Lineage-specific variations in the trigger loop modulate RNA proofreading by bacterial RNA polymerases

Daria Esyunina, Matti Turtola, Danil Pupov, Irina Bass, Saulius Klimašauskas, Georgi Belogurov, Andrey Kulbachinskiy

SUPPLEMENTARY DATA

Supplementary Methods

Analysis of nucleotide addition and RNAP translocation (Fig. 2 and Fig. S2)

TEC assembly. TECs (1 μM final concentration) were assembled by a procedure described in (1). RNA primer labeled with Atto680 fluorescent dye at the 5′-end was annealed to template DNA, and incubated with 1.5 μM RNAP for 10 min at 25 °C in TB10 buffer (10 mM $MgCl_2$, 40 mM HEPES-KOH pH 7.5, 80 mM KCl, 5% glycerol, 0.1 mM EDTA, and 0.1 mM DTT) and with 2 μM of the non-template DNA for 20 min at 25 °C. For TECs used in nucleotide addition measurements, RNA was the limiting component at 1 μM, and the template strand was used at 1.4 μM, whereas for TECs used in translocation the template strand was limiting at 1 μM, and RNA was added at 1.4 μM.

Nucleotide addition measurements. To determine the incorporation efficiency of GMP and 2′-dGMP substrates, 0.2 μM TEC in 20 μl of TB10 buffer was incubated for 5 min with indicated amounts of substrates (Fig. 2C) at 25 °C and quenched by adding 80 μl of loading buffer (94% formamide, 4 mM Li_3-EDTA, and 0.2% Orange G). Time-resolved measurements were performed in an RQF 3 quench-flow instrument (KinTek Corporation, Austin, TX, USA). The reaction was initiated by rapid mixing of 14 μl of 0.4 μM TEC with 14 μl of 400 μM NTP. Both TEC and NTP solutions were prepared in TB10 buffer. The reaction was allowed to proceed for 0.004–10 s at 25 °C, quenched with 86 μl of 0.5 M HCl and immediately neutralized by adding 171 μl of loading buffer (290 mM Tris base, 13 mM Li_3-EDTA, 0.2% Orange G, 94% formamide). RNAs were separated on 16% denaturing polyacrylamide gels and visualized with Odyssey Infrared Imager (Li-Cor Biosciences, Lincoln, NE, USA); band intensities were quantified using ImageJ software (2).

Translocation measurements. RNAP translocation was assayed by monitoring changes in fluorescence of 6-methyl-isoxanthopterin (6-MI) base incorporated into template DNA (3). Equilibrium levels of fluorescence were determined by recording emission spectra of 6-MI (excitation at 340 nm) with an LS-55 spectrofluorometer (PerkinElmer, Waltham, MA, USA) at 25 °C. The fluorescence at peak emission wavelength (420 nm) was used for data analysis and representation. Preassembled TECs were diluted at 200 nM with TB10 buffer. Prior to spectral measurements, TEC was incubated for 5 min with indicated amounts of substrates (Fig. 2C) in 100 μl of TB10 buffer at 25 °C. Time-resolved measurements were performed in an Applied Photophysics (Leatherhead, UK) SX.18MV stopped-flow instrument at 25 °C. The reaction was initiated by mixing 60 μl of 0.2 μM TEC with 60 μl of 400 μM NTP. Both solutions were prepared in TB10 buffer. 6-MI fluorophore was excited at 340 nm and emitted light was collected through 400 nm longpass filters. At least three individual traces were averaged for each reported curve.

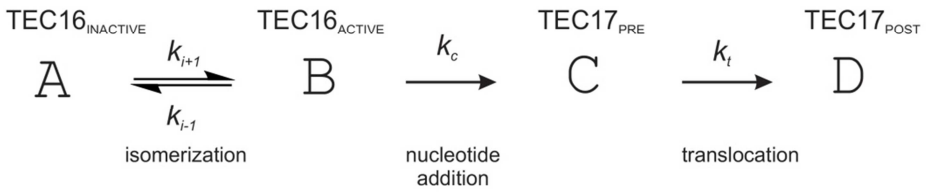
Data analyses. Time-resolved nucleotide incorporation and translocation data were simultaneously fit to a three-step model using the numerical integration capabilities of KinTek Explorer software version 5.1 (4) (KinTek Corporation, Austin, TX, USA). During fitting nucleotide incorporation data were assigned a weight of 50 to equalize their contribution to the sum of residuals with the translocation time-traces that contained 50 times more data points. The model postulated that the assembled TEC16 slowly and reversibly interconverts between inactive and active states and, upon the addition of the GTP substrate, undergoes an irreversible transition to TEC17, followed by irreversible translocation. The reversible isomerization step was needed to model a small (10-20%) fraction of slowly extending complex in *Eco* TEC preparations (5). The above step is not depicted in the abbreviated version of the analysis scheme shown in Fig. 2. The full model used for fitting the data is described in detail in the next section.

1. Komissarova, N., Kireeva, M.L., Becker, J., Sidorenkov, I. and Kashlev, M. (2003) Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol.*, **371**, 233-251.
2. Abramoff, M.D., Magalhaes, P.J. and Ram, S.J. (2004) Image Processing with ImageJ. *Biophotonics Int.*, **11**, 36-42.
3. Malinen, A.M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M.S. and Belogurov, G.A. (2012) Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res*, **40**, 7442-7451.
4. Johnson, K.A. (2009) Fitting enzyme kinetic data with KinTek Global Kinetic Explorer. *Methods Enzymol*, **467**, 601-626.
5. Malinen, A.M., Nandymazumdar, M., Turtola, M., Malmi, H., Grocholski, T., Artsimovitch, I. and Belogurov, G.A. (2014) CBR antimicrobials alter coupling between the bridge helix and the beta subunit in RNA polymerase. *Nature communications*, **5**, 3408.

Model employed for fitting of nucleotide addition and translocation data (Fig. 2)

Initial conditions (See Note 1):

T=0 A = p (~0.2) B = 1-p (~0.8) C = 0 D = 0



Rate equations (See Note2):

$A' = -A*k_{i+1}+B*k_{i-1}$

$B' = -B*k_c - B*k_{i-1} + A*k_{i+1}$

$C' = -C*k_t + B*k_c$

$D' = C*k_t$

Equations for dependent variables:

RNA17=F1*(C+D)

SF_fluorescence=F2*(A+B+C)+F3*D

Independent variables: T -time

Dependent variables:
RNA17 - RNA17 band intensities from quench flow experiment
SF_fluorescence - Fluorescent trace from stopped-flow experiment

Parameters:
k_{i+1} -rate of isomerization of inactive TEC into active TEC
k_{i-1} -rate of isomerization of active TEC into inactive TEC
k_c -rate of nucleotide incorporation
k_t -rate of forward translocation

F1 -normalization coefficient for quench flow data
F2 -normalization coefficient for stopped-flow data (fluorescence at T= 0)
F3 -normalization coefficient for stopped-flow data (fluorescence at T= ∞)
p -fraction of inactive TEC (See Note 1)

Note 1: The fraction of inactive TEC (”p” parameter) is determined by k_{i+1} and k_{i-1} because TEC16 is assembled and reaches equilibrium in the absence of NTP. To reflect this condition we introduced a 100 s mixing step without NTP into the Kintek Explorer virtual experiment setup. The initial fractions of active and inactive TEC16 can then be chosen arbitrary, the ”p” parameter is redundant and is not explicitly used.

Note 2: Rate equations are uniquely defined by the reaction scheme and do not need to be explicitly specified when fitting data with Kintek Explorer.

Note 3: During fitting nucleotide incorporation data were assigned a weight of 50 to equalize their contribution to the sum of residuals with the translocation time-traces that contained 50 times more data points.

Parameters describing kinetics of nucleotide addition and translocation (Fig. 2)

	k_{i+1} (isomerisation)		k_{i-1} (isomerisation)		k_c (nucleotide addition)		k_t (translocation)	
	bestfit	range	bestfit	range	bestfit	range	bestfit	range
Eco RNAP	1.0 s^{-1}	0.4 - 2.0 s^{-1}	0.1 s^{-1}	0.05 - 0.3 s^{-1}	30 s^{-1}	29 - 33 s^{-1}	76 s^{-1}	70 - 85 s^{-1}
Dra RNAP	fixed=0	n/a	fixed=0	n/a	15 s^{-1}	14 - 16 s^{-1}	115 s^{-1}	90 - 140 s^{-1}

Lower and upper bounds of parameters (range) were calculated at a 10% increase in Chi^2 over the minimal value using FitSpace routine of Kintek Explorer

Supplementary Tables and Figures

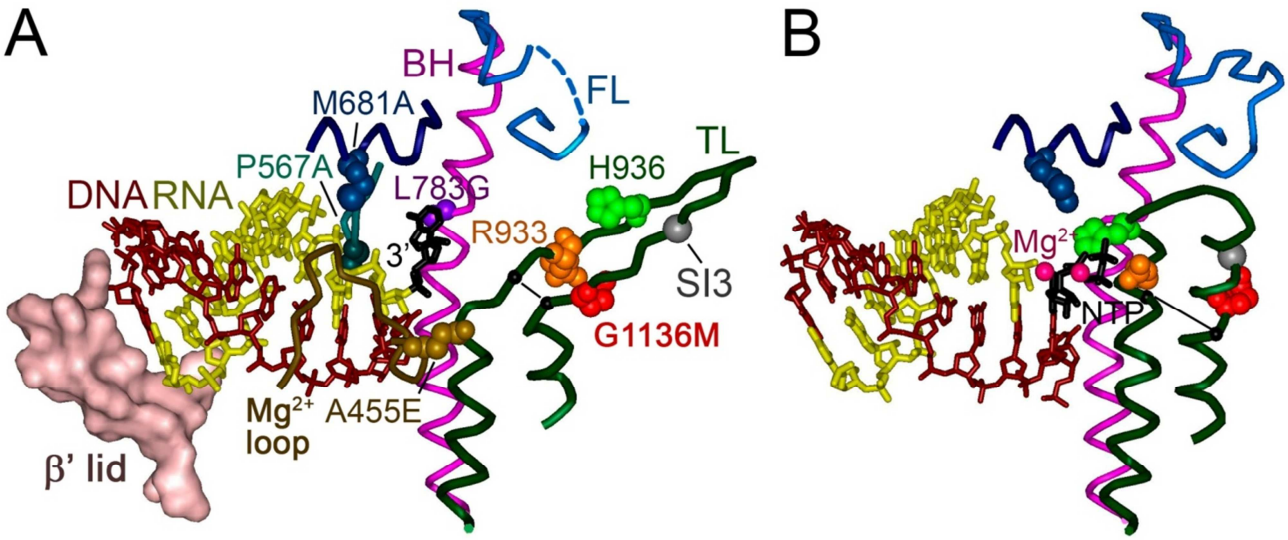


Figure S1. Comparison of the RNAP structures in the backtracked and active TECs. (A) Backtracked TEC structure (4WQS, (1)). (B) Active TEC structure with ATP bound in the active site (2O5J, (2)). Positions of all RNAP elements analyzed in this study are shown on the backtracked TEC structure. All designations correspond to Fig. 1.

1. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. and Yokoyama, S. (2015) The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol Cell*, **57**, 408-421.
2. Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I. and Landick, R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, **448**, 163-168.

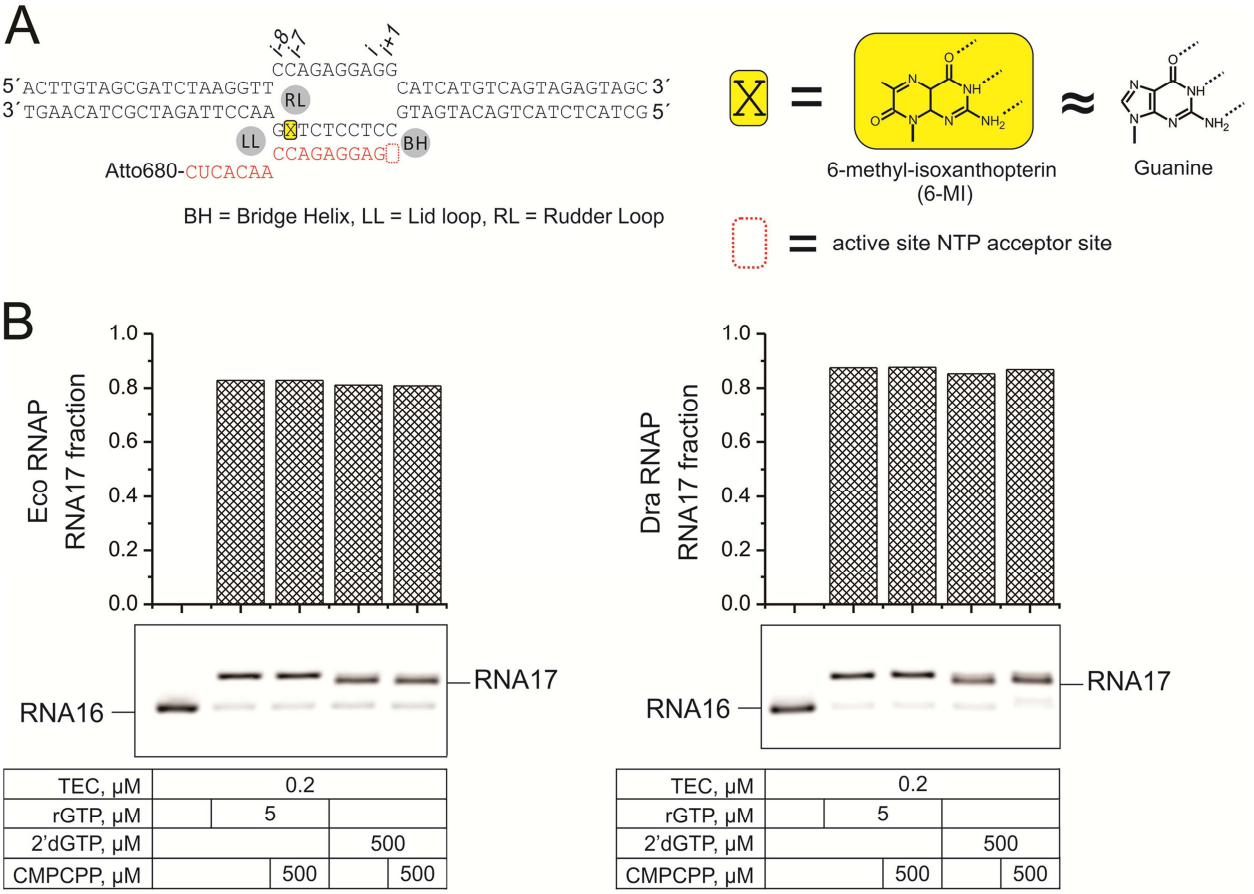


Figure S2. The experimental system used for comparative analysis of nucleotide addition and translocation by *Dra* and *Eco* RNAPs.

(A) The schematic of the TEC. The infrared fluorophore Atto680 at the 5′-end of the RNA primer is used for detection of RNA products in PAGE gels. The 6-MI fluorescent base is positioned at register *i-7* in the initial TEC, and its fluorescence is quenched by the upstream guanine base. After nucleotide addition and forward translocation along the DNA, 6-MI migrates to the edge of the RNA:DNA hybrid (register *i-8*) and unstacks from the upstream guanine. As a result, the 6-MI fluorescence increases 2-5 fold and this increase is employed for monitoring RNAP translocation along the DNA. The absolute levels of 6-MI fluorescent intensities of GMP-extended *Dra* TECs were nearly twofold higher than those of *Eco* TECs (main text Fig. 2B), whereas the fluorescent intensities of the initial TECs and the TEC extension efficiencies (80–90 %) were similar for both RNAPs. While the exact reasons for these differences remain to be investigated, they may be possibly explained by potentially dissimilar conformations of the upstream DNA and/or protein environments of the fluorophore in the two RNAPs.

(B) Extension efficiencies of *Eco* and *Dra* TECs. The RNA extension experiments were performed under the same conditions as equilibrium fluorescence experiments depicted in the main text Fig. 2C except that in RNA extension experiments the TECs were assembled on scaffolds with the excess of DNA over RNA.

A

B

RNAP	UTP (k_{obs}, s^{-1}, 25°C)
***Eco* WT**	26 ± 4
***Eco* ΔSI3**	28 ± 4
***Eco*-G1136M(ΔSI3)**	44 ± 7
Dra	48 ± 8
***Dra*-M1271G**	28 ± 6

C

RNAP	UTP (k_{obs}, s^{-1}, 10°C)		CTP (k_{obs}, s^{-1}, 25°C)	
	Mg^{2+}	Mn^{2+}	Mg^{2+}	Mn^{2+}
Eco	9 ± 4 **1**	38 ± 6 **4.2**	0.016 **1**	1.2 ± 0.1 **75**
Dra	23 ± 3 **1**	55 ± 11 **2.4**	0.024 **1**	10 ± 2 **420**

Figure S3. Nucleotide incorporation by *Eco*, *Dra* RNAPs and their mutant variants on the minimal nucleic acid scaffold template. (A) Scheme of the template. (B) The rates of cognate UTP addition by wild-type *Eco*, *Dra* RNAPs and their mutant variants. The experiments were performed at 25 °C at 1 mM UTP concentration in transcription buffer containing 40 mM Tris-HCl, pH 7.9, 40 mM NaCl and 10 mM $MgCl_2$. At each time point, RNA extension was calculated as a ratio of extended 9 nt RNA product to the sum of the 8 nt and 9 nt RNAs and k_{obs} values were calculated from single-exponential fits of the data. (C) The rates of cognate UTP (1 mM, measured at 10 °C) and non-cognate CTP (1 mM, measured at 25 °C) addition in the presence of $MgCl_2$ or $MnCl_2$ (10 mM).

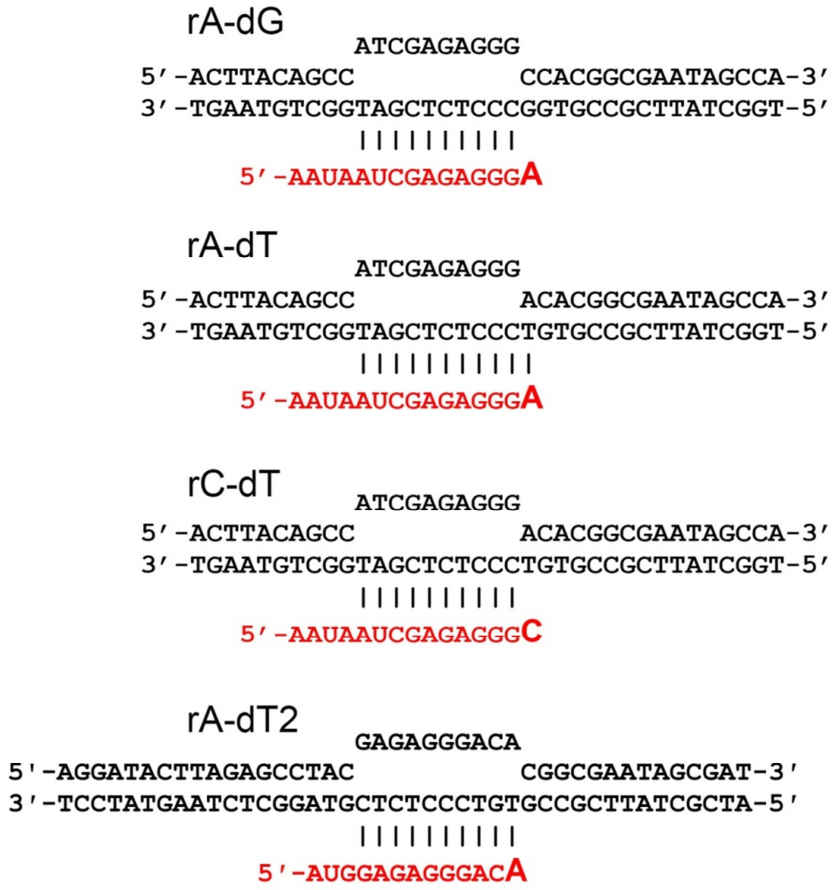


Figure S4. Schematics of nucleic acid scaffolds used for analysis of RNA cleavage.

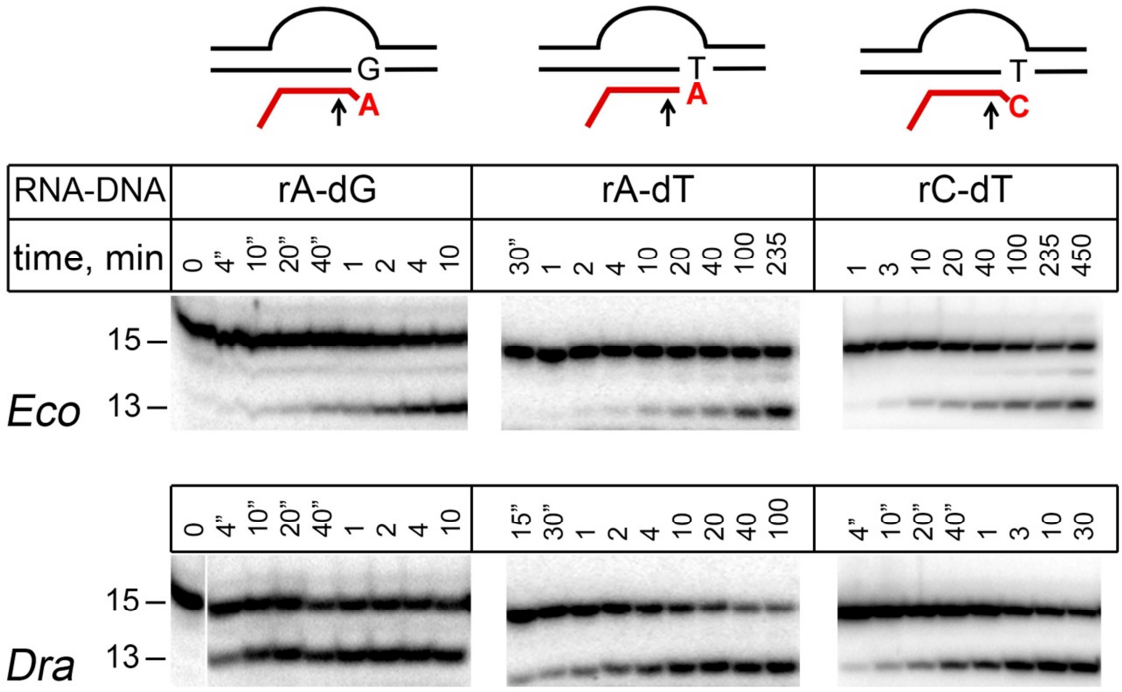


Figure S5. Time course of RNA cleavage by *Eco* and *Dra* RNAPs in mismatched and correct TECs. The unreacted 15 nt RNAs at long reaction times correspond to RNA oligonucleotides that were not assembled into the TECs during the reconstitution procedure.

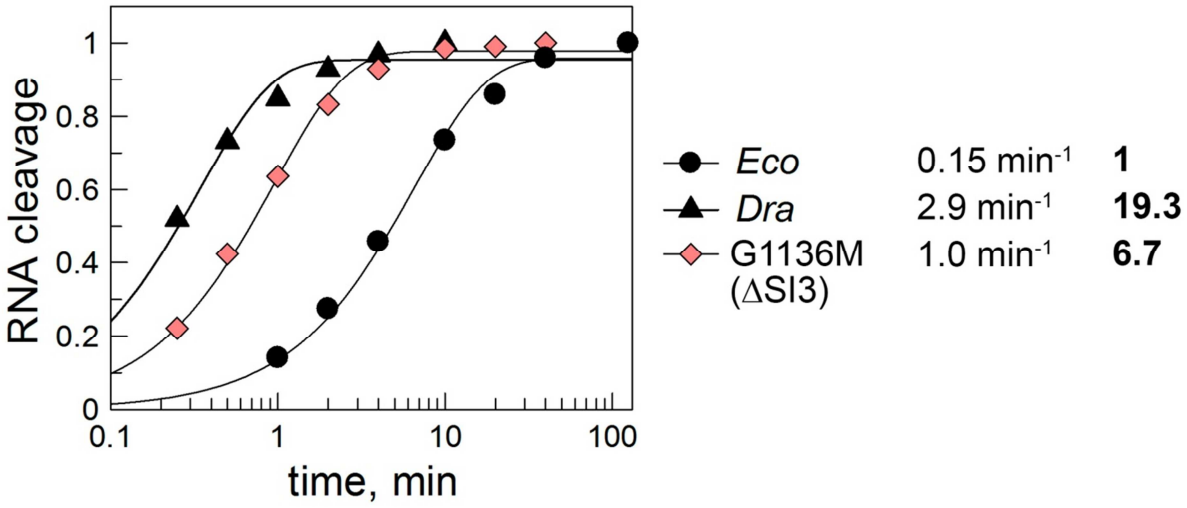


Figure S6. Kinetics of RNA cleavage by *Eco*, *Dra* and *Eco*-G1136M(ΔSI3) RNAPs in the rA-dG TEC at 25 °C. The amounts of cleaved RNA are normalized to the maximal cleavage observed for each RNAP. The observed catalytic rates (k_{obs}, min^{-1}) are shown on the right; the numbers in bold show fold-differences in the rates in comparison with wild-type *Eco* RNAP.

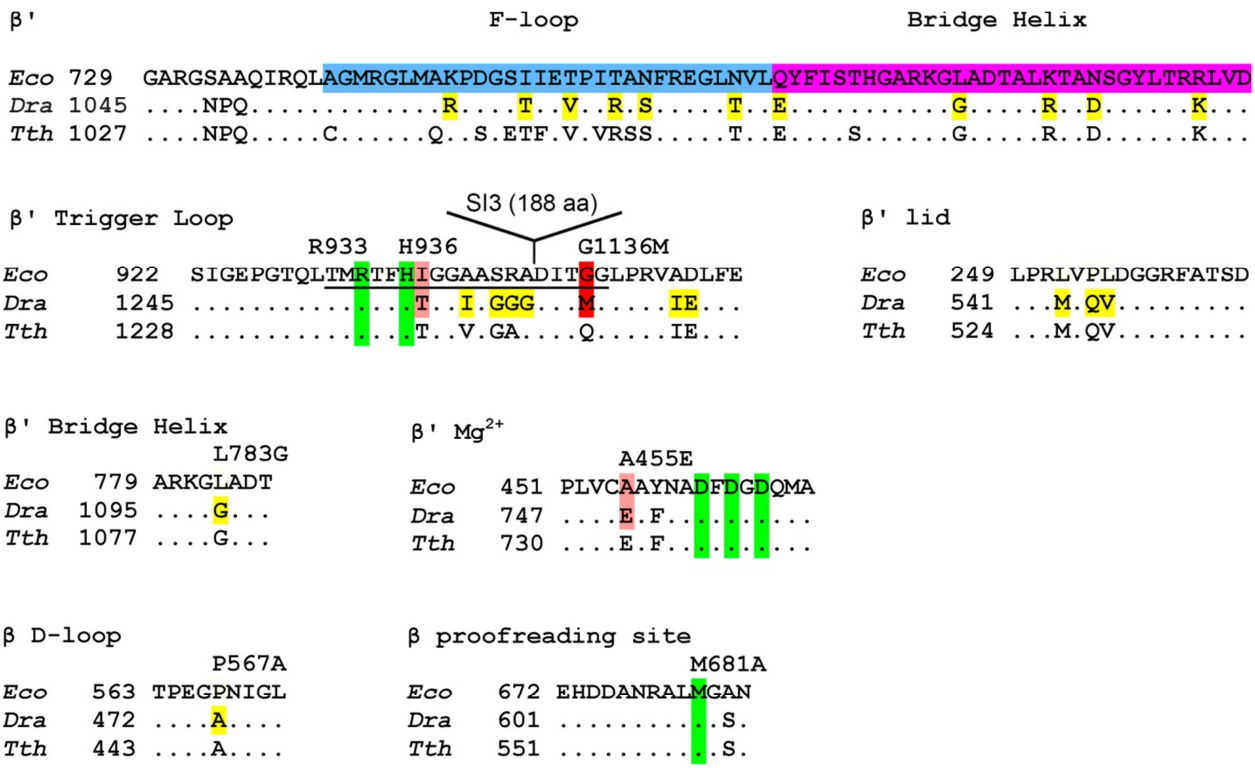


Figure S7. Alignments of the β′ and β subunit regions replaced in *Eco* RNAP with *Dra* sequences. Regions replaced in the FL and BH are shown in blue and magenta, respectively. Amino acid substitutions introduced in mosaic *Eco* RNAP variants are shown in yellow; substitutions that affect RNA cleavage are shown in red/pink. Conserved residues implicated in RNAP catalysis are shown in green.

Table S1. Kinetic parameters of RNA cleavage by *Eco*, *Dra* RNAPs and mutant *Eco* RNAP variants in the rA-dT2 TEC.

RNAP	Cleavage rate (rA-dT2)	
	k_{obs} (min^{-1}) 20°C	Fold change
Eco	0.029 ± 0.007	**1**
Dra	0.83 ± 0.13	**28.3**
Eco*-lid-*Dra	0.038 ± 0.01	**1.3**
Eco*-F-*Dra	0.011 ± 0.002	**0.36**
***Eco*-β′L783G**	0.035 ± 0.01	**1.2**
***Eco*-βP567A**	0.037 ± 0.01	**1.2**

The reactions were performed at 20°C and 10 mM $MgCl_2$. The *Eco*-F-*Dra* RNAP contained substitution of the whole FL-BH segment.

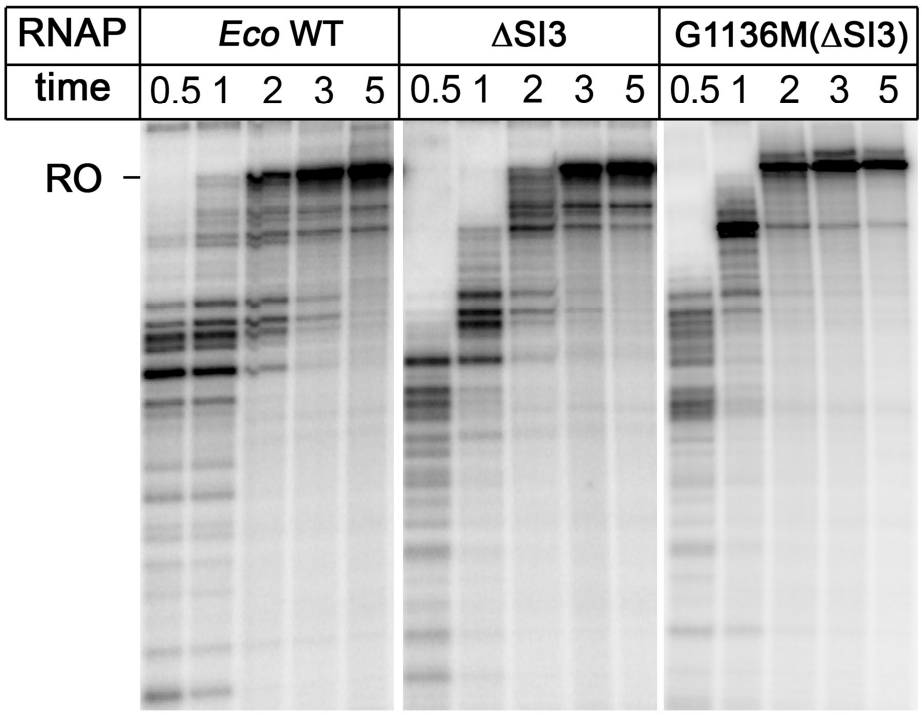


Figure S8. Elongation time course analysis of *Eco*, *Eco* ΔSI3 and *Eco*-G1136M(ΔSI3) RNAPs. TECs stalled at position +26 of the λP_R-*rpoB* template were obtained by the addition of a limited substrate set to preformed promoter complexes at 37 °C. The samples were transferred to 20 °C and transcription was restarted by the addition of all four NTPs (200 μM each) and heparin (15 μg/ml). The reactions were terminated after different time intervals and RNA products were analyzed by 10% denaturing PAGE. The position of the 500 nt run-off RNA (RO) is indicated.

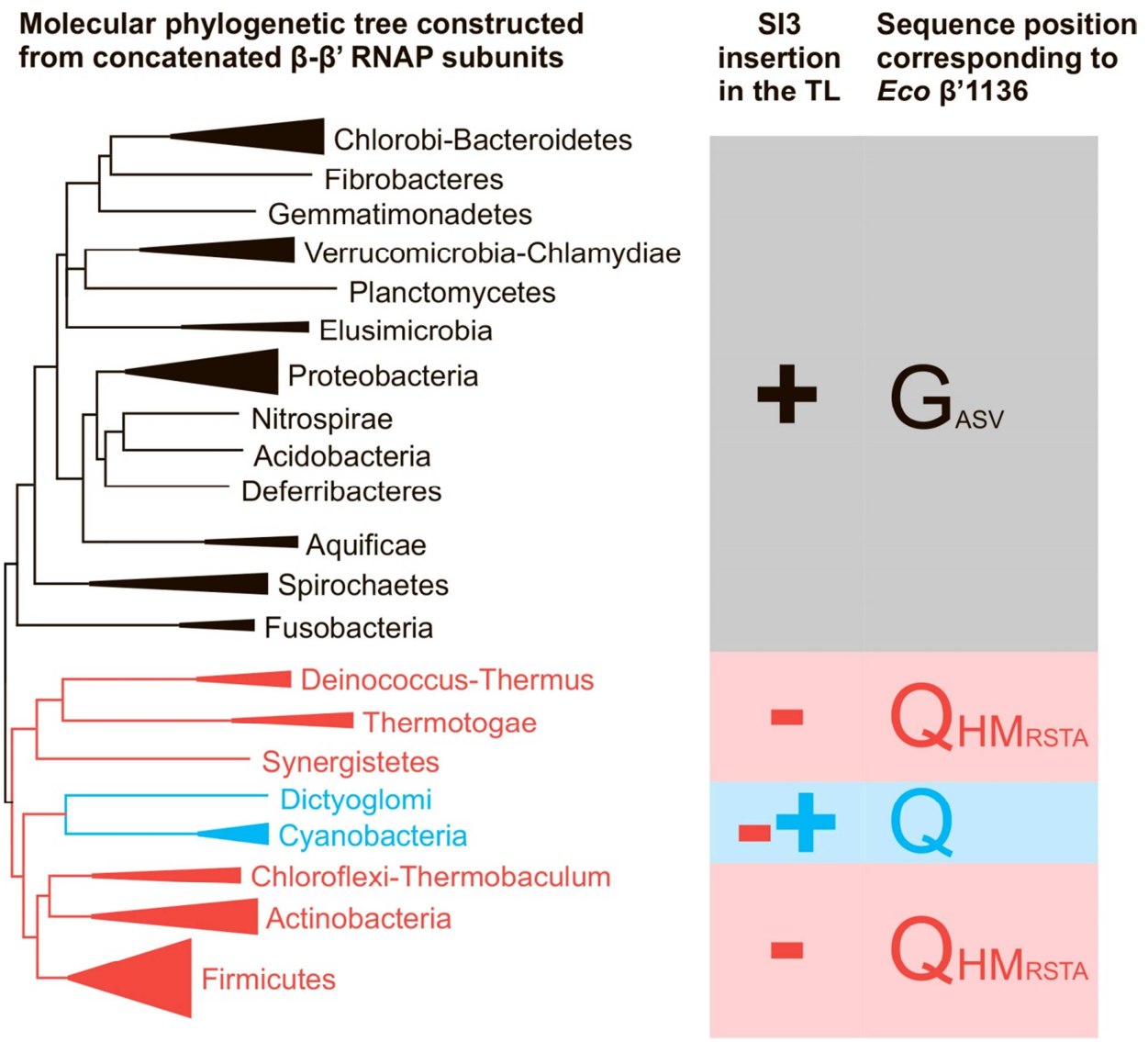


Figure S9. Amino acid substitutions at the RNAP β′ subunit position corresponding to *Eco* G1136 in various bacterial lineages. The presence of a Gly residue at this position correlates with the presence of the SI3 insertion in the TL. The molecular phylogenetic tree was constructed using MrBayes (v. 3.1.2) from 1882 amino acid x 76 sequences block extracted by Gblocks (v. 0.91b) from Muscle (v. 3.6) alignment of bacterial RNAP sequences (β-β′ concatenations). The height of the triangles at the vertexes of the tree branches is proportional to the number of sequences representing the phylum during the analysis. Bacterial phyla possessing RNAPs with and without the SI3 insert in the TL are colored black and red, respectively. Cyanobateria and Dictyoglomi phyla are colored cyan because their position in the molecular phylogenetic tree and the amino acid sequences of their SI3 inserts suggest that the latter are distinct from the SI3 insert in other bacterial phyla (black). The rightmost column depicts amino acid residues observed at the position corresponding to β′G1136 of *Eco* RNAP. The predominant amino acid (observed in >90 % of sequences from the corresponding clades) is depicted with the largest font. Other amino acids found in the position are depicted with the smaller fonts. The size of the font correlates with the prevalence of the corresponding amino acid but is not proportional to the latter to maintain legibility.