CHAPTER TWO

SEQUENCE ANALYSIS OF *Pseudomonas aeruginosa* DSM1707 ATTACHMENT-INDUCED GENETIC ELEMENTS

2.1 INTRODUCTION

Biofilms are complex, highly organized communities of microorganisms attached to a surface and their formation occurs in response to a variety of environmental signals (Costerton *et al.*, 1995; O'Toole and Kolter, 1998a, 1998b; Pratt and Kolter, 1998; Wimpenny and Colasanti, 1997) which lead to the expression of new phenotypes that distinguish the attached cells from their planktonically growing counterparts (O'Toole *et al.*, 2000b; Steyn *et al.*, 2001; Sauer and Camper, 2001). Most notably, biofilm bacteria have been reported to have a greatly enhanced tolerance to stresses and antimicrobial agents (Nickel *et al.*, 1985; Cochran *et al.*, 2000).

In recent years, a number of different approaches have been used to identify genes involved in the formation, development and maintenance of biofilms. Most biofilm-regulated genes have been identified by screening for mutants defective in biofilm formation (O'Toole *et al.*, 1999). By screening a library of *Tn* insertion mutants of *P. aeruginosa* for altered adhesion abilities, O'Toole and Kolter (1998a) identified 13 surface attachment defective (*sad*) mutants. The majority of these mutants were affected in motility, indicating that flagella and type IV pili play an important role in the early events of biofilm development in *P. aeruginosa*. Although it represents a powerful approach, the information is limited to genes obligately related to biofilm formation, while those genes contributing to the biofilm phenotype as well as genes regulated in the attached mode of growth, may be omitted if not obligately required for biofilm development *per se*.

Early studies by Brözel *et al.* (1995), in which global gene expression patterns in attached *P. aeruginosa* cells were investigated, have indicated that the expression levels of 11 proteins were altered during various stages of attachment. Proteomic studies involving the examination of the expression pattern of cellular proteins under different conditions have consequently become a popular method whereby global gene regulation can be studied (Miller and Diaz-Torres, 1999).

Two-dimensional gel electrophoresis of proteins extracted from both attached (biofilm) and planktonic cells, followed by amino acid sequencing of selected spots using either Edman-degradation or Matrix-assisted laser desorption / ionisation-time of flight mass spectrometry (MALDI-TOF MS), has lead to the identification of numerous proteins that are differentially expressed, up-regulated and/or repressed in *Pseudomonas* spp. growing as a biofilm (Steyn *et al.*, 2001; Sauer and Camper, 2001). Using a proteomic approach, in combination with cDNA subtractive hybridization of mRNA, Sauer and Camper (2001) recently identified several proteins involved in amino acid metabolism, membrane proteins involved in transport processes and proteins involved in the production of extracellular polymers that are differentially expressed in *P. putida* biofilm cells.

The use of DNA microarrays to study gene expression in *P. aeruginosa* biofilms has also been reported (Whiteley *et al.*, 2001). Differential expression of 73 genes in *P. aeruginosa* biofilm cells, in comparison to planktonic cells, was noted. These included, amongst other, bacteriophage genes, genes involved in motility, genes involved in translation and metabolism as well as genes encoding for membrane and secretory proteins.

From the above, it is apparent that numerous genes are differentially expressed by *P. aeruginosa* growing as a biofilm. However, in most cases, the mechanism(s) that regulate this differential gene expression remain to be elucidated. In an earlier study undertaken by Weyers (1999), attempts were made to isolate genetic elements which may be important in regulating the expression of genes required for the attachment of *P. aeruginosa* to a surface and the subsequent formation of well-developed biofilms. The approach was based on the use of *lacZ* reporter gene technology to identify genomic sequences responsible for up-regulated expression of the reporter gene upon attachment to a glass wool substratum. These studies resulted in the isolation of 131 elements displaying significant up-regulation of the reporter gene expression. To identify and to characterize the genetic elements, 24 of these putative regulatory elements were selected for use in this study.

The aims of this part of the investigation were therefore (i) to determine the nucleotide sequence of the selected genetic elements, (ii) to identify putative promoter sequences and (iii) to map the

DNA fragments to the *P. aeruginosa* genome in order to determine which genes might be expressed by these regulatory elements.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains and plasmids

Escherichia coli JM109 was routinely propagated in Luria-Bertani broth (LB: 1% tryptone; 0.5% yeast extract; 1% NaCl; pH 7.4) with shaking for 16 to 18 h at 37°C or maintained on LB-agar plates. The recombinant pALacZsd plasmids used in this study have been described previously (Weyers, 1999) and were kindly supplied by J.B. Weyers (Department of Microbiology and Plant Pathology, University of Pretoria).

2.2.2 Preparation of competent *E. coli* JM109 cells

Competent *E. coli* JM109 cells were prepared essentially as described by Sambrook *et al.* (1989). An overnight culture of *E. coli* JM109 cells was prepared by inoculating 3 ml LB-broth with a single colony of the bacterial strain and shaking overnight at 37°C. One ml of this culture was used to inoculate 100 ml of preheated (at 37°C) sterile LB-broth after which the cells were grown at 37°C with agitation to mid-logarithmic phase (OD_{550} of 0.4 to 0.5) and then placed on ice for 20 min. The cells from 30 ml of the culture were pelleted in a Corex tube by centrifugation at 4 000 rpm for 10 min at 4°C in a pre-cooled Sorvall HB 4 rotor. The pellet was suspended in 10 ml filter-sterilized ice-cold 50 mM $CaCl_2$, incubated on ice for 1 h, pelleted as before and gently resuspended in 1 ml of the $CaCl_2$ solution. The cells were incubated on ice for at least 1 h before being used in transformation experiments.

2.2.3 Transformation of competent *E. coli* cells

The prepared competent *E. coli* JM109 cells were transformed with recombinant pALacZsd plasmid DNA using the heat shock-method as described by Sambrook *et al.* (1989). The competent cells (200 µl) were mixed with 200 ng plasmid DNA in a sterile transformation tube

and maintained on ice for 30 min. The cells were then incubated at 42°C for 90 s and immediately chilled on ice for 2 min. After addition of 800 µl prewarmed (37°C) LB-broth, the transformation mixtures were incubated at 37°C for 1 h to allow the cells to recuperate and express the tetracycline resistance gene. As controls, the competent cells were either transformed with 10 ng pUC18 plasmid DNA to determine the transformation efficiency, or directly plated onto LB-agar plates to check for contamination. The cells were plated in aliquots of 100-200 µl onto 1.2% LB-agar plates supplemented with 20 µg/ml tetracycline or 100 µg/ml ampicillin (for the transformation of the pUC plasmid DNA) and incubated overnight at 37°C. Transformants were selected for further characterization and grown overnight at 37°C in 10 ml LB-broth supplemented with 20 µg/ml tetracycline.

2.2.4 Plasmid DNA extraction

The recombinant plasmid DNA was isolated using a modified alkaline lysis method (Birnboim and Doly, 1979). The cells from 3 ml of the overnight cultures were harvested by centrifugation for 1 min at 15 000 rpm. The bacterial pellets were suspended in 400 µl of a solution containing 50 mM glucose; 25 mM Tris-HCl (pH 8.0); 10 mM EDTA and 100 µg RNaseA and incubated at room temperature for 10 min. The resultant spheroplasts were lysed following the addition of 400 µl of a freshly prepared solution containing 1% SDS and 0.2 M NaOH. The mixture was neutralized by the addition of 300 µl of 7.5 M NH_4OAc (pH 7.6), resulting in the precipitation of chromosomal DNA, high molecular weight RNA and proteins. After incubation on ice for 10 min and centrifugation for 10 min at 15 000 rpm, the plasmid DNA was precipitated from the recovered supernatants by the addition of 650 µl isopropanol and incubated at room temperature for 10 min. The precipitated DNA was collected by centrifugation and resuspended in 100 µl 2 M NH_4OAc (pH 7.4). After incubation for 10 min on ice, the excess proteins were removed by centrifugation at 15 000 rpm for 10 min and the plasmid DNA in the supernatant precipitated by the addition of 110 µl isopropanol. The plasmid DNA was pelleted by centrifugation, rinsed with 1 ml of 70% ethanol, dried under vacuum and resuspended in 25 µl UHQ water.

2.2.5 Agarose gel electrophoresis

An aliquot of each plasmid DNA preparation was analyzed by agarose gel electrophoresis. For this purpose, horizontal 1% (w/v) agarose slab gels were cast and electrophoresed at 100 V in 1 × TAE buffer (40 mM Tris-HCl; 20 mM NaOAc; 1 mM EDTA; pH 8.5) using a Biorad Mini Sub™ electrophoresis unit. The agarose gels were stained with ethidium bromide (EtBr) at a final concentration of 1 mg/ml and the DNA visualized by UV-induced fluorescence on a transilluminator.

2.2.6 Nucleic acid sequencing

The recombinant plasmid DNA was of sufficient quality to allow its direct use in nucleic acid sequencing reactions. The recombinant plasmid DNA was sequenced using an ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems, Perkin Elmer). The forward primer JB-1 (5'-GAATTCGAGCTGGGTAC-3') and the reverse primer JB-2 (5'-GTTTTCCCAGTCACGAC-3') were used in the sequencing reactions. The JB-1 and JB-2 primers anneal to the 5' and 3' ends of the multiple cloning site of pALacZsd, respectively. Each reaction mixture contained 200-500 ng of template DNA, 1.6 pmol sequencing primer, 2 µl Terminator Ready Reaction Mix, 3 µl 5 × buffer and UHQ water in a final reaction volume of 20 µl. Cycle sequencing was performed using the following program for 25 cycles: denaturation at 96°C for 30 s; annealing at 50°C for 15 s and extension at 60°C for 4 min. Following brief centrifugation, the extension products were precipitated by the addition of 20 µl 80% ethanol. The tubes were incubated at room temperature for 20 min in the dark, centrifuged at 15 000 rpm for 30 min and the supernatant carefully aspirated. The pellets were rinsed twice with 50 µl 70% ethanol, vacuum-dried for 10 min and then stored at -20°C. Prior to electrophoresis, the purified extension products were resuspended in 3.5 µl Blue dextran/EDTA loading buffer, denatured for 2 min at 90°C and loaded onto a model 377 automated DNA sequencer (Perkin-Elmer). The obtained nucleotide sequences were edited using the ABI PRISM Sequencing Analysis 3.1 and Sequencing Navigator 1.0.1 software programs (PE Applied Biosystems).

2.2.7 Computer analysis

The non-redundant GenBank database as well as the *P. aeruginosa* genome sequence (Stover *et al.*, 2000) was searched for nucleotide sequences similar to the obtained sequences using the BLAST alignment program (Altschul *et al.*, 1997). Multiple alignments were carried out with ClustalW (Thompson *et al.*, 1994). Putative integration host factor (IHF) and sigma 54 (σ^{54}) recognition sites were identified using SEQSCAN, while DNAMAN was used to identify both repeat sequences and σ^{70} promoters within the obtained sequences. The origin and location of the programs used in the above analysis are indicated in Table 2.1.

Table 2.1 Web-based and other computer programs used in this study

Name	Address or distributor
BLAST	http://www.ncbi.nlm.nih.gov/BLAST
Pseudomonas genome project	http://www.pseudomonas.com
ClustalW	http://www.dot.imgen.bcm.tmc.edu:9331/multi-align
SEQSCAN	http://www.bmb.psu.edu/seqscan/seqform1.htm
DNAMAN	Linchon Biosoft

2.3 RESULTS

2.3.1 Identification of attachment-induced regulatory elements

A chromosomal library of *P. aeruginosa* DSM1707 had been previously constructed by digestion of the isolated chromosomal DNA with *Sau* 3AI, and DNA fragments of approximately 200-300 bp in length were cloned into the promoterless broad-host-range reporter vector pALacZsd (Weyers, 1999). Recombinant plasmids displaying promoter activity, as indicated by *lacZ* gene expression, were subsequently analyzed for up-regulation of the reporter gene expression following attachment to a glass wool substratum. From these, a total of 24 clones were selected and further characterized in this study. Recombinant plasmid DNA corresponding to the selected

clones was transformed into competent *E. coli* JM109 cells and following plasmid DNA extraction, the nucleotide sequence of the cloned insert DNA was determined by automated sequencing procedures as described under Materials and Methods (Section 2.2.7). The length of the cloned DNA fragments varied between 61 bp (clone 278) to 445 bp (clone 65). Sequence analysis of these clones furthermore indicated that 8 of the 24 recombinant clones (clones 30, 56, 65, 205, 267, 461, 569 and 899) contained more than one cloned DNA fragment. The formation of such chimeric DNA sequences was not surprising, since *Sau* 3AI digestion of the genomic DNA results in DNA fragments with sticky ends which would be able to concatemerize prior to ligation into the reporter vector molecule. Thus, the nucleotide sequence of a total of 42 distinct putative regulatory elements were determined. These sequences are supplied at the back of this dissertation (Appendix 1).

A BLAST search against all nucleotide sequences contained in the GenBank database revealed significant identity between the obtained sequences and sequences originating from *P. aeruginosa*. A subsequent BLAST search of the *P. aeruginosa* genome (www.pseudomonas.com) indicated that all of the obtained sequences, except one (clone 498), displayed 98-100% sequence identity to sequences located in the *P. aeruginosa* genome. In the case of clone 498, the sequence displayed 51.5% identity to a sequence located in the *P. aeruginosa* genome. As expected, the nucleotide sequences mapped to regions all over the *P. aeruginosa* PAO1 genome. The obtained sequences could, however, be classified into two broad groups. The first group (23/42) consisted of cloned DNA fragments mapping to regions within genes encoding hypothetical proteins (8/23), genes encoding proteins with functions in the transport of molecules, translation, amino acid biosynthesis and metabolism as well as genes encoding for probable two component sensors (10/23) and probable transposases (5/23). The second group (17/42) consisted of cloned DNA fragments mapping to regions upstream of genes encoding hypothetical proteins (5/17), probable transcription regulators (3/17), proteins involved in motility (2/17) or genes encoding different enzymes and membrane proteins (7/17). In addition to the above, one sequence each mapped to a region adjacent to a gene encoding HisM, which is responsible for the transport of small molecules, and to a region downstream of a gene encoding protein PA1136, a probable transcription terminator. The location of these fragments in the *P. aeruginosa* genome is summarized in Table 2.2.

Table 2.2 Location of the sequenced attachment-induced regulatory elements in the *Pseudomonas aeruginosa* genome

Fragment	Position on fragment	Position on *Pseudomonas* genome map	Open reading frame and encoded protein
~30(1)	1-80	3 157 028-3 156 948	Inside PA2800, conserved hypothetical protein
30(2)	77-343	4 006 350-4 006 616	From 159 bp upstream to 107 bp of PA3574, probable transcription regulator
30(3)	339-437	2 365 733-2 365 831	275 bp upstream from PA2151, conserved hypothetical protein
55	1-391	4 627 158-4 627 548	Inside PA4137, probable porin
56(1)	1-65	5 783 132-5 783 068	Inside PA5134, probable carboxyterminal protease (Translation, post translational modification, degradation)
56(2)	62-278	5 896 475-5 896 259	330 bp upstream from PA5238, probable O-antigen acetylase (Cell wall/LPS/capsule)
56(3)	275-377	609 388-609 286	Inside PA0550, hypothetical protein
65(1)	1-226	2 781 439-2 781 214	601 bp upstream from PA2463, hypothetical protein
65(2)	223-445	4 288 237-4 288 459	483 bp upstream from PA3831, leucine aminopeptidase (Translation, post-translational modification, degradation)
73	1-95	6 090 565-6 090 471	Inside PA5412, hypothetical protein
205(1)	1-196	3 890 086-3 890 281	368 bp upstream from PA3477, transcriptional regulator RhlR
205(2)	193-260	4 068 865-4 068 934	Inside PA3635, enolase
267(1)	1-79	3 216 518 -3 216 596	Inside PA2864, conserved hypothetical protein
267(2)	76-238	967 242-967 080	Inside PA0892, AotP, transport of small molecules
278	1-61	2 209 087-2 209 147	Inside PA2018, RND (mex) multidrug efflux transporter

457	1-115	3 661 781-3 661 895	Inside PA3271, probable two component sensor
461(1)	1-120	3 281 560-3 281 441	Next to PA2925, HisM, transport of small molecules
461(2)	116-258	4 219 431-4 219 572	268 bp upstream from PA3764, conserved hypothetical protein
461(3)	255-330	2 746 448-2 746 373	Inside PA2445, amino acid biosynthesis and metabolism
498	285-303	1 165 803-1 165 821	62 bp upstream from PA1080, FlgE, flagellar hook protein
560	1-90	4 538 347-4 538 436	252 bp upstream from PA4058, hypothetical protein
561	1-148	4 589 529-4 589 676	Inside PA4106, conserved hypothetical protein
569(1)	1-100	4 578 510-4 578 609	900 bp upstream from PA4097, probable alcohol dehydrogenase
569(2)	94-166	542 818-542 890	Inside PA0482, carbon compound catabolism
570	1-97	3 868 548-3 868 644	826 bp upstream from PA3461, hypothetical protein
614(a)	1-70	500 902-500 833	Inside PA0445, probable transposase
614(b)	1-70	4 473 840-4 473 909	Inside PA3993, probable transposase
614(c)	1-70	3 843 071-3 843 002	Inside PA3434, probable transposase
614(d)	1-70	5 383 593-5 383 524	Inside PA4797, probable transposase
614(e)	1-70	2 557 166-2 557 235	Inside PA2319, probable transposase
~633(1)	1-157	1 572 726-1 572 882	669 bp upstream from PA1444, FliN, motility and attachment
~699	1-441	5 277 262-5 277 702	Inside PA4699, hypothetical protein
~703	1-429	3 254 326-3 253 897	808 bp upstream from PA2897, probable transcription regulator
707	1-106	524 761-524 866	603 bp upstream from CreD (PA0465), a membrane protein

877	1-193	1 228 563-1 228 371	Downstream from PA1136, probable transcriptional regulator
~889(1)	1-188	1 933 112-1 933 302	Inside PA1785, hypothetical protein
889(2)	183-259	942 036-941 960	Inside PA0861, hypothetical protein
889(3)	254-414	1 227 597-1 227 440	Inside PA1136, probable transcriptional regulator
~921	1-147	4 456 550-4 456 404	829 bp upstream from PA3974, probable two-component sensor
~930(1)	1-64	2 052 889-2 052 825	215 bp upstream from PA1881, probable oxidoreductase
930(2)	60-164	1 925 695-1 925 799	Map start from PA1779, assimilatory nitrate reductase, opposite direction
930(3)	160-261	4 172 627-4 172 526	42 bp upstream from PA3725 (RecJ), single stranded - DNA - specific exonuclease (DNA replication, recombination, modification and repair)

~: Indicates that the sequence is incomplete at the 5' end
( ): The numbers and letters in brackets denote chimeric DNA fragments and fragments located more than once on the genome, respectively

2.3.2 Characterization of the obtained nucleic acid sequences

2.3.2.1 Identification of conserved motifs and repeat sequences

The obtained sequences were characterized in order to identify conserved motifs and inverted repeat sequences. Although multiple sequence alignment analysis may provide a powerful tool for identifying conserved domains in homologous sequences, ClustalW alignment analysis of the sequences obtained in this study did not reveal any significant conserved regions or domains. This may have been due to the low levels of sequence identity observed between the respective DNA fragments as they originated from widely different regions within the *P. aeruginosa* genome.

Previous reports have indicated that inverted repeat sequences within the upstream region of the *algD* gene of *P. aeruginosa* enhance transcription of the gene (Mohr *et al.*, 1990). *AlgD* is involved in alginate synthesis and is also up-regulated by attachment of the cells to a surface (Hoyle *et al.*, 1993; Rice *et al.*, 1995). Consequently, the sequences of the elements were analysed using DNAMAN to identify direct and/or inverted repeat sequences. Although several of the clones contained direct and/or inverted repeat sequences of varying lengths, none of these displayed a resemblance to the types found upstream of the *algD* promoter. The significance, if any, of the identified repeat and inverted repeat sequences will have to be experimentally verified through further mutational analysis.

2.3.2.2 Comparison of the sequences to promoter specificity elements in quorum sensing-controlled genes

P. aeruginosa possesses two quorum sensing systems, the LasR-LasI system and the RhlR-RhlI system, which are globally regulatory elements that control the expression of numerous genes (Fuqua *et al.*, 1996; Whiteley *et al.*, 1999). Davies *et al.* (1998) have reported that the *lasI* system is required for maturation of *P. aeruginosa* biofilms. Recently, Whiteley and Greenberg (2001) described LasR- and RhlR-dependent determinants in promoters of several quorum sensing-controlled (qsc) genes. The promoters possessed inverted repeat sequences, termed lux-

box-like sequences, that display considerable dyad symmetry (16 to 18 bp of 20 bp) and extensive sequence variation. A minimal consensus sequence for the lux-box-like element of *P. aeruginosa* qsc promoters was defined as NNCT-N_{12}-AGNN following sequence analysis of the promoters of several *P. aeruginosa* qsc genes.

Analysis of the sequences obtained from the attachment-induced regulatory elements revealed the presence of the lux-box-like element in the sequences derived from clones 65(2), 498 and 560 (Fig. 2.1). These sequences mapped to the upstream regions of genes encoding a leucine aminopeptidase, FlgE and a hypothetical protein of unknown function, respectively (Table 2.2). Induction of promoters containing an A at position 8 and a complementary T at position 13 of the lux-box-like sequence (Fig. 2.1) have been reported to require a functional LasI and a functional RhlR for full induction, while promoter regions lacking the 8A-13T motif respond to LasR specifically (Whiteley and Greenberg, 2001). Although the sequences of the respective attachment-induced regulatory elements contained the lux-box-like element, they possessed either a 8C-13T or 8G-13A motif, but not the 8A-13T motif. This may therefore suggest that LasR could play a role in the induction of these putative promoter elements.

2.3.2.3 Identification of consensus promoter sequences

To identify consensus promoter sequences, the respective sequences were analyzed using the DNAMAN and SEQSCAN programs. None of the sequences revealed homology with the typical sigma 70 (σ^{70}) family of promoter sequences of *E. coli*. However, a number of the clones were found to contain sequences corresponding to that of σ^{54} promoters (Fig. 2.2). Although the dinucleotides GG-N_{10}-GC are generally highly conserved in σ^{54} promoters (Cases and De Lorenzo, 2001; Barrios *et al.*, 1999), note should be taken that exceptions do occur. The σ^{54} promoters of the *nifH* gene of *Rhizobium leguminosarum* biovar *viciae* and the *glnB* gene of *Rhodospirillum rubrum* both contain GG-N_{10}-GA, while the σ^{54} promoters of the *E. coli glnH* gene and the *P. aeruginosa oprE* gene contain sequences of GG-N_{10}-TC and GG-N_{10}-CC, respectively (Yamano *et al.*, 1998; Nohno and Saito, 1987; Roelvink *et al.*, 1990; Johansson and Nordlund, 1996). By taking these exceptions into account, 6 clones were found to contain putative σ^{54} promoter sequences (Fig. 2.2).

However, when the putative σ^{54} promoters were analyzed for their orientation and proximity relative to *P. aeruginosa* open reading frames (ORFs) (Table 2.2), only two of these were located within 400 bp upstream of *P. aeruginosa* genes. The respective genes encode a probable O-antigen acetylase (clone 56(2)) and FlgE, a flagellar hook protein (clone 498). A third putative σ^{54} promoter was located approximately 600 bp upstream of a gene encoding a hypothetical protein of unknown function (clone 65(1)). The remainder of the putative σ^{54}promoter sequences were located within the open reading frames of a probable transposase, a hypothetical protein and a probable transcriptional regulator. Since *P. aeruginosa* has a G + C content of 66.6% (Stover *et al.*, 2000), the identification of putative σ^{54} consensus sequences in these latter sequences may have been fortuitous.

2.3.2.4 Identification of motifs associated with σ^{54} promoters

Transcription initiation from σ^{54}promoters is a multistep process involving the recognition of the promoter by σ^{54}, binding of the core RNA polymerase to the σ^{54} to form a closed complex, and subsequent activation to an open complex following binding by an enhancer binding protein (EBP) (Dworkin *et al.*, 1997). In most cases, the EBP is brought into contact with the σ^{54}-RNA polymerase complex by DNA looping, an event mediated by the integration host factor (IHF) (Hoover *et al.*, 1990). Since IHF binding sites are therefore often found in the proximity of σ^{54} promoters, the above sequences were examined for putative IHF binding sites by searching for the consensus WATCAA-N4-WTR sequence (Hoover *et al.*, 1990; Taylor *et al.*, 1996). A single IHF binding site was identified only in the upstream sequence of clone 498, 70 nucleotides upstream of the putative σ^{54} promoter.

Gene/fragment	Nucleotide sequence	Dyad symmetry
*gsc*102	A C **C T** G C C C G G A A G G G C **A G** G T	16
*rsa*L	A A **C T** A G C A A A T G A G A T **A G** A T	12
*las*I	A T **C T** A T C T C A T T T G C T **A G** T T	12
*qsc*117	C A **C T** G C C A G A T C T G G C **A G** T T	18
*phz*A	A C **C T** A C C A G A T C T T G T **A G** T T	16
*las*B OP1	A C **C T** G C C A G T T C T G G C **A G** G T	18
*hcn*A	A C **C T** A C C A G A A T T G G C **A G** G G	12
498	G C **C T** G G G C A G C A T C C C **A G** C A	12
65(2)	G A **C T** G C G G C C G G A C A T **A G** A C	12
560	A G **C T** C G G C G G T C T C G A **A G** G T	10

Fig. 2.1 An alignment of the lux-box-like elements (NNCT-N_{12}-AGNN) in quorum sensing-controlled genes (Whiteley and Greenberg, 2001) and the attachment-induced regulatory elements (this study). Highly conserved nucleotides are indicated in bold, while specificity determinants are boxed. Nucleotides of dyad symmetry are indicated to the right.

Fragment	Nucleotide sequence
56(2)	T **G G** C G C C G A A G G T **G C** T
65(1)	T **G G** G A C T G C T C C T **G C** C
498	T **G G** C A A G G T C A C T **T C** A
614	T **G A** C A A G C A C G A T **G C** A
699	C **G G** C G C G C G A C C T **G C** T
889(3)*	T **G G** T A G C G G C C T T **G C** G
889(3)*	T **G G** C G C G G G C A T T **C C** T

Fig. 2.2 Identification of putative σ^{54} promoter sequences (GG/A-N_{10}-G/T/C) in six attachment-induced regulatory elements using the SEQSCAN software program. Highly conserved nucleotides are indicated in bold. An asterisk (*) denotes a single DNA fragment containing two putative σ^{54} promoter sequences.

2.4 DISCUSSION

Although significant advances have been made toward identifying environmental conditions, surface properties and bacterial surface structures that may affect bacterial attachment to surfaces, the molecular mechanisms underlying the switch(es) from planktonic to attached growth and regulation of such mechanisms, have yet to be completely elucidated. To investigate, initial approaches in our laboratory relied on the isolation of regulatory genetic elements from *P. aeruginosa* in response to surface growth. It was envisaged that this may assist in the identification of genetic elements required for bacterial attachment and would yield information on how these elements are regulated.

In this part of the investigation, the nucleotide sequence of 24 clones selected from a group of clones that displayed up-expression of a *lacZ* reporter gene upon attachment of *P. aeruginosa* to a glass wool substratum (Weyers, 1999), was determined. As a consequence of the cloning strategy used, several of the obtained sequences were found to consist of ligated DNA fragments originating from different locations within the *P. aeruginosa* genome. Thus, although 24 clones were initially selected, a total of 42 distinct putative regulatory elements could be identified following sequence analysis (Table 2.2). Since it was not known which of the cloned elements within the chimeric DNA sequences was responsible for the observed up-regulated expression of the *lacZ* reporter gene, all of the obtained sequences were included in the subsequent analysis of the sequence data.

The identity of the attachment-induced genetic elements was determined by BLAST homology searches against the *P. aeruginosa* PAO1 genome sequence. With the exception of the sequence obtained from clone 498, all of the obtained sequences displayed 98-100% sequence identity to sequences within the *P. aeruginosa* genome (Table 2.2). The sequence obtained from clone 498 did however reveal identity to a region upstream of the *P. aeruginosa flgE* gene. The low level of sequence identity may be due to differences in the genomic DNA sequences of *P. aeruginosa* strain DSM1707 used in this study and strain PAO1 of which the genome has been completely sequenced. For example, the genomes of the these two clonally derived *P. aeruginosa* prototype strains have been shown to differ by the presence of a 2.19 Mbp inversion between the *rrnA* and

rrnB operons including *oriC* (Stover *et al.*, 2000). It is therefore possible that other sequence differences may exist between these two *P. aeruginosa* strains. A rather surprising finding was that approximately half of the attachment-induced genetic elements are localized to regions within the ORF of genes. The significance of these sequences being responsible for up-expression of the reporter gene, is therefore unclear. However, a recent report has indicated the presence of a functional promoter within the coding region of another gene. Ochsner and co-workers (1999) reported that the proximal *furP1* promoter of *P. aeruginosa* was located within the *omlA* coding sequence. The encoded proteins had no influence on each other's expression, excluding *trans*-acting cross-regulation between *fur* and *omlA*. The significance of this type of promoter structure awaits future experimentation by for example primer extension analysis.

Recently, the promoters of several *P. aeruginosa* quorum sensing-controlled genes were analyzed and shown to possess conserved quorum control elements that were shown to be important in quorum sensing-controlled transcription (Whiteley and Greenberg, 2001). It has been suggested, that the quorum control elements may serve as activator binding sites and that the activators function by making contact with the RNA polymerase α C-terminal domain and with some other part of the RNA polymerase (Whiteley and Greenberg, 2001). Analysis of the sequences indicated the presence of a quorum control element, NNCT-N_{12}-AGNN, in the sequence obtained from three different clones. These mapped to the upstream regions of *flgE* (clone 498) encoding a flagellar hook protein, and a gene encoding a leucine aminopeptidase (clone 65(2)). Clone 560 mapped to the upstream region of a gene encoding a hypothetical protein. These genes do not correspond to the quorum sensing-controlled genes described by Whiteley *et al.* (1999). Since a saturation mutagenesis was performed and there may be conditions other than those used by Whiteley *et al.* (1999) that can reveal other genes not detected in their screen, it may be possible that other quorum sensing-controlled genes exist. In addition, Whiteley and Greenberg (2001) indicated that some genes with an upstream NNCT-N_{12}-AGNN motif as well as an A at position 8 and a T at position 13, required both the N-(3-oxododecanoyl-$_L$-homserine lactone ($3OC_{12}$-HSL) and N-butyryl-$_L$-homoserine lactone (C_4-HSL) signals for full induction, but that they can respond to either of the signal receptors LasR and RhlR. Thus, the 8A-13T element appears to relax specificity of the quorum sensing-controlled promoters so that it will respond to either LasR or RhlR. By contrast, quorum sensing-controlled genes which lacked the 8A-13T motif responds

only to LasR. Since none of the sequences analyzed contained the 8A-13T motif, it is tempting to speculate that these genes may be regulated by LasR. However, detailed analysis is required before a final conclusion can be drawn.

Analysis of the sequences for consensus promoter sequences indicated that 6 of the 24 clones contained putative σ^{54} promoter sequences (Barrios *et al.*, 1999; Cases and De Lorenzo, 2001). Although the RpoN sigma factor (σ^{54}) was originally discovered in *E. coli* as being essential for nitrogen metabolism (Hirschman *et al.*, 1985), it has since been shown to be involved in various functions in other organisms, amongst other flagellar assembly, pilin synthesis (Arora *et al.*, 1997; Kinsella *et al.*, 1997), alginate production (Zielinski *et al.*, 1992) and rhamnolipid production (Pearson *et al.*, 1997). Bacterial genes transcribed by σ^{54}-RNA polymerases are invariably regulated at a distance by a class of transcription regulators known as prokaryotic enhancer binding proteins (EBP). These regulatory proteins, which catalyze the isomerization of the enzyme-σ^{54} promoter complex from the closed to the open state, bind to upstream activator sequences (UAS) located 80 to 200 bp upstream of the promoter and act through DNA looping to contact the holoenzyme (Dworkin *et al.*, 1997). DNA looping may, however, also be mediated by intrinsic DNA bends. The putative σ^{54} promoter sequence identified in clone 498 was the only promoter sequence identified containing an IHF consensus sequence. Notably, this sequence mapped to the upstream region of the *P. aeruginosa flgE* gene, which encodes a flagellar hook protein, one of the structural components of flagella. In several bacteria, flagella and pili have been described as major structures required for either stable cell-to-surface attachment and/or cell-to-cell interactions required in the formation of microcolonies (DeFlaun *et al.*, 1994; De Weger *et al.*, 1987; O'Toole and Kolter, 1998a; 1998b; Watnick *et al.*, 1999; Pratt and Kolter, 1998; Genevaux *et al.*, 1996). Recently, Sauer and Camper (2001) proposed a role for flagella in the detachment of cells from well-developed biofilms as flagella could be detected in 3- to 7-day old biofilms. Of the other clones containing consensus σ^{54} promoter sequences, clone 56(2) mapped to the upstream region of a gene encoding a probable O-antigen acetylase. This enzyme may O-acetylate polysaccharides, e.g. alginate (Gacesa *et al.*, 1998). Adhesiveness of *Pseudomonas* species has been reported to be related to the presence and composition of polysaccharides (Williams and Fletcher, 1996) and substantially reduced attachment to biotic and abiotic surfaces has been observed in O-polysaccharide-deficient *Pseudomonas* spp. (DeFlaun

et al., 1999; Dekkers *et al.*, 1998). None of the other sequences containing consensus σ^{54} promoter sequences mapped to regions upstream of *P. aeruginosa* genes and as *P. aeruginosa* has a high G + C content, their identification may have been fortuitous.

In conclusion, sequence analysis of *P. aeruginosa* genomic DNA fragments responsible for up-regulating the expression of a reporter molecule upon attachment to a glass wool substratum indicated no homology with σ^{70}-dependent promoter sequences. By contrast, several of the sequences displayed significant homology with σ^{54}-dependent promoter sequences. This may indicate that σ^{54} plays a role in the formation of biofilms in *P. aeruginosa* by regulating transcription of specific genes whose products are required for biofilm formation. The sequence of one of the DNA fragments (clone 498) mapped upstream of a flagellar gene and contained in addition to a consensus σ^{54} promoter sequence, an IHF binding site. This clone was selected and the transcriptional activity from this regulatory element was characterized at a single cell level in both biofilm and planktonic cells using reporter gene technology, as described in the following chapter.