

# **Structural origins of *Escherichia coli* RNA polymerase open promoter complex stability**

**Ruth M. Saecker<sup>a,\*</sup>, James Chen<sup>a</sup>, Courtney E. Chiu<sup>a</sup>, Brandon Malone<sup>a</sup>, Johanna Sotiris<sup>b</sup>, Mark Ebrahim<sup>b</sup>, Laura Y. Yen<sup>c</sup>, Edward T. Eng<sup>c</sup>, Seth A. Darst<sup>a,\*</sup>**

<sup>a</sup>Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10065 USA.

<sup>b</sup>The Evelyn Gruss Lipper Cryo-Electron Microscopy Resource Center, The Rockefeller University, 1230 York Avenue, New York, NY 10065 USA.

<sup>c</sup>The National Resource of Automated Molecular Microscopy, Simons Electron Microscopy Center, New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027 USA

\*Corresponding authors: Ruth M. Saecker, Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA. [rsaecker@rockefeller.edu](mailto:rsaecker@rockefeller.edu). 608-234-8038; Seth A. Darst, Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA. [darst@rockefeller.edu](mailto:darst@rockefeller.edu). 212-327-7479.

**Email:** [rsaecker@rockefeller.edu](mailto:rsaecker@rockefeller.edu) or [darst@rockefeller.edu](mailto:darst@rockefeller.edu)

**Author Contributions:** Conceptualization, R.M.S. and S.A.D.; Methodology, R.M.S., J.C., C.E.C. and S.A.D.; Investigation, R.M.S., J.C., C.E.C, B.M., J. S., M.E., L.Y.Y., and E.T.E.; Writing – Original Draft, R.M.S. and S.A.D.; Writing – Review & Editing, R.M.S. and S.A.D.; Supervision, S.A.D.; Funding Acquisition, R.M.S. and S.A.D.

# Abstract

The first step of gene expression in all organisms requires opening the DNA duplex to expose one strand for templated RNA synthesis. In *Escherichia coli*, promoter DNA sequence fundamentally determines how fast the RNA polymerase (RNAP) forms “open” complexes (R<sub>Po</sub>), whether R<sub>Po</sub> persists for seconds or hours, and how quickly RNAP transitions from initiation to elongation. These rates control promoter strength *in vivo* but their structural origins remain largely unknown. Here we use cryo-electron microscopy to determine structures of R<sub>Po</sub> formed *de novo* at three promoters with widely differing lifetimes at 37°C:  $\lambda$ P<sub>R</sub> ( $t_{1/2} \sim 10$  hours), T7A1 ( $t_{1/2} \sim 4$  minutes), and a point mutant in  $\lambda$ P<sub>R</sub> ( $\lambda$ P<sub>R-5C</sub>) ( $t_{1/2} \sim 2$  hours). Two distinct R<sub>Po</sub> conformers are populated at  $\lambda$ P<sub>R</sub>, likely representing productive and unproductive forms of R<sub>Po</sub> observed in solution studies. We find that changes in the sequence and length of DNA in the transcription bubble just upstream of the start site (+1) globally alter the network of DNA-RNAP interactions, base stacking, and strand order in the single-stranded DNA of the transcription bubble; these differences propagate beyond the bubble to upstream and downstream DNA. After expanding the transcription bubble by one base (T7A1), the nontemplate-strand “scrunches” inside the active site cleft; the template-strand bulges outside the cleft at the upstream edge of the bubble. The structures illustrate how limited sequence changes trigger global alterations in the transcription bubble that modulate R<sub>Po</sub> lifetime and affect the subsequent steps of the transcription cycle.

# Introduction

Transcription by DNA-dependent RNA polymerases (RNAP) releases information stored in duplex DNA in the form of RNA transcripts. Appropriate responses to changing cellular conditions and growth rates require rapid and tight control of cellular RNA levels. In the model organism *Escherichia coli*, the rate of productive initiation events largely determines RNA transcript amount (1). RNA chain initiation frequencies vary over four orders of magnitude *in vivo*; during exponential growth  $10^3$ - $10^4$  ribosomal transcripts are synthesized per generation whereas other transcripts may appear once or not at all (1-3). Intensive investigations of how *E. coli* achieves this extraordinary range are ongoing, recently revealing novel modes of transcription initiation (4) and termination (5, 6) despite decades of study.

In bacteria, a single “core” enzyme (E), comprising five subunits ( $\alpha_2\beta\beta'\omega$ ), catalyzes all templated phosphodiester bond synthesis. Operon-specific output is orchestrated by the addition of a sixth dissociable subunit sigma ( $\sigma$ ), forming the holoenzyme (E $\sigma$ ) [cf. (3, 7)]. The vast majority of studies have focused on the “housekeeping” group I  $\sigma$  factors [ $\sigma^{70}$  in *Eco*;  $\sigma^A$  in other bacteria (8)]. Regions of high sequence conservation in the  $\sigma^{70}$ -family (numbered sequentially 1.1-4.2) correspond to structural domains linked by flexible linkers (9, 10), each playing distinct roles in promoter DNA recognition and strand separation [cf. (3)].

Exposing and positioning the start site (+1) near the RNAP catalytic Mg<sup>2+</sup> in the RNAP open promoter complex (R<sub>Po</sub>) requires unwinding of over a turn of the DNA helix (11). Because binding free energy drives these steps, promoter DNA sequence

intrinsically determines how quickly RPo forms and how long it persists (1, 12-14). These kinetic differences critically underlie cell function. For example, differences in RPo lifetime allow the RNAP-binding factors dksA and ppGpp to “discriminate” between the operons they regulate and those they do not, and to have opposing effects at those they do (15, 16).

A striking discovery of ensemble and single molecule mechanistic investigations is that RPo is not a singular universal complex; multiple forms of RPo can exist at the same promoter (cf. (17-22)). At the model phage promoter  $\lambda P_R$ , a series of distinct open complexes form after the rate-limiting step [I1 to I2; (21, 22)]:



[Eq. 1. Minimal mechanism of RPo formation at  $\lambda P_R$ . Steps that contribute to the observed dissociation rate constant shown in red (11, 21)]. I2 is only transiently populated; conversions to I3 and RPo successively stabilize the strand-separated state. The structural and functional differences between these complexes are largely unknown.

While the conserved -35 and -10 elements and the length of the “spacer” between them affect RPo formation and thus its overall stability, both the sequence and length of the “discriminator” [Fig. 1A; (16)] has emerged as a primary determinant of RPo lifetime (23-25). Haugen et al. (23) demonstrated that the critical sequence runs from -6 to -4 (numbering with respect to the transcription start site at +1), with 5'-GGG-3' [nontemplate strand (nt-strand) sequence] yielding the longest RPo half-lives ( $t_{1/2}$ ) (23). The cornerstone of “discrimination” was pinpointed to -5 on the nt-strand, where the presence of G [G<sub>-5</sub>(nt)] increases RPo  $t_{1/2}$  by 10 to 50-fold at the *rrnB* P1,  $\lambda P_R$ ,  $\lambda P_L$  and *Pgal* promoters. Crosslinking, mutational and other biochemical approaches mapped these large effects on RPo stability to specific interactions with  $\sigma^{70}$  conserved region 1.2 ( $\sigma^{70}_{1.2}$ ) (26, 27). Subsequent work found that the  $\beta$  “gate loop” ( $\beta$ GL) also plays a crucial role in nt-strand discriminator interactions (28). How do these interactions dictate RPo lifetimes that range from seconds to many hours?

The understanding of how the transcription bubble is differentially stabilized is currently limited. The first atomic resolution structures of RPo determined by X-ray crystallography used promoters with consensus -35 and -10 promoter elements and pre-formed bubble templates with short regions of downstream duplex (29, 30) or short downstream “fork” constructs (31, 32). Resulting structures often revealed strand disorder; to improve resolution, NTPs or short RNA primers were added, forming transcription initiation complexes (RPinit). Until very recently (33-36), the structure of RPo was mostly inferred from these RPinit's formed at nonnative sequences/structures.

Similarly, detailed biochemical studies of RPo formation exist for only a small handful of promoters. Mechanistic and biochemical investigations of DNA opening at two phage promoters ( $\lambda P_R$  and T7A1) have largely defined the critical steps in forming the transcription bubble [cf. (11, 14, 37-39)]. However, no high resolution structural data exist for any complex formed at either promoter. Here we begin to address these gaps by using cryo-electron microscopy (cryo-EM) to visualize RPo formed *de novo* at  $\lambda P_R$

and T7A1. We also studied a single point mutant in  $\lambda P_R$  ( $\lambda P_{R-5C}$ ) that significantly decreases RPo half-life. Below we present the structural differences between these complexes, providing insights into how changes in promoter sequence, even at a single base, give rise to orders of magnitude changes in RPo lifetimes.

## Results

### Cryo-EM structures of $E\sigma^{70}$ RPo at the $\lambda P_R$ , $\lambda P_{R-5C}$ , and T7A1 promoters

To understand how promoter sequence dictates widely differing RPo lifetimes, we analyzed three *de novo* DNA-melted  $E\sigma^{70}$ -promoter complexes by single particle cryo-EM [T7A1,  $\lambda P_R$ , and a point mutant at -5 in  $\lambda P_R$  ( $\lambda P_{R-5C}$ ; Fig. 1A). To reduce particle orientation bias at the liquid-air interface, the cryo-EM buffer contained 8 mM CHAPSO (40). CHAPSO decreases RPo lifetime (2 to 3-fold) but does not change the relative stabilities of each RPo (SI Appendix, Fig. S1 and Table S1). Steps of maximum likelihood classification (41) revealed two distinct conformational classes populated at  $\lambda P_R$  (75% of the particles fall into class I, 25% in class II) whereas only a single class was found for T7A1 and  $\lambda P_{R-5C}$  (Fig. 1B; SI Appendix, Figs. S2-S8 and Table S2).

In all complexes: i. the transcription bubble is fully open; and ii. interactions with the -35 element, the spacer region and the -10 element bases on the nt-strand are largely the same. Because these interfaces do not significantly differ from each other or from previously reported structures [cf. (10, 29-31, 34, 42)] they will not be discussed further.

Relative to  $\lambda P_R$  class I, the DNA strands in transcription bubbles of the other RPo are more dynamic. Every single base in the nt- and t-strands within the transcription bubble is well-resolved in  $\lambda P_R$  class I (Figs. 1B and 2A) whereas the entire t-strand (-11 to +2) and -4 to +2 on the nt-strand have little to no map density in class II (Figs. 1B and 2B). Similarly, the mid-regions of the  $\lambda P_{R-5C}$  bubble (Figs. 1B and 2C) and the T7A1 t-strand from -4 to +2 (Figs. 1B and 2D) could not be modeled.

Duplex DNA regions distant from the transcription bubble also exhibit distinct differences. Unlike the relatively unstable RPos (SI Appendix, Fig. S1 and Table S1), an additional helical turn could be modeled downstream of +13 (to +23) in  $\lambda P_R$  class I. Upstream of the -35 element, map density for the two flexibly-tethered  $\alpha$ CTDs bound to DNA (-38 to -55) exists for all RPo but only the “proximal”  $\alpha$ CTD in  $\lambda P_R$  class I and T7A1 RPo was well-resolved [Fig. 2 and SI Appendix, Figs. S3 and S8, respectively; see also (43, 44)]. “Distal”  $\alpha$ CTD (UP) sequences exist at both promoters (3), prompting a focused classification of this region. This approach extracted classes with different DNA trajectories but not distinct bound states of the second  $\alpha$ CTD (cf. (35)). Thus despite the presence of UP elements in these promoters and in the ribosomal promoters *rpsT* P2 and *rrnB* P1, by the time RPo has formed, the DNA upstream of -45 is dynamic and the DNA binding mode of the second  $\alpha$ CTD is heterogeneous (34-36, 45).

Superposition of the cryo-EM RPo structures here and published high resolution cryo-EM RPo complexes [*rpsT* P2 (34) and *rmnB* P1 (36)] revealed small to moderate differences in the conformation of the clamp and the  $\beta$ lobe (SI Appendix, Table S3). With respect to  $\lambda P_R$  class I, the clamp and the  $\beta$ lobe are more open in all other RPo. Additional narrowing of the cleft in  $\lambda P_R$  class I appears to be stabilized by partial ordering of five  $\sigma^{70}$  residues ( $\sigma^{70}$ S85 to S89; SI Appendix, Fig. S9) on the  $\beta$ lobe. Extension across the  $\beta$ lobe creates a “clasp” between  $\beta'$  and  $\beta$  similar to that observed in mycobacterial  $E\sigma^A$  (33); the clasp must be undone for DNA to leave the cleft.

$\sigma^{70}$ S85 to S89 lie at the C-terminus of the 37-residue linker connecting  $\sigma^{70}_{1.2}$  to  $\sigma^{70}_{1.1}$  [ $\sigma^{70}_{1.1}$ -linker; (46)]. During RPo formation, DNA displaces  $\sigma^{70}_{1.1}$  bound in the cleft [cf. (35, 47)]. No density exists for  $\sigma^{70}_{1.1}$  in any RPo structure to date, indicating it does not rebind elsewhere. Instead, ejection creates a high local concentration of the flexibly-tethered  $\sigma^{70}_{1.1}$  above the cleft. Interactions between the linker and the  $\beta$ lobe may increase stability in  $\lambda P_R$  class I by directing  $\sigma^{70}_{1.1}$  away from the channel, effectively reducing its concentration near the downstream DNA and disfavoring its re-entry relative to other RPo (see SI Appendix for a discussion of relationship of class I and class II to the intermediates in Eq. 1).

## Differences in base stacking in the transcription bubble

To illustrate the differences in strand/base resolution and in the extent of base stacking, each transcription bubble is presented schematically in Fig. 2. At the upstream end, the nt-strand single-stranded -10 element hexamer exhibits the same conserved interactions seen in all RPo structures to date:  $A_{-11}(\text{nt})$  and  $T_{-7}(\text{nt})$  are flipped out into protein pockets on  $\sigma$ , the intervening bases from -10 to -8 stack with each other, facing into the channel with only the backbone atoms making interactions with  $\sigma$ . However, downstream of the -10 element, differences in the discriminator length and sequence impact both single-stranded base stacking and strand order. For  $\lambda P_R$  class I, almost every base in the bubble (nt- and t-strands) has a stacking partner (Fig. 2A). The single base change from G to C (nt-strand) at -5 of  $\lambda P_{R-5C}$  affects the entire bubble, disordering regions of both strands (Fig. 2C). Addition of an additional nucleotide to the discriminator (T7A1) and placing an A at the same position as  $G_{-5}(\text{nt})$  in  $\lambda P_R$  [ $A_{-6}(\text{nt})$  T7A1 numbering] disrupts all base stacking interactions in the nt-strand downstream of -5 (Fig. 2D).

## Interactions with the nt-strand: structural consequences of base identity at -5 and of a 7 base vs 6 base discriminator

How do three nt-strand bases out of all the bases that define a promoter profoundly effect RPo lifetime (23)? DNA opening positions bases from -6 to -4 near highly conserved residues in  $\sigma^{70}_{1.2}$ ,  $\sigma^{70}_2$  (9, 10) and the opposing  $\beta\text{GL}$  [cf. (31, 48, 49)]. These interactions close off the top of the active site channel, effectively trapping the strand-separated DNA inside (see SI Appendix, Figs. S3C, S4C, S6C and S8C). As detailed below, changes away from guanine (G) and increasing discriminator length significantly reduce the extent of these contacts, resulting in global changes in RPo structure.



At  $\lambda P_R$ , base stacking and pairing interactions from -6 to -4 in the duplex DNA are replaced in RPo by an extensive network of polar,  $\pi$ , and van der Waals contacts with highly conserved  $E\sigma^{70}$  residues (Fig. 3A). Notably,  $\sigma^{70}M102$ ,  $\sigma^{70}R99$ , and  $\beta R371$  interact with multiple bases in  $\lambda P_R$  class I RPo but not in T7A1 (Fig. 3B) or  $\lambda P_{R-5C}$  (SI Appendix, Fig. S11). The “keystone”  $G_{-5}(nt)$  (23) is tightly held by distinct chemical interactions that include two base-specific hydrogen bonds [ $\sigma^{70}R99(NE)-O6$  and  $\sigma^{70}D96(OD2)-N1$ ; Fig. 3A]. Replacing  $\lambda P_R G_{-5}(nt)$  with C not only eliminates these local contacts but abolishes the complex set of interactions that constrain the bases and sugar phosphate backbone from -6 to -4 in  $\lambda P_R$  (SI Appendix, Fig. S11). The quality and quantity of interactions in the T7A1 nt-strand interface are diminished as well (Fig. 3B). Because  $E\sigma^{70}$  intimately reads out the bases at -6 to -4 (Fig. 3A), deviations from the optimal sequence (G) as well as length alter the interface cooperatively. As a result, the driving force for the isomerizations that stabilize RPo changes nonadditively with sequence, yielding widespread differences in RPo structure (the extent of strand and downstream DNA order, clamp/lobe/ $\sigma^{70}_{1.1}$ -linker positions; Fig. 3, SI Appendix, Figs. S10 and S11, Table S3) and lifetime.

One key similarity unites all RPo characterized to date: the majority of the nt-strand discriminator contacts are with the upstream and downstream ends, leaving the mid-region fairly unconstrained and indeed observed to be dynamic in  $\lambda P_{R-5C}$  and in  $\lambda P_R$  class II. As seen in T7A1, adding another base to the canonical six base discriminator is accommodated by unstacking and flipping bases out of the backbone relative to one another in the “middle” of the discriminator (Fig. 3B). In this “scrunch”,  $A_{-4}(nt)$  and  $G_{-3}(nt)$  face into the RNAP channel where they are solvent-exposed, and  $C_{-2}(nt)$  occupies the downstream end of the channel where it interacts with  $\beta R201$ .

## DNA scrunches in the t-strand at the upstream end of the DNA channel in RPo

Entry of the t-strand into the active site channel at  $T_{-11}(t)$  [ $T_{-12}(t)$  in T7A1] distorts the DNA backbone, unstacking and flipping  $T_{-11}(t)$  out at  $\lambda P_R$  and both  $T_{-12}(t)$  and  $A_{-11}(t)$  at T7A1, placing the t-strand “scrunch” at the upstream end of the channel (Fig. 4). While map density exists for these bases in locally filtered maps, it is weaker than that for the corresponding base partners on the nt-strand, consistent with the accessibility of these conserved thymines to permanganate ions (20, 50). As the strand descends further into the cleft, both  $\lambda P_R$  and T7A1 exhibit an intriguing interaction with the single-stranded -10 element t-strand at  $A_{-10}(t)/T_{-9}(t)$  or  $T_{-10}(t)/G_{-9}(t)$ , respectively. In both RPo, these bases are stacked and captured by the loop between the  $\alpha$ -helices formed by  $\sigma^{70}_{2.1}$  and  $\sigma^{70}_{2.2}$ , and the opposing residues on the  $\beta$ protrusion (Fig. 4, SI Appendix, Fig. S12). At the upstream end of this “sandwich”,  $\sigma^{70}R397$  and  $\sigma^{70}N396$  interact with the -10 base, the -10 phosphate oxygen makes polar ionic contacts with  $\sigma^{70}R468$  ( $\sigma^{70}_3$  helix). At the downstream end, the -9 base and  $\beta R470$  form a cation- $\pi$  stacking interaction and  $\beta K496$  makes both ionic and nonpolar interactions with the DNA backbone (Fig. 4). Whether the “sandwich” causes the scrunch or whether it helps stabilize the scrunch is unknown. However, placing the additional base (relative to 6 base discriminator) at the

upstream end of the channel equalizes the length of the t-strand from -8 to the active site (see Discussion).

## **$\lambda P_R$ t-strand is largely stacked and tightly held in the active site channel**

The relatively high resolution of all the transcription bubble bases of the  $\lambda P_R$  class I RPo and of the residues in the active site channel allows a molecular visualization of the interactions that direct the t-strand from the ds/ss upstream fork to the  $Mg^{2+}$  active site on the “floor” of the cleft some 65 Å away. As illustrated in the stick/cartoon representation (Fig. 5A) and schematic (Fig. 5B), every base in the upstream half of the bubble makes multiple favorable interactions with at least one residue of  $\sigma^{70}$ ,  $\beta$  and/or  $\beta'$ . While flipped out of the helix at the ds/ss junction,  $T_{-11}(t)$  is not captured in a protein pocket like its base pairing partner on the nt-strand. Nonetheless,  $\sigma^{70}$  appears to constrain its position via numerous interactions with the base and backbone atoms: i. polar contacts between thymine and a trio of residues in the  $\sigma^{70}_3$  helix, including a H-bond between  $\sigma^{70}N461$  and  $N3$ ; ii. salt-bridges between the phosphate oxygens of  $C_{-12}(t)$  and  $T_{-11}(t)$  and  $\sigma^{70}R397$  and  $\sigma^{70}R468$ , respectively; and iii. multiple polar and nonpolar contacts with sugar atoms. With the exception of  $\sigma^{70}N461$ , these residues are invariant or nearly invariant in the housekeeping  $\sigma$  factors. Multiple interactions form between  $E\sigma^{70}$  and the bases and the sugar phosphate backbone from -11 to -6, including H-bonds with  $T_{-11}(t)$ ,  $A_{-7}(t)$ , and  $C_{-6}(t)$ . As seen for the nt-strand,  $\pi$  interactions form with unstacked bases at  $T_{-9}(t)$  ( $\beta R470$ ),  $C_{-5}(t)$  ( $\sigma^{70}F514$ ), and  $A_{-4}(t)$  ( $\sigma^{70}F522$ ).

Perhaps surprisingly, no positively charged (or negatively) charged groups fall within 4.5 Å (nor within 6 Å) of the phosphate oxygens from  $T_{-8}(t)$  to  $C_{-3}(t)$ . By contrast, all interactions with the t-strand DNA near the active site (-2 to +2) are with the DNA backbone [with the exception of  $\sigma^{70}P427$  N near  $O2$  of  $T_{+1}(t)$ ]; salt bridges to each of the phosphate oxygens at -2, -1, and +1 constrain the orientation of the bases with respect to the active site  $Mg^{2+}$ , leaving the bases free for pairing with substrate.

## **$\lambda P_R$ class I may be an unproductive RPo**

The relatively high resolution of the t-strand in the  $\lambda P_R$  class I complex (Fig. 1B, SI Appendix, Figs. S3 and S12A; Table S2) obtained in the absence of NTPs led us to examine whether the observed order corresponds to a site that is pre-organized to bind the first two initiating NTPs. To address this question, we took advantage of the high-resolution X-ray structure [2.9 Å, PDB ID 4Q4Z, (32)] of an initiation complex (RPinit) between *Thermus thermophilus* (*Tth*)  $E\sigma^A$  at a downstream fork promoter construct with the same t-strand sequence from -3 to +1 as  $\lambda P_R$ , and the first two initiating NTPs. In RPinit, the +1 and +2 t-strand bases pair with the initiating nucleotide ( $iNTP=ATP$ ) and a nonhydrolyzable nucleotide analog of the  $i+1NTP$  (CMPCPP), respectively. Despite differences between the two RNAPs, the DNA constructs used to form RPo and the methods used (X-ray, cryo-EM), the t-strand from -3 to +3 largely superposes (Fig. 5C), reflecting the evolutionary conservation of this site in multi-subunit RNAP (48, 49). While the path of the DNA backbone and the plane of the bases at all positions are strikingly

similar,  $G_{-1}(t)$  is a notable exception. In RP<sub>init</sub>,  $G_{-1}(t)$  forms an inter-strand stack with the  $i$ NTP (32). However, the base tilt of  $G_{-1}(t)$  in  $\lambda$ P<sub>R</sub> differs: modeling predicts that instead of forming a stabilizing interaction,  $G_{-1}(t)$  would clash with the incoming  $i$ NTP, suggesting that this form of RPo at  $\lambda$ P<sub>R</sub> may be recalcitrant to  $i$ NTP binding (see SI Appendix).

Given the persistent production of abortive products at  $\lambda$ P<sub>R</sub> in transcription assays that prevent RNAP rebinding (single round) (25, 51), we hypothesize that a relatively small conformational change in base tilt at -1 may convert class I  $\lambda$ P<sub>R</sub> to a conformer competent to bind the initiating NTP. Such a movement would occur on a time scale faster than the conversion to a productive complex. At  $\lambda$ P<sub>R</sub>, changing the  $i$ [NTP] from 5  $\mu$ M to 200  $\mu$ M increased abortive products by ~20-fold but had no effect on the amount of full-length RNA (51). One interpretation of these results is that unproductive complexes have a lower NTP binding affinity (51, 52). These data are consistent with the requirement of a conformational change to bind the  $i$ NTP (Fig. 5C).

## Discussion

Our findings illuminate the structural strategies that *Eco* RNAP uses to stabilize the transcription bubble and reveal how small differences in DNA sequence and/or discriminator length globally alter RPo structure. Because the next steps of NTP addition largely disrupt these contacts, their quality and extent impacts how quickly and efficiently RNAP breaks its promoter interactions as it converts to a processive elongation complex (53). Promoter escape is a complex function of DNA sequence and other extrinsic variables [cf. (53) and references therein; (13, 54-57)]. However, in general, initiation from promoters that form highly stable RPo is often rate-limited at escape whereas unstable RPo are limited at the steps of DNA binding and opening [cf. (13, 53)]. As a consequence, unstable RPo (e.g. *rmb* P1, T7A1) typically produce full length transcripts with few abortive products (58, 59) in single round assays. By contrast, highly stable RPo (e.g.  $\lambda$ P<sub>R</sub>) produce smaller amounts of full-length products while continuing to synthesize short RNAs (25, 51, 53, 58, 60).

At all promoters, initiation of RNA synthesis drives translocation of the nascent RNA/DNA hybrid and further unwinding of the downstream duplex DNA. Because  $E\sigma^{70}$  maintains its interactions with the upstream ds/ss fork junction, the transcription bubble progressively enlarges with each NTP addition, a process termed 'scrunching' (61-63). Given the volume constraints of the active site cleft, the cost of DNA unwinding and compaction during initiation was proposed to create "stressed" intermediates and that an accumulation of stress could drive promoter escape (25, 53, 57, 64).

Based on the structures described here, the structural barriers to initially "scrunching" the nt-strand seem low as few or no contacts are made from -3 to -1 in either  $\lambda$ P<sub>R</sub> or T7A1 (Fig. 3). At other promoters, the map density for DNA in this region is poor [ $\lambda$ P<sub>R-5C</sub> (Fig. 2); *rmb* P1 (36); *rspT* P2 (34)], indicating structural heterogeneity. We note that reducing interactions with the nt-strand discriminator region in RPo appears to favor a state where regions of *both* strands in the bubble are dynamic [cf. (18, 19)]. The T7A1 RPo structure suggests that the first translocation step on a six base discriminator



promoter extends the nt-strand downstream of -4 into the channel without steric opposition; subsequent steps are proposed to extrude the nt-strand out between the  $\beta$ protrusion and  $\beta$ lobe (65). While mid-region contacts are minimal on the t-strand compared to those upstream (Fig. 5A and 5B), the growing RNA-DNA hybrid pushes the t-strand further into the upstream end of the cleft (65). The t-strand “sandwich” at -10/-9 observed for  $\lambda P_R$  and T7A1 (Fig. 4) and *rrnB* P1 RPo (36) may provide an additional constraint that helps direct the bulge as scrunching in the t-strand is proposed to disrupt regions ( $\beta$ ’lid and the  $\sigma$ -finger) that block RNA entry into its exit channel (65).

## Proposed structural differences between productive and moribund RPo

Seminal work by Shimamoto, Hsu, Chamberlin and colleagues discovered that two forms of RPo can be populated at a given promoter: one that ultimately escapes to make full length RNA, and one stuck in iterative rounds of abortive cycling (51, 52, 60). The latter “moribund” complex often backtracks (3’-OH RNA no longer correctly positioned in the active site), forming a dead-end complex in the absence of Gre factors [cf. (53)]. The relative amounts of these functionally distinct complexes are promoter-sequence dependent (53, 58, 60). One of the striking implications of this work is that core promoter sequence fundamentally sets the degree to which the initiation branches *before* NTP binding [cf. (53) and references therein]. Functionally, branching allows additional regulation of transcriptional output independent of the rate-limiting step of initiation (51) and is modulated by the Gre factors *in vivo* [cf. (66)].

While the structural basis of moribund complex has been unknown, the unambiguous finding of two distinct cryo-EM classes at  $\lambda P_R$  but not at T7A1 or at  $\lambda P_{R-5C}$  may provide an answer. Based on studies of  $\lambda P_R$  by Shimamoto and colleagues (51, 66) and by the Record lab (21, 22), we suggest that class I and class II represent moribund and productive complexes, respectively (Eq. 1, RPo and I3, see SI Appendix). If so, our data indicate that favorable interactions with the discriminator region (e.g. G<sub>-5</sub>(nt)) drive formation of additional barriers to translocation (increased strand and downstream DNA order,  $\beta$ lobe/ $\sigma^{70}_{1.1}$ -linker/discriminator contacts, a more tightly closed clamp). We hypothesize that in this form of RPo, the next step of DNA unwinding is strongly disfavored, leaving the RNA/DNA hybrid largely pre-translocated (the newly added nucleotide remains in the active site). The longer in this state, the greater probability that the 3’ RNA hydroxyl disengages, creating a backtracked complex [cf. (67-69)] that either resets by releasing short RNA products or becomes dead-end complex.

## Implications of this study for the regulation of transcription initiation

Stable transcription bubble formation requires establishing RNAP/DNA interactions that disfavor reannealing, rewinding and DNA dissociation. Promoter sequence intrinsically dictates the cost of disrupting base stacking and base pairing and the degree to which unfavorable conformational changes are driven by forming favorable protein interactions. Perhaps not surprisingly, in the most stable RPo studied here ( $\lambda P_R$ ), bases on the separated strands are largely stacked (Fig. 2A) and DNA (both base and the

phosphate backbone) interactions with  $E\sigma^{70}$  appear to be maximized relative to the faster dissociating RPo ( $\lambda P_{R-5C}$ , T7A1). What is more remarkable is how the sequence and length of the nt-strand discriminator impact the overall RPo structure. In particular, the promoter-dependent degrees of disorder of the strands in the bubble described here presumably not only affect RPo lifetime but also start site selection and promoter escape.

The recent cryo-EM study of the steps of DNA opening at the *rpsT* P2 promoter revealed that bubble formation is not a simple progression in strand unwinding and base unstacking: DNA disruptions are dynamic and protein-DNA interactions form and unform (35). Chen et al. found that DNA entry into the active site channel is facilitated by unpairing and unstacking the DNA base pair at -12, which then repairs and restacks with -13 in the subsequent intermediate. In addition, T<sub>-9</sub>(t) binds in a pocket on the  $\beta$ protrusion at an intermediate step, but then leaves the pocket in subsequent intermediates and RPo (35).

Based on these observations and the comparison between  $\lambda P_R$  class I and II (Figs. 2A and 2B), we highlight what may have not been fully appreciated before this structural work: in a multistep mechanism (cf. Eq. 1), if base restacking occurs, it provides an enthalpic driving force for conformational rearrangements in that step, even if the overall net cost to RPo is zero. For example, if unwinding the upstream bubble partially (or fully) unstacks bases in the -10 hexamer [e.g. T<sub>-10</sub>(nt)/A<sub>-9</sub>(nt)/A<sub>-8</sub>(nt) and/or T<sub>-9</sub>(t)] as part of entry into the cleft, the cost is reflected in a slower forward rate or faster back rate (depending whether unstacking occurs before or after the transition state) relative to no disruption. Restacking in a later step will then have the opposite effect on the forward and back rates, respectively. If this scenario is applicable, we note that unstacking purines is more costly than unstacking pyrimidines, suggesting that the base sequence may play hidden roles in the individual steps of DNA opening and in the subsequent steps of promoter escape.

## Materials and Methods

Detailed descriptions of  $E\sigma^{70}$  purification, assembly of  $E\sigma^{70}$  – promoter DNA complexes, specimen preparation for cryo-EM, cryo-EM data acquisition and processing, model building and refinement, and abortive initiation transcription assays are provided in SI Appendix.

## Data Availability

The cryo-EM density maps have been deposited in the EMDataBank under accession codes EMD-23892 [*Eco*  $E\sigma^{70}$ - $\lambda P_R$  class I (RPo)], EMD-23893 [*Eco*  $E\sigma^{70}$ - $\lambda P_R$  class II (I3)], EMD-23895 (*Eco*  $E\sigma^{70}$ - $\lambda P_{R-5C}$  RPo), and EMD-23897 (*Eco*  $E\sigma^{70}$ -T7A1 RPo). The atomic coordinates have been deposited in the Protein Data Bank under accession codes 7MKD [*Eco*  $E\sigma^{70}$ - $\lambda P_R$  class I (RPo)], 7MKE [*Eco*  $E\sigma^{70}$ - $\lambda P_R$  class II (I3)], 7MKI (*Eco*  $E\sigma^{70}$ - $\lambda P_{R-5C}$  RPo), and 7MKJ (*Eco*  $E\sigma^{70}$ -T7A1 RPo).

## Acknowledgments

We thank E.A. Campbell, M. Lilic and members of the Darst-Campbell Laboratory for experimental advice and helpful discussions. R.M.S. thanks M.T. Record, Jr. and former members of the Record lab for many fruitful collaborations, and T.M. Lohman and D. Jensen for stimulating conversations and encouragement. We are grateful to R.L. Gourse and W. Ross for advice, insightful discussions and inspiring work. Some of the work was performed at the Simons Electron Microscopy Center and National Resource for Automated Molecular Microscopy, located at the New York Structural Biology Center, supported by grants from the Simons Foundation (SF349247), New York State Office of Science, Technology and Academic Research, and the NIH National Institute of General Medical Sciences (GM103310), with additional support from the Agouron Institute (F00316). This work was supported by NIH grant R35 GM118130 and Re-Entry Supplement R35 GM118130-04S1 to S.A.D.

## References

1. M. T. Record, Jr., W. S. Reznikoff, M. L. Craig, K. L. McQuade, P. J. Schlax, *Escherichia coli* RNA Polymerase ( $E\sigma^{70}$ ), *Promoters, and the Kinetics of the Steps of Transcription Initiation, Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (ed. 2nd, 1996), vol. 2.
2. W. R. McClure, Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* **54**, 171-204 (1985).
3. S. P. Haugen, W. Ross, R. L. Gourse, Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev. Microbiol.* **6**, 507-519 (2008).

4. E. A. Warman *et al.*, Widespread divergent transcription from bacterial and archaeal promoters is a consequence of DNA-sequence symmetry. *Nat. Micro.* 10.1038/s41564-021-00898-9 (2021).
5. X. Ju, D. Li, S. Liu, Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat. Micro.* **4**, 1907-1918 (2019).
6. T. T. Harden *et al.*, Alternative transcription cycle for bacterial RNA polymerase. *Nat. Comm.* **11**, 448 (2020).
7. A. Feklistov, B. D. Sharon, S. A. Darst, C. A. Gross, Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* **68**, 357-376 (2014).
8. T. M. Gruber, C. A. Gross, Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* **57**, 441-466 (2003).
9. M. Lonetto, M. Gribskov, C. A. Gross, The  $\sigma^{70}$  family: sequence conservation and evolutionary relationships. *J. Bacteriol.* **174**, 3843-3849 (1992).
10. E. A. Campbell *et al.*, Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell* **9**, 527-539 (2002).
11. R. M. Saecker, M. T. Record, P. L. Dehaseth, Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* **412**, 754-771 (2011).
12. E. A. Galburt, The calculation of transcript flux ratios reveals single regulatory mechanisms capable of activation and repression. *Proc. Natl. Acad. Sci. USA* **115**, E11604-E11613 (2018).
13. D. Jensen, E. A. Galburt, The context-dependent influence of promoter sequence motifs on transcription Initiation kinetics and regulation. *J. Bacteriol.* **203** (2021).
14. E. F. Ruff, M. T. Record, I. Artsimovitch, Initial events in bacterial transcription initiation. *Biomolecules* **5**, 1035-1062 (2015).
15. R. L. Gourse *et al.*, Transcriptional responses to ppGpp and DksA. *Annu. Rev. Microbiol.* **72**, 163-184 (2018).
16. A. A. Travers, Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *J. Bacteriol.* **141**, 973-976 (1980).
17. D. C. Straney, D. M. Crothers, Intermediates in transcription initiation from the *E. coli lac* UV5 promoter. *Cell* **43**, 449-459 (1985).
18. N. C. Robb *et al.*, The transcription bubble of the RNA polymerase-promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J. Mol. Biol.* **425**, 875-885 (2013).
19. E. Lerner, A. Ingargiola, S. Weiss, Characterizing highly dynamic conformational states: The transcription bubble in RNAP-promoter open complex as an example. *J. Chem. Phys.* **148**, 123315 (2018).
20. W. C. Suh, W. Ross, M. T. Record, Two open complexes and a requirement for  $Mg^{2+}$  to open the  $\lambda P_R$  transcription start site. *Science* **259**, 358-361 (1993).
21. W. S. Kontur, R. M. Saecker, M. W. Capp, M. T. Record, Late steps in the formation of *E. coli* RNA polymerase- $\lambda P_R$  promoter open complexes: characterization of conformational changes by rapid [perturbant] upshift experiments. *J. Mol. Biol.* **376**, 1034-1047 (2008).

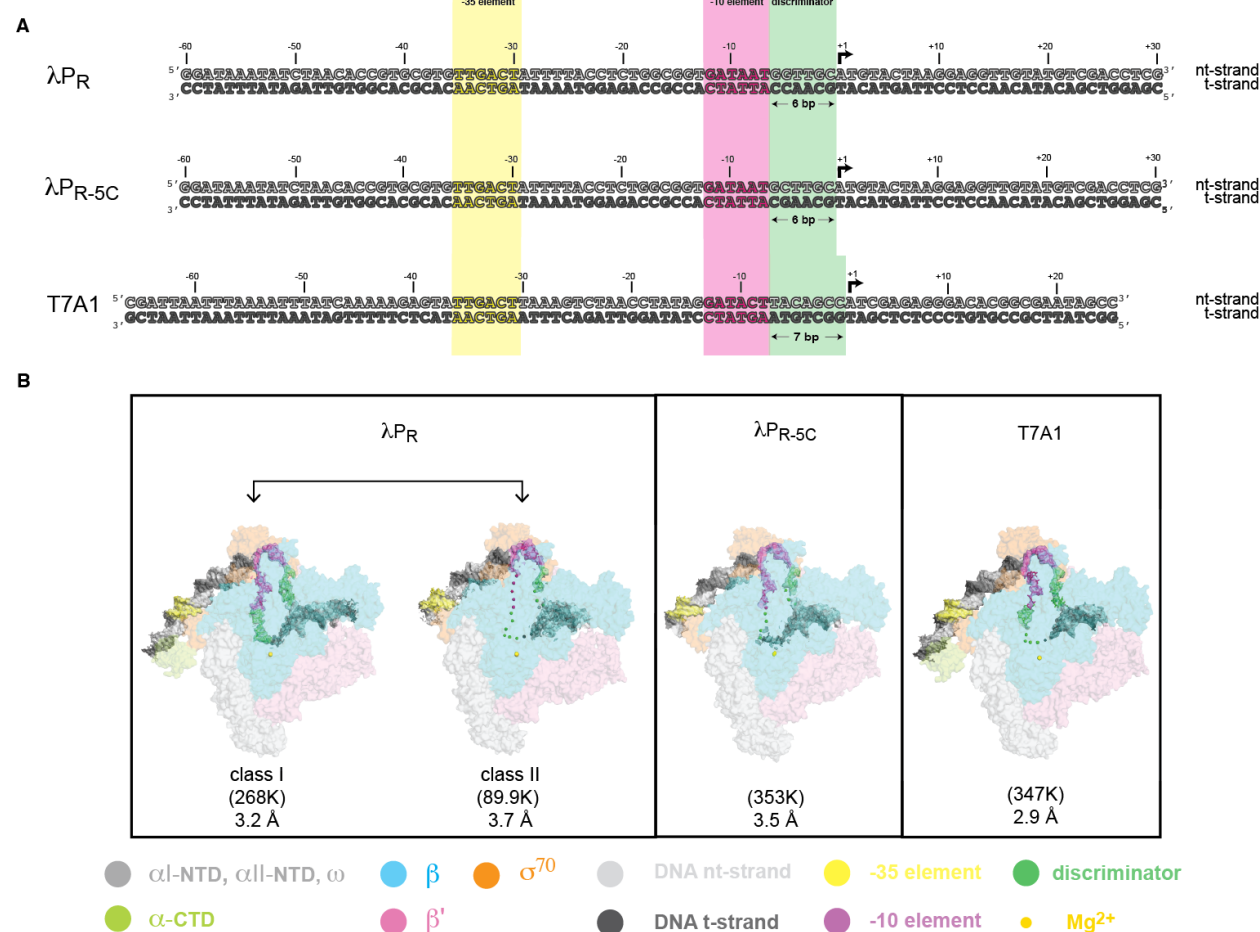
22. T. J. Gries, W. S. Kontur, M. W. Capp, R. M. Saecker, M. T. Record, One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proc. Natl. Acad. Sci. USA* **107**, 10418-10423 (2010).
23. S. P. Haugen *et al.*, rRNA promoter regulation by nonoptimal binding of  $\sigma$  region 1.2: an additional recognition element for RNA polymerase. *Cell* **125**, 1069-1082 (2006).
24. M. M. Barker, T. Gaal, C. A. Josaitis, R. L. Gourse, 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 (2001).
25. K. L. Henderson *et al.*, Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA* **114**, E3032-E3040 (2017).
26. S. P. Haugen, W. Ross, M. Manrique, R. L. Gourse, Fine structure of the promoter- $\sigma$  region 1.2 interaction. *Proc. Natl. Acad. Sci. USA* **105**, 3292-3297 (2008).
27. N. Zenkin *et al.*, Region 1.2 of the RNA polymerase  $\sigma$  subunit controls recognition of the -10 promoter element. *EMBO J.* **26**, 955-964 (2007).
28. M. NandyMazumdar *et al.*, RNA polymerase gate loop guides the nontemplate DNA strand in transcription complexes. *Proc. Natl. Acad. Sci. USA* **113**, 14994-14999 (2016).
29. B. Bae, A. Feklistov, A. Lass-Napiorkowska, R. Landick, S. A. Darst, Structure of a bacterial RNA polymerase holoenzyme open promoter complex. *eLife* **4** (2015).
30. Y. Zuo, T. A. Steitz, Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell* **58**, 534-540 (2015).
31. Y. Zhang *et al.*, Structural basis of transcription initiation. *Science* **338**, 1076-1080 (2012).
32. R. S. Basu *et al.*, Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J. Biol. Chem.* **289**, 24549-24559 (2014).
33. H. Boyaci, J. Chen, R. Jansen, S. A. Darst, E. A. Campbell, Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* **565**, 382-385 (2019).
34. J. Chen *et al.*, TraR allosterically regulates transcription initiation by altering RNA polymerase conformation. *eLife* **8** (2019).
35. J. Chen *et al.*, Stepwise Promoter Melting by Bacterial RNA Polymerase. *Mol. Cell* **78**, 275-288.e276 (2020).
36. Y. Shin *et al.*, Structural basis of ribosomal RNA transcription regulation. *Nat. Comm.* **12**, 528 (2021).
37. A. Rogozina, E. Zaychikov, M. Buckle, H. Heumann, B. Sclavi, DNA melting by RNA polymerase at the T7A1 promoter precedes the rate-limiting step at 37 degrees C and results in the accumulation of an off-pathway intermediate. *Nucleic Acids Res.* **37**, 5390-5404 (2009).
38. B. Sclavi *et al.*, Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. *Proc. Natl. Acad. Sci. USA* **102**, 4706-4711 (2005).
39. P. Schickor, W. Metzger, W. Werel, H. Lederer, H. Heumann, Topography of intermediates in transcription initiation of *E. coli*. *EMBO J.* **9**, 2215-2220 (1990).



40. J. Chen, A. J. Noble, J. Y. Kang, S. A. Darst, Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J. Struct. Biol. X* **1** (2019).
41. S. H. Scheres, RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519-530 (2012).
42. A. Feklistov, S. A. Darst, Structural basis for promoter -10 element recognition by the bacterial RNA polymerase  $\sigma$  subunit. *Cell* **147**, 1257-1269 (2011).
43. W. Ross, A. Ernst, R. L. Gourse, Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev.* **15**, 491-506 (2001).
44. W. Ross, D. A. Schneider, B. J. Paul, A. Mertens, R. L. Gourse, An intersubunit contact stimulating transcription initiation by *E. coli* RNA polymerase: interaction of the  $\alpha$  C-terminal domain and  $\sigma$  region 4. *Genes Dev.* **17**, 1293-1307 (2003).
45. N. Naryshkin, A. Revyakin, Y. Kim, V. Mekler, R. H. Ebright, Structural organization of the RNA polymerase-promoter open complex. *Cell* **101**, 601-611 (2000).
46. B. Bae *et al.*, Phage T7 Gp2 inhibition of *Escherichia coli* RNA polymerase involves misappropriation of  $\sigma^{70}$  domain 1.1. *Proc. Natl. Acad. Sci. USA* **110**, 19772-19777 (2013).
47. V. Mekler *et al.*, Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599-614 (2002).
48. W. J. Lane, S. A. Darst, Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J. Mol. Biol.* **395**, 671-685 (2010).
49. W. J. Lane, S. A. Darst, Molecular evolution of multisubunit RNA polymerases: structural analysis. *J. Mol. Biol.* **395**, 686-704 (2010).
50. E. Zaychikov, L. Denissova, T. Meier, M. Götte, H. Heumann, Influence of  $Mg^{2+}$  and temperature on formation of the transcription bubble. *J. Biol. Chem.* **272**, 2259-2267 (1997).
51. T. Kubori, N. Shimamoto, A branched pathway in the early stage of transcription by *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **256**, 449-457 (1996).
52. N. V. Vo, L. M. Hsu, C. M. Kane, M. J. Chamberlin, *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 2. Formation and characterization of two distinct classes of initial transcribing complexes. *Biochemistry* **42**, 3787-3797 (2003).
53. L. M. Hsu, Promoter Escape by *Escherichia coli* RNA Polymerase. *EcoSal Plus* **3** (2008).
54. J. Skancke, N. Bar, M. Kuiper, L. M. Hsu, Sequence-dependent promoter escape efficiency is strongly influenced by bias for the pretranslocated state during initial transcription. *Biochemistry* **54**, 4267-4275 (2015).
55. E. Heyduk, T. Heyduk, DNA template sequence control of bacterial RNA polymerase escape from the promoter. *Nucleic Acids Res.* **46**, 4469-4486 (2018).
56. J. Ko, T. Heyduk, Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. *Biochem. J.* **463**, 135-144 (2014).
57. A. Mazumder, A. N. Kapanidis, Recent Advances in Understanding  $\sigma^{70}$ -Dependent Transcription Initiation Mechanisms. *J. Mol. Biol.* **431**, 3947-3959 (2019).

58. R. Sen, H. Nagai, V. J. Hernandez, N. Shimamoto, Reduction in abortive transcription from the  $\lambda P_R$  promoter by mutations in region 3 of the  $\sigma^{70}$  subunit of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **273**, 9872-9877 (1998).
59. R. L. Gourse, Visualization and quantitative analysis of complex formation between *E. coli* RNA polymerase and an rRNA promoter *in vitro*. *Nucleic Acids Res.* **16**, 9789-9809 (1988).
60. M. Susa, R. Sen, N. Shimamoto, Generality of the branched pathway in transcription initiation by *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **277**, 15407-15412 (2002).
61. A. Revyakin, C. Liu, R. H. Ebright, T. R. Strick, Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139-1143 (2006).
62. A. N. Kapanidis *et al.*, Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144-1147 (2006).
63. G. M. Cheetham, T. A. Steitz, Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305-2309 (1999).
64. D. C. Straney, D. M. Crothers, A stressed intermediate in the formation of stably initiated RNA chains at the *Escherichia coli lac UV5* promoter. *J. Mol. Biol.* **193**, 267-278 (1987).
65. J. T. Winkelman *et al.*, Crosslink Mapping at Amino Acid-Base Resolution Reveals the Path of Scrunched DNA in Initial Transcribing Complexes. *Mol. Cell* **59**, 768-780 (2015).
66. M. Susa, T. Kubori, N. Shimamoto, A pathway branching in transcription initiation in *Escherichia coli*. *Mol. Microbiol.* **59**, 1807-1817 (2006).
67. E. Lerner *et al.*, Different types of pausing modes during transcription initiation. *Transcription* **8**, 242-253 (2017).
68. M. Turtola, J. J. Mäkinen, G. A. Belogurov, Active site closure stabilizes the backtracked state of RNA polymerase. *Nucleic Acids Res.* **46**, 10870-10887 (2018).
69. J. Saba *et al.*, The elemental mechanism of transcriptional pausing. *eLife* **8** (2019).

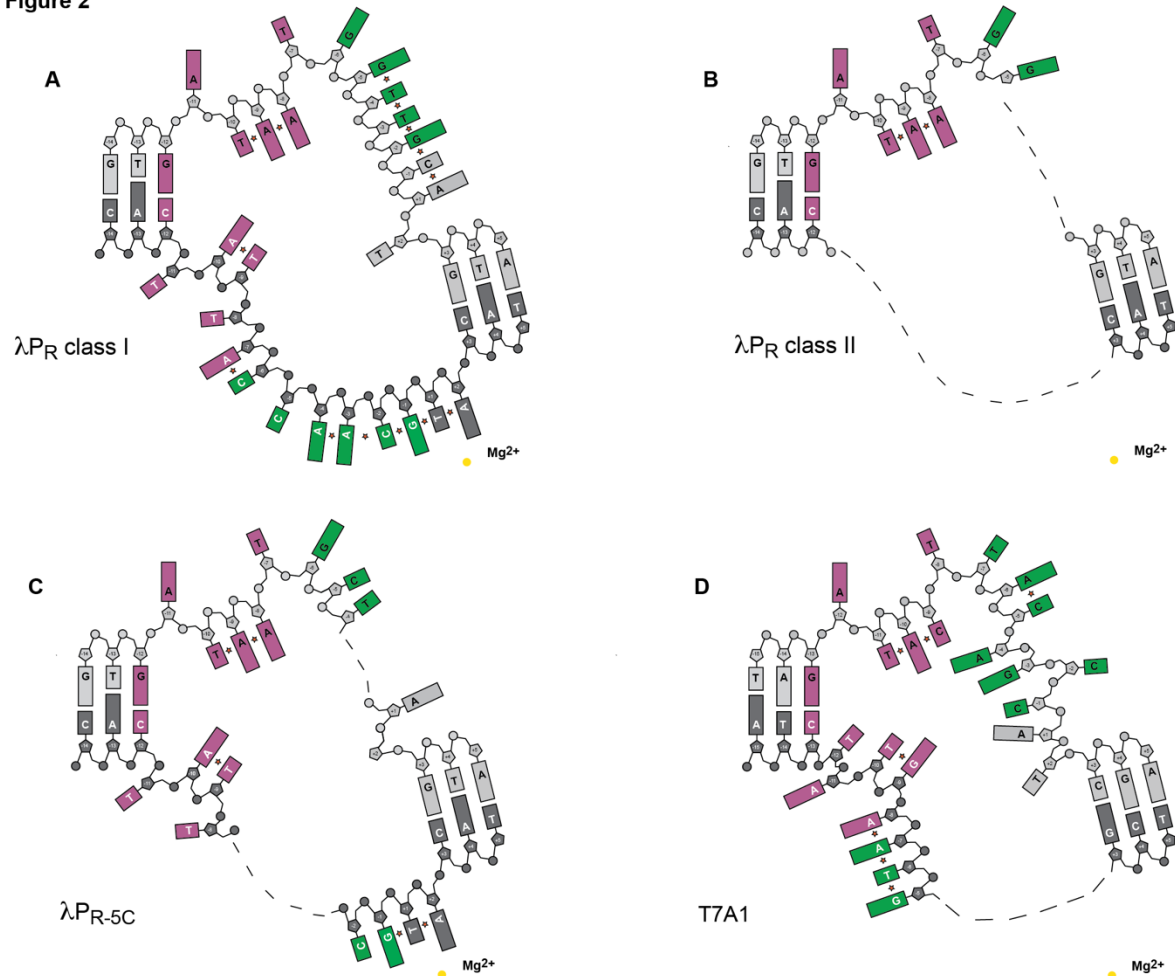
**Figure 1**



**Fig. 1.** Promoter DNA constructs used for cryo-EM studies and overall cryo-EM structures of RPos. (A) Promoter sequences studied by cryo-EM [nt-strand DNA (top strand), light grey; t-strand (bottom strand), dark grey]. Numbers above the DNA sequences denote positions with respect to the transcription start site (+1, denoted by the black arrow). Shaded colors highlight key promoter regions: -35 element (yellow), -10 element (magenta), and the discriminator (pale green). (top)  $\lambda P_R$  (-60 to +30). -60 to +20 are native  $\lambda P_R$  sequences. Sequences downstream of +20 originate from the plasmid construct used in extensive kinetic and DNA footprinting studies of  $\lambda P_R$  (11); (middle)  $\lambda P_{R-5C}$  (-60 to +30). Single base pair inversion of  $\lambda P_R$  G-5 to C. (bottom) T7A1 (-66 to +20). While T7A1 and  $\lambda P_R$  share the same -35 element sequence, key differences exist in both the -10 element and the sequence and length (7 vs 6 nucleotides, respectively) of the discriminator. At the upstream end of the constructs used here, the T7A1 promoter has both proximal and distal UP elements (tight-binding  $\alpha$ CTD binding sites), and  $\lambda P_R$  (and  $\lambda P_{R-5C}$ ) has a distal UP (3). (B) *Eco* RNAP subunits are shown as transparent surfaces ( $\alpha$ I,  $\alpha$ II,  $\omega$ , light gray;  $\alpha$ CTD, pale green;  $\beta$ , pale cyan;  $\beta'$ , light pink;  $\sigma^{70}$ , light orange) with the active site  $Mg^{2+}$  shown as a sphere (pale yellow). Promoter DNA is shown as cryo-EM difference density (-35 element, -10 element and discriminator colored as in A). Particle classification revealed two distinct

RNAP-DNA complexes populated at  $\lambda P_R$ , and a single class at T7A1 and at  $\lambda P_{R-5C}$ . The number of particles in each class and nominal resolution are shown. In  $\lambda P_R$  class I, good map density allowed all bases in the nt- and t-strands to be modeled. For comparison, disordered bases in the other RPo are shown as spheres positioned approximately at the corresponding position of the phosphate backbone in  $\lambda P_R$  class I. DNA in RPo modeled from -45 to +23 in  $\lambda P_R$  class I, -37 to +13 in  $\lambda P_R$  class II, -37 to +15 in  $\lambda P_{R-5C}$ , and -47 to +15 in T7A1.

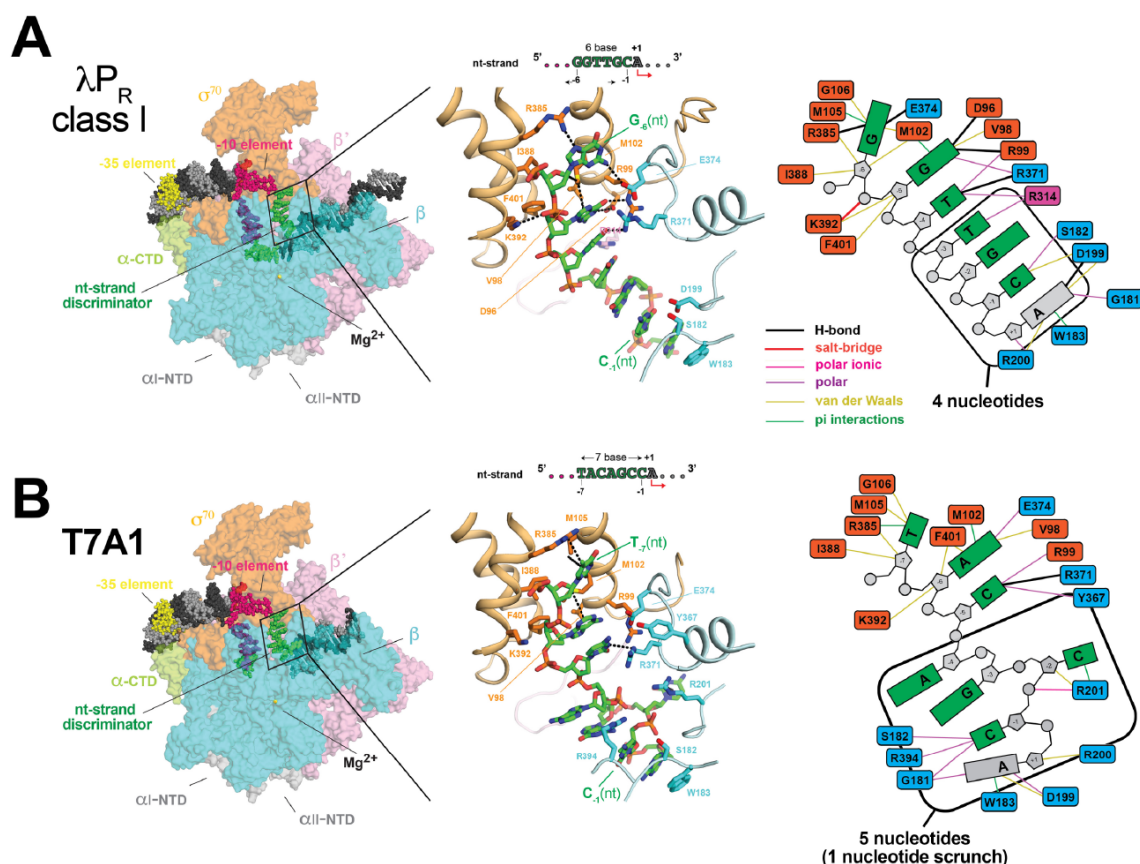
**Figure 2**



**Fig. 2.** Comparison of differences in base stacking in the transcription bubble. The schematic illustrates: i. position of bases in each open complex, missing bases shown as dashes; and ii. base stacking pairs indicated by a star symbol. DNA backbone and bases colored as in Fig. 1. (A)  $\lambda P_R$  class I. (B)  $\lambda P_R$  class II. (C)  $\lambda P_R$ -5C. (D) T7A1.

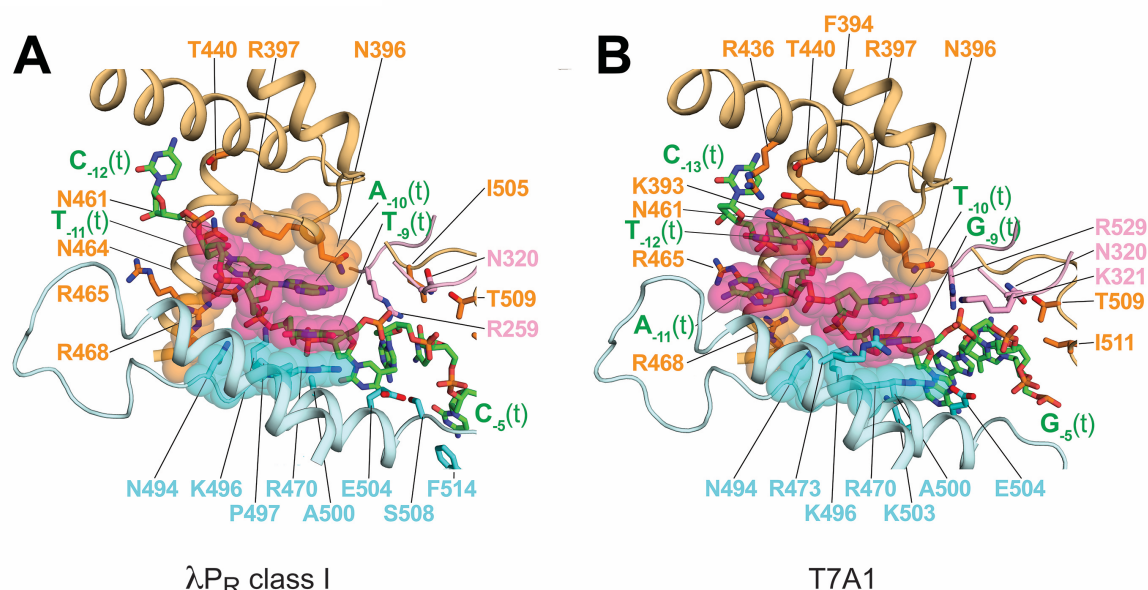


Figure 3



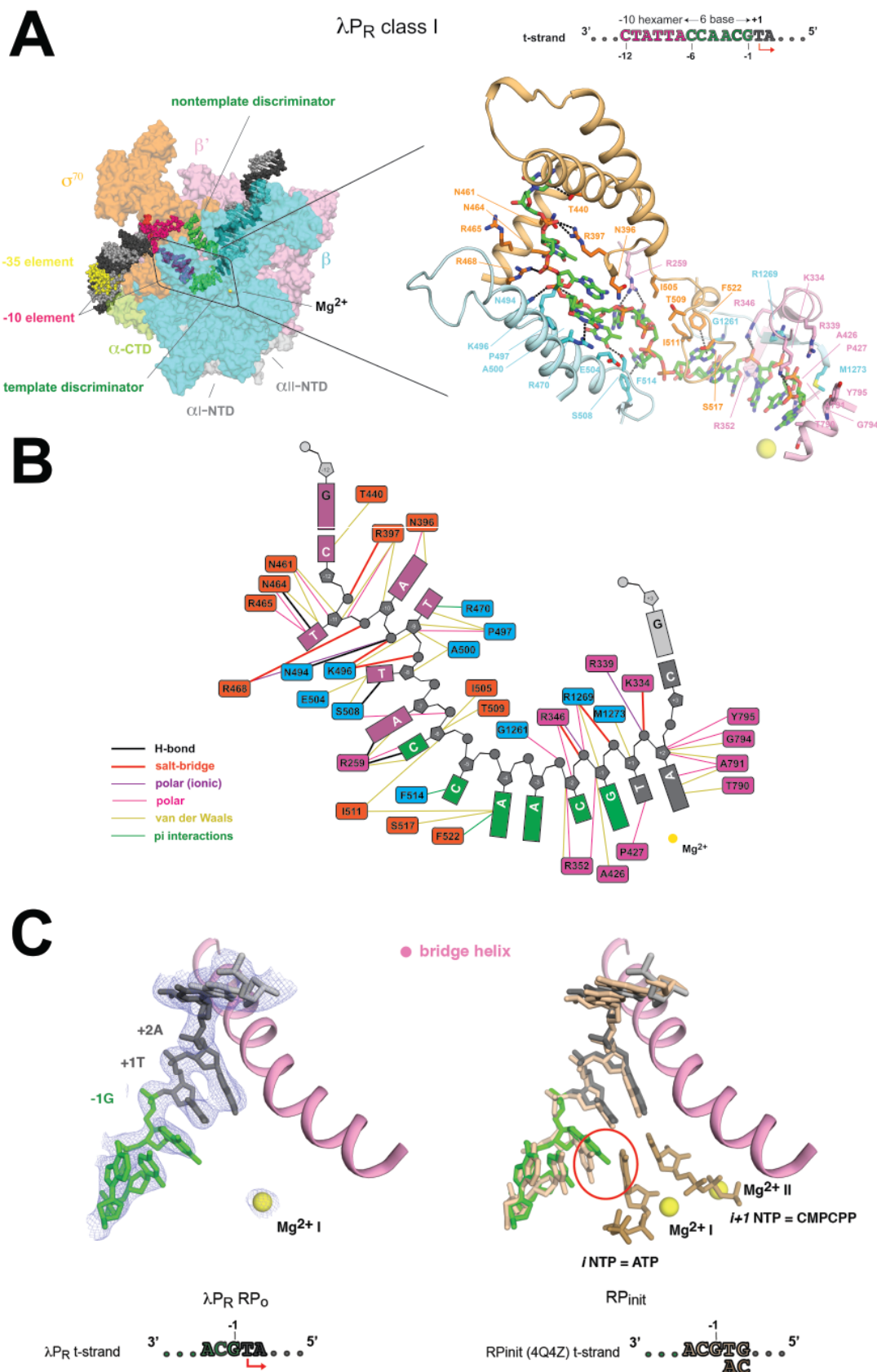
**Fig. 3.** Differences in  $E\sigma^{70}$  interactions with the nt-strand discriminator region between (A)  $\lambda P_R$  class I (6 base discriminator) and (B) T7A1 (7 base discriminator). (Left) Overall view of RPO (similar to Fig. 1B with same color scheme) with RNAP subunits shown in surface representation and DNA as atomic spheres. The boxed area is magnified in the middle. (Middle) Magnified view showing interactions with the nt-strand from the discriminator to +1. RNAP subunits shown in backbone worm ( $\beta$ , pale cyan;  $\beta'$ , light pink;  $\sigma^{70}$ , light orange); sidechains of atoms within 4.5 Å of nucleic acid atoms [ $\lambda P_R$ (-6 to +1) or T7A1(-7 to +1)] are shown as sticks. Atomic distances within 3.5 Å and interactions with the  $\pi$  electrons of the DNA bases (sulfur- $\pi$ , and cation- $\pi$ ) are shown by black dashed lines. The DNA carbon atoms are colored green. (Right) Schematic comparing the network of interactions between the nt-discriminator (and +1) and  $E\sigma^{70}$ . Favorable interactions within 3.5 Å (hydrogen bonds, salt bridges) are shown by heavier lines (see color key) than those within 4.5 Å [polar, ionic, van der Waals,  $\pi$  (sulfur, cation,  $\pi$ )]. Corresponding cryo-EM density is shown in SI Appendix, Fig. S10. Although the nature and extent of interactions with the nt-strand discriminator differ between  $\lambda P_R$  and T7A1, the DNA backbone is largely solvent-exposed. Few favorable interactions exist with the sugar atoms (at upstream and downstream ends of the discriminator region) and only one positively charged amino acid falls within 4.5 Å of the DNA phosphate backbone of either promoter.

Figure 4



**Fig. 4.** Scrunching in the t-strand. (A)  $\lambda P_R$  class I RPo flips out T<sub>-11</sub>(t) at the ss/ds junction. (B) T7A1 flips out two bases, T<sub>-12</sub>(t) and A<sub>-11</sub>(t), scrunching the t-strand at the upstream entrance to the active site channel. Residues within 4.5 Å of nucleic acid atoms of the t-strand are shown as sticks. *Eco* E $\sigma^{70}$  subunits shown in backbone worm ( $\beta$ , pale cyan;  $\beta'$ , light pink;  $\sigma^{70}$ , light orange). Single-stranded bases from -11 to -9 (A;  $\lambda P_R$ ) or -12 to -9 (B; T7A1) are shown as sticks (same color coding as Fig. 3) and also transparent atomic spheres (hot pink). The same  $\sigma$  and  $\beta$  residues stabilize the A<sub>-10</sub>(t)/T<sub>-9</sub>(t) or T<sub>-10</sub>(t)/G<sub>-9</sub>(t) stacking pairs in  $\lambda P_R$  and T7A1, respectively, noted and shown as transparent CPK atoms ( $\sigma$ , orange;  $\beta$ , cyan). Corresponding cryo-EM density is shown in SI Appendix, Fig. S12.

Figure 5



**Fig. 5.**  $\lambda P_R$  class I t-strand interactions and positioning in the active site. (A) (Left) Overall representation of RPo as in Fig. 3. The boxed area is magnified to the right. (B) Schematic detailing the interactions between the t-strand and residues within 4.5 Å. (C) Comparison of the t-strand in  $\lambda P_R$  class I RPo and in RPinit. (Left) Cryo-EM density (blue mesh) defining the modeled position of single-strand bases on the t-strand (-3 to +2), the downstream ss/ds junction at +3 (shown as sticks) and the  $Mg^{2+}$  bound in the RPo active site (yellow sphere). The conserved RNAP bridge helix ( $\beta'$ ; pink) is shown as a point of reference. Colors as in Fig. 1. (Right) Result of aligning  $\lambda P_R$  class I RPo with a high-resolution X-ray structure of RPinit [2.9 Å resolution, PDB 4Q4Z; (32)]. Initiating triphosphate ribonucleotides (dark beige) bound at +1 ( $i$  NTP = ATP) and +2 ( $i+1$  NTP = CMPCPP) pair with the corresponding t-strand bases. The backbone and bases from -3 to +3 in RPinit (beige) largely superpose with the equivalent positions in  $\lambda P_R$  with the exception of G<sub>-1</sub>(t) where a change in base tilt leads to a clash (red circle) with the iNTP (ATP).