Giant DNA Virus Mimivirus Encodes Pathway for Biosynthesis of Unusual Sugar 4-Amino-4,6-dideoxy-β-glucose (Viosamine)*§

Received for publication, October 25, 2011, and in revised form, December 7, 2011 Published, JBC Papers in Press, December 8, 2011, DOI 10.1074/jbc.M111.314559

Francesco Piacente¹, Margherita Marin², Antonio Molinaro³, Cristina De Castro⁴, Virginie Seltzer⁴, Annalisa Salis⁵, Gianluca Damonte⁶, Cinzia Bernardi⁶, Jean-Michel Claverie⁶, Chantal Abergel⁷, and Michela Tonetti¹²

From the ¹Department of Experimental Medicine and Center of Excellence for Biomedical Research, University of Genova; Viale Benedetto XV/1, 16132 Genova, Italy, the ²Department of Organic Chemistry and Biochemistry, University of Napoli "Federico II," Napoli, Italy, and ³Information Génomique et Structurale, CNRS-UPR 2589, Aix-Marseille University, Institut de Microbiologie de la Méditerranée, Parc Scientifique de Luminy, Case 934, FR-13288 Marseille, France

Background: Mimivirus is highly glycosylated; however, nothing is known about its glycan composition and structure.

Results: We identified a Mimivirus UDP-viosamine synthetic pathway, and we determined the sugar composition of viral fibers.

Conclusion: Our data give further support to the presence of a Mimivirus-encoded glycosylation machinery.

Significance: These results contribute to shed light on the origin of viral glycosylation systems.

Mimivirus is one of the largest DNA viruses identified so far, infecting several Acanthamoeba species. Analysis of its genome revealed the presence of a nine-gene cluster containing genes potentially involved in glycan formation. All of these genes are co-expressed at late stages of infection, suggesting their role in the formation of the long fibers covering the viral surface. Among them, we identified the L136 gene as a pyridoxal phosphate-dependent sugar aminotransferase. This enzyme was shown to catalyze the formation of UDP-4-amino-4,6-dideoxy-β-glucose (UDP-viosamine) from UDP-4-keto-6-deoxy-β-glucose, a key compound involved also in the biosynthesis of L-rhamnose. This finding further supports the hypothesis that Mimivirus encodes a glycosylation system that is completely independent of the amoebal host. Viosamine, together with rhamnose, (N-acetyl)glucosamine, and glucose, was found as a major component of the viral glycans. Most of the sugars were associated with the fibers, confirming a capsular-like nature of the viral surface. Phylogenetic analysis clearly indicated that L136 was not a recent acquisition from bacteria through horizontal gene transfer, but it was acquired very early during evolution. Implications for the origin of the glycosylation machinery in giant DNA virus are also discussed.

Mimivirus is a member of the nucleocyttoplasmic large DNA virus group (1) and one of the largest viruses ever described, with a diameter of 700 nm and a 1.2-Mbp genome encoding more than 1000 genes (2, 3). The large size and the Gram-positive staining of the viral particles led to its initial misidentification as a small intracellular bacterium, named Bradford coccus. It was later recognized as a giant virus with a typical icosahedral capsid surrounded by long fibers ∼120 nm in length, supposed to be highly glycosylated, this explaining the Gram staining (4–7). The Mimivirus mode of infection is unique because it penetrates the phagocytic cells using internalization vacuoles. Later, its nucleocapsid is transferred into its host cytoplasm through the opening of a specific structure, the stargate, and the fusion of one of the two virion internal membranes with the vacuole one (8). From the very beginning, it was predicted that the glycosylated fibers could be of a capsular nature and used by the virion to promote the phagocytosis of host cells by mimicking bacterial polysaccharide structures. They can also contribute to the increased virion resistance to proteases and other environmental stresses. A detailed analysis of the composition and structure of Mimivirus glycans is thus needed to understand its mode of infection.

On a more general level, it is now becoming increasingly obvious that, in contrast to other viruses, some members of the nucleocyttoplasmic large DNA virus encode at least part, if not all, of the enzymes required for the glycosylation of their structural proteins. For instance, Chlorella viruses (Phycodnaviridae), which infect unicellular green algae, encode functional enzymes involved in the production of modified nucleotide sugars, glycosyltransferases, and glycosidases (9–12). Indeed, much evidence has indicated that the glycosylation of their major capsid occurs independently of the algal host endoplasmic reticulum/Golgi machinery (13). Interestingly, Mimivirus genome analysis revealed many genes potentially encoding proteins involved in glycan formation, including glycosyltransferases and sugar-modifying enzymes (2). This strongly suggests that this virus also encodes an independent glycosylation system. We have previously demonstrated that two Mimivirus genes, one being part of this cluster, encode two functional enzymes required for the biosynthesis of UDP-L-Rha³ (14). Putative enzymes involved in the synthesis of 3-deoxy-D-
Viosamine Synthesis in Mimivirus

manno-octulosonate were also found in the recently sequenced marine giant virus CroV, which infects the microflagellate grazer *Cafeteria roenbergensis* (15).

In this study, we characterized the Mimivirus *L136* gene as a new enzyme responsible for the production of the unusual monosaccharide 4-amino-4,6-dideoxy-\(\alpha\)-glucopyranose (vios-amine (Vio)) or 4-aminoquinovose. Vio is found in several bacterial glycosans, such as the LPS O-antigens of *Shigella dysenteriae* type 7 and of the Shiga toxin-producing *Escherichia coli* O121 (16), the fiber-associated pentasaccharide of *Bacillus anthracis* exosporium (17), the O-polysaccharide of *Francisella tularen- sis* and of the emerging pathogen * Photonabidus asymbiotica* (18), the glycols of *Pseudomonas syringae* flagellin (19), and the capsular polysaccharides of several marine bacteria belonging to the genus *Pseudoalteromonas* and *Shewanella* (20–21). We also analyzed the sugar composition of Mimivirus by GC-MS and were able to demonstrate that, together with Vio, major components of the viral glycosans are Rha, Glc, and (N-acetyl-)glucosamine (GlcN(Ac)). With the exception of Glc, most of the sugars were found associated with the fibers covering the viral capsids, thus confirming the capsular-like nature of Mimivirus fibers.

**EXPERIMENTAL PROCEDURES**

*L136 Sequence and Phylogenetic Analyses*—The most similar homologues of the Mimivirus *L136* PLP-dependent transaminase were first identified using the Blast-Explorer tool (22) on the Phylogeny.fr server (23). The environmental sequences were specifically retrieved using the NCBI server (24) and the “env_nr” database (25). A subset of sequences was selected based on the alignment quality (preserving enough informative positions). An optimal multiple alignment was then computed using MAFFT version 6 (26) on the CBRC-AIST server. The neighbor-joining tree was computed from 234 unzipped positions using the JTT default substitution model.

For structural comparison of PLP-dependent aminotransferases, reference structures were retrieved from the Protein Data Bank (27). The *L136* model was downloaded from the ModBase server (28). The *L136* model and the reference structures were superimposed using the Coot software (29), and the biggest ligand coordinates were retrieved and inserted in all coordinate files to identify in all structures the contacting residues less than 3.5 Å from it.

*Mimivirus 9 Gene Cluster Expression Profiles*—The Mimivirus transcriptome was performed by RNA-seq as described previously (30). Briefly, cells were infected by Mimivirus with a multiplicity of infection of 1000 to obtain synchronization. Total RNA was extracted from infected cells collected every hour from the beginning of the infectious cycle until 7 h to sample efficiently the early and intermediate phase of the infection (31). A last point was collected at 11 h. We thus generated nine barcoded transcriptome libraries from 1 mmol/L Tris-buffered saline (50 mM Tris-HCl, pH 7.5, 150 mM NaCl) containing 1 mM DTT (TBSD). The absorbance spectrum of the purified protein was obtained using a Beckman Coulter DU800 spectrophotometer, and concentration was estimated by absorbance at 280 nm using a computed extinction coefficient of 36330 M\(^{-1}\) cm\(^{-1}\) (32). Protein purity, determined by SDS-PAGE, exceeded 95% in all preparations.

**Enzymatic Activity Assays**—The *L136* enzymatic activity was assayed in TBSD at 25 °C using 0.025–1.5 mM UDP(dTDP)-4-keto-6-deoxy-Glc and 1–10 mM glutamate as substrates. Unless otherwise indicated, all chemicals were obtained from Sigma-Aldrich. UDP(dTDP)-4-keto-6-deoxy-\(\alpha\)-Glc was produced from UDP-\(\beta\)-Glc and dTDP-\(\alpha\)-Glc by ATCV-1 and Mimivirus UDP-\(\beta\)-Glc 4,6-dehydratase (UDG) (14). The reaction was stopped by heat inactivation for 3 min at 80 °C, and the denatured enzyme was removed by ultrafiltration (Microcon YM10, Millipore). *L136* enzymatic activity was determined in a discontinuous assay. Aliquots were removed at each time point, and the reaction was immediately stopped by heating at 80 °C for 3 min. After clarification, the conversion of UDP(dTDP)-4-keto-6-deoxy-\(\alpha\)-Glc to the product was determined by ion exchange HPLC, as described (9). The formation of \(\alpha\)-ketoglutarate derived from the transamination of glutamate was followed by a coupled assay with glutamate dehydrogenase by monitoring the disappearance of NADH at 340 nm. To monitor the reverse reaction, the purified UDP(dTDP)-4-amino-6-deoxyhexose produced by *L136* was purified and reacted again with *L136* in the presence of \(\alpha\)-ketoglutarate. Kinetic parameters were determined by the Michaelis–Menten equation using nonlinear regression (GraphPad Prism).

**Purification of L136 Product**—*L136* product was purified by anion exchange HPLC using the described procedure (9) and then subjected to solid phase extraction using Carbograph ultraclean columns (150 mg/4 ml; Alltech). The solid phase extraction columns were pretreated with 3 ml of 60% acetonitrile/H\(_2\)O containing 0.3% ammonium formate (pH 9), followed by one wash with 2 ml of H\(_2\)O and one with 3 ml of 250 mM NH\(_4\)HCO\(_3\). The HPLC-purified compound was then applied to the solid phase extraction column. After a wash with 3 ml of H\(_2\)O, *L136* product was eluted using 1 ml of 60% acetonitrile/H\(_2\)O containing 0.3% ammonium formate and dried under vacuum. The amount of the UDP-sugar recovered after
purification was determined by UV absorbance using $\epsilon_{262} = 10,000 \, \text{m}^{-1} \, \text{cm}^{-1}$ for UTP.

**Structural Characterization of L136 Product—**Electrospray ionization MS analysis was performed as described (14). The structural assignment of the UDP-4-amino-6-deoxyhexosamine was obtained by NMR spectroscopy. One- and two-dimensional NMR spectra were recorded on a Bruker 600 DRX equipped with a cryoprobe on a solution of 500 $\mu$L of D$_2$O. Double quantum-filtered phase-sensitive COSY experiments were performed using data sets of 2048 × 512 points (33, 34); the data matrix was zero-filled in both dimensions to give a matrix of 4K × 2K points and was resolution-enhanced in both dimensions by a cosine-bell function before Fourier transformation. Coupling constants were determined on a first order basis from high resolution one-dimensional spectra or by two-dimensional phase-sensitive DQF-COSY. Heteronuclear single quantum correlation spectroscopy and heteronuclear multiple-quantum correlation spectroscopy were measured in the $^1$H-detected mode via single quantum coherence with proton decoupling in the $^{13}$C domain, using data sets of 2048 × 512 points. Experiments were carried out in the phase-sensitive mode (35), and the data matrix was extended to 2048 × 1024 points using forward linear prediction extrapolation.

**Analysis of Mimivirus Monosaccharide Composition—**Mimivirus was purified from infected Acanthamoeba castellanii culture as described (31). Briefly, the virus was collected by 10 min of centrifugation at 15,000 × g and 4 °C and resuspended into PBS with CsCl to obtain a density of 1.15. 20 ml of the suspension were layered onto a discontinuous gradient of CsCl (1.45/1.25) and centrifuged at 18,000 $\times g$ for 40 min. The virus band was collected and washed/centrifuged 5 times with 50 mM Tris, pH 8.0, storage buffer. Removal of fibers was achieved using the published procedure (5), and the treated sample was controlled by electron microscopy. Approximately 0.5–1 × 10$^{11}$ viral particles were used for each analysis. Briefly, 100 mM DTT was added to 200 $\mu$L of virus suspensions; the tubes were then heated at 99 °C for 15 min. After cooling, 5 units of Benzonase were added and incubated for 60 min at 37 °C to degrade DNA. Proteins were then precipitated using 5 volumes of acetone at −20 °C overnight. After removal of acetone, the samples were resuspended in H$_2$O and briefly sonicated in order to disrupt protein aggregates. Hydrolysis was performed with 2 $\mu$L trifluoroacetic acid (TFA) at 121 °C for 2 h under N$_2$ with continuous stirring (36). Parallel experiments also with 4 $\mu$L TFA gave comparable sugar recoveries. TFA was then removed under vacuum, and samples were subjected to three washes with methanol. Alditols were obtained by the addition of 300 $\mu$L of 0.25 $M$ NaBH$_4$ dissolved in 1 $mM$ NH$_4$OH, followed by incubation at 40 °C for 2 h. The reaction was stopped with 30% acetic acid in methanol, and methyl borates were removed by four washes with methanol/acetic acid (200:1). Acetylation was carried on with 200 $\mu$L of acetic anhydride/pyridine (1:1) for 1 h at 80 °C, followed by drying and postderivatization cleaning four times, using H$_2$O/ethyl acetate (1:0.5). Samples resuspended in ethyl acetate were injected in a HP5890 series II gas chromatograph coupled to a HP5889A mass spectrometer equipped with an electron impact ionization source (Hewlett-Packard). Separation was performed on an SE54 capillary column (0.32 mm × 30 m; Alltech); the helium gas flow was 2 ml/min. The oven temperature gradient was as follows: initial temperature of 170 °C, isothermal for 4 min, 170–280 °C (rate, 5 °C/min), and isothermal to 36 min. The MS analysis was performed in full-scan mode.

Sugar quantification was done by comparing the area of sample peaks with that obtained for standard sugars, using selected ions. The most abundant ions were used for each monosaccharide; m/z 115 was used for 6-deoxyhexoses, pentoses, and hexoses and m/z 84 for the amino sugars. Response proved to be linear in the range of the sugar concentrations used for the analysis.

**RESULTS**

**Analysis of L136 Sequence—**We previously characterized the Mimivirus R141 as a dehydratase and L780 as an epimerase/reductase, both involved in the synthesis of UDP-$\alpha$-Rha (14). We then searched the viral genome for other putative enzymes able to catalyze the formation of modified monosaccharides and identified a nine-gene cluster, including R141 (Fig. 1A). L136, the second in this cluster, has been annotated as a PLP-dependent aminotransferase and exhibited a weak homology with bacterial sugar aminotransferases belonging to the DegT/Dnr/EryC1/Strs family. The best BLAST hits correspond to a putative sugar aminotransferase from F. tularensis SCHU S4 (29% identity over 361 amino acids) and from Shewanella sediminis HAW-EB3 (27% identity over 359 amino acids). Identity with B. anthracis AntC, which produces dTDP-4-amino-4,6-dideoxy-Glc in the anthrose biosynthetic pathway (17), was 26% over 250 amino acids. E. coli and Salmonella enterica WecE, which are involved in the synthesis of $\alpha$-fucosamine found in the enterobacterial common antigen (37), were 25% identical with L136 over 354 and 227 amino acids, respectively. Finally, the BLAST search against the environmental sequence database revealed closer homologues (up to 46% identity over 354 amino acids) with hypothetical proteins from marine metagenomic samples (25). To obtain some information on the L136 specificity, we retrieved its ModBase model (28), based on the Caulobacter crescentus perosamine synthase structure (Protein Data Bank codes 3DR4 or 3DR7) (38) with whom it shares 26% identity over 236 amino acids. We then compared the model with the reference structure of PLP-dependent aminotransferase from Campylobacter jejuni (Protein Data Bank code 1O61) (39) and the Desi PLP-dependent aminotransferase from Streptomyces venezuelae (Protein Data Bank code 2PO3) (40). The multiple alignment of L136 with the bacterial enzymes and the reference structures is reported in supplemental Fig. S1. All residues that have been identified in transamases and are involved in PLP binding and in the catalytic mechanism (40) are all well conserved in the L136 sequence. The ligand (T4K) of S. venezuelae aminotransferase 2PO3 (40) was used to measure contacting residues less than 3.5 Å from T4K in all structures. They are highlighted in supplemental Fig. S1.

The phylogenetic tree built using the L136 sequence, reference bacterial sequences, and selected environmental sequences exhibited two main clusters (supplemental Fig. S2). One includes bacterial sequences from different classes (cluster...
Viosamine Synthesis in Mimivirus

A

Mimivirus

\[
\begin{array}{ccc}
\text{R135} & \text{L136} & \text{L137} \\
\text{L138} & \text{R139} & \text{L140} \\
\text{R141} & \text{L142} & \text{L143}
\end{array}
\]

Megavirus

\[
\begin{array}{c}
\text{mg878} \\
\text{mg538}
\end{array}
\]

M4

\[
\begin{array}{c}
\text{R135} \\
L142a, L142b, L143
\end{array}
\]

B

FIGURE 1. The Mimivirus nine-gene cluster. A, conservation of the Mimivirus nine-gene cluster in Megavirus and in the Mimivirus isolate M4. The arrows indicate the coding strand. The only two conserved genes are at the two edges of the cluster. B, Mimivirus gene expression profiles over the infection cycle. The three expression classes are color-coded (yellow for early, blue for intermediate, and gray for late). The early class of expression mostly corresponds to unknown function genes, the intermediate class corresponds to genes involved in DNA replication, and the late class of expression corresponds to genes involved in capsid production (structural proteins and proteins involved in fibril formation) and to genes encoding proteins that are associated to the capsid, such as the viral transcription machinery. L140 and L143 genes were co-expressed all along the intermediate and late infection cycle; L138, R139, L142, and L780 were expressed 1 h later until the end of the infection; L137, R141, and L136 were co-expressed at 4 h until the end; and R135 was the latest to be expressed, until the end of the infection cycle (see the Mimivirus Genome Browser Web site). The putative functions of these genes are indicated under “Results.” L780 is not part of the nine-gene cluster; however, it has been reported because it catalyzes the last step of UDP-L-rhamnose synthesis.

2), whereas cluster 1 includes the L136 sequence, environmental sequences, and one of the two S. sediminis paralogous sequences of PLP-dependent transaminases, the other one belonging to cluster 2. We selected the environmental sequences from contigs encompassing more than one gene in order to obtain some hints about their putative origins using a direct BLAST search against the Refseq database. With this approach, we were able to discriminate between possible viral or prokaryotic sequences. Details are reported in supplemental Fig. S2. Results from the phylogenetic analysis strongly suggest that the Mimivirus L136 sequence is ancestral and, if originating from bacteria, was acquired early in evolution.

L136 is encoded in a nine-gene cluster including several putative enzymes involved in glycan formation (Fig. 1A). In particular, R135 has been proposed to be a component of Mimivirus fibers and proved to be glycosylated and to be involved in antigenic response (5, 41). L137, R139, L142, and L780 were expressed 1 h later until the end of the infection; L137, R141, and L136 were co-expressed at 4 h until the end; and R135 was the latest to be expressed, until the end of the infection cycle (see the Mimivirus Genome Browser Web site). The putative functions of these genes are indicated under “Results.” L780 is not part of the nine-gene cluster; however, it has been reported because it catalyzes the last step of UDP-L-rhamnose synthesis.

Biochemical Characterization of L136 Protein—We hypothesized that the L136 protein could catalyze the transfer of an amino group from an amino acid donor to the C4 of UDP-4-keto-6-deoxy-β-Glc (Fig. 2). This latter compound is formed by UGD as an intermediate in the pathway leading to the production of UDP-1-Rha (14). Approximately 2 mg of pure L136 protein per liter of bacterial culture were obtained after GST cleavage. UV spectral analysis of the protein revealed an additional peak of light absorbance with a maximum at 330 nm, which is consistent with the presence of the cofactor mainly as pyridoxamine phosphate (supplemental Fig. S3) (46).

To verify the activity of the L136 protein, we incubated the enzyme with UDP-4-keto-6-deoxy-β-Glc in the presence of glutamate and observed the formation of a new compound by anion exchange HPLC (Fig. 3, peak B). Formation of peak B was not observed when L136 was heat-inactivated before incubation (not shown). The aminotransferase activity was also confirmed by monitoring α-ketoglutarate formation in the reaction mixture, using a glutamate dehydrogenase-coupled assay. The enzyme exhibited a $k_{\text{cat}}$ of 0.9 ± 0.06 s$^{-1}$ and a $K_{\text{m}}$ of 44.3 μM for UDP-4-keto-6-deoxy-β-Glc. dtUDP-4-keto-6-deoxy-β-Glc could also be used as a substrate, but the resulting activity was clearly reduced (about one-third compared with the UDP-bound substrate). The higher affinity of the L136 protein for the UDP-bound substrate is consistent with the preference of
Mimivirus UGD for UDP-\(\beta\)-Glc (14). At variance with bacterial sugar aminotransferases (16), when glutamate was substituted with glutamine as the amino group donor, the activity was negligible. No effects were observed after the addition of different concentrations of PLP, confirming that the co-factor remained tightly bound to the enzyme during the purification procedure, as already suggested by the UV spectral analysis of the protein. We addressed the possibility that L136 could also perform the reverse reaction, converting the UDP-amino-6-deoxysugar into the UDP-4-keto-6-deoxy compound. We thus purified the UDP-6-deoxyhexosamine product and assayed the L136 enzymatic activity in the presence of \(\alpha\)-ketoglutarate. No activity was detected as the production of either UDP-4-keto-6-deoxy-\(\beta\)-Glc or glutamate, demonstrating that the sugar transamination is irreversible.

**Structural Characterization of UDP-4-keto-6-deoxyhexosamine**—HPLC analysis indicated that L136 was able to catalyze the transamination of UDP-4-keto-6-deoxy-\(\beta\)-Glc with the expected formation of a UDP-4-amino-4,6-dideoxyhexose. Actually, electrospray ionization MS indicated an \(m/z\) (mass/charge) of 548 for the L136 product, consistent with the expected formation of a UDP-4-amino-4,6-dideoxyhexose. We addressed the possibility that L136 could also perform the reverse reaction, converting the UDP-amino-6-deoxysugar into the UDP-4-keto-6-deoxy compound. We thus purified the UDP-6-deoxyhexosamine product and assayed the L136 enzymatic activity in the presence of \(\alpha\)-ketoglutarate. No activity was detected as the production of either UDP-4-keto-6-deoxy-\(\beta\)-Glc or glutamate, demonstrating that the sugar transamination is irreversible.

**FIGURE 2. Proposed pathway for UDP-Vio synthesis in Mimivirus**. The pathway utilizes UDP-4-keto-6-deoxy-\(\alpha\)-Glc, the intermediate in UDP-\(\alpha\)-Rha synthesis, and glutamate to produce UDP-\(\alpha\)-Vio. UGD, encoded by Mimivirus R141, catalyzes the first step in this pathway.

**FIGURE 3. Anion exchange HPLC analysis of nucleotide sugars**. UDP-4-keto-6-deoxy-\(\alpha\)-Glc (peak A) was incubated with glutamate in the presence of L136 protein at 25 °C. The progressive formation of a new compound (peak B) was observed. The chromatogram was obtained after a 15-min incubation.

NMR experiments (DQF-COSY and \(^1\)H-\(^13\)C heteronuclear single quantum correlation spectroscopy (Fig. 4, B and C) and heteronuclear multiple-bond correlation spectroscopy) was executed to assign the three spin systems of the molecule: uracil (U), ribose (R), and the sugar residue (V). In fact, \(^1\)H and \(^13\)C chemical shifts for uracil and ribose were in good agreement with the literature (34), whereas sugar residue V was identified as Vio on the basis of its \(^1\)H and \(^13\)C chemical shifts (supplemental Table S1) and \(J_{1H,1H}\) vicinal coupling constants (47). In agreement, with the exception of the anomeric proton signal, which is on a \(\alpha\)-configured carbon atom (signal \(V_1\) in Fig. 4A), each ring proton displayed a large coupling constant (about 10 Hz) with the neighboring protons, a clear indication that these protons occupy the axial position of the pyranose sugar ring. Therefore, the Vio residue monosaccharide possesses the \textit{gluco} stereochemistry (i.e. all substituents on the ring are oriented in the equatorial position as is the case of Vio). Additionally, C-4 chemical shift (58.2 ppm) of the V residue indicated that this was a nitrogen-bearing carbon atom; this information completed the identification of this monosaccharide as Vio.

**Sugar Composition of Mimivirus Glycans**—Neutral sugar and amino sugar presence in viral particles was analyzed by GC-MS. Following acid hydrolysis, the free monosaccharides were converted to alditol acetates. Clear identification of Mimivirus sugars was obtained by comparing the retention times and the fragmentation spectra with standard monosaccharides. Retention times of standard sugars and of the viral samples are reported in supplemental Table S2. The main components of Mimivirus glycans were Rha, Glc, and GlcN(Ac) (Fig. 5). Low amounts of ribose, arabinose, xylose, mannose, and galactose were also detected. Two peaks corresponding to unknown sugars were also contributing to the Mimivirus glycan composition (indicated as \textit{peaks I} and \textit{2} in Fig. 5). Given the enzymatic activity of the L136 protein and the identification of its product as a Vio, we ran a control using this compound. As already reported, Vio was not substantially decomposed during acid hydrolysis (48). The Vio peak corresponded to a retention time of 20.7 min (Fig. 6A), which was also observed in the virus sample chromatogram (Fig. 6C, peak 2); in agreement, the ion peak in the virus sample has a fragmentation pattern identical to the Vio standard (Fig. 6, B and D), definitely suggesting that Vio is a component of Mimivirus glycans. Another peak with a retention time of 18.2 min could not be assigned to any available sugar standards. However, the fragmentation spectrum indicated the presence of an amino sugar, and, in particular, ion fragments were highly suggestive of the presence of Vio modified with a 3-O-methyl group (supplemental Fig. S5). This hypothesis needs to...
be confirmed by the production of a standard sugar and NMR analysis. The modified bacterial amino sugars muramic acid and bacillosamine were not found, excluding the presence of typical bacterial peptidoglycan or of bacteria-like N-linked glycans (49).

Because it has been postulated that the Mimivirus fibers are highly glycosylated, we have also analyzed the sugar content of the defibered viral particles. The electron microscopy analysis of Mimivirus particles confirmed that the treatment (5) led to a majority of bald particles (>80%). As a consequence, we measured a significant reduction (about 90%) of Rha, Vio, GlcN(Ac), and the unidentified sugar, whereas the decrease in Glc and in the other neutral sugars was less pronounced (Fig. 7). This demonstrates that most of the Mimivirus glycans are indeed associated with the fibers.

As indicated previously in Fig. 1A, the nine-gene cluster found in Mimivirus is not present in Megavirus. Megavirus also lacks a homologue of the Mimivirus L780 gene, which encodes the epimerase/reductase involved in the last step of UDP-L-Rha synthesis. These differences might be at the origin of the distinct appearances of the fiber layers covering the Megavirus and Mimivirus particles. We thus performed a comparison of the sugar composition of Megavirus with Mimivirus using GC-MS (supplemental Fig. S6). As expected from the comparison of their gene contents, their glycan composition appeared completely different. The Megavirus lacks Rha, Vio, and the puta-
we provided the first characterization of neutral and amino fiber-associated oligo(poly)saccharides, and indeed, also demonstrated that Vio is one of the main components of rus encodes a host-independent glycosylation machinery. We (37). This finding further supports the hypothesis that Mimivirus fiber glycan composition. Because Vio often occurs in bacterial polysaccharides, generally carrying N-acylation. In the O-antigen of S. disenteriae type 7 and of the E. coli O121 and in the O-antigen of the emerging human pathogen Photobacterium asymbiotica, the amino group of Vio can be acylated by an acetyl or an acetyl-glycyl group (16, 18), whereas F. tularensis LPS O-antigen contains a formyl group (18). In B. anthracis spores, Vio is further modified to obtain 2-O-methy-L-4-(3-hydroxy-3-methyl butanamido)-4,6-dideoxy-D-Glc, which is named anthrose (17). This sugar, together with N-acylation, also contains a 2-O-methyl group. The function of this substituent might serve to block further chain extension. Anthrose is a component of the pentasaccharide that modifies the collagen-like BclA protein of the exosporium external hairlike nap (51). Serological cross-reactivity between sera of patients affected by tularemia and Mimivirus has been observed (41), it would be interesting to verify if this cross-reactivity could be also ascribed to the presence of Vio.

We cannot presently confirm that the Mimivirus glycans also contain modified forms of Vio (i.e. presenting N-acylation and/or O-methylation). The L136 gene is located in a Mimivirus genomic region containing several other genes possibly involved in glycan formation (Fig. 1A). These include the N-terminal region of L142, which could be involved in the acylation of the sugar backbone. Putative O- and N-acyl(acetyl)transferases are often strictly associated with the aminotransferase in bacterial Vio gene clusters (i.e. VioB is localized immediately downstream of the aminotransferase VioA in E. coli O7, and it has been shown to catalyze the formation of an acetamido group on C-4 of the 6-deoxy sugar) (16). In addition, L143 encodes a putative exonuclease V-like ketal pyruvate transferase, which has been involved in the formation of acidic exopolysaccharides in symbiotic and environmental bacteria and in human pathogens (44, 52, 53). Moreover, putative O-methyltransferases are also contained in the Mimivirus genome (2). Cloning and expression of these proteins, which could be involved in Vio modification, are currently under way, and they will help us to identify the presence of derivatives of this sugar in Mimivirus glycans. All of the genes included in this genomic
region are expressed late during the viral infection, as already observed for the enzymes involved in the UDP-\(\text{L-}\)Rha biosynthetic pathway (14). Because fibrils are added at the last stage of viral particle formation, this finding is in agreement with a role for all these proteins in fibrillogenesis.

L136 belongs to the aspartate aminotransferase fold type I (AAT_I) superfamily of PLP-dependent enzymes and in particular to the DegT/DnrJ/EryC1/StrS family (pfam01041). Members of this family are involved in the transfer of amino groups to several different substrates to obtain modified amino sugars, which are frequently found in bacterial polysaccharides or in antibiotics, such as erythromycin and other macrolides. L136 displays relatively low sequence homology with the characterized enzymes of this family, in particular with VioA from \textit{E. coli} and DesI from \textit{S. venezuelae}, which catalyze a similar reaction, from dTDP-4-keto-6-deoxy-D-Glc to dTDP-Vio (16, 40). However, all of the residues previously recognized as essential for PLP binding are conserved in L136. This is also confirmed by comparison of the L136 model with the published structures of sugar transaminases.

The phylogenetic analysis indicates with strong support that the L136 gene was not acquired from a bacteria by a recent horizontal transfer. However, the L136 sequence is even more distant from eukaryotic homologues. Thus, its origin, as well as the one of its closest homologues found in the environmental data set, remains unknown, as for the many “virus-only” genes found in Mimivirus. On the other hand, the same genomic cluster encodes R141, which catalyzes the first step of Vio biosynthesis and, together with L780, also leads to L-Rha. Both R141 and L780 exhibit a clear affinity with protist homologues, suggesting a lateral gene transfer from a cellular host to Mimivirus (14). A mixture of enzymes of unknown, eukaryotic, or bacterial origins appears to constitute a Mimivirus-encoded UDP-\(N\)-acetylglucosamine biosynthetic pathway\(^4\) as well as the GDP-\(\text{L-}\)fucose biosynthetic pathways in Chlorella viruses (9, 10). This could be explained by the co-existence of genes from a very ancestral glycosylation machinery predating the emergence of the eukarya, more recently complemented or replaced by genes of extant protozoans and bacteria. The recently discovered

\(^4\) M. Tonetti, F. Piacente, and C. Abergel, unpublished results.
Megavirus (45) supports this hypothesis by suggesting that giant viruses, most probably derived from an ancestral cellular genome, might have conserved genes from an ancestral glycosylation machinery. The subsequent process of reductive evolution and lineage-specific losses, combined with episodic lateral gene transfer (including non-orthologous replacements), would then have resulted in the complex picture we are seeing today. Thus, the study of the giant DNA virus-encoded glycosyltransferases as well as the characterization of their glycan structures might turn out to be of general interest in probing the origin and ancient evolution of today’s cellular glycosylation pathways.

Acknowledgments—We thank Dr. Martin Young (Institute for Biological Sciences, National Research Council of Canada) for providing the bacillosamine standard, Prof. Antonio De Flora and Prof. Umberto Benatti (University of Genova) for helpful discussion, and Dr. Matthieu Legendre for the transcriptions data. We acknowledge the use of the PACA-Bioinf (IBISA) bioinformatic platform.

REFERENCES

1. Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006) Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 117, 156–184
2. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scala, B., Suzan, M., and Claverie, J. M. (2004) The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350
3. Claverie, J. M., and Abergel, C. (2010) Mimivirus. The emerging paradox of quasi-autonomous viruses. Trends Genet. 26, 431–437
4. Renesto, P., Abergel, C., Decloquement, P., Moinier, D., Azza, S., Ogata, H., Fourquet, P., Gorvel, J. P., and Claverie, J. M. (2006) Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. J. Virol. 80, 11678–11685
5. Xiao, C., Kuznetsov, Y. G., Sun, S., Hafenstein, S. L., Kostyuchenko, V. A., Chipman, P. R., Suzan-Monti, M., Raoult, D., Mcpherson, A., and Rossman, M. G. (2009) Structural studies of the giant mimivirus. PLoS Biol. 7, e92
6. Kuznetsov, Y. G., Xiao, C., Sun, S., Raoult, D., Rossmann, M., and Mcpherson, A. (2010) Atomic force microscopy investigation of the giant mimivirus. Virology 404, 127–137
7. Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., Fournous, G., Borg, A., Robert, C., Zhang, X., Desnues, C., Henrisatt, B., Rossman, M. G., La Scala, B., and Raoult, D. (2011) Mimivirus shows dramatic genome reduction after intraamoebal culture. J. Virol. 85, 10296–10301
8. Nazarenko, E. L., Komandrovna, N. A., Grishkova, R. P., Tomshich, S. V., Zubkov, V. A., Kilcoyne, M., and Savage, A. V. (2003) Structures of polysaccharides and oligosaccharides of some Gram-negative marine proteobacteria. Carbohydr. Res. 338, 2449–2457
9. Leone, S., Silipo, A., Nazarenko, E. L., Lanzetta, R., Parrilli, M., and Molinaro, A. (2007) Molecular structure of endotoxins from Gram-negative marine bacteria. An update. Marine Drugs 5, 85–112
10. Dereeper, A., Audic, S., Claverie, J. M., and Blanc, G. (2010) BLAST-EXPLORER helps you building data sets for phylogenetic analysis. BMC Evol. Biol. 10, 8
11. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., and Gascuel, O. (2008) Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465–W469
12. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402
13. Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusc, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealon, K., White, O., Peterson, J., Hoff- man, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H., and Smith, H. O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74
14. Katoh, K., and Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298
15. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242
16. Pieper, U., Webb, B. M., Barkan, D. T., Schneidman-Duhovny, D., Schlessinger, A., Brasberg, H., Yang, Z., Meng, E. C., Pettersen, E. F., Huang, C. C., Datta, R. S., Sampaikumar, P., Madhusudhan, M. S., Sjölander, K., Ferrin, T. E., Burley, S. K., and Sali, A. (2011) ModBase. A database of annotated comparative protein structure models and associated resources. Nucleic Acids Res. 39, D465–D474
17. Emsley, P., and Cowtan, K. (2004) Coot. Model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
18. Legendre, M., Santini, S., Rico, A., Abergel, C., and Claverie, J. M. (2011) Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome
Viosamine Synthesis in Mimivirus

and transcriptome sequencing. Virol. J. 8, 99

31. Byrne, D., Grzela, R., Lartigue, A., Audic, S., Chenivesse, S., Encinas, S., Claverie, J. M., and Abergel, C. (2009) The polyadenylation site of Mimivirus transcripts obeys a stringent “hairpin rule”. Genome Res. 19, 1233–1242

32. Gasteiger, E., Haag, J., Ratsch, G., Bissantz, C., Engel, T., Baur, M., and Hofmann, I. L. (2008) Somatic antigens of Escherichia coli. J. Biol. Chem. 283, 1150–1155

33. Piantini, U., Sørensen, O. W., and Ernst, R. R. (1982) Multiple quantum filters for elucidating NMR coupling networks. J. Am. Chem. Soc. 104, 6800–6801

34. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., and Wüthrich, K. (1983) Improved spectral resolution in COSY 1H NMR spectra of proteins via double quantum filtering. Biochemistry 22, 6800–6801

35. States, D. J., Haberkorn, R. A., and Ruben, D. J. (1982) A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. J. Magn. Reson. 48, 286–292

36. Manzi, A. (2001) Acid hydrolysis for release of monosaccharides. Carbohydr. Res. 328, 1779–1802

37. Hwang, B. Y., Lee, H. J., Yang, Y. H., Joo, H. S., and Kim, B. G. (2004) Characterization and investigation of substrate specificity of the sugar aminotransferase WecE from Bacillus subtilis. J. Biol. Chem. 279, 43701–43709

38. Coutinho, P. M., Deleury, E., Davies, G. J., and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 328, 307–317

39. Lewis, A. L., Hensler, M. E., Varki, A., and Nizet, V. (2006) The group B streptococcal sialic acid O-acetyltransferase is encoded by neuD, a conserved component of bacterial sialic acid biosynthetic gene clusters. J. Biol. Chem. 281, 11186–11192

40. Arslan, D., Legendre, M., Seltzer, V., Abergel, C., and Claverie J. M. (2011) Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc. Natl. Acad. Sci. U.S.A. 108, 17486–17491

41. Dhara, J., Bousová, I., and Malon, P. (2005) Determination of quality of pyridoxal-5’-phosphate enzyme preparations by spectroscopic methods. J. Pharm. Biomed. Anal. 37, 1173–1177

42. Knirel, Y. A., Dashunin, V. V., Shashkov, A. S., Kochetkov, N. K., Dmitriev, B. A., and Hofman, I. L. (1988) Somatic antigens of Shigella. Structure of the O-specific polysaccharide chain of the Shigella dysenteriae type 7 lipopolysaccharide. Carbohydr. Res. 179, 51–60

43. Jann, B., and Jann, K. (1967) 4-Amino-4,6-dideoxyhexoses isolated from lipopolysaccharides of Escherichia coli. Eur. J. Biochem. 2, 26–31

44. Sharon, N. (2007) Celebrating the golden anniversary of the discovery of bacillosamine, the diamino sugar of a Bacillus. Glycobiology 17, 1150–1155

45. Luther, K. B., Hülsmeier, A. J., Schegg, B., Deuber, S. A., Raoul, D., and Hennet, T. (2011) Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. J. Biol. Chem. 286, 43701–43709

46. Kubler-Kielb, J., Vinogradov, E., Hu, H., Leplla, S. H., Robbins, J. B., and Schneerson, R. (2008) Saccharides cross-reactive with Bacillus anthracis spore glycoprotein as an anthrax vaccine component. Proc. Natl. Acad. Sci. U.S.A. 105, 8709–8712

47. Poli, A., Anzelmo, G., and Nicolaus, B. (2010) Bacterial exopolysaccharides from extreme marine habitats. Production, characterization, and biological activities. Marine Drugs 8, 1779–1802

48. Viosamine Synthesis in Mimivirus