Supporting Information

Cardarelli et al. 10.1073/pnas.1005822107

SI Methods

Structure Refinement. During analysis of the {¹H}-¹⁵N heteronuclear NOE values, it was noted that residues 92 to 103, which were previously defined as an unstructured loop, exhibited relaxation values consistent with a structured region. Reanalysis of the previously collected NMR data allowed us to identify a number of NOEs that had been ambiguously assigned by ARIA (1) and assign them to these residues. Using the previously derived protein chemical shift values, dihedral angle restraints, and an additional three hydrogen bond restraints, new structures were calculated using CYANA 2.1 (2). We were able to determine that this previously under-restrained region forms an antiparallel β -sheet that pairs with the β -strand created by residues 63 to 67. Structure validation was performed using PROCHECK-NMR (3). This structure has been deposited into the PDB as 2KX4.pdb and supercedes the original 1KOH.pdb. An ensemble of 10 lowenergy structures from this new calculation are shown in Fig. S14. The most significant change in the structure can be seen Fig. S1B, where residues 101 to 106 are incorporated into a β -sheet to complete the barrel-like structure of gpFII. Arg92, a key residue on tail-binding surface of gpFII changes conformation in this new structure, and becomes exposed on the bottom surface (Fig. S1B).

Analysis of the Dominant Negative Phenotype in Vitro. To determine whether the $\Delta 53-61$ mutant inhibited reactions by binding to heads or tails, we performed an in vitro assembly reaction in two stages. First an excess amount of the mutant protein was added such that assembly would be severely inhibited. Then, either an excess of tails plus additional WT gpFII, or an excess of gpFIIcontaining heads were added to the reactions. The inhibited reaction showed no increase in infectious phage particle yield when additional gpFII and tails were added (Table S2). However, the addition of phage heads containing WT gpFII did lead to the production of large numbers of infectious phage particles. This result demonstrated that the inhibited reaction did not contain active heads, but did contain active tails, implying that the Δ 53–61 mutant acted by binding to heads and not to tails. To definitively eliminate the possibility that the $\Delta 53-61$ mutant might be able to bind tails but mediate the assembly of complete phage particles that are noninfective, we examined a lysate produced by inducing a λ FIIam prophage in the presence of a plasmid expressing the Δ 53–61 mutant. We found that this lysate contained almost exclusively unjoined heads and tails with less than 1% of particles observed displaying joined heads and tails. This appearance was very similar to the lysate produced by induction in the presence of a plasmid expressing the $\Delta 1$ -24 mutant or no plasmid at all (Table S3). The same analysis was performed with the C-terminal truncation mutant $\Delta 115-117$. Like the $\Delta 53-61$ mutant, the $\Delta 115-117$ mutant inhibited in vitro reactions showed no increase in activity when additional WT gpFII and tails were added, although the addition of gpFII-containing heads did increase activity (Table S2). As well, lysates from a λ FIIam prophage induction in the presence of a plasmid expressing the Δ 115–117 truncation mutant displayed no observable head-tail joining (Table S3). Expression of WT gpFII, on the contrary, resulted in production of a large number of mature phage particles (Table S3). These observations led us to conclude that the $\Delta 53$ -61 and $\Delta 115$ -117 mutants are able to bind to heads, but lack the surface required to bind to tails.

Stability and Expression Level of Mutant Proteins It should be noted that the behavior of the $\Delta 1-24$ deletion mutant was not a result of its being unfolded or profoundly altered in its folding behavior.

The Δ 1–24 mutant and other key mutants that lacked activity possessed thermodynamic stabilities that were similar to WT as assessed by temperature-induced unfolding experiments monitored by circular dichroism spectroscopy (Table 1). All mutants were found to be monomeric as determined by gel filtration chromatography. Finally, WT gpFII and all mutants were purified by the same Ni-affinity protocol and very similar yields of pure soluble protein were obtained for each, demonstrating that the expression levels of each mutant were similar to WT.

Analysis of Protein Structure by Circular Dichroism Spectroscopy. Pure gpFII protein was analyzed by circular dichroism spectroscopy to determine both the thermal stability and secondary structure content of mutant protein compared with WT protein. A $10-\mu$ M sample of the protein was analyzed using a 0.5-cm path length cuvette and a model 202 circular dichroism spectrometer (Aviv). Temperature melts were monitored at 227 nm.

Heteronuclear NOE Analysis. $[U^{-15}N]\lambda$ gpFII was obtained by IPTGinduced over-expression in Escherichia coli BL21(DE3) in 1 l M9 minimal medium with 1.0 g 15 NH₄Cl as described previously (4). The cells were harvested and lysed by sonication in 0.5 mg/mL lysozyme, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, and 50 mM sodium phosphate (pH 8.0). The expression product was purified using nickel affinity chromatography, eluted with 300 mM NaCl, 250 mM imidazole, 50 mM sodium phosphate (pH 8.0), dialyzed against 200 mM NaCl, 2.5 mM CaCl₂, 2 mM DTT, 50 mM Tris/HCl (pH 8.0), followed by cleavage with 2.5 U thrombin (Sigma) at 4 °C overnight, removal of cleaved hexahistidine-tag and uncleaved protein by nickel affinity chromatography, and final purification by size-exclusion chromatography. The NMR sample contained 300 μ L 0.6 mM [U-¹⁵N] λ gpFII, 150 mM NaCl, 2 mM DTT, 50 mM sodium phosphate (pH 7.0) in H_2O/D_2O (9:1) in a Shigemi microtube. Sequence-specific ¹H and ¹⁵N resonance assignments (4) were verified using conventional ¹H-¹⁵N heteronuclear NMR spectroscopy. Briefly, [¹H, ¹⁵N]-TOCSY-HSQC with a DIPSI-2rc mixing scheme (5) (73 ms mixing time) and [¹H, ¹⁵N]-NOESY-HSQC (150 ms mixing time) spectra (6) were recorded at 25 °C on a Unity INOVA 500 MHz NMR spectrometer (Varian) equipped with a room temperature probe with z axis pulsed field gradient capabilities. The H₂O resonance was suppressed by gradient coherence selection with quadrature detection in the indirect ¹H and ¹⁵N dimensions achieved by States-TPPI (7) and the echo-antiecho method (8, 9), respectively. All NMR spectra were processed with NMRPipe (10) software and analyzed with NMRView 5.2.2 (11). Hydrogen 1 chemical shifts were referenced with respect to external DSS in D₂O and ¹⁵N chemical shifts were referenced indirectly (12). Backbone mobility on the pico- to nanosecond time scale was probed by measuring ¹⁵N relaxation rates (13) at 25 °C on a Varian Unity INOVA 800-MHz spectrometer equipped with a room-temperature probe. {¹H}¹⁵N heteronuclear NOE values are reported as the signal intensity ratio in the spectra with and without proton saturation by applying a train of 120° high-power pulses 5 ms apart from each other for the final 6.0 s of the relaxation delay of 15.0 s (14) quantified in NMRView 5.2.2 (11); errors were estimated from the spectral noise. For the determination of ¹⁵N longitudinal relaxation rates R_1 (14), 10 data points were collected at delays ranging from 0.0 s to 1.2 s, for the determination of ¹⁵N rotating frame (spin-lock field 2.0 kHz) relaxation rate $R_{1\rho}$ (15) eight data points (two of them in duplicate) at delays ranging from 2 ms to 100 ms. Amide resonance intensities were quantified by three-way

decomposition using MUNIN (16, 17) and converted into relaxation rates by fitting monoexponential decays; errors were estimated from the uncertainties of the fitting parameters. Nitrogen 15 single quantum CPMG relaxation dispersion experiments (18, 19) with 14 different CPMG frequencies $\nu_{\text{CPMG}} = 1/(2\delta)$, where δ is the time between consecutive refocusing pulses, ranging from 41.7 Hz to 1,000.0 Hz during a constant-time relaxation interval of T_{CPMG} = 48 ms did not reveal any substantial exchange processes on the millisecond time scale.

In Vitro Analysis of the Dominant Negative Phenotype. To determine which assembly intermediates may accumulate in mutant protein in in vitro complementation reactions, we tested for rescue by adding an equal volume of complete tails or complete heads to in vitro complementation reactions after an initial 20-min incubation with mutant protein (Table S2). Semipurified heads were made by growing a $594(\lambda Jam_{421}cI_{857}Sam_7)$ lysogen in 500 mL of LB at 30 °C, inducing at 45 °C, and growing for a further 3 h at 37 °C. The cells were centrifuged and the pellet was resuspended in 15 mL λ -dilution buffer (10 mM Tris, pH 7.5, 10 mM MgSO₄) supplemented with 20 mM putrescine and sonicated in 15-s bursts until clear. The lysate was clarified with a low-speed spin, and the supernatant was spun at 45,000 rpm in a Beckman 70Ti rotor for 90 min. The supernatant was removed and the glassy pellet was incubated in 15 mL λ -dilution buffer supplemented with 20 mM putrescine O/N at 4 °C and resuspended by trituration and frozen at -70 °C. A typical preparation of heads had a complementing activity of 10^{10} pfu/mL. A crude tail extract was prepared by growing BL21 cells transformed with the pETail plasmid (a pET vector encoding all of the λ tail genes, a gift from R. Hendrix, University of Pittsburgh, Pittsburgh), induced with IPTG, grown for a further 3 h, and pelleted. The pelleted cells were resuspended in λ -dilution buffer, incubated in lysozyme, sonicated in 15-s bursts until clear, and centrifuged to remove cell debris. The clarified lysate was

- Rieping W, et al. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23:381–382.
- Güntert P (2004) Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353–378.
- Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486.
- Maxwell KL, Yee AA, Arrowsmith CH, Gold M, Davidson AR (2002) The solution structure of the bacteriophage lambda head-tail joining protein, gpFII. J Mol Biol 318: 1395–1404.
- Cavanagh J, Rance M (1992) Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. J Magn Reson 96:670–678.
- Zhang O, Kay LE, Olivier JP, Forman-Kay JD (1994) Backbone ¹H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR 4:845–858.
- Marion D, Ikura M, Tschudin R, Bax A (1989) Rapid recording of 2D NMR spectra without phase cycling: Application to the study of hydrogen exchange in proteins. J Magn Reson 85:393–399.
- Kay LE, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc 114:10663–10665.
- Schleucher J, Sattler M, Griesinger C (1993) Coherence selection by gradients without signal attenuation: Application to the three-dimensional HNCO experiment. Angew Chem Int Ed Engl 32:1489–1491.
- Delaglio F, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293.
- 11. Johnson BA, Blevins RA (1994) NMRView: A computer program for the visualization and analysis of NMR data. J Biomol NMR 4:603–614.

frozen at -70 °C in aliquots. This gave a complementing activity of approximately 10^{10} pfu/mL.

Electron Microscopy. To localize gpW to the λ connector, the W gene was expressed from pMal-C4× (New England Biolabs), which expressed gpW fused at its N terminus with a cytosollocalizing form of MalE (MBP). Two liters of a W mutant lysogen [594 (λW amcI₈₅₇Sam₇)] carrying the plasmid expressing MBP-gpW were induced by high temperature to produce phage particles that had incorporated the MBP-gpW protein. These cells were lysed by addition of chloroform and the phages were isolated using a standard PEG precipitation followed by CsCl ultracentrifugation as described previously (20). Phage were applied to the surface of a freshly glow-discharged continuous carbon film-coated EM grid and stained with 2% (wt/vol) uranyl acetate. Specimens were examined with a Tecnai F20 electron microscope (FEI). Images were recorded on Kodak SO-163 film and digitized with an Intergraph photoscan densitometer (Intergraph). Images of the connector regions were selected interactively with Ximdisp (21) and aligned and averaged with SPIDER (22).

Analysis of Mutant gpFII Protein Activity by TEM. To directly visualize the effect of mutant protein on λ assembly, we grew 594 (λ FIIam₄₂₃cI₈₅₇Sam₇) freshly transformed with pAD100 plasmids encoding WT and mutant constructs of gpFII in KH medium (23) supplemented with ampicillin at 30 °C with shaking. The cells were grown to an OD₆₀₀ of 0.6 and simultaneously induced for prophage induction with shaking at 45 °C and protein production with 0.1 mg/mL IPTG. The cells were grown for a further 2 h at 37 °C and then centrifuged. The pellet was resuspended in 3 mL KH supplemented with 20 mM putrescine and lysed by the addition of chloroform. The lysate was clarified by centrifugation and visualized as described elsewhere (24).

- Markley JL, et al. (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids (recommendations 1998). Pure Appl Chem 70:117–142.
- Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* 28:8972–8979.
- Farrow NA, et al. (1994) Backbone dynamics of a free and phosphopeptidecomplexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* 33: 5984–6003.
- Korzhnev DM, Skrynnikov NR, Millet O, Torchia DA, Kay LE (2002) An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. J Am Chem Soc 124:10743–10753.
- Orekhov VY, Ibraghimov IV, Billeter M (2001) MUNIN: A new approach to multidimensional NMR spectra interpretation. J Biomol NMR 20:49–60.
- Korzhneva DM, Ibraghimov IV, Billeter M, Orekhov VY (2001) MUNIN: Application of three-way decomposition to the analysis of heteronuclear NMR relaxation data. *J Biomol NMR* 21:263–268.
- Loria JP, Rance M, Palmer AGI (1999) A relaxation-compensated Carr-Purcell-Meiboom-Gill sequence for characterizing chemical exchange by NMR Spectroscopy. J Am Chem Soc 121:2331–2332.
- Tollinger M, Skrynnikov NR, Mulder FA, Forman-Kay JD, Kay LE (2001) Slow dynamics in folded and unfolded states of an SH3 domain. J Am Chem Soc 123:11341–11352.
- Cardarelli L, et al. (2010) The crystal structure of bacteriophage HK97 gp6: Defining a large family of head-tail connector proteins. J Mol Biol 395:754–768.
- Crowther RA, Henderson R, Smith JM (1996) MRC image processing programs. J Struct Biol 116:9–16.
- Frank J, et al. (1996) SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116:190–199.
- Casjens S, Horn T, Kaiser AD (1972) Head assembly steps controlled by genes F and W in bacteriophage lambda. J Mol Biol 64:551–563.

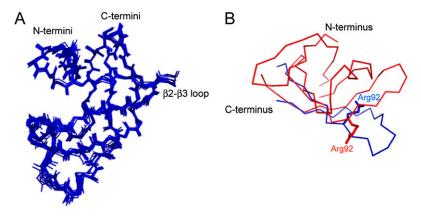


Fig. S1. The refined NMR solution structure of gpFII. (*A*) An overlay of 10 low-energy structures of gpFII. Only the structured portions of the protein are shown, and these overlay with an average rmsd of 0.9 Å (Table S1). These structures have been deposited into the PDB as 2KX4.pdb. Although residues of the three unstructured regions were removed for clarity, their positions are indicated. (*B*) An overlay of gpFII NMR structures 1K0H (the old gpFII structure) and 2KX4 (our newly deposited gpFII structure). Residues 85–108 of 1K0H (blue) are overlayed onto the 2KX4 structure (red). The overlay reveals the incorporation of the previously under-refined region into a β -sheet, as well as the change in conformation of residue Arg92 from a buried to an exposed position.

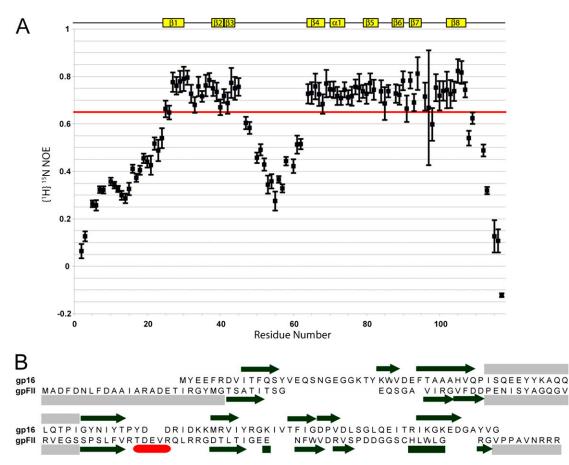


Fig. 52. The NMR solution structure of gpFII displays significant disorder. (*A*) Values of the heteronuclear $\{^{1}H\}^{15}N$ NOE as a function of position in gpFII. Residues that fall below the red horizontal line representing a cutoff of +0.65 show increased dynamics on the fast, ns-ps time scale (1). These include residues 1–24, 46–62, and 109–117. The positions of gpFII's regular secondary elements are indicated as yellow bars above. (*B*) The sequences of gpFII and gp16 are aligned according to the structural alignment. The secondary structure of gp16 is indicated above its sequence and that of gpFII is shown below its sequence (β -strands are represented as black arrows, the α -helix as a red cylinder, and unstructured regions as gray boxes). Unstructured regions of gpFII (Fig. S2A) and gp16 (2) were accurately delineated by measuring $\{^{1}H\}$ -15N heteronuclear NOEs.

^{1.} Pawley NH, Wang C, Koide S, Nicholson LK (2001) An improved method for distinguishing between anisotropic tumbling and chemical exchange in analysis of 15N relaxation parameters. J Biomol NMR 20:149–165.

^{2.} Lhuillier S, et al. (2009) Structure of bacteriophage SPP1 head-to-tail connection reveals mechanism for viral DNA gating. Proc Natl Acad Sci USA 106:8507–8512.

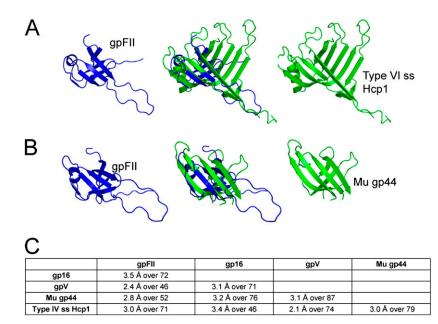


Fig. S3. Overlays of gpFII used to build gpFII oligomeric models. (A) Overlay of gpFII to an Hcp1 monomer. (B) Overlay of gpFII to residues 167–326 of Mu gp44. The structural overlay of gpFII and Hcp1 and gp44 is displayed in the center and the individual monomers are displayed on either side. (C) Summary of all structural alignments discussed in this work. Initial alignments were made using DaliLite (1) or FATCAT (2) and then refined in Swiss-PDBViewer (3).

1. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16:566-567.

- 2. Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19 (suppl 2):ii246-ii255.
- 3. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18:2714–2723.

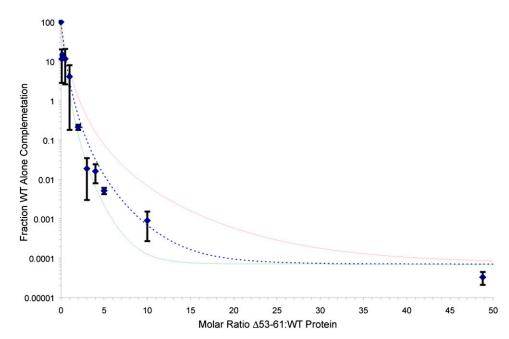


Fig. 54. In vitro titration of the $\Delta 53-61$ gpFII protein. Equal volumes of purified WT and $\Delta 53-61$ gpFII protein were mixed in vitro before testing for complementation of an *Fllam* extract. WT protein was kept at a concentration of 1 µM and $\Delta 53-61$ gpFII protein concentration was varied from 0.1 to 50 µM. Even at very low ratios, e.g., 1:1, the $\Delta 53-61$ gpFII protein was able to exert a dominant negative effect on WT protein. The results were normalized to WT protein-only and no-protein-added controls and represent the average of three experiments. These data could be fit to a simple exponential equation. We have shown theoretical curves of models in which one (green), two (dashed blue), or three (red) mutant monomers (assuming that gpFII is a hexamer) are required to abrogate function of a mixed ring of WT and $\Delta 53-61$ gpFII protein. Theoretical curves were based on a simple model that if a single mutant monomer is sufficient to abrogate function only hexamers that are made up entirely of WT protein will function. The probability of the formation of a functional hexamer is equivalent to the probability of a monomer being WT at all six positions or P(WT)⁶. This probability was calculated for all concentrations of mutant protein tested, normalized to a maximum activity of 100 and a minimum as defined by the "no gpFII added" control, and plotted with experimental data. This was repeated for a model in which two [P(WT)⁵] and three [P(WT)⁴] monomers would be required to abrogate function. Our data support a model in which one or two mutant monomers of six are sufficient to block head-tail binding activity. Although the model is simple and variables such as relative affinities or positional effects of mutants within a hexamer are not accounted for, it predicts the dose-dependence effects well.

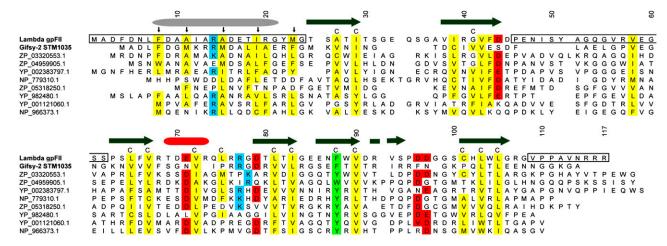


Fig. S5. Sequence and structure comparison of gpFII and its homologues. Sequence alignment of a diverse set of sequences from phages and prophages that display similarity to gpFII as detected through PSI-BLAST (1) searches. The average pair-wise identity of these sequences to gpFII is 18% and no pair of sequences displays greater than 30% identity. The homology of these proteins with gpFII was supported by the genomic position of the genes encoding them and their co-occurrence with genes encoding λ gpW homologues. Each sequence possesses an N-terminal region that is predicted to be helical. The amphipathic character of these putative helices is indicated by the *i*+3, *i*+4 spacing of conserved hydrophobic positions as indicated by small arrows. The secondary structure of gpFII are boxed within the sequence alignment. Secondary structure predictions were performed using the Jpred server (2). The Gifsy-2 protein accession number is YP_001700680.

1. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nuc Acids Res 25 (17):3389–3402.

2. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. Nuc Acids Res 36:W197-201.

	λαρEll		50
T. Det Balancia T. MARE OND DET IL. 14.14 YE YE I. TE CE IM MARE IS A THE IS A THE OS INTER IN SUBJECT			
M. 00000001 I M. NOW ALV M. D. OWA V. 00-10000000000000000000000000000000000			
9.92014 1 MARK & TOP ALL & LE KOWAT ALL EL KOWAT ALL		1	57
T. D. 20139271 1 N. M.S. S. R. NO, C. Y. A. S. L. M.W. H. E. M. NO, C. A. N. L. M.W. H. E. T. D. A. S. S. P. M. S. M. NU, E. T. D. A. S. S. P. M. S. M. M. S. M		MIDEDKLALEPCMATFA - EPAKWRSGRLADWVA - ITGVYDDAYFPLDPLGGDDGMGGVHITT (0
M. C. DOD, OT R. J. D. DOT, M. T. BUL, DOT, M. T. BUL, M. C. D. L. M. C. BUL, M. K. BUL, K. BUL			
1			
9.992921 1	YP_002419153.1		
mp. mp. 460213.1 1		1	6
1			
27.03280201 1		1	52
P. 01062311 1		1	56
P. 00232941 1		1	53
P. 023214641 1		1	5
P. 0019309101 1			
Imp. 34921.1 Imp. 34 Will OR K. MT MA PLO LL DOV DE V F OV V F A V LA - DO V K LE GAS - V K CAM PA RA OU RO K. SA WILL OR K. SA		1	0
P. 02278733.1 1			
P2-9348041 1		1	54
1		1	48
2P.9031703.1 1			
P. 0.0121060:1 1		1	4
YP.2043601 1			
AC13411.		1	52
2P.0328863.1 1 ••••••••••••••••••••••••••••••••••••			
2P.0329368.1 1			
2P_01259588.1 1			
AppFil 61 S S S S F W R T D E V R - Q R - R G DT L T G			
GFW_225MI035 44 N V V F A C N V V F A C S E T V T R I A F R F N C R C V T L E N N G G G A 99 20.3320571 59 V A F V F B S A W V F A C N V V A A C G V V V A A O L V V V K D O G O T L R G K A A V V T B K A W V F G S S A A S I S Y 118 22.0332051 59 T A F V D O A C K C K A F N M H C K V V M A C G V V V A A G O A G C K K L H N V D G 118 118 22.0332051 59 T A F V D O A C K C K A F N M H C K N V M A A C G D V V D A G O C C K K S G A A D F K A B V R S K R V G S O A A 110 72.0101951 53 V L F V S D V C A O A C V K K C O G O F L V T V A G S D R V V D I R O G A K L I H R V D A 107 70.0403246.1 57 V S A V C G A C A V K K C O G O F L V T A A G C A K L I H R V D A 110 110 72.01019531 53 V L R V L G A C A V K K C O G O F L V A T A G C A K L V K V D A C A A C V K K C O C O F L V A A C A C K K K V O B A C A C K K K V O B A C A C K K K V O B A C A C K K K V O B A C A C K K K V O C A C A K A A A V O C A C A K A A A A V C C K A A A V C C K A A A A V C C K A A A A C A V K K C A C A C K K K V O C A C A K A A A A A A V A A C A A K K K A A C A V K K C A A C A K K K A A C A V K K C A A C A K K K A A C A V K K C A A C A K K K A A C A V K K C A A C A K K K A A A A A V C A C A A K K K A A A A A A A A A A A A			
2P. 03205331 59 Y A P R F V Y S D A - CMT PK A R V D G			
2P.0939905.1 55 S E P L Y R D K D A K - G L K - I R Q K T Y A			
2P.03043761.1 99 TA M V LF RS A WV A - GL X - K V D R VM II H		56 SEPELYLROKDAK-GLK-IROKLTVA	
2P.0019631 33 VL		59 T A P V L F <mark>T</mark> R S A W V A - G L K - K Y <mark>D</mark> R V M I H G E P Y Q V V D P - G W D E S G T A G H G V I T I T L A R G E P G R S T P A A P V R P S K R Y G S Q R A 1	133
PP.001462546.1 SP TL PV LG W R DAAC RVK PK K. QG DIFL V W T		57 T A <mark>P V L D V L L A D L P</mark> V E <mark>P</mark> E - E <mark>G</mark> Y T L T V A G S D <mark>Y</mark> R V V D I - R <mark>T</mark> DG M G G A K L I L H R V Q P	
YP.B66788.1 S8 YR			
YP_00168106.1 57 YS_NV_LG_VOLS Q_F_PS_D_PM_ON_QOLS_VASV NTT LV KE_V_R_VDS R_GAKLLLSKVSS P YP_192733.1 G1 SR_PV_LG_VOLS Q_F_PS_D_PM_ON_QOLS_VASV SKM_R_R_HEV_QOLSG_F_R_R_R_K_R_VDS R_RC_QALL_V_NAATG_SDA			
YP_192733.1 61 S R P V G VQ L S A L W E P E Q G D L G V R G K M R I H E M Q A D G R G G A L L V N D A T G E S D A L S C D H P 121 YP_001445177.1 77 E H G V L T V F NM R I E K Q L V - S G A S V D P N N R M R V T W D D E T S L I L S M K P N G Q R V S P N D H L 138 YP_00139822.1 SS T S L R O L P S M Q V P V V N N R M R V T W G A N E S R S E L I V F K Q R Q Q H S W L D E N 120 YP_001039822.1 SS T S L R O L P S M Q V P V V V N - A V T E Y G V G V G T R I T A V S O V S T E A G S G R S E L I V F K Q R Q Q H S W L D E N 121 YP_001039823.1 SS D M L D V R L S V A L E D A L E D A L E D G A V T E Y C V G G L T R I D T R H L V R T C E V A V L M D D D D D D D D D D D D D D D D D D			
YP_002383797.1 76 HA A A F S A MTT D I V. G L S - R H D E W V N N		61 S R P V L G V Q L S A L A V E P E - Q G D L I G V R G K M Y R I H E V - Q A D G R G G A L L V L N D A T G E S D A L S C D H P 1	121
YP_00103982211 SE TS C P V V Q Q V P V Q Q V P Q Q V E P AG S G R E I V F Q Q I D R E D			
YP_002419153.1 55 55 D A M L I D V R L S E V A E P A E G D A T D L N E D G A V T E T V Q V I G L T R D T R R V R T C E V A P V L P D D P D E D D E P 121 118 YP_001608955.1 57 A R K L D V C I N D I G G I P R K Q D S V V W T L E S T K D S F S Q E H F I V T D V - Q A S E S G M W K L I L R AQ E R 116 116 D P o 55792.1 61 V Q V Q R L G Q V Q Q E A A R V - Q D Q L H I R R T S C R V R W H T P L R D N S G M W K L Q A S C 116 110 ZP_05786505.1 53 D T M F I D V R V S E A P - Q D V - Q G S V F E L L D H A G Q P T G E S F E V Q D E P R R D G C G F L R L L L H D G A S P 101 111 ZP_05786505.1 53 D T M F I D V R V S E A P - Q D V - Q G S V F E L L D H A G Q P T G E S F E V Q D E P R R D G C G F L R L L L H D G A S P 103 101 YP_001232141 54 M A P T W R D D L F P V - E G A V T V R G R V E V D L I P P A - T P A A D A L R C T L R V A P 108 108 NP_779310.1 56 P E P S F T C K E S D W - D F K - K H D Y A R I E D R H V R L T H D Q Q D G T G M A L V R L A P M A P P 107 108 NP_2012322464.1 61 - G R V E V S F L N V V E P E - Q G R V E V G G E V F V I S K K L S D D G S K S W U V R R G 107 107 YP_00137310.1 58 R V E V G D N A G V E V C R V V O I P P E C G G G T T T I A G G S C S K I W K I Q A G E G S V S 100 107			
YP_001608965.1 S7 A R K E L D V C I N D I G G I P P K P Q D S V V WILLE S T K D S F S Q E H F I V T D V Q A S E S G M Y K L I L R A Q E R 118 YP_051605.1 S1 V Q P L G I Q V S Q F E A A P V Q D D Q L H K R T G L I V V R E A A L D G WG G G C M L N L I G Q G G G 101 ZP_05786505.1 S3 D T M F I D V R V S E A P - V M - Q G S V F E L L D H A G - Q P T G E S F E V Q D P R R D G - E R L V W K A E A V R E V 101 ZP_05786505.1 S3 D T M F I D V R V S E A P - V M - Q G S V F E L L D H A G - Q P T G E S F E V Q D P R R D G - E R L V W K A E A V R R E V 101 ZP_05786505.1 S3 D T M F I D V R V S E A P - D M - Q G S V F E L L D H A G - Q P T G E S F E V Q D E P R R D G - E R L V W K A E A V R R E V 101 ZP_05786505.1 S3 D T M V R D L V P V R D L V P V R D L V P R D G - Q A V T R A A D G R N U V V D V R P D G - E R L V W K A E A V R R E V 108 NP_70310.1 S6 P E P S F T C K E S D V M - D F K - K H D Y A R I E O R H R R L T H D P Q P D G - T G M A L V R L A P M A P P 107 YP_001327324.1 S9 T H L A V S A S T V P - G L A S Q R D S V A I G G V Y V I N I - D D A - R A M R I S L A G D I 108 YP_002237633.1 S9 <td< td=""><td></td><td></td><td></td></td<>			
NP_966373.1 S2 E I L L L E V S V F DV L K P M - V G DT F S I G			
ZP_05786505.1 S3 DT MF I D V R V S E A P D V M - Q G S V F E L L D H A G - Q P T G E S F E V Q D E P R P Q - E R L V W K A E A V R E V 111 ZP_05786505.1 S7 V Q P L L E V W K O D C T P P V E Q A V T V R - G R N L I V D V - R P D G C G F L R L L H L D G A S P 109 VP_0010425232402.1 57 W A P T W V R D L V P P D R R R A D		61 VQPRLGIQVSQFEAAPV-QDDQLHIKRTGLIYVVREA-RLDGWGGGGLMLNLIGGDGG	
ZP_05292402.1 S7 VQ_PL_L E V × E Q_R_V_L V × E Q_R_V_L V V V × R P_Q_G			
YP_001045231.1 S4 MAP YW VR R D VR VR VV V P E A R R VV VV VV P P A R VV VV P P A D VV VV VV P P A T P A D VV P A T P A D A T P A D P P A VV P A D P P A D A P A D A P A D A D A D A D A D A D A D A D D D D D A D			
NP_779310.1 56 PE P S F T C K E S D V M - D F K - K H D Y A R I E D R H Y R L T H D P O P D G T G M A L Y R L A P M A P P			
YP_001327324.1 59 T T H L L A V S A S T V P G L A S Q R D S V A L G - - G V T V Q V I N I - D D A - R M L R I S L A G D I 108 YP_001327324.1 61 - G R V E V S F L R V D V E P E - Q G D R V E V G - - G E V F V I S K K L B D D - - G S R S V V V R G - 107 YP_00197109.1 53 A S L E D D D V I K E - S R F V K I F P P L K D S - S K I W K I Q V G E S S V S 106 NP_745192.1 52 R E P H L V I R V G D N A - G V E V G D V V D L P P E - - G T F F E I A P H L A S - L R K A V - 106 YP_419741.1 49 G T A V F E I R R E V P A - P A Q A G D V I V H D - - - - A E D F I A P H I A D S - L R L V V V D L P A A 98 ZP_05844704.1 55 F T R L D A R I A D L P - T P A - P G D L I R I G C - - A E D F I L Q G E P K L D A - - R L I W V C A A D C C S V V Q C A A - - 88 ZP_05844704.1 55 F T R L D A R I A D L P - T P A - P G D L I R I R G V P - - - - A E D F I L Q G E P K L D R - - R L W V V D L P A A			
YP_001975109.1 S3 Q A S E R D Q D V T S F S - I G D V I K E	YP_001327324.1	59 TT H L L A V S A S T V P - G L A S Q R D S V A I G	108
NP_745192.1 52 R E P L V I R V G D N A - G V E - V K Q R V V D D P P E D G G T Y T I A G S E P G G D G L V T L L R K A V		61 - GRVEVSFLRVDVEPE-QGDRVEVGGEVEV <mark>G</mark> SRSVWLVRRGGSRSVWLVRRGGSRSVWLVRRG	
YP_002287833.1 55 P T V L I D V R R S E V A E P A - S G D T V E F D G D T F E I I A P P I A D S L R L V W V C E A S L Q G		53 Q I AS I E UR DQ DV T · S FIS - I G G V I K I E · · · · · K R F K K F F P P I K DS · · · · S S K I WK Q A V G E G S V S · · · · 1	
YP_419741.1 49 G T A V F E I R R G E V P A P Q - A G D V I V H D - - G D S F V V Q G E P R L D A - E R L V W T L D T R P A - 98 ZP_05844704.1 55 E T I R L D A R I A D L P - T P A A P Q D L I R G - A E D F I L Q E P K L D A - E R L V W T L D T R P A 98 ZP_05031703.1 55 N O P I C G I L S D L P V T P K · V Q D R L I V R - - A E D F I L Q G E P K L D R - E R L I V W T L D T R P A 103 ZP_05031703.1 53 T S E V I L V R S E V P - S P A - A G N I V T L A E S - G L Q R V T G D P K V Q R D R - R G R W I C V V A P V 103 YP_001121660.1 53 A T H R F D V M A D V A - D P R - E G D R F T V A - - G T N V R V G D P L V D - R C R W I C V V A P V 102 YP_0015565651.1 56 L A - C R L P A D M V P - G I A - A G E P I D D - - - - - C M V R Q V F P E A 102 YP_0015565651.1 56 L A - C R L P A D M V P - G I A - G G P I D D - - - - - - - - 102 XP_0315656951.1 56			
ZP_05844704.1 S5 ET R L D A R I A D L P T P A P C D L I R I C - - A E D F I L Q C E P K L D R - - R - 104 VP_171505.1 55 N C P I C G I R L S D L P V T P K V O D R L I V R - - - A E D F I L Q C E P K L D R - N V G V L V L Q N A - 103 2P_05031703.1 53 T S E V I L V R S E V P S N O P I C G I R K D N V T A E S - - - - N V G V L V L Q N A - - 103 YP_001121060.1 53 A T H R F D V M A R D V A - P R - E G D R F V A - - - - - R V C V A P V - 102 YP_0015565651.3 5 L A - C R I P A O M K P - - - - - - - R V R Q V F P E A - 102 ACY34311.1 53 C Q T V L P A O K C P - A R V V V V V V V R - - - - - - - - - - - - - - - - - - -			
YP_171505.1 55 NQ P I C G I R L S DL P V T P K · V G D R L I V R · · · · G V N V R I A E G · Q E D G · · · · N V G V L L V L Q N A · · · · 103 103 ZP_05031703.1 53 T 5 E V I L V R R S E V P · S P A · A G N I V T L A E S · · · G L Q Y R V T G D P K V D R · · · K G R W I C V A P V · · · · · 103 103 YP_001121060.1 53 A T M R F D V M A R D V A · D P R · E G D R F T V A · · · · · G Q P V G V G D P L V D · · · · R D R L I W T L T G A P V · · · · · · · 102 102 YP_001121060.1 53 S A R T C S L D L A L V P · G I A · A G I L V I N · · · · · · · G T N R V S G G V E P D E · · · · T G W V R L Q V F P E A · · · · · · 102 102 YP_001565695.1 56 L A - · C R L P A DM V P · G I A · Q G E P I D I D · · · · · · · · · · · G K A R I A E S · D P D E · · · · · · · · · · · · · · · · · ·		55 ET <mark>T</mark> RLDARIADLP-TPA-PGDLIRIGAEDFILQGEPKLDRERLIWSLDTRPA1	
YP_001121060.1 S3 A T HR F DVM A R DV A - DP R - E G D R F T V A	YP_171505.1	55 NQ P I C G I R L S DL P VT P K - V G D R L I V R	103
YP_982480.1 S3 SART CSLDLAL VP - GIA - AGGILVIN GTN VR VSGV PDE TGWVRLQVFPEA		53 T S E V I L V R R S E V P - S P A - A G N I V T L A E S G L Q Y R V T G D P K V D R K G R W I G V V A P V	
YP_001565695.1 56 LA - CRLPADMVP - GIA - QGPIDID		SATHREOWMARDVA - DPR - EDDRETVA	
ACY34311.1 53 ΤΟ PTV LV PLO K C P - A N P - V G LP IS IN			
ZP_05318250.1 47 A D <mark>P Q I I V T E D D L P</mark> - E D V - K S V V V T V R G K R Y A V A E T - D F D G C G M V V V Q L R A I H D K <mark>P</mark> T Y 100 ZP_02408866.1 51 V A P M L H V I D A D A A - G I Q - P G S V V T A R S G R Y R V V E A - Q P D G T G W T I L T L Q 96			
	ACY34311.1		
ZP_01259588.1 50 A R K L T F K K A D N V - R P K - K G T P I E F V S S G R K L T V T S G - P Y P E D G N I V V I L 96 N I V V I L 96		47 A D P Q I I V T E D D L P - E D V - K S V V V T V R	
	ZP_05318250.1 ZP_02408866.1	47 A D <mark>P</mark> Q I I V T E D D L P - E D V - K S V V V T V R G K R Y A V A E <mark>T</mark> - D F D G C G M V V V Q L R A I H D K <mark>P</mark> T Y	100 96

Fig. S6. An alignment of gpFII homologues. These proteins were detected through PSI-BLAST (1) searches and all display detectable sequence similarity to gpFII. Conserved positions are highlighted with the "ClustalX" coloring scheme as implemented in Jalview (2). All of these homologues possess an N-terminal region that is similar in sequence to the unstructured N-terminal region of gpFII. The conserved hydrophobic positions in this region, as indicated by arrows, support our hypothesis that this region forms an amphipathic α -helix in the assembled particle.

1. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

2. Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. Bioinformatics 20:426-427.

SANG SAN

Ging, 2 TM 103.1n 5 Ging, 2 TM 103.1n 6 Fint, 2 TM 103.1n <	λgpW	1 MTRQEELAAARAALHDLMTGKRVA - TVQK DGTRVEFTAT SVS DLKKYIAELEVOTGMTQRRGPA	
22 032054841 1 NMM ME R R V H A OCU H C GV H GV GV <td< td=""><td></td><td></td><td></td></td<>			
YP 00233801.1 2 T K S E 00 ML V T RO M RO S D G K S V S - F GV - N G A T TH - D V A R K LE V W K R MO A V MG R T - N F K - L N F L			
27 03033221 1			
YP 1927261 23 T 10 A D A A D A A D A AD A D A AD A D A AD A D A D A AD A D </td <td></td> <td></td> <td></td>			
YP_000341914.1 B DL AR R TO A'S KL R DA M A'S CL L Y CG A DO T G YL TR R Y A M RL A CD L MOQ NAL DAQ Y GG - YT R R TR Q V W TG R S GW 44 44 YP_001393161 12 S K E Q L KAAL NE A Q A T I E L S G R R V F S M A Q D G T T Y T Y S Q A L A D L L A L A T L D R A C L E C A CL R R R R S S F I R R S S F I N R N Y Y Y N N S R Y C S S Y Y Y N N S R Y C S S Y Y Y Y N N S R Y C S S Y Y Y Y N N S R Y C S S Y Y Y Y Y N N S R Y C S S Y Y Y Y Y Y N N S R Y C S S Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y			
YP_001039816.1 102 A R S HA K RM D A LE ALL E K R AT R DO G T T YT S O A M DELL K L R DR Y R E HO R LE Q A G L R F R Q S F I R YS L S			
YP_0014625421 12 S K E Q L K A A L N E A Q A A Y I E L S G R - R G Y E S A Q G D G T Y Y S Q A S R A D L L A L TA Y L D R A L G L R T R S L			
YP_001608970.1 14 R L E S L K R R E Q L E A L			-
22 049599011 102 PK AS DAEKA 0 A I R M L A N R AT A DQLKLS S F 0 C S I S L K K F R A V M V A K K R K S G. R G L L K I T K V R R R 175 YP, 5579661 14 T D A QLQ A A L A A Q Q M I D L R T G D N K K V V A Y A Q G D S I S V F Q T T N L A V M M I R D L I A A L N G V K M C K R R N M V V K K K K R R N M V V K Y A Q G D S I S V F Q T T N L A V M M I R D L I R A L R M K K R K R N M V I F S T N K G V D R I M PP.960585011 1 - M Y W Q A E L D A L K A A R A Q A M T D L S G R K V I N A Y A Q G D T K S V Y Q Q S D I T I R K A L E K I A A M M K R R R N R V K M C K R R V G A K D R G R 69 YP.001851131 S T A D M Q S A L A A A A A Y D A S M T D L S G R K I V R Y N O G D T K S Y Y Q Q S D I T I K A L E M M A G R R M A R A R A R N K K R K A R R R R V A S K G L - 70 70 YP.001045221 2 T L D D M E R R T G L D A I R A R - 5 K S M L 7 (R A T D Q Q R Y R I - N N R E W T - P I A E L L K R T F N A V A V Q R E T T D T H - R S W N V V R F V G - 70 70 YP.001327331.1 10 V L L A R L D A R A I A D A I N A V L Q R A R T A R Q V R - 1 G L S Y - N N K E S Y T N A A N I G A L R Q V R I T - 7 R S W G N V V V R F V G - 78 78 ZP.01019671 1 M A DY E T D A A I R S A I - S K G A R - F R R V A - 1 G K R I A N A A V R E F R - 7 T I N R A I A R A R A R S R - 6 K Y R T T K R A I R R A R - 8 S R - 6 K Y R T T K R A I R R A R A R S R R - 7 Y V Y A A Q R R I R A R A R R S R R - 7 Y Y A Y A Q R R I R A R A R A R R			
YP_557966.1 14 T DAQLQAALAAAQQAY I DLRTGD-KAVV VAYAQCDGS VT QTTNLAAVRMFIRDLGAALNGV VKMCKRR- RMV PLF- 89 NP_966367.1 1-MAW Q DL DALSAN C D V E KAIK KU N G C - RVV I AV A Q ODS N S VT QTTNLAAVRMFIRDLGAR ALNGV VKMCKRR- KI V F STNKGV DKIM 77 ZP_0578661.1 1-MAW Q AL DALSAN C A V T AV A Q ODS T L STNKGV DKIM 78 66 KK P K KI V S STNKGV DKIM 76 ZP_0578661.1 1-MAW Q AL DALSAN C A V T AV A Q ODS T K STV V Q S D IT DIRKA FEM LONDILS R Q R KK P K A R W V S F AK F DR R GP 80 YP_001858113.1 8 S A DMQ S R LAAL Q AY F DL S G A - K I V A T N NG D STK STV Y Q S D IT DIRKA FEM LONDIL R V AN ALG I I C NY R A R R V K A S G G K K P K A R R V K A S G G K K P K A R R V K A S G G K K P K N R P A X K G L - 70 YP_00185201.1 1 D D M R R L T G L L D I R - M R G V R G K R K T A R - N S G G K V M A A N D L R V A X A V C R T T D T P R S W N V P V F V N R P A X K D L - 69 N P Y AV K D L - 60 90 NP 773314.1 100 E R S L A R A I A A K E T A R N O L A A I A A V E P Y A V K D L - 60 90 N P Y A V K D L - 60 90 N P Y A V K D L - 60 90 90 90 10 L A R V Y A V A V C R K R A A - A N A V C R K R A A - A N A V C R K R R A - R - R R R K T A R - N K V V V V V V V V V V V V V V V V V V			
NP_96367.1 1 - MY N E DY L 10 V E E AT K K L 0 N G E - RY V L 1 A Y - G DH V V R A E V D I N D L L S L R Q R I K A E L K I A G M K R K R - K I V F S T N K V V K I M 77 ZP_05786501.1 1 - MAW C Q E L D A L K S K Y - R Q G T L - R V S Y - E G K L V F Y Q S D I D L N S L R Q R I K A E L K I A G M K R K R - V G F A K F D R R G P - 69 ZP_0528357.1 2 T L E L K Q L D A L R A R A R - Y S M - T Y N Q G D T K Y Y Q S D I D I R K I F M L D M A L G I C Y F R - R I R Y A S K G L - 70 YP_00185226.1 2 T L D DM E R L T G L D I R R A R A A L - Y S M G S K Y E N G S E R Y E N S A A L O D L R K I A R K T A R R V I R Y A S K G L - 70 YP_001327331.1 10 V L A R D E A R E A L HO M E L G R A E I E L S Y - N R E W R E A L K Q Y A K L G L R F A R A - R S R - G V I F G 28 - 72 ZP_001327331.1 10 V L A R D E A R E A L HO M E L G R A E I E L S Y - N R E W R E A L K Q Y A K L G L R F A R A - R S R - G V I F G 72 - 72 ZP_001327331.1 10 V L A R D E A R A I - A S Q G L V K A O D Q I Y Q - S V E A M L A A I A A V E E Y A E			-
2P.05786501.1 1 MAWT QAEL DALKSAY AQGTL RVSY EAGLLRRIRTIE EAGLLRRIRTIE EGELAASGGGKKEK K V V OBS6113.1 BS ADMQSKLAALQAAYFDLSSGA K V AT N QGDTKSVTQQSDITE DITDIRKALENDALGILRKA EALQMALGILRKK K V ARRVLF BI VP.0010452261 2 T EELKAALQAAYFDLSSGA K V AAR QAAYFDLSSGA GS K S ARRVLF ARRVLF BI QAALGILRKK ARRVLF ARRVLF BI QAALGILKK ARRVLF ARRVLF ARRVLF BI QAALGILRKK ARRVFTAR <			
YP_001858113.1 8 S A A M Q S R L A A Q A A Y F D L S G A - K I V T A T N N G D G T S S V T Q Q S D T D I R K A T E M L D M A L G I I C N Y R - A A R V L F			
2P_05292397.1 2 T E E L K A Q L D A L R A R Y S G M L T V K A Y K A O A L Q A L A A Y L Q Q A L A A Y L Q A L A A Y L Q A L A A Y L Q A L A A Y L Q A L A A Y L Q A L A A Y L Q A L A A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Q A L A Y L Y S S G L S Y R H G D R Y Y A . G D H Y N Y G L Y A A A T L Q R Y A A T L Q R Y A A T L Q R Y A A T L Q R Y A A T L Q R Y A A Y Q Q R L Y A Y Q Q R L Y A Y Q Q R L Y A Y Q Q R L Y A Y Q Q R L Y A Y Y Y A Y Q Q R L Y A Y			
YP_001045226.1 2 T D DM E R R L T G L L D I R			
NP_779314.1 109 G R S D NQ R A L D A L NA V L Q K R A T Q D Q Q R Y R I			
YP_001327331.1 10 V L L A R L D E A R E A L H Q M - E L G R A E I - S L S M - NG E S V T Y A A A N L G A L R Q Y V R D L E A K L G L R R F A R A - R S R - G V I F G			
2P_02190589.1 2 T P DR L A AWR E AL LEAR FRG VL TVK A - G DK S V TYR S DA E LAA A I AAV E F P VAE			3
ZP_01011967.1 1 M A D Y E T D L A A I R S A I S K G A R S V Q M N G E R V E F R S L S E M Q R I E R K I ME D L G L S A T T R R Q V P T T K S Q W 71 71 VP_002321456.1 1 M S W T K D D V Q R L K A A I A S G Q L S V A M G D R Q I T Y Q S V E A M I T A L D R M A E V A A T T A G R R K S A T R R W F T T I K S F 71 71 VP_001975114.1 2 Y E E Y L T Q V E K A I Q K L 0 S G E R V V S I A M G D R Q I T Y Q S V E A M I T A L D R M A E K K A G M R P K R K K Y T S T S K G I L 74 74 NP_745187.1 1 M A Y I Q A H L A A V E R A I A R G E R V V R Y S D R T V E Y R M V D E L I K A R D L I R T E L S Q S A G P R S R V K I Y H G - G K G I L 74 76 VP_866795.1 3 T A E L Q I R R D K L L E R L E S L Q T R V S H G D R T V O F D L S Q A R E A L D L D R I A N G S R R I A R H I R V R S N K D L 71 71 VP_002287828.1 2 T L E E M K A R R K K L A R R N S L V A R W T V G D R T V O F D L T Q A M H L A D D R I A M T E G G R K R R R I L A V A T K G L 69 70 ZP_05844699.1 2 T L A E M K A E K K V L A R R N S L V A R W T V D G R V T Y A S D A E L A T A I F D L R I A A T R E G R K R R Y I R F A S K O L 71 71 YP_03173.1 1 M A F T Q E D L Q A L E A A 1 A S G C D E V A Y S D R N Y R K - T L S O M L Q A R D M K K E L A V A R R F K S L A V F E G D 69 70 YP_01121065.1 9 A M T Y P E H L Q A L R E A L A S G C D E V A Y S D R N Y R K - T	YP_001327331.1		
YP_002321456.1 1 M S WT K D D V Q R L K A A I - A S G Q L - S V R M - - G D R Q I T Y Q - S V R M L T A L D R M E A E V A A T T A G R R - K S A T - R R Y R F T T L R G F - 71 YP_001975114.1 2 Y E E Y L T Q V E K A I Q K L - - O S G E - R V V S I A Y - - D H V R Y G E V Q I N D L I K A R Q I I K A Q I I K A Q I I K A C I K K A L I K A K L I K K A L I K K A Q I R K A L K V A G M R P K R - - K I V F S T S K G I L - 74 NP_745171 1 M Y Q A H A A Y E R A I R G V R - - C D H V V R Y G E VQ I N D L I K A R Q I I K A E L K V A G M R P K R - - V F S T S K G I L - 74 NP_7451731 1 A E Q I R D K L L E R L - E S L Q T - - R V S H - - D N V Y A - D D A W A A T T A G R R - - - 60 K T V Q F D - L S Q A R E A L D L L D R I A A R T E G G R K R - - N L R A S K G L - 71 YP_002287828.1 2 T L E E M L A Q R D A L L A A R - F R G V R - I V Y - D R V Y A - D A E M A A L T D E R I A A T D E R I A A S D P A V A - R P R - V R I L A V A K G L - 71 YP_02387361 2 T L A E M Q A L L S A L L G M R - F G G V R - I T Y - D G R Q I	ZP_02190589.1	2 T P D R L A AWR E A L L E AR F R G V L T V K A G D K S V T Y R S D A E L A A A A A A V E E P V A E	
YP_001975114.1 2 Y E E M I Q C E N Y E Y N Y Y E Y N Y	ZP_01011967.1	1 M A D Y <mark>T</mark> E T D L A A T R S A T S K G A R <mark>S</mark> V Q M <u>NG E</u> R V E F R <mark>S L</mark> S E M Q R T E R K T M E D L G L S A T T	
NP_745187.1 1 M A Y I Q A H L A A Y E R A I - A R G E R - V R Y - S D R T Y E Y R - I Y D E L I K A R D L I R T E L S Q S A G P R S R - V R G S R - 1 - A R G L C R R Y T - S D R T Y E Y R - I Y D E L I K A R D L I R T E L S Q S A G P R S R - - V R G S R Y T - - 71 YP_002287828.1 2 T L E E M L A Q R D A L L A R - - E S Q R Y T Y A - S Q R T Y Q A H L A A T R E G R K R - R I A T R E G R K R - - R I A T R E G R K R - - R I A T R E G R K R - - R I A A T R E G R K R - - R I A T R E G R K R - - R I A T R E G R K R - - R I A T R E G R K R - - R I A A T R E G R K R - - R I A T R E G R K R - - R I A T R E G R K R - - R I A T R E G R K R - - R I A A T R E G R K R - - R I A A T R E G R K R - - R I A A T R E G R K R - - R I A A T R E G R K R - - R I A A T R A S G L - - D R A T Y Y A - D R N Y Y A - D R N Y Y A - D R N Y Y A - D R N Y Y A - T I A A N M L A A T R E A R A A - N R A S K G L - - N R A T R	YP_002321456.1	1 - <u>M</u> SW <mark>T</mark> KDDVQRLKAAIAS <mark>G</mark> QL <mark>S</mark> VRHGD <mark>R</mark> QITYQSVEAMLTALDRMEAEVAATTAGRRKSATRRYRFTTLRGF 71	
YP_866795.1 3 T I A E L Q I R R D K L L E R L	YP_001975114.1	2 Y <mark>S</mark> EE Y LTQVEKAIQKL QSGE RVV SIAY GDHVVRYGEVQINDLLNLRQRIKAELKVAGMRPKR KIVESTSKGIL 74	
YP_002287828.1 2 T L E E M L A Q R D A L L A A R F R G V R F R G V R F R G V R F R G V R F R G V R	NP_745187.1	1 - MAY TQAHLAAVERAI ARGER VVRY <u>5</u> D <mark>R</mark> TVEY R TVDELIKARDLIRTELSQSAGPRSR VVRLYHG - GKGL 68	
YP_419736.1 2 T L D E M K A E R E K V L A R R · · · N S L V A · · · R V T V · · · G D R T V Q Y D · · · L T Q A N H V L A D L D R R I A M L E G K K P R R · · · · R I L A V A T K G L · · · 69 69 ZP_05844699.1 2 T L A E MQ A L I S A L G M R · · · F G G V R · · · S I Y V · · · G D R V Y G · · G D R L A T A I F D E R R I A A S D A V A · · R P R · · · · V F R F A S K D L · · · 71 71 YP_171509.1 1 · · M A F T Q E D L Q A L E A A · · · A S G C D · · · E V A Y · · · G D R V Y R · · F T S O L L K M L D Y R T K A S L S K R T T T R · · · · Y R F A S K D L · · · · Y Q Y D S D R · · · F Q O S D R · · · F Q O S D R · · · Y R · · · · Y R F A S K D L · · · · Y Q Y D · · · L T S O M L Q A R D L M R K E L L G K G R T T T R · · · · Y R V F A S L S C D · · · E V A Y · · · · G D R V Y R · · · T S S O L L K M L D Y F R T K A S L S A S P T P P S · · · · · L A L V F E G D · · · 69 YP_001121065.1 9 S M T Y T P E H Q A L R E A A · · · A S G E L · · · R Y T Y · · · E G K S I E Y R · · S V A D L K A A I A E V E A T I A R E S G A P K S · R Q I · · · R V A S · · K A L · · · 77 YP_902165699.1 3 T Q D Q M R L A R I N A I · · · H S G E R · · · T I T T · · · E D G A S VT Y R · · S L D M I Q K G R E WE A K V T A E A T A A G · V P T F G G L R W K F R F A G L R D R · · · 74 YP_003279612.1 1 · · · · M V K Y L E A E A A L · · · L G K · · · · I V S F · · · G G R T L T V E · · N L D S I R K G R Q E W E R A A A E Q R R S G · · V G G L · · S Y S V A R F D G P D R · 68 ZP_02408871.1 137 · M A Y T R Q D L R (Q A E A A H · · · L G G K M V · · S Q K · · · G Q R K V F · · · A D R R K Y R · S R N I H E L R A Y I D D L R G Q L G L S S V R R R G · · · P A G V S · · · · R I R A - G K S V	YP_866795.1	3 T I A E L Q I R R D K L L E R L E S L Q T R V S H G D K T V Q F D L S Q A R E A L D L L D R E I A R N G G S R R I A R H L R V R S N K D L 71	
ZP_05844699.1 2 T L A E MQ A L L S A L L G M R F G G V R F G G V R F G G V R F G G V R F G G V R F G G V R V A R P R V F R F A S A D P A V A R P R V F R F A S K D L 71 VP_171509.1 1 - M A F T Q D L Q A L E A A I A S G C D E V A V F G D R N Y K - T L S O M L Q A R D M R K E L G K G R T T T R Y I Q Y R P D Q S D R 69 ZP_0503375.1 2 P A P D Y A V I A A E I A A A A S G E L E V A V F O D R N Y K T S S O L L K M D Y F R K A S I A S P P P S	YP_002287828.1	2 T L E E M L A Q R D A L L A A R F R G V R T V E I D G R R V T Y A S D A E M A A A L T D L E R R I A A T R E G G R K R	
YP_171509.1 1 - MAFTQEDLQALEAAT - ASGCD EVAY - GDKRVRYK TLSQMLQARDLMRKELLGKGRTTTRR YIQYRPDQSDR 69 ZP_005033715.1 2 FAPDYAVEIAALEAAA ASGEL EVAY - COK FV - SDLLKMLDYRTKAS ASTTTRR YIQYRPDQSDR 69 YP_0012105.1 9 MYTPEHLQALEAAA ASGEL EVAY - COK FV - SDLLKMLDYRTKAS ASTTTRR YIQYRPDQSDR 69 YP_0012105.1 9 MYTPEHLQALEAAA ASGEL EVAY - COK FV - SVADLKAAIAEV ATIARESGAPKS RQI RVAYS KAL 77 YP_982484.1 3 H NIMLQLKLDAEAAV - LAGK	YP_419736.1	2 T L DEMKAEREKVLARR NSLVA RVTV G D AT V Q Y D L T Q ANHVLADLDRR I AMLE G KK P R R R T L AV AT K G L 69	
2P_05033715.1 2PAP D YAVE I AAL E AAA AS G E L IVE S D G D R V T Y R - S T S D L L K M L D Y F RT K AS I AS P T P P S L AIL L A V F E G D 69 69 VP_001121065.1 9 S M T Y TP E H L Q AL RE AL AS G E L R Y T Y E G K S I E Y R - S V AD L K AA I AE VE AT I AR E S G AP K S - R Q I R V A S K AL 77 79 VP_00125065.1 9 S M T Y TP E H L Q AL RE AL AS G E H R Y T Y E G K S I E Y R - S V AD L K AA I AE VE AT I AR E S G AP K S - R Q I R V A S	ZP_05844699.1	2 T L A E M Q A L L S A L L G M R F G G V R <mark>S</mark> I T Y D G R Q I T Y G <mark>S</mark> D A E L A T A I F D L E R R I A A S D P A V A R P R V F R T F A S K D L 71	
YP_001121065.1 9 S M T Y TP E H L Q A L R E A L A S G E H R Y T Y E G K S I E Y R S V A D L K A A I A E V E A T I A R E S G A P K S - R Q I R V A T S K A L 77 YP_902184.1 3 H N I M L Q L Y L D A E A A V L A G K T Q G E A G R T H T M E - S L D M I Q K G R R WE A K Y TA E A T R A A G - V P T F G G L R Y K A R F N 74 YP_001526569.1 3 T L Q L Q MR L A R L NA I H S G E R H T I T E D G A S Y T Y E D G A S Y T Y	YP_171509.1	1 - MAFTQEDLQALEAAIASGCD <u>EVAY</u> GDKRVRYKTLSQMLQARDLMRKELLGK <u>GRTTT</u> RRYLQYRPDQSDR69	
YP_982484.1 3 H N I M L Q L Y L D A E A A V - · · L A G K · · · · T I Q F · · · A G R T H T M E · S L D M I Q K G R R E WE A K V T A E A T R A A G · V P T F G G L R Y K T A R F N · · · · · 74 YP_001565699.1 3 T L Q D L Q M R L A R L N A A I · · · H S G E R · · · T I T T · · · E D G A S V T Y R · · S L D E M M G A R R D L H T Q I S A V A G T G Q · · · S R A L V A R F R F A G L R D R · · · 76 YP_001565699.1 1 · · · M V A K Y L E A E A A I · · · L G K · · · · T V S F · · · G G R T L T V E · NL D S I R K G R Q E WE R R A A A E Q R S S · · V G G L · · S Y S V A R F D S P D · · 68 ZP_002279612.1 1 · · · · M V A K Y L E A E A A I · · · L G K · · · · T V S F · · · · G G R T L T V E · · NL D S I R K G R Q E WE R R A A A E Q R S S · V G G L · · S Y S V A R F D S P D · · 68 ZP_02408871.1 137 · M A Y T R Q D L D R I Q S A I · · · A K G E L · · · E Q Y · · · · A D R R V K Y R · S I L E L R E A Q T E I I R A L D G A S G R S R I V · · · · · R R H A G K G V R · · · · G D R K V K Y R · S I L E L R E A V I D D I R G Q L G A S G R S R I V · · · · · · P A G V F · · · · · · · R R H A G K G V F · · · · · · · · · R R H A G K G V F · · · · · · · · · · · · · · · · R R H A G K G V F · · · · · · · · · · · · · · · · · ·	ZP_05033715.1	2 PAP DY AVE TAALEAAA A SGEL TVE S DG DR VTYR ST SDLLKML DYFRTKASTAS PTPPS LATLAVFEGD 69	
YP_001565699.1 3 TLQDLQMRLARLNAAI HSGER TITT EDGASVTYR - SLDEMMGARRDLHTQ ISAVAGTGQ SRALVARFR AGLRDR 76 YP_003279612.1 1 MVAKYLEAEAAL LLGK TVSF GGRTLTVE - NLDSIRKGRQEWERAAAEQRSG - VGGL SYSYARFDGPDR - 68 ZP_02408871.1 137 - MAYTRQDLDRIQSAI AKGEL EVQY ADRVKYR - SLLELREAQEUEIRAAAEQRSG RSRIV RLRHA - GKGVR - 204 YP_001454517.1 137 - MAYTRQDLQRIQSAI AKGEL EVQY ADRVKYR - SLLELREAQEUE IRAADEQRSG RSRIV RLRHA - GKGVR 204 ZP_02408871.1 1 MKQE MQQAEAAFHSLQTGKMAV - SVQK GDRKVEYSRAN HELRAY IDVLRGQLGLSSVRRG PAGVSF	YP_001121065.1	9 S M T Y <mark>T P</mark> E H L Q A L R E A L A S G E H R V T Y E G K S I E Y R <mark>S</mark> V A D L K A A I A E V E A T I A R E S G A P K S - R Q I R V A T S K A L 77	
YP_003279612.1 1 MV A K Y LE A E A A L L L G K T V S F G G R T L T V E NL D S I R K G R Q E WE R R A A A E Q R R S G V G G L S Y S V A R F D G P D R - 68 ZP_02408871.1 137 - MA Y T R Q D D R I Q S A I A K G E L E V Q Y A D R R V K Y R - B L E L R E A T E I I R A D G A S G R S R I V R I R H A - G K K V R - 204 YP_001445171.1 1 - MT K Q E M L Q A E A A F H S L Q T G K MA V - S V Q K G D R K V E Y S R A NI H E L R A Y I D D L R G Q L G L S S V R R G P A G V S F 69 ZP_0125982.1 3 K Q E M L Q A E A A F H S L Q T G K MA V - S V Q K G D R K V E Y S R A NI H E L R A Y I D D L R G Q L G L S S V R R G P A G V S F 69	YP_982484.1	3 H <mark>S</mark> N I M L Q L <mark>Y</mark> L D A E A A V L A G K T I Q F A G R T H T M E <mark>S L</mark> D M I Q K G R R E W E A K V T A E A T R A A G - V P T F G G L R Y K T A R F N 74	
ZP_02408871.1 137 · MAY TRQ DL DR I Q S A I · · · A K G E L · · · E V Q Y · · · · A D R V K Y R · · SI L E L R E A Q T E I I R A L DG A S G R S R I V · · · · · R L R H A · G K G V R · · · 204 YP_001445171.1 1 · · MT K Q E M L Q Q A E A A F H S L Q T G K M A V · S V Q K · · · · G D R K V E Y S R A N I H E L R A Y I D D L R G Q L G L S S V R R G · · · · P A G V S F · · · · · · 69 ZP_01259582.1 3 · · · · K Q E M L Q Q A E A A F H S L Q T G K M A V · S V Q K · · · · G D R K V E Y S R A N I H E L R A Y I D D L R G Q L G L S S V R R G · · · · P A G V S F · · · · · · 69	YP_001565699.1	3 T L Q D L Q M R L A R L N A A I H S G E R T I T T E D G A S V T Y R S L D E M M G A R R D L H T Q I S A V A G T G Q S R A L V A R F R F A G L R D R 76	
ZP_02408871.1 137 · MAY TRQ DL DR I QS A I · · · A K G E L · · · E VQ Y · · · · A D R V K Y R · · SI L E L R E A Q T E I I R A L DG A S G R S R I V · · · · · R L R H A · G K G V R · · · 204 YP_001445171.1 1 · · MT K Q E M L QQ A E A A F H S L Q T G K M A V · S V Q K · · · · G D R K V E Y S R A N I H E L R A Y I D D L R G Q L G L S S V R R G · · · · P A G V S F · · · · · · 69 ZP_01259582.1 3 · · · · K Q E M L QQ A E A A F H S L Q T G K M A V · S V Q K · · · · G D R K V E Y S R A N I H E L R A Y I D D L R G Q L G L S S V R R G · · · · P A G V S F · · · · · · 69	YP 003279612.1	1 M V A K Y L E A E A A L L L G K T V S F G G R L L T V E N L D S I R K G R O E WE R A A A E O R R S G V G G L S Y S V A R F D G P D R 68	
YP_001445171.1 1 MT K Q E ML Q Q A E A A F H S L Q T <mark>G</mark> K MA V - <mark>S</mark> V Q K G D <mark>R</mark> K V E Y S R A N I H E L R A Y I D D L R G Q L G L S S V R R G P A G V S F 69 ZP_01259582.1 3 K Q E ML Q Q A E A A F H S L Q T G K MA V <mark>S</mark> V Q K G D <mark>R</mark> K V E Y S R A N I H E L R A Y I D Y L R G Q L G L S S V R R G P A G V S F 69			4
ZP_01259582.1 3 KQEMLQQAEAAFHSLQT <mark>G</mark> KMAV - <mark>9</mark> VQK GD <mark>R</mark> KVEYSRANIHELRAKIDYLRGQLGLSSVRR <mark>G</mark> PAGVSF 69			ŝ
	ZP 05318245.1		

Fig. 57. Alignment of putative gpW homologues. These proteins were identified through examination of prophage genomes containing genes encoding homologues of gpFII as detected in PSI-BLAST (1) searches. These proteins were encoded between the genes encoding the large terminase subunit and the portal protein. The alignment above shows that these proteins are significantly similar in sequence to gpW and are likely to be homologous. Conserved positions are highlighted with the ClustalX coloring scheme as implemented in Jalview (2). The gpW homologues shown in this alignment correspond to the gpFII homologues shown in Fig. S6 and are displayed in the same order.

1. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. 2. Clamp M, Cuff J, Searle SM, Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* 20:426–427.

	Statistic
Distance restraints	
Short rage, $ i-j \le 1$	1026
Medium range, 1 < li-jl < 5	81
Long range, li-jl ≥ 5	337
Hydrogen bond pairs (HN-O, N-O)	14
Dihedral angle restraints	
φ/ψ pairs	51
Pairwise RMSD	
Assigned atoms*	
All backbone atoms	0.9 Å
All heavy atoms	1.6 Å
Ramachandran statistics [†]	
Most favored regions	77.6%
Additional allowed regions	22.4%
Generously allowed regions	0%
Disallowed regions	0%

Table S1. Structural statistics for the ensemble of 10 structures of gpFII

*RMSD for residues 25-46, 60-96, and 100-116.

[†]Determined using PROCHECK-NMR. Residues selected based on dihedral angle order parameter with $S(\phi)+S(\psi) \ge 1.8$.

VAS PNAS

Table S2. In vitro addition experiments

Experiment	+Heads	+Tails, gpFII	+Buffer
WT	100	100	100
∆1–9	72	71	0.00079
∆53–61	58	0.0076	0.0019
Δ115–117	61	0.042	0.00090
No protein	53	62	0.0011

Table S3. TEM particle counting

PNAS PNAS

Count	Heads	Tails	Phage	% Phage
WT	88	69	44	22
∆1–24	71	117	1	0.5
∆53–61	207	106	2	0.6
∆115–117	129	97	0	0
(-) control	105	78	0	0