FEBS 29066

FEBS Letters 579 (2005) 899–903

Minireview

Structural basis of eukaryotic gene transcription

Hinrich Boeger, David A. Bushnell, Ralph Davis, Joachim Griesenbeck, Yahli Lorch
J. Seth Strattan, Kenneth D. Westover, Roger D. Kornberg*

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

Received 2 November 2004; accepted 2 November 2004

Available online 23 November 2004

Edited by Gunnar von Heijne and Anders Liljas

Abstract **An RNA polymerase II promoter has been isolated in transcriptionally activated and repressed states. Topological and nuclease digestion analyses have revealed a dynamic equilibrium between nucleosome removal and reassembly upon transcriptional activation, and have further shown that nucleosomes are removed by eviction of histone octamers rather than by sliding. The promoter, once exposed, assembles with RNA polymerase II, general transcription factors, and Mediator in a ~3 MDa transcription initiation complex. X-ray crystallography has revealed the structure of RNA polymerase II, in the act of transcription, at atomic resolution. Extension of this analysis has shown how nucleotides undergo selection, polymerization, and eventual release from the transcribing complex. X-ray and electron crystallography have led to a picture of the entire transcription initiation complex, elucidating the mechanisms of promoter recognition, DNA unwinding, abortive initiation, and promoter escape.**
© 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: Nucleosome; RNA polymerase II; General transcription factors; Mediator

1. Introduction

The many thousands of genes coding for proteins in eukaryotes are transcribed by a common multiprotein machinery. Biochemical studies have revealed the components of the machinery and their generality. Structural studies have provided a basis for understanding the biochemical results, explaining the roles of all the components in the transcription process. Here, we review structural studies of the past year, defining the nature of the transcription template and leading to a nearly complete picture of a transcription initiation complex.

2. The transcription template

The DNA of eukaryotes is ordinarily refractory to transcription because of its organization in nucleosomes (reviewed in [1]). The DNA is wrapped around a histone octamer, which interferes with many DNA transactions. Nucleosomes thus serve as general gene repressors. They help assure the inactivity of all genes in eukaryotes except those whose transcription is brought about by specific positive regulatory mechanisms.

A first indication of how repression by the nucleosome is relieved came from nuclease digestion of chromatin, showing an increase in accessibility of promoter DNA upon transcriptional activation (reviewed in [2]). Recently, it was found that active promoters are associated with histones modified in various ways, including acetylation, phosphorylation, and methylation. Some of these modifications are correlated with and apparently required for transcription (reviewed in [3]). Such findings have led to the widespread belief that active promoters are associated with an altered form of the nucleosome.

We wished to isolate and analyze any altered form of the nucleosome, and employed homologous recombination for the purpose [4]. We inserted recognition sites for the R recombinase on either side of the *PHO5* gene of yeast. Induction of the recombinase caused excision of the gene as a small circle, bearing 12 nucleosomes. Three of these nucleosomes are located in the promoter region and were shown long ago to be altered in some way upon transcriptional activation [5].

An alteration of promoter nucleosome structure could be detected in a sensitive manner by topological analysis. The wrapping of DNA around a histone octamer imparts approximately one left-handed superhelical turn, and if the DNA is in circular form then the superhelicity is retained even after the octamer is removed. The superhelicity of the circle can then be measured by gel electrophoresis. We found a change in superhelicity upon *PHO5* activation corresponding to the loss of 1.85 nucleosomes, which could be due to an altered state of the promoter in which all three nucleosomes are partially unfolded, or *to the complete unfolding of 1.85 out of the three nucleosomes* [6]. We could distinguish between these possibilities by a quantitative analysis of the nuclease digestion of *PHO5* chromatin.

An enzyme such as micrococcal nuclease rapidly removes the DNA between nucleosome core particles and then more slowly degrades the core particles themselves. The degradation process obeys a simple second order rate equation, and it can easily be shown that the ratio of DNA remaining from activated and repressed promoters, as revealed by hybridization with a labeled probe, approaches the initial ratio of core particles on the promoters, in the limit of complete digestion.

The results of experiments exhibit such asymptotic behavior, with a limiting value of 0.37 detected by a probe for three *PHO5* promoter nucleosomes, corresponding to $0.37 \times 3 = 1.1$ core particles retained upon activation. The experiment

*Corresponding author. Fax: +1 650 723 8464.
E-mail address: kornberg@stanford.edu (R.D. Kornberg).

0014-5793/$30.00 © 2004 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.febslet.2004.11.027

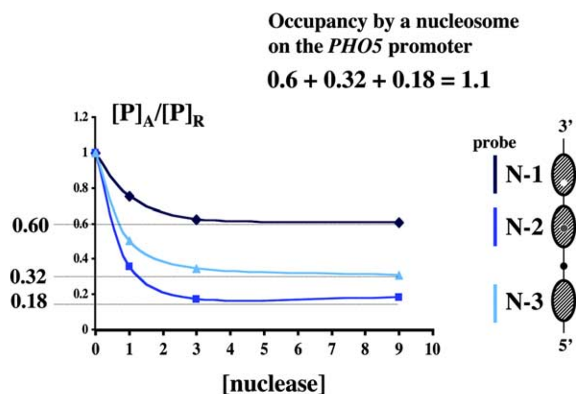

Fig. 1. Probability of a nucleosome at various positions on the transcriptionally active *PHO5* promoter. *PHO5* gene circles, isolated from yeast before or after transcriptional activation, designated R and A, respectively, were digested with increasing amounts of micrococcal nuclease as indicated, for 5 min at 30 °C. DNA was extracted, electrophoresed in an agarose gel, blotted, and hybridized with probes for the N-1, N-2, and N-3 nucleosome regions, as diagrammed on the right. The total hybridization signal P was measured by integrating the profile of radioactivity in each lane of the blot and the ratio of hybridization signals was plotted against the amount of nuclease in the reaction. In the diagram of the promoter, the white circle indicates the TATA box, the gray circle indicates upstream activation sequence 2, and the black circle indicates upstream activation sequence 1.

was repeated with three individual probes for the three promoter nucleosomes. Asymptotic behavior was again observed, with the limiting values for retention of the three individual core particles adding up to exactly the same value of 1.1 as from the previous analysis with a single promoter probe covering all three particles ([6], Fig. 1).

The retention of 1.1 core particles corresponds to *the loss of 1.9 of the original three nucleosomes*, in striking agreement with the results from topological analysis. This close quantitative agreement between two very different methods, along with additional evidence, leads to the conclusion that nucleosomes are lost from the transcriptionally activated promoter, and not retained in an altered form [6].

We infer from the results that transcriptional activation does not produce a persistently altered state of promoter chromatin. Rather there is a dynamic process, in which nucleosomes are continually removed from an activated promoter and reassembled. What we measure are steady state values, reflecting the balance between removal and reassembly. We can easily reconcile our conclusion that promoter nucleosomes are removed with the current evidence that modified histones remain associated with active promoters: histones are indeed removed but they are also reformed; modified histones may represent intermediates in the cycle of removal and reformation. Two recent reports, as well as our own unpublished work, indicate that our conclusion is general – nucleosomes are removed from most, if not all, promoters in the course of transcriptional activation.

If nucleosomes are removed, then what is the mechanism? There have been reports of nucleosomes sliding away from the TATA box and start site of transcriptionally active promoters [7]. In this case, the histone octamer remains associated with the DNA. The alternative is that the histone octamer is dislodged from the DNA. We could distinguish between these possibilities by recourse to chromatin circles. In our previous experiments, we activated the *PHO5* gene in yeast and afterwards induced recombinase to form gene circles. If, instead, we first form circles and then activate transcription, we can answer the question about the mechanism. The reason is that a circle is a closed domain, so if nucleosomes are removed by sliding, their number on the circle will remain the same. If, on the other hand, nucleosomes are removed by disassembly, their number will be diminished. The results from topological analysis and limited digestion were again in agreement: the number of nucleosomes declined to almost the same extent as before. We conclude that nucleosomes are removed by disassembly, not by sliding [8].

What is the mechanism of disassembly? Recent findings have implicated chromatin-remodeling complexes, such as the 11-subunit SWI/SNF complex [9,10], and the related but more abundant and essential 15-subunit RSC complex [11]. RSC can disassemble a nucleosome by transferring the histone octamer to another DNA molecule [12]. There is recent genetic evidence for a histone octamer-acceptor protein required for *PHO5* gene activation in yeast, the previously identified Asf1 histone "chaperone" protein [13]. It remains to be determined whether RSC or another chromatin-remodeling complex transfers histone octamers from the *PHO5* promoter to Asf1 in the course of transcriptional activation.

3. The RNA polymerase II transcription machinery

Due to the loss of nucleosomes from transcriptionally active promoters, the RNA polymerase II transcription initiation complex is assembled on a naked DNA molecule in vivo. It is for this reason that we and others have been able, over the years, to identify the components of the transcription machinery by fractionation of cell extracts, guided by transcription assays with naked DNA in vitro.

The RNA polymerase II transcription machinery defined in this way is made up of three components: a 12-subunit polymerase, capable of synthesizing RNA and proofreading the nascent transcript; a set of five general transcription factors, denoted TFIIB, -D, -E, -F, and –H, responsible for promoter recognition and for unwinding the promoter DNA [14]; and a 20-subunit Mediator, which transduces regulatory information from

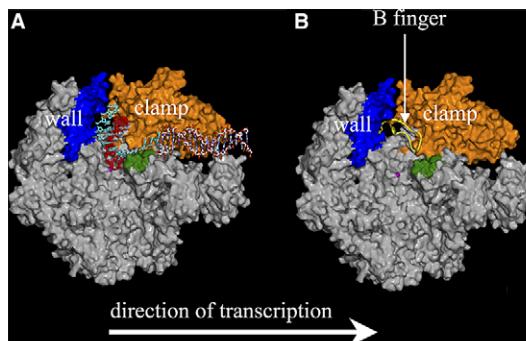

Fig. 2. Structure of RNA polymerase II. Cutaway view, to reveal contents of active center cleft. Surface representation of atomic model, with features relevant to the discussion colored as follows: clamp, orange; wall, blue; bridge helix, green; active center Mg ion, pink; and remainder of polymerase, gray. (A) Transcribing complex, with coding strand of DNA in active center region in turquoise, and RNA in red (PDB 1I6H). (B) RNA polymerase II – TFIIB complex, with backbone model of TFIIB in yellow (PDB 1R5U).

activator and repressor proteins to RNA polymerase II [15–17]. Whereas the polymerase and general transcription factors have counterparts in bacteria, Mediator is unique to eukaryotes, and must enable the far more intricate regulation that underlies the development of complex multicellular organisms.

With nearly 60 subunits and a mass in excess of 3 million Daltons, the RNA polymerase II transcription machinery poses a formidable challenge for structural analysis. We began with the structure determination of the polymerase because it forms the core of the machinery. It is the platform upon which all components are assembled. As will shortly be seen, the polymerase structure is a key for understanding eukaryotic gene transcription.

The structure was initially at 2.8 Å [18] and has since been extended to 2.3 Å resolution (Westover et al., Cell in press). It was determined both alone and with DNA and RNA, in the form of a transcribing complex ([19], Fig. 2A). The nucleic acids occupy a deep cleft across the surface of a transcribing complex. The DNA is unwound, with 9 bp of DNA–RNA hybrid in the active center region. The axis of the hybrid helix is at nearly 90° to that of the entering DNA duplex, due to a *wall* of protein density that prevents straight passage of nucleic acids through the cleft. A massive protein element, termed the *clamp*, swings more than 30 Å over the nucleic acids in the course of forming the transcribing complex.

Two recent results have illuminated the critical events in transcription at the ends of the DNA–RNA hybrid. At the downstream end, where RNA synthesis occurs, the original transcribing complex structure contained the nucleotide just added to the growing chain still in the nucleotide addition, or *A*, site. We therefore refer to this structure as a *pre-translocation complex* (Fig. 3). We have since learned to trap and solve the complex at the next step, following translocation. In the new structure, of a *post-translocation complex* (Fig. 3), the nucleotide just added to the RNA has moved to the next position, leaving the A site open for entry of a nucleoside triphosphate (NTP). We added an NTP, appropriately matched for base-pairing with the coding base in the DNA, and found difference density for this NTP in the A site (Fig. 3). To our surprise, when we performed the control and added an unmatched nucleoside triphosphate, that could not pair with the coding base, it also gave rise to difference density corresponding to a bound NTP, but in an inverted orientation, defining a new site that we refer to as the entry or *E* site (Fig. 3). Additional structures demonstrated binding of all four NTPs to the E site. This series of crystal structures (Westover et al., Cell in press) has revealed a nucleotide addition cycle (Fig. 3): entry in the E site, rotation to sample base pairing in the A site, phosphodiester bond formation, and translocation to repeat the cycle.

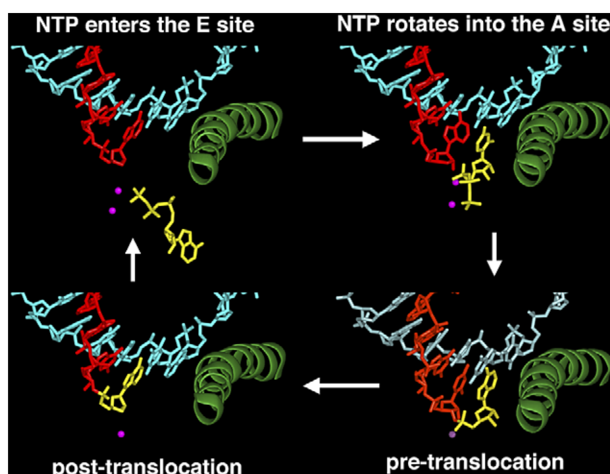

Fig. 3. Four crystal structures of RNA polymerase II transcribing complexes. Only nucleic acids in the active center region, bridge helix, and Mg ions are shown. (Upper left) Post-translocation complex with unmatched NTP bound in E site (PDB 1R9T). (Upper right) Post-translocation complex with matched NTP bound in E site (PDB 1R9S). (Lower right) pre-translocation complex (PDB 116H). (Lower left) Post-translocation complex (PDB 1SFO). Color code as in Fig. 2a, with nucleotide just added to the RNA chain or NTP soaked into the crystals in yellow.

Our structures of pre- and post-translocation complexes could be compared with those of the small single subunit polymerases, for example bacteriophage T7 RNA polymerase. The structures had virtually nothing in common. Both the protein and the configuration of the coding strand of the DNA were unrelated between the eukaryotic and bacteriophage complexes (Westover et al., Cell in press). We conclude that the very large multisubunit RNA polymerases responsible for cellular gene transcription and regulation evolved independently of the small single subunit DNA and RNA polymerases of viruses and bacteria.

Our new transcribing complex structure also revealed the way in which newly synthesized RNA is separated from the DNA template at the upstream end of the DNA–RNA hybrid [20]. The bases at position 7 were separated by the distance of a hydrogen bond, but the bases at positions 8, 9, and 10 were increasingly far apart. This stable melting of the DNA–RNA hybrid was due to the intervention of three protein loops, termed *fork loop 1*, *rudder*, and *lid*. Rudder and lid were located between the DNA and RNA strands, with rudder contacting the DNA and lid the RNA. A phenylalanine side chain of the lid appeared to serve as a wedge to maintain separation of the strands. Fork loop 1 contacted base pairs 6 and 7, limiting the strand separation to positions 8 and beyond, preventing the DNA–RNA hybrid from unraveling further. The three loops interacted with one another as well as with the DNA and RNA, forming a strand-loop network, whose stability must drive the melting process. Finally, it may be noted, for the purposes of the discussion to follow, that the strand-loop network guides the RNA to a region between the clamp and wall that we term the *saddle.*

As already mentioned, the full significance of the RNA polymerase II structure lies in the implications for the higher complexes formed with general transcription factors and Mediator. It is at this level that the polymerase structure transforms our view of transcription and of the regulation of the process. An illustration of the point comes from the structure of a cocrystal of RNA polymerase II with the general factor TFIIB.

In this structure [21], the N-terminal domain of TFIIB starts with a Zn ribbon that binds the polymerase *dock* domain. Then, the N-terminal domain does a surprising thing – rather than pass back out into solution, it continues on across the saddle between the clamp and wall and plunges down into the polymerase active center, from which it then loops back and remerges across the saddle (Fig. 2B). We refer to the loop as the *B finger*.

It will be apparent that the B finger occupies the same location as the DNA–RNA hybrid in a transcribing complex. Closer

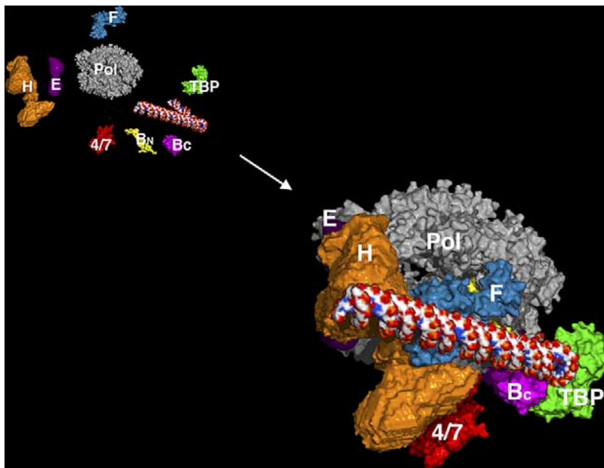

Fig. 4. RNA polymerase II transcription initiation complex. X-ray and electron microscope structures (upper left) were assembled in a complete transcription initiation complex (lower right) as described in the text.

examination reveals that the clash is with the RNA beyond about position 5, not with the DNA. Indeed, a binding study has shown that TFIIB actually stabilizes a complex containing a 5-residue RNA. Rather than interfering, TFIIB may enhance the formation of an early transcribing complex before a length of 9 base pairs, required for optimal stability, is attained. On the other hand, when the RNA grows beyond 9 residues and the strand-loop network is formed, a clash with TFIIB is unavoidable. RNA and TFIIB must compete for space on the polymerase saddle. If RNA wins the competition, TFIIB is ejected and the polymerase is released from the promoter to complete transcription of the gene. If TFIIB wins, initiation aborts and must be tried again. The location of the B finger thus explains two crucial but until now completely mysterious aspects of transcription, abortive initiation and promoter escape.

Turning to the C-terminal domain of TFIIB, only one α-helical region was clearly resolved in the cocrystal with the polymerase, but this region was sufficient to dock the previously determined structure of the C-terminal domain in a complex with the TBP subunit of TFIID and a TATA box fragment of promoter DNA. TBP bends the TATA box region, and when the fragment is extended with straight B-form DNA, the role of the bend becomes apparent (Fig. 4): it configures the DNA to the contours of the polymerase. The DNA fits perfectly to the polymerase surface.

What is more, the trajectory of the DNA leads right by the saddle, crossing it about 15 bp from the TATA box (Fig. 4). We know from the transcribing complex structure that after the DNA unwinds, there are about 10 residues from the saddle to the active site. The sum of 15 bp from the TATA box plus 10 residues to the active site corresponds well with the spacing of 25 bp from the TATA box to the transcription start site characteristic of almost all RNA polymerase II promoters.

Finally, we could assemble a first picture of a "minimal" initiation complex of RNA polymerase II and all the general transcription factors from the combined results of X-ray diffraction and electron microscopy ([21], Fig. 4). Two polymerase subunits, Rpb4 and Rpb7, omitted from the original polymerase structure, were included on the basis of more recent crystallographic results. The location of TFIIF was based on electron microscopy of an RNA polymerase II – TFIIF complex, with the structure of TFIIF derived from that of bacterial sigma factor, in light of sequence and structural homologies. The location of TFIIE was from electron microscopy of a 2-D RNA polymerase II – TFIIE cocrystal. The structure of TFIIH was also from electron microscopy of a 2-D crystal.

Our current understanding of transcription initiation in terms of this first structural picture may be summarized as follows. TBP bends TATA box DNA around the C-terminal domain of TFIIB. The N-terminal domain of TFIIB brings the complex to a point on the polymerase surface from which the DNA need only follow a straight path and, by virtue of the conserved spacing from TATA box to transcription start site in pol II promoters, the start site will be juxtaposed with the active center. Entry of TFIIB and promoter DNA in the complex leads to binding of TFIIE which, in turn, recruits TFIIH. An ATPase/helicase subunit of TFIIH torques the DNA, introducing negative superhelical tension in the region over the active center cleft. Consequent thermal unwinding is followed by capture of the non-template strand by TFIIF, as it is in bacterial initiation by sigma, and the template strand descends to the active site, where it interacts with the B finger. Initiation and the synthesis of RNA greater than about 10 residues in length lead to the displacement of TFIIB, promoter escape, and the completion of gene transcription.

We thus arrive at a solution, in outline, of the transcription initiation problem. Each of the GTFs is seen to play a simple, essential role in the initiation mechanism. Testing and refinement will come from structure determination of additional RNA polymerase II – general transcription factor cocrystals. After that, it will remain to solve the Mediator and gain an understanding of transcriptional regulation. We have a first indication of Mediator structure from electron microscopy of an RNA polymerase II – Mediator complex [22]. In this low resolution structure, the density due to RNA polymerase II can be identified by a fit to the atomic structure. Mediator is then seen to form a crescent-shaped density largely enveloping the polymerase. There are many points of contact through which regulatory information may be delivered to the polymerase. Extension of polymerase crystallography to the Mediator complex will illuminate the regulatory mechanism.

References

[1] Kornberg, R.D. and Lorch, Y. (1999) Cell 98, 285–294.
[2] Elgin, S.C. (1988) J. Biol. Chem. 263, 19259–19262.
[3] Wu, J. and Grunstein, M. (2000) Trends Biochem. Sci. 25, 619–623.
[4] Ansari, A. and Gartenberg, M.R. (1999) Proc. Natl. Acad. Sci. USA 96, 343–348.
[5] Svaren, J. and Horz, W. (1997) Trends Biochem. Sci. 22, 93–97.
[6] Boeger, H., Griesenbeck, J., Strattan, J.S. and Kornberg, R.D. (2003) Mol. Cell. 11, 1587–1598.
[7] Lomvardas, S. and Thanos, D. (2001) Cell 106, 685–696.
[8] Boeger, H., Griesenbeck, J., Strattan, J.S. and Kornberg, R.D. (2004) Mol. Cell. 14, 667–673.
[9] Cairns, B.R., Kim, Y.J., Sayre, M.H., Laurent, B.C. and Kornberg, R.D. (1994) Proc. Natl. Acad. Sci. USA 91, 1950–1954.
[10] Peterson, C.L., Dingwall, A. and Scott, M.P. (1994) Proc. Natl. Acad. Sci. USA 91, 2905–2908.

[11] Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B. and Kornberg, R.D. (1996) Cell 87, 1249–1260.
[12] Lorch, Y., Zhang, M. and Kornberg, R.D. (1999) Cell 96, 389–392.
[13] Adkins, M.W., Howar, S.R. and Tyler, J.K. (2004) Mol. Cell 14, 657–666.
[14] Conaway, R. and Conaway, J. (1997) Prog. Nuc. Acid Res. Mol. Biol. 56, 327–346.
[15] Kelleher, I.R.J., Flanagan, P.M. and Kornberg, R.D. (1990) Cell 61, 1209–1215.
[16] Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H. and Kornberg, R.D. (1994) Cell 77, 599–608.
[17] Gustafsson, C.M., Myers, L.C., Beve, J., Spahr, H., Lui, M., Erdjument-Bromage, H., Tempst, P. and Kornberg, R.D. (1998) J. Biol. Chem. 273, 30851–30854.
[18] Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001) Science 292, 1863–1876.
[19] Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A. and Kornberg, R.D. (2001) Science 292, 1876–1882.
[20] Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004) Science 303, 1014–1016.
[21] Bushnell, D.A., Westover, K.D., Davis, R.E. and Kornberg, R.D. (2004) Science 303, 983–988.
[22] Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M. and Kornberg, R.D. (1999) Science 283, 985–987.