PNAS
www.pnas.org

Supplementary Information for

Structural and mechanistic basis of σ-dependent transcriptional pausing

Chirangini Pukhrambam[a,b,1], Vadim Molodtsov[a,c,1], Mahdi Kooshkbaghi[d], Ammar Tareen[d], Hoa Vu[a,b], Kyle S. Skalenko[a,b], Min Su[e], Zhou Yin[a,c], Jared T. Winkelman[a,b,c], Justin B. Kinney[d], Richard H. Ebright[a,c,2], Bryce E. Nickels[a,b,2]

corresponding authors: Richard H. Ebright and Bryce E. Nickels

Email: ebright@waksman.rutgers.edu and bnickels@waksman.rutgers.edu

This PDF file includes:

SI Materials and Methods

Strains

Plasmids were maintained in *E. coli* strain DH10B (Thermo Fisher Scientific) and NEB 5-alpha (New England Biolabs). Protein expression was performed using *E. coli* strain NiCo21(DE3) (New England Biolabs) or *E. coli* strain BL21 Star (DE3) (Invitrogen). *E. coli* strain ML149 (1) is a derivative of MG1655 that contains a kanamycin cassette downstream of *rpoD*. *E. coli* strain ML176 (1) is a *greA*$^{-}$ derivative of ML149. Strains ML149 and ML176 were used for analysis of σ-dependent pausing *in vivo*.

Oligodeoxyribonucleotides

Oligonucleotides (Table S1) were dissolved in nuclease-free water to 1 mM and stored at -80°C.

Plasmids

Plasmid pLHN12-His (2) contains the gene for *E. coli* σ^{70} with an N-terminal histidine coding sequence (6xHis) under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible. Plasmid pLHN12-His-(R541C, L607P) is a derivative of pLHN12-His with substitutions that disrupt interaction between σ^{70} region 4 and the β flap (3).

Plasmid pIA900-RNAP-$\beta'^{R1148Bpa}$ and pIA900-RNAP-β'^{T48Bpa} contains genes for the RNAP α subunit, RNAP β subunit, RNAP ω subunit, and RNAP β′ subunit with a nonsense codon (TAG) at position 1148 or 48, respectively, with an N-terminal decahistidine coding sequence (4).

Plasmid pEVOL-pBpF [Addgene; (5)] contains genes directing the synthesis of an engineered Bpa-specific UAG-suppressor tRNA and an engineered Bpa-specific aminoacyl-tRNA synthetase that charges the amber suppressor tRNA with Bpa.

Plasmid pσ^{70}, is a derivative of pBAD24 [(6); American Type Culture Collection, ATCC], containing the gene for *E. coli* σ^{70} with an N-terminal decahistidine coding sequence under the control of an arabinose-inducible promoter (pBAD). To construct plasmid pσ^{70}, we used PCR to generate a ~4.6 kb DNA fragment containing a decahistidine coding sequence, pBAD promoter, *araC*, pBR322 *ori*, and ampicillin resistance gene and a ~2 kb DNA fragment containing the gene for *E. coli* σ^{70}. To generate the pBAD-containing fragment, reactions contained plasmid pBAD24 (0.04 μM), primer JW119 (0.5 μM), primer JW270 (0.5 μM), and 1X Phusion HF Master Mix (Thermo Fisher Scientific). To generate the σ^{70}-containing fragment, reactions contained pET28-*rpoD* (0.04 μM; gift of J. Roberts), primer JW155 (0.5 μM), primer JW154 (0.5 μM), and 1X Phusion HF Master Mix. (30 cycles at 95°C, 10 sec; 55°C, 10 sec, 72°C, 4 min). After PCR, 20 U of DpnI (New England Biolabs) was added, reactions were incubated at 37°C for 16 hrs, and nondigested DNA fragments were isolated using a PCR purification kit (Qiagen). Recovered fragments (~50 ng) were mixed with 1X Gibson assembly master mix (New England Biolabs) in a 20 μl reaction and incubated at 50°C for 20 min. Next, 1 μl of the Gibson assembly reaction was introduced by electroporation into DH10B cells (Invitrogen), cells were plated on LB agar containing 100 μg/ml carbenicillin, recombinant plasmid DNA was isolated from individual transformants, and plasmid sequences verified by Sanger sequencing (Macrogen/Psomagen).

Plasmid pσ^{70}-R448Bpa-His^{10}, is a derivative of plasmid pσ^{70} with a nonsense codon (TAG) at position 448 of the gene for *E. coli* σ^{70}. Plasmid pσ^{70}-R448Bpa-His^{10} was generated by site-directed mutagenesis. Reactions (12.5 μl volume) contained 0.04 μM of pσ^{70}, 0.5 μM oligo HV75 and 1 X Phusion HF Master Mix (Thermo Fisher Scientific) (95°C for 2 min; 95°C for 15 sec, 55°C for 15 sec, 72°C for 2 min; 30 cycles). After PCR, 20 U of DpnI (New England Biolabs) was added, reactions were incubated at 37°C for 16 hrs. Next, 1 μl of reaction was introduced by electroporation into DH10B cells (Invitrogen), cells were plated on LB agar containing 100 μg/ml

carbenicillin, recombinant plasmid DNA was isolated from individual transformants, and plasmid sequences verified by Sanger sequencing (Macrogen/Psomagen).

Plasmid pCDF-λPR′ is a derivative of pCDF-CP (7) that contains λPR′ sequences from positions -65 to +50 inserted into BglI-digested pCDF-CP. Plasmid pCDF-CP contains a CloDF13 replication origin, a selectable marker conferring resistance to spectinomycin and streptomycin, and two BglI recognition sites used to introduce DNA fragments upstream of transcription terminator tR2.

The +14-20 library was generated using procedures described in (8, 9) with JW615 as template and s1219 and s1220 as amplification primers. The amplification primers have 5′ end sequences that introduce BglI recognition sequences and 3′ end sequences complementary to the template oligo. PCR amplicons were treated with BglI and BglI-digested fragments were ligated into BglI-digested pCDF-CP. The ligation mixture was transformed into NEB 5-alpha cells (New England Biolabs), cells plated on LB agar plates containing 50 μg/ml spectinomycin and 50 μg/ml streptomycin, and recombinant plasmid DNA (pCDF-λPR′-N7) was isolated from ~0.5 x 10^6 transformants.

Proteins

RNAP core enzyme was prepared from *E. coli* strain NiCo21(DE3) (New England Biolabs) containing plasmid pIA900 (10) using procedures described in (4).

Bpa-containing RNAP core enzyme derivatives, RNAP-$\beta'^{R1148Bpa}$ and RNAP-β'^{T48Bpa}, were prepared from *E. coli* strain NiCo21(DE3) (New England Biolabs) containing plasmid pIA900-RNAP-$\beta'^{R1148Bpa}$ or plasmid pIA900-RNAP-β'^{T48Bpa} (4) and plasmid pEVOL-pBpF (5), using procedures described in (4).

Wild-type σ^{70} was purified using procedures described in (11). Bpa-containing σ^{70} derivative, $\sigma^{70\ R448Bpa}$, was prepared from *E. coli* strain DH10B (Invitrogen) containing plasmid pσ^{70}-R448Bpa-His^{10} using procedures described in (4). $\sigma^{70\ C541,\ P607}$ was prepared from *E. coli* strain BL21 Star (DE3) (Invitrogen) containing pLHN12-His-(R541C, L607P) using procedures described in (2).

RNAP σ^{70} holoenzyme was prepared by incubating 1 μM *E. coli* RNAP core enzyme and 5 μM *E. coli* σ^{70} in 10 mM Tris-Cl (pH 8.0), 100 mM KCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, and 50% glycerol for 30 min at 25°C.

GreB was purified using procedures described in (12).

DNA templates for *in vitro* assays

DNA templates used for *in vitro* assays in Figures 1C, 3A, and S1-S4, which contain sequences from positions -86 to +96 of λPR′ promoter, were generated by PCR amplification of ~1 pg pCDF-λPR′ with 0.4 μM primer JW521 and 0.4 μM primer JW544 in reactions containing 1 X Phusion HF Master Mix (Thermo Fisher Scientific). Amplicons were purified using a PCR purification kit (Qiagen).

DNA templates used for *in vitro* assays in Figure 5, which contain sequences from positions -65 to +79 of λPR′ promoter with a T, G or X (where X is an abasic site) at nontemplate strand position +14, were prepared by mixing 0.4 μM of template-strand oligo (JW680) with 0.4 μM nontemplate strand oligo (JW679, JW683 or JW687) in 1 X Phusion HF Master Mix (Thermo Fisher Scientific) and performing 30 cycles of 95°C for 2 min; 95°C for 15 sec, 55°C for 15 sec, and 72°C for 2 min. Double stranded products were gel purified to remove non-annealed oligos.

σ-dependent pausing *in vitro*

In vitro transcription assays in Figure 1C and S1 were performed essentially as described in (3, 13). Reactions in Figure 1C (120 μL total volume) contained 10 nM of λPR′ template, 40 nM of RNAP holoenzyme (wild-type), 200 μM each of ATP, GTP, CTP and UTP supplemented with

0.03 mCi α-^{32}P-UTP (Perkin Elmer; 3000 Ci/mmol) in 1X reaction buffer (RB; 10 mM Tris-HCl, pH 8.0; 70 mM NaCl, 0.1 mg/ml bovine serum albumin, BSA; and 5% glycerol) and 100 nM GreB (where indicated) were incubated for 10 min at 37°C. Next, a mixture of 10 mM $MgCl_2$ and 10 μg/ml rifampin (Gold Biotech) was added, 20 μL aliquots were removed at the indicated times (20s, 40s, 60s, 80s), and mixed with five volumes of stop solution (500 mM Tris-HCl, pH 8.0; 10 mM EDTA; and 0.1 mg/ml glycogen). 120 μL of phenol:chloroform (pH 4.5; Ambion) was added, samples were mixed, the aqueous layer (~100 μL) was removed, 100% ethanol was added (~300 μL), samples were placed at −80°C for 12 h, precipitated nucleic acids were recovered by centrifugation, re-suspended in 5 μl nuclease free water, and mixed with 5 μL loading dye (1 X TBE, 8 M urea, 0.025% xylene cyanol, and 0.025% bromophenol blue). Samples were heated at 95°C for 2 min, cooled to 25°C, and analyzed by electrophoresis on 20%, 8 M urea, 1 X TBE polyacrylamide gels (Urea Gel System; National Diagnostics). Radiolabeled bands were visualized by storage phosphor screen (GE Healthcare) and phosphorimagery (Typhoon 9400 variable mode imager, GE Healthcare). Sizes of RNA products were estimated by comparison to radiolabeled Decade Marker (Thermo Fisher Scientific).

Reactions in Figure S1 (20 μL total volume) contained 4 nM of the λPR′ linear DNA template, 40 nM of the indicated RNAP holoenzyme, 1X RB (10 mM Tris–HCl, pH 8.0; 70 mM NaCl; 10 mM $MgCl_2$; 0.1 mg/ml BSA; and 5% glycerol) and 100 nM GreB (where indicated) were incubated for 10 min at 37°C, 200 μM ATP, 200 μM GTP, 200 μM CTP and 200 μM UTP supplemented with 0.02 mCi α-^{32}P-UTP (Perkin Elmer; 3000 Ci/mmol) were added, reactions were incubated at 37°C for 10 min, 100 μL of stop solution was added, and samples further processed as described in the preceding paragraph.

σ-dependent pausing *in vivo*

ML149 (*greA*$^+$) and ML176 (*greA*$^-$) cells containing plasmid pCDF-λPR′ were grown in 25 ml LB containing 100 μg/ml carbenicillin and 1 mM IPTG in 125 ml DeLong flasks (Bellco Glass), shaken at 37°C on an orbital platform shaker at 220 rpm. When cultures reached an OD_{600} of ~0.5, 1 ml aliquots of cell suspensions were removed before or at the indicated time points after addition of 500 μg/ml rifampin (Gold Biotech). 1 ml aliquots of cell suspensions were mixed with 3 ml of RNAlater solution (Thermo Fisher Scientific) in 50 ml Oakridge tubes by inversion several times and incubated overnight at 4°C. The mixture was centrifuged at 17,000 × g for 20 min at 4°C, supernatant was removed, 1 ml of Tri-reagent (Molecular Research Center) was added and pellets were dispersed by vortexing. Cell suspensions in Tri-reagent were transferred to 1.7 ml low binding tubes (Axygen), incubated at 70°C for 10 min, centrifuged at 21,000 × g at 4°C for 10 min, and the supernatants were recovered into fresh tubes. 200 μl of chloroform was added to each tube and mixed by vigorous shaking for 15 s. Phases were separated by centrifugation at 21,000 × g at 4°C for 15 min. 500 μl of the upper, aqueous phase was recovered and transferred to a fresh tube to which 167 μl of 100% ethanol was added. Subsequent removal of RNA >200 nt and recovery of RNA <200 nt was performed using the mirVana microRNA Isolation kit (Thermo Fisher Scientific) according to the manufacturer's protocol. After elution from mirVana columns, eluates were concentrated by ethanol precipitation and resuspended directly into formamide loading dye (95% deionized formamide, 18 mM EDTA, and 0.025% SDS, xylene cyanol, bromophenol blue, amaranth).

50 pmol of a locked nucleic acid (LNA) probe complementary to PR' sequences +1 to +17 (LNA8) was incubated in a 20 μl volume with 5 μl γ-^{32}P-ATP (EasyTide; Perkin Elmer), 2 μl 10X T4 PNK buffer, 9 μl nuclease free water (Thermo Fisher Scientific), and 2 μl T4 PNK (New England Biolabs) at 37°C for 1 hr followed by 95°C for 10 min. Labeled probe was separated from unincorporated radiolabeled nucleotide using a size-exclusion spin column (Cytiva Illustra Microspin G-25; Thermo Fisher Scientific).

Pause RNAs and full-length RNAs generated from λPR′ *in vivo* were detected by hybridization as described in (14, 15). RNA isolated from cells was subjected to electrophoresis on 20% 8M urea slab gels (equilibrated and run in 50 mM MOPS, pH 7), transferred to a neutral

nylon membrane (Whatman Nytran N; GE Healthcare Life Sciences) using a semi-dry electroblotting apparatus (Biorad) operating at 20V for 25 min using chilled 20 mM MOPS (pH 7) as conductive medium. RNA was crosslinked to the membrane using 157 mM N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide hydrochloride (EDC) (Sigma–Aldrich) in 0.97% 1-methylimidazole (pH 8) (Alfa Aesar) for 60 min at 55°C. Membrane was washed with 20 mM MOPS (pH 7) at 25°C, placed on nylon hybridization mesh, the membrane mesh stack was placed into a hybridization bottle (70 × 150 mm) at 50°C and 50 ml of pre-hybridization solution [5X SSC, 5% SDS, 2X Denhardt's solution, 40 µg/ml sheared salmon sperm DNA solution (Thermo Fisher Scientific), 20 mM Na_2HPO_4 (pH 7.2) in diethylpyrocarbonate (DEPC) treated water] at 50°C was added. The hybridization bottle was incubated at 50°C for 30 min with constant rotation, the solution was decanted and replaced by a 50 ml of pre-warmed hybridization solution with radiolabeled LNA probe prepared above and continue incubation at 50°C for 16 hr. The membrane was washed in non-stringent wash buffer [3X SS, 5% SDS, 10X Denhardt's solution, 20 mM Na_2HPO_4 (pH 7.2) in DEPC treated water] two times for 10 min, two times for 30 min and in stringent wash buffer (1X SSC, 1% SDS, in DEPC treated water) one time for 5 min before it was blotted dry, wrapped in plastic film, and radiolabeled bands were visualized by storage phosphor screen (GE Healthcare) and phosphorimagery (Typhoon 9400 variable mode imager, GE Healthcare).

In vitro protein-DNA photocrosslinking

In vitro protein-DNA photocrosslinking and crosslink mapping experiments were performed as described in (7, 16). Reactions contained 40 nM RNAP holoenzyme (β′T48Bpa, σ^{70}R448Bpa or β′R1148Bpa), 4 nM template, 1 X RB (10 mM Tris-Cl, pH 8.0; 70 mM NaCl; 10 mM $MgCl_2$; 0.1 mg/ml BSA; and 5% glycerol) and 100 nM GreB (where indicated) were incubated for 2 min at 37°C, 200 µM ATP, 200 µM CTP, 200 µM GTP, and 200 µM UTP were added, reactions were further incubated for 10 min, and subjected to UV irradiation for 10 min at 25°C in a Rayonet RPR-100 photochemical reactor equipped with 16 x 350 nm tubes (Southern New England Ultraviolet).

Reactions were mixed with 15 µL 5 M NaCl and 6 µL 100 µg/µl heparin, incubated for 5 min at 95°C and then cooled to 4°C. RNAP-DNA crosslinked complexes were isolated by adding 20 µl MagneHis Ni-particles (Promega) equilibrated and suspended in 1 X Taq DNA polymerase buffer, 10 mg/ml heparin, and 0.1 mg/ml BSA; MagneHis Ni-particles were collected using a magnetic microfuge tube rack; particles were washed with 1 X Taq DNA polymerase buffer, 10 mg/ml heparin, and 0.1 mg/ml BSA, washed twice with 50 µl 1 X Taq DNA polymerase buffer (New England Biolabs), and particles (which contained bound RNAP-DNA complexes) were resuspended in 10 µl 1 X Taq DNA polymerase buffer.

Primer extension reactions (12.5 µl) were performed by combining 2 µl of the recovered RNAP-DNA complexes, 1 µl of 1 µM ^{32}P-5′ end-labeled primer JW77 (for leading edge position) or JW544 (for trailing edge and σR2), 1 µL 10 X dNTPs (2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM dTTP), 0.25 µL 5 U/ml Taq DNA polymerase (New England Biolabs), 5 µL 5 M betaine, 0.625 µL 100% dimethyl sulfoxide, and 1.25 µl 10 X Taq DNA polymerase buffer; 40 cycles of 30 sec at 95°C, 30 sec at 55°C, and 30 sec at 72°C. Reactions were stopped by addition of 12.5 µL 1 X TBE, 8 M urea, 0.025% xylene cyanol, and 0.025% bromophenol blue. Radiolabeled products were separated by electrophoresis on 8% 8M urea slab gels (equilibrated and run in 1 X TBE) and visualized by storage-phosphor imaging (Typhoon 9400 variable-mode imager; GE Life Science). Positions of RNAP-DNA crosslinks were determined by comparison to products of a DNA-nucleotide sequencing reaction generated using oligo JW77 or JW544 and a DNA template containing sequences from positions -86 to +96 of pCDF-λPR′ (Thermo Sequenase Cycle Sequencing Kit; Affymetrix).

In vivo protein-DNA photocrosslinking

In vivo protein-DNA photocrosslinking and crosslink mapping experiments were done essentially as in (16, 17). Experiments in Figure 3A and S2-S4 were performed by sequential introduction of plasmid pCDF-λPR′, plasmid pIA900-RNAP-β′^{T48Bpa} or pIA900-RNAP-β′^{R1148Bpa} or pσ$^{70\ R448Bpa}$ and plasmid pEVOL-pBpF into electrocompetent *E. coli* strain NiCo21(DE3) by transformation. After the final transformation step, cells were plated on LB agar containing 100 μg/ml carbenicillin, 50 μg/ml spectinomycin, 50 μg/ml streptomycin, and 25 μg/ml chloramphenicol; at least 1,000 individual colonies were scraped from the plate, combined, and used to inoculate 250 ml LB broth containing 1 mM Bpa (Bachem), 100 μg/ml carbenicillin, 50 μg/ml spectinomycin, 50 μg/ml streptomycin, and 25 μg/ml chloramphenicol in a 1000 mL flask (Bellco Glass) to yield OD_{600} = 0.3; the culture was placed in the dark and shaken (220 rpm) for 1 h at 37°C; 1 mM IPTG (for pIA900-RNAP-β′^{R1148Bpa} or pIA900-RNAP-β′^{T48Bpa}) or 0.2% L-arabinose (for σ$^{70\ R448Bpa}$) was added to induce expression; and the culture was placed in the dark and shaken (220 rpm) for 3 hrs at 37°C (for pIA900-RNAP-β′^{R1148Bpa} or pIA900-RNAP-β′^{T48Bpa} expression) or 3 hrs at 30°C (for σ$^{70\ R448Bpa}$ expression).

UV irradiation of cell suspension, purification of RNAP-DNA photocrosslinked complexes from cell cultures, denaturation, isolation, primer extension and electrophoresis for mapping RNAP-DNA crosslinked complexes were done following the procedures as described in (16, 17).

Analysis of σ-dependent pausing *in vivo* for +14-20 library transcription complexes (Figure 4) was performed by sequential introduction of plasmid pCDF-λPR′-N7 (+14-20) library (yielding ~28 million transformants), plasmid pIA900-RNAP-β'^{R1148Bpa} (yielding ~8 million transformants), and plasmid pEVOL-pBpF (yielding ~5 million transformants) into electrocompetent *E. coli* strain NiCo21(DE3). After the final transformation step, cells were plated on ~4-6 LB agar plates containing 100 μg/ml carbenicillin, 50 μg/ml spectinomycin, 50 μg/ml streptomycin, and 25 μg/ml chloramphenicol to yield a lawn. Colonies were scraped from the surface of the plates, combined, and used to inoculate 150 ml LB broth containing 1mM Bpa, 100 μg/ml carbenicillin, 50 μg/ml spectinomycin, 50 μg/ml streptomycin, and 25 μg/ml chloramphenicol in a 1000 mL flask to yield OD_{600} = 0.3; the culture was placed in the dark and shaken (220 rpm) for 1 h at 37°C; IPTG was added to 1 mM and the culture was placed in the dark and shaken (220 rpm) for 3 h at 37°C.

To measure background signal, a portion of the cell cultures containing pCDF-λPR′ or pCDF-λPR′-N7 (+14-20 library) were removed, rifampin (Gold Biotech) was added to a final concentration of 200 μg/ml, and the culture was shaken at 37°C for 10 min prior to UV irradiation.

UV irradiation of cell suspension, purification of RNAP-DNA photocrosslinked complexes from cell cultures, denaturation, isolation, primer extension and electrophoresis for mapping RNAP-DNA crosslinked complexes were done following the procedures as described in (16, 17).

XACT-seq experiments (see below) were performed using denatured RNAP-DNA complexes isolated from cells containing the +14-20 library.

XACT-seq: primer extension

Primer extension was performed in 50 μl reactions containing 8 μl of recovered RNAP-DNA complexes, 1 μl of 10 μM primer s128a, 5 μl 10 X dNTPs (2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM dTTP), 1 μl 5 U/μl *Taq* DNA polymerase, 20 μl 5 M betaine, 2.5 μl 100% dimethyl sulfoxide, and 5 μl 10 X *Taq* DNA polymerase buffer, and cycling 40 times through 30 s at 95°C, 30 s at 55°C, and 30 s at 72°C. Primer extension products were isolated by phenol:chloroform:IAA pH 8.0 extraction followed by ethanol precipitation, washed twice with 80% cold ethanol, resuspended in 20 μl water, and mixed with 20 μl of 2 X RNA loading dye (95% deionized formamide, 18 mM EDTA, 0.25% SDS, xylene cyanol, bromophenol blue, amaranth).

Primer extension products were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1 X TBE), stained with SYBR Gold nucleic acid gel stain (Life Technologies) and ssDNA products ~50- to ~100-nt in size were excised from the gel. The gel fragment was crushed to elute nucleic acid from gel as described in (9), 350 μl of 0.3 M NaCl in 1 X TE buffer was added, the mixture was incubated for 10 min at 70°C, and the supernatant was

collected using a Spin-X column (Corning). The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 5 µl of nuclease-free water.

XACT-seq: 3′-adapter ligation and library amplification

The recovered primer extension products (5 µl) were combined with 1 µl 10 X NEBuffer 1, ~0.8 µM 3′-adapter oligo s1248 [5′ adenylated and 3′-end blocked oligo containing ten randomized nucleotides (10N) at the 5′ end], 5 mM $MnCl_2$ and 1 µM of 5′-AppDNA/RNA ligase (New England Biolabs) in a final volume of 10 µl. The mixture was incubated for 1 h at 65°C followed by 3 min at 90°C, and cooled to 4°C for 5 min. The reaction was combined with 15 µl of mixture containing 10 U of T4 RNA ligase 1 (New England Biolabs), 1 X T4 RNA ligase 1 reaction buffer, 12% PEG 8000, 10 mM DTT, 60 µg/mL BSA. Reactions were incubated at 16°C for 16 h.

Adapter-ligated products were separated by electrophoresis on 10% 7M urea slab gels (equilibrated and run in 1 X TBE), stained with SYBR Gold nucleic acid gel stain and species ranging from ~80 to ~150-nt were isolated by gel excision. The gel fragment was crushed, 400 µl of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column (Corning). The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by ethanol precipitation, washed twice with 80% cold ethanol, and resuspended in 13 µl of nuclease-free water.

Adapter-ligated DNA (1 µl) were used as template in emulsion PCR (ePCR). Reactions contained 1 X Detergent-free Phusion HF reaction buffer containing 5 µg/ml BSA, 0.4 mM dNTPs, 0.5 µM Illumina RP1 primer, 0.5 µM Illumina index primer and 0.04 U/µl Phusion HF polymerase [95°C for 10 s, 95°C for 5 s, 60°C for 5 s, 72°C for 15 s (20 cycles), 72°C for 5 min]. Amplicons were recovered using a Micellula DNA Emulsion and Purification Kit. The emulsion was broken, DNA was purified according to the manufacturer's recommendations, recovered by ethanol precipitation, and resuspended in 15 µl of nuclease-free water. Reaction products were separated by electrophoresis on a non-denaturing 10% slab gel (equilibrated and run in 1 X TBE), and amplicons between ~150 bp and ~220 bp were isolated by gel excision. The gel fragment was crushed, 400 µl of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column. The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by isopropanol precipitation, washed twice with 80% cold ethanol, and resuspended in 15 µl of nuclease-free water.

Libraries generated by this procedure are: CP49, CP50, CP51, CP52, CP53 and CP54.

XACT-seq: analysis of template sequences in the +14-20 library

To identify template sequences present in the +14-20 library, we performed ePCR in reactions containing ~10^9 molecules of the pCDF-λPR′-N7 (+14-20) plasmid library, 1 X Detergent-free Phusion HF reaction buffer with 5 µg/ml BSA, 0.4 mM dNTPs, 0.5 µM Illumina RP1 primer, 0.5 µM Illumina index primer and 0.04 U/µl Phusion HF polymerase [95°C for 10 s, 95°C for 5 s, 60°C for 5 s, 72°C for 15 s (30 cycles), 72°C for 5 min]. Amplicons were recovered using a Micellula DNA Emulsion and Purification Kit. The emulsion was broken, DNA was purified according to the manufacturer's recommendations, recovered by ethanol precipitation, and resuspended in 15 µl of nuclease-free water. Products were subjected to electrophoresis on a non-denaturing 10% slab gel (equilibrated and run in 1 X TBE), the 251 bp fragment was excised from the gel. The gel fragment was crushed, 400 µl of 0.3M NaCl in 1 X TE buffer was added, the mixture was incubated for 2 h at 37°C, the supernatant was collected using a Spin-X column. The elution procedure was repeated, supernatants were combined, and nucleic acids were recovered by isopropanol precipitation, washed twice with 80% cold ethanol, and resuspended in 15 µl of nuclease-free water. The library generated by this procedure is CP61T.

XACT-seq: high-throughput sequencing

Barcoded libraries were pooled and sequenced on an Illumina NextSeq500 platform in high-output mode using custom sequencing primer s1115.

XACT-seq: sample serial numbers

CP52, CP53 and CP54 are the samples used for the identification of RNAP-active-center A-site positions in active transcription complexes in vivo (- Rifampin). CP49, CP50 and CP51 are the samples used for the identification of RNAP-active-center A-site positions in static transcription complexes in vivo (+ Rifampin). Sample CP61T was used to identify template sequences present in the +14-20 library.

XACT-seq data analysis: ligation and crosslinking bias correction

To quantify crosslinking and ligation bias in the XACT-seq protocol, we reanalyzed the results of reference (16). In that study, XACT-seq was performed in the presence and absence of rifampin (Rif+ and Rif-, respectively) on a p*lac*CONS promoter library that contained an 11 bp variable region (*lac*VR) spanning positions +3 to +13 relative to the transcription start site (TSS).

We obtained Illumina read counts from three Rif+ samples (CP22, CP24, and CP28). For each 11-nt *lac*VR sequence s, each crosslinked position p within the larger p*lac*CONS promoter (defined as the 3'-most nucleotide of the primer extension product, with $p = 1$ denoting the first position of the *lac*VR), and each sample X, we denote this read count quantity by $\tilde{c}_s^{\mathrm{X},p,2020}$. We also obtained corresponding Illumina read counts, $\tilde{c}_s^{\mathrm{tem},2020}$, representing template abundance, i.e., the relative abundance of each DNA sequence s in the plasmid library (sample CP26T). From these counts we computed the marginal nucleotide counts of each base $b = \mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}$ at each position $q = 1, \ldots, 11$ within the *lac*VR:

$$\tilde{c}_{b,q}^{\mathrm{X},p,2020} = \sum_s \tilde{c}_s^{\mathrm{X},p,2020} s_{b,q},$$
$$\tilde{c}_{b,q}^{\mathrm{tem},2020} = \sum_s \tilde{c}_s^{\mathrm{tem},2020} s_{b,q}.$$

Here, $s_{b,q}$ is a one-hot encoding of sequence s, i.e., $s_{b,q} = 1$ if base b occurs at position q in sequence s, and $s_{b,q} = 0$ otherwise.

From these counts we computed a weight matrix that quantifies the relative $\log_2$ probability of crosslinking and ligation based on the DNA sequence in the vicinity of the crosslink position. Specifically, let $i = q - p$ denote the relative coordinate of position q in the *lac*VR with respect to the crosslink position p. The elements $w_{b,i}^{X,p}$ of this weight matrix were then computed as

$$w_{b,i}^{X,p} = \widetilde{w}_{b,i}^{X,p} - \frac{1}{4}\sum_{b'} \widetilde{w}_{b',i}^{X,p}, \qquad \widetilde{w}_{b,i}^{X,p} = \log_2\left[\frac{\tilde{c}_{b,p+i}^{\mathrm{X},p,2020} + 1}{\tilde{c}_{b,p+i}^{\mathrm{tem},2020} + 1}\right].$$

We restricted our attention to crosslink positions associated with the rifampin-inhibited initiation complex, namely, positions $p =$ +7 and +8 (which correspond to RNAP-active-center A-site positions +2 and +3, respectively). We reasoned that any sequence-dependence observed in our data at $p =$ +7 and +8 in the rifampin-inhibited complex would primarily reflect crosslinking and ligation bias.

At $p =$ +7 and +8, we found that the relative nucleotide positions $i = -2, -1, 0, +1, +2$ contributed substantially and reproducibly to the weight matrix elements $w_{b,i}^{X,p}$, whereas other relative nucleotide positions i did not. We therefore restricted the bias weight matrix to this range

of values for i. We then averaged together the six matrices (corresponding to p = +7, +8, and X = CP22, CP24, CP28) to generate a weight matrix quantifying both the crosslink and ligation bias:

$$w_{b,i}^{\text{xlink+lig}} = \frac{1}{6}\sum_{X}\sum_{p} w_{b,i}^{X,p}.$$

XACT-seq data analysis: template bias correction

To correct for template bias, we obtained Illumina read counts, $\tilde{c}_s^{\text{tem}}$, for the underlying plasmid library (sample CP61T). We then computed the marginal base counts at each nucleotide position $q = 1, \ldots, 7$ of the λPR' variable region (λVR; promoter positions +14 to +20):

$$\tilde{c}_{b,q}^{\text{tem}} = \sum_{s} \tilde{c}_s^{\text{tem}} s_{b,q}.$$

These counts were converted to a weight matrix that quantifies the relative $\log_2$ probability of each base at each position within the λVR:

$$w_{b,q}^{\text{tem}} = \widetilde{w}_{b,q}^{\text{tem}} - \frac{1}{4}\sum_{b'} \widetilde{w}_{b',q}^{\text{tem}}, \qquad \widetilde{w}_{b,q}^{\text{tem}} = \log_2\left[\tilde{c}_{b,q}^{\text{tem}} + 1\right].$$

XACT-seq data analysis: computation of reweighted sequence counts

To correct for bias in our XACT-seq data (both "xlink+lig" and "tem"), we reweighted the counts $c_s^{X,p}$ for each sequence s, sample X, and crosslink position p (with position $p = 1$ corresponding to the first position of λVR) using,

$$c_s^{X,p} = \tilde{c}_s^{X,p} \times \exp_2[-w_{s,p}^{\text{bias}}],$$

where

$$w_{s,p}^{\text{bias}} = \sum_{i=-2}^{2}\sum_{b} w_{b,i}^{\text{xlink+lig}} s_{b,p+i} + \sum_{q=1}^{7}\sum_{b} w_{b,q}^{\text{tem}} s_{b,q}$$

is the relative $\log_2$ probability of a sequence s being both present in the variant λPR' plasmid library and successfully crosslinked and ligated at position p in the corresponding rifampin-inhibited initiation complex.

A similar correction was also performed for XACT-seq data reported in reference (16) that was obtained in the absence of rifampin:

$$c_s^{X,p,2020} = \tilde{c}_s^{X,p,2020} \times \exp_2[-w_{s,p}^{\text{bias},2020}],$$

where

$$w_{s,p}^{\text{bias},2020} = \sum_{i=-2}^{2}\sum_{b} w_{b,i}^{\text{xlink+lig}} s_{b,p+i} + \sum_{q=1}^{11}\sum_{b} w_{b,q}^{\text{tem},2020} s_{b,q},$$

$$w_{b,p}^{\text{tem},2020} = \widetilde{w}_{b,p}^{\text{tem},2020} - \frac{1}{4}\sum_{b'} \widetilde{w}_{b',p}^{\text{tem},2020},$$

$$\widetilde{w}_{b,p}^{\text{tem},2020} = \log_2\left[\tilde{c}_{b,p}^{\text{tem},2020} + 1\right].$$

XACT-seq data analysis: sequence logo generation

All sequence logos in this manuscript were created using Logomaker (18) and quantify the relative $\log_2$ enrichment of each possible base at each position within a sequence. Specifically, to compute a sequence logo from the (potentially reweighted) counts c_s of a set of sequences s, we first computed the marginal base counts at each position q within the variable region,

$$c_{b,q} = \sum_s c_s \, s_{b,q}.$$

A logo was then generated in which the height $h_{b,q}$ of each base b at each position q within the logo was given by the centered $\log_2$ ratio

$$h_{b,q} = \tilde{h}_{b,q} - \frac{1}{4}\sum_{b'} \tilde{h}_{b',q}, \qquad \tilde{h}_{b,q} = \log_2\left[c_{b,q} + 1\right].$$

The logo for the σ-dependent paused complex (Figure 4C, top) was created by first generating three different logos, corresponding to position $p = +3$ (A site position +16 relative to the TSS) and Rif- samples $\mathrm{X} = \mathrm{CP52}, \mathrm{CP53}, \mathrm{CP54}$. The logo shown was then created by averaging these heights $h_{b,q}$ (for $q = 1,2,3,4$) across the three replicates. Fig. S5B displays the same information but for all seven randomized positions ($q = 1, \ldots, 7$). The logo for the initial-transcription paused complex (Figure 4C, center) was created by first generating three different logos, corresponding to position $p = +4$ (A site position +6 relative to the TSS) and Rif- samples $\mathrm{X} = \mathrm{CP21}, \mathrm{CP23}, \mathrm{CP27}$. The logo shown was then created by averaging these heights $h_{b,q}$ (for $q = 2,3,4,5$) across the three replicates. The logo for the elemental paused complex (Figure 4C, bottom), was created from pause sites identified by NET-seq and reported in reference (19), with each sequence s assigned a count $c_s = 1$ and the position of the A site corresponding to the RNA 3' end.

Cryo-EM structure determination: sample preparation

The σ-dependent paused transcription elongation complex was prepared by incubating 12 μM *E. coli* RNAP σ^{70} core enzyme with 22 μM nucleic-acid scaffold (Figure 6) in 60 μl transcription buffer (20 mM Tris-HCl, pH 7.9, 30 mM KCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol) for 15 min at 25°C, followed by adding 20 μl 120 μM *E. coli* [C541; P607]-σ^{70} and incubating 10 min at 25°C. The sample was concentrated to 50 μl using 0.5 ml Amicon Ultra 30 kDa MWCO centrifugal concentrator (Millipore), mixed with 6 μl 80 mM CHAPSO, and stored on ice prior to applying on grids. 3 μl samples were applied to QuantiFoil 1.2/1.3 Cu 300-mesh grids (glow-discharged for 50 s) using a Vitrobot Mark IV autoplunger (FEI), with the environmental chamber set at 22°C and 100% relative humidity. Grids were blotted with filter discs #595 (Ted Pella) for 8 sec at blotting force 4, and flash-frozen in liquid ethane cooled with liquid nitrogen; grids were stored in liquid nitrogen.

Cryo-EM structure determination: data collection and data processing

Cryo-EM data were collected at the University of Michigan Cryo-Electron Microscopy Facility, using a 300 kV Titan Krios G4i (Thermo Fisher Scientific) electron microscope equipped with a K3 direct electron detector and BioQuantum energy filter (Gatan). Data were collected automatically in counting mode using Leginon (20), a nominal magnification of 105,000x in EFTEM mode (actual magnification 595,238x), a calibrated pixel size of 0.84 Å per pixel, and a dose rate of 15 electrons/pixel/s. Movies were recorded at 100 ms/frame for 3 s (30 frames), resulting in a total radiation dose of 45 electrons/$Å^2$. Defocus range varied between -1.5 μm and -3.5 μm. A total of 11,611 micrographs were recorded from one grid over five days.

Micrographs were saved in Tiff format upon pre-processing for gain normalization and defect correction.

Data were processed as in Figure S6. Dose weighting and motion correction (5x5 tiles; b-factor = 150) were performed using Motioncor2 (21). CTF estimation was performed using CTFFIND4 (22). Subsequent image processing was performed using cryoSPARC (23). Automatic particle picking with blob picker yielded an initial set of 2,156,960 particles. Particles were binned 2x, extracted into 192 x 192-pixel boxes, and subjected to four rounds of reference-free 2D classification and removal of poorly populated classes, yielding a selected set of 921,867 particles. *Ab initio* models were generated, and heterogeneous refinement was performed for further classification of particles. One class, comprising particles with intact transcription complexes, was selected, and was subjected to homogeneous refinement, yielding a reconstruction with a global resolution of 3.8 Å as determined from gold-standard Fourier shell correlation.

The initial atomic model for protein components of the σ-dependent paused transcription elongation complex was built by manual docking of a cryo-EM structure of the *E. coli* Q21 transcription anti-termination loading complex Q21-QBE [PDB 6P18; (24)], with Q21 omitted, to the map in Chimera (25). Initial atomic models for DNA and RNA of the σ-dependent paused transcription elongation complex were built manually using Coot (26). The initial model of the of the σ-dependent paused transcription elongation complex was real-space rigid-body refined in Phenix (27) and subsequently refined with secondary-structure, geometry, Ramachandran, rotamer, Cβ, non-crystallographic-symmetry, and reference-model restraints. Molecular graphics representations were created using PyMOL and Chimera. EM density maps were visualized using PyMOL and Chimera.

Data and software availability

Sequencing reads have been deposited in the NIH/NCBI Sequence Read Archive under accession number PRJNA797396. Python analysis scripts and corresponding documentation are provided at http://www.github.com/jbkinney/21_nickels. The version of this repository used in the final publication is available on Zenodo (DOI:10.5281/zenodo.6359975).

The final atomic model and map of the σ-dependent paused transcription elongation complex have been deposited in the Protein Data Bank (PDB) and Electron Microscopy Data Bank (EMD) with accession codes 7N4E and EMD-24148, respectively.

SI Figures

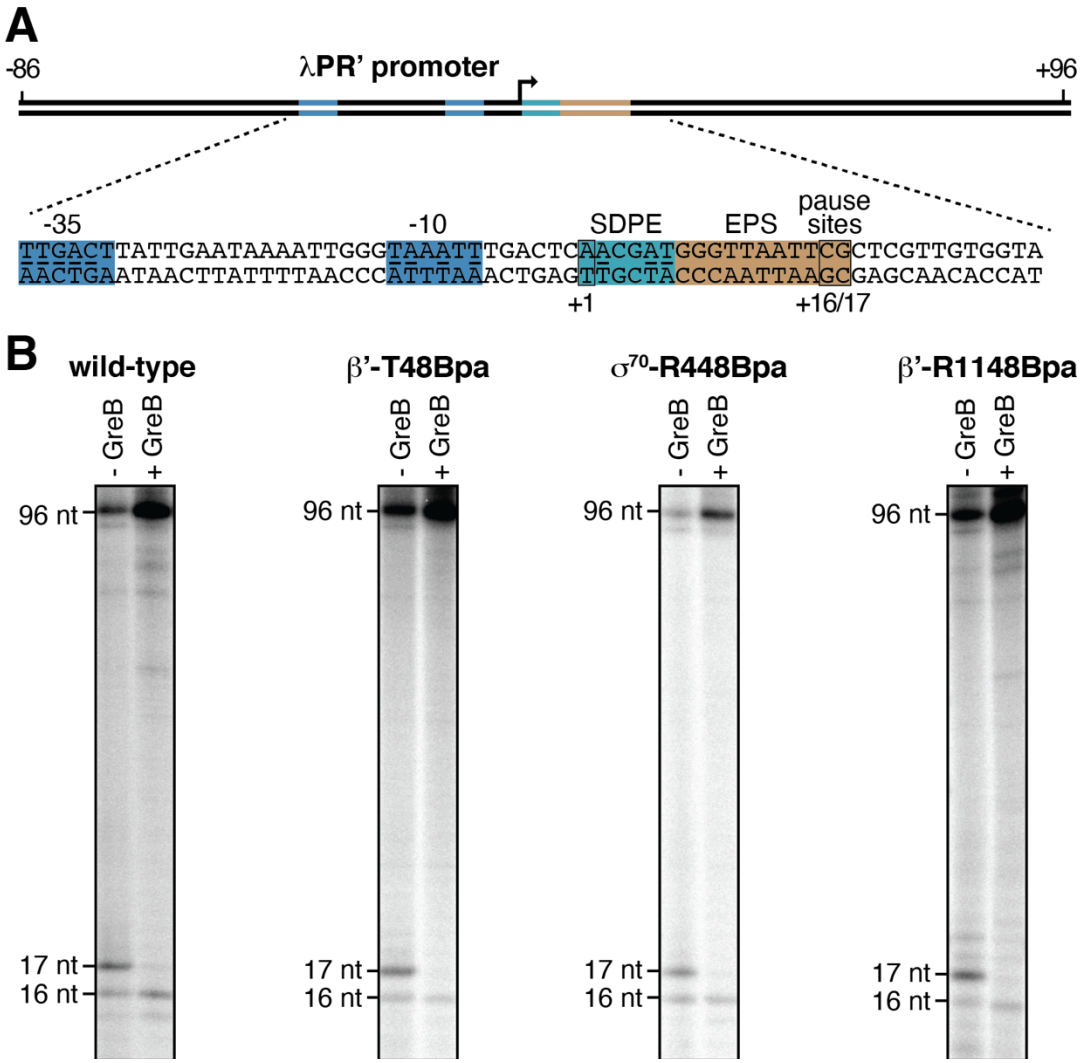

Figure S1. Use of site-specific protein-DNA photocrosslinking to define positions of RNAP trailing and leading edges and of σ relative to DNA at λPR': pausing properties of Bpa-labeled RNAP and σ derivatives.
(A) DNA template containing λPR' promoter. Colors as in Figure 1A.
(B) Gel images of PAGE analysis of RNA products for *in vitro* transcription reactions performed with wild-type RNAP, RNAP-β'^{T48Bpa} (Bpa at RNAP trailing edge), RNAP-σ$^{70\ R448Bpa}$ (Bpa in σR2), and RNAP-β'^{R1148Bpa} (Bpa at RNAP leading edge). Positions of 16- and 17-nt RNA products and 96-nt full-length RNA products are indicated.

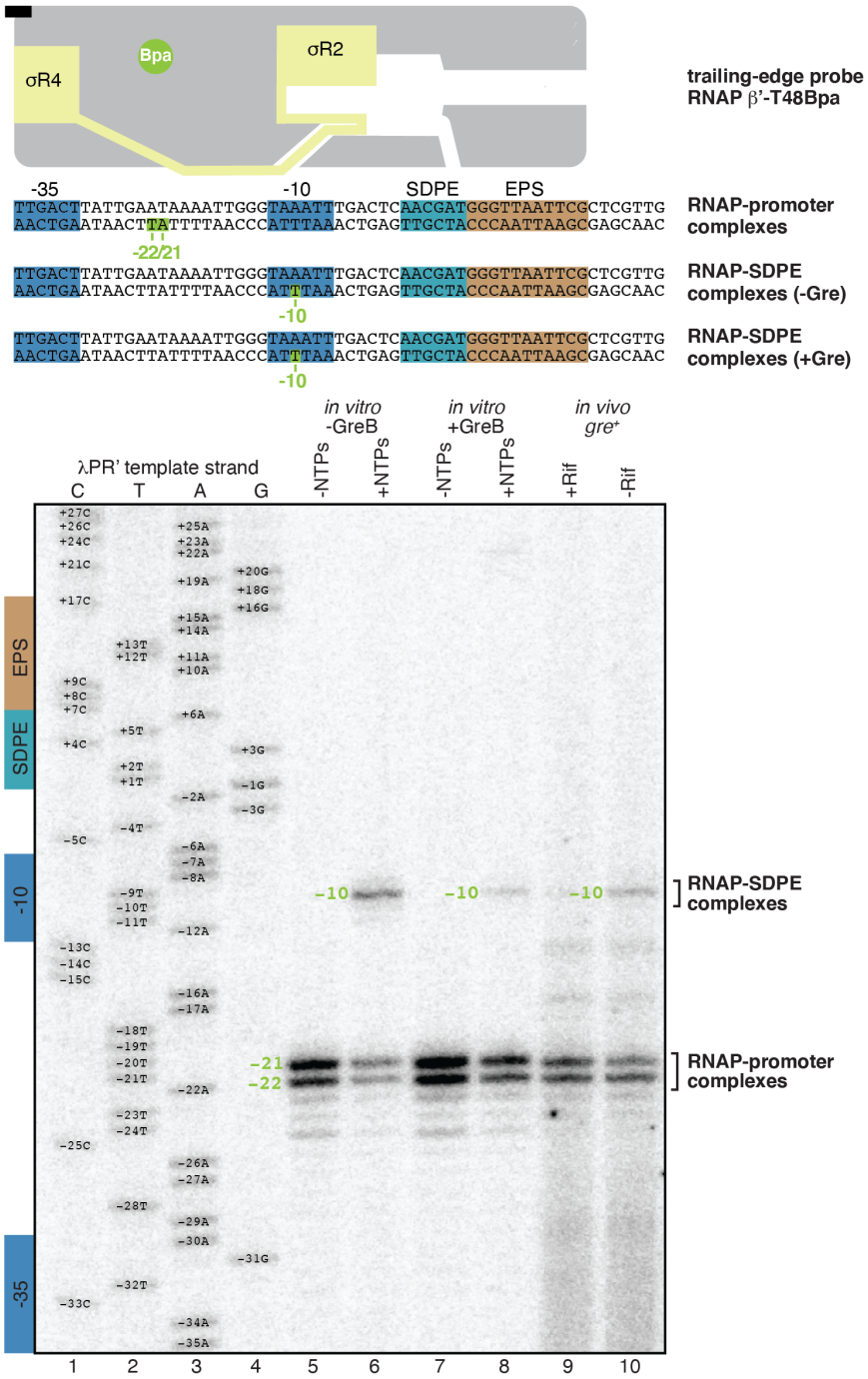

Figure S2. Use of site-specific protein-DNA photocrosslinking to define position of the RNAP trailing edge at λPR': results.

Top, λPR' promoter. Observed trailing-edge crosslinking sites are in olive green. Other colors as in Figure 1A. Bottom, positions of RNAP trailing edge in RNAP-promoter complexes or RNAP-SDPE complexes at λPR'. Figure shows sequence ladder generated using λPR' (lanes 1-4) and primer-extension mapping of crosslinking sites for each experimental condition *in vitro* and *in vivo*, identified at top (lanes 5-10).

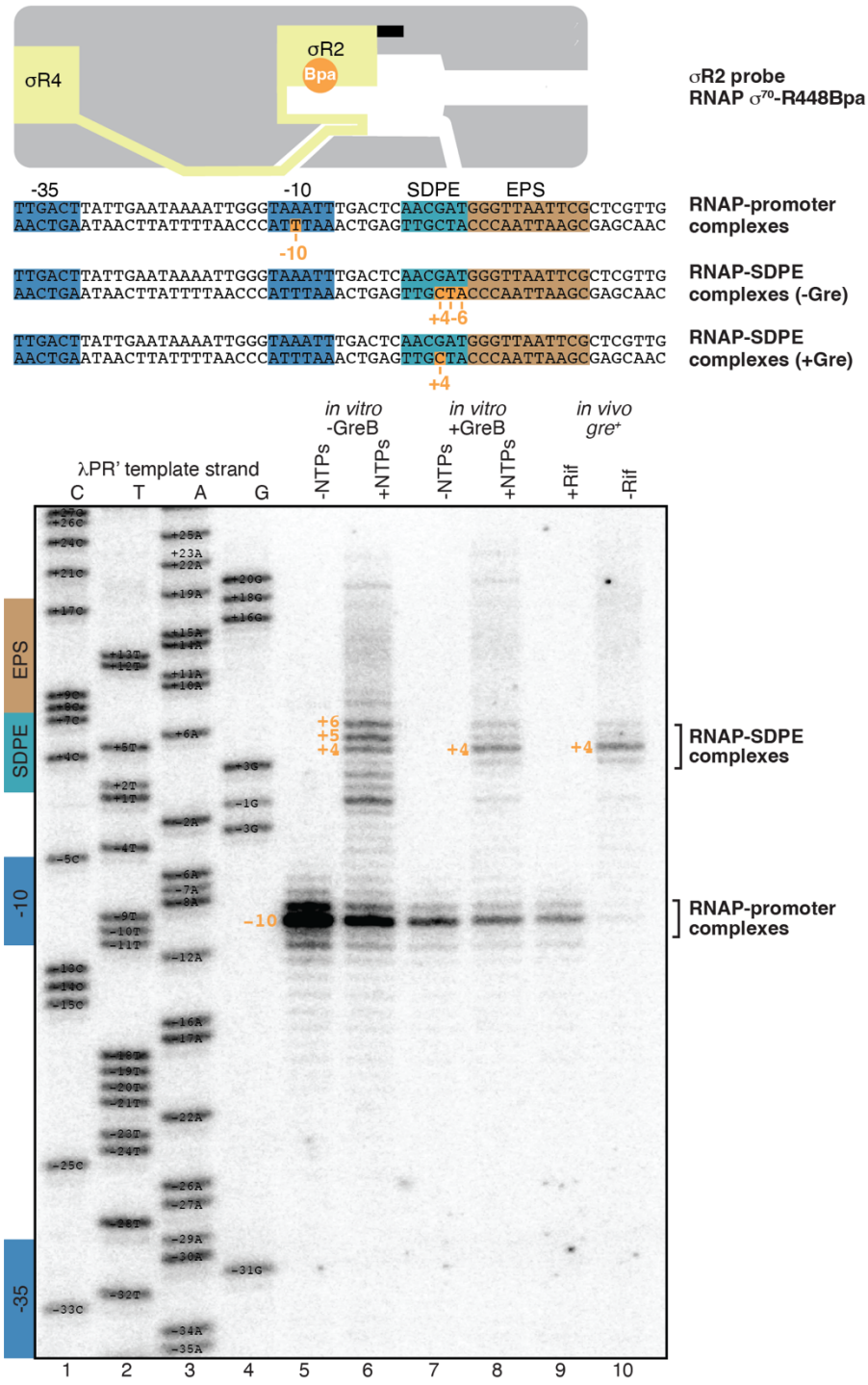

Figure S3. Use of site-specific protein-DNA photocrosslinking to define position of σR2 at λPR': results.

Top, λPR' promoter. Observed σR2 crosslinking sites are in orange. Other colors as in Figure 1A. Bottom, positions of σR2 in RNAP-promoter complexes or RNAP-SDPE complexes at λPR'. Figure shows sequence ladder generated using λPR' (lanes 1-4) and primer-extension mapping of crosslinking sites for each experimental condition *in vitro* and *in vivo*, identified at top (lanes 5-10).

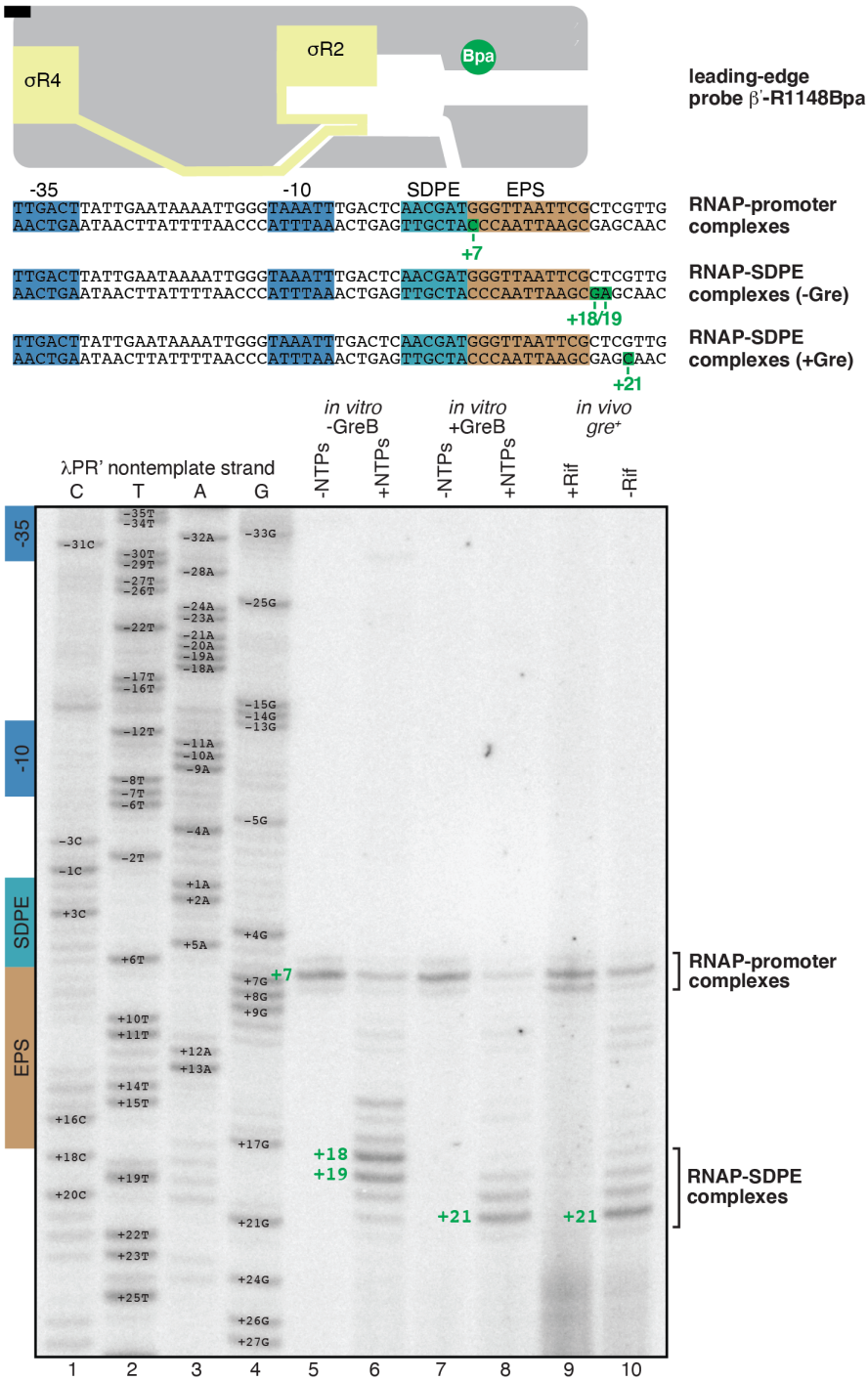

Figure S4. Use of site-specific protein-DNA photocrosslinking to define position of the RNAP leading edge at λPR': results.

Top, λPR' promoter. Observed leading-edge crosslinking sites are in forest green. Other colors as in Figure 1A. Bottom, positions of RNAP leading edge in RNAP-promoter complexes or RNAP-SDPE complexes at λPR'. Figure shows sequence ladder generated using λPR' (lanes 1-4) and primer-extension mapping of crosslinking sites for each experimental condition *in vitro* and *in vivo*, identified at top (lanes 5-10).

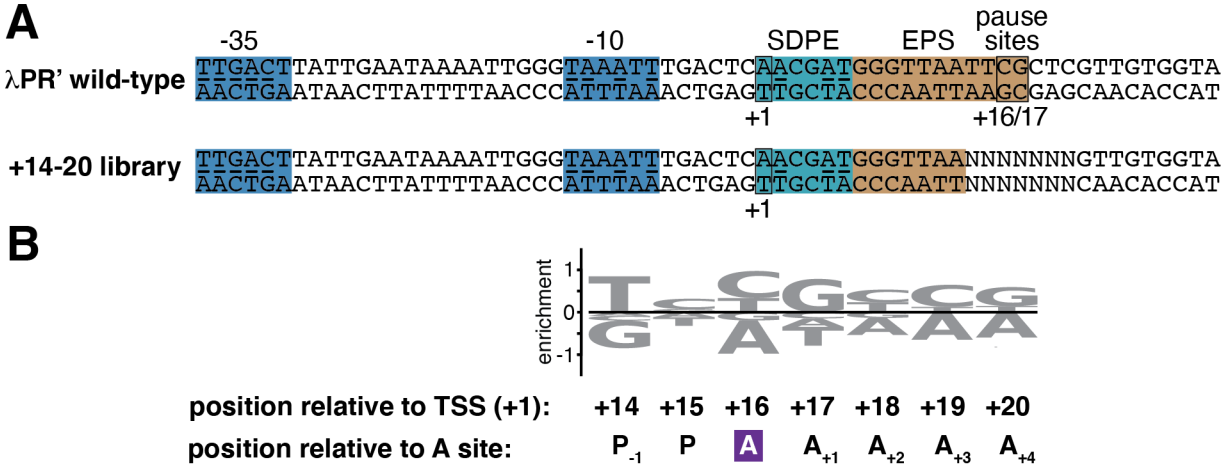

Figure S5. Sequence determinants for scrunching in σ-dependent pausing.
(A) DNA templates containing wild-type λPR' or +14-20 library. NNNNNNN, randomized nucleotides of +14-20 library. Other colors as in Figure 1A.
(B) Full sequence logo (positions +14 to +20) quantifying the formation and/or stability of major scrunched σ-dependent paused complex (RNAP-active-center A-site at position +16). The height of each base "X" at each position "Y" represents the $\log_2$ enrichment computed across sequences containing nontemplate strand X at position Y.

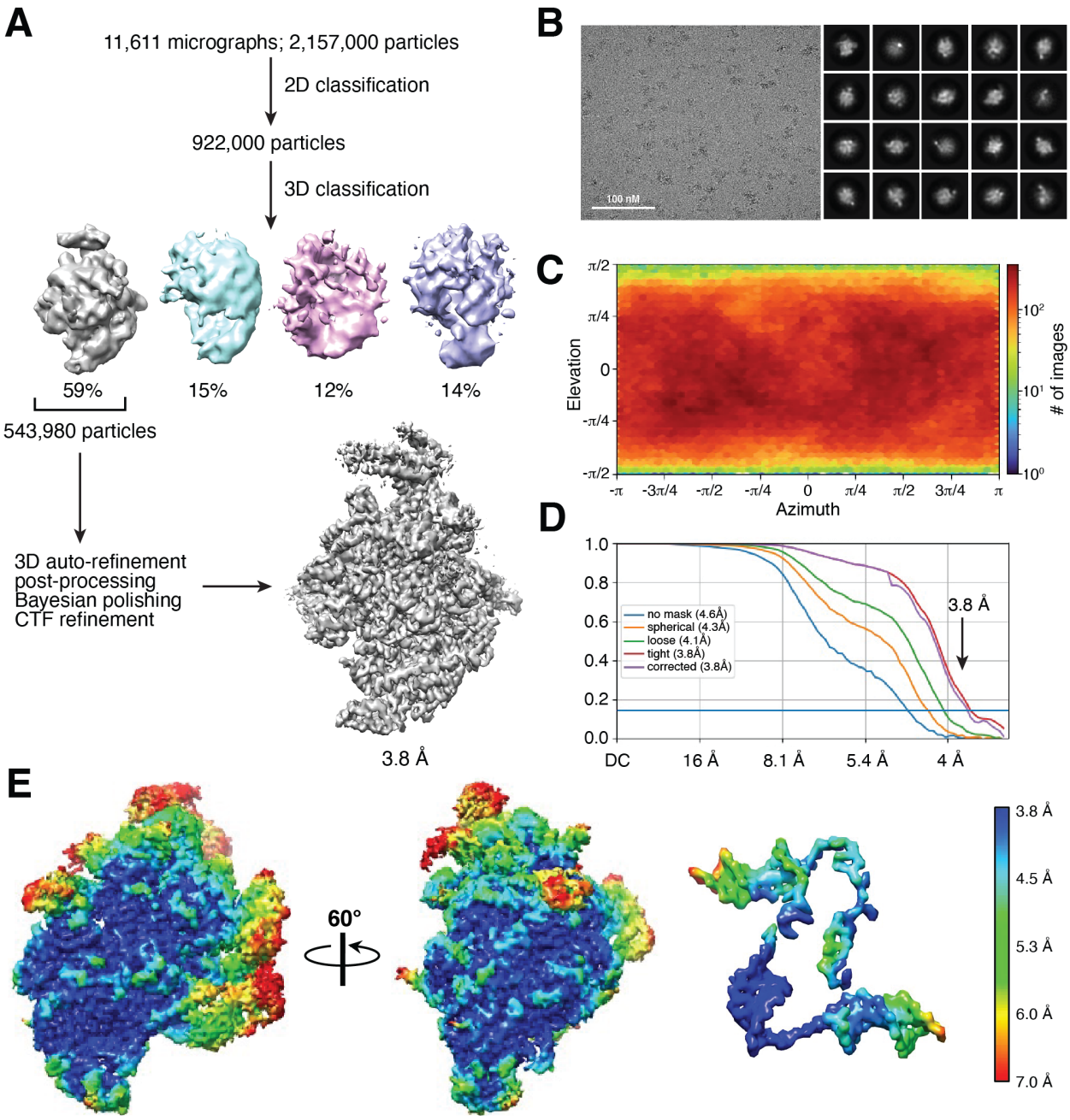

Figure S6. Structural basis for scrunching in σ-dependent pausing: cryo-EM structure determination.
(A) Data processing scheme.
(B) Representative electron micrograph (left; 100 nm scale bar) and representative class averages (right).
(C) Orientational distribution.
(D) Gold-standard Fourier shell correlation (GSFSC) resolution plot.
(E) EM density maps colored by local resolution. Left, overall structure (two views; orientations as in Figure 6B). Right, DNA and RNA in structure (view orientation as in Figure 6B, left).

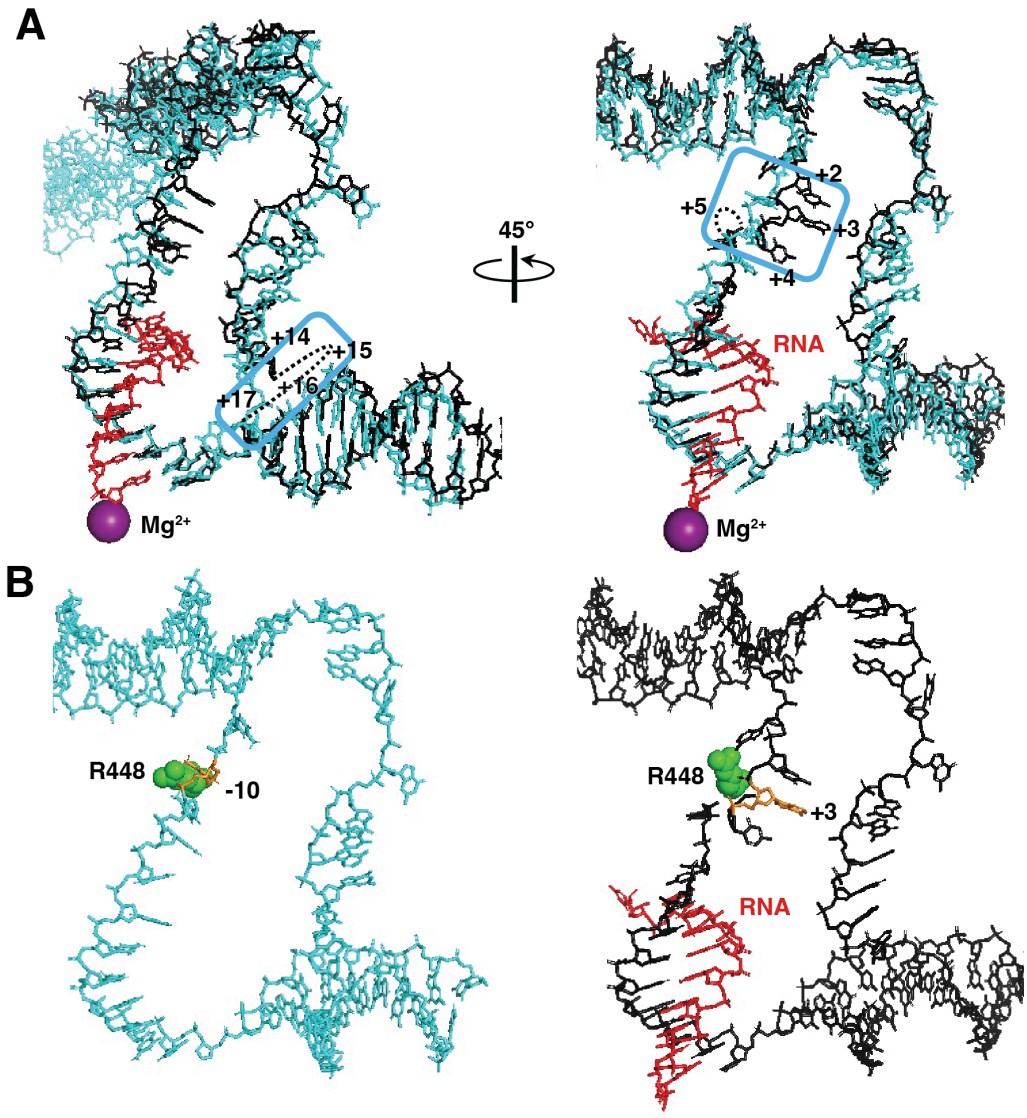

Figure S7. Structural basis for scrunching in σ-dependent pausing: scrunched nontemplate-strand and template-strand DNA nucleotides in pTEC.
(A) Superimposition of DNA (black), RNA (red), and RNAP-active-center Mg^{2+} (violet sphere) in structure of pTEC (3 bp of scrunching; Figure 6) on DNA (cyan) in structure of RPo (no scrunching; (28); PDB 5I2D) nontemplate-strand (left) and template-strand (right) DNA nucleotides in pTEC (view orientations as in Figure 6C). Blue boxes, DNA nucleotides disordered or repositioned due to DNA scrunching; black dots, DNA nucleotides disordered due to DNA scrunching. Nucleotides numbered as in λPR'.
(B) Comparison of relative positions of σR2 R448 and template strand of -10 element in RPo (left; (28); PDB 5I2D) to relative positions of σR2 R448 and template strand of SDPE in pTEC (right; Figure 6). Green, σR2 R448; orange, template-strand nucleotide at third position of -10 element (left, λPR' position -10) or template-strand nucleotide at third position of SDPE (right, λPR' position +3). View orientation and other colors as in panel A, left.

Table S1. Oligonucleotides.

name	sequence (5' to 3')	description
JW77	CTGTTCGATTGGGATGGCTATTCGG	oligo for primer-extension mapping of leading-edge crosslinks
JW119	GGGGATCCTCTAGATCCCAATCGAACAGGCC	forward primer for amplifying pBAD24
JW154	GGCCTGGAAGTTCTGTTCCAGGGGCCCCATATGGAGCAAAACCCGC	forward primer for amplifying pET28-*rpoD* for cloning *rpoD*
JW155	GATTGGGATCTAGAGGATCCCCGGGTACCTTAATCGTCCAGGAAGCTACG	reverse primer for amplifying pET28-*rpoD* for cloning *rpoD*
JW270	ATGGGGCCCCTGGAACAGAACTTCCAGGCCGCTGCTATGGTGATGGTGATGGTGATGGTGATGGTGGCTGCTGCCCATTGAATTCCTCCTGCTAGCCC	reverse primer for amplifying pBAD24
JW521	GGATCCGCCGCTGGGGCCTGTTCGATTGGGATGGCTATTCGG	PCR primer for production of linear DNA templates containing λPR' promoter
JW544	GTACCCTAGAGCCTGACCGGC	PCR primer for production of linear DNA templates containing λPR' promoter and oligo for primer-extension mapping of trailing-edge and σR2 crosslinks
JW615	GTTCAGAGTTCTACAGTCCGACGATCATGGCAACATATTAACGGCATGATA**TTGACT**TATTGAATAAAATTGGG**TAAATT**TGACTCAACGATGGGTTAANNNNNNNGTTGTGGTAGTGAGATGAAAAGAGGCGGCGCCTGCAGGNNNNNNNNNNNNNTGGAATTCTCGGGTGCCAAGG	template oligo containing λPR' sequence, a 7-nt randomized region (+14 to +20), and a 15-nt randomized region (promoter -35 and -10 elements are in bold)
JW679	CACGGATGGCAACATATTAACGGCATGATATTGACTTATTGAATAAAATTGGGTAAATTTGACTCAACGATGGGTTAAtTCGCTCGTTG	wild-type λPR' non-template strand
JW680	CTGTTCGATTGGGATGGCTATTCGGATCCCGCCGCCTCTTTTCATCTCACTACCACAACGAGCGAaTTAACCCATC	wild-type λPR' template strand
JW683	CACGGATGGCAACATATTAACGGCATGATATTGACTTATTGAATAAAATTGGGTAAATTTGACTCAACGATGGGTTAAgTCGCTCGTTG	T+14G λPR' non-template strand
JW687	CACGGATGGCAACATATTAACGGCATGATATTGACTTATTGAATAAAATTGGGTAAATTTGACTCAACGATGGGTTAA/idSp/TCGCTCGTTG	T+14X λPR' non-template strand
HV75	GCGGATCAGGCGTAGACCATCCGTATTC	Primer to incorporate TAG stop codon into plasmid $p\sigma^{70}$
s1219	TATAATGCCTGACCGGCGTTCAGAGTTCTACAGTCCGACGATC	oligo for amplifying and cloning pCDF-λPR′-N7 library
s1220	AATTAAGCCGCTGGGGCCCTTGGCACCCGAGAATTCC	oligo for amplifying and cloning pCDF-λPR′-N7 library
s128a	CCTTGGCACCCGAGAATTCCA	oligo for primer extension on +14-20 library templates
s1248	/5′Phos/NNNNNNNNNNGATCGTCGGACTGTAGAACTCTGAAC/3ddC/	3′ adapter with N10 at 5′ end (HPLC purified)
s1115	CTACACGTTCAGAGTTCTACAGTCCGACGATC	custom Illumina sequencing primer

LNA8	cgAatTaaCccAtcGtt	hybridization probe for detection of RNA transcripts *in vivo* (LNA bases are capitalized)

SI References

1. M. Leibman, A. Hochschild, A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *EMBO J* **26**, 1579-1590 (2007).
2. G. Panaghie, S. E. Aiyar, K. L. Bobb, R. S. Hayward, P. L. de Haseth, Aromatic amino acids in region 2.3 of *Escherichia coli* sigma 70 participate collectively in the formation of an RNA polymerase-promoter open complex. *J Mol Biol* **299**, 1217-1230 (2000).
3. B. E. Nickels *et al.*, The interaction between σ^{70} and the β-flap of *Escherichia coli* RNA polymerase inhibits extension of nascent RNA during early elongation. *Proc Natl Acad Sci U S A* **102**, 4488-4493 (2005).
4. J. T. Winkelman *et al.*, Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Mol Cell* **59**, 768-780 (2015).
5. J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc Natl Acad Sci U S A* **99**, 11020-11024 (2002).
6. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-4130 (1995).
7. L. Yu *et al.*, The mechanism of variability in transcription start site selection. *eLife* **6**, e32038 (2017).
8. I. O. Vvedenskaya *et al.*, Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol Cell* **60**, 953-965 (2015).
9. I. O. Vvedenskaya, S. R. Goldman, B. E. Nickels, Analysis of Bacterial Transcription by "Massively Systematic Transcript End Readout," MASTER. *Methods Enzymol* **612**, 269-302 (2018).
10. I. Artsimovitch, V. Svetlov, K. S. Murakami, R. Landick, Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* **278**, 12344-12355 (2003).
11. M. T. Marr, J. W. Roberts, Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* **276**, 1258-1260 (1997).
12. S. Borukhov, V. Sagitov, A. Goldfarb, Transcript cleavage factors from E. coli. *Cell* **72**, 459-466 (1993).
13. B. E. Nickels, J. Mukhopadhyay, S. J. Garrity, R. H. Ebright, A. Hochschild, The σ 70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. *Nat Struct Mol Biol* **11**, 544 (2004).
14. P. Deighan, C. Pukhrambam, B. E. Nickels, A. Hochschild, Initial transcribed region sequences influence the composition and functional properties of the bacterial elongation complex. *Genes Dev* **25**, 77-88 (2011).
15. S. R. Goldman, R. H. Ebright, B. E. Nickels, Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927-928 (2009).
16. J. T. Winkelman *et al.*, XACT-Seq Comprehensively Defines the Promoter-Position and Promoter-Sequence Determinants for Initial-Transcription Pausing. *Mol Cell* **79**, 797-811 e798 (2020).
17. C. Pukhrambam, I. O. Vvedenskaya, B. E. Nickels, XACT-seq: A photocrosslinking-based technique for detection of the RNA polymerase active-center position relative to DNA in *Escherichia coli*. *STAR Protoc* **2**, 100858 (2021).
18. A. Tareen, J. B. Kinney, Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272-2274 (2020).
19. I. O. Vvedenskaya *et al.*, Interactions between RNA polymerase and the “core recognition element” counteract pausing. *Science* **344**, 1285-1289 (2014).
20. C. Suloway *et al.*, Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60 (2005).

21. S. Q. Zheng *et al.*, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332 (2017).
22. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).
23. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
24. Z. Yin, J. T. Kaelber, R. H. Ebright, Structural basis of Q-dependent antitermination. *Proc Natl Acad Sci U S A* **116**, 18384-18390 (2019).
25. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
26. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
27. P. V. Afonine *et al.*, Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **66**, 1153-1163 (2010).
28. Y. Feng, Y. Zhang, R. H. Ebright, Structural basis of transcription activation. *Science* **352**, 1330-1333 (2016).