Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex

Katsuhiko S. Murakami, Shoko Masuda, Elizabeth A. Campbell, Oriana Muzzin, Seth A. Darst*

The crystal structure of *Thermus aquaticus* RNA polymerase holoenzyme ($\alpha_2\beta\beta'\omega\sigma^A$) complexed with a fork-junction promoter DNA fragment has been determined by fitting high-resolution x-ray structures of individual components into a 6.5-angstrom resolution map. The DNA lies across one face of the holoenzyme, completely outside the RNA polymerase active site channel. All sequence-specific contacts with core promoter elements are mediated by the σ subunit. A universally conserved tryptophan is ideally positioned to stack on the exposed face of the base pair at the upstream edge of the transcription bubble. Universally conserved basic residues of the σ subunit provide critical contacts with the DNA phosphate backbone and play a role in directing the melted DNA template strand into the RNA polymerase active site. The structure explains how holoenzyme recognizes promoters containing variably spaced −10 and −35 elements and provides the basis for models of the closed and open promoter complexes.

The structure of the bacterial RNA polymerase (RNAP) holoenzyme (*1*) reveals the architecture of the molecular machinery at the heart of gene expression. Nevertheless, many questions about transcription initiation remain: How does holoenzyme recognize variably spaced −10 and −35 hexamers of promoter DNA? How is the transcription bubble formed? Here, we report the crystal structure of *Thermus aquaticus* (*Taq*) RNAP holoenzyme (subunit composition $\alpha_2\beta\beta'\omega\sigma^A$) complexed with a promoter DNA fragment at 6.5 Å resolution, which begins to answer these questions.

The interaction of RNAP holoenzyme (R) with promoter DNA (P) initiates a series of structural transitions from the initial closed promoter complex (RP_c) to the transcription-competent open complex (RP_o). The double-stranded DNA is melted over a region spanning the transcription start site at +1, in the following process [reviewed in (*2*)]:

$$R + P \leftrightarrow RP_c \leftrightarrow I \leftrightarrow RP_o$$

I represents at least one, and possibly more, intermediate states of the complex. RP_o is in rapid equilibrium with I, and this equilibrium depends on many factors, including temperature, Mg^{2+} concentration, and promoter sequence (*2*). This presents challenges for structure determination, where large quantities of a homogeneous complex must be isolated. To overcome this, we used fork-junction DNA (*3*) as the template for complex preparation. Fork-junction DNA, which contains a double-strand-

The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: darst@rockefeller.edu

ed −35 element and a mostly single-stranded −10 element (Fig. 1A), forms stable complexes with RNAP holoenzyme (fig. S1A), independent of temperature. The base pair at −12 is required for optimal binding (*3*), which suggests that the fork-junction template mimics the structure of the upstream edge of the transcription bubble in RP_o. The fork-junction sequence used here (*4*) also contains an extended −10 motif (*5*). The holoenzyme/fork-junction DNA complex (RF) mimics many properties of RP_o, including the following: (i) RF, like RP_o, is resistant to the inhibitor heparin (fig. S1A), (ii) mutations in both the promoter and the RNAP that are deleterious to open-complex formation cause parallel reductions in fork-junction binding affinity, and (iii) formation of RF is a multistep process, and some of the intermediates along the pathway share common properties with the intermediates in RP_o formation (*6*). Moreover, we also crystallized a complex of holoenzyme with the same promoter fragment except that it was entirely double-stranded from −40 to −7. Difference Fourier analysis revealed no changes; the −12 position was base paired, and the added template strand from −11 to −7 lacked density because of disorder. Thus, the structure of the fork junction is an excellent model system to study the structural properties of RP_o.

Crystallization and structure determination. RF formed large single crystals, but diffraction from the crystals was highly anisotropic (4.5 × 6.5 Å). The structure was solved at 6.5 Å resolution by the method of multiple isomorphous replacement (Table 1). At this resolution, the major and minor grooves of the double-stranded DNA, as well as the 3′ single-stranded tail of the fork-

junction nontemplate strand, were readily apparent in the electron density maps (Figs. 1B and S1B). Moreover, the protein/solvent boundary was clearly discernible, allowing the *Taq* core RNAP structure (*7*, *8*) to be placed manually into the density. This rough placement of the core RNAP, combined with a close examination of the map, revealed an essentially one-to-one correspondence between rodlike densities in the map and α-helices in the protein (Figs. 1B and S1B). The core RNAP structure was divided into five "domains" [table s1 of (*1*)] that were fit manually as rigid bodies into the density, mainly by aligning α-helices with the rodlike densities. Substantial conformational changes in the core RNAP structure were required to fit the density map (*1*). After rigid body refinement of the core RNAP domains, the three structural domains of σ^A (*9*), as well as the DNA model, were readily placed into the map. Even though individual base pairs were not visible in the map, the alignment of the DNA along the double-helical axis (with respect to the protein) was fixed based on symmetry arguments (*10*). A second round of rigid body refinement, in which the σ domains and DNA were individually positioned along with the original core RNAP domains, resulted in the model that was used for the molecular replacement search to solve the 4 Å resolution *Taq* RNAP holoenzyme structure (*1*). Eventually, the final holoenzyme model derived at 4-Å resolution was combined with the DNA model, conformations of some domains were altered to fit the density map, and a final round of rigid body refinement was executed.

Overall structure. In RF, the fork-junction DNA lies across one holoenzyme face, completely outside the RNAP active site channel (Fig. 2). All of the sequence-specific RNAP contacts with the core promoter elements (−10, extended −10, and −35) are mediated by the σ subunit. The β′ NH_2-terminal Zn^{2+}-binding domain (β′ZBD) may contact the DNA backbone in the spacer between the extended −10 and −35 elements at −22 on the template strand and −27 on the nontemplate strand (Figs. 1B and 2).

The DNA is contacted by RNAP from only one side (Fig. 2), which explains footprinting data (*11*–*14*). The B-form DNA is relatively straight from −41 to −26 and −24 to −17. Centered at −25, the DNA bends 8° toward the major groove facing the RNAP β′ZBD (Fig. 2B). At −16, the DNA makes a sharp 37° bend toward the RNAP.

Conformational changes. A comparison of holoenzyme within RF to holoenzyme alone (*1*) reveals conformational changes in two mobile modules (Fig. 3) [table s1 of (*1*)]. In RF, the RNAP clamp domain (magenta, in Fig. 3), along with σ_2 bound to it, rotates in toward the RNAP channel by 3°, closing the

RNAP channel even further by about 3 Å relative to the holoenzyme alone. The β flap (blue, in Fig. 3), along with σ_4 bound to it, rotates 4°, resulting in the movement of the σ_4 –35 element recognition helix (9) by about 6 Å in the downstream direction (Fig. 3).

Holoenzyme-promoter interactions. At 6.5 Å resolution, protein-DNA contacts were analyzed in terms of the vicinity of amino acid side chains to the DNA and whether alterations of rotamers could bring the side chains near the DNA. Within σ conserved regions 2.2 to 3.0, four distinct determinants have been implicated in direct interactions with promoter DNA. In region 3.0, two residues (His[278] and Glu[281] of *Taq* σ^A, corresponding to His[455] and Glu[458] of *Escherichia coli* σ^{70}) are involved in recognition of the extended –10 element (15). Glu[281] was critical for recognition of the extended –10 sequence, whereas His[278] appeared to play a nonspecific DNA binding role. Both residues (red, in Fig. 4) are exposed on the surface of the σ region 3.0 α-helix, facing the major groove of the extended –10 DNA. Glu[281] may be within reach of the nontemplate strand T at –13. His[278] appears to be too far away from the extended –10 bases but could interact with the phosphate backbone of the nontemplate strand at positions –17/–18.

Region 2.4 of σ contains allele-specific suppressors of promoter mutations in the –10 element, implicating these residues in base-specific interactions with the –10 element [reviewed in (16)]. In group 1 σs, Gln[260] and Asn[263] (corresponding to Gln[437] and Thr[440] in *E. coli* σ^{70}) suppress promoter mutations at –12 within the –10 element. In the structure, both residues are exposed, facing the major groove of the DNA near –12. Gln[260] is easily within reach of the –12 bases and could interact with the nontemplate strand T or the template strand A (Fig. 4). An amino acid at position 263 (whether Asn or Thr), is out of reach of the –12 bases, but small adjustments of the structure might bring it within reach of the nontemplate strand T.

In σ region 2.3, highly conserved aromatic residues (Phe[248], Tyr[253], and Trp[256], corresponding to Tyr[425], Tyr[430], and Trp[433] of *E. coli* σ^{70}) play a role in promoter melting,

at least partly through sequence-specific binding of the nontemplate strand of the melted –10 element in the open complex (17–19). These residues appear ideally positioned to interact with unpaired bases of the single-stranded tail of the nontemplate strand DNA, which crosses the surface-exposed aromatic residues (Fig. 4). Phe[248] is closest to bases at the –8/–9 positions, whereas Tyr[253] is closest to bases at –9/–10 (Fig. 4). Most intriguing is Trp[256], which is positioned to stack on the exposed face of the –12 base pair, forming
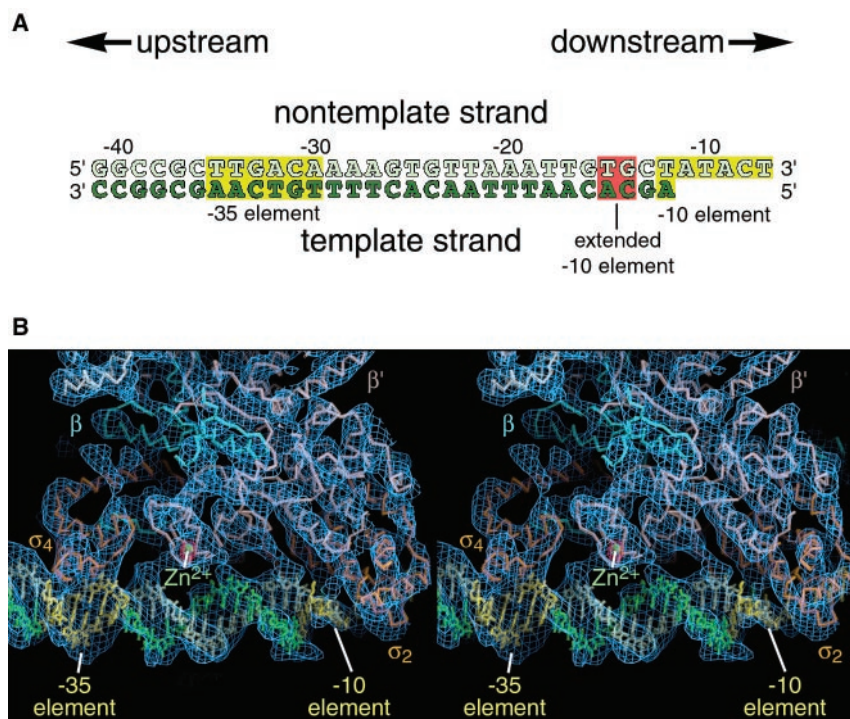

Fig. 1. Fork-junction DNA and electron density map. (**A**) Synthetic DNA oligonucleotides used for complex formation and crystallization. The numbers above denote the DNA position with respect to the transcription start site at +1. Downstream corresponds to the direction of RNAP movement during transcription. Mutations in the bottom DNA strand cause corresponding mutations in the RNA transcript, defining it as the template (versus the nontemplate) strand. The DNA sequence is derived from the full con promoter (4), with –35 and –10 elements (shaded yellow and labeled) as well as an extended –10 element (shaded red and labeled). (**B**) Stereo view of the *Taq* RNAP holoenzyme/fork-junction DNA complex. The α-carbon backbone of ω is colored white, β cyan, β′ pink, and σ orange (the α subunits are not visible). The DNA template strand is colored dark green, and the nontemplate strand is light green, except for the –35 and –10 elements, which are colored yellow. The visible structural domains of σ (σ_2 and σ_4) (1, 9) are labeled. The direction of transcription (downstream) is to the right. The experimental electron density map, calculated using observed amplitude (F_o) coefficients, is shown (blue net, contoured at 1.5σ), and was computed using multiple isomorphous replacement phases (Table 1), followed by density modification. The view is sliced at a level just in front of the DNA to reveal the β′ NH_2-terminal Zn^{2+}-binding domain and the associated Zn^{2+} (labeled, shown as a green sphere). Shown in red is a difference Fourier map, calculated using ($|F_o^{EMTS} - F_o^{native}|$) coefficients (Table 1), revealing the Hg-binding site that was used to locate the Zn^{2+}-site.

Table 1. Crystallographic analysis.

Data set*	Wavelength (Å)	Resolution (Å)	Redundancy (total/last shell)	Completeness (%) (total/last shell)	$I/\sigma(I)$ (total/last shell)	R_{sym}[†] (%) (total/last shell)	No. of sites	Phasing power[‡]
				Diffraction data				
Native	0.94799	80–6.5	5.7/2.4§	90.9/78.5	11.5/2.4	8.5/24.1	–	–
W_{12}‖	0.94799	80–6.5	4.0	89.3/73.5	7.2/1.5	11.6/37.9	3	1.29/1.67
Ta_6Br_{14}	0.94799	80–6.5	4.2	73.5/29.8	10.4/1.6	11.3/34.8	2	0.46/0.59
EMTS	0.94799	80–6.5	6.5	89.7/78.8	9.5/2.8	10.9/29.2	2	0.34/0.54
$PbCl_2$	0.94799	60–7.0	2.9	87.6/78.9	11.1/1.7	9.6/38.9	1	0.23/0.34
Figure of merit‡ (80 to 6.5 Å resolution):			0.374					

*Data sets were collected at CHESS-F1, Ithaca, NY.   †$R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I is observed intensity and $\langle I \rangle$ is average intensity obtained from multiple observations of symmetry related reflections.   ‡Phasing power and figure of merit as calculated by MLPHARE (39).   §Last shell: 6.73 to 6.5 Å.   ‖$K_5HCoW_{12}O_{40}$ (40).

the upstream edge of the transcription bubble (Fig. 4). Trp[256] may also be able to interact with the exposed base at the –11 position.

Universally conserved basic residues in regions 2.2 and 2.3 (Arg[237] and Lys[241], corresponding to Arg[414] and Lys[418] of *E. coli* σ[70]) are critical for DNA binding, probably in a non–sequence specific manner (*20*). Both positively charged residues are positioned to interact with the negatively charged DNA backbone of the nontemplate strand at the –13/–14 positions (Arg[237]) or at –15 (Lys[241]) (Fig. 4).

The 2.4 Å resolution structure of *Taq* σ$^A_4$ complexed with –35 element DNA (*9*) confirmed earlier genetic studies (*16*) indicating that residues of the recognition helix in the σ$_4$ helix-turn-helix motif bind the –35 element se-quence. In that structure, the DNA is bent 36° around the recognition helix, consistent with footprinting data (*21*). In contrast, in the RF complex, the DNA from –41 to –26, which includes the –35 element, is straight (Fig. 2B). In addition, the σ$_4$ recognition helix is shifted upstream about 6 Å, so that the sequence-spe-cific interactions with the –35 element (*9*) could not occur. There are two possible explanations for this discrepancy between the high-resolu-tion σ$_4$-DNA structure and the RF structure: (i) The crystals of the RF complex may have cap-tured a state, likely late in the series of steps involved in open-complex formation, in which σ$_4$ interactions with the –35 element DNA are no longer sequence specific. However, the available evidence from footprinting and crosslinking studies indicates that the specific σ$_4$/–35 element interactions are the first RNAP/promoter interactions to be established and that these persist throughout the process of RP$_o$ formation (*12*, *14*, *22–24*), making this expla-nation seem unlikely. (ii) Our preferred expla-nation is that the σ$_4$/DNA interactions observed in the RF complex are distorted by crystal packing. The upstream end of the fork-junction DNA packs against the upstream end of a crys-tallographically related DNA molecule, form-ing a pseudo-continuous double helix (*10*). This may not be compatible with a bend in the DNA near the upstream end. We suggest that crystal packing forces dictate that the upstream DNA be straight, and these forces are of sufficient strength to distort the sequence-specific σ$_4$/–35 element DNA interactions.

Model of the complete open complex. The RF structure reveals the disposition of the double-stranded promoter DNA from –41 to –12 and that of the nontemplate strand from –11 to –7 (Fig. 1A). However, the –35 element DNA and upstream may be distorted by crystal packing, and additional upstream and downstream DNA that interacts with ho-loenzyme is absent. Therefore, we construct-ed a model of RP$_o$ (Fig. 5), containing RNAP holoenzyme from the RF structure, as well as both strands of the DNA from –60 to +25.

The RF structure, combined with footprint-ing data (*12*, *14*, *23*, *24*), also suggests the nature of RP$_c$. In RP$_c$, where there is no detect-able strand separation of the DNA, the holoen-zyme protects promoter DNA from DNase I and hydroxyl radical cleavage with an approx-imately 10 base-pair periodic pattern from about –54 to –6. However, there is no protec-tion downstream of –6 by RP$_c$, in contrast to RP$_o$, which protects both strands of the down-stream DNA to about +20 (consistent with the enclosure of this DNA within the RNAP main channel) (Fig. 5). The RP$_c$ footprinting data indicate that the RNAP engages the promoter DNA upstream of the –10 element from one face of the DNA in a very similar manner as in RP$_o$. However, in RP$_c$, the DNA downstream of the –10 element is exposed, indicating that it
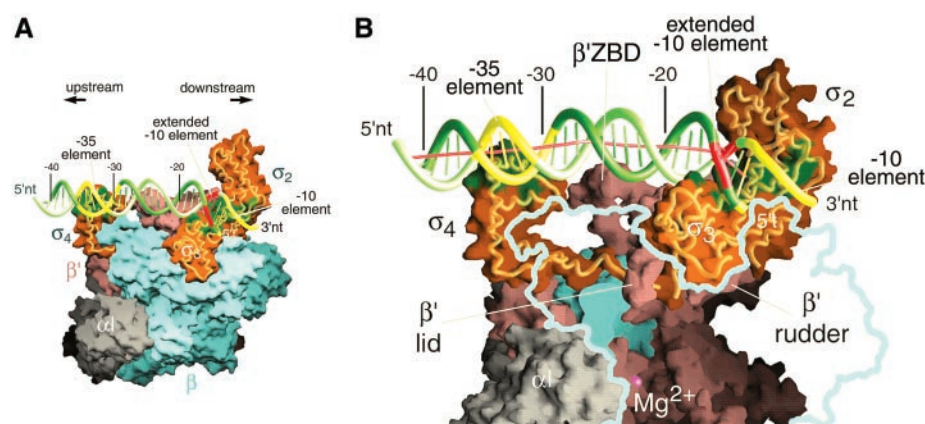

Fig. 2. *Taq* RNAP holoenzyme/fork-junction DNA structure. Views of the holoenzyme/fork-junction DNA complex. The RNAP holoenzyme is shown as a molecular surface, color coded as follows: αI, αII, ω, gray; β, cyan; β′, pink; and σ, orange. The molecular surface of σ is rendered partially transparent, revealing the α-carbon backbone worm (bright orange) inside. Protein surfaces in contact with the DNA (<4 Å) are colored green. These occur exclusively on σ. The DNA phosphate backbones are shown as worms, with the template strand (t) dark green, the nontemplate strand (nt) light green, except the –35 and –10 elements are yellow, and the extended –10 element is red. (**A**) Overall view of the complex. The view is similar to that of fig. 1A of (*1*), except that it is rotated about 45° conterclockwise around the *z* axis (perpendicular to the page) such that the DNA double-helical axis is roughly horizontal. The downstream and upstream directions are indicated. A numbering scale for the DNA position with respect to the transcription start site (at +1) is shown above the DNA. (**B**) Magnified view showing only a part of the complex, similar to the view of (A), with the downstream direction to the right. A numbering scale for the DNA position with respect to the transcription start site (at +1) is shown above the DNA. A pink line denotes the direction of the DNA double-helical axis (*38*), with a small kink at about –25, and a sharp bend at about –16. Obscuring parts of the β subunit in front have been removed (the outline of β is shown as a cyan line), revealing the structural features inside the main RNAP channel. The β′ZBD, β′ lid, and β′ rudder are labeled. The active site Mg^{2+} is shown as a magenta sphere.

Fig. 3. Conformational changes. The superimposed α-carbon backbones of the *Taq* RNAP ho-loenzyme alone (*1*) and the holoenzyme within the fork-junc-tion DNA complex are shown as worms (view the same as Fig. 2A). The structure of holoen-zyme alone is colored gray (core RNAP) and black (σ). The two modules that move in the ho-loenzyme-DNA complex as com-pared with the holoenzyme alone are colored as follows: clamp + σ$_2$, magenta and or-ange (respectively); β flap + σ$_4$, blue and orange (respectively). The phosphate backbones of the DNA in the holoenzyme/DNA complex are shown as ribbons and colored green (template strand, t) and light green (non-template strand, nt). The downstream direction is indicated. The movements of the mobile modules from the holoenzyme structure to their positions in the holoenzyme-DNA complex are indicated by the arrows.
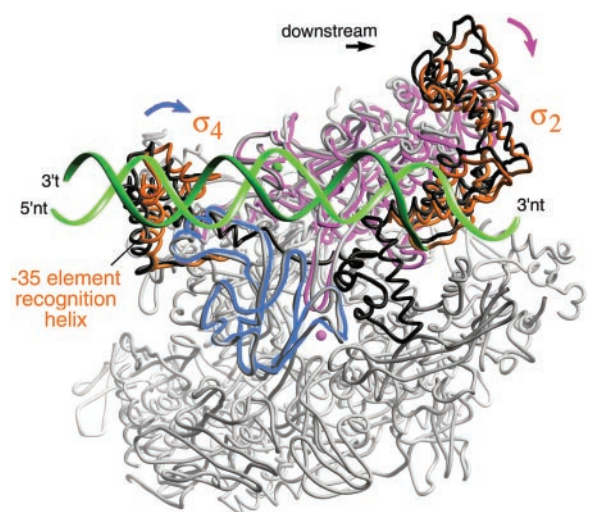
is not engaged with the RNAP. The simplest interpretation is that, in RP_c, the downstream DNA extends as a double helix straight past RNAP (Fig. 5A).

In addition to the −10, extended −10, and −35 elements, some promoters contain an UP element (an upstream, A/T-rich sequence, roughly −40 to −60) that is bound by the α-subunit C-terminal domains (αCTDs) in the DNA minor groove [reviewed in (25)]. The αCTD is an 80-residue, globular domain (26) connected to the α NH_2-terminal domain by a flexible, 14-residue linker (27). In addition to stimulating transcription up to several hundredfold by binding to the UP element, the αCTDs serve as targets for a wide array of transcription activators [reviewed in (28)]. In the bacterial RNAP structures [core RNAP (7, 8), holoenzyme (1), and the RF structure], the αCTDs and linkers are disordered, but their possible disposition is schematically illustrated in Fig. 5, along with the sites bound by the αCTDs in an UP element–containing promoter (29).

In the RP_o model, the upstream DNA bends around the RNAP, primarily due to kinks at about −35 (36°) and −25 (8°). At −16, the DNA makes another sharp 37° bend toward the RNAP. The two DNA strands separate and take different paths, beginning at −11, the upstream edge of the transcription bubble, which extends 15 nucleotides downstream.

The single-stranded, nontemplate DNA of the −10 element (−11 to −7) crosses σ_2, where it can interact with the exposed aromatic residues of σ region 2.3 (Fig. 4). The strand from −2 to +4 is held in a groove between two lobes of β (β1 and β2) (Fig. 5A).

The single-stranded template strand, which must enter the RNAP active site channel to base pair with the initiating nucleotide substrates bound in the i and i+1 sites (Fig. 5), is diverted through a tunnel, completely enclosed on all sides by protein (Fig. 6). The tunnel is formed by parts of σ_2, σ_3, β1, the β′ lid, and the β′ rudder (Figs. 5B and 6). Universally conserved

Fig. 4. Holoenzyme contacts with downstream promoter elements. At the lower right, a view of the holoenzyme/fork-junction DNA complex is shown [view the same as Fig. 1A of (1)]. The boxed area denotes the portion magnified in the main part of the image. The holoenzyme is shown as a molecular surface (color coding: αI, αII, ω, gray; β, cyan; β′, pink; and σ, orange). The template DNA (t) is green, the nontemplate DNA (nt) is light green, except for the −35 and −10 elements, which are yellow, and the extended −10 element is red. In the single-stranded part of the nontemplate strand (−11 to −7), the DNA bases have been removed because their positions were not determined at the resolution of this analysis. Residues of σ, determined from genetic studies to be important for downstream promoter binding, are labeled and color coded as follows: extended −10 element recognition, red (15); −10 element recognition, green (16); −10 element melting/nontemplate strand binding, yellow (17–19); and universally conserved basic residues important for DNA binding, blue (20). The downstream direction is toward the lower right.
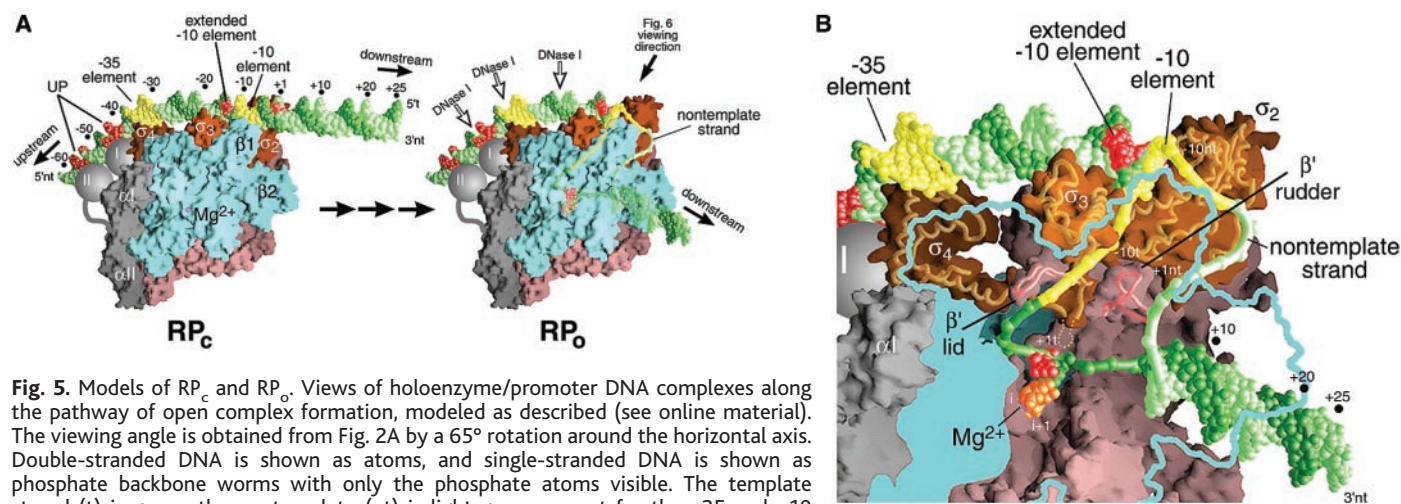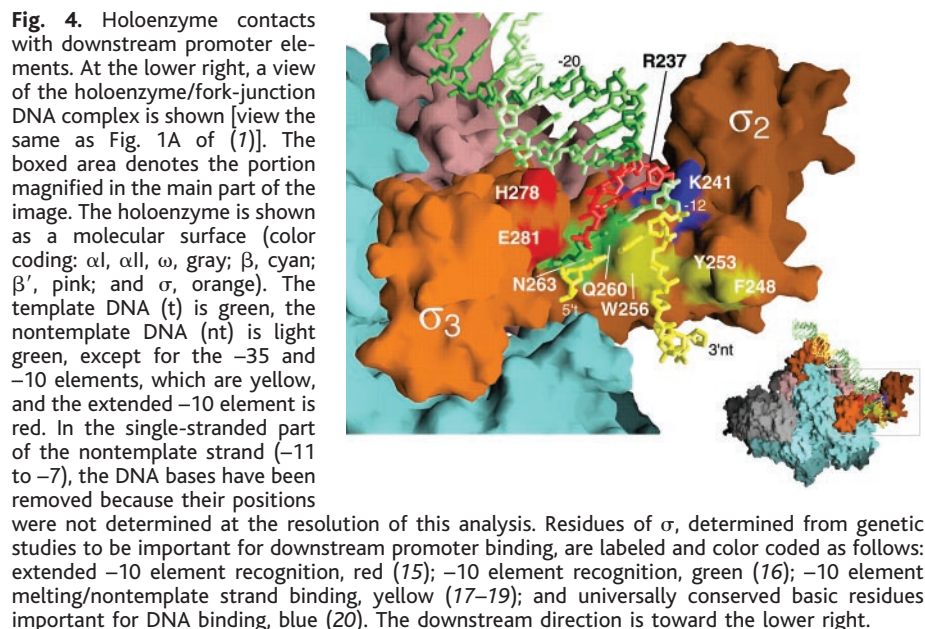
Fig. 5. Models of RP_c and RP_o. Views of holoenzyme/promoter DNA complexes along the pathway of open complex formation, modeled as described (see online material). The viewing angle is obtained from Fig. 2A by a 65° rotation around the horizontal axis. Double-stranded DNA is shown as atoms, and single-stranded DNA is shown as phosphate backbone worms with only the phosphate atoms visible. The template strand (t) is green, the nontemplate (nt) is light green, except for the −35 and −10 elements, which are yellow; and the UP elements, extended −10 element, and transcription start site on the template strand (+1) are red. The direction of transcription (downstream) is to the right. RNAP holoenzyme is shown as a molecular surface, color coded as follows: αI, αII, ω, gray; β, cyan; β′, pink; and σ, orange. The possible disposition of the αCTDs (drawn as gray spheres, labeled "I" and "II") on the UP elements (29) is illustrated. (**A**) Models of RP_c (left) and the final RP_o (right). The arrows in between denote that several intermediate steps exist along the pathway between these two states (2). The β subunit is rendered partially transparent to reveal the RNAP active site Mg^{2+} (magenta sphere) inside the main channel and the transcription bubble and downstream DNA enclosed inside the channel in RP_o. In RP_c, a numbering scale for the DNA position (−60 to +25) with respect to the transcription start site (+1) is shown above the DNA. The labeled domains of β (β1 and β2) are defined in table s1 of (1). In RP_o, RNA occupying the i and i+1 sites is shown as orange atoms. Sites of DNase I hypersensitivity in footprints of open complexes are denoted by open arrows in exposed minor grooves at approximately −45, −35, and −25 (21). The direction of view for Fig. 6 is also denoted. (**B**) Magnified view of RP_o, showing the details of the core promoter interactions, transcription bubble, and downstream DNA. Obscuring portions of the β subunit in front have been removed (the outline of β is shown as a cyan line) to reveal the structural elements inside the main RNAP channel. The numbering of downstream DNA positions (with respect to the transcription start site at +1) is shown. RNA occupying the i and i+1 sites is shown as orange atoms. The molecular surfaces of the entire σ subunit, as well as of the β′ lid and β′ rudder are rendered transparent, revealing the α-carbon backbone worms (bright orange and pink, respectively) inside. A 9-residue disordered segment of σ (residues 337 to 345) is shown as a dotted orange line (1). The template strand DNA within the transcription bubble is directed through a protein tunnel framed by σ_2 and the σ_3-σ_4 linker underneath, an α-helix of σ_3 and the β′ lid on one side, and σ_2 and the β′ rudder on the other side, and a domain of β (β1) in front, closest to the viewer, but seen only in outline (Fig. 6).

basic amino acids of σ regions 2.4 and 3.0 (Arg[259], Lys[285], Arg[288], and Arg[291], corresponding to Arg[436], Lys[462], Arg[465], and Arg[468] of *E. coli* σ[70]) are exposed at the entrance of the tunnel, contributing to the electrostatic potential that may direct the template strand into the tunnel (Fig. 6). The path of the template strand then takes it past the σ_2–σ_3 linker and elements of the β subunit that make up the back wall of the RNAP active site channel, until it passes near the active site (Fig. 5B). The downstream, double-stranded DNA from +5 to roughly +12 is clamped inside another protein tunnel between the β and β′ subunits (Figs. 5B and 6), consistent with footprinting (*11–14*) as well as functional studies demonstrating the importance of this downstream double-stranded DNA to the stability of the complex (*30*).

A survey of DNase I footprinting data of binary complexes of *E. coli* RNAP holoenzyme with 33 different promoters revealed a nonran-dom pattern of sites that became hypersensitive to nuclease cleavage after RNAP binding (*21*). The DNase I hypersensitive sites are [(nontem-plate/template strands); (−45/−47), ([−33, −34]/−37), and (−25/[−25,−26]). In the model, these sites are centered at three successive exposed minor grooves at about −45, −35, and −25 (Fig. 5A). At −35 and −25, the DNA is kinked (about 36° and 8°, respectively) toward the major groove facing the RNAP, thereby widening the exposed minor groove. Crystal structures of DNase I bound to DNA reveals that the nuclease attacks by binding in the minor groove, widening it by about 3 Å, and bending the DNA about 20° toward the major groove, away from the enzyme (*31*). Thus, kinking of the DNA in RP$_o$ at −35 and −25 by the RNAP in this manner would facilitate DNase I attack, explaining the tendency for DNase I hypersensitivity at these sites. The hypersensitive site around −45 suggests that the DNA contains an additional kink induced by the binding of αCTDI that would further bend the DNA around the RNAP (Fig. 5), but this kink has not been modeled because of the lack of structural information on the αCTD/DNA complex.

Conclusions. The structure of the *Taq* RNAP holoenzyme/fork-junction DNA complex reveals the disposition of the upstream promoter DNA and the nature of interactions forming the upstream edge of the transcription bubble. All of the sequence-specific contacts with core promoter elements are mediated by conserved regions of the σ subunit. A Trp residue, universally conserved among group 1 σ's (*32, 33*), is ideally positioned to stack on the exposed downstream face of the base pair at −12, forming the upstream edge of the transcription bubble. Universally conserved basic residues in σ regions 2 and 3 provide critical contacts with the DNA phosphate backbone and play a role in directing the melted template strand of the DNA through a tunnel into the RNAP active site. The structure also provides the basis for models of the initial closed and open complexes that are consistent with most of the available biochemical data. In these complexes, the RNAP introduces a series of discrete kinks in the upstream DNA, bending the DNA around the RNAP to increase the available binding interface.

The structure explains how holoenzyme recognizes promoters containing variably spaced −10 and −35 elements. First, plasticity that appears to be inherent in holoenzyme allows repositioning of the β flap and the bound σ_4 with respect to the DNA by at least 6 Å (Fig. 3). Second, the RNAP can kink the intervening DNA to correctly position the −10 and −35 elements with respect to each other (Fig. 2B). A close examination of data available in the literature reveals a strong correlation between −10/−35 spacer length and DNase I hypersensitivity around −25/−26 (table S1). Promoters with a 16 base-pair spacer rarely show DNase I hypersensitivity in this region, whereas promoters with a spacer of 18 base pairs or more always show hypersensitivity.

In the holoenzyme of the fork-junction complex, the claws around the RNAP main channel are essentially completely closed. Elements of the protein on each claw, such as the β′ rudder and parts of the β1 domain, interact across the channel by poking through the middle of the transcription bubble, sealing the DNA strands apart (Fig. 5B). This arrangement, presumably maintained during elongation, dictates that DNA must rotate through the RNAP structure (or vice versa) during translocation, like a screw inside a nut, perfectly tracking the helical pitch of the DNA double helix. This has been observed experimentally (*34*).

In the model of the complete open complex, the DNA template strand is enclosed in a protein
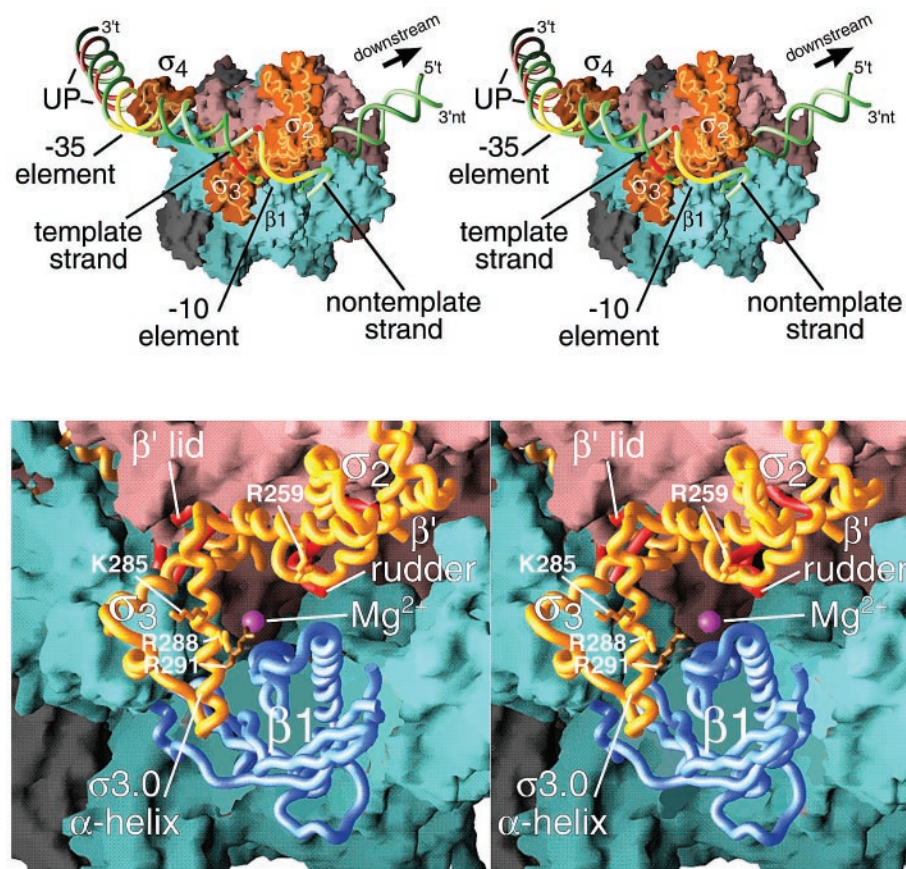

Fig. 6. Template strand tunnel. The RP$_o$ model (with αCTDs omitted) is viewed along the direction denoted in Fig. 5A, parallel with the template strand as it enters its protein tunnel. (**A**) Stereo view of the entire RP$_o$ model. The RNAP holoenzyme is shown as a molecular surface, color coded as follows: αI, αII, ω, gray; β, cyan; β′, pink; and σ, orange. The molecular surface of σ is rendered partially transparent, revealing the α-carbon backbone worm (bright orange) inside. The DNA phosphate backbones are shown as worms, with the template strand (t) dark green, the nontemplate strand (nt) light green, except for the −35 and −10 elements, which are colored yellow, and the UP elements and extended −10 element, which are red. (**B**) Magnified stereo view, centered on the template strand tunnel, with the DNA removed. The main protein elements framing the template strand tunnel are shown as α-carbon backbone worms without the associated molecular surface. These are the β1 domain (blue), the β′ lid and β′ rudder (red), and σ_2 and σ_3 (orange). The side chains of four universally conserved basic residues of σ that frame the entrance to the tunnel are also labeled and shown. At the far end of the tunnel lies the RNAP active site, denoted by the Mg[2+] ion (magenta sphere).

tunnel framed by universally conserved basic amino acids (Fig. 6B). Because open complex formation occurs without breaking covalent bonds in the DNA, the RNAP claws must open at some point during the process of open complex formation to allow the template strand to slip into its channel. Subsequent closure of the claws would then establish the tunnel. This requirement for prior states (intermediates) during the steps of open complex formation with different conformations of the enzyme, combined with the good match between footprinting data and the complete open complex model, leads us to suggest that the complex represented in this holoenzyme/fork-junction structure closely resembles the final RP_o.

The RF structure, and the models derived from it, raise key questions that are central to understanding transcription initiation. How is RP_o generated from RP_c (Fig. 5A)? How do transcription activators interact with the complex to enhance the rate of transcription initiation? The structures and models presented here provide a basis for designing more decisive experiments probing these questions and more.

References and Notes

1. K. S. Murakami, S. Masuda, S. A. Darst, *Science* **296**, 1280 (2002).
2. P. L. deHaseth, M. L. Zupancic, M. T. J. Record, *J. Bacteriol.* **180**, 3019 (1998).
3. Y. Guo, J. D. Gralla, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11655 (1998).
4. T. Gaal et al., *Mol. Microbiol.* **42**, 939 (2001).
5. S. Keilty, M. Rosenberg, *J. Biol. Chem.* **262**, 6389 (1987).
6. L. Tsujikawa, O. Tsodikov, P. de Haseth, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3493 (2002).
7. E. A. Campbell et al., *Cell* **104**, 901 (2001).
8. G. Zhang et al., *Cell* **98**, 811 (1999).
9. E. A. Campbell et al., *Mol. Cell* **9**, 527 (2002).
10. In the crystal, two DNA fork junctions pack against each other at their blunt, upstream ends (Fig. 1A), forming a pseudo-continuous double-helix related by a crystallographic twofold symmetry axis perpendicular to the DNA helical axis. Thus, the double-stranded upstream ends of the symmetry-related fork junctions were placed with an end-to-end spacing of 3.4 Å (corresponding to base-pair stacking in double-stranded B-form DNA), with the symmetry axis in the middle.
11. M. L. Craig, W.-C. Suh, M. T. J. Record, *Biochemistry* **34**, 15624 (1995).
12. J. Mecsas, D. W. Cowing, C. A. Gross, *J. Mol. Biol.* **220**, 585 (1991).
13. W. Metzger, P. Schickor, H. Heumann, *EMBO J.* **8**, 2745 (1989).
14. P. Schickor, W. Metzger, W. Wladyslaw, H. Lederer, H. Heumann, *EMBO J.* **9**, 2215 (1990).
15. K. A. Barne, J. A. Bown, S. J. W. Busby, S. D. Minchin, *EMBO J.* **16**, 4034 (1997).
16. C. A. Gross et al., *Cold Spring Harbor Symp. Quant. Biol.* **63**, 141 (1998).
17. S. A. Darst et al., in *Nucleic Acids & Molecular Biology*, F. Ekstein, D. M. J. Lilley, Eds. (Springer, London, 1997), vol. 11, pp. 27–40.
18. B. A. Young et al., *Cell* **105**, 935 (2001).
19. X. Huang, F. J. Lopez de Saro, J. D. Helmann, *Nucleic Acids Res.* **25**, 2603 (1997).
20. M. Tomsic et al., *J. Biol. Chem.* **276**, 31891 (2001).
21. O. N. Ozoline, M. A. Tsyganov, *Nucleic Acids Res.* **23**, 4533 (1995).
22. M. Buckle, I. K. Pemberton, M.-A. Jacquet, H. Buc, *J. Mol. Biol.* **285**, 955 (1999).
23. D. W. Cowing, J. Mecsas, M. T. J. Record, C. A. Gross, *J. Mol. Biol.* **210**, 521 (1989).
24. R. T. Kovacic, *J. Biol. Chem.* **262**, 13654 (1987).
25. R. L. Gourse, W. Ross, T. Gaal, *Mol. Microbiol.* **37**, 687 (2000).
26. Y. H. Jeon et al., *Science* **270**, 1495 (1995).
27. Y. H. Jeon, T. Yamazaki, T. Otomo, A. Ishihama, Y. Kyogoku, *J. Mol. Biol.* **267**, 953 (1997).
28. R. H. Ebright, S. Busby, *Curr. Opin. Genet. Devel.* **5**, 197 (1995).
29. W. Ross, A. Ernst, R. L. Gourse, *Genes Dev.* **15**, 491 (2001).
30. E. Nudler, E. Avetissova, V. Markovtsov, A. Goldfarb, *Science* **273**, 211 (1996).
31. S. A. Weston, A. Lahm, D. Suck, *J. Mol. Biol.* **226**, 1237 (1992).
32. T. M. Gruber, D. A. Bryant, *J. Bacteriol.* **179**, 1734 (1997).
33. M. Lonetto, M. Gribskov, C. A. Gross, *J. Bacteriol.* **174**, 3843 (1992).
34. Y. Harada et al., *Nature* **409**, 113- (2001).
35. T. A. Jones, J.-Y. Zou, S. Cowan, M. Kjeldgaard, *Acta Crystallogr. A* **47**, 110 (1991).
36. A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).
37. M. Carson, *J. Appl. Crystallogr.* **24**, 958 (1991).
38. R. Lavery, H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 63 (1988).
39. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend*, W. Wolf, P. R. Evans, A. G. W. Leslie, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1991) pp. 80–86.
40. N. Casan-Pastor, P. Gomez-Romero, G. B. Jameson, L. C. W. Baker, *J. Am. Chem. Soc.* **113**, 5685 (1991).
41. We thank M. Pope for the gift of the W_{12} cluster, G. Schneider for the gift of Ta_6Br_{14}, and T. Gaal and R. Gourse for providing the full con promoter sequence before publication; D. Thiel and the staff at the Cornell High Energy Synchrotron Source and M. Becker and L. Berman at NSLS X25 for support during data collection; K. Kinosita and R. Landick for important discussion; and D. Jeruzalmi, S. Nair, and H. Yamaguchi for invaluable advice. Figure 1B was made using the program O (*35*). Figures 2, 4, 5, and 6 were made using the program GRASP (*36*). Figure 3 was made using the program RIBBONS (*37*). Supported by a Norman and Rosita Winston Postdoctoral Fellowship and a Human Frontiers Sciences Program Postdoctoral Fellowship (K.S.M.); National Research Service Award number NIH GM20470 (E.A.C.); and, in part, by NIH grants GM53759 and GM61898 (S.A.D.)

Formation of a Matter-Wave Bright Soliton

L. Khaykovich,[1] F. Schreck,[1] G. Ferrari,[1,2] T. Bourdel,[1] J. Cubizolles,[1] L. D. Carr,[1] Y. Castin,[1] C. Salomon[1]*

We report the production of matter-wave solitons in an ultracold lithium-7 gas. The effective interaction between atoms in a Bose-Einstein condensate is tuned with a Feshbach resonance from repulsive to attractive before release in a one-dimensional optical waveguide. Propagation of the soliton without dispersion over a macroscopic distance of 1.1 millimeter is observed. A simple theoretical model explains the stability region of the soliton. These matter-wave solitons open possibilities for future applications in coherent atom optics, atom interferometry, and atom transport.

Solitons are localized waves that travel over long distances with neither attenuation nor change of shape, as their dispersion is compensated by nonlinear effects. Soliton research has been conducted in fields as diverse as particle physics, molecular biology, geology, oceanography, astrophysics, and nonlinear optics. Perhaps the most prominent application of solitons is in high-rate telecommunications with optical fibers (*1*).

We use a Bose-Einstein condensate (BEC) of a dilute atomic gas of lithium atoms as a macroscopic matter-wave to form a soliton. Nonlinearity is provided by binary atomic interactions leading to the mean-field potential $U(\vec{r}) = gn(\vec{r}) = 4\pi\hbar^2 an(\vec{r})/m$, where a is the scattering length, $n(\vec{r})$ the spatial density, and m the atomic mass. For $a < 0$ the effective interaction is attractive, and a trapped BEC is only stable for a number of atoms less than a critical number above which collapse occurs (*2–4*). When the BEC is confined in only two directions, matter-waves have dispersion in the free direction owing to their kinetic energy, $E_{kin} \propto k^2$, where k is the atomic wave vector. The balance between this dispersion and the attractive mean-field energy can lead to the formation of bright solitons as shown theoretically (*5–7*). Until now, only dark solitons have been observed in BECs with repulsive interactions ($a > 0$)

[1]Laboratoire Kastler Brossel, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France. [2]European Laboratory for Non-Linear Spectroscopy–Istituto Nazíonale per la Fisica della Materia, Largo E. Fermi 2, Firenze 50125, Italy.

*To whom correspondence should be addressed. E-mail: salomon@lkb.ens.fr