Supplementary Information

Supplementary Methods

Estimation of frequency of arrest due to CTG and CAG slip-outs.

The frequency of arrest was calculated as the average of two independent experiments. The intensities of the bands were measured using a PhosphorImager. The intensity of each band is proportional to the product of the molar amount of the corresponding product and its length. We divided the intensity of each band by its length to normalize intensities. These normalized intensities were then used for percentage calculations. Adding the intensities of the bands of all truncated products and dividing the result by the sum of truncated products and the runoff band corresponds to the percentage of total arrest.

Nuclease Activity Assay

DNA substrates containing a CAG slip-out on the transcribed strand were radioactively labeled using Klenow fragment of DNA Polymerase I (NEB). The labeling reaction was performed in a total volume of 35μL, containing 20μL of DNA substrate, 1X NEB2 buffer, 20units of Klenow fragment, and 75μCi of (α-^{32}P) dATP. The reaction was incubated at room temperature for 15min, followed by addition of EDTA to a final concentration of 10mM and heating at 75°C for 20min. Purification was performed using the QIAquick nucleotide removal kit. To test for nuclease activity, labeled DNA was used for the RNAPII transcription protocol with the following modification: either cold nucleotides or no nucleotides were used. Samples were run on a 5% denaturing polyacrylamide gel at 2000 volts, the gel was dried and visualized using a phosphorimager.

Electrophoretic mobility of structures

Different oligonucleotide strands were radioactively labeled using T4 DNA kinase. The labeling reaction was carried out in a total volume of 20μL containing 5μM of single oligonucleotides, 1X forward buffer, 25μCi of (α-32P) dATP, 1μL of T4 kinase. The labeling reaction was incubated at 37° for 20min and the labeled oligonucleotides were purified using the nucleotide kit removal protocol (QIAGEN). The labeled strand was then annealed to an unlabeled oligonucleotide, using the same protocol for annealing as mentioned previously. The different constructs were then run on a native 8% polyacrylamide gel either at 11 or 33 watts and visualized using a PhosphorImager.

Supplementary Figure Legends

Fig. SI1. Analysis of DNA substrates by agarose gel electrophoresis. DNA susbtrates were analyzed on a 1.5% agarose gel. Annealed oligonucleotides were used in excess to promote ligation of most promoter fragments to an oligonucleotide. The final products appeared as expected at around 300 for no insert and around 360 for those with an insert. A minor fraction appeared (grey arrow) consistent with the length of two promoter fragments ligated to each other (expected around 518nt). Annealed oligonucleotides containing an insert ran slower than expected, consistent with the formation of a slip-out (lanes 5-12). Substrates containing CTG, which form more stable mismatched hairpins than CAG, showed higher staining, possibly due to higher intercalation of ethidium bromide into hairpin.

Fig. SI2. RNAPII transcription is arrested by CTG and CAG slip-outs and hairpins. A. DNA substrates containing a longer promoter fragment (+45nt) were obtained and used for transcription. Expected size of full-length transcripts was 306nt for an insert on TS and 246nt for no insert on TS. Note that the band still appears at 190nt. These longer substrates also revealed some truncated products when using substrates containing CAG or CTG on both strands that had not been observed in previous experiments because the position overlapped with the 190nt band. B. Alternative pattern of blockage for hairpin structures.

RNAPII transcription shows that a hairpin on the TS always gave rise to truncated products but the pattern of arrest varied. A distinct pattern of arrest for hairpin on the TS is shown here, with white arrows indicating truncated products that are comparable to the runoff. A hairpin on the NTS always gave truncated products at the same position, at around 175nt.

Fig. SI3 Truncated products are not caused by nuclease activity in the extracts. A. The transcribed strand of DNA substrates containing a CAG slip-out was radioactively labeled and used for the RNAPII transcription protocol using either unlabeled nucleotides or no nucleotides to prevent transcription. Samples were run on a denaturing 5% polyacrylamide gel. Full length DNA appeared at around 363nt, as expected. If nuclease activity was present in the extract then DNA products would appear at 343 and 283nt, corresponding to the beginning and end of the slip-out. However, no shorter DNA products appear indicating that the DNA was not cut. A band at around 500nt appeared for all samples, likely corresponding to the minor fraction of our DNA in which promoter fragments were ligated through the BamHI site (Fig. 1A) B. Overexposure of gel A shows that no distinctive shorter DNA products appear.

Fig. SI4. Electrophoretic mobility of structures in polyacrylamide gels. A. Radioactively labeled single strands (lanes 1-5) or double stranded DNA with different slip-outs (lanes 6-12) were run on a native 8% polyacrylamide at room temperature (11watts) to test whether structures were forming. All single strands containing an insert were 100nt long but had a different mobility in the polyacrylamide gel due to the formation of different structures. Sequences with an insert capable of forming a perfect hairpin (H2) ran faster than those capable of forming a mismatched hairpin (CTG and CAG) (lanes 2, 3, 4) as expected for formation of hairpins (1). CTG ran faster than CAG confirming that it forms a more stable hairpin (lanes 2, 3). All doubles stranded constructs containing slip-outs moved slower than expected from size (lanes 6,7,8,9), consistent with the formation of slip-outs (1). Double stranded constructs with CAG-CAG and CTG-CTG ran closer to 100nt, indicating that they spend more time in a duplex-like conformation (lanes 10,11). B. Single-stranded samples were run on a native 8% polyacrylamide gel at a higher temperature (by increasing the power 3 fold to 33watts). Note that the CTG single strand moves slower than at room temperature (closer to CAG), consistent with the CTG strand spending less time as a hairpin at a higher temperature (lanes 2, 3).

Fig. SI5. Preparation and evaluation of substrates with no nick on the non-transcribed strand (NTS). A. Preparation of substrates without a nick on the NTS. To prepare substrates that did not have a nick on the NTS, the oligonucleotide corresponding to the NTS was 5'-end phosphorylated (5' (P)). The promoter fragment was obtained by digestion of plasmid G4, a derivative of pUCGTG plasmid (2) with the BsgI restriction enzyme, which leaves non-palindromic sticky ends, allowing the oligonucleotides to be designed to prevent self-ligation. B. Scheme used to label only the NTS. To test whether there was a nick on the NTS, we labeled only the NTS using T4 kinase to introduce a radiolabeled phosphate (5' (*P)). First the substrates were dephosphorylated using antarctic phosphatase (NEB), so that both 5' ends would be labeled with the same efficiency. Dephosphorylation was followed by a T4 kinase forward reaction to phosphorylate the 5' ends using a radiolabeled phosphate. The substrates were then restricted using SalI to remove the label from the TS. C. Denaturing polyacrylamide gel to assay for ligation of NTS. Substrates with SalI restriction (SalI +) were labeled only on the NTS through a 5'-end radiolabeled phosphate. For each substrate there are four possibilities: (I) Both NTS and TS are ligated, (II) NTS ligated but TS has a nick, (III) NTS has a nick but TS is ligated, (IV) Neither NTS nor TS are ligated (free promoter fragment). These four possibilities are illustrated for the No Insert (-) No insert (-) substrate, where the red line indicates the labeled strand and the triangle (▽) indicates a nick. For substrates with No Insert (-) on the NTS, ligated and restricted NTS would result in a band at around 280nt (note that this includes I and II), while a non-ligated (nick) NTS would result in a band at around 250nts (note that this includes III and IV). The dashed arrow indicates the unrestricted substrate (at around 290nts). For substrates with an insert on the NTS, ligated and restricted NTS would result in a band at around 340nts while a non-ligated (nick) NTS would result in a band at around 250nts. D. Non-denaturing

polyacrylamide gel to assay for free promoter fragment. To discriminate between (III) (NTS has a nick but TS is ligated) and (IV) (neither NTS nor TS are ligated), the labeled substrates were run on a non-denaturing polyacrylamide gel. Since the duplex remains annealed, any substrates that have at least one ligated strand (I, II and III) will migrate together, while substrates where neither strand is ligated (IV) (free promoter fragment) will migrate faster. The estimation of the percentage of ligation of the NTS can be obtained using the information obtained from both gels. The fraction of interest comprises of the amount of substrates where the NTS is ligated out of the substrates that produce long transcription products. There are two possible limiting cases: 1) Substrate II (NTS is ligated but TS has a nick) produces long transcription products with 100% probability or 2) The nick on the TS present in substrate II causes a complete block to transcription and thus substrate B does not produce long transcription products.

In the first case, the fraction of interest is (I+II)/(I+II+III) since I and II have a ligated NTS and I+II+III all produce long transcription products. In this case, the fraction of interest can be estimated directly from the gels since the percentage of the slower migrating band in the denaturing (P_D) gel is:

P_D= I+II

The percentage of the slower migrating band in the non-denaturing gel (P_{ND}) is:

P_{ND}= I+II+III

Therefore, the fraction of interest = (I+II)/(I+II+III)=(P_D)/(P_{ND}).

In the second case, where substrate II does not contribute to long transcription, the fraction of interest is I/(I+III) since out of the products that produce long transcription (I+III), only I has a ligated NTS. To use the information from the gels to find this fraction, we can define the probability of ligation of each strand as P_{TS}=probability of ligation of TS and P_{NTS}=probability of ligation of NTS. Using these probabilities, we can find the probability of occurrence of each of the substrates:

I=$P_{TS}P_{NTS}$ II=$P_{NTS}(1-P_{TS})$ III=$P_{TS}(1-P_{NTS})$ IV=$(1-P_{TS})(1-P_{NTS})$

We can then substitute these probabilities into

P_D = I+II = $(P_{TS}P_{NTS})+P_{NTS}(1-P_{TS}) = P_{TS}P_{NTS} + P_{NTS} - P_{TS}P_{NTS} = P_{NTS}$

Therefore, $P_D=P_{NTS}$

We can also substitute the ligation probabilities into:

Fraction of interest = I/(I+III) = $(P_{TS}P_{NTS})/(P_{TS}P_{NTS}+P_{TS}(1-P_{NTS}))= (P_{TS}P_{NTS})/(P_{TS}P_{NTS}+P_{TS}-P_{TS}P_{NTS})=$ $(P_{TS}P_{NTS})/(P_{TS})=P_{NTS}$

Therefore, fraction of interest = P_{NTS} = P_D

Using this information we can estimate the interval for the percentage of our fraction of interest as P_D to (P_D/P_{ND}).

For:

No insert (TS) No insert (NTS): 63.13% to 83.00%

CAG (TS) and No insert (NTS): 61.01% to 90.40%

No insert (TS) and CTG (NTS): 61.19% to 100%

CAG (TS) and CTG (NTS): 61.81% to 97.86%

Fig. SI6. The presence of a nick on the non-transcribed strand has no effect on the RNAPII transcription arrest due to CTG or CAG slip-outs. A. RNAPII transcription was performed with substrates with or without a nick on the NTS. Substrates were tested for the presence of a nick on the NTS by labeling of the NTS only and separating the products on a denaturing polyacrylamide gel (See FigSI 5). B. Substrates designated as containing no nick on the NTS were found to always be over 60% ligated (See Fig. SI5), however, the percentage of arrest when comparing these substrates to the ones containing a nick was very similar, indicating that a nick on the NTS has no effect on the RNAPII transcription arrest caused by CAG or CTG slip-outs.

Fig. SI7. T7 RNAP transcription arrest in the presence of nuclear extracts found when a slip-out is on the TS is independent of T7 RNAP concentration. T7 RNAP transcription was performed in the presence of nuclear extracts with substrates containing only the T7 RNAP promoter and following the RNAPII protocol (See Materials and Methods). T7 RNAP was added in the concentration used in previous experiments (1X) or was diluted 10X in T7 RNAP storage buffer (0.1X) before addition to reaction. A black arrow indicates the runoff while a white arrow indicates transcription arrest, which is only present when the substrate has a slip-out and the reaction is performed in the presence of nuclear extract. For a T7 RNAP concentration of 1X or 0.1X, arrest represents 19.41% or 19.20% of total transcription products, respectively, indicating that the amount of arrest is independent of T7 RNAP concentration. Since a lower T7 RNAP concentration yields a lower concentration of transcription products, the lanes are separated to show comparable exposures.

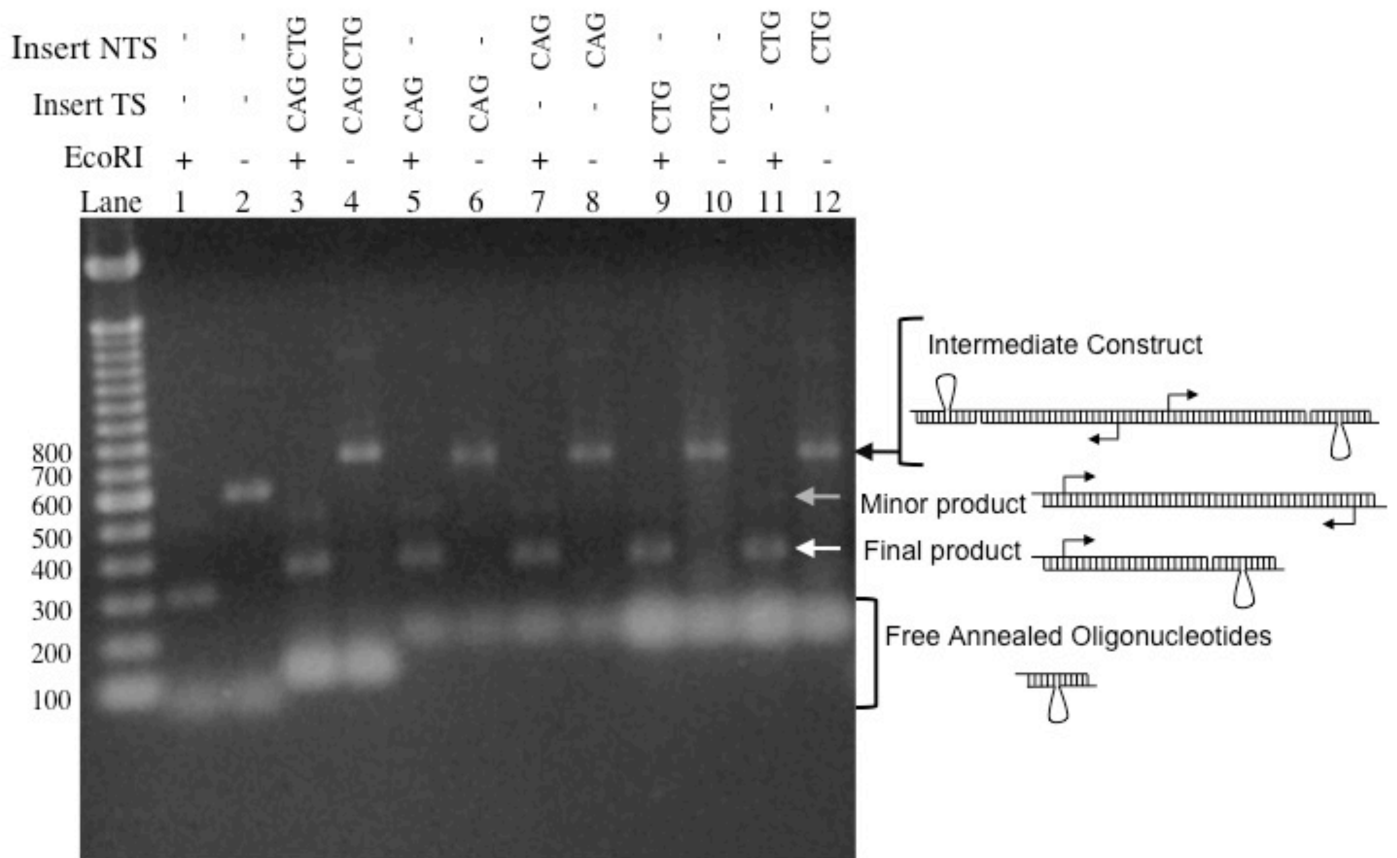
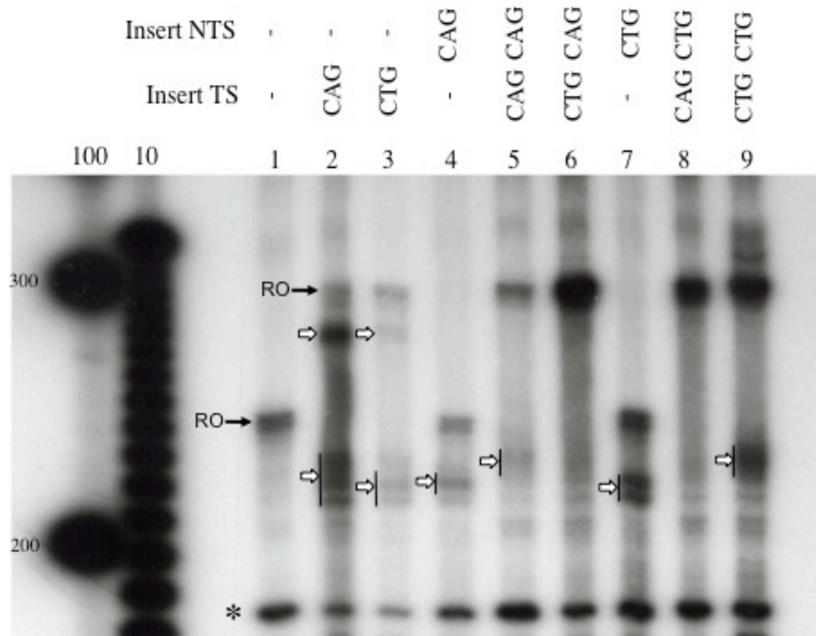


Fig. SI1



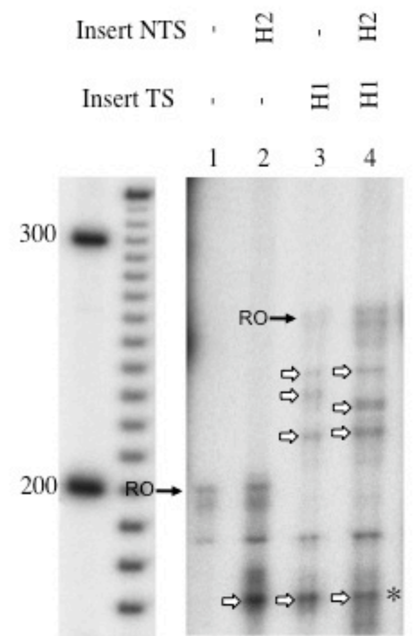
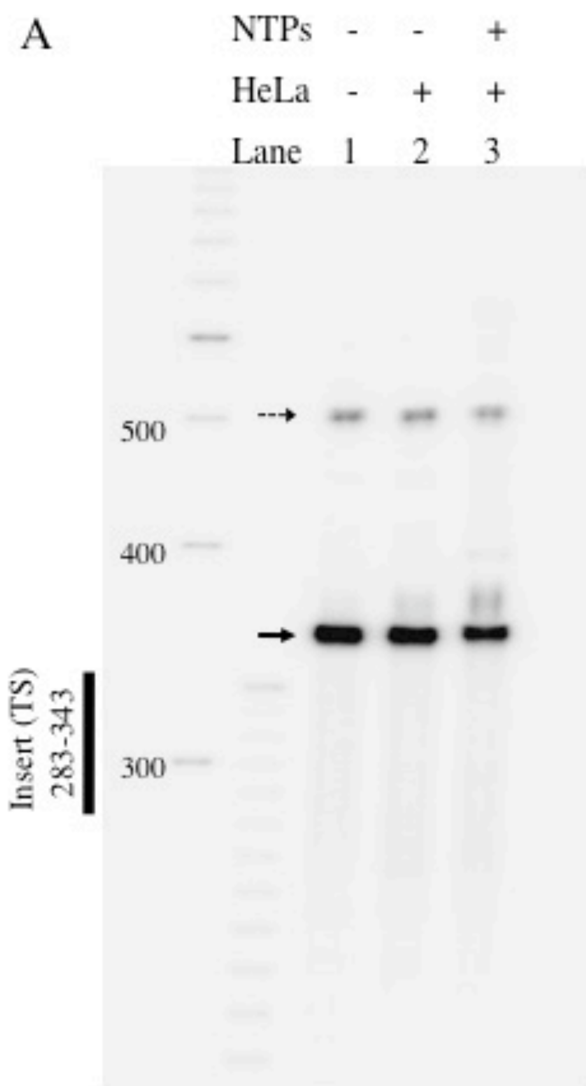


Fig. SI2



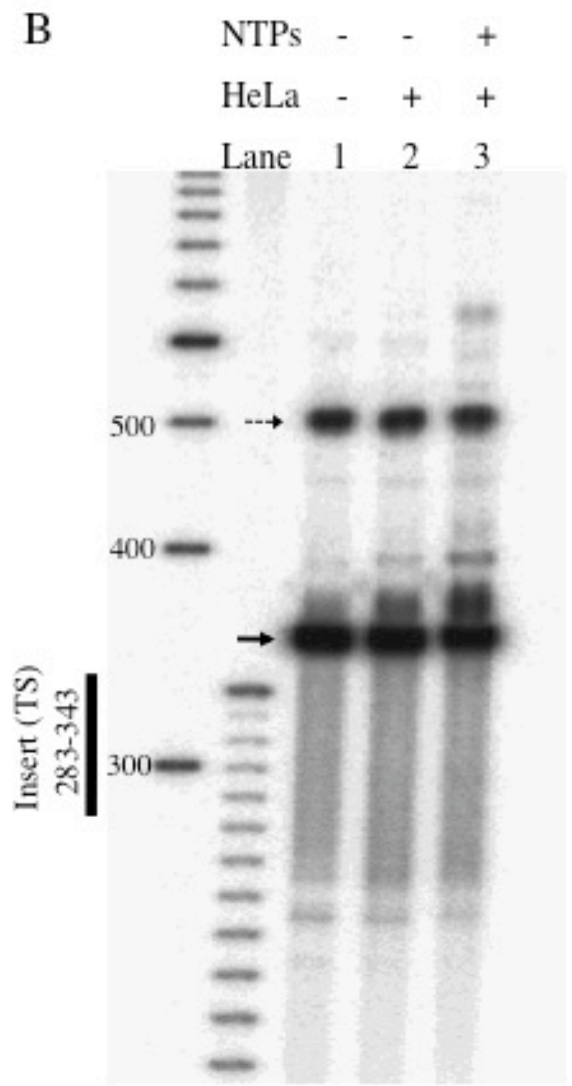


Fig. SI3

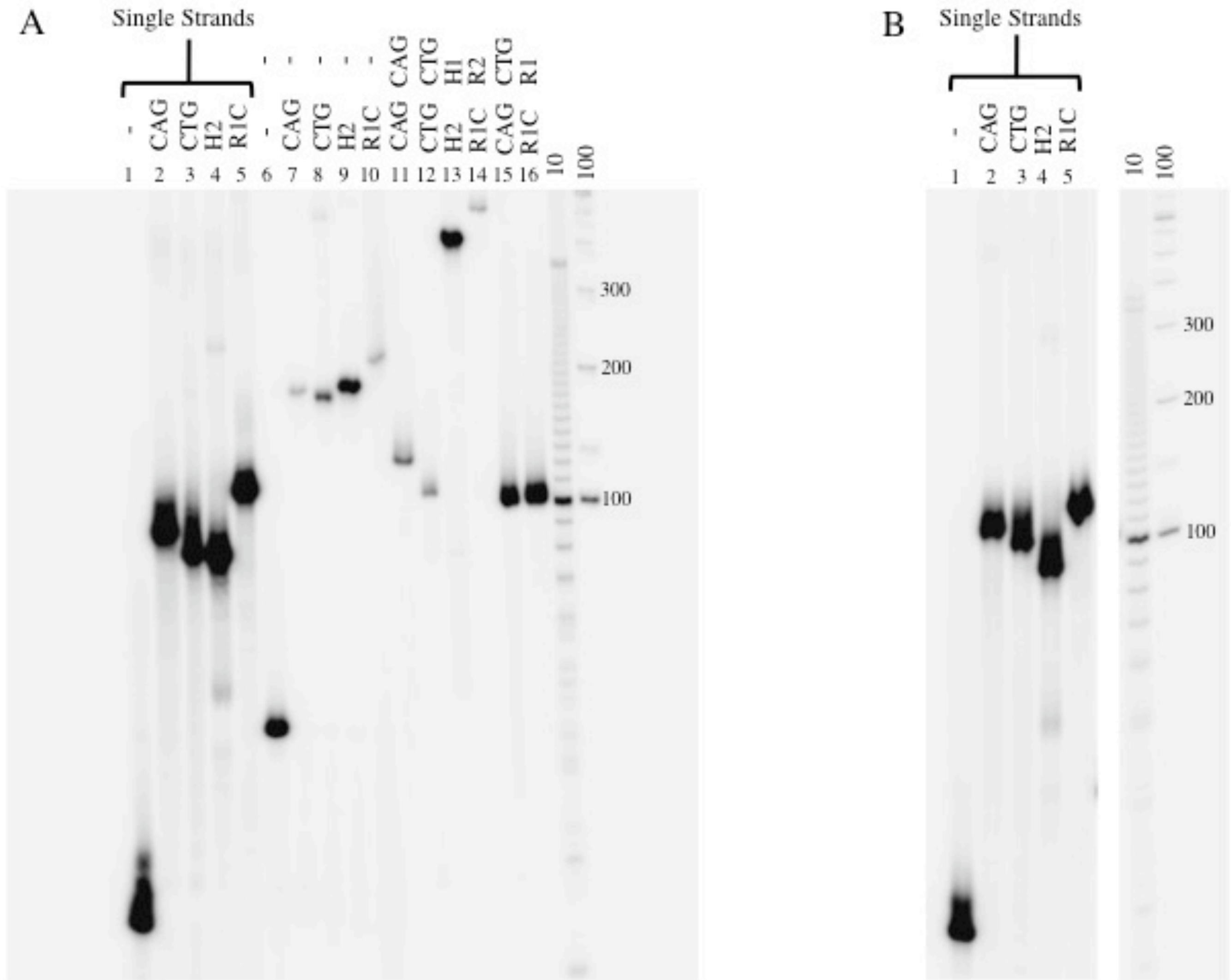


Fig. SI4

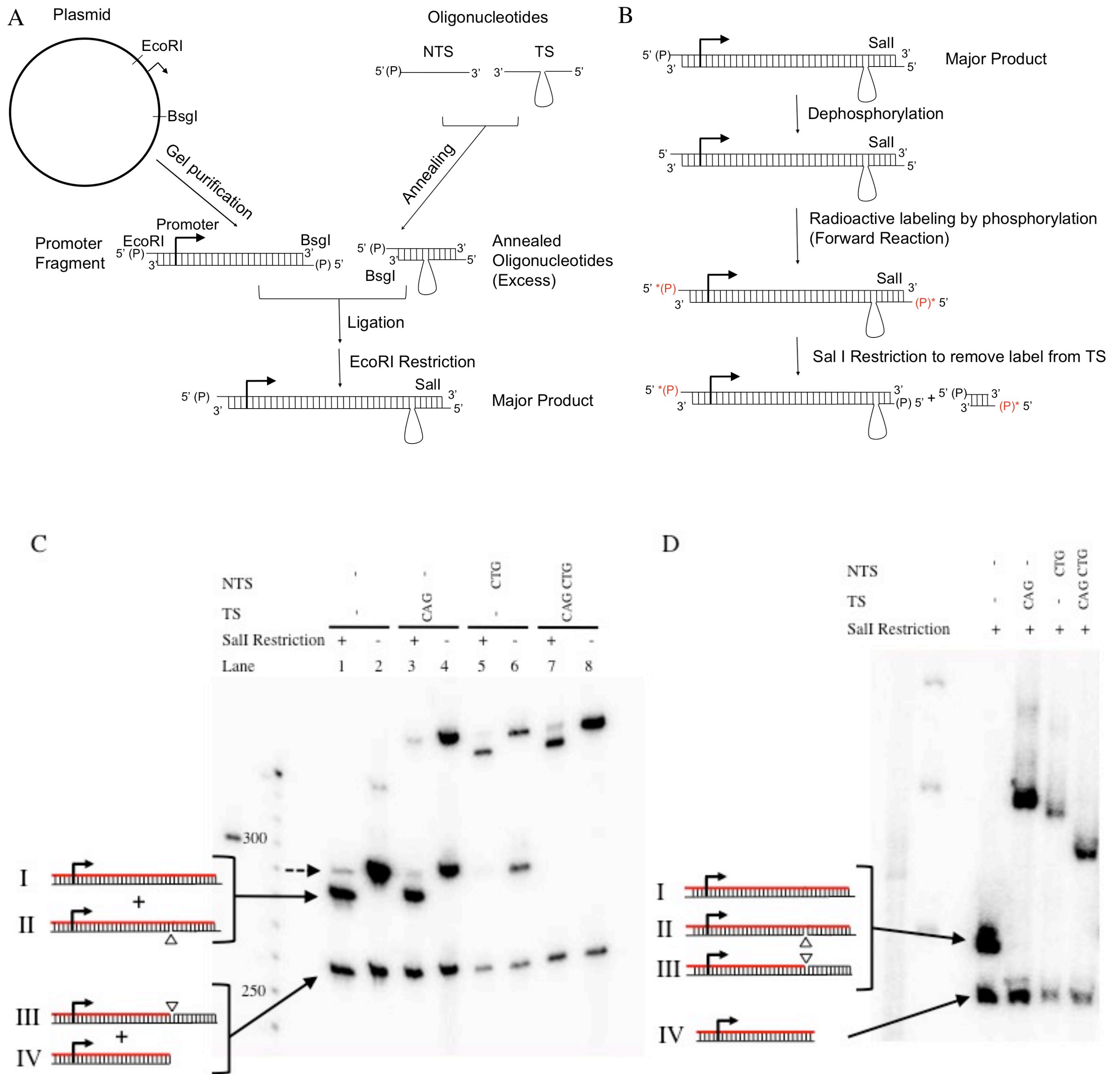


Fig. SI5

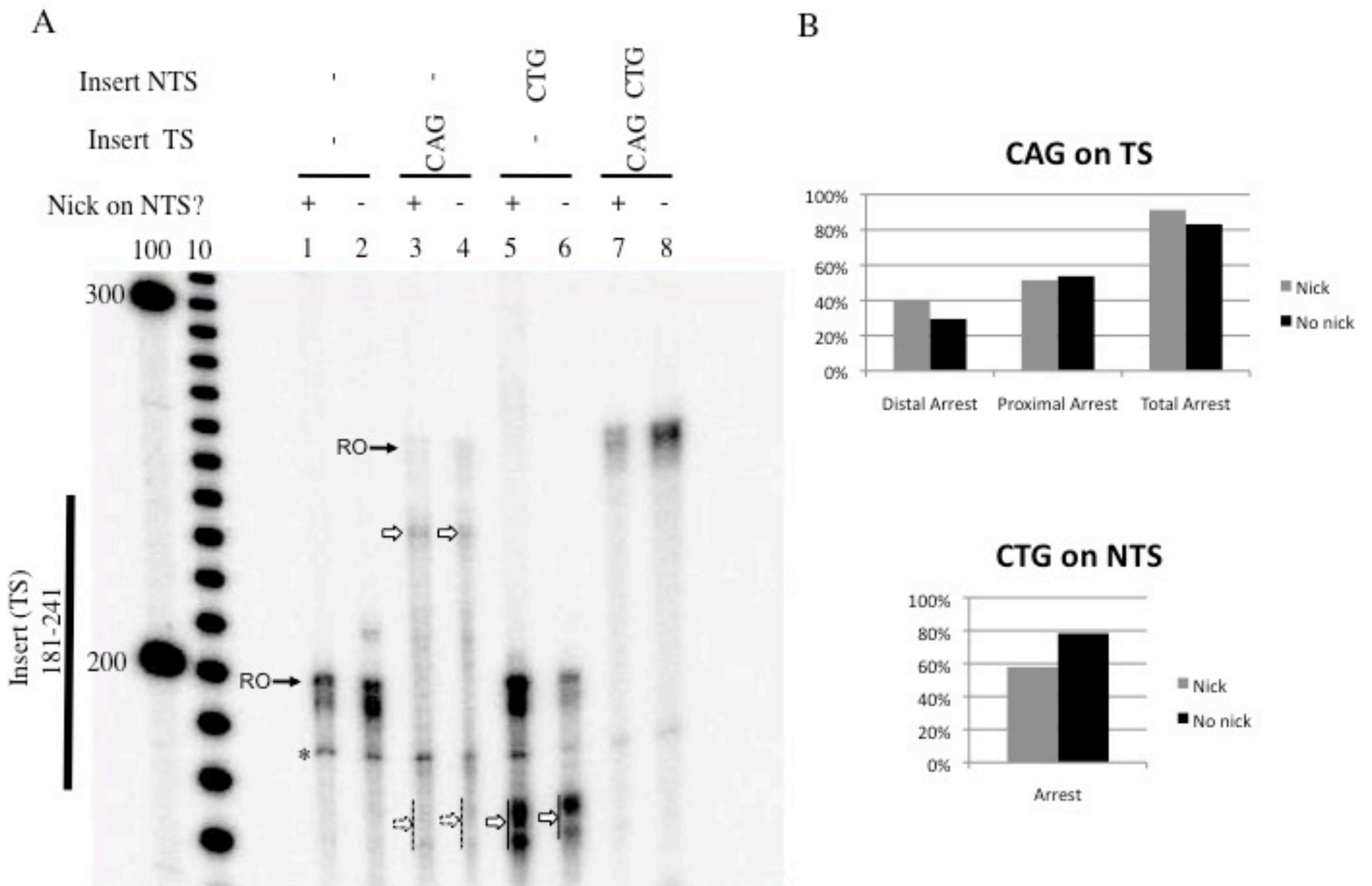


Fig. SI6

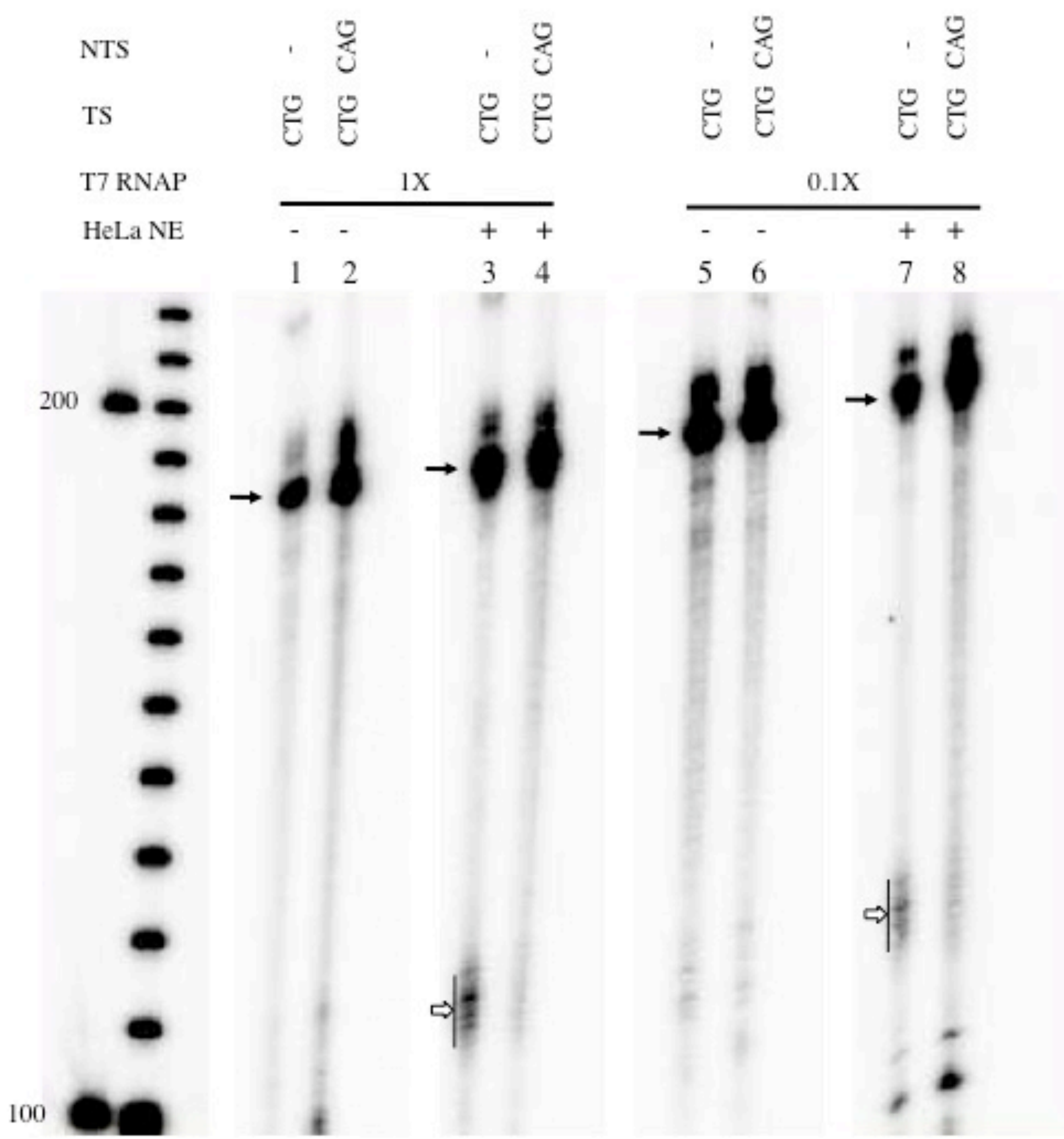


Fig. SI7

References:

1. Oussatcheva, E.A., Shlyakhtenko, L.S., Glass, R., Sinden, R.R., Lyubchenko, Y.L. and Potaman, V.N. (1999) Structure of branched DNA molecules: gel retardation and atomic force microscopy studies. *J Mol Biol*, **292**, 75-86.
2. Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M. and Hanawalt, P.C. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc Natl Acad Sci U S A*, **107**, 12816-12821.