The basal initiation machinery: beyond the general transcription factors
Timothy W Sikorski and Stephen Buratowski

In vitro experiments led to a simple model in which basal transcription factors sequentially assembled with RNA Polymerase II to generate a preinitiation complex (PIC). Emerging evidence indicates that PIC composition is not universal, but promoter-dependent. Active promoters are occupied by a mixed population of complexes, including regulatory factors such as NC2, Mot1, Mediator, and TFII S. Recent studies are expanding our understanding of the roles of these factors, demonstrating that their functions are both broader and more context dependent than previously realized.

Address
Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, United States

Corresponding author: Buratowski, Stephen (SteveB@hms.harvard.edu)

Different pathways to basal promoter recognition
Early studies identified the TATA element as a common basal promoter element. This sequence is recognized by the TATA-binding protein (TBP) subunit of TFIIID, the first basal factor to engage the promoter. Indeed, TBP can support basal transcription in vitro without any of the other TAF subunits. However, as attention expanded beyond a small set of strong promoters it became clear that many promoters do not have a recognizable TATA element. Other basal elements identified include the downstream promoter element (DPE), TFIIB recognition elements (BREs), and the initiator element (INR) [4]. Any given promoter may have one or more of these elements, but rarely are they all seen together. The other subunits of TFIIID (the TBP-associated factors or TAFs) appear to interact with INR and DPEs. Besides promoter elements, specific chromatin modifications may play a role in basal factor recruitment. Recent work suggests that at some promoters, trimethylation at the lysine 4 residue of histone H3 stimulates binding of TFIIID [5]. Surprisingly, TAF1 promoter occupancy and gene expression levels correlate at only ~75% of genes [6], casting doubt upon the simple assumption that TFIIID binding necessarily leads to transcription.

In higher eukaryotes, there are multiple genes encoding TBP-related factors (TRFs) and variant TAFs [1,7]. It is presumed that these subunits change the promoter specificity of the different TFIIID variants, a model consistent with the fact that many of them are expressed in specific cell types or developmental stages. Deato et al. characterized a particularly striking example of TFIIID variation [8,9]. They showed that differentiated muscle cells have very low levels of canonical TFIIID and instead found that several muscle-specific genes instead utilize a complex consisting only of the TRF-like protein TRF3 and TAF3. TRF3 is also necessary for development of the hematopoietic lineage in zebrafish, where it activates key differentiation genes [10]. It will be necessary to further characterize the DNA binding properties of the different TFIIID variants to see if their gene specificity is tied to specific promoter elements. Alternatively, they could be targeted using specific upstream activators or somehow compete with each other for similar sequences.

Another mechanism for delivering TBP may be via the SAGA complex, a factor better known as a histone acetyltransferase complex that is recruited to upstream
activating sequences [11,12]. Both genetic and co-immunoprecipitation experiments suggested an interaction between SAGA and TBP [13,14], but a stable complex has not been isolated. However, recent crosslinking experiments support a direct interaction between the Spt3 subunit of SAGA and TBP in vivo [15]. SAGA and TFIID have several subunits in common and are somewhat similar in overall shape [16,17], suggesting they may have evolved from a common ancestor. Gene expression studies in Saccharomyces cerevisiae mutants indicate that ~10% of genes are dependent upon SAGA rather than TFIID for expression [18]. Interestingly, highly inducible genes with clear TATA boxes tend to be SAGA-dependent while TFIID appears to be preferentially used at housekeeping genes without recognizable TATA sequences [19].

The expanding PIC
The minimal set of basal factors does not respond to activators and is insufficient for transcription of chromatin templates, a finding that led to the discovery of a multitude of co-activators and chromatin-modifying complexes. To help define factors directly associated with the basal transcription machinery, a proteomics study of

Many paths to the PIC. The factors and assembly pathways used to form transcriptionally competent preinitiation complexes can be promoter dependent [3,73]. (1) TBP assembling onto promoter regions via TFIID leads to recruitment of the other basal initiation factors, as outlined in the stepwise assembly pathway [1]. In S. cerevisiae, this pathway is most often utilized at TATA-less genes. At some mammalian promoters, histone H3K4 trimethylation helps to recruit the TFIID complex [5]. (2) Mediator bridges interactions between activators and the basal initiation machinery, and can stimulate basal transcription as well. At some promoters Mediator can recruit TFIIH and TFIIE independently of RNApol [45]. (3) TBP can also be brought to promoters by the SAGA complex, in S. cerevisiae, this pathway is most utilized at TATA containing promoters. The Mot1 and NC2 complexes can repress this pathway by actively removing TBP from the TATA element [73]. (4) Mot1 and NC2 can also have a positive role in transcription by removing non-productive TBP complexes from DNA, thereby allowing functional PICs to form [60,65,77].
S. cerevisiae PICs was performed to find factors dependent upon TBP for promoter association. Most known components of the RNApolII transcription machinery were found, as well as several novel components [20]. A new subunit of basal factor TFIIH (Tb5) was discovered, and this protein is necessary for efficient transcription initiation and transcription-coupled repair (TCR) [21,22]. Surprisingly, PICs also contained the elongation factor TFIIIS (discussed below) [23*]. In vitro studies of PIC assembly have been complemented by chromatin immunoprecipitation coupled with DNA microarray hybridization or deep sequencing (ChIP-ChIP or ChIP-Seq). As expected, basal factors are greatly enriched in TBP-dependent promoters [24*,25], although the correlation is not perfect [6]. Importantly, other factors such as TFIIIS, NC2, and Mot1 also show correlation between promoter occupancy and transcription activity (discussed below).

In vitro assembly and in vivo crosslinking experiments have led to suggestions that different promoters utilize distinct subsets of basal factors, or that factors that inhibit transcription are paradoxically found in PICs and must therefore act positively. However, it should be noted that both of these experimental approaches analyze a complex mixture. Just because two factors both bind to immobilized templates or crosslink to the same promoter in vivo, this does not necessarily mean they are present at the same time in the same complex. Multiple complexes are dynamically assembling and disassembling, so promoters may be enriched for certain basal factors because of different rate limiting steps in PIC assembly. Transcription inhibitors may be present at promoters in complexes that are not on the reaction pathway to productive transcription.

**Mediator in activator-independent transcription**

The Mediator complex was isolated as a co-activator that bridges regulatory factors and the basal machinery to allow high levels of activator-dependent transcription [1,26]. However, mounting evidence suggests that Mediator also contributes to basal (activator-independent) transcription, leading to a debate about whether Mediator should be classified as a GTF [27–33]. Some genomewide location analyses in S. cerevisiae and S. pombe found Mediator upstream of almost all active genes and some inactive genes [34,35], but under different growth conditions Mediator did not localize to many active

---

**Table 1**

<table>
<thead>
<tr>
<th>Protein complex</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNApolII</td>
<td>12 Subunits; catalyzes transcription of all mRNAs and a subset of noncoding RNAs including snoRNAs and miRNAs</td>
</tr>
<tr>
<td>TFIIA</td>
<td>2-3 subunits; functions to counteract repressive effects of negative cofactors like NC2; acts as a coactivator by interacting with activators and components of the basal initiation machinery</td>
</tr>
<tr>
<td>TFIIB</td>
<td>Single subunit; stabilizes TFIIA-promoter binding; aids in recruitment of TFIIA/Pol II to the promoter; directs accurate start site selection</td>
</tr>
<tr>
<td>TFIID</td>
<td>14 subunits including TBP and TBP Associated Factors (TAFs); nucleates PIC assembly either through TBP binding to TATA sequences or TAF binding to other promoter sequences; coactivator activity through direct interaction of TAFs and gene specific activators</td>
</tr>
<tr>
<td>TFIIE</td>
<td>2 subunits; helps recruit TFIIH to promoters; stimulates helicase and kinase activities of TFIIH; binds ssDNA and is essential for promoter melting</td>
</tr>
<tr>
<td>TFIIF</td>
<td>2-3 subunits; tightly associates with RNApolII; enhances affinity of RNApolII for TBP-TFIIB-promoter complex; necessary for recruitment of TFIIE/TFIIH to the PIC; aids in start site selection and promoter escape; enhances elongation efficiency</td>
</tr>
<tr>
<td>TFIIH</td>
<td>10 subunits; ATPase/helicase necessary for promoter opening and promoter clearance; helicase activity for transcription coupled DNA repair; kinase activity required for phosphorylation of RNApolII CTD; facilitates transition from initiation to elongation</td>
</tr>
<tr>
<td>Mediator</td>
<td>At least 24 subunits; bridges interaction between activators and basal factors; stimulates both activator dependent and basal transcription; required for transcription from most RNApolII dependent promoters</td>
</tr>
<tr>
<td>SAGA</td>
<td>20 subunits; interacts with activators, histone H3, and TBP; histone acetyltransferase activity; deubiquitinating activity</td>
</tr>
<tr>
<td>Trf1</td>
<td>TBP related factor identified in Drosophila; upregulated in CNS and gonads during development; can bind TATA sequences; mostly found at RNApolII dependent promoters as part of TFIIH but also required at a subset of RNApolII dependent promoters</td>
</tr>
<tr>
<td>Trf2</td>
<td>TBP related factor identified in all metazoans; cannot bind TATA sequences; important for histone gene expression in Drosophila</td>
</tr>
<tr>
<td>Trf3</td>
<td>TBP related factor identified in vertebrates; can bind TATA sequences; important for differentiation of muscle cells in mammals and for haematopoietic cell development in zebrafish</td>
</tr>
<tr>
<td>TFIIIS</td>
<td>1 subunit; stimulates intrinsic transcript cleavage activity of RNApolII allowing backtracking to resume RNA synthesis after transcription arrest; stimulates PIC assembly at some promoters</td>
</tr>
<tr>
<td>NC2</td>
<td>2 subunits; binds TBP/DNA complexes and blocks PIC assembly; can have both positive and negative effects on transcription</td>
</tr>
<tr>
<td>Mot1/bTAF1</td>
<td>1 subunit; induces dissociation of TBP/DNA complexes in ATP dependent manner; can have both positive and negative effects on transcription</td>
</tr>
</tbody>
</table>
promoters [36]. Thus, unlike the basal initiation factors and RNAPII, the correlation between Mediator presence and transcription activity is less clear and Mediator functions may be promoter-specific.

Mediator is not required for basal transcription in purified systems, but can stimulate transcription in these systems even in the absence of activators (see [37] and references therein). It directly interacts with RNAPII and its binding to immobilized in vitro templates is TBP-stimulated, suggesting it assembles with basal factors as a component of PICs. Although in vivo crosslinking suggests Mediator can be recruited to promoters prior to basal factors and RNAPII [38–40] this early association is presumably mediated by activators bound at upstream sites followed by transfer of Mediator to the PIC at the basal promoter. The precise pathway of assembly could be promoter-dependent, explaining the variability in Mediator crosslinking. At some promoters, interdependent recruitment of TFIIB, Pol II, and Mediator in vitro suggests these factors are cooperatively recruited [41]. Indeed, ‘holoenzyme’ forms of Pol II have been co-purified with various Mediator components and basal factors (see [1] and references therein). However, in other in vitro systems Mediator binding is required for TFIIB recruitment but not vice versa [31]. Future work is still needed to determine if specific promoter elements, cofactors, or growth conditions drive a given assembly pathway.

Irrespective of how it is recruited, how does Mediator promote basal transcription? One mechanism may involve stimulating phosphorylation of the RNAPII largest subunit C-terminal domain (CTD) by TFIIH kinase [42]. Although CTD phosphorylation is not required for initiation, this modification leads to release of RNAPII from Mediator and so may promote escape into elongation [43]. Mediator may also directly stabilize PIC assembly intermediates [31,44]. Yeast strains with conditional mutations of the essential Mediator subunit Med11 have reduced RNAPII occupancy, but normal levels of TFIIE and TFIIFII at several constitutively active promoters [45**, suggesting these promoters may not follow the classical stepwise assembly pathway in which TFIIE and TFIIFII are dependent upon RNAPII for incorporation into the PIC (see [1] and references therein). A different point mutation of Med11 decreased the occupancy of the TFIIF submodule of TFIIFII at some but not all active promoters, suggesting that Mediator’s role in PIC recruitment could be promoter dependent as well [45**]. Importantly, Mediator has been shown to stabilize a subcomplex of basal factors at the promoter after initiation in vitro. This ‘ Scaffold’ would then promote subsequent reinitiations [44,46]. The imbalance seen in Med11 strains between RNAPII and basal factors TFIIE and TFIIFII might be due to an inability to utilize Scaffolds for reinitiation in vivo.

Importantly, multiple forms of Mediator can exist within cells, each bearing slightly different subunit compositions and stoichiometries [26]. Future research will focus on where each of these specific forms is recruited, and how they differ in function. B-Med, a form of Mediator isolated from mammalian cell extracts, has been shown to specifically regulate basal transcription in vitro [29,30], although it remains to be seen if this is a physiologically relevant form of Mediator in vivo.

**TFIIS: an elongation factor’s new role at the promoter**

TFIIS is a well-characterized transcription elongation factor that allows arrested RNAPII elongation complexes to backtrack via RNA cleavage and generation of a new 3’ transcript end [47]. Surprisingly, a growing body of evidence indicates this protein also plays a role in initiation. In vitro, TFIIS was found in complexes containing RNAPII and basal factors TFIIFII and TFIIE [48,49], it can directly interact with the promoter-associated factors Spt8 and Med13 [50], and it associated with in vitro assembled PICs [23**]. In vivo, deletion of TFIIS is synthetic lethal with loss of subunits from Mediator and the Swi/Snf chromatin remodeling complex [51,52]. Furthermore, TFIIS was recruited to the promoter of the galactose-inducible gene Gal1, dependent upon Mediator and SAGA but not RNAPII. Loss of TFIIS resulted in reduced recruitment of TBP and RNAPII to the GAL1 promoter in vivo [53].

Yeast TFIIS stimulates in vitro PIC formation on the HIS4 promoter [23**]. This initiation function requires the TFIIS polymerase interaction domain [54], but is independent of transcript cleavage activity [23**]. The TFIIS N-terminal domain, which may interact with Mediator and SAGA subunits, also contributes to PIC assembly in vitro. In agreement, although deletion of TFIIS and Med31 are synthetically lethal, this can be complemented by expression of a TFIIS truncation that interacts with RNAPII but does not stimulate elongation [55]. A single point mutation in the RNAPII interaction domain of TFIIS decreases polymerase recruitment to three promoters tested in vivo. In mammalian cells, transcriptional activators are required for TFIIS-mediated induction of some reporter genes, but not others [56]. Whether this is related to TFIIS’ role in initiation is unclear.

Interestingly, TFIIS has also recently been shown to promote accurate transcription initiation by RNA Polymerase III (RNAPIII), although it appears transcript cleavage activity may be important in this case [57**]. It is unexpected that the RNAPIII system would involve TFIIS because the RPC11 subunit of this polymerase is thought to be homologous to TFIIS [58]. Mammalian genomes contain several TFIIS orthologues, some of which are expressed in a tissue specific manner [47]. It is interesting to speculate about whether multiple TFIIS
molecules function in initiation and elongation independently.

**NC2 and BTA1/Mot1: repressors at the PIC**
Two transcription repressors, Mot1/BTA1 and NC2, act through direct interactions with TBP. Mot1/BTA1 is a Snf2 family ATPase that removes TBP from promoters. It also behaves genetically as a repressor. NC2 is a heterodimer that blocks TFIIA and TFIIIB from associating with the TBP–TATA complex. Its genetic properties are also consistent with transcription repression (see [59,60] and references therein).

Paradoxically, there have been indications that Mot1/BTA1 and NC2 can positively affect gene expression in some contexts. Genomewide crosslinking analyses in yeast show that Mot1, NC2 and TBP are all found at most active promoters [25,61–65]. In mammalian cells, genomewide occupancy of the NC2α subunit correlates with gene activity as well [66*]. The correlation between Mot1 and NC2 is particularly high (>97%) [65**], and the proteins can physically interact [67], suggesting they might function together. Microarray expression analyses in yeast suggest that about 10% of genes are upregulated by Mot1 [63], and about 8% of genes by NC2 [61].

To explain the apparent paradox of inhibitory complexes binding to active promoters, it has been proposed that these repressors (particularly Mot1) displace TBP from cryptic TATA sequences or other inappropriate genomic locations in order to make it available to weaker promoters [62,63,68]. Several recent reports support this model. NC2 alters the conformation of the TBP/DNA complex, allowing it to slide along DNA away from TATA boxes [69†]. TBP mutants with decreased ability to form PICs suppress the gene expression defects seen in a mot1 mutant [70]. FRAP experiments show that the rapid exchange of TBP associated with chromatin is dependent upon Mot1 [71].

Further supporting the promoter redistribution model is the observation that basal promoter sequence strongly affects response to the repressors. In yeast, NC2 and Mot1 repress TATA-containing and activate TATA-less promoters [62,72,73**]. In metazoans, NC2 binding is antagonized by the presence of INR [74*] and BREs [66*,75**]. In Drosophila extracts, NC2 stimulates transcription from DPE-containing promoters, but represses TATA-containing promoters [76]. Manipulation of TBP, Mot1, and NC2 levels in vivo also show opposing effects on TATA versus DPE promoters, leading to the suggestion that NC2 and Mot1 stimulate DPE-dependent transcription by removing TBP from these promoters [77**]. The exact biochemical function of Mot1 and NC2 in this context remains unclear but one interesting possibility to be explored is whether removal of TBP could allow binding of alternative TBP-related factors or TAF complexes to these promoters.

In addition to facilitating transfer of TBP between promoters, NC2 and Mot1 could upregulate expression by displacing transcriptionally inactive forms of TBP from promoters. A dynamic exchange of positive and negative complexes may allow a rapid response to physiological signals [65**,73**]. ChIP experiments at Mot1-dependent promoters show reduced PIC assembly despite increased TBP levels following loss of Mot1 function [78]. NC2 mutants show similar decreases of PIC components at active promoters, although its unclear if the mechanism is related to the removal of inactive TBP from these promoters [79]. Recent RNA sequencing studies have shown that most eukaryotic promoters produce the expected transcripts but also a set of short unstable transcripts synthesized in the opposite direction [80*–83*]. This suggests that basal promoter regions often contain multiple TBP-binding sites or are largely bidirectional [84*]. It has recently been shown that Mot1 can remove TBP bound in the ‘wrong’ direction to free the promoter for productive TBP binding [85*].

Of course, it remains possible that NC2 and Mot1 directly participate in PIC formation. Although these repressors are not typically found in complexes with basal factors other than TBP, ChIP experiments suggest Mot1 can co-occupy promoters with TFIIH and Pol II under heat stress conditions [64]. Mot1 acts in conjunction with SAGA to remodel chromatin at the Gal1 promoter [86] and can physically interact with Mediator and several other chromatin remodeling complexes [87]. Although these observations are not easy to reconcile with structural studies, future experiments may reveal new surprises.

**Conclusion and future prospects**
It is becoming increasingly clear that transcription initiation at basal promoters is not a simple linear reaction. Recent genome and proteome scale analyses of active promoters implicate multiple factors that can both positively and negatively regulate initiation. Assembly pathways may be branched with several non-productive complexes leading to transcription inhibition. In the future it will be important to consider the dynamics of PIC assembly, since chromatin immunoprecipitation and proteomic studies do not provide temporal resolution for distinguishing multiple complexes that can occupy promoters in a population of cells. While non-basal factors such as TFIIIS, NC2 and Mot1 are found at most promoters in vivo, loss of function only affects a small subset of genes. Progress is being made in defining the promoter sequences that determine responsiveness. As there appear to be many varieties of basal sequence elements, it will not be surprising to find heterogeneity in the factors present.
References and recommended reading

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

* of special interest
* of outstanding interest


This review discusses the large variety of promoter types found throughout the genome and provide models for how diverse sets of promoter recognition factors are utilized to match these core promoter elements.


These papers [81st. 93] demonstrate that the loss of TFII D and its replacement by the novel TAF3/TRF3 core promoter complex is required for activation of MyoD responsive genes and the differentiation of muscle cell precursors into myotubes.


This work shows the TBP related factor TRF3 is required for development of the haematopoietic system in the zebrafish embryo and provides a pathway from TRF3 to hematopoietic specific gene expression.


The authors insert a nonnatural photo-reactive amino acid at specific sites on TBP to map direct protein-protein interactions in the context of the PIC. Using this technique they identify a direct interaction between TBP and Spt3, and demonstrate that the interaction is necessary for optimal activation at SAGA dependent promoters.


The authors identify TFII S as a component of PICs in a proteomic screen. TFII S stimulates PIC assembly and transcription in vitro, and IIS cleavage activity is dispensable for this function.


The data presents a genomewide localization map of a diverse set of proteins involved in transcription initiation, including sequence specific activators, chromatin remodelers, mediator, and basal factors, in the context of nucleosomal positions. The authors suggest a model for PIC assembly and concurrent nucleosome removal at promoters.


This paper demonstrates a role for TFIIIS in stimulation of PIC assembly at several active promoters in vivo. This function is particularly important in strains containing mutations of Mediator subunits, thereby explaining previous genetic interactions between these two factors.


This work provides a genomewide occupancy map of TFIIIS in S. cerevisiae, and finds that it localizes to most RNA Polymerase III genes. In vitro and in vivo data suggest TFIIIS is important for initiation for these genes, possibly by aiding in recruitment of TFIIIB and RNApolyll to the promoters.


A genomewide occupancy map of the NC2 complex and Mot1 demonstrates strong co-localization of NC2, Mot1, and TBP at many active promoters. A complex composed of NC2, Mot1, TBP and DNA was isolated from chromatin extracts suggesting these complexes function together to regulate promoter bound TBP.


A genomewide analysis of NC2 localization at promoters in mammalian cells shows that NC2 promoter binding correlates positively with gene expression levels, and that NC2 occupancy is negatively correlated with the presence of TFIIB recognition elements.

67. Kleijman MP, Pereira LA, van Zeeburg HJ, Gillillan S, Meisterernst M, Timmers HT: NC2alpha interacts with BTF1


Using a variety of biophysical methods, the authors find that NC2 induces rapid dynamic changes in the conformation of the TBP-DNA complex, and can mobilize TBP away from TATA regions of promoters. The data suggest that NC2 not only binds TBP to inhibit PIC formation, but also to mobilize TBP on the DNA.


A screen for TBP mutants in yeast finds that strains with defects in PIC assembly no longer require Mot1. The authors suggest that Mot1 may help to maintain appropriate levels of PIC instability which is critical for the regulation of transcription in vivo.


The authors use a computational approach to model PIC assembly regulation in terms of defined biochemical interactions that regulate the function of TBP. Perturbations to the TBP regulatory network are simulated and tested experimentally via genetic mutations. The work shows that pathways to PIC assembly can be modeled with biochemically defined regulatory pathways.


This work shows that presence of INR elements on TATA containing promoters (but not TATA-less promoters) confers resistance of these promoters to NC2 mediated inhibition in vitro. This is an example of how a specific set of promoter elements restrict the binding of a transcriptional regulatory factor.


The authors find that TFIIA preferentially associates with BRE-containing promoters while NC2 is recruited to promoters that lack consensus BRES. Moreover, TFIIA assembly at BRE-containing promoters results in reduced transcriptional activity, while NC2 acts as a positive factor at promoters that lack functional BRES, suggesting a model where by promoter sequence elements can direct the positive or negative functions of initiation factors.


The work shows that TBP overexpression inhibits DPE-dependent transcription. Depletion of Mot1 and NC2 decreases transcription more strongly from TPE-dependent promoters compared to TATA-dependent promoters. The authors suggest a model whereby NC2 and Mot1 stimulate DPE-dependent transcription by removing TBP from these promoters.


This group of papers [80-83] describes a new class of transcripts that initiate near the expected transcription start sites of protein-encoding genes. These RNAs are short, present at low abundance, and often are transcribed in the opposite direction of the protein-encoding mRNA. This challenges our models of how promoter sequence directs initiation of transcription.


The authors provide an analysis of TBP binding sites in human cells, and then obtain binding profiles for 26 transcription factors at these locations, including several subunits of the basal initiation machinery. The authors find many TBP binding sites outside of canonical promoter regions, including in introns, and provide distinct profiles for basal factors at promoters in CpG and non-CpG islands.


The authors show that TBP can bind in the wrong orientation at some promoters in vitro, which inhibits PIC formation. Moreover, Mot1 functions positively at these promoters in part by facilitating redistribution of TBP binding orientation.
