Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution

Dmitry G. Vassylyev*†, Shun-ichi Sekine*†, Oleg Laptenko‡, Jookyung Lee‡, Marina N. Vassylyeva†§, Sergei Borukhov‡ & Shigeyuki Yokoyama*†§‖

** Cellular Signaling Laboratory, and † Structurome Research Group, RIKEN Harima Institute at Spring-8, 1-1-1 Kouto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan*
‡ 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

In bacteria, the binding of a single protein, the initiation factor σ, to a multi-subunit RNA polymerase core enzyme results in the formation of a holoenzyme, the active form of RNA polymerase essential for transcription initiation. Here we report the crystal structure of a bacterial RNA polymerase holoenzyme from *Thermus thermophilus* at 2.6 Å resolution. In the structure, two amino-terminal domains of the σ subunit form a V-shaped structure near the opening of the upstream DNA-binding channel of the active site cleft. The carboxy-terminal domain of σ is near the outlet of the RNA-exit channel, about 57 Å from the N-terminal domains. The extended linker domain forms a hairpin protruding into the active site cleft, then stretching through the RNA-exit channel to connect the N- and C-terminal domains. The holoenzyme structure provides insight into the structural organization of transcription intermediate complexes and into the mechanism of transcription initiation.

The DNA-dependent RNA polymerase (RNAP) is the principal enzyme of the transcription process, and is a final target in many regulatory pathways that control gene expression in all living organisms. In bacteria, RNAP is responsible for the synthesis of all RNAs in the cell (messenger RNA, ribosomal RNA, transfer RNA, and so on). The bacterial RNAP exists in two forms: core and holoenzyme. The core enzyme has a relative molecular mass of about 400,000, and consists of five subunits: α-dimer (α_2), β, β′ and ω. They are evolutionarily conserved in sequence, structure and function from bacteria to humans[1,2].

The transcription cycle in bacterial cells can be divided into three main stages: initiation, elongation and termination. Although it is catalytically active, the core enzyme is incapable of initiating transcription efficiently and with specificity. For this, it must bind an initiation factor, σ, to form a holoenzyme that can recognize specific DNA sequences (promoters)[3,4]. During initiation, the holoenzyme specifically binds to two conserved hexamers in the promoter at nucleotide (nt) positions −35 and −10 relative to the transcription start site (+1), to form a closed promoter complex. Then, it unwinds the double-stranded DNA (dsDNA) around the −10 region (between nt −12 and +2), resulting in the open promoter complex, and initiates transcription in the presence of nucleoside triphosphate substrates[5]. After the synthesis of an RNA 9–12 nt long, of which 8 or 9 are base-paired with the DNA template strand (DNA–RNA hybrid), the transcription complex passes from the initiation to elongation stage[6]. This transition is characterized by the escape of RNAP from the promoter, the dissociation of σ from the core, and the formation of a highly processive ternary elongation complex. An alternative mechanism has been proposed, in which σ factor remains bound to the core in the ternary complex during the elongation stage[7,8]. During the transition from initiation to elongation, RNAP undergoes a significant conformational change[5,9]. The σ factor has a central role in initiation, being directly involved in promoter recognition, DNA melting and promoter escape and clearance[4,10].

Among several distinct σ factors[4,11], σ70 is primarily responsible for the transcription of most 'housekeeping' genes in the cell during exponential growth. The family of related σ70 proteins shares four regions of sequence homology, designated 1–4, which are further divided into several subregions[12]. The regions 4.2, 2.3–2.4 and 2.5 were shown to recognize the −35, −10, and the so-called 'extended −10' (around nt −17 to −13) elements of the promoter, respectively[13–15]. In addition, regions 2.3–2.4 are known to interact specifically with the −10 region of the nontemplate strand in the open promoter complex[16,17]. Finally, genetic and biochemical data have shown that the portions of σ that contact the core include most of these conserved regions[18–20].

The crystal structures of the *Thermus aquaticus* RNAP core[21], the yeast *Saccharomyces cerevisiae* RNAP (polymerase II, Pol II)[22], and the Pol II ternary elongation complex with a 9-base pair (bp) DNA–RNA hybrid[23] revealed the high degree of structural similarity between the prokaryotic and eukaryotic RNAPs. Together with information gained from a wide range of biochemical, biophysical and genetic studies, these structures allow a plausible modelling of bacterial elongation and open promoter transcription complexes[24,25].

However, many aspects of the σ-dependent transcription initiation still remain obscure. These include: the promoter recognition by the holoenzyme; the formation of a closed promoter complex and its transition to an open complex; and the promoter escape and dissociation of σ factor from RNAP. The high-resolution structure of the *T. thermophilus* RNAP holoenzyme reported here provides information crucial to addressing these questions. It allows us to identify the structural changes in the core induced by σ binding, and to correlate them with the known functional, biologically relevant properties of the holoenzyme. Finally, the holoenzyme structure substantially improves the atomic model of the RNAP core, which is currently available at a modest resolution.

Structure determination and overall structure

The holoenzyme crystals belong to the $P3_2$ space group with unit cell dimensions $a = b = 236.35$, $c = 249.04$ Å. The structure was determined by molecular replacement and refined against data to 2.6 Å resolution (Fig. 1a, b; see also Supplementary Information).

The holoenzyme maintains the overall crab-claw-like shape (Fig. 1c) that was observed in the structures of the *T. aquaticus* core and *S. cerevisiae* Pol II[21,22]. However, the presence of the σ70 subunit and the large portion of the nonconserved N-terminal domain of the β′ subunit in the model greatly extends the β′ pincer, thus making the analogy to a crab claw even more fitting. The σ70 subunit is located almost entirely on the core surface, except for a short segment (σ313–342), which is buried within the core molecule. The binding of σ to the RNAP core does not seem to result in the widening of the DNA-binding channel, as was previously thought. On the contrary, certain structural elements of σ reduce the available space of functionally important cavities in RNAP that are presumed to accommodate the transcription bubble and the RNA product. These elements include: (1) the second N-terminal domain of σ, which bridges the β and β′ pincers of the claw on the upstream DNA side of RNAP, creating a wall that blocks the opening of the DNA-binding channel; (2) a hairpin-like motif of σ, which protrudes into the cleft of the active site, towards the possible location of the RNA–DNA hybrid; and (3) an extended fragment following the hairpin-like motif, which occupies the RNA-exit channel.

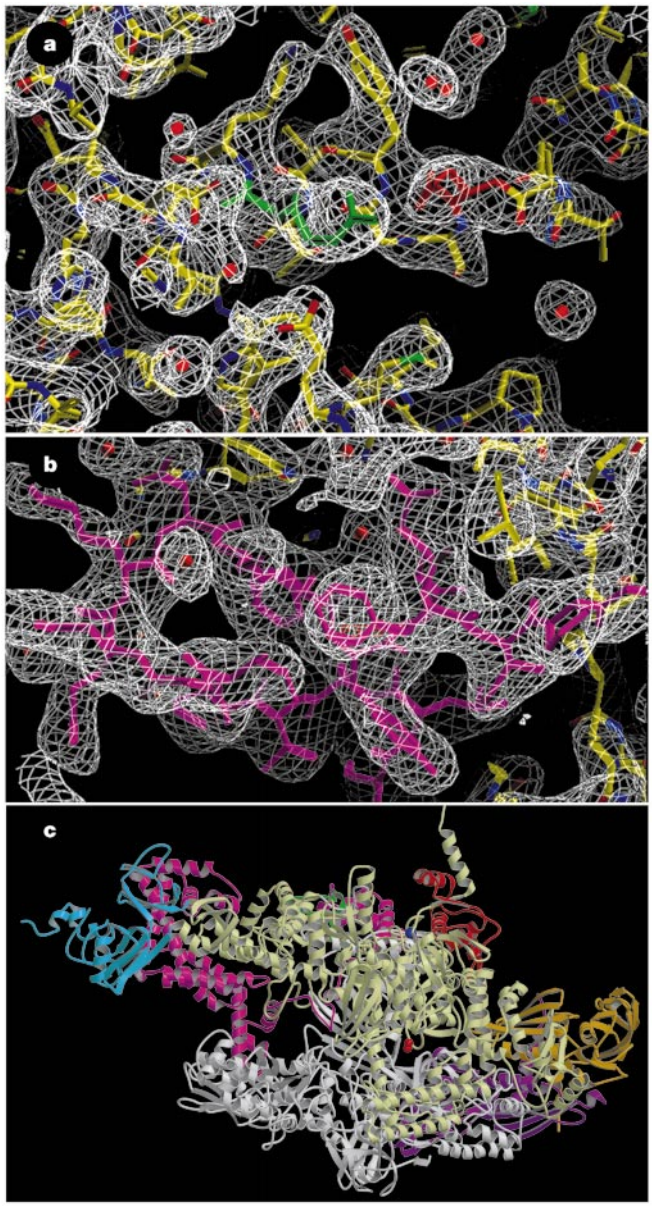

Figure 1 Holoenzyme crystal structure. **a**, **b**, The β′F bridge helix (**a**) and the hairpin loop of σ (**b**). The $|2F_O - F_C|$ electron density map (1.3σ contour level; white) is superimposed on the atomic model. Protein atoms have atom-dependent colours, except for the σ-subunit (magenta) and the β′ residues Asp 1,090 (red) and Arg 1,096 (green). Water molecules are represented by red spheres. **c**, The overall structure. The colours of the subunits are: β, khaki; β′, white (β′ 163–452, cyan; β′ Zn finger, green); αI, blue; αII, light orange; σ, magenta; and ω, red. Two catalytic Mg^{2+} (red) and two Pb^{2+} (Zn^{2+}) ions (blue) are shown as spheres.

The core-to-holoenzyme conversion is accompanied by changes in all structural domains of the core enzyme. The most noticeable include the βG flap domain (β702–828), the αI subunit, the large N-terminal portion of β (β20–395), and the β′ coiled-coil domain (β′540–585) (Fig. 2). The RNA-exit channel is most significantly affected by σ70 binding. The channel is constricted by the β flap domain, which is shifted by about 5–6 Å towards the σ C-terminal domain compared with the *T. aquaticus* core[21]. The change in the orientation of the 'flap tip' helix, which is now squeezed between three α-helices of σ (Fig. 2), is even more dramatic (~11 Å shift). The change in the orientation of individual domains of the RNAP subunits (αI and β20–395) that are adjacent to the β flap domain seems to mirror the flap domain rotation.

In the crystal structure, we observed many Mg^{2+} ions bound to holoenzyme, most of which formed a coat on the protein surface. These metal ions could be important in the binding and bending of the DNA molecule, which is proposed to be wrapped around RNAP during transcription[26].

Structure of σ70

The atomic model of the σ70 subunit lacks the disordered N-terminal domain (σ1–73), which includes the conserved σ region 1.1. This is a self-inhibitory domain, which is known to mask the DNA-binding regions of σ before it binds the core[27]. Its precise role in transcription initiation is not understood.

The modelled structure of σ70 consists entirely of α-helices connected by either turns or loops (Fig. 3). It can be divided into four structural domains: N-terminal domain 1 (ND1), N-terminal domain 2 (ND2), 'linker' domain (LD) and C-terminal domain (CD).

ND1 consists of eight α-helices (σ74–254) comprising four helix–turn–helix (HtH) motifs: HtH1 (σ75–121), HtH2 (σ94–137), HtH3 (σ123–147) and HtH4 (σ152–203) (Figs 3 and 4a). This domain encompasses region 1.2 up to the N-terminal half of region 2.4, including the nonconserved segment between regions 1 and 2. ND1 has a U-shaped structure, and is connected to ND2 (σ261–312) by a short linker loop (σ255–260), which lies at the C terminus of region 2.4.

ND2, corresponding to regions 2.4–2.5 and 3.1, consists of three α-helices that fold into an α-helical bundle (Figs 3 and 4a). The C-terminal helix of ND1 and the N-terminal helix of ND2 (σ234–281) form a V-shaped structure (Fig. 4a) near the opening of the upstream DNA-binding channel. The inner surface of the V is lined by the residues from the C- and N-terminal α-helices of ND1 and ND2, respectively. These α-helices correspond to regions 2.3–2.5 (Fig. 3c), which are known to be involved in the specific binding of σ to the −10 and extended −10 elements of the promoter[14,15].

The 30-residue LD (σ313–339; region 3.2) intervenes between the globular N- and C-terminal portions of σ, and has mostly an extended, unfolded conformation (Figs 3b and 4b). Roughly at its midpoint, the LD forms a hairpin loop (σ318–329) that protrudes into the active site cleft, between the 'rudder' loop (β′584–601) and the 'lid' loop (β′526–538), which are believed to interact directly with nucleic acids[21,22]. The rest of the LD stretches along the RNA-exit channel towards σ's CD.

The CD (σ340–423), which includes conserved regions 4.1 and 4.2, contains four α-helices, which are arranged as a pair of HtH motifs (HtH5, σ341–376; and HtH6, σ387–418) (Figs 3 and 4c). The conserved σ region 4 was shown to recognize the −35 promoter element specifically[28,29]. The CD is located on the core surface in the

vicinity of the opening of the RNA-exit channel. This site is about 57 Å from the N-terminal half of the protein, containing regions 2.3–2.5.

All of the conserved sequence motifs between the *T. thermophilus* σ70 and the *Escherichia coli* σ70 fragment[30] are also conserved in three dimensions with a root-mean-square deviation (r.m.s.d.) of 2.9 Å over 125 α-carbon (Cα) atoms[31]. Therefore, it is likely that the *E. coli* σ70 conformation and mode of core binding are similar to those of the *T. thermophilus* protein. Interestingly, we detected an intrinsic structural similarity between the CD and the nonconserved portion of ND1 in *T. thermophilus* σ (Fig. 4d), although they do not exhibit any sequence homology. Structural similarity also exists, to a lesser extent, between the CD and the nonconserved region of the *E. coli* σ70 fragment (σ274–338). This observation raises the possibility that these structural segments have evolved from a common ancestor. Their structural similarity and the presence of HtH motifs suggest that the nonconserved ND1 region of σ might have served as an alternative DNA-binding site at some stage of evolution.

β′ zinc finger and N-terminal domains

The previously reported structure of the *T. aquaticus* core enzyme lacks the 'zinc finger' domain (β′53–81) and the N-terminal nonconserved region of the β′ subunit (β′163–452). These domains are fixed in the holoenzyme structure.

The zinc finger domain has no secondary structure, and folds into a U-shaped structure, with two long loops harbouring one Pb^{2+} ion (Zn^{2+} in the native enzyme) between them. In the holoenzyme, this domain is located between the N- and C-terminal domains of σ70, in the vicinity of σ's CD, which implies a potential role in promoter binding (Fig. 4a, c).

We traced a large portion of the β′ N-terminal nonconserved domain, spanning residues β′163–251 and β′364–452. This domain consists mostly of β-strands, with one α-helix near its N terminus, and two short α-helices at the end of its N-terminal portion, just before the gap. The nonconserved β′ domain is close to the σ ND1 domain, which suggests that its primary function is as an anchor for the σ70 subunit (Figs 1c and 4a).

Interactions between σ70 and the core

In summary, among the four structural domains of σ70, ND1 makes the most extensive contacts with the RNAP core. Its U-shaped structure folds around the nearby α-helical coiled-coil of β′ (β′540–585)[18,20,32,33]. Residues from both nonconserved and conserved (1.2, 2.1, 2.2, 2.4) segments of ND1 make multiple polar interactions with these β′ α-helices (Fig. 4a). The outer surface of the U interacts with the N-terminal portion of the β′ subunit (β′95–460), which includes a large portion of the nonconserved segment.

The ND2 of σ70 (Fig. 4a) is loosely bound to the core. Although it bridges the ββ′ pincers of the claw, it has only a few direct interactions with the core molecule. However, the binding is enhanced by multiple water-mediated interactions between ND2 and the core. The putative flexibility of ND2 may be important in the −10 hexamer recognition and/or in the discrimination between promoters with −10 or extended −10 elements[14,15].

The σ70 LD makes few contacts at its N terminus with the β′ loops that form the 'rudder' and 'lid' structures (Fig. 4b). The hairpin loop (σ318–329) interacts with the 'lid' loop, as well as the short α-helix (β′612–619) that follows the rudder in β′. The C-terminal half of the LD, which is entirely buried within the RNA-exit channel, makes a couple of contacts with the N-terminal segment of β′ and the C terminus of β.

The interaction of σ CD with the core is limited to its contacts with the flap tip helix of the β subunit (β768–779) on the one side, and the β′ zinc finger domain on the other (Fig. 4c).

Altogether, there are surprisingly few direct polar interactions between the residues of the σ and core molecules, despite their large interface area.

Catalytic site

As in the structures of the *T. aquaticus* core[21] and *S. cerevisiae* Pol II[22], the active site of *T. thermophilus* RNAP is defined by the three invariant aspartate residues of the β′ subunit (Asp 739, Asp 741 and Asp 743), located within a highly conserved sequence motif (NADFDGD)[34]. In the holoenzyme structure, one of the two molecules in the asymmetric unit has one Mg^{2+} ion chelated by three active-site aspartates. The second molecule of the holo-

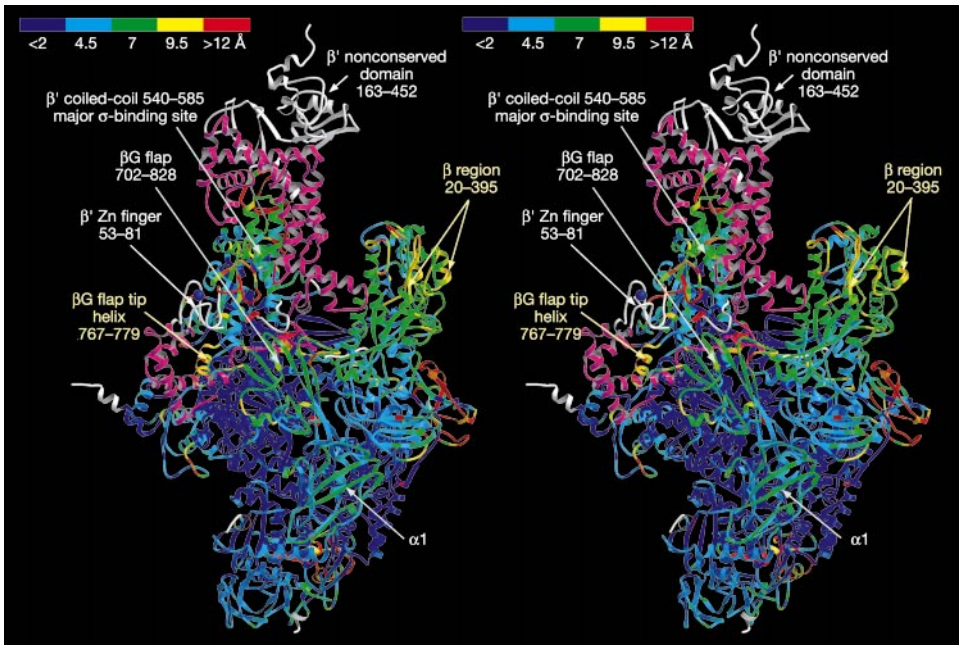

Figure 2 Comparison between the holoenzyme and *T. aquaticus* core structures. In the stereo view shown, the holoenzyme is coloured from blue (<2 Å) to red (>12 Å), according to the deviation in the positions of the Cα atoms of the corresponding holoenzyme and core residues. The σ subunit and other structural elements that are absent from the *T. aquaticus* structure are shown in magenta and white, respectively. Most of the deviations that are more than 12 Å (red) correspond to the mis-assigned regions in the *T. aquaticus* structure. The structures were superimposed by the β′ subunits (r.m.s.d. = 1.65 Å over 700 Cα atoms).

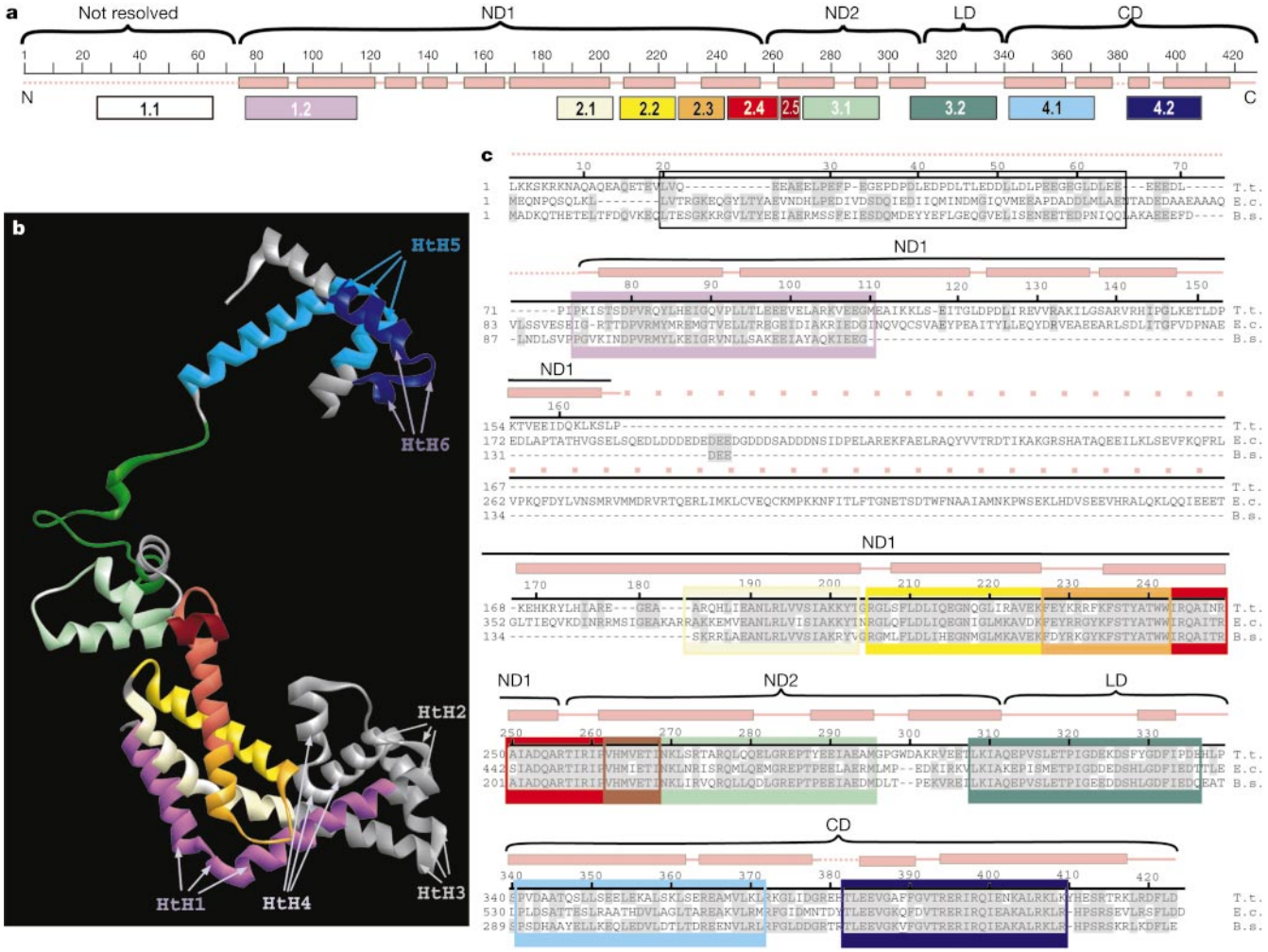

Figure 3 Organization and structure of the σ. **a**, Scheme of structural domains and conserved regions. **b**, Ribbon diagram of σ. The colour coding is the same as in **a** except for the nonconserved regions (grey). **c**, Secondary structure and sequence alignment. Numbers of *T. thermophilus* residues are used. In **a** and **c**, the secondary structure is indicated in pink as boxes (α-helices), solid lines (coils) and dotted lines (disordered structures). The σ factors from *T. thermophilus* (T.t.), *E. coli* (E.c.) and *Bacillus subtilis* (B.s.) were aligned using the LaserGene program. Identical residues are greyed. Conserved regions of σ (1.1–4.2 in **a**) are boxed and coloured as in **a**.

enzyme contains two Mg^{2+} ions (with *B*-factors of 26 Å^2 and 27 Å^2) bound at the catalytic centre. The same three conserved aspartates, together with several neighbouring water molecules, form a network of interactions responsible for holding these two Mg^{2+} ions (Fig. 5a). Two metal ions are separated by 3.0 Å, and the distance between each metal ion and the oxygen ligand is in the range of 2.5–4.0 Å. These results are consistent with the two-metal-ion mechanism proposed for all nucleic acid polymerases[35].

The bridge helix

The β′ bridge helix (β′1,067–1,102), which separates the internal space occupied by the transcription bubble and the downstream DNA, had somewhat different conformations in the structures of the *T. aquaticus* core and the eukaryotic Pol II[21–23]. This difference may reflect two possible states of the bridge helix in both enzymes. It has been proposed that the transition between the two conformations triggers the translocation of RNAP during RNA polymerization[22]. Here, we present a more detailed view of this process, based on the high-resolution structure of the holoenzyme.

The overall shape of the bridge helices in the bacterial holoenzyme and Pol II is nearly identical (r.m.s.d. = 1.1 Å over 33 Cα atoms), with both helices having a bend of about 28° (Fig. 5b). The conformational difference between the two enzymes is localized to the central region of the helix. In Pol II, the bridge helix maintains a uniform α-helical conformation, whereas in the holoenzyme, the backbones of two residues (Asp 1,090, Ser 1,091) in the middle are flipped out of the α-helix (not partially unwound, as observed in the *T. aquaticus* core structure). This pair of flipped-out residues could serve as a molecular 'switch' in RNAP translocation. In the holoenzyme structure, the observed unstable, flipped-out conformation of the bridge helix centre is stabilized by intrahelical hydrogen bonding between Asp 1,090 and the invariant Arg 1,096 (Fig. 5c). In the Pol II elongation complex structure (Protein Data Bank (PDB) accessions number 1I6H), the counterpart of this arginine bridges two adjacent phosphates of the template strand DNA at positions i and i_{+2} (conventionally, i and i_{+1} refer to the product-binding and substrate-binding sites, respectively) (Fig. 5d). In addition to the bridge helix, a conserved β′G loop comprising residues 1,238–1,250 may also be involved in the translocation process. This loop is traced in the holoenzyme structure, but is disordered in both the *T. aquaticus* core and Pol II structures[21,22]. The β′G loop forms part of the floor of the downstream DNA-binding cavity, and is located close to the distorted, central region of the bridge helix (Fig. 5c). If we model the bridge helix in a uniform α-helical conformation (as observed in the Pol II structure), then the Asp 1,090 side chain would clash with a patch of highly conserved hydrophobic residues (Met 1,238, Phe 1,241 and Leu 1,256) located in this flexible β′G loop (Fig. 5d). Thus, it appears that the bridge

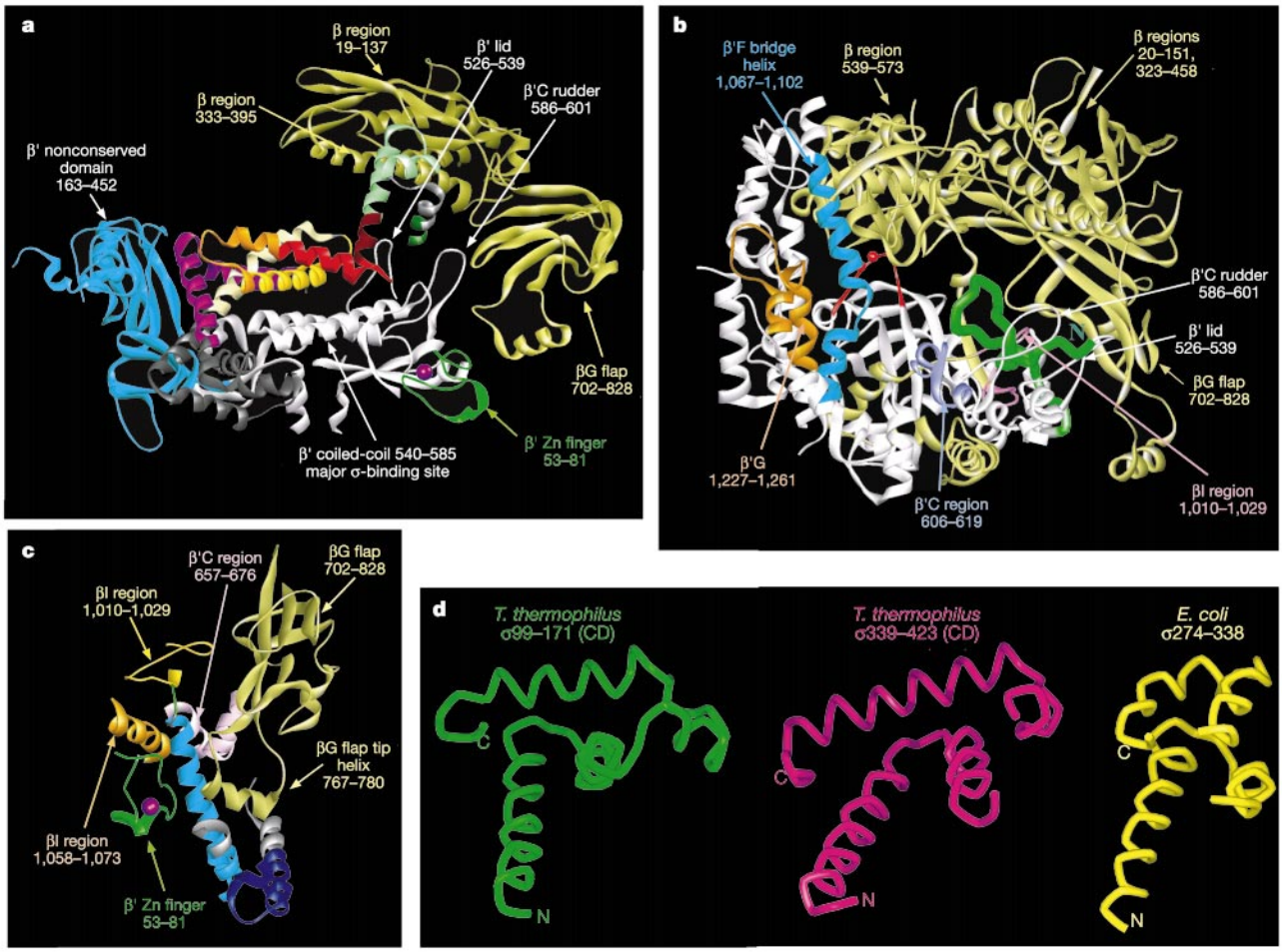

Figure 4 Individual σ domains and their interactions with the core. The conserved regions of σ have the same colours as in Fig. 3a. **a**, The σ domains ND1 and ND2 and the neighbouring core regions are shown. **b**, Partial view of the active centre cleft, showing the buried location of σ LD (green) and its hairpin loop. The catalytic loop and the Mg^{2+} ions are shown in red. **c**, The σ CD is shown together with the closely interacting regions of β and β′. **d**, Structural similarity between the *T. thermophilus* conserved σ CD, and the nonconserved regions of the *T. thermophilus* σ ND1 and *E. coli* σ 70 fragment.

helix may alternate between the flipped-out and normal α-helical conformations, depending on the position of the β′G loop (the 'trigger' loop).

Taken together, these structural data allow us to propose a possible mechanism for RNAP translocation during RNA synthesis. At the step of 'relaxation' (after translocation, before the next reaction), the template base at position i_{+1} is paired with the substrate nucleoside triphosphate (NTP), the bridge helix is in an all α-helical conformation, the Arg 1,096 bridges the i and i_{+2} DNA phosphates, and the flexible trigger loop is distal (rather than proximal) to the bridge α-helix. After phosphodiester bond formation, a signal induces the movement of the trigger loop towards the bridge α-helix, pushing out the 'switch' residues. In their flipped-out conformation, the switch residues may engage the DNA phosphate at position i_{+1} and bring the bridge α-helix under the DNA backbone towards the i_{+2} nucleotide. During this step, Arg 1,096 may also switch its interacting partner from DNA phosphates to the side chain of an acidic (polar) switch residue, thus simultaneously stabilizing the flipped-out conformation of the switch residues and facilitating the translocation of the enzyme (Fig. 5c, d).

Promoter recognition by the RNAP holoenzyme

A working model of bacterial open promoter complex has been constructed that is largely consistent with the genetic, biochemical and structural data[25]. In this model, a straight dsDNA fragment is placed to indicate the orientation of the promoter region upstream of its −10 element on the surface of the RNAP core enzyme. The presence of the σ subunit in our structure allows us to further improve on this model. After superposition of the core and holoenzyme structures, a simple translation of this dsDNA fragment is enough to bring it into plausible assembly with the holoenzyme (Fig. 6a). In this model, the −35 promoter hexamer is recognized by the recognition helix in HtH6 of the σ CD (region 4.2). The −10 promoter region apparently passes between the two central domains of σ, ND1 and ND2, which form a V-shaped nook at the entrance to the active site cleft. However, the 17-bp spacer between the −35 and −10 regions appears to be too long to be accommodated as a straight dsDNA fragment between the two promoter recognition modules of σ. Therefore, a local bend or a kink in the DNA seems necessary to bring the −10 promoter element and region 2.4 of σ into contact (Fig. 6a), making the model consistent with the biochemical and genetic data[14,15,36,37]. In particular, in this model, the residues from regions 2.3–2.4 (σ241–255) and 2.4–2.5 (σ256–269) are inserted in the major groove of the DNA, in the vicinity of the −12, −11 and −13 to −16 base pairs, respectively, which would allow base-specific recognition of the −10 and extended −10 regions of the promoter (Fig. 6a).

DNA melting by RNAP holoenzyme

After forming a closed complex at the promoter, the holoenzyme melts and unwinds the DNA near the −10 region (nt −12 to +2) to

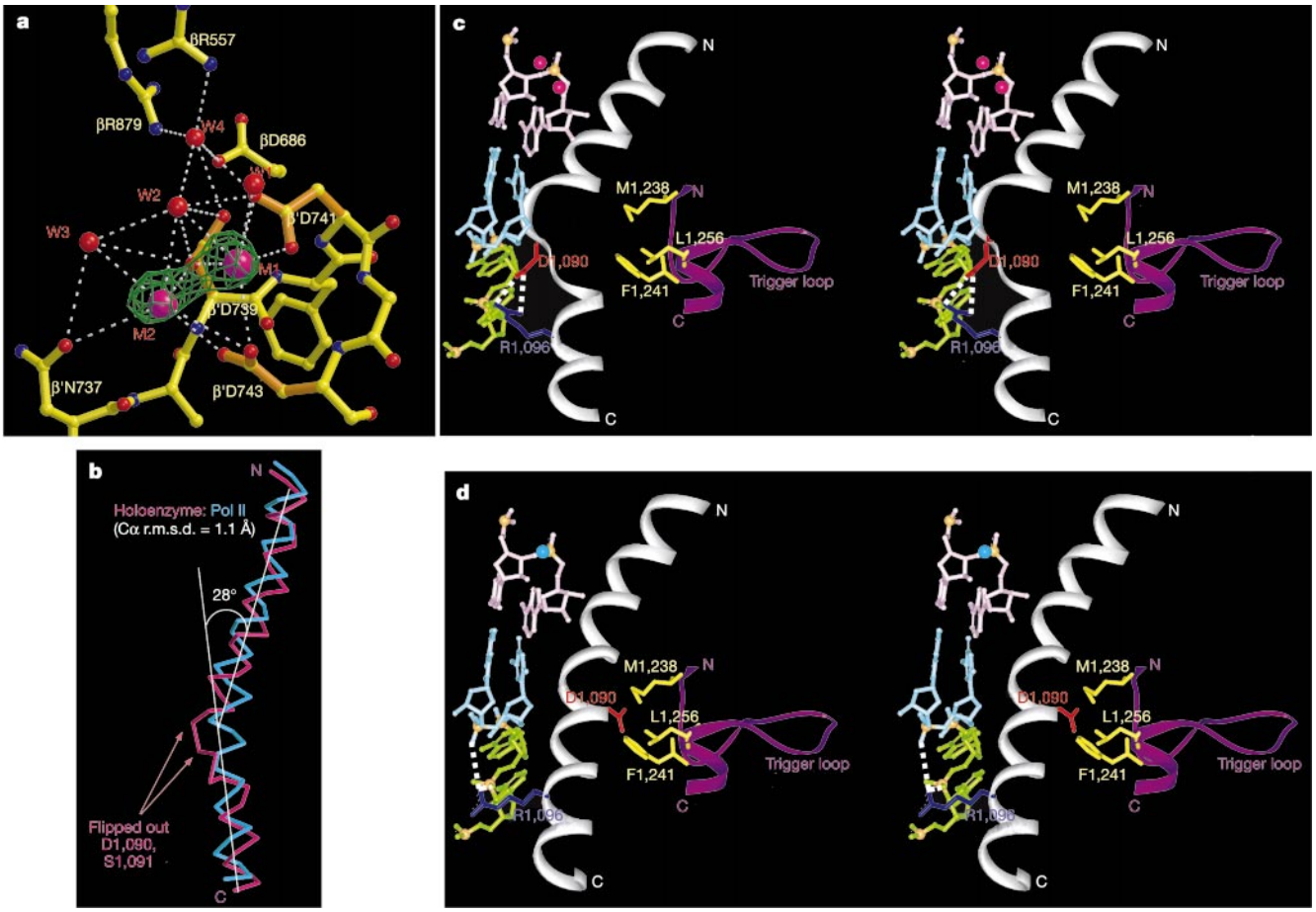

Figure 5 Catalytic centre. **a**, The Mg^{2+}-binding site. The slow annealing omit $|F_o - F_c|$ map contoured at 4.5σ (green) is shown for Mg^{2+} ions. Mg^{2+} ions (magenta) and water molecules (red) are shown as spheres. **b**, Superposition of the holoenzyme (magenta) and Pol II[23] (cyan) bridge helices. **c**, **d**, Two conformational states (stereo views) of the bridge helix (white); in the holoenzyme (**c**) and in Pol II (**d**). The G-loop (purple), modelled for Pol II in **d**, and the DNA/RNA (cyan/pink) hybrid[23] (downstream DNA, green), modelled for the holoenzyme in **c**, are shown. The catalytic Mg^{2+} ions are shown as magenta (**c**) and cyan (**d**) spheres.

form an open promoter complex. The analysis of the holoenzyme structure and the model for the closed promoter complex discussed above allow us to hypothesize as to how the holoenzyme may catalyse promoter opening. Once the closed promoter complex is formed, the downstream DNA would eventually move towards its binding site on the RNAP, which was previously localized in a deep cavity[23–25], about 30 Å away from the location of the −10 recognition site (Fig. 6b, c). To achieve this, the dsDNA is likely to be kinked or bent near the −12 to −11 region of the promoter, in the immediate vicinity of the entrance to the active site cleft (Fig. 6b). Disruption of base-pairing at the −12 to −10 region and DNA strand separation (nucleation of promoter melting[38]) should follow the specific recognition of this region by σ. We speculate that the melting of two or three DNA base pairs would sufficiently increase the DNA flexibility to direct the downstream dsDNA towards its binding site on the RNAP. However, the downstream dsDNA access to the binding site is still obstructed by the presence of the conserved σ α-helix (σ74–92, region 1.2) and the β helix–loop–helix fragment (β232–259) (Fig. 6b, c). The distance between these elements (~16 Å) is much shorter than the dsDNA diameter of about 22 Å (Fig. 6c). Interestingly, in our model, the β-subunit loop (β242–253) protruding from the helix–loop–helix motif fits nicely in the dsDNA major groove, in close proximity to the +1 to +3 nt positions, which are known to be the end point of promoter melting (Fig. 6a, b). The sequence analysis shows a consensus motif of six residues—(L/M)RPGEP—in bacterial RNAPs (β242–247 in *T. thermophilus*), which are spatially located at the tip of the β loop. This β loop may serve as a gate that allows single-stranded, but not double-stranded, DNA to go through. By interacting with DNA, this 'gate' loop may also stabilize the DNA orientation to favour melting, and/or to facilitate the downstream DNA rotation or oscillation around the helical axis to unwind the dsDNA. It should be noted that in the core enzyme, which is incapable of melting dsDNA, the absence of σ opens enough space at this region for the dsDNA to go through (Fig. 6d). After the DNA melting reaches the β loop (around nt +2 or +3), the unwound DNA and the downstream dsDNA may continue their movement towards the active site to complete the transcription bubble formation.

The mechanism of promoter DNA melting is largely unknown, and the 'gate' loop of the β subunit described here has never been the subject of genetic or biochemical studies. Future analyses of this region are required to shed light on its possible function.

Promoter escape

To better understand the mechanism of promoter escape by RNAP and σ release, we constructed a model of a DNA–RNA hybrid (9 bp) bound to the holoenzyme, based on a simple superposition of the conserved catalytic domains from the crystal structures of the eukaryotic elongation complex[23] and the *T. thermophilus* holoenzyme. According to our model, the DNA–RNA hybrid probably fits in a narrow passage (~11 Å wide) formed by the helix–loop–helix β fragment (β375–408), and a hairpin loop from σ70 LD (σ318–329) protruding into the major groove of the DNA–RNA hybrid double helix (Fig. 6e). The putative interactions of the σ hairpin loop with the nucleotides in the hybrid may further deform the unusual 'underwound' conformation of the DNA–RNA

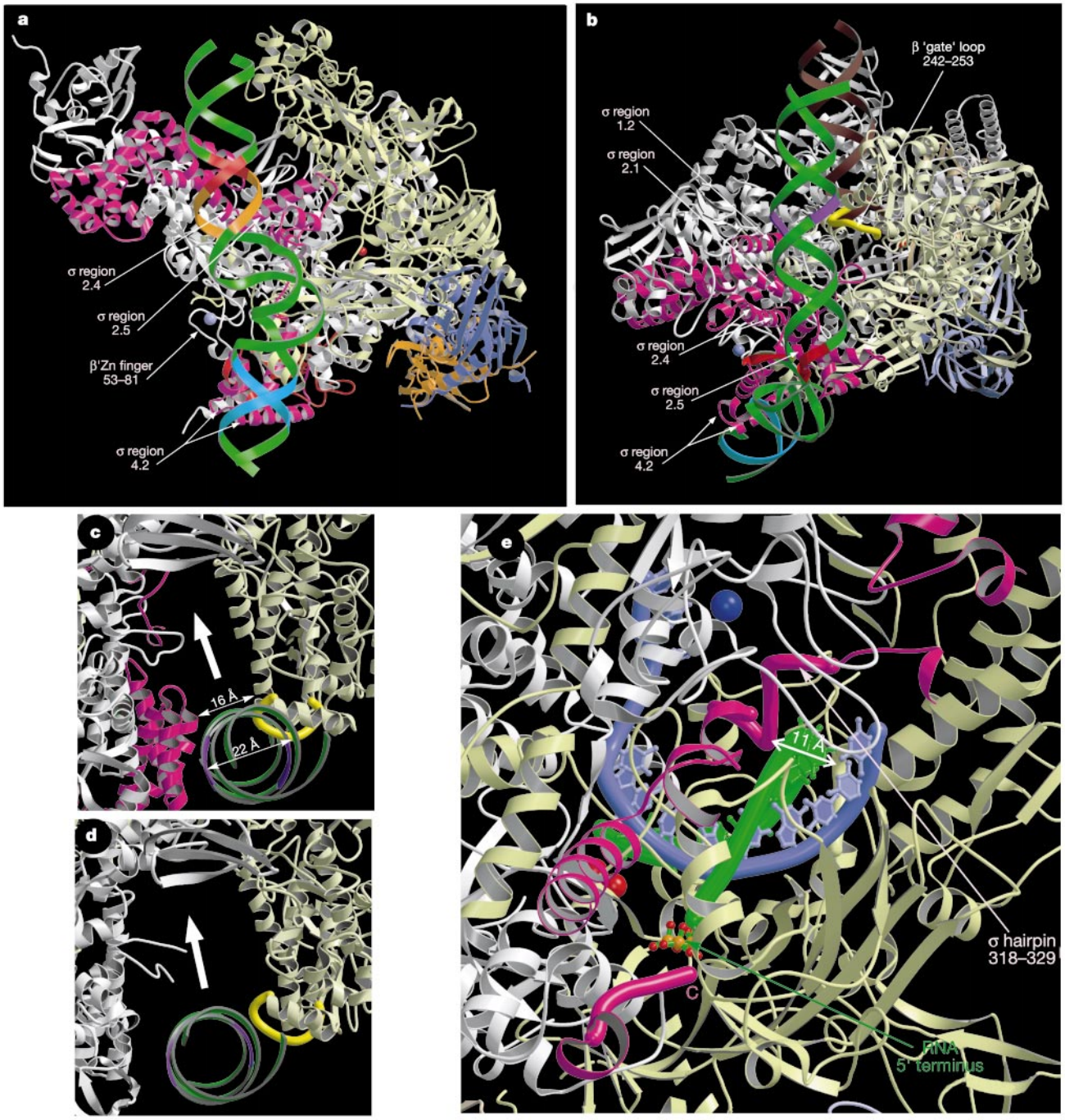

Figure 6 Models of the holoenzyme–nucleic acid complexes. The protein colour coding is the same as in Fig. 1c, except all of β′ is shown in white. **a**, **b**, The closed (**a**) and partially melted (**b**) promoter complexes. **c**, **d,** The same DNA model as in **b**, but looking down the dsDNA axis for the holoenzyme (**c**) and for the *T. aquaticus* core (**d**). The dsDNA (green) contains the −35 (cyan), −10 (red), extended −10 (orange) and +1 to +3 (blue) promoter regions. A brown dsDNA (**b**) indicates the downstream DNA-binding site in the open promoter complex. **e**, The model of the holoenzyme complex with the DNA/RNA (blue/green) hybrid and the ssRNA transcript. The Mg^{2+} (red) and Pb^{2+} (blue) ions are shown as spheres.

hybrid[23]. Thus, the loop may act as a destabilizer to facilitate the disruption of the DNA–RNA hybrid base pairs. At the same time, the hybrid, competing with the σ destabilizer loop for limited space, may displace the loop from the active site cavity, thereby initiating the release of σ from the promoter and/or RNAP. This putative competition may also result in the release of abortive RNA products. After strand separation, the RNA chain is likely to enter the RNA-exit channel in an extended, rather than stacked, conformation. This is because the channel is substantially narrowed in the holoenzyme structure (10–12 Å in diameter, compared with 16–19 Å in the *T. aquaticus* core) by the presence of the σ LD and the more closed conformation of the β flap domain. When RNAP synthesizes a 12-nt RNA, the 5′-terminal 3 or 4 nt in the extended conformation may easily reach the outlet of the RNA-exit channel, which in the holoenzyme structure is blocked by the σ C terminus (Fig. 6e). The potential steric collision between σ CD and nascent RNA may facilitate the displacement of the σ CD from the RNAP and from the −35 region of the promoter, to trigger the dissociation of σ70 and promoter escape. This mechanism is consistent with recent biochemical data[39]. □

Methods

Crystallization and data collection

The *T. thermophilus* holoenzyme was purified as described[40]. The holoenzyme crystals were grown by the hanging-drop vapour-diffusion technique. For crystallization, 2 μl of well solution were mixed with 2 μl of protein solution (protein concentration 10 mg ml^{-1}). The final conditions, 13 mM magnesium formate, 5% polyethylene glycol 400, 2 mM spermine, 2 mM dithiothreitol and 20 mM MES buffer (pH 5.8), yielded high-quality hexagonal crystals[40]. The crystals of the native enzyme and several heavy-atom derivatives were subjected to data collection at SPring-8 synchrotron beam lines. Among these, the crystal of the lead heavy-atom derivative—$(CH_3)_3Pb(CH_3COO)_2$—showed the best diffraction pattern, and the data from this crystal were used for refinement (see Supplementary Information). To increase the data completeness, data from the same crystal were collected at the BL44XU and BL45XU beam lines. The data were processed by the HKL2000 program package[41].

Structure determination and refinement

Careful analysis of the diffraction data, as described previously[40], revealed that the holoenzyme crystals belong to the $P3_2$ space group, have perfect merohedral twinning along a three-fold axis[42], and contain two molecules in the asymmetric unit, the orientation of which closely resembles crystallographic $P6_5$ symmetry. The molecular replacement solution, which was obtained previously for the 3 Å native data using the *T. aquaticus* RNAP core enzyme structure (PDB number 1I6V) as a search model[40], was used as a starting point for the holoenzyme structure determination. The rigid body refinement using the CNS program[43] converged to an R_{free} of 39.2% at 2.6 Å resolution. At this stage, the solvent flipping procedure[44], as implemented in the CNS program[43], greatly improved the electron density map of the σ subunit and the regions of the core enzyme that were absent from the *T. aquaticus* protein structure. The refinement and model building revealed several 'local' mistakes in the sequence assignment of the *T. aquaticus* core, with the largest misassigned region of 70 amino acids (β460–530) in the β subunit. The model was gradually improved by alternate cycles of CNS refinement and model building using the O program[45]. The two lead ions in each holoenzyme molecule substitute for zinc in the two zinc-binding sites. This is revealed by the more than 20σ peaks in the difference electron density map at 3.5 Å resolution produced between the lead derivative and native data (data not shown). At the final stage, water molecules were added to the model, using alternate steps of the 'water pick' and 'water delete' procedures as implemented in the CNS program[43]. The refinement converged to a final *R*-factor value of 22.8% (R_{free} = 27.4%) for all data at 2.6 Å resolution (see Supplementary Information). The asymmetric unit contains two holoenzyme molecules, with r.m.s.d. = 1.05 Å over all main-chain atoms. Figures 1, 2, 5a and 6 were prepared using the programs O[45], Molscript[46], Bobscript[47] and Raster3D[48]. Other figures were prepared by the WebLab ViewerPro program (MSI).

Received 26 February; accepted 25 April 2002; doi:10.1038/nature752.
Published online 8 May 2002.

1. Sweetser, D., Nonet, M. & Young, R. A. Prokaryotic and eukaryotic RNA polymerase have homologous core subunits. *Proc. Natl Acad. Sci. USA* **84,** 1192–1196 (1987).
2. Ebright, R. H. RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* **304,** 687–698 (2000).
3. Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. F. Factors stimulating transcription by RNA polymerase. *Nature* **221,** 43–44 (1969).
4. Gross, C., Lonetto, M. & Losick, R. in *Transcriptional Regulation* (eds McKnight, S. R. & Yamamoto, K. R.) 129–176 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1992).
5. Record, M. T. J., Reznikoff, W., Craig, M., McQuade, K. & Schlax, P. in *Escherichia coli and Salmonella* (ed. Neidhart, F. C.) 792–820 (ASM, Washington DC, 1996).
6. von Hippel, P. An integrated model of the transcription complex in elongation, termination, and editing. *Science* **281,** 660–665 (1998).
7. Bar-Nahum, G. & Nudler, E. Isolation and characterization of σ^{70}-retaining transcription elongation complexes from *Escherichia coli*. *Cell* **106,** 443–451 (2001).
8. Mukhopadhyay, K. *et al.* Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **106,** 453–463 (2001).
9. Craig, M. L. *et al.* DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase–promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. *J. Mol. Biol.* **283,** 741–756 (1998).
10. Gross, C. *et al.* The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harbor Symp. Quant. Biol.* **63,** 141–155 (1998).
11. Ishihama, A. Functional modulation of *Escherichia coli* RNA polymerase. *Annu. Rev. Microbiol.* **54,** 499–518 (2000).
12. Lonetto, M., Gribskov, M. & Gross, C. The σ^{70} family: sequence conservation and evolutionary relationships. *J. Bacteriol.* **174,** 3843–3849 (1992).
13. Dombroski, A. J., Walter, W. A., Record, M. T., Siegele, D. A. & Gross, C. A. Polypeptides containing highly conserved region of transcription initiation factor Σ 70 exhibit specificity of binding to promoter DNA. *Cell* **70,** 501–512 (1992).
14. Barne, K. A., Bown, J. A., Busby, S. J. W. & Minchin, S. D. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended −10' motif at promoters. *EMBO J.* **16,** 4034–4040 (1997).
15. Fenton, M. S., Lee, S. J. & Gralla, J. D. *Escherichia coli* promoter opening and −10 recognition: mutational analysis of σ70. *EMBO J.* **19,** 1130–1137 (2000).
16. Marr, M. T. & Roberts, J. W. Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. *Science* **276,** 1258–1260 (1997).
17. Huang, X., Lopez de Saro, F. J. & Helmann, J. D. Sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with −10 region single-stranded DNA. *Nucleic Acids Res.* **25,** 2603–2609 (1997).
18. Sharp, M. *et al.* The interface of σ with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev.* **13,** 3015–3026 (1999).
19. Lesley, S. A. & Burgess, R. R. Characterization of the *Escherichia coli* transcription factor sigma 70: localization of a region involved in the interaction with core RNA polymerase. *Biochemistry* **28,** 7728–7734 (1989).
20. Owens, J. T. *et al.* Mapping the σ^{70} subunit contact sites on *Escherichia coli* RNA polymerase with a σ^{70}-conjugated chemical protease. *Proc. Natl Acad. Sci. USA* **95,** 6021–6026 (1998).
21. Zhang, G. *et al.* A crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–824 (1999).
22. Cramer, P., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* **292,** 1863–1876 (2001).
23. 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 Å resolution. *Science* **292,** 1876–1882 (2001).
24. Korzheva, N. *et al.* A structural model of transcription elongation. *Science* **289,** 619–625 (2000).
25. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V. & Ebright, R. H. Structural organization of the RNA polymerase–promoter open complex. *Cell* **101,** 601–611 (2000).
26. Coulombe, B. & Burton, Z. F. DNA bending and wrapping around RNA polymerase: a "revolutionary" model describing transcriptional mechanisms. *Microbiol. Mol. Biol. Rev.* **63,** 457–478 (1999).
27. Dombroski, A. J., Walter, W. A. & Gross, C. A. Amino-terminal amino acids modulate σ-factor DNA-binding activity. *Genes Dev.* **7,** 2446–2455 (1993).
28. Siegele, D. A., Hu, J. C., Walter, W. A. & Gross, C. A. Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J. Mol. Biol.* **206,** 591–603 (1989).
29. Gardella, T., Moyle, H. & Susskind, M. M. A mutant *Escherichia coli* sigma 70 subunit of RNA polymerase with altered promoter specificity. *J. Mol. Biol.* **206,** 579–590 (1989).
30. Malhotra, A., Severinova, E. & Darst, S. A. Crystal structure of a σ70 subunit fragment from *E. coli* RNA polymerase. *Cell* **87,** 127–136 (1996).
31. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233,** 123–138 (1993).
32. Arthur, T. M., Anthony, L. C. & Burgess, R. R. Mutational analysis of $\beta'_{260-309}$, a σ^{70} binding site located on *Escherichia coli* core RNA polymerase. *J. Biol. Chem.* **275,** 23113–23119 (2000).
33. Young, B. A. *et al.* A coiled-coil from the RNA polymerase β′ subunit allosterically induces selective nontemplate strand binding by σ^{70}. *Cell* **105,** 935–944 (2001).
34. Dieci, G. *et al.* A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J.* **14,** 3766–3776 (1995).
35. Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **1,** 54–63 (1998).
36. Bown, J. A. *et al.* Organization of open complexes at *Escherichia coli* promoters. *J. Biol. Chem.* **274,** 2263–2270 (1999).
37. Owens, J. T. *et al.* Mapping the promoter DNA sites proximal to conserved regions of σ^{70} in an *Escherichia coli* RNA polymerase–*lac*UV5 open promoter complex. *Biochemistry* **37,** 7670–7675 (1998).
38. Lim, H. M., Lee, H. J., Roy, S. & Adhya, S. A "master" in base unpairing during isomerization of a promoter upon RNA polymerase binding. *Proc. Natl Acad. Sci. USA* **98,** 14849–14852 (2001).
39. Daube, S. S. & von Hippel, P. H. Interactions of *Escherichia coli* σ^{70} within the transcription elongation complex. *Proc. Natl Acad. Sci. USA* **96,** 8390–8395 (1999).
40. Vassylyeva, M. N. *et al.* Purification, crystallization and initial crystallographic analysis of RNA polymerase holoenzyme from *Thermus thermophilus*. *Acta. Crystallogr.* (submitted).
41. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
42. Yeates, T. D. Detecting and overcoming crystal twinning. *Methods Enzymol.* **276,** 344–358 (1997).
43. Brünger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
44. Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370,** 621–628 (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. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
47. 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).
48. Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).

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

Acknowledgements

We thank M. Yamamoto for assistance during the data collection at the SPring-8 synchrotron beam line, BL45. We are grateful to T. Yeates for discussions and advice concerning the merohedral twinning problem. This work was supported in part by a grant from the National Institutes of Health to S.B. and by a grant from the Organized Research Combination System of Science and Technology Agency (Japan) to 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), S.Y. (e-mail: yokoyama@biochem.s.u-tokyo.ac.jp) or S.B. (e-mail: serbor@asan.com). The atomic coordinates have been deposited in the Protein Data Bank under accession number 1IW7.