Experimental evidence for proteins constituting virion components and particle morphogenesis of bacteriophage ZF40

Natalia Korol¹, An Van den Bossche²,³, Liudmyla Romaniuk¹, Jean-Paul Noben⁴, Rob Lavigne², Fedor Tovkach¹

¹ Department of Bacteriophage Molecular Genetics, D.K. Zabolotny Institute of Microbiology and Virology, NAS of Ukraine, Kyiv, Ukraine
² Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Leuven, Belgium
³ Division of Bacterial Diseases, Scientific Institute of Public Health (WIV-ISP), 1050 Brussels, Belgium
⁴ Biomedical Research Institute and Transnationale Universiteit Limburg, School of Life Sciences, Hasselt University, Belgium

Abstract

Bacteriophage ZF40 is the only currently available, temperate Myoviridae phage infecting the potato pathogen Pectobacterium carotovorum subsp. carotovorum. Despite its unusual tail morphology, its major tail sheath and tube proteins remained uncharacterized after the initial genome annotation. Using ESI tandem mass-spectrometry, 24 structural proteins of the ZF40 virion were identified, with a sequence coverage ranging between 15,8 and 87,8%. The putative function of sixteen proteins could be elucidated based on secondary structure analysis and conservative domain searches. The experimental annotation of 35% of the encoded gene products within the structural region of the genome represents a complete view of the virion structure, which can serve as the basis for future structural analysis as a model phage.
Key words: bacteriophage ZF40, ESI-MS/MS, virion morphogenesis, proteome analysis, *Pectobacterium carotovorum*

Running title: Bacteriophage ZF40 virion proteins and morphogenesis

**Introduction**

Bacteriophage ZF40 is the only temperate *Myoviridae* phage of the potato pathogen *Pectobacterium carotovorum* subsp. *carotovorum* (previously called *Erwinia carotovora* subsp, *carotovora*) available today. This strain was isolated in Ukraine from a potato tuber soft rot (Tovkach 2002). The phage causes lambda-type lysogeny in a number of closely related pectobacterium strains (Kushkina and Tovkach 2006). Although its genome sequence does not provide evidence for the presence of genes determining pathogenicity factors, lysogens carrying the ZF40 prophage show increased levels of pectate lyase secretion and are more sensitive to mutagenic factors. Phage ZF40 has a relatively small, circularly permuted genome consisting of 48 kb, containing 68 predicted ORFs and no tRNAs (Comeau et al. 2012). The function of 24 genes (35%) could be predicted by the initial annotation. Although the phage lysogenic life strategy is determined by the encoded lambda-like repressor-antirepressor system, ZF40 surprisingly includes a putative transposase gene (Comeau et al. 2012).

The virion morphology is a distinguishing feature based upon which phage ZF40 was classified into a group of so called ‘dwarf’ phiPLPE-like myoviruses (Comeau et al. 2012). The average phage particle consists of an isometric icosahedral capsid 58 nm in diameter and a relatively short contractile tail 86 nm in length (Tovkach 2002). TEM studies revealed a criss-crossed assembly pattern of its sheath protein subunits. Interestingly, neither the major tail sheath nor the major tail tube protein genes were identified by BLAST similarity searches.

Previous research showed that ZF40 phage particles are very unstable and that their morphogenesis can be easily influenced by the host. Wild type virions proved to
be highly unstable when preserved for more than a month (Tovkach et al. 2012) and were shown to be sensitive to osmotic shock when concentrated in CsCl gradients (Korol et al. 2011). Interestingly, long term laboratory observations of some ZF40 variants have revealed an ability to accumulate separate virion components (capsids and tails) when propagated on non-typical indicators, which is an indicative of an abortive infection for such strains (Korol et al. 2012; Romaniuk et al. 2010). This feature makes this phage particularly attractive for structural research.

A deeper understanding of the virion organization is necessary to explain these unique morphogenetic features of bacteriophage ZF40 and to enable us to hypothesize on the observed assembly defects. In this study, we experimentally identify structural proteins included in the virion, predict their coding ORFs and analyze their putative functions.

**Materials and methods**

**Bacteria and bacteriophage.** *P. carotovorum* subsp. *carotovorum* strain 62A (UCM¹) and its derivatives are normally used as sensitivity indicators for bacteriophage ZF40. The virulent phage variant (ZF40-421AK) was isolated in the process of *in vivo* mutagenesis with MitC of the strain 62A-d1 lysogenized with ZF40wt (A. Kushkina, personal communication). It was observed that propagation of ZF40-421AK in the cells of a non-standard strain RC5297 (a bacteriocin stable dissociant of 62A) results in an abortive infection leading to the accumulation of separate phage tails (Korol et al. 2012). A similar feature was detected for another variant – ZF40-RT80 that was obtained by propagation of ZF40 on an alternative host strain *P. carotovorum* J2 (NCPPB² 1744). This strain generates excessive numbers of discrete capsids and procapsids in the corresponding lysate (Romaniuk et al. 2010).

Strains RC5297 and J2 were used to prepare protein samples enriched for separate structural components of the phage ZF40. Preliminary analysis to assess the presence of defective prophage structural genes was performed by PCR using primer

---

¹ Ukrainian collection of microorganisms
² National collection of plant pathogenic bacteria, UK
sequences (TTGAAGCGGCAACATCAACG and CATAGTGGCCAGCTTTCCGA), amplifying the ZF40 major capsid protein gene.

**Purification and concentration of virion components.** Samples containing separate phage tails and capsids along with native phage particles were separated and concentrated by different techniques: (1) ion-exchange chromatography on a column filled with DEAE-fibrous cellulose SS23 (SERVA), (2) ultracentrifugation of the obtained peak fractions (at 30 000 rpm for 2 h at 10°C in Beckmann rotor SW55) and (3) size-exclusion chromatographic separation on the ToyoPearl HW75 (Tosoh Biosciences, Japan) stationary phase (Korol et al. 2013). The resulting samples were analyzed by TEM (JEOL JEM 1400 (Japan).

**Proteomic analysis.** The protein contents of the phage variants were separated by size by standard SDS-PAGE prior to the MS procedure. Protein bands were isolated and analyzed by electron spray ionization tandem mass spectrometry exactly as described previously (Van den Bossche et al. 2014).

**Results and discussion**

**Obtaining samples containing intact phage and its virion components**

The PCR analysis with primer sequences to the ZF40 MCP (major capsid protein) gene revealed no relevant amplicons in both host strains RC5297 and J2 (Supplemental file 1). This limits the possibility of additional expression of phage capsid proteins from the bacterial chromosome, which could lead to the excessive formation of phage heads. It was also shown previously that DNA restriction patterns of ZF40-421AK and RT80 phages with *BamHI, HindIII* and *HpaI* endonucleases are identical to the wild type revealing no deletions or insertions of significant size (Korol’ et al., 2011; Romaniuk et al., 2010).

Three preparations of 2 ml each were obtained: one containing only native phage ZF40-421AK particles, (Fig. 1A), one enriched with its separate tails (Fig. 1B) and one enriched with separate capsids of ZF40-RT80 (although sporadic native virions were detected by TEM in both samples). The final concentration of viable
phages in the native phage sample was about $1 \times 10^{12}$ PFU/ml, while other two samples contained $1 \times 10^9$ and $1 \times 10^8$ PFU/ml respectively.

**General characterization of the phage proteome**

Due to different concentrations, the number of protein bands visualized on the SDS-PAGE gel varied significantly in the samples tested (Fig. 1C, D). The lane containing intact phages, which includes 11 bands, was used as a control. The capsid preparation was characterized by the highest protein content, while the tails-enriched sample revealed only three distinguishable bands corresponding to the most abundant structural proteins with calculated molecular weights: 40, 33 and 16 kDa (Fig. 1C).

Mass spectrometry samples taken systematically (covering the entire lane) for all three samples, allowed to analyse the structural proteome of the phage and to assign functions to the predicted structural genes. Overall, 24 proteins with a sequence coverage ranging from 15.8 to 87.8 % were detected in the sample containing native particles (Fig. 2A). The functions of sixteen proteins were predicted based on secondary structure analysis and conservative domain search with HHpred and Phyre2 tools (Table 1; Supplemental file 2) (Remmert et al. 2012; Kelley et al. 2015). HHpred and Phyre2 identity levels with template amino acid sequences reached respectively 11-37% and 10-34% for proteins of different function, which indicates distant evolutionary relationships with the well-studied phages available in the databases. However, the identity levels of major structural proteins of bacteriophage ZF40 and other phiPLPE-like phages are higher – 23-47% (Supplemental file 2). This supports the idea proposed by Comeau et al. that these dwarf phages can be organized into a coherent group based on the similarity of their morphogenetically significant genes (Comeau et al. 2012).

Subsequently, all of the identified proteins can be analyzed according to their involvement in head, tail, baseplate and fiber assembly process.

**Capsid morphogenesis proteins**
Experimental data indicates at least five proteins involved in bacteriophage ZF40 head morphogenesis (Fig. 2B). The most abundant band in the sample enriched with capsids corresponds to the sequence of gp46, previously identified as the major capsid protein. Its secondary structure prediction reveals a conservative domain (aa 26-309, 15% identity), characteristic for bacteriophage capsid proteins of different families of Caudovirales. A corresponding putative head protease ORF (gp44) was located upstream of the ZF40 MCP gene. Its low sequence coverage (15%) and prevalence in the protein band with a mobility of approximately 30-32 kDa (compared to the estimated by MS/MS molecular weight of 44 kDa) may be explained by the proteinase self-cleavage following phage capsid maturation.

The predicted molecular weight identified by MS/MS for the ZF40 MCP is about 34 kDa, yet its mobility in gel is slightly lower (31 kDa). Since this protein has a moderate pl of 5.04, a shifted mobility in the gel based on the overall charge is unlikely. Combined with the availability of a prohead protease, one may assume that approximately 3 kDa of the MCP may be cleaved in the process of a prohead maturation. Previously, the TEM analysis revealed that the ZF40-RT80 sample is characterized, not only by high numbers of capsids, but also by the presence of procapsids (Romaniuk et al. 2010). The SDS-PAGE profile of that sample included an additional major band of 33 kDa, absent in the wild type phage profile. Possibly, this band represents an uncleaved form of the MCP subunit.

Remarkably, ZF40 capsids might lack a sufficient scaffolding protein or use a different mechanism for the formation of a prohead. The relevant genes are usually located between the protease and the MCP genes (Oh et al. 2014). However, in this case, these ORFs are interrupted by an ORF of bacterial origin (Fig. 2B). A similar insertion was observed in the model representative of ‘dwarf’ phages – phiPLPE (Comeau et al. 2012). Possibly, the scaffolding role may be played by a protein coded by one of the uncharacterized ORFs located upstream of the MCP gene.

The phage ZF40 connector structure is predicted to include three proteins: gp42, gp43 and gp48. The putative portal protein gp42 (15% identity) and putative
head morphogenesis protein gp43 (24% identity) that may form a complex similar to SPP1 gp6-gp7. In this complex, gp7 is responsible for controlling the phage DNA injection into the bacterial cell. It should be noted that the putative head to tail connector protein (gp48) was observed by the identification of only two peptides (sequence coverage 15%) in the tails sample compared to eight (sequence coverage 66%) in the intact phage virion (Table 1) which indicates its lower concentration in the sample. A defect in gp48 could cause the improper head-tail joining leading to the observed abundance of separate phage tails in the ZF40-421AK sample.

**Tail structural proteins**

The contractile tail structure of bacteriophage ZF40 is interesting considering its unusual morphology. There are eleven ZF40 ORFs coding for proteins presumed to be involved in the assembly of tail components (Fig. 2C). As mentioned previously, genes responsible for the phage tail sheath and tail tube assembly were not predicted by BLAST search. However, MS/MS analysis of the sample enriched with phage tails revealed the most abundant protein, having a molecular weight of 42,25 kDa, encoded by ORF52. The amino acid sequence contains a fold (E-value $8.4\times10^{-10}$ and 15% identity) corresponding to the tail sheath proteins of *Myoviridae* phages (including T4, Mu and P2). This data was supported by Phyre2, demonstrating that 34% of the aa sequence (residues 186-319) creates a fold matching the C-terminal fragment of gp18 – the major sheath protein (MSP) of phage T4. According to its mobility in gel, the actual protein size is about 40 kDa, which may also indicate a post-translational modification.

The identification of the major tail tube (TTP) protein was more difficult, due to the small quantity of these proteins in the sample. Generally, its MW ranges from 12 to 22 kDa in different *Myoviridae* representatives (Supplemental file 3), and its coding sequence is located downstream of the MSP gene (Pell *et al.* 2009). β-sheets are the prevailing structures forming the internal lining of the tube, which is necessary for proper DNA sliding during the injection process. We calculated that the amount of β-structures in the tail tube proteins of model myoviruses stays between
33-39% while α-helixes constitute 9-16% (Supplemental file 3). The MW of the gp53 identified by MS/MS is 14.7 kDa and this protein is characterized by a ratio of β-sheaths and α-helixes within these boundaries. In addition, although the e-value of the alignment with the phage Mu tail tube is not significant (Table 1), one may consider the ORF53 to encode ZF40 tail tube protein.

The phage ZF40 tape measure protein (TMP) gp57 consists of 586 aa which corresponds to the tail length of 86 nm (it consists of α-helixes one turn of which is 0.54 nm high and consists of 3.6 aa residues (Pauling et al., 1951)). It includes a lysozyme motif known to facilitate efficient infection and DNA injection into stationary phase cells (Piuri and Hatfull 2006). It is noteworthy that myoviruses in general contain tail-associated chaperones that promote TTP polymerization and tube assembly (Xu, Hendrix and Duda 2014). They are coded by two ORFs located in between the TTP and TMP. In ZF40, the three relevantly located ORFs 54-56 were previously identified as two genes of bacterial origin and one ORFan (Comeau et al. 2012). Other dwarf phages, such as phiPLPE and PY100, are also characterized by ORFan-rich regions between the TTP and TMP genes. Therefore, one may assume that tail chaperones may be coded by the ORFan or other ORFs of unknown function localized within the structural module (Fig. 2C).

Other tail proteins demonstrate similarity to the relevant proteins of enterobacterial bacteriophages. Gp58 and gp49 are predicted to be involved in tail completion and its attachment to the capsid (Linderoth et al. 1994). The bacteriophage ZF40 baseplate cluster shows syntenic to the P2 late gene operon. Three proteins encoded by ORFs 61-63, have similar molecular weights and share structural similarity (18-31% identity) with P2 gpV, W and J (Haggard-Ljungquist et al. 1995). They are predicted to form a tail spike, a baseplate wedge and a baseplate distal edge, respectively.

Finally, bacteriophage ZF40 tail fibers appear to include at least two proteins: gp66 and gp67. The latter is assumed to form the distal part of tail fibers and contains an Ig-like domain belonging to the Big2 family, which are predicted to enhance
phage interaction with carbohydrates on the bacterial cell surface and facilitate phage adsorption (Fraser, Maxwell and Davidson 2007). Alternatively, this might even be indicative of an association with the plant, similar to observations made for phages active in the gut (Barr et al. 2013). While all other structural proteins share conservative domains with some or all PLPE-like phages, it is not surprising that ZF40 fiber proteins do not show any similarity to fiber proteins of other representatives of the group. This is not unexpected since these phage have a different host specificity.

**Conclusion**

The functional characterization of the ZF40 structural region of the genome presented in this study supports the earlier idea suggested by Comeau et al. about the phages’ relation to the group of phiPLPE-like bacteriophages (Comeau et al. 2012). Proteins involved in the phage morphogenesis demonstrate sufficiently higher levels of identity with the corresponding protein sequences within this group of viruses rather than with the representatives of other genera.

The proteomic data analyzed in this manuscript provided the experimental proof of gene products constituting a ZF40 phage particle, enabled us to improve the initial genome annotation by 10% and to experimentally verify the existence of 24 out of the 68 encoded proteins using ESI-MS/MS. By analyzing the mass spectrometry data of the virion components, the function of seven previously uncharacterized proteins could now be predicted, as well as the function of nine other proteins detected, based on their secondary structure prediction. Such functional analysis is valuable for more accurate structural characterization of the newly emerging genomes of viruses related to the phiPLPE group. However, further studies are necessary to determine the actual pathway of bacteriophage ZF40 particle assembly and to identify the exact reasons for the morphogenetic malfunctions leading to the formation of discrete structural components of the virion.

**Acknowledgements**
The authors thank Erik Royackers (UHasselt, Belgium) for technical assistance with the mass spectrometric analysis, Dr. Sergiy Voichuk (IMV NANU, Ukraine) for the assistance with electron microscopy, Katarzyna Danis-Wlodarczyk (KU Leuven, Belgium) for the help with bioinformatic programs, Dr. Alla Kushkina (IMV NANU, Ukraine) for the fruitful discussions on the topic of the article and Dr. V. Samoilenko for PCR primer design.

References


---

**Fig. 1.** Transmission electron micrographs and protein content of samples containing intact phage particles and samples enriched with phage tails and capsids. A, B – TEM images of the native phage and its tails respectively. C, D – SDS-PAGE of the concentrated native particles (1), samples enriched with phage

Fig. 2. Genetic map of bacteriophage ZF40 and detailed structure of its morphogenetic region. A – Full circular genome map demonstrating identified and uncharacterized genes. B. Capsid gene cluster: P – portal, H – minor head protein, P’ase – protease, MCP – major capsid protein, Con – connector protein. C. Genes

Table 1. Bacteriophage ZF40 structural proteins identified by ESI-MS/MS

<table>
<thead>
<tr>
<th>Protein</th>
<th>Identified function</th>
<th>Band No</th>
<th>Molecular weight, kDa</th>
<th>Number of unique peptides</th>
<th>Sequence coverage, %</th>
<th>HHpred significant match (model name)</th>
<th>E-value(^4)</th>
<th>Identity(^5), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>gp4 2</td>
<td>Portal protein</td>
<td>6</td>
<td>51,92</td>
<td>35</td>
<td>30</td>
<td>20</td>
<td>81.7</td>
<td>3.8 e-08</td>
</tr>
<tr>
<td>gp4 3</td>
<td>Head morphogenesis protein</td>
<td>12</td>
<td>36,01</td>
<td>–</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>4.9 e-25</td>
</tr>
<tr>
<td>gp4 4</td>
<td>Capsid protein protease</td>
<td>10</td>
<td>44,12</td>
<td>–</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1.2</td>
</tr>
<tr>
<td>gp4 6</td>
<td>Major capsid protein</td>
<td>10</td>
<td>34,35</td>
<td>31</td>
<td>10</td>
<td>13</td>
<td>8</td>
<td>1.6 e-05</td>
</tr>
<tr>
<td>gp4 8</td>
<td>Head to tail connecto r protein</td>
<td>17</td>
<td>18,74</td>
<td>–</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>0.3 9</td>
</tr>
<tr>
<td>gp4 9</td>
<td>Phage tail compone nt</td>
<td>16, 17</td>
<td>19,58</td>
<td>–</td>
<td>3</td>
<td>5</td>
<td>–</td>
<td>9.5 e-09</td>
</tr>
<tr>
<td>gp5 2</td>
<td>Major tail sheath protein</td>
<td>6</td>
<td>42,25</td>
<td>40</td>
<td>17</td>
<td>21</td>
<td>19</td>
<td>8.4 e-10</td>
</tr>
<tr>
<td>gp5 3</td>
<td>Putative tail tube protein</td>
<td>17</td>
<td>15,71</td>
<td>16</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4.6</td>
</tr>
<tr>
<td>gp5 7</td>
<td>Tape measure protein/ly sin</td>
<td>4,5</td>
<td>63,95</td>
<td>66</td>
<td>25</td>
<td>28</td>
<td>23</td>
<td>4.2 e-47</td>
</tr>
<tr>
<td>gp5 8</td>
<td>Tail protein</td>
<td>12</td>
<td>22,39</td>
<td>–</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>58.7</td>
</tr>
</tbody>
</table>

\(^3\) Band number in the sample containing native phage particles  
\(^4\) HHpred E-value of the most significant matching protein model  
\(^5\) Identity % - the HHpred percent sequence identity for a pairwise aa alignment with the template protein.
<table>
<thead>
<tr>
<th>gp6</th>
<th>Protein Type</th>
<th>Mol. Wt</th>
<th>Ext. Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Baseplate protein</td>
<td>9</td>
<td>34.2</td>
</tr>
<tr>
<td>1</td>
<td>Baseplate spike protein</td>
<td>12</td>
<td>24.5</td>
</tr>
<tr>
<td>2</td>
<td>Baseplate wedge protein</td>
<td>17</td>
<td>15.0</td>
</tr>
<tr>
<td>3</td>
<td>Baseplate assembly protein</td>
<td>6</td>
<td>42.8</td>
</tr>
<tr>
<td>6</td>
<td>Tail fiber protein</td>
<td>16</td>
<td>19.7</td>
</tr>
<tr>
<td>7</td>
<td>Tail fiber protein/I G-like domain</td>
<td>3</td>
<td>71.5</td>
</tr>
</tbody>
</table>

- **gp6 0**: 43 kDa tail protein; baseplate, gene product MU, baseplate, gene product 44.
- **gp6 1**: GPV, baseplate assembly protein V Enterobacteria phage P2.
- **gp6 2**: Tail lysozyme; GP25-like fold, initiation of sheath polymerization Enterobacteria phage T4.
- **gp6 3**: Baseplate J-like protein Enterobacteria phage P2.
- **gp6 6**: Long tail fiber protein P37; viral protein; 2.20A Enterobacteria phage T4.

---

6 The function of this protein was suggested by Comeau et al.