Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution

Tahir H. Tahirov*†, Dmitry Temiakov†‡, Michael Anikin‡, Vsevolod Patlan§, William T. McAllister‡, Dmitry G. Vassylyev§‖ & Shigeyuki Yokoyama§‖¶#

** High Throughput Factory, § Structurome Research Group, and ‖ Cellular Signaling Laboratory, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan*
‡ Morse Institute for Molecular Genetics, Department of Microbiology, SUNY Health Science Center, 450 Clarkson Avenue, Brooklyn, New York 11203, USA
¶ RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan
 # Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
† These authors contributed equally to this work

The single-subunit bacteriophage T7 RNA polymerase carries out the transcription cycle in an identical manner to that of bacterial and eukaryotic multisubunit enzymes. Here we report the crystal structure of a T7 RNA polymerase elongation complex, which shows that incorporation of an 8-base-pair RNA–DNA hybrid into the active site of the enzyme induces a marked rearrangement of the amino-terminal domain. This rearrangement involves alternative folding of about 130 residues and a marked reorientation (about 130° rotation) of a stable core subdomain, resulting in a structure that provides elements required for stable transcription elongation. A wide opening on the enzyme surface that is probably an RNA exit pathway is formed, and the RNA–DNA hybrid is completely buried in a newly formed, deep protein cavity. Binding of 10 base pairs of downstream DNA is stabilized mostly by long-distance electrostatic interactions. The structure implies plausible mechanisms for the various phases of the transcription cycle, and reveals important structural similarities with the multisubunit RNA polymerases.

Gene expression in all organisms requires messenger RNA synthesis by DNA-dependent RNA polymerase (RNAP). These enzymes can be divided into two classes: multisubunit (bacteria, archaea, eukaryotes) and single-subunit (some bacteriophages, mitochondria, chloroplasts) RNAPs. Although they share no apparent sequence or structural homology, the RNAPs of both classes carry out the basic steps of transcription in an identical manner[1]. To initiate RNA synthesis, the enzyme must bind to a specific promoter DNA sequence that lies upstream of the start site for transcription, separate (melt) the two strands of the DNA in the vicinity of the start site (forming a transcription bubble), and begin RNA synthesis using the coding strand of the downstream DNA as a template and a single ribonucleotide as a primer. During the early stages of transcription, contacts between the RNAP and the upstream promoter sequences are maintained while the active site translocates downstream, resulting in the formation of a short RNA–DNA hybrid and extension of the transcription bubble[2–5]. The T7 RNAP initiation complex (IC) is unstable and repeatedly releases short (3–8 nucleotides) abortive RNA products before it undergoes a transition to form a stable elongation complex (EC)[6]. The transition starts when the RNA–DNA hybrid reaches a length of 8–9 base pairs (bp) and results in promoter release, collapse of the melted promoter region, and displacement of the 5′ end of the nascent RNA[4,5,7]. During elongation the length of the RNA–DNA hybrid is maintained at 7–8 bp and the transcription bubble closes just after the RNA chain peels away from the DNA template[8–10].

The solution of several RNAP structures has resulted in a breakthrough in our understanding of the functional aspects of transcription[11–20]. The structures of T7 RNAP–nucleic acid complexes revealed that promoter binding involves three principal structural motifs of the RNAP: an (A + T)-rich recognition loop interacts with the −17 region; a specificity loop makes base-specific contacts around −9; and an intercalation loop involving Val 237 facilitates promoter melting and stabilizes the upstream edge of the separated transcription bubble between −5 and −4 (the start site for transcription is designated as +1). The structure of an early IC in which the first three bases of the template strand have been transcribed is essentially unchanged from that of the binary promoter complex, indicating that the initial stages of transcription may be achieved without major changes in enzyme structure. However, the structure of the IC did not allow space in the active site for an RNA–DNA hybrid greater than 3 bp, and therefore did not provide a plausible mechanism for further transcription progress. The crystal structure of the T7 RNAP elongation complex reported here reveals that incorporation of an 8-bp-long RNA–DNA hybrid into the active site results in unprecedented structural alterations of protein structure, thereby resolving the principal differences between previous biochemical and structural data[9,14,21]. Although the conformation of the carboxy-terminal portion of T7 RNAP (which includes the active site) resembles that of polymerase I-like DNA polymerases[11,22,23], the overall organization of the EC is remarkably similar to that of the multisubunit RNAPs, reflecting the common functions of these two classes of RNAPs.

Structure determination and overall structure

Elongation complexes were formed and crystallized as described[24,25]. The structure was refined at 2.9 Å resolution to a final $R = 23.5\%$ and $R_{free} = 28.4\%$ (Fig. 1a, b; see also Supplementary Table 1).

The RNAP in the EC has a shell-like, highly porous architecture (Fig. 1c–e). The downstream DNA is bound in a deep groove and enters through a wide passage to a cavity that contains an 8-bp RNA–DNA hybrid. The axis of the hybrid is nearly perpendicular to the entering DNA, as in the yeast RNA polymerase II elongation complex[17]. The structure contains partly accessible channels, which presumably correspond to binding sites for the template and non-template DNA strands, and prominent pores for entry of the substrate and exit of the RNA product. The positive charge covering almost the entire interior of the molecule extends through the pores and channels to the external surface. The channels and pores are features characteristic only to the EC.

Protein structure

During the formation of the EC, the N-terminal domain of the

RNAP (residues 2–266) undergoes remarkable structural reorganization as compared with the IC (Fig. 2). This involves a marked reorientation of a core subdomain (72–151, 206–257), which remains unaltered, and an alternative folding of approximately 130 residues. These changes result in the formation of three structural elements: an N-terminal extension (N-subdomain; residues 2–71), a central flap-like subdomain (152–205), and a C-terminal linker that connects the N-terminal domain to the C-terminal portion of the enzyme (C-linker; 258–266) (Fig. 2).

Relative to the IC (Protein Data Bank accession number 1QLN[14]; Fig. 2b, c), the core subdomain does not change its internal structure (root mean square deviation, r.m.s.d. = 0.7 Å over 524 main chain atoms) but moves as a rigid body (35 Å translation and 130° rotation) to its final position in the EC (Fig. 2a, c). This reorientation opens space in the active site to accommodate the expanding RNA–DNA hybrid. Of note, the core exhibits high intrinsic structural similarity (but lacks sequence homology) to a C-terminal region of T7 RNAP and to a lesser extent to the corresponding segment of pol I-like DNA polymerases (Taq DNAP; Protein Data Bank accession number 1TAQ[26]) (Fig. 3). This suggests that the core subdomain evolved from duplication of a conserved structural element.

In the IC (Fig. 2b, c), the N-subdomain does not interact with nucleic acids, and its fold consists of seven structural motifs: loop-helix-loop-helix-loop-helix-disordered region. In the EC (Fig. 2a, c), the N-subdomain adopts a more compact loop-helix-loop-helix-loop conformation that constitutes a portion of the binding site for the RNA–DNA hybrid. Despite the marked difference in folding, the N-subdomains in the IC and EC are located on the same side of the RNAP molecule, so that a short α-helical fragment (31–43) coincides in both structures. This suggests that the N-subdomain keeps a fixed orientation during the structural transition of the RNAP, and that all alterations in this domain result from alternative folding.

A long, unfolded loop (165–205) intervenes between the N- and C-terminal portions of the core in the IC (Fig. 2b, c). In the EC (Fig. 2a, c) this loop is folded into the flap subdomain, which consists of a helix-turn-helix (HTH) motif followed by a C-terminal loop. In the EC, the flap subdomain spans the interval between the upstream end of the RNA–DNA hybrid and downstream DNA.

Finally, whereas the short C-linker connecting the N-terminal domain to the C-terminal domain has a random coil conformation in the IC (Fig. 2b, c), it adopts an α-helical conformation in the EC (Fig. 2a, c) and extends the C-terminal α-helix of the core subdomain towards the end point of the N-terminal domain.

The C-terminal, DNA polymerase-like portion of the RNAP, which consists of three major domains designated as 'fingers', 'palm' and 'thumb'[11,12], remains largely unaltered in the EC as compared with the IC (Fig. 2a, b). The primary exception is the specificity loop (739–772)[13,27,28] (Fig. 2a, b). A substantial shortening of the loop hairpin (by about 13 Å) is probably the result

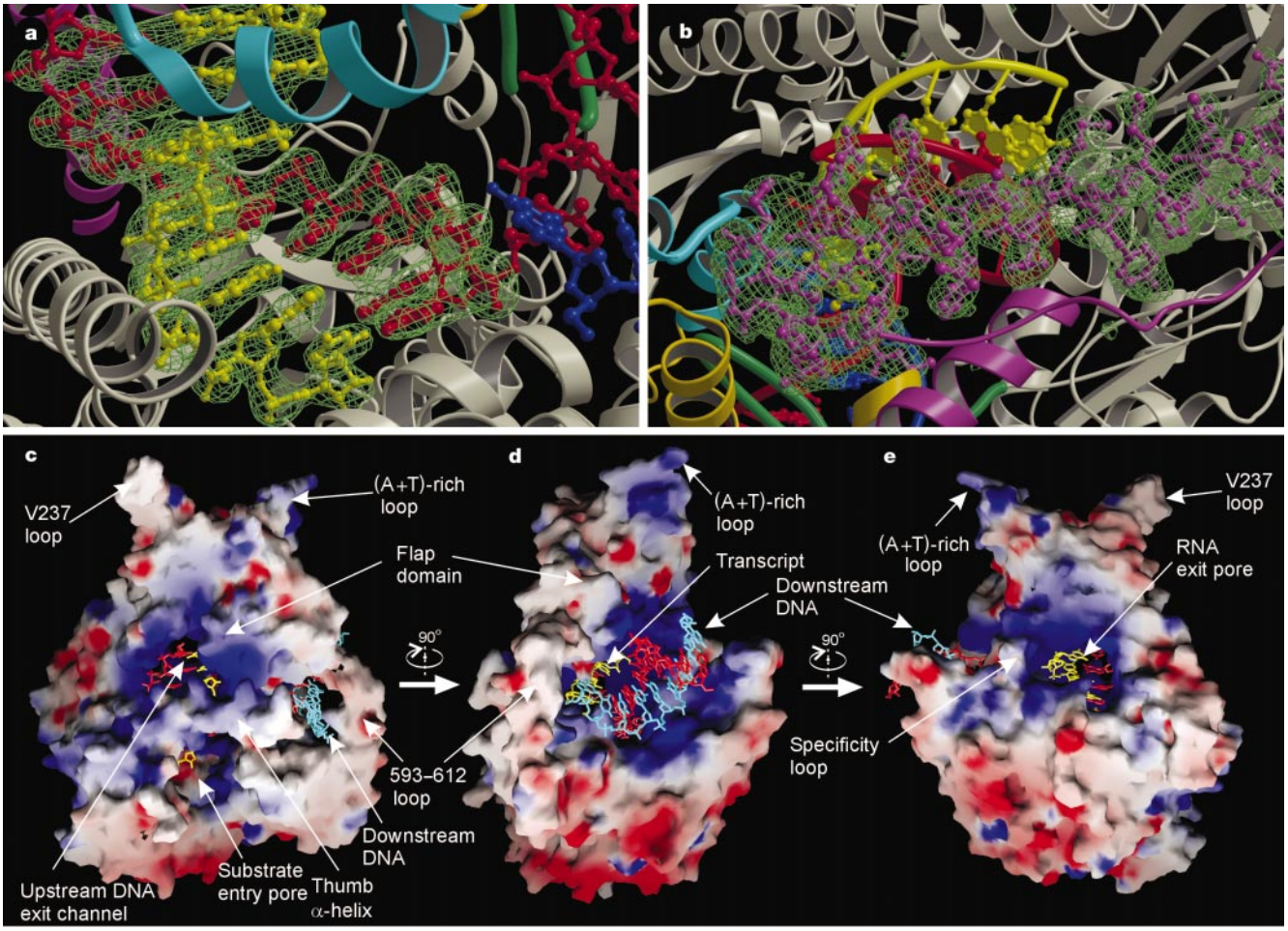

Figure 1 The T7 RNAP EC crystal structure. The RNA (light yellow), DNA template (red) and non-template (blue) strands are shown as a 'ball-and-stick' model. **a**, **b**, The $|F_o - F_c|$ omit electron density map (3.2σ contour level; green) produced for the RNA–DNA hybrid (**a**) and for the long N-terminal α-helix (residues 31–62, ball-and-stick model; magenta) (**b**) is superimposed on the atomic model. The rest of the protein is represented by a ribbon diagram. **c**–**e**, Three views of the EC. The protein surface is coloured according to the electrostatic potentials (positive, negative and neutral are dark blue, red and white, respectively).

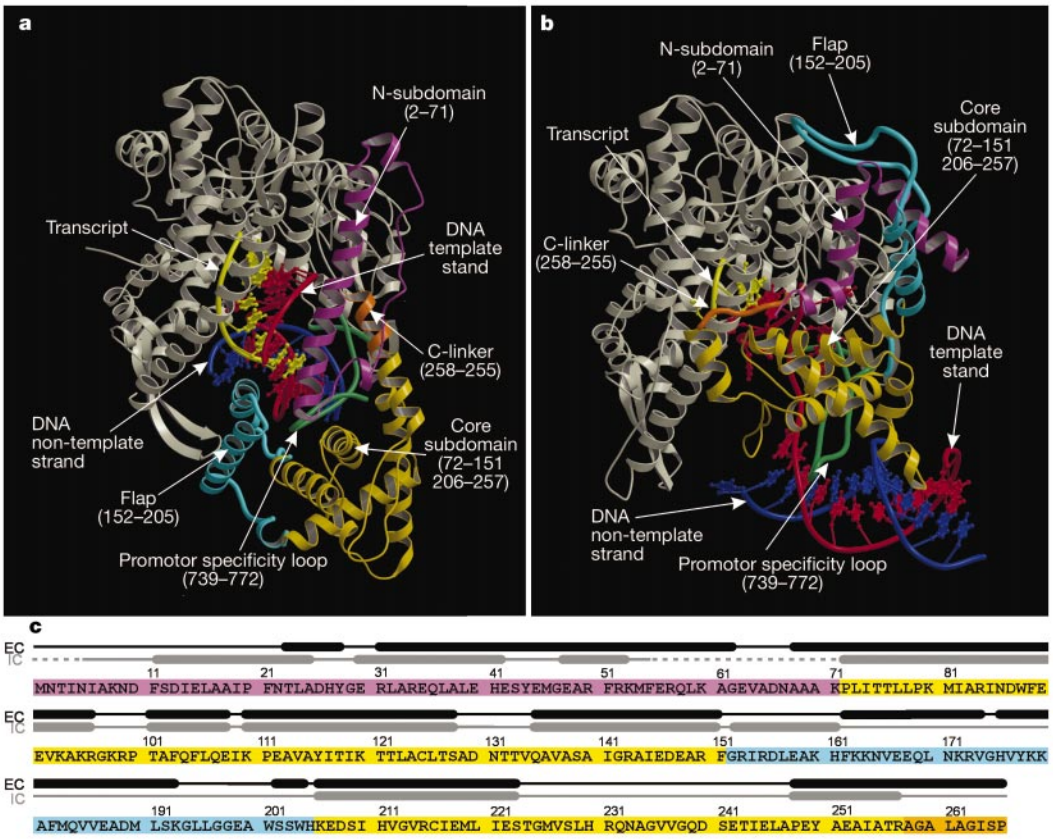

Figure 2 Comparison between the T7 RNAP EC and IC structures. **a**, **b**, The EC (**a**) and the early IC (**b**) ribbon diagrams are shown in the same orientation with respect to the unaltered C-terminal portion (white). The view in **a** corresponds to that of Fig. 1c, but rotated 180° around the horizontal axis. The nucleic acids are shown as ribbon (phosphate backbone) and ball-and-stick (bases) models. **c**, Amino acid sequence of the N-terminal domain; secondary structures for the EC and the early IC are shown above (cylinders, α-helices; thin lines, coils; dashed lines, disordered regions). The colour scheme is the same as in **a** and **b**.

of outlooping of the C-terminal strand in the vicinity of residues 757–767, as indicated by weak but visible electron density in this region. Furthermore, the tip of the loop is buried in a hydrophobic cavity formed by the flap and core subdomains, and the region of the loop that is used for base-specific promoter recognition in the IC is shifted. As a result of these changes, and reorientation of the core subdomain, the specificity loop becomes separated from the two elements in the core subdomain that were involved in promoter binding in the IC.

There are two other noticeable alterations in the C-terminal portion of the enzyme. First, the loop at the tip of the fingers domain (593–612) moves towards the downstream DNA to form part of the DNA-binding site (Fig. 1c, d). Second, the N-terminal end of the long α-helix in the thumb domain bends towards the fingers domain to complete the formation of the pore that leads to the active site (Fig. 1c).

Protein–nucleic acid interactions

The position of the downstream DNA is determined largely by long-range electrostatic interactions (Fig. 4a) with parts of the fingers

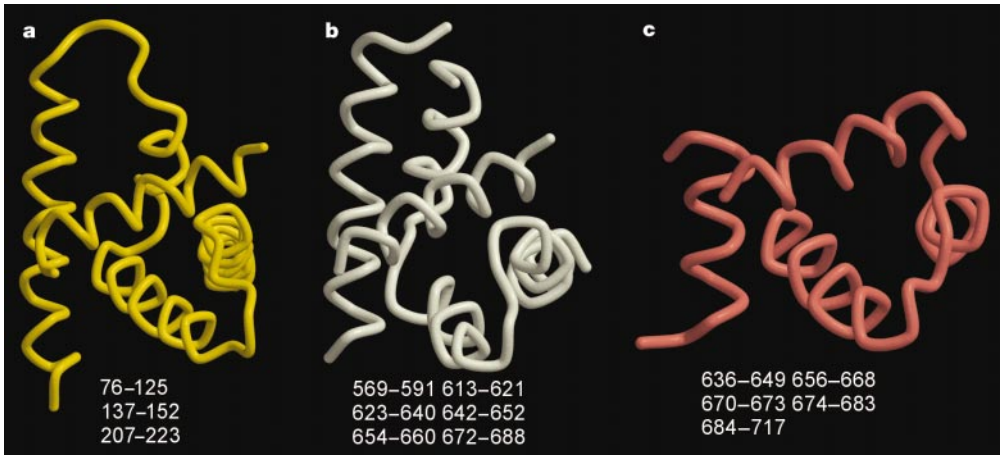

Figure 3 Structural homology of the T7 RNAP core subdomain. **a**–**c**, The T7 RNAP core (**a**), the C-terminal homologous region (**b**), and the homologous fragment of the Taq DNA polymerase (**c**) are shown in the same orientation. The numbers of residues used for the superimposition are shown at the bottom.

subdomain and with residues in the first turn of the flap subdomain. The DNA strands are separated just before the template strand enters the active site by interaction with the N-terminal part of a helix in the fingers subdomain (642–662), which we designate a 'downstream DNA zipper'. DNA melting involves stacking of the template base at +2 on the Phe 644 side chain of the zipper, which appears to cause the base at +1 to flip out, thereby directing the template strand towards the active site (Fig. 4b). (Nucleotide positions in the EC are designated relative to the active site at +1.) The first unpaired nucleotide of the non-template strand (at +1) is directed towards a deep groove formed by a loop (593–610) and a helix-turn-helix fragment (647–675) from the fingers subdomain, part of the flap subdomain, and the N-terminal portion of the long α-helix from the thumb subdomain (Fig. 1c, d). This groove extends to the template strand exit located between the second turn of the flap subdomain and the C-terminus of the long helix of the N-terminal extension (Fig. 1c).

The 8-bp RNA–DNA hybrid in the EC adopts an underwound A-form conformation of the double helix that is similar to the conformation of the RNA–DNA hybrid in the yeast RNA polymerase II EC (Fig. 1a; r.m.s.d. 0.6 Å)[17]. The hybrid-binding cavity complements this shape and has a surface that consists of alternating hydrophobic and positively charged residues, which may facilitate translocation (Fig. 4a). The first 3 bp at the downstream end of the hybrid interact with the protein in a similar fashion as in the T7 IC (Fig. 4a). However, the interactions of the 3 bp at the upstream end of the hybrid involve newly folded elements of the EC. Most notable are interactions of template strand nucleotides at positions −6, −7 and −8 with the C-terminal portion (50–60) of the long helix from the N-subdomain, and the RNA nucleotides at positions −6 and −7 with the flap subdomain (Fig. 4a).

The upstream boundary of the RNA–DNA hybrid cavity is defined by the C-terminal part of a loop-helix motif (62–71) that connects the long helix of the N-subdomain with the core subdomain. The Gln 58, Glu 63 and Asp 66 side chains in this loop project towards the plane of the base pair at the upstream end of the hybrid at a distance of about 7.3 Å (Fig. 4a, c), which may allow hybrid extension by 1 bp without serious protein alterations.

RNA exit and substrate entry pores

In the EC, a highly positive pore leads from the upstream end of the RNA–DNA hybrid to the surface, and probably provides an RNA exit pathway[29,30] (Fig. 1e). The pore is formed by part of the reorganized specificity loop, the newly formed C-linker, and part of the C-terminal portion of the enzyme (291–303). A final step in the transition from the IC to EC occurs when 10–14 nucleotides of RNA have been synthesized[31,32], and may result from interactions of the displaced RNA with the surface of the pore, thereby stabilizing the transcription bubble.

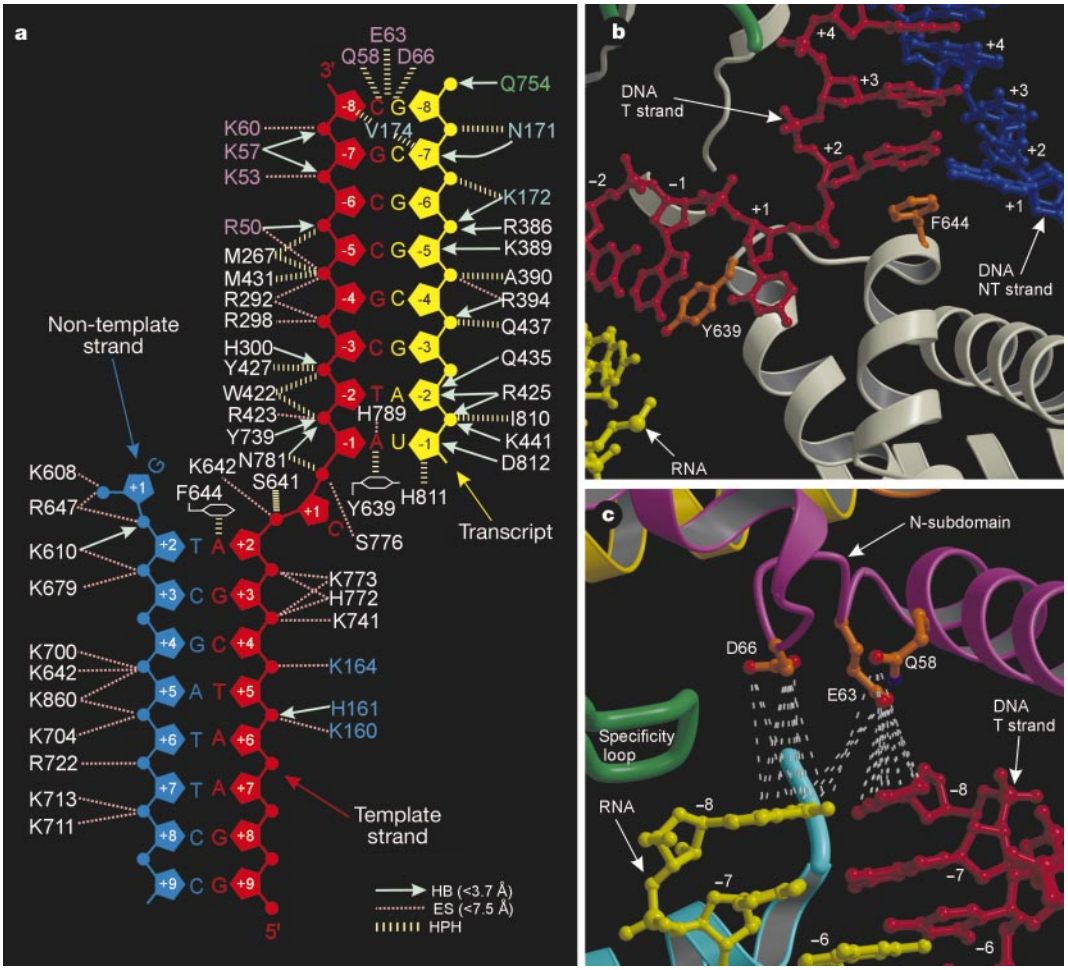

Figure 4 Protein–nucleic acid interactions. **a**, Schematic drawing of the protein–nucleic acid interface. HB, hydrogen bonds; ES, electrostatic interactions; HPH, hydrophobic interactions and closely approaching residues with the nucleotide indicated (including phosphate and sugar). **b**, Junction between downstream DNA and the RNA–DNA hybrid. **c**, Upstream end of the RNA–DNA hybrid. Colours are the same as in Fig. 2. The side chains of Tyr 639 and Phe 644 (**b**), and Gln 58, Glu 63 and Asp 66 (**c**) are shown in orange. Proximity of the N-subdomain residues to the upstream hybrid base pair (**c**) are indicated by dashed lines. T, template; NT, non-template.

Another pore provides access to the active site, and is probably the pathway for entering substrate (Fig. 1c). Although this opening is also present in the IC, it becomes more accessible in the EC as a result of a 12° bend of the thumb domain α-helix (372–409) (Fig. 1d). Substitutions of residues that line the surface of the pore result in marked effects on catalytic activity[33–36], which would be consistent with a role for this region in providing a path for entering substrate, or, owing to its proximity to the active site, on catalysis.

The presence of these pores in the EC provides a clear parallel to the multisubunit RNAPs. However, whereas these exit and entry pathways are short in the T7 EC, they are considerably longer (about 30 Å) in the multisubunit RNAPs[15,16]. This difference may account, in part, for the high rate of polymerization exhibited by T7 RNAP (200 nucleotides per s for T7 compared with 30 nucleotides per s for *Escherichia coli* RNAP)[1].

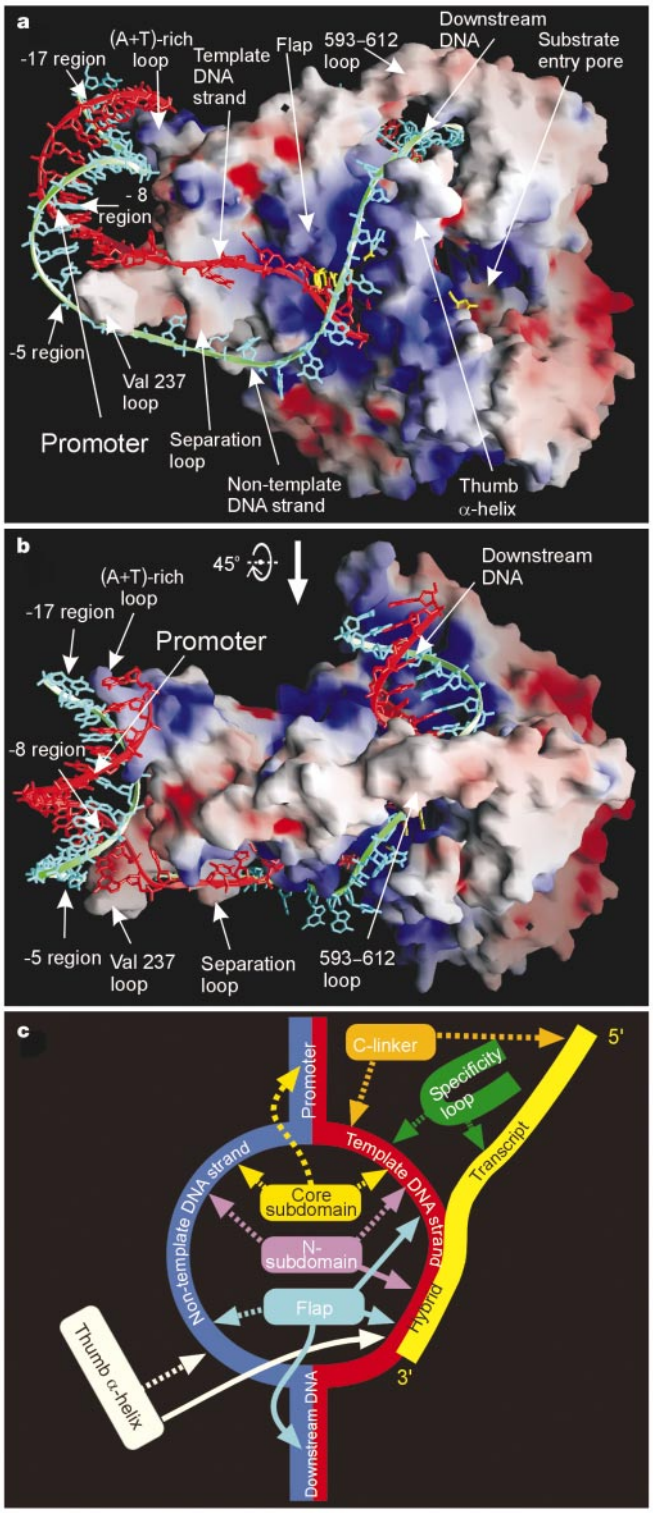

Figure 5 Model of the late IC. Two different views are shown in **a** and **b**. The protein is represented by the same electrostatic potential surface as in Fig. 1c–e. The nucleic acids are shown as a combination of ribbon phosphate backbone and ball-and-stick models with the RNA, and DNA template and non-template strands coloured in light yellow, red and light blue, respectively. **c**, Interaction of the protein domains that have principal roles in the structural organization of the late IC (dashed arrows) or the EC (solid arrows), with the components of the transcription bubble shown schematically. The colour scheme is the same as in Fig. 2.

The transition from an IC to an EC

Biochemical experiments have shown that during the late stages of initiation, just before promoter release, the length of the RNA–DNA hybrid is about 8 bp, as is observed in the EC[7]. It therefore seems likely that the conformation of the RNAP in the late IC resembles that of the EC. A number of questions arise concerning the transition from an early IC to a late IC. Does the transition occur in one step, or is it a multistep process? Furthermore, how are the upstream promoter contacts maintained while the hybrid grows to 8 bp?

Modelling of the RNA–DNA hybrid as it is extended from 3 bp in the early IC indicates that the first clash of the hybrid would occur with elements of the core domain at 4 bp, but that rotation of the core domain, together with specificity loop and promoter, could accommodate extension of the hybrid to 5 bp without substantial refolding (consistent with biochemical data[2,37]). After 5 bp, however, the structure of the EC shows that the hybrid is greatly stabilized by interactions with the refolded elements of the N-terminal domain (Fig. 4a), suggesting that reorganization starts at this point. This is consistent with the observation that mutations affecting the elements in the core domain that interact with the hybrid at this length inhibit extension beyond 5 nucleotides[38,39]. The (A + T)-rich recognition and the Val 237 intercalation loops are part of the unaltered core subdomain, and the promoter contacts mediated by these elements may be retained during the reorientation of this subdomain (assuming that the upstream DNA moves along with the core) and might persist up to 8–9 bp. On the other hand, modelling suggests that this reorganization cannot occur without loss of contacts between the specificity loop and the promoter (Fig. 2a). This is consistent with the results of ultraviolet laser crosslinking experiments, which show that promoter contacts at −7 and −8 are lost before the contact at −17 (ref. 5).

The pattern described above suggests that a simple superposition of the IC and EC core subdomains could provide a plausible model for continued binding of the upstream promoter region in the late IC (Fig. 5). The model shows that a transcription bubble of 9–10 nucleotides (which corresponds to a hybrid length of 5–6 bp) would be too short to connect the promoter in the late IC with the hybrid and downstream DNA (not shown). Thus, the rearrangement that is proposed to commence at 6 bp cannot occur as a single step, and is probably a multistep process in which gradual reorganization of the N-terminal domain is accompanied by incremental movement of its core. The 13–14 nucleotide length of the bubble in the late IC (which corresponds to a hybrid of 8–9 bp, and is in agreement with recent fluorescence experiments[7]) suggests a probable pathway for the template and non-template strands that connect the upstream region of the promoter with the hybrid and downstream DNA (Fig. 5). A 25 Å-long positively charged groove, formed by the thumb domain α-helix and the flap subdomain, is likely to form a binding site for approximately 6–8 nucleotides of the downstream non-template DNA strand[29,30] (Fig. 5). At the upstream edge of the transcription bubble there are two hydrophobic cavities that can accommodate 4 bases of template and 4–5 bases of non-template strand DNA (Fig. 5). The central wall separating these two cavities

(the separation loop) is formed by a portion of the core domain (128–133), whereas the remaining portions are formed by newly formed structural elements. If the movement of the core domain is fixed at this point, further extension of the RNA–DNA hybrid would result in promoter release as a consequence of the strain associated with extension, and/or as a result of compression and extrusion of the template and non-template strands from the hydrophobic channels. Biochemical data suggest that promoter release may precede the final collapse of the transcription bubble and displacement of the 5′ end of the RNA from the hybrid[7]. Interaction of the separation loop with the template and non-template strands just downstream of the promoter may prevent complete collapse of the bubble until later in the initiation process. In the model, the non-template base at −11, which would later re-anneal to the template strand to form the upstream edge of the bubble after promoter release, is on the protein surface in close proximity to the upstream edge of the RNA–DNA hybrid. This suggests that the final transition to the EC may occur without further major alterations of protein structure.

Structural organization of the EC

The EC structure, in combination with modelled promoter DNA and a plausible trace of the template and non-template strands, comprises a transcription intermediate that corresponds to the late IC. Although the modelled nucleic acids are not suitable for detailed analysis of the interactions, they allow a discussion of the role and significance of new structural elements that are acquired by the RNAP in the course of refolding.

The EC structure shows that the rearranged segments of the N-terminal domain (N- and flap-subdomains, and C-linker) are principal elements of the complex. These motifs, which do not have an essential functional role in the early IC, become multi-functional in the late IC and in the EC, and are extensively involved in interactions with various components of the transcription bubble (Fig. 5c).

The N terminus of the N-subdomain probably provides a pathway for the template and non-template DNA strands in both the late IC and the EC. The long α-helix of the N-subdomain interacts with the RNA–DNA hybrid in the EC and may also contribute to the formation of the upstream template and non-template DNA-binding sites in the late IC (Fig. 5c). Importantly, the C-terminal linker region of the N-subdomain is in close proximity to the upstream boundary of the hybrid in the EC, and extension of the hybrid by 1 bp would clash with the hydrophilic residues clustered at the tip of this loop (Fig. 4a, c). The loop may therefore act as a 'hybrid zipper' to facilitate strand separation of the hybrid. In bacterial RNAPs, the σ-subunit 'destabilizer' loop, which is also likely to interact with the upstream edge of the hybrid, contains two highly conserved acidic residues at the tip that may have essentially the same role[19].

In the EC structure, the flap subdomain interacts both with the upstream end of the hybrid and with the downstream DNA. In the late IC, it is probably involved in binding of the upstream portion of the template strand. The flap subdomain also forms part of a potential non-template binding site in both the late IC and the EC (Fig. 5c). Finally, in the EC, this subdomain is required to fix the position of the specificity loop (Fig. 2a), which is indispensable for the formation of the RNA exit channel. Thus, the flap subdomain seems to be of central importance to the integrity and function of the late IC and the EC. Consistent with this, mutations within this motif result in a failure to resolve the transcription bubble[31,40–42].

Although very short (8 residues), the C-linker also probably interacts with two distinct elements of the transcription bubble: with the upstream DNA template strand in the late IC and with the displaced RNA product in the RNA exit pore (Fig. 5c).

Other than the N-terminal domain, the principal element whose functional purpose is switched during the transition to an EC is the specificity loop. While losing its role in promoter recognition it forms a portion of the RNA exit channel and is probably involved in interactions with the displaced RNA product. The flexible out-looped segment of the loop, which is rich in hydrophilic residues, may intrude into the RNA exit pore. We speculate that this segment may be important for responding to pausing and termination signals. The tip of the specificity loop may also interact with the DNA template strand in the late IC (Fig. 5c).

Abortive cycling

Four structures of T7 RNA have been determined previously: free enzyme; RNAP bound to the transcription inhibitor T7 lysozyme; a binary RNAP–promoter complex; and an early IC[11–14]. All of these structures display very similar conformation of the enzyme (Fig. 2b), which implies that this conformation probably represents the most stable organization of the protein. In contrast, the conformation of RNAP in the EC (Fig. 2a) has never been observed before. The extensive protein–nucleic acid interface (about 2,500 Å^2) in the EC, which is lacking in the other structures, suggests that the protein conformation of the EC is greatly stabilized by these interactions. During the transition to an EC the exposed hydrophobic surface of the N-terminal domain increases by roughly 1,400 Å^2, which, in the absence of the nucleic acid–protein interface, may favour a return to the more stable IC conformation. Given the probable multistep nature of the transition from the early IC to the late IC, it seems probable that the intermediate enzyme conformations would be less stable than the one observed in the EC, because they lack the full extent of the final nucleic acid interface. In addition, modelling shows that the principal N-subdomain and C-linker α-helices can only be gradually folded during the course of the transition, and that the unfolded IC loop, comprising the flap domain in the EC, would lose its stabilizing interactions with the protein during the initial stage of the transformation. We speculate that competition between the nascent RNA–DNA hybrid for space in the active site (which induces the reorganization of the RNAP) against the tendency of the partially folded protein structures to fold back to the initial conformation is a major factor contributing to abortive cycling. A similar mechanism has been proposed for bacterial multisubunit enzymes, in which the growing RNA–DNA hybrid competes with initiation factor-σ resulting either in abortive RNA synthesis or displacement of σ to form a stable EC[18,19].

Discussion

On the basis of footprinting experiments, two principal models were proposed for the progress of transcription initiation: 'polymerase inchworming', in which the RNAP structure extends to include the expanding transcription bubble, and 'DNA scrunching', in which the RNAP structure remains unaltered while the bubble is accumulated in the active site. The interpretation of previous structural data favoured the scrunching model, in which the RNA–DNA hybrid length was limited to 3 bp and the strain associated with packing of the transcription bubble in the unaltered enzyme led to abortive initiation[14].

Although not excluding a role for scrunching during the early stages of transcription (when the hybrid is ≤4 bp), the T7 RNAP EC structure reported here provides strong evidence in favour of polymerase inchworming during the later stages of initiation, as the structure of the RNAP is expected to change progressively to accommodate the RNA–DNA hybrid as it grows to a final length of about 8–9 bp.

During the maturation of the EC, the elaboration of the components of the transcription bubble both induce and stabilize the reorganization of the protein structure. We suggest that a similar role for nucleic acids may also be expected for other multifunctional nucleic-acid-binding proteins, including the multisubunit RNAPs during their transition from an IC to an EC.

The overall organization of the T7 RNAP EC bears a marked resemblance to that of the multisubunit RNAPs. Moreover, the structural mechanism by which T7 RNAP achieves this final state is similar to the steps that are observed in bacterial RNAPs. In both cases the active site that accommodates the components of the mature transcription bubble is greatly reduced at the beginning stages of transcription, and the RNA exit channel is either not formed or is blocked. These elements probably form gradually in a process that involves competition between the growing RNA–DNA hybrid and protein structural elements. The reorganization observed in T7 RNAP therefore resembles the displacement and release of the bacterial σ-factor, and we suggest that a similar situation may exist in the multisubunit eukaryotic RNAPs. The induced reorganization of the enzyme as it matures and makes the transition to an EC is therefore probably a consistent theme among DNA-dependent RNAPs. □

Methods

Crystallization and data collection

Stable, transcriptionally competent complexes were formed by mixing RNAP at a 1:1 molar ratio with a nucleic acid scaffold consisting of an 18-nucleotide template DNA strand annealed to 10 nucleotides of downstream DNA and a 12-nucleotide RNA primer, of which 4 nucleotides at the 5′ end were unpaired with the template. Such complexes exhibit all of the properties of a promoter-initiated EC[24].

Crystallization was carried out by the sitting-drop vapour diffusion technique at 20 °C (ref. 25). The drop, containing 2 μl of 10 mg ml^{-1} complex solution, was mixed with 2 μl of well solution containing 10% PEG 8000, 10% glycerol, 5 mM β-mercaptoethanol, 100 mM Tris buffer, pH 8.1. Thin, plate-like crystals (0.5 × 0.5 × 0.05 mm) were grown within 3–5 days and diffracted to 2.6 Å resolution using synchrotron X-ray radiation. The T7 RNAP EC crystals belong to the *P*1 space group with unit cell dimensions $a = 79.91$, $b = 84.97$, $c = 202$ Å, $\alpha = 90.36$, $\beta = 92.97$, $\gamma = 109.94°$. A complete diffraction data set was collected at 100 K on SPring-8 beam line BL45XU, and processed with the HKL2000 program package[43] (Supplementary Table 1).

Structure determination and refinement

The structure was determined by molecular replacement using the coordinates of the T7 RNAP IC (Protein Data Bank accession number 1QLN[14]) as a search model[44]. Clear molecular replacement solutions for four crystallographically independent molecules were obtained only after removal of the N-terminal domain (266 residues) and all protruding segments of the C-terminal domain from the initial model, which indicated large structural alterations of the omitted portions of the molecule[25]. The phases obtained from the initial structure were improved by density modification and non-crystallographic symmetry (NCS) averaging using the DM program[44], resulting in an interpretable electron density for the omitted N-terminal domain and the RNA–DNA hybrid. The atomic model was built on the basis of this density using the O program[45] and then refined by the CNS program[46]. At this stage, however, the refinement was unstable, resulting in large portions of density that disappeared from the $(2F_o - F_c)$ electron density maps after additional refinement cycles. Careful inspection of diffraction data revealed barely separated twin spots at medium resolution, which become more profound at high resolution. This was characteristic of all crystals examined. To avoid this problem, the size of integration spots was increased at the stage of data processing to include both portions of the spots twinned at high resolution, and the diffraction intensities were calculated using summation rather than profile fitting to allow for better evaluation of the intensities for the two peaks in the integrated area. The four-fold NCS averaging appeared to be a powerful tool for the evaluation of the quality of the diffraction data (see Supplementary Information section 'NCS averaging'). The reprocessed data at 2.9 Å resolution resulted in clear electron density, which allowed for unambiguous modelling of the missing regions of the molecules with almost all protein side chains. The four independent molecules in the asymmetric unit are almost identical, with r.m.s.d. between the molecules ranging from 0.77 to 1.12 Å for the protein main chain and nucleic acid atoms. The downstream DNA is flexible in two molecules, and two regions of the protein were not resolved (757–765 and 352–371). Only weak electron density was visible for the four unpaired nucleotides (AACU) in the 5′ region of the RNA and for the last base pair of the downstream DNA (G•C, position +10). Figures 1a, b, 2a, b, 3 and 4b, c, were prepared using the programs Molscript[47], Bobscript[48] and Raster3D[49]. Figures 1c–e and 5 were prepared using the GRASP program[50].

Received 29 July; accepted 19 September 2002; doi:10.1038/nature01129.
Published online 9 October 2002.

1. McAllister, W. T. Transcription by T7 RNA polymerase. *Nucleic Acids Mol. Biol.* **11,** 15–25 (1997).
2. Brieba, L. G. & Sousa, R. T7 promoter release mediated by DNA scrunching. *EMBO J.* **20,** 6826–6835 (2001).
3. Gunderson, S. I., Chapman, K. A. & Burgess, R. R. Interactions of T7 RNA polymerase with T7 late promoters measured by footprinting with methidiumpropyl-EDTA-iron(II). *Biochemistry* **26,** 1539–1546 (1987).
4. Ikeda, R. A. & Richardson, C. C. 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 (1986).
5. Place, C., Oddos, J., Buc, H., McAllister, W. T. & Buckle, M. Studies of contacts between T7 RNA polymerase and its promoter reveal features in common with multisubunit RNA polymerases. *Biochemistry* **38,** 4948–4957 (2000).
6. Martin, C. T., Muller, D. K. & Coleman, J. E. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **27,** 3966–3974 (1988).
7. Liu, C. & Martin, C. T. Promoter clearance by T7 RNA polymerase. *J. Biol. Chem.* **277,** 2725–2731 (2002).
8. Liu, C. & Martin, C. T. Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase. *J. Mol. Biol.* **308,** 465–475 (2001).
9. Temiakov, D. *et al.* The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl Acad. Sci. USA* **97,** 14109–14114 (2000).
10. Huang, J. & Sousa, R. T7 RNA polymerase elongation complex structure and movement. *J. Mol. Biol.* **303,** 347–358 (2000).
11. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B. C. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **364,** 593–599 (1993).
12. Jeruzalmi, D. & Steitz, T. A. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J.* **17,** 4101–4113 (1998).
13. Cheetham, G., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399,** 80–83 (1999).
14. Cheetham, G. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286,** 2305–2309 (1999).
15. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–824 (1999).
16. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Ångstrom resolution. *Science* **292,** 1863–1876 (2001).
17. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Ångstrom resolution. *Science* **292,** 1876–1882 (2001).
18. Murakami, K. S., Musada, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296,** 1280–1284 (2002).
19. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å reslolution. *Nature* **417,** 712–719 (2002).
20. Cheetham, G. & Steitz, T. A. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* **10,** 117–123 (2000).
21. Severinov, K. T7 RNA polymerase transcription complex: what you see is not what you get. *Proc. Natl Acad. Sci. USA* **98,** 5–7 (2000).
22. Steitz, T. A. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.* **3,** 31–38 (1993).
23. McAllister, W. T. & Raskin, C. A. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol. Microbiol.* **10,** 1–6 (1993).
24. Temiakov, D., Anikin, M. & McAllister, W. T. Characterization of T7 RNA polymerase transcription complexes assembled on nucleic acid scaffolds. *J. Biol. Chem.* (in the press).
25. Temiakov, D. *et al.* Crystallization and preliminary crystallographic analysis of T7 RNA polymerase elongation complex assembled on an RNA:DNA scaffold. *Acta Crystallogr.* (submitted).
26. Kim, Y. *et al.* Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376,** 612–616 (1995).
27. Raskin, C. A., Diaz, G. A. & McAllister, W. T. T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl Acad. Sci. USA* **90,** 3147–3151 (1993).
28. Rong, M., He, B., McAllister, W. T. & Durbin, R. K. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **95,** 515–519 (1998).
29. Ma, K., Temiakov, D., Jiang, M., Anikin, M. & McAllister, W. T. Major conformational changes occur during the transition from an initiation complex to an elongation complex by T7 RNA polymerase. *J. Biol. Chem* (in the press).
30. Mukherjee, S., Brieba, L. G. & Sousa, R. Structural transitions mediating transcription initiation by T7 RNA polymerase. *Cell* **110,** 1–20 (2002).
31. Mentesana, P. E., Chin-Bow, S. T., Sousa, R. & McAllister, W. T. Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. *J. Mol. Biol.* **302,** 1049–1062 (2000).
32. Jiang, M., Rong, M., Martin, C. T. & McAllister, W. T. Interrupting the template strand of the T7 promoter facilitates translocation of the DNA during initiation, reducing transcript slippage and the release of abortive products. *J. Mol. Biol.* **310,** 509–522 (2000).
33. Bonner, G., Lafer, E. M. & Sousa, R. Characterization of a set of T7 RNA polymerase active site mutants. *J. Biol. Chem.* **269,** 25120–25128 (1994).
34. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by wild type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39,** 11571–11580 (2000).
35. Osumi-Davis, P. A., de Aguilera, M. C., Woody, R. W. & Woody, A. Y. Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. *J. Mol. Biol.* **226,** 37–45 (1992).
36. Woody, A. Y., Osumi-Davis, P. A., Hiremath, M. M. & Woody, R. W. Pre-steady state and steady-state kinetic studies on transcription initiation catalyed by T7 RNA polymerase and its active site mutants K631R and Y639F. *Biochemistry* **37,** 15958–15964 (1999).
37. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39,** 10419–10430 (2000).
38. He, B., Rong, M., Durbin, R. K. & McAllister, W. T. A mutant T7 RNA polymerase that is defective in RNA binding and blocked in the early stages of transcription. *J. Mol. Biol.* **265,** 275–288 (1997).
39. Brieba, L. G., Gopal, V. & Sousa, R. Scanning mutagenesis reveals roles for helix N of the bacteriophage T7 RNA polymerase thumb subdomain in transcription complex stability, pausing, and termination. *J. Biol. Chem.* **276,** 10306–10313 (2000).
40. Lyakhov, D. L. *et al.* Mutant T7 RNA polymerases with altered termination properties. *J. Mol. Biol.* **269,** 28–40 (1997).
41. Macdonald, L. E., Durbin, R. K., Dunn, J. J. & McAllister, W. T. Characterization of two types of

termination signal for bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **238,** 145–158 (1994).
42. Gopal, V., Brieba, L. G., Guajardo, R., McAllister, W. T. & Sousa, R. Characterization of structural features important for T7 RNAP elongation complex stability reveals competing complex conformations and a role for the non-template strand in RNA displacement. *J. Mol. Biol.* **290,** 411–431 (1999).
43. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
44. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
45. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
46. Brunger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
47. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
48. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55,** 938–940 (1999).
49. Merrit, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).
50. Nichols, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11,** 281–296 (1991).

Supplementary Information accompanies the paper on *Nature*'s website (**http://www.nature.com/nature**).

Acknowledgements We thank Y. Kawano for assistance during the data collection at the SPring-8 synchrotron beam line, BL45. We are grateful to A. Murzin for discussions and advice concerning the analysis of the structure. This work was supported in part by grants from the NIH (USA) (W.T.M.) and the Organized Research Combination System of Science and Technology Agency (Japan) (S.Y.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.G.V. (e-mail: dmitry@yumiyoshi.harima.riken.go.jp) or S.Y. (e-mail: yokoyama@biochem.s.u-tokyo.ac.jp). The atomic coordinates have been deposited in the Protein Data Bank under accession number 1H38.