Mutational and Functional Analysis of a Segment of the Sigma Family Bacteriophage T4 Late Promoter Recognition Protein gp55*

Received for publication, November 10, 2002, and in revised form, December 18, 2002
Published, JBC Papers in Press, December 20, 2002, DOI 10.1074/jbc.M211447200

Kevin Wong†, George A. Kassavetis‡, Jean-Paul Léonetti¶, and E. Peter Geiduschek∥
From the Division of Biological Sciences and Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92037-0634

Bacteriophage T4 late promoters, which consist of a simple 8-base pair TATA box, are recognized by the gene 55 protein (gp55), a small, highly diverged member of the σ family proteins that replaces σ70 during the final phase of the T4 multiplication cycle. A 16-amino acid segment of gp55 that is proposed to be homologous to the σ70 region 2.2 has been subjected to alanine scanning and other mutagenesis. The corresponding proteins have been examined in vitro for binding to Escherichia coli RNA polymerase core enzyme and for the ability to generate accurately initiating basal as well as sliding clamp-activated T4 late transcription. Mutations in the amino acid 68–83 segment of gp55 generate a wide range of effects on these functions. The changes are interpreted in terms of the multiple steps of involvement of gp55, like other σ proteins, in transcription. Effects of mutations on RNA polymerase core binding are consistent with the previously proposed homology of amino acids 68–82 of gp55 with σ70 region 2.2 and the recently determined structures of the Thermus thermophilus and Thermus aquaticus σ70-RNA polymerase holoenzymes.

All multisubunit RNA polymerases require proteins that are specialized for promoter identification and for initiation of transcription. The eukaryotic and archaeal RNA polymerases employ extrinsic proteins whose assemblies on DNA mark the promoter and recruit RNA polymerase, whereas the bacterial RNA polymerases use their tightly bound σ subunits for a strictly comparable purpose. A single major σ protein, σ70 in Escherichia coli, is used to recognize most of a genome’s promoters, but even the smallest bacterial genomes encode additional σ proteins recognizing distinct promoter sequences, and some of the larger genomes encode dazzling numbers of accessory σ factors (1). All σ proteins are multivalent and multifunctional; they bind to RNA polymerase core enzymes and to DNA, almost invariably recognizing two separate DNA sites. Many of the proteins, possibly all of them, also are targets of accessory ligands that regulate their activity and/or cellular compartmentation. The segmented pattern of amino acid sequence conservation among σ proteins (2, 3), defining homology segments 1.1, 1.2, 2.1–2.5, 3.1, 3.2, 4.1, and 4.2, is associated with common functions of these proteins.

Much of the information about the function of σ proteins and much of the recent key information about mechanism of action comes from the analysis of E. coli σ70. σ70 is a four-domain protein; each structural domain occupies a separate site on the surface of the RNA polymerase core enzyme (4–9). Homology segments 2 and 4, which are located in separate structural domains, also bind specifically to separate DNA sites (the −10 and −35 promoter elements) centered −1 and 3.2 turns upstream of the transcriptional start site. The core of the polymerase holoenzyme provides the scaffold that constrains the appropriate spacing of homology segments 2 and 4 for promoter recognition. Polymerase core binding also changes the internal structure of σ70, disrupting an interaction between homology segments 1.1 and 4 that blocks DNA binding by homology segment 4 (10), separating segments 2 and 4 (4), and allowing site-specific binding to the melted nontranscribed strand as well as double-stranded DNA of the −10 promoter element (11–13). The structures of E. coli σ70 structure domain 2 (extending from homology segments 1.2 to 2.4) and, very recently, of three structure domains of Thermus aquaticus (Taq) σ70 (comprising homology segments 1.2–2.4, 3.0–3.1, and 4.1–4.2, respectively) have been determined (5, 14). The locations of E. coli σ70 homology segments 1.1 (comprising the fourth structure domain of σ), 2, 3.1, 3.2, and 4 have been modeled onto the structure of RNA polymerase core in the holoenzyme and in open promoter complexes on the basis of an extensive survey of spatial separations in solution, determined by fluorescence resonance energy transfer (6). The culmination of all this effort has been the determination of structure of σ70 holoenzymes from Taq and Thermus thermophilus and of a Taq holoenzyme complex with fork junction DNA representing an open promoter complex (8, 9, 15).

The experiments that are presented here deal with the bacteriophage T4 gene 55 protein, gp55, a highly diverged, small (185 amino acids) σ family protein. gp55 confers the ability to recognize T4 late promoters, which consist of an 8-base pair TATA box (TATAAATA in the nontranscribed strand) centered one helical turn upstream of the transcriptional start site. T4 late promoters entirely lack the −35 binding site that is characteristic of the other σ family promoters, and gp55 has no segment homologous to σ region 4, which contains the corresponding −35 site-binding domain (homology segment 4.2). T4 late genes are activated by an apparently unique mechanism that connects their transcription to concurrent DNA replication through the action of the sliding clamp (gp45) of the phage DNA polymerase holoenzyme. From the point of view of mechanism, the salient feature of gp45 is that it activates T4 late
transcription in a topologically but not physically DNA-bound state. Consequently, gp45 does not play a direct role in marking its conjugate T4 late promoters. Like its cellular homologs (16, 17), the T-even phage family sliding clamp is ring-shaped (18, 19). Loading the sliding clamp activator onto DNA at primer-template junctions or single strand nicks and gaps in DNA is done by its clamp loader, the T4 gene 44/62 protein complex (gp44/62). The gp45 sliding clamp interacts with a C-terminal hydrophobic-acidic epitope of gp55; gp33, the T4-encoded, RNA polymerase core-bound co-activator/core transcription factor T4 DNA polymerase, have similar C-terminal epitopes (20, 21). The C termini of gp55 and gp33 are both required for sliding clamp activation of T4 late transcription (21–24).

The recognizable but weak homology of gp55 with the σ family proteins is confined to segment 2 (amino acids 381–451 in E. coli σ70). The corresponding amino acid 45–115 segment of gp55 includes those parts of the protein that are most strongly protected from peptide bond cleavage by hydroxyl radical when bound to RNA polymerase core and that are supposed to be homologous to σ70 segments 2.1 and 2.2 (25). This σ70 region is known to be involved in binding to the RNA polymerase core (12, 13, 26–29).

We have subjected the region of gp55 that is supposed to be homologous to σ70 segment 2.2 (amino acids 68–83) to alamine scanning and other mutagenesis. The corresponding mutant proteins have been examined in vitro for ability to bind to E. coli RNA polymerase core and also for function in basal as well as sliding clamp-activated T4 late transcription. The very wide range of phenotypes that is generated by mutations in this short segment is interpreted in terms of the multiple roles that σ plays in promoter recognition, initiation of transcription, and transcriptional activation.

**EXPERIMENTAL PROCEDURES**

**Proteins—**Gene 55 has been modified for this work by insertion of an N-terminal kinase tag and a C-terminal His tag into the wild type gene in expression vector pET21b. The corresponding protein is referred to throughout as “wild type” gp55. Oligonucleotide-directed mutagenesis was used to generate mutations at amino acids 68–83 that are specified in Fig. 1. (The mutant protein with amino acid 68 changed from lysine to alanine is referred to as gp55-K68A, for example.) The wild type and mutant genes 55 were overexpressed in E. coli BL21(DE3) grown at 37°C to an absorbance at 600 nm of 0.6–0.8 and purified by nickel-nitriol triacetic acid agarose. The insoluble gp55 proteins were recovered from inclusion bodies that were washed twice in buffer containing 50 mM NaHepes, pH 7.8, 5 mM β-mercaptoethanol, 1 mM NaCl, and 0.2% (v/v) Triton X-100 before solubilization in buffer containing 40 mM Tris-HCl, pH 8, 6 mM guanidine HCl, 10 mM β-mercaptoethanol, and 10% (v/v) glycerol. The extracted proteins were applied to nickel-nitriotriacetic acid-agarose, and washed on the column with solubilization buffer. gp55 was eluted in solubilization buffer with 200 mM imidazole, stored at −20°C at concentrations of 150–700 μM, and diluted from this storage buffer directly into transcription buffer (specified below) before each transcription or RNA polymerase binding experiment (introducing 3–18 μM guanidine into the reaction medium, as specified below).

**Untagged E. coli RNA polymerase core, RNA polymerase core with a His6-tagged β subunit, T4 gp45 (sliding clamp), gp44/62 complex (clamp loader), gp32 (single-stranded DNA-binding protein), and gp53 (late transcription activator) were purified or referenced elsewhere (23, 30) and stored at −20°C.**

**Restriction endonuclease, exoIII, protein A-Coupled Sepharose beads, and protein kinase A were purchased.**

**Co-immunoprecipitation of gp55 with RNA Polymerase Core—**Wild type gp55 was phosphorylated in its N-terminal tag with bovine heart protein kinase A (5 units) in buffer containing 20 mM Tris-HCl, pH 7.8, 100 mM MgCl2, and 0.01% (v/v) Tween 20. Monoclonal antibody RA2 to the α subunit of E. coli RNA polymerase core, a generous gift from R. R. Burgess, was conjugated to protein A-Sepharose beads following a standard method (31). The gp55 probe was overexpressed in E. coli DH5α, was used to generate mutations at amino acid 45 (v/v) Tween 20, and 66 μg/ml bovine serum albumin (BSA) for 30 min in the presence or absence of 3 or 6 pmol of unlabeled mutant competitor gp55 in 5-cm siliconized Eppendorf tubes that had been preblocked at 0°C overnight with IP buffer, 20 μl of a 25% slurry of anti-α-conjugated protein A-Sepharose beads suspended in buffer IP was added to each tube and allowed to equilibrate at 4°C for 1 h, with rocking. Supernatant fluid was removed from pelleted beads, and the latter were washed three times with 600-μl aliquots of IP buffer. The pelleted beads were aliquoted for scintillation counting with 200 μl of IP buffer.

**Single-round Transcription—**DNA for single-round transcription was derived from pDH310, which contains a transcription unit defined by the T4 gene 23 late promoter and the phage T7 early transcription terminator and yields an ~420-nucleotide transcript. Supercorrelated or blunt end linear DNA was used for basal transcription. For sliding clamp activated transcription, pDH310 was first linearized with EcoRI and endonuclease and reacted with exoIII to 1 μl guanidine HCl to create 5′-overhanging ends. The overhanging end upstream of the T4 late promoter was removed with SmaI endonuclease, and the resulting transcription template was purified as described (32).

**Single-round transcription at 25°C was performed as described (23).** For basal transcription, 36 pmol of C-terminally His-tagged wild type or mutant gp55 was incubated on ice with 6 μl of RNA polymerase in 40 μl of transcription buffer (240 mM potassium acetate, 33 mM KHepes, pH 7.8, 10 mM magnesium acetate, 1 mM dithiothreitol, 150 μg/ml bovine serum albumin), transferred to 25°C, and added to 0.75 pmol of linear pDH310 DNA in 40 μl of transcription buffer (pre-equilibrated at 25°C). At specified subsequent times, transcription was initiated by transferring a 1 μl aliquot to 40 μl of transcription buffer (37°C to 4°C to analyze), 1 mM GTP, 1 mM ATP, 0.1 mM CTP, 0.1 mM [32P]UTP (4000 cpm/pmol), 25 μg/ml rifampicin in transcription buffer). Transcription was allowed to proceed for 8 min and halted by adding 150 μl of stop buffer (20 mM Na3EDTA, 40 mM Tris-HCl, pH 8, 250 mM NaCl, 0.4% (v/v) SDS, 250 μg/ml yeast RNA) containing a small quantity of labeled DNA fragment as a sample recovery marker. The transcripts were purified, resolved, and quantified as described (25). For sliding clamp-activated transcription, the DNA mix contained, in addition to 0.6 pmol of DNA, 6.1 μg of gp32 and 3 μl of ATP in 40 μl of transcription buffer, and the protein mixture contained, in addition to RNA polymerase core and gp55, 30 pmol of gp35, 40 pmol of gp44/62 complex, and 22 pmol of unlabeled mutant 32P-gp55. Background was determined with samples lacking core RNA polymerase. Each experiment was also accompanied by control samples to determine competition by 3, 6, or 12 pmol of wild type unlabeled gp55 and to examine the effect of 6 pmol of each unlabeled mutant gp55 on competition by 6 pmol of unlabeled wild type gp55.

**KmO4 Footprinting—**DNA (hp – 150 to +150 relative to the transcriptional start site in pDH310), 32P End-labeled in the nontranscribed strand, was generated by PCR and purified by non-denaturing gel electrophoresis, essentially as described (33). The samples, in 10 μl of transcription buffer containing 10 fmol of probe DNA, 50 ng of poly[d(G- 

**dC):poly[d(G-dC)], 30 fmol of** E. coli RNA polymerase core, and 1.2 pmol of wild type or mutant gp55, were incubated for 20 min at 25°C. KMnO4 (final concentration, 12 mM) was added to each reaction mixture followed by 1 min rotation, 3 μl of buffer (20 μl Na3EDTA, 40 mM Tris-HCl, pH 8, 350 mM NaCl, 0.5% (v/v) SDS, 200 μl of 10 μg/ml RNAase, and 200 μl of 200 mM β-mercaptoethanol). The reaction mixtures were extracted with phenol-chloroform. DNA was precipitated with ethanol, cleaved with piperidine, and analyzed on 6% polyacrylamide gel containing 7 M urea, as described (33). DNA cleavage at T-4 of the nontranscribed strand was quantified by phosphorimaging scanning and normalized to cleavage generated by wild type gp55 holomerase. A control for background DNA cleavage in a sample lacking RNA polymerase core.

**Aggregation of Mutant gp55—**Aggregation was assessed by gel exclusion chromatography on Superose 12 essentially as described (23). Briefly, gp55 was diluted (400–600-fold) out of its storage buffer into column buffer (20 mM NaHepes, pH 7.8, 7 mM MgCl2, 500 mM NaCl, 0.2% (v/v) Tween 20, 0.4% (v/v) glycerol, and 0.01% (v/v) Tween 20. 14 m M MgCl2, 10 mM β-mercaptoethanol, 5 μM phenylmethylsulfonyl fluoride, and 0.05 μg/ml each of leupeptin and pepstatin) containing cytochrome c (12.5 kDa), ovalbumin (45 kDa), bovine serum albumin (68 kDa), and ferritin (450 kDa) as size markers. A 200-μl volume (containing 100 μmol of
protein) was loaded onto a Superose 12 column equilibrated and developed in column buffer. Column fractions were analyzed by SDS-PAGE with silver staining. Monomeric gp55 elutes between cytochrome c and ovalbumin.

**Modeling—**gp55 amino acids 42–119 were aligned to Tth αw amino acids 184–290 with a single gap opposite gp55 residue Trp72. gp55 residues 42–66 and 68–119 were threaded into the Tth holoenzyme structure (11W7) by sequential mutagenesis and rotamer optimization, followed by energy minimization using the GROMOS96 forcefield module of Swiss-Pdb Viewer (34). gp55 Trp72, which was not included in the modeling, would lie in the loop between the region 2.1 and region 2.2 α helices. This threaded structure was used for examining the effects of individual gp55 mutations following the same process as above. The Nonecovalent Bond Finder Module of Protein Explorer and visual inspection was used to assess the effects of gp55 mutation.

**RESULTS**

An alanine scan mutagenesis of amino acids 68–83, the gp55 segment previously aligned with αw segment 2.2, was carried out; alanine at amino acid 78 was replaced with Gly, the putatively corresponding residue at αw amino acid 411. Charge-reversing mutations at amino acids 77 (Glu → Arg) and 81 (Lys → Asp) were also introduced (Fig. 1). The corresponding C-terminally His₆-tagged proteins with an additional N-terminal kinase tag (RRASV) and purified the corresponding wild type protein (Procedures).

**Binding to RNA Polymerase Core—**Each of these proteins was assessed for its ability to bind to E. coli RNA polymerase core, using a co-immunoprecipitation/competition assay devised by Sharp and co-workers (28) to screen αw mutations for the same function. Wild type gp55 was 32P-labeled in its N-terminal kinase tag and used at a concentration sufficient to nearly saturate the RNA polymerase core. The ability of a 2- and 4-fold excess of unlabeled (unphosphorylated) wild type and mutant gp55 to compete with this binding was compared by incubating labeled and unlabeled gp55 with core and then separating core-bound gp55 on protein A-Sepharose beads coated with monoclonal antibody directed against the RNA polymerase α subunit (the generous gift of R. R. Burgess). The quantity of 32P-labeled gp55 co-immunoprecipitating with core decreased by ~70% in the presence of a 2-fold excess of unlabeled gp55 (Fig. 2). Mutant proteins considered to be defective in core binding were those that diminished binding of 32P-gp55 by less than 50% at 2-fold excess. As a control, mutant proteins were also tested for interference with competition by unlabeled wild type gp55 at high excess (a 4-fold excess each of the wild type and mutant proteins, compared with an 8-fold excess of wild type protein).

The results of the analysis are presented in Fig. 2 and in Table I, where degrees of defective binding are indicated by a + to +++ scale, compared with +++ for the wild type gp55. The mutant proteins found to be most defective in core binding were gp55-E70A, gp55-D74A, gp55-G82A, and gp55-L83A, but gp55-G82A and gp55-L83A also showed evidence of interference with competition by unlabeled wild type gp55 at high concentration. Such interference might be due to protein aggregation, perhaps resulting from misfolding under these conditions (see “Experimental Procedures”). Accordingly, these most core-binding-defective proteins (with the exception of the transcriptionally highly active gp55-L83A), as well as gp55-M71A and gp55-I72K, were also tested for aggregation by size exclusion on Superose 12. Only gp55-G82A showed substantial aggregation, with ~50% of protein eluting as multimeric complexes (data not shown). Despite an apparent tendency to aggregate at the relatively high concentrations used for column chromatography, the severe defect in core binding of gp55-G82A is unlikely to be entirely due to aggregation.

**Basal Transcription—**Each of these proteins was examined for basal transcription of linear duplex DNA containing the T4 late transcription unit in the previously constructed and extensively analyzed plasmid pDH310 (35). The T4 late promoter in pDH310 (from T4 gene 23, which encodes the major phage head protein) and the transcriptional terminator from phage T7 gene 1 define a transcription unit that yields an ~420-nucleotide RNA product. Promoter complexes were formed at 25 °C for the times indicated in Fig. 3 (A and B) and then allowed to execute a single round of transcription as specified under “Experimental Procedures.” Most of these mutant proteins were quantitatively deficient in basal transcription, relative to wild type gp55, but some, including gp55-Q69A and gp55-I76A, were comparably active with the corresponding wild type protein. All of these proteins are compared quantitatively (for activity after 20 min of promoter opening) in Fig. 4 and Table I.

It was especially interesting to see that gp55-E77A and gp55-L83A were respectively 3- and 2-fold more active transcriptionally than wild type gp55 (Table I and Fig. 4), because they form transcriptionally competent promoter complexes more rapidly (Fig. 3B and data not shown). The existence of this surprising but consistently observed advantage was also confirmed in experiments with gp55-E77A to examine the pseudo-first order rate constant of formation of DNA competitor-resistant nitro...
cellulose membrane-retained polymerase-promoter complexes. In the example that is shown in Fig. 5, the gp55-E77A holoenzyme formed these complexes 3.7 times more rapidly at 37 °C than did the wild type gp55 holoenzyme.

Basal transcription of negatively supercoiled pDH310 DNA was also surveyed in a standard assay, in which a single round of transcription at 25 °C in standard reaction buffer was quantified after allowing 10 min for formation of open promoter complexes (Fig. 4, open columns, Table I, and “Experimental Procedures”). DNA supercoiling was found to diminish the transcriptional defects of gp55 variant proteins. Restoration of activity was only partial for the more deficient proteins (gp55-

| Protein       | Basal transcription | Sliding clamp-activated transcription | Core binding | KMnO₄ footprinting |
|---------------|---------------------|---------------------------------------|--------------|--------------------|
|               | Linear              | Supercoiled                          |              |                    |
| gp55          | (1.00)              | (1.00)                                | (1.00)       | (1.00)             |
| gp55-K68A     | 0.06                | 0.17                                  | 0.97         | ++                 |
| gp55-Q69A     | 0.81                | 1.04                                  | 0.03         | ++                 |
| gp55-E70A     | 0.04                | 0.26                                  | 0.71         | +                  |
| gp55-M71A     | 0.01                | 0.12                                  | 1.15         | ++                 |
| gp55-I72A     | 0.30                | 0.48                                  | 1.20         | ++                 |
| gp55-I72K     | 0.20                | 0.43                                  | 0.70         | ++                 |
| gp55-D74A     | 0.08                | 0.35                                  | 1.15         | ++                 |
| gp55-G75A     | 0.57                | 0.83                                  | 1.11         | ++++               |
| gp55-I76A     | 0.79                | 1.03                                  | 0.95         | ++                 |
| gp55-E77A     | 2.95                | 1.38                                  | 0.38         | ND                 |
| gp55-E77R     | 0.08                | 0.09                                  | 0.95         | ++                 |
| gp55-A78G     | 0.23                | 0.45                                  | 0.93         | ++                 |
| gp55-S79A     | 0.49                | 0.73                                  | 1.02         | ++++               |
| gp55-I76A     | 0.28                | 0.46                                  | 0.81         | ++++               |
| gp55-K51A     | 0.62                | ND                                    | 0.38         | ND                 |
| gp55-G82A     | 0.89                | 0.61                                  | 0.98         | +                  |
| gp55-L83A     | 1.97                | 1.19                                  | 0.69         | +                  |

**Table I**

Summary of the effects of gp55 mutations on RNA polymerase core binding, basal transcription of linear and supercoiled DNA, gp45-activated transcription, and promoter opening.

All of the values are expressed relative to wild type gp55 in the same assay. The data for transcription are from Fig. 4, the data for RNA polymerase core binding are from Fig. 2, and the data for promoter opening assayed by KMnO₄ footprinting are from Fig. 7. ND, not determined.

**Fig. 3.** Rates of acquisition of the capacity for a single round of productive transcription at the T4 late promoter. A and B, basal transcription at 25 °C by gp55 holoenzyme assembled with wild type (wt) and mutant proteins. The time allotted to formation of promoter complexes is specified above each lane in A and in the abscissa in B. Full-length and readthrough transcripts and the recovery marker are identified at the left side of A. The 420-nucleotide full-length transcript is quantified in B. C and D, sliding clamp-activated transcription at 25 °C by T4 late holoenzyme assembled with wild type or mutant gp55 (together with wild type gp33 co-activator). Presentation of data follows A and B.
Survey of activities in basal and sliding clamp-activated transcription at 25 °C. Yields of single round transcription are shown for gp55 mutant proteins relative to wild type gp55 (gp55-wt) for basal transcription of linear DNA (allotting 20 min to promoter opening; closed bars), basal transcription of supercoiled DNA (allotting 10 min to promoter opening; open bars) and gp45-activated (enhanced) transcription (allotting 1 min to promoter opening; gray bars). The averages of at least two experiments are shown. The error bars indicate the standard deviations.

Fig. 5. gp55-E77A forms stable T4 late promoter complexes more rapidly than does the wild type protein; rate of formation of stable gp55 holoenzyme complexes with an ~300 bp DNA fragment at 37 °C. gp55 holoenzyme (40 ns) and 32P-labeled DNA (0.33 ns) were incubated at 37 °C for the indicated times, a vast excess of poly(dAT):poly(dAT) was added for 20 min, and the mixture was filtered through nitrocellulose membrane. Retained radioactivity, expressed as a fraction of input 32P-DNA, is shown on the abscissa. ○, gp55-E77A; □, wild type gp55 (WT).

K80A, gp55-E70A, gp55-M71A, gp55-I72A, gp55-I72K, gp55-D74A, gp55-A78G, and gp55-K81A). The rest were active at 60% or greater of wild type level. Equally striking was the fact that the hyperactivity of gp55-E77A was diminished (~60% or greater of wild type level). Equally striking was the fact that the hyperactivity of gp55-E77A was diminished (~60% or greater of wild type level. The time allotted to formation of open promoter complexes is noted above each lane. The 490-nucleotide full-length transcript and recovery marker are identified at the side.

Fig. 6. Activation by the sliding clamp enables a core binding-defective mutant gp55 to form transcriptionally competent promoter complexes. The acquisition of capacity for a single round of gp45-activated (enhanced) transcription at 25 °C is compared for T4 late holoenzyme assembled with wild type gp55 (wt, lanes 2–6) and gp55-E70A (lanes 7–11). Lanes 1 and 12, single-round basal transcription of linear DNA by gp55 holoenzyme (U) assembled with wild type gp55 and gp55-E70A, respectively. The time allotted to formation of open promoter complexes is noted above each lane. The 490-nucleotide full-length transcript and recovery marker are identified at the side.

The T4 late promoter opens extremely rapidly under the influence of the sliding clamp activator and gp33 co-activator (21) (Fig. 3D). Measuring the capacity for a single round of transcription after allowing only 1 min for sliding clamp loading, promoter complex formation, and opening thus is a way of monitoring the ability of variant gp55 to respond to the transcriptional activator. An experiment comparing rates of formation of transcriptionally competent promoter complexes with wild type gp55, gp55-E70A, and gp55-E77A is shown in Fig. 3C. The results of the gp55 mutant screen are compiled in Fig. 4 and Table I. Most of the mutant T4 late holoenzymes, including those assembled with gp55-K68A, gp55-I72A, gp55-I72K, gp55-A78G, gp55-S79A, gp55-I80A, gp55-K81A, and gp55-G82A, are indistinguishable from the wild type late holoenzyme in activated transcription, despite significant defects of the mutant gp55 in basal transcription of linear DNA. Even the highly defective gp55-M71A and gp55-D74A yield comparable activity (~70% of wild type), and gp55-E77R is also relatively active in sliding clamp-activated transcription. gp55-E77A and gp55-E77R, which are hyperactive for basal transcription of linear DNA, lose this advantage for gp45-activated transcription. gp55-K81D is less active in gp45-activated than in basal transcription of linear DNA (relative to the wild type). This is a potentially interesting phenotype because it suggests a defect that is specific to the activation mechanism. Only gp55-E70A is largely defective in gp45-activated transcription, although a very slowly accumulating transcription capacity can be detected even for this protein (Fig. 6).

Promoter Opening and Promoter Clearance—Open T4 late promoter complexes undergo multiple rounds of abortive synthesis of short oligonucleotides before they produce each transcript (43). At certain promoters, the σ70 holoenzyme forms an appreciable fraction of open complexes that remain in the abortive mode and never yield complete transcripts (44, 45). Thus, a quantitative comparison of the ability to form open promoter complexes and complete transcripts provides a way of determining whether a gp55 mutation primarily impairs promoter opening or promoter clearance.
Promoter opening was screened by KMnO$_4$ footprinting of the nontranscribed strand in basal promoter complexes of gp55 holoenzyme formed under the conditions of the basal transcription screen, that is, during 20 min at 25 °C, and quantified as described under “Experimental Procedures.” The principal outcome of the analysis is that promoter opening and production of transcripts are closely correlated (Fig. 7 and Table I). This indicates that defects in promoter clearance are not determining for inactivity generated by these homology segment 2.2 mutations of gp55.

**DISCUSSION**

σ$^{70}$ binds to the RNA polymerase core through separate sites that are located in its homology segments 1, 2.2, 3, and 4 (7, 12, 46, 47). Formation of the σ$^{70}$ holoenzyme is a multi-step process that is accompanied by a reorganization of interactions within σ$^{70}$ (4, 12) leading to the segregation of σ$^{70}$ structure domains to well separated sites on the surface of the core polymerase. Homology segments 1.2–2.4 constitute one of these structure domains (5).

The weak homology of gp55 with σ$^{70}$ is essentially confined to homology segment 2. The part of gp55 that is homologous with σ$^{70}$ segments 2.1 and 2.2 is most strongly protected by RNA polymerase from proteolytic cleavage by hydroxyl radical (25), implying tight binding. Our analysis of mutations in gp55 focuses on amino acids 68–83, comprising a segment that is homologous with σ$^{70}$ segment 2.2 (Fig. 1). This segment is completely conserved in the newly sequenced phage RB69. Indeed, the entire amino acid sequence of gp55 is highly conserved between phages T4 and RB69, with only 11.3% divergence, and an additional 4-amino acid insertion near the N terminus of RB69 gp55 (phage.bioc.tulane.edu).

Mutations at multiple sites in σ$^{70}$ homology segment 2.2 (L402F, D403N, D403A, Q406A, E407K, N409D, and M413T) disrupt the σ$^{70}$-core interaction (28). gp55 mutations at amino acids 71 and 72 (M71A and I72K), putatively corresponding with amino acids 404 and 405 of σ$^{70}$, were previously shown to generate defects in binding to RNA polymerase core (25). Alanine substitution mutants at all possible sites in the amino acid 68–83 segment have been examined. Ala$^{78}$ has been replaced with Gly, the amino acid at the putatively corresponding location in σ$^{70}$ and charge-reversing mutations at amino acids 77 (Glu → Lys) and 81 (Lys → Asp) have also been introduced.

As anticipated, many of these mutations affect gp55 binding to the RNA polymerase core (Fig. 2 and Table I). In addition to the previously analyzed gp55-M71A and gp55-I72K, mutant proteins with significant core binding defects include gp55-E70A, gp55-D74A, gp55-E77A, gp55-A78G, gp55-G82A, and gp55-L83A. Although aggregation could contribute to the core-binding defects of gp55-G82A and gp55-L83A, it is unlikely that it is solely responsible for very weak binding. Indeed, gp55-L83A is transcriptionally highly active under all of the tested conditions (Figs. 3A and 4). The scale that has been used to designate binding defects in Table I can be calibrated by reference to gp55-M71A and gp55-I72K, which have also been examined by affinity chromatography and found not to bind core enzyme in presence of 400 mM NaCl. (Core binding by gp55-I72K is also partly defective in presence of 250 mM NaCl (25).) E70A and D74A generate even more defective core binding than M71A and I72K.

These effects of mutations on core binding can be interpreted in terms of the recently determined structures of T. aquaticus (Taq) and T. thermophilus (Tth) α$^{70}$RNA polymerase holoenzymes (8, 9, 15), which retain the four-helix structure previously determined for E. coli σ$^{70}$ and Taq α$^{70}$ domain 2 (σ$_{2}$), comprising homology segments 1.2–2.4 (5, 14). The α-helical homology segments 2.2 in these holoenzyme structures interact directly with an antiparallel coiled-coil segment of β$^{70}$ (E. coli amino acids 265–310; Tth amino acids 540–585).

The proposed alignment of gp55 amino acids 68–82 with σ homology segment 2.2 (Fig. 1) has been tested by comparing the outcome of the gp55 mutagenesis (Table I) with structure predictions based on α$^{70}$ segment 2.2-β$^{70}$ coiled-coil interactions within the Tth and Taq holoenzymes. Tth and Taq α$^{70}$ homology segments 2.2 are identical (Fig. 1), and the β$^{70}$-coils differ only by two conservative substitutions (Tth Leu$^{681}$ is Val$^{581}$ in Taq β$^{70}$; Tth Leu$^{682}$ is Ile$^{582}$ in Taq β$^{70}$). The comments that follow refer directly to the higher resolution Tth holoenzyme structure but can be inferred from either structure.

There is a periodicity of effect of alanine substitutions in gp55 on E. coli RNA polymerase core binding, with the E70A, D74A, and E77A mutations generating the greatest defects (Table I). In the gp55-α$^{70}$ alignment, gp55 Glu$^{70}$, Asp$^{74}$, and Glu$^{77}$ correspond with Tth α$^{70}$ Asp$^{211}$, Glu$^{215}$, and Glu$^{218}$, whose side chains face the β$^{70}$-coiled-coil.

Tth α$^{70}$ Asp$^{211}$ interacts with Tth β$^{70}$ Arg$^{560}$ and Arg$^{563}$ corresponding to Arg$^{275}$ and Arg$^{278}$ in E. coli β$^{70}$, respectively. A region 2.2. helix-β$^{70}$ coiled-coil energy-minimized model with gp55 amino acids 68–83 threaded into the structure in place of the corresponding segment of α$^{70}$ has gp55 Glu$^{70}$ also interacting with the same Arg side chains in the β$^{70}$-coiled-coil. In a mutational analysis of the E. coli β$^{70}$-coiled-coil region, Arthur and co-workers (47) found Arg → Gln at amino acid 275 greatly diminishing the ability to assemble σ$^{70}$ holoenzyme in vivo and essentially eliminating the ability of the β$^{70}$ amino acids 1–319 segment (which includes the entire coiled-coil) to bind σ$^{70}$ in far Western blotting.

Tth α$^{70}$ Glu$^{215}$ interacts with β$^{70}$ Arg$^{563}$, in the energy-minimized gp55 model, the corresponding Asp$^{77}$ is similarly juxtaposed with β$^{70}$ Arg$^{563}$. Tth α$^{70}$ Glu$^{218}$ interacts with β$^{70}$ Lys$^{566}$ (Arg$^{278}$ in E. coli β$^{70}$); in the model, the corresponding Glu$^{77}$ of gp55 is also located in proximity to β$^{70}$ Lys$^{563}$. Thus, in terms of the Tth holoenzyme structure, it is plausible that eliminating any of these three favorable charge interactions between gp55 region 2.2 and β$^{70}$ and substituting a shorter side chain in gp55 would diminish core binding.

The three-amino-acid phasing of α coiled-coil interaction is interrupted beyond this point. gp55 Ile$^{80}$ corresponds to Tth α$^{70}$
Ile221, which is in vicinity of β' Gin560, corresponding to Leu295 in E. coli β'. Interactions between these side chains should not contribute significantly to affinity. Gin560 is in the loop connecting the antiparallel α helices of the β' coiled-coil, and the lack of effect of the gp55-I80A mutation on core binding is consistent with the Thr holoenzyme structure. Gly392 and Leu393 of gp55 correspond with Thr σα Ala293 and Val294, which are not in the immediate vicinity of β'. In the gp55 model, the Gly392 and Leu393 side chains face in the direction of the crossing region 2.1 and 2.3 helices. Thus, examination of the model suggests that weak core binding of the gp55-O82A and gp55-L82A mutants may be due to effects of these mutations on the folding of gp55 region 2.

What is striking about the properties of these mutants is their diversity of phenotype for basal and activated transcription and the diversity of relationship between RNA polymerase core binding and transcriptional activity. The existence of such a range of defect is conceptually congruent with recent insights into the function of σ70 homology segment 2.2 (13). σ70 binds the nontranscribed strand and duplex DNA of the −10 promoter region (11) through its homology segment 2.4. The switch from double-stranded DNA binding to single-stranded DNA binding requires a structural transition that is generated by interaction of σ70 with the β' subunit of the core enzyme (48). The β' coiled-coil interacts with σ homology segment 2.2, as already specified; segments 2.2 and 2.4 are apposed (14). Evidently, polymerase core binding by homology segment 2.2 is coupled to the promoter opening capacity of σ70 homology segment 2.4 (13), but how that coupling is achieved is not yet clear.

Most of the transcriptional defects that are detected with linear DNA templates are strongly mitigated by the sliding clamp activator. Several factors may operate to produce this effect. First, the additional interactions of the sliding clamp-activated promoter complex, between gp55 and DNA-confined gp45 and between gp45 and core-bound gp33, should increase the effective affinity of gp55 for RNA polymerase core. Second, transcription is assayed with a 6-fold excess of gp55 over core. For sliding clamp-activated transcription, this is an at least 4-fold excess over the core-saturating concentration of wild type gp55. The excess can compensate for considerably weaker binding by mutant gp55. Third, it is conceivable that gp45 interaction exerts a direct function-restoring effect on the structure of some of these mutant proteins. The extremely weakly binding gp55-D74A could be a candidate for such an effect.

Overall, these gp55 homology segment 2.2 mutations generate an ~300-fold range of activity in basal transcription of linear DNA. The close correlation between promoter opening, as determined by permanganate footprinting, and transcriptional activity specifies that these gp55 mutations neither create nor eliminate barriers to the transition from abortive to productive transcription elongation. Among the other phenotypes, three are particularly interesting. First, gp55-E70A is almost completely inactive for basal transcription of linear DNA and for sliding clamp-activated transcription but significantly active for transcription of supercoiled DNA. The gp55-E70A RNA polymerase opens the promoter slowly in supercoiled DNA but ultimately yields a nearly wild type level of single round transcription (data not shown). This result suggests that the E70A mutation also affects the ability to respond to the transcriptional activator. The primary interaction site of gp55 with the gp45 sliding clamp is its C-terminal hydrophobic/acidic epitope (21, 23). The transcriptional defect generated by the E70A mutation is therefore likely to reside in a step of the reaction pathway that follows gp55-gp45 interaction, most probably promoter opening. The fact that the defect of gp55-E70A manifests itself most severely in basal as well as activated transcription of linear DNA suggests that the sliding clamp activator and DNA underwinding relieve distinguishable rate limitations on the opening of T4 late promoters by the gp55-RNA polymerase holoenzyme.

Second, gp55-E77A and gp55-L83A are significantly defective in core binding but hyperactive in forming open promoter complexes for basal transcription of linear DNA (Figs. 2–4 and Table I). The "advantage" of these mutant proteins is almost completely lost in the context of basal transcription of supercoiled DNA (and cannot be tested properly for gp45-activated transcription under our assay conditions because the late promoter opens so rapidly with wild type gp55 (24)). An inverse correlation between core binding and promoter opening in basal transcription implies that the interaction of wild type gp55 segment 2.2 with polymerase core, which is essential to its function, nevertheless limits the rate of promoter opening. This is reminiscent of the just-published observation that disulfide bridge-locking σ70 segment 2.1 or 2.3 with segment 2.2 yields holoenzymes that are functional but quantitatively deficient in formation and stability of open promoter complexes. Evidently, conformational flexibility within structure domain 2 of σ70 facilitates DNA strand opening in the promoter complex and initiation of transcription (49). Comparable structure adaptations probably figure in gp55-dependent T4 late transcription.

Third, conversely, certain gp55 mutations significantly diminish activity for basal transcription without significantly diminishing core binding. gp55-I80A and gp55-K81A fall into this category, with the K81A mutant protein showing the most pronounced phenotype. The K68A mutation only barely affects core binding but makes unenhanced transcription grossly defective. Activation by gp45 restores transcription to the wild type level. Under reaction conditions that differ only slightly, the sliding clamp activator increases the second order rate constant for T4 late promoter opening by gp55-RNA polymerase ~340-fold (24). For gp55-K68A, the activation ratio must be an additional order of magnitude greater.

Acknowledgments—We thank R. R. Burgess, N. Thompson, C. A. Gross, and M. M. Sharp for generous gifts of antibodies, M. M. Sharp for guidance, J. D. Karam, and H. Krisch for sharing information in advance of publication, and S. Kolesy and M. Ouhannob for advice and helpful discussions.

REFERENCES
1. Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Chulis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, W., Warren, T., Wietzorek, A., Woodward, J., Barrett, B. G., Parkhill, J., and Hopwood, D. A. (2002) Nature 417, 141–147.
2. Grubskov, M., and Burgess, R. R. (1996) Nucleic Acids Res. 24, 6745–6763.
3. Helmann, J. D., and Chamberlin, M. J. (1988) Annu. Rev. Biochem. 57, 839–872.
4. Callaci, S., Heyduk, E., and Heyduk, T. (1999) Mol. Cell 3, 229–238.
5. Campbell, R. A., Muzzon, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinnman, O., Trester-Zelditz, M. L., and Darst, S. A. (2002) Mol. Cell 9, 527–539.
6. Meckler, V., Korkhoff, J., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R., and Ebright, R. H. (2002) Cell 109, 599–614.
7. Mukhopadhyay, J., Kapanidis, A. N., Meckler, V., Korkhoff, J., Ebright, Y. W., and Ebright, R. H. (2001) Cell 106, 454–463.
8. Murakami, K. S., Masuda, S., and Darst, S. A. (2002) Science 296, 1280–1284.
9. Vassilyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassilyev, M. N., Borukhov, S., and Yokoyama, S. (2002) Nature 417, 712–719.
10. Domrnski, A. J., Walter, W. A., and Gross, C. A. (1993) Genes Dev. 7, 2446–2455.
11. Marx, M. T., and Roberts, J. W. (1997) Science 276, 1288–1290.
12. Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., Artsimovitch, I., Geszvain, K. M., Arthur, T. M., Burgess, R. R., Landick, R., Severinov, K., and Gross, C. A. (2001) Mol. Cell 8, 21–31.
13. Young, B. A., Anthony, L. C., Gruber, T. M., Arthur, T. M., Heyduk, E., Lu, C. Z., Sharp, M. M., Heyduk, T., Burgess, R. R., and Gross, C. A. (2001) Cell 105, 935–944.
