The structural proteome of *Pseudomonas aeruginosa* bacteriophage φKMV

Rob Lavigne,1 Jean-Paul Noben,2 Kirsten Hertveldt,1 Pieter-Jan Ceyssens,1 Yves Briers,1 Debora Dumont,2 Bart Roucourt,1 Victor N. Krylov,3 Vadim V. Mesyanzhinov,1,4 Johan Robben2 and Guido Volckaert1

1Laboratory of Gene Technology, Katholieke Universiteit Leuven, Kasteelpark Arenberg 21, Leuven, B-3001, Belgium
2Biomedical Research Institute, Limburgs Universiteits Centrum and School of Life Sciences, University Hasselt, Diepenbeek, B-3590, Belgium
3State Institute for Genetics and Selection of Industrial Micro-organisms, 1st Dorozhnii proezd 1, Moscow, 113545 Russia
4Howard Hughes Medical Institute; Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Miklukho-Maklaya Street 16/10, Moscow, 117991 Russia

The structural proteome of φKMV, a lytic bacteriophage infecting *Pseudomonas aeruginosa*, was analysed using two approaches. In one approach, structural proteins of the phage were fractionated by SDS-PAGE for identification by liquid chromatography-mass spectrometry (LC-MS). In a second approach, a whole-phage shotgun analysis (WSA) was applied. WSA uses trypsin digestion of whole phage particles, followed by reversed-phase HPLC and gas-phase fractionation of the complex peptide mixture prior to MS. The results yield a comprehensive view of structure-related proteins in φKMV and suggest subtle structural differences from phage T7.

INTRODUCTION

Differential mass spectrometry (MS) techniques after 1-D or 2-D polyacrylamide protein fractionation have been applied to the identification and characterization of proteins in bacterial lysates (Goodacre et al., 1998; Zhang et al., 2004), and they have recently been applied to study viruses (Roberts et al., 2004). With the advent of high-throughput sequencing, providing over 300 complete genome sequences of phages infecting different bacterial hosts, MS protein identification allows comparative genomics based on experimentally determined structural proteins and proper genome sequence annotation.

Bacteriophage φKMV is a strongly lytic, T7-like bacteriophage infecting *Pseudomonas aeruginosa*. In silico analysis of the 42519 bp double-stranded DNA sequence reveals that this phage shares genome architecture with *Escherichia coli* phage T7 and contains 11 ORFs similar to those of the latter phage. Apart from the major capsid protein, which was identified by N-terminal protein sequencing, some putative structural proteins, including tail, connector and core proteins, have been hinted at by sequence similarity searches (Lavigne et al., 2003). In addition, phylogenetic analysis based on the putative RNA polymerase and on the DNA ligase revealed SP6 as the phage most closely related to φKMV (Lavigne et al., 2003, 2005). The completed nucleotide sequences of *Podoviridae* SP6 and K1-5 (phages infecting *Salmonella typhimurium* and *E. coli* K1-5, respectively) have allowed more in-depth comparisons (Dobbins et al., 2004; Scholl et al., 2004). These sequences and phage tail morphology seem to place SP6 and K1-5 in a distinct subgroup within the T7 supergroup. However, evidence such as the late localization of the φKMV RNA polymerase suggests a distinct replication strategy and gene regulation, compared to members of the SP6 group.

We used different approaches to analyse the structural proteome of φKMV by MS, to verify sequence annotation and to gain insights into the phage particle structure.

METHODS

Phage propagation and purification. Bacteriophage particles were purified using sequential step CsCl-gradient centrifugation as described by Sambrook & Russell (2001) in which φKMV has a buoyant density of about 1·50 g cm⁻³. Before disruption of the phage particles by a triple freeze–thawing cycle, the Microcon-procedure (Millipore) was used to concentrate the sample 12-fold.
To visualize bands, gels were silver-stained as described by Shevchenko et al. (1996a) and subsequently scanned and analysed for band size using Un-Scan-It gel v5.1 software (Silk Scientific Corporation), compared to a standard low-molecular-mass protein ladder (GE Healthcare). Subsequently, heat-denatured samples (5 min at 95 °C) were loaded onto a standard 12 % SDS-PAGE gel. Samples for tryptic treatment (Shevchenko et al., 1996b) and MS analysis were picked from Coomassie-stained gels (Simply Blue Safestain; Invitrogen).

Liquid chromatography-MS (LC-MS) and data analysis. Protein digests were analysed by electrospray ionization (ESI)-MS/MS on an LCQ Classic (ThermoFinnigan) equipped with a nano-LC column switching system. The trapped sample was separated over 55 min on an analytical column (Biosphere C18, 200 mm × 0.05 mm i.d., 5 μm; NanoSeparations) using a linear gradient from 5 to 60 % (v/v) acetonitrile in water containing 100 mM acetic acid. The eluate from the analytical column was introduced by a nanoelectrospray device (ThermoFinnigan) and sprayed from a gold-coated fused silica emitter (5 μm i.d.; NanoSeparations). The mass spectrometer was operated in a data-dependent acquisition mode to automatically switch between MS (m/z 300–1500 Thompson in centroid mode at a maximum injection time of 150 ms) and MS/MS acquisition on the three most intense precursor ions, controlled by Xcalibur 1.3 software. All MS/MS data were searched using Mascot (Matrix Sciences) and Sequest (ThermoFinnigan) against the GenBank non-redundant protein database (31 March 2005) and against a local database of all possible φKMV gene products.

For the Sequest parameters, a cross correlation value (Xcorr) was set at ≥1.8, ≥2.5 or ≥3.5 for singly, doubly or triply charged peptide ions, respectively. The delta correlation value (ΔCn) was >0.1, while the parent and fragment ion mass tolerance allowed ±3.0 and ±1 Da variation, respectively. Possible static and chemical modifications included were cysteine carbamidomethylation and oxidation of methionine, histidine and tryptophan.

Search parameters included all organisms, with a parent and peptide ion mass tolerance of ±3 and ±0.5 Da, respectively. The significance threshold was set at P<0.05 and one missed tryptic cleavage was allowed.

To ensure validity of the identified proteins, single- and double-peptide protein identifications were re-examined with the de novo sequencing algorithm Lutefsk1900 v.1.3.2. (Taylor & Johnson, 1997), utilizing the database sequence option. In doing so, the de novo-derived sequence was compared to all theoretical peptide spectra deduced from all structural proteins (gp29, 38, 40, 47 and 48) having little or no similarity to known phage proteins can now be classified as structural proteins. In this manner, protein annotation of six -predicted structural proteins is confirmed. In addition, five gene products (gp29, 38, 40, 47 and 48) having little or no similarity to known phage proteins can now be classified as structural proteins (Fig. 2).

The combined data of the two identification approaches allowed us to assign 12 φKMV phage structural proteins to annotated ORFs (Table 1) and to verify the corresponding ORF predictions, since the peptide coverage is spread throughout previously predicted protein sequences. In this manner, protein annotation of six in silico-predicted structural proteins is confirmed. In addition, five gene products (gp29, 38, 40, 47 and 48) having little or no similarity to known phage proteins can now be classified as structural proteins (Fig. 2).

During data analysis, peptide MS/MS spectra were compared to all theoretical peptide spectra deduced from all possible ORFs in the six frames of the φKMV genome sequence. This did not reveal previously unpredicted gene products, nor translational frame-shifts like those observed in the major capsid protein of T7 (Condron et al., 1991).

Identification of proteins after SDS-PAGE separation

Phage particles were analysed by SDS-PAGE after one or two CsCl gradient purification steps and Microcon concentration using a 100 kDa filter (Fig. 1). Subsequently, visible protein bands were analysed by MS. It should be noted that LC-MS does not allow quantification of relative amounts of the identified proteins. Hence, visible protein bands on SDS-PAGE were used to verify the relative decrease of
contaminating proteins during subsequent purification steps, while phage particle proteins remain in constant relative amounts. As shown in Fig. 1, MS analysis identified several P. aeruginosa proteins of the outer-membrane porin family. These include OmpC, F, G, H1, OprF and OprL, and a type B flagellin protein. The number of different proteins detected by MS are listed with their predicted (Mol. mass) and observed (Obs. mol. mass) molecular mass. For both the SDS-PAGE and WSA approaches, the number of identified peptides in each protein and the corresponding protein sequence coverage are indicated. gp29 was found at the electrophoretic front (EF), while other gene products could not be determined (ND) from the SDS-PAGE analysis. The identification analyses of the proteins (Mascot and Lutefisk files) can be verified in the supplementary table of data available with the online version of this paper.

### Table 1. Mass spectrometry data for \( \Phi \)KMV

<table>
<thead>
<tr>
<th>( \Phi )KMV protein</th>
<th>Predicted function</th>
<th>Mol. mass (kDa)</th>
<th>Detected proteins</th>
<th>MS identification data</th>
<th>SDS-PAGE approach</th>
<th>WSA approach</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Obs. mol. mass (kDa)</td>
<td>No. of peptides</td>
<td>Sequence coverage (%)</td>
</tr>
<tr>
<td>gp29</td>
<td>Unknown protein</td>
<td>10 528</td>
<td>EF</td>
<td>5</td>
<td>74·5</td>
<td>1</td>
</tr>
<tr>
<td>gp30</td>
<td>Putative head–tail connector protein</td>
<td>56 194</td>
<td>60</td>
<td>24</td>
<td>62·0</td>
<td>2</td>
</tr>
<tr>
<td>gp31</td>
<td>Scaffolding protein</td>
<td>33 144</td>
<td>ND</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>gp32</td>
<td>Capsid protein</td>
<td>37 710</td>
<td>38</td>
<td>23</td>
<td>63·9</td>
<td>16</td>
</tr>
<tr>
<td>gp33</td>
<td>Putative tail protein A</td>
<td>21 188</td>
<td>ND</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>gp34</td>
<td>Putative tail protein B</td>
<td>92 004</td>
<td>94</td>
<td>23</td>
<td>38·3</td>
<td>1</td>
</tr>
<tr>
<td>gp36</td>
<td>Phage lysin</td>
<td>98 250</td>
<td>94</td>
<td>38</td>
<td>51·9</td>
<td>1</td>
</tr>
<tr>
<td>gp37</td>
<td>Putative internal virion protein</td>
<td>143 885</td>
<td>&gt;100</td>
<td>36</td>
<td>34·3</td>
<td>2</td>
</tr>
<tr>
<td>gp38</td>
<td>Putative tail fibre protein</td>
<td>28 487</td>
<td>32</td>
<td>8</td>
<td>47·8</td>
<td>3</td>
</tr>
<tr>
<td>gp40</td>
<td>Putative tail fibre protein</td>
<td>31 813</td>
<td>29</td>
<td>8</td>
<td>45·2</td>
<td>4</td>
</tr>
<tr>
<td>gp47</td>
<td>Unknown protein</td>
<td>10 564</td>
<td>ND</td>
<td>–</td>
<td>–</td>
<td>2</td>
</tr>
<tr>
<td>gp48</td>
<td>Unknown protein</td>
<td>6 866</td>
<td>ND</td>
<td>–</td>
<td>–</td>
<td>1</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** SDS-PAGE analysis of disrupted \( \Phi \)KMV particles. Proteins derived from disrupted \( \Phi \)KMV particles were separated on SDS-PAGE (12%) after different steps of purification. After silver-staining, each band was excised and identified by MS. For the structure-related \( \Phi \)KMV proteins, the relative amount of protein between bands remains constant in the different gels. In contrast, the disappearance of OprF after a second CsCl purification step exemplifies the relative decrease in the amount of contaminants. While a single purification step allowed identification of five \( \Phi \)KMV proteins, purification and concentration allowed visualization and identification of additional structural proteins (gp37 and 40 after two CsCl steps and gp29 after further Microcon concentration).
P. aeruginosa proteins decreased gradually in the consecutive purification cycles, exemplified most clearly by the disappearance of the OprF band after the second CsCl purification. Although the number of contaminating proteins decreased drastically, final phage samples still contained contaminating porins OprE and OprF, as well as the presumed outer-membrane protein PA2760. This is presumably due to the presence of P. aeruginosa membrane vesicles (between 50 and 250 nm), which are continuously discharged from the cell surface during bacterial growth (Mayrand & Grenier, 1989; reviewed by Beveridge, 1999) and probably some fractions may have a similar density/size as intact φKMV phage particles during CsCl centrifugation. Alternatively, aspecific association between the phage and cellular debris or a specific phage receptor/host interaction cannot be excluded.

In Fig. 1, the proteins identified after SDS-PAGE separation, including tail protein B, the head–tail connector, the capsid and internal core proteins, are shown. Table 1 displays the number of identified peptides and their amino acid coverage within each protein. The most abundant protein in the SDS-PAGE gel, corresponding to the capsid protein (gp32), was previously identified by N-terminal sequencing (Lavigne et al., 2003) and is now confirmed by MS analysis. Because of its abundance, gp32 peptides are present throughout the gel (Fig. 1), most probably due to aspecific adsorption to the gel (in bands of higher molecular mass) and to partially degraded protein (in bands of lower molecular mass).

The recombinant C-terminal domain of gp36 (M737 and E898) was recently described as a functional peptidoglycan-degrading lysozyme with strong thermostable properties (Lavigne et al., 2004). For gp36, one of the phage proteins migrating around 94 kDa, MS analysis revealed a total of 38 significant peptides, originating from throughout the entire gp36 sequence with an amino acid coverage of 51–9 %. Additionally, to verify the enzymic activity of the entire gene product, ORF36 was cloned into pBAD ThioTOPO (Invitrogen), expressed in E. coli and purified as a thioredoxin/gp36/V5/His6 fusion protein by His6-based affinity chromatography. A mean yield of approximately 0.3 μg (gp36) soluble protein (ml cell culture)\(^{-1}\) was obtained. Enzymic activity of the purified recombinant protein was verified by turbidimetric assays on P. aeruginosa protoplasts.
At optimal conditions, an activity of approximately 2820 U mg⁻¹ was measured. Compared to the enzymic activity of the C-terminal end of this protein (gp36C, 20-9 kDa) described previously (Lavigne et al., 2004), the expressed gp36 protein (115 kDa) has a threefold higher activity, taking into account the molecular mass of both proteins and comparing the activity per molecule of enzyme.

WSA

In Table 1, WSA is compared to protein identifications from gel bands. Interestingly, the number of identified peptides within individual proteins bands (and thus amino acid coverage) in SDS-PAGE bands is considerably higher than in the WSA approach. Using WSA, all proteins seen in the SDS-PAGE-based approach analysis were found, but supplemented by three other protein species, i.e. the tail protein A (gp33) and two minor polypeptides, gp47 (10-6 kDa) and gp48 (6-9 kDa). Corresponding ORFs 47 and 48 are located at the extreme right end of the genome, downstream of the lysis cassette, and have no similarity to any protein present in the NCBI database. Since neither these proteins nor gp33 (±21 kDa) has a visible gel band on SDS-PAGE, it is clear that the optimized WSA allows the identification of peptides from small, low-abundance proteins. These identified proteins do not appear to be artefacts, since they are found after all three purification steps. Conversely, the disappearance of gp31 after a second CsCl centrifugation indicates the absence of this protein in the mature phage particle. However, this observation could be consistent with the possible role of this protein as a scaffolding protein, especially in view of its genomic localization compared to T7 (Fig. 2).

Two gene products (gp35 and gp39) are located within the cluster encoding structural proteins, but neither was identified by either technique in the final phage samples used for analysis. If gene annotation is presumed to be accurate, either the absence of these proteins in the mature phage particle or the presence therein at very low abundance is possible. For gp35 and gp39 the latter is expected to be the case.

DISCUSSION

Phage morphology as visualized by electron microscopy has so far served as a strong basis for classification, while DNA-based classification has proved difficult for diverse bacteriophage genomes. However, an alternative classification method based on protein similarity has been suggested (Rohwer & Edwards, 2002), and this method indicates that SP6 is the closest relative of SP6 is the closest relative of KMV (e.g. genome structure and the absence of consensus phage promoter sequences) and the structural differences stated above suggest that KMV should be considered as a separate entity within the T7-like Podoviridae.

Conclusion

MS analysis of purified phage particles allows genome-wide analysis of the structural phage proteins and provides useful complementary information to the available genome sequence. LC-MS/MS after SDS-PAGE separation allowed identification of the dominant structural proteins in KMV with high amino acid coverage. Tryptic digestion of whole phage particles followed by reversed-phase LC and a subsequent fine-tuned gas-phase chromatography pre-fractionation step before MS/MS additionally revealed smaller, less abundant structural proteins. These methods have allowed us to experimentally establish the structural proteome in bacteriophages, allowing classification based on structural proteins, and to provide preliminary insights into phage particle architecture. For KMV, the MS and sequence data suggest subtle structural differences to T7. These differences, also present in phage SP6, include domain
reorganizations in the internal core and injection needle, as well as the absence of ribosomal slippage in the capsid protein of δKMV. The flexibility of the localization of the structural lysin, present on either the C terminus of gp15 or at the N terminus of gp16 (in T7, corresponding to gp36 in δKMV) and observed in other T7-like phages, such as gh-1, suggests that both proteins (gp15 and 16 in T7, and gp35 and 36 in δKMV) interact to form a protruding mechanism of injection.

ACKNOWLEDGEMENTS

K. H., B. R. and Y. B. are FWO postdoctoral and TWO IWT predoctoral fellows, respectively. V. V. M. is supported by the Howard Hughes Medical Institute and Human Frontiers Science Program. This work was financed by the Research Council of the K.U. Leuven (grant OT/04/30) and by the IQR-fund of the Laboratory of Gene Technology.

REFERENCES


