

論文 / 著書情報  
Article / Book Information

Title	ORF334 in Vibrio phage KVP40 plays the role of gp27 in T4 phage to form a heterohexameric complex
Author	Shuji Kanamaru, FUMIO ARISAKA, Mai Nemoto
Journal/Book name	J Bacteriol, vol. 190, , p. 3606-12
Issue date	2008, 9

## ORF334 in *Vibrio* Phage KVP40 Plays the Role of gp27 in T4 Phage To Form a Heterohexameric Complex<sup>∇</sup>

Mai Nemoto,<sup>1</sup> Kazuhiro Mio,<sup>2</sup> Shuji Kanamaru,<sup>1</sup> and Fumio Arisaka<sup>1\*</sup>

Department of Biomolecular Engineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4359-B39 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan,<sup>1</sup> and National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 2, Umezono 1-1-4, Tsukuba Science City, Ibaraki 305-8568, Japan<sup>2</sup>

Received 18 January 2008/Accepted 28 February 2008

**KVP40 is a T4-related phage, composed of 386 open reading frames (ORFs), that has a broad host range. Here, we overexpressed, purified, and biophysically characterized two of the proteins encoded in the KVP40 genome, namely, gp5 and ORF334. Homology-based comparison between KVP40 and its better-characterized sister phage, T4, was used to estimate the two KVP40 proteins' functions. KVP40 gp5 shared significant homology with T4 gp5 in the N- and C-terminal domains. Unlike T4 gp5, KVP40 gp5 lacked the internal lysozyme domain. Like T4 gp5, KVP40 gp5 was found to form a homotrimer in solution. In stark contrast, KVP40 ORF334 shared no significant homology with any known proteins from T4-related phages. KVP40 ORF334 was found to form a heterohexamer with KVP40 gp5 in solution in a fashion nearly identical to the interaction between the T4 gp5 and gp27 proteins. Electron microscope image analysis of the KVP40 gp5-ORF334 complex indicated that it had dimensions very similar to those of the T4 gp5-gp27 structure. On the basis of our biophysical characterization, along with positional genome information, we propose that ORF334 is the ortholog of T4 gp27 and that it plays the role of a linker between gp5 and the phage baseplate.**

Tailed phages reside in the order *Caudovirales*, a taxonomic classification that includes more than 95% of all bacteriophages (17). The tailed phages constitute three of the familial groups making up the order *Caudovirales*, *Myoviridae* (phages having a long contractile tail), *Siphoviridae* (phages having a noncontractile, flexible tail), and *Podoviridae* (phages having a short, noncontractile tail). Among the family *Myoviridae*, the T4-related phages are characterized by their elongated icosahedral heads and large baseplates. So far, the complete genome sequences of 10 T4-related phages (DNA Data Bank of Japan [http://www.ddbj.nig.ac.jp]) have been determined. The most characterized and best-understood member of the T4-related family is the phage T4, from which the group derives its name. Members of the T4-related phage family have been classified into four groups based on the sequence similarity of gp23 (a major capsid protein; gp, gene product) and gp18 (a tail sheath protein); these four groups are T-even, Pseudo T-even, Schizo T-even, and Exo T-even (3, 28). Between T-even and Pseudo T-even there is a high degree of conservation of gp5 at the amino acid sequence level (>50%) (4). Between T-even and Schizo T-even, the amino acid similarity between gp5 proteins is between 20% and 50%. Between T-even and Exo T-even, the similarity is <20%. KVP40, a *Vibrio* phage, is a member of the Schizo T-even phage genus. It has a more elongated head than bacteriophage T4 (140 by 70 nm) (18, 21). Phage KVP40 was isolated from seawater, and it has a broad host range. The genome size of phage KVP40 is 245 kbp, which

includes 386 ORFs. About 30% of the ORFs in KVP40 have significant homology with those of T4 (21).

The gp5 protein of bacteriophage T4 is an essential structural component of the phage baseplate. In the T4 phage, the gp5 protein possesses a lytic activity (8–10, 13, 14, 19), and because of this, it was originally referred to as a lysozyme that caused “lysis from without” (6, 7, 22). During the tail assembly of phage T4, gp5 first interacts with gp27 and forms a heterohexameric complex, (gp5)<sub>3</sub>(gp27)<sub>3</sub>, (1, 4, 8–10). X-ray crystallography of the complex in combination with electron microscopy (EM)-based three-dimensional image reconstruction has unambiguously localized the gp5-gp27 complex to the central region of the baseplate at the tip of the tail tube (2, 4, 13). The gp5 protein has three distinct domains, namely, the N-terminal domain (gp5N), the lysozyme domain (gp5Lys), and the C-terminal domain (gp5C). The gp5C domain is responsible for forming the extraordinary triple-stranded β-helix, which plays a major role in both puncturing the outer membrane of the host, *Escherichia coli*, and locally degrading the peptidoglycan layer. The gp27 trimer in the complex forms a cuplike structure, together with gp5N at the base (cup inner and outer diameters, ~30 and 80 nm, respectively). Trimeric gp27 has a pseudo-sixfold symmetry. It connects the threefold-symmetrical tail lysozyme complex with the sixfold-symmetrical baseplate (4). The upper part of the cup is thought to connect the tail lysozyme complex with the tail tube via two tail-associated proteins, gp48 and gp54 (2, 11, 13).

The gp5 protein of KVP40 has a high degree of sequence similarity in the N-terminal and C-terminal domains to the T4 gp5 protein (46% and 35%, respectively). It also possesses the multiple-repeat VXGXXXXX sequence in the C-terminal domain; however, it lacks the lysozyme domain that the T4 gp5 protein possesses (Fig. 1). No KVP40 ORF product having

\* Corresponding author. Mailing address: Department of Biomolecular Engineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4359-B39 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan. Phone and fax: 81-45-924-5713. E-mail: farisaka@bio.titech.ac.jp.

<sup>∇</sup> Published ahead of print on 7 March 2008.

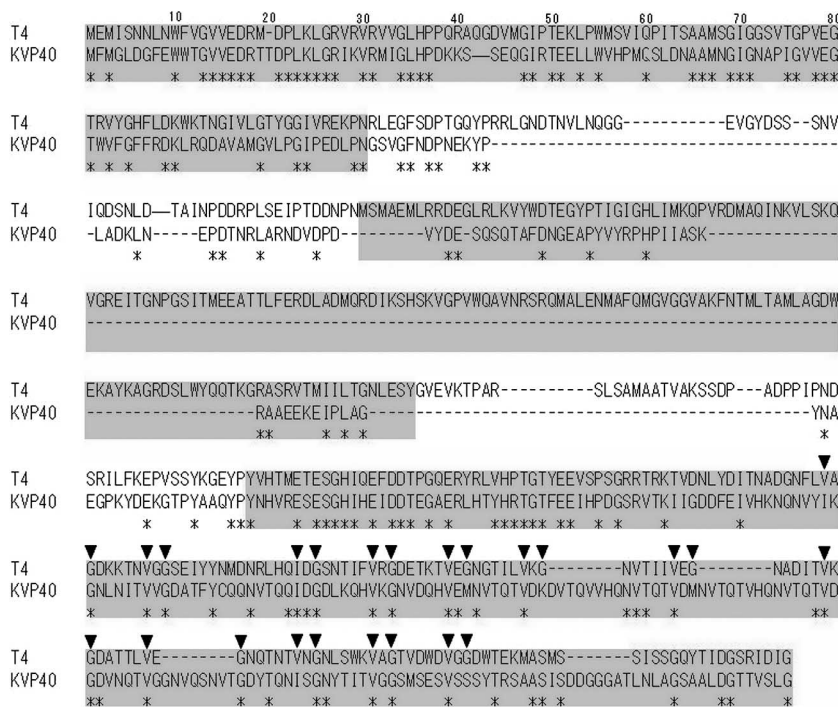


FIG. 1. Sequence comparison of gp5 proteins from phages T4 and KVP40. Sequences of gp5 proteins from six T4-related phages, T4 and RB69 (T-even), RB49 and 44RR2.8t (Pseudo), and Aeh1 and KVP40 (Schizo), were aligned using ClustalW. Only the results for gp5 proteins from T4 and KVP40 are shown. Shading indicates the N-terminal, lysozyme, and C-terminal domains of T4. The asterisks indicate identical amino acids in the two phages, and the arrowheads denote V and G in the 8-residue repeats, VXGXXXXX, in the C-terminal domain of T4.

homology to the T4 gp27 has been detected, in spite of the fact that a number of other baseplate proteins in the two phages share significant homology (21). We surmised that there exists a gene encoding a T4 gp27 homolog in the KVP40 genome and that the three-dimensional structure is better preserved than the amino acid sequence. We chose ORF334 as a candidate for the T4 gp27 homolog, as the gene product is of a size similar to that of T4 gp27 and it is located in the baseplate gene cluster (the arrangement of baseplate genes in KVP40 is different from that of phage T4, a point taken up in Discussion below). In the present study, we expressed and analyzed KVP40 gp5 and ORF334. It was shown that KVP40 gp5 forms a trimer in solution and that it forms a heterohexamer with KVP40 ORF334. We propose that KVP40 ORF334 is the homolog of T4 gp27.

**MATERIALS AND METHODS**

**Vector construction.** Gene 5 in the genome of KVP40 was amplified by PCR using two primers, 5'-CCCCCGGATCCATATCGCAATTGCGGGGAGC-3' (the BamHI site is in boldface) and 5'-CCCCGTCGACTGAACCTAGACTTACTGTGTGC-3' (the SalI site is in boldface; the italicized anticodon TGA [ser] was changed from a stop anticodon, TTA, for a C-terminal histidine tag). For the expression of gene 5, a plasmid, pMNK, was created by inserting the PCR product into pET32, which had been digested with BamHI and SalI. For the formation of the gp5-ORF334 complex, the gene 48-5 cluster (Fig. 2) was amplified by PCR with the primers 5'-CCCCGGATCCTATTTGCCAGAATATG C-3' (the BamHI site is in boldface) and 5'-CCCCGTCGACTGAACCTAGACTTACTGTGTGC-3' (the SalI site is in boldface; the italicized anticodon TGA [ser] was changed from a stop anticodon, TTA, for a C-terminal histidine tag) using the KVP40 genome as the template. Plasmid pMNC, which was designed to express gene 48, gene 53, orf334, and gene 5, was created by inserting the PCR product into pET29 after digesting it with BamHI and SalI.

**DNA sequencing.** The CEQ2000 DNA analysis system (Beckman-Coulter) was used for DNA sequencing in combination with a Dye Terminator Cycle Sequencing Quick Start kit. Five primers were synthesized in order to confirm the DNA sequence: 5'-GGAGCAAGCAAACCGAGTC-3', 5'-CTGACTCTTAACGAT TAC-3', 5'-GAAATACCCGGGAACAC-3', 5'-GCGTTTGATAACGGTGAA GCGCC-3', and 5'-CGTTACTCAACAGATTGACGGGG-3'.

**Amino acid sequence alignment of gp5 proteins from T4-related phages.** The amino acid sequences of the gp5 proteins from six T4-related phages, T4, RB49, RB69, 44RR2.8t, Aeh1, and KVP40, were aligned using ClustalW (30). In Fig. 1, only the sequences from T4 and KVP40 are shown.

**Expression and purification of gp5.** For expression of gene 5, *E. coli* BL21(DE3) cells containing pMNK were cultivated in LB medium with 200 µg/ml ampicillin at 37°C. When the optical density of the culture at 600 nm was 0.4 to 0.5, protein expression was induced by 1 mM IPTG (isopropyl-β-D-thio-galactopyranoside). The cells were pelleted at 2,140 × g for 20 min 4 hours after induction.

The purification steps for gp5 tagged with His<sub>6</sub> at the C terminus were based on those for T4 (5) and were modified as follows. Harvested cells were resuspended in a 10× volume of buffer 1 (50 mM Tris-Cl, 5 mM imidazole, pH 8.0) and sonicated with phenylmethylsulfonyl fluoride at a final concentration of 1 mM. After centrifugation at 20,000 × g for 20 min, the supernatant was loaded onto a HiTrap chelating column charged with nickel (GE Healthcare) that had been equilibrated with buffer 1, and the proteins were eluted by a linear gradient of 5 to 500 mM imidazole in EDTA-containing solution. The fractions containing the desired proteins were collected and applied to a HiTrap Q HP column (GE Healthcare) equilibrated with buffer 2 (50 mM Tris-Cl, pH 8.0). The targeted proteins were eluted at an NaCl concentration of 0.40 to 0.45 M by use of a linear gradient of 0 to 1 M. The fractions containing the pertinent proteins were collected and concentrated to 2 to 5 ml by Amicon Ultra 50K (Millipore) and then loaded onto a Hiload 16/60 Superdex 200-pg column (GE Healthcare) that had been equilibrated with buffer 3 (50 mM Tris-Cl, 100 mM NaCl, pH 8.0). The proteins were collected after elution with buffer 3.

**Expression and purification of the gp5-ORF334 complex.** For coexpression of gene 48, gene 53, orf334, and gene 5, *E. coli* BL21(DE3) cells containing the pMNC plasmid were cultivated in LB medium in the presence of 50 µg/ml kanamycin at 37°C. Expression was induced by the addition of 1 mM IPTG when

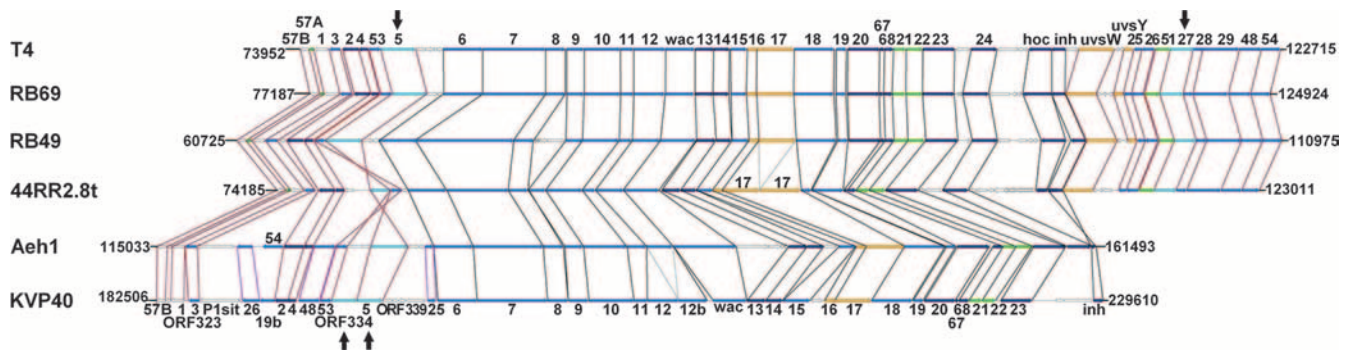


FIG. 2. Alignment of six T4-related phage genomes. Tail and head gene alignments among six T4-related phages, T4, RB69, RB49, 44RR2.8t, Aeh1, and KVP40, are shown. The color definitions are as follows: light blue, tail protein (gp5 and ORF334 are shown in sky blue); dark blue, head and neck proteins; yellow, DNA association protein; and green, chaperonin and assembly catalysis. The lines among phages show the orthologs: red is baseplate-related proteins, black is others. The arrows indicate the positions of gene 5 and gene 27, and gene 5 and ORF334, of T4 and KVP40, respectively.

the culture reached an optical density of 0.4 at 600 nm, and the cells were then incubated at 20°C for 1 h. The cells were harvested at  $2,140 \times g$  for 15 min after overnight incubation at 20°C.

The purification steps were the same as those for gp5 except that (i) buffer 1 was 50 mM Tris-Cl, 5 mM imidazole, 50 mM NaCl; (ii) buffer 2 was 50 mM Tris-Cl, 50 mM NaCl, pH 7.5; and (iii) the gradient of NaCl was 0.05 to 1 M (eluted at 0.4 to 0.5 M).

**SDS-PAGE and N-terminal amino acid sequence determination.** Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the method of Laemmli (15) with a vertical minislab gel (9 by 7.5 cm). The gel was stained with staining buffer (0.1% Coomassie brilliant blue, 10% acetic acid). Proteins separated on SDS-PAGE were transferred to polyvinylidene difluoride membranes electrophoretically. After proteins on the membranes were visualized by Coomassie brilliant blue, bands were cut out and the sequences were confirmed by a protein sequencer (PPSQ-21 protein sequencer; Shimadzu).

**CD spectrum.** The far-UV circular-dichroism (CD) spectrum of gp5 was measured at 20°C with a J-720 spectropolarimeter (Jasco) in a 1-mm-path-length cell. The protein concentration of gp5 was 0.26 mg/ml. The reference solvent was buffer 3, which was also used for prior exhaustive dialysis of the protein sample. The CD spectrum obtained between 198 and 240 nm was analyzed using the program CONTINLL (24) in order to estimate the secondary structure.

**Analytical ultracentrifugation.** Sedimentation velocity and equilibrium experiments were conducted with an Optima XL-I (Beckman-Coulter) using a four-hole An60Ti or an eight-hole An50Ti rotor at 20°C. gp5 was dialyzed against buffer 3, and the dialysate was used as the reference solution. For the gp5-ORF334 complex, the equilibration buffer for gel filtration was used as the reference solution, because extended periods of dialysis tended to result in nonspecific protein aggregation. Sedimentation velocity data were acquired at a rotor speed of 40,000 rpm for gp5 and 35,000 rpm for the complex without specification of time intervals between successive scans. The sedimentation coefficient distribution function,  $c(s)$  was obtained using the SEDFIT program (26, 27). The molecular mass distribution  $c(M)$  was obtained by converting  $c(s)$  on the assumption that the frictional ratio  $f/f_0$  was common to all the molecular species (as implemented in SEDFIT).

Sedimentation equilibrium was carried out at starting absorbances at 280 nm ( $A_{280}$ ) of 0.15, 0.3, and 0.5 at rotor speeds of 6,000, 8,000, and 10,000 rpm for gp5 and at  $A_{280}$  of 0.2, 0.3, and 0.4 at rotor speeds of 4,500, 7,000, and 8,500 rpm for the complex. For each experiment, the data were globally fitted to a single-species model to determine the molecular weight. The protein partial specific volumes ( $\bar{v}$ ) were determined based on the amino acid sequence by the program SEDNTERP (16; J. Philo, unpublished data). The  $\bar{v}$  value of the complex was calculated using the amino acid composition of a hypothetical tandemly connected ORF334 with gp5 (as the complex contains an equal number of moles of gp5 and ORF334, as determined by SDS-PAGE [see Results]). The buffer density ( $\rho$ ) and viscosity ( $\eta$ ) were also calculated using the SEDNTERP program.

**EM.** A solution of the gp5-ORF334 complex with approximately 0.03- $\mu$ g/ml total protein concentration in buffer 3 (defined above) was adsorbed onto a thin carbon film (supported on top of copper mesh grids) that had been previously rendered hydrophilic by glow discharge in a partial vacuum. Samples were

washed with 5 drops of double-distilled water, negatively stained twice with 2% uranyl acetate solution for 30 s each time, blotted, and then dried in air. Micrographs of negatively stained particles were recorded in a JEOL 100CX transmission electron microscope at  $\times 53,000$  magnification using a 100-kV acceleration voltage. The images were recorded on SO-163 films (Eastman Kodak), developed with a D19 developer (Eastman Kodak), and digitized with a Scitex Leafscan 45 scanner (Leaf Systems Inc.) at a pixel size of 1.82 Å at the specimen level.

## RESULTS

gp5 and ORF334 of phage KVP40 were amplified, cloned, and overexpressed in order to establish the respective subunit stoichiometries of gp5 and ORF334 and to determine the nature of their interaction with each other. This information was used to assess whether ORF334 is the ortholog of the T4 phage gp27.

**Purification of gp5.** gp5 expressed in BL21(DE3) cells was purified with Ni affinity, anion-exchange, and gel filtration chromatography to homogeneity (Fig. 3a) (see Materials and Methods). In order to confirm that the purified protein was gp5, the N-terminal amino acid sequence was determined by

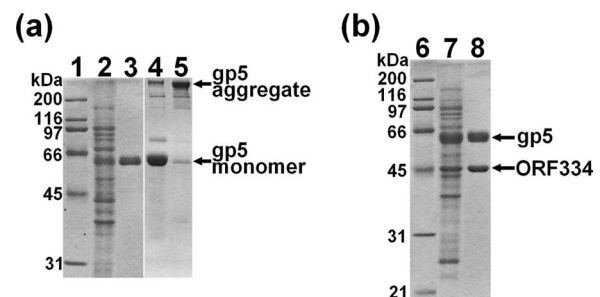


FIG. 3. SDS-PAGE of purified gp5 and the gp5-ORF334 complex. The gels (12.5%) were stained with Coomassie brilliant blue. (a) Purification of gp5 and change in the migration pattern in the presence of urea. Lane 1, standard molecular mass marker; lane 2, soluble fraction of the cell that overexpressed gp5 using pMNK; lane 3, purified gp5 boiled in the presence of urea; lane 4, purified gp5 boiled in the presence of urea; lane 5, same sample as lane 4 boiled in the absence of urea. (b) Purification of the gp5-ORF334 complex. Lane 6, standard molecular mass marker; lane 7, soluble fraction of cells that overexpressed gp5, gp53, ORF334, and gp5 using pMNC; lane 8, purified gp5-ORF334 complex boiled in the presence of urea.

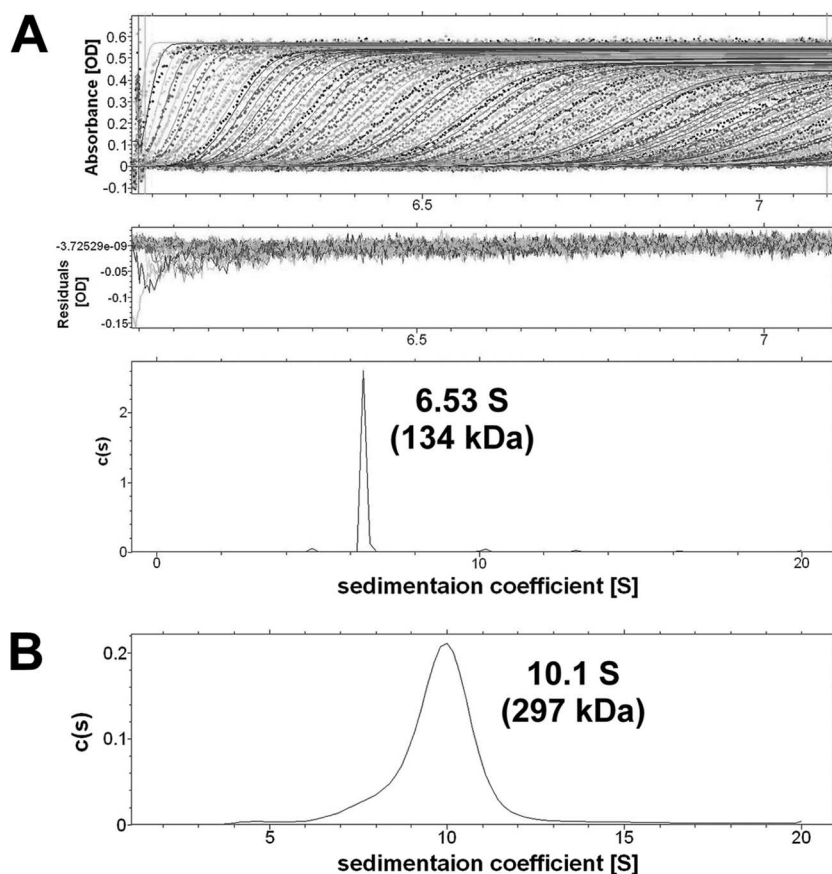


FIG. 4. Sedimentation velocity. Moving boundaries were measured using  $A_{280}$  (20°C). The sedimentation coefficient at the peak top of  $c(s)$  was obtained by SEDFIT analysis, and the molecular mass at the peak top of  $c(M)$  was converted from  $c(s)$ . (Top) Raw data on moving boundaries. (Middle) Residuals between raw data points and the fitted theoretical curve. (Bottom two panels)  $c(s)$ . (A) gp5. The rotor speed was 40,000 rpm. The sedimentation coefficient was  $6.53 \pm 0.14$  S, and the molecular weight was  $134,000 \pm 4,000$ . (B) gp5-ORF334 complex. The rotor speed was 35,000 rpm. The sedimentation coefficient and the molecular weight were  $10.1 \pm 0.4$  S and  $297,000 \pm 14,000$ , respectively.

protein sequencing. The N-terminal 7 amino acid residues were MFMGLDG, which confirmed that the protein was indeed gp5.

The SDS-PAGE indicated that significant amounts of the protein remained at the border between the stacking and separation gels, indicating that purified gp5 formed large aggregates when it was boiled with Laemmli sample buffer (Fig. 3a). Production of similar insoluble aggregates was also observed in the case of gp5 of phage T4 and appears to be a noted characteristic of proteins having the C-terminal triple-stranded  $\beta$ -helix. To confirm that the aggregated protein was gp5, 4 M urea was added to the sample solution for SDS-PAGE before it was boiled (Fig. 3a). After such treatment, the bands of the aggregated protein migrated as monomeric gp5, strongly indicating that the aggregates were indeed gp5.

**gp5 forms a trimer rich in  $\beta$ -structure.** In order to establish the nature of the quaternary state of gp5, analytical ultracentrifugation (AUC) and far-UV CD measurements were carried out.

Sedimentation velocity experiments indicated that gp5 exists in solution in a single quaternary state that has a sedimentation coefficient of  $6.53 \pm 0.14$  S. The molecular weight corresponding to this  $s$  value, 134,000, is 2.8 times the value calculated based on the amino acid sequence, indicating that it is a trimer (Fig. 4A). Complemen-

tary sedimentation equilibrium experiments gave a molecular weight of  $145,000 \pm 3,000$ , which is 3.03 times that of the molecular weight of the monomer (data not shown). From these measurements, it was concluded that gp5 is a trimer in solution.

The far-UV CD spectrum of gp5 was measured to estimate the secondary structure of the protein. The secondary-structure content as estimated by the CONTINLL program (24) indicated that gp5 was rich in  $\beta$ -structure ( $\alpha$ -helix, 23%;  $\beta$ -structure, 69%; turn, 15%) (Table 1). As KVP40 gp5 shares significant sequence similarity with T4 gp5 (except for the lysozyme domain [Fig. 1], which is absent in KVP40 gp5), the

TABLE 1. Estimated secondary structure of gp5

Protein	$\alpha$ -helix (%)	$\beta$ -strand (%)	Turn (%)
gp5-KVP40 <sup>a</sup>	23	69	15
gp5-T4, except for lysozyme domain <sup>b</sup>	8	58	1

<sup>a</sup> The secondary structure of gp5 was estimated by the CONTINLL program over the wavelength range 198 to 240 nm.

<sup>b</sup> The estimated contents based on the tertiary structure determined by X-ray crystallography (Protein Data Bank, 1k28).

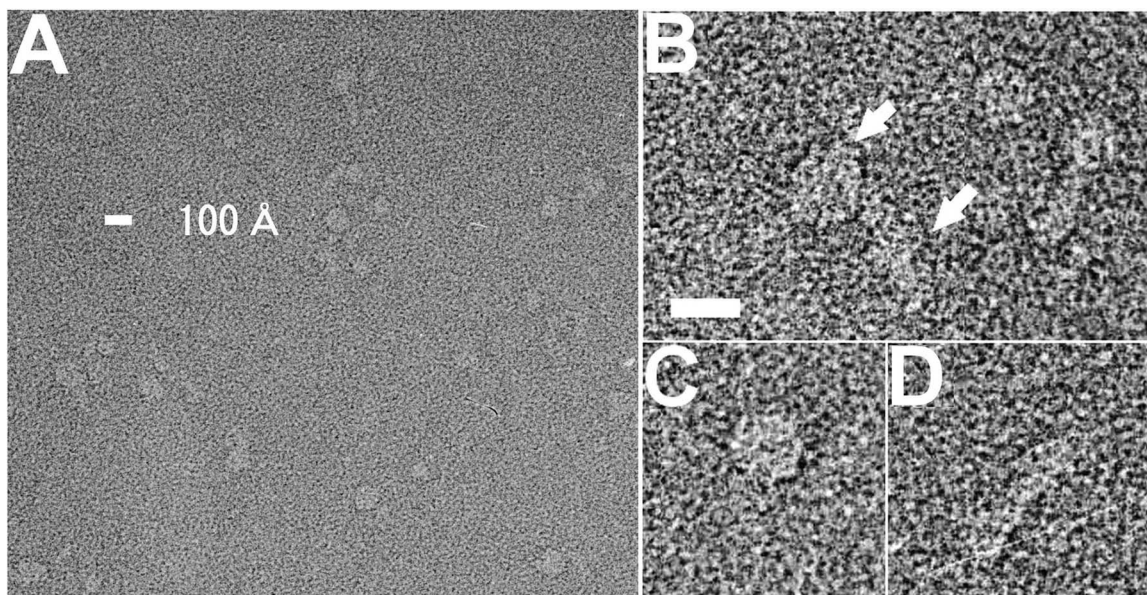


FIG. 5. EM images of the gp5-ORF334 complex. (A) A typical image of the gp5-ORF334 complex. (B to D) Magnified images. (B) Combination of a globe (upper arrow) and a rod (lower arrow). (C) Globe. (D) Rod. The scale bars in panels A and B are 100 Å.

secondary-structure contents of the N- and C-terminal domains of T4 gp5 were calculated based on the reported three-dimensional structure of T4 gp5 (4). This result was used for comparative analysis (Table 1). As can be seen, T4 gp5 and KVP40 gp5 showed similar secondary-structure contents, except for the lysozyme domain.

**Interaction between gp5 and ORF334.** When ORF334 was overexpressed in BL21(DE3) cells, it aggregated and was mostly present in the insoluble fractions. In order to check if this aggregation occurred after cell sonication, ORF334-expressing cells and gp5-expressing cells were mixed, resuspended, and then sonicated to make the lysate. In this case, some portion of ORF334 associated with gp5 and coeluted from the Ni column, but the yield of the complex was low compared with the amount of expressed proteins. We therefore constructed another expression vector, pMNC, for coexpression of ORF334 and gp5 within a single bacterial host (see Materials and Methods). Using this expression pathway, the soluble fraction containing both gp5 and ORF334 increased about 30%, which was enough to continue purification. The complex thus obtained was purified with Ni affinity, anion-exchange, and gel filtration chromatography (Fig. 3b). To confirm that the purified complex was the gp5-ORF334 complex, the ORF334 band in SDS-PAGE was cut out, and the N-terminal amino acid sequence was determined with a protein sequencer (see Materials and Methods). The amino acid sequence of the N-terminal first 7 amino acids was MFEMRSA, which is identical to that of ORF334.

**The oligomerization state of the gp5-ORF334 complex.** In order to determine the association state of the gp5-ORF334 complex, sedimentation velocity experiments were carried out using the peak fraction isolated with the combination expression system. The results of the data analysis using the program SEDFIT are shown in Fig. 4B. There is some shoulder to the left, but the median peak  $s$  value was  $10.1 \pm 0.4$  S. When  $c(s)$

was converted into  $c(M)$ , the peak molecular weight was  $297,000 \pm 14,000$ , which is close to the expected molecular weight of the heterohexamer,  $(gp5)_3(ORF334)_3$ , namely, 287,000. In order to confirm the molecular weight of the complex, sedimentation equilibrium experiments were carried out. The result of the analysis using a single-species model based on nine datasets (see Materials and Methods) gave the molecular weight of  $301,000 \pm 12,000$ , which added further weight to the heterohexamer model (data not shown).

**Observation of the complex by EM.** The gp5-ORF334 complex was negatively stained and examined by EM (Fig. 5A). The gp5-ORF334 complex appeared similar in shape to the gp5-gp27 complex of phage T4 (Fig. 5B), in which the total length of the globe and the rod is about  $253 \pm 27$  Å. Figure 5C and D very likely represent the dissociation products of the globe and the rod, respectively, where the diameter of the globe and the length of the rod are about  $110 \pm 10$  Å and  $190 \pm 20$  Å. The longer rod measurement is consistent with the fact that ORF334 is predicted to have 17 repeats of VXGXXXXX in the  $\beta$ -helix motif compared with 12 such repeats in T4 gp5, which is 110 Å (Fig. 1).

## DISCUSSION

T4 phage gp5 is a remarkable protein. It possesses two domains that are important for infection, namely, the lysozyme domain, which degrades peptidoglycan locally at the phage adsorption site, and the triple-stranded  $\beta$ -helix domain, which punctures the outer membrane of *E. coli*. The highly ordered triple-stranded  $\beta$ -helix of T4 gp5 is based on the regular repeat of 8 residues, VXGXXXXX. A triple-stranded  $\beta$ -helix has also been found in gp12 of phage T4 (the short tail fiber), but it is shorter and more irregular than the triple-stranded  $\beta$ -helix of gp5 (29). When the gp5 proteins from six T4-related phages were aligned, it was found that KVP40 gp5 had significant

homology with T4 gp5 in the N-terminal (46%) and C-terminal (35%) domains. In this work, we found that like T4 gp5, KVP40 gp5 possesses the regular 8-residue repeat for the triple-stranded  $\beta$ -helix. However, the lysozyme domain is apparently absent in KVP40 gp5. In the present study, we have experimentally demonstrated that in a manner identical to that of T4 phage gp5, KVP40 gp5 both exists as a trimer in solution and is rich in  $\beta$ -structure (as evidenced by AUC and CD measurements, respectively). The fact that the purified gp5 forms aggregate upon being boiled with SDS sample buffer is reminiscent of gp5 from phage T4 and further supports the notion that the C-terminal domain of KVP40 gp5 forms a triple-stranded  $\beta$ -helix.

In the T4 phage, gp5 associates with gp27 to form a heterohexamer, (gp5)<sub>3</sub>(gp27)<sub>3</sub> (2). The baseplate assembly in KVP40 has not been investigated, but due to their relatedness, it is expected to be similar to that of phage T4. A BLAST search did not find any ORF in KVP40 sharing significant homology to T4 gp27. On the assumption that such a gp27 homolog existed in KVP40, we chose a number of possible candidates based on the likely ORF size and direction of transcription. From our initial candidate pool (ORFP1sit, ORF339, and ORF334), ORF334 was chosen as the most likely candidate. ORF334 alone was cloned, expressed, and purified. In this work, we demonstrated that ORF334 interacted with gp5 to form a heterohexamer (Fig. 3 and 4). Furthermore, in a similar fashion to gp27, purified ORF334 tends to form aggregates upon standing in solution, suggesting it has somewhat similar properties.

AUC experiments indicated that the gp5-ORF334 complex was not as stable as the gp5-gp27 complex derived from phage T4. This result may indicate that the heterohexamer requires further stabilization by other baseplate proteins involved in subsequent stages of baseplate assembly. The EM images revealed that the KVP40 gp5-ORF334 heterohexameric complex formed a globe (ORF334) binding a rod (gp5), similar to the structures observed in EM micrographs of the gp5-gp27 complex derived from phage T4. The top view of the globe-like structure is shown in Fig. 5C. At 110 Å, the diameter of the globe is slightly larger than that of the corresponding structure in T4 (4). Upon closer examination, the globe appears to be a hexagon containing a central hole. We suggest that this hole is likely to be the center of the ORF334 trimer, as is the case with the T4 gp27 trimeric complex (2, 4). Based on the above observations, we concluded that ORF334 of KVP40 is very likely a homolog of T4 gp27. The amino acid homology between KVP40 ORF334 and T4 gp27 reveals only 15% identity. Such a low value is commonly taken as showing no relatedness; however, such an assessment based on the linear sequence information gives no consideration to the existence of possible three-dimensional structural similarities. Recently, the crystal structure of gp44 from phage Mu was reported (12). This gp44 three-dimensional structure was extremely similar to that of T4 gp27, and the root mean square deviation of the distances between equivalent C $\alpha$  atoms was 2.7 Å. It is also known that both proteins form trimers. Such a tertiary- and quaternary-based structural similarity was unexpected, because (i) Mu and T4 belong to different subgroups of the family *Myoviridae*, (ii) Mu does not have a gp5-like protein, and (iii) the amino acid homology between the two phage

proteins is only 14%. This finding reinforces the importance of comparing the three-dimensional structures rather than just the amino acid sequences. In future, we plan to apply X-ray crystallography to solve the structure of the gp5-ORF334 complex so that we may further prove that ORF334 is indeed the homolog of T4 gp27.

Phage KVP40 is categorized as a Schizo T-even phage (28). Although about 30% of all 386 ORFs in the KVP40 genome have some similarity to those of phage T4, the homology at the amino acid level is less than 50%. The remaining 70% of KVP40 ORFs have no significant homology at the amino acid sequence level with any estimated gene products in the database (21, 23). With regard to the baseplate genes, there is some variety in the locations of gene clusters in the genomes of the KVP40 and T4 phages. In general, functionally related genes form a cluster in the genome. In the case of T4, clusters of related genes are apparent, and the baseplate genes indeed form clusters; however, the cluster which encodes wedge proteins and that which encodes hub proteins are separated by head gene clusters (20) (Fig. 2). It has also been noted that gene 5, which encodes one of the hub proteins, is located in the wedge cluster, and gene 25, a gene that encodes a wedge protein, is located in the hub cluster. Such an arrangement of genes is conserved among the phages closely related to T4, such as RB69 (T-even), RB49, and 44RR2.8t (Pseudo T-even). On the other hand, distantly related phages, such as KVP40 and Aeh1 (Schizo T-even) (23), have only a single cluster of the baseplate genes, and the swapped gene cluster arrangement for gene 5 and gene 25 as seen in phage T4 is not observed in KVP40 or Aeh1 (For T-even, Pseudo T-even, Schizo T-even, and Exo T-even, see the introduction). Both KVP40 and the Aeh1 phage have a larger genome than T4, namely, 245 kbp and 233 kbp, respectively, whereas that of T4 is 169 kbp. Exo T-even phages are more distant from T4 than other T4-related phages and up to now have only included the cyanophages. The inclusion of this group in the T4-related phages is solely based on the similarities of gp23, the head major protein, and gp18, the contractile tail sheath protein. A noticeable difference between these Exo T-even phages and other T4-related phages is that the head is not elongated but rather icosahedral. The homolog of T4 gp5 in these phages is not clear, but orf211 of S-PM2 encodes a protein whose N-terminal domain shows significant homology to T4 gp5. It is a matter of considerable interest whether ORF211 plays the role of the structural protein in the baseplate complex in a manner analogous to the role of T4 gp5. We have cloned orf211 into an expression vector and plan to isolate the gene product for crystallization and subsequent X-ray-based structural analysis.

Recently, Pukatzki et al. reported that the type 6 secretion system of *Vibrio cholerae* secretes (extracellularly) three related proteins, VgrG-1, VgrG-2, and VgrG-3, and that they are structurally related to the gp5-gp27 cell-puncturing device of bacteriophage T4 (25). In these VgrG proteins, the N-terminal domain resembles gp27 and the middle domain resembles the C-terminal  $\beta$ -helix domain. These proteins do not have the corresponding oligonucleotide/oligosaccharide binding fold in the N-terminal domain or the lysozyme domain of gp5 (25). In this regard, it is interesting that the KVP40 gp5-ORF334 complex resembles the T4 gp5-gp27 complex yet the lysozyme domain is missing in the KVP40 gp5 protein. *V. cholerae* is one

of the hosts of phage KVP40, and the VgrG proteins secreted from the type 6 secretion system are highly conserved in many pathogenic gram-negative bacterial species, including *E. coli*, which is the host of phage T4. In light of the work presented in this paper, it is fascinating to speculate that the cell-puncturing devices of T4-related bacteriophages and the type 6 secretion system of the host bacteria might in some way be related through the course of evolution. We plan to continue investigating this intriguing possibility.

In summary, we have shown that KVP40 gp5, like T4 gp5, forms a trimer in solution, despite the fact that it lacks the lysozyme domain. No T4 gp27 homologs were detected in KVP40 in a BLAST search; however, KVP40 ORF334, an ORF resident in the baseplate gene cluster, was shown to form a heterohexameric complex with KVP40 gp5 in a manner highly similar to the T4 gp5-gp27 system. Recent discoveries, such as the finding (12) that Mu phage gp44 has a quaternary structure nearly identical to that of T4 gp27 despite sharing no significant sequence homology and the finding (25) that the *V. cholerae* secretion system is structurally similar to the gp5-gp27 cell-puncturing device of bacteriophage T4, further reinforce the importance of structure-based comparative studies (such as the present effort) for the investigation of the phylogeny and origin of bacteriophages.

#### ACKNOWLEDGMENTS

This work was supported in part by a Grant-in-Aid for Scientific Research in Priority Areas (no. 16087204) and Scientific Research (C) (no. 18570147) to F.A. and a Grant-in-Aid for Young Scientist (A) (no. 17687014) to S.K. from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. This work was supported in part by the 21st Century COE Program "How To Build Habitable Planets," Tokyo Institute of Technology, sponsored by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan.

We thank Damien Hall for assistance with improving the clarity of expression.

#### REFERENCES

1. Arisaka, F. 2005. Assembly and infection process of bacteriophage T4. *Chaos* **15**:047502.
2. Arisaka, F., S. Kanamaru, P. Leiman, and M. G. Rossmann. 2003. The tail lysozyme complex of bacteriophage T4. *J. Biochem Cell Biol.* **35**:16–21.
3. Harbly, E., F. Tetart, C. Desplats, W. H. Wilson, H. M. Krisch, and N. H. Mann. 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. *Proc. Natl. Acad. Sci. USA* **98**:11411–11416.
4. Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann. 2002. Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**:553–557.
5. Kanamaru, S., Y. Ishiwata, T. Suzuki, M. G. Rossmann, and F. Arisaka. 2005. Control of bacteriophage T4 tail lysozyme activity during the infection process. *J. Mol. Biol.* **346**:1013–1020.
6. Kao, S. H., and W. H. McClain. 1980. Baseplate Protein of Bacteriophage T4 with both structural and lytic function. *J. Virol.* **34**:95–103.
7. Kao, S. H., and W. H. McClain. 1980. Roles of bacteriophage T4 gene 5 and gene 8 products in cell lysis. *J. Virol.* **34**:104–107.
8. Kikuchi, Y., and J. King. 1975. Genetic control of bacteriophage T4 baseplate morphogenesis. 1. Sequential assembly of the major precursor, in vivo and in vitro. *J. Mol. Biol.* **99**:645–672.
9. Kikuchi, Y., and J. King. 1975. Genetic control of bacteriophage T4 baseplate morphogenesis. 2. Mutants unable to form the central part of the baseplate. *J. Mol. Biol.* **99**:673–694.
10. Kikuchi, Y., and J. King. 1975. Genetic control of bacteriophage T4 baseplate morphogenesis. 3. Formation of the central plug and overall assembly pathway. *J. Mol. Biol.* **99**:695–716.
11. King, J. 1971. Bacteriophage T4 tail assembly: four steps in core formation. *J. Mol. Biol.* **58**:693–709.
12. Kondou, Y., D. Kitazawa, S. Takeda, Y. Tsuchiya, E. Yamashita, M. Mizuguchi, K. Kawano, and T. Tsukihara. 2005. Structure of the central hub of bacteriophage Mu baseplate determined by X-ray crystallography of gp44. *J. Mol. Biol.* **352**:976–985.
13. Kostyuchenko, V. A., P. G. Leiman, P. R. Chipman, S. Kanamaru, M. J. van Raaij, F. Arisaka, V. V. Mesyanzhinov, and M. G. Rossmann. 2003. Three-dimensional structure of bacteriophage T4 baseplate. *NaT. Struct. Biol.* **10**:688–693.
14. Kozloff, L. M., and J. Zorzopulos. 1981. Dual functions of bacteriophage T4D gene 28 product: structural component of the viral tail baseplate central plug and cleavage enzyme for folyl polyglutamates. *J. Virol.* **40**:635–644.
15. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
16. Laue, T. M., B. D. Shah, T. M. Ridgeway, and S. L. Pelletier. 1992. Computer-aided interpretation of analytical sedimentation data for proteins, p. 90–125. *In* S. E. Harding, A. J. Rowe, and J. C. Horton (ed.), *Analytical ultracentrifugation in biochemistry and polymer science*. Royal Society of Chemistry, Cambridge, United Kingdom.
17. Maniloff, J., and H. W. Ackermann. 1998. Taxonomy of bacterial viruses: establishment of tailed virus genera and the order Caudovirales. *Arch. Virol.* **143**:2051–2063.
18. Matsuzaki, S., T. Inoue, and S. Tanaka. 1998. A vibriophage, KVP40, with major capsid protein homologous to gp23\* of coliphage T4. *Virology* **242**:314–318.
19. Meezan, E., and W. B. Wood. 1970. The sequence of gene product interaction in bacteriophage T4 tail core assembly. *J. Mol. Biol.* **58**:685–692.
20. Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger. 2003. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**:86–156.
21. Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum, O. White, I. T. Paulsen, W. C. Nierman, J. Lee, B. Szczypinski, and C. M. Fraser. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J. Bacteriol.* **185**:5220–5233.
22. Nakagawa, H., F. Arisaka, and S. Ishii. 1985. Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. *J. Virol.* **54**:460–466.
23. Nolan, J. M., V. Prteov, C. Bertrand, H. M. Krisch, and J. D. Karam. 2006. Genetic diversity among five T4-like bacteriophages. *Virol. J.* doi:10.1186/1743-422X-3-30.
24. Provencher, S. W., and J. Glockner. 1981. Estimation of protein secondary structure from circular dichroism. *Biochemistry* **20**:1085–1094.
25. Pukatzki, S., A. T. Ma, A. T. Revel, D. Sturtevant, and J. J. Mekalanos. 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl. Acad. Sci. USA* **104**:15508–15513.
26. Schuck, P. 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**:1606–1619.
27. Schuck, P., M. A. Perugini, N. R. Gonzales, G. J. Howlett, and D. Schubert. 2002. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys. J.* **82**:1096–1111.
28. Tetart, F., C. Desplats, M. Kutateladze, C. Monod, H.-W. Ackermann, and H. M. Krischi. 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. *J. Bacteriol.* **183**:358–366.
29. Thomassen, E., G. Gielen, M. Schütz, G. Schoehn, J. P. Abrahams, S. Miller, and Mark J. van Raaij. 2003. The structure of the receptor-binding domain of the bacteriophage T4 short tail fiber reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331**:361–373.
30. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.