DATASET BRIEF

Identification and comparative analysis of the structural proteomes of ϕKZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages

Elke Lecoutere[1], Pieter-Jan Ceyssens[1]*, Konstantin A. Miroshnikov[2], Vadim V. Mesyanzhinov[2], Victor N. Krylov[3], Jean-Paul Noben[4], Johan Robben[4], Kirsten Hertveldt[1], Guido Volckaert[1] and Rob Lavigne[1]*

[1] Division of Gene Technology, Katholieke Universiteit Leuven, Leuven, Belgium
[2] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia
[3] State Institute for Genetics and Selection of Industrial Microorganisms, Moscow, Russia
[4] Biomedical Research Institute and transnational University Limburg, School of Life Sciences, University Hasselt, Diepenbeek, Belgium

Giant bacteriophages ϕKZ and EL of *Pseudomonas aeruginosa* contain 62 and 64 structural proteins, respectively, identified by ESI-MS/MS on total virion particle proteins. These identifications verify gene predictions and delineate the genomic regions dedicated to phage assembly and capsid formation (30 proteins were identified from a tailless ϕKZ mutant). These data form the basis for future structural studies and provide insights into the relatedness of these large phages. The ϕKZ structural proteome strongly correlates to that of *Pseudomonas chlororaphis* bacteriophage 201ϕ2-1. Phage EL is more distantly related, shown by its 26 non-conserved structural proteins and the presence of genomic inversions.

Received: September 10, 2008
Revised: January 27, 2009
Accepted: February 23, 2009

Keywords:
Bacteriophages / ESI-MS/MS / Structural proteome

The Myoviridae are characterized by a long, contractile tail and represent 25% of all isolated tailed phages [1]. Among these, the ϕKZ-related phages (ϕKZ, phage EL and 201ϕ2-1) share a highly similar morphology and taxonomically relevant similarity at the proteome level [2]. As the type representative, ϕKZ has been extensively studied, has been used in phage therapy [3] and is a source of useful enzymes [4–7]. These phages provide a convenient model for molecular and structural biology [8, 9].

Bacteriophages ϕKZ, 201ϕ2-1 and EL have permuted, terminally redundant genomes comprising 280 334 bp (306 encoded genes), 316 674 bp (468 genes) and 211 215 bp (201 genes), respectively [3, 10, 11]. Although these phages lack homology at the DNA level, they appear morphologically identical. Whereas *Pseudomonas chlororaphis* phage 201ϕ2-1 is closely related to ϕKZ (167 similar proteins) [10], ϕKZ and EL share only protein similarity in 66 gene products [11]. We here analyze the structural proteomes of phages ϕKZ and EL using ESI-MS/MS and compare these proteomes to that of the recently analyzed 201ϕ2-1.

Both phages were amplified to a titer of 10^{11} pfu/mL and purified by a double CsCl density gradient centrifugation as described previously [3]. After disruption of 10^{11} phage particles by reduction in 2 mM β-mercaptoethanol followed by heat-denaturation (95°C, 5 min) [12], particle proteins were separated by 1-D SDS-PAGE. For ϕKZ, 38 bands were visualized in the range of 20–150 kDa by Coomassie blue staining (Fig. 1A). Analogously, 31 bands from phage EL were visualized (Fig. 1C). For ϕKZ, a temperature-sensitive, tailless mutant phage [8] was used for a further discrimination of head-specific proteins (Fig. 1B). For protein identification, entire gel lanes were cut into slices and subjected to an overnight trypsin digestion [13, 14]. Peptides

Correspondence: Dr. Rob Lavigne, Division of Gene Technology, Katholieke Universiteit Leuven, Kasteelpark Arenberg 21-box 2462, B-3001 Leuven, Belgium
E-mail: rob.lavigne@biw.kuleuven.be
Fax: +32-16-32-19-65

*These authors contributed equally to this work.

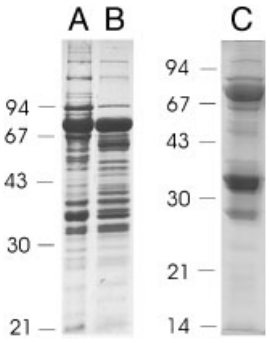

Figure 1. ϕKZ (A) and EL (C) structural proteome separated by SDS-PAGE and visualized by Coomassie blue. The ϕKZ heads (B) were from a tailless mutant. For ϕ, since no clear bands were visible below 20 kDa, this part of the gel lane was not analyzed.

were separated by LC with a linear 5–60% v/v ACN gradient and subsequently identified by ESI-MS/MS (LCQ Classic, Thermofinnigan) in an *m/z* range of 300–1500. All MS data were analyzed using Sequest (ThermoFinnigan) against the NCBI database considering a minimal cross correlation value of 1.8, 2.5 and 3.5 for single, double or triple charged peptide ions, respectively. The delta correlation 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 trypsin cleavage was allowed.

In total, 62 ϕKZ structural proteins were identified (Table 1, PRIDE database ♯8648), at least 30 of which are part of the phage head, as identified from the ϕKZ tailless mutant. In total, 64 EL structural proteins were identified (Table 1, PRIDE database ♯8648). Among the 64 ϕKZ proteins, five (KZ089, KZ120, KZ145, KZ146 and KZ181) have previously been identified as structural proteins, as had the main capsid protein (EL078). These previous identifications were made by N-terminal sequencing [3, 15].

Sequence-based similarity with proteins of other phages had originally predicted only the major capsid proteins for EL and ϕKZ. However, a 2008 proteomic study of phage 201ϕ2-1, found 69 structural proteins with similarity to ϕKZ proteins. The present study confirms directly that 60 of these proteins are, indeed, ϕKZ structural proteins (Supporting Information Table 1). In addition, two structural proteins (KZ164 en KZ224) were detected in our study, though they were not reported as structural proteins in Thomas *et al.* [10]. This high overall correlation and the fact that this analysis was performed in parallel and independently of the Thomas *et al.* analysis, suggests that the number of non-structural gene products (due to co-purification) is low and that the identified proteins are indeed true particle proteins. Comparison of EL and 201ϕ2-1 reveals 38 EL virion proteins with similarity to 201ϕ2-1 structural proteins. All of these EL proteins are confirmed here to be structural proteins. However, 25 structural EL proteins were detected without homology to either ϕKZ or 201ϕ2-1 proteins, consistent with the lower proteomic correlation of EL to ϕKZ compared with 201ϕ2-1 [11]. Apparently, the morphological differences caused by these proteins are too small to be visible by electron microscopy of a negatively stained specimen. Another possibility might be that they are residing within the phage particle as internal proteins.

In many phage genomes (<100 kb, mainly Podoviridae and Siphoviridae), structural genes are exclusively clustered in a single genome region [*e.g.* 12]. However, as encountered in Myoviridae like T4, substructures of larger and more complex phage particles are often encoded in separate clusters, which are distributed throughout the genome [16]. This genome-wide distribution is also very pronounced in the giant *Pseudomonas* phages described in this study, as illustrated in Fig. 2. It can be argued that this spread of structural gene clusters provides an efficient barrier for extensive gene shuffling and horizontal gene transfer. Most interesting is also the reshuffling of nine structural regions in EL compared with ϕKZ and 201ϕ2-1, a phenomenon that has not been observed to this extent, to our knowledge. The orientation of both single genes (*e.g.* EL175) and larger genomic regions (*e.g.* EL5-9, EL182-192) is inverted (Fig. 2). No bacterial promoters or other regulatory elements could be detected in this region, and possible causes of these genomic inversions remain unknown.

Specific functional information cannot be obtained from these data, although the positive identification of regions encoding head-specific proteins in ϕKZ is a first step. The major capsid proteins of both phages (EL078 and KZ120) are the most abundant. EL063 and KZ145, the second most abundant proteins, have an unknown function. The major sheath proteins, EL6 and KZ029, and the immediate downstream EL005 and KZ030 are also strongly represented. Proteins EL113, EL114 and EL115 (predicted tail fiber proteins) were detected, whereas EL117 was not found, despite weak similarity with baseplate proteins. The predicted function of KZ030 as tail tube protein is reinforced by the complete absence of this protein in phage heads. In contrast, EL156 and KZ146 were proposed as potential tail fiber proteins, based on collagen-like motifs within the protein sequences [3, 11]. However, KZ146 is still clearly present in the ϕKZ heads, contradicting this hypothesis. The presence of the endolysin (KZ144) as part of the structural proteome is not entirely unexpected as it has been located in the phage tail by antibodies [17] and since 201ϕ2-1 has a structural homologue [10].

In conclusion, the high-throughput analysis of the structural proteins of phage ϕKZ, ϕKZ-head and EL allowed identification of the majority of the encoded structural proteins (62, 30 and 64

Table 1. Proteins identified by MS for bacteriophages ϕKZ (A) and EL (B)[a)]

ORF	Processing	MW (Da)	AA	%Cov	Pept	ORF	Processing	MW (Da)	AA	%Cov	Pept
A. Identified ϕKZ proteins											
26		61 994	552	14.5	7	131		84 606	771	7.0	4
27		105 242	898	1.4	1	132	N-terminal	12 284	110	42.7	3
28*		36 458	320	38.8	13	133		52 685	462	2.6	1
29*	N-terminal	77 641	695	33.8	17	134		51 612	457	7.9	3
30		32 659	293	51.9	11	135		52 173	463	13.0	5
32*		49 365	424	13.7	5	139	N-terminal	32 659	298	40.6	8
34		29 864	256	14.1	2	143		20 102	177	29.9	6
35*		29 699	271	3.7	1	144		28 815	260	17.7	3
52*		42 180	363	9.6	2	145*	Edman	84 116	789	54.6	38
79*		33 212	296	17.2	5	146*		115 835	1093	54.2	45
80*		50 520	449	9.8	3	149*		25 308	224	21.9	3
83*		54 546	504	10.7	4	153	N-terminal	34 099	304	29.6	7
84*		46 877	413	4.8	2	157		51 102	445	32.4	15
85		16 462	149	51.0	6	160		17 783	157	21.0	2
86*		48 880	428	10.0	5	162*		57 633	522	31.6	16
87	C-terminal	110 375	971	14.4	8	163*	N-terminal	44 184	395	15.9	4
89*		42 200	387	29.7	11	164		32 922	297	5.4	1
90*		35 913	309	31.4	9	174	N-terminal	39 093	354	25.7	6
91		20 799	181	6.6	1	175*	C-terminal	30 771	270	42.2	8
92		50 620	440	4.1	2	176*		27 734	242	15.3	3
93*	N-terminal	47 611	435	22.1	11	177*		58 753	519	11.4	5
94*		56 949	505	8.9	4	178*		164 886	1451	23.1	27
95*	N-terminal	58 906	547	37.1	21	180*		56 200	490	2.2	1
97*	N-terminal	87 816	816	21.8	16	181	Edman	245 756	2237	24.0	38
99*		54 751	469	6.2	2	184		25 277	230	3.5	1
101		51 962	460	2.4	1	201		72 571	649	4.2	2
119		20 234	177	13.6	3	202		16 495	156	19.9	2
120*	Edman	82 929	747	43.8	29	224		16 644	144	7.6	1
127		33 607	290	4.5	1	244		13 431	118	7.6	1
128*		84 218	724	1.2	1	298		36 277	335	11.0	3
130		48 589	427	1.4	1	303*	N-terminal	72 095	646	22.4	12
B. Identified phage EL proteins											
5	N-terminal	33 371	302	58.6	10	103	C-terminal	110 287	990	35.8	23
6		79 822	722	59.7	40	104		89 921	773	21.0	11
7		32 627	292	69.9	17	106		34 969	315	65.4	12
8		10 0547	870	30.0	21	112		47 895	413	17.7	7
9	N-terminal	64 953	574	57.7	27	113		27 529	247	45.3	8
13		52 791	453	19.2	8	114	N-terminal	131 988	1184	25.4	19
14		31 237	282	26.2	6	115	N-terminal	132 098	1193	28.2	23
20		15 586	140	60.0	8	116		104 898	956	29.2	20
22		42 364	366	52.7	12	144		9 252	86	51.2	3
28		15 638	135	51.9	5	147		22 795	209	66.5	13
44	N-terminal	48 222	433	31.6	10	154		24 894	222	58.6	11
45	N-terminal	34 134	299	24.7	5	155	N-terminal	132 097	500	64.6	19
50	N-terminal	39 521	358	40.8	17	156	C-terminal	113 218	1046	56.3	46
52	N-terminal	50 278	453	46.1	13	157	N-terminal	51 496	480	55.4	17
53		16 863	148	60.1	6	158		12 295	109	24.8	2
54	N-terminal	57 113	505	50.7	16	165	N-terminal	38 691	351	55.8	21
55		27 588	243	85.6	17	166		107 728	940	7.1	6
56		110 398	976	34.1	24	168		23 220	210	71.4	11
57		41 016	357	20.7	6	169		50 263	448	45.1	16
58	N-terminal	39 988	364	72.3	19	172	N-terminal	31 312	275	39.3	13
59	C-terminal	33 411	294	77.6	19	175		25 201	225	45.3	9
60		19 189	170	72.9	11	181		40 748	369	16.3	4
61	N-terminal	48 148	431	64.3	23	182		80 088	730	15.6	7
62	N-terminal	36 862	337	59.1	18	183		277 884	2543	20.1	29
63	N-terminal	44 659	418	48.1	28	184		54 659	481	33.5	12
64	N-terminal	34 653	312	37.2	10	186		121 789	1067	33.2	28

Table 1. Continued

ORF	Processing	MW (Da)	AA	%Cov	Pept	ORF	Processing	MW (Da)	AA	%Cov	Pept
66	N-terminal	50 990	468	46.2	20	187		40 252	357	17.6	5
68		58 911	497	14.3	6	188		32 472	292	8.2	1
69	C-terminal	49 842	431	43.4	17	189		6 361	58	65.5	4
71		50 335	445	21.3	7	190		57 616	516	70.3	28
77	N-terminal	19 580	181	49.7	7	191		27 337	242	21.9	4
78	Edman	79 869	741	66.8	36	192	C-terminal	32 985	287	52.6	12

a) Structure-related ORFs are indicated, with their molecular weight (MW, in Da) and number of amino acids (AA). The coverage percentage (%Cov) based on the number of identified peptides (Pept) is given, as well as presumed N-terminal or C-terminal processing. Protein processing predictions were based on protein coverage in mass spectrometric identification and correlate with the predicted and confirmed processing of 201ϕ2-1 virion proteins [10]. In some cases, the N-terminal was already verified by Edman degradation [3]. Proteins marked with an asterisk were identified from proteins of the tailless ϕKZ mutant.

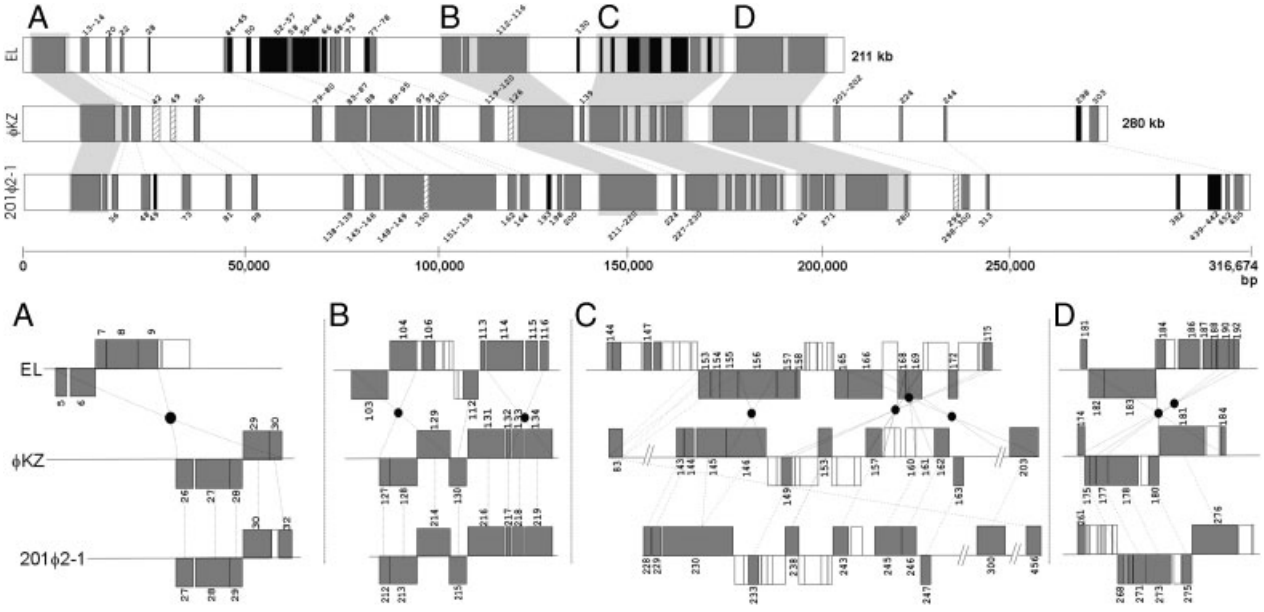

Figure 2. Comparison of structural proteomes of phages EL, ϕKZ and 201ϕ2-1. Identified proteins are indicated as gray boxes, proteins unique to a phage are marked in black, related proteins are connected with dotted lines. Four regions (A–D) containing genomic inversions (black dots) are shown in detail.

gene products, respectively). This is supported by the high correlation to the analysis of 201ϕ2-1 by Thomas *et al.* [10]. These data confirm the original open reading frame predictions after genome sequencing of both phages and provide a necessary first step toward their functional characterization in structural studies. From this perspective, the observed differences between the EL and the cryo-EM analyzed ϕKZ proteomes are interesting and elicit the need for further study on phage EL.

J. C. and E. L. hold predoctoral fellowships of the 'Instituut voor aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen' (I.W.T., Belgium). R. L. holds a postdoctoral fellowship of the 'Fonds voor Wetenschappelijk Onderzoek-Vlaanderen' (FWO-Vlaanderen, Belgium). This work was financially supported by the research council of the K.U. Leuven (OT/05/47 and BIL/05/46).

The authors have declared no conflict of interest.

References

[1] Ackermann, H. W., 5500 phages examined in the electron microscope. *Arch. Virol.* 2007, *89*, 227–243.

[2] Lavigne, R., Summer, E., Seto, D., Mahadevan, P. *et al.*, Classification of *Myoviridae* bacteriophages using BLASTP-tools. *Res. Microbiol.*, submitted.

[3] Mesyanzhinov, V. V., Robben, J., Grymonprez, B., Kostyuchenko, V. A. *et al.*, The genome of bacteriophage

ϕKZ of *Pseudomonas aeruginosa*. *J. Mol. Biol.* 2002, *317*, 1–19.

[4] Miroshnikov, K. A., Faizullina, N. M., Sykilinda, N. N., Mesyanzhinov, V. V., Properties of the endolytic transglycosylase encoded by gene 144 of *Pseudomonas aeruginosa* bacteriophage ϕKZ. *Biochemistry* 2006, *71*, 300–305.

[5] Briers, Y., Volckaert, G., Cornelissen, A., Lagaert, S. *et al.*, Muralytic activity and modular structure of the endolysins of *Pseudomonas aeruginosa* bacteriophages ϕKZ and EL. *Mol. Microbiol.* 2007, *65*, 1334–1344.

[6] Paradis-Bleau, C., Cloutier, I., Lemieux, L., Sanschagrin, F. *et al.*, Peptidoglycan lytic activity of the *Pseudomonas aeruginosa* phage ϕKZ gp144 lytic transglycosylase. *FEMS Microbiol. Lett.* 2007, *266*, 201–209.

[7] Briers, Y., Miroshnikov, K., Chertkov, O., Nekrasov, A. *et al.*, The structural peptidoglycan hydrolase gp181 of bacteriophage phiKZ. *Biochem. Biophys. Res. Commun.* 2008, *PMID*: 18671939.

[8] Fokine, A., Kostyuchenko, V. A., Efimov, A. V., Kurochkina, L. P. *et al.*, A three-dimensional cryo-electron microscopy structure of the bacteriophage ϕKZ head. *J. Mol. Biol.* 2005, *352*, 117–124.

[9] Fokine, A., Battisti, A. J., Bowman, V. D., Efimov, A. V. *et al.*, Cryo-EM study of the *Pseudomonas* bacteriophage ϕKZ. *Structure* 2007, *15*, 1099–1104.

[10] Thomas, J. A., Rolando, M. R., Carroll, C. A., Shen, P. S. *et al.*, Characterization of *Pseudomonas chlororaphis* myovirus 201ϕ2-1 via genomic sequencing, mass spectromety, and electron microscopy. *Virology* 2008, *376*, 330–338.

[11] Hertveldt, K., Lavigne, R., Pleteneva, E., Sernova, N. *et al.*, Genome comparison of *Pseudomonas aeruginosa* large phages. *J. Mol. Biol.* 2005, *354*, 536–545.

[12] Ceyssens, P.-J., Mesyanzhinov, V., Sykilinda, N., Briers, Y. *et al.*, The genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage resembling M6. *Bacteriol.* 2008, *190*, 1429–1435.

[13] Dumont, D., Noben, J.-P., Raus, J., Stinissen, P., Robben, J., Proteomic analysis of cerebrospinal fluid from multiple sclerosis patients. *Proteomics* 2004, *4*, 2117–2124.

[14] Lavigne, R., Noben, J.-P., hertveldt, K., Ceyssens, P.-J. *et al.*, The structural proteome of *Pseudomonas aeruginosa* bacteriophage KMV. *Microbiology* 2006, *152*, 529–534.

[15] Bourkaltseva, M. V., Krylov, V. N., Pleteneva, E. A., Shaburova, O. V. *et al.*, Phenogenetic characterization of a group of giant ϕKZ-like bacteriophages of *Pseudomonas aeruginosa*. *Russ. J. Genet.* 2002, *38*, 1242–1250

[16] Miller, E. S., Kutter, E., Mosig, G., Arisaka, F. *et al.*, Bacteriophage T4 genome. *Micorobiol. Mol. Biol. Rev.* 2003, *67*, 86–156.

[17] Miroshnikov, K. A., Faizullina, N. M., Sykilinda, N. N., Mesyanzhinov, V. V., Properties of the endolytic transglycosylase encoded by gene 144 of *Pseudomonas aeruginosa* bacteriophage phiKZ. *Biokhimiia* 06, *71*, 379–385.

