gpwac of the T4-Type Bacteriophages: Structure, Function, and Evolution of a Segmented Coiled-Coil Protein That Controls Viral Infectivity

A. Letarov,† X. Manival,‡ C. Desplats,§ and H. M. Krisch*

Laboratoire de Microbiologie et Génétique Moléculaire du CNRS, UMR 5100, Toulouse, France

The *wac* gene product (gpwac) or fibritin of bacteriophage T4 forms the six fibers that radiate from the phage neck. During phage morphogenesis these whiskers bind the long tail fibers (LTFs) and facilitate their attachment to the phage baseplate. After the cell lysis, the gpwac fibers function as part of an environmental sensing device that retains the LTFs in a retracted configuration and thus prevents phage adsorption in unfavorable conditions. A comparative analysis of the sequences of 5 *wac* gene orthologs from various T4-type phages reveals that the ~50-amino-acid N-terminal domain is the only highly conserved segment of the protein. This sequence conservation is probably a direct consequence of the domain's strong and specific interactions with the neck proteins. The sequence of the central fibrous region of gpwac is highly plastic, with only the heptad periodicity of the coiled-coil structure being conserved. In the various gpwac sequences, the small C-terminal domain essential for initiation of the folding of T4 gpwac is replaced by unrelated sequences of unknown origin. When a distant T4-type phage has a novel C-terminal gpwac sequence, the phage's gp36 sequence that is located at the knee joint of the LTF invariably has a novel domain in its C terminus as well. The covariance of these two sequences is compatible with genetic data suggesting that the C termini of gpwac and gp36 engage in a protein-protein interaction that controls phage infectivity. These results add to the limited evidence for domain swapping in the evolution of phage structural proteins.

Modular swapping of genetic material is thought to play an important role in the evolution of phages (3, 20). Such swaps can involve either large blocks of genes, such as those encoding entire phage tail or head structures, or much smaller units, such as a single gene. It was demonstrated more than 20 years ago that lambdoid phages often exchange large modules encoding groups of genes with related function (3, 8). Similarly, in the virulent T4-type phages, the loss, acquisition, or replacement of segments of DNA also occurs (11, 12, 22) but in these phages modular exchanges usually involve smaller units, such as individual genes or protein domains. In the T4-type phages, there is extensive and convincing evidence for modular shuffling of domains in the tail fiber proteins and their associated adhesins (37). These proteins have a critical role in determining phage host range; they directly interact with the receptor molecules on the surface of the bacterial cell (37, 40). The T4 virion actually contains three sets of fibrous structures with distinct functions: the long tail fibers (LTFs), which are involved in reversible adsorption; the short tail fibers, which are involved in the final, irreversible step of adsorption; and the gpwac-encoded "collar-whiskers" (40). gpwac forms a collar

(10, 13) on the phage neck of six radiating 480-Å fibrils that can bind transiently to the LTFs and properly position them for attachment to the baseplate (14, 15, 40). These whiskers also have a potentially important regulatory function; they act as part of an environmental sensing device that prevents phage absorption in conditions unsuitable for infection (e.g., low pH). In such conditions the ends of the whisker fibers are thought to bind to the middle of the LTFs and hold them in a retracted configuration that prevents phage infection (9, 15). Both LTFs and short tail fibers have a similar trimeric parallel β-helical protein structure (26, 39). However, the 486-amino-acid (aa) T4 gpwac polypeptide trimerizes into a coiled-coil α-helical structure that is flanked by small N- and C-terminal globular domains (14, 28, 33). It is the N-terminal 50-aa domain of T4 gpwac that binds to the neck of the phage, whereas the 30-aa C-terminal domain of the protein (the "foldon") is essential for the initiation of trimerization and the correct folding of the polypeptide chains in the fiber (14, 23). The nucleotide sequences of the *wac* genes from a few T-even phages very closely related to T4 have been determined previously (M. M. Schneider et al., unpublished data; our unpublished results), but the amino acid sequences of all of these proteins are virtually identical to that of T4 gpwac. Recently, a phylogeny for the T4-type phages has been established that includes a number of much more distant phages (18, 27, 38). Four subgroups of T4-type phages can be distinguished (18, 38): the T-evens (very closely related to T4), the PseudoT-evens (morphologically very similar to T4 but phylogenetically distant from it) and SchizoT-evens (having limited but clear morphological differences from T4 and phylogenetically quite distant from it). The fourth subgroup, the ExoT-evens, are phages of marine cyanobacteria that have a virion morphology that dif-

---

* Corresponding author. Mailing address: Laboratoire de Microbiologie et Génétique Moléculaire, CNRS UMR 5100, 118 Route de Narbonne, 31062 Toulouse Cedex, France. Phone: (33) 5-61-33-58-81. Fax: (33) 5-61-33-58-86. E-mail: krisch@ibcg.biotoul.fr.
† Present address: S. N. Winogradsky Institute of Microbiology, Russian Academy of Science, 117312 Moscow, Russia.
‡ Present address: Centre de Biochimie Structurale, CNRS-UMR 9955, INSERM-U554, Universite de Montpellier I, F-34090 Montpellier, France.
§ Present address: Départment de Biochimie Médicale, University Medical Centre, 1211 Geneva 4, Switzerland.

fers considerably from T4 (isometric heads and longer contractile tail structures); nevertheless, many of the ExoT-even genes are homologs of T4 genes (18). Surprisingly, in view of the modular shuffling model for phage genomes, initial genomic sequencing (12, 18; unpublished results) revealed a homogeneous level of protein sequence divergence among the various "core" genes within the members of the T4 phage family. Thus, a phylogeny of the T4-type phages based on the major capsid protein reflects reasonably well the relatedness of almost all of their structural proteins and many of the nonstructural proteins.

We have now determined the *wac* gene sequence from both PseudoT-even and SchizoT-even phages. The N-terminal globular domain that interacts specifically with the neck structure is largely conserved while the central domain diverges considerably in amino acid sequence, retaining only the repeated heptad motifs of the fibrous structure. In contrast to our expectations, however, the C-terminal domain, which is essential for efficient initiation of T4 gpwac folding, was frequently subjected to modular replacement. All of the distant *wac* gene sequences had different and unrelated C-terminal domains. The possible role of the modular swapping of this domain is discussed here.

MATERIALS AND METHODS

Phages and bacteria. All of the T4-type phages are from the Toulouse collection (>80 isolates). Phages T4 and T6 were obtained from R. H. Epstein of the University of Geneva in Switzerland. The sources of all of the other T4-type phages whose wac sequence was studied (coliphages RB6, RB11, RB16, RB17, RB18, RB27, RB33, RB42, RB43, RB49, RB69, and Tu1a; *Enterobacter cloacae* phage 1; *Aeromonas salmonicida* phages 44RR2.8t [synonym 44RR] and 65; *Aeromonas hydrophila* phage Aeh1; *Burkholderia cepacia* phage 42; *Vibrio natriegens* phage nt-1 and *Vibrio parahaemolyticus* phages KVP20 and KVP40) are found in references 1 and 37. For phage that grow on *Escherichia coli* B^E, standard techniques were used for their propagation and the preparation of their DNA (27). The T4-type phages of non-*E. coli* hosts were grown as suggested in the information sheets provided with the phages by the Félix d'Hérelle Reference Center for Bacterial Viruses (Quebec, Quebec, Canada).

PCRs, oligonucleotide primers, and sequencing. PCR amplification of the DNAs was performed with the Long Expand kit (Roche, Penzberg, Germany) according to the instructions provided by the manufacturer. When inosine-containing primers were used, the PCR amplifications were done with HotTub polymerase (Amersham). For degenerate primers, a 50°C hybridization temperature was used. For specific primers the hybridization temperatures were chosen to match the lowest T_m of the two primers used.

Sequencing was done on the Beckman Seq2000XL automated sequencer by using Beckman DTCS standard chemistry kits according to the supplier's recommendations.

The primers used to amplify the fragments between gene *10* and the previously determined sequence (accession number Z78092) upstream of the gene *13* in bacteriophage RB49 were as follows: G10dir (5′-TCACAATTCGCTGGTTTA ATAATGAT-3′) and 13Wrev (5′-TTATACCGGAGAAATCCCAAAGAGG-3′). The "consensus" inosine-containing primers designed for amplification of gene *10* to gene *13* fragments from various PseudoT-even and SchizoT-even phages were AL10d (5′-CCTICACAIACCGCAGGTGG-3′) and AL13r (5′-A GAAAGTCIGTIAACCACGGATA-3′). With these primers fragments of 4.5 kb were obtained for phages RB42 and RB43, as well as for T4 and RB49 used as a control. For some other distant T4-type phages, no amplification was obtained. Thus, to obtain the sequences in the region from gene *10* to gene *13* of phage 42, we had to use a more complicated approach. We had previously obtained a number of sequences of random PCR fragments of the genome phage 42 (details are available from the authors). We selected sequences flanking the region from gene *11* to gene *13* and used these to synthesize the primers 42-11d (5′-TTTA CTAGTGTCGAAGCTCC-3′) and 42-13R (5′-TCTAGGGACCGCTGTATA-3′).

The "universal" primers used in the attempt to amplify the same region in PseudoT-evens and SchizoT-evens, as well as several different T-even phages

(see Results), were Tev-10d (5′-TTTGTCG(T/C)TGGATAAGG(A/G)T(A/T) G-3′) and Tev-13r (5′-CTAGGGCACGCTGNATACAATC(G/A)TA-3′).

To amplify the fragment from phage Aeh1 from gene *10* to gene *18*, we used the primers A10d (5′-AATTCGATCCAGACGATCAG-3′) and A18R (5′-GA TATAGAGCAGAGGAGTTAC-3′). These primers had Aeh1 gene*10* and gene*18* sequences obtained by a strategy (see below) involving sequencing of random genome fragments. The PCR amplification of the Aeh1 genome using these primers yields a fragment >14 kb, so we digested this large fragment with HindIII and ligated the digest to the DNA of the pBluescript SK(+) vector digested with the same enzyme. We then performed a PCR with this ligation as a template and A10d and standard M13 reverse as the primers. We obtained an 8-kb fragment that was sequenced from the M13 reverse primer. From the sequence obtained, we designed the new primer AH3R (5′-TAAAGGTGTCC GTCTGGATA-3′) that was used together with A10d primer to generate an 8-kb fragment from gene *10* to the point of hybridization of the AH3R primer. This fragment was sequenced by primer walking, and one of the primers generated for this purposes, Awd6 (5′-CTCAACGGTAAAGTTATGGAT-3′), was used with A18R primer to obtain 7-kb fragment that was partially sequenced by the primer walking from Awd6 primer to the beginning of the N terminus of the gene *13* homolog.

The primers used to amplify the RB 49 *wac* sequence for cloning in the expression vector were Wstart-NcoI (5′-GGGTGCGC**CCATGG**TTGAACAAT TAAACATCC-3′) and WBamHI-rev (5′-TCTTG**GGATCC**TGGTTACTATCC TGCA-3′). The restriction sites are in boldface. For the C-truncated deletion variant of RB49 *wac* (RBW-N), we used the primer combination Wstart-NcoI and F-stop-W, which introduces a BamHI site (5′-CTCAAG**GGATCC**TCAAA TATCTGCGGTATTTTTAAATGC-3′). To create the N terminally truncated deletion variant of the same gene (RBW-CE), we used the primer WBamHI-rev and the primer WEdir that introduces an NcoI site (5′-CTCTAG**CCATGG**GT ATTAAAGTTGTAGAAAACACTC-3′). To clone the phage 42 *wac* into an expression vector, we used the primers PCW-NcoI (5′-GGGTTCAG**CCATGG** AAATTCTTCCATTTGTAAATAGC −3′) and PCW-BamHI (5′-ACCTCGA T**GGATCC**GAAGGCCCTAA-3′).

To amplify, clone, and to sequence gene *36* of the phage T4, we used the primers T436NcoD (AAGGGGCATACCATGGCTGA) and T436BamR (CAT**GGATCC**TCTTAATAATAG CCGA). To create the amber mutation *am*S197 in gene *36* of phage T4, we used the primers T436CamD (ATCCTGC ATAGCAACCCTCACA) and T436CamR (GAGGGTTGCTATGCAGGATT CT).

Random sequencing procedure. To obtain the sequences flanking the regions of interest in the T4-type genomes, we constructed random genome libraries and sequenced about 30 fragments per genome. To construct the library, the samples containing 1 to 5 μg of the phage genome DNA were digested simultaneously with EcoRV and SspI restriction enzymes or, alternatively, with HaeIII. All of these enzymes produce blunt-ended fragments. The DNA was purified from restriction mixtures by using a QuiaGene PCR purification kit and incubated with 3 U of *Taq* polymerase per 10 μl of reaction volume in the presence of 0.2 mM ATP. This treatment leads to a template-independent addition of the single A residue at the 3′ termini of the fragments. The fragments were then cloned according to the instructions of supplier by using the T-System kit for direct cloning of PCR fragments (Promega). The plasmid insertions obtained were sequenced by using standard M13 sequencing reverse primer. In some cases when large insertions were obtained, additional sequencing from opposite end of the insert was done by using the M13 sequencing primer.

The GenBank accession numbers of the nucleotide sequences were as follows: AY266304, *B. cepacia* phage 42 fibritin (*wac*) gene and a part of gene 13; AY266305, enterobacterium phage RB43 fibritin (*wac*) gene; AY266307, enterobacterium phage RB43 genes *36–37.1*, *36–37.2*, and *38*. The sequences from phages RB49 and Aeh1 are now included in the complete genome sequences of these viruses, (AY343333 and AY266303, respectively).

Cloning, protein expression, and protein analysis. The cloning, expression, and protein analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were done as described by Sambrook et al. (31)

Purification of proteins. The same method was used for purification of the phage RB49 gpwac and of its variants. Competent cells (Novagen) of *E. coli* BL21(DE3) were transformed with the appropriate chimeric plasmid. Then, 100 ml of the Luria-Bertani medium plus 0.1% glucose was inoculated with a transformed colony, and the cells were grown at 37°C with agitation for 3 h. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to 1 mM, and the incubation was continued for 2.5 h. The cells were then collected by centrifugation at 6,000 rpm for 15 min and resuspended in 5 ml of 50 mM Tris-HCl (pH 7.5). Lysozyme and DNase were added to final concentrations of 300 and 5 μg/ml, respectively. The extract was frozen and thawed three times, incubated at 20°C
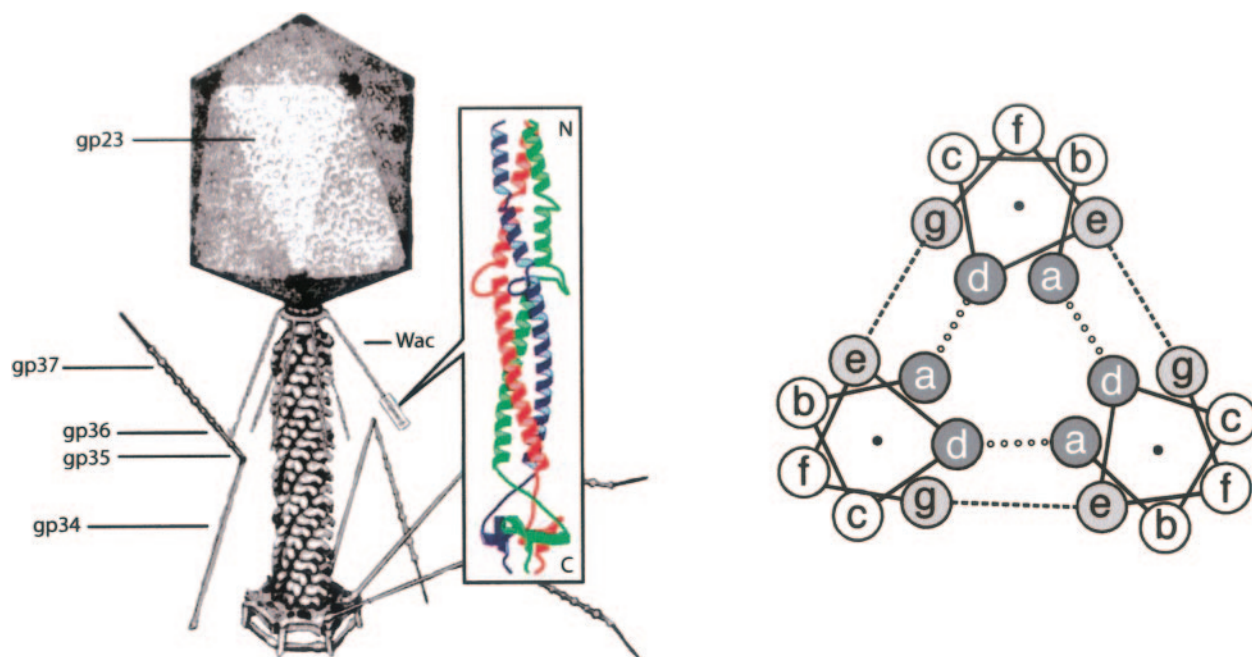
FIG. 1. (A) Bacteriophage T4 virion. One of the LTFs at the left is shown in the course of its attachment. It has bound to the gpwac but not yet been fixed into position on the baseplate. The geometry of the virion and the sizes of its constituent components suggest that the interaction between LTF and gpwac occurs between the "knee" region of the LTF and the C terminus of the gpwac. The enlargement shows the X-ray structure (34) of N-truncated gpwac mutant protein (gpwac E). This coiled-coil trimeric α-helical structure corresponds to the most distal part of the gpwac fiber that is terminated by the small foldon domain at the C terminus. (B) Structural basis of heptad periodicity in coiled-coil α-helical segments. The view shown is along the axis of the α-helices in the triple-stranded parallel-coiled coil. In this superhelical structure, the α-helices are distorted to have only 3.5 aa per turn instead of 3.6 aa. As a result, the hydrophobic residues in positions a and d (dark gray) form the hydrophobic core of the structure. The superhelix could be further stabilized by electrostatic interactions of the charged residues in positions e and g (light gray).

for 30 min, and then treated with an ultrasonic disintegrator for 1 min. Phenyl-methylsulfonyl fluoride was then added to a final concentration of 1 mM, strep-tomycin sulfate was added to a final concentration of 3% (wt/vol), and the extract was placed on ice for 30 min. The nucleic acids and cell debris were removed by centrifugation at 10,000 rpm for 15 min. $(NH_4)_2SO_4$ was slowly added to a final concentration of 25% saturation with continual magnetic stirring. After 30 min of further incubation on ice, the precipitated proteins were obtained by a 15-min centrifugation at 10,000 rpm, and the pellet was resuspended in 1 ml of 50 mM Tris-HCl (pH 7.5). The supernatant's concentration of $(NH_4)_2SO_4$ was then increased to 35% of saturation or until the turbidity increased notably, and the solution was incubated for a further 30 min on the ice. After 20 min on ice the protein pellet was centrifuged and collected as described above. The supernatant was carefully removed with a Pasteur pipette, and the pellet was redissolved in 1 ml of 50 mM Tris-HCl (pH 7.5).

Further purification was performed, in some cases, by chromatography on a hydroxylapatite column. The column containing 3 to 4 ml of hydroxylapatite equilibrated overnight in 10 mM sodium phosphate buffer (pH 7.6) was loaded with 0.5 to 1 ml of the partially purified protein solution, and the protein was eluted by using the following step gradient of sodium phosphate buffer concentration: 10, 20, 50, 100, and 200 mM. All fractions were then analyzed by SDS-PAGE.

Circular dichroism of wild-type phage RB49 *wac* protein. Circular dichroic spectra were recorded at 25°C with a Jobin-Yvon VI dichrograph. A cell with a 1-mm optical path length was used to record the spectra of purified recombinant protein preparations at a peptide concentration of 0.4 mg/ml in 10 or 50 mM phosphate buffer (pH 7.5) in the UV region (260 to 190 nm).

Mutagenesis of phage T4 gene 36. To obtain the gene *36* amber mutation *am*S197, the N-terminal and C-terminal parts of the gene were amplified separately by PCR with primers T4g36CamR and T436NcoD for the N-terminal fragment and primers T4g36CamD and T436BamR for the C-terminal fragment. These PCR fragments were purified and mixed, and an additional 10 cycles of PCR were performed with the primers T436NcoD and T436BamR. The 6.6-kb fragment obtained was treated with *Taq* DNA polymerase in the presence of

dATP and cloned into the pGEM-T vector. Sequencing verified that the amber mutation introduced by the primers T4g36CamR and T4g36CamD was present in this plasmid. This plasmid was used for transformation of the *E. coli* B^E strain. A culture of the plasmid-containing strain was infected with T4D$^+$ phage, and the titer of the lysate was determined on strain CR63, which contains an amber suppressor. Individual plaques were then replicated with toothpicks to lawns of strains CR63 and B^E. About 1% of the progeny phages grew well on CR63 but not on B^E. Phage stocks were grown from two such plaques, after repeated single plaque isolation. The gene *36* sequence of these phages was PCR amplified and then sequenced to confirm the presence of the mutation, which was named T4g36*am*C1.

Test of the ability of the phage mutants to grow in presence of PEG. A liquid culture of *E. coli* B^E was grown for 2 to 3 h with vigorous agitation to a cell density of ca. 10^7, and a sterilized 30% solution of polyethylene glycol (PEG) 6000 was then added to a final concentration of either 5 or 6.4%. The culture was inoculated with phage at a multiplicity of infection of 30 to 50 PFU/cell. The incubation was continued at 37°C with agitation until the lysis of the culture was observed.

RESULTS

Structure and function of the Wac protein of phage T4. Extensive biochemical and structural analysis (24, 40) of the T4 gpwac has defined a clear role for this protein in the T4 life cycle (Fig. 1, left panel). This molecule can self-fold into a tight trimeric structure with an extended α-helical coiled-coil domain (Fig. 1). The atomic structure of its N-terminal globular domain, which interacts with the phage neck, has recently been determined (5). The foldon of the C-terminal domain is required for the assembly and folding of the trimeric T4 gpwac

Domain N-terminal

```
            ↓
T4_wac     --MTDIVLNDLPFVDGPEAEG-SRISWIKNGEEILGADTQYGSEGSMNRPTVSVLRNVE   57
RB69       ----MIELKSLPYVDGPEDEG-KRLNWIKNSEEITGADTLYGSEGVMNRPITEVQRNVE   55
RB49_wac   MIEQLN-IHEIPHVDGVESEG-HRIIWIKNGECLSGATTQTGNDGTLNRAPVAIQENVA   58
RB42_wac   MISKLNPLPKIPFVDGVSDTT-TPINWIINGETLDGAKTKVTNEGSLNRCGVLVQKNAV   60
42         -------MKILPFVNSVPNDG-QRIQWIRNGDALMGASTKFGSDGNLNIATTQVQANVV   52
Aeh1_wac   MAIEKVSQSQLPFVDGVPAVG-TRIDWIQNGEYTDGAVSATDNGGALNRASVKVQKNVD   59
```

```
              Segment 1                    Loop 1                Segment 2
T4_wac     VLDKNIGILKTSLETANSDIKTIQGILDVSGDIEALAQIGINKKDISDLKTLTSEHTEIL  117
RB69       TINDNVKTIAESLDTANADIVTIKSILDVSGDVDALAQIGHNTDDIEVLKHTVNSHGVDI  115
RB49_wac   RVNENAKSTVAKVNEVVTAVNDIKTTLGVISDGSIIDVVNQTVAKVESVEENLGETARTL  117
RB42_wac   QLETNTITQTAKINEVIDQVNLISENLAAISDESVIDKLDQVVADVEVLKVDMTAVQGST  120
42         VLDENLQELNLGFDANTQKIQEIEALLEASGSLDLVQVVQTNTTDIVQLKADSEDAKTRL  112
Aeh1_wac   TINQELNKTKDSLSKVIDLAETHDRILNAQGDG---TNLYDRVTAVEVVNKAQQDRLAKI  116
```

```
                    Loop 2              Segment 3    Loop 3
T4_wac     NGTNNTVDSILADIGPFNAEANSVYRTIRNDLLWIKRELGQYTGQDINGLPVVGNPSSG  176
RB69       LNTEEKLDDTIANIGVVNPETDSVYRTVRNDLLWIKTELGQYTGQDINGVPTEGNESTG  174
RB49_wac   QTTVDAVSELNIKIGDR-PAADPSTRTIYDDISFIKQEMGSYPNFNINGSPDAGSSGSG  176
RB42_wac   ASNTLKIQDLDKEIGTWDASTDPKHRTIRKDIIFLKGEMGAYPGFNENGDVDPTSTGSG  179
42         TSVETTVRENKTMLGEAGPNFVG---TVNSEIEFLKARIGNNTNETVNGTPSPGTSATG  168
Aeh1_wac   ET------DEIEHVAGQEAINKRITDTVDPQITNIKKMLGNYKGFDQDGNPDINAPGLG  169
```

```
           Segment 4a              Loop 4a      Segment 4b   Loop 4b
T4_wac     MKHRIINNTDVITSQGIRLSELETKFIESDVGSLTIEVGNLREELGPKPPSFSQNVYSRL  236
RB69       MKRRIITNSSVLVDQGVRLTELENKFADSDVGALTTEVENLRQEIGPRP-SLTVPVYTRL  233
RB49_wac   IKYRVIQNTAALNDHGRRIVQLENDWQDSDVGELTRQVNKMRSEIGDTSLATQENVYVRL  236
RB42_wac   MKYKIMTNAQAISIHEGRITRLEDDWANSDVGHLTDEVTDLRNEMGPKHLATFESVYVRL  239
42         IKLSINNVILELNAVKEDNLELNTKLDLFDVPSIKQDIVQIRRELGQTPVSQPT-VYERL  227
Aeh1_wac   VKGQVERNWENVVQHDNRIKSLESFWQGDTPDVLIRNTRDLRAEMGPSSEKTDKTIYQRV  229
```

```
             Segment 5            Loop 5        Segment 6      Loop 6
T4_wac     NEIDTKQTTVESDISAIKTSIGYPGNN---SIITSVNTNTDNIASINLELN-QSGGIKQR  292
RB69       SGIDSSISIQTRDIAALKDFVGYPNST---AIKTQVEANRLSISTINSDIN-SPGGIKPR  289
RB49_wac   RTNARNIANLSNEVLEVKKAILFDSIP---DIGTKVTTLGTNYELLNNDVNGVDSGIRPR  293
RB42_wac   NTIGTEITGINSELTSINQYIGRTGTGGTGSISARMTTAESNIATLQTQMNDSATGVLPR  299
42         NSLNNXSNTFSTDIEAIKVKINFDNPI---DIAPRVFTLETNYSTLDNVIN-GPNGITTK  283
Aeh1_wac   KESEDHVTTANKHLTDIDTAIGKIEFPAKYVTSTGTEVTAKDIWEYTKLVRGDVGANQNA  289
```

```
           Segment 7 Loop 7              Segment 8     Loop 8  Segment 9 Loop 9
T4_wac     LTVIETSIGSDDIPSS-IKGQIKDNTTSIESLNGIVGENTSSGLRANVSWLNQIVGTDSS  351
RB69       LTTLETTIGSPDLPTT-LQGKIKLNTDSISGINTVLGVDSSSGLRFNVAWLNQVVGVDSN  348
RB49_wac   LLAVENKIGTLQSAGS-IDYRLTNTEREFLDLKQVVGVSGADGLRGEVVRLSNEIGTDSD  352
RB42_wac   VISLENAVGNSQIPGS-IRYDISAIKRDIGDINMIVGESSDDGLRGEMAQVMTDIGTDSQ  358
42         VTVLESEMALVKTSG------VDTLELQVDTIDSTLDG--ASGLTTQVQEIKTYIGIVPN  335
Aeh1_wac   LNTAINDILLNKSDIKTMKEFDTSSGAKITNLESLVGLQPTDGMRGDIKALDLRESNNAT  349
```

```
              Segment 10                  Loop 10 Segment 11
T4_wac     GGQPSPPGSLLNRVSTIETSVSGLNNAVQNLQVEIGNNSAGIKGQVVALNTLVNGTNP  409
RB69       GGQPEPAGSLLYRTRILETGVTDLGNNIQNVQTELGTNSSGIKGQVTSLNKLISGTNP  406
RB49_wac   -------RHTLAGRIKVVENTQGRLEASVQDMEQIVGDRESGLTAGVIKLSSDIYGDDG  404
RB42_wac   ------PTSIKGRLLNVENTQRDDHQKLIDVESVVGNTSSGLVAANIAMGKAVYGDAT  410
42         G-TPVPPTSLEGRFITLQGIQNNTASTVQDLQVEVQDLRSQLTALTARVAALESNP*  390
Aeh1_wac   RLEGMVKTVTDVVNLETTPAVRNLQEDMTKAKPDI♦  384
```

```
           Loop 11          Segment 12
T4_wac     NGSTVEERGLTNSIKANETNIASVTQEVNT-AKGNISSLQGDVQALQEAGYIPEA  463
RB69       NGQTIEERGILPTVKNHDTSIMALTTRVTT-LETDLAAAEAEIQALKEAGYIEDA  460
RB49_wac   SEDPFTQKGVIKIVKELQTSNGN--------------------------KVDDV  432
RB42_wac   SSDPFLKDGIQKTARDSKAAIGVNTTGSETGIYKLISDLTARVAALES♦  458
```

```
T4_wac     PRDGQAYVRKDGEWVFLSTFLSPA*  487
RB69       PSDGKFYVRKDGAWVELPTA*  480
RB49       PDDGFHYLRKRGEWVQVA♦  450
```
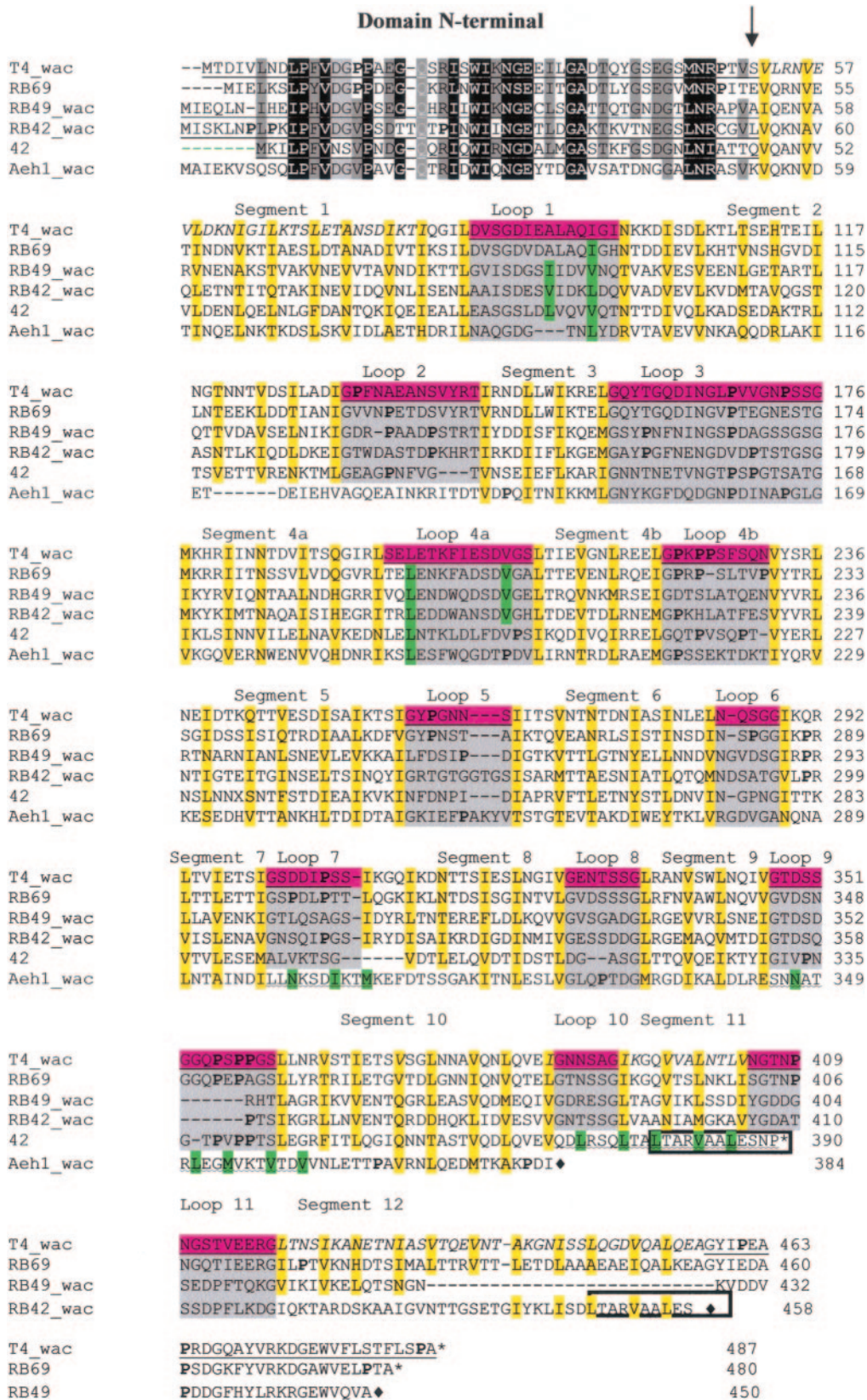
FIG. 2. Alignment of the amino acid sequences of the N-terminal domains and coiled-coil shafts of the gpwacs of the T-even phages T4 and RB69; the PseudoT-even phages RB42, RB49, and 42; and the SchizoT-even phage Aeh1. This alignment was done by using the CLUSTAL W program (http://www2.ebi.ac.uk/clustalw/). Shading indicates the conserved amino acids of the N-terminal domain that are identical or similar in at least four of the six sequences. In the region of the coiled-coil shaft the positions of conserved heptad residues *a* and *d* are

fibril. The structure of truncation mutants containing only the last portion of the coiled-coil domain has also been determined at the atomic resolution level by X-ray crystallography (34, 36).

The exact site on T4 gpwac that interacts with the T4 LTF remains uncertain. The physicochemical data suggest that the foldon itself is not directly involved since the C-terminal domain's structure is not altered by changes in the chemical environment (4). However, the coiled-coil domain responds to modifications of both the pH and the ionic strength. Since the interactions of gpwac and LTF are sensitive to such environmental influences, it seemed probable that the site of binding to the LTF is located either in the most distal coiled-coil segments or in the intervening loops.

Genomic sequence of the region from g*10* to g*13* of the T-even, PseudoT-even, and SchizoT-even phages. Using a variety of consensus and unique PCR primers (see Materials and Methods), we isolated and sequenced the genomic segment (g*10* to g*13*) that contains the *wac* gene of diverse T4-type phages. These included a number of T-even phages, as well as the PseudoT-even phages RB49, 42, RB42, and RB43. Using a genome scanning technique (12), we also obtained the sequence of the gpwac region of the SchizoT-even phage genome Aeh1. In all cases, sequences homologous to the T4 phage genes *11*, *12*, and *wac* were found in the segment between the genes *10* and *13*.

Conservation of the N-terminal domain of gpwac. The N-terminal domain of the *wac* gene was conserved in all of the phages we analyzed. For example, the N-terminal domain of phage 42 wac has 48% sequence identity to T4 wac, whereas the downstream coiled-coil domain has just 27% identity (Fig. 2).

Periodicity of the hydrophobic residues in the central domain of gpwac. The formation of a coiled-coil structure requires that hydrophobic residues occupy positions *a* and *d* in a series of repeated motifs containing seven amino acids (*abc-d*efg). The α-helices in a coiled-coil structure are strained, so that there are only 3.5 aa per helix turn rather than the normal 3.6 aa per turn. As a consequence, the residues *a* and *d* in the heptad (indicated in Fig. 2 by yellow shading) form a hydrophobic band along one side of the helix. In the intertwined trimeric helical structure, these residues form a hydrophobic core that stabilizes the fibril (Fig. 1, right panel).

In T4 gpwac, there are 13 sequence segments that could form coiled-coil domains. However, this T4 coiled-coil structure is interrupted by looped-out sequences (33) that lack the heptad periodicity; these loops frequently begin with a Gly residue and often contain Pro residues. In T4 the structure and position of only the last two C-terminal loops has been confirmed by X-ray crystallography (36). The amino acid sequences of the coiled-coil domain of the Pseudo-even and SchizoT-even gpwac's diverge substantially from those of T4,

although the heptad periodicity is well conserved (Fig. 2). Furthermore, the positions of the interruptions in this periodicity can usually be aligned to the loop positions predicted in the T4 gpwac (33, 36). In the gpwac of phage 42 the T4 loop 11 is deleted (Fig. 2) and in phage Aeh1 there are additional perturbations of the loop-coil pattern.

The divergence of the central fibrous domain of the gpwac protein is consistent with this part of the protein not being involved in strong protein-protein interactions. The only constraints on this sequence appear to be the periodicity of the hydrophobic residues. The presence of different hydrophobic amino acid residues in position *a* and *d* of the heptad can influence the number of α-helixes forming the superhelix, but the C-terminal domain may superimpose the trimeric fiber organization. Thus, the amino acid interactions within the coiled-coil domain could be less important in determining the oligomeric structure of the fibril than in polypeptide structures that form without a nucleating foldon domain. This could also explain why T4 gpwac has significantly more coiled-coil structure than is predicted by the COILS program (http://www.ch.embnet.org/software/COILS_form.html).

To confirm that the wac fibrils from one of the distant T4-type phages were composed of gpwac trimers, we performed in vitro refolding experiment with an equimolar mixture of a gpwac extract obtained from RB49 wild type and from a small wac-deletion mutant. The SDS-PAGE analysis of the unheated refolded complexes (see below) indicated that, as in T4, the RB49 fibrils are composed of gpwac trimers (data not shown).

Similarly, we were able to show that these novel wac genes were functional and are not pseudogenes. Antibody was made to the cloned Wac protein from phage Aeh1, and this was shown in a Western blot to interact with the same-size protein present in purified Aeh1 virions (data not shown). This result demonstrates that the Aeh1 wac gene is functional. It could still be argued that the more bizarre version of wac that we sequenced was not the functional gene copy and that another, functional, copy of wac might exist elsewhere in the phage genome. Recently, a number of T4-type phage genomes have been completely sequenced (H. M. Krisch and J. D. Karam, http://phage.bioc.tulane.edu/index.html), including those analyzed here. No duplication of wac was found in any of these phages.

Modular swapping of the fibritin C terminus in phages phylogenetically distant from T4. Since the 30-aa C-terminal domain of the T4 gpwac is required for the initiation of trimerization and folding (14, 23), this segment was expected to be conserved among all gpwac homologs. The C-terminal amino acid sequences of the *wac* gene from 13 closely related T-even phages were virtually identical to that of the T4 protein (sequence data not shown). However, all of the phylogenetically

---

highlighted in yellow. The residues in putative heptad positions that do not align with their counterparts in T4 gpwac sequence are marked in green. The downward arrow indicates the position a residue of the first heptad repeat in the coiled-coil segment. The heptad positions shown in the T4 gpwac sequence are those proposed by Sobolev and Mesyanzhinov (33) with the corrections of Tao et al. (36). The loops of the gpwac of bacteriophage T4 (33, 36) are highlighted in red. The corresponding putative loop sequences in the other gpwacs are shown in light gray. The "♦" symbol at the end of the RB42, RB49 and Aeh1 sequence indicates that downstream of this point there are nonhomologous sequences that are not shown. The asterisks indicate the positions of the translational stop signals. The conserved 10-aa sequence motifs of phage 42 and RB42 gpwacs are boxed.

distant (<80% amino acid identity) T4-type phages had unrelated C-terminal domains (Fig. 3). In phage RB49, for example, the homology with T4 gpwac ceases after the residue Trp476 (T4 *wac* sequence coordinates), and the terminal 10 aa are replaced by a 143-aa sequence of unknown origin. In phages RB42 and RB43, at the residue aligned with Gln455 in T4 gpwac, there is a 323-aa insertion with weak homology to proteins containing an immunoglobulin-like fold domain (17). BLAST analysis (data not shown) of this novel 323-aa extension reveals that it actually contains three imperfect tandem repetitions of such an immunoglobulin-like domain.

In phage 42 the C-terminal domain of the fibritin has been completely deleted and the protein terminates (Fig. 2) within the region of homology to the coiled-coil domain of T4 gpwac. In some of the phages the position of the transition between the coiled-coil domain and the novel C-terminal module is clear. In phage RB49 this transition point can be localized precisely, just after the conserved sequence motif GEWV within the sequence homology to the T4 foldon (Fig. 2), but in other *wac* genes the transition point can only be situated approximately. Interestingly, phage RB43 has a 10-aa motif LT ARVAALES at the end of its homology to the T4 fibritin sequence that is identical to an amino acid sequence near the terminus of phage 42 fibritin. The function of this motif is unknown, but its appearance in similar locations in two otherwise-divergent *wac* sequences is probably not coincidental. Among the various possibilities, the corresponding DNA sequence could be the substrate for a site-specific recombination system that mediates swapping of the different types of the C-terminal wac domains.

The SchizoT-even phage, Aeh1, encodes an extremely large gpwac (1,035 aa) and only the N-terminal 423 aa of this protein show homology with the T4 sequence. The long C-terminal domain of the Aeh1 gpwac has weak homology (23% identity) with the putative host specificity determinants of some *E. coli* prophages. This large chimeric *wac* gene is followed by two open reading frames (ORFs) of 534 and 201 aa, both of which are database orphans. The published electron micrographs of this phage (38) reveal fibrous structures running along the contractile tail to the baseplate. The quality of these images is not sufficient to conclude a possible relationship between these fibers and the collar structure. Nonetheless, the sequence data alone suggest that the structure and function of the Aeh1 gpwac protein may differ from that in T4. This possibility needs to be explored, as does the organization of the Aeh1 tail fibers. Such studies are now in progress as part of the sequence analysis of the complete Aeh1 genome (Krisch and Karam, http://phage.bioc.tulane.edu/index.html).

The sequences reported here, together with the previously published sequence of phage T4 gpwac, reveal at least five distinct types of C-terminal domains (Fig. 3).

Overexpression and biochemical characterization of the RB49 *wac* protein and its deletion derivatives. We have characterized the phage RB49 gpwac protein to determine whether it has biochemical properties similar to those of the T4 protein. We first determined whether this protein mediates its own trimerization as does the T4 gpwac. To do this, we constructed an expression vector pWRB49, which allows high-level expression of the intact RB49 gpwac under the control of the T7 promoter. To determine whether this protein could be cor-
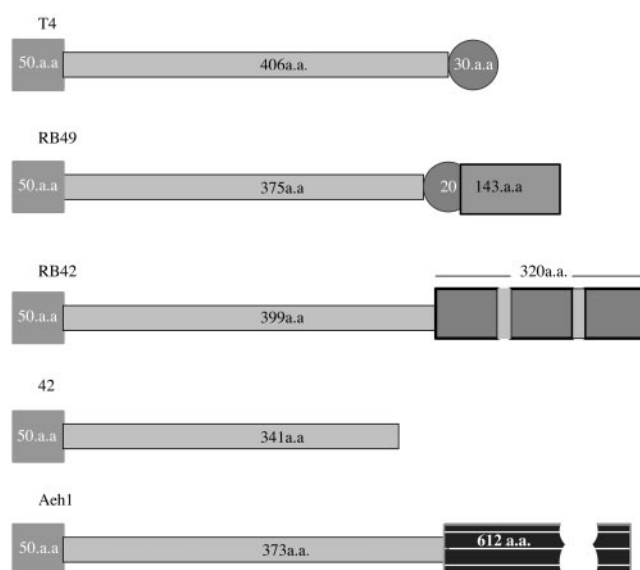

FIG. 3. Domain organization of the gpwac's from T-even, PseudoT-even, and SchizoT-even phages. Proposed structural domains are shaded as follows: medium gray, conserved N-terminal domain; light gray, coiled-coil α-helical domain; and dark gray, small C-terminal foldon domain. The phage RB49 has a chimeric C-terminal sequence: the 20 first aa of the 30-aa T4 gpwac foldon is followed by the 143-aa sequence of unknown origin. Novel C termini are indicated by different shades of gray. The repeated C-terminal motifs in RB42 are indicated by three boxes.

rectly folded, we used the assay previously developed for the T4 homolog (14, 23). Correctly folded T4 gpwac has a series of properties that make it easily identifiable. It is soluble and only precipitates at relatively high $(NH_4)_2SO_4$ concentration (>30% of saturation), but the precipitate can be redissolved easily in a low-salt buffer. Properly folded gpwac is eluted by low phosphate (10 mM) on hydroxylapatite chromatography. When lysates of cells overexpressing T4 gpwac are separated by SDS-PAGE without prior heating of the sample, the protein remains trimeric and migrates as a low-mobility band. This protein is resistant to trypsin digestion even at enzyme/protein ratios of 1:20. T4 gpwac mutants that are unable to correctly fold have a tendency to aggregate and to form insoluble inclusion bodies. They precipitate at low (18 to 22% of saturation) concentrations of $(NH_4)_2SO_4$ and are hard to redissolve. They interact strongly with hydroxylapatite columns and cannot be eluted by 500 mM phosphate buffer. The protein is sensitive to trypsin digestion and does not form a characteristic low-mobility band when subjected to SDS-PAGE.

gpwac is well expressed from the pWRB49 plasmid (constituting ~30% of the total cell protein). This overexpressed gpwac protein is completely soluble and, when denatured, its mobility on SDS-PAGE corresponds to the size expected for a monomer (~65 kDa). When samples containing gpwac were not heat denatured prior to electrophoresis, a much lower mobility band was detected upon SDS-PAGE (Fig. 4). The overproduced gpwac protein precipitates at an $(NH_4)_2SO_4$ concentration of 35% saturation but can be easily redissolved in 50 mM Tris-HCl (pH 7.5). Furthermore, it can be eluted from hydroxylapatite in 50 mM phosphate buffer. Treatment of
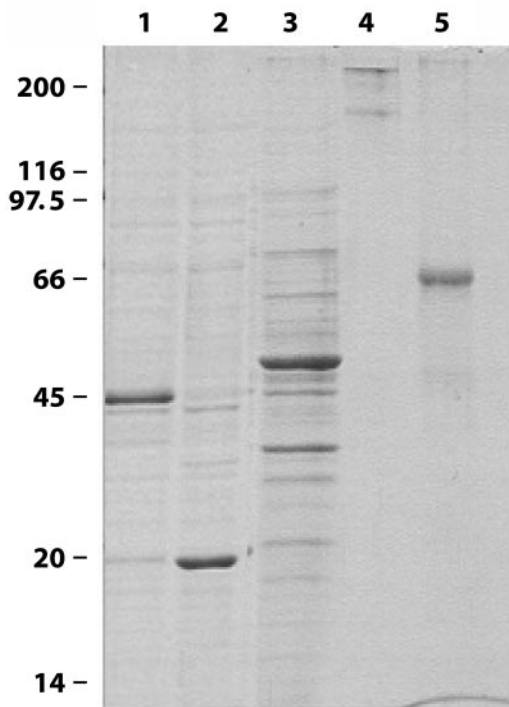
FIG. 4. PAGE analysis of the RB49 gpwac and its deletion derivatives. Partially purified extracts of the overexpressing cultures were applied to the gel. Lane 1, RBW-C nonheated sample; lane 2, RBW-C heated sample; lane 3, RBW-N nonheated sample (the mobility of the gpwac in heated samples was the same); lane 4, wild-type RB49 gpwac nonheated sample. The band of the monomeric form is absent, and a low-mobility oligomeric band can be seen in the upper part of the gel. Lane 5, wild-type RB49 gpwac. The sample was heated before application to the gel.

the crude lysate with trypsin for 30 min at 37°C does not have any effect on the protein that can be detected by SDS-PAGE. However, if the sample was first heated for 1 min at 95°C before addition of the enzyme, it was completely degraded (data not shown). A circular dichroic spectrum of RB49 gpwac purified by hydroxylapatite chromatography allowed us to estimate the protein's secondary structure. The gpwac structure was partitioned as follows: 47% α-helix, 18% β structure, 13% turns, and 22% other structures. This percentage of α-helical structure is compatible with the RB49 and the T4 gpwac having coiled-coil domains of similar size.

This ensemble of biochemical and structural data convincingly demonstrates that the RB49 gpwac can be folded into a fibrous structure with a large α-helical coiled-coil domain. The functions of the various domains of the *wac* protein were dissected by deletion analysis. We constructed a deletion mutant (RBW-C) of RB49 fibritin that lacks the first 358 aa of this protein. The mutant makes a 238-aa polypeptide from a translation initiation sequence inserted in phase at residue 359. Based on the structural model for gpwac of T4 (Fig. 2), this N terminally truncated protein should contain the last three coiled-coil segments and the C-terminal globular domain. The overexpressed RBW-C protein is soluble and, after denaturation, its mobility in SDS-PAGE corresponds to ~20 kDa. In

unheated protein samples this gpwac deletion protein migrates as an ~47-kDa protein (Fig. 4). Precipitation of the RBW-C protein requires an $(NH_4)_2SO_4$ concentration of 37% of saturation. The protein is trypsin resistant, as is the intact RB49 fibritin. These data clearly indicate that RBW-C has a structure similar to the intact gpwac. Thus, as for T4, the N-terminal domain is not required for the RB49 protein to fold into the coiled-coil structure.

We have also constructed a RB49 mutant (RBW-N) that produces a fibritin that is terminated by a TGA stop codon 20 aa downstream of the region of homology to T4 fibritin. This C terminally truncated protein, unlike the RBW-C protein, is totally insoluble and forms inclusion bodies. No SDS-resistant high-molecular-weight complex is present in unheated samples of this extract, and the protein was not resistant to trypsin. These results indicate that the RBW-N protein does not fold properly and that the absence of the C-terminal domain is probably responsible.

gpwac of phage 42. The gpwac of the phage 42 is interesting because it lacks the C-terminal foldon domain and contains no obvious replacement. In the initial analysis of phage 42 gpwac, we simply determined whether the polypeptide could be correctly folded when overexpressed in *E. coli*. The *wac* gene was amplified by PCR and cloned into the pET11d vector. *E. coli* cells overproducing this protein had inclusion bodies, and in unheated samples subjected to SDS-PAGE there was no band corresponding to a defined oligomeric species of the protein. Although the phage 42 fibritin is not being folded correctly under our conditions, it may be properly folded in *B. cepacia* either by a host- or phage-encoded chaperone or because of the intracellular conditions during infection. Another possibility is that even a much-reduced level of self-folding of the coiled-coil domain produces sufficient active gpwac to make infective phage 42 (V. P. Efimov, unpublished data; see also reference 4). Phage 42 grows very poorly and, for this reason, we have been unable to verify the presence of gpwac in the virion.

Target of gpwac interaction with the LTF. The geometry of the T4 phage particle (Fig. 1) suggests that the region near the tip of the wac fiber interacts with the knee region of the LTF, implying that the end of g*34*, g*35*, or g*36* probably encodes the wac-binding site. The existence of a mutation in g*36* with a phenotype very similar to the original wac mutant (15) lends support to a direct interaction between gpwac and gp36. If this hypothesis is correct, then distant T4-type phages with novel C-terminal wac sequences should also have novel motifs within their LTFs. Therefore, we focused our analysis on the LTF sequences of the T4-type phages that were known to have different versions of gpwac. For example, the LTF locus of the PseudoT-even bacteriophage RB49 had been sequenced previously (12). Interestingly, only the N-terminal portion of the g*36* homolog retains homology to the T4 gp36 sequence (12); the C-terminal sequence is novel (12) and substantially longer than in the T4 protein. We also examined the LTF region of the phages RB43 and Aeh1 by first performing a genome scan to localize the LTF locus (11) and then sequencing this segment by primer walking. When a 10-kb fragment encompassing the entire LTF locus of RB43 bacteriophage was analyzed by BLASTX, it was found to encode proteins homologous to T4 gp34 and gp35 (Fig. 5), followed by two large ORFs of 756 and
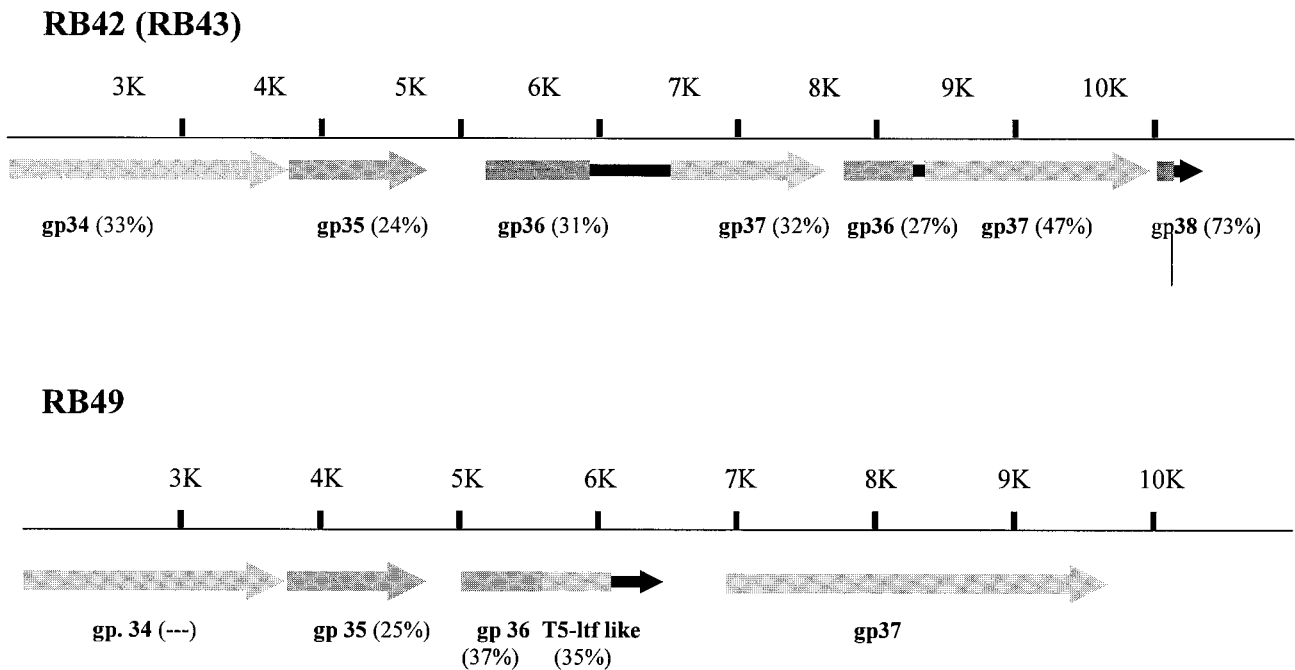
RB42 (RB43)

RB49

FIG. 5. Organization of the tail fiber loci in phages RB43 (and RB42) and RB49. The percentages of amino acid identity with corresponding regions of the phage T4 genome are shown in parentheses. For the homologs for gene *37* and gene *38*, the reference sequences are T2 homologs, and for the extension of the C-terminal domain of RB49 gp36 the reference sequence is T5 LTF.

816 aa. Both of these sequences begin with a short segment with good homology to the N-terminal sequence of T4 gp36, and the remainder of the ORF sequence is related to the central and C-terminal portions of the gp37 of various T-even bacteriophages. The sequences of these two ORFs are related but not identical. A putative transcription terminator and a putative late promoter separate them. Finally, downstream of the putative g36/g37 fusion proteins is a 174-aa ORF whose first 45 aa are 73% identical to the N-terminal sequence of the gp38 adhesin of T2 bacteriophage.

The LTF region of Aeh1 bacteriophage presents an organization very similar to that of RB43, except that the 174-aa gp38 homolog is absent (Fig. 5). For the moment we can offer no compelling explanation why, in both of these bacteriophages, the g36 and g37 are present as a single chimeric gene that has been duplicated. Nevertheless, the principal conclusion of this sequencing work is that the modular replacement of the C-terminal domain in gpwac in phages RB49, Aeh1, and RB43 is accompanied by substantial changes in the gp36 sequence. These observations are in accordance with the postulated interaction between the C-terminal part of gpwac and the sequences that form the LTF knee joint. In support of this conclusion, we have recently learned from P. Leiman (unpublished data) of his cryo-electron microscopy three-dimensional reconstruction of the T4 tail that indicates an interaction between wac and either g35 or g36.

Genetic analysis of the interaction of gp36 and gpwac. The alignment of the gp36 amino acid sequences of the T-even bacteriophages and the PseudoT-even phage RB49 (Fig. 6) reveals a conserved GDT(M/L)TG motif (aa 179 to 184 in T4) that separates the C-terminal domain extension of RB49 from the part of gp36 homologous to the T-even phages. In the

T-even phages, this motif is followed by a conserved 37-aa sequence containing four Pro residues. Such an abundance of Pro residues is incompatible with this part of T4 gp36 forming a compacted fibrous structure like the proposed β-helix structure of the LTF. Thus, the GDT(M/L)TG motif could be the boundary between the fibrous part of gp36 and a C-terminal globular domain mediating the gpwac-LTF interaction. To test this hypothesis, we created the mutation in gene *36*, amS197, that removes the last 24 residues of the gp36. If this C-terminal sequence is only involved in gp36's interaction with gpwac, then such a mutant should be partially viable in the nonsuppressing host strain *E. coli* B^E. A plasmid carrying this mutated g36 sequence was created as described in Materials and Methods. A culture of the plasmid-containing *E. coli* CR63 (su$^+$) cells was used to prepare a stock of wild-type T4 phage that contained rare recombinants with the plasmid mutated g36 sequence. This phage stock was plated on CR63 (su$^+$), and 150 individual plaques were tested on B^E (su$^-$) to identify recombinants having reduced viability on this host. Two plaques thus identified were purified by single plaque isolation on CR63 (su$^+$), and stocks were prepared on the same host. DNA sequencing confirmed that both of these phages had the expected g36 amber mutation (named T4g36amC1). The mutant T4g36amC1 was inviable in host strains that did not suppress this nonsense mutation. Thus, the C-terminal part of the gp36 must provide some indispensable function, perhaps being required for the correct folding of the entire protein or its interaction with gp37.

We examined a series of T4g36amC1 revertants by plating the T4g36amC1 stock on the B^E strain at a low dilution. Two of these revertants are able, in the presence of 6.4% PEG 6000, to lyse an *E. coli* culture in 4 to 6 h, a time intermediate
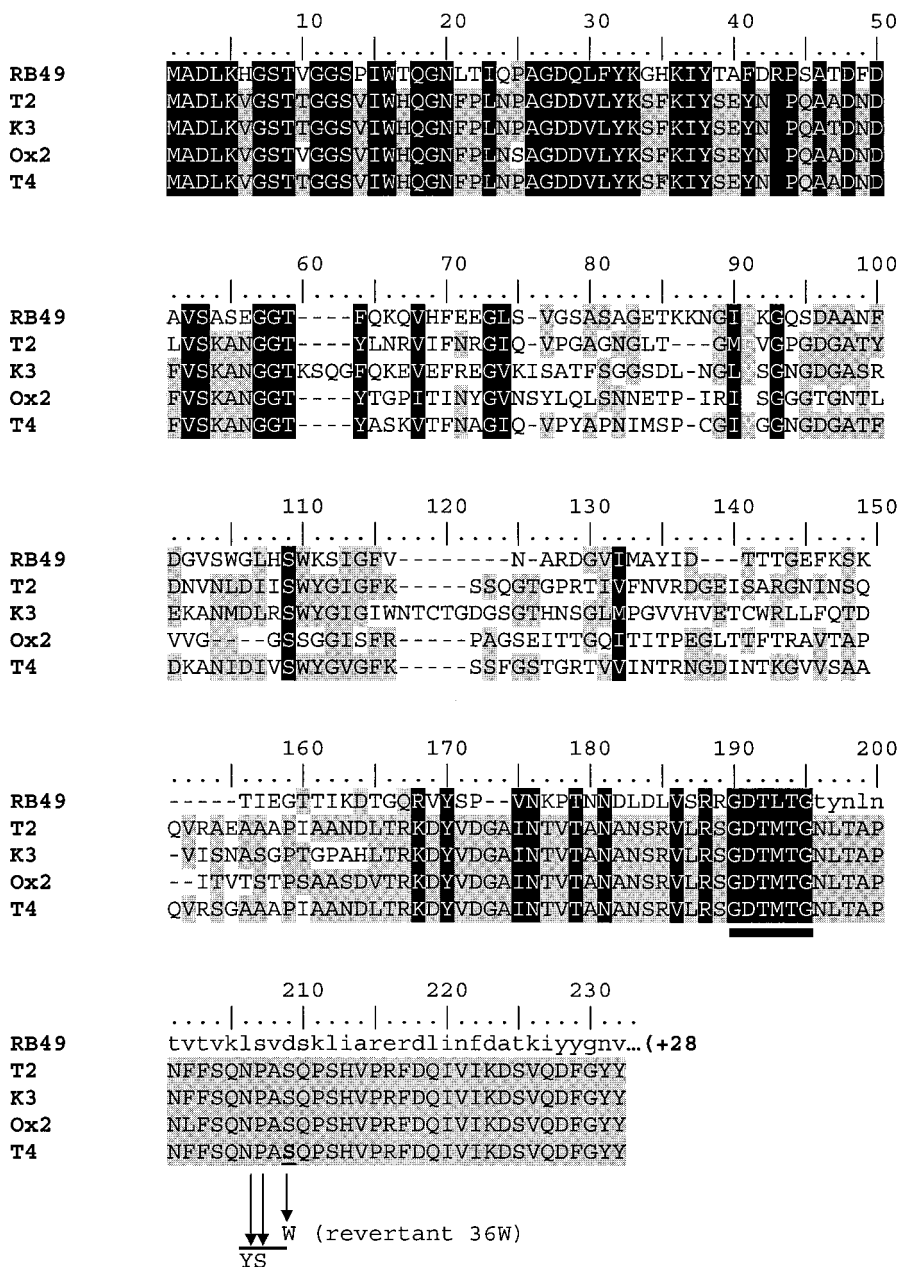
```
                     10        20        30        40        50
              ....|....|....|....|....|....|....|....|....|....|
       RB49   MADLKHGSTVGGSPIWTQGNLTIQPAGDQLFYKGHKIYTAFDRFSATDFD
         T2   MADLKVGSTTGGSVIWHQGNFPLNPAGDDVLYKSFKIYSEYN PQAADND
         K3   MADLKVGSTTGGSVIWHQGNFPLNPAGDDVLYKSFKIYSEYN PQATDND
        Ox2   MADLKVGSTVGGSVIWHQGNFPLNSAGDDVLYKSFKIYSEYN PQAADND
         T4   MADLKVGSTTGGSVIWHQGNFPLNPAGDDVLYKSFKIYSEYN PQAADND

                     60        70        80        90       100
              ....|....|....|....|....|....|....|....|....|....|
       RB49   AVSASEGGT----FQKQVHFEEGLS-VGSASAGETKKNGI KGQSDAANF
         T2   LVSKANGGT----YLNRVIFNRGIQ-VPGAGNGLT---GM VGPGDGATY
         K3   FVSKANGGTKSQGFQKEVEFREGVKISATFSGGSDL-NGL SGNGDGASR
        Ox2   FVSKANGGT----YTGPITINYGVNSYLQLSNNETP-IRI SGGGTGNTL
         T4   FVSKANGGT----YASKVTFNAGIQ-VPYAPNIMSP-CGI GCNGDGATF

                    110       120       130       140       150
              ....|....|....|....|....|....|....|....|....|....|
       RB49   DGVSWGLHSWKSIGFV--------N-ARDGVIMAYID---TTTGEFKSK
         T2   DNVNLDIISWYGIGFK-----SSQGTGPRTIVFNVRDGEISARGNINSQ
         K3   EKANMDLRSWYGIGIWNTCTGDGSGTHNSGLMPGVVHVETCWRLLFQTD
        Ox2   VVG----GSSGGISFR-----PAGSEITTGQITITPEGLTTFTRAVTAP
         T4   DKANIDIVSWYGVGFK-----SSFGSTGRTVVINTRNGDINTKGVVSAA

                    160       170       180       190       200
              ....|....|....|....|....|....|....|....|....|....|
       RB49   -----TIEGTTIKDTGQRVYSP--VNKPTNNDLDLVSRRGDTLTGtynln
         T2   QVRAEAAAPIAANDLTRKDYVDGAINTVTANANSRVLRSGDTMTGNLTAP
         K3   -VISNASGPTGPAHLTRKDYVDGAINTVTANANSRVLRSGDTMTGNLTAP
        Ox2   --ITVTSTPSAASDVTRKDYVDGAINTVTANANSRVLRSGDTMTGNLTAP
         T4   QVRSGAAAPIAANDLTRKDYVDGAINTVTANANSRVLRSGDTMTGNLTAP
                                                     ‾‾‾‾‾‾‾‾‾‾

                    210       220       230
              ....|....|....|....|....|....|...
       RB49   tvtvklsvdskliarerdlinfdatkiyygnv...(+28
         T2   NFFSQNPASQPSHVPRFDQIVIKDSVQDFGYY
         K3   NFFSQNPASQPSHVPRFDQIVIKDSVQDFGYY
        Ox2   NLFSQNPASQPSHVPRFDQIVIKDSVQDFGYY
         T4   NFFSQNPASQPSHVPRFDQIVIKDSVQDFGYY

                          ↓↓↓↓ W (revertant 36W)
                            YS
```

FIG. 6. Alignment of amino acid sequences of the various gp36 from T-even phages and the PseudoT-even phage RB49. The thick black line indicates the conserved motif GDT(M/L)TG that separates putative structural domains in gp36 (see the text). The downward arrows indicate the different mutations introduced into the T4 genome.

between that taken by wild type and the T4 *wac⁻* PEG-resistant mutants. On plates the revertants form smaller plaques than does the T4 wild type. Sequencing of g*36* from these phages revealed that they both carried a reversion replacing the TAG amber codon by TGG codon (Trp). Thus, it appears that alterations in g*36* sequence in the vicinity of the Ser197 residue can generate a phenotype similar to that of the classical *wac* mutations. Since we were unable to isolate gene *36* mutations with a PEG-resistant phenotype exactly like that originally described (15), we obtained a T4 Δ*wac* mutant constructed by V. Mesyanzhinov. This mutant resulted from a

genetic exchange between the T4 mutant *am12⁻am13⁻* and a plasmid containing the gene *12-wac-13* segment of the T4 genome but with a large HindIII-HindIII deletion within the *wac* gene. No PEG resistance selection was used to isolate this *wac* deletion phage. The Δ*wac* mutant makes small plaques, as do the other *wac* mutants. When we compared the growth of T4 wild-type and Δ*wac* phages in 6.4% PEG, we unexpectedly found that neither phage was able to lyse a culture in 5 h. Nonetheless, the control experiment with two PEG-resistant phages obtained by using a slightly modified version of the PEG selection procedure of Follansbee et al. (15) gave com-

plete lysis in 2 to 2.5 h. We then repeated this experiment with only 5% PEG and, under these less restrictive conditions, although the wild-type phage does not lyse the culture after a 4-h incubation, both the Δwac mutant and the gene 36 mutant (T4g36W) lyse cells in 2 h. Thus, the wac-null mutant is only partially PEG resistant; it can adsorb efficiently in 5% PEG but not in 6.4% PEG. "Full" resistance to PEG of some wac-null mutants (15) appears to be due to additional genetic lesions selected in presence of high levels of PEG. These data are fully compatible with our suggestion that the C-terminal part of the gp36 (or the corresponding segment in the g36–37 fusion proteins in phages with LTF organization such as RB42) is the site of interaction with fibritin.

DISCUSSION

Susskind and Boutstein (35) first postulated modular evolution on the basis of comparisons between the genomes of the lambdoid phages (reviewed in reference 8; see also references7 and 29). These authors postulated that these phage genomes could easily exchange modules and that this modular shuffling played an important role in phage evolution.

The complete genome sequence is now known for six T4-type phages (see http://phage.bioc.tulane.edu/index.html). Although the acquisition, deletion, and exchange of small non-essential genetic elements appear to be frequent (11, 12, 19, 22), there is no evidence for the exchange of large modules with essential functions as in the phage λ family. Modular swapping among the T4-type phages has been convincingly documented in only a few genomic loci involved in host range adaptation (30, 32, 37). One of these involves the tail fiber genes (32, 37). In the LTF locus, modular shuffling requires only limited regions of nucleotide sequence homology and can occur between very distant or even unrelated phages (16, 32). Since the structures of these fibrous proteins do not involve long-range tertiary interactions, there are probably few sequence constraints on their genetic shuffling with other fibrous sequences of the same type. The mutual compatibility of sequences encoding fibrous structures, coupled with the strong selective advantage derived from acquisition of novel adhesin domains (37), probably suffices to explain the frequent modular shuffling within the LTFs.

The modular exchange of the C-terminal domain of gpwac documented here provides an additional clear example of a modular replacement in a T4-type phage structural gene. The functional and structural similarities of the fibers encoded by wac and the LTF locus indicate that their sequence plasticity may be the consequence of related mechanisms. Although gpwac belongs to a different structural class of fibrous proteins (α-helical) from that of tail fibers (β-structured), it does have a receptor recognition-like function, binding specifically to a component in the middle of the LTFs. The most probable explanation for the plasticity of the gpwac C-terminal region is the variability of the LTF target sequence to which it must bind.

In the T-even bacteriophages, rearrangements of LTF genes occur frequently within the g37 and g38 sequences (37, 40), which encode the most distal portion of the tail fiber (Fig. 3) and the proximal fiber components (gp34, gp35, and gp36) are relatively conserved. The genetic plasticity in the distal part of the LTF probably results from the fact that swapping in this region can lead to advantageous changes in viral adsorption specificity. Interestingly, the protein-protein interaction between the LTFs and gpwac could exclude exchanges that extend into the proximal portion of the LTFs, since such exchanges would require a coordinated alteration in the binding specificity of the wac gene that is in another region of the genome.

The reorganization of LTF regions that we observed in the bacteriophages depicted in Fig. 5 significantly alters gene 36 (or its functional equivalent). This observation, coupled with the correlated swapping of gpwac C-terminal domains in these phages, is compatible with our explanation that the domains of these two proteins interact. However, the apparent correlation of the plasticity in gpwac and gp36, by itself, does not allow us to rigorously conclude that the changes in the C-terminal domains of gpwac are necessitated by the alterations in g36 structure, although the ensemble of our data supports such a conclusion. The examination of the structure and function of a series of gp36 proteins certainly leads to a similar suggestion. Based on the alignments of these gp36 sequences, we predicted that the C-terminal region of this protein is a probable target of interaction with gpwac in T4 phage and, furthermore, we showed that a point mutation in this region of gp36 has a phenotype identical to that of the wac deletion. Ultimate proof of the physical interaction between the C termini gp36 and the gpwac requires the isolation of intergenic suppressors of our g36 mutants and the demonstration that these are located in the predicted region of the wac gene. No such mutants have yet been isolated, but attempts to obtain them are in progress.

Our observation that phage carrying a null mutation in the gpwac protein have only partial resistance to PEG contradicts the original findings of Follansbee et al. (15) that wac-deficient mutants are resistant to 6.4% PEG. Nevertheless, we were able to obtain phages completely resistant to high concentration PEG by a slightly modified adaptation of their isolation procedure (15; data not shown). The simplest explanation for the difference between our results and those of Follansbee et al. is that their "wac" mutants had a secondary mutation that augmented the level of the phage's resistance to PEG. Such multiple mutants could have been erroneously identified as having single mutation in the wac gene. The observed partial phenotype of Δwac, as well as of our g36 mutants, suggest that the environmental sensing function might actually occur in two steps. The LTF could interact only transiently with fibritin and this interaction facilitates (but is not absolutely required for) a much more stable interaction between LTF and some other phage component (the most probable candidate being the tail shaft protein gp18). This second interaction could play a key role in the sensitivity to PEG and to other environmental agents that have long been known to influence phage infectivity, such as tryptophan, indole, and potassium glutamate (6, 21). It would be extremely interesting to fully characterize the mechanism by which such environmental agents activate or inhibit the adsorption of various T4-type phages strains. In the case of tryptophan or indole (T4B and T2H), it seems likely that a protein-protein interaction between the tail sheath (gp18) and the LTF is responsible. In other examples of environmental sensing (pH, ionic strength, and temperature) by the T4-type phage, gpwac is involved. As one considers pro-

gressively more distant T4-type phage, the environmental sensing functions may become more exotic. For example, it is known that the T4-type phage S-PM2 only infects its host, the photosynthetic cyanobacterium *Synechococcus* spp., under conditions of high illumination (N. H. Mann, unpublished data). A light sensitive interaction between the LTFs and another virion component such as gpwac could explain such sensing.

While the present study was in preparation, the genome sequences of the PseudoT-even phage 44rr2.8t (Krisch and Karam, http://phage.bioc.tulane.edu/index.html) and SchizoT-even KVP40 (25) appeared. The sequences of the *wac* genes of these phages are fully compatible with all of the conclusions we have drawn here about the domain structure of gpwac. In particular, these additional gpwac sequences both contain novel C-terminal domains that are not related to any of those described here, and in both cases significant rearrangements can be observed in the genes coding for the distal half of the LTFs.

In conclusion, the present study has used phylogenetic, genetic, biochemical, and biophysical methods to analyze the structure, function, and evolution of the T4-type phage collar fiber, gpwac. Sequence comparisons convincingly argue that the *wac* gene has a modular construction. Surprisingly, the sequences of the C-terminal domain of gpwac are completely unrelated in a number of diverse T4-type phages. This implies a large gene pool of alternative modules encoding the gpwac C-terminal domain. This portion of the wac protein has two functions in T4: it initiates the folding of gpwac subunits into a coiled-coil structure, and it binds specifically to the LTFs. In the mature virion, the gpwac-LTF interaction contributes to holding the LTFs in a retracted conformation that renders the phage noninfectious. The retraction of the LTFs is sensitive to ambient conditions such as ionic strength, pH, and temperature and, as a consequence, T4 infectivity can be adapted to its environmental situation. The considerable natural diversity of the gpwac C-terminal domains suggests that they could encode different environmental sensing capacities. These results justify a more profound examination of the sensing mechanisms used in the diverse T4-type phages to adapt their infectivity to the environmental conditions.

ACKNOWLEDGMENTS

REFERENCES

1. Ackermann, H.-W., and H. M. Krisch. 1997. A catalogue of T4-type bacteriophages. Arch. Virol. **142:**2329–2345.
2. Belle, A., M. Landthaler, and D. A. Shub. 2002. Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. Genes Dev. **16:**351–362.
3. Botstein, D. 1980. A theory of modular evolution for bacteriophages. Ann. N. Y. Acad. Sci. **354:**484–490.
4. Boudko, S. P., Y. Y. Londer, A. V. Letarov, N. V. Sernova, J. Engel, and V. V. Mesyanzhinov. 2002. Domain organization, folding and stability of bacteriophage T4 fibritin, a segmented coiled-coil protein. Eur. J. Biochem. **269:**833–841.
5. Boudko, S. P., S. V. Strelkov, J. Engel, and J. Stetefeld. 2004. Design and crystal structure of bacteriophage T4 mini-fibritin NCCF. J. Mol. Biol. **339:**927–935.
6. Brenner, S., S. P. Champe, G. Streisinger, and L. Barnett. 1962. On the interaction of adsorption cofactors with bacteriophage T2 and T4. Virology **17:**30–39.
7. Brussow, H., and F. Desiere. 2001. Comparative phage genomics and the evolution of *Siphoviridae*: insights from dairy phages. Mol. Microbiol. **39:**213–222.
8. Campbell, A. 1994. Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. **48:**193–222.
9. Conley, M. P., and W. B. Wood. 1975. Bacteriophage T4 whiskers: a rudimentary environment-sensing device. Proc. Natl. Acad. Sci. USA **72:**3701–3705.
10. Coombs, D. H., and F. A. Eiserling. 1977. Studies on the structure, protein composition, and assembly of the neck of bacteriophage T4. J. Mol. Biol. **116:**375–405.
11. Desplats, C., and H. M. Krisch. 2003. The diversity and evolution of the T4-type bacteriophages. Res. Microbiol. **154:**259–267.
12. Desplats, C., C. Dez, F. Tetart, H. Eleaume, and H. Krisch. 2002. A snapshot of the genome of the Pseudo-T-even bacteriophage RB49. J. Bacteriol. **184:**2789–2804.
13. Dewey, M. J., J. S. Wiberg, and F. R. Frankel. 1974. Genetic control of whisker antigen of bacteriophage T4D. J. Mol. Biol. **84:**625–634.
14. Efimov, V. P., I. V. Nepluev, B. N. Sobolev, T. G. Zurabishvili, T. Schulthess, A. Lustig, J. Engel, M. Haener, U. Aebi, S. Venyaminov, et al. 1994. Fibritin encoded by bacteriophage T4 gene *wac* has a parallel triple-stranded alpha-helical coiled-coil structure. J. Mol. Biol. **242:**470–486.
15. Follansbee, S. E., R. W. Vanderslice, L. G. Chavez, and C. D. Yegian. 1974. A new set of adsorption mutants of bacteriophage T4D, identification of a new gene. Virology **58:**180–199.
16. Haggärd-Ljungquist, E., C. Halling, and R. Calendar. 1992. DNA sequence of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. J. Bacteriol. **174:**1462–1477.
17. Halaby, D. M., A. Poupon, and J. Mornon. 1999. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. Protein Eng. **12:**563–571.
18. Hambly, 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.
19. Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. Trends Microbiol. **8:**504–508.
20. Hendrix, R. W., M. C. Smith, R. N. Burns, M. E. Ford, and G. F. Hatfull. 1999. Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc. Natl. Acad. Sci. USA **96:**2192–2197.
21. Kellenberger, E., E. Stauffer, M. Haner, A. Lustig, and D. Karamata. 1996. Mechanism of the long tail-fiber deployment of bacteriophages T-even and its role in adsorption, infection and sedimentation. Biophys. Chem. **59:**41–59.
22. Kutter, E., K. Gachechiladze, A. Poglazov, E. Marusich, M. Shneider, P. Aronsson, A. Napuli, D. Porter, and V. Mesyanzhinov. 1995. Evolution of T4-related phages. Virus Genes **11:**285–297.
23. Letarov, A. V., Y. Y. Londer, S. P. Boudko, and V. V. Mesyanzhinov. 1999. The carboxy-terminal domain initiates trimerization of bacteriophage T4 fibritin. Biochemistry **64:**817–823.
24. Mathews, C. K. 1994. Overview of the T4 developmental program, p. 1–11. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. American Society for Microbiology, Washington, D.C.
25. 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.
26. Mitraki, A., S. Miller, and M. J. van Raaij. 2002. Review: conformation and folding of novel beta-structural elements in viral fiber proteins: the triple beta-spiral and triple beta-helix. J. Struct. Biol. **137:**236–247.
27. Monod, C., F. Repoila, M. Kutateladze, F. Tetart, and H. M. Krisch. 1997. The genome of the pseudo-T-even bacteriophages, a diverse group that resembles the T-even phages. J. Mol. Biol. **267:**237–249.

28. **Prilipov, A. G., N. A. Selivanov, L. I. Nikolaeva, and V. V. Mesyanzhinov.** 1988. Nucleotide and deduced amino acid sequence of bacteriophage T4 gene wac. Nucleic Acids Res. **16:**10361.

29. **Ravin, V., N. Ravin, S. Casjens, M. E. Ford, G. F. Hatfull, and R. W. Hendrix.** 2000. Genomic sequence and analysis of the atypical temperate bacteriophage N15. J. Mol. Biol. **299:**53–73.

30. **Repoila, F., F. Tetart, J.-Y. Bouet, and H. M. Krisch.** 1994. Genomic polymorphism in the T-even bacteriophages. EMBO J. **13:**4181–4192.

31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

32. **Sandmeier, H.** 1994. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers. Mol. Microb. **12:**343–350.

33. **Sobolev, B. N., and V. V. Mesyanzhinov.** 1991. The wac gene product of bacteriophage T4 contains coiled-coil structural patterns. J. Biomol. Struct. Dyn. **8:**953–965.

34. **Strelkov, S. V., Y. Tao, M. M. Shneider, V. V. Mesyanzhinov, and M. G. Rossmann.** 1998. Structure of bacteriophage T4 fibritin M: a troublesome packing arrangement. Acta Crystallogr. D Biol. Crystallogr. **54:**805–816.

35. **Susskind, M. M., and D. Boutstein.** 1978. Molecular genetics of the bacteriophage P22. Microbiol. Rev. **42:**385–413.

36. **Tao, Y., S. V. Strelkov, V. V. Mesyanzhinov, and M. G. Rossmann.** 1997. Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. Structure **15:**789–798.

37. **Tetart, F., C. Desplats, and H. M. Krisch.** 1998. Genome plasticity in the distal tail fiber Locus of the T-even bacteriophage: recombination between conserved motifs swaps adhesin specificity. J. Mol. Biol. **282:**543–556.

38. **Tetart, F., C. Desplats, M. Kutateladze, C. Monod, H.-W. Ackermann, and H. K. Krisch.** 2001. Phylogeny of the major head and tail genes of the wide ranging T4-type group bacteriophages. J. Bacteriol. **183:**358–366.

39. **van Raaij, M. J., G. Schoehn, M. R. Burda, and S. Miller.** 2001. Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. J. Mol. Biol. **314:**1137–1146.

40. **Wood, W. B., F. A. Eiserling, and R. A. Crowther.** 1994. Long tail fibers: genes, proteins, structure and assembly, p. 282–290. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. American Society for Microbiology, Washington, D.C.