A genomic island in *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation

Shiwani K. Arora*, Mahalaxmi Bangera†, Stephen Lory†, and Reuben Ramphal*‡

*Department of Medicine/Infectious Diseases, University of Florida, Gainesville, FL 32610; and †Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115

Edited by John J. Mekalanos, Harvard Medical School, Boston, MA, and approved June 12, 2001 (received for review May 18, 2001)

Protein glycosylation has been long recognized as an important posttranslational modification process in eukaryotic cells. Glycoproteins, predominantly secreted or surface localized, have also been identified in bacteria. We have identified a cluster of 14 genes, encoding the determinants of the flagellin glycosylation machinery in *Pseudomonas aeruginosa* PAK, which we called the flagellin glycosylation island. Flagellin glycosylation can be detected only in bacteria expressing the a-type flagellin sequence variants, and the survey of 30 *P. aeruginosa* isolates revealed coinheritance of the a-type flagellin genes with at least one of the flagellin glycosylation island genes. Expression of the b-type flagellin in PAK, an a-type strain carrying the glycosylation island, did not lead to glycosylation of the b-type flagellin of PAO1, suggesting that flagellins expressed by b-type bacteria not only lack the glycosylation island, they cannot serve as substrates for glycosylation. Providing the entire glycosylation island of PAK, including its a-type flagellin in a flagellin mutant of a b-type strain, results in glycosylation of the heterologous flagellin. These results suggest that some or all of the 14 genes on the glycosylation island are the genes that are missing from strain PAO1 to allow glycosylation of an appropriate flagellin. Inactivation of either one of the two flanking genes present on this island abolished flagellin glycosylation. Based on the limited homologies of these gene products with enzymes involved in glycosylation, we propose that the island encodes similar proteins involved in synthesis, activation, or polymerization of sugars that are necessary for flagellin glycosylation.

Protein glycosylation in prokaryotic organisms is a relatively uncommon posttranslation modification process, first described in Archea, where the S layer proteins were shown to contain covalently attached sugars (1). Several bacterial proteins are now known to undergo glycosylation, including potential virulence factors of pathogenic bacteria, such as the pilins of *Neisseria gonorrhoeae* (2), *Neisseria meningitidis* (3), and one strain of *Pseudomonas aeruginosa* (4), an adhesin of *Chlamydia* (5), a surface-exposed immunodominant protein of two *Ehrlichia* species (6), and the TiBA adhesin of ETEC (7). Moreover, the subunits of the flagellar filaments in a variety of bacterial species are similarly modified, including *P. aeruginosa* (8), *Campylobacter coli* and *Campylobacter jejuni* (9), *Treponema pallidum* (10), *Borrelia burgdorferi* (11), *Helicobacter felis* (12), *Caulobacter crescentus* (13), and *Agrobacterium tumefaciens* (14). However, glycosylation is not restricted to proteins found within bacterial surface appendages. A broad range of glycoproteins has been identified in *C. jejuni* (15) and in *Mycobacterium tuberculosis* (16), indicating that glycosylation in bacteria may be as common as it is in higher cells.

The genetic basis of glycosylation of selected bacterial proteins has been only recently approached. Glycosylation of meningococcal pilin requires the product of the *pglA* gene; although related to a number of glycosyltransferases, it is not involved in the biosynthesis of lipopolysaccharide (LPS) (3). Similarly, the proteins of the general glycosylation system in *C. jejuni* (15) encoded by the *pglA-G* genes are involved in the modification of a number of proteins, including flagellin, and do not participate in the synthesis of LPS, despite their sequence similarity with the enzymes of LPS and capsule biosynthesis. Therefore, glycosylation of bacterial proteins represents a specific modification pathway and is not simply carried out as a secondary activity of enzymes involved in the biosynthesis of polysaccharides or glycolipids. Moreover, bacterial glycosylation appears to resemble the analogous process in eukaryotic cells with asparagine, serine, and threonine residues serving as the acceptors for the first sugar on the protein (17, 18).

P. aeruginosa can be classified into two groups (a- and b-types) (19) on the basis of the expression of the flagellin protein of a specific primary amino acid sequence (20). The observed difference in the mobility of the flagellin protein from the strain PAK (a-type) from its predicted size by sequence (21) was assumed to be because of posttranslational modification. This conclusion was subsequently verified by Brimer and Montie (8), who showed that flagellins of the a-type but not b-type are glycosylated. The a-type flagellins showed variability in size as well as considerable sequence diversity in the central region (8). The deduced amino acid sequences predict sizes around 40–41 kDa, but these flagellins, when analyzed by SDS/PAGE, migrate with an apparent size of 45–52 kDa (8). Carbohydrates have been detected on three of four strains examined; however, the fourth strain also showed discrepant migration on SDS/PAGE. In contrast, the amino acid sequences of b-type flagellins show remarkable conservation. Eight b-type flagellin genes showed identical sizes and deduced amino acid sequences, and five strains showed single synonymous nucleotide substitutions (20). The migration of b-type flagellins on SDS gels is consistent with their deduced molecular masses, and glycosyl groups have not been detected on these proteins (8). Therefore, the absence of carbohydrates on the b-type flagellin may be because of the absence of the genetic determinants encoding the glycosylation machinery in these strains and/or because the b-type flagellins lack the amino acid sequences necessary for glycosylation.

The goal of this work was to elucidate the genetic basis for selective glycosylation by strains expressing a-type flagellin. Here, we describe the presence of a genomic island in *P. aeruginosa* strains expressing a-type flagellin, which was likely acquired by these strains through horizontal gene exchange. Sequence analysis of this island identified genes for several putative enzymes involved in glycosylation, and the role of two of these in flagellin modification was demonstrated after mutagenesis of the respective genes. Furthermore, by constructing strains that express the heterologous flagellins, and introduction of the entire glycosylation island into a b-type strain, we demonstrate that the failure of b-type strains to glycosylate their

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: LPS, lipopolysaccharide.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF332547).

‡To whom reprint requests should be addressed. E-mail: RAMPHR@medmac.ufl.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

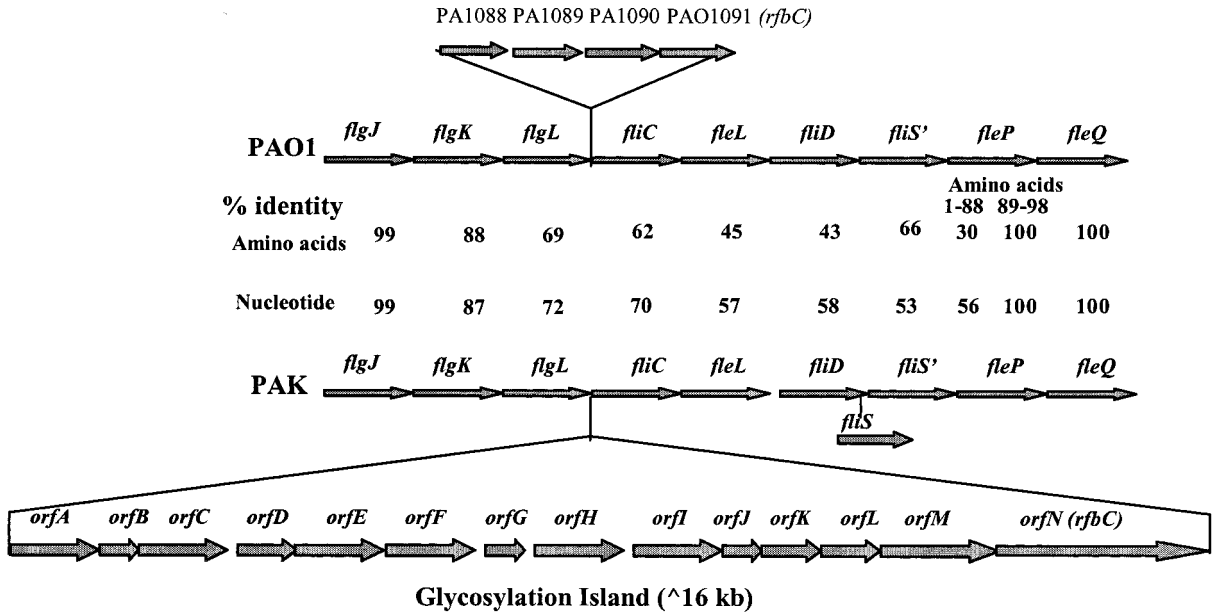

Fig. 1. Genetic organization of the polymorphic chromosomal region of *P. aeruginosa* strains PAO1 and PAK involved in the biosynthesis and assembly of flagellar components. Indicated are segments of DNA that are inserted at specific locations within individual genomes, including a large segment of DNA (≈16 kb) containing a cluster of 14 ORFs that are associated with glycosylation of flagellin, present in PAK, and an apparent duplication of the *fliS* gene. The genome of PAO1 has a ≈6-kb DNA segment containing three genes of unknown function followed by PA1091, a homologue of the *rfbC* gene that is 35% identical to the PAK *orfN*.

cognate flagellins is because of two factors: inability of the b-type flagellin to serve as a substrate and the absence of the glycosylation island-encoded genes in the b-type strains.

Materials and Methods

Bacterial Strains, Cosmids, Plasmids, and Media. *P. aeruginosa* strains PAK and PAO1 were used in this study. *Escherichia coli* strain DH5α was used as the host strain for clonings. The PAK*fliC* mutant (*pilA*::Tc^r, *fliC*::Gm^r) was described previously (22). The antibiotics used were as follows: for *E. coli*, ampicillin (200 μg/ml), tetracycline (25 μg/ml), and gentamicin (10 μg/ml); for *P. aeruginosa*, carbenicillin (300 μg/ml), tetracycline (100 μg/ml), and gentamicin (100 μg/ml).

PCR Amplification and Primers. PCR was used to obtain specific amplification products. Restriction enzyme recognition sites were added at the ends of primers to facilitate subsequent cloning of the PCR products if desired. Primer RER83 (with a *Hin*dIII site) was used as the 5′ primer, and primer RER84 (with a *Sst*I site) was used as the 3′ primer to amplify a 2.3-kb segment of DNA carrying an incomplete *orfN* gene from *P. aeruginosa* PAK. Primers RER69 (with a *Hin*dIII site) and RER70 (with a *Sst*I site) were used for the amplification of the *fliC* gene from PAO1 DNA, and the product was used to engineer a mutation in this gene and to make the complementing plasmid. The distribution of the *orfA* gene in *P. aeruginosa* isolates was analyzed by PCR, by using primers for an internal sequence of *orfA*. The generation of a 555-bp product was assessed after agarose gel electrophoresis of an aliquot from the PCR.

DNA Sequencing. Nucleotide sequence of a ≈23-kb DNA segment of *P. aeruginosa* PAK genome was determined by cosmid walking. The templates used for the sequencing reactions consisted of cosmids pRR194 (23) and cosmid D1, isolated from the same library by colony hybridization by using an *orfA*-specific DNA probe. Double-stranded sequences were aligned and assembled by using programs in the SEQUENCER software package (Gene Codes, Ann Arbor, MI).

Plasmid and Mutant Strain Constructions. A 2.3-kb insert, containing the PAK homologue of a gene annotated as *rfbC* (24) having N-terminal 526 aa missing, was cloned into the *Hin*dIII and *Sst*I sites of pBluescript KS(+) to give pBS8384. This plasmid was linearized at the unique *Eco*RV site present in the PAK homologue of the *rfbC* gene (*orfN*), and a gentamicin resistance gene cassette excised from pUC7G was inserted at that site, leading to the construction of pBS8384G. This construct was used to generate a chromosomal mutation in the PAK *orfN* (*rfbC*) gene by marker exchange. A 3.2-kb PCR product with *Eco*RI/*Xho*I ends containing the *orfA* gene was first cloned into pBluescript, followed by deletion of 981 bp, which deleted 327 aa from *orfA*. A chromosomal deletion mutant of *orfA* was subsequently obtained by allelic exchange in PAK by using the counterselectable *sacB* marker. Construction of the plasmid pBS6970 with the PAO1 *fliC* gene was described previously (25). A 1.4-kb *Eco*RI/*Bam*HI fragment containing the tetracycline resistance cassette was excised from pBSTet and was cloned into the *Eco*RI/*Bgl*II sites in the PAO1 *fliC* gene on pBS6970. The resulting plasmid pBS6970Tet was used to construct a tetracycline-resistant PAO1*fliC* mutant. The construct pBSTet was obtained by cloning a filled-in blunt-ended 1.4-kb *Eco*RI/*Ava*I fragment containing the tetracycline resistance gene into the *Sma*I site of pBluescript. Plasmid pDN186970 used for complementation of PAO1*fliC* was constructed by cloning the 2.5-kb *Hin*dIII/*Sst*I fragment from pBS6970 into the broad host range cloning vector pDN18 (26). Plasmid pPT244 (21) containing the PAK*fliC* gene was used to complement the PAK*fliC* mutant. The plasmid used for the complementation of the PAK *rfbC* mutant, pMMB194H, was obtained by cloning an 8-kb *Hin*dIII fragment from the cosmid pRR194 (23) into the broad host range cloning vector pMMB67HE (27). This 8-kb *Hin*dIII insert contained the complete *rfbC* homolog and the *fliC* gene of strain PAK and ORFs L and M (See Fig. 1). Plasmid pSP329Gm10EH, used for the

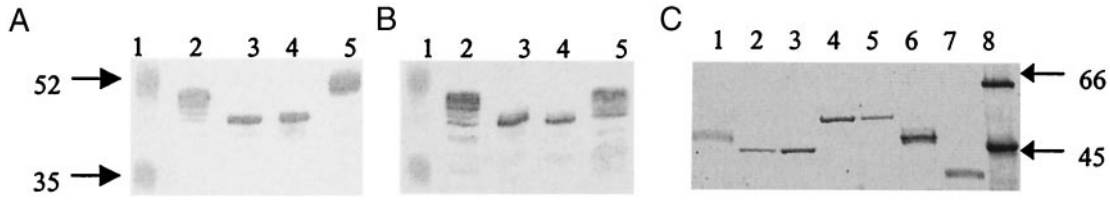

Fig. 2. (*A*) Glycosylation status of the flagellin protein from PAK*orfN* mutant and the complemented strains assessed by Western blots. The wild-type PAK flagellin migrates as a diffuse band indicating glycosylation of flagellin (lane 2). The flagellin from the PAK*orfN* mutant migrates much faster, suggesting loss of glycosylation (lane 3). Introduction of the complete *orfN* gene on a plasmid (pMMB194H) complements the glycosylation defect of the PAK*orfN* mutant (lane 5), whereas the vector control (pMMB66) remains nonglycosylated (lane 4). Size markers (lane 1) are in kDa. (*B*) Glycosylation status of the flagellins from PAK*orfA* mutant and the complemented strains assessed by Western blots. The wild-type PAK flagellin migrates as a diffuse band indicating glycosylation of flagellin (lane 2). The flagellin from the PAK*orfA* mutant migrates faster than the wild-type flagellin, suggesting loss of glycosylation (lane 3). Introduction of the complete *orfA* gene on a plasmid (pSP329Gm10EH) restores flagellin glycosylation in the PAK*orfA* mutant (lane 5) but not in mutants carrying the vector control (pSP329Gm) (lane 4). Lane 1 contains the size markers. (*C*) Effect of chemical deglycosylation on flagellin mobility during SDS/PAGE. Lane 1, PAK flagellin; lane 2, deglycosylated PAK flagellin; lane 3, PAK*orfA* mutant flagellin; lane 4, PAO1 flagellin; lane 5, deglycosylated PAO1 flagellin; lane 6, horseradish peroxidase; lane 7, deglycosylated horseradish peroxidase; and lane 8, molecular mass markers in kDa.

complementation of the PAK*orfA* mutant, was constructed by cloning a 10-kb *Eco*RI/*Hin*dIII fragment (excised from the cosmid D1) into the *Eco*RI/*Hin*dIII sites of the broad host range cloning vector pSP329Gm. This 10-kb insert carried *orfA,* a homolog of the *E. coli* vioA (28), as well as *orfs* B–J (see Fig. 1). Further addition of the 8-kb *Hin*dIII fragment into the unique *Hin*dIII site of pSP329Gm10EH gave rise to pSP329Gm10EH8, which harbors the whole cluster of putative glycosylation genes that was entirely absent from the b-type strain PAO1.

SDS/PAGE and Western Blots. SDS/PAGE and Western blots were used to assess the glycosylation status of *P. aeruginosa* strains. Poly(vinylidene difluoride) membranes with the transferred proteins were incubated with 1:5,000 dilutions of either anti-FlaA or anti-FlaB polyclonal antibodies followed by detection of alkaline phosphatase activity by color reaction.

Deglycosylation Assay. Carbohydrates were removed from the glycoproteins by treatment with anhydrous trifluoromethanesulfonic acid by using the Glyco Free Kit (Glyco, Williamsport, PA). Deglycosylation was monitored by increased mobility of the treated proteins during SDS/PAGE compared with the untreated samples.

LPS Analysis. LPS was prepared as described by Hitchcock and Brown (29), from the wild-type strain PAK, and from strains carrying mutations in the first and last genes of the putative glycosylation island. These preparations were subjected to SDS/PAGE and silver stained for analysis of the banding patterns.

Results

Identification, Sequence Analysis, and Genetic Organization of a Putative Glycosylation Region Linked to the Flagellar Genes of Strain *P. aeruginosa* PAK. Using the differential hybridization analysis (30), we identified several clones from a variety of clinical strains that defined novel sequences not present in strain PAO1. End-sequencing of several of these clones revealed that they were located adjacent to the flagellin locus of *P. aeruginosa*. The presence of unique sequences was verified by analysis of the region upstream from the *fliC* gene of *P. aeruginosa* PAK, suggesting that a new determinant was present in this chromosomal location of strain PAK, and perhaps in other *P. aeruginosa* isolates. The complete nucleotide sequence of the region between *flgJ* and *fliC* genes of strain PAK was determined. This 20,978-bp sequence was combined with several previously sequenced regions that contained genes required for the biogenesis of the *P. aeruginosa* flagellum. The location of various genes within this *ca.* 25-kbp segment of the PAK chromosome is shown in Fig. 1, along with the corresponding region of the PAO1 genome (www.pseudomonas.com). Most notably, the PAK genome contains a genomic island consisting of a cluster of 13 genes (designated *orfA–orfM*), which are absent from PAO1, that are located between flagellar structural genes *flgL* and *fliC*. The PAK genome also includes a second copy of the *fliS* gene, as reported previously (31). In contrast, the PAO1 genome has three genes of unknown function (PA1088, PA1089, and PA1090) directly upstream of the *fliC* gene that are absent in PAK. The 13 genes, in addition to the adjacent gene named *orfN*, the homologue of the *rfbC* gene, contain some of the determinants of glycosylation of the flagellin protein (see below), and hereafter it will be referred to as the glycosylation island. These genes were not present at another location on the chromosome. Therefore, in PAK or its ancestral strain, this island was likely placed in the flagellar biosynthetic gene cluster as a consequence of its acquisition from another bacterial genome, resulting in the replacement of a small preexisting DNA segment.

Nucleotide Sequence Accession Number. The nucleotide sequence of the glycosylation island is part of that in GenBank accession no. AF332547.

Characterization of Chromosomal Mutants of *P. aeruginosa orfA* and *orfN*, Present at the Boundaries of the Flagellar Intergenic Region. Because the genomic island in strain PAK is linked to the flagellin gene, which is glycosylated in this a-type strain, we examined the role of the gene products encoded within this island in flagellin glycosylation. Mutations in two genes located at the 5′ end and at the 3′ end of the island, respectively, were engineered by making a deletion in the *orfA* gene and by insertion of the gentamicin resistance cassette in *orfN*. Each of these mutations abolished the ability of *P. aeruginosa* PAK to glycosylate flagellin as determined by its faster migration, compared with the wild-type (Fig. 2 *A* and *B*). The mobility of the flagellin in the mutants was consistent with the predicted size of the protein from the sequence of the *fliC* gene. This suggested that the genes in this unique region were involved in flagellin glycosylation. To further confirm that the faster migration of the flagellins of *orfA* and *orfN* mutants was because of lack of glycosylation, the wild-type PAK flagellin was chemically deglycosylated by using trifluoromethanesulfonic acid. Treatment with this reagent chemically removes N- and O-linked glycosyl moieties from proteins. After trifluoromethanesulfonic acid treatment, the PAK flagellin migrated faster with the electrophoretic mobility of the nonglycosylated flagellin from the *orfA* mutant (Fig. 2*C*). The defect in flagellin glycosylation in these mutants was complemented when the mutated genes were

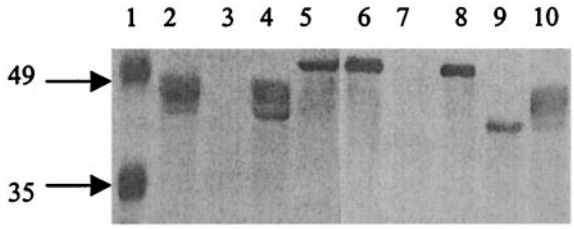

Fig. 3. Flagellin specificity of the glycosylation apparatus. Lane 2, wild-type strain PAK (a-type); lane 3, PAK*fliC* mutant; lane 4, PAK*fliC* mutant carrying a plasmid pPT244 expressing the a-type flagellin; lane 5, PAK flagellin mutant carrying a plasmid pDN186970 expressing the b-type flagellin; lane 6, wild-type strain PAO1 (b-type); lane 7, PAO1*fliC* mutant; lane 8, PAO1*fliC* mutant carrying a plasmid pDN186970 expressing the b-type flagellin; lane 9, PAO1*fliC* mutant carrying a plasmid pPT244 expressing the a-type flagellin; and lane 10, wild-type strain PAK (a-type). Size markers, in kDa, are in lane 1.

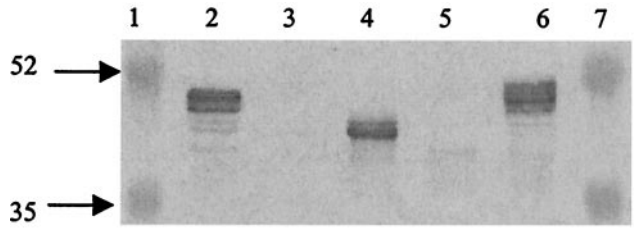

Fig. 4. The glycosylation island contains all of the missing requirements for flagellin glycosylation in the b-type strain PAO1. Lanes 1 and 7 contain molecular size markers. Lane 2, wild-type strain PAK; lane 3, PAO1 flagellin mutant carrying the vector (pMMB66); lane 4, PAO1 flagellin mutant carrying the plasmid pMMB194H containing incomplete glycosylation island and the a-type flagellin; lane 5, PAO1 flagellin mutant carrying the vector control (pSP329Gm); lane 6, PAO1 flagellin mutant carrying the plasmid pSP329Gm10EH8 containing the complete glycosylation island along with the a-type flagellin gene.

provided on plasmids (Fig. 2). Examination of these mutants on motility agar indicated that the mutations had no effect on motility as measured by the zone of migration compared with the wild-type strain (data not shown). Furthermore, because *orfA* and *orfN* mutants produced functional flagella, these mutations did not exert polar effects on genes specifying flagellum assembly components. Because a number of the genes in the island shared similarities to genes involved in LPS biosynthesis in some organisms, a preparation of LPS from the two mutants was examined for structural changes based on the patterns of migration, following SDS/PAGE. No defect in LPS structure could be detected by silver staining of these preparations (data not shown).

Coinheritance of the Glycosylation Island with a-Type Flagellins. Using antiflagellin antibody, we assembled a set of 15 flagellar a-type and 15 flagellar b-type isolates of *P. aeruginosa*. They represent independent clinical isolates from blood, urine, and cystic fibrosis sputum, as well as a limited number of common laboratory strains. These strains were examined blindly for the presence of the *orfA* gene by PCR. There was an absolute concordance between the expression of a-type flagellin and the presence of this gene, and all b-type strains lacked *orfA*. The absence of *orfA* in b-type strains was verified by Southern blot analysis (data not shown). This finding suggests that the strains of *P. aeruginosa* that glycosylate their flagellin evolved from a common ancestor that carried an a-type *fliC* gene and acquired the genes encoding the glycosylation machinery very early. A more extensive survey of the various *P. aeruginosa* isolates showed that the a-type strains are more common, regardless of the source of the bacteria. In our collection of 437 independent isolates, a-type strains represented 65% of blood isolates, 74% nonmucoid and 80% mucoid isolates from cystic fibrosis patients, and 69% of environmental isolates.

Only a-Type Flagellins Are Substrates for Glycosylation. To examine whether the glycosylation defect in b-type flagellins is solely because of the absence of the genes of the glycosylation island, we introduced the *fliC* gene of a b-type *P. aeruginosa* (PAO1) into a *fliC* mutant of an a-type strain (PAK). Conversely, the normally glycosylated PAK flagellin was introduced into the *fliC* mutant of PAO1. Each strain expressed the heterologous flagellin, and the motility defect of each *fliC* mutant was complemented (data not shown). However, Western immunoblot analysis of *fliC* mutants carrying the flagellin gene of the opposite type showed that neither strain was capable of glycosylating heterologous flagellin (Fig. 3). This result was consistent with the interpretation that the b-type flagellin was lacking in the appropriate glycosylation sites, as even the a-type strain did not glycosylate the b-type flagellin. Furthermore, these findings suggested that the b-type cells did not possess other genes that were capable of substituting for the products of the genes in the glycosylation island of the a-type strains.

The PAK Glycosylation Island Contains the Genetic Elements Missing from Strain PAO1. The observation that a large DNA element had been inserted into the genes of the flagellar regulon of PAK and other a-type strains suggested that this element conferred glycosylation on these strains and that it may contain all of the essential glycosylation functions that were missing from the b-type strain PAO1. To test this hypothesis, the region of the PAK chromosome containing the entire glycosylation island and the PAK flagellin gene was reconstructed, and it was introduced into a *fliC* mutant of PAO1. The recombinant plasmid pSP329Gm10EH8, carrying the glycosylation island genes as an 18-kb insert, was capable of complementing the motility defect of the *fliC* PAO1. Moreover, in this background, the heterologous flagellin is glycosylated (Fig. 4). Therefore, this segment of DNA contains the missing glycosylation functions of PAO1, which are very likely absent in all b-type strains.

Annotation of the PAK-Specific Glycosylation Island. The annotation of the region between *flgL* and *fliC* is summarized in Table 1 (24, 28, 32–38). The island contains 14 ORFs (*orf*s). Two of these (*orfI* and *orfK*) encode products of no known function, and these are designated as hypothetical unknown genes. *OrfH* and *orfL* had low level of homology with nodulation protein NoeI of *Sinorhizobium fredii* and methyltransferase of *N. meningitidis,* respectively (37). A group of genes within the island is similar to genes encoding proteins that function in various biosynthetic pathways. These include *orfB*, *orfC*, and *orfD*, which encode homologues of carriers or modifiers of acyl groups. More relevant to the function of the glycosylation island is the presence of genes that encode products with similarities to enzymes that are involved in the biosynthesis or modification of polysaccharides. These include the product of *orfA*, a homologue of nucleotide–sugar aminotransferase; *orfE*, encoding an acetyltransferase (potentially involved in modification of sugars or serine), and *orfJ*, a sugar–cytidine transferase similar to the enzymes involved in the biosynthesis of the core portion of the Gram-negative LPS. Finally, *orfN*, present in both PAK and PAO1, specifies a homologue of glycosyltransferase, with a significant similarity to the enzymes involved in the addition of sugars to the O-side chains of LPS (27) or formation of polysaccharides by *Enterococcus faecalis* (39).

Table 1. Similarities of *P. aeruginosa* PAK ORFs to proteins in the databases

ORF	Size, aa	Homologue/organism	% identity, aa	Function of homologues	Acc. no. of homologue (ref.)
A	467	VioA/*E. coli*	58	Synthesis of dTDP-4-amino-4,6 dideoxyglucose (dTDP-viosamine) (nucleotide sugar transaminase)	AF125322 (28)
B	75	Putative acyl carrier protein/*C. jejuni*	33	Carrier of the growing fatty acid chain in fatty acid biosynthesis	AL139078 (32)
C	334	FABH, 3-oxoacyl-[acyl-carrier-protein]-synthase/*Streptomyces coelicolor*	36	Catalyzes the condensation reaction of fatty acid synthesis by the addition of an acyl acceptor of two carbons from malonyl-ACP	T34914
D	220	FABG, 3-oxoacyl-[acyl-carrier-protein]-reductase/*Chlamydia pneumonia*	36	First reduction step in the fatty acid biosynthesis pathway	Q9Z8P2 (33)
E	211	Serine *O*-acetyltransferase/*Bacillus subtilis*	33	Involved in methylation of Nod factors (NoeI)	E70037 (34)
F	376	Putative aromatic ring hydroxylating dioxygenase alpha subunit/*Sphingomonas sp.*	24	Involved in napthalene catabolic pathway	U65001
G	207	Putative acetyl transferase/ *Legionella pneumophilia*	21	Involved in LPS biosynthesis	AJ007311 (35)
H	380	Nodulation protein NoeI/*S. fredii*	33	Involved in Nod factor biosynthesis	AF228683
I	426	No homologue			
J	256	2-deoxy-manno-octulosonate cytidyl transferase/*E. coli*	35	Activates KDO for incorporation into LPS in Gram negative bacteria	P42216 (36)
K	190	No homologue			
L	368	3-demethylubiquinone-9 3-methyltransferase/*N. meningitidis*	24	Involved in ubiquinone biosynthesis pathway	AL162753 (37)
M	539	*cmtG*, 4-hydroxy-2-oxovalerate aldolase/*Pseudomonas putida*	27	Involved in *p*-isopropylbenzoate (*p*-cumate) catabolic pathway	Q51983 (38)
N	1138	O-antigen biosynthesis protein RfbC/*Myxococcus xanthus*	29	O-antigen biosynthesis	Q50864 (24)

Discussion

P. aeruginosa expresses a large number of virulence factors, including the organelles of motility and attachment, which are specified by genes of the flagellar biosynthetic locus (40). The regulatory network that controls the coordinate expression of proteins necessary for the formation of the flagellum expression of the flagellum by this organism appears to be significantly different from that observed for the well studied *Enterobacteriaceae* (41). Furthermore, unlike in the Enteric bacteria, the flagellin subunit of one of the two variants of *P. aeruginosa* flagellins (the a-type) undergoes posttranslational modification by covalent attachment of glycosyl moieties to one or several sites within the polypeptide.

In this report, we have identified a unique genomic island that is present in *P. aeruginosa* PAK and very likely in all strains expressing a-type flagellin. The proteins encoded within this island are capable of glycosylating exclusively a-type flagellins. Furthermore, the comparison of the genes within this region of the *P. aeruginosa* chromosome indicates that it is highly polymorphic. In addition to the presence of the glycosylation island in PAK (and the absence of three genes from the corresponding site), the region contains several genes that are poorly conserved. Specifically, *flgL*, *fliC*, *fleL*, *fliD*, *fliS*, and the 3′part of *fleP* share only modest levels of sequence similarity between PAK and PAO1, even though they encode proteins that almost certainly play the same role in flagellar biosynthesis (Fig. 1). The comparison of the sequences of this region suggests that horizontal gene transfer to the flagellar a-type strains, such as PAK used in this study, and b-type strains (which includes PAO1) may be responsible for the observed sequence variability. The most likely mechanism of genetic exchange appears to be the acquisition of distinct a- or b-type cassettes followed by incorporation of these cassettes into the genome by reciprocal recombination. Based on the comparison of the highly conserved genes between PAK and PAO1, the most likely sites of recombination can be mapped in the coding sequence of *fleP* and *flgJ*, which flank this highly polymorphic segment of the bacterial chromosome (Fig. 1). The last 31 nucleotides of the *fleP* gene in PAK and the downstream sequences containing *fleQ* and *fleSR* are identical. The precise site of recombination within *flgJ* is more difficult to determine. The sequences of the *flgK* genes in PAK and PAO1 share 86% sequence identity, with two segments corresponding to the first 250 nucleotides of the coding sequence and the middle portion of 212 being nearly identical, which would provide significant regions of homology for recombination following acquisition of DNA from a heterologous strain. Because the %(G + C) content of the island is 63.3%, which is not significantly different from the overall base composition of the genome of strains PAO1 (66%) (42), the origin of the island is possibly another *Pseudomonas* or a bacterium of a comparable GC-rich DNA.

The proteins encoded within the glycosylation island in strain PAK do not show identity to other bacterial proteins, yet a number of them share significant similarities to enzymes involved in the synthesis, activation, or polymerization of sugars. Interestingly, the only gene encoding a *fliC*-linked glycosyltransferase (*orfN*) is present in both PAK and PAO1. However, the level of identity between the two proteins is only 37%. It is

conceivable that the synthesis and attachment of the complete sugar moiety on an a-type flagellin polypeptide proceeds through a pathway that includes not only enzymes within the glycosylation island but additional products such as OrfN, and enzymes encoded at different locations within the chromosome.

There is increasing evidence that glycosylation of surface appendages is fairly common among different Gram-negative bacterial species. These include addition of various sugars to pili expressed by a wide range of bacteria (2, 4, 43, 44) and to flagellins of *A. tumefaciens* (14), *H. felis* and possibly *Helicobacter pylori* (12), *P. aeruginosa* (4), and *Clostridium tyrobutyricum* (45). Furthermore, a homologue of *orfN*, a putative glycosyltransferase gene, is present in a symbiont of *Riftia*, immediately adjacent to its *fliC* gene (46). The precise role of glycosylation in flagellar function is unclear at this time. Because the subunits are not modified in *P. aeruginosa* expressing the b-type flagellin, and the bacteria are motile by standard motility assay, glycosylation is very likely not necessary for the motility function when assayed under laboratory conditions. However, the strong association of the glycosylation with certain pathogenic strains of *P. aeruginosa* suggests that there may be certain functions for glycosylation *in vivo*. For example, *P. aeruginosa* may use a different mechanism to move in mucosal tissues from those that are manifested by swarming behavior in soft agar.

Although the location of the glycosylation island within the flagellar locus suggests that flagellin is the primary substrate for this modification process, we cannot exclude the possibility that other proteins in *P. aeruginosa* may be also glycosylated and that other accessory genes provide substrates for this pathway.

A comprehensive analysis of the different proteins that are modified through the activities of the glycosylation island-encoded enzymes may provide new insights into the biologic function of these proteins. Moreover, the glycosylation of proteins by pathogenic microorganisms may endow these with new effector functions or aid in the evasion of host defense, which may be an important component of disease-causing capabilities of a number of bacterial pathogens.

This work was supported by Public Health Service Grant AI 45014 (to R.R.).

1. Lechner, J. & Wieland, F. (1989) *Annu. Rev. Biochem.* **58,** 173–194.
2. Stimson, E., Makepeace, K., Dell, A., Morris, H. R., Payne, G. & Saunders, J. R. (1995) *Mol. Microbiol.* **17,** 1201–1214.
3. Parge, H. E., Forest, K. T., Hickey, M. J., Christensen, D. A., Getzoff, E. D. & Tainer, J. A. (1995) *Nature (London)* **378,** 32–38.
4. Castric, P. (1995) *Microbiology* **141,** 1247–1254.
5. Kuo, C., Takahashi, N., Swanson, A. F., Ozeki, Y. & Hakomori, S. (1996) *J. Clin. Invest.* **98,** 2813–2818.
6. McBride, J. W., Yu, X. J. & Walker, D. H. (2000) *Infect. Immun.* **68,** 13–18.
7. Lindenthal, C. & Elsinghorst, E. A. (1999) *Infect. Immun.* **67,** 4084–4091.
8. Brimer, C. D. & Montie, T. C. (1998) *J. Bacteriol.* **178,** 3209–3217.
9. Doig, P., Kinsella, N., Guerry, P. & Trust, T. J. (1996) *Mol. Microbiol.* **19,** 379–387.
10. Wyss, C. (1998) *Infect. Immun.* **66,** 5751–5754.
11. Ge, Y., Li, C., Corum, L., Slaughter, C. A. & Charon, N. W. (1998) *J. Bacteriol.* **180,** 2418–2425.
12. Josenhans, C., Ferrero, R. L., Labigne, A. & Suerbaum, S. (1999) *Mol. Microbiol.* **33,** 350–362.
13. LeClerc, G., Wang, S. P. & Ely, B. (1998) *J. Bacteriol.* **180,** 5010–5019.
14. Deakin, W. J., Parker, V. E., Wright, E. L., Ashcroft, K. J., Loake, G. J. & Shaw, C. H. (1999) *Microbiology* **145,** 1397–1407.
15. Syzmanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. (1999) *Mol. Microbiol.* **32,** 1022–1030.
16. Romain, F., Horn, C., Pescher, P., Abdelkader, N., Riviere, M., Puzo, G., Barzu, O. & Marchal, G. (1999) *Infect. Immun.* **67,** 5567–5572.
17. Brockl, G., Behr, M., Fabry, S., Hensel, R., Kaulewitz, H., Biendl, E. & Konig, H. (1991) *Eur. J. Biochem.* **199,** 147–152.
18. Reinhold, B. B., Hauer, C. R., Plummer, T. H. & Reinhold, V. N. (1995) *J. Biol. Chem.* **270,** 13197–13203.
19. Lanyi, B. (1970) *Acta Microbiol. Acad. Sci. Hung.* **17,** 35–48.
20. Spangenberg, C., Heuer, T., Burger, C. & Tummler, B. (1996) *FEBS Lett.* **396,** 213–217.
21. Totten, P. A. & Lory, S. (1990) *J. Bacteriol.* **172,** 7188–7199.
22. Simpson, D. A., Ramphal, R. & Lory, S. (1992) *Infect. Immun.* **60,** 3771–3779.
23. Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S. & Ramphal, R. (1996) *Infect. Immun.* **64,** 2130–2136.
24. Guo, D., Bowden, M. G., Pershad, R. & Kaplan, H. B. (1996) *J. Bacteriol.* **178,** 1631–1639.
25. Fleiszig, S. M. J., Arora, S. K., Van, R. & Ramphal, R. (2001) *Infect. Immun.* **69,** 4931–4937.
26. Nunn, D., Bergman, S. & Lory S. (1990) *J. Bacteriol.* **172,** 2911–2919.
27. Furste, J. P., Pansegrau, W., Frank, R., Blocker, H., Scholz, P., Bagdasarian, M. & Lanka, E. (1986) *Gene* **48,** 119–131.
28. Marolda, C. L. & Valvano, M. A. (1993) *J. Bacteriol.* **175,** 148–158.
29. Hitchcock, P. J. & Brown, T. M. (1983) *J. Bacteriol.* **154,** 269–277.
30. Liang, X., Pham, Q. T., Olson, M. V. & Lory, S. (2001) *J. Bacteriol.* **183,** 843–853.
31. Arora, S. K., Dasgupta, N., Lory, S. & Ramphal, R. (2000) *Infect. Immun.* **68,** 1474–1479.
32. Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., *et al*. (2000) *Nature (London)* **403,** 665–668.
33. Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. & Stephens, R. S. (1999) *Nat. Genet.* **21,** 385–389.
34. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature (London)* **390,** 249–256.
35. Luneberg, E., Zetzmann, N., Alber, D., Knirel, Y. A., Kooistra, O., Zahringer, U. & Frosch, M. (2000) *Int. J. Med. Microbiol.* **290,** 37–49.
36. Pazzani, C., Rosenow, C., Boulnois, G. J., Bronner, D., Jann, K. & Roberts, I. S. (1993) *J. Bacteriol.* **175,** 5978–5983.
37. Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., *et al*. (2000) *Nature (London)* **404,** 502–506.
38. Eaton, R. W. (1996) *J. Bacteriol.* **178,** 1351–1362.
39. Xu, Y., Murray, B. E. & Weinstock, G. M. (1998) *Infect. Immun.* **66,** 4313–4323.
40. Ramphal, R., Arora, S. K. & Ritchings, B. W. (1996) *Am. J. Respir. Crit. Care Med.* **154,** S170–S174.
41. Macnab, R. M. (1999) *J. Bacteriol.* **181,** 7149–7153.
42. Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., *et al.* (2000) *Nature (London)* **406,** 959–964.
43. Tomoeda, M., Inuzuka, M. & Date, T. (1975) *Prog. Biophys. Mol. Biol.* **30,** 23–56.
44. Gubish, E. R., Chen, K. C. S. & Buchanan, T. M. (1982) *Infect. Immun.* **37,** 189–194.
45. Arnold, F., Bedouet, L., Batina, P., Robreau, G., Talbot, F., Lecher, P. & Malcoste, R. (1998) *Microbiol. Immunol.* **42,** 23–31.
46. Millikan, D. S., Felbeck, H. & Stein, J. L. (1999) *Appl. Environ. Microbiol.* **65,** 3129–3133.