Supplemental information

Content:

Supplemental Text

Determination of the lyso structure from low-GC Gram-positive gut microbes

In addition to *Enterococcus faecalis* and *Bacillus cereus* as described in the main text, we determined lyso-structures of lipoproteins from *Lactobacillus delbrueckii* ssp. *bulgaricus* (*L. bulgaricus*), a probiotic and lactic acid-producing Gram-positive bacterium originally isolated from Bulgarian yogurt. The lipoprotein-enriched TX114 phase fraction of *L. bulgaricus* cells was separated using SDS-PAGE, and a 38-kDa band containing Ldb2183 (Lbd number from strain ATCC11842) (1) was analyzed. The MALDI-TOF MS spectrum showed a peak at *m*/*z* 996.7 (Fig. S1*A*), which corresponds to the N-terminal CSGK lipopeptide of the Ldb2183 containing lipid-modified cysteine with possibly two fatty acids, 36:2 in total. Ions corresponding to triacyl form were not observed (Fig. S1*A*). More, ions at *m*/*z* 996.7 had very low intensities of ions with different lengths of fatty acids (Fig. S1*A*, insert), which was similar lipidated pattern to *E. faecalis* PnrA, suggesting the *L. bulgaricus* lipopeptide was modified with particular fatty acids as discussed later and Table S1. Lipoprotein lipase removed only one 18:1 fatty acid from the lipopeptide (Fig. S1, *B* and *C*), indicating that the lipopeptide has one ester-linked 18:1 fatty acid. The MS/MS spectrum of the ion at *m*/*z* 996.7 provided y-series ions that confirmed the SGK peptide sequence of *L. bulgaricus* Ldb2183 (Fig. S1, *D* and *E*). Ions at *m*/*z* 732.3 corresponded to a product fragment ion that lost 18:1 fatty acid, and a characteristic fragment ion at *m*/*z* 624.2 corresponded to an *N*-acyl(18:1)-dehydroalanyl peptide generated by the neutral loss of monoacyl(18:1)thioglycerol. Oxidation of the lipopeptide by H_2O_2 enhanced the neutral loss (Fig. S1*D*, insert), in which the thioether sulfur at the lipidated cysteine

residue of lipoproteins was added by oxygen and resulting sulfoxide group underwent facile neutral loss of acyl(18:1)oxyhydroxypropane-1-sulfenic acid from the lipopeptides by MS/MS (2). Together, these results provide the existence of the lyso-form of this *L. bulgaricus* lipoproteins, having an *N*-acyl(18:1)-*O*-acyl(18:1)-*S*-glyceryl-cysteine structure (Fig. S1*E*). We could not determine the exact *O*-acylation position; however, we assumeed that *sn*-1 (3-position)-lyso form is the major component because of the same reason as described in the main text for *E. faecalis* PnrA. Additionally, phosphonate ABC transporter substrate-binding PhnD (Ldb0202) protein was identified as the same lyso form of the *N*-acyl(18:1)-*O*-acyl(18:1)-*S*-glyceryl-cysteine structure (Table 1).

Next we show lyso-form structures of *Streptococcus sanguinis*, a Gram-positive bacterium commonly found in oral cavities. The lipoprotein-enriched TX114 phase fraction of *S. sanguinis* cells was separated by SDS-PAGE, and a 35.5-kDa band containing SSA_1038 (number refers to strain SK36) (3) was digested in-gel with trypsin, extracted using organic solvent, and analyzed with MALDI-TOF MS. The obtained MS spectrum showed a peak at *m/z* 1025.7 (Fig. S2*A*), which corresponds to the N-terminal CGNR lipopeptide of SSA_1038 that contains a lipid-modified cysteine with one saturated and one mono-unsaturated fatty acids, 34:1 in total. Ions corresponding to the conventional triacyl form were not observed (Fig. S2*A*). This *S. sanguinis* lipopeptide ion at *m/z* 1025.7 showed very low amounts of 14-Da interval ions (Fig. S2*A*, insert), being similar to *E. faecalis* PnrA (Fig. 1*B*, insert) but being different from *S. aureus* lipoproteins (2), suggesting the lipopeptide obtained from SSA_1038 was modified with specific fatty acids as discussed later. The lipoprotein lipase removed only one 18:1 fatty acid from the lipopeptide (Fig. S2, *B* and *C*), suggesting that this lipopeptide has one *O*-esterified 18:1 fatty acid. The MS/MS spectrum of the ion at *m/z* 1025.7 provided C-terminus-containing y- and y*-series ions that confirmed the GNR peptide sequence of *S. sanguinis* SSA_1038 (Fig. S2, *D* and *E*). Ions at *m/z* 743.4 and 787.5 corresponded to product ions that had lost 18:1 fatty acid and 16:0 ketene, respectively, and a characteristic fragment ion at *m/z* 653.4 corresponded to *N*-acyl(16:0)-dehydroalanyl peptide generated by the neutral loss of monoacyl(18:1)thioglycerol. Oxidation of the lipopeptide by H_2O_2 enhanced the neutral loss (Fig. S2*D*, insert), as described above for *L. bulgaricus* Lbd2183. Taken together, these results provide the existence of the lyso-form of this *S. sanguinis* lipoprotein; one 18:1 fatty acid is linked via an *O*-ester

bond and one acyl(16:0) group is linked via an amide bond (Fig. S2*E*). Additionally, the SSA_0375 was identified to have the same lyso form at its N-terminal *N*-acyl(16:0)-*O*-acyl(18:1)-*S*-glyceryl-CGGSSSK lipopeptide (Table 1). These results suggest that the lyso-form is a well-distributed structure of lipoproteins in low-GC content Gram-positive bacteria, and that at least several Gram-positive bacterial species in gut microbes and a probiotic strain produce the lyso form lipoproteins.

Determination of the *N*-acetyl structure

In addition to lipoproteins from *B. licheniformis*, a soil bacterium capable of causing food poisoning, described in main text, we found the second new *N*-acetyl forms of lipoproteins in other Gram-positive bacteria. Regarding *Bacillus subtilis*, a best characterized low-GC content Gram-positive bacteria that is used in the production of Natto, a traditional Japanese dish of fermented soya beans, MS analysis of the in-gel-digested 34-kDa band revealed a peak at *m*/*z* 1188.7 (Fig. S4*A*), which corresponded to the N-terminal CSSGDK lipopeptide of peptidylprolyl isomerase PrsA (BSU09950 of strain 168) (4) with two possible saturated fatty acids, 35:0 in total. Ions having different lengths of fatty acids were obvious in this lipopeptide (Fig. S4*A*, insert). The lipoprotein lipase removed two *O*-esterified fatty acids; however, the two released fatty acids were 15:0 and 17:0 (Fig. S4, *B*-*D*). Thus, the total carbon number of the two was 32:0, which was smaller than the estimated number of 35:0. Further MS/MS spectrum of the ion at *m/z* 1188.7 provided y-series ions clarifying the SSGDK peptide sequence of *B. subtilis* PrsA (Fig. S4, *E* and *F*). The ion at *m/z* 946.6 in Fig. S4*E* corresponded to a product ion that had lost a 15:0 fatty acid. Importantly, the MS/MS spectrum provided a characteristic fragment ion at *m/z* 604.3 that corresponded to the *N*-acetyl-dehydroalanyl peptide generated by the neutral loss of diacyl(15:0/17:0)thioglycerol. Edman degradation did not provide the N-terminal sequence. These results indicate that the acetyl group is linked at the α-amino group of the lipidated cysteine via an amide bond, and two saturated fatty acids (15:0/17:0) are linked to the *S*-glyceryl group (Fig. S4*F*). In general, modification of the *sn*-1 position of the glyceryl group in phospholipid is more stable than *sn*-2 modification (5), as described in the main text. Thus, only the ion at *m*/*z* 946.6 (17:0-containing), and not the 15:0-containing ion, was

represented in the MS/MS spectrum (Fig. S4*E*), suggesting that the 2 and 3 positions of glyceryl group could be modified by the 15:0 and 17:0 fatty acids, respectively (Fig. S4*F*). Additionally, BSU0630 lipoprotein was identified to take the same *N*-acetyl form, *N*-acetyl-*O*-diacyl(17:0/15:0)-*S*-glyceryl-cysteine, as its major form (Table 1).

To examine how the *N*-acetyl form is ubiquitous in *Bacillus*-related strains, we determined N-terminal structures of lipoproteins from three other strains surviving in extremophilic conditions. As described in the main text, the *N*-acetyl form was identified from alkaliphilic and extremely halotolerant *Oceanobacillus iheyensis* grown at pH 9.5, and thermophilic *Geobacillus kaustophilus* grown at 60 °C that was isolated from the deep sea of the Mariana Trench (Fig. 5, Table 1). In addition, we determined the structures of lipoproteins of *Bacillus halodurans* strain C-125, an alkaliphilic Gram-positive soil bacterium used in biotechnology fields (6). *B. halodurans* cells were grown aerobically at pH 9.5 until late log phase. MALDI-TOF MS analysis of an in-gel-digested band revealed a main ion at *m*/*z* 1111.7 and 14-Da interval ions (Fig. S5*A*). The main ion corresponded to the N-terminal CGPDR lipopeptide of maltose/maltodextrin transporter substrate-binding MalE (BH2926 of *B. halodurans* C-125 genome) (6) with two possible saturated fatty acids, 33:0 in total. The MS/MS spectrum of the ion at *m/z* 1111.7 provided y- and y*-series ions that confirmed the GPDR peptide sequence of *B. halodurans* MalE (Fig. S5*B*). The ion at *m/z* 869.6 corresponded to a product ion that had lost a 15:0 fatty acid. The MS/MS spectrum also provided a characteristic fragment ion at *m/z* 555.3 that corresponded to the *N*-acetyl-dehydroalanyl peptide generated by the neutral loss of diacyl(15:0/15:0)thioglycerol. Oxidation of the lipopeptide by H_2O_2 enhanced the neutral loss (Fig. S5*B*, insert), as described above for *L. bulgaricus* Lbd2183 (Fig. S1*D*, insert) and *S. sanguinis* SSA_1038 (Fig. S2*D*, insert). These results indicate that MalE in *B. halodurans* is the *N*-acetyl form: an *N*-acetyl-*O*-diacyl(15:0/15:0)-*S*-glyceryl-cysteine structure (Fig. S5*C*). Additionally, BH3460 was identified to take the same *N*-acetyl form, *N*-acetyl-*O*-diacyl(15:0/15:0)-*S*-glyceryl-cysteine, as its major form (Table 1). These results suggest that the *N*-acetyl form is also a well-distributed structure of lipoproteins in *Bacillus*-related strains. These *N*-acetyl form lipoprotein-producing bacteria grown in environments or food sources can be exposed to cutaneous immunity or to mucosal immune systems of the gut after ingestion.

N-terminal structure of the lipoproteins of mycoplasmas

TLR2-stimulating lipoprotein structures of mycoplasmas have been disputed. First, we analyzed lipoproteins of *Mycoplasma genitalium* and *Mycoplasma pneumoniae*, which were suggested to be triacyl-form stimulating TLR-1/2 heterodimer (7, 8), but they have no *E. coli*-type Lnt homologue and their detailed molecular structures of lipoproteins have not yet been investigated. The MS spectrum of the in-gel digest of the 29-kDa band from *M. genitalium* cells represented a main ion at *m/z* 1294.0 (Fig. S6*A*), which corresponded to the N-terminal CGATK lipopeptide of the hypothetical MG_040 lipoprotein with possibly two saturated and one mono-unsaturated fatty acids, 50:1 in total (MG number referred to strain G37) (9). Ions having different lengths of fatty acids were not observed: the 13.7-Da smaller ion at *m/z* 1280.3 than the main ion was a contaminated fragment ion (Fig. S6*A*, insert). The MS/MS spectrum of the ion at *m/z* 1294.0 provided y- and y°-series ions describing the GATK peptide sequence of *M. genitalium* MG_040 (Fig. S6, *B* and *C*). The ion at *m/z* 1037.5 corresponded to the product ion that had lost a 16:0 fatty acid. Importantly, the MS/MS spectrum of an oxidized lipopeptide fraction provided a characteristic fragment ion at *m/z* 683.5 that corresponded to the *N*-acyl(16:0)-dehydroalanyl peptide generated by the neutral loss of 2,3-diacyloxy(18:1/16:0)propane sulfenic acid (Fig. S6*B*, insert). These results indicate that this lipopeptide is the conventional triacyl form, having an *N*-acyl(16:0)-*O*-diacyl(18:1/16:0)-*S*-glyceryl-cysteine structure (Fig. S6*C*). Regarding modification of the 2 and 3 positions of *S*-glyceryl group, only the ion at *m/z* 1037.5 (18:1-containing), but not the 16:0-containing ion, was represented in the MS/MS spectrum (Fig. S6*B*)**,** suggesting that the 2 and 3 positions of the *S*-glyceryl group may be modified by a 16:0 and 18:1 fatty acid, respectively (Fig. S6*C*).

Lipoproteins of *M. pneumoniae* were then analyzed. The MS spectrum of the in-gel digest of the 30-kDa band represented a main ion at *m/z* 1322.0 (Fig. S6*D*), which corresponded to the N-terminal CGATR lipopeptide of the hypothetical MPN052 lipoprotein with possibly two saturated and one mono-unsaturated fatty acids, 50:1 in total (MPN number referred to strain M129) (10). This lipopeptide ion had little 14-Da interval ions (Fig. S6*D*, insert). The lipoprotein lipase removed two *O*-esterified fatty acids of 18:1 and 16:0 (Fig. S6, *D* and *E*). The MS/MS spectrum of the ion at *m/z*

1322.0 provided y- and y*-series ions clarifying the GATR peptide sequence of the *M. pneumoniae* MPN052 (Fig. S6, *F* and *G*). The ion at *m/z* 1065.8 corresponded to the product ion that had lost a 16:0 fatty acid. Importantly, the MS/MS spectrum provided a characteristic fragment ion at *m/z* 711.4 that corresponded to the *N*-acyl(16:0)-dehydroalanyl peptide generated by the neutral loss of diacyl(18:1/16:0)thioglycerol. These results indicate that the MPN052 is the conventional triacyl lipopeptide having the *N*-acyl(16:0)-*O*-diacyl(18:1/16:0)-*S*-glyceryl-cysteine structure (Fig. S6*G*). Regarding modification of the 2 and 3 positions of *S*-glyceryl group, only the ion at *m/z* 1065.8 (18:1-containing), and not the 16:0-containing ion, was observed in the MS/MS spectrum (Fig. S6*F*), suggesting that the 2 and 3 positions of the *S*-glyceryl group might be modified by 16:0 and 18:1 fatty acid, respectively (Fig. S6*G*). The same conventional triacyl structure was also determined for the MPN415 (high affinity transport system protein P37) of *M. pneumoniae* (Table 1). Taken together, these results demonstrate that *M. genitalium* and *M. pneumoniae* have the conventional triacyl form of lipoproteins, consistent with our prediction (7, 8). Also, these two mycoplasmas must have new-type Lnt enzyme, being consistent with a recent *Acholeplasma laidlawii* study (11).

Lipoproteins of *M. fermentans* were then analyzed, which is known to release MALP2 diacyl lipopeptide to stimulate TLR2 (12). We identified a protein band on an SDS-PAGE gel containing MALP2 precursor (MBIO_0763 of strain PG18) via LC-MS/MS (Fig. 6*A*). MALDI-TOF MS spectrum of an in-gel digest of this band represented a main ion at *m/z* 1906.1 (Fig. S7*A*), which corresponded to the N-terminal CGNNDESNISFK lipopeptide of the MALP2 precursor with possibly two saturated fatty acids, 34:0 in total. This main ion consisted of 14-Da interval ions with little intensities (Fig. S7*A*, insert). The MS/MS spectrum of the ion at *m/z* 1906.1 provided y-series ions (Fig. S7*B*) that confirmed the NNDESNISFK peptide sequence of *M. fermentans* MALP2 precursor. The MS/MS spectrum provided a characteristic fragment ion at *m/z* 1293.2 that corresponded to the dehydroalanyl peptide generated by the neutral loss of diacyl(34:0)thioglycerol. These results suggest that the MALP2 precursor in *M. fermentans* is conventional diacyl form, as reported previously (12), consisting of an *O*-diacyl(34:0 in total)-*S*-glyceryl-cysteine structure (Fig. S7*C*). The fatty acid modification of the 3 and 2 positions of the MALP2 precursor *S*-glyceryl group may be analogous to *M. fermentans* MBIO_0319 N-terminal structure (Fig. 6*A*). The conventional diacyl form was also

determined for the MBIO_0869 in *M. fermentans* (Table 1). We could not detect the peptidyl structure for both MALP2 precursor lipoprotein and MBIO_0869 lipoprotein in the MS spectrum. As described in main text, our analyses of the four *M. fermentans* lipoproteins found both the conventional diacylated form and the peptidyl form. In addition, lipoproteins in *M. fermentans* take only one of two structures.

Conventional diacyl structure in *Listeria monocytogenes*

Finally, we describe the N-terminal structure of lipoproteins of *L. monocytogenes*, a food-poisoning pathogen that is able to survive as an intracellular parasite: after ingestion of contaminated food, *Listeria* disseminates from the intestinal lumen to the central nervous system and the fetoplacental unit. Studies using lipoprotein-deleted mutant showed important roles of lipoproteins for virulence and TLR2-medated immune activation in *L. monocytogenes* (13). *L. monocytogenes* cells were grown aerobically at 37 °C until late log phase. MALDI-TOF MS analysis of an in-gel-digested band revealed a main ion at *m*/*z* 1260.7 with a little 14-Da interval ions (Fig. S7*E*). The main ion corresponded to the N-terminal CGGGSDSK lipopeptide of Lmo2196 protein (protein ID from strain EGD-e) (14) with two possible saturated fatty acids, 32:0 in total. The MS/MS spectrum of the ion at *m/z* 1260.7 provided y-series ions that confirmed the GGGSD peptide sequence of the *L. monocytogenes* Lmo2196 (Fig. S7*F*). Ions at *m/z* 1018.4 and 748.1 corresponded to product ions that had lost a 15:0 fatty acid and two fatty acids of 32:0 in total, respectively, and a characteristic fragment ion at *m/z* 676.0 corresponded to the dehydroalanyl peptide generated by the neutral loss of diacyl(17:0/15:0)thioglycerol. Oxidation of the lipopeptide by H_2O_2 enhanced the neutral loss (Fig. S7*F*, insert). These results indicate that Lmo2196 in *L. monocytogenes* is the conventional diacyl form: an *S*-diacyl(17:0/15:0)glyceryl-cysteine structure (Fig. S7*G*). Regarding modification of the 2 and 3 positions of *S*-glyceryl group, only the ion at *m*/*z* 1018.4 (17:0-containing), and not the 15:0-containing ion, was observed in the MS/MS spectrum (Fig. S7*F*), suggesting that the 2 and 3 positions of the *S*-glyceryl group might be modified by 15:0 and 17:0 fatty acid, respectively (Fig. S7*G*). Additionally, Lmo0135 and Lmo2219 lipoproteins were identified to take the same conventional diacyl form as its major form (Table 1). These results suggest that the conventional

diacyl form is the TLR2-activating lipoprotein structure of *L. monocytogenes*, and the *N*-acylation is not essential for *Listeria* cell growth and that the *N*-acyl state is not absolute structure of lipoproteins in *Firmicutes* species.

Putative lipoprotein modification enzymes

As described in main text, this study provides the evidences of presence of yet-to-be identified key enzymes involved in the bacterial lipoprotein biosynthesis. Here, we describe the probable biochemical features of the enzymes based on the structures of bacterial lipoproteins. We categorized the N-terminus lipid-modified structures of bacterial lipoproteins into the three classes A to C (*see* main text and Table 1). The class B lipoproteins are *N*-acylated despite a lack of *E. coli*-type Lnt homologue, suggesting the existence of new-type Lnt (Fig. 7, step 3 or 7). Although the *E. coli*-type Lnt uses membrane phospholipids as substrates to supply lipoproteins with fatty acids, the new-type Lnt producing the *N*-acetyl form (*N*-acetylation) must not use phospholipids as a substrate (step 7). It is likely that the new-type Lnt for the *N*-acetylation (step 7) shares protein structure and catalytic mechanism with Lnt producing the conventional triacyl form (*N*-long chain acylation) (step 3). According to this prediction, the new-type Lnt for *N*-long chain acylation may not use membrane phospholipids as substrates. Nonetheless, we cannot deny a possibility that the new-type Lnt for *N*-long chain acylation uses phospholipids as a fatty acid source, as like as *E. coli*-type Lnt.

In addition, structural information of lipoproteins implies a fatty acid preference of the new-type Lnt during biosynthesis of the conventional triacyl form (Fig. 7, step 3). Regarding *S. aureus*, 48% of fatty acids bound to α-amino group of the lipidated cysteine (R3 position) are 18:0, while the 18:0 occupied only 3.2% in total membrane or 13.4% in phosphatidylglycerol, and other fatty acids such as a 15:0 and 20:0 occupy 40% and 22% of total membrane fatty acids, respectively, in *S. aureus* (Table S1). These results suggest that Lnt of *S. aureus* seems to prefer the 18:0 fatty acid even though the source of fatty acids cannot be specified at this time. Similarly, the R3 position of MPN052 lipoprotein of *M. pneumoniae* was only a 16:0, while total membrane of *M. pneumoniae* has other two major fatty acids, 18:0 and 18:1, both occupy around 20% (Table S1). Thus, Lnt in *M. pneumoniae* may prefer the 16:0 fatty acid. These results suggest that new-type Lnt has fatty acid preference,

which is different among bacterial species.

The fatty acid preference was also observed in the lyso form lipoproteins. In *B. cereus* PrsA, 70% of fatty acids at the R3 position were 15:0, which was twice higher than the 15:0 content in total membrane (31.8%) (Table S1). Similarly, 92% of fatty acids at the R3 position of *E. faecalis* PnrA were 16:0, which was three-times higher than the 16:0 content in total membrane (27%) (Table S1). Also, more than 95% of fatty acids at the R3 position of *L. bulgaricus* Ldb2183 were the 18:1, while 16:0 (18.8%) and 16:1 (13.8%), along with 18:1 (35.8%), are major fatty acids of *L. bulgaricus* total membrane (Table S1). Thus, the enzyme for *N*-long chain acylation of the lyso form lipoproteins has the fatty acid preference in each bacterium. Two possible enzymes are predicted for *N*-long chain acylation of the lyso form lipoproteins (*see* Fig. 7): the new-type Lnt (step 3) or a putative lipoprotein transacylase (step 5) that transfers a fatty acid from the *S*-diacylglyceryl group to the α-amino group of the lipid modified cysteine (*see* main text). In low-GC-content monoderm bacteria, the *sn*-1 (R1) position of phospholipids is predominantly filled with C16 to C20 fatty acids and the *sn*-2 (R2) position of phospholipids is C14 to C16 fatty acid (15, 16). Therefore, the enzyme responsible for *N*-long chain acylation of *B. cereus* or *E. faecalis* lipoproteins may transfer a fatty acid from the *sn*-2 position of phospholipids to the α-amino group of the lipidated cysteine if the enzyme uses phospholipids as substrates. On the other hand, the enzyme of *L. bulgaricus* may do it from the *sn*-1 position of phospholipids.

Interestingly, a diderm spirochete *Borrelia burgdorferi*, the etiologic agent of Lyme disease, is known to have an unusual structure of lipoproteins where one of the ester-linked fatty acids is replaced by an acetyl group (17), suggesting that *B. burgdorferi* should have other lipoprotein modification enzyme(s). Thus, structural remodeling of acyl moieties in bacterial lipoproteins may take place in a wide variety of bacteria.

In conclusion, we provide solid structural information on bacterial lipoproteins and describe new modes of lipoprotein modification via putative unidentified enzymes. We also pose related questions on the physiological roles of these modifications as discussed in the main text. To solve these questions, identification of presumed key enzymes followed by biochemical, genetic, and immunological characterizations are necessary.

Supplemental References

1. van de Guchte, M., Penaud, S., Grimaldi, C., Barbe, V., Bryson, K., Nicolas, P., Robert, C., Oztas, S., Mangenot, S., Couloux, A., Loux, V., Dervyn, R., Bossy, R., Bolotin, A., Batto, J. M., Walunas, T., Gibrat, J. F., Bessieres, P., Weissenbach, J., Ehrlich, S. D., and Maguin, E. (2006) *Proc Natl Acad Sci U S A* **103,** 9274-9279
2. Asanuma, M., Kurokawa, K., Ichikawa, R., Ryu, K. H., Chae, J. H., Dohmae, N., Lee, B. L., and Nakayama, H. (2011) *FEBS J* **278,** 716-728
3. Xu, P., Alves, J. M., Kitten, T., Brown, A., Chen, Z., Ozaki, L. S., Manque, P., Ge, X., Serrano, M. G., Puiu, D., Hendricks, S., Wang, Y., Chaplin, M. D., Akan, D., Paik, S., Peterson, D. L., Macrina, F. L., and Buck, G. A. (2007) *J Bacteriol* **189,** 3166-3175
4. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhöft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.-F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.-M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.-H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B.-S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.-F., Zumstein, E., Yoshikawa, H., and Danchin, A. (1997) *Nature* **390,** 249-256
5. Nakanishi, H., Iida, Y., Shimizu, T., and Taguchi, R. (2010) *J Biochem* **147,** 245-256
6. Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S., and Horikoshi, K. (2000) *Nucleic Acids Res* **28,** 4317-4331
7. Shimizu, T., Kida, Y., and Kuwano, K. (2008) *Infect Immun* **76,** 3672-3678
8. Shimizu, T., Kida, Y., and Kuwano, K. (2007) *Immunology* **121,** 473-483
9. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, R. D., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M.,

Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., 3rd, and Venter, J. C. (1995) *Science* **270,** 397-403

10. Dandekar, T., Huynen, M., Regula, J. T., Ueberle, B., Zimmermann, C. U., Andrade, M. A., Doerks, T., Sánchez-Pulido, L., Snel, B., Suyama, M., Yuan, Y. P., Herrmann, R., and Bork, P. (2000) *Nucleic Acids Res* **28,** 3278-3288
11. Serebryakova, M., Demina, I., Galyamina, M., Kondratov, I., Ladygina, V., and Govorun, V. (2011) *J Biol Chem* **286,** 22769-22776
12. Mühlradt, P. F., Kieß, M., Meyer, H., Süßmuth, R., and Jung, G. (1997) *J Exp Med* **185,** 1951-1958
13. Machata, S., Tchatalbachev, S., Mohamed, W., Jänsch, L., Hain, T., and Chakraborty, T. (2008) *J Immunol* **181,** 2028-2035
14. Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A., Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Chetouani, F., Couvé, E., de Daruvar, A., Dehoux, P., Domann, E., Domínguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K. D., Fsihi, H., Garcia-Del Portillo, F., Garrido, P., Gautier, L., Goebel, W., Gómez-López, N., Hain, T., Hauf, J., Jackson, D., Jones, L. M., Kaerst, U., Kreft, J., Kuhn, M., Kunst, F., Kurapkat, G., Madueño, E., Maitournam, A., Vicente, J. M., Ng, E., Nedjari, H., Nordsiek, G., Novella, S., de Pablos, B., Pérez-Diaz, J. C., Purcell, R., Remmel, B., Rose, M., Schlueter, T., Simoes, N., Tierrez, A., Vázquez-Boland, J. A., Voss, H., Wehland, J., and Cossart, P. (2001) *Science* **294,** 849-852
15. Kaneda, T. (1977) *Bacteriol Rev* **41,** 391-418
16. Fischer, W. (1994) *Med Microbiol Immunol* **183,** 61-76
17. Beermann, C., Lochnit, G., Geyer, R., Groscurth, P., and Filgueira, L. (2000) *Biochem Biophys Res Commun* **267,** 897-905

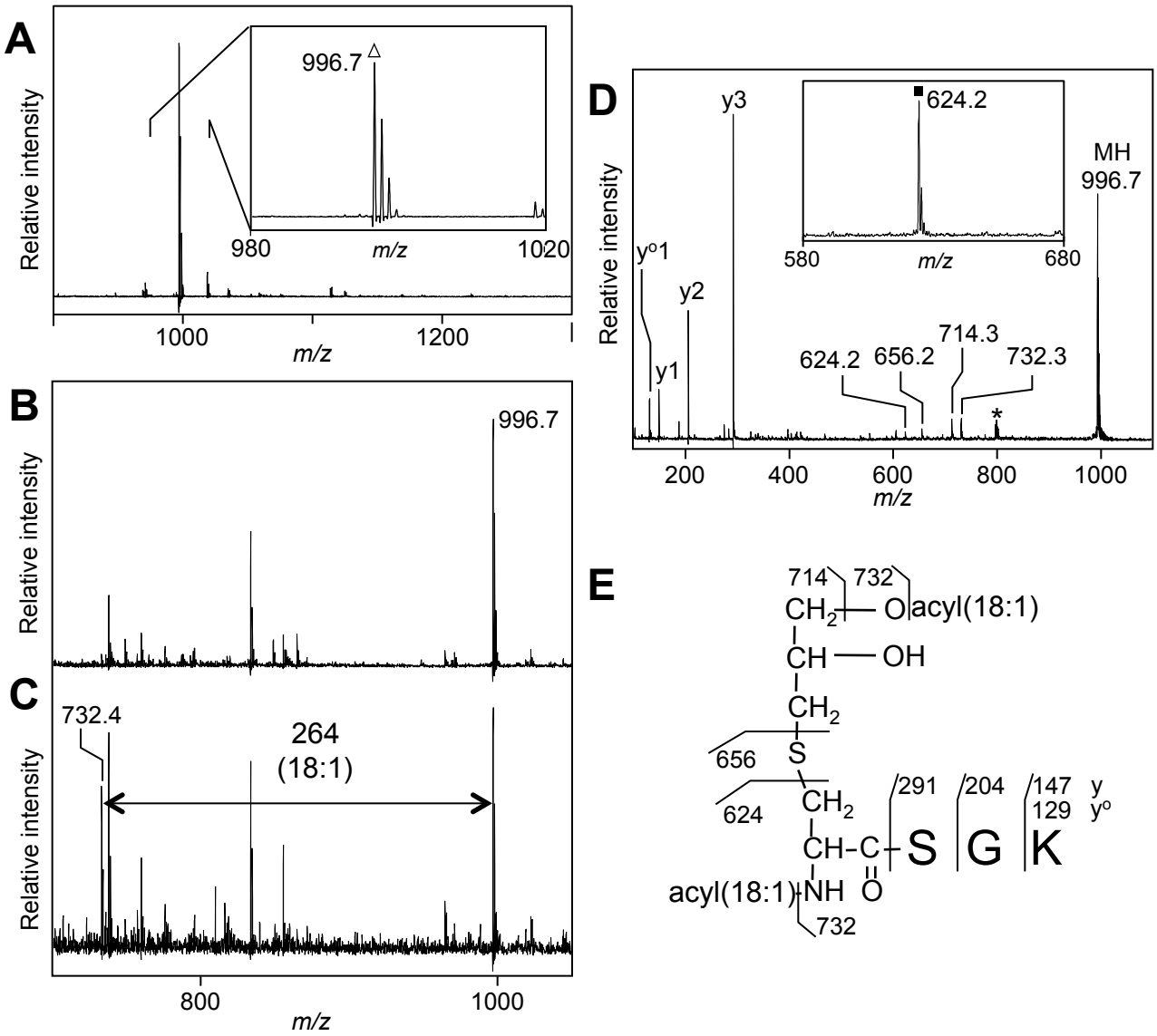

Figure S1. Lyso structure of lipoprotein from *Lactobacillus bulgaricus*.
Shown are MALDI-TOF MS spectra of a fraction containing in-gel-digested Ldb2183 lipoprotein of *L. bulgaricus* (*A*) or of the fraction incubated with lipoprotein lipase for 0 h (*B*) or 41 h (*C*). A deacylated product ion is highlighted. The insert in (*A*) is the magnified view. Shown in right hand are MS/MS spectra of the N-terminal lipopeptide ion at *m*/*z* 996.7 (*D*) or of the ion after oxidation (*D*, insert) and elucidated lyso structure of *L. bulgaricus* Ldb2183 (*E*). The y-series ions, y°1 ion that has lost H_2O from y1 ion, and product ions that have lost a 18:1 fatty acid, 18:1 ketene or monoacyl(18:1)thioglycerol are highlighted. The asterisk indicates a contaminated ion derived from the matrix. The closed square and open triangle indicate the mono-isotopic lipopeptide ion containing one or two acyl chain(s) with total number of double bonds of one and two, respectively.

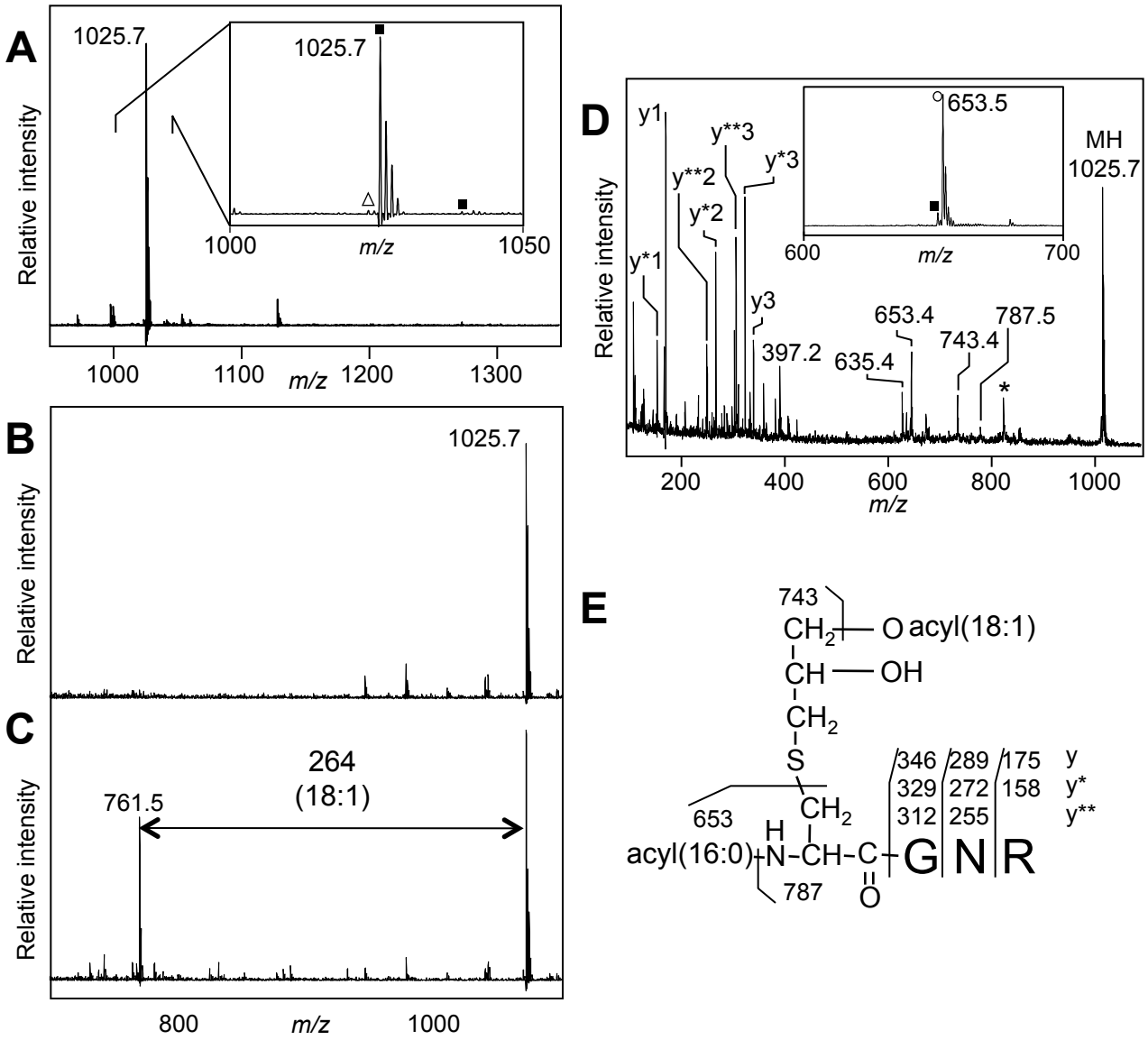

Figure S2. Lyso structure of lipoprotein from *Streptococcus sanguinis*.
MALDI-TOF MS spectra of a fraction containing in-gel-digested SSA_1038 lipoprotein of *S. sanguinis* (*A*) or of the fraction incubated with lipoprotein lipase for 0 h (*B*) or 72 h (*C*). A deacylated product is highlighted. The insert in (*A*) is the magnified view. Shown in right hand are MS/MS spectra of the ion at *m/z* 1025.7 (*D*) or of the ion after oxidation (*D*, insert) and elucidated lyso structure of *S. sanguinis* SSA_1038 (*E*). The C-terminus-containing y-series ions, y* ions that have lost ammonium from y ions, product ions that have lost a 18:1 fatty acid, 16:0 ketene or monoacyl(18:1)thioglycerol are highlighted. The asterisk indicates a contaminated ion derived from the matrix. The open circle, closed square and open triangle indicate the mono-isotopic lipopeptide ion containing one or two acyl chain(s) with total number of double bonds of zero, one and two, respectively.

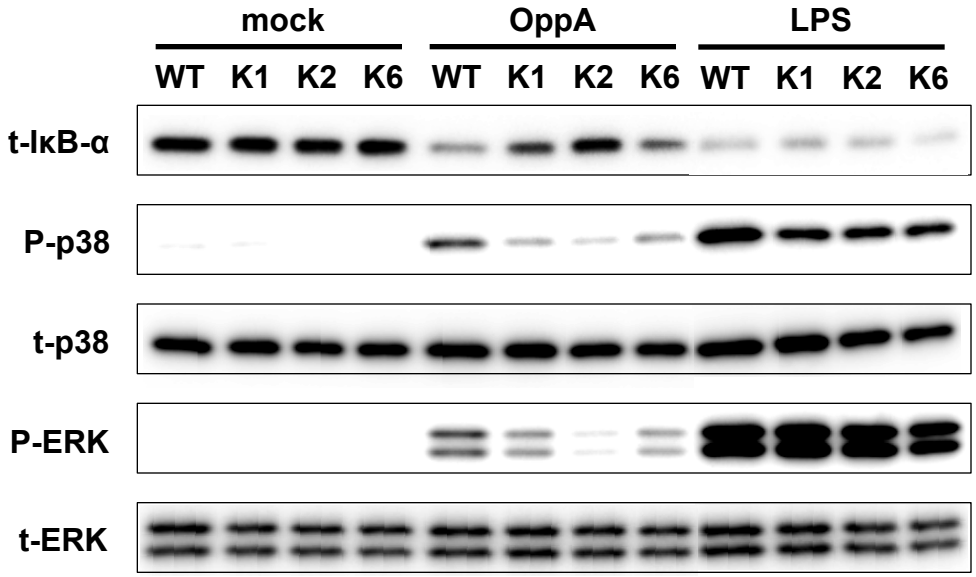

Figure S3. IκB-α degradation and MAP kinase activation induced by *B. cereus* lyso form OppA lipoprotein in peritoneal macrophages.
Thioglycollate-induced peritoneal macrophages (2.5 x 10^5 cells) prepared from the parental C57BL/6 (WT). TLR1-/- (K1), TLR2-/- (K2), or TLR6-/- (K6) mouse were stimulated for 30 min with or without (mock) 1 μg/ml of purified *B. cereus* OppA lipoprotein or 100 ng/ml LPS. Cell lysates were separated using 5-20% gradient SDS-PAGE and immunoblotted for total IκB-α (t-IκB-α), phosphorylated p38 MAPK (P-p38), total p38 MAPK (t-p38), phosphorylated ERK1/2 (P-ERK) and total ERK1/2 (t-ERK). The level of t-ERK1/2 were used to validate loading amount. Images are grouped from different parts of the same membrane exposure.

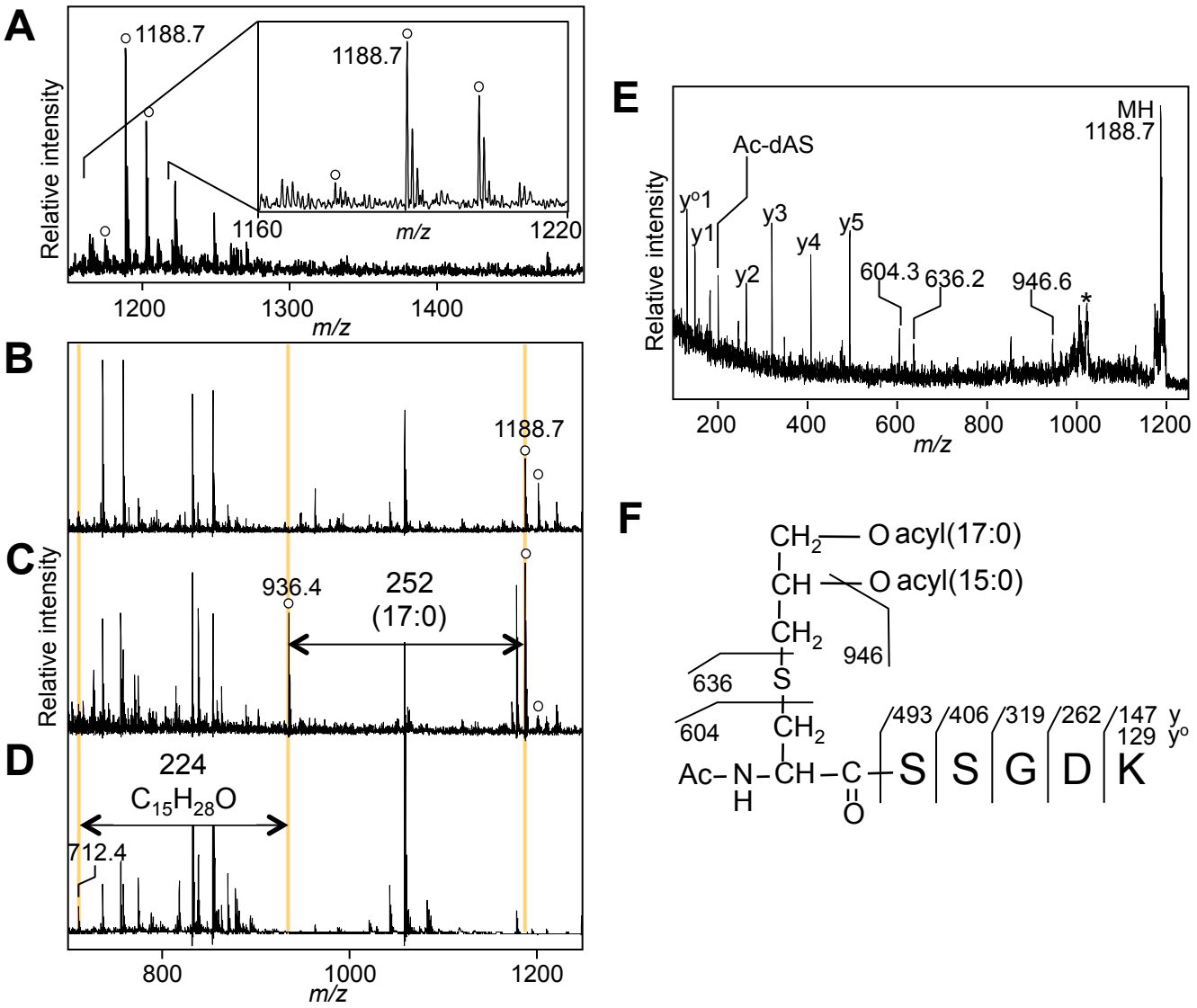

Figure S4. *N*-acetyl structure of PrsA lipoprotein from *Bacillus subtilis*.
MALDI-TOF MS spectra for a fraction containing in-gel-digested PrsA lipoprotein of *B. subtilis* (*A*) or for the fraction incubated with lipoprotein lipase for 0 h (*B*), 47 h (*C*) or 97 h (*D*). N-terminal PrsA fragment ion at *m*/*z* 1188.7 and its deacylated product ions are highlighted. The insert in (*A*) is the magnified view. The open circle indicate the mono-isotopic lipopeptide ion containing three saturated fatty acids. Further shown are MS/MS spectrum of the ion at *m*/*z* 1188.7 (*E*) and elucidated *N*-acetyl structure of *B. subtilis* PrsA (*F*). The y-series ions, y°1 ion that has lost H_2O from y1 ion, and product ions that have lost a 15:0 fatty acid or diacyl(17:0/15:0)thioglycerol are highlighted. The asterisk indicate the contaminated ions derived from the matrix.

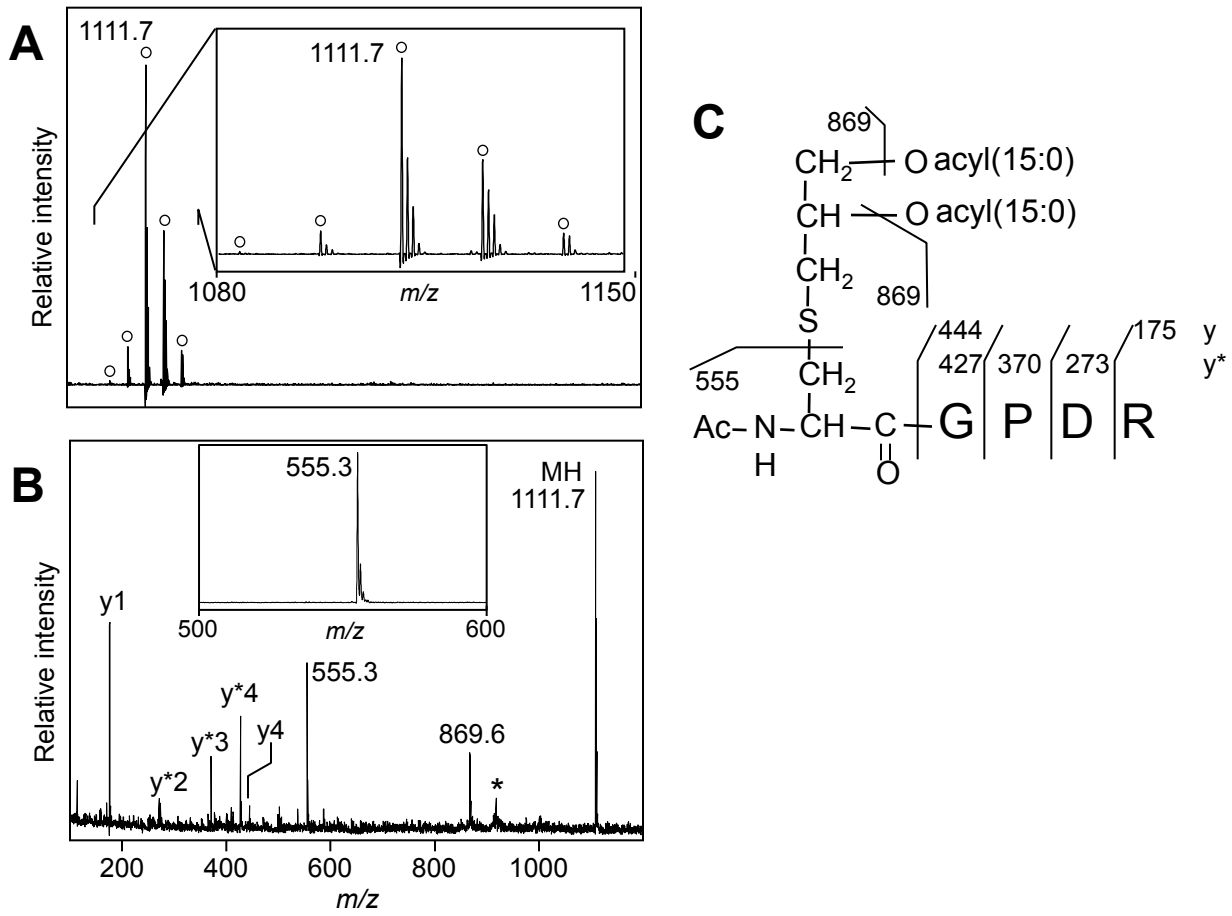

Figure S5. *N*-acetyl structure of MalE lipoprotein from *Bacillus halodurans*.
Shown are MS spectrum of a fraction containing in-gel-digested MalE from *B. halodurans* cells grown in alkaline medium (*A*), MS/MS spectra of an ion obtained at *m/z* 1111.7 (*B*) or of the ion after oxidation (*B*, insert), and elucidated *N*-acetyl structure of MalE (*C*). The y-series ions, y* ions that lost ammonium from y ions, and product ions that have lost a 15:0 fatty acid or diacyl(15:0/15:0)thioglycerol are highlighted. The insert in (*A*) is the magnified view. The open circle and asterisk indicate the mono-isotopic lipopeptide ion containing three saturated fatty acids and contaminated ions derived from the matrix, respectively.

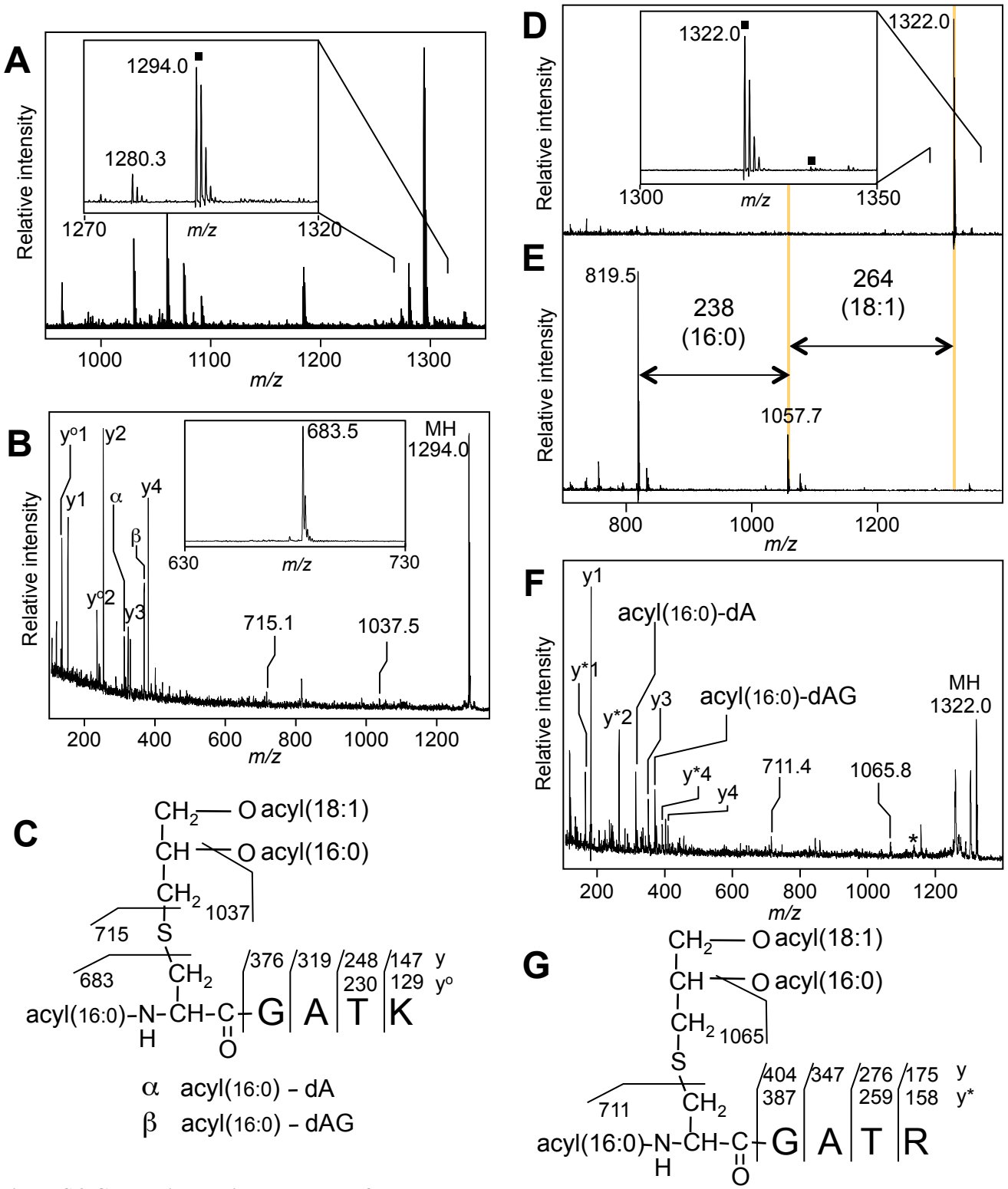

Figure S6. Conventional triacyl structure from mycoplasmas.

(*A*-*C*) Analysis of MG_040 lipoprotein of *Mycoplasma genitalium*. Shown are MS spectrum of a fraction containing in-gel-digested MG_040 of *M. genitalium* (*A*), MS/MS spectra of a major ion obtained at *m*/*z* 1294.0 (*B*) or of the ion after oxidation (*B*, insert), and elucidated conventional triacyl structure of *M. genitalium* MG_040 (*C*). The y-series ions, y° ions that have lost H_2O from y ions, and other product ions are highlighted.

(*D*-*G*) Analysis of MPN052 lipoprotein of *Mycoplasma pneumoniae*. MS spectra of a fraction containing in-gel-digested MPN052 incubated with lipoprotein lipase for 0 h (*D*) or 264 h (*E*) is shown. An N-terminal MPN052 fragment ion and its deacylated product ions are highlighted. Further shown are MS/MS spectrum of the N-terminal fragment ion at *m*/*z* 1322.0 (*F*) and elucidated triacyl structure of *M. pneumoniae* MPN052 (*G*). The y-series ions, y* ions that have lost ammonium from the y ions, and other product ions are highlighted. The asterisk indicates a contaminated ion derived from the matrix. The insert in (*A*) or (*D*) is the magnified view. The closed square indicates the mono-isotopic lipopeptide ion containing two saturated and one mono-unsaturated fatty acids. Ion at *m*/*z* 1280.3 was not the MG_040 lipopeptide.

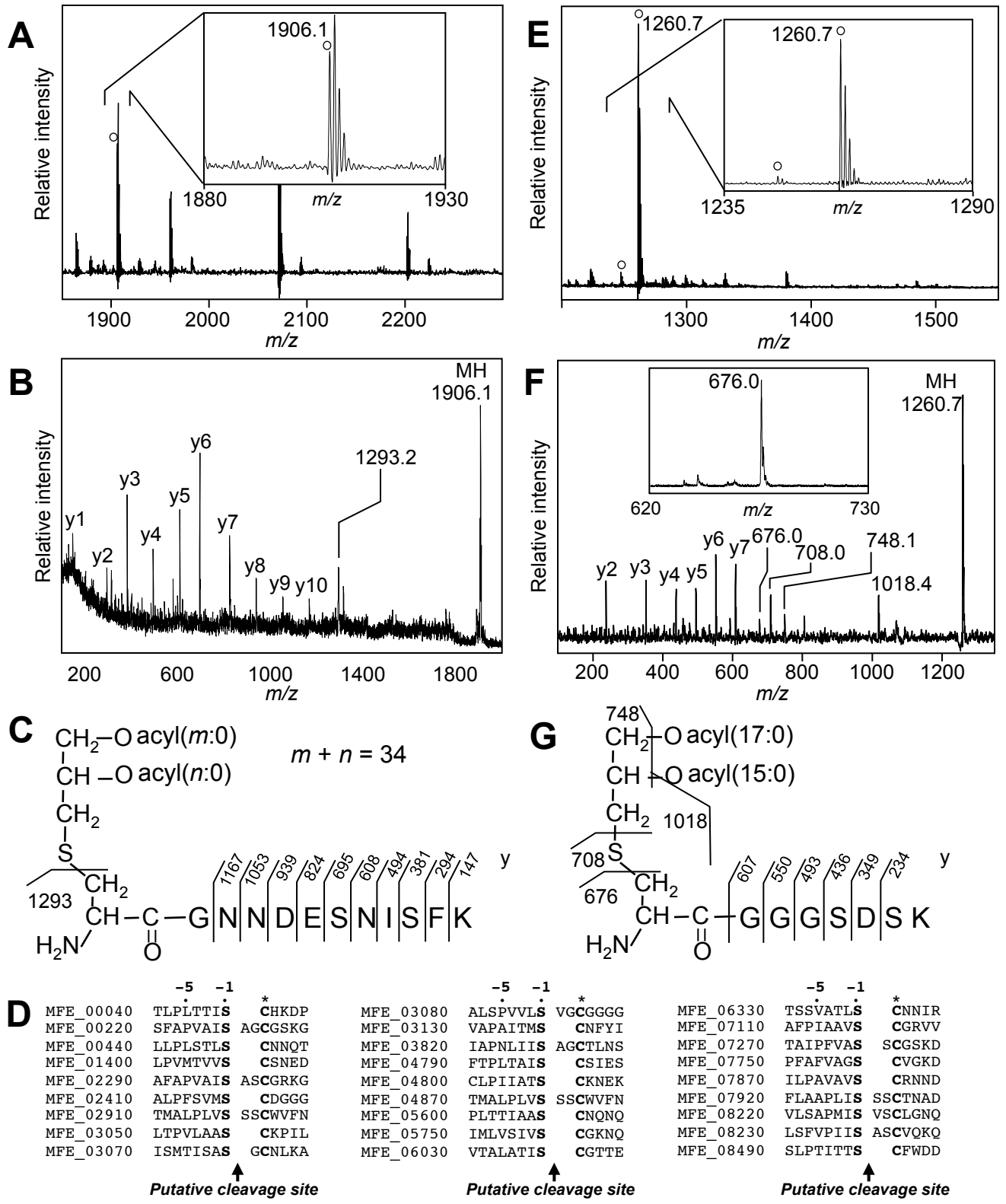

Figure S7. Conventional diacyl structure from *Mycoplasma fermentans* and *Listeria monocytogenes*.
(*A*-*C*) Analysis of MALP2 precursor of *M. fermentans*. Shown are MS spectrum of a fraction containing in-gel-digested MALP2 precursor of *M. fermentans* (*A*), MS/MS spectrum of an ion obtained at *m*/*z* 1906.1 (*B*), and elucidated conventional diacyl structure of *M. fermentans* MALP2 precursor (*C*). The C-terminus-containing y-series ions and a product ion generated by the neutral loss of diacyl(34:0)thioglycerol are highlighted.
(*D*) Putative cleavage sites of 27 lipoproteins from *M. fermentans* strain JER. Protein ID of strain JER. The predicted lipid-modified cysteine is marked with an asterisk.
(*E*-*G*) Analysis of Lmo2196 lipoprotein of *L. monocytogenes*. Shown are MS spectrum of a fraction containing in-gel-digested Lmo2196 of *L. monocytogenes* (*E*), MS/MS spectra of an ion obtained at *m*/*z* 1260.7 (*F*) or of the ion after oxidation (*F*, insert), and elucidated conventional diacyl structure of *L. monocytogenes* Lmo2196 (*G*). The y-series ions and product ions that have lost a 15:0 fatty acid, 17:0 fatty acid, two fatty acids (17:0 and 15:0), or diacyl(17:0/15:0) thioglycerol are highlighted. The insert in (*A*) or (*E*) is the magnified view. The open circle indicates the mono-isotopic lipopeptide ion containing two saturated fatty acids.