Snapshots of RNA polymerase II transcription initiation
Stephen Buratowski

Several papers published within the last year utilize innovative techniques for characterizing intermediates in RNA polymerase II transcription. Structural studies of polymerase and its associated factors provide a detailed picture of the transcription machinery, and studies of transcription complex assembly both in vitro and in vivo provide insights into the mechanism of gene expression. A high resolution picture of the transcription complex is likely to be available within the foreseeable future. The challenge is to determine the roles of individual proteins within this surprisingly large molecular machine.

Introduction
A huge amount of work by many laboratories has advanced the RNA polymerase II (pol II) field to the stage that the proteins required to initiate transcription in vitro are completely defined. Remarkably, transcription requires several dozen proteins, most of which are components of multiple complexes and subunits. The biochemical functions of each complex and of individual subunits within them are only now being explored in detail. Within cells, however, regulation of transcription involves many more factors, and these affect the transcription complex both directly and indirectly via chromatin.

The following interactions are generally believed to occur within the transcription complex. Firstly, the transcription factor TFIID binds to the promoter via the direct interaction of its subunit TBP (TATA element). This interaction also requires TBP-associated factors (TAFs) and other basal promoter elements; TFIIA encourages this binding [1]. Secondly, TFIIIB bridges TBP and recruits pol II in collaboration with TFIIIE. Thirdly, TFIIIE and an ATP-dependent helicase within TFIIH participate in unwinding of promoter DNA. Fourthly, following initiation, escape into elongation is concomitant with phosphorylation of the polymerase carboxy-terminal domain (CTD).

Using purified factors, these interactions can be observed as discrete steps in a linear transcription pathway. However, it is necessary to define the assembly pathway(s) used in vivo, since this has important implications for how genes are regulated. There are several models for this assembly; at one extreme, each factor is proposed to enter the assembling transcription complex individually, whereas at the opposite end of the spectrum, a huge RNA polymerase II "holoenzyme" is proposed to contain all basal and regulatory transcription factors, as well as chromatin-remodeling complexes, DNA repair factors, and DNA replication factors. The majority of studies support a mechanism in between these extremes. Considerable evidence suggests that some, but not all, basal factors are associated with polymerase before binding to the promoter [2]. In addition, a mediator complex associates with the polymerase to transmit the effects of transcription regulatory proteins [2]. Reports within the last year further define the structure of initiation complex components and probe the assembly reaction both in vitro and in vivo and these will be discussed in this review.

Structure
Structural studies continue to provide insights into the architecture of the transcription complex. Earlier high resolution structures of TBP, TFIIA, and TFIIIB greatly illuminated biochemical and genetic studies. Recently, several new structures have been presented. Perhaps the most significant are several views of polymerase itself. Two structures are of yeast pol II: one derived from X-ray diffraction at 5 Å resolution [3**] and another from electron diffraction of an elongating polymerase [4**]. An even higher resolution structure of RNA polymerase from the bacterium Thermus aquaticus has also been solved [5**]. Because conservation of the catalytic polymerase subunits is evident from sequence alignments, the bacterial and eukaryotic polymerases were predicted to share structural features. Some of these can be seen when the new structures are compared. Coupled with extensive mutagenic and biochemical studies, a rather detailed model of the transcription elongation complex can be assembled.

RNA polymerase can be envisaged as a roughly "crab claw" or "clam" shaped molecule, with the bottom half consisting mainly of the largest subunit β and the upper half containing most of the second largest subunit β. Several channels lead into the internal active site. Double-stranded DNA downstream of the transcription bubble enters the polymerase through one channel, while upstream DNA exits through a second. The net result is that the DNA undergoes a bend of roughly 90°. This has several important implications. First, the topological constraints strongly predict that DNA threads through the transcribing polymerase, rather than the polymerase rotating around
the axis of the DNA. Second, to reconcile models in which DNA wraps around the transcription initiation complex [6], dramatic rearrangements upon transition from initiation to elongation phase must be invoked.

The regions of highest evolutionary conservation map to the active site. Approximately 8 to 9 base pairs of DNA–RNA hybrid can be accommodated there, and a magnesium ion is in position to aid polymerization [5••]. Interestingly, both the prokaryotic and eukaryotic polymerases contain a ‘tunnel’ on the opposite side from the DNA leading to the active site [3•••–5••]. This channel is predicted to allow NTPs to diffuse into place for incorporation into RNA. Near the upstream boundary of the transcription bubble, there is a protrusion from the active site floor of the prokaryotic enzyme [5••]. This ‘rudder’ is proposed to aid in splitting the RNA–DNA hybrid, thus allowing the two DNA strands to renature. RNA is predicted by crosslinking studies to exit through yet another channel opposite to the entering DNA. This tunnel is defined on one side by a protein ‘flap’ that may contact RNA as it emerges from the active site. A similar, flexible domain is seen in the eukaryotic enzyme [3•••,4••]. This proposed RNA–protein contact may aid in elongation and termination. The RNA chain would need to reach a critical length to contact the flap and it will be interesting to see whether this length corresponds to the point at which the transcription complex ceases abortive initiations and escapes into productive elongation. This domain is also predicted to mediate the functions of stem–loop RNA structures and protein factors in control of transcription elongation and termination [7].

Electron micrographs of individual molecules were averaged to determine the overall shape of the yeast and mouse mediator complexes, both alone and bound to pol II [8•]. Free mediator appears to be folded upon itself but can accommodate pre-initiation complexes; other examples include the yeast SAGA complex and the mammalian P/CAF complex [19]. The TFIID and TFTC molecules resemble each other, undoubtedly reflecting their shared and homologous components. There are multiple globular lobes arranged in a roughly horseshoe shape. These are connected by flexible regions. Antibodies against TBP bind to a central lobe, which is also where TFIIA and TFIIB dock. If one assumes that the concave face of the central lobe corresponds to the DNA-binding surface of TBP, one set of lobes may be poised to interact with DNA downstream of the DNase I footprinting). Another lobe may be oriented towards the upstream side of the transcription complex. It should be possible to map individual TAF subunits onto the structure in future studies.

**Transcription complex assembly**

Reconstituted *in vitro* transcription systems have been essential for establishing the contacts between various transcription factors. However, it can be argued that crude extracts more closely resemble nuclear conditions in terms of factor ratios and preassembled complexes. Using an immobilized template assay and nuclear extracts, two stable pre-initiation intermediates were isolated [20•]. In agreement with earlier footprinting and template commitment studies, TFIID and TFIIA associate with a promoter in the absence of other factors. The role of TFIIA in encouraging the interaction between TFIID and the promoter has now been extensively documented both *in vivo* and *in vitro* [21,22,23•,24]. The second stable intermediate contained TFIIB, polymerase, and the Srb4 (suppressor of RNA polymerase B) protein (a component of the yeast mediator complex). These results suggest that either polymerase enters the initiation complex while bound to the mediator and TFIIIB or that complex assembly is highly cooperative. The second stable intermediate also contained TFIIE and TFIIH, but it is not known whether these factors are obligatory components.

To study transcription complex assembly *in vivo*, the technique of chromatim immunoprecipitation or ChiP has been adapted from earlier studies on chromatin and replication [25]. Whole cells are treated with formaldehyde to rapidly crosslink proteins and DNA. Chromatin is isolated and sheared, followed by immunoprecipitation with proteins against the transcription factor of interest. After reversing the crosslink, pellets are assayed for the presence of promoter sequences by PCR.
The events that follow the induction of the well-studied HO promoter were monitored by the ChIP assay [26••]. HO expression requires the activators Swi5 and SBF (Swi4–Swi6 cell cycle box factor), the SWI/SNF chromatin remodeling complex, and the SAGA histone acetyltransferase complex. As monitored by ChIP, Swi5 binds the promoter first and is required to recruit SWI/SNF to regulatory regions of the promoter. SWI/SNF activity is in turn required for SAGA and SBF to bind. Interestingly, continued Swi5 binding is not required to maintain the association of SWI/SNF and SAGA complexes. Obviously, the next step is to monitor the binding of the transcription machinery itself.

The ChIP technique was used to monitor occupancy of TBP at several promoters [27•,28••]. Crosslinking of TBP correlates strongly with levels of transcription and is dependent upon the presence of activators, pol II, and the mediator component Srb4. These results suggest that either TBP crosslinking to DNA is indirect via one of these other transcription factors or, more likely, that TBP is not permitted to dwell on inactive promoters in vivo. The second hypothesis is supported by the observation that mutation of Mot1, a protein that can dissociate DNA-bound TBP [29,30], leads to increased TBP crosslinking at many promoters, as well as at nonphysiological sites within a coding region [28••]. The ChIP experiments indicate that transcription complex assembly in vivo is highly cooperative and support earlier models of gene regulation via control of TBP binding.

A basic understanding of interactions between basal factors has been established for a typical pol II promoter, but interesting exceptions to the rules may exist. The snRNA genes from higher eukaryotes use an alternative TBP-containing complex known as SNAPc (snRNA activating protein complex) to recognize snRNA-specific basal promoter sequences rather than TFIH [32•]. Genome sequencing has also uncovered genes encoding proteins closely related to TBP, TFIH, and several TAFs. Interestingly, many of these are expressed in specific tissues. They are likely to be alternative forms of the basal factors that recognize alternative promoter sequences and/or respond to specific activators [33–36]. Therefore, to differentially regulate groups of genes, it appears that nature has evolved several variations on the ‘standard’ mechanism of transcription. So far, these appear to be confined to the factors involved in promoter recognition (e.g. TFIID and TFIIB), analogous to the use of alternative sigma factors in bacteria.

**Post-assembly steps**

Assembly of the initiation complex is only one step in the transcription reaction, and many of the basal factors have essential post-assembly functions. In vivo and in vitro studies with TFIH mutant subunits led to the conclusion that TFIH DNA helicase activity is required for melting the promoter [37,38,39•]. Both TFIH [38,39•] and TFIIF [40–42] are also required after assembly for the polymerase to escape into productive elongation phase.

TFIIB mutants competent for transcription complex assembly, but not for in vitro transcription, indicate that TFIIB has an important post-assembly function [20•,43–45]. Many of these alleles cause start-site shifts in vivo [43,45–48], suggesting that TFIIB may participate in the binding of template DNA in the polymerase active site. Interestingly, the region containing these TFIIB mutants has also been implicated in the response to transcription activators, suggesting that this step might be regulated [49,50].

Recently, there has been remarkable progress in our understanding of other post-initiation events, including regulation of elongation and the link between transcription and RNA processing. These topics are beyond the scope of this article, but were recently reviewed [51,52]. What has become abundantly clear is that transcription is intimately linked to many other processes within the cell and must be studied in that context for a full understanding of gene expression.

**Mechanisms of activation**

Regulation of eukaryotic transcription can occur by at least three mechanisms. The first is often referred to as ‘recruitment’: essentially cooperative binding mediated by an interaction between an activator and some component of the transcription machinery. This appears to be the primary mechanism used in prokaryotes. The second mechanism involves localized chromatin remodeling and modification, again via protein–protein contacts between the activator and the chromatin-modifying enzymes [26••,53•–59•]. The third is by control of elongation processivity, which may be accomplished by events both at the promoter and during elongation itself (reviewed in [49]).

For each mechanism, the challenge has been to unambiguously identify the direct targets of the activator proteins. Reconstitution of activated transcription in vitro led to the identification of multiple coactivator activities defined by different assays [2,60]. Only recently, a coherent picture has begun to emerge. Things have been greatly simplified by the finding that several different coactivator activities (TRAP, DRIP, ARC, CRSP, SMCC, NAT1) are in fact identical or closely related mammalian mediator complexes (reviewed in [2,60]). In several cases, direct interactions with activators were used to facilitate mediator purification. Little is understood about activation domains and what determines their target(s). However, recent experiments in vivo and in vitro have begun to delineate the contacts between activation domains and specific mediator subcomplexes [54•,61•–64•]. A strong case also continues to be made for direct interactions between activators and TAFs within
the TFIID and HAT complexes ([65–67] and references therein). This may be one explanation for why the two complexes share many subunits. Defining the determinants for activator–target interactions at the molecular and structural level is an important future goal.

**Conclusions and future directions**

Our picture of the transcription complex continues to be refined, and future structural studies will undoubtedly provide new insight into the functions of individual factors. In parallel, earlier successes in identifying, purifying, and cloning the transcription factors are now being translated into the tools necessary for analyzing factors in their natural context. In *in vitro* studies will continue to be essential, but *in vivo* studies will become increasingly important for testing models of gene regulation.

**Update**

A high resolution structure of yeast RNA polymerase II has been solved, providing details of the core subunits as well as the smaller eukaryotic-specific subunits (RD Kornberg, personal communication).

**Acknowledgements**

Space limitations prevent this from being a comprehensive review and I apologize for the omission of many interesting papers. I would like to remember Paul Sigler for many entertaining and informative conversations about TBP within the last few years. He showed us many spectacular protein structures and his professional and personal contributions will be missed.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest
** of outstanding interest


3. Fu J, Gnatt AL, Bushnell DA, Jensen GJ, Thompson NE, Burgess RR, ** David FR, Kornberg RD: Yeast RNA polymerase II at 5 Å resolution. *Cell* 1999, 98:799-810. A medium resolution X-ray structure is presented in this paper, which along with references [4**] and [5**], leads to an emerging model of the transcription elongation complex. The polymerase is ‘crab claw’ shaped, with four channels for entering DNA, exiting DNA, NTPs, and the existing transcript.


18. Brand M, Leonard C, Mallouh V, Tora L, Schultz P: Three-dimensional structures of the TAF(II) containing complexes TFIID and TFC. *Tat* 1999, 286:2151-2153. Similar to [17**], this paper shows low resolution structures of both TFIID and a TAF-containing HAT complex. The two structures are remarkably similar.


The authors performed an *in vitro* analysis of transcription complex assembly using an immobilized template assay and various yeast transcription factor mutants and make a case for two kinetically stable intermediates. The first intermediate consists of TFIIID and TFIIB bound to the promoter, the second is a fully assembled transcription complex.


23. Coleman RA, Taggart AK, Buma S, Chicca JJ II, Pugh BF: TFIID regulates TBP and TFIID dimers. *Mol Cell* 1999, 4:451-457. Although TBP can dimerize *in vitro*, there was substantial skepticism about whether this has any physiological relevance. This paper and others from the same laboratory make a case for dimerization being important *in vivo*.
addition to its role in stabilization of TBP at the promoter, TFIIA is found to affect TBP’s monomer–dimer transition.


27. Kuraa L, Struhl K: Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holenzyme. Nature 1999, 399:609-613. Along with [28*], the authors monitored the association of TBP with promoters using chromatin immunoprecipitation. Cross-linking of TBP to promoters correlates strongly with transcription levels and is generally dependent upon activators and the mediator complex.

28. Li XY, Virbasius A, Zhu X, Green MR: Enhancement of TBP binding by activators and general transcription factors. Nature 1999, 399:605-609. These results are similar to [27*], but they also show that TBP crosslinking is dependent upon TFIIA. TFIIA crosslinking shows promoter-specific effects in TATAAF mutation. Furthermore, the Mot1 protein is shown to reduce levels of TBP crosslinking at uninduced genes and at nonpromoter sequences.

29. Auble DT, Steggerda SM: Testing for DNA tracking by MOT1, a TAF mutation. Furthermore, the Mot1 protein is shown to reduce levels of TBP crosslinking at uninduced genes and at nonpromoter sequences.


39. Tirole E, Bussu D, Coif F, Egly JM: Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XBP, XPD, and cdk2. Mol Cell 1999, 3:87-95. The remarkable technical feat of reconstitution of the multifunctional TFIIH complex is documented here; it will open the door to interesting mutational analyses of the individual subunits.


53. Yudkovsky N, Logie C, Hahn S, Peterson CL: Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators. Genes Dev 1999, 13:2369-2374. In this paper, as well as [26*], the authors monitored the association of the SWI/SNF complex and the SRB/mediator body of evidence indicates that activators work not only through cooperative binding with the basal transcription machinery, but also by recruiting chromatin remodeling factors and histone acetyltransferases that act in a highly localized fashion.


  specific and general transcription stimulation by native histone
  See annotation [53*].

60. Berk AJ: Activation of RNA polymerase II transcription. Curr Opin

61. Han SJ, Lee YC, Gim BS, Ryu GH, Park SJ, Lane WS, Kim YJ:
  Activator-specific requirement of yeast mediator proteins for RNA
  polymerase II transcriptional activation. Mol Cell Biol 1999,
  19:979-988.
  Along with [53*•62••64•], as well as mammalian mediator studies summa-
  rized in [2,60], a strong case is made that specific mediator subunits are
  the direct targets of specific activators.

62. Lee YC, Park JM, Min S, Han SJ, Kim YJ: An activator binding
  module of yeast RNA polymerase II holoenzyme. Mol Cell Biol
  See annotation [61*].

63. Lee D, Kim S, Lis JT: Different upstream transcriptional activators
  have distinct coactivator requirements. Genes Dev 1999,
  13:2934-2939.
  The Cup1 promoter in yeast is unusual in its apparent lack of requirement
  for several coactivators, such as TAFs and the mediator component Srb4.
  Using an isogenic set of activator fusions, this paper shows that this inde-
  pendence from coactivators maps to the activation domain of upstream reg-
  ulators. This provides a clear demonstration that not all activation domains
  are equivalent, and they are likely to target different components of the
  transcription reaction.

64. Myers LC, Gustafsson CM, Hayashibara KC, Brown PO,
  Kornberg RD: Mediator protein mutations that selectively abolish
  See annotation [61*].

65. Natarajan K, Jackson BM, Rhee E, Hinnebusch AG: yTAFII61 has a
  general role in RNA polymerase II transcription and is required
  by Gcn4p to recruit the SAGA coactivator complex. Mol Cell

66. Massari ME, Grant PA, Pray-Grant MG, Berger SL, Workman JL,
  Murre C: A conserved motif present in a class of helix-loop-helix
  proteins activates transcription by direct recruitment of the SAGA

67. Wu SY, Thomas MC, Hou SY, Likhite V, Chiang CM: Isolation of
  mouse TFIID and functional characterization of TBP and TFIID in
  mediating estrogen receptor and chromatin transcription. J Biol