

Mechanism for *De Novo* RNA Synthesis and Initiating Nucleotide Specificity by T7 RNA Polymerase

William P. Kennedy, Jamila R. Momand and Y. Whitney Yin*

Department of Chemistry
and Biochemistry, Institute of
Cellular and Molecular Biology
University of Texas at Austin
Austin, TX 78712, USA

DNA-directed RNA polymerases are capable of initiating synthesis of RNA without primers, the first catalytic stage of initiation is referred to as *de novo* RNA synthesis. *De novo* synthesis is a unique phase in the transcription cycle where the RNA polymerase binds two nucleotides rather than a nascent RNA polymer and a single nucleotide. For bacteriophage T7 RNA polymerase, transcription begins with a marked preference for GTP at the +1 and +2 positions. We determined the crystal structures of T7 RNA polymerase complexes captured during the *de novo* RNA synthesis. The DNA substrates in the structures in the complexes contain a common $\Phi 10$ duplex promoter followed by a unique five base single-stranded extension of template DNA whose sequences varied at positions +1 and +2, thereby allowing for different pairs of initiating nucleotides GTP, ATP, CTP or UTP to bind. The structures show that the initiating nucleotides bind RNA polymerase in locations distinct from those described previously for elongation complexes. Selection bias in favor of GTP as an initiating nucleotide is accomplished by shape complementarity, extensive protein side-chain and strong base-stacking interactions for the guanine moiety in the enzyme active site. Consequently, an initiating GTP provides the largest stabilization force for the open promoter conformation.

Published by Elsevier Ltd.

*Corresponding author

Keywords: T7 RNA polymerase; *de novo* RNA synthesis; GTP specificity

Introduction

DNA-directed RNA polymerases (RNAP) are essential enzymes in transcribing genetic information from DNA into RNA. Unlike DNA polymerases, RNA polymerases initiate RNA synthesis in the absence of a primer. The first step in initiation is called *de novo* RNA synthesis, in which RNAP recognizes a specific sequence on the DNA template, selects the first pair of nucleotide triphosphates complementary to template residues at positions +1 and +2, and catalyzes the formation of a phosphodiester bond to form a dinucleotide.

Several aspects of *de novo* RNA synthesis by bacteriophage T7 RNAP differ from subsequent steps of RNA synthesis. First, the initiating nucleotides have lower affinities for the polymerase than those used during elongation. The K_d value is 2 mM for the first initiating NTP and

80 μM for the second,¹ whereas the K_d is $\sim 5 \mu\text{M}$ for NTPs during elongation.² Second, RNAP exhibits relatively slow chemistry during *de novo* synthesis. The rate of formation of the first phosphodiester bond by T7 RNAP is $\sim 7.8 \text{ s}^{-1}$, in contrast to 220 s^{-1} for formation of the same bond between an NTP and the 3' end of the nascent RNA during elongation.³ In fact, *de novo* synthesis is the rate-limiting step during transcription. The difference in rates of bond formation implies a difference in the mechanistic details of catalysis during the two phases of transcription. Third, T7 RNAP exhibits a strong bias for GTP as the initiating nucleotide.⁴ Among the 17 T7 promoters in the genome, 15 initiate with GTP (and 13 with pppGpG), whereas there is no obvious NTP preference during transcription elongation.

The mechanism for GTP selection as the initiating nucleotide is not known. However, such selection seems broadly used: DNA-directed-RNA polymerases from other members of the T7 supergroup of bacteriophages (e.g. T3, SP6, K1-5, K1E, K1F, and K11), and the RNA-directed-RNA polymerases from many pathogenic *Flaviviridae* family viruses (e.g. Dengue, West Nile, hepatitis C, and bovine

Abbreviations used: RNAP, RNA polymerase; HCV, hepatitis C virus; BVDV, bovine viral diarrhea virus.

E-mail address of the corresponding author:
whitney.yin@mail.utexas.edu

viral diarrhoea viruses) all initiate RNA synthesis with GTP.⁵

In order to understand the structural basis for nucleotide selection by T7 RNAP, we determined the crystal structures of T7 RNAP binary complexes captured during *de novo* synthesis with four promoter variants, together with their ternary complexes containing the corresponding incoming nucleotides; and we measured the stabilization of open promoter complexes by different initiating nucleotides. The structures of the ternary complexes provide a rational basis for initiating nucleotide selection and add to our understanding of the rate-limiting step of transcription.

Results

Transcription initiation by T7 RNAP can be divided into at least four steps: (1) promoter DNA binding, (2) promoter unwinding to form an initial transcription bubble, (3) initiating nucleotide binding and (4) the first phosphodiester bond formation. See equation (1):



where DNA_c and DNA_o are closed and open promoter DNA, respectively.

The structures of a T7 RNAP promoter complex corresponding to the product of step 2, and an initiation complex with a 3-mer RNA corresponding to the product of step 5 have been reported.^{6,7} The 3-mer RNA initiation complex represents the next completed reaction after *de novo* synthesis. The structures described here correspond to the products of steps 2 and 3 of the transcription reaction. The difference between the step 2 structures reported previously and those described here is that the present promoters, like those employed to elucidate the structure corresponding to step 5,⁷ contain a single-stranded template extension. In order to capture *de novo* complexes with various initiating nucleotides, DNA substrates were constructed to contain a common duplex $\Phi 10$ promoter followed by a 5 nt single-stranded template whose sequences are unique at positions +1 and +2: CC, TT, GG, or AA. They are therefore complementary to the initiating nucleotides GTP, ATP, CTP or UTP, respectively. The DNA constructs are named accordingly: T-CC, T-TT, T-GG or T-AA (Table 1).

Four binary complexes of T7 RNAP were formed with the respective DNAs. In order to capture T7 RNAP at the pre-chemistry stage, ternary complexes were made by mixing the binary complexes with appropriate 3'-deoxynucleotide triphosphates (3'-dNTP) as non-reactive analogues of the normal initiating rNTPs. As 3'-dNTPs are unable to complete phosphodiester bond formation, T7 RNAP is stalled at the pre-chemistry stage.

Table 1. Sequences of DNA substrates for crystallographic studies

Substrates	Sequence	Initiating nucleotide
T-CC	TAATACGACTCACTATA ATTATGCTGAGTGATATCCTTC	3'-dGTP
T-AA	TAATCGACTCACTATA ATTATGCTGAGTGATATAATTC	3'-dUTP
T-GG	TAATCGACTCACTATA ATTATGCTGAGTGATATGGTTC	3'-dCTP
T-TT	TAATCGACTCACTATA ATTATGCTGAGTGATATTTAAC	3'-dATP

Although all constructs could be crystallized, variation in crystal quality was observed. The crystals diffracted from 2.2–3.2 Å resolution, the best-diffracting crystals were obtained from the ternary complex T-CC with 3'-dGTP. All complex crystals belong to the primitive orthorhombic space group $P2_12_12$ with unit cell dimensions of approximately $a=220$ Å, $b=73$ Å, $c=81$ Å, $\alpha=\beta=\gamma=90^\circ$, except for the T-GG binary complex, which crystallized in the space group $P222_1$ with unit cell dimensions $a=73.25$ Å, $b=160.58$ Å, $c=225.24$ Å, $\alpha=\beta=\gamma=90^\circ$.

The smaller unit cell contains one complex per asymmetric unit, whereas the larger unit cell contains two complexes per asymmetric unit. The structures were solved by molecular replacement using the T7 RNAP 3-mer initiation complex as the starting model (PDB accession code 1QLN) and further refined to the working R -factor $\sim 27\%$ ($R_{\text{free}} \sim 30\%$). The diffraction data and refinement statistics are presented in Table 2.

The T7RNAP/T-CC/3'-dGTP ternary complex structure will be discussed in detail because it gave the highest resolution, is the physiologically most relevant structure, and reveals numerous features of *de novo* RNA synthesis and of GTP selection. The other complexes of T7 RNAP provide insights into the structural basis for the discrimination against initiating nucleotides other than GTP. T7RNAP can be described by the canonical "right-hand" configuration, with thumb, palm and fingers domains. The palm domain contains the active site that harbors the catalytic residues (D812 and D537) in an accessible, open conformation.

No induced-fit conformational change by the initiating nucleotide

The polymerase remains in a nearly unchanged conformation in all binary and ternary *de novo* complexes. The rmsd for the backbone of the polymerase in all structures is less than 1.2 Å, suggesting that binding of initiating nucleotides does not induce conformational changes in RNAP. This is different from elongation, where binding of

Table 2. Summary of crystallographic analysis

	T-CC ternary complex	T-GG ternary complex	T-AA ternary complex	T-TT ternary complex
A. Data collection				
Incoming nucleotide	3'dGTP	3'dCTP	3'dUTP	3'dATP
Resolution (Å)	2.4	3.0	3.2	2.5
Space group	<i>P</i> ₂ ₁ ₂ ₁ ²			
Cell dimensions				
<i>a</i> (Å)	224.65	219.06	221.32	219.76
<i>b</i> (Å)	73.59	75.47	75.82	79.74
<i>c</i> (Å)	79.51	80.63	81.46	73.30
α (deg.)	90	90	90	90
β (deg.)	90	90	90	90
γ (deg.)	90	90	90	90
Number of reflections	236,540	82,121	149,615	26,968
<i>R</i> _{sym} ^a (%)	10.8	11.2	7.0	10.5
Completeness ^b	98 (77)	77 (45)	99.3 (99.9)	72 (44)
B. Refinement				
<i>R</i> _{work} ^c (%)	27.9	26.5	28.2	28.8
<i>R</i> _{free} ^d (%)	29.1	28.7	30.1	31.4
RMS deviations from ideal values				
Bond (Å)	0.0078	0.0081	0.0097	0.0087
Angle (°)	1.54	1.67	1.34	1.67
C. Data collection				
Incoming nucleotide	T-CC binary complex	T-GG binary complex	T-AA binary complex	T-TT binary complex
	None	None	None	None
Resolution (Å)	3.2	2.2	2.6	2.6
Space group	<i>P</i> ₂ ₁ ₂ ₁ ²	<i>P</i> ₂ ₂ ₂ ₁	<i>P</i> ₂ ₁ ₂ ₁ ²	<i>P</i> ₂ ₁ ₂ ₁ ²
Cell dimensions				
<i>a</i> (Å)	222.67	73.25	221.67	220.31
<i>b</i> (Å)	82.71	160.58	73.38	73.36
<i>c</i> (Å)	77.67	225.24	79.58	78.26
α (deg.)	90	90	90	90
β (deg.)	90	90	90	90
γ (deg.)	90	90	90	90
Reflections	86,540	538,541	214,012	218,570
<i>R</i> _{sym} (%)	10.8	11.4	10.6	10.6
Completeness	98 (77)	94.8 (89.4)	81(43)	95.2 (88.5)
D. Refinement				
<i>R</i> _{work} (%)	27.4	25.2	26.6	27.7
<i>R</i> _{free} (%)	29.3	28.3	29.5	30.4
RMS deviations from ideal				
Bond lengths (Å)	0.0108	0.0088	0.0115	0.0090
Bond angles (deg.)	1.57	1.54	1.61	1.71

^a $R_{\text{sym}} = \sum |I_i - \langle I \rangle| / \sum I_i$ where I_i is the i th measurement and $\langle I \rangle$ is the weighted mean of all measurements of I .

^b Values in parentheses are for the highest resolution shell.

^c $R_{\text{work}} = \sum_{hkl} |F_{\text{obs}}(hkl) - F_{\text{calc}}(hkl)| / \sum_{hkl} |F_{\text{obs}}(hkl)|$ for reflections in the working data set.

^d R_{free} is the same as R_{work} for 5% of the data randomly omitted from refinement.

the incoming correct nucleotide causes the polymerase to undergo conformational changes and adopt a closed structure. The open conformation of the RNAP before NTP binding is termed the pre-insertion state, while the NTP-bound closed conformation is termed post-insertion. The two states are so named because both structures are different from the true open conformation of T7 RNAP in the absence of DNA. RNAP in the pre-insertion conformation is unable to bind to NTP because its binding site (N-site) is partially precluded by the active site residue Y639.¹⁰ In addition, the templating residue $n+1$ is in a flipped-out position and is thus unable to form a base-pair with the incoming nucleotide. In the post-insertion conformation, RNAP is competent for NTP binding because the conformational change in the DNA-binding fingers domain causes rotation of the O helix (residues

627–638), the binding site for the triphosphate of incoming nucleotides, towards the palm active site domain. In addition, residue Y639 rotates, allowing the incoming nucleotide to access the N-site, and the flipped-out template $n+1$ nucleotide to reposition (Figure 1). The overall result of the concerted conformational changes is that an incoming NTP is able to bind in the N-site, and the $n+1$ nucleotide can base-pair with the NTP.¹⁰ Because the set of conformational changes occurs only when RNAP binds to the correct nucleotide, it constitutes a mechanism for nucleotide discrimination. During *de novo* synthesis, however, RNAP shows no initiating nucleotide-induced conformational change, suggesting a different mechanism for NTP selection. The template sequence is likely to play only a relatively minor role in nucleotide selection because the rigid active

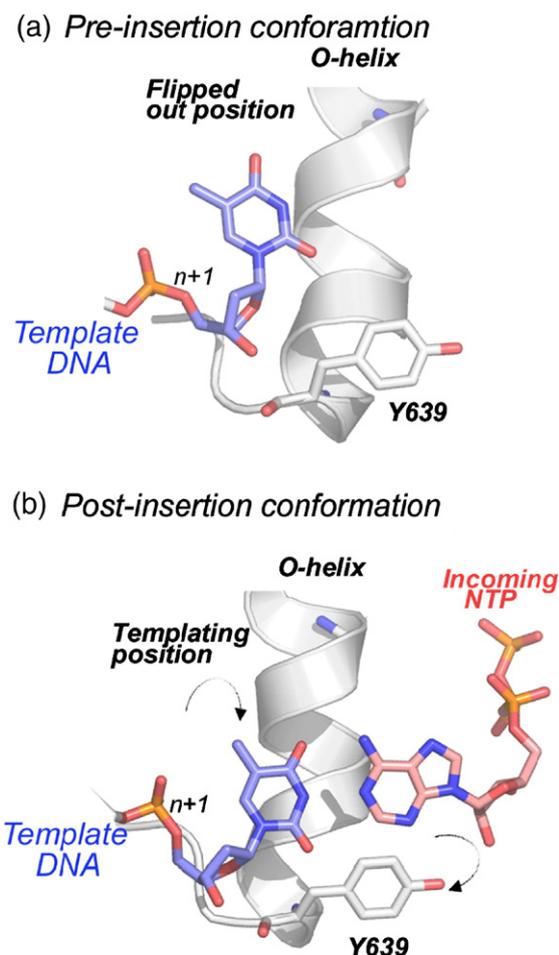


Figure 1. Configuration of the active site of T7 RNAP before and after NTP binding. (a) The pre-insertion conformation that is incompetent for NTP binding. The templating residue n is in a flipped-out position; the NTP-binding site (N-site) is occluded by Y639 and the O-helix is inward, away from the NTP site. (b) The post insertion conformation induced by NTP binding. The conformational changes move Y639 away from the N-site, repositions the templating residue n , and rotates the O-helix towards the active site.

site cannot accommodate the various base-pairs equally well.

Unique template DNA conformation

The DNA from positions -17 to -1 in the *de novo* complex adopts a conformation identical with that in the binary 17 bp open promoter complex (PDB code 1cez). The locations of the $+1$ and $+2$ template residues are restricted by the promoter and, relative to the -1 residue, are fully extended towards their 5' end. The $+1$ and $+2$ template nucleotides are flanked by Y639 of the polymerase and the -1 template base, and are poised to form hydrogen bonds with two incoming nucleotides simultaneously (Figure 2(a)). Superposition of the *de novo* and 3-mer initiation complexes suggests that the templating residue

during elongation is equivalent to the $+3$ position in the *de novo* complex. In *de novo* complexes the $+3$ template residue is in an inactive, flipped-out position, as in a typical pre-insertion complex. However, the conformation of this residue has no effect on binding of the two initiating nucleotides. The template DNA forms an arched structure from positions -1 to $+3$, which may stabilize the single-stranded DNA region in the open promoter for initiating nucleotide binding. The template conformation in subsequent steps, as exemplified in the 3-mer initiation complex, adopts the *A*-form in the heteroduplex region.

Novel binding sites for initiating nucleotides

Template positioning within RNAP determines the unique binding sites of the initiating nucleotides. Since the templating residue during elongation is located downstream of the two initiating nucleotides during *de novo* synthesis, the latter necessarily bind at locations different from that of the incoming NTP during elongation. Previous studies on elongating T7 RNAP defined the incoming NTP binding site as the N-site and the 3' end of the RNA product site as the P-site (Figure 3(b)).^{10,11} In *de novo* complexes, neither of the initiating GTP nucleotides binds in the N-site; rather, one binds to the P-site and the second is found in a novel site located upstream of the P-site. This new substrate-binding site is named the D-site, for *de novo* synthesis (Figure 3(a)). The lack of flexibility in the D and P-sites suggests that a nucleotide that can form the most interactions with the residues lining those sites, and thus has the highest affinity, will be preferred for *de novo* synthesis.

In addition to Watson–Crick base-pairing interactions with the template, the two initiating GTP nucleotides form guanine-specific interactions with the polymerase. The first nucleotide, 3'-dGTP(1), interacts with the polymerase primarily through the guanine moiety, whereas the second nucleotide, 3'-dGTP(2), interacts with the polymerase through both base and triphosphate moieties. Specifically, interactions between guanine of 3'-dGTP(1) and the polymerase include: the 2-NH₂ forms bipartite H-bonds with 2NH+ of H811 and the guanidino group of R425; O⁶ and N7 make water-mediated H-bonds with the guanidino group of R632; and van der Waals interactions are made with H811 (Figure 4(a)). These data provide an atomic explanation for the preference of T7 RNAP in initiating synthesis with a purine, specifically a guanine, nucleotide and the discrimination against 7-deazaGTP.¹² No direct contact between the triphosphate of 3'-dGTP(1) is observed, despite the presence of the positively charged residue R394 in the vicinity (6.0 Å) of the γ -phosphate. Any electrostatic interaction may be neutralized by the presence of the negatively charged residue D351 located at an equal distance (6.2 Å) (not shown in Figure 4(a)). Selection for

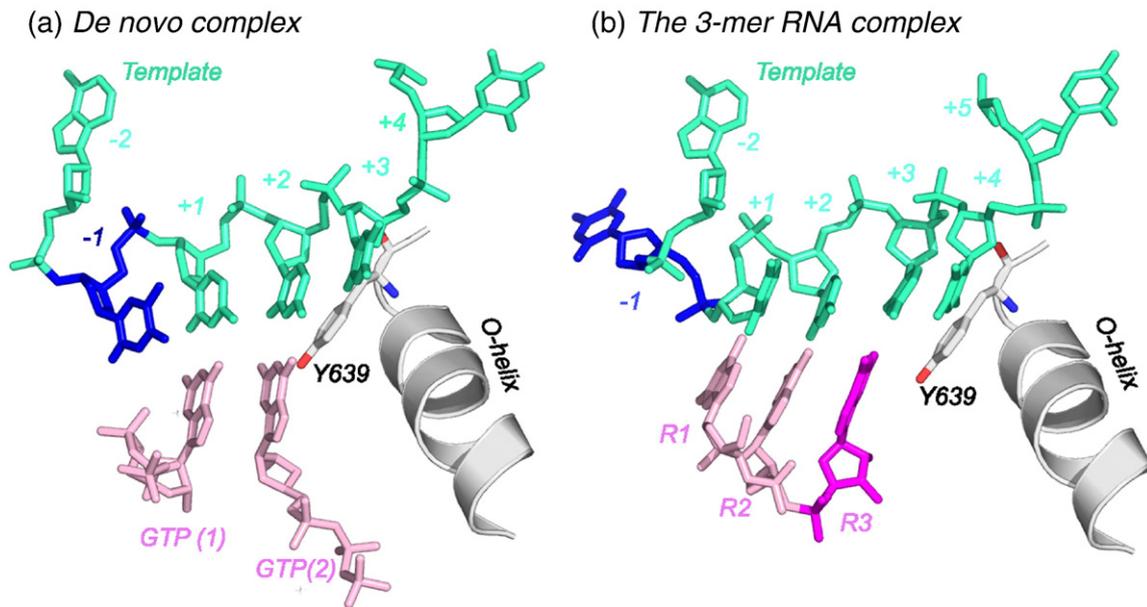


Figure 2. Comparison of the *de novo* RNA complex and a 3-mer RNA complex shows the novel initiating nucleotide-binding sites, different DNA template conformation and DNA scrunching. Both complexes are in the pre-insertion configuration. (a) The DNA template (light blue) forms an arched conformation during *de novo* synthesis. The +1 and +2 template residues are flanked by the -1 nucleotide on the template and Y639, and are poised to form base-pairing interactions with two initiating GTP nucleotides (pink). The +3 template nucleotide is in a flipped-out position. (b) The DNA template in the 3-mer complex adapts an A-form conformation in the heteroduplex with RNA transcription in the 3-mer RNA (pink) initiation complex. The -1 template residue forms a bulge from its position in the *de novo* complex, showing that DNA scrunching originates from the -1 position when the RNA transcript reaches three nucleotides in length.

rNTP rather than dNTP by RNAP may be achieved by K441 though an ionic interaction with the ribose 2'-OH. The second initiating nucleotide, 3'-dGTP(2), is seen forming charged interactions at the triphosphate moiety with the positively charged residues of the O-helix. The guanidino group of R627 and the ϵ -NH₃⁺ of K631 are located 2.9 Å and 3.3 Å, respectively, from the γ -phosphate of 3'-dGTP(2) (Figure 4(b)). Base-specific interactions with 3'-dGTP(2) include: R632 forms bipartite H-bonds with O⁶ and N7; and both H784 and R425 make van der Waals interactions with 2-NH₂. Selectivity for rGTP is likely *via* a hydrogen bonding interaction between the ribose 2'-OH and γ -O⁻ of D812.

The two 3'-dGTP nucleotides are bound nearly parallel with each other in the active site, with a 7° twisting angle and a separation of 3.6 Å, an arrangement that provides maximum base-stacking free energy. It appears that the affinity for the first nucleotide is strengthened by the second nucleotide; when we examined the structure of a complex formed with GMP at concentration equivalent to that of GTP in the T-CC ternary complex, no electron density corresponding to GMP was found (data not shown). The selectivity for two GTP nucleotides as initiating nucleotides is achieved primarily by the free energy from base-stacking, plus specific interactions between the polymerase residues, the guanine moieties of the initiating nucleotides and base complementarity interactions.

No translocation until after synthesis of a 3-mer RNA

During *de novo* synthesis, the first two nucleotides are located in the D and P-sites, while binding of a third nucleotide to the N-site is prohibited by interactions of the O-helix with the triphosphate of 3'-dGTP(2). However, after dinucleotide formation and release of PPI, and the conformational changes described above, the third NTP can bind in the N-site and be incorporated without requiring translocation of the dinucleotide. Therefore, T7 RNAP can synthesize a 3-mer RNA without translocation. After the 3-mer has been synthesized, its 3'-end extends into the N-site, and thus translocation is necessary for incorporation of the fourth nucleotide. Incorporation of the third nucleotide therefore begins the reiterative nucleotide addition cycle.

To visualize the initial translocation step, we compared structures of the DNA templates in the *de novo* complex described here and the 3-mer complex.^{6,7} The two complexes correspond to the steps before binding of the third and fourth NTP, and both structures are in the pre-insertion conformation. While the -17 to -2 region of the promoter is bound identically in the enzyme in the two complexes, the -1 to +5 region of the 3-mer complex has been translocated by one nucleotide. Consequently, residue -1 is bulged out in the 3-mer complex (Figure 2(b)), suggesting that DNA scrunching coincides with 3-mer RNA

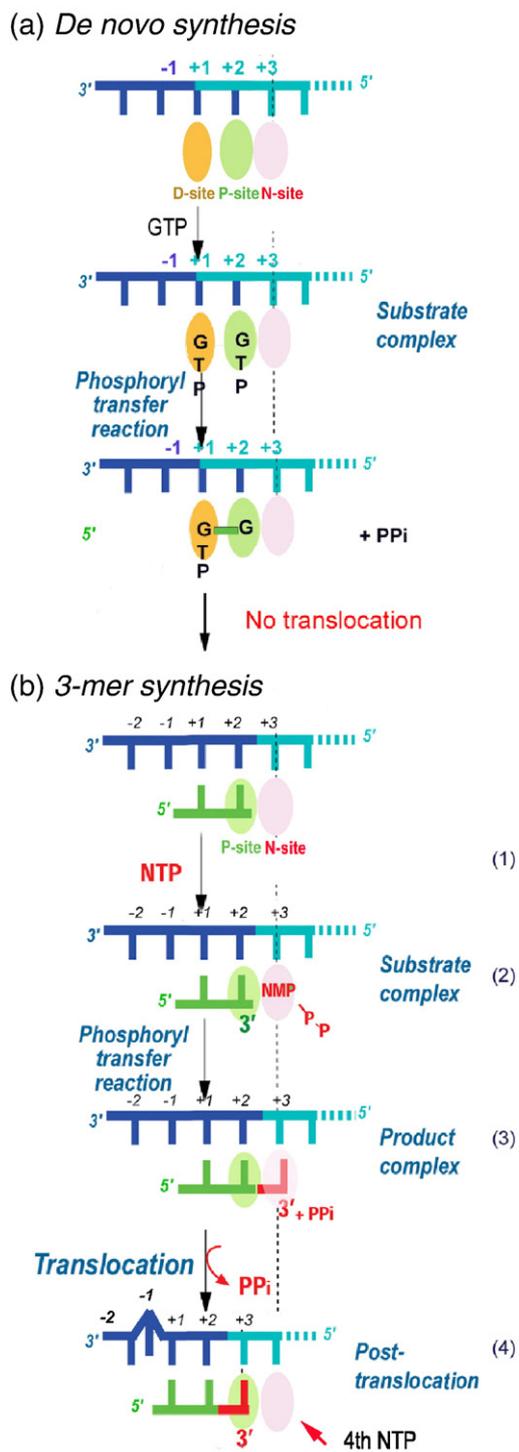


Figure 3. Schemes for nucleotide incorporation during *de novo* synthesis and 3-mer RNA formation. (a) Arrangement of the nucleotide-binding sites during *de novo* synthesis, where the D-site (orange oval) and P-site (green oval) denote the binding sites for the first and second initiating nucleotides, respectively. These sites are upstream of the elongating nucleotide-binding site (N-site, pink oval). The transcription start site on the template DNA (blue) is numbered +1. (b) Incorporation of the third nucleotide. The third NTP binds in the N-site (pink oval) without translocation of the dinucleotide, and then reacts with the 3'-OH of the dinucleotide in the P-site (green oval) to form the 3-mer RNA. RNAP uses only the P-site and N-site for subsequent cycles of nucleotide incorporation.

synthesis and originates from the -1 position. The bulged nucleotide fits into a pocket in RNAP bounded by residues R143, W201, L294, O417, and I761. These structures provide the first direct evidence for DNA scrunching.

Enzyme catalysis

During elongation, the phosphoryltransfer reaction occurs when the nucleophilic RNA 3'-OH in the P-site attacks the α -phosphate of the incoming NTP in the N-site. The reaction is catalyzed by two divalent metal ions: metal A facilitates nucleophilic attack by lowering the pK_a of the 3'-OH, whereas metal B stabilizes the pyrophosphate leaving group.¹³ The locations of the initiating nucleotides in the D and P-sites position the two reacting groups away from usual reacting locations (P and N-sites). This raises an interesting question about the mechanism of catalysis because of the known fixed locations of the catalytic residues.

In the *de novo* T7RNAP/T-CC/3'-dGTP ternary complex, two metal ions are observed: each is associated with the triphosphate moiety of the two 3'-dGTP nucleotides. The location of both ions is consistent with that of metal B in the elongation complex, so that only one metal ion is found in the active site of the *de novo* complex. The metal ion associated with 3'-dGTP(1) does not participate in any reaction. No metal ion corresponding to metal A is observed in the *de novo* structure due to the absence of the 3'-OH in 3'-dGTP.

We wanted to reconstruct the active site geometry for formation of the first phosphodiester bond. We superposed the active site palm domain of the only T7 RNAP ternary complex that contains an incoming nucleotide in the active site and two metal ions (PDB accession code 1s77), and the same domain of the *de novo* ternary complex. Alignment of the incoming nucleotide in 1s77 with 3'-dGTP(2) then allowed metal A to be modeled into the *de novo* complex (Figure 5). Although structure 1s77 is an elongation complex, modeling of metal A into the *de novo* complex should be reliable, because the polymerase palm domain that contains the active site maintains the same conformation throughout the entire transcription cycle. The active site geometry with a bound incoming nucleotide will therefore be almost identical in initiation or elongation complexes.⁶⁻⁹ Modeling suggests that both metal A and metal B can be coordinated by the catalytic residues D537 and D812 in the *de novo* complex. However, in the *de novo* complex, the attacking 3'-OH of GTP(1) and the leaving PPi of GTP(2) are skewed relative to their counterparts during elongation. As a result, the distance between metal A and the 3'-OH of GTP(1) is predicted to be ~ 3.0 Å in the *de novo* complex (Figure 5). This distance is significantly greater than the 2.3 Å found in the elongation complex, suggesting that T7 RNAP is optimized for catalyzing elongation using the P and N-sites, rather than the D and P-sites necessary for *de novo* synthesis.

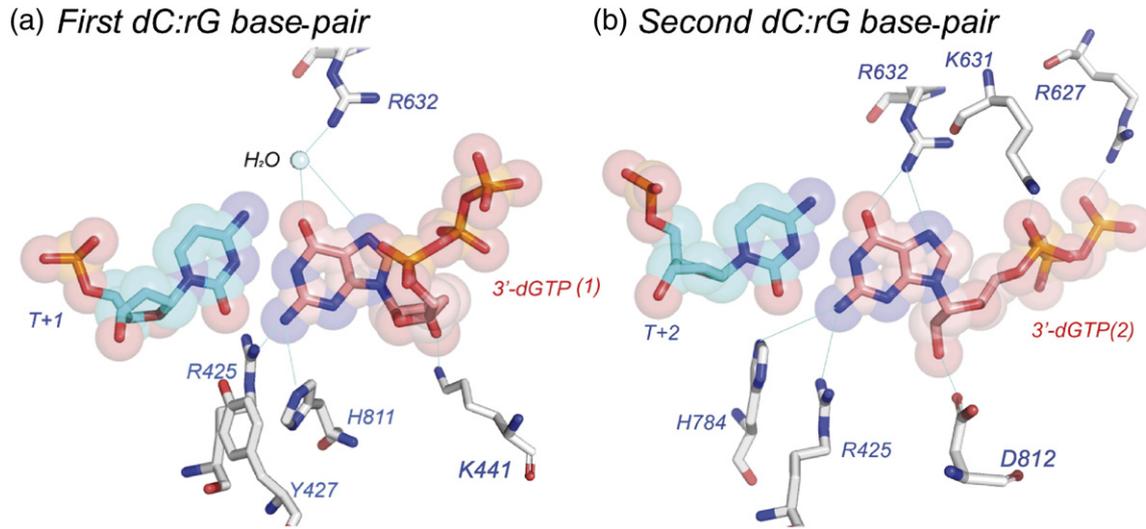


Figure 4. Guanine-specific interactions between the initiating GTP nucleotides and the active site residues of the polymerase. (a) Interaction of the first G:C base-pair between GTP (gold) and the +1 cytosine on the template (blue) with RNAP. (b) Interaction of the second G:C base-pair between GTP (gold) and the +2 cytosine (blue) on the template with the RNAP.

The increased distance between metal A and the 3'-OH of GTP(1) provides an atomic explanation for the slow formation of the first phosphodiester bond, relative to the formation of the same bond during elongation.

Non-consensus initial transcribed sequence

The initial transcribed sequence refers to the region of DNA template from positions +1 to +6,

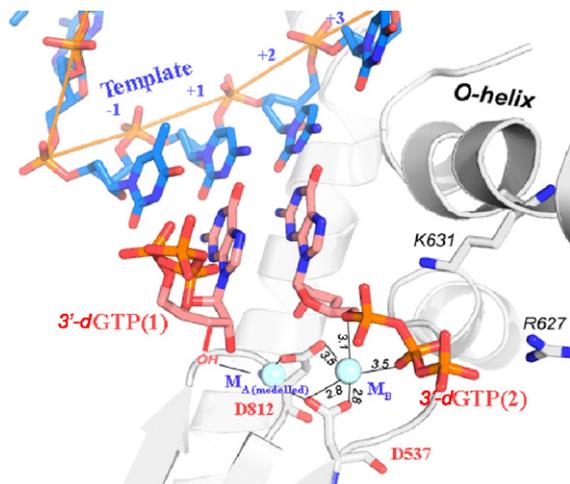


Figure 5. Formation of the first phosphodiester bond. The metal ion associated with 3'-dGTP(2) whose position is consistent with metal B (M_B) is observed in the *de novo* complex. The metal ion associated with 3'-dGTP(1) is irrelevant to the reaction and is omitted for clarity. The metal ion A (M_A) is absent and is modeled on the basis of a T7 RNAP elongation complex containing an incoming nucleotide (PDB code 1s77) after superposition of the two active site domains. The metal coordination shows increased distance between the 3'-OH of the ribose and metal A, suggesting slower chemistry than phosphodiester bond formation during elongation.

which for T7 promoters is a conserved sequence 3'-⁺¹CCCTCT⁺⁶-5'. To ask whether the preference for GTP as the first initiating nucleotide arises from the fact that T7 RNAP interacts directly with the consensus initial transcribed sequence, we examined the DNA conformation in various binary complexes. Both consensus and non-consensus initial transcribed sequence showed poor electron density in binary complexes, suggesting that polymerase has weak affinity for the single-stranded template strand for this region of the template.

Among the ternary complexes, only the T-CC complex shows clear density for the initiating nucleotides. Weak density that likely corresponds to a low occupancy of 3'-dATP was observed in the T-TT ternary complex. T7 RNAP uses ATP as the initiating nucleotide at two natural promoters on the T7 genome. The σ^{OL} and $\sigma^{2.5}$ promoters are, however, rather weak,^{14–16} observations that are consistent with reduced binding of ATP in the *de novo* initiation complex. The ternary complexes with T-AA and T-GG showed no density for 3'-dUTP or 3'-dCTP, suggesting that their dissociation constants may be lower than the concentration (5 mM) used for crystallographic studies.

Initiating nucleotides in promoter melting

We measured the effect of initiating nucleotides on promoter melting by using conformation-sensitive DNA substrates that contains a 2-aminopurine incorporated at the -4 position, which is where promoter unwinding begins, during which base-unstacking of 2-aminopurine increases the quantum yield of fluorescence decay.^{17–19} Fluorescence intensity is therefore proportional to the amount of promoter DNA unwound. The DNA substrates used in these experiments are similar to those used in structural studies, except that they were perfect duplexes (Figure 6).

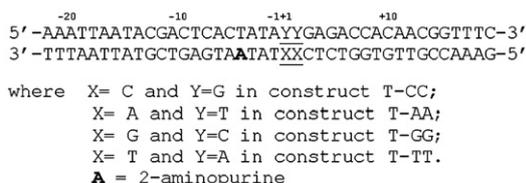


Figure 6. Substrates used in pre-steady-state kinetic studies.

The affinity of each promoter variant for T7 RNAP was determined in the presence or in the absence of the corresponding initiating nucleotides using an established procedure.¹ Despite variations in sequence, all four DNA substrates showed affinity similar to that of T7 RNAP in the absence of an initiating nucleotide (Figure 7). The K_d values were determined to be: T-CC, 0.63 μ M; T-GG, 0.60 μ M; T-AA, 0.65 μ M; and T-TT, 0.72 μ M. These data confirm the conclusion made from the binary complex structures: residues at positions +1 and +2 on the template strand contribute little to overall promoter binding by RNAP. However, significant differences are observed in the presence of initiating nucleotides. The largest fluorescence change was observed with the T-CC promoter, where the amplitude of fluorescence increased twofold over that seen in the absence of GTP (Figure 7). This leads to a 4.8-fold decrease in K_d to 0.13 μ M, suggesting that the presence of GTP strengthens DNA binding and increases promoter opening by RNAP. Similarly, the presence of ATP results in a 1.8-fold decrease in K_d value to 0.39 μ M at the T-TT promoter. The smaller decrease, relative to that seen with GTP, is consistent with the lower occupancy of 3'-dATP in the T-TT ternary complex. UTP or CTP had little effect on fluorescence changes at their respective promoters, suggesting that they do not stabilize an open promoter significantly.

Discussion

We have presented detailed structural analyses of ternary complexes of T7 RNAP captured during *de novo* synthesis and before formation of the first phosphodiester bond. These complexes are distinct from previously determined complexes of T7RNAP captured during the initiation⁶ and elongation^{8,9} phases of transcription.

The DNA substrates used in our structural studies are only partial duplexes. Nonetheless, the complexes made with single-stranded template DNA faithfully represent conformations of substrates containing downstream DNA and polymerase during *de novo* synthesis. The DNA from -17 to -1 in the *de novo* complex adopts a conformation identical with that in the binary 17 bp open promoter complex (PDB code 1cez). The locations of the +1 and +2 template residues within the active site of T7 RNAP are therefore fixed by the promoter. The registry of the single-stranded DNA template shows

that binding of a single-stranded DNA template in the partial duplex within the active site of the polymerase is identical with that of the corresponding sequence of an active transcription bubble; the register and orientation of the template is not changed in the absence of the non-template strand and downstream duplex DNA.^{8,9} Furthermore, T7 RNAP does not have a sequence-specific interaction with downstream DNA.⁸⁻¹¹ The interactions are exclusively between the phosphate backbone and the positively charged DNA-binding cleft in the polymerase. T7 RNAP is therefore able to recognize and position the DNA template in the absence of a downstream duplex, a feature that may contribute to T7 RNAP's ability to transcribe both single- and double-stranded DNA templates.

Transcription by T7 RNAP is relatively fast during elongation, and Watson-Crick pairing rules are followed for nucleotide incorporation. In contrast, *de novo* synthesis is slow and there is a strong bias in nucleotide utilization. Fifteen of the 17 promoters in the bacteriophage T7 genome initiate synthesis with GTP, and 13 do so with pppGpG.²⁰ On the basis of the ternary complexes of T7RNAP reported here, the structural basis for the compositional bias and slowed rate of nucleotide incorporation during *de novo* synthesis derives from several features: (1) utilization of a novel NTP-binding site, the D-site, adjacent to the previously characterized P-site and N-site;¹⁰ (2) exploitation of favorable stacking interactions between the incoming purine nucleotides in the D and P-sites as they are positioned by residues in the active site; and (3) less favorable geometry during the formation of the first phosphodiester bond compared to subsequent rounds.^{9,21,22}

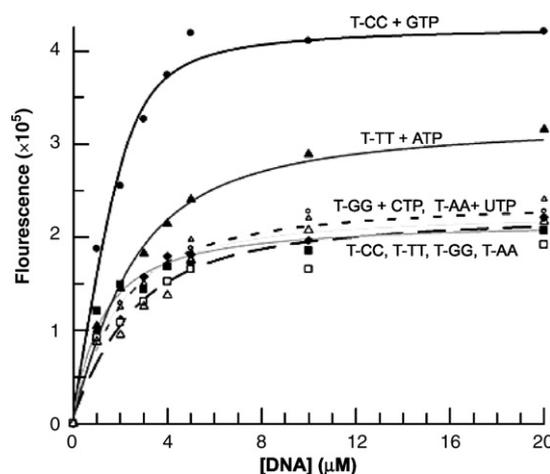


Figure 7. Effects of initiating NTP on promoter melting. Observed fluorescence changes are plotted as a function of DNA concentration at a constant concentration of the polymerase, in the absence (open symbols) or in the presence (filled symbols) of initiating nucleotides. Four DNA substrates are measured, T-CC (circle), T-TT (triangle), T-AA (square) and T-GG (diamond), either in the presence or in the absence of their cognate initiating nucleotides.

The independent binding sites for the initiating NTPs are distinct from the NTP-binding site during elongation and reveal several unique features of *de novo* RNA synthesis that clearly differentiate this phase from subsequent steps. The mechanism for selection of initiating NTPs differs from that of an elongating NTP. During elongation, the incoming nucleotide induces conformational changes in the fingers domain of T7 RNAP, which is thought to be important for discriminating against incorrect nucleotides. This ability is critical for polymerase fidelity. Studies of members of the DNA Pol1 family, to which T7 RNAP belongs, suggest that conformational changes induced by binding of the correct dNTP provide the mechanism for discrimination. The desolvation effect associated with the conformational change is thought to enlarge the affinity difference between the correct and incorrect nucleotides.^{23–27}

Binding of the initiating nucleotides, however, is not accompanied by structural changes, suggesting that a different mechanism exists for correct initiating NTP selection. The initiating NTPs make direct contacts with the active-site residues of RNAP. Nucleotides other than GTP will be discriminated against due to the lack of these specific interactions. For example, the interaction between the 2-NH₂ of guanine and R425 and Y427 of RNAP is absent if the initiating nucleotide is adenine. Although the H-bond between O⁶ of guanine and R632 of RNAP can be formed using the 6-NH₂ of adenine, the 2-NH₂ interaction has been shown to be important. ITP has a tenfold higher K_m value during initiation than GTP.²⁸ Similarly, the interaction with N7 of GTP(1) with R632 of T7 RNAP explains the biochemical observation¹² that usage of 7-deazaGTP is inefficient during *de novo* synthesis. The differences between NTP discrimination during elongation and initiation can be likened to the former using an induced-fit mechanism, whereas an initiating NTP uses a lock-and-key mechanism.

In the enzyme active site, non-polar contributions to base-stacking interactions predominate. Non-polar contributions to base-stacking have been calculated by combining van der Waals interactions with the hydrophobic effect in a thermodynamic transfer cycle.²⁹ Purine:purine stacking provides far more free energy than purine:pyrimidine stacking, with pyrimidine:pyrimidine stacking providing the least. Among all stacked purines, G:G stacking provides the most favorable free energy (−9.74 kcal/mol), followed by G:A (−9.31 kcal/mol), A:G (−8.65 kcal/mol) and A:A (−8.52 kcal/mol). These data suggest that the natural promoters initiating with GA or AG are weaker than promoters initiating with GG. This is known to be the case. Comparison to the direct interaction with the polymerase, where one H-bond is estimated to contribute to ~1 kcal/mol to binding energy,³⁰ the purine:purine stacking interactions would appear to dominate ground-state stabilization for *de novo* RNA synthesis. The general trend of these calculated free energies is in accord with the apparent stabilities of the ternary

complexes examined in this study. Those containing 3'-dGTP are most stable, those containing 3'-dATP have only partial nucleotide occupancy, while ternary complexes containing 3'-dCTP or 3'-dUTP could not be detected.

Biochemical evidence in support of the importance of stacking interactions comes from studies on promoter mutagenesis. Substitution of cytosine for guanosine at template position +1 results in a ~15-fold decrease in promoter strength.³¹ C:G stacking provides only −6.22 kcal/mol, significantly less than for G:G, and the reduction in promoter activity argues against Watson-Crick base-pairing being the primary driving force for incoming nucleotide selection. Switching the +1 base-pair from dC:rG to dG:rC would not be predicted *a priori* to cause such a significant change in promoter strength.

Different binding modes are observed for the initiating and elongating NTP. While the polymerase interacts with elongating NTPs through their common triphosphate moiety, it makes base-specific interactions with the initiating GTP. T7 RNAP does not interact with the triphosphate of the first nucleotide, GTP(1), which is reflected in the different K_d values for the first and second NTP.^{1,32} The lack of any interaction with the triphosphate may actually benefit translocation. As the product RNA retains the triphosphate of GTP(1), a strong interaction between it and the polymerase may impede movement of the DNA-RNA heteroduplex. The lack of interaction with the triphosphate of GTP(1) also explains why both GMP and GDP can initiate transcription *in vitro*.²⁸

T7 RNAP exploits specific enzyme-nucleotide substrate interactions and innate base-stacking differences in stabilizing heteroduplex formation during *de novo* synthesis. However, preferential binding of GTP would likely lead to misincorporation in subsequent steps of transcription. By separating *de novo* RNA synthesis, which requires two initiating NTPs to bind in the D and P-sites, from elongation synthesis, where the 3'-OH of the nascent RNA and incoming NTP are in the P and N-sites, the enzyme is able to accommodate both nucleotide-specific initiation and nucleotide non-specific elongation. A trade-off for having separate binding sites for the initiating NTPs is a slower chemistry of phosphodiester synthesis. This slower rate is explained by the off-set locations of both the 3'-OH of GTP(1) in the D-site and the α -phosphate of GTP(2) in the P-site, relative to the fixed positions of the catalytic aspartate residues and the magnesium ions.^{1,32}

Although T7 RNAP is capable of unwinding the promoter DNA, binding to the initiating nucleotide shifts the equilibrium towards open promoter conformation. Structures of all binary complexes showed disordered, single-stranded templates, even though they mimic an open promoter configuration. This suggests that the open promoter is not poised properly for transcription in the absence of NTPs. The template strand becomes ordered in the presence of GTP because the initiating nucleotides

provide additional stabilization interactions. However, the positions of the +1 and +2 template nucleotides during *de novo* synthesis are still governed by the -17 to -1 promoter DNA, which is tightly bound to RNAP. As a consequence, the +1 and +2 template residues in the *de novo* complex must be upstream of the templating position used during elongation. The structural data shown here and previous kinetic studies are consistent in showing that GTP best stabilizes the open promoter conformation. Initiating nucleotides with lower affinity for the polymerase, e.g. CTP and UTP, are unable to stabilize and guide the template adequately in an open promoter. Both structural and kinetic data indicate that stabilization of the template DNA by an NTP is related directly to its affinity for RNAP.

Site-directed mutagenesis of residues surrounding the D-site or P-site of RNAP, R425A and Y427A, results in reduced polymerase activity and a reduced ability to open the transcription bubble.^{32–34} We reason that these mutations decrease the affinity of the initiating nucleotides for the polymerase, which in turn reduces promoter unwinding. The initiating nucleotide is known to stabilize the open promoter configuration,^{1,32,35–37} we have now shown that the initiating nucleotide does so by guiding the open promoter into the active site as stronger interactions are formed in the ternary complex.

The footprint of *Escherichia coli* and T7 RNAP on DNA enlarges during initiation, and a model where the DNA becomes scrunched within the enzyme has been proposed.^{38–40} Using single-molecule FRET as a probe for structural changes, it has been shown recently that scrunching in *E. coli* RNAP occurs when the RNA product is >2-mer.^{22,41} Our structural data show conclusively that scrunching in T7 RNAP occurs at the 3-mer stage. The initiating NTP-binding sites (D and P-sites) allow T7 RNAP to synthesize a 3-mer RNA without translocation, because the third NTP can bind to the N-site after synthesis of the dinucleotide is completed. This feature of the initiation reaction may serve to increase transcription efficiency by stabilizing the 2-mer heteroduplex in the active site.

When the 3-mer RNA is translocated, the template DNA at the -1 position is scrunched into a pocket of RNAP. Assuming that the size of the pocket does not change, we estimate that four to five nucleotides of template DNA can be scrunched in this manner. No equivalent pocket for the non-template strand is apparent in the structure.

The innate stability of G:G stacks may have been exploited by enzymes other than T7 RNAP for initiation. The preference for GTP as an initiating nucleotide extends to many RNA-directed RNA polymerases from pathogenic *Flaviviridae* family members, including hepatitis C virus (HCV), West Nile virus and bovine viral diarrhoea virus (BVDV). Bacteriophage Φ 6 RNA polymerase, whose mechanism of initiation has been characterized structurally

with a series of crystal structures captured in *de novo* synthesis, also uses GTP as the initiating nucleotide.^{42,43} In an attempt to identify the common features in polymerases for initiating *de novo* RNA synthesis with GTP, we compared the T7 RNAP *de novo* complex with the structures of HCV, BVDV and Φ 6 RNA polymerases.^{44–46,34} There is no corresponding ternary structure for HCV and BVDV polymerases, and we therefore docked the initiating GTP nucleotides from the T7 RNAP complex onto the HCV or BVDV RNA polymerase structures after superposition of the active site palm domains. The initiating nucleotides were positioned into the HCV and BVDV active sites without steric clash (Figure 8).

Two common features in the active sites of these RdRP polymerases are noteworthy in regard to *de novo* synthesis. There is a positive charged residue, often arginine (Figure 8), that can make a stabilizing hydrogen bond with O⁶ of the initiating GTP nucleotides and thereby provides discrimination against adenine during *de novo* synthesis. In addition, a conserved tyrosine is stacked against the first G:C base-pair, which may stabilize the binding of the first NTP and formation of the *de novo* complex.

The tyrosine in BVDV RNAP has been substituted with alanine; mutant Y581A gave the predicted decrease in transcription initiation but also showed an unanticipated increase in activity for initiation with a primer.⁴⁷ An RNA primer enables the polymerase to bypass the initiation phase and enter directly into the elongation phase.⁴⁸ The effect of the Y581A mutation is to decrease the ability of the enzyme to perform *de novo* synthesis but to retain full capacity to synthesize RNA in the elongation mode. The different effects are consistent with the idea that *de novo* synthesis is a unique phase of transcription for RNA-directed RNA polymerases, as it is for the bacteriophage DNA-directed RNA polymerases.

Experimental Procedures

T7 RNA polymerase and substrate DNA preparation

T7RNAP was purified as described⁴⁹ from *E. coli* strain BL21 containing the plasmid pAR1219.⁵⁰ All DNA oligonucleotides were purchased from Integrated DNA Technology Inc., and purified by HPLC on a reverse phase C4 column. The 3'-deoxynucleoside triphosphates were purchased from the TriLink BioTechnologies. Ultrapure GTP, ATP, GTP and UTP were purchased from GE Healthcare Life Sciences.

Crystallographic studies

Four DNA partial duplexes, whose sequence are shown in Table 1, were formed by annealing equal molar amounts of template and non-template strands in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, heated to 75 °C for 5 min and cooled to 20 °C over the course of 2 h.

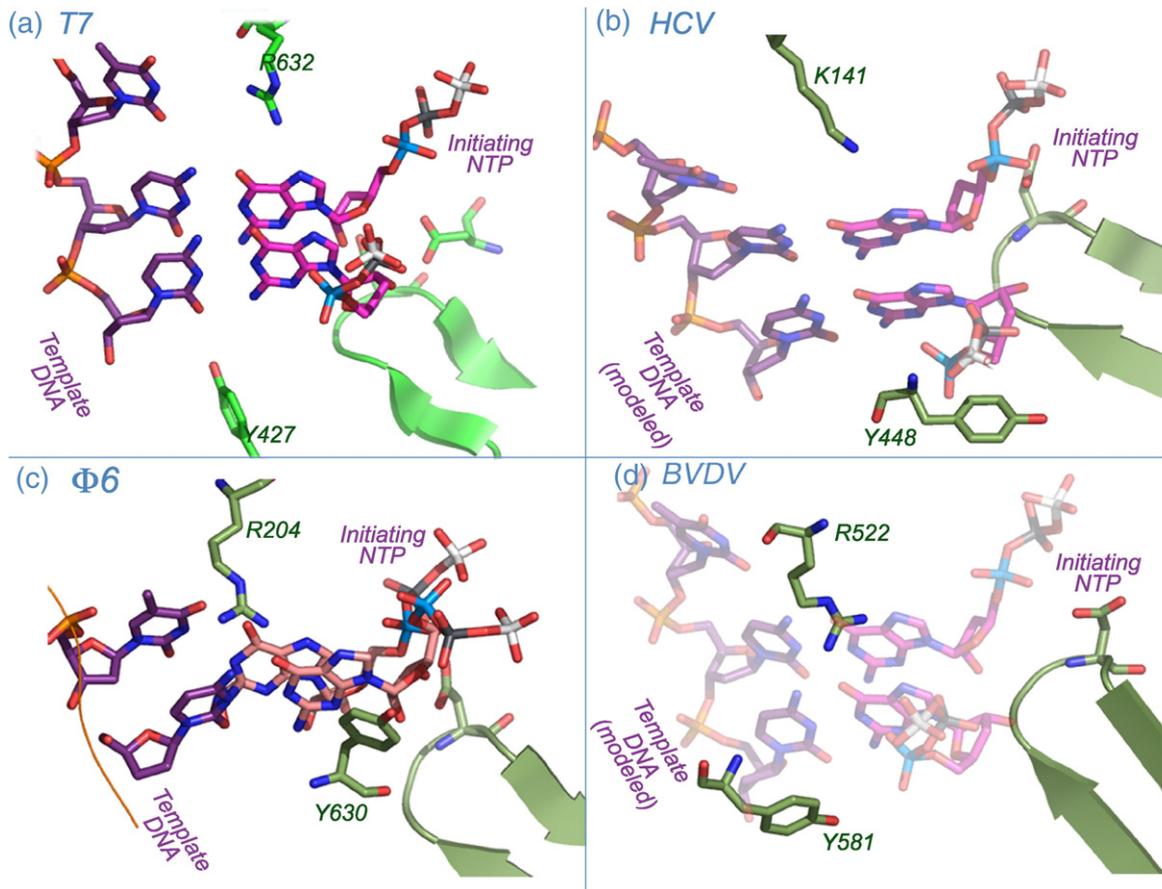


Figure 8. Conserved interactions in RNAP complexes during *de novo* synthesis. (a) T7 RNAP; (b) hepatitis C virus (HCV) RNA polymerase; (c) $\Phi 6$ RNA polymerase; (d) bovine viral diarrhea virus (BVDV) RNA polymerase. These RNAPs have a conserved positively charged residue (T7: R632; HCV: K141; $\Phi 6$: R204; BVDV: R522) that interacts with the second initiating nucleotide, and a conserved tyrosine residue (T7: Y427; HCV: Y448; $\Phi 6$: Y630; BVDV: Y581) that positions the first nascent base-pair between the initiating nucleotide and the +1 template residue.

T7 RNAP-DNA binary complexes were prepared by mixing 200 μM T7 RNAP and 240 μM DNA in 50 mM Tris-HCl, (pH7.5), 100 mM NaCl, 10 mM MgCl_2 , 1 mM DTT, and incubating at 25 $^\circ\text{C}$ for 5 min. Complexes were crystallized using 50 mM Tris-HCl (pH 7.5–8.0), 200 mM Li_2SO_4 , 10 mM MgCl_2 , 1 mM DTT, 20–30% (w/v), PEG 8000 using the vapor-diffusion method at 20 $^\circ\text{C}$. Ternary complexes of T7RNAP/DNA were prepared using 250 μM T7RNAP/DNA complex with the corresponding 3'-dNTP (5 mM). Complexes formed extremely thin crystals with dimensions of 500 nm \times 300 nm \times 5 nm. Diffraction data were collected at Advanced Photon Source (APS, ID14) at 100K using X-rays of wavelength 1.05 Å , and were recorded on an ADQX charge-coupled device detector. Diffraction data were reduced with the program HKL2000.⁵¹ Structures were solved by the molecular replacement method using programs AMoRe,⁵² and CNS.⁵³ The T7 RNAP-promoter complex (PDB code 1QLN) was used as a search model for phase calculations. Structures were refined with programs CNS and Refmac,⁵⁴ and rebuilt with program O.⁵⁵

Kinetic measurements

The sequences of the substrates are shown in Figure 6: 2-aminopurine was incorporated into the template strand

by substitution of the –4 adenine. Duplex DNAs were prepared by annealing equal molar amounts of the two DNA strands at 75 $^\circ\text{C}$ for 5 min, followed by cooling to 20 $^\circ\text{C}$ over the course of 2 h.

Fluorescence measurements

Equilibrium DNA-binding experiments were conducted at 25 $^\circ\text{C}$ in a 2 ml quartz cuvette using a spectrofluorimeter (Photon Technology international, Inc.); fluorescence signals were analyzed using the software Felix. Fluorescence titrations varied the amount of DNA added to a constant amount of T7 RNAP (5 μM) in 200 μL of the reaction buffer (40 mM Tris-HCl (pH 7.9), 10 mM MgCl_2 , 10 mM DTT, 2 mM spermidine). The effect of the initiating nucleotides on promoter opening was measured in a second titration that varied the amount of DNA added to a constant amount of T7 RNAP-NTP mixture (2.5 μM T7 RNAP + 5 mM 3'-dNTP) in 200 μL of reaction buffer. Fluorescence values were corrected as necessary by the dilution factor and by subtracting the corresponding fluorescence values of 2-aminopurine DNA.

The observed fluorescence changes were plotted *versus* DNA concentration in the presence or in the absence of the initiating NTP; in order to obtain K_d

values, the data were transformed using the program GraFit:

$$F_c F_{\max} \{ [(k_d + E_t + D_t) - \sqrt{((k_d + E_t + D_t)^2 - 4(E_t D_t))}] / 2 \}$$

where E_t and D_t are the concentrations of enzyme and DNA, K_d is the equilibrium dissociation for the T7 RNAP-DNA complex, and F_{\max} is the maximum fluorescence change at saturating concentrations of DNA.

Protein Data Bank accession code

The coordinates for T-CC binary and ternary complexes have been deposited to the Protein Data Bank with accession codes 2PI5 and 2PI4, respectively.

Acknowledgements

We thank I. Molineux for critical reading of the manuscript, and K. Johnson for advice and use of his spectrofluorimeter. We thank and acknowledge the staff and facilities of the Advanced Photon Source ID14. This work was supported, in part, by a grant from the Welch Foundation (F-1592).

References

- Bandwar, R. P., Jia, Y., Stano, N. M. & Patel, S. S. (2002). Kinetic and thermodynamic basis of promoter strength: multiple steps of transcription initiation by T7 RNA polymerase are modulated by the promoter sequence. *Biochemistry*, **41**, 3586–3595.
- Guajardo, R., Lopez, P., Dreyfus, M. & Sousa, R. (1998). NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation. *J. Mol. Biol.* **281**, 777–792.
- Anand, V. S. & Patel, S. S. (2006). Transient state kinetics of transcription elongation by T7 RNA polymerase. *J. Biol. Chem.* **281**, 35677–35685.
- Chamberlin, M. & Ring, J. (1973). Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template specificity of the enzyme. *J. Biol. Chem.* **248**, 2235–2244.
- Selisko, B., Dutartre, H., Guillemot, J. C., Debarnot, C., Benarroch, D., Khromykh, A. *et al.* (2006). Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. *Virology*, **351**, 145–158.
- Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature*, **399**, 80–83.
- Cheetham, G. M. & Steitz, T. A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science*, **286**, 2305–2309.
- Yin, Y. W. & Steitz, T. A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science*, **298**, 1387–1395.
- Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassilyev, D. G. & Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature*, **420**, 43–50.
- Yin, Y. W. & Steitz, T. A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, **116**, 393–404.
- Temiakov, D., Patlan, V., Anikin, M., McAllister, W. T., Yokoyama, S. & Vassilyev, D. G. (2004). Structural basis for substrate selection by t7 RNA polymerase. *Cell*, **116**, 381–391.
- Kuzmine, I., Gottlieb, P. A. & Martin, C. T. (2003). Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J. Biol. Chem.* **278**, 2819–2823.
- Steitz, T. A. & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA*, **90**, 6498–6502.
- Golomb, M. & Chamberlin, M. (1974). Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major *in vitro* transcripts by gel electrophoresis. *J. Biol. Chem.* **249**, 2858–2863.
- Niles, E. G. & Condit, R. C. (1975). Translational mapping of bacteriophage T7 RNAs synthesized *in vitro* by purified T7 RNA polymerase. *J. Mol. Biol.* **98**, 57–67.
- McAllister, W. T. & Carter, A. D. (1980). Regulation of promoter selection by the bacteriophage T7 RNA polymerase *in vitro*. *Nucl. Acids Res.* **8**, 4821–4837.
- Ward, D. C., Reich, E. & Stryer, L. (1969). Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives. *J. Biol. Chem.* **244**, 1228–1237.
- Guest, C. R., Hochstrasser, R. A., Sowers, L. C. & Millar, D. P. (1991). Dynamics of mismatched base pairs in DNA. *Biochemistry*, **30**, 3271–3279.
- Xu, D., Evans, K. O. & Nordlund, T. M. (1994). Melting and premelting transitions of an oligomer measured by DNA base fluorescence and absorption. *Biochemistry*, **33**, 9592–9599.
- Dunn, J. J. & Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166**, 477–535.
- Theis, K., Gong, P. & Martin, C. T. (2004). Topological and conformational analysis of the initiation and elongation complex of T7 RNA polymerase suggests a new twist. *Biochemistry*, **43**, 12709–12715.
- Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, **314**, 1139–1143.
- Kunkel, T. A. (2004). DNA replication fidelity. *J. Biol. Chem.* **279**, 16895–16898.
- Doublet, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, **391**, 251–258.
- Franklin, M. C., Wang, J. & Steitz, T. A. (2001). Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell*, **105**, 657–667.
- Li, Y., Korolev, S. & Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525.
- Li, Y. & Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP-, and ddGTP-trapped ternary complex of KlenTaq1: insights into nucleotide incorporation and selectivity. *Protein Sci.* **10**, 1225–1233.

28. Martin, C. T. & Coleman, J. E. (1989). T7 RNA polymerase does not interact with the 5'-phosphate of the initiating nucleotide. *Biochemistry*, **28**, 2760–2762.
29. Friedman, R. A. & Honig, B. (1995). A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.* **69**, 1528–1535.
30. Fersht, A. (1985). *Enzyme Structure and Mechanism*, 2nd edit, W.H. Freeman and Company, New York.
31. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry*, **39**, 10419–10430.
32. Stano, N. M., Levin, M. K. & Patel, S. S. (2002). The +2 NTP binding drives open complex formation in T7 RNA polymerase. *J. Biol. Chem.* **277**, 37292–37300.
33. Imburgio, D., Anikin, M. & McAllister, W. T. (2002). Effects of substitutions in a conserved DX(2)GR sequence motif, found in many DNA-dependent nucleotide polymerases, on transcription by T7 RNA polymerase. *J. Mol. Biol.* **319**, 37–51.
34. Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H. & Stuart, D. I. (2001). A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, **410**, 235–240.
35. Sastry, S. S. & Ross, B. M. (1997). Probing the mechanisms of T7 RNA polymerase transcription initiation using photochemical conjugation of psoralen to a promoter. *Biochemistry*, **36**, 3133–3144.
36. Villemain, J., Guajardo, R. & Sousa, R. (1997). Role of open complex instability in kinetic promoter selection by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **273**, 958–977.
37. Kuzmine, I. & Martin, C. T. (2001). Pre-steady-state kinetics of initiation of transcription by T7 RNA polymerase: a new kinetic model. *J. Mol. Biol.* **305**, 559–566.
38. Ikeda, R. A. & Richardson, C. C. (1986). Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proc. Natl Acad. Sci. USA*, **83**, 3614–3618.
39. Carpousis, A. J. & Gralla, J. D. (1985). Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation. Footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation. *J. Mol. Biol.* **183**, 165–177.
40. Straney, D. C. & Crothers, D. M. (1987). Comparison of the open complexes formed by RNA polymerase at the Escherichia coli lac UV5 promoter. *J. Mol. Biol.* **193**, 279–292.
41. Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S. & Ebricht, R. H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, **314**, 1144–1147.
42. Frilander, M., Poranen, M. & Bamford, D. H. (1995). The large genome segment of dsRNA bacteriophage phi6 is the key regulator in the in vitro minus and plus strand synthesis. *Rna*, **1**, 510–518.
43. Makeyev, E. V. & Bamford, D. H. (2000). Replicase activity of purified recombinant protein P2 of double-stranded RNA bacteriophage phi6. *EMBO J.* **19**, 124–133.
44. Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. & Miyano, M. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure*, **7**, 1417–1426.
45. Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. F. & Weber, P. C. (1999). Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature Struct. Biol.* **6**, 937–943.
46. Choi, K. H., Groarke, J. M., Young, D. C., Kuhn, R. J., Smith, J. L., Pevear, D. C. & Rossmann, M. G. (2004). The structure of the RNA-dependent RNA polymerase from bovine viral diarrhea virus establishes the role of GTP in de novo initiation. *Proc. Natl Acad. Sci. USA*, **101**, 4425–4430.
47. Laurila, M. R., Makeyev, E. V. & Bamford, D. H. (2002). Bacteriophage phi 6 RNA-dependent RNA polymerase: molecular details of initiating nucleic acid synthesis without primer. *J. Biol. Chem.* **277**, 17117–17124.
48. Daube, S. S. & von Hippel, P. H. (1992). Functional transcription elongation complexes from synthetic RNA-DNA bubble duplexes. *Science*, **258**, 1320–1324.
49. Jeruzalmi, D. & Steitz, T. A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* **17**, 4101–4113.
50. Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA*, **81**, 2035–2039.
51. Otwinowski, Z. & Minor, M. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
52. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallog. sect. D*, **57**, 1367–1372.
53. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
54. Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog. sect. D*, **57**, 122–133.
55. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.

Edited by R. Ebricht

(Received 16 October 2006; received in revised form 14 March 2007; accepted 14 March 2007)
Available online 21 March 2007