Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase

Vladimir Svetlov\textsuperscript{1}, Dmitry G. Vassylyev\textsuperscript{2,3}, Irina Artsimovitch\textsuperscript{1*}

\textsuperscript{1}Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, OH 43210, U. S. A.

\textsuperscript{2}Structural and Molecular Biology Laboratory and \textsuperscript{3}Structurome Research Group, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan

* To whom correspondence should be addressed.
artsimovitch.1@osu.edu
SUMMARY

Nucleic acid polymerases have evolved elaborate mechanisms that prevent incorporation of the non-cognate substrates, which are distinguished by both the base and the sugar moieties. While the mechanisms of substrate selection have been studied in single-subunit DNA and RNA polymerases (DNAPs and RNAPs), the determinants of substrate binding in the multi-subunit RNAPs are not yet known. Molecular modeling of *Thermus thermophilus* RNAP/substrate NTP complex identified a conserved β’ subunit Asn737 residue in the active site that could play an essential role in selection of the substrate ribose. We utilized the *Escherichia coli* RNAP model system to assess this prediction. Functional *in vitro* analysis demonstrates that the substitutions of the corresponding β’Asn458 residue lead to the loss of discrimination between ribo- and deoxyribonucleotide substrates as well as to defects in RNA chain extension. Thus, in contrast to the mechanism utilized by the single-subunit T7 RNAP where substrate selection commences in the inactive pre-insertion site prior to its delivery to the catalytic center, the bacterial RNAPs likely recognize the sugar moiety in the active (insertion) site.
INTRODUCTION

Copying (replication) and read-out (transcription) of genetic information contained within cellular genomes are carried out by an array of DNAPs and RNAPs, respectively. These enzymes operate within the same compartment, accessed by both types of substrates, r- and dNTPs, and control their fidelity at both co- and post-synthetic steps in order to avoid the so-called ‘error catastrophe’, when the amount of mistakes exceeds redundancy of the system. The first mechanism relies on the discrimination against the “wrong” nucleotide, whereas the second mechanism is invoked after the incorporation of a mismatched nucleotide.

The initial selection of the cognate substrate is comprised of two parts: the selection of the nucleotide complementary to the template DNA base and the selection of the correct sugar. The mechanism for sugar selection has been extensively studied in several DNAPs and in the single-subunit T7 RNAP. DNAPs actively discriminate against the rNTP via a "steric gate" formed by the Glu and Phe side chains (1-5), which sandwich the substrate sugar moiety and exclude the 2’-OH, while the substrate is positioned in the insertion site that has "closed" (active) configuration. In single-subunit T7 RNAP, Tyr639 hydroxyl has been implicated in the positive selection of the ribose via formation of a hydrogen with the 2’-OH group (6-8). In contrast to DNAPs, T7 RNAP commences the substrate selection in the inactive "open" conformation, while the substrate and the Tyr639 side chain are located in the so called pre-insertion site far away from the catalytic center (9,10). This conclusion is consistent with the mutational analysis of T7 RNAP where amino acid changes that affect transcription fidelity cluster around the pre-insertion site (6,7,11). By contrast, no mutations affecting the fidelity or discrimination between r- and dNTPs by multi-subunit RNAPs have been characterized.

Single- and multi-subunit RNAPs possess no sequence or apparent structural similarity,
yet carry out the transcription cycle in a nearly identical manner and could utilize analogous structural elements during catalysis (12-14). These similarities suggest that the basic substrate selection mechanism might also be conserved, and that a residue analogous to the Tyr639 would play a critical role in maintaining the ribose specificity in multi-subunit RNAPs. However, modeling of the substrate NTP bound to the *T. thermophilus* RNAP active site (15) suggests that β’ Asn458 (*E. coli* numbering is used throughout) within a highly conserved sequence motif NΔ458ADFDGD464 that includes the catalytic Asp triad (β’Asp460, 462, 464) (16-18) could mediate specific recognition of the O2’ ribose atom (Fig. 1A). Thus, Asn458 is likely not a structural analog of Tyr639, as the latter initially recognizes the ribose in the pre-insertion site.

To evaluate the role of β’ Asn458 in substrate selection by a bacterial RNAP we prepared and tested the conservative substitutions of this residue for their effects on utilization of the non-cognate substrates by the *E. coli* RNAP. Here we demonstrate that, in accordance with the structural predictions, these substitutions lead to a loss of discrimination of sugar moiety.
EXPERIMENTAL PROCEDURES

Reagents - All general reagents were obtained from Sigma and Fisher; rNTPs and dNTPs from Amersham, PCR reagents from Eppendorf, and restriction and modification enzymes from NEB. Oligonucleotides were obtained from Integrated DNA Technologies, $[^{32}\text{P}]-\text{NTPs}$ from NEN. Plasmid DNAs and PCR products were purified using spin kits from Qiagen.

Construction and purification of altered RNAPs - The *E. coli* *rpoC* gene was subjected to site-directed mutagenesis with two fully complementary oligonucleotides that defined the desired mutation. The shortest corresponding restriction fragment of the *rpoC* gene was completely sequenced at Genewiz Inc., and transferred into the overexpression vector pIA299, which encodes the *rpoA-rpoB-rpoC* gene cassette under the control of T7 promoter. The wild-type and altered RNAPs were expressed and purified as described previously (19).

In Vitro Transcription Reactions - All templates for transcription reactions were generated by PCR amplification. Plasmid pIA171 encodes a 29-nt U-less transcribed region under control of T7A1 promoter followed by a *his* pause site (20). Plasmid pIA349 encodes a 37-nt U-less transcribed region under control of T7A1 promoter (21). To form halted TECs, linear DNA template (40 nM), RNAP (50 nM), ApU (100 µM), and starting NTPs (2.5 µM ATP and GTP, 1 µM CTP, 10µCi of $[^{32}\text{P}]-\text{CTP}$ (3000 Ci/mmol)) were mixed on ice in 50 µl of TGA buffer (20 mM Tris-acetate, 20 mM Na-acetate, 10 mM Mg-acetate, 5% glycerol, 14 mM 2-mercaptoethanol, 0.1 mM EDTA, pH 8.0). Reactions were incubated at 37°C for 15 min and purified from nucleotides by gel-filtration through G50 spin columns (GE Health) equilibrated in TGA buffer. Transcription was restarted by addition of substrates indicated in figure legends, samples were removed at selected times and quenched by the addition of an equal volume of STOP buffer (10 M urea, 20 mM EDTA, 45 mM Tris-borate; pH 8.3). Samples were heated for 2
min at 90 °C, separated by electrophoresis in denaturing 10% acrylamide (19:1) gels (7 M Urea, 0.5X TBE), visualized and quantitated using PhosphorImager (GE Health) and bundled software.
RESULTS

Experimental set-up - To determine whether β′ Asn458 residue is critical for the interaction with the 2’ OH, we changed Asn458 to Asp and Ser; these are the conservative changes that would not be expected to substantially alter the structure of the protein but would change either the chemical properties (Asp) or the size (Ser) of the discriminating Asn side chain. The altered enzymes were overexpressed from a polycistronic vector that allows assembly of the core αββ′ RNAP in vivo (19), purified, and tested for the ability to discriminate between the cognate and non-cognate substrates. Similarly purified wild-type (WT) RNAP was used as a control.

We assembled transcription elongation complexes (TECs) on pIA349 template that encodes a T7A1 promoter (21). On this template, TECs can be initially halted at position 37 by withholding UTP from the reaction mix (Fig. 2A), purified by gel-filtration to remove the unincorporated substrates, and then “walked” to the next template position during addition of a subset of NTPs. This approach allows to test whether RNAP will efficiently incorporate a given substrate, and whether it will extend the nascent RNA transcript after a misincorporation event. We used a similar approach to measure rC/dC selectivity using pIA171 template (20) instead.

Substitution of Asn458 leads to increased dNTP utilization - We formed a halted TEC on an appropriate template and tested extension of 32P-labeled RNA upon addition of different unlabeled substrates. For each r/dNTP pair assayed, selected substrate was added to a concentration ranging between 0.1-62.5 µM for NTPs and 2-1250 µM for dNTPs, reactions were allowed to proceed for 2 min at 37°C, and quenched with STOP buffer (Fig. 2B).

As can be seen from representative gel images (Fig. 2B), the WT RNAP exhibits strong preference towards rNTP substrates, while in the N458D variant these preferences are relaxed.
For example, a 2-min incubation with 0.1 µM rGTP led to efficient extension of the nascent RNA by the WT RNAP, whereas ~100-fold higher concentration of dGTP was required to achieve the same degree of extension. For the purpose of quantitative comparison we selected the ratio between concentrations of dNTP and rNTP, when half of the transcripts are extended by one nucleotide, as the discrimination quotient. The selectivity of incorporation of different r/d NTP pairs by WT RNAP ranges from 120- to 1200-fold (Fig. 2C) suggesting that the substrate selection depends not only on the identity of the sugar moiety but also on the base structure, even when the Watson-Crick base pairing is maintained. Similar variations have been reported for T7 RNAP (6) and DNAP I (2); in the latter case selectivity correlates with the strength of base pairing, consistent with the proposed mechanism to selectively stabilize incoming A and T substrates to balance their weaker interaction with the template base (22). The same could also be true for the *E. coli* RNAP as discrimination capacity generally increases with the weakening of the base pair except for rU/dT selection, which could be additionally influenced by the presence of the 5-methyl group on thymine.

As expected, substitution of Asn458 for Asp led to substantially relaxed sugar discrimination ranging from 4.8-fold loss of discrimination for rG/dG pair to 26-fold loss in case of rU/dT pair. Substitution of Asn458 for Ser led to a smaller loss of preference for the ribose, whereas substitution of an adjacent Tyr457 residue did not alter sugar selection (Fig. 2C). The maximal observed effect for N458D RNAP was somewhat less than that reported for the T7 Y639F RNAP (6,8); the base-specific order of r/dNTP selectivity was also different between the bacterial and T7 enzymes (Fig. 2C and ref(6)). The direct comparison of these data is not straightforward, however, as we have used pre-steady state assays of substrate incorporation by the TEC, whereas Sousa *et al* used multi-round assays that include both the initiation and
elongation steps, which resulted in different discrimination efficiencies for the same RNAP depending on the assay design (6,8). In T7 RNAP Tyr to Phe substitution likely completely disrupts polar discriminative interactions with the substrate ribose, whereas Asn for Asp and Ser substitutions in the bacterial enzyme might still maintain specific hydrogen bonding with rNTP sugar moiety. In addition, Asn458 might be not the only residue participating in the ribose recognition in the active site. Indeed, according to the modeling guanidinium group of a highly conserved in bacteria and eukaryotes β’Arg425 appears proximal to the substrate sugar (Fig. 1B). The experiments are now underway to test the role of β’Arg425 in substrate selection.

Molecular modeling of the multi-subunit enzyme pre-insertion site (10) suggested that *E. coli* β’Thr790 residue could play a role in ribose selection analogous to that of the T7 Y639 residue. However, the T790V RNAP not only did not loose preference for rNTP substrates, it actually exhibited more stringency in substrate selection relative to the WT, ranging from 1.1-fold effect in case of rC/dC to 4.8-fold effect for rU/dT (Fig. 2C). These data underscore the importance of Thr790 residue in substrate selection but argue against involvement of its hydroxyl group in positive selection of the substrate ribose; less conservative substitutions of Thr to Ala and Leu led to gross defects in catalysis (data not shown).

A greater loss of discrimination upon the Asn to Asp substitution suggests that charge distribution on the discriminating protein group is more important than the side chain size. In the absence of the experimental crystallographic data on the RNAP/substrate complex structure we propose a model that is certainly too tentative to predict the exact scheme of the hydrogen-bonding with the substrate but nonetheless allows to zero in on the three protein groups that likely interact with the substrate ribose: Asn458 side chain, Asn458 main chain carbonyl oxygen, and Arg425 guanidinium group. In the high-resolution crystal structure of the *T. thermophilus*
RNAP holoenzyme (17) these three groups form internal hydrogen bonding network (Fig. 1B). Interestingly, the main chain conformation of Asn458 does not fall in the most favorable region on the Ramachandran plot ($\phi=-68^\circ; \psi=77^\circ$). The stabilization of this unfavorable conformation likely comes from the hydrogen bonding of the Asn458 main chain carbonyl with Arg425 and with its own side chain (Fig. 1B). The latter interaction would prevent flipping of the Asn458 side chain amido group that might be crucial for the proper sensing of the substrate ribose. This network of interaction would be enhanced upon binding of the rNTP substrate. In the model, Arg425 as well as the main chain carbonyl oxygen and the side chain amide of Asn458 make contacts with the O3' group, whereas the Asn458 side chain oxygen likely recognizes O2' group of the substrate ribose. The modeling of N458S substitution showed that though the interactions with the sugar would be weakened due to the smaller size of the Ser side chain, its hydroxyl group may preserve the framework of protein-protein and protein-substrate contacts similar to that of Asn (data not shown). In contrast, the negatively charged Asp side chain would lack the interactions with the main chain carbonyl potentially affecting both the main chain and side chain conformations and thus distorting the optimal orientation of the discriminating residue. In addition, the acidic Asp458 side chain might also form a salt bridge with the adjacent Arg425 that would further perturb the Asp interactions with the substrate ribose.

Substitution of Asn458 leads to profound defects in transcript elongation - In addition to being instrumental for the elucidation of the molecular mechanisms of substrate selection, RNAP variants with altered substrate selection properties could be used to determine the contribution of individual transcript bases to recognition of regulatory signals. This approach is particularly important for the functional analysis of transient kinetic intermediates, such as those occurring during transcription termination, when artificial stalling of complexes leads to trapping of the
off-pathway species (23), and relies on the ability of nucleotide analogs to affect RNAP response to a particular signal (nucleotide analogs interference mapping; NAIM). This analysis requires that RNAP is able to incorporate substrate analogs bearing modifications at various positions (24). For example, T7 RNAP with double substitution Y639F/H784A has been used to demonstrate the effect of replacement of individual residues in the nascent RNA with dNTP analogs on termination (24). However, to apply this strategy one needs to ascertain that the mutation in RNAP, while allowing incorporation of unusual substrates, does not lead to a defect in recognition of the transcription signals intended for study.

We wanted to test if the N458D RNAP could prove similarly useful in mapping the requirements in the nascent transcript during elongation/termination. We studied the elongation properties of the N458D RNAP on pIA349 template that encodes several well-characterized pause sites (Fig. 3). We found that N458D substitution confers a strong elongation defect: the rate of transcription elongation was dramatically reduced (more than 20-fold), and the enzyme paused strongly early in the transcribed sequence. We conclude that N458D RNAP, albeit able to efficiently incorporate dNTPs into the nascent RNA, is not suitable for NAIM due to its profound elongation defects. N458S RNAP also displayed a reduced elongation rate but the defect was less pronounced (5-fold, Fig. 3 and data not shown). These results are also in a good agreement with our modeling. In the RNAP holoenzyme structure β’Arg425 makes strong hydrogen bonds with β’Asp464 from the catalytic triad. We presume that these interactions are crucial for the proper positioning of the Asp464 side chain that is likely required to optimize the coordination of the major catalytic Mg^{2+} ion (cMg1, Fig. 1B). In N458D enzyme side chains of Arg425 and Asp458 may form a salt bridge (see above), thereby disrupting functionally important Arg425/Asp464 contacts and subsequently violating proper Mg^{2+} coordination and
compromising catalysis. Consistently, the smaller polar Ser residue would not affect strongly the orientation of the Arg425 side chain and would not exhibit dramatic effect on catalysis.

Asn458 is located on the catalytic loop next to the active site. Thus the alternative possible explanation of the observed functional alterations in the substrate selection and catalysis exhibited by the mutant enzymes would be allosteric effects on the active site configuration conferred by substitutions of Asn to the other residues. However, substitutions of an adjacent Tyr457 residue for His or Val did not confer either altered substrate selection or defects in transcription elongation (Fig. 2C and data not shown).
DISCUSSION

The major conclusion of this work is that β’ Asn458 residue in the \textit{E. coli} RNAP provides the recognition of sugar moiety on the incoming NTP substrate. No significant changes in misincorporation of non-templated NTP substrates arose from Asn458 substitutions (see Supplemental Data), indicating that the determinants for the sugar and base selection are non-overlapping, as was observed for the single-subunit T7 RNAP (11). The proposed ribose selection mechanism would be also likely valid for eukaryotic RNAPs that contain a highly conserved counterpart of β’Asn458; indeed this role for the corresponding Asn479 residue in Rpb1 had been proposed earlier (14).

For both T7 and \textit{E. coli} RNAPs the selectivity is far below the values reported for the DNAPs that exclude the rNTP substrates with several thousand- to a million-fold efficiency (1,2). These differences could be explained, on one hand, by the fact that the levels of rNTPs are at least 10-fold higher in the cell than the levels of the corresponding dNTPs (25). Thus, to prevent incorporation of rNTPs that could prove to be lethal (26), DNAPs must impose the strict discrimination mechanism, which is achieved by the steric exclusion of the ribose 2’ hydroxyl. On the other hand, RNAPs might face the opposite problem: the relatively inefficient (e.g. via a single H bond) discrimination between r and dNTPs could result not only in the synthesis of compromised messages but also in “draining” of the dNTPs pool. Cellular RNAPs would be expected to evolve tighter control mechanisms as compared to the phage ones since at the time phage RNAP becomes engaged in transcription the host cell is usually moribund.

In DNAPs, the substrate recognition is thought to occur exclusively in the insertion site (27). In contrast, in the T7 TECs substrate can be bound in either the pre-insertion or in the
insertion sites located 10 Å apart, suggesting that RNAPs may select substrates in two rather than in a single site (9,10). The presence of a pre-insertion site in which the substrate can be “sampled” prior to catalysis was also proposed recently for multi-subunit RNAPs (10,28). Our present data on Asn458, which is adjacent to the active site, suggest that in bacterial RNAPs substrate selection, at least in part of rNTP/dNTP selection, occurs in the insertion site. Thus the sugar selection may principally occur in the pre-insertion site in T7 but in the insertion site in bacterial RNAPs. If discrimination in favor of rNTP binding were to occur predominantly in the insertion site, persistence of these interactions after catalysis might hinder the movement of the incorporated nucleotide from the n to (n-1) site, thereby slowing translocation and the rate of polymerization. In T7 RNAP, where the pre-insertion site binding seems to be preferable, the Tyr639-2’OH contact might be compromised during the transition to the closed form. This might explain a faster rate of T7 RNAP transcription as compared to the multi-subunit cellular enzymes (29) which, facing more regulatory and fidelity constraints, might have evolved somewhat different mechanisms for substrate selection.

One cannot, however, rule out the possibility that rNTP/dNTP discrimination occurs both in pre-insertion and insertion sites by the different sets of residues. Indeed, substitution of the E. coli β’Thr790 that is located in the bridge helix and likely belongs to the putative pre-insertion site for Val not only did not decrease the enzyme preference for the ribose, but instead increased discrimination up to ~5-fold (Fig. 2C), implying the direct interactions of β’Thr790 (or an adjacent) residue with the substrate. Since β’Thr790 is located 18Å away from the active (insertion) site and ~8Å from the modeled substrate ribose, this result provides strong support to the hypothesis of existence of the substrate pre-insertion site in multi-subunit RNAPs. It is therefore possible that some other residue from the pre-insertion site may play a ribose
discriminating role similar to that of Asn458 in the insertion site. Alternatively, upon the substrate binding to the insertion site the structural elements that we are now assigning to the pre-insertion site might move towards the RNAP active center to constitute a single "closed" insertion site. We are currently dissecting the roles of residues in the pre-insertion site of the bacterial RNAP in selection of sugar and the base moieties of the incoming substrate.

ACKNOWLEDGEMENTS

The authors apologize for any omission of references due to space restrictions. This work was supported in part by grant from National Institutes of Health (to IA) and by RIKEN Institute (to DGV).

*Keywords*: RNA polymerase, substrate selection, transcription

*Running Title*: Substrate selection by bacterial RNA polymerase
FIGURE LEGENDS

**Fig. 1.** A model of the substrate bound to the *T. thermophilus* RNAP active (insertion) site. The numbering corresponds to the *E. coli* RNAP. The bridge helix (orange) is shown in the uniform conformation that has been observed in the yeast RNAP (16); in an alternative distorted state the central portion of the helix (dark blue) is flipped out (17). The DNA template strand and the RNA nucleotides in the RNA/DNA hybrid are shown in red and yellow, respectively. Two catalytic Mg$_2^+$ ions (cMg1 and cMg2, magenta spheres) are coordinated (white dashed lines) by the four catalytic RNAP Asp residues (white) and by the substrate phosphates. β’ Asn458 (cyan) located in the β’ active site loop (white ribbon) and β’ Arg425 (cyan) forming internal hydrogen bonding network in the crystal structure (cyan dashed lines) make putative contacts (cyan dashed lines) with the substrate ribose. (A). Overall view of the substrate-binding site. (B). Close-up view of the active site.

**Fig. 2.** Assay for the incorporation of the r/dNMPs. (A) Linear pIA349 template used to generate radiolabeled TECs halted at position G37 with the start site indicated by an arrow) (top) and the schematic representation of the assays used to measure the utilization of r/dNTP substrates (bottom). (B). Representative gel panels illustrating the extension of the nascent RNA in complexes halted at A39 or U38 upon addition of the increasing concentrations of r/dGTP and r/dATP substrates. Selection of rUTP vs dTTP was measured by extension of G37 RNA; selection of r/dCTP was assayed on a similar T7A1 promoter template, pIA171. (C). Comparison of the discrimination efficiencies between four r/d nucleotide combinations. The assays were repeated 2-4 times for each enzyme-r/dNTP combination, the discrimination efficiencies varied within 20%; NT, not tested. In *E. coli* RNAP, all substitutions are in the β’ subunit; the effects of selected substitutions in T7 RNAP (6) are presented for comparison.
FIG. 3. Transcription elongation by the N458 mutant enzymes. (Top) Transcript generated from the T7A1 promoter on a linear pIA349 template; transcription start site (+1), transcript end (run-off) and positions of $ops$ (U43) and $P_{his}$ (U145) pause sites are indicated. (Bottom) Halted G37 TECs were challenged with heparin at 50 µg/ml and NTPs at low (20 µM GTP, 100 µM ATP, CTP, UTP) or high (100 µM GTP, 500 µM ATP, CTP, UTP) concentrations. Aliquots were withdrawn at times indicated above each lane, followed by the high NTP chase (1 mM each NTP; C) and quenched as above. Positions of the DNA size markers ($^{32}$P-labeled pBR $Msp$I digest) are shown on the right (in nt).
REREFENCES

1. Joyce, C. M. (1997) Proc Natl Acad Sci U S A 94, 1619-1622
2. Astatke, M., Ng, K., Grindley, N. D., and Joyce, C. M. (1998) Proc Natl Acad Sci U S A 95, 3402-3407
3. Boyer, P. L., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (2000) Proc Natl Acad Sci U S A 97, 3056-3061
4. Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A., and Goff, S. P. (1997) Proc Natl Acad Sci U S A 94, 407-411
5. Yang, G., Franklin, M., Li, J., Lin, T. C., and Konigsberg, W. (2002) Biochemistry 41, 10256-10261
6. Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997) Biochemistry 36, 8231-8242
7. Brieba, L. G., and Sousa, R. (2000) Biochemistry 39, 919-923
8. Sousa, R., and Padilla, R. (1995) Embo J 14, 4609-4621
9. Yin, Y. W., and Steitz, T. A. (2004) Cell 116, 393-404
10. Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S., and Vassylyev, D. G. (2004) Cell 116, 381-391
11. Huang, J., Brieba, L. G., and Sousa, R. (2000) Biochemistry 39, 11571-11580
12. Yin, Y. W., and Steitz, T. A. (2002) Science 298, 1387-1395
13. Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G., and Yokoyama, S. (2002) Nature 420, 43-50
14. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A., and Kornberg, R. D. (2001) Science 292, 1876-1882
15. Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M. N., Hosaka, T., Ochi, K., Yokoyama, S., and Vassylyev, D. G. (2004) Cell 117, 299-310
16. Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001) Science 292, 1863-1876
17. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002) Nature 417, 712-719
18. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999) Cell 98, 811-824
19. Artsimovitch, I., Svetlov, V., Murakami, K. S., and Landick, R. (2003) J Biol Chem 278, 12344-12355
20. Artsimovitch, I., and Landick, R. (2000) Proc Natl Acad Sci U S A 97, 7090-7095
21. Artsimovitch, I., and Landick, R. (2002) Cell 109, 193-203
22. Donlin, M. J., and Johnson, K. A. (1994) Biochemistry 33, 14908-14917
23. Kashlev, M., and Komissarova, N. (2002) J Biol Chem 277, 14501-14508
24. Schwartz, A., Rahmouni, A. R., and Boudvillain, M. (2003) Embo J 22, 3385-3394
25. Kornberg, A., and Baker, T. (1992) DNA replication, Freeman and Company, New York
26. Gao, G., and Goff, S. P. (1998) J Virol 72, 5905-5911
27. Kunkel, T. A., and Bebenek, K. (2000) Annu Rev Biochem 69, 497-529
28. Holmes, S. F., and Erie, D. A. (2003) J Biol Chem 278, 35597-35608
29. Uptain, S., Kane, C., and Chamberlin, M. (1997) Annu. Rev. Biochem. 66, 117-172
A. 

![Diagram of RNA polymerase (RNAP) and DNA interaction with A-U+ A+G to G37 and gel filtration]

B. 

![Gel electrophoresis images showing differences between wild-type RNAP and N458D RNAP with respect to A39-G40, A39-U38, and G37 concentrations]

C. 

| RNAP      | A  | C  | U  | G  |
|-----------|----|----|----|----|
| EC WT     | 1200 | 850 | 550 | 120 |
| EC N458D  | 100  | 110 | 20  | 25 |
| EC N458S  | 660  | 290 | 300 | 60 |
| EC Y457H  | 1300 | 810 | 530 | NT |
| EC T790V  | 2400 | 950 | 2500 | 540 |
| T7 WT     | 121  | 89  | 34  | 60 |
| T7 Y639F  | 5    | 7.5 | 2   | 2  |
| T7 Y639V  | 21   | 21  | 26  | 11 |
SUPPLEMENTARY DATA

Substitution of Asn458 does not alter misincorporation

The recognition of the cognate substrate relies on selection of both the correct sugar moiety and the complementary base. To analyze the miscoding properties of N458D RNAP we have utilized a similar approach where we added a large excess of a non-complementary rNTP substrate and monitored the RNA chain extension by the WT and altered enzymes (Fig S1A). As expected, RNAP incorporates a wrong base very reluctantly; however, no notable differences were observed among the three enzymes. This result is reminiscent of the observations made for the T7 Y639F RNAP and phage φ29 Y254V DNAP, which display a relaxed sugar discrimination but have no defects in base recognition (1,2).

Substitution of Asn458 does not alter extension of the dNTP primer

Upon addition of the wrong substrate nucleotide, RNAP could also fail to extend the misincorporated transcript efficiently and edit it through the preferential cleavage of the compromised 3’ end fragment. This mechanism was proposed to contribute to fidelity of transcription in both bacterial and eukaryotic systems (3,4). We tested if the substitution of β’ Asn458 would alter the efficiency of the extension of 3’ deoxynucleotide (Fig S1B). Again, the extension efficiency of misincorporated dNMP did not differ significantly between the WT and mutant enzymes.
FIGURE LEGENDS

FIG. S1. (A) Misincorporation of rAMP in place of rGMP. The schematic representation of the assay used to measure the misincorporation propensity (top). Linear pIA349 template (ref. 5 and Fig. 2) was used to generate radiolabeled TECs halted at position G37, which are indicated as 0 complexes. To generate TECs halted at position 38, UTP was added to 5 µM followed by a 5-min incubation at 37°C; the U38 complexes are marked as 0*. Assay was performed as in Fig 2 except that a higher concentration of ATP (25µM) and prolonged incubation times were used to 'force' misincorporation. A representative 10% denaturing gel with reactions generated by the wild-type and N458D enzymes; positions of each transcript are indicated.

(B). Extension of the nascent RNA following incorporation of a single dTMP residue. Linear pIA349 template was used to generate radiolabeled G37 TECs (marked as 0). The schematic representation of the assay used to measure extension of a single dTMP residue at position 38 is shown on top; T38 TEC is marked as 0*. Reaction was initiated by the addition of ATP at 0.1 µM, aliquots were withdrawn at times indicated above each lane and quenched as above. A representative 10% denaturing gel with reactions generated by the wild-type and N458D enzymes; positions of each transcript are indicated.

REFERENCES
1. Huang, J., Brieba, L. G., and Sousa, R. (2000) Biochemistry 39, 11571-11580
2. Bonnin, A., Lazaro, J. M., Blanco, L., and Salas, M. (1999) J Mol Biol 290, 241-251
3. Thomas, M. J., Platas, A. A., and Hawley, D. K. (1998) Cell 93, 627-637
4. Erie, D. A., Hajiseyedjavadi, O., Young, M. C., and von Hippel, P. H. (1993) Science 262, 867-873
5. Artsimovitch, I., and Landick, R. (2002) Cell 109, 193-203
Figure S1.
