Uncoupling the TFIIH Core and Kinase Modules Leads To Unregulated RNA Polymerase II CTD Serine 5 Phosphorylation

Gabriela Giordano[1], Robin Buratowski[1], Célia Jeronimo[2], Christian Poitras[2]
François Robert[2,3,4], and Stephen Buratowski[1,*]

[1]Department of Biological Chemistry and Molecular Pharmacology,
Harvard Medical School, Boston, MA.

[2]Institut de recherches cliniques de Montréal (IRCM),
Montréal, Québec, Canada

[3]Département de Médecine, Université de Montréal, Montréal, Québec, Canada

[4]Division of Experimental Medicine, Medicine, McGill University, Montréal, Québec, Canada

*Corresponding Author:
steveb@hms.harvard.edu
1-617-432-0696

Abstract

TFIIH is an essential transcription initiation factor for RNA polymerase II (RNApII). This multi-subunit complex comprises two modules that are physically linked by the subunit Tfb3 (MAT1 in metazoans). The TFIIH Core Module, with two DNA-dependent ATPases and several additional subunits, promotes DNA unwinding. The TFIIH Kinase Module phosphorylates Serine 5 of the C-terminal domain (CTD) of RNApII subunit Rpb1, a modification that coordinates exchange of initiation and early elongation factors. While it is not obvious why these two disparate activities are bundled into one factor, the connection may provide temporal coordination during early initiation. Here we show that Tfb3 can be split into two parts to uncouple the TFIIH modules. The resulting cells grow slower than normal, but are viable. Chromatin immunoprecipitation of the split TFIIH shows that the Core Module, but not the Kinase, is properly recruited to promoters. Instead of the normal promoter-proximal peak, high CTD Serine 5 phosphorylation is seen throughout transcribed regions. Therefore, coupling the TFIIH modules is necessary to localize and limit CTD kinase activity to early stages of transcription. These results are consistent with the idea that the two TFIIH modules began as independent functional entities that became connected by Tfb3 during early eukaryotic evolution.

INTRODUCTION

The three nuclear RNA polymerases in eukaryotes share many features, reflecting their evolution from a single precursor (reviewed in (Vannini and Cramer, 2012)). Their multiple subunits are either homologous or identical, and the initiation factors needed to direct all three polymerases to their appropriate target promoters include both the TATA-Binding Protein (TBP) and a TFIIB-like factor. Extending the parallels, RNA polymerases I and III contain intrinsic subunits that resemble the RNA polymerase II (RNApII) initiation factors TFIIE or TFIIF, which help direct the template DNA into the active site cleft.

In contrast to this overall mechanistic conservation, RNApII initiation factor TFIIH has no obvious counterparts in the other polymerase systems. Accordingly, TFIIH performs two functions that are unique to RNApII (Thomas and Chiang, 2006). The first is the use of ATP hydrolysis to accelerate promoter melting, and the second is phosphorylation of a unique C-terminal domain (CTD) on Rpb1, the largest RNApII subunit. These two activities presumably evolved after the divergence of the three RNA polymerases.

The two TFIIH functions map to distinct sub-complexes, sometimes called modules, that are apparent in the TFIIH structure (**Fig. 1A**). Ssl2 (XPB in metazoans) is the DNA-dependent ATPase that pushes double-stranded DNA towards the RNApII active site, leading to strand separation and placement of the template strand in the polymerase active site (Grünberg et al., 2012). Ssl2/XPB interacts with DNA downstream of RNApII, held in place by six other subunits (Tfb1, Tfb2, Tfb4, Tfb5, Rad3/XPD, and Ssl1) that make up the "Core" TFIIH subcomplex. TFIIH binding to the pre-initiation complex (PIC) is mediated by interactions of the Core with TFIIE and the Rpb4/7 stalk of RNApII (He et al., 2016; Robinson et al., 2016; Schilbach et al., 2017). It is not obvious why so many TFIIH subunits should be necessary for the promoter melting function. Interestingly, the Rad3 (XPD in metazoans) subunit is also a DNA-dependent ATPase, but it translocates along DNA in the opposite direction to Ssl2/XPB. Surprisingly, Rad3 makes no contact with DNA in PIC structures, and catalytically inactive Rad3 mutants do not affect transcription (Sung et al., 1988; Winkler et al., 2000). The apparent explanation for the existence of the "extra" subunits and functions in the TFIIH Core is that it also functions as a DNA nucleotide excision repair (NER) enzyme. In that process, both ATPases unwind DNA to expose a short stretch of the damaged strand for removal (Kim et al., 2023; van Eeuwen et al., 2021b; Yu et al., 2023). It therefore seems most likely that the TFIIH Core module originally evolved for DNA repair, and only later was co-opted for RNApII transcription.

The second TFIIH sub-complex carries the CTD phosphorylation activity. The kinase Kin28 (CDK7 in metazoans) and its cyclin partner Ccl1 (Cyclin H) phosphorylate CTD Serine 5. PIC structures show that the Mediator complex helps guide the CTD to the kinase module (Robinson et al., 2016; Schier and Taatjes, 2021; Schilbach et al., 2017; Schilbach et al., 2023). CTD phosphorylation triggers release of Mediator, as well as subsequent recruitment of mRNA capping enzyme and other factors that function near 5' ends of mRNAs (reviewed in (Buratowski, 2009)). Like the Core, the Kinase module also has a second function unrelated to transcription. At least in metazoans, CDK7/cyclin H serves as a CDK-activating kinase (CAK), phosphorylating the T-loop of other cyclin-dependent kinases to promote an active conformation (Fisher and Morgan, 1994). This is not a universal function, as *Saccharomyces cerevisiae* and some other species have a separate CAK enzyme (Cismowski et al., 1995;

Espinoza et al., 1996; Kaldis et al., 1996). Nonetheless, it seems likely that the activity of CDK7 on RNApII evolved from a pre-existing CDK that had either CAK activity or some other function.

The two TFIIH subcomplexes are linked by the Tfb3 subunit (known as MAT1 in metazoans). Tfb3/MAT1 consists of two globular domains and a linker region. The N-terminal domain is part of the Core module, where it helps incorporate TFIIH into the PIC through direct interactions with RNApII and TFIIE (Schier and Taatjes, 2021). The Tfb3 C-terminal domain contacts both the kinase and cyclin subunits, and this interaction promotes an active kinase conformation (Greber et al., 2020; Peissert et al., 2020; van Eeuwen et al., 2021a). PIC structures show the linker emerging from the N-terminal domain as a long alpha-helix running along the interface between the two ATPase subunits, followed by a turn and a short stretch of helix just N-terminal to a disordered region that connects to the C-terminal region (see schematic in **Fig. 1A**). A deletion study in yeast showed that Tfb3 tolerates 30 amino acid deletions throughout the linker, but that the N- and C-terminal domains were essential for viability (Warfield et al., 2016).

Given its modular structure, we tested whether other CTD kinases could be fused to Tfb3 to bypass the requirement for Kin28/CDK7. While the results have so far been negative, we unexpectedly found that cells lacking the C-terminal domain of Tfb3 were actually viable, albeit very slow growing. Remarkably, growth is significantly rescued when the two parts of Tfb3 are expressed as separate proteins. In these cells, the TFIIH kinase module is no longer tethered to the Core TFIIH. ChIP experiments show that the Core Module, but not the Kinase, is still promoter localized. Surprisingly, phosphorylation of CTD Serine 5 is now seen at high levels throughout transcribed regions, indicating that the untethered kinase accesses RNApII throughout the cell cycle. These results are consistent with the idea that the two TFIIH modules began as independent entities that only later became fused into a single complex.

RESULTS

Tethering other kinase modules to Tfb3 does not bypass the requirement for Kin28. Given the apparent modularity of TFIIH, we asked whether tethering other kinases to Tfb3 could functionally replace Kin28. Although Kin28 phosphorylates CTD Serine 5 in vivo, other kinases can phosphorylate this residue in vitro. Two other transcription kinase complexes, Bur1/Bur2 (the yeast homologs of CDK9/cyclin T) and Ctk1/Ctk2/Ctk3 (yeast CDK12/Cyclin K), phosphorylate the Rpb1 linker (Chun et al., 2019) and CTD Serine 2 (Cho et al., 2001), respectively. However, both are reported to also have activity on Serine 5, at least in vitro (Corden, 2013; Eick and Geyer, 2013; Glover-Cutter et al., 2009). Accordingly, we constructed Tfb3 fusions **(Fig. 1B, Supplemental Table 1)** in which the C-terminal region was replaced with either Bur2 (the cyclin partner of Bur1) or Ctk3 (the third subunit within the Ctk1/Ctk2 CDK/cyclin complex). We also created a fusion of the Tfb3 N-terminal domain and linker to Mpk1, a single subunit kinase also reported to have activity on CTD Serine 5, as well as Ser2 and Tyr1 (Yurko et al., 2017).

The three fusion constructs were introduced into a yeast strain lacking the chromosomal copy of *KIN28*. Expression of the fusion proteins was confirmed by immunoblotting (**Supplemental Fig. S1**). None of the fusions supported viability after plasmid shuffling to remove *KIN28* (**Fig. 1C**), indicating that tethering the other kinases could not substitute for Kin28 function. As a control experiment, the Tfb3 fusions were also tested for complementation of a *TFB3* deletion. Surprisingly, these shuffled strains proved to be viable, albeit slow growing **(Fig. 1D)**. Complementation of *tfb3*Δ indicates that the fusion proteins were incorporated into TFIIH. This result was unexpected because a *TFB3* allele encoding amino acids 1-275, and therefore lacking the C-terminal domain, did not support growth, consistent with earlier reports that the C-terminal domain is essential for Tfb3 function and cell viability (Feaver et al., 2000; Warfield et al., 2016).

The N-terminal domain is the only essential part of Tfb3. The unexpected complementation of *tfb3*Δ by the Tfb3 fusions raised the question of whether the Tfb3 C-terminal domain was truly essential for viability. We created a new set of Tfb3 truncations, each lacking either one or two of the three major Tfb3 domains **(Fig. 2A)**, and used plasmid shuffling to introduce them into a *tfb3*Δ strain. Cells lacking the N-terminal domain failed to grow, no matter how long the plates were incubated (data not shown). In contrast, strains expressing only the N-terminal (N-term) domain, or the N-terminal and Linker (N-term-Linker) domains, exhibited slow growth **(Fig. 2B**). These deletions grew at rates similar to the fusion proteins tested above, arguing that the fused kinase module did not provide an additional function or stability relative to a Tfb3 completely lacking any sequence C-terminal to the Linker region. We therefore conclude that, while important for normal growth, the C-terminal domain of Tfb3 (which is part of the Kinase Module) is not essential for viability.

Expression of the TFIIH Core and Kinase Modules as independent complexes. Recent structural studies show that the C-terminal domain of Tfb3, or its metazoan homolog MAT1, stabilizes an active kinase conformation when bound to Kin28/CDK7 and its partner cyclin (Greber et al., 2020; Peissert et al., 2020; van Eeuwen et al., 2021a). Therefore, the slow growth of strains lacking the Tfb3 C-terminal domain could be due to loss of kinase tethering to the Core Module, to loss of Kin28/CDK7 activation, or both.

To test the contribution of the Tfb3 kinase activation function separately from kinase tethering, parts of Tfb3 were expressed as two separate proteins (**Fig. 2C**). Strains expressing either the N-term (left panel) or N-term-Linker (right panel) derivatives were transformed with plasmids expressing the indicated C-terminal (C-term) domain constructs. These strains were tested for growth by serial dilution spotting, where the size of single colonies serves as a measure of relative growth rate. Again, strains lacking any C-term domain grew very slowly (Vector, top row). In the N-term strain (left panel), adding wild-type Tfb3 restored growth to normal. Co-expression of the separate N-term and C-term domains noticeably improved growth compared to N-term alone, despite the fact that this strain completely lacks the Linker region. Co-expression of the Linker-C-term construct rescued growth to an even greater extent, although not quite to wild-type rates. Therefore, the Linker does contribute to full function of the Kinase Module.

Unexpectedly, expression of the N-term-Linker construct (**Fig. 2C**, right panel) had a dominant negative effect, as it caused slow growth even in the presence of WT Tfb3 (right panel, second row). Nonetheless, some growth improvement was seen by co-expressing C-term, but only using a high copy plasmid likely to produce more protein. Co-expressing the Linker-C-term construct gave a similar growth improvement, but at both high and low copy. Altogether, the co-expression results show that the Tfb3 N-terminal and C-terminal domains can function independently of covalent physical linkage, and suggest that the Linker region primarily contributes to function of the Kinase Module.

We considered the possibility that, despite the split Tfb3, the TFIIH Kinase and Core Modules could remain associated through other subunit contacts. To test this possibility, a tandem affinity purification (TAP) tag was used to precipitate full-length, C-term, or Linker-C-term Tfb3 derivatives using IgG beads. While there was some variation in protein levels (**Fig. 3A**, left panel, lanes 1-4), similar amounts of all three TAP fusions were precipitated with IgG beads (lanes 5-8). The Kinase Module subunits Kin28 and Ccl1 interacted with full-length or Tfb3 C-terminal constructs equally well (**Fig. 3B**). An in vitro phosphorylation assay using GST-CTD as substrate showed that Kinase Modules from the truncated Tfb3 alleles retained activities similar to that of full-length Tfb3 (**Supplemental Fig. S2**).

In contrast to the Kinase Module, the TFIIH Core Module (monitored by Core subunit Tfb1) no longer associates strongly with the uncoupled Tfb3 C-term derivatives (**Fig. 4A**, compare lanes 3 and 4 to full-length Tfb3 in lane 2). The Linker-C-term derivative may retain some weak interaction with the Core, but the Tfb1 levels were just slightly above background binding in the untagged strain (lane 1), and far below that seen with full length Tfb3. We conclude that the split Tfb3 does indeed dissociate the two TFIIH modules.

The basal transcription factor TFIIE directly contacts several Core TFIIH subunits, including Tfb3, within the initiation complex. Furthermore, TFIIE stimulates TFIIH enzymatic activities, including the kinase (Watanabe et al., 2003). To test how the TFIIE-TFIIH association was affected by the split Tfb3, the TAP-precipitated samples were also immunoblotted for the TFIIE subunits Tfa1 and Tfa2. Whereas full length Tfb3 efficiently pulls down both TFIIE subunits, the separated C-term or Linker-C-term derivatives do not **(Fig. 4B)**. This result further indicates that the Kinase Module is no longer associated with the TFIIH Core upon splitting Tfb3 into two parts.

Uncoupling the TFIIH modules disrupts proper kinase localization. Given the viability of the split Tfb3 strains, we wondered if the untethered Kinase Module was still delivered to promoters, perhaps via interaction with Mediator. Genome-wide chromatin immunoprecipitation (ChIP-Seq) was used to compare localization of the TFIIH modules in the wild-type and split Tfb3 strains. In cells expressing full-length Tfb3, the Core TFIIH subunit Tfb1 showed a strong peak near the transcription start site, reflecting promoter localization in the PIC (**Fig 5A and Supplemental Fig. S3A**, red line**)**. In all three split Tfb3 strains (yellow, blue, and green), the Tfb1 signal was reduced relative to the full-length, but still clearly promoter-localized. Perhaps reflecting reduced PIC formation, a similar reduction was seen for RNApII subunit Rpb1 (**Fig 5D and Supplemental Fig. S3D**). Interestingly, the Tfb1 signal was closest to normal in the strain in which the Linker was fused to the N-terminal domain, despite the fact that this strain grew markedly slower than strains with the Linker on the C-terminal domain or missing completely (**Fig 2C**). In contrast to Core TFIIH, the wild-type Kin28 ChIP signal localized at promoters was completely absent in all the split Tfb3 strains (**Fig 5B** and **Supplemental Fig. S3B**). Therefore, although Core TFIIH is still properly localized, the untethered Kinase Module is no longer recruited to promoters.

We expected that loss of Kinase Module recruitment would affect CTD Serine 5 phosphorylation (Ser5P) negatively. This prediction was first tested using immunoblotting of whole cell extracts. Cells lacking any Tfb3 C-term were reduced for both Ser5P and Ser2P (**Supplemental Fig. S4**, lanes 2 and 3). Similar to the growth phenotypes, adding back the C-term as a separate protein greatly improved CTD phosphorylation levels, albeit not to full wild-type levels (lanes 4 and 5). However, ChIP-seq revealed that Ser5P was abnormally distributed in cells with split Tfb3. Normally highest at promoters and at low levels throughout transcribed regions, Ser5P was now more uniform throughout the transcribed region in the mutants (**Fig 5C and Supplemental Fig. S3C**). Normalization to total RNApII (**Fig 5D and Supplemental Fig. S3D**) showed that Ser5P was maintained at high levels throughout transcription (**Fig 5E**). Plots of changes at individual genes (**Supplemental Fig. S3E, F**) showed that the metagene analysis reflected the response of virtually all genes. While the normalized Ser5P signals at promoters are reduced by 50% or less at strongly transcribed genes (i.e. those with highest Rpb1 signal), the distribution for the mutant/WT ratio is centered around one (zero in log2, **Supplemental Fig. S3E**, left panels). In contrast, the downstream signal in open reading frames (right panels) are nearly all increased by two-fold or more in the mutants. Interestingly, while transcription is clearly down overall in the split Tfb3 strains (**Supplemental Fig. S3F**), there were seven genes with reproducibly increased Rpb1 signals. All but one are known heat shock response genes (**Supplemental Fig. S3G**), suggesting these slow growing cells are under stress.

The surprising increase of Ser5P downstream of the promoter suggests that the Tfb3 linkage not only delivers the Kinase Module to the promoter, but may also inhibit its activity outside of the PIC. Transcription-associated phosphatases normally cause Ser5P to drop soon after RNApII enters the body of the gene (reviewed in (Buratowski, 2009; Corden, 2013; Eick and Geyer, 2013)), so the higher downstream Ser5P signal suggests untethered kinase module may be able to continually access the CTD during elongation. This could be due to inhibition of TFIIH Core-associated Kinase outside of the PIC, or indicate that the majority of TFIIH is at promoters and therefore unavailable at later stages of transcription.

Disconnected Core TFIIH does not cause an NER defect. The TFIIH Core Module also functions as a key component of the NER machinery. While many mutations in Core subunits therefore cause sensitivity to ultraviolet (UV) light, this is not the case for Kinase Module mutants. Recent structures (reviewed in (Yu et al., 2023)) show that the Kinase Module would block interactions between the Core Module and other NER factors. Therefore, TFIIH either enters into the NER complex as free Core Module, or the Kinase Module must dissociate soon after. We tested the response to UV light in strains where the split Tfb3 dissociates the Core and Kinase Modules (**Fig. 6**). Two wild-type Tfb3 strains acted as controls for normal UV response, including one strain lacking the Tfb6 protein that competes with the Core for Ssl2 binding (Murakami et al., 2012), as this background was used for many experiments here (see **Supplemental Table S2**). None of the split Tfb3 configurations caused UV sensitivity, indicating that the separated N-term of Tfb3 is sufficient for the Core NER function. In fact, we noticed that strains with free N-term domain, where the linker is either absent or connected to the C-terminal domain, may even have slight resistance to UV light, consistent with a model in which the Kinase Module inhibits the TFIIH NER function.

DISCUSSION

We show here that the Core and Kinase Modules of TFIIH can be separated by expressing the Tfb3/MAT1 protein as two separate parts (**Fig. 2**). Although it was previously reported that both the N-term and C-term of Tfb3 are essential for viability in *S. cerevisiae* (Warfield et al., 2016), we found that cells lacking the C-terminal domain are viable but extremely slow growing. Interestingly, accurate transcription initiation could be reconstituted in vitro using a Tfb3 truncation consisting of only N-term residues 1-148 (Yang et al., 2022). However, we found that adding back the C-terminal domain as a separate protein dramatically improved growth (**Fig. 2**). Notably, Mouse Embryonic Fibroblasts and cardiomyocytes with a homozygous deletion of *MAT1* were shown to be viable but abnormal (Helenius et al., 2011; Sano et al., 2007). In these cells, Core TFIIH remains intact, but levels of the Kinase Module proteins CDK7 and Cyclin H are severely reduced. *MAT1* deletion leads to major defects in CTD phosphorylation and transcription, but the effects are partially compensated by reduced mRNA degradation rates. The residual CTD Serine 5 phosphorylation is presumably due to the remaining untethered kinase. Therefore, our yeast results splitting the Core and Kinase Modules are likely also applicable to mammalian TFIIH.

Separating the Core and Kinase Modules leads to deregulated CTD phosphorylation. ChIP-Seq showed that Kin28 kinase was no longer recruited to promoters in the split Tfb3 strains, even though Core TFIIH still showed proper localization (**Fig. 5**). Having uncoupled the Kinase Module from TFIIH, we expected CTD Ser5P to be severely reduced in the split Tfb3 strains. Instead, Ser5P was evenly distributed throughout the transcribed region, at levels below the normal promoter peak but higher than normal downstream. Therefore, the untethered Kinase Module, but not free TFIIH, can apparently access the CTD throughout the transcription cycle. A surprising implication is that, in addition to delivering the kinase to the initiation complex, the connected TFIIH Core may help limit CTD phosphorylation to early transcription. We previously showed that CTD kinase activity of free trimer Kinase Module was less sensitive than TFIIH complex to mutations in the activating T-loop (Keogh et al., 2002). Therefore, the Core module may have some inhibitory effect on CTD kinase activity that is relieved when TFIIH is incorporated into the PIC. However, precipitates with full-length Tfb3 did not have less kinase activity than those from split Tfb3 cells (**Supplemental Fig. S2**). It is also possible that most Kinase Module is tethered to the Core at the promoter and therefore limiting for access downstream RNApII.

Increased downstream Ser5 phosphorylation in the split Tfb3 strain raises an interesting question about normal CTD phosphorylation patterns. The earliest ChIP experiments (Komarnitsky et al., 2000), as well as subsequent ChIP-seq studies, detected some Ser5P throughout transcribed regions, despite removal of most Ser5P by phosphatases soon after promoter escape. Mass spectrometry of RNApII pulled down with an mRNA termination factor also detected significant Ser5P in late elongation (Suh et al., 2016). It has been presumed that this downstream signal represented residual phosphatase-resistant Ser5P acquired at the promoter. However, an alternative model that must now be considered is that free Kin28/Cdk7 kinase trimer continues to carry out de novo phosphorylation throughout the transcription cycle. Free CDK7/Cyclin H/MAT1 trimer exists in metazoans to provide CAK activity, but free yeast Kin28/Ccl1/Tfb3 trimer is also detected in yeast (Keogh et al., 2002). A dynamic equilibrium between kinase and phosphatase activities regulate CTD Serine 2 phosphorylation

during elongation (Cho et al., 2001), and a similar dynamic may exist for Ser5P well downstream of promoters.

Communication between the Core and Kinase Modules. The viability of a split Tfb3 strain completely lacking the linker might suggest this central region serves only to tether the N- and C-terminal domains. However, available TFIIH structures and our data suggest otherwise. Cells expressing a Linker-C-term fusion grew better than cells completely lacking the linker. In contrast, The N-term-Linker fusion protein acts as a dominant negative allele, slowing growth even in the presence of wild-type Tfb3 (**Fig. 2C**). These results suggest a linker function in addition to tethering the two modules, and TFIIH structural studies suggest a model for what that function might be.

Structures are available for free TFIIH complex, TFIIH incorporated into the PIC, and TFIIH as part of the NER machinery (reviewed in (Yu et al., 2023)). The Core Module has a roughly horseshoe shape, with the TFIIH ATPases, Ssl2/XPB and Rad3/XPD, at the two ends. However, the spacing between the ATPases changes substantially depending on context. In free TFIIH, the long alpha-helix in the Tfb3/MAT1 Linker spans from Tfb3 N-term, which is contacting Rad3/XPD, to the other side of the horseshoe, where it directly contacts Ssl2/XPB. Here, the central Linker domain serves to lock together the two TFIIH ATPases in a "closed" conformation. In the PIC, the Tfb3/MAT1 alpha-helix is disconnected from Ssl2/XPB, producing an "open" conformation with the Ssl2/XPB rotated outward, its active site bound to the double-stranded DNA downstream of RNApII. TFIIH within the NER complex lacks the Kinase Module, and in this case the NER factors bring Rad3/XPD and Ssl2/XPB even closer together to facilitate the interaction of the damaged DNA with both ATPase active sites (Yu et al., 2023).

A recent report describing two new PIC structures (Abril-Garrido et al., 2023) shows that when the +1 nucleosome blocks Ssl2/XPB from downstream DNA, TFIIH is in the closed conformation, with the Tfb3/MAT1 Linker contacting both ATPases. Positioning the nucleosome further downstream allows the open TFIIH conformation, resulting in loss of the contact between the Tfb3/MAT1 Linker and Ssl2/XPB. Notably, much of the Tfb3 long alpha-helix becomes disordered when disconnected from Ssl2/XPB. These two new PIC structures lack Mediator, and it hasn't been shown that the closed TFIIH PIC is actually a necessary precursor to the open TFIIH. However, it seems quite plausible that TFIIH enters the PIC in the closed conformation, but transitions to the open conformation upon DNA binding in order to proceed with DNA translocation and promoter melting.

Based on our findings, we speculate that the Tfb3 Linker could provide a mechanism for temporal coupling of TFIIH kinase and promoter melting functions. PICs with closed TFIIH are predicted to be inhibited for ATP hydrolysis by Ssl2/XPB (Abril-Garrido et al., 2023), and as described above, may also be partially inhibited for kinase activity. We propose that movement of the Kinase Module in the PIC repositions the Tfb3 Linker to release the inhibitory contact with Ssl2/XPB, freeing the translocase for promoter melting. This movement may involve capture of the Kinase Module by Mediator, and perhaps even CTD phosphorylation. Supporting the idea that the Kinase Module is mobile, PICs formed in vitro without Mediator show the Kinase Module close to the Rpb4/Rpb7 stalk of RNApII (Abril-Garrido et al., 2023; Chen et al., 2021), far from its position seen when interacting with Mediator. CTD phosphorylation triggers release of Mediator (Søgaard and Svejstrup, 2007), and in vivo experiments with TFIIH kinase inhibition suggests this transition occurs very rapidly after PIC formation (Jeronimo and Robert,

2014; Wong et al., 2014). Notably, release of the Tfb3 Linker contact also results in the long alpha-helix becoming disordered, which could allow the kinase access to a far larger radius of area. This flexibility could help the kinase reach both proximal and distal repeats within the CTD, which can theoretically extend quite far from the RNApII body.

It has long been known that yeast PICs access transcription start sites (TSSs) far further downstream than metazoan PICs. Enzymatic studies show that for yeast TFIIH, but not its mammalian homolog, the Kinase Module greatly stimulates the processivity of the Ssl2 translocase activity necessary for downstream TSS scanning (Tomko et al., 2021). This effect is dependent only on the presence of the Kinase Module, not kinase activity. Interestingly, yeast transcription reconstituted in vitro with TFIIH lacking Tfb3 and the Kinase Module causes a dramatic upstream shift in transcription start sites, with the TATA-TSS spacing now resembling that of metazoans (Murakami et al., 2015). This effect can be reversed by adding back only the Tfb3 N-term (Yang et al., 2022), arguing against a model where the contact between Tfb3 and Ssl2 is necessary for downstream TSS usage. Nevertheless, there is clearly some communication between the Kinase and Core Modules, and Tfb3 is likely to be a key link.

Evolutionary considerations. We speculate that cells with a complete deletion of the Tfb3 Linker might resemble an evolutionarily ancient state where TFIIH Kinase and Core existed as completely distinct enzymes. Initially, the primary role of TFIIH Core was likely NER, while the Kinase Module precursor functioned to phosphorylate CDKs or other substrates. It is likely each was independently co-opted for RNApII transcription. Supporting this hypothesis, TFIIH in the early diverged eukaryote *Trypanosome brucei* appears to consist of only the Core Module, as no Tfb3 or other Kinase Module homologs are found in the purified enzyme complex or by sequence similarity search (Lee et al., 2010; Lee et al., 2009). The *T. brucei* CTD is quite diverged, without the typical repeat sequence found in most eukaryotes (Badjatia et al., 2013; Das and Bellofatto, 2009). In the main eukaryotic lineage, the Tfb3 Linker likely evolved over time to coordinate the timing of CTD phosphorylation and promoter melting.

In the split Tfb3 strains, fusion of Linker to the C-term domain may improve growth relative to the complete Linker deletion by partially restoring kinase-translocase coupling. In contrast, the observed toxicity of the Tfb3 N-term-Linker fusion may result from the Core TFIIH being locked into the closed conformation. Our ChIP-Seq results suggest an ancestral, free CTD kinase may have acted throughout elongation. Tethering the kinase to TFIIH may have guaranteed high levels of CTD phosphorylation during early transcription, but may also have opened the door for the CTD Code as we know it today, where distinct phosphorylation patterns to mark the progression of the transcription cycle.

Acknowledgements

We are grateful to Steve Hahn (FHCRC, Seattle) and John Feaver (Stanford) for providing several Tfb3 strains and plasmids, and to Jason Tanny (McGill Univ., Montreal) for *S. pombe* cells. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR; PJT-162334) to F.R., and by US NIH grants GM046498 and GM056663 to S.B.

METHODS

Plasmid construction and Yeast Strains. Plasmids used in this study, and their construction, are listed in **Supplemental Table S1**. Yeast strains used in this study are listed in **Supplemental Table S2**.

Antibodies and Immunoblotting. The following primary antisera were used: anti-Ccl1 (Keogh et al., 2002), anti-Kin28 (Covance, PRB-250C-500), anti-Tfb1 (Matsui et al., 1995), anti-Tfa1 and anti-Tfa2 (Kuldell and Buratowski, 1997), anti-Ser2P (3E10), anti-Ser5P (3E8) (gift from Dirk Eick (Chapman et al., 2007)), anti-CTD (8WG16) , anti-Rpb3 monoclonal 1Y26 (BioLegend). For immunoblotting, primary antibodies were used at a 1:1000 dilution. Secondary antibodies coupled to horse radish peroxidase (goat anti-rabbit, goat anti-rat, and goat anti-mouse) were purchased from Sigma and used at the recommended dilution.

Anti-Tfb3 antibodies (Keogh et al., 2002) were affinity purified from sera to reduce cross-reacting signals. Briefly, recombinant Tfb3 protein was expressed from plasmid pSBEThis7-Tfb3 in *E. coli* BL21(DE3). Cells were grown at 37°C to an OD_{600} of 0.4-0.6 in 4 L of selective media. Expression was induced with 0.2mM IPTG and cultures were incubated shaking at 30°C for 4 hrs. Cells were pelleted and resuspended in 100 mL lysis buffer (1X PBS, 1% SDS, pH 7.4). After sonication, cell debris was pelleted at 23,970 g for 20 min. The supernatant was filtered and then incubated with nutation at 4°C for 1 hr with Ni-NTA agarose (Gold Biotechnology #H-350-100) that was pre-equilibrated with lysis buffer. Lysate and beads were drained through a column and washed extensively with wash buffer (1X Phosphate Buffered Saline (PBS), 0.1% Sarkosyl, 10 mM imidazole, pH 7.4). Bound protein was serially eluted with elution buffer (1X PBS, 0.1% Sarkosyl, pH 7.4) supplemented with increasing imidazole concentrations (1 mL each): 50 mM, 100 mM, 150 mM (2X), 200 mM, 300 mM. Fractions with the highest Tfb3 yield were pooled. Roughly 2 mg of purified Tfb3 was covalently linked to 1 ml AminoLink® Plus Coupling Resin (Thermo Scientific #PI20505) following the manufacturer's protocol for crosslinking at pH 10 with cyanoborohydride solution. Specific anti-Tfb3 antibodies were affinity purified from crude antiserum as described (Harlow and Lane, 1988). Tfb3-coupled beads were serially washed with 10 bed volumes each of 10 mM Tris pH 7.5, 100 mM glycine pH 2.5, 10 mM Tris pH 8.8, and then 100 mM triethylamine pH 11.5. Finally, beads were poured into a small column and washed with 10 mM Tris pH 7.5 until pH reached 7.5. Antiserum was diluted 1:10 with 10 mM Tris pH 7.5 and passed through the column three times. The column was then washed with 20 bed volumes each of 10 mM Tris pH 7.5 and 500 mM NaCl, and then 10 mM Tris pH 7.5. Bound antibodies were serially eluted with small volumes of high (100 mM triethylamine pH 11.5) and low (100 mM glycine pH 2.5) pH, followed by rapid neutralization. Peak fractions were pooled and concentrated using Amicon Ultra 0.5 10K Concentrator, with PBS as the replacement buffer.

Yeast whole cell extracts. Two types of extracts were used for gel electrophoresis and immunoblotting, with similar results. The first is an alkaline lysis method. 1.5 mL of late log phase liquid culture was pelleted at 4,600 g for 90 sec at room temperature. Media was removed, and the pellet was resuspended in 0.1 M NaOH and incubated at room temperature for 5 min. Cells were pelleted at 9,300 g for 90 sec at room temperature, supernatant was removed, and pellet was resuspended in 60 µL lysis buffer (0.06 M Tris-HCl pH 6.8, 5% glycerol, 2% SDS, 4% β-mercaptoethanol, 0.0025% bromophenol blue) (Kushnirov, 2000). Samples were heated at 95°C for 5 min, pelleted at 9,300 g for 1:30 min at room temperature, and supernatant was collected to run on an SDS-PAGE gel.

For some gels, a non-denaturing whole cell extract was used. First, 1.0-1.5 mL of liquid culture was pelleted (all centrifugations were at 2,300 g) at 4°C for 2 min and washed with 1 mL cold distilled water. Pellets were resuspended in 100 µL cold lysis buffer (50 mM HEPES pH 7.6, 1 mM EDTA, 300 mM NaCl, 20% glycerol, 1 mM DTT, 1 µg/mL leupeptin, 1 µg/mL aprotinin, 1 µg/mL pepstatin A, and 1 µg/mL antipain). The resuspension was mixed with 100 µL 425-600 µm glass beads (Sigma, #G8772-1KG) and vortexed at 4°C for 10 min. Extract was centrifuged for 10 min at 4°C and supernatant collected. Total protein concentration of extracts was measured using NanoDrop 2000 Spectrophotometer (ThermoScientific)**.**

TAP purification and co-precipitation assay. TAP-tagged Tfb3 derivatives were used for assaying Kinase Module association and in vitro kinase assays. Yeast strains were grown in 200 mL of selective media at 30°C to an OD_{600} of ~2. Cells were collected by centrifugation, flash frozen in liquid nitrogen, and lysed using a SPEX SamplePrep Freezer Mill 6875D with small chambers. Pulverized cells were resuspended by rotation at 4°C in 1 mL of lysis buffer (20 mM HEPES pH 7.6, 10% glycerol, 200 mM KoAc, 1 mM EDTA, 1 mM DTT, 1 mM NaF, 1 mM Na3VO4, 1 µg/mL leupeptin, 1 µg/mL aprotinin, 1 µg/mL pepstatin A, 1 µg/mL antipain, 1 mM PMSF) until homogeneous. Cell debris was pelleted in a benchtop microfuge at 4,500 g for 25 min, and supernatant was collected and further cleared in a microfuge at maximum speed for 10 min. All centrifugations were done at 4°C. The clear layer of lysate was transferred to a new tube, 10 µL of IgG Sepharose (GE Healthcare, #17096901) pre-equilibrated with lysis buffer was added, and the suspension was rotated at 4°C for 1 hr. After binding, IgG beads were gently pelleted, supernatant was removed, and beads were washed 3 times with lysis buffer.

For analysis of bound proteins, 50 µL of 2X SDS-PAGE sample buffer (0.1% bromophenol blue, 0.2 M DTT, 20% glycerol, 4% SDS, 0.1 M Tris-Cl pH 6.8) were added to the beads and the sample was heated at 100°C for 5 min. The beads were then pelleted by centrifugation, supernatant was collected, and samples were analyzed by immunoblotting. Input, flowthrough, and final samples from the co-immunoprecipitation protocol were run on 10% polyacrylamide SDS-PAGE gels. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes. Membranes were blocked with 5% powdered milk in Tris Buffered Saline with 0.1% Tween and then probed with the appropriate primary antibody. Blocking and antibody incubations were done for either 1 hr at room temperature or overnight at 4°C. Following several washes, the blots were incubated with secondary antibody at a 1:10,000 dilution for 1 hr. Reacting bands were detected by chemiluminescent detection using the Thermo SuperSignal Plus Pico kit according to the manufacturer's instructions.

In vitro CTD phosphorylation assay. IgG beads were used to immobilize the Tfb3-TAP derivatives as described above. After normalizing for the amount of Kin28, beads were

incubated with 100 ng GST-CTD in 1x kinase reaction buffer (20 mM HEPES pH 7.6, 7.5 mM MgOAc, 2% Glycerol, 100 mM KOAc, 2 mM DTT, and 25 μM ATP). Samples were taken at 15, 30, and 60 min, and phosphorylation was assayed by immunoblotting with anti-Ser5P (3E8) antibody.

ChIP-seq. ChIP experiments were performed from two independent biological replicates. Yeast cultures were grown in 50 mL of selective medium at 30°C to an OD_{600} of 0.7-0.9 before crosslinking with 1% formaldehyde (Fisher Scientific, BP531-500) at room temperature for 30 min and quenched with 125 mM glycine. Crosslinked cells were collected by centrifugation and washed twice with 1X TBS (20 mM Tris-HCl pH 7.5, 150 mM NaCl). Cell pellets were then resuspended in 700 µL Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and protease inhibitor cocktail (1 mM PMSF, 1 mM benzamidine, 10 µg/mL aprotinin, 1 µg/mL leupeptin, 1 µg/mL pepstatin A)). About the same number of OD_{600} units was used for all samples. A fixed amount of crosslinked *Schizosaccharomyces pombe* cells (a gift from Jason Tanny), at 10% the amount of *S. cerevisiae* cells, was added for normalization as previously described (Jeronimo et al., 2019). Cells were lysed by bead beating and the lysate was sonicated with a Model 100 Sonic dismembrator equipped with a microprobe (Fisher Scientific), 4 x 20 sec at output 7 Watts, with a 1 min break between sonication cycles. Soluble fragmented chromatin was recovered by centrifugation. 600 µL of the chromatin sample was taken per immunoprecipitation (IP) and 6 µL (1%) was saved as an Input sample. The following amounts of antibody per IP were used: anti-RNApII Rpb1 8WG16 (2 µg), anti-Ser5P 3E8 (5 µL), anti-Tfb1 (4 µL) and anti-Kin28 (4 µL). All antibodies have been validated for ChIP. Anti-Ser5P 3E8, anti-Tfb1, and anti-Kin28 antibodies were coupled to Dynabeads coated with Protein G (Thermo Fisher Scientific, 10004D) and anti-RNApII Rpb1 8WG16 antibody was coupled to Dynabeads coated with Pan Mouse IgG antibodies (Thermo Fisher Scientific, 11042). 50 µL of the appropriate Dynabeads pre-coupled with the indicated antibody were added to the chromatin sample and incubated overnight at 4°C. Beads were washed twice with Lysis buffer, twice with Lysis buffer 500 (Lysis buffer + 360 mM NaCl), twice with Wash buffer (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP40, 0.5% Na-deoxycholate, 1 mM EDTA) and once with TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA). Input and immunoprecipitated chromatin were eluted and reverse-crosslinked with 50 µL TE/SDS (TE + 1% SDS) by incubating overnight at 65°C. Samples were treated with RNase A (345 µL TE, 3 µL 10 mg/mL RNAse A (Sigma-Aldrich, R6513), 2 µL 20 mg/mL Glycogen (Roche, 10901393001)) at 37°C for 2 hr and subsequently subjected to Proteinase K (15 µL 10% SDS, 7.5 µL 20 mg/mL Proteinase K (Thermo Fisher Scientific, 25530049)) digestion at 37°C for 2 hr. Samples were extracted twice with phenol/chloroform/isoamyl alcohol (25:24:1), followed by precipitation with 200 mM NaCl and 70% ethanol. Precipitated DNA was resuspended in 50 µL of TE before being used in ChIP-seq libraries.

All ChIP and Input samples were subjected to a 1.7X cleanup using KAPA Pure Beads (Roche, 07983280001) according to the manufacturer's instructions. Samples were eluted with 40 µL of Elution buffer (10 mM Tris pH 8.0). DNA concentration of ChIP and Input samples was determined by qPCR using a standard curve made with a fragment size control sample. 1-10 ng of ChIP and Input DNA were used for ChIP-seq library preparation as follows. The ends of DNA were repaired by incubating in 70 µL of 1X NEBuffer 2 containing 0.6 units of T4 DNA polymerase (NEB, M0203S), 2 units of T4 polynucleotide kinase (NEB, M0201S), 0.09 nM dNTPs

and 0.045 µg/µL of BSA at 12°C for 30 min. Repaired DNA was then subjected to a 1.7X cleanup using KAPA Pure Beads before dA tailing as follows. Beads containing the repaired DNA were resuspended in 50 µL of 1X NEBuffer 2 containing 0.1 mM dATP and 25 units of Klenow Fragment (3'→5' exo-) (NEB, M0212M), and incubated at 37°C for 30 min. After a 1.8X cleanup with KAPA Pure Beads, the A-tailed DNA was ligated to index adapters (Illumina Truseq DNA UD Indexes (20023784)) as follows. The beads were resuspended in 45 µL of 1X Ligase buffer containing 8 nM of adapter and 2.5 units of T4 DNA ligase (ThermoFisher, 15224041) and incubated at room temperature for 60 min. The ligated DNA was then subjected to a 1X cleanup with KAPA Pure Beads, followed by a double size selection (0.52X-1X) leading to fragments in the 200-600 bp range. Libraries were PCR-amplified with 12-13 cycles using KOD Hot Start DNA polymerase (Millipore, 71086-3) and cleaned up using 1X KAPA Pure Beads. Libraries were qualified on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit and quantified by qPCR using NEBNext Library Quant Kit for Illumina (NEB, E7630). Equal molarity of each library was pooled and subjected to sequencing on Illumina NovaSeq 6000 platform at the McGill University and Génome Québec Innovation Centre to generate 100 bp paired-end reads.

ChIP-seq Data analysis. Adapter sequences were removed from paired-end reads for each biological replicate using Trimmomatic, version 0.36 (Bolger et al., 2014) with parameters "ILLUMINACLIP:adapters.fa:2:30:10: TRAILING:3 MINLEN:25". Paired-end reads for each biological replicate were independently aligned to the *S. cerevisiae* (UCSC sacCer3) and *S. pombe* (Downloaded from https://www.pombase.org/ on December 7th 2022) reference genomes using the short read aligner Bowtie 2 (version 2.3.4.3) (Langmead et al., 2009). Only reads mapped in proper pairs and primary alignments were kept in aligned files using Samtools, version 1.9 (Li et al., 2009) with parameters "-F 2048 -F 256 -f 2". Duplicate reads were removed from aligned files using Samtools "fixmate" and "markdup -r". Coverage for each base pair of the *S. cerevisiae* genome was computed using genomeCoverageBed from BEDTools, version 2.27.1 (Quinlan and Hall, 2010) and normalized using *S. pombe* reads as follows. The read density at each position of the *S. cerevisiae* genome was multiplied by a normalization factor N defined as:

$N = 1{,}000{,}000 * P / R * C$

where:

1,000,000 is an arbitrarily chosen number used for convenience;

R is the total number of reads in the IP sample that mapped to the *S. pombe* genome;

C is the total number of reads in the Input sample that mapped to the *S. cerevisiae* genome;

P is the total number of reads in the Input sample that mapped to the *S. pombe* genome.

The deduplicated BAM files of replicates were merged using Samtools "merge".

Correlation between replicates were generated using deepTools, version 3.5.1 (Ramirez et al., 2016) with parameters "multiBigwigSummary bins --binSize 100" and "plotCorrelation --corMethod pearson --whatToPlot heatmap --colorMap PiYG --plotNumbers".

The genome coverage files were converted to the bigWig format using the utilities from UCSC Genome Browser (Casper et al., 2018).

Statistics regarding the reads (including the number of reads, mapped reads, the correlation between replicates, etc.) are shown in **Supplemental Table S3**.

The RNApII-adjusted Ser5P coverage file (used in **Fig. 5E** and **Supplemental Fig. S3E**) was generated as follows, starting from *S. pombe*-normalized coverage files. For each genomic position, the Ser5P occupancy value of a given Tfb3 mutant was multiplied by the RNApII occupancy value in WT and divided by the RNApII occupancy value in the mutant.

Aggregate profiles (metagene analyses) were generated using the Versatile Aggregate Profiler (VAP) (Brunelle et al., 2015; Coulombe et al., 2014). The aggregate profiles were generated by aligning data on the transcription start site (TSS) and the polyadenylation (pA) signal of all verified genes that have a signal of RNApII >4 and are longer than 1 kb (n=95). The data were averaged over 10 bp bins. TSS and pA site coordinates were from Pelechano et al. (Pelechano et al., 2013).

In **Supplemental Fig. S3E**, the values for "promoters" are the average signal over a 200 bp window centered on the TSS whereas the values for "ORFs" are the average of the signal between position +200 and the pA site. The region between +100 and +200 was excluded to avoid signal contamination from promoters for ORFs and vice versa.

All raw and processed ChIP-seq data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE241436.

UV sensitivity assay: Yeast strains were grown in liquid media selective for the *TFB3* plasmids, and overnight cultures were diluted to an OD_{600} = 0.005 and 0.0005 for plating. For each test plate, 100 μL of diluted cells were spread on selective media. Plated cells were then exposed to one of four doses of UV radiation: 0, 30, 60, or 90 J/m^2. UV intensity was calibrated using a UVX Radiometer UVP and dosage calculated by varying time of exposure to a short-wave UV transilluminator. After irradiation, plates were grown at 30°C for 2 days in the dark to avoid photoreactivation repair, and then at room temperature for 5 days. Individual colonies were counted and fraction survivals were calculated and plotted.

REFERENCES

Abril-Garrido, J., Dienemann, C., Grabbe, F., Velychko, T., Lidschreiber, M., Wang, H., and Cramer, P. (2023). Structural basis of transcription reduction by a promoter-proximal +1 nucleosome. Mol Cell *83*, 1798-1809 e1797.

Badjatia, N., Ambrosio, D.L., Lee, J.H., and Gunzl, A. (2013). Trypanosome cdc2-related kinase 9 controls spliced leader RNA cap4 methylation and phosphorylation of RNA polymerase II subunit RPB1. Mol Cell Biol *33*, 1965-1975.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Brunelle, M., Coulombe, C., Poitras, C., Robert, M.A., Markovits, A.N., Robert, F., and Jacques, P.E. (2015). Aggregate and Heatmap Representations of Genome-Wide Localization Data Using VAP, a Versatile Aggregate Profiler. Methods Mol Biol *1334*, 273-298.

Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. Molecular Cell *36*, 541-546.

Casper, J., Zweig, A.S., Villarreal, C., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Karolchik, D*., et al.* (2018). The UCSC Genome Browser database: 2018 update. Nucleic Acids Res *46*, D762-D769.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science (New York, NY) *318*, 1780-1782.

Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H., Li, Y., Yu, Z., Liu, W*., et al.* (2021). Structural insights into preinitiation complex assembly on core promoters. Science *372*.

Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes and Development *15*, 3319-3329.

Chun, Y., Joo, Y.J., Suh, H., Batot, G., Hill, C.P., Formosa, T., and Buratowski, S. (2019). Selective Kinase Inhibition Shows That Bur1 (Cdk9) Phosphorylates the Rpb1 Linker In Vivo. Molecular and Cellular Biology *39*, e00602-00618.

Cismowski, M.J., Laff, G.M., Solomon, M.J., and Reed, S.I. (1995). KIN28 encodes a C-terminal domain kinase that controls mRNA transcription in Saccharomyces cerevisiae but lacks cyclin-dependent kinase-activating kinase (CAK) activity. Mol Cell Biol *15*, 2983-2992.

Corden, J.L. (2013). RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. Chemical reviews *113*, 8423-8455.

Coulombe, C., Poitras, C., Nordell-Markovits, A., Brunelle, M., Lavoie, M.A., Robert, F., and Jacques, P.E. (2014). VAP: a versatile aggregate profiler for efficient genome-wide data representation and discovery. Nucleic Acids Res *42*, W485-493.

Das, A., and Bellofatto, V. (2009). The non-canonical CTD of RNAP-II is essential for productive RNA synthesis in Trypanosoma brucei. PLoS One *4*, e6959.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chemical reviews *113*, 8456-8490.

Espinoza, F.H., Farrell, A., Erdjument-Bromage, H., Tempst, P., and Morgan, D.O. (1996). A cyclin-dependent kinase-activating kinase (CAK) in budding yeast unrelated to vertebrate CAK. Science *273*, 1714-1717.

Feaver, W.J., Huang, W., Gileadi, O., Myers, L., Gustafsson, C.M., Kornberg, R.D., and Friedberg, E.C. (2000). Subunit interactions in yeast transcription/repair factor TFIIH. Requirement for Tfb3 subunit in nucleotide excision repair. Journal of Biological Chemistry *275*, 5941-5946.
Fisher, R.P., and Morgan, D.O. (1994). A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. Cell *78*, 713-724.
Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Molecular and Cellular Biology *29*, 5455-5464.
Greber, B.J., Perez-Bertoldi, J.M., Lim, K., Iavarone, A.T., Toso, D.B., and Nogales, E. (2020). The cryoelectron microscopy structure of the human CDK-activating kinase. Proc Natl Acad Sci U S A *117*, 22849-22857.
Grünberg, S., Warfield, L., and Hahn, S. (2012). Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. Nature structural & molecular biology *19*, 788-796.
Harlow, E., and Lane, D.P. (1988). Antibodies: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Inc.).
He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., and Nogales, E. (2016). Near-atomic resolution visualization of human transcription promoter opening. Nature *533*, 359-365.
Helenius, K., Yang, Y., Tselykh, T.V., Pessa, H.K.J., Frilander, M.J., and Mäkelä, T.P. (2011). Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. Nucleic Acids Research *39*, 5025-5035.
Jeronimo, C., Poitras, C., and Robert, F. (2019). Histone Recycling by FACT and Spt6 during Transcription Prevents the Scrambling of Histone Modifications. Cell reports *28*, 1206-1218.e1208.
Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nature structural & molecular biology *21*, 449-455.
Kaldis, P., Sutton, A., and Solomon, M.J. (1996). The Cdk-activating kinase (CAK) from budding yeast. Cell *86*, 553-564.
Keogh, M.C., Cho, E.-J., Podolny, V., and Buratowski, S. (2002). Kin28 is found within TFIIH and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. Molecular and Cellular Biology *22*, 1288-1297.
Kim, J., Li, C.L., Chen, X., Cui, Y., Golebiowski, F.M., Wang, H., Hanaoka, F., Sugasawa, K., and Yang, W. (2023). Lesion recognition by XPC, TFIIH and XPA in DNA excision repair. Nature *617*, 170-175.
Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes and Development *14*, 2452-2460.
Kuldell, N.H., and Buratowski, S. (1997). Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct functions. Molecular and Cellular Biology *17*, 5288-5298.
Kushnirov, V.V. (2000). Rapid and reliable protein extraction from yeast. Yeast *16*, 857-860.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Lee, J.H., Cai, G., Panigrahi, A.K., Dunham-Ems, S., Nguyen, T.N., Radolf, J.D., Asturias, F.J., and Gunzl, A. (2010). A TFIIH-Associated Mediator Head Is a Basal Factor of Small Nuclear Spliced Leader RNA Gene Transcription in Early-Diverged Trypanosomes. Molecular and Cellular Biology *30*, 5502-5513.
Lee, J.H., Jung, H.S., and Günzl, A. (2009). Transcriptionally active TFIIH of the early-diverged eukaryote Trypanosoma brucei harbors two novel core subunits but not a cyclin-activating kinase complex. Nucleic Acids Research *37*, 3811-3820.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) *25*, 2078-2079.
Matsui, P., DePaulo, J., and Buratowski, S. (1995). An interaction between the Tfb1 and Ssl1 subunits of yeast TFIIH correlates with DNA repair activity. Nucleic Acids Research *23*, 767-772.
Murakami, K., Gibbons, B.J., Davis, R.E., Nagai, S., Liu, X., Robinson, P.J.J., Wu, T., Kaplan, C.D., and Kornberg, R.D. (2012). Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. Proceedings of the National Academy of Sciences of the United States of America *109*, 4816-4821.
Murakami, K., Mattei, P.J., Davis, R.E., Jin, H., Kaplan, C.D., and Kornberg, R.D. (2015). Uncoupling Promoter Opening from Start-Site Scanning. Mol Cell *59*, 133-138.
Peissert, S., Schlosser, A., Kendel, R., Kuper, J., and Kisker, C. (2020). Structural basis for CDK7 activation by MAT1 and Cyclin H. Proc Natl Acad Sci U S A *117*, 26739-26748.
Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127-131.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.
Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.
Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.-J., Burlingame, A.L., and Kornberg, R.D. (2016). Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. Cell *166*, 1411-1414.e1416.
Sano, M., Izumi, Y., Helenius, K., Asakura, M., Rossi, D.J., Xie, M., Taffet, G., Hu, L., Pautler, R.G., Wilson, C.R*., et al.* (2007). Menage-a-trois 1 is critical for the transcriptional function of PPARgamma coactivator 1. Cell Metab *5*, 129-142.
Schier, A.C., and Taatjes, D.J. (2021). Everything at once: cryo-EM yields remarkable insights into human RNA polymerase II transcription. Nat Struct Mol Biol *28*, 540-543.
Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., and Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIIH and Mediator. Nature *551*, 204-209.
Schilbach, S., Wang, H., Dienemann, C., and Cramer, P. (2023). Yeast PIC-Mediator structure with RNA polymerase II C-terminal domain. Proc Natl Acad Sci U S A *120*, e2220542120.
Søgaard, T.M.M., and Svejstrup, J.Q. (2007). Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. Journal of Biological Chemistry *282*, 14113-14120.

Suh, H., Ficarro, S.B., Kang, U.-B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Molecular Cell *61*, 297-304.
Sung, P., Higgins, D., Prakash, L., and Prakash, S. (1988). Mutation of lysine-48 to arginine in the yeast RAD3 protein abolishes its ATPase and DNA helicase activities but not the ability to bind ATP. EMBO J *7*, 3263-3269.
Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Critical reviews in biochemistry and molecular biology *41*, 105-178.
Tomko, E.J., Luyties, O., Rimel, J.K., Tsai, C.L., Fuss, J.O., Fishburn, J., Hahn, S., Tsutakawa, S.E., Taatjes, D.J., and Galburt, E.A. (2021). The Role of XPB/Ssl2 dsDNA Translocase Processivity in Transcription Start-site Scanning. J Mol Biol *433*, 166813.
van Eeuwen, T., Li, T., Kim, H.J., Gorbea Colon, J.J., Parker, M.I., Dunbrack, R.L., Garcia, B.A., Tsai, K.L., and Murakami, K. (2021a). Structure of TFIIK for phosphorylation of CTD of RNA polymerase II. Sci Adv *7*.
van Eeuwen, T., Shim, Y., Kim, H.J., Zhao, T., Basu, S., Garcia, B.A., Kaplan, C.D., Min, J.H., and Murakami, K. (2021b). Cryo-EM structure of TFIIH/Rad4-Rad23-Rad33 in damaged DNA opening in nucleotide excision repair. Nat Commun *12*, 3338.
Vannini, A., and Cramer, P. (2012). Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. Molecular Cell *45*, 439-446.
Warfield, L., Luo, J., Ranish, J., and Hahn, S. (2016). Function of Conserved Topological Regions within the Saccharomyces cerevisiae Basal Transcription Factor TFIIH. Molecular and Cellular Biology *36*, 2464-2475.
Watanabe, T., Hayashi, K., Tanaka, A., Furumoto, T., Hanaoka, F., and Ohkuma, Y. (2003). The carboxy terminus of the small subunit of TFIIE regulates the transition from transcription initiation to elongation by RNA polymerase II. Molecular and Cellular Biology *23*, 2914-2926.
Winkler, G.S., Araujo, S.J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J.M., Hoeijmakers, J.H., Wood, R.D., Timmers, H.T., and Weeda, G. (2000). TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. J Biol Chem *275*, 4258-4266.
Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. Molecular Cell *54*, 601-612.
Yang, C., Fujiwara, R., Kim, H.J., Basnet, P., Zhu, Y., Gorbea Colon, J.J., Steimle, S., Garcia, B.A., Kaplan, C.D., and Murakami, K. (2022). Structural visualization of de novo transcription initiation by Saccharomyces cerevisiae RNA polymerase II. Mol Cell *82*, 660-676 e669.
Yu, J., Yan, C., Dodd, T., Tsai, C.L., Tainer, J.A., Tsutakawa, S.E., and Ivanov, I. (2023). Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases. Nat Commun *14*, 2758.
Yurko, N., Liu, X., Yamazaki, T., Hoque, M., Tian, B., and Manley, J.L. (2017). MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD. Molecular Cell *68*, 913-925.e913.

Figure Legends

Fig. 1. Tfb3 N-term-Linker fusions to other kinase modules can replace Tfb3, but not Kin28. **(A)** Schematic of TFIIH structure within the PIC. The domain structure of Tfb3 (green) is shown, connecting the TFIIH Kinase (cyan) and Core (blue) Modules. **(B)** Schematic of Tfb3 fusions to other kinase modules. (**C)** The *kin28Δ* strain YSB744 was transformed with pRS425 (Vector), pRS415-*KIN28*, pRS425-*TFB3-BUR2*, pRS425-*TFB3-CTK3*, or pRS425-*TFB3-MPK1*. Transformants were tested for the ability to replace a *KIN28/URA3* construct by plasmid shuffling. Cells were streaked on -LEU-TRP (center) or -LEU-TRP+5-FOA (right) plates for 3 days at 30℃. **(D)** pRS315 (Vector), pRS425-*TFB3*, pRS315/TFB3Δ2 (aa 1-275) (Feaver et al., 2000), pRS425-*TFB3-BUR2*, pRS425-*TFB3-CTK3*, and pRS425-*TFB3-MPK1* were transformed into *tfb3Δ* strain SHY907 (Warfield et al., 2016), replacing the plasmid-expressed wild-type gene by plasmid shuffling. Streaks shown were grown on –LEU for 4 days (center) or on -LEU+5-FOA plates (right) for 9 days.

Fig. 1

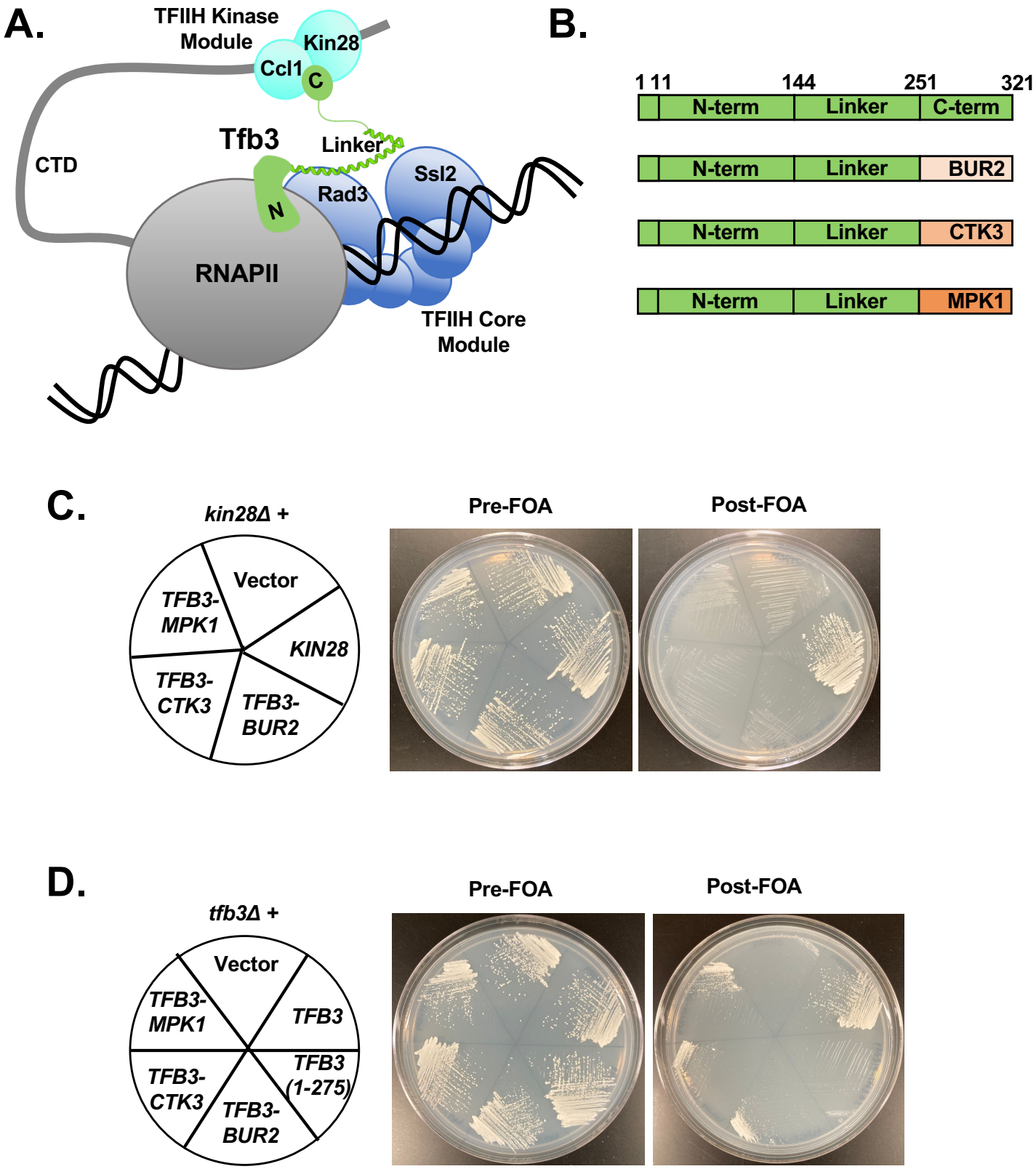
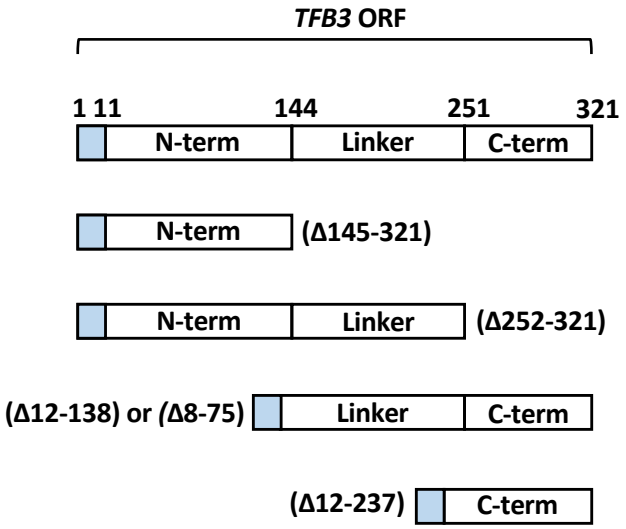

A.
TFIIH Kinase Module
Kin28
Ccl1
C
CTD
Tfb3
Linker
N
Rad3
Ssl2
RNAPII
TFIIH Core Module
B.
1 11
144
251
321
N-term
Linker
C-term
N-term
Linker
BUR2
N-term
Linker
CTK3
N-term
Linker
MPK1
C.
kin28Δ +
Vector
TFB3-MPK1
KIN28
TFB3-CTK3
TFB3-BUR2
Pre-FOA
Post-FOA
D.
tfb3Δ +
Vector
TFB3-MPK1
TFB3
TFB3-CTK3
TFB3 (1-275)
TFB3-BUR2
Pre-FOA
Post-FOA

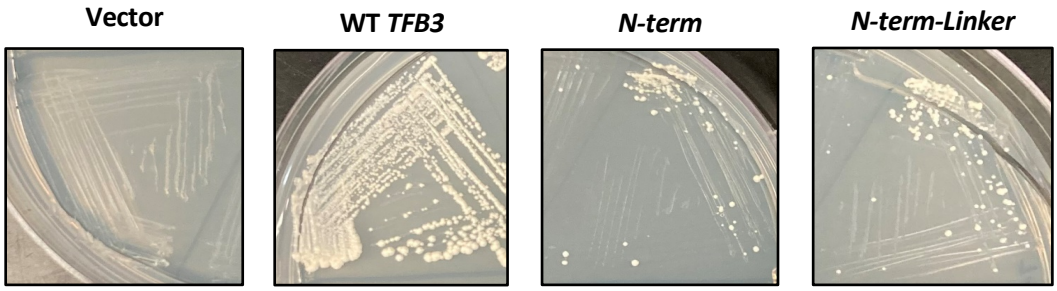
Fig. 2. Tfb3 can be split into two functional parts.
(A) Schematic of Tfb3 deletion constructs. **(B)** Plasmid shuffling was used to test pRS424 (Vector), pRS424-*TFB3* (WT *TFB3*), pRS424-*TFB3(*Δ145-321) (N-term), and pRS424-*TFB3* (Δ252-321) (N-term-Linker) for the ability to support growth of *tfb3Δ* shuffling strain SHY907/YF2456 (Warfield et al., 2016). Streaks shown were grown on -TRP+5-FOA plates for 5 days at 30°C **(C)** Spot assay for growth. Tfb3 N-term (YSB3704) and N-term-Linker (YSB3722) strains were transformed with pRS425 (Vector), pSH1542 (WT *TFB3*), pRS315-TFB3(1-11, 238-Stop)-Flag1-TAP (C-term low copy), pRS425-TFB3(1-11, 238-Stop) (C-term high copy), pRS425-TFB3(1-11, 139-Stop), or pLH366 (Linker-C-term high copy). Ten-fold serial dilutions were spotted onto –LEU-TRP plates for 3 days. For comparison to normal growth, an isogenic strain carrying WT *TFB3* (YSB3788) was spotted in parallel (top strip).

Fig. 2

A.


TFB3 ORF
1 11
144
251
321
N-term
Linker
C-term
N-term
(Δ145-321)
N-term
Linker
(Δ252-321)
(Δ12-138) or (Δ8-75)
Linker
C-term
(Δ12-237)
C-term

B.


Vector
WT TFB3
N-term
N-term-Linker

C.

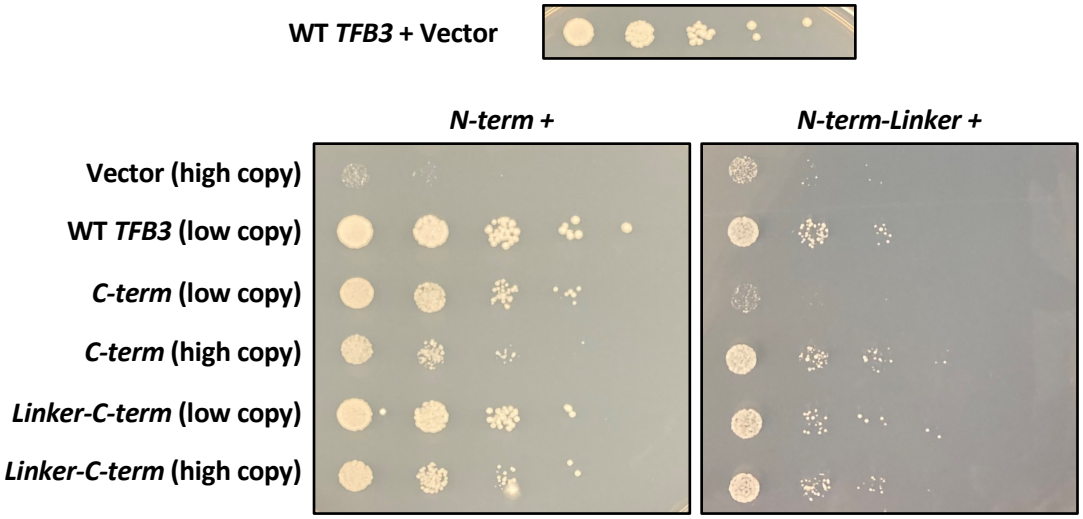

WT TFB3 + Vector
N-term +
N-term-Linker +
Vector (high copy)
WT TFB3 (low copy)
C-term (low copy)
C-term (high copy)
Linker-C-term (low copy)
Linker-C-term (high copy)

Fig. 3. The TFIIH Kinase Module associates with the untethered Tfb3 C-terminal domain. (A) Whole cell extracts were made from YSB3788 (*TFB3* untagged), YSB3723 (*TFB3*-TAP + N-term), YSB3732 (N-term-Linker + C-term-TAP), and YSB3728 (N-term + Linker-C-term-TAP). Both Input extracts (lanes 1-4) and IgG-Agarose precipitated fractions (lanes 5-8) were resolved by SDS-PAGE and blotting for Tfb3. **(B)** To probe for other Kinase Module subunits, IgG precipitates were probed with anti-Ccl1 and anti-Kin28 antibodies. In all panels, asterisks mark TAP-tagged Tfb3 derivatives (which also react with secondary antibody due to the protein A component).

Fig. 3

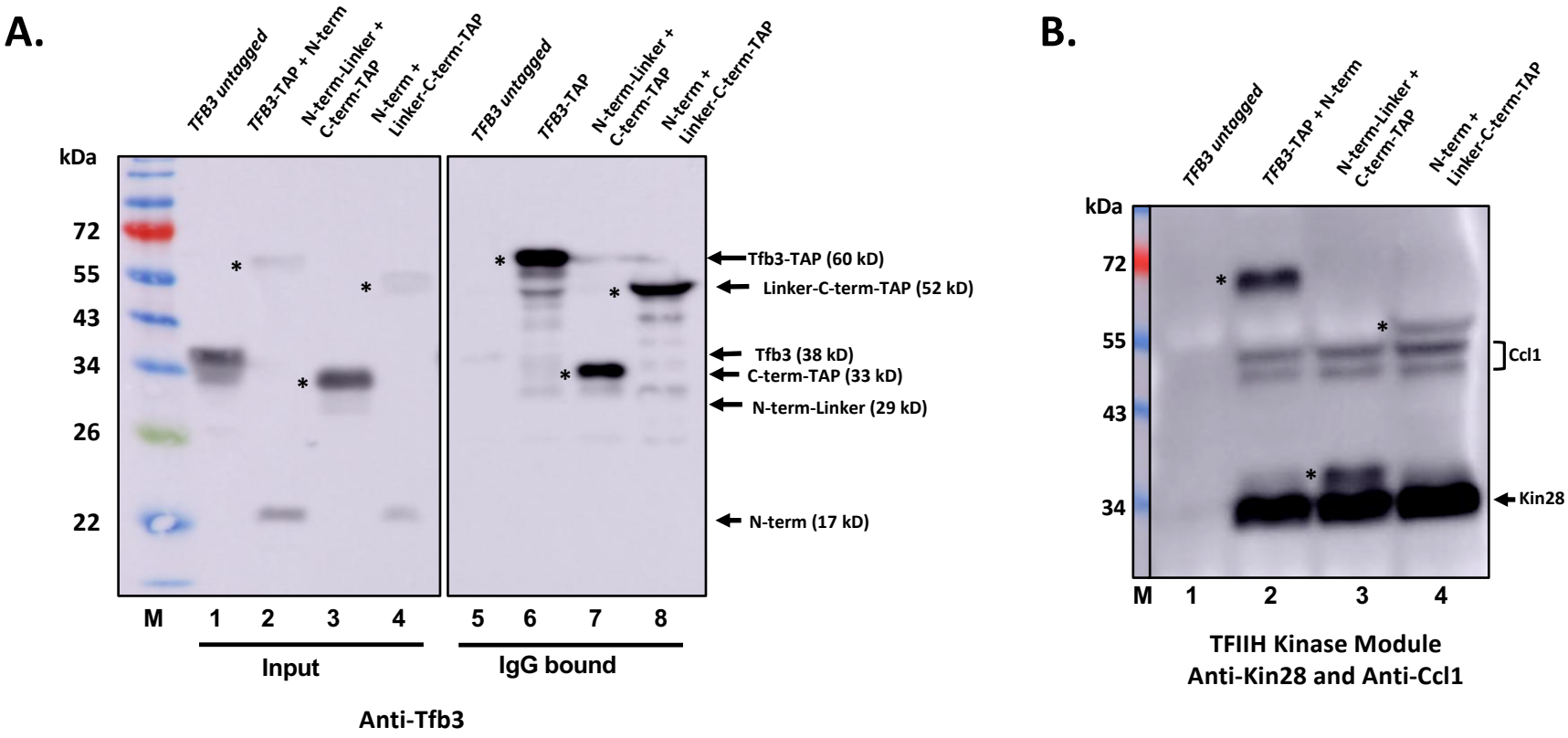
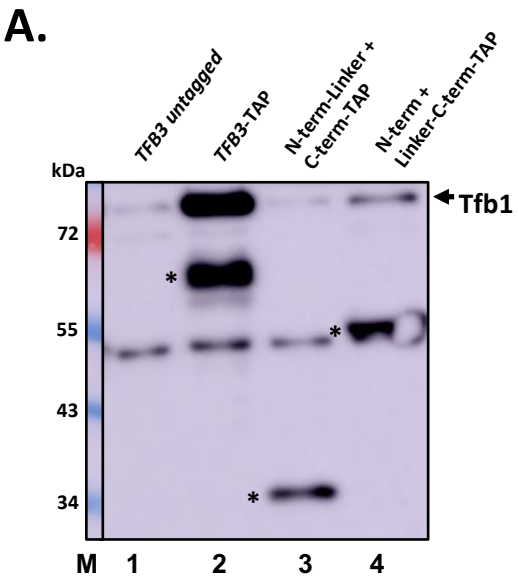
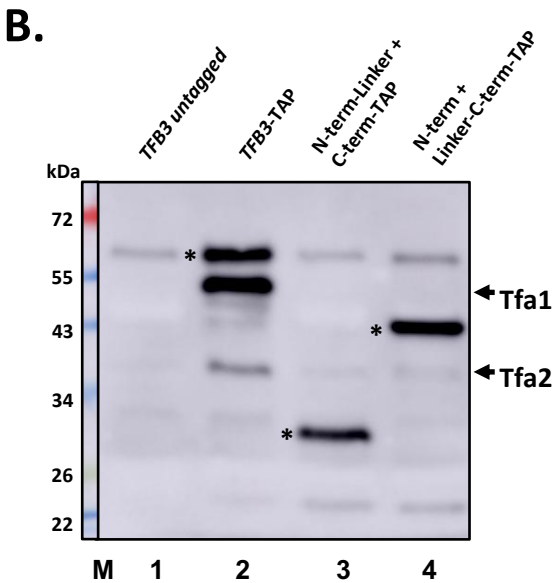

A.
TFB3 untagged
TFB3-TAP + N-term
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
TFB3 untagged
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
kDa
72
55
43
34
26
22
Tfb3-TAP (60 kD)
Linker-C-term-TAP (52 kD)
Tfb3 (38 kD)
C-term-TAP (33 kD)
N-term-Linker (29 kD)
N-term (17 kD)
M 1 2 3 4 5 6 7 8
Input
IgG bound
Anti-Tfb3
B.
TFB3 untagged
TFB3-TAP + N-term
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
kDa
72
55
43
34
Ccl1
Kin28
M 1 2 3 4
TFIIH Kinase Module
Anti-Kin28 and Anti-Ccl1

Fig. 4. Splitting Tfb3 uncouples the TFIIH Kinase Module from the Core Module and TFIIE. The same IgG-Agarose precipitates analyzed in **Fig. 3** were also probed with the following antibodies against: **(A)** TFIIH Core Module subunit Tfb1 (note background binding of Tfb1 to beads as seen in lane 1, and non-specific band ~50 kDa), and **(B)** TFIIE subunits Tfa1 and Tfa2 (note non-specific band ~60 kDa). In all panels, asterisks mark TAP-tagged Tfb3 derivatives reacting with secondary antibody due to the protein A component.

Fig. 4



TFIIH Core Module
Anti-Tfb1



TFIIE
Anti-Tfa1 and Anti-Tfa2

Figure 5. **Loss of proper Kinase Module and Ser5P promoter localization in split Tfb3 cells**. Crosslinked chromatin was prepared from strains expressing full length Tfb3 (YSB3788), the separated N-term-Linker and C-term derivatives (YSB3731), the N-term and Linker-C-term derivatives (YSB3727), or N-term and C-term with no Linker (YSB3725). Samples were immunoprecipitated and processed for sequencing of crosslinked DNA as described in Methods. Two biological replicates for each antibody were used to generate metagene profiles of **(A)** TFIIH Core subunit Tfb1, **(B)** TFIIH Kinase subunit Kin28, **(C)** Rpb1 CTD Ser5P, **(D)** RNApII subunit Rpb1, and **(E)** Ser5P normalized to total Rpb1. The metagene analysis used only those nuclear genes with an average RNApII occupancy greater than 4 reads per million (rpm) and longer than 1 kb (n=95). Individual replicates are shown in **Supplemental Fig. S3**.

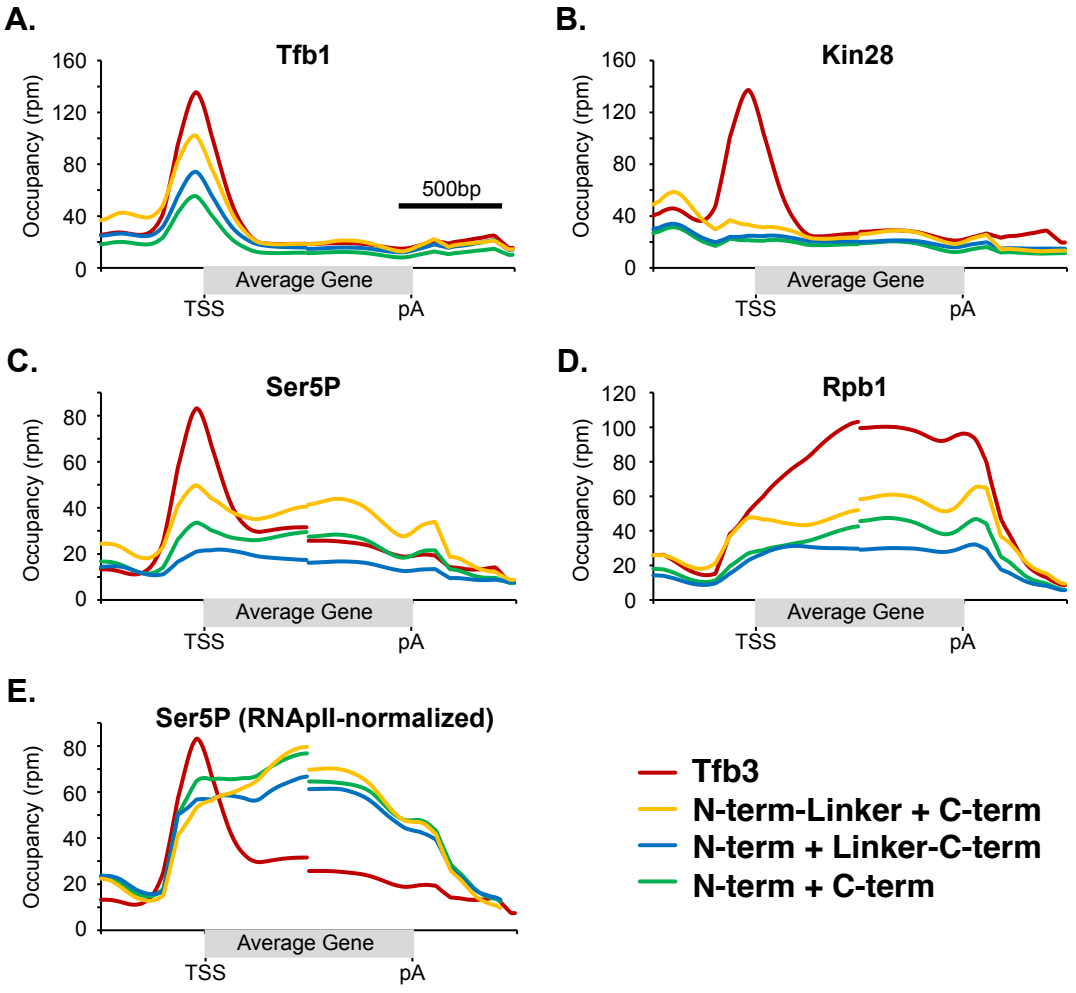
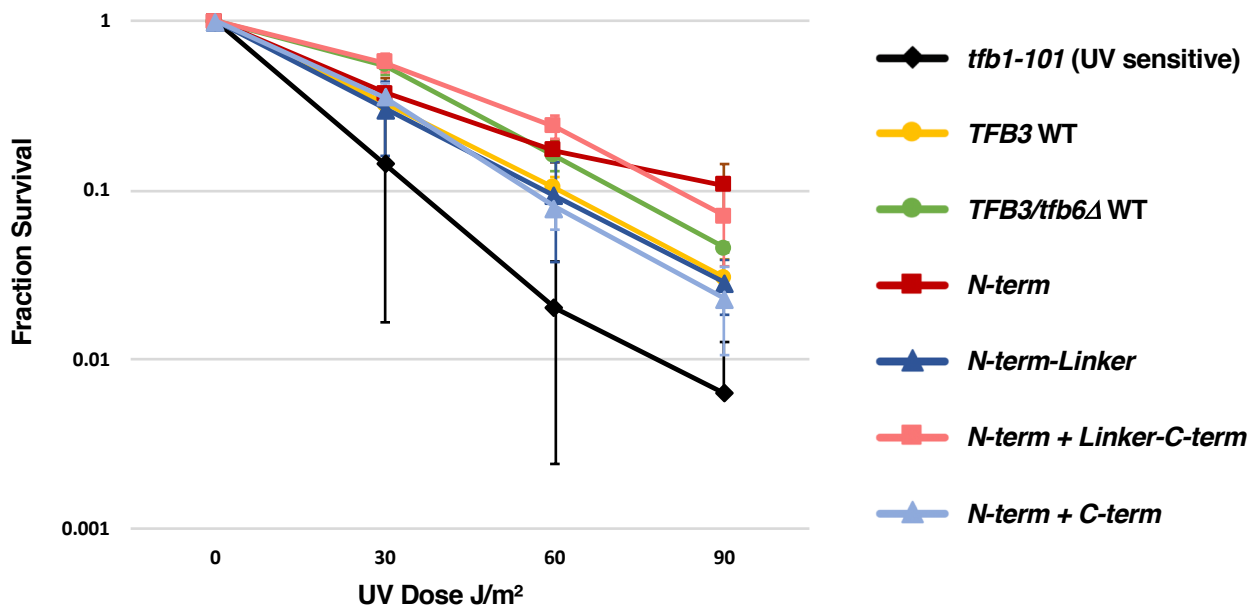

A.
Tfb1
Occupancy (rpm)
160
120
80
40
0
500bp
Average Gene
TSS
pA
B.
Kin28
Occupancy (rpm)
160
120
80
40
0
Average Gene
TSS
pA
C.
Ser5P
Occupancy (rpm)
80
60
40
20
0
Average Gene
TSS
pA
D.
Rpb1
Occupancy (rpm)
120
100
80
60
40
20
0
Average Gene
TSS
pA
E.
Ser5P (RNApII-normalized)
Occupancy (rpm)
80
60
40
20
0
Average Gene
TSS
pA
Tfb3
N-term-Linker + C-term
N-term + Linker-C-term
N-term + C-term

Fig 5

Figure 6. **Split Tfb3 strains are not hyper-sensitive to UV light**. Yeast strains carrying the indicated Tfb3 configurations were plated at various densities and exposed to the indicated dosages of UV light. Surviving colonies were counted after several days. The graph shows the averaged values from multiple replicates, and error bars represent standard deviation. Note that there are two strain backgrounds. The *tfb1-101* strain (YSB250, black) is a positive control for an NER-defective strain sensitive to UV light, and its isogenic WT control is YSB207 (yellow). The other five isogenic strains are in a *tfb6Δ* background, lacking the Tfb6 protein that competes with the TFIIH for Ssl2 binding (Murakami et al., 2012): WT/tfb6Δ (YSB3788, green), N-term (YSB3724, red), N-term-Linker (YSB3729, blue), N-term + Linker-C-term (YSB3727, orange), and N-term-Linker + C-term (YSB3731, slate).

Fig. 6

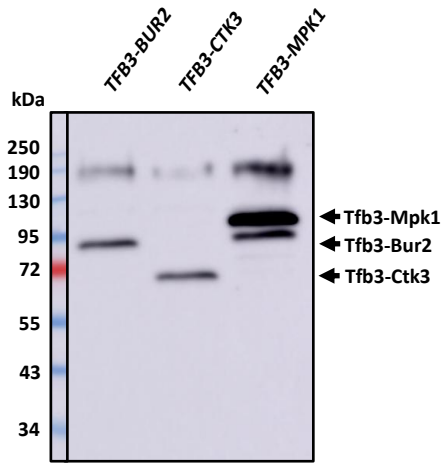

1
0.1
0.01
0.001
Fraction Survival
0
30
60
90
UV Dose J/m²
tfb1-101 (UV sensitive)
TFB3 WT
TFB3/tfb6Δ WT
N-term
N-term-Linker
N-term + Linker-C-term
N-term + C-term

Supplemental Information for Giordano et al.

Four supplemental figures and three supplemental tables.

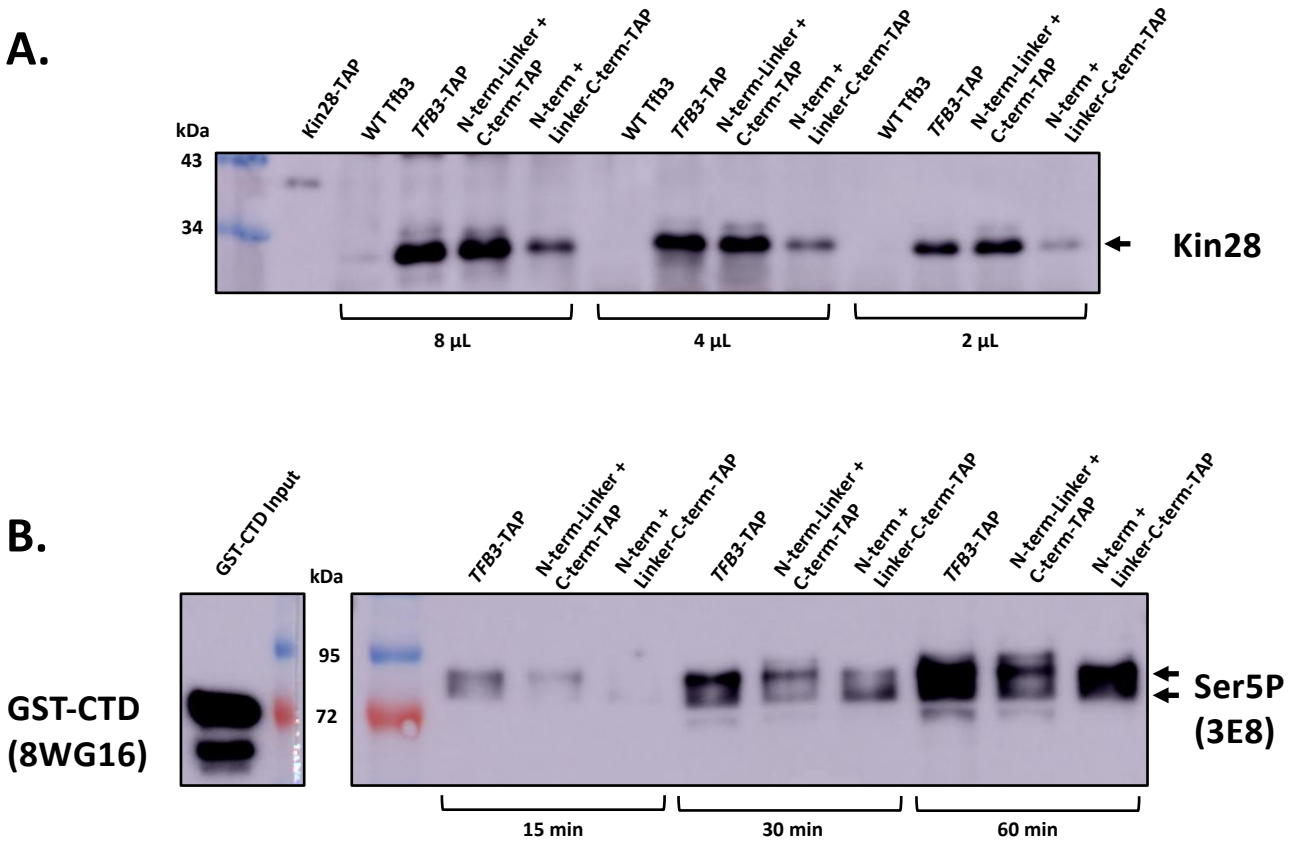
Supplemental Figure S1 (related to Figure 1). Immunoblot of Tfb3 fusions from **Fig. 1C**. Whole extracts were made from the post-FOA strains constructed by shuffling pRS425-*TFB3-BUR2*, pRS425-*TFB3-CTK3*, or pRS425-*TFB3-MPK1* into *tfb3Δ* strain SHY907/YF2456 (Warfield et al., 2016). Extracts were resolved by SDS-PAGE, blotted to a membrane, and probed with anti-Tfb3 antiserum. Left-most lane shows markers imaged with visible light, and the rest of the blot shows chemiluminescence detection.

Sup. Fig. 1

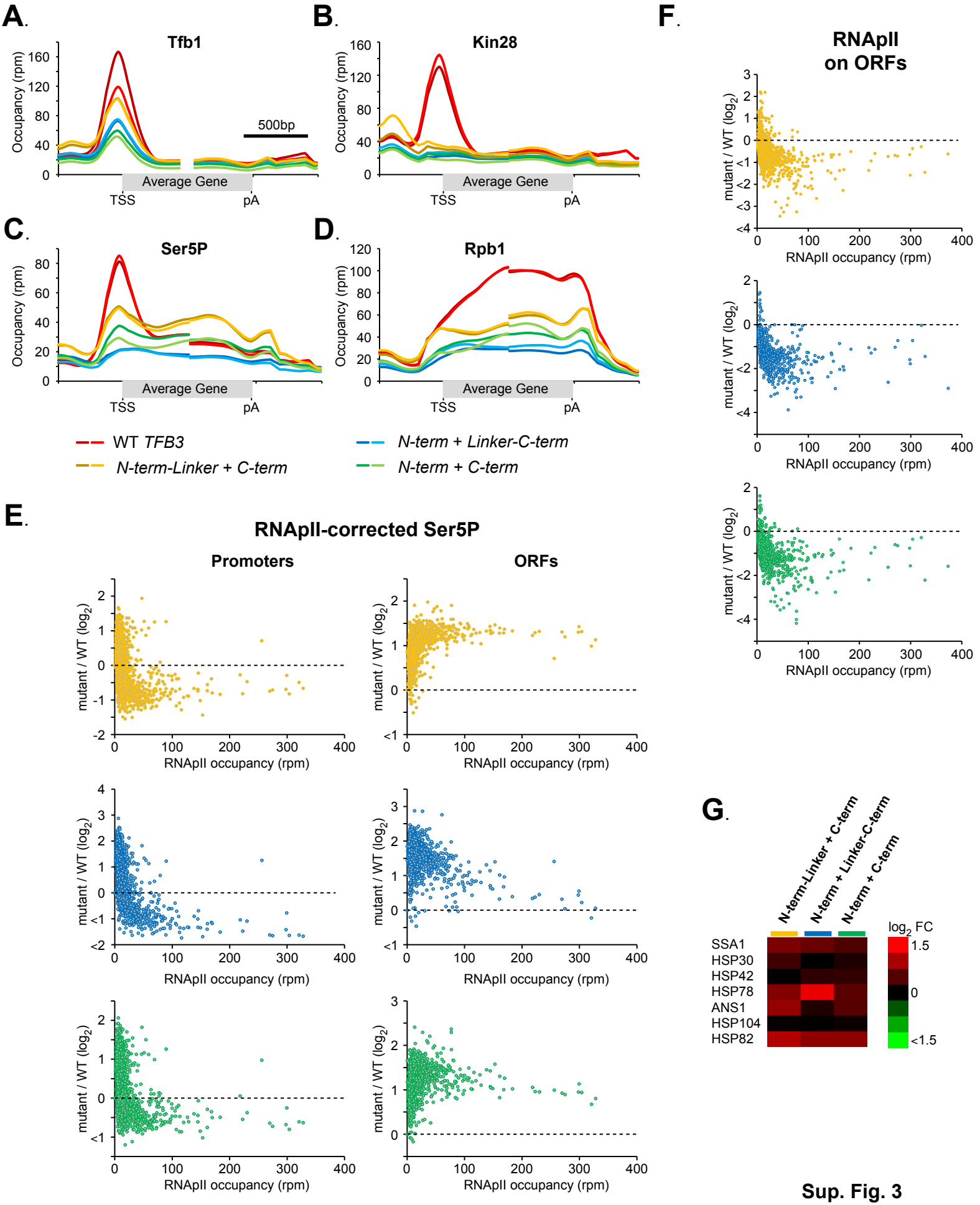

TFB3-BUR2
TFB3-CTK3
TFB3-MPK1
kDa
250
190
130
95
72
55
43
34
Tfb3-Mpk1
Tfb3-Bur2
Tfb3-Ctk3

Supplemental Figure S2 (related to Figure 3). Tethered and untethered Kinase Module have similar CTD kinase specific activity in vitro.
(A) Kinase Module was isolated from YSB3707 (TFB3-TAP), YSB3732 (N-term-Linker + C-term-TAP), and YSB3728 (N-term + Linker-C-term-TAP) using the TAP tag. A parallel isolation from the untagged strain YSB3788 (WT TFB3) showed that binding was dependent on the TAP tag. The second lane shows a Kin28-TAP precipitate (not done in parallel) as size marker. Relative levels of the Kin28 subunit on beads were determined by immunoblotting 8uL, 4uL, and 2uL of beads.
(B) In vitro CTD phosphorylation assay. After roughly normalizing bead volumes for amount of bound Kin28, kinase reactions were performed with GST-CTD. Samples were taken at the indicated time points and tested for CTD Ser5 phosphorylation using monoclonal antibody 3E8 (Chapman et al., 2007). Note the phosphorylated CTD as a series of bands, with slower mobility believed to represent denser phosphorylation. To mark the position of unphosphorylated GST-CTD, a negative control reaction lane from the blot was probed with 8WG16 (left panel).

Sup. Fig. 2


A.
kDa
43
34
Kin28-TAP
WT Tfb3
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
WT Tfb3
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
WT Tfb3
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
Kin28
8 µL
4 µL
2 µL
B.
GST-CTD Input
kDa
95
72
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
TFB3-TAP
N-term-Linker + C-term-TAP
N-term + Linker-C-term-TAP
GST-CTD (8WG16)
Ser5P (3E8)
15 min
30 min
60 min

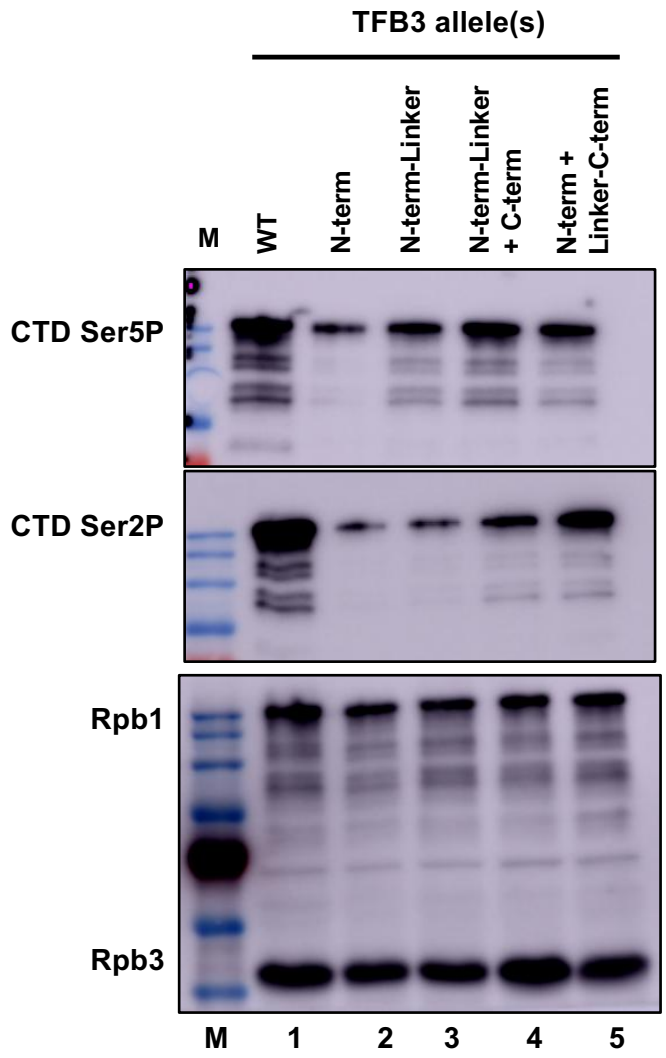
Supplementary Figure S3 (related to Figure 5). Additional ChIP-seq analysis. (**A-D**) These panels are the same metagene analysis as Figure 5A-D, but showing the individual replicates as separate lines. The Tfb3 configurations are indicated in the key at the bottom: wild-type Tfb3 in reds, N-term-Linker + C-term in yellows, N-term + Linker-C-term in blues, and N-term + C-term in greens. (**E**) For each split Tfb3 mutant (color coded as in A-D), a plot of the mutant/WT $\log_2$ RNApII-adjusted Ser5P ratio versus WT RNApII occupancy, for all verified genes (n=5587), over promoters (left) and ORFs (right) (see Methods). In panels E and F, the x-axis is cut at 400 rpm for clarity, excluding only three genes with higher occupancy. (**F**) Plots of the mutant/WT $\log_2$ ratio versus WT level for RNApII occupancy, over ORFs for all verified genes (n=5573). Yellow: N-term-Linker + C-term, Blue: N-term + Linker-C-term, and Green: N-term + C-term. (**G**) Heatmap of the mutant/WT $\log_2$ RNApII ratio for the seven genes that are systematically induced (mutant/WT $\log_2$ RNApII ratio>0) in all three split mutants. All but *ANS1* are known to be involved in stress-response.


A.
Tfb1
Occupancy (rpm)
500bp
Average Gene
TSS
pA
B.
Kin28
C.
Ser5P
D.
Rpb1
WT TFB3
N-term-Linker + C-term
N-term + Linker-C-term
N-term + C-term
E.
RNApII-corrected Ser5P
Promoters
ORFs
mutant / WT (log2)
RNApII occupancy (rpm)
F.
RNApII on ORFs
G.
N-term-Linker + C-term
N-term + Linker-C-term
N-term + C-term
log2 FC
SSA1
HSP30
HSP42
HSP78
ANS1
HSP104
HSP82

Sup. Fig. 3

Supplemental Figure S4 (related to Figure 5). Immunoblot for CTD phosphorylation levels in strains with split Tfb3 or completely lacking the C-term. Whole cell extracts from YSB3712 (WT), YSB3710 (N-term), YSB 3711 (N-term-Linker), YSB3731 (N-term-Linker + C-term), and YSB3727 (N-term + Linker-C-term) were resolved by SDS-PAGE and immunoblotted. The top blot shows CTD Ser5P (monoclonal antibody 3E8), the middle blot shows Ser2P (monoclonal antibody 3E10), and the bottom blot shows both Rpb1 (monoclonal antibody 8WG16) and Rpb3 (monoclonal antibody 1Y26).

Sup. Fig. 4


TFB3 allele(s)
M
WT
N-term
N-term-Linker
N-term-Linker + C-term
N-term + Linker-C-term
CTD Ser5P
CTD Ser2P
Rpb1
Rpb3
M
1
2
3
4
5

Supplemental Table S1. Plasmids used in this study.

Supplemental Table S2. Yeast strains used in this study.

Supplemental Table S3. ChIP-Seq parameters.

Supplemental Table 1. Plasmids used in this study.

Plasmid	Features	Construction or source reference
pSBEThis7-TFB3	T7 promoter driving N-terminal his-tagged TFB3 (1-285) [yeast TFIIH subunit], ArgU, KanR	962 bp NcoI-Bgl II fragment bearing TFB3 open reading frame from pBS-TFB3 ORF (SB835) was cloned into NcoI + BamHI sites of pSBEThis7. Created by Michael Keogh.
pBS-TFB3 A	TFB3 gene with 5' flanking sequence [yeast TFIIH subunit], f1+ ori, AmpR	TFB3 gene and promoter 2.059 kb fragment PCR'd from yeast genomic DNA using primers TFB3-A (O #411) and TFB3-C (O #413) and Pfu polymerase, and blunt ligated into SrfI site of pCR-Script SK+. Created by M. Keogh.
pRS415-KIN28	KIN28, LEU2, CEN/ARS, f1+ ori, AmpR	pRS426-KIN28 was digested with BamHI + HindIII. The ~1.3 kb fragment was cloned into BamHI + HindIII digested pRS415.
pBS-TFB3-BUR2	TFB3 promoter driving TFB3 (1-251)-BUR2 (2-395) fusion with TFB3 5' and 3' flanking sequence, f1+ ori, AmpR	~1.2 kb BUR2 fragment was amplified from pRS316-BUR2 (SB1209) using oligos Tfb3-Bur2 F (O #4287) and Bur2 stop Tfb3 3UTR R (O #4288). ~4.8 kb pBS-Tfb3 backbone was amplified using oligos Tfb3-3UTR F (O #4283) and Tfb3-1753 R (O #4284) on plasmid pBS-TFB3 A (SB836a). Fragments assembled through Gibson Assembly.
pBS-TFB3-CTK3	TFB3 promoter driving TFB3 (1-251)-CTK3 (2-296) fusion with TFB3 5' and 3' flanking sequence, f1+ ori, AmpR	~900 bp CTK3 fragment was amplified from pJYC4501 (F716) using oligos Tfb3-Ctk3 F (O #4285) and Ctk3 stop Tfb3 3UTR R (O #4286). ~4.8 kb pBS-Tfb3 backbone was amplified using oligos Tfb3-3UTR F (O #4283) and Tfb3-1753 R (O #4284) on plasmid pBS-TFB3 A (SB836a). Fragments assembled through Gibson Assembly.
pRS425-TFB3 (1-144)	TFB3 (1-144), LEU2, 2 μ ori, f1+ ori, AmpR	Oligos Tfb3- 3UTR F (O #4283) and TFB3 Ile144-Stop (O #4335) were used for inverse PCR on pRS425-TFB3 (SB1968). The resulting ~8.2 kb fragment was ligated by intramolecular Gibson isothermal assembly.
pRS425-TFB3 (1-251)	TFB3 (1-251), LEU2, 2 μ ori, f1+ ori, AmpR	Oligos Tfb3- 3UTR F (O #4283) and TFB3 Leu251-Stop (O #4336) were used for inverse PCR on pRS425-TFB3 (SB1968). The resulting ~8.5 kb fragment was ligated by intramolecular Gibson isothermal assembly.
pRS425-TFB3	TFB3, LEU2, 2 μ ori, f1+ ori, AmpR	~1.9 kb BamHI-Sal I fragment from pBS-TFB3 A (SB836a) cloned into the BamHI - Sal I sites of pRS425 (YV27).
pRS425-TFB3-MPK1	TFB3 promoter driving TFB3 (1-251)-MPK1 (2-484) fusion with TFB3 5' and 3' flanking sequence, LEU2, 2 μ ori, f1+ ori, AmpR	~8.5 kb pRS425-Tfb3 backbone fragment was amplified from pRS425-TFB3 (SB1968) using oligos Tfb3- 3UTR F (O #4283) and Tfb3- 1753 R (O #4284). ~1.5 kb Mpk1 fragment was amplified from genomic DNA with oligos Tfb3-Mpk1 for (O #4323) and Tfb3-Mpk1 rev (O #4324). Fragments were assembled through Gibson isothermal assembly.
pRS425-TFB3-BUR2	TFB3 promoter driving TFB3 (1-251)-BUR2 (2-395) fusion with TFB3 5' and 3' flanking sequence, LEU2, 2 μ ori, f1+ ori, AmpR	~3 kb fragment was amplified from pBS-TFB3-BUR2 (SB1973) and using oligos T3 sequencing primer (O #242) and KS poly: EcoRV-EcoRI-Pst (O #4260). ~6.8 kb pRS425 (YV27) backbone was digested with BamHI and SacI. Fragments assembled through Gibson isothermal assembly.
pRS425-TFB3-CTK3	TFB3 promoter driving TFB3 (1-251)-CTK3 (2-296) fusion with TFB3 5' and 3' flanking	~2.7 kb fragment was amplified from pBS-Tfb3-CTK3 (SB1974) using oligos T3 sequencing primer (O #242) and KS poly: EcoRV-EcoRI-Pst (O #4260). ~6.8 kb

	sequence, LEU2, 2 μ ori, f1+ ori, AmpR	pRS425 (YV27) backbone was digested with BamHI and SacI. Fragments assembled through Gibson isothermal assembly.
pRS424-TFB3	TFB3, TRP1, 2 μ ori, f1+ ori, AmpR	~1.9 kb BamHI-Sal I fragment from pRS425-TFB3 (SB1968) cloned into the BamHI - Sal I sites of pRS424 (YV26).
pRS424-TFB3 (1-251)	TFB3 (1-251), TRP1, 2 μ ori, f1+ ori, AmpR	~1.7 kb BamHI-SalI fragment from pRS425-TFB3 (1-251) (SB1994) and was cloned into the BamHI - Sal I sites of pRS424 (YV26).
pRS424-TFB3 (1-144)	TFB3 (1-144), TRP1, 2 μ ori, f1+ ori, AmpR	~1.4 kb BamHI-SalI fragment from pRS425-TFB3 (1-144) (SB1991) was cloned into the BamHI - Sal I sites of pRS424 (YV26).
pRS425	LEU2, 2 μ ori, f1+ ori, pBluescriptII SK polylinker with T7 and T3 promoters flanking, blue-white color selection, AmpR	Sikorski and Hieter (1989) Genetics 122: 19-27.
pRS315	LEU2, CEN/ARS, f1+ ori, pBluescript KS+ polylinker, blue-white color selection, AmpR	Sikorski and Hieter (1989) Genetics 122: 19-27.
pRS425-TFB3 (1-11, 238-Stop)	TFB3 (1-11, 238-Stop), LEU2, 2 μ ori, f1+ ori, AmpR	Oligos TFB3 33 R (O #4334) and TFB3 Asp11-Pro238 (O #4339) were used for inverse PCR on pRS425-TFB3 (SB1968). The resulting ~8 kb fragment was ligated by intramolecular Gibson isothermal assembly.
pRS315-TFB3 (1-11, 238-Stop)-Flag1-TAP	TFB3 (1-11, 238-Stop)-Flag1-TAP, LEU2, CEN/ARS, f1+ ori, AmpR	Primers TFB3 33 R (O #4334) and TFB3 Asp11-Pro238 (O #4339) were used for inverse PCR on pSH1542 (F1255). The resulting ~7.3 kb fragment was ligated by intramolecular Gibson assembly.
pRS425-TFB3 (1-11, 139-Stop)	TFB3 (1-11, 139-Stop), LEU2, 2 μ ori, f1+ ori, AmpR	Primers TFB3 33 R (O #4334) and TFB3 Asp11-Leu139 (O #4337) were used for inverse PCR on pRS425-TFB3 (SB1968). The resulting ~8.3 kb fragment was ligated by intramolecular Gibson assembly.
pLH366	TFB3 (Δ8-75)-Flag1-TAP, LEU2, CEN/ARS, AmpR	Warfield et al. (2016) MCB 36 (19): 2464-2475.
pSH1597	TFB3, URA3, CEN/ARS, AmpR	Warfield et al. (2016) MCB 36 (19): 2464-2475.
pSH1542	TFB3-Flag1-TAP, LEU2, CEN/ARS, AmpR	Warfield et al. (2016) MCB 36 (19): 2464-2475.
pRS315/TFB3Δ2	TFB3Δ2 (1-275), LEU2, CEN/ARS, AmpR, f1+ ori	Feaver et al. (2000) J. Biol. Chem. 275, 5941-5946

Supplemental Table 2. Yeast strains used in this study.

Strain	Genotype
YSB744	*MATα ura3-1 leu2-3,112 trp1-1 his3-11,15 kin28Δ::leu2Δ::TRP1 ade2-1 ade3-22 can1-100* [pRS426-KIN28]
SHY907/ YF2456	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pSH1597]. From Warfield et al. (2016) MCB 36 (19): 2464-2475.
YSB3707	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pSH1542]
YSB3710	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS425-TFB3 (1-144)]
YSB3712	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS425-TFB3]
YSB3713	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS425-TFB3-MPK1]
YSB3715	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS425-TFB3-BUR2]
YSB3717	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS425-TFB3-CTK3]
YSB3722	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251)]
YSB3704	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144)]
YSB3723	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144), pSH1542]
YSB3724	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144), pRS425]
YSB3725	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144), pRS425-TFB3 (1-11, 238-Stop)]
YSB3726	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144), pRS315-TFB3 (1-11, 238-Stop)-Flag1-TAP]
YSB3727	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144), pRS425-TFB3 (1-11, 139-Stop)]
YSB3728	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-144), pLH366]
YSB3729	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251), pSH1542]
YSB3730	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251), pRS425]
YSB3731	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251), pRS425-TFB3 (1-11, 238-Stop)]
YSB3732	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251), pRS315-TFB3 (1-11, 238-Stop)-Flag1-TAP]
YSB3733	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251), pRS425-TFB3 (1-11, 139-Stop)]
YSB3734	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3 (1-251), pLH366]
YSB3786	*MATα ura3-1 leu2-3,112 trp1-1 his3-11,15 kin28Δ::leu2Δ::TRP1 ade2-1 ade3-22 can1-100* [pRS415-KIN28]
YSB3787	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3]
YSB3788	*MATα ura3Δ0 leu2Δ0 trp1Δ63 his3Δ200 ade2Δ::hisG lys2Δ0 met15Δ0 tfb3Δ::HygR tfb6Δ::KanMX* [pRS424-TFB3, pRS425]
YSB207	*MATa ura3-52 leu2-3,112 his3Δ200 tfb1Δ::LEU2* [pRS316-TFB1] From Matsui et al., (1995)

	Nucleic Acids Res 23, 767-772.
YSB260	*MATa ura3-52 leu2-3,112 his3Δ200 tfb1Δ::LEU2* [pRS313-tfb1-101] From Matsui et al., (1995) Nucleic Acids Res 23, 767-772.