Thermodynamic and Functional Characterization of Protein W from Bacteriophage λ

THE THREE C-TERMINAL RESIDUES ARE CRITICAL FOR ACTIVITY*

Received for publication, February 11, 2000, and in revised form, March 30, 2000
Published, JBC Papers in Press, April 18, 2000, DOI 10.1074/jbc.M001178200

Karen L. Maxwell‡, Alan R. Davidson§§, Helios Murialdo¶, and Marvin Gold‡‡

From the Departments of ‡Molecular and Medical Genetics and §§Biochemistry, University of Toronto, Toronto, Ontario
¶SASS 1A8, Canada and the ¶Fundacion Ciencia para la Vida, Avenida Marathon 1943, Santiago, Chile

Gene product W (gpW), the head-tail joining protein from bacteriophage λ, provides a fascinating model for studying protein interactions. Composed of only 68 residues, it must interact with at least two other proteins in the phage, and probably with DNA. To study the structural and functional properties of gpW, plasmids were constructed expressing gpW with hexahistidine tag sequences at either the N or C terminus. The purified wild type protein was found to be stably folded and biologically active. The protein is monomeric as judged by equilibrium ultracentrifugation, and appears to unfold by a cooperative two-state mechanism. Circular dichroism studies indicate that the protein is 47% helical, with a Tm of 71.3 °C, and a ΔGf of 3.01 kcal/mol at 25 °C. Mutagenesis of the three hydrophobic C-terminal residues of gpW showed that they are critical for activity, even though they do not contribute to the thermodynamic stability of the protein. Using secondary structure prediction as a guide, we also designed destabilized gpW mutants. The hydrophobic nature of the gpW C-terminus caused these mutants to be degraded by the ClpP-containing proteases in Escherichia coli.

Macromolecular assembly processes, which proceed along ordered morphogenetic pathways, are fundamental to all biological systems. One such system that has been well characterized is the morphogenesis of the E. coli bacteriophage λ. The 48-kilobase pair linear double-stranded DNA genome of bacteriophage λ is contained within an icosahedral head that is assembled with the participation of 10 phage proteins and the Escherichia coli GroEL and GroES proteins. Head morphogenesis initiates with the assembly of the preconnector, a doughnut-shaped structure composed of 12 subunits of uncleaved gene product gpW in a reaction that requires the presence of gpNu3 and gpGroEL, and gpGroES (1). This structure then interacts with gpC, and acts as the starting point for the polymerization of 420 molecules of gpE into the head (2). Next, about three quarters of the gpB molecules are cleaved, all of the gpNu3 is degraded, and each of the gpC molecules participates in a fusion/cleavage reaction with gpE to form the proteins pX1 and pX2. DNA packaging by terminase (made up of gpNu1 and gpA) then occurs, during which time the head expands by approximately 20%, and 420 molecules of gpD are incorporated into the surface lattice (3). Once packaging is complete, gpW interacts with the DNA-filled heads, which allows the incorporation of gpFII and completion of the head. Finally, the tail, which is formed in a separate pathway (for review, see Ref. 4), spontaneously attaches to the completed head to form a mature infectious virion.

gpW, the 68-residue head stabilizing protein, has a molecular weight of 7,614, and a calculated pI of 10.8. W+ phage produce heads of the same diameter as wild type λ, DNA molecules of monomer length with cohesive ends, and active tails that are capable of efficiently complementing heads from a λ− infection (5, 6). Katsura and Tsugita (7) detected a gpW complementing activity of approximately 10 kDa in guanidine hydrochloride dissociated phage ghosts, suggesting that it is a structural protein, although its precise stoichiometry is unknown. Partially purified gpW has been shown to interact with DNA-filled proheads in vitro after the incorporation of gpD. In the completed virion gpW probably interacts with both gpFII, as incorporation of gpW is necessary before gpFII can act, and some other component of the head. The most likely candidate is gpB, since this forms the portal collar where the tail attaches. The DNA inside heads in W+ phage has an increased sensitivities to DNase treatment (8), and is more likely to spontaneously come out of the head, suggesting that gpW may play a role either as a structural protein that forms a plug at the portal prohead vertex, or as a DNA-binding protein that interacts directly with the right end of the λ DNA molecule, which protrudes into the tail in fully assembled phage.

The multiple roles played by gpW, despite its small size, make it an attractive subject for detailed in vitro studies. Here we show that large amounts of active gpW can be easily purified using N- or C-terminal hexahistidine tags and that these tagged proteins display reversible two-state unfolding behavior indistinguishable from the untagged protein. Furthermore, we show that the three C-terminal residues of gpW are crucial for in vitro activity, but not for thermodynamic stability.

EXPERIMENTAL PROCEDURES

Plasmid Construction—Gene W was amplified by polymerase chain reaction from bacteriophage λ DNA. Residues 1–68 were cloned into the expression vector pET21d (Novagen), which contains a T7 RNA polymerase promoter, in such a way that the second residue was changed from Thr to Val, and 16 extra residues encoding the FLAG epitope (9)
and hexahistidine sequence used for protein purification (RLDYKDDK-His$_6$) were fused to the C terminus. Changing the second residue from Thr to Val was considered a conservative mutation as Val is observed at this position in 3 closely related bacteriophage 21 (10). An N-terminal tagged fusion construct was also created, which expressed a native N terminus, G$\alpha_{TP}$, and the 21-residue T2V DYDITPTTENLYFQ G tag at the N terminus for the E. coli lacZ $\Delta_{D::}$zhf $D_{Tet}$ construct. Thermal denaturation was monitored by measuring the CD signal at 225 nm, averaged over 30 s. Both thermal and urea denaturation experiments were performed at rates of 0.2 °C/min, using a Jasco J-810 spectropolarimeter. Samples of gpW(CT) at concentrations of 40, 100, and 300 μM in 1 mM NaCl were run, with the major protein peak eluting at 0.4 M NaH$_2$PO$_4$, 10 mM Tris-HCl, 10 mM imidazole, pH 8.0, and purified in the same buffer via batch method using nickel-nitrilotriacetic acid-agarose resin (Qiagen). The pure proteins were eluted with 6 M GdnHCl, 0.2 M acetic acid, and were refolded by dialysis into 10 mM Tris-HCl, pH 8.0, 0.2 mM EDTA, 200 mM NaCl. All other experiments described in this work were performed in this buffer. After dialysis of the N-terminal tagged protein, TEV cleavage was performed according to the manufacturer’s directions (Life Technologies, Inc.), and the proteins were loaded onto a second nickel-nitrilotriacetic acid column in the TEV buffer, and were shaken at 4 °C for 1 h. The TEV protease is 6-His-tagged, so the nickel-nitrilotriacetic acid resin binds the TEV protease, any uncleaved gpW(NT), and the cleaved 6-His tag. The unbound protein was dialyzed against a high concentration of urea (8.5–9.0 M) at 25 °C. After the second cleavage, the isolated fragment was loaded onto a second nickel-nitrilotriacetic acid column in the TEV buffer, and was washed with buffer with a 1 ml/min flow rate. A 2 mM sample of gpW(NT) was isolated from the tobacco etch virus (TEV) and the 6-His tag is isolated from the lysogenic strain 594(amgD ammG60), which prevents the expression of a 21-kDa histidine-rich protease cleavage site, and is the buffer density, $\rho$ is the partial specific volume of the protein, $\Omega$ is the angular velocity, $M$ is the molecular weight of the protein. The partial specific volumes as calculated from the amino acid compositions of gpW(CT) and gpW(NT) were 0.7030 cm$^3$ g$^{-1}$ and 0.7115 cm$^3$ g$^{-1}$, respectively, and the solution density was measured to be 1.011 g/ml. Gel filtration chromatography was performed at room temperature by FPLC on a 25-ml Superdex-75 column (Amersham Pharmacia Biotech) calibrated with bovine serum albumin (67 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), myoglobin (17 kDa), lysozyme (14kDa), and insulin (3.5 kDa). The column was washed with two volumes of buffer with a 1 ml/min flow rate. A 2 mM sample of gpW(CT) was loaded on the column and eluted at 1 ml/min, using absorbance at 280 nm to detect protein.

**In Vitro Assay of gpW Activity**—Virion extracts lacking gpW were prepared from the lysogenic E. coli strain 5044(lacW $\Delta_{lacW}$ lacS::lacZ), 50 ml of LB was inoculated with 1 ml of overnight culture, and the cells were grown to an optical density at 600 nm of approximately 0.4, heat-induced at 45 °C for 15 min, and then grown 1 h at 37 °C. At the end of the induction period, the cells were collected by centrifugation, and resuspended in 1 ml of RMR buffer (16.8 mM NH$_4$Cl, 0.9 mM MgSO$_4$, 18 mM KCl, 2.7 μM FeCl$_3$, 44.5 mM Na$_2$HPO$_4$, 19.8 mM KH$_2$PO$_4$, 0.25 M Na$_2$EDTA, 10% glycerol, 0.1% β-mercaptoethanol, 10 mM putrescine) and 150 μl of chloroform to induce lysis of the cells. Cellular debris was collected by centrifuging the tubes for 5 min at 15,000 rpm, 4°C, and the upper aqueous phase was collected and stored at −80 °C. Equal volumes of acceptor extract and purified gpW at known concentrations were mixed in an Eppendorf tube and incubated at 37 °C for 1 h, then placed on ice, and serial dilutions of the hemagglutinating units were made, and titers were determined using the plaque assay.

**Thermodynamic and Functional Characterization of Protein W**

Thermal and chemical denaturation experiments were performed in a 1-cm cuvette using a Microlab 500 series automated titrator, and the software program IgorPro™. A 4 μM protein solution in buffer was mixed in 2 μl increments into a 160 μl protein solution that had a high concentration of urea (8.5–9.0 M) at 25 °C. After each injection, the sample was mixed for 60 s, and the CD measurement at 225 nm was averaged over 30 s. Both thermal and urea denaturation curves were completely reversible under all conditions. These data were fit to the standard thermodynamic equations for a monomeric protein. A minimum of five thermal and three urea denaturation experiments were performed on each of the wild type constructs. The average error associated with these experiments was 0.7 °C and 0.24 kcal/mol, respectively.

The change in heat capacity upon unfolding ($\Delta C_p$) of gpW(NT) was determined according to the method of Pace and Laurent (15). Averaged $\Delta C_p$ values determined from the thermal unfolding transition zone of 9000 M$^{-1}$ cm$^{-1}$, a 7°C value determined by performing urea denaturation experiments at temperatures between 15 °C and 35 °C were used to determine $\Delta C_p$ values over a range of temperatures. A plot of $\Delta C_p$ versus temperature was fit using the following equation, where $\Delta G(T)$ is the $\Delta G$ at a temperature $T$, $T_m$ is the midpoint of the thermal denaturation curve, $\Delta H_u$ is the unfolding enthalpy measured at the $T_m$, and $\Delta C_p(T)$ is the change in heat capacity upon unfolding (15).

$$\Delta G(T) = \Delta H_u - \Delta C_p(T - T_m) + T \ln(\text{In})$$

Fitting was performed using Kaleidagraph with $\Delta C_p$ as the only free parameter. The $T_m$ and $\Delta H_u$ values used were 72.4 °C and 36.4 kcal/mol, respectively, as averaged from the manual fitting of two temperature-induced unfolding curves of gpW(NT) by the method of Breslauer (28).

**Molecular Weight Determination**—Sedimentation equilibrium centrifugation experiments were performed at 20 °C in a Beckman model XL-I analytical ultracentrifuge equipped with UV-visible absorbance optics. Samples of gpW(CT) at concentrations of 40, 100, and 300 μM in buffer (10 mM Tris, pH 8.0, 0.2 mM EDTA, 200 mM NaCl) were centrifuged at 25,000, 29,000, and 30,000 rpm until equilibrium was reached at each speed, and the absorption profiles at the appropriate average speed were then recorded. The oligomeric state of gpW(NT) was also examined by equilibrium centrifugation. This protein was examined at a single concentration, 20 μM, as at lower concentrations the scatter in the measurement was too high, and at higher concentrations the protein began to aggregate in the cuvette. Data were collected at 25,000, 30,000, and 35,000 rpm. The molecular weights of the weights were calculated by fitting the resulting data sets using non-linear least squares regression (using the program Kaleidagraph) to Equation 2.

$$A = A_0 \exp[- \omega M/2RT (1 - \nu r^2 - \alpha r^3) \nu R^2 \Omega T / M T]$$

$A$ is the absorbance at 280 nm at radius $r$, $A_0$ is the absorbance at 280 nm at reference radius $r_0$, $R$ is the gas constant, $T$ is the temperature, $\nu$ is the buffer density, $\alpha$ is the partial specific volume of the protein calculated from the amino acid composition, $\omega$ is the angular velocity, and $M$ is the molecular weight of the protein. The partial specific volumes as calculated from the amino acid compositions of gpW(CT) and gpW(NT) were 0.7030 cm$^3$ g$^{-1}$ and 0.7115 cm$^3$ g$^{-1}$, respectively, and the solution density was measured to be 1.011 g/ml.

Gel filtration chromatography was performed at room temperature by FPLC on a 25-ml Superdex-75 column (Amersham Pharmacia Biotech) calibrated with bovine serum albumin (67 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), myoglobin (17 kDa), lysozyme (14 kDa), and insulin (3.5 kDa). The column was washed with two volumes of buffer with a 1 ml/min flow rate. A 2 mM sample of gpW(NT) was loaded on the column and eluted at 1 ml/min, using absorbance at 280 nm to detect protein.

**Determination of Thermodynamic Parameters by CD Spectroscopy**—Thermal and chemical denaturation experiments were performed by fitting the results of the thermal unfolding experiments to a two-state model of unfolding.

$$\Delta C_p(T) = \Delta C_p^0 (1 - r^2 - r^3) \nu R^2 \Omega T / M T$$

$\Delta C_p^0$ is the change in heat capacity upon unfolding at lower concentrations the scatter in the measurement was too high, and at higher concentrations the protein began to aggregate in the cuvette. Data were collected at 25,000, 30,000, and 35,000 rpm. The molecular weights of the weights were calculated by fitting the receiving data sets using non-linear least squares regression (using the program Kaleidagraph) to Equation 2.
leaving only one extra Gly residue at the N terminus of the protein (Fig. 1B). A purified preparation of this protein, which was named gpW(+G), displayed a high level of activity in an in vitro assay of gpW function. Since its level of activity was similar to that observed with untagged wild type gpW (data not shown), we concluded that this purification strategy would provide an effective means to study gpW structure and function. Surprisingly, the uncleaved 6-His-tagged construct, gpW(NT), displayed equivalent activity to gpW(+G), even though it possesses 25 extra residues fused to its N terminus (Table I). Since gpW(NT) could be produced even more easily than gpW(+G), further functional studies were performed with this protein.

The reaction kinetics and concentration dependence of the gpW-mediated in vitro reaction were studied in more detail. Fig. 2A shows that the maximal yield of phage is observed within 12 min of incubation of purified gpW with the extract, while an increase of 4 orders of magnitude is observed in only 2 min. In Fig. 2B it can be seen that, as the concentration of gpW increases, the number of pfu produced increases in a logarithmic fashion, reaching a plateau at a concentration of 10 μM protein. To ensure that the concentration of gpW was truly limiting when gpW was present at low concentrations, and that the results observed were not due to the phage components becoming inactivated before the reaction had time to proceed to completion, we added more phage extract to reaction mixtures after they had been incubated for 1 h (data not shown). When extra phage extract was added to reactions in which gpW was under limiting conditions (i.e., <10 μM gpW), no increase in the number of phage particles produced was observed. Conversely, when extra phage extract was added to the reactions in which gpW was in excess (>10 μM gpW), there was an increase in the number of phage produced.

To ensure that the protein purification performed under denaturing conditions was not inactivating a large percentage of the molecules in our gpW preparations, we performed packaging assays on crude E. coli extracts containing gpW(NT). gpW(NT) was found predominantly in the soluble fraction of these extracts. E. coli cells expressing gpW(NT) were diluted and assayed for in vitro activity. The intensity of the Coomassie Blue-stained band corresponding to gpW(NT) in the extract, as visualized by SDS-PAGE, was compared with known concentrations of purified gpW(NT) to estimate the concentration. As shown in Fig. 2B, the level of activity in the crude gpW(NT) extract and its concentration dependence was identical to that of the purified gpW(NT), demonstrating that the purification protocol is not inactivating the protein.

To assess whether the C terminus of gpW could also be tagged, we created a second construct, gpW(CT), which encodes full-length gpW fused to the FLAG epitope and six histidines (Fig. 1B). This tag had previously been found to affect neither the structure or function of another protein studied in our laboratory (18). However, the in vitro activity of gpW(CT) was decreased by 80-fold compared with gpW(NT) (Table I) and its activity reached a plateau at 1.0 × 10^{-7} pfu/ml, even at protein concentrations as high as 170 μM. Although gpW(CT) is not as active as the N-terminal tagged constructs, it is soluble at protein concentrations as high as 2 mM, while gpW(+G) and gpW(NT) are soluble only to concentrations of approximately 40 μM.

**Determination of the Native Molecular Weight of gpW**—The high order dependence of the gpW-mediated assembly reaction on protein concentration (Fig. 2B) implies that multiple copies of gpW are required for the production of one phage particle. For this reason, we hypothesized that gpW may be required to oligomerize for its assembly into phage, as do many structural

---

2 H. Murialdo and M. Gold, unpublished results.
In vitro assay. The soluble fraction of crude E. coli extract prepared from a number of pfu/ml formed by the addition of purified gpW to a phage extract and incubated at 37 °C. Aliquots were removed at various intervals, and plated on QD5003 cells. The activity leveled at approximately 15 min, and did not increase further by 60 min. B, gpW(NT) (●) was added to the phage extracts at varying concentrations between 0.125 and 28 μM, and the reaction was allowed to proceed for 60 min before plating on QD5003 cells. The soluble fraction of crude E. coli extract containing induced gpW(NT) (○) was also tested for activity in the in vitro assay. The concentration of gpW(NT) in the extract was estimated by visualizing the band corresponding to gpW(NT) on an SDS-PAGE gel stained with Coomassie Blue.

FIG. 2. In vitro activity of gpW(NT) as a function of time and protein concentration. Activity was monitored as a function of the number of pfu/ml formed by the addition of purified gpW to a phage extract prepared from a W* lysogenic strain. A, 20 μM gpWN(T) was mixed with a W* extract and incubated at 37 °C. Aliquots were removed at various intervals, and plated on QD5003 cells. The activity leveled at approximately 15 min, and did not increase further by 60 min. B, gpW(NT) (●) was added to the phage extracts at varying concentrations between 0.125 and 28 μM, and the reaction was allowed to proceed for 60 min before plating on QD5003 cells. The soluble fraction of crude E. coli extract containing induced gpW(NT) (○) was also tested for activity in the in vitro assay. The concentration of gpW(NT) in the extract was estimated by visualizing the band corresponding to gpW(NT) on an SDS-PAGE gel stained with Coomassie Blue.

proteins in viruses. To address this issue, analytical ultracentrifugation experiments were used to determine whether gpW exists as a multimer in solution. Both gpW(CT) and gpW(NT) were used for these experiments. Although gpW(CT) is somewhat less active than the N-terminally tagged gpW constructs, it possesses high solubility allowing native molecular weight to be determined at a wide range of protein concentrations. Sedimentation equilibrium experiments were performed at protein concentrations of 40, 100, and 300 μM (Fig. 3A). When the data from these experiments were fitted as a single species, a molecular weight of 9,235 ± 299 was calculated. The actual molecular weight of the monomer calculated from its amino acid sequence is 9,436, a deviation of only 2%. These experiments demonstrate that gpW(CT) remains monomeric even at concentrations 20-fold higher than that used in the in vitro gpW assay. Furthermore, gel filtration experiments confirmed that gpW(CT) exists in solution as a monomer at concentrations as high as 2 mM (data not shown). Since gpW(CT) is not fully active, we also determined the native molecular weight of gpW(NT). These experiments could be performed only at low protein concentration (20 μM) due to the low solubility of this construct. When the data were fit as a single species, a molecular weight of 10,738 ± 132 was calculated (Fig. 3B). The expected molecular weight of a gpW(NT) monomer is 10,707, showing that gpW(NT) is monomeric at 20 μM, which is higher than the concentration required for maximal specific activity in the in vivo assay. These results imply that gpW activity does not require prior oligomerization of the protein.

**Thermodynamic Characterization of His-tagged gpW Constructs**—To determine whether the N- or C-terminal tags had any effect on the structure or stability of gpW, circular dichroism (CD) studies were undertaken. The far UV folded spectrum of gpW(CT) is typical of a helical protein (Fig. 4A), with minima observed at 222 and 208 nm. Identical spectra were observed for gpW(NT) and gpW(+G), suggesting that all three assume the same structure with no contribution from the affinity tags (data not shown). Near UV CD spectra were also collected for gpW(CT). As the signal in the near UV region is much weaker than far UV gpW(CT) was the only protein that was soluble enough to be examined by this technique. gpW(CT) displays significant near UV protein spectrum, with a single maximum observed at 280 nm (Fig. 4B).

Since a large difference was seen between its folded and unfolded spectra (Fig. 4, A and B), CD provided a means to monitor gpW unfolding. All three tagged gpW variants showed fully reversible thermal denaturation curves that were independent of protein concentration (Fig. 5A). The three proteins displayed almost identical transition midpoint temperature (Tm) values, which averaged to 71.5 °C. Additionally, the thermal denaturation of gpW(CT) following both near and far UV CD signal gives identical curves (Fig. 5A), illustrating that the thermal unfolding transition is independent of the monitoring method. To fully analyze the data from the thermal melts, the molar heat capacity change upon unfolding (ΔCp) was determined using the method of Pace and Laurents (15). For this purpose, eight urea melts were performed at various temperatures between 15 °C and 35 °C. The ΔCp values calculated from these melts were plotted and fit in combination with ΔGm values calculated from the transition regions of thermal dena-

![Table I](image)

**Table I**

| Construct      | Tm (°C) | ΔGm (kcal/mol) | m (M) | In vitro activity at (pfu/ml) protein concentration |
|----------------|---------|----------------|-------|-----------------------------------------------------|
| gpW(CT)       | 71.6 ± 0.4 | 2.91 ± 0.11 | 647 ± 25 | 8.5 × 10³ |
| gpW(NT)       | 71.8 ± 0.7 | 3.27 ± 0.06 | 657 ± 17 | 5.0 × 10⁶ |
| gpW(+G)       | 70.8 ± 0.7 | 3.77 ± 0.11 | 648 ± 42 | 5.0 × 10⁸ |
| Average       | 71.5 ± 0.7 | 3.01 ± 0.24 | 662 ± 30 | 5.0 × 10⁷ |

a Averaged from seven experiments performed at protein concentrations of 10–300 μM.
b Averaged from five experiments at 10–50 μM.
c Averaged from three experiments at 10–50 μM.
d Averaged from the N-terminal tagged constructs only.
measured at two concentrations: 5°C-terminal residues. The activity of the mutant proteins was 1000-fold decreases in activity were displayed by single substitutions at all three logical activity were observed (Table II). A single substitution reduction is thermodynamic stability, dramatic effects on biochemical ranges occurred upon substitution. Thermodynamic and Functional Characterization of Protein W

Fig. 3. Analytical ultracentrifugation of gpW(NT). Sedimentation equilibrium ultracentrifugation data are shown for purified gpW(CT) at a protein concentration of 300 mM (A), and gpW(NT) at a protein concentration of 20 μM (B). Data points were collected at 25,000 rpm (20°C) by measuring absorbance at 250 and 280 nm, respectively. The lines joining the points in the graphs are the theoretical fits to the data using a one-species function. Residual deviations from the theoretical fits for these data sets are shown in the upper panels. The residuals in panel B appear less scattered because the optical density of the sample is much lower.

Urea-induced denaturation curves were obtained for each of the tagged gpW constructs, and in each case a single transition between the folded and unfolded states was observed (Fig. 5B). The free energy of unfolding in water (ΔGf) and the dependence of the free energy of unfolding on urea concentration (m) were found to be independent of protein concentration (data not shown) and were similar for each of the wild type tagged proteins (Table I), giving average values of 3.01 kcal mol⁻¹ and 0.66 kcal mol⁻¹ m⁻¹ respectively. All thermodynamic data were fitted with the assumption, based on the native molecular weight determination described above, that gpW folds and unfolds as a monomer. The conclusion that gpW folds as a monomer is also supported by the observation that its stability, as measured by thermal and chemical denaturation, was independent of protein concentration.

Mutagenesis of the C Terminus of gpW—Since gpW(CT) was less active than both untagged and N-terminal tagged gpW, we postulated that the C terminus may be important for protein function. To further investigate the role of the C terminus in gpW stability and function, each of the last three positions of gpW(NT) were replaced with several different amino acids and an amber codon. The mutant proteins were subsequently purified and characterized. Each variant displayed the same CD spectrum as WT (data not shown), and retained WT thermal stability (Table II), indicating that no large structural rearrangements occurred upon substitution.

While the substitutions at the C terminus of gpW caused no reduction in thermodynamic stability, dramatic effects on biological activity were observed (Table II). A single substitution of the terminal residue (V68E) completely abolished activity, as did truncations of the protein (resulting from amber mutations) by more than one residue. Greater than 1000-fold decreases in activity were displayed by single substitutions at all three C-terminal residues. The activity of the mutant proteins was measured at two concentrations: 5 μM, where the concentration of WT gpW is limiting in the in vitro reaction; and 15 μM, where the concentration of WT gpW is in excess (Fig. 2A). The relative activity of most of the mutants compared with WT increased markedly when they were added to the reaction in excess. For example, the activity of the Y67S mutant is reduced by 10⁴-fold compared with WT when assayed at 5 μM, but is less than 3-fold reduced when assayed at 15 μM, indicating that this mutant is able to function at close to normal level when present in the reaction in excess. The increased relative activity of mutants when present in excess may indicate that these have a reduced affinity for one or more of the components in the assembly reaction; thus, a higher concentration of protein can partially or completely compensate for the defect. Additionally, some mutants may have an intrinsically reduced activity once incorporated into the phage, which could lead to a lower plateau of infectivity at saturating concentrations of protein.

Destabilizing Substitutions in gpW—To provide a contrast to the C-terminal gpW substitutions, which all retained WT thermodynamic stability, we set out to design mutations that would decrease the stability of gpW. In the strongly predicted helix including residues 40–53 (Fig. 1A), the periodicity of hydrophobic residues suggested that the residues at positions 40, 43, 47, and 50 would form part of the hydrophobic core of the protein. Three of these residues, Val-40, Leu-43, and Leu-50, were substituted with Ser or Trp, two of the substitutions made in the C-terminal residues. If these positions were buried in the hydrophobic core of the protein, their substitution with Ser, a polar residue, or Trp, a very bulky aromatic residue, would be expected to decrease stability. Substitutions at these positions did indeed lead to large decreases in stability (Table II). Most dramatically, the L43W substitution resulted in a protein that was completely unfolded at room temperature. The V40S and L50W mutants displayed close to 20°C decreases in their Tm values. In contrast, substitution of Val-52, which is not predicted to form part of the hydrophobic core, caused no decrease in thermostability (Table II). The V40S, V52L, and V52W mutants all retained close to WT in vitro activity, even when assayed at a protein concentration of 5 μM. The high activity of the V40S mutant demonstrates that a large destabilization of gpW does not necessarily lead to a reduction in biological activity. The Val-52 position does not appear to play a crucial role in either the structure or function of gpW. The L50W mutant activity is reduced by approximately 20-fold, indicating that structural rearrangements induced by this putative hydrophobic core substitution cause some alter-
ation of the functional surface of the protein. Surprisingly, the L43W mutant retains some activity even though it is unfolded under the conditions of the assay. This observation suggests that the assembly of this mutant into phage helps to stabilize its native structure. Its failure to display an increase in relative activity when assayed at high concentration could indicate that the phage particles containing this unstable mutant protein are themselves unstable (i.e. the ability of this mutant to produce plaque-forming units is limited because the phage produced are too unstable to infect cells efficiently).

When the V40S, L50W, and L43W mutant proteins were expressed in WT E. coli, they did not accumulate to a high level and could not be purified. Since thermodynamically unstable proteins are generally degraded rapidly in E. coli (19), this result was not surprising. Unstable proteins spend a greater amount of time in the unfolded state, and thus are better targets for intracellular proteases than stable proteins. In order to obtain purified protein for in vitro studies, we expressed the unstable mutants in the E. coli strain SG1146A (12), which lacks the ClpP protease, which is responsible for selectively degrading proteins with hydrophobic C termini (20). Using this strain we were able to purify the unstable mutants at levels comparable to the wild type protein.

**DISCUSSION**

Although gpW was partially purified previously, its structural, biophysical, and functional properties had been largely unstudied. The construction of 6-His-tagged forms of gpW described here and the demonstration that these constructs possess thermodynamic stability and in vitro activity at levels similar to the untagged protein has allowed a detailed analysis of gpW structure and function.

**Structural Features of gpW**—Our results indicate that gpW is a monomeric, helical protein. The degree of ellipticity seen at 222 nm (Fig. 4) corresponds to a protein with approximately 47% helical content, which is similar to the 43% helical content predicted by the PHD structure prediction program. The validity of this structure prediction is also supported by our ability to successfully predict destabilizing substitutions in gpW (Table II).

Denaturation experiments of the tagged and untagged proteins show cooperative, fully reversible unfolding curves with a single transition, suggesting that gpW folds by a two-state process in which only the native and denatured states are significantly populated. Thermal denaturation curves of gpW(CT) were monitored by the change in CD signal in both the near and far UV wavelengths. As the far UV CD signal reports primarily on secondary structure, and near UV CD signal provides tertiary structure information (21), monitoring protein denaturation in these two regions provides distinct probes with which to assess protein unfolding. The fact that the near and far UV CD thermal denaturation curves are coincident supports a two-state unfolding model.

It is generally assumed that protein affinity tags will be
unstructured in solution and thus have little effect on the stability or folding properties of the tagged protein. This assumption holds true with gpW, as the values calculated for various thermodynamic parameters obtained for temperature and urea melts of the two tagged proteins and the cleaved protein shows that the thermodynamic stability of gpW is not affected by the presence of the tags. Thus, the stability data obtained from the tagged versions of the protein is comparable to the results that would be obtained from untagged protein.

Earlier work from our laboratory led to the conclusion that gpW is a small, heat-stable protein (8). Although the protein is only 68 residues long and has no disulfide bonds, temperature-induced unfolding experiments on the three WT constructs monitored by CD show that it is indeed thermostable, with a denaturation midpoint of 71.5 °C. ΔGd values obtained for the three proteins were similar, with an average value of 3.0 kcal/mol at 25 °C. This property of possessing high thermal stability accompanied by moderate stability at ambient temperatures is typical of small proteins as they have a low mol at 25 °C. This property of possessing high thermal stability accompanied by moderate stability at ambient temperatures is typical of small proteins as they have a low mol at 25 °C. This property of possessing high thermal stability accompanied by moderate stability at ambient temperatures is typical of small proteins as they have a low mol at 25 °C. This property of possessing high thermal stability accompanied by moderate stability at ambient temperatures is typical of small proteins as they have a low mol at 25 °C.

Functional Characterization of gpW—The most striking feature of the in vitro activity of gpW is its extreme concentration dependence (Fig. 2B). The most likely explanation for this phenomenon is that many molecules of gpW are required for the formation of a single phage particle. The kinetic measurements of the gpW-mediated reaction support this supposition in that the reaction occurs very quickly at first when the concentration of free gpW is highest, but decreases rapidly as large numbers of gpW molecules are incorporated irreversibly into phage particles. The incorporation of gpW into the phage particle is assumed to be an irreversible reaction as once the phage are formed, they are stable for long periods of time. In addition, large dilutions of the reactions required for plating the phage do not cause the particles to dissociate. The demonstration by analytical ultracentrifugation and size exclusion chromatography that free gpW is monomeric, even at very high concentrations, suggests that, if gpW oligomerizes, it must do so as it is being incorporated into phage particles.

An enigmatic aspect of the in vitro gpW reaction is the necessity to add gpW in vast excess compared with the other reaction components. A typical reaction contains approximately 10^15 molecules/ml of gpW to produce a maximum of only 10^6 to 10^8 phage/ml. Despite this excess of gpW, the concentration of gpW is clearly the limiting factor in the yield of the reaction when it is present at a concentration of 5 μM or below. One explanation for this observation is that a large proportion of gpW molecules are inactive. However, gpW(NT) in a crude E. coli extract produces the same number of phage particles as purified gpW(NT), showing that the protein is not becoming inactivated during the purification process. A second possibility, that a large proportion of the protein molecules are not folded, is discounted by the CD data, which show that almost all of the protein is folded at 37 °C. A third possibility is that the protein is subtly proteolyzed in E. coli, while it is being expressed. This theory can also be ruled out, as the mutagenic data showing distinct differences in the C-terminal mutations show that the C terminus of the protein is not being proteolyzed and the purification of the protein using the hexahistidine tag shows that the N terminus is intact. For these reasons we think that it is unlikely that a large percentage of gpW molecules are inactive. Thus, the most likely explanation for the data is that the in vitro assembly of each phage particle requires many molecules of gpW. The requirement for a large excess of gpW in vitro appears to differ from that inside phage-infected cells, where the level of gpW appears to be relatively low. It was never possible to identify a band on SDS-polyacrylamide gels corresponding to gpW in lysates from phage-infected cells or from plasmid-transformed cells expressing gpW under the control of its natural translation initiation site (24).

It is known that the low expression level of certain other λ morphogenetic genes is mediated by poor translation initiation (25). The low yields of gpW purified from phage-infected cells also indicate that gpW is present at relatively low levels in vivo (8). Clearly, some distinct properties of the phage assembly process in vivo must exist to allow gpW to function efficiently at low concentrations.

The Role of the Three C-terminal Residues—The sequence alignment of the gpW homologs from phage 21 and N15 shows that the hydrophobic nature of the C terminus of the protein is conserved. In all three proteins positions 66 and 68 are hydrophobic, while position 67 is hydrophilic in λ and N15, and is an Arg, which has significant aliphatic character, in phage 21. Although conserved, substitution of the three C-terminal residues of gpW does not decrease the stability of the protein, nor

| Protein | Tm °C | ND, not determined. |
|---------|-------|---------------------|
| WT      | 71.8  | 100 100             |
| V40S    | 54.0  | 65  ND               |
| L45W    | Unfolded | 0.011 | 0.011 |
| L50W    | 52.4  | 4.5  ND              |
| V52L    | 71.8  | 91   ND               |
| V52W    | 73.2  | 28   ND              |
| F66Am   | 71.7  | <10^-5 <10^-5        |
| F66L    | 71.0  | <10^-5 0.015        |
| F66W    | 70.5  | 0.0024 1.4          |
| Y67Am   | 73.1  | <10^-5 <10^-5        |
| Y67L    | 72.5  | 0.046 41            |
| Y67S    | 73.2  | 0.001 40            |
| Y67W    | 73.2  | 0.018 2.0           |
| V68Am   | 73.6  | <10^-5 0.012        |
| V68L    | 72.7  | 34   75             |
| V68S    | 73.6  | 0.088 1.1           |
| V68W    | 72.7  | 0.020 2.5           |
| V68E    | 71.8  | <10^-5 <10^-5        |
does truncation of the protein by insertion of an amber codon at any of these positions. This suggests that the C terminus is not buried in the interior of the protein as might be expected for a hydrophobic region, but is solvent exposed. This idea is supported by our observation that the gpW constructs with wild type C termini (gpW(NT) and gpW(+G)) are much less soluble than gpW(CT), which has charged residues at its C terminus. When the native hydrophobic C terminus is present, low solubility results from the exposure of these hydrophobic residues, which could cause protein aggregation. In examining the gpW sequence, it can be seen that the last 12 residues consist of mostly polar and charged residues with no large hydrophobic residues except the last three, suggesting that this whole region may be unstructured (as predicted by the PHD program). The possibility that gpW may contain some unstructured regions is supported by the \( \Delta C_p \) and \( m \) values calculated from our unfolding data, which are considerably lower than would be expected for a fully folded 68-residue protein (26).

The results of our mutagenesis experiments clearly demonstrate that the last three residues of gpW are crucial for its activity (Table II). Even relatively conservative substitutions at these positions (e.g. F66W, Y67W) cause large decreases in activity, and other substitutions totally abrogate function (e.g. V68E). The dependence of a protein’s function upon the character of its last few residues is quite unusual. We hypothesize that the C terminus of gpW is disordered and may serve as a binding site for another protein involved in a particular process. For example, gpW has an acidic C-terminal carboxylate group as part of its recognition site. In E. coli, proteins with hydrophobic C termini have been shown to be specifically degraded by a variety of proteases present in the cytoplasm and periplasm (19, 20). The ClpP-containing proteases, which we have found are able to degrade our destabilized gpW mutants, comprise a group of these tail-specific proteases. Interestingly, this protease is thought to recognize its substrates using PDZ-like domains (20). Future studies will be directed at determining whether a PDZ-like interaction may be important in the function of gpW.

Acknowledgments—We are very grateful to Andrew Becker for advice and material support of this project. We also thank Avi Chakrabarty for the use of the analytical ultracentrifuge, Sandy Koh for centrifuging the gpW(NT) sample, Eric Roche for the gift of the ClpP protease mutant strain, and Arianna Rath for calibration of the gel filtration column.

REFERENCES
1. Murialdo, H. (1979) Virology 96, 341–367
2. Murialdo, H. & Becker, A. (1978) J. Mol. Biol. 125, 57–74
3. Imber, R., Taugata, A., Wurtz, M. & Hohn, T. (1980) J. Mol. Biol. 139, 277–295
4. Katsura, I. (1983) J. Mol. Biol. 171, 297–317
5. Casjens, S., Hohn, T. & Kaiser, A. D. (1972) J. Mol. Biol. 44, 551–563
6. McClure, S. C., MacHattie, L. & Gold, M. (1973) Virology 54, 1–18
7. Katsura, I. & Taugata, A. (1977) Virology 76, 129–145
8. Peruccetti, R., Parris, W., Becker, A. & Gold, M. (1988) Virology 165, 103–114
9. Hopp, T. P., Prickett, R. S., Price, V. L., Cerretti, D. P., Urdal, D. L. & Conlon, P. J. (1988) Bio/Technology 6, 1204–1210
10. Siegle, D. A., Frackman, S., Sippy, J., Momany, T., Howard, T. M., Tilly, K., Georgopoulos, C. & Feiss, M. (1983) Virology 129, 484–489
11. Bhandari, P. & Gowrishankar, J. (1997) J. Bacteriol. 179, 4403–4406
12. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. (1998) Genes Dev. 12, 1338–1347
13. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995) Protein Sci. 4, 2411–2423
14. Chen, Y. H., Yang, J. T. & Martinez, H. M. (1972) Biochemistry 11, 4120–4131
15. Pace, C. N. & Laurents, D. V. (1989) Biochemistry 28, 2520–2525
16. Yanofsky, C. & Roder, J. (1966) J. Mol. Biol. 21, 313–334
17. Bost, B. & Sander, C. (1994) Proteins Struct. Funct. Genet. 19, 55–72
18. Maxwell, K. L. & Davidson, A. R. (1998) Biochemistry 37, 16172–16182
19. Parsell, D. A. & Sauer, R. T. (1989) J. Biol. Chem. 264, 7590–7595
20. Levchenko, I., Smith, C. K., Walsh, N. P., Sauer, R. T. & Baker, T. A. (1997) Cell 91, 939–947
21. Towell, J. F. & Manning, M. C. (1994) in Analytical Applications of Circular Dichroism (Purdy, N. & Brittain, H. G., eds) pp. 175–205, Elsevier Science, Oxford
22. Alexander, P., Fahnstock, S., Lee, T., Orban, J. & Bryan, P. (1992) Biochemistry 31, 3597–603
23. Lim, W. A., Fox, R. O. & Richards, F. M. (1994) Protein Sci. 3, 1261–1266
24. Murialdo, H. & Simioneau, L. (1972) Virology 48, 785–823
25. Murialdo, H., Davidson, A., Chow, S. & Gold, M. (1978) Nucleic Acids Res. 15, 119–140
26. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995) Protein Sci. 4, 2138–2148
27. Fanning, A. S. & Anderson, J. M. (1996) Curr. Biol. 6, 1385–1388
28. Breslauer, K. J. (1995) Methods Enzymol. 259, 221–242
