RNA Polymerase II – The Transcription Machine

Nobel Prize in Chemistry 2006

Jiyoti Verma[1], *Aruna Naorem*[2], *Anand Kumar*[3], *Manimala Sen*[4] *and Parag Sadhale*[5]

It was 2005 December. The distinguished speaker at the "Transcription Assembly" meeting at the Center for Cellular and Molecular Biology in Hyderabad was about to deliver his plenary lecture. We engaged ourselves in small talk with him. Since he had just arrived in the early hours of the day and probably had been traveling continuously to reach Hyderabad, an obvious opening question for the conversation was "How are you doing? You must have hardly got any sleep". The answer came rather sharply "Well, it wasn't too bad. Any way I sleep only for about four hours a day!" We listened to the marvelous exposition of the structure of one of the most complex and intriguing machines in biology, the eukaryotic RNA polymerase II by Prof. Roger Kornberg. Little did we know then that the very next year he was to receive the coveted Noble Prize in Chemistry for his work on elucidating the structure and function of RNA polymerase II!

RNA polymerases are enzymes that carry out the first step of the 'Central Dogma' of molecular biology (*Figure* 1), i.e. copying the genetic information present in the DNA into RNA, a macromolecule that is either translated into proteins or participates in the translation of other RNAs. This class of enzymes has gone through a large number of different forms during evolution. The simplest of these enzymes are found in bacteriophages (bacteria eaters), viruses that infect bacteria and are entities on the borderline of the living and the non-living. Bacteriophages exist as complexes of proteins and nucleic acids and in that sense are not alive. But once inside the host they reproduce as if they are alive. They have RNA polymerases made up of single polypeptides large enough to carry out all the biochemical reactions to faithfully copy the genetic material of the virus.

Jiyoti Verma, Aruna Naorem and Manimala Sen are research students, S Anandkumar is a postdoctoral fellow and Parag Sadhale is Associate Professor in the Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore. The research interest of the group is transcriptional regulation of stress response, involving polymerase subunits.

Key words
RNA polymerase II, Mediator, Roger Kornberg.

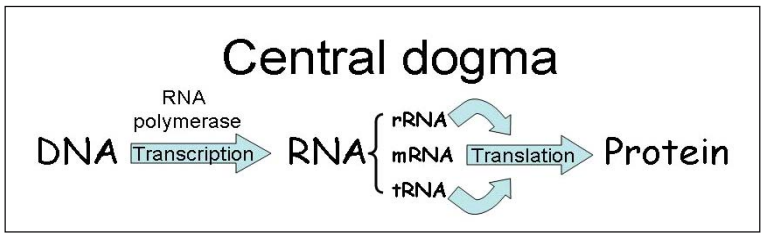

Figure 1. The 'Central Dogma' of molecular biology. Genetic information encoded in the DNA is first transcribed into RNA which is subsequently translated into proteins (information flows from DNA to protein via RNA). The step of translation is irreversible, i.e., information cannot flow from proteins to the genes.

The hosts of bacteriophages, the bacteria, are among the simplest living organisms and have about 100-1000 fold more DNA than the bacteriophages. Therefore, not surprisingly, they encode a lot more genetic information that needs to be converted into RNA at the appropriate time and in appropriate amounts. This increase in the complexity of the DNA has necessitated the bacterial RNA polymerase to increase in complexity to a great extent. The typical RNA polymerase in bacteria is made up of 5 different polypeptides ranging from 150 to 35 kDa in size[1]. The core enzyme can carry out all the biochemical reactions required to copy DNA into RNA, but is unable to recognize the specific sequences called promoters from where the copying process or transcription is initiated. The RNA polymerase depends on a subunit called the Sigma factor which helps it to recognize and specifically bind to the promoter. This binding allows the polymerase to make RNA with the correct beginning that ensures that the correct protein is synthesized from it.

[1] These are α, β, β', σ and ω. The core enzyme is the polymerase that does not have the σ subunit.

The eukaryotic genomes are anywhere from 3-4 to a thousand times bigger than the typical bacterial genome. The increased size of the genome has led to an increase in the complexity of organization and regulation of these genomes. In fact the eukaryotes have three different RNA polymerases to carry out the synthesis of the three types of RNAs[2]. Each of these three polymerases called Pol I, Pol II and Pol III in turn are made up of 10 or more subunit proteins. These three activities were identified decades ago in mammalian extracts based on their difference in their sensitivity to the antibiotic amanitin. It was later realized that these enzymes are present in all eukaryotic organisms. A fourth polymerase, Pol IV, has recently been discovered in plants and mammalian cells.

[2] These are ribosomal RNA (rRNA), messenger RNA (mRNA) and transfer RNA (tRNA).

Studies on Eukaryotic Transcription

If one follows Kornberg's research spanning about 40 years through close to 200 publications, one sees almost a directed effort towards a deeper understanding, almost visualization, of the components of eukaryotic transcription. Earlier he provided evidence to support the concept of the nucleosome hypothesis which states that eukaryotic DNA is tightly packaged around a core of structural proteins called histones to generate an array of nucleosomes to form chromatin, the functional state of the eukaryotic genome. His findings provided more insights into the structure of the nucleosome. He showed that the histones oligomerize resulting in a histone octamer that interacts with a defined length of DNA to form nucleosomes. His studies helped to establish correspondence between the beads-on-a-string appearance of chromatin in electron micrographs and the biochemical studies on nucleosomes. After a significant gap he has revisited the nucleosomes. He recently showed their interactions with highly refined systems and demonstrated that during transcriptional activation, the nucleosomes actually are disassembled from the promoters and not just moved away by sliding.

The Yeast as Model for Eukaryotic Transcription

After several years of working with mammalian systems such as rat liver extract for the biochemical analyses of chromatin and associated proteins, he realized early in his career the power of yeast as a model system for biochemical analyses. His major contribution in the field of eukaryotic transcription started three decades ago when he exploited yeast to study the biochemical and structural basis of eukaryotic transcription. His initial work helped elucidate the binding site as well as other biochemical features of one of the best characterized eukaryotic transcriptional activators, GAL4. The major hurdle in studying transcription was the isolation of transcriptional machinery, i.e., RNA polymerase II and its associated proteins, which is responsible for the synthesis of all mRNA. This was partly due to their larger size and low abundance in the cell.

Born in 1947, Roger Kornberg is the first of three children of Arthur Kornberg (who won the Nobel Prize in 1959 for his discovery and characterization of DNA Polymerase, the enzyme responsible for DNA replication) and Sylvy who is also a biochemist. Kornberg once said. "Science was a part of dinner conversation and an activity in the afternoons and on weekends. Scientific reasoning became second nature. Above all, the joy of science became evident to my brothers and me."

In order to assign functions to purified components of the yeast transcription machinery *in vitro*, he needed an assay for their transcriptional activity. He developed a method for the isolation of transcriptionally active yeast extracts in the late 1980s. This accelerated the biochemical studies of yeast transcription and helped establish the universality of the transcription machinery among the eukaryotic systems as evolutionarily distant as yeast and mammals. This in other words validated the use of the yeast system as a model for studying eukaryotic transcription in general. This further led to the purification of various factors required during transcription initiation by RNA polymerase II. After many years of hard work, Kornberg and his coworkers succeeded in the biochemical and genetic characterization of factors, now known as the 'General Transcription Factors' (GTFs), like TFIID, TFIIB, and TFIIE, which are required by the polymerase II at all promoters. Several other investigators have also contributed to this area of research.

During that time it was generally accepted that similar to prokaryotic systems, eukaryotic transcriptional activators make contact with enhancer DNA element and stimulate the initiation of transcription by interacting with general transcription factors at a promoter. In the course of studies to understand how basal transcription is enhanced by the presence of activators, Kornberg's group proposed the requirement of an intermediary molecule(s) distinct from the general transcription factors for the function of activator, which they termed as 'Mediator'[3]. Subsequently, their investigations led to the purification of the multi-protein mediator complex and reconstitution of first *in vitro* activator-dependent transcription system using homogenous basal factors and RNA polymerase II. Interestingly the development of the mediator concept through biochemical analyses was also supported by genetic studies by others which identified through genetic screens, components of the mediator that suppressed defects in mutants of Pol II. Mediator as we know today is a large complex comprising ~20 proteins with a total mass of 1 million Daltons. Kornberg and coworkers have contributed largely to the

[3] Mediator is a component of eukaryotic transcription machinery that was originally thought to be required only for regulated transcription but is now believed to be required for transcription from most if not all promoters in eukaryotic systems.

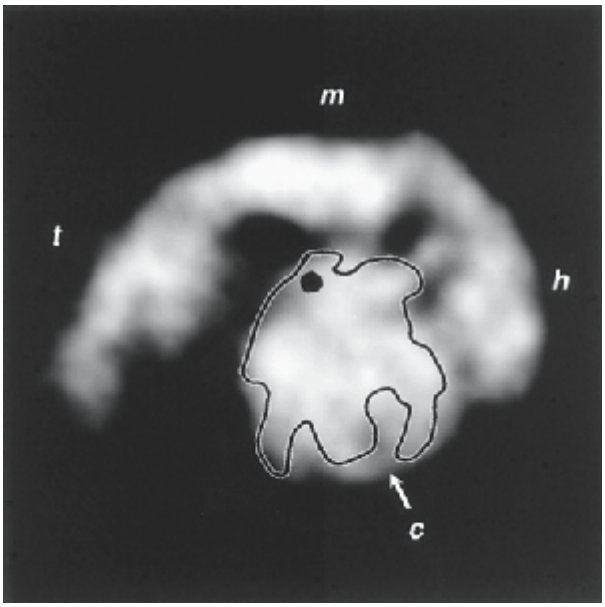

Figure 2. Electron microscope picture of the purified complex of RNA polymerase II and the mediator complex, in total containing about 35 proteins. The bar represents 200A. (1Å = 10^{-10} meter.)
Courtesy:R Kornberg

biochemical characterization of several mediator proteins and helped in providing the molecular basis for understanding the functions of several mediator proteins. Moreover, they also carried out various structural studies for the mediator complex and its interaction with RNA polymerase II to understand how the mediator complex interacts with the RNA polymerase II and relays to it the signals from regulator proteins recruited at regulatory elements on the DNA.

Solving the RNA polymerase II structure

Kornberg initiated his quest to solve RNA polymerase II structure and its transcribing complex in the early 1990s. Using lipid as a platform for the formation of two dimensional crystals of purified RNA polymerase II, the Kornberg group was able to provide the first low resolution structure (30Å resolution) of yeast RNA polymerase II. This two dimensional structure of RNA polymerase yielded the possibilities of obtaining the three dimensional structure of this important protein. However, to achieve this goal, several hurdles had to be overcome. First and foremost was to isolate yeast RNA polymerase II to homogeneity in milligram amounts. The second major problem to obtain struc-

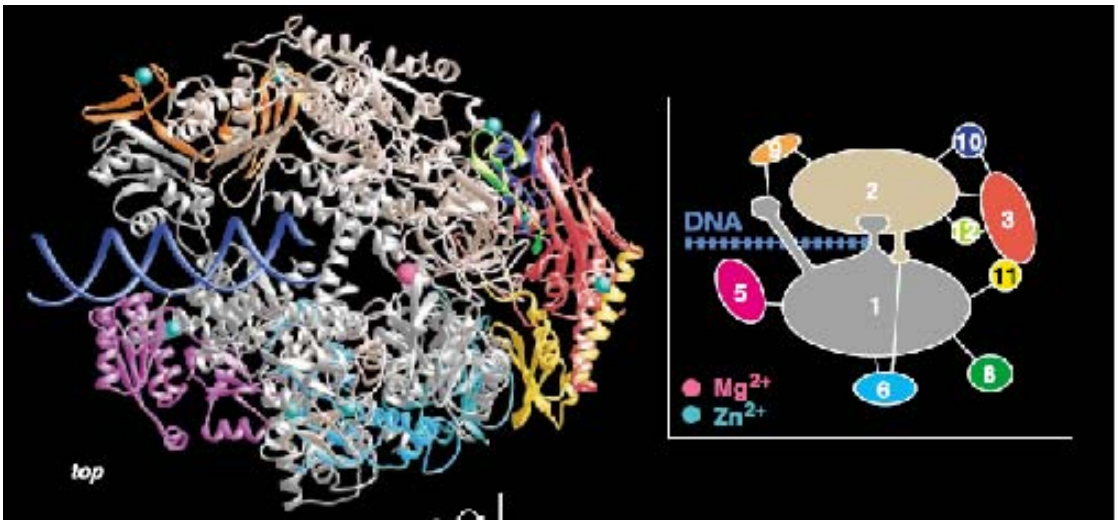

Figure 3. High resolution crystal structure of RNA polymerase II. This structure was solved from crystals that contained only 10 out of the12 subunits of the yeast polymerase. The schematic on the right shows the number of each subunit as seen in the crystal structure. The DNA was not present in the actual crystal but has been introduced in the model based on earlier experiments.
Courtesy:R Kornberg

ture of this 12 subunit enzyme was the sub-stoichiometric amounts of the fourth and seventh largest subunit (Rpb4 and Rpb7 respectively). So the first high resolution (2.8 Å resolution) structure of yeast RNA polymerase was only for ten subunits. This structure gave several insights about RNA polymerase and suggested that polymerase can be divided into four modules including the core, the clamp, the shelf and the jaw- lobe. It also suggested that the clamp serves as a multifunctional element, permits entry of the promoter DNA, sensing the DNA-RNA hybrid conformation and separating DNA and RNA strands at the upstream end of the transcription bubble. The structure of the complete 12 subunit enzyme was also contributed by Kornberg and coworkers at 4.1Å resolution and then further even at higher resolution. This complete structure provided greater insights into polymerase function in transcription.

Kornberg and several of the post-docs who trained with him and who have gone on to be independent investigators together have contributed tremendously to our understanding of the structural details of the transcription machinery in eukaryotes. Kornberg also achieved making an entirely synthetic yeast transcription system comprising of over 40 proteins in the late 90s, an evidence of our level of understanding of the requirements of transcription

Box 1. Other Contributions of Kornberg

Although not directly related to the work that won him the Nobel Prize, his contribution to an important phenomenon in biology is invaluable. One of the interesting questions that cell biologists have pondered over the years is how proteins destined to different compartments in a given cell reach their destination. The nucleus being the seat of regulation of gene expression, the localization of proteins involved in gene expression into the nucleus has attracted a lot of attention. In the mid 1980s, in one of Roger Kornberg's studies, he described how nucleoplasmin, one of the most abundant proteins in the nucleus of *Xenopus laevis* (frog) oocyte, is targeted to the nucleus. For this he constructed a fusion protein between the enzyme beta-galactosidase and nucleoplasmin using recombinant DNA technology. By following the distribution of this fusion protein injected in *X. laevis* oocytes as a function of time, he demonstrated that the signal sequence that directs the protein to the nucleus is located close to one end of the protein. He could also map the signal sequence to a stretch of four lysine residues. He further showed that the process is highly specific and a single point mutation that converts Lysine 128 to Threonine or Asparagine prevents nuclear import. Conversely, a synthetic peptide with the signal sequence could cause non-nuclear proteins and even gold particles to be localized to the nucleus. These studies have had great impact on our understanding of the nuclear localization of proteins

in a eukaryotic system. Beyond solving the RNA polymerase II structure, Kornberg has also solved the structure of RNA polymerase with general transcription factors, mediator as well as with promoter DNA. These studies have had tremendous impact on our understanding of the mechanism of eukaryotic transcription as we know today.

Suggested Reading

[1] http://nobelprize.org/nobel_prizes/chemistry/laureates/2006/kornberg-lecture.html

[2] P Cramer, D A Bushnell and R D Kornberg, *Science,* Vol.292, No.5523, pp.1863–76, 2001.

[3] D A Bushnell and R D Kornberg, *Proc. Natl. Acad Sci. U S A,* Vol. 100, No.12, pp.6969–73, 2003.

[4] R D Kornberg, *Trends Biochem Sci.,* May, Vol.30, No.5, pp.235–9, 2005.

Address for Correspondence
Parag Sadhale
Department of Microbiology and Cell biology,
Indian Institute of Science,
Bangalore 560 012, India
Email:pps@mcbl.iisc.ernet.in