

# Genetic diversity of flagellins of *Pseudomonas aeruginosa*

Claudia Spangenberg\*, Thomas Heuer, Christiane Bürger, Burkhard Tümmler

Klinische Forschergruppe, Zentrum Biochemie und Zentrum Kinderheilkunde, Medizinische Hochschule Hannover and Max-Planck-Institut für experimentelle Endokrinologie, D-30625 Hannover, Germany

Received 8 September 1996

**Abstract** Physical genome analysis of the virulence-associated *fliC* locus in 20 *Pseudomonas aeruginosa* strains by mapping and sequencing revealed groups of heterologous a-type (1164 bp; 1185 bp) and highly conserved b-type (1467 bp) flagellin genes. Whereas only two synonymous nucleotide substitutions were detected in eight b-type *fliC* sequences, the 12 a-type sequences exhibited 57 nucleotide substitutions, of which 39 occurred within a variable central region. Although a-type and b-type flagellins differ by 35% in their primary structure, they share strong homology in their predicted features, implying that the polymorphic proteins fold into similar structures during polymerization of the flagella.

**Key words:** Flagellum; Flagellin gene; Genome diversity; Gene mapping; Yeast artificial chromosome; *Pseudomonas aeruginosa*

## 1. Introduction

The opportunistic pathogen *Pseudomonas aeruginosa* is a Gram-negative bacterium found ubiquitously in nature at low frequency [1]. *P. aeruginosa* possesses a single polar flagellum which confers motility and chemotaxis, facilitates adherence to cells and inanimate surfaces and contributes to colonization and invasion during the early phase of infection in predisposed hosts [2–4]. *P. aeruginosa* strains express either a-type or b-type flagella [5]. This classification is based on the antigenicity and apparent molecular weight of the flagellin subunits encoded by *fliC*. b-type flagellins were found to comprise a homogeneous group of proteins, whereas the heterogeneous a-types were divided into several subgroups.

We report on the genetic basis of the diversity of flagellins. Twenty *P. aeruginosa* strains from various clinical and environmental habitats were compared in their *fliC* nucleotide and deduced amino acid sequences. The previously undescribed b-type *fliC* gene of reference strain PAO was cloned and localized on the chromosome by high-resolution restriction mapping. Although the a-type and b-type genes differ by 35% in primary structure, sequence alignments demonstrate a high homology between all *P. aeruginosa* flagellins, which substantiates the intention to use flagella antigens as antipseudomonal vaccines [6,7].

## 2. Materials and methods

### 2.1. Strains and plasmids

In order to analyze the diversity of flagellin genes, 19 *P. aeruginosa* strains from different habitats and strain PAK were selected as re-

ported previously [8]. A library of *SpeI*-restricted genomic DNA of *P. aeruginosa* PAO was maintained as pYAC4-derived [9] artificial chromosomes in *Saccharomyces cerevisiae* strain AB1380 [9] (T. Heuer, to be published). The subcloned terminal *XhoI* fragments of the YACs were stored as circular plasmids in the host *Escherichia coli* DH5 $\alpha$  [10].

### 2.2. Preparation of chromosomal DNA

Genomic DNA was prepared using a rapid method for Gram-negative bacteria [11].

### 2.3. PCR and sequencing

PCR was performed from purified DNA as described previously [8]. Consensus oligonucleotide primers *fliA* in the 5'-region of *fliC* (5'-GCCTGCAGATCGCCAACC) and *fliB* in the 3'-region (5'-GGCAGCTGGTTGGCCTG) enabled amplification of all flagellin genes of the analyzed strains. The complete *fliC* genes were sequenced in both directions by primer walking with 10 additional primers.

After purification by ultrafiltration with Ultrafree-MC Filter Units (Millipore), the PCR products were sequenced by the dideoxy chain termination method [12] using the Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems Inc.) and analyzed on a 373A automatic sequencer (ABI).

### 2.4. Mapping and cloning

*P. aeruginosa* bacteria grown to late exponential phase were encapsulated into agarose blocks, lysed with detergents and proteinase K, and the intact chromosomes were cleaved with *SpeI* as described previously [13]. The *SpeI* digests were separated by pulsed-field gel electrophoresis (PFGE) in a BioRad DR<sup>II</sup> cell (U=200 V, 37 h, 10°C, two linear ramps of 5–25 s and 5–60 s in 1 s increments) and transferred onto nylon membranes by capillary blotting [13]. Truncated a-type or b-type *fliC* sequences were amplified from *P. aeruginosa* DNA by PCR, labelled with digoxigenin-dUTP [13] and hybridized with the pulsed-field blot. Hybridized fragments were detected by chemiluminescence using an alkaline phosphatase-conjugated anti-digoxigenin antibody and subsequently CDP-Star (Tropix) as substrates [13]. In the case of *P. aeruginosa* PAO, the precise map position of *fliC* was determined by repetitive hybridizations of PFGE-separated *SpeI* complete/n partial double digestions (n: *EcoRI*, *BglII*, *PstI*, *XhoI*, *NdeI*) with cloned *SpeI* fragment ends and the *fliC* PCR probe. A *SpeI* fragment end subclone was identified to carry the complete *fliC* gene and sequenced.

### 2.5. Nucleotide sequence analysis

DNA sequence data were analyzed using the Genetics Computer Group (GCG) Sequence Analysis Software Package (University of Wisconsin, Madison, WI) [14]. Homology searches were conducted against the GenBank, EMBL and Swiss-Prot databases with the programs 'BlastN', 'BlastP' and 'FastA'. Pairwise sequence comparisons and multiple alignments were generated using the 'Gap' and the 'ClustAl' programs, respectively. Structural features of the primary sequences as hydrophilicity [15], surface probability [16], chain flexibility [17] and antigenicity index [18] were calculated with 'Peptide-structure'. The *fliC* sequences of strain DSM 1128 and strain ATCC 15691 have been assigned GenBank accession numbers L81146 and L81147.

## 3. Results and discussion

### 3.1. Cloning and sequence of the b-type *fliC* gene of *P. aeruginosa* PAO

The flagellin gene of the genetic reference strain PAO was

\*Corresponding author. Klinische Forschergruppe, OE 4350, Medizinische Hochschule Hannover, D-30623 Hannover, Germany.  
Fax: (49) (511) 5325966.

TTCGAAGCATGTAAACCACTGAAGAGGA -151  
 AGAGAAAAAGAAATGTTGATTTTCTCTTAAAGCTCCGCGGAAACGC -101  
CGATAACACCATGAACGCGAATTTCTGGGGACCTGAGCAAGCAGGCCG -51  
 AGAGATCGCAAGCTCAGGTAAACGAAATAGTCTCTTGGAGGAAATCACC -1  
 ATGGCCCTTACAGTCAACACGAACATTGCTTCCCTGAACACTCAGCGCAA 50  
 M A L T V N T N I A S L N T Q R N 17  
 CCTGAATGCTTCTTCCAACGACCTCAACACCTCGTTGCAGCGTCTGACCA 100  
 L N A S S S N D L N T S L Q R L T T 34  
 CCGGCTACCGCATCAACAGTGCACGAGCATGCTGCCGCGCTGCAGATC 150  
 G Y R I N S A K D D A A G L Q I 50  
 TCCAACGCGCTGTCCAACGAGTCAAGCGTCTGAACGTTGCCACCGCAA 200  
 S N R L S N Q I S G L N V A T R N 67  
 CGCCAACGCGCATCTCCCTGGCGCAGACCGCTGAAGGTGCCCTGCAGC 250  
 A N D G I S L A Q T A E G A L Q Q 84  
 AGTCCACCAATATCTGCAGCGTATCCGCGACCTGGCCCTGCAATCCGCC 300  
 S T N I L Q R I R D L A L Q S A 100  
 AACGGCTCCAACAGCGACCGACCTGCGCCCTGCAGAAAGAGTCCG 350  
 N G S N S D A D A G S L Q K E V A 117  
 TGCGCAACAGCGCGAAGTACCGGTATCTCCGATACCAACCTTCGGTG 400  
 A Q Q A E L T R I S L D T T T F G G 134  
 GCCCAAGCTGTCTGACGGCTCTCCGACCAACGAGTTCAGGTCCGGT 450  
 R K L L D G S F G T T S F Q V G 150  
 TCCAACGCTACGAGACCATTGACATCAGCCTGCAGAAATGCCTCTGCCAG 500  
 S N A Y E T I D I S L Q A N S A S 167  
 CGCCATCGGTTCTTACAGGTCCGCGACCAACGCGCGGTACCGTCGCCA 550  
 A I G S Y Q V G S N G A G T V A S 184  
 GCGTAGCGGCGACCGCGACCGCTTCGGGCATCGCCTCGGGCACCGTCAAC 600  
 V A G T A T A T V G I N A S G T V N 200  
 CTGGTCGGTGGCGGTGAGTGAAGAACATCGCCATCGCCGCAGCGATAG 650  
 L V G G G Q V K N I A I A A G D S 217  
 CGCCAGGCCATCGCCGAGAAGATGGACGGTGCATCCCGAACCTGTCCG 700  
 A K A I A E A G A G A I P N L S A 234  
 CTCGTGCCGTACCGTGTTCACCGCTGATGTGACGCGGTGACCGGTGGT 750  
 R A R T V F T A D V S G V T G G 250  
 TCGCTGAACCTTCGACGTAAACCGTTGGCAGCAACACCGTGAGCCTGGCAGG 800  
 S L A T D V T V G S I N A S L A G 267  
 CGTGACCTCACTCAGGATCTGGCCGACCAACTGAACTCCAACCTCGTGA 850  
 V T S T Q D L A D Q L N S N S S K 284  
 AGCTGGGCTACTGCCAGCATCAACGACGAGGTGTACTGACCATCAC 900  
 L G I T A S I N D K G V L T I T 300  
 TCCGCTACCGCGAGAACGTCAGTTCGGTGCAGACCGGTACCGCTAC 950  
 S A T G E N V K F G A Q T G T A T 317  
 TGCCGGTCAAGTTCGAGTGAAGTTCAGGTTCCGACGCGAAGTTCGAG 1000  
 A G Q V A V K V Q G S D G K F E A 334  
 CGGCCCGCAAGACGTGGTGTGTCGGTACTGCGCTACCAACCACTC 1050  
 A A K N V V A A G T A A T T T I 350  
 GTGACCGGCTACGTGCAACTGAACTCGCCGACCGCTACTCGGTCAAGCGG 1100  
 V T G Y V Q L N S P T A Y S V S G 367  
 TACCGGCACCCAGGCTTCGACAGTCTTCGGCAACGCCAGCGCCGCGCAGA 1150  
 T G T Q A S T V F G N A S A A Q K 384  
 AGAGCAGCGTTGCCAGCGTGCACATCTCCACTGCCGACGGCGCCGAGAAC 1200  
 S S V A S V D I S T A D G A Q N 400  
 GCCATCGCGGTAGTCGATAACGCCCTGGCTGCGATCGACGCCACCGTGC 1250  
 A I A V V D N A I D A I D A Q R A 417  
 TGACCTCGGTGCTGTTTCAAGAACCTATCGACAACCTGA 1300  
 D L G A V Q N R F K N T I D N L T 434  
 CCAACATTCGGAACCGTACCAACCGTCTAGCCGATCAAGACACC 1350  
 N I S E N A T N R S R I K D T 450  
 GACTTCGCTGCCGAAACCGCGCGCTGTGCAAGAACAGGTGCTGCAACA 1400  
 D F A A E T A A L S K N Q V L Q Q 467  
 GGCCGGTACCGCATCTGGCCAGGCCAACAGCTGCCGAGGCGGTCC 1450  
 A G T A I L A Q A N Q L P Q A V L 484  
 TGAGCCTGCTCGCTAAGCCCGGAACGGTCACTCAGCGGTACTGGGAG 1500  
 S L L R \*\*\* 488  
 GAAGGGGTGAGCCCTCTCCCTTTCCCTTTGCGAGGCGATGAGAAATGGA 1550  
 CGTCGGAATATCACTTCCCTTTCTACGTTCAGACCGCGCAGGCGCCCGG 1600  
 AGGCCAGCGCGATATCTTTCGCGCGCGCAGCGAGCGGATGGCAGCGGC 1650  
 AAACCGTTGCCGGAAGTGACGGCTTCCCGGAGGCCAGCGAATCTCGCGA 1700  
 TGACCTGGGCTCGCGCTCAGCGACATCCAGTCTTTCGTGCAGAGCGTCA 1750  
 AGCGCAACTTGAACCTCAGCATCGACGA 1778

Fig. 1. Nucleotide and deduced amino acid sequence of b-type *fliC* and flanking regions of *P. aeruginosa* PAO. The putative recognition sequence for ribosome binding (GGAG<sub>-10</sub>, numbering refers to the first nucleotide of the start codon ATG) and the consensus motif of RpoF-dependent promoters (TAAA-(N)<sub>15</sub>-GCCGATAA<sub>-95</sub>) [35] upstream from *fliC* are boxed. The two dimorphic nucleotide sites in the coding region are in bold face. The palindromic sequence forming a potential stem of a termination loop is underlined. Another inverted repeat with unknown function in the upstream region is overlined. The arrows indicate the complementary sequences of the consensus primers fla1 and fla2.

a 1300 bp large product from genomic PAO DNA which exhibited more than 70% nucleotide sequence identity with *fliC* sequence of strain PAK. Using the PCR product as a probe, *fliC* was positioned on the physical map of the PAO chromosome [23] by Southern hybridization. The full-length gene was retrieved from the 4.8 kb terminal *SpeI/XhoI* fragment of PAO *SpeI* fragment SpU subcloned in *E. coli* from a yeast artificial chromosome carrying SpU as insert (see below).

Fig. 1 shows the 1467 bp long *fliC* sequence that encodes a 488 amino acid large protein. Codon usage and the GC content of 63% are typical for the GC-rich *P. aeruginosa* genome [24]. Like other previously described flagellins [19–22], *P. aeruginosa* b-type flagellin mainly consists of aliphatic uncharged amino acids (83% of total), a nearly equal number of acidic (7%) and basic (7%) residues and a few aromatic amino acids (3%). It does not contain any cysteine, histidine and only three prolines, i.e. those amino acids are counter-selected that could exert kinetic and/or structural constraints on the folding and multimerization of flagellin subunits during the assembly of the flagella [4].

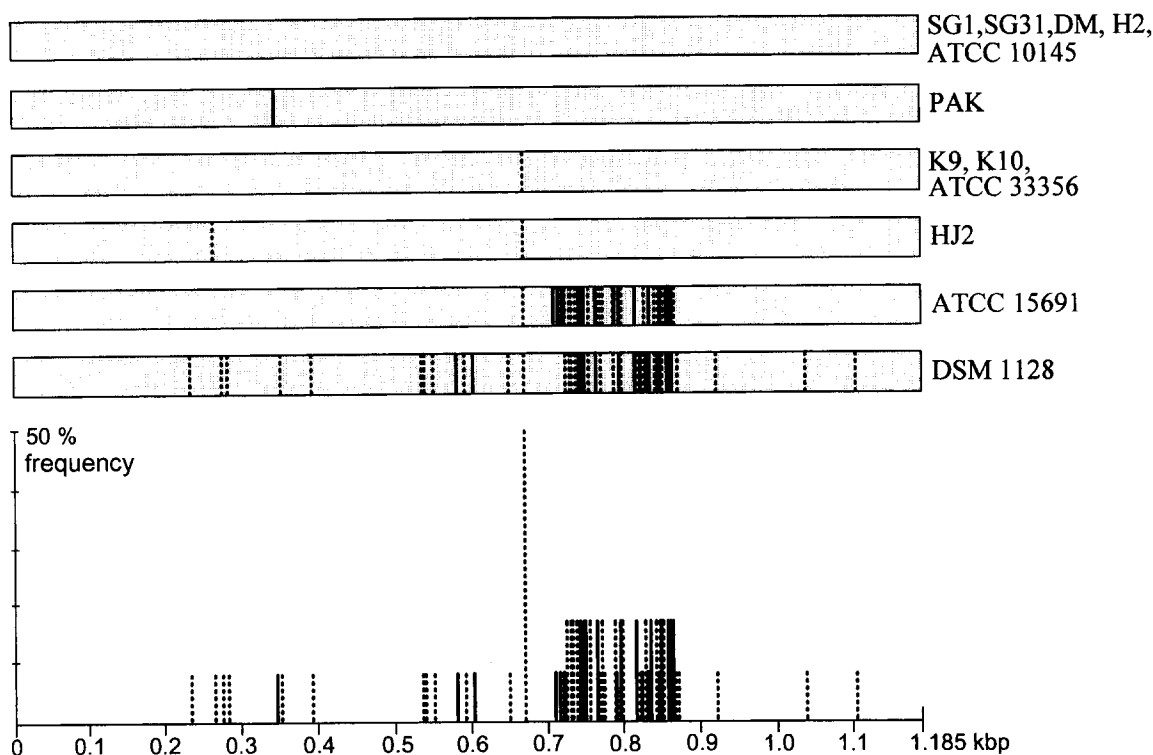
### 3.2. Sequence diversity of a-type and b-type flagellins

Twenty *P. aeruginosa* strains were analyzed in *fliC*, of which 12 strains harbor the a-type and eight strains the b-type *fliC* gene. The amino acid sequence was identical in the eight strains that encode a b-type flagellin. At the level of the nucleotide sequence (Fig. 2), single synonymous C-to-T substitutions of the PAO sequence were detected at position 1386 in four strains and position 642 in one strain, whereby a 7-mer palindromic sequence is destroyed. The b-type *fliC* gene is more conserved than housekeeping genes of *P. aeruginosa* (own unpublished data), *E. coli* and *Salmonella* [25–27] that have been subjected to comparative sequence analysis.

The sequences of the 12 analyzed a-type *fliC* genes were less conserved in accordance with the known classification of a-type flagellins into different antigen subtypes (Fig. 2). The open reading frame varies between 1164 bp and 1185 bp in length. Thus, a-type flagellins are 94 or 101 amino acids smaller in size than the 488 amino acid large b-type flagellin. a- and b-type flagellins share nearly identical N- and C-terminal sequences, whereas the central region is variable in size and primary structure (Fig. 3). This central part is also the major region of sequence variation amongst a-type *fliC* genes (Figs. 2 and 3). Within a 141 bp cassette, the type strain ATCC 15691 and the ear isolate DSM 1128 show 28% nucleotide and 40% amino acid diversity compared to the other ten analyzed a-flagellins. Outside this variable central region only 18 nucleotide substitutions were in total identified in the 12 strains, of which three give rise to conservative amino acid

identified by the combination of PCR, mapping and cloning. Consensus oligonucleotide primers fla1 and fla2 were designed from the sequence alignment of the closely related *fliC* genes of *P. aeruginosa* strain PAK (a-type *fliC*) [19], *P. putida* [20], *Serratia marcescens* [21] and *Bacillus subtilis* [22]. PCR yielded

## a) a-type



## b) b-type

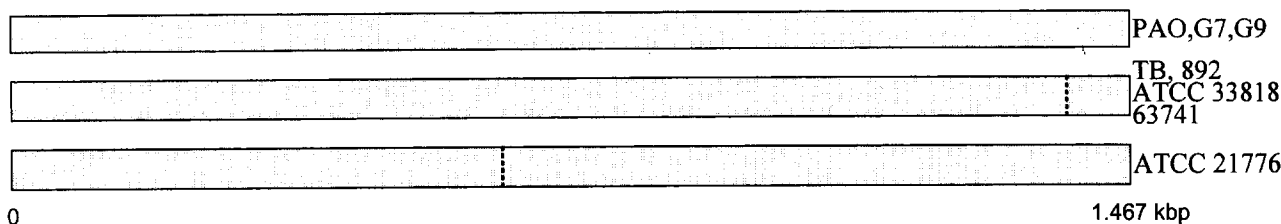


Fig. 2. Sequence polymorphisms of a-type (a) and b-type (b) flagellin genes. Synonymous nucleotide substitutions are indicated by dotted bars, non-synonymous substitutions by solid bars.

exchanges (A116V, I195L, A202T). The 12 a-type *fliC* sequences were compiled into six groups by their diagnostic nucleotide pattern at dimorphic sites (Fig. 2). The most frequent dimorphism is an A-to-T substitution at position 672. The thermodynamic stability of its four adjacent base pairs probably facilitated the uncommon transversion of the center nucleotide in the palindrome CCGCWGCGG.

All a-type sequences share 96–98% identity in both their nucleotide and amino acid sequences. Pairwise sequence comparisons of a- and b-type flagellins of *P. aeruginosa* gave a 74% identity of the nucleotide sequences and a 63–65% identity of the amino acid sequences. A search in several protein databases revealed that amongst all known flagellins the a-type and b-type flagellins of *P. aeruginosa* are most homologous within themselves. The next closest relatives at the gene and protein levels are flagellins from other bacteria, i.e. ordered by decreasing relatedness the flagellins of *Pseudomonas putida* > *Legionella micdadei* ≥ *Salmonella* ≅ *Escherichia coli*. The rank order roughly reflects the phylogenetic distance of *P. aeruginosa* to other prokaryotes [28].

### 3.3. *fliC* location on the chromosome

The b-type *fliC* gene was assigned to the 114 kb large fragment SpU of *P. aeruginosa* PAO [23] and the a-type *fliC* gene was mapped onto the 123 kb large fragment SpU of strain C (SG 1) [29] by Southern hybridization of pulsed-field blots (Fig. 4). Both types of flagellin genes are located at analogous positions in the auxotroph-poor region of the chromosome [29]. The two *fliC* probes detected a broad range of *SpeI* fragment sizes among unrelated strains (a-type, 86–485 kb; b-type, 83–298 kb; Fig. 4). The same random distribution of *SpeI* fragment length has been observed in a larger panel of *P. aeruginosa* strains with PAO fragment SpU as a probe [10], suggesting that genome organization around the *fli* operon [30] is not conserved in *P. aeruginosa* although the *fliC* b gene itself does not display extensive sequence polymorphism (see above).

The chromosomal localization of *fliC* in strain PAO was refined by Smith-Birnsteil [31] partial/*SpeI* complete restriction mapping with enzymes *EcoRI*, *BglII*, *PstI*, *XhoI* and *NdeI* (Fig. 5). Successive hybridizations of the restricted ge-



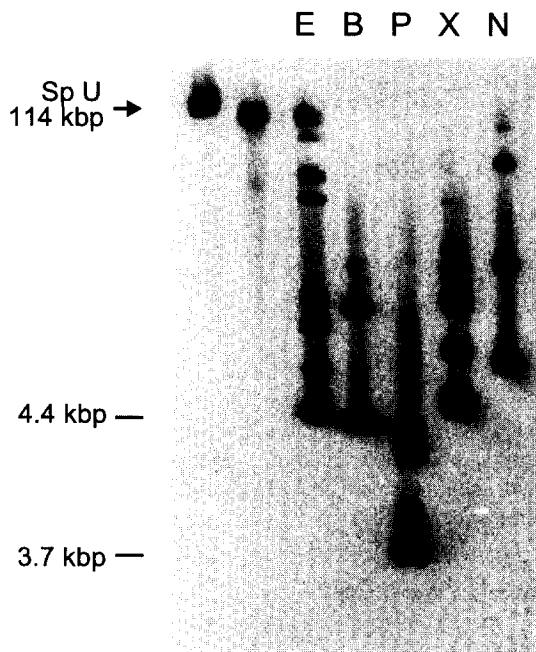


Fig. 5. Smith-Birnstiel high resolution restriction mapping of the *fliC* locus of *P. aeruginosa* PAO. The autoradiogram shows the *SpeI* complete/partial fragment patterns (*EcoRI* (E), *BglII* (B), *PstI* (P), *XhoI* (X) or *NdeI* (N)) hybridized with *fliC* PCR probe.

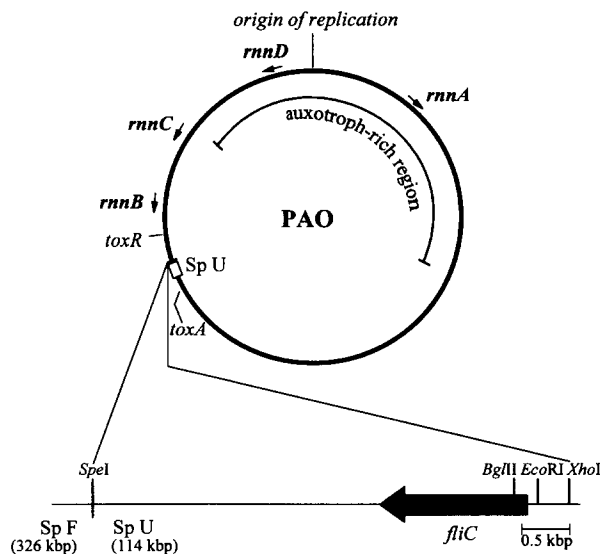


Fig. 6. Physical map position of type-b flagellin gene of *P. aeruginosa* PAO. The 4.8 kb terminal *SpeI/XhoI* fragment of SpU is enlarged showing the chromosomal localization and orientation of *fliC*.

constraints for the efficient self-assembly of many flagellin subunits to a flagellum are probably so tight that the polymorphic proteins fold into a similar three-dimensional structure.

**Acknowledgements:** We would like to cordially thank Peter Pietrzyk for technical support and Karen Schmidt and Lutz Wiehlmann for

helpful discussions. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

## References

- [1] Botzenhart, K. and Döring, G. (1993) in: *Pseudomonas aeruginosa* as an Opportunistic Pathogen (Laupa, M., Bendinelli, M. and Friedman, H., Eds.), pp. 1–18, Plenum Press, New York.
- [2] Montie, T.C., Drake, D., Sellin, H., Slater, O. and Edmonds, S. (1987) *Antibiot. Chemother.* 39, 233–248.
- [3] Drake, D. and Montie, T.C. (1988) *J. Gen. Microbiol.* 134, 43–52.
- [4] Macnab, R.M. (1992) *Annu. Rev. Genet.* 26, 131–158.
- [5] Allison, J.S., Dawson, M., Drake, D. and Montie, T.C. (1985) *Infect. Immun.* 49, 770–774.
- [6] Crowe, B.A., Enzensberger, O., Schober-Bendixen, S., Mitterer, A., Mundt, W., Livey, I., Pabst, H., Kaeser, R., Eibl, M., Eibl, J. and Dorner, F. (1991) *Antibiot. Chemother.* 44, 143–156.
- [7] Döring, G., Pfeiffer, C., Weber, U., Mohr-Pennert, A. and Dorner, F. (1995) *Am. J. Respir. Crit. Care Med.* 151, 983–985.
- [8] Spangenberg, C., Fislage, R., Sierralta, W., Tümmler, B. and Römling, U. (1995) *FEMS Microbiol. Lett.* 125, 265–274.
- [9] Burke, D.T., Carle, G.F. and Olson, M.V. (1988) *Science* 236, 806–812.
- [10] Heuer, T. (1994) PhD Thesis, Hannover.
- [11] Chen, W. and Kuo, T. (1993) *Nucleic Acids Res.* 21, 2260.
- [12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [13] Römling, U., Heuer, T. and Tümmler, B. (1994) in: *Advances in Electrophoresis* (Chrambach, A., Dunn, M.J. and Radola, B.J., Eds.), Vol. 7, pp. 355–395, VCH, Weinheim.
- [14] Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* 12, 387–395.
- [15] Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 110.
- [16] Emini, E.A., Hughes, J.V., Perlow, D.S. and Boger, J. (1985) *J. Virol.* 55, 836–839.
- [17] Karplus, P.A. and Schulz, G.E. (1985) *Naturwissenschaften* 72, 212–213.
- [18] Jameson, B.A. and Wolf, H. (1988) *Comput. Appl. Biol. Sci.* 4, 181–186.
- [19] Totten, P.A. and Lory, S. (1990) *J. Bacteriol.* 172, 7188–7199.
- [20] Winstanley, C., Morgan, J.A.W., Pickup, R.W. and Saunders, J.R. (1994) *Microbiology* 140, 2019–2031.
- [21] Harshey, R.M., Estepa, G. and Yanagi, H. (1989) *Gene* 79, 1–8.
- [22] LaVallie, E.R. and Stahl, M.L. (1989) *J. Bacteriol.* 171, 3085–3094.
- [23] Römling, U. and Tümmler, B. (1991) *Nucleic Acids Res.* 19, 3199–3206.
- [24] West, S.E.H. and Iglewski, B.H. (1988) *Nucleic Acids Res.* 16, 9323–9335.
- [25] Nelson, K., Whittam, T.S. and Selander, R.K. (1991) *Proc. Natl. Acad. Sci. USA* 88, 6667–6671.
- [26] Smith, N.H. and Selander, R.K. (1991) *Proc. Natl. Acad. Sci. USA* 88, 956–960.
- [27] Boyd, E.F., Nelson, K., Wang, F.-S., Whittam, T.S. and Selander, R.K. (1994) *Proc. Natl. Acad. Sci. USA* 91, 1280–1284.
- [28] Olsen, G.J., Woese, C.R. and Overbeek, R. (1994) *J. Bacteriol.* 176, 1–6.
- [29] Schmidt, K.D., Tümmler, B. and Römling, U. (1996) *J. Bacteriol.* 178, 85–93.
- [30] Tsuda, M. and Iino, T. (1983) *J. Bacteriol.* 153, 1008–1017.
- [31] Smith, H.O. and Birnstiel, M.L. (1976) *Nucleic Acids Res.* 3, 2387–2398.
- [32] Strom, M.S. and Lory, S. (1993) *Annu. Rev. Microbiol.* 47, 565–596.
- [33] Römling, U., Greipel, J. and Tümmler, B. (1995) *Mol. Microbiol.* 17, 323–332.
- [34] Paranchych, W. (1990) in: *Molecular Basis of Bacterial Pathogenesis* (Iglewski, B.H. and Clark, V.L., Eds.), The Bacteria Vol. XI, pp. 61–78, Academic Press, San Diego, FL.
- [35] Starnbach, M.N. and Lory, S. (1992) *Mol. Microbiol.* 6, 459–469.