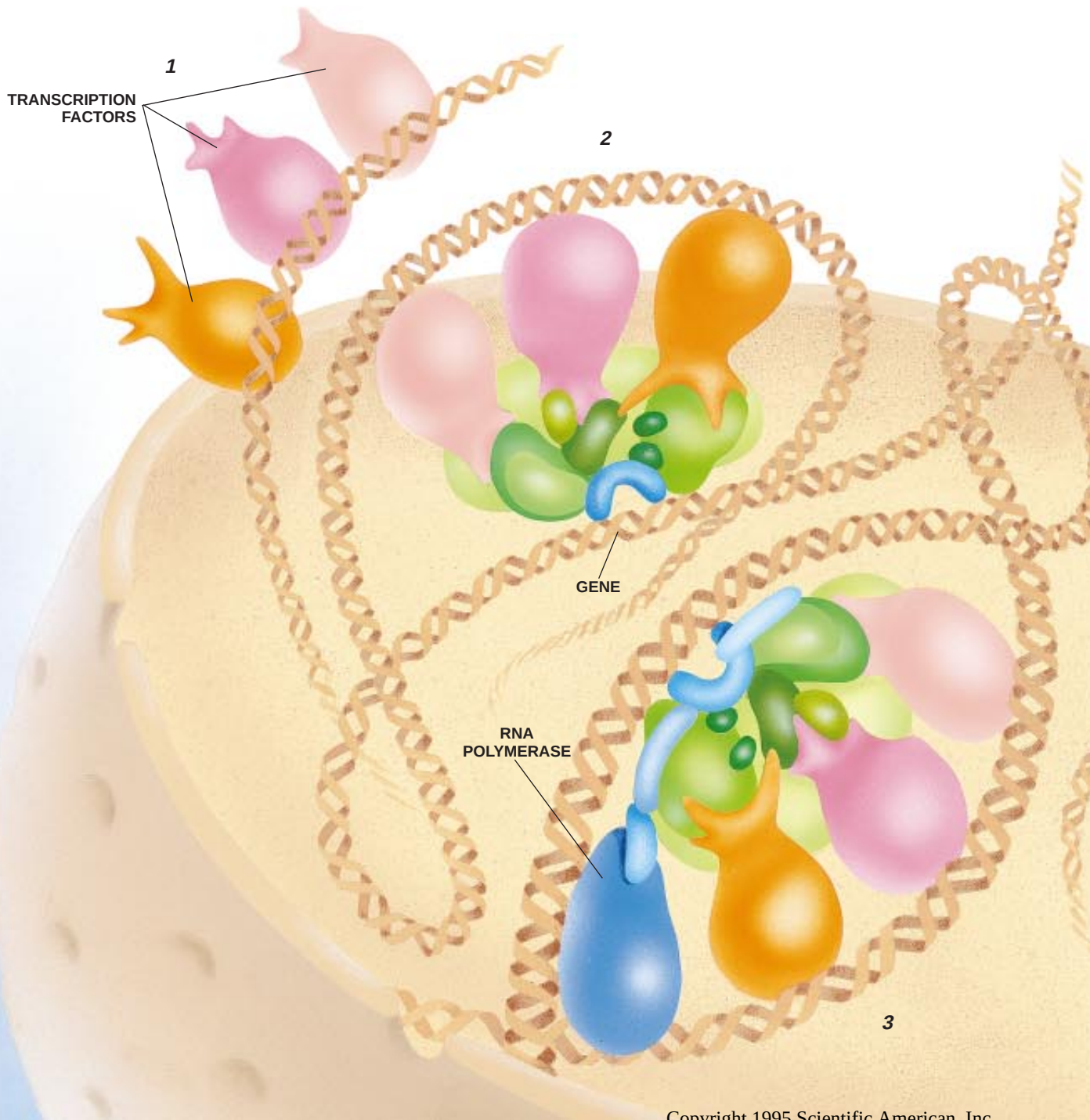


Molecular Machines That Control Genes

The activities of our genes are tightly regulated by elaborate complexes of proteins that assemble on DNA. Perturbations in the normal operation of these assemblies can lead to disease

by Robert Tjian



Asthma, cancer, heart disease, immune disorders and viral infections are seemingly disparate conditions. Yet they turn out to share a surprising feature. All arise to a great extent from overproduction or underproduction of one or more proteins, the molecules that carry out most reactions in the body. This realization has recently lent new urgency to research aimed at understanding, and ultimately manipulating, the fascinating biochemical machinery that regulates an essential step in protein synthesis: the transcription of genes. For a protein to be generated, the gene that specifies its composition must be transcribed, or copied, from DNA into strands of messenger RNA, which later serve as the templates from

which the protein is manufactured.

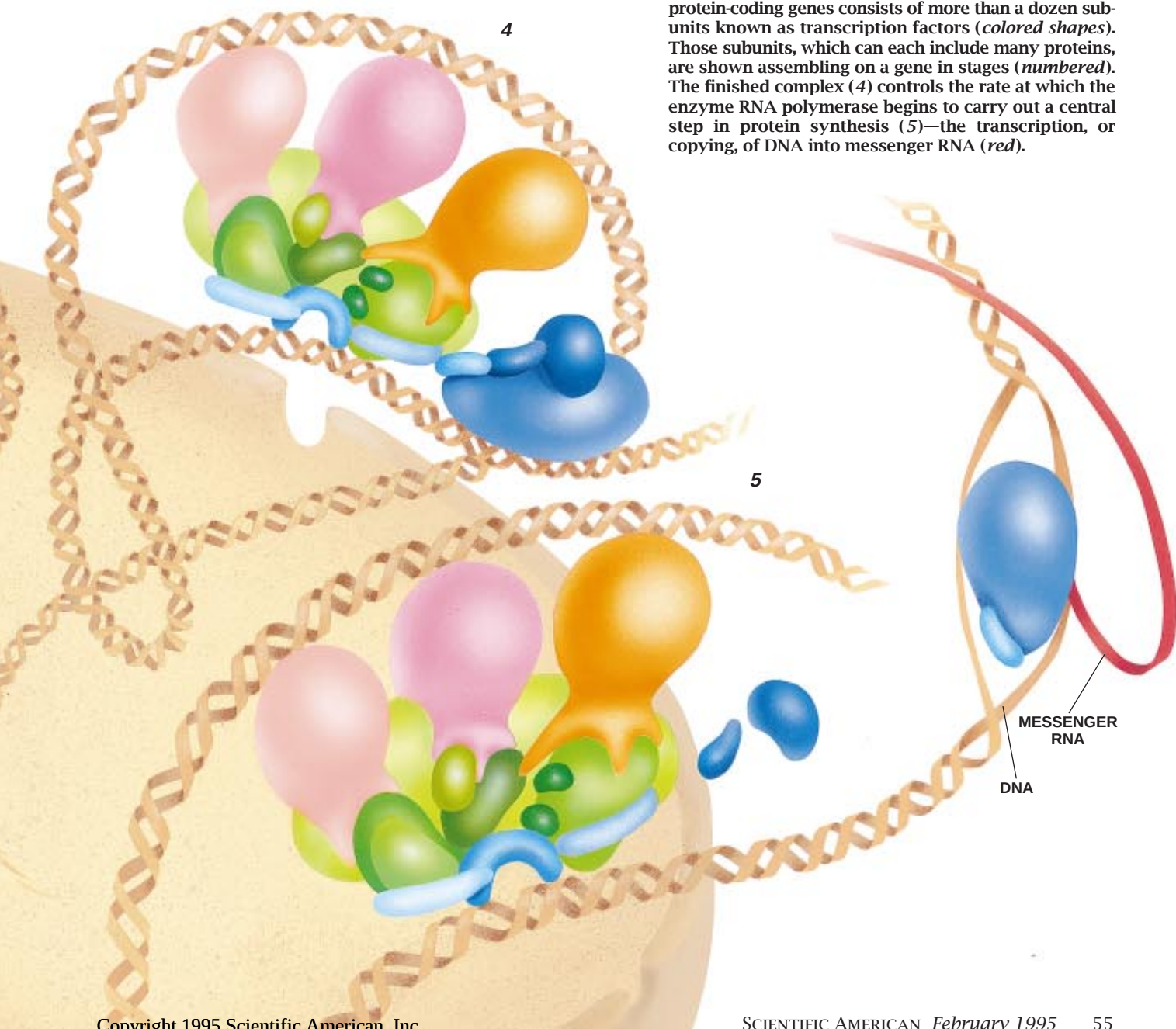
Even before therapy became a goal, transcription had long captivated scientists for another reason: knowledge of how this process is regulated promises to clarify some central mysteries of life. Each cell in the body contains the same genome, the complement of some 150,000 genes that form the blueprint for a human being. How is it that the original cell of an organism—

the fertilized egg—gives rise to a myriad of cell types, each using somewhat different subsets of those genes to produce different mixtures of proteins? And how do the cells of a fully formed body maintain themselves, increasing and decreasing the amounts of proteins they manufacture in response to their own needs and those of the larger organism?

To answer these questions and design

ROBERT TJIAN, who was born in Hong Kong, is an investigator with the Howard Hughes Medical Institute and professor of molecular and cell biology at the University of California, Berkeley. He earned his Ph.D. in biochemistry and molecular biology from Harvard University in 1976 and conducted research at the Cold Spring Harbor Laboratory in New York State before joining the faculty of Berkeley in 1979.

MOLECULAR MACHINERY that regulates the activity of protein-coding genes consists of more than a dozen subunits known as transcription factors (*colored shapes*). Those subunits, which can each include many proteins, are shown assembling on a gene in stages (*numbered*). The finished complex (4) controls the rate at which the enzyme RNA polymerase begins to carry out a central step in protein synthesis (5)—the transcription, or copying, of DNA into messenger RNA (*red*).



drugs able to modulate transcription, investigators need to know something about the makeup of the apparatus that controls reading of the genetic code in human cells. After some 25 years of exploration, the overall structure of that apparatus is becoming clear. Work in my laboratory at the University of California at Berkeley and at other institutions has revealed that one part of the apparatus—the engine driving transcription of most, if not all, human genes—consists of some 50 distinct proteins. These proteins must assemble into a tight complex on DNA before a special enzyme, RNA polymerase, can begin to copy DNA into messenger RNA. The putative constituents have now been combined in the test tube to yield a fully operational transcription engine. Still other proteins essentially plug into receptive sockets on the engine and, in so doing, “program” it, telling it which genes should be transcribed and how quickly. Critical details of these interactions are emerging as well.

Clues from Bacteria

When my colleagues and I at Berkeley began focusing on human genes in the late 1970s, little was known about the transcription machinery in our cells. But studies begun early in that decade had provided a fairly clear picture of transcription in prokaryotes—bacteria and other primitive single-celled organisms that lack a defined nucleus. That work eventually lent insight into human and other eukaryotic (nucleated) cells and helped to define features of transcription that hold true for virtually all organisms.

The bacterial research showed that genes are essentially divided into two functionally distinct regions. The coding region specifies the sequence of amino acids that must be linked together to make a particular protein. This sequence is spelled out by the nucleotides (the building blocks of DNA) in one strand of the DNA double helix; the nucleotides are distinguished from one another by the nitrogen-rich base they carry—adenine (A), thymine (T), cytosine (C) or guanine (G). The other region of a gene has regulatory duties. It controls the rate at which RNA polymerase transcribes the coding region of a gene into messenger RNA.

In bacteria, as in most prokaryotes, the regulatory region, called the promoter, resides within a stretch of nucleotides located a short distance—often as few as 10 nucleotides—in front of (upstream from) the start of the coding region. For transcription to proceed accurately and efficiently, RNA polymerase

must attach to the promoter. Once it is so positioned, it slides over to the start of the coding region and rides along the DNA, like a train on a track, constructing an RNA replica of the coding sequence. Except in very long genes, the number of RNA molecules made at any moment depends mainly on the rate at which molecules of RNA polymerase attach to the promoter and initiate transcription.

Interestingly, RNA polymerase is a rather promiscuous molecule, unable to distinguish between the promoter and other DNA sequences. To direct the enzyme to promoters of specific genes, bacteria produce a variety of proteins, known as sigma factors, that bind to RNA polymerase. The resulting complexes are able to recognize and attach to selected nucleotide sequences in promoters. In this way, sigma factors program RNA polymerase to bypass all nonpromoter sequences and to linger only at designated promoters.

Considering the importance of sigma factors to the differential activation of genes in bacteria, my colleagues and I began our inquiry into the human transcription apparatus by searching for sigma-like molecules in human cells. But we had underestimated the complexity of the machinery that had evolved to retrieve genetic information from our elaborate genome. It soon became apparent that human sigma factors might not exist or might not take the same form as they do in bacteria.

Surprising Complexity

If there were no simple sigma factors in eukaryotes, how did such cells ensure that RNA polymerase transcribed the right genes at the right time and at the right rate? We began to see glimmers of an answer once the unusual design of eukaryotic genes was delineated.

By 1983 investigators had established that three kinds of genetic elements, consisting of discrete sequences of nucleotides, control the ability of RNA polymerase to initiate transcription in all eukaryotes—from the single-celled yeast to complex multicellular organisms. One of these elements, generally located close to the coding region, had been found to function much like a bacterial promoter. Called a core promoter, it is the site from which the polymerase begins its journey along the coding region. Many genes in a cell have similar core promoters.

Walter Schaffner of the University of Zurich and Steven Lanier McKnight of the Carnegie Institution of Washington, among others, had additionally identified an unusual set of regulatory ele-

Anatomy of the Transcription Apparatus

The molecular apparatus controlling transcription in human cells consists of four kinds of components. Basal factors (*blue shapes at bottom*), generally named by single letters, are essential for transcription but cannot by themselves increase or decrease its rate. That task falls to regulatory molecules known as activators (*red*) and repressors (*gray*); these can vary from gene to gene. Activators, and possibly repressors, communicate with the basal factors through coactivators (*green*)—proteins that are linked in a tight complex to the TATA binding protein (TBP), the first of the basal factors to land on a regulatory region of genes known as the core promoter. Coactivators are named according to their molecular weights (in kilodaltons).

ments called enhancers, which facilitate transcription. These sequences can be located thousands of nucleotides upstream or downstream from the core promoter—that is, incredibly far from it. And subsequent studies had uncovered the existence of silencers, which help to inhibit transcription and, again, can be located a long distance from the core promoter.

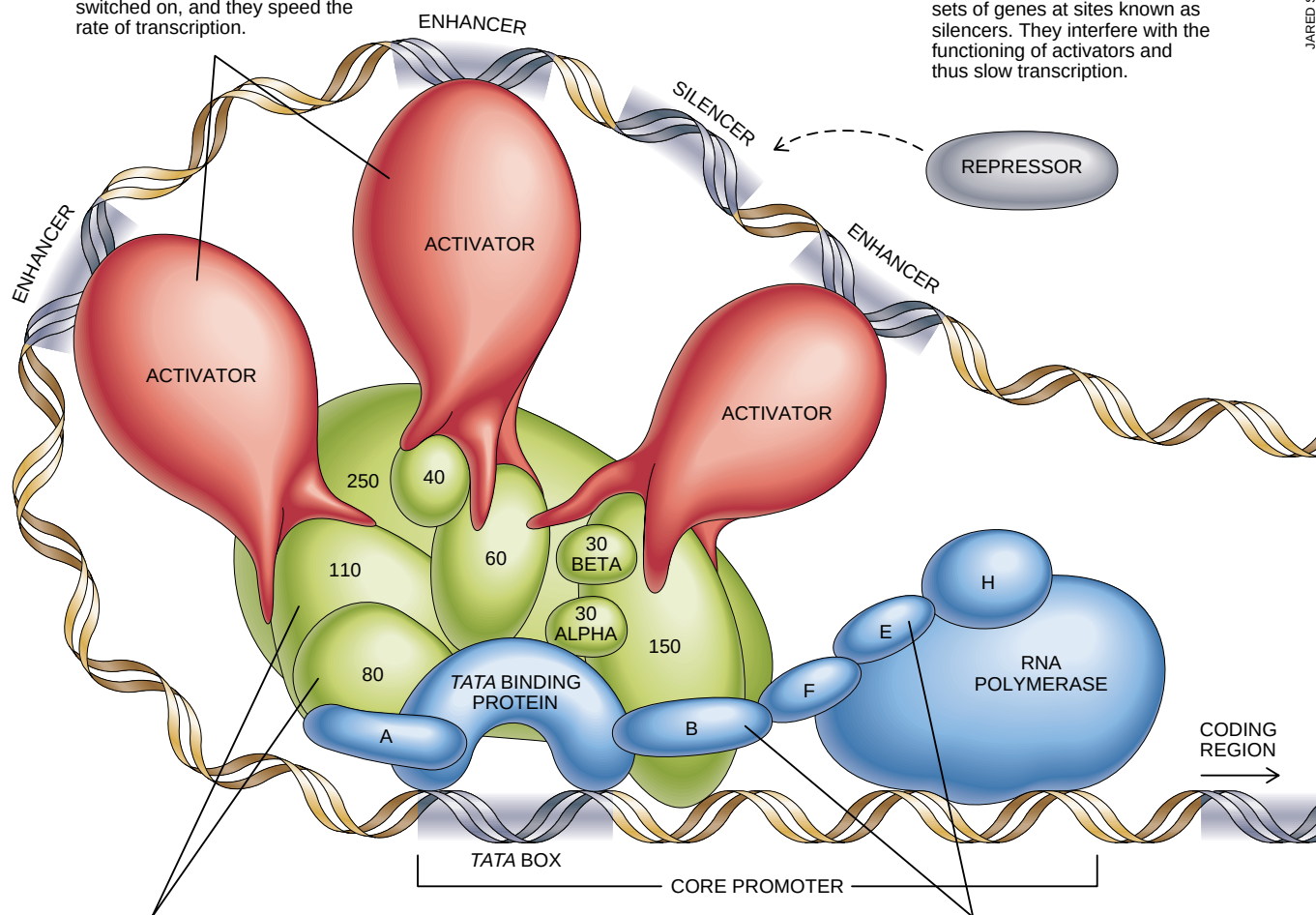
In a somewhat imperfect analogy, if the core promoter were the ignition switch of a car engine, enhancers would act as the accelerator, and silencers as the brakes. Eukaryotic genes can include several enhancers and silencers, and two genes may contain some identical enhancer or silencer elements, but no two genes are precisely alike in the combination of enhancers and silencers they

ACTIVATORS

These proteins bind to genes at sites known as enhancers. Activators help to determine which genes will be switched on, and they speed the rate of transcription.

REPRESSORS

These proteins bind to selected sets of genes at sites known as silencers. They interfere with the functioning of activators and thus slow transcription.

**COACTIVATORS**

These "adapter" molecules integrate signals from activators and perhaps repressors and relay the results to the basal factors.

BASAL FACTORS

In response to injunctions from activators, these factors position RNA polymerase at the start of the protein-coding region of a gene and send the enzyme on its way.

carry. This arrangement enables cells to control transcription of every gene individually.

Discovery of these elements led to two related—and, at the time, highly surprising—conclusions. It was evident that enhancers and silencers could not control the activity of RNA polymerase by themselves. Presumably they served as docking sites for a large family of proteins. The proteins that bound to enhancers and silencers—now called activators and repressors—then carried stimulatory or repressive messages directly or indirectly to RNA polymerase (that is, pressed on the accelerator or on the brakes). It also seemed likely that the rate at which a gene was transcribed would be dictated by the combined activity of all the proteins—or transcrip-

tion factors—bound to its various regulatory elements.

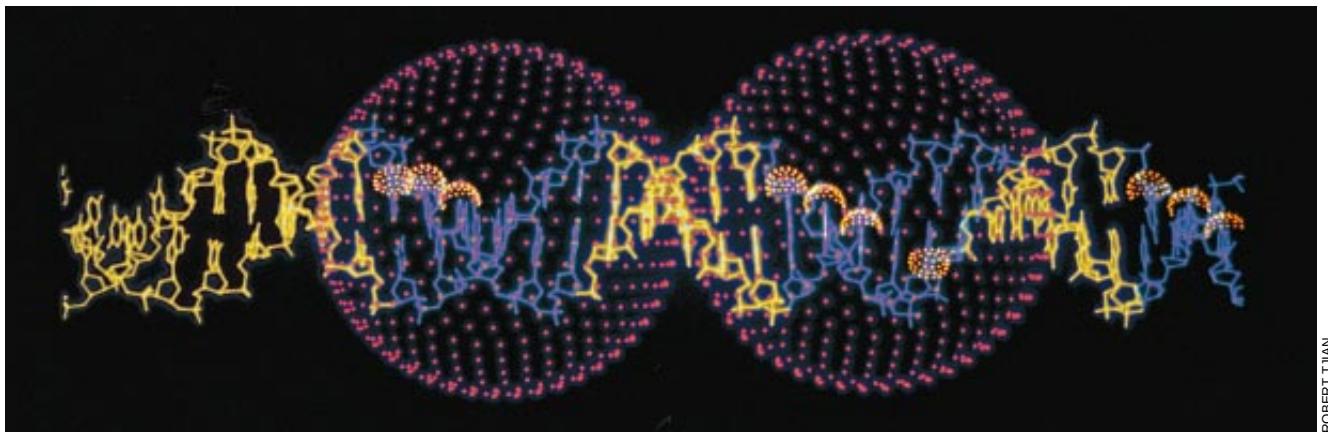
A Human Factor Is Discovered

Nevertheless, we were hard-pressed to explain how proteins that bound to DNA sequences far from the core promoter of a gene could influence transcription of that gene. As is true of other laboratories, we began attacking this puzzle by trying to isolate human transcription factors, none of which had yet been found (with the exception of RNA polymerase itself). We assumed that once we had pure copies of the factors we would be able to gain more insight into exactly how they function.

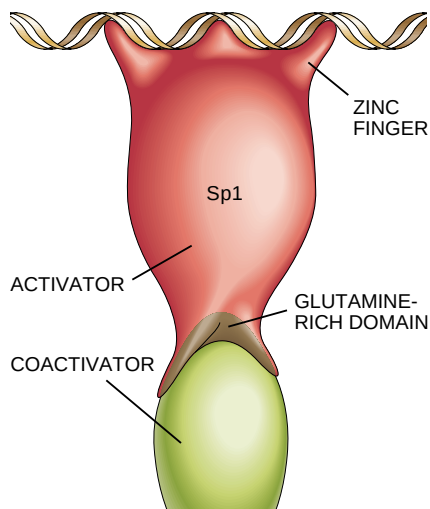
Because many proteins that bind to DNA play no role in reading genes, we

could not find transcription factors efficiently by screening nuclear proteins solely according to their ability to associate with DNA. My group therefore adopted a more discriminating strategy, looking for proteins that in a test-tube reaction both combined with DNA and stimulated transcription.

In 1982 William S. Dynan, a postdoctoral fellow in my laboratory, determined that some protein in a mixture of nuclear proteins fit all the requirements of a transcription factor. It bound to a regulatory element common to a select set of genes—an enhancer sequence known as the *GC box* (because of its abundance of *G* and *C* nucleotides). More important, when added to a preparation of nuclear proteins that included RNA polymerase, the substance mar-



ROBERT TJIAN



TWO MOLECULES of the activator protein Sp1 (represented above as large dotted spheres) have each attached to enhancer sequences called GC boxes (purple regions) by means of protrusions known as zinc fingers; the points of contact with DNA are highlighted by orange hemispheres. After Sp1 grabs on to DNA, it uses a region rich in the amino acid glutamine (brown in diagram) to convey transcription-stimulating signals to a specific coactivator (green shape).

JARED SCHNEIDMAN

edly increased the transcription only of genes carrying the GC box. Thus, we had identified the first human transcription factor able to recognize a specific regulatory sequence. We called it specificity protein 1 (Sp1).

We immediately set out to purify the molecule. One daunting aspect of this work was the fact that transcription factors tend to appear only in minuscule quantities in cells. Typically, less than a thousandth of a percent of the total protein content of a human cell consists of any particular factor. In 1985 James T. Kadonaga in my laboratory found a way to overcome this substantial technical barrier—and in the process introduced a powerful new tool that has since been used to purify countless transcription factors and other scarce DNA binding proteins.

Because Sp1 selectively recognized the GC box, Kadonaga synthesized DNA molecules composed entirely of that box and chemically anchored them to solid beads. Then he passed a complex mixture of human nuclear proteins over the DNA, predicting that only Sp1 would stick to it. True to plan, when he separated the bound proteins from the synthetic DNA, he had pure Sp1.

From studies carried out by Mark Ptashne and his colleagues at Harvard University, we knew that bacterial transcription regulators are modular proteins, in which separate regions perform distinct tasks. Once we learned the sequence of amino acids in Sp1, we therefore looked for evidence of distinct modules and noted at least two interesting ones.

One end of the molecule contained a region that obviously folded up into three “zinc fingers.” Zinc-finger structures, in which parts of a protein fold around a zinc atom, are now known to act as the “hooks” that attach many activator proteins to DNA. But at the time Sp1 was only the second protein found to use them. Aaron Klug and his colleagues at the Medical Research Council in England had discovered zinc fingers, in a frog transcription factor, just a short time before [see “Zinc Fingers,” by Daniela Rhodes and Aaron Klug; SCIENTIFIC AMERICAN, February 1993].

The other end of Sp1 contained a domain consisting of two discrete segments filled with a preponderance of the amino acid glutamine. We strongly suspected that this region played an important role during transcription because of a striking finding. In test-tube experiments, mutant Sp1 molecules lacking the domain could bind to DNA perfectly well, but they failed to stimulate gene transcription. This outcome indicated that Sp1 did not affect transcription solely by combining with DNA; it worked by using its glutamine-rich segment—now known as an activation domain—to interact with some other part of the transcription machinery. The question was, which part?

In 1988 when we began searching for

the target of Sp1, we had some idea of where it lay. Our guess was based on an emerging understanding of the so-called basal transcription complex, one part of which seemed to be a likely target.

Closing in on a Target

In the mid-1980s Robert G. Roeder and his colleagues at the Rockefeller University had shown that RNA polymerase cannot transcribe eukaryotic genes unless several other transcription factors—now called basal factors—also collect on the core promoter. And over the course of the 1980s, Roeder's laboratory and others had identified at least six of those essential factors, called A, B, D, E, F and H.

In a test tube, this assembly of factors enabled RNA polymerase to transcribe a bound gene at a basal—low and invariant—rate, but it could not by itself modulate that rate. It was as if someone had constructed and switched on the engine of a car but had lost use of the steering wheel, the accelerator and the brakes. For instance, when my group mixed the components of the complex (including RNA polymerase) with a gene containing a GC box, we obtained a low, unchanging level of transcription. We saw a marked increase in transcription only when we incorporated Sp1 into the mixture.

By the late 1980s it was apparent that human cells harbor at least two separate classes of transcription factors. Basal factors are required for initiation of transcription in all genes; other proteins—activators and repressors—dictate the rate at which the basal complex initiates transcription. Different genes are controlled by distinct combinations

of activators and repressors. We now suspect that in the body the basal complex arises spontaneously only rarely; most of the time, cells depend on activators to initiate its construction.

These various discoveries suggested that the glutamine-rich domain of Sp1 enhanced transcription by contacting a basal factor. More specifically, we suspected that Sp1 latched on to factor D, and facilitated its attachment to the promoter. We focused on this subunit because Phillip A. Sharp and Stephen Buratowski of the Massachusetts Institute of Technology had shown that it can land on the core promoter before all other basal factors and can facilitate assembly of the complete basal engine. In fact, factor D is the only basal component able to recognize DNA. It binds selectively to a sequence called the *TATA* box, found in the core promoters of many eukaryotic genes.

To pursue our hypothesis, we needed to know more about the composition of factor D, which we assumed was a solitary protein. Other investigators also wanted to know its makeup, and so the race was on to attain pure copies. Isolation from human cells proved more challenging than anyone anticipated. Consequently, many groups eventually tried their luck with yeast cells. Finally, in 1989, several laboratories independently succeeded in isolating a yeast protein that displayed the expected properties of factor D. The protein, named TBP (for *TATA* binding protein), recognized and bound selectively to the *TATA* box and led to a low level of transcription when it was joined at the core promoter by RNA polymerase and other constituents of the basal machinery.

Believing that the TBP protein was factor D itself, we undertook to test this idea in additional studies. Once we did that, we intended to determine exactly which regions of TBP were contacted by Sp1 and other regulators. Little did we know that we were about to be completely thwarted—and to make a critical discovery.

Unexpected Trouble

When B. Franklin Pugh in our laboratory replaced the impure preparations of factor D previously used in our test-tube reactions with purified molecules of TBP, he had no trouble replicating the earlier finding that such substitution in no way disrupted basal transcription. To our surprise and consternation, though, he found that Sp1 was no longer able to influence the basal machinery. We had to conclude that factor D and TBP were not, in fact, equivalent and that factor D actually

consisted of TBP plus other subunits. (It is now known that many transcription factors consist of more than one protein.) Apparently, those subunits were not needed for operation of the basal machinery, but they were essential to regulation of that machinery by activators.

In other words, these additional components were not themselves activators, for they did not bind to specific sequences in DNA. Nor were they basal factors, because low, unregulated levels of transcription could be achieved without them. They seemed to constitute a third class of transcription factor, which we called coactivators. We further proposed that coactivators, not TBP, were the targets for the protein binding domains of activators. We envisioned that activators would bind to selected coactivators to speed up the rate at which the basal complex set molecules of RNA polymerase in motion.

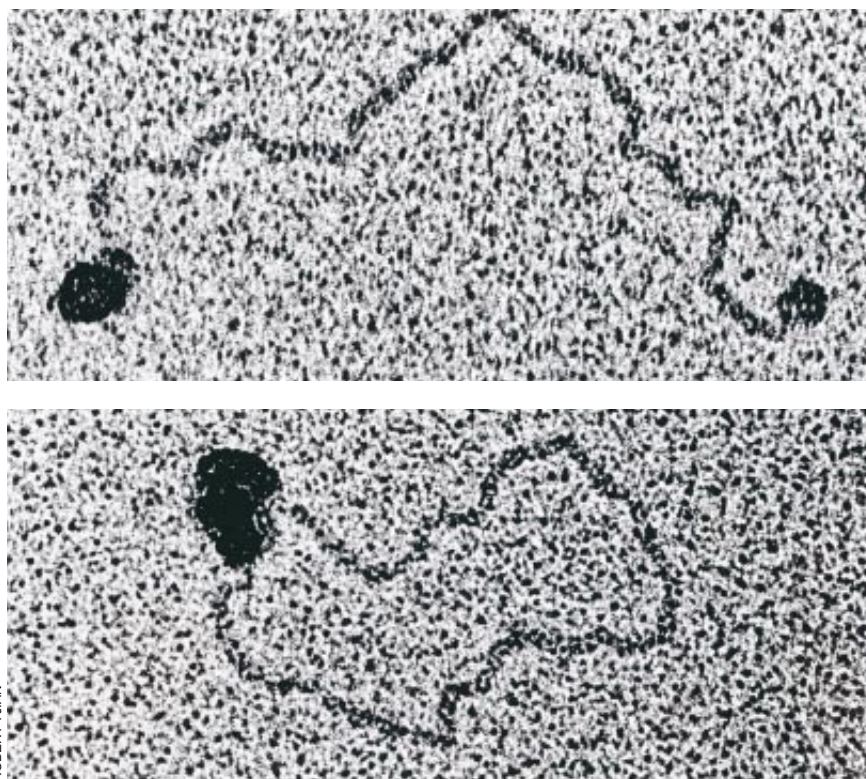
We were attracted to this scenario because we had difficulty imagining how a single protein, TBP, would have enough binding sites to accommodate all the activators made by human cells. But if the coactivators that were tightly linked to TBP bore multiple binding domains, the coactivators could collectively provide the docking sites needed to relay

messages from hundreds or thousands of activators to the transcription engine.

It was Pugh who originally proposed that coactivators might function as such adapter molecules. His data soon convinced me he was probably correct, but not everyone in our laboratory agreed. Indeed, our weekly meetings in early 1990 were often punctuated by heated discussions. Not surprisingly, when the coactivator concept was presented to other workers in the field, they, too, expressed considerable skepticism. This reaction to an unexpected and complicating result was probably justified at that stage, because our data were only suggestive, not conclusive. We had not yet isolated a single coactivator.

Coactivators: The Missing Links

To satisfy ourselves and the scientific community that we were correct, we had to devise an experimental procedure that would unambiguously establish whether coactivators existed and operated as the relays we envisioned. For approximately two years after Pugh formulated the coactivator hypothesis, we struggled to purify an intact and functional complex containing TBP and all the other associated constituents of factor D. I must admit to some dark mo-



ACTIVATORS CAN COMMUNICATE with one another, not only with DNA and coactivators, as is demonstrated in an experiment involving Sp1. Copies of the protein (*dark dots*) each recognized and bound to *GC* boxes at the tips of a string of DNA. After binding, they joined together—causing DNA to loop out in the process.

ments when it seemed the rather unpopular coactivator hypothesis might be based on some error in our studies.

The breakthrough finally came in 1991, when Brian D. Dynlacht, Timothy Hoey, Naoko Tanese and Robert Weinzierl—graduate students and postdoctoral fellows in our laboratory—found an ingenious way to isolate pure copies of factor D. Subsequent biochemical analyses revealed that, aside from TBP, the complete unit included eight previously unknown proteins. Because we did not yet have proof that these proteins could function as coactivators, we referred to them more generically as TBP-associated factors, or TAFs.

We became convinced that TAFs do indeed convey molecular signals from activators to the basal transcription ap-

paratus after we separated the bound proteins from TBP and completed several more experiments. For instance, we were able to show that mixing of the activator Sp1 with basal factors and RNA polymerase enhanced production of messenger RNA from a gene containing a GC box only when TAFs were added as well. Later, Jin-Long Chen, a graduate student, combined purified TBP and the eight isolated TAFs in a test tube along with a human gene and the rest of the basal transcription machinery. The various proteins assembled on the gene and proved able to respond to several different types of activator proteins. These activators, we later showed, produced their effects by coupling directly with selected TAFs. Together the coactivators in factor D do indeed consti-

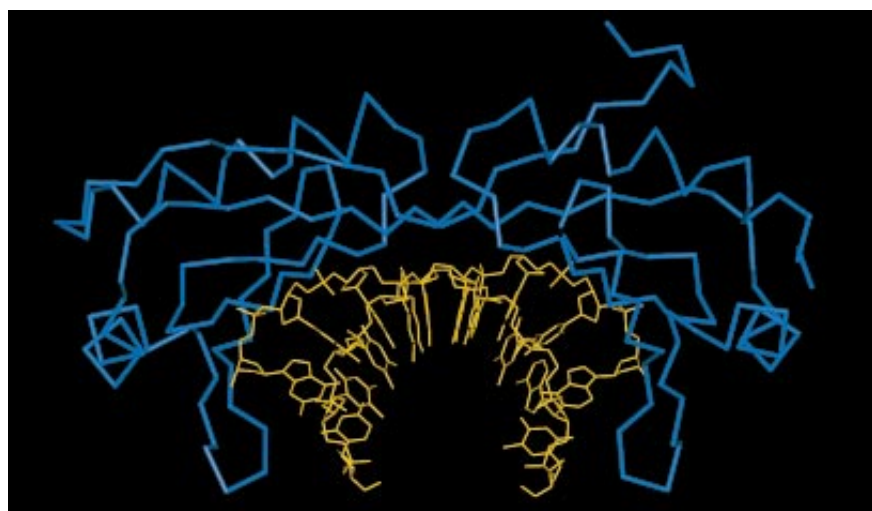
tute a kind of central processing unit that integrates the regulatory signals issued by DNA-bound activators.

A Universal Theme

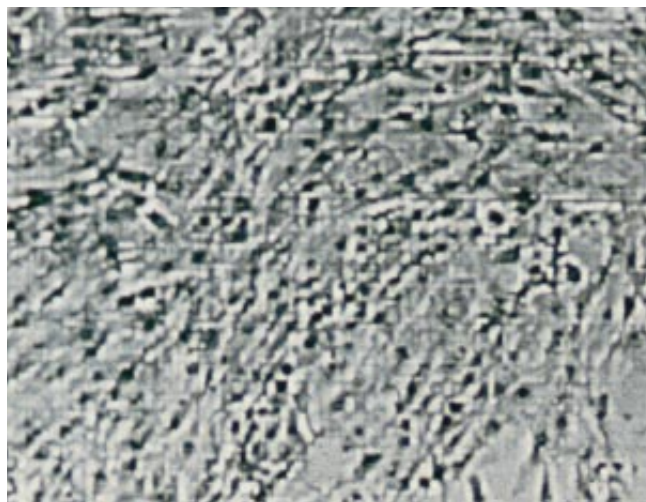
The complexes formed by activators, coactivators and the basal machinery appear to be human equivalents of sigma factors; they, too, draw RNA polymerase to specific genes at specific rates. In a way, the complexes can be viewed as sigma factors that have been elaborated into many subunits. Gratifyingly, recent evidence from our group and others suggests we have uncovered a universal mode of gene regulation in eukaryotes. Those studies confirm that coactivators also exist in yeast and that factor D consists of multiple subunits in fungi as well as in humans.

As satisfying as these results are, they do not fully explain how binding of activators to enhancers and to coactivators influences the rate at which RNA polymerase transcribes genes in living cells. It may be that linkage of activators to enhancers causes DNA to bend in a way that brings the enhancers closer to one another and to the core promoter. This arrangement may help activators (alone or in concert with one another) to dock with coactivators and position factor D on the promoter. This step, in turn, would facilitate assembly of the complete basal complex. Formation of this complex may distort the underlying DNA in a way that enables RNA polymerase to begin its journey along the coding region.

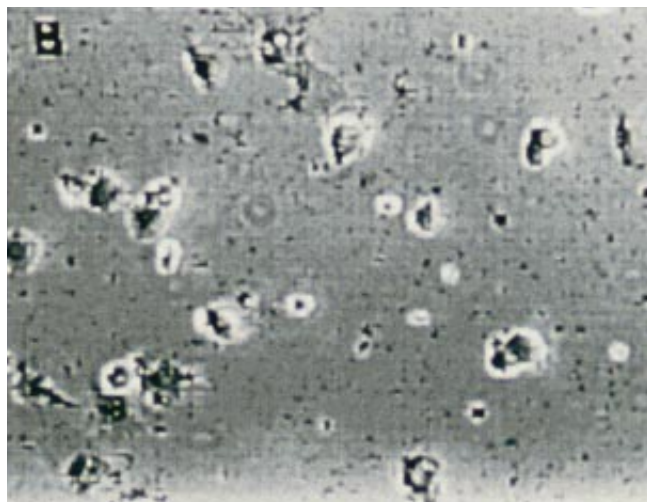
Researchers know less about the functioning of repressors. Nevertheless, many of us think repressors may also bind to coactivators at times. This bind-



TATA BINDING PROTEIN (blue) is a remarkably symmetrical, saddle-shaped molecule. Its underside rides on DNA (yellow) and seems to bend it. This bending may somehow facilitate assembly of the complex that initiates transcription. Coactivators, which are not depicted, are thought to bind tightly to the upper surface.



CELLS GROWING IN CULTURE (left) stopped proliferating and began to die (right) when they were blocked from making a coactivator. They failed because they could no longer



produce proteins critical to survival. This finding is one of many indicating that coactivators are essential to the transcription of most, if not all, genes.

ing could inhibit transcription by preventing activators from attaching to their usual sites on coactivators. Other times repressors might bypass the basal machinery, blocking transcription by preventing activators from connecting with enhancers.

Although there are gaps in our knowledge, we can now begin to sketch out an explanation as to why different cells make different mixtures of proteins during embryonic development and in mature organisms. A gene will be transcribed at a measurable rate only if the various activators it needs are present and can successfully overcome the inhibitory effects of repressors. Cells vary in the proteins they make because they contain distinct batteries of activators and repressors. Of course, this scenario begs the question of how cells decide which transcription factors to produce in the first place, but progress is being made on that front as well.

Therapies of Tomorrow

How might investigators use our newly acquired knowledge of gene regulation to develop drugs for combating life-threatening diseases involving excessive or inadequate transcription of a gene? In theory, blocking selected activators from attaching to enhancers or coactivators should depress unwanted transcription, and stabilizing the transcription machinery on a gene should counteract undesirably weak transcription.

Blockade could be achieved by fitting a molecular "plug" into an activator, thereby preventing its interaction with a coactivator, or by enticing an activator to attach to a decoy that resembles a coactivator. Stabilization of a complex might be achieved by deploying molecules that would strengthen the interaction between activators and DNA or between activators and coactivators. Such approaches are remote today, but it is exciting to consider a sampling of the applications that might eventually be possible.

Take, for example, the human immunodeficiency virus (HIV), which causes AIDS. To reproduce itself in human cells, HIV needs the viral transcription factor *TAT* to enhance transcription of HIV genes. If *TAT* could be inhibited by some agent that recognized *TAT* but ignored human transcription factors, replication of the virus might be halted without affecting production of proteins needed by the patient.

Conversely, treatment of some disorders—for instance, hypercholesterolemia—might involve enhancing the transcription of selected genes. Hypercho-



BETTINA CIRONE Photo Researchers, Inc.

ONE OF THE WORLD'S TALLEST WOMEN reportedly reached her height of seven feet, seven and a quarter inches because her pituitary gland produced an excess of growth hormone. Investigators hope eventually to treat this and many other disorders by increasing or decreasing the transcription of selected genes.

lesterolemia increases a person's risk for heart disease. Cholesterol accumulates to destructive levels in the blood when low-density lipoprotein (LDL), otherwise known as the bad cholesterol, is not removed efficiently. In theory, the disease could be corrected by turning up transcription of the gene for the LDL receptor in liver cells. This receptor helps to clear LDL from the blood. This idea may soon be testable, because studies by Michael S. Brown and Joseph L. Goldstein of the University of Texas Health Science Center at Dallas are teasing apart the specific molecular constituents of the apparatus that regulates transcription of the receptor gene.

Until recently, no one put much effort into screening small molecules, natural products or other compounds for their ability to modulate transcription. Even so, a number of drugs already on the market have been found by chance to work by altering the activity of transcription factors. One of these, RU 486 (the French "abortion" pill), represses the function of particular steroid receptors, a class of activators that direct embryonic development. Similarly, the immunosuppressants cyclosporine and FK506 suppress transcription of a gene whose protein product is needed by certain cells of the immune system. These drugs act indirectly, however. They activate an enzyme that impedes the functioning of a transcription factor for the gene.

As time goes by, the precise combination of transcription factors that reg-

ulate individual genes is sure to be identified. And drug developers will probably use this information to devise sophisticated compounds for fighting cancer, heart disease, immune disorders, viral infections, Alzheimer's disease and perhaps even the aging process. How well these agents will succeed is anybody's guess, but it is likely that therapies of the future will benefit in one way or another from basic research into transcription—research that began not out of a wish to design drugs but rather out of a simple desire to get to the heart of the molecular machinery that controls the activity of our genes.

FURTHER READING

TRANSCRIPTIONAL SELECTIVITY OF VIRAL GENES IN MAMMALIAN CELLS. Steven McKnight and Robert Tjian in *Cell*, Vol. 46, No. 6, pages 795-805; September 12, 1986.

TRANSCRIPTIONAL REGULATION IN MAMMALIAN CELLS BY SEQUENCE-SPECIFIC DNA BINDING PROTEINS. Pamela J. Mitchell and Robert Tjian in *Science*, Vol. 245, pages 371-378; July 28, 1989.

EUKARYOTIC COACTIVATORS ASSOCIATED WITH THE TATA BOX BINDING PROTEIN. G. Gill and R. Tjian in *Current Opinion in Genetics and Development*, Vol. 2, No. 2, pages 236-242; April 1992.

TRANSCRIPTIONAL ACTIVATION: A COMPLEX PUZZLE WITH FEW EASY PIECES. R. Tjian and T. Maniatis in *Cell*, Vol. 77, No. 1, pages 5-8; April 8, 1994.