4 October 2006

KUNGL.
VETENSKAPSAKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Advanced information on the Nobel Prize in Chemistry 2006

Molecular basis of eukaryotic transcription

Information Department, Box 50005, SE-104 05 Stockholm, Sweden
Phone: +46 8 673 95 00, Fax: +46 8 15 56 70, E-mail: info@kva.se, Website: www.kva.se

Molecular basis of eukaryotic transcription

The Nobel Prize in Chemistry for 2006 is awarded to **Roger Kornberg** for his fundamental studies of the molecular basis of eukaryotic transcription. Transcription is the process in a cell in which the genetic information stored in DNA is activated by the synthesis of complementary mRNA by enzymes called RNA polymerases. Eventually, the mRNA is translated by ribosomes into functional cell proteins. Transcription is one of the most central processes of life, and is controlled by a sophisticated and complex regulatory system. The current needs for proteins of different kinds in the cell, determine when the regulatory system triggers the activation of specific genes. Kornberg has made breakthrough progress in the molecular understanding of transcription and its regulation in eukaryotic cells. His combination of advanced biochemical techniques with structural determinations has enabled the atomic level reconstruction of RNA polymerase from yeast in isolation as well as in a number of functionally relevant complexes with template DNA, product mRNA, substrate nucleotides and regulatory proteins.

Introduction

The early history of eukaryotic transcription starts almost five decades ago with the discovery by Weiss and Gladstone of an RNA polymerase activity in rat liver nuclei (Weiss and Gladstone, 1959). This publication opened transcription as a field for scientific research of crucial importance for processes such as cell differentiation and the development of multicellular organisms as well as for cellular responses to external signals and stimuli.

While it proved difficult to purify RNA polymerase from rat liver nuclei, it was easier to purify the enzyme from bacterial extracts. Therefore, there was an early upsurge in prokaryotic transcription research recognized by the award of the Nobel Prize in physiology or medicine to Jacob, Monod and Lwoff in 1965 for their studies on transcriptional regulation of gene expression in bacteria.

There was a long standing preconception that the gene structure and transcriptional machinery identified in bacteria would be the same in all cells. However, we now know that in eukaryotes, such as yeast and humans, the chromosomal DNA is bound to proteins and packaged in nucleosomes and higher-order forms of chromatin not seen in bacteria. Furthermore, the transcriptional machinery in eukaryotes is much more complex than in bacteria and contains additional levels of regulation. Understanding the mechanism and regulation of RNA polymerase is vital for the understanding of the regulation of transcription, the most important regulatory level for the control of gene expression in a cell.

In contrast to bacteria, eukaryotic cells contain three different forms of RNA polymerase (I-III) as demonstrated by Roeder and Rutter (1969) and by Chambon (Kedinger et al., 1970). All protein-coding genes are transcribed by RNA polymerase II, which is the main target of transcriptional regulation. During the 1970's, continued work on eukaryotic RNA polymerases demonstrated that they are composed of multiple subunits. However,

unlike the bacterial RNA polymerase, the purified eukaryotic enzymes displayed no capacity for selective transcription of purified DNA. Bacterial RNA polymerase consists of a four-subunit core plus a variable fifth subunit called sigma (Burgess et al., 1969). The sigma subunit is required for the core polymerase to recognize a promoter and to initiate transcription. The promoter is a specific sequence of nucleotides in DNA that serves as the start site for RNA synthesis. Surprisingly, no eukaryotic sigma-like factors could be identified. Further progress in eukaryotic transcription therefore required the development of a reconstituted cell-free system capable of RNA polymerase II dependent, promoter-specific transcription on exogenous DNA.

A human tissue culture cell extract which was able, together with purified RNA polymerase II, to specifically initiate transcription at a viral promoter was reported in 1979 (Weil et al., 1979). Biochemical fractionation of this extract revealed the existence of multiple transcription factors for RNA polymerase II (Matsui et al., 1980). These were called general transcription factors since they are involved in the transcription of virtually all genes. The detailed characterization of the RNA polymerase II general transcription factors (TFIIB, D, E, F and H) involved many investigators. Assisted by the five general transcription factors, RNA polymerase II in eukaryotes recognizes the start site on a gene, separates the strands of the DNA template, copies one strand into RNA using ribonucleoside triphosphates as building blocks and finally re-unites the two DNA strands while it translocates along the DNA.

The promoter concept was originally derived from studies of transcription in prokaryotes but eukaryotic are much more extended than the prokaryotic promoters. Studies by Schaffner (Banerji et al., 1981) and Chambon (Moreau et al., 1981) led to the identification of gene-specific ‘enhancer’ elements, i.e. DNA sequences that bind gene activator proteins, which in turn control the transcription of specific genes (Johnson and McKnight, 1989).

Kornberg´s contributions

Roger Kornberg entered the field of transcription when he worked on the structure of chromatin as a post-doctoral student at the MRC in Cambridge, U.K., with Francis Crick and Aaron Klug. At that time, X-ray studies had shown that chromatin consists of repeated units with a size of around 100Å. Fragmentation of chromatin with nuclease resulted in cleavage-products being multiples of a certain size (Hewish and Burgoyne, 1973). Kornberg and Thomas (1974) published the observation that the histones H3 and H4 in solution form a tetramer of the type $(H3)_2(H4)_2$. The same year Kornberg proposed that the basal unit of chromatin (the nucleosome) is a histone octamer and 200 base pairs of DNA (Kornberg, 1974).

After his return to Stanford, the main objective of Kornberg’s research continued to be the understanding of transcriptional regulation in eukaryotes. To take advantage of the versatility of Baker’s yeast *Saccharomyces cerevisiae* as a model eukaryotic organism, he decided to develop an *in vitro* yeast transcription system (Lue and Kornberg, 1987, Sayre et al., 1992). It was soon apparent that the reconstituted system, containing highly purified RNA polymerase II and the five general transcription factors TFIIB, E, F, H and

the TATA-binding protein (TBP), only supported basal transcription and did not respond to the addition of gene-specific activator proteins. This observation led to the unexpected discovery and purification of Mediator, a multiprotein complex composed of ~20 different proteins (Kelleher et al., 1990, Flanagan et al., 1991, Kim et al., 1994). This discovery explained a number of earlier genetic data in eukaryotic transcription. The role of Mediator in all eukaryotes from yeast to humans is to transfer positive and negative signals from DNA-binding, gene-specific transcription factors to RNA polymerase II and the general transcription factors.

With the isolation of Mediator, the three essential components of gene regulation and transcription in eukaryotes had been established; namely the general transcription factors, Mediator and RNA polymerase II. In bacteria, transcriptional repressors and activators directly contact RNA polymerase and affect its binding to the promoter. In eukaryotes, chromatin and Mediator constitute new layers of regulation between the gene-specific transcription factors and RNA polymerase, leading to a much greater complexity of regulation. However, even with the knowledge of all the proteins involved in eukaryotic transcription, there was still little molecular understanding of the process and all the participating polypeptides could merely be depicted as ball-shaped structures assembled around the promoter DNA (Fig. 1)

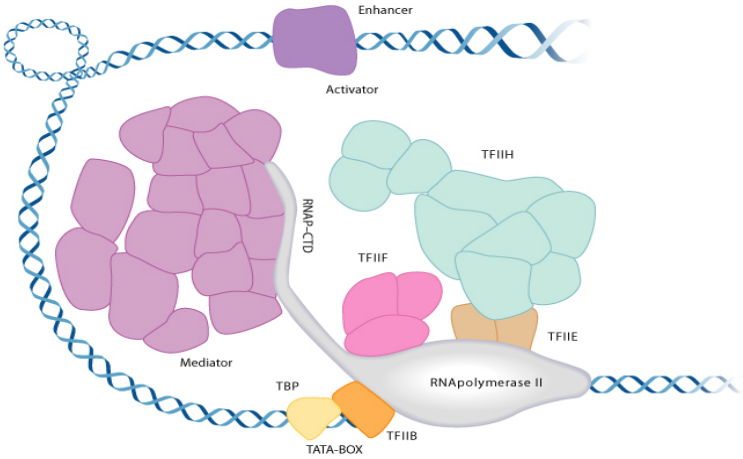

Fig. 1. A cartoon of a eukaryotic transcription initiation complex consisting of DNA, the general transcription factors TBP, TFIIB, E, F and H, Mediator, RNA polymerase II and a specific transcription factor binding to an enhancer element.

The structures of TBP in complex with a piece of TATA-box DNA and the ternary complex of TFIIB,TBP and DNA were solved by X-ray crystallography more than a decade ago (Kim, J.L. et al., 1993, Kim, Y. et al., 1993, Chasman et al., 1993, Nikolov et al., 1995). This provided some information about promoter recognition but very little information about the rest of the transcriptional apparatus. Kornberg realized early on that RNA polymerase might be the platform around which the whole eukaryotic transcription machinery is built. However, the large size of yeast RNA polymerase II – 12 subunits, 0.5x106 daltons – and the scarcity and the instability of the purified complex made structural studies very difficult. Kornberg approached the problem using a

combination of electron microscopy and X-ray crystallography and finally solved the problem after 20 years of biochemical work with protein expression and purification along with the development of a method to form two-dimensional protein crystals on a lipid surface (Uzgiris and Kornberg, 1983). Another important prerequisite was the establishment of the in vitro yeast transcription system mentioned above.
The breakthrough came in 2001. In two papers in *Science*, the structure of a 10-subunit yeast RNA polymerase at 2.8 Å resolution was described as well as an elongating complex consisting of RNA polymerase, template DNA and product RNA (Cramer et al., 2001, Gnatt et al., 2001). In the first structure, the two largest subunits lie in the center on either side of the apparent nucleic acid-binding cleft, with the many smaller subunits on the outside. The cleft is bridged by an α-helix from one of the two large subunits passing through the active site for phosphodiester bond formation and RNA chain elongation. Binding of nucleic acid in the cleft was directly demonstrated in the second structure (Fig 2). Due to the extensive homology between RNA polymerase II in yeast and the corresponding enzymes in mammalian cells, the work on yeast RNA polymerase is expected to offer an excellent model for all multi-subunit eukaryotic RNA polymerases, including those from human cells.

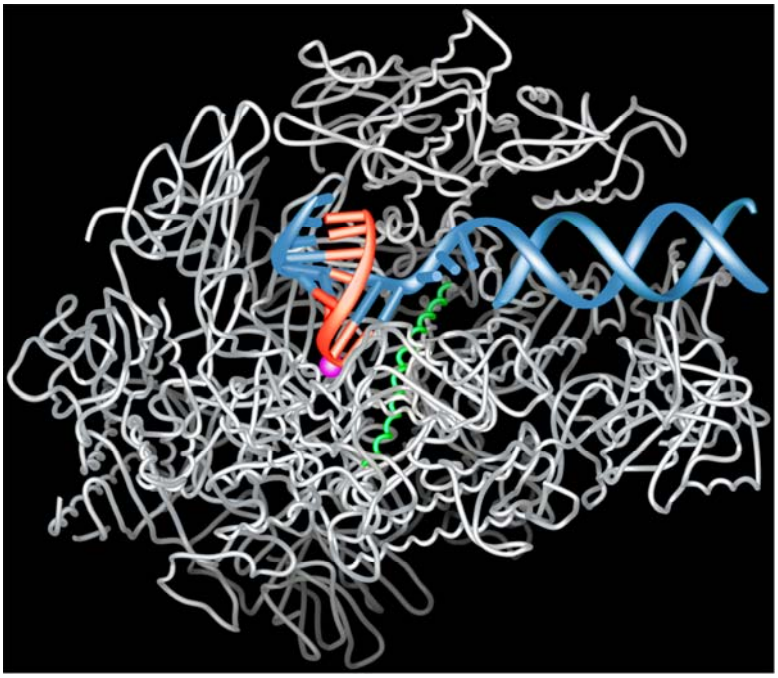

Fig. 2. Structure of an RNA polymerase II transcribing complex. The bridge helix is shown in green and the active site metal as a pink sphere. The DNA helix is colored blue while the newly synthesized RNA is in red.

In a number of later publications from Kornberg's laboratory, nearly a dozen new crystal structures of RNA polymerase have been solved describing different functional complexes with DNA, RNA, nucleotides or other proteins. This multitude of data has enabled a dynamic interpretation of the transcription process. The first crystal structure of an elongating complex was obtained using a DNA duplex with a long 3′ overhang on one

of the strands. The transcribing polymerase molecules were paused at a specific site on the template by withholding one of the four ribonucleotides. The structure represents the molecular conditions immediately after the formation of a phosphodiester bond between the 3´-end of the RNA chain and the incoming ribonucleotide *before* translocation of the DNA-RNA hybrid has occurred (Gnatt et al., 2001). In a recent crystal structure of a transcribing complex, the enzyme was combined with a 15-nucleotide DNA and a 9-nucleotide complementary RNA in the presence of a chain-terminating 3´-deoxyadenosin nucleotide. In this complex, the binding site for the incoming nucleotide is unoccupied and it represents a stage *after* translocation of the DNA-RNA hybrid (Westover et al., 2004a).

Translocation. Translocation of the DNA-RNA hybrid helix during transcription is proposed to involve the earlier mentioned bridge helix acting as a ratchet (Gnatt et al., 2001, Bushnell et al., 2002). The side chains of two residues in the bridge helix beneath the active center – threonine-831 and alanine-832 – contact the nucleotide base at the end of the hybrid region. The helix is straight in yeast RNA polymerase crystal structures but bent in the structure of a bacterial RNA polymerase reported by Zhang et al. (1999). Oscillation of the bridge helix between straight and bent states would move the DNA-RNA hybrid by 3 Å or one nucleotide step. This mechanism fits in with later biochemical and genetic studies (Epshtein et al., 2002, Bar-Nahum et al., 2005).

Strand separation. The post-translocation state structure (Westover et al., 2004a) shows how a lid-like loop in the polymerase separates the newly formed RNA-chain from the DNA template at position -8, resulting in a complete separation of the two strands at positions -9 and -10. The last nucleotide added to RNA has the position -1. Together with two other protein loops, the lid loop prevents the reassociation of the RNA and DNA strands.

Nucleotide selection. Nucleoside triphosphates diffuse to the active center of RNA polymerase through a funnel-shaped opening in the structure. By soaking crystals of the transcribing complex in the post-translocation state with ribonucleoside triphosphates matched to the DNA base at position +1 as well as with unmatched nucleotides, it was possible to study how nucleotides are selected in the active site of the enzyme (Westover et al., 2004b). The selection appears to occur via a two-step mechanism, where the nucleotide is initially bound in an inverted orientation to an entry site below the active site. This binding greatly enhances the lifetime of a nucleotide in the active site region (Batada et al., 2004). Binding to the entry site is followed by a rotation of the nucleotide into the active site where it can bind selectively to the DNA template. If there is a match between the incoming and template bases, a phosphodiester bond is formed by two-metal ion catalysis as proposed for all nucleic acid polymerases (Steitz, 1998). The fidelity in transcription is determined by basepairing and properties of the RNA-DNA hybrid. The protein-bound hybrid helix adapts a non-standard conformation intermediate between the conformations of A- and B-DNA. This form is underwound compared to the structure of a free DNA-RNA hybrid (Gnatt et al., 2001). Incorporation of a mismatched base would lead to a destabilization of the hybrid helix, and is thus selected against.

Promoter recognition, abortive initiation and promoter clearance. The RNA polymerase II structure is important not only for understanding the transcription mechanism. It also opens up the possibility of extending the structural studies to larger functional complexes with general transcription factors and Mediator which is a key to understanding transcription regulation in eukaryotes. Kornberg has already taken the first step in this direction by determining the structure of the general transcription factor TFIIB in complex with RNA polymerase II (Bushnell et al., 2004). Under special conditions, TFIIB and TBP can form a minimal initiation complex with RNA polymerase and DNA. The crystal structure illustrates the mechanism of transcription initiation. Only one α-helical region of the C-terminal domain of TFIIB is clearly resolved in the co-crystal but this allowed the previously determined crystal structure of a TFIIB-TBP-DNA complex (Tsai and Sigler, 2000) to be docked to the structure. TBP bends the TATA box region directing the downstream DNA towards the active site. The bend fits the DNA to the contours of the polymerase and a spacing of about 25 nucleotides is observed between the TATA box and the transcription start site. This spacing is characteristic of most RNA polymerase II promoters.

Previous observations such as abortive initiation and promoter escape can now be explained. Abortive initiation is the failure of many transcripts to grow past 10 residues and promoter escape is the departure of the transcribing RNA polymerase from the promoter after synthesis of about 10 residues of RNA. A loop formed from the N-terminal part of TFIIB stretches into the active site of the polymerase where it competes with the newly formed DNA/RNA hybrid. If the loop wins, the RNA chain is displaced and this results in abortive initiation. However, if the RNA chain wins, the loop is displaced together with TFIIB and promoter DNA, allowing RNA polymerase to escape from the promoter and start elongation.

Outline of the transcription initiation complex. Combining results of X-ray diffraction studies and electron microscopy, Kornberg and coworkers have also presented a model of the transcription initiation complex including all five general transcription factors (Bushnell et al., 2004). A remarkable observation is that the five general transcription factors appear to bind directly to the double-stranded DNA promoter while RNA polymerase can bind to DNA only after melting of the helix. The melting allows bending of the DNA template by 90^{o}, making it possible for the template strand to enter the active site cleft in the polymerase.

Electron microscopy of an RNA polymerase II – Mediator complex depicts Mediator as a crescent-shaped structure around the polymerase (Asturias et al., 1999). Mediator consists of ~20 polypeptides with a total mass of 1 MDa making it difficult to study. Recently, the expression and purification of a functional Mediator head module was reported. This complex consists of only 7 subunits with a mass of 223 kDa (Takagi et al., 2006). Interaction between the Mediator head module and RNA polymerase requires the presence of the general transcription factor TFIIF. However, a full molecular understanding of the role of Mediator will require an extension of polymerase crystallography to include the Mediator complex.

Final remarks

As a result of Roger Kornberg´s mechanistic studies of eukaryotic transcription, for the first time we now have a molecular understanding of promoter recognition, the mechanism of transcription initiation, how the DNA-RNA hybrid is translocated after the addition of a nucleotide, how the newly synthesized RNA strand is separated from the DNA template, and of the structural basis for accurate selection of an incoming ribonucleotide that is complementary to the DNA template. Furthermore, the structure of RNA polymerase II is the basis for the next generation of research to determine the precise operations of the general transcription factors and Mediator in transcriptional regulation.

Kornberg has already reported the X-ray structure of a co-crystal between RNA polymerase and TFIIB and electron microscopy structures of complexes between the polymerase, other general transcription factors and Mediator. High resolution X-ray structures of a complete functional transcription machinery will ultimately reveal the molecular mechanisms of transcriptional regulation. This knowledge is important since disturbances in transcriptional regulation contribute directly to cancer, inflammation, heart disease and metabolic diseases. Regulation of transcription also controls the development of stem cells into specific cells with a certain function in different organs. To benefit fully from the potential of stem cell transplantation to cure disease, it is essential to be able to direct this development to form only the wanted type of cells.

References

Asturias, F.J., Jiang, Y.W., Myers, L.C., Gustafsson, C.M. and Kornberg, R.D. (1999) Conserved structures of Mediator and RNA polymerase II holoenzyme. Science 283, 985-987.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27, 299-308.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A. and Nudler, E. (2005) A rachet mechanism of transcription elongation and its control. Cell 120, 183-193.

Batada, N.N., Westover, K.D., Bushnell, D.A., Levitt, M. and Kornberg, R.D. (2004) Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. Proc. Natl. Acad. Sci. USA 101, 17361-17364.

Burgess, R.R., Travers, A.A., Dunn, J.J. and Bautz, E.K. (1969) Factor stimulating transcription by RNA polymerase. Nature 221, 43-46.

Bushnell, D.A., Cramer, P. and Kornberg, R:D. (2002) Structural basis of transcription: α-amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. Proc. Natl. Acad. Sci. USA 99, 1218-1222.

Bushnell, D.A., Westover, K.D., Davis, R.E. and Kornberg, R.D. (2004) Structural basis of transcription: An RNA polymerase II – TFIIB cocrystal at 4.5 angstroms. Science 303, 983-988.

Chasman, D.I., Flaherty, K.M., Sharp, P.A. and Kornberg, R.D. (1993) Crystal structure of yeast TATA-binding protein and model for interaction with DNA. Proc. Natl. Acad. Sci. USA 90, 8174-8178.

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

Epshtein, V., Mustaev, A., Markovtsov, V., Bereshchenko, O., Nikiforov, V. and Goldfarb, A. (2002) Swing-gate model of nucleotide entry into the RNA polymerase active center. Mol. Cell 10, 623-634.

Flanagan, P.M., Kelleher III, R.J., Sayre, M.H., Tschochner, H. and Kornberg, R.D. (1991) A mediator for activation of RNA polymerase II transcription *in vitro*. Nature, 350, 436-438.

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.

Hewish, D.R. and Burgoyne, L.A. (1973) Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. Biochem. Biophys. Res. Commun. 52, 504-510.

Johnson, P.F. and McKnight, S.L. (1989) Eukaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. 58, 799-839.

Kedinger, C., Gniazdowski, M., Mandel, J.L., Gissinger, F. and Chambon, P. (1970) α-amanitin: a specific inhibitor of one of the two DNA-dependent RNA polymerase activities from calf thymus. Biochem. Biophys. Res. Commun. 38, 165-171.

Kelleher III, R.J., Flanagan, P.M. and Kornberg, R.D. (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. Cell 61, 1209-1215.

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H. and Kornberg, R.D. (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell 77, 599-608.

Kim, J.L., Nikolov, D.B. and Burley, S.K. (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature 365, 520-527.

Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA box complex. Nature 365, 512-520.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. Science 184, 868-871.

Kornberg, R.D. and Thomas, J.O. (1974) Chromatin structure; oligomers of the histones. Science 184, 865-868.

Lue, N.F. and Kornberg, R.D. (1987) Accurate initiation at RNA polymerase II promoters in extracts from *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA 84, 8839-8843.

Matsui, T., Segall, J., Weil, P.A. and Roeder, R.G. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. 255, 11992-11996.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. (1981) The SV40 72 base pair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acid Res. 9, 6047-6068.

Nikolov, D.M., Chen, H., Halay, E.D., 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.

Roeder, R.G. and Rutter, W.J. (1969) Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature 224, 234-237.

Sayre, M.H., Tschnochner, H. and Kornberg, R.D. (1992) Reconstitution of transcription with five purified protein factors and RNA polymerase II from *Saccharomyces cerevisiae.* J. Biol. Chem. 267, 23376-23382.

Steitz, T.A. (1998) A mechanism for all polymerases. Nature 391, 231-232.

Takagi, Y., Calero, G., Komori, H., Brown, J.A., Ehrensberger, A.H., Hudmon, A., Asturias, F.J. and Kornberg, R.D. (2006) Head module control of Mediator interactions. Mol. Cell 23, 355-364.

Tsai, F.T.F. and Sigler, P.B. (2000) Structural basis of preinitiation complex assembly on human Pol II promoters. EMBO J. 19, 25-36.

Uzgiris, E.E. and Kornberg, R.D. (1983) Two-dimensional crystallization technique for imaging macromolecules, with an application to antigen-antibody-complement complexes. Nature 301, 125-129.

Weil, P.A., Luse, D.S., Segall, J. and Roeder, R.G. (1979) Selective and accurate initiation of transcription at the Ad_2 major late promoter in a soluble system dependant on purified RNA polymerase II and DNA. Cell 18, 469-484.

Weiss, S.B. and Gladstone, L. (1959) A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J. Am. Chem. Soc. 81, 4118-4119.

Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004a) Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II. Science 303, 1014-1016.

Westover, K.D., Bushnell, D.A. and Kornberg, R.D. (2004b) Structural basis of transcription: Nucleotide selection by rotation in the RNA polymerase II active center. Cell 119, 481-489.

Zang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K. and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell 98, 811-824.

Further reading

Boeger, H., Bushnell, D.A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J.S., Westover, K.D. and Kornberg, R.D. (2005). Structural basis of eukaryotic gene transcription. FEBS Lett. 579, 899-903.

Kornberg, R.D. (2005) Mediator and the mechanism of transcriptional activation. Trends Biochem. Sci. 30, 235-239.

Film of transcription: The Dolan DNA learning center – genes in education.
http://www.dnalc.org/home.html
Media showcase; Transcription: DNA codes for mRNA, 3D animation.

Lars Thelander
Department of Medical Biochemistry and Biophysics
Umeå University

Member of the Nobel Committee for Chemistry