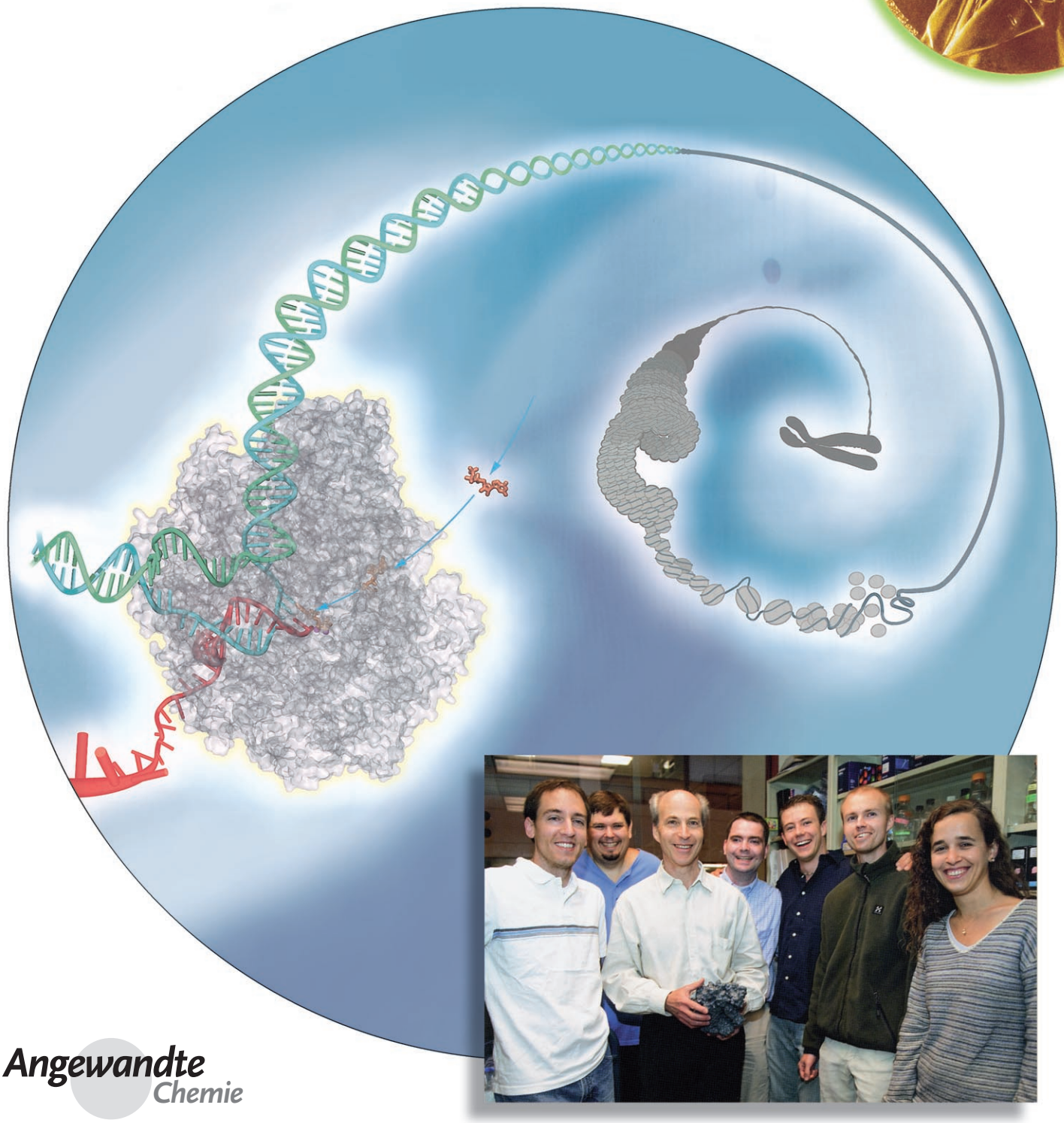
THE NOBEL PRIZE IN CHEMISTRY 2006



Angewandte Chemie

Mechanism of Gene Transcription

DOI: 10.1002/anie.200701832

The Molecular Basis of Eukaryotic Transcription (Nobel Lecture)**

*Roger Kornberg**

Keywords:
Nobel Lecture · nucleosomes · RNA polymerases · structure elucidation · transcription

I am deeply grateful for the honor bestowed on me by the Nobel Committee for Chemistry and the Royal Swedish Academy of Sciences. It is an honor I share with my collaborators. It is also recognition of the many who have contributed over the past quarter century to the study of transcription.

The Nucleosome

My own involvement in studies of transcription began with the discovery of the nucleosome, the basic unit of DNA coiling in eukaryote chromosomes.[1] X-ray studies and protein chemistry led me to propose the wrapping of DNA around a set of eight histone molecules in the nucleosome (Figure 1). Some years later, Yahli Lorch and I found this wrapping of DNA prevents the initiation of transcription in vitro.[2] Michael Grunstein and colleagues showed that nucleosomes interfere with transcription in vivo.[3] The nucleosome serves as a general gene repressor. It assures the inactivity of all the many thousands of genes in eukaryotic cells except those whose transcription is brought about by specific positive regulatory mechanisms. What are these positive regulatory mechanisms? How is repression by the nucleosome overcome for transcription? Our recent work has shown that the promoter chromatin is transformed from a static to a dynamic state upon gene activation.[4] Nucleosomes are rapidly removed and reassembled in the activated state, while promoter DNA is made transiently available for interaction with the transcription machinery.

Figure 1. The nucleosome, fundamental particle of the eukaryote chromosome. The schematic representation shows the coiling of DNA around a set of eight histones in the nucleosome, the further coiling in condensed (transcriptionally inactive) chromatin, and uncoiling for interaction with the RNA polymerase II (pol II) transcription machinery.

RNA Polymerase II Transcription

Our studies have focused on the RNA polymerase II (pol II) transcription machinery. Pol II is responsible for all messenger RNA synthesis in eukaryotes. As the first step in gene expression, pol II transcription is an end point of a great many signal-transduction pathways. The intricate regulation of pol II transcription underlies cell differentiation and development.

Since nucleosomes are removed from promoter DNA for transcription in vivo, we and others have been able to fractionate the components of the transcription machinery, guided by transcription assays performed with naked DNA in vitro. Robert Roeder and colleagues initiated the isolation

[*] Prof. Dr. R. Kornberg
Department of Structural Biology
Stanford University Medical School
299 Campus Dr. West, Stanford CA, 94305-5126 (USA)
Fax: (+1) 650-723-8464
E-mail: kornberg@stanford.edu

of pol II transcription proteins from human HeLa cell extracts.[5] This effort was brought to fruition by Ronald and Joan Conaway, who took advantage of the greater abundance of starting material available from a rat liver extract.[6] We at Stanford isolated the pol II machinery from yeast, in work begun by Neal Lue in 1987, who solved the long-standing problem of preparing an extract active in pol II transcription from yeast.[7] In retrospect, our pursuit of the problem in yeast was a fortunate choice. It proved crucial for unraveling both the structure and regulation of the pol II machinery. There were serious doubts when we began whether findings in yeast would prove relevant to human cells, but upon fractionation of yeast and mammalian systems, the results were the same.

Both systems comprise six proteins: pol II and a set of five general transcription factors known as TFIIB, -D, -E, -F, and -H.[8] Pol II is capable of unwinding DNA, synthesizing RNA, and rewinding DNA, but pol II alone is incapable of recognizing a promoter and initiating transcription. For these essential functions, the participation of the general transcription factors is required.

Mediator of Transcriptional Regulation

It was at first thought that the set of six proteins constituted a complete transcription system, and that it would support not only accurately initiated but also appropriately regulated transcription. Communication from a regulatory protein to the transcription machinery at a promoter was believed to be direct. We found, however, that an additional crude fraction was required for regulation in the yeast system (Figure 2). We referred to this activity as

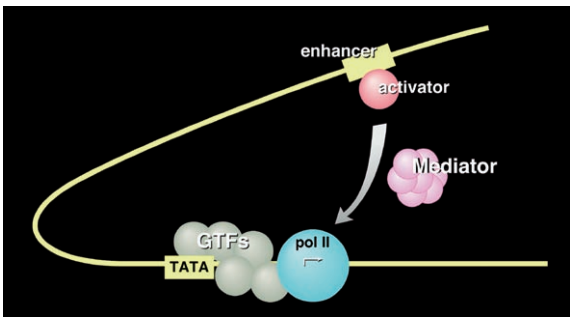


Figure 2. The Mediator of transcriptional regulation. The schematic representation shows the transduction of regulatory information from a gene activator protein bound to an enhancer DNA element to the pol II transcription machinery at a promoter.

the "Mediator",[9,10] and in 1994 Stefan Bjorklund and Young-Joon Kim isolated the active protein as an assembly of more than twenty subunits, with a total mass in excess of a million Daltons.[11] Thirteen of the subunits were products of genes previously identified in screens for molecules involved in transcriptional regulation. These were disparate screens, done on different promoters in different laboratories at different times. With the isolation of the Mediator, the products of all the screens were united in a common biochemical entity. Still the Mediator idea did not gain wide acceptance, as regulation in higher organisms was believed to be direct, through interaction of gene activator proteins with the so-called TAF subunits (TAF: TBP-associated factor, TBP: TATA binding protein) of the general transcription factor TFIID. Finally in 1998, we and others isolated mammalian counterparts of the yeast Mediator (reviewed in Refs. [12] and [13]), and TAFs were shown to be important for promoter recognition, not regulation.[14–16] As others have shown, 22 of the 25 yeast Mediator subunits have demonstrable homologues in higher cells.[17,18]

We are just beginning to fathom the complexity of the Mediator, but already three points are clear: 1) the Mediator is not only the basis for regulated transcription; it is absolutely required for almost all transcription of almost all pol II promoters.[19,20] The Mediator is no less essential for transcription than pol II itself. 2) The Mediator interacts directly with both activator proteins and with pol II.[13] It forms a tight complex with an activator at an enhancer, and it subsequently contacts pol II and the general transcription factors at the promoter to stimulate the initiation of transcription.[12] 3) The Mediator is important not only for positive but also for negative regulation of transcription.

Although the Mediator is commonly referred to as a co-activator, this is a misnomer. The Mediator is a co-activator, a co-repressor, and a general transcription factor all in one. The Mediator may be viewed as a signal-processing device. It transduces regulatory information from enhancers to promoters in the whole range of organisms from yeast to humans.

Structural Studies of the Pol II Transcription Machinery

It has been said: "if you want to understand function, study structure". The challenge in the case of the pol II transcription machinery lies in the great size of the structure. A giant complex of nearly 60 proteins, with a total mass in excess of three million Daltons, assembles at every pol II promoter prior to the initiation of transcription (Figure 3). We began with the structure determination of pol II because it forms the core of the transcription complex. In retrospect, this was another fortunate choice. It was preferable to pursuing the smaller, simpler transcription factors because pol II is the platform upon which all factors are assembled. It emerged from our work that some of the general factor proteins adopt

	Subunits	Mass (kD)
RNA polymerase II	12	515
General transcription factors	26	1560
Mediator	21	1005
Pre-initiation complex	58	3080

Figure 3. The RNA polymerase II transcription machinery. Masses are round figures for proteins from the yeast *S. cerevisiae*.

their fully folded structures only upon interaction with pol II. Knowledge of the pol II structure has proved key to understanding eukaryotic gene transcription.

Two-Dimensional Protein Crystallography

The story of the pol II structure began in my graduate work in physical chemistry, with nuclear resonance experiments revealing the rapid lateral diffusion of lipid molecules in multilayers.[21] Some years later I thought of exploiting lateral diffusion for the formation of single-layer-thick or two-dimensional (2D) protein crystals. The idea was to bind a protein to a lipid layer, through interaction with the lipid head groups. The bound protein would be constrained in two dimensions but free to diffuse in the plane and crystallize (Ref. [22] and Figure 4). Seth Darst and Al Edwards succeeded in forming 2D crystals of pol II in this way.[23] The crystals were initially small and poorly ordered. They were of little use for structure determination, but they provided a starting point. We could take advantage of the ease, rapidity, and small amount of material required by the crystallographic approach and use it as a structural assay, thus guiding the preparation of pol II that would form better crystals. We soon found that heterogeneity arising from a substoichiometric amount of two small polymerase subunits was the problem—although they accounted for only 8% of the mass of the enzyme, their presence in a variable amount was the impediment to crystallization. We turned to a deletion mutant of yeast from which we could isolate pol II lacking both small subunits. The resulting homogeneous pol II preparation formed very large, exceedingly well ordered 2D crystals.[24] Even at the low protein concentration (about 50 $\mu g\,mL^{-1}$) used for 2D crystal growth the crystals could often be seen to add additional layers in register with the first. This propensity for epitaxial growth could be exploited and the 2D crystals used to seed the formation of 3D crystals for X-ray analysis.

Figure 4. Two-dimensional protein crystallization on lipid layers. The schematic diagram shows the binding of a protein of interest (oval symbols) to the head groups (red triangles) of lipid molecules in a monolayer at the air–water interface. Rapid lateral diffusion of the lipids leads to protein crystallization.

X-Ray Crystal Structure of RNA Polymerase II

We were excited when the first 3D crystals were obtained. I also recall a chill of anxiety. The largest X-ray structure of an asymmetric particle at the time—17 years ago—was a fifth the size of pol II, and for good reason: X-ray beam intensities, detectors, and computational capabilities were all limiting factors. Actually, the limitations of diffraction technology need not have concerned us because the pol II crystals failed to diffract! The project would have ended there, were it not for Al Edwards noticing the crystals were faintly tinged with yellow. The problem was oxidation, and the solution was to grow and maintain the crystals in a glove box under argon.

Then it emerged that the crystals were profoundly polymorphic, varying by more than 10 Å along one unit cell direction. This variation, and the sheer size of pol II, posed challenges for phasing the diffraction by multiple isomorphous replacement with heavy atoms. Jianhua Fu, who discovered the best diffracting form of pol II crystals, addressed the phase problem by data collection from a large number of crystals and with the use of heavy atom clusters developed by others. He found matched pairs of native and derivative crystals from which phases to 5-Å resolution could be derived.[25] The resulting electron-density map corresponded closely to the structure of pol II at 16-Å resolution determined from 2D crystals by electron microscopy and 3D reconstruction. This marked a turning point in the solution of the pol II problem. With reliable 5-Å phases it was possible, in principle, to locate individual heavy atoms and solve the structure to near atomic resolution.

Patrick Cramer and Dave Bushnell sought additional heavy atom derivatives. When they changed the mother liquor of the crystals for the purpose, the crystals shrank by 11 Å along the previously variable unit cell direction, thereby eliminating the problem of polymorphism and extending the diffraction to 2.8-Å resolution. None of the 50 heavy atom compounds commonly used for phase determination gave useful derivatives, but an iridium compound, identified by Fu, and rhenium compounds were eventually effective. The resulting structure comprised some 3500 amino acids, with 28000 non-hydrogen atoms (Refs. [26,27] and Figure 5).

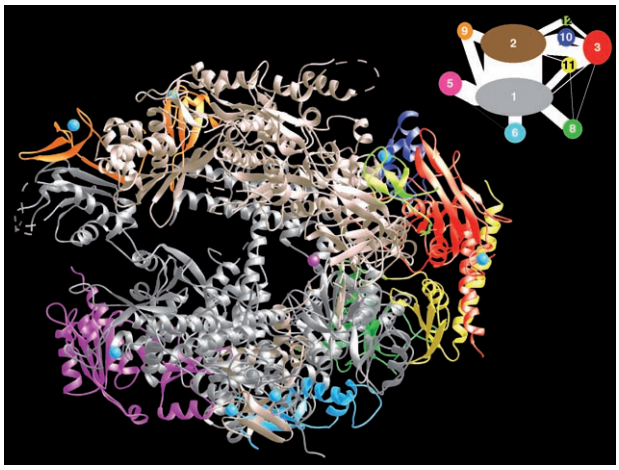


Figure 5. Structure of RNA polymerase II at 2.8-Å resolution. The protein is shown in ribbon representation, with the various subunits color coded and a Mg ion at the active center depicted as a purple sphere. The graphic in the top right corner shows an interaction diagram.

Where do DNA and RNA bind to pol II? The answer came from the structure determination of the polymerase in the form of a transcribing complex. We tried to crystallize such a complex all the while we pursued the structure of the polymerase alone. The problem was that even very highly purified pol II contains many inactive molecules and these would contaminate any preparation of transcribing complexes. Finally, Avi Gnatt discovered a way of removing the inactive molecules and succeeded in growing transcribing complex crystals.[28] Transcription had been paused by withholding one of the four nucleoside triphosphates (NTPs), and upon soaking the crystals in the missing NTP, transcription resumed without loss of crystal morphology. The crystals were very thin and gave diffraction complete to only about 6-Å resolution. After years of trying, Avi collected a dataset complete to 3.3-Å resolution, which was solved by molecular replacement with the 2.8-Å polymerase structure.[29]

DNA enters the transcribing complex in duplex form and unwinds three bases before the active site (Figure 6). Then the template strand makes a sharp bend, and as a result, the next base is flipped, thus pointing down towards the active site. This base is paired with that of the ribonucleotide just added to the RNA strand. The structure reveals eight more DNA–RNA hybrid base pairs and one additional base on the template DNA strand. The remainder of the template strand, the RNA, and the nontemplate DNA strand are not seen, as a result of motion or disorder.

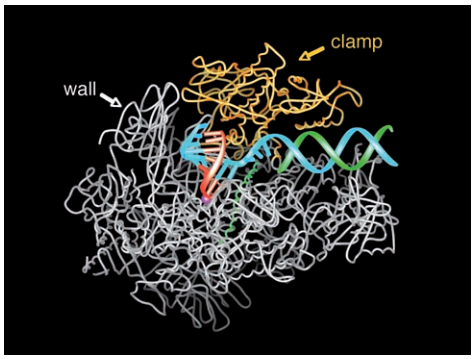


Figure 6. Structure of RNA polymerase II (at near atomic resolution) in the act of gene transcription. The polypeptide chain is shown in white, orange (mobile "clamp"), and green (bridge helix connecting the two largest subunits). Backbone models of the nucleic acids are shown in blue (template DNA strand), green (nontemplate DNA strand), and red (RNA).

Fidelity of Transcription

How does pol II select the right nucleotide for addition to the RNA chain? This is the essence of transcription, the accurate readout of the genetic code. Our most recent work has shown how accurate readout is achieved. In the original transcribing complex structure, the nucleotide just added to the RNA was still in the active center. In subsequent structures, we were able to trap the complex following translocation of the DNA and RNA across the enzyme surface, thereby creating an empty site at the active center, available for binding the next NTP (Figure 7). Soaking crystals of this "post-translocation" complex with NTPs resulted in additional electron density at two sites.[30] All four NTPs were seen to bind an entry (E) site, while only the NTP correctly matched for base pairing with the coding base in the DNA was seen to bind in the active center, at the nucleotide addition (A) site. The orientation of the NTP in the E site was inverted with respect to that in the A site, which leads to the suggestion that NTPs in the E site rotate to sample base pairing in the A site.

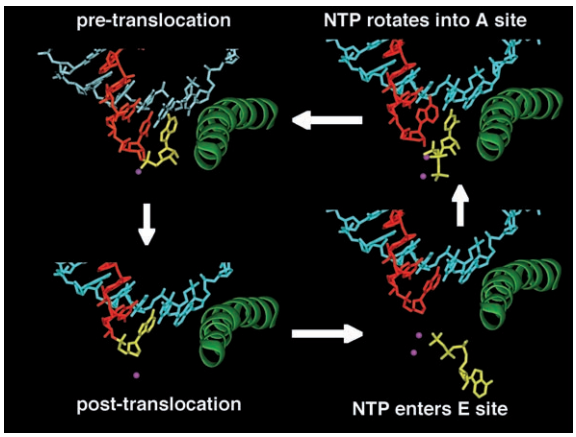


Figure 7. A cycle of nucleotide addition by RNA polymerase II. Top left: The structure from Figure 6 is shown, omitting all but the DNA (blue) and RNA (red) near the active center (purple) and the bridge helix (green). The ribonucleotide in the active center, just added to the RNA chain, is shown in yellow. Bottom left: Structure after translocation of DNA and RNA across the pol II surface. Bottom right: Structure with an unmatched NTP in the entry (E) site. Top right: Structure with an NTP, matched for pairing to the coding base in the template strand, in the addition (A) site.

These structures failed to reveal the basis for the fidelity of transcription. The energy of base pairing—through formation of two or three hydrogen bonds with the template DNA—is far less than required to account for the selectivity of the polymerase reaction. The mystery remained until this year, when the screening of many hundreds of crystals by Dong Wang and Dave Bushnell led to improved resolution and data quality. The structure of a transcribing complex with correctly matched NTP in the A site now included a feature termed the trigger loop beneath the A site (Ref. [31] and Figure 8). The trigger loop had been seen before in many pol II structures, but only in the two solved this year, with correctly matched NTP in the A site, did it appear in proximity to the A site. In all previous structures, it was located 30 Å or more distant from the A site. The trigger loop is evidently a mobile element, swinging like a trap door beneath correctly matched NTP in the A site.

The trigger loop contacts all moieties of the NTP—the base, the phosphate groups, and through other pol II residues, the sugar as well (Figure 9). The resulting network of interactions even includes the 2′-OH group of the nucleotide just added to the end of the RNA strand. The importance of

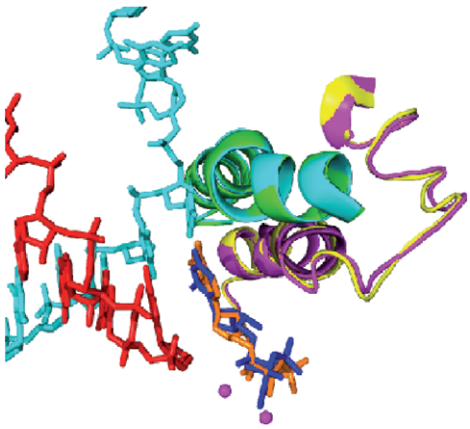


Figure 8. The trigger loop. Transcribing complex structures with purine nucleotide (orange) or pyrimidine nucleotide (dark blue) in the addition site (compare with that in the top right corner of Figure 7) are shown superimposed. The corresponding trigger loops are shown in purple and yellow, and the bridge helices are shown in green and light blue.

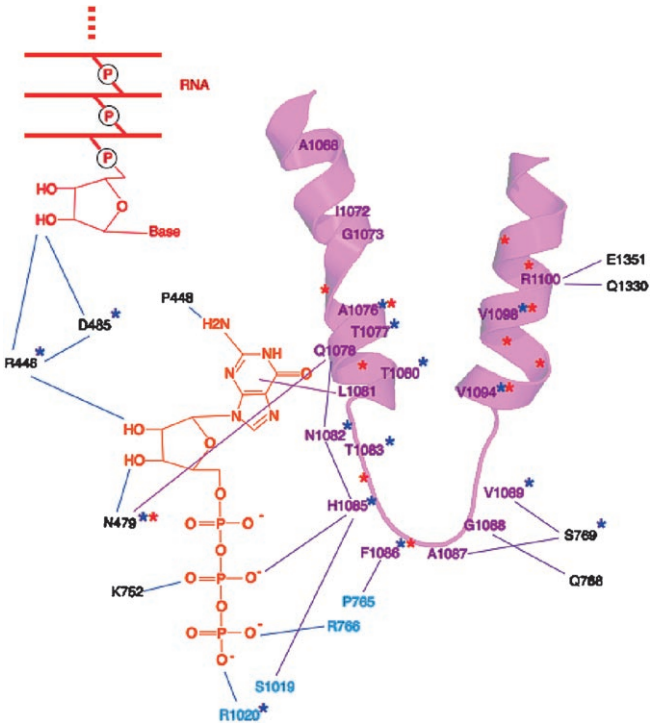


Figure 9. The trigger-loop network. Trigger loop: magenta, GTP: orange, 3′end of the RNA: red; other residues of Rpb1 and Rpb2 are indicated in black and cyan. Asterisks identify residues whose mutation impairs transcript elongation in vivo (blue from the literature, red from an unpublished screen by Craig Kaplan).

these interactions is shown by mutations in the trigger loop affecting transcription. For example, mutation of asparagine 479, which is hydrogen bonded to the 3′-OH group of the NTP, results in a loss of discrimination between the normal nucleotide and one lacking a 3′-OH group. The discrimination is not great, only about tenfold, of the magnitude expected for the energy of a single hydrogen bond. In contrast, discrimination between a normal ribo NTP and a 2′-deoxy NTP is very great, at least 1000-fold, and is unaffected by mutation of asparagine 479. How is such extraordinary specificity for a single OH group achieved? The answer lies in the alignment of the trigger loop with the NTP and the precise positioning of a histidine side chain 3.5 Å from the β-phosphate group (Figure 10). The histidine promotes the flow of electrons

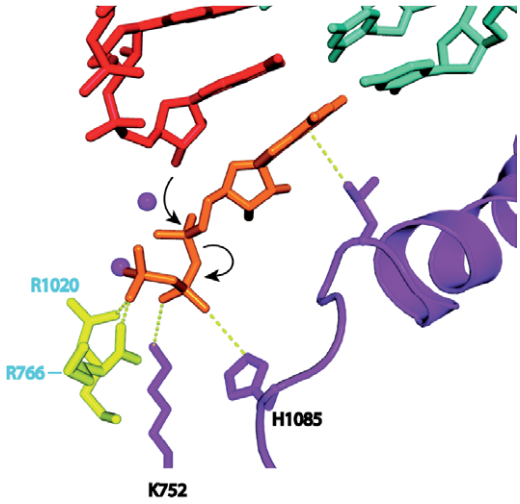


Figure 10. The trigger loop couples NTP recognition to phosphodiester bond formation. Color scheme as in Figure 8, with side chains of Rpb2 Arg1020 and Rpb2 Arg766 shown in yellow. Contacts responsible for alignment, and for the contact of histidine 1085 with the NTP that promotes catalysis, are indicated by dashed lines. Nucleophilic attack and phosphoanhydride bond breakage are indicated by arrows.

during nucleophilic attack of the 3′-OH group at the chain terminus and breakage of the phosphoanhydride bond. It serves as a proton donor for the pyrophosphate leaving group. It literally triggers formation of a phosphodiester bond and couples nucleotide selection to catalysis.

The electronic transactions involved in trigger-loop function require precise alignment of the interacting moieties. This is achieved for a correct NTP by formation of the trigger-loop network. In the case of an incorrect NTP—for example, a 2′-deoxy NTP—misalignment is profound. A double helix formed with a 2′-deoxy nucleotide is 2 Å narrower than that formed by a ribonucleotide. The resulting misalignment with the catalytic histidine residue is as great as in the case of a pyrimidine–pyrimidine base mismatch, thus leading to a thousandfold reduction in the rate of phosphodiester bond formation.

Nucleic Acid Translocation

The significance of the trigger-loop network extends beyond nucleotide selection and catalysis. The network includes many contacts with the bridge helix which, in turn, contacts the coding base in the template DNA strand. The structure of bacterial polymerase, determined by Seth Darst, also includes a bridge helix, but in contrast with that in pol II, the bridge helix in the bacterial enzyme is bent (Ref. [32] and Figure 11). The bend produces a movement of about 3 Å in

Figure 11. Straight and bent states of the bridge helix in RNA polymerase II and bacterial RNA polymerase structures, proposed to underlie nucleic acid translocation during transcription. The color code is as in Figure 6 except that the bridge helix is shown in purple.

the direction of the template strand, which corresponds to a step of one base pair along the strand. This led us to suggest that transitions of the bridge helix between straight and bent states underlie the translocation step in transcription. The bridge helix may serve as a kind of molecular ratchet that allows the polymerase to let go of the DNA and RNA for translocation, while retaining a grip on the growing end of the DNA–RNA hybrid helix to preserve the register of transcription. There is now a good deal of biochemical and genetic evidence in support of this idea.

RNA Release

In the final step of transcription, the RNA is released. The question arises as to how the RNA is peeled off the template DNA. How is the very stable RNA–DNA hybrid helix disrupted and the RNA discharged into solution? Our original transcribing complex structure gave no indication of the mechanism of this important process. A subsequent structure, determined by Ken Westover, revealed RNA release taking place.[33] Base pair 7 of the DNA–RNA hybrid in this structure appears normal—the bases are coplanar, with a distance appropriate for hydrogen bonding between them (Figure 12). Base pairs 8, 9, and 10, however, show increasing deviations, and consequent splaying apart of the DNA and RNA strands. The strand separation is due to

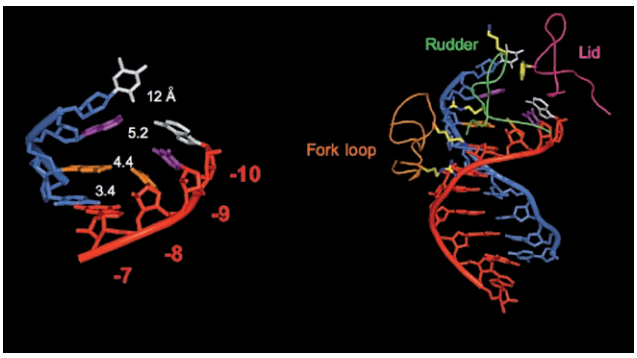


Figure 12. Release of RNA transcript from the DNA-RNA hybrid revealed in the structure of an RNA polymerase II transcribing complex. Left: The upstream end of the DNA–RNA hybrid helix, 7–10 residues from the active center; distances between the DNA and RNA bases are indicated. Right: The entire DNA–RNA hybrid helix, along with protein loops involved in helix melting (rudder and lid) and stabilization (fork loop).

the intervention of three protein loops, termed fork loop 1, rudder, and lid (Figure 12). These loops were disordered in all previous pol II structures. The rudder and lid lie between the DNA and RNA strands, with the rudder contacting the DNA, and the lid the RNA. A phenylalanine side chain of the lid serves as an actual wedge to maintain separation of the strands. Fork loop 1 contacts the sugar–phosphate backbone of the hybrid helix at base pairs 6 and 7, thereby stabilizing the helix, limiting strand separation to positions 8 and beyond, preventing the DNA–RNA hybrid from unraveling further, and inhibiting transcription.

The Pol II–TFIIB Complex

It may be asked how the transcribing complex is formed in the first place. How is straight duplex promoter DNA melted, bent, and inserted in the pol II active center, to enable the initiation of transcription? These DNA transactions are brought about by the general transcription factors TFIIB, -D, -E, -F, and -H. Our solution of the X-ray structures of pol II–TFIIB and pol II–TFIIF complexes has shed light on the initiation mechanism.

The structure of the pol II–TFIIB complex[34] revealed distinct functions of the N- and C-terminal domains of TFIIB. A polypeptide chain trace of the N-terminal domain (yellow in Figure 13) begins with a Zn ribbon that binds the pol II

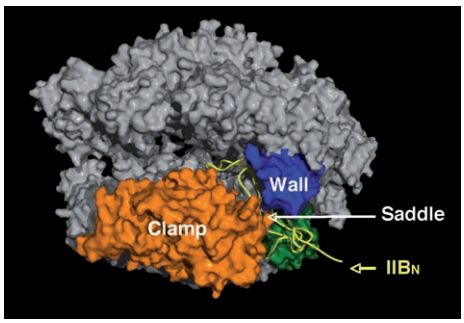


Figure 13. Structure of an RNA polymerase II–TFIIB complex. A surface representation of pol II is shown, with the clamp and wall as in Figure 6, a polypeptide chain trace of the amino terminal region of TFIIB (IIB_N), shown in yellow, and the region of the pol II surface interacting with IIB_N in green.

surface adjacent to the clamp and wall. Then the polypeptide does an amazing thing: rather than pass back into solution, it continues across the saddle between the clamp and wall and plunges towards the active center, from which it loops back and remerges across the saddle. The loop, which we have termed the B finger, occupies almost the same location as the DNA–RNA hybrid in a transcribing complex. Superimposing the B finger on the DNA–RNA hybrid from the transcribing complex structure reveals no interference with the template DNA strand or with the RNA up to position 5, but a steric clash with the RNA at positions 6 and beyond (Figure 14).

Biochemical experiments show the B finger is not only compatible with a hybrid containing five residues of RNA, but is required for the stability of such a complex. When the

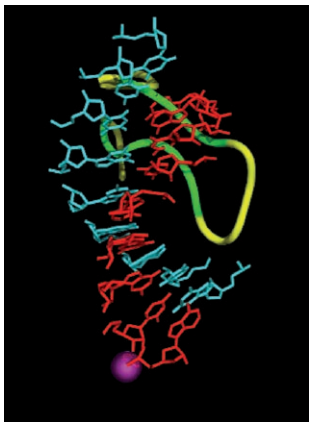


Figure 14. Superposition of the DNA–hybrid helix from an RNA polymerase II transcribing complex (Figure 6) and the B finger from an RNA polymerase II–TFIIB complex (Figure 13). The conserved region of the B finger is shown in green.

RNA grows beyond five or six residues, however, it must compete with TFIIB for space on the pol II saddle. If TFIIB wins the competition, initiation is aborted and must be tried again. If the RNA wins, TFIIB is ejected and pol II is released from the promoter to continue and complete transcription. The B finger thus explains two crucial but until now mysterious aspects of pol II transcription: abortive initiation and promoter escape. In these respects it resembles the sigma factor in bacterial transcription.[35,36]

Turning to the C-terminal domain of TFIIB, its location in the structure of the pol II–TFIIB complex served as a guide for docking a structure determined long ago[37] of a ternary complex of a C-terminal TFIIB fragment, the TATA-binding protein (TBP) subunit of TFIID, and a TATA-box DNA fragment (Figure 15). This modeling exercise led to a moment of truth. The TATA-box DNA fragment is sharply bent by TBP. What would happen if the ends of the bent fragment were extended with straight B-form DNA? The result was remarkable in two respects. First, the DNA fit snugly against the protein (Figure 15, left)—the TBP evidently configures the promoter DNA to the contours of the pol II surface. Second, the DNA downstream of the TATA box ran past the saddle (Figure 15, right). The distance from the TATA box to the saddle is about 1.5 turns of the double helix, or 15 base pairs. We know from the transcribing complex structure that about 12 residues are required to cross the saddle to the active site. The sum of 15 base pairs from the TATA box and 12 residues to the active site is 27 base pairs, closely coincident with the spacing of 25–30 base pairs from the TATA box to the transcription start site of almost all pol II promoters. In this way, the pol II–TFIIB interaction may determine the location of the transcription start site.

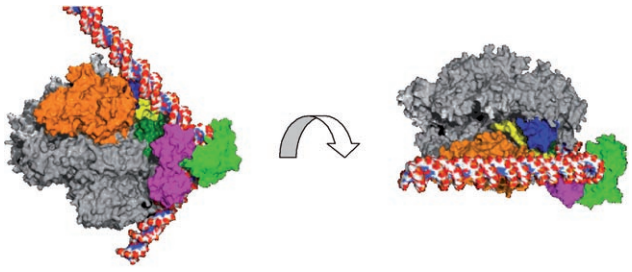


Figure 15. Model of an RNA polymerase II–TBP–TFIIB–DNA complex. The structure of the C-terminal region of TFIIB (purple) complexed with TBP (green) and TATA box containing DNA (red, white, blue) was docked to the structure of the pol II–TFIIB complex (as shown in Figure 13). The views on the left and right are related by a 90° rotation about an axis between them, as indicated by the curved arrow. The direction of view in the right model is the same as that in Figure 13.

The Pol II–TFIIF Complex

A cocrystal structure of a transcribing complex with the central subunit of TFIIF is also informative about the initiation of transcription. This work, which is still in progress, of Guillermo Calero was made possible by his mastery of the biochemical behavior of TFIIF. The structure includes a complete transcription bubble—not only the template DNA strand with associated RNA seen in previous structures, but also the nontemplate DNA strand and the region upstream where duplex DNA is reformed following transcription (Figure 16). The nontemplate strand and upstream duplex DNA were revealed because of their interaction with TFIIF, which constrains their mobility. This interaction of the nontemplate strand with TFIIF may trap a transient bubble in the promoter DNA, thereby leading to the initiation of transcription.

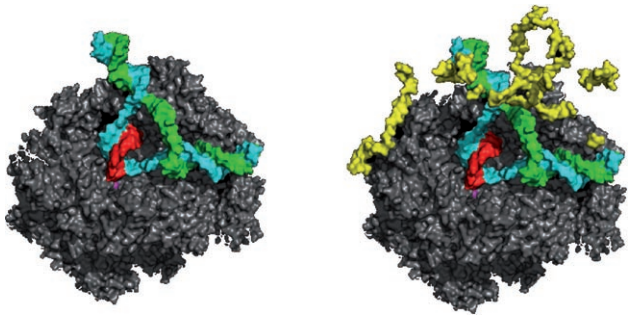


Figure 16. Structure of an RNA polymerase II transcribing complex with the central subunit of TFIIF (work in progress by Guillermo Calero). Left: Structure of pol II and the nucleic acids; right: structure including the TFIIF subunit (known as Tfg2, in yellow). The direction of view and color scheme as in Figure 6.

Pol II Transcription Initiation Complex

The available structural information may be assembled in a preliminary picture of the pol II transcription initiation complex (Figure 17). The structures of pol II, TBP, and TFIIB come from X-ray crystallographic analysis, as I've described above. The structures of TFIIE, TFIIF, and TFIIH are from electron crystallography, and from cryoelectron microscopy and single particle analysis.

The picture of the complete complex provides a solution, in outline, of the transcription initiation problem. Each of the

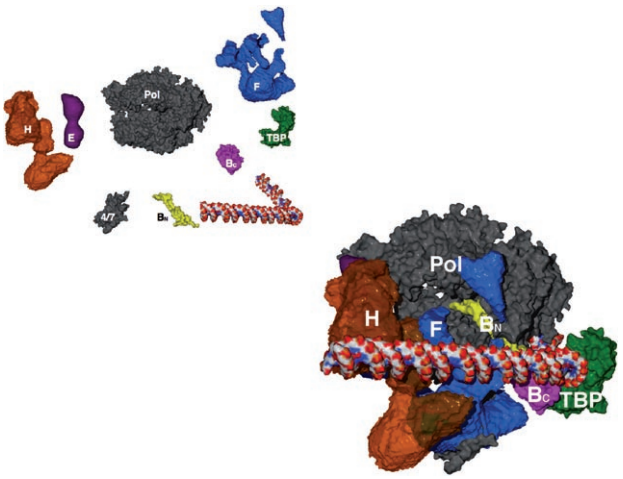


Figure 17. Model of an RNA polymerase II pre-initiation complex. The structures top left were assembled in the complex shown bottom right. Direction of view as in Figures 5 and 13. "4/7" indicates pol II subunits Rpb4 and Rpb7, which were omitted from the original structure (Figure 5).

general factors is seen to play a simple, essential role in the initiation process: TBP bends the promoter DNA around the polymerase and the C-terminal domain of TFIIB. The N-terminal domain of TFIIB brings the DNA to a point on the polymerase surface from which it need only follow a straight path and, by virtue of the conserved spacing from the TATA box to the transcription start site in pol II promoters, the start site is juxtaposed with the active center. TFIIE enters the complex and recruits TFIIH, whose ATPase/helicase subunit introduces negative superhelical tension in the DNA. Thermal unwinding produces a transient bubble, which is captured by the TFIIF binding to the nontemplate strand. The DNA can now bend in the single-stranded region and descend into the pol II active center. Initiation and the synthesis of RNA ensue, initially stabilized by the B finger. Synthesis of a transcript greater than about six residues in length leads to the displacement of TFIIB, promoter escape, and the completion of transcription.

The Mediator and the Regulation of Transcription

It remains to solve the structure of the Mediator, and thus to understand transcriptional regulation. The structure of a pol II–Mediator complex has been determined at low resolution by Francisco Asturias, by cryoelectron microscopy and single-particle analysis (Ref. [38] and Figure 18). The portion of the structure corresponding to pol II could be identified by docking the atomic model of the polymerase. The remainder of the structure, arising from the Mediator, was in the form of a crescent, largely enveloping pol II, with many points of contact through which regulatory information may be transmitted. Extension of the structure to atomic resolution will one day reveal the regulatory mechanism.

The charge for this lecture was to tell the story of the work leading to the Nobel Prize. It is also the story of many students

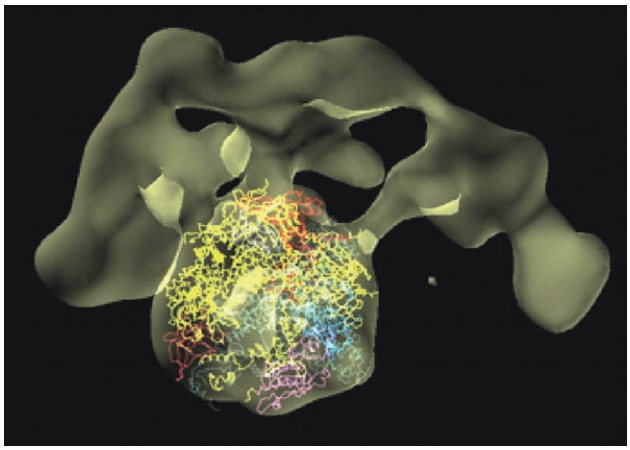


Figure 18. Cryo-EM structure of an RNA polymerase II–Mediator complex. The pol II structure was docked in the central electron density, and is shown in a similar direction of view and color scheme as Figure 5.

and postdoctoral fellows, whose skill, great effort, and willing suspension of disbelief, over a quarter century, transformed a scientific dream into a reality. I could only mention a few by name without interrupting the flow of the story, but the contributions of others were not less important. I pay heartfelt tribute to them, and to all who have shared part of their scientific lives with me (Figure 19), in pursuit not only of the

Chris Ackerson	Opher Gileadi	Barbara Maier-Davis
Francisco Asturias	Avi Gnatt	Lynne Mercer
Maia Azubel	Dave Goldfarb	Gavin Meredith
Brad Baer	Joachim Griesenbeck	Larry Myers
Avital Baraket-Samish	Claes Gustafsson	Kazuko Nishikura
Stefan Björklund	Lynn Henry	Hiroshi Nojima
Hinrich Boeger	John Heumann	Claudia Poglitsch
Tilman Borggrefe	Glenn Horn	Ariel Prunell
Rick Bram	Pablo Jadzinsky	Jeff Reidler
Andy Buchman	Grant Jensen	Hans Ribi
Dave Bushnell	Yi-Wei Jiang	Phil Robinson
Brad Cairns	Craig Kaplan	Alan Sachs
Guillermo Calero	Lin-Woo Kang	Mike Sayre
Wei-hau Chang	Don Katcoff	Jonny Sexton
Dan Chasman	Ray Kelleher	Dong Shin
Joan Conaway	Young-Joon Kim	Henrik Spahr
Patrick Cramer	Hirofumi Komori	Seth Strattan
Seth Darst	Guy Lorch Kornberg	Nori Sugimoto
Peter David	Liz Kubalek	Jesper Svejstrup
Ralph Davis	Yoshikazu Kurashima	Yuro Takagi
Al Edwards	Jan LaPointe	Herbert Tschochner
Andy Ehrensberger	Karl-Magnus Larsson	Ed Uzgiris.
John Feaver	Kerstin Leuther	Philippe Veschambre
Martha Fedor	Alan Levine	Daguang Wang
Peter Flanagan	Yang Li	Dong Wang
Jianhua Fu	Yahli Lorch	Ken Westover
Brian Gibbons	Neal Lue	Mincheng Zhang

Figure 19.

pol II structure, but also of chromatin, transcription biochemistry, and EM methodology. I should add that this is a lecture, not a review, and so is personal rather than referential. The pol II structure is a culmination of research done in many laboratories, and I am deeply indebted to those involved.

Received: April 25, 2007
Published online: August 2, 2007

[1] R. D. Kornberg, *Science* **1974**, *184*, 868 – 871.
[2] Y. Lorch, J. W. LaPointe, R. D. Kornberg, *Cell* **1987**, *49*, 203 – 210.
[3] M. Han, M. Grunstein, *Cell* **1988**, *55*, 1137 – 1145.
[4] H. Boeger, J. Griesenbeck, J. S. Strattan, R. D. Kornberg, *Mol. Cell* **2003**, *11*, 1587 – 1598.
[5] T. Matsui, J. Segall, P. A. Weil, R. G. Roeder, *J. Biol. Chem.* **1980**, *255*, 11992 – 11996.
[6] J. W. Conaway, D. Reines, R. C. Conaway, *J. Biol. Chem.* **1990**, *265*, 7552 – 7558.
[7] N. F. Lue, R. D. Kornberg, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8839 – 8843.
[8] R. Conaway, J. Conaway, *Prog. Nucleic Acid Res. Mol. Biol.* **1997**, *56*, 327 – 346.
[9] R. J. Kelleher, P. M. Flanagan, R. D. Kornberg, *Cell* **1990**, *61*, 1209 – 1215.
[10] P. M. Flanagan, R. J. Kelleher, M. H. Sayre, H. Tschochner, R. D. Kornberg, *Nature* **1991**, *350*, 436 – 438.
[11] Y.-J. Kim, S. Björklund, Y. Li, M. H. Sayre, R. D. Kornberg, *Cell* **1994**, *77*, 599 – 608.
[12] R. D. Kornberg, *Trends Biochem. Sci.* **2005**, *30*, 235 – 239.
[13] S. Malik, R. G. Roeder, *Trends Biochem. Sci.* **2005**, *30*, 256 – 263.
[14] Z. Moqtaderi, Y. Bai, D. Poon, P. A. Weil, K. Struhl, *Nature* **1996**, *383*, 188 – 191.
[15] S. S. Walker, J. C. Reese, L. M. Apone, M. R. Green, *Nature* **1996**, *383*, 185 – 188.
[16] W.-C. Shen, M. R. Green, *Cell* **1997**, *90*, 615 – 624.
[17] R. C. Conaway, S. Sato, C. Tomomori-Sato, T. Yao, J. W. Conaway, *Trends Biochem. Sci.* **2005**, *30*, 250 – 255.
[18] H. M. Bourbon et al., *Mol. Cell* **2004**, *14*, 553 – 557.
[19] C. M. Thompson, R. A. Young, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 4587 – 4590.
[20] J. Takagi, R. D. Kornberg, *J. Biol. Chem.* **2005**, *281*, 80 – 89.
[21] R. D. Kornberg, H. M. McConnell, *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 2564 – 2568.
[22] E. E. Uzgiris, R. D. Kornberg, *Nature* **1983**, *301*, 125 – 129.
[23] A. M. Edwards, S. A. Darst, W. J. Feaver, N. E. Thompson, R. R. Burgess, R. D. Kornberg , *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2122 – 2126.
[24] S. A. Darst, E. W. Kubalek, A. M. Edwards, R. D. Kornberg, *J. Mol. Biol.* **1991**, *221*, 347 – 357.
[25] J. Fu, A. L. Gnatt, D. A. Bushnell, G. J. Jensen, N. E. Thompson, R. R. Burgess, P. R. David, R. D. Kornberg, *Cell* **1999**, *98*, 799 – 810.
[26] P. Cramer, D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis, N. E. Thompson, R. R. Burgess, A. M. Edwards, P. R. David, R. D. Kornberg *Science* **2000**, *288*, 640 – 649.
[27] P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, *292*, 1863 – 1876.
[28] A. Gnatt, J. Fu, R. D. Kornberg, *J. Biol. Chem.* **1997**, *272*, 30799 – 30805.
[29] A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, *292*, 1876 – 1882.
[30] K. D. Westover, D. A. Bushnell, R. D. Kornberg, *Cell* **2004**, *119*, 481 – 489.
[31] D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, *Cell* **2006**, *127*, 941 – 954.
[32] G. Zhang, E. A Campbell, L. Minakhin, C. Richter, K. Severinov, S. A Darst, *Cell* **1999**, *98*, 811 – 824.
[33] K. D. Westover, D. A. Bushnell, R. D. Kornberg, *Science* **2004**, *303*, 1014 – 1016.
[34] D. A. Bushnell, K. D. Westover, R. E. Davis, R. D. Kornberg, *Science* **2004**, *303*, 983 – 988.
[35] K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, *Science* **2002**, *296*, 1285 – 1290.
[36] K. S. Murakami, S. Masuda, S. A. Darst, *Science* **2002**, *296*, 1280 – 1284.
[37] D. B. Nikolov, H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. Kun Lee, R. G. Roeder, S. K. Burley, *Nature* **1995**, *377*, 119 – 128.
[38] J. A. Davis, Y. Takagi, R. D. Kornberg, F. J. Asturias, *Mol. Cell* **2002**, *10*, 409 – 415.