

Lawrence Berkeley National Laboratory

LBL Publications

Title

Combined Solution and Crystal Methods Reveal the Electrostatic Tethers That Provide a Flexible Platform for Replication Activities in the Bacteriophage T7 Replisome

Permalink

<https://escholarship.org/uc/item/1vr8k9r2>

Journal

Biochemistry, 58(45)

ISSN

0006-2960

Authors

Foster, Brittni M

Rosenberg, Daniel

Salvo, Henry

et al.

Publication Date

2019-11-12

DOI

10.1021/acs.biochem.9b00525

Peer reviewed



HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2020 November 10.

Published in final edited form as:

Biochemistry. 2019 November 12; 58(45): 4466–4479. doi:10.1021/acs.biochem.9b00525.

Combined Solution and Crystal Methods Reveal the Electrostatic Tethers That Provide a Flexible Platform for Replication Activities in the Bacteriophage T7 Replisome

Brittni M. Foster^{1,&}, Daniel Rosenberg^{2,3}, Henry Salvo¹, Kasie L. Stephens^{1,#}, Brittanica J. Bintz¹, Michal Hammel², Tom Ellenberger⁴, Maria D. Gainey¹, Jamie R. Wallen^{1,*}

¹Department of Chemistry & Physics, Western Carolina University, Cullowhee, NC 28723, USA

²Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

³Graduate Group in Biophysics, University of California, Berkeley, Berkeley, CA 94720, USA.

⁴Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA

Abstract

Recent structural studies on the bacteriophage T7 DNA replication system have shed light on how multiple proteins assemble to copy two antiparallel DNA strands. In T7, acidic C-terminal tails of both the primase-helicase and single-stranded DNA binding protein bind to two basic patches on the DNA polymerase to aid in replisome assembly, processivity, and coordinated DNA synthesis. Although these electrostatic interactions are essential for DNA replication, the molecular details for how these tails bind the polymerase are unknown. We have determined an X-ray crystal structure of the T7 DNA polymerase bound to both a primer/template DNA and a peptide that mimics the C-terminal tail of the primase-helicase. The structure reveals that the essential C-terminal phenylalanine of the tail binds to a hydrophobic pocket that is surrounded by positive charge on the surface of the polymerase. We show that alterations of polymerase residues that engage the tail lead to defects in viral replication. In the structure we also observe dTTP bound in the exonuclease active site and stacked against tryptophan 160. Using both primer/extension assays and high throughput sequencing, we show how mutations in the exonuclease active site lead to defects in mismatch repair and an increase in mutagenesis of the T7 genome. Finally, using small angle X-ray scattering we provide the first solution structures of a complex between the

*Corresponding Author: jamiewallen@email.wcu.edu. Tel: +828-227-2491.

&Department of Cancer Biology, Wake Forest University School of Medicine, Winston-Salem, NC 27103, USA

#College of Pharmacy and Health Sciences, Campbell University, Buies Creek, NC 27506, USA

The authors declare no competing financial interest.

Supporting Information.

The Supporting Information is available free of charge on the ACS Publications website.

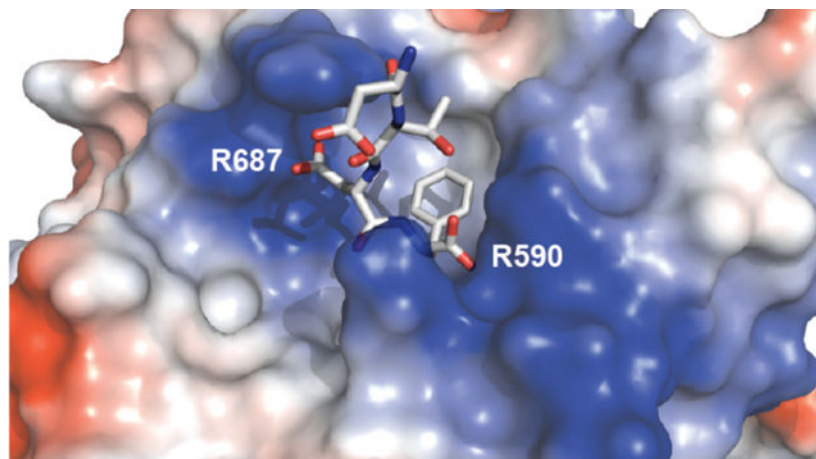
Full results of whole-genome sequencing of the T7 5 phage stock, as well as the three wild-type, W160A, and 5A7A passage ten samples, are provided in Supplemental Table S1.

Accession Codes

The coordinates of gp5 bound to a primer/template DNA and the gp4 peptide have been deposited in the Protein Data Bank as entry 6P7E, Gp5: P00581; Trx: P0AA25; Gp4: P03692; Gp2.5: P03696

single-stranded DNA binding protein and the DNA polymerase and show how a single-stranded DNA binding protein dimer engages both one and two copies of DNA polymerase.

Graphical Abstract



Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. Consult the journal's Instructions for Authors for TOC graphic specifications.

Bacteriophage T7 has long served as a model system for studying the complex reactions of DNA replication, as only four proteins are required to efficiently synthesize both the leading and lagging strands of DNA^{1,2}. The T7 DNA polymerase (gp5) forms a stable 1:1 complex with its processivity factor, *Escherichia coli* thioredoxin (trx), and synthesizes both the leading and lagging strands of DNA. The primase-helicase (gp4) catalyzes nucleotide-dependent DNA unwinding via its C-terminal helicase domain to provide the single-stranded DNA (ssDNA) template needed by gp5/trx. Additionally, the N-terminal primase domain of gp4 synthesizes short four-nucleotide RNA primers required to initiate the synthesis of Okazaki fragments. The single-stranded DNA binding protein (gp2.5) serves to coat and protect the ssDNA generated during replication and also directly contacts both gp4 and gp5/trx to coordinate DNA replication^{3,4}. The fully assembled T7 replisome carries out DNA replication with a processivity of ~17,000 nucleotides per DNA binding event^{1,5}.

Although a relatively simple system, biochemical and structural data have revealed that the T7 replisome is highly dynamic, with the four proteins forming multiple transient complexes during different stages of DNA replication. For example, the acidic C-terminal tail of gp4 binds two basic patches on gp5 known as the front basic patch (FBP) and the thioredoxin binding domain basic patch (TBD), and current data suggests that these electrostatic interactions are required for both the loading of gp5/trx on DNA and to keep a gp5/trx molecule that has disengaged from a DNA substrate tethered to a moving replisome⁶. Both the acidic nature of the gp4 tail (seven of nineteen residues are either a D or E residue) and the C-terminal aromatic phenylalanine are required for gp5 binding⁷. When gp5/trx is loaded on DNA the binding to gp4 changes to a high affinity interaction that no longer requires the

gp4 acidic tail⁸, although recent structural reports show that the tail of gp4 still engages the polymerase in the presence of DNA^{9,10}. A third, distinct complex between gp4 and gp5/trx, termed the priming mode, occurs during the initiation of lagging strand DNA synthesis. The four-residue ribonucleotide primer generated by the gp4 primase domain is too short for gp5/trx to efficiently extend alone, and a stable interaction between the zinc binding domain (ZBD) of the gp4 primase and a small loop in gp5 (residues 401–404) allows for efficient primer utilization by gp5/trx^{11–14}. The gp2.5 protein also contains an acidic C-terminal tail that harbors a phenylalanine residue at the carboxyl terminus, and this tail also binds to basic patches on gp5/trx^{3,15}. As with gp4, both the acidic nature and C-terminal F residue of the gp2.5 tail are required for gp5/trx binding¹⁶. Unlike gp4, the interaction between gp2.5 and gp5/trx requires the acidic tail both in the presence and absence of DNA³.

Although structures of all four T7 replisome components have been known for some time^{17–19}, crystal structures of T7 replisome complexes have been difficult to obtain, likely due to multiple transient protein interactions. We recently determined a low resolution crystal structure of the electrostatic interaction between a heptameric gp4 ring and three copies of gp5/trx in the absence of DNA²⁰. The structure revealed asymmetric binding between the gp4 ring and gp5/trx molecules, with two of the polymerases loading on to gp4 side by side and forming a weak interface via the fingers subdomain. Two of the three polymerases were positioned where they could engage the acidic C-terminal tail of gp4, and although clear electron density was present for the gp4 tails leading to the FBP⁸ of the gp5 molecules, the low resolution of our structure prevented accurate modeling of the gp4 tail. Our attempts to crystallize the complex in the presence of DNA failed to produce crystals, and subsequent small angle X-ray scattering (SAXS) studies revealed that the shape of the gp4:gp5/trx complex changed dramatically in response to DNA substrates²⁰.

In addition to our crystal structure, an ~14 Å cryoEM structure of a hexameric gp4 ring bound to two gp5/trx molecules on a fork-shaped DNA substrate was also published, and this structure provided the first insight into T7 replisome assembly on DNA²¹. As in our structure, the cryoEM structure showed the two gp5/trx molecules contacting each other via their fingers subdomains, suggesting a possible mechanism for how the polymerases communicate to coordinate replication of both strands. Most recently, high-resolution cryoEM structures of gp4:gp5/trx complexes engaging in the synthesis of both DNA strands revealed three polymerases bound to a helicase hexamer¹⁰. While one gp5/trx molecule engages the C-terminal helicase domain of gp4 to synthesize the leading strand of DNA, two polymerases were observed situated on the N-terminal side of gp4¹⁰. Such an orientation positions these polymerases to engage the N-terminal primase domain for Okazaki fragment synthesis, and indeed the lagging strand polymerase was shown to engage the ZBD of gp4 during primer handoff.

The recent structures of the T7 replisome have greatly expanded our knowledge of how replication proteins coordinate leading and lagging strand DNA synthesis. However, there are still several factors that are poorly understood. For example, current efforts to obtain structures of the T7 replisome have focused specifically on complexes between gp4 and gp5/trx, and to date there is no structural information on how the gp2.5 protein engages either gp4 or gp5/trx. Additionally, although it is well-established that the acidic tails of both gp2.5

and gp4 are essential for gp5/trx binding, the gp4 tails have been disordered in all structures published to date, and current data fail to pinpoint the exact site of tail binding on gp5/trx. Recent work by Zhang *et al.* identified four basic residues (K587, K589, R590, and R591) in the FBP of gp5 that are required for gp4 tail binding⁸, and in our crystal structure we observed density for the acidic tail near these residues²⁰. The recent high-resolution cryoEM structures also show that the tails of gp4 are in close proximity to both the FBP and TBD of gp5, but again the gp4 tails appear disordered and could not be modeled¹⁰.

To observe gp4 tail binding to gp5/trx, we crystallized a complex of gp5/trx bound to a peptide that mimics the acidic gp4 tail. Although the majority of the tail is disordered in our structure, we have identified the binding site of the essential C-terminal F566 residue of gp4 on gp5. Our structure reveals that F566 is anchored to the polymerase by both a salt bridge between the main chain carboxylate of F566 and R590 of gp5 as well as several hydrophobic residues of gp5 that bind the F566 aromatic ring. The intimate hydrophobic interaction observed provides a molecular explanation for a requirement of an aromatic residue at the C-terminus of gp4. Additionally, although high concentrations of dTTP were included in the crystallization setup to lock the DNA polymerase on DNA, our crystal structure reveals dTTP bound in the gp5/trx exonuclease active site. The dTTP complex suggests an important role for gp5 residue W160 in exonuclease activity. Using whole-genome sequencing, we show how mutagenesis of key residues in the gp5 exonuclease active site impacts replication of the T7 genome *in vivo*. Finally, to begin to understand how gp2.5 assembles in the T7 replisome, we provide SAXS data of a complex between gp2.5 and gp5/trx on DNA. The solution structures support a model in which a gp2.5 dimer uses its flexible C-terminal tails to engage one or two polymerase molecules.

Materials and Methods

Oligonucleotides, Peptides, and Proteins.

All deoxynucleotides were either synthesized using an Applied Biosystems 394 DNA/RNA synthesizer or were purchased from Integrated DNA Technologies (IDT). To generate primer/template substrates for gp5/trx, equal concentrations of complementary oligonucleotides were mixed in 10 mM Tris pH 8.0, 50 mM NaCl, and 1 mM EDTA, heated to 95 °C, and then allowed to slow cool to room temperature in a water bath. A peptide of sequence SGEEESHSESTDWSNDTDF, which mimics the C-terminal tail of gp4, was synthesized on a CEM Liberty Automated Microwave Peptide Synthesizer.

All gp5 mutants were generated by site-directed mutagenesis using wild-type gp5 in pET21b as a template. Correct mutations were confirmed by sequencing. All gp5/trx proteins were expressed and purified as described previously¹⁷. Full-length gp2.5 was subcloned into a pET21b vector between NdeI and XhoI sites, verified by sequencing, and expressed in BL21(DE3) cells. The pET21b vector has an engineered precision protease site designed to remove the N-terminal poly-His tag. Gp2.5 was purified by loading onto a 5 mL HiTrap Nickel column equilibrated in 50 mM Tris pH 8.0, 200 mM NaCl, 10 mM imidazole, 1mM 2-mercaptoethanol, and 10% glycerol. The column was washed and gp2.5 was eluted using 125 mM imidazole. Fractions containing gp2.5 were identified by SDS-PAGE, pooled, mixed with precision protease in a 1:100 precision protease:gp2.5 ratio, then dialyzed

overnight in 20 mM Tris pH 7.5, 200 mM NaCl, 10 mM imidazole, 1mM dithiothreitol, and 5% glycerol. The protein was then passed back over a 5 mL HiTrap Nickel column to remove both the cleaved His tag and the His-tagged precision protease. The resulting flow-through containing gp2.5 was diluted to lower the NaCl concentration to 100 mM, then loaded onto a 10/100 MonoQ column equilibrated in 20 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA, 1mM DTT, and 5% glycerol. The column was washed and gp2.5 was eluted using a 100–800 mM NaCl gradient. Fractions containing gp2.5 were pooled, concentrated, and then loaded onto a Superdex 200 column equilibrated in 20 mM Tris pH 7.5, 200 mM NaCl, 0.5 mM EDTA, 1mM DTT, and 5% glycerol. Fractions containing gp2.5 were pooled, concentrated, flash-frozen in liquid nitrogen, and stored at -80°C .

Structure Determination and Refinement.

Crystals of gp5/trx complexed with a primer/template DNA and a peptide mimicking the C-terminal tail of gp4 were formed by mixing 100 μM gp5/trx with 120 μM primer/template DNA, 0.5 mM ddATP, 5 mM dTTP, and 0.5 mM gp4 peptide in a buffer containing 20 mM Tris pH 7.5, 100 mM NaCl, 5 mM MgCl_2 , and 2 mM DTT. The primer (5'-GGCAGGTGGTCTTGCCGGTG) and template (5'-CCCATCACCGCAAGACCCTGCC) oligonucleotides provide a five-base 5' overhang once annealed. To form a stable gp5/trx:DNA complex, the gp5/trx terminates the primer by addition of ddAMP, and the next incoming nucleotide (dTTP) locks the complex in place. Crystals were grown by hanging-drop vapor diffusion at 22°C by mixing 1 μL protein with 1 μL reservoir placed over a 500- μL reservoir containing 25% PEG 4000, 0.1M LiSO_4 , and 0.1 M Tris pH 7.5. Plate-like crystals appeared within one day and grew to full-size (~ 300 – $500\ \mu\text{m}$) in one week. For preparation for data collection, crystals were transferred to a solution that contained 30% PEG 4000, 0.1M LiSO_4 , 0.1 M Tris pH 7.5, 5 mM MgCl_2 , 5 mM dTTP, and 0.25 mM gp4 peptide for two hours, then flash-cooled in a gaseous nitrogen stream at 100 K. X-ray data were collected at a wavelength of 0.979180 at beamline 24-ID-E of the Northeastern Collaborative Access Team (NECAT) at the Advanced Photon Source using an ADSC QUANTUM 315 detector. Data were processed using XDS²².

The structure of the gp5/trx:gp4 peptide complex was determined by molecular replacement using the program Phaser²³ in the Phenix software suite²⁴. Gp5 polymerase coordinates from PDB 2AJQ were used as the starting model, and for the molecular replacement search the flexible TBD of gp5 (resides 264–328), DNA, and trx were removed. Phaser correctly located the positions of all four gp5 molecules in the asymmetric unit. The trx subunits of each gp5 molecule were positioned in the molecular replacement structure using density-modified electron density maps generated in Phenix. Rigid body refinement was initially performed on the gp5/trx structure, then additional components of the structure including the primer/template DNA, TBD, and gp4 peptide were built using density-modified maps. Manual structure rebuilding was followed by xyz coordinate, real space, and individual B-factor refinement using phenix.refine. Both non-crystallographic symmetry and secondary structure restraints (using PDB 1T7P) were included during refinement. All structure figures were generated using the program Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). A summary of data collection and refinement statistics are provided in Table 1.

Phage Complementation Assays.

The T7 5 bacteriophage used for all complementation assays was a gift from Bill Studier, Brookhaven National Laboratory. The T7 5 bacteriophage was plaque purified to generate a stock that was used for subsequent experiments. BL21(DE3) cells harboring a pET21b plasmid containing either wild-type gp5 or a gp5 mutant were grown overnight at 37 °C in LB broth containing ampicillin. 0.5 mL of saturated bacterial culture was mixed with 4.5 mL melted 0.7% LB agar containing ampicillin at 55 °C, then plated on solid 1% LB agar containing ampicillin and allowed to solidify. A stock of T7 5 phage (titer of 7.8×10^9 pfu/mL) was serially diluted in LB broth, then two μ L of each dilution was spotted in triplicate onto the agar plates containing the bacterial cells within 15 minutes of plating. Plates were stored at 37 °C for seven hours, then viral titers were counted. All titers shown in Table 2 are from replicate experiments performed in triplicate.

Primer Extension Assays.

For primer extension assays, either a 25-base normal Watson-Crick primer (5'-TGGCAGGCAGGTGGTCTTGCCGGTC) or a 25-base primer containing a 3'-terminal mismatch (5'-TGGCAGGCAGGTGGTCTTGCCGGTA) was annealed to a 35-base template (5'-GCGATCCCTAGACCGGCAAGACCACCTGCCTGCCA). Primers were radiolabeled at the 5'-end using adenosine triphosphate, labeled on the gamma phosphate with 32 P, and polynucleotide kinase. 50 nM of either wild-type gp5/trx, 5A7A gp5/trx, or W160A gp5/trx were mixed with 20 μ M of either normal or mismatched primer/template DNA and 1mM dNTPs in 20 mM Tris pH 7.5, 100 mM NaCl, 5 mM MgCl₂, 2 mM DTT, and 0.1 mg/mL BSA at room temperature. Timepoints were collected at 1, 10, and 60 minutes and quenched in 20 mM EDTA and formamide, then loaded on a 14% denaturing urea PAGE gel. DNA extension products were visualized by phosphorescence.

Viral Passaging and Whole-Genome Sequencing.

The stock T7 5 virus was serially passaged ten times at a multiplicity of infection (MOI) of 0.001 in *E. coli* BL21(DE3) cells harboring a pET21b plasmid containing either wild-type, 5A7A, or W160A gp5. For passaging, two mL of an overnight culture of *E. coli* containing a pET21b gp5 plasmid were added to eighteen mL of LB broth containing ampicillin to give a starting cell OD₆₀₀ of ~0.4 (~ 3.2×10^8 cells/mL). Stock T7 5 phage was then added to the bacterial cell culture at an MOI of 0.001, and the cultures were allowed to shake at 37 °C for four hours. After four hours, the cultures were spun at 2,600 RPM for 10 minutes, and the resulting supernatant containing virus (passage one) was tittered and stored at 4 °C. For each polymerase, this process was repeated for a total of ten passages (passage one virus was subsequently used to infect a fresh culture of cells at a MOI=0.001, and four hours later passage two virus was collected and tittered, etc.). Each passage was stored at 4 °C after harvest.

To prepare for whole-genome sequencing a plaque assay was performed using *E. coli* BL21(DE3) cells harboring a pET21b plasmid containing wild-type gp5. Three individual, well-isolated plaques from each polymerase passage ten sample were picked and transferred to LB broth. Viral stocks were then grown from each plaque pick by serially diluting each sample and infecting 0.5 mL of BL21(DE3) cells containing the pET21b wild-type gp5

plasmid. Webbed-plates for each plaque sample were flooded with 8 mL of buffer containing 10 mM Tris pH 7.5, 10 mM MgSO₄, 68 mM NaCl, and 1 mM CaCl₂. After 24 hours at 4 °C the buffer solution was pipetted off the LB agar plate and sterile filtered. One mL of the phage solution for each sample was used to purify T7 genomic DNA by first performing a DNase/RNase pretreatment followed by a phenol/chloroform extraction. For a reference control sample, we also purified genomic DNA from the original T7 5 stock that was used for viral passaging.

The NEBNext UltraII DNA Library Prep Kit and NEBNext index primers were used to create libraries from viral genomic DNA following manufacturer's protocol. AMPure XP beads were used for size selection. Ten viral genomic libraries were then pooled, denatured, and sequenced using a MiSeq Reagent Nano Kit v2 according to Illumina protocols. Illumina sequencing was performed at the Western Carolina University Chemistry and Physics Biotechnology Core using an Illumina MiSeq FGx in RUO mode. Sequence data was demultiplexed and adapter sequences were trimmed during primary analysis using Illumina® Real Time Analysis (RTA) software. Resulting raw fastq files were uploaded into CLC Genomics Workbench software v12.0. All failed reads were filtered from the data set prior to secondary analysis. Remaining reads were quality trimmed when a) base accuracy decreased below 95% b) two or more ambiguous nucleotides were called sequentially and 3) when a sequence read was less than 15 bp in length or greater than 1000 bp in length. Retained reads were mapped against a bacteriophage T7 reference genome (NC_001604.1–2) to maintain consistency in position number across the genome using the standard CLC mapping algorithm. Variant tables were constructed using the Basic Variant Detection algorithm v2.02 with a specified minimum coverage of 10, minimum count of 2, and frequency threshold of 10%. Four variants from the reference genome were consistently found in all T7 data sets. These variants did not represent mutations of interest and were therefore removed from the data set.

SAXS and MALS data acquisition in line with Size-Exclusion Chromatography.

For small angle x-ray scattering coupled with multi-angle light scattering in line with size-exclusion chromatography (SEC-SAXS-MALS) experiments, 60 µL samples containing either 5 mg/mL (55 µM) of gp5/trx on a primer/template DNA, 5 mg/mL (195 µM) gp2.5 on a dT15 ssDNA, and a mixture of 195 µM gp2.5 and 55 µM gp5/trx in the presence of primer/template and dT15 DNAs were prepared in 20 mM Tris 7.0, 100 mM NaCl, 5mM MgCl₂, and 1mM DTT. The primer (5'-TGGCAGGCAGGTGGTCTTGCCGGTddC) and template (5'-GATCCCTCGACCGCAAGACCACCTGCCTGCCA) sequences for gp5/trx differed from those used in crystallization in that the primer was purchased from IDT terminated by ddCMP, and the next incoming nucleotide specified by the template strand was dGTP. In all samples containing gp5/trx on DNA, 5 mM dGTP was included to lock the polymerase on the primer/template. SEC-SAXS-MALS were collected at the ALS beamline 12.3.1 LBNL Berkeley, California²⁵. X-ray wavelength was set at $\lambda=1.127$ Å and the sample-to-detector distance was 2100 mm resulting in scattering vectors, q , ranging from 0.01 Å⁻¹ to 0.4 Å⁻¹. The scattering vector is defined as $q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20°C and data was processed as described^{26,27}. Briefly, a SAXS flow cell was directly coupled with an online Agilent 1260 Infinity HPLC system

using a Shodex KW803 column. The column was equilibrated with running buffer (20 mM Tris, pH 7.0, 100 mM NaCl, 5mM MgCl₂, 1mM DTT, 0.1 mM dGTP) with a flow rate of 0.6 mL/min. The buffer included 0.1 mM dGTP to keep the polymerase locked on DNA during column elution. 55 μ L of each sample was run through the SEC and three second X-ray exposures were collected continuously during a 30-minute elution. The SAXS frames recorded prior to the protein elution peak were used to subtract all other frames. The subtracted frames were investigated by radius of gyration (R_G) derived by the Guinier approximation $I(q) = I(0) \exp(-q^2 R_G^2/3)$ with the limits $qR_G < 1.5$. The elution peak was mapped by comparing integral of ratios to background and R_G relative to the recorded frame using the program SCATTER (Figure 7A). While uniform R_G values across an elution peak represent a homogenous assembly, a gradual decline of R_G values across an elution peak indicates intrinsic sample heterogeneity. Due to the heterogeneity of the sample, multiple sections of the SEC peak delivered different SAXS profiles that were used for further analysis. Final merged SAXS profiles used for further analysis include a Guinier plot which determined aggregation free states (Figure 7C). The program SCATTER was used to compute the pair distribution function ($P(r)$) (Figure 7B). The distance r where $P(r)$ approach zero intensity identifies the maximal dimension of the macromolecule (D_{max}). $P(r)$ functions were normalized based on the molecular weight of the assemblies as determined by SCATTER using volume of correlation V_c^{28} (Table 3). Eluent was subsequently split 4 to 1 between SAXS line and a series of UV at 280 and 260 nm, multi-angle light scattering (MALS), quasi-elastic light scattering (QELS), and refractometer detector. MALS experiments were performed using an 18-angle DAWN HELEOS II light scattering detector connected in tandem to an Optilab refractive index concentration detector (Wyatt Technology). System normalization and calibration was performed with bovine serum albumin using a 45 μ L sample at 10 mg/mL in the same SEC running buffer and a dn/dc value of 0.19. The light scattering experiments were used to perform analytical scale chromatographic separations for MW determination of the principle peaks in the SEC analysis (Figure 7A). UV, MALS, and differential refractive index data was analyzed using Wyatt Astra 7 software to monitor the homogeneity of the sample across the elution peak complementary to the above-mentioned SEC-SAXS signal validation (Figure 7A).

SAXS Solution Structure Modeling.

Minimal molecular dynamics (MD) simulations were performed on flexible regions in the crystal structure by the rigid body modeling strategy BILBOMD in order to explore conformational space²⁹. The experimental SAXS profiles were then compared to theoretical scattering curves generated from atomistic models using FOXS^{30,31}. Ensembles of conformers for the complex of gp5/trx bound to gp2.5 with DNA present were determined by MultiFOXS³².

Results

Crystals of gp5/trx in complex with a primer/template DNA and a peptide that mimics the C-terminal tail of gp4 diffracted to 3.0 \AA resolution, and the structure was determined using molecular replacement (Table 1). Crystals of the gp5/trx:peptide complex in the absence of DNA were not obtained, likely due to movement of the fingers subdomain, and previous

reports show that the gp4 tail binds gp5 in the presence of DNA⁹. The crystallized complex contains four gp5/trx molecules in the asymmetric unit, and all four polymerases are bound to DNA with the next incoming nucleotide (dTTP) specified by the template positioned in the polymerase active site (Figure 1). Segments of weak electron density prevented complete modeling of the TBD in all four gp5/trx molecules, and therefore residues 302–311 of chain A, 302–308 of chain B, 289–315 of chain C, and 266–321 of chain D were omitted from the final refined model.

Structure of the C-terminus of gp4 bound to gp5/trx.

Gp5/trx was crystallized with a peptide that mimics the C-terminal nineteen amino acids of gp4. This region of the helicase corresponds to the acidic C-terminal tail that has been shown to be essential for interaction with gp5/trx during DNA replication, and the peptide includes the essential C-terminal aromatic F566 (Figure 1). Electron density was observed at the base of the FBP in three of the four gp5/trx molecules in the asymmetric unit, and we have modeled the last four C-terminal residues (D563, T564, D565, and F566) of the peptide into this electron density (Figure 1B). We do observe some electron density at the FBP of the fourth gp5/trx, but not enough to accurately model. Only four of the nineteen residues of the peptide could be modeled, which points to disorder for the majority of the peptide. The peptide is bound with F566 anchored in a hydrophobic pocket on the surface of gp5, with several gp5 non-polar residues providing an intimate fit for the gp4 aromatic side chain (Figures 1C and 2A). This hydrophobic pocket, which is formed by gp5 helices L, P and Q, as well as beta strands 10 and 11 from the fingers subdomain¹⁷, lies near the gp5 FBP⁸ and is surrounded by regions of positive charge (Figure 1B and 1C). The peptide is stabilized by two salt bridges observed between the C-terminal carboxylate of gp4 F566 and R590 of gp5 as well as the penultimate D565 side chain of the peptide and R687 of gp5 (Figure 1B). The F566 carboxyl group is also in an optimal geometry to interact with the main chain nitrogen of K594. The C-terminal F and penultimate D residues are conserved in both the gp2.5 and gp4 acidic tails, so the peptide binding orientation observed in our structure may be a conserved mode of binding for both gp2.5 and gp4. The binding orientation of the gp4 peptide on gp5 is similar to that observed in the structure of a peptide mimicking the C-terminal tail of the *E. coli* SSB protein bound to exonuclease I³³. In that report two SSB peptides were observed bound, and both peptides contained a C-terminal F residue buried in a hydrophobic cleft on the surface of exonuclease I. Additionally, the C-terminal main-chain carboxylate of each peptide formed a salt bridge with an exonuclease I arginine residue.

The C-terminal F566 of Gp4 is Recognized by a Hydrophobic Pocket on the Surface of Gp5.

Previous studies have shown that the C-terminal F566 of gp4 is essential for the recognition of gp5/trx⁷. In the structure the aromatic side chain of F566 forms an intimate fit with a hydrophobic pocket that is formed by residues F487, L565, I569, T572, I593, and L613 of gp5 (Figure 2A). This hydrophobic pocket is optimally shaped for the aromatic side chain of F566, which explains why a substantial loss of gp5/trx binding was observed in both an F566D and an F566L mutant⁷. Mutation of the phenylalanine with an aromatic tyrosine residue shows near wild-type levels of gp5/trx binding⁷, and a model in which we changed F566 to Y shows that in our structure the hydroxyl group of gp5 S568 is well positioned to hydrogen bond to the Y side chain (Figure 2B). An F566L model, on the other hand, reveals

that the leucine side chain would result in a smaller contact surface with the gp5 pocket (Figure 2C). Our structure provides the first high-resolution description of how F566 docks on the DNA polymerase and offers insight on why the size and aromatic nature of the C-terminal residue is critical for interaction with the polymerase.

Mutation of the Peptide Binding Pocket Results in Defective Viral DNA Replication.

To validate the interactions observed between gp5/trx and the gp4 peptide in the crystal structure, residues in gp5 that bind the gp4 peptide were mutated and tested for their ability to complement T7 phage that lacks the gene 5 encoding for DNA polymerase (T7^{Δ5}). For mutagenesis we focused on residues R590 and R687, which bind the C-terminal carboxylate of F566 and the side chain of D565 of gp4, respectively, as well as residues that make up the hydrophobic binding pocket. Complementation results are shown in Figure 3 and Table 2.

Mutation of R590 to alanine (R590A) leads to a complete defect in viral replication, which points to the significance of R590 in binding to the C-terminal F566. Additionally, a R590Q mutation is also defective in viral growth, which points to the essential nature of the positive charge at residue 590. An MD simulation with the C-terminal tail of gp4 modeled into the T7 replisome structure revealed that the highest occupancy interaction observed occurred between the R590 side chain and the main chain carboxylate of F566²⁰, supporting the notion that this interaction is crucial for the recognition of the gp4 tail by gp5. In contrast to the essential nature of R590, mutation of R687 to either alanine or glutamine results in a one to two-log difference in growth as compared to wild-type (Table 2). Despite only a modest defect in viral growth, the R687 mutations resulted in T7 plaques with a smaller plaque morphology as compared to wild-type gp5. In the T7 replisome MD simulations described above, R687 formed a salt bridge with D297 of gp4, while the D565 side chain of gp4 formed a low-occupancy interaction (6.9%) with R490 of gp5²⁰. While R490 lies close to D565 in our peptide structure (Figure 1B), R687 is optimally positioned to form a salt bridge with the side chain of D565. The complementation results support the essential role of R590 in binding the main chain carboxylate of F566; however, the exact role of R687 in binding gp4 remains unclear.

Select residues of gp5 (F487, L565, I569, and T572) shown to contact the aromatic ring of F566 were also mutated to alanine and tested for their ability to complement T7^{Δ5}. The results in Figure 3 and Table 2 show that with the exception of T572 all mutants show a severe defect in phage complementation as compared to wild-type gp5; the T572A mutant shows wild-type levels of phage growth. Of the hydrophobic pocket mutants, the L565A mutation was lethal to the virus, while the F487A and I569A mutants both showed an ~4-log defect in viral growth and a smaller plaque morphology as compared to wild-type gp5. These results provide strong support that the gp5 residues contacting F566 in the crystal structure are critical for the recognition of gp4 during DNA replication.

Nucleotide Binding in the T7 Polymerase Exonuclease Active Site.

To form a stable gp5/trx:DNA complex for crystallization, the polymerase was mixed with a primer/template DNA containing a dideoxy-terminated primer and high concentrations of the next incoming nucleotide (dTTP) specified by the template. During refinement we

observed electron density in the exonuclease active site that we have attributed to dTTP binding (Figure 4A). By including high concentrations of dTTP for the polymerase active site we fortuitously obtained the first structure of gp5/trx with nucleotide present in the exonuclease active site. A comparison of the bound dTTP in the gp5/trx exonuclease active site to the well-studied Klenow fragment editing complexes^{34,35} reveals that the position of the dTTP is equivalent to the 3'-terminal base where phosphodiester bond cleavage occurs (Figure 4B). In Klenow the base of the 3'-terminal nucleotide is stacked between the aromatic ring of F473 and L361, while Y497 forms a hydrogen bond with the scissile phosphate of the terminal nucleotide. Mutation of all three residues severely impairs the exonuclease activity of Klenow, demonstrating that they are critical for proper alignment of the 3'-terminal nucleotide for catalysis³⁶. Equivalent to the stacking observed between the 3'-terminal base and F473 in Klenow, in the T7 structure the base of dTTP is stacked with W160 (Figure 4). A superposition of the structures shows that dTTP does not superimpose perfectly with the 3'-terminal nucleotide observed in Klenow (Figure 4B). As a result, rather than stacking with L11 in T7 (L361 in Klenow), the base of dTTP is stacked between W160 and Y64 (Y423 in Klenow), and in the T7 structure Y170 (Y497 in Klenow) is pointing away from the nucleotide and toward solvent. Given that only a single nucleotide is present in the T7 structure, we would predict that ssDNA in the T7 exonuclease active site would adopt a conformation more similar to that observed in Klenow.

To confirm a functional role for W160 in gp5/trx exonuclease activity, we compared DNA synthesis and repair activities of wild-type, W160A, and the well-characterized exonuclease-deficient D5A, E7A (5A7A) double mutant³⁷ on short primer/template substrates (Figure 5). On a normal primer/template substrate all three enzymes show robust polymerase activity and extend the 25-base primer to the full 35 nucleotide product. A second primer/template containing a G:A mismatch at the 3'-terminus of the primer was also assayed, with the requirement that the polymerase must correct the mismatch before it can extend the primer to a full product. The results of Figure 5 show that while wild-type gp5/trx can efficiently repair the mismatch and extend the primer to a full product, both the W160A and 5A7A mutants are defective in repairing the G:A mismatch. These results support the conclusion that like F473 in Klenow, W160 plays an essential role in properly aligning the 3'-terminal base in the exonuclease active site.

T7 Phage Can Propagate in the Presence of Exonuclease-Deficient Gp5/Trx.

With both the W160A and 5A7A mutants established as exonuclease-deficient enzymes *in vitro*, we tested the ability of these mutations to complement T7 ϕ with the mindset that accumulation of mutations would reduce viral fitness. T7 ϕ was serially diluted onto top agar containing *E. coli* with either wild-type, W160A, or 5A7A gp5, and the results are shown in Figure 3 and Table 2. To our surprise, T7 ϕ grows well in the presence of both exonuclease-deficient enzymes. The 5A7A mutant displays a one-log reduction in titer while the W160A mutant titer levels are the same as wild-type. Along with the slight reduction in viral titer, we also see evidence of a reduced plaque size for the 5A7A mutant as compared to wild-type.

To further investigate T7 viral fitness in the presence of exonuclease-deficient polymerases, we serially passaged the T7 5 virus ten times at an MOI of 0.001 using *E. coli* cells containing a plasmid with either the wild-type, W160A, or 5A7A gp5 polymerases. Surprisingly, not only can both the W160A and 5A7A mutants complement T7 5 *in vivo*, no decline in viral titer was observed with subsequent passages. To confirm that a loss of gp5 exonuclease activity results in mutation of the T7 5 genome, we performed whole-genome sequencing of three plaque-picked wild-type, W160A, and 5A7A passage ten T7 5 viruses. The original T7 5 viral stock was also sequenced and used as a reference. Results of the whole-genome sequencing experiments are shown in Figure 6 and Supplemental Table S1. Our T7 5 viral stock contains four mutations as compared to reference genome NC_001604 (Supplemental Table S1), and therefore these four mutations are observed in all sequenced samples and were excluded in our analysis. Two of the wild-type gp5 passage ten samples were free of mutations, while one wild-type genome contains a T insertion in a non-coding region of the genome. Two of the W160A genomes sequenced are identical and contain six point mutations in both essential and non-essential genes³⁸. The mutations include one in a non-coding region, two silent mutations, an I45F mutation in the essential gene 3 (endonuclease), a V21I mutation in gene 17.5 (holin), and a V39A mutation in gene 19.3 (unknown function). Along with these six mutations, the third W160A genome contains an additional seventh mutation, a K106Q alteration in gene 5.7 (unknown function).

Although the results of Figure 5 show that both W160A and 5A7A are defective in repairing a mismatch *in vitro*, we see a dramatic difference in the number of mutations in the genomes of these two enzymes. Whereas the W160A genomes contain only very few point mutations, the three 5A7A genomes contain 68, 63, and 68 mutations, respectively, throughout the entire T7 5 genome and include point mutations, deletions, and insertions (Figure 6 and Supplemental Table S1). Along with several mutations in non-essential genes, the essential genes 1 (RNA polymerase), 3 (endonuclease), 6.7 (unknown function), 9 (scaffolding protein), 12 (tail tubular protein B), 14 (internal virion protein B), 15 (internal virion protein C), 16 (internal virion protein D), 17 (tail fiber), and 19 (DNA maturation protein) are all mutated in the 5A7A genomes. As observed in the W160A results, two of the 5A7A genomes are identical and contain the same mutations. What is striking is that the mutations observed in the third 5A7A sample are completely different than those observed in the other two 5A7A genomes. Although only a small sample size was sequenced, the results show that in the 5A7A background the mutations are more frequent and they appear to be random. Additionally, although both W160A and 5A7A are defective at repairing a mismatch in the *in vitro* assay, the whole-genome sequencing results show that the 5A7A mutation is more detrimental to gp5 exonuclease activity.

Solution Structures of Gp2.5:Gp5/Trx Complexes.

To obtain structural data of complexes between gp2.5 single-stranded DNA binding protein and gp5/trx, we mixed the proteins together in the presence of DNA and monitored complex formation using SEC-SAXS-MALS³⁹. Gp2.5 loaded on a dT15 ssDNA is predicted to be dimeric in solution¹⁸, with a mass of ~55 kDa. MALS and SAXS data agree with a gp2.5 dimer in solution (Figure 7 and Table 3), and the slightly larger masses observed for gp2.5 can be explained by flexible C-terminal acidic tails present for both gp2.5 monomers (Figure

7D). The C-terminal tails of gp2.5 had to be removed to obtain a high-resolution crystal structure¹⁸, and our SAXS data show the unfolded character of these tails in solution. Our SEC-SAXS-MALS data also show that gp5/trx alone on DNA is monomeric in solution, and the gp5/trx SAXS scattering profile match well with crystal structures of gp5/trx loaded on DNA (Figures 7C and 7D)¹⁷. When gp2.5 and gp5/trx are mixed in the presence of both a dT15 oligo for gp2.5 and a primer/template for the polymerase, we observe a shift in the elution time of SEC-MALS and SEC-SAXS indicating complex formation (Figure 7A). The best fit to the SAXS curve is obtained with an ensemble of models that include a gp2.5 dimer:gp5/trx monomer complex (64% of the total population) and a gp2.5 dimer bound to two polymerase molecules (36% of the population, Figures 7C and 7D). The complexes are driven by interactions between the flexible C-terminal tails of gp2.5 and the polymerases (Figure 7D). In the 1:1 complex model the gp2.5 tail binds the polymerase near the FBP and appears to be extended, which explains the large R_G . The complex containing two gp5/trx molecules is more compact in shape, and although the low resolution limits accurate placement of the gp2.5 tails on gp5/trx, the compact structural model places the two acidic tails of the gp2.5 dimer such that they could bind both the FBP and TBD of gp5. Gp2.5 is critical for establishing coordinated leading and lagging strand synthesis at the T7 replisome^{40,41}, and our SAXS analysis provide the first view of how a gp2.5 dimer could bring multiple gp5/trx molecules together for coordinated DNA synthesis.

Discussion

Due to its overall simplicity, the DNA replication system of bacteriophage T7 has long served as a model for understanding how multiple proteins communicate to carry out coordinated DNA synthesis⁴². Several recent reports have used a variety of structural methods to provide the first images of how the gp4 and gp5/trx proteins of the T7 replisome assemble during different stages of DNA replication^{10,20,21}. Although it is well established that interactions between the acidic C-terminal tails of both the gp2.5 and gp4 proteins and two basic patches on gp5 are essential for T7 DNA replication, current published structures have been unable to show how these acidic tails engage the polymerase. Our crystal structure provides the first glimpse in atomic detail of how the tail of gp4 binds the FBP of gp5. Additionally, we provide the first solution structure of a complex between gp2.5 and gp5/trx, and the data support that the observed complexes are driven by electrostatic interactions.

We recently determined the structure of the T7 replisome that contained three molecules of gp5/trx bound to a heptameric gp4 ring in the absence of DNA²⁰. In the structure, two of the three gp5/trx molecules contain a patch of F_o-F_c electron density that leads from the last residue of gp4 modeled to the FBP of gp5, and we hypothesized that this density corresponded to the acidic C-terminal tail of gp4. A superposition of the T7 replisome structure on our peptide structure reveals that our peptide sits at the base of where the F_o-F_c electron density ends, and these results show that both structures corroborate the location of the gp5 binding site for the C-terminus of gp4.

In the first report that characterized the FBP of gp5, four positively charged residues (K587, K589, R590, and R591) were all simultaneously mutated to glutamine, and this tetramutant

was found to be defective in strand-displacement DNA synthesis in the presence of wild-type gp4⁸. As a result, this mutant gp5 was lethal and did not complement T7 5 phage. In our structure R590 is shown to play a critical role in recognizing the C-terminal F566 residue of gp4 (Figure 1), and mutation of R590 to either alanine or glutamine is lethal in complementation studies with T7 5 (Figure 3 and Table 2). We did not observe peptide electron density near the other three front basic patch residues (K587, K589, and R591), so the roles of these residues in gp4 tail recognition is unclear. However, given the acidic and flexible nature of the gp4 tail it is possible that the tail could wrap around the front basic patch to contact these residues during replication. Indeed, the pocket where the peptide is bound is surrounded by positive charge (Figure 1C) that could support favorable electrostatic interactions with the acidic residues of gp4.

It is well-established that the acidic tail of gp4 binds to two different basic patches on the gp5/trx polymerase; the FBP at the base of the DNA binding cleft⁸ and to two basic loops in the TBD⁶. The recent cryoEM structure of the T7 replisome on DNA revealed that C-terminal tails of the gp4 hexamer were positioned to interact with both of these gp5 basic patches, although the tails were disordered in the structure¹⁰. A gp4 tail truncated at residue 554 and missing the last twelve C-terminal residues was positioned next to the FBP of gp5 in the cryoEM structure. Also in the cryoEM structure, a gp4 tail truncated at residue 553 lies next to residues K281, K285, and R287 in the TBD, which make up one of the two basic loops that are known to bind the tail of gp4⁶. We see no clear evidence of gp4 peptide density near the TBD in our crystal structure reported here. The TBD of gp5 is flexible, and in our structure part of this region is disordered in each of the four subunits. With the limited resolution of our structure it is unclear whether the peptide did not bind or is disordered in the complex with the TBD.

Given the sequence similarity between the gp2.5 and gp4 acidic tails, we hypothesize that the gp2.5 tail will also bind the FBP in the same manner that gp4 binds gp5 in our crystal structure. Our current knowledge of gp2.5 binding to gp5/trx is that in the absence of DNA the tail of gp2.5 engages the TBD, while in the presence of DNA the gp2.5 tail is required for gp5/trx binding, but the binding site for gp2.5 lies outside the TBD³. We hypothesize that this additional binding site is the FBP of gp5/trx, and this is supported by a structural model of the gp2.5:gp5/trx complex based on our SAXS data (Figure 7D). This structural model of the gp2.5:gp5/trx complex on DNA positions the flexible tails of gp2.5 near both the FBP and TBD of gp5. Additionally, it shows how a dimer of gp2.5 could engage two polymerase molecules in order to coordinate DNA synthesis at the replication fork^{40,41}.

The SAXS modeling also shows interactions between two gp5/trx molecules (Figure 7D) that is reminiscent of the gp5/trx interactions observed in our T7 replisome crystal structure in the absence of DNA²⁰ as well as an earlier cryoEM structure of the replisome on a fork-shaped DNA substrate²¹. All three structures suggest that coordination of DNA polymerase activities within the T7 replisome could be accomplished through direct interactions of gp5/trx subunits. In contrast, the recent cryoEM structure of the gp4:gp5/trx complex captured in coordinate leading and lagging strand synthesis shows all three polymerase molecules spaced apart from each other around the ring-shaped gp4 hexamer¹⁰. Although we have obtained the first solution structure of a gp2.5:gp5/trx complex, the low resolution

limits our ability to define specific protein interactions. Future work that focuses on obtaining structures of the full T7 replisome (gp2.5, gp4, and gp5/trx) on and off DNA will aim to clarify how and when gp2.5 and gp4 compete for binding to the same sites of gp5/trx, as well as how gp2.5 binds multiple gp5/trx molecules in order to coordinate DNA replication.

The presence of dTTP in the exonuclease active site of our structure provides the first details of how DNA binds gp5 for 3'-terminal base removal. Our structure agrees well with the extensive biochemical and structural data of Klenow fragment editing complexes³⁴⁻³⁶, with W160 of gp5 serving an analogous role to Klenow residue F473 in stacking against the 3'-terminal base to properly position the nucleotide for hydrolysis (Figure 4). Primer extension assays show that like the well-characterized gp5 5A7A exonuclease-deficient mutant³⁷, a W160A mutant is defective in repairing a G:A mismatch (Figure 5). Despite the inability to repair a mismatch *in vitro*, both the 5A7A and W160A mutants are able to efficiently complement a T7 phage lacking the gene 5 DNA polymerase (Figure 3 and Table 2). These results suggest that T7 is able to adapt and survive despite the accumulation of point mutations in the genome, and this hypothesis is supported by our whole-genome sequencing studies in which we propagated T7 ϕ in the presence of either wild-type or mutant polymerases. Surprisingly, the three W160A viral populations show very few mutations per genome after ten passages, whereas the 5A7A samples show a much higher mutation rate that includes insertions, deletions, and point mutations. Although both the 5A7A and W160A mutants appear equally defective in repairing a mismatch in our *in vitro* assay, our *in vivo* results show they differ in their ability to fix DNA replication errors, with mutation of two residues (D5, E7) that are needed to coordinate divalent cations being much more detrimental than alteration of W160. The W160A mutant may function more efficiently when editing occurs in the context of additional protein-protein and protein-DNA interactions provided by the replisome. These results highlight the advantage of using multiple approaches to monitor defects in DNA synthesis and repair. By combining the two assays we were able to show a difference in the ability to repair replication mistakes *in vivo*.

We initially hypothesized that in the presence of the 5A7A and W160A polymerase mutants the T7 ϕ titer would drop with each passage due to the accumulation of mutations that would lead to a decrease in viral fitness. However, despite the appearance of mutations in essential genes, the titers for both mutants remained comparable to wild-type in all ten passages, showing that the mutations that did occur were not lethal. Our results are similar to a study in which T7 was exposed to the DNA damaging agent N-methyl-N'-nitro-N-nitrosoguanidine⁴³. Although the T7 titer was expected to drop in the presence of the DNA damaging agent after several passages, the virus adapted to the stress through mutagenesis, with the mutated virus displaying an enhanced fitness relative to the starting viral population. Sequencing of the mutant T7 stocks revealed that most mutations occurred in DNA metabolism genes, with the majority of mutations in the gp5 protein⁴³. Although these authors hypothesized that given the high mutation rate of gp5, the observed fitness gain could be blocked by repeating the same experiment with the T7 ϕ phage lacking the DNA polymerase, subsequent experiments where T7 ϕ was exposed to N-methyl-N'-nitro-N-nitrosoguanidine for multiple generations revealed no loss in viral fitness in the absence of

gp5. This result was unexpected, and the exact mechanism by which T7 adapted to the damaging agent is unknown.

For our studies, we also are uncertain as to how the observed mutations have allowed the virus to adapt to the exonuclease deficiencies. For W160A, the I45F mutation of the essential gene 3 endonuclease that is observed in all three sequenced samples could be a source of viral adaptation, as mutations in gene 3 have been shown previously to rescue gp5 mutations that result in low polymerase processivity⁴⁴. Given that gene 3 is essential for phage survival, mutations observed in gene 3, as well as other essential genes, are predicted to alter activity but not completely inhibit function. The 5A7A results, on the other hand, are more complex, as several essential and non-essential genes throughout the genome have been mutated (Supplemental Table S1), and the I45F mutation observed in the W160A sequences is not conserved in the 5A7A samples. Indeed, there is no gene 3 mutation that is conserved in all three 5A7A sequenced samples. Previous reports have provided evidence that T7 can alter DNA metabolism genes through mutagenesis in order to accommodate alterations in essential genes, but that these alterations must be precisely controlled in order for the virus to adapt⁴⁴⁻⁴⁶. Our results agree with these observations, and additional work is required to understand how the observed mutations allow T7 to adapt to deficiencies in gp5 exonuclease activity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

We thank Gregory Hura for assistance in SEC-SAXS-MALS data collection and analysis.

Funding Sources

This work was supported by startup funds provided by Western Carolina University to J.R.W., a Provost Internal Support Grant provided by Western Carolina University to J.R.W., M.D.G., and B.J.B., and NIH grant GM055390 to T.E. SAXS data collection at SIBYLS is funded through DOE BER Integrated Diffraction Analysis Technologies (IDAT) program and NIGMS grant P30 GM124169-01, ALS-ENABLE.

ABBREVIATIONS

Gp5	DNA polymerase
trx	<i>E. coli</i> thioredoxin
gp4	primase-helicase
ssDNA	single-stranded DNA
gp2.5	single-stranded DNA binding protein
FBP	front basic patch
TBD	thioredoxin binding domain basic patch
ZBD	zinc binding domain

SAXS	small angle X-ray scattering
MOI	multiplicity of infection
SEC	size-exclusion chromatography
MALS	multi-angle light scattering
R_G	radius of gyration
P(r)	pair-distribution function
D_{max}	maximal particle dimension
QELS	quasi-elastic light scattering
MD	molecular dynamics
T7 5	T7 phage that lacks gene 5

REFERENCES

- (1). Lee S-J, and Richardson CC (2011) Choreography of bacteriophage T7 DNA replication. *Curr. Opin. Chem. Biol* 15, 580–586. [PubMed: 21907611]
- (2). Hamdan SM, and Richardson CC (2009) Motors, Switches, and Contacts in the Replisome. *Annu. Rev. Biochem* 78, 205–243. [PubMed: 19298182]
- (3). Ghosh S, Hamdan SM, and Richardson CC (2010) Two Modes of Interaction of the Single-stranded DNA-binding Protein of Bacteriophage T7 with the DNA Polymerase-Thioredoxin Complex. *J. Biol. Chem* 285, 18103–18112. [PubMed: 20375019]
- (4). He Z-G, and Richardson CC (2004) Effect of Single-stranded DNA-binding Proteins on the Helicase and Primase Activities of the Bacteriophage T7 Gene 4 Protein. *J. Biol. Chem* 279, 22190–22197. [PubMed: 15044449]
- (5). Lee J-B, Hite RK, Hamdan SM, Xie SX, Richardson CC, and van Oijen AM (2006) DNA primase acts as a molecular brake in DNA replication. *Nature* 439, 621–624. [PubMed: 16452983]
- (6). Hamdan SM, Johnson DE, Tanner NA, Lee J-B, Qimron U, Tabor S, van Oijen AM, and Richardson CC (2007) Dynamic DNA Helicase-DNA Polymerase Interactions Assure Processive Replication Fork Movement. *Mol. Cell* 27, 539–549. [PubMed: 17707227]
- (7). Lee S-J, Marintcheva B, Hamdan SM, and Richardson CC (2006) The C-terminal Residues of Bacteriophage T7 Gene 4 Helicase-Primase Coordinate Helicase and DNA Polymerase Activities. *J. Biol. Chem* 281, 25841–25849. [PubMed: 16807231]
- (8). Zhang H, Lee S-J, Zhu B, Tran NQ, Tabor S, and Richardson CC (2011) Helicase-DNA polymerase interaction is critical to initiate leading-strand DNA synthesis. *Proc. National Acad. Sci* 108, 9372–9377.
- (9). Kulczyk AW, Akabayov B, Lee S-J, Bostina M, Berkowitz SA, and Richardson CC (2012) An Interaction between DNA Polymerase and Helicase Is Essential for the High Processivity of the Bacteriophage T7 Replisome. *J. Biol. Chem* 287, 39050–39060. [PubMed: 22977246]
- (10). Gao Y, Cui Y, Fox T, Lin S, Wang H, de Val N, Zhou HZ, and Yang W (2019) Structures and operating principles of the replisome. *Science* eaav7003. [PubMed: 30679383]
- (11). Wallen JR, Majka J, and Ellenberger T (2013) Discrete interactions between bacteriophage T7 primase-helicase and DNA polymerase drive the formation of a priming complex containing two copies of DNA polymerase. *Biochemistry* 52, 4026–4036. [PubMed: 23675753]
- (12). Kato M, Ito T, Wagner G, and Ellenberger T (2004) A Molecular Handoff between Bacteriophage T7 DNA Primase and T7 DNA Polymerase Initiates DNA Synthesis. *J. Biol. Chem* 279, 30554–30562. [PubMed: 15133047]

- (13). Kato M, Ito T, Wagner G, Richardson CC, and Ellenberger T (2003) Modular Architecture of the Bacteriophage T7 Primase Couples RNA Primer Synthesis to DNA Synthesis. *Mol. Cell* 11, 1349–1360. [PubMed: 12769857]
- (14). Kato M, Frick DN, Lee J, Tabor S, Richardson CC, and Ellenberger T (2001) A Complex of the Bacteriophage T7 Primase-Helicase and DNA Polymerase Directs Primer Utilization. *J. Biol. Chem* 276, 21809–21820. [PubMed: 11279245]
- (15). Ghosh S, Marintcheva B, Takahashi M, and Richardson CC (2009) C-terminal Phenylalanine of Bacteriophage T7 Single-stranded DNA-binding Protein Is Essential for Strand Displacement Synthesis by T7 DNA Polymerase at a Nick in DNA. *J. Biol. Chem* 284, 30339–30349. [PubMed: 19726688]
- (16). Marintcheva B, Hamdan SM, Lee S-J, and Richardson CC (2006) Essential Residues in the C Terminus of the Bacteriophage T7 Gene 2.5 Single-stranded DNA-binding Protein. *J. Biol. Chem* 281, 25831–25840. [PubMed: 16807232]
- (17). Doublíé S, Tabor S, Long AM, Richardson CC, and Ellenberger T (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* 391, 251–258. [PubMed: 9440688]
- (18). Hollis T, Stattel JM, Walther DS, Richardson CC, and Ellenberger T (2001) Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7. *Proc. National Acad. Sci* 98, 9557–9562.
- (19). Toth EA, Li Y, Sawaya MR, Cheng Y, and Ellenberger T (2003) The Crystal Structure of the Bifunctional Primase-Helicase of Bacteriophage T7. *Mol. Cell* 12, 1113–1123. [PubMed: 14636571]
- (20). Wallen JR, Zhang H, Weis C, Cui W, Foster BM, Ho C, Hammel M, Tainer JA, Gross ML, and Ellenberger T (2017) Hybrid Methods Reveal Multiple Flexibly Linked DNA Polymerases within the Bacteriophage T7 Replisome. *Structure* 25, 157–166. [PubMed: 28052235]
- (21). Kulczyk AW, Moeller A, Meyer P, Sliz P, and Richardson CC (2017) Cryo-EM structure of the replisome reveals multiple interactions coordinating DNA synthesis. *Proc. National Acad. Sci* 114, E1848–E1856.
- (22). Kabsch W (2010) XDS. *Acta Crystallogr. Sect. D Biological Crystallogr* 66, 125–132.
- (23). McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr* 40, 658–674. [PubMed: 19461840]
- (24). Adams PD, Afonine PV, Bunkóczy G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, and Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. Sect D Biological Crystallogr* 66, 213–221.
- (25). Classen S, Hura GL, Holton JM, Rambo RP, Rodic I, McGuire PJ, Dyer K, Hammel M, Meigs G, Frankel KA, and Tainer JA (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr* 46, 1–13. [PubMed: 23396808]
- (26). Dyer KN, Hammel M, Rambo RP, Tsutakawa SE, Rodic I, Classen S, Tainer JA, and Hura GL (2013) Structural Genomics, General Applications. *Methods Mol. Biology Clifton N J* 1091, 245–258.
- (27). Hura GL, Menon AL, Hammel M, Rambo RP, Farris PL, Tsutakawa SE, Jr FE, Classen S, Frankel KA, Hopkins RC, Yang S, Scott JW, Dillard BD, Adams MW, and Tainer JA (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* 6, nmeth.1353.
- (28). Rambo RP, and Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 496, 477–481. [PubMed: 23619693]
- (29). Pelikan M, Hura G, and Hammel M (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys* 28, 174–189.
- (30). Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A (2013) Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys. J* 105, 962–74. [PubMed: 23972848]

- (31). Schneidman-Duhovny D, Hammel M, and Sali A (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38, W540–W544. [PubMed: 20507903]
- (32). Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* 44, W424–W429. [PubMed: 27151198]
- (33). Lu D, and Keck JL (2008) Structural basis of Escherichia coli single-stranded DNA-binding protein stimulation of exonuclease I. *Proc. National Acad. Sci* 105, 9169–9174.
- (34). Brautigam CA, and Steitz TA (1998) Structural principles for the inhibition of the 3′–5′ exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates 11 Edited by R. Huber. *J. Mol. Biol* 277, 363–377. [PubMed: 9514742]
- (35). Beese L, Derbyshire V, and Steitz T (1993) Structure of DNA polymerase I Klenow fragment bound to duplex DNA. *Science* 260, 352–355. [PubMed: 8469987]
- (36). Lam W-C, Thompson EH, Potapova O, Sun X, Joyce CM, and Millar DP (2002) 3′–5′ Exonuclease of Klenow Fragment: Role of Amino Acid Residues within the Single-Stranded DNA Binding Region in Exonucleolysis and Duplex DNA Melting. *Biochemistry* 41, 3943–3951. [PubMed: 11900537]
- (37). Patel SS, Wong I, and Johnson KA (1991) Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry* 30, 511–525. [PubMed: 1846298]
- (38). Qimron U, Tabor S, and Richardson CC (2010) New Details about Bacteriophage T7-Host Interactions. *Microbe Mag.* 5, 117–122.
- (39). Putnam CD, Hammel M, Hura GL, and Tainer JA (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev. Biophys* 40, 191–285. [PubMed: 18078545]
- (40). Lee J, Chastain PD, Kusakabe T, Griffith JD, and Richardson CC (1998) Coordinated Leading and Lagging Strand DNA Synthesis on a Minicircular Template. *Mol. Cell* 1, 1001–1010. [PubMed: 9651583]
- (41). Hamdan SM, Loparo JJ, Takahashi M, Richardson CC, and van Oijen AM (2008) Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. *Nature* 457, 336–339. [PubMed: 19029884]
- (42). Kulczyk AW, and Richardson CC (2016) The Replication System of Bacteriophage T7. *Enzymes* 39, 89–136. [PubMed: 27241928]
- (43). Springman R, Keller T, Molineux I, and Bull J (2010) Evolution at a High Imposed Mutation Rate: Adaptation Obscures the Load in Phage T7. *Genetics* 184, 221–232. [PubMed: 19858285]
- (44). Lee S-J, Chowdhury K, Tabor S, and Richardson CC (2009) Rescue of Bacteriophage T7 DNA Polymerase of Low Processivity by Suppressor Mutations Affecting Gene 3 Endonuclease. *J. Virol* 83, 8418–8427. [PubMed: 19535436]
- (45). Bull J, and Molineux I (2008) Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity* 100, 453–463. [PubMed: 18212807]
- (46). Sadowski P (1974) Suppression of a mutation in gene 3 of bacteriophage T7 (T7 endonuclease I) by mutations in phage and host polynucleotide ligase. *J. Virol* 13, 226–229. [PubMed: 4590018]

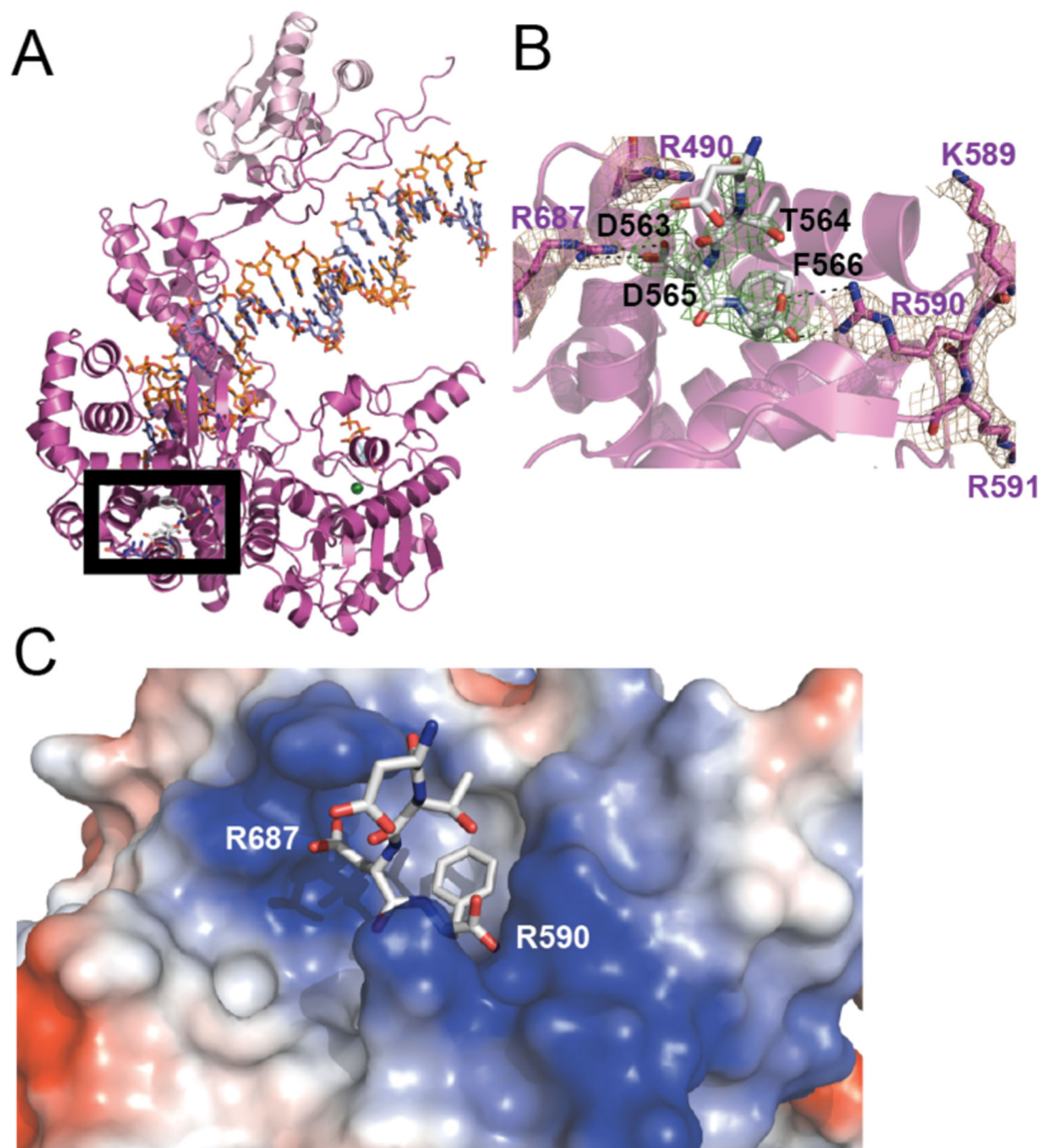


Figure 1:

A) Structure of gp5/Trx bound to a primer/template DNA and a peptide mimicking the C-terminal tail of gp4. The peptide binds at the front basic patch of the polymerase, which is indicated with a black box. Color coding is as follows: polymerase domain: magenta, thioredoxin: light pink, primer and template DNAs: slate blue and orange, respectively, gp4 peptide: white, and magnesium: green spheres. **B)** Zoomed-in view of the gp4 peptide binding site that is indicated by the black box shown in figure A). Gp5 residues are colored magenta, while gp4 peptide residues are colored white and labeled in black. *2Fo-Fc* density, contoured at 1 sigma, is shown for gp5 residues and is colored wheat. *Fo-Fc* density, contoured at 2.5 sigma, is shown for the gp4 peptide and is colored green. **C)** Electrostatic surface potential map for gp5 shows that the C-terminal F566 of gp4 binds at the base of a

hydrophobic cleft that is flanked by regions of positive charge. R590 and R687 of gp5 are labeled for clarity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

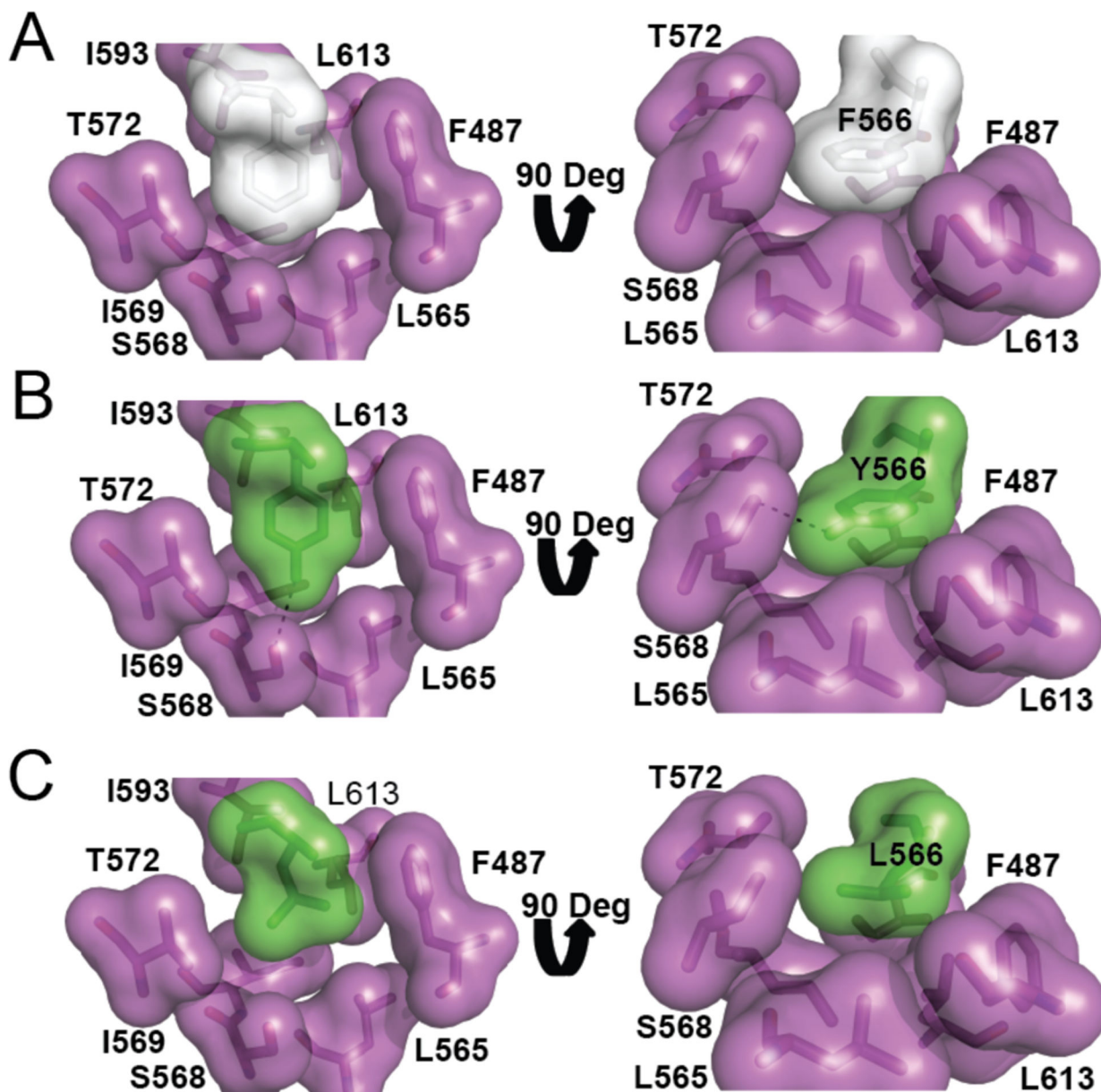


Figure 2:

The C-terminal F566 of gp4 tightly packs in a hydrophobic cleft on the surface of gp5. **A)** Surface representation showing the intimate fit between F566 of gp4 and multiple residues of gp5. **B)** A model of a F566Y mutation shows that tyrosine can be accommodated by the hydrophobic cleft. The -OH of the tyrosine can form a hydrogen bond with S568. **C)** Unlike the F566Y mutation, a model of an F566L mutation shows a less intimate fit between leucine and the gp5 pocket.

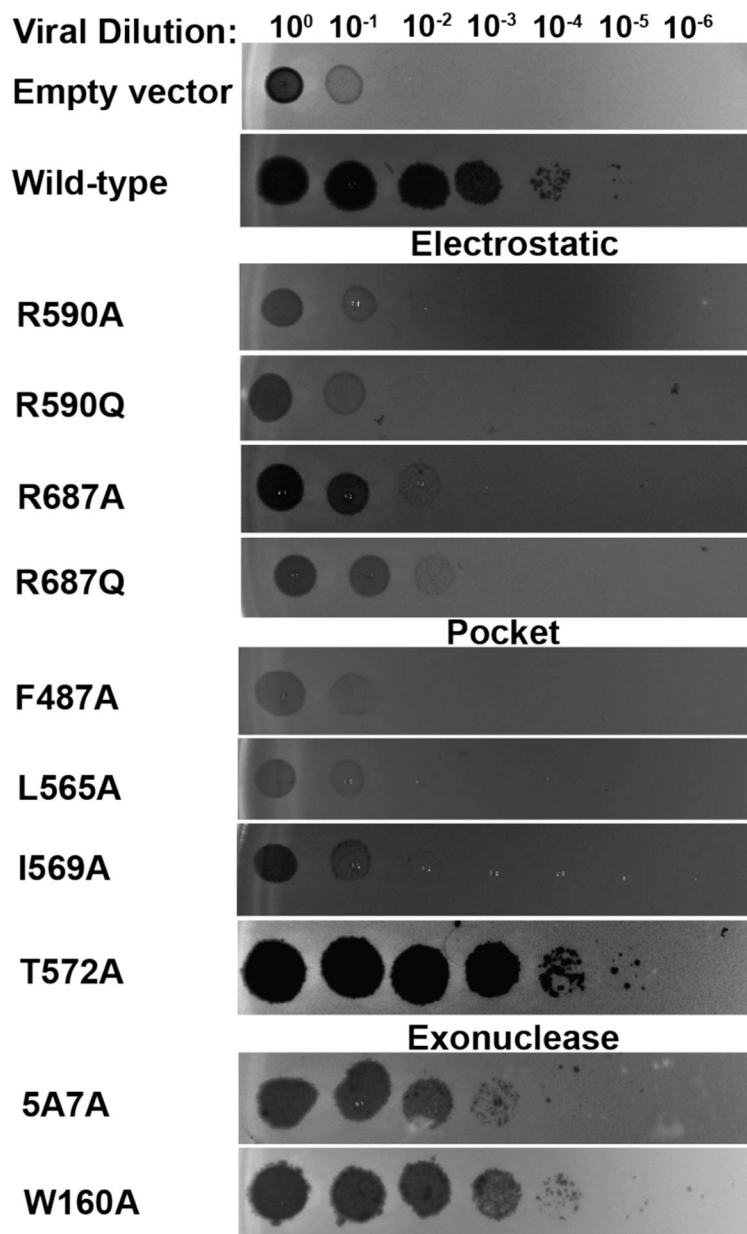


Figure 3:
Gp5 residues observed to bind the gp4 peptide in the front basic patch or to dTTP located in the exonuclease active site were mutated and tested for their ability to complement a T7 phage that lacks gene 5. Stock virus was serially diluted, then spotted on a lawn of *E. coli* harboring a plasmid with either wild-type or mutant gp5. The empty vector lane shows T7 growth in the absence of gp5.

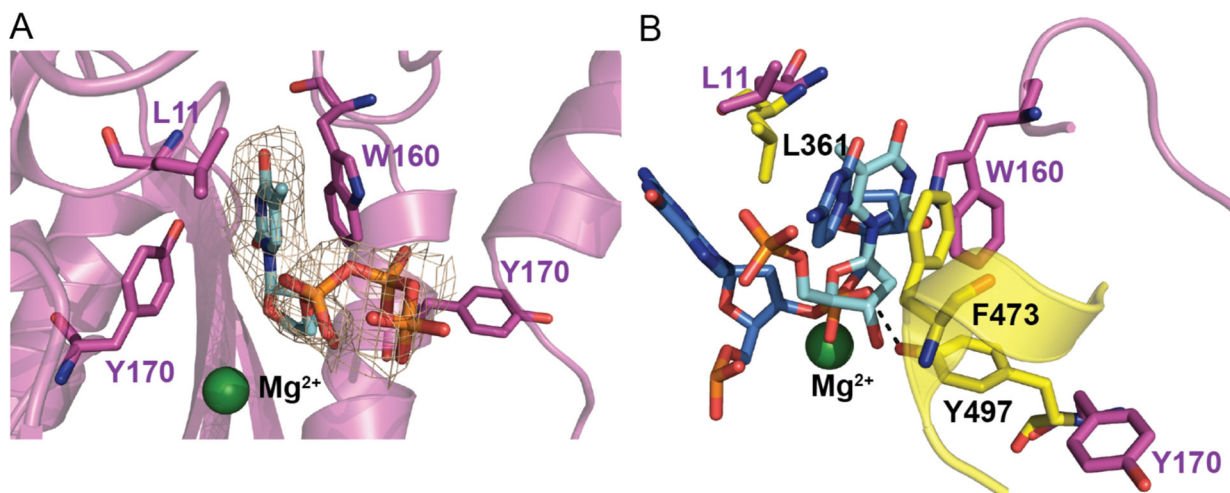
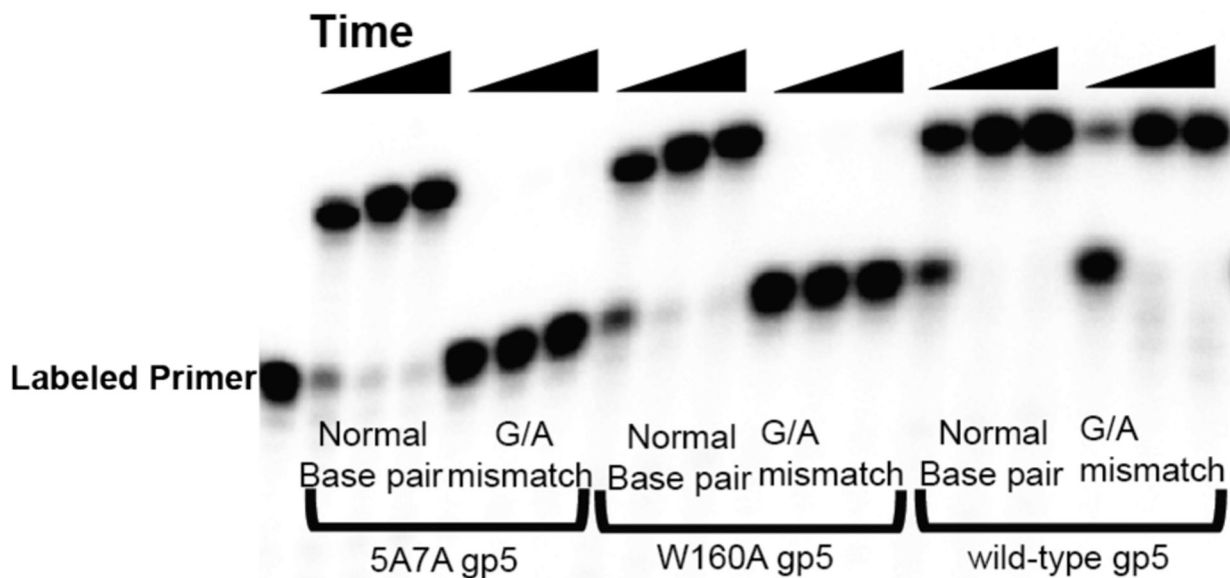


Figure 4:

dTTP bound to the exonuclease active site of gp5. **A)** dTTP (colored cyan), which was added to the crystallization setup at high concentrations to lock the polymerase on DNA in the polymerase active site, was observed bound in the exonuclease active site of gp5. $2F_o-F_c$ density, contoured at 1 sigma, is shown for dTTP and is colored wheat. The dTTP base is observed wedged between the aromatic Y170 and W160 side chains. **B)** Superposition of the T7 exonuclease active site with the Klenow fragment editing complex (pdbID 1KLN). T7 residues are colored magenta, with the bound dTTP colored cyan. Klenow residues are colored yellow, and the DNA from the Klenow complex is colored dark blue. Only the 3'-terminal and the penultimate nucleotides of DNA from the Klenow structure are shown for clarity. The location of the dTTP mimics the binding site of the 3'-terminal base in the Klenow editing complexes, and W160 of T7 functions similar to F473 of Klenow in stacking with the 3'-terminal base.



Normal Base Pair Sequence:

3' -ACCGTCCGTCCACCAGAACGGCCAGATCCCTAGCG
 5' -TGGCAGGCAGGTGGTCTTGCCGGTC

Mismatch Sequence:

3' -ACCGTCCGTCCACCAGAACGGCCAGATCCCTAGCG
 5' -TGGCAGGCAGGTGGTCTTGCCGGTA

Figure 5:

Gp5 residue W160 is critical for gp5 exonuclease activity. Wild-type gp5 and both the 5A7A and W160A gp5 mutants were tested for their ability to extend a primer on both a normal Watson-Crick base pair and a G:A terminal mismatch substrate. Polymerase was mixed with substrate, and time points at 1, 10, and 60 minutes were collected and analyzed on a 14% denaturing Urea PAGE gel. While all three enzymes can efficiently extend a primer on a normal base-paired substrate, both the 5A7A and W160A mutants fail to extend a mismatched primer.

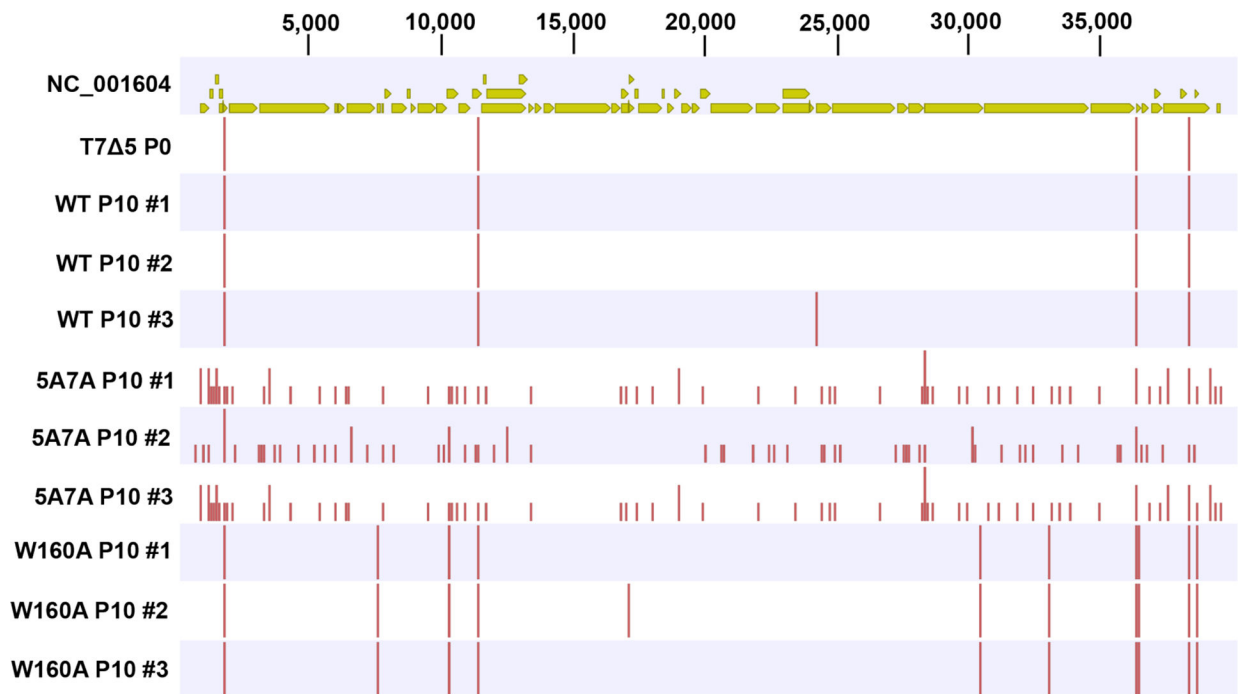


Figure 6:

Whole-genome sequencing analysis of T7 5 in the presence of wild-type and exonuclease-deficient gp5 polymerases. Shown at the top is the CDS annotation of reference T7 genome NC_001604, with numerical genome position indicated above. Sequencing results show that our stock T7 5 virus (T7 5 P0) has four mutations as compared to the reference genome, with the location of these mutations designated with vertical bars. The height of the vertical bars corresponds to the number of mutations located within a 100 bp range of the genome. Analysis of the three wild-type passage 10 samples show that wild-type samples 1 and 2 have no additional mutations relative to the T7 5 stock, while wild-type sample 3 contains an insertion in a non-coding region of the genome. Whereas the W160A passage 10 samples contain a small number of additional point mutations as compared to wild-type, the 5A7A samples contain a variety of insertions, deletions, and point mutations throughout the entire genome. A description of the location and type of all mutations highlighted here are provided in the supplement.

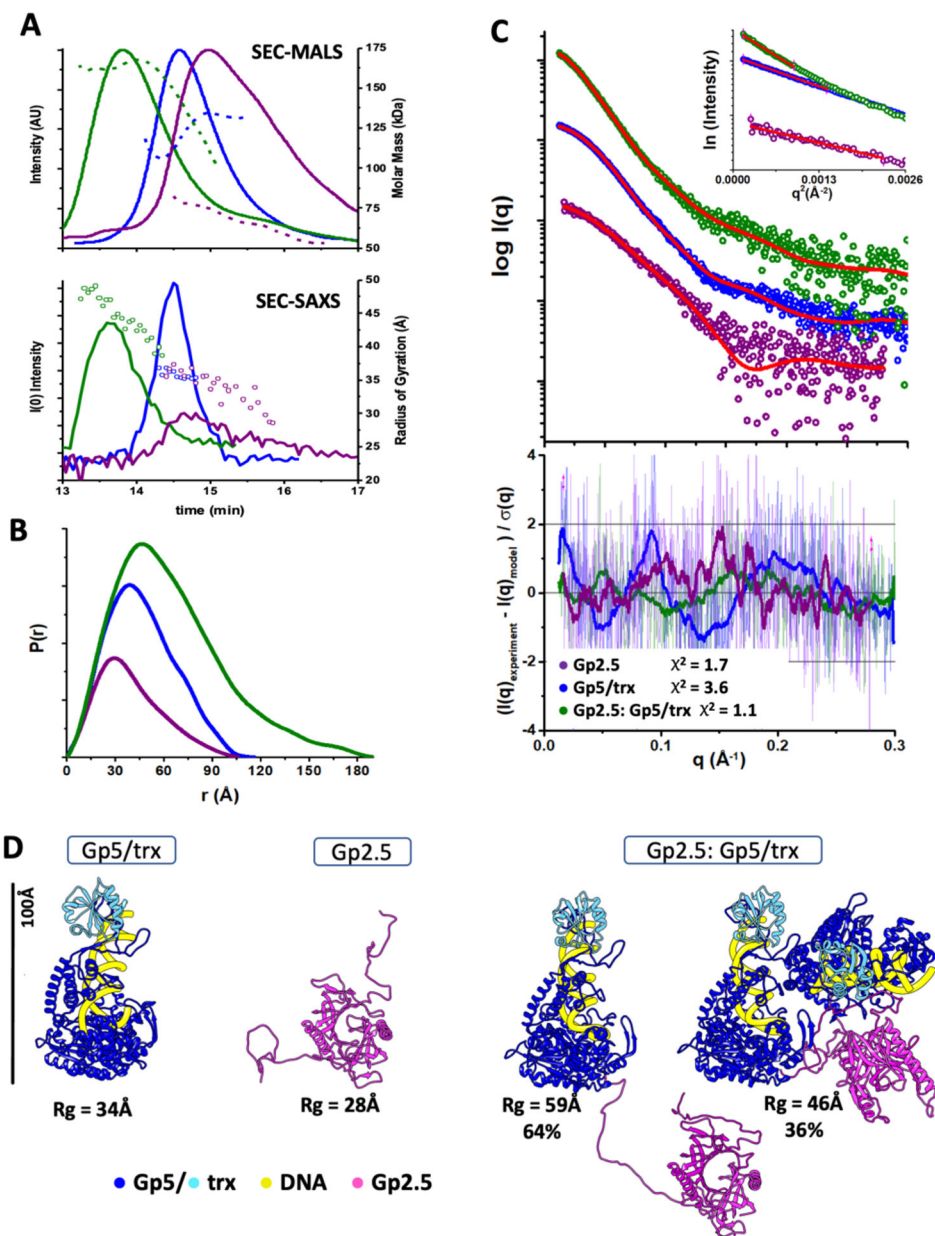


Figure 7: SAXS analyses of gp5/trx, gp2.5, and complexes with DNA. **A)** Top: SEC-MALS chromatographs for gp2.5 with DNA (Purple), gp5/trx with DNA (Blue), and a complex of gp5/trx and gp2.5 (Green). Solid lines represent the light scattering signal (Rayleigh ratio in arbitrary units), while dashed lines represent molecular mass versus elution time. Bottom: SEC-SAXS profiles showing I(0) (lines) and R_G (symbols) values are shown for each collected frame across the SEC peak colored according to the top panel of **A**). **B)** P(r) functions calculated for the experimental data shown in panel **C**). The area of the P(r) is normalized relative to the MW estimated by SAXS and is listed in Table 3. **C)** Experimental and theoretical SAXS curves (red) for corresponding models shown in panel **D**) are provided together with the fit-residuals. Inset, Guinier plots for the SAXS curves. **D)** Best fit

models for gp5/trx-DNA, gp2.5, and ensemble models of gp5/trx -DNA-gp2.5 complexes are shown together with model R_G values and weights for the ensemble models.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

X-Ray Diffraction Data Collection and Refinement Statistics for the Gp5/Trx:Gp4 Peptide Structure

<u>Data collection</u>	
Space group	P1
Cell dimensions	
<i>a, b, c</i> (Å)	100.72, 102.70, 148.89
α, β, γ (°)	91.36, 96.83, 113.11
Resolution (Å)	3.00
Complexes per ASU	4
Reflections	392,815
Unique Reflections	105,150
CC1/2	0.974 (0.344)
CC*	0.993 (0.715)
<i>R</i> _{merge}	22.2 (127.8)
<i>I</i> σ <i>I</i>	7.64(1.00)
Completeness (%)	98.8 (97.8)
Redundancy	3.7 (3.8)
<u>Refinement</u>	
Resolution (Å)	47.44–3.00
No. reflections	107,345
<i>R</i> _{work} / <i>R</i> _{free}	21.4(36.5)/24.8(39.3)
No. atoms	28,831
Protein	28,556
Ligand/ion	244
Water	31
<i>B</i> -factors	
Protein	61.75
R.M.S. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.17

* Values in parentheses are for highest-resolution shell.

Table 2:

Viral titers calculated from spot tests shown in Figure 3.

Mutation	Viral (pfu/mL) *	Titer
Empty Vector	Lethal	
Wild-type	5.5×10^8	
R590A	Lethal	
R590Q	Lethal	
R687A	7.2×10^6 *	
R687Q	1.9×10^7 *	
F487A	6.0×10^4 *	
L565A	Lethal	
I569A	1.3×10^4 *	
T572A	2.8×10^8	
5A7A	7.5×10^7 *	
W160A	5.9×10^8	

* Indicates small plaques as compared to wild-type

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SEC-MALS and SAXS parameters for gp2.5 on DNA, gp5/trx on DNA, and the gp2.5 gp5/trx complex.

Table 3:

Sample	Theoretical MW(kDa)	SEC-MALS MW(kDa)	SAXS MW(kDa)	R_g (Å)	D_{max} (Å)	P_x
Gp2.5	54.6	59–78	48	28	~100	3.9
Gp5/trx	107.9	111–132	97	36	~110	4.0
Gp5/trx / Gp2.5	162.5	116–162	170	49	~180	3.5