Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution

Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution

Patrick Cramer,* David A. Bushnell, Roger D. Kornberg†

Structures of a 10-subunit yeast RNA polymerase II have been derived from two crystal forms at 2.8 and 3.1 angstrom resolution. Comparison of the structures reveals a division of the polymerase into four mobile modules, including a clamp, shown previously to swing over the active center. In the 2.8 angstrom structure, the clamp is in an open state, allowing entry of straight promoter DNA for the initiation of transcription. Three loops extending from the clamp may play roles in RNA unwinding and DNA rewinding during transcription. A 2.8 angstrom difference Fourier map reveals two metal ions at the active site, one persistently bound and the other possibly exchangeable during RNA synthesis. The results also provide evidence for RNA exit in the vicinity of the carboxyl-terminal repeat domain, coupling synthesis to RNA processing by enzymes bound to this domain.

RNA polymerase II (Pol II) is responsible for all mRNA synthesis in eukaryotes. Pol II transcription is the first step in gene expression and a focal point of cell regulation. It is a target of many signal transduction pathways, and a molecular switch for cell differentiation in development.

Pol II stands at the center of a complex machinery, whose composition changes in the course of gene transcription. As many as six general transcription factors assemble with Pol II for promoter recognition and melting (1–3). A multiprotein Mediator transduces regulatory information from activators and repressors (4). Additional regulatory proteins interact with Pol II during RNA chain elongation (5, 6), as do enzymes for RNA capping, splicing, and cleavage/polyadenylation (7, 8).

Pol II comprises 12 subunits, with a total mass of >0.5 MD. A backbone model of a 10-subunit yeast Pol II (lacking two small subunits dispensable for transcription) was previously obtained by x-ray diffraction and phase determination to ~3.5 Å resolution (9). The model revealed the general architecture of the enzyme and led to proposals for interactions with DNA and RNA in a transcribing complex. DNA was suggested to enter a cleft down the middle of the enzyme, passing between a pair of mobile elements termed "jaws." Beyond the active site, marked by a Mg^{2+} ion in the floor of the cleft, the DNA path was blocked by a protein "wall." DNA-

RNA hybrid emanating from the active site would have to pass up the wall, at nearly right angles to the incoming DNA in the cleft. Both DNA and RNA would be held in place by a massive "clamp" swinging over the cleft and active-center region. A hole in the floor of the cleft beneath the active site ("pore 1") would allow entry of substrate nucleoside triphosphates and would also allow exit of RNA during retrograde movement of the polymerase on the DNA.

The Pol II backbone model thus revealed the gross structural elements of the enzyme and their likely roles in transcription. To understand how the elements perform their roles, an atomic structure is required. Here we present atomic structures determined from the previous crystal form at 3.1 Å resolution and from a new crystal form, containing the enzyme in a different conformation, at 2.8 Å resolution. The structures illuminate the transcription mechanism. They provide a basis for understanding both transcription initiation and RNA chain elongation. They permit the identification of protein features and amino acid residues crucial in the structure of an actively transcribing complex (10).

Atomic structures of Pol II. The Pol II crystals from which the previous backbone model was derived were grown and then shrunk by transfer to a solution of different composition (9). Shrinkage reduced the **a** axis of the unit cell by 11 Å and improved the diffraction from about 6.0 to 3.0 Å resolution (crystal form 1). We subsequently found that addition of Mn^{2+}, Pb^{2+}, or other metal ions induced a further shrinkage by 8 Å along the same unit cell direction and improved diffraction to 2.6 Å resolution in favorable cases (crystal form 2, Table 1) (11). Differences in Pol II conformation between form 1 and form 2, as well as atomic details most visible in

form 2, led to the conclusions reported here.

An atomic model was initially built in electron density maps from crystal form 1, for which phase information from multiple isomorphous heavy-atom derivatives was available (9). Model building was facilitated by the use of sequence markers (12), especially 94 selenomethionine residues, and maps were gradually improved by phase combination (13). The model was refined at 3.1 Å resolution by classical positional and B-factor minimization, alternating with manual rebuilding (14). The resulting structure was placed in crystal form 2 and further refined at 2.8 Å resolution to a free R factor of 28.2% (Table 1) (15). Electron density maps at that resolution revealed side-chain conformations and the orientations of backbone carbonyl groups (Fig. 1A).

Both form 1 and form 2 structures contain over 3500 amino acid residues, with more than 28,000 nonhydrogen atoms and 8 Zn^{2+} ions (Table 1). The Mg^{2+} ion in form 1 is replaced by a Mn^{2+} ion in form 2, and several additional loops, as well as 78 structural water molecules, are also seen in form 2. The stereochemical quality of the structures is high (Table 1), with 98.0% of the residues in form 2 in allowed regions of the Ramachandran plot (16), and all residues in disallowed regions located in mobile loops for which only main-chain density was observed (17). Disordered regions in the structures are limited to the COOH-terminal repeat domain (CTD) of the largest subunit, Rpb1, to the nonconserved NH_2-terminal tails of Rpb6 and Rpb12, and to several short exposed loops in Rpb1, Rpb2, and Rpb8 (18).

Over 53,000 Å^2 of surface area is buried in subunit interfaces (Fig. 1B and Table 2), about a third of it between Rpb1 and Rpb2, accounting for the high stability of Pol II (19). Many salt bridges and hydrogen bonds, and some structural water molecules, five at 2.8 Å resolution, are observed in the interfaces. There are seven instances of a "β-addition motif," in which a strand from one subunit is added to a β sheet of another. The COOH-terminal region of Rpb12, which bridges between Rpb2 and Rpb3, participates in two such β-addition motifs (Table 2). The importance of one of these motifs is shown by deletion of two residues from the COOH-terminus of Rpb12, which confers a lethal phenotype (20). Termini of Rpb10 and Rpb11 also play structural roles (21), whereas the remaining 17 subunit termini extend outwards into solvent.

For ease of display and discussion, we represent all Pol II subunits as arrays of domains or domainlike regions, named according to their locations or presumed functional roles (Figs. 2 to 5). In many cases, however, these domains and regions do not appear to be independently folded. For exam-

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305–5126, USA.

*Present address: Institute of Biochemistry, Gene Center, University of Munich, 81377 Munich, Germany.
†To whom correspondence should be addressed. E-mail: kornberg@stanford.edu

ple, the "active site" region of Rpb1 and the "hybrid-binding" region of Rpb2 combine in a single fold that forms the active center of the enzyme (Figs. 1B, 2, and 3). None of the folds in Rpb1 and Rpb2 could be found in the protein structure database and so all are evidently unique (22). Rpb3, Rpb5, and Rpb9 each consist of two independent domains, whereas the remaining small subunits form single domains (Figs. 4 and 5).

The surface charge of Pol II is almost entirely negative, except for a uniformly positively charged lining of the cleft, the active center, the wall, and a "saddle" between the clamp and the wall (Fig. 6). This strongly asymmetric charge distribution accords with previous proposals for the paths of DNA and RNA in a transcribing complex. It is also consistent with previous evidence for an electrostatic component of the polymerase-DNA interaction (23). The positively charged environment of the cleft may help to localize

DNA without restraining movement toward the active site for transcription. The positive charge on the saddle supports the proposal that it serves as an exit path for RNA (9). Homology modeling of human Pol II reveals that the overall surface charge distribution is well conserved (24).

Four mobile modules. Comparison of the form 1 and form 2 structures reveals a division of the polymerase into four mobile modules (Fig. 7 and Table 3). Half the mass of the enzyme lies in a "core" module, containing the regions of Rpb1 and Rpb2 that form the active center and subunits Rpb3, Rpb10, Rpb11, and Rpb12, which have been implicated in Pol II assembly. Three additional modules, whose positions relative to the core module change between form 1 and form 2, lie along the sides of the DNA-binding cleft, before the active center. The "jaw-lobe" module contains the "upper jaw" (9), made up of regions of

Rpb1 and Rpb9, and the "lobe" of Rpb2 (Figs. 3 and 4). The "shelf" module contains the "lower jaw" (9) (a domain of Rpb5), the "assembly" domain of Rpb5, Rpb6, and the "foot" and "cleft" regions of Rpb1 (Figs. 3 and 4). The remaining module, the "clamp," was originally identified as a mobile element in a Pol II map at 6 Å resolution (25).

The changes observed between form 1 and form 2 structures are small rotations of the jaw-lobe and shelf modules about axes roughly parallel to the cleft (perpendicular to the plane of the page in Fig. 7B), producing movements of individual amino acid residues of up to 4 Å, and a larger swinging motion of the clamp, resulting in movements of as much as 14 Å (Table 3). The mobility of the clamp is also evidenced by its high overall temperature factor (Table 4). Rotations of the jaw-lobe and shelf modules might contribute to a helical screw rotation of the DNA as it advances toward the active center.

The clamp and transcription bubble interaction. The swinging motion of the clamp produces a greater opening of the cleft in form 2 than form 1, which may permit the entry of promoter DNA for the initiation of transcription (see below). Features seen in the form 2 structure suggest that, upon closure in a transcribing complex, the clamp serves as a multifunctional element, sensing the DNA-RNA hybrid conformation and separating DNA and RNA strands at the upstream end of the transcription bubble. The unique clamp fold is formed by NH_2- and COOH-terminal regions of Rpb1 and the COOH-terminal region of Rpb2 (26). It is stabilized by three Zn^{2+} ions, two within the "clamp core" and one underlying a distinct region at the upper end, termed the "clamp head" (27). Mutations of the Zn^{2+}-coordinating cysteine residues in the clamp confer a lethal phenotype (28, 29). At its base, the clamp is connected to the "cleft" region of Rpb1, to the "anchor" region of Rpb2, and to Rpb6 through a set of "switch" regions that are flexible and enable clamp movement (Figs. 2 and 3). Whereas the shorter switches (4 and 5) are well ordered, the longer switches are poorly ordered (switches 1 and 2) or disordered (switch 3). All five switches undergo conformational changes in the transition to a transcribing complex, and switches 1, 2, and 3 contact the DNA-RNA hybrid in the active center (10). The switches therefore couple closure of the clamp to the presence of the DNA-RNA hybrid, which is key to the processivity of transcription. Interaction with the DNA-RNA hybrid may also be instrumental in the readout of the template DNA sequence in the active center (10).

Weak electron density is seen for three loops extending from the clamp that may interact with DNA and RNA upstream of the
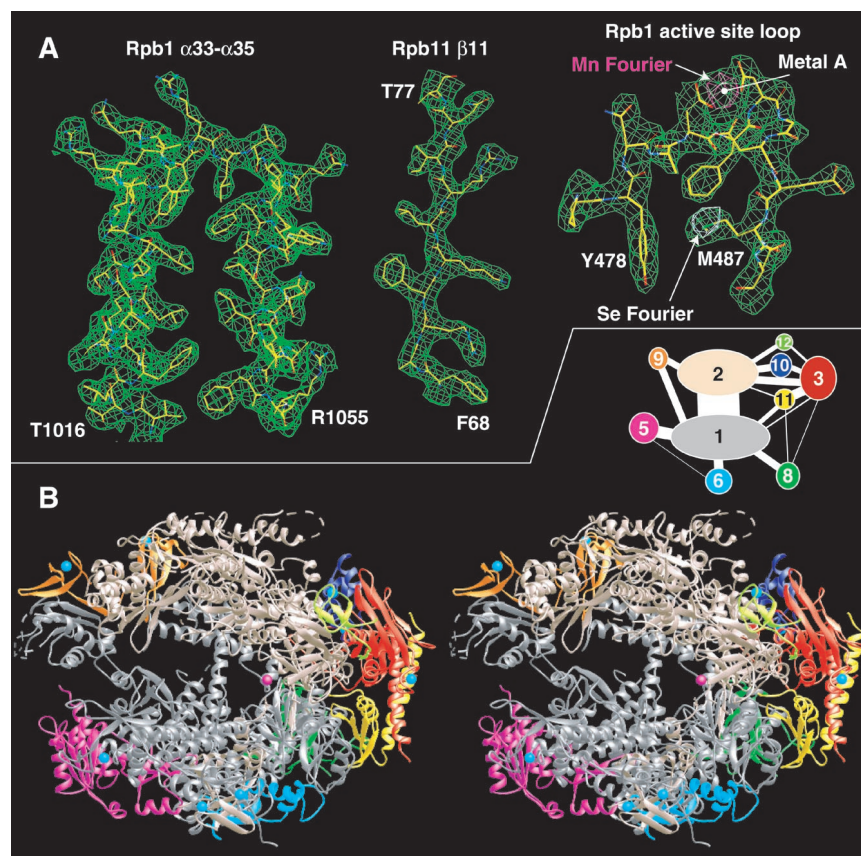

Fig. 1. Refined Pol II structure. (**A**) σ_A-weighted $2mF_{obs} - DF_{calc}$ electron density at 2.8 Å resolution (green) superimposed on the final structure in crystal form 2. Three areas of the structure are shown: the packing of α helices in the foot region of Rpb1, a β strand in Rpb11, and the active-site loop in Rpb1. Backbone carbonyl oxygens are revealed in the map. An anomalous difference Fourier of the Mn^{2+}-soaked crystal reveals the location of the active-site metal A (magenta, contoured at 10σ). An anomalous difference Fourier of a crystal of partially selenomethionine-substituted polymerase reveals the location of the S atom in residue M487 (white, contoured at 2.5σ). This figure was prepared with O (75). (**B**) Stereoview of a ribbon representation of the Pol II structure in form 2. Secondary structure was assigned by inspection. The diagram in the upper right corner is a key to the color code and an interaction diagram for the 10 subunits. The thickness of the connecting lines corresponds to the surface area buried in the corresponding subunit interface. This figure and others were prepared with RIBBONS (81).

active-center region. The loop nearest the active center corresponds to a "rudder" previously noted in the structure of bacterial RNA polymerase and suggested to participate in the separation of RNA from DNA and maintenance of the upstream end of the RNA-DNA hybrid (*30–32*). The second and third loops, here termed "lid" and "zipper" (Fig. 2D, "Clamp core, Linker," viewed in stereo), may be involved in these processes as well. Although disordered in the bacterial polymerase structure, both lid and zipper are apparently conserved (*33*). They lie 10 to 20 Å, corresponding to roughly three to six nucleotides, beyond the rudder. We speculate that the rudder and lid are involved in the separation of RNA from DNA, whereas the lid and zipper maintain the upstream end of the transcription bubble (*10*). In keeping with this idea, a region in the largest subunit of the *Escherichia coli* enzyme containing residues corresponding to the zipper has been cross-linked to the upstream end of the bubble (*32*). A disordered loop on top of the wall, termed the "flap loop" (Fig. 3) (*34*), may cooperate with the lid and zipper in the maintenance of the bubble.

Two metal ions at the active site. A Mg^{2+} ion, bound by the invariant aspartates D481, D483, and D485 of Rpb1, identifies the active site of Pol II (*9*) and is here referred to as metal A. At the corresponding position in the structure of a bacterial RNA polymerase, a metal ion was previously detected as well (*30*). The presence of only a single metal ion was unexpected, because a two-metal-ion mechanism had been proposed for all nucleic acid polymerases on the basis of x-ray studies of single-subunit enzymes (*35–40*). We now present evidence at the higher resolution of the form 2 data for a second metal ion in the Pol II active site. A difference Fourier map computed with only the protein structure and no metals contained two peaks, one at 21.0σ, owing to metal A, and a second at 4.6σ, designated metal B (Fig. 8). Peaks with comparable relative intensities were observed at the same locations in anomalous difference Fourier maps computed for the Mn^{2+}-soaked crystal. Metal B was not included in the structure because of its low occupancy.

Three observations suggest that metal B is part of the active site and that it corresponds to the second metal ion of single-subunit polymerases. (i) Metal B is in the vicinity of metal A, at a distance of 5.8 Å, compared with about 4 Å in the single-subunit polymerases (*36, 41*). (ii) Metal B is located near three invariant acidic residues—D481 in Rpb1, and E836 and

Table 1. Crystallographic data and structure statistics.

Crystal form	1	2
	*Data collection**	
Space group	I222	I222
Unit cell dimensions (Å)	130.7 by 224.8 by 369.4	122.7 by 223.0 by 376.1
Wavelength (Å)	1.283†	1.291†
Resolution (Å)	40–3.1 (3.2–3.1)‡	40–2.8 (2.9–2.8)‡
Unique reflections	98,315 (9,073)‡	125,251 (12,023)‡
Completeness (%)	99.2 (92.7)‡	99.0 (96.2)‡
Redundancy	4.7	3.6
Mosaicity (°)	0.44	0.36
R_{sym} (%)§	8.4 (29.8)‡	5.8 (34.4)‡
	Refinement	
Nonhydrogen atoms	28,173	28,379
Protein residues	3543	3559
Water molecules	0	78
Metal ions	8 Zn^{2+}, 1 Mg^{2+}	8 Zn^{2+}, 1 Mn^{2+}
Anisotropic scaling (B_{11}, B_{22}, B_{33})	−7.9, 11.3, 6.7	−14.2, 4.3, 9.9
rmsd bonds (Å)	0.008	0.007
rmsd angles (°)	1.50	1.43
Reflections in test set (%)	4,778 (4.8)	3,800 (3.0)
R_{cryst}/R_{free}‖	22.9/28.3	22.9/28.2

*Data for form 1 are from (*9*). Data collection for form 2 was carried out at 100 K as described (*9*) with an ADSC Quantum 4 charge-coupled device detector at beamline 9-2 of SSRL. Diffraction data were processed with DENZO and SCALEPACK (*79*). †Data for form 1 were collected at the Zn^{2+} anomalous peak to reveal native Zn^{2+} sites. Data for form 2 were collected below the Zn^{2+} anomalous peak energy to localize the Mn^{2+} ion at the active center. ‡Values in parentheses correspond to the highest resolution shells. §$R_{sym} = \Sigma_{i,h}|I(i,h) - \langle I(h) \rangle|/\Sigma_{i,h}|I(i,h)|$, where $\langle I(h) \rangle$ is the mean of the I observations of reflection h. R_{sym} was calculated with anomalous pairs merged; no σ cut-off was applied. ‖$R_{cryst/free} = \Sigma_h||F_{obs}(h)| - |F_{calc}(h)||/\Sigma_h|F_{obs}(h)|$. R_{cryst} and R_{free} were calculated from the working and test reflection set, respectively.

Table 2. Subunit interactions.

Subunit interface	Buried surface area (Å²)*	Salt bridges†	Hydrogen bonds‡	β-addition motifs§
Rpb1-Rpb2	17,178	6	58	Rpb2-β41-Rpb1-β7; Rpb2-β45-Rpb1-β1
Rpb1-Rpb3	608	1	3	-
Rpb1-Rpb5	4,768	5	19	-
Rpb1-Rpb6	3,797	3	12	Rpb1-β35-Rpb6-β3
Rpb1-Rpb8	3,056	3	6	Rpb8-β6-Rpb1-β18
Rpb1-Rpb9	3,011	2	21	Rpb9-β4-Rpb1-β28
Rpb1-Rpb11	1,913	-	8	-
Rpb2-Rpb3	3,070	5	26	-
Rpb2-Rpb9	2,705	1	5	-
Rpb2-Rpb10	2,941	1	11	-
Rpb2-Rpb11	608	1	2	-
Rpb2-Rpb12	1,923	4	14	Rpb12-β3-Rpb2-β32
Rpb3-Rpb8	333	1	1	-
Rpb3-Rpb10	2,175	4	15	-
Rpb3-Rpb11	3,899	4	6	-
Rpb3-Rpb12	993	3	7	Rpb12-β4-Rpb3-β3
Rpb5-Rpb6	204	1	3	-
Rpb8-Rpb11	396	-	-	-
Total	53,578	45	217	7 instances

*Calculated with programs AREAIMOL and RESAREA (*73*) with a standard probe radius of 1.4 Å. †A conservative distance cut-off of 3.6 Å was used [program CONTACT (*73*)]. ‡Potential hydrogen bonds with a donor-acceptor distance below 3.3 Å were included. §The order of strands in a β-addition motif is added β strand–accepting strand of a β sheet. Biochemical mapping (*80*) suggests that the β-addition motif formed by Rpb1 and Rpb9 may be largely responsible for the interaction of these subunits. The β-addition motif formed between Rpb1 and Rpb6 restrains clamp mobility.
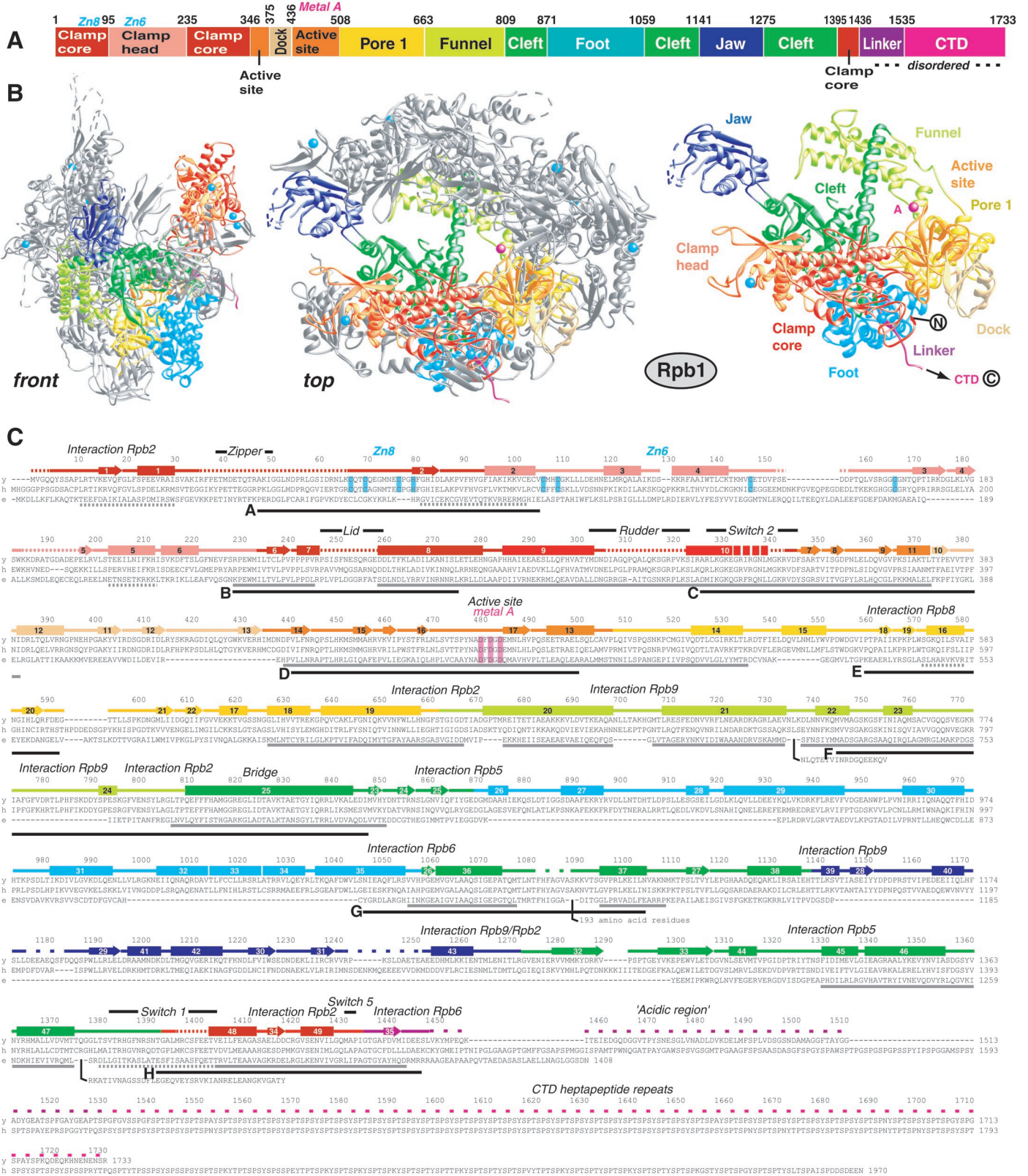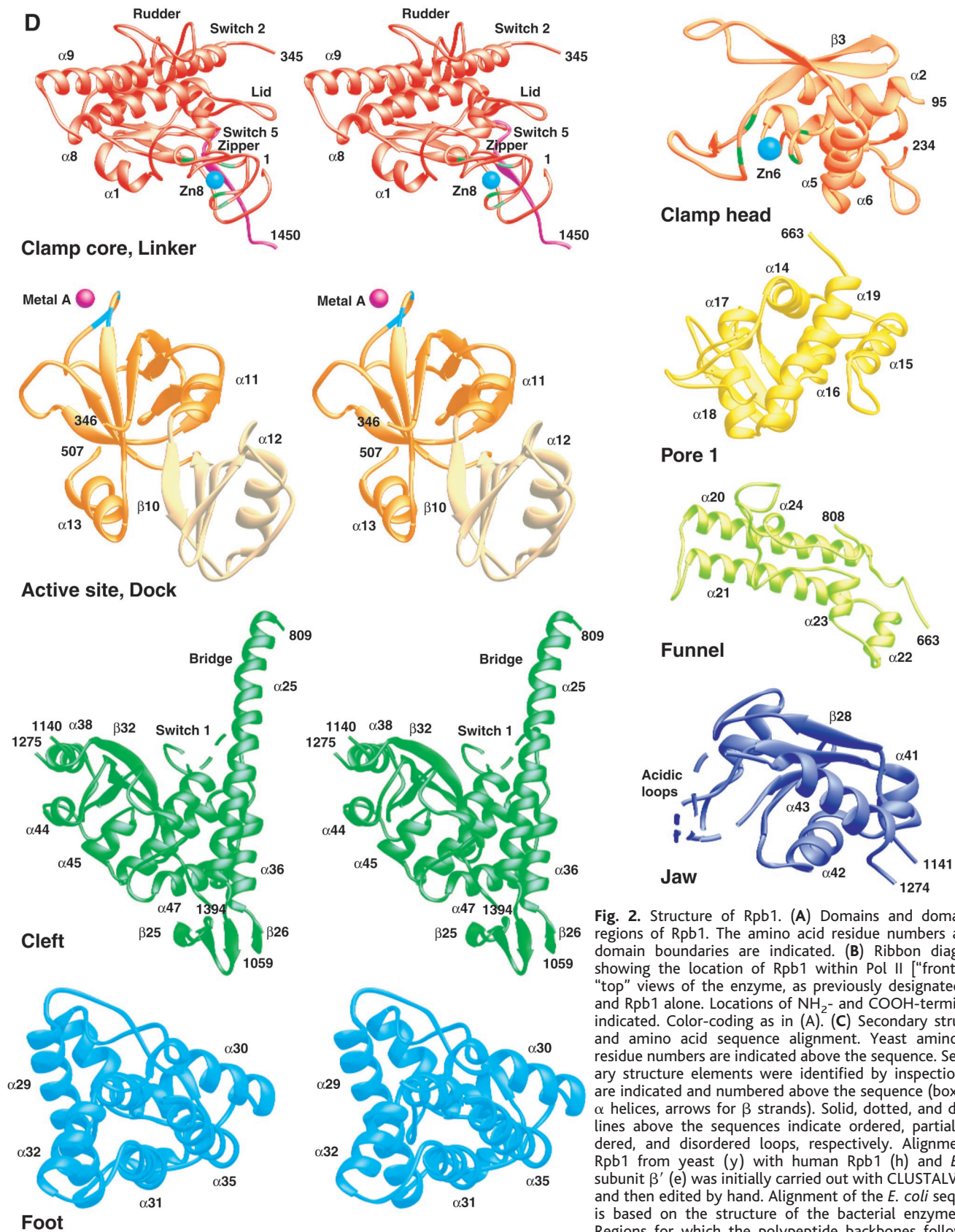
Fig. 2.

Fig. 2. Structure of Rpb1. (**A**) Domains and domainlike regions of Rpb1. The amino acid residue numbers at the domain boundaries are indicated. (**B**) Ribbon diagrams, showing the location of Rpb1 within Pol II ["front" and "top" views of the enzyme, as previously designated (9)], and Rpb1 alone. Locations of NH_2- and COOH-termini are indicated. Color-coding as in (A). (**C**) Secondary structure and amino acid sequence alignment. Yeast amino acid residue numbers are indicated above the sequence. Secondary structure elements were identified by inspection and are indicated and numbered above the sequence (boxes for α helices, arrows for β strands). Solid, dotted, and dashed lines above the sequences indicate ordered, partially ordered, and disordered loops, respectively. Alignment of Rpb1 from yeast (y) with human Rpb1 (h) and *E. coli* subunit β′ (e) was initially carried out with CLUSTALW (82) and then edited by hand. Alignment of the *E. coli* sequence is based on the structure of the bacterial enzyme (30). Regions for which the polypeptide backbones follow the same course are indicated by gray bars below the sequences (dotted when uncertain). The remaining regions could not be aligned because of disorder or because they differ in structure so that alignment is meaningless. Sequence homology blocks A to H (70) are indicated below the sequences by black bars. Important structural elements and prominent regions involved in subunit interactions are also noted. Residues involved in Zn^{2+} and Mg^{2+} coordination are highlighted in blue and pink, respectively. (**D**) Views of the domains and domainlike regions of Rpb1 (stereo on the left, mono on the right). These views reveal the entire course of the polypeptide chain from NH_2- to COOH-terminus and the locations of all secondary structure elements.

D837 in Rpb2 (Fig. 8), with aspartate D481 located between the two metals—resembling the situation in several single-subunit polymerases. The distance from metal B to the acidic residues, 3 to 4 Å, is too great for coordination, but may change during transcription (see below). (iii) The general organization of the active center resembles that of T7 RNA polymerase (42–44) and DNA polymerases of various families (36, 37, 45, 46). The two metal ions in Pol II are accessible to substrates from one side, and the Rpb1 helix bridging the cleft to Rpb2 is in about the same location relative to the metal ions as a helix in several single-subunit polymerases, generally referred to as the "O-helix."

The location of the two metals is consistent with the geometry of substrate binding inferred from structures of a Pol II transcription elongation complex (10) and of some single-subunit polymerases (35–40). In the single-subunit structures, metal A coordinates the 3′-OH group at the growing end of the RNA and the α-phosphate of the substrate nucleoside triphosphate, whereas metal B coordinates all three phosphate groups of the triphosphate. Both metals stabilize the transition state during phosphodiester bond formation. In Pol II, only metal A is persistently bound, at the upper edge of pore 1, whereas metal B, located further down in the pore, may enter with the substrate nucleotide. Ori-

entation of the nucleotide by base pairing with the template may enable complete coordination of metal B, leading to phosphodiester bond formation.

Possible structural changes during translocation. A central mystery of all processive enzyme-polymer interactions is how the enzyme translocates along the polymer between catalytic steps without dissociation. Comparison of the Pol II structure with that of bacterial RNA polymerase has given unexpected insight into this aspect of the transcription mechanism. The bridge helix, highly conserved in sequence, is straight in Pol II but bent and partially unfolded in the bacterial polymerase structure (30). The bridge
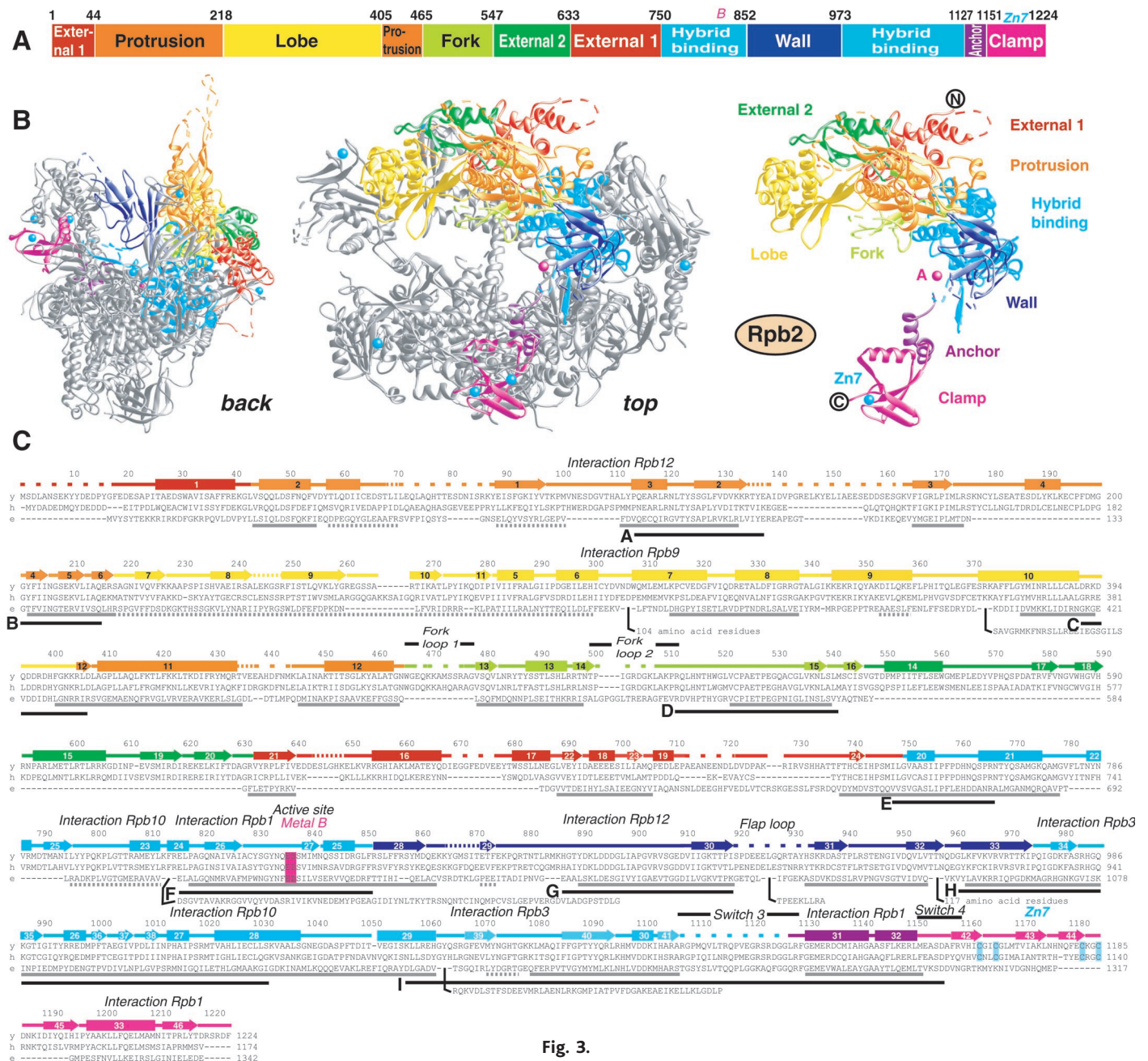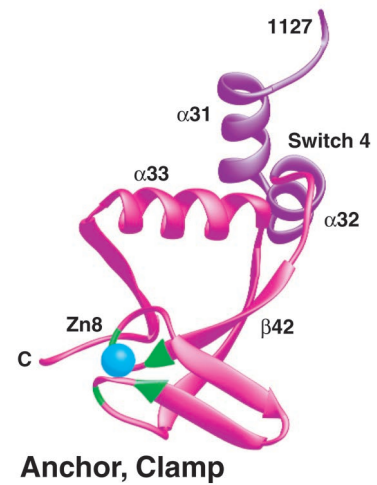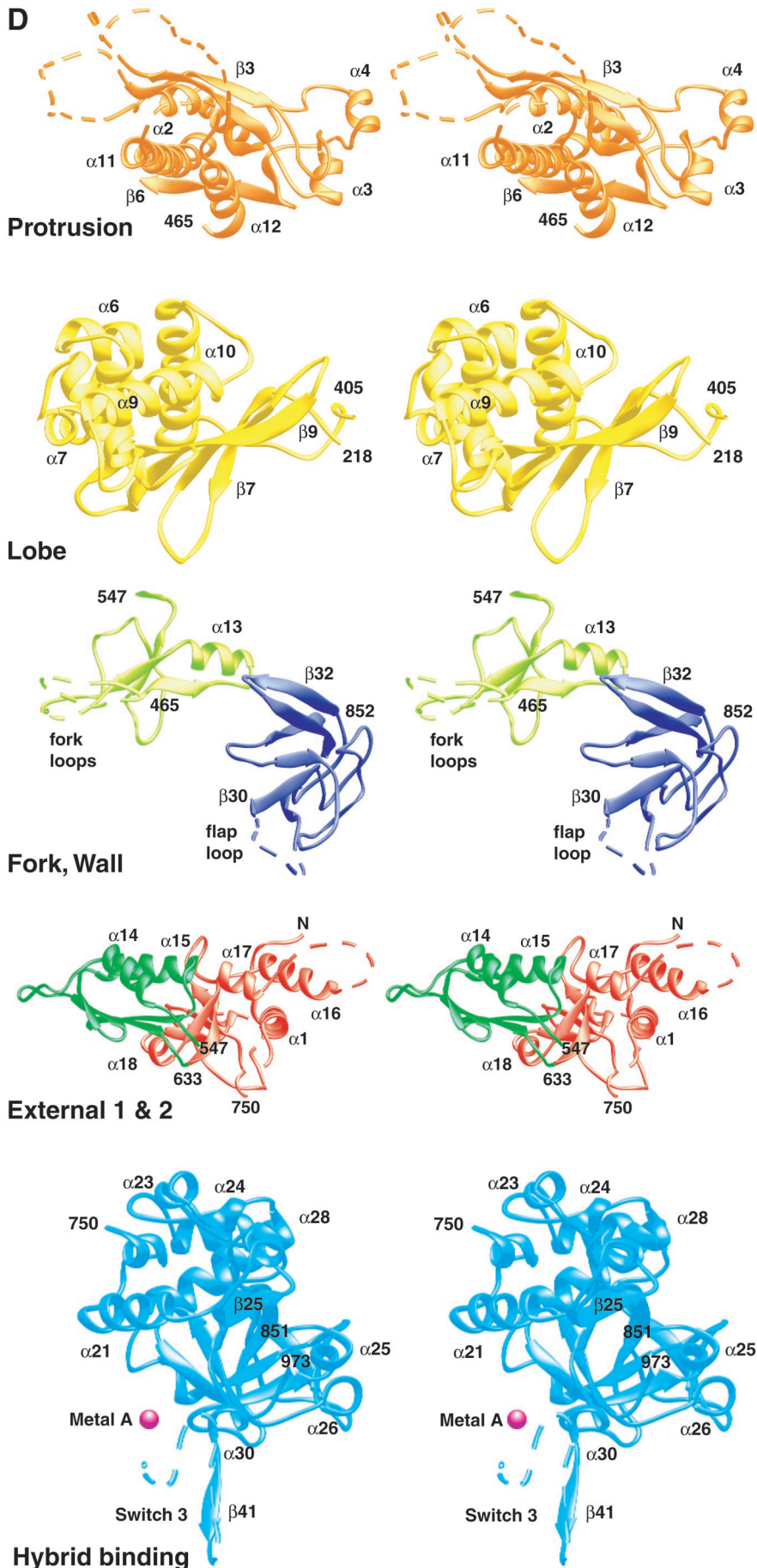

Fig. 3.

Fig. 3. (**A** to **D**) Structure of Rpb2. Organization and notation as in Fig. 2, except that the sequence alignment in (C) is with *E. coli* subunit β and its homology blocks A to I (*71*).

helix contacts the end of the DNA-RNA hybrid in a Pol II transcription elongation complex (10), and we speculate that bending of the helix is important for maintaining nucleic acid–protein interaction during translocation.

RNA exit, the CTD, and coupling of transcription to RNA processing. Two grooves in the Pol II surface were previously noted as possible paths for RNA exiting from the active-center region: "groove 1," at the

base of the clamp, and "groove 2," passing alongside the wall (Fig. 9A) (9). The atomic structure, together with a result from RNA-protein cross-linking, argue in favor of groove 1. A cross-link is formed to the NH2-terminal region of β′, the homolog of Rpb1, in an E. coli transcription elongation complex (32, 47). The corresponding residues in Rpb1 are located on the side of the clamp core above the beginning of groove 1 (Fig. 9A).

The length of RNA in groove 1 may be short, because it enters at about residue 12 and becomes accessible to nuclease digestion at about residue 18 in Pol II (48) and at about residue 15 in the bacterial enzyme (49). RNA in this part of groove 1 would lie on the saddle, beneath the Rpb1 lid and Rpb2 "flap loop." As noted above, the surface of the saddle is positively charged, appropriate for nucleic acid interaction.
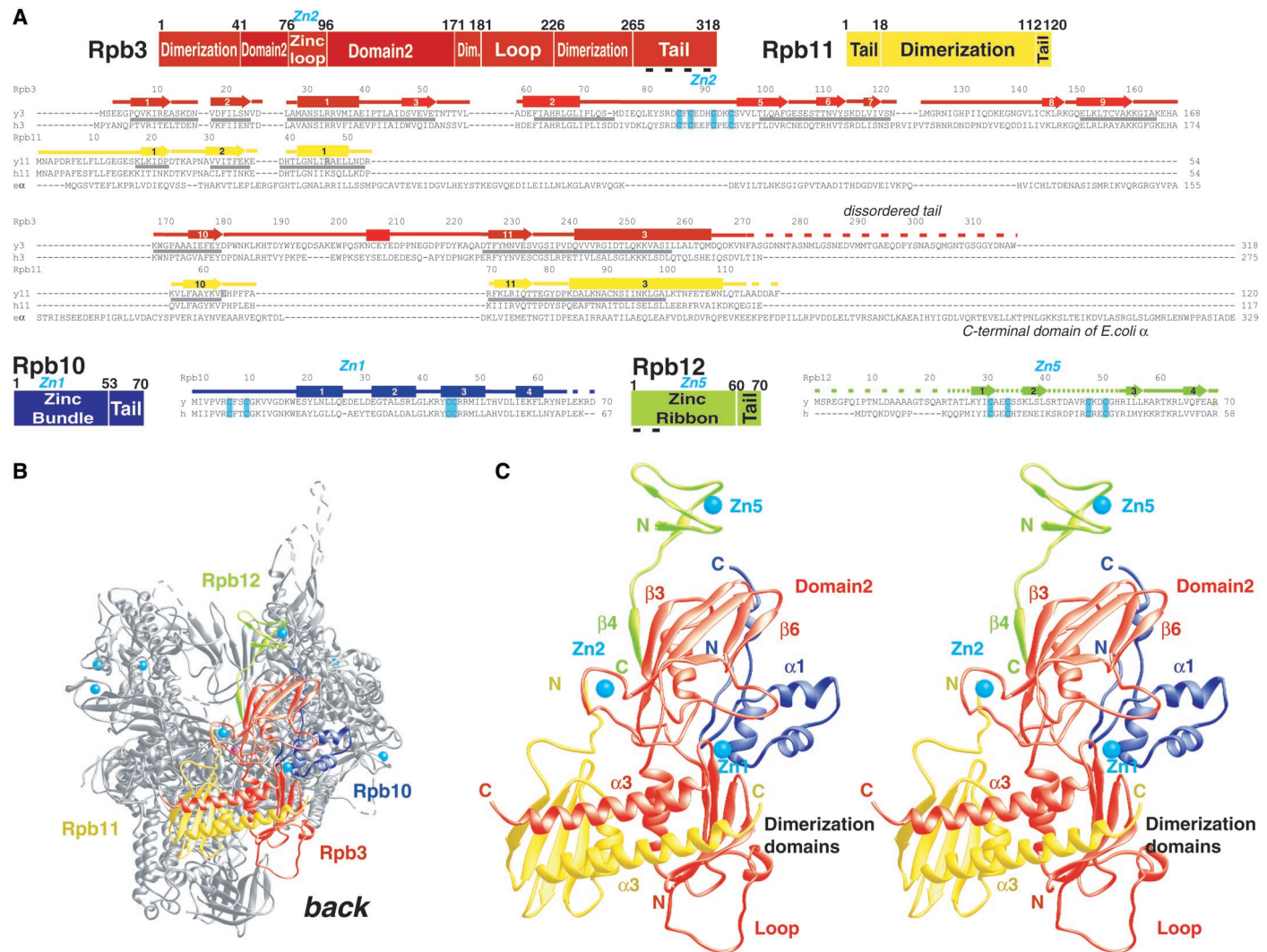

Fig. 4. Structure and location of the Rpb3/10/11/12 subassembly. (**A**) Domain structure and sequence alignments. Rpb3 and Rpb11 from yeast (y3, y11) and human (h3, h11) were aligned with E. coli subunit α (eα) on the basis of comparison with the bacterial structure (30). Regions for which the polypeptide backbones follow the same course are indicated by gray bars. Rpb10 and Rpb12 from yeast (y) were aligned with the human subunits (h). See Fig. 2 for details. (**B**) Location of the Rpb3/10/11/12 subassembly in Pol II ["back" view, of the enzyme, as previously designated (9)]. (**C**) Stereoview of the subassembly from the same direction as in (B).

Table 3. Mobile modules.

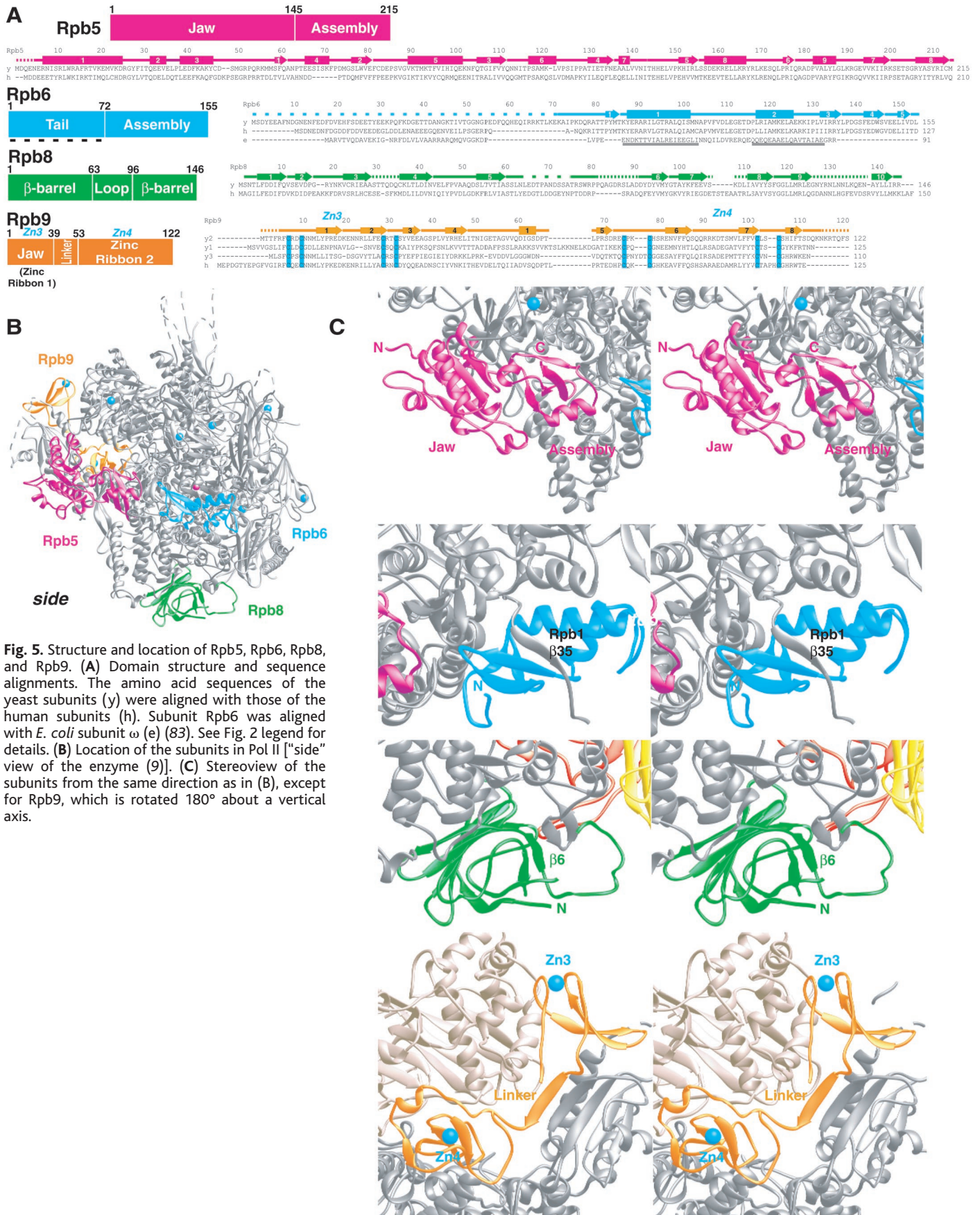| Module	Subunits and regions	Percentage of total mass	Maximum Cα atom displacement (Å) (residue number)
Core	All except other three modules	57	-
Shelf	Rpb1 cleft, Rpb1 foot, Rpb5, Rpb6	21	3.3 (N903 of Rpb1)
Clamp	Rpb1 clamp core and clamp head, Rpb2 clamp	12	14.2 (D193 of Rpb1); 14.4 (G283 of Rpb1)
Jaw-lobe	Rpb1 jaw, Rpb9 jaw, Rpb2 lobe	10	4.3 (K347 of Rpb2)

Fig. 5. Structure and location of Rpb5, Rpb6, Rpb8, and Rpb9. (**A**) Domain structure and sequence alignments. The amino acid sequences of the yeast subunits (y) were aligned with those of the human subunits (h). Subunit Rpb6 was aligned with *E. coli* subunit ω (e) (*83*). See Fig. 2 legend for details. (**B**) Location of the subunits in Pol II ["side" view of the enzyme (*9*)]. (**C**) Stereoview of the subunits from the same direction as in (B), except for Rpb9, which is rotated 180° about a vertical axis.

Soon after exiting from the polymerase, RNA must be available for processing, because capping occurs upon reaching a length of about 25 residues (*50, 51*). Consistent with this requirement, the exit from groove 1 is located near the last ordered residue of Rpb1, L1450, at the beginning of the linker to the CTD (Fig. 9B), and capping and other RNA processing enzymes interact with the phosphorylated form of the CTD (*7, 8, 52*). It may be argued that the length of the linker would allow the CTD to reach any point on the Pol II surface (Fig. 9B), and nuclear magnetic resonance (NMR) and circular dichroism studies have demonstrated a disordered state of a free, unphosphorylated CTD-derived peptide (*53*). The absence of electron density in Pol II maps owing to the linker and CTD provides evidence of motion or disorder, but even if disordered, the linker and CTD are unlikely to be in an extended conformation. The linker and CTD regions of four neighboring Pol II molecules share a space in the crystal sufficient to accommodate them only in a compact conformation (Fig. 9B).

Whereas the 5′ end of the RNA exits through groove 1 during RNA synthesis and forward movement of Pol II, the 3′ end of the RNA is extruded during retrograde movement of the enzyme. The previous backbone model suggested extrusion through pore 1 into a "funnel" on the back side of the enzyme (*9*). Transcription factor TFIIS, which provokes cleavage of extruded RNA, was thought to bind in the funnel as well (*9*). The atomic structure of Pol II lends support to these previous suggestions. A fragment of the largest bacterial polymerase subunit that can be cross-linked to the end of extruded RNA (*54*) is located in the funnel (Fig. 6). Further, Rpb1 residues that interact either physically or genetically with TFIIS (*55, 56*) cluster on the outer rim of the funnel (Fig. 6). The Gre proteins, bacterial counterparts of TFIIS, also bind to the rim of the funnel (*32*). A cluster of mutations that cause resistance to the mushroom toxin α-amanitin is located in the funnel as well (Fig. 6) (*57, 58*),

Implications for the initiation of transcription. The previous Pol II backbone model posed a problem for initiation because DNA entering the cleft and passing through the model would have to bend at the wall, whereas promoter DNA around the start site of transcription must be essentially straight (before binding to the enzyme and melting to form a transcription bubble). The only apparent solution to the problem, passage of promoter DNA over the wall, was unappealing because the DNA would be suspended over the cleft, far above the active center. A large movement of the DNA would be required for the initiation of transcription.

The form 2 structure suggests a new and more plausible solution of the initiation problem. In form 2, the clamp has swung further away from the active-center region, opening a wider gap than in form 1. A path is created for straight duplex DNA through the cleft from one side of the enzyme to the other (Fig. 10) (*59*). Following this path, the DNA contacts the jaw domain of Rpb9, fits into a concave surface of the Rpb2 lobe, and passes over the saddle, where it is surrounded by switch 2, switch 3, the rudder, and the flap loop. These surrounding elements probably do not impede entry of DNA, because they are all poorly ordered or disordered.

Fig. 6. Surface charge distribution and factor binding sites. The surface of Pol II is colored according to the electrostatic surface potential (*84*), with negative, neutral, and positive charges shown in red, white, and blue, respectively. The active site is marked by a pink sphere. The asterisk indicates the location of the conserved start of a fragment of *E. coli* RNA polymerase subunit β′ that has been cross-linked to an extruded RNA 3′ end (*54*).
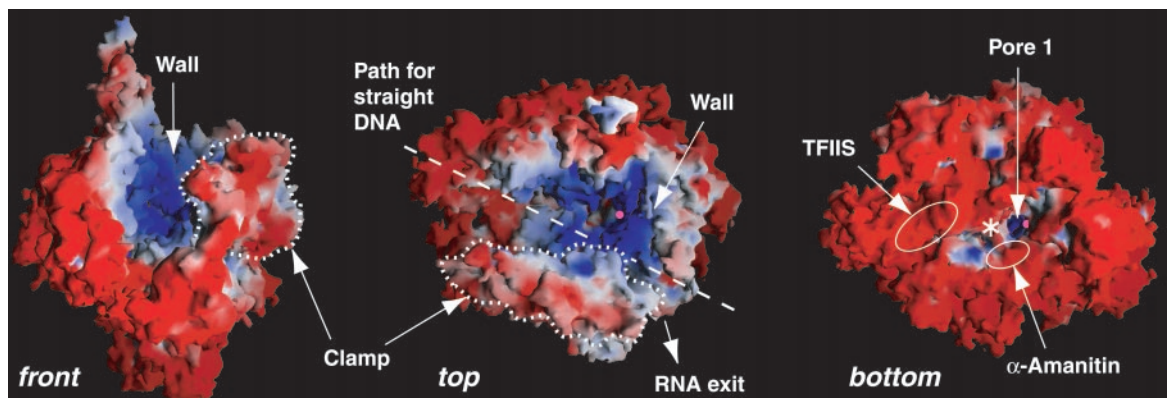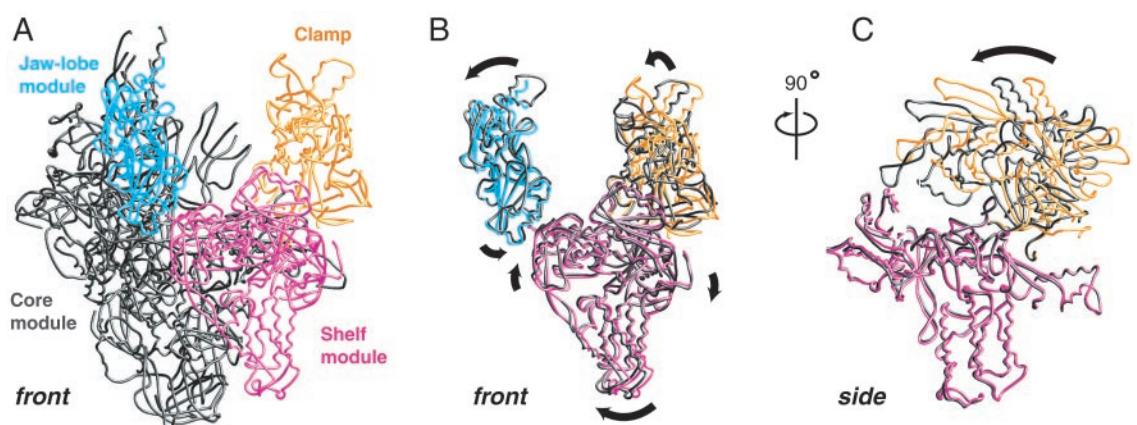

Fig. 7. Four mobile modules of the Pol II structure. (**A**) Backbone traces of the core, jaw-lobe, clamp, and shelf modules of the form 1 structure, shown in gray, blue, yellow, and pink, respectively. (**B**) Changes in the position of the jaw-lobe, clamp, and shelf modules between form 1 (colored) and form 2 structures (gray). The arrows indicate the direction of charges from form 1 to form 2. The core modules in the two crystal forms were superimposed and then omitted for clarity. (**C**) The view in (B) rotated 90° about a vertical axis. The core and jaw-lobe modules are omitted for clarity. In form 2, the clamp has swung to the left, opening a wider gap between its edge and the wall located further to the right (not shown).

Genetic evidence supports the proposed path for straight DNA during the initiation of transcription. A Pol II mutant lacking Rpb9 is defective in transcription start site selection (*60*), and complementation of the mutant with the Rpb9 jaw domain relieves the defect (*61*). Mutations in Rpb1 and Rpb2 affecting start site selection or otherwise altering initiation (*62*, *63*) lie along the proposed path as

well (Fig. 10). Some of these mutations are in residues that could contact the DNA, whereas others are in residues that may interact with general transcription factors.

Previous biochemical studies have suggested that the general transcription factor TFIIB bridges between the TATA box of the promoter and Pol II during initiation (*64*). Structural studies led to the suggestion that TFIIB brings a

TFIID-TATA box complex to a point on the Pol II surface from which the DNA can run straight to the active center (*65*). A conserved

Table 4. Crystallographic temperature factors.

Selection of model atoms	Average atomic *B* factor (Å²)	
	Crystal form 1	Crystal form 2
Rpb1	71.8	64.0
Rpb2	70.4	61.5
Rpb3	59.1	59.5
Rpb5	78.6	69.1
Rpb6	59.5	51.8
Rpb8	101.7	100.0
Rpb9	75.1	67.6
Rpb10	57.6	51.2
Rpb11	56.2	62.0
Rpb12	108.0	97.7
Clamp	113.3	81.6
Water molecules	-	39.4
Active-site metal A	58.4 (Mg^{2+})	40.7 (Mn^{2+})
Zn^{2+} ions	119.1	84.9
Overall	71.5	64.5

Fig. 8. Active center. Stereoview from the Rpb2 side toward the clamp. Two metal ions are revealed in a σ_A-weighted $mF_{obs} - DF_{calc}$ difference Fourier map (shown for metal B in green, contoured at 3.0σ) and in a Mn^{2+} anomalous difference Fourier map (shown for metal A in blue, contoured at 4.0σ). This figure was prepared with BOBSCRIPT (*85*) and MOLSCRIPT (*86*).
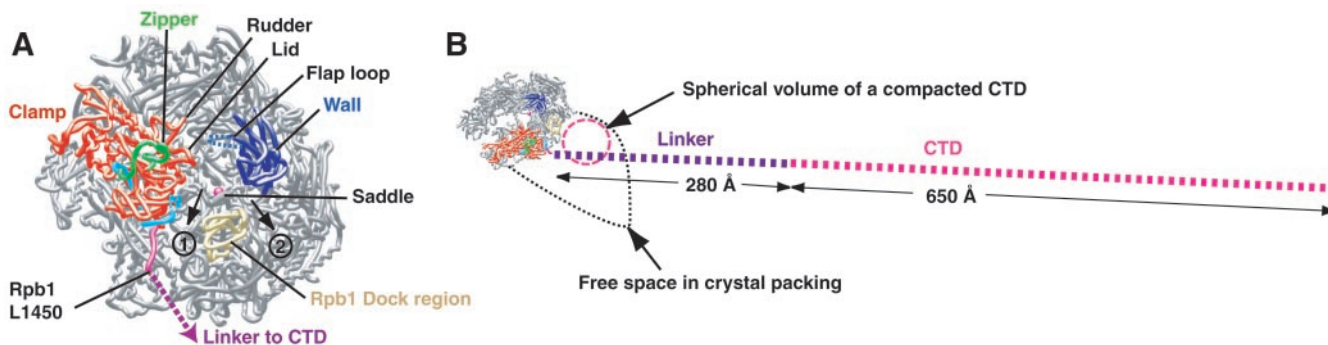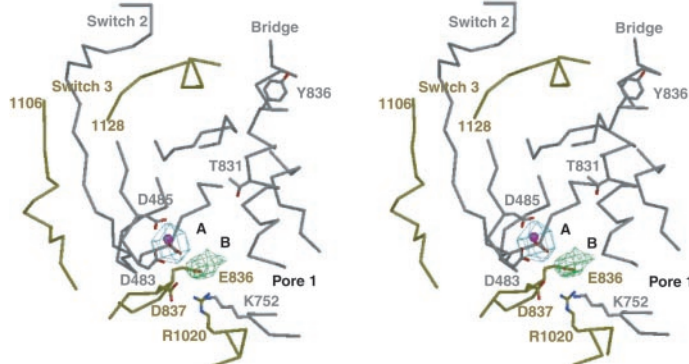

Fig. 9. RNA exit and Rpb1 COOH-terminal repeat domain (CTD). (**A**) Previously proposed RNA exit grooves 1 and 2 (*9*). The two grooves begin at the saddle between the clamp and wall and continue on either side of the Rpb1 dock region. The last ordered residue in Rpb1 (L1450) is indicated. The NH$_2$-terminal 25 residues of Rpb1 are highlighted in blue and correspond to an *E. coli* RNA polymerase fragment that was cross-linked to exiting RNA (*32, 47*). The next 30 residues of Rpb1, which form the zipper, are highlighted in green and likely mark the location of *E. coli* residues that have been cross-linked to exiting RNA (*47*) and to the upstream end of the transcription bubble (*32*). (**B**) Size and location of the CTD. The space available in the crystal lattice for the CTDs from four neighboring polymerases is indicated. The dashed line represents the length of a fully extended linker and CTD. The pink dashed circle indicates the size of a compacted random coil with the mass of the CTD.

Fig. 10. Proposed path for straight DNA in an initiation complex. (**A**) Top view. A B-DNA duplex was placed as indicated by the dashed cylinder. Rpb9 regions involved in start site selection are shown in orange. The location of mutations that affect initiation or start site selection are marked in yellow. The presumed location of general transcription factor TFIIB in a preinitiation complex is indicated by a dashed circle. (**B**) Back view. DNA may pass through the enzyme over the saddle between the wide open clamp (red) and the wall (blue). The circle corresponds in size to a B-DNA duplex viewed end-on.
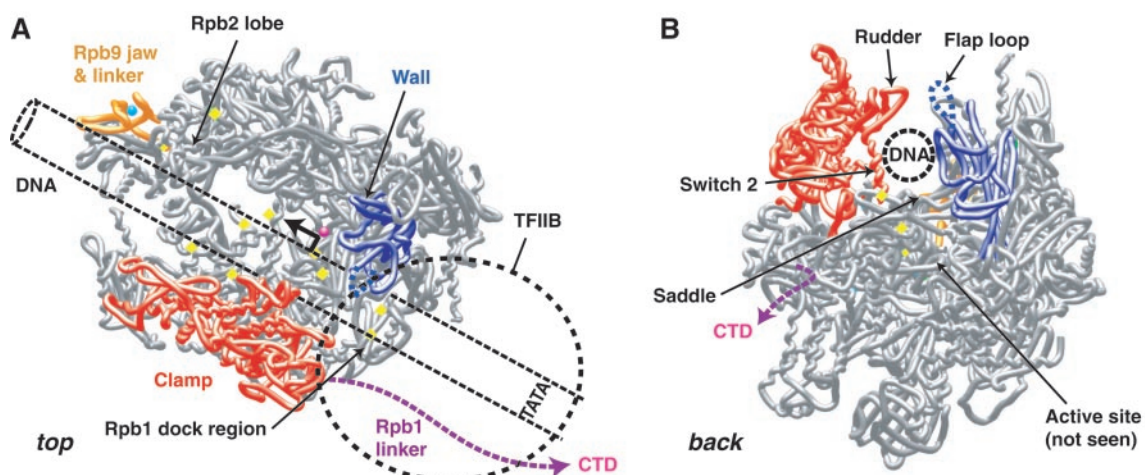
spacing of about 25 base pairs between the TATA box and transcription start site in Pol II promoters would correspond to the straight distance to the active center. This hypothesis for transcription start site determination is consistent with the path for straight DNA proposed here. There is space appropriate for a protein the size of TFIIB between a TATA box some 25 base pairs (85 Å) from the active center and the Pol II surface (Fig. 10). TFIIB in this location would contact a region of Pol II around the Rpb1 "dock" domain that is not conserved in the bacterial polymerase sequence or structure (66). Binding of TFIIB in this area would also explain its interaction with an acidic region of Rpb1 that includes the adjacent "linker" (67).

Once bound to Pol II, promoter DNA must be melted for the initiation of transcription by the adenosine 5′-triphosphate–dependent helicase activity of general transcription factor TFIIH. The region to be melted, extending from the transcription start site about half way to the TATA box, passes close to the active center and across the saddle. As the template single strand emerges, it can bind to nearby sites in the active center, on the floor of the cleft and along the wall, where it is localized in a transcribing complex (10). The

transition from duplex to melted promoter would thus be effected with minimal movement of protein and DNA. The transition would also remove duplex DNA from the saddle, clearing the way for RNA, whose exit path crosses the saddle.

Conservation of RNA polymerase structure. All 10 subunits in the Pol II structure are identical or closely homologous to subunits of RNA polymerases I and III (68). Pol II is also highly conserved across species. Yeast and human Pol II sequences exhibit 53% overall identity, and the conserved residues are distributed over the entire structure (Fig. 11A). The yeast Pol II structure is therefore applicable to all eukaryotic RNA polymerases.

Some of the amino acid differences between Pol I, Pol II, and Pol III may relate to the specificity of assembly. A complex of Rpb3, Rpb10, Rpb11, and Rpb12 anchors Rpb1 and Rpb2 in Pol II and appears to direct their assembly (19, 20, 69). Rpb10 and Rpb12 are also present in Pol I and Pol III, together with homologs of Rpb3 and Rpb11, designated AC40 and AC19. Residues that interact with the common subunits Rpb10 and Rpb12 are conserved between the three polymerases. Most residues in the interface between Rpb3 and Rpb11 differ

in the homologs, accounting for the specificity of heterodimer formation (24). Moreover, an important part of the Rpb2-Rpb3 interface (strand β10 of Rpb2 and "loop" region of Rpb3) is not conserved, which may account for the specificity of AC40 (Rpb3 homolog) interaction with the second largest subunits of Pol I and Pol III.

Sequence conservation between yeast and bacterial RNA polymerases is far less than for yeast and human enzymes. Identical residues are scattered throughout the structure (Fig. 11B). Regions of sequence homology between eukaryotic and bacterial RNA polymerases (70, 71), however, cluster around the active center (Fig. 12A). Structural homology, determined by comparison of the Pol II protein folds with the bacterial RNA polymerase structure (30), is even more extensive (Fig. 12B). Yeast Pol II evidently shares a core structure, and thus a conserved catalytic mechanism, with the bacterial enzyme, but differs entirely in peripheral and surface structure, where interactions with other proteins, such as general transcription factors and regulatory factors, take place.

Conclusions and prospects. The immediate implications of the atomic Pol II structure are for understanding the transcription mechanism. The structure has given insight into the formation of an initiation complex, the transition to a transcribing complex, the mechanism of the catalytic step in transcription, a possible structural change accompanying the translocation step, the unwinding of RNA and rewinding of DNA, and the coupling of transcription to RNA processing. No less important are the implications for future genetic and biochemical studies of all RNA polymerases. The atomic structure provides a basis for interpretation of available data and the design of experiments to test hypotheses, such as those advanced here, for the transcription mechanism. Amino acid residues of structural elements such as the bridge helix, rudder, lid, zipper, and so forth may be altered by site-directed mutagenesis to assess their roles. Homology modeling of human RNA polymerase II may enable structure-based drug design.
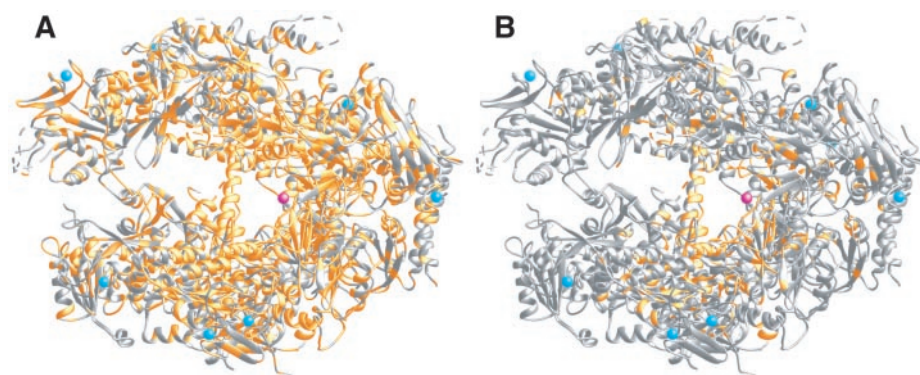

Fig. 11. Sequence identity between RNA polymerases. (**A**) Residues identical in yeast and human Pol II sequences are highlighted in orange. (**B**) Residues identical in the corresponding yeast and *E. coli* sequences are highlighted in orange.
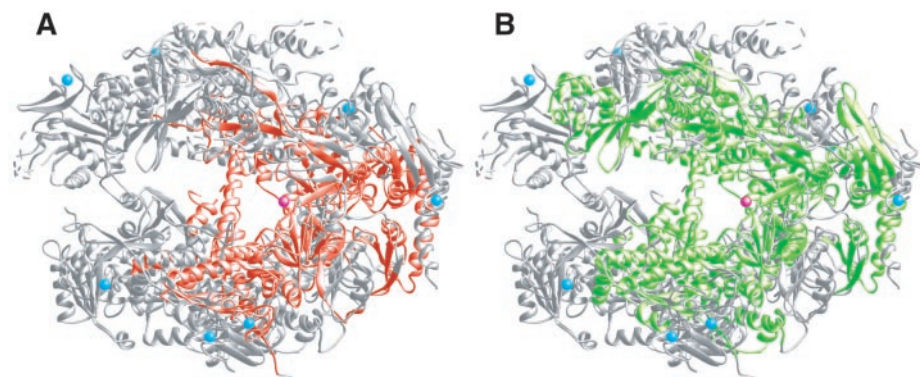

Fig. 12. A conserved RNA polymerase core structure. (**A**) Blocks of sequence homology between the two largest subunits of bacterial and eukaryotic RNA polymerases (70, 71) are in red. (**B**) Regions of structural homology between Pol II and bacterial RNA polymerase (30), as judged from a corresponding course of the polypeptide backbone, are in green.

References and Notes
1. R. G. Roeder, *Trends Biochem. Sci.* **21**, 327 (1996); R. Conaway, J. Conaway, *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 327 (1997).
2. R. D. Kornberg, *Trends Cell Biol.* **9**, M46 (1999).
3. T. I. Lee, R. A. Young, *Annu. Rev. Genet.* **34**, 77 (2000).
4. L. C. Myers, R. D. Kornberg, *Annu. Rev. Biochem.* **69**, 729 (2000).
5. J. W. Conaway, R. C. Conaway, *Annu. Rev. Biochem.* **68**, 301 (1999).
6. A. Shilatifard, *FASEB J.* **12**, 1437 (1998).
7. Y. Hirose, J. L. Manley, *Genes Dev.* **14**, 1415 (2000).
8. N. Proudfoot, *Trends Biochem. Sci.* **25**, 290 (2000).
9. P. Cramer *et al.*, *Science* **288**, 640 (2000).
10. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1876 (2001); published online 19 April 2001 (10.1126/science.1059495).
11. Addition of 1 to 10 mM Mg^{2+}, Mn^{2+}, Pb^{2+}, or lanthanide ions led to further shrinkage. The resulting form 2 crystals had a slightly lower solvent content

and lower mosaicity. Shrinkage of form 1 to form 2 results in additional crystal contacts of the mobile clamp and jaw-lobe module (see below), which may account for the improvement in diffraction.

12. A total of 141 amino acid residues were located by sequence markers. Out of 103 methionine residues in the final structure, 94 were revealed as peaks of greater than 3.3σ in a 4 Å anomalous difference Fourier map calculated with data from partially selenomethionine-substituted Pol II and with experimental multiple isomorphous replacement with anomalous scattering (MIRAS) phases (72). The few remaining methionines are located in poorly ordered regions. In the selenomethionine-substituted Pol II map, three cysteine residues, C520 and C1400 in Rpb1 and C207 in Rpb3, also showed peaks. Eight Zn^{2+} ions (9) confirmed the location of 31 cysteine residues and one histidine residue (Figs. 2 to 5). The active-site metal A is coordinated by three invariant aspartate residues in Rpb1 (Fig. 2). Two different Hg derivatives (9) revealed the location of 10 surface cysteine residues (Rpb1, C1400, C1421; Rpb2, C64, C302, C388, C533; Rpb3, C207; Rpb5, C83; Rpb8, C24, C36).

13. MIRAS phases were combined by the program SIGMAA (73) with phases from the initial polyalanine model. Phase combination was followed by solvent flattening with DM (74). This led to an electron density map at 3.1 Å resolution in which many side chains were visible. Improved maps were obtained by combination of the MIRAS phases with improved phases from combined polyalanine and atomic models in an iterative process.

14. Model building was carried out with the program O (75), and refinement, with the program CNS (76). After bulk solvent correction and anisotropic scaling, the model was subjected to positional minimization in CNS with experimental phase restraints (MLHL target). After several rounds of model building into the resulting σ_A-weighted electron density maps and subsequent further refinement, the maximum likelihood target function (MLF) was used and restrained atomic B-factor refinement was carried out. With the resulting phase-combined maps, poorly ordered regions such as parts of the clamp and the Rpb2 lobe region could be built. Extensive rebuilding and refinement of atomic positions and B factors lowered the free R factor to 29.8%. Inclusion in the form 1 structure of fine stereochemical adjustments that were achieved in refinement of the form 2 structure (15) lowered the free R factor to 28.3%.

15. The form 1 structure was manually placed with experimental Zn^{2+}-ion positions and the positon of the active-site metal in form 2. The clamp was adjusted to its new position relative to the rest of Pol II. After initial rigid body refinement of the entire polymerase in CNS, σ_A-weighted difference electron density maps revealed regions that had moved. Manual adjustment of these regions was followed by rigid body refinement in groups and positional and atomic B-factor refinement. The structure in form 2 was further confirmed with the use of sequence markers, including selenomethionine (24). After several rounds of fine adjustment of the model stereochemistry and further refinement, 78 water molecules could be included.

16. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, J. Appl. Crystallogr. 26, 283 (1993).

17. Regions showing only main-chain electron density: Rpb1, amino acids 1 to 4, 36 to 66, 154 to 157, 186 to 197, 248 to 266, 307 to 323, 330 to 338, 1388 to 1403; Rpb2, 69 to 70, 133 to 138, 241 to 251, 434 to 437, 643 to 649, 864 to 872, 915 to 919, 933 to 935, 1104 to 1110; Rpb5, 1 to 5; Rpb8, 29 to 35, 82 to 91, 107 to 113, 127 to 139; Rpb9, 1 to 4, 116 to 122; Rpb12, 24 to 53.

18. Disordered regions: Rpb1, amino acids 1082 to 1091, 1177 to 1186, 1244 to 1253, 1451 to 1733; Rpb2, 1 to 17, 71 to 88, 139 to 163, 438 to 445, 468 to 476, 503 to 508, 669 to 677, 713 to 721, 920 to 932, 1111 to 1126; Rpb3, 1 to 2, 269 to 318; Rpb6, 1 to 71; Rpb8, 1, 64 to 75; Rpb10, 66 to 70, Rpb11, 115 to 120; Rpb12, 1 to 23.

19. M. Kimura, A. Ishiguro, A. Ishihama, J. Biol. Chem. 272, 25851 (1997).

20. L. Rubbi, S. Labarre-Mariotte, S. Chedin, P. Thuriaux, J. Biol. Chem. 274, 31485 (1999).

21. The NH_2-terminal methionine of Rpb10 is inserted in a hydrophobic pocket lined by Rpb2, Rpb3, and Rpb11. The NH_2-terminus of Rpb11 binds in the previously proposed RNA exit groove 2 (9). The charge of its terminal amino group is neutralized by the conserved residue D1100 of Rpb2. The COOH-terminal residue R70 of Rpb12 is linked by a salt-bridge to the conserved residue E166 of Rpb3, whereas the charge of its carboxylate is neutralized by the conserved residue R852 of Rpb2.

22. Domains and domainlike regions of Rpb1 and Rpb2 did not produce any significant matches when submitted to the DALI server (77). The unique folds of the large subunits appear to depend on extensive contacts with small subunits on the periphery (Table 2).

23. P. H. von Hippel, D. G. Bear, W. D. Morgan, J. A. McSwiggen, Annu. Rev. Biochem. 53, 389 (1984).

24. P. Cramer, D. A. Bushnell, R. D. Kornberg, data not shown.

25. J. Fu et al., Cell 98, 799 (1999).

26. At the base of the clamp, these regions are held together in a β sheet made up of one strand from each region (Rpb1 β1, Rpb1 β34, and Rpb2 β46). Not included at the base of the clamp is the NH_2-terminal tail of Rpb6, the only change in subunit assignment of a density feature between the atomic structures and the previous backbone model. Incorporation of the Rpb6 tail in the backbone model was based on early electron density maps and the NMR structure of free Rpb6 (78). Several residues in the NH_2-terminal tail form an outer strand of a β sheet in the NMR structure. In the course of building the previous Pol II backbone model, the NMR structure was placed in the available electron density and the outer strand of the Rpb6 β sheet was extended toward the NH_2-terminus, following continuous density into the base of the clamp. The current, improved maps and sequence markers show that the continuous density near the base of the clamp instead corresponds to part of conserved region H of Rpb1, and that the NH_2-terminal tail of Rpb6 is disordered.

27. Zinc ions Zn7 and Zn8 in the clamp core are bound by residues in the common motif $CX_2CX_nCX_2C/H$ (where X is any amino acid). Zinc ion Zn6 shows an unusual coordination that underlies the clamp head fold (Fig. 2).

28. I. M. Donaldson, J. D. Friesen, J. Biol. Chem. 275, 13780 (2000).

29. M. Werner, S. Hermann-Le Denmat, I. Treich, A. Sentenac, P. Thuriaux, Mol. Cell. Biol. 12, 1087 (1992).

30. G. Zhang et al., Cell 98, 811 (1999).

31. The rudder, corresponding to Rpb1 residues 304 to 324, was not detected in early electron density maps of Pol II and so is absent from the previous backbone model of Pol II (9). Main-chain density for the rudder is clearly revealed in the improved, phase-combined electron density maps reported here.

32. N. Korzheva et al., Science 289, 619 (2000).

33. The lid and zipper are located in sequence homology blocks B and A, respectively. The lid is also flanked by regions of conserved structure.

34. The region termed the "wall" in Pol II (9) corresponds to a feature referred to as the "flap" in the bacterial RNA polymerase structure (30). The "flap loop" extending from the top of the wall, disordered in Pol II, corresponds to a loop six residues longer in E. coli that is ordered in the bacterial polymerase structure.

35. L. S. Beese, T. A. Steitz, EMBO J. 10, 25 (1991).

36. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, Nature 391, 251 (1998).

37. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, Nature 391, 304 (1998).

38. M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut, H. Pelletier, Biochemistry 36, 11205 (1997).

39. T. A. Steitz, Curr. Opin. Struct. Biol. 3, 31 (1993).

40. _____, Nature 391, 231 (1998).

41. J. F. Davies, R. J. Almassy, Z. Hostomska, R. A. Ferre, Z. Hostomsky, Cell 76, 1123 (1994).

42. G. M. Cheetham, T. A. Steitz, Science 286, 2305 (1999).

43. G. M. Cheetham, D. Jeruzalmi, T. A. Steitz, Nature 399, 80 (1999).

44. D. Jeruzalmi, T. A. Steitz, EMBO J. 17, 4101 (1998).

45. K. P. Hopfner et al., Proc. Natl. Acad. Sci. U.S.A. 96, 3600 (1999).

46. A. C. Rodriguez, H. W. Park, C. Mao, L. S. Beese, J. Mol. Biol. 299, 447 (2000).

47. E. Nudler, I. Gusarov, E. Avetissova, M. Kozlov, A. Goldfarb, Science 281, 424 (1998).

48. W. Gu, M. Wind, D. Reines, Proc. Natl. Acad. Sci. U.S.A. 93, 6935 (1996).

49. N. Komissarova, M. Kashlev, Proc. Natl. Acad. Sci. U.S.A. 95, 14699 (1998).

50. Y. Furuichi, A. J. Shatkin, Adv. Virus Res. 55, 135 (2000).

51. S. Shuman, Prog. Nucleic Acid Res. Mol. Biol. 66, 1 (2000).

52. E. J. Steinmetz, Cell 89, 491 (1997).

53. P. M. Cagas, J. L. Corden, Proteins 21, 149 (1995).

54. V. Markovtsov, A. Mustaev, A. Goldfarb, Proc. Natl. Acad. Sci. U.S.A. 93, 3221 (1996).

55. J. Archambault, F. Lacroute, A. Ruet, J. D. Friesen, Mol. Cell. Biol. 12, 4142 (1992).

56. J. Wu, D. E. Awrey, A. M. Edwards, J. Archambault, J. D. Friesen, Proc. Natl. Acad. Sci. U.S.A. 93, 11552 (1996).

57. M. S. Bartolomei, J. L. Corden, Mol. Gen. Genet. 246, 778 (1995).

58. _____, Mol. Cell. Biol. 7, 586 (1987).

59. The path for straight DNA is offset by 20° to 30° from the path of DNA entering a transcribing complex (10). Movement of DNA to this extent in the transition from an initiating to a transcribing complex seems plausible, because the DNA in this region is loosely held in the transcribing complex; the jaws, lobe, and clamp surrounding it are mobile; and a far larger movement of upstream DNA occurs upon promoter melting.

60. W. W. Hull, K. McKune, N. A. Woychik, Genes Dev. 9, 481 (1995).

61. S. A. Hemming et al., J. Biol. Chem. 275, 35506 (2000).

62. R. W. Berroteran, D. E. Ware, M. Hampsey, Mol. Cell. Biol. 14, 226 (1994).

63. D. S. Hekmatpanah, R. A. Young, Mol. Cell. Biol. 11, 5781 (1991).

64. S. Buratowski, H. Zhou, Proc. Natl. Acad. Sci. U.S.A. 90, 5633 (1993).

65. K. K. Leuther, D. A. Bushnell, R. D. Kornberg, Cell 85, 773 (1996).

66. The proposed site of interaction with TFIIB, in the vicinity of the "dock" domain, is unrelated to a site seen previously in a difference Fourier map of a two-dimensional TFIIB-Pol II cocrystal (65). The difference peak attributed to TFIIB was small and may have been misleading.

67. H. Xiao, J. D. Friesen, J. T. Lis, Mol. Cell. Biol. 14, 7507 (1994).

68. A. Sentenac et al., Yeast RNA Polymerase Subunits and Genes, Transcriptional Regulation (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992); E. P. Geiduschek, M. S. Bartlett, Nature Struct. Biol. 7, 437 (2000).

69. D. Lalo, C. Carles, A. Sentenac, P. Thuriaux, Proc. Natl. Acad. Sci. U.S.A. 90, 5524 (1993).

70. R. S. Jokerst, J. R. Weeks, W. A. Zehring, A. L. Greenleaf, Mol. Gen. Genet. 215, 266 (1989).

71. D. Sweetser, M. Nonet, R. A. Young, Proc. Natl. Acad. Sci. U.S.A. 84, 1192 (1987).

72. D. A. Bushnell, P. Cramer, R. D. Kornberg, Structure 9, R11 (2001).

73. CCP4, Acta Crystallogr. D50, 760 (1994).

74. K. Cowtan, Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr. 31, 34 (1994).

75. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, Acta Crystallogr. A47, 110 (1991).

76. A. T. Brünger, et al., Acta Crystallogr. D54, 905 (1998).

77. L. Holm, C. Sander, Trends Biochem. Sci. 20, 478 (1995).

78. F. del Rio-Portilla, A. Gaskell, D. Gilbert, J. A. Ladias, G. Wagner, Nature Struct. Biol. 6, 1039 (1999).

79. Z. Otwinowski, W. Minor, Methods Enzymol. 276, 307 (1996).

80. S. A. Hemming, A. M. Edwards, J. Biol. Chem. 275, 2288 (2000).

81. M. Carson, Methods Enzymol. 277, 493 (1997).

82. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
83. L. Minakhin et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 892 (2001).
84. A. Nicholls, K. A. Sharp, B. Honig, *Proteins* **11**, 281 (1991).
85. R. M. Esnouf, *J. Mol. Graph.* **15**, 132 (1997).
86. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
87. We thank P. Ellis, P. Kuhn, T. McPhillips, and M. Soltis for assistance at beamline 9-2 of the Stanford Synchrotron Radiation Laboratory (SSRL). This research is based in part on work done at SSRL, which is funded

Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution

Averell L. Gnatt,* Patrick Cramer,† Jianhua Fu,‡ David A. Bushnell, Roger D. Kornberg§

The crystal structure of RNA polymerase II in the act of transcription was determined at 3.3 Å resolution. Duplex DNA is seen entering the main cleft of the enzyme and unwinding before the active site. Nine base pairs of DNA-RNA hybrid extend from the active center at nearly right angles to the entering DNA, with the 3′ end of the RNA in the nucleotide addition site. The 3′ end is positioned above a pore, through which nucleotides may enter and through which RNA may be extruded during back-tracking. The 5′-most residue of the RNA is close to the point of entry to an exit groove. Changes in protein structure between the transcribing complex and free enzyme include closure of a clamp over the DNA and RNA and ordering of a series of "switches" at the base of the clamp to create a binding site complementary to the DNA-RNA hybrid. Protein–nucleic acid contacts help explain DNA and RNA strand separation, the specificity of RNA synthesis, "abortive cycling" during transcription initiation, and RNA and DNA translocation during transcription elongation.

The recent structure determination of yeast RNA polymerase II at 2.8 Å resolution and of bacterial RNA polymerase at 3.3 Å resolution has led to proposals for polymerase-DNA and -RNA interactions (*1–3*). A DNA duplex was suggested to enter a positively charged cleft between the two largest subunits and to make a right angle bend at the active center, where the DNA strands are separated and from which a DNA-RNA hybrid emerges. Avenues for entry of substrate nucleoside triphosphates and for exit of RNA could also be surmised. Although consistent with results of cross-linking experiments (*4–8*), these general proposals for polymerase–nucleic acid interaction have not been proven, and they do not address key questions about the transcription process: How is an unwound "bubble" of DNA established and maintained in the active center? Why does the enzyme initiate repeatedly, generating many short transcripts, before a transition is made to a stable elongating complex? What is the nature of the presumptive DNA-RNA hybrid duplex? How are DNA and RNA translocated across the surface of the enzyme, forward and backward, during RNA synthesis and back-tracking? Here we report the crystal structure determination of yeast RNA polymerase II in the form of an actively transcribing complex, from which answers to these questions and additional insights into the transcription mechanism are derived.

The main technical challenge of this work was the isolation and crystallization of a transcribing complex. Initiation at an RNA polymerase II promoter requires a complex set of general transcription factors and is poorly efficient in reconstituted systems (*9, 10*). Moreover, most preparations contain many inactive polymerases, and the transcribing complexes obtained would have to be purified by mild methods to preserve their integrity (*11*). The initiation problem was overcome with the use of a DNA duplex bearing a single-stranded "tail" at one 3′-end (Fig. 1A) (*12, 13*). Pol II starts transcription in the tail, two to three nucleotides from the junction with duplex DNA, with no requirement for general transcription factors. All active polymerase molecules are converted to transcribing complexes, which pause at a specific site when one of the four nucleoside triphosphates is withheld. The problem of contamination by inactive polymerases was solved by passage through a heparin column (*13*); inactive molecules were adsorbed, whereas transcribing complexes flowed through, presumably because heparin binds in the positively charged cleft of the enzyme, which is occupied by DNA and RNA in transcribing complexes. The purified complexes formed crystals diffracting anisotropically to 3.1 Å resolution (*14*).

Structure of a pol II transcribing complex. Diffraction data complete to 3.3 Å resolution were used for structure determination by molecular replacement with the 2.8 Å pol II structure (*15*). A native zinc anomalous difference Fourier map showed peaks coinciding with five of the eight zinc ions of the pol II structure, confirming the molecular replacement solution (*16*). The remaining three zinc ions were located in the clamp, a region shown previously to undergo a large conformational change between different pol II crystal forms (*3*). The locations of the three zinc ions served as a guide for manual repositioning of the clamp in the transcribing complex structure. An initial electron density map revealed nucleic acids in the vicinity of the active center. After adjustment of the protein model, the nucleic acid density improved and nine base pairs of DNA-RNA hybrid could be built (*17*). Additional density along the DNA template strand allowed another three nucleotides downstream and one nucleotide upstream to be built. Modeling of the nucleic acids assumed the 3′-end of the RNA at the biochemically defined pause site (Fig. 1A), because the nucleic acid se-

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305–5126, USA.

*Present address: Department of Pharmacology and Experimental Therapy, University of Maryland, 655 West Baltimore Street, HH403, Baltimore, MD 21201, USA.
†Present address: Institute of Biochemistry, Gene Center, University of Munich, 81377 Munich, Germany.
‡Present address: Department of Molecular Biology and Genetics, Cornell University, 223 Biotechnology Building, Ithaca, NY 14853, USA.
§To whom correspondence should be addressed. E-mail: kornberg@stanford.edu