Full structural characterization of *Shigella flexneri* M90T serotype 5 wild-type R-LPS and its ΔgalU mutant: glycine residue location in the inner core of the lipopolysaccharide

**Introduction**

Shigellosis is an acute rectocolitis caused by enteroinvasive Gram-negative bacteria belonging to *Shigella* genus. Infected individuals may become convalescent carriers, acting as reservoirs of this organism. Mucosally invasive *Shigella*, which could cause dysentery, are less amenable to the effects of the oral rehydration than noninvasive pathogens that cause watery diarrhea, such as *Vibrio cholerae* and *Escherichia coli* (Levine and Cutts 2007). The genus *Shigella* is divided into four serogroups each of which consists of a different serotypes: *Shigella dysenteriae* (group A), which have 15 serotypes; *Shigella flexneri* (group B), which has 14 serotypes; *Shigella boydii* (group C), which has 20 serotypes; and *Shigella sonnei* (group D), which has a single serotype. In particular, *S. flexneri* is known as the major causative agent of the endemic form of shigellosis or bacillary dysentery responsible for approximately one million fatalities annually among infants, mostly in developed countries (Sansonetti 2006a). The disease is characterized by the entry of the bacterium into colonic epithelial cells following to intracellular multiplication and intercellular spreading which allows bacterium to infect neighboring cells (Sansonetti 2006b; Phalipon and Sansonetti 2007).

Lipopolysaccharides (LPSs) are an important virulence factor that play a key role in the pathogenesis and in the toxic manifestation of Gram-negative infection (Alexander and Rietschel 2001; Raetz and Whitfield 2002) promoting the activation of the immune system. LPSs, also called endotoxins, are the main components of the outer membrane of almost all Gram-negative bacteria. They are amphiphilic macromolecules typically composed in their smooth form (S-LPSs) of three structurally, chemically, and biogenetically distinct regions: the O-antigenic polysaccharide (O-specific chain); the oligosaccharide core region composed of up to 15 monosaccharides (Holst 1999) divided into a relatively conserved inner core and a distal and slightly variable outer core region and a lipophilic portion, termed lipid A, which anchors the LPS molecule to the bacterial outer membrane. LPSs not containing O-chain are termed rough LPSs (R-LPSs) or lipooligosaccharides (LOSs). LOSs have been found either in wild-type strains or in mutant strains bearing mutations in the genes encoding enzymes of the biosynthesis and/or the transfer of the O-specific polysaccharide. The only carbohydrate residue present in all LPSs is the Kdo residue (3-deoxy-D-manno-oct-2-ulosonic acid) that covalently links both the core oligosaccharide and lipid A whereas the major part of core oligosaccharides also bears the L-glycero-D-manno-heptose residue (1→6)-Hep residue in the inner part. The outer core region of LPSs is more variable and it is usually composed of hexoses (Holst 1999; Raetz and Whitfield 2002).

Antonio Molinaro1,2, Alba Silipo2, Cristina De Castro2, Luisa Sturiale3, Giulia Nigro4, Domenico Garozzo3, Maria Lina Bernardini4,5, Rosa Lanzetta5, and Michelangelo Parrilli2

1Dipartimento di Chimica Organica e Biochimica, Università degli Studi di Napoli Federico II, Via Cintia 4, I-80126 Napoli; 2CNR Istituto per la Chimica e la Tecnologia dei Materiali Polimerici, Viale A. Doria 6, 95125 Catania; 3Dipartimento di Chimica Organica e Biochimica, Università degli Studi di Roma, Via dei Sardi 70, 00185 Roma; and 4Istituto Pasteur-Fondazione Cenci Bolognetti, Sapienza-Università di Roma, Piazzale Aldo Moro 5, 00185 Roma, Italy

Received on December 4, 2007; revised on December 28, 2007; accepted on December 29, 2007

*Shigella flexneri* is a Gram-negative bacterium responsible for serious enteric infections that occur mainly in the terminal ileum and colon. High interest in *Shigella*, as a human pathogen, is driven by its antibiotic resistance and the necessity to develop a vaccine against its infections. Vaccines of the last generation use carbohydrate moieties of the lipopolysaccharide as probable candidates. For this reason, the primary structure of the core oligosaccharide from the R-LPS produced by *S. flexneri* M90T serotype 5 using chemical analysis, nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MALDI), is herein reported. This is the first time that the core oligosaccharide primary structure by *S. flexneri* M90T is established in an unambiguous multidisciplinary approach. Chemical and spectroscopical investigation of the de-acetylated LPS showed that the inner core structure is characterized by a 1→6-Hep-(1→7)-1→8-Hep-(1→3)-1→8-Hep-(1→5)-Kdo sequence that is the common structural theme identified in *Enterobacteriaceae*. In particular, in *S. flexneri* M90T serotype 5 LPS, a glucosamine residue is additionally sitting at O-7 of the last heptose whereas the outer core is characterized by glucose and galactose residues. Also, in order to exactly define the position of glycine that is an integral constituent of the core region of the LPS, we created a *S. flexneri* M90T ΔgalU mutant and studied its LOS. In this way it was possible to establish that glycine is sitting at O-6 of the second heptose in the inner core.

**Keywords:** Glycine/lipooligosaccharide/mass spectrometry/NMR spectroscopy/*Shigella flexneri"
The R-LPS of *S. flexneri* serotype 5 M90T was isolated in a very small amount from the LPS fraction by gel permeation chromatography. In the present paper, we have carried out the structural determination of the carbohydrate and noncarbohydrate components of a core region from the R-LPS of *S. flexneri* serotype 5 M90T by compositional analysis, 2D NMR spectroscopy and matrix-assisted laser desorption/ionization mass spectrometry. A previous paper of the early 1980s (Katzenellenbogen and Romanowska 1980) already described the structure of the core-oligosaccharide region of *S. flexneri* serotype 6 obtained by compositional and methylation analyses, Smith degradation and enzymatic modifications that we now completely confirm and complete by state-of-the-art techniques. Moreover, we have also addressed the issue of glycine identification and location in the inner core of *Shigella* LPS as, in fact, glycine (Gly) has been reported as a constitutive component of the LPS core from *Shigella* (even though with no precise information). Since both the small amount of wild-type R-LPS and that of Gly in it impaired any hypothesis for its position we have created and analyzed an M90T ΔgalU mutant. The mutation galU gene blocks the biosynthesis of UDP glucose from glucose-1-P, a precursor needed for the glucosyl residues into its LPS (Maurelli and Sansonetti 1988). The virulence of *S. flexneri* 2 ΔgalU mutants has been extensively analyzed in vitro as well as in vivo (Sandlin et al. 1995; Köhler et al. 2002). This mutation in the outer core region of *S. flexneri* M90T LPS returned a truncated lipooligosaccharide in which the nonstoichiometric presence of glycine was detectable at position O-6 of the second heptose residue of the inner core.

Furthermore, while writing this paper another one was published dealing with the finding and positioning of the Gly residue in the LOS of *Campilobacter jejuni* and interestingly, it was found to be placed on the same second distal heptose residue in every LPS containing glycine from that bacterium. In that case, however, it was not possible to get further information on the precise position of the amino acid (Dzieciatkowska et al. 2007), i.e., which heptose carbon Gly residue is sitting on.

**Results**

**Cell culture, extraction, and compositional analysis of the lipooligosaccharide**

The dried cells of *S. flexneri* M90T were sequentially extracted by a phenol–chloroform–light petroleum extraction method (Galanos et al. 1969) and by a hot phenol–water method (Westphal and Jahn 1965). The LPS material was found exclusively in the water phase of the second extraction as shown by sodium deoxycholate–polyacrylamide gel electrophoresis (SDS–PAGE), which revealed a typical ladder appearance located in the upper part of the gel, in accordance with the high molecular mass of the LPS form (S-LPS). The S-LPS was further purified from other cell components by enzymatic hydrolysis with DNase, RNase, and proteinase. The sample was dialyzed and chromatographed on a Sephacryl HR-500 that yielded two peaks; one of the two was in very low amount but contained lower molecular mass species, i.e., LOS, as detected by SDS–PAGE, and was analyzed for the core oligosaccharide primary structure.

Combining the information derived from monosaccharide composition, absolute configuration, and methylation analyses, we found that the LOS fraction contained 7-substituted, 3,7-di-substituted, and 3-substituted deoxy-glucurono-manno-heptose (Hep); terminal and 6-substituted 2-amino-2-deoxy-d-glucose (d-GlcN); terminal and 2-substituted d-galactose (Gal); terminal, 3-substituted and 2,3-di-substituted d-glucose (Glc); 4,5-substituted and terminal Kdo. All monosaccharides were in D configuration and heptose residues in L,D configuration. Fatty acids analysis showed the presence of C14:0 (3-0H) either in amide or in ester linkages while C12:0 and C14:0 exclusively found as ester linked, thereby confirming the structure of LPS lipid A from *Shigella* genus. A MALDI MS spectrum confirmed this hypothesis containing the main molecular ion related to typical exa-acylated lipid A from *Shigella* (m/z 1797.1). Strong alkaline treatment led to the fully de-acylated product (OS) that was further purified by HPLC yielding an oligosaccharide that was elucidated by NMR.

**NMR and MALDI analysis of OS**

The 1H NMR spectra of oligosaccharide are shown in Figure 1. A combination of homo- and heteronuclear 2D NMR experiments (double quantum-filtered phase-sensitive correlation spectroscopy (DQF-COSY), total correlation spectroscopy (TOCSY), nuclear Overhauser enhancement spectroscopy (NOESY), heteronuclear single quantum coherence (1H, 13C HSQC), and heteronuclear multiple bond correlation (1H, 13C HMBC)) were executed in order to assign all the spin systems and the monosaccharide sequence. In the anomic region of the 1H NMR spectrum, 11 signals corresponding to 11 different spin systems were present. Spin systems were denoted with a capital letter according to their decreasing chemical shift values (A–M, Table I), and, in addition, the signals at 1.78/2.16 and 1.97/2.13 ppm were identified as the H-3 methyl protons of two Kdo residues (N and O). Anomeric configurations were assigned on the basis of 1JCH values measured by a coupled HSQC experiment whereas relative configurations were established on the basis of the chemical shifts and the 2JHH values obtained from the DQF-COSY spectrum. All monosaccharide residues were present as pyranose rings, according either to 13C chemical shift values or to the occurrence of long-range correlation between C-1/H-1 and H-5/C-5 in the 1H, 13C HMBC spectrum (for Kdo residues between C-2 and H-6). By DQF-COSY, TOCSY, and ROESY (rotating frame overhauser enhancement spectroscopy) spectra, the full assignment of proton resonances of each spin system was possible and afterward, 13C chemical shifts were assigned from observed correlation in the 1H, 13C HSQC spectrum. Even though the 1H NMR spectrum looked like a heterogeneous mixture of oligosaccharides, through 2D NMR we were able to assign each spin system and to ascertain the origin of such heterogeneity as nonstoichiometric phosphate substitution.

Spin systems A and C (5.91/96.1 and 5.57/92.7 ppm, respectively) possessed a galacto-configuration as indicated by the typical small 3JHH < 2 Hz as read by the DQF-COSY spectrum. From H-2 of each residue it was possible to identify in the TOCSY spectrum all

Core oligosaccharide structure from *Shigella flexneri*
the other resonances of ring protons up to H-7 proton signals, leading to the identification of these spin systems as an α-heptose ($J_{H-1,C-1} = 174$ Hz). Spin systems G, I, and L were identified as glucose residues, as indicated by their large ring $^3J_{H1,H2}$ coupling constant (above 10 Hz). In particular, as for residues I and L, the $^3J_{H1,H2}$ value (7 Hz) together with the strong intra-residual NOE connectivity from H-1 to H-3 and to H-5 were diagnostic of a β-configuration whereas, for residue G, the intra-residue NOE contact of H-1 only to H-2 and the $^3J_{H1,H2}$ coupling constant (3 Hz) were indicative of an α-anomeric configuration.

Finally, spin systems B, F, and M were attributed to GlcN units because of all large ring $^3J_{H1,H2}$ coupling constants and the correlation of their H-2 protons at 3.41, 3.34, and 3.07 ppm, respectively, to the nitrogen-bearing carbons at 54.4, 55.3, and 56.7. Spin system B was identified as the α-GlcN I of the lipid A skeleton because of its chemical shift and multiplicity of

Table I. $^1$H and $^{13}$C NMR chemical shifts (ppm) of the oligosaccharide OS derived from strong alkaline treatment of the LOS from wild-type S. flexneri M90T serotype 5. $^{31}$P NMR resonances for phosphate substitution at O-1 B, O-4 D, and O-4 M are at 2.11, 3.12, and 2.60 ppm, respectively.

<table>
<thead>
<tr>
<th>Residue</th>
<th>H-1/C-1</th>
<th>H-2/C-2</th>
<th>H-3/C-3</th>
<th>H-4/C-4</th>
<th>H-5/C-5</th>
<th>H-6/C-6</th>
<th>H-7/C-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-Gal</td>
<td>5.91</td>
<td>4.22</td>
<td>4.31</td>
<td>4.41</td>
<td>4.15</td>
<td>3.83/3.90</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>96.0</td>
<td>69.1</td>
<td>72.2</td>
<td>71.1</td>
<td>7.15</td>
<td>61.3</td>
<td></td>
</tr>
<tr>
<td>6-GlcN</td>
<td>5.65</td>
<td>3.41</td>
<td>3.62</td>
<td>3.43</td>
<td>3.91</td>
<td>4.15/3.74</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>91.1</td>
<td>54.4</td>
<td>71.5</td>
<td>69.8</td>
<td>71.5</td>
<td>71.5</td>
<td></td>
</tr>
<tr>
<td>2-Gal</td>
<td>5.57</td>
<td>3.76</td>
<td>3.78</td>
<td>3.90</td>
<td>4.08</td>
<td>3.73/3.90</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>92.7</td>
<td>76.6</td>
<td>71.9</td>
<td>70.5</td>
<td>72.2</td>
<td>61.3</td>
<td></td>
</tr>
<tr>
<td>7-Hep</td>
<td>5.30</td>
<td>4.08</td>
<td>4.00</td>
<td>3.98</td>
<td>4.05</td>
<td>4.15</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>100.0</td>
<td>72.9</td>
<td>72.1</td>
<td>70.8</td>
<td>69.9</td>
<td>70.0</td>
<td></td>
</tr>
<tr>
<td>3-Hep4P</td>
<td>5.28</td>
<td>4.09</td>
<td>4.12</td>
<td>4.40</td>
<td>4.12</td>
<td>4.20</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>100.5</td>
<td>72.1</td>
<td>76.9</td>
<td>71.5</td>
<td>69.8</td>
<td>69.7</td>
<td></td>
</tr>
<tr>
<td>t-GlcN</td>
<td>5.23</td>
<td>3.34</td>
<td>3.94</td>
<td>3.49</td>
<td>3.76</td>
<td>3.75/3.92</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>97.3</td>
<td>55.3</td>
<td>72.7</td>
<td>70.4</td>
<td>71.8</td>
<td>61.5</td>
<td></td>
</tr>
<tr>
<td>2.3-Glc</td>
<td>5.21</td>
<td>3.58</td>
<td>3.99</td>
<td>3.60</td>
<td>3.84</td>
<td>3.78/3.92</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>96.4</td>
<td>76.1</td>
<td>76.9</td>
<td>68.7</td>
<td>69.1</td>
<td>61.5</td>
<td></td>
</tr>
<tr>
<td>3,7-Hep</td>
<td>5.12</td>
<td>4.31</td>
<td>4.07</td>
<td>4.41</td>
<td>4.00</td>
<td>4.25</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>102.9</td>
<td>72.4</td>
<td>78.8</td>
<td>71.3</td>
<td>69.8</td>
<td>68.8</td>
<td></td>
</tr>
<tr>
<td>3-β-Glc</td>
<td>4.98</td>
<td>3.48</td>
<td>3.77</td>
<td>3.59</td>
<td>3.57</td>
<td>3.88/3.90</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>102.6</td>
<td>73.2</td>
<td>82.7</td>
<td>69.5</td>
<td>76.4</td>
<td>61.4</td>
<td></td>
</tr>
<tr>
<td>t-β-Glc</td>
<td>4.95</td>
<td>3.43</td>
<td>3.44</td>
<td>3.45</td>
<td>3.43</td>
<td>3.76/3.95</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>102.0</td>
<td>73.7</td>
<td>75.7</td>
<td>69.8</td>
<td>76.5</td>
<td>61.2</td>
<td></td>
</tr>
<tr>
<td>β-GlcN</td>
<td>4.86</td>
<td>3.07</td>
<td>3.85</td>
<td>3.74</td>
<td>3.49</td>
<td>3.69/3.45</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>100.9</td>
<td>56.7</td>
<td>73.3</td>
<td>73.2</td>
<td>77.1</td>
<td>63.9</td>
<td></td>
</tr>
<tr>
<td>H-3/C-3</td>
<td></td>
<td>74.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-4/C-4</td>
<td></td>
<td>75.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-5/C-5</td>
<td></td>
<td>76.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-6/C-6</td>
<td></td>
<td>77.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-7/C-7</td>
<td></td>
<td>78.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-Kdo</td>
<td>1.78/2.16</td>
<td>4.14</td>
<td>4.04</td>
<td>3.75</td>
<td>4.01</td>
<td>3.80/3.70</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>35.1</td>
<td>66.9</td>
<td>67.2</td>
<td>73.0</td>
<td>70.1</td>
<td>63.9</td>
<td></td>
</tr>
<tr>
<td>4,5-Kdo</td>
<td>1.97/2.13</td>
<td>4.16</td>
<td>4.25</td>
<td>3.67</td>
<td>3.98</td>
<td>3.86/3.60</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>34.8</td>
<td>71.8</td>
<td>68.6</td>
<td>73.6</td>
<td>70.3</td>
<td>63.9</td>
<td></td>
</tr>
</tbody>
</table>
the anomic signal (double doublet, $J_{H-1,H-2} = 3.1$ Hz, and $J_{H-1,P} = 7.6$ Hz). The $O$-phosphorylation at the anomic position was confirmed by the presence of a cross peak at 5.65/2.11 ppm in the $^1$H, $^{31}$P HSQC spectrum (Table I). Residue M was attributed to the second monose residue (GlcN II) of the lipid A backbone, as inferred by the value of 8.1 Hz of its $J_{H-1,H-2}$, and as also indicated, in the ROESY spectrum, by the spatial proximity of H-6 of GlcN I with the anomic proton of residue M (GlcN II). In agreement with the identification of this residue as GlcN II of the lipid A backbone, its C-6 chemical shift experienced a mild glycosylation effect, consistent with a substitution position: O-6 of residue GlcN I, whereas residue C was substituted at O-2 by terminal β-galactose A, as indicated by the NOE contact of H-1 and H-2 signals. The sequence of residues and their attachment points was fully confirmed by scalar long-range correlations presented in the $^1$H, $^{13}$C HMBC spectrum.

The $^{31}$P NMR spectrum showed the presence of three monophosphate monoester signals (Table I). The site of substitution was deduced by the $^1$H, $^{31}$P HSQC spectrum, which showed correlations of $^{31}$P signals with H-1 and H-2 (the dipolar coupling between two anomeric protons of two monosaccharides in an oligo/polysaccharide is a typical consequence of a 1→2 glycosidic linkage). Finally, residue C was substituted at O-2 by terminal α-galactose A, given the NOE effect among their H-1 and H-2 signals. The sequence of residues and their attachment points was fully confirmed by scalar long-range correlations presented in the $^1$H, $^{13}$C HMBC spectrum.

Fig. 2. Section of the NOESY spectrum of the oligosaccharide obtained by alkaline treatment. Monosaccharide labels are as indicated in Table I. The relevant inter-residual NOE cross peaks are indicated.

The down-field of carbon resonance identified the glycosylation position: O-6 of residue B, O-2 of C, O-3 of D, O-7 of E, O-2 and O-3 of G, O-3 and O-7 of H, and O-3 of I, whereas residues A, F, and L were nonreducing terminal sugar, in full agreement with the methylation analysis data.

The sequence of residues was inferred by inter-residual dipolar correlation found in the NOESY spectrum (Figure 2). As mentioned above the lipid A carbohydrate backbone was composed of residues B and M. The linkage of heptose D to O-5 of Kdo unit O was deduced by the NOE contacts found between H-1 D and H-5 and H-7 O, and in addition between H-5 D and H-3eq O. Kdo O was further substituted at O-4 by terminal Kdo N as demonstrated by the NOE effect between H-6 N and H-3eq O. These NOE data together with the information on an absolute configuration of heptose D also indicated the D absolute configuration of residues O and N (Bock et al. 1994): in summary, the inner core backbone of Shigella LPS was built up of α-L,D-Hep-(1→5)-[α-D-Kdo-(2→4)]-α-D-Kdo.

Heptose D was substituted at O-3 by heptose H as shown by the NOE effect between anomic signal H-1 H and H-3 D. Residue H was in turn substituted at O-3 and O-7, actually, the NOE contact of its H-3 with H-1 I evidenced that it was glycosylated at O-3 by I residue whereas residue E was sitting at O-7 according to the NOE of its H-7 with H-1 E. Residue E was in turn substituted at O-7 by terminal GlcN F as confirmed by the NOE connectivity between H-1 F and H-7 E. Residue I was also glycosylated by glucose G as indicated by the NOE contact between H-1 G and H-3 I. Unit G was substituted by the terminal β-glucose residue L as indicated by the NOE contact of the H-1 L with H-3 G that, in turn, was also substituted at O-2 by residue C according to the characteristic NOEs of H-1 G and H-2 G with H-1 C (the dipolar coupling between two anomic protons of two monosaccharides in an oligo/polysaccharide is a typical consequence of a 1→2 glycosidic linkage). Finally, residue C was substituted at O-2 by terminal α-galactose A, given the NOE effect among their H-1 and H-2 signals. The sequence of residues and their attachment points was fully confirmed by scalar long-range correlations presented in the $^1$H, $^{13}$C HMBC spectrum.

The $^{31}$P NMR spectrum showed the presence of three monophosphate monoester signals (Table I). The site of substitution was deduced by the $^1$H, $^{31}$P HSQC spectrum, which showed correlations of $^{31}$P signals with H-1 B (GlcN I), H-4 D (3-Hep), and H-4 M (GlcN II) (Table I).

Thus, methylation data and NMR analysis allowed us to establish the following primary structure of the OS oligosaccharide:

\[
\alpha\text{-Gal(1→2)α-Gal(1→2)α-Glc(1→3)β-Glc(1→3)α-Hep(1→3)α-HepP(1→5)α-Kdo(2→4)\text{-Lipid A}} \\
\beta\text{-Glc(1→)} \\
\alpha\text{-Kdo(1→4)} \\
\beta\text{-GlcN(1→7)} \\
\alpha\text{-Hep(1→7)} \\
\beta\text{-GlcN(1→7)} \]
The negative-ion MALDI mass spectrum of the OS obtained from the alkaline treatment is reported in Figure 3. It showed a base peak at \( m/z \) 2567.8 built up of three HexN, three Hep, two Kdo, five hexose, and three phosphate groups that confirmed the above structural hypothesis. At \( m/z \) 2067.3 the B-type ion (Domon and Costello 1988) of the oligosaccharide core arising from the in-source cleavage of the labile glycosidic linkage between Kdo and the \( \beta \)-GlcN II belonging to lipid A was observed (Gibson et al. 1997; Sturiale et al. 2005). Ions corresponding to species differing by phosphate groups were also present.

The LOS molecule, in wild-type LPS blend of \( S. \) flexneri serotype 5 M90T, was present in a very heterogeneous and low amount. In fact, even though we found a detectable (but very low) glycine amount in the chemical analysis of intact LOS, the MALDI MS measurements of an intact LOS molecule (spectrum not shown) was too heterogeneous to appreciate and assign glycine presence. Thus, we created and analyzed a \( galU \) mutant that lacks all the outer core region of LPS in order to detect in a clear-cut manner the presence of Glycine residue in the inner core.

**Construction and analysis of the \( S. \) flexneri 5 M90T \( \Delta galU \) mutant**

In order to obtain a mutant with a deletion in the \( galU \) locus encoding glucose-1-phosphate uridylyltransferase, the \( galU \) gene was amplified from M90T DNA, mutagenized in vitro and reintroduced into the M90T genome. The resulting strain M90T \( \Delta galU \) showed a characteristic “rough” phenotype, i.e., a typical broad diffuse pellet which was formed on gravity sedimentation of an overnight culture in a growth medium left at room temperature for several hours. Moreover, as expected, it failed to agglutinate in typing serum specific for \( S. \) flexneri 5.

The lipooligosaccharide fraction (LOS) was isolated by M90T \( \Delta galU \) by a phenol–chloroform–light petroleum extraction method typical for LOS molecules (Galanos et al. 1969) and, in fact, SDS–PAGE analysis showed a low molecular mass band as expected on the basis of the mutation. The monosaccharide analysis along with the linkage analysis revealed the presence of 6-substituted D-GlcN; terminal and 3-substituted L,D-Hep; 4,5-di-substituted and terminal Kdo. The NMR analysis of the fully de-acylated LOS was very simple and allowed us to establish the presence of the following two oligosaccharide structures in nearly 1:1 ratio; obviously, the strong alkaline treatment removed any ester or amide bond group, and so also the glycine linked residue if present.

The MALDI mass spectrum of the intact molecule acquired in negative polarity (Figure 4A) showed an array of molecular ions due to the considerable heterogeneity of the LOS but in this case glycine was clearly detectable. At a lower mass range, two ions related to the lipid A moiety were identified raising from the known fragmentation between the lipid A moiety and the Kdo-containing oligosaccharide core (Gibson et al. 1997; Sturiale et al. 2005). In particular, the ion peak at \( m/z \) 1797.3 was attributed to the hexa-acylated bis-phosphorylated lipid A species, carrying in ester linkage two C14:O (3-OH) residues and in amide linkages two C14:O (3-OH) as primary fatty acids and one C12:O and one C14:O as secondary fatty acids. The presence of the ion at \( m/z \) 1877.1 revealed a lipid A backbone possessing an additional phosphate group, i.e., a pyrophosphate, which was not possible to place and thus is not discussed here. The core oligosaccharide molecular mass was inferred by the mass differences between the LOS molecular ions and the fragment related to lipid A. In particular, ion peak A at \( m/z \) 2621.1 correlated with lipid A at \( m/z \) 1797.3, whereas molecular ion B at \( m/z \) 2701.0 referred to lipid A at \( m/z \) 1877.1. In both cases mass differences led us to identify a core unit constituted by two Hep and two Kdo residues and these results matched with the information obtained by chemical analysis and NMR data of the fully de-acylated product. Furthermore, two other ion peaks (A’ and B’) were visible in the mass region between 2400 and 3000 Th (Figure 4B) that differed by 57 Th from A and B ions, respectively, and exactly matched with the nonstoichiometric
presence of a glycine residue. A careful evaluation of the mass spectrum allowed us to assign two very minor ion peaks, A'' and B'', owing to the presence of a second glycine, which unfortunately we were not able to locate. Further glycoforms related to B/B' species and including further 2-aminoethyl-phosphate (PEtN, Δm/z = 123) groups were also visible (m/z 2824.0 and 2880.9 respectively, Figure 4), together with ion peaks due to additional phosphate groups.

In order to confirm and define the glycine attachment to the inner core, an aliquot of the sample was hydrolyzed with 1% acetic acid, in this way leaving almost unaffected any ester bound substituent on the carbohydrate chain. The mild acid treatment produced a heterogeneous mixture of core oligosaccharides descending from one basic structure still decorated with nonstoichiometric noncarbohydrate substituents that was subjected to NMR studies (Table II). The assignment of most important signals within the spin systems by a combination of homonuclear and heteronuclear experiments allowed us to identify the glycine residue as the acyl substituent of O-6 of the terminal heptose (spin system Z). Actually, the H-6 signal of this latter heptose residue was notably deshielded at lower fields as a consequence of acylation (4.84 ppm), and in the HMBC spectrum (Figure 5A) both glycine methylene (3.92 ppm) proton signal and H-6 signal of heptose Z were correlated to glycine carbonyl 13C resonance at 168.5 ppm, thus univocally indicating by a scalar correlation the covalent appendage of glycine. The terminal heptose bearing glycine of the S. flexneri M90T ΔgalU mutant corresponds to the second heptose in the wild-type LOS. The 2D NMR spectra also gave account of the presence of a PEtN group that was located by means of 31P NMR spectroscopy. In fact, the 31P NMR spectrum (Figure 5B) showed the presence of a pyrophosphate group at −9.8 ppm and its existence was only possible if PEtN was linked to the phosphate substituting Hep at O-4. This group correlated in the 1H, 31P-HSQC spectrum to H-4 of the internal heptose and to both methylene signals of PEtN. In the spectra it was also possible to detect a different oligosaccharide deprived by PEtN in which the heptose only carried a phosphate group.

**Discussion**

*S. flexneri* is a Gram-negative human pathogen bacterium able to colonize the intestinal mucosa. LPSs are well-recognized...
pathogen associated molecular patterns (PAMPs) responsible for stimulation of the innate immune response. Actually, it has been demonstrated that the Toll-like receptor 4 (TLR4), a member of the Toll-like receptor family, serves as the main upstream sensor for the LPS effect in vitro and in vivo (Medzhitov 2001; Akira et al. 2006). TLR-4 is part of a trimolecular LPS receptor complex that contains two co-receptors: MD2, a secreted accessory protein required for the correct positioning of TLR4 on the cell surface and for TLR4 transmembrane signal transduction, and soluble- or membrane-bound CD14 (Alexander and Rietschel 2001; Medzhitov 2001; Akira et al. 2006). Therefore, it is crucial to understand the full chemical nature of LPSs in order to dissect the molecular processes underlying the interaction of this pathogen with the components of the innate immunity system. A previous work by others has already addressed the issue of LPS O-chain glucosylation of S. flexneri M90T and its outstanding biological consequence (West et al. 2005).

In the present work we continue on this issue unambiguously identifying for the first time the carbohydrate core oligosaccharide obtained by the R-LPS of S. flexneri M90T. We consider this finding very important since the core region has a crucial role in the molecular interaction of LPS with its cognate TLR4 receptor. It is also worth noting that the whole structural elucidation was inferred using the small amount of R-type LPS that S. flexneri M90T produces under canonical conditions and that this is the first time that the core oligosaccharide primary structure of Shigella LPS is established by a clear-cut approach. Chemical and spectroscopical investigation of the de-acetylated LPS showed that the inner core structure is characterized by a L,D-Hep-(1→7)-L,D-Hep-(1→3)-L,D-Hep-(1→5)-Kdo-(2→4)]-Kdo sequence that is the common structural theme identified in Enterobacteriaceae. In addition, in S. flexneri M90T LPS, a glucosamine residue is additionally sitting at O-7 position of the third heptose whereas the outer core is characterized by glucose and galactose residues.

Several studies also indicate that the core oligosaccharide is usually characterized by the presence of noncarbohydrate substituents as phosphate, acyl groups, and amino acids. Indeed, glycine is a common component present in the LPS of several Gram-negative bacteria including a survey of LPSs from over 30 strains of Escherichia, Salmonella, Hafnia, Campilobacter, Citrobacter, and also Shigella species (Gamian et al. 1996; Li et al. 2001; Dzieciatkowska et al. 2007) but nevertheless its precise chemical location is very seldom clear. Even if we found glycine in the compositional analysis of S. flexneri M90T wild-type LOS, the presence of Gly residue would hardly be detectable by other approaches, possibly due to the low amount of this amino acid within a very long carbohydrate chain. To confirm this glycine finding, we decided to create and analyze the M90T ΔgalU mutant LPS. The galU mutation generates a lipooligosaccharide molecule built up of a deeply truncated structure with the core restricted to two Kdo and two heptoses. Combining the information obtained from NMR and MALDI analyses it was possible to definitely confirm the presence of Gly and its position in the inner core as substituent at O-6 of the second heptose (Figure 6). Interestingly, in a very recent paper dealing with structural elucidation of LOS of Campilobacter jejuni Gly was found to be located on the same distal heptose residue in each glycine containing strain (Dzieciatkowska et al. 2007). However, it was not possible in that case to get further information about the linkage of glycine residue.

The biological function of this substituent is still unknown but we speculate that it may also play a key role in pathogenesis since the presence of this residue modifies the net charge surface on the external membrane rendering it positively charged or in an isoelectric state. Residues such as Ara4N, PCho, PEtN all bearing a free amino group that under a physiological condition...
is positively charged are able to confer resistance to antibi-
otic compounds and host cationic antimicrobial peptides. These
residues prevent the antimicrobial action polymyxin B of in-
creasing permeability in the bacterial outer membrane (Trent,
Zhou et al. 2001). Likewise, we hypothesize that also Gly might
play this role. Interestingly and in agreement, Gly is not the first
amino acid found in the core region of LPSs but in this case, its
amino group is present in a free form and thus positively charged
at physiological pH, whereas, for example, alanine is present in
the inner core of P. aeruginosa LPS (Knirel et al. 2006), but
amino function is “naturally” protected by an acetyl group.

Finally, from the biosynthetical point of view, it was not
surprising to find that the S. flexneri M90T ΔgalU LPS lacks
the third heptose residue and its attached GlcN substituent in
the inner core. Actually, the core region of S. flexneri LPS is
very similar to E. coli R-1 core type in which the third heptose
is substituted by a nonstoichiometric amount of GlcN (Holst
1999). In full agreement, the E. coli J-5 rough mutant LPS,
which shares the same inner core carbohydrate composition, is
constituted by a blend of oligosaccharides in which the third
heptose and GlcN are nonstoichiometrically present (Müller-
Loennies et al. 1994). It is very likely that this Hep residue
is added at a later stage of core oligosaccharide biosynthesis
by another heptosyl transferase and that this enzyme does not
recognize the core oligosaccharide of S. flexneri M90T ΔgalU
LPS as a substrate, i.e., a larger oligosaccharide with the outer
core is required.

Bacteria were cultured in the Trypticase Soy Broth
(TSB) (BBL, Becton Dickinson and Co., Cockeysville, MD)
or agar (TSA). Recombinant DNA techniques were car-
ried out following standard procedures. To obtain M90T ΔgalU two DNA fragments from M90T galU were am-
plified by PCR with the following primers: galUF1 (5′-
GCTCTAGAGGCCGTTATCCCCGTTGC-3′) and galUR1 (5′-
GCTGCAAGGACCTGACGAACTTG-3′) for fragment 1; galUF2 (5′-
GCTGCAGGACCTGACGAACTTG-3′) and galUR2 (5′-
GCTGCAAGGACCTGACGAACTTG-3′) for fragment 2. The two fragments were then ligated with an 829-bp
PsI cassette encoding chloramphenicol resistance from pGEM-
CAT, cloned into the suicide plasmid pGP704 and introduced
into M90T to replace the wild-type galU through allelic ex-
change.

Isolation and purification of the LPS and LOS fraction
Dried cells from M90T and M90T ΔgalU were extracted ac-
ording to the phenol–water method (Westphal and Jahn 1965)
and the Galanos method, respectively (Galanos et al. 1969)
yield, 10 mg and 60 mg, respectively). SDS–PAGE (12%) was
performed and stained with silver nitrate for detection of LPS
and LOS (Kittelberger and Hilbink 1993). The extracts were
both digested with DNase, RNase, and Proteinase K, dialyzed,
lyophilized, and further purified by gel filtration chromatogra-
phy (Sephacryl S-500 50 × 1.5 cm, water as eluent).

The LOS fraction obtained from M90T (10 mg) and an aliquot
M90T ΔgalU LOS (20 mg) was dissolved in anhydrous hy-
drazine (1 mL), stirred at 37°C for 90 min, cooled, poured
into ice-cold acetone (20 mL), and allowed to precipitate. The
precipitate was then centrifuged (3000 × g, 30 min), washed
twice with ice-cold acetone, dried, and then dissolved in wa-
ter and lyophilized (10 mg, 80% of LOS). This material was
subsequently de-N-acylated with 4 M KOH and desalted using
a column (50 × 1.5 cm) of Sephadex G-10 (Pharmacia). The

Materials and methods

Bacteria and construction of S. flexneri 5 ΔgalU

S. flexneri 5, M90T, was originally isolated by Sansonetti et al.
(1982).
oligosaccharide fraction obtained from a wild-type sample was further purified by HPLC (TSK gel G3000 PWXL 0.78 × 30, water as eluent).

General and analytical methods
Monosaccharide analysis was obtained by GLC analysis of their O-methyl glycoside derivatives whereas the absolute configuration was assigned by GLC analysis of their (+)-O-oct-2-yl glycoside derivatives (Leontein and Lönngren 1978).

The methylation analysis was carried out on a dephosphorylated sample obtained with 48% HF (4°C, 48 h). For methylation analysis of a Kdo region, lipooligosaccharide was carboxy-methylated with methanolic HCl (0.1 M, 5 min) and consecutively with diazomethane in order to improve its solubility in DMSO. Methylation was carried out as described (Ciucanu and Kerek 1984). Lipooligosaccharide was hydrolyzed with 2 M trifluoroacetic acid (100°C, 1 h), carbonyl-reduced with NaBD₄, carboxymethylated as before, carboxyl-reduced with NaBD₄ (4°C, 18 h), and acetylated and analyzed by GLC-MS. Methylation of the complete core region was carried out as described (Ciucanu and Kerek 1984), and the sample was hydrolyzed with 4 M trifluoroacetic acid (100°C, 4 h), carbonyl-reduced with NaBD₄, carboxy-methylated, carboxyl-reduced, and acetylated and analyzed by GLC-MS. Fatty acid analyses were executed as described (Rietschel 1976; Molinaro et al. 2003). GLC and GLC-MS were all carried out on a Hewlett-Packard 5890 instrument, SPB-5 capillary column (0.25 mm × 30 m, Supelco); for sugar methylation analysis and O-methyl glycosides derivatives the temperature program was 150°C for 2 min, then 2°C min⁻¹ to 200°C for 0 min, then 10°C min⁻¹ to 260°C for 11 min, and then 8°C min⁻¹ to 300°C for 20 min. For fatty acids analysis the temperature program was 80°C for 2 min and then 8°C min⁻¹ to 300°C for 15 min.

NMR spectroscopy
For structural assignments of oligosaccharide 1D and 2D ¹H NMR spectra were recorded on a solution of 2 mg in 0.5 mL of D₂O, at 298 K; on Bruker 600 DRX equipped with a cryo probe. Spectra were calibrated with internal acetone [δH 2.225, δC 31.45]. ³¹P NMR experiments were carried out using a Bruker DRX-400 spectrometer; aqueous 85% phosphoric acid was used as an external reference (0.00 ppm). ROESY was measured using data sets (t₁ × t₂) of 4096 × 512 points; a mixing time of 200 ms was used whereas for TOCSY experiments a spinlock time of 120 ms was used with a data set (t₁ × t₂) of 4096 × 512 points. A DQF-COSY experiment was performed using data sets of 4096 × 1024 points. In all homonuclear experiments the data matrix was zero-filled in the F₁ dimension to give a matrix of 4096 × 2048 points and resolution-enhanced in both dimensions by a shifted sine-bell function before Fourier transformation. Coupling constants were determined on a first-order basis from DQF-COSY (States et al. 1982; Rance et al. 1983). HSQC and HMBC experiments were measured in the ¹H-detected mode via single quantum coherence with proton decoupling in the ¹³C domain, using data sets of 2048 × 256 points. Experiments were carried out in the phase-sensitive mode according to the method of States et al. (1982). ¹H,¹³C HMBC was optimized for 6 Hz coupling constant and ¹H,³¹P HSQC for 8 Hz coupling constant.

MALDI TOF analysis
MALDI-TOF analyses were performed using a Voyager STR instrument (Applied Biosystems, Framingham, MA) equipped with nitrogen laser (λ = 337 nm) and provided with delayed extraction technology. Ions generated by the pulsed laser beam were accelerated through 24 kV. Mass spectra reported are the result of 256 laser shots.

MALDI preparation of native LOS was performed according to the thin layer procedure described (Sturiale et al. 2005) by using 2,4,6-trihydroxyacetophenone (THAP) as matrix, while the oligosaccharide sample for MS analysis was prepared utilizing a matrix solution of dihydroxybenzoic acid (DHB) 50 mg/mL in TFA 0.1%-ACN 80/20, by the classic dried drop method: 1 µL of a sample/matrix solution mixture (1:1, v/v) was deposited onto a stainless-steel MALDI sample plate and left to dry at room temperature.

Acknowledgements
Financial support to D.G. from the Research Project “Innovazione e Tecnologie per il Miglioramento della Sostenibilità Agroindustriale, della Sicurezza e Qualità Alimentare” financed in the framework of the agreement between “Regione Sicilia,” MIUR, and MEF (APQ, project code RS-20).

Conflict of interest statement
None declared.

Abbreviations
DQF-COSY, double quantum-filtered phase-sensitive correlation spectroscopy; HMBC, heteronuclear multiple bond correlation; HSQC, heteronuclear single quantum coherence; LOSs, lipooligosaccharide; LPSs, lipopolysaccharides; NOESY, nuclear overhauser enhancement spectroscopy; ROESY, rotating-frame overhauser enhancement spectroscopy; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TOCSY, total correlation spectroscopy experiments.

References


