

Biology

Microbiology & Immunology fields

Okayama University

Year 2007

Flagellin Glycans from two pathovars of
Pseudomonas syringae contain rhamnose
in D and L configurations in different
ratios and modified
4-amino-4,6-dideoxyglucose

Kasumi Takeuchi, *National Institute of Agrobiological Sciences*

Hiroshi Ono, *National Food Research Institute*

Mitsuru Yoshida, *National Food Research Institute*

Tadashi Ishii, *Forestry and Forest Products Research Institute*

Etsuko Katoh, *National Institute of Agrobiological Sciences*

Fumiko Taguchi, *Okayama University*

Ryuji Miki, *Okayama University*

Katsuyoshi Murata, *National Institute of Agrobiological Sciences*

Hanae Kaku, *Meiji University*

Yuki Ichinose, *Okayama University*

This paper is posted at eScholarship@OUDIR : Okayama University Digital Information Repository.

http://escholarship.lib.okayama-u.ac.jp/microbiology_and_immunology/6

1 **Title: Flagellin glycans from two pathovars of *Pseudomonas syringae* contain rhamnose**
2 **at different ratios of D- and L- configurations and modified 4-amino-4,6-dideoxyglucose**

3
4 Author's names: Kasumi Takeuchi,^{1*§} Hiroshi Ono,^{2§} Mitsuru Yoshida,² Tadashi Ishii,³ Etsuko
5 Katoh,¹ Fumiko Taguchi,⁴ Ryuji Miki,⁴ Katsuyoshi Murata,¹ Hanae Kaku,⁵ and Yuki Ichinose⁴

6
7 i) Affiliations and addresses: ¹National Institute of Agrobiological Sciences, Kannondai 2-1-2,
8 Tsukuba, Ibaraki 305-8602, Japan. ²National Food Research Institute, 2-1-12 Kannondai,
9 Tsukuba, Ibaraki 305-8642, Japan. ³Forestry and Forest Products Research Institute, 1,
10 Matsunosato, Tsukuba, Ibaraki 305-8687, Japan. ⁴Graduate School of Natural Science and
11 Technology, Okayama University, Tsushima-naka 1-1-1, Okayama 700-8530 Japan.
12 ⁵Faculty of Agriculture, Meiji University, Kawasaki, Kanagawa 214-8571, Japan

13
14 ii) *For correspondence: E-mail kasumit@affrc.go.jp; TEL (+81) 29 838 7005; FAX. (+81) 29
15 838 7408

16
17 iii) Running title: Glycosylation of *P. syringae* flagellin

18
19 iv) Footnotes: [§] The first two authors contributed equally to this work.

20
21

1 Abstract

2 Flagellins from *Pseudomonas syringae* pv. *glycinea* race 4 (Pgl4) and *P. s.* pv. *tabaci* 6605
3 (Pta6605) have been found to be glycosylated. Glycosylation of flagellin is essential for
4 bacterial virulence and is also involved in the determination of host specificity. Flagellin
5 glycans from both pathovars were characterized and common sites of glycosylation were
6 identified on six serine residues (positions 143, 164, 176, 183, 193 and 201). The structure of
7 the glycan at serine 201 (S201) of flagellin from each pathovar was determined by sugar
8 composition analysis, mass spectrometry and ¹H and ¹³C NMR spectroscopy. These analyses
9 showed that the S201 glycans from both pathovars were composed of a common unique
10 trisaccharide consisting of two rhamnosyl (Rha) residues and one modified
11 4-amino-4,6-dideoxyglucosyl (Qui4N) residue,
12 β -D-Quip4N(3-hydroxy-1-oxobutyl)2Me-(1→3)- α -L-Rhap-(1→2)- α -L-Rhap. Furthermore,
13 mass analysis suggests that the glycans on each of the six serine residues are composed of a
14 similar trisaccharide unit. Determination of the enantiomeric ratio of Rha from the flagellin
15 proteins showed that flagellin from Pta6605 consisted solely of L-Rha, whereas Pgl4 flagellin
16 contained both L-Rha and D-Rha at a molar ratio of about 4:1. Together with our previous
17 study, we conclude that these structures of the flagellin glycans may be important for virulence
18 and host specificity of *P. syringae*.

19

20 Introduction

21 Glycosylation of pathogenic bacterial cell surface proteins, such as flagellin and pilin, has
22 recently been recognized as an important factor in host-pathogen interactions (3, 19). Flagellin
23 glycosylation is found in animal pathogens and the genes required for glycosylation and
24 glycan structure have been characterized in several bacteria, such as *Campylobacter jejuni* (10,
25 32), *Pseudomonas aeruginosa* (1, 25, 34) and *Helicobacter pylori* (24). Similarly, flagellins of

1 some plant pathogens were found to be glycosylated in *P. syringae* pv. *tabaci* 6605 (Pta6605),
2 pv. *glycinea* race 4 (Pgl4) and pv. *tomato* DC3000 (28) as well as *Acidovorax avenae* (4).
3 Although the possible biological significance of flagellin glycosylation is frequently discussed,
4 experimental evidence has been restricted to our study on *P. syringae* (13, 30, 31) and the
5 studies by other groups on *P. aeruginosa* (2, 33) and on *Campylobacter jejuni* and *C. coli* (11).

6 The phytopathogenic bacterium *P. syringae* is classified as a pathovar by its virulence
7 toward different host plant species. In our previous study, flagellin from *P. syringae* was found
8 to be an elicitor that causes a hypersensitive reaction (HR) of non-host plants (28). Moreover,
9 the HR-inducing activity is thought to be dependent on glycosylation (29). The significance of
10 glycosylation is particularly notable in the two pathovars of *P. syringae*, Pgl4 and Pta6605,
11 because although the respective flagellins display absolute amino acid sequence conservation,
12 the HR-inducing activities are different. Recently, we found that a flagellin glycosylation
13 island, which possesses putative glycosyltransferase genes, is required for flagellin
14 glycosylation in Pgl4 and Pta6605, and that deletion of these genes reduced both virulence to
15 their respective host plants and HR-inducing activity for non-host plants (13, 30, 31). These
16 results demonstrate that flagellin glycosylation plays an important role in determining host
17 specificity of each pathovar of *P. syringae*. We have identified six glycosylated serine residues
18 in flagellin from Pta6605 (30). These serine residues are all localized on the predicted
19 surface-exposed domain when the flagellin folds as a monomer in the assembled filament.
20 Based on studies of Ser/Ala-substituted mutants and glycosylation island deletion mutants, we
21 demonstrated that flagellin glycosylation is essential for bacterial adhesion, swarming motility
22 and virulence on host tobacco leaves. Thus, flagellin glycosylation plays a key role not only as
23 the determinant of HR-induction activity, but also in virulence-related bacterial characteristics.

24 Although biological and mutational studies in *P. syringae* emphasized the importance
25 of flagellin glycosylation for bacterial virulence and host specificity, there was no direct

1 structural information on the flagellin glycans. Here we report structural characterization of
2 the flagellin glycans in *P. syringae*.

3

4 **Materials and methods**

5 **Bacterial strains and culture conditions**

6 The bacteria used in this study are listed in Table 1. Pgl4 and Pta6605 and their
7 derivative mutants were maintained in King's B (KB) medium at 27°C. *Escherichia coli* strains
8 were grown at 37°C in Luria-Bertani (LB) medium.

9 **Site-directed mutagenesis of glycosylated residues of flagellin in Pgl4**

10 Ser/Ala substituted mutants of Pgl4 were obtained by first generating a deletion mutant
11 of the flagellin coding region (*fliC*) of Pgl4. The resultant mutant ($\Delta fliC$) was then
12 complemented with the *fliC* region possessing the desired point mutation(s). The $\Delta fliC$ mutant of
13 Pgl4 was made using a previously reported method (26) with a slight modification. One of the
14 primers for the downstream region of *fliC*, designated PC4, was modified to
15 5'-GATCGCGTAAGTACCGTTGA-3'. Methods for site-directed mutagenesis and
16 complementation by homologous recombination were described previously (30). The Ser/Ala
17 substituted mutants were designated as follows: race 4-S143A, race 4-S164A, race 4-S176A,
18 race 4-S183A, race 4-S193A and race 4-S201A. A mutant (race 4-6 S/A) with six serine
19 substitutions (race 4-S143A, S164A, S176A, S183A, S193A and S201A) was also constructed
20 by the same method.

21 **Purification of flagellin and preparation of glycosylated peptides**

22 *P. syringae* was incubated in LB medium containing 10 mM MgCl₂ for 48 h at 25°C.
23 The cells were harvested by centrifugation and resuspended in 1/3 volume of minimal medium
24 (MM; 50 mM potassium phosphate buffer, 7.6 mM (NH₄)₂SO₄, 1.7 mM MgCl₂ and 1.7 mM
25 NaCl, pH 5.7) supplemented with 10 mM each of mannitol and fructose, and then incubated for

1 24 h at 23°C. Flagellin was purified by the method of Taguchi et al. (28). For the identification
2 of glycan components of flagellin proteins, purified proteins were subjected to SDS-PAGE and a
3 band at 32 kDa was excised, crushed and mixed with distilled water. The extracted flagellin was
4 then concentrated using spin columns (Vivaspin VS0403 10,000 MWCO, Vivascience,
5 Hannover, Germany). For purification of glycosylated peptides, purified flagellin from each
6 pathovar was digested with aspartic N-peptidase (Boehringer Mannheim, Mannheim, Germany)
7 at 35°C for 20 h in Tris-HCl buffer (pH 8.0). The resultant peptides with 0.1% (v/v)
8 trifluoroacetic acid (TFA) were subjected to reverse-phase high performance liquid
9 chromatography (HPLC) using a 2.0 × 250 mm TSKgel ODS-80TS column (Tosoh, Tokyo,
10 Japan) as reported by Taguchi et al., 2006. For large-scale preparation of the peptides,
11 approximately 4 mg of digested flagellin was applied to a TSKgel ODS-120TS (4.6 × 150 mm,
12 Tosoh) and eluted at a flow rate of 1.0 ml/min with a linear gradient of 9-90% aqueous
13 acetonitrile (0.1% TFA) for 87 min. UV detection was carried out at 210 nm and fractions were
14 collected every minute. The target peptide (D200-A211) was identified by N-terminal amino
15 acid sequencing using a protein sequencer (Procise 494 HT protein sequencing system, Applied
16 Biosystems, Tokyo, Japan).

17 **Mass Spectrometry: Comparison of mass spectra in intact flagellins and the peptides**
18 **N136-K255 from Pta6605 and Pgl4**

19 Flagellins from wild-type and mutant strains of Pta6605 and Pgl4 were digested with
20 lysyl endopeptidase (Wako, Osaka, Japan) at 37°C overnight in 10 mM Tris-HCl buffer (pH 9.0).
21 Each intact or digested protein was dissolved in water with 0.1% TFA and mixed with an equal
22 volume of matrix solution [a saturated solution of sinapinic acid in 33% acetonitrile/water with
23 0.1% TFA (v/v)], and deposited on a target plate. Samples were analyzed using a Biflex III
24 spectrometer (Bruker Daltonik GmbH, Bremen, Germany) and matrix-assisted laser
25 desorption/ionization time of flight (MALDI-TOF) mass spectra were recorded in a linear,

1 positive-ion mode with mass accuracy of 0.1%.

2 **MALDI-QIT-TOF MS analysis of the glycopeptide D200-A211**

3 The HPLC eluate of the glycopeptide D200-A211 from each pathovar (0.5 μ l) was
4 mixed with an equal volume of matrix solution [10 mg/ml of 2,5-dihydroxybenzoic acid in
5 0.06% (v/v) TFA and 40% (v/v) acetonitrile] and deposited on a sample target plate. The mass
6 and MS/MS spectra of the glycopeptide D200-A211 were recorded on an AXIMA quadrupole
7 ion trap (QIT) MALDI-TOF mass spectrometer (Shimadzu, Kyoto, Japan). Both MALDI-TOF
8 mass spectrometers were calibrated using a standard mixture of peptides (Bruker Daltonics,
9 MA).

10 **ESI-Q-TOF MS analysis of the glycopeptide D200-A211**

11 Electrospray ionization (ESI)-TOF MS experiments were conducted using a
12 quadrupole (Q)-TOF mass spectrometer (QSTAR XL, Applied Biosystems) equipped with a
13 nanospray ESI source. The ion-spray voltage was set to 1000 V. For accurate mass
14 measurements, the instrument was calibrated using γ -series fragment ions (22) of
15 Glu-fibrinopeptide B and mass accuracy was within 5 ppm. Prior to analyses, samples were
16 prepared by dissolving them in 30% acetonitrile and 0.1% formic acid. All mass spectra were
17 obtained in positive-ion mode.

18 **Sugar composition analysis of flagellin glycan**

19 The monosaccharide composition of glycans from purified flagellin proteins was
20 analyzed using an ABEE (*p*-aminobenzoic acid ethyl ester) labeling kit (J-oil mills, Tokyo,
21 Japan). Sialic acid, being a non-reducing sugar, is not converted by ABEE. To assess the
22 presence/absence of sialic acid in the flagellin preparation, purified flagellin was treated with
23 *N*-acetylneuraminic acid aldolase in order to release sialic acid residues prior to acid hydrolysis,
24 thereby enabling detection as *N*-acetyl-mannosamine. The subsequent processes of acid
25 hydrolysis, *N*-acetylation and conversion with ABEE were carried out according to the method

1 of Yasuno et al. (36). The resultant ABEE-converted monosaccharide(s) in the aqueous layer
2 were analyzed by reverse-phase HPLC using a Honenpak C18 column (75 mm × 4.6 mm i.d.,
3 J-oil mills), according to the manufacturer's instructions. For quantification of monosaccharides,
4 a set of monosaccharides including glucose, galactose, mannose, arabinose, ribose, fucose,
5 xylose, rhamnose (Rha), *N*-acetylglucosamine, *N*-acetylgalactosamine and
6 *N*-acetylmannosamine, was used as standards.

7 **Determination of D-Rha/L-Rha ratios in flagellins**

8 Enantiomeric ratios of the Rha residues in glycopeptides and intact flagellins were
9 determined using gas chromatography (GC) according to the method of Gerwig et al. (8) with a
10 slight modification. The glycopeptide D200-A211 and intact flagellin protein from each
11 pathovar were subjected to acidic solvolysis with 1N HCl in (*S*)-2-butanol for 16 h at 80°C. The
12 (*S*)-2-butyl glycosides formed were then converted into their trimethylsilyl (TMS) derivatives
13 and analyzed by GC (GC-17A, Shimadzu) using a DB-1 column (30 m × 0.25 mm, J&W
14 Scientific, Folsom, CA) (37). Because D-glycosides of (*S*)-2-butanol and L-glycosides of
15 (*R*)-2-butanol have the same retention time by GC analysis (non-chiral stationary phase
16 separation), L-rhamnoside of (*R*) and (*S*)-2-butanol were prepared as standards for determination
17 of (*S*)-2-butyl-D-rhamnoside and (*S*)-2-butyl-L-rhamnoside, respectively. To confirm two peaks
18 assigned as D- and L-rhamnosides of (*S*)-2-butanol from Pgl4 flagellin, GC-MS was performed
19 according to the method of McNeil and Albersheim (20) with a slight modification. A JMS
20 DX-303 mass spectrometer (JEOL, Tokyo, Japan) was interfaced with a Hewlett-Packard 5890
21 gas chromatograph (Hewlett-Packard, Palo Alto, CA) using an SPB-1 column (30 m × 0.32 mm,
22 Supelco Inc., Bellefonte, PA). GC-MS was performed by Toray Research Center Inc., Kamakura,
23 Japan.

24 **NMR spectroscopy**

25 Lyophilized glycopeptide D200-A211 prepared from Pta6605 flagellin was dissolved

1 in 300 μl of D_2O to give a final concentration of 150 nmol/L (pH 4.2). Glycopeptide D200-A211
2 from Pgl4 was dissolved in 300 μl of D_2O to give a final concentration of 50 nmol/L (pH 2.9).
3 ^1H NMR spectra of the glycopeptides, including ^1H - ^1H correlation spectra (DQF-COSY,
4 TOCSY and NOESY) and ^1H - ^{13}C correlation spectra (HSQC and HMBC), were obtained at
5 800.33 MHz on a Bruker Avance 800 spectrometer with a TCI ($^1\text{H}/^{13}\text{C}\{^{15}\text{N}\}$) CryoProbeTM
6 (Bruker Biospin, Karlsruhe, Germany) at 298K. ^{13}C NMR spectra were obtained at 125.76 MHz
7 on a Bruker Avance 500 spectrometer with a Dual ($^1\text{H}\{^{13}\text{C}\}$) CryoProbeTM (Bruker Biospin) at
8 298K. Methyl signals of 2-methyl-2-propanol, δ_{H} at 1.23 ppm and δ_{C} at 31.3 ppm were used as
9 references for ^1H and ^{13}C chemical shifts.

10

11 **Results**

12 **Identification of the glycosylated amino acid residues in Pgl4 flagellin**

13 Monomeric flagellin of Pgl4 and the corresponding lysyl endopeptidase-digested
14 peptides were subjected to MALDI-TOF MS in order to characterize the modification pattern
15 (Table 2). Intact Pgl4 flagellin purified from the wild-type strain exhibited $[\text{M}+\text{H}]^+$ around m/z
16 32,380 as a broad peak reflecting quantitative heterogeneity in glycosylation (Table 2 and Fig.
17 1C). The m/z values of the three main species (32,380, 32,515 and 32,668) within the broad peak
18 were larger than the predicted values for $[\text{M}+\text{H}]^+$ (i.e., m/z 29,148), based on amino acid
19 sequences, by 3,232 Da 3,367 Da and 3,520 Da, respectively. The molecular mass of flagellin
20 from the $\Delta orf1$ mutant, which lacks the ability to glycosylate (31), was 29,154 Da. This value
21 corresponds within the margins of error of the system to the molecular mass predicted from the
22 deduced amino acid sequence, confirming that flagellin from the $\Delta orf1$ mutant is not
23 glycosylated.

24 In a lysyl endopeptidase-digested peptide mixture from the wild-type strain, three
25 sharp peaks were observed at m/z 15,296, 15,444 and 15,591 (Table 2 and Fig. 1D). These

1 corresponded to N136-K255 peptides, which without modification should give $[M+H]^+$ of m/z
2 12,074. Thus, our results indicate that the N136-K255 peptide is modified with glycan of total
3 molecular mass 3,222 Da, 3,370 Da or 3,517 Da. Because the differences between the delta
4 values ($[\text{observed } m/z \text{ value}] - [m/z \text{ value calculated from the peptide sequence}]$) of intact
5 flagellin and the N136-K255 peptides are almost the same (Table 2), the sites of modification
6 appear to be located between N136-K255 in the primary amino acid sequence of flagellin.

7 Previous studies on flagellin from Pta6605 identified 6 serine residues (S143, S164,
8 S176, S183, S193 and S201) as sites of glycosylation (30). To evaluate the effect of point
9 mutations at each of the 6 serine residues on the glycosylation status of flagellin in Pgl4, one or
10 all of the 6 serine residues were substituted with alanine. The molecular mass of flagellin from
11 each substituted mutant was then determined. Substitution of each serine by alanine decreased
12 the molecular mass of the modified peptide (indicated as the Δ value in Table 2) by an average
13 of 534 Da. The mutated form of flagellin in which all six serine residues had been substituted by
14 alanine (i.e., 6 S/A) gave a molecular mass corresponding to unmodified peptide (Table 2).
15 These results suggest that the sites of glycosylation in Pgl4 and Pta6605 flagellin are identical.
16 Mass spectra of the peptide N136-K255 from each of the six Ser/Ala-substituted mutants
17 showed the same heterogeneity at three major peaks, as observed for the wild-type strain (Table
18 2). The mass average of the intervals between the three peaks in these seven strains was 147 Da.
19 This mass difference suggests that heterogeneity is derived from the number of deoxyhexose
20 units, which is predicted to give rise to a molecular mass difference of 146 Da per unit.

21 The mass spectra of intact flagellin or peptide N136-K255 from Pgl4 and Pta6605
22 were compared (Fig. 1). In our previous study (30), mass spectra of intact flagellin and peptide
23 from wild-type strain Pta6605 showed heterogeneity at 2-3 major peaks (Fig. 1A and 1B).
24 However, in the case of the wild-type strain Pgl4, although the positions of the peaks indicating
25 glycosylation heterogeneity are similar to those of Pta6605, their relative intensities at higher

1 mass values are significantly greater (Fig. 1C and 1D). The molecular mass of flagellin from the
2 $\Delta orf2$ mutant of Pta6605 was reported to be quite variable, with a value intermediate between
3 that of the wild-type and the $\Delta orf1$ mutant (30). The molecular mass of flagellin from the $\Delta orf2$
4 mutant of Pgl4 also showed heterogeneity in the m/z range of 13,292 to 14,858 with more than
5 15 peaks.

6 **Peptide mapping of the Pgl4 flagellin**

7 An HPLC profile of the proteolytic fragments of flagellin was generated by digestion
8 with endoproteinase Asp-N. When flagellin from Pta6605 was digested with this endoproteinase,
9 peptides containing glycosylation sites were mapped to 3 fractions (Fr. 41 for D200-A211, Fr.
10 50 for D168-T187, and Fr. 66 for D139-F167 and E189-I199) (30). Here, we performed the
11 same analysis on flagellin from Pgl4. N-terminal amino acid sequencing confirmed the presence
12 of the former two peptides in Fr. 43 (D200-A211) and Fr. 51 (D168-T187). In Fr. 67 of Pgl4, we
13 detected two peptide sequences DGSAxTMTFQVGS and ETNFxAAIAA (where 'x' denotes an
14 unidentified residue), corresponding to the N-terminal amino acids of D139-F167 (i.e.
15 D139-S151) and almost the entire sequence of E189-I199, respectively. It was not possible to
16 determine the residual C terminal sequence of D139-F167 (i.e. N152-F167) because the
17 concentration of peptide was too low. During sequence analysis, S143 (Fr. 67), S176 (Fr. 51),
18 S183 (Fr. 51), S193 (Fr. 67) and S201 (Fr. 43) were found to have an anomalous retention time,
19 suggesting that the serine residues had undergone modification. These results are consistent with
20 the MS analysis of Ser/Ala-substituted mutants. We were unable to verify whether S164 also
21 runs anomalously because this residue is located too far from the N-terminus (D139) of the
22 peptide for analysis.

23 Among these peptides, D200-A211 bears a single glycosylation site at S201 and a
24 sufficient amount of material for structural analysis could be obtained by preparative
25 chromatography. Therefore, peptide D200-A211 derived from either Pta6605 or Pgl4 was

1 analyzed further by MS/MS and NMR spectroscopy in order to determine the structure of the
2 modification site.

3 **Sugar composition analysis**

4 Sugar composition analysis was carried out on intact flagellin from Pta6605 and Pgl4.
5 In both pathovars, Rha was identified by correspondence of the retention time of its derivative
6 with that of the Rha standard as described in “Materials and Methods” (data not shown).

7 **Determination of D-Rha/L-Rha ratios in glycopeptide D200-A211 and intact flagellin** 8 **proteins**

9 For determination of the enantiomeric ratio of Rha residues on S201, glycopeptide
10 D200-A211 was treated with HCl in (*S*)-2-butanol to form diastereomeric glycosides. In both
11 pathovars, GC analysis of trimethylsilyl derivatives of the (*S*)-2-butyl rhamnoside yielded a peak
12 corresponding to the L-rhamnoside diastereomer, showing that Rha residues on this peptide were
13 exclusively of the L form (data not shown).

14 To elucidate the enantiomeric ratio of the Rha residues of the whole flagellin protein in
15 each pathovar, the intact flagellin proteins were also subjected to solvolysis and converted into
16 trimethylsilyl derivatives of the (*S*)-2-butyl rhamnoside. In the case of Pta6605, the configuration
17 of Rha was shown to be solely of the L form with a retention time of 22.0 min. (Fig. 2A). By
18 contrast, the flagellin glycan from Pgl4 yielded two peaks corresponding to D- and L-Rha at a
19 molar ratio of about 1:4 with retention times of 21.5 min. and 22.0 min., respectively (Fig. 2B).
20 The assignment of these two peaks was confirmed by GC-MS analysis. MS fragmentation
21 patterns of these derivatives were identical to those of the L-Rha standard. Fig. 2C shows the
22 fragmentation patterns from a GC peak with a retention time of 21.5 min.

23 **Structural characterization of glycopeptide D200-A211 by MS analysis**

24 To characterize the structure of flagellin glycan, glycopeptides D200-A211 from
25 Pta6605 and Pgl4 were subjected to MALDI-QIT-TOF MS and MS/MS analyses. The mass

1 spectra of both pathovars' glycans were essentially identical. Fig. 3 shows the mass spectrum of
2 D200-A211 from Pta6605. The $[M+H]^+$ of the peptide was observed at m/z 1814 (inset of Fig.3)
3 while the corresponding calculated value of the sequence DSALQTINSTRA is 1276. Thus,
4 modification of this peptide increased the molecular mass by 538 Da. This result is consistent
5 with the mass difference in the modified part of WT and the S201A-substituted mutant (Table 2
6 for Pgl4; Taguchi et al. (30) for Pta6605). Fig. 3 shows the MS/MS spectrum of the peak at m/z
7 1814 ($[M+H]^+$). An ion observed at m/z 1277 corresponds to the calculated value for the ion
8 DSALQTINSTRA. Those found at m/z 1699 ($[M-115+H]^+$) and 1162
9 ($[DSALQTINSTRA-115+H]^+$) are presumably generated by cleavage of the N-terminal Asp.
10 Ions observed at m/z 1423 and 1569 correspond to peptides with one and two glycosyl residues
11 of molecular mass 146 Da, respectively. These results suggest that the 538 Da glycan is
12 composed of three residues with masses of 246, 146, and 146 Da. The molecular mass difference
13 of 146 Da suggests the presence of a deoxyhexose, consistent with the result from
14 monosaccharide analysis in which only Rha was detected. Thus, the glycan is linked to the
15 serine via two Rha residues. The product with a molecular mass of 246 Da does not correspond
16 to a known saccharide.

17 To obtain structural information on this unidentified residue, accurate mass
18 measurements by ESI-Q-TOF MS analysis were performed. The mass spectra of the peptides
19 derived from the two pathovars were identical (Fig. 4 shows data from Pta6605). The initial
20 mass spectrum of the glycopeptide D200-A211 showed that the glycan of 538 Da is composed
21 of two deoxyhexose residues and one unknown residue with a mass of 246 Da (Fig. 4A).
22 Furthermore, the unit of two deoxyhexose residues is directly attached to the peptide (at S201).
23 This result is in accordance with the MALDI-QIT-TOF MS/MS analysis (Fig. 3). The
24 fragmentation pattern suggests that these residues are located linearly from the distal end in the
25 order unknown residue followed by two deoxyhexose residues. The MS/MS spectrum of the

1 [M+2H]²⁺ observed at *m/z* 907.5 for the peptide D200-A211 (Fig. 4) yielded an intense product
2 ion peak at *m/z* 246.1 (Fig. 4A). The MS/MS analysis of [M+3H]³⁺ observed at *m/z* 605.3 for the
3 peptide D200-A211 (Fig. 4) also yielded an intense product ion at *m/z* 246.1 (data not shown).
4 These data are consistent with the proposal that the glycopeptide (1813 Da) includes a residue of
5 246 Da. By contrast, MS/MS spectra of fragment ions at *m/z* 784.9, 711.9 and 638.8 did not
6 include the product ion at *m/z* 246.1 (Fig. 4B-4D). These observations confirm that the unit of
7 246.1 Da is located at the distal end of the glycan. Accurate mass analysis within 5 ppm of the
8 calculated value of this unknown unit exhibited *m/z* 246.134 (C₁₁H₂₀NO₅, calculated 246.1335)
9 for the ion.

10 **Structure determination of glycan on D200-A211 by NMR analysis**

11 The structure of glycan on D200-A211 from Pta6605 flagellin was elucidated by NMR
12 spectroscopy. Signals of two Rha and the unidentified substructure were observed in addition to
13 those derived from the peptide backbone. Assignment of these signals was performed based on
14 ¹H-¹H correlations on DQF-COSY and TOCSY, and ¹H-¹³C correlations on HSQC and HMBC
15 spectra (Table 3, Fig 5, Fig. 6B). Binding of an L-Rhap to S201 was confirmed by HMBC
16 correlation between β-carbon of S201 and H1 of the L-Rhap (Fig. 5, Fig. 6B). The second
17 L-Rhap is attached at C2 of the L-Rhap linked to S201, as deduced by HMBC correlation of
18 C2-H1 (Fig. 5, Fig. 6B) and low field shift of the C2 (79.9 ppm) (Table 3). The structure of the
19 terminal saccharide was identified as 4-amino-4,6-dideoxyglucose (Qui4N, trivial name
20 viosamine) from the presence of a C6 methyl group, large H-H coupling constants, and high
21 field shift of C4 (58.3 ppm) (Table 3). The attachment position of the Qui4N to L-Rhap was
22 shown to be C3 by HMBC correlations of L-Rhap C3 to Qui4N H1 and Qui4N C1 to L-Rhap
23 H3 (Fig. 5, Fig. 6B). Further modification of the Qui4N residue by *O*-methylation of C2 was
24 indicated by HMBC correlations (Fig. 5, Fig. 6B). The presence of these links was supported by
25 data from the NOESY experiments (Fig. 5, Fig. 6A). The presence of the 3-hydroxybutyryl

1 group and its attachment to Quip4N through an amide link was revealed by COSY and HMBC
2 analyses (Table 3, Fig. 5, Fig. 6B). Moreover, the elemental formula of the modified Quip4N
3 residue (C₁₁H₂₀NO₅) coincides with the result of accurate mass analysis.

4 Quip4N was estimated to be in the D-configuration based on the ¹³C NMR chemical
5 shift of the C1 signal (18, 35). The value of 105.20 ppm indicated an opposing absolute
6 configuration of β-Quip4N and α-L-Rhap in (1→3) linkage. Alternative configuration of Quip4N
7 would lead to a smaller chemical shift by ca. 3 ppm. Therefore, the structure of the glycan was
8 determined to be β-D-Quip4N(3-hydroxy-1-oxobutyl)2Me-(1→3)-α-L-Rhap-(1→2)-α-L-Rhap
9 (Fig. 5). Structural identity of the glycan on D200-A211 from Pgl4 with that from Pta6605 was
10 confirmed by NMR analyses.

11

12 Discussion

13 Bacterial flagellin is one of the best studied molecules containing pathogen- or
14 microbe-associated molecular patterns that can activate basal defense in the form of non-host
15 resistance in plants (14). The synthetic oligopeptide flg22, which was designed from an
16 N-terminal conserved sequence in the D0 interior domain of flagellin from *P. aeruginosa*, elicits
17 plant basal defense responses (7). Thus, flg22 has been defined as a general elicitor. In
18 *Arabidopsis thaliana*, recognition of flg22 is mediated by its binding to a leucine-rich repeat
19 transmembrane receptor kinase FLS2 (5, 9). We have demonstrated the importance of flagellin
20 glycosylation, located on the surface of the flagellar filament, as a determinant of host specificity
21 (29, 30 31). Plant hypersensitive reactions caused by flagellin are attenuated when the flagellin is
22 modified by glycosylation in the manner of a compatible bacterium (29). Thus, there is a
23 plausible hypothesis that flagellin glycosylation plays an important role in masking this pattern
24 in order to avoid recognition by the host plant.

25 Our analysis of Pgl4 flagellin shows that 6 serine residues at positions 143, 164, 176,

1 183, 193 and 201 are glycosylated. The mass value of the major glycan on each serine was about
2 534, while heterogeneity due to the addition of one or two units with a mass value of about 147
3 was also observed. These results are concordant with our previous work on Pta6605 flagellin
4 (30). Analysis of intact flagellin showed that D-Rha is present in the glycan of Pgl4, whereas
5 Rha from Pta6605 flagellin glycan is exclusively of the L-form. Studies on the glycopeptide
6 D200-A211 derived from either Pta6605 or Pgl4 flagellin revealed structural identity of the
7 major glycan on S201, where only L-Rha was found. It is conceivable that minor glycan species
8 comprising one or two more Rha residues might attach via S201. We believe that heterogeneity
9 of glycosylation is not confined to one particular modification site because it was present in each
10 of the six Ser/Ala-substituted mutants. Differences between Pta6605 and Pgl4 may exist in the
11 structure of the glycan with the extra Rha residue(s). Further investigation of glycan moieties
12 attached to serine residues other than S201 is also required in order to explain host specificity of
13 these pathovars in terms of the structure of flagellin.

14 It is intriguing to explore the relationship between the content of D-Rha and
15 glycosylation heterogeneity. For example, the relative intensities of the peaks observed in
16 MALDI-TOF MS analyses might reflect the content of D-Rha. In accordance with heterogeneity
17 of flagellin glycosylation, D-Rha may be dispersed amongst all or some of the six glycans.
18 Alternatively, D-Rha may be attached to specific serine residue(s) in Pgl4 flagellin. If so, such
19 residue(s) may be significant in determining host specificity. More precise analysis of each
20 glycan and utilization of Ser/Ala-substituted mutants will be helpful in determining the
21 localization of D-Rha in Pgl4 flagellin.

22 Rha is also reported to be a common major component of lipopolysaccharides (LPS) in
23 these two pseudomonad pathogens (17). Although naturally-occurring Rha is mainly present in
24 the L-form, both the D- and L-forms have been found in O-antigens of *P. syringae*. Furthermore,
25 emerging patterns in the chain of rhamnan are thought to correlate with serogroups (17). This

1 may explain the significance of the chirality of Rha in LPS. LPS is recognized by mammals
2 differently according to its constituent parts. Internal/conserved domains, such as lipid A, are
3 important for the innate immune response, whereas the surface exposed/highly variable domains,
4 such as O-antigen, determine antigenic specificity (21). This prototype of domain distinction of
5 activities is also applicable to flagellin. Thus, it is reasonable to propose that flagellin glycans on
6 the putative surface exposed domain are responsible for determining host specificity (38). A
7 complete structural characterization of flagellin should help to elucidate how plants recognize
8 flagellin glycan and how bacteria evade such recognition.

9 In animals, flagellin recognition is mediated by Toll-like receptor 5 (TLR5; 12). The
10 conserved N-terminal region of flagellin is reported to be important for binding to TLR5 (27).
11 Notably, the role of flagellin glycosylation of animal pathogenic bacteria in innate immunity is
12 just beginning to be elucidated. In the case of *P. aeruginosa* strains PAK and PAO1, the
13 virulence of flagellar glycosylation mutants in mice was significantly attenuated (2). In both of
14 these strains, flagellin glycosylation plays an important role in the ability of flagellin to stimulate
15 interleukin-8 release from human lung carcinoma cells (33). These results suggest that flagellin
16 glycans might be responsible for stimulation of inflammation. In *P. aeruginosa* strain PAK, a
17 glycan consisting of 11 monosaccharides is linked to the flagellin protein through a Rha residue
18 at the two glycosylation sites (25). Thus, Rha is a common component of flagellin glycan in *P.*
19 *syringae* and *P. aeruginosa*.

20 The distal residue of the glycan on S201 is the modified unique amino sugar, Qui4N.
21 MALDI-TOF MS analyses of the six Ser/Ala-substituted mutants suggest that the major glycan
22 on each serine residue is a trisaccharide composed of modified Qui4N and two Rha residues.
23 Only Rha was detected in sugar composition and enantiomeric ratio analyses by GC despite the
24 presence of the modified Qui4N. In *Vibrio* LPS, 4-amino-4,6-dideoxymannose was produced in
25 abundance by mild acid hydrolysis (15). Thus, 4-amino-4,6-dideoxyhexoses may be destroyed

1 by more vigorous chemical procedures. On the other hand, in *Bacillus anthracis* exosporium,
2 Qui4N(3-hydroxy-3-methyl-1-oxobutyl)2Me was detected after methanolysis and acetylation (6).
3 It might be possible to detect modified Qui4N in our study by adopting such methods. LPS of *P.*
4 *aeruginosa* is known to be rich in unusual amino sugars, some of them with hydroxybutyryl
5 groups instead of acetylation (16). This suggests the existence of a common synthetic pathway
6 for flagellin glycans and LPS. Qui4N was also detected as one of the components of *P.*
7 *aeruginosa* PAK flagellin glycan (25). Thus, there is a clear structural similarity of the glycans
8 of *P. syringae* and *P. aeruginosa* a-type flagellins, in addition to the attachment of Rha to the
9 peptide backbone. The Orf1 and Orf2 products of the flagellin glycosylation island of *P.*
10 *syringae* are similar to OrfN (fgtA, for flagellar glycosyltransferase) of *P. aeruginosa* a-type
11 strain PAK (i.e. 32 and 38% identity, respectively) (30, 31). Orf1 and Orf2 also show similarity
12 to PA1091 (fgtA) in *P. aeruginosa* b-type strain PAO (43 and 34% identity, respectively). OrfN
13 and PA1091 are considered to transfer deoxyhexose to the protein backbone (25, 34), indicating
14 similarity to the function of Orf1 in *P. syringae*. Thus, these putative glycosyltransferases may
15 possess a common enzymatic activity in pseudomonad pathogens. We propose that the genes
16 *orf1* and *orf2* in the *P. syringae* glycosylation island be renamed *fgt-1* and *fgt-2*, respectively.
17 Notably, the homologue of *orfA*, which belongs to the PAK glycosylation island but is not found
18 in PAO1 of *P. aeruginosa*, is located upstream of the flagella gene cluster of two pathovars of
19 whole genome sequenced *P. syringae*, *P. s. pv. tomato* DC3000 and *pv. phaseolicola* 1448A.
20 The homologues from both of these two pathovars display 68% identity to *orfA* of PAK at the
21 amino acid level. *OrfA* possesses homology to the *vioA* gene which is responsible for the
22 synthesis of viosamine, i.e., Qui4N (1). Qui4N is one of the PAK strain-specific glycan
23 components in *P. aeruginosa* flagellin, but was not detected in the PAO1 strain (25, 34). Thus, it
24 will be interesting to investigate the role of the *orfA* homologue in flagellin glycan synthesis in *P.*
25 *syringae*.

1 Although the structures of Pta6605 and Pgl4 flagellin glycans are similar, differences
2 were observed in the content of D-Rha. Chirality of the Rha residues may be one of the
3 significant determinants of flagellin's elicitor activity. Interaction of *P. syringae* with its host or
4 non-host plant is the most advanced system for elucidating the biological significance of
5 flagellin glycosylation in the interaction of bacteria with eukaryotes. Therefore, our findings are
6 important in defining biological activity, such as bacterial virulence and host specificity, in terms
7 of molecular structure.

10 **Acknowledgements**

11 We thank Dr. A. Collmer (Cornell University, U.S.A.) and Japan Tobacco Inc., Leaf Tobacco
12 Research Laboratory (Tochigi, Japan) for providing *P. syringae* pv. *glycinea* race 4 and pv.
13 *tabaci* 6605 respectively. We are grateful to Dr. E. Minami (National Institute of Agrobiological
14 Sciences, Japan), Dr. M. Ohnishi-Kameyama (National Food Research Institute) and Dr. N.
15 Shibuya (Meiji University, Japan) for general discussions. We also thank Ms. I. Maeda (National
16 Food Research Institute) for technical assistance with the NMR measurements. This work was
17 supported in part by Grants-in-Aid for Scientific Researches (S) (No. 15108001) and (B) (No.
18 18380035) from the Ministry of Education, Culture, Sports, Science and Technology of Japan
19 and the Okayama University COE program "Establishment of Plant Health Science".

21 **References**

- 22 1. **Arora, S. K., M. Banger, S. Lory, and R. Ramphal.** 2001. A genomic island in
23 *Pseudomonas aeruginosa* carries the determinants of flagellin glycosylation. Proc. Natl.
24 Acad. Sci. USA **98**:9342-9347.
- 25 2. **Arora, S. K., A. N. Neely, B. Blair, S. Lory, and R. Ramphal.** 2005. Role of motility and

- 1 flagellin glycosylation in the pathogenesis of *Pseudomonas aeruginosa* burn wound
2 infections. Infect Immun. **73**:4395-4398.
- 3 3. **Benz, I., and M. A. Schmidt.** 2002. Never say never again: protein glycosylation in
4 pathogenic bacteria. Mol. Microbiol. **45**:267-276.
- 5 4. **Che, F. S., Y. Nakajima, N. Tanaka, M. Iwano, T. Yoshida, S. Takayama, I. Kadota,**
6 **and A. Isogai.** 2000. Flagellin from an incompatible strain of *Pseudomonas avenae* induces
7 a resistance response in cultured rice cells. J. Biol. Chem. **275**:32347-32356.
- 8 5. **Chinchilla, D, Z. Bauer, M. Regenass, T. Boller, and G. Felix.** 2006. The Arabidopsis
9 receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant
10 Cell **18**:465-476.
- 11 6. **Daubenspeck, J. M., H. Zeng, P. Chen, S. Dong, C. T. Steichen, N. R. Krishna, D. G.**
12 **Pritchard, and C. L. Thurnbough, Jr.** 2004. Novel Oligosaccharide Side Chains of the
13 Collagen-like Region of BclA, the Major Glycoprotein of the *Bacillus anthracis* Exosporium.
14 J. Biol. Chem. **279**:30945-30953.
- 15 7. **Felix, G., J. D. Duran, S. Volko, and T. Boller.** 1999. Plants have a sensitive perception
16 system for the most conserved domain of bacterial flagellin. Plant J. **18**:265-276.
- 17 8. **Gerwig G. J., J. P. Kamerling, and J. F. Vliegenthart.** 1979. Determination of the
18 absolute configuration of mono-saccharides in complex carbohydrates by capillary G.L.C.
19 Carbohydr. Res. **77**:1-7.
- 20 9. **Gomez-Gomez, L., and T. Boller.** 2000. FLS2: an LRR receptor-like kinase involved in the
21 perception of the bacterial elicitor flagellin in *Arabidopsis*. Mol. Cell. **5**:1003–1011.
- 22 10. **Guerry, P., P. Doig, R. A. Alm, D. H. Burr, N. Kinsella, and T. J. Trust.** 1996.
23 Identification and characterization of genes required for post-translational modification of
24 *Campylobacter coli* VC167 flagellin. Mol. Microbiol. **19**:369-378.
- 25 11. **Guerry, P., C. P. Ewing, M. Schirm, M. Lorenzo, J. Kelly, D. Pattarini, G. Majam, P.**

- 1 **Thibault, and S. M. Logan.** 2006. Changes in flagellin glycosylation affect *Campylobacter*
2 autoagglutination and virulence. *Mol. Microbiol.* **60**:299-311.
- 3 12. **Hayashi, F., K. D. Smith, A. Ozinsky, T. R. Hawn, E. C. Yi, D. R. Goodlett, J. K. Eng, S.**
4 **Akira, D. M. Underhill, and A. Aderem.** 2001. The innate immune response to bacterial
5 flagellin is mediated by Toll-like receptor 5. *Nature* **410**:1099-1103.
- 6 13. **Ishiga, Y., K. Takeuchi, F. Taguchi, Y. Inagaki, K. Toyoda, T. Shiraishi, and Y.**
7 **Ichinose.** 2005. Defense responses of *Arabidopsis thaliana* inoculated with *Pseudomonas*
8 *syringae* pv. *tabaci* wild type and defective mutants for flagellin (Δ *fliC*) and
9 flagellin-glycosylation (Δ *orf1*). *J. Gen. Plant Pathol.* **71**:302-307.
- 10 14. **Jones, D. A., and D. Takemoto.** 2004. Plant innate immunity – direct and indirect
11 recognition of general and specific pathogen-associated molecules. *Curr. Opin. Immunol.*
12 **16**:48-62.
- 13 15. **Kenne, L., B. Lindberg, P. Unger, B. Gustafsson, and T. Holme.** 1982. Structural studies
14 of the *Vibrio cholerae* O-antigen. *Carbohydr. Res.* **100**:341-349.
- 15 16. **Knirel, Y. A.** 1990. Polysaccharide antigens of *Pseudomonas aeruginosa*. *Crit. Rev.*
16 *Microbiol.* **17**:273-304.
- 17 17. **Knirel, Y. A., and G. M. Zdorovenko.** 1997. Structures of O-polysaccharide chains of
18 lipopolysaccharides as the basis for classification of *Pseudomonas syringae* and related
19 strains. p. 475–480. *In* K. Rudolph, T. J. Burr, J. W. Mansfield, D. Stead, A. Vivian, and J.
20 von Kietzell (ed.), *Pseudomonas Syringae Pathovars and Related Pathogens*, Kluwer
21 Academic Publishers, Dordrecht, Boston, London.
- 22 18. **Lipkind, G. M., A. S. Shashkov, Y. A. Knirel, E. V. Vinogradov, and N. K. Kochetkov.**
23 1988. A computer-assisted structural analysis of regular polysaccharides on the basis of
24 ¹³C-n.m.r. data. *Carbohydr. Res.* **175**:59-75.
- 25 19. **Logan, S. M.** 2006. Flagellar glycosylation – a new component of the morility repertoire?

- 1 Microbiol. **156**:1249-1262.
- 2 20. **McNeil, M., and P. Albersheim.** 1977. Chemical ionization mass spectrometry of
3 methylated hexitol acetates. Carbohydr. Res. **56**:239-248.
- 4 21. **Miyake, K.** 2004. Innate recognition of lipopolysaccharide by Toll-like receptor 4-MD-2.
5 Trends Microbiol. **12**:186-192.
- 6 22. **Roepstorff, P., and J. Fohlman.** 1984. Proposal for a common nomenclature for sequence
7 ions in mass spectra of peptides. Biomed. Mass Spectrom. **11**:601.
- 8 23. **Scafer, A., A. Tauch, W. Jager, J. Kalinowski, G. Thierbach, and A. Puhler.** 1994.
9 Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids
10 pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium*
11 *glutamicum*. Gene **145**:69-73.
- 12 24. **Schirm, M., E. C. Soo, A. J. Aubry, J. Austin, P. Thibault, and S. M. Logan.** 2003.
13 Structural, genetic and functional characterization of the flagellin glycosylation process in
14 *Helicobacter pylori*. Mol. Microbiol. **48**:1579-1592.
- 15 25. **Schirm, M., S. K. Arora, A. Verma, E. Vinogradov, P. Thibault, R. Ramphal, and S. M.**
16 **Logan.** 2004. Structural and genetic characterization of glycosylation of type a flagellin in
17 *Pseudomonas aeruginosa*. J. Bacteriol. **186**:2523-31.
- 18 26. **Shimizu, R., F. Taguchi, M. Marutani, T. Mukaihara, Y. Inagaki, K. Toyoda, T.**
19 **Shiraishi, and Y. Ichinose.** 2003. The Δ *fliD* mutant of *Pseudomonas syringae* pv. *tabaci*,
20 which secretes flagellin monomers, induces a strong hypersensitive reaction (HR) in
21 non-host tomato cells. Mol. Genet. Genomics **269**:21-30.
- 22 27. **Smith, K. D., E. Andersen-Nissen, F. Hayashi, K. Strobe, M. A. Bergman, S. L. R.**
23 **Barrett, B. T. Cookson, and A. Aderem.** 2003. Toll-like receptor 5 recognizes a conserved
24 site on flagellin required for protofilament formation and bacterial motility. Nat. Immunol.
25 **4**:1247-1253.

- 1 28. **Taguchi, F., R. Shimizu, R. Nakajima, K. Toyoda, T. Shiraishi, and Y. Ichinose.** 2003a.
2 Differential effects of flagellins from *Pseudomonas syringae* pv. *tabaci*, *tomato* and *glycinea*
3 on plant defense response. *Plant Physiol. Biochem.* **41**:165-174.
- 4 29. **Taguchi, F., R. Shimizu, Y. Inagaki, K. Toyoda, T. Shiraishi, and Y. Ichinose.** 2003b
5 Post-translational modification of flagellin determines the specificity of HR induction. *Plant*
6 *Cell Physiol.* **44**:342-349.
- 7 30. **Taguchi, F., K. Takeuchi, E. Katoh, K. Murata, T. Suzuki, M. Marutani, T. Kawasaki,**
8 **M. Eguchi, S. Katoh, H. Kaku, C. Yasuda, Y. Inagaki, K. Toyoda, T. Shiraishi, and Y.**
9 **Ichinose.** 2006. Identification of glycosylation genes and glycosylated amino acids of
10 flagellin in *Pseudomonas syringae* pv. *tabaci*. *Cell. Microbiol.* **8**:923-938.
- 11 31. **Takeuchi, K., F. Taguchi, Y. Inagaki, K. Toyoda, T. Shiraishi, and Y. Ichinose.** 2003.
12 Flagellin glycosylation island in *Pseudomonas syringae* pv. *glycinea* and its role in host
13 specificity. *J. Bacteriol.* **185**:6658-6665.
- 14 32. **Thibault, P., S. M. Logan, J. F. Kelly, J. R. Brisson, C. P. Ewing, T. J. Trust, and P.**
15 **Guerry.** 2001. Identification of the carbohydrate moieties and glycosylation motifs in
16 *Campylobacter jejuni* flagellin. *J. Biol. Chem.* **276**:34862-34870.
- 17 33. **Verma, A., S. K. Arora, S. K. Kuravi, and R. Ramphal.** 2005. Roles of specific amino
18 acids in the N terminus of *Pseudomonas aeruginosa* flagellin and of flagellin glycosylation
19 in the innate immune response. *Infect. Immun.* **73**:8237-8246.
- 20 34. **Verma, A., M. Schirm, S. K. Arora, P. Thibault, S. M. Logan, and R. Ramphal.** 2006.
21 Glycosylation of b-type flagellin of *Pseudomonas aeruginosa*: Structural and Genetic Basis.
22 *J. Bacteriol.* **188**:4395-4403.
- 23 35. **Vinogradov, E., and M. B. Perry.** 2000. Structural analysis of the core region of
24 lipopolysaccharides from *Proteus mirabilis* serotypes O6, O48 and O57. *Eur. J. Biochem.*
25 **267**:2439-2446.

- 1 36. Yasuno, S., K. Kokubo, and M. Kamei. 1999. New method for determining the sugar
2 composition of glycoproteins, glycolipids, and oligosaccharides by high-performance liquid
3 chromatography. *Biosci. Biotechnol. Biochem.* **63**:1353-1359.
- 4 37. York, W. S., A. G. Darvill, M. McNeil, T. T. Stevenson, and P. Albersheim. 1985.
5 Isolation and characterization of plant cell walls and cell wall components. *Methods*
6 *Enzymol.* **118**: 3-40.
- 7 38. Zipfel, C. and G. Felix. 2005. Plants and animals: a different taste for microbes? *Curr. Opin.*
8 *Plant Biol.* **8**:353-360.

9

10

11 **Figure Legends**

12 **FIG. 1.** MALDI-TOF MS analysis of intact flagellin and the peptide N136-K255 from Pta6605
13 and Pgl4. (A) Intact flagellin from the wild-type strain of Pta6605. (B) The peptide N136-K255
14 from the wild-type strain of Pta6605. (C) Intact flagellin from the wild-type strain of Pgl4. (D)
15 The peptide N136-K255 from the wild-type strain of Pgl4.

16

17 **FIG. 2.** Determination of D-Rha/L-Rha ratios in intact flagellin proteins. (A) GC pattern of the
18 trimethylsilylated (*S*)-2-butyl rhamnosides obtained from Pta6605 flagellin. (B) GC pattern of
19 the trimethylsilylated (*S*)-2-butyl rhamnosides obtained from Pgl4 flagellin. (C) GC-MS
20 fragmentation patterns from a GC peak with a retention time corresponding to that of authentic
21 trimethylsilylated (*R*)-2-butyl L-rhamnose in Pgl4 flagellin. The inset in (C) shows the structure
22 of rhamnose derivative and the expected primary fragment ions.

23

24 **FIG. 3.** MALDI-QIT-TOF MS/MS spectrum of the glycopeptide D200-A211 from Pta6605
25 flagellin with the respective fragmentation scheme. The glycopeptide D200-A211 showing

1 $[M+H]^+$ at m/z 1814 (*inset*) corresponds to a glycopeptide with the sequence
 2 $^{200}\text{DSALQTINSTRA}^{211}$ in which S201 is modified with a 538 Da moiety. The MS/MS
 3 experiment gave product ions at m/z 1796 ($[M+H-H_2O]^+$), 1699 ($[M-\text{Asp}+H]^+$), 1569
 4 ($[M-246+H]^+$), 1423 ($[M-246-\text{Rha}+H]^+$), 1308 ($[M-246-\text{Rha}-\text{Asp}+H]^+$), 1277
 5 ($[M-246-2\text{Rha}+H]^+$) and 1162 ($[M-246-2\text{Rha}-\text{Asp}+H]^+$). X in the fragmentation scheme stands
 6 for a substructure of 246 Da. This substructure was assigned to
 7 Qui4N(3-hydroxy-1-oxobutyl)2Me by subsequent NMR experiments.

8
 9 **FIG. 4.** ESI-Q-TOF mass spectra of the glycopeptide D200-A211 from Pta6605 flagellin. (A-D)
 10 MS/MS spectra of ions observed at m/z 907.5 ($[M+2H]^{2+}$), 784.9 ($[M-246+2H]^{2+}$), 711.9
 11 ($[M-246-\text{Rha}+2H]^{2+}$) and 638.8 ($[M-246-2\text{Rha}+2H]^{2+}$), respectively. The observed peak at m/z
 12 605.3 corresponds to $[M+3H]^{3+}$. The *b*- and *y*-series ions, shown both in the sequence of this
 13 peptide and in the MS/MS spectra, originated from the N- and C-terminus (Roepstorff and
 14 Fohlman, 1984), respectively. A substructure of 246 Da was assigned to
 15 Qui4N(3-hydroxy-1-oxobutyl)2Me by subsequent NMR experiments.

16
 17 **FIG. 5.** Chemical structure and selected NMR correlations (HMBC and NOESY) of glycan
 18 attached to S201 of the peptide D200-A211 from Pta6605 and Pgl4.

19
 20 **FIG. 6.** Two dimensional NMR spectra of
 21 β -D-Quip4N(3-hydroxyburyryl)2Me-(1 \rightarrow 3)- α -L-Rhap-(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow O^{S201})-peptide
 22 D200-A211. A, NOESY. B, HMBC.

23

TABLE 1. Bacterial strains used in this study

Bacterial strain	Relevant characteristics ^a	Reference or source
<i>E. coli</i>		
DH5 α	<i>FλØ80dlacZΔM15 Δ (lacZYA-argF)U169 recA1 endA1 hsdR17(rK⁻ mK⁺) supE44 thi-1 gyrA relA1</i>	Takara, Kyoto, Japan
S17-1	<i>thi pro hsdR⁻ hsdM⁺ recA</i> [<i>chr::RP4-2-Tc::Mu-Km::Tn7</i>]	23
<i>P. syringae</i> pv. <i>glycinea</i>		
race 4	Wild type	Collmer, A.
race 4-d1	Isolate race 4 Δ <i>orf1</i>	31
race 4-d2	Isolate race 4 Δ <i>orf2</i>	31
race 4-d3	Isolate race 4 Δ <i>orf3</i>	31
race 4-dfliC	Isolate race 4 Δ <i>fliC</i>	This study
race 4-S143A	Isolate race 4 S143A	This study
race 4-S164A	Isolate race 4 S164A	This study
race 4-S176A	Isolate race 4 S176A	This study
race 4-S183A	Isolate race 4 S183A	This study
race 4-S193A	Isolate race 4 S193A	This study
race 4-S201A	Isolate race 4 S201A	This study
race 4-6 S/A	Isolate race 4 S143A, S164A, S176A, S183A, S193A, S201A	This study
<i>P. syringae</i> pv. <i>tabaci</i>		
Isolate 6605	Wild type	30
Plasmids		
pMC	1.8 kb chimeric PCR product deleting <i>fliC</i> cloned into pK18mobSacB at EcoRI site, Km ^r	This study, 26
pK18 <i>mobsacB</i>	Small mobilizable vector, Km ^r , sucrose sensitive (<i>sacB</i>)	23

^a Km^r = kanamycin resistance

TABLE 2. Mass values of intact flagellin and peptide fragment (N¹³⁶-K²⁵⁵)

<i>P. s. pv. glycinea</i>	[M+H] ⁺		Δ (*2-*3)
	observed* ²	calculated* ³	
intact flagellin (A ² -Q ²⁸²)* ¹			
WT	32380, 32515, 32668	29148	3232, 3367, 3520
Δorf1	29154	29148	6
peptide fragment (N ¹³⁶ -K ²⁵⁵)			
WT	15296, 15444, 15591	12074	3222, 3370, 3517
S143A	14742, 14890, 15036	12058	2684, 2832, 2978
S164A	14746, 14894, 15040	12058	2688, 2836, 2982
S176A	14748, 14893, 15038	12058	2690, 2835, 2980
S183A	14747, 14894, 15044	12058	2689, 2836, 2986
S193A	14751, 14899, 15045	12058	2693, 2841, 2987
S201A	14747, 14894, 15038	12058	2689, 2836, 2980
6 S/A	11976	11978	-2
Δorf1	12072	12074	-2
Δorf2* ⁴	13292-14858	12074	1218-2784
Δorf3	15298, 15446, 15591	12074	3224, 3372, 3517

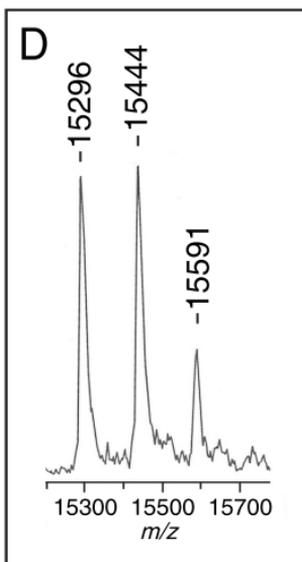
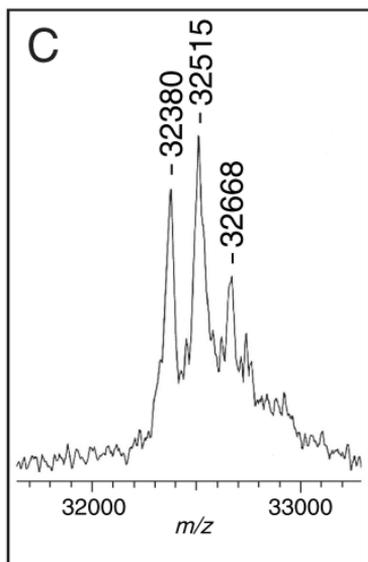
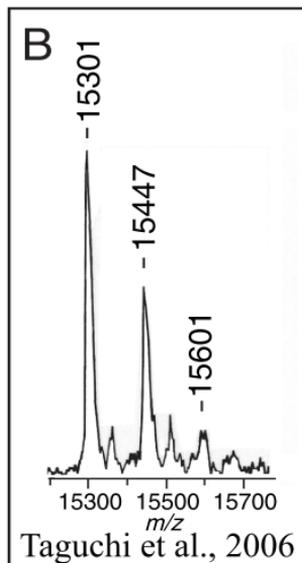
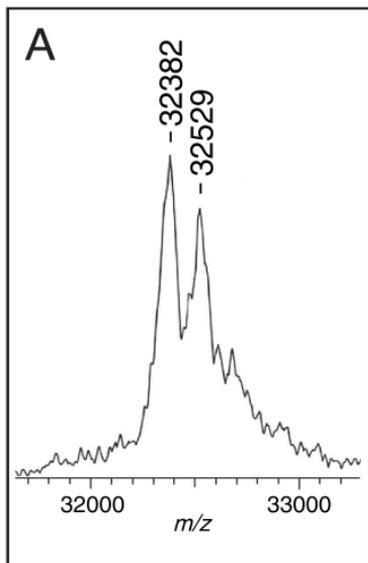
*¹: N-terminal methionine of flagellin is subject to posttranslational cleavage (data not shown).

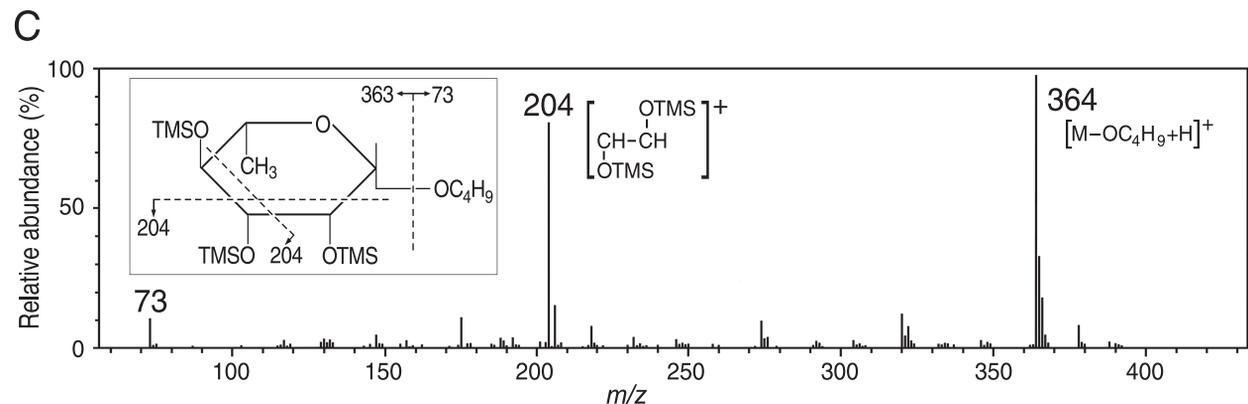
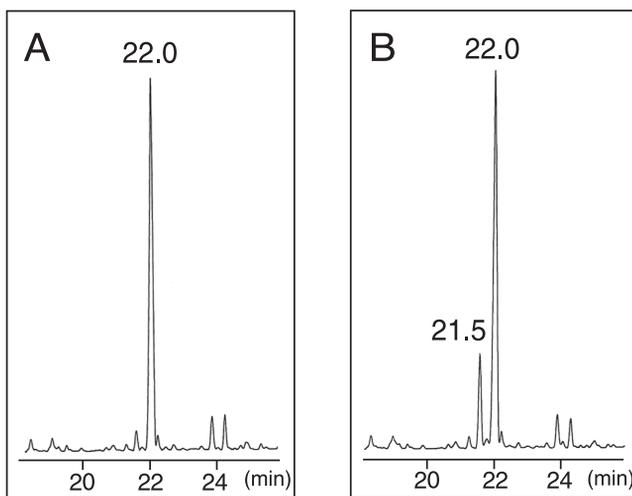
*²: Major peaks in MALDI-TOF mass spectra. *³: Calculated value by deduced amino acid sequence. *⁴: Peaks were observed as broad peaks by heterogeneity.

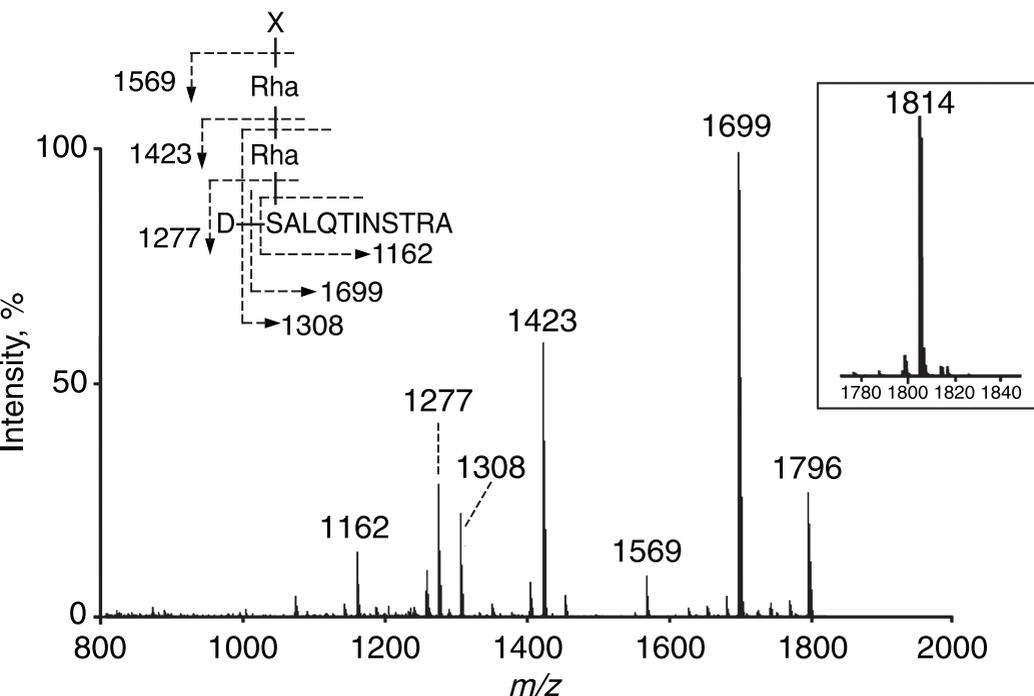
Table 3. Assignment of NMR signals^a of glycan on S201 of the peptide D200-A211. Numbers of rhamnosyl residues are based on the order from the reducing end.

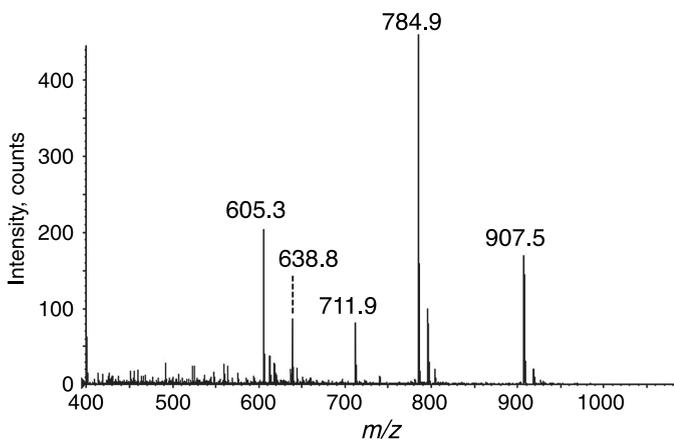
residue	¹ H			¹³ C	
	position	δ (ppm)	multiplicity ^b	<i>J</i> (Hz)	δ (ppm)
Ser201					
	amide carbonyl				172.5 ^c
	α	4.599	dd	4.3, 3.7	55.53
	β	4.055	dd	10.0, 4.3	67.67
		3.732	dd	10.0, 3.7	
α-L-Rhap 1					
	1	4.888	br s (<i>W</i> _{1/2} = 4.5Hz)		99.81
	2	3.977	br s (<i>W</i> _{1/2} = 7.2Hz)		79.90
	3	3.780	dd	9.7, 3.1	71.66
	4	3.440	dd	9.7, 9.6	73.71
	5	3.541	dq	9.6, 6.0	70.84
	6	1.273	d	6.0	18.51
α-L-Rhap 2					
	1	4.979	br s (<i>W</i> _{1/2} = 4.5Hz)		103.55
	2	4.264	br dd	3.1, 1.6	71.38
	3	3.900	dd	9.8, 3.1	81.13
	4	3.591	dd	9.8, 9.8	72.83
	5	3.756	dq	9.8, 6.1	71.04
	6	1.271	d	6.1	18.39
β-D-Quip 4N (4-amino-4,6-dideoxy-β-D-glucopyranosyl)					
	1	4.726	d	8.0	105.20
	2	3.126	dd	9.2, 8.0	84.94
	3	3.517	dd	10.2, 9.2	74.50
	4	3.619	dd	10.2, 10.0	58.26
	5	3.535	dq	10.0, 6.0	72.56
	6	1.199	d	6.0	18.64
	2- <i>O</i> -Me	3.620	s		61.77
	<i>N</i> -(3-hydroxy-1-oxobutyl)				
	1'				176.2 ^c
	2'	2.402	dd	14.1, 7.9	46.87
		2.422	dd	14.1, 5.5	
	3'	4.188	ddq	7.9, 5.5, 6.3	66.73
	4'	1.218	d	6.3	23.77

^a 800.33 MHz for ¹H and 125.76 MHz for ¹³C, in D₂O at 298K^b s, singlet; d, doublet; t, triplet; q, quartet; br, broad; *W*_{1/2}, width at half height^c read from a position of an HMBC cross-peak

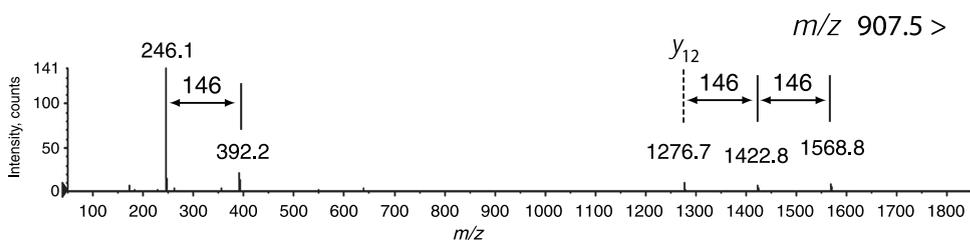




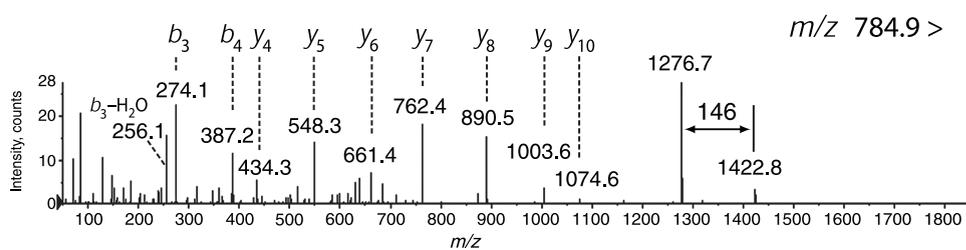




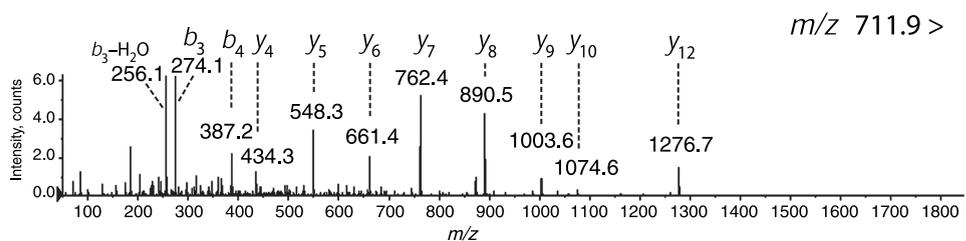
A



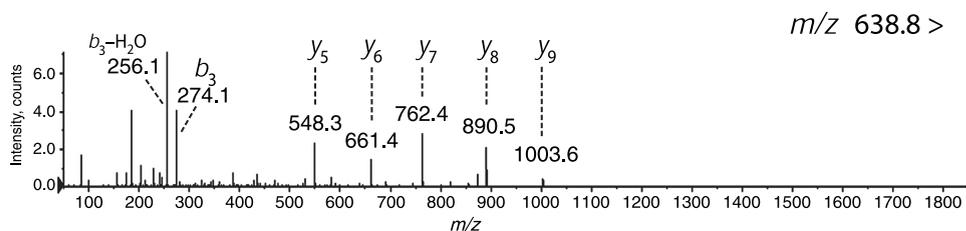
B



C



D



β -D-Quip4N(3-hydroxy-1-oxobutyl)2Me-(1 \rightarrow 3)- α -L-Rhap-
(1 \rightarrow 2)- α -L-Rhap-(1 \rightarrow O^{S201})-peptide D200-A211

