Biol. Chem., Vol. 391, pp. 731–735, July 2010 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2010.094

Minireview

Towards molecular systems biology of gene transcription and regulation

Patrick Cramer

Gene Center and Department of Biochemistry, Center for Integrated Protein Science CIPSM, Ludwig Maximilians University, Feodor-Lynen-Straße 25, D-81377 Munich, Germany

e-mail: cramer@LMB.uni-muenchen.de

Abstract

Ten years after the determination of the RNA polymerase II structure, the basic mechanism of mRNA synthesis during gene transcription is known. In the future, the initiation and regulation of transcription must be studied with a combination of structural biology, biochemistry, functional genomics, and computational methods. In this article, the efforts of our laboratory to move from an integrated structural biology of gene transcription towards molecular systems biology of gene regulation are reviewed.

Keywords: gene transcription and regulation; integrated structural biology; molecular systems biology; RNA polymerase; yeast.

Introduction

Transcription is the first step in gene expression and is highly regulated in living cells. During transcription, the enzyme RNA polymerase II (Pol II) synthesizes messenger-RNA (mRNA). Pol II is at the heart of a large multiprotein machinery that changes composition during the transcription cycle. To recognize promoters at the beginning of genes, initiation factors assemble with Pol II into the initiation complex. A large multiprotein complex, the Mediator, integrates regulatory signals from activators and repressors bound to adjacent DNA sites, and stimulates the initiation of transcription. Other protein complexes bind Pol II during elongation and processing of the RNA chain, and induce changes in chromatin modification and structure.

Transcription elongation is well understood

A detailed structure-function analysis of transcription was initiated ten years ago, when the crystal structure of the Pol II core enzyme from yeast was reported (Cramer et al., 2000, 2001). The Pol II core enzyme comprises ten subunits and has a molecular weight of approximately 0.5 MDa. The first structure of an elongation complex of Pol II with a minimal DNA template and RNA product was also obtained (Gnatt et al., 2001). These structures revealed the architecture of Pol II. At the floor of a central cleft resides the active center of the enzyme, which holds a DNA-RNA hybrid duplex (Figure 1). These breakthrough discoveries were decisive for awarding the Nobel Prize for Chemistry to Roger Kornberg in 2006. In the following years, many structures of functional complexes of Pol II were determined (Gnatt et al., 2001; Kettenberger et al., 2004; Wang et al., 2006; Vassylyev et al., 2007; Brueckner and Cramer, 2008) that together with functional data led to the first video of the nucleotide addition cycle during transcription elongation (Brueckner et al., 2009). The video facilitates teaching and can be downloaded from the internet at www.LMB.uni-muenchen.de/cramer/PR-materials/.

During elongation, the nucleotide addition cycle is repeated over and over again. The cycle begins with binding of a nucleoside triphosphate (NTP) substrate at the active center. The subsequent catalytic addition of the nucleotide to the growing 3′-end of the mRNA chain is followed by release of a pyrophosphate ion. DNA and RNA are then translocated over the Pol II surface. Translocation generates a free NTP-binding site and the cycle is repeated. An important role is played by the so-called trigger loop, which closes over the active center to establish a catalytically competent conformation (see Figure 1; Wang et al., 2006). The NTP binds two catalytic metal ions, A and B. Metal A is persistently bound to the enzyme and metal B enters with the NTP. The mechanism of translocation remains poorly understood, but unexpected insights were obtained from the structure of the Pol II elongation complex bound to α-amanitin, the toxin of the amanita mushroom (Figure 1). Amanitin forms a cyclic peptide that stabilizes Pol II in a state that can be described as an intermediate of translocation (Brueckner and Cramer, 2008). In this state, the next DNA template base resides above the so-called bridge helix, a conserved element of the active center. This helix is shifted such that it blocks entry of the template base into the active site. A wedge-like conformation of the trigger loop stabilizes the bridge helix in this state and is part of a Brownian ratchet that is thought to underlie the translocation mechanism.

Over the past few years the mechanisms of RNA proofreading and the detection of DNA lesions were also eluci-

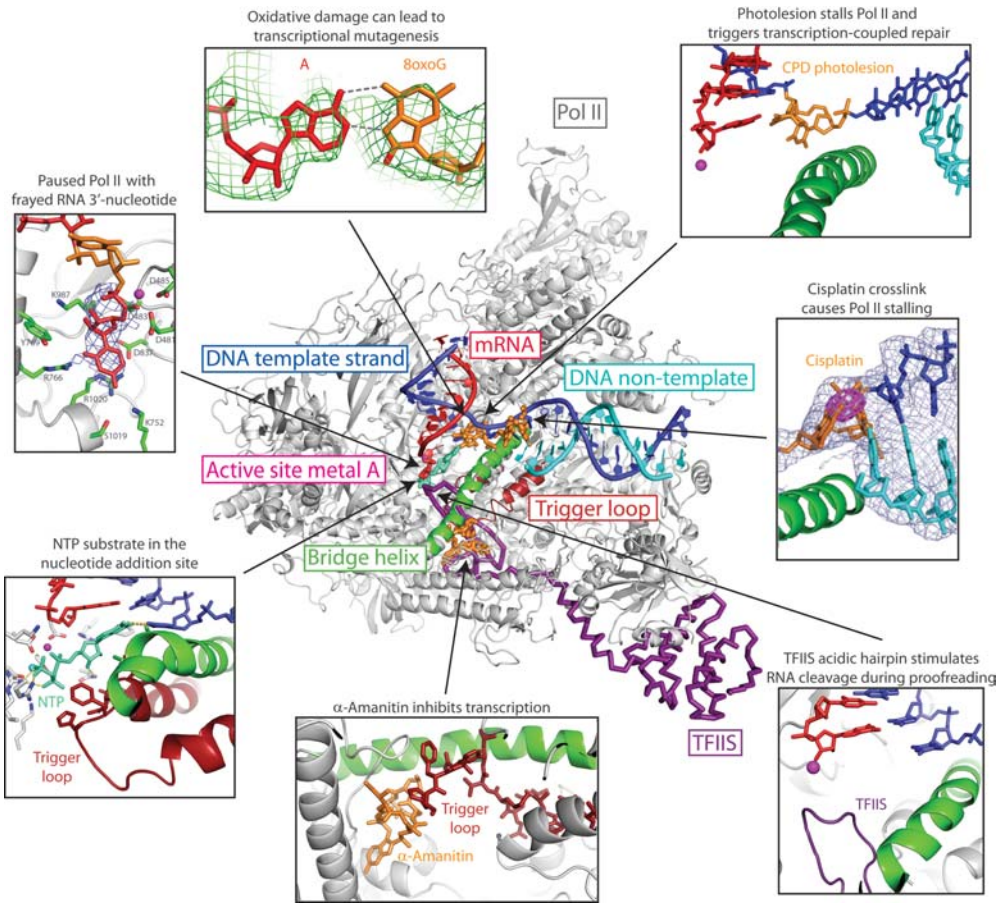

Figure 1 Structures elucidate transcription elongation.
In the elongation complex structure, Pol II is shown as a ribbon model in silver, the DNA template and non-template strands are in blue and cyan, respectively, and RNA is in red. The active site metal ion A is in pink, the bridge helix is green. During elongation, Pol II adds a nucleotide to the growing RNA 3′-end. For catalysis, the trigger loop (dark red) closes over the NTP substrate (Wang et al., 2006). Trigger loop closure is prevented by the mushroom toxin α-amanitin (orange) that inhibits transcription (Brueckner and Cramer, 2008). Misincorporations can be corrected with the help of the RNA cleavage stimulatory factor TFIIS (orange) (Kettenberger et al., 2003, 2004). Proofreading begins with pausing of Pol II, mediated by fraying of the 3′-terminal RNA nucleotide (Sydow et al., 2009). DNA lesions such as a cyclobutan pyrimidine dimer (CPD, orange) or a cisplatinum crosslink (orange) can stall transcription and trigger DNA repair (Brueckner et al., 2007; Damsma et al., 2007). In the case of the small 8-oxoguanine lesion (8oxoG) the lesion remains unrepaired and the RNA can undergo mutagenesis (Damsma and Cramer, 2009).

dated. Pol II has an intrinsic RNA cleavage activity that is strongly stimulated by the transcription factor IIS (TFIIS), which approaches the active site through a pore (Figure 1) (Kettenberger et al., 2003, 2004). In the case of nucleotide misincorporation, Pol II can remove the erroneous nucleotide. During such proofreading, the polymerase first pauses, and then translocates backwards by one step to induce cleavage of a 3′-terminal RNA dinucleotide that comprises the erroneous nucleotide. The high fidelity of Pol II for the incorporation of the correct nucleotide is required to avoid the production of erroneous mRNA chains. Pol II uses different strategies to recognize and remove misincorporated nucleotides (Sydow et al., 2009). One reason for nucleotide misincorporation into the growing RNA chain can be the presence of DNA lesions within the template strand. DNA gets damaged by ultraviolet light, chemical agents, or oxidative stress. The resulting lesions can block transcription and induce DNA repair. Structure-function studies of elongation complexes with big lesions were able to resolve the first step in such transcription-coupled repair (Figure 1) (Brueckner et al., 2007; Damsma et al., 2007). Small lesions such as 8-oxoguanin can trigger transcriptional mutagenesis (Figure 1) (Damsma and Cramer, 2009). Mutagenesis occurs because a misincorporated adenine forms a stable Hoogsteen base pair with the lesion and escapes proofreading.

Transcription initiation has recently been elucidated

Transcription starts with recognition of the promoter region by initiation factors, which assemble with Pol II into the initiation complex near the transcription start site. Transcription initiation is a very complex and highly regulated event.

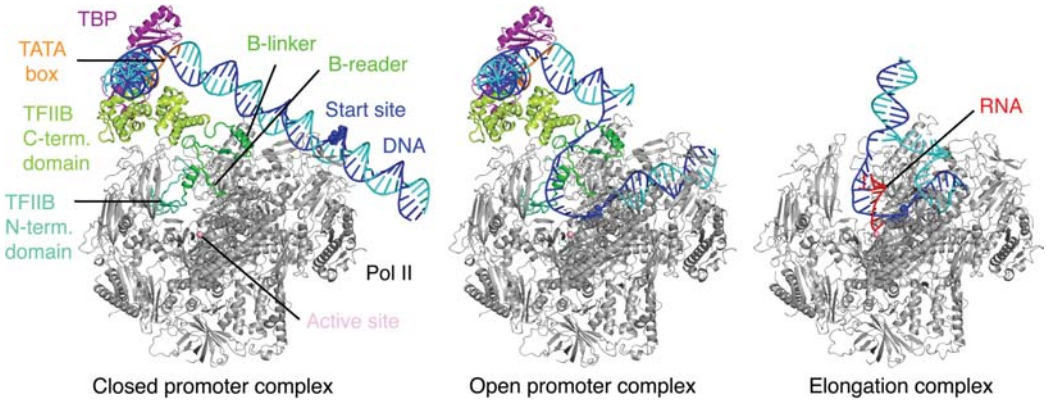

Figure 2 Model of transcription initiation.
Initiation begins with formation of a closed promoter complex that minimally contains TBP, TFIIB, and Pol II on promoter DNA (left). The DNA duplex is then melted with the help of the B-linker and the template strand is positioned at the polymerase active center to form the open promoter complex (middle). Then the DNA template strand is scanned with the help of the B-reader for an Initiator sequence motif. Models are based on details in Kostrewa et al. (2009). Synthesis of the RNA leads to a clash with TFIIB, release of TFIIB, and escape of Pol II from the promoter, and formation of the elongation complex (Andrecka et al., 2009) (right).

The following model of initiation (Figure 2) is based on superposition of two crystal structures, one comprising the TATA box-binding protein (TBP), transcription factor IIB (TFIIB), DNA, and one recent structure that contains Pol II and TFIIB (Nikolov et al., 1995; Kostrewa et al., 2009). Many promoters contain a so-called TATA box approximately 30 base pairs upstream of the transcription start site. The TATA box is bound by TFIID, which comprises TBP. TBP bends the DNA by 90 degrees. The initiation factor TFIIB binds with its C-terminal domain to TBP and to DNA regions flanking TBP, and with its N-terminal domain to Pol II, thereby recruiting Pol II to promoter DNA near the transcription start site. The resulting complex of Pol II, TFIIB, TBP, and promoter DNA is called a minimal closed promoter complex (Figure 2). An independent structure of the Pol II-TFIIB complex arrived at a similar model of the closed complex (Liu et al., 2010). In archaeal cells, this minimal complex suffices to promote transcription, whereas the additional initiation factors TFIIE, -F, and -H are required in eukaryotic cells.

Pol II alone is unable to locate the transcription start site. Instead, it requires continued help from TFIIB. First, the two DNA strands must be unwound. A structural element of TFIIB, the B-linker, is apparently involved in DNA opening before the transcription start site (Kostrewa et al., 2009). Second, the template single strand must be threaded into the active center and the transcription start site must be detected. TFIIB is again required at this step and helps to recognize the so-called Initiator element in the DNA. For this task, TFIIB uses the so-called B-reader, another structural element that consists of a helix followed by a mobile loop. Now the polymerase synthesizes short RNA products, which often dissociate again from the DNA template. The growing RNA strand interferes with TFIIB and helps to release TFIIB from Pol II. Now the polymerase can escape the promoter and enter the elongation phase of transcription (Andrecka et al., 2009; Kostrewa et al., 2009).

Understanding transcription regulation remains a major challenge

It is the initiation phase and the initiation-elongation transition that are highly regulated in cells. Therefore it is very important to obtain a mechanistic view of this process. The initiation-elongation transition requires co-activator complexes such as the Mediator, which integrate signals from regulatory factors. Mediator transmits signals to the Pol II machinery, but it is unknown how this is achieved. Mediator consists of 25 proteins and has a molecular weight of over 1 MDa. A combination of structural analysis and differential gene expression analysis revealed that Mediator contains functional submodules that are involved in the regulation of subsets of genes (Figure 3). For example, yeast mutations that destroy the integrity of certain Mediator submodules lead to submodule-specific changes in the transcriptome (Lariviere et al., 2006; Koschubs et al., 2009).

The goal of transcription research is to unravel the regulatory principles not only on a molecular level but genome-wide and in a quantitative manner. It will not be sufficient to measure mRNA abundance. Instead, the rates of synthesis and decay for each mRNA in the cell must be determined. In addition, the underlying DNA-protein interactions must be mapped genome-wide at high resolution. With the use of chromatin immunoprecipitation and high-resolution tiling microarray analysis, which cover the genome with short DNA probes, occupancy profiles of factors can be obtained (Jasiak et al., 2008). If the response of a cell to a change in external conditions is followed with kinetic data, primary and secondary target genes can be distinguished. The evaluation of such genome-wide data is often based on a correlation analysis that can lead to testable hypotheses on the structure and behavior of a system.

In the future, continued structure-function analysis *in vitro* will allow us to unravel further details of the mechanisms of transcription and its regulation. Structure-based mutations

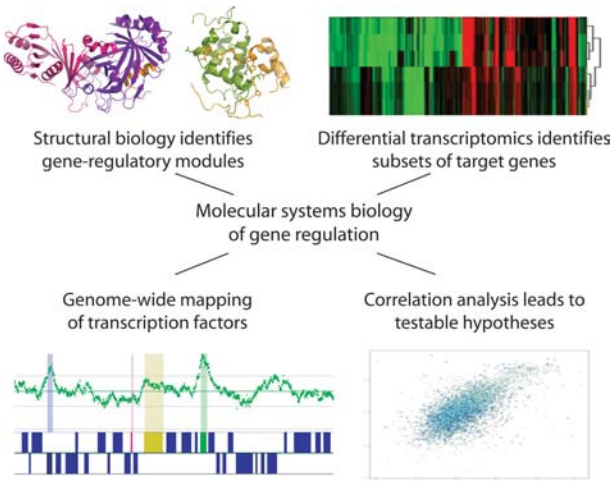

Figure 3 Towards molecular systems biology of gene regulation. Structural subcomplexes of the Mediator complex are identified and characterized by structural biology methods (Lariviere et al., 2006; Larivière et al., 2008; Koschubs et al., 2009) (top left). Target genes of these subcomplexes are identified by differential transcriptome analysis, identifying the subcomplexes as functional submodules (Larivière et al., 2008; Koschubs et al., 2009) (top right). Chromatin immunoprecipitation coupled to high-resolution tiling microarrays is used to map Pol II or transcription factors genome-wide (Jasiak et al., 2008) (bottom left). Correlation analysis allows for generation of testable hypotheses (bottom right).

will then be used to establish the relevance of regulatory mechanisms *in vivo*. Such analysis could be extended from individual genes to the cellular system. With the help of functional genomics, in particular the determination of mRNA synthesis rates and the genome-wide mapping of transcription factors, regulatory networks might be uncovered. Comparative studies of various mutants might lead to a detailed understanding of transcriptional networks on the mechanistic and cellular level. For such research on molecular biosystems, structural and molecular biology must be combined with functional genomics and computational biology in interactive institutions that host experts in various disciplines.

Acknowledgments

I thank all previous and current members of my laboratory for their very good work. I thank Jasmin Sydow, Dirk Kostrewa, and Dietmar Martin for help with preparing the figures. Research in our laboratory is funded by the Deutsche Forschungsgemeinschaft, the SFB646, the SFB TR5, the Nanosystems Initiative Munich NIM, EMBO, the Fonds der Chemischen Industrie, the Boehringer Ingelheim Fonds, the EU network 3D repertoire, the Marie-Curie training network, and the Jung-Stiftung.

References

Andrecka, J., Treutlein, B., Arcusa, M.A., Muschielok, A., Lewis, R., Cheung, A.C., Cramer, P., and Michaelis, J. (2009). Nano-positioning system reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex. Nucleic Acids Res. *17*, 5803–5809.

Brueckner, F. and Cramer, P. (2008). Structural basis of transcription inhibition by α-amanitin and implications for RNA polymerase II translocation. Nat. Struct. Mol. Biol. *15*, 811–818.

Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. Science *315*, 859–862.

Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. Curr. Opin. Struct. Biol. *19*, 294–299.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640–649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. Science *292*, 1863–1876.

Damsma, G.E. and Cramer, P. (2009). Molecular basis of transcriptional mutagenesis at 8-oxoguanine. J. Biol. Chem. *284*, 31658–31663.

Damsma, G.E., Alt, A., Brueckner, F., Carell, T., and Cramer, P. (2007). Mechanism of transcriptional stalling at cisplatin-damaged DNA. Nat. Struct. Mol. Biol. *14*, 1127–1133.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science *292*, 1876–1882.

Jasiak, A.J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Kremmer, E., Strasser, K., Martin, D.E., Soding, J., and Cramer, P. (2008). Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. J. Biol. Chem. *283*, 26423–26427.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell *114*, 347–357.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol. Cell *16*, 955–965.

Koschubs, T., Seizl, M., Lariviere, L., Kurth, F., Baumli, S., Martin, D.E., and Cramer, P. (2009). Identification, structure, and functional requirement of the Mediator submodule Med7N/31. EMBO J. *28*, 69–80.

Kostrewa, D., Zeller, M.E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P. (2009). RNA polymerase II-TFIIB structure and mechanism of transcription initiation. Nature *462*, 323–330.

Lariviere, L., Geiger, S., Hoeppner, S., Rother, S., Strasser, K., and Cramer, P. (2006). Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. Nat. Struct. Mol. Biol. *13*, 895–901.

Larivière, L., Seizl, M., von Wageningen, S., Röther, S., van de Pasch, L., Feldmann, H., Sträßer, K., Hahn, S., Holstege, F.C.P., and Cramer, P. (2008). Structure-system correlation identifies a gene regulatory Mediator submodule. Genes Dev. *22*, 872–877.

Liu, X., Bushnell, D.A., Wang, D., Calero, G., and Kornberg, R.D. (2010). Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism. Science *327*, 206–209.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. Nature *377*, 119–128.

Sydow, J.F., Brueckner, F., Cheung, A.C., Damsma, G.E., Dengl, S., Lehmann, E., Vassylyev, D., and Cramer, P. (2009). Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. Mol. Cell *34*, 710–721.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163–168.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell *127*, 941–954.

Received January 26, 2010; accepted March 31, 2010