the 890th nucleotide in the 297th codon. In 14 strains the deletion event was observed at the same position which in Table 4.2 is denoted by an '*.' The deletion entailed the loss of the first 630 nucleotides of the *ycgG* sequence relative to the wild-type strain and continued further -41bps upstream of the ATG start site, and included the Shine-Dalgarno sequence. The deleted 5' end of *ycgG* encoded the N-terminal sensory domain, including the CSS motif. Based on the work of Spurbeck *et al.*, (2012) this truncated version of the YcgG protein (283 AAs) is expressed in the UPEC CFT073 strain and is an active PDE (Spurbeck *et al.*, 2012). If this is also the case for the other strains with this variant, an interesting phenomenon would be uncovered (Discussion section 5.3). Finally, the O127:H6 E2348/69 (EPEC) had an additional deletion in the 3' end of the *ycgG* gene after the 1506th nucleotide in the 502nd codon relative to the wild-type variant, so that a frameshift was observed and a later stop codon. This strain is among the strains with the potentially truncated version of

the YcgG protein and this deletion would then extend the encoded protein to 285 AAs, assuming a new start site exists.

4.1.2.1 Four novel EAL domain proteins in pathogenic *E. coli*: PdeT (VmpA), PdeX, PdeY and PdeZ

A novel EAL domain protein was discovered; it is described as Rtn-like in the NCBI database and is here named PdeT, for phosphodiesterase and ‘T’ since it seems to be in the same operon as *ycdT* (Fig. 4.1). All of the EHECs of the O157:H7 serotype had this additional protein as well as the EPECs: *E. coli* O55:H7 str. CB9615 and *E. coli* O55:H7 str. RM12579. Table 4.4 lists the designations and GI numbers of the novel EAL domain-containing proteins within the native strains. PdeT was most similar to the Rtn protein out of all of the compared EAL proteins (Fig. 4.26) and the two proteins were 31% identical (data not shown). PdeT shared 100% identity (536/536) between the seven strains. Recently, Branchu *et al.*, (2013) have demonstrated that PdeT (there named VmpA) from the EHEC EDL933 strain is an active PDE and is encoded in the same operon as *ycdT* (Branchu *et al.*, 2013). Since the operon encoding these genes is identical in the other EHEC strains of the O157:H7 serotype, it is a valid assumption that the PdeT in those strains is also active. In the K-12 and ExPEC *E. coli* strains YcdT was encoded as its own transcriptional unit, while in the UPEC *E. coli* strains another gene (P4 family integrase) was encoded immediately downstream of *pdeT* and the two genes seemed to be within the same operon. In all of the strains the *pdeT* encoding region was adjacent to the *pgaABCD* locus (Fig. 4.18).

Table 4.4: The novel EAL domain-containing proteins and their native in strain designation.

| Designation in this study | <i>E. coli</i> Strain | Designation in strain | GI number |
|---------------------------|-----------------------|----------------------------|-----------|
| PdeT | O157:H7 str. TW14359 | ECSP_1197 | 254792284 |
| | O157:H7 str. EC4115 | ECH74115_1268 | 209399126 |
| | O157:H7 str. EDL933 | Z1528 | 15801017 |
| | O157:H7 str. Sakai | ECs1272 | 15830526 |
| | O157:H7 str. Xuzhou21 | CDCO157_1207 | 387881786 |
| | O55:H7 str. CB9615 | Rtn-like protein GI1N-1259 | 291282023 |
| | O55:H7 str. RM12579 | ECO55CA74_06170 | 387506136 |
| PdeX | 536 | ECP_2965 | 110643119 |
| PdeY | 536 | ECP_0300 | 110342098 |
| | IHE3034 | ECOK1_1105 | 386598807 |

Results

| Designation in this study | <i>E. coli</i> Strain | Designation in strain | GI number |
|---------------------------|-----------------------|-----------------------|-----------|
| | CFT073 | c1246 | 26247120 |
| | UTI89 | UTI89_C1116 | 91210145 |
| | UM146 | UM146_12325 | 386605046 |
| | ABU 83972 | ECABU_c12040 | 386638503 |
| PdeZ | E24377A | EcE24377A_E0054 | 157149510 |

In the TW14359 (EHEC) and EDL933 (EHEC) strains, PdeT was 536 AAs long. In the Sakai (EHEC) strain it was annotated to be 510 AAs long, while in the EC4115 (EHEC) strain it was annotated to be 504 AA but this difference in length turned out to be artificial as different start codons were annotated for each of the variants and when the DNA sequence was examined it was identical among these strains. Based on the Shine-Dalgarno sequence, the start codon chosen, as well as the subsequent topology, in the 536 residue variation was the most probable. In four STEC strains of the O145:H28 serotype (RM13516, RM13514, RM12761 and RM12581) only the last half of the *pdeT* gene was found and the *ycdT* gene was missing altogether due to a large deletion, making it unlikely that PdeT would be expressed.

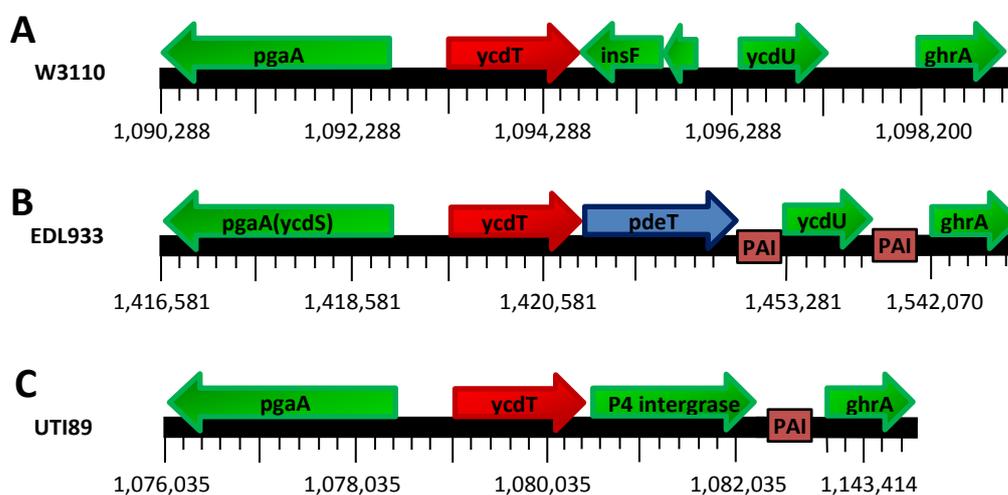
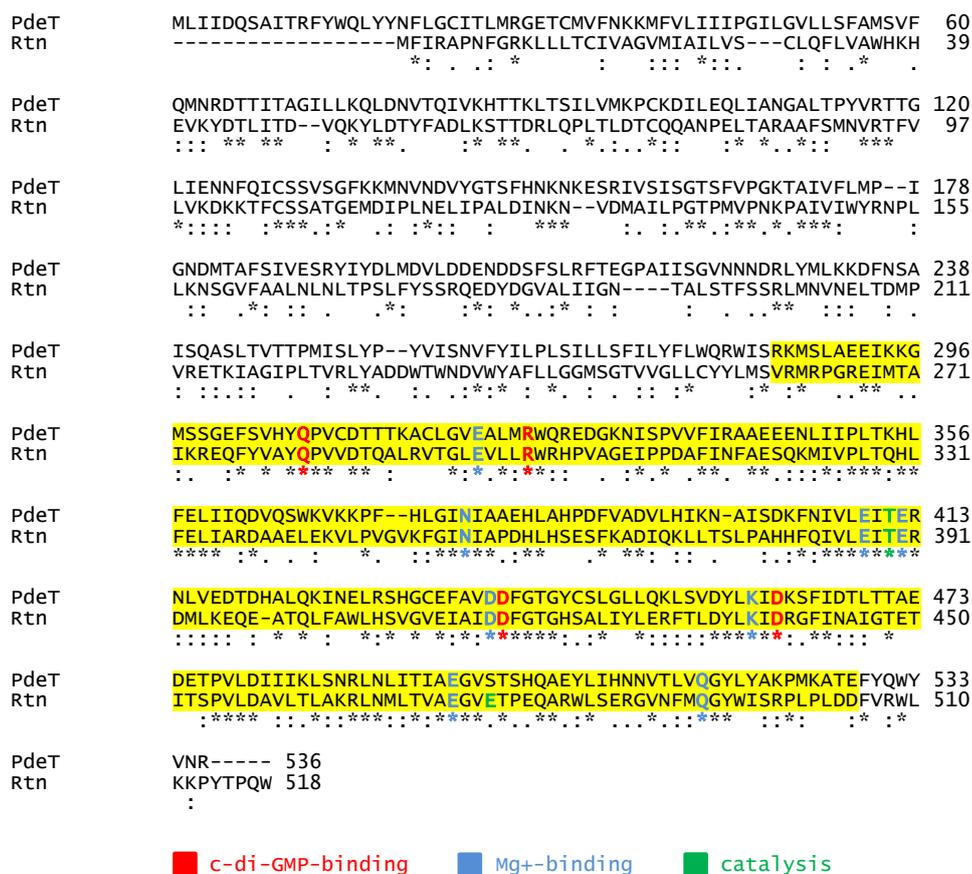


Figure 4.18: Operon layout of the *ycdT* gene. A) Depicts the *ycdT* operon layout for *E. coli* str. K-12 substr. W3110 and is a representative layout for all K-12 strains. Here the *ycdT* is its own transcriptional unit. B) depicts the layout found in *E. coli* O157:H7 EDL933 (EHEC) and is identical to the layouts found in all other EHEC O157:H7 strains, where *ycdT* is in the same operon as the novel PDE gene *pdeT*. In O55:H7 str. EPECs: CB9615 and RM12579, the layout is similar to that in EHECs, only the *ycdT* gene has a 69 nucleotide insertion that leads to frameshift and early termination (not pictured). C) is the layout of the *ycdT* operon in the UPEC *E. coli* UTI89 and is identical to the layout of the other UPECs. Here, *ycdT* exists in the same operon as a P4 family integrase protein, named UTI89_C1089 in the UTI89 strain.

Results

The EAL domain of PdeT is made up of 238 AAs. According to the NCBI database the protein also contains the CSS domain in the N-terminal sensory domain which spans an interval of 223 AAs from 61-284 residues. The CSS domain is associated with PDEs and 5 other PDEs in this study also contain the CSS domain (Table 4.1). PdeT was predicted by the TMHMM program to have 4 TMSs, but based on its homology to Rtn (2 TMSs) and the location of the CSS domain, it seems more probable that PdeT also has 2 transmembrane segments. PdeT clusters closest with the Rtn protein, followed by YoaD (AdrB) (Fig. 4.26). The alignment of PdeT and Rtn is shown in figure 4.19 with the EAL domain highlighted in yellow. Close alignment of the N-terminal is due to both of the proteins containing the CSS domain.



Results

the Mg^{2+} binding site which are required for PDE activity and had all four of the amino acids associated with c-di-GMP binding (Table 4.1). Based on sequence criteria, PdeX looks to be an active PDE and is clustered most closely with the YliE protein (Fig. 4.26) specifically its EAL domain as PdeX was only 260 AAs long and consisted only of the EAL domain (Fig. 4.21).

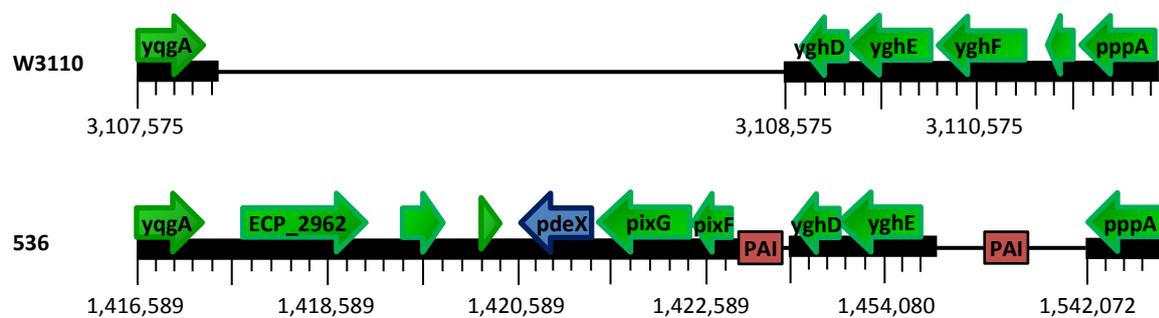


Figure 4.20: Genomic alignment of *pdeX* in *E. coli* 536 (UPEC) with reference strain W3110. The location of the novel gene *pdeX* (*ECP_2965*) is highlighted.

Results

The PdeZ protein was found in *E. coli* E24377A (ETEC). PdeZ contained 226 residues and consisted of only the EAL domain. The protein was actually encoded on uncharacterized plasmid (plasmid 2) which was composed of 71,915 nucleotides. Plasmid 1 of *E. coli* E24377A contained heat labile enterotoxin genes and plasmid 5 had genes for streptomycin resistance (Ecocyc database). PdeZ was encoded as its own transcriptional unit but its closest neighboring gene upstream encoded the YdeO protein [not to be confused with YdeO (DcyD) found in W3110 chromosomal DNA], a HTH-type transcriptional regulator and the gene downstream of PdeZ encoded a resolvase domain-containing protein (EcE24377A_E0053) (Fig. 4.24). The closest homologue of PdeZ, which aligned throughout the proteins length, was found in *E. fergusonii* ECD227 and from the EAL domain-containing proteins analyzed in this study it clustered most closely with CsrD and YciR proteins on the phylogenetic tree (Fig. 4.26, Fig. 4.25). PdeZ had one mutation in the amino acid sequence required for c-di-GMP binding (E/R/D/D instead of Q/R/D/D) and one mutation in the amino acid sequence required for magnesium binding (L instead of N). How these mutations affect PDE activity is unclear. YhjH's EAL domain also has two mutations in its c-di-GMP binding site and two mutations in the non-contiguous sequence that makes up the Mg²⁺ chelating site (see Table 4.1 for the exact mutations) and has been shown to be an active PDE capable of binding c-di-GMP (Pesavento *et al.*, 2008).

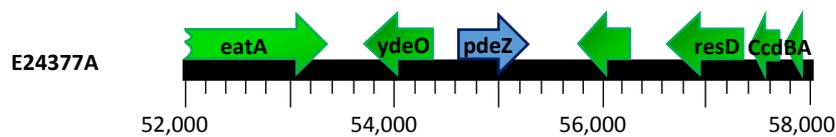


Figure 4.24: Genomic layout of *pdeZ* in *E. coli* E24377A (ETEC). The location of the novel gene *pdeZ* (*ecE24377A_E0054*) is indicated by the blue arrow. *pdeZ* is encoded on Plasmid 2, i.e. on an extrachromosomal element, absent in K-12 strains.

Results

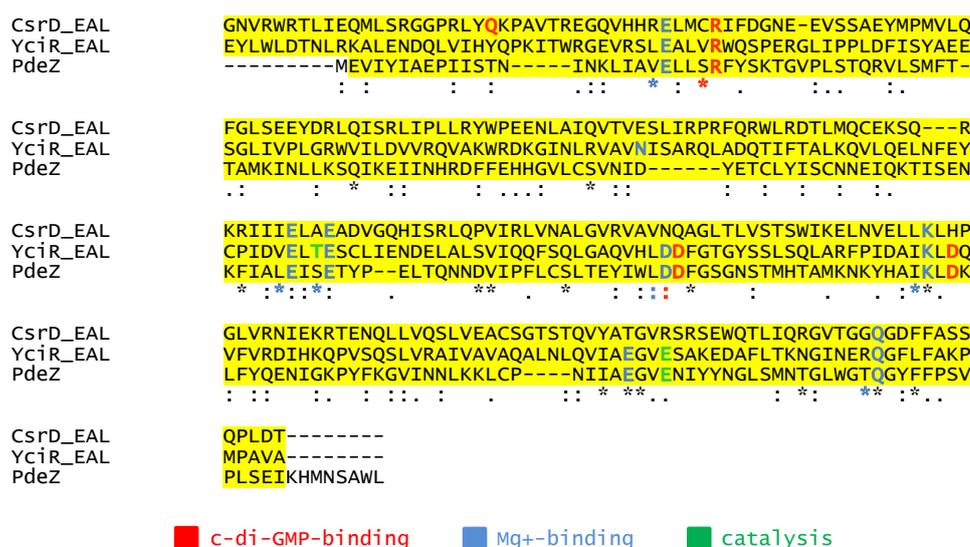


Figure 4.25: EAL-domain sequence alignment of the novel PdeZ (putative PDE) and the EAL domains of its closest homologs CsrD (degenerate GGDEF/EAL domain protein) and YciR (PDE), as identified by the SMART EMBL database. The alignment was generated using CLUSTAL X (Thompson *et al.*, 1997) and then edited manually to highlight key amino acid positions associated with PDE activity. PdeZ consists of 226 residues which encompass the EAL domain, thus the alignment with CsrD and YciR highlights the similarity of the three proteins' EAL domain.

All three of the novel PDEs, PdeX, PdeY and PdeZ had no additional sensory domains and were large enough to only contain the EAL domain. Both PdeX and PdeZ were found once within all 61 *E. coli* genomes analyzed and did not appear in any other bacteria. One can conclude that these may be very recent additions to the *E. coli* genome, especially as PdeZ was found only on a plasmid and not within the chromosomal DNA of its carrier strain.

4.1.2.2 Phylogenetic distribution of EAL domain-containing proteins

Figure 4.12 shows the phylogenetic relationship based on the entire protein amino acid sequence of all of the EAL domain-containing proteins. Therefore the difference in the N-terminal sensory domains dictates the phylogenetic relationship between the proteins. The included 20 proteins group in 8 clusters, numbered 1-8 in figure 4.26. The first cluster consists of the proteins YjcC, YlaB, PdeX and YliE, with YjcC and YlaB clustering closer together and PdeX and YliE clustering closer together. Both YjcC and YlaB have 2 TMSs and contain the CSS domain. YliE also has 2 TMSs in its N-terminus and aligns with PdeX in its C-terminus (EAL-domain) as PdeX is 246 AAs long and consists of only the EAL domain. Thus the four proteins may be clustering together due to their EAL domain alignment, while YliE is clustering close to YjcC and YlaB because of its 2 TMS topology.

Results

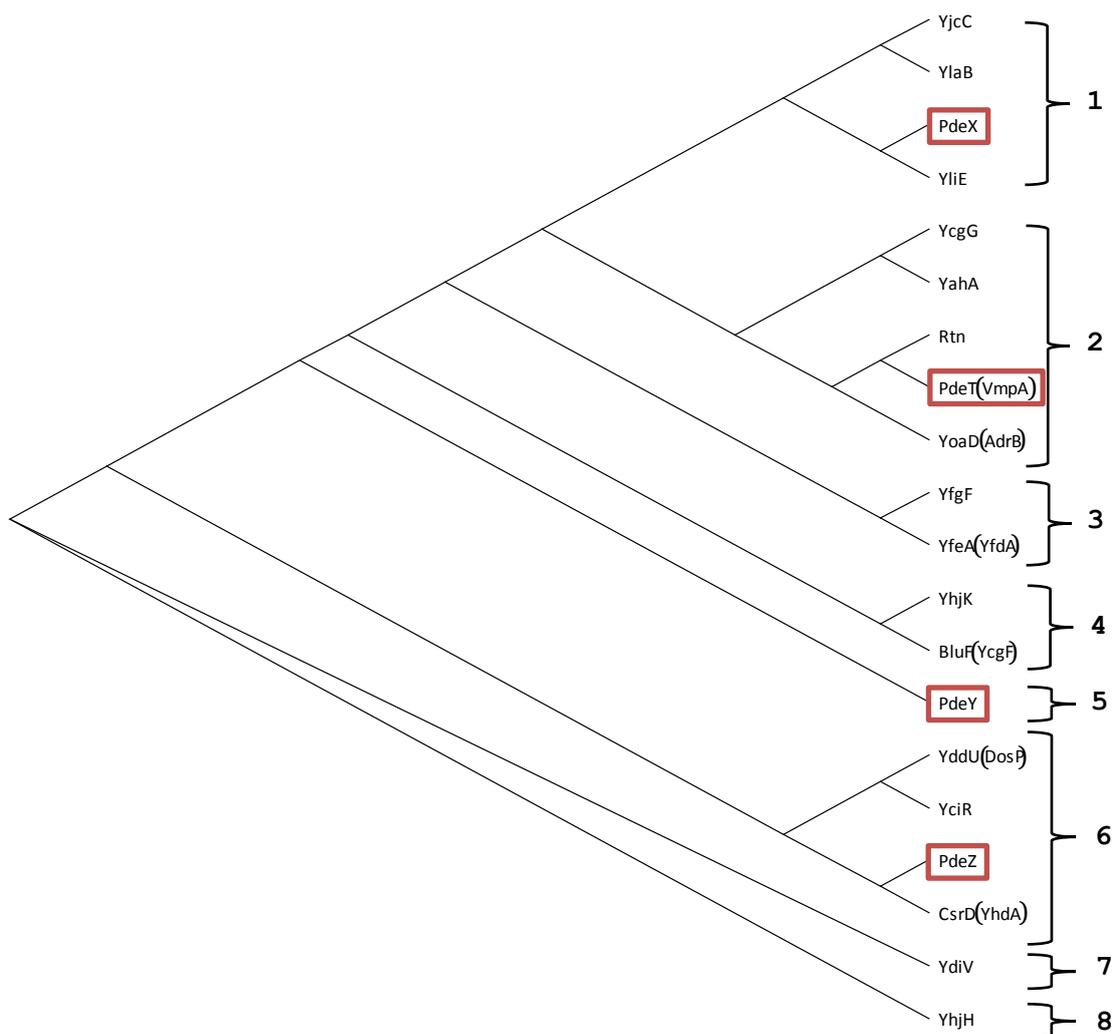


Figure 4.26: Phylogenetic tree of the EAL domain-containing proteins (20). The tree is based on the Clustal X multiple alignment of EAL domain-containing proteins in their full entirety and drawn with the TreeViewX program. The novel proteins are boxed in red. The 20 proteins cluster in 8 clusters, numbered 1-8. The protein CsrD has a degenerate GGDEF domain and a degenerate EAL and was included in both of the phylogenetic trees.

The proteins YcgG, YahA, Rtn, PdeT (VmpA) and YoaD (AdrB) can be found in cluster 2. YcgG and YahA, and Rtn, PdeT and YoaD form a subcluster. Four of these five proteins (YcgG, Rtn, PdeT and YoaD) have 2 TMSs and contain the CSS domain. YahA contains the lux-c-like DNA binding domain in its N-terminus and based on its close clustering with CSS domain-containing proteins, the two domains may have similar features.

Cluster 3 consists of YfgF and YfeA (YfdA) which both contain the MASE1 domain in their N-terminus. Cluster 4 is made up of YhjK and BluF (YcgG). YhjK contains the HAMP sensory domain and BluF contains the BluF sensory domain. It is unclear why the two cluster together and the relationship between the two domains needs further study.

Clusters 5, 7 and 8 each consist of single proteins, PdeY, YdiV and YhjH, respectively. These proteins did not align in a significant manner with any of the other EAL domain-containing

Results

proteins found in *E. coli*. YdiV is a degenerate GGDEF/EAL domain-containing protein and this degeneration may account for its dissimilarity from the others. PdeY and YhjH consist of only their EAL domains and have no sensory domains. As the phylogenetic tree is generated from the alignment of all of the EAL domain-containing proteins, the similarity of the EAL domains in YhjH and PdeY were not enough to group them with any of the other proteins analyzed in this section.

Cluster 6 is made up of four proteins, YddU (DosP), YciR, PdeZ and CsrD (YhdA) with YddU and YciR forming a subcluster and PdeZ and CsrD forming another one. Both YddU and YciR share the PAS domain. As shown in figure 4.25, PdeZ aligns well with CsrD and YciR but specifically in the EAL domain region because PdeZ is 226 residues long and consists of the EAL domain.

4.1.3 Degenerate GGDEF/EAL domain proteins

Genes encoding degenerate GGDEF/EAL domain-containing proteins were conserved with varying frequency. Figure 4.27 displays the frequency of protein conservation among the 61 analyzed strains. Only proteins for which the gene is free from deleterious corruptions were included in the figure. All corruptions are described thoroughly in the results and were compared in reference to the wild-type variant from the K-12 W3110 strain.

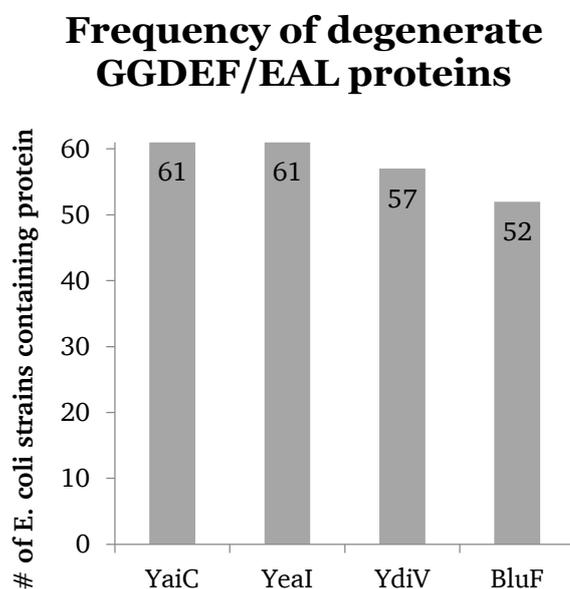


Figure 4.27: Frequency of protein conservation of degenerate GGDEF/EAL domain-containing proteins in the 61 *E. coli* strains analyzed. Displayed are the proteins in relation to the number of times their genes are found fully intact.

There were only four proteins that fell within this category and of the four, the genes encoding two of these proteins were conserved in all of the 61 strains analyzed; *yeaI* and *csrD* (*yhdA*). CsrD had both the GGDEF and EAL domains present but the sequences essential for

Results

enzymatic activity were degenerate in both of the domains. The GGDEF domain's A-site sequence was mutated to HRSDF and the EAL domain had mutations in two of the amino acids needed for c-di-GMP binding and two mutations in the Mg²⁺ chelating site (see Table 4.1). These mutations appear to be critical as CsrD has no PDE activity, instead promoting motility via its indirect regulation of CsrA (Babitzke *et al.*, 2007). YeaI contained a GGDEF domain with a degenerate A-site (EGEVF vs GGDEF) but an intact I-site (see Table 4.1) and its function remains to be elucidated.

The *ydiV* gene was disrupted in four strains. In the ETEC H10407 strain a point mutation occurred after the 247th nucleotide, where a G in the wild-type variant was replaced by an A. This led to the 83rd codon becoming a stop codon (TAG). In the commensal strains: REL606, BL21(DE3) and 'BL21-Gold(DE3)pLysS AG' a deletion event was observed at the same location. The deletion of a single T in a string of seven T's after the 218th nucleotide in the 73rd codon resulted in a frameshift and early termination after 85 AAs.

The *bluF* gene was conserved in 52 of the 61 analyzed strains. The gene was altogether missing from six strains (the five EHECs of the O157:H7 serotype and the commensal K-12 substr. MDS42). In the ETEC H10407 strain a point mutation occurred after the 811th nucleotide, where a C in the wild-type variant was replaced by a T. This led to the 271st codon becoming a stop codon (TAA). In the EPEC O55:H7 str. CB9615 a deletion of 4 nucleotides after the 716th nucleotide in the 238th codon led to a frameshift and early termination after 244 AAs. In *E. coli* O7:K1 str. CE10 (ExPEC, NMEC) an insertion of an IS3411 insertion element disrupted the *ydiV* gene. The insertion was 1313 nucleotides, encoded transposases A and B, and was flanked by the sequence 'ATT.' It occurred after the 731st nucleotide in the 244th codon and led to frameshift and early termination after 259 AAs.

The five O157:H7 EHECs were missing the entire genomic region surrounding *bluF*, including the *bluF* gene itself. The five O157:H7 strains were missing the region of 1197950-1225272 bp of the *E. coli* K-12 str. W3110. Not only are the operons containing the *ycgG* and *bluF* genes located there but also operons from a ϕ 14 prophage. This genome region was flanked by *rhuE* and *icd* genes on the 5' end and *minCDE* operon on the 3' end. These flanking genes were also found in the O157:H7 EHECs but the genomic content that these genes were flanking was different between the EHECs and the K-12 W3110. In the EHECs the majority of the genes in this region were from prophage CP-933X. The span of the actual region flanked by *rhuE* and *minCDE* in the O157:H7 EHECs differs slightly between the five strains. In the TW14359 strain the region spanned from 1555175-1601039 bp; in the Sakai strain the region spanned from 1918212-1665394 bp; in EC4115 the spanning region was from 1554887-1602065 bp; in EDL933 the spanning region was from 1702185-1756800 bp.

4.1.4 Biofilm matrix components and transcriptional regulators

4.1.4.1 Curli synthesis: *csgDEFG*, *csgBAC* and *mlrA*

csg genes and *mlrA* are important for curli formation and the *csg* genes were among the best conserved among the genes encoding biofilm matrix components. Figure 4.28 displays the frequency of protein conservation among the 61 analyzed strains. Only proteins for which the gene is free from deleterious corruptions were included in the figure. All corruptions are described thoroughly in the results and are compared in reference to the wild-type variant from the K-12 W3110 strain.

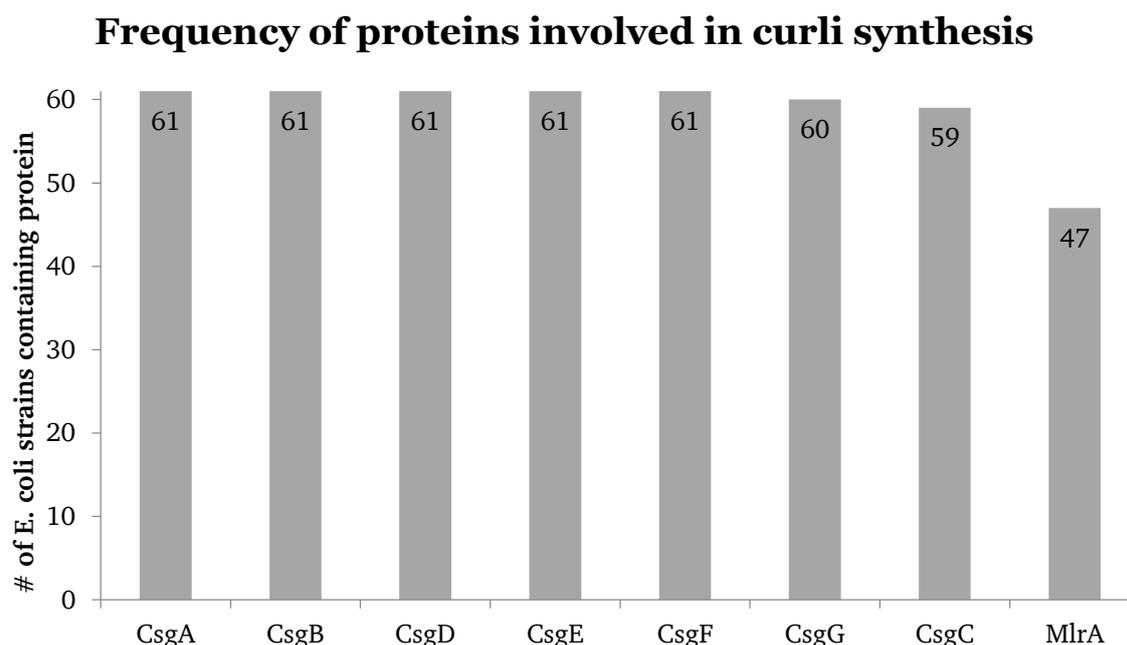


Figure 4.28: Frequency of protein conservation involved in curli synthesis in the 61 *E. coli* strains analyzed. Displayed are the proteins in relation to how often their genes are found fully intact.

The *csg* genes: *csgD*, *csgE*, *csgF*, *csgB* and *csgA* were conserved in all 61 strains. The *csgG* gene was only disrupted in a single strain, the UPEC 536. A single nucleotide insertion of a T after the 218th nucleotide in the 73rd codon led to a frameshift and early termination after 105 AAs. The *csgC* gene was disrupted in two cases: in the STECs RM13516 and RM12761. The same insertion event corrupted the gene in both of the strains. The insertion consisted of 1314 nucleotides, occurred after the 267th nucleotide in the 90th codon and brought with it a stop codon (TGA) at the site of the insertion so that the predicted protein would terminate after 89 residues. The insertion was flanked by the sequence ‘TGTT’ and was identified as outer membrane protein precursor Lom by the NCBI database. It is unclear what result this truncation of the protein may have on its function, the CxC motif that is hypothesized to be necessary for its redox function (Taylor *et al.*, 2011) is still intact in the truncated version.

Results

The gene encoding the transcriptional regulator MlrA was affected in 14 instances. The *mlrA* gene was disrupted by prophages in the same location in 12 strains (all five O157:H7 serotype EHECs: EC4115, EDL933, Sakai, TW14359, Xuzhou21; the STECs of the O145:H28 serotype: RM13514, RM12581, RM12761, RM13516; ExPEC IAI39, and the commensal ED1a). The insertion took place after the 68th nucleotide in the 23rd codon. The prophage insertion was accompanied by a deletion event of 222 bps corresponding to nucleotides 69-290 relative to the wild-type variant. The remaining part of the *mlrA* was found following the prophage insertion. The prophage size varied by strain, the STECs RM13514 and RM12581 had a prophage consisting of 37,121 nucleotides. The STECs RM12761 and RM13516 had a prophage consisting of 57,281 nucleotides. The EHECs Xuzhou21, Sakai and EDL933 had a prophage consisting of 47,862 nucleotides, while the EHEC TW14359 had a prophage consisting of 48,305 nucleotides and the EHEC 4115 had a 93,728 nucleotide prophage. The EPEC RM12579 had a 51,119 nucleotide prophage and the ExPEC IAI39 had a prophage that was 17,440 nucleotides. Finally, in the commensal ED1a the prophage consisted of 62,850 nucleotides.

In the O7:K1 str. CE10 (ExPEC, NMEC) strain had a prophage insertion of 36,979 nucleotides which occurred after the 60th nucleotide in the 21st codon. In the *E. coli* HS (commensal) strain an 8 nucleotide insertion disrupted the *mlrA* gene after the 535th nucleotide in the 179th codon. The disruption led to a frameshift and later termination after 262 AAs compared to 243 AAs in the wild-type variant.

It turns out that all of the EHECs of the O157:H7 serotype and STEC strains analyzed in this study have a lambdoid phage insertion carrying the *stxIA* and *stxIB* genes (encoding shiga toxin) in the *mlrA* (*yehV*) gene (Fig. 4.29). In fact, according to the work of Serra-Moreno *et al.*, (2007) this is the preferred locus for insertion of shiga toxin carrying phages in the O157:H7 strains (Serra-Moreno *et al.*, 2007; Shaikh & Tarr, 2003). In the database the *mlrA* gene is annotated as intact in the O157:H7 strains but upon inspection of the nucleotide sequence it is clear that the prophage insertion disrupts the *mlrA* gene as further confirmed by the work of Uhlich *et al.*, 2013.

Results

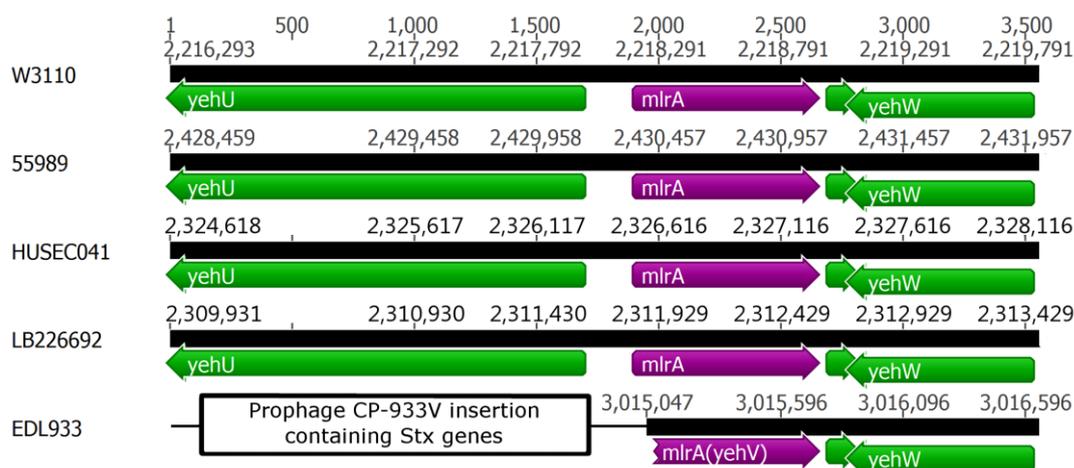


Figure 4.29: Alignment of genomic regions of the *mlrA* gene compared in 5 strains of *E. coli*. At the top is the K-12 W3110 strain used as a reference, followed by the EAECs 55989, HUSEC041, the outbreak strain LB226692 and the EHEC EDL933. The *mlrA* gene (purple) shows the insertion of prophage after codon 28 of the *mlrA* gene in the EHEC O157:H7 EDL933 strain, thus disrupting the gene. The same insertion is found within all EHECs of the O157:H7 serotype analyzed in this study.

4.1.4.2 Cellulose biogenesis: *bcsQABZC* and *bcsEFG*

The *bcs* genes are involved in cellulose biogenesis. The *bcsQABZC* operon contains the cellulose synthase genes (*bcsA* and *bcsB*) as well as several accessory factors (see 1.3.7). The *bcsEFG* operon is also essential for cellulose formation. Non-polar deletions in each of the three of the genes led to loss of cellulose formation in macrocolonies (Richter & Povolotsky, submitted).

The *bcs* genes were conserved with varying frequency. Figure 4.30 displays the frequency of protein conservation among the 61 analyzed strains. Only proteins for which the gene is free from deleterious corruptions were included in the figure. All corruptions are described thoroughly in the results and apart from one exception were compared in reference to the wild-type variant from the K-12 W3110 strain. The exception is the *bcsQ* gene where the wild-type variant used for gene comparison came from the EAEC 55989 strain.

The *bcsG* gene was affected in three strains. In the strains O7:K1 str. CE10 (ExPEC, NMEC) and IAI39 (ExPEC) the same type of insertion event took place in the same location. The 1426 nucleotide insertion occurred after the 361st nucleotide in the 121st codon of the *bcsG* gene which led to a frameshift and early termination after 125 AAs. The insertion identified as KpLE2 phage-like element/Transposase *insG* for insertion sequence element IS4 and DNA-binding transcriptional regulator CueR (DNA-binding transcriptional activator of copper-responsive regulon genes). It was flanked by the sequence ‘CAGATGATTGGGG.’ In the HS (commensal) strain a deletion of 1148 nucleotides occurred in the 5’ end of the *bcsG* gene. Alignment relative to the wild-type variant starts in the 383rd codon. The HS strain had a large deletion starting in the *bcsG* gene and ending in the *bcsC* gene so that all of the other *bcs* genes in between (see Fig. 1.6) were absent altogether.

Frequency of proteins involved in cellulose biogenesis

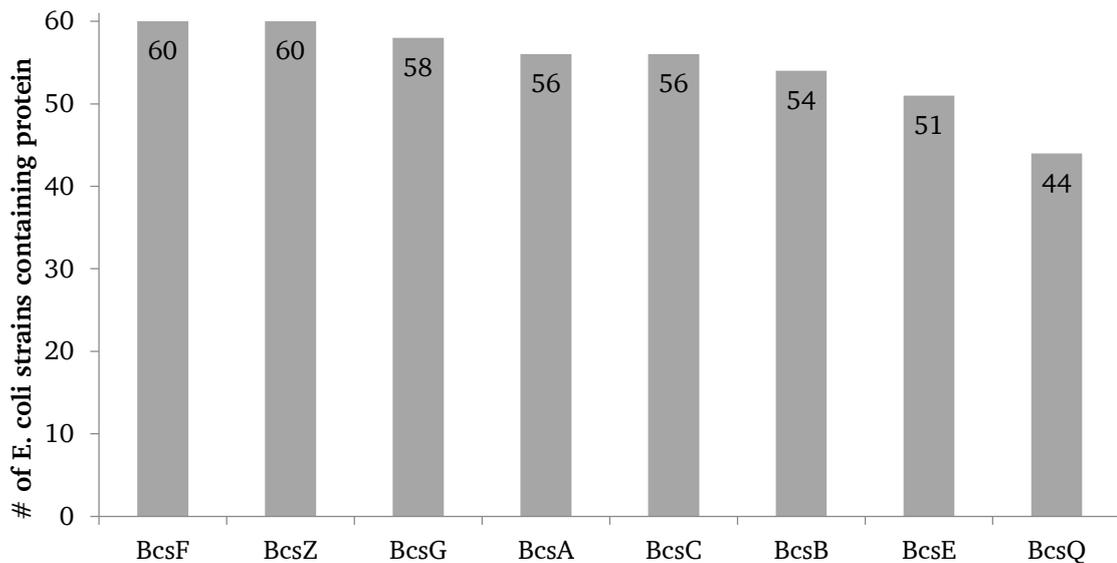


Figure 4.30: Frequency of protein conservation involved in cellulose biosynthesis in the 61 *E. coli* strains analyzed. Displayed are the proteins in relation to how often their genes are found fully intact.

The *bcsF* gene was only affected in the HS strain where it was completely absent. The *bcsE* gene was affected in 10 strains and completely absent from the genome of *E. coli* HS. A point mutation occurred in the commensal ATCC 8739 where the 303rd nucleotide was changed from a T in the wild-type variant to an A. This led to the 101st codon becoming a stop codon (TGA). In 4 of the EAECs of the O104:H4 serotype (2009EL-2050, 2009EL-2071, 2011C-3493 and LB226692) an insertion of a single C in a string of C's after the 1340th nucleotide in the 447th codon disrupted the *bcsE* gene. This led to a frameshift and early termination after 462 AAs. This insertion was confirmed by resequencing of a PCR fragment in the 2011 outbreak strain LB226692. The same deletion event affected all of the STEC strains (RM13516, RM13514, RM12761 and RM12581). The deletion affected the 5' end of the *bcsE* gene and consisted of the first 779 nucleotides. Alignment relative to the wild-type variant starts in the 260th codon. The deletion included the intergenic region between *bcsE* and *yhjR*, including all of *yhjR* and part of *bcsQ* (reported below).

The *bcsQ* gene was conserved in 44 strains. It was missing from the genome of the HS strain and suffered the same deletion in the four STEC strains (RM13516, RM13514, RM12761 and RM12581) where a deletion of the first 299 nucleotides occurred in the 5' end of the gene. Three different types of point mutations were observed in 11 strains that disrupted the *bcsQ* gene. In the commensal ED1a, the point mutation occurred after the 637th nucleotide that changed a C (wild-type variant) into a T and led to the 213th codon becoming a stop (TAA). In all five of the EHECs of the O157:H7 serotype (EC4115, EDL933, Sakai, TW14359, Xuzhou21) the point mutation

Results

took place in the 196th nucleotide where a C (wild-type variant) was changed into a T and the 66th codon became a stop (TAG). The last type of point mutation affected six strains, including all five of the K-12 derivatives (BW2952, DH10B, MG1655, W3110 and MDS42), and the commensal DH1. The point mutation changed the 17th nucleotide from a T (in the wild-type variant) to an A which led to the 6th codon becoming a stop (TAG).

The *bcsA* gene was affected in five strains and missing altogether from the genome *E. coli* HS. It had the same type of deletion in the strains of the O7:K1 serotype: IAI39 (ExPEC) and CE10 (ExPEC, NMEC). The deletion occurred after the 83rd nucleotide in the 28th codon and consisted of 19 noncontiguous nucleotides. The deleted nucleotides correspond to the nucleotides numbering 84-85, 88, 90, 92-94, 96-100, 103-104, 108-109, 111, 113 and 117 in the wild-type variant. This deletion led to a frameshift and early termination after 83 AAs before the stop codon was reached. In the ExPEC S88 a deletion of a G in a string of 6 G's after the 550th nucleotide in the 184th codon disrupted the *bcsA* gene and caused a frameshift and early termination after 185 AAs. Finally, in the ExPEC UMN026 two deletions and one insertion took place. The first deletion of an A in a string of 5 A's occurred after the 186th nucleotide in the 62nd codon. The second deletion of a G in a string of 2 G's took place after the 698th nucleotide in the 233rd codon relative to the wild-type variant. The insertion of an A occurred after the 702nd nucleotide in the 234th codon relative to the wild-type variant. These deletion/insertion events led to a frameshift and early termination after 89 AAs.

The *bcsB* gene was affected in seven strains and missing altogether from the genome of *E. coli* HS (commensal). It had a point mutation in the O111:H- str. 11128 (EHEC) strain where the 88th nucleotide was mutated from a C (wild-type variant) to a T. This led to the 30th codon becoming a stop (TAA). All of the EHECs of the O157:H7 serotype had in-frame insertion that would extend the length of the predicted protein. The insertion took place after the 2254th nucleotide in the 752nd codon. In the strains EDL933, Sakai and Xuzhou21, it consisted of 18 nucleotides which were a direct repetition of the immediately preceding nucleotide sequence 'TGCTGGCGGTGCTGGCGG.' In EC4115 and TW14359 the insertion went on for 36 nucleotides with the direct repetition occurring twice so that the nucleotide sequence 'TGCTGGCGGTGCTGGCGG' was encountered consecutively a total of three times in the *bcsB* gene.

The *bcsZ* gene was conserved in 60 of the 61 analyzed strains and was only affected in the *E. coli* HS, where it was missing altogether. Finally, the *bcsC* gene was affected in five strains. In the EHECs: EDL933, Sakai and Xuzhou21, the deletion was 10 nucleotides long and occurred after the 399th nucleotide in the 134th codon, leading to a frameshift and early termination after 166 AAs. In the ExPEC UMN026 the point mutation occurred in the 1195th nucleotide where an A (wild-type variant) was changed to a T which resulted in a subsequent stop codon (TAA) in the

399th codon. Finally, the *E. coli* HS was affected by a deletion in the 5' end that deleted the first 2603 nucleotides of the *bcsZ* gene. As described above this deletion spanned into the *bcsG* gene and included all of the genes in between (Fig 1.6).

4.1.4.3 PGA biosynthesis: *pgaABCD*

The *pga* genes were conserved with varying frequency. Figure 4.31 displays the frequency of protein conservation among the 61 analyzed strains. Only proteins for which the gene is free from deleterious corruptions were included in the figure. All corruptions are described thoroughly in the results and were compared in reference to the wild-type variant from the K-12 W3110 strain.

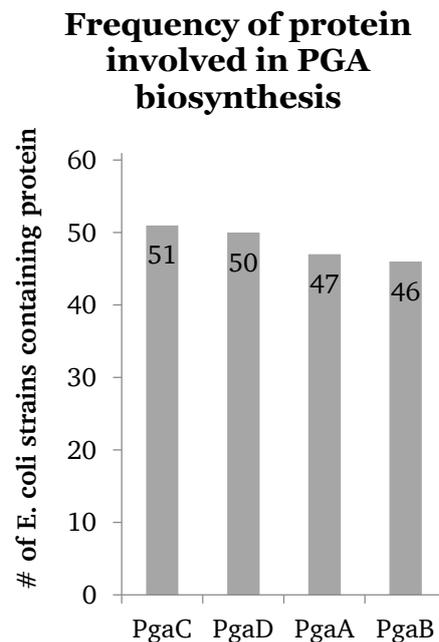


Figure 4.31: Frequency of protein conservation involved in PGA biosynthesis in the 61 *E. coli* strains analyzed. Displayed are the proteins and how often their genes are found fully intact.

The *pgaABCD* operon is required for the synthesis and excretion of the PGA exopolysaccharide (see Section 1.3.7 for mechanistic details) which is used as a biofilm matrix component by some *E. coli*, notably UPECs. In ten of the genomes investigated in this study, it was deleted as a whole operon and not deactivated by a mutation/deletion or insertion within a single gene. The ten affected strains were 042 (EAEC), UMNK88 (porcine ETEC), O111:H- str. 11128 (EHEC), O26:H11 str. 11368 (EHEC), O127:H6 str. E2348/69 (EPEC), UMN026 (ExPEC), SMS-3-5 (environmental), ED1a (commensal), IAI1 (commensal) and K-12 substr. MDS42 (commensal).

The *pgaA* gene was affected in 14 strains, including were it was missing altogether in the 10 strains listed above. It had a point mutation in the ETEC H10407 strain in the 790th nucleotide,

Results

where a G in the wild-type variant was changed to a T. The mutation caused the 264th codon to become a stop (TAA). In the ExPEC S88 the *pgaA* gene was also disrupted by a point mutation, this time in the 2424th nucleotide. At this location, a G in the wild-type variant was mutated into a T which led to the 808th codon becoming a stop (TAA). In the commensal strain DH1 (K-12 derivative) a 1199 nucleotide insertion disrupted the *pgaA* gene. The insertion occurred after the 1117th nucleotide in the 373rd codon and identified as ‘transposase *insH* for insertion sequence element IS5.’ The insertion was flanked by the sequence ‘CTAA’ and resulted in a shift in the reading frame of the *pgaA* gene and early termination after 379 AAs. In the commensal HS strain a different insertion disrupted the gene. It took place after the 777th nucleotide in the 453rd codon and was identified as IS1 insertion element. The insertion element was 777 nucleotides long and consisted of transposase *orfA* and transposase *orfB*. The insertion was flanked by the sequence ‘GGCCTCGTG.’

The *pgaB* gene was affected in 15 strains, including were it was missing altogether in the 10 strains listed above. In the STECs (RM13516, RM13514, RM12761 and RM12581) a deletion of a T in a string of 5 T’s after the 1141st nucleotide in the 381st codon caused a frameshift and early termination after 394 AAs. In the EPEC O55:H7 str. CB9615 a deletion of an A in a string of 7 A’s after the 788th nucleotide in the 263rd codon led to a frameshift and early termination after 290 AAs before a stop codon was reached.

The *pgaC* gene was only affected in the 10 strains where it was missing altogether (listed above). The *pgaD* gene was affected in 11 strains, including the 10 strains listed above where it was missing altogether. In the ETEC H10407 strain the *pgaD* gene was corrupted by a point mutation in the 364th nucleotide that changed a C (wild-type variant) into a T and turned the 122nd codon into a stop (TAG).

4.2 Regulation of newly identified DGCs and PDEs

Based on the results of the bioinformatic analysis, three proteins were chosen for further investigation. These were DgcX, YneF and PdeT (VmpA). Expression patterns of a gene can often give hints to their function. In order to study their regulation, single copy chromosomal *lacZ* fusions to all three genes were made. LacZ is used as a reporter gene that codes for β -galactosidase, and it was fused downstream of the promoter region, in frame with the gene encoding the protein of interest, in this case *dgcX*, *yneF* and *pdeT*. The created fusions were translational so that the activity of the promoter could be coupled to the genes’ of interest native Shine-Dalgarno sequences including the native start codons of the genes. A couple of beginning codons of the genes were fused in-frame with the *lacZ* gene. Thereby β -galactosidase expression was under control of the fused promoter and by performing β -galactosidase assays the activity

level of the expressed gene could be measured. A single copy of the chromosomal *lacZ* fusion was necessary to investigate gene regulation under physiological conditions, as there is only a single copy of the gene in the bacterial genome. The resulting *lacZ* reporter strains were grown under various conditions (28°C/37°C) and contained various background deletions (wt/ $\Delta rpoS::kan$ / $\Delta csgD::kan$) in order to try to uncover the mode of regulation of these genes. In *E. coli* RpoS is the general stress response regulator and enables the transition into the stationary phase and biofilm formation, while CsgD is an RpoS controlled transcriptional regulator that mediates the expression of both curli fibers and cellulose (through its activation of the DGC YaiC) (Hengge 2009). Since a number of the 28 characterized GGDEF/EAL domain-containing genes in the *E. coli* K-12 strain W3110 are under RpoS and CsgD control and in order to check whether the newly discovered genes are involved in biofilm formation, these two knockouts are routinely included in regulation testing.

4.2.1 Expression of *dgcX*

The *dgcX* gene was highly expressed at both 28°C and 37°C, though about twice as much at 28°C (Fig 4.34). It was the most strongly expressed DGC gene so far observed in *E. coli* (Fig 4.34). Due to its strong expression, it was tested in additional conditions (28°C/37°C/Salt/No Salt) and with additional background deletions [wt (cellulose-) $\Delta rpoS::kan$ / $\Delta csgD::kan$ /cellulose+/cellulose+ $\Delta rpoS::kan$ /cellulose+ $\Delta csgD::kan$]. The gene was equally well expressed in the presence and absence of salt and the background strain's ability to make cellulose had no effect on *dgcX* expression (data not shown). Based on primer extension experiments (Fig 4.33, right), DgcX had a classical vegetative (σ^{70} /RpoD) promoter (Fig 4.32) with a perfect -35 region consensus sequence (TTGACA) and a passable -10 region consensus sequence (TAGTAT). Figure 4.35 shows the expression of *dgcX::lacZ* in a $\Delta rpoS$ background, though still strong, to be half of that as in the wild-type background. This result suggested that *dgcX* was also partially regulated by the stationary phase regulator RpoS. In fact, RpoS is very flexible in the promoter sequence to which it can bind and has been shown to be able to bind to RpoD promoters (Typas *et al.*, 2007). Expression of *dgcX* starts in the exponential phase and is further induced during entry into the stationary phase in an RpoS-dependent manner (Fig 4.35).

Results

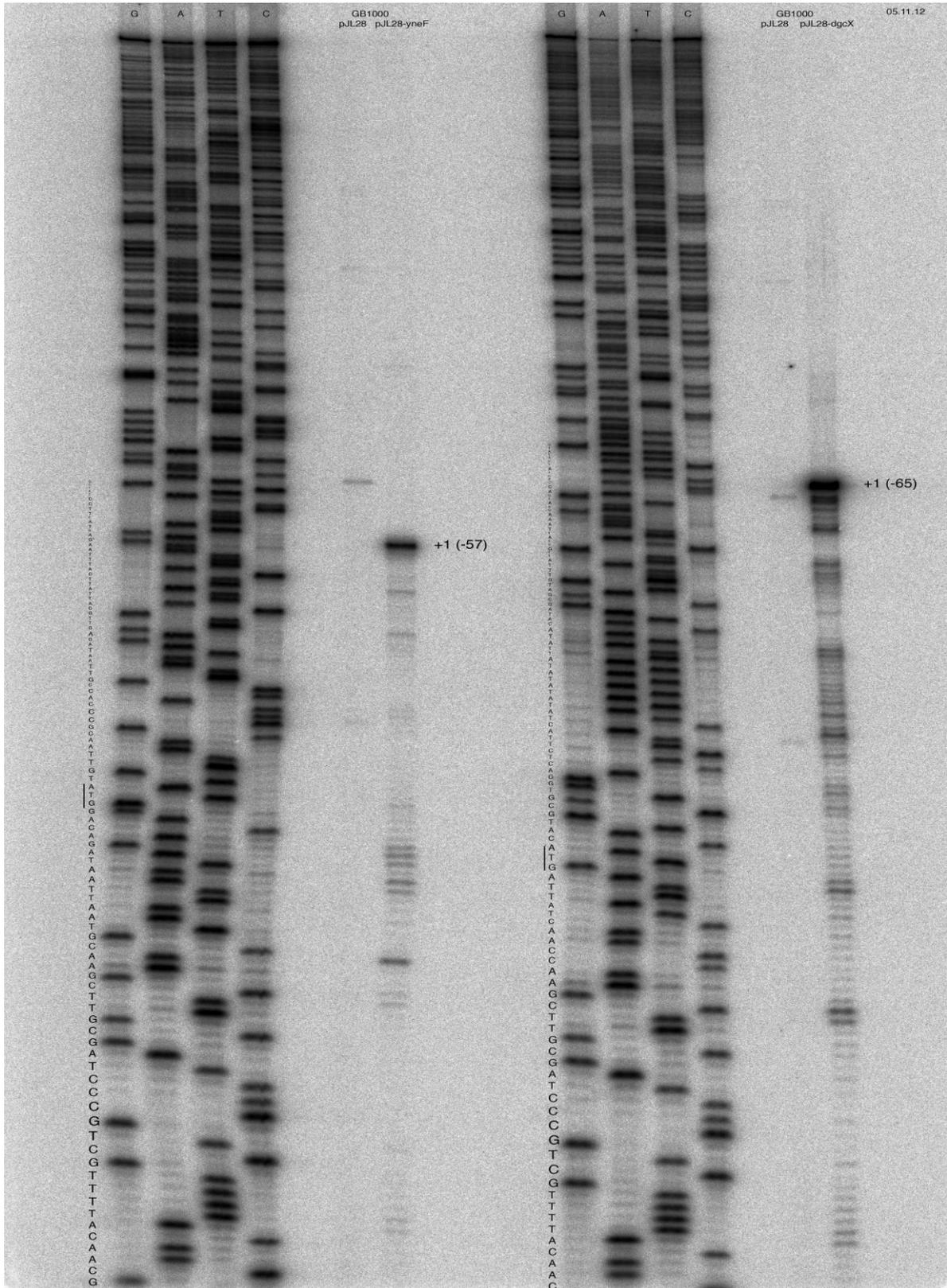


Figure 4.33: Determination of transcriptional start sites by primer extension for *yneF* (left) and *dgxX* (right) 5' end mRNA. Positions are numbered relative to the translational start site and the sequence is also displayed in figure 4.29. The underlined regions mark the start codon of the genes. (Experiment was performed by Alexandra Poßling and is shown here with permission).

Results

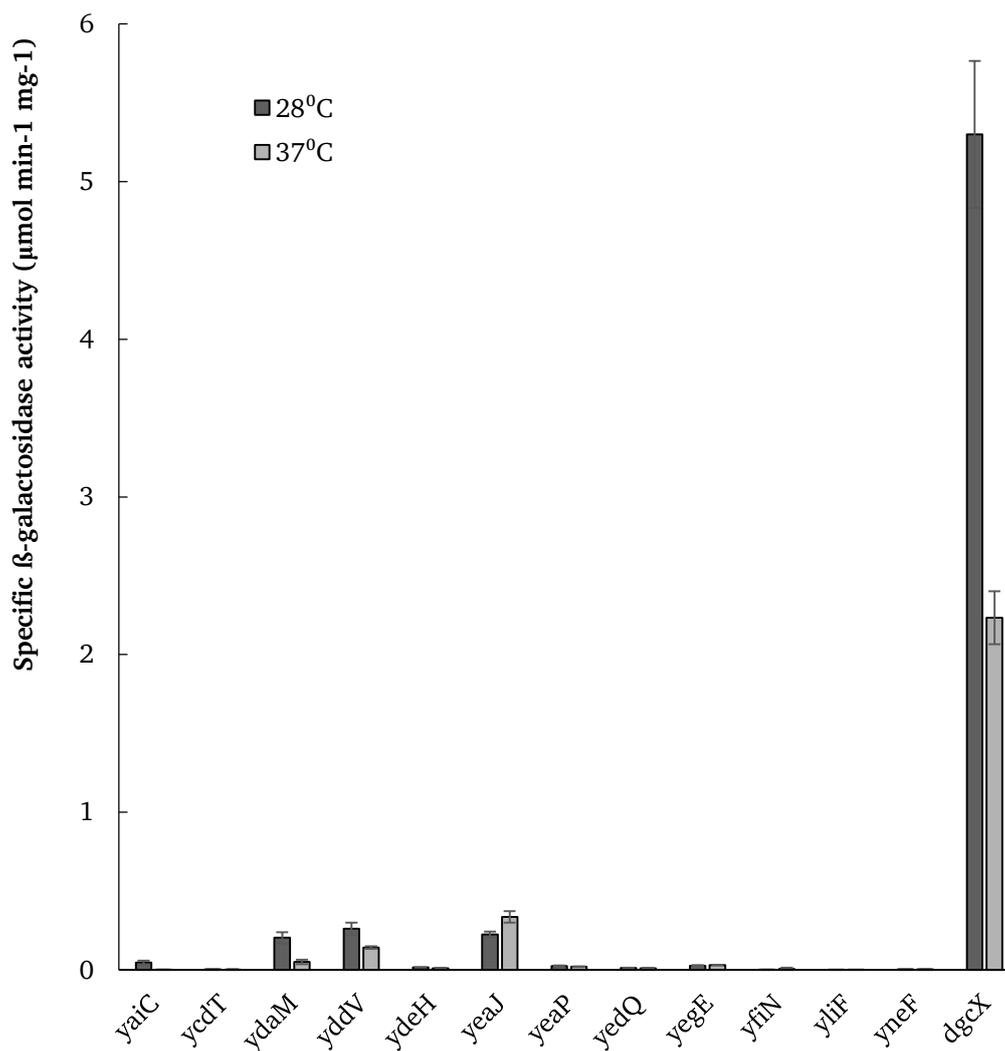


Figure 4.34: Expression of *lacZ* reporter fusions to *dgx* and other *E. coli* DGC genes. Derivative of strain W3110 carrying single-copy *lacZ* fusions in *dgx* and *yneF* (relevant sequences obtained from the O104:H4 outbreak strain) or in all other DGC genes previously described in *E. coli* K-12 (Sommerfeldt *et al.*, 2009) were grown in LB at 28°C and 37°C, and specific β -galactosidase activities were determined in overnight cultures.

Results

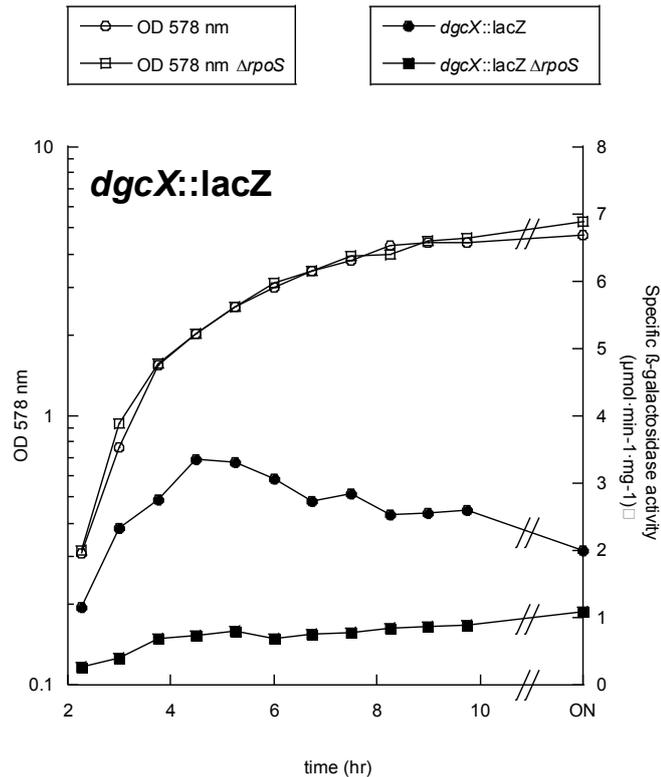


Figure 4.35: DgcX strongly expressed and under partial control of RpoS in K-12. Derivative of strain W3110 (circles) and an otherwise isogenic *rpoS* mutant (squares) carrying a single-copy *dgcX::lacZ* fusion (relevant sequence obtained from the O104:H4 outbreak strain) were grown in LB at 37°C. Optical densities (open symbols) and specific β -galactosidase activities (closed symbols) were determined.

4.2.2 Expression of *yneF*

The *yneF* gene was poorly expressed at both 28°C and 37°C but had stronger expression in the stationary phase (Fig. 4.37). This result was further confirmed by the primer extension experiment displayed in figure 4.33 which revealed that *yneF* had a classical σ^S (RpoS) promoter with an extended -10 region consensus sequence (Fig. 4.32). YneF appeared to be under RpoS regulation, the expression level with a $\Delta rpoS$ background dropped almost to zero (Fig 4.36; Fig. 4.37) at both 28°C and 37°C. YneF also appeared to be under partial control of CsgD, with expression levels of the *yneF::lacZ* fusion in a $\Delta csgD$ background dropping to half of that found in the wild-type background (Fig. 4.37).

Results

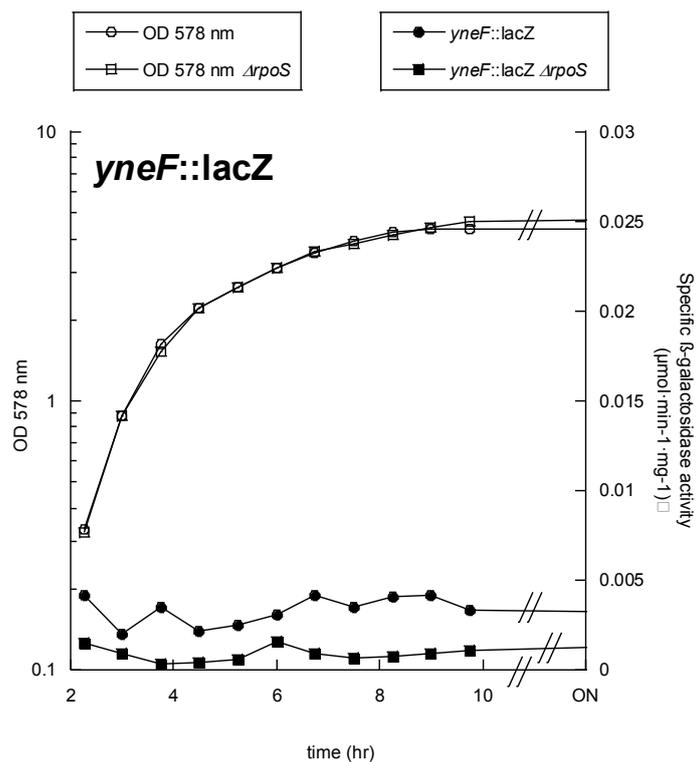


Figure 4.36: YneF weakly expressed and under RpoS control in K-12. Derivative of strain W3110 (circles) and an otherwise isogenic *rpoS* mutant (squares) carrying a single-copy *yneF::lacZ* fusion (relevant sequence obtained from the O104:H4 outbreak strain) were grown in LB at 37°C. Optical densities (open symbols) and specific β -galactosidase activities (closed symbols) were determined.

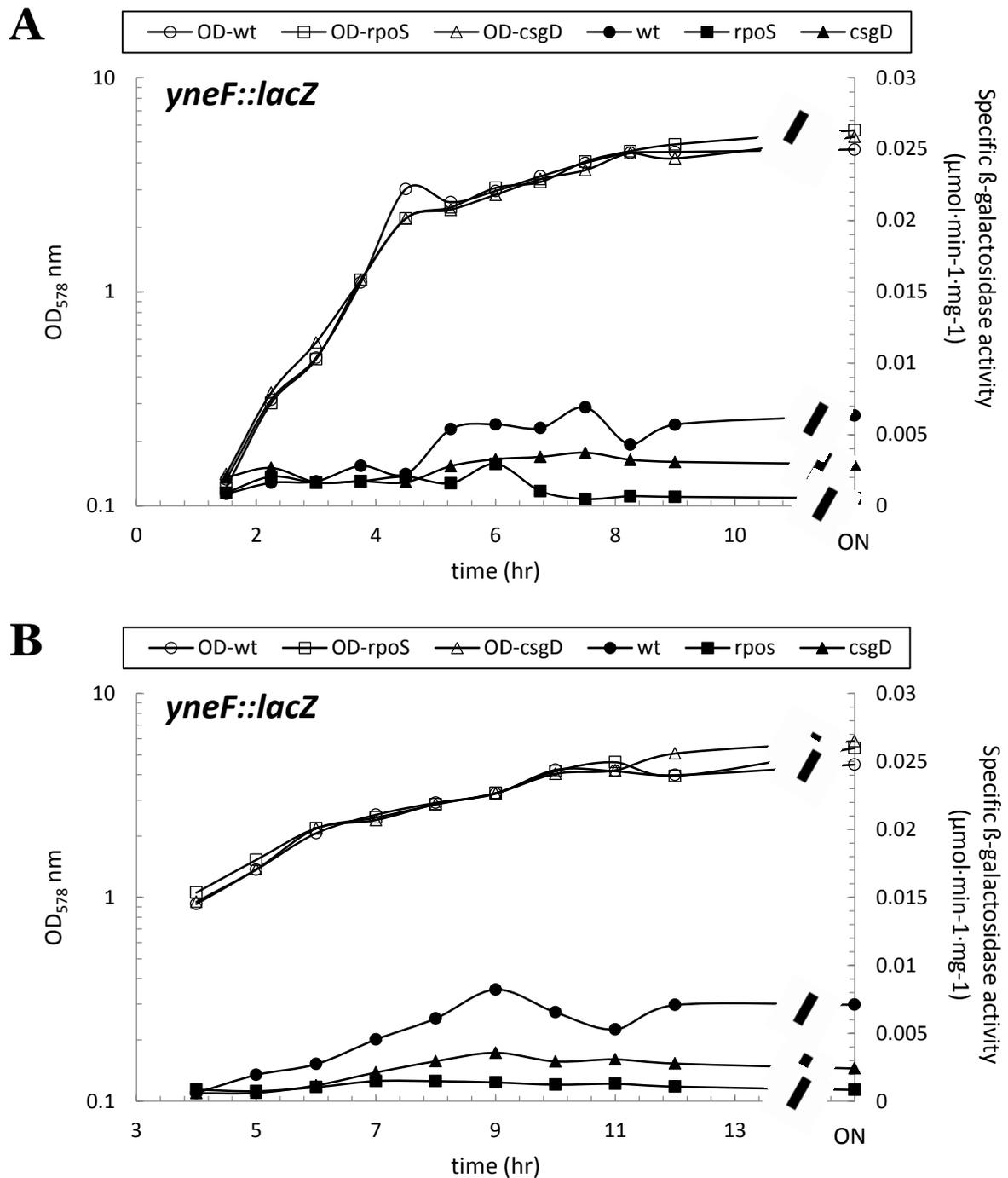


Figure 4.37: *YneF* weakly expressed but under *RpoS* regulation and partial *CsgD* control in K-12. Derivative of strain W3110 (circles), an otherwise isogenic *rpoS* mutant (squares) and an otherwise isogenic *csgD* mutant (triangles) carrying a single-copy *yneF::lacZ* fusion (relevant sequence obtained from the O104:H4 outbreak strain), were grown in LB at A) 37°C and B) 28°C. Optical densities (open symbols) and specific β -galactosidase activities (closed symbols) were determined.

4.2.3 Expression of *pdeT*

The *PdeT* protein was found in all EHECs of the O157:H7 serotype, as well as in EPECs of the O55:H7 serotype. The *pdeT* gene had practically no expression under various conditions at both 28°C and 37°C, and in the absence or presence of salt. With such poor expression it is difficult to

Results

form any conclusions. At the start of this study it was unclear whether *ycdT* and *pdeT* were in the same operon as suggested by their proximal location in the *E. coli* genome. In the O55:H7 EPECs was also a large 5' deletion in the *ycdT* gene. To see whether *ycdT* and *pdeT* were in the same operon and whether *pdeT* was at all expressed in the EPEC strains, two *lacZ* fusions were constructed. The first construct treated *pdeT* as its own transcriptional unit, with 500 nucleotides upstream of the start codon fused to the *lacZ* gene. The second construct treated *pdeT* as part of the same operon as *ycdT* and the promoter region fused with the *lacZ* gene included 500 nucleotides upstream of the start codon of the *ycdT* gene, including the entire *ycdT* gene. Both of the constructs were tested and the poor expression of the *pdeT* gene as part of the same operon *ycdT*, labeled as *ycdTpdeT::lacZ*, at both 28°C and 37°C is displayed in figure 4.38. This construct was also tested for CsgD and RpoS dependence, but with such poor expression levels no definite conclusions could be drawn. The construct that treated *pdeT* as a single transcriptional unit (*pdeT::lacZ*) had practically no expression as well (Fig. 4.39) at 37°C but seemed to have a little expression at 28°C. It was not clear whether the expression was genuine or an artifact of the experimental procedure. Branchu and colleagues (2013) showed that in the EHEC EDL933 strain the *pdeT* (there named *vmpA*) gene was co-transcribed with *ycdT*. In the K-12 strains, YcdT is only expressed in *csrA* mutants, as CsrA directly binds *ycdT* mRNA and inhibits protein translation (Jonas *et al.*, 2009). It is possible that PdeT also requires a *csrA* mutation in order to be expressed in the K-12 strain. If the *pdeT* gene has another minor promoter, as suggested by my tenuous results then it could still be expressed in the O55:H7 EPECs and contribute to their regulatory network of c-di-GMP regulation.

Results

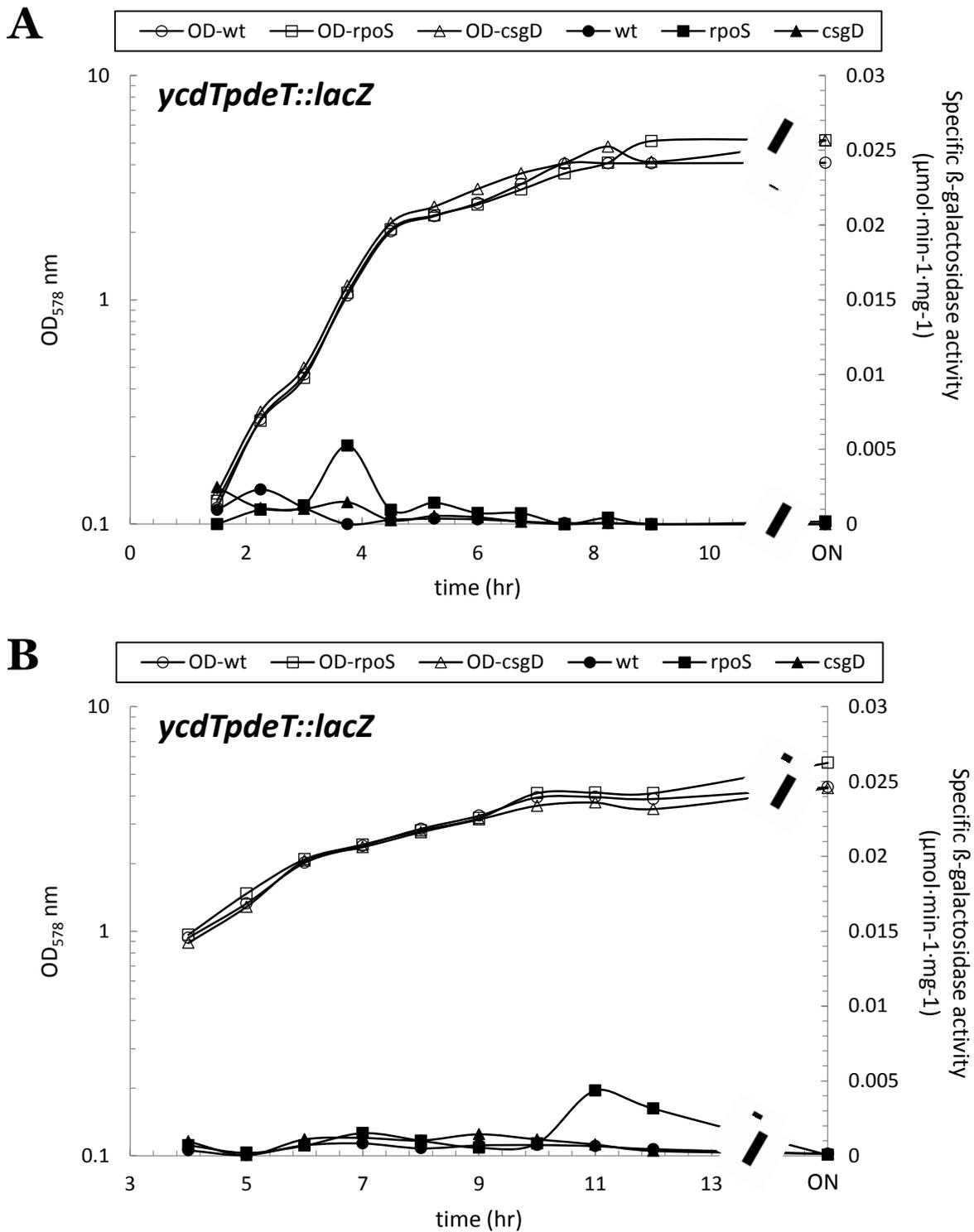


Figure 4.38: PdeT in joint operon with YcdT shows almost no expression in K-12. Derivative of strain W3110 (circles), an otherwise isogenic *rpoS* mutant (squares) and an otherwise isogenic *csgD* mutant (triangles) carrying a single-copy *ycdTpdeT::lacZ* fusion (relevant sequence obtained from the EHEC EDL933 strain), were grown in LB at A) 37°C and B) 28°C. Optical densities (open symbols) and specific β-galactosidase activities (closed symbols) were determined.

Results

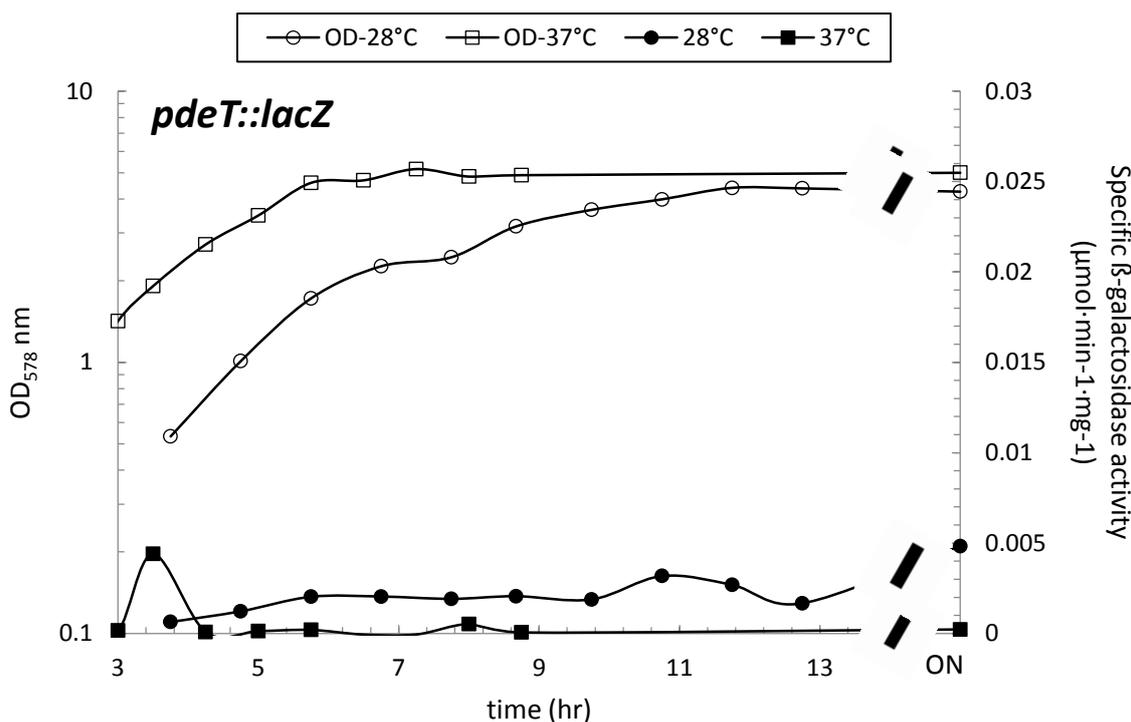


Figure 4.39: PdeT as its own transcriptional unit shows weak expression only at 28°C in K-12. Derivatives of strain W3110 carrying a single-copy *pdeT::lacZ* fusion (relevant sequence obtained from the EHEC EDL933 strain), were grown in LB at 37°C (squares) and 28°C (circles). Optical densities (open symbols) and specific β -galactosidase activities (closed symbols) were determined.

4.3 Functional characterization of newly identified DGCs and PDEs

4.3.1 Protein-protein interaction: PdeT (VmpA) does not interact with YcdT

Branchu *et al.*, (2013) showed that the *pdeT* (PDE) gene was encoded in the same operon as *ycdT* (DGC). Due to this and because there have been other examples of PDEs and DGCs having direct interaction, I tested whether this was also true for PdeT and YcdT. The BacterioMatch II two-hybrid system was used to detect *in vivo* protein-protein interactions. The system detects protein-protein interactions via the transcriptional activation of the HIS3 reporter gene (Materials & Methods 3.5.18). The bait protein (expressed from the pBT vector) is fused to the full-length bacteriophage λ repressor. The target protein (expressed from the pTRG vector) is fused to the N-terminal domain of the A-subunit of RNA polymerase. If an interaction between the two proteins occurs then the HIS3 reporter gene is transcribed and will complement the histidine auxotrophy of the reporter strain. Thus the reporter strain, where the bait and target proteins interact, will be able to grow in the presence of a competitive inhibitor of the HIS3 enzyme [2-amino-1,2,4-triazole (3-AT)].

Primers for the construction of pBT-*ycdT*, pTRG-*pdeT*, pBT-*pdeT* and pTRG-*ycdT* are listed in table 3.7. All four possible configurations of each protein acting as bait and target were tested. Each protein was also tested as bait or target with the counterpart empty vector in order to detect

Results

potential false positives. No growth on selective media was observed for any of the configurations (Fig. 4.40). Plasmids pBT-*rpoS* and pTRG-*rssB* obtained from Eberhard Klauck were used as positive controls since RpoS and RssB have a strong interaction (Klauck & Hengge, unpublished data). Plasmids pBT-*rpoS*_K173E and pTRG-*rssB* were also used as a negative control since the mutation in RpoS of K173E interferes with RssB binding. Only the positive control grew on the selective media, while all of the tested configurations grew on the non-selective media. The pTRG-*ycdT* had trouble growing (growth was spotty and not as uniform as other interactions) (Fig. 4.40) on the non-selective media plates and when cloning the plasmid (data not shown). The pTRG plasmid has a higher copy number (20-30) than the pBT plasmid (5-10) (BacterioMatch II two-hybrid system manual). The observation that the reporter strain had trouble growing with YcdT encoded on pTRG rather than pBT may be a hint that overexpression of YcdT has a toxic or growth inhibiting effect on the cell. PdeT (VmpA) may still be antagonizing YcdT, but if so then it is doing it indirectly. Due to these results and the 2013 publication of Branchu and colleagues, all subsequent research of PdeT (VmpA) was discontinued.

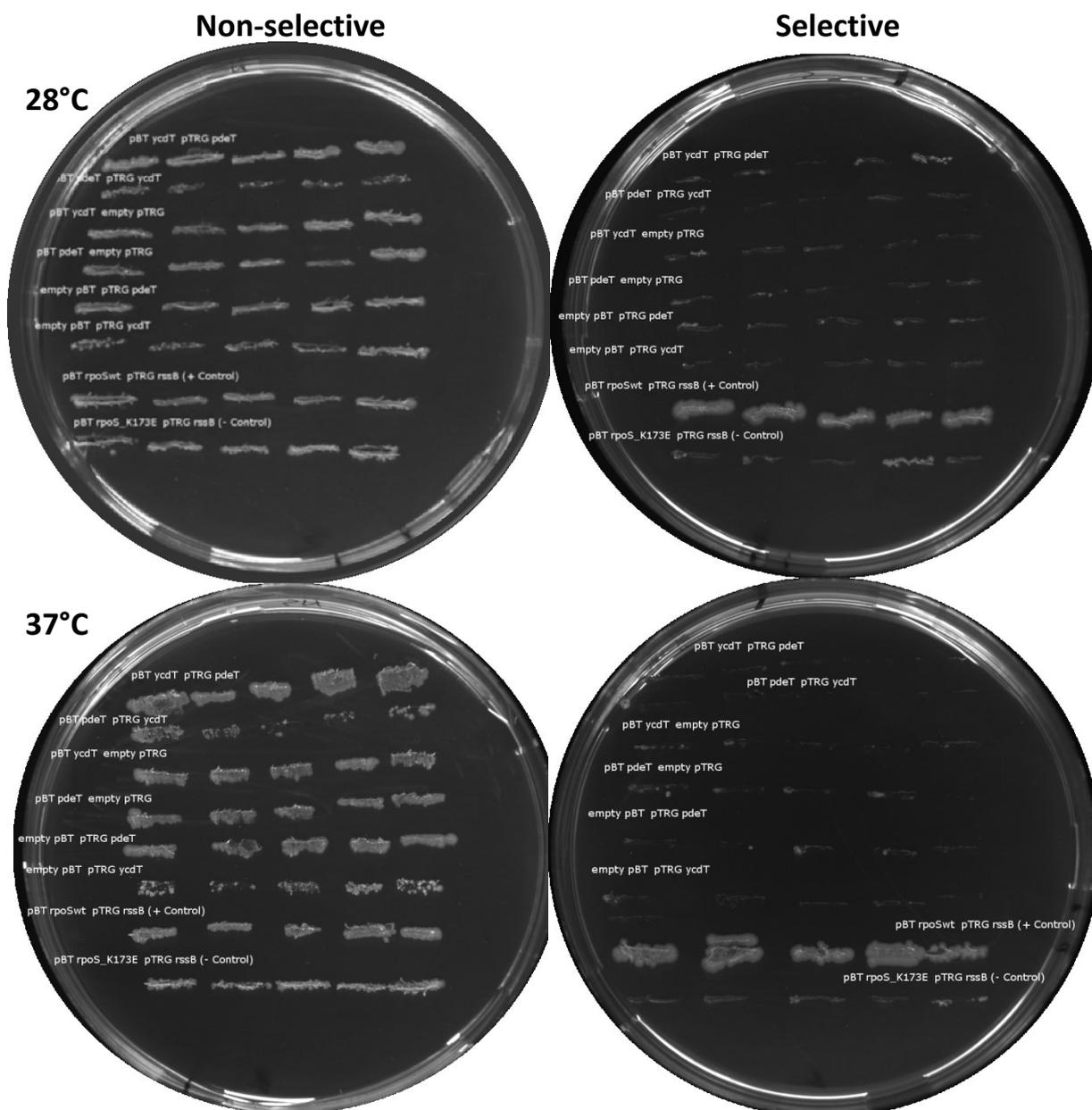


Figure 4.40: PdeT (VmpA) does not directly interact with YcdT. In the top row plates were grown at 28°C and in the bottom row plates were grown at 37°C. The first column contains non-selective media plates and the second column contains selective media plates. Only the positive control pBT-*rpoS* and pTRG-*rssB* showed interaction (grew on selective media). All other configurations tested, including the negative control pBT-*rpoS*_K173E (mutation prevents binding with RssB) and pTRG-*rssB*, grew on the non-selective media.

4.3.2 Diguanilate cyclase activity of DgcX and YneF

Both DgcX and YneF are predicted DGCs (based on the conservation of specific residues required for DGC activity) that contain 8 transmembrane spanning segments (TMSs). In an attempt to confirm that both DgcX and YneF are functional DGCs, I cloned the GGDEF domain encoding 3'-end of the respective genes onto a HIS-tagged high copy plasmid, pQE30. Only the C-terminal portions (GGDEF domain-containing regions) of both DgcX and YneF were cloned in

Results

order to ensure overexpression and purification (Materials & Methods 3.5.16.1), using the primers listed in table 3.7. A DGC assay was performed with the two proteins to test whether the proteins could synthesize c-di-GMP. Radioactive labeled ^{33}P -GTP was combined with purified DgcX, YneF and the highly active soluble DGC PleD (positive control) to see whether the proteins could synthesize ^{33}P -c-di-GMP. DGC activity could not be detected for either DgcX or YneF, while PleD (positive control) was able to synthesize c-di-GMP (Fig. 4.41). As both of the proteins have intact A-sites in their GGDEF domains (see Table 4.1), it is more likely that they are active DGCs but that the proteins were not purified in their active forms or that they need their sensory domains for activation. The proteins were overexpressed and able to be purified at a high enough concentration to perform the assay (data not shown). This suggests that the sensory domains are needed for stability, possibly proper domain folding or prevention of degradation and perhaps even activation of the DGC activity by dimerization.

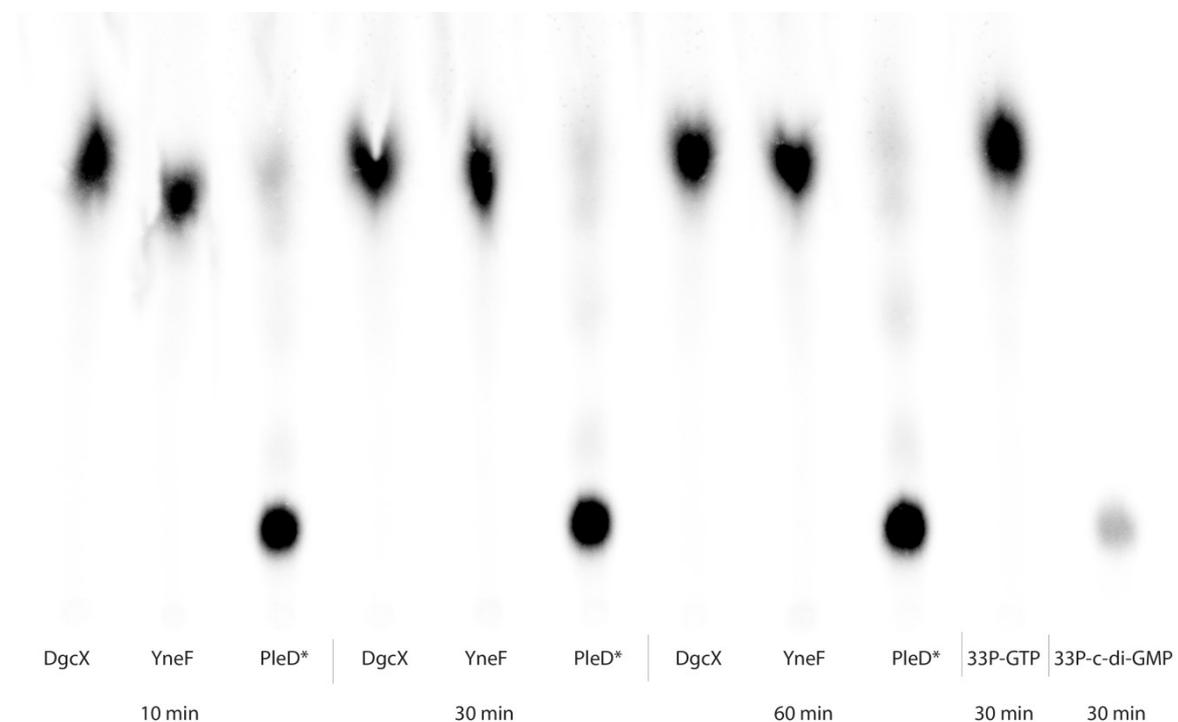


Figure 4.41: DGC assay shows no c-di-GMP synthesis by DgcX or YneF. DGC activity was measured in 10, 30 and 60 minute intervals. The highly active DGC PleD* was used as a positive control and shows ^{33}P -c-di-GMP synthesis from ^{33}P -GTP for each of the time intervals. (Experiment was performed by Alexandra Poßling and is shown here with permission).

4.3.3 Can DgcX complement a *yegE* knockout?

As previously demonstrated (see Results 4.2.1), *dgcX* has a very strong promoter. This led to the idea that it may be acting as a DGC that regulates a global c-di-GMP pool within the cell. As shown in section 4.3.2, direct testing of whether DgcX is an active DGC by performing a DGC assay failed, so a different approach was attempted. In order for CsgB, the minor component of curli fibers fibers (see Introduction 1.3.7), to be expressed, the cell needs c-di-GMP to be

Results

produced. Commonly YegE is one of the main DGCs that synthesize this global c-di-GMP pool. By knocking out YegE and compensating with DgcX, I wanted to test whether DgcX could also act as a global c-di-GMP regulator. The experiment was set up to measure the β -galactosidase activity of *csgB::lacZ*.

The *dgcX* gene was cloned onto an IPTG controlled plasmid (pCAB18), using the primers listed in table 3.7. The plasmid was then transformed into the K-12 GB1000 *csgB::lacZ* Δ *yegE* and GB1000 *csgB::lacZ* strains. As a positive control the pCAB18-*yaiC* was created, as YaiC is a DGC that is able to complement a *yegE* knockout. As a negative control the empty vector was also transformed into the two strains and all of the transformed strains were grown with and without IPTG from which samples for β -galactosidase activity testing were collected (see Material & Methods 3.5.9).

The results of the first β -galactosidase assay, where the tested strains were grown for 24 hours at 28°C and samples were taken along the growth curve, indicated that *dgcX* was not able to complement Δ *yegE* (Fig. 4.42). The YaiC protein (positive control) showed complementation of the *yegE* knockout (Fig. 4.42B), the ON β -galactosidase activity of *csgB::lacZ* in a Δ *yegE* background carrying pCAB18-*yaiC* was 0.799 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$ and the activity of *csgB::lacZ* in a wild-type background carrying the empty pCAB18 was 0.344 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$. In the *csgB::lacZ* in a wild-type background strain, YaiC considerably increased the β -galactosidase activity (2.22 $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$), because higher than encountered *in vivo* levels of c-di-GMP further boosted *csgB::lacZ* expression (Fig. 4.42A). The pCAB18 plasmid has a 'leaky' promoter, i.e. genes are still expressed even in the absence of IPTG to bind with LacI, so that LacI can no longer repress the promoter of the plasmids. This results in a small difference between the expression of proteins with and without the addition of IPTG (Fig. 4.42).

An SDS-PAGE gel was unable to clearly demonstrate whether DgcX was even expressed, as many other proteins fall into the size range of DgcX (52.9kDa) (data not shown). A C-terminally HIS-tagged *dgcX* construct (pQE60-*dgcX*) was then made along with the positive control (pQE60-*yaiC*) using the primers listed in table 3.7. Using these new constructs, the whole complementation experiment was once again repeated according to the protocols described in Materials & Methods 3.5.9. Yet again complementation could not be shown. Detection via immunoblotting of DgcX expression in this new construct failed. A total of two HIS-tagged constructs were created; in the first one I was tried to maintain the original amino acid sequence of the DgcX protein. However, I could not use the synthetic Shine-Dalgarno sequence of the pQE60 vector and so inserted the entire *dgcX* gene along with the intergenic region 54 nucleotides upstream of the native ATG start site. The plasmid was restricted at the EcoRI and BglII restriction sites so that the synthetic Shine-Dalgarno sequence and synthetic start site were thrown out of it. The 54 bps of the intergenic region cloned into the pQE60-*dgcX* construct included the

Results

native Shine-Dalgarno sequence of *dgcX* and the native ATG start codon. The reasoning behind this decision was since the *dgcX::lacZ* translational fusion was very strongly expressed, this construct should also be strongly expressed. This did not yield a detectable expression; I therefore generated a second pQE60 construct, using the plasmids' synthetic Shine-Dalgarno sequence and synthetic start codon with the primers listed in table 3.7. This changed the second amino acid of the DgcX protein from an isoleucine to a valine. The plasmid was restricted at the NcoI and BglII restriction sites. Once again, the full complementation experiment was repeated and once again complementation was not detected (data not shown). This new and final C-terminally HIS-tagged *dgcX* construct did not result in detectable expression of DgcX as determined by immunoblotting (data not shown).

This outcome may be the result of the instability of the protein on its own and the possible degradation of DgcX in the absence of a ligand that it is sensing in its N-terminal sensory domain.

Results

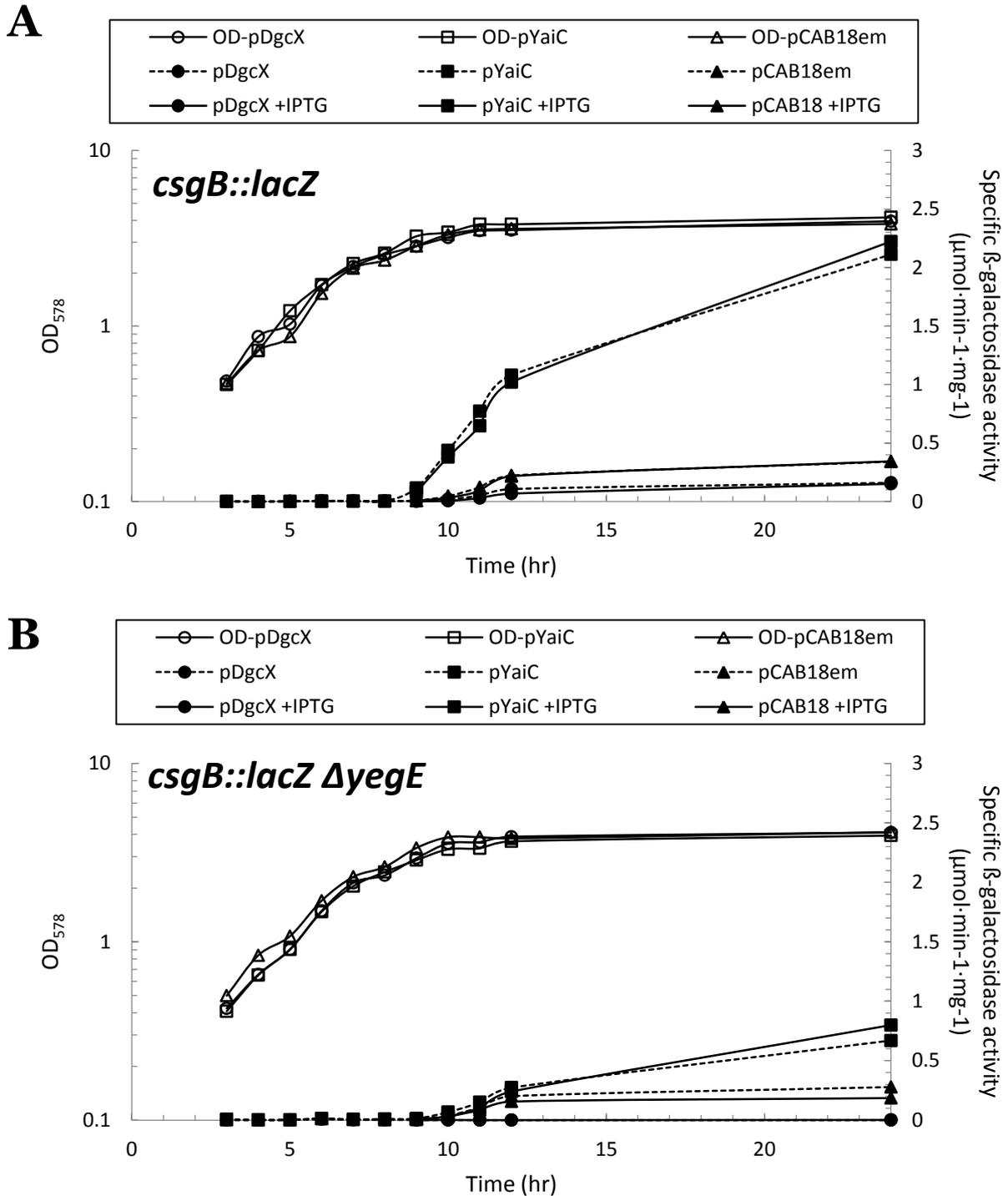


Figure 4.42: DgcX unable to complement $\Delta yegE$. DgcX (pDgcX, circles & dashed lines), YaiC (pYaiC, squares & dashed lines, positive control) and empty vector (pCAB18em, triangles & dashed lines, negative control) transformed into A) a derivative of strain W3110 carrying a single-copy *csgB::lacZ* fusion and B) a derivative of strain W3110 with $\Delta yegE::kan$ carrying a single-copy *csgB::lacZ* fusion were grown in LB at 28°C. The plasmids were induced with IPTG (solid lines) 9 hours after inoculation when the bacterial cultures were split. OD_{578} (open symbols) and specific β -galactosidase activities (closed symbols) were determined.

4.3.4 Morphology of macrocolony biofilms in K-12 strains with *dgcX* integrated into the chromosome

E. coli can produce biofilms when grown on agar plates for prolonged periods of time. The biofilms can contain complex morphological structures and in *E. coli* these structures are present due to the expression of biofilm matrix components like the exopolysaccharide cellulose and curli fibers. The addition of Congo red (CR) is used to visualize the presence and location of these matrix components as both substances readily bind CR. The presence or absence of each of these matrix components have distinct morphological phenotypes associated with them. Production of only curli fibers with no cellulose results in a phenotype of concentric rings of wrinkles in the colony originating from the colony center and symmetrically radiating outwards (Serra *et al.*, 2013b). The concentric rings arise because the non-elastic curli fibers break under the pressure of gravity and bacterial cells are trying to expand in all directions, including upwards. In the presence of both curli and cellulose, the biofilm structure takes on an interconnected network-like phenotype with smaller wrinkles, connected with large ridges that are formed by the elastic cellulose (Serra & Hengge, 2014; Serra *et al.*, 2013a; Serra *et al.*, 2013b). This morphotype has been dubbed ‘rdar’ for red, dry and rough (on CR plates) (Römling *et al.*, 1998b). The absence of both curli fibers and cellulose in biofilm formation results in a shiny, smooth and white morphotype (Hammar *et al.*, 1995). Typically, biosynthesis of both matrix components in *E. coli* occur at temperatures below 30°C and in low osmolarity conditions during growth in stationary phase (Olsen *et al.*, 1989, 1993; Maurer *et al.*, 1998). The expression of both curli fibers and cellulose depends on a regulatory cascade that involves the stationary phase RNA polymerase subunit, RpoS, the transcription factors MlrA and CsgD and a number of DGCs (Pesavento *et al.*, 2008; Lindenberg *et al.*, 2013).

Since attempts to complement a $\Delta yegE$ with *DgcX* encoded on a plasmid proved futile, I tested whether *DgcX* could complement a *yegE* knockout via different means. I was able to integrate *dgcX* into the chromosomal DNA of the *E. coli* K-12 strain GB1000 in an arrangement that corresponds to the natural genetic location of *dgcX* in EAEC, at the lambda *attB* site between *ybhB* and *ybhC* following the construction of the chromosomal *lacZ* fusion protocol (section 3.5.6) with minor adjustments. The λ RS45 bacteriophage lysat was used to transduce *dgcX*(-412, +1318)::*lacZ* construct, the entire *dgcX* gene, including the stop codon and 412 nucleotides upstream of the ATG start site, into the GB1000 chromosome. White colonies were selected for and multiple lysogeny tests were performed using PCR. The chromosomal integration was confirmed by re-sequencing PCR extracted DNA fragments containing *dgcX*. Upon verification of the correct *dgcX* sequence, I P1 transduced a deletion in *yegE* (W3110 *yegE*::*kan*) in this newly created strain and then proceeded to look at the macrocolony biofilm morphology of these

Results

generated strains with respect to the control strains that did not have the *dgcX* gene insertion. In addition, swim plate assays were performed to test what effect *dgcX* may have on motility.

A total of 8 strains were created and tested. These were GB1000 (cellulose negative strain) with a *dgcX* gene insertion (with and without the $\Delta yegE$), GB1000 wild-type (with and without the $\Delta yegE$), TLP18 (cellulose positive strain with a *dgcX* gene insertion, with and without the $\Delta yegE$) and TLP30 (cellulose positive strain without a *dgcX* gene insertion, with and without the $\Delta yegE$). The cellulose positive strain was created by a P1 transduction of a *bcsQ*^{TTG} mutation (cellulose producing strain, described in table 3.6, labeled as cell⁺, originally created by Anja Richter).

There was no detectable difference between the strains with or without *dgcX* in either the cellulose⁺ or cellulose⁻ backgrounds and complementation of the $\Delta yegE$ was not observed (Fig. 4.43). At 37°C, colony structuring was suppressed as both curli fibers and cellulose are produced at a temperature below 30°C (Olsen *et al.*, 1989, 1993). The colonies were mostly white, indicating no binding of CR by curli fiber or cellulose. At 28°C, the cellulose and curli producing strain (wt-cell⁺) was forming the expected network-like branching pattern, while the curli-only producing strain (wt-cell⁻) was forming the expected concentric ring pattern and the colonies were actively binding CR as visualized by the uniform red coloring throughout the colonies. Colonies with a $\Delta yegE$ were smaller in size and had less structuring as compared to their wild-type counterparts at 28°C. This result is expected as YegE is a major globally regulating diguanylate cyclase, that in the late post-exponential phase synthesizes c-di-GMP to high enough cellular levels to enable the transition into the stationary phase, the eventual production of biofilm matrix components like curli fibers and the exopolysaccharide cellulose (see Introduction 1.3.6). In each of the tested conditions, the presence of DgcX did not seem to impact the colony morphology and no complementation was seen by *dgcX* of the *yegE* knockout (Fig. 4.43).

Results

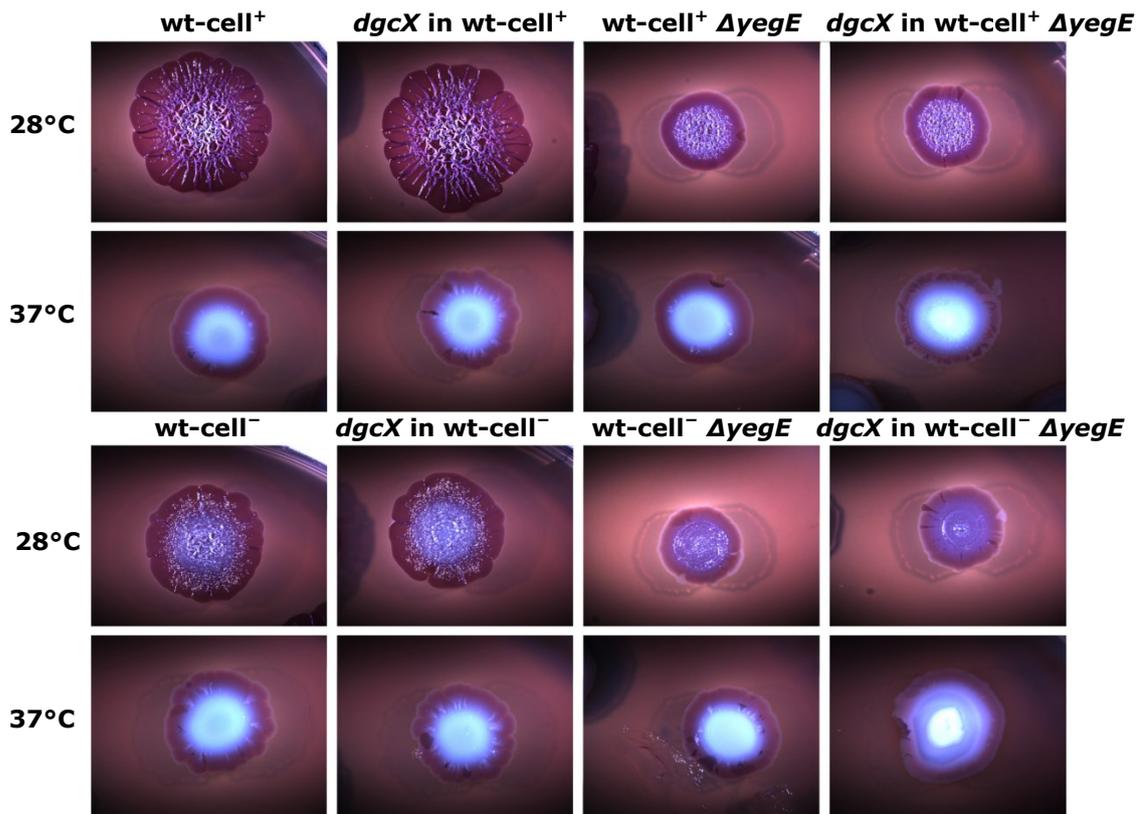


Figure 4.43: Macrocolonies on CR plates at 28°C & 37°C after 6 days. At 28°C *wt-cell*⁺ (described in table 3.6) produce network-like structures indicating curli and cellulose synthesis, while *wt-cell*⁻ (GB1000) produce concentric ring-like structures indicating curli only expression. At 37°C there is a general absence of structure which is due to very low CsgD expression and translucent pink color on CR. The $\Delta yegE$ deletion results in low levels of cellular c-di-GMP and suppresses curli and cellulose production. The presence of *dgcX* gene in each of the eight conditions does not impact colony morphology and the $\Delta yegE$ is not complemented.

As *dgcX* is found in intestinal pathogens, it was hypothesized that mucin or a mucin derivative could be the activating molecule of *dgcX*, which potentially interacts with its sensory domain. The same 8 strains described in the paragraph above were then grown for 6 days at 28°C on CR LB no salt plates containing MUC2 from porcine stomach (Roth). MUC2 is among the most common types of mucin found in the body (Malin *et al.*, 2010). MUC2 type mucin did not seem to affect the macrocolony biofilm morphology of the *dgcX*-containing strains in comparison to the control *dgcX*-lacking strains (data not shown).

4.4 Biofilm producing properties of O104:H4 outbreak strain (LB226692)

The 2011 outbreak of the Shiga toxin (Stx)-producing *Escherichia coli* O104:H4 in Germany had a high incidence of HUS (more than 20% of patients). The outbreak strain is genetically most similar to an EAEC but has acquired a Stx carrying phage from EHEC (Mellmann *et al.*, 2011). Al Safadi *et al.*, (2012) demonstrated that the May 2011 German outbreak strain O104:H4 (LB226692) has excellent biofilm forming potential, but c-di-GMP signaling has been largely

Results

neglected in EAECs which prompted the idea to test the biofilm forming properties of six *E. coli* strains. These were RKI II-2027 (EAEC), LB226692 (EAEC), HUSEC041 (EAEC), 55989 (EAEC), EDL933 (EHEC) and the commensal K-12 W3110. The RKI II-2027 and LB226692 are clonal isolates of the outbreak strain; the RKI II-2027 is the official German isolate and thus was included in this analysis, but this isolate has never been sequenced, so the sequenced LB226692 strain was also included in this study. The classical EAEC 55989 was chosen to be included in the laboratory experiments because the outbreak strain shares 98% of its genome with it. HUSEC041 (01-09591) was included because it is a historic isolate from 2001, closely related to the outbreak strain LB226692. The classical EHEC EDL933 strain was included because both the outbreak strain (LB226692) and HUSEC041 harbor a Stx2 containing prophage inserted into the *wrbA* gene, the same integration site where the phage is found in EDL933. The EAEC 55989 strain has an intact *wrbA* gene not occupied by a prophage. The K-12 laboratory strain W3110 was included as reference strain that is a commensal and not a pathogen. All of the selected pathogenic strains had an extra DGC, YneF, compared to the K-12 W3110 strain and all three of the tested EAECs had another additional DGC, DgcX. All of the wetlab experiments in this section were performed together with Anja Richter.

4.4.1 Morphology of macrocolony biofilms of O104:H4 outbreak strain (LB226692)

At 28°C the 55989 and HUSEC041 strains produced branching web-like structures, while the LB226692 and K-12 W3110 strains produced concentric ring patterns typical of curli only utilization (Fig. 4.44A and B at 28°C). The EHEC EDL933 produced shiny, white, flat unstructured colonies, typical of curli and cellulose negative strains. As mentioned earlier, this is due to the prophage containing Stx1 being inserted in the *mlrA* gene, thereby disrupting it and thus abolishing production of the two biofilm matrix components. CR staining produced dark red coloring (typical for high curli production) of the colonies not affecting the web-like structures of the 55989 and HUSEC041 and the concentric ring structure of LB226692 and K-12 W3110. The EHEC EDL933 was unstained white, characteristic of curli and cellulose minus strains. In all of the strains a thick non-CR staining white layer was also observed to varying degrees. This could be representative of unknown additional biofilm matrix components (Fig. 4.44B at 28°C).

At 37°C the K-12 W3110 strain lost structuring and produced unstained flat colonies. The EHEC EDL933 remained unstructured as well. The EAECs continued to produce structures and be stained red, but in a differential manner. The 55989 strain produced a regular pattern of staining but the outbreak LB226692 and HUSEC041 stained red only in certain sectors with the rest of the colony remaining white. The HUSEC041 strain produced significantly fewer of these stained sectors as compared to the outbreak strain (Fig. 4.44B at 37°C).

Results

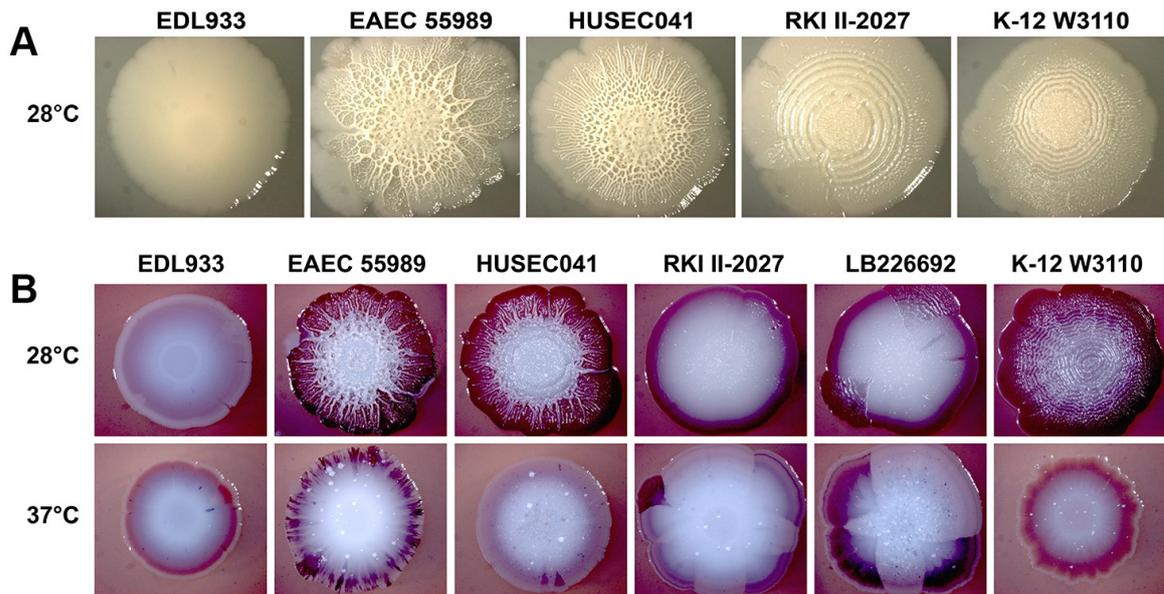


Figure 4.44: Macrocolonies on A) LB no salt at 28°C and B) CR plates at 28°C & 37°C after 7 days. 55989 & HUSEC041 produce network-like structures indicating curli and cellulose. EDL933's absence of structure is most likely due to prophage insertion in *mlrA*, resulting in very low CsgD expression and translucent pink color on CR. Both the LB226692 & RKI II-2027 lack cellulose expression due to single nucleotide insertion in the *bcsE* gene which results in a frameshift and early termination and shows ring formation typical for curli only strains (e.g., *E. coli* K-12 W3110). (Experiment was performed together with Anja Richter; the figure was created by Anja Richter and is shown here with permission).

4.4.2 Expression of biofilm regulator CsgD in O104:H4 outbreak strain (LB226692)

CsgD expression is integral for the production of both curli fibers and cellulose. The 7 day old macrocolonies were tested for CsgD expression by scraping the entire colony and immunoblotting. The amount of CsgD for each strain was calculated (by Anja Richter) by applying known amounts of CsgD (first four lanes in Fig. 4.45) and by measuring the band intensities and correlating them to the known values using the ImageGauge V3.45 software (data not shown). At 28°C no salt condition, all of the tested strains except for EDL933 (EHEC) had high levels of CsgD. At 37°C, only the EAECs 55989 and LB226692 produced high levels of CsgD even in the presence of salt (Fig. 4.45). It is likely that in these strains the mechanism that controls CsgD expression has lost its sensitivity to temperature and salt concentration. Examination of the CsgD promoter did not show any mutations, thus the regulatory mutations responsible for this switch remain to be elucidated.

Results

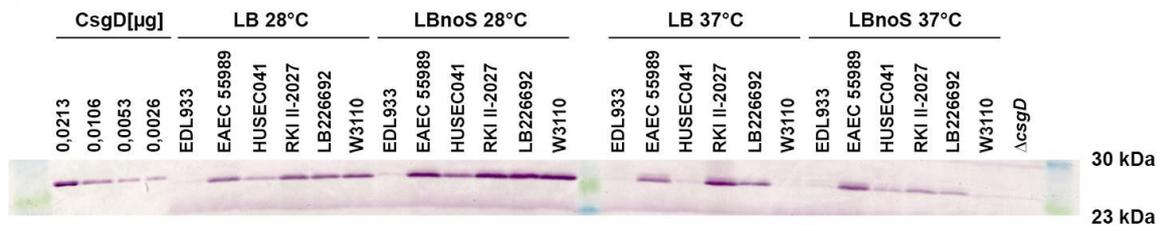


Figure 4.45: Cellular CsgD levels. Immunoblotting of colony grown samples revealed that all of the strains except for EDL933 produce CsgD at 28°C, but 55989 and the outbreak strains (LB226692 & RKI II-2027) also produce CsgD at 37°C. (Experiment was performed together with Anja Richter, the figure was created by Anja Richter and is shown here with permission).

5 Discussion

5.1 Acquisition of a novel DGC, DgcX, and changes in other DGC/PDE genes in EAECs including outbreak strain O104:H4 LB226692

The acquisition of a novel DGC, DgcX and the extended version of the *yneF* gene in the EAEC strains of the O104:H4 serotype increases the overall number of DGCs carried by these strains to 13 as compared to the 11 carried by the K-12 strains W3110 and MG1655. All of these EAEC strains have the anomaly of possessing the DGC. YedQ encoded in two parts where a point mutation in the 252nd codon has generated a stop codon but that is followed immediately by a start codon within the same reading frame which is highly likely to trigger a translational restart via translational coupling, that is if read-through does not occur automatically. Only the 55989 strain has an additional corrupted DGC, YcdT, which is distorted by a 7 nucleotide insertion that results in a frameshift and early termination of the protein without the GGDEF domain, thereby bringing its total DGC count to 12. Tagliabue and colleagues have shown that the YddV protein regulates the *pgaABCD* operon (Tagliabue *et al.*, 2010) and the 55989 strain has the *yddV* gene intact. Recently, it has been shown that the EAEC 55989 strain has a high level of expression of *pga* genes (Al Safadi *et al.*, 2012). This may be the result of an insertion element, IS1 (Fig. 4.1) between *pgaA* and *ycdT*. The *pgaA* promoter in the 55989 strain compared to the characterization in the *E. coli* K-12 (Wang *et al.*, 2005) is augmented to contain not only a -35 region that may combine with the still present -10 region but also an additional full promoter further upstream. This double promoter of IS1 has been shown to increase expression of other genes (Prentki *et al.*, 1986; Barker *et al.*, 2004) and may be responsible for the enhanced production of PGA in EAEC 55989.

The total number of PDEs within the strains of the O104:H4 serotype remain unaltered as compared to the K-12 W3110 strain, all have a total of 13 PDEs, but the increase of DGCs suggests an increased potential for biofilm production. Macrocolony experiments presented above (Results 4.4.1), showed that all of the tested EAEC strains of the O104:H4 serotype (55989, RKI-II-2027, LB226692 and HUSEC041) formed excellent biofilms.

The strong expression of DgcX compared to the other 12 DGCs, as observed by regulation experiments of single copy *lacZ* reporter fusion inserted into the W3110 chromosome, where DgcX showed the highest expression at both 28°C and 37°C, further purports these EAECs biofilm forming potential. Moreover, the *dgcX* promoter seems to be under both vegetative and stationary phase RNA polymerase regulation. The fusion was strongly expressed in the growing phase but showed to also be partially RpoS-dependent. Primer extension experiments showed a classic vegetative promoter sequence. E σ^S can also optimally recognize and bind -10 and -35

Discussion

hexamers *in vitro* (Typas *et al.*, 2007; Gaal *et al.*, 2001). Thus it looks like DgcX would be expressed in all cells of the EAECs regardless of their access to nutrients. In terms of macrocolony morphology, the work of Serra *et al.*, (2013a; 2013b) has illustrated that the cells in a macrocolony forming biofilm are not undergoing a homogeneous growth phase but rather specific differentiation experienced by each cell based on the cells' location within the macrocolony. Cells along the base, interfacing with the agar, have access to nutrients and are predominantly growing and under the regulation of the vegetative promoter σ^{70} and the cells along the surface, the interface between the air and colony, have limited nutrient supplies and are in stationary phase under the regulation of the σ^S promoter (Serra *et al.*, 2013a; 2013b). Thus DgcX may be expressed in all of the locations regardless of whether the cells are under σ^{70} or σ^S control.

Experiments showed no evidence of complementation in $\Delta yegE$ mutants or any detectable expression of the DgcX protein, DGC activity (despite the conservation of specific residues required for DGC activity) or morphological changes in macrocolonies. Together these results suggest that the protein may be unstable on its own. The discovery of a motif (Results 4.1.1.2) in the sensory domain of DgcX coupled together with the other negative results further reinforces the hypothesis that DgcX may be sensing some ligand in its N-terminal sensory domain that is required for protein stability, prevention of degradation and activity.

Because *dgcX* is found in interintestinal pathogenic *E. coli* and mucin is specifically secreted onto mucosal surfaces (Malin *et al.*, 2010), mucin was hypothesized as a possible activating ligand that DgcX could interact with, in its N-terminal sensory domain. Mucin did not change the macrocolony biofilm morphology of the *dgcX* containing strains when compared to the *dgcX* lacking control strains. EAECs like the outbreak strain produce mucinases while the K-12 W3110 strain does not, thus it is possible that DgcX is sensing some byproduct of the mucin digestion. Gluconate is a major byproduct of mucin that has been treated with mucinase, therefore testing the macrocolony biofilm morphology in the presence of gluconate should be performed to test whether this molecule can activate *dgcX*.

Finally, whenever the *dgcX* gene was found within a strain it was always located at the end of a prophage insertion (Fig. 4.6) and a total of four distinct prophages were observed that carried the *dgcX* gene. Thus it appears that *dgcX* is passed along strains of *E. coli* via horizontal gene transfer by hitchhiking on prophages.

5.2 Acquisition of a novel PDE, PdeT (VmpA) and changes in other DGC/PDE genes in EHECs of the O157:H7 serotype

The EHECs of the O157:H7 serotype have acquired an additional PDE, PdeT (VmpA). Branchu *et al.*, (2013) demonstrated that PdeT (VmpA) is an active PDE, that if knocked out increases biofilm formation and reduces swimming motility, as would be expected for a PDE and that it is encoded within the same operon as *ycdT*. Overexpression of the *pdeT (vmpA)* gene on a plasmid in a wild type background of EHEC EDL933 did not affect biofilm formation as tested with crystal violet assays. The group also uncovered frequent suppressor mutations arising in $\Delta pdeT$ (*vmpA*) backgrounds, so that even if the gene is brought back into the genome, swimming motility was not restored. Based on these observations the group put forth two possible interpretations of this phenomenon: 1) either PdeT (VmpA) is a global regulator of the *E. coli* of the O157:H7 serotype and without suppressor mutations the deletion of *pdeT (vmpA)* severely affects the growth of EDL933 or 2) high levels of c-di-GMP are toxic for *E. coli* of the O157:H7 serotype and PdeT (VmpA) is a strong regulator of those levels, so that its absence forces suppressor mutations to compensate (Branchu *et al.*, 2013).

The EHECs of the O157:H7 serotype are all missing YcgG (PDE) and BluF (degenerate EAL), as well as have corruptions in genes encoding DosP (PDE) and YegE (DGC). EDL933, Sakai and Xuzhou21 also had corruptions in the *yeaP* gene (encoding a DGC), while EC4115 and TW14359 contained corruptions in the *yciR* gene (encoding a PDE). A PDE like PdeT (VmpA) that has a phenotypic affect when it is deleted may be compensating for the loss of the PDEs YcgG and DosP (and additionally YciR in EC4115 and TW14359) in the DGC/PDE complement of the O157:H7 EHECs. In the EDL933, Sakai and Xuzhou21 strains the ratio of PDE-to-DGC is 12-to-10 and in the EC4115 and TW14359 strains the ratio of PDE-to-DGC is 11-to-11. The findings that EHECs primarily reside in the mammalian digestive tract and utilize a completely different method for colonization, namely through the production of Intiman and pedestal structures via which they adhere in the host (Kaper *et al.*, 2004), may point to them favoring a low c-di-GMP cellular environment and could be the reason behind their acquisition of a novel PDE like PdeT (VmpA) in order to not have more DGCs present.

Tagliabue *et al.*, (2010) showed that it was the YddV protein that induces the synthesis of PGA by stimulating the expression of the *pgaABCD* operon and not YcdT as suggested by Jonas *et al.*, (2008). However, CsrA seems to regulate both YcdT and YddV (Jonas *et al.*, 2008) and like in the EHECs of the O157:H7 serotype, YcdT is encoded in the same operon as PdeT (VmpA), and then it too would be regulated by CsrA. As mentioned above, all of the O157:H7 EHECs have a disrupted *dosP (yddU)* gene. It was disrupted by a 10 nucleotide deletion in the 405th codon that caused a frame shift and early termination after 412 AAs as compared to 799 AAs in the wild-type variant. The *yddV* gene is encoded in the same operon as *dosP*, but upstream

(Fig. 4.4), so theoretically the deletion should not have an effect on *yddV* expression. All of the aforementioned EHECs had an intact *pgaABCD* operon. The macrocolony experiments also showed that the EHEC EDL933 does not produce either curli or cellulose (Results 4.4.1). Based on the similar genomic profile of disruption in *bcs* and *mlrA* genes, the other O157:H7 EHECs are likely cellulose and curli deficient as well, instead utilizing PGA as a biofilm matrix component, probably outside of its mammalian host, as they use a different invasion and attachment strategy inside (see Introduction 1.1).

5.3 The uncoupling of the PDE YcgG's sensory domain from the EAL domain

The gene encoding YcgG was missing altogether in eight strains of *E. coli*, including all of the O157:H7 EHECs. It was disrupted by a single nucleotide deletion in the 297th codon in its EAL domain portion in the porcine ETEC UMNK88 and by a point mutation in the 27th codon that resulted in a stop codon in the N-terminus in the ExPEC UMN026. It was fully intact in 37 *E. coli* strains and in the remaining 15 strains an interesting phenomenon was discovered: these strains had a large 5' end deletion of 630 nucleotides, so that only the EAL domain-containing C-terminal was encoded in their genomes.

At first, it appeared like the protein would not be expressed, but based on the work of Spurbeck *et al.*, (2012) the shortened variant of the *ycgG* is expressed and the truncated YcgG protein is functional. The work described how a single deletion of *c1610* (shortened version of the *ycgG* gene found in the UPEC CFT073) led to increased adherence to bladder epithelial cells (240% in $\Delta c1610$ vs 100% in wild-type). Further, the group was able to complement the deletion with the overexpression of an active exogenous PDE (VC1086) cloned onto a plasmid and claimed that this confirmed that C1610 was an active PDE (Spurbeck *et al.*, 2012). Unfortunately, though these results may hint that C1610 may be an active PDE, they certainly do not prove this, as these results could also be explained by C1610 being part of a regulation cascade that ultimately leads to the expression of a PDE rather than an active PDE itself. The $\Delta c1610$ also led to increased expression of type 1 fimbria as determined by immunoblotting for FimA, the major subunit of type 1 fimbriae (Spurbeck *et al.*, 2012). UPECs actively switch between the planktonic, motile lifestyle and the adherent sedentary lifestyle during their colonization process (Introduction 1.1) of ascending the urethra, colonizing the bladder and then ascending the ureters to colonize the kidneys (Spurbeck *et al.*, 2012; Kaper *et al.*, 2006). This switch of lifestyles is mediated by cellular levels of c-di-GMP where high cellular levels lead to adherence via type 1 fimbria and biofilm formation (Spurbeck *et al.*, 2012).

In the UPEC CFT073 and 14 other *E. coli* strains, YcgG was encoded as a 283 AA variant that only contained the EAL domain. A deletion of 630 bps in the 5' end of the gene uncoupled the EAL domain-encoding portion from its transmembrane sensory domain-encoding portion

which contains the CSS-motif. Figure 5.1 shows the potential start site of the shortened *ycgG* gene (highlighted in red) and an underlined consensus of the Shine-Dalgarno sequence. Highlighted in magenta is the beginning point where the shortened *ycgG* gene begins aligning with the wild-type variant. The region upstream of the highlighted in magenta nucleotides corresponds to the intergenic region between *ycgG* and *yngC*. Finally, this trend of YcgG truncation was observed mostly in pathogens and occurred in only two commensal strains: *E. coli* SE15 and *E. coli* ED1a. Thus it may be advantageous for pathogenic *E. coli* to have a strong unregulated PDE.

ycgG



Figure 5.1: 15 of 61 *E. coli* strains feature a 5'-truncated *ycgG* gene that most likely encodes an active PDE. *ycgG* is truncated by a deletion that includes 630 nucleotides of its 5' end in 15 strains. Spurbeck *et al.*, (2012) showed that $\Delta c1610$ in UPEC CFT073 led to increased adherence to bladder epithelial cells. *c1610* is the truncated version of *ycgG*. The potential internal start site is highlighted in red and a putative Shine-Dalgarno sequence is underlined.

5.4 Acquisition of a novel PDE, PdeY and changes in other DGC/PDE genes in UPECs

PdeY was found in six strains, three of which were UPECs. All of the strains that had PdeY also had corrupted versions of *yddV* where only a fragment of the gene remained (Results 4.1.1; Fig. 4.4). This is noteworthy because the work of Tagliabue *et al.*, (2010) showed that it was the YddV protein that induces the synthesis of PGA by stimulating the expression of the *pgaABCD* operon (Tagliabue *et al.*, 2010) and PGA is expressed in the host by UPECs (Cerca *et al.*, 2007). UPECs use PGA as a component in their biofilm formation that helps promote the structural integrity and mediates cell-to-surface and cell-to-cell adhesion, thereby enabling them to colonize the host (Itoh *et al.*, 2008). This finding points to the possibility that at least in UPECs a different protein may be regulating PGA expression.

The *yddV* gene is encoded upstream of the *dosP* gene and in the same operon (Fig. 4.4), thus its disruption potentially also prevents the expression of the DosP protein. However, the work of Spurbeck *et al.*, (2012) showed that a $\Delta dosP$ (there named $\Delta c1918$) mutant had decreased biofilm formation in M9 (glycerol) media as compared to the wild-type UPEC CFT073 strain but the deletion could not be complemented by the overexpression of an active exogenous DGC (Spurbeck *et al.*, 2012). The UPEC CFT073 also had a point mutation in the 518th codon in its

ycjC gene which resulted in a stop codon and early termination and two insertions in its *dosP* gene in the 438th and 771st codons which led to frameshift and early termination. Spurbeck *et al.*, (2012) could not detect any phenotypic effects in a $\Delta yjcC$ mutant as compared to the wild-type UPEC CFT073 strain, but as mentioned above did see an effect in a $\Delta dosP$ mutant (Spurbeck *et al.*, 2012). The explanation might be that they could not observe complementation of a $\Delta dosP$ mutant by the overexpression of an active exogenous DGC because the EAL domain of DosP is not expressed in the CFT073 strain due to the insertions.

The UPEC UTI89 was missing the *ycgG* gene altogether, while the other two (CFT073 and 536) had the truncated uncoupled version of the YcgG protein discussed in the section above, where YcgG's EAL domain is uncoupled from its sensory domain. The 536 strain also uniquely had an additional novel PDE, PdeX. PdeX is 260 AAs long and consisted of only the EAL domain, while PdeY is 246 AAs long and also consisted of only the EAL domain. Thus it seems that the UPECs maintain strong acting PDE's unregulated by sensory domains. These PDEs may be maintained due to the way that UPECs infect and colonize their host, where they actively switch between the planktonic, motile lifestyle and the adherent sedentary lifestyle during their colonization process (Introduction 1.1) of ascending the urethra, colonizing the bladder and then ascending the ureters to colonize the kidneys (Spurbeck *et al.*, 2012; Kaper *et al.*, 2006).

5.5 DGC/PDE complement reduced in K-12 strains W3110, MG1655, MDS42, DH10B, DH1 and BW2952

It has been reported that the K-12 strains complement of DGC/PDE encoding genes, which contain a GGDEF and/or an EAL domain, consists of 29 proteins. This study uncovered that one of the previously counted DGCs, YneF, is actually miss-annotated in the K-12 genomes. The K-12 annotated genomes claimed that the *yneF* gene was intact and encoded a protein of 315 AAs. Comparing this region in the K-12 strains to the corresponding region in other *E. coli* strains showed that the full variant of YneF is 472 residues and that a substantial deletion (433 bps) occurred in the protein's N-terminus. This helps explain why *yneF::lacZ* fusions in the K-12 W3110 strain failed to show expression at both 37°C and 28°C (Sommerfeldt *et al.*, 2009). Of the 61 *E. coli* genomes analyzed, 14 had deletions or were missing the *yneF* gene entirely (see Table 4.2), and of these, six were the K-12 strains.

Three of the K-12 strains were either entirely missing or had corruptions in other DGC/PDE encoding genes. DH10B had its *yedT* (DGC) gene disrupted by an IS2 insertion element. K-12 MDS42 was entirely missing *ydaM*, *yedT*, *yfeA*, *ycgG* and *ycgF*, and had a 1425 nucleotide 5' end deletion in its *yeaJ* (DGC) gene. Finally, BW2952 was also completely missing the *yahA* (PDE) gene.

Being aware of the difference in DGC/PDE complement between *E. coli* strains is critical, as most c-di-GMP research centers on the use of laboratory K-12 strains such as W3110 and MG1655, yet often larger conclusions are drawn to include pathogens as well. Moreover, additional GGDEF/EAL proteins specifically in certain classes of pathogenic *E. coli* may reflect distinct necessities in the host associated niches as well as a potential involvement of c-di-GMP in specific pathogenicity.

5.6 Loss of curli through prophage insertion in the *mlrA* gene

Curli synthesis is regulated via a signaling cascade that involves several DGCs, PDEs, the stationary phase RNA polymerase σ^S , the transcription factor MlrA and the expression of both of the *csgDEFG* and *csgBAC* operons. *csgBA* genes directly code curli fiber subunits while the *csgDEFG* operon is required for the expression and translocation of the curli fiber subunits (see Section 1.3.7). CsgC on the other hand is not indispensable for curli fiber production (Hammar *et al.*, 1995). Finally, the expression of the *csgDEFG* operon is under the control of the transcriptional regulator MlrA.

The *csg* genes were strongly conserved in all 61 of the *E. coli* strains analyzed. The only exceptions were the clonal isolates of STEC strains RM12761 and RM13516, and the UPEC 536 strain. The STEC strains had a corruption in the *csgC* gene which had an insertion in the C-terminus leading to repeated stop codons after the 89th residue instead of the wild-type variant of 110 AAs. As mentioned above, the *csgC* gene has also been shown to be not essential for the curli formation (Hammar *et al.*, 1995). In the UPEC 536 an insertion in the 73rd codon led to a shift in the reading frame and early termination after the 105th residue instead of the 277 AAs in the wild-type variant.

Despite the conservation of the *csg* genes, not all of the strains are able to synthesize curli. This begs the question of why these genes are then maintained and what causes the loss of curli formation. As mentioned in section 4.1.4.1, the *mlrA* gene was corrupted in 14 strains. Since MlrA is the transcriptional regulator of the *csgDEFG* operon this can explain why there is a loss of curli formation in some of the strains. Only the *E. coli* HS strain had a corruption caused by an insertion of 8 nucleotides in the 179th codon which actually led to late termination after 262 AAs as compared to 243 AAs in the wild-type strain. In the other affected *E. coli* strains, the corruption of the *mlrA* gene occurred via a prophage insertion, predominantly the one carrying the shiga toxin genes, which happens to favor *mlrA* as an insertion site (Serra-Moreno *et al.*, 2007; Shaikh & Tarr, 2003). Thus if a strain was to lose the ability to form curli, its primary mechanism for doing this is an *mlrA* knockout rather than corruption of the *csg* genes. Uhlich *et al.*, (2013) showed this to be the strategy employed by EHECs of the O157:H7 serotype. MlrA is also important for cellulose expression via its regulation of CsgD, which regulates cellulose synthesis,

yet there no strong conservation observed among the *bcs* genes. This seems to indicate that in order to get rid of its ability to form cellulose, *E. coli* undergo disruption of the *bcs* genes directly rather than using a more global regulation factor.

E. coli is able to bypass MlrA regulation altogether by mutations of the CsgD promoter. Uhlich and colleagues (2001) were able to show that a mutation of a single nucleotide in the -10 promoter region of CsgD from 'TAGATT' to either 'TATATT' or 'TAGAAT' was enough to restore curli production in EHECs of the O157:H7 serotype, independent of temperature. This mutation was observed in EHECs that were forming subpopulations and was instable. Strains tended to keep forming subpopulations of those carrying the mutations and reverted back to the original sequence even if plated from a homogeneous population carrying the mutation (Uhlich *et al.*, 2001). Römling *et al.*, (1998b) showed that an insertion of a single nucleotide in the spacer region (between the -10 and -35 box) of the *csgD* promoter, changing it from 16 bps to 17 bps was enough to make *csgD* no longer σ^S -dependent. All of the analyzed strains with a corrupted *mlrA* gene had a non-mutated form of the -10 promoter region of CsgD and had a 16 bps spacer region of the *csgD* promoter. Thus these specific mechanisms for bypassing MlrA regulation were not observed in the strains with the disruptions in the *mlrA* gene. However, MlrA regulation of the *csg* genes could in principle be bypassed by other mutations yet to be discovered which may be the reason why the *csg* genes are so well conserved among the strains.

Furthermore, some *E. coli*, particularly EHECs of the O157:H7 serotype, may forgo curli or cellulose formation altogether as part of their colonization strategy. O157:H7 EHECs are considered commensals in ruminates and it has been shown that curli fibers are pro-inflammatory in mammals (Tükel *et al.*, 2005, 2009). EHECs strategy for colonization of the recto-anal junction (RAJ) of the colon in ruminates involves attaching and effacing (A/E) lesions on mucosal epithelium which consists of the structural rearrangement of the normal microvillar architecture to induce the formation of pedestals onto which the bacterium adheres (Kaper *et al.*, 2004; Nguyen *et al.*, 2012; Etcheverría *et al.*, 2013,). In other words, the O157:H7 *E. coli* may have evolved to prevent curli expression and rely on other adhesins for biofilm formation outside of the colon and in the environment in order to avoid inflammation and triggering the immune response.

5.7 Cellulose and PGA, biofilm matrix components least maintained among *E. coli* strains

The *bcs* genes are involved in cellulose biogenesis. The *bcsQABZC* operon contains the cellulose synthase genes (*bcsA* and *bcsB*) as well as several accessory factors (see Introduction 1.3.7). The *bcsEFG* operon is also essential for cellulose formation. Non-polar deletions in each of the three genes led to loss of cellulose formation in macrocolonies (Richter & Povolotsky, submitted). The *pgaABCD* operon is required for the synthesis and excretion of the PGA exopolysaccharide

Discussion

(Introduction 1.3.7 for mechanistic details) which is used as a biofilm matrix component by some *E. coli*, notably UPECs.

The *bcs* and *pga* genes were more frequently corrupted compared to the *csg* genes. There were 24 strains that conserved all of the *bcs* and *pga* genes free of any deleterious mutations, insertions or deletions. These strains were in principle capable of synthesizing both PGA and cellulose. In 22 other strains either the complete *pgaABCD* operon was intact or the complete *bcsEFG* and *bcsQABZC* operons were intact. These strains were in principle only capable of synthesizing either PGA or cellulose, but not both. 15 strains were corrupted in both *pga* and *bcs* genes, so that in principle these strains could neither synthesize PGA nor cellulose. Of these 15 cases only 2 strains also had disruptions in their curli synthesizing genes (STECs RM12761 and RM13516). As discussed above, these two strains had a disrupted *csgC* gene and it has been shown that *csgC* is not essential for the curli formation (Hammar *et al.*, 1995). Thus none of the 61 strains had deleterious disruptions that prevent the synthesis of biofilm matrix components in all three *csg*, *bcs* and *pga* genes. Also discussed in the section above, curli was almost always maintained. Macarisin *et al.*, (2012) demonstrated that for adhesion to surfaces, like spinach leaves, curli is much more essential than cellulose. Thus, as most *E. coli* strains have to spend at least some part of their lives surviving in the environment outside of the host, they may therefore conserve curli genes better than cellulose and PGA synthesizing genes.

In ten strains the *pgaABCD* operon and the *ycdT* gene were missing in their entirety and only in the *pga* genes an entire operon was deleted. This finding is not unexpected since the two genomic regions are neighboring each other (Fig. 1.6). There was no pattern associated with the loss of the entire *pgaABCD* operon; with the exception of UPECs all types of pathogenic and commensal *E. coli* were affected. An additional nine *E. coli* strains had single genes within the operon deactivated via deleterious insertion, deletion or point mutation events. The deactivated genes did not overlap and were either *pgaA* or *pgaB*. Only the *E. coli* ETEC H10407 strain had a point mutation in its *pgaD* gene in addition to a point mutation in its *pgaA* gene (see Results 4.1.4.3 for details). The *pgaC* gene was always maintained unless the whole genomic region had been deleted.

The *bcsQ* gene was disrupted in 17 of the 61 analyzed strains. Three types of disruptions were observed repeatedly. The deletion of 299 nucleotides in the 5' end was observed in the STEC strains of the O145:H28 serotype (RM13516, RM13514, RM12761 and RM12581). The point mutation in the 66th codon was observed in all five of the EHECs of the O157:H7 serotype (EC4115, EDL933, Sakai, TW14359 and Xuzhou21). Finally, another point mutation in the 6th codon was observed in all five of the K-12 strains (BW2952, DH10B, MG1655, W3110 and MDS42). This point mutation was discovered when aligning the W3110 strain with the EAEC 55989, and as discussed in section 4.4.1, 55989 is able to form cellulose. In the K-12 W3110

strain, this point mutation was altered to the intact sequence of the *bcsQ* gene in the 55989 strain and the W3110 strain regained its ability to form cellulose (Serra *et al.*, 2013a). In the NCBI database in all of the K-12 strains the *bcsQ* gene is annotated as intact as a 242 AAs variant starting after the ‘TAG’ mutation, thereby masking the mutation. But based on the findings with the K-12 W3110 strains, these other K-12 strains are most likely also cellulose negative.

The finding that the *bcsQ* gene was inactivated the most often, compared to the other *bcs* genes and by differing modes, suggests that this mechanism might be a ‘favored’ way for cellulose inactivation and that these inactivation events were distinct and independent from each other, occurring several times during the evolutionary process, not a single event in a progenitor strain that was subsequently passed on to all of the later descendants.

5.8 Conservation of a core eight c-di-GMP signaling housekeeping genes in *E. coli*

The systematic analysis of the 61 *E. coli* genomes uncovered the ubiquitous conservation of eight GGDEF/EAL domain-containing proteins encoded within all of the genomes analyzed in this study (Fig. 5.1). The eight genes *yaiC* (*adrA*), *yliF*, *rtn*, *yhjH*, *yhjK*, *ylaB*, *yeaI* and *csrD* (*yhdA*) may represent the housekeeping genes for c-di-GMP regulation. An additional three genes, *yeaJ*, *yfiN* and *yoaD*, were corrupted in only singular instances. The *yeaJ* gene had a large 5’ end deletion of 1425 nucleotides in the laboratory strain K-12 substr. MDS42. This strain was missing a number of other proteins (discussed in Section 5.5) and its domestication process may have caused the loss of genes that may be crucial for its survival outside of the lab environment. The *yfiN* gene was only corrupted in the EPEC RM12579 strain by the deletion of a G in the 371st codon which caused a shift in reading frame and early termination after 377 AAs, after the GGDEF domain’s A-site. It is not clear how this omission of the last 31 residues relative to the wild-type variant affects YfiN stability or function but it could be hypothesized that the protein is still active since the deletion occurred after the GGDEF domain’s A-site. The *yoaD* gene was corrupted in the ETEC E24377A, where a claimed deletion in the 118th codon of a C nucleotide in a string of three C’s resulted in a frameshift and early termination; this may be a sequencing error and not the case.

The *yliF* gene was intact in all of the 61 analyzed strains, while the *yliE* gene was corrupted in six strains. *yliF* and *yliE* are encoded within the same operon and *yliF* is located downstream of the *yliE*. In five of the six instances where *yliE* was affected, it has been disrupted by a point mutation (see Results 4.1.2 for exact codon positions). In the sixth instance the corruption was caused by an in-frame insertion of the 9 nucleotides in the EHEC O111:H- str. 11128. Thus, the predicted protein gains 3 residues and would consist of 785 AAs. As these mutations do not affect the promoter, Shine-Dalgarno or ATG start site of the operon, it may be that *yliF* is still expressed in the strains that carry disruptions in the *yliE* gene.

Discussion

YaiC is a CsgD regulated DGC that is involved in cellulose synthesis (Ogasawara *et al.*, 2011, Zogaj *et al.*, 2001). YaiC initiates cellulose production via direct synthesis of c-di-GMP. It is high levels of c-di-GMP that activate cellulose biosynthesis (Römling *et al.*, 2000; Simm *et al.*, 2004). Under specific environmental conditions in certain strains, CsgD expression and cellulose biosynthesis is uncoupled (Monteiro *et al.*, 2009). Furthermore, since YaiC regulates cellulose biosynthesis via its production of c-di-GMP, it too could potentially be uncoupled from this process. In fact, when *Salmonella* is grown in minimal medium and synthesizes cellulose, it does this by bypassing CsgD regulation via its use of the DGC STM1987 (Garcia *et al.*, 2004) and in *E. coli*, STM1987's homologue is YedQ. The ubiquitous conservation of the *yaiC* gene while at the same time the non-conservation of the *bcs* genes needed, cellulose biosynthesis brings forth the question whether *yaiC* could have an additional function. The pathotypes of the strains affected by *bcs* gene corruption include EHECs, STECs, commensal lab strains, ExPECs, NMEC and EAECs from the 2011 outbreak in Germany and the 2009 outbreak in Georgia. It could be argued that the corruption of the *bcs* genes is a recent evolutionary step among the affected strains where the *yaiC* gene had not time yet to be also affected.

YeaJ is a DGC that down regulates motility at 37°C in *E. coli* and enables the transition into the stationary phase (Pesavento *et al.*, 2008). It may be conserved because *E. coli* have adapted to also surviving in the mammalian body. For most of the *E. coli* analyzed in this study, this is the preferred habitat, therefore it is logical to conserve genes that help its survival in the host.

YhjH is a PDE that promotes motility in the early post-exponential phase by degrading c-di-GMP and keeping the cellular levels of the second messenger low. The ability to be motile is vital for *E. coli*'s survival, to perform chemotaxis, forage for food and relocate to more preferential locations. Therefore the conservation of this gene is also logical.

CsrD (YhdA) has both the EAL and GGDEF domain in a degenerate form. It indirectly regulates the flagellar master regulator, FlhDC, PGA inhibition, and CsgD expression. CsrD targets two small non-coding RNAs, CsrB and CsrC, for degradation by RNaseE (Suzuki *et al.*, 2006). When not degraded, these two small RNAs sequester CsrA which is an RNA binding protein that controls carbon metabolism, PGA synthesis and motility. The degradation of CsrB and CsrC releases CsrA, which in turn stabilizes the mRNA of the flagellar master regulator FlhDC and redirects central carbon flux. CsrA also interferes with glycogen gene expression and, through down-regulating mRNA levels of the GGDEF domain proteins YcdT and YdeH, with PGA production (Babitzke & Romeo, 2007).

The functions of YliF, Rtn, YhjK, YlaB and YeaI remain to be elucidated and their universal conservation by all of the 61 strains may propel them as prime candidates for further investigation that may yield greater understanding of c-di-GMP regulation. The functions of the other

conserved proteins influence major processes such as cellulose formation, promoting or down regulating motility or global regulation of other processes via turnover of RNAs CsrB/CsrC. Therefore their conservation seems fitting for proper c-di-GMP regulation under various environmental conditions.

5.9 O104:H4 outbreak strain's (LB226692) unique ensemble of biofilm forming properties may contribute to its high virulence

Melmann *et al.*, (2011) showed that the O104:H4 outbreak strain (LB226692) and the closely related HUSEC041 (01-09591) (both capable of causing HUS) carry genes typically coming from two different pathotypes observed in *E. coli*, namely EAEC and EHEC. Both of these strains harbor the Stx2 containing prophage in the *wrbA* gene, as well as tellurite resistance genes (*ter*) and encode the *iha* gene that is responsible for epithelial cell adherence. The EHEC EDL933 also harbors the *iha* adhesin and tellurite resistance genes (Melmann *et al.*, 2011), thus these features are not unique to the outbreak strain. The LB226692 and the HUSEC041 harbor plasmids containing genes, encoding aggregative adherence fimbriae, a hallmark of EAECs. The HUSEC041 as well as the 55989 strain has a plasmid encoding aggregative adherence fimbriae type III (AAF/III) and the heat-stable enterotoxin, AstA, while the outbreak LB226692 strain contains the plasmid encoding aggregative adherence fimbriae type I (AAF/I) without the AstA enterotoxin (Melmann *et al.*, 2011). These classical virulence factors do not fully explain why the May 2011 outbreak strain was more virulent than the other strains that it is most closely related to. The outbreak strain was demonstrated to form thick biofilms (Al Safadi *et al.*, 2012), prompting the idea to check the complement of c-di-GMP regulating genes and to test the biofilm forming potential of the outbreak strain in comparison to the other related *E. coli* strains that share genetic material.

The O104:H4 outbreak strain (LB226692) proved to have unique properties with regards to c-di-GMP signaling and biofilm formation. The strain possessed two extra DGCs as compared to the K-12 W3110 strain, DgcX and YneF. This acquisition coupled with DgcX's strong expression (highest among all DGCs tested) at both 28°C and 37°C supports the higher potential of biofilm formation. Unlike the other EAECs tested, the outbreak strain clonal-isolates (LB226692 and RKI II-2027) did not form cellulose due to a single nucleotide insertion in the *bcsE* gene which resulted in a frameshift and early termination omitting the last 78 amino residues of the protein. This single nucleotide insertion of a C in the 448th codon in a string of three Cs was confirmed by re-sequencing. It has been shown in *Salmonella* (Solano *et al.*, 2002) and in *E. coli* (Richter, Povolotsky, submitted) that *bcsE* is critical for cellulose expression. The macrocolonies of the strains also showed concentric ring patterning typical of curli-only biofilm formation. Several studies have shown that the concomitant expression of both cellulose and curli fibers led to less

Discussion

inflammation and suppressed the immune response as compared to curli production alone (Kai-Larsen *et al.*, 2010; Wang *et al.*, 2006), thus cellulose is able to mask the proinflammatory and adhesive properties of curli. Furthermore, the outbreak strain produced high levels of CsgD, the essential transcriptional regulator of both curli and cellulose at both 37°C and 28°C in contrast to its closely related Stx-producing EAEC strain HUSEC041. Thus it is highly likely that LB226692 (outbreak strain) is producing exposed curli fibers in the body that trigger a strong inflammatory response, contributing to the outbreak strain's high virulence. The outbreak strain also has the interesting property of generating sector mutants with increased CsgD expression, especially at 37°C. When the strain was grown on Congo red for 7 days at 37°C, sectors appeared with only portions of the colony staining red. These various sectors were further tested by Anja Richter for their level of CsgD expression and unsurprisingly the red stained sectors had the strongest CsgD expression (Richter, Povolotsky *et al.*, submitted). Thus the outbreak strain forms subpopulations capable of highly expressing CsgD and curli fibers. Taking together the high potential of c-di-GMP accumulation, the strong expression of CsgD at 37°C, the ability to form subpopulations, the curli expression but absence of cellulose, represent a unique set of properties of the outbreak strain that may have boosted host colonization, adhesion and inflammatory response and may very well be an important factor of its high virulence and pathogenicity.

6 References

- Adler, J.; B. Templeton (1967) The effect of environmental conditions on the motility of *Escherichia coli*. *J Gen Microbiol* 46(2): 175-184
- Agladze, K., Wang, X. & Romeo, T. (2005) Spatial periodicity of *Escherichia coli* K-12 biofilm microstructure initiates during a reversible, polar attachment phase of development and requires the polysaccharide adhesin PGA. *J Bacteriol* 187, 8237–8246
- Ahmed SA, Awosika J, Baldwin C, Bishop-Lilly KA, Biswas B, Broomall S, Chain PS, Chertkov O, Chokoshvili O, Coyne S, et al., (2012) Genomic comparison of *Escherichia coli* O104:H4 isolates from 2009 and 2011 reveals plasmid, and prophage heterogeneity, including shiga toxin encoding phage stx2. *PLoS one* 7(11):e48228
- Aldridge, P.; Hughes, K. T. (2002) Regulation of flagellar assembly. *Curr Opin Microbiol* 5(2): 160-165
- Al Safadi R, Abu-Ali GS, Sloup RE, Rudrik JT, Waters CM, Eaton KA, Manning SD (2012) Correlation between in vivo biofilm formation and virulence gene expression in *Escherichia coli* O104:H4. *PLoS One* 7: e41628
- Altschul SF, GishW, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J mol Biol* 215:403-410
- Amikam, D., Galperin, M.Y., (2006) PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* 22, 3-6
- Amsler, C. D., M. Cho and P. Matsumura (1993) Multiple factors underlying the maximum motility of *Escherichia coli* as cultures enter post-exponential growth. *J Bacteriol* 175(19): 6238-6244
- Anderson, G. G. et al., (2003) Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* 301, 105–107
- Andrade MO, Alegria MC, Guzzo CR, Docena C, Rosa MC, Ramos CH, Farah CS. (2006) The HD-GYP domain of RpfG mediates a direct linkage between the Rpf quorum-sensing pathway and a subset of diguanylate cyclase proteins in the phytopathogen *Xanthomonas axonopodis* pv *citri*. *Mol. Microbiol.* 62:537–551
- Andreoli, S. P., Trachtman, H., Acheson, D. W., Siegler, R. L. & Obrig, T. G. (2002) Hemolytic uremic syndrome: epidemiology, pathophysiology, and therapy. *Pediatr. Nephrol.* 17, 293–298
- Arnqvist, A., A. Olsen and S. Normark (1994) σ S-dependent growth-phase induction of the *csgBA* promoter in *Escherichia coli* can be achieved in vivo by σ 70 in the absence of the nucleoid-associated protein H-NS. *Mol Microbiol* 13(6): 1021-1032
- Babitzke, P. & Romeo, T. (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology*, 10:2:156-163
- Baghal et al., (2010) Production and immunogenicity of recombinant ferric enterobactin protein (FepA), *International Journal of Infectious Diseases* 14 pp. e166-e170.
- Baikalov, I., I. Schroder, M. Kaczor-Grzeskowiak, K. Grzeskowiak, R.P. Gunsalus, R.E. Dickerson (1996) Structure of the *Escherichia coli* response regulator NarL. *Biochemistry* 35:11053-11061
- Bailey TL, Elkan C (1995) The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol.* 3:21-29
- Barker CS, Prüss BM, Matsumura P (2004) Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the *flhD* operon. *J Bacteriol* 186: 7529-7537
- Barnhart, M.M., Chapman, M.R., (2006) Curli biogenesis and function. *Annu. Rev. Microbiol.* 60, 131-147
- Beloin, C.; Roux, A.; Ghigo, J. M. (2008) *Escherichia coli* biofilms. *Curr Top Microbiol Immunol* 322: 249-289

References

- Bernet-Carnard, M. F., Coconnier, M. H., Hudault, S. & Servin, A. L. (1996) Pathogenicity of the diffusely adhering strain *Escherichia coli* C1845: F1845 adhesin-decay accelerating factor interaction, brush border microvillus injury, and actin disassembly in cultured human intestinal epithelial cells. *Infect. Immun.* 64, 1918–1928
- Bettelheim, K. A.; Breadon, A.; Faiers, M.; O'Farrell, S. M.; Shooter, R. A. (1974) The origin of 'O' serotypes of *Escherichia coli* in babies after normal delivery. *J Hyg* 72, 67–70.
- Bielaszewska, M., Mellmann, A., Zhang, W., Köck, R., Fruth, A., Bauwens, A., Peters, G., H., K., (2011) Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect. Dis* 11, 671-676
- Boehm, A., Kaiser, M., Li, H., Spangler, C., C.A., K., Ackerman, M., Kaefer, V., Sourjik, V., Roth, V., Jenal, U., (2010) Second messenger-mediated adjustment of bacterial swimming velocity. *Cell* 141, 107-116
- Boehm, A., Steiner, S., Zaehring, F., Casanova, A., Hamburger, F., Ritz, D., Keck, W., Ackermann, M., Schirmer, T. and Jenal, U. (2009) Second messenger signalling governs *Escherichia coli* biofilm induction upon ribosomal stress. *Molecular Microbiology*, 72: 1500–1516
- Braeken, K., Moris, M., Daniels, R., Vanderleyden, J., Michiels, J. (2006) New horizons for (p)ppGpp in bacterial and plant physiology. *Trends Microbiol.* 14:45-54
- Branchu P, Hindré T, Fang X, Thomas R, Gomelsky M, Claret L, Harel J, Gobert AP, Martin C (2013) The c-di-GMP phosphodiesterase VmpA absent in *Escherichia coli* K12 strains affects motility and biofilm formation in the enterohemorrhagic O157:H7 serotype. *Vet Immunol Immunopathol* 152: 132-140
- Bray, J. (1945) Isolation of antigenically homogeneous strains of *Bact. Coli neapolitanum* from summer diarrhoea of infants. *J. Pathol. Bacteriol.* 57, 239-247
- Brombacher, E., Dorel, C., Zehnder, A.J.B., Landini, P., (2003) The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in *Escherichia coli*. *Microbiology* 149, 2847–2857
- Brombacher E., Baratto A., Dorel C., Landini P. (2006) Gene expression regulation by the curli activator CsgD protein: modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion. *J. Bacteriol.* 188:2027–2037
- Brown, P. K., C. M. Dozois, C. A. Nickerson, A. Zuppardo, J. Terlonge and R. Curtiss, 3rd (2001) MlrA, a novel regulator of curli (AgF) and extracellular matrix synthesis by *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* 41(2): 349-363
- Casper-Lindley, C. and Yildiz, F.H. (2004) VpsT is a transcriptional regulator required for expression of vps biosynthesis genes and the development of rugose colonial morphology in *Vibrio cholerae* O1 El Tor. *J. Bacteriol.* 186, 1574–1578
- Cassels, F. J. & Wolf, M. K. (1995) Colonization factors of diarrheagenic *E. coli* and their intestinal receptors. *J. Ind. Microbiol.* 15, 214–226
- Chan, C., Paul, R., Samoray, D., Amiot, N., Giese, B., Jenal, U., Schirmer, T., (2004) Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl. Acad. Sci. USA* 101, 17084-17089
- Chang, A.L., Tuckerman, J.R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., Gilles-Gonzalez, M.-A., (2001) Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. *Biochemistry* 40, 3420-3426
- Chevance, F. F.; Hughes, K. T. (2008) Coordinating assembly of a bacterial macromolecular machine. *Nat Rev Microbiol* 6(6): 455-465
- Chilcott, G.S., and Hughes, K.T. (2000) Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. *Microbiol Mol Biol Rev* 64: 694–708

References

- Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., Jenal, U., (2006) Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* 281, 32015-32024
- Christen, M.; Christen, B.; Allan, M. G.; Folcher, M.; Jenoe, P.; et al., (2007) DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci.* 104:4112–4117
- Christen, M., Christen, B., Folcher, M., Schauerte, A., Jenal, U., (2005) Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J. Biol. Chem.* 280, 30829-30837
- Chumakov AM, Silla A, Williamson EA, Koeffler HP. (2007) Modulation of DNA binding properties of CCAAT/enhancer binding protein alt epsilon by heterodimer formation and interactions with NF-kappaB pathway. *Blood* 109: 4209–4219
- Chung, C. T.; Niemela, S. L.; Miller, R. H. (1989) One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* 86, 2172
- Cerca, N., Maira-Litran, T., Jefferson, K.K., Grout, M., Goldmann, D.A., and Pier, G.B. (2007) Protection against *Escherichia coli* infection by antibody to the *Staphylococcus aureus* poly-N-acetylglucosamine surface polysaccharide. *Proc Natl Acad Sci USA* 104: 7528–7533
- Cerca, N., Jefferson, K. K., (2008) Effect of growth conditions on poly-N –acetylglucosamine expression and biofilm formation in *Escherichia coli*. *FEMS Microbiol Lett* 283(1);36-41. PMID: 18445167
- Clermont, O.; Christenson, J. K.; Denamur, E.; Gordon, D. M. (2012) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports.* 5(1):58-65
- Cooper, K. K.; Mandrell, R. E.; Louie, J. W.; Korlach, J.; Clark, T. A.; Parker, C. T.; Huynh, S.; Chain, P. S.; Ahmed, S.; Carter, M. Q. (2014) Comparative genomics of enterohemorrhagic *Escherichia coli* O145:H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157:H7. *BMC Genomics* 15:17
- Costerton, J.W., Stewart, P.S., and Greenberg, E.P. (1999) Bacterial biofilms: a common cause of persistent infections. *Science* 284: 1318–1322
- Cotter, P. A. & Stibitz, S. (2007) c-di-GMP-mediated regulation of virulence and biofilm formation. *Curr Opin Microbiol* 10, 17–23
- Croxen, M. A.; Finlay, B. B. (2009) Molecular mechanisms of *Escherichia coli* pathogenicity. *Nature Reviews Microbiology*, 8(1), 26-38
- Czeczulin, J. R., Balepur, S., Hicks, S., Phillips, A., Hall, R., Kothary, M. H., Nataro, J. P. (1997) Aggregative adherence fimbria II, a second fimbrial antigen mediating aggregative adherence in enteroaggregative *Escherichia coli*. *Infection and immunity*, 65(10), 4135-4145
- Da Re, S. & Ghigo, M. (2006) A CsgD-Independent Pathway for Cellulose Production and Biofilm Formation in *Escherichia coli*. *J. of Bac.* vol. 188 no. 8 3073-3087
- Darfeuille-Michaud, A. (2002) Adherent-invasive *Escherichia coli*: a putative new *E. coli* pathotype associated with Crohn’s disease. *Int. J. Med. Microbiol.* 292, 185–193
- Datsenko, K. A. and B. L. Wanner (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97(12): 6640- 6645
- Dawson, K. G., Emerson, J. C. & Burns, J. L. (1999) Fifteen years of experience with bacterial meningitis. *Pediatr. Infect. Dis. J.* 18, 816–822
- Donnenberg, M. S.; Yu, J.; Kaper, J. B. (1993). A second chromosomal gene necessary for intimate attachment of enteropathogenic *Escherichia coli* to epithelial cells. *J Bacteriol* 175, 4670–4680
- Dow, J.M., Fouhy, Y., Lucey, J.F., Ryan, R.P., (2006) The HD-GYP domain, cyclic di-GMP signaling, and bacterial virulence to plants. *Mol. Plant Microbe Interact.* 19, 1378-1384

References

- Duerig, A., Folcher, M., Abel, S., Nicollier, M., Schwede, T., Amiot, N., Giese, B., Jenal, U., (2009) Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. *Genes Dev.* 23, 93-104
- Dyer, J. G.; Sriranganathan, N.; Nickerson, S. C.; Elvinger, F. (2007) Curli Production and Genetic Relationships Among *Escherichia coli* from Cases of Bovine Mastitis. *Journal of Dairy Science* Volume 90, Issue 1, Pages 193–201
- Egland KA, Greenberg EP. (1999) Quorum sensing in *Vibrio fischeri*: elements of the luxI promoter. *Mol Microbiol.* 31:1197–1204
- Eliot AC, Sandmark J, Schneider G, Kirsch JF (2002) "The dual-specific active site of 7,8-diaminopelargonic acid synthase and the effect of the R391A mutation." *Biochemistry* 41(42):12582-9. PMID: 12379100
- Elliott, S. J., Wainwright, L. A., McDaniel, T. K., Jarvis, K. G., Deng, Y. K., Lai, L. C., McNamara, B. P., Sonnenberg, M. S.; Kaper, J. B. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* 28, 1-4
- Etcheverria, A.I.; Padola, N. L. (2013) Shiga toxin-producing *Escherichia coli*-Factors involved in virulence and cattle colonization. *Virulence* 4(5): 366–372
- Fang, X., Gomelsky, L., (2010) A post-translational, c-di-GMP-dependent mechanism regulating flagellar motility. *Mol. Microbiol.* 76, 1295-1305
- Flynn, J. M., S. B. Neher, et al., (2003) "Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals." *Mol Cell* 11(3): 671-83
- Foxman, B. (2002) Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am. J. Med.* 113(Suppl. 1A):5S-13S
- Frank, C.; Werber, D.; Cramer, J.P.; Askar, M.; Faber, M.; an der Heiden, M.; Bernard, H.; Fruth, A.; Prager, R.; Spode, A.; Wadl, M.; Zoufaly, A.; Jordan, S.; Kemper, M.J.; Follin, P.; Müller, L.; King, L.A.; Rosner, B.; Buchholz, U.; Stark, K.; Krause, G. (2011) Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med.* 10;365(19):1771-80
- Fricke, F. W.; Wright, M. S.; Lindell, A. H.; Harkins, D. M.; Baker-Austin, C.; Ravel, J.; Stepanauskas, R. (2008) Insights into the Environmental Resistance Gene Pool from the Genome Sequence of the Multidrug-Resistant Environmental Isolate *Escherichia coli* SMS-3-5. *J Bacteriol.* 190(20): 6779-6794
- Friedman L, Kolter R. 2004. Two genetic loci produce distinct carbohydrate-rich structural components of the *Pseudomonas aeruginosa* biofilm matrix. *J. Bacteriol.* 186:4457–4465
- Gaal, T., W. Ross, S. T. Estrem, L. H. Nguyen, R. R. Burgess and R. L. Gourse (2001) Promoter recognition and discrimination by EoS RNA polymerase. *Mol Microbiol* 42(4): 939-954
- Gaastra, W.; Svennerholm, A. M. (1996) Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends. Microbiol.* 61, 1267-1282
- Galperin, M.Y., (2004) Bacterial signal transduction network in a genomic perspective. *Environ. Microbiol.* 6, 552-567
- Galperin, M.Y., (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol.* 5, 35
- Galperin, M.Y., Nikolskaya, A.N., Koonin, E.V., (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.* 203, 11-21
- Garcia B, Latasa C, Solano C, Garcia-del Portillo F, Gamazo C, Lasa I. (2004) Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. *Mol. Microbiol.* 54:264–77
- Girgis, H.S., Liu, Y., Ryu, W.S., Tavazoie, S., (2007) A comprehensive genetic characterization of bacterial motility. *PLoS Genetics* 3, e154

References

- Goller C, Wang X, Itoh Y & Romeo T (2006) The cationresponsive protein NhaR of *Escherichia coli* activates pgaABCD transcription, required for production of the biofilm adhesin poly-b-1,6-N-acetyl-D-glucosamine. *J Bacteriol* 188: 8022–8032
- Govan, J.R., and Deretic, V. (1996) Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol Rev* 60: 539–574
- Grigorova, I. L., N. J. Phleger, V. K. Mutalik and C. A. Gross (2006) Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proc Natl Acad Sci U S A* 103(14): 5332-5337
- Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S. (1995) Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in *Escherichia coli* K-12. *Mol Microbiol.* 18(4):661-70
- Hammar, M., Z. Bian and S. Normark (1996) Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*. *Proc Natl Acad Sci U S A* 93(13): 6562- 6566
- Hammer, B. K. & Bassler, B. L. (2009) Distinct sensory pathways in *Vibrio cholerae* El Tor and classical biotypes modulate cyclic dimeric GMP levels to control biofilm formation. *J Bacteriol* 191, 169–177
- Harrison, J. J., Ceri, H. & Turner, R. J. (2007) Multimetal resistance and tolerance in microbial biofilms. *Nat Rev Microbiol* 5, 928–938
- Harrison, J. J., Wade, W. D., Akierman, S., Vacchi-Suzzi, C., Stremick, C. A., Turner, R. J. & Ceri, H. (2009) The chromosomal toxin gene yafQ is a determinant of multidrug tolerance for *Escherichia coli* growing in a biofilm. *Antimicrob Agents Chemother* 53, 2253–2258
- Hatt JK, Rather PN. (2008) Role of bacterial biofilms in urinary tract infections. in *Bacterial Biofilms* (ed. Romeo T.), Springer, Heidelberg 163–192
- Hasegawa, K., Masuda, S., Ono, T.A., (2006) Light induced structural changes of a full-length protein and its BLUF domain in YcgF (Blrp), a blue-light sensing protein that uses FAD (BLUF). *Biochemistry* 45, 3785-3793.
- Hayashi, K., N. Morooka, Y. Yamamoto, K. Fujita, K. Isono, S. Choi, E. Ohtsubo, T. Baba, B. L. Wanner, H. Mori and T. Horiuchi (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* 2: 2006 0007
- He YW, Ng AY, Xu M, Lin K, Wang LH, Dong YH, Zhang LH. (2007) *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol. Microbiol.* 64:281–292
- Hengge, R., (2009) Principles of cyclic-di-GMP signaling. *Nature Rev. Microbiol.* 7, 263-273.
- Hengge, R., (2010) Cyclic-di-GMP reaches out into the bacterial RNA world. *Sci. Signal.* 3, pe44
- Hengge, R., (2011) The general stress response in Gram-negative bacteria. In: Storz, G., Hengge, R. (Eds.), *Bacterial Stress Responses*. ASM Press, Washington, D.C., pp. 251-289
- Hengge-Aronis, R., (2002) Signal transduction and regulatory mechanisms involved in control of the σ^S subunit of RNA polymerase in *Escherichia coli*. *Microbiol. Molec. Biol. Rev.* 66, 373-395
- Hickman, J.W., Harwood, C.S., (2008) Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. *Mol. Microbiol.* 69, 376-389
- Hickman, J.W., Tifrea, D.F., Harwood, C.S., (2005) A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc. Natl. Acad. Sci. USA* 102, 14422-14427
- Hinnebusch, B. J.; Erickson, D. L. (2008) *Yersinia pestis* biofilm in the flea vector and its role in the transmission of plague. In *Bacterial biofilms* (pp. 229-248). Springer Berlin Heidelberg
- Ho, Y.S., Burden, L.M., Hurley, J.H., (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO J.* 19, 5288-5299.
- Hoffman, J. A., Wass, C., Stins, M. F. & Kim, K. S. (1999) The capsule supports survival but not traversal of *Escherichia coli* K1 across the blood–brain barrier. *Infect. Immun.* 67, 3566–3570

References

- Højby, N. (2002) New antimicrobials in the management of cysticfibrosis. *J. Antimicrob. Chemother.* 49: 235-238
- Itoh Y, Rice JD, Goller C, Pannuri A, Taylor J, Meisner J, Beveridge TJ, Preston JFr, Romeo T (2008) Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. *J Bacteriol* 190: 3670-3680
- Jackson KD, Starkey M, Kremer S, Parsek MR, Wozniak DJ. (2004) Identification of psl, a locus encoding a potential exopolysaccharide that is essential for *Pseudomonas aeruginosa* PAO1 biofilm formation. *J. Bacteriol.* 186:4466–75
- Jenal, U., (2004) Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? *Curr. Opin. Microbiol.* 7, 185-191
- Jenal, U., Malone, J., (2006) Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu. Rev. Genet.* 40, 385-407
- Johansson, M. E.V; Holmen Larsson, J. M.; Hansson, G.C. (2010) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host–microbial interactions. *PNAS* 180: 4659-4665
- Johnson JR, Russo TA. (2002) Extraintestinal pathogenic *Escherichia coli*: “the other bad *E coli*.” *J. Lab. Clin. Med.* 139:155–162
- Johnson JR, Russo TA. (2005) Molecular epidemiology of extraintestinal pathogenic (uropathogenic) *Escherichia coli*. *Int. J. Med. Microbiol.* 295:383–404
- Jonas, K., Edwards, A.N., Simm, R., Romeo, T., Römling, U., Melefors, O., (2008) The RNA binding protein CsrA controls c-di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol. Microbiol.* 70, 236-257
- Kader, A., R. Simm, U. Gerstel, M. Morr and U. Römling (2006) Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of *Salmonella enterica* serovar Typhimurium. *Mol Microbiol* 60(3): 602-616
- Kaper, J.B.; Nataro, J.P.; Mobley, H.L.T. (2004) Pathogenic *Escherichia coli* *Nature Reviews Microbiology*, 2 (2), pp. 123-140
- Karlinsky, J. E., S. Tanaka, V. Bettenworth, S. Yamaguchi, W. Boos, S. I. Aizawa; Hughes, K. T. (2000) Completion of the hook-basal body complex of the *Salmonella typhimurium* flagellum is coupled to FlgM secretion and fliC transcription. *Mol Microbiol* 37(5): 1220-1231
- Karmali M A, Steele B T, Petric M, Lim C. (1983) Sporadic cases of haemolytic uremic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet.* 1(8325):619-620
- Krasteva, P. V., Fong, J. C., Shikuma, N. J., Beyhan, S., Navarro, M. V., Yildiz, F. H. & Sondermann, H. (2010) *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* 327, 866–868
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567-580
- Kropec, A.; Maira-Litran, T.; Jefferson, K. K.; Grout, M.; Cramton, S. E.; Gotz, F.; Goldmann, D. A.; Pier, G.B. (2005) *Infect Immun* 73:6868–6876
- Kulesekara, H., Lee, V., Brencic, A., Liberati, N., Urbach, J., Miyata, S., et al., (2006) Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5') -cyclic-GMP in virulence. *Proc Natl Acad Sci USA* 103: 2839–2844
- Kutsukake, K. (1994) Excretion of the anti-sigma factor through a flagellar substructure couples flagellar gene expression with flagellar assembly in *Salmonella typhimurium*. *Mol Gen Genet* 243(6): 605-612
- Lacey, M.; Agasing, A.; Lowry, R.; Green, J. (2013) Identification of the YfgF MASE1 domain as a modulator of bacterial responses to aspartate, *Open Biol* 3: 130046

References

- Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., Breaker, R.R., (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science*, (in press)
- Lee, V.T., Matewish, J.M., Kessler, J.L., Hyodo, M., Hayakawa, Y., Lory, S., (2007) A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. *Mol. Microbiol.* 65, 1474-1484
- Le Quéré, B. & Ghigo, J. (2009) BcsQ is an essential component of the *Escherichia coli* cellulose biosynthesis apparatus that localizes at the bacterial cell pole. *Molecular Microbiology* Volume 72, Issue 3, pages 724–740
- Lindenberg S, Klauck G, Pesavento C, Klauck E, Hengge R (2013) The EAL domain phosphodiesterase YciR acts as a trigger enzyme in a c-di-GMP signaling cascade in *E. coli* biofilm control. *EMBO J* 32: 2001-2014
- Lu, S.; Zhang, X.; Zhu, Y.; Kim, k. S.; Yang, J.; Jin, Q. (2011) Complete Genome Sequence of the Neonatal-Meningitis-Associated *Escherichia coli* Strain CE10. *J of Bacteriology* 193:24 7005
- Lucht, J. M., P. Dersch, B. Kempf and E. Bremer (1994) Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled proU operon of *Escherichia coli*. *J Biol Chem* 269(9): 6578-6578
- Macarasin, D.; Patel, J.; Bauchan, G.; Giron, J.A.; Sharma, V.K. (2012) Role of curli and cellulose expression in adherence of *Escherichia coli* O157:H7 to spinach leaves. *Foodborne Pathog Dis.* 9(2):160-7
- Maeda, H., N. Fujita and A. Ishihama. (2000) Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res* 28(18): 3497-3503
- Malone, J., Williams, R., Christen, M., Jenal, U., Spiers, A.J., Rainey, P.B., (2007) The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain. *Microbiology* 153, 980-994
- Masuda N, Church GM (2002). "Escherichia coli gene expression responsive to levels of the response regulator EvgA." *J Bacteriol* 184(22);6225-34. PMID: 12399493.
- Maurer, J. J., Brown, T. P., Steffens, W. L. & Thayer, S. G. (1998) The occurrence of ambient temperature-regulated adhesins, curli, and the temperature-sensitive hemagglutinin tsh among avian *Escherichia coli*. *Avian Dis* 42, 106–118
- Mazur, O.; Zimmer, J. (2011) Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. *J Biol Chem.* 286(20):17601-6
- McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. & Kaper, J. B. (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl Acad. Sci. USA* 92, 1664–1668
- Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, et al., (2011) Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. *PLoS ONE* 6(7): e22751. doi:10.1371/journal.pone.0022751
- Melton-Celsa, A. R. & O'Brien, A. D. in *Escherichia coli* O157:H7 and Other Shiga Toxin-Producing *E. coli* Strains (eds Kaper, J. B. & O'Brien, A. D.) 121–128 (ASM Press, Washington DC, USA, 1998)
- Méndez-Ortiz, M. M., Hyodo, M., Hayakawa, Y. & Membrillo-Hernández, J. (2006) Genome-wide transcriptional profile of *Escherichia coli* in response to high levels of the second messenger 3',5'-cyclic diguanylic acid. *J Biol Chem* 281, 8090–8099
- Merighi, M., Lee, V.T., Hyodo, M., Hayakawa, Y., Lory, S., (2007) The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* 65, 876-895.
- Miller JH (1972) *Experiments in molecular genetics*, Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory

References

- Minamino, T., Ferris, H.U., Moriya, N., Kihara, M., and Namba, K. (2006) Two parts of the T3S4 domain of the hook-length control protein FliK are essential for the substrate specificity switching of the flagellar type III export apparatus. *J Mol Biol* 362: 1148–1158
- Monteiro C, Saxena I, Wang X, Kader A, Bokranz W, Simm R, Nobles D, Chromek M, Brauner A, Brown Jr RM et al (2009) Characterization of cellulose production in *Escherichia coli* Nissle 1917 and its biological consequences. *Environ Microbiol* 11: 1105- 1116
- Mossoro, C.; Glaziou, P.; Yassibanda, S.; Lan, N. T. P.; Bekondi, C.; Minssart, P.; Bernier, C.; Le Bouguenec, C.; Germani, Y. (2002) Chronic Diarrhea, Hemorrhagic Colitis, and Hemolytic-Uremic Syndrome Associated with HEP-2 Adherent *Escherichia coli* in Adults Infected with Human Immunodeficiency Virus in Bangui, Central African Republic. *J Clin Microbiol.* 40(8): 3086–3088
- Mulvey, M. A. et al., (1998) Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* 282, 1494–1497
- Nash J. H.; Villegas, A.; Kropinski, A. M.; Aguilar-Valenzuela, R.; Konczyk, P.; Mascarenhas, M.; Ziebell, K.; Torres, A. G.; Karmali, M. A.; Coombes, B. K. (2010) Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genomics* 11: 667
- Nataro, J. P., Deng, Y., Maneval, D. R., German, A. L., Martin, W. C., & Levine, M. M. (1992) Aggregative adherence fimbriae I of enteroaggregative *Escherichia coli* mediate adherence to HEP-2 cells and hemagglutination of human erythrocytes. *Infection and immunity*, 60(6), 2297-2304
- Nataro, J. P., Yikang, D., Yingkang, D., & Walker, K. (1994) AggR, a transcriptional activator of aggregative adherence fimbria I expression in enteroaggregative *Escherichia coli*. *Journal of bacteriology*, 176(15), 4691-4699
- Nataro, J. P. & Kaper, J. B. (1998) Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* 11, 142–201
- Nenninger, A. A., L. S. Robinson and S. J. Hultgren (2009). Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc Natl Acad Sci U S A* 106(3): 900-905
- Nickel, J. C., J. W. Costerton, R. J. C. McLean, and M. Olson. 1994. Bacterial biofilms: influence on the pathogenesis, diagnosis, and treatment of urinary tract infections. *J. Antimicrob. Chemother.* 33(Suppl. A):31-41
- Nguyen, Y.; Sperandio, V. (2012) Enterohemorrhagic *E. coli* (EHEC) pathogenesis. *Front Cell Infect Microbiol.* 2012; 2: 90
- Ogasawara, H.; Yamamoto, K.; Ishihama, A. (2011) Role of the Biofilm Master Regulator CsgD in Cross-Regulation between Biofilm Formation and Flagellar Synthesis. *J Bacteriol.*; 193(10): 2587–2597
- Olsen, A., A. Arnqvist, M. Hammar, S. Sukupolvi and S. Normark (1993). The RpoS sigma factor relieves H-NS-mediated transcriptional repression of *csgA*, the subunit gene of fibronectin-binding curli in *Escherichia coli*. *Mol Microbiol* 7(4): 523-536
- Olsen, A., A. Jonsson and S. Normark (1989). Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*. *Nature* 338(6217): 652-655
- Omadjela, O.; Narahari, A.; Strumillo, J.; Melida, H; Mazur, O; Bulone, V.; Zimmer, J. (2013) BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for in vitro cellulose synthesis. *Proc Natl Acad Sci U S A.* 110(44): 17856–17861
- Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* 18, 715–727
- Paul, K., Nieto, V., Carlquist, W.C., Blair, D.F., Harshey, R.M., (2010) The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a “backstop brake” mechanism. *Mol. Cell* 38, 128-139

References

- Pesavento C, Becker G, Sommerfeldt N, Possling A, Tschowri N, Mehli A, Hengge R (2008) Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. *Genes Dev* 22(17): 2434-46. PMID: 18765794
- Picard, B., Garcia, J.S.; Gouriou, S; Duriez, P.; Brahimi, N.; Bingen, E.; Elion, J.; Denamur, E. (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 67: 546-553
- Povolotsky, T.L.; Hengge, R. (2012) 'Life-style' control networks in *Escherichia coli*: Signaling by the second messenger c-di-GMP. *J of Biotechnology* 160(1-2): 10-16
- Powell, B. S., M. P. Rivas, D. L. Court, Y. Nakamura and C. L. Turnbough, Jr. (1994). Rapid confirmation of single copy lambda prophage integration by PCR. *Nucleic Acids Res* 22(25): 5765-5766
- Prentki P, Teter B, Chandler M, Galas DJ (1986) Functional promoters created by the insertion of transposable element IS1. *J Mol Biol* 191: 383-393
- Prigent-Combaret, C., Brombacher, E., Vidal, O., Ambert, A., Lejeune, P., Landini, P., Dorel, C., 2001. Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the *csgD* gene. *J. Bacteriol.* 2001, 7213–7223
- Pristovsek, P.; Sengupta, K.; Löhr, F.; Schäfer, B.; von Trebra, M.W.; Rüterjans, H.; Bernhard, F. (2001) Structural analysis of the DNA-binding domain of the *Erwinia amylovora* RcsB protein and its interaction with the RcsAB box. *J. Biol. Chem.* 278: 17752–17759
- Pruimboom-Brees, I. M.; Morgan, T. W.; Ackermann, M. R.; Nystrom, E. D.; Samuel, J. E.; Cornick, N. A.; Moon, H. W. (2000) Cattle lack vascular receptors for *Escherichia coli* O157:H7 Shiga toxins. *PNAS* vol. 97 no. 19 10325–10329
- Rao, F., Yang, Y., Qi, Y. and Liang, Z.X. (2008) Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*. *J. Bacteriol.*, 190, 3622–3631
- Rasko, D.A.; Rosovitz, M.J.; Myers, G.S.; Mongodin, E.F.; Fricke, W.F.; Gajer, P.; Crabtree, J.; Sebahia, M.; Thomson, N.R.; Chaudhuri, R.; Henderson, I.R.; Sperandio, V.; Ravel, J. (2008) The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* 190:6881–6893
- Riley, L. W.; Remis, R. S.; Helgerson, S. D.; McGee, H. B.; Wells, J. G.; Davis, B. R.; Hebert, R. J.; Olcott, E. S.; Johnson, L. M.; Hargrett, N. T.; Blake, P. A.; Cohen, M. L. (1983) Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* 308(12):681-5
- Robinson, L. S., E. M. Ashman, S. J. Hultgren and M. R. Chapman (2006). Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol Microbiol* 59(3): 870-881
- Römling, U., Amikam, D., (2006) Cyclic di-GMP as a second messenger. *Curr. Opin. Microbiol.* 9, 218-228
- Römling U, Gomelsky M, Galperin MY. (2013) Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiol. Mol. Biol. Rev.* 2013, 77(1):1
- Römling, U., Rohde, M., Olsén, A., Normark, S., Reinköster, J., (2000) AgfD, the checkpoint of multicellular and aggregative behaviour in *Salmonella typhimurium* regulates at least two independent pathways. *Mol. Microbiol.* 36, 10-23
- Römling, U., Sierralta, W.D., Eriksson, K., Normark, S., (1998a) Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the *agfD* promoter. *Mol. Microbiol.* 28, 249-264.
- Römling, U.; Bian, Z.; Hammar, M.; Sierralta, W. D.; Normark, S. (1998b) Curli Fibers Are Highly Conserved between *Salmonella typhimurium* and *Escherichia coli* with Respect to Operon Structure and Regulation. *J Bact.* 180(3): 722-731

References

- Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de Vroom, E., van der Marel, G.A., van Boom, J.H., Benziman, M., (1987) Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylate. *Nature* 325, 279-281
- Ross, P., Mayer, R. & Benziman, M. (1991) Cellulose biosynthesis and function in bacteria. *Microbiol Rev* 55, 35-58
- Rupp, M. E., Ulphani, J. S., Fey, P. D., Mack, D. (1999) Characterisation of staphylococcus epidermidis polysaccharide intercellular adhesin/haemagglutinin in the pathogenesis of intravascular catheter associated infection in a rat model. *Infect. Immun.* 67: 2656-2659
- Russo, T. A.; Johnson, J.R. (2000) Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. *The Journal of infectious diseases* 181: 1753-1754
- Ryan, R.P., Fouhy, Y., Lucey, J.F., Crossman, L.C., Spiro, S., He, Y.-W., Zhang, L.-H., Heeb, S., Williams, P., Dow, J.M., (2006) Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci. USA* 103, 6712-6717
- Ryan, R.P., McCarthy, Y., Andrade, M., Farah, C.S., Armitage, J.P., Dow, J.M., (2010) Cell-cell signal-dependent dynamic interactions between HD-GYP and GGDEF domain proteins mediate virulence in *Xanthomonas campestris*. *Proc. Natl. Acad. Sci. USA* 107, 5989-5994
- Ryjenkov, D.A., Simm, R., Römling, U., Gomelsky, M., (2006) The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ protein YcgR controls motility in enterobacteria. *J. Biol. Chem.* 281, 30310-30314
- Ryjenkov, D.A., Tarutina, M., Moskvina, O.V., Gomelsky, M., (2005) Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.* 187, 1792-1798
- Sahl, J.W.; Steinsland, H.; Redman, J.C. Angiuoli, S. V.; Nataro, J.P.; Sommerfelt, D.; Rasko, D. A. (2011) A comparative genomic analysis of diverse clonal types of enterotoxigenic *Escherichia coli* reveals pathovar-specific conservation. *Infect Immun* 79: 950-960. doi: 10.1128/iai.00932-10
- Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular Cloning: a laboratory manual*, 2nd edition. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.
- Scaletsky, I. C. A.; Fabbicotti, S. H.; Aranda, K. R.; Morais, M. B.; Fagundes-Neto, U. (2002) Comparison of DNA hybridization and PCR assays for detection of putative pathogenic enteroadherent *Escherichia coli*. *J Clin Microbiol* 40, 1254-1258
- Schierack, P.; Romer, A.; Jores, J.; Kaspar, H.; Guenther, S.; Filter, M.; Eichberg, J.; Wieler, L.H. (2009) Isolation and characterization of intestinal *Escherichia coli* clones from wild boars in Germany. *Appl Environ Microbiol* 75: 695-702
- Schirmer, T., Jenal, U., (2009) Structural and mechanistic determinants of c-di-GMP signalling. *Nat. Rev. Microbiol.* 7, 724-735
- Schmidt, A.J., Ryjenkov, D.A., Gomelsky, M., (2005) The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J. Bacteriol.* 187, 4774-4781
- Serra-Moreno, R.; Jofre, J.; Muniesa, M. (2007) Insertion site occupancy by stx2 bacteriophages depends on the locus availability of the host strain chromosome. *J Bacteriol* 189: 6645-6654.
- Serra, D. O.; Hengge, R. (2014) Stress responses go three dimensional - the spatial order of physiological differentiation in bacterial macrocolony biofilms. *Environ Microbiol.* 16(6):1455-71
- Serra, D. O.; Richter, A. M.; Hengge, R. (2013a) Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms. *J Bacteriol* 195: 5540-5554

References

- Serra DO, Richter AM, Klauck G, Mika F, Hengge R (2013b) Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. *MBio* 4(2): e00103-00113
- Serre L, Pereira de Jesus K, Zelwer C, Bureaud N, Schoentgen F, Benedetti H (2001). "Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein." *J Mol Biol* 2001; 310(3):617-34. PMID: 11439028
- Silhavy, T. J.; Berman, M. L.; Enquist, L. W. (1984). *Experiments with gene fusions*. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press
- Simm R., Morr M., Kader A., Nimtz M., Römling U. (2004) GGDEF and EAL domains inversely regulate cyclic d-GMP levels and transition from sessility to motility. *Mol. Microbiol.* 53:1123–1134
- Simm, R., Lusch, A., Kader, A., Andersson, M., Römling, U., (2007) Role of EAL-containing proteins in multicellular behavior of *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* 189, 3613-3623
- Simons, R. W., F. Houman and N. Kleckner (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene* 53(1): 85-96
- So, M.; Boyer, H. W.; Betlach, M.; Falkow, S. (1976) Molecular cloning of an *Escherichia coli* plasmid determinant that encodes for the production of heat-stable enterotoxin. *Journal of bacteriology*, 128(1), 463-472
- So, M., Dallas, W. S., & Falkow, S. (1978) Characterization of an *Escherichia coli* plasmid encoding for synthesis of heat-labile toxin: molecular cloning of the toxin determinant. *Infection and immunity*, 21(2), 405-411
- Solano, C., B. Garcia, J. Valle, C. Berasain, J. M. Ghigo, C. Gamazo, and I. Lasa. (2002) Genetic analysis of *Salmonella enteritidis* biofilm formation: critical role of cellulose. *Mol. Microbiol.* 43:793-808
- Soutourina, O. A.; Bertin, P. N. (2003) Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiol Rev* 27(4): 505-523
- Smith, J. L.; Fratamico, P. M.; Gunther, N. W. (2007) Extraintestinal pathogenic *Escherichia coli*. *Foodborne pathogens and disease* 4: 134-163
- Smith, R. F.; Wiese, B. A.; Wojzynski, M. K.; Davison, D. B.; Worley, K. C. (1996) BCM Search Launcher—an integrated interface to molecular biology data base search and analysis services available on the World Wide Web. *Genome Res* 6:454 -462
- Spurbeck RR, Tarrien RJ, Mobley HL. (2012) Enzymatically active and inactive phosphodiesterases and diguanylate cyclases are involved in regulation of motility or sessility in *Escherichia coli* CFT073. *MBio.*;13
- Standal, R., Iversen, T.G., Coucheron, D.H., Fjaervik, E., Blatny, J.M., and Valla, S. (1994) A new gene required for cellulose production and a gene encoding cellulolytic activity in *Acetobacter xylinum* are colocalized with the *bcs* operon. *J Bacteriol* 176: 665–672.
- Stewart, P. S., Costerton, J. W. (2001) Antibiotic resistance of bacteria in biofilms. *Lancet* 358, 135 ± 138
- Stoner GL, Eisenberg MA (1975) "Biosynthesis of 7, 8-diaminopelargonic acid from 7-keto-8-aminopelargonic acid and S-adenosyl-L-methionine. The kinetics of the reaction." *J Biol Chem* 1975;250(11);4037-43. PMID: 1092682
- Stoner GL, Eisenberg MA (1975) (1975) "Purification and properties of 7, 8-diaminopelargonic acid aminotransferase." *J Biol Chem.*;250(11);4029-36. PMID: 1092681
- Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., Breaker, R.R. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321, 411-413
- Suzuki, K., Babitzke, P., Kushner, S.R., Romeo, T. (2006) Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev.* 20, 2605-2617

References

- Sweeney, N. J., P. Klemm, B. A. McCormick, E. Moller-Nielsen, M. Utey, M. A. Schembri, D. C. Laux, and P. S. Cohen. (1996) The *Escherichia coli* K-12 *gntP* gene allows *E. coli* F-18 to occupy a distinct nutritional niche in the streptomycin-treated mouse large intestine. *Infect. Immun.* 64:3497-3503
- Takahashi, H. and T. Shimizu (2006) Phosphodiesterase activity of Ec DOS, a hemeregulated enzyme from *Escherichia coli*, toward 3',5'-cyclic diguanylic acid is obviously enhanced by O₂ and CO binding. *Chem Lett* 35(8): 970-971
- Takaya, A.; Erhardt, M.; Karata, K.; Winterberg, K.; Yamamoto, T.; Hughes, K. T. (2012) YdiV: a dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex. *Mol. Micro.* 83(6): 1268-1284
- Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P., Benziman, M., (1998) Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.* 180, 4416-4425
- Tamayo, R., Tischler, A.D., Camilli, A. (2005) The EAL domain protein *VieA* is a cyclic diguanylate phosphodiesterase. *J. Biol. Chem.* 280, 33324-33330
- Tani, T. H., A. Khodursky, R. M. Blumenthal, P. O. Brown and R. G. Matthews (2002) Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci U S A* 99(21): 13471-13476
- Tarutina, M., Ryjenkov, D. A. & Gomelsky, L. An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. *J. Biol. Chem.* 2006;281:34751–34758
- Taylor BL. 2007. Aer on the inside looking out: paradigm for a PAS-HAMP role in sensing oxygen, redox and energy. *Mol. Microbiol.* 65:1415–24
- Taylor, J. D., Zhou, Y., Salgado, P. S., Patwardhan, A., McGuffie, M., Pape, T., Grabe, G., Ashman, E., Constable, S. C., Simpson, P. J. et al., (2011) Atomic Resolution Insights into Curli Fiber Biogenesis. *Structure* 19: 1307–1316
- Tischler, A.D., Camilli, A., (2005) Cyclic diguanylate regulates *Vibrio cholerae* virulence gene expression. *Infect. Immun.* 73, 5873-5882
- Tagliabue, L., Antoniani, D., Maciag, A., Bocci, P., Raffaelli, N., & Landini, P. (2010) The diguanylate cyclase YddV controls production of the exopolysaccharide poly-N-acetylglucosamine (PNAG) through regulation of the PNAG biosynthetic *pgaABCD* operon. *Microbiology*, 156(10), 2901-2911
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. (1997) The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acid.s Res* 25:4876-4882
- Tomoyasu, T., A. Takaya, E. Isogai and T. Yamamoto (2003) Turnover of FlhD and FlhC, master regulator proteins for *Salmonella* flagellum biogenesis, by the ATP-dependent ClpXP protease. *Mol Microbiol* 48(2): 443-452
- Tschowri, N., S. Busse, and R. Hengge. 2009. The BLUF-EAL protein YcgF acts as a direct anti-repressor in a blue-light response of *Escherichia coli*. *Genes Dev.* 23:522-534
- Tuckerman, J.R., Gonzalez, G., Gilles-Gonzalez, M.-A. (2011) Cyclic di-GMP activation of polynucleotide phosphorylase signal-dependent RNA processing. *J. Mol. Biol.* 407, 633-639
- Tükel, C.; Raffatellu, M.; Humphries, A. D.; Wilson, R. P.; Andrews-Polymeris, H. L.; Gull, T.; Figueiredo, J. F.; Wong, M. H.; Michelsen, K. S.; Akçelik, M. et al., (2005) CsgA is a pathogen-associated molecular pattern of *Salmonella enterica* serotype Typhimurium that is recognized by Toll-like receptor 2. *Mol. Microbiol.*, 58 (2005), pp. 289–304

References

- Tükel, C.; Wilson, P.; Nishimori, J. H.; Pezeshki, M.; Chromy, B. A.; Bäumlner, A. J. (2009) Responses to Amyloids of Microbial and Host Origin Are Mediated through Toll-like Receptor 2. *Cell Host & Microbe* 6(1):45-53
- Tusnady GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849-850
- Typas, A., G. Becker and R. Hengge (2007) The molecular basis of selective promoter activation by the σ^S subunit of RNA polymerase. *Mol Microbiol* 63(5): 1296-1306
- Typas, A. & R. Hengge (2005) Differential ability of σ^S and σ^{70} of *Escherichia coli* to utilize promoters containing half or full UP-element sites. *Mol Microbiol* 55(1): 250- 260
- Typas, A. & R. Hengge (2006) Role of the spacer between the -35 and -10 regions in sigma promoters selectivity in *Escherichia coli*. *Mol Microbiol* 59(3): 1037-1051
- Uhlich, G. A.; Keen, J. E.; Elder, R. O. (2001) Mutations in the *csgD* Promoter Associated with Variations in Curli Expression in Certain Strains of *Escherichia coli* O157:H7. *Appl Environ Microbiol.* 67(5): 2367–2370
- Uhlich, G. A.; Chen, C.; Cottrell, B. J.; Hofmann, C. S.; Strobaugh, T. P.; Nguyen, L. (2013) Phage insertion in *mlrA* and variations in *rpoS* limit curli expression and biofilm formation in *Escherichia coli* serotype O157:H7. *Microbiology* 159:1586-1596
- Von Heijne G, Gavel Y (1988) Topogenic signals in integral membrane proteins. *Eur J Biochem* 174:671 -678
- Wada, T.; Morizane, T.; Abo, T.; Tominaga, A.; Inoue-Tanaka, K.; Katsukake, K. (2011). EAL domain protein YdiV acts as an anti-FlhD4C2 factor responsible for nutritional control of the flagellar regulon in *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* 193:1600–1611
- Wada T, Hatamoto Y, Kutsukake K. (2012) Functional and expressional analyses of the anti-FlhD4C2 factor gene *ydiV* in *Escherichia coli*. *Microbiology* 158:1533–1542
- Wan, X., Tuckerman, J.R., Saito, J.A., Freitas, T.A.K., Newhouse, J.S., Denery, J.R., Galperin, M.Y., Gonzalez, G., Gilles-Gonzalez, M.-A., Alam, M., (2009) Globins synthesize the second messenger bis-(3'-5')-cyclic diguanosine monophosphate in bacteria. *J. Mol. Biol.* 388, 262-270
- Wang X, Preston JF, Romeo T (2004) "The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation." *J Bacteriol* 186(9);2724-34. PMID: 15090514
- Waters CM, Lu W, Rabinowitz JD, Bassler BL. (2008) Quorum sensing controls biofilm formation in *Vibrio cholerae* through modulation of cyclic di-GMP levels and repression of *vpsT*. *J. Bacteriol.* 190:2527–2536
- Weber, H., Pesavento, C., Possling, A., Tischendorf, G., Hengge, R., (2006) Cyclic-di-GMP-mediated signaling within the σ^S network of *Escherichia coli*. *Mol. Microbiol.* 62, 1014-1034
- Weber, H., T. Polen, J. Heuveling, V. F. Wendisch and R. Hengge (2005) Genome-wide analysis of the general stress response network in *Escherichia coli*: σ^S -dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* 187(5): 1591-1603
- Whittam T.S. (1996) Genetic variation and evolutionary processes in natural populations of *Escherichia coli*. In: Neidhardt F.C., editor. *Escherichia coli and Salmonella: Cellular and molecular biology*. ASM Press; Washington, DC: pp. 2708–2720
- Wolf, M. K. (1997) Occurrence, distribution, and associations of O and H serogroups, colonization factor antigens, and toxins of enterotoxigenic *Escherichia coli*. *Clin. Microbiol. Rev.* 10, 569–584
- Yildiz, F.H. et al., (2001) *VpsR*, a Member of the Response Regulators of the Two-Component Regulatory Systems, Is Required for Expression of *vps* Biosynthesis Genes and EPS(ETr)-Associated Phenotypes in *Vibrio cholerae* O1 El Tor. *J. Bacteriol.* 183, 1716-1726

References

- Zhai Y, Saier MH Jr (2002) A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. *J Mol Microbiol Biotechnol* 4:375-377
- Zhai Y, Tchieu J, Saier MH Jr (2002) A Web-based Tree View (TV) program for the visualization of phylogenetic trees. *J Mol Microbiol Biotechnol* 4:69-70
- Zhao, K., Liu, M., Burgess, R.R., (2007) Adaptation in bacterial flagellar and motility systems: from regulon members to "foraging"-like behavior in *E. coli*. *Nucl. Acids Res.* 35, 4441-4452
- Zogaj X., Nimtz N., Rohde M., Bokranz W., Romling U. (2001) The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as second component of the extracellular matrix. *Mol. Microbiol.* 39:1452–1463