A study of sequence and function divergence of pilz proteins

Cheang, Qing Wei

2018

Cheang, Q. W. (2018). A study of sequence and function divergence of pilz proteins.
Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73213>

<https://doi.org/10.32657/10356/73213>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**A STUDY OF SEQUENCE AND FUNCTION
DIVERGENCE OF PILZ PROTEINS**

Cheang Qing Wei

School of Biological Sciences

2017

A STUDY OF SEQUENCE AND FUNCTION DIVERGENCE OF PILZ PROTEINS

Cheang Qing Wei

School of Biological Sciences

A thesis submitted to the Nanyang Technological
University in fulfillment of the requirement for the
degree of Doctor of Philosophy

2017

Acknowledgements

First and foremost, I would like to thank my supervisor, Associate Professor Liang Zhao-xun for the invaluable guidance, advice, support and limitless patience he has shown me during these years. I am indebted to NTU and MOE Singapore for the scholarship and grants necessary for me to pursue my research interests.

I would also like to thank my former and current colleagues, Limei, Lingyi, Zhenjie, Hoa, Howard, Alolika, Prabha, Linghui, Ishin, Sheng Shuo, Ruijuan, Lifeng and Hui Hua. Their invaluable intellectual input was always helpful and their companionship in and outside of the laboratory never fails to brighten my day.

I would also want to thank our collaborators, Zi Jing from SCELSE, Yukai from BII and finally Linghui who is currently in China, much of this work would not have been possible without the help from all these great people.

Next, I would like to express my gratitude to my Thesis Advisory Committee, Prof. Gruber and Prof. Oliver, their advice helped me massively during the period when I was feeling lost.

Finally, I would like to thank everyone I've met during this adventure in NTU, this experience would not have been the same without you people. This was a journey well-travelled.

Contents

List of Figures	v
List of Tables.....	viii
Abbreviations	ix
Abstract	1
Chapter 1 – Introduction.....	3
1.1 Cellular cyclic di-GMP concentration is controlled by diguanylate cyclases (DGCs) and phosphodiesterases (PDEs).....	3
1.1.1 Diguanylate cyclases	4
1.1.2 Cyclic di-GMP Phosphodiesterases	6
1.2 PilZ family of cyclic di-GMP receptors.....	9
1.2.1 Single-domain PilZ proteins	10
1.2.2 Di-domain PilZ proteins.....	13
1.2.3 Multi-domain PilZ proteins	15
1.3 Cyclic di-GMP signaling in the model organism <i>P. aeruginosa</i> PAO1	16
1.4 Objectives.....	18
Chapter 2 – Sequence and function divergence of PilZ domains revealed by sequence similarity networking (SSN) analysis	20
2.1 Introduction	20
2.2 Materials and Methods.....	24
2.3 Results and Discussion	26
2.3.1 Classification of PilZ proteins into single, di- and multi-domain proteins	26
2.3.2 SSN of single domain PilZ proteins.....	27
2.3.2.1 Cluster 1	28
2.3.2.2 Cluster 2	30
2.3.2.3 Cluster 3	32
2.3.2.4 Cluster 4	34
2.3.2.5 Cluster 5	35
2.3.2.6 Cluster 6	36
2.3.2.7 Cluster 7	37
2.3.2.8 Cluster 8	37
2.3.2.9 Cluster 10	38
2.3.2.10 Cluster 9, 11 and PA0012	39
2.3.2.11 DgrA and DgrB clusters	40

2.3.3	SSN of di-domain PilZ proteins.....	40
2.3.3.1	Cluster 1	40
2.3.3.2	Cluster 2	43
2.3.3.3	Cluster 3	43
2.3.3.4	Cluster 4	45
2.3.3.5	Cluster 5	45
2.3.3.6	Cluster 6	48
2.3.3.7	Cluster 7	50
2.3.3.8	Cluster 8	50
2.3.3.9	Cluster 9	51
2.3.3.10	PlzA (BB0733) cluster	51
2.3.3.11	xcc-b100_2234 cluster	52
2.4	Discussion.....	52
2.5	Conclusion.....	55
Chapter 3 – Probing the function of two single-domain PilZ proteins in <i>P. aeruginosa</i> PAO1.....		56
3.1	Introduction	56
3.2	Materials and Methods.....	58
3.3	Results.....	65
3.3.1	Phenotypes of Δ PA4324 and Δ PA0012 mutants.....	65
3.3.2	Deletion of PA0012 and PA4324 had little impact on bacteria motility	65
3.3.3	The PA0012 and PA4324 mutants exhibit similar colony morphology as the wild type PAO1.....	67
3.3.4	Deletion of PA0012 and PA4324 did not cause significant changes in antibiotic susceptibility	68
3.3.5	Deletion of PA0012 and PA4324 changed gene expression profile based on transcriptome analysis ^b	69
3.3.6	Gene expression profile of PA0012 mutants	70
3.3.7	T6SS bacterial competition assay.....	72
3.3.8	Gene expression profile of PA4324 mutants	74
3.3.9	Quantification of pyocyanin and PQS and analysis of the transcriptional activity of the pqs operon ^c	76
3.3.10	Identification of protein partners of PA0012 and PA4324 by protein pull-down assays	78
3.4	Discussion.....	79
3.5	Conclusion.....	81

Chapter 4 – Probing the function of the two di-domain PilZ proteins in <i>P. aeruginosa</i> PAO1	82
4.1 Introduction	82
4.2 Materials and Methods.....	84
4.3 Results.....	86
4.3.1 PA2989 and FlgZ mutants have altered motility modes compared to PAO1	86
4.3.2 Deletion of PA2989 results in a wrinkled colony	89
4.3.3 Deletion of PA2989 causes changes to antibiotic susceptibility.....	90
4.3.4 Deletion of PA2989 down-regulates protective genes as shown by transcriptome analysis ^b	91
4.3.5 FlgZ regulates flagellar rotation speed and torque in PAO1.....	94
4.3.6 PA2989 represents a class of structurally novel di-domain PilZ proteins with an inserted α -helical domain	96
4.4 Discussion.....	99
4.5 Conclusion.....	100
Chapter 5 – Conclusion and Future directions.....	101
References	105

List of Figures

Figure	Title	Page
1	Cellular functions regulated by cyclic di-GMP	4
2	Structure of the diguanylate cyclase PleD	5
3	Structures and active sites of EAL domain and HD-GYP domain based phosphodiesterases.	8
4	Cyclic di-GMP binding proteins in <i>P. aeruginosa</i> PAO1	10
5	Structures of single and di-domain PilZ proteins.	11
6	Comparison of a phylogenetic tree and its corresponding SSN.	21
7	Histogram of the sequence length of PilZ proteins	27
8	The single-domain PilZ SSN	29
9	Clusters from the single domain SSN that contain more than 60 nodes.	31
10	HapZ is located in single-domain PilZ cluster 1	32
11	MapZ and PlzA are located in single-domain PilZ cluster 3.	34
12	Single-domain PilZ cluster 5 contains PXO_02374 and PA4324	35
13	Sequence logo for single-domain PilZ cluster 6 from the order <i>Bacillales</i> .	36
14	Closeup of the interaction site between XAC1133 and FimX _{xac} and sequence logo for single-domain PilZ cluster 8.	38
15	Sequence logo for single-domain PilZ cluster 10 that is predicted to not bind cyclic di-GMP	39
16	The di-domain PilZ SSN	42
17	Clusters from the single domain SSN that contain more than 60 nodes.	42
18	Di-domain PilZ cluster 3 and structure of PlzD	44
19	The residue prior to the RxxxR motif determines the cyclic di-GMP binding mode of the di-domain PilZ proteins	45
20	Sequence logo for di-domain PilZ cluster 5a.	47
21	Structure of MrkH in complex with cyclic di-GMP and sequence logo for MrkH and its neighboring nodes.	48
22	Sequence logo for di-domain PilZ cluster 6a with the putative binding motifs indicated	49
23	Sequence logo for di-domain PilZ cluster 7a which has two sets of cyclic di-GMP binding motifs	50
24	Sequence logo for di-domain PilZ cluster 9 that contains incomplete cyclic di-GMP binding motifs and HTH region.	51
25	Structure of xcc-b100_2234 homo-tetramer	52
26	Representative images of PA0012 and PA4324 swarming motility plates	66
27	Representative images of PA0012 and PA4324 swimming motility plates	66
28	Representative images of PA0012 and PA4324 twitching motility plates	67
29	Representative images of PA0012 and PA4324 colonies after 5 days.	68
30	Total gene change in the Δ PA0012 mutant.	71

List of Figures (Continued)

Figure	Title	Page
31	Deletion of <i>PA0012</i> increases expression of the Type VI Secretion System HSI-I genes cluster	72
32	T6SS bacterial competition assay	73
33	Total gene change in the $\Delta PA4324$ mutant strain	74
34	Deletion of <i>PA4324</i> increases expression of the <i>pqs</i> operon.	76
35	Quantification of pyocyanin and PQS; and transcriptional activity of the <i>pqsA</i> promoter over 45 hours.	77
36	Representative images of silver-stained gels of pull down experiments.	79
37	Biosynthetic pathways for PQS and pyocyanin share a common substrate	81
38	Representative images of <i>PA2989</i> and <i>FlgZ</i> swarming motility plates	87
39	Representative images of <i>PA2989</i> and <i>FlgZ</i> swimming motility plates	88
40	Representative images of <i>PA2989</i> and <i>FlgZ</i> twitching motility plates	89
41	Representative images of <i>PA2989</i> and <i>FlgZ</i> colonies after 5 days	90
42	Total gene change in $\Delta PA2989$ mutant.	92
43	Deletion of <i>PA2989</i> decreases expression of protective genes	93
44	The flagellar motor of $\Delta FlgZ$ mutants rotate at a slower speed than wild type PAO1.	96
45	Predicted 3D structure of PA2989	98
46	Model of PA2989 in complex with cyclic di-GMP	99

Appendix

Figure	Title	Page
S1	Sequence logos for single-domain PilZ cluster 1.	115
S2	Sequence logos for single-domain PilZ cluster 2	116
S3	Sequence logos for single-domain PilZ cluster 3.	118
S4	Sequence logos for single-domain PilZ cluster 4.	120
S5	Sequence logos for single-domain PilZ cluster 5.	121
S6	Sequence logos for single-domain PilZ cluster 6	123
S7	Sequence logos for single-domain PilZ cluster 7	124
S8	Sequence logo for single-domain PilZ cluster 8	125
S9	Sequence logo for single-domain PilZ cluster 9	127
S10	Sequence logo for single-domain PilZ cluster 10	128
S11	Sequence logo for single-domain PilZ cluster 11	129
S12	Sequence logo for the cluster that includes PA0012.	130
S13	Sequence logo for the single-domain PilZ clusters that include DgrA and DgrB	131
S14	Sequence logo for di-domain PilZ cluster 1.	132

List of Figures (continued)

Figure	Title	Page
S15	Sequence logo for di-domain PilZ cluster 2	133
S16	Sequence logo for di-domain PilZ cluster 3	135
S17	Sequence logo for di-domain PilZ cluster 4	138
S18	Sequence logo for di-domain PilZ cluster 5	139
S19	Sequence logo for di-domain PilZ cluster 6	142
S20	Sequence logo for di-domain PilZ cluster 7.	144
S21	Sequence logo for di-domain PilZ cluster 8.	146
S22	Sequence logo for di-domain PilZ cluster 9	147
S23	Sequence logo for the di-domain PilZ cluster that includes PlzA.	148
S24	Sequence logo for the di-domain PilZ cluster that includes xcc-b100_2234.	149
S25	Complete histogram of the sequence length of PilZ proteins.	150
S26	Bacteria two-hybrid screen of potential FlgZ partners.	151

List of Tables

Table	Title	Page
1	Susceptibility of $\Delta PA0012$, $\Delta PA4324$ and PAO1 to various antibiotics.	69
2	Susceptibility of $\Delta PA2989$ and PAO1 strain to various antibiotics.	90
3	Di-domain PilZ proteins and number of residues between their RxxxR and DxSxxG motifs.	97
Appendix		
Table	Title	Page
S1	List of bacteria genera in single-domain PilZ cluster 1	115
S2	List of bacteria genera in single-domain PilZ cluster 2	117
S3	List of bacteria genera in single-domain PilZ cluster 3	119
S4	List of bacteria genus in single-domain PilZ cluster 4	120
S5	List of bacteria genera in single-domain PilZ cluster 5	122
S6	List of bacteria genera in single-domain PilZ cluster 6	123
S7	List of bacteria genera in single-domain PilZ cluster 7	124
S8	List of bacteria genera in single-domain PilZ cluster 8	126
S9	List of bacteria genera in single-domain PilZ cluster 9	127
S10	List of bacteria genera in single-domain PilZ cluster 10	128
S11	List of bacteria genera in single-domain PilZ cluster 11	129
S12	List of bacteria genera in the single-domain PilZ cluster that includes PA0012	130
S13	List of bacteria genus in the single-domain PilZ clusters that includes DgrA and DgrB	131
S14	List of bacteria genera in di-domain PilZ cluster 1	132
S15	List of bacteria genera in di-domain PilZ cluster 2	134
S16	List of bacteria genera in di-domain PilZ cluster 3	137
S17	List of bacteria genera in di-domain PilZ cluster 4	138
S18	List of bacteria genera in di-domain PilZ cluster 5	141
S19	List of bacteria genera in di-domain PilZ cluster 6	143
S20	List of bacteria genera in di-domain PilZ cluster 7	145
S21	List of bacteria genera in di-domain PilZ cluster 8	146
S22	List of bacteria genera in di-domain PilZ cluster 9	147
S23	List of bacteria genera in the di-domain PilZ cluster that includes PlzA (BB0733)	148
S24	List of bacteria genera in the di-domain PilZ cluster that includes xcc-b100_2234	149
S25	List of bacteria strains and plasmids used.	152
S26	List of primers used.	154
S27	Genes up-regulated in $\Delta PA0012$	156
S28	Genes down-regulated in $\Delta PA0012$	159
S29	Genes up-regulated in $\Delta PA324$	162
S30	Genes down-regulated in $\Delta PA4324$	167
S31	Genes up-regulated in $\Delta PA2989$	170
S32	Genes down-regulated in $\Delta PA2989$	173

Abbreviations

cyclic di-GMP	bis-(3'-5')-cyclic dimeric guanosine monophosphate
DGC	Diguanylate cyclase
PDE	Phosphodiesterase
<i>G. xylinus</i>	<i>Gluconacetobacter xylinus</i>
GTP	Guanosine Triphosphate
<i>C. crescentus</i>	<i>Caulobacter crescentus</i>
I-site	Inhibitory site
<i>X. campestris</i>	<i>Xanthomonas campestris</i>
<i>E. coli</i>	<i>Escherichia coli</i>
<i>K. pneumonia</i>	<i>Klebsiella pneumonia</i>
T4P	Type IV Pili
<i>Xcc</i>	<i>Xanthomonas campestris</i> pv. <i>campestris</i>
<i>Xac</i>	<i>Xanthomonas axonopodis</i> pv. <i>citri</i>
PAO1	<i>Pseudomonas aeruginosa</i> PAO1
<i>Xoo</i>	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
<i>B. burgdorferi</i>	<i>Borrelia burgdorferi</i>
<i>V. cholerae</i>	<i>Vibrio cholerae</i>
SSN	Sequence Similarity Network
Uniprot	UniProt Knowledgebase
<i>S. purpureogeniscleroticus</i>	<i>Streptomyces purpureogeniscleroticus</i>
REC	Response receiver domain
HTH	helix-turn-helix
NDM-1	New Delhi metallo- β -lactamase-1
LB	lysogeny broth
PCR	Polymerase Chain Reaction
HA	haemagglutinin
CFU	colony forming units
T6SS	Type 6 secretion system
RNA seq	RNA sequencing
HHQ	2-heptyl-4-quinolone
PQS	<i>Pseudomonas</i> quinolone signal
PCA-1	phenazine-1-carboxylic acid

Abstract

Cyclic di-GMP is a near-ubiquitous second messenger that mediates important cellular functions in many species of bacteria. Cyclic di-GMP binds to a diverse range of proteins receptors/effectors. The PilZ protein family is one of the largest families of specialized cyclic di-GMP receptors. Currently more than 26, 000 PilZ proteins are recorded in the InterPro database and most these proteins have unknown function. In this research project, I performed a large-scale analysis of single and di-domain PilZ proteins using the sequence similarity networking (SSN) tool and then analysed the largest clusters in detail; and generated sequence logos to unveil highly conserved residues that may be important for the function of the PilZ proteins. One of the most important observations is that PilZ proteins exhibit enormous sequence divergence and share little in conservation amongst different clusters, which suggest that PilZ proteins are likely to have different cellular functions.

In addition of the large-scale bioinformatic analysis, I also performed biochemical and genetic experiments to probe the biological function of four uncharacterized PilZ proteins (PA0012, PA4324, PA2989 and PA3353) from the opportunistic pathogen *P. aeruginosa* PAO1. By transcriptome analysis, I found that the single domain PilZ protein PA0012 specifically suppresses genes of the anti-bacterial HSI-I gene cluster. This observation is in accordance with bacterial competition assays which showed that the $\Delta PA0012$ strains displayed enhanced killing of the prey bacteria whereas the *PA0012* overexpression strain displayed reduced killing of prey bacteria. Meanwhile, the second single-domain PilZ protein PA4324 was found to regulate the expression of the *pqs* operon and other genes under the control of the PQS signalling network.

I discovered that the di-domain PilZ protein PA2989 likely plays a protective role because many genes that play protective roles were down-regulated when *PA2989* was deleted. Furthermore, I found that PA2989 likely represents a unique class of di-domain PilZ proteins that contain an inserted α -helical domain between the two cyclic di-GMP-binding motifs. For the second di-domain protein PA3353, we found that the protein is involved in the regulation of flagellar motor output by using cell-tethering assay and swimming speed analysis. Based on the observations, I have

proposed a model where PA3353 (also known as FlgZ) mediates flagellar motor torque and speed output by interacting with flagellar stator proteins. Collectively, the bioinformatical and experimental work described in this thesis bring us one step closer towards understanding the functional diversity of the entire PilZ protein superfamily and elucidating the functions of PilZ proteins in the model organism *P. aeruginosa* PAO1.

Chapter 1 – Introduction

The role of cyclic nucleotides as second messengers in cellular signal transduction is a commonality in all three branches of Life and is likely a trait from our common ancestor. Amongst them, the cyclic di-nucleotide bis-(3'-5')-cyclic dimeric guanosine monophosphate (cyclic di-GMP) is one of the most important prokaryotic second messenger. Cyclic di-GMP has been demonstrated to control various biological processes such as motility and biofilm formation, cell cycle, differentiation, virulence and production of natural products. It has thus been regarded as the master regulator in many aspects of the prokaryotic life cycle. Equally important are the receptors which bind this molecule and effect changes within the bacterial cell. This chapter will begin with a background on cyclic di-GMP and its associated metabolism in bacteria and then focus on protein receptors that contain the cyclic di-GMP-binding PilZ domain and cover the various physiological functions modulated by these proteins.

1.1 Cellular cyclic di-GMP concentration is controlled by diguanylate cyclases (DGCs) and phosphodiesterases (PDEs)

Cyclic di-GMP was first discovered in 1987 as a molecule which is required for activation of cellulose production in the bacterium *Gluconacetobacter xylinus* (*G. xylinus*)¹. Building on their pioneering work, Benziman et al. subsequently identified the same system in *Agrobacterium tumefaciens*² thereby demonstrating that this new cyclic dinucleotide was not unique to *G. xylinus*. They were also the first to report on the diguanylate cyclase (DGC) and phosphodiesterase (PDE) enzymes responsible for the metabolism of cyclic di-GMP in *G. xylinus* by identifying three operons which contain genes encoding DGCs and PDEs³. They observed that the DGCs possessed conserved Gly-Gly-Asp-Glu-Phe motifs while the PDEs possessed Glu-Ala-Leu motifs and the authors subsequently named these domains after their signature motifs (GGDEF⁴ and EAL³). At that point in time, the functional implications of those domains were unknown. Today, we know that these domains are important for the catalytic activity of these enzymes and have identified three domain families⁴

involved with cyclic di-GMP metabolism in Prokaryotes – EAL and HD-GYP break while GGDEF make cyclic di-GMP (Figure 1A).

Figure 1. Cellular functions regulated by cyclic di-GMP (A) Cyclic di-GMP is synthesized from GTP by DGCs with a conserved GGDEF motif and it is degraded by PDEs with conserved EAL or HD-GYP motifs. Within the cell it has the overall effect of suppressing motility and virulence while enhancing biofilm formation. It also plays a role in cell cycle progression and differentiation^{5,6}. (B) Conformations of cyclic di-GMP molecules in protein crystal structures.

1.1.1 Diguanylate cyclases

A DGC homodimer condenses two molecules of Guanosine Triphosphate (GTP) into a single cyclic di-GMP molecule. When existing as a dimer, each DGC is bound to a GTP molecule and is oriented anti-parallel relative to each other; their catalytic GGDEF domains then cooperate to synthesize their cyclic di-GMP product⁷. Specifically, within the active site, the 3'-OH group of each GTP substrate is deprotonated and performs a nucleophilic substitution on the α -phosphate of the other GTP, forming a 3'-5' phosphodiester bond with the release of a pyrophosphate^{7,8}. The GGDEF residues play important roles in this process at a DGC's active site, which explains their conservation throughout Evolution. For example, the crystal structure of the DGC PleD from *Caulobacter crescentus* (*C. crescentus*) shows that the glycine residues

(G368 and G369) are required to accommodate the β and γ phosphates of a GTP molecule⁷. In addition, the glutamates (GGEEF; PleD has a GGEEF motif) contribute the negative charge necessary to coordinate Mg^{2+} ions which binds the GTP phosphates and are also required for catalysis (Figure 2B)

Figure 2. Structure of the diguanylate cyclase PleD (A) Crystal structure of the PleD enzyme (PDB: 2V0N) in complex with GTP at its active site and intercalated cyclic di-GMP at its I-site. (B) Half active site of PleD with magnesium ions shown as cyan spheres. Residues important for coordinating the magnesium ions or GTP is shown in stick. (C) Cartoon representation of the regulation of DgcZ activity by zinc ions.

The activity of DGCs are strictly controlled within the cell and one mode of control is by facilitating or disrupting the formation of the functional homodimer. For instance, non-competitive inhibition by cyclic di-GMP occurs when a pair of base-intercalated cyclic di-GMP binds to the inhibitory site (I-site) of a DGC and its conformation is disrupted into a non-functional one^{7,9} (Figure 1B, 2A). This mechanism is present in approximately half of all GGDEF domain-based domain DGCs which possess an I-site containing an RxxD motif that binds the intercalated cyclic di-GMP molecules^{10,11}. Product inhibition is also present in DGCs without an I-site, Chao *et al.* reported on the competitive inhibition of XCC4471 from *Xanthomonas campestris* (*X. campestris*)

by a partially intercalated cyclic di-GMP dimer (Figure 1B) which binds to the active site of XCC4471 and inhibits its activity¹².

Another mechanism of DGC regulation involves the sensory domains to which many GGDEF domains are linked. The GGDEF domain in PleD is linked to a receiver domain and phosphorylation of this receiver domain triggers the dimerization of the monomers into a functional DGC⁸. The DgcZ (YdeH) from *Escherichia coli* (*E. coli*) exists as a dimer *in-vivo* and consists of a GGDEF catalytic domain and a zinc-binding CZB domain. Although dimerized, the GGDEF domains are held in a non-functional conformation as long as zinc ions are bound to the CZB domain¹³ (Figure 2C). In this manner, cyclic di-GMP production can be tied to the concentrations of zinc within *E. coli*. By linking the catalytic GGDEF domain to receiver domains, the biosynthesis of cyclic di-GMP is tightly regulated in a tempo-spatial manner in response to signals including oxygen¹⁴, L-arginine¹⁵ and even physical surfaces¹⁶.

1.1.2 Cyclic di-GMP Phosphodiesterases

While the GGDEF remains the only catalytic family for synthesizing cyclic di-GMP in Prokaryotes, two distinct families of PDEs were found to degrade cyclic di-GMP in bacteria. Originally discovered by Benizman and colleagues³, biochemical evidence demonstrating that the EAL domain specifically hydrolyzes cyclic di-GMP was provided by Schmidt et al in 2005¹⁷. The EAL-domain based PDEs are the more numerous of the two PDE families, accounting for 6946 proteins encoded in all bacterial genomes where they are frequently found as part of GGDEF-EAL duet or linked to other protein domains⁴. It should also be noted that some EAL-domains based PDEs are catalytically inactive and retain only their binding affinity for cyclic di-GMP, thereby serving as receptors not PDEs.

The EAL domain-based PDEs require Mg^{2+} or Mn^{2+} for their activity and are highly specific, possessing low, sub micro-molar K_m ¹⁸⁻²⁰ which corresponds to cellular concentrations of cyclic di-GMP^{21,22}. They catalyze an asymmetrical hydrolysis of the ester bond in cyclic di-GMP and open the cyclic di-GMP macrocycle to produce a 5'-phosphoguanylyl-(3'-5')- guanosine (pGpG) molecule. Although monomeric EAL domains are able to degrade cyclic d-GMP¹⁷, most EAL PDEs exist as dimers or

oligomers *in vitro*^{20,23-25}. Structural studies on EAL PDEs have revealed their catalytic mechanisms in detail. In BlrP1 from *Klebsiella pneumonia* (*K. pneumonia*), the two antiparallel EAL domains interact through three α -helices and bind its substrate using two metal cations (Figure 3A). Within the EAL domain, cyclic di-GMP adopts its open conformation (Figure 1B) which is believed to facilitate the hydrolysis of the phosphodiester linkage^{25,26}. The activity of BlrP1 depends critically on a two-cation cluster in which the metal cations coordinate a water molecule each and one of the water molecule is involved in the hydrolysis of the phosphodiester bond (Figure 3B). The importance of the cation cluster is demonstrated clearly in BlrP1, if the Mn^{2+} ions and higher pH required for optimal hydrolysis is replaced with Mg^{2+} and low pH, the cluster is distorted from its optimum conformation^{25,26}. Another divalent cation, Ca^{2+} , inhibits the activity of EAL PDEs altogether by distorting distances within the cluster. In EAL domain-based PDEs, the Glu residue of the EAL motif contributes to the coordination of the metal cations, which explains its strict conservation in EAL domain-based PDEs^{17-20,24-26}.

In addition to EAL PDEs, another family of cyclic di-GMP-specific PDE can be found, this family of proteins contains a conserved HD-GYP motif. 1974 genes encoding these proteins can be found in all bacteria genomes studied⁴. Analogously to EAL proteins, not all HD-GYP domain-containing proteins are functional enzymes, some serve as cyclic di-GMP receptors. The HD-GYP domain-based PDEs are part of the large HD family of hydrolytic enzymes and HD-GYP domains were initially predicted to have cyclic di-GMP-specific activity from genetic evidence due to linkage with GGDEF domains akin to the frequently seen GGDEF-EAL proteins^{27,28}. In 2006, using the HD-GYP domain of the protein RpfG from *X. campetris*, Ryan et al. demonstrated experimentally that this domain is able to specifically degrade cyclic di-GMP²⁹ into GMP.

The initial detection of GMP as a product of the hydrolysis was the first indication that the HD-GYP domain-based PDEs use a different catalytic mechanism from EAL domain-based PDEs. The only structure of a functional HD-GYP PDE was solved recently and revealed a novel trinuclear iron binding site in its catalytic core³⁰. Within the active site, cyclic di-GMP takes up a V-shaped conformation in contrast to its

extended conformation in the EAL domain-based PDEs. Some of the conserved amino acids serve to bind the Fe^{2+} co-factors required for function: H221 and Y285 (HD-GYP) are in close proximity with the two peripheral Fe^{2+} ions while D222 (HD-GYP) contributes toward ligand binding (Figure 3C and 3D). Interestingly, a P286A (HD-GYP) mutation did not prove too detrimental to the activity of the enzyme³⁰.

Figure 3. Structures and active sites of EAL domain and HD-GYP domain based phosphodiesterases. (A) Crystal structure of the EAL PDE BlrP1 (PDB: 3GG0). The light-sensing flavin-mononucleotide (FMN) cofactor of the BLUF-domain is indicated. (B) Active site of BlrP1. Cyclic di-GMP and the residues needed to coordinate the manganese ions are shown in stick. Manganese ions are shown as blue spheres. (C) Crystal structure of the HD-GYP domain-based PDE from *Persephonella marina*. (PDB: 4M34) (D) Active site of the same HD-GYP PDE with its unique iron center. The residues which contribute charges to coordinate the three iron ions are shown in stick. The iron ions are shown as red spheres.

1.2 PilZ family of cyclic di-GMP receptors

In bacterial cells, DGCs and PDEs keep tight control on the concentration of cyclic di-GMP, cyclic di-GMP-binding receptors then sense changes in its concentration and interact with the cellular machinery to bring about physiological changes in the cell. A large number of different protein families that bind cyclic di-GMP have since been identified and in addition to protein receptors, cyclic di-GMP also binds to RNA riboswitches⁴. The PilZ-domain proteins were the first dedicated family of cyclic di-GMP receptors to be discovered and appear to have the highest affinities for cyclic di-GMP compared to other receptors. Named after the PilZ protein (PA2960) in *Pseudomonas aeruginosa* PAO1 (*P. aeruginosa* PAO1), the PilZ-domain is one of the most widespread family of cyclic di-GMP binding receptor in bacteria³¹ and the domain can be found as a standalone single-domain protein, linked to other domains as a di-domain protein or as part of a multi-domain protein. (The term “PilZ proteins” will be used to refer to PilZ-domain containing proteins from here on.)

Bioinformatic and structural studies of PilZ proteins have uncovered the amino acids within this domain that are necessary for cyclic di-GMP binding, two motifs, RxxxR and (D/N)xSxxG that are separated by 20-30 amino acids form the primary cyclic di-GMP-binding structure⁴. The solved structure of PilZ proteins indicate that it forms an eight-stranded antiparallel β -barrel fold and the (D/N)xSxxG motif is located on the surface of this barrel where it binds to cyclic di-GMP. The RxxxR motif forms a loop near the cyclic di-GMP molecule where the side chains of the two arginine residues form hydrogen-bonds with cyclic di-GMP^{32,33} (Figure 5). Typically, cyclic di-GMP acts a glue that brings two domains or proteins closer together, causing the conformation change that is required to initiate downstream signaling. Interestingly, different PilZ proteins can bind to different forms of cyclic di-GMP, MapZ (PA4608), for example, binds to intercalated cyclic di-GMP while VCA0042 binds to a cyclic di-GMP monomer³⁴. Particularly surprising is the flexibility with which PilZ proteins can switch from binding one form of cyclic di-GMP to another with minimal sequence and conformation changes³⁵. Finally, there are also PilZ proteins which are unable to bind cyclic di-GMP and PA2960, the namesake protein for the PilZ family is ironically one of them.

Figure 4. Cyclic di-GMP binding proteins in *Pseudomonas aeruginosa* PAO1

1.2.1 Single-domain PilZ proteins

The single domain PilZ proteins are the largest group of PilZ proteins and regulate a range of physiological process ranging from motility, biofilm formation, virulence and Type IV Pili (T4P) biogenesis³⁶⁻³⁹. These PilZ proteins typically comprise of 80-160 amino acids and mediate their control of cellular processes by binding to another protein upon cyclic di-GMP binding. It can be argued that the single-domain PilZ proteins are the hardest to study as there are no clues to their function compared to multi-domain PilZ proteins whose other domains may provide clues to their functions. At the time of writing, the most extensively characterized species with regards to the function of their PilZ repertoire is the model organism *P. aeruginosa* PAO1. Eight PilZ proteins are encoded in its genome, of which five are single-domain PilZ proteins. Of these five proteins, the functions of PilZ (PA2960), HapZ (PA2799) and MapZ (PA4608) have been elucidated. The remaining two, PA0012 and PA4324, are the subject of investigation in this thesis.

Figure 5. Structures of single and di-domain PilZ proteins. (A) Structure of PA4608 in complex with cyclic di-GMP. (PDB: 2L74) (B) Crystal structure of PilZD in complex with cyclic di-GMP. (PDB: 2RDE) The RxxxR loop is colored blue and the DxSxxG motif on the barrel is colored magenta in both images.

Together with the degenerate GGDEF-EAL protein FimX, PilZ is required for T4P biogenesis^{40,41} even though the two proteins do not interact physically⁴². Homologs of the same proteins in *Xanthomonas campestris* *pv. campestris* (*Xcc*) and *Xanthomonas axonopodis* *pv citri* (*Xac*) have also been shown to be essential for T4P biogenesis but structural studies have also shown that PilZ and FimX from *Xcc* and *Xac* physically interact to form a complex that binds cyclic di-GMP^{39,43}. The PilZ_{*Xac*}-FimX_{*Xac*} complex was also found to bind the ATPase PilB in *Xac*⁴⁴ but this was not investigated in *Xcc*. Given the close phylogeny between the two species, it is likely that PilZ_{*Xcc*}-FimX_{*Xcc*} also binds to PilB_{*Xcc*}. This structural divergence between the PilZ proteins in *Pseudomonas* and *Xanthomonas* species without a corresponding divergence in their functions is a curious phenomenon that has not been observed elsewhere so far.

HapZ has recently been shown to directly bind to the hybrid histidine kinase SagS in a manner that is enhanced by cyclic di-GMP. Through this interaction, HapZ blocks the phosphotransfer from SagS to its downstream phosphoreceptor HptB³⁷. SagS is a

two-component signaling protein that regulates the transition from a planktonic to a surface-attached biofilm lifestyle, under planktonic conditions, SagS interferes with GacS indirectly through HtpB; SagS also interferes directly with BfiS, suppressing the activity of both proteins. By blocking the phosphotransfer between SagS and HtpB, HapZ alleviates the suppression of GacS and leads to a decrease in sRNA (rsmY, rsmZ) levels⁴⁵. Additionally, by preventing the phosphorelay between SagS and HtpB, HapZ may also push Sags towards transferring its phosphate group to BfiS, leading to the activation to biofilm-related genes.

Like HapZ, the PilZ protein MapZ also functions as an inhibitor of a signaling pathway, it binds to the methyltransferase CheR1 and prevents it from methylating the chemoreceptor PctA. Methylation of PctA increases its affinity to chemoattractants and promotes chemotaxis of the cell and by preventing this process from taking place, MapZ suppresses chemotaxis of PAO1 in response to cyclic di-GMP³⁶. The most significant effect of MapZ was observed in the frequency of direction switching of flagella rotation, when overexpressed, MapZ promotes the unidirectional swimming of PAO1 cells, leading to increased surface attachment and biofilm formation²⁶.

The PilZ proteins from *C. crescentus*, DgrA and DgrB, also affects motility when overexpressed. Christen et al reported that both proteins were able to inhibit the motility of *C. crescentus* on semi-solid agar plates when overexpressed. In addition, when both proteins were deleted, artificially raising cyclic di-GMP levels in *C. crescentus* could not inhibit motility of the bacteria, motility was also not affected in knock-out mutants when cyclic di-GMP levels were low. This lead Christen et al to conclude that DgrA and DgrB mediate motility suppression in *C. crescentus* under elevated cyclic di-GMP levels. Their efforts also linked the mode of inhibition by DgrA to the protein FliL, which is required for flagella rotation. The cellular concentration of FliL was reduced when DgrA was overexpressed and overexpressing FliL alleviates the suppression by DgrA⁴⁶.

In the genome of the pathogen *Vibrio Cholerae* (*V. cholerae*), five PilZ encoding genes can be found, of which PlzA (VC0697) is a single-domain PilZ protein. Pratt et al concluded that PlzA is an essential gene as they were unable to knock out the gene.

They were also unable to detect cyclic di-GMP binding by PlzA and overexpression of the gene did not cause any discernible effects on the motility of *V. cholerae* on semi-solid agar plates⁴⁷. The functions of PlzA in *V. cholerae* remain an enigma and warrants investigation as this bacteria is the causative agent of cholera, a disease that can be life-threatening.

The genome of the rice pathogen *Xanthomonas oryzae* pv. *oryzae* (*Xoo*), contains three PilZ genes of which PXO_02715 and PXO_02374 are single-domain PilZ proteins. PXO_02715 is a homolog of PilZ and binds to FlpP, the FimX homolog in *Xoo*; these two proteins were found to regulate the Type III secretion system (T3SS) of the bacteria and deletions of these two genes lead to a decrease in virulence⁴⁸. This represents a shift from the well-reported paradigm where PilZ and FimX regulate T4P, unfortunately, Yang et al did not investigate if deletions of either gene had any effects on the bacteria's T4P⁴⁸. Thus, it is unknown if the two proteins also affect T4P biogenesis in addition to T3SS. PXO_02374 was reported to repress the T3SS system of *Xoo* by down-regulating the genes involved, deletion of PXO_02374 also affected sliding motility of the bacteria but not biofilm or exopolysaccharide production³⁸.

1.2.2 Di-domain PilZ proteins

Of the eight PilZ proteins encoded in the PAO1 genome, two are di-domain proteins, PA2989 and FlgZ (PA3353). The functions of PA2989 within the cell is current unknown and it is also the subject of investigation in this thesis. FlgZ, on the other hand, has been demonstrated to bind specifically to the MotC stator protein in PAO1 to regulate the swarming motility of the bacteria under high cyclic di-GMP levels⁴⁹. PAO1 possesses two sets of stator proteins, MotAB and MotCD and the authors found that the Mot AB stators are unable to drive swarming motility. By binding to MotC and displacing it from the flagella motor complex, FlgZ allows MotA to replace MotC in the complex and results in a flagellar complex that is unable to drive swarming motility.

The PilZ YpfA performs a similar function as FlgZ in the gram-positive species of *Bacillus subtilis* (*B. subtilis*), YpfA was demonstrated to inhibit swarming motility in *B. subtilis* by binding to the stator protein MotA⁵⁰. The di-domain PilZ YcgR also inhibits

swimming motility in *E. coli* by binding to motor proteins, it interacts with the flagellar switch complex proteins FliG and FliM and the stator MotA to disrupt torque generation by the motor complex and act as a “backstop brake” to suppress motility^{51,52}. Mutagenesis experiments also showed that YcgR binds to FliG using its N-terminal domain and to FliM using its C-terminal domain⁵¹, therefore it is likely cyclic di-GMP acts as the intramolecular “glue” that brings these two domains closer together, allowing YcgR to interact with FliG and FliM at the same time.

The majority of the PilZ proteins discovered so far mediate their activity by binding to other proteins in the cell but MrkH and PlzA (BB0733) are two exceptions. MrkH is a transcription factor in the opportunistic pathogen *K. pneumoniae*⁵³. MrkH consists of a PilZ domain and a LuxR-type transcriptional regulator domain and functions as a positive regulator of T4P biogenesis in *K. pneumoniae* by activating the genes needed. When bound to cyclic di-GMP, MrkH activates transcription of the *mrkABCD* gene cluster by binding to the *mrkA* promoter and RNA polymerase⁵⁴. Structural studies revealed that MrkH possesses PilZ-like folds in its LuxR-type domain which also binds to cyclic di-GMP, in addition, the positive residues in its C-terminal domain are responsible for interacting with DNA⁵⁵. PlzA is the only PilZ domain protein in the obligate parasite *Borrelia burgdorferi* (*B. burgdorferi*)⁵⁶ and has been shown to also be a positive regulator of transcription⁵⁷. PlzA also appears to be a multifunctional protein as it has been shown to affect swimming motility and virulence when deleted⁵⁸. Although the mechanism through which PlzA regulates motility is unknown, He et al have identified the mechanism through which PlzA regulates virulence: PlzA is able to positively regulate the transcription levels of BosR which then increases expression of the protein RpoS, the global regulator for virulence genes in *B. burgdorferi*⁵⁷.

The genome of the pathogen *V. cholerae* contains two di-domain PilZ proteins: PlzD (VCA0042) and PlzE (VCA0735). PlzE appears to be an essential gene in this bacteria but its function is currently unknown⁴⁷. Deletion of PlzD did not have an effect on *V. cholerae* motility but overexpression PlzD and PlzD^{R140A} had the same negative effect on motility on the bacteria; this suggests that PlzD regulates motility in *V. cholerae* in a cyclic di-GMP independent manner⁴⁷. The crystal structure of PlzD revealed that the

protein exists as a dimer *in vitro* and binds to monomeric cyclic di-GMP³³. PP4397 is a similar protein to PlzD from *Pseudomonas putida* which binds to intercalated cyclic di-GMP instead, and interestingly it exists as a dimer in the apo-form and upon binding to its ligand, it transits into a monomer³⁵. Importantly, the authors were able to switch the cyclic di-GMP binding modes of PlzD and PP4397 (i.e. PlzD binding dimeric cyclic di-GMP and PP4397 binding to monomeric cyclic di-GMP) simply by mutating a single amino acid situated immediately before the RxxxR motif³⁵. This experiment demonstrates the ease with which PilZ proteins with different cyclic di-GMP binding stoichiometry and structures can evolve to play diverse roles in the cyclic di-GMP network within the cell.

PXO_00049 is the only di-domain PilZ protein in the genome of *Xoo* and appears to bind cyclic di-GMP in a 1:1 stoichiometry. It appears to have similar functionality as PXO_02374 as it was reported that deletion of PXO_00049 resulted in increased expression of T3SS genes and a reduction in sliding motility of the bacteria³⁸.

1.2.3 Multi-domain PilZ proteins

Alg44 is the only multi-domain PilZ protein in PAO1 and is comprised of a PilZ domain, a transmembrane domain and a membrane fusion protein (MFP) domain; it is required for the biosynthesis of alginate by the bacteria⁴⁴. The work of Oglesby et al demonstrated that Alg44 acts as a co-polymerase with Alg8 for the polymerization of alginate, importantly they also showed that while cyclic di-GMP is necessary for Alg44's activity, it was the MFP domain that was critical for alginate production⁵⁹. Subsequently the solved structure of Alg44 showed that it binds to intercalated cyclic di-GMP and forms a homodimer after binding, however unlike other PilZ proteins, this dimerization is not mediated by cyclic di-GMP but rather the β -barrel of the PilZ domain and the helix which contributes hydrophobic residues for the dimerization interface⁶⁰.

Like Alg44, BcsA is a PilZ protein responsible for the production of an extracellular polysaccharide, which in this case is cellulose¹. Cellulose is often produced during biofilm formation and thus it is not surprising that a PilZ protein would be responsible for mediating cyclic di-GMP's control over this process. The PilZ domain of BcsA is

situated within its C-terminal and structural studies have revealed that cyclic di-GMP binding to this domain induces a conformational change that allows the UDP-glucose subunits access to the catalytic site⁶¹. In addition to being a matrix component, cellulose also plays a role in accentuating the motility suppression of YcgR in *Salmonella*, Zorraquino et al found that cellulose accumulation around cells results in steric hindrance of flagella rotation⁶². In this manner, BcsA and YcgR act in concert to inhibit motility and push the cell towards biofilm formation.

The first PilZ to be characterized in the pathogen *Treponema denticola* is TDE0214 which was shown to have a high affinity for cyclic di-GMP and regulates bacteria motility and virulence. Deleting TDE0214 results in a less motile mutant which also had more reversing during swimming motility. The authors also noted that the deletion mutant exhibits reduced biofilm formation and host invasiveness in a mouse model⁶³. Interestingly, the multi-domain PilZ protein Tlp1 in *Azospirillum brasilense* which is a aerotaxis and chemotaxis receptor, mediates motility in the bacteria in a similar manner. Mutants of the bacteria that expressed a mutant Tlp1 which is unable to bind cyclic di-GMP had decreased swimming velocity and increased frequency of swim reversals⁶⁴. The functions of these two PilZ proteins appear to be the opposite of the cyclic di-GMP paradigm where high levels of cyclic di-GMP trigger a suppression of motility. The modes of action of these two proteins are reminiscent of that of MapZ, which also affects the frequency of swim reversals. The effect of MapZ is however completely opposite to that of Tlp1 and TDE0214; MapZ causes a net reduction in swim motility while these two proteins cause a net increase in the motility of the bacteria. It seems that in bacteria whose lifestyle are highly reliant on being motile, cyclic di-GMP signaling may also be co-opted to increase motility upon stimulus.

1.3 Cyclic di-GMP signaling in the model organism *P. aeruginosa* PAO1

PAO1 is a ubiquitous Gram-negative bacterium that can cause infection as an opportunistic pathogen. It has been used extensively as a model organism to study the microbiology of gram-negative bacteria and biofilm formation. Genome sequencing have revealed that the genome of PAO1 encodes one of the highest number of cyclic di-GMP-related proteins: 17 GGDEF, 5 EAL, 16 GGDEF-EAL, 3 HD-GYP,

8 PilZ, 2 MshEN and 1 FleQ protein³¹. This abundance of cyclic di-GMP-related proteins demonstrates the number of cellular functions that are regulated by cyclic di-GMP in this opportunistic pathogen. In PAO1, like in several other species, low cyclic di-GMP concentration in the cell is correlated to motility while high cyclic di-GMP concentration in the cell is correlated to sessility and biofilm. Many of the processes regulated by cyclic di-GMP is therefore related to these stages in the PAO1 life cycle and the transition between them.

Although it is to be expected that functional DGCs increase cyclic di-GMP concentration in the cell and trigger the transition to biofilm, some DGCs have been linked to biofilm dispersal instead. Unlike the DGCs WspR, SadC, RoeA and SiaD that have been reported to specifically control the transition to biofilm-associated growth, GcbA and NicD are functional DGCs that contribute towards biofilm dispersal instead⁶⁵⁻⁶⁹. Interestingly, the signaling mechanisms of GcbA and NicD converge at the chemosensory protein BdlA, which regulates the activity of the PDEs DipA and RbdA. In response to an extracellular signal such as heavy metals (or nutrients in the case of NicD), the DGC is activated and produce cyclic di-GMP. This transient increase in cyclic di-GMP triggers the proteolysis and activation of BdlA which in turn activates DipA and RbdA. The two PDEs then reduce cellular cyclic di-GMP concentration leading to biofilm dispersal^{68,69}. MucR and NbdA are also two PDEs that trigger biofilm dispersal in PAO1 upon exposure to nitric oxide. Both proteins are membrane-associated and consists of MHYT-GGDEF-EAL domains, however NbdA appears to possess only PDE activity while MucR has been shown to display both DGC and PDE activity in vitro⁷⁰.

In addition to the PilZ proteins discussed earlier, many other types of receptors that detect the cellular concentration of cyclic di-GMP have been reported. Most of these proteins fit broadly into the scheme of cyclic di-GMP signalling in PAO1 and serves to suppress motility and promote biofilm formation. The receptor PelD contains a degenerated GGDEF domain with an I-site and cyclic di-GMP binding to its I-site is required for Pel polysaccharide production during biofilm formation⁷¹. The protein FimX contains GGDEF and EAL domains and like PelD, its GGDEF domain is degenerated, but its PDE activity appears to be active⁴¹. As mentioned previously,

FimX and PilZ is required for T4P biogenesis when cyclic di-GMP concentration is low. Recently, FimA has been discovered as a novel cyclic di-GMP receptor crucial for T4P-based surface attachment of PAO1 during high cyclic di-GMP concentrations and biofilm formation⁷².

Transcriptional regulators have also been identified as cyclic di-GMP effectors. FleQ is a transcriptional regulator that is regulated by cyclic di-GMP in two different modes: at low cyclic di-GMP concentration, FleQ activates flagella biosynthesis genes and suppresses genes involved in EPS biosynthesis; high concentrations of cyclic di-GMP competitively inhibits the ATPase activity of FleQ and converts it into an activator of biofilm-related genes⁷³. BrlR is another transcription factor that responds to cyclic di-GMP, however unlike FleQ or most other receptors, it does not regulate motility or biofilm-related functions. BrlR is a member of the MerR family of multidrug transport activators and when bound to cyclic di-GMP, it activates expression of genes encoding multidrug efflux pumps and its own *brlR* gene⁷⁴. Thus BrlR contributes towards the high level of tolerance to antibiotics in PAO1 biofilms.

Cyclic di-GMP has recently been shown to bind ATPases and regulate their activity. Trampari et al discovered that cyclic di-GMP acts as a competitive inhibitor to the ATPases FliI, HrcN and ClpB2 and directly interfere with the activity of these enzymes⁷⁵. Cyclic di-GMP was also reported to bind the MshE-type ATPases with extremely high affinities (0.5 μ M to 14 nM) and activates their ATPase-activity^{76,77}. This discovery suggests a novel role for cyclic di-GMP in controlling many bacterial export pathways via the associated ATPases.

1.4 Objectives

We have made significant advances in our understanding of cyclic di-GMP signaling over the 25 years since its discovery. While an increasing number of receptors for this second messenger have been discovered, our current knowledge regarding the largest family of cyclic di-GMP receptors, the PilZ proteins, remain scarce. Only a small number of PilZ proteins have been characterized with known biological functions. Of the PilZ proteins with known bacterial phenotypes, such as altered motility, virulence and polysaccharide production, only very few have their molecular

mechanism studied and characterized. Given the large number of PilZ proteins encoded by bacterial genomes and their enormous functional diversity indicated by the low sequence homology, the physiological functions of the majority of PilZ proteins remain hidden today.

One objective of my research project is to have a large-scale and comprehensive survey of the PilZ proteins encoded by bacterial genomes. I will use sequence similarity networking to analyze the functional diversity of PilZ proteins to understand the evolutionary divergence of PilZ proteins (Chapter 2). Our lab has previously established the biological functions of the two single-domain PilZ proteins, HapZ and MapZ in the model organism *P. aeruginosa* PAO1^{36,37}. *P. aeruginosa* PAO1 has four more PilZ proteins with unknown functions. The four PilZ proteins include two single-domain PilZ proteins: PA0012 and PA4324; and two di-domain PilZ proteins: PA3353 and PA4324. The other objective of my research project is to continue our endeavor to unearth the biological function of the four functionally unknown PilZ proteins of PAO1 (Chapters 3 and 4).

Chapter 2 – Sequence and function divergence of PilZ domains revealed by sequence similarity networking (SSN) analysis

2.1 Introduction

Cyclic di-GMP is an intracellular messenger found in many environmental and pathogenic bacteria. High cellular cyclic di-GMP concentration is believed to promote the sessile lifestyle and formation of surface-associated biofilm. While the cellular concentration of cyclic di-GMP is modulated by diguanylate cyclases (DGC) and cyclic di-GMP phosphodiesterases (PDE), cyclic di-GMP exerts its biological effect by binding a wide variety of protein and riboswitch targets. PilZ domain proteins are the first discovered c-di-GMP binding proteins and arguably the most prevalent ones⁷⁸. The PilZ domain is characterized by an eight-strand β -barrel consisting of two antiparallel β -sheets and a long and flexible N-terminal loop. PilZ domains bind monomeric or dimeric cyclic di-GMP using the arginine-rich motif (RxxxR) from the N terminal loop and the (D/N)xSxxG motif from the β -barrel. Apart from the canonical PilZ domains that exhibit c-di-GMP-binding ability, several non-canonical PilZ domain proteins that lack c-di-GMP-binding capability have also been reported.

Studies suggest that PilZ proteins form a superfamily of proteins evolved to bind cyclic di-GMP with different affinities and potentially interact with different protein partners. As of today, there are 26,643 PilZ domain (IPR009875) containing proteins and 255 atypical PilZ domain (IPR031800) proteins listed in the Interpro database⁷⁹. Apart from the small number of PilZ proteins that have been characterized experimentally (See Chapter 1), the biological functions and mechanism of action for most PilZ domain proteins remain unknown. Even for the cyclic di-GMP signaling network of the model systems (e.g. *P. aeruginosa*, *V. cholerae* and *X. campestris pv. campestris*) that have been under intensive investigation, the physiological roles of most PilZ proteins remain to be unveiled. Given the large number of PilZ proteins and the great diversity in their sequence and function, it remains a challenge to predict the function of PilZ proteins based on sequence-structure-function relationship.

Today, newly discovered proteins are routinely annotated with predicted function based on the function of experimentally characterized homologous proteins⁸⁰. This

approach is based on the assumption that the sequence of a protein determines its structure and function and thus, proteins with high sequence similarity are likely to share the same or similar functions. Using this approach, the functions of a few experimentally characterized proteins can be extrapolated to a large number of homologs or orthologues. To this end, many computational tools have been developed to predict function by surveying the relationship between the sequence-function space and homology of proteins. But for the resultant information to be truly meaningful, it should be presented in a format that is accessible to not just computational but also experimental biologists. To present such data, dendrograms and phylogenetic trees are typically used (Figure 5). However, the construction of trees consisting of large sequence sets is computationally intensive and interpretation gets progressively difficult as the tree increases in size and complexity⁸¹.

Figure 6. Comparison of a phylogenetic tree and its corresponding SSN. (A) A simulated phylogenetic tree containing 10 protein sequences in 3 classes. (B) A SSN that may result from the tree in (A) and how its topology changes as the stringency of the threshold value is increased. As the stringency or threshold value is increased, only nodes with high sequence similarity remains connected. While the relative positions of the unconnected groups have no meaning, the lengths of edges are inversely related to the relative similarity of the connected nodes.

To address this problem, Atkinson et al developed the use of Sequence Similarity Networks (SSN) for the visualization of complex relationships and functional trends across protein sequences. Within the SSN, the relationship between proteins is measured by independent pairwise alignments between sequences and the entire context is visualized in the form of a network. In a SSN, protein(s) are represented by a node and is connected to other nodes (protein) by edges (lines) which represent the sequence similarity of the nodes⁸². Only nodes which share pairwise sequence alignments that are greater than a user-specified value are connected by edges, creating a network of protein relationships. Figure 6 depicts a simplified overview of a simulated phylogenetic tree and a SSN that may result from the same sequences.

Generating SSNs are less computationally demanding and very large amounts of sequence data can be displayed. In addition, SSNs can show multiple neighbors for a given sequence whose linkage may be lost in phylogenetic trees. The display of sequence relationship in a two-dimensional topology is also more visually intuitive and thereby confers a greater ease of use. In addition, the use of software like Cytoscape⁸³ allows the user explore a SSN interactively. Hyperlinks to additional information such as structural and biochemical databases also can help the user evaluate the relationship between proteins in a network. Although a SSN captures most of the information in a phylogenetic tree, they are not calculated based on Evolutionary models and thus cannot be used for inferring Evolutionary history⁸². Today, SSNs are most widely used to visualize sequence relatedness and segregate diverse superfamilies of proteins into putative isofunctional groups⁸². This has had the most impact in the exploration of enzyme families, where SSN used in sequence-function investigations has led to the discovery of new chemistry within enzyme families⁸⁴⁻⁸⁶. Studies using SSNs in conjunction with protein structure has also been used to probe substrate specificities of enzyme families^{87,88}.

In this chapter, a large-scale analysis of sequence, structure and functional relationship for single and di-domain PilZ proteins was undertaken. SSNs were constructed for single and di-domain PilZ proteins with characterized PilZ proteins mapped onto the networks. Sequence logos were also generated to unveil residues or motifs that are unique to a group of PilZ proteins. In combination with structural

information for several PilZ proteins, the network and sequence logo reveal a great diversity of sequence and potentially function of the PilZ proteins.

The intention for this work is two-fold. First, in conjugation with our biochemical and functional studies on the PilZ proteins of *P. aeruginosa* PAO1, we expect the SSN to uncover their functional homologs to probe their biological functions. Second, considering that only a small fraction of single and di-domain PilZ proteins have been experimentally characterized, we hope that the SSNs will provide an overarching view of PilZ proteins and generate a useful framework for formulating hypothesis about the function of PilZ proteins.

2.2 Materials and Methods

2.2.1 Construction of SSN

Interpro database (v62.0) and UniProt Knowledgebase (Uniprot, March 2017 version) were used for sequence mining. The SSN for the Interpro PilZ domain (IPR009875) was generated with Option C of the Enzyme Function Initiative (EFI) Enzyme Similarity Tool (EST), which accepts Interpro/Pfam numbers and user sequences (FASTA file) as input⁸¹. The entire IPR009875 category of proteins, along with the FASTA sequence of three characterized PilZ proteins (PlzA⁵⁶ (Uniprot ID: O51675; MrkH⁵³ (Uniprot ID: G3FT00) and xcc-b100_2234^{89*} (Uniprot ID: B0RT03)) which were not included under the IPR009875 family, was submitted for SSN construction. Sequences that contain 80-180 residues were submitted to generate the SSN for single domain PilZ proteins and sequences that contain 140-280 residues were submitted for di-domain PilZ proteins. Given the large number of nodes in the full network, 80% representative node networks that contain “metanodes” were used in the following analysis. In the representative node networks, sequences that share 80% or higher sequence similarity are merged into a single node (metanode).

All networks were visualized in Cytoscape (v3.5.0) using the “organic” layout whereby nodes are clustered more tightly together if they are more connected⁸³, more connection within clusters indicate higher similarity. In this layout, the relative edge lengths have been shown to be close to BLAST E-value distances⁸². The alignment scores chosen for visualization of the PilZ SSNs were determined by stepping through alignment score values in annotated networks until major clusters containing known PilZs of distinct function were separated from other unrelated clusters. The alignment score at which clusters were sufficiently distinct but not overly fractionated was chosen to avoid missing any potential distant relationships between proteins.

**xcc-b100_2243 was identified as xcc6012 in its original publication but the Uniprot entry has since been changed to xcc-b100_2234. For cross-referencing purposes, all gene names will be based on Uniprot entries.*

2.2.2 Generation of protein sequence logos

Protein sequence logos were generated using the Weblogo tool⁹⁰ by inputting a multiple sequence alignment file generated from all protein sequences within the given cluster. To retrieve the protein sequences, the cluster of interest was first created into a new network (Select cluster, File > New Network > From selected nodes, all edges). The nodes within this new network were then exported into a table which contains the Uniprot ID of proteins (in addition to other information) represented by this network. The Uniprot IDs were used to retrieve the sequences in question from Uniprot and then used for Multiple Sequence Alignment using Clustal Omega with default settings. Gaps in the alignment were deleted manually and all residues are colored by hydrophobicity.

2.2.3 Databases and Servers

The EFI-EST and Weblogo tools are available at: <http://efi.igb.illinois.edu/efi-est/stepa.php> and <http://weblogo.threeplusone.com/> respectively. The multiple sequence alignment server Clustal Omega can be found at: <http://www.ebi.ac.uk/Tools/msa/clustalo/>. Uniprot Knowledgebase can be accessed at: <http://www.uniprot.org/>. Structures of proteins were retrieved from the RCSB Protein Databank at <http://www.rcsb.org/pdb/home/home.do>.

2.3 Results and Discussion

2.3.1 Classification of PilZ proteins into single, di- and multi-domain proteins

Using the EFI-EST⁸¹ we collected 24,950 sequences of PilZ domain proteins sequences from the Interpro protein database. The sequence length profile of these proteins shows that the PilZ protein family consists of a large number of single-domain PilZ proteins that have length between 79 – 175 residues (Figure 6). The core of PilZ domains contains an eight-stranded β -barrel of ~72 residues and most of the single-domain PilZ proteins have an N-terminal loop (12-18 amino acids) that is essential for cyclic di-GMP binding. The variation of protein length in single-domain PilZ proteins is mainly due to the difference in the C-terminal sequences. The second largest group of PilZ proteins are di-domain PilZ proteins (180 – 280 residues) that possesses an additional C- or N-terminal domain. The YcgR_N domain is by far the most common domain identified to be the additional domain in di-domain PilZ proteins. In addition to single and di-domain PilZ proteins, the remaining PilZ proteins feature much longer protein sequences (>300 amino acids) and are predicted to be multi-domain proteins based on Interpro and Pfam domain analysis. The multi-domain PilZ proteins contain a variety of protein domains such as glycosyltransferase, GGDEF, EAL, methyl-accepting chemotaxis protein domains. The variation in protein length and domain composition is a testament to how the PilZ domain has been harnessed to regulate different cellular functions through the course of Evolution.

In this study, we focus on the analysis of single and di-domain PilZ proteins using the SSNs. In comparison to the single and di-domain PilZ proteins, the biological function of many multi-domain PilZ proteins can be annotated with confidence because of the high sequence homology with already characterized homologs. For example, the largest group of multi-domain PilZ proteins, which has an average length of 880 residues, contain a glycosyltransferase domain and share high similarity with the cellulose synthase subunit of *E. coli* and *G. xylinum*. These homologs are likely to be responsible for the production of cellulose and other types of extracellular polysaccharides.

Figure 7. Histogram of the sequence length of PilZ proteins. The initial plot generated by the EFI-EST was cropped to better display the area occupied by the majority of PilZ proteins, a full plot is available in the Appendix (Figure S25).

2.3.2 SSN of single domain PilZ proteins

The SSN for single-domain PilZ proteins was generated using an alignment score of 16 (~35% sequence similarity cutoff, E-value 10^{-16}). This score was chosen as it segregates functionally distinct known PilZ proteins into different clusters and retains known functional homologs in the same cluster. Hence, nodes in the final SSN shown in Figure 8 are potentially sorted into isofunctional clusters consisting of functional homologs. The entire network consists of 11,781 sequences represented by 5803 metanodes in clusters various sizes (Figure 8). Most clusters contain sequences from bacteria of different orders or genera, suggesting that the clustering is not due to species (See Appendix for list of genera in each cluster). In this section, I focus on the analysis of the relatively large clusters (Cluster 1-11) that contain more than 60 nodes (Figure 9). Clusters with less than 60 nodes will not be discussed, unless they contain an experimentally characterized PilZ protein. Some of the clusters appear compact

with short edges connecting tightly packed nodes; while others feature smaller sub-clusters with the nodes between sub-clusters connected by long edges (e.g. Clusters 1a and 1b). The lack of compactness indicates lower sequence homology shared between the sequences of the sub-clusters and may reflect a functional divergence.

For all the sub-clusters from the 11 large clusters, we generated sequence logos in an effort to identify conserved residues or motifs that are likely to be important for protein function. The PilZ domain is known to contain two essential motifs (RxxxR and (N/D)xSxxG) for cyclic di-GMP binding. The two motifs can be identified for all the sequences from the network except for the ones in clusters 8 and 10. As we discuss in detail below, PilZ proteins from the two clusters are likely to be unorthodox PilZ proteins that lack cyclic di-GMP binding ability.

2.3.2.1 Cluster 1

As the fourth largest cluster, cluster 1 contains 601 sequences and can be further segregated into clusters 1a and 1b (Figure 10A). The two motifs (RxxxR, (N/D)xSxxG) important for cyclic di-GMP binding can be found in sequences in cluster 1, indicating that the proteins are capable of binding cyclic di-GMP. In addition, observation of a lysine residue upstream of the RxxxR motif indicates that the proteins of cluster 1a are likely to bind dimeric cyclic di-GMP as the (R/K)RxxxR motif is known to favor the binding of dimeric cyclic di-GMP³⁴.

Cluster 1a contains 405 sequences from 14 genera, with the majority of them from the genus *Pseudomonas*. The sequences in the sub-cluster share high homology as reflected by the compactness of the cluster. The only protein from this cluster with known biological function is HapZ protein (PA2799). HapZ is a 99-residue long single domain PilZ protein from PAO1. HapZ interacts with the histidine kinase SagS at elevated cyclic di-GMP concentrations³⁷ to block the transfer of a phosphate group from SagS to its downstream proteins. Hence, HapZ functions as a cyclic di-GMP binding adaptor protein and regulate the SagS-associated two-component signaling.

We previously built a structural model of HapZ in complex with dimeric cyclic di-GMP and verified the importance of the residues from the two motifs with bacterial two-

Figure 8. The single-domain PilZ SSN. This network contains 11,781 sequences represented by 5,803 nodes organized into 555 clusters of various sizes and 1024 single nodes at ~35% sequence similarity. All nodes are colored according to the phylogenetic order of bacteria from which the sequences are originated. The two clusters that contain sequences of PilZ proteins predicted to not bind cyclic di-GMP-binding are indicated with the box. The legend shows the top 20 orders of bacteria which account for 93% of the total sequences in the network. The SSN is displayed with an E-value threshold of 10^{-16} .

hybrid assay⁹¹. In addition to the residues (R12, R16, D38 and S40) from the two motifs, two additional residues (R80 and G92) located in the putative cyclic di-GMP binding pocket were

also found to be essential for cyclic di-GMP binding (Figure 10B and C, Liang lab, unpublished data). Notably, the position of R80 is occupied by a tryptophan residue in two thirds of the proteins from cluster 1a (Figure 10C, residue 88 in the sequence logo). This tryptophan residue may be involved in cyclic di-GMP binding through hydrophobic interactions. The high sequence similarity shared by sequences in cluster 1a suggests that the proteins in this cluster potentially bind to SagS homologs. In addition to the conserved residues that are likely involved in cyclic di-GMP binding, the sequence logo also shows several hydrophobic residues (Val, Leu, Phe) that are unique to this cluster of proteins and potentially may be involved in the interaction with their protein partners.

Cluster 1b contains 196 sequences from 29 genera. Aside from the cyclic di-GMP binding residues (Figure S1 in Appendix), very few other residues are conserved, indicating greater sequence and possibly functional divergence than the proteins in cluster 1a.

2.3.2.2 Cluster 2

Cluster 2 is the fifth largest cluster that contains 404 sequences with many of them coming from bacteria in the *Rhizobales* order. All proteins from Cluster 2 likely binds dimeric cyclic di-GMP as the conserved motif is present in all three sequence logos (Figure S2) with a K/R before the RxxxR motif³⁴. A point of interest in Cluster 2 is the presence of PilZ proteins from *Streptomyces purpurogeneiscleroticus* in cluster 2a and 2c, sharing a node with bacteria from the *Methylobacterium* genus. To my knowledge, *S. purpurogeneiscleroticus* is the only *Streptomyces* currently known to harbor PilZ genes. The high sequence similarity between the proteins lends credence to a hypothesis of horizontal gene transfer⁹² between the two species of bacteria which are both soil-dwelling.

Figure 9. Clusters from the single domain SSN that contain more than 60 nodes. Clusters which contain less than 60 nodes but contain reported single-domain PilZ proteins are also shown here.

Figure 10. HapZ is located in single-domain PilZ cluster 1 (A) Cluster 1; nodes colored according to the phylogenetic order of the bacteria from which the sequences were originated. (B) A structural model of HapZ in complex with dimeric cyclic di-GMP. Residues that directly interact with cyclic di-GMP are shown in stick. (C) Sequence logo for cluster 1a. Residues in HapZ that are involved in cyclic di-GMP binding are indicated with red arrows. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.2.3 Cluster 3

Cluster 3 contains a total of 1,186 sequences and is the second largest cluster. The cluster is divided into two sub-clusters 3a and 3b (Figure 11A). Cluster 3a is a compact cluster with nodes clustered closely indicating higher similarity between the sequences. In contrast, the nodes in cluster 3b are more scattered which indicates lower sequence homology. Based on the conserved residues for cyclic di-GMP

binding in sequence logos for cluster 3a and 3b (Figure S3), all proteins from this cluster are predicted to bind cyclic di-GMP. Cluster 3a represents a group of PilZ proteins found in diverse bacteria species from 90 genera. Within this cluster are two studied PilZ proteins (PlzA (VC0697) from *V. cholerae* and MapZ (PA4608) from PAO1) that have been characterized. Binding assays demonstrated that MapZ binds cyclic di-GMP with a K_D of 6-8 μM ^{36,93}.

MapZ is one of the few PilZ proteins with their protein partners identified. It was found by our lab that MapZ interacts with the chemotaxis methyltransferase CheR1 at elevated cyclic di-GMP concentrations. In addition, co-crystallization studies on MapZ and CheR1 has identified that the two C-terminal helices of MapZ as key motifs in mediating MapZ-CheR1 interaction (Liang lab, to be published). The amino acid residues that are crucial in this interaction are labeled in Figure 11B and C. Importantly, the sequence logo shows that the C-terminal segment of MapZ is strongly conserved, including the four invariant residues involved in CheR1 interaction. The conservation of the functionally important residues strongly suggests that the proteins in this cluster have similar functions by binding to CheR1-like proteins and thus, confirms cluster 3a as an isofunctional cluster.

Although the phenotype associated with PlzA remains unknown, our results point towards a role of this protein in the regulation of chemotaxis and motility in *V. cholera*; such a hypothesis can be readily tested experimentally. In their report, Pratt et al were unable to knockout PlzA in *V. cholerae*, leading them to conclude that it is an essential gene. Overexpressing PlzA also had no discernable effects on motility⁴⁷ whereas overexpression of MapZ causes a suppression of flagellar motor switching and enhanced directional swimming in PAO1 cells³⁶. It should be noted that the motility assays described in the two studies were different and are not directly comparable.

Cluster 3b is a scattered conglomerate of 122 sequences from 17 genera. It should be noted that 50% of the sequences in this cluster were derived from metagenomic studies and are therefore from unknown bacteria species. The remaining known species in this cluster are sulfate reducing bacteria. Because of the low similarity

between the 122 sequences in cluster 3b and those in cluster 3a, the proteins in cluster 3b likely have a different function.

Figure 11. MapZ and PilZ are located in single-domain PilZ cluster 3. (A) Cluster 3 with the nodes colored according to the phylogenetic order of the bacteria from which the sequences were originated. (B) MapZ in complex with cyclic di-GMP (PDB: 2L74), the residues suspected to be important for interacting with CheR1 are highlighted. (C) Sequence logo 3a for protein sequences in Cluster 3a, the same residues labeled in (B) are indicated by red arrows. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.2.4 Cluster 4

Cluster 4 contains two highly scattered clusters of nodes (Figure 9) and all 97 sequences from this cluster are exclusively from the genera *Nitrospira*. The two cyclic di-GMP binding motifs are conserved in all the sequences (Figure S4); and the PilZ

proteins in 4b are predicted to bind monomeric cyclic di-GMP as hinted by the leucine residue upstream of the RxxxR motif³⁴. Other than the cyclic di-GMP-binding residues, very few other residues are conserved in this cluster.

2.3.2.5 Cluster 5

With a total of 815 sequences, cluster 5 consists of three sub-clusters of very different sizes (Figure 12). Cluster 5a is a compact cluster representing 184 sequences from 17 genera; whereas Cluster 5b is comprised of 584 sequences from 24 genera and cluster 5c contains 47 sequences from the genera *Pseudomonas* and *Simiduia*. PilZ proteins in cluster 5c are predicted to bind dimeric cyclic di-GMP.

Cluster 5 contains the proteins PXO_02374 from *Xoo* and PA4324 from PAO1. PXO_02374 codes for a cyclic d-GMP binding PilZ protein which appears to negatively regulate virulence of *Xoo* by downregulating T3SS genes³⁸. Given the high sequence similarity of proteins in cluster 5a, they may share similar functions with PXO_02374. PA4324 is a focus on my PhD study and will be discussed in detail in Chapter 3 of this thesis.

Figure 12. Single-domain PilZ cluster 5 containing PXO_02374 and PA4324. The nodes are colored according to the phylogenetic order of bacteria from which the PilZ sequences were originated.

2.3.2.6 Cluster 6

Cluster 6 is a relatively small cluster that includes 203 sequences from the *Bacillales* Order exclusively. This cluster can be further divided into cluster 6a that include 46 sequences exclusively from the genera *Paenibacillus* and cluster 6b that includes 157 sequences from 53 genera of *Bacilli*. Many bacteria from the genus *Paenibacillus* promote plant growth and are used as biofertilizer⁹⁴. Notably, PilZ proteins from this cluster seem to have diverged slightly from the canonical sequences by having an RxxxFR motif in contrast to the RxxxR motif that is required for cyclic di-GMP binding (Figure 13). As none of the PilZ proteins from this cluster has been experimentally characterized, it is currently unknown how this change affects cyclic di-GMP binding.

Figure 13. Sequence logo for single-domain PilZ cluster 6 from the order *Bacillales*. Sequence logos for (A) cluster 6a and (B) cluster 6b. The signature cyclic di-GMP binding motifs are indicated by the asterisk and line; the extra phenylalanine residue in both sequences is highlighted. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.2.7 Cluster 7

Cluster 7 consists of 180 sequences and all the PilZ proteins are predicted to bind dimeric cyclic di-GMP (Figure S7). This cluster is further divided into two sub-clusters. Cluster 7a consists of 144 sequences from 10 genera and cluster 7b consists of 36 sequences from the single genera of *Sphingomonas*. The sequences from the two sub-clusters share very low sequence similarity apart from the cyclic di-GMP binding motifs. None of the PilZ proteins from this cluster has been characterized experimentally.

2.3.2.8 Cluster 8

Cluster 8 is the largest cluster in this SSN, consisting of 1596 sequences from 175 genera. The relative compactness of this cluster is consistent with the high similarity shared by the sequences. Importantly, all the sequences from this cluster lack the RxxxR and (D/N)xSxxG motif for cyclic di-GMP binding and are predicted to be incapable of binding cyclic di-GMP (Figure 14B). The cluster contains four experimentally characterized proteins: PilZ (*PA2960*), XCC1024, XAC1133 and PXO_02715. XCC1024, XAC1133 and PXO_02715 share high sequence similarity and hence are in the same node in the cluster.

Structural studies have found that the two orthologues, XCC1024 and XAC1133, form complexes with FimX and cyclic di-GMP to regulate T4P biogenesis^{39,95}. PXO_02715 was found to upregulate T3SS genes by binding the FimX homolog Flip in *Xoo*. Similarity between PilZ proteins of this particular node suggests that PXO_02715 is likely to also regulate T4P biogenesis, however this was not investigated in the original study⁴⁸. Meanwhile, PilZ has been shown to not bind FimX_{PAO1} due to replacement of several key residues in FimX_{PAO1}⁴² required for PilZ binding. Despite this, PilZ and FimX_{PAO1} are still needed for T4P assembly at low cyclic di-GMP concentrations in PAO1⁴¹.

The residues that mediate this protein-protein interaction have remained largely invariant for all the proteins from this cluster (Figure 14B). In addition, the C-terminal

region that mediates the interaction between PilZ and PilB in *Xac* is also conserved⁴³. Given the high sequence homolog shared by the proteins in this cluster, we predict that the PilZ proteins from cluster 8 share the same function with the characterized proteins in regulating T4P biogenesis. However, the mechanism of regulation may differ by interacting with different protein partners.

Figure 14. (A) Close-up of the interaction site between XAC1133 and FimX_{xac} (PDB: 4FOU), XAC1133 is colored in light brown and FimX_{xac} is colored in grey. Residues reported to mediate PilZ-cyclic di-GMP interactions are presented in stick and labelled (B) Sequence logo for cluster 8. Residues reported to mediate PilZ-cyclic di-GMP^{39,95}. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.2.9 Cluster 10

This cluster encompasses 275 sequences from 26 genera. The sequence logo reveal that the PilZ proteins are lacking the RxxxR and (D/N)xSxxG motifs and are unlikely to bind cyclic di-GMP (Figure 15). This makes cluster 10 the second cluster of non-canonical PilZ proteins in the single domain PilZ SSN. Furthermore, sequences in

cluster share low similarity with those from cluster 8, indicating the two clusters of PilZ proteins have different biological functions and partners.

Figure 15. Sequence logo for single-domain PilZ cluster 10 that is predicted to not bind cyclic di-GMP. No sequences that bear resemblance to the signature PilZ cyclic di-GMP-binding motifs can be identified. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.2.10 Cluster 9, 11 and PA0012

Cluster 9 consists of 101 sequences and two sub-clusters. Cluster 9a comprises of 42 sequences from 19 genera; and cluster 9b comprises 59 sequences from 6 genera. About 65% of sequences in 9a and 50% in 9b were obtained in metagenomic studies of uncharacterized bacteria.

Cluster 11 contains 183 sequences from 12 genera. The PilZ proteins represented by this cluster likely bind cyclic di-GMP as suggested by the conserved RxxxR and (D/N)xSxxG motifs in the sequence logo (Figure S11).

Cluster PA0012 contains 153 sequences from 12 genera (Figure 8). The protein of interest in this case is the PilZ PA0012 from PAO1, which is a relatively small PilZ⁹³ protein that lacks the C-terminal region in MapZ which mediates interaction with CheR1. Little is known about the functions regulated by this PilZ protein and it is the subject of investigation in Chapter 3.

2.3.2.11 DgrA and DgrB clusters

The sequences from the two small clusters are exclusively from the *Caulobacter* genus. The DgrA cluster encompasses 12 PilZ proteins and the DgrB cluster contains seven sequences only. Hence, it appears that these PilZ proteins from *Caulobacter* form a distant class of single domain PilZ proteins apart from other single domain PilZs in the network. Both proteins were shown to suppress bacterial motility in the bacteria in response to increasing levels of cyclic di-GMP⁴⁶. Mutations to DgrA at the C-terminal of DgrA abolishes its ability to suppress motility⁴⁶, none of the residues map to a conserved residue on the sequence logo however.

2.3.3 SSN of di-domain PilZ proteins

The di-domain PilZ SSN is comprised of a total of 8,766 protein sequences. The sequences are segregated into 339 clusters of various sizes at the chosen alignment score of 19 (~37% sequence similarity cut-off, E-value of 10^{-19}). 805 sequences share low homology with other sequences and appeared as singletons in the network. The nine clusters that contain the most sequences are labeled in Figure 16 and will be discussed in detail in this section. Apart from sub-cluster 7b and cluster 10, all other clusters are predicted to bind cyclic di-GMP based on the two conserved binding motifs discussed above. The best-characterized di-domain PilZ protein today is the flagellar motoring interacting YcgR protein in *E. coli*. A large number of di-domain PilZ proteins are considered as YcgR homologs consisting of a N-terminal YcgR_N-like domain and a C-terminal PilZ domain. Apart from the YcgR_N domain, the other domains commonly found in the di-domain PilZ proteins include response receiver (REC) domain, DNA-binding domains.

2.3.3.1 Cluster 1

Cluster 1 consists of 119 sequences from 34 genera. The degree of conservation appears to be higher in the N-terminal domain which features a stretch of conserved

Figure 16. The di-domain PilZ SSN. This network contains 8,766 sequences represented by 4,138 nodes organized into 339 clusters of various sizes and 678 singletons at ~37% sequences similarity. All nodes are colored according to the phylogenetic order of bacteria from which the sequences are originated. The two clusters that contain sequences of PilZ proteins predicted to not bind cyclic di-GMP-binding are indicated with the box. The legend shows the top 20 orders of bacteria which account for 89% of the total sequences in the network. The SSN is displayed with an E-value threshold of 10^{-19} .

hydrophobic residues (amino acids 121-132, Figure S14) that may contribute to protein-protein interaction with its cellular partner. The strongest conservation in the C-terminal PilZ domain is present only in the cyclic di-GMP binding motifs.

Figure 17. Clusters from the di-domain SSN that contain more than 60 nodes. Clusters which contain less than 60 nodes but contain reported di-domain PilZ proteins are also shown here.

2.3.3.2 Cluster 2

Cluster 2a is comprised of 1,012 sequences and is the second largest cluster in the SSN. It is further sub-divided into cluster 2a and 2b. Cluster 2a is compact cluster consisting of 510 sequences from the Orders of *Bacillales* and *Lactobacillales*. Cluster 2a contains YpfA, the YcgR homolog in *B. subtilis*. YpfA suppresses motility in a cyclic di-GMP dependent manner by binding to the stator MotA in a mechanism analogous to YcgR in *E. coli*⁵⁰. Cluster 2b is a scattered cluster containing 502 sequences from 87 genera. Notably, the bacteria in this cluster consist of anaerobes from a variety of habitats.

2.3.3.3 Cluster 3

Cluster 3 is the fourth largest cluster in the SSN and comprises of 580 sequences and is further divided into four sub-clusters: 3a, 3b, 3c and 3d (Figure 18A). The PilZ domains in all sub-clusters presumably bind monomeric cyclic di-GMP as hinted from the signature motifs (Figure S16). Two experimentally characterized di-domain PilZ proteins are located in clusters 3b (PlzD VCA0042) and 3d (PlzE VCA0735).

Cluster 3a comprises of 68 sequences from sulfate-reducing bacteria whereas cluster 3b consists of 204 sequences from 7 genera, including PlzD from *V. cholerae*. Structural studies show that PlzD is a dimeric protein that undergoes a large conformational change upon binding to cyclic di-GMP. Cyclic di-GMP is bound at the interface between the two domains, with both domains contributing residues that coordinate the ligand³³ (Figure 18 B and C).

As mentioned earlier, a lysine or arginine residue before the signature RxxxR motif favors an interaction with dimeric cyclic di-GMP. In the case of PlzD, which binds monomeric cyclic di-GMP, a leucine is conserved before the canonical motif (Figure 19A). Remarkably when this leucine residue in PlzD was mutated to an arginine, the mutant protein binds intercalated cyclic di-GMP dimers with a change in the orientation of its domains (see Figure 19C). The authors then investigated if the reverse was true by mutating PP4397 [Cluster 5c] which is a di-domain PilZ protein

that binds dimeric cyclic di-GMP. When a reciprocal mutant PP4397 was analyzed, its binding stoichiometry to cyclic di-GMP was changed to 1:1 from 2:1, indicating that it bound monomers³⁵.

Figure 18. Di-domain PilZ cluster 3 and structure of PilZ (A) Cluster 3 with nodes colored according to the phylogenetic order of the bacteria where the PilZ proteins were originated. The top 6 Genera of bacteria are shown in the legend. The nodes containing the characterized PilZ proteins, PlzD and PlzE, are colored yellow. Space-filling structure of (B) apo-PilZD and (C) cyclic di-GMP-bound holo-PilZD; demonstrating the conformational change upon binding. The C-terminal PilZ domain is colored blue, the cyclic di-GMP binding loop is colored yellow, the N-terminal domain is colored green and cyclic di-GMP is colored red.

Cluster 3c contains 151 sequences from 17 genera and cluster 3d is comprised of 157 sequences from 8 genera, including PlzE which is also from *V. cholerae*. The original report on PlzE suggests that it is essential for the growth of *V. cholerae* and does not bind cyclic di-GMP⁴⁷. The sequence logo for this cluster (Figure S16D) however indicates that this protein would bind monomeric cyclic di-GMP like its neighbours in this cluster. This discordance can only be resolved with more experimental verification.

Figure 19. The residue prior to the RxxxR motif determines the cyclic di-GMP binding mode of the di-domain PilZ proteins (A) The RxxxR motifs in WT PilZD and the PP4397^{R122L} mutant which binds monomeric cyclic di-GMP and; (B) the same motif in PilZD^{L135R} mutant and PP4397 wild type which binds dimeric cyclic di-GMP. The mutated residue is highlighted and colored in red. (C) Structural alignment of holo-PilZD wt (green) and holo-PilZD^{L135R} (blue) which demonstrates the differences in their conformation after the mutation, cyclic di-GMP is not shown.

2.3.3.4 Cluster 4

Cluster 4 consists of 177 sequences from 12 genera. The sequences in this cluster have relatively low similarity as can be seen by the scattered nodes. Most nodes in this cluster represent only one sequence, signifying that sequences in this cluster do not have 80% sequence similarity with each other. The N-terminal domains in this cluster are recognized as a REC domain by both Pfam (PF00072) and Interpro (IPR001789). The low sequence similarity may reflect the variation in the REC domain for recognizing different upstream histidine kinases. It should also be noted that 74% of the sequences in this cluster are from metagenomic studies and thus are from uncharacterized bacteria.

2.3.3.5 Cluster 5

Cluster 5 is the largest cluster in the di-domain SSN accounting for 1,906 sequences in this network. The majority of the proteins from this cluster are predicted to be YcgR homologs containing an N-terminal YcgR_N domain. The cluster can be divided into three sub-clusters which contain a few currently best-characterized di-domain PilZ proteins.

Cluster 5a consists of 888 sequences from 41 genera and the proteins YcgR from *E. coli* in the main cluster and MrkH from *K. pneumonia* as an offshoot from the main cluster. This organization suggests possible distant linkage between these two proteins that have very different functions. YcgR interacts with the stator proteins in the flagella switch-complex acting as a “backstop brake” in *E. coli*^{51,52} to suppress swimming motility; whereas MrkH is a transcription factor that binds to DNA and directly activates transcription through interaction with RNA polymerase⁵⁴.

The work of Paul et al⁵¹ showed that YcgR uses residues on its C-domain to interact with FliM and its N-domain to interact with FliG (Figure 20). The sequence logo for cluster 5a indicates a high degree of conservation amongst the 888 sequences in this cluster including the residues involved in binding FliG/FliM. This indicates that the proteins in this cluster are likely functional homologs of YcgR and could play similar biological roles in regulating flagellar motor output.

The crystal structure of MrkH show that it does not contain an YcgR_N domain⁵⁵ (Figure 21B). We therefore generated a separate sequence logo for the 24 MrkH homologs found in the neighborhood of MrkH in the SSN (Figure 21C). The MrkH protein is composed of a C-terminal PilZ domain and an N-terminal domain with a PilZ-like fold. The MrkH N-terminal domain is able to bind cyclic di-GMP through the sequence of amino acids that form a β -turn- β structure that is similar to the signature DxSxxG motif of a true PilZ domain⁵⁵ (Figure 21B). This motif is not absolutely conserved in the 24 sequences neighboring MrkH and this may suggest that the other proteins may bind cyclic di-GMP using a different set of residues. Another point of conservation amongst these 24 sequences is at the C-terminal helix which was proposed to responsible for binding DNA⁵⁴ (Figure 21C). The DNA binding residues

are conserved in the sequence logo and this indicates that the proteins can potentially bind DNA.

Figure 20. Sequence logo for di-domain PilZ cluster 5a. Amino acids identified to mediate interaction with FliG are indicated with red arrows and those which mediate interaction with FliM indicated with yellow arrows. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Cluster 5b consists of 456 sequences from six genera of bacteria. Cluster 5b contains PXO_00049, which was found to regulate the virulence of *Xoo* by suppressing the expression of the T3SS genes³⁸. Cluster 5c is a compact cluster consisting of 562 sequences from 26 genera. The cluster contains two characterized proteins: FlgZ (PA3353) from PAO1 and PP4397 from *Pseudomonas putida*. The compactness of the cluster and the high conservation seem in its sequence logo indicates that the proteins in this cluster are likely true functional homologs of FlgZ. Unlike YcgR, FlgZ suppresses motility by binding to the stator protein MotC at high cyclic di-GMP levels⁴⁹. The other experimentally characterized PilZ protein in cluster 5c is PP4397 which exists as a monomer and forms a dimer upon binding an intercalated cyclic di-

GMP dimer³⁵. The clustering of PP4397 with FlgZ indicates that PP4397 probably also functions by binding to MotC homologs.

Figure 21. Structure of MrkH in complex with cyclic di-GMP and sequence logo for MrkH and its neighboring nodes. (A) MrkH (yellow hexagon) and neighbouring nodes. (B) Structure of MrkH (PDB: 5KGO), the novel cyclic di-GMP interaction motif is colored red while the canonical cyclic di-GMP-binding motif is in blue and presented in stick. (C) Sequence logo for the 24 sequences in the same grouping as MrkH. The novel cyclic di-GMP interaction motif in the N-terminal domain of MrkH is labeled and red arrows indicate other residues that are also involved in binding cyclic di-GMP⁵⁵. The DNA binding residues in the C-terminal are also labeled. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Cluster 6 consists of two sub-clusters. Cluster 6a is comprised of 691 sequences from 45 genera. PA2989, a PilZ protein which is the subject of study in chapter 4 also found in this cluster. An RxxxR motif can be found in positions 10-14 in the sequence logo, separated from the DxSxxG equivalent motif by about 100 amino acids. Current knowledge however dictates that the two cyclic di-GMP-binding motifs are separated by 20-30 residues and if residues 10-14 are the true cyclic di-GMP binding motif, then PA2989 may have a unique conformation to bridge the increased distance between the motifs. To this end, a structural prediction using the I-TASSER server⁹⁶ on full length PA2989 was performed and the result suggests that there is a possibility where the motif in position 10-14 may indeed bind cyclic di-GMP in an hitherto unreported fold in PilZ proteins. This is further explored in chapter 4.

Cluster 6b consists of 60 sequences from 13 genera, it should also be noted that 56% of the sequence in this sub-cluster were obtained in metagenomic studies and were from uncharacterized bacteria. The sequence logo does not show any recognizable cyclic di-GMP binding motifs and thus I am unable to ascertain the cyclic di-GMP affinity of the PilZ proteins in this sub-cluster (Figure S19B).

Figure 22. Sequence logo for di-domain PilZ cluster 6a. Putative cyclic di-binding motifs are indicated. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.3.7 Cluster 7

Cluster 7 consists of PilZ sequences in bacteria from the Order *Spingomonadales* exclusively and is divided into two sub-clusters. Cluster 7a contains 64 sequences from 5 genera of bacteria in this order and cluster 7b contains 72 sequences from 5 genera of bacteria as well. In the sequence logo of both sub-clusters (Figure 23 and Figure S20), the complete signature cyclic di-GMP binding motifs are in the N-terminal. Interestingly, the C-terminal domain is also predicted as PilZ and contains a shortened, incomplete RxxR motif (residues 134-137) and DxSxxG motif (residues 160-165, Figure 23). Observation of the motifs raises the possibility that the C-terminal region can also bind cyclic di-GMP or another structurally related nucleotide.

Figure 23. Sequence logo for di-domain PilZ cluster 7a which contains two sets of cyclic di-GMP binding motifs. The complete cyclic di-GMP-binding motifs are indicated in black and the putative incomplete motif is indicated in red. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.3.8 Cluster 8

Cluster 8 contains 579 sequences from 54 genera of bacteria in a compact cluster. The sequences in the cluster are quite well conserved (Figure S21) and only the PilZ domain have been recognized by Interpro and Pfam. This cluster also contains a di-domain PilZ sequence from *S. purpurogeneiscleroticus*, which was mentioned earlier (Single-domain SSN cluster 2) for being the only *Streptomyces* bacteria whose genome encodes PilZ proteins.

2.3.3.9 Cluster 9

Cluster 9 is also a highly compact cluster comprised of 91 sequences from 7 genera. A notable feature of this cluster is the absence of the RxxxR motif for cyclic di-GMP binding, suggesting that the cluster of proteins may not be able to bind cyclic di-GMP. Apart from the PilZ domain, the proteins from this cluster also contain a predicted DNA-binding domain (IPR001387 and IPR010982). Pfam recognized the C-terminal domain to contain a XRE family helix-turn-helix (HTH) DNA binding region (Figure 24).

Figure 24. Sequence logo for di-domain PilZ cluster 9 that contains an incomplete cyclic di-GMP binding motif and HTH region. The predicted cyclic di-GMP binding motif and the predicted XRE family HTH domain is indicated. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

2.3.3.10 PilZ (BB0733) cluster

The characterized PlzA protein and ten other PilZ proteins from the *Spirochaetes*, including several pathogenic strains from the *Treponema* genus, form a small cluster on their own. PlzA contains the two cyclic di-GMP binding motifs and regulates virulence gene expressions in response to cyclic di-GMP⁵⁷ while regulating mRNA levels independently of cyclic di-GMP.

2.3.3.11 xcc-b100_2234 cluster

This cluster only contains three sequences, including the structurally characterized PilZ protein Xcc-b100_2234. xcc-b100_2234 is an atypical PilZ domain protein that forms a tetramer and does not bind cyclic di-GMP. The protein's N-terminal domain contains a long α -helix that interacts with other monomers to form the four-helix bundle to drive the formation of the homotetramer (Figure 25). Apart from the observation that the protein contributes to Xcc pathogenicity, little is known about the physiological functions and mechanism of action for this protein.

Figure 25. Structure of xcc-b100_2234 homo-tetramer. (A) side view showing its “house-like” structure and (B) top view of the 4 monomers that interact via their α 3 helices.

2.4 Discussion

In this chapter, we performed a large-scale analysis of single and di-domain PilZ proteins using sequence similarity networking. One of the most important

observations is that the PilZ proteins exhibit extraordinary sequence divergence, hinting that PilZ proteins are likely to have very different biological functions and protein partners. The SSNs are most informative when nodes are segregated into isofunctional clusters. Although we do not expect that all the nodes to be grouped into isofunctional clusters due to the enormous evolutionary divergence of the sequences, many clusters from the two SSNs, especially the ones consisting of highly similar sequences are likely isofunctional clusters. This conclusion is supported by the high sequence similarity shared by sequences within individual clusters. This conclusion is further reinforced by the observation that functionally important residues are conserved for clusters containing structurally or biochemically characterized PilZ proteins such as MapZ, YcgR and PilZ.

Despite the fact that single-domain are the most abundant PilZ proteins, they are arguably the most divergent and poorly characterized ones. As of today, only a handful of single-domain PilZ proteins have been characterized biochemically or structurally. Among the 24 sub-clusters examined in my study, 17 contained no characterized proteins. In addition, 4,917 belong to clusters with less than 40 nodes, indicating that those PilZ proteins may have evolved for very specific functions in a narrow range of bacteria species.

Although it is known that the homologs of PilZ (PA2960) are not able to bind cyclic di-GMP, our study reveals that a surprisingly large number of the single-domain PilZ proteins from the two large clusters (clusters 8 and 10) are likely to function without binding cyclic di-GMP. Sequences for the two clusters account for ~16% of the total number of sequences in the single-domain SSN. As the largest cluster in the SSN, cluster 8 contains homologs of PilZ which are likely to mediate T4P assembly through protein-protein interactions. Meanwhile, none of the proteins from cluster 10 has been experimentally characterized. Considering that the PilZ proteins from cluster 10 possess a different set of conserved residues than the ones from cluster 10, they are likely to have different biological functions.

One of the motivations for us to generate the SSNs was to identify the homologs of PilZ proteins of PAO1 to assess their distribution in various bacterial species. The

homologs of the single-domain PilZ (PA2960), MapZ, HapZ, PA0012 and PA4324 can be found in a variety of bacterial species, indicating that the pathways involving these PilZ proteins are likely widespread. In comparison to single-domain PilZ proteins, di-domain PilZ proteins display less sequence divergence as reflected by fewer numbers of clusters and singletons in the SSN. Another salient feature of the di-domain PilZ proteins is that the YcgR-type proteins, which are defined as the di-domain PilZ proteins that contain an N-terminal YcgR_N domain, outnumber the other di-domain PilZ proteins. In the SSN, 5 out of the 9 large clusters belong to the YcgR-type proteins. It should be noted that the functional homologs of the *E. coli* YcgR protein are only found in cluster 5 and that the other YcgR-type proteins are likely to have different biological functions other than regulating flagellar motor activity.

In addition to the YcgR-type proteins, our SSN and sequence analysis revealed the presence of a large number of poorly understood di-domain PilZ proteins containing an uncharacterized protein domain. Cluster 6, represented by the PAO1 protein PA2989, consists of proteins that are predicted to contain a PilZ domain and a largely α -helical domain. As I will discuss in Chapter 4, the α -helical domain is likely to be incorporated into the PilZ domain.

In comparison to the well-known fact that some single-domain PilZ proteins do not bind cyclic di-GMP due to the lack of the RxxxR and DxSxxG motifs, the only known di-domain PilZ protein that lacks cyclic di-GMP binding ability is *Xcc-b100_2234*. Our SSN and sequence analysis revealed additional families of di-domain PilZ proteins that are likely to lack cyclic di-GMP binding ability. The proteins in cluster 6b and 9 do not contain the complete motif and are unlikely to bind cyclic di-GMP. In addition, cluster 7 contains an orthodox N-terminal PilZ domain and an unusual C-terminal PilZ domain with the cyclic di-GMP-binding RxxxR motif replaced by an RxxR motif. It remains to be seen whether the C-terminal PilZ domain can still bind cyclic di-GMP or a different nucleotide.

With the exception of cluster 2b from the di-domain SSN, we did not detect any unique distribution of bacteria from a specific niche in any of the other individual clusters. However, as mentioned earlier, the PilZ proteins have extensively diverged

over the course of Evolution and it would not be surprising if there are PilZ proteins which have evolved to regulated specialized biochemical pathways in bacteria adapted to various niches. It is likely that these “niche-adapted” PilZ proteins are clustered into small clusters which were not analyzed.

2.5 Conclusion

The work described in this chapter provides a panoramic perspective of the “PilZ landscape” and reveals several new families of PilZ proteins which deviate from the canonical PilZ archetype. It also highlights the functional divergence of PilZ proteins. Finally I hope the observations from this chapter prove useful to microbiologists for the discovery of new PilZ protein functions and pathways.

Chapter 3 – Probing the function of two single-domain PilZ proteins in *P. aeruginosa* PAO1

3.1 Introduction

P. aeruginosa is an opportunistic pathogen that is metabolically versatile and able to thrive under many conditions. This trait has allowed this species to colonize nearly all environments and as such it is considered a ubiquitous bacterium⁹⁷. Because of this, *P. aeruginosa* can colonize biotic and abiotic surfaces in the healthcare environment, leading to nosocomial infections in immunocompromised individuals⁹⁸. Compounding this situation its intrinsic resistance to many antibiotics due to the presence of a large number of efflux pumps⁹⁹. When forming biofilm, the cells are even more resistant to anti-microbial and disinfectant treatment and at times, even vigorous application of standard medical disinfection procedures can fail to eliminate the bacteria from medical devices^{100,101}.

Outbreaks of *P. aeruginosa* nosocomial infections occur regularly in hospitals¹⁰²⁻¹⁰⁴ and Singaporean hospitals are no exception. *P. aeruginosa* is the second-most prevalent species of resistant gram-negative bacteria¹⁰⁵ and the first case of New Delhi metallo- β -lactamase-1 (NDM-1) producing *P. aeruginosa* in Singapore was reported in 2015¹⁰⁶. In April 2016, there was an outbreak of *P. aeruginosa* conjunctivitis in the Neonatal Intensive Care unit of a Private Hospital in Singapore that was traced to the sink drains in the wards¹⁰⁷. Hospital water systems and associated fixtures have been identified as key areas where microbial biofilms develop, and thus prevention or disinfection of these reservoirs of microbial contamination is an important issue¹⁰⁸. When exposed to antimicrobials or disinfectants, many bacteria including *P. aeruginosa*, is induced to form biofilm¹⁰⁹⁻¹¹¹ that often exhibit increased resistance to treatment¹¹¹⁻¹¹³. Understanding the cellular pathways that govern biofilm formation and dispersal is therefore crucial to augment existing disinfection procedures and develop new strategies to treat biofilms-type infections.

Cyclic di-GMP-associated signaling proteins are being considered as potential targets to prevent biofilm formation or disperse biofilm of *P. aeruginosa*. Cyclic di-GMP has

been established as an important regulator in most bacteria species, controlling the motile and sessile transition in many bacteria. It is also involved in the regulation of the cell cycle, cell differentiation and other developmental processes⁴. The genome of *P. aeruginosa* codes for a large number of DGC and PDE proteins and cyclic di-GMP binding effectors, which included the PilZ proteins. In *P. aeruginosa* PAO1, there are seven cyclic di-GMP binding PilZ proteins and four of them (PA0012, HapZ, PA4324 and MapZ) are single domain PilZ proteins.

It was discovered by our lab recently that HapZ interacts with the histidine kinase SagS at high cyclic di-GMP levels to regulate the motile-sessile switch and biofilm growth in PAO1³⁷; and that MapZ binds to the chemotaxis methyltransferase CheR1 at high cyclic di-GMP levels to modulate chemotaxis and flagellar motor switching³⁶. The binding partners of HapZ and MapZ were identified using bacterial two-hybrid screening. However, our two-hybrid screening failed to identify the binding partners of PA0012 and PA4324. Although PA0012 and PA4324 have been shown to bind cyclic di-GMP⁹³, the physiological function of the two proteins remain unknown.

In this chapter, I present the data from my PhD work to probe the biological function of PA0012 and PA4324. I will first compare the phenotype of the PA0012 and PA4324 null mutants with that of the wild type PAO1. Then I will present transcriptome data obtained from RNA sequencing studies to show the two genes likely affect gene expression. Finally, I will discuss affinity pull-down experiments designed to capture the potential protein partners of these PilZ proteins.

3.2 Materials and Methods

3.2.1 Bacteria strains and plasmids

Strains and plasmids used in this thesis are listed in Table S25 and S26 respectively. Unless otherwise stated, bacteria were grown in lysogeny broth (LB) at 37°C and if necessary, the medium was solidified with 1.5% agar (BD Difco). All media were also sterilized by autoclaving. LB media consisted of 10 g/L Tryptone (BD Difco), 5 g/L yeast extract (BD Difco) and 10 g/L NaCl (Merck) in deionized water.

Where necessary, antibiotics were used at the following concentrations when culturing PAO1 strains: carbenicillin (300 µg/ml, MP Biomedical) and gentamicin (30 µg/ml, Life Technologies) and; when culturing *E. coli*: ampicillin (100 µg/ml, Sigma Aldrich), gentamicin (10 µg/ml) and apramycin (50 µg/ml, Sigma Aldrich).

3.2.2 Cloning of PilZ genes into overexpression plasmids

PA0012 and *PA4324* genes were cloned from PAO1 genomic DNA using Polymerase Chain Reaction (PCR) with Q5 polymerase (New England Biolabs (NEB)). The product was separated using electrophoresis on 1% agarose gel and the bands of the correct size were excised and purified using a Gel and PCR clean up kit (Macherey-Nagel). The restriction enzymes *SacI* and *BamHI* (NEB) were used cut the purified PCR fragments and the plasmid pUCP18. Following that, the fragments were ligated into pUCP18 using T4 ligase (NEB) and the mixture transformed into *E. coli* DH5α cells. Plasmids extracted from the transformed cells were verified by sequencing.

To clone *PA0012* and *PA4324* with haemagglutinin (HA) tag into the inducible plasmid pJN105, the in-frame coding sequence for the HA-tag was introduced upstream of the genes by means of the primers used in the PCR reaction. To ligate HA-*PA0012* into pJN105, *EcoRI* and *XbaI* (NEB) restriction enzymes were used to restriction both PCR fragments and plasmid; to clone HA-*PA4324*, *NheI* and *SacI* (NEB) enzymes were used. DNA constructs of *PA0012*^{R5A} and *PA4324*^{R9A} were designed to include the same restriction sites and synthesized by Integrated DNA Technologies for cloning into pJN105.

List of primers used can be found in Table S27 in the Appendix.

3.2.3 Swimming and swarming assay

Swimming motility was analyzed on M8 media which consists of 6 g/L Na_2HPO_4 , 3 g/L KH_2PO_4 and 0.5 g/L NaCl (Sigma Aldrich) dissolved in deionized water. The M8 media was supplemented with glucose to 0.2%, casamino acids to 0.3% and MgSO_4 (Sigma Aldrich) to 1 mM and solidified using 3 g of agar. Swarming motility plates are comprised of the same components but 5 g of agar was used to solidify the medium instead. Antibiotics were used where necessary. Strains to be assayed were grown overnight at 37 °C with shaking and used to inoculate swimming and swarming motility plates. For swarming motility, 1 μl of the overnight liquid culture was inoculated on to the surface of the agar plates and allowed to dry. For swimming motility, a 10 μl pipette tip was first dipped into the overnight culture and then stabbed halfway into the agar. Plates were then incubated for 18 hours at 37 °C in a humidified incubator.

3.2.4 Twitching motility assay

Twitching motility was analyzed on thin LB agar cast in tissue culture coated petri dishes (Fisher Scientific). Strains to be assayed were grown overnight at 37 °C with shaking and used to inoculate twitching motility plates. Plates were inoculated using 10 μl pipette tips by first dipping the tip into the overnight liquid culture and then stabbing the tip through the agar to touch the bottom of the petri dish. Inoculated plates were cultured in a humidified container at 37 °C for 24 hours. To visualize the interstitial colony, plates were flooded with cold developer solution (50% methanol and 10% acetic acid (Sigma Aldrich) in deionized water) and incubated at room temperature for 30 minutes. The developer solution and top layer agar was removed before images of the colonies were taken.

3.2.5 Colony morphology assay

Colony morphology agar consisted of 10 g/L of Tryptone in deionized water solidified with 1% agar. Congo red (40 mg/L) and Coomassie blue (20 mg/L) was added to the medium before pouring into agar plates. Strains to be assayed were grown overnight

in tryptone broth (10 g/L tryptone in deionized water) with shaking at 37 °C. The overnight cultures were diluted to an OD_{600nm} of 0.025 using tryptone broth before 10 µl was inoculated onto the agar plates and allowed to dry. Plates were wrapped with parafilm and incubated at room temperature in the dark for 5 days.

3.2.6 Broth dilution Minimum Inhibitory Concentration (MIC) assay

The MIC assay was modified from Wiegand et al¹¹⁴. Cation-adjusted Mueller-Hinton broth (MH broth, Sigma Aldrich) was used for this assay. Strains to be tested were streaked on MH agar and incubated at 37 °C overnight for single colonies. The following day, a single colony was inoculated into MH broth and incubated at 37 °C with shaking until OD_{600nm} of 0.6 (4.8×10^8 colony forming units (CFU)). The culture was then diluted to 1×10^6 CFU/ml and used within 30 minutes to prevent changes in the CFU concentration. A sample of the diluted culture was also diluted by a factor of 10^3 and plated on LB agar plates and incubated at 37 °C overnight for CFU count the following day.

To perform serial dilution of the antibiotic, 50 µl of MH broth was added to wells in columns 1-11 of a 96-well plate (Nunc). Following that, 50 µl of MH broth containing antibiotics at 1,024 µg/ml was added to the first row of wells (column 1) and mixed thoroughly. 50 µl of this mixture was then transferred to the neighboring well (column 2) and mixed thoroughly. This process was repeated for 10 columns in total (Columns 1-10). Column 11 was reserved as a growth control and no antibiotics were added to it. Column 12 was used as a sterility control and contains 100 µl of MH broth without inoculum or antibiotics. 50 µl of the diluted culture described above was then added to wells in columns 1-11 and mixed well. Plates were then covered and incubated in a humidified container at 37 °C for 18 hours.

The following day, OD_{600nm} of plates were measured using microplate reader (Tecan).

3.2.7 Transcriptome analysis by RNA- sequencing

Single colonies of PAO1 and PilZ deletion and complementation strains were inoculated into M9 medium (3.39 g/L Na₂HPO₄, 1.5 g/L KH₂PO₄, 0.5 g/L NH₄Cl and 0.25 g/L NaCl dissolved in deionized water) supplemented with MgSO₄ to 2 mM,

CaCl₂ to 0.1 mM, and glucose to 0.4% and grown overnight at 37 °C. The cultures were then diluted to an OD_{600nm} of 0.01 and cultured in 24-well plates at 37 °C for 7 hours and mixed immediately with RNAProtect Bacteria reagent (Qiagen). After incubation for 5 minutes at room temperature, samples were pelleted by centrifugation. To extract RNA, pellets were treated with lysozyme and total RNA was extracted with RNeasy Mini Purification kit (Qiagen). DNA was eliminated by on-column digestion using RNase-free DNase (Qiagen). Ribosomal rRNAs were removed using the Ribo-Zero Magnetic kit (Epicentre).

Gene expression analysis was performed in duplicates by RNA sequencing. The rRNA-depleted RNA was fragmented to 200-300 base pairs fragments and first and second strand cDNA synthesis was followed by end repair and adaptor ligation. The libraries were sequenced using the Illumina HiSeq2000 platform with paired-end protocol and read lengths of 100 nucleotides. The sequence reads were assembled and analyzed using RNA-seq and the expression analysis application of CLC genomics Workbench 6.0 (CLC Bio). The following criteria were used to filter the unique sequence reads: minimum length fraction of 0.9, minimum similarity fraction of 0.8, and maximum number of two mismatches. Data were normalized by calculating the reads per kilobase per million mapped reads for each gene. Analysis of Variance (ANOVA) and t-test were performed on transformed data to identify the genes with significant changes in expression ($p < 0.05$).

3.2.8 Quantification of pyocyanin

PAO1, $\Delta PA4324$, $PA4324^+$ and $\Delta PA4324^c$ strains were cultured in PB medium¹¹⁵ (20 g/L peptone, 1.4 g/L MgCl₂ and 10 g/L K₂SO₄ dissolved in deionized water) and grown with shaking at 220 RPM at 37 °C for 24 hours. Following that, cells were removed using centrifugation followed by filtration through a 0.2 µm filter (Satorius). Chloroform (Sigma Aldrich) was then used to extract pyocyanin from the culture supernatant. Chloroform was mixed with culture supernatant and vortexed for 2 minutes at high speed. The mixture was then separated by centrifugation (8 000 xg for 10 minutes) and the lower blue-colored organic layer was transferred to a fresh tube. 0.2M HCl (Sigma Aldrich) was then added to this layer and mixed thoroughly

before centrifugation (8 000 xg for 10 minutes). The top aqueous layer was then transferred to a cuvette and measured for absorbance at 520nm. Pyocyanin concentration was calculated using the following formula: $A_{520nm} \times 17.072 =$ pyocyanin concentration ($\mu\text{g/ml}$). Final pyocyanin concentration was derived by normalizing calculated pyocyanin concentration with OD_{600nm} cell densities of its respective culture.\

3.2.9 Quantification of PQS

PQS produced were extracted using solvent extraction and separated using High Performance Liquid Chromatography (Agilent systems). PAO1, $\Delta PA4324$, $PA4324^+$ and $\Delta PA4324^c$ strains were cultured in M9 medium with shaking for 24 hours at 37 °C. Following that, cells were removed using centrifugation followed by filtration through a 0.2 μm filter. Culture supernatant was then extracted three times with acidified ethyl acetate (1 ml acetic acid/liter). The organic layers were then combined and dried to completion and the resultant residue dissolved in methanol. HPLC was performed using a 250 mm by 4.6 mm Agilent XDB-C18 column at 35 °C. Extracts were separated using a linear gradient of 40% (vol/vol) methanol in water to 100% methanol at a flow rate of 0.5 ml/min over 35 minutes. All eluents were acidified with 1 g/L of citric acid. A PQS standard was used to determine the retention time and calibrate the run for quantification of PQS. Final PQS amount was derived by normalizing with OD_{600nm} cell densities of its respective culture.

3.2.10 *pqs* operon transcriptional activity

$PA4324$ mutant strains and PAO1 wild type control were transformed with the reporter plasmid¹¹⁶ and cultured overnight in LB medium at 37 °C with shaking. Overnight cultures were then diluted to OD_{600nm} of 0.01 using M9 medium and 200 μl of each strain were loaded into each well of a 96-well plate. Six replicates were prepared. The plate was then loaded into a microplate reader and monitored over 45 hours, readings of optical density and fluorescence (excitation 486 nm, emission 535 nm) was recorded every 15 minutes.

3.2.11 Type 6 Secretion System (T6SS)-mediated competition assay

The assay was performed as described by Hachani et al¹¹⁷ with slight modifications. The bacteria Prey was prepared by transforming DH5 α cells with pCRISPomyces-2 plasmid that allows complementation of the lacZ gene. The transformed cells were plated on LB agar plates containing apramycin and X-gal (40 μ g/ml, Goldbio) and plates were incubated at 37 °C overnight. The following day, a single blue colony from the transformed cells (Prey) was inoculated into LB broth with apramycin and incubated at 37 °C with shaking for 16 hours. PAO1, Δ PA0012, PA0012⁺ and Δ PA0012^c strains (Killer) were also inoculated into LB broth and incubated similarly.

The following day, a volume of culture equivalent to a cell density of OD_{600nm} 1.0 was collected from each culture and cells were pelleted using centrifugation and the supernatant was discarded. The cells were then resuspended using 100 μ l fresh LB broth. To prepare assay control plates, 10 μ l of each strain was inoculated as a spot onto LB agar plates and dried completely. To prepare competition assay plates, 30 μ l of Prey was mixed with 30 μ l of Killer (PAO1, Δ PA0012, PA0012⁺ or Δ PA0012^c) and mixed thoroughly. 20 μ l of the Prey-Killer mixture was then inoculated as a spot onto LB agar plates and dried completely. Plates were then incubated at 37 °C for 5 hours.

After the incubation to allow for bacterial killing, the individual spots were collected using sterile loops and resuspended in 1 ml of LB broth each. The tubes were then vortexed at low speeds for 15 minutes in order to resuspend the bacteria thoroughly. Following that, each resuspended spot were serially diluted up to 10⁻³ with LB broth. 20 μ l of these dilutions were then spotted in triplicates within a single quadrant of a LB agar plate containing 40 μ g/ml of X-gal. To enable CFU count of the competition assay, resuspended Prey-Killer spots (for example: DH5 α -PAO1, DH5 α -PA0012⁺) were further diluted to 10⁻⁵ and 100 μ l was spread on LB agar plates containing X-gal in triplicates. All plates were then incubated at 37 °C overnight. The following day, images of plates were taken and blue DH5 α colonies (Prey cells that survived) were counted and used to calculate CFU/ml.

3.2.12 Affinity pull-down experiments

To capture proteins interacting with PA0012 or PA4324, pJN105-*HA-PA0012*, pJN105-*HA-PA4324*, pJN105-*HA-PA0012*^{R5A} and pJN105-*HA-PA4324*^{R9A} were transformed into PAO1 cells. To express the recombinant proteins, the transformed strains were inoculated into LB broth and incubated with shaking at 37 °C overnight. The cultures were then used to inoculate fresh LB broth and incubated with shaking at 37 °C until OD_{600nm} of 0.2, at which point 20% L-arabinose (MP Biomedicals) was added to each culture to achieve a final concentration of 0.2%. Cultures were continued to be incubated with shaking at 37 °C until OD_{600nm} of 1.0. Cells were then harvested by centrifugation and resuspended with ice-cold lysis buffer (10% glycerol, 1% NP-40 (MP Biomedical), Protease inhibitor cocktail (Thermo Scientific) in Phosphate-buffered saline).

Cells were lysed on ice using sonication at 30% amplitude power for 5-10 minutes. Total cell lysates were then clarified by centrifugation (15 000 xg for 20 minutes) at 4 °C and the supernatant pre-cleared with protein G Agarose beads (Sigma Aldrich) for 1 hour at 4 °C. After removing the protein G beads, HA-affinity Agarose beads (Sigma Aldrich) were added to the supernatant and incubated for 2 hours at 4 °C. Following incubation, beads were collected by centrifugation and washed five times with lysis buffer at 4 °C. 2x SDS-PAGE sample buffer was then added to the beads and the mixture heated at 95 °C for 10 minutes. Samples were then subjected to SDS-PAGE and gels were stained using Silver Stain Plus Kit (Biorad).

3.3 Results

3.3.1 Phenotypes of $\Delta PA4324$ and $\Delta PA0012$ mutants

PA0012 is one of the smallest PilZ proteins and consists of 88 amino acids; whereas PA4324 consists of 119 amino acids due to an additional C-terminal motif. Both PilZ proteins have been shown to bind cyclic di-GMP⁹³. We performed various phenotypic assays to probe the function of these two genes and their protein products in PAO1. In addition to in-frame deletion mutants^a of these genes ($\Delta PA0012$, $\Delta PA4324$), we also prepared overexpression ($PA0012^+$, $PA4324^+$) and complemented strains ($\Delta PA0012^C$, $\Delta PA4324^C$) by expressing the PilZ genes using the overexpression plasmid pUCP18. The phenotypes of the deletion, overexpression and complementary strains will be compared to reveal the functions of PA0012 and PA4324.

3.3.2 Deletion of $PA0012$ and $PA4324$ had little impact on bacteria motility

Neither $PA0012$ nor $PA4324$ is required for growth as deletion of either genes had negligible effect on the growth rate of PAO1. Cyclic di-GMP is known to regulate bacterial motility and two other PilZ proteins (FlgZ and MapZ) in PAO1 are also known to regulate flagellum-dependent motility. To find out whether $PA4324$ and $PA0012$ are also involved in motility regulation, I used agar plate assays to examine potential differences in swarming, swimming and twitching for the two mutants.

Swarming motility is movement of bacteria across a semi-solid surface and requires not just flagellar activity, but also quorum sensing and production of biosurfactants. Hence any defects in flagellar activity, quorum sensing or biosurfactant production can be detected using a swarming motility assay. No significant differences were observed between $PA0012$ mutant strains and the PAO1 strain (Figure 26A). The overexpression of $PA4324$ results in significantly increased swarming whereas deletion of $PA4324$ caused a modest decrease in swarming (Figure 26B).

^a The mutants were from our collaborator, Professor Xu Linghui, from South China Agricultural University

Figure 26. Representative images of PA0012 and PA4324 swarming motility plates. The graph on the right depicts the area covered by each strain as a percentage of wild type PAO1 (100%). (A) Swarming motility plates of PA0012 strains, no differences in swarming motility were observed between the mutant strains and PAO1. (B) Swarming motility plates of PA4324 strains. Overexpression of PA4324 (middle plate) resulted in increased swarming motility ($p < 0.001$) compared to the PAO1 wild type. No differences in swarming motility were observed between wild type and other mutant strains. All statistical analysis were determined by One-way ANOVA with Dunnett's multiple comparison posttest.

Swimming motility is flagellum-dependent movement in liquid or low viscosity (such as 0.3% agar). We sought to detect any defects in flagella function by using a soft agar swimming motility assay (Figure 27). No differences in swimming motility were observed between the wild type and the mutant strains.

Figure 27. Representative images of PA0012 and PA4324 swimming motility plates. The graph on the right depicts the area covered by each strain as a percentage of wild type PAO1 (100%). (A) Swimming motility plates of PA0012 strains, no differences in swimming motility were observed between the mutant strains and PAO1. (B) Swimming motility plates of PA4324 strains, no differences in swimming motility were observed between the mutant strains and PAO1 as well. All statistical analysis were determined by One-way ANOVA with Dunnett's multiple comparison posttest.

Twitching motility is a form of motility on solid or semi-solid surfaces powered by T4P. The cell extends and retracts its T4P to pull itself on surfaces. This form of motility manifests as an interstitial colony at the interface between an agar media and a surface¹¹⁸. We used the twitching assay to determine if the PilZ mutant strains retained the ability to form interstitial colonies at the interface of a LB agar and the petri dish (Figure 28). No differences were observed between the PilZ mutants and the PAO1 strain.

Figure 28. Representative images of PA0012 and PA4324 twitching motility plates. The graph on the right depicts the area covered by each strain as a percentage of wild type PAO1 (100%). (A) Twitching motility plates of PA0012 strains, no differences in twitching motility were observed between the mutant strains and PAO1. (B) Twitching motility plates of PA4324 strains, no differences in twitching motility were observed between the mutant strains and PAO1 as well. All statistical analysis were determined by One-way ANOVA with Dunnett's multiple comparison posttest.

3.3.3 The *PA0012* and *PA4324* mutants exhibit similar colony morphology as the wild type PAO1

During biofilm development, PAO1 cells display strong auto-aggregation and this ability leads to several macroscopic phenotypes including growth into colonies with unique structures on agar plates¹¹⁹. Deviations in colony morphology from PAO1 may therefore indicate disruptions to the biofilm forming processes. Our lab has previously characterized the functions of the PilZ protein HapZ and found that deletion of the *HapZ* gene leads to a wrinkled colony morphology. Also during late

stage biofilm development, PAO1 produces an extracellular matrix which comprises proteins, DNA and polysaccharides. Certain parts of this process are regulated by cyclic di-GMP¹²⁰ and the PilZ alg44, is required for alginate biosynthesis¹²¹. Congo red is a dye that binds to polysaccharides and can indicate disruptions to carbohydrate synthesis during matrix production. Therefore, we screened for defects in biofilm or polysaccharide production in the strains by observing colony morphology on congo red agar plates. No discernible differences between PA0012 and PA4324 mutants and PAO1 were observed after 5 days of growth at room temperature (Figure 29).

Figure 29. Representative images of PA0012 and PA4324 colonies after 5 days. (A) Colony images of PA0012 strains, no differences in macroscopic morphology can be observed between the mutant strains and PAO1. (B) Colony images of PA4324 strains, no differences in macroscopic morphology can be observed between

3.3.4 Deletion of *PA0012* and *PA4324* did not cause significant changes in antibiotic susceptibility

PAO1 is known to possess intrinsic resistance to many classes of antibiotics. At least 12 multi-drug efflux pumps have been identified in its genome¹²² and expression of the genes encoding these efflux pumps are under the regulation of many processes. It was reported that genes encoding efflux pumps were up-regulated when cellular concentration of cyclic di-GMP is elevated¹²³. Nicastro et al also found that high cyclic di-GMP enhances survival of planktonic cells in the presence of Imipenem¹²⁴, a

drug of choice for treating clinical *P. aeruginosa* infections. We therefore speculated that one of the PilZ proteins in PAO1 may function as a cyclic di-GMP receptor that regulates the expression of efflux pumps genes. To test this, we examined the susceptibility of *PA0012* and *PA4324* mutant strains towards various antibiotics.

Deletions of *PA0012* and *PA4324* do not appear to affect the susceptibility of the bacteria towards any of the antibiotics tested. We did however notice a decreased susceptibility to streptomycin in the $\Delta PA4324$ mutant and increased susceptibility towards it in the $\Delta PA0012$ mutant and the effect is considered to be moderate¹¹⁴.

Table 1. Susceptibility of $\Delta PA0012$, $\Delta PA4324$ and PAO1 strains to various antibiotics.

	MIC ($\mu\text{g/ml}$)		
	PAO1	$\Delta PA0012$	$\Delta PA4324$
Apramycin	8	2	8
Kanamycin	32	32	32
Chloramphenicol	>256	256	128
Erythromycin	256	256	256
Gentamicin	32	16	16
Tetracycline	16	32	32
Lincomycin	>256	128	128
Novobiocin	256	256	128
Neomycin	4	8	8
Spectinomycin	256	256	128
Streptomycin	64	32	128
Nalidixic acid	>256	128	128
Ciprofloxacin	<0.5	<0.5	<0.5

3.3.5 Deletion of *PA0012* and *PA4324* changed gene expression profile based on transcriptome analysis^b

As cyclic di-GMP often exerts its effect by modulating gene expression, there is a possibility that *PA0012* and *PA4324* may affect gene expression and mRNA levels. Using RNA sequencing (RNA seq), we examined the gene expression profiles of PAO1, $\Delta PA0012$ and $\Delta PA4324$ strains and their respective complemented strains ($\Delta PA0012^c$, $\Delta PA4324^c$).

^b RNA sequencing was done in collaboration with SCELSE, NTU

3.3.6 Gene expression profile of *PA0012* mutants

A total of 401 genes were differentially expressed in $\Delta PA0012$ compared to PAO1 and these genes were grouped according to their Pseudocap Function Class and summarized in Figure 30. We also found no significant differences between the gene expression profiles of $\Delta PA0012^c$ and PAO1 strains, suggesting that the differential gene expression observed in the $\Delta PA0012$ mutant were due to deletion of *PA0012*. We chose a fold change of ± 3.5 as a cutoff to filter the differentially expressed genes and 72 genes meet the criteria (Figure 31A); these genes were grouped into 10 functional categories (Figure 31B). Not including hypothetical genes, the biggest category of genes with ≥ 3.5 fold change were genes involved in secretion and export.

We found that all the genes in this highly up-regulated category belong to the Hcp Secretion Island I (HSI-I) gene cluster. The HSI-I is one of three Type VI secretion systems (T6SS) in PAO1 and experimental evidence suggests it has evolved to kill bacteria specifically¹²⁵. Unsurprisingly, the Type VI exported (Tse) and Type VI secreted immunity (Tsi) proteins 1-3, were also up-regulated (2-7 fold). Tse 1-3 genes code for toxin proteins and Tsi 1-3 are the corresponding protective proteins which are found in the same operon as the toxin genes¹²⁶. Figure 30C and D shows the gene organization of HSI-I gene cluster, Tsi and Tse genes and their relative fold changes upon *PA0012* deletion. The HSI-I-encoded T6SS proteins mediate the export of Tse proteins to other bacteria and the Tsi proteins serve as protective proteins against these toxins. The HSI-I gene cluster encodes membrane proteins which crosses the cell envelope and anchor the T6SS complex¹²⁷. It also encodes proteins which are related to bacteriophage T4 tail complex that acts as a syringe and spike to penetrate target cell and inject the toxin proteins^{128,129}.

Figure 30. Pie chart depicting percentage ratios of the functional groups of genes perturbed by the deletion of *PA0012*. Based on predicted functions, genes that were expressed differentially in the $\Delta PA0012$ mutant were classified into 21 functional groups and their respective percentage ratios were represented using a pie chart. Pie chart was generated using Graphpad Prism software. (A) Percentage ratios of genes up-regulated in the $\Delta PA0012$ mutant, the total number of up-regulated genes is denoted below the pie chart; and (B) Percentage ratios of genes down-regulated in the $\Delta PA0012$ mutant. The total number of genes in each case is denoted below the pie chart.

We also found the non-coding RNA, *rsmZ*, to be up-regulated 8.7-fold and *rsmZ* has been reported to increase expression of HSI-I genes¹²⁶; this increase likely explains the increased activation of the HSI-I and associated gene clusters. The transcriptome study suggests that *PA0012* regulates the expression of the anti-bacterial T6SS HSI-I gene cluster and other functionally related genes via *rsmZ*. Intriguingly, no other components of the Gac/*rsm* pathway were affected by *PA0012* and further research is necessary to identify the branch of this pathway on which *PA0012* acts to regulated these genes.

turn blue in the presence of X-gal. By co-culturing the killer strain with the prey strain and assessing the survival of the prey bacteria, the T6SS-mediated activity of the killer strain can be qualitatively assessed. In addition, a quantitative measurement of survival is also conducted by measuring the CFU of the prey strain post-killing.

Figure 32. T6SS bacterial competition assay (A) Quantitative assessment of the survival of prey bacteria in CFU/ml after a 5 hour contact time with the killer strains. Significantly decreased survival is observed in prey incubated with $\Delta PA0012$ ($p = 0.004$) and the opposite is observed with $PA0012^+$ ($p = 0.005$). All statistical analyses were determined using one-way ANOVA with Dunnett's multiple-comparison posttest. (B) Killer/prey co-cultures following an overnight incubation. The degree of survival can be evaluated qualitatively by the degree of blue color (surviving prey) in the spots.

The outcome of the assay is summarized in Figure 32, Figure 32A depicts the survival of prey bacteria after a 5 hour contact time as determined by CFU count the following day. The killer/prey co-cultures were also serially diluted and plated on LB

agar plates in triplicates to allow killing to continue overnight, the degree of survival of the prey bacteria can be evaluated by the ratio of blue/white color in the bacteria spots following the overnight incubation (Figure 32B). We observed a significant decrease in the survival of prey bacteria co-cultured with $\Delta PA0012$ killer (16% survival) and a reciprocal increase in the survival of those co-cultured with $PA0012^+$ killer (180% survival) compared to PAO1 control. This observation agrees with the gene expression profile derived from transcriptome analysis which suggests that PA0012 represses the HSI-I and associated genes.

3.3.8 Gene expression profile of $PA4324$ mutants

Figure 33. Pie chart depicting percentage ratios of the functional groups of genes perturbed by the deletion of $PA4324$. Based on predicted functions, genes that were expressed differentially in the $\Delta PA4324$ mutant were classified into 21 functional groups and their respective percentage ratios were represented using a pie chart. Pie chart was generated using Graphpad Prism software. (A) Percentage ratios of genes up-regulated in the $\Delta PA4324$ mutant and (B) Percentage ratios of genes down-regulated in the $\Delta PA4324$ mutant. The total number of genes in each case is denoted below the pie chart.

In the case of PA4324, a total of 547 genes were differentially expressed in the Δ PA4324 mutant, a total of 326 genes were up-regulated and 221 genes were down-regulated (Figure 33). After applying a cut-off of ≥ 3.5 fold change, 35 genes remained and were organized into 8 functional categories. It was observed that genes from the biosynthesis of cofactors and prosthetic groups and secretion and export categories were over-represented in this group of genes (Figure 34B).

The *pqs* gene cluster was amongst the genes from these two categories and were highly up-regulated with an average increase of 25-fold (Figure 34C). The *pqs* operon codes for enzymes that converts anthranilic acid to 2-heptyl-4-quinolone (HHQ), which is then converted into the *Pseudomonas* quinolone signal (PQS) by an enzyme encoded by the *pqsH* gene¹³⁰. PQS is one of the quorum sensing molecules in *Pseudomonas*, and serves as an intercellular communication signal and plays a role in coordinating expression of genes for biofilm formation and virulence¹³¹. Studies have also revealed that PQS and its precursor HHQ regulate the expression of *phz1* operon and *phz2* operon respectively^{132,133} and regulate pyocyanin production¹³⁴.

In accordance with this, both *phz1* and *phz2* operons were up-regulated in Δ PA4324 mutant, genes in the *phz1* operon were up-regulated by an average of 2.6-fold and genes in the *phz2* operon were up-regulated by an average of 4.2 fold when compared to the PAO1 wild type (Figure 34D). The *phz* operons codes for biosynthetic enzymes that produce phenazine-1-carboxylic acid (PCA-1) which is then converted by other enzymes into phenazines. Phenazines such as pyocyanin are biologically active secondary metabolites that enhance host competitiveness and virulence during infection.

The number of genes perturbed by the deletion of PA4324 is about 20% more than those affected in Δ PA0012 mutant. This is likely due to the differences in the downstream processes that are regulated by the top-level functions affected by PA0012 and PA4324. PA0012 regulates the expression of T6SS genes, which lead to a very specific phenotype but in the case of PA4324, genes from the *pqs* gene cluster

are affected and the products of this gene cluster, HHQ and PQS are known to regulate many processes in PAO1.

Figure 34. Deletion of *PA4324* increases expression of the *pqs* operon. (A) Heat map of genes with ≥ 3.5 fold change in $\Delta PA4324$ and $\Delta PA4324^c$ mutants and; (B) Functional categories of these 35 genes. Genes involved in PQS and phenazine production are circled. (C) Gene organization of the *pqs* and *phn* operon; the graph above indicates their expression levels in $\Delta PA4324$ mutant. (D) Gene organization of the *phz1* and *phz2* operons, likewise, the graph above indicates their expression levels in $\Delta PA4324$ mutant.

3.3.9 Quantification of pyocyanin and PQS and analysis of the transcriptional activity of the *pqs* operon^C

We sought to measure pyocyanin produced by PAO1 and *PA4324* mutant strains as an initial step to confirming the gene expression profiles. Pyocyanin produced after

24-hour culture in liquid medium was extracted and quantified using its absorbance at 520nm. In agreement with the transcriptome data, deletion of *PA4324* leads to increased production of pyocyanin (~30%) when compared to the wild type strain. Overexpression of *PA4324* also causes a reciprocal reduction in pyocyanin production (Figure 35A). After confirming that pyocyanin is indeed produced in higher amounts by the $\Delta PA4324$ mutant, we next sought to confirm that increased expression of the *pqs* operon translates to increased PQS in culture broth. The strains were cultured for 24 hours at 37 °C in liquid medium and PQS was extracted from the culture medium and quantified using HPLC by comparing against a PQS standard (Figure 35B and C).

Figure 35. Quantification of pyocyanin and PQS; and transcriptional activity of the *pqsA* promoter over 45 hours. (A) OD_{600nm} normalized pyocyanin concentration after 24 hours at 37°C. $\Delta PA4324$ produces more pyocyanin than PAO1 in liquid culture, and overexpressing *PA4324* appears to lower pyocyanin production by the bacteria. (B) PQS produced after a 24 hour culture in liquid medium at 37°C expressed as a percentage of PAO1 control. (C) Absorption spectrum of PQS standard, its retention time during HPLC is shown in the inset. (D) Induction of the *pqsA-gfp(ASV)* transcriptional reporter in PAO1 and $\Delta PA4324$ mutant. Statistical analysis were determined by One-way ANOVA with Dunnett posttest.

c Analysis of the transcriptional activity of the *pqs* operon was done in collaboration with SCELSE, NTU.

We were unable to quantify HHQ as we could not obtain a HHQ standard, however since PQS is converted from HHQ, it would not be unreasonable to draw parallels between the amounts of PQS and HHQ produced. The amount of PQS produced mirrors the trend observed in the pyocyanin experiment (Figure 35B). Finally, we also examined the expression pattern of the *pqs* operon over 45 hours by transforming the PAO1 and $\Delta PA4324$ strains with a plasmid-based *pqsA* promoter : : *gfp(ASV)* transcriptional reporter. This plasmid encodes an unstable version of GFP and is thus useful for monitoring transcriptional activity of the *pqs* operon¹³⁵. The results indicate that the *pqs* operon in the $\Delta PA4324$ mutant has consistently higher expression levels when compared to the wild type PAO1 strain over the period of the experiment (Figure 35D). Our accumulated experimental evidence points to PA4324 functioning as a regulator controlling the expression of the *pqs* operon and as a consequence, it also regulates other genes affected by PQS and pyocyanin levels.

3.3.10 Identification of protein partners of PA0012 and PA4324 by protein pull-down assays

We have previously screened for potential protein partners of PA0012 and PA4324 using Bacterial two-hybrid but failed to find any interacting proteins. We then speculated that the two proteins may only bind their partners in a complex consisting of three or more proteins. We therefore decided to use an affinity pull-down approach which will enable us to capture the entire protein complex (if any) from a cell lysate of PAO1 cells by using PA0012 and PA4324 as bait proteins. HA-tagged PA4324 and PA0012 were expressed in PAO1 cells as bait proteins and HA-affinity agarose beads were used to capture these proteins together with any putative interaction partners. In addition, HA-tagged PA0012^{R5A} and PA4324^{R9A} mutants were used as control. We reasoned these mutants that cannot bind cyclic di-GMP would be unable to interact with their true partners in the cell but would retain non-specific interactions with other protein; this may then reduce presence of false positive

background bands. Unfortunately, our pull-down experiments did not reveal any potential binding partners of PA0012 or PA4324 (Figure 36).

Figure 36. Representative images of silver-stained gels of pull down experiments. (A) HA-tagged PA4324 or PA4324^{R9A} expressed in PAO1 as bait. The arrow indicates the position of the PA4324 or PA4324^{R9A} protein. No differences were found between either lane. (B) HA-tagged PA0012 or PA0012^{R5A} expressed in PAO1 as bait. The arrow indicates the position of the PA0012 or PA0012^{R5A} protein. No differences were found between either lane.

3.4 Discussion

This chapter describes my research work to probe the biological function of the two single domain PilZ proteins PA0012 and PA4324. The swimming, swarming and twitching motility modes of *PA0012* and *PA4324* mutants were evaluated. We observed a significant increase in swarming motility when *PA4324* was overexpressed and a moderate decrease when *PA4324* was deleted, suggesting that PA4324 may promote swarming. Meanwhile, colony morphology and antibiotics susceptibility of the deletion mutants did not show significant differences from the wild type strain. These observations suggest that the two PilZ proteins are not involved in the regulation of EPS production and expression of efflux pump proteins.

Transcriptome analysis by RNA-sequencing showed that deletion of *PA0012* and *PA4324* had distinct effect on gene expression profile of the bacterium. This strongly

suggests that cyclic di-GMP regulates gene expression through the two PilZ proteins and presumably two different signaling pathways. For *PA0012*, the most significant observation is that HSI-I, one of the three T6SS responsible for killing other bacteria, is up-regulated. Follow-up bacterial competition assay confirmed that the $\Delta PA0012$ strain exhibited enhanced killing of prey bacteria and the overexpression strain exhibited suppressed killing efficiency. Together, the observations led us to conclude that *PA0012* is part of a cyclic di-GMP signaling pathway that regulates the T6SS in PAO1. Considering that high cellular cyclic di-GMP levels generally lead to biofilm formation, the putative role of *PA0012* and its associated pathway is to “disarm” PAO1 cells prior to a more passive community living style in a biofilm.

The differences in the gene expression profile of $\Delta PA4324$ mutant suggests that *PA4324* is likely to regulate the expression of the *pqs* operon in PAO1, considering that an average increase expression of 25-fold was observed for the *pqs* operon. Increased expression in the *phz1* and *phz2* operons, which are regulated by HHQ and PQS were also observed, which is consistent with the increased production of pyocyanin by $\Delta PA4324$ mutant in liquid medium. Together, the results suggest that cyclic di-GMP regulate the expression of the *pqs* operon through *PA4324*. Because the expression of *phz* genes is regulated by HHQ and PQS levels^{132,133}, the increased production of pyocyanin is likely the consequence of increased HHQ/PQS production. As the ability of *P. aeruginosa* to swarm is also affected by PQS^{136,137}, perturbations to the *pqs* genes also explain the changes of swarming motility for the *PA4324* mutants. The observation that the increase in pyocyanin and PQS was not as drastic as the corresponding increase in gene expression suggests that an unknown mechanism is restricting the output of both pathways. One possibility may be due to substrate depletion as both pyocyanin and PQS pathways use chorismic acid as the starting material^{138,139} (Figure 37). Alternatively, there may be regulation at the translational stage and the increased mRNA levels are only translated into slightly increased protein levels in both pathways.

The attempt to utilize protein-pulldown experiments to identify potential binding partners of *PA0012* and *PA4324* was unsuccessful. This could possibly be due to low expression levels of the partner proteins. Also, if a prey protein was indeed captured

but has the same molecular weight as the antibody light or heavy chains, its band would be masked by the stronger antibody bands on the SDS-PAGE gel. If this was indeed the case, using dissociation buffers that do not denature the antibody light and heavy chains could be considered. Finally, the possibility that PA0012 and PA4324 may bind nucleic acids like the PilZ proteins MrkH and BB0733 should be also considered.

Figure 37. Biosynthetic pathways for PQS and pyocyanin share a common substrate.

3.5 Conclusion

The aim of this PhD research project was to unveil the functions of the uncharacterized PilZ proteins of *P. aeruginosa*. The research work described in this chapter shed light on the functions of the two single-domain PilZ proteins, PA0012 and PA4324, in regulating the expression of genes related to T6SS and quorum-sensing molecule production respectively. The results expand our current understanding of the function of cyclic di-GMP in *P. aeruginosa*. Although we have not identified the protein partners of the two proteins to unveil the detailed regulatory mechanism, the work described in this chapter represents important advances in our understanding of the previously elusive roles of the two PilZ proteins.

Chapter 4 – Probing the function of the two di-domain PilZ proteins in *P. aeruginosa* PAO1

4.1 Introduction

Di-domain PilZ proteins consists of a PilZ-domain joined to another domain and they are the most numerous group of PilZ proteins after the single-domain PilZ proteins. They are also better studied amongst PilZ proteins, with seven di-domain PilZ proteins characterized and two crystal structures published. Similar to the single-domain PilZ proteins, there is extensive divergence in the di-domain PilZ proteins and they regulate many different cellular processes. One of the best studied di-domain PilZ proteins is YcgR, which acts as a back-stop brake interacting with the flagella motor of *E. coli* to inhibit swimming motility^{51,52}. MrkH, another di-domain PilZ protein, binds to DNA to act as a transcriptional activator of Pili genes⁵³. The crystal structures of di-domain PilZ proteins PP4397 and PlzD show that these proteins typically undergo a conformational change upon cyclic di-GMP-binding and both domains provide residues to bind cyclic di-GMP directly^{33,35}. Importantly, it was also demonstrated that a di-domain PilZ protein (PlzD) can switch from binding monomeric cyclic d-GMP to dimeric cyclic di-GMP through the mutation of a single residue³⁵, this hints the versatility of the di-domain PilZ proteins at evolving and acquiring new functions or binding partners.

PA2989 and PA3353 (at the time when we started, the protein had not been named as FlgZ) are the only two di-domain PilZ proteins encoded in the genome of PAO1 and the physiological role of PA2989 remains unknown although some studies have shown that its expression is increased in antibiotic-resistant strains of *P. aeruginosa*^{140,141}. Apart from these two studies, very little information on PA2989 can be found in literature. As described in chapter 2, we found that PA2989 has a unique sequence whereby its canonical cyclic di-GMP binding motifs were separated by more residues than other PilZ proteins. However, this insertion appears to have no deleterious effect on its cyclic di-GMP binding affinity and in fact PA2989 has the highest binding affinity amongst PilZ proteins in PAO1⁹³. Elucidating the function of PA2989 and its cyclic di-GMP-binding mechanism which may differ significantly from

other di-domain PilZ proteins would possibly contribute much to our understanding of PilZ proteins.

With a YcgR_N domain, PA3353 was presumed to regulate swimming motility by controlling flagellar output. Our sequence similarity network analysis (Chapter 2) suggests that PA3353 may have a different function than YcgR. YcgR functions by interacting with proteins of the flagella motor complex, including motA⁸³ and PAO1 possesses two sets of stator proteins, MotAB (*PA4954/PA4953*) and MotCD (*PA1460/PA1461*). While my research was in progress, Baker et al reported that PA3353 (renamed as FlgZ) has been shown to bind MotC at elevated cyclic di-GMP levels and inhibit swarming motility⁴⁹. The precise mechanism of this motility suppression is still unknown and how it affects the dynamics of the flagella motor output and subsequently motility is still unclear.

In this chapter, I present my work to probe the biological function of PA2989 and FlgZ in PAO1. The phenotypes of *PA2989* and *FlgZ* null mutants are compared with that of the wild type PAO1. Transcriptome analysis of *PA2989* mutants were also performed using RNA sequencing to study its potential role in gene expression. Building upon the discoveries from chapter 2, I constructed a structural model for PA2989 to reveal a novel di-domain PilZ protein architecture with an inserted helix domain. I will also present my findings that FlgZ impacts flagellar rotational speed and torque by analyzing the *FlgZ* mutants using cell tethering and swimming trajectory assays.

4.2 Materials and Methods

4.2.1 Cloning of PilZ genes into overexpression plasmids

PA2989 and *PA3353* genes were cloned from PAO1 genomic DNA using Polymerase Chain Reaction (PCR) with Q5 polymerase. The product was separated using electrophoresis on 1% agarose gel and the bands of the correct size were excised and purified using a Gel and PCR clean up kit. The restriction enzymes *SacI* and *XbaI* (NEB) were used to cut the purified PCR fragments and the plasmid pUCP18. Following that, the fragments were ligated into pUCP18 using T4 ligase and the mixture transformed into *E. coli* DH5 α cells. Plasmids extracted from the transformed cells were verified by sequencing.

4.2.2 Bacteria two-hybrid screening

The BacterioMatch II Two-Hybrid System (Stratagene) was used to probe the interactions between *FlgZ* and *FliG*, *FliM* and *MotA*. The genes were cloned from PAO1 genomic DNA using PCR and cloned into the respective plasmids by restriction digest and ligation. *FlgZ* ligated into pBT vector using restriction enzymes *EcoRI* and *XhoI*. *FliG* and *FliM* were ligated into pTRG vector using the restriction enzymes *EcoRI* and *XhoI* and; *MotA* was ligated into the same vector using the enzymes *EcoRI* and *SpeI*. Constructs were verified by sequencing. The bait and target plasmids were co-transformed into the reporter strain by electroporation. Co-transformed cells were grown on non-selective agar medium at 37 °C containing tetracycline (10 μ g/ml, Sigma Aldrich) and chloramphenicol (25 μ g/ml, MP Biomedicals,) overnight. The following day, three co-transformed clones were picked and resuspended in sterile water and 3 μ l was inoculated on M9⁺ His-deficient medium containing 5 mM 3-amino-1,2,4-triazole (Sigma Aldrich) and non-selective medium containing tetracycline, chloramphenicol. Plates were then incubated at 37°C for up to 48 hours.

4.2.3 Cell tethering assay

Single colonies of PAO1, Δ *FlgZ* and *FlgZ*⁺ strains were inoculated into 10 ml of LB media and incubated at 37 °C overnight with shaking. Overnight cultures were then diluted to OD_{600nm} of 0.2 using M9 medium and growth at 37 °C with shaking until

exponential growth phase at OD_{600nm} of 0.8. Cell cultures were then diluted 10-fold and then flagella were sheared off by passing the cells through a 34-gauge blunt needle. Cells were then loaded into a microfluidic stagnation flow device¹⁴² which was pre-coated with flagellar antibodies. An inverted microscope (Nikon TE2000-U) was used to visualize the cells and capture videos of the tethered bacteria; video was taken at 40 fps for 1-5 minutes. Videos were processed using ImageJ (NIH) and analyzed using the in-house analysis program BTAP.

4.2.4 Swimming trajectory analysis

Free swimming PAO1 and *FlgZ* mutant cells with their intact flagellum were grown to exponential phase of OD_{600nm} of 1.0 in LB medium and diluted 100-fold with LB medium. 20 ml of the diluted culture was then added to 80 ml of separately prepared 6.25%, 12.5%, 18.75% and 25% glycerol-LB solutions each to obtain final 100 ml of 5%, 10%, 15% and 20 % glycerol-LB bacteria solutions respectively. These solutions were then loaded in cell chambers with a channel height of 20 µm and observed in the middle of the chamber depth (away from the coverslip). Cells were visualized using an inverted microscope (Nikon TE2000U) and videos of bacteria were taken at 25 fps using a CMOS camera (Thorlabs DCC1645). Image were processed using ImageJ (NIH) and bacterial swimming trajectories were obtained using particle segmentation and tracking algorithms as described previously^{143,144} and implemented through the software Trackmate Plugin in Fiji¹⁴⁵.

4.2.5 3D structure prediction using I-TASSER software suite

Full length protein sequence of PA2989 was submitted for structure prediction at the following URL: <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>. The software Pymol was used to view and manipulate the pdb file generated by the server.

4.3 Results

The experimental strategy to probe the functions of PA2989 and FlgZ is similar to that used for the single-domain PilZ proteins PA0012 and PA4324. We examined the phenotypic variation between the in-frame deletion mutants^a ($\Delta PA2989$, $\Delta FlgZ$) and the wild type strain. The differences in phenotype and gene expression will allow us to formulate hypothesis regarding the cellular function of the two PilZ proteins, which will be subjected to further experimental validation. In addition to the deletion mutants, we also generated PA2989 and FlgZ overexpressing mutants ($PA2989^+$, $FlgZ^+$) and complementary strains ($\Delta PA2989^c$, $\Delta FlgZ^c$) by using an exogenous plasmid to express the PilZ genes.

This investigation was initiated when the exact functions of PA2989 and FlgZ was unknown and although Baker et al has elucidated the role of FlgZ in suppressing swarming motility⁴⁹ since then, our experiments reveal additional data that furthers the knowledge regarding the role of this protein in the cell.

4.3.1 PA2989 and FlgZ mutants have altered motility modes compared to PAO1

We first examined the swarming, swimming and twitching motility of the mutants considering that motilities, including the flagellum-dependent swimming, swarming and T4P-dependent twitching motility, are some of the most common bacteria behaviors regulated by cyclic di-GMP.

Swarming motility is dependent on flagella, quorum sensing and biosurfactant production. Deletion of PA2989 abolishes swarming on M8 agar plates and overexpression of the gene causes an opposite increase in swarming motility. Complementation using an overexpression plasmid recovered the swarming phenotype but the swarm pattern of the complementary strain appears slightly different from the PAO1 wild type strain with its radiating tendrils (Figure 38A).

^a The mutants were from our collaborator, Professor Xu Linghui, from South China Agricultural University

Deleting or overexpressing *FlgZ* in PAO1 results in slight changes to its swarming motility which are not statistically significant as determined by One-way ANOVA (Figure 38B). This observation seems to agree with a recent study which reported that deleting *FlgZ* only leads to an 11% increase in swarming area in PA14 cells⁴⁹.

Figure 38. Representative images of *PA2989* and *FlgZ* swarming motility plates. The graph on the right depicts the area covered by each strain as a percentage of wild type PAO1 (100%). (A) Swarming motility plates of *PA2989* mutant strains. The $\Delta PA2989$ strain loses its ability to swarm (left plate, $p = 0.04$) while overexpressing *PA2989* in PAO1 results in increased swarming motility (middle plate, $p = 0.003$). (B) Swarming motility plates of *FlgZ* strains. Deleting *FlgZ* appears to cause a slight increase in swarming motility (left plate) while overexpressing the gene causes a decrease in swarming (middle plate), but the differences were statistically insignificant. All statistical analysis were determined by One-way ANOVA with Dunnett's multiple comparison posttest.

Swimming motility was analyzed using soft agar plates which would allow us to observed changes in flagellar function and associated chemotaxis system caused by the deletion and overexpression of *PA2989* and *FlgZ*. A statistically significant increase in swimming motility

Figure 39. Representative images of *PA2989* and *FlgZ* swimming motility plates. The graph on the right depicts the area covered by each strain as a percentage of wild type PAO1 (100%). (A) Swimming motility plates of *PA2989* strains, overexpressing *PA2989* leads to an increased swim area ($p = 0.002$). (B) Swimming motility plates of *FlgZ* strains, overexpressing *FlgZ* leads to a large decrease in swimming motility of PAO1 cells ($p < 0.001$). All statistical analysis were determined by One-way ANOVA with Dunnett's multiple comparison posttest.

was observed in the $\Delta PA2989$ strain but the opposite effect was not observed when *PA2989* was overexpressed (Figure 39A). Deletion of *FlgZ* resulted in slightly increased swimming area but the largest effect was observed when *FlgZ* was overexpressed, a reduction of ~50% in swimming area was observed ($p < 0.001$, Figure 39B). Complementation of $\Delta FlgZ$ strains was unsuccessful ($p = 0.0021$) and the swim area was reduced compared to PAO1, the plasmid-expressed *FlgZ* levels was probably artificially high due to the constitutive promoter and resulted in an overexpression phenotype.

Twitching motility requires a functional T4P in the bacteria. No changes to twitching motility was observed in the *FlgZ* mutant strains, which is in accordance with the genetic observations that the *flgZ* is clustered with flagellar biogenesis genes (Figure 40A). Deletion of *PA2989* led to reduced twitching motility that cannot be complemented using an exogenous plasmid expressing *PA2989* (Figure 40B).

Figure 40. Representative images of PA2989 and FlgZ twitching motility plates. The graph on the right depicts the area covered by each strain as a percentage of wild type PAO1 (100%). (A) Twitching motility plates of PA2989 strains, statistically significant reductions in twitching area were observed in deletion and complementary strains. (B) Twitching motility plates of FlgZ strains, no differences in twitching motility were observed between the mutant strains and PAO1. All statistical analysis were determined by One-way ANOVA with Dunnett's multiple comparison posttest.

We speculated that this phenotype is not a direct effect of PA2989 deletion but rather a result of the polar effect from the deletion of this gene. PA2989 is not only in the same operon as *lolC* (PA2986), *lolD* (PA2987) and *lolE* (PA2988) but also overlaps with *lolE*. However, our subsequent transcriptome analysis indicated that the expression levels of *lol* genes in the Δ PA2989^c strain was the same as PAO1, this suggests that the changed twitching motility is not due to these genes and that the PA2989 level in the complementary strain may be more similar to the deletion strain than the wild type strain under the experimental condition.

4.3.2 Deletion of PA2989 results in a wrinkled colony

The colony morphology of *P. aeruginosa* is sensitive to changes in many cellular functions, including changes in the production of EPS, phenazines and quorum sensing autoinducers. I performed colony morphology assay to test whether PA2989 and FlgZ regulate those cellular functions. I found that the Δ PA2989 strain developed into a wrinkled colony that was smaller than those formed by the other strains (Figure 41A). Again, this phenotype could not be complemented completely with an exogenous plasmid. Instead, the complementary strain developed into a rough and glossy-looking colony that was larger than the Δ PA2989 colony but this phenotype

was not observed in the overexpression strain. No discernible differences were observed in between either of the *FlgZ* strains and the wild type control.

Figure 41. Representative images of *PA2989* and *FlgZ* colonies after 5 days. (A) Colony images of *PA2989* strains, the deletion strain presents a wrinkled morphology (second image) which cannot be complemented completely (last image). (B) Colony images of *FlgZ* strains, no differences in macroscopic morphology can be observed between either of the strains.

4.3.3 Deletion of *PA2989* causes changes to antibiotic susceptibility

Table 2. Susceptibility of $\Delta PA2989$ and PAO1 strain to various antibiotics.

	MIC ($\mu\text{g/ml}$)	
	PAO1	$\Delta PA2989$
Apramycin	8	8
Kanamycin	32	32
Chloramphenicol	>256	128
Erythromycin	256	256
Gentamicin	32	2
Tetracycline	16	8
Neomycin	4	8
Spectinomycin	256	256
Streptomycin	64	32
Nalidixic acid	>256	>256
Ciprofloxacin	<0.5	<0.5

Mutant PA14 strains with increased expression of the *PA2989* homolog has been shown to display increased resistance to β -lactam¹⁴⁰ and cephalosporin-class¹⁴¹ of antibiotics. Our susceptibility screen indicates that the $\Delta PA2989$ mutant appear to have increased susceptibility towards gentamicin, which is an aminoglycoside. No differences in the susceptibility towards other aminoglycosides were detected (Table 2).

4.3.4 Deletion of *PA2989* down-regulates protective genes as shown by transcriptome analysis^b

As cyclic di-GMP also exerts its effect by modulating gene expression, we also performed transcriptome analysis by RNA sequencing to test whether *PA2989* affects gene expression. We found that a total of 431 genes were differentially expressed in the $\Delta PA0012$ strain when compared the PAO1 control with 211 up-regulated and 220 down-regulated genes (Figure 42).

109 genes had a fold change of ≥ 3.5 (Figure 43A and B), of which 34 genes were not complemented in the $\Delta PA2989^c$ strain. Of the remaining 75 genes, 11 genes were known whilst the rest are hypothetical genes with unknown functions, the fold change of these 11 genes in the $\Delta PA2989$ strain is depicted in Figure 43C. Five genes, *dctA*, *leuA*, *nalC*, *phr* and *rsmZ* were up-regulated when *PA2989* was deleted. *NalC* codes for a transcriptional repressor of the MexAB-OprM efflux pump and this efflux pump has been reported to export β -lactam antibiotics¹⁴⁶ and may explain why *PA2989* has been implicated in resistance to β -lactam-class antibiotics. The gene *rsmZ*, which is known to suppresses swarming motility, was also up-regulated¹⁴⁷ and may explain the loss of swarming motility. Interestingly, although *rsmZ* was also up-regulated in the $\Delta PA0012$ strains, its effect was limited only to the *HCI-I* genes and swarming motility was not affected. These two PilZ proteins may act on the same

^b RNA sequencing was done in collaboration with SCELSE, NTU

Figure 42. Pie chart depicting percentage ratios of the functional groups of genes perturbed by the deletion of *PA2989*. Based on predicted functions, genes that were expressed differentially in the $\Delta PA4324$ mutant were classified into 21 functional groups and their respective percentage ratios were represented using a pie chart. Pie chart was generated using Graphpad Prism software. (A) Percentage ratios of genes up-regulated in the $\Delta PA2989$ mutant and (B) Percentage ratios of genes down-regulated in the $\Delta PA2989$ mutant. The total number of genes in each case is denoted below the pie chart.

pathway but are somehow able to restrict the resultant phenotype to a very specific one. *DctA* and *leuA* are involved in carbon metabolism while *phr* encodes a photolyase^{148,149}. On the other hand, all of the six known genes down-regulated by the deletion of *PA2989* (Figure 43C) have been reported to exert protective effects. Cytochrome p450 (*PA3331*) belongs to the family of enzymes that are involved in diverse metabolic activities, including the degradation of xenobiotics.

Figure 43. Deletion of PA2989 decreases expression of protective genes. (A) Heat map of genes with ≥ 3.5 fold change in $\Delta PA2989$ and $\Delta PA2989^c$ when compared to PAO1. (B) Gene categories of the 109 genes that have a fold change of 3.5 or greater. (C) Fold change of the 11 known genes which were completely complemented in $\Delta PA2989^c$.

Both *osmC* and *osmE* are also down-regulated, OsmC is a peroxiredoxin which is believed to defend against oxidative stress¹⁵⁰ and OsmE is a lipoprotein that contains SmpA/OmlA domains (Interpro annotation) and is likely to be involved in maintaining the structural integrity of the cell membrane. LptF is also a lipoprotein that is found on the outer membrane and was found to mediate resistance to oxidants¹⁵¹. *PfpI* is an anti-mutator gene and has been shown to play an important role in DNA

protection¹⁵². The gene, *dadA*, which showed the greatest change in expression level in the $\Delta PA2989$ strain, encodes a D-Amino acid dehydrogenase enzyme necessary for D-amino acid catabolism. Importantly, it was reported that the *dadA* was highly expressed in cystic fibrosis-adapted *P. aeruginosa* strains¹⁵³ and is required for optimal production of cyanide¹⁵⁴. DadA was also subsequently reported to be required for the optimal production of pyocyanin, pyoverdine and rhamnolipid and also for biofilm formation in cystic fibrosis-adapted *P. aeruginosa* strains¹⁵⁵.

Many of the differentially expressed genes in the $\Delta PA2989$ strain appear to have protective effects and points to *PA2989* having a role in enhancing many protective functions in PAO1. All six genes that were down-regulated when *PA2989* was deleted mediate some form of protection, in particular DadA mediates the production of many factors that protect the bacteria and serve as virulence factors against the host in cystic fibrosis. The loss of swarming motility when *PA2989* was deleted may be explained by the up-regulated *rsmZ* but none of the genes that were differentially expressed can explain the increased swimming motility observed, and it is likely that some uncharacterized gene is causing this phenotype. Collectively, our data suggest that *PA2989* serve as a cyclic di-GMP adaptor protein that activates the protective mechanism of *P. aeruginosa* when the level of the local cyclic di-GMP pool is high.

4.3.5 FlgZ regulates flagellar rotation speed and torque in PAO1

Our phenotypic assays indicate that FlgZ affects both swarming and swimming motility but its greatest effect was seen on swimming motility. This indicates that FlgZ may function like YcgR in PAO1 cells. However, PAO1 possess two sets of stator motors which contribute differently to the swimming and swarming motility¹⁵⁶. Furthermore, FlgZ was found in a different isofunctional cluster from YcgR in our SSN (Chapter 2). We therefore postulated that FlgZ may regulate motility in PAO1 in a different manner.

To validate our hypothesis, we first sought to determine if FlgZ interacts with the reported partners of YcgR by performing bacteria two-hybrid screen of FlgZ against FliG, FliM and MotA (Figure S26). The two-hybrid screen confirms that FlgZ does not function by interacting with FliG, FliM or MotA of the flagella complex. We also

concurrently analyzed the flagella rotation and output in FlgZ mutants by cell-tethering assays and swimming speed analysis^c. PAO1 and *FlgZ* mutant strains were tethered onto a fixed surface using flagellar antibodies, when the flagellar motor generates a rotational force, the cell bodies rotate (Figure 44B) and this rotation was recorded with a camera to determine the rotational behavior of the strains. No rotational bias was observed when *FlgZ* was deleted or overexpressed (Figure 44A). Unexpectedly, we observed a drop in the flagella rotation speeds when *FlgZ* was deleted and a slight increase when it was overexpressed (Figure 44C). Our observations at this point confirmed our postulations that FlgZ regulates motility differently from YcgR.

It has been reported that the MotAB stators are able to drive faster swimming speeds than the MotCD stators while the MotCD stators are required for swarming motility and swimming under load^{156,157}. By incorporating these observations with our experimental data, we were able to formulate a hypothesis on the mechanism of flagellar output regulation by FlgZ. We propose that FlgZ regulates the torque output of the PAO1 flagellar motor by biasing stator selection towards the MotAB stators.

Since the speed of rotation of the flagellar motor is inversely proportional to the torque generated^{158,159}, it is possible that the MotAB stators have evolved to generate more speed over torque while the MotCD stators generate torque over speed. Thus, when the flagellar motor consists of MotAB stators, it is able to rotate at a higher speed, as observed in the case of the *FlgZ* overexpression strains. Conversely, a flagellar motor consisting of MotCD stators would rotate slower but generate more torque and therefore would be able to drive swarming motility. By binding to MotC, FlgZ allows greater incorporation of the MotAB stators into the flagellar motor, which explains the increased rotation speed observed in the *FlgZ*⁺ strain (Figure 44C). Similarly, increased MotCD incorporation into the flagellar motor results in the reduced rotation speed observed in the Δ *FlgZ* strain.

^c Cell tethering assays and swimming speed analysis were conducted in collaboration with BII, ASTAR.

Figure 44. The flagellar motor of $\Delta PA3353$ mutants rotate at a slower speed than wild type PAO1. (A) Rotational bias of $FlgZ$ mutant cells as determined by cell-tethering assay; $FlgZ$ does not cause any rotational bias in PAO1. (B) A cartoon depicting flagella tethering to study rotation of the cell body (C) Rotation speed of $FlgZ$ mutants. Deletion of $FlgZ$ causes an unexpected drop in rotation speeds. (D) Swimming speed of $FlgZ$ deletion mutant under conditions of increasing load (viscosity).

This hypothesis would also explain the trend observed in Figure 44D, in low viscosity, wild type cells will likely incorporate more MotAB stators and are able to swim faster while the $\Delta FlgZ$ mutant will incorporate more MotCD stators and swim slower. As viscosity increases, the difference in swimming speed reduces as the wild type cells switch to utilizing MotCD stators to increase torque generation. This difference between MotCD/AB utilization may also explain the observations of the swimming motility assay (Figure 38B). Even if soft agar was used, it presents a more viscous environment compared to liquids, hence $\Delta FlgZ$ mutants which generate more torque are able to cover more area in the soft agar compared to PAO1 or $FlgZ^+$ strains.

4.3.6 PA2989 represents a class of structurally novel di-domain PilZ proteins with an inserted α -helical domain

The bioinformatic analysis carried out in chapter 2 led us to realize that the cyclic di-GMP binding motifs in PA2989 were different from other PilZ proteins (Figure 22). Typically, the RxxxR and DxSxxG motifs in PilZ proteins are separated by 20-30

residues but the two motifs in *PA2989* are separated by 89 amino acids instead (Table 3). In addition, the bioinformatic analysis also indicates that rather than an DxSxxG motif, the second motif is SEGG in *PA2989*. Importantly *PA2989* also has the highest affinity ($K_D = 0.288 \mu\text{M}$) for cyclic di-GMP amongst the PilZ proteins in PAO1⁹³ which suggests that these changes did not impair its cyclic di-GMP binding ability. We hypothesized that these deviations in *PA2989* would require an altered cyclic di-GMP binding pocket that contributes to the high affinity to cyclic di-GMP. As there is no crystal structure for *PA2989* homologs, we used the I-TASSER software suite⁹⁶ to predict the 3D structure of full length *PA2989* protein (Figure 45A).

Table 3. Di-domain PilZ proteins and number of residues between their RxxxR and DxSxxG motifs.

PilZ	Length	Residues between RxxxR and DxSxxG
PA2989	254	89
FlgZ	263	27
PP4397	247	29
PlzD	252	21
YcgR	244	26
MrkH	234	26

The predicted structure of *PA2989* consists of PilZ β -barrel core, three helices and several unstructured coils (Figure 45). The 89-residue region that spans the two cyclic di-GMP binding motifs was predicted to form helices and part of the PilZ β -barrel structure (Figure 45). Our bioinformatic analysis in chapter 2 also suggests that *PA2989* binds to dimeric cyclic di-GMP. Thus, we chose PP4397 which is a di-domain PilZ protein that binds to intercalated cyclic di-GMP to model the cyclic di-GMP binding mode of the predicted *PA2989* structure (Figure 46).

Our approximated cyclic di-GMP binding model of *PA2989* offers a clue to why *PA2989* has a high affinity to cyclic di-GMP (Figure 46D). The short $\alpha 1$ helix immediately before the REYYR motif contains aspartic acid residues which can interact with the nitrogen atoms on the cyclic di-GMP molecule. In addition, there is a stretch of basic amino acids (residues 45-52, Figure 45B) that are located on the random coil region that may also interact with cyclic di-GMP. Finally, PilZ proteins

typically undergo a fairly large conformational change upon binding to cyclic di-GMP^{33,35,160} and so it is entirely possible that the random coil region can change conformation to enhance the interaction between the protein and cyclic di-GMP further.

Figure 45. Predicted 3D structure of PA2989. (A) The structural model of PA2989. The inserted helices ($\alpha 2$ and $\alpha 3$) are colored in red. (B) The protein sequence of PA2989 and its predicted secondary structure. Residues within the inserted helices are depicted with red font.

4.4 Discussion

Collectively, my experimental data suggest that PA2989 plays a role in the protective cellular functions of PAO1 by regulating gene expression. Many of the 109 genes that showed a fold change of ≥ 3.5 were hypothetical proteins with unknown functions. Among the annotated genes, six genes that play putatively protective roles were down-regulated and *naIC* which suppresses the MexAB-OprM efflux pump was up-regulated. I also found that *PA2989* mediates swarming motility, presumably by modulating *rsmZ* levels and the signaling pathways under its control.

Figure 46. Model of PA2989 in complex with cyclic di-GMP. (A) Overlay of the cyclic di-GMP-binding pocket of PP4397 onto PA2989. PA2989 is colored beige and PP4397 is colored light purple. The RxxxR motifs of PP4397 and PA2989 are colored blue and green respectively while the DxSxxG motifs are colored yellow and cyan respectively. Cyclic di-GMP shown is bound to PP4397. (B) Ribbon-structure of PA2989 in complex with cyclic di-GMP. (C) Space-filling model of PA2989 bound to cyclic di-GMP (D) The modeled binding pocket of PA2989 in complex with cyclic di-GMP (shown as lines), possible hydrogen bonds or electrostatic interactions between residues and cyclic di-GMP are indicated with blue dotted lines.

Furthermore, I discovered that PA2989 is likely to represent a unique class of di-domain PilZ proteins that contain an inserted α -helical domain. The two cyclic di-GMP-binding motifs are separated by more than 80 residues in sequence but closer in space. The binding mode of cyclic di-GMP is likely to be rather different from that in other di-domain PilZ proteins.

The *flgZ* gene is part of the flagellar gene cluster, and thus, FlgZ is likely to play a role in regulating flagellar function. While my research was in progress, O'toole and co-workers published their work to show that FlgZ interacts with MotC to suppress swarming but not swimming motility⁴⁹. My results show that the Δ *FlgZ* mutant exhibited increased swimming motility when compared to wild type PAO1. Most unexpectedly, the cell tethering experiments showed that the flagellar motor rotation speed was reduced when *FlgZ* was deleted. To rationalize our experimental observations in light of the proposed interaction between FlgZ and MotC, we propose that the MotAB stators are able to drive the rotation of the flagellum at high speeds under conditions of low load but are unable to generate sufficient torque to propel the bacterium through viscous medium (high load) whereas the MotCD stators can generate higher torque but cannot rotate the flagella at high speed. Switching or mixing between these two sets of stators likely allows the bacteria to quickly adapt its motility machinery for movement through mediums of various viscosity, in addition to suppressing swarming motility during conditions of high cyclic di-GMP.

4.5 Conclusion

Here I describe my research to uncover the cellular functions of the di-domain PilZ proteins PA2989 and FlgZ. Although a full disclosure of the biological function of the two proteins remain to be achieved, the studies represent a substantial advance towards that goal. The phenotypic and gene expression studies have yielded important clues about the function of the two proteins. The hypotheses generated based on the current results will guide future research work and experimental design to fully unveil the function of the two *P. aeruginosa* proteins and their functional homologs in other bacterial species.

Chapter 5 – Conclusion and Future directions

This dissertation, the results of a large-scale sequence similarity networking analysis for single and di-domain PilZ proteins and experimental analysis of uncharacterized PilZ proteins in *P. aeruginosa* PAO1 are presented. The SSNs in Chapter 2 show that these proteins have undergone extensive divergent evolution and have diverse ligand-binding properties and cellular functions. The SSNs provide a panoramic perspective on the sequence-structure-function relationships in PilZ proteins and reveal many putative isofunctional clusters that are unexplored. Several unique groups of PilZ proteins were also uncovered through the SSNs and would be worth further studying. Groups of PilZ proteins which do not bind cyclic di-GMP were identified in both single and di-domain PilZ proteins (Figures 15 and 24). The SSNs also enabled the discovery of two new groups of PilZ proteins which differ from the canonical PilZ proteins and would merit further research into their biochemical properties. Cluster 6 from the single-domain PilZ SSN is comprised of PilZ proteins from *Bacillales* bacteria that contain a phenylalanine residue inserted into the cyclic di-GMP-binding RxxxR motif (Figure 13); because none of the PilZ proteins in this cluster have been studied, it is not known if the PilZ proteins retain their cyclic di-GMP-binding abilities. Biochemical studies to ascertain if these PilZ proteins binds cyclic di-GMP can be performed. And if these proteins do bind cyclic di-GMP, they may possibly utilize a modified ligand binding motif and structure elucidation may be worthwhile.

PA2989 belongs to a cluster of di-domain PilZ proteins that have a fold that differs from the canonical PilZ protein fold (Figure 22). Compared to canonical PilZ proteins which have between 20-30 residues between the two cyclic di-GMP binding motifs, there are 96 residues between the two putative cyclic di-GMP-binding motifs in all proteins from this cluster. Since PA2989 has been shown to have a high affinity to cyclic di-GMP, these proteins likely form the cyclic di-GMP-binding pocket using slightly different binding motifs. Future work regarding PA2989 will be discussed in detail further on in this chapter.

My work on the single-domain PilZ proteins, PA0012 and PA4324 was discussed in Chapter 3. Through a combination of phenotypical and genetic analysis, PA0012 was discovered to regulate the gene expression of T6SS in PAO1. In the $\Delta PA0012$ strain of the bacteria, the HSI-I gene cluster of the T6SS was up-regulated several folds. Consequently, a T6SS bacterial competition assay confirmed that this increased gene expression correlates with increased killing of prey bacteria. Since the non-coding RNA, *rsmZ*, was also up-regulated when *PA0012* is deleted, it is likely that PA0012 executes its regulatory functions through *rsmZ*. The work on PA0012 represents the first steps towards characterizing its biological functions and the mechanisms underlying them. The experiments indicate that PA0012 does specifically inhibit the HSI-I genes, however it was not established if this function is regulated by cyclic di-GMP concentration in the cell. To test this hypothesis, a mutant form of PA0012 protein, PA0012^{R5A}, which does not bind cyclic di-GMP can be used for phenotypical analysis in the experiments described in Chapter 3. Failure of the exogenous PA0012^{R5A}-expressing plasmid to complement $\Delta PA0012$ strains or suppress bacteria killing in wild type strains would indicate the requirement of cyclic di-GMP in activating PA0012.

The single-domain PilZ protein, PA4324 which was also covered in Chapter 3 regulates gene expression in PAO1 as well. In the PAO1 strain where *PA4324* was deleted, the expression levels of the *pqs* operon increased an average of 25-fold (Figure 34C) and was consistently up-regulated over a period of 45 hours (Figure 35D) when compared to the wild type strain. PQS and pyocyanin production by the $\Delta PA4324$ strain was correspondingly increased as a result. Several questions remain to be answered. Similar to PA0012, the involvement of cyclic di-GMP in regulating the function of PA4324 needs to be established by using non-cyclic di-GMP-binding mutants of PA0012. Also, as discussed in Chapter 3, the discrepancy between the gene expression levels and PQS and pyocyanin levels may be due to substrate depletion (Figure 37) or regulation at the translational level. Feeding experiments can be used to determine if substrate depletion is indeed the cause of the discrepancy between gene expression levels and PQS/pyocyanin levels. Supplementing cultures with anthranilic acid or PCA would bypass the need for chorismic acid for PQS and

pyocyanin production respectively and allow the quantities to be indicative of gene/protein expression levels. To measure the protein levels of the *pqs* and *phz* gene products, these genes can be replaced at their native chromosomal loci with recombinant genes linked to epitope tags. Protein techniques like western blotting can then be used to quantify protein levels.

Protein pulldown experiments using HA-tagged PA0012 and PA4324 did not yield positive results. We suspect that PA0012 and PA4324 may interact with nucleic acids instead. To verify this hypothesis, chromatin immunoprecipitation or ribonucleoprotein immunoprecipitation¹⁶¹ can be used to isolate DNA or RNA that may interact with these PilZ proteins respectively.

Chapter 4 discusses the work to elucidate the function of the di-domain PilZ proteins PA2989 and FlgZ. Based on our results, we proposed that FlgZ regulates the torque output of the bacterial flagellar in response to cyclic di-GMP by biasing stator selection towards the MotAB stators which generates higher speeds but lower torque. Since the MotAB stators are unable to support motility under load, FlgZ effectively inhibits swarming motility through this mechanism. However, it is likely that there is additional intricacy in this mode of regulation. Rather than an all-or-nothing mode of regulation, the cell may possibly use FlgZ to fine-tune the speed-torque output of its flagellar motor in order to navigate environments with varying viscosity.

Bioinformatic analysis revealed that PA2989 may be a unique di-domain PilZ protein whose cyclic di-GMP-binding motifs are separated by 89 amino acids instead of 20-30 amino acids as per other PilZ proteins. Structural modeling of the full-length PA2989 reveals that the predicted cyclic di-GMP-binding motifs of REYYR and SEGG does indeed form the ligand-binding pocket with inserted α -helical domains. This suggests that PA2989 represents a hitherto unknown di-domain PilZ protein with a different fold and possibly different ligand binding mode. Structural elucidation of this protein would therefore be an important follow-up to the studies in Chapter 4. In addition, mutagenesis experiments should be conducted to confirm the predicted cyclic di-GMP-binding residues. Finally, once the cyclic di-GMP-binding residues have been determined, the *PA2989* gene can be replaced at its native chromosomal locus with a

non-binding mutant *PA2989* gene to generate a functionally *PA2989*-null mutant. This mutant strain can be subjected to phenotypical and gene expression analysis to elucidate the biological functions of *PA2989*.

The sequence similarity networks provide an unprecedented global view on the PilZ superfamily. It also highlights our lack of knowledge on this family of proteins as many clusters in the networks are still unannotated and the biological functions of many PilZ proteins are unknown. The work described in this thesis brings us one step closer towards fully elucidating the functions of PilZ proteins in the model organism *P. aeruginosa* PAO1, which will significantly expand our knowledge about this superfamily of proteins that regulates many cyclic di-GMP-dependent and independent pathways.

References

- 1 Ross, P. *et al.* Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279-281 (1987).
- 2 Amikam, D. & Benziman, M. Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. *J Bacteriol* **171**, 6649-6655 (1989).
- 3 Tal, R. *et al.* 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 (1998).
- 4 Romling, U., Galperin, M. Y. & Gomelsky, M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77**, 1-52, doi:10.1128/MMBR.00043-12 (2013).
- 5 Tschowri, N. *et al.* Tetrameric c-di-GMP mediates effective transcription factor dimerization to control *Streptomyces* development. *Cell* **158**, 1136-1147, doi:10.1016/j.cell.2014.07.022 (2014).
- 6 Lori, C. *et al.* Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* **523**, 236-239, doi:10.1038/nature14473 (2015).
- 7 Chan, C. *et al.* Structural basis of activity and allosteric control of diguanylate cyclase. *Proc Natl Acad Sci U S A* **101**, 17084-17089, doi:10.1073/pnas.0406134101 (2004).
- 8 Wassmann, P. *et al.* Structure of BeF₃⁻-modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure* **15**, 915-927, doi:10.1016/j.str.2007.06.016 (2007).
- 9 De, N., Navarro, M. V., Raghavan, R. V. & Sondermann, H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J Mol Biol* **393**, 619-633, doi:10.1016/j.jmb.2009.08.030 (2009).
- 10 Seshasayee, A. S., Fraser, G. M. & Luscombe, N. M. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Res* **38**, 5970-5981, doi:10.1093/nar/gkq382 (2010).
- 11 Christen, B. *et al.* Allosteric control of cyclic di-GMP signaling. *J Biol Chem* **281**, 32015-32024, doi:10.1074/jbc.M603589200 (2006).
- 12 Yang, C. Y. *et al.* The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)₂ at the active site. *Acta Crystallogr D Biol Crystallogr* **67**, 997-1008, doi:10.1107/S090744491104039X (2011).
- 13 Zahringer, F., Lacanna, E., Jenal, U., Schirmer, T. & Boehm, A. Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure* **21**, 1149-1157, doi:10.1016/j.str.2013.04.026 (2013).
- 14 Tuckerman, J. R. *et al.* An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control. *Biochemistry* **48**, 9764-9774, doi:10.1021/bi901409g (2009).
- 15 Mills, E., Petersen, E., Kulasekara, B. R. & Miller, S. I. A direct screen for c-di-GMP modulators reveals a *Salmonella Typhimurium* periplasmic L-arginine-sensing pathway. *Sci Signal* **8**, ra57, doi:10.1126/scisignal.aaa1796 (2015).
- 16 O'Connor, J. R., Kuwada, N. J., Huangyutitham, V., Wiggins, P. A. & Harwood, C. S. Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol Microbiol* **86**, 720-729, doi:10.1111/mmi.12013 (2012).
- 17 Schmidt, A. J., Ryjenkov, D. A. & Gomelsky, M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *J Bacteriol* **187**, 4774-4781, doi:10.1128/JB.187.14.4774-4781.2005 (2005).

- 18 Tamayo, R., Tischler, A. D. & Camilli, A. The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *J Biol Chem* **280**, 33324-33330, doi:10.1074/jbc.M506500200 (2005).
- 19 Christen, M., Christen, B., Folcher, M., Schauerte, A. & Jenal, U. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J Biol Chem* **280**, 30829-30837, doi:10.1074/jbc.M504429200 (2005).
- 20 Rao, F., Yang, Y., Qi, Y. & Liang, Z. X. Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*. *J Bacteriol* **190**, 3622-3631, doi:10.1128/JB.00165-08 (2008).
- 21 Paul, R. *et al.* Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev* **18**, 715-727, doi:10.1101/gad.289504 (2004).
- 22 Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol Microbiol* **53**, 1123-1134, doi:10.1111/j.1365-2958.2004.04206.x (2004).
- 23 Tarutina, M., Ryjenkov, D. A. & Gomelsky, M. An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. *J Biol Chem* **281**, 34751-34758, doi:10.1074/jbc.M604819200 (2006).
- 24 Tchigvintsev, A. *et al.* Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. *J Mol Biol* **402**, 524-538, doi:10.1016/j.jmb.2010.07.050 (2010).
- 25 Barends, T. R. *et al.* Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* **459**, 1015-1018, doi:10.1038/nature07966 (2009).
- 26 Gomelsky, M. & Klug, G. BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. *Trends Biochem Sci* **27**, 497-500 (2002).
- 27 Galperin, M. Y., Natale, D. A., Aravind, L. & Koonin, E. V. A specialized version of the HD hydrolase domain implicated in signal transduction. *J Mol Microbiol Biotechnol* **1**, 303-305 (1999).
- 28 Ausmees, N. *et al.* Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiol Lett* **204**, 163-167 (2001).
- 29 Ryan, R. P. *et al.* Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc Natl Acad Sci U S A* **103**, 6712-6717, doi:10.1073/pnas.0600345103 (2006).
- 30 Bellini, D. *et al.* Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Mol Microbiol* **91**, 26-38, doi:10.1111/mmi.12447 (2014).
- 31 *Census of cyclic di-GMP binding receptors*, <https://www.ncbi.nlm.nih.gov/Complete_Genomes/c-di-GMP.html> (2016).
- 32 Habazettl, J., Allan, M. G., Jenal, U. & Grzesiek, S. Solution structure of the PilZ domain protein PA4608 complex with cyclic di-GMP identifies charge clustering as molecular readout. *J Biol Chem* **286**, 14304-14314, doi:10.1074/jbc.M110.209007 (2011).
- 33 Benach, J. *et al.* The structural basis of cyclic diguanylate signal transduction by PilZ domains. *EMBO J* **26**, 5153-5166, doi:10.1038/sj.emboj.7601918 (2007).
- 34 Chou, S. H. & Galperin, M. Y. Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. *J Bacteriol* **198**, 32-46, doi:10.1128/JB.00333-15 (2016).
- 35 Ko, J. *et al.* Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by Pilz domain proteins. *J Mol Biol* **398**, 97-110, doi:10.1016/j.jmb.2010.03.007 (2010).

- 36 Xu, L. *et al.* A cyclic di-GMP-binding adaptor protein interacts with a chemotaxis methyltransferase to control flagellar motor switching. *Sci Signal* **9**, ra102, doi:10.1126/scisignal.aaf7584 (2016).
- 37 Xu, L. *et al.* A Cyclic di-GMP-binding Adaptor Protein Interacts with Histidine Kinase to Regulate Two-component Signaling. *J Biol Chem* **291**, 16112-16123, doi:10.1074/jbc.M116.730887 (2016).
- 38 Yang, F. *et al.* The *Xanthomonas oryzae* pv. *oryzae* PilZ Domain Proteins Function Differentially in Cyclic di-GMP Binding and Regulation of Virulence and Motility. *Appl Environ Microbiol* **81**, 4358-4367, doi:10.1128/AEM.04044-14 (2015).
- 39 Chin, K. H. *et al.* Structural polymorphism of c-di-GMP bound to an EAL domain and in complex with a type II PilZ-domain protein. *Acta Crystallogr D Biol Crystallogr* **68**, 1380-1392, doi:10.1107/S0907444912030594 (2012).
- 40 Alm, R. A., Boder, A. J., Free, P. D. & Mattick, J. S. Identification of a novel gene, pilZ, essential for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. *J Bacteriol* **178**, 46-53 (1996).
- 41 Jain, R., Behrens, A. J., Kaeber, V. & Kazmierczak, B. I. Type IV pilus assembly in *Pseudomonas aeruginosa* over a broad range of cyclic di-GMP concentrations. *J Bacteriol* **194**, 4285-4294, doi:10.1128/JB.00803-12 (2012).
- 42 Qi, Y. *et al.* Functional divergence of FimX in PilZ binding and type IV pilus regulation. *J Bacteriol* **194**, 5922-5931, doi:10.1128/JB.00767-12 (2012).
- 43 Guzzo, C. R., Salinas, R. K., Andrade, M. O. & Farah, C. S. PILZ protein structure and interactions with PILB and the FIMX EAL domain: implications for control of type IV pilus biogenesis. *J Mol Biol* **393**, 848-866, doi:10.1016/j.jmb.2009.07.065 (2009).
- 44 Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y. & Lory, S. 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, doi:10.1111/j.1365-2958.2007.05817.x (2007).
- 45 Petrova, O. E. & Sauer, K. SagS contributes to the motile-sessile switch and acts in concert with BfiSR to enable *Pseudomonas aeruginosa* biofilm formation. *J Bacteriol* **193**, 6614-6628, doi:10.1128/JB.00305-11 (2011).
- 46 Christen, M. *et al.* 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 U S A* **104**, 4112-4117, doi:10.1073/pnas.0607738104 (2007).
- 47 Pratt, J. T., Tamayo, R., Tischler, A. D. & Camilli, A. PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in *Vibrio cholerae*. *J Biol Chem* **282**, 12860-12870, doi:10.1074/jbc.M611593200 (2007).
- 48 Yang, F. *et al.* The degenerate EAL-GGDEF domain protein Filp functions as a cyclic di-GMP receptor and specifically interacts with the PilZ-domain protein PXO_02715 to regulate virulence in *Xanthomonas oryzae* pv. *oryzae*. *Mol Plant Microbe Interact* **27**, 578-589, doi:10.1094/MPMI-12-13-0371-R (2014).
- 49 Baker, A. E. *et al.* PilZ Domain Protein FlgZ Mediates Cyclic Di-GMP-Dependent Swarming Motility Control in *Pseudomonas aeruginosa*. *J Bacteriol* **198**, 1837-1846, doi:10.1128/JB.00196-16 (2016).
- 50 Chen, Y., Chai, Y., Guo, J. H. & Losick, R. Evidence for cyclic Di-GMP-mediated signaling in *Bacillus subtilis*. *J Bacteriol* **194**, 5080-5090, doi:10.1128/JB.01092-12 (2012).
- 51 Paul, K., Nieto, V., Carlquist, W. C., Blair, D. F. & Harshey, R. M. 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, doi:10.1016/j.molcel.2010.03.001 (2010).

- 52 Boehm, A. *et al.* Second messenger-mediated adjustment of bacterial swimming velocity. *Cell* **141**, 107-116, doi:10.1016/j.cell.2010.01.018 (2010).
- 53 Wilksch, J. J. *et al.* MrkH, a novel c-di-GMP-dependent transcriptional activator, controls *Klebsiella pneumoniae* biofilm formation by regulating type 3 fimbriae expression. *PLoS Pathog* **7**, e1002204, doi:10.1371/journal.ppat.1002204 (2011).
- 54 Yang, J. *et al.* Transcriptional activation of the mrkA promoter of the *Klebsiella pneumoniae* type 3 fimbrial operon by the c-di-GMP-dependent MrkH protein. *PLoS One* **8**, e79038, doi:10.1371/journal.pone.0079038 (2013).
- 55 Schumacher, M. A. & Zeng, W. Structures of the activator of *K. pneumoniae* biofilm formation, MrkH, indicates PilZ domains involved in c-di-GMP and DNA binding. *Proc Natl Acad Sci U S A* **113**, 10067-10072, doi:10.1073/pnas.1607503113 (2016).
- 56 Freedman, J. C. *et al.* Identification and molecular characterization of a cyclic-di-GMP effector protein, PlzA (BB0733): additional evidence for the existence of a functional cyclic-di-GMP regulatory network in the Lyme disease spirochete, *Borrelia burgdorferi*. *FEMS Immunol Med Microbiol* **58**, 285-294, doi:10.1111/j.1574-695X.2009.00635.x (2010).
- 57 He, M., Zhang, J. J., Ye, M., Lou, Y. & Yang, X. F. Cyclic Di-GMP receptor PlzA controls virulence gene expression through RpoS in *Borrelia burgdorferi*. *Infect Immun* **82**, 445-452, doi:10.1128/IAI.01238-13 (2014).
- 58 Pitzer, J. E. *et al.* Analysis of the *Borrelia burgdorferi* cyclic-di-GMP-binding protein PlzA reveals a role in motility and virulence. *Infect Immun* **79**, 1815-1825, doi:10.1128/IAI.00075-11 (2011).
- 59 Oglesby, L. L., Jain, S. & Ohman, D. E. Membrane topology and roles of *Pseudomonas aeruginosa* Alg8 and Alg44 in alginate polymerization. *Microbiology* **154**, 1605-1615, doi:10.1099/mic.0.2007/015305-0 (2008).
- 60 Whitney, J. C. *et al.* Dimeric c-di-GMP is required for post-translational regulation of alginate production in *Pseudomonas aeruginosa*. *J Biol Chem* **290**, 12451-12462, doi:10.1074/jbc.M115.645051 (2015).
- 61 Morgan, J. L., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181-186, doi:10.1038/nature11744 (2013).
- 62 Zorraquino, V. *et al.* Coordinated cyclic-di-GMP repression of *Salmonella* motility through YcgR and cellulose. *J Bacteriol* **195**, 417-428, doi:10.1128/JB.01789-12 (2013).
- 63 Bian, J., Liu, X., Cheng, Y. Q. & Li, C. Inactivation of cyclic Di-GMP binding protein TDE0214 affects the motility, biofilm formation, and virulence of *Treponema denticola*. *J Bacteriol* **195**, 3897-3905, doi:10.1128/JB.00610-13 (2013).
- 64 Russell, M. H. *et al.* Integration of the second messenger c-di-GMP into the chemotactic signaling pathway. *MBio* **4**, e00001-00013, doi:10.1128/mBio.00001-13 (2013).
- 65 Huangyutitham, V., Guvener, Z. T. & Harwood, C. S. Subcellular clustering of the phosphorylated WspR response regulator protein stimulates its diguanylate cyclase activity. *MBio* **4**, e00242-00213, doi:10.1128/mBio.00242-13 (2013).
- 66 Merritt, J. H. *et al.* Specific control of *Pseudomonas aeruginosa* surface-associated behaviors by two c-di-GMP diguanylate cyclases. *MBio* **1**, doi:10.1128/mBio.00183-10 (2010).
- 67 Irie, Y. *et al.* Self-produced exopolysaccharide is a signal that stimulates biofilm formation in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* **109**, 20632-20636, doi:10.1073/pnas.1217993109 (2012).
- 68 Basu Roy, A. & Sauer, K. Diguanylate cyclase NicD-based signalling mechanism of nutrient-induced dispersion by *Pseudomonas aeruginosa*. *Mol Microbiol* **94**, 771-793, doi:10.1111/mmi.12802 (2014).

- 69 Petrova, O. E., Cherny, K. E. & Sauer, K. The diguanylate cyclase GcbA facilitates *Pseudomonas aeruginosa* biofilm dispersion by activating BdlA. *J Bacteriol* **197**, 174-187, doi:10.1128/JB.02244-14 (2015).
- 70 Li, Y., Heine, S., Entian, M., Sauer, K. & Frankenberg-Dinkel, N. NO-induced biofilm dispersion in *Pseudomonas aeruginosa* is mediated by an MHYT domain-coupled phosphodiesterase. *J Bacteriol* **195**, 3531-3542, doi:10.1128/JB.01156-12 (2013).
- 71 Lee, V. T. *et al.* A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. *Mol Microbiol* **65**, 1474-1484, doi:10.1111/j.1365-2958.2007.05879.x (2007).
- 72 Hengge, R., Grundling, A., Jenal, U., Ryan, R. & Yildiz, F. Bacterial Signal Transduction by Cyclic Di-GMP and Other Nucleotide Second Messengers. *J Bacteriol* **198**, 15-26, doi:10.1128/JB.00331-15 (2016).
- 73 Hickman, J. W. & Harwood, C. S. Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. *Mol Microbiol* **69**, 376-389, doi:10.1111/j.1365-2958.2008.06281.x (2008).
- 74 Chambers, J. R., Liao, J., Schurr, M. J. & Sauer, K. BrlR from *Pseudomonas aeruginosa* is a c-di-GMP-responsive transcription factor. *Mol Microbiol* **92**, 471-487, doi:10.1111/mmi.12562 (2014).
- 75 Trampari, E. *et al.* Bacterial rotary export ATPases are allosterically regulated by the nucleotide second messenger cyclic-di-GMP. *J Biol Chem* **290**, 24470-24483, doi:10.1074/jbc.M115.661439 (2015).
- 76 Roelofs, K. G. *et al.* Systematic Identification of Cyclic-di-GMP Binding Proteins in *Vibrio cholerae* Reveals a Novel Class of Cyclic-di-GMP-Binding ATPases Associated with Type II Secretion Systems. *PLoS Pathog* **11**, e1005232, doi:10.1371/journal.ppat.1005232 (2015).
- 77 Wang, Y. C. *et al.* Nucleotide binding by the widespread high-affinity cyclic di-GMP receptor MshEN domain. *Nat Commun* **7**, 12481, doi:10.1038/ncomms12481 (2016).
- 78 Amikam, D. & Galperin, M. Y. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22**, 3-6 (2006).
- 79 Institute, E. B. PilZ domain (IPR009875), Interpro, <<https://www.ebi.ac.uk/interpro/entry/IPR009875?q=pilZ>> (
- 80 Sangar, V., Blankenberg, D. J., Altman, N. & Lesk, A. M. Quantitative sequence-function relationships in proteins based on gene ontology. *BMC Bioinformatics* **8**, 294, doi:10.1186/1471-2105-8-294 (2007).
- 81 Gerlt, J. A. *et al.* Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **1854**, 1019-1037, doi:10.1016/j.bbapap.2015.04.015 (2015).
- 82 Atkinson, H. J., Morris, J. H., Ferrin, T. E. & Babbitt, P. C. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* **4**, e4345, doi:10.1371/journal.pone.0004345 (2009).
- 83 Cline, M. S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols* **2**, 2366-2382, doi:http://www.nature.com/nprot/journal/v2/n10/supinfo/nprot.2007.324_S1.html (2007).
- 84 Peterhoff, D. *et al.* A comprehensive analysis of the geranylgeranylglycerol phosphate synthase enzyme family identifies novel members and reveals mechanisms of substrate specificity and quaternary structure organization. *Mol Microbiol* **92**, 885-899, doi:10.1111/mmi.12596 (2014).
- 85 Goble, A. M., Feng, Y., Raushel, F. M. & Cronan, J. E. Discovery of a cAMP deaminase that quenches cyclic AMP-dependent regulation. *ACS Chem Biol* **8**, 2622-2629, doi:10.1021/cb4004628 (2013).

- 86 Cummings, J. A. *et al.* Prospecting for unannotated enzymes: discovery of a 3',5'-nucleotide bisphosphate phosphatase within the amidohydrolase superfamily. *Biochemistry* **53**, 591-600, doi:10.1021/bi401640r (2014).
- 87 Dong, G. Q. *et al.* Prediction of substrates for glutathione transferases by covalent docking. *J Chem Inf Model* **54**, 1687-1699, doi:10.1021/ci5001554 (2014).
- 88 Tian, B. X. *et al.* Predicting the functions and specificity of triterpenoid synthases: a mechanism-based multi-intermediate docking approach. *PLoS Comput Biol* **10**, e1003874, doi:10.1371/journal.pcbi.1003874 (2014).
- 89 Li, T. N. *et al.* A novel tetrameric PilZ domain structure from xanthomonads. *PLoS One* **6**, e22036, doi:10.1371/journal.pone.0022036 (2011).
- 90 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).
- 91 Venkataramani, P. *Regulation of two-component signalling by a cyclic di-GMP-binding PilZ protein* PhD thesis, Nanyang Technological University, (2016).
- 92 Berthold, T. *et al.* Mycelia as a focal point for horizontal gene transfer among soil bacteria. *Sci Rep* **6**, 36390, doi:10.1038/srep36390 (2016).
- 93 Pultz, I. S. *et al.* The response threshold of Salmonella PilZ domain proteins is determined by their binding affinities for c-di-GMP. *Mol Microbiol* **86**, 1424-1440, doi:10.1111/mmi.12066 (2012).
- 94 Grady, E. N., MacDonald, J., Liu, L., Richman, A. & Yuan, Z. C. Current knowledge and perspectives of Paenibacillus: a review. *Microb Cell Fact* **15**, 203, doi:10.1186/s12934-016-0603-7 (2016).
- 95 Guzzo, C. R., Dunger, G., Salinas, R. K. & Farah, C. S. Structure of the PilZ-FimXEAL-c-di-GMP Complex Responsible for the Regulation of Bacterial Type IV Pilus Biogenesis. *J Mol Biol* **425**, 2174-2197, doi:10.1016/j.jmb.2013.03.021 (2013).
- 96 Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8, doi:10.1038/nmeth.3213 (2015).
- 97 Klockgether, J. *et al.* Genome diversity of Pseudomonas aeruginosa PAO1 laboratory strains. *J Bacteriol* **192**, 1113-1121, doi:10.1128/JB.01515-09 (2010).
- 98 Quick, J. *et al.* Seeking the source of Pseudomonas aeruginosa infections in a recently opened hospital: an observational study using whole-genome sequencing. *BMJ Open* **4**, e006278, doi:10.1136/bmjopen-2014-006278 (2014).
- 99 Poole, K. Pseudomonas aeruginosa: resistance to the max. *Front Microbiol* **2**, 65, doi:10.3389/fmicb.2011.00065 (2011).
- 100 Qiu, L. *et al.* Investigating the failure of repeated standard cleaning and disinfection of a Pseudomonas aeruginosa-infected pancreatic and biliary endoscope. *Am J Infect Control* **43**, e43-46, doi:10.1016/j.ajic.2015.04.204 (2015).
- 101 Kovaleva, J., Peters, F. T., van der Mei, H. C. & Degener, J. E. Transmission of infection by flexible gastrointestinal endoscopy and bronchoscopy. *Clin Microbiol Rev* **26**, 231-254, doi:10.1128/CMR.00085-12 (2013).
- 102 Blanc, D. S., Francioli, P. & Zanetti, G. Molecular Epidemiology of Pseudomonas aeruginosa in the Intensive Care Units - A Review. *Open Microbiol J* **1**, 8-11, doi:10.2174/1874285800701010008 (2007).
- 103 Lanini, S. *et al.* Molecular epidemiology of a Pseudomonas aeruginosa hospital outbreak driven by a contaminated disinfectant-soap dispenser. *PLoS One* **6**, e17064, doi:10.1371/journal.pone.0017064 (2011).
- 104 Suarez, C. *et al.* A large sustained endemic outbreak of multiresistant Pseudomonas aeruginosa: a new epidemiological scenario for nosocomial acquisition. *BMC Infect Dis* **11**, 272, doi:10.1186/1471-2334-11-272 (2011).

- 105 Suwantararat, N. & Carroll, K. C. Epidemiology and molecular characterization of multidrug-resistant Gram-negative bacteria in Southeast Asia. *Antimicrob Resist Infect Control* **5**, 15, doi:10.1186/s13756-016-0115-6 (2016).
- 106 Teo, J. W., La, M. V., Jureen, R. & Lin, R. T. Emergence of a New Delhi metallo-beta-lactamase-1-producing *Pseudomonas aeruginosa* in Singapore. *Emerg Microbes Infect* **4**, e72, doi:10.1038/emi.2015.72 (2015).
- 107 Singapore, M. o. H. Epidemiological News Bulletin. **43**, 17-21 (2017).
<https://www.moh.gov.sg/content/moh_web/home/Publications/epidemiological_news_bulletin/2017/january-2017-vol--43-no--1.html>.
- 108 Swan, J. S. *et al.* Elimination of biofilm and microbial contamination reservoirs in hospital washbasin U-bends by automated cleaning and disinfection with electrochemically activated solutions. *J Hosp Infect* **94**, 169-174, doi:10.1016/j.jhin.2016.07.007 (2016).
- 109 Kaplan, J. B. Antibiotic-induced biofilm formation. *Int J Artif Organs* **34**, 737-751, doi:10.5301/ijao.5000027 (2011).
- 110 Hoffman, L. R. *et al.* Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature* **436**, 1171-1175, doi:10.1038/nature03912 (2005).
- 111 Knobloch, J. K., Horstkotte, M. A., Rohde, H., Kaulfers, P. M. & Mack, D. Alcoholic ingredients in skin disinfectants increase biofilm expression of *Staphylococcus epidermidis*. *J Antimicrob Chemother* **49**, 683-687 (2002).
- 112 Simões, M., Simões, L. C. & Vieira, M. J. A review of current and emergent biofilm control strategies. *LWT - Food Science and Technology* **43**, 573-583, doi:<https://doi.org/10.1016/j.lwt.2009.12.008> (2010).
- 113 Hoiby, N., Bjarnsholt, T., Givskov, M., Molin, S. & Ciofu, O. Antibiotic resistance of bacterial biofilms. *Int J Antimicrob Agents* **35**, 322-332, doi:10.1016/j.ijantimicag.2009.12.011 (2010).
- 114 Wiegand, I., Hilpert, K. & Hancock, R. E. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3**, 163-175, doi:10.1038/nprot.2007.521 (2008).
- 115 Mavrodi, D. V. *et al.* Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **183**, 6454-6465, doi:10.1128/JB.183.21.6454-6465.2001 (2001).
- 116 Yang, L. *et al.* Effects of iron on DNA release and biofilm development by *Pseudomonas aeruginosa*. *Microbiology* **153**, 1318-1328, doi:10.1099/mic.0.2006/004911-0 (2007).
- 117 Hachani, A., Lossi, N. S. & Filloux, A. A visual assay to monitor T6SS-mediated bacterial competition. *J Vis Exp*, e50103, doi:10.3791/50103 (2013).
- 118 Turnbull, L. & Whitchurch, C. B. Motility assay: twitching motility. *Methods Mol Biol* **1149**, 73-86, doi:10.1007/978-1-4939-0473-0_9 (2014).
- 119 Friedman, L. & Kolter, R. Two genetic loci produce distinct carbohydrate-rich structural components of the *Pseudomonas aeruginosa* biofilm matrix. *J Bacteriol* **186**, 4457-4465, doi:10.1128/JB.186.14.4457-4465.2004 (2004).
- 120 Jenal, U., Reinders, A. & Lori, C. Cyclic di-GMP: second messenger extraordinaire. *Nat Rev Microbiol* **15**, 271-284, doi:10.1038/nrmicro.2016.190 (2017).
- 121 Remminghorst, U. & Rehm, B. H. Alg44, a unique protein required for alginate biosynthesis in *Pseudomonas aeruginosa*. *FEBS Lett* **580**, 3883-3888, doi:10.1016/j.febslet.2006.05.077 (2006).
- 122 Li, X.-Z. & Plésiat, P. in *Efflux-Mediated Antimicrobial Resistance in Bacteria: Mechanisms, Regulation and Clinical Implications* (eds Xian-Zhi Li, Christopher A. Elkins, & Helen I. Zgurskaya) 359-400 (Springer International Publishing, 2016).

- 123 Li, X. Z., Plesiat, P. & Nikaido, H. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin Microbiol Rev* **28**, 337-418, doi:10.1128/CMR.00117-14 (2015).
- 124 Nicastro, G. G. *et al.* Cyclic-di-GMP levels affect *Pseudomonas aeruginosa* fitness in the presence of imipenem. *Environ Microbiol* **16**, 1321-1333 (2014).
- 125 Chen, L., Zou, Y., She, P. & Wu, Y. Composition, function, and regulation of T6SS in *Pseudomonas aeruginosa*. *Microbiol Res* **172**, 19-25, doi:10.1016/j.micres.2015.01.004 (2015).
- 126 Hood, R. D. *et al.* A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host Microbe* **7**, 25-37, doi:10.1016/j.chom.2009.12.007 (2010).
- 127 Zoued, A. *et al.* TssK is a trimeric cytoplasmic protein interacting with components of both phage-like and membrane anchoring complexes of the type VI secretion system. *J Biol Chem* **288**, 27031-27041, doi:10.1074/jbc.M113.499772 (2013).
- 128 Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**, 4154-4159, doi:10.1073/pnas.0813360106 (2009).
- 129 Bonemann, G., Pietrosiuk, A. & Mogk, A. Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76**, 815-821, doi:10.1111/j.1365-2958.2010.07171.x (2010).
- 130 Schertzer, J. W., Brown, S. A. & Whiteley, M. Oxygen levels rapidly modulate *Pseudomonas aeruginosa* social behaviours via substrate limitation of PqsH. *Mol Microbiol* **77**, 1527-1538, doi:10.1111/j.1365-2958.2010.07303.x (2010).
- 131 Lee, J. & Zhang, L. The hierarchy quorum sensing network in *Pseudomonas aeruginosa*. *Protein Cell* **6**, 26-41, doi:10.1007/s13238-014-0100-x (2015).
- 132 Deziel, E. *et al.* The contribution of MvfR to *Pseudomonas aeruginosa* pathogenesis and quorum sensing circuitry regulation: multiple quorum sensing-regulated genes are modulated without affecting lasRI, rhIRI or the production of N-acyl-L-homoserine lactones. *Mol Microbiol* **55**, 998-1014, doi:10.1111/j.1365-2958.2004.04448.x (2005).
- 133 Recinos, D. A. *et al.* Redundant phenazine operons in *Pseudomonas aeruginosa* exhibit environment-dependent expression and differential roles in pathogenicity. *Proc Natl Acad Sci U S A* **109**, 19420-19425, doi:10.1073/pnas.1213901109 (2012).
- 134 Xiao, G. *et al.* MvfR, a key *Pseudomonas aeruginosa* pathogenicity LTTR-class regulatory protein, has dual ligands. *Mol Microbiol* **62**, 1689-1699, doi:10.1111/j.1365-2958.2006.05462.x (2006).
- 135 Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* **64**, 2240-2246 (1998).
- 136 Guo, Q. *et al.* PqsR-dependent and PqsR-independent regulation of motility and biofilm formation by PQS in *Pseudomonas aeruginosa* PAO1. *J Basic Microbiol* **54**, 633-643, doi:10.1002/jobm.201300091 (2014).
- 137 Ha, D. G. *et al.* 2-Heptyl-4-quinolone, a precursor of the *Pseudomonas* quinolone signal molecule, modulates swarming motility in *Pseudomonas aeruginosa*. *J Bacteriol* **193**, 6770-6780, doi:10.1128/JB.05929-11 (2011).
- 138 Deziel, E. *et al.* Analysis of *Pseudomonas aeruginosa* 4-hydroxy-2-alkylquinolines (HAQs) reveals a role for 4-hydroxy-2-heptylquinoline in cell-to-cell communication. *Proc Natl Acad Sci U S A* **101**, 1339-1344, doi:10.1073/pnas.0307694100 (2004).
- 139 Farrow, J. M., 3rd & Pesci, E. C. Two distinct pathways supply anthranilate as a precursor of the *Pseudomonas* quinolone signal. *J Bacteriol* **189**, 3425-3433, doi:10.1128/JB.00209-07 (2007).
- 140 Alvarez-Ortega, C., Wiegand, I., Olivares, J., Hancock, R. E. & Martinez, J. L. Genetic determinants involved in the susceptibility of *Pseudomonas aeruginosa* to beta-

- lactam antibiotics. *Antimicrob Agents Chemother* **54**, 4159-4167, doi:10.1128/AAC.00257-10 (2010).
- 141 Dotsch, A. *et al.* Genomewide identification of genetic determinants of antimicrobial drug resistance in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **53**, 2522-2531, doi:10.1128/AAC.00035-09 (2009).
- 142 Alicia, T. G., Yang, C., Wang, Z. & Nguyen, N. T. Combinational concentration gradient confinement through stagnation flow. *Lab Chip* **16**, 368-376, doi:10.1039/c5lc01137j (2016).
- 143 Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* **60**, 91-110, doi:10.1023/b:visi.0000029664.99615.94 (2004).
- 144 Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods* **5**, 695-702, doi:10.1038/nmeth.1237 (2008).
- 145 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
- 146 Masuda, N. *et al.* Substrate specificities of MexAB-OprM, MexCD-OprJ, and MexXY-oprM efflux pumps in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* **44**, 3322-3327 (2000).
- 147 Heurlier, K. *et al.* Positive control of swarming, rhamnolipid synthesis, and lipase production by the posttranscriptional RsmA/RsmZ system in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **186**, 2936-2945 (2004).
- 148 Valentini, M., Storelli, N. & Lapouge, K. Identification of C(4)-dicarboxylate transport systems in *Pseudomonas aeruginosa* PAO1. *J Bacteriol* **193**, 4307-4316, doi:10.1128/JB.05074-11 (2011).
- 149 Kim, J. J. & Sundin, G. W. Construction and analysis of photolyase mutants of *Pseudomonas aeruginosa* and *Pseudomonas syringae*: contribution of photoreactivation, nucleotide excision repair, and mutagenic DNA repair to cell survival and mutability following exposure to UV-B radiation. *Appl Environ Microbiol* **67**, 1405-1411, doi:10.1128/AEM.67.4.1405-1411.2001 (2001).
- 150 Lesniak, J., Barton, W. A. & Nikolov, D. B. Structural and functional features of the *Escherichia coli* hydroperoxide resistance protein OsmC. *Protein Sci* **12**, 2838-2843, doi:10.1110/ps.03375603 (2003).
- 151 Damron, F. H., Napper, J., Teter, M. A. & Yu, H. D. Lipotoxin F of *Pseudomonas aeruginosa* is an AlgU-dependent and alginate-independent outer membrane protein involved in resistance to oxidative stress and adhesion to A549 human lung epithelia. *Microbiology* **155**, 1028-1038, doi:10.1099/mic.0.025833-0 (2009).
- 152 Rodriguez-Rojas, A. & Blazquez, J. The *Pseudomonas aeruginosa* pfpl gene plays an antimutator role and provides general stress protection. *J Bacteriol* **191**, 844-850, doi:10.1128/JB.01081-08 (2009).
- 153 Son, M. S., Matthews, W. J., Jr., Kang, Y., Nguyen, D. T. & Hoang, T. T. In vivo evidence of *Pseudomonas aeruginosa* nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. *Infect Immun* **75**, 5313-5324, doi:10.1128/IAI.01807-06 (2007).
- 154 Hagins, J. M., Locy, R. & Silo-Suh, L. Isocitrate lyase supplies precursors for hydrogen cyanide production in a cystic fibrosis isolate of *Pseudomonas aeruginosa*. *J Bacteriol* **191**, 6335-6339, doi:10.1128/JB.00692-09 (2009).
- 155 Oliver, K. E. & Silo-Suh, L. Impact of D-amino acid dehydrogenase on virulence factor production by a *Pseudomonas aeruginosa*. *Can J Microbiol* **59**, 598-603, doi:10.1139/cjm-2013-0289 (2013).
- 156 Doyle, T. B., Hawkins, A. C. & McCarter, L. L. The complex flagellar torque generator of *Pseudomonas aeruginosa*. *J Bacteriol* **186**, 6341-6350, doi:10.1128/JB.186.19.6341-6350.2004 (2004).

- 157 Kuchma, S. L. *et al.* Cyclic di-GMP-mediated repression of swarming motility by
Pseudomonas aeruginosa PA14 requires the MotAB stator. *J Bacteriol* **197**, 420-430,
doi:10.1128/JB.02130-14 (2015).
- 158 Tipping, M. J., Delalez, N. J., Lim, R., Berry, R. M. & Armitage, J. P. Load-dependent
assembly of the bacterial flagellar motor. *MBio* **4**, doi:10.1128/mBio.00551-13 (2013).
- 159 Yuan, J., Fahrner, K. A., Turner, L. & Berg, H. C. Asymmetry in the clockwise and
counterclockwise rotation of the bacterial flagellar motor. *Proc Natl Acad Sci U S A*
107, 12846-12849, doi:10.1073/pnas.1007333107 (2010).
- 160 Shin, J. S., Ryu, K. S., Ko, J., Lee, A. & Choi, B. S. Structural characterization reveals
that a PilZ domain protein undergoes substantial conformational change upon
binding to cyclic dimeric guanosine monophosphate. *Protein Sci* **20**, 270-277,
doi:10.1002/pro.557 (2011).
- 161 Keene, J. D., Komisarow, J. M. & Friedersdorf, M. B. RIP-Chip: the isolation and
identification of mRNAs, microRNAs and protein components of ribonucleoprotein
complexes from cell extracts. *Nat Protoc* **1**, 302-307, doi:10.1038/nprot.2006.47
(2006).

Appendix

Single-domain PilZ SSN sequence logos and bacteria genera

Single-domain PilZ SSN cluster 1

Figure S1. Sequence logos for single-domain PilZ cluster 1. (A) cluster 1a and (B) cluster 1b. Residues identified to be important for SagS interaction are marked with blue arrows in (A). The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S1. List of bacteria genera in single-domain PilZ cluster 1

Cluster	Genus			
1a	<i>Pseudomonas</i>	<i>Saccharophagus</i>	<i>Thalassolituus</i>	<i>Oleiphilus</i>
	<i>Oceanobacter</i>	<i>Oleispira</i>	<i>Enterobacter</i>	<i>Acinetobacter</i>
	<i>Streptococcus</i>	<i>Reinekea</i>	<i>Teredinibacter</i>	<i>Stenotrophomonas</i>
	<i>Oleibacter</i>	<i>Cellvibrio</i>		
1b	<i>Marinobacter</i>	<i>Halorhodospira</i>	<i>Alishewanella</i>	<i>Piscirickettsia</i>
	<i>Sedimenticola</i>	<i>Rheinheimera</i>	<i>Cycloclasticus</i>	<i>Glaciecola</i>
	<i>Oleiphilus</i>	<i>Paraglaciecola</i>	<i>Aliiglaciecola</i>	<i>Lacimicrobium</i>
	<i>Alteromonas</i>	<i>Methylobacter</i>	<i>Catenovulum</i>	<i>Methylomonas</i>
	<i>Ectothiorhodospira</i>	<i>Aestuariaibacter</i>	<i>Thalassomonas</i>	<i>Thalassotalea</i>
	<i>Hahella</i>	<i>Simiduia</i>	<i>Thiorhodospira</i>	<i>Methylomicrobium</i>
	<i>Colwellia</i>	<i>Acidihalobacter</i>	<i>Thioalkalivibrio</i>	<i>Pseudoalteromonas</i>
	<i>Arsukibacterium</i>			

Single-domain PilZ SSN cluster 2

Figure S2. Sequence logos for single-domain PilZ cluster 2. (A) cluster 2a; (B) cluster 2b and (C) cluster 2c. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S2. List of bacteria genera in single-domain PilZ cluster 2

Cluster	Genus			
2a	<i>Bradyrhizobium</i>	<i>Oligotropha</i>	<i>Afipia</i>	<i>Roseibium</i>
	<i>Labrenzia</i>	<i>Pseudolabrys</i>	<i>Pannonibacter</i>	<i>Acinetobacter</i>
	<i>Methylobacterium</i>	<i>Rhodoplanes</i>	<i>Rhodopseudomonas</i>	<i>Streptomyces</i>
	<i>Nitrobacter</i>			
2b	<i>Rhizobium</i>	<i>Devosia</i>	<i>Hyphomicrobium</i>	<i>Labrenzia</i>
	<i>Hyphomonas</i>	<i>Pseudovibrio</i>	<i>Prosthecomicrobium</i>	<i>Pleomorphomonas</i>
	<i>Polymorphum</i>	<i>Stappia</i>	<i>Pannonibacter</i>	<i>Methylobacterium</i>
	<i>Asticcacaulis</i>	<i>Agrobacterium</i>	<i>Bradyrhizobium</i>	<i>Neorhizobium</i>
2c	<i>Rhodoplanes</i>	<i>Pleomorphomonas</i>	<i>Streptomyces</i>	<i>Thioclava</i>
	<i>Mesorhizobium</i>	<i>Aureimonas</i>	<i>Bradyrhizobium</i>	<i>Aurantimonas</i>
	<i>Tardiphaga</i>	<i>Methylobacterium</i>	<i>Fulvimarina</i>	<i>Bosea</i>
	<i>Nitrobacter</i>	<i>Rhodopseudomonas</i>	<i>Afipia</i>	

Single-domain PilZ cluster 3

Figure S3. Sequence logo for single-domain PilZ cluster 3. (A) cluster 3a and (B) cluster 3b. Residues suspected to be involved in interaction between MapZ and CheR1 are marked by blue arrows in (A). The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S3. List of bacteria genera in single-domain PilZ cluster 3

Cluster	Genus			
3a	<i>Marinobacter</i>	<i>Halorhodospira</i>	<i>Curvibacter</i>	<i>Sulfuricella</i>
	<i>Azoarcus</i>	<i>Ectothiorhodospira</i>	<i>Hylemonella</i>	<i>Colwellia</i>
	<i>Halioglobus</i>	<i>Stenotrophomonas</i>	<i>Catenovulum</i>	<i>Methylobacter</i>
	<i>Aeromonas</i>	<i>Oleispira</i>	<i>Alcanivorax</i>	<i>Pelomonas</i>
	<i>Shewanella</i>	<i>Aliivibrio</i>	<i>Thauera</i>	<i>Spirochaeta</i>
	<i>Psychromonas</i>	<i>Paucibacter</i>	<i>Thalassomonas</i>	<i>Moritella</i>
	<i>Neptunomonas</i>	<i>Enterovibrio</i>	<i>Acidihalobacter</i>	<i>Teredinibacter</i>
	<i>Pseudomonas</i>	<i>Cyanothece</i>	<i>Thiorhodospira</i>	<i>Methylomicrobium</i>
	<i>Rheinheimera</i>	<i>Arsukibacterium</i>	<i>Geobacter</i>	<i>Mizugakiibacter</i>
	<i>Dechloromonas</i>	<i>Marinobacterium</i>	<i>Acidovorax</i>	<i>Alkalilimnicola</i>
	<i>Neptuniibacter</i>	<i>Pseudohalaea</i>	<i>Lacimicrobium</i>	<i>Gallaecimonas</i>
	<i>Photobacterium</i>	<i>Alishewanella</i>	<i>Methylothermus</i>	<i>Enterobacter</i>
	<i>Vibrio</i>	<i>Anaeromyxobacter</i>	<i>Desulfuromonas</i>	<i>Acinetobacter</i>
	<i>Pseudoalteromonas</i>	<i>Methylophaga</i>	<i>Methylomonas</i>	<i>Saccharophagus</i>
	<i>Gynerella</i>	<i>Grimontia</i>	<i>Pseudohongiella</i>	<i>Oceanobacter</i>
	<i>Candidatus Brocadia</i>	<i>Cellvibrio</i>	<i>Thiorhodovibrio</i>	<i>Oleibacter</i>
	<i>Malonomonas</i>	<i>Sedimenticola</i>	<i>Oleiphilus</i>	<i>Ideonella</i>
	<i>Thalassolituus</i>	<i>Thioalkalivibrio</i>	<i>Reinekea</i>	<i>Rhodoferrax</i>
	<i>Congregibacter</i>	<i>Simiduia</i>	<i>Sulfuritalea</i>	<i>Roseateles</i>
	<i>Nitrocola</i>	<i>Salinivibrio</i>	<i>Oceanimonas</i>	<i>Bermanella</i>
	<i>Geoalkalibacter</i>	<i>Agarivorans</i>	<i>Streptococcus</i>	<i>Aliiglaciecola</i>
	<i>Hahella</i>	<i>Azospira</i>	<i>Sterolibacterium</i>	
	<i>Candidatus Accumulibacter</i>		<i>Candidatus Symbiobacter</i>	
	<i>Candidatus Kuenenia</i>			
3b	<i>Syntrophobacter</i>	<i>Desulfobulbus</i>	<i>Desulfurivibrio</i>	<i>Geobacter</i>
	<i>Desulfonatronospira</i>	<i>Desulfarculus</i>	<i>Desulfotalea</i>	<i>Desulfomicrobium</i>
	<i>Desulfovibrio</i>	<i>Desulfobacterium</i>	<i>Pelobacter</i>	<i>Desulfocarbo</i>
	<i>Desulfopila</i>	<i>Desulfobacter</i>	<i>Dethiosulfatibacter</i>	<i>Nitrospina</i>
	<i>Candidatus Magnetomorum</i>			

Single-domain PilZ SSN cluster 4

Figure S4. Sequence logo for single-domain PilZ cluster 4. (A) cluster 4a and (B) cluster 4b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S4. List of bacteria genera in single-domain PilZ cluster 4

Cluster	Genus
4a	<i>Nitrospira</i>
4b	

Single-domain PilZ SSN cluster 5

Figure S5. Sequence logos for single-domain PilZ cluster 5. (A) cluster 5a, (B) cluster 5b and (C) cluster 5c. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S5. List of bacteria genera in single-domain PilZ cluster 5.

Cluster	Genus			
5a	<i>Luteimonas</i>	<i>Thermomonas</i>	<i>Thiolapillus</i>	<i>Pseudoxanthomonas</i>
	<i>Arenimonas</i>	<i>Sedimenticola</i>	<i>Pseudomonas</i>	<i>Dokdonella</i>
	<i>Stenotrophomonas</i>	<i>Sphingomonas</i>	<i>Rhodanobacter</i>	<i>Xylella</i>
	<i>Oleigrimonas</i>	<i>Xanthomonas</i>	<i>Alcanivorax</i>	<i>Dyella</i>
	<i>Lysobacter</i>			
5b	<i>Pseudomonas</i>	<i>Azoarcus</i>	<i>Enterobacter</i>	<i>Thiomargarita</i>
	<i>Chitinivibrio</i>	<i>Desulfofustis</i>	<i>Chitinispirillum</i>	<i>Desulfurivibrio</i>
	<i>Streptococcus</i>	<i>Endozoicomonas</i>	<i>Prochlorothrix</i>	<i>Desulfocapsa</i>
	<i>Desulfopila</i>	<i>Stenotrophomonas</i>	<i>Halorhodospira</i>	<i>Acinetobacter</i>
	<i>Beggiatoa</i>	<i>Azotobacter</i>	<i>Pleurocapsa</i>	<i>Geoalkalibacter</i>
	<i>Desulfurispirillum</i>	<i>Caldithrix</i>	<i>Candidatus Contendobacter</i>	
	<i>Candidatus Competibacter</i>			
5c	<i>Pseudomonas</i>	<i>Simiduia</i>		

Single-domain PilZ SSN cluster 6

Figure S6. Sequence logos for single-domain PilZ cluster 6. (A) cluster 6a and (B) cluster 6b. The conserved RxxxFR and DxSxxG motifs are indicated by the asterisk and line; the extra phenylalanine residue in the sequence is also highlighted. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S6. List of bacteria genera in single-domain PilZ cluster 6

Cluster	Genus			
6a	<i>Paenibacillus</i>			
6b	<i>Halobacillus</i>	<i>Domibacillus</i>	<i>Exiguobacterium</i>	<i>Paraliobacillus</i>
	<i>Geobacillus</i>	<i>Pontibacillus</i>	<i>Lysinibacillus</i>	<i>Virgibacillus</i>
	<i>Bacillus</i>	<i>Sporosarcina</i>	<i>Gracilibacillus</i>	<i>Jeotgalibacillus</i>
	<i>Anoxybacillus</i>	<i>Planococcus</i>	<i>Kurthia</i>	<i>Aneurinibacillus</i>
	<i>Viridibacillus</i>	<i>Planomicrobium</i>	<i>Ornithinibacillus</i>	<i>Psychrobacillus</i>
	<i>Parageobacillus</i>	<i>Solibacillus</i>	<i>Rummeliibacillus</i>	

Single-domain PilZ SSN cluster 7

Figure S7. Sequence logos for single-domain PilZ cluster 7. (A) cluster 7a and (B) cluster 7b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S7. List of bacteria genera in single-domain PilZ cluster 7

Cluster	Genus			
7a	<i>Altererythrobacter</i>	<i>Sphingomonas</i>	<i>Sphingorhabdus</i>	<i>Porphyrobacter</i>
	<i>Erythrobacter</i>	<i>Sphingobium</i>	<i>Pseudomonas</i>	<i>Novosphingobium</i>
	<i>Sphingopyxis</i>	<i>Croceicoccus</i>		
7b	<i>Sphingomonas</i>			

Single-domain PilZ SSN cluster 8

Figure S8. Sequence logo for single-domain PilZ cluster 8. Residues reported to mediate protein-protein interactions with FimX are indicated by red arrows. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S8. List of bacteria genera in single-domain PilZ cluster 8.

Cluster	Genus			
8	<i>Salinisphaera</i>	<i>Dichelobacter</i>	<i>Beggiatoa</i>	<i>Halolamina</i>
	<i>Dyella</i>	<i>Morococcus</i>	<i>Methylobacillus</i>	<i>Pseudoxanthomonas</i>
	<i>Methyломicrobium</i>	<i>Thiobacillus</i>	<i>Oleiphilus</i>	<i>Moraxella</i>
	<i>Chitinimonas</i>	<i>Acinetobacter</i>	<i>Hahella</i>	<i>Wenzhouxiangella</i>
	<i>Methylophilus</i>	<i>Microbulbifer</i>	<i>Cellvibrio</i>	<i>Hydrocarboniphaga</i>
	<i>Pelomonas</i>	<i>Neisseria</i>	<i>Saccharophagus</i>	<i>Psychrobacter</i>
	<i>Paucibacter</i>	<i>Idiomarina</i>	<i>Marinobacter</i>	<i>Collimonas</i>
	<i>Methylophaga</i>	<i>Thermomonas</i>	<i>Simiduia</i>	<i>Thiothrix</i>
	<i>Pseudomonas</i>	<i>Sphingomonas</i>	<i>Polaromonas</i>	<i>Ralstonia</i>
	<i>Pseudorhodoferax</i>	<i>Alcanivorax</i>	<i>Ramlibacter</i>	<i>Frateuria</i>
	<i>Rhodoferax</i>	<i>Rhodanobacter</i>	<i>Thiomargarita</i>	<i>Thiomonas</i>
	<i>Halioglobus</i>	<i>Thiocystis</i>	<i>Lysobacter</i>	<i>Azospira</i>
	<i>Pseudohalaea</i>	<i>Marichromatium</i>	<i>Crenobacter</i>	<i>Oleiagrimonas</i>
	<i>Luminiphilus</i>	<i>Allochromatium</i>	<i>Methyloversatilis</i>	<i>Marinospirillum</i>
	<i>Congregibacter</i>	<i>Thioflavococcus</i>	<i>Achromatium</i>	<i>Lautropia</i>
	<i>Pseudalteromonas</i>	<i>Thalassomonas</i>	<i>Oblitimonas</i>	<i>Ectothiorhodospira</i>
	<i>Thalassotalea</i>	<i>Halorhodospira</i>	<i>Thiorhodospira</i>	<i>Alishewanella</i>
	<i>Janthinobacterium</i>	<i>Thiocapsa</i>	<i>Sphaerotilus</i>	<i>Cycloclasticus</i>
	<i>Duganella</i>	<i>Legionella</i>	<i>Sulfuricella</i>	<i>Azotobacter</i>
	<i>Vogesella</i>	<i>Aquabacterium</i>	<i>Neptuniibacter</i>	<i>Xylella</i>
	<i>Methylomonas</i>	<i>Rubrivivax</i>	<i>Chromobacterium</i>	<i>Thioalkalivibrio</i>
	<i>Methyloglobulus</i>	<i>Mycobacterium</i>	<i>Cupriavidus</i>	<i>Thalassolituus</i>
	<i>Methylobacter</i>	<i>Methylothena</i>	<i>Thiorhodovibrio</i>	<i>Oleispira</i>
	<i>Xanthomonas</i>	<i>Limnobacter</i>	<i>Luteimonas</i>	<i>Dokdonella</i>
	<i>Stenotrophomonas</i>	<i>Methylococcus</i>	<i>Thauera</i>	<i>Ideonella</i>
	<i>Catenovulum</i>	<i>Cobetia</i>	<i>Acidithiobacillus</i>	<i>Rhizobacter</i>
	<i>Nitrincola</i>	<i>Endozoicomonas</i>	<i>Luteibacter</i>	<i>Leptothrix</i>
	<i>Rheinheimera</i>	<i>Arenimonas</i>	<i>Acidihalobacter</i>	<i>Marinobacterium</i>
	<i>Ottowia</i>	<i>Arsukibacterium</i>	<i>Comamonas</i>	<i>Teredinibacter</i>
	<i>Sedimenticola</i>	<i>Oceanobacter</i>	<i>Delftia</i>	<i>Thiolapillus</i>
	<i>Noviherbaspirillum</i>	<i>Bermanella</i>	<i>Azoarcus</i>	<i>Ferrimonas</i>
	<i>Herbaspirillum</i>	<i>Enterobacter</i>	<i>Aromatoleum</i>	<i>Amantichitinum</i>
	<i>Herminiimonas</i>	<i>Oleibacter</i>	<i>Ventosimonas</i>	<i>Woeseia</i>
	<i>Variovorax</i>	<i>Streptococcus</i>	<i>Halothiobacillus</i>	<i>Eikenella</i>
	<i>Pseudacidovorax</i>	<i>Methylibium</i>	<i>Nitrococcus</i>	<i>Methylovorus</i>
	<i>Thioploca</i>	<i>Shewanella</i>	<i>Alkalilimnicola</i>	<i>Snodgrassella</i>
	<i>Tepidiphilus</i>	<i>Kangiella</i>	<i>Nitrosococcus</i>	<i>Puccinia</i>
	<i>Spongiibacter</i>	<i>Neptunomonas</i>	<i>Lampropedia</i>	<i>Hydrogenophaga</i>
	<i>Acidovorax</i>	<i>Aquitalea</i>	<i>Fluoribacter</i>	<i>Xylophilus</i>
	<i>Diaphorobacter</i>	<i>Pseudogulbenkiania</i>	<i>Reinekea</i>	<i>Curvibacter</i>

Single-domain PilZ SSN cluster 9

Figure S9. Sequence logos for single-domain PilZ cluster 9. (A) cluster 9a and (B) cluster 9b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S9. List of bacteria genera in single-domain PilZ cluster 9.

Cluster	Genus			
9a	<i>Lachnospira</i>	<i>Catonella</i>	<i>Caminicella</i>	<i>Butyrivibrio</i>
	<i>Eubacterium</i>	<i>Paramaledivibacter</i>	<i>Acetivibrio</i>	<i>Geosporobacter</i>
	<i>Lachnoclostridium</i>	<i>Anaerosporebacter</i>		
9b	<i>Roseburia</i>	<i>Butyrivibrio</i>	<i>Pseudobutyrvibrio</i>	<i>Lachnospira</i>
	<i>Eubacterium</i>	<i>Clostridium</i>		

Single-domain PilZ SSN cluster 10

Figure S10. Sequence logo for single-domain PilZ cluster 10. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S10. List of bacteria genera in single-domain PilZ cluster 10

Cluster	Genus			
10	<i>Pseudomonas</i>	<i>Thiorhodococcus</i>	<i>Gynuella</i>	<i>Reinekea</i>
	<i>Methylibium</i>	<i>Methylophaga</i>	<i>Oleiphilus</i>	<i>Marinomonas</i>
	<i>Leptothrix</i>	<i>Spongiibacter</i>	<i>Cellvibrio</i>	<i>Allochromatium</i>
	<i>Thiocystis</i>	<i>Thiorhodovibrio</i>	<i>Candidatus Endobugula</i>	<i>Hahella</i>
	<i>Mariprofundus</i>	<i>Marichromatium</i>	<i>Marinobacterium</i>	<i>Thioploca</i>
	<i>Rhizobacter</i>	<i>Beggiatoa</i>	<i>Marinobacter</i>	<i>Neptuniibacter</i>
	<i>Candidatus Competibacter</i>		<i>Candidatus Accumulibacter</i>	

Single-domain PilZ SSN cluster 11

Figure S11. Sequence logo for single-domain PilZ cluster 11. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S11. List of bacteria genera in single-domain PilZ cluster 11.

Cluster	Genus			
11	<i>Novosphingobium</i>	<i>Blastomonas</i>	<i>Sphingomonas</i>	<i>Erythrobacter</i>
	<i>Altererythrobacter</i>	<i>Mastigocladus</i>	<i>Porphyrobacter</i>	<i>Sphingopyxis</i>
	<i>Sphingobium</i>	<i>Citromicrobium</i>	<i>Croceicoccus</i>	<i>Sphingorhabdus</i>

Single domain PilZ PA0012 cluster

Figure S12. Sequence logo for the single-domain PilZ cluster that includes PA0012. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S12. List of bacteria genera in the single-domain PilZ cluster that includes PA0012.

Cluster	Genus			
PA0012	<i>Pseudomonas</i>	<i>Cellvibrio</i>	<i>Enterobacter</i>	<i>Oleiphilus</i>
	<i>Streptococcus</i>	<i>Simiduia</i>	<i>Acinetobacter</i>	<i>Micrococcus</i>
	<i>Marinobacter</i>	<i>Saccharophagus</i>		

Single domain PilZ DgrA and DgrB clusters

Figure S13. Sequence logo for the single-domain PilZ clusters that includes DgrA and DgrB. (A) DgrA cluster and (B) DgrB cluster. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S13. List of bacteria genera in the single-domain PilZ clusters that includes DgrA and DgrB.

Cluster	Genus
DgrA	<i>Caulobacter</i>
DgrB	

Di-domain PilZ SSN sequence logos and bacteria genera

Di-domain PilZ SSN cluster 1

Figure S14. Sequence logo for di-domain PilZ cluster 1. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S14. List of bacteria genera in di-domain PilZ cluster 1

Cluster	Genus			
1	<i>Nitrincola</i>	<i>Pseudomonas</i>	<i>Neptunomonas</i>	<i>Marinobacterium</i>
	<i>Oleiphilus</i>	<i>Cupriavidus</i>	<i>Oleispira</i>	<i>Janthinobacterium</i>
	<i>Reinekea</i>	<i>Ferriphaselus</i>	<i>Methyloversatilis</i>	<i>Duganella</i>
	<i>Azospira</i>	<i>Neptuniibacter</i>	<i>Methylobacter</i>	<i>Massilia</i>
	<i>Methylobacillus</i>	<i>Herminiimonas</i>	<i>Marinomonas</i>	<i>Noviherbaspirillum</i>
	<i>Hahella</i>	<i>Sedimenticola</i>	<i>Nitrosomonas</i>	<i>Thalassolituus</i>
	<i>Teredinibacter</i>	<i>Cellvibrio</i>	<i>Pseudorhodoferrax</i>	<i>Pandora</i>
	<i>Saccharophagus</i>	<i>Simiduia</i>	<i>Gyruella</i>	<i>Curvibacter</i>
	<i>Thauera</i>	<i>Bermanella</i>		

Di-domain PilZ SSN cluster 2

Figure S15. Sequence logos for di-domain PilZ cluster 2. (A) cluster 2a and (B) cluster 2b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S15. List of bacteria genera in di-domain PilZ cluster 2.

Cluster	Genus			
2a	<i>Fictibacillus</i>	<i>Rummeliibacillus</i>	<i>Paenibacillus</i>	<i>Virgibacillus</i>
	<i>Tumebacillus</i>	<i>Kurthia</i>	<i>Bacillus</i>	<i>Terribacillus</i>
	<i>Salimicrobium</i>	<i>Exiguobacterium</i>	<i>Streptococcus</i>	<i>Thermobacillus</i>
	<i>Hydrogenibacillus</i>	<i>Ornithinibacillus</i>	<i>Lysinibacillus</i>	<i>Alicyclobacillus</i>
	<i>Aeribacillus</i>	<i>Tepidibacillus</i>	<i>Paraliobacillus</i>	<i>Oceanobacillus</i>
	<i>Psychrobacillus</i>	<i>Anoxybacillus</i>	<i>Geobacillus</i>	<i>Pontibacillus</i>
	<i>Domibacillus</i>	<i>Brevibacillus</i>	<i>Viridibacillus</i>	<i>Solibacillus</i>
	<i>Aneurinibacillus</i>	<i>Sporolactobacillus</i>	<i>Gracilibacillus</i>	<i>Sporosarcina</i>
	<i>Parageobacillus</i>	<i>Jeotgalibacillus</i>	<i>Halobacillus</i>	<i>Amphibacillus</i>
	<i>Cohnella</i>	<i>Paraliobacillus</i>		
2b	<i>Caloramator</i>	<i>Anaerocolumna</i>	<i>Gottschalkia</i>	<i>Defluviitoga</i>
	<i>Centipeda</i>	<i>Lachnoclostridium</i>	<i>Anaerovibrio</i>	<i>Lachnospira</i>
	<i>Acetohalobium</i>	<i>Thermosipho</i>	<i>Peptoanaerobacter</i>	<i>Selenomonas</i>
	<i>Oxobacter</i>	<i>Caldicellulosiruptor</i>	<i>Pseudothermotoga</i>	<i>Desulfotomaculum</i>
	<i>Fervidicola</i>	<i>Thermovenabulum</i>	<i>Mahella</i>	<i>Ruminiclostridium</i>
	<i>Caldanaerovirga</i>	<i>Thermincola</i>	<i>Pelotomaculum</i>	<i>Syntrophomonas</i>
	<i>Caminicella</i>	<i>Mitsuokella</i>	<i>Eubacterium</i>	<i>Clostridium</i>
	<i>Paraclostridium</i>	<i>Thermacetogenium</i>	<i>Thermanaerovibrio</i>	<i>Tepidibacter</i>
	<i>Halanaerobium</i>	<i>Thermosinus</i>	<i>Dethiobacter</i>	<i>Halothermothrix</i>
	<i>Tepidanaerobacter</i>	<i>Thermotoga</i>	<i>Denitrovibrio</i>	<i>Pseudobutyrvibrio</i>
	<i>Fervidicella</i>	<i>Aminomonas</i>	<i>Hippea</i>	<i>Desulfosporosinus</i>
	<i>Anaerosporobacter</i>	<i>Sporanaerobacter</i>	<i>Heliobacterium</i>	<i>Syntrophothermus</i>
	<i>Thermotalea</i>	<i>Natranaerobius</i>	<i>Orenia</i>	<i>Caloranaerobacter</i>
	<i>Schwartzia</i>	<i>Acetomicrobium</i>	<i>Acetonema</i>	<i>Roseburia</i>
	<i>Tissierella</i>	<i>Cellulosilyticum</i>	<i>Thermoanaerobacterium</i>	<i>Desulfurella</i>
	<i>Syntrophobotulus</i>	<i>Pelosinus</i>	<i>Paramaledivibacter</i>	<i>Peptoclostridium</i>
	<i>Dehalobacter</i>	<i>Propionispora</i>	<i>Fervidobacterium</i>	<i>Anaerosporomusa</i>
	<i>Caldisalibacter</i>	<i>Alkaliphilus</i>	<i>Thermaerobacter</i>	<i>Dethiosulfovibrio</i>
	<i>Acetivibrio</i>	<i>Ammonifex</i>	<i>Mucispirillum</i>	<i>Thermosyntropha</i>
	<i>Petrotoga</i>	<i>Moorella</i>	<i>Caldanaerobacter</i>	<i>Anaerobium</i>
	<i>Carboxydotherrmus</i>	<i>Butyrvibrio</i>	<i>Catonella</i>	<i>Desulfitobacterium</i>
	<i>Marinitoga</i>	<i>Thermosediminibacter</i>		
	<i>Thermoanaerobacter</i>			

Di-domain PilZ SSN cluster 3

Figure S16. Sequence logos for di-domain PilZ cluster 3. (A) cluster 3a and (B) cluster 3b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Figure S16 (continued). Sequence logos for di-domain PilZ cluster 3. (C) cluster 3c and (D) cluster 3d. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S16. List of bacteria genera in di-domain PilZ cluster 3.

Cluster	Genus			
3a	<i>Desulfovibrio</i>	<i>Moritella</i>	<i>Desulfocarbo</i>	<i>Aliivibrio</i>
	<i>Halodesulfovibrio</i>	<i>Desulfotignum</i>	<i>Syntrophobacter</i>	<i>Desulfarculus</i>
	<i>Desulfobacula</i>	<i>Desulfobacterium</i>	<i>Candidatus Magnetomorum</i>	
3b	<i>Salinivibrio</i>	<i>Grimontia</i>	<i>Enterovibrio</i>	<i>Thalassomonas</i>
	<i>Vibrio</i>	<i>Aeromonas</i>	<i>Photobacterium</i>	
3c	<i>Aliiglaciecola</i>	<i>Alishewanella</i>	<i>Shewanella</i>	<i>Glaciecola</i>
	<i>Idiomarina</i>	<i>Alteromonas</i>	<i>Rheinheimera</i>	<i>Marinomonas</i>
	<i>Pseudoalteromonas</i>	<i>Colwellia</i>	<i>Lacimicrobium</i>	<i>Ferrimonas</i>
	<i>Agarivorans</i>	<i>Aestuariibacter</i>	<i>Paraglaciecola</i>	<i>Moritella</i>
	<i>Thalassomonas</i>			
3d	<i>Vibrio</i>	<i>Nitrococcus</i>	<i>Acidihalobacter</i>	<i>Shewanella</i>
	<i>Enterovibrio</i>	<i>Nitrospira</i>	<i>Grimontia</i>	
	<i>Candidatus Magnetoovum</i>			

Di-domain PilZ SSN cluster 4

Figure S17. Sequence logo for di-domain PilZ cluster 4. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S17. List of bacteria genera in di-domain PilZ cluster 4.

Cluster	Genus			
4	<i>Geobacter</i>	<i>Pelobacter</i>	<i>Malonomonas</i>	<i>Desulfurispirillum</i>
	<i>Desulfuromonas</i>	<i>Halorhodospira</i>	<i>Pseudoalteromonas</i>	<i>Terasakiispira</i>
	<i>Geoalkalibacter</i>	<i>Marinospirillum</i>	<i>Candidatus Magnetoovum</i>	
	<i>Candidatus Koribacter</i>			

Di-domain PilZ SSN cluster 5

Figure S18. Sequence logos for di-domain PilZ cluster 5. (A) cluster 5a; residues that were identified to mediate interaction with FliG are annotated with red arrows and residues that mediate interaction with FliM are annotated with yellow arrows. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Figure S18 (continued). Sequence logos for di-domain PilZ cluster 5. (C) cluster 5c. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S18. List of bacteria genera in di-domain PilZ cluster 5.

Cluster	Genus			
5a	<i>Dickeya</i>	<i>Cronobacter</i>	<i>Pectobacterium</i>	<i>Mumia</i>
	<i>Caballeronia</i>	<i>Lelliottia</i>	<i>Rahnella</i>	<i>Shigella</i>
	<i>Burkholderia</i>	<i>Leclercia</i>	<i>Enterobacter</i>	<i>Methyloglobulus</i>
	<i>Serratia</i>	<i>Kluyvera</i>	<i>Atlantibacter</i>	<i>Collimonas</i>
	<i>Klebsiella</i>	<i>Cedecea</i>	<i>Franconibacter</i>	<i>Kosakonia</i>
	<i>Shimwellia</i>	<i>Yersinia</i>	<i>Trabulsiella</i>	<i>Salmonella</i>
	<i>Paraburkholderia</i>	<i>Vibrio</i>	<i>Ralstonia</i>	<i>Escherichia</i>
	<i>Cupriavidus</i>	<i>Citrobacter</i>	<i>Mangrovibacter</i>	<i>Achromobacter</i>
	<i>Buttiauxella</i>	<i>Variovorax</i>	<i>Pantoea</i>	<i>Yokenella</i>
	<i>Ewingella</i>	<i>Beauveria</i>	<i>Erwinia</i>	<i>Brenneria</i>
5b	<i>Raoultella</i>			
	<i>Alcaligenes</i>	<i>Janthinobacterium</i>	<i>Methylomonas</i>	<i>Ferrovum</i>
	<i>Achromobacter</i>	<i>Rhizobacter</i>	<i>Stenotrophomonas</i>	<i>Laribacter</i>
	<i>Dechloromonas</i>	<i>Xanthomonas</i>	<i>Chromobacterium</i>	<i>Acidovorax</i>
	<i>Massilia</i>	<i>Bordetella</i>	<i>Ideonella</i>	<i>Azospira</i>
	<i>Tepidiphilus</i>	<i>Vogesella</i>	<i>Gallionella</i>	<i>Thermomonas</i>
	<i>Thauera</i>	<i>Pseudorhodoferax</i>	<i>Pseudogulbenkiania</i>	<i>Nitrosomonas</i>
5c	<i>Candidatus Accumolibacter</i>			
	<i>Thioalkalivibrio</i>	<i>Thiorhodospira</i>	<i>Thiomonas</i>	<i>Oleibacter</i>
	<i>Pseudomonas</i>	<i>Stenotrophomonas</i>	<i>Acinetobacter</i>	<i>Oceanobacter</i>
	<i>Acidithiobacillus</i>	<i>Marichromatium</i>	<i>Enterobacter</i>	<i>Marinospirillum</i>
	<i>Ectothiorhodospira</i>	<i>Methylomonas</i>	<i>Streptococcus</i>	<i>Thalassolituus</i>
	<i>Halorhodospira</i>	<i>Terasakiispira</i>	<i>Frateuria</i>	<i>Thiorhodococcus</i>
	<i>Bermanella</i>	<i>Acidihalobacter</i>	<i>Oleiphilus</i>	

Di-domain PilZ SSN cluster 6

Figure S19. Sequence logo for di-domain PilZ cluster 6. (A) cluster 6a; putative cyclic di-binding motifs are indicated and (B) cluster 6b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S19. List of bacteria genera in di-domain PilZ cluster 6

Cluster	Genus			
6a	<i>Saccharophagus</i>	<i>Streptococcus</i>	<i>Thalassolituus</i>	<i>Thermodesulfatator</i>
	<i>Methylophaga</i>	<i>Hahella</i>	<i>Cycloclasticus</i>	<i>Alkalilimnicola</i>
	<i>Pseudomonas</i>	<i>Candidatus Endobugula</i>	<i>Cellvibrio</i>	<i>Methyломicrobium</i>
	<i>Oleiphilus</i>	<i>Oceanobacter</i>	<i>Fluoribacter</i>	<i>Thiorhodococcus</i>
	<i>Methylobacter</i>	<i>Oleibacter</i>	<i>Stenotrophomonas</i>	<i>Piscirickettsia</i>
	<i>Thiocystis</i>	<i>Oleispira</i>	<i>Gynuella</i>	<i>Neptuniibacter</i>
	<i>Nitrincola</i>	<i>Thiolapillus</i>	<i>Acinetobacter</i>	<i>Ventrosimonas</i>
	<i>Marichromatium</i>	<i>Legionella</i>	<i>Enterobacter</i>	<i>Congregibacter</i>
	<i>Teredinibacter</i>	<i>Allochromatium</i>	<i>Methylomonas</i>	<i>Simiduia</i>
	<i>Neptunomonas</i>	<i>Reinekea</i>	<i>Marinobacter</i>	<i>Nitrococcus</i>
	<i>Tatlockia</i>	<i>Magnetococcus</i>	<i>Sedimenticola</i>	<i>Desulfurivibrio</i>
	<i>Marinobacterium</i>			
6b	<i>Thalassomonas</i>	<i>Plesiomonas</i>	<i>Grimontia</i>	<i>Aliiglaciecola</i>
	<i>Colwellia</i>	<i>Shewanella</i>	<i>Arsukibacterium</i>	<i>Enterovibrio</i>
	<i>Rheinheimera</i>	<i>Alteromonas</i>	<i>Pseudoalteromonas</i>	<i>Aestuariibacter</i>

Di-domain PilZ SSN cluster 7

Figure S20. Sequence logo for di-domain PilZ cluster 7 which contains two sets of cyclic di-GMP binding motifs The predicted complete cyclic di-GMP-binding motifs in both sequence logos are indicated in black and the predicted incomplete motif is indicated in red. (A) cluster 7a and (B) cluster 7b. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S20. List of bacteria genera in di-domain PilZ cluster 7

Cluster	Genus			
7a	<i>Sphingopyxis</i>	<i>Sphingomonas</i>	<i>Novosphingobium</i>	<i>Sphingobium</i>
	<i>Altererythrobacter</i>			
7b	<i>Novosphingobium</i>	<i>Sphingomonas</i>	<i>Sphingorhabdus</i>	<i>Altererythrobacter</i>
	<i>Erythrobacter</i>	<i>Porphyrobacter</i>		

Di-domain PilZ SSN cluster 8

Figure S21. Sequence logo for di-domain PilZ cluster 8. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S21. List of bacteria genera in di-domain PilZ cluster 8

Cluster	Genus			
8	<i>Devosia</i>	<i>Afipia</i>	<i>Chelatococcus</i>	<i>Hoeflea</i>
	<i>Phyllobacterium</i>	<i>Bradyrhizobium</i>	<i>Pannonibacter</i>	<i>Pleomorphomonas</i>
	<i>Prosthecomicrobium</i>	<i>Sinorhizobium</i>	<i>Stappia</i>	<i>Neorhizobium</i>
	<i>Methyloceanibacter</i>	<i>Ensifer</i>	<i>Pseudolabrys</i>	<i>Aurantimonas</i>
	<i>Chelativorans</i>	<i>Pelagibacterium</i>	<i>Tardiphaga</i>	<i>Shinella</i>
	<i>Hyphomicrobium</i>	<i>Oceanicaulis</i>	<i>Maricaulis</i>	<i>Ochrobactrum</i>
	<i>Rhodopseudomonas</i>	<i>Parvularcula</i>	<i>Aureimonas</i>	<i>Hirschia</i>
	<i>Nitratireductor</i>	<i>Bosea</i>	<i>Polymorphum</i>	<i>Lutibaculum</i>
	<i>Methylobacterium</i>	<i>Labrenzia</i>	<i>Oligotropha</i>	<i>Streptomyces</i>
	<i>Roseibium</i>	<i>Pseudovibrio</i>	<i>Blastochloris</i>	<i>Rhodoplanes</i>
	<i>Rhizobium</i>	<i>Rhodovulum</i>	<i>Ahrensia</i>	<i>Parvibaculum</i>
	<i>Agrobacterium</i>	<i>Fulvimarina</i>	<i>Filomicrobium</i>	<i>Nitrobacter</i>
	<i>Hyphomonas</i>	<i>Tepidicaulis</i>	<i>Candidatus Liberibacter</i>	
	<i>Candidatus Phaeomarinobacter</i>			

Di-domain PilZ SSN cluster 9

Figure S22. Sequence logo for di-domain PilZ cluster 9. The predicted cyclic di-GMP binding motif and the predicted XRE family HTH region in the C-terminal region are indicated. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S22. List of bacteria genera in di-domain PilZ cluster 9

Cluster	Genus			
9	<i>Novosphingobium</i>	<i>Sphingopyxis</i>	<i>Porphyrobacter</i>	<i>Sphingomonas</i>
	<i>Sphingobium</i>	<i>Altererythrobacter</i>	<i>Erythrobacter</i>	

Di-domain PilZ SSN PlzA (BB0733) cluster

Figure S23. Sequence logo for the di-domain PilZ cluster that includes PlzA. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S23. List of bacteria genera in the di-domain PilZ cluster that includes PlzA (BB0733)

Cluster	Genus		
PlzA	<i>Spirochaeta</i>	<i>Treponema</i>	<i>Salinispira</i>

Di-domain SSN xcc-b10_2234 cluster

Figure S24. Sequence logo for the di-domain PilZ cluster that includes xcc-b100_2234. The residues are colored according to hydrophobicity and the height of each residue corresponds to its degree of conservation.

Table S24. List of bacteria genera in the di-domain PilZ cluster that includes xcc-b100_2234

Cluster		Genus	
xcc-b100_2234	<i>Lysobacter</i>	<i>Luteimonas</i>	<i>Xanthomonas</i>

Figure S25. Complete histogram of the sequence length of PilZ proteins.

Figure S26. Bacteria two-hybrid screen of potential FlgZ partners.

Table S25. List of bacteria strains and plasmids used

No.	Bacterial Strain/plasmid	Description	Source
1	<i>P. aeruginosa</i> PAO1	Wild type <i>P. aeruginosa</i> PAO1	Lab stock
2	<i>E. coli</i> DH5 α	Cells for routine cloning experiments	Stratagene
3	Δ PA0012	PAO1 strain with in-frame deleted PA0012	Collaborator
4	Δ PA2989	PAO1 strain with in-frame deleted PA2989 (base 166-765)	Collaborator
5	Δ FlgZ	PAO1 strain with in-frame deleted PA3353	Collaborator
6	Δ PA4324	PAO1 strain with in-frame deleted PA4324	Collaborator
7	PA0012+	PAO1 strain transformed with pUCP18-PA0012	This study
8	PA2989+	PAO1 strain transformed with pUCP18-PA2989	This study
9	FlgZ+	PAO1 strain transformed with pUCP18-PA3353	This study
10	PA4324+	PAO1 strain transformed with pUCP18-PA4324	This study
11	Δ PA0012c	Δ PA0012 strain transformed with pUCP18-PA0012	This study
12	Δ PA2989c	Δ PA2989 strain transformed with pUCP18-PA2989	This study
13	Δ FlgZc	Δ PA3353 strain transformed with pUCP18-PA3353	This study
14	Δ PA4324c	Δ PA4324 strain transformed with pUCP18-PA4324	This study
15	HA-0012	PAO1 strain transformed with pJN105-HA-PA0012	This study
16	HA-0012R5A	PAO1 strain transformed with pJN105-HA-PA0012	This study
17	HA-4324	PAO1 strain transformed with pJN105-HA-PA4324	This study
18	HA-4324R9A	PAO1 strain transformed with pJN105-HA-PA4324R9A	This study
19	DH5 α Prey	DH5 α cells transformed with pCRISPOmyces-2 plasmid	This study
20	BacterioMatch II reporter strain	Host strain used for two-hybrid screening	Stratagene
21	pUCP18	<i>E. coli</i> - <i>P. aeruginosa</i> shuttle expression plasmid with Plac, Amp ^r , Car ^r	Addgene
22	pUCP18-PA0012	DNA fragment encoding full length PA0012 cloned into pUCP18 at SacI and BamHI site	This study
23	pUCP18-PA2989	DNA fragment encoding full length PA2989 cloned into pUCP18 at SacI and XbaI site	This study
24	pUCP18-PA3353	DNA fragment encoding full length PA3353 (flgZ) cloned into pUCP18 at SacI and BamHI site	This study
25	pUCP18-PA4324	DNA fragment encoding full length PA4324 cloned into pUCP18 at SacI and BamHI site	This study
26	pJN105	<i>E. coli</i> - <i>P. aeruginosa</i> Inducible expression plasmid with Gen ^r	Addgene

Table S25 (continued). List of bacteria strains and plasmids used

No.	Bacterial Strain/plasmid	Description	Source
27	pJN105- <i>HA-PA0012</i>	DNA fragment encoding full length <i>PA0012</i> with in-frame N-terminal HA tag cloned into pJN105 at EcoRI and XbaI site	This study
28	pJN105- <i>HA-PA0012^{RSA}</i>	DNA fragment encoding full length <i>PA0012^{RSA}</i> mutant with in-frame N-terminal HA tag cloned into pJN105 at EcoRI and XbaI site	This study
29	pJN105- <i>HA-PA4324</i>	DNA fragment encoding full length <i>PA4324</i> with in-frame N-terminal HA tag cloned into pJN105 at NheI and SacI site	This study
30	pJN105- <i>HA-PA4324^{RSA}</i>	DNA fragment encoding full length <i>PA4324^{RSA}</i> mutant with in-frame N-terminal HA tag cloned into pJN105 at NheI and SacI site	This study
31	<i>pqsA :: gfp(ASV)</i>	pqsA reporter plasmid	Collaborator
32	pCRISPomyces-2	Plasmid for complementation of lacZ gene	Addgene
33	pBT	Empty bait plasmid for bacterial two-hybrid assay	Stratagene
34	pTRG	Empty target plasmid for bacterial two-hybrid assay	Stratagene
35	pBT- <i>PA3353</i>	DNA fragment encoding full length <i>PA3353 (flgZ)</i> cloned into pBT at EcoRI and XhoI site	Stratagene
36	pTRG- <i>FliG</i>	DNA fragment encoding full length <i>FliG</i> cloned into pTRG at EcoRI and XhoI site	Stratagene
37	pTRG- <i>FliM</i>	DNA fragment encoding full length <i>FliM</i> cloned into pTRG at EcoRI and XhoI site	Stratagene
38	pTRG- <i>MotA</i>	DNA fragment encoding full length <i>FliM</i> cloned into pTRG at EcoRI and SpeI site	Stratagene
39	pBT- <i>LGF2</i>	Positive interaction control plasmid for two-hybrid screening; encodes dimerization domain of <i>Gal4</i> transcriptional activator	Stratagene
40	pTRG- <i>Gal 11P</i>	Positive interaction control plasmid for two-hybrid screening; encodes a domain of mutant <i>Gal 11</i>	Stratagene

Table S26. List of primers used

No.	Primer	Sequence	Description
1	pUC-PA0012 Forward	GATCGAGCTCATGGATAATCAACGACAATAC	For construction of pUCP18-PA0012
2	pUC-PA0012 Reverse	GATCGGATCCTCAGGCTTCGCTGAGAAAGC	For construction of pUCP18-PA0012
3	pUC-PA4324 Forward	GATCGAGCTCATGGGTCTTCAACGACATATC	For construction of pUCP18-PA4324
4	pUC-PA4324 Reverse	GATCGGATCCTCAGACCGACTGCGGCTGGC	For construction of pUCP18-PA4324
5	pJn-HA-PA0012 Forward	GATCGAATTCATGTATCCATATGACGTTCCAGATTACG CTACCCGGGTACGTATGGATAATCAACGACAATAC	For construction of pJN105-HA-PA0012
6	pJn-HA-PA0012 Reverse	GATCTCTAGATCAGGCTTCGCTGAGAAAGC	For construction of pJN105-HA-PA0012
7	pJn-HA-PA4324 Forward	GATCGCTAGCATGTATCCATATGACGTTCCAGATTACG CTACCCGGGTACGTATGGGTCTTCAACGACATATC	For construction of pJN105-HA-PA4324
8	pJn-HA-PA4324 Reverse	GATCGAGCTCTCAGACCGACTGCGGCTGGC	For construction of pJN105-HA-PA4324
9	pUC-PA2989 Forward	GATCGAGCTCATGGCTGCGACGTTTGGCCC	For construction of pUCP18-PA2989
10	pUC-PA2989 Reverse	GATCTCTAGATCAGGTCGAAGAAGGTTGGC	For construction of pUCP18-PA2989
11	pUC-PA3353 Forward	GATCGAGCTCGTGCTATCATTGAGGCATTC	For construction of pUCP18-PA3353
12	pUC-PA3353 Reverse	GATCTCTAGATCAGAACAGTTCGTCTTTCT	For construction of pUCP18-PA3353
13	pBT-PA3353 Forward	CTAGTCGAATTCGCTGCTATCATTGAGGCATTCGTGCC	For construction of pBT-PA3353
14	pBT-PA3353 Reverse	CTAGTCCTCGAGTCAGAACAGTTCGTCTTTCTCGAAG	For construction of pBT-PA3353
15	pTRG-Flig Forward	CTAGTCGAATTCAGATGAGTGAGAATCGTCTCGCCGC	For construction of pTRG-Flig
16	pTRG-Flig Reverse	CTAGTCCTCGAGTCAGATCATCTCCTCGCCACCC	For construction of pTRG-Flig
17	pTRG-Flim Forward	CTAGTCGAATTCAGATGGCCGTGCAAGATCTGCTTTC	For construction of pTRG-Flim
18	pTRG-Flim Reverse	CTAGTCCTCGAGTCAGCGCGAGCGCTCGAC	For construction of pTRG-Flim

Table S26 (continued). List of primers used

No.	Primer	Sequence	Description
19	pTRG- <i>MotA</i> Forward	CTAGTCGAATTCATGTCAAAAATCATCGGCATCATCG	For construction of pTRG- <i>MotA</i>
20	pTRG- <i>MotA</i> Reverse	CTAGTCACTAGTTCAGCGGCCGCGCACGGCCT	For construction of pTRG- <i>MotA</i>
19	pTRG- <i>MotA</i> Forward	CTAGTCGAATTCATGTCAAAAATCATCGGCATCATCG	For construction of pTRG- <i>MotA</i>
20	pTRG- <i>MotA</i> Reverse	CTAGTCACTAGTTCAGCGGCCGCGCACGGCCT	For construction of pTRG- <i>MotA</i>

Table S27. Genes up-regulated in $\Delta PA0012$ strain.

Gene	Fold Change	padj	Gene	Fold Change	padj
<i>PA0087</i>	13.608395	6.6E-58	<i>PA3732</i>	3.7914426	8.8E-76
<i>rsmZ</i>	8.7757052	1.2E-26	<i>leuA</i>	3.7157689	1.8E-73
<i>PA3661</i>	8.1960013	2.2E-45	<i>PA1845</i>	3.7023977	2.5E-14
<i>PA0089</i>	7.958125	1.8E-64	<i>PA3731</i>	3.688844	1E-45
<i>PA0082</i>	7.6036617	6.4E-98	<i>PA0047</i>	3.6328133	1.9E-11
<i>PA0083</i>	7.3987255	2E-148	<i>PA2464</i>	3.5811384	8.3E-32
<i>PA0086</i>	6.8304644	2E-121	<i>PA0046</i>	3.517809	3.5E-26
<i>PA0096</i>	6.6134568	5.7E-22	<i>PA3730</i>	3.5095784	1.3E-53
<i>PA0097</i>	6.5703456	2.7E-59	<i>PA0095</i>	3.4948699	8.1E-94
<i>PA3484</i>	6.4992722	3.8E-90	<i>PA0099</i>	3.4428575	1.1E-25
<i>PA0078</i>	6.4377066	3E-31	<i>PA3444</i>	3.4384258	7.3E-22
<i>PA4659</i>	6.161663	4.7E-46	<i>PA0563</i>	3.4034212	2.5E-46
<i>PA0080</i>	6.1433612	3.7E-55	<i>PA3729</i>	3.3942374	7E-110
<i>PA0098</i>	6.0263671	5.3E-33	<i>vgrG1</i>	3.3069039	3.8E-58
<i>PA0079</i>	5.9777308	2.4E-45	<i>leuC</i>	3.2702288	7E-101
<i>PA0088</i>	5.8946966	5.8E-78	<i>PA2550</i>	3.2654756	4.3E-15
<i>PA0840</i>	5.8577417	3E-43	<i>PA2774</i>	3.2648534	9E-20
<i>PA0084</i>	5.7015504	6E-101	<i>PA2287</i>	3.2210929	2.1E-10
<i>PA3485</i>	5.2046091	1.1E-16	<i>PA0173</i>	3.2156102	1.5E-12
<i>PA2703</i>	5.0527874	9.3E-39	<i>PA1396</i>	3.1911938	1.6E-25
<i>PA0040</i>	4.9351301	9.4E-60	<i>PA0528</i>	3.1778561	2.5E-09
<i>PA4658</i>	4.922971	7.5E-73	<i>PA0380</i>	3.1599605	4.1E-05
<i>PA4318</i>	4.8164716	2.2E-33	<i>PA2928</i>	3.1555516	2.7E-09
<i>icmF1</i>	4.7854422	3.7E-46	<i>PA3442</i>	3.1214433	1.3E-08
<i>PA0101</i>	4.732979	3.3E-25	<i>PA0093</i>	3.1176432	1.1E-40
<i>PA2540</i>	4.6765087	7.1E-96	<i>PA4491</i>	3.0866986	5.6E-67
<i>PA0045</i>	4.5128872	3.8E-57	<i>PA2538</i>	2.9863821	5.7E-12
<i>PA1639</i>	4.3392806	6.9E-50	<i>bioD</i>	2.9706453	0.00012
<i>dctA</i>	4.2694223	4.4E-11	<i>PA4889</i>	2.9611207	0.00046
<i>PA2376</i>	4.2033031	2.2E-15	<i>PA0578</i>	2.9606674	9E-11
<i>phr</i>	4.1775102	7.2E-73	<i>PA0562</i>	2.960086	4E-07
<i>pppA</i>	4.1198946	1.1E-19	<i>ribC</i>	2.8880885	7E-14
<i>PA0070</i>	4.0909043	8E-106	<i>PA4321</i>	2.866457	3.1E-23
<i>PA2539</i>	4.0231129	7.9E-56	<i>PA0094</i>	2.860721	3.1E-37
<i>clpV1</i>	3.9542438	2.2E-51	<i>hcp1</i>	2.8490186	1.4E-30
<i>PA0855</i>	3.9045762	6.4E-17	<i>PA0174</i>	2.8423109	1E-12
<i>PA0076</i>	3.8284957	8.5E-15	<i>bkdA1</i>	2.842132	8.4E-25
<i>lpdV</i>	3.8201565	1.5E-13	<i>PA2408</i>	2.8287423	5E-05
<i>PA5135</i>	3.808102	2E-08	<i>glpF</i>	2.8276073	4.1E-47

Table S27 (continued). Genes up-regulated in $\Delta PA0012$ strain.

Gene	Fold Change	padj	Gene	Fold Change	padj
<i>PA0072</i>	2.79849896	1.2E-05	<i>lexA</i>	2.42303866	3.9E-53
<i>PA2792</i>	2.79348256	1E-15	<i>PA4690.5</i>	2.39518545	0.01064
<i>PA5369.5</i>	2.78526647	0.00216	<i>PA0333</i>	2.39385147	1.1E-08
<i>PA0100</i>	2.74454392	2.8E-43	<i>PA4219</i>	2.3906034	9.3E-07
<i>pdxB</i>	2.743854	3.3E-06	<i>PA0369</i>	2.3800002	0.00599
<i>ppkA</i>	2.72032346	3.4E-41	<i>PA4320</i>	2.3738946	1.3E-24
<i>PA3928</i>	2.71826573	2.2E-06	<i>PA2655</i>	2.36914868	0.00404
<i>PA4319</i>	2.71497931	1E-23	<i>PA3427</i>	2.36508383	2E-15
<i>PA0730</i>	2.70184271	5.9E-40	<i>PA3685</i>	2.35074018	3.5E-06
<i>femR</i>	2.69910711	1.5E-05	<i>pslK</i>	2.35055153	5.5E-05
<i>PA2913</i>	2.68263344	1.2E-18	<i>arnD</i>	2.34465754	3.4E-05
<i>PA4317</i>	2.67844547	7.1E-37	<i>PA4492</i>	2.33617303	3E-48
<i>PA3483</i>	2.6758487	1.2E-15	<i>PA3307</i>	2.33469198	0.00428
<i>PA0092</i>	2.65617368	1.8E-20	<i>PA1645</i>	2.33396458	4.3E-05
<i>PA4490</i>	2.65117488	6E-45	<i>PA2368</i>	2.33292208	1.1E-10
<i>PA0123</i>	2.64902055	4.8E-18	<i>PA3727</i>	2.32704377	3.9E-22
<i>PA0069</i>	2.64341491	3.2E-10	<i>PA4707</i>	2.32658601	5.7E-14
<i>PA4489</i>	2.63602285	3.1E-93	<i>glpD</i>	2.32429538	2.7E-25
<i>bkdA2</i>	2.62436259	4.5E-13	<i>PA2541</i>	2.32041274	5.4E-52
<i>PA3564</i>	2.62360559	0.00066	<i>PA2406</i>	2.31807082	7.4E-10
<i>fepG</i>	2.62250709	5.3E-06	<i>PA2560</i>	2.31460116	2.5E-23
<i>pfeR</i>	2.60153969	3.7E-13	<i>PA2537</i>	2.30101757	1.1E-39
<i>PA0839</i>	2.58643935	1E-11	<i>arnT</i>	2.29801327	1.6E-08
<i>PA5114</i>	2.58493879	1.1E-16	<i>PA2362</i>	2.28998155	4.4E-13
<i>folK</i>	2.57493138	3.7E-06	<i>PA3964</i>	2.28991634	0.0009
<i>arnC</i>	2.56473232	1.6E-13	<i>PA2288</i>	2.28612192	0.0001
<i>PA3728</i>	2.56398886	4.3E-47	<i>PA4516</i>	2.28160589	8.3E-20
<i>PA2702</i>	2.55847564	5E-16	<i>PA1112.1</i>	2.27711288	1.3E-13
<i>PA2684</i>	2.55799126	4.5E-20	<i>PA4331</i>	2.27679773	4.5E-05
<i>PA4045</i>	2.5472464	5.4E-08	<i>nalC</i>	2.26078195	3.8E-26
<i>PA3936</i>	2.54364068	5E-06	<i>PA3021</i>	2.26057977	4.4E-33
<i>PA3448</i>	2.53136507	9.5E-05	<i>PA3067</i>	2.24377253	1.8E-06
<i>PA4884</i>	2.51351033	0.00028	<i>PA5113</i>	2.24360301	8.8E-12
<i>PA3443</i>	2.50794358	7.9E-15	<i>PA2369</i>	2.23431641	1.3E-26
<i>ribB</i>	2.50632399	3.2E-11	<i>PA0142</i>	2.23379982	0.0003
<i>PA2929</i>	2.50455617	1.1E-24	<i>PA4895</i>	2.2310692	3.6E-06
<i>PA3883</i>	2.50079835	0.00097	<i>PA4322</i>	2.22996745	1.4E-24
<i>PA3850</i>	2.47944595	2.5E-35	<i>PA3208</i>	2.22984483	3.7E-05
<i>PA2947</i>	2.44365466	0.00049	<i>PA2203</i>	2.22749564	4.5E-11

Table S27 (continued). Genes up-regulated in $\Delta PA0012$ strain.

Gene	Fold Change	padj	Gene	Fold Change	padj
<i>PA3449</i>	2.22584	1.6E-08	<i>PA0570</i>	2.06466	0.00162
<i>PA4871</i>	2.22426	0.00175	<i>PA2789</i>	2.06331	1.5E-05
<i>PA2365</i>	2.21339	3.7E-46	<i>PA4849</i>	2.06249	0.00029
<i>PA1364</i>	2.2107	2.8E-08	<i>cysW</i>	2.0611	2.5E-27
<i>PA3939</i>	2.20781	1.8E-05	<i>PA5273</i>	2.05854	8.5E-15
<i>PA5183</i>	2.20457	0.0002	<i>PA4668</i>	2.05474	1.7E-11
<i>PA0073</i>	2.20278	0.00842	<i>PA2912</i>	2.05249	5.5E-08
<i>PA3008</i>	2.19732	3.2E-25	<i>PA4280.5</i>	2.049	0.03181
<i>PA4894</i>	2.19495	0.00339	<i>PA2914</i>	2.04609	5.6E-05
<i>fepD</i>	2.1761	0.0007	<i>PA3317</i>	2.03741	1.2E-11
<i>PA2370</i>	2.16829	1.4E-18	<i>PA1844</i>	2.03568	4.7E-05
<i>PA0539</i>	2.16581	0.00012	<i>PA3900</i>	2.03539	4.9E-05
<i>PA1676</i>	2.1633	0.01118	<i>PA0680</i>	2.03433	0.02907
<i>PA2503</i>	2.15975	3.7E-10	<i>cobM</i>	2.02631	5.2E-07
<i>PA1451</i>	2.15119	8.3E-11	<i>PA2449</i>	2.02393	6.6E-13
<i>PA4883</i>	2.14482	0.01225	<i>PA0238</i>	2.02252	0.01182
<i>foxR</i>	2.14282	4.1E-07	<i>ispA</i>	2.01919	0.00011
<i>PA1394</i>	2.14182	0.01367	<i>pgl</i>	2.01762	4E-05
<i>PA4220</i>	2.1417	0.0007	<i>ilvI</i>	2.01742	3.8E-19
<i>napF</i>	2.14033	0.00155	<i>PA2366</i>	2.01602	9E-32
<i>PA0050</i>	2.14016	5E-10	<i>PA5441</i>	2.01354	1.5E-39
<i>PA3722</i>	2.1341	6.7E-14	<i>PA4610</i>	2.01316	0.02118
<i>PA0530</i>	2.13156	2E-08	<i>PA2775</i>	2.01289	1.4E-10
<i>cysA</i>	2.12689	2.5E-21	<i>PA1495</i>	2.01109	0.0168
<i>PA5071</i>	2.10597	2.7E-05	<i>rrmA</i>	2.00856	1.5E-07
<i>PA3726</i>	2.10272	1.3E-06	<i>pchC</i>	2.00803	1.7E-09
<i>PA2606</i>	2.10103	1.5E-06	<i>pabC</i>	2.00793	2.9E-05
<i>PA2724</i>	2.10068	0.02356	<i>PA2627</i>	2.00742	1.9E-07
<i>PA2358</i>	2.09465	7.8E-09	<i>PA1547</i>	2.00641	0.00468
<i>arnB</i>	2.09298	2.9E-23	<i>PA1657</i>	2.00288	4.9E-08
<i>PA2746</i>	2.09131	7E-42	<i>PA0471</i>	2.00169	4.3E-12
<i>PA3490</i>	2.08764	0.01003			
<i>PA4323</i>	2.0871	3.5E-08			
<i>PA5289</i>	2.08574	2.8E-14			
<i>PA4064</i>	2.08512	2E-19			
<i>thiD</i>	2.08494	1.8E-07			
<i>glpK</i>	2.08286	4.2E-31			
<i>PA0671</i>	2.08178	0.00813			
<i>PA1149</i>	2.07019	0.01262			

Table S28. Genes down-regulated in $\Delta PA0012$ strain.

Gene	Fold change	padj	Gene	Fold change	padj
PA0012	-53.17092921	7.9178E-107	PA0713	-3.032571914	3.86374E-20
PA4738	-11.55220474	1.3287E-195	<i>glgB</i>	-3.0191111	1.27929E-30
PA4739	-10.00558618	1.3035E-187	PA2150	-3.003135441	3.43427E-12
PA2166	-8.059697181	7.5785E-128	<i>ibpA</i>	-2.958609144	2.35737E-60
<i>pfpl</i>	-5.895329199	4.5257E-134	PA3819	-2.941034789	5.39926E-66
PA5482	-5.685945647	1.86003E-76	PA2154	-2.940177638	3.19151E-16
<i>osmE</i>	-5.516051467	4.3683E-158	PA3370	-2.916650393	3.68822E-18
PA3274	-5.479793582	6.64354E-39	PA4877	-2.844430085	2.51527E-13
PA3273	-4.738233073	8.55699E-17	PA2169	-2.815804237	2.33167E-16
PA2134	-4.640686074	6.36758E-41	PA2167	-2.810125483	1.23583E-19
PA1404	-4.548640866	1.19212E-45	<i>mucA</i>	-2.807093412	1.49249E-62
<i>osmC</i>	-4.523111178	9.5321E-99	<i>imm2</i>	-2.781913901	5.97499E-10
PA3231	-4.17245296	5.05987E-18	PA2190	-2.771523117	1.6368E-16
PA1324	-4.161067719	1.4124E-165	PA2160	-2.770973157	1.49446E-48
PA0567	-3.957892999	6.47785E-52	PA5212	-2.762436503	1.10197E-43
PA4141	-3.948920001	2.53297E-42	<i>phrS</i>	-2.74872797	4.62356E-14
PA1323	-3.897036623	2.81694E-66	PA2754	-2.730001849	6.56086E-29
PA2184	-3.887644658	1.28288E-41	PA3040	-2.70798069	3.32139E-29
<i>chiC</i>	-3.721136387	3.41279E-66	PA0346	-2.689611973	7.25037E-17
PA5481	-3.639050489	4.02346E-59	PA1111	-2.68383674	6.95916E-08
PA2171	-3.57742261	3.94325E-20	PA2377	-2.663774189	1.16925E-52
PA3432	-3.56271989	5.96429E-52	PA2485	-2.63937476	1.40964E-20
PA2176	-3.536560402	1.24253E-19	PA3334	-2.636247537	2.50144E-35
PA5061	-3.528159234	7.3073E-45	PA0572	-2.625274629	4.06158E-31
<i>lptF</i>	-3.527758478	6.97215E-77	<i>ahpC</i>	-2.572334849	1.78121E-32
PA2433	-3.462854539	8.8895E-53	PA0329	-2.564198893	4.22322E-23
PA3691	-3.389767789	7.66256E-89	PA4139	-2.550782412	2.10622E-14
PA1324.1	-3.387804184	3.3391E-08	PA0941	-2.535194188	4.1555E-18
PA3529	-3.376835448	4.0322E-115	PA1606	-2.518399768	5.82169E-10
<i>groES</i>	-3.338103516	1.08522E-72	<i>htpG</i>	-2.513264642	7.88753E-38
PA3891	-3.329057785	8.93209E-33	<i>azu</i>	-2.509251724	2.64642E-24
PA1592	-3.309313814	2.91653E-68	PA1221	-2.495037998	3.43139E-09
<i>aprA</i>	-3.281943594	4.94842E-53	<i>dnaK</i>	-2.474927312	1.62745E-48
PA4880	-3.234651898	2.11252E-45	PA0962	-2.468033341	2.31715E-35
PA2777	-3.177300643	1.44569E-30	PA2165	-2.467389574	5.07693E-30
PA2422	-3.156319359	1.02594E-12	PA2180	-2.460854898	5.29551E-08
PA0039	-3.137487166	4.86673E-35	PA2021	-2.450450168	5.03378E-07
PA1830	-3.131362319	2.76295E-27	PA3952	-2.448516537	1.87761E-17
<i>groEL</i>	-3.081759153	5.0334E-126	PA3889	-2.440659802	2.3933E-19
PA3431	-3.078603349	9.66975E-30	<i>algU</i>	-2.436806222	8.61388E-45
PA2146	-3.07685545	1.99722E-09	PA2151	-2.428222507	2.60784E-14
PA2747	-3.057986272	3.32998E-87	PA4172	-2.419367897	1.93023E-05

Table S28 (continued). Genes down-regulated in $\Delta PA0012$ strain.

Gene	Fold change	padj	Gene	Fold change	padj
PA3890	-2.415919044	9.23722E-13	<i>pscF</i>	-2.160720643	3.15482E-05
PA0623	-2.410341512	2.18751E-16	PA0634	-2.153641955	1.67097E-16
PA4674	-2.406193371	2.01476E-24	PA4611	-2.151229791	2.13447E-12
PA3284	-2.404159298	4.0461E-09	PA0315	-2.146767483	1.24482E-17
PA3888	-2.384357529	2.12107E-10	PA1114	-2.146761237	0.000150376
PA0940	-2.382918927	1.67097E-16	PA3795	-2.140414304	2.40698E-47
PA3144	-2.3795796	3.16829E-06	PA4394	-2.139483344	1.55133E-24
PA3496	-2.376087869	5.03816E-11	PA0627	-2.137928319	1.54431E-08
PA2381	-2.336970473	5.95771E-14	PA0038	-2.134807021	1.52329E-34
PA5424	-2.333259821	1.46543E-11	<i>pslN</i>	-2.130703385	0.000118318
<i>lasA</i>	-2.328628503	2.09551E-36	PA0613	-2.12281382	1.13943E-11
<i>cynS</i>	-2.324781089	7.87268E-11	PA3041	-2.117697911	3.84115E-47
<i>clpB</i>	-2.320021477	3.6287E-36	<i>hupB</i>	-2.117671244	1.21465E-21
PA3461	-2.31862454	5.37189E-39	<i>grpE</i>	-2.117453543	1.49392E-16
<i>poxB</i>	-2.318366386	3.68674E-18	PA2182	-2.109762683	1.02329E-09
PA0642	-2.314528994	7.08134E-08	PA0490	-2.102418538	1.94376E-12
PA3278	-2.312514396	6.14218E-11	PA0633	-2.095647247	2.45632E-18
PA1545	-2.278625147	1.24054E-14	PA1516	-2.093540553	1.55965E-10
PA3369	-2.27841244	1.12012E-14	PA2531	-2.091360633	3.41977E-37
PA1535	-2.278004536	6.56327E-22	PA1617	-2.090123938	4.92076E-23
<i>pcaC</i>	-2.27512175	3.59902E-16	PA1245	-2.089541932	4.0521E-36
PA1244	-2.271409401	3.0456E-34	PA0648	-2.087516403	1.85393E-09
PA2708	-2.270407996	4.12826E-17	<i>oprl</i>	-2.086132832	4.07227E-18
PA4797	-2.267389072	0.012671871	PA0747	-2.085382628	7.26138E-30
PA2168	-2.253843782	9.33119E-07	PA3140	-2.082776051	1.01573E-13
<i>hasA</i>	-2.243836743	1.30989E-27	PA0714	-2.072604169	1.66077E-07
PA0573	-2.243755526	5.97311E-11	<i>pvdJ</i>	-2.070693371	3.25821E-58
PA1348	-2.226001227	1.74754E-19	PA2698	-2.063641361	3.74243E-10
PA0060	-2.21829078	1.95908E-12	PA2414	-2.056214421	3.36153E-16
PA5178	-2.215207034	2.80876E-26	<i>opdB</i>	-2.054796286	4.40626E-13
PA2486	-2.211363923	5.95436E-07	<i>mexG</i>	-2.053282979	1.90121E-07
PA3332	-2.204386671	3.36415E-28	PA4173	-2.052133716	0.024046825
PA0776	-2.203748851	3.45588E-11	PA1330	-2.050608602	6.77565E-09
<i>gloA3</i>	-2.203123211	3.14217E-27	PA0449	-2.048678375	6.35932E-13
PA2152	-2.202546479	9.95356E-17	PA1242	-2.048428324	4.10902E-07
PA2158	-2.201505011	1.28936E-08	PA2751	-2.040568025	2.4734E-11
PA2136	-2.201470248	0.00870476	PA2159	-2.03631096	2.51294E-05
PA0745	-2.185040286	1.98605E-19	PA2575	-2.035965156	3.82718E-15
PA2149	-2.184866566	0.012209551	PA2161	-2.032180406	0.015174301
<i>rnpB</i>	-2.178937404	1.95042E-21	PA0647	-2.025356136	4.85933E-09
PA2313	-2.170605748	6.00549E-12	PA2453	-2.021270969	2.6675E-08
<i>cupC2</i>	-2.168707208	0.005953499	PA3460	-2.01732108	6.4224E-18

Table S28 (continued). Genes down-regulated in $\Delta PA0012$ strain

Gene	Fold change	padj
<i>exoS</i>	-2.011011741	1.57266E-25
<i>rsmA</i>	-2.008663443	0.000124715
<i>PA2143</i>	-2.00741398	4.45623E-05
<i>PA2918</i>	-2.006051766	1.44606E-38
<i>PA2274</i>	-2.005015424	0.000150668
<i>PA4171</i>	-2.00468392	3.64853E-10

Table S29. Genes up-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>phnA</i>	32.49044772	2.3483E-280	<i>PA0332</i>	3.32627713	9.4E-13
<i>pqsB</i>	32.32593913	1.5337E-164	<i>phzD1</i>	3.32525013	0.00013
<i>pqsA</i>	29.89633452	2.6803E-285	<i>PA5135</i>	3.32336772	3.3E-07
<i>pqsE</i>	26.46024625	2.6803E-285	<i>exsE</i>	3.31182024	1.7E-17
<i>pqsD</i>	20.3743102	0	<i>PA3428</i>	3.29468015	1.5E-05
<i>pqsC</i>	18.25512636	1.4172E-232	<i>PA0671</i>	3.26072993	1.8E-06
<i>PA4325</i>	17.49656904	7.0613E-215	<i>pcrR</i>	3.25588998	7.1E-05
<i>phzG2</i>	6.807380617	5.06303E-14	<i>PA3444</i>	3.25220293	3.8E-20
<i>phzC2</i>	5.673725585	3.28907E-09	<i>PA3448</i>	3.21816755	1.8E-07
<i>PA0123</i>	5.302089929	3.23469E-52	<i>bkdA1</i>	3.20952152	7E-32
<i>PA4142</i>	4.931301964	1.05101E-59	<i>PA4155</i>	3.20603607	2E-28
<i>phzF1</i>	4.575135283	6.12849E-07	<i>PA4818</i>	3.2057681	6E-05
<i>PA3415</i>	4.567868871	3.24328E-15	<i>PA4917</i>	3.18573079	7.2E-44
<i>phzE1</i>	4.418702808	1.93493E-06	<i>PA2433</i>	3.15971454	5.9E-60
<i>phzE2</i>	4.378968347	2.04285E-06	<i>PA1906</i>	3.1270875	1.8E-18
<i>PA4916</i>	4.238951935	4.10593E-30	<i>pntAA</i>	3.12020952	2.5E-14
<i>PA0878</i>	4.115888802	2.12932E-12	<i>PA3429</i>	3.1086927	1.2E-11
<i>PA1052</i>	4.102972985	4.2864E-10	<i>PA4323</i>	3.08180342	2.5E-18
<i>PA3928</i>	3.888182411	1.34408E-11	<i>phzD2</i>	3.06849309	0.00028
<i>phzF2</i>	3.789080967	2.1762E-05	<i>PA2268</i>	3.05337829	1.3E-05
<i>PA3416</i>	3.738990655	9.61205E-24	<i>pcrG</i>	3.02966879	0.00042
<i>PA4889</i>	3.722591771	6.56981E-06	<i>PA3939</i>	3.02293659	2.9E-10
<i>PA3414</i>	3.721798976	2.75749E-10	<i>bkdA2</i>	3.02170618	1.4E-17
<i>PA3564</i>	3.709269798	7.58422E-07	<i>PA2170</i>	3.00861716	3.8E-05
<i>PA0468</i>	3.664560572	3.14063E-13	<i>PA3442</i>	3.00574009	2E-08
<i>PA4326</i>	3.651732318	9.33621E-33	<i>pslK</i>	2.96389116	5.4E-08
<i>phnB</i>	3.649017164	4.604E-136	<i>PA0813</i>	2.95138457	4E-10
<i>fepD</i>	3.645395075	1.05271E-09	<i>femR</i>	2.95100576	1.3E-06
<i>PA2655</i>	3.568645426	3.46433E-06	<i>PA3417</i>	2.93855264	1.6E-19
<i>PA3307</i>	3.529110977	3.69205E-06	<i>pfeR</i>	2.91807856	2E-16
<i>phzA2</i>	3.482005375	8.12687E-08	<i>PA0578</i>	2.90165825	9.5E-11
<i>PA3936</i>	3.479937627	1.13014E-10	<i>PA3298</i>	2.89217062	5.1E-05
<i>PA0451</i>	3.479042034	1.06498E-18	<i>PA1251</i>	2.88578854	7.1E-10
<i>ptrB</i>	3.419697553	8.87344E-05	<i>PA2069</i>	2.88098759	2.7E-29
<i>eraS</i>	3.415579487	2.48864E-05	<i>PA2109</i>	2.8535101	1.7E-05
<i>fepG</i>	3.410126239	1.2087E-09	<i>PA4144</i>	2.83558188	3.2E-08
<i>bioD</i>	3.381422966	5.85531E-06	<i>PA5027</i>	2.83465829	2.3E-09
<i>nrdG</i>	3.36709755	1.07817E-08	<i>moeB</i>	2.82953969	1.1E-12

Table S29 (continued). Genes up-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>PA1547</i>	2.82871	3.6E-06	<i>PA2179</i>	2.54770417	3.7E-08
<i>PA0380</i>	2.80366859	0.00017	<i>PA0112</i>	2.54484708	5.6E-07
<i>PA3354</i>	2.78176727	6.4E-16	<i>pcrH</i>	2.54161745	2.7E-11
<i>PA0539</i>	2.77544764	6.1E-08	<i>PA0881</i>	2.53721968	0.00181
<i>PA3208</i>	2.77456686	3.1E-08	<i>PA0345</i>	2.53006452	4.5E-06
<i>PA4045</i>	2.77197912	7.1E-10	<i>PA0144</i>	2.51689329	0.00032
<i>PA2173</i>	2.7660788	8.6E-05	<i>opmD</i>	2.5061889	0.00029
<i>PA0850</i>	2.76054815	0.00013	<i>PA2067</i>	2.50397008	3.8E-06
<i>PA2287</i>	2.7479302	3.4E-08	<i>PA4766</i>	2.49390569	2.9E-10
<i>PA3957</i>	2.73145129	2.4E-19	<i>PA5357</i>	2.49055185	5.3E-11
<i>pdxB</i>	2.71979328	1.8E-06	<i>PA3973</i>	2.48425196	6.2E-09
<i>PA4895</i>	2.71582007	2.1E-09	<i>lpdV</i>	2.47563992	9.3E-07
<i>PA2135</i>	2.71550452	3.6E-07	<i>PA3490</i>	2.46818453	0.00066
<i>PA0791</i>	2.69506025	4.9E-11	<i>glpF</i>	2.46663651	1.1E-35
<i>PA2689</i>	2.69280564	6.5E-07	<i>waaA</i>	2.45512747	4.7E-06
<i>PA1676</i>	2.686355	0.00042	<i>PA4871</i>	2.45502478	0.00017
<i>PA2110</i>	2.68166536	0.0016	<i>PA2919</i>	2.45452875	4E-07
<i>PA3426</i>	2.67551474	1.9E-08	<i>PA0879</i>	2.45256712	9E-11
<i>PA1732</i>	2.66817345	3E-12	<i>PA1945</i>	2.45079733	7.9E-08
<i>PA4111</i>	2.66751609	3.5E-10	<i>PA4894</i>	2.44818121	0.00043
<i>PA4143</i>	2.65426113	7.7E-23	<i>PA1147</i>	2.43927968	4.3E-09
<i>PA4331</i>	2.65007611	4E-07	<i>PA3275</i>	2.43738874	1.3E-05
<i>PA0880</i>	2.64346821	0.00205	<i>PA1470</i>	2.43231414	2.4E-08
<i>PA1261</i>	2.64252943	4.9E-08	<i>PA0110</i>	2.43122756	2.7E-10
<i>PA0562</i>	2.64091979	4E-06	<i>ambE</i>	2.42116293	1E-18
<i>PA2500</i>	2.63938434	1.6E-06	<i>pcaQ</i>	2.42083575	9.6E-07
<i>nosF</i>	2.62506044	0.00034	<i>PA0274</i>	2.42074004	0.00014
<i>PA0142</i>	2.62357332	3.7E-06	<i>PA3316</i>	2.4151437	5.7E-10
<i>PA3883</i>	2.62305991	0.00025	<i>PA1220</i>	2.39166087	0.00407
<i>PA3971</i>	2.62224794	2.7E-08	<i>PA0756</i>	2.38632444	0.00027
<i>PA1356</i>	2.62053996	1.3E-08	<i>PA1876</i>	2.38435099	3E-10
<i>PA0787</i>	2.60284182	5.4E-11	<i>PA2141</i>	2.37691549	0.00117
<i>PA4018</i>	2.60136837	1.4E-05	<i>PA5343</i>	2.37058353	3.1E-06
<i>popD</i>	2.58478708	1.5E-34	<i>pslH</i>	2.36631274	6.7E-13
<i>PA3845</i>	2.57662857	3.1E-08	<i>mdcE</i>	2.35803885	0.00425
<i>PA0173</i>	2.57114687	1.2E-08	<i>PA2292</i>	2.35024941	0.0054
<i>PA2872</i>	2.56939766	7.6E-06	<i>alkA</i>	2.34689539	3.1E-12
<i>PA1700</i>	2.55952431	0.00344	<i>PA0680</i>	2.34404496	0.00441
<i>pslF</i>	2.55536123	2.8E-12	<i>PA2480</i>	2.341792	2.1E-05

Table S29 (continued). Genes up-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>PA0631</i>	2.33906205	3.7E-05	<i>PA4096</i>	2.21760094	0.00968
<i>PA2753</i>	2.33688895	0.00028	<i>pgl</i>	2.21161958	1.8E-06
<i>PA5446</i>	2.3302619	2.7E-20	<i>mtlZ</i>	2.20971466	0.00116
<i>PA2330</i>	2.32917795	0.00199	<i>PA0238</i>	2.20962942	0.00227
<i>PA4059</i>	2.32760498	0.00019	<i>PA3090</i>	2.20872773	0.00039
<i>vanB</i>	2.32694556	0.00041	<i>PA0882</i>	2.20456047	0.00022
<i>PA0202</i>	2.32670839	4.5E-14	<i>PA0842</i>	2.20100446	0.00059
<i>PA3677</i>	2.32595931	0.00041	<i>PA2898</i>	2.20055412	0.00702
<i>PA2288</i>	2.32391805	3.4E-05	<i>pcaG</i>	2.19850132	8.7E-09
<i>PA3067</i>	2.32278784	2.3E-07	<i>PA4610</i>	2.19010322	0.00524
<i>PA0790</i>	2.32142507	0.00194	<i>desT</i>	2.18666271	0.00031
<i>PA2070</i>	2.31853586	4.6E-20	<i>PA4350</i>	2.18575302	1.4E-06
<i>PA2606</i>	2.3171168	1.5E-08	<i>PA1109</i>	2.18485952	1.1E-06
<i>PA0510</i>	2.31601808	0.00926	<i>pmtA</i>	2.18474769	1E-12
<i>PA5407</i>	2.31447555	5.8E-06	<i>PA2778</i>	2.17859164	1.1E-10
<i>PA0033</i>	2.31025844	4.4E-06	<i>PA5116</i>	2.17823314	0.00229
<i>PA0700</i>	2.30965634	0.00912	<i>PA3786</i>	2.17758868	2.4E-06
<i>PA2933</i>	2.30600778	0.00893	<i>PA0377</i>	2.1759762	4.2E-05
<i>PA4220</i>	2.29885934	0.00013	<i>PA0753</i>	2.17527667	0.01079
<i>PA1740</i>	2.29322731	1.8E-06	<i>exsB</i>	2.17484013	1.3E-16
<i>PA5408</i>	2.29121795	0.00175	<i>foxR</i>	2.17477373	1.4E-07
<i>PA5149</i>	2.29049435	4.1E-05	<i>PA3207</i>	2.16967155	0.00218
<i>PA0209</i>	2.2895843	0.01004	<i>PA0503</i>	2.1684059	6.9E-05
<i>PA4365</i>	2.28773026	0.00118	<i>PA3203</i>	2.16753261	2.8E-07
<i>PA0311</i>	2.28527905	0.00448	<i>PA1933</i>	2.16642033	2.5E-11
<i>PA1645</i>	2.27981575	3.8E-05	<i>PA1932</i>	2.16290989	1.5E-07
<i>PA0630</i>	2.27371426	1.7E-06	<i>glpK</i>	2.16154227	9.7E-35
<i>PA2750</i>	2.26950615	0.00068	<i>PA0163</i>	2.15774002	1.8E-06
<i>PA1465</i>	2.26670586	6.2E-06	<i>PA2406</i>	2.15667126	1.1E-08
<i>phzB2</i>	2.25751129	2.3E-06	<i>PA0386</i>	2.15252771	2.5E-05
<i>PA2066</i>	2.25626733	1.3E-09	<i>PA2163</i>	2.15209348	2E-06
<i>PA4881</i>	2.2452124	0.00264	<i>vreA</i>	2.14888737	0.0004
<i>napF</i>	2.24420588	0.00037	<i>PA2439</i>	2.14647819	4.6E-05
<i>PA0168</i>	2.24326364	0.00045	<i>PA3314</i>	2.14528153	8.9E-05
<i>PA1238</i>	2.23962623	0.00199	<i>nirM</i>	2.14524782	0.01467
<i>dgcB</i>	2.23178939	2.6E-05	<i>PA1274</i>	2.14432587	0.0068
<i>PA1495</i>	2.22545973	0.0033	<i>PA1112.1</i>	2.14275436	6.6E-12
<i>PA0816</i>	2.22400872	2.5E-05	<i>PA4134</i>	2.14097014	0.00415
<i>xcpW</i>	2.22189016	2.6E-07	<i>hutH</i>	2.13906261	0.00017

Table S29 (continued). Genes up-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>stk1</i>	2.13534387	2.1E-05	<i>PA2324</i>	2.07055519	0.00857
<i>PA2534</i>	2.13459556	3.6E-07	<i>PA2947</i>	2.07003601	0.00369
<i>PA3259</i>	2.13371467	6.5E-07	<i>PA2746</i>	2.06991107	2.3E-41
<i>PA0864</i>	2.13274885	2.8E-05	<i>PA0419</i>	2.06840765	3.9E-05
<i>stp1</i>	2.13174463	0.00048	<i>PA1149</i>	2.06826994	0.00797
<i>PA1601</i>	2.13170002	0.00062	<i>PA4509</i>	2.06672315	0.00016
<i>PA0530</i>	2.1313784	9.4E-09	<i>rfaD</i>	2.06624921	1E-19
<i>PA4978</i>	2.13133042	0.00028	<i>mdcD</i>	2.06602152	0.00765
<i>PA3911</i>	2.12843818	0.01479	<i>PA1262</i>	2.06463353	0.0066
<i>phzA1</i>	2.12776721	0.02058	<i>PA3943</i>	2.06385357	1.2E-10
<i>PA4659</i>	2.12664104	2.7E-08	<i>PA0029</i>	2.0632355	0.00188
<i>PA2478</i>	2.12605954	6.9E-05	<i>PA0639</i>	2.06239687	1.8E-09
<i>PA2159</i>	2.12600677	1.8E-07	<i>ispA</i>	2.05719437	3.1E-05
<i>thiD</i>	2.12239446	3.5E-08	<i>PA4873</i>	2.05578606	5.7E-08
<i>PA3065</i>	2.11837472	0.00029	<i>PA5400</i>	2.05472122	0.01999
<i>PA2315</i>	2.11739974	0.00621	<i>PA4677</i>	2.05307288	0.0002
<i>PA2984</i>	2.11691156	7E-07	<i>PA3684</i>	2.05264315	8.1E-05
<i>bfiS</i>	2.11685678	5.1E-06	<i>PA3685</i>	2.05227379	6.9E-05
<i>PA1616</i>	2.11602139	2.1E-06	<i>PA3175</i>	2.05223051	6.6E-05
<i>PA5085</i>	2.11410922	2E-07	<i>PA1335</i>	2.0509509	2.3E-07
<i>PA1237</i>	2.1131965	0.01407	<i>popN</i>	2.05080604	0.00616
<i>PA2046</i>	2.10667965	0.00943	<i>tadD</i>	2.04915854	0.00436
<i>PA2307</i>	2.1042768	2E-05	<i>PA2116</i>	2.04852027	0.00151
<i>PA2143</i>	2.1039069	2.3E-08	<i>antC</i>	2.04718754	0.00023
<i>PA0072</i>	2.1038114	0.00158	<i>cpg2</i>	2.03976834	4.3E-15
<i>PA2084</i>	2.10158153	3.5E-07	<i>PA3846</i>	2.0381086	1.3E-11
<i>PA5211</i>	2.09910713	1.7E-06	<i>PA3920</i>	2.03647029	3.5E-06
<i>str</i>	2.09869225	1.7E-05	<i>PA4635</i>	2.0363279	0.0008
<i>PA3669</i>	2.09585616	3.5E-09	<i>PA3576</i>	2.03326496	1.9E-09
<i>PA2660</i>	2.09356126	0.00023	<i>PA4966</i>	2.03295367	1.3E-05
<i>phzG1</i>	2.08703319	0.00262	<i>PA0333</i>	2.03128793	3E-06
<i>PA4362</i>	2.08441008	8.9E-08	<i>PA4004</i>	2.03084358	3.9E-06
<i>PA4219</i>	2.08418359	3.1E-05	<i>pabC</i>	2.03018052	8.6E-06
<i>pseE</i>	2.08341288	0.00188	<i>PA4062</i>	2.02991255	2.8E-05
<i>PA2769</i>	2.08183071	0.00015	<i>PA2374</i>	2.0296969	3.2E-07
<i>PA1566</i>	2.07982092	0.00046	<i>PA1332</i>	2.02958717	8.1E-06
<i>dadX</i>	2.07475725	1E-17	<i>PA0446</i>	2.02694936	8.2E-15
<i>aer2</i>	2.07344798	1.2E-20	<i>PA4618</i>	2.02677739	0.0004
<i>PA0136</i>	2.07282345	0.00549	<i>rbsA</i>	2.02625393	0.00206

Table S29 (continued). Genes up-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj
<i>PA1351</i>	2.02504588	1.9E-05
<i>PA1884</i>	2.02401498	0.0028
<i>fliO</i>	2.02118483	6.8E-06
<i>PA0369</i>	2.02085218	0.02152
<i>PA4149</i>	2.01428465	0.00124
<i>PA2408</i>	2.01390242	0.00624
<i>lecB</i>	2.01084161	1.9E-22
<i>PA2068</i>	2.01008956	5.5E-11
<i>phzB1</i>	2.00937864	0.00012
<i>PA2680</i>	2.0088443	0.01282
<i>xcpU</i>	2.00697803	4.1E-05
<i>nirJ</i>	2.00608587	0.00329
<i>PA0777</i>	2.00487606	0.01704
<i>PA3125</i>	2.00387232	3.6E-05
<i>folK</i>	2.002171	0.00066
<i>PA2172</i>	2.00043244	2.6E-06

Table S30. Genes down-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>phhA</i>	-2.000930052	2.04118E-23	<i>PA0264</i>	-2.055773	2.3E-19
<i>PA1143</i>	-2.00200257	0.020212979	<i>aceA</i>	-2.0565777	1.2E-29
<i>PA3451</i>	-2.002458226	8.49059E-10	<i>PA0392</i>	-2.0587357	5E-18
<i>sucD</i>	-2.002675453	1.50331E-27	<i>PA2372</i>	-2.0598041	4.1E-06
<i>pvdN</i>	-2.003746678	6.29485E-18	<i>PA2367</i>	-2.0628211	5.6E-14
<i>PA1935</i>	-2.007112355	1.82478E-09	<i>PA4451</i>	-2.0633848	1.7E-08
<i>fppA</i>	-2.00824113	0.000247136	<i>htpG</i>	-2.0644111	8.2E-25
<i>rplS</i>	-2.009377058	3.80654E-22	<i>ilvI</i>	-2.0658024	8.7E-20
<i>PA3510</i>	-2.011425736	0.000939913	<i>PA3370</i>	-2.0665564	1.1E-10
<i>tatA</i>	-2.013431285	3.4556E-10	<i>pilY2</i>	-2.066895	1.1E-12
<i>rubA1</i>	-2.014022619	5.91861E-06	<i>sucB</i>	-2.0714393	2.8E-26
<i>PA1033</i>	-2.014970607	6.83844E-18	<i>himA</i>	-2.0715658	1.6E-15
<i>PA0160</i>	-2.016370917	1.6046E-05	<i>rpsR</i>	-2.0722434	3E-11
<i>rpoS</i>	-2.018027429	3.34342E-15	<i>PA3119</i>	-2.0743522	1.3E-15
<i>PA5061</i>	-2.019237801	1.08841E-14	<i>PA0262</i>	-2.0755202	1.9E-22
<i>pilC</i>	-2.020140718	2.85842E-14	<i>pldA</i>	-2.0775246	1.1E-23
<i>PA2101</i>	-2.020290324	5.80186E-08	<i>algP</i>	-2.081997	4.6E-23
<i>PA3730</i>	-2.021515895	1.03427E-12	<i>rplL</i>	-2.0825952	1.2E-16
<i>PA2591</i>	-2.022055269	4.58254E-12	<i>PA0567</i>	-2.0867914	1.2E-17
<i>leuA</i>	-2.023965077	4.82196E-21	<i>rpsQ</i>	-2.0872642	5.7E-24
<i>rplK</i>	-2.024392777	4.4506E-17	<i>PA0315</i>	-2.0897312	1.6E-16
<i>PA4671</i>	-2.025251208	4.43012E-16	<i>PA4066</i>	-2.0914902	1.9E-27
<i>PA4335</i>	-2.02590863	1.42437E-26	<i>PA5226</i>	-2.0991792	1.6E-08
<i>azu</i>	-2.026580904	1.12846E-14	<i>PA2730</i>	-2.103881	1.1E-10
<i>PA3205</i>	-2.027765655	8.78899E-17	<i>rpmA</i>	-2.1066489	1.2E-12
<i>PA2604</i>	-2.028921715	1.2154E-16	<i>PA1842</i>	-2.1102477	3E-20
<i>rplQ</i>	-2.029312001	1.1295E-17	<i>folX</i>	-2.1120529	9.3E-18
<i>rnpB</i>	-2.032755552	7.68089E-18	<i>PA4782</i>	-2.1175661	6.4E-13
<i>ffs</i>	-2.034512243	2.78971E-10	<i>PA0604</i>	-2.1192897	9E-13
<i>gltA</i>	-2.036956827	1.70963E-12	<i>PA2313</i>	-2.1204392	7.9E-12
<i>pilG</i>	-2.037040944	5.22072E-11	<i>sucA</i>	-2.1221537	6.8E-40
<i>PA1369</i>	-2.037490611	9.07498E-08	<i>PA3085</i>	-2.127951	7.8E-12
<i>purC</i>	-2.039641322	1.88472E-19	<i>PA1579</i>	-2.1292858	1.7E-18
<i>PA1545</i>	-2.039820969	2.51872E-11	<i>PA0126</i>	-2.135207	8.5E-13
<i>PA3756</i>	-2.047565186	1.67933E-14	<i>infC</i>	-2.1364935	4.3E-11
<i>PA3326</i>	-2.04901187	1.17388E-21	<i>PA0046</i>	-2.1413827	1.9E-07
<i>ibpA</i>	-2.051889972	8.42637E-29	<i>rpmH</i>	-2.1413995	2.3E-11
<i>PA4697</i>	-2.055192543	4.3332E-09	<i>PA0042</i>	-2.1434374	4.8E-13
<i>PA5460</i>	-2.05540031	2.70727E-08	<i>PA0039</i>	-2.1439805	5.7E-16

Table S30 (continued). Genes down-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>PA1511</i>	-2.1441674	1.2E-16	<i>PA2826</i>	-2.2740973	2.3E-16
<i>fusA1</i>	-2.1451858	2.3E-34	<i>mvaT</i>	-2.2771299	3.1E-13
<i>galE</i>	-2.1452286	1.1E-06	<i>PA1517</i>	-2.2782505	1.9E-29
<i>PA2860</i>	-2.1487327	4.6E-18	<i>rplE</i>	-2.2868016	1.6E-20
<i>PA0722</i>	-2.1524295	1.6E-09	<i>PA3647</i>	-2.3090922	2E-19
<i>PA3998</i>	-2.1531722	1.2E-11	<i>ssrA</i>	-2.3118579	7.2E-14
<i>tufB</i>	-2.1561471	5E-28	<i>PA4390</i>	-2.313289	2.4E-21
<i>pvdA</i>	-2.1598347	1.6E-45	<i>rmf</i>	-2.3152372	3.1E-15
<i>wbpG</i>	-2.16015	4.2E-10	<i>PA1134</i>	-2.3232855	1.9E-28
<i>rplV</i>	-2.1662205	1.8E-19	<i>PA4272.1</i>	-2.3250268	1.9E-13
<i>PA4797</i>	-2.1671565	0.01316	<i>lysC</i>	-2.3288213	3.4E-29
<i>asd</i>	-2.1714772	3.4E-24	<i>PA2531</i>	-2.3319171	3E-49
<i>PA2083</i>	-2.1746735	5.5E-22	<i>rplX</i>	-2.3319424	1.4E-25
<i>PA1510</i>	-2.1768057	5.3E-18	<i>PA4638</i>	-2.3460864	6.3E-10
<i>PA2453</i>	-2.1772503	1.9E-10	<i>PA5285</i>	-2.3464957	4.1E-30
<i>ahpC</i>	-2.1791981	3.7E-22	<i>PA1372</i>	-2.3487018	4.5E-12
<i>rpsD</i>	-2.1955992	1.4E-24	<i>PA0720</i>	-2.3516502	8.2E-11
<i>rpmD</i>	-2.196807	5.8E-16	<i>PA2575</i>	-2.3544343	1.4E-21
<i>leuC</i>	-2.2126421	2.3E-42	<i>PA5301</i>	-2.35449	1.9E-17
<i>pys2</i>	-2.2250935	9.1E-13	<i>PA5191</i>	-2.357173	7.4E-16
<i>rpmC</i>	-2.2280644	5.5E-18	<i>mliC</i>	-2.3584424	4.7E-30
<i>pilE</i>	-2.229175	3.7E-16	<i>PA0727</i>	-2.3673823	2.4E-22
<i>PA4139</i>	-2.2293971	3.5E-12	<i>tolB</i>	-2.3724071	2.6E-23
<i>trxA</i>	-2.2294724	2.4E-16	<i>PA0435</i>	-2.3744572	2.2E-17
<i>rplN</i>	-2.2338269	1.2E-28	<i>PA0991</i>	-2.3832475	1E-09
<i>PA2935</i>	-2.2352422	0.00083	<i>PA1509</i>	-2.384791	1.3E-16
<i>PA1664</i>	-2.2431892	0.01438	<i>eco</i>	-2.3880323	3.8E-15
<i>PA0071</i>	-2.2453422	8.4E-15	<i>leuB</i>	-2.3880827	2.5E-39
<i>PA0329</i>	-2.2462792	2.7E-17	<i>PA2166</i>	-2.388397	2E-33
<i>PA1847</i>	-2.2468054	1.7E-18	<i>PA5088</i>	-2.4026394	9.2E-10
<i>PA1035</i>	-2.2492433	3.2E-19	<i>mscL</i>	-2.4083269	2.5E-15
<i>phhB</i>	-2.2528407	2E-17	<i>ppiB</i>	-2.4132266	6.9E-20
<i>PA1159</i>	-2.25584	3E-19	<i>PA3031</i>	-2.4150671	1.8E-14
<i>rsmA</i>	-2.2613931	3.7E-06	<i>PA4578</i>	-2.4151177	1.2E-37
<i>rpmB</i>	-2.2643095	6.4E-12	<i>rpsS</i>	-2.4238704	3.7E-16
<i>rpmJ</i>	-2.267173	3.3E-11	<i>PA2422</i>	-2.4243977	4.9E-09
<i>rpsC</i>	-2.2672667	1E-21	<i>pvdF</i>	-2.4248689	1.7E-28
<i>rgsA</i>	-2.2687105	1E-14	<i>pasP</i>	-2.4328296	2.1E-22
<i>PA5481</i>	-2.2700307	2E-25	<i>PA1533</i>	-2.4511633	1.4E-22

Table S30 (continued). Genes down-regulated in $\Delta PA4324$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>secB</i>	-2.4529941	4.7E-21	<i>PA1371</i>	-2.7728668	2.9E-15
<i>PA1518</i>	-2.4570471	8.7E-25	<i>acpP</i>	-2.7733432	1.7E-10
<i>PA0099</i>	-2.4574041	2.1E-13	<i>PA3931</i>	-2.7834447	5.1E-16
<i>PA2581</i>	-2.4623384	2.3E-31	<i>prfF1</i>	-2.8068156	1.6E-14
<i>PA5339</i>	-2.4790554	9.5E-27	<i>PA4773</i>	-2.8142656	9E-19
<i>PA4738</i>	-2.486206	2.9E-28	<i>PA3440</i>	-2.815253	5.6E-21
<i>cupC2</i>	-2.491192	0.00059	<i>rpmF</i>	-2.838743	3.2E-16
<i>PA0093</i>	-2.4916608	1E-22	<i>cspD</i>	-2.8583502	2.4E-20
<i>PA0083</i>	-2.4943747	6.2E-28	<i>PA2780</i>	-2.877613	3.2E-15
<i>prfF2</i>	-2.4987378	2.8E-11	<i>PA5087</i>	-2.8994094	4.5E-18
<i>PA4739</i>	-2.5016847	1.6E-30	<i>nrdB</i>	-2.9037448	4.4E-48
<i>minE</i>	-2.5048247	2.2E-15	<i>PA3291</i>	-2.9196498	4E-08
<i>PA1942</i>	-2.506656	6.7E-14	<i>PA0092</i>	-2.9354793	6.9E-21
<i>fdx1</i>	-2.5079828	2.7E-16	<i>PA4570</i>	-3.0269497	2.1E-27
<i>rplP</i>	-2.5102768	6.8E-33	<i>leuD</i>	-3.0793338	9.9E-88
<i>PA0603</i>	-2.5179913	3.2E-19	<i>ilvC</i>	-3.1184275	2.4E-47
<i>PA1516</i>	-2.5182978	2.9E-16	<i>PA3441</i>	-3.1277554	1.4E-46
<i>PA3057</i>	-2.5200667	5.5E-08	<i>ilvH</i>	-3.3381102	3.7E-34
<i>PA3572</i>	-2.556733	7.5E-19	<i>imm2</i>	-3.4143638	5.1E-14
<i>PA4463</i>	-2.5759976	1.7E-23	<i>hcp1</i>	-3.4213646	1.3E-40
<i>PA2146</i>	-2.5794817	3.7E-07	<i>PA3866</i>	-3.4678735	6.2E-33
<i>PA2775</i>	-2.5874571	1.6E-15	<i>PA3496</i>	-3.4849062	8E-22
<i>PA0434</i>	-2.598101	1.1E-40	<i>fpvA</i>	-3.5728243	2.4E-49
<i>PA0563</i>	-2.5982767	1.5E-27	<i>PA3784</i>	-3.7377774	1.3E-84
<i>PA4611</i>	-2.6053038	7.4E-19	<i>PA3785</i>	-4.5717396	1.1E-88
<i>PA1152</i>	-2.6112485	7.3E-07	<i>crcZ</i>	-6.3565923	8.7E-28
<i>pvdO</i>	-2.6313839	2.9E-44	<i>PA4324</i>	-185.17234	2E-239
<i>tolR</i>	-2.6345551	9.7E-21			
<i>pvdE</i>	-2.6380278	1.8E-36			
<i>PA1830</i>	-2.6432154	5.1E-20			
<i>gdhA</i>	-2.6621203	2.2E-27			
<i>PA2412</i>	-2.6642045	1.8E-33			
<i>PA4336</i>	-2.6669479	4.1E-33			
<i>PA0284</i>	-2.6677477	9.5E-23			
<i>rpmE</i>	-2.6689086	4.2E-18			
<i>PA0084</i>	-2.7124117	4.3E-32			
<i>tufA</i>	-2.7234316	4.9E-40			
<i>pyoS5</i>	-2.730119	1.9E-17			
<i>PA3292</i>	-2.7389657	1.3E-09			

Table S31. Genes up-regulated in $\Delta PA2989$ strain

Gene	Fold change	padj	Gene	Fold change	padj
PA3661	26.6061627	3.539E-113	PA3485	4.875476847	2.00602E-15
PA0083	12.22745732	1.5542E-233	PA0047	4.856304774	3.53194E-17
PA0082	12.18272158	8.4836E-150	PA1844	4.84510841	3.14226E-24
PA0840	10.27226507	1.19067E-75	PA2376	4.796857103	2.09133E-18
PA0087	10.03175501	4.41859E-45	PA0092	4.735054764	5.31301E-52
PA2703	9.970381596	1.15832E-80	PA2539	4.639507923	3.10311E-68
PA3484	9.840952659	7.7194E-136	PA3731	4.61113731	4.01286E-63
PA2550	8.528562545	2.36467E-49	PA0095	4.604348516	2.6898E-140
PA0080	8.505837537	1.25302E-77	PA0563	4.482192927	1.51987E-69
PA1845	8.407734023	1.49634E-39	PA3729	4.320905457	2.2625E-158
PA0084	7.932295125	3.3691E-143	PA2464	4.308118483	6.79813E-42
PA1639	7.713929449	5.6221E-100	PA0089	4.133259629	1.38985E-29
PA0086	7.558823385	3.3611E-135	PA2775	4.120584756	7.92201E-43
PA0079	7.374794525	7.05502E-57	PA2702	4.114637207	2.28915E-36
PA0045	7.308961345	2.6024E-102	<i>phr</i>	3.943848839	8.96436E-67
<i>rsmZ</i>	7.229721229	4.01043E-22	PA4321	3.903072753	2.12755E-39
PA0040	7.199048223	1.38384E-92	<i>nalC</i>	3.8587354	2.82299E-73
PA0097	7.024082065	8.3756E-64	PA3722	3.823217868	1.32195E-42
PA4318	6.996768543	1.25146E-51	PA1395	3.813513557	5.0666E-26
<i>dctA</i>	6.677239106	1.46439E-18	PA3427	3.688094578	1.01892E-35
PA0046	6.524784457	4.01497E-60	PA3483	3.637226549	5.7331E-27
PA4317	6.039040244	3.4583E-123	PA0855	3.635834075	3.10642E-15
PA0096	6.030427228	6.02968E-20	PA4491	3.624109925	8.13018E-88
PA0094	5.995545432	1.2477E-110	PA4658	3.594363887	2.47423E-46
PA2540	5.826963369	1.0482E-125	PA0528	3.587059495	2.25444E-11
PA0093	5.783119529	7.8763E-98	PA2560	3.577196485	1.0272E-53
PA3730	5.733485107	6.4181E-107	PA4490	3.473046677	3.81299E-74
PA0078	5.733391545	1.68161E-27	PA1341	3.460761784	9.9509E-45
PA0098	5.686538311	6.8094E-31	PA2685	3.451251648	1.44317E-66
PA1396	5.478060102	7.24671E-57	PA0099	3.449506541	6.31937E-26
PA0070	5.305803797	4.9085E-149	<i>ilvI</i>	3.44793975	8.12614E-59
PA3732	5.299920919	1.4195E-120	<i>ribC</i>	3.418784382	1.15718E-18
PA2774	5.138451487	1.50663E-38	<i>arnE</i>	3.414373676	3.44242E-08
PA2792	5.111898231	8.16692E-41	PA3277	3.414230444	3.99423E-32
PA4659	5.08823936	1.103E-36	<i>icmF1</i>	3.363616236	7.65798E-28
PA0088	5.054366828	5.69091E-65	PA4320	3.326693353	3.97564E-48
<i>leuA</i>	5.051874614	3.3487E-112	PA5114	3.293487178	3.33417E-26
PA4319	4.980909118	8.06677E-63	<i>ggt</i>	3.265930183	1.37816E-45

Table S31 (continued). Genes up-regulated in Δ PA2989 strain

Gene	Fold change	padj	Gene	Fold change	padj
PA0839	3.232407008	6.56308E-18	PA5273	2.44692091	7.95541E-23
PA2793	3.220593935	3.56222E-34	PA4582	2.429263215	3.26586E-32
PA2581	3.194122095	1.27524E-56	PA2365	2.429108536	1.03306E-57
PA3294	3.151064644	4.21801E-19	PA0050	2.420378739	1.83841E-13
PA3749	3.136521722	6.52299E-31	PA0069	2.418388684	1.75705E-08
PA2684	3.132982998	7.94776E-30	<i>nhaP</i>	2.409014443	1.09932E-34
PA4774	3.127962893	1.96649E-53	<i>lexA</i>	2.40860029	3.37683E-52
<i>arnC</i>	3.10097106	1.24952E-19	PA0173	2.406723035	3.13543E-07
PA4773	3.078260788	8.41471E-24	PA2912	2.3892853	1.53397E-11
PA1791	3.07733677	2.60481E-26	PA0603	2.385609358	1.52323E-20
<i>ribB</i>	3.057439629	1.73354E-16	PA3454	2.384811129	5.94313E-14
PA3850	3.047374313	1.69635E-53	PA2929	2.383436998	4.54133E-22
PA1394	3.033539943	0.00013157	PA3314	2.379331118	1.47285E-05
PA2773	3.004023113	4.03574E-24	<i>glyA2</i>	2.376185725	7.3421E-22
PA5135	2.966059923	1.08803E-05	<i>cysW</i>	2.352271109	2.20604E-38
<i>hcp1</i>	2.869510936	3.71952E-31	PA2537	2.347573936	1.65947E-41
PA0989	2.853366004	8.40721E-15	PA2538	2.346263131	1.90887E-07
PA5441	2.852757641	5.55495E-90	PA4883	2.331489734	0.00469486
PA1451	2.846833559	4.22903E-20	PA1363	2.330261552	1.56963E-37
<i>lpdV</i>	2.829700797	1.91266E-08	PA2913	2.327895241	1.25348E-13
PA1068	2.779833325	1.22081E-35	<i>braZ</i>	2.320685377	2.1865E-14
<i>arnT</i>	2.747372866	2.02986E-12	PA2362	2.319850408	1.75332E-13
<i>arnB</i>	2.743580487	1.76824E-44	PA1274	2.311448598	0.004424615
PA0578	2.742982715	2.41335E-09	PA3420	2.311329108	6.57281E-05
<i>ilvH</i>	2.738221673	5.0666E-26	<i>ispA</i>	2.310789495	2.06664E-06
PA1339	2.684678446	6.41927E-34	PA4584	2.306088283	9.14041E-23
<i>arnD</i>	2.658036429	1.12987E-06	PA5289	2.301219238	2.54473E-18
PA5020	2.645813756	1.86711E-24	<i>infA</i>	2.296116482	2.10586E-09
PA2541	2.64155559	1.07667E-69	PA4064	2.287943483	1.18159E-24
PA4492	2.639281253	2.22105E-63	PA2503	2.276879832	1.45716E-11
PA5136	2.62863326	8.21136E-34	PA2366	2.274895787	1.18263E-43
PA1365	2.628358505	1.62761E-92	PA3559	2.274213702	2.50842E-39
PA4583	2.57634009	6.67054E-36	PA3008	2.254562782	8.06276E-27
PA4489	2.550877937	4.65937E-87	<i>arnA</i>	2.243860273	4.99935E-38
PA3021	2.536297857	3.47311E-43	PA2203	2.242320623	3.32939E-11
PA1340	2.526677302	7.50676E-26	<i>cysA</i>	2.236774373	2.62365E-24
PA5234	2.513777856	3.54792E-27	PA2449	2.22964377	1.2042E-16
PA0562	2.507962877	2.74634E-05	PA2393	2.226965614	6.54481E-19
PA3965	2.482553609	4.92631E-19	<i>arnF</i>	2.224840072	4.7358E-08

Table S31 (continued). Genes up-regulated in Δ PA2989 strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>PA3317</i>	2.221231005	1.40319E-14	<i>PA4746</i>	2.099074522	2.50161E-12
<i>rrmA</i>	2.213410806	1.10954E-09	<i>pvdP</i>	2.093695847	1.42391E-19
<i>leuD</i>	2.213399133	5.82881E-49	<i>PA0418</i>	2.08776537	1.66822E-12
<i>PA1559</i>	2.207868909	2.61349E-29	<i>pvdN</i>	2.081008448	1.44609E-20
<i>pmrA</i>	2.203194545	2.48063E-13	<i>PA3728</i>	2.078505802	1.20232E-28
<i>PA2368</i>	2.202502827	2.23E-09	<i>PA3293</i>	2.077295422	0.000202928
<i>nrdA</i>	2.200843717	3.97368E-60	<i>pdxB</i>	2.072857334	0.00151188
<i>PA2655</i>	2.199857662	0.009484178	<i>PA4707</i>	2.053257448	2.38435E-10
<i>PA1657</i>	2.198175273	3.19553E-10	<i>PA1828</i>	2.042013581	3.81984E-22
<i>PA4580</i>	2.194136153	4.43488E-15	<i>metE</i>	2.041587939	3.36975E-08
<i>PA0236</i>	2.189348666	0.000748309	<i>PA3685</i>	2.039190449	0.000188263
<i>PA0126</i>	2.189218156	3.23446E-19	<i>grx</i>	2.038492808	1.91016E-08
<i>PA2452</i>	2.189105607	4.33995E-12	<i>PA1336</i>	2.029080632	5.68917E-09
<i>PA3534</i>	2.188899955	3.74483E-07	<i>PA1560</i>	2.022101126	1.35992E-24
<i>PA2380</i>	2.187787385	2.48489E-20	<i>PA5120</i>	2.019683445	0.004292641
<i>PA5113</i>	2.185720358	3.85012E-11	<i>bioD</i>	2.018520888	0.021171583
<i>PA3443</i>	2.182678896	9.36279E-11	<i>PA3453</i>	2.014350307	1.09498E-06
<i>fha1</i>	2.18230264	9.73265E-28	<i>PA1450</i>	2.008358772	1.68114E-11
<i>PA4654</i>	2.179763481	9.12066E-05	<i>thiD</i>	2.003018191	9.39174E-07
<i>glmS</i>	2.178654573	5.49814E-36	<i>PA1342</i>	2.001609235	1.20096E-17
<i>PA0730</i>	2.169269617	4.95699E-24			
<i>PA0171</i>	2.167535841	0.000146567			
<i>PA2789</i>	2.165850659	3.08317E-06			
<i>PA3119</i>	2.165137521	3.37446E-22			
<i>PA3444</i>	2.162465232	1.11005E-08			
<i>PA0680</i>	2.161521988	0.016535629			
<i>rpsT</i>	2.157607764	6.05087E-06			
<i>PA4684</i>	2.153776061	4.49792E-10			
<i>PA0736</i>	2.152684948	1.28096E-17			
<i>PA4782</i>	2.147938364	8.30749E-15			
<i>PA0277</i>	2.146231489	5.41653E-09			
<i>PA4322</i>	2.142192226	5.06948E-22			
<i>PA0174</i>	2.117606488	9.91592E-07			
<i>PA4015</i>	2.116152727	4.16633E-17			
<i>PA4668</i>	2.111905908	2.4786E-12			
<i>PA1829</i>	2.111619283	5.30372E-15			
<i>PA0100</i>	2.106659259	9.53108E-24			
<i>PA1645</i>	2.103211672	0.000431787			
<i>PA0659</i>	2.099376063	1.1746E-38			

Table S32. Genes down-regulated in $\Delta PA2989$ strain

Gene	Fold change	padj	Gene	Fold change	padj
<i>fimT</i>	-2.003992879	0.001571626	<i>ldh</i>	-2.157616249	4.08E-29
<i>ambE</i>	-2.007059609	3.52892E-10	<i>PA0788</i>	-2.175798331	2.35E-30
<i>PA4172</i>	-2.00830452	0.000703929	<i>imm2</i>	-2.179151898	3.25E-06
<i>PA0637</i>	-2.009719152	3.4311E-08	<i>aprE</i>	-2.182412635	6.24E-41
<i>PA0714</i>	-2.012813795	6.17552E-07	<i>cynS</i>	-2.19541925	1.55E-09
<i>PA3284</i>	-2.016946106	2.32472E-06	<i>PA2986</i>	-2.20125018	3.72E-15
<i>PA0639</i>	-2.027994243	3.62932E-07	<i>PA3041</i>	-2.213020589	6.04E-52
<i>PA0940</i>	-2.030399403	1.43882E-11	<i>PA0643</i>	-2.225710529	5.73E-17
<i>PA2486</i>	-2.035018595	8.10001E-06	<i>PA2165</i>	-2.228164703	5.21E-24
<i>PA3952</i>	-2.04144324	6.97418E-12	<i>PA4141</i>	-2.230674225	4.93E-15
<i>azu</i>	-2.043066423	4.46573E-15	<i>PA2174</i>	-2.230859573	4.68E-12
<i>PA2187</i>	-2.043410651	0.000103945	<i>PA1114</i>	-2.231187165	8.1E-05
<i>fepB</i>	-2.045028909	1.94425E-12	<i>PA4683</i>	-2.240595157	6E-10
<i>algZ</i>	-2.047393136	3.2256E-21	<i>PA0568</i>	-2.244504581	1.64E-14
<i>PA3460</i>	-2.055357586	8.50061E-19	<i>PA1221</i>	-2.246900471	1.6E-07
<i>ampDh3</i>	-2.056851917	4.04564E-16	<i>algA</i>	-2.249120485	8.26E-17
<i>PA3808</i>	-2.06202279	2.99333E-18	<i>PA2179</i>	-2.252131685	0.000322
<i>PA2177</i>	-2.067028237	1.16469E-26	<i>PA0621</i>	-2.25234772	1.46E-13
<i>PA0631</i>	-2.068489287	0.005365457	<i>PA2883</i>	-2.254950337	8.42E-14
<i>PA0941</i>	-2.069143856	8.28797E-12	<i>PA2175</i>	-2.256028588	1.17E-09
<i>PA2414</i>	-2.0738789	4.16606E-16	<i>PA0952</i>	-2.256929766	6.8E-17
<i>PA0911</i>	-2.076051239	2.06343E-06	<i>PA5473</i>	-2.266112308	1.8E-38
<i>PA0588</i>	-2.078842239	1.15665E-19	<i>PA2108</i>	-2.268308618	5.7E-09
<i>acnA</i>	-2.086586718	1.03001E-67	<i>PA2178</i>	-2.271392997	1.87E-11
<i>PA1244</i>	-2.090874312	6.01354E-28	<i>phzG1</i>	-2.271973104	0.006157
<i>PA2751</i>	-2.09666177	4.8831E-12	<i>pilC</i>	-2.289347879	2.1E-19
<i>PA1545</i>	-2.109197762	2.79137E-12	<i>napE</i>	-2.298683184	2.2E-16
<i>dnaJ</i>	-2.113861105	6.66039E-27	<i>PA3369</i>	-2.298721355	8.34E-15
<i>PA0038</i>	-2.114299189	1.32985E-33	<i>PA0614</i>	-2.299759045	3.71E-05
<i>PA3459</i>	-2.120009104	6.79087E-35	<i>PA2024</i>	-2.309036229	5E-36
<i>PA2135</i>	-2.121069165	0.003397809	<i>PA1245</i>	-2.322311281	4.29E-47
<i>PA3451</i>	-2.128063324	2.5902E-11	<i>PA0632</i>	-2.33706499	2.54E-08
<i>PA1597</i>	-2.136436317	2.24789E-07	<i>PA3890</i>	-2.351741063	6.38E-12
<i>PA0191</i>	-2.136978004	8.06646E-12	<i>PA0616</i>	-2.352835484	8.14E-15
<i>PA2163</i>	-2.138556836	0.000157434	<i>PA2779</i>	-2.366772995	8.96E-17
<i>glpD</i>	-2.151665705	1.57778E-19	<i>PA0620</i>	-2.368236506	3.58E-26
<i>PA3450</i>	-2.152029701	5.70553E-17	<i>PA0449</i>	-2.375082447	2.59E-18
<i>PA2145</i>	-2.153702019	1.29216E-05	<i>PA2156</i>	-2.377152725	0.000123
<i>PA4155</i>	-2.156070748	9.02743E-11	<i>PA2377</i>	-2.381426451	6.05E-42

Table S32 (continued). Genes down-regulated in Δ PA2989 strain

Gene	Fold change	padj	Gene	Fold change	padj
PA0626	-2.392892088	3.98E-18	PA1242	-2.697933995	1.54E-11
PA0039	-2.403748295	5.27E-21	PA3278	-2.698421637	3.37E-14
PA2422	-2.408296038	2.87E-08	PA2146	-2.699419304	1.17E-07
PA3040	-2.409781312	4.76E-23	PA0618	-2.70091701	4.06E-21
PA3461	-2.411677093	8.07E-42	PA0629	-2.731552426	6.34E-20
PA2162	-2.420432797	5.46E-10	PA2149	-2.734943014	0.001133
PA2157	-2.441168679	3.3E-08	PA3888	-2.737424181	1.03E-12
PA0624	-2.447406322	1.36E-14	<i>fabH2</i>	-2.741935097	6.11E-73
<i>groES</i>	-2.450671368	2.98E-42	PA0638	-2.773754418	1.93E-21
PA0572	-2.460259268	3.23E-27	PA0201	-2.77428821	4.65E-30
<i>ibpA</i>	-2.462872369	9.31E-43	PA2158	-2.785365059	1.75E-12
PA0641	-2.464333905	1.5E-36	<i>rfaD</i>	-2.793844484	4.79E-23
<i>cmpX</i>	-2.476017146	1.61E-35	PA3931	-2.821979826	3.6E-16
PA0747	-2.487032617	3.11E-42	PA2143	-2.822539386	9.31E-09
PA5424	-2.502629133	1.71E-13	PA0625	-2.832327855	7.17E-47
PA3529	-2.505256463	1.27E-66	PA0640	-2.864196196	5.42E-29
PA0619	-2.50558184	1.01E-19	PA0636	-2.873772155	1.68E-42
PA1324.1	-2.513834085	2.12E-05	<i>pvdJ</i>	-2.883823116	1.1E-122
PA2152	-2.514115914	7.81E-22	PA0642	-2.910126967	4.24E-12
PA0635	-2.515443478	2.8E-11	PA5061	-2.947102222	2.1E-33
PA0346	-2.520194249	7.48E-15	PA3330	-2.950291373	1.14E-18
PA1870	-2.530000827	5.29E-05	PA0634	-2.962292305	6.06E-30
PA2170	-2.541634019	0.004111	<i>chiC</i>	-2.971355596	3.19E-48
<i>algU</i>	-2.549143419	3.71E-49	<i>aprA</i>	-2.984597478	3.39E-45
PA3446	-2.55193968	6.9E-16	PA3370	-3.022904452	1.54E-18
PA2151	-2.573431533	1.04E-15	PA2747	-3.061433259	2.2E-86
<i>poxB</i>	-2.585979388	9.18E-22	PA2150	-3.065556706	2.96E-12
<i>dnaK</i>	-2.588745332	3.62E-53	PA2485	-3.068399667	5.73E-26
<i>plcR</i>	-2.599573447	4.81E-06	PA2021	-3.085751888	1.44E-09
PA4294	-2.60241287	8.12E-11	PA1592	-3.10511557	1.84E-61
PA0617	-2.614196225	2.16E-14	<i>pslN</i>	-3.105378164	2.08E-08
PA2708	-2.620076874	1.61E-21	PA2155	-3.11740458	2.68E-10
<i>clpB</i>	-2.62362381	3.6E-46	PA2190	-3.121892716	3.23E-19
PA3329	-2.628262977	3.92E-56	<i>phrS</i>	-3.124912856	2.57E-17
PA0613	-2.633183289	5.47E-17	PA0713	-3.126314302	1.2E-20
<i>xdhA</i>	-2.635414293	4.34E-31	PA2754	-3.148892899	3E-36
PA3889	-2.679942553	5.99E-22	<i>mucA</i>	-3.163108818	1.91E-76
<i>groEL</i>	-2.686497939	7.12E-98	PA0622	-3.190871572	4.69E-68
PA0628	-2.687943763	1.37E-40	PA3819	-3.200019279	9.96E-76

Table S32 (continued). Genes down-regulated in Δ PA2989 strain

Gene	Fold change	padj	Gene	Fold change	padj
PA5383	-3.207804896	4.36E-16	PA3331	-4.506896121	1.2E-161
PA2180	-3.210393969	1.49E-11	PA1324	-4.51037499	9.7E-182
PA2988	-3.222344094	4.97E-34	PA1323	-4.519004308	1.33E-79
PA2298	-3.255260165	2.32E-33	<i>osmE</i>	-4.650963177	2.4E-129
PA5212	-3.264869084	7.77E-56	PA1111	-4.739269871	3.52E-14
PA2159	-3.266989443	1.76E-11	PA5303	-4.832993101	3.75E-73
PA2299	-3.27865226	4.08E-16	PA3332	-5.066224587	6E-105
PA2160	-3.293134404	6.83E-62	PA5482	-5.224880308	1.52E-69
PA2169	-3.327397623	1.55E-19	<i>osmC</i>	-5.343806454	3.5E-113
PA0627	-3.346334532	1.43E-17	PA2176	-5.431958269	2.05E-29
PA2184	-3.349822813	4.5E-34	PA3274	-5.791819281	3E-39
PA0567	-3.367206254	1.73E-41	PA5446	-5.974822375	2.12E-80
<i>glgB</i>	-3.391156946	4.55E-35	<i>pfpl</i>	-6.133536578	2.8E-135
PA3891	-3.464448707	6.56E-34	PA3430	-7.774487756	3.58E-91
PA0623	-3.476508059	2.62E-31	PA2166	-7.848237647	7.7E-122
PA4877	-3.494769544	9.42E-17	PA4739	-8.585612637	2.8E-164
PA5481	-3.502335764	1.05E-55	PA4738	-9.362476348	7.2E-165
PA5460	-3.504274277	1.35E-19	<i>dadA</i>	-14.26809745	9.5E-242
PA3336	-3.538469487	3.75E-44	PA2989	-17.1677541	1.8E-114
PA4523	-3.582489386	0	PA3431	-46.13670519	5.8E-201
PA2777	-3.603185337	1.68E-35	PA3432	-58.91469639	6.8E-160
PA2154	-3.605196962	4.99E-20			
PA2171	-3.622709832	4.39E-20			
PA3273	-3.62743651	9.3E-13			
PA3335	-3.634524734	3.05E-54			
PA0633	-3.669973481	4.26E-51			
PA4880	-3.869560228	4.84E-55			
PA2168	-3.893806421	8.18E-14			
PA3334	-3.976808245	7.75E-68			
PA2987	-3.982922071	6.27E-24			
PA2148	-4.010183917	1.02E-07			
PA1404	-4.025468676	2.05E-39			
<i>lptF</i>	-4.138411119	4.41E-97			
PA2167	-4.229063287	1.45E-30			
PA3691	-4.248660722	4.2E-122			
PA4156	-4.330004843	1.37E-91			
PA3231	-4.345760696	2.55E-18			
PA2134	-4.444560426	4.28E-38			
PA2433	-4.487526574	1.1E-69			