Two quite different types of research have been invaluable in elucidating principles in the field of molecular biology: biochemistry and genetics. When those two approaches intersect, as they have in recent work on eukaryotic transcription control, the resulting reverberation has the ring of truth. In a recent issue of these *Proceedings*, Li et al. (1) presented the latest example of the convergence of biochemical and genetic studies that have led to the recent discovery of a eukaryotic RNA polymerase holoenzyme. The holoenzyme is a multiprotein complex containing the multimeric RNA polymerase II protein that transcribes protein coding genes, most of the general transcription factors required for the polymerase to locate eukaryotic promoters and initiate transcription, and a newly discovered multimeric protein complex called mediator required for the proper temporal regulation of transcription in response to regulatory transcription factors. Together with the transcription initiation factors TFIIH and TFIIA, the holoenzyme forms an approximately ribosome-sized complex of more than 50 polypeptides that assembles on eukaryotic promoters. This huge preinitiation complex integrates signals from regulatory transcription factors bound at enhancer and silencer DNA sequences to determine how frequently RNA polymerase II initiates transcription of the neighboring gene.

Even though RNA polymerase II is a complex multimeric protein of 14 subunits (2), by itself it is incapable of starting transcription at the specific initiation sites used in intact cells. Specification of the transcription initiation site is a critical aspect of transcription control, since it determines which portions of genomic DNA are transcribed into mRNA. Consequently, the initial efforts of biochemists went toward identifying a set of protein factors called general transcription factors (named TFIIA, -B, -D, -E, and -H) that direct the polymerase to initiate transcription at sites in DNA that are used in the cell (3–5). However, although this complex set of general transcription factors can act together to specify the correct starts for RNA polymerase II, they are not sufficient to reproduce the regulation of transcription initiation observed in cells.

A combination of genetic and biochemical studies established that, in vivo, activator proteins bind to regulatory DNA sequences called enhancers and increase the frequency of transcription initiation at neighboring promoters, even when the closest promoter is 50 kb away in the linear DNA sequence (6). Activators stimulate transcription in cell-free in vitro reactions using complex extracts prepared from isolated cell nuclei. However, they fail to stimulate transcription in reactions with the purified general transcription factors. Working in the budding yeast system, Roger Kornberg and co-workers went back to the complex nuclear extract to search for an activity that would allow activators to stimulate transcription in the in vitro reaction with purified general transcription factors. They succeeded in identifying such an activity, calling it mediator (7). As they purified and analyzed mediator (8), it became apparent that they were characterizing the same multiprotein complex that was concurrently discovered by Richard Young and colleagues (24) through what was initially a genetic strategy for the analysis of transcription control. This was the initial intersection of the biochemical and genetic studies that led to the discovery of the holoenzyme.

Young and colleague’s genetic strategy began with an analysis of a unique structural aspect of RNA polymerase II, a specific C-terminal domain (CTD) associated with its largest subunit (9, 10). The CTD is composed of 26–52 tandem copies (depending on the organism) of nearly perfect repeats of a seven amino acid residue sequence, Tyr-Ser-Pro-Thr-Ser-Pro-Ser. When RNA polymerase II is assembled into the massive preinitiation complex at eukaryotic promoters, its CTD is unmodified. As the polymerase initiates transcription and then transcribes away from the initiation site on the template, the CTD is phosphorylated at multiple sites (11–14). A number of protein kinases have been identified that can phosphorylate the CTD, but the Cdk7 subunit of TFIIH appears to be responsible for most CTD phosphorylations in in vitro transcription reactions (15). In these in vitro reactions, Cdk7 phosphorylation of the CTD is required for transcription from some promoters (16) but not others (15).

Genetic experiments using the methods of site-directed mutagenesis revealed that the CTD is essential for the viability of budding yeast, *Drosophila*, and mammalian cells (17–20). To explore the function of the CTD through genetic methods, Young and his colleagues (17) identified the minimal number of heptadepptide repeats that would allow a single cell to grow into a colony on a plate of rich medium, 10, whereas the number found in wild-type *Saccharomyces cerevisiae* is about 26. They found that, unlike cells with the full-length CTD, cells with this mutant form of RNA polymerase II were unable to form colonies at low temperature, most likely because of generalized defects these mutants exhibit in regulating transcription of multiple genes. This cold sensitivity of cells with a partially deleted CTD allowed Young and his colleagues to take advantage of the genetic technique of identifying suppressing mutations. Such suppressors can result from mutations in additional proteins involved in the same process.

The identification of suppressing mutations that allow the CTD partial deletion mutant to grow at low temperature (21–23) led to the discovery of nine genes called SRB2, -4, . . . -11 (for suppressor of RNA polymerase B). When the Young group isolated these genes and used them to prepare specific antibody against the encoded SrB proteins, they discovered that all the SrB proteins were part of the same, very large, macromolecular complex that included the subunits of RNA polymerase II and most of the subunits of the general transcription factors (22, 24). [However, the critical TFIIH factor that binds to the TATA-box sequence of eukaryotic promoters was not present in the complex (24)]. The multiprotein complex they purified from yeast nuclear extracts was very similar to the complex identified by Kornberg and co-workers during their purification of mediator activity (8). Both groups agreed to call the complex the holoenzyme. While there are potentially important differences between the composition of the holoenzymes described by the two groups, the many similarities seem more significant than the differences. In experiments with cells having temperature-sensitive srb4 or srb6 mutations, the transcription of multiple genes shut down rapidly when the cells were shifted to the nonpermissive temperature (25). This important observation argues that the holoenzyme is required for the transcription of most, if not all, yeast genes.
The discovery of the holoenzyme was surprising because earlier work had shown that the purified general transcription factors and RNA polymerase II can be assembled into holoenzyme in a specific, ordered, stepwise fashion (3–5). Purification of the holoenzyme indicates that, in the cell, most of the general transcription factors are preassembled into holoenzyme complexes with RNA polymerase II and a mediator/Srb complex. Consequently, assembly of a preinitiation complex on promoter DNA may involve only two binding steps: binding of TFII D and TFIIA to promoter DNA, followed by binding of the holoenzyme to the TFII D–TFIIA-promoter DNA complex. Until recently, the holoenzyme had been observed only in budding yeast. But fundamental cellular processes such as transcription are generally highly conserved among all eukaryotes, and a recent paper reports the characterization of a holoenzyme from mammalian liver cells (26). Thus, it is a good bet that the yeast mediator/Srb complex proteins will have functional counterparts in higher cells. The current descriptions of holoenzyme preparations differ in the number of associated general transcription factors (8, 24, 26). It will be a challenging problem to determine precisely which general transcription factors associate with the holoenzyme in vivo before it binds to promoter DNA. The holoenzyme isolated from mammalian cells appears to include TFII D, raising the possibility that preinitiation complex assembly on promoter DNA may involve only a single protein–DNA-binding step (26, 27).

Now, Li et al. (1) report additional instances where biochemistry and genetics cross paths in the holoenzyme, providing further insight into its function. When the holoenzyme is treated with a monoclonal antibody that binds to the CTD, ~20 polypeptides are displaced in a subcomplex with mediator activity (8). The subcomplex includes the three subunits of yeast TFII F and at least four of the Srb proteins. Another of these polypeptides is encoded by the GAL11 gene (8), a gene discovered in 1980 because mutations in it reduce the ability of yeast cells to ferment galactose (28). Normal fermentation of galactose requires the function of the Gal4 activator protein, which stimulates transcription of genes encoding enzymes required to metabolize galactose. In gal11 mutants, Gal4 activation is impaired (29). Additional studies revealed that gal11 mutations have a general defect that reduces the ability of multiple activators to stimulate transcription (30, 31). With this precedent in mind, Li et al. (1) tested whether any of the unidentified mediator/Srb complex polypeptides might be encoded by other genes which have generalized effects on transcription.

Li et al. (1) focused on the genes SIN4 and RGR1 because they realized that the published characteristics of cells with mutations in these genes were similar to those of gal11 mutants (32–34). Indeed, specific antibody to Sin4 protein revealed that it purifies with the holoenzyme, and peptide microsequencing revealed that another previously unidentified polypeptide in the complex is encoded by the RGR1 gene (1). Thus, earlier studies indicating that the Gal11, Sin4, and Rgr1 proteins are required for maximal transcriptional activation by a number of different enhancers can now be explained by the association of these proteins with the holoenzyme, which is required for the transcription of many, if not all, genes (25). Sugar first identified through genetic studies (35), is also a component of the holoenzyme (1). Moreover, as Li et al. point out, the association of Sin4 and Rgr1 with the holoenzyme has significant additional implications for the function of the mediator/Srb complex in the holoenzyme. Sin4 and Rgr1 proteins are also required for repression by a number of different repressors (32, 33, 35–37). Similarly, recent genetic studies implicate the Srb10 and Srb11 proteins in mechanisms of repression by several different repressors (38–40). Consequently, the mediator/Srb subcomplex of the holoenzyme likely functions in negative as well as positive control of transcription initiation. The Srb10 and Srb11 proteins are particularly interesting because they have sequence homology to Cdk–cyclin protein kinases (23), which regulate target proteins by phosphorylating them at specific sites.

Our understanding of the RNA polymerase II preinitiation complex has grown enormously over the past few years. In addition to the 14 polypeptides of the polymerase (2) and the ~20 polypeptides of the general transcription factors TFIIA, -B, -E, -F, and -H (3–5), there are multiple TAFs required for activated transcription (41) that associate with the TATA-box-binding polypeptide to form the multisubunit TFII D general transcription factor. And now we can add to the preinitiation complex the ~20 additional polypeptides in the mediator/Srb complex. The picture that emerges is one of an intricate molecular machine positioned at a transcription start site consisting of some 60–70 different polypeptides. This appropriately named preinitiation ‘‘complex’’ functions to integrate multiple positive and negative regulatory signals from other multiprotein complexes bound to enhancer DNA sequences (e.g., see ref. 42) to control the timing of transcription initiation. Still other coactivator proteins not tightly associated with DNA-bound protein complexes also participate (e.g., see ref. 43). In addition to this complexity, chromatin proteins and their states of modification contribute to transcription control (44, 45). The number of potential regulatory interactions in such large complexes seems staggering. But we should not have expected less from the mechanisms that control transcription initiation, processes central to the development of organisms as complex as ourselves.


