Structural and Biochemical Characterization of Phage λ FI Protein (gpFI) Reveals a Novel Mechanism of DNA Packaging Chaperone Activity*

Ana Popovic†, Bin Wu‡*, Cheryl H. Arrowsmith§†, Aled M. Edwards§, Alan R. Davidson||, and Karen L. Maxwell‡**

From the Departments of† Biochemistry and§ Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada, § Division of Cancer Genomics and Proteomics, Ontario Cancer Institute, Toronto M5G 2M9, Canada, ||Department of Biophysics, University of Toronto, Ontario M5G 2M9, Canada, and **Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada

Background: gpFI is required for efficient packaging of the phage genome.
Results: gpFI binds the λ major head protein, and the NMR structure reveals previously unidentified homologues.
Conclusion: gpFI homologues are widely spread through myo- and siphophages, and the head binding activity of the λ protein is required for its function.
Significance: The role of gpFI in phage morphogenesis is more conserved than was previously appreciated.

One of the final steps in the morphogenetic pathway of phage λ is the packaging of a single genome into a preformed empty head structure. In addition to the terminase enzyme, the packaging chaperone, FI protein (gpFI), is required for efficient DNA packaging. In this study, we demonstrate an interaction between gpFI and the major head protein, gpE. Amino acid substitutions in gpFI that reduced the strength of this interaction also decreased the biological activity of gpFI, implying that this head binding activity is essential for the function of gpFI. We also show that gpFI is a two-domain protein, and the C-terminal domain is responsible for the head binding activity. Using nuclear magnetic resonance spectroscopy, we determined the three-dimensional structure of the C-terminal domain and characterized the helical nature of the N-terminal domain. Through structural comparisons, we were able to identify two previously unannotated prophage-encoded proteins with tertiary structures similar to gpFI, although they lack significant pairwise sequence identity. Sequence analysis of these diverse homologues led us to identify related proteins in a variety of myoviruses and siphophages, revealing that gpFI function has a more highly conserved role in phage morphogenesis than was previously appreciated. Finally, we present a novel model for the mechanism of gpFI chaperone activity in the DNA packaging reaction of phage λ.

Over the past four decades, studies on the morphogenesis of the bacteriophage λ head have shed light on the mechanisms by which large DNA-protein complexes are assembled, providing considerable insight into the diverse areas of viral head structure, DNA packaging, and molecular chaperone activity. Among the proteins that participate in λ head assembly, the mechanism of activity of the FI protein (gpFI), which appears to function as a chaperone in the DNA packaging process, has remained enigmatic.

During the late stages of its lytic life cycle, bacteriophage λ DNA is replicated into long linear concatemers comprising many genomes joined end to end. At the same time, empty head precursors, known as proheads, are assembled. Proheads are composed primarily of gpE, the major head protein, but also contain a ring-shaped structure known as the portal protein, which is composed of 12 monomers of gpB. This structure provides the point of entry and egress for the λ genome. The terminase enzyme, a hetero-oligomer of the proteins gpNu1 and gpA, catalyzes the packaging of λ DNA into proheads and the concomitant cleavage of the DNA concatemers into unit length genomes. Terminase first forms a specific complex with λ DNA, which then binds to proheads to initiate packaging (1, 2). The binding of proheads to the DNA-terminase complex is stimulated by the presence of gpFI, and in the absence of gpFI, empty proheads and uncleaved phage DNA accumulate (3–5).

Although gpFI is one of the most abundantly expressed morphogenetic proteins, it has never been found as a component of mature viruses (6, 7), and when terminase, proheads, and λ DNA concatemers are present in high concentrations in vitro, phages are assembled normally even in the absence of gpFI (8). This suggests that gpFI is not an essential component of the phage particle but acts to facilitate the packaging reaction.

Although phages bearing FI amber (Flam) mutations cannot form plaques on non-permissive host cells, they do produce between 0.1 and 0.5 particles/cell (9). This number is much higher than that observed for amber mutants of completely

* This work was supported in part by Canadian Institutes for Health Research Operating Grants MOP-6279 (to K. L. M. and A. M. E.) and MOP-77680 (to A. R. D.) and by the Natural Sciences and Engineering Research Council of Canada (to C. H. A.).

The atomic coordinates and structure factors (code 2L5M) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

Chemical shift assignments have been deposited in the BioMagResBank database under accession number 18437.

1 Holds a Canada Research Chair in Structural Genomics.
2 To whom correspondence should be addressed: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St., Rm. 508, Toronto, Ontario M5S 3E1, Canada. Tel.: 416-978-0872; E-mail: karen.maxwell@utoronto.ca.

This is an Open Access article under the CC BY license.

SEPTEMBER 14, 2012 • VOLUME 287 • NUMBER 38

JOURNAL OF BIOLOGICAL CHEMISTRY 32085
Structural and Biochemical Characterization of \( \lambda \) gpFI

essential genes like gpE that produce fewer than \( 10^{-6} \) particles/cell (10). Mutants of \( \lambda \) that are able to grow in the absence of gpFI have been isolated and produce small plaques, suggesting a partial rescue of activity due to inefficient packaging (9, 11–13). These mutations, termed \( \text{fin} \) (FI-independent), map to the terminase and major head genes. \( \text{finA} \) mutants, which induce a 4-fold increase in gpA expression and result in a 10-fold increase in terminase activity, map to the terminase genes \( \text{Nu1} \) and \( A \). \( \text{finB} \) mutants do not affect the expression of gpA and constitute missense mutations in \( \text{Nu1} \) and \( E \). The effects of the \( \text{Nu1} \text{finB} \) mutants are unknown, but it is speculated that they increase terminase cleavage activity (12). The \( \text{finB} \) mutations that map to gene \( E \) cluster in a region encoding a 26-amino acid stretch termed the EFi domain (11). Each of the mutations involves a single amino acid substitution that results in a net increase in positive charge on the capsid and has a phenotype only in the absence of \( \text{FI} \). Further support for the interaction of gpFI with gpA and gpE has been provided by a study of \( \lambda \) protein-protein interactions using the yeast two-hybrid screen in which these interactions were detected (14).

It has been suggested that gpFI stimulates the interaction of proheads with the terminase-DNA complex, thereby increasing the overall rate of DNA packaging under conditions of limiting terminase and/or proheads (4, 15, 16). gpFI was also found to stimulate the extent rather than the rate of DNA cos cleavage by terminase independently of proheads (17). As this result was obtained under conditions of limiting terminase, gpFI may destabilize the terminase-DNA complex, thereby allowing enzyme turnover. To date, no detailed physical evidence of interactions between gpFI and other phage components exists. The goal of this work was to uncover the specific mechanism of activity of gpFI as a chaperone in \( \lambda \) DNA packaging through structural and functional studies.

**EXPERIMENTAL PROCEDURES**

**Recombinant Protein Cloning**—Gene \( \text{FI} \) was amplified by PCR from \( \lambda \) genomic DNA and cloned into a pET15b expression vector (Novagen), which encodes a 20-amino acid N-terminal hexahistidine tag. All subsequent mutants and protein fragments were generated using this plasmid as a template. For co-expression, the C-terminal domain was cloned into the compatible pCDF-1b vector (Novagen), and for tandem affinity purification assays, gpFI was cloned into the pAD100 plasmid encoding a C-terminal sequential peptide affinity (SPA) tag (18).

**Expression and Purification of Recombinant Proteins**—Protein expression plasmids were grown in *Escherichia coli* BL21 \( \lambda \text{DE3} \) and induced with isopropyl \( \beta \)-d-galactopropyranoside. Cultures were grown in LB for unlabeled protein or in M9 minimal medium containing \( ^{15}\text{NH}_4\text{Cl} \) and \( ^{13}\text{C}\text{glucose} \) (Iso tec) for labeled proteins for nuclear magnetic resonance (NMR) study of phenotype only in the absence of in a net increase in positive charge on the capsid and has a 26-amino acid stretch termed the EFi domain (11). Each of the \( \text{finB} \) mutants that map to gene \( E \) cluster in a region encoding the EFi domain (11). Each of the mutations involves a single amino acid substitution that results in a net increase in positive charge on the capsid and has a phenotype only in the absence of \( \text{FI} \). Further support for the interaction of gpFI with gpA and gpE has been provided by a study of \( \lambda \) protein-protein interactions using the yeast two-hybrid screen in which these interactions were detected (14).

It has been suggested that gpFI stimulates the interaction of proheads with the terminase-DNA complex, thereby increasing the overall rate of DNA packaging under conditions of limiting terminase and/or proheads (4, 15, 16). gpFI was also found to stimulate the extent rather than the rate of DNA cos cleavage by terminase independently of proheads (17). As this result was obtained under conditions of limiting terminase, gpFI may destabilize the terminase-DNA complex, thereby allowing enzyme turnover. To date, no detailed physical evidence of interactions between gpFI and other phage components exists. The goal of this work was to uncover the specific mechanism of activity of gpFI as a chaperone in \( \lambda \) DNA packaging through structural and functional studies.

**EXPERIMENTAL PROCEDURES**

**Recombinant Protein Cloning**—Gene \( \text{FI} \) was amplified by PCR from \( \lambda \) genomic DNA and cloned into a pET15b expression vector (Novagen), which encodes a 20-amino acid N-terminal hexahistidine tag. All subsequent mutants and protein fragments were generated using this plasmid as a template. For co-expression, the C-terminal domain was cloned into the compatible pCDF-1b vector (Novagen), and for tandem affinity purification assays, gpFI was cloned into the pAD100 plasmid encoding a C-terminal sequential peptide affinity (SPA) tag (18).

**Expression and Purification of Recombinant Proteins**—Protein expression plasmids were grown in *Escherichia coli* BL21 \( \lambda \text{DE3} \) and induced with isopropyl \( \beta \)-d-galactopropyranoside. Cultures were grown in LB for unlabeled protein or in M9 minimal medium containing \( ^{15}\text{NH}_4\text{Cl} \) and \( ^{13}\text{C}\text{glucose} \) (Iso tec) for labeled proteins for nuclear magnetic resonance (NMR) study. Recombinant His\(_x\)-tagged proteins were purified by nickel affinity chromatography by standard procedures and dialyzed against phosphate buffer (25 mM Na₂HPO₄, 250 mM NaCl, pH 6.8). Protein concentration was determined using the BCA protein assay kit (Pierce), and protein samples were concentrated using Ultracel 5000 molecular weight cutoff centrifugal filter devices (Millipore).

**Partial Proteolysis and Domain Analysis**—gpFI at a concentration of \( 10–20 \mu\text{M} \) was proteolyzed with 2.5 \( \mu\text{g/ml} \) trypsin for 1 h, and the reaction was quenched with 5 mg/ml PMSF. Proteolyzed products were separated by FPLC on a Superdex 75 gel filtration column, and eluted fractions were analyzed on a Coomassie-stained 15% SDS-polyacrylamide gel. Band intensities were quantified by densitometry and plotted to compare elution peaks for proteolyzed and full-length gpFI. gpFI samples proteolyzed for various lengths of time were also run at 40 V on a native 15% polyacrylamide gel buffered by 25 mM Tris, pH 8.8.

**Circular Dichroism Spectroscopy**—Elliptyc of purified proteins was measured in an Aviv Circular Dichroism Model 202 spectrometer. Scans were collected from 200 to 260 nm at 25 °C (or 95 °C for denatured protein) to assess secondary structure. Thermal denaturation was monitored at 222 nm from 15 to 105 °C in 2 °C increments, and the ellipticity was converted to the fraction of protein folded, normalized folded and unfolded protein base lines before fitting the data to a two-state curve using the following equation: \( \Delta G^\circ = \Delta H^\circ (1 - T/T_m) + \Delta C_p(T \times \ln(T/T_m) + (T - T_m)) \).

**NMR Spectroscopy**—All spectra were collected with 1 mM protein in 25 mM sodium phosphate buffer, pH 6.8, 250 mM NaCl, 10% (v/v) \( D_2\text{O} \) or (100% (v/v) \( D_2\text{O} \) for the \( ^{15}\text{N} \) NMR experiment) at 25 °C on a Bruker 800-MHz spectrometer (full-length gpFI) or Varian INOVA 500-MHz spectrometer at the Quebec/Eastern Canada High Field NMR Facility (gpFI C-terminal domain). All chemical shifts were assigned using SPARKY 3.114 software (19), and backbone and aliphatic resonance assignments were attained using a combination of standard triple resonance experiments (20, 21). Structure calculations were carried out with CYANA 2.0 (22) using automatically assigned and manually verified distance restraints from \( ^{15}\text{N} \) and \( ^{13}\text{C} \) NOESY experiments, and dihedral angle restraints were derived from assigned secondary structure via TALOS (23). One hundred structures were calculated from which the 20 lowest energy structures were chosen for the ensemble and submitted to PROCHECK-NMR analysis via PSVS (Protein Structure Validation Suite) for quality validation (24).

**Protein Interaction Assays**—pAD100 gpFI-SPA was transformed into *E. coli* \( \lambda cl_{\text{r57}} S_{\text{am7}} \) containing either wild-type morphogenetic genes or an amber mutation in one gene important for head assembly. Cells were cultured to midlog phase at 30 °C in LB + ampicillin, and phage production was heat-induced by incubating the culture at 42 °C for 20 min and returning it to 37 °C for 3 h to allow gpFI-SPa expression and phage assembly. Cells were lysed by sonication, and the proteins were purified as described by Zeghouf et al. (18). Purified products were separated by 15% SDS-PAGE and silver-stained; the gel bands were excised; and the products were extracted, treated with trypsin, and sent for mass spectrometry analysis. Resulting peptide masses were subjected to Profound for protein identification.

To study the interaction of gpFI with polymerized gpE, *E. coli* \( \lambda cl_{\text{r57}} S_{\text{am7}} \) lysogens were transformed with pAD100 His\(_x\)-gpFI and cultured as described above. Half of each cell extract was centrifuged for 2 h at 40,000 rpm to pellet large complexes. The supernatant was removed, and the pellet was resuspended.
in an equal volume of buffer. The tagged protein in each fraction was purified using nickel-nitrilotriacetic acid beads by standard procedures, separated by 15% SDS-PAGE, and visualized by Coomassie Blue.

To detect the interaction of endogenous gpFI with gpE, the experiment was carried out with untransformed lysogens, and the fractions were run on an SDS-polyacrylamide gel, transferred to a nitrocellulose membrane, and probed with a polyclonal anti-FI antibody (a kind gift from the laboratory of Michael Feiss, University of Iowa).

**In Vivo Activity Assays**—λ FI$_{am}$ phage was prepared from *E. coli* QD5003 containing an FI$_{am}$ prophage with a temperature-sensitive CI repressor. Cells were cultured in LB to midlog phase at 30 °C, phage production was heat-induced at 42 °C, and the cells were further incubated at 37 °C to allow phage assembly and cell lysis. Several drops of chloroform were added, cell debris was pelleted, and the lysate was filtered. *E. coli* BL21 λDE3 Δtail cells were transformed with pET15b empty vector, wild type, or mutant His$_6$-gpFI constructs. Cultures were grown to midlog phase, and 250 μl of cells were infected with various dilutions of λ FI$_{am}$ phage. Infections were plated in soft LB agar (LB + 0.75% agar + 10 mM MgSO$_4$) on LB + 10 mM MgSO$_4$ plates. Plates were incubated overnight at 37 °C to allow phage infection and uninduced low level of protein expression from the plasmid, and the following day, phage plaques were counted as a measure of phage activity. For co-expressed domains, cells were co-transformed with pET15b His$_6$-gpFI,His$_6$-N-terminal domain, or empty vector and pCDF-1b empty vector or His$_6$-C-terminal domain. Cells were selected and grown under double ampicillin/streptomycin selection and infected with λ FI$_{am}$ phage as above.

**RESULTS**

**gpFI Is a Two-domain Protein**—To gain further insight into the mechanism of action of gpFI, we initiated NMR studies. The H$^1$-N$^{15}$ heteronuclear single quantum correlation spectrum revealed the expected number of well dispersed peaks (Fig. 1A), and a combination of two- and three-dimensional NMR experiments was used to identify and sequentially assign the backbone and side chain chemical shifts of the full-length protein. Two α-helices were identified in the N-terminal region extending from residues 4 to 17 (helix 1) and from residues 28 to 41 (helix 2). These helices were defined by the presence of strong dNN($i, i + 1$) NOEs and by NOEs to hydrogen atoms three and four residues distant. Amide proton chemical shift values consistent with unstructured protein and a lack of medium and
long range NOEs indicated that residues 42–80 were unstructured. The secondary structure of the C-terminal region as determined by chemical shift analysis and mapping of backbone NOEs indicated the presence of five β-strands and a single α-helix. No NOEs were identified between protons in the N-terminal and C-terminal regions, suggesting that these two regions do not interact in the full-length protein.

Because the NMR data strongly suggested that gpFI is a two-domain protein, we endeavored to define the domain boundaries. A sequence alignment of gpFI and homologues from related bacteriophages revealed two areas of high conservation at the N and C termini of gpFI separated by a region of variable length (25–50 residues) and low sequence identity (Fig. 2). A partial trypsin digestion reaction of full-length gpFI by SDS-PAGE revealed the presence of a single band of ~10 kDa that was resolved as two protein fragments by native PAGE. This further supports the presence of two independently folded domains in this protein with cleavage likely taking place at the single Arg residue present at position 56 in the unstructured loop between the two domains.

To confirm the finding that gpFI comprises two independent domains, the N- and C-terminal domains (residues 1–41 and 72–132, respectively, as determined by boundaries of conservation from the sequence alignment with gpFI homologues; Fig. 2) were subcloned into plasmids with an N-terminal hexahistidine affinity tag. Each domain was expressed, purified to homogeneity, and subjected to circular dichroism (CD) spectroscopy. The CD spectrum of full-length gpFI is indicative of the presence of α-helical structure with minima observed at ~222 and 208 nm (Fig. 3B). The spectra of the folded and thermally denatured states show a clear difference, and after cooling, the protein refolded into a native structure with a CD spectrum that was indistinguishable from the original native spectrum, indicating that unfolding was reversible (data not shown). By fitting the data using non-linear least square regression assuming a two-state unfolding mechanism, the transition midpoint temperature ($T_{m}$) was determined to be 60.4 °C (Fig. 3C). The N-terminal domain spectrum was typical of a helical protein with a $T_{m}$ of 61.8 °C, whereas the C-terminal domain revealed a spectrum that may indicate the presence of a β-strand and unstructured regions. This domain displayed a $T_{m}$ of 60.1 °C (Fig. 3C). The cooperative thermal unfolding transitions seen for each of these domains suggest that they are able to adopt independently stable tertiary structures.
Structure of the gpFI C-terminal Domain—We had difficulty assigning NOEs to side chain chemical shifts for the residues in the N-terminal domain and unstructured loop because of significant spectral overlap and stretches of repetitive sequence. Therefore, we did not pursue further structural characterization of these regions. Because the heteronuclear single quantum correlation of the C-terminal domain of gpFI expressed on its own proved to be highly amenable to NMR studies, we determined the tertiary structure of this domain. The heteronuclear single quantum correlation spectra (Fig. 1A) and chemical shift assignments of the full-length protein and C-terminal domain alone were very similar. For example, amide proton chemical shifts determined for residues Asp73–Gln132 in the full length versus C-terminal domain alone displayed a maximum difference of 0.1 ppm (Fig. 1B). This implies that the C-terminal domain is stably folded in the same manner in the intact protein as in the absence of the N-terminal domain. The unstructured loop does not appear to interact with the C-terminal domain of the protein.

The NMR solution structure of the C-terminal domain was determined using a total of 1607 experimental restraints. We were able to assign greater than 95% of the $^{1}H$, $^{13}C$, and $^{15}N$ resonances of the backbone and side chain atoms for residues 73–132. The ensemble of 20 lowest energy structures calculated for the C-terminal domain of gpFI is presented in Fig. 1C. This implies that the C-terminal domain interacts with the major head protein, gpE. Lysogens with amber substitutions in a number of morphogenetic genes induced for phage production, and proteins that interact with gpFI-SPA were purified and identified by mass spectrometry. The phage variants tested were WT and mutants lacking the scaffold protein (gpNu3–), the head protease (gpC–), and the large terminase subunit (gpA–). B, gpFI interacts with polymerized gpE. His6-tagged gpFI was expressed from a plasmid in a lysogen induced for phage production, and large complexes were pelleted by ultracentrifugation. The His6-tagged gpFI and interacting proteins were purified from samples of the whole extract (ex), and supernatant (sn), and resuspended pellet (pe) following centrifugation, and the purified proteins were analyzed by SDS-PAGE. Note that the larger amount of gpFI relative to gpE seen here as compared with A is likely due to the different purification protocols used in the two experiments. C, wild-type phage lysates were analyzed as in B, and endogenous gpFI was detected by a polyclonal anti-gpFI antibody.

**TABLE 1**

Structural statistics for the ensemble of 20 structures for gpFI C-terminal domain

<table>
<thead>
<tr>
<th>Distance restraints*</th>
<th>gpFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short range, $</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium range, 1 &lt; $</td>
<td>i-j</td>
</tr>
<tr>
<td>Long range, $</td>
<td>i-j</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dihedral angle restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>f, y pairs (74 angles)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pairwise r.m.s.d. (Å)*</th>
<th>gpFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backbone atoms</td>
<td>0.28</td>
</tr>
<tr>
<td>All heavy atoms</td>
<td>0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ramachandran statistics (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favored regions</td>
</tr>
<tr>
<td>Additional allowed regions</td>
</tr>
<tr>
<td>Generously allowed regions</td>
</tr>
<tr>
<td>Disallowed regions</td>
</tr>
</tbody>
</table>

* Determined using PROCHECK-NMR (24).

---

**FIGURE 4.** A, tandem affinity chromatography purification of tagged gpFI shows that the gpF1C-terminal domain interacts with the major head protein, gpE. Lysogens with amber substitutions in a variety of morphogenetic genes induced for phage production, and proteins that interact with gpFI-SPA were purified and identified by mass spectrometry. The phage variants tested were WT and mutants lacking the scaffold protein (gpNu3–), the head protease (gpC–), and the large terminase subunit (gpA–). B, gpFI interacts with polymerized gpE. His6-tagged gpFI was expressed from a plasmid in a lysogen induced for phage production, and large complexes were pelleted by ultracentrifugation. The His6-tagged gpFI and interacting proteins were purified from samples of the whole extract (ex), and supernatant (sn), and resuspended pellet (pe) following centrifugation, and the purified proteins were analyzed by SDS-PAGE. Note that the larger amount of gpFI relative to gpE seen here as compared with A is likely due to the different purification protocols used in the two experiments. C, wild-type phage lysates were analyzed as in B, and endogenous gpFI was detected by a polyclonal anti-gpFI antibody.
Amino Acid Substitutions Affecting gpE Binding by gpFI Also Affect in Vivo Activity—To determine whether the N- or C-terminal domains alone are involved in gpE binding, we mixed His$_6$-tagged full-length gpFI, N-terminal domain, or C-terminal domain proteins with Nu3$_{am}$ lysates and subjected the complexes to nickel affinity chromatography. These extracts contain small oligomers of gpE, immature prohead complexes, and aberrant gpE aggregates (26). As shown in Fig. 5A, gpE co-purified with full-length gpFI and the C-terminal domain but not the N-terminal domain (NTD). B, the structure of the C-terminal domain of gpFI revealed two residues, His$^{92}$ and Phe$^{106}$, that are highly conserved in a sequence alignment and exposed on the surface of the protein. These residues are represented as blue sticks in the structure. C, in vivo complementation assays with the H92A and F106A mutants reveal that the loss of biological activity of these proteins correlates with the loss of ability to bind to gpE complexes as illustrated by SDS-PAGE analysis (D). By contrast, the substitution H97A on an adjacent surface is able to fully complement an FI$_{am}$ mutant in vivo and retains its ability to bind to gpE. Error bars represent in vivo activities from three independent experiments.

The importance of the N-terminal domain was further characterized by the creation of mutants that resulted in the charge reversal of a number of residues. These substitutions, which include K3D, R33E, and E36R, lay within the boundaries of the helices determined from analysis of chemical shift and NOE data. Additionally, we substituted a conserved Gly in the turn

affinity co-purification experiments revealed that the H92A protein was unable to bind gpE, whereas the F106A protein showed significantly less binding than the wild type. The substitution of a third residue, His$^{97}$, caused no decrease in binding of gpE (Fig. 5D). Because His$^{92}$ and Phe$^{106}$ are adjacent to each other in the gpFI structure, we concluded that these residues form part of the gpE-binding interface.

To evaluate the biological relevance of the gpE binding activity of gpFI, the gpFI H92A and F106A mutants lacking this activity were assayed for the ability to complement an FI$_{am}$ phage in vivo. As shown in Fig. 5C, the H92A mutant was unable to complement a λ FI$_{am}$ phage. By contrast, a mutant protein in which a third surface-exposed residue was substituted with Ala (H97A) was able to bind proheads in vitro and also displayed close to wild-type levels of in vivo complementation. The F106A mutant that was partially deficient in prohead binding also displayed a partial decrease in in vivo complementation activity. These results imply that prohead binding mediated by the C-terminal domain is required for the biological activity of gpFI.

The N-terminal Domain of gpFI Is Also Required for Biological Activity—Although we were able to define the requirement of the gpFI C-terminal domain for binding proheads, no function has yet been ascribed to the N-terminal domain. To assess the biological importance of the N-terminal domain, plasmids expressing the N- or C-terminal domain were assayed for the ability to complement an FI$_{am}$ phage in vivo. As shown in Fig. 6A, each domain expressed on its own had background levels of biological activity, and co-expression of the two domains from separate plasmids in the same cell did not restore activity. This indicates that both domains are required for gpFI function and that they must be linked within the same protein. Deletion of 24 amino acids from the linker region between the N- and C-terminal domains had no effect on complementation activity, indicating that this region is dispensable (Fig. 6A) and serves only to tether the two domains. This result is consistent with the low level of sequence conservation observed in the linker region.
region that separates the two helices. Each of these mutants was tested for the ability to complement an F1\text{viv0} phage in vivo (Fig. 6B). Although the R9E substitution was able to complement at the wild-type level, the activities of K3D, R33E, and E36R were decreased by 3–7-fold, and the activity of G25A was decreased to the wild-type level, the activities of K3D, R33E, and E36R were decreased by 3–7-fold, and the activity of G25A was decreased 4-fold, suggesting the presence of a restrained turn in this region.

Putative gpFI Homologues Are Found in Contractile and Noncontractile Tailed Phages and Prophages—To identify structural homologues of the gpFI C-terminal domain fold, we performed a DALI (27) search. This search yielded significant hits to two phage-encoded proteins: Bacillus subtilis YqbF (Protein Data Bank code 2HIJ) and Haemophilus influenzae HI1506 (Protein Data Bank code 2OUT). The SCOP database (28) classifies both YqbF and HI1506 as belonging to the GINS/PriA/YqbF domain family. These proteins consist of a three-helix bundle domain.

FIGURE 7. A, the C-terminal domain of gpFI (red) exhibits significant structural similarity with the N-terminal domains (NTD) of YqbF (green) and HI1506 (blue). B, genome maps showing the head morphogenesis regions of sipho- and myophages encoding gpFI homologues. In each case, gpFI is located between the major head protein and the connector proteins. C, the structure of the small, helical C-terminal domains of YqbF (green) and HI1506 (blue) are overlaid with a SAP domain (yellow; Protein Data Bank code 1H1J). YqbF and HI1506 overlap with an r.m.s.d. of 2 Å over 30 backbone positions. D, an alignment of the helical domains of very distant homologues of gpFI that were identified through structural similarity and PSI-BLAST searches. The secondary structure designations are above the sequence indicate the helical secondary structure of the solved proteins. The secondary structure designations are above or below the sequence to which they refer except for the HI1506 secondary structure, which is at the top of the alignment. The two sequences at the bottom are not putative gpFI homologues, but they possess the same structure as HI1506 and YqbF. Arrows indicate the positions of amino acid substitutions of gpFI that were tested in C. These positions are conserved in the homologues of gpFI that can be identified by amino acid sequence similarity. P. alcalifaciens, Providencia alcalifaciens.

The gene encoding YqbF lies within a B. subtilis PBSX-like prophage element called skin (29, 30). The PBSX prophage, which is also found in B. subtilis, encodes contractile tailed phage-like particles that are released following DNA damage (31–33). The position of the yqbF gene is analogous to the position of the FI gene in λ (Fig. 7B), lying immediately adjacent to the gene encoding the major head protein (YqbE), which is 80% identical to XkdG, the major head protein of phage PBSX. Immediately downstream of yqbF are the genes yqbG and yqbH whose protein products are 50 and 54% identical to YkzL and XkdH, respectively, which are protein homologues of the phage SP1 connector proteins (29). Interestingly, PBSX does not contain an open reading frame corresponding to yqbF despite its similarity to the YqbF-containing prophage. The gene encoding HI1506 is also embedded in an H. influenzae prophage known as FluMu (34, 35). The gene immediately upstream of HI1506 encodes a protein that is 51% identical to the major head protein of E. coli phage Mu (Mup34), and the gene immediately downstream encodes a protein that shares 38% sequence identity with Mup36, a connector protein (36). Thus, HI1506 also shares a common genomic position with gpFI.

Although the structural similarity and genomic positioning of gpFI, YqbF, and HI1506 suggest an evolutionary relationship...
Structural and Biochemical Characterization of λ gpFI

between these proteins, no significant sequence similarity could be detected in pairwise alignments of these proteins. In addition, despite multiple iterations of PSI-BLAST utilizing each of these proteins as queries, we could not detect sequence links among these proteins. When the λ gpFI sequence was used to initiate PSI-BLAST searches, most of the sequences identified were from closely related phages or prophages in Escherichia, Shigella, and Salmonella. The B. subtilis protein YqbF showed similarity only to proteins from Bacillus species. By contrast, searches utilizing HI1506 as a query revealed similarity to proteins from a wide variety of phages and prophages in Gram-positive and –negative bacterial species, including putative homologues in 40 long tailed phages. The genes encoding 37 of these homologues are positioned immediately adjacent to the genes encoding the major head protein, and the remaining three are within four ORFs. Interestingly, the sequence similarities among these homologues do not always extend through both the head-binding and helical domains. For example, the head-binding domains of HI1506-like proteins from phages D3112 and Mmp38 are 44% identical, whereas their helical domains are less than 15% identical. These domains are also in the reverse order with respect to each other (i.e. in phage D3112, the helical domain is at the N terminus similar to gpFI, whereas in phage Mmp38, the helical domain at the C terminus is similar to HI1506). HI1506 and its homologue from phage Mu also display clearly related head-binding domains (35% identity), but their helical domains are highly diverged (<20% identity). The phages described above all possess major head proteins within the same sequence family (i.e. >30% identity), which is expected because their head-binding domains display pairwise identities of greater than 30%. By contrast, some phages with very similar HI1506 homologues possess highly diverged major head protein sequences. For example, the HI1506 homologues of Staphylococcus aureus phages CNPH82 and PH15 are 95% identical, but their major head proteins are less than 15% identical.

Surprisingly, the HI1506 homologues of ~10 phages comprised only a helical domain and possessed no domain similar to the gpFI head-binding domain. In the cases of Listeria phages PSA, A500, and A118 and Lactococcus phage TP901-1, these helical domain proteins have been shown or hypothesized to be appended to the C terminus of the major head proteins by both +1 (phage PSA) and −1 (phages A500, A118, and TP901-1) programmed translational frameshifts (37, 38). Despite the clear relatedness of the helical domains from these four phages (pairwise sequence identities range from 35 to 85%), their corresponding major head proteins display sequence identities of less than 15%. On the other hand, the major head proteins of several phages, such as TP901-1 and phage 55, display sequence identities of ~50%; however, TP901-1 uses a frameshifted helical domain, whereas phage 55 possesses an HI1506 homologue with both a head-binding domain and a helical domain. Furthermore, Brochothrix phage NF5 possesses a major head protein that is 70% identical to that of TP901-1, but it frameshifts an Ig-like domain (39) onto its C terminus instead of an FI-like helical domain. Another strategy is used by Lactobacillus phage Kc5a in which a helical domain is fused in-frame onto the major head protein. Interestingly, some phages with major head proteins that are greater than 40% identical to that of Kc5a (e.g. Bacillus phage IEBH) do not display any fusion to a helical domain and do not appear to encode any FI-like protein. The sporadic occurrence and varying mechanisms of association with the major head protein suggest that the genes encoding gpFI-like proteins may have spread through horizontal gene transfer events.

DISCUSSION

gpFI is a molecular chaperone that facilitates the interaction of the terminase-DNA complex with proheads and thus increases packaging efficiency (4, 15, 16). The present results provide insight into the mechanism by which gpFI carries out this function. We show that gpFI is a two-domain protein and that both domains are required for biological activity (Fig. 6A). The C-terminal domain of gpFI interacts with gpE, and amino acid substitutions decreasing this interaction caused a corresponding decrease in the biological activity of gpFI. This implies that gpE binding is a requirement for gpFI function. An interaction between gpE and gpFI is consistent with the contiguous positions of genes E and FI in the λ genome as well as with previous genetic evidence (11, 14).

Although we did not determine the tertiary structure of the gpFI N-terminal domain by NMR, our CD and NMR data demonstrated that this domain adopts a stable helical structure. Our discovery of putative homologues of gpFI in prophages FluMu and skin led us to hypothesize a structure and function for this domain. As shown in Fig. 7D, the helical domain of HI1506 displays sequence similarity to the SAP and Rho_N DNA-binding domain Pfam families PF02037 and PF07498 respectively. In addition, the tertiary structures of the helical domains of both HI1506 and YqbF overlay well with the SAP DNA-binding domain (Fig. 7C). Sequence alignment based on the structures of the proteins revealed similarities in the helical domains of gpFI, HI1506, YqbF, a SAP domain, and a Rho_N domain protein (Fig. 7D). This alignment also showed that the helices of gpFI mapped by NMR using chemical shift index and NOE data are comparable in length with the known structures and are separated by a putative turn region of similar length. The substitution of a conserved Gly in this putative turn in gpFI abrogated activity, supporting an important role of this region in maintaining the structure of the domain (Fig. 6B). Taken together, the sequence and structural similarities between the gpFI-like proteins and the SAP and Rho_N family of DNA-binding domains suggest that the helical domain of gpFI may be a DNA-binding domain. This conclusion is supported by the observation that gpFI co-purified with DNA through a sucrose gradient (40). This gpFI-mediated DNA binding activity could also play a role in the observed stimulation of cos site cleavage by terminase (17).

Given that the C-terminal domain of gpFI binds to gpE and the N-terminal domain may bind DNA, we propose that gpFI facilitates the interaction between proheads and the terminase-DNA complex. Because gpFI is expressed at a level similar to gpE during λ infection and our SPA tag affinity chromatography experiments showed that similar amounts of gpFI and gpE are found in complexes of those proteins (Fig. 4A), gpFI likely coats the surface of the prohead through binding mediated by its C-terminal domain. This would place its helical N-terminal
domain on the prohead surface where even a relatively weak and nonspecific DNA binding activity could aid in attracting the terminase-DNA complex to the prohead. Supporting our model of gpFI action, many proteins adopting the SAP fold bind DNA nonspecifically (41). Also consistent with a nonspecific DNA binding role for the N-terminal domain of gpFI, the amino acid substitutions in gpE that decrease the requirement for gpFI caused a net increase in the positive charge of gpE (11). Modeling of gpE using the structure of a closely related prophage-encoded protein shows that these residues are positioned on the prohead surface (Fig. 8). This increase in surface positive charge might also serve to increase the attraction of the terminase-DNA complex to the proheads, thereby bypassing the necessity for gpFI to coat the prohead. The same effect would occur in the cases of those phages (e.g. PSA and TP901-1) that append putative DNA-binding helical domains directly to their major head proteins through frameshifting. Another possible activity for the N-terminal domain of gpFI is a direct interaction with the terminase...
protein complex. This interaction, which was detected in a two-hybrid assay (14), could also serve to bring together proheads and the terminase-DNA complex. The head expansion that occurs during DNA packaging may disrupt binding of gpFI, explaining why this protein is not found in mature λ particles (7).

The analysis of the HI1506 family of putative gpFI homologues showed that very similar homologues may be associated with distantly related major head proteins, and among phages with closely related major head proteins, some may possess full-length homologues, whereas others append a helical domain through translational frameshifting. Furthermore, major head proteins closely related to ones associated with a gpFI homologue may frameshift an unrelated domain, such as an Ig-like domain, and others lack a detectable gpFI homologue. This sporadic occurrence of gpFI homologues among unrelated myo- and siphophages and their varying mechanisms of association with major head proteins argue that the genes encoding these proteins have spread via horizontal gene transfer events. Whether all of the identified gpFI-like proteins are related through divergent evolution or whether they arose from several distinct progenitors is not clear. The distribution of gpFI-like proteins is reminiscent of phage Ig-like domains, which are also found sporadically in a variety of phage structural proteins and are often appended to their C termini through translational frameshifting (42, 43).

gpFI was previously considered to be unique to phage λ and its close relatives φ80, N15, and 21 because homologues could not be identified outside of this group. However, our determination of the tertiary structure of the C-terminal domain of gpFI has allowed us to identify two prophage-encoded proteins, HI1506 and YqbG, with structures very similar to gpFI. These proteins both contain one domain that closely resembles the C-terminal domain of gpFI and another that is small and helical like the N-terminal domain of gpFI. Like gpFI, these proteins are encoded adjacent to genes encoding major head proteins. The similarities in structure and genome position of gpFI, HI1506, and YqbG imply that these proteins are homologues. It is notable that HI1506 homologues are found in well-characterized phages, such as Escherichia coli phage Mu and Lactococcus lactis phage TP901-1, even though gpFI-like packaging chaperones have not been proposed in these phages. The widespread occurrence of putative gpFI homologues in diverse phages suggests that a gpFI-like function is more generally important for phage morphogenesis than has been previously appreciated. Functional studies on these proteins will be of great interest.

Acknowledgments—We thank Diane Bona for technical assistance, Mike Feiss for the gift of the gpFI antibody, Gabe Lander for providing the coordinates of Protein Data Bank code 3BQW modeled into the λ prohead cryo-electron density, Jonathan Lukin and Tara Sprules for NMR data collection, and Paul Sadowski and Lisa Pell for critical reading of the manuscript. NMR experiments were recorded at the Québec/Eastern Canada High Field NMR Facility, supported by grants from the Canada Foundation for Innovation, the Québec Ministère de la Recherche en Science et Technologie, and McGill University.

REFERENCES


