Figure S1: *rrnB* con characterization

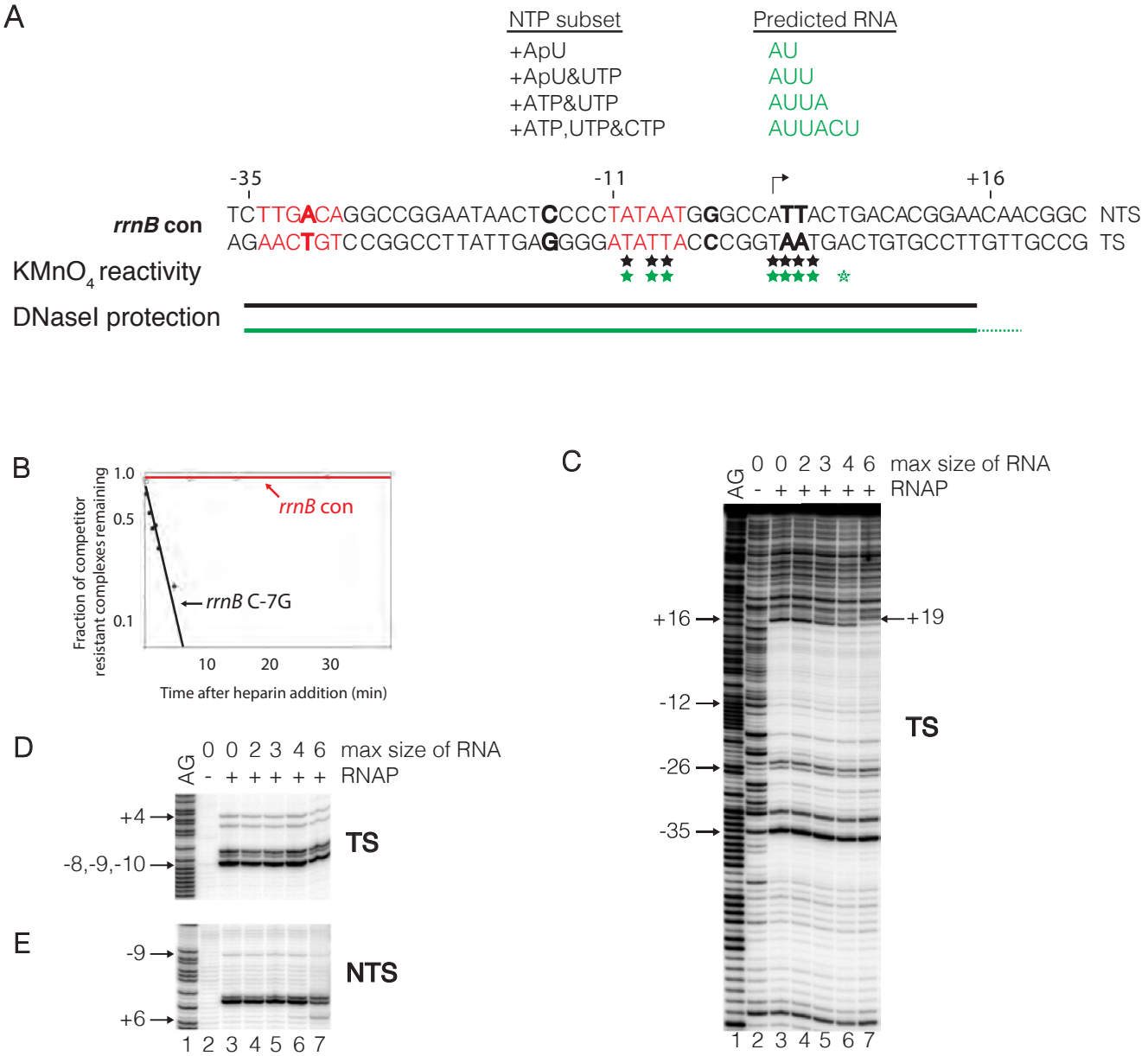
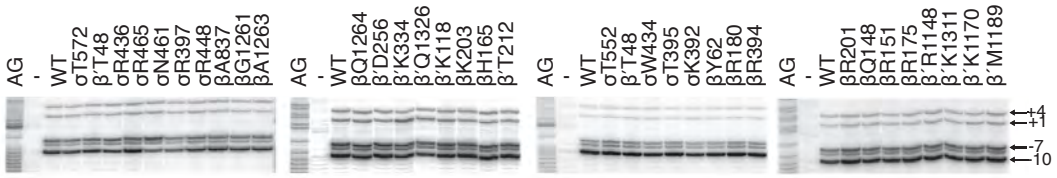


Figure S1. Characterization of *rrnB* con open and initial transcribing complexes, Related to Figure 2.

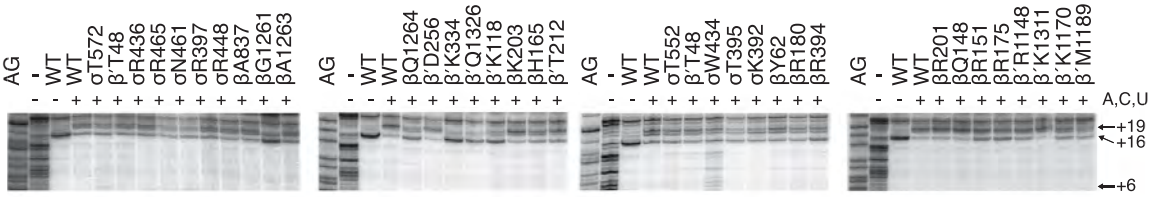
Panel (A): Sequence and footprinting summary of *rrnB* con, a derivative of *rrnB* P1 with 5 open complex-stabilizing mutations shown in bold. Predicted RNAs synthesized in the presence of subsets of nucleotides are shown above the promoter sequence. Lines below promoter sequence indicate region protected in DNaseI footprints in RP_O (black) or in the presence of ATP, CTP, and UTP (green). Asterisks indicate thymines reactive with $KMnO_4$ in RP_O (black) or in the presence of ATP, CTP, and UTP (green). (B) Lifetimes of RNAP-*rrnB* C-7G or RNAP-*rrnB* con complexes measured by in vitro transcription at times following competitor (heparin) addition. Under these solution conditions, the wild-type *rrnB* P1 complex half-life is too short to measure. (C) DNase I footprints, or (D, E) $KMnO_4$ footprints of wild-type RNAP-*rrnB* con complexes formed in the absence of NTPs (lanes 3), or in the presence of ApU (lanes 4), ApU and UTP (lanes 5), ATP and UTP (lanes 6), or ATP, CTP, and UTP (lanes 7). A+G sequencing ladders are in lanes 1, and controls without RNAP are in lanes 2. The template strand is 3' end-labeled in (C) and (D), and the non-template strand is 3' end-labeled in (E).

Figure S2: *rrnB* con footprinting and crosslink mapping

A KMnO$_4$ footprints



B DNaseI footprints



C Crosslink mapping

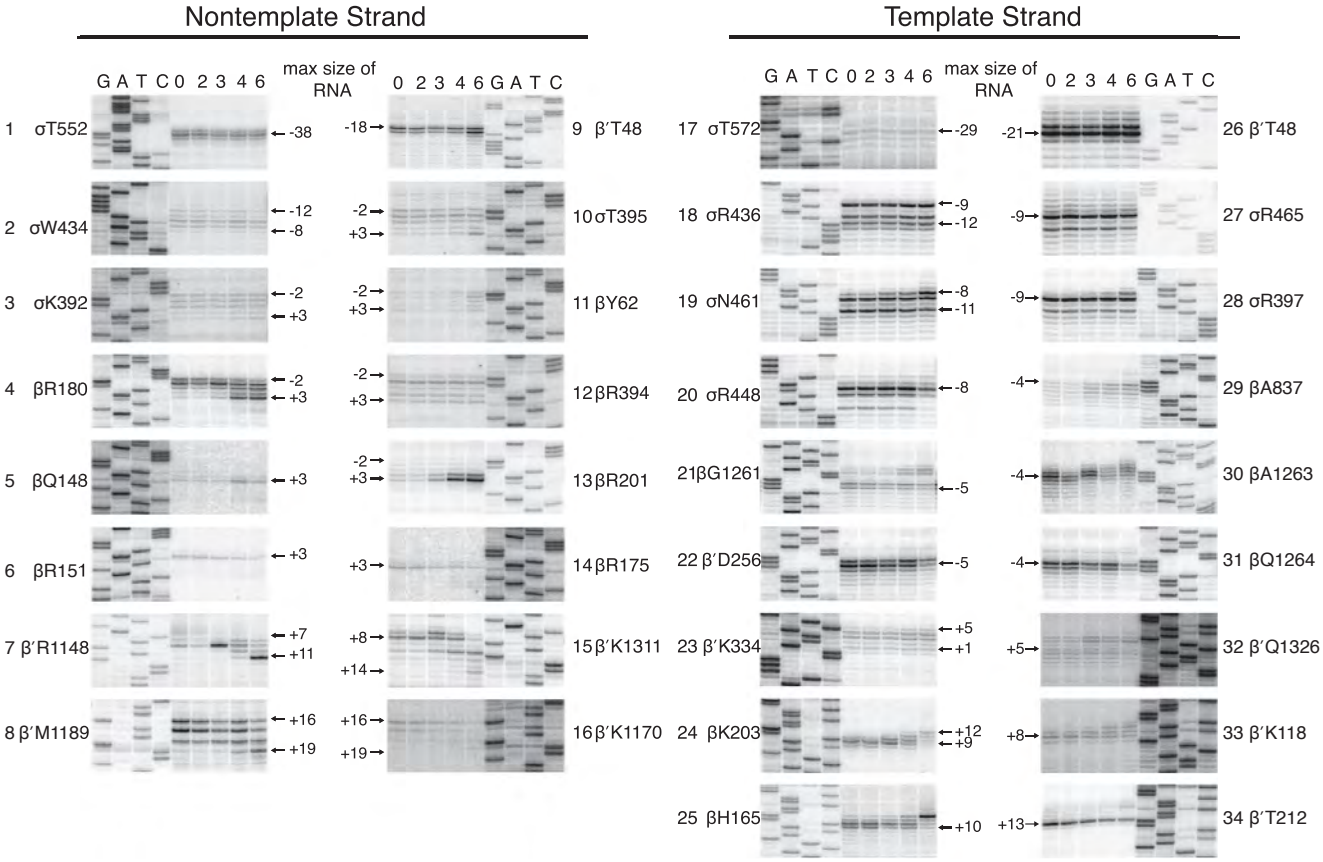


Figure S2. Footprinting and crosslink mapping of Bpa-containing RNAP-*rrnB* con complexes. Related to Figure 2. (A) $KMnO_4$ footprints of open complexes formed with wild-type or Bpa-containing RNAPs and *rrnB* con fragment 3' end-labeled in the template strand. $KMnO_4$ reactive thymines are indicated. (B) DNaseI footprints of initial transcribing complexes formed with wild-type or Bpa-containing RNAPs and *rrnB* con in the presence or absence of of ATP, CTP and UTP. Template strand is 3'-end labeled. Downstream footprint boundaries of +16 or +19 are indicated. (C) Primer extension crosslink mapping of complexes formed with Bpa-RNAPs that crosslink to the non-template strand (panels 1-16) or to the template strand (panels 17-34). The expected RNA product lengths are indicated above the lanes: complexes were formed in the absence of NTPs (labeled 0), in the presence of 500 μM ApU (labeled 2), in the presence of 500 μM ApU and 500 μM UTP (labeled 3), in the presence of 500 μM ATP and 500 μM UTP (labeled 4), or in the presence of ATP, UTP, and CTP (500 μM each). GATC sequencing reactions are shown for each gel. For orientation, primer extension products are indicated, with positions numbered as in RP_O.

Figure S3: Bpa-containing RNAPs form RP_O and RP_{ITC5} on *rrnB* C-7G

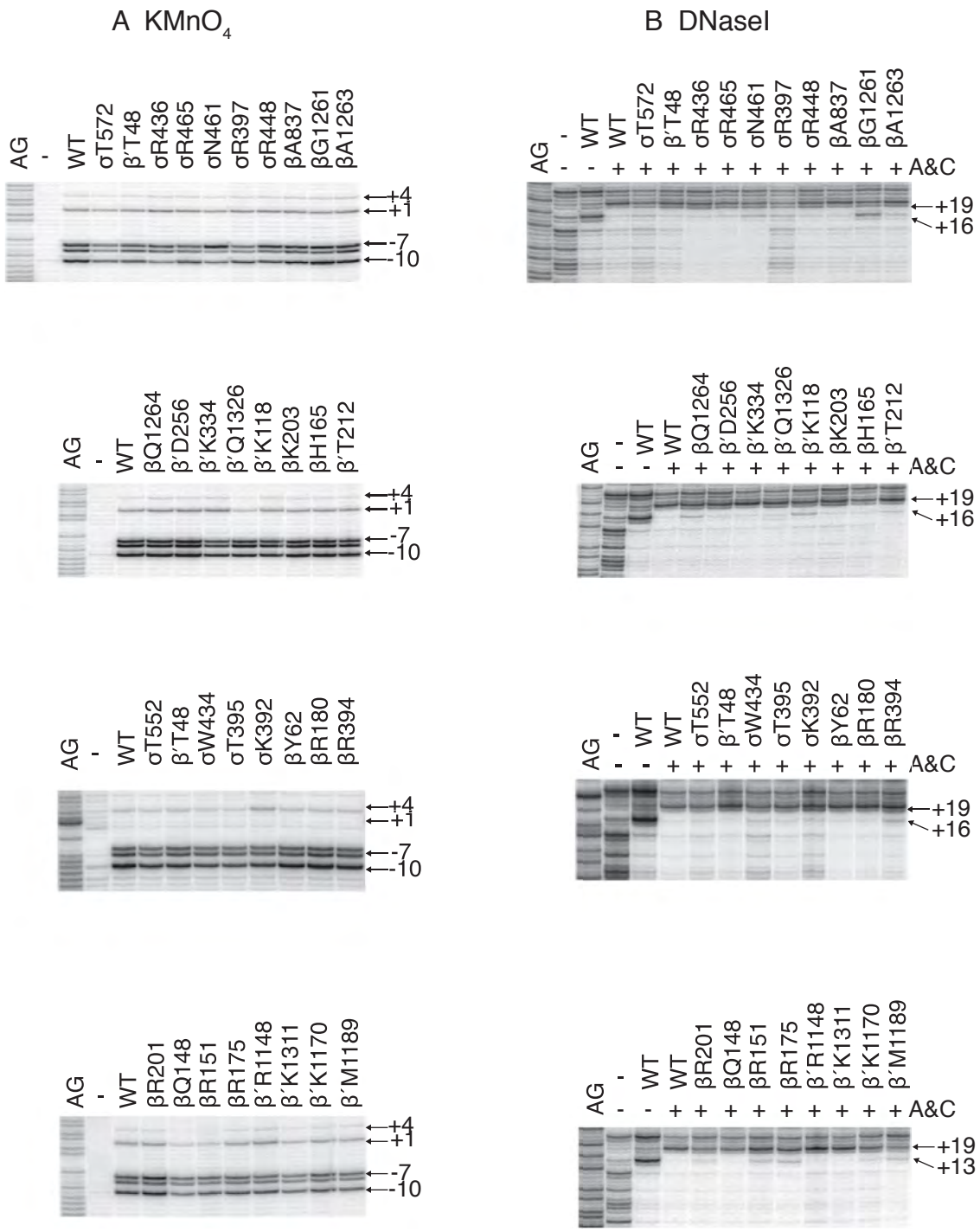


Figure S3. Bpa-RNAPs form RP_O and RP_{ITC5} on *rrnB* C-7G, Related to Figure 3. (A) $KMnO_4$ reactivity of Bpa-RNAP-*rrnB* C-7G complexes. DNA was 3'-end labeled in the template strand. (B) DNaseI footprints of initial transcribing complexes formed between Bpa-containing RNAPs and *rrnB* C-7G (template strand labeled). Complexes were formed in the presence of ATP (500 μM) and CTP (200 μM), except for control and wild-type RNAP lanes. Downstream boundary of protected region at +16 or +19 is indicated.

Figure S4: *rrnB* C-7G Crosslink Mapping

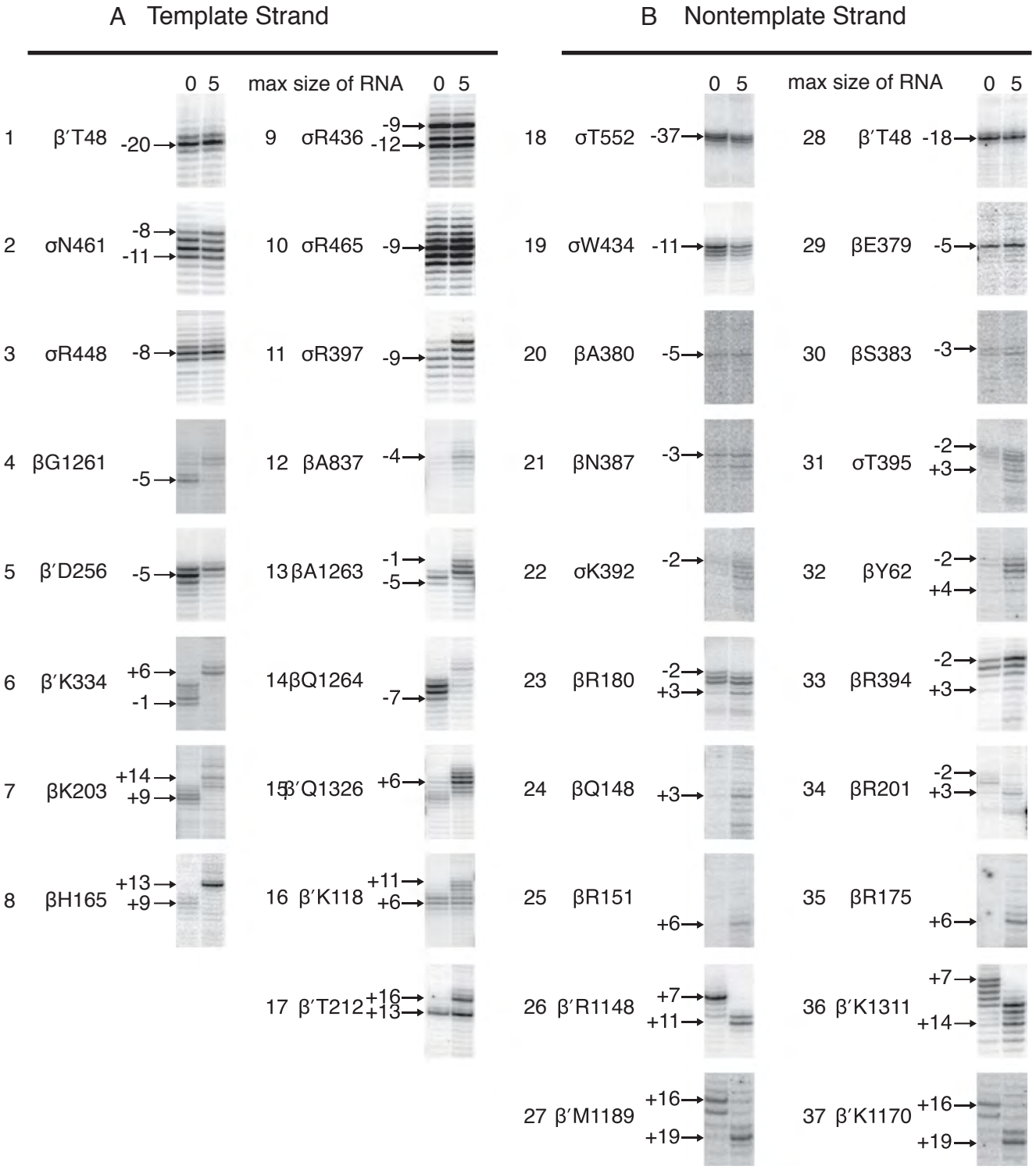


Figure S4. Crosslink mapping of RPo and RP_{ITC5} complexes formed by Bpa-RNAPs and the *rrnB* C-7G promoter, Related to Figures 4 and 5.

Primer extension mapping of the position of the crosslink formed in *rrnB* C-7G complexes by Bpa-RNAPs. Panels 1-17, non-template strand. Panels 18-27, template strand. Complexes were formed in the absence of NTPs (lanes labeled 0), or in the presence of 500 μM ATP and 200 μM CTP (lanes labeled 5). GATC sequencing ladders were run on each gel but are not shown. We note that the RNAPs containing Bpa at σK392, βQ148, βR151, and βR175 did not crosslink to DNA in the RP_O formed on *rrnB* C-7G, although they did crosslink to DNA in RP_O formed on *rrnB* con (see Figure S3).

Figure S5: Comparison of RP_O formed on *rrnB* con and *rrnB* C-7G

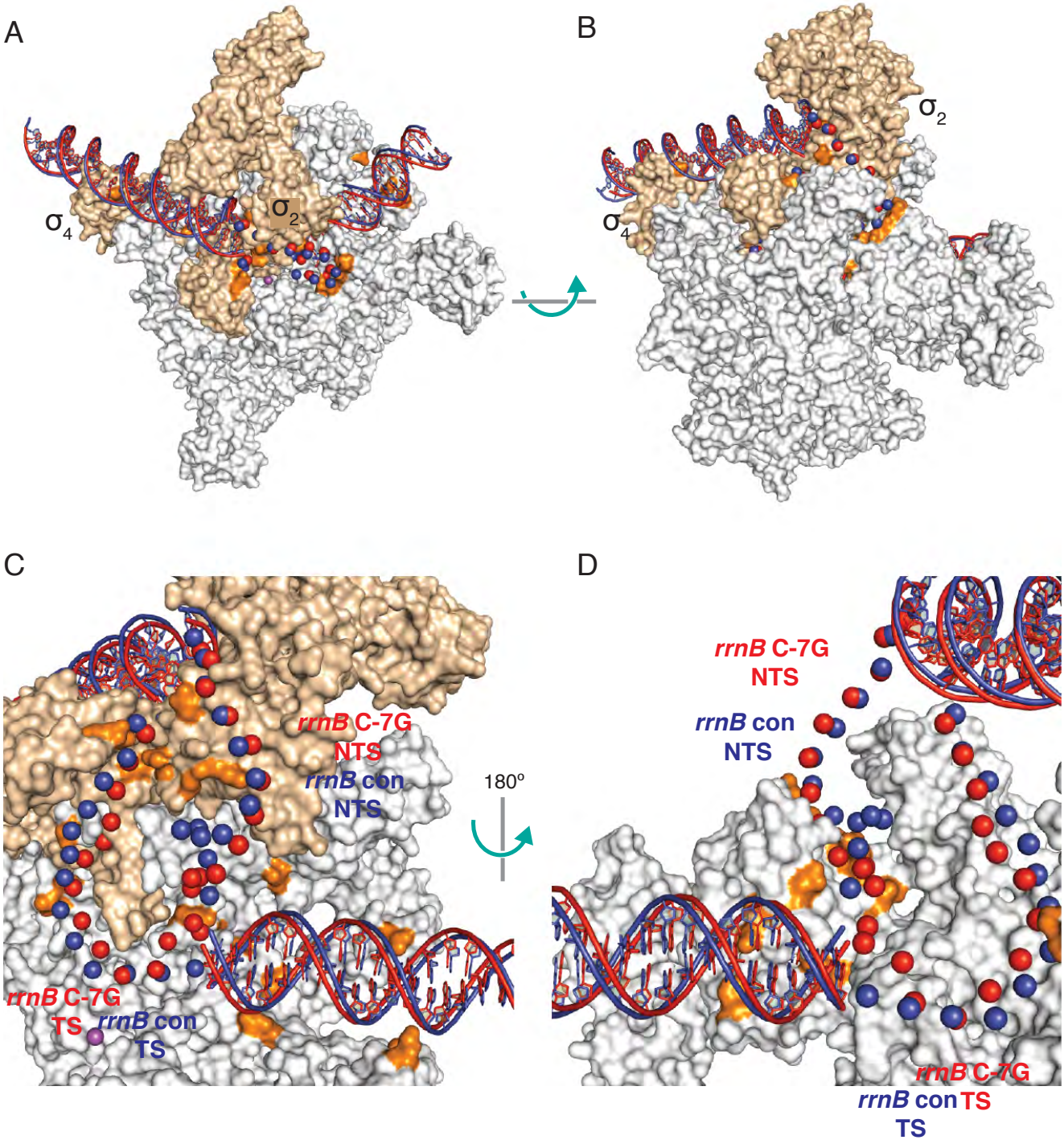


Figure S5. Comparison of the DNA path in the RP_O models of *rrnB* con and *rrnB* C-7G, Related to Figure 2. (A-B) Overlays of the path of DNA in models of RP_O complexes containing the *rrnB* con and *rrnB* C-7G promoters, based on the footprinting and crosslinking data in Figures S2 and S4 and summarized in Table S2. (B) is the same as (A) but rotated by ~90°. (C-D) Closeups showing the path of the single-stranded DNA bubble and double-stranded downstream DNA. DNA is blue in the *rrnB* con model and red in the *rrnB* C-7G model. Double-stranded regions are in ribbon, single-stranded regions are illustrated by spheres. The amino acid residues where Bpa substitutions resulted in crosslinks to DNA are in orange. In (C), β' and σ^{70} are shown, but most of β is eliminated for clarity. In (D), β is shown, but β' and σ are eliminated for clarity and the complex is rotated by 180° from the view shown in (C).

Table S1. Oligonucleotides used for stop codon mutagenesis, Related to Figure 1.

The first set was used for creating Bpa substitutions at the indicated positions in subunits of RNAP. Bpa substitutions that resulted in crosslinks to DNA are in red. Oligonucleotides used for PCR amplification and other purposes are also listed.

Table S2: Summary of crosslinking

		***rrnB* con RP_O**	***rrnB* C-7G RP_O**	***rrnB* C-7G RP_{ITC}**
Position of Bpa	**xlinked strand**	**Xlinked positions (% total)**	**Xlinked positions (% total)**	**Xlinked positions (% total)**
βY62	NT	Not detected	Not detected	-2(15),-1(30), +1(30),+2(15), +4(10)
βQ148	NT	+1(20), +2(30), +3(50)	Not detected	+3(100)
βR151	NT	+3(100)	Not detected	+6(100)
βH165	T	+10(50), +11(50)	+8(30), +9(40), +10(30)	+13(100)
βR175	NT	+3(100)	Not Detected	+5(20),+6(80)
βR180	NT	-2(30),-1(45), +1(25)	-2(25), -1(35), +1(40)	-2(15),-1(30),+1(35),+3(20)
βR201	NT	-2(15),-1(15), +1(15), +2(28), +3(27)	-2(20),-1(20),+1(50),	-2(10), -1(10),+1(15), +2(15), +3(50)
βK203	T	+9(25), +10(50), +11(25)	+8(20), +9(40), +10(40)	+12(30), +13(10), +14(30), +16(30)
βE379	NT	Not tested	-5 (100)	-5 (100)
βA380	NT	Not tested	-5(100)	-5(100)
βS383	NT	Not tested	-3(50), -2(50)	-3(40), -2(50),+1(10)
βN387	NT	Not tested	-3(70), -2(10),-1(20)	-3(50), -1(40),+1(5),+2(5)
βR394	NT	-2(13),-1(35), +1(12), +2(20), +3(20)	-2(30),-1(40),+1(30)	-2(40),-1(30),+1(20),+3(10)
βA837	T	Not detected	Not detected	-6(23), -5(23), -4(30), -3(23)
βG1261	T	-5(50), -4(50)	-5(60), -4(40)	-2(50), -1(50)
βA1263	T	-5(25), -4(50), -3(25)	-5(25), -4(50), -3(25)	-4(15), -3(45), -2(25), -1(15)
βQ1264	T	-5(40), -4(60)	-7(10), -6(40), -5(40), -4(10)	Not detected

β'T48	T/NT	T -21(75), -20(25): NT -18(60), -17(40)	T -21(25), -20(50), -19(25); NT -19(15), -18(70), -17(15)	T -21(25), -20(50), -19(25); NT -19(20), -18(70), -17(10)
β'K118	T	+7(20), +8(50), +9(30)	+6(50), +7(50)	+6(20), +7(20), +9(20), +10(20), +11(20)
β'T212	T	+13(100)	+13(100)	+13(50), +16(50)
β'D256	T	-6(20), -5(50), -4(30)	-6(25), -5(50), -4(25)	-5(20), -4(60), -3(20)
β'K334	T	-1(6), +1(8), +2(8), +3(35), +4(35), +5(8)	-1(30), +1(30), +2(5), +3(5), +4(30)	+6(60), +7(40)
β'R1148	NT	+7(33), +8(34), +9(33)	+7(70), +8(10), +9(10), +10(10)	+10(50), +11(50)
β'K1170	NT	+16(65), +17(35)	+16(50), +17(50)	+18(50), +19(50)
β'M1189	NT	+16(40), +17(40), +18(20)	+16(50), +17(50)	+18(20), +19(80)
β'K1311	NT	+7(40), +8(50), +10(10)	+7(20), +8 (20), +9(20), +10, (20), +11(20)	+11(25), +12(20), +13(15), +14(20), +15(10), +16(10)
β'Q1326	T	+1(13),+2(13),+3(13), +4(13),+5(14),+6(13),+7(13)	+1(13),+2(13),+3(13), +4(13),+5(14),+6(13),+7(13)	+4(11), +5(11), +6(30), +7(25), +8(15), +9(8)
σK392	NT	-2(25), -1(25), +1(25), +2(12), +3(13)	Not detected	-2(12), -1(25), +1(13), +2(25), +4(25)
σT395	NT	-2(30), -1(30), +1(25), +2(8),+3(7)	-2(30), -1(30), +1(30),+2(5),+3(5)	-2(10), -1(18), +1(18), +2(10), +3(16), +4(10), +5(16)
σR397	T	-11(20), -9(80)	-11(10), -10(25) -9(35), -8(25), -7(5)	-11(5), -10(10), -9(25), -8(25),-7(35)
σW434	NT	-12(20),-11(20),-10(30),-9(20),-8(10)	-11(40), -10(30), -9(30)	-11(40), -10(30), -9(30)
σR436	T	-13(10), -12(20), -11(20), -9(50)	-12(30), -11(30), -9(40)	-12(30), -11(30), -9(40)
σR448	T	-11(15), -9(30), -8(45),-7(10)	-9(25), -8(50), -7(25)	-9(25), -8(50), -7(25)
σN461	T	-11(40), -9(40), -8(20)	-11(15), -10(35), -9(35), -8(5)	-11(30), -10(35), -9(15), -8(20)
σR465	T	-11(35), -9(65)	-11(15), -10(15), -9(55), -8(15)	-11(15), -10(15), -9(55), -8(15)
σT552	NT	-39(30),-38(40), -37(30)	-38(30), -37(40), -36(30)	-38(30), -37(40), -36(30)
σT572	T	-31(20), -30(20), -29(60)	Not detected	Not detected

Table S2. Crosslink efficiencies for mapped crosslink positions in RP_O and RP_{ITC5}, Related to Figures 2,4, and 5.

NT, non-template strand. T, template strand. The numbers in parentheses represent the approximate percentage of the total crosslinked population that crosslinked to the indicated position in the DNA. These numbers were used to generate the percentages of crosslinked DNA positions illustrated in Figure 5.

Table S3: Comparison of RP_O crosslinking data to RP_O structure from Zhang et. al. 2012

	***rrnB* con RP_O crosslinking data**		***rrnB* C-7G RP_O crosslinking data**		**Crystallography data**	
***E. coli* residue**	**strand xlinked**	**Xlinked positions**	**strand xlinked**	**Xlinked positions**	**strand**	**Positions within 10 angstroms**
βY62	-	-	-	-	-	None
βQ148	NT	+1,+2,+3	-	-	-	None
βR151	NT	+3	-	-	NT	+1,+2,+3
βH165	T	+10,+11	T	+8,+9,+10	T	+9,+10,+11
βR175	NT	+3	-	-	NT	+1,+2,+3
βR180	NT	-2,-1,+1	NT	-2,-1,+1	NT	+1
βR201	NT	-2,-1,+1,+2,+3	NT	-2,-1,+1	NT	-4,-3,-2,-1,+1
βK203	T	+9,+10,+11	T	+8,+9,+10	T	+8,+9
βE379	-	Not tested	NT	-5	-	None
βA380	-	Not tested	NT	-5	NT	-6,-5,-4,-3,-2
βS383	-	Not tested	NT	-3,-2	NT	-3,-2
βN387	-	Not tested	NT	-3,-1	NT	-3,-2
βR394	NT	-2,-1, +1, +2, +3	NT	-2,-1,+1,+2,+3	NT	-3,-2
βA837	-	-	-	-	None	None
βG1261	T	-5,-4	T	-5,-4	T	-4,-3
βA1263	T	-5,-4,-3	T	-5,-4,-3	T	None
βQ1264	T	-5,-4	T	-7,-6,-5,-4	T	-4,-3
β′T48	T	-21,-20	T	-21,-20,-19	-	None
β′T48	NT	-18,-17	NT	-19,-18,-17	-	None
β′K118	T	+7, +8, +9	T	+6,+7	T	+5,+6,+7
β′T212	T	+13	T	+13	-	None
β′D256	T	-6,-5,-4	T	-6,-5,-4	-	None
β′K334	T	-1, +1, +2, +3, +4, +5	T	-1,+1,+4	T	+1,+2,+3
β′R1148	NT	+7, +8, +9	NT	+7	NT	+5,+6,+7
β′K1170	NT	+16,+17	NT	+16,+17	-	None
β′M1189	NT	+16,+17,+18	NT	+16,+17	-	None
β′K1311	NT	+7, +8, +10	NT	+7,+8,+9,+10,+11	NT	+6,+7,+8,+9
β′Q1326	T	+1,+2,+3, +4,+5,+6,+7	T	+1,+2,+3,+4,+5,+6	T	+3,+4,+5,+6
σK392	NT	-2, -1, +1, +2, +3	-	-	NT	-8,-7,-6,-5,-4
σT395	NT	-2,-1,+1,+2,+3	NT	-2,-1,+1	NT	-5,-4
σR397	T	-11,-9	T	-11,-10,-9,-8	-	None
σW434	NT	-12,-11,-10,-9,-8	NT	-11,-10,-9	NT	-12,-11
σR436	T	-13, -12, -11, -9	T	-12,-11,-9	-	None
σR448	T	-11,-9,-8,-7	T	-9,-8,-7	-	None
σN461	T	-11,-9,-8	T	-11,-10,-9,-8	-	None
σR465	T	-11,-9	T	-11,-10,-9,-8	-	None
σT552	NT	-39,-38,-37	NT	-38,-37,-36	-	None
σT572	T	-31,-30,-29	-	-	-	None

Table S3. Comparison of RP_O crosslinking results to DNA-protein interactions in X-ray structure, Related to Figure 2. Positions that crosslinked to individual Bpa residues in the RP_O complexes formed by *rrnB* con and *rrnB* C-7G were compared to the DNA positions in closest proximity to the corresponding wild-type amino acid residues in the downstream fork-junction crystal structure (Zhang et al., 2012). Crosslinking data for the *rrnB* con and *rrnB* C-7G complexes are in columns 2-5. NT, non-template strand, TS, template strand. DNA positions within 10 angstroms of the corresponding wild-type amino acid residues are in columns 6-7. The *E. coli* RNAP (PDB 4LJZ) and *T. thermophilus* RNAP structures with downstream fork-junction DNA (PDB 4G7Z) were aligned using Pymol, and then the DNA positions were identified that were within 10 angstroms of the amino acid residue corresponding to each Bpa residue that crosslinked to DNA (column 7). Since the crystal structure did not contain DNA downstream of +12, template strand DNA upstream of -4, or non-template strand DNA upstream of -12, for some amino acid residues there were no promoter positions in the proximity of the amino acid where Bpa substitutions resulted in crosslinks.

Supplemental Experimental Procedures

Plasmids for overexpression of WT or Bpa-containing RNAPs

Plasmids used to overexpress WT or Bpa containing core RNAP were derived from pIA299 (encoding α,β, β' with His_6 on the C-terminus of β'; (Artsimovitch et al., 2003)or pIA900 (encoding α,β,β',and ω with His_{10} on the C-terminus of β' (Svetlov and Artsimovitch, 2015). Plasmids used to overexpress WT or Bpa-containing σ^{70} proteins were derived from pRLG13105 (encoding σ^{70} with and His_{10} at the N-terminus and a PreScission protease cleavage site between the His tag and σ^{70} sequence). pRLG13105 is a pBAD24 derivative that was constructed in several steps. First, DNA encoding an N-terminal His_{10} tagged σ^{70} was amplified from pRLG12073 using primers 6307 and 3355, digested with PciI and HindIII, and ligated into pBAD24 that had been digested with NcoI and HindIII. This generated plasmid pRLG13101. An out of frame translation start site 8 bp downstream from the Shine–Dalgarno sequence in pRLG13101 was created with primer 6380 to generate pRLG13105, which expressed Bpa-containing σ^{70}.

Stop codons (TAG) were introduced into *rpoB*, *rpoC*, or *rpoD* at the position chosen for Bpa incorporation (see Table S1 for primers). Mutagenesis was performed using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent). Positions chosen were based on the structure of *E. coli* RNAP (Zuo et al., 2013), i.e. solvent-exposed residues in σ regions 2, 3, or 4 close to promoter DNA or residues in β or β' lining the main channel. Transformants were streaked and restreaked for single colonies, and the overexpression plasmids were purified and sequenced to verify the identity of the mutation.

Incorporation of Bpa and protein expression

For protein purification, cotransformation of the overexpression plasmid and the tRNA/tRNA synthetase plasmid into the host strain (BL21(DE3) for pIA299 or pIA900; DH10B for pRLG13105) was performed fresh for each experiment. Plasmids were cotransformed by electroporation, selecting for both ampicillin and chloramphenicol resistance. Fresh transformants were scraped from plates for use as an inoculum, generating a relatively high starting culture density (OD_{600} ~0.3), and grown at 30°(for σ^{70} overexpression) or 37°C (for core RNAP overexpression) in LB with Bpa (1 mM), ampicillin (100 μg/ml), and chloramphenicol (25 μg/ml). Using a large inoculum from plates avoided suppressor accumulation from extended growth in liquid culture. The culture medium was prepared by addition of Bpa to LB medium dropwise from a freshly made 100 mM Bpa stock in 1 M NaOH. 1 M HCl was added to produce a final pH of ~7.2. After 1 hr of growth with Bpa, expression was induced with 1 mM IPTG or 0.2% L-arabinose, depending on the expression system. Core RNAP overexpressing cells were grown in the dark at 37°C for 6-20 hr after induction, and σ^{70}-overexpressing cells were grown in the dark at 30° for 1-1.5 hr after induction. Cells were harvested by centrifugation and stored at -20° for up to one week.

Purification of Bpa-containing overexpressed core RNAP and σ^{70}

Core RNAPs with a His_6 or His_{10} tag at the C-terminus of the β′ subunit were purified using Ni-agarose and heparin affinity chromatography sequentially. Cell pellets harvested from 250 ml cultures were suspended in 5 ml of BugBuster (Novagen), phenylmethylsulfonyl fluoride (PMSF) to a final concentration of 23 μg/ml, and 5 μl Lysonase (Novagen). Resuspended pellets were incubated at room temperature with

gentle rocking for 30 min before adding 15 ml of RNAP resuspension buffer (1.4 M NaCl, 40 mM Tris-Cl pH 8.0, 30 mM imidazole), followed by centrifugation at 14,000 rpm for 40 min at 4°C. The cleared lysate was then added to 0.5 ml pre-equilibrated Ni resin, the column was washed with RNAP wash buffer (300 mM NaCl, 40 mM Tris-Cl pH 8.0, 30 mM imidazole), and the protein was eluted with wash buffer containing 300 mM imidazole. The eluate was diluted to 200 mM NaCl with TGED (10 mM Tris-Cl pH 8.0, 5% glycerol, 0.1 mM EDTA, 0.1 mM DTT; (Burgess and Jendrisak, 1975)and 100 mM NaCl, loaded onto 0.4 ml of heparin resin column that had been pre-equilibrated in TGED plus 200 mM NaCl. The column was washed with 5 ml of TGED plus 200 mM NaCl, and RNAP was eluted with 1 ml of TGED plus 600 mM NaCl. RNAPs were concentrated using 5 ml Microcon centrifugal filtration units with a 100 kDa molecular weight cutoff, and exchanged with 2X storage buffer without glycerol (20 mM Tris-Cl at pH 7.9, 200 mM NaCl, 0.2 mM DTT, 0.2 mM EDTA). The final volume was measured and an equal volume of 100% glycerol was added. Proteins were stored at −20°C. Protein concentrations were measured using the Bradford assay reagent (Bio-Rad) using bovine serum albumen (BSA) as a standard.

σ^{70} RNAPs with a His_{10} tag at the N-terminus were purified using Ni-agarose. Cell pellets harvested from 250 ml cultures were suspended in 15 ml buffer A (40 mM Tris-Cl pH 7.9, 200 mM NaCl, 10 mM imidazole), 1X HALT protease inhibitor (Pierce), and PMSF to a final concentration of 23 μg/ml. Resuspended pellets were lysed by sonication before being centrifugation at 14,000 rpm for 20 min at 4°C. The cleared lysate was then added to 0.5 ml Ni resin pre-equilibrated in buffer A, the column was washed sequentially with 10 ml buffer A, 5 ml of buffer B (40 mM Tris-Cl pH 7.9, 1 M NaCl) with 10 mM imidazole, 5 ml buffer B with 20 mM imidazole, and 5 ml buffer B with 50 mM imidazole, before elution with 1.5 ml buffer B with 300 mM imidazole. The eluate

was dialyzed into PPX buffer (50 mM Tris-Cl pH 6.5, 150 mM NaCl, 1 mM EDTA, 1mM DTT) for 12 hr at 4°. PreScission protease (4 u; GE Healthcare) was added to the dialyzed protein and incubated for another 12 hr at 4°C before the sample was applied to a 0.5 ml of Ni-NTA column equilibrated in PPX buffer. The flowthrough was collected and concentrated using 5 ml Microcon centrifugal filtration units with a 10 kDa molecular weight cutoff, exchanged with storage buffer (50% glycerol, 10 mM Tris-Cl at pH 7.9, 100 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA), and stored at −20°C. Protein concentrations were measured using the Bradford assay reagent (Bio-Rad) using bovine serum albumen (BSA) as a standard. Holoenzymes were formed with 4-10 fold excess Bpa-containing σ^{70}.

Crosslinking and primer extension mapping

10 μl crosslinking reactions were performed by incubating 40 nM Bpa-containing RNAP with 2 nM plasmid or linear PCR DNA containing the *rrnB* C-7G or *rrnB* con promoters in transcription buffer [10 mM Tris-Cl pH 8.0, 30 mM KCl, 10 mM $MgCl_2$, 0.2 mg/ml BSA (NEB) and 1 mM DTT] in the absence or presence of NTPs for 5 min in a 37° water bath. Reactions in 1.5 ml microfuge tubes were placed directly onto the surface of a UV transilluminator with two 15 watt bulbs and irradiated with 365 nm UV light for 1 min. Samples were then returned to a water bath for 1 min while the lamp was turned off to prevent UV-bulb overheating. Irradiation and water bath incubation were repeated for a total of 10 min of UV-exposure.

Crosslink-proficient positions were initially identified by forming complexes on radiolabeled DNA fragments containing either the wild-type *rrnB* P1 or *rrnB* C-7G mutant promoter, irradiating the complexes with 365 nM UV light for 10 total minutes (as described above), adding 10 μl of 2X SDS-loading buffer 2X Protein Gel Loading

Solution (0.125M Tris-Cl pH 7.4, 4 % SDS, 20% Glycerol, 1.4 M β-mercaptoethanol, 0.1% bromophenol blue), and electrophoresing the reactions on 4-12% acrylamide gels (Invitrogen) in MES buffer, followed by gel drying and phosphorimaging.

For Bpa-RNAPs that crosslinked with low efficiency (βQ148, βR201, βH165, βK203, β'Q1326), we enriched for crosslinked DNA by scaling up to 25 μl, and 2 reactions were performed for each Bpa-containing RNAP in the absence or presence of the appropriate NTPs. After UV-irradiating for 10 min, as described above, the 2 crosslinking reactions were pooled and added to 10 μl of MagneHis Ni-particle slurry (Promega) that had been equilibrated in wash buffer (1 M NaCl, 40 mM Tris-Cl pH 8.0). After incubating 10 min at 25° the supernatant was removed and 100 μl of wash buffer was added to the beads. The wash was incubated for 10 min with gentle mixing every minute before the supernatant was removed. This step was repeated before washing with 50 μl of 1X Taq DNA polymerase buffer (NEB). After removing the supernatant, the beads were resuspended in 10 μl of 1X Taq buffer and used directly in primer extension reactions.

2 μl of each crosslinking reaction (or of the resuspended beads for the scaled-up crosslinking reaction) was used as a template in 12.5 μl primer extension reactions. Reactions also contained 1.25 units of *Taq* DNA polymerase (NEB), 1X Taq buffer (NEB), 250 μM of each dNTP, 2M betaine, 5% DMSO, and ~1 pmol of radiolabeled primer (primer 5910 to monitor crosslinks to the non-template strand and primer 5853 to monitor crosslinks to the template strand). Primer 5910 annealed to the non-template strand of the plasmid backbone, 51-76 nt downstream from the transcription start site. Primer 5853 annealed to the template strand from –83 to -61 relative to the start site (-83 to -74 was from the plasmid backbone sequence and -73 to -61 was from the promoter sequence). Extension products were amplified by 18 cycles of PCR (30 s at 95°C, 30 s at

53°, and 30 s at 72°). An equal volume of primer extension reaction and loading solution (8 M urea, 0.5X TBE, 0.05% bromophenol blue, 0.05% xylene cyanol) were mixed, loaded onto a 40 cm, 9.5% acrylamide, 0.5X TBE, 7M urea gel, and electrophoresed for ~2.5 hr at 2000 V. GATC sequencing ladders were generated with primer 5910 or 5853 and the same template DNAs used for crosslinking, using the Thermo Sequenase Cycle Sequencing Kit (Affymetrix).

DNaseI and $KMnO_4$ Footprinting

For DNaseI footprinting (Bartlett et al., 1998), the template strand from either plasmid pRLG12825 (containing an *rrnB* C-7G promoter fragment with endpoints -73 to +50) or pRLG13829 (containing an *rrnB* con promoter fragment with endpoints -74 to +50) was digested with NcoI (NEB), end-labeled by filling-in with [α-^{32}P] dCTP (Perkin-Elmer) using Sequenase (USB), and digested with NheI (NEB). To label the non-template strand, the above plasmids were first digested with NheI, labeled by filling-in the end of the promoter fragment with [α-^{32}P] dCTP, then digested with NcoI. The DNA was concentrated after each step by ethanol precipitation, and electrophoresed on a 5% acrylamide gel. The promoter fragments were then excised from the gel, diffused overnight into low salt elution buffer (0.2 M NaCl, 20 mM Tris-Cl pH 7.4, 1 mM EDTA), purified using a Qiagen PCR purification kit, ethanol precipitated, and resuspended in 100 µl 10 mM Tris-HCl pH 8.0.

For footprints of *rrnB* con, 20 nM RNAP was added to ~0.2 nM template DNA in 100 mM KCl transcription buffer (10 mM Tris-HCl, pH 8.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1 µg/µl BSA). ApU (600 μM), and UTP (100 μM). For footprints of *rrnB* C-7G, 20-80 nM RNAP was added to ~0.2 nM template DNA in 30 mM KCl transcription buffer (10 mM Tris-HCl, pH 8.0, 30 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1 µg/µl BSA).

ATP (500 μM) and CTP (200 μM) were included in the reaction where indicated. 1 μl of DNaseI (to a final concentration of 0.18-0.5 μg/ml was added for 30 sec before the reaction was stopped by addition of 10 mM EDTA, 0.3 M sodium acetate, and phenol. Glycogen was added to the aqueous fraction, and the DNA was precipitated with ethanol, washed with 100% ethanol, dried, and suspended in 4 μl loading buffer (7 M urea, 0.5x TBE 0.05% bromophenol blue, 0.05% xylene cyanole). Control reactions were performed without DNaseI and RNAP.

For $KMnO_4$ footprinting, (Newlands et al., 1991)RNAP-promoter complexes were formed with promoter fragments radiolabeled on the template or non-template strand and treated with $KMnO_4$ (2.5 mM) for 2 min at 37°C. Reactions were terminated with β-mercaptoethanol (0.34 M final concentration) and precipitated with 0.5 M sodium acetate, glycogen, and 2 vol of ethanol. DNA was resuspended in 2 M ammonium acetate and precipitated again in ethanol. The pellet was washed with ethanol, dried, and suspended in 100 μl 1 M piperidine, heated at 90°C for 30 min, precipitated with 0.3 M sodium acetate and ethanol, washed extensively with ethanol, air-dried, and resuspended in 4 μl loading buffer.

A+G ladders used as markers were made using the same templates as those used for footprinting. 12 μl of template DNA (in H_2O) was depurinated with 50 μl formic acid at room temperature for 7 min(Ross et al., 2001). The DNA was precipitated with ethanol, washed, dried, and suspended in 100 μl 1 M piperidine, heated at 90°C for 30 min, precipitated with ethanol, washed, air-dried, and suspended in 15 μl loading solution.

All samples were heated briefly to 90°C before loading on a 9.5% acrylamide, 7 M urea sequencing gel. Gels were dried under vacuum at 80°C and exposed to a phosphorimager screen overnight.

RNAP-promoter dissociation assays

Promoter decay assays were performed as described previously (Berkmen et al., 2001) using in vitro transcription as a measure of the open complex at various times after heparin addition. Supercoiled plasmid pRLG14034 containing the control promoter *RNA1* and the *rrnB* con promoter (endpoints -73 to +50) or pRLG6791 containing the control promoter *RNA1* and the *rrnB* C-7G promoter (endpoints -66 to +50) were used as templates. RNAP-promoter complexes were formed in transcription buffer containing 100 mM NaCl at 25°C with 10 nM RNAP. After addition of the competitor heparin (40 μg/ml for *rrnB* C-7G or 80 μg/ml for *rrnB* con), samples were removed and added to NTPs (500 μM ATP, 200 μM GTP, 200 μM CTP, 10 μM UTP, and 1 μCi [α-^{32}P] UTP). Samples were incubated for 15 min at 25° before adding an equal volume of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Transcripts were separated on 6% acrylamide-7 M urea denaturing gels, analyzed by phosphorimaging, and quantified using ImageQuant software. The fraction of RNA product at t=0 was plotted vs time, and data points were fit to a single exponential decay curve using SigmaPlot.

Structure modeling

Our modeling aimed to provide an objective, data-driven positioning of the DNA relative to RNAP, based on optimization that aimed to minimize the distance between amino acid-nucleotide crosslinking pairs (summed over all pairs; pairs defined in table S2) while simultaneously minimizing the overlap of RNAP and DNA atoms. Because the configuration of the individual DNA and RNAP molecules in their bound state is unknown, certain assumptions were made about the forms of each of the molecules within the model. First, it was assumed that the RNAP conformation was the same as

that defined by PDB 4LJZ (Bae et al., 2013). This conformation (excluding atoms in σ region 1.1) was used for all models. Second, the DNA molecule was broken into three separate molecules: the upstream double-stranded DNA, the downstream double-stranded DNA, and the single-stranded portion. The upstream and downstream double-stranded DNAs were modeled as straight B-form using the 3D-DART webserver (van Dijk and Bonvin, 2009; van Dijk et al., 2006)This was an oversimplification that prevented precise positioning of the upstream duplex DNA because our model did not allow for necessary DNA bends that have been reported during open complex formation (Campbell et al., 2002; Murakami et al., 2002). The single-stranded portion was modeled as spheres with a 4 angstrom radius connected by 6.7 angstroms. We assumed that the single stranded DNA was completely flexible, except we prevented angles of >90° between three adjacent individual spheres. We used van der Waals radii for all other atoms in the model. We assumed the DNA was melted from -10 to +3 in open complexes formed on both *rrnB* con and *rrnB* C-7G and from -10 to +6 in RP_{ITC5} formed on *rrnB* C-7G. Our $KMnO_4$ footprinting experiments support these endpoints. The precise atoms that were crosslinked were unknown; however, crosslinks in our model were defined as occurring between the α-carbon of the amino acid and the phosphorus atom of the DNA nucleotide for double stranded DNA or between the α-carbon and the sphere for single-stranded DNA.

DNA molecules used in each model:

rrnB con RP_O:

Upstream duplex DNA from -41 to -11 with *rrnB* con with non-template strand sequence 5′-TTTCCTCTTGACAGGCCGGAATAACTCCCCT-3′; single-stranded DNA spheres from -10 to +3; downstream DNA from +4 to +21 with non-template strand sequence 5′-ACTGACACGGAACAACGGC-3′

rrnB C-7G RP_O:

upstream DNA from -40 to -11 with *rrnB* C-7G non-template strand sequence 5′-TTTCCTCTTGACAGGCCGGAATAACTCCCT-3′; single-stranded DNA spheres from -10 to +3; downstream DNA from +4 to +21 with non-template strand sequence 5′-GACACGGAACAACGGC-3′

rrnB C-7G RP_{ITC5}:

Upstream duplex DNA from -40 to -11 with *rrnB* C-7G non-template strand sequence 5′-TTTCCTCTTGACAGGCCGGAATAACTCCCT-3′; single-stranded DNA spheres from -10 to +6; downstream DNA from +7 to +21 with non-template strand sequence 5′-GACACGGAACAACGGC-3′.

Mathematical Formulation:

Initially, the DNA and RNA-Polymerase were randomly positioned. RNAP remained fixed while the DNA was rotated about three axes (x, y, z) and translated to fit the molecules together.

The starting xyz-coordinates of the RNAP atoms are given by $RNAP^{i}_{x,y,z} \in \mathbb{R}^3$ for $i \in I = \{1,2, \dots, N_{RNAP}\}$ and N_{RNAP} indicating the total number of atoms in the RNAP molecule. Similarly, the starting coordinates of the upstream and downstream double-stranded DNA molecules are referred to as $Upstream^{j}_{x,y,z} \in \mathbb{R}^3$ with $j \in J = \{1,2, \dots, N_{Upstream}\}$, and $Downstream^{l}_{x,y,z} \in \mathbb{R}^3$ with $l \in L = \{1,2, \dots, N_{Downstream}\}$. Because the DNA for the melted region is modeled as "beads on a string," its elements do not correspond to atoms, but rather to nucleotides. Similar to upstream and downstream DNA, though, the starting positions of these elements are given by $Melted^{k}_{x,y,z} \in \mathbb{R}^3$ with $k \in K = \{1,2, \dots, N_{Melted}\}$. After being

repositioned, the new coordinates of the DNA segments are given by $NewUpstream^{j}_{x,y,z}$, $NewMelted^{k}_{x,y,z}$, and $NewDownsream^{l}_{x,y,z}$.

Let the crosslinks be indexed by $Link_{Upstream,DNA} \subset J$ with $Link_{Upstream,RNAP} \subset I$, $Link_{Melted,DNA} \subset K$ with $Link_{Melted,RNAP} \subset I$, and $Link_{Downstream,DNA} \subset L$ with $Link_{Downstream,RNAP} \subset I$. Each of the DNA index sets corresponds to an index set from the RNAP atoms. So, for example, the first element of $Link_{Upstream,DNA}$ gives the atom number of the upstream DNA molecule that has been experimentally crosslinked to the RNAP atom given by the first element of $Link_{Upstream,RNAP}$. Let the number of crosslinks be given by $M_{Upstream}$, M_{Melted}, and $M_{Downstream}$.

Optimization Problem: Fit One

Because it is assumed that upstream and downstream DNA maintained their forms during binding, both sections of DNA were repositioned using a rigid transformation matrix and translation vector:

$$NewUpstream^{j}_{x,y,z} = A * Upstream^{j}_{x,y,z} + a \qquad 1)$$

$$NewDownstream^{l}_{x,y,z} = B * Downstream^{l}_{x,y,z} + b \qquad 2)$$

Here A and B are matrices in $\mathbb{R}^{3x3}$ while a and b are vectors in $\mathbb{R}^3$.

The melted DNA was not subject to the same rigid transformation constraints. Instead, each of the beads, representing a nucleotide, was free to be positioned anywhere as long as the distance from its neighboring beads remained the same (i.e. the string could not be stretched). This distance between adjacent nucleotides was defined as d_{Melted} and given a value of 6.7 angstroms. In addition, the movements of the three DNA sections were restricted such that the DNA endpoints matched the endpoints of the adjacent section. For example, the last upstream atom on the non-template strand must be within a specified distance $d_{Endpoints}$ of the first non-template strand nucleotide of the

melted DNA. In order to keep the melted DNA from bending too sharply, a final smoothing constraint was applied. This constraint ensured that the distance between any two nucleotides that are separated by only one nucleotide was greater than some multiple (greater than one but less than two) of d_{Melted}

The objective, as mentioned above, was to minimize the distance between all of the crosslinks. However, because the distance between two crosslinked atoms was likely not zero, a buffer zone was allowed. The size of the buffer is determined by the parameter d_{Buffer}. A distance between two crosslinked atoms was penalized only when the distance became greater than d_{Buffer}.

Decision Variables:

1.	$A \in \mathbb{R}^{3x3}$	:	Transformation matrix
2.	$a \in \mathbb{R}^{3}$	:	Translation vector
3.	$B \in \mathbb{R}^{3x3}$	:	Transformation matrix
4.	$b \in \mathbb{R}^{3}$	:	Translation vector
5.	$NewMelted^{k}_{x,y,z} \in \mathbb{R}^{3}, \forall\, k \in K$	:	Melted DNA nucleotide coordinates
6.	$Gap^{t}_{Upstream}, \forall\, t \in \{1, \dots, M_{Upstream}\}$	:	Upstream Crosslink Distance Penalty
7.	$Gap^{t}_{Melted}, \forall\, t \in \{1, \dots, M_{Melted}\}$	:	Melted Crosslink Distance Penalty
8.	$Gap^{t}_{Downstream}, \forall\, t \in \{1, \dots, M_{Downstream}\}$	:	Downstream Crosslink Distance Penalty

Constraints:

Crosslink Buffer:

For $\forall\, t \in \{1,2, \dots, M_{upstream}\}$:

$$\left\| A * \ Upstream_{x,y,z}^{Link_{Upstream,DNA}^{t}} + a - RNAP_{x,y,z}^{Link_{Upstream,RNAP}^{t}} \right\|^2 - \ Gap_{Upstream}^{t} \leq \left(d_{Buffer}\right)^2 \quad 3)$$

For $\forall\ t \in \{1,2,\dots,M_{Downtream}\}$:

$$\left\| B * \ Downstream_{x,y,z}^{Link_{Downstream,DNA}^{t}} + b - RNAP_{x,y,z}^{Link_{Downstream,RNAP}^{t}} \right\|^2 - \ Gap_{Downstream}^{t} \leq \left(d_{Buffer}\right)^2 \quad 4)$$

For $\forall\ t \in \{1,2,\dots,M_{Melted}\}$:

$$\left\| NewMelted_{x,y,z}^{Link_{Melted,DNA}^{t}} - RNAP_{x,y,z}^{Link_{Melted,RNAP}^{t}} \right\|^2 - \ Gap_{Melted}^{t} \leq \left(d_{Buffer}\right)^2 \quad 5)$$

Distances between Melted DNA Nucleotides:

For $\forall\ t \in \{2,3,\dots,\frac{M_{Melted}}{2},\frac{M_{Melted}}{2}+2,\frac{M_{Melted}}{2}+3,\dots,M_{Melted}\}$:

$$\left\| NewMelted_{x,y,z}^{t} - NewMelted_{x,y,z}^{t-1} \right\| = d_{Melted}$$

This constraint was also applied between the end nucleotides of the two melted strands and atoms representing the end nucleotides of the two strands of upstream and downstream DNA.

Melted DNA Smoothing:

For $\forall\ t \in \{3,4,\dots,\frac{M_{Melted}}{2},\frac{M_{Melted}}{2}+3,\frac{M_{Melted}}{2}+4,\dots,M_{Melted}\}$:

$$\left\| NewMelted_{x,y,z}^{t} - NewMelted_{x,y,z}^{t-3} \right\| \geq k * d_{Melted}$$

This constraint was also applied between the end nucleotides of the two melted strands and atoms representing the end nucleotides of the two strands of upstream and downstream DNA.

Rigid Transformation:

The matrices A and B must have the form:

$$\begin{bmatrix} cos\beta cos\gamma & sin\alpha sin\beta cos\gamma - cos\alpha sin\gamma & cos\alpha sin\beta cos\gamma + sin\alpha sin\gamma \\ cos\beta sin\gamma & sin\alpha sin\beta sin\gamma + cos\alpha cos\gamma & cos\alpha sin\beta sin\gamma - sin\alpha cos\gamma \\ -sin\beta & sin\alpha cos\beta & cos\alpha cos\beta \end{bmatrix} \quad 6)$$

For some α, γ, and β, representing rotation about the axes.

Objective Function:

$$z = \sum_{t=1}^{M_{Upstream}} Gap_{Upstream}^{t} + \sum_{t=1}^{M_{Melted}} Gap_{Melted}^{t} + \sum_{t=1}^{M_{Downstream}} Gap_{Downstream}^{t} \quad 7)$$

Optimization Problem: Fit Two

The first fit moved the crosslinked atoms as close as possible while maintaining assumptions about the structure of the two molecules. The second fit enhanced the first fit by preventing overlap of the two molecules' atoms when possible. In order to prevent overlap, a penalty was imposed whenever one atom from the DNA molecule overlapped with an atom from RNAP. This was implemented by treating each atom as a sphere with a radius determined by the type of atom. Specifically, van der Waals radii were used. These radii are designated by $Radius_{RNAP}^{i}$ for $i \in I$, $Radius_{Upstream}^{j}$ for $j \in J$, $Radius_{Melted}^{k}$ for $k \in K$, and $Radius_{Downstream}^{l}$ for $l \in L$. When two atoms were within a distance less than the sum of their radii they were overlapping and a penalty was incurred.

It was computationally difficult to consider all DNA-RNAP atom pairs when attempting to prevent overlap. In order to keep the problem manageable, the result from the first fit was used to determine which pairs of atoms were likely to overlap. After the first fit, if an RNAP atom was within $nbrhdSize$ of an upstream DNA atom, the pair (the

RNAP atom and the upstream DNA atom) was included in the set $CLOSE_{Upstream} \subset I \times J$. Similarly, the sets $CLOSE_{Melted} \subset I \times K$, and $CLOSE_{Downstream} \subset I \times L$ were created. Only pairs of atoms within these sets were considered when penalizing overlap.

Finally, the objectives of overlap minimization and crosslink distance minimization often clashed. Thus, weights were assigned to the two objectives to allow a solution to be found. The parameters $w_{Upstream}$, w_{Melted}, and $w_{Downstream}$ were used for this purpose, with each applied as a factor to the corresponding overlap term of the objective function.

Constraints:

Overlap:

For $\forall\ (i,j) \in CLOSE_{Upstream}$:

$$\left\|RNAP^{i}_{x,y,z} - NewUpstream^{j}_{x,y,z}\right\| \geq Radius^{i}_{RNAP} + Radius^{j}_{Upstream} - Overlap_{Upstream}{}^{i,j}$$

For $\forall\ (i,k) \in CLOSE_{Melted}$:

$$\left\|RNAP^{i}_{x,y,z} - NewMelted^{k}_{x,y,z}\right\| \geq Radius^{i}_{RNAP} + Radius^{k}_{Melted} - Overlap^{i,k}_{Melted}$$

For $\forall\ (i,l) \in CLOSE_{Downstream}$:

$$\left\|RNAP^{i}_{x,y,z} - NewDownstream^{l}_{x,y,z}\right\| \geq Radius^{i}_{RNAP} + Radius^{l}_{Downstream} - Overlap^{i,l}_{Downstream}$$

Objective Function:

$$z = \sum_{t=1}^{M_{Upstream}} Gap_{Upstream}^{t} + \sum_{t=1}^{M_{Melted}} Gap_{Melted}^{t} + \sum_{t=1}^{M_{Downstream}} Gap_{Downstream}^{t} + w_{Upstream} * \sum_{(i,j)\in CLOSE_{Upstream}} Overlap_{Upstream}^{i,j} + w_{Melted} * \sum_{(i,k)\in CLOSE_{Melted}} Overlap_{Melted}^{i,k} + w_{Downstream} * \sum_{(i,l)\in CLOSE_{Downstream}} Overlap_{Downstream}^{i,l} \quad 8)$$

Decision Variables:

1.	$A \in \mathbb{R}^{3x3}$	:	Transformation matrix
2.	$a \in \mathbb{R}^{3}$	:	Translation vector
3.	$B \in \mathbb{R}^{3x3}$	:	Transformation matrix
4.	$b \in \mathbb{R}^{3}$	:	Translation vector
5.	$NewMelted_{x,y,z}^{k} \in \mathbb{R}^{3}, \forall\, k \in K$	:	Melted DNA nucleotide coordinates
6.	$Overlap_{Upstream}^{i,j}, \forall\, (i,j) \in CLOSE_{Upstream}$	:	Upstream overlap with RNAP
7.	$Overlap_{Melted}^{i,k}, \forall\, (i,k) \in CLOSE_{Melted}$	:	Melted overlap with RNAP
8.	$Overlap_{Downstream}^{i,l}, \forall\, (i,l) \in CLOSE_{Downstream}$	:	Downstream overlap with RNAP
9.	$Gap_{Upstream}^{t}, \forall\, t \in \{1, \dots, M_{Upstream}\}$	:	Upstream Crosslink Distance Penalty
10.	$Gap_{Melted}^{t}, \forall\, t \in \{1, \dots, M_{Melted}\}$	:	Melted Crosslink Distance Penalty
11.	$Gap_{Downstream}^{t}, \forall\, t \in \{1, \dots, M_{Downstream}\}$	:	Downstream Crosslink Distance Penalty

For the three models we used the following parameter values: d^{buffer} was 44; $nbrhdSize$ was 10; and $w_{Upstream}$, w_{Melted}, and $w_{Downstream}$ were 100, 1000, and 100, respectively. These parameters were chosen after a period of trial and error, since we did not know how the value of each variable would affect the output of the models *a priori*. The solution was sensitive to some parameter changes. We found that the value of d^{buffer} sometimes affected the ability of the algorithm to identify a minimum. Increasing $nbrhdSize$ values, on the other hand, quickly resulted in solve times that were not useful for analysis. The parameters that had the largest effect on the final DNA positioning were the weight factors, $w_{Upstream}$, w_{Melted}, and $w_{Downstream}$, which balanced the two obectives of overlap minimization and cross-link distance minimization. In general, giving too little weight to overlap minimization resulted in biologically impossible positioning of the DNA molecules e.g. through the middle of protein domains. Increasing the weight too much, however, would result in complete separation of the relevant atoms of the two molecules (relevant atoms being those in the sets $CLOSE$). Finally, the smoothing parameter k was set to a value of 1.75.

Supplemental References

Artsimovitch, I., Svetlov, V., Murakami, K., and Landick, R. (2003). 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.

Bae, B., Davis, E., Brown, D., Campbell, E.A., Wigneshweraraj, S., and Darst, S.A. (2013). Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves

misappropriation of σ70 domain 1.1. Proc. Natl. Acad. Sci. U.S.a. *110*, 19772–19777.

Bartlett, M.S., Gaal, T., Ross, W., and Gourse, R.L. (1998). RNA polymerase mutants that destabilize RNA polymerase-promoter complexes alter NTP-sensing by rrn P1 promoters. J. Mol. Biol. *279*, 331–345.

Berkmen, M.B., Gaal, T., Josaitis, C.A., and Gourse, R.L. (2001). Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. J. Mol. Biol. *305*, 673–688.

Burgess, R.R., and Jendrisak, J.J. (1975). A procedure for the rapid, large-scall purification of Escherichia coli DNA-dependent RNA polymerase involving Polymin P precipitation and DNA-cellulose chromatography. Biochemistry *14*, 4634–4638.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol. Cell *9*, 527–539.

Murakami, K., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science *296*, 1285–1290.

Newlands, J.T., Ross, W., Gosink, K.K., and Gourse, R.L. (1991). Factor-independent activation of Escherichia coli rRNA transcription. II. characterization of complexes of rrnB P1 promoters containing or lacking the upstream activator region with Escherichia coli RNA polymerase. J. Mol. Biol. *220*, 569–583.

Ross, W., Ernst, A., and Gourse, R.L. (2001). Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. Genes Dev. *15*, 491–506.

Svetlov, V., and Artsimovitch, I. (2015). Purification of bacterial RNA polymerase: tools and protocols. Methods Mol. Biol. *1276*, 13–29.

van Dijk, M., and Bonvin, A.M.J.J. (2009). 3D-DART: a DNA structure modelling server. Nucleic Acids Research *37*, W235–W239.

van Dijk, M., van Dijk, A.D.J., Hsu, V., Boelens, R., and Bonvin, A.M.J.J. (2006). Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. Nucleic Acids Research *34*, 3317–3325.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. Science *338*, 1076–1080.

Zuo, Y., Wang, Y., and Steitz, T.A. (2013). The mechanism of E. coli RNA polymerase regulation by ppGpp is suggested by the structure of their complex. Mol. Cell *50*, 430–436.