Two complementary X-ray studies of the interaction between RNA polymerase II and nucleic acids have improved our understanding of mRNA elongation. These studies suggest how RNA polymerase II unwinds DNA, how it separates the RNA product from the DNA template and how it incorporates nucleoside triphosphate (NTP) substrates into the growing RNA chain. The tunable polymerase active center apparently allows repositioning of a catalytic metal ion, rotation of NTPs before their incorporation, RNA repositioning by a transcript cleavage factor, and modulation of enzyme activity by a bacterial small molecule regulator and its protein cofactor.

Introduction

During active transcription of protein-coding genes, RNA polymerase II (Pol II) forms a stable elongation complex (EC) (for a recent review, see [1]). In the EC, incoming (downstream) DNA is unwound before the polymerase active site and is rewound beyond it to form the exiting (upstream) duplex. In the unwound region (the transcription bubble), the DNA template strand forms a hybrid duplex with the growing mRNA. Pol II maintains the bubble, selects nucleoside triphosphates (NTPs) in a template-directed manner, synthesizes RNA, translocates along the DNA and separates RNA from DNA at the upstream end of the hybrid. Pol II also has weak RNA nuclease activity, which is stimulated by the transcript cleavage factor TFIIS [2,3]. Recent reviews of Pol II structure concentrate on TFIIS function and transcription initiation [4–7]. Here, we review advances in our understanding of Pol II elongation based on two independent structural studies of yeast Pol II ECs.

All crystal structure determinations of Pol II ECs relied on phasing with refined Pol II atomic models. Although an atomic model of the 10-subunit core of Pol II was available [8,9], only backbone models of complete 12-subunit Pol II, which comprises the core and the Rpb4/7 subcomplex, had been reported [10,11]. A new structure of free Pol II subcomplex Rpb4/7 has now enabled refinement of an atomic model of complete Pol II [12*]. This structure revealed that the Pol II core–Rpb4/7 interaction involves local folding [12*], similar to the folding transitions that occur when Pol II binds elongation factor TFIIS [13] or initiation factor TFIIIB [14**,15**].

RNA polymerase II elongation complex structures

The first Pol II EC to be analyzed crystallographically was obtained by transcription of a ‘tailed’ DNA template (duplex DNA with a single-strand extension, Figure 1c). This structure revealed that the DNA–RNA hybrid is present in the enzyme cleft, and provided important insights into elongation [9]. Experimental limitations of this system have been overcome in recent studies by reconstituting ECs with synthetic nucleic acids. One group reported X-ray models of the Pol II core bound to a hybrid with separating upstream strands [16*], and bound to the same hybrid, downstream DNA and NTPs [17**] (‘core EC’, Figure 1b). Another group determined the structure of a complete Pol II EC, with a 42-mer of DNA with a mismatch bubble, a 20-mer of RNA, and an NTP substrate analogue, both with and without the elongation factor TFIIS (‘complete EC’) [18**] (Figures 1a and 2). The new studies [17**,18**] revealed the hybrid and also three upstream nucleotides of both the DNA and RNA, as well as the downstream DNA duplex. The non-template strand in the bubble region and the upstream DNA are present in the complete EC crystals, but are not seen because of mobility [18**].

Comparison of the independently determined EC structures reveals nearly identical positions of the backbone phosphates in the hybrid and in the separating DNA and RNA strands (Figure 1). The location of downstream DNA is also very similar. Because the two studies used different crystal forms and different nucleic acid constructs and sequences, the location of the nucleic acids is apparently determined by the interaction with Pol II rather than being the result of the experimental design. All new EC structures revealed the enzyme in the post-translocation state, with an empty NTP-binding site (position +1, Figure 1), allowing complementation with an NTP substrate (see below).
Downstream DNA unwinding

The two recent structures reveal for the first time continuous electron density for the downstream DNA duplex, which runs along the cleft on the side of the largest Pol II subunit (Figure 2) [17**,18**]. The two structures differ in the point of DNA strand separation. In the complete EC, DNA base pairs are observed until position +3 before the active site, and position +2 corresponds to the point of strand separation (Figure 1a). In the core EC, DNA base pairs are observed until position +6 and the DNA strands separate at position +5 (Figure 1b). However, the DNA construct used in this study lacked...
complementary bases at positions +3 and +4 of the non-template strand; the authors predict that the point of DNA strand separation is further upstream in a natural EC [17**].

Two mechanisms could underlie downstream DNA strand separation. First, a polymerase-induced destabilization of the incoming DNA might drive duplex unwinding [18**]. Charged Pol II residues appear to induce a distortion of the duplex, which facilitates DNA strand separation. Second, the polymerase fork loop 2 sterically blocks duplex binding, interferes with the non-template strand as it approaches the active center, and prevents reassociation of separated DNA strands [9,18**].

RNA separation and exit
Separation of RNA from DNA at the upstream end of the hybrid was not observed in the first EC structure [9], probably because a persistent DNA–RNA hybrid was formed (Figure 1c). The new EC structures, however, show several nucleotides of separating DNA and RNA upstream of the hybrid. Strand separation starts at register –8 as the edges of the bases are a little too far apart for Watson–Crick pairing [16*,18**]. Thus, the hybrid contains seven base pairs in the post-translocation state and would be eight base pairs long after nucleotide incorporation (in the pre-translocation state). The exact length of the hybrid might vary slightly with the construct design. DNA–RNA strand separation involves the polymerase ‘lid’ loop, which forms a wedge between the two strands [16*,18**]. In addition, conserved Pol II residues bind the separating strands, directing them on diverging paths [18**]. Because the amino acid sequences of the lid and fork loop 2 are not well conserved among the three nuclear RNA polymerases (I, II and III), these loops lack defined interactions with nucleic acids and play a topological role [18**]. After separation from DNA, RNA can extend through a narrow exit pore (or exit tunnel) [16*,18**], formed between the lid and the ‘saddle’ [8]. RNA exits near Rpb4/7 and the linker to the Pol II C-terminal domain (CTD), which binds RNA processing factors.

Three NTP sites in the pore
The nucleotide addition cycle minimally involves the binding of an NTP to the EC, the incorporation of the nucleotide into the growing RNA by phosphodiester bond formation and the translocation of nucleic acids to free the NTP insertion site (‘addition site’ or ‘A-site’) for the next cycle. The recent structural studies not only defined the NTP insertion site but also revealed two additional overlapping NTP-binding sites, suggesting that the nucleotide addition cycle contains intermediary steps. The three observed NTP sites are located in the active center pore between the RNA 3′ end and the bridge helix (Figure 3).
An NTP bound to the insertion site when soaked into core EC crystals that contain a chain-terminating 3'-deoxy RNA end [17**]. This NTP paired with the template base at +1, bound to two metal ions in the active site and was positioned for catalysis. This NTP site had been modeled previously for bacterial RNA polymerase, based on the location of two metal ions [19**]. By contrast, a non-reactive NTP analog, which carried a methylene group between the a- and the b-phosphate groups, bound slightly too far away from the active site to support catalysis, although it did pair with the template base at +1 [18**]. This site can be described as a ‘pre-insertion site’ that is near but not identical to the insertion site. Finally, an NTP that was not complementary to the base at position +1 (‘mismatched NTP’) bound to a distinct ‘entry site’ (E-site) [17**]. The mismatched NTP was flipped with respect to the matched NTPs (so that the base points away from the template), although the triphosphate moiety was also engaged with the active site metals.

NTP selection and incorporation

The specificity of Pol II for RNA synthesis (rather than DNA synthesis) was expected to require recognition of the ribose 2'-OH group that is not present in dNTPs. Indeed, the ribose 2'-OH group of the NTP in the pre-insertion site can form a hydrogen bond with residue N479 of the largest Pol II subunit [18**], consistent with the observation that the corresponding residue in bacterial RNA polymerase is involved in the discrimination between NTPs and dNTPs [20]. The 2'-OH groups of NTPs in the A- or E-sites are, however, too far from Pol II residues to explain sugar discrimination [17**].

Considering the structures in the light of biochemical data, a possible pathway of NTP incorporation emerges. NTPs could enter through the pore, and then bind to the E-site. Indeed, NTPs bind close to the E-site in free Pol II [17**]; this interaction could help to overcome restricted NTP diffusion through the pore [21]. The NTP might then rotate into the pre-insertion site, for sampling of correct pairing with the template base and for discrimination from dNTPs [18**]. Only correctly paired NTPs can transiently bind the insertion site, leading to catalytic phosphodiester bond formation.

In an alternative model, NTPs pair with downstream template bases and enter from the Pol II cleft [22,23]. Indeed, the template base at position +2 is unpaired in the complete EC structure and could bind an incoming NTP. However, movement of a paired NTP from position +2 to +1 requires conformational changes in fork loop 2. A further possible line-up of template-paired NTPs before the active site, including at positions +3 and +4, disagrees with the structural observations [18**].

Catalysis

The active site of the free Pol II core structure revealed two metal ions: metal ion A, bound by aspartates D481, D483 and D485 of the largest Pol II subunit [18**], consistent with the observation that the corresponding residue in bacterial RNA polymerase is involved in the discrimination between NTPs and dNTPs [20]. The 2'-OH groups of NTPs in the A- or E-sites are, however, too far from Pol II residues to explain sugar discrimination [17**].

Considering the structures in the light of biochemical data, a possible pathway of NTP incorporation emerges. NTPs could enter through the pore, and then bind to the E-site. Indeed, NTPs bind close to the E-site in free Pol II [17**]; this interaction could help to overcome restricted NTP diffusion through the pore [21]. The NTP might then rotate into the pre-insertion site, for sampling of correct pairing with the template base and for discrimination from dNTPs [18**]. Only correctly paired NTPs can transiently bind the insertion site, leading to catalytic phosphodiester bond formation.

In an alternative model, NTPs pair with downstream template bases and enter from the Pol II cleft [22,23]. Indeed, the template base at position +2 is unpaired in the complete EC structure and could bind an incoming NTP. However, movement of a paired NTP from position +2 to +1 requires conformational changes in fork loop 2. A further possible line-up of template-paired NTPs before the active site, including at positions +3 and +4, disagrees with the structural observations [18**].

Catalysis

The active site of the free Pol II core structure revealed two metal ions: metal ion A, bound by aspartates D481, D483 and D485 of the largest Pol II subunit; and metal ion B, located 5–6 Å from metal A, near residues D836 and E837 of the second largest subunit [8] (Figure 3).
Metal B is mobile and is thought to enter Pol II with the NTP substrate [8]. The new core EC study supports and extends this view. The NTP binds to both metal ions within the insertion site [17**]. Whereas metal A remains at the previously detected location, metal B binds closer to metal A in the presence of NTP, interacting with D481 and D483 of the largest subunit, with D836 of the second largest subunit and with two NTP phosphates [17**]. Details of the mechanisms of catalysis or of NTP incorporation, however, remain uncertain, as the core EC and complete EC structures with matched NTPs have been determined at 4.2 Å and 4.5 Å resolution, respectively [17**,18**]. At higher resolution, structures of the free Pol II core with bound NTP and of a bacterial RNA polymerase bound to the small molecule regulator (alarmone) ppGpp have revealed two metal ions at a similar location, supporting the two metal ion mechanism [17**,19**].

Comparison with a single-subunit RNA polymerase
Comparison of the new Pol II EC structures with EC structures of the unrelated single-subunit RNA polymerase from phage T7 [24**,25**] shows that two catalytic metal ions bind in the same relative orientation, supporting a general two metal ion mechanism for all nucleic acid polymerases [26]. The comparison with T7 RNA polymerase EC structures further reveals a similar overall arrangement of DNA, RNA and NTP substrates, but also shows clear differences. The length of the DNA–RNA hybrid is the same in both polymerase ECs. There is also a similar 90° bend between axes of the incoming DNA and the exiting DNA–RNA hybrid [27]; however, the two duplex axes intersect in the Pol II EC, whereas they are offset in the T7 EC (Figure 4). Furthermore, there is a similar 90° twist between template bases at positions +2 and +1. The distance between the insertion and the pre-insertion sites is small in the Pol II EC, but much larger in the T7 EC. The E-site in Pol II apparently does not exist in the T7 enzyme. The Pol II EC structures did not reveal conformational changes upon NTP binding, but the T7 EC structures showed structural changes in the active center that accompany NTP incorporation and nucleic acid translocation (reviewed in [28,29]).

**Tunable active site and functional states**

In contrast to the T7 RNA polymerase, Pol II contains a ‘tunable’ active center [13], in which both RNA synthesis and cleavage occur. Repositioning of metal B could underlie the tunability of the active site [7,13,30]; this is supported by the comparison of structures of the Pol II core and the core EC [17**]. Pol II also contains a pore that accommodates extruded RNA during backtracking, transcriptional arrest and transcription-coupled DNA repair. The transcript cleavage factor TFIIS inserts a hairpin loop into the pore, to complement the active site with two acidic residues that might position a metal ion.
and a nucleophilic water molecule for RNA cleavage [13]. New results further revealed that TFIIS induces a realignment of the RNA strand in the active center of the complete EC, which could facilitate the in-line attack of the scissile phosphodiester bond by a nucleophilic water molecule [18**]. TFIIS-induced repositioning of the RNA is the first structural indication of distinct EC conformations that might correspond to different functional states of the enzyme [31,32].

The tunable active center and the pore (or secondary channel) are also important for the modulation of bacterial RNA polymerase function. The Gre factors are functional homologs of TFIIS and enter bacterial RNA polymerase through the pore, to complement the active site with acidic residues [33]. Recent studies show that the alarmone ppGpp also binds to a site in the pore and might increase the affinity of the active site for the catalytic metal ions [19**]. The effect of ppGpp is potentiated by DksA, a protein that also enters the pore and apparently stabilizes ppGpp binding [34,35*,36].

Conclusions
Recent studies of Pol II ECs have elucidated the nucleotide addition cycle. The enzyme has now been studied when trapped in the presumed states of NTP entry [17**], NTP pre-insertion [18**], NTP insertion [17**], pre-translocation [9] and post-translocation [16*,17**,18**]. Additional studies of bacterial and eukaryotic polymerases have provided insights into the modulation of enzyme function by accessory factors and small molecules. To obtain insights into the translocation of nucleic acids, we need to isolate Pol II in other conformational states. We also need to study the mechanisms of coupling mRNA elongation to mRNA processing, for example by analyzing how processing factors recognize the phosphorylated Pol II CTD [37**].

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest
** of outstanding interest


The authors present the detailed structure of a bacterial RNA polymerase holoenzyme bound to the alarmone ppGpp. Active site metal ions were revealed, which allowed modeling of the NTP substrate in the insertion site. Evidence for base pairing of the alarmone with cytidine bases in the non-template strand of DNA is presented.


24. Temiakov D, Patlau V, Anikin M, McAllister WT, Yokoyama S, Vassylyev DG: Structural basis for substrate selection by T7 RNA polymerase. Cell 2004, 116:381-391. The authors present a detailed structure of the bacteriophage T7 RNA polymerase EC with a bound NTP analogue. The NTP is, however, not bound to the insertion site, but rather to a pre-insertion site: this allows base pairing with the template bases and discrimination of the ribose versus the deoxyribose sugar. The authors suggest that substrate selectivity occurs before isomerization to a catalytically competent conformation.

25. Yin YW, Steitz TA: The structural mechanism of translocation and helicase function in T7 RNA polymerase. Cell 2004, 116:393-404. The authors present detailed structures of the bacteriophage T7 RNA polymerase EC with the incoming NTP substrate and with pyrophosphate. Conformational differences between the two structures prompted the authors to suggest that dissociation of the pyrophosphate product after nucleotide addition produces the conformational change that might be coupled to translocation of nucleic acids.


34. Perederina A, Svetlov V, Vassylyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassylyev DG: Regulation through the secondary channel: structural framework for ppGpp-DksA synergism during transcription. Cell 2004, 118:297-309. The authors present the high-resolution structure of the DksA protein. The structure includes a coiled coil with acidic residues at its tip and resembles the Gre factors that penetrate the pore (secondary channel) of the polymerase. Biochemical analysis suggests that DksA inserts into the pore and stabilizes the alarmone ppGpp, potentiating its function.

35. Paul BJ, Barker MM, Ross W, Schneider DA, Webb C, Foster JW, Gourse RL: DksA: a critical component of the transcription initiation machinery that potentiates the regulation of tRNA promoters by ppGpp and the initiating NTP. Cell 2004, 118:311-322. The authors demonstrate that the function of two small molecule signals that regulate bacterial transcription, ppGpp and the initiating NTP, is modulated by DksA. DksA is shown to bind to tRNA polymerase, to reduce the open complex lifetime and to inhibit the activity of the tRNA promoter.


37. Meinhart A, Cramer P: Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature 2004, 430:223-226. The authors present the high-resolution structure of the CTD-interacting domain of the 3'-RNA processing factor Pcf11 bound to a CTD peptide that is phosphorylated at a Ser2 residue. The structure suggests that this phosphorylation pattern is recognized indirectly because the phosphate is not bound by the Pcf11 domain. The authors additionally present a compact β-spiral model of the CTD.