Upstream, Downstream: The Intricate Affairs Of Busy Transcription Factors

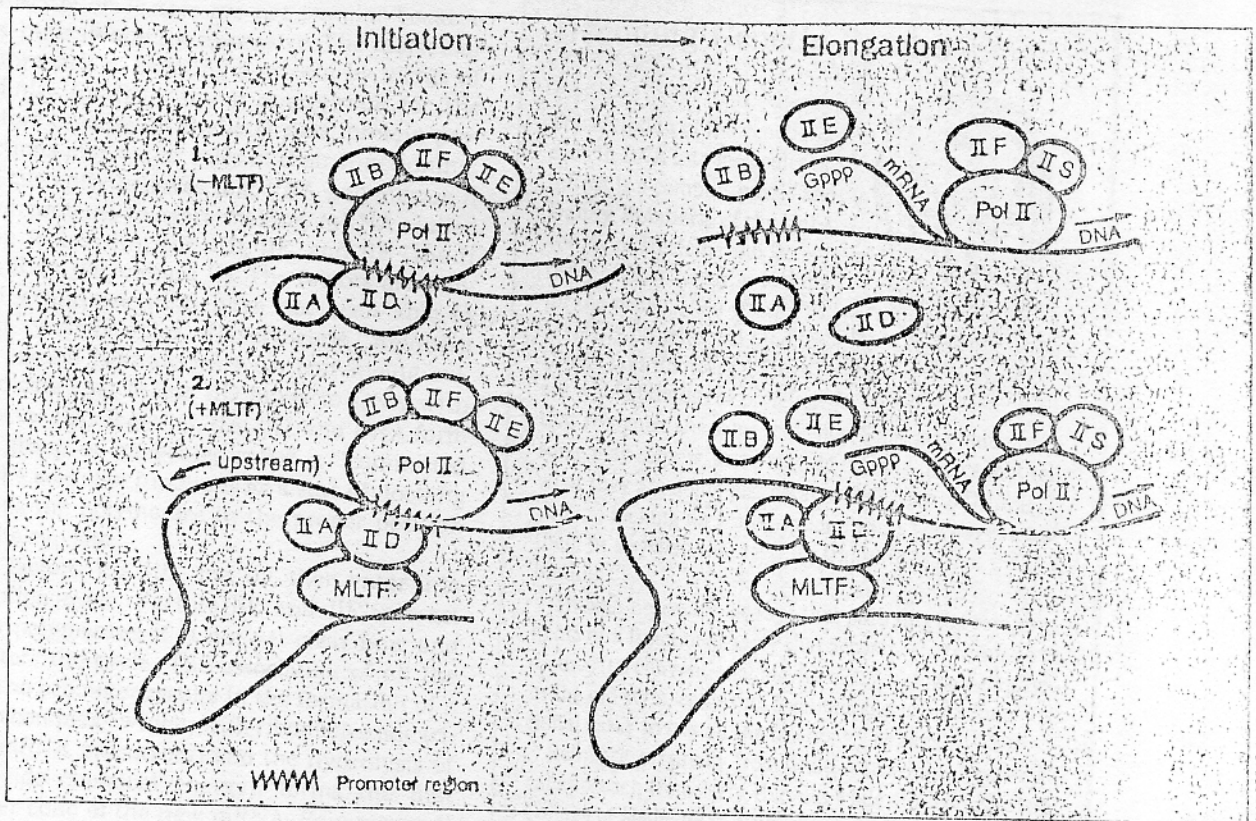
...the ongoing drama revolving around a dedicated protein family, meddling neighbors, and the genes they serve. In this week's episode, MLTF, an upstream DNA-binding protein, pressures transcription factor IID into working overtime at an adenovirus promoter.

J. Carcamo, S. Lobos, A. Merino, L. Buckbinder, R. Weinmann, V. Natarajan, and D. Reinberg, "Factors involved in specific transcription by mammalian RNA polymerase II—role of factors IID and MLTF in transcription from the adenovirus major late and IVa2 promoters," *J. Biol. Chem.* 264, 7704 (1989).

ALMOST EVERY WEEK, IT SEEMS, another character appears, expanding the cast of DNA-binding proteins. Some of these proteins are destined for major roles, binding to DNA sequences a few hundred basepairs upstream of key genes and controlling the work of cells by regulating transcription of DNA to messenger RNA. Other dramatis personae in the cellular theater take their places in enhancer regions thousands of basepairs upstream or downstream from the transcribed gene on center stage. Clearly, untwisting the labyrinthine affairs of the protein families that regulate gene expression will not be easy, but Danny Reinberg and colleagues at the University of Medicine and Dentistry of New Jersey think a plot resolution may soon be in sight.

In a recent installment, published in the *Journal of Biological Chemistry*, Reinberg, Juan Carcamo, also of the University of Medicine and Dentistry, and their colleagues identify a new role for a transcription factor that has appeared before. Known as a general transcription factor, IID has now been spotted interacting with an unusual promoter region in an in vitro system from adenovirus. The promoter region figures into the prologue of the transcription drama: It is the collection of DNA sequences at which transcription begins. In his review of the protein players and their script, Ron Conaway at the University of Texas at Austin anticipates that the new IID twist will open questions about the promoter region and IID's interaction with the region. He concludes on a philosophical note: There is much to learn about this old and vital biological drama.

The new results also confirm a previous observation of transcription watchers. In this case the character involved is the major late transcription factor (MLTF). Unlike the general factor IID, which binds at the promoter, MLTF binds to DNA at a distance before (upstream from) the promoter region. Even from this re-



Model of transcription, based on studies by Reinberg and others, of the adenovirus promoter in a reconstituted system. *(1) The general transcription factors IIA and IID form a docking site at which RNA polymerase II (pol II) associates to begin transcription of mRNA from DNA (initiation). The general transcription factors IIB, IIF, and IIE interact in an unknown manner with RNA polymerase. When the polymerase moves down the DNA to extend the initiated mRNA (elongation), the IIA-IID docking site falls apart, IIB and IIE fall off the polymerase, and factors IIF and IIS now bind to keep RNA polymerase II transcribing. (2) When an upstream factor such as the major late transcription factor (MLTF) binds to an upstream element on DNA, it may also contact the IIA-IID docking site. Under these conditions, the docking site remains stably anchored to the DNA when the polymerase moves on to elongate mRNA, and is waiting for another polymerase molecule to start another round of transcription. In the adenovirus promoter system, this results in the production of 10 to 12 times more mRNA than would result in the absence of MLTF.*

ILLUSTRATION BY MARY PUDUPE, RESEARCH BY DAVE LOWE GALLAGHER

mote location a few hundred base pairs away, MLTF interacts with factor IID at the promoter, increasing the amount of RNA that is made. Scientists expect they will eventually see more examples of such interplay between upstream or downstream factors and events unfolding at the promoter region.

The ultimate goal, says Reinberg, is to understand how the enzyme RNA polymerase II makes messenger RNA (mRNA) from genes during transcription. The key, he says, is to know more about events at the promoter region of the DNA where RNA polymerase II starts making mRNA. Then, researchers can understand how specific DNA-binding proteins such as MLTF operate from upstream or downstream locations to regulate the amounts of specific mRNAs produced.

Reinberg and colleagues study several proteins called "general transcription factors" that associate with RNA polymerase II as it makes mRNA from class II genes. These are genes that code for proteins, as opposed to class I genes—which code for large ribosomal RNA (rRNA)—and class III genes—which code for transfer RNA (tRNA) and other small rRNAs. Class I and class III genes are transcribed by RNA polymerases I and III, respectively.

These general factors, including IID, were initially discovered in Robert G. Roeder's laboratory at Rockefeller University in New York and are necessary for transcription of all class II genes. (This distinguishes them from the upstream DNA-binding factors such as MLTF, which are specialized; each may be involved in the regulation of only one or a few genes.) The general factors can bind to the DNA, to each other, or to RNA polymerase II, but they apparently function only in the promoter region. This group of DNA elements is located just before, and may include, the so-called "cap" site that marks the beginning of each mRNA. Each mRNA starts with a guanylate triphosphate, Gppp, that is modified into a cap structure before splicing events remove intervening sequences called introns.

Reinberg says that the general factors have been difficult to isolate because, unlike the specialized DNA-binding proteins, there is no high affinity probe that can be used to fish them out of nuclear extracts. Michael Chamberlin, of the University of California at Berkeley agrees. In defense of criticisms that the field is

progressing too slowly, he says that experiments with general transcription factors are not popular because they are too difficult and too expensive. The molecular biological approach of isolating a protein by virtue of its high affinity for DNA is not appropriate for the general factors because, in the later case, protein-protein interactions play major roles in their functioning.

Reinberg currently believes that RNA polymerase II requires five proteins to start making mRNA, called transcription factors IIA, IIB, IID, IIE, and IIF (the II comes from RNA polymerase II). The factors are difficult to purify, and scientists have not been able to clarify their various roles. Nevertheless, Reinberg thinks that several laboratories have convincing evidence to indicate that factor IID binds to DNA with help from factor IIA. Together, IIA, IID, and DNA form a stabilized group of molecules called a "committed complex." The complex functions as a sort of docking site that binds the polymerase in the proper position relative to the cap site.

The other factors, IIB, IIE, and IIF, then enter into the transcription cycle. Factors IIE and IIF interact directly with polymerase; it is not yet clear how and when IIB enters the cycle. The association of all the factors with the promoter results in the formation of a "preinitiation complex" which, with ribonucleoside triphosphates (mRNA precursors) and the hydrolysis of ATP (to provide chemical energy), forms the first phosphodiester bond of the new mRNA. Polymerase II is then released from the initiation complex and begins to elongate the mRNA by continued polymerization of nucleoside triphosphates with help from transcription factors IIF and IIS.

In addition to the general factors, the promoter region on the DNA helps to determine the efficiency of transcription. Keith Yamamoto of the University of California at San Francisco describes the promoter as a region made like a patchwork quilt of sequence elements: Promoter regions vary, and a given promoter is made of multiple patches.

Many promoters of class II genes have a sequence element with repeated thymidines and adenosines called a TATA box. It, or variations such as TATAAA or ATAAA, is located about 30 base pairs before the cap site. In several previous studies, Reinberg and his colleagues used a promoter with a TATA box, the adenovirus major late promoter, to assay the general transcription factors in vitro. In vivo, it promotes transcription of the mRNA for most of the proteins required in the late part of the adenovirus life cycle.

Adenoviruses have several promoters that could be used in vitro to assay general transcription factors, but researchers often use the major late promoter because they believe it to be relatively simple and efficient. Other investigators reported that factor IID binds to its TATA box when it forms the docking site with factor IIA, but the new studies show that IID does not always require a TATA box.

For some time scientists have known that not all class II genes have a TATA recognition sequence in their promoter region. In the new work, Reinberg and his colleagues studied such a promoter, also from adenovirus, called the IVa2 promoter, to find out how RNA polymerase II binds to IVa2 in the absence of a TATA box. The researchers wanted to identify the transcription factors that this TATA-less promoter requires, and they initially expected that without a TATA box, the IIA-IID docking site would have nowhere to bind. But their initial instincts were wrong.

Decoding Genetic Puzzles

Transcription of the genetic code in DNA into molecules of mRNA is a complex puzzle and Danny Reinberg, of the University of Medicine and Dentistry of New Jersey, is looking for the pieces.

"Many proteins can bind to DNA," says Reinberg. "There are whole families of proteins that recognize the same sequence. Some may enhance, some may repress, some may do nothing. You cannot tell what they are doing just because they bind to DNA. If you are looking for transcription factors you have to test them in a system that tests for transcription, not DNA binding."

Though necessary, he says that purifying the components of a functional assay that accomplishes this is tedious. "It's conventional biochemistry, chromatography in the old style, collecting fractions and using a functional assay to test all the fractions," he says.

Reinberg, originally from Chile, learned this biochemical approach from Jerry Hurwitz while at Albert Einstein College of Medicine in New York. Later, he worked with Robert G. Roeder at Rockefeller University. "There is where I learned most of the tricks for tackling transcription," he says. "Without that background, none of the current work would have been possible." He then taught at the State University of New York at Stony Brook before settling in at the University of Medicine and Dentistry of New Jersey, formerly part of Rutgers Medical School.

Reinberg has been in New Jersey three and a half years and it took three of those years to save enough extracts to purify the human RNA polymerase II. He says, "I have to admit, we were lucky." Once they had enough starting material the purification went smoothly.

Reinberg hopes that most of the general transcription factors and RNA polymerase II will be purified, cloned, and available soon. He wants to be able to look at the roles of other specific factors (upstream and downstream) in the reconstituted functional assay.

He would also like to compare transcription in various species. "We know that much of the transcription process has been evolutionarily conserved. But even though sequences are somewhat conserved, proteins from yeast do not necessarily substitute for proteins from mammals. Protein-protein interactions appear to be much more specific than DNA-protein interactions."

—G.L.G.

Reinberg and his co-workers set up a functional assay system that contained a DNA template (with the TATA-less IVa2 promoter and a sequence to copy), polymerase II, and each of the general transcription factors except IID, in addition to the ribonucleoside triphosphates and energy sources. The researchers observed no transcription until they added a small amount of nuclear extracts from HeLa cells.

This result told them that a factor required for transcription from a TATA-less promoter was in the nuclear extract. To identify the mystery factor they fractionated nuclear extracts and tested the fractions in the same assay system. To their surprise, they found they were using the same procedures that are required to purify factor IID, the factor they thought would not work with a TATA-less promoter.

Because they have not yet completely purified factor IID, they cannot be sure if the mystery factor really is IID or if it just co-purifies with IID. So the researchers did some additional controls to convince themselves that the factors really is IID. Now they are wondering how IID binds in the absence of the TATA box.

"This implies that IID may be a degenerate protein that recognizes more than one sequence," says Reinberg. "Or it could mean that IID interacts with many other proteins to form heterodimers (molecules made of two different polypeptides), and the heterodimer is regulating the binding specificity."

Conaway offers another explanation. He says that Reinberg's results bring up the question of what constitutes the structure of the promoter. Conaway suggests that it may not be just the TATA box that is recognized by IID, even in promoters that have the recognition element. He cites a recent paper from David Baltimore of the Whitehead Institute in Boston and his colleagues, which suggests that other sequences are recognized by the transcription apparatus.

"The history of what constitutes a promoter is long and not complete yet," says Conaway. "There is substantial evidence that the TATA box and sequences around the cap site constitute the promoter, that you need two sets of sequences for the promoter to function. Putting a mutation in the TATA box decreases the strength of that promoter but you still have a promoter [because of the sequences around the cap site]. And if you put a mutation around the cap site you still have a promoter [because of the TATA box].

"Also, what constitutes the TATA box is not well defined. The TATA box got a name for itself because it's an AT-rich sequence that can be easily picked out. But the [nucleotides that define the] boundaries are not well established. Ultimately, to understand transcription we will have to understand the structure of the promoters."

The second finding of the Reinberg group concerns a specific DNA-binding protein that comes from HeLa cells, not adenovirus. It recognizes a DNA sequence upstream from the adenovirus major late promoter region and stimulates mRNA synthesis by affecting a general transcription factor. This cellular factor is called major late transcription factor or MLTF. In vivo, the adenovirus genome has a sequence upstream from its major late promoter (the efficient promoter with a TATA box that Reinberg has used to isolate the general transcription factors) that binds MLTF.

In the new work, Reinberg and coauthors extended an observation originally made by Roeder and Michelle Sawadogo, also of Rockefeller. They showed that purified MLTF appears to stabilize the committed complex on the adenovirus major late promoter. When polymerase II is released to continue elongating the new message, the committed complex is available to bind a new polymerase II molecule and start another round of transcription. This increases RNA synthesis 10- to 12-fold. Reinberg has not proven that MLTF binds directly to the general factors IIA or IID, but he cites the Roeder and Sawadogo results indicating that IID and MLTF cooperate to bind to DNA more stably.

Other researchers expect that other upstream DNA-binding proteins such as MLTF interact with general transcription factors to affect the amount of mRNA synthesized. Recent reports from Roeder's laboratory support the idea that such interactions may be a common occurrence. Yamamoto says these kinds of interactions give the appearance of "contextual effects" at the promoter region. Promoter efficiency, he says, depends on the sequences in and around the promoter region.

"The emerging picture of transcription," says Chamberlin, "is as a multi-step process, or pathway, in which each step depends on one or more of the general transcription factors and the efficiency of the promoter." Chamberlin thinks that each step will probably be susceptible to influences by one or more factors that bind somewhere else on the DNA and can loop over to affect a general transcription factor, or even RNA polymerase itself. If each step with the general factors is as efficient as it can be, the promoter cannot be enhanced—but it can be repressed. An inefficient promoter is one that needs enhancers at one or more steps.

"Because transcription involves protein-protein interactions, there is lots of flexibility for building lots of regulative steps, resulting in an enormous number of possible 'regulatory circuits' that might control transcription of any given gene," says Chamberlin. Ultimately, scientists will want to understand as many regulatory circuits as they can. "What has held up progress," he says, "is the lack of the general factors. With milligram quantities of them, we could work out each of these steps pretty quickly. Isolating the general factors will be critical to understanding the role of other factors in the control of transcription."

—GAYL LOHSE GALLAGHER

Additional Reading

S.T. Smale and D. Baltimore, "The 'initiator' as a transcriptional control element," *Cell* 57, 103 (1989).

F. Mermelstein, O. Flores, and D. Reinberg, "Initiation of transcription by RNA polymerase II," *Biochim. Biophys. Acta* (in press).

M. Horikoshi, T. Hai, Y.S. Lin, M.R. Green, and R.G. Roeder, "Transcription factor ATF interacts with TATA factor to facilitate establishment of preinitiation complex," *Cell* 54, 1033 (1988).

M. Sawadogo and R.G. Roeder, "Interaction of a gene specific transcription factor with adenovirus major late promoter upstream of the TATA box region," *Cell* 43, 165 (1985).

S.G. Saltzman and R. Weinmann, "Promoter specificity and modulation of RNA polymerase II transcription," *FASEB J.* 3, 1723 (1989).

P.J. Mitchell and R. Tjian, "Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins," *Science* 245, 371 (1989).