Table S1. Cleavage sites of ϕKZ proteins detected by a semi-tryptic peptide.

gp	Residue[1]	Site[2]
84	10	ISNE VLDV
86	122	ISQE AVMA
89	122	FSAE EYSD
89	156	ASME ANDE
92[3,4]	85	VSLE DLSK
93[3,4]	13	LSLE SIED
93	156	ISLE DLKE
94[3]	11	DSLE GYGV
94	95	VASE EYPA
94	275	VSNE RVTP
95	13	DSLE DYTE
95	117	IAVE TVIP
95[4]	133	QGGE TVTQ
95[3]	260	ISTE DIKE
96	65	VSTE DISD
97	13	ASLE NQDG
97	66	ISAE EHEE
97	468	ASNE SMLE
97	532	TSLE SFNS
120[3,4]	54	ISHE NFGT
120	84	YGFE RVSS
120	163	VGLE NYNE

129	262	IATE THHF
153	52	ISQE ANFA
162[3]	17	ASLE RVSD
162	137	PSNE DISI
162	200	TALE NYIG
162	258	ASIE VLDN
163	54	FALE SIDP
175	210	PALE NYED
177	60	ASVE EFFE
177	297	LASE ANNV
180_180_2	505	YSEE AIAE
184	108	YGHE HFDE
203[3]	653	ESLE SYVD
303	60	YSNE EWVE
303[3]	153	IFVE ADTT
303	268	AALE EFIE
303	366	VSNE GLID

[1]The coordinate of the conserved glutamate immediately preceding the cleavage.

[2]Sequence identified in a semi-tryptic fragment is underlined.

[3]Peptide assigned with a minimum peptide identification probability <95% according to Scaffold

[4] low abundance peptide.

Table S2. Conservation of cleavage sites detected by mass spectrometry in ϕKZ and 201ϕ2-1 in homologous proteins.

ϕKZ gp	ϕKZ cleavage site/motif[1]	201ϕ2-1 gp	201ϕ2-1 cleavage site/motif[1]	ϕPA3 gp	ϕPA3 cleavage motif[1]
84	ISNE-10	145	ISTE-9	86	LSME-10
86	ISQE-122	148	ISQE-113	89	VSNE-110
89	FSAE-122 ASME-156	151	LAFE-112nr ASLE-166	92	FSKE-118 YSME-145nr
92	VSLE-85	154	VSNE-79	95	ISTE-84
93*	LSLE-13 ISLE-156	155	VSTE-13 NA AALE-327	100	LSQE-12 NA VSTE-216
94*	DSLE-11 VSNE-275	156	PSNE-157	99	PANE-156
95*	DSLE-13 IAVE-117 QGGE-133 ISTE-260	157	VSNE-15 NA NA NA IALE-207	100	LSQE-12 nr nr VSTE-216
96	VSTE-65	158	LSLE-60	101	VSTE-71
97	ASLE-13 ISAE-66 ASNE-468 TSLE-532	159	NA NA ELSE-567 ASLE-630nr	102	NA NA TTVE-494 TSLE-555nr

no homolog		193	GSLE-139	no homolog	
120	ISHE-54 YGFE-84 VGLE-163	200	IGTE-57 YGFE-86 VALE-161	136	IGNE-52 YGFE-81 VGLE-163
129	IATE-262	214	MEAE-272	148	LEAE-186
153	ISQE-52	238	ISTE-64	176	VSTE-68
162*	ASLE-17 PSNE-137 TALE-200 ASIE-258	246	IANE-16 NA TSVE-128nr ASIE-203 VSLE-258	186	ISME-16 TSVE-124nr AAVE-193 sites nr
163*	FALE-54	247	TALE-50	187	IGLE-50
175	PALE-210	268	PALE-218	205	PALE-216
177	ASVE-60 LASE-297	271	VAVE-61 LRNE-293	208	ASVE-60
178		274/3	VTFE-275	211	
180_180_2	YSEE-505	275	YSEE-504	212	YSKE-504
184	ASLE-13 NA NA YGHE-108	280	KGLE-15 VSAE-72 NA	217	LSLE-11 NA NA NA
203	ELSE-653	300	nr	233	ASFE-102
303	NA YSNE-60 NA IFVE-153	455	VSNE-20 NA PGME-151 NA	375	NA NA ISIE-16 NA

AALE-268	**NA**	**NA**
VSNE-366	**ISNE-179**	**NA**

[1] Red indicates cleavage confirmed by detection of semi-tryptic peptide in mass spectral analyses. Identified cleavages in 201ϕ2-1 are from (Thomas *et al.*, 2010). Black indicates putative cleavage sites in homologous proteins determined by Blastp alignment, cleavages/motifs aligning perfectly are underlined. nr, refers to putative cleavage motif found near (usually within 10 residues), but not aligning with, a known cleavage site in an homologous protein. sites nr, refers to putative cleavage motifs near, but not aligning with, a known cleavage site in an homologous protein. NA, refers to no alignment in corresponding region of homolog [some putative cleavage motifs were identified in these regions based on their being within ~20 residues of the N-terminus of the protein and therefore candidates for a capsid targeting sequence or CTS (Mullaney and Black, 1996)]. Blank, refers to no identifiable cleavage site. Asterisks indicates members of the PF12699 family.

Table S3. Conservation of T4 processed proteins among representatives of the T4-like phage subfamily. Taxonomically, JS98, RB43, RB49 and 44RR2.8t are members of the T4-like phage genus, whereas KVP40 and P-SSM2 are outside of the T4-like phage genus (Lavigne *et al.*, 2009). Divergence is measured as diverged from 100-percent identity of a BlastP match. The terminase protein, tail sheath and tube proteins are included as being representatives of highly conserved morphogenesis proteins. Proteins shaded in the same color have genes located in the same module. Proteins with names/gps in red indicate high abundance capsid proteins (>100 copies). A dash indicates no homologs were identified using BlastP.

	Divergence of homologs to T4 proteins (%)													
T4 protein, functional classification:	terminase gp17,	sheath gp18,	tube gp19,	portal gp20,	core gp67,	core gp68,	protease gp21,	scaffold gp22,	**capsid gp23,**	vertex gp24,	**IPI,**	**IPII,**	**IPIII,**	alt,
	essen	essen	essen	essen	essen	semi-E	essen	essen	essen	essen	aux	aux	aux	aux
JS98	18	29	29	25	47	29	11	22	16	30	-	-	14	-
RB43	31	39	39	47	50	48	38	47	30	58	-	-	-	-
RB49	31	38	51	43	60	43	44	46	33	55	-	-	-	-
44RR2.8t	37	40	36	43	47	50	45	53	30	56	-	-	-	-
KVP40	46	52	44	54	80	60	51	62	40	70	-	-	-	-
P-SSM2	65	73	68	62	-	-	58	76	65		-	-	-	-

[1]Classification of gene function as essential (essen), semi-essential (semi-E) or auxillary (aux) was obtained from (Miller *et al.*, 2003)

Table S4. Conservation of the homologs to ϕKZ processed proteins among representatives of the ϕKZ-related phages. Divergence is measured as 100-percent identity (determined using BlastP). The terminase protein, tail sheath and tube proteins are included as being representatives of highly conserved morphogenesis proteins. Gp90 is also included as it is an abundant inner head protein. Proteins shaded in the same color have genes located in the same module. Proteins with names/gps in red indicate high abundance capsid proteins (>100 copies). A dash indicates no homologs were identified using BlastP.

				Divergence of homologs to ϕKZ proteins (%)										
ϕKZ protein:	terminase gp25	sheath gp29	tube gp30	RNAP beta'' gp80	gp84	gp86	**gp89**	**gp90**	gp92	**gp93**	gp94	**gp95**	gp96	**gp97**
201ϕ2-1	40	27	29	44	66	76	46	53	69	73	71	79	73	70
ϕPA3	31	30	27	35	64	76	37	52	62	62	79	74	67	70
EL	73	76	79	77	78	-	74	84	-	-	-	-	-	-

ϕKZ protein:	**capsid** **gp120**	gp129	gp153	**gp162**	gp163	protease gp175	gp177	RNAP gp178	RNAP gp180_180_2	184P	helicase gp203	gp303
201ϕ2-1	36	49	73	76	55	45	59	47	32	41	49	72
ϕPA3	32	44	68	73	47	41	53	42	25	45	42	70
EL	79	80	-	-	-	71	-	70	68	-	71	-

References

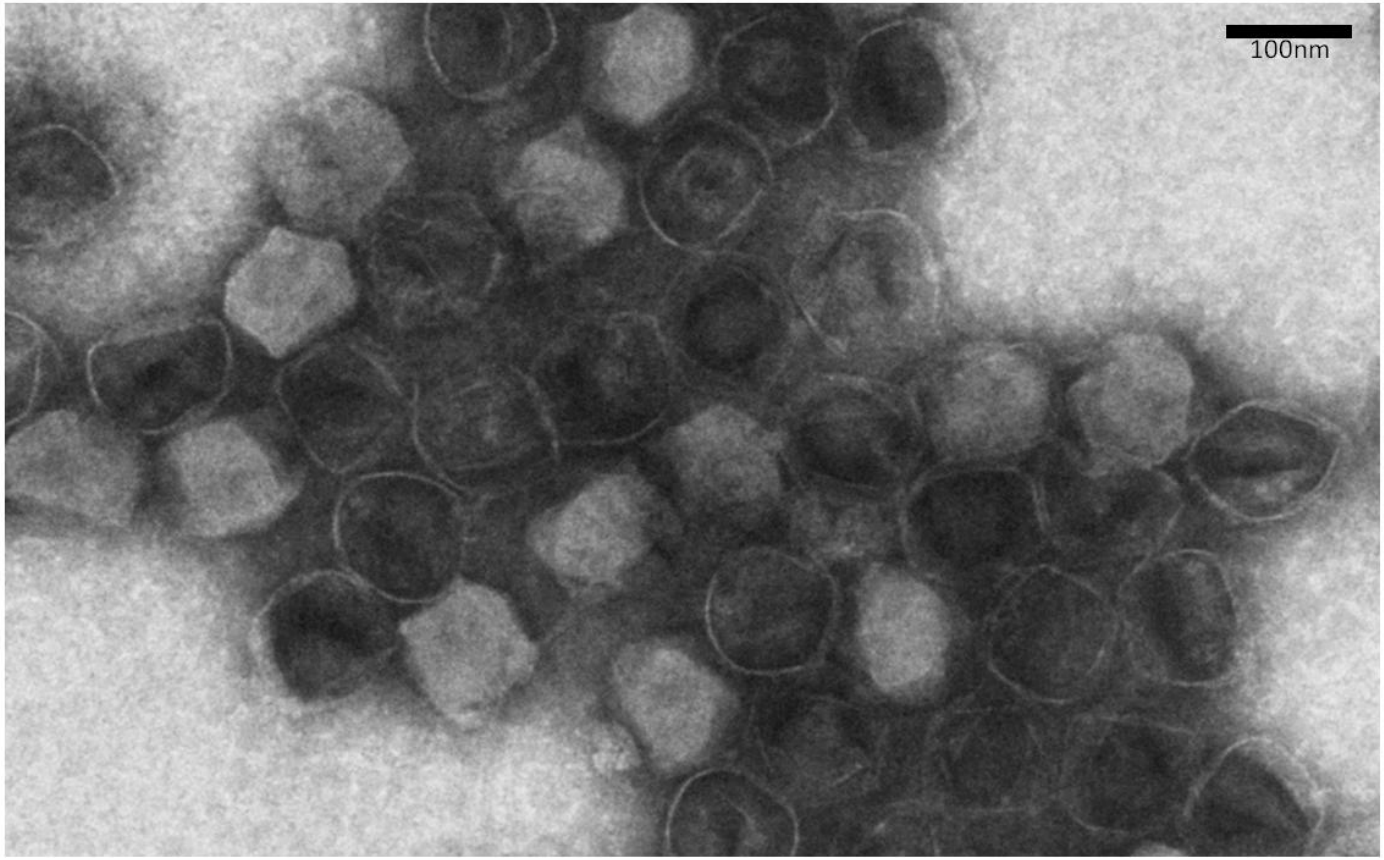
Lavigne, R., Darius, P., Summer, E.J., Seto, D., Mahadevan, P., Nilsson, A.S., *et al.* (2009) Classification of *Myoviridae* bacteriophages using protein sequence similarity. *Bmc Microbiol* **9**: 224.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. and Ruger, W. (2003) Bacteriophage T4 Genome. *Microbiol. Mol. Biol. Rev.* **67**: 86-156.

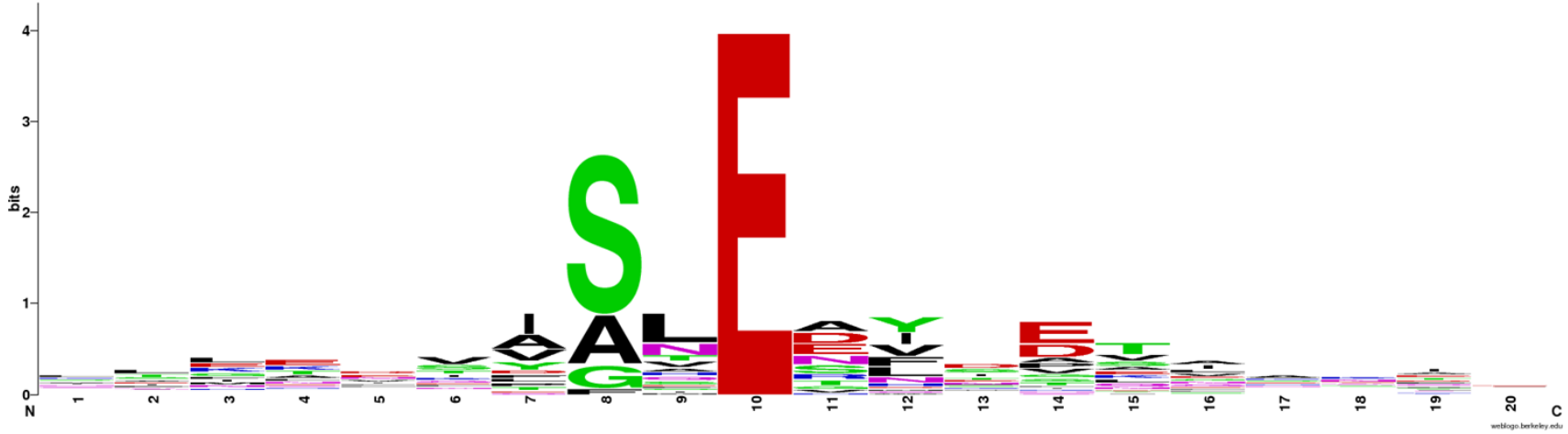
Mullaney, J.M. and Black, L.W. (1996) Capsid targeting sequence targets foreign proteins into bacteriophage T4 and permits proteolytic processing. *J. Mol. Biol.* **261**: 372-385.

Thomas, J.A., Weintraub, S.T., Hakala, K., Serwer, P. and Hardies, S.C. (2010) Proteome of the large *Pseudomonas* myovirus 201phi2-1: Delineation of proteolytically processed virion proteins. *Molecular & Cellular Proteomics* **9**: 940-951.

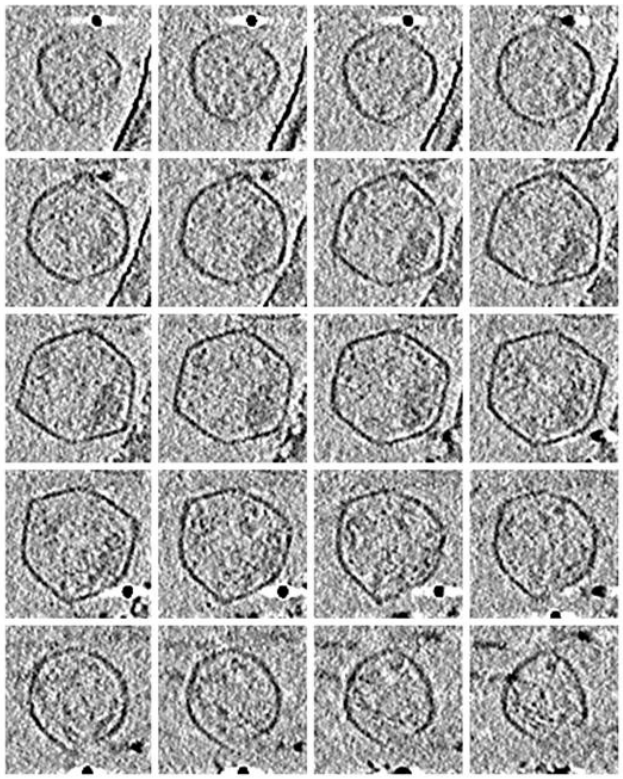
Supplemental Figure 1. Electron micrograph of purified tailless ϕKZ heads negatively stained with 1% phosphotungstic acid.



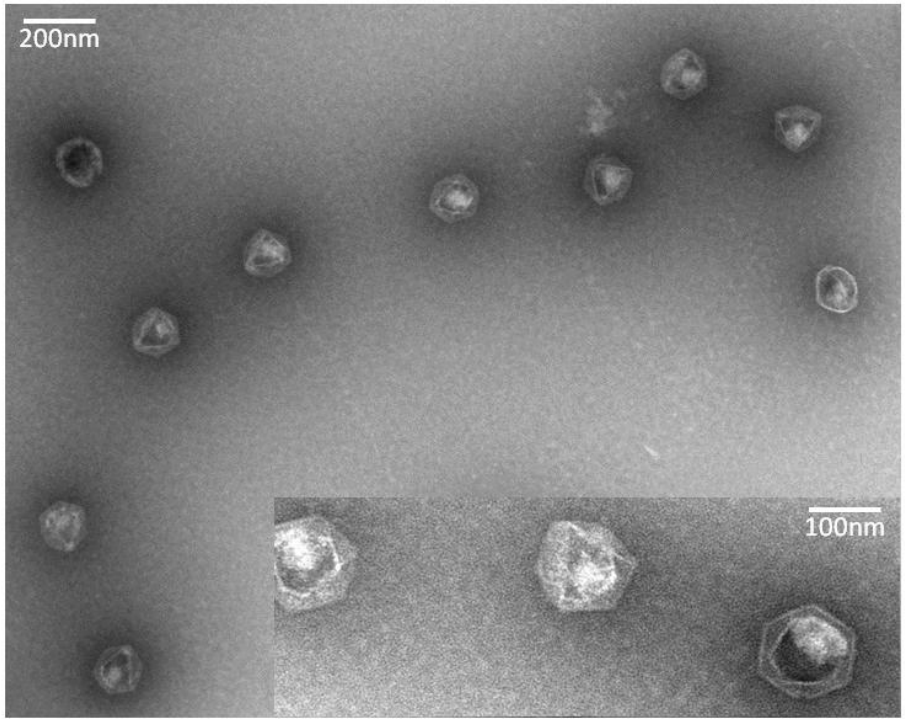
Supplemental Figure 2. Sequence logo representing a total of 39 cleavage sites in ϕKZ determined by detection of a semi-tryptic peptide. Logo was created using Weblogo (http://weblogo.berkeley.edu/). The co-ordinates of the cleavage sites used for construction of the logo are listed in Table S1.



Supplemental Figure 3. Montage of serial slices through a cryo-electron tomogram of a ϕKZ capsid produced in the presence of 9-aminoacridine. Each slice is 172 nm across and 1.6 nm thick, with a separation between slices of 4.7 nm.



Supplemental Figure 4. Purified ϕKZ capsid-related particles produced in the presence of 9-aminoacridine stained with 2.5% ammonium molybdate and 10% trehalose.



Supplemental Figure 5. Electron micrograph of empty ϕKZ capsid shells after osmotic shock treatment. Sample is Fraction 6 of the glycerol gradient in Figure 7A. Sample was negatively stained with 1% phosphotungstic acid. Space bar represents 100 nm.