Molecular Genetic Dissection of the RNA Polymerase II Transcription Factor D (TFIID) Subunit Taf2

By

Jordan T. Feigerle

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

In

Molecular Physiology & Biophysics

May, 2017

Nashville, Tennessee

Approved:

P. Anthony Weil, Ph.D.

Roland Stein, Ph.D.

William P. Tansey, Ph.D.

Bryan Venters, Ph.D.

Roger Colbran, Ph.D.

ACKNOWLEDGEMENTS

I would like to thank the Cellular, Biochemical and Molecular Sciences Training grant for financial support and for providing student and faculty seminar series that enriched my education during my graduate studies at Vanderbilt University. I would also like to thank the National Institute of General Medical Sciences (NIGMS) at the National Institutes of Health (NIH) for funding the work in the Weil laboratory. Our lab benefitted from the NIGMS decision to increase funding distribution to small labs with single NIH grants as the sole funding source. Furthermore, I am grateful to the Department of Molecular Physiology & Biophysics for providing bridge funding while our lab was in between grants.

I appreciate the advice and counsel I have received from all of my committee members as well as my labmates, Amanda Johnson and Dr. Chirie Sumanasekera. All were very interested in my experiments when I would discuss them (sometimes for longer than they would like) and provided excellent advice and encouragement. I owe particular thanks to Dr.'s Bill Tansey and Roland Stein for writing me reference letters for fellowship applications and my post-doctoral fellowship position with Dr. Roger Kornberg at Stanford.

I would not have become the scientist I am today without the mentorship I received from Dr. Tony Weil. Tony allowed me the freedom to pursue my projects in the manner of my choosing. And he gave me the space I needed to both succeed and fail on the bench. Importantly, whenever I needed experimental advice, no matter how detailed, Tony was available and eager to help. I also owe my systematic approach and understanding of important experimental controls from Tony's research approach. I am confident that this training will serve me well as I move forward in my research career. I also tremendously enjoyed our non-science discussions. While we agreed on much related to current events

and politics, our disagreements were always illuminating and provided me an educated perspective on the world, in a way I often had not considered. I will certainly miss our conversations, even the ones that sometimes turned combative.

I would like to thank my family. I am fortunate to be the son of two parents who valued my education and my freedom to make my own choices. My mother Minki was always understanding about my graduate studies and would listen to all of the details of my project, often without understanding much of what I was saying. I imagine it was easier for her than most because she had supported my father in much the same way when he was earning his doctorate. My father Chuck offered similar support and was able to provide me advice both as a Ph.D. and a principal investigator who has trained many students in his research career. I am thankful for both of their encouragement throughout all of these years during graduate school and over the course of my life. My brother Travis also provided an important outlet for me. He would listen to my research and understand the best he could. But we would also text back and forth a few times a month about fun stuff and hobbies. Travis also made a point to visit Tennessee multiple times a year even though he lives in Houston and it is a significant time investment. I also can't thank him enough for being my best man at my wedding.

Finally, I have to thank my wife Dr. Katie Hutchinson-Feigerle. Katie and I both entered Vanderbilt for graduate school at the same time. She finished a few years earlier than I did but stayed at Vanderbilt with me for a post-doc even though she could have gone anywhere she wanted to based on her success in research. I needed her support, both emotionally and in my experiments. I would often overestimate how long something in the lab would take me, including performing literally thousands of transformation reactions. Katie would

help me plate all of those transformations on certain occasions or stay with me in the lab to keep me company so I would not have to be there late on my own. Even when I told her that I wanted to go to Stanford and work in the Kornberg lab, she did not hesitate to support me, even though she knew she would be moving 2000 miles away from her family. I am very grateful to her for the sacrifices she has made for me and I hope that I can repay the favor throughout our new adventure in California.

TABLE OF CONTENTS

Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Overview

In all domains of life, cellular function depends upon appropriate regulation of the genes encoded in the DNA genome. Historically, this regulation was described in terms of the "central dogma" of molecular biology (Crick, 1970) whereby DNA is transcribed into RNA, which is in turn translated into proteins, the functional workhorses of the cell. This simplistic view of molecular biology – simplistic in the sense that the central dogma primarily applies to coding RNA (messenger RNAs (mRNA)) and not non-coding RNAs (transfer RNA (tRNA), ribosomal RNA (rRNA), small non-coding RNAs (ncRNA) and long non-coding RNA (lncRNAs)) which perform important cellular functions without being translated into proteins – provides an important framework to understand how the cell can regulate the expression of the DNA genome. Specifically, gene activity can be increased or decreased based on processes that regulate the synthesis or degradation of RNA (transcription), the conversion of RNA to protein (translation) or the activity or stability of the protein product. At the cellular level, appropriate gene regulation dictates cellular function and identity, which is especially important in multicellular organisms. Considering the human genome consists of approximately 3 billion base pairs encoding at least 20,000 mRNA genes that must be coordinately and differentially regulated in all of the tissue types of the body, the importance of gene regulation in cellular and organismal

function cannot be overstated (International Human Genome Sequencing Consortium, 2004).

The focus of this dissertation is on the regulation of the transcription process, the first step in the gene regulatory pathway. In yeast and metazoans, transcription of the nuclear genome is performed by three DNA-dependent RNA polymerases (Vannini and Cramer, 2012). RNA polymerase I (Pol I) transcribes the 28s and 18s rRNA genes; RNA polymerase II (Pol II) transcribes the mRNA genes as well as certain ncRNA and lncRNAs genes; and RNA polymerase III (Pol III) transcribes the 5s rRNA gene, the tRNA genes and certain ncRNA genes. The transcription process for all of these RNA polymerases includes three distinct steps, each of which is a target for important modes of regulation: transcription initiation, elongation and termination. More specifically, I will focus on the regulation of transcription initiation of the mRNA genes by Pol II, in particular the role the general transcription factor and coactivator TFIID plays in Pol II transcription initiation. Functional similarities between eukaryotic Pol II transcription and prokaryotic transcription are discussed.

In addition, the work presented herein was performed in the model system *Saccharomyces cerevisiae*, also known as baker's yeast. While certain differences exist between yeast and higher eukaryotic systems with regard to transcription initiation (some of which will be described in detail below), the vast majority of the proteins and gene structures that play important roles in transcription regulation are highly evolutionarily conserved at the amino acid, structure and functional levels. Furthermore, *S. cerevisiae* is an excellent model organism for hypothesis testing due to its short doubling time (90 minutes) and facile genetics, without which my project would not have been possible.

Structure of an mRNA Gene

The structures of mRNA genes include both the sequences that are critical for regulating transcription initiation, elongation and termination as well as the coding sequence that is recognized by the translational machinery to synthesize the encoded protein product (Figure 1.1). Transcription initiation is guided by two structures within the mRNA gene sequence: the enhancer/silencer and the core promoter. Enhancers are sequences that function to stimulate transcription while silencers are sequences that repress transcription. In general, while the core promoter is positioned 5' of the transcription start site (TSS) and oriented in a manner to direct transcription through the mRNA transcription unit, the enhancer/silencer is not constrained by these position or orientation limitations (Li et al., 2016). Enhancers/silencers can be located often kilobases to megabases away from the TSS and these sequences can be 5' of the TSS, within the transcription unit such as within introns or 3' of the transcription termination sequence (TTS). However, in *S. cerevisiae*, likely as a result of a more compact genome, enhancers/silencers are known as upstream activating sequences (UAS) or upstream repressing sequences (URS) and are located 5' to the core promoter, often within 1000bp of the TSS (Hahn and Young, 2011). These sequences are collections of binding sites for trans-acting factors. These trans-acting factors can both stimulate transcription (trans-activators) and repress transcription (trans-repressors).

In most eukaryotic systems including metazoans and the fission yeast *Schizosaccharomyces pombe*, which has been argued based on sequence conservation to more similar to metazoans than to budding yeast (Sipiczki, 2000), the core promoter is

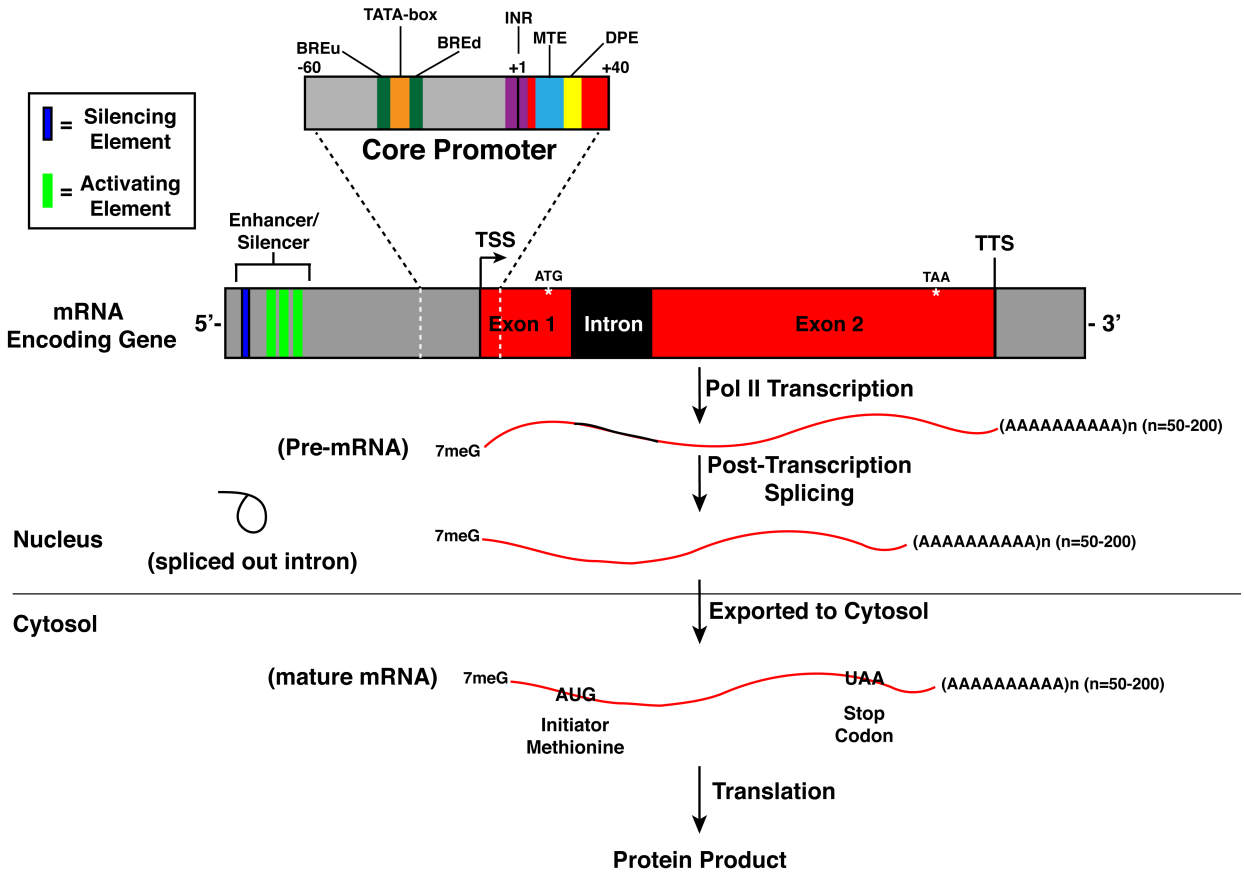


Figure 1.1. Structure and regulation of an mRNA gene. A model for the structure of a eukaryotic gene is presented. Within the DNA genome, housed in the nucleus, a gene consists of the cis-elements that regulate the association between RNA polymerase II and the transcription unit, including the transcription start site (TSS) and the transcription termination site (TTS). The cis-regulatory elements include the enhancer/silencer and the core promoter. A number of sequence elements within the core promoter have been identified, the most common of which are displayed. These elements are often not all present within the same core promoter but are displayed as such for simplicity. The transcription unit often includes protein coding exons interspersed with non-coding introns. The first exon begins with the transcription start site while the final exon contains the TTS and a polyadenylation sequence. RNA polymerase II uses the DNA template to transcribe the RNA from the TSS to the TTS and past until the polymerase is subject to active transcription termination mechanisms (not discussed here). The resulting mRNA is post-transcriptionally 7-methylguanisine (7meG) capped at its 5' end and polyadenylated at its 3' end. The resulting transcript, including the intron, is the pre-mRNA and it is maintained in the nucleus. The intron is spliced out of the pre-mRNA to form the mature mRNA, which then exits the nucleus. Protein translation occurs in the cytosol where ribosomes, reading from the start codon (initiator methionine) to the stop codon, synthesize the protein product. Importantly, each step of the process is subject to regulation including transcription, post-transcriptional processing, nuclear export and translation. BREu= TFIIB recognition element upstream, BREd= TFIIB recognition element downstream, INR= initiator element, MTE= motif-ten element, DPE=downstream promoter element.

located approximately 30bp 5' of the transcription start site. However, in *S. cerevisiae* the core promoter location varies between 40bp and 120bp upstream of the TSS likely because of inherent differences in yeast Pol II (Li et al., 1994). The core promoter is a degenerate sequence containing several documented binding elements: the TATA-box, the initiator (INR), the downstream promoter element (DPE), the motif-ten element (MTE) and the TFIIB recognition element (BRE), among others that have been identified for specific gene classes (Burke and Kadonaga, 1997; Deng and Roberts, 2006; Kadonaga, 2012; Lagrange et al., 1998; Lim et al., 2004; Smale and Baltimore, 1989; Thomas and Chiang, 2006). Core promoters do not contain consensus sequences for all of these binding elements but instead contain some combination of these elements. Other than the TATA-box, none of the other core promoter elements have been unambiguously identified in the yeast system. Furthermore, only 20% of genes contain consensus TATA-boxes while the remaining genes have historically been considered TATA-less (Basehoar et al., 2004) although recent studies have demonstrated that the majority of yeast core promoters contain TATA-like elements (Rhee and Pugh, 2012). The functions of these core promoter elements and how they interface with the RNA polymerase II transcription machinery will be described more below.

Following transcription initiation by RNA polymerase II, the polymerase will transcribe the entire transcription unit. This transcription unit contains the TSS, the 5' untranslated region (UTR), a start codon (AUG, initiator methionine), a stop codon, a 3' UTR and a TTS. In addition, mRNA genes are divided into introns and exons. Introns are non-coding sequences that are interspersed between coding exons. These introns must be excised from the RNA and the exons must be ligated together to form the mature mRNA

that can be transported out of the nucleus and used as a template for protein synthesis. The sequence encompassed by the translation start and stop codon of the mature mRNA is the open reading frame (ORF). The translation machinery initiates protein synthesis at the start codon and terminates at the stop codon to synthesize the protein encoded by the mRNA. Once RNA polymerase II has completed transcription of the 3' UTR and terminated transcription, it is free to reinitiate on the same mRNA gene or move to a different mRNA gene and begin an entirely new transcription cycle (Hahn, 2004).

The RNA Polymerase II General Transcription Factors and Their Relation to Prokaryotic Transcription

In prokaryotes, RNA polymerase (RNAP) holoenzyme can specifically bind to promoter sequences and accurately initiate transcription (Murakami, 2015). RNAP holoenzyme is composed of 6 subunits: σ, ω, α (two copies), β and β'. While w is not essential for RNAP function, the remaining five polypeptides are all that is required for transcription initiation. The eukaryotic transcription machinery is exponentially more complicated. As mentioned above, instead of one RNAP, eukaryotes have 3 nuclear RNA polymerases (Roeder and Rutter, 1969). In addition, eukaryotes contain mitochondrial and chloroplast RNA polymerases that are responsible for transcription of the mitochondrial and chloroplast genomes (Arnold et al., 2012; Börner et al., 2015). Plants contain two additional RNA polymerase (RNA polymerase IV and RNA Polymerase V) that transcribe the genes encoding non-coding RNAs that are involved in gene silencing functions (Haag and Pikaard, 2011). The increased complexity of the eukaryotic transcription machinery provides a vast increase in the modes through which eukaryotes can regulate gene

expression, likely allowing for evolution of the many diverse and highly complex metazoan and plant species.

Focusing solely on mRNA gene transcription, Pol II cannot perform promoter specific transcription on its own (Weil et al., 1979). Instead, six additional protein factors are required for Pol II to initiate transcription (Flores et al., 1992; Reinberg et al., 1987; Reinberg and Roeder, 1987a). These co-factors are known as the general transcription factors: transcription factor for RNA polymerase II A (TFIIA), TFIIB, TFIID, TFIIE, TFIIF and TFIIH. These six along with Pol II form the pre-initiation complex (summarized in Table 1.1), which is required for promoter-specific transcription by Pol II (Figure 1.2). The functions and subunit structures of each of the transcription factors and Pol II are briefly described below:

TFIID: TFIID is a multi-protein complex containing 14 conserved subunits. These subunits include TATA-box binding protein (TBP) and 13 TBP-associated factors (Taf1→13). While TBP is required for transcription by all three nuclear DNA-dependent RNA polymerases, TFIID specifically functions with RNA polymerase II (Kim and Roeder, 1994). In *S. cerevisiae*, TFIID contains an extra subunit, Taf14, that is also a component of 6 other transcription related complexes including the general transcription factor TFIIF (Cairns et al., 1996; Henry et al., 1994; John et al., 2000; Kabani et al., 2005; Lim et al., 2007; Sanders et al., 2002b). In addition to being a multi-subunit complex, several of the TFIID subunits are present at 2 moles per mole of TFIID including Tafs 5, 6, 9, 10, 12 and 14 (potentially greater than 2 moles per mole of TFIID) (Sanders et al., 2002a). Thus, TFIID is composed of at least 19 individual polypeptides, depending on the species (Tora, 2002). In *S. cerevisiae*, with the exception of *TAF14*, all of the genes encoding TFIID subunits are

single-copy essential genes (Klebanow et al., 1997, 1996; Moqtaderi et al., 1996; Poon et al., 1995; Ray et al., 1991; Reese et al., 1994; Sanders and Weil, 2000). *taf14* null cells display a slow growth phenotype and temperature sensitive (Ts) growth.

Pol II Subunits	kDa	TFIIA Subunits	kDa	TFIID Subunits	kDa	TFIIH Subunits	kDa
Rpb1	192	Toa1	32	Taf1	121	Ssl2	95
Rpb2	139	Toa2	13	Taf2	161	Rad3	90
Rpb3	35	**TFIIB Subunits**	**kDa**	Taf3	40		
Rpb4	25			Taf4	42	Tfb1	73
Rpb5	25	Sua7	38	Taf5	89	Tfb2	59
Rpb6	18	**TFIIE Subunits**	**kDa**	Taf6	58		
Rpb7	19			Taf7	67	Ssl1	52
Rpb8	17	Tfa1	55	Taf8	58	Tfb3	38
Rpb9	14	Tfa2	37	Taf9	17		
Rpb10	8	**TFIIF Subunits**	**kDa**	Taf10	23	Ccl1	45
Rpb11	14			Taf11	41	Tfb4	37
Rpb12	8	Tfg1	82	Taf12	61		
		Tfg2	47	Taf13	19	Kin28	35
		Tfg3	27	Taf14	27		
				TBP	27	Tfb5	8

Table 1.1: The RNA polymerase II General Transcription Factors in *S. cerevisiae*. Subunits and molecular weights of general transcription factors. In total, the PIC is composed of ~49 polypeptides that amounts to ~2.5 MDa in size, not including additional co-activators such as SAGA and mediator, or DNA. Values derived from Saccharomyces Genome Database.

TFIID serves two known transcription initiation functions: promoter recognition and transcription co-activation (Albright and Tjian, 2000). As indicated by its name, TBP binds to the TATA-box in gene core promoters. TBP-promoter association is absolutely required for PIC formation because the rest of the general transcription factors cannot specifically associate with the core promoter in the absence of TBP (Buratowski et al., 1989). In this sense, TBP most resembles σ factor from prokaryotic RNAP. The σ subunit of RNAP is required to recognize the core promoter in prokaryotic gene promoters (Sugiura et al., 1970). Without σ, RNAP cannot perform promoter-specific transcription. When core promoters contain consensus TATA-boxes, TBP can function independently of Tafs to nucleate PIC formation. In fact, TBP has been used extensively for *in vitro*

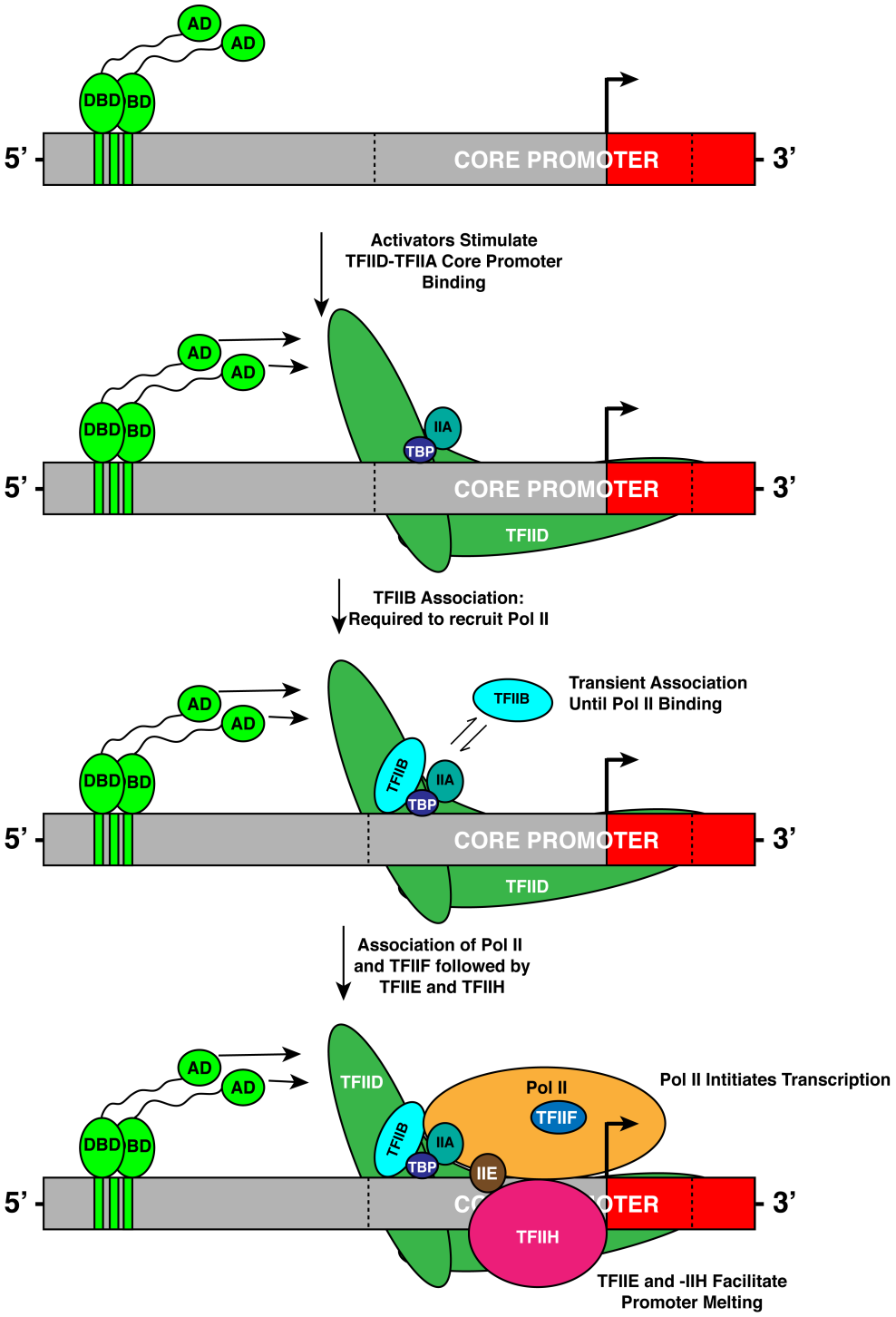


Figure 1.2: Model for Step-wise Formation of the Pre-Initiation Complex. The trans-activator associates with its cognate binding site within the enhancer via the DNA binding domain (DBD). The activation domain (AD) stimulates TFIID and TFIIA association with the promoter. TFIID and TFIIA binding to DNA is stable while TFIIB binding is dynamic. Pol II-TFIIF association through TFIIB, stabilizes TFIIB binding. TFIIH and TFIIE associate last and are required for promoter melting. The fully formed PIC enables Pol II to initiate transcription.

transcription with purified components (He et al., 2016; Murakami et al., 2013; Myers et al., 1997). However, in the absence of a consensus TATA-box, TFIID Tafs are likely to play an important role in core promoter recognition through direct binding to sequence elements within the core promoter (Figure 1.3). When TBP is associated with TFIID, TBP is incapable of binding to DNA because the Taf1 N-terminal Domain (TAND) binds to the DNA binding cleft of TBP and mimics TATA-DNA (Bai et al., 1997; Kokubo et al., 1998; Kotani et al., 1998). This repression of TBP binding can be alleviated through the binding of TFIIA.

Prokaryotes utilize alternative σ factors that have different core promoter binding specificities to regulate different transcriptional programs. These alternative σ factors allow prokaryotes to respond to external stress or nutrient availability or initiate alternative expression programs distinct from vegetative growth such as in response to viral infection or sporulation (Burgess et al., 1969; Duffy and Geiduschek, 1977; Fox et al., 1976; Haldenwang et al., 1981; Haldenwang and Losick, 1980, 1979). Similarly, eukaryotes also contain alternate forms of both Tafs and TBP that are required for expression of transcription programs within differentiated tissue types. For example, drosophila testes express cannonball/can (Taf5 paralog), no hitter/nht (Taf4 paralog), meiosis I arrest/mia (Taf6 paralog), spermatocyte arrest/sa (Taf8 paralog) and ryan express/rye (Taf12 paralog) (Hiller et al., 2004). Alternative forms of TBP (TBP-related factors or TRFs) have functional roles in muscle cell-specific transcriptional programs as well as transcription of the ribosomal protein genes in drosophila and human cells (Deato et al., 2008; Wang et al., 2011). However, a recent study by the labs of Tora and Puri have called into question the role of TRF2 in muscle cell differentiation (Malecova et al., 2016). Regardless, the current model in the field is that alternative TFIID subunits play specialized roles to increase the

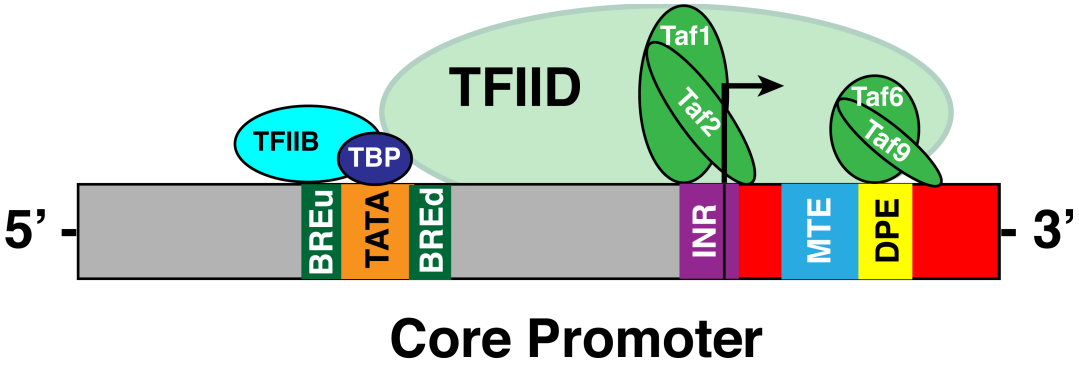


Figure 1.3: Association of TFIIB and TFIID with core promoter sequence elements. Contacts made by TBP, Tafs and TFIIB with core promoter are shown. Protein-DNA contacts identified by various methods including protein-DNA cross-linking and DNaseI footprinting.

functional diversity within tissue types of multi-organ species and allow for the increasingly complex modes of gene regulation that are not required in prokaryotes.

The other function of TFIID is to serve as a transcription co-activator (Pugh and Tjian, 1990). In this role, TFIID directly interacts with trans-activator proteins (Chen et al., 1994; Garbett et al., 2007; Gill et al., 1994; Goodrich et al., 1993; Hoey et al., 1993; Liu et al., 2009; Tanese et al., 1991; Thut et al., 1995). While the exact mechanism(s) of how TFIID-activator interactions stimulate transcription has not been well explored, the predominant model is known as "recruitment" (Pugh and Tjian, 1991, 1990, Sauer et al., 1995a, 1995b). While recruitment is an imprecise term, TFIID recruitment refers to the concept whereby TFIID-core promoter DNA binding affinity is weak in the absence of a trans-activator protein leading to an inability of TFIID to promote PIC formation and transcription initiation. However, trans-activator binding to the cis-linked enhancer can facilitate TFIID association with the core promoter, which can be achieved through either direct trans-activator-TFIID binding that results in cooperative binding to the core promoter or indirectly through a protein intermediate that interacts with both TFIID and the trans-

activator (Johnson et al., 2002; Sauer et al., 1995a). Ultimately, an increase in TFIID-promoter DNA binding affinity leads to an increase in TFIID residency on promoters resulting in stimulation of PIC formation. Prokaryotes utilize similar mechanisms to stimulate RNAP association with promoters. For example, catabolite activator protein (CAP) binds to cis-linked DNA sequences upstream of the core promoter and stimulates RNAP-promoter association through direct interaction with the C-terminal domain of the α subunit (Heyduk et al., 1993).

TFIIA: TFIIA is composed of two polypeptides in yeast, Toa1 and Toa2 (Ranish and Hahn, 1991). Both *TOA1* and *TOA2* are single copy essential genes (Ranish et al., 1992). In metazoans, the large subunit of TFIIA is synthesized as single polypeptide (TFIIAαβ) but post-translationally processed to form a three subunit complex, TFIIAα, β, γ (Ma et al., 1993; Zhou et al., 2006). Biochemically, TFIIA directly binds to TBP and can stabilize binding of TBP to DNA although TFIIA does not display significant DNA binding activity on its own (Yokomori et al., 1994). In addition, TFIIA has been shown to serve an anti-repressor function by displacing interactions between the Taf1-TAND and the TBP DNA binding domain, allowing TBP to access DNA (Ozer et al., 1998). Furthermore, TFIIA directly interacts with trans-activator proteins and can stimulate activated transcription in a TFIID-dependent manner (Kobayashi et al., 1995; Ozer et al., 1996). Despite the role of TFIIA in activated transcription, when TBP is substituted for TFIID, TFIIA is not required for *in vitro* transcription performed with purified Pol II and the rest of the general transcription factors (Murakami et al., 2013).

TFIIB: *SUA7* encodes TFIIB in *S. cerevisiae* (Pinto et al., 1992). *SUA7* is a single copy essential gene. TFIIB is the one general transcription factor that is composed of a

single polypeptide instead of a multi-subunit complex. Similar to TFIIA, TFIIB can also stabilize TBP-DNA binding (Imbalzano et al., 1994). TFIIB mediates the interaction between TBP and Pol II (Barberis et al., 1993; Buratowski et al., 1989; Buratowski and Zhou, 1993; Killeen et al., 1992; Maldonado et al., 1990). This interaction between TFIIB and Pol II also stimulates Pol II transcription initiation activity and TSS selection (Pardee et al., 1998). TFIIB may also function to maintain the transcription bubble (He et al., 2016).

TFIIE: TFIIE is composed of two polypeptides, Tfa1 and Tfa2 (Feaver et al., 1994). TFIIE forms a heterotetramer in solution (Inostroza et al., 1991). Both *TFA1* and *TFA2* are single copy essential genes. TFIIE binds directly to both Pol II and TFIIH (Maxon et al., 1994). Binding of TFIIE to TFIIH stimulates TFIIH enzymatic activity (described below) (Lu et al., 1992; Ohkuma and Roeder, 1994). In addition, TFIIE is thought to help Pol II maintain an open transcription bubble (Buratowski et al., 1991; Forget et al., 2004; Holstege et al., 1996; Plaschka et al., 2016).

TFIIF: TFIIF is composed of three polypeptides in *S. cerevisiae*, Tfg1, Tfg2 and Tfg3 (Henry et al., 1992). Tfg3 is identical to Taf14. *TFG1* and *TFG2* are both single copy essential genes (Henry et al., 1994). In metazoans, Tfg3 is not present in TFIIF (Flores et al., 1989). TFIIF is tightly associated with Pol II (Rani et al., 2004) and functions in a similar manner to TFIIB in the sense that it is important for Pol II incorporation into the pre-initiation complex (Flores et al., 1991), initiation of transcription and TSS selection (Ghazy et al., 2004). TFIIF can also stimulate transcription elongation (Conaway et al., 2000).

TFIIH: TFIIH is a ten subunit complex and in *S. cerevisiae* is composed of Ssl2, Rad3, Tfb1, Tfb2, Ssl1, Tfb3, Ccl1, Tfb4, Kin28 and Tfb5 (Murakami et al., 2012; Ranish et al., 2004; Svejstrup et al., 1994). With the exception of *TFB5*, all of the genes encoding

TFIIH subunits are essential for viability (Feaver et al., 1997; Gileadi et al., 1992; Gulyas and Donahue, 1992; Higgins et al., 1983; Naumovski and Friedberg, 1983; Park et al., 1992; Simon et al., 1986; Valay et al., 1993; Yoon et al., 1992). TFIIH has two enzymatic activities that are critical for transcription initiation. First, Ssl2 acts as a helicase and melts the promoter DNA such that Pol II can begin scanning for the transcription start site (Guzder et al., 1994). Prokaryotic RNAP contains this helicase activity within the σ and β subunits (Fisher and Blumenthal, 1980; Wigneshweraraj et al., 2003). Second, Kin28 has kinase activity and phosphorylates Ser5 and Ser7 in the heptad repeat of the Rpb1 C-terminal domain (CTD) (Komarnitsky et al., 2000; Lu et al., 1992) (importance of the Ser5 modification is described more below). As mentioned above, direct interaction between TFIIE and TFIIH stimulates the enzymatic activities of TFIIH.

Pol II: Pol II contains 12 subunits, Rpb1-Rpb12 (Edwards et al., 1990; Kolodziej et al., 1990; Treich et al., 1992; Vannini and Cramer, 2012; Woychik et al., 1993). With the exception of Rpb4 and Rpb9, all of the Pol II subunits are essential for life (Woychik et al., 1993, 1991; Woychik and Young, 1989). The three nuclear RNA polymerases all contain a 10 subunit catalytic core, five of which are shared among the polymerases: Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12. Rpb6 is similar to ω, both subunits serve structural roles within the RNA polymerase (Minakhin et al., 2001). Together, Rpb1 and Rpb2 compromise the active center, which contains the nucleotide binding pocket and catalyzes phosphodiester bond formation (Riva et al., 1987), and are similar to the β' and β subunits, respectively (Allison et al., 1985; Cramer et al., 2001; Sweetser et al., 1987). Subunits Rpb3 and Rpb11 are the remaining subunits of the catalytic core and most resemble the two α subunits (Zhang and Darst, 1998). Rpb4 and Rpb7 are less stably associated with the catalytic core.

These subunits are critical for transcription initiation (Edwards et al., 1991) but are thought to dissociate during transcription elongation (Mosley et al., 2013). Pol II also has another accessory factor that is important for transcription elongation, TFIIS (Rappaport et al., 1987; Reinberg and Roeder, 1987b; Sekimizu et al., 1979). This accessory factor can stimulate the RNA cleavage activity of Rpb9 as a mechanism to backtrack during transcriptional stalling (Awrey et al., 1997; Reines, 1992). The additional subunits of Pol I (A49) and Pol III (C37, C53 C82, C34 and C31) function in a similar manner to TFIIE and TFIIF, as described above (Carter and Drouin, 2010; Geiger et al., 2010; Kuhn et al., 2007).

The Rpb1 C-Terminal Domain

A unique characteristic of Pol II is the C-terminal domain (CTD) of Rpb1 (Allison et al., 1985). Rpb1, the largest subunit in Pol II, contains a repeating sequence known as the heptad repeat ($Tyr^1Ser^2Pro^3Thr^4Ser^5Pro^6Ser^7$) at its C-terminus. The repeat length varies by species with *S. cerevisiae* containing 26 repeats and mammals containing 52 repeats, although many of the mammalian heptad repeats are non-consensus in sequence (Corden et al., 1985; Nonet et al., 1987; Wintzerith et al., 1992). While not all of these repeats are essential for life, partial truncation of the CTD results in Ts growth in yeast and strains harboring a complete deletion of the Rpb1 CTD are inviable (Allison et al., 1988). Similar mutations in metazoan cells cause loss in viability (Zehring et al., 1988).

Importantly, the Rpb1 CTD is the target for significant post-translational modification (Zaborowska et al., 2016). Both proline residues are subject to isomerization (Hanes, 2014) while the remaining consensus residues can be both phosphorylated (Kolodziej et al., 1990) and glycosylated (Kelly et al., 1993). Some of the non-consensus

residues in the mammalian CTD can be ubiquitylated (Li et al., 2007), methylated (Sims et al., 2011) and acetylated (Voss et al., 2015). In general, site-specific phosphorylation has been most intensively studied. For example, ChIP-seq studies have shown that Ser5 phosphorylation (Ser5-p) is localized closest to the 5' end of the transcription unit while Ser2 is localized more towards the 3' end of the transcription unit (Komarnitsky et al., 2000). Ser5-p is catalyzed by the Kin28 subunit of TFIIH while Ser2 is phosphorylated by the Bur1/Bur2 and Ctk1 kinases (Keogh et al., 2003; Lee and Greenleaf, 1991; Qiu et al., 2009).

Using a truncated Rpb1 CTD in *S. pombe*, Schwer and Shuman exhaustively characterized the contribution of each residue of the heptad repeat to life (Schwer and Shuman, 2011). Importantly, they show that Ser5 is essential for life. However, the function of this residue can be largely bypassed if an mRNA capping enzyme is fused to the Rpb1 CTD. Thus, a major function of Ser5-p is the recruitment of mRNA capping enzymes (Cho et al., 1997; McCracken et al., 1997a). While prokaryotes do not cap their mRNAs, eukaryotes use a 5' 7-methylguanosine (7meG) cap that serves to both protect the RNA from 5' to 3' nuclease digestion as well as provide a binding site for the translation machinery (see Figure 1.1) (Both et al., 1975; De Kloet and Andrean, 1976; Muthukrishnan et al., 1975; Wei and Moss, 1975). Thus, this cap is a critical RNA post-transcriptional modification for eukaryotic translation.

Also, in Schwer and Shuman's work, they demonstrate Ser2 can be mutated to alanine. However, mutating Ser2 to the phosphomimetic Glu results in lethality. This residue it thought to play a significant role in co-transcriptional splicing and transcription termination including recruitment of the polyadenlyation machinery (Barillà et al., 2001;

Gu et al., 2013; McCracken et al., 1997b). mRNA polyadenylation is thought to play an important role in RNA stability as well as translation activity (Dreyfus and Régnier, 2002). While eukaryotes contain polyA sequences between 50-200nt (see Figure 1.1), prokaryotes do contain polyA sequences but they are less than 10 nucleotides and are thought to promote RNA degradation rather than stability.

Role of Co-activators in Transcription Initiation

While the general transcription factors and Pol II are sufficient for basal transcription *in vitro*, PIC formation *in vivo* requires communication between enhancer bound trans-activators and proteins at the core promoter. In addition to TFIID, two additional co-activator complexes contribute importantly to general transcription initiation mechanisms: the Spt-Ada-Gcn5-acetyltransferase (SAGA) complex and Mediator (Grant et al., 1997; Kim et al., 1994). SAGA, a complex composed of at least 20 different subunits, shares the Taf 5, 6, 9, 10 and 12 subunits with TFIID including (Grant et al., 1998). SAGA can also chaperone TBP to mRNA gene core promoters (Dudley et al., 1999). The mechanisms by which SAGA contributes to transcription activation include direct interaction between SAGA subunits Tra1 with trans-activators including Gal4 (Bhaumik et al., 2004) and Gcn4 (Fishburn et al., 2005), chromatin based mechanisms (described more below) and aids in the recruitment of mediator (Bryant and Ptashne, 2003; Larschan and Winston, 2005).

Given the shared subunits, TBP chaperone activities and activator binding activities, TFIID and SAGA are believed to play complementary roles. Genome-wide analyses have demonstrated that SAGA is the pre-dominant co-activator on stress response

genes (around 10% of yeast genes) which often contain consensus TATA-boxes such as the *GAL* genes, the *PHO* genes and genes that respond to amino acid starvation (Gcn4-regulated genes) (Huisinga and Pugh, 2004). Meanwhile, TFIID dominates the other 90% of yeast gene promoters, which predominantly lack consensus TATA-boxes. The terms "TFIID-dominated" and "SAGA-dominated" are used pervasively in the literature to describe which complex is more responsible for the regulation of a gene. However, recent studies have expanded the role SAGA plays in regulating TFIID-dominated genes. Using 4-thiouracil RNA pulse labeling in cells containing either WT SAGA or deletion of subunits required for SAGA structural integrity (Spt7 and Spt20), Tora and colleagues showed that RNA synthesis was dramatically reduced in TFIID-dominated genes despite near WT levels of steady-state transcripts (Bonnet et al., 2014). Thus, SAGA appears to play an important role in the transcription for all yeast and human Pol II transcribed genes. Furthermore, Hahn and colleagues recently demonstrated that Taf1 also occupies SAGA-dominated genes using the newly developed chromatin endogenous cleavage and high throughput sequencing (ChEC-seq) method (Grünberg et al., 2016; Zentner et al., 2015). Altogether, TFIID and SAGA appear to have significant functional overlap that will need to be explored further in future research.

Mediator is another multisubunit complex containing at least 21 subunits in *S. cerevisiae*. Kornberg and colleagues discovered Mediator as a biochemical activity that could alleviate the ability of Gal4-VP16 to interfere with the ability of other trans-activators to stimulate transcription (Kelleher et al., 1990). Co-incident with the Kornberg lab's discovery, Richard Young and colleagues genetically identified mutations in several Mediator subunits (not known as Mediator subunits at the time) that could suppress the Ts

phenotype of strains harboring a truncated C-terminal domain of Rpb1 (Thompson et al., 1993). Subsequently, both the Kornberg lab and the Young lab have postulated that Mediator and Pol II come together to form the Pol II holoenzyme that is required for transcription initiation *in vivo* (Koleske and Young, 1994). Mediator has been shown to contribute to all phases of the transcription cycle – initiation, elongation and termination – by physically and functionally interacting with all of the general transcription factors, TFIIS and Pol II (Allen and Taatjes, 2015; Baek et al., 2006; Conaway and Conaway, 2013; Guglielmi et al., 2007; Johnson and Carey, 2003; Kremer et al., 2012; Malik et al., 2007; Mukundan and Ansari, 2011; Sakurai et al., 1996; Sakurai and Fukasawa, 1998; Takagi et al., 2006; Takahashi et al., 2011; Wery et al., 2004).

The Role of Chromatin and Nucleosome Positioning in Transcription Regulation

While much of the work done to define the components required for transcription initiation has been performed using naked DNA, the eukaryotic DNA genome is packaged into structures called chromatin (Olins and Olins, 2003). The building block of chromatin is the nucleosome: ~147bp of DNA wrapped around a histone octamer composed of two copies of heterodimeric pairs of histones H2A/H2B and H3/H4 (Kornberg, 1974; Luger et al., 1997). In general, the nucleosome is inhibitory to transcription because most DNA binding proteins are unable to bind their cognate cis-elements when it is wrapped around the histone octamer. However, core promoters for constitutively active genes are contained within a region known as the nucleosome free region (NFR) (Choder et al., 1984; Jakobovits et al., 1980; Saragosti et al., 1980, p. 40). This NFR is maintained through two likely interdependent mechanisms: first, ATP- dependent

chromatin remodelers such as RSC, Swi/Snf and Ino80 can slide histones upstream and downstream of the core promoter leaving it open (Lorch and Kornberg, 2015; Yen et al., 2012) and second, binding of the general transcription factors and Pol II can occupy the core promoter and thus prevent nucleosome formation. Targeting of these ATP-dependent chromatin remodelers to chromatin is thought to occur through pioneer transcription factors (Swinstead et al., 2016). These factors can bind to chromatinized DNA enhancers (Cirillo et al., 2002, 1998) and recruit ATP-dependent remodelers to open up the chromatin leading to gene activation (Figure 1.4) (Hu et al., 2011; L. Wang et al., 2014).

Recently, the labs of Pugh and Korber used purified components in an attempt to reconstitute the distribution of chromatin on yeast genomic DNA (Krietenstein et al., 2016). They observed that, in the presence of global transcription regulators (Abf1/Reb1), chromatin-remodeling enzymes (RSC, INO80, ISW2 and ISW1a) order the nucleosomes in a sequential process. This process entails clearing of the nucleosome free region, establishing the appropriate positioning of the +1 nucleosome and arranging the downstream nucleosomes to have the appropriate internucleosomal spacing. While this study made significant progress toward defining the mechanism of nucleosome positioning genome wide, none of assays the labs of Korber and Pugh used ever reconstituted the nucleosome positioning observed *in vivo* and, thus, only informs us of the contributions that ATP-dependent chromatin remodelers can make. To truly test their model, they need to perform these studies in cells with conditionally (Anchor Away or auxin-inducible degron) depleted chromatin remodeling enzymes, either individually or in combination, and map nucleosome positioning genome-wide (Haruki et al., 2008; Nishimura et al., 2009).

The *PHO5* System as a Model for Chromatin-Dependent Gene Activation

Arguably, the most well characterized system with regard to chromatin remodeling as a gene activation mechanism is the activation of the *PHO5* gene in *S. cerevisiae* (Korber and Barbaric, 2014). Specifically, under inorganic phosphate rich conditions, *PHO5* is repressed. Repression is achieved through two mechanisms. First, nucleosomes occlude one of the binding sites for the Pho4 trans-activator and the TATA-box (Almer et al., 1986; Almer and Hörz, 1986; Barbarić et al., 1996; Bergman et al., 1986; Bergman and Kramer, 1983; Venter et al., 1994). Second, Pho4 is sequestered to the cytosol through phosphorylation (Kaffman et al., 1998, 1994; O'Neill et al., 1996).

PHO5 activation is achieved through the following mechanism. Under inorganic phosphate starvation conditions, Pho4 is no longer phosphorylated (Schneider et al., 1994), can enter the nucleus and bind cooperatively with Pho2 at its cognate binding site on *PHO5* (Barbaric et al., 1998). Pho4 can then direct remodeling of the histones within the *PHO5* core promoter by ATP-dependent chromatin remodeling complexes exposing a second Pho4 binding site and the TATA-box (Barbaric et al., 2007; Gregory et al., 1999; Svaren et al., 1994). SAGA is also critically important for this process serving two distinct roles: histone acetylation through its Gcn5 subunit and a TBP chaperone function through its Spt3 subunit (Barbaric et al., 2003; Brownell et al., 1996, p. 5; Gregory et al., 1998). Together, SAGA and chromatin remodelers activate the *PHO5* promoter such that PIC formation can occur and Pol II can initiate transcription. In this example, nucleosomes play a direct role in repressing transcription by occluding the TATA-box and preventing association of TBP and subsequent PIC formation. In support of this model, Adkins and Tyler have shown that if histone chaperones involved in reassembling nucleosomes at the

PHO5 promoter are disrupted, even under repressive conditions where both Pho2 and Pho4 are displaced from the promoter, transcription continues (Adkins and Tyler, 2006). This data highlights the importance of DNA accessibility to transcription activation.

Histone Post-Translational Modification and Transcription Activation

Alternatively, nucleosomes also play a role in transcription activation and promoter specific transcription. Histones contain unstructured regions that protrude out of the nucleosome known as histone tails. These tails are heavily post-translationally modified (PTM) by methylation, phosphorylation, acylation (most commonly acetylation) and ubiquitylation, among other PTMs (Hottiger, 2015; Shilatifard, 2006). The dominant model for how these histone marks regulate transcription is that co-activator and co-repressor proteins recognize these marks and stimulate either PIC formation or chromatin condensation, respectively, to make gene promoters less accessible. For transcription activation, the +1 nucleosome (the nucleosome immediately downstream of the NFR) is a hub for post-translational modification (Figure 1.4). For example, Rad6/Bre1 ubiquitylates lysine 123 of H2B (H2Bub123) at the +1 nucleosome (Dover et al., 2002; Hwang et al., 2003; Sun and Allis, 2002; Wood et al., 2003). This H2Bub123 mark is required for the SET1/COMPASS complex to tri-methylate lysine 4 on histone H3 (Krogan et al., 2002; Miller et al., 2001) (H3K4me3; see Figure 1.4). The H3K4me3 mark is highly correlated with actively transcribing genes (Briggs et al., 2001; Roguev et al., 2001; Santos-Rosa et al., 2002; Strahl et al., 1999). Both the plant homeobox domain (PHD) finger in the TFIID subunit Taf3 as well as the Tudor domain of Sgf29 in SAGA can specifically bind to H3K4me3 (Bian et al., 2011; Vermeulen et al., 2007). Furthermore, active genes are highly

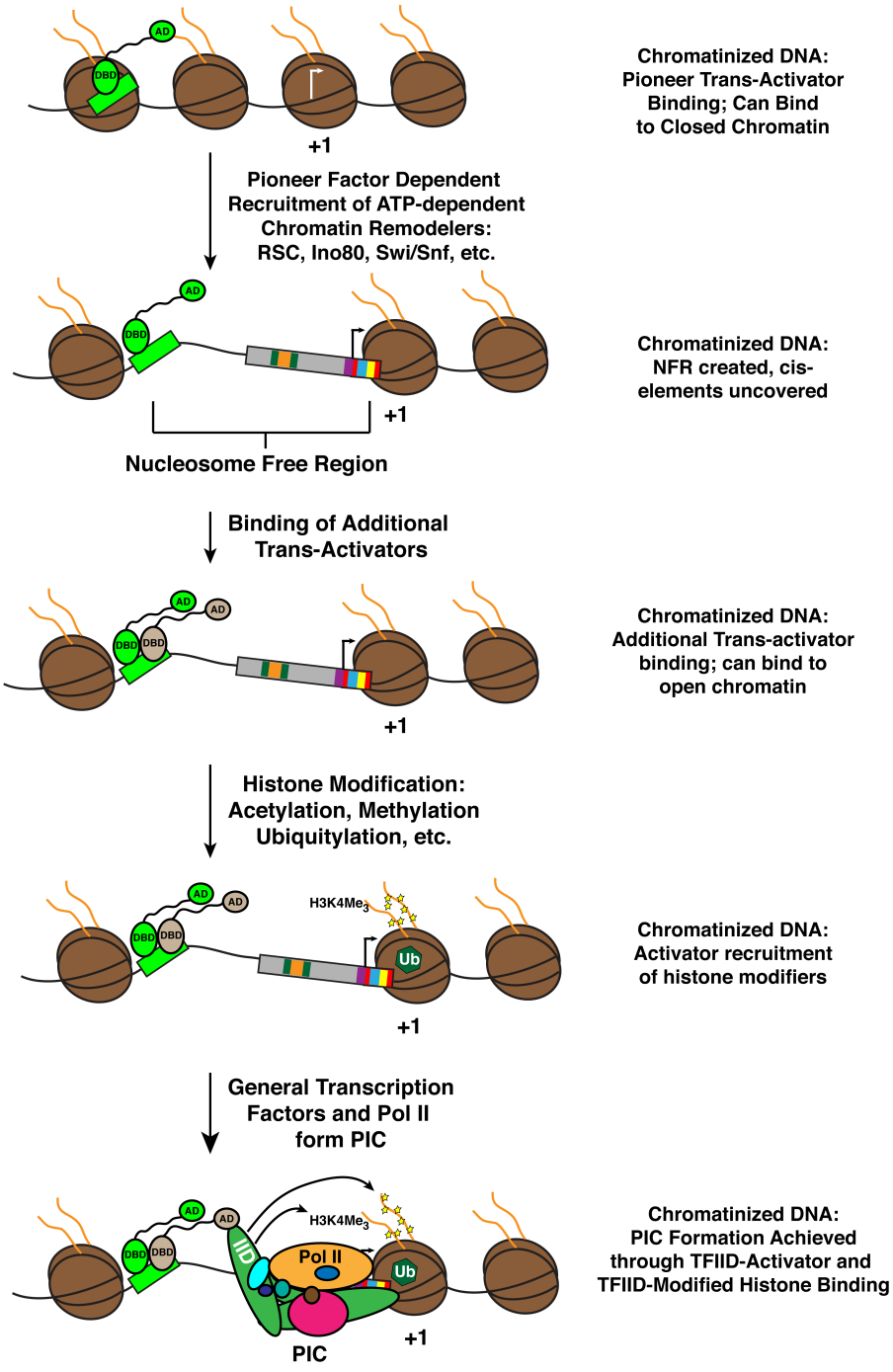


Figure 1.4: Model for Gene Activation in Chromatinized Environment. Assuming a ground state of chromatinized DNA, a pioneer transcription factor (in lime green) must associate with either nucleosomal or internucleosomal DNA to facilitate nucleosome remodeling by ATP-dependent chromatin remodelers. This remodeling forms a nucleosome free region (NFR) with a phased +1 nucleosome in close proximity to the transcription start site. Additional trans-activator proteins (grey) can associate with DNA and stimulate co-activator promoter association and histone modification. TFIID promoter binding is stimulated by access to the core promoter, interaction with activators and interaction with modified histones. Yellow Star=acetylation, Green hexagon=ubiquitylation, PIC Constituents: Royal Blue=TFIIF, Light Blue=TFIIB, Dark Blue=TBP, Aqua green=TFIIA, Pink=TFIIH, Brown= TFIIE

enriched in acetylation on histones H3 and H4. Three of the major acetyltransferases in yeast are Nucleosomal Acetyltransferase of Histone H3 (NuA3) (John et al., 2000), Nucleosomal Acetyltransferase of Histone H4 (NuA4) (Allard et al., 1999) and the Gcn5 acetyltransferase containing SAGA and transcriptional ADAptor (ADA) co-activatorcomplexes (Grant et al., 1997), all three of which play a role in stimulating transcription activation. One of the lysines Gcn5 acetylates is lysine 9 of histone H3 (H3K9ac), another mark that is localized to the 5' ends of genes and correlated with actively transcribed genes (Bonnet et al., 2014; Cieniewicz et al., 2014; Grant et al., 1999; Ruiz-García et al., 1997). The Taf1 subunit of TFIID contains a doublebromo domain that directly interacts with acetylated histones (Jacobson et al., 2000) and Taf14 has been shown to preferentially interact with acetylated histone H3 over unmodified histone H3 including H3K9ac (Shanle et al., 2015). Based on these biochemical activities, the simplest model is that both TFIID and SAGA can recognize histone marks, which results in stimulation of TBP-promoter binding and PIC formation.

Post-translational modification of histone tails can also repress spurious transcription (Figure 1.5). For example, in the wake of actively transcribing Pol II, intragenic nucleosomes are both acetylated (a mark of active transcription) as well as di- and tri-methylated on lysine 36 of H3 (H3K36me) by Set2 (Kizer et al., 2005; Krogan et al., 2003b; Strahl et al., 2002), which is bound to the Ser2-p form of the Rpb1 CTD (Li et al., 2003). Acetylated histones are readily removed and/or exchanged with other histones potentially resulting in free DNA and subsequent PIC formation at cryptic core promoters at the 3' end of genes (Schwabish and Struhl, 2004; Venkatesh et al., 2012). This H3K36me mark is recognized by the Rpd3S histone deacetylase complex, which acts to

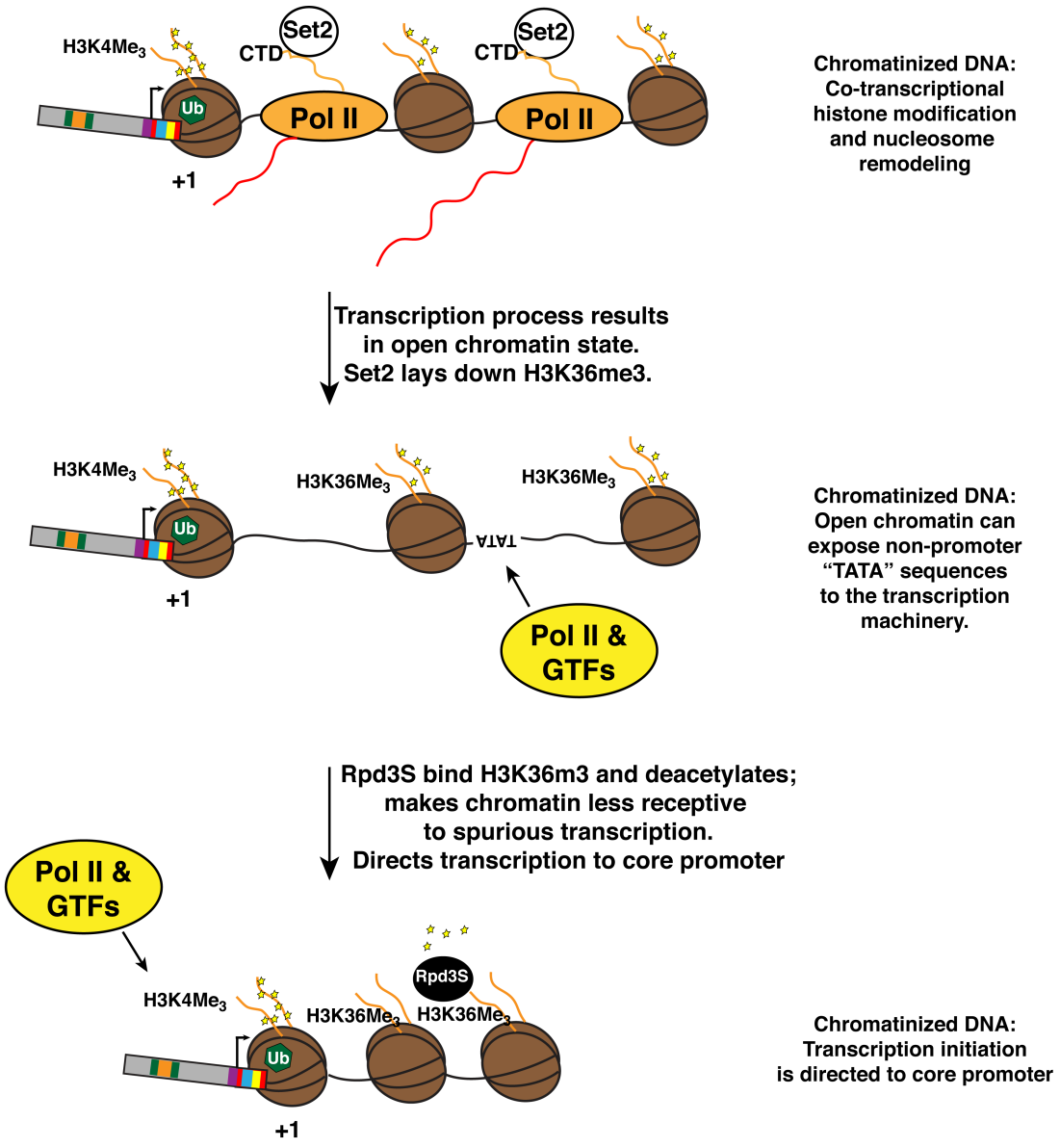


Figure 1.5: Model for Intragenic Transcription Suppression. Active transcription by RNA polymerase II results in nucleosome modification: either nucleosome eviction, nucleosome translation or co-transcriptional histone post-translational modification. Acetylation results in a more permissive transcription environment, likely through increased histone turn-over and DNA accessibility. However, Set2, associated with the Rpb1 CTD, di- and tri-methylates H3K36 during the process of transcription. This mark is recognized by the Rpd3S de-acetylation complex resulting in removal of intragenic acetylation marks. Along with histone deposition by histone chaperone proteins (not shown), this suppresses spurious transcription and favors transcription initiation from the core promoter.

remove the acetylation mark (Carrozza et al., 2005; Keogh et al., 2005). Functionally, this represses spurious initiation of transcription within intragenic regions. Thus, these observations have led to the model that, within the context of a transcribed gene, nucleosomes play an active repressive role in minimizing transcription initiation outside of the core promoter. However, nucleosome modifications such as acetylation and H3K4me3 at the +1 nucleosome stimulate PIC formation at the core promoter.

The Structure of the 15-Subunit TFIID Complex

As described above, TFIID is critical for regulating transcription of 90% of mRNA genes likely through its co-activator function as well as its ability to recognize both promoter DNA and active histone marks. As such, the elucidation of the structure and biochemical properties of TFIID will shed light on how this complex accomplishes its role in transcription regulation.

S. cerevisiae TFIID (scTFIID) is a 1.2 megadalton 15-subunit complex (Figure 1.5) (Sanders et al., 2002a). Prior to joining the lab, our lab in collaboration with the lab of Patrick Schultz from the IGBMC (Institut Génétique Biologie Moléculaire Cellulaire) in Strasbourg, France had described the structure of yeast TFIID using cryo-electron microscopy at approximately 25 Angstrom resolution (Papai et al., 2010). The shape of TFIID can be divided into 4 lobes: lobe A, lobe B, lobe C (containing C1 and C2 subdomains) and lobe D (Figure 1.6). These lobes form an asymmetric structure that could be described as a catcher's mitt. Lobe B forms the smaller thumb part of the mitt, lobe C1 and C2 form the base of the palm and lobe A and lobe D form the top part of the mitt where the four fingers would be located. The interface of these lobes is the cleft of the complex where the

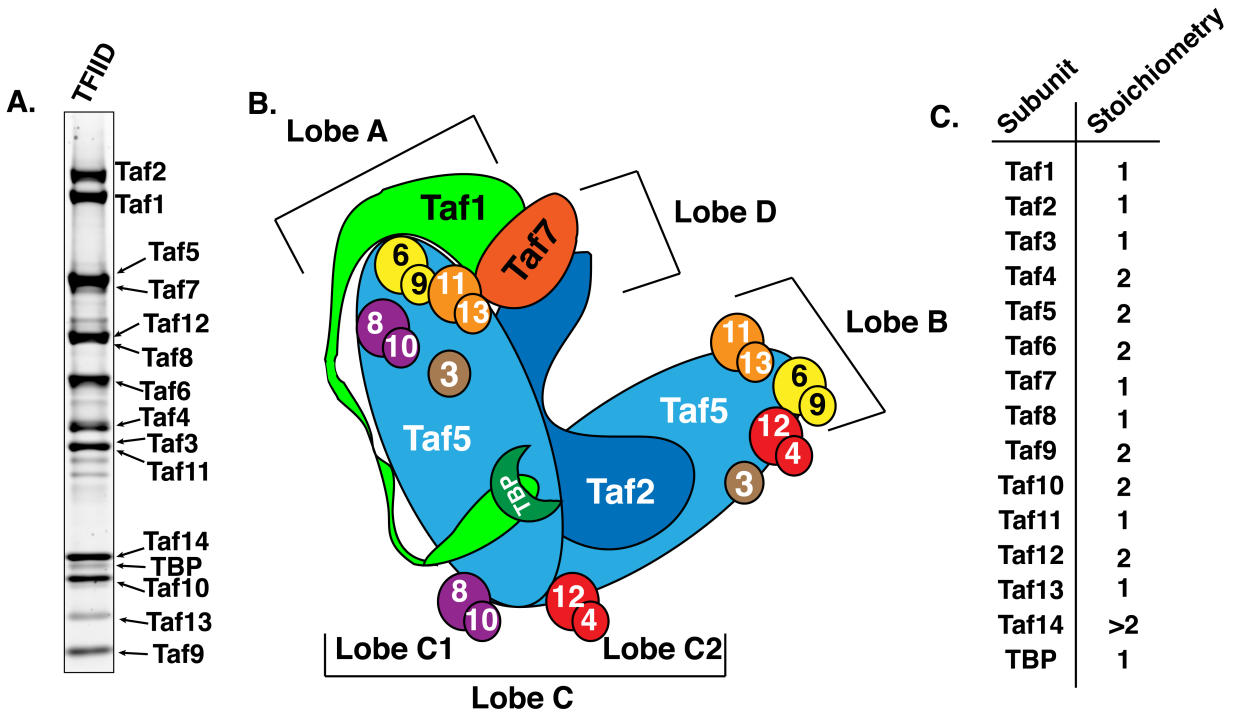


Figure 1.6: **TFIID Subunit Stoichiometry and Structure.** ***A.*** Purified TFIID separated on a SDS-PAGE gel. Subunits (Taf1→Taf14 and TBP) were visualized by Sypro Ruby gel stain. ***B.*** Schematic of TFIID structure based on immunolabeling and difference mapping. Note: Location of Taf14 within the complex is currently unknown. ***C.*** Stoichiometry of the subunits within the TFIID complex (mole subunit/mole TFIID). Values based on staining measurements and *in vivo* immunoprecipitation measurements.

DNA associates. Through this work, our lab along with the Schultz lab was able to identify the rough location of all the Tafs (with the exception of Taf14) using electron microscopy techniques coupled with immunolabeling and difference mapping (Leurent et al., 2004, 2002; Papai et al., 2009). Two copies of Taf5 form the scaffold of the molecule onto which other TFIID subunits assemble. The N-terminal domains (NTDs) of Taf5 are located in lobe C1 and C2 while the C-terminal WD40 motifs are located in both the A lobe and the B lobe. Lobe A also contains Taf1, Taf3, Taf6, Taf7, Taf8, Taf9, Taf10, Taf11 and Taf13. Lobe B also contains Taf3, Taf4, Taf12, Taf6, Taf9, Taf11 and Taf13. Lobe D contains the C-terminus of Taf2. The remaining Taf2 density extends throughout the central part of the molecule with its N-terminus positioned between lobe C2 and lobe B. The TAND of Taf1 extends from

lobe A to lobe C1 where it binds to TBP, which is located in the central cleft of the molecule. In addition to the Taf5 NTDs, lobe C1 and C2 also contain Taf4, Taf8, Taf10 and Taf12.

In general, the stoichiometry and biochemical binding data for the TFIID subunits are in agreement with the identified positions within the TFIID structure. Many of these subunits contain what is known as a histone-fold domain, which allows for specific heterodimerization among the TFIID subunits. The TFIID histone fold pairs include Taf3-Taf10, Taf4-12, Taf6-9, Taf8-10 and Taf11-13 (Birck et al., 1998; Gangloff et al., 2001b; Hisatake et al., 1995; Selleck et al., 2001; Werten et al., 2002). With the exception of Taf3 that localizes to lobe B, all of the histone fold domain-containing Tafs localize to lobes that contain their cognate binding partners. However, the stoichiometry of the complex and the locations for Taf3, Taf8, Taf11 and Taf13 subunits are inconsistent if the structure of TFIID is monolithic. First, Taf3, Taf8, Taf11 and Taf13 are present in one mole per mole of TFIID so they should not be present in multiple copies, as the immunolabeling might suggest. It is possible that Taf3 and the Taf11-Taf13 heterodimer can bind to both lobe A and lobe B while Taf8 can bind to both lobe A and lobe C. This possibility would result in TFIID assembly isomers that could explain the immunolabeling results. To more accurately determine the location of all of the TFIID subunits and if these isomers exist, we would need a TFIID cryoEM or x-ray crystallographic structure with improved resolution.

Recently, the Nogales lab published two cryoEM structures of human TFIID in complex with promoter DNA and TFIIA (Cianfrocco et al., 2013; Louder et al., 2016). In their first report, Nogales and colleagues suggested a lobe of TFIID is highly mobile, capable of moving from one side of the molecule to the other, further confounding our understanding of the TFIID structure. In their second report, Nogales et al. refined their

structure by docking existing crystal structure data or homology models (Taf2 in particular) into the structure and refining the amino acids to fit the densities observed in the cryoEM reconstruction. Using this method, they were able to resolve TBP, TFIIA, Taf2, Taf7 and parts of Taf1 and Taf6 at the atomic level. They also speculate about the position of Taf8, although the resolution of the structure makes it difficult to conclude with confidence that the density is truly Taf8. The rest of the structure is not at a high enough resolution to determine the location of the remaining subunits and, therefore, does not solve the discrepancy related to the position of Taf3, Taf8 and the Taf11-Taf13 heterodimer. Furthermore, Nogales and colleagues could not locate any of the histone fold domain-containing heterodimers in her structures and suggests that TFIID undergoes a significant rearrangement once either completely assembled or bound to promoter DNA/TFIIA.

In parallel with the Nogales work, the labs of Imre Berger, Lazlo Tora and Patrick Schultz began to reconstitute the human TFIID complex using purified Taf subunits (Bieniossek et al., 2013). They attempted to begin with Taf5 as the scaffold. However, Taf5 is insoluble when expressed in *E. coli* or *Sf9* cells. To overcome this hurdle, they discovered that Taf5, Taf6 and Taf9, when co-expressed, from a soluble and stable complex. They used this complex as the scaffold onto which they could assemble several TFIID subunits. First, Taf5-Taf6-Taf9 complex can associate with the Taf4-Taf12 heterodimer to form a 5-Taf complex. Consistent with this observation, using siRNA knockdown of Taf1, the Tjian lab was able to isolate a similar 5-Taf complex from drosophila S2 cells (Wright et al., 2006). The Schultz lab went on to determine the structure of this 5-Taf complex using cryoEM. This complex contains two copies of each subunit and is arranged as a symmetrical

complex, inconsistent with the asymmetric structure of TFIID. Berger and colleagues termed this 5-Taf complex as the TFIID core. Once the core was formed, Berger et al. could add a Taf8-Taf10 complex resulting in an 7-Taf assembly. Addition of this Taf8-Taf10 complex to the 5-Taf TFIID core complex breaks its symmetry such that the structure begins to resemble TFIID and can be docked into the yeast TFIID complex structure. As a result, Berger and Schultz have proposed that the TFIID assembly pathway is regulated by the association of this Taf8-Taf10 complex with the TFIID core and this break in symmetry is important for the final steps of TFIID assembly. These data were further supported by the identification of a cytoplasmic Taf2-Taf8-Taf10 subcomplex that is distinct from TFIID and whose entry into the nucleus is regulated by the Taf8 nuclear localization sequence and importin-α1 (Trowitzsch et al., 2015). This Taf2-Taf8-Taf10 complex can also associate and co-purify with the 5-Taf TFIID core suggesting that the Taf2-Taf8-Taf10 submodule may stimulate this conformational change and subsequent TFIID assembly *in vivo*. Additional TFIID structural intermediates generated *in vitro* and, if possible, *in vivo* would help to understand the TFIID assembly pathway.

When comparing the structures generated by our lab in collaboration with Patrick Schultz, the collaboration of Berger and Schultz, and the Nogales lab, there remain some discrepancies. Neither the 5-Taf TFIID core nor the 8-Taf complex could be docked into the Nogales human TFIID structure. In fact, in the Nogales structure, the subunits constituting the 5-Taf core appeared to be the most disordered region of her structure. Furthermore, Nogales and colleagues could not dock any of the crystal structures of the histone fold heterodimers. Even though they could identify the position of Taf6, she could not detect the Taf6/Taf9 heterodimer. This led Nogales et al. to speculate that TFIID

undergoes a dramatic rearrangement where the histone fold domain Tafs are separated into two distinct domains, one of which is exceptionally mobile. These results may reflect a rearrangement stemming from this break in asymmetry described by the work of the Berger and Schultz labs. As mentioned above, this could be consistent with observation Nogales and colleagues made that an entire domain of TFIID dissociates from one part of the complex and reassociates with another part of the complex. It is important to note that the labs of Weil and Schultz never observed these structural rearrangements in scTFIID, either by itself or in complex with TFIIA and promoter DNA. In addition, the Nogales TFIID model was not supported by either genetic (i.e. GFP or MBP tags or domain deletion variants) or biochemical (i.e. immunolabeling) labeling or chemical cross-linking/mass spectrometry techniques that could help to unambiguously identify the location of these subunits. While these new structural studies are exciting and could lead to important revelations about the subunit architecture of TFIID as well as provide some functional implications that could be dissected biochemically and genetically, our understanding of the TFIID structure still remains quite poor.

TFIID Functions as a Transcription Co-activator and Binds the Core Promoter

Functionally, the Tjian lab has been instrumental in establishing TFIID as a co-activator. In their early investigations, Tjian and colleagues formed TBP-Taf subcomplexes and used these complexes for *in vitro* transcription reactions and DNA binding studies with and without various trans-activator proteins. Using TFIID depleted, partially purified general transcription factors, Tjian et al. demonstrated that a TBP/Taf1/Taf2/Taf4 complex, in the presence of the trans-activator Sp1, could stimulate transcription from an

enhancer/promoter DNA containing Sp1 binding sites (Chen et al., 1994). This stimulation did not occur with TBP only, a TBP/Taf1/Taf2 complex or a TBP/Taf1/Taf4 complex. These workers then demonstrated that Sp1 directly interacts with Taf4. Using similar experimental paradigms, Tjian and colleagues have demonstrated direct interaction between the trans-activator NTF-1 and Tafs 2 and 6, the trans-activator p53 with Taf6 and Taf9 and the trans-activator c-Jun with Taf1 and Taf6 (Thut et al., 1995). These protein-protein interaction assays were complemented with *in vitro* transcription functional approaches showing activator-TFIID Taf dependent transcription stimulation. But, the Tjian lab did not follow up these results with iterative complementary genetic and biochemical dissection of these Taf-trans-activator interactions to determine how they contribute to gene regulatory mechanism *in vivo*. While the Tjian lab's work was performed with metazoan TFIID, the Weil lab has demonstrated that in *S. cerevisiae* Rap1 directly interacts with TFIID through Tafs 4, 5 and 12 supporting a role for TFIID as a co-activator across all eukaryotes (Garbett et al., 2007; Layer et al., 2010; Layer and Weil, 2013). Furthermore, the Weil lab demonstrated that the interaction between Rap1 and TFIID was sensitive to mutagenesis and important for the regulation of the ribosomal protein-encoding genes. Altogether these data implicate a direct interaction between trans-activator and TFIID as a mechanism for activated transcription. Moreover, multiple Tafs can directly interact with distinct trans-activators providing multiple hubs for trans-activator/co-activator communication (Chen et al., 1994; Hoey et al., 1993; Liu et al., 2009; Thut et al., 1995).

TFIID also plays a critical role in promoter recognition through direct interactions between TFIID Tafs and both core promoter DNA elements as well as modified

histones associated with active transcription (described in detail in the next section). The Smale lab first demonstrated that TFIID directly and functionally interacts with the initiator (INR) core promoter element (Kaufmann and Smale, 1994). Functional assays performed by Smale and colleagues focused on identifying a co-factor that could help stimulate transcription from the terminal deoxynucleotidyl transferase core promoter fragment engineered to contain a consensus TATA box and either a WT or a mutant INR element (Kaufmann et al., 1998, 1996). Impetus for these experiments derived from a failure in the field to observe robust INR-dependent stimulation of transcription with purified components when this stimulation had been observed using nuclear extracts (Smale and Baltimore, 1989). Ultimately, Smale and colleagues purified an activity that could stimulate transcription in an INR-dependent manner. He dubbed the protein that displayed this activity CIF150 and, surprisingly, this protein displayed similar reactivity to antibodies raised against drosophila Taf2 (Kaufmann et al., 1996). Concurrent with these studies, Tjian and Verrijzer demonstrated that Taf2 can directly bind to DNA and, using DNaseI footprinting, demonstrated that Taf2 binding resulted in significant hypersensitivity around the TSS and protection immediately downstream of the TSS, a position that is in close proximity to the INR (Verrijzer et al., 1995, 1994). Ultimately, Chalkley and Verrizjer advanced these results by performing systematic evolution of ligands by exponential enrichment (SELEX) with a Taf1/Taf2 complex and demonstrated that the preferred specificity was for the INR (Chalkley and Verrijzer, 1999). The Kadonaga lab performed similar footprinting analyses as well as protein-DNA cross-linking to identify that Tafs 6 and 9 directly interact with the DPE core promoter element (Burke and Kadonaga, 1997). Also, TFIID footprints, performed with either metazoan or yeast TFIID

using either DNaseI or hydroxyl radical DNA cleavage methods, span between 60bp to up to 100bp in length (Juven-Gershon et al., 2006; Sanders et al., 2002a). While metazoan TFIID footprints tend to display robust protection from DNA cleavage agents, scTFIID shows an increase in DNaseI hypersensitivity.

Functionally, the binding of TFIID to downstream promoter elements has raised concerns about how Pol II can traverse the DNA when TFIID may sterically hinder access to the TSS. From a Taf2 perspective, this concern is underscored by the fact that in metazoans, the initiator is also the site of transcription initiation, with the starting nucleotide residing directly within the Taf2 binding site. Two labs have provided evidence that TFIID structural rearrangement and dissociation from the INR is important for Pol II to clear the promoter and engage in productive transcription. First, cross-linking, footprinting and *in vitro* transcription studies by the Goodrich and Kugel labs suggested that following recruitment of Pol II to the PIC, TFIID undergoes a slow isomerization (Yakovchuk et al., 2010). This isomerization may result in a change in TFIID structure that is more permissive for Pol II to engage the DNA and transcribe the gene. Second, Tjian and colleagues identified a small molecule that targeted a predicted intrinsically disordered region (IDR) of the Taf2 C-terminus (Zhang et al., 2015). They utilized this inhibitor to interrogate the mechanism of Taf2 action using *in vitro* transcription as a read-out. The Tjian group observed that this inhibitor did not block TFIID association with promoter DNA, but instead it enhanced it. As a result, if the inhibitor was used prior to the first round of transcription (pre-Pol II addition), transcription was inhibited, presumably because TFIID could not undergo the isomerization step proposed by Yakovchuk et al. However, if the inhibitor was used after the first round of transcription, transcription re-initiation was

not impacted. Together, these biochemical studies have led to a putative model for Taf2 action whereby it is important for promoter recognition and Pol II association but then must dissociate from or alter its DNA contacts such that Pol II can escape the promoter and engage in productive transcription initiation. However, as of yet, there have been no studies to determine if this mechanism is operative and/or important *in vivo* because Taf2 has not been molecularly genetically dissected in any system.

Differences between Metazoan and Yeast TFIID

Despite the functional similarities between metazoan and scTFIID described above, the TFIID subunits do display significant differences. A comparison of the scTFIID and the human TFIID subunits is provided in Figure 1.7 (Bai et al., 1997; Birck et al., 1998; Durso et al., 2001; Gangloff et al., 2001b; Hisatake et al., 1995; Klebanow et al., 1997; Matangkasombut et al., 2000; Papai et al., 2009; Scheer et al., 2012; Werten et al., 2002).. Of note, several functional domains that exist in human TFIID Tafs are not conserved in the scTFIID homologs. For example, while human Taf3 contains a PHD finger involved in H3K4me3 binding (described above), scTaf3 does not contain this domain (Gangloff et al., 2001a). In addition, human Taf1 is a 213kDa protein that, in addition to structural domains required for Taf1-Taf and Taf1-TBP interactions, contains two kinase domains at its N- and C-termini, an acetyltransferase domain and a doublebromo domain, which binds to acetylated histones (Bai et al., 1997; Dikstein et al., 1996; Mizzen et al., 1996; Poon et al., 1995). scTaf1 is only a 121kDa protein. Sequence alignments have demonstrated scTaf1 corresponds to the N-terminal half of hTaf1. Thus, based on homology, scTaf1 contains only one of the kinase domains, which has never been shown to possess phosphorylation

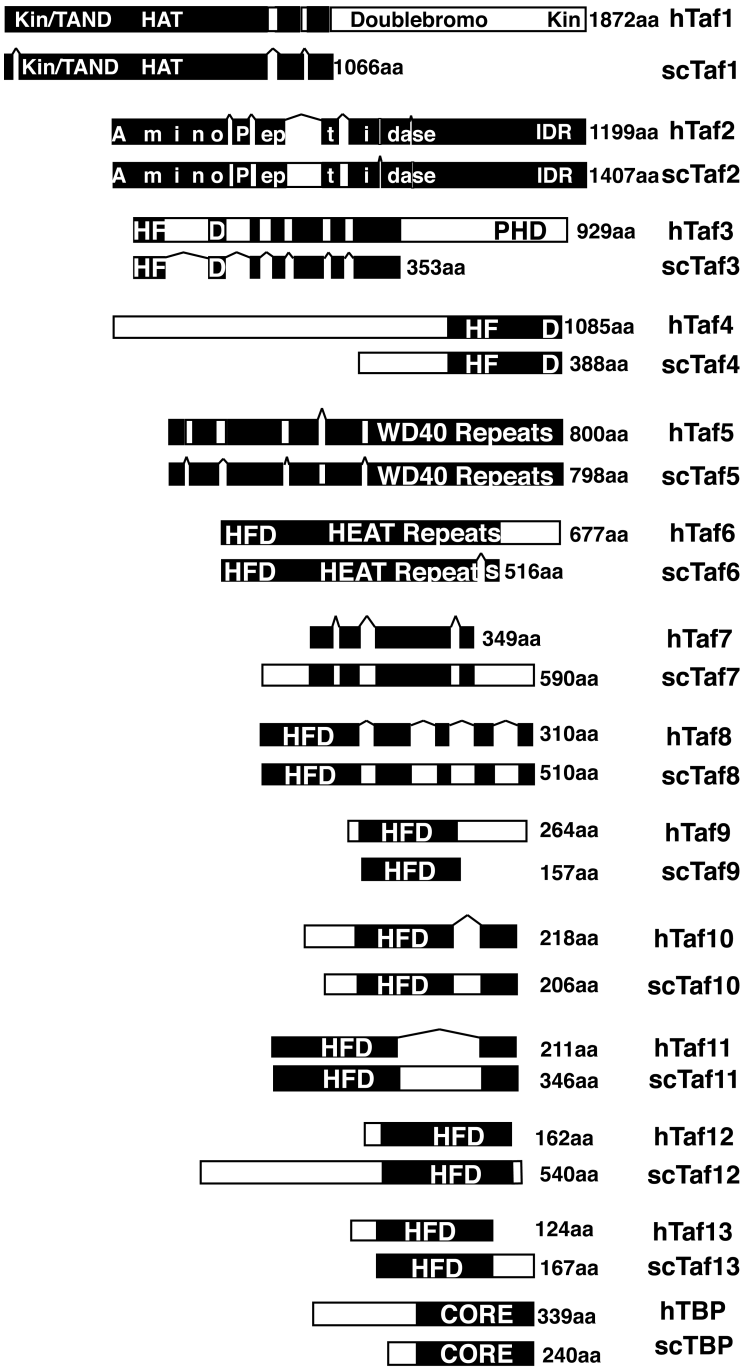


Figure 1.7: Conservation and Domain Structure of the Evolutionarily Conserved TFIID Subunits. Schematics of the primary structures of TFIID subunits for human (h) and *Saccharomyces cerevisiae* (sc). Black indicates regions of homology. White boxes or empty regions indicate regions of dissimilar or missing sequences, respectively. Schematics derived from alignments of primary amino acid sequence using ClustalW analysis in MacVector. Domains are labeled. Kin = Kinase, TAND = Taf1 N-Terminal Domain, HAT = histone acetyltransferase, IDR = intrinsically disordered region, WD40 repeats = 40 amino acid stretches with characteristic Tryptophan-Aspartic Acid repeats, HEAT repeats = domains common to Huntingtin, Elongation factor 3, protein phosphatase 2A and Tor1, HFD = histone fold domain. CORE = highly conserved (~90%) TBP C-terminal sequence. Sequence information derived from UniProt.

activity, and the acetyltransferase domain, which only possess weak acetyltransferase activity. To account for this missing doublebromo domain, Buratowski and colleagues has suggested that the TFIID accessory protein Bdf1, which contains a doublebromo domain that binds to acetylated H3 and H4 *in vivo* (Ladurner et al., 2003), is the missing C-terminal piece of Taf1 (Matangkasombut et al., 2000). Furthermore, genome-wide chromatin precipitation (ChIP) analyses have demonstrated that Bdf1 and Taf1 share similar occupancy profiles on gene promoters (Rhee and Pugh, 2012). Still, if Bdf1 does in fact functionally correspond to the missing Taf1 C-terminus, then the question remains why human evolution favored a Taf1 protein containing a covalently fused doublebromo domain while *S. cerevisiae* favored a Taf1 protein that weakly associates with Bdf1. There are two likely scenarios that could explain this phenomenon. First, *S. cerevisiae* utilizes a mechanism not required in humans to modulate the interaction between Taf1 and Bdf1 which could regulate transcription activation potential. Second, in *S. cerevisiae* Bdf1 is required in multiple transcription related complexes while in humans its particular function is only required in the Taf1 C-terminus. Both scenarios are supported by experimental evidence including that Bdf1 is subject to phosphorylation (Sawa et al., 2004), Bdf1 is also associated with the SWR-C complex (Krogan et al., 2003a) and Bdf1 can associate with chromatin independent of TFIID (Ladurner et al., 2003).

The other major difference between scTFIID and its metazoan homologs is that scTFIID contains a fifteenth subunit, Taf14 (Henry et al., 1994). In *S. cerevisiae,* Taf14 is a stoichiometric member or associated protein for at least 7 different transcription related complexes: the general transcription factors TFIID and TFIIF (Henry et al., 1994; Kabani et al., 2005; Sanders et al., 2002b); the Mediator co-activator complex (Henry et al., 1994; Lim

et al., 2007; Sanders et al., 2002b); the ATP-dependent chromatin remodeling complexes Swi/Snf (Cairns et al., 1996; Sanders et al., 2002b), Ino80 (Kabani et al., 2005; Sanders et al., 2002b) and RSC (Kabani et al., 2005); and the histone acetylation complex NuA3 (John et al., 2000; Kabani et al., 2005; Sanders et al., 2002b). Until recently, the role Taf14 played in these complexes was largely unknown. *TAF14* is not an essential gene although strains that are *taf14* null display slow growth and a temperature sensitive phenotype. The N-terminus of Taf14 contains an evolutionarily conserved YEATS domain, named after the proteins that contain this domain: Yaf9, ENL, AF9, Taf14, Sas5 (Schulze et al., 2009). However, this N-terminal YEATS domain is dispensable for growth with minimal impact on growth rate or thermoresistance (Schulze et al., 2010). The Taf14 C-terminus is required for interaction with the transcription related complexes listed above. Deletion of the Taf14 C-terminus phenocopies a *taf14* null strain. Together, these data suggest that the C-terminal domain plays a critical role in the function of these transcription related complexes that is distinct from the function of the YEATS domain. Still, the role of the YEATS domain in transcription is likely conserved between *S. cerevisiae* and humans considering both ENL and AF9 are members of the human super elongation complex (Lin et al., 2010).

In the last two years, scientific interest in YEATS domain containing proteins has increased because both Taf14 and AF9 were shown to specifically bind to the histone tail of acetylated lysine 9 of H3 (H3K9ac), a histone modification that is highly correlated with the promoters of actively transcribing genes (Li et al., 2014; Shanle et al., 2015). Further structural and biochemical analyses demonstrated that the YEATS domains for Taf14 and AF9 contain unusually large hydrophobic/aromatic binding pockets that can accommodate

larger histone acylation marks including propionylation, butyrylation and crotonylation (Andrews et al., 2016; Y. Li et al., 2016; Tan et al., 2011; Zhang et al., 2016). In fact, the binding affinity of Taf14 and AF9 to crotonyl-lysine (Kd=10μM) is substantially higher than acetyl lysine (Kd=150μM). Similar to H3K9ac, crotonylated lysine 9 of H3 (H3K9cr) also correlates with promoters of actively transcribing genes. Altogether, these observations suggest that one important role of Taf14 in transcription regulation is its ability to bind modified chromatin allowing the complexes with which it associates to regulate transcription. In support of this hypothesis, Taf14 mutants that are defective in H3K9ac and H3K9cr binding display defects in steady-state RNA transcripts for DNA repair pathways and cell cycle signaling pathways. Still, two issues remain with regard to understanding the role of Taf14 in transcription regulation. First, considering Taf14 is a stoichiometric member or associated protein of multiple transcription related complexes, it is impossible to know how a mutation in the YEATS domain of Taf14 specifically impacts the function of any one of those complexes. Second, considering the Taf14 YEATS domain is dispensable for normal growth, what is the function of the Taf14 C-terminus? Does it mediate interactions between the different proteins within each complex (intra-complex interactions), bridge interactions between different proteins in different complexes (inter-complex interactions) or serve some other function such as stimulate proper folding or mask domains that are inhibitory if exposed in the wrong cellular context. Ultimately, despite these exciting advancements in our understanding of Taf14 function, its role in *S. cerevisiae* remains quite poorly understood.

While differences in TFIID function have not been identified with regard to the ability of specific Tafs to bind to DNA, the core promoter architecture is significantly

different between *S. cerevisiae* and metazoans. As described above, metazoan core promoters contain several elements that are directly bound by TFIID subunits, functionally stimulate transcription in a TFIID-dependent manner or both. Those elements include the BREu, BREd, INR, the DPE, the downstream core element (DCE) and the motif 10 element (MTE) (Figures 1.2 and 1.3)(Lewis et al., 2000). None of these elements have been unambiguously identified in *S. cerevisiae*. In addition, studies that demonstrate that yeast Tafs bind DNA do not show that these Tafs bind with any sequence specificity (Shao et al., 2005). Given that ~80% of *S. cerevisiae* genes do not contain a consensus TATA box, one must question how scTFIID specifically binds the core promoter. Of course, scTFIID Tafs could bind to the core promoter in a sequence specific manner but the cognate cis-elements are too degenerate to identify using bioinformatic methods. Otherwise, there exist two likely, potentially interdependent mechanisms that could answer this conundrum.

First, TFIID promoter occupancy could be directed by its co-activator function, meaning that TFIID interaction with enhancer-bound transactivators could direct TFIID to TFIID-dominated gene core promoters. In support of this hypothesis, Struhl and colleagues have demonstrated TATA-containing, SAGA-dominated enhancer/promoters can be converted to TFIID-dominated enhancer/promoters by fusing the enhancer of a TFIID-dominated gene to the core promoter from a SAGA-dominated gene (Mencía et al., 2002). The Green lab has made similar observations (Li et al., 2002). These data suggest that the core promoter does not necessarily dictate whether a gene is TFIID-dominated or SAGA-dominated in *S. cerevisiae*. Furthermore, our lab has published structural data where TFIID, when in complex with enhancer/promoter DNA (both the enhancer/promoter fragments used by Struhl and colleagues as well as from the native *RPS1A* gene), the trans-

activator Rap1 and TFIIA, appears to make direct contacts with TFIID along the cleft of the molecule near TBP and in lobe D near the Taf2 C-terminus. On the other hand, the Ptashne lab came to the opposite conclusion. Ptashne and colleagues suggest that activators can stimulate the association of both SAGA and TFIID to promoters but the promoter is what dictates which co-activator is required for activation (Cheng et al., 2002). Thus, the roles enhancers and promoters play in coordinating activator-co-activator communication are still poorly understood.

Second, TFIID promoter binding could result from the ability of TFIID to interact with the +1 nucleosome. Rhee and Pugh have shown that the distance from the location of the PIC to the +1 nucleosome is much more uniform across TFIID-dominated genes and much less uniform across SAGA-dominated genes. They speculate that TFIID acts like a ruler, measuring the distance between its interaction with the +1 nucleosome on one side of the molecule and where the TATA-box (either consensus or non-consensus) is positioned on the other side of the molecule. In this second model, TFIID-chromatin binding activities are essential. Thus, Taf14-H3K9ac or Taf14-H3K9cr interactions and Taf1-Bdf1 interaction would be particularly important for TFIID promoter binding function. In both cases, TFIID can still interact with DNA through TBP, non-specific or low specificity Taf-DNA interactions or both; however, the ability of TFIID to specifically associate with the promoter is directed by activator and/or +1 nucleosome interactions and not by TBP-TATA binding. In fact, both models are consistent with the observation that TATA-like sequences are present in essentially all *S. cerevisiae* genes (Rhee and Pugh, 2012) because TFIID-Tafs could act to position TBP in close proximity to low affinity TATA-like sequences.

Alternatively, it is possible that TBP-DNA binding, in the context of TFIID, is not required for TFIID-dependent transcription. Two lines of evidence support this hypothesis. Multiple groups have shown that in strains containing a *tbp*^{ts} allele, following temperature shift, Taf occupancy for TFIID-dominated genes does not change while both TBP occupancy and Pol II occupancy is markedly decreased (Li et al., 2000; Mencía et al., 2002, p. 1). Such data demonstrate that Taf stability on genes (off rate) is not significantly changed upon TBP inactivation but it does not show that TFIID can associate with gene promoters in the absence of TBP. Hahn and colleagues attempted to address this problem, to a certain extent, by identifying TBP variants that displayed a reduction in the ability to bind DNA and failed to support SAGA-dominated gene transcription but did not display a significant loss in TFIID-dominated gene transcription (Kamenova et al., 2014). Using TBP-inactivated transcription extracts, addition of these TBP variants could reconstitute *in vitro* transcription of a *RPS5* reporter gene (does not contain a consensus TATA) but could not fully reconstitute *in vitro* transcription of a *HIS4* reporter gene (does contain a consensus TATA). These results suggest that TBP binding is not required for transcription of the TFIID-dependent *RPS5* gene and could support an increased role for TFIID-DNA binding in *RPS5* gene transcription. However, Hahn et al. also states that he observed no TFIID-specific DNA binding, which directly contradicts DNaseI footprinting data published by our laboratory (Sanders et al., 2002a).

Ultimately, the preponderance of data supports a role for both TBP-DNA binding, even to non-consensus TATA sequences, as well as TFIID-Taf DNA binding in promoter recognition. Nonetheless, the field will have to seriously address the Hahn lab's data to understand the role of TBP and Tafs in TFIID-dependent transcription and promoter

recognition. Importantly, we need to generate separation-of-function variants in Tafs that specifically disrupt Taf-DNA binding or Taf-chromatin binding without impacting other essential Taf functions. With these reagents, we can determine the role of Taf-promoter interactions on TFIID's functions including promoter recognition (i.e. DNA binding *in vitro* and promoter occupancy *in vivo*) and transcription activation activity (i.e. RNA output both *in vivo* and *in vitro*).

The Role of Taf2, a Putative INR Binding Protein, in TFIID Biology

As described above, Taf2 appears to play a significant role in TFIID-promoter recognition based on its ability to bind to the metazoan INR element. However, other than *in vitro* biochemical data demonstrating Taf2 DNA binding activity, we have minimal understanding of how Taf2 functions in the context of TFIID. In fact, structure-function analyses of Taf2 have never been performed in any system so we do not even understand the domain structure of this critically important protein. Based on this dearth of information, I decided to focus my dissertation research on the function of Taf2 in *S. cerevisiae*.

While we do not know much about Taf2 function, several labs have published descriptive findings regarding Taf2 in *S. cerevisiae*. To identify important protein-protein interactions, two methods are commonly used in *S. cerevisiae*. First, 2-hybrid genetic screens are commonly used. Two studies that utilized 2-hybrid screens have identified potential Taf2-TFIID subunit interaction in yeast. The first study described a genome-wide 2-hybrid screen to identify proteins that could interact with Taf14 (Kabani et al., 2005). In this study, while no data were shown, Kabani et al. reported that a Taf2 aa 1313-1407

fragment interacts with Taf14. The second study, performed by Stargell and colleagues, used a directed 2-hybrid screen approach where each TFIID subunit was screened against every other TFIID subunit to create a TFIID interaction map (Yatherajam et al., 2003). This study found that a Taf2-Gal4$_{DBD}$ fusion interacted strongly with a Taf4-Gal4$_{AD}$ fusion and interacted weakly with Taf3-Gal4$_{AD}$ fusion and a Taf8-Gal4$_{AD}$ fusion. Alternatively, a Taf2-Gal4$_{AD}$ fusion only interacted weakly with Taf7-Gal4$_{DBD}$ and Taf10-Gal4$_{DBD}$ fusions. Importantly, Taf14 was not assayed in this study. While not performed with yeast Tafs, an *Arabidopsis* TFIID subunit directed 2-hybrid screen showed that a Taf2-Gal4$_{DBD}$ fusion only weakly interacted with a Taf1-Gal4$_{AD}$ fusion and found no observable interaction with any other Tafs (Lawit et al., 2007). The second method commonly used to identify interacting proteins is to purify the protein in question along with any other associated proteins and perform proteomics analysis. The Weil lab created affinity purified antibodies against every TFIID subunit and used these antibodies to immunopurify each protein from a partially purified chromatographic fraction (Sanders et al., 2002b). Immunopurification of Taf2 of course purified TFIID subunits but also specifically purified the cleavage and polyadenylation specificity factor (CPSF). None of the other immunopurifications showed enrichment in CPSF suggesting that Taf2 may directly interact with CPSF or perform other functions outside of TFIID in the CPSF complex. One other study indicated that CPSF interacts with TFIID (Dantonel et al., 1997), so, taken together, it is possible that CPSF interacts with TFIID through Taf2. Considering our interest in the Weil lab is on transcription initiation mechanism, the interaction between Taf2 and CPSF was not further explored.

At the start of this project, the existence of a Taf2-Taf8-Taf10 complex in humans had not yet been discovered. However, the Weil lab had generated some data to suggest that Taf2 and Taf8 likely form a subcomplex in *S. cerevisiae*. Post-doc Madhu Singh used immunoprecipitation of deletion variants of Taf1 to identify the domains involved with Taf1-TFIID subunit interactions (Singh et al., 2004). He found that three deletion variants in Taf1 (aa deletions: Δ300-367, Δ365-435 and Δ430-495) could co-precipitate all of the TFIID subunits as efficiently as WT Taf1 with the exception of Taf2, Taf8 and Taf14. Surprisingly, Taf10 did not show a defect in co-precipitation in these Taf1 deletion variants. Similarly, Post-doc Manish Tripathi purified TFIID using a Taf1-TAP tag protocol. This Taf1-TAP purified TFIID consistently resulted in substoichiometric levels of Taf2 in the complex (Papai et al., 2009). These preparations were used to identify the location of Taf2 in the TFIID complex, as described above. However, upon closer inspection, both Taf8 and Taf14 were substoichiometric in these TFIID preparations while Taf10 levels were unchanged. These data suggest that instead of the Taf2-Taf8-Taf10 subcomplex that exists in metazoans, *S. cerevisiae* may have a Taf2-Taf8-Taf14 complex. This hypothesis is consistent with both biochemical data in metazoans which shows that Taf2 directly interacts with Taf8 and not Taf10 and genetic data in yeast which shows that the histone fold domain of Taf8 is not essential for life whereas the C-terminal portion of Taf8 (the domain for Taf2-Taf8 interaction in metazoans) is essential for life (Gangloff et al., 2001b).

The role of Taf2 in regulating TFIID-dependent transcription regulation in *S. cerevisiae* is even less well understood. As mentioned above, *TAF2* is a single copy essential gene (Ray et al., 1991). Thus, the best way to study essential gene function is through conditional loss-of-function methods. While in recent years new, less invasive techniques

have been developed such as Anchor Away (Haruki et al., 2008) and the auxin inducible degron (Nishimura et al., 2009), historically yeast biologists have used Ts variants, which can be abruptly inactivated following shift to the non-permissive temperature. Green and colleagues created Ts alleles for all of the TFIID Tafs, performed acute heat inactivation and used microarrays to determine which genes were misregulated upon acute inactivation of the TFIID subunit (Shen et al., 2003). The results for Taf2 were not particularly informative for a few reasons. First, the time it took to achieve growth arrest for the strain containing a *taf2*^{ts} allele was the slowest of all the strains harboring Taf Ts alleles suggesting that acute heat inactivation was the least efficient for the *taf2*^{ts} allele harboring strain. Based on this result, not surprisingly, Green and colleagues' results suggested that Taf2 only contributed to the regulation of 3% of yeast genes whereas another TFIID-specific subunit, Taf8, contributed to the regulation of 51% of yeast genes. Considering Taf2 and Taf8 likely interact within the TFIID complex, it is highly unlikely that Taf2 only contributes to 3% of yeast genes. In addition, the study that generated this *taf2*^{ts} allele did not work to determine the mechanism through which this allele disrupted Taf2 function. Thus, it is possible that this particular *taf2*^{ts} allele does result in a disregulation of 3% of yeast genes because a certain function that is essential for the regulation of those genes is inactivated whereas the rest of Taf2 function remains intact. The other data point that could illuminate Taf2 function was derived from a ChIP-microarray study using a *taf1*^{ts} allele (Ohtsuki et al., 2010). In this study, Kokubo and et al. acutely heat inactivated strains carrying either *TAF1* or a *taf1*^{ts} allele and then perform ChIP-array on every TFIID subunit. They found that on certain genes, upon gene heat inactivation, strains carrying the *taf1*^{ts} allele displayed significant reductions in occupancy of Taf1 and the rest of the TFIID

specific Tafs when compared to the *TAF1* strain, with the exception of Taf2. Indeed, when comparing the two strains, Taf2 occupancy did not change or increased. The result could indicate that Taf2 can stably associate with the promoter in the absence of Taf1 function whereas the rest of the TFIID subunits cannot. While this is consistent with Taf2 having a DNA binding activity, on its own this result is merely suggestive.

Another strong impetus for investigating Taf2 came from an observation the Weil lab made when describing the structure of the transactivator Rap1 in complex with enhancer/promoter DNA, TFIIA and TFIID (Papai et al., 2010). This structure implicated Taf2 in direct Taf2-promoter DNA binding to a sequence downstream of the TATA-box. This result suggests that scTaf2 may also have DNA binding activity similar to the metazoan Taf2-INR binding function.

Specific Aims

Considering Taf2 had never been subjected to structure-function analysis, I believed that a systematic molecular genetic dissection of *TAF2* would be an important contribution to our understanding of TFIID biology. This project had the promise to define the domain structure of Taf2, which of those domains are essential for viability, which other TFIID subunits Taf2 binds in the complex and if those interactions are necessary for Taf2 function and the importance of Taf2-DNA binding to TFIID function.

To address these goals, I proposed two specific aims:

Aim 1: Develop methods to purify Taf2 mutant forms, both as individual polypeptides as well as in the context of TFIID.

Aim 2: Perform a molecular genetic dissection of Taf2 to identify domains or individual amino acids that are critical for Taf2 function *in vivo* and *in vitro*.

The results of the studies to achieve these aims are documented in Chapters 2 and 3 of this dissertation.

CHAPTER II

METHOD DEVELOPMENT FOR THE PURIFICATION OF WILD-TYPE AND MUTANT TAF2 AS BOTH A SOLUBLE PROTEIN AND WITHIN THE TFIID COMPLEX

Current Requirements and Difficulties Involved with Purification of Taf2 and TFIID

Considering Taf2 is a component of the multi-subunit TFIID complex, the simplest and most straightforward method of deciphering the biochemical activities of Taf2 vis-à-vis Taf2-DNA interactions or Taf2-protein interactions is by using purified proteins. These studies can be performed with holoTFIID using protein-DNA label transfer approaches, as performed by the Buratowski laboratory (Auty et al., 2004), or through chemical cross-linking/mass-spectrometry. While crude extracts containing exogenously expressed Taf2 without any other TFIID subunits could also be employed, the most versatile and most easily interpretable approach is to work with purified Taf2.

When I entered the Weil laboratory, no protocols existed for the purification of full length soluble Taf2. As described above, Taf2 functions involve both DNA binding and Taf-Taf interactions. When generating Taf2 antigen for anti-Taf2 antibodies, previous Weil lab members collaborated to make His_6-$TEV_{cleavage\ site}$-Taf2 baculovirus. They used this virus to infect High Five cells (*Trichoplusia ni*) and express Taf2. When they generated Taf2 using this method, full length Taf2 accumulated at the lowest level of any other Taf expressed in this manner. Furthermore, they used denaturing methods to purify the protein so I had no information with regard to the solubility of the generated protein. Considering these experiments were performed nearly 10 years prior to when I joined the lab, I decided that I

should attempt all methods at my disposal to generate purified Taf2 including expressing the protein and/or protein fragments using *Escherichia coli*, baculoviral infection or galactose (Gal) induction methods in *S. cerevisiae*.

In addition, to truly study Taf2 function in transcription mechanism, Taf2 should be studied in the context of TFIID. For example, if I were to generate a Taf2 variant with a documented separation-of-function molecular phenotype such as a loss of DNA binding activity or a loss in the ability to interact with one TFIID subunit but not with the rest of the complex, then I should be able to purify Taf2-mutant TFIID and assay its function using *in vitro* transcription. However, generating purified TFIID remains a significant hurdle. Current literature suggests that TFIID is the most difficult general transcription factor to purify. Specifically, tandem affinity purification (TAP)-tagging purification procedures have increased the accessibility and efficiency of purification of yeast multi-subunit complexes (Rigaut et al., 1999). RNA polymerase II is highly abundant and easily purified using the TAP-tag procedure (Borggrefe et al., 2001). Kornberg and colleagues have developed genetic and biochemical methods to purify milligram quantities of TFIIH and TFIIF, respectively (Murakami et al., 2012). Recombinant TFIIA, TFIIB and TFIIE can all be expressed and purified from *E. coli*. But, the standard TAP-purification method has not proven effective for TFIID because the resulting material is substoichiometric for Taf2, Taf8 and Taf14. In addition, using Weil lab published HA-immunopurification methods to purify TFIID results in a yield of ~1mg/2.5kg of yeast pellet. Purifying TFIIH, the next most difficult general transcription factor to purify, with the TAP-tag method yields ~8mg/2.5kg of yeast pellet, an 8-fold difference. In addition, the financial cost of the TFIID HA-immunopurification procedure is considerably more expensive than the TAP-tag

purification procedure. To overcome both the financial and yield difficulties involved with the HA-immunopurification procedure, I employed two distinct strategies: 1) reconstitute TFIID using recombinant protein expression technologies and 2) improve the TAP-tag purification protocol to eliminate the purification step that results in disruption of Taf2, Taf8 and Taf14 from the complex.

METHODS

Expression Vector Design and Cloning

All cloning was performed using restriction enzyme-based methods. Appropriate restriction ends were added to all cloned sequences using PCR with either *Pyrococcus woesei* DNA polymerase as described (Dabrowski and Kur, 1998) or Q5 DNA polymerase according to manufacturers instructions (NEB). Individual details for each cloned fragment will be described below. Both plasmid DNA and PCR-generated DNA fragments were digested with the appropriate restriction enzymes, dephosphorylated with Antarctic phosphatase (NEB) and gel purified using Qiagen Gel Extraction buffers with either Qiagen gel extraction columns or Denville Scientific gel extraction columns, according to manufacturers instructions. The digested plasmid DNA was ligated to the digested PCR product in a 5μL reaction containing approximately a 3 to 1 molar ratio of PCR product to plasmid DNA. Total mass of plasmid DNA was ~10-20ng per reaction with varying amounts of PCR product depending on the size of the fragment. In addition, this reaction contained 1x T4 DNA ligase buffer and 100units of T4 DNA ligase (NEB). The ligation reactions were allowed to proceed for 2hrs to overnight at room temperature

(~20°) prior to transformation into competent *E. coli* (cell manipulations described below). Isolated recombinant expression plasmids were analyzed by restriction digestion to confirm the presence of the correct sized inserts. Plasmids containing the correct sized inserts were sequence verified.

For expression in *E. coli*, all Taf2 variants were expressed as N-terminal His_{x6}-tagged proteins using pET28a (Novagen). Full length and *TAF2* fragments were all engineered to contain an in-frame 5' SpeI restriction enzyme cut site and SalI cut site at the 3' end. Six Taf2 protein fragments were generated: full length Taf2 (aa 1-1407), Taf2 aa 1-407, Taf2 aa 1-1007, Taf2 aa 401-1007, Taf2 aa 401-1407 and Taf2 aa 1001-1407. The multiple cloning site (MCS) of pET28a contained the restriction enzyme cut sites for NheI and XhoI. NheI is in-frame with the N-terminal His_6-tag of pET28a and the NheI digested overhang is compatible for the digested overhang for a SpeI digested fragment. Similarly, a XhoI digested overhang is compatible with the SalI digested overhang.

For expression using baculovirus infected *Sf9* cells, all of the TFIID subunits were individually cloned into pFASTBac1 or pFASTBAC HTc (Invitrogen) as untagged expression cassettes, N-terminally tagged HA_3-PreScission site ($3C_{pro}$)- expression cassettes or N-terminally tagged $FLAG_3$-TEV protease cleavage site (TEV_{pro})- expression cassettes. The 3Cpro site and the TEVpro site are protease recognition sequences for the 3C and TEV proteases, respectively. These enzymes can be used to liberate the tag from the purified protein. All of the TFIID subunits ORF sequences, with the exception of *TAF2* which was subcloned as a 5'-SpeI-*TAF2*-SalI-3' fragment as described above, were synthesized using PCR from either existing plasmids containing the TFIID subunit sequence or from yeast genomic DNA. The restriction enzymes cut sites for each of the TFIID subunits are as

follows: 5'-SpeI-*TAF1*-XhoI-3', 5'-NheI-*TAF3*-KpnI-3', 5'-SpeI-*TAF4*-HindIII-3', 5'-*SpeI*-TAF5-Xho-3', 5'-SpeI-*TAF6*-HindIII-3', 5'-SpeI-*TAF7*-XhoI-3', 5'-NheI-*TAF8*-SphI-3', 5'-NheI-*TAF9*-KpnI-3', 5'-SpeI-*TAF10*-HindIII-3', 5'-SpeI-*TAF11*-XhoI-3', 5'-NheI-*TAF12*-KpnI-3', 5'-NheI-*TAF13*-KpnI-3', 5'-SpeI-*TAF14*-XhoI-3' and 5'-SpeI-*SPT15*(TBP)-XhoI-3'. All of the 5' restriction enzyme cut sites were in-frame with the start codons of the TFIID subunit ORFs. The design intent was to have complementary restriction enzyme sites for both cloning into a single MCS and into a dual expression vector system such as the one described below for yeast overexpression. For the HA_3-$3C_{pro}$-tag and the $FLAG_3$-TEV_{pro}-tag, complementary oligos (top strand and bottom strand) for each tag were annealed together to form "cut" 5'-RsrII- HA_3-$3C_{pro}$-NheI-3' and a 5'-RsrII- $FLAG_3$-TEV_{pro}-NheI-3'. When annealed together, the ends of these oligos mimic overhangs that would be generated by restriction enzyme digestion and can be ligated with compatible overhang ends. To create double stranded DNA fragments for cloning of synthesized oligos into plasmids, 10μg of both top and bottom strand oligos were phosphorylated by T4 PNK (NEB) per manufacturers instructions in separate reactions and mixed together. The reactions were then put in a pot of water and heated to a boil with a Bunsen burner. Once the water was vigorously boiling, the pot was removed from the heat and allowed to cool slowly on the benchtop to allow the two oligos to anneal. The annealed double stranded DNA was separated on a 1x TBE, 1.5% agarose gel, stained with ethidium bromide and the most intense band was excised and gel purified as described above. This procedure was used to anneal and purify all complementary oligos described in this dissertation. For both tags, the NheI restriction enzyme cut site is in-frame with the tag. To clone the tagged variants into pFASTBac HTc, the plasmids were digested with RsrII and an appropriate 3' restriction end for the subunit

to be cloned (XhoI (SalI compatible), KpnI, SphI, HindIII). The cut vector, either the HA_3-$3C_{pro}$-tag or the $FLAG_3$-TEV_{pro}-tag and the TFIID subunit ORF was ligated together in a 3-piece ligation reaction to form the expression plasmid containing tagged *TAF* variant. For untagged variants, the TFIID ORFs were ligated into the corresponding restriction sites in the pFASTBac1 MCS with the exception of *TAF2*. For *TAF2*, the 5'-SpeI-*TAF2*-SalI-3' fragment was ligated into the SpeI and XhoI restriction ends of the pFASTBac1 MCS such that the *TAF2* ORF would be in the correct orientation in the plasmid.

The pFASTBac plasmid series works within the Bac-to-Bac protein expression system (Invitrogen). This system required the DH10Bac *E. coli* strain. This strain expresses a baculovirus shuttle vector (bacmid) that encodes the proteins required for efficient virus production and a helper plasmid that helps transpose the pFASTBac expression sequence into the bacmid. DH10Bac competent cells were generated using the Inoue method (Inoue et al., 1990). Recombinant bacmids were generated according to manufacturers instructions and transfected into *Sf9* cells (described below).

For expression in *S. cerevisiae*, *TAF2* and tagged-*TAF2* variants were cloned into p425*GAL1* (Mumberg et al., 1994). First, the same SpeI to SalI full length Taf2 fragment was cloned into p425*GAL1* at the SpeI to SalI restriction sites within the MCS. This plasmid was used to overexpress an untagged variant of Taf2. However, in order to purify overexpressed Taf2 from yeast, two distinct MBP-tagged Taf2 constructs were generated, one with a maltose binding protein (MBP) tag at the N-terminus of Taf2 and the other with an MBP tag at the C-terminus of Taf2. To generate the N-terminal MBP tag, PCR based cloning methods were used to amplify the MBP coding sequence from the pLM302 plasmid, a plasmid obtained from the Vanderbilt Center for Structural biology. pLM302 is a

derivative of pET27 (Novagen) engineered to contain MBP (derived from pMAL (NEB)) followed by the $3C_{pro}$ site and a multiple cloning site. The primers used to clone MBP introduced an XbaI restriction end at the 5' end of the MBP coding sequence and an in-frame SpeI restriction end at the 3' end of the $3C_{pro}$ site. Two MBP-$3C_{pro}$ variants were generated in this manner, one with a stop codon for purification of just MBP-$3C_{pro}$ (termed MBP hereafter) and one without a stop codon for purification of MBP-$3C_{pro}$-Taf2 (termed MBP-Taf2 hereafter). The PCR generated MBP DNA fragments were cloned into the SpeI restriction site in p425*GAL1* (Mumberg et al., 1994). DNA cleaved with XbaI generates a compatible end with SpeI digested DNA. After verification that the MBP was cloned into the p425*GAL1* plasmid correctly, the full length SpeI to SalI Taf2 fragment was cloned into p425*GAL1 MBP* (no stop codon) cut with SpeI to SalI restriction endonucleases to generate the p425*GAL1 MBP-TAF2* plasmid. For the 3' MBP-tag, using PCR, a *TAF2* 3' fragment was generated containing a 5' HindIII restriction end and an in-frame AvrII restriction end without a stop codon. The 5' HindIII end was generated using a naturally occurring internal HindIII site within the *TAF2* coding region (+3303). The MBP tag was generated by adding a 5' in-frame AvrII restriction end, inserting a stop codon at the end of the MBP coding sequence and adding a 3' SalI restriction end. Initially, this tag was cloned into a p415*TAF2* plasmid, a derivative of pRS415 (Sikorski and Hieter, 1989) containing the enhancer/promoter and ORF of the *TAF2* gene. The p415*TAF2* plasmid was digested HindIII to SalI to remove the 3' end of the *TAF2* coding region. This digested plasmid was then ligated to both the HindIII-*TAF2*-AvrII fragment and the AvrII-*MBP*(STOP)-SalI fragments in a 3-piece ligation. The resulting plasmid was p415*TAF2 TAF2-MBP*. However, this plasmid did not have a $3C_{pro}$ cleavage site between *TAF2* and MBP so the MBP tag could

not be cleaved off. To insert a $3C_{pro}$ site, complementary oligonucleotides (top strand and bottom strand) were generated that encoded 5'-SpeI-SacI-His_6-$3C_{pro}$-SpeI-3' insert and annealed as described above. The annealed oligos were then ligated into p415*TAF2-MBP*, which had been linearized with AvrII. AvrII overhangs are compatible with SpeI overhangs. Considering the oligo could be ligated into the plasmid in either direction, the SacI restriction end was included to determine the orientation of the insert in an analytical restriction digestion. Both the 5'-SpeI-*TAF2-MBP*-SalI-3' and the 5'-SpeI-*TAF2*-SacI-His_6-$3C_{pro}$-*MBP*-SalI-3' fragments were then subcloned into p425*GAL1* at the SpeI to SalI sites within its MCS. The p425*GAL1 TAF2-His6-3C$_{pro}$-MBP* construct was used for overexpression by Gal induction. The protein product generated by this plasmid will be described as Taf2-MBP hereafter.

In addition to overexpressing Taf2 by *GAL* induction, a multi-protein expression system was designed based on the bi-directional *GAL1-GAL10* enhancer/promoter sequence for the purpose of Taf2-TFIID subunit co-expression and purification (Burgers, 1999; Gerik et al., 1997). This Gal induction expression system is a two-plasmid system (Figure 2.1). Both plasmids were derived from pBluescript II SK+ (Agilent). Plasmid #1 was designed as the yeast expression plasmid, containing a 2μm origin of replication, a selectable marker and an MCS containing 7- and 8-bp restriction endonuclease cut sites (BlpI, RsrII, PmeI and PacI). Three variants of Plasmid #1 were generated, one containing a *TRP1* selectable marker (pJFTRP1), one containing a *LEU2* selectable marker (pJFLEU2) and one containing a *URA3* selectable marker (pJFURA3). The 2μm replication site was PCR amplified from pRS425 (Christianson et al., 1992), engineered to contain 5' and 3' KpnI ends and cloned into the KpnI restriction site within pBluescript II SK+. *TRP1*, *URA3* and

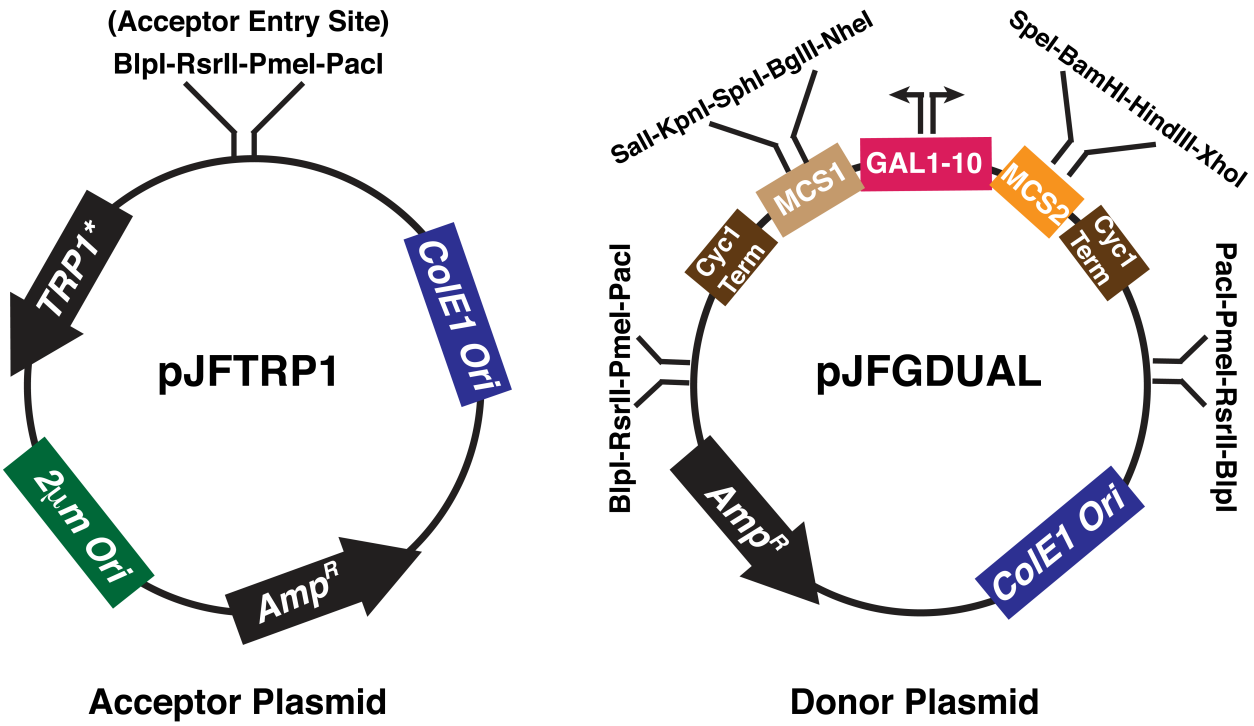


Figure 2.1: Multi-ORF Overexpression System for *S. cerevisiae*. Diagram of elements engineered into donor and acceptor plasmids. * = Three different acceptor donors were generated: pJFTRP1, pJFLEU2, pJFURA3.

LEU2 were PCR amplified from pRS424, pRS426 and pRS425, respectively, engineered to have SpeI to XhoI restriction ends and ligated into pBluescript II SK+ at the SpeI and XhoI restriction sites within its MCS. A Not1 – BlpI – RsrII – PmeI – PacI – NotI fragment was generated from complementary oligos (top strand and bottom strand) as described above and ligated into pBluescript II SK+ at the NotI restriction site within its MCS. These 7- to 8-bp restriction enzyme cut sites would be used as the entry point for the multi-TFIID subunit dual overexpression cassettes, described below.

Plasmid #2 (pJFGDUAL) was an entry plasmid to which the TFIID subunits were initially cloned. This plasmid was engineered in stages. Step #1, the MCS of pBluescript II SK+ was replaced with an MCS containing the following restriction enzyme cut sites: Blp1-

RsrII-PmeI-PacI-EcoRI-SalI-KpnI-SphI-BglII-NheI-SpeI-BamHI-HindIII-XhoI-EcoRV-PacI-PmeI-RsrII-BlpI-SacI. Step #2, the *GAL1-GAL10* bi-direction enhancer/promoter was PCR amplified from genomic DNA, engineered to include a 5' NheI end and a 3' SpeI end and ligated into these same restriction sites within the new MCS. Step #3, the *CYC1* terminator (*CYC1*$_{term}$) sequence was PCR amplified from p415ADH and engineered to have two sets of restriction ends. The first variant contained XhoI and EcoRV 5' and 3' restriction ends, respectively. The second variant contained SalI to EcoRI 5' and 3' restriction ends, respectively. These fragments were sequentially cloned into plasmid #2 to complete the plasmid. The result is a plasmid with two distinct MCS's to which TFIID subunit ORFs can be cloned (see Figure 2.1). As an alternative to using the *CYC1*$_{term}$ sequences, the TFIID subunits *TAF1*, *TAF3*, *TAF4*, *TAF5*, *TAF6*, *TAF7*, *TAF8*, *TAF9*, *TAF10*, *TAF11*, *TAF12*, *TAF13*, *TAF14* and *SPT15* were PCR amplified from genomic DNA including ~100bp beyond the 3' UTR sequence. Coordinates for the end of the transcription unit for these genes was derived from RNAseq data from the yeast genome (Nagalakshmi et al., 2008). In addition, a 5'-SpeI-His$_6$-*MBP*-3C$_{pro}$-NheI-3' fragment and a 5'-His$_6$-*GST*-3C$_{pro}$-3' fragment were generated through PCR from pLM302 and pBG101 plasmids. pBG101 was acquired from the Vanderbilt Center for Structural Biology. pBG101 was constructed in a similar manner as pLM302 except the GST sequence derived from pGEX (GE Healthcare).

Ultimately, several multi-TFIID subunit expression plasmids were generated. As mentioned above, TFIID subunits that were known to interact were cloned into the same expression cassette. For example, Taf1 and Taf7 were cloned into the same expression cassette (Bhattacharya et al., 2014; Singh et al., 2004; H. Wang et al., 2014). Plasmids that included the *CYC1*$_{term}$ were: pJFURA3 *TAF5*; pJFURA3 *TAF11*,*TAF13;* pJFLEU2 *TAF8*,*TAF10*;

pJFLEU2 *His6-MBP-3C$_{pro}$-TAF1*,*TAF7*; pJFLEU2 *TAF1*,*TAF7*; pJFLEU2 *TBP*,*TAF2*; pJFLEU2 *TBP*, *His6-MBP-3C$_{pro}$-TAF2* pJFLEU2 *TAF4*, HA$_3$-3C$_{pro}$-*TAF12*; pJFTRP1 *TAF6*,*TAF9*; pJFTRP1 *His6-GST-3C$_{pro}$-TAF8*, *TAF10*; pJFTRP1 *His6-MBP-3C$_{pro}$-TAF8*,*TAF10*; pJFTRP1 *His6-GST-3C$_{pro}$-TAF3*,*TAF10*; pJFTRP1 *His6-MBP-3C$_{pro}$-TAF3*,*TAF10*; pJFTRP1 *TBP,TAF2*; pJFTRP1 *TAF5*; pJFTRP1 *His6-GST-3C$_{pro}$-TAF1,TAF7*; pJFTRP1 *His6-MBP-3C$_{pro}$-TAF1,TAF7*; pJFTRP1 *TAF1,TAF7*. The plasmids containing native TAF terminator sequences: pJFTRP1 *TAF6*, *TAF9*; pJFTRP1 *TAF6*, *TAF9*, *TAF5*, *TAF14*; pJFTRP1 *TAF6*, *TAF9*, *TAF5*, *TAF14*, *TAF8*, *TAF10*; pJFTRP1 *TAF6*, *TAF9*, *TAF5*, *TAF14*, *TAF8*, *TAF10, TAF4, TAF12*; pJFTRP1 *TAF6*, *TAF9*, *TAF5*, *TAF14*, *TAF8*, *TAF10, TAF4, HA$_3$-3C$_{pro}$TAF12*; pJFURA3 *TAF11*, *TAF13*; pJFURA3 *TAF11*, *TAF13*, *TAF3*, *TAF10*; and pJFURA3 *TAF11*, *TAF13*, *TAF3*, *TAF10*, *TAF1*, *TAF7*.

Bacterial, Yeast and Insect Cell Growth Manipulations

E. coli manipulations - DH5α was used for all cloning and propagation of plasmids. DH5α competent cells were acquired from the Vanderbilt Molecular and Cellular Biology core. Cells were grown in Luria-Bertani (LB) broth (1% w/v Tryptone, 0.5% yeast extract and 0.5% sodium chloride). During transformation reactions, cells were resuspended in Super Optimal broth with Catabolite repression (SOC) medium (2% tryptone, 0.5% yeast extract, 10mM sodium chloride, 2.5mM potassium chloride, 10mM magnesium chloride, 10mM magnesium sulfate and 20mM glucose). All transformation reactions were performed using 1.5mL snap cap tubes. For simply propagating plasmids, a modified one-minute transformation protocol was used (Golub, 1988). Briefly, 50-200ng of plasmid DNA was added to 10uL of competent cells that had been thawed on ice. The cells were immediately heat shocked at 42^o for 30s. The cells were then resuspended in 100μL

of SOC medium. If the plasmid resistance marker encoded β-lactamase (Amp^R), the entire reaction was immediately plated to LB + 100mg/mL Ampicillin agar plates. However, if the plasmid resistance cassette encoded the kanamycin resistance enzyme (Kan^R), the reaction was incubated at 37° for 30min prior to plating to LB + 50mg/mL Kanamycin. For ligation reactions, the entire 5μL ligation reaction (set-up described below) was mixed with 50uL of competent cells thawed on ice. The cells were then incubated on ice for 15-30min, heat shocked at 42° for 30s. After 5min rest on ice, the cells were resuspended in 500μL of SOC and incubated on a tiltboard at 30°. After 1hour, 20-40% of the reaction was plated on LB + the appropriate antibiotic (Ampicillin or Kanamycin). Plated cells were incubated at either 37° for 14-20hrs or 30° for 24-36hrs prior to isolating individual colonies. To isolate plasmids, individual colonies were grown in 5mL of LB + appropriate antibiotic at 37° overnight (12-16 hours) shaking at 250rpm. For small-scale plasmid purifications, the entire 5mL of cells were harvested by centrifugation in a Beckman Coulter J6-HC centrifuge for 5min at 4.2k rpm at room temperature. The supernatant was discarded and the plasmid DNA was harvested using Qiagen Mini-Prep kit according to manufacturers instructions. Alternatively, the buffers from the Qiagen Mini-Prep kit were used along with the purification columns from Denville Scientific. For large-scale plasmid purifications, the entire 5mL of cells were used to inoculate 250mL of LB + the appropriate antibiotic and grown at 37° for 12 hours shaking at 250rpm. The cells were harvested in J6-HC for 5min at 4.2k rpm at room temperature. The supernatant was discarded and the pellets were either immediately processed or stored at -80° until ready to process. DNA was harvested from the pellet using Qiagen Maxi-Prep filter kit according to manufacturers instructions.

Recombinant proteins were expressed in *E. coli* Rosetta2 (DE3) strains (Novagen). Competent cells were generated using the Inoue method (Inoue et al., 1990). For transformation, 20-50mL of Rosetta2 (DE3) competent cells, thawed on ice, were mixed with ~500ng of plasmid DNA. The cells were incubated on ice for 30-45min and heat shocked at 42^{o} for 30s. After 5min rest on ice, the cells were mixed with 500μL of SOC medium and incubated on a tiltboard at 30^{o}. After 1hour, 40-80% of cells were plated to LB + 34μg/mL (dissolved in ethanol) Chloramphenicol + 50μg/mL Kanamycin agar plates and grown at 37^{o} overnight (12-16) hours. When expressing proteins, individual colonies were picked to 5mL of LB + 34μg/mL Chloramphenicol + 50μg/mL Kanamycin liquid medium and grown overnight at 37^{o} shaking at 250rpm. The overnight grown culture was used to inoculate fresh LB + 34μg/mL Chloramphenicol + 50μg/mL Kanamycin liquid medium to an optical density at 600nm (OD_{600})=0.1. The liquid culture was grown at 37^{o} shaking at 250rpm to an OD_{600}=0.6-0.9 and then the temperature was subsequently lowered to 30^{o}. After 15min to allow the temperature to equilibrate, 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the cultures to a final concentration at 1mM to induce protein expression. Induction was allowed to proceed for 4hours at 30^{o} prior to harvest. Cells were harvested by centrifugation in a J6-HC centrifuge for 10min at 4200 rpm at room temperature (20^{o}). For small-scale inductions (5-50mL), the supernatant was discarded and the pellets were immediately stored at -80^{o}. For large-scale inductions, the pellets were resuspended in fresh LB, transferred to 50mL conical tubes and centrifuged again. The supernatant was discarded and the pellets were stored at -80^{o}.

Sf9 Cell Manipulations – *Spotoptera frugiperda* (*Sf9*) cells were obtained from Invitrogen Gibco Catalog No. 11496-015. The cells were received on dry ice, thawed at 37^{o}

until the residual ice just completed melting and used to inoculate 25mL of SF900 II serum free medium (SFM) (Invitrogen) + 100U/mL Penicillin/Streptomycin antibiotic (Gibco), pre-warmed to room temperature. For normal passage and cell growth, the cultures were grown in SF900 II SFM + 100U/mL Penicillin/Streptomycin antibiotic at 28^{o} in 250mL vent closed shaker flasks, shaking 100-125rpm. All cell measurements were performed with trypan blue to distinguish between living and dead cells and counted with a hemocytometer. Once the original inoculum reached $2x10^{6}$ cells/mL, the cell volume was gradually increased so the cell concentration was $1x10^{6}$ cells/mL in 100mL medium and allowed to continue to grow. Once the cells doubled two additional times (~48 hrs), half of the culture was used to make freezer stocks ($20x10^{7}$ cells in total). The cells were centrifuged for 5min at 1k rpm in J6-HC. To make the freezing medium, 10mL of supernatant (conditioned medium) was mixed with 10mL of fresh SF900 II SFM containing 14% DMSO and placed on ice. The rest of the supernatant was discarded and the cells were resuspended in freezing medium at a final concentration of $1x10^{7}$ cells/mL. The cells were distributed to 2mL screw cap cryogenic vials and stored at -80^{o}. The rest of the cells were split to $1x10^{6}$ cells/mL and allowed to continue to grow. Cells were only maintained in continuous culture for a maximum of 3 months prior to thawing a fresh vial of cells.

For bacmid transfection, actively growing cells were diluted to $5x10^{5}$ cells/mL in SF900 II SFM and 5mL were plated to T25 Nunclon vent closed flasks (ThermoFisher). Cells were allowed to adhere for 1 hour at room temperature, during which the transfection reactions were prepared. In two separate sterile 1.5mL snap cap tubes, 8μL of Cellfectin II reagent (Life Technologies) were mixed with 100μL of SF900 II SFM and 10μL of Bacmid mini-prep were mixed with 100μL of SF900 II SFM. The two solutions were

mixed together and DNA:lipid transfection reagent complex was allowed to form for 30min. The transfection reaction was added to 2.8mL of SF900 II SFM. Once the cells adhered, the residual medium was aspirated from the T25 flasks, the cells were washed once with 3mL of SF900 II SFM and the transfection solution was overlayed onto the cell monolayer. The cells were then incubated at 28^{o} for 5hrs. After 5 hours, the transfection solution was aspirated from the cells, 6mL of SF900 II SFM was added to the cells and the cells were incubated at 28^{o} for 60hrs. After 60hrs, the viral supernatant was harvested and filtered through 0.45μm filter into sterile 15mL conical tubes. This original viral supernatant (V_o) was used to amplify the virus.

In order to amplify the virus for protein expression, 250μL of V_o was added to 2x100mL of Sf9 cells at a concentration of 1x10^{6} cells/mL in two shaker flasks and allowed to grow under the following growth paradigm. Ideally, the cells would double 2 to 4 times and then arrest. Following arrest, the cells would significantly swell in size. Once the cells take on this characteristic phenotype, the cells were allowed to produce virus for an additional 24 hours prior to harvesting the virus. If the cells reached 5x10^{6} cells/mL prior to showing any sign of arrest, the cells were split 1:2 into fresh medium until the cells arrested. The final cell concentration at the time of arrest was ideally between 3-6x10^{6} cells/mL. The entire process would take between 3-7 days. The virus was harvested by centrifuging the cells at 1k rpm for 5min in a J6-HC centrifuge at room temperature. The supernatant was filtered with a 0.45μm bottle top filter assembly into sterile plastic bottles. This amplified virus (V_1) could then be used for protein expression.

To determine the viral titer of V_1, the end-point dilution method was employed. This method identifies the tissue culture infectious dose at 50% (TCID50) as described by

Reed and Muench (Reed and Muench, 1938). In essence, the V_1 virus is serially diluted and then each dilution is used to infect multiple wells (experimental replicates) of plated cells. Over 5-7 days, the infected replicates will show severe cytological effects or die whereas uninfected replicates will look normal and healthy. High concentrations of V_1 will completely infect all of the experimental replicates while low concentration will not infect any of the experimental replicates. However, in the middle of the dilution series, a few viral dilutions will result in a mix of infected and uninfected replicates. These dilutions bracket the end-point and are used to determine the infectious units. In general, the viral amplification procedure highlighted above resulted in viral titers between $1x10^8$ to $1x10^9$ infectious units/mL.

For protein expression, cells were infected at a multiplicity of infection (MOI) of ~5 of each virus. MOI is calculated based on the relationship between infectious units and total cells. For example, if one wanted to infect 10^8 cells at an MOI of 5, then s/he would need to infect with $5x10^8$ infectious units. In essence, this insures that greater than 90% of cells are infected with the virus and maximizes protein expression. Upon infection, cell growth arrests and the cells display the characteristic swollen phenotype after 24-36 hours. Cells were harvested after 48-60hrs of infection by centrifugation at 1k rpm for 5min in J6-HC at room temperature. The supernatant was discarded and the cells were washed in 1x phosphate-buffered saline (PBS), pH 7.4 and spun again. The supernatant was discarded and the cells were stored at -80°.

Yeast Manipulations – BY4741, BY4743 (Brachmann et al., 1998), BY4741 *trp1Δ::KANMX6* and FM113 (Gary and Burgers, 1995) was used for overexpression of TFIID subunits. YLSTAF1 was used for Taf1-TAP TFIID purification (Papai et al., 2009). Both

strains were propagated in YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v dextrose), either plate or liquid, for normal maintenance. All strains were grown at 30^{o} shaking at 250rpm. For TFIID purification, YLSTAF1 was grown to approximately mid-logarithmic (mig-log) phase. Mid-log phase was determined experimentally by performing growth curves with each strain to determine the inflection point at which time exponential growth begins to slow prior to stationary phase. Synthetic complete (SC) medium (0.67% w/v yeast nitrogen base without amino acids, 2% w/v dextrose (or 1% w/v raffinose where indicated), 0.2% w/v amino acid dropout mix) without (-) Leucine (Leu), Uracil (Ura) or Tryptophan (Trp) or a combination of the three, where appropriate, was used to grow strains transformed with expression plasmids.

All transformations were performed using the lithium acetate/PEG/salmon sperm carrier DNA transformation protocol (Gietz and Schiestl, 2007). Briefly, single yeast colonies were used to inoculate YPD and grown overnight at 30^{o} shaking at 250rpm. The following day, the overnight culture was used to inoculate 2.5mL per transformation reaction of YPD to an OD_{600} of approximately 0.2 and cells were allowed to grow for 5hours at 30^{o} shaking at 250 rpm. Cell pellets were harvested by centrifugation in J6-HC centrifuge, spinning at 4.2k rpm for 5min at room temperature. The supernatant was discarded and the pellet was washed with 1.25mL sterile deionized distilled H_2O (ddH_2O) per transformation reaction and centrifuged as above. Again, the supernatant was discarded the cell pellet was resuspended in transformation master mix: 120μL 50% w/v polyethylene glycol (PEG) MW 3350, 18μL 1M lithium acetate, 18μL sterile ddH2O and 25μL 2mg/mL single stranded salmon sperm DNA; total of 180μL per reaction. The cell transformation mix was then transferred to 1.5mL snap cap tubes containing the plasmid

DNA of interest. For both individual transformations (1 plasmid) and co-transformations (up to 3 plasmids), a minimum of 50ng of each plasmid was used. A no plasmid control was always to used to ensure the reagents used were free of contaminating DNA or cells that are prototrophic for the essential nutrient missing from the selection plates (Leucine, Uracil or Tryptophan). The transformation reaction was then incubated at 42^{o} for 40min for heat shock. Following the heat shock incubation, the cells were pelleted at 5k rpm for 2min in a table-top mini-centrifuge (Eppendorf Centrifuge 5417C). The supernatant was discarded and the pellet was resuspended in 500μL of sterile ddH_2O. For single transformations, 10% of the reactions was plated to SC – Leu, Ura or Trp agar and grown at 30^{o} for 36-48 hrs. For co-transformations, 80% of the transformation reactions were plated to SC – the appropriate combination of Leu, Ura or Trp agar and grown at 30^{o} for 36-48 hours. If the "no DNA control" reaction resulted in no transformants, I would proceed with the experimental transformations.

Colonies that grew on selective media agar were used to inoculate 5mL of SC – appropriate nutrient and grown at 30^{o} with shaking at 250rpm overnight. For Gal induction, 1mL of overnight culture was diluted 1:10 in sterile ddH_2O and centrifuged in J6-HC for 5min at 4.2k rpm. The supernatant was discarded and the cells were resuspended in 5mL of SC (+1% w/v raffinose) – appropriate nutrient and grown at 30^{o} with shaking at 250 rpm overnight. The raffinose grown culture was used to inoculate fresh raffinose medium to an OD_{600}=1 and grown to an OD_{600} between 2-3 prior to galactose addition. Once the cells reached OD_{600}=2-3, galactose was added to a final concentration of 2% w/v. For small-scale induction, 20% w/v galactose liquid was used to induce the cultures. For large-scale induction, solid galacose was used to induce the cultures. Inductions were

allowed to proceed for between 2 and 8 hours (4hrs was optimal) at 30° shaking at 250rpm. Following induction, cells were centrifuged at 4.2k rpm for 5min in J6-HC centrifuge at room temperature. The cell pellets were washed in ddH_2O. For small-scale inductions, the cells were transferred to 2mL snap-cap tubes, centrifuged for 1min at 14k rpm in table-top mini-centrifuge and, after discarding the supernatant, immediately processed for immunoblotting (described below). For large-scale inductions, the cells in ddH_2O wash were distributed to 50mL conical tubes and centrifuged again as above. The supernatant was discarded the cell pellets were stored at -80° prior to purification.

It was my observation that growth of cells in raffinose is particularly finicky. The doubling time of cells grown in raffinose is quite slow, ~3hrs instead of the standard 90min. In addition, cells that are allowed to reach stationary phase growth have a significant and unpredictable lag time before they can resume doubling in fresh raffinose medium. Thus, it is imperative that the cells be actively growing in raffinose for timing of the experiments to be consistent.

Analytical Immunoblotting for Overexpressed Protein Detection

To determine if the bacmids effectively recombined such that they could produce TFIID subunits, anti-TFIID subunit immunoblots were performed. Following transfection of the recombinant bacmid into Sf9 cells, the cells were washed off the T25 flasks with the virus conditioned medium and pelleted by centrifugation at 1k rpm for 5min in J6-HC centrifuge at room temperature. The cell pellet was resuspended in 500μL of 1x lithium dodecyl sulfate (LDS) NuPAGE buffer (Invitrogen) with 100mM DTT and heated at 75° for 10min. Between 0.5% and 2% of the cell extract was used for

immunoblotting. To determine if TFIID subunits were overexpressed upon Gal induction, total cell protein was extracted from $2OD_{600}$ units of cells of early to mid-log grown culture using sodium hydroxide based lysis (Kushnirov, 2000). In both instances, cell extracts were separated via SDS-PAGE using 4-12% NuPAGE Bis-Tris gradient gels (Life Technologies), run with 1x MOPS running buffer and electro-transferred to PVDF membranes using a wet transfer system. Polyclonal anti-Taf and anti-TBP antibodies (Papai et al., 2010, 2009; Sanders et al., 2002b; Singh et al., 2004) were used in the following manner. All antibodies were diluted in 1% non-fat milk (Carnation) in 1x tris buffered saline (25mM Tris-HCl pH 7.5, 150mM NaCl). The antibody dilutions used are as follows: anti-Taf1 1:1000, anti-Taf2 full length 1:500, anti-Taf2 N-term peptide 1:1000, anti-Taf2 C-term peptide 1:1000, anti-Taf3 1:1000, anti-Taf4 1:1000, anti-Taf5 1:2500, anti-Taf6 1:1000, anti-Taf7 1:5000, anti-Taf8 1:2000, anti-Taf9 1:500, anti-Taf10 1:2500, anti-Taf11 1:1000, anti-Taf12 1:5000, anti-Taf13 1:500, anti-Taf14 1:5000 and anti-TBP 1:2500. Anti-HA horseradish peroxidase (HRP) conjugate (3F10) antibody was procured from Roche and used at 1:5000 dilution. Goat anti-Rabbit (Fc) HRP conjugate was procured from ThermoFisher and used at 1:10,000 dilution. Protein signal was detected using Amersham ECL Western Blotting Detection Reagent (GE Healthcare) and exposed to Blue Basic Double Emulsion Autoradiography Film (GeneMate).

Recombinant Protein Purification

All purification steps were performed at 4^{o} or on ice unless noted. For each His_6-Taf2 expression construct (full length Taf2 (aa1-1407), Taf2 aa1-407, Taf2 aa1-1007, Taf2 aa401-1007, Taf2 aa401-1407, Taf2 aa1001-1407), approximately 3-5g of *E. coli* pellets

were generated from 1L of induced *E. coli* cultures. The cell pellets were lysed in 40mL of *E. coli* lysis buffer (20mM Tris-HCl pH 7.9, 200mM NaCl, 20mM imidazole, 0.1% Triton X-100, 10% glycerol, 1mM DTT and 1x protease inhibitors) in the presence of approx. 100μg/mL lysozyme. Lysates were sonicated and insoluble material was pelleted by spinning at 14k rpm in Sorvall RC 5C Plus centrifuge with an SS-34 rotor for 15min. Soluble lysate was then bound in batch to 2mL of Ni^{2+}-NTA agarose (Qiagen) for 2 hours at 4^{o} with mixing. Ni^{2+}-NTA agarose and bound proteins were transferred to disposable 20mL chromatography columns (Bio-Rad) and were washed in column format with >10 column volumes of *E. coli* lysis buffer. Proteins were eluted with elution buffer (lysis buffer without T4 lysozyme but with 200mM imidazole) and peak fractions were collected. Quantity and purity of purified material were assessed by SDS-PAGE followed by coomassie brilliant blue staining (0.1% coomassie brilliant blue, 50% methanol and 10% acetic acid).

For purification of Taf2 and TFIID subcomplexes from Sf9 cells, between $2x10^8$ and $1x10^9$ cells were infected with baculovirus according to the paradigm described above. Baculoviral infections were performed with HA_3-$3C_{pro}$-Taf2 expressing virus; a combination of Taf1, HA_3-$3C_{pro}$-Taf2 and TBP expressing viruses; a combination of Taf4 and HA_3-$3C_{pro}$-Taf12 expressing viruses; a combination of Taf4, Taf5 and HA_3-$3C_{pro}$-Taf12 expressing virus; a combination of HA_3-$3C_{pro}$-Taf5, Taf6 and Taf9 expression viruses; a combination of Taf4, Taf5, Taf6, Taf9 and HA_3-$3C_{pro}$-Taf12 expressing viruses; a combination of Taf3, Taf4, Taf5, Taf6, Taf9, Taf10 and HA_3-$3C_{pro}$-Taf12 expressing viruses; a combination of Taf4, Taf5, Taf6, Taf8, Taf9, Taf10 and HA_3-$3C_{pro}$-Taf12 expressing viruses; a combination of Taf3, Taf4, Taf5, Taf6, Taf8, Taf9, Taf10 and HA_3-$3C_{pro}$-Taf12 expressing viruses; a combination of Taf3, Taf4, Taf5, Taf6, Taf9, Taf10, Taf11, HA_3-$3C_{pro}$-

Taf12 and Taf13 expressing viruses; a combination of Taf3, Taf4, Taf5, Taf6, Taf8, Taf9, Taf10, Taf11, HA_3-$3C_{pro}$-Taf12 and Taf13 expressing viruses; and a combination of all TFIID subunit expressing viruses, all of which expressed untagged TFIID subunits with the exception of the HA_3-$3C_{pro}$-Taf12-expressing virus. The infected cells were aliquoted for trial purifications ($2x10^7$ cells) or for preparative scale purifications ($2x10^8$). For each type of purification, the following *Sf9* lysis buffer was used: 20mM HEPES-KOH pH 7.9, 200mM KOAc, 10% glycerol, 0.1% NP40 substitute, 2mM DTT, 0.2mM PMSF, 2mM Benzamidine-HCl, 1mM EDTA, 1μg/mL pepstatin A, 2.5μg/mL aprotinin and 2.5μg/mL leupeptin. In small-scale purification, pellets were resuspended in 1mL of lysis buffer and lysed by sonication. Cell debris was removed by centrifugation at 14k rpm for 15min in the tabletop mini-centrifuge. The supernatant was mixed with 20-60μg of 12CA5 monoclonal antibody (anti-HA) covalently cross-linked to proteinA sepharose (Invitrogen) at 4mg/mL antibody to bead volume for 3hrs on a tiltboard. Following binding, the beads were washed 2x with 1mL of ice-cold lysis buffer. The bound protein complexes were then eluted with 600ng of 3C protease in 50μL lysis buffer at room temperature with intermittent mixing. The eluted material was then visualized with SDS-PAGE followed by coomassie brilliant blue staining and immunoblotting. For large-scale purification, $2x10^8$ cells were resuspended in 20mL of *Sf9* lysis buffer and lysed by sonication. Cellular debris was removed by centrifugation at 14k rpm in Sorvall RC 5C Plus centrifuge with an SS-34 rotor for 15min. The supernatant was mixed with 4mg of 12CA5 antibody crosslinked to 1mL of proteinA sepharose and allowed to bind on tiltboard for 3hrs. Following binding, the beads were transferred to a 10mL disposable chromatography column (Bio-Rad) and washed with >10 bed volumes of lysis buffer. The column was capped and subjected to 3C elution overnight on a tiltboard

with 1mL of elution buffer at concentration of 100ng 3C protease/mL. The 3C eluate was extracted from the beads by centrifugation at 1k rpm for 5min in J6-HC centrifuge into a 15mL screw cap conical tube (Denville Scientific). The eluate was further purified using a Superose6 gel filtration column. Final purified material was analyzed by SDS-PAGE followed by SyproRuby gel staining. The gels were imaged using PharosFX Scanner (Bio-Rad).

Purification of Taf2 variants from S. cerevisiae: MBP-Taf2 and Taf2-MBP were expressed in BY4741 by Gal induction as described above and 3g of cell pellet/1L of culture was harvested. For purification, all manipulations were performed at 4^{o} or on ice unless specified. Thirty grams of cell pellet were lysed in 30mL of Taf2 purification buffer (20mM HEPES-KOH, 500mM potassium acetate, 0.5% NP40 substitute, 10% glycerol, 2mM DTT, and 2X protease) using glass bead lysis. Lysate was then centrifuged at 14k rpm in Sorvall RC 5C Plus centrifuge with an SS-34 rotor for 15min and soluble cell extract was mixed with 5mL of DE-52 resin equilibrated in Taf2 purification buffer for 10min with mixing at 20^{o}. DE-52 flowthrough was then diluted with 20mM HEPES-KOH pH 7.9, 10% glycerol, 1mM DTT and 1X protease inhibitors to reduce the potassium acetate and the NP40 substitute concentration to 200mM and 0.2%, respectively, and bound to 5mL amylose resin (NEB) in batch with mixing for 2 hours. The amylose resin-bound proteins were transferred to a disposable chromatography column, washed with >10 column volumes amylose wash buffer (20mM HEPES-KOH pH 7.9, 200mM potassium acetate, 0.1% NP40 substitute, 10% glycerol, 1mM DTT and 1X protease inhibitors). MBP-Taf2 and Taf2-MBP were eluted with wash buffer + 10mM maltose and peak fractions were collected. Alternatively, Taf2 was eluted with lab generated 3C protease at 100ng/mL concentration

at 4° for 16hrs. MBP-Taf2 or free Taf2 was further purified using a MonoQ column (Pharmacia). Peak eluate fractions from the amylose resin purification were loaded onto MonoQ with BA200 (BA = 20mM HEPES-KOH pH 7.9, 10% glycerol, 1mM DTT and 1x protease inhibitors with variable concentrations of potassium acetate, eg. BA200 contains 200mM potassium acetate). Proteins were eluted with a linear gradient of BA200 to BA1500. Peak fractions eluted at approx. BA1200. Peak fractions were pooled and dialyzed extensively against dialysis buffer.

HA_1-Taf1 and Taf1-TAP TFIID Purification

HA_1-Taf1 TFIID was purified essentially as described (Poon et al., 1995; Sanders et al., 2002a) with the following modifications. Woontner extracts (Woontner et al., 1991) were dialyzed against 20mM HEPES-KOH pH 7.9, 50mM potassium acetate, 20% glycerol, 5mM DTT, 1X protease inhibitor mix (0.1mM PMSF, 1mM Benzamidine HCl, 2.5mg/mL aprotinin, 2.5mg/mL leupeptin, 1mg/mL pepstatinA) until the dialysate reached a conductivity equivalent to BA300. Dialyzed extract derived from a maximum of 300g of yeast cell pellet was chromatographed over a 200mL Bio-Rex70 100-200 mesh column. For immunopurification, 10% SurfactAmps NP40 was added to Bio-Rex70 1M fraction to a final concentration of 0.2%. The Bio-Rex70 1M fraction was subsequently diluted 1:2 in BA0, ethidium bromide was added to final concentration of 50μg/mL and subjected to anti-HA affinity chromatography with 10mg of anti-HA 12CA5 mAb covalently coupled to 2.5mL proteinA sepharose beads (Life Technologies) at 4° for 16 hours with mixing. Bound proteins were transferred to a 10mL disposable chromatography column, washed with >5 column volumes of BA300 with 0.1%, SurfactAmps NP40, >5 column volumes of BA300

with 0.01% SurfactAmps NP40 and >5 column volumes of BA300 with 0.001% SurfactAmps NP40. TFIID was eluted from the column two times with 2.5mL of elution buffer (BA300 with 0.001% SurfactAmps NP40 plus 2mg/mL HA_1 peptide) for 30min at 20° with mixing. The HA_1 peptide eluate was immediately subjected to ion exchange chromatography on a 1.2mL UnoS column (Bio-Rad). Following binding of the HA_1 peptide eluate, the UnoS column was washed with >5 column volumes of BA300 and then subjected to a linear gradient of BA300 to BA1000. TFIID elutes at approx. BA650.

Taf1-TAP TFIID was purified according to a modified tandem affinity purification protocol. The following protocol was performed multiple times with different amounts of yeast cell pellet with comparable results. One kilogram of YLSTAF1 pellet was resuspended in 500mL of 3X TAP buffer (0.45M Tris-Acetate pH 7.8, 0.15M potassium acetate, 60% glycerol, 3mM DTT, 3mM EDTA, 3x protease inhibitors) and lysed by glass bead beating. Then, the salt concentration of the lysate was adjusted with 5M potassium acetate to a conductivity equivalent of 300mM potassium acetate and centrifuged at 42k rpm in Beckman Optima LE-80K Ultracentrifuge with a 45-Ti rotor for 90min. The post-ribosomal supernatant was collected, avoiding the turbid material at the bottom of the centrifugation tube and aliquoted for both analytical and large-scale purification. The ~800mL of post-ribosomal supernatant was bound to 80mL IgG Sepharose Fast Flow, in batch, at 4° for 2hr with mixing. Following binding, the IgG sepharose and bound protein was washed extensively in IgG Sepharose Binding buffer (20mM Tris-Acetate pH 7.8, 300mM potassium acetate, 10% glycerol, 0.5mM EDTA, 1mM DTT, 1x protease inhibitors). TFIID was eluted from the IgG sepharose in 100mL of IgG Sepharose Binding buffer plus TEV protease at 250ng/mL for 2 hours at 4°. Following TEV elution from IgG sepharose,

the eluate was immediately subjected to cation exchange chromatography. The TEV eluate was loaded onto the UnoS column at a rate of 1mL/min with BA300 buffer (BA = 20mM HEPES-KOH pH 7.9, 10% glycerol, 1mM DTT and 1x protease inhibitors with variable concentrations of potassium acetate, eg. BA300 contains 300mM potassium acetate). After loading the protein onto the UnoS column, the column was washed with >5 column volumes of BA300 and then the protein was eluted with a 10 column linear salt gradient from BA300 to BA1000. TFIID elutes at approx. BA650. Purification fractions were analyzed for conductivity to determine the salt concentration, immunoblotting and SyproRuby gel staining.

RESULTS

Full Length Soluble Taf2 Could Only Be Purified from *S. cerevisiae*

In order to purify soluble full-length Taf2 for *in vitro* biochemical assays, three different recombinant expression systems were employed. First, full length Taf2 and Taf2 fragments were expressed in and purified from *E. coli* as His_6 N-terminally tagged protein variants (Figure 2.2a). While all protein variants accumulated upon IPTG induction, expression attempts for all of the fragments that contained an intact Taf2 C-terminus resulted in reduced protein accumulation in *E. coli* whole cell extracts (data not shown). Upon Ni^{2+}-NTA agarose purification, the yield of soluble Taf2 aa 401-1407 and Taf2 aa 1001-1407 fragments was considerably lower than the Taf2 aa 1-407, Taf2 aa 1-1007 and Taf2 aa 401-1007 fragments. Furthermore, the yield of full length Taf2 was essentially non-existent. Attempts to purify full length Taf2 using either alternative tags (MBP) or by

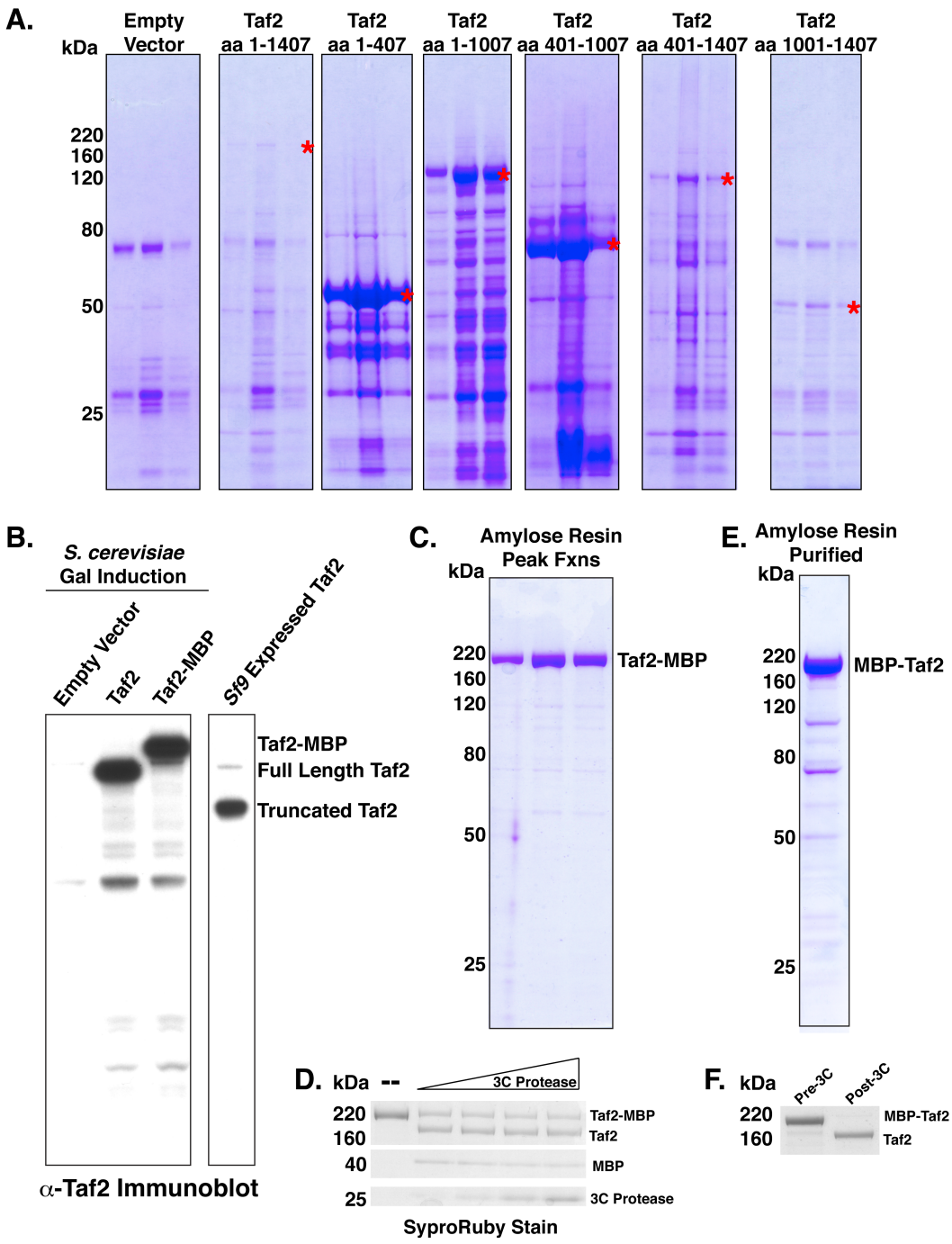


Figure 2.2: Recombinant Taf2 Protein Production from Various Sources. ***A.*** Partially purified His_6-tagged Taf2 variants, expressed in *E. coli* and purified via Ni^{2+}-NTA agarose, are presented. The peak fractions were analyzed via SDS-PAGE and coomassie brilliant blue (CBB) staining. Red asterisks indicate unit length Taf2 products. ***B.*** Anti-Taf2 immunoblot depicting protein overexpression of Taf2 forms in *S. cerevisiae* and *Sf9* cells. ***C.*** SDS-PAGE and CBB staining of amylose resin purified Taf2-MBP. ***D.*** Proteolytic cleavage of Taf2-MBP with between 2 and 15ng/μL of 3C protease to remove the MBP tag. ***E.*** SDS-PAGE and CBB staining analysis of amylose resin purified MBP-Taf2. ***F.*** Proteolytic cleavage of MBP-Taf2 with 2ng/μL 3C protease to remove the MBP tag.

employing a denaturing purification followed by a renaturation step failed to yield any soluble full length Taf2 (data not shown). In addition, attempts to use these Taf2 fragments in DNA binding assays failed to yield any DNA binding activity (data not shown). These results convinced me that *E. coli* was not a viable organism for production of Taf2. Thus, I began to explore alternative avenues of recombinant protein purification.

Recent successes in reconstituting milligram quantities of recombinant TFIID subcomplexes (Bieniossek et al., 2013; Trowitzsch et al., 2015) as well as transcriptionally active TFIIH (Fukuda et al., 2001) using baculoviral infection of insect cells suggested that this method of recombinant protein expression and purification could prove productive. Furthermore, the Tjian lab had employed baculovirus-based methods to express and purify drosophila Taf2 (Chen et al., 1994). Unfortunately, attempts at expressing full length Taf2 in insect cells resulted in a soluble but truncated Taf2 protein product (Figure 2.2b). Our lab had previously derived anti-Taf2 N- and C-terminal peptide affinity purified antibodies for the purposes of immunolabeling Taf2 in the TFIID complex (Papai et al., 2010, 2009). These reagents were used to determine that the truncated region of Taf2 lied at its C-terminus (data not shown). I hypothesized that co-expression of Taf2 with other TFIID subunits may protect the Taf2 C-terminus from degradation. Thus, to test this hypothesis, Taf2 was co-expressed with Taf1, Taf1 and TBP and with all of the TFIID subunits at one time. These efforts failed to produce full length Taf2 in insect cells (data not shown). Thus, insect cell expression was ruled out as a feasible method to produce full length soluble Taf2.

Considering neither *E. coli* nor *Sf9* cells naturally produce yeast Taf2, I hypothesized that the most effective way of producing recombinant Taf2 may be through

expression in its native organism, *S. cerevisiae*. The work of Peter Burgers and colleagues had previously demonstrated that multi-protein complexes could be overexpressed and purified from *S. cerevisiae* to great effect using a high-copy plasmid Gal induction-based system (Gerik et al., 1997). Gal induction of full length Taf2, either in untagged form or containing an MBP-tag at either its N-terminus or its C-terminus resulted in significant protein overexpression (Figure 2.2b and data not shown). Furthermore, amylose resin purification of this overexpressed Taf2 resulted in highly pure full length MBP-tagged Taf2 (Figure 2.2c,e). Attempts to remove the MBP tag by 3C protease cleavage were successful to varying degrees. Specifically, cleavage of amylose resin-bound Taf2-3C_{pro}-MBP resulted in liberation of the majority of the Taf2 protein from the resin. However, the liberated protein was a mixture of free Taf2 and residual Taf2-3C_{pro}-MBP (Figure 2.2d). Attempts at prolonged digestion of free protein in solution with higher concentration of 3C protease failed to achieve greater than a 50:50 mix of free Taf2 to Taf2-MBP. Alternatively, cleavage of MBP-Taf2 with 3C protease resulted in far more efficient cleavage resulting in a near complete cleavage and liberation of free Taf2 (Figure 2.2f). As a result, the MBP-Taf2 variant was used for *in vitro* biochemical experiments (described in Chapter III).

Progress on Generating an Alternative TFIID Purification Strategy

Generating Recombinant TFIID in Sf9 Cells - In addition to generating soluble Taf2, a goal of this dissertation was to develop a method for facile purification of mutant forms of TFIID, specifically to test the function of mutant Taf2 in TFIID function. As described above, current protocols for purification of TFIID are both time and cost intensive. Thus, if I could develop a method where TFIID could be generated using

recombinant expression technologies such that milligram quantities of TFIID could be produced from as little as one liter of insect cell culture or 10 liters of yeast cell cultures instead of the 200-300 liters currently required, this technological development would dramatically increase the ability of scientists to genetically and biochemically dissect the function of TFIID.

While attempting to express Taf2 in *Sf9* cells either by itself or in complex with other TFIID subunits, I generated baculoviruses encoding all of the yeast TFIID subunits. With the exception of Taf2, each baculovirus induced the expression of the encoded TFIID subunits at high levels without any obvious N- or C-terminal truncations (Figure 2.3a). To begin to construct recombinant TFIID, mixtures of virus encoding different TFIID subunits were used to infect Sf9 cells to form subcomplexes that could be subsequently purified. Initially, attempts were made to reconstitute the 5-Taf TFIID core complex (Tafs 4, 5, 6, 9 and 12), described by the Berger and Schultz labs for human TFIID, to determine if I could replicate published data (Bieniossek et al., 2013). According to the Berger and Schultz labs' observations, Taf4 and Taf12 do not co-purify with Taf6 or Taf9 in the absence of Taf5. To test this observation, *Sf9* cells were infected with baculovirus encoding Taf4 and HA_3-$3C_{pro}$-Taf12; Taf4, Taf6, Taf9 and HA_3-$3C_{pro}$-Taf12; or Taf4, Taf5, Taf6, Taf9 and HA_3-$3C_{pro}$-Taf12 and subjected to anti-HA immunopurification (Figure 2.3b). The results of the purification demonstrated that in all three experimental conditions, levels of Taf4 and Taf12 in the extract were depleted upon anti-HA immunopurification and enriched in the immunopurified material (Figure 2.3c and data not shown). However, levels of Taf6 and Taf9 were only depleted from the extract and enriched in the immunopurified material when Taf5 was also co-expressed (Figure 2.3c and data not shown). The resulting 5-Taf

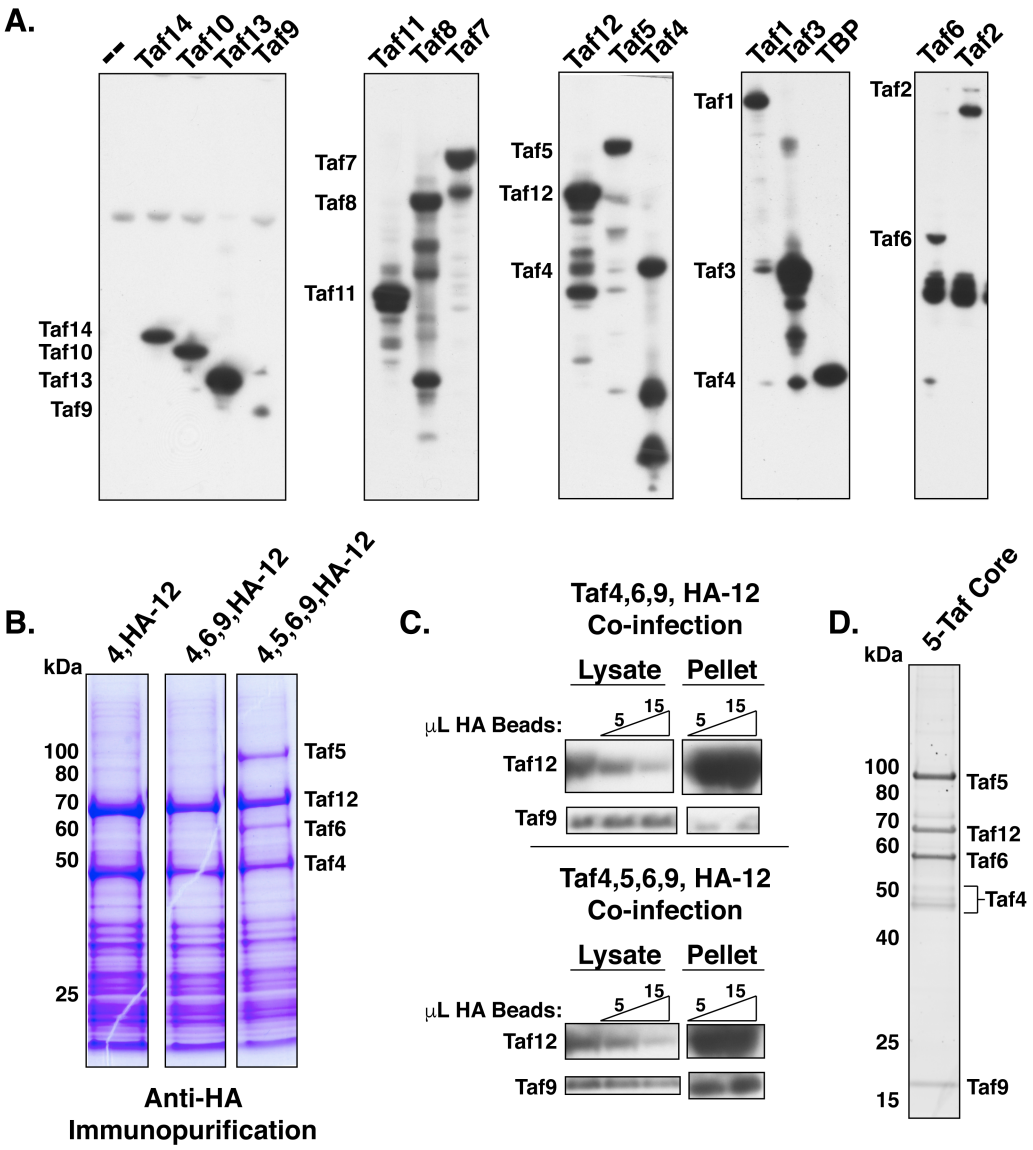


Figure 2.3: Reconstitution of Yeast 5-Taf TFIID Core by *Sf9* Cell Co-Infection. ***A.*** Anti-TFIID subunit immunoblots of *Sf9* cells singly infected with recombinant baculovirus encoding the TFIID subunits. ***B.*** SDS-PAGE and CBB staining of immunopurified Taf subcomplexes generated by co-infection of Sf9 cell with recombinant baculoviruses encoding *TAF4* and HA_3-$3C_{pro}$ (HA)-*TAF12*; *TAF4, TAF6, TAF9* and HA_3-$3C_{pro}$ (HA)-*TAF12;* or *TAF4, TAF5, TAF6, TAF9* and HA_3-$3C_{pro}$ (HA)-*TAF12*. ***C.*** Immunoblots of *Sf9* cell lysates and immunopurified material following co-infection with either *TAF4, TAF6, TAF9* and HA_3-$3C_{pro}$ (HA)-*TAF12* or *TAF4, TAF5, TAF6, TAF9* and HA_3-$3C_{pro}$ (HA)-*TAF12*-encoding baculoviruses. Input lysate or lysates following binding to either 5 or 15μL of 4mg/mL anti-HA-coupled proteinA sepharose were assayed for Taf12 and Taf9 as a measure of antigen and co-purifying protein depletion, respectively. Proteins bound to anti-HA beads were assayed to demonstrate level of enrichment in the pull-down. ***D.*** SyproRuby gel stain of highly purified 5-Taf TFIID core complex derived from *Sf9* cell co-infection.

TFIID core complex was then subjected to size-exclusion chromatography for further purification. This analysis separated the intact 5-Taf complex from excess Taf4-Taf12 complex and revealed that the 5-Taf complex displays equal stoichiometry for all of the different components (Figure 2.3d). These results recapitulated the Schultz and Berger labs' published data and suggest that the yeast TFIID core likely assembles in a similar manner to the human TFIID core complex.

Also, according to the Berger and Schultz labs, addition of Taf8 and Taf10 to the TFIID core is important for a conformational change in the complex. They hypothesized that this structural transition is critical for the assembly of the holo-TFIID complex. Thus, to recapitulate their ability to form the 7-Taf complex (5-Taf TFIID core + Taf8 and Taf10), I co-infected Sf9 cells with viruses encoding all of the 5-Taf TFIID core complex constituents as well as viruses encoding Taf8 and Taf10. This 7-Taf complex was subjected to anti-HA immunopurification and size-exclusion chromatography. While I was able to immunopurify a 7-Taf complex, this complex displayed characteristics of protein aggregation. First, despite achieving near complete cleavage of the HA_3-$3C_{pro}$-tag on Taf12 when purifying the 5-Taf TFIID core complex, the 7-Taf complex displayed both a mix of cleaved and uncleaved Taf12 (data not shown). Second, this 7-Taf complex would not enter the size-exclusion chromatography column matrix but instead eluted within the void volume of the column (data not shown). While additional Tafs could be added to this 7-Taf assembly, namely Taf3, Taf11 and Taf13, all of these complexes displayed the same signs of aggregation. In an effort to create larger Taf assemblies that did not display this protein aggregation phenotype, *Sf9* cells were infected with viruses encoding the 5-Taf TFIID core complex constituents

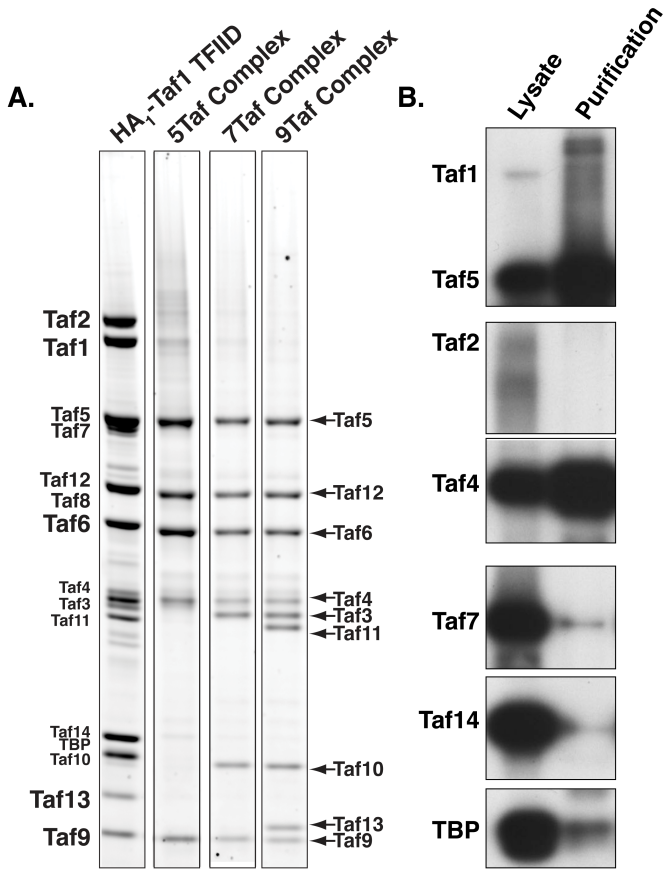


Figure 2.4: Reconstitution of Taf Subcomplexes in Sf9 cells. ***A.*** SyproRuby stain of highly purified native HA_1-Taf1 TFIID from *S. cerevisiae* and Taf subcomlexes from baculovirally co-infected *Sf9* cells. ***B.*** Sf9 cells co-infected with baculoviruses encoding every TFIID subunit were subjected to anti-HA immunopurification, as described in Methods section. Both input (lysate) and purified material were analyzed by immunobloting with anti-Taf and anti-TBP IgGs.

and either Tafs 3 and 10 or Tafs 3, 10, 11 and 13. The resulting complexes were subjected to anti-HA immunopurification and size-exclusion chromatography. Unlike the Taf8/Taf10 containing subcomplexes, both the 5-Taf TFIID core + Taf 3 and 10 (7Taf) complex as well as the 5-Taf TFIID core + Taf 3, 10, 11 and 13 (9Taf) complex entered the size-exclusion chromatography column and eluted as a homogeneous complex (data not shown). When analyzing these complexes by SDS-PAGE and gel staining (Figure 2.4a), with the exception of Taf13, all of the TFIID subunits ran at the same mobility as TFIID purified from *S. cerevisiae*. Taf13 appeared to migrate at a faster mobility likely due to proteolysis. In addition, neither complex displayed the stoichiometry observed in TFIID purified from *S. cerevisiae*. Specifically, Taf3, Taf11 and Taf13 are thought to be one mole per mole of TFIID while Tafs 4, 5, 6, 9, 10 and 12 are thought to be two moles per mole of TFIID. The recombinant TFIID subcomplexes display an approximate subunit stoichiometry of 1:1 for all of the subunits.

Attempts at adding additional TFIID subunits to this complex were unsuccessful for several reasons. First, when infecting with all 15 viruses, production of both Taf1 and the truncated form of Taf2 severely decreased (Figure 2.4b). Second, addition of Taf8, as described above caused the complexes to adopt biochemical phenotypes suggestive of protein aggregation (data not shown). Third, despite being expressed at high levels, Tafs 7 and 14 and TBP failed co-purify with the 9Taf complex (Figure 2.4b). Regardless, these results represent significant progress in attempts to generate recombinant TFIID.

Generating Recombinant TFIID in S. cerevisiae - Considering that yeast were capable of overexpressing large quantities of soluble Taf2 by Gal induction, I hypothesized that I could instead use multi-subunit overexpression in yeast to generate recombinant TFIID. Taking a similar approach as described for Sf9 cells, the goal would be to dramatically overexpress every TFIID subunit in the same cell at the same time. This would allow native yeast protein assembly pathways to construct the TFIID complex *in vivo* leading to dramatically increased yield.

To this end, I constructed a two-plasmid system (Figure 2.1; described in Methods). The rationale for this two-plasmid system was that one plasmid would be the donor plasmid and the other plasmid would the acceptor plasmid. The donor plasmid (pJFGDUAL) would contain a fully competent dual expression cassette to which I would clone the ORFs of two distinct TFIID subunits. The acceptor plasmid (pJFLEU2, pJFTRP1 or pJFURA3) would contain a high copy origin of replication and the selectable marker required for growth in amino acid drop out medium. In addition, the acceptor plasmid

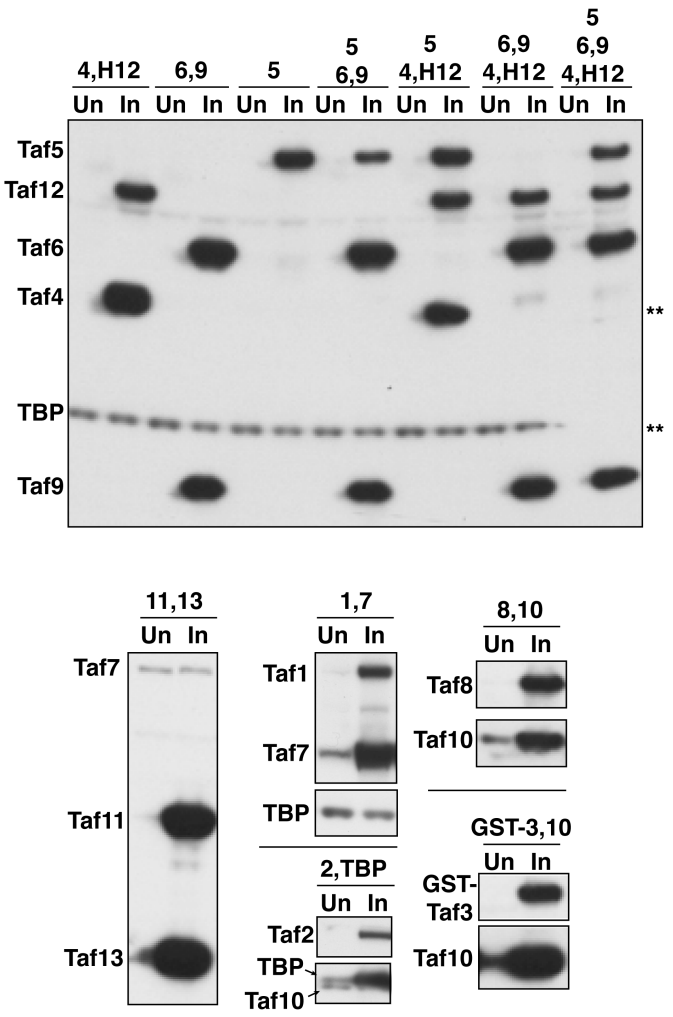


Figure 2.5: Multi-subunit Overexpression in Yeast. Yeast strains harboring individual or combinations of galactose inducible dual expression plasmids were harvested before (Un) and after 4hrs of galactose induction (In). Whole cell extracts were analyzed by immunoblotting with anti-Taf and anti-TBP IgGs. Numbers correspond to Tafs encoded in dual expression cassette (e.g. 4=Taf4). H=HA_3-$3C_{pro}$-tag. ** = loss of signal because of inappropriate cut in PVDF membrane.

would contain four distinct restriction enzyme sites that could be used as entry points for the dual expression cassette. In theory, a single acceptor plasmid could overexpress 8 TFIID subunits. By using multiple plasmids with different selectable markers, all of the TFIID subunits could be expressed in the same cell at the same time.

The initial plasmid design included two $CYC1_{term}$ sequences for each dual expression cassette. Using these plasmids, overexpression of one or two TFIID subunits in a single cell at the same time was highly successful (Figure 2.5). Furthermore, expressing two and three different plasmids in the same cell did not markedly diminish total protein output. However, despite repeated attempts to create multiple dual expression cassettes into the same acceptor plasmids, I was never able to produce the correct sized plasmid. In fact, the plasmids that were purified from *E. coli* showed significant recombination (data not shown). In the one

case where a plasmid was generated that purportedly contained four Tafs in the same plasmid, Gal induction experiments revealed that only three of the Tafs were expressed, likely as a result of recombination (data not shown). These results led me to hypothesize that recombination was a result of the multiple copies of the *CYC1*$_{term}$ in the dual expression cassettes. Indeed, despite ultimately having success generating the pJFGDUAL plasmid, I initially had difficulty ligating the second copy of the *CYC1*$_{term}$ into the dual expression cassette and subsequently sequencing the plasmid.

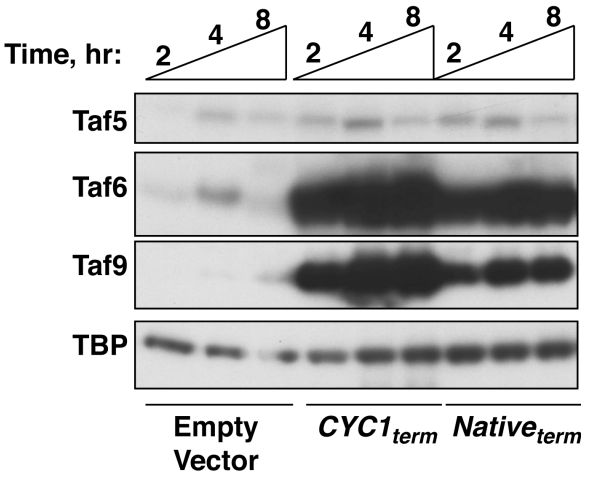


Figure 2.6: Comparison of Protein Output Between *CYC1* and Native Terminator Containing Expression Vectors. *S. cerevisiae* strains harboring an empty vector and dual expression plasmids encoding *TAF6* and *TAF9* containing either the *CYC1*$_{term}$ or the natural *TAF6*/*TAF9* terminators were induced with galactose for 2, 4 or 8 hrs. Whole cell extracts derived from the induced cells were analyzed by immunoblotting with anti-Taf and anti-TBP IgGs

As an alternative to the *CYC1*$_{term}$, I re- cloned all of the TFIID subunits (except for Taf2) such that they would contain ~100bp beyond the documented 3' untranslated sequence. Then, the Taf ORF + terminator sequences were ligated into pJFGDUAL sans the *CYC1*$_{term}$ sequences and the resulting dual expression cassettes were transferred to the acceptor plasmids. This change in plasmid design then allowed me to insert multiple dual expression cassettes into the same acceptor plasmid. To determine if the native termination sequences were as effective as the *CYC1*$_{term}$ sequence, overexpression of acceptor plasmids containing *CYC1*$_{term}$ sequences with Taf6 and Taf9 was compared to overexpression of acceptor plasmids containing the native Taf6 and Taf9 terminator sequences (Figure 2.6). While the *CYC1*$_{term}$ sequences performed better than the native

terminator sequences, considering I was unable to express more than two TFIID subunits in a single acceptor plasmid containing the *CYC1*$_{term}$ sequences, I moved forward with the native terminator sequences. Three acceptor plasmids, and associated intermediate plasmids, were generated that contained all of the TFIID subunits. The following dual expression cassettes were sequentially cloned into pJFTRP1: 1) Taf6 and Taf9, 2) Taf5 and Taf14, 3) Taf8 and Taf10 and 4) Taf4 and (HA_3-$3C_{pro}$-)Taf12. The following dual expression cassettes were sequentially cloned into pJFURA3: 1) Taf11 and Taf13, 2) Taf3 and Taf10 and 3) Taf1 and Taf7. MBP-$3C_{pro}$-Taf2 and TBP were cloned into the pJFGDUAL containing the *CYC1*$_{term}$ sequences and the dual expression cassette was inserted into pJFLEU2. All of these plasmids were transformed into yeast in combination or individually and tested for ability to overexpress TFIID subunits by Gal induction (Figure 2.7). Overall, I was able to observe significant protein induction for all of the TFIID subunits when they were expressed on a single plasmid or in two plasmid combinations. However, I did not observe induction of all of the TFIID subunits in the three-plasmid combination. In addition, when focusing on the Tafs encoded in the first dual expression cassettes introduced into the acceptor plasmids (Taf6 and Taf9 or Taf11 and Taf13), as additional dual expression cassettes were introduced into the acceptor plasmid, the level of protein overexpression was significantly reduced. In totality, these results demonstrate that in separate strains, I could achieve significant overexpression of all of the TFIID subunits; however, I could not achieve significant overexpression of all of the TFIID subunits in the same strain.

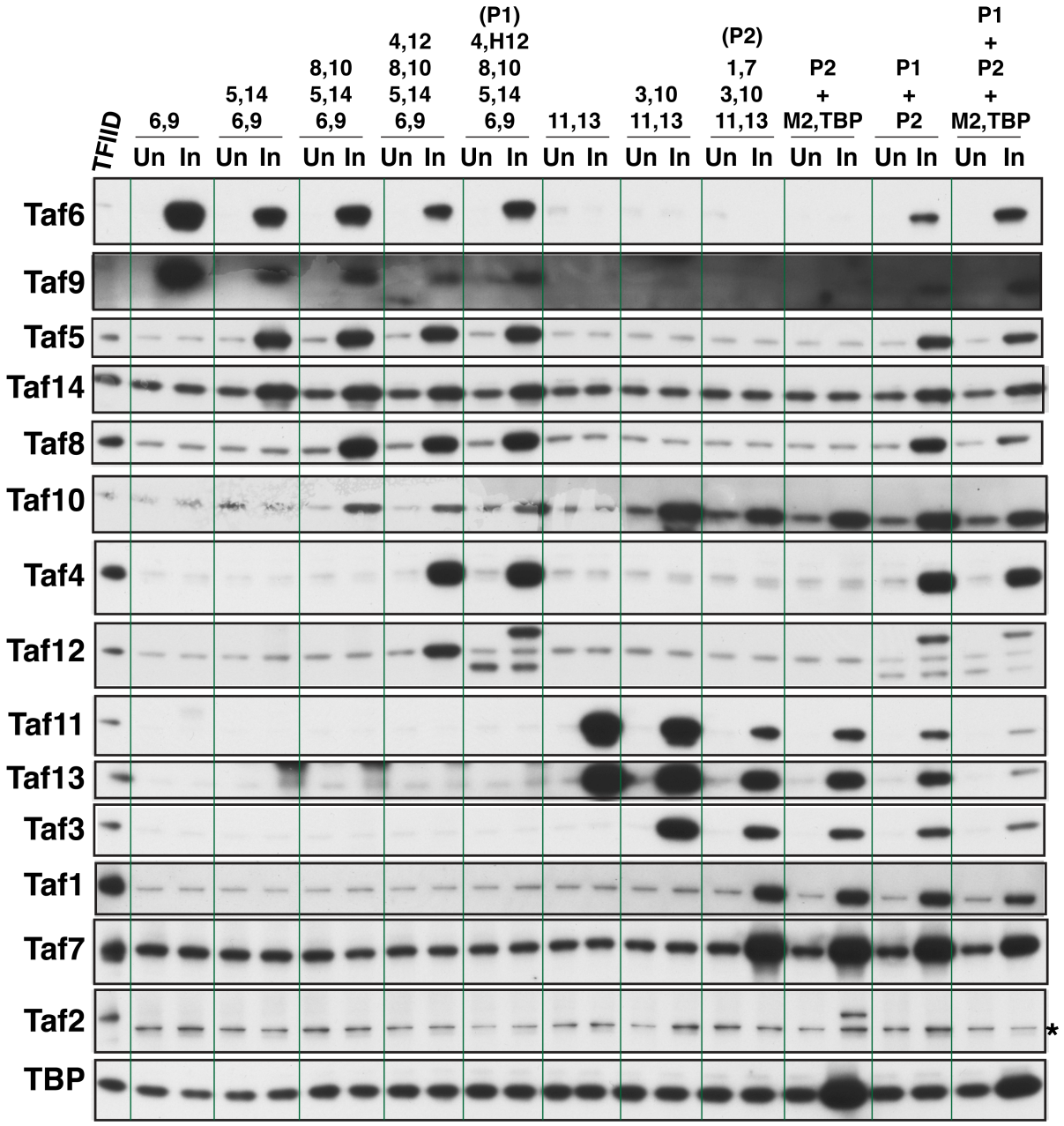


Figure 2.7: Iterative Addition of Dual Expression Cassettes for Overexpression of All TFIID Subunits in the Same Cell at the Same Time. *S. cerevisiae* strains, transformed with multi-subunit overexpression plasmids either individually or in combination, were harvested pre-galactose induction (Un) or after 4hrs galactose induction (In). Whole cell extracts derived from these strains were subjected to immunoblotting with anti-Taf and anti-TBP IgGs. Overexpression plasmids were derived from dual expression cassettes containing the native terminators for all TFIID subunits with the exception of the MBP-*TAF2*,*TBP (SPT15)* encoding plasmid. Plasmid 1 (P1) was derived from iteratively integrating dual galactose inducible expression cassettes encoding 1) Tafs 6 and 9, 2) Tafs 5 anad 14, 3) Tafs 8 and 10 and 4) Taf4 and (HA_3-$3C_{pro}$)-Taf12. Plasmid 2 (P2) was derived from iteratively integrating dual galactose inducible exrpression cassettes encoding 1) Tafs 11 and 13, 2) Tafs 3 and 10 and 3) Tafs 1 and 7. Purified TFIID was included as a positive control. Un = uninduced, In = induced, H= HA_3-$3C_{pro}$-tag , M=MBP-$3C_{pro}$-tag, *=non-specific band.

Alternative Taf1-TAP TFIID Purification Optimization – As an alternative to generating recombinant TFIID or purifying TFIID using our anti-HA immunopurification lab protocol, I hypothesized that if I were to optimize the Taf1-TAP TFIID purification protocol, then I would be able to at least decrease the cost of TFIID purification. IgG sepharose is significantly less expensive than anti-HA (12CA5) covalently coupled to proteinA sepharose. In addition, rabbit IgG can easily be covalently coupled to cyanogen bromide activated sepharose 4B beads in-house, decreasing the cost further (March et al., 1974). The challenge with optimizing this protocol is that I would need to figure out how to eliminate the step that results in dissociation of Taf2, Taf8 and Taf14 from the complex

To determine which step is likely to result in dissociation of Taf2, Taf8 and Taf14, I investigated the unpublished data of Post-doc Manish Tripathi who had invested extensive time and energy into the purification of Taf1-TAP TFIID. When looking through his data, I observed that he had already identified the step in the purification that leads to Taf2, Taf8 and Taf14 dissociation, he just had not realized it. In the first step of the purification, he was able to bind Taf1-TAP TFIID to IgG sepharose and elute with TEV protease without any loss of Taf2. However, in the next step (calmodulin sepharose), he observed that if he used too high of a salt concentration, Taf2 would dissociate from the complex (data not shown). Instead of omitting this step in the purification, he lowered the salt concentration and proceeded. However, if he tried to separate the complex further using ion exchange chromatography, Taf2, Taf8 and Taf14 would dissociate from the complex. This particular result was quite illuminating to me because TFIID purified using anti-HA immunopurification can be subjected to ion exchange chromatography without disrupting complex integrity (Sanders et al., 2002a). Based on this body of data, I

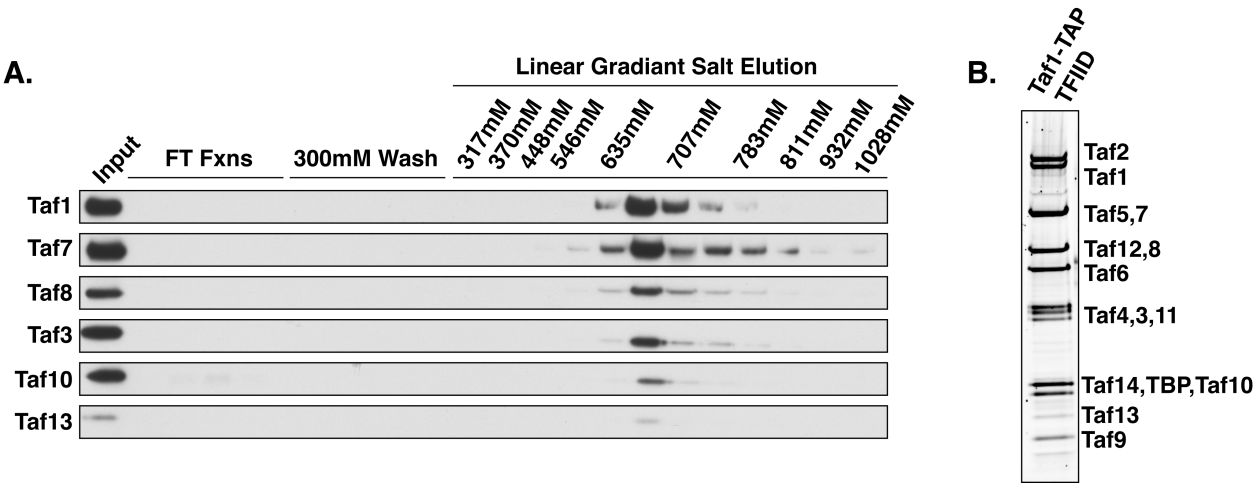


Figure 2.8: Alternative Taf1-TAP Purification Protocol Yields Intact TFIID. ***A.*** *S. cerevisiae* cells expressing Taf1-TAP were lysed and extracts bound to IgG sepharose and eluted via TEV cleavage. The TEV eluate was then bound to a cation exchange column, washed and eluted with a linear gradient (300mM to 1M) of potassium acetate. The fractions were analyzed for conductivity to determine potassium acetate concentration as well as analyzed by immunoblot using anti-Taf IgGs. Input= TEV eluate, FT Fxns= Flow Through (unbound) fractions. ***B.*** SyproRuby stained gel of Taf1-TAP TFIID purified with alternative TAP purification protocol.

hypothesized that the step that was disrupting complex integrity was the calmodulin sepharose binding step. Therefore, if this step were removed, TFIID would remain intact.

For simplicity, I utilized the buffer system and initial purification steps developed by graduate student Steven Sanders (unpublished data; outlined in Methods section). After optimizing the volume of beads required for IgG sepharose binding and concentration of TEV protease required to elute Taf1-TAP TFIID, instead of proceeding to the calmodulin sepharose step, I immediately bound the TEV eluate to the cation exchange column that our lab routinely uses following HA_1-Taf1-TFIID anti-HA immunpurification. The column was developed with a linear potassium acetate salt gradient from 300mM to 1M and the resulting fractions were immunoblotted for TFIID subunits, including Taf8 as a surrogate for the putative Taf2, Taf8 and Taf14 subcomplex (Figure 2.13). In addition, the TFIID peak fraction was analyzed by SyproRuby gel stain. These analyses suggested that

my hypothesis was correct. Following a protocol that omitted the calmodulin binding step allowed me to separate the IgG sepharose purified Taf1-TAP TFIID on a cation exchange column without resulting in dissociation of Taf8 from the rest of the complex. In fact, the TFIID complex eluted at approximately 650mM potassium acetate, the same elution point observed for cation exchange chromatography of anti-HA immunopurified TFIID (data not shown). Based on these results, omitting the calmodulin binding step makes the TAP purification procedure more reliable at producing TFIID with the correct stoichiometry of Taf2, Taf8 and Taf14.

DISCUSSION

In this chapter, I document attempts to purify soluble full length Taf2 and to identify alternative methods for efficient TFIID purification. Successful purification of full length Taf2 would enable to me to perform *in vitro* biochemical assays to identify functions of Taf2. These functions could then be genetically and biochemically dissected to identify loss- or gain-of-function variants. Loss- and gain-of-function reagents are key for understanding protein functions *in vivo* and the identification of such reagents would allow me to determine the role of Taf2 in TFIID-dependent transcription regulation. Additionally, developing an efficient and robust method for the purification of TFIID in both yeast and humans is an urgent need for scientists studying transcription initiation mechanism *in vitro*. Such a method would enable more laboratories to analyze the structure and function of the TFIID complex *in vitro*. Ultimately, the ability to generate recombinant TFIID would lead to the generation of mutant TFIID variants, important tools

for dissecting its mechanism of action regarding its promoter recognition and co-activator function. Currently, only one laboratory has ever published biochemical analysis of purified mutant TFIID forms (Lauberth et al., 2013) and their results have provided important insights into how TFIID engages chromatinized DNA. This study underscores the need for innovation in TFIID purification in order to advance TFIID research.

Attempts to express recombinant Taf2 in both *E. coli* and *Sf9* cells resulted in similar and potentially informative results. In *E. coli*, I was able to successfully express, in large quantities, soluble Taf2 fragments that did not include the Taf2 C-terminus (Figure 2.2a). However, I was not able to produce soluble full length Taf2 at all or significant quantities of fragments of Taf2 that contained the Taf2 C-terminus. In *Sf9* cells, attempts at expression and purification of full length Taf2 led to a C-terminally truncated version of the protein (Figure 2.2b). Taken together, it is possible that the C-terminus of Taf2 adopts a conformation that renders it insoluble in *E. coli* and causes proteolysis in *Sf9* cells. Consistent with this hypothesis, Kabani et al. reported that their attempts to express a Taf2 C-terminal fragment in *E.coli* resulted in an insoluble protein (Kabani et al., 2005). In addition, this region of yeast Taf2 corresponds to the predicted intrinsically disordered region of metazoan Taf2 (Zhang et al., 2015) which could play a role in the proteins solubility or integrity in exogenous expression systems.

I was able to resolve the issues of solubility and C-terminal truncation by overexpressing Taf2 in and purifying from *S. cerevisiae* (Figure 2.2d,c,e). Expressing MBP-Taf2 using a high copy plasmid with the *GAL1* inducible promoter produced significant amounts of protein that could be subsequently purified using amylose resin. In general, the amylose resin does not appear to specifically bind endogenous yeast cell proteins. As a

result, a one step amylose resin purification of MBP-Taf2 yields highly pure material for subsequent biochemical analysis.

In order to generate recombinant TFIID, I employed two similar overexpression strategies. First, I attempted to infect Sf9 cells with baculovirus encoding TFIID subunits. Using this method, I was able to individually overexpress the full-length versions of every TFIID subunit with the exception of Taf2 (Figure 2.3a). Then, I used these viruses to express and purify Taf subcomplexes containing up to 9 subunits (Figure 2.4a). Importantly, I also demonstrated that yeast Tafs 4, 5, 6, 9 and 12 can form the TFIID core complex in a similar manner to what has been published for the human TFIID core complex (Figure 2.3b-d). The existence of this core structure strengthens the conclusion that yeast TFIID is highly evolutionarily conserved at the amino acid, structural and functional level.

The main hurdle that I could not clear with regard to growing this complex beyond 9 subunits was the inability to efficiently overexpress both Taf1 and Taf2 when the rest of the other TFIID subunits were overexpressed (Figure 2.4b). Our lab has demonstrated that even when the ability of Taf1 to interact with holo-TFIID is disrupted, it still maintains its ability to bind Taf7 (Singh et al., 2004). In addition, we have also shown the Taf1 TAND directly interacts with TBP (Bai et al., 1997). Thus, if Taf1 is not expressed, it is unlikely that either Taf7 or TBP stably associate with the complex. At the time I performed this study, the finding that a Taf2-Taf8-Taf10 subcomplex existed had not been published (Trowitzsch et al., 2015). In addition, the fact that the Taf1-TAP TFIID purification protocol resulted in a preparation with not only substoichiometric levels of Taf2 but also substoichiometric levels Taf8 and Taf14 had escaped my attention (Papai et al., 2009). Together with the fact that a mutant form of Taf1 can co-precipitate the entire

TFIID complex with the exception of Taf2, Taf8 and Taf14 (Singh et al., 2004), these data suggest that in the absence of Taf1 and Taf2, neither Taf8 nor Taf14 should properly associate with the TFIID complex. Therefore, it is my supposition that in the absence of Taf1 and Taf2, at least some of the binding surfaces required for Tafs 7, 8 and 14 as well as TBP are missing such that these subunits cannot stably associate with the remaining TFIID subunits.

The fact that a Taf8-Taf10 subcomplex could co-purify with the 5-Taf TFIID core but in an aggregated state is puzzling. It is possible that this aggregated state reflects the conformational change proposed by the Schultz and Berger labs insofar as binding of Taf8 and Taf10 to the 5-Taf TFIID core results in the exposure of surfaces that result in aggregation if the appropriate TFIID subunit is not available to bind them. Alternatively, it is possible that Taf10 mediates the association of the 5-Taf TFIID core with the Taf8-Taf10 subcomplex, and as a result, in the absence of Taf2, Taf8 facilitates protein aggregation. Considering full length Taf2 could not be efficiently expressed in Sf9 cells, this hypothesis could not be tested, at least within an *Sf9* co-expression paradigm.

The other method I employed to generate recombinant TFIID was through simultaneous overexpression of every TFIID subunit in yeast by Gal induction. This method was exceptionally tantalizing because I had demonstrated that I could massively overexpress and purify Taf2 in this manner (Figure 2.2b,c,e). In addition, the Burgers lab utilized this approach to co-express and purify the 5-subunit replication factor C complex (Gerik et al., 1997). Even though the 15-subunit TFIID complex is significantly more complicated, in theory overexpression of every TFIID subunit in yeast should be possible.

I successfully designed, constructed and implemented a multi-ORF overexpression system with significant success. The dual ORF expression cassettes dramatically overexpressed all TFIID subunits tested (Figure 2.5). In addition, I was able to transform a single strain with multiple co-expression plasmids, each of which drove overexpression of the encoded subunits in the same strain. As a result, this system can massively overexpress up to 6 individual TFIID subunits all at once. Attempts to massively overexpress all of the TFIID subunits in the same strain at the same time were not successful for two reasons (Figure 2.7). First, Taf2 was not overexpressed when all of the TFIID subunits were expressed at the same time. Second, as the number of TFIID subunit ORFs included in a single overexpression vector increased, the total amount of protein produced decreased. Even if Taf2 were overexpressed to similar levels of the other proteins, without dramatic overexpression, the utility of this approach is diminished.

Despite being unable to reconstitute TFIID in Sf9 yeast cells through concurrent overexpression and purification, the overexpression systems described in this chapter still have the potential for TFIID reconstitution. Instead of trying to reconstitute TFIID all at the same time, Taf subcomplexes could be used to reconstitute the complex *in vitro*. Sf9 and (likely) yeast cells readily express the 5-Taf TFIID core complex. In addition, I am able to overexpress full-length versions of every additional TFIID subunit in yeast. By iterative overexpression and pull-down, it is likely that I would be able to identify additional Taf subcomplexes that could be mixed with the 5-Taf TFIID core resulting in recombinant TFIID. This method was employed by the Berger, Tora and Schultz labs to form a 8-Taf complex composed of Tafs 2, 4, 5, 6, 8, 9, 10 and 12 (Trowitzsch et al., 2015). The first such complex I would attempt to reconstitute is a Taf2, Taf8, Taf10 and Taf14 complex. In yeast,

there is evidence that Taf2, Taf8 and Taf14 form a subcomplex (Papai et al., 2009; Singh et al., 2004). In addition, the yeast Taf8 and Taf10 histone fold domains directly interact (Gangloff et al., 2001b). Therefore, it is likely that through co-expression of all four subunits, I would be able to purify this subcomplex in a soluble form. Similar approaches have been successful in reconstituting the modules of the Mediator co-activator complex (Baumli et al., 2005; Imasaki et al., 2011; Koschubs et al., 2010, 2009, Larivière et al., 2012, 2008, 2006; Takagi et al., 2006). In fact, the Cramer laboratory has reconstituted a 15-subunit Mediator complex that is active for *in vitro* transcription assays (Plaschka et al., 2015). Alternatively, considering I am able to overexpress all of the TFIID subunits in different yeast strains, it may be possible to mix lysates containing massively overexpressed TFIID subunits, allow TFIID to form in the extract and then purify the complex. I am hopeful that through persistence, reconstitution of TFIID using the methodology developed herein is possible.

When I was unable to produce recombinant TFIID through overexpression techniques, I hypothesized that if I could optimize the Taf1-TAP TFIID purification protocol, I could at least make strides in making TFIID more accessible to scientists interested in transcription initiation mechanism. The TAP tag purification procedure is widely used, having been cited over 900 times in PubMed (Rigaut et al., 1999). Our lab had preferred not to use the TAP-tag protocol because the quality of TFIID was not as high compared to TFIID purified by our anti-HA immunopurification protocol. After examining old unpublished data from our laboratory, I surmised that I could improve the TAP-tag purification protocol, with regard to TFIID, by eliminating the calmodulin sepharose binding step. By removing this step, I was able to eliminate the dissociation of subunits

Taf2, Taf8 and Taf14 our lab had observed when purifying TFIID by TAP tag (Figure 2.8). This alternative, two-step TFIID purification protocol does effectively purify TFIID, albeit not to the same purity achieved by our standard anti-HA immunopurification. In addition, the yield is not as robust and the amount of IgG sepharose required for purification can be quite expensive, even if the IgG sepharose is generated in the laboratory rather than purchased commercially. Thus, in order for this protocol to be an improvement over our current anti-HA immunopurification protocol, the protocol must be optimized to increase yield and purity.

I believe the Taf1-TAP protocol can be improved in the following ways. First, I never optimized cell lysis conditions for the Taf1-TAP protocol. The lysis buffer I used for the TAP purification procedure was dramatically different than the protocol used for our standard HA_1-Taf1 TFIID purification protocol. The primary difference between these two buffers is salt concentration. The salt concentration is considerably higher in the HA_1-Taf1 TFIID purification protocol, which may lead to increased TFIID extraction from the cells. Second, the HA_1-Taf1 TFIID purification protocol includes an ammonium sulfate precipitation step that concentrates the protein up to 5-fold. It is possible that by performing this step, I could use a smaller volume of beads to capture as much TFIID as possible. This would reduce the cost and technical difficulty of the purification and make it more accessible to other researchers. Third, the BioRex70 ion exchange chromatography step is exceptionally good at depleting TFIID from yeast extract and results in a roughly 10-fold purification of TFIID (Poon and Weil, 1993). Thus, it may be prudent to pre-fractionate the yeast extract on the BioRex70 column prior to IgG sepharose binding. This may both increase yield as well as improve the purity of the final preparation. Ultimately, I believe

that the Taf1-TAP TFIID protocol is a viable method for the purification of TFIID and, with minor modifications, may become the dominant and reliable method used by laboratories wishing to study the biochemistry of this important complex.

CHAPTER III

THE C-TERMINUS OF TAF2 MEDIATES STABLE ASSOCIATION OF TAF14 INTO THE YEAST TFIID COMPLEX

Considerations to Deciphering the Role Taf2 Plays in TFIID-dependent Transcription Regulation

As described in Chapter I, while significant similarities exist between yeast and metazoan TFIID, yeast display some significant differences to its metazoan counterparts with regard to TFIID structure and promoter architecture. First, the metazoan Taf1 double bromodomain and the Taf3 plant homeobox domain (PHD) finger, domains that directly interact with modified chromatin, are missing from the yeast homologs (Gangloff et al., 2001b; Poon et al., 1995; Reese et al., 1994). Second, yeast TFIID contains an additional subunit, Taf14 (Henry et al., 1994; Sanders et al., 2002a). Third, in metazoan TFIID, Tafs 1 and 2 bind to the initiator (INR) core promoter element (Chalkley and Verrijzer, 1999; Verrijzer et al., 1995, 1994) and Tafs 6 and 9 bind to the downstream promoter element (DPE) (Burke and Kadonaga, 1997). Neither the INR nor the DPE have been unambiguously identified in the yeast system (Hahn and Young, 2011).

Despite these differences, yeast have evolved mechanisms to achieve similar TFIID promoter-DNA binding and chromatin recognition activities. Instead of having a double bromodomain covalently attached to Taf1, the double bromodomain protein Bdf1 is a TFIID associated protein (Ladurner et al., 2003; Matangkasombut et al., 2000) and its occupancy in genome-wide analyses correlates with Taf1 occupancy (Rhee and Pugh,

2012). While not conserved as a TFIID subunit, Taf14 is conserved through its YEATS domain (Schulze et al., 2009). This domain, which is also present in the super elongation complex proteins AF-9 and ENL (Cairns et al., 1996; Lin et al., 2010; Welch and Drubin, 1994), was recently shown to bind to acetylated and crotonylated lysine 9 of histone H3 (H3K9) (Andrews et al., 2016; Shanle et al., 2015), marks associated with active transcription (Bonnet et al., 2014; Li et al., 2016). In regards to *S. cereivisae* (Sc) TFIID promoter recognition, footprinting analyses have demonstrated that both metazoan and ScTFIID display extended footprints with contacts spanning nearly 100bp (Cianfrocco et al., 2013; Oelgeschläger et al., 1996; Sanders et al., 2002a). ScTFIID histone fold pairs Taf4/Taf12 and Taf6/Taf9 also display *in vitro* DNA binding activities, although binding has not been shown to be sequence specific (Shao et al., 2005). In addition, structural analyses with ScTFIID-TFIIA-activator in complex with enhancer/promoter-DNA positions DNA in contact with the C-terminus of Taf2 (Papai et al., 2010). Taken together, these observations demonstrate that TFIID promoter-DNA and chromatin interaction activities are maintained in the yeast system. Yet, how these activities contribute to TFIID transcriptional activation function remains undefined.

In addition to its role in TFIID, Taf14 is also a subunit or associated protein of six other transcription-related complexes (Cairns et al., 1996; Henry et al., 1994; Kabani et al., 2005; Sanders et al., 2002b). While Tafs 1 through 13 are essential for life (Klebanow et al., 1997, 1996; Moqtaderi et al., 1996; Poon et al., 1995; Ray et al., 1991; Sanders and Weil, 2000), *taf14* null cells display temperature sensitive (Ts) growth and defects in expression of *GAL* and DNA repair genes (Cairns et al., 1996; Shanle et al., 2015). Still, how Taf14 contributes to TFIID structure or function remains poorly understood because mutations in *TAF14*

could impact the function of all of the transcription related complexes with which it associates. To begin to understand *TAF14* function *in vivo*, we need true separation-of-function variants that can specifically dissociate Taf14 from a single complex without disrupting its ability to perform its other functions.

We, as a field, have minimal understanding of Taf2 function, despite the fact that it was the first yeast Taf discovered (Poon et al., 1995; Ray et al., 1991; Verrijzer et al., 1994). Our lab was the first to discover the location of Taf2 within the TFIID complex (Papai et al., 2009). In addition to its INR binding function, metazoan Taf2 directly interacts with Taf8 and this interaction is critical for Taf2 to localize to the nucleus (Trowitzsch et al., 2015). For RNA polymerase II to clear the promoter and begin productive elongation, TFIID appears to isomerize in a Taf2-dependent manner to release TFIID from downstream promoter sequences (Yakovchuk et al., 2010; Zhang et al., 2015). However, none of these biochemical activities have been genetically dissected. In fact, Taf2 has never been subjected to structure-function analysis in any system. During the execution of this dissertation project, a cryoEM structure of human TFIID allowed Louder et al. to describe the structure of Taf2 in molecular detail (Louder et al., 2016). But, this structure was not interrogated genetically or biochemically. Consequently, we have minimal understanding of how Taf2 contributes to TFIID-dependent transcription *in vivo*.

In this chapter, I describe a systematic molecular genetic dissection of Taf2 to address this gap in knowledge. These analyses identified a genetic interaction between *TAF2* and *TAF14*. I demonstrate that these two subunits directly interact and define the interaction domains in both subunits. Despite Taf14 being present in multiple copies per TFIID subunit, mutation in Taf2 can completely disrupt the ability of Taf14 to associate with the

TFIID complex. Taf14-less TFIID containing cells display defects in growth and transcript abundance for the highly transcribed TFIID-dominated ribosomal protein-encoding genes. Furthermore, my data indicate that the Taf14 YEATS domain contributes to TFIID function.

METHODS

Yeast Expression Vector Construction and Cell Manipulations

All general cloning techniques (i.e. ligations, transformation, PCR with proofreading polymerases and addition of restriction ends by PCR) were performed as described in Chapter II Methods. Site-directed mutants were generated by PCR using the gene sequence overlap extension method (Ho et al., 1989) and ligated into Taf2 using either the engineered 5'-SpeI or 3' SalI restriction sites or the natural restriction sites located within the body of the *TAF2* coding sequence (EcoRI, EcoRV, BamHI, XhoI, HindIII). The site-directed mutants were cloned into the p415ADH expression plasmid (Mumberg et al., 1995) containing an N-terminal HA_3-SV40 Large T antigen nuclear localization sequence (NLS)-tag (HA_3NLS). Each TFIID subunit ORF (DNA fragments described in Chapter II Methods) was cloned into p413GPD MCS using compatible restriction ends. For all DNA fragments containing a 3' KpnI restriction end, the fragment overhang was filled in with Klenow fragment (NEB) and blunt-end ligated into the EcoRV restriction site within the MCS. Furthermore, a 5'XbaI-ClaI-$FLAG_2$-NLS-SpeI-3' fragment was cloned 5' and in-frame of the *TAF14* ORF in p413GPD *TAF14* and all *TAF14* site-directed mutant variants. The *TAF2-TAF14* chimeras were generated by PCR-based sequence overlap extension (Ho et al., 1989). The chimeras were engineered to contain three glycine residues separating

the *TAF2* sequence from the *TAF14* sequence. The fragment was designed as a 5'-SpeI-*TAF2*-Gly_3-*TAF14*-SalI-3' fragment and ligated into the corresponding restriction sites within p415ADH HA_3NLS plasmid. All constructs were sequence verified.

Cell growth and transformation protocols were performed as described in Chapter II Methods. For plasmid shuffle assays, cells were grown on SC -Leu, His or both including 0.1% 5-fluoroorotic (5-FOA) acid (Boeke et al., 1987).

All strains generated for this chapter were derivatives of BY4741. A *taf2* null strain (JFTAF2del) was generated by co-transforming BY4741 with a p416ADH *TAF2* covering plasmid and a linearized *TAF2* deletion cassette. The *TAF2* deletion cassette replaced -233 to +4224 relative to the start codon of the *TAF2* ORF with the hygromycin B resistance cassette from pAG32 (Goldstein and McCusker, 1999). Co-transformants were sequentially plated on SC-Ura followed by replica plating to YPD + 300mg/mL hygromycin B. Hygromycin B^+ and Ura^+ clonal isolates were confirmed as *taf2* null using plasmid shuffle genetic complementation. An HA_1-TAF1 strain (JFHATAF1) was generated using ends-in integration of a HA_1-*TAF1* N-terminal tagging cassette into the *TAF1* locus using a similar strategy described previously for Mot1 (Poon et al., 1994). The resulting strain harbors a $G418^R$ cassette and an N-terminal MYPYDVPDYAGVE tag (HA epitope underlined) at the *TAF1* chromosomal locus (additional details upon request). JFHAT1T2delC was generated by applying PCR-based homologous recombination methods to *TAF2* deleting sequences coding for amino acids 1261-1407 at its chromosomal locus. The Taf2-ΔC deletion was confirmed by PCR and immunoblotting.

I-TASSER Structural Prediction and TAF2 Site-Directed Mutagenesis

Primary amino acid sequences for *Saccharomyces cerevisiae* (yeast) Taf2, *Drosophila melanogaster* (fly) Taf2 and *Homo sapiens* (human) Taf2 were submitted to the I-TASSER server for 3-D structure prediction (Yang and Zhang, 2015). Resulting 3D models of yeast, fly and human were imported into Pymol, displayed in cartoon format and colored based on secondary structure. The first of five models generated for yeast Taf2 was used to define solvent exposed residues.

To assess amino acid conservation, yeast, fly and human Taf2 primary sequences were aligned in MacVector using ClustalW. Fifty-eight mutants were designed based on solvent accessibility and proximity to amino acids that are either similar or identical among yeast, fly and human Taf2. Twenty-nine additional mutants were designed based on amino acids predicted to be solvent inaccessible but highly conserved among yeast, fly and human Taf2. All mutations were arbitrarily limited to a maximum of 8 contiguous amino acids. A list of the mutants is provided (Table 3.3).

Plasmid Shuffle and Overexpression Suppression

All *TAF2* site-directed mutants, deletion mutants and *TAF2-TAF14* chimeras were expressed from p415ADH with an N-terminal HA_3NLS tag. For plasmid shuffle analyses, JFTAF2del was transformed with an empty p415ADH plasmid, a p415ADH *TAF2* plasmid, a p415ADH HA_3NLS *TAF2* plasmid or a p415ADH HA_3NLS *TAF2* mutant plasmid and transformants were grown on SC-Leu plates. Leu^+ colonies were grown to saturation in SC-Leu at 30^o, serially diluted 1/4 in sterile ddH_2O, spotted to SC-Leu or SC-Leu + 5-FOA 15cm plates using a pinning tool and grown at various temperatures (20^o, 25^o, 30^o, 34^o and

37°) to assess temperature sensitive growth. Duration of growth ranged from 48 hours to 96 hours, as indicated.

For overexpression suppression screening of all TFIID subunits, plasmid shuffle analyses were performed essentially as above with the following exception. JFTAF2del was co-transformed with *LEU2*-marked *TAF2* plasmids and an empty p413GPD or p413GPD containing the ORF for each of the TFIID subunits and grown on SC-His,Leu plates. For each *LEU2*-marked *TAF2* and p413GPD TFIID subunit combination, two His^+,Leu^+ colonies were spotted undiluted to SC-His,-Leu or SC-His,-Leu + 5-FOA. TFIID subunits that scored positive for overexpression suppression displayed uniformly improved growth for both colonies tested.

Immunoblotting and Co-immunoprecipitation

For steady-state protein immunoblotting, protein was extracted from approx. $1.2x10^7$ cells of early to mid-log grown culture using sodium hydroxide based lysis (Kushnirov, 2000). For co-immunoprecipitations, JFTAF2del strains were grown as pseudodiploids containing both a wild-type *URA3*-marked *TAF2* gene as well as a WT or mutant *LEU2*-marked *TAF2* gene. In these analyses, 50mL of yeast cells were grown at 30° to early to mid-log phase (approx. 2.4-3.6 x 10^7 cells/mL), harvested by centrifugation and lysed in Co-IP buffer (20mM HEPES-KOH pH 7.9, 200mM potassium acetate, 10% glycerol, 0.1% NP40 substitute (Sigma-Aldrich), 1mM DTT + 1X protease inhibitors) using glass bead beating. Soluble protein was separated from insoluble material by centrifugation at 14k rpm for 15min in table-top mini-centrifuge. Protein concentrations were determined using a BSA standard curve with Bio-Rad Protein Assay. Two milligrams of soluble protein

extract were incubated with 2.5μg of anti-HA 12CA5 mAb and 50ng/μL ethidium bromide in a total volume of 412μL at 4° for 2 hours. Immunecomplexes were captured with 10μL of proteinA sepharose beads (Life Technologies) for 30min with mixing at 4°. Captured protein complexes were washed two times with 1mL ice-cold Co-IP buffer and eluted with 2x LDS-NuPAGE Sample Buffer (Life Technologies) and heating at 75° for 10min. Proteins for both steady-state immunoblotting and co-immunoprecipations were separated via SDS-PAGE and processed for immunoblotting as described in Chapter II Methods. Anti-actin (mAb8224) was procured from Abcam. Horse anti-mouse IgG HRP conjugate was procured from Cell Signaling. For immunoblot loading controls, the blot was stained with Ponceau-S (0.5% w/v 1% v/v glacial acetic acid) following the electro-transfer and destained with ddH_2O prior to imaging.

Protein Overexpression and Purification

All His_6-tagged proteins were expressed from pET28a (Novagen). All His_6-GST-tagged proteins were plasmid generated from pBG101 in Rosetta2 (DE3) *E. coli* (Novagen). His_6-Taf14 and Taf2 fragments were N-terminally tagged with MGSS**HHHHHH**SSGLVPAGSHMAS (bold indicating hexa-histidine tag). *E. coli* expression strains were grown and induced with IPTG as described in Chapter II Methods. His_6-Taf14, His_6-GST, His_6-GST-Taf14 aa 1-244, His_6-GST-Taf14 aa 1-123 and His_6-GST-Taf14 aa 124-244 were purified as described in Chapter II Methods. Taf2 fragments were purified in a similar fashion with modifications for denaturing purification. Approximately 200mg of *E. coli* pellet was lysed in 6mL of denaturing lysis buffer (20mM Tris-HCl pH 7.9, 1x PBS, 6M guanidinium HCl, 10mM imidazole). Denatured cell extract was mixed with 200μL of Ni^{2+}-

NTA agarose for 2 hours at 20°. Ni^{2+}-NTA agarose and bound proteins were washed with >10 column volumes of freshly made denaturing wash buffer (20mM Tris-HCl pH 7.9, 1x PBS, 7M urea, 10mM imidazole). His_6-Taf2 fragments were eluted with freshly made denaturing elution buffer (denaturing wash buffer except 200mM imidazole). For all purified proteins, protein concentration was determined via in-gel quantitation using a BSA standard curve.

MBP, MBP-Taf2 and MBP-Taf2ΔC were expressed in *S. cerevisiae* using Gal induction and purified as described in Chapter II Methods with the exception that MBP was not subjected to MonoQ ion exchange chromatography.

Taf2/Taf14 Co-expression Solubilization Analyses

For Taf2 and Taf14 co-expression, a bi-cistronic expression plasmid was generated by cloning in order either full length Taf14 or Taf14 aa 164-244, an internal ribosome binding sequence (RBS) (Lutzmann et al., 2002) and Taf2 C-terminal fragments (aa 1301-1407) into pET28a. Taf14 full length and aa 164-244 fragments were engineered through PCR to contain 5' NcoI and 3' XhoI restriction ends. The internal RBS double stranded DNA insert was generated by annealing complementary oligos, as described in Chapter II Methods. The RBS insert included a 5' XhoI compatible end and 3' SpeI compatible end. The Taf2 C-terminal fragments were generated by PCR and engineered to contain a 5' SpeI compatible end, no stop codon and a 3' HindIII compatible end. The final plasmid was generated by performing a 4-piece ligation containing pET28a digested with NcoI-HF and HindIII-HF restriction endonucleases (NEB) and the three inserts (5'-NcoI-*TAF14* (full-length or aa 164-244)-XhoI-3' + 5'-XhoI-internal RBS-SpeI-3' + 5' SpeI-*TAF2* aa

1301-1407-(No STOP)-3' HindIII). For all cloning of all additional *TAF2* C-terminal fragments, the resulting pET28a *TAF14-TAF2* co-expression plasmid (either *TAF14* full length or aa164-244) was digested with SpeI-HF and HindIII-HF restriction endonucleases to dropout the *TAF2* aa1301-1407 fragment. This fragment was subsequently replaced with a family of *TAF2* C-terminal fragment N- and C-terminal truncation variants. The removal of the stop codon made the coding sequence of the *TAF2* C-terminal fragments in-frame with the His_6-tag (KLAAALE**HHHHHH**(STOP)) encoded in the pET28a backbone. Taf2/Taf14 co-expression solubilization assays were performed by growing 6mL of *E. coli* expression strains at 37° shaking at 250rpm to an OD600 = 0.6-0.9, shifted to 30° and induced for 2 hours with 1mM IPTG. One-sixth of the cell pellet was lysed in *E. coli* denaturing wash buffer for total cellular protein. The remaining cell pellet was lysed in 1mL *E. coli* lysis buffer with lysozyme as described above. Soluble protein was mixed with 7.5μL of Q-Sepharose to remove nucleic acids for 30min at 4°. The flow-through was mixed with 20μL of Ni^{2+}-NTA agarose for 2 hours at 4°. Bound proteins were washed 2X with 500μL of *E. coli* lysis buffer and eluted with 200μL of *E. coli* elution buffer. Total cellular protein and purified Taf2/Taf14 complexes were separated via SDS-PAGE as described above except run with 1X MES buffer (Life Technologies) to separate smaller protein species. The proteins were visualized using coomassie brilliant blue.

TFIID Purification

HA_1-Taf1 TFIID, HA_1-Taf1 Taf2-ΔC TFIID and Taf1-Tap TFIID were purified as described in Chapter II Methods. To determine relative subunit stoichiometry of purified HA_1-Taf1 TFIID and HA_1-Taf1 Taf2-ΔC TFIID, the peak fractions of each TFIID variant were

analyzed by SDS-PAGE and SyproRuby gel staining. The gel was scanned using PharosFX scanner (Bio-Rad), band peak intensity was calculated in QuantityOne (Bio-Rad) and the values were graphed in GraphPad Prism. Relative subunit stoichiometry between the HA_1-Taf1 TFIID and the HA_1-Taf1 Taf2-ΔC TFIID was calculated based on calculating the area under the curve using ImageJ software and normalizing to the Taf6 signal within each preparation.

Far-western Blotting

0.5pmol of Taf1-TAP TFIID, 5pmol purified MBP, 1pmol MBP-Taf2 and between 1 and 5pmol Taf2 truncation variants were subjected to Far-western analysis with a His_6-Taf14 overlay essentially as described (Garbett et al., 2007) with the following modifications. For all analyses, proteins samples were subjected to SDS-PAGE in triplicate: one gel for SyproRuby gel staining, one mock overlay control and one His_6-Taf14 binding assay. During the blotting process, all binding steps and washes were performed in renaturation buffer: 20mM HEPES-KOH pH 7.6, 75mM potassium chloride, 2.5mM magnesium chloride, 0.25mM EDTA and 0.05% Triton X-100 with 1mM DTT freshly added. The overlay was performed with 7nM His_6-Taf14 with 1% BSA as a non-specific competitor or with just the BSA competitor for the mock control. Bound His_6-Taf14 was detected using a standard immunoblotting protocol (primary antibody: antigen affinity purified anti-Taf14 polyclonal rabbit IgG at a concentration of 0.1ng/mL and secondary antibody: goat-anti-rabbit Fc-HRP used per manufacturers instructions). Prior to treatment with ECL and exposure to film, the blots were washed once with tris-buffered saline.

GST Pull Down Assays

Typically, binding reactions were performed in a total volume of 200mL in the following reaction buffer: 20mM HEPES-KOH pH 7.9, 300mM potassium acetate, 10% glycerol, 1mM DTT, 0.1% NP40 substitute, 0.1mg/mL BSA. MBP-Taf2 and MBP-Taf2-ΔC GST-pull downs were performed twice, either with 2pmol His_6-GST-Taf14 or 5pmol His_6-GST-Taf14 and twice that amount with His_6-GST only pull-downs. Between 1pmol and 32 pmol of MBP-Taf2 or MBP-Taf2-ΔC were used to generate dose response curves. For Taf2 binding assays, 16pmol His_6-GST, 8pmol His_6-GST-Taf14, 12pmol His_6-GST-Taf14 aa 1-123 and 12pmol His_6-GST-Taf14 aa 124-244 were used. These proteins were mixed with either no Taf2 or between 0.78pmol and 6.25pmol Taf2. Binding reactions were allowed to proceed at 20^o for 1hour followed by 30min capture at 20^o with 10mL 1:1 slurry of glutathione agarose Fast-Flow (GE Healthcare) equilibrated in reaction buffer. Glutathione agarose bound complexes were pelleted by centrifugation and the supernatant was discarded. The pellet was washed with 500mL binding buffer sans BSA, pelleted again and the wash buffer was discarded. Bound proteins were eluted with 2x LDS sample buffer with 100mM DTT and heating at 75^o for 10min. Proteins were separated with 4-12% NuPAGE Bis-Tris gradient gels (Life Technologies), run with 1X MOPS running buffer and stained with SyproRuby per manufacturers instructions. Proteins were imaged with PharosFX Scanner (Bio-Rad).

Quantitative Reverse Transcription PCR (qRT-PCR)

For all RNA analyses, JFTAF2del was used as the parent strain. For temperature shift experiments, this strain was co-transformed with p415ADH-*HA_3NLS-TAF2* or

p415ADH-*HA$_3$NLS-taf2-ts7* and either p413GPD or p413GPD- *FLAG$_2$NLS-TAF14*. Leu$^+$His$^+$ colonies were subjected to plasmid shuffle on SC-His,Leu + 5-FOA. For each co-transformed shuffled strain, two independent colonies were grown in SC-His,Leu at 25^o until they reached a cell density of ~1.2x10^7 cells/mL and then abruptly shifted to 37^o by adding an equal volume of 50^o heated SC-His,-Leu and then grown at 37^o for 2 hours. Cells were harvested by centrifugation and pellets were stored at -80^o. For steady-state RNA experiments, JFTAF2del was transformed with either p415ADH-*HA$_3$NLS-TAF2* or p415ADH-*HA$_3$NLS-taf2-DC*. Leu$^+$ colonies were subjected to plasmid shuffle by growth on SC-Leu + 5-FOA. For each shuffled strain, four independent colonies were grown in YPD until the cells reached ~2.4x10^7 cells/mL and harvested by filtration.

For all samples, RNA was extracted using hot acidic phenol as described (Schmitt et al., 1990). Reverse transcription was performed with 1μg of total RNA using Superscript III according to manufacturers instructions. cDNA was generated using oligo dT-16 and 1pmol each of gene specific reverse primers for U3, RDN58 and SNR6. The primer sequences used in these analyses are as follows: (U3-F: CAAAAGAGCCACTGAATCCAACT-TGG), (U3-R: GTACCCACCCATAGAGCCCTATCCCTTC), (RDN58-F: AACGGATCTCTTGGTTC-TCG), (RDN58-R: GTGCGTTCAAAGATTCGATG), (SNR6-F: CGAAGTAACCCTTCGTGGAC), (SNR6-R: TCATCCTTATGCAGGGGAAC) (Bonnet et al., 2014), (RPS3-F: TACGGTGTCG-TCAGATACG), (RPS3-R: GACCAGAGTGAATCAAGAAACC), (RPS5-F: GGATGCTTCTTT-GGTTGACTAC), (RPS5-R: GGACATTGAGCCTTTCTGAATCTC), (RPS8A-F: AAAGATCCG-CTACCGGTGCCAAG), (RPS8A-R: TCTTGGAGATACCTTCAGAAGCCC), (RPS9B-F: CGGT-TTGAAGAACAAGAGA), (RPS9B-R: GCATTACCTTCGAACAATC), (PGK1-F: TGCTGCTT-

TGCCAACCATC), (PGK1-R: GTGACATCCTTACCCAACAATG), (PYK1-F: CCAACCTCCA-CCACCGAAAC), (PYK1-R: GGGCTTCAACATCATCAGTCCA) (Garbett et al., 2007).

Quantitative PCR was performed using Sybr Green Supermix (Bio-Rad) according to manufacturers instructions. Samples were quantified using the relative standard curve method, normalized to U3 and expressed as a percentage of the average of the HA-*TAF2* strain. The standard curve was generated by mixing equal amounts of RNA from each of the samples tested prior to reverse transcription. This RNA mix was used both as the standard curve and as the no RT control. Three values were used for each standard curve based on the dilution of the cDNA. For example, if a dilution of 1:50 was used to measure the experimental cDNA samples, then a standard curve of 1:5, 1:50 and 1:500 was used. The no RT reactions displayed either no observable signal or required an additional 10 C_t values above the +RT samples to achieve measureable signal. Thus, we concluded that contaminating genomic DNA was negligible in the samples. The dilutions used for each gene are as follows: 1:50 for *RPS3*, *RPS5*, *RPS8A*, *RPS9B*, and *SNR6*; 1:1000 for *PGK1*, *PYK1* and *U3*; and 1:20,000 for *RDN58*. All qPCR reactions were performed in triplicate. Individual technical replicates were only discarded as outliers if C_t values were different from the other technical replicates by greater than a full C_t value and the amplification trace displayed apparent aberrations in amplification efficiency.

RESULTS

I-TASSER Prediction of Taf2 Structures

To understand the role of Taf2 in TFIID function, I initially pursued classical approaches that have been successfully used to identify TFIID subunit functional domains. First, primary amino acid sequences from *S. cerevisiae* (yeast), *H. sapiens* (human) and *D. melanogaster* (fly) Taf2 were subjected to ClustalW sequence alignment to identify "hotspots" of amino acid sequence homology that we could target for mutagenesis. These analyses have been successfully applied to the histone fold domain containing Tafs as well as the conserved TBP C-terminus (Gangloff et al., 2001b; Klebanow et al., 1997, 1996; Poon et al., 1991; Sanders and Weil, 2000). Yeast Taf2, which encodes a 1407 aa protein, only displays weak sequence conservation with human (1199 aa's) and fly (1221 aa's) Taf2, ~15% sequence identity and ~32% sequence similarity (Figure 3.1a and 3.1c). In addition, the amino acids that are conserved do not cluster in "hotspots" but instead are distributed throughout the amino acid sequence. Second, I performed systematic ~100 amino acid N-terminal, C-terminal and internal deletion mutagenesis as previously described (Bai et al., 1997; Garbett et al., 2007; Layer et al., 2010; Singh et al., 2004) to generate 49 deletion variants. *TAF2* is a single-copy essential gene. Thus, to perform genetic complementation assays, I used a pseudodiploid *TAF2* WT strain for plasmid shuffle analysis, as previously described (Gangloff et al., 2001b; Garbett et al., 2007; Layer et al., 2010; Layer and Weil, 2013). Forty-eight of these variants were unable to complement a *taf2* null strain and failed to stably incorporate into the TFIID complex (summarized in Table 3.1). The single variant that could complement a *taf2* null strain (Taf2 aa 1-1307) displayed no discernable adverse

Table 3.1

TAF2 VARIANT	*taf2Δ* COMPLEMENTATION	STEADY-STATE PROTEIN LEVELS	ABILITY TO CO-IP TFIID (Taf1 and Taf4 scored)
Deletion Analyses			
WT Taf2	YES; no growth defects	(++)	YES
Taf2 aa 1-1307	YES; no growth defects	(+++)	YES
Taf2 aa 1-1207	NO, inviable	(+)	NO
Taf2 aa 1-1107	NO, inviable	(+/-)	NO
Taf2 aa 1-1007	NO, inviable	(++++)	NO
Taf2 aa 1-907	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1-807	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1-707	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1-607	NO, inviable	(++)	NO
Taf2 aa 1-507	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1-407	NO, inviable	(++++)	NO
Taf2 aa 1-307	NO, inviable	(++)	NO
Taf2 aa 1-207	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1-107	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1301-1407	NO, inviable	(++)	NO
Taf2 aa 1201-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1101-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 1001-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 901-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 801-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 701-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 601-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 501-1407	NO, inviable	Protein Not Detectable	NO
Taf2 aa 401-1407	NO, inviable	(+)	NO
Taf2 aa 301-1407	NO, inviable	(+)	NO
Taf2 aa 201-1407	NO, inviable	(+)	NO
Taf2 aa 101-1407	NO, inviable	(+)	NO
Taf2 1-93,304-1407	NO, inviable	(+/-)	NO
Taf2 1-193,304-1407	NO, inviable	(+/-)	NO
Taf2 1-293,304-1407	NO, inviable	(+/-)	NO
Taf2 1-396, 497-1407	NO, inviable	(+/-)	NO
Taf2 1-396, 597-1407	NO, inviable	(+/-)	NO
Taf2 1-396, 697-1407	NO, inviable	(+/-)	NO
Taf2 1-443,744-1407	NO, inviable	(+/-)	NO
Taf2 1-543,744-1407	NO, inviable	(+/-)	NO
Taf2 1-643,744-1407	NO, inviable	(+/-)	NO
Taf2 1-746,847-1407	NO, inviable	(+/-)	NO
Taf2 1-746,947-1407	NO, inviable	(+/-)	NO
Taf2 1-746,1047-1407	NO, inviable	(+/-)	NO
Taf2 1-800, 1101-1407	NO, inviable	(+/-)	NO
Taf2 1-900,1101-1407	NO, inviable	(+/-)	NO
Taf2 1-1000,1101-1407	NO, inviable	(+/-)	NO
Taf2 1-1102,1203-1407	NO, inviable	(+/-)	NO
Taf2 1-1102,1303-1407	NO, inviable	(+/-)	NO
Taf2 101-1307	NO, inviable	(+/-)	NO
Taf2 201-1207	NO, inviable	(+/-)	NO
Taf2 301-1107	NO, inviable	(+/-)	NO
Taf2 401-1007	NO, inviable	(++)	NO
Taf2 501-907	NO, inviable	Protein Not Detectable	NO
Taf2 601-807	NO, inviable	Protein Not Detectable	NO

Table 3.2

TAF2 VARIANT	*taf2Δ* COMPLEMENTATION	STEADY-STATE PROTEIN LEVELS	ABILITY TO CO-IP TFIID (Taf1 and Taf4 scored)
Random Mutagenesis Ts+ Variants	**P = permissive, @NP = Non-permissi**	(++)	YES
taf2- A12T, E571K	(@P= normal, @NP = no growth)	(++)	YES
taf2 - E403K, G642D, D861N	(@P= normal, @NP = no growth)	(++)	YES
taf2 - E407K, A518V, D861N	(@P= normal, @NP = no growth)	(++)	YES
taf2 - D537N, S828P	(@P= normal, @NP = no growth)	(++)	YES
taf2 - S595F, A919V	(@P= normal, @NP = no growth)	(++)	YES
taf2- E672K, T751I	(@P= normal, @NP = no growth)	(++)	YES
taf2 - S677F, E1080K	(@P= normal, @NP = no growth)	(++)	YES
taf2 - V679G	(@P= normal, @NP = very slow)	(++)	YES
taf2 - P798L, P987L	(@P= normal, @NP = no growth)	(++)	YES
taf2 - S900N	(@P= normal, @NP = no growth)	(++)	YES
taf2 - R904K	(@P= normal, @NP = no growth)	(++)	YES
taf2 - T905M	(@P= normal, @NP = no growth)	(++)	YES
taf2 - C920Y	(@P= normal, @NP = no growth)	(++)	YES
taf2 - G934D, G935S	(@P= normal, @NP = no growth)	(++)	YES
taf2 - G935S	(@P= normal, @NP = no growth)	(++)	YES
taf2 - G935C	(@P= slow, @NP = no growth)	(++)	YES
taf2 - H938Y	(@P= normal, @NP = no growth)	(++)	YES
taf2 - I987L, S1053L	(@P= normal, @NP = no growth)	(++)	YES
taf2 - E1137k, D1138N, V1301I	(@P= normal, @NP = no growth)	(++)	YES
taf2 - D1138N	(@P= normal, @NP = no growth)	(++)	YES

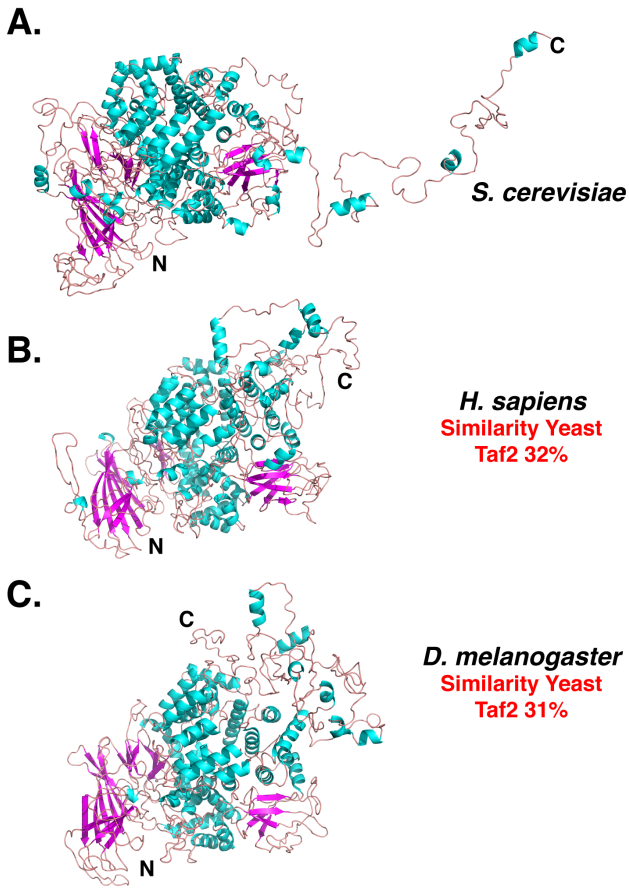


Figure 3.1: 3-D Structural Prediction of Taf2 Homologs Suggests Similar Structures Despite Low Overall Sequence Conservation. A-C. The primary amino acid sequences of **A.** *S. cerevisiae* Taf2, **B.** *H. sapiens* Taf2 and **C.** *D. melanogaster* Taf2 were analyzed using ClustalW sequence alignment algorithm and the 3-D structural prediction program I-TASSER. Evolutionary sequence conservation and predicted 3-D structural models indicating secondary structure (pink = b-strand, cyan = a-helix and tan = random coil) are shown. Amino (N)- and carboxy (C)- termini are labeled. Sequence alignment similarity score of *H. sapiens* and *D. melanogaster* compared to *S. cerevisiae* displayed next to the predicted structures in **B.** and **C.**, respectively.

growth, steady-state protein levels or TFIID stable incorporation phenotypes. Third, I created a hydroxylamine mutagenized *TAF2* plasmid mutant library (Klebanow et al., 1996) to generate *taf2*^{ts} variants (summarized in Table 3.2). However, none of the twenty *taf2*^{ts} alleles identified in a plate growth-based screen displayed acute loss-of-function in liquid culture when shifted to the non-permissive temperature (37^{o}) (data not shown) as has been shown to occur for Ts alleles for every essential yeast Taf (Shen et al., 2003).

To overcome these technical hurdles, I hypothesized that a structure-guided site-directed mutagenesis strategy would successfully identify Taf2 amino acids critical for Taf2 function. However, at the time of this study, no 3-D structural information existed for Taf2. Thus, to generate a putative 3-D model for Taf2, I employed the 3-D structural prediction program I-TASSER to generate models for yeast, human and drosophila Taf2 structures (Figure 3.1a,

3.1b and 3.1c, respectively) (Yang and Zhang, 2015). I-TASSER has been successfully used in the gene regulation field to model structurally intractable proteins (Louder et al., 2016; Plaschka et al., 2016; Reiss and Mobley, 2011). Despite the weak amino acid sequence conservation, I-TASSER predicts that the Taf2 structures from these three different organisms display the same general features. Particularly, the N-terminus of Taf2 is predicted to contain a β-sheet motif, the central portion and majority of the molecule is composed of HEAT repeats and the C-terminus is predicted to form another β-sheet motif along with an unstructured region. This putative C-terminal unstructured region is also predicted to be an intrinsically disordered region in metazoans (Zhang et al., 2015). As noted above, while this manuscript was in preparation, a high-resolution cryo-EM structure of human Taf2 was described and it exhibits the same general architecture predicted by our analysis (Louder et al., 2016). Given the similarities among the yeast, human and drosophila Taf2 predicted structures, I concluded that our *in silico* generated yeast Taf2 structure was a suitable model from which to design a structure-based site-directed mutagenesis screen.

Site-directed Mutagenesis of *TAF2*

To further genetically interrogate y*TAF2*, I designed two classes of *TAF2* site-directed variants comprising eighty-seven mutants. These variants primarily consisted of ala block substitution mutations but also included charge reversal mutations (see Table 3.3). Class I mutants, designed based on predicted solvent accessibility and proximity to evolutionarily conserved residues, contained fifty-eight variants. Class II mutants, designed based on predicted solvent inaccessibility and included mutations of groups of conserved

Table 3.3

TAF2 VARIANT	*taf2Δ* COMPLEMENTATION	STEADY-STATE PROTEIN LEVELS	ABILITY TO CO-IP TFIID
Site-Directed Mutagenesis	(@P= permissive 25C, @NP=nonpermissive 37C)		(Taf7 and Taf8 scored)
Class I Mutants: Solvent Accessible			
taf2-m1 ala 25-28	YES, no defect	Similar to WT	Not Tested
taf2-m2 ala 29-30,32	(@P = Slow, @NP = very slow growth)	Similar to WT	Loss of Function
taf2-m3 ala 34,36,38	YES, no defect	Similar to WT	Not Tested
taf2-m4 ala 45,47,49	YES, no defect	Similar to WT	Not Tested
taf2-m5 ala 72-77	No, inviable	Similar to WT	Loss of Function
taf2-m6 ala 120,123-127	YES, no defect	Similar to WT	Not Tested
taf2-m7 ala 139-143	YES, no defect	Similar to WT	Not Tested
taf2-m8 ala 145, 147-150, 152	YES, no defect	Similar to WT	Not Tested
taf2-m9 ala 230-231, 233	YES, no defect	Similar to WT	Not Tested
taf2-m10 ala 236, 238	YES, no defect	Similar to WT	Not Tested
taf2-m11 ala 243-246, 248	YES, no defect	Similar to WT	Not Tested
taf2-m12 ala 364-368	NO, inviable	Similar to WT	Loss of Function
taf2-m13 ala 370-374	YES, no defect	Similar to WT	Not Tested
taf2-m14 ala 394-398	YES, no defect	Similar to WT	Not Tested
taf2-m15 ala 439-445	YES, no defect	Similar to WT	Not Tested
taf2-m16 ala 449-452	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested
taf2-m17 ala 454-458	YES, no defect	Similar to WT	Not Tested
taf2-m18 ala 462-466	(@P = very slow growth, @NP = No growth)	Similar to WT	Loss of Function
taf2-m19 ala 469, 471-474	(@P = very slow growth, @NP = No growth)	Similar to WT	Loss of Function
taf2-m20 ala 554-557, 559	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested
taf2-m21 ala 563-564	YES, no defect	Similar to WT	Not Tested
taf2-m22 ala 588, 592-592, 594-595	YES, no defect	Similar to WT	Not Tested
taf2-m23 ala 633-634, 637-639	YES, no defect	Similar to WT	Not Tested
taf2-m24 ala 640-644	(@P = slow growth, @NP = No growth)	Similar to WT	Loss of Function
taf2-m25 ala 671-672, 675	(@P = slow growth, @NP = No growth)	Similar to WT	Loss of Function
taf2-m26 ala 693-696	(@P = slow growth, @NP =very slow growth)	Similar to WT	Loss of Function
taf2-m27 ala 710-717	YES, no defect	Similar to WT	Not Tested
taf2-m28 ala 727-733	YES, no defect	Similar to WT	Not Tested
taf2-m29 ala 746-747, 751-753	YES, no defect	Similar to WT	Not Tested
taf2-m30 ala 754, 756-758	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested
taf2-m32 ala 775-780	Slow at all temperatures	Similar to WT	Similar to WT
taf2-m33 E829K	YES, no defect	Similar to WT	Not Tested
taf2-m34 ala 833-837	YES, no defect	Similar to WT	Not Tested
taf2-m35 D848K, E850K	(@P = slow growth, @NP = No growth)	Similar to WT	Loss of Function
taf2-m36 ala 858-859	YES, no defect	Similar to WT	Not Tested
taf2-m37 ala 861-863, 866	YES, no defect	Similar to WT	Not Tested
taf2-m38 ala 867-871, 873-874	No, inviable	Similar to WT	Loss of Function
taf2-m39 ala R910E	YES, no defect	Similar to WT	Not Tested
taf2-m40 ala 917, 921 gly 918-919,922	No, inviable	Similar to WT	Loss of Function
taf2-m41 ala 924-925, 927-929	YES, no defect	Similar to WT	Not Tested
taf2-m42 ala 947-952	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested
taf2-m43 ala 953-958	YES, no defect	Similar to WT	Not Tested
taf2-m44 ala 965-966, 968-971	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested
taf2-m46 ala 999, 1001, 1003, 1005	YES, no defect	Similar to WT	Not Tested
taf2-m47 ala 1006-1007, 1010	YES, no defect	Similar to WT	Not Tested
taf2-m48 ala 1027-1032	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested
taf2-m49 ala 1036-1037, 1039-1041	YES, no defect	Similar to WT	Not Tested
taf2-m50 ala 1043-1044, 1047-1049	YES, no defect	Similar to WT	Not Tested
taf2-m51 ala 1051-1055	YES, no defect	Similar to WT	Not Tested
taf2-m52 ala 1077-1080	YES, no defect	Similar to WT	Not Tested
taf2-m53 ala 1084, 1086-1088	YES, no defect	Similar to WT	Not Tested
taf2-m54 ala 1090-1091, 1093-1094	YES, no defect	Similar to WT	Not Tested
taf2-m55 ala 1105-1107, 1112	(@P = slow growth, @NP = No growth)	Slight Reduction	Loss of Function
taf2-m56 ala 1136-1140	No, inviable	Similar to WT	Loss of Function
taf2-m57 ala 1143-1144, 1146-1149	(@P = slow growth, @NP = No growth)	Similar to WT	Loss of Function
taf2-m58 ala 1151-1155	YES, no defect	Similar to WT	Not Tested
taf2-m59 ala 1181-1188	YES, no defect	Similar to WT	Not Tested
taf2-m60 ala 1238-1242	(@P= No defect, @NP=Slow Growth)	Similar to WT	Not Tested

Table 3.3, continued

TAF2 VARIANT	*taf2Δ* COMPLEMENTATION	STEADY-STATE PROTEIN LEVELS	ABILITY TO CO-IP TFIID
Site-Directed Mutagenesis	(@P= permissive 25C, @NP=nonpermissive 37C)		(Taf7 and Taf8 scored)
Class II Mutants: Solvent Inaccessible			
taf2-m67 ala 129-133	YES, no defect	Similar to WT	Not Tested
taf2-m68 ala 134-138	YES, no defect	Similar to WT	Not Tested
taf2-m69 ala 180-185	No, inviable	Similar to WT	Loss of Function
taf2-m70 ala 186-190	YES, no defect	Similar to WT	Not Tested
taf2-m71 ala 191-195	No, inviable	Similar to WT	Loss of Function
taf2-m72 ala 218, 220-223	(@P= No defect, @NP=Slow Growth)	Similar to WT	Loss of Function
taf2-m73 ala 225-229	(@P= Slow, @NP=Slow Growth)	Similar to WT	Loss of Function
taf2-m74 ala 239-241	No, inviable	Similar to WT	Loss of Function
taf2-m76 ala 358-363	YES, no defect	Similar to WT	Not Tested
taf2-m77 ala 375, 377-380	(@P= Slow, @NP=No growth)	Similar to WT	Loss of Function
taf2-m79 ala 388-393	YES, no defect	Similar to WT	Not Tested
taf2-m80 ala 434-438	No, inviable	Similar to WT	Loss of Function
taf2-m81 ala 475-480	No, inviable	Similar to WT	Loss of Function
taf2-m82 ala 490, 493-495	No, inviable	Similar to WT	Loss of Function
taf2-m83 ala 506-509	YES, no defect	Similar to WT	Not Tested
taf2-m84 ala 517, 519, 521, 523-524	No, inviable	Similar to WT	Not Tested
taf2-m85 ala 537-542	No, inviable	Similar to WT	Loss of Function
taf2-m86 ala 543-544,546-548	No, inviable	Similar to WT	Loss of Function
taf2-m87 ala 842-846	No, inviable	Similar to WT	Loss of Function
taf2-m88 ala 909-913	(@P= Slow, @NP=No growth)	Similar to WT	Loss of Function
taf2-m89 ala 914-917	(@P= Slow, @NP=No growth)	Slight Reduction	Loss of Function
taf2-m90 ala 972-976	(@P= No defect, @NP=Slow Growth)	Similar to WT	Similar to WT
taf2-m91 ala 978-983	YES, no defect	Similar to WT	Not Tested
taf2-m92 ala 984-990	(@P= No defect, @NP=Slow Growth)	Similar to WT	Similar to WT
taf2-m94 ala 1016-1020	(@P= No defect, @NP=Slow Growth)	Similar to WT	Similar to WT
taf2-m95 ala 1021-1024, 1026	YES, no defect	Similar to WT	Not Tested
taf2-m96 ala 1056-1061	YES, no defect	Similar to WT	Not Tested
taf2-m97 ala 1062-1065, 1067	YES, no defect	Similar to WT	Not Tested
taf2-m98 ala 1068-1074	YES, no defect	Similar to WT	Not Tested

amino acids, contained twenty-nine variants. All variants were engineered to contain an HA_3NLS (HA) N-terminal tag. Importantly, HA-Taf2 phenocopies untagged-Taf2 in genetic complementation assays and can efficiently and stably incorporate into the TFIID complex (Figure 3.2a and 3.2b). Every variant was scored for its ability to genetically complement a *taf2* null strain at both permissive (25^o) and non-permissive temperatures (37^o) as well as by anti-HA immunoblotting to ensure each variant accumulated to levels similar to wild-type (WT) Taf2.

Results for these analyses are described in detail in Table 3.3. In summary, of the Class I variants, thirty-six percent displayed a genetic complementation defect including five that were inviable and eight variants that were both slow growing at the permissive temperature and Ts at the non-permissive temperature (Figure 3.2; *taf2-ts1*

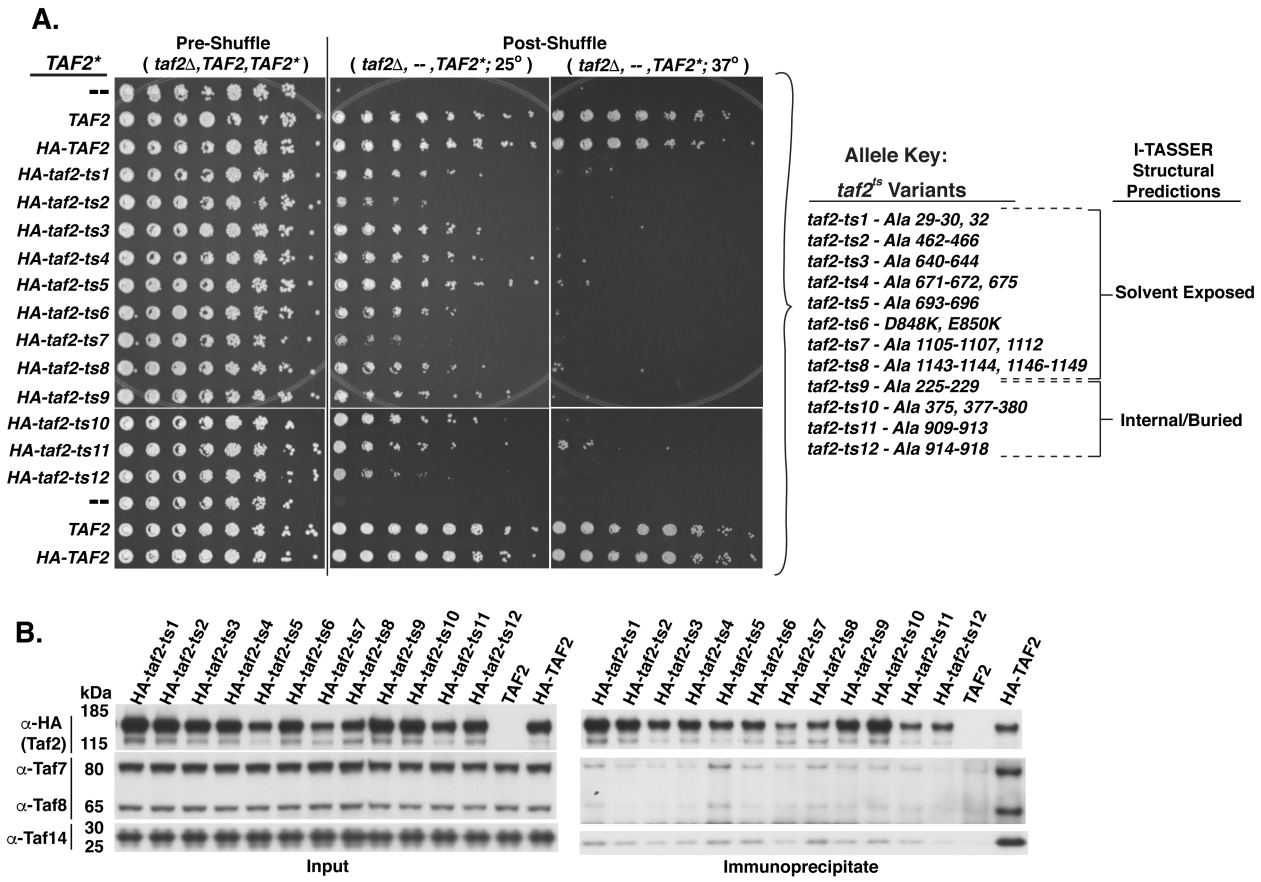


Figure 3.2: Identification of *taf2ts* Alleles Using Structural Prediction and Amino Acid Conservation-Guided Site-Directed Mutagenesis. ***A.*** Genetic complementation assay. *taf2* null cells harboring both *URA3*-marked *TAF2* and *LEU2*-marked *TAF2** plasmids were serially diluted 1:4 from left to right and spotted with a pinning tool onto either synthetic complete (SC) medium without leucine (-L) (Pre-Shuffle) or SC-L + 5-FluorOrotic Acid (5-FOA) to select for cells that have spontaneously lost the *URA3*-marked *TAF2* plasmid (Post-Shuffle). Plates were grown for 72 hours at 37°C or 96 hours at 25°C prior to imaging. *TAF2** = Test gene, either no ORF (--), *TAF2*, *HA$_3$NLS-TAF2* or various *HA$_3$NLS*-tagged *taf2*-temperature sensitive (Ts) alleles (*HA$_3$NLS-taf2-ts1* through *ts12*). The variant amino acids in the *taf2ts* alleles are shown in the Allele Key to the right (*taf2ts* Variants). Representative plate images from at least two biological replicates are presented. HA$_{x3}$NLS-tag labeled as HA in all figures. ***B.*** Taf2-TFIID Co-Immunoprecipitation (CoIP). Whole cell extracts derived from the strains containing TAF2*: TAF2, *HA$_{x3}$NLS*-TAF2, taf2-ts1 through ts12 described in panel ***A.*** were used for immunoprecipitation with the anti-HA mAb 12CA5 IgG. One percent of the lysate (Input) and 33% of the pellet (Immunoprecipitate) were separated via SDS-PAGE, blotted to PVDF membrane and probed with anti-HA (α-HA(Taf2)) and the indicated anti-Taf IgGs (α-Taf7, α-Taf8 and α-Taf14). Representative immunoblots from at least two biological replicates are presented.

through *taf2-ts8*). Of the Class II variants, sixty-two percent displayed a genetic complementation defect including ten that were inviable and four variants that were both slow growing at the permissive temperature and Ts at the non-permissive temperature (Figure 3.2; *taf2-ts9* through *taf2-ts12*). Of note, none of the variants displayed dominant negative phenotypes. Surprisingly, none of the variants displayed drastic reduction in steady-state protein levels (>2x), even variants that could not complement the *taf2* null strain.

To assess the mechanism by which the loss-of-function Taf2 variants fail to complement the *taf2* null strain, inviable and $taf2^{ts}$ variants were subjected to anti-HA Co-ImmunoPrecipitation (Co-IP) (Figure 3.2b and Table 3.2). A strain expressing untagged WT Taf2 was used as a negative control. Despite similar steady-state protein levels and IP efficiency, none of the loss-of-function *taf2* mutant variants tested could co-precipitate either the TFIID-specific subunits Taf7 and Taf8 or Taf14 as efficiently as HA-Taf2. These results suggest that these mutations disrupt Taf2-TFIID subunit interactions, contributing to growth defects.

TAF14 Overexpression Suppresses Select $taf2^{ts}$ Growth Phenotypes

Considering all of the loss-of-function *taf2* variants display defects in stable incorporation into the TFIID complex, likely as a result of disruption of specific Taf2-TFIID subunit interactions, I hypothesized that artificially driving Taf2-TFIID stable incorporation would suppress the $taf2^{ts}$-associated growth defects. Ideally, if a $taf2^{ts}$ mutant variant has a reduction in binding affinity to a TFIID subunit, increasing concentration of that subunit could drive complex formation and rescue the ability of Taf2 to stably associate with TFIID.

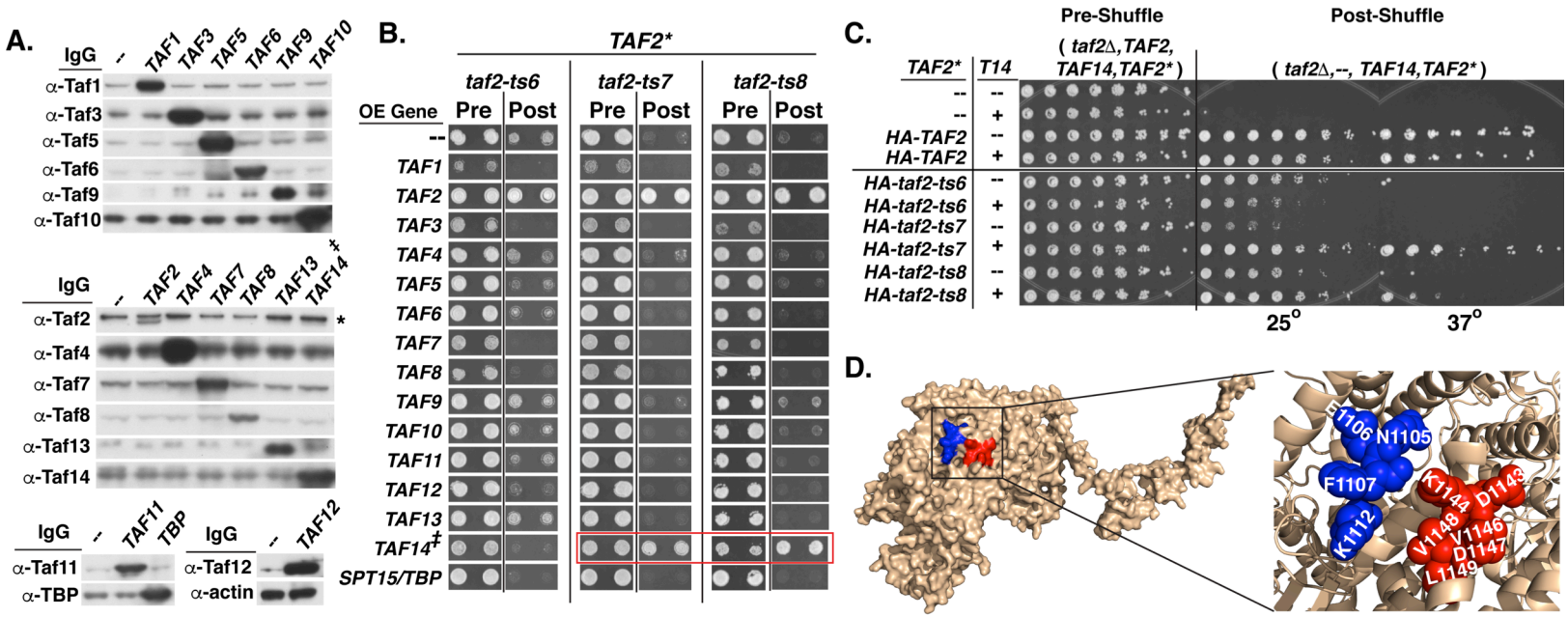


Figure 3.3: Overexpression of *TAF14* Suppresses the Slow and Ts Growth Phenotypes Associated with Select *taf2*^{ts} Mutant Variants. ***A.*** Overexpression of all TFIID Subunits in yeast. Whole Cell Extracts derived from *Saccharomyces cerevisiae* strains harboring TFIID subunit overexpression (OE) plasmids (no ORF (--), Taf1 to Taf14 and TBP as indicated) were separated via SDS-PAGE and immunoblotted using anti-TFIID subunit specific IgGs (α-Taf1 to α-Taf14 and α-TBP) or anti-actin IgG as a loading control for the α-TAF12 blot: *non-specific band, ‡denotes the fact that the cDNA for *TAF14* was used because *TAF14* contains an intron. Image derived from one replicate. ***B.*** Overexpression suppression of select *taf2*^{ts} mutant strains. Two biological replicates of *taf2* null cells harboring a *URA3*-marked *TAF2* plasmid, a *LEU2*-marked *TAF2** plasmid (*HA*$_3$*NLS-taf2-ts6*, *HA*$_3$*NLS-taf2-ts7* and *HA*$_3$*NLS-taf2-ts8*) and a *HIS3*-marked TFIID OE plasmid (same TFIID subunits as in **A.**) were spotted on SC-L,-Histidine (H) medium (pre-shuffle) or SC-L,-H +5-FOA (post-shuffle) medium and grown for 72 hours at 30° for *taf2-ts6* and *taf2-ts7*, and 34° for *taf2-ts8*. Qualitatively improved growth in both biological replicates beyond the no ORF control is indicated with a red box. ***C.*** *TAF14* Overexpression Suppression of *taf2*^{ts} alleles. Plasmid shuffle performed as in panel ***B.*** except the *HIS3*-linked OE plasmid contains a FLAG$_2$NLS-tagged *TAF14* cDNA and cells were serially diluted 1:4 from left to right prior to plating. Plates were grown at 25° for 96 hours or at 37° for 72 hours prior to imaging. *TAF2**: no ORF (--), *HA*$_3$*NLS-TAF2*, *HA*$_3$*NLS-taf2-ts6*, *HA*$_3$*NLS-taf2-ts7* and *HA*$_3$*NLS-taf2-ts8*. ***D.*** A surface rendering of the Taf2 I-TASSER predicted structure (tan) highlighting the position of the amino acids mutated in *taf2-ts7* (blue) and *taf2-ts8* (red). Note: amino acids mutated in *taf2-ts6* are located on the reverse side of the Taf2 predicted structure and hence not visible in this image. Blow-up: Identities of the amino acids mutated in *taf2-ts7* and *taf2-ts8* in spheres with surrounding secondary structure represented by ribbons.

To this end, I individually overexpressed every TFIID subunit (Figure 3.3a) in strains harboring the *taf2*^{ts} variants. Our lab has previously shown that overexpression of TFIID subunits *TAF4* and *TAF11* have displayed positive genetic interactions with *TOA2*, a protein that directly interacts with the TFIID complex, demonstrating that overexpression of individual TFIID subunits is a viable method for identifying genetic interactions (Layer and Weil, 2013).

Overexpression of *TAF1*, *TAF3*, *TAF7*, *TAF8* and *TAF14* resulted in dominant negative growth phenotypes, consistent with published results (Yoshikawa et al., 2011). As expected, overexpression of *TAF2* could complement growth for all *taf2*^{ts} variants (Figure 3.3b, data not shown). In addition, I observed strong *taf2*^{ts} allele-specific suppression when I overexpressed *TAF14* but not for any of the other non-*TAF2* TFIID subunits (Figure 3.3b and 3.3c; see *taf2-ts6* vs. *taf2-ts7* and *ts8*). *TAF14* overexpression suppression of the *taf2-ts7*-associated growth phenotype was particularly potent, rescuing growth to near WT levels at both permissive and non-permissive temperatures.

While five of the twelve *taf2*^{ts} variants displayed mild growth improvement, *TAF14* overexpression suppression of *taf2-ts7* and *ts8* significantly improved growth at all temperatures tested. The amino acids mutated in the *taf2-ts7* and *ts8* variants map to adjacent α-helixes in the yTaf2 predicted structure whereas *taf2-ts6* is predicted to reside on the opposite side of the molecule (Figure 3.3d). These results suggest that amino acids mutated in the *taf2-ts7* and *ts8* form a domain that is likely involved in Taf2-TFIID subunit physical interactions, potentially Taf2-Taf14 interaction. Due to the potency of the genetic interaction between *TAF14* and *taf2-ts7*, further molecular and genetic analyses were performed with *taf2-ts7.*

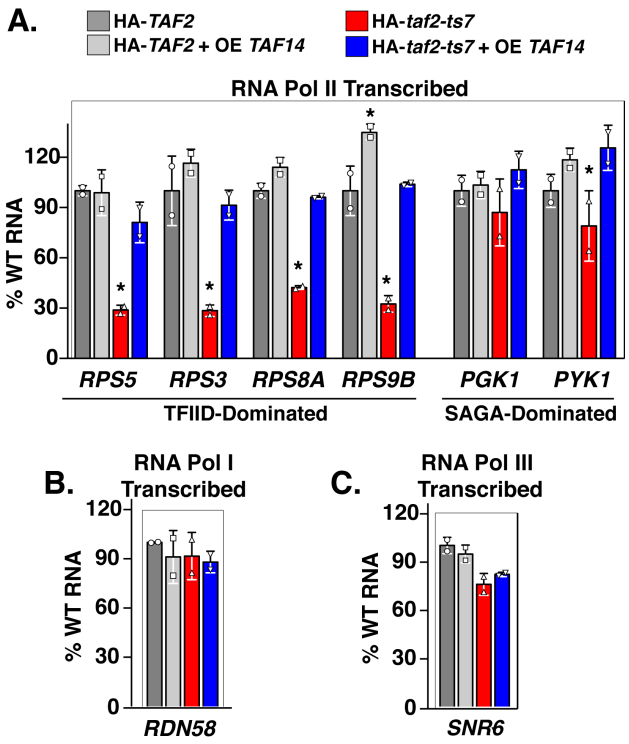


Figure 3.4: Overexpression of *TAF14* Suppresses the Ribosomal Protein Gene Transcription Defect Associated with *taf2-ts7*. *A-C.* Analysis of RNAs from cells shifted to the non-permissive temperature. Shuffled strains harboring *HA$_3$NLS-TAF2* or *HA$_3$NLS-taf2-ts7* with either an empty *HIS3*-marked overexpression (OE) plasmid or a *HIS3*-marked OE plasmid containing an N-terminally FLAG$_2$NLS tagged *TAF14* cDNA were grown at 25^o to early-mid-log phase and then shifted to the non-permissive temperature (37^o) for 2 hours. RNAs were extracted and analyzed by qRT-PCR scoring for ***A.*** RNA Pol II, ***B.*** RNA Pol I and ***C.*** RNA Pol III transcribed genes. Data were generated from two biological replicates. Each data point in the graph represents one biological replicate and is generated from the average of three technical replicates. Results were statistically analyzed using a 2-way ANOVA with Dunnett's multiple comparisons test (GraphPad Prism). Mean ± SD depicted. * = p<0.05. Dark Gray = *HA$_3$NLS-TAF2*, Light Gray = *HA$_3$NLS-TAF2* + OE *TAF14*, Red = *HA$_3$NLS-taf2-ts7* and Blue = *HA$_3$NLS-taf2-ts7* + OE *TAF14*.

TAF14 Overexpression Suppresses Defects in Ribosomal Protein Gene Transcript Abundance Associated with *taf2-ts7*

I hypothesized strong *taf2-ts7* growth defects likely result in a reduction in ribosomal protein-encoding gene (RPG) transcript abundance, a class of genes that is TFIID-dominated (Basehoar et al., 2004; Huisinga and Pugh, 2004; Li et al., 2000; Rhee and Pugh, 2012; Shen et al., 2003). To assess RPG transcripts, log phase growing yeast strains, harboring either *TAF2* or *taf2-ts7* and containing either an empty overexpression vector or a *TAF14* overexpression vector, were abruptly shifted to the non-permissive temperature for two hours followed by RNA extraction and quantitative reverse transcription PCR (qRT-PCR). Our lab has assessed RPG transcript abundance using this temperature shift paradigm using Ts variants for *TAF1* through *TAF13*, *TOA1* and *TOA2* (Garbett et al., 2007; Layer et al.,

2010; Layer and Weil, 2013).

I found that *taf2-ts7* without *TAF14* overexpression displays a statistically significant ~3-fold reduction compared to *TAF2* in TFIID-dominated RPG transcripts (*RPS3*, *RPS5*, *RPS8A* and *RPS9B*; Figure 3.4a) without a concomitant significant decrease in SAGA-dominated transcripts (*PGK1* and *PYK1*; Figure 3.4a), RNA polymerase I transcripts (*RDN58*; Figure 3.4b) or RNA polymerase III transcripts (*SNR6*; Figure 3.4c) (Bonnet et al., 2014). The RPG transcript defects were ameliorated when *TAF14* was overexpressed in the *taf2-ts7* strain. These data suggest the ability of *taf2-ts7* to appropriately regulate RPG transcript abundance is aided by elevated *TAF14* levels.

Taf2 and Taf14 Directly Interact *in vitro*

The simplest model for the genetic interaction between Taf2 and Taf14 is that these two proteins directly interact. Consistent with this hypothesis, Taf2 was identified as a Taf14 interacting protein in a genome-wide yeast two-hybrid screen but was not authenticated as a direct interaction (Kabani et al., 2005). To determine if Taf2 and Taf14 directly interact, Taf1-TAP purified TFIID, maltose binding protein (MBP) and MBP-Taf2, all purified from yeast, were subjected to Far-Western blotting with and without purified recombinant His_6-Taf14 (Figure 3.5a). His_6-Taf14 bound proteins were detected with antigen affinity purified anti-Taf14 IgG (Sanders et al., 2002b). When His_6-Taf14 was omitted from the Far-Western blot, the only signals observed were for purified Taf14 from the TFIID complex as well as trace amounts of Taf14 that co-purify with the yeast generated MBP-Taf2. Upon overlay with His_6-Taf14, His_6-Taf14 bound to both Taf2 from TFIID as well as MBP-Taf2, but not MBP or other TFIID subunits. To extend these analyses

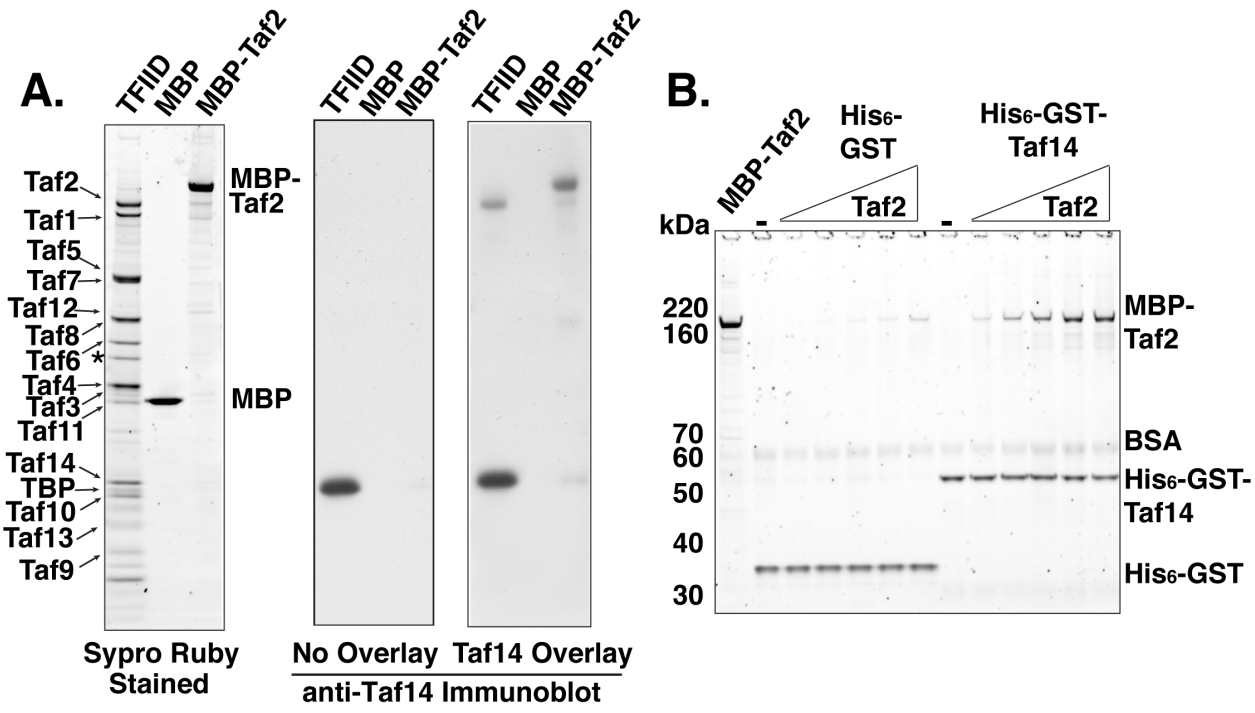


Figure 3.5: Taf2 and Taf14 Directly Interact *in vitro*. *A.* Far-Western Blotting. Approx. 500 nanograms of Taf1-TAP purified TFIID, 200 nanograms of recombinant MBP and 200 nanograms of recombinant MBP-Taf2 were separated via SDS-PAGE in triplicate. One gel was stained with Sypro Ruby gel stain while the other two were analyzed by anti-Taf14 blotting, with and without 7nM overlay of purified recombinant His_6-Taf14. Representative image presented from more than three technical replicates. ***B.*** Glutathione S-Transferase (GST) Pull-Downs. Either four picomoles of purified recombinant His_6-GST or two picomoles of His_6-GST-Taf14 were mixed with 0.1mg/mL BSA and either no MBP-Taf2 or between 2 and 32pmol of MBP-Taf2. His_6-GST or His_6-GST-Taf14 bound proteins were pulled-down with glutathione agarose, washed with binding buffer, analyzed via SDS-PAGE and stained with Sypro Ruby gel stain. One picomole of purified MBP-Taf2 (lane 1) was loaded to demonstrate purity of input material. One technical replicate was performed.

to solution binding assays, increasing concentrations of MBP-Taf2 were mixed with either purified recombinant His_6-GST or His_6-GST-Taf14, in the presence of BSA as a non-specific competitor, and subjected to GST pull-down. His_6-GST-Taf14 specifically bound MBP-Taf2 in a dose-dependent and saturable manner (Figure 3.5b). Thus, I have shown through two independent methods that Taf2 and Taf14 specifically and directly interact *in vitro*.

The Taf2 C-terminus Is Necessary and Sufficient for Binding Taf14 *in vitro* and *in vivo*

Using the Far-Western assay, I determined the domain of Taf2 where Taf2 and Taf14 directly interact (Figure 3.6). To this end, purified MBP-Taf2 (lane 1), His_6-Taf2 aa 1-407 (lane 2), His_6-Taf2 aa 401-1007 (lane 3), His_6-Taf2 aa 1001-1407 (lane 4), His_6-Taf2 aa 1001-1207 (lane 5) and His_6-Taf2 aa 1201-1407 (lane 6) were subjected to Far Western blotting, as described above (Taf2 purified forms Figure 3.6a Left, constructs diagrammed in 3.6b). When His_6-Taf14 was omitted from the overlay, the only signal present was the co-purifying Taf14 in the MBP-Taf2 sample. Upon His_6-Taf14 overlay, His_6-Taf14 bound to MBP-Taf2 (lane1), His_6-Taf2 aa 1001-1407 (lane 4) and His_6-Taf2 aa 1201-1407 (lane 6). These results suggest that the Taf14 binding domain resides in Taf2 aa 1201-1407.

As described above, my systematic 100aa *TAF2* truncation analysis showed that Taf2 aa 1-1307 could complement a *taf2* null strain and stably incorporate into the TFIID complex whereas Taf2 aa 1-1207 could do neither (Table 3.1). Since these C-terminal amino acids contain the Taf14 binding domain, I hypothesized that a finer truncation analysis of the Taf2 C-terminus may define the amino acids necessary for Taf14 binding *in vivo*. *TAF2* was subjected to 10 aa serial truncations of its C-terminus. These variants were analyzed for their ability to complement a *taf2* null strain, steady-state protein levels and ability to coIP TFIID subunits (Figure 3.6c and 3.6d). I found that Taf2 aa 1-1250 could not complement a *taf2* null strain, had reduced steady-state protein levels compared to HA-Taf2 and could not coIP TFIID subunits Taf4, Taf7, Taf8, Taf9 or Taf14. Smaller truncations (Taf2 aa 1-1260 to 1300), still maintained their ability to complement the *taf2* null strain, displayed elevated steady-state protein levels compared to HA-Taf2 and could coIP TFIID subunits Taf4, Taf7, Taf8 and Taf9. In regards to Taf14 coIP, Taf2 aa 1-1280, Taf2 aa 1-

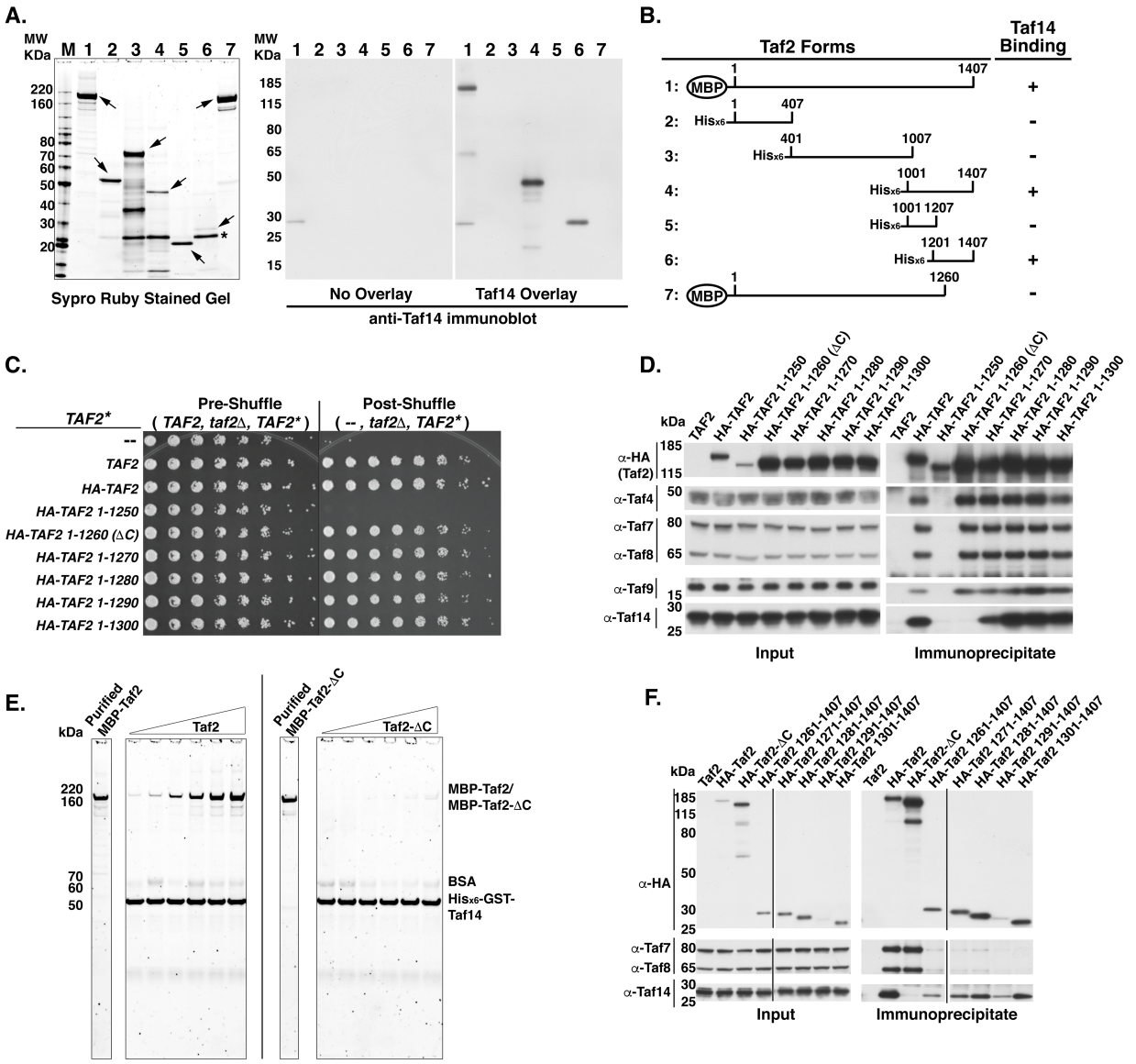


Figure 3.6: The Taf2 C-terminus Is Necessary and Sufficient for Binding to Taf14 *in vitro* and *in vivo*. *A.* Far Western-Blotting. Purified Taf2 forms were analyzed in triplicate as in Figure 3.5, one replicate analyzed by Sypro Ruby gel stain (shown in ***A, Left***) while the other two were analyzed by anti-Taf14 Far-Western blotting, with and without His_6-Taf14 overlay (shown in ***A, Right***). Two technical replicates were performed. ***B.*** Schematic of Taf2 forms used in panel ***A***; numbers refer to the proteins loaded in Lanes 1-7 panel ***A***. ***C,D.*** Plasmid Shuffle Complementation and Taf2-TFIID Co-IP. Taf2 C-terminal truncation variants (TAF2*: no ORF (--), *TAF2* full length aa 1-1407, *HA_3NLS-TAF2* full length, and successive *HA_3NLS-TAF2* truncation variants ranging from aa 1-1251 to aa 1-1300) were analyzed for their ability to complement a *taf2* null strain via plasmid shuffle (***D.***) and to coIP TFIID subunits, as described in Figure 3.2. Additional anti-Taf IgGs were used in these analyses (α-Taf4, α-Taf9).

1290 and Taf2 aa 1-1300 all maintained the ability to co-precipitate Taf14 at levels similar to WT. However, Taf2 aa 1-1270 showed a mild reduction in the ability to co-precipitate Taf14 and Taf2 aa 1-1260 (hereafter referred to as Taf2-ΔC) completely lost the ability to coIP Taf14. Based on this data, I conclude that Taf2 aa 1261-1407 are necessary for Taf2-Taf14 interaction *in vivo*.

Purified MBP-Taf2-DC was tested for its ability to directly interact with Taf14 *in vitro*. MBP-Taf2-DC was subjected to both Far Western blotting (Figure 3.6a lane 7) as well as GST-pull downs (Figure 3.6e), as described above. MBP-Taf2-ΔC displayed no observable binding to His_6-Taf14 in the Far Western blot or His_6-GST-Taf14 in the GST-pull down. Thus, Taf2 aa 1261-1407 are necessary for Taf14 binding *in vitro*.

To assess if the Taf2 C-terminus is sufficient for binding to Taf14 *in vivo*, untagged Taf2, HA-Taf2, HA-Taf2-ΔC and HA-Taf2 C-terminal fragments (aa 1261-1407, 1271-1407, 1281-1407, 1291-1407 and 1301-1407) were subjected to anti-HA coIP, as described above (Figure 3.6f). As expected, HA-Taf2 could co-IP Taf7, Taf8 and Taf14 while HA-Taf2-ΔC could coIP Taf7 and Taf8 but could not coIP Taf14. Surprisingly, all of the Taf2 C-terminal fragments displayed the ability to coIP Taf14. Despite Taf2 aa 1261-1407 being necessary for binding to Taf14 *in vivo*, aa 1261-1407 and smaller C-terminal fragments are sufficient for binding to Taf14 *in vivo*. Considering TFIID contains multiple copies of Taf14

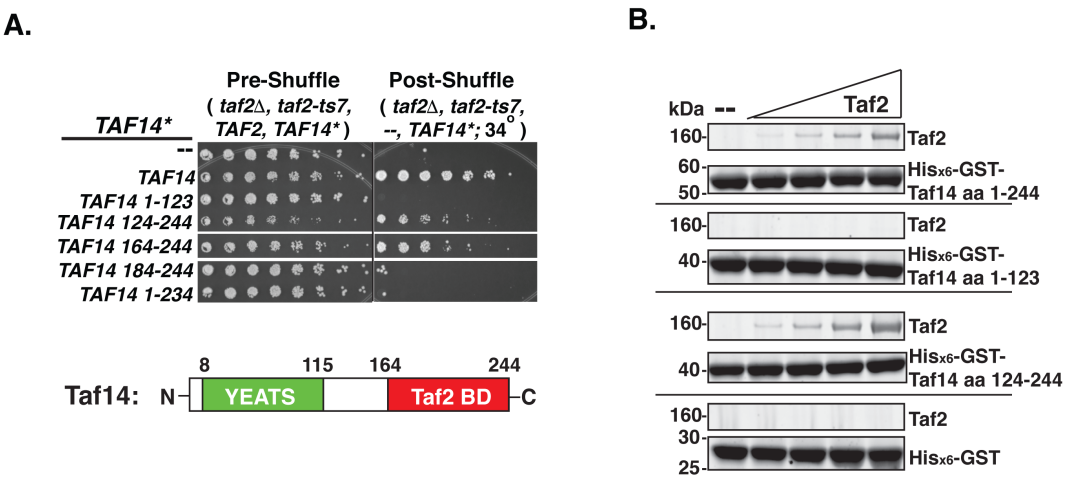
Figure 3.6 Continued:
E. GST Pull-downs. Similar to Figure 5B except 5 picomole of His_6-GST-Taf14 was incubated with 0.1mg/mL BSA and between 1 and 32 pmol MBP-Taf2 or MBP-Taf2-ΔC. Two independent binding assays were performed, once with 2 pmol His_6-GST-Taf14 and once with 5 pmol His_6-GST-Taf14, both displaying high affinity saturable binding for MBP-Taf2 and a lack of specific binding to MBP-Taf2-ΔC. One picomole of purified MBP-Taf2 and one picomole of purified MBP-Taf2-ΔC were loaded to demonstrate purity of input material. ***F.*** Taf2-TFIID Co-IP. Performed as described in Figure 3.2. Black line depicts non-contiguous lanes from the same SDS-PAGE gel and from the same film exposures. One technical replicate was performed. ΔC = Taf2 aa 1-1260.

per TFIID molecule (Sanders et al., 2002a), these data allowed me to hypothesize that the Taf2 C-terminus contains multiple Taf14 binding domains that are independently capable of promoting Taf14-TFIID association.

Defining the Minimal Taf2 Interaction Domain in Taf14

In parallel to my analyses to map the Taf14 interaction domain in Taf2, I also subjected *TAF14* to systematic truncation mutagenesis to identify the Taf14 domain operative in suppressing the *taf2-ts7* Ts phenotype. Previous Taf14 studies have shown that the N-terminus contains the YEATS domain while the Taf14 C-terminus can both complement the growth deficiencies associated with a *taf14* null strain and associate with transcription related complexes such as TFIID and TFIIF *in vivo* (Schulze et al., 2010). In addition, C-terminally-tagged Taf14 variants display defects in growth, likely because the tag negatively impacts the ability of Taf14 to interact with transcription-related complexes (Kabani et al., 2005).

In strains harboring *taf2-ts7*, Taf14 full length (aa 1-244), the Taf14 N-terminus (aa 1-123) and the Taf14 C-terminus (aa 124-244) were overexpressed to determine if these fragments could suppress the Ts phenotype (Figure 3.7a). All variants were expressed with a $FLAG_2$-NLS N-terminal tag. While Taf14 aa 1-244 displayed the most robust suppression of the *taf2-ts7* Ts phenotype, Taf14 aa 124-244 also suppressed the *taf2-ts7* Ts phenotype. The N-terminal YEATS domain-containing fragment (aa 1-123) could not suppress the *taf2-ts7* Ts phenotype. I further dissected the Taf14 C-terminus in this overexpression suppression assay by performing systematic N- and C-terminal truncations of Taf14 aa 124-244 (Figure 3.7a and data not shown). These analyses identified that the



C.

***TAF14* ala Scan Mutants**

	*TAF14**	*taf2-ts7 Suppression*	
	TAF14	Suppression Up to 37°	
m1	V177A, D178A, L179A	No Suppression @ 37°	
	V181A, D182A, A183	No Suppression @ 37°	
	L186A, T187A, K188A	No Suppression @ 37°	
	M200A, V201A, T202A	No Suppression @ 37°	
	G218A, E219A, F220A	No Suppression @ 37°	
	F220A, I221A, I222A	No Suppression @ 30°, 34° or 37°	
	I222A, D223A, L224A	No Suppression @ Any Temp.	Mutational Hotspot
	L224A, Y225A, S226A	No Supression @ 34° or 37°	
	S226A, L227A, P228A	No Suppression @ 37°	
	P228A, E229A, G230A	No Suppression @ 37°	
m11	S234A, L235A, W236A	No Suppression @ 37°	

Figure 3.7: The Taf14 C-terminus Physically and Functionally Interacts with Taf2. ***A.*** *TAF14* Overexpression Suppression of *taf2-ts7*. (Upper) Plasmid shuffle performed as in Figure 3.3C. *taf2* null cells harboring a *URA3*-marked *TAF2* plasmid, a *LEU2*-marked *HA$_3$NLS-taf2-ts7* plasmid and a *HIS3*-marked *TAF14** overexpression plasmid (*TAF14* and *TAF14* N- and C-terminal truncation fragments; all *TAF14* forms contained an N-terminal FLAG$_2$NLS-tag) were spotted on SC-L,-H medium (pre-shuffle) or SC-L,-H +5-FOA (post-shuffle) medium and grown for 72 hours at 34°. Representative image of two biological replicates is presented. (Lower) Schematic represents the domain structure of Taf14. Coordinates for YEATS domain derived from (Shanle et al., 2015). Coordinates for Taf2 Binding Domain (BD) defined in this study. ***B.*** Glutathione S-Transferase (GST) Pull-Downs. 16pmol of purified recombinant His$_6$-GST, 8pmol of His$_6$-GST-Taf14, 12pmol of His$_6$-GST-Taf14 aa 1-123 or 12pmol of His$_6$-GST-Taf14 aa 124-244 were mixed with 0.1mg/mL BSA and either no Taf2 or 0.78pmol Taf2, 1.56pmol Taf2, 3.125pmol Taf2 and 6.25pmol of Taf2. His$_6$-GST or His$_6$-GST-Taf14 variant bound proteins were pulled-down with glutathione agarose, washed with binding buffer, analyzed via SDS-PAGE and stained with Sypro Ruby gel stain. One technical replicate was performed. ***C.*** Summary of overexpression suppression of *taf2-ts7* with *FLAG$_2$NLS-TAF14* ala scan variants. Horizontal lines separate non-contiguous mutant variants. Results were derived from serial dilution spot assays as described in ***A*** (not shown). Two biological replicates were performed.

domain minimally required for suppression of the *taf2-ts7* Ts phenotype lies within Taf14 aa 164-244.

Again, the simplest model for the ability of the Taf14 C-terminus to suppress the *taf2-ts7* Ts phenotype is that the C-terminus of Taf14 directly interacts with Taf2. To test this hypothesis, purified recombinant His_6-GST, His_6-GST-Taf14 aa 1-244, His_6-GST-Taf14 aa 1-123 and His_6-GST-Taf14 aa 124-244 were mixed with purified Taf2 and, in the presence of BSA, subjected to GST-pull down, as described above (Figure 3.7b). His_6-GST-Taf14 aa 1-244 and His_6-GST-Taf14 aa 124-244 were able to pull-down purified Taf2 in a dose-dependent manner whereas His_6-GST and His_6-GST-Taf14 aa 1-123 did not pull-down any detectable Taf2. These data confirm that the Taf14 C-terminus, the domain required for suppression of the *taf2-ts7* Ts phenotype, is necessary and sufficient for direct interaction with Taf2 *in vitro*.

To identify point mutants that disrupt the *TAF2-TAF14* genetic interaction, I performed 3ala scan mutagenesis with one aa overlap of Taf14 aa 164-244. These variants were overexpressed in the context of full length Taf14 to determine if they could suppress the *taf2-ts7* Ts growth phenotype. Of the 39 ala variants generated, I identified 11 that displayed defects in the ability to suppress the *taf2-ts7*-associated Ts phenotype at 37^{o}. The results of the suppression analyses for these 11 *taf14-ala* variants are summarized in Figure 3.7c. A particularly sensitive hotspot was identified between Taf14 aa 218-230 where every mutant variant (m5 through m10) displayed a defect in *taf2-ts7* Ts growth suppression. These amino acids are likely critical for the function of the Taf14 C-terminus.

Fine Mapping of the Taf2 C-terminus Reveals Two Taf14 Interaction Domains

A previous study reported that the Taf2 C-terminus is insoluble when expressed in *E. coli* (Kabani et al., 2005). My studies confirm this observation. Expression of Taf2 aa 1301-1407-His_6 is largely insoluble and refractory to native purification with Ni^{2+}-NTA agarose (Figure 3.8c, lane 1). Our lab previously demonstrated that co-expression of insoluble Tafs with their cognate binding partner results in solubilization (Gangloff et al., 2001b). Therefore, I hypothesized that co-expression of the Taf2 C-terminus with either full length Taf14 or Taf14 aa 164-244 would result in solubilization of the Taf2 C-terminal fragment. To this end, I co-expressed Taf2 aa 1301-1407-His_6 with either Taf14 aa 1-244 or Taf14 aa 164-244 and subjected these complexes to Ni^{2+}-NTA agarose purification. Using this strategy, I could generate soluble Taf2-Taf14 complexes in milligram quantities/liter of *E. coli* culture (Figure 3.8a and 3.8b, respectively). I also attempted to co-express full length Taf2 with Taf14 aa 164-244 in *E. coli* but these attempts were unsuccessful (data not shown).

I then employed this co-purification assay to fine map the Taf14 binding domain within Taf2 aa 1301-1407. N- and C-terminal truncations of Taf2 aa 1301-1407-His_6 were co-expressed with Taf14 aa 164-244 in *E. coli* and subjected to Ni^{2+}-NTA agarose purification (Figure 3.8c, diagrammed in 3.8d). These analyses revealed that Taf2 aa 1381-1407 were sufficient for Taf14 co-purification while Taf2 aa 1362-1407 were necessary for Taf14 co-purification. Henceforth, I will refer to Taf2 aa 1363-1407 as Domain 1. Surprisingly, Domain 1 not only binds to Taf14 but also contributes to the insolubility of the Taf2 1301-1407 fragment since fragments deleted for Domain 1 can be purified from *E. coli* without co-purifying Taf14.

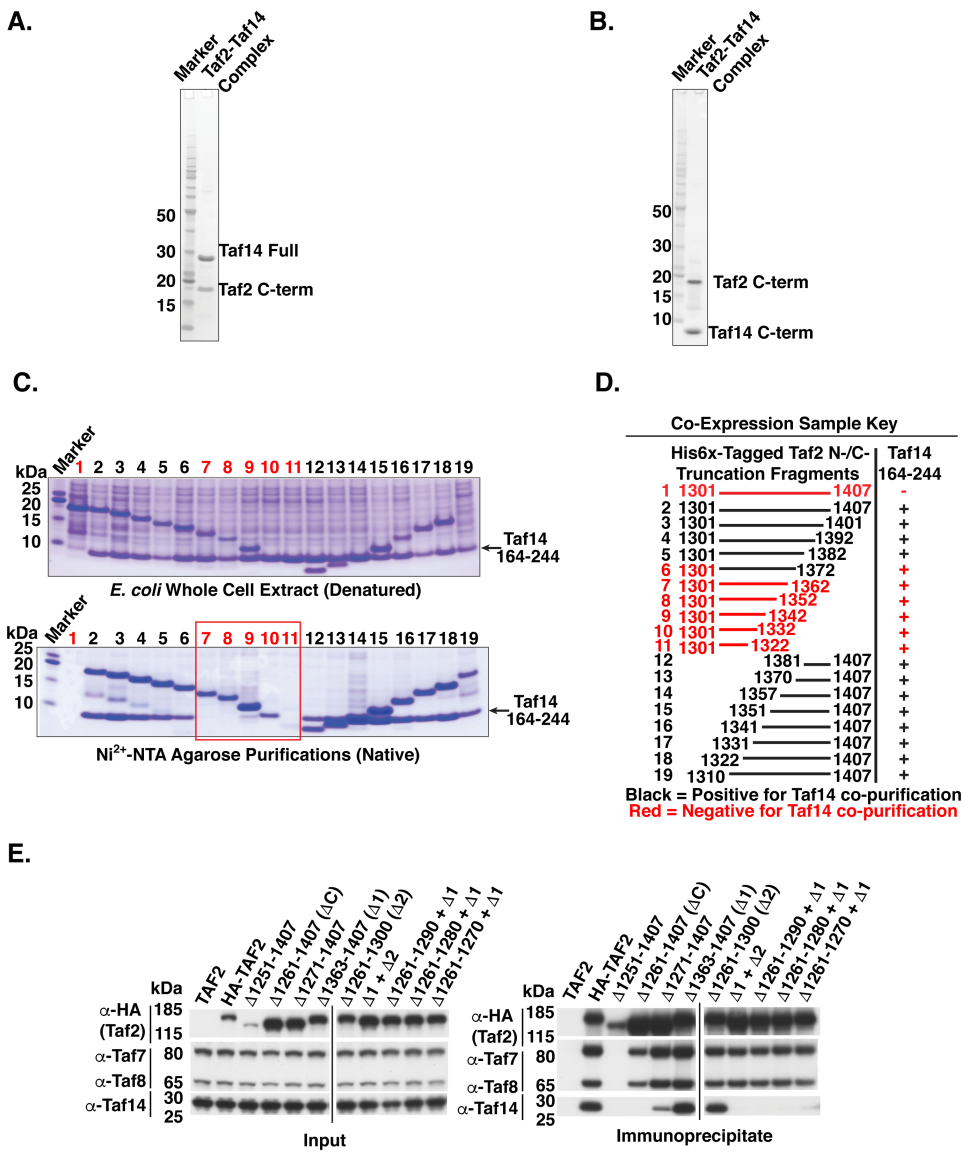


Figure 3.8: Taf2 Amino Acids 1261-1407 Contain Two Domains that Contribute to Taf14 Binding. ***A,B.*** Co-purification of Taf2 and Taf14. Taf2 aa 1301-1407-His_6 was co-expressed in *E. coli* with either Taf14 aa 1-244 or Taf14 aa 164-244 and purified using Ni^{2+}-NTA agarose. Purified Taf2 aa 1301-1407/Taf14 aa 1-244 (shown in **A.**) and Taf2 aa 1301-1407/Taf14 aa 164-244 (shown in **B.**) were separated via SDS-PAGE and stained with coomassie blue. ***C,D.*** Fine mapping of the Taf14 interaction domain in Taf2 aa 1301-1407. C-terminally hexa-histidine tagged Taf2 aa 1301-1407 and N- and C-terminal truncation variants of Taf2 aa 1301-1407 (diagrammed in ***D.***) were co-expressed with Taf14 aa 164-244 in *E. coli*. As a negative control, C-terminally hexa-histidine tagged Taf2 aa 1301-1407 was expressed alone. Denatured whole cell extracts or Ni^{2+}-NTA agarose purified complexes were separated on SDS-PAGE and stained with coomassie blue (**C.**). Samples that co-purify Taf14 aa 164-244 are shown in black while samples that fail to co-purify Taf14 aa 164-244 are shown in red. A representative image is shown of two technical replicates. ***E.*** Taf2-TFIID Co-IP. Performed as described in Figure 3.2. All deletion variants contain an HA_{x3}NLS-tag. Black lines separate non-contiguous lanes from the same SDS-PAGE gel and from the same film exposures. One technical replicate was performed. ΔC = deletion of Taf2 aa 1261-1407. Δ1 = deletion of Taf2 aa 1363-1407. Δ2 = deletion of Taf2 aa 1261-1300.

To assess the relevance of Taf2 Domain 1 to binding to Taf14 *in vivo*, I performed anti-HA coIPs with a series of Taf2 C-terminal deletion variants, as described above (Figure 3.8e). Considering Taf2-ΔC (aa 1-1260) fails to interact with Taf14 *in vivo* and *in vitro* and the Taf2 aa 1301-1362-His_6 fragment fails to co-purify Taf14, I reasoned that a second Taf14 binding domain likely resides within Taf2 aa 1261-1300 (hereafter, Domain 2). Deletion of either Domain 1 (Δ1) or Domain 2 (Δ2) had no impact on the ability of Taf2 to co-precipitate TFIID subunits Taf7, Taf8 or Taf14. However, a Taf2 double deletion variant (Δ1+Δ2) could co-precipitate Taf7 and Taf8 but failed to co-precipitate Taf14. Furthermore, successively smaller deletion within Domain 2 (Δ1261-1291, Δ1261-1281 and Δ1261-1271), when combined with Δ1, displayed strong defects in the ability to co-precipitate Taf14. Taken together, these data are consistent with the Taf2 C-terminus containing at least two distinct domains that can independently facilitate incorporation of Taf14 into the TFIID complex. Thus, to completely abrogate association of Taf14 with the TFIID complex, both of these domains must be disrupted.

The Taf14 Binding Domains in Taf2 are Necessary for *TAF14* Overexpression Suppression of the *taf2-ts7* Growth Defect

My overarching hypothesis has been that *TAF14* overexpression suppression of the *taf2-ts7* Ts phenotype occurs via a Taf2-Taf14 direct interaction. By extension, if Taf2 and Taf14 could no longer physically interact, then *TAF14* overexpression would no longer be able to suppress the *taf2-ts7* Ts phenotype.

To test this hypothesis, I performed genetic complementation assays with *TAF2, taf2-ΔC, taf2-ts7* and *taf2-ts7-ΔC*, with or without *TAF14* overexpression (Figure 3.9a). As

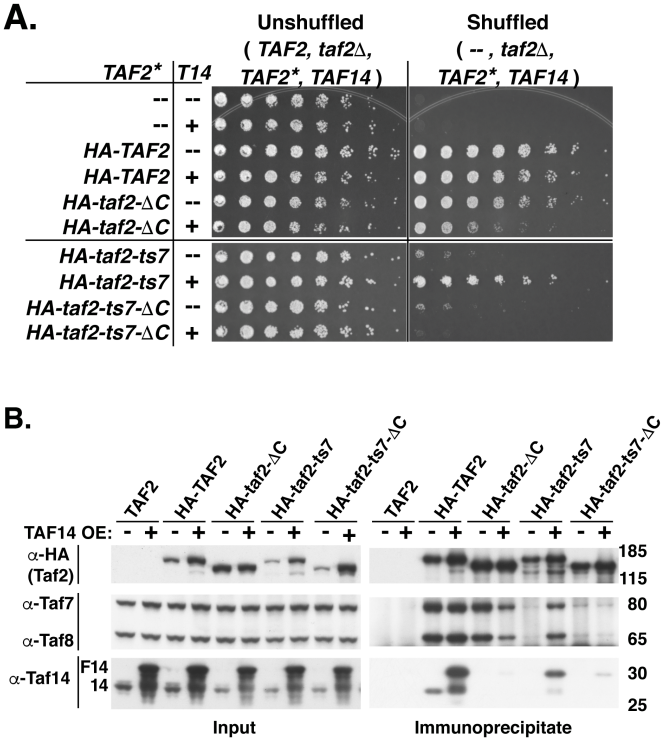


Figure 3.9: The Taf14 Binding Domain in Taf2 is Necessary for *TAF14* Overexpression-Mediated Suppression of *taf2-ts7*. *A.* Intragenic *TAF2* Synthetic Genetics. Plasmid Shuffle complementation assays performed as described in Figure 3C. Cells were grown for 72 hours at 25° prior to imaging. Representative images from at least two biological replicates are presented. *TAF2**: no ORF (--), *TAF2*, *HA$_3$NLS-TAF2, HA$_3$NLS-TAF2-ΔC, HA$_3$NLS-taf2-ts7* and *HA$_3$NLS-taf2-ts7-ΔC*. ***B.*** Taf2-TFIID Co-immunoprecipitation. Performed as described in Figure 3.2 except strains also contained either an empty *HIS3*-marked overexpression plasmid or a *FLAG$_2$NLS-TAF14* cDNA containing *HIS3*-marked overexpression plasmid. Representative image of at least two biological replicates. *TAF2**: no ORF (--), *TAF2*, *HA$_3$NLS-TAF2, HA$_3$NLS-TAF2-ΔC, HA$_3$NLS-taf2-ts7* and *HA$_3$NLS-taf2-ts7-ΔC*. Taf14 WT labeled 14; FLAG$_2$NLS-Taf14 labeled F14.

shown before, overexpression of *TAF14* did not have a strong impact on growth of strains harboring *TAF2*. However, overexpression of *TAF14* in strains harboring *taf2-ΔC* resulted in a synthetic slow growth phenotype. Similarly, while *TAF14* overexpression suppressed the *taf2-ts7* Ts phenotype, it did not suppress the *taf2-ts7-ΔC* phenotype but, instead, caused synthetic lethality. Consistent with my hypothesis, the Taf14 binding domains in the Taf2 C-terminus are required for suppression of the *taf2-ts7* Ts phenotype.

Furthermore, my hypothesis also predicts *TAF14* overexpression drives Taf2-Taf14 complex formation resulting in stable incorporation into the TFIID complex. To test this hypothesis, I performed anti-HA coIP analysis, as described above, to determine the impact of *TAF14* overexpression on the ability of Taf2, taf2-ΔC, taf2-ts7 and taf2-ts7-ΔC to co-precipitate TFIID subunits (Figure 3.9b). The strains used for these analyses were

pseudodiploid for both *TAF2*, containing WT and a test *TAF2* allele (*TAF2**), as well as *TAF14,* containing genomically-encoded WT *TAF14* and either an empty overexpression plasmid or an expression plasmid containing $FLAG_2$-NLS-tagged *TAF14*. *TAF14* overexpression had no impact on the ability of Taf2 to coIP Taf7 and Taf8; however, Taf2 co-precipitated elevated levels of Taf14, compared to the no *TAF14* overexpression strain. Consistent with the synthetic sick growth phenotype, taf2-ΔC reproducibly displayed a modest reduction in the ability to coIP Taf7 and Taf8 in strains that overexpressed *TAF14*. Validating my hypothesis, *TAF14* overexpression rescued the ability of taf2-ts7 to coIP TFIID subunits Taf7, Taf8 and Taf14. However, taf2-ts7-ΔC was not responsive to *TAF14* overexpression and still failed to efficiently coIP TFIID subunits Taf7, Taf8 and Taf14.

Replacing the Taf14 Binding Domain in Taf2 with Taf14 via Gene Fusion Partially Suppresses the *taf2-ts7*-associated Growth Defects

My data suggests that the domain necessary for Taf2-Taf14 interaction resides within Taf2 aa 1261-1407 while the location of the mutations in *taf2-ts7* and *ts8* reside between Taf2 aa 1100-1150. Based on this discrepancy, I conceived of two possible mechanisms through which *TAF14* overexpression can suppress the growth phenotypes associated with these Ts mutants. First, saturating the C-terminus of taf2-ts7 and ts8 with Taf14 induces a conformational change that allows these variants to more stably associate with the TFIID complex. Second, when at saturating levels, Taf14 can tether taf2-ts7 and ts8 to the TFIID complex, likely through direct interaction between Taf14 and another TFIID subunit(s). The first model inherently requires a binding event where Taf14 binding

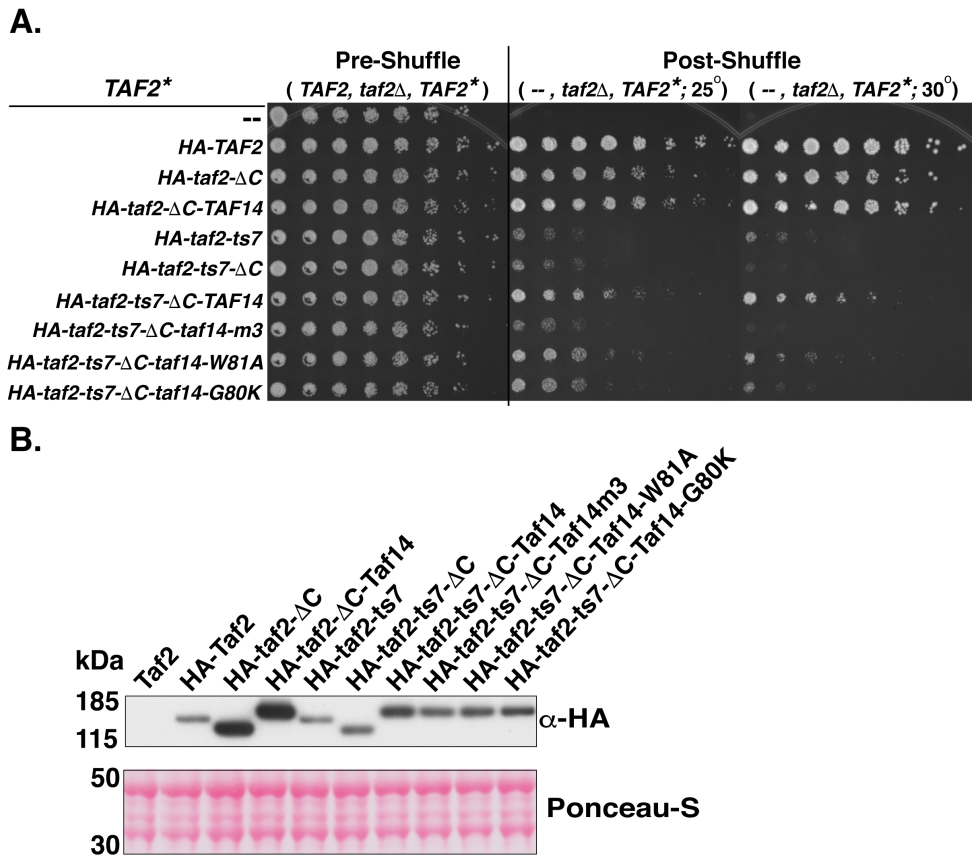


Figure 3.10: Replacing the Taf14 Binding Domain in Taf2 with Taf14 Partially Suppresses the *taf2-ts7* growth defect. ***A.*** Plasmid Shuffle Genetic Complementation. Performed as described in Figure 3.2. Cells were grown for 60 hours at 25° and 30° prior to imaging. Plate images representative of at least two biological replicates. TAF2*: no ORF (--), *TAF2*, *HA$_3$NLS-TAF2*, *HA$_3$NLS-TAF2-ΔC*, *HA$_3$NLS-TAF2-ΔC-TAF14* chimeric fusion, *HA$_3$NLS-taf2-ts7*, *HA$_3$NLS-taf2-ts7-ΔC*, *HA$_3$NLS-taf2-ts7-ΔC-TAF14* chimeric fusion, *HA$_3$NLS-taf2-ts7-ΔC-taf14 m3* chimeric fusion, *HA$_3$NLS-taf2-ts7-ΔC-taf14-W81A* chimeric fusion and *HA$_3$NLS-taf2-ts7-ΔC-taf14-G80K* chimeric fusion. ***B.*** Steady-State Protein Immunoblot. Total cell protein was fractionated via SDS-PAGE and transferred to PVDF membrane. High molecular weight part of the membrane was probed with anti-HA antibody and total protein on the lower molecular weight part of the membrane was visualized with Ponceau-S as a loading control.

to taf2-ts7 or ts8 induces a conformational change. The second model involves Taf14 playing a role outside of simply Taf2 binding.

To distinguish between these two models, I constructed Taf2-Taf14 chimeras where *taf2-ΔC* and *taf2-ts7-ΔC* were fused to the *TAF14* ORF. Ideally, these chimeric fusions would bypass the need for *TAF14* overexpression to achieve saturable binding to the Taf2 C-terminus since these chimeras contain covalently attached Taf14. These

chimeras were tested to see if they could complement a *taf2* null strain (Figure 3.10a). A "WT" chimera (*taf2-ΔC-Taf14)* supported growth at a level similar to *TAF2* suggesting the chimera does not negatively impact Taf2 function. This *taf2-ts7-ΔC-TAF14* chimera suppressed the slow growth phenotype at 30° associated with *taf2-ts7* and *taf2-ts7-ΔC*.

I then sought to determine if this suppression was Taf14-dependent and, if so, which domain(s) was involved. To this end, I made *taf2-ts7-ΔC-taf14* chimeras with mutations either in the Taf2 Binding Domain of Taf14 (*taf14-m3*: L186A, T187A, K188A) or in the Taf14 YEATS domain that disrupt its ability to bind to acetylated or crotonylated H3K9 (*taf14-W81A* and *taf14-G80K*) (Andrews et al., 2016; Shanle et al., 2015). Mutations in the Taf14 C-terminus completely abrogated the ability of the chimera to suppress the growth defect associated with *taf2-ts7-ΔC.* In addition, while both YEATS domain mutant fusions displayed some reduction in the ability to suppress the growth defects associated with *taf2-ts7-ΔC*, the G80K mutation displayed a significant loss in the ability to suppress. The loss of suppression cannot be attributed to a reduction in steady-state protein levels since all constructs were expressed at least as well as WT Taf2 (Figure 3.10b).

Taf14-less TFIID Mutant Cells Display a Slow Growth Phenotype and Defects in RPG Transcript Abundance

My results suggest that the Taf2 C-terminus is not only required for Taf2-Taf14 interaction but it is also required for association of Taf14 with the TFIID complex. To test this hypothesis, I engineered two strains: one that genomically encodes an HA_1-tag at the N-terminus of Taf1 for anti-HA immunopurification and one that encodes both an HA_1-Taf1 and genomic deletion of Taf2 aa 1261-1407. TFIID was purified from these two strains as

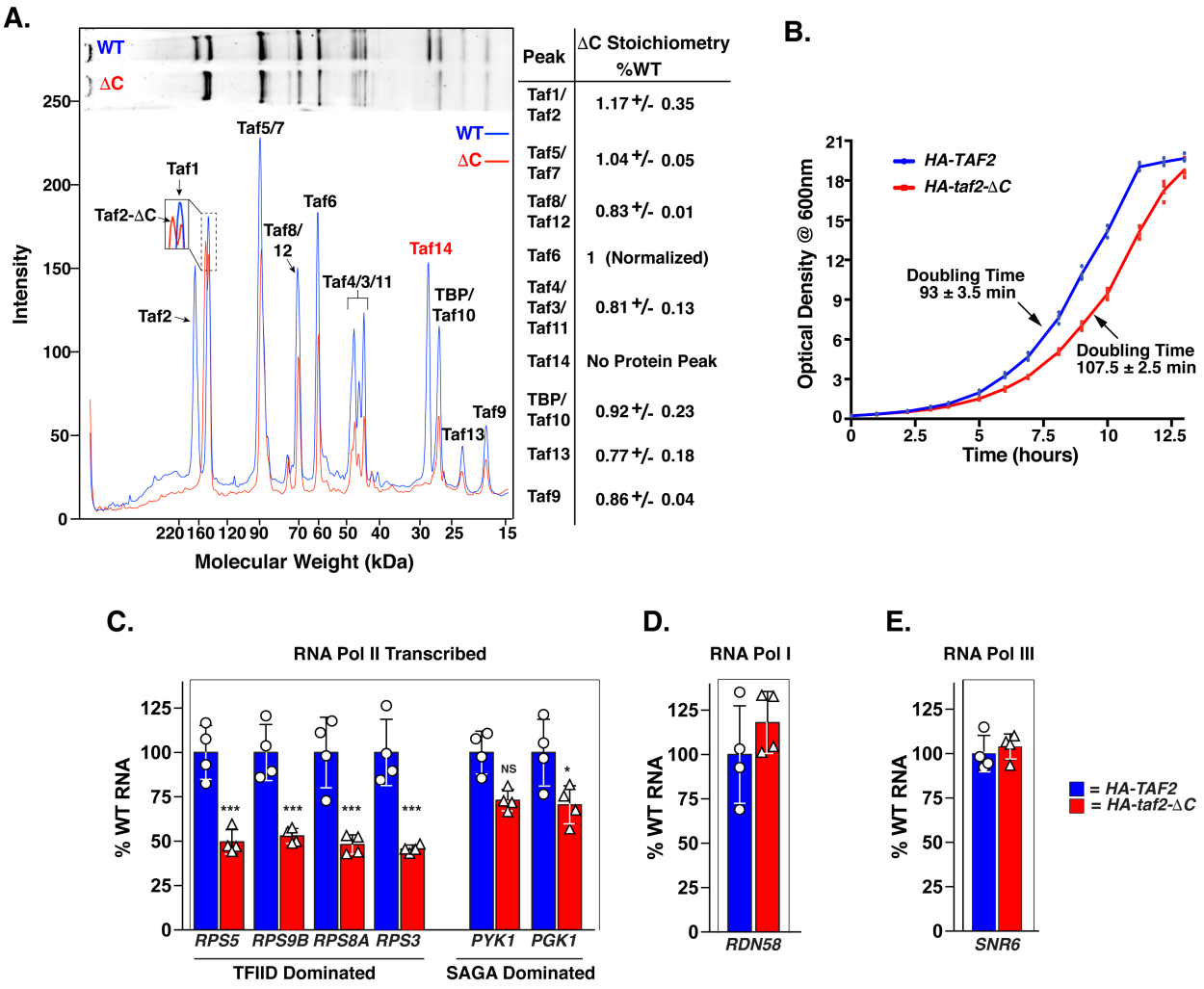


Figure 3.11: Taf14-less TFIID Mutant Cells Display a Slow Growth Phenotype and Defects in Ribosomal Protein Gene Transcription. ***A.*** Purified TFIID Forms. Between 300 and 600 nanograms of HA_1-Taf1 purified TFIID and HA_1-Taf1 Taf2-ΔC purified TFIID were separated via SDS-PAGE and stained with Sypro Ruby gel stain. Peak traces were generated using Quantity One (Bio-Rad). TFIID subunit peaks were quantified in ImageJ by calculating the area under the curve. The signal intensity of Taf6 was used to normalize each TFIID preparation to determine relative subunit stoichiometry for each peak. Mean ± SD is depicted. Quantitation generated from two technical replicates. ***B.*** Growth Rate Analysis. Log phase growing *taf2* null cells shuffled to contain either plasmid-borne *HA_3NLS-TAF2* or plasmid-borne *HA_3NLS-TAF2-ΔC* were diluted to an optical density (OD) at 600nm= ~0.2 (*HA_3NLS-TAF2*, average start OD_{600} = 0.229; *HA_3NLS-TAF2-ΔC*, average start OD_{600} = 0.239), grown at 30º and OD_{600} was measured approximately every hour until the strains reached stationary phase growth. Doubling time was calculated using a non-linear exponential growth fit (GraphPad Prism) for the early phase of the growth curve (first 7 time points). Growth curve was performed with four biological replicates. Experimental error in doubling time is derived from S.E.M.

described in Chapter II,III Methods and then subjected to SDS-PAGE and gel staining to score TFIID subunit composition and stoichiometry (Figure 3.11a). When comparing the TFIID variants, two differences were apparent. First, in the Taf2-ΔC TFIID, the size of Taf2 was reduced reflecting the genomic Taf2 C-terminal deletion. However, Taf2-ΔC is maintained at an apparent 1:1 Taf2-TFIID stoichiometric ratio relative to Taf1, similar to HA_1-Taf1 TFIID, indicating that the deletion does not negatively impact the stability of Taf2 in the TFIID complex. Second, in the Taf2-ΔC TFIID, Taf14 is completely absent, consistent with our Taf2 C-terminus truncation analyses. Quantitation of these TFIID preparations demonstrates that the stoichiometry for the rest of the TFIID subunits is similar between the two TFIID forms. Based on these data, I can conclude that Taf2 aa 1261-1407 are necessary for Taf14 stable incorporation into the TFIID complex and that strains lacking Taf2 aa 1261-1407 have TFIID devoid of Taf14 (Taf14-less TFIID).

When I streaked these TFIID purification strains onto rich medium to isolate single colonies, I found that the Taf14-less TFIID strain displayed a reduced growth rate as measured by colony size (data not shown). To determine if this phenotype was directly attributable to the Taf2-ΔC variant, growth curves were performed in yeast strains that contain only HA-*TAF2* or HA-*taf2-ΔC*. These growth curves revealed that strains harboring

---

Figure 3.11 Continued:
C,D,E. qRT-PCR. Steady-state RNA was analyzed by qRT-PCR scoring for ***C.*** RNA Pol II, ***D.*** RNA Pol I and ***E.*** RNA Pol III transcribed genes. Data were generated from four biological replicates. Each data point in the graph represents one biological replicate and is generated from the average of three technical replicates. Results were statistically analyzed using a 2-way ANOVA with Sidak's multiple comparisons test (GraphPad Prism). Mean ± SD is depicted. *= $p<0.05$, ***= $p\leq0.0001$. Blue = shuffled strain harboring plasmid-borne *HA_3NLS-TAF2*, Red = shuffled strains harboring plasmid-borne *HA_3NLS-TAF2-ΔC*.

HA-*TAF2-ΔC* displayed a 14.5 min (15.6%) slower growth rate during log phase at 30^{o} (Figure 3.11b).

I then assessed steady-state transcript abundance for these two strains. RNA was extracted from mid-log phase cells growing at 25^{o} and analyzed using qRT-PCR. Despite no reduction in transcript abundance for RNA polymerase I transcribed *RDN58* (Figure 3.11d) and RNA polymerase III transcribed *SNR6* (Figure 3.11e), I reproducibly observed a statistically significant ~2-fold reduction in RPG transcript abundance (*RPS5*, *RPS9B*, *RPS8A*, *RPS3*) (Figure 3.11c). In addition, I observed a moderate (~25%), though not statistically significant for both genes, reduction in steady-state transcript abundance for the SAGA-dominated glycolytic *PGK1* and *PYK1* genes. Although the glycolytic genes are not considered TFIID-dependent, my data are consistent with previously observed modest reductions in *PGK1* steady-state transcripts in *taf4^{ts}* variants (Layer and Weil, 2013).

DISCUSSION

Multiple structural and biochemical studies have attributed specific Tafs with promoter-DNA or modified chromatin-binding capabilities (Andrews et al., 2016; Burke and Kadonaga, 1997; Chalkley and Verrijzer, 1999; Gazit et al., 2009; Jacobson et al., 2000; Kaufmann et al., 1998, 1996; Lauberth et al., 2013; Oelgeschläger et al., 1996; Shanle et al., 2015; Shao et al., 2005; Verrijzer et al., 1994). These activities provide convenient mechanisms through which TFIID can engage with genes to facilitate PIC formation. Indeed, an activator-TFIID-TFIIA promoter DNA quaternary complex, with distinct TFIID-DNA interactions, displays a locked DNA conformational state that likely serves as a

platform for general transcription factor and Pol II binding (Andrews et al., 2016; Burke and Kadonaga, 1997; Chalkley and Verrijzer, 1999; Gazit et al., 2009; Jacobson et al., 2000; Kaufmann et al., 1998, 1996; Lauberth et al., 2013; Oelgeschläger et al., 1996; Shanle et al., 2015; Shao et al., 2005; Verrijzer et al., 1994). However, these many biochemical activities have rarely been interrogated genetically to establish their importance *in vivo* (Lauberth et al., 2013). This lack of knowledge was the impetus for the work reported in this dissertation.

Specifically, the functional role of Taf2 *in vivo* was largely unknown, despite longstanding evidence for its *in vitro* INR binding activity (Chalkley and Verrijzer, 1999; Kaufmann et al., 1998, 1996). A lack of Taf2 molecular genetic studies is likely a result of technical challenges. In addition to being essential for life, Taf2 is a large protein and particularly labile in ScTFIID purifications (Papai et al., 2009; Ray et al., 1991). As this chapter demonstrates, systematic deletion analyses largely disrupt Taf2 protein stability and the ability of Taf2 to stably incorporate into the TFIID complex, precluding conventional methods of genetically interrogating large proteins (Table 3.1) (Bai et al., 1997; Knutson and Hahn, 2011; Singh et al., 2004). Our site-directed mutagenesis approach generated variants that displayed similar TFIID-incorporation defects, limiting our ability to interpret their precise molecular functions (Figure 3.2). Of note, while this manuscript was in preparation, specific amino acids were predicted to be important for the INR binding function of Taf2 (Louder et al., 2016). One of the site-directed mutants generated in this study, *taf2-m32,* targeted a subset of these residues but displayed only a mild slow growth phenotype and was not further pursued (Table 3.2). Genetic

interrogation of this putative INR binding domain will likely shed insights into the role of INR binding in TFIID transcriptional activation.

The key finding in this chapter is the discovery of a genetic interaction between *TAF2* and *TAF14*. Individual overexpression of each TFIID subunit identified *TAF14* overexpression as a mechanism to achieve suppression of select *taf2*^{ts} alleles (*taf2-ts7* and *ts8*). The location of the residues mutated in these Ts variants is suggestive of a functional domain. Furthermore, I demonstrated that Taf2 and Taf14 directly interact. Molecular genetic dissection of both Taf2 and Taf14 led to the identification of the domains responsible for physical and functional interaction *in vivo* and *in vitro*.

Our lab's previous structural and biochemical characterization of the TFIID complex identified the stoichiometry and location of all of the TFIID subunits with the exception of Taf14 (Leurent et al., 2004, 2002; Papai et al., 2009; Sanders et al., 2002a). Purified TFIID displays a stoichiometry of at least two copies of Taf14 per TFIID molecule. In addition, Taf14 self-associates *in vivo*. However, gel staining of purified SWI/SNF and TFIIF show one copy per complex, so this multi-copy per complex phenotype is likely to be specific to TFIID (Cairns et al., 1996; Henry et al., 1994). While deletion of the Taf2 C-terminus completely disrupts association of Taf14 with the TFIID complex, fine mapping of the Taf2 C-terminus identified two domains that can independently facilitate Taf14 incorporation into the TFIID complex. This data is consistent with a genome wide two-hybrid screen, which identified part of the Taf2 C-terminus as a Taf14 interacting protein (Kabani et al., 2005). Yet, Taf2 variants that fail to stably incorporate into the TFIID complex as well as Taf2 C-terminal fragments display reductions in Taf14 co-precipitation relative to WT Taf2. This observation suggests that when Taf2 incorporates into the TFIID

complex, Taf14 binding is enhanced, potentially through a multivalent binding site between Taf2 and another TFIID subunit(s).

Interestingly, the Taf2 domain identified to directly interact with Taf14 does not contain the amino acids mutated in *taf2-ts7* or *ts8*, despite this domain being required for *TAF14* overexpression suppression of these variants. To test the possibility of a multivalent Taf14 binding site between Taf2 and another TFIID subunit, I fused *TAF14* to a $taf2^{ts}$ variant deleted for the Taf14 binding domain. This chimeric fusion improved growth in a *TAF14*-dependent manner consistent with the existence of a multivalent binding site. The TFIID subunit responsible is likely to be Taf8. Human Taf2 directly interacts with Taf8 (Trowitzsch et al., 2015). Taf1-TAP purifications of TFIID result in substoichiometric levels of Taf2, Taf8 and Taf14 (Papai et al., 2009). Furthermore, specific deletions in Taf1 result in dissociation of Taf2, Taf8 and Taf14 from the TFIID complex suggesting that these three subunits form a subcomplex (Singh et al., 2004). As with human TFIID, the association of Taf2 and Taf8, along with Taf14, with the TFIID core may stimulate a structural transition in the TFIID assembly pathway (Bieniossek et al., 2013).

The function of Taf14 in transcription regulation has remained enigmatic. Functional interpretation of *TAF14* mutant variants is limited because of its presence in multiple transcription-related complex assemblies (Cairns et al., 1996; Henry et al., 1994; Kabani et al., 2005; Sanders et al., 2002b). Specifically, molecular defects in *taf14* null strains or strains harboring *taf14* mutant variants unable to bind modified H3K9 could be attributed to TFIID-promoter interactions, the role of TFIIF in PIC or elongation function, ATP-dependent chromatin remodeling or the myriad transcription related functions associated with Mediator. The results presented in this chapter begins to decipher the role

of Taf14 in transcription regulation through the identification of *taf2-ΔC*, a separation-of-function Taf2 variant that can stably incorporate into the TFIID complex but precludes incorporation of Taf14. The existence of this variant is consistent with the model presented by Kabani et al. that Taf14 has a particular entry point protein in each complex with which it is associated (Kabani et al., 2005). Thus, it may be possible to generate genetic reagents that specifically dissociate Taf14 from the TFIIF, Ino80, Swi/Snf, NuA3, Mediator and RSC without perturbing complex integrity or the other functions of the complex. *taf2-ΔC* mutant cells display a modest slow-growth phenotype as well as a reduction in transcript abundance for the TFIID-dominated RPGs. In addition, purification of TFIID from strains harboring *taf2-ΔC* yields a Taf14-less TFIID complex. Structural analyses of this complex in contrast with WT TFIID may yield to important insights into the location of Taf14 in TFIID as well as the mechanism by which two copies of Taf14 can associate with a single TFIID molecule.

My data suggests that *taf2-ΔC* is a true separation-of-function variant whose cellular and molecular phenotypes reflect the contribution of Taf14 to TFIID function (Figure 3.12). However, the mechanism by which Taf14 contributes to TFIID transcription activation mechanism remains speculative. The Taf14 YEATS domain responsible for binding to modified chromatin (Andrews et al., 2016; Shanle et al., 2015) may enhance TFIID occupancy of active genes by increasing the number of contact points between TFIID and gene promoters. In addition, the Taf2-Taf14 chimeric fusion analyses suggest that the YEATS domain does contribute to TFIID function, likely through its ability to interact with modified chromatin. This hypothesis is consistent with the observation that the Taf3 PHD finger H3K4me3 binding activity stimulates transcription, especially in context of a mutant

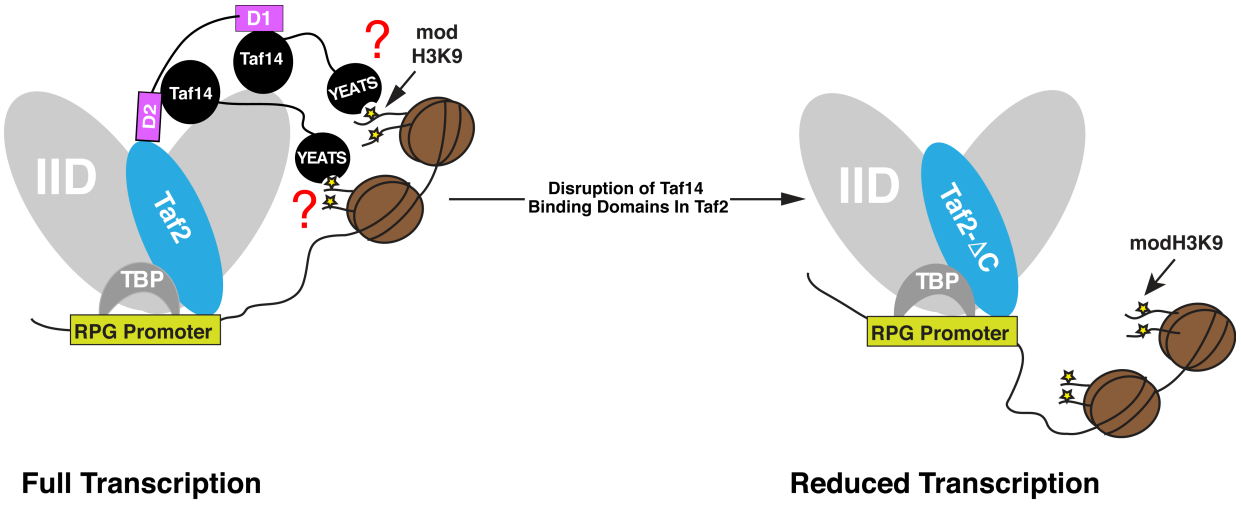


Figure 3.12: Model of Taf2-Taf14 Interaction in TFIID-Dependent Transcription. (Left) Depicts role of Taf2 C-terminus in facilitating incorporation of two Taf14 molecules per TFIID complex. Taf14 may then bind to active chromatin through its YEATS domain. Brown circles = histone octamer, yellow stars = modified H3K9, D1 = the Taf2 Taf14 Binding Domain 1, D2 = the Taf2 Taf14 Binding Domain 2. (Right) Taf14 cannot incorporate into the TFIID complex and as a result TFIID may lose the ability communicate with active chromatin through the Taf14 YEATS domain.

TATA-box (Lauberth et al., 2013). Similar *in vitro* transcription experiments with chromatin templates need to be performed with WT and Taf14-less TFIID to assess the validity of this model for ScTFIID considering the Taf3 PHD finger is not present in the yeast system.

On the other hand, the Taf14 YEATS domain is non-essential and displays minimal growth defects when deleted whereas deletion of the Taf14 C-terminal domain phenocopies the *taf14* null strain (Schulze et al., 2010). Considering the Taf14 C-terminal domain is responsible for the interaction with Taf14-associated complexes, this domain could mediate interactions among these complexes that have thus far not been explored. We do not know if a single Taf14 molecule can bind to multiple transcription-related complexes at the same time or if binding of Taf14 to TFIID or TFIIF, for example, are mutually exclusive. If these interactions are not mutually exclusive, Taf14 could serve as a

bridge between the transcription machinery that could play a key role in the transcription process.

In summary, through systematic mutagenesis of the TFIID subunit Taf2, I have uncovered important physical and functional interactions between Taf2 and Taf14. These discoveries have shed light on the role Taf14 plays in TFIID function, including a putative role in TFIID-chromatin interaction. I believe this work could provide a model for disambiguating the role Taf14 plays in gene regulatory mechanisms.

CHAPTER IV

FUTURE DIRECTIONS

Summary

The work presented thus far, at its nature, reflects my desire to understand the structural organization and function of the TFIID general transcription factor and co-activator complex. In particular, when I started as a graduate student in the Weil laboratory, our knowledge of the largest TFIID subunit in yeast and second largest subunit in metazoans, Taf2, was woefully incomplete. My systematic genetic and biochemical dissection of Taf2 helped to define the interaction domains between Taf2 and Taf14 and the importance of this interaction to Taf2-TFIID association and ribosomal protein-encoding gene transcription. Yet, there is still much to learn about Taf2 and TFIID function. The findings described in this dissertation will serve as the foundation onto which multiple avenues of study could be built.

Specifically, our understanding of *TAF2* function within the TFIID complex remains poorly understood *in vivo*. I will describe additional studies that have allowed me to generate *taf2*^{ts} variants that could be used to acutely disrupt the stable association of Taf2 with the TFIID complex. The implications for proposed future studies utilizing this type of mutant variant are discussed. Alternatively, the recent structural data of Taf2 in complex with promoter DNA make very specific predictions regarding which Taf2 amino acids are critical for this function (Louder et al., 2016). Mutation of these amino acids may prove insightful regarding Taf2/TFIID DNA binding activity and its role in transcription

activation *in vivo.* In addition, the link between the Taf14 H3K9ac- or H3K9cr-modified chromatin binding activity and TFIID function remains to be defined. The identification of the *taf2-ΔC* mutant variant and Taf14-less TFIID will likely be useful for addressing the role of Taf14 chromatin binding in TFIID-dependent transcription activation *in vitro*.

Attempts at reconstituting TFIID in cells using recombinant protein overexpression, while unsuccessful, have provided an important framework for probing TFIID submodule assembly and structure. I will discuss ideas to utilize the multi-subunit Gal inducible overexpression system I have developed to define the organizing principles for TFIID assembly as well as methods to identify the existence of TFIID assembly intermediates *in vivo*.

What is the Fate of TFIID Without Taf2?

The initial site-directed mutagenesis screen I performed on Taf2 yielded a number of Ts and inviable mutant variants. While I was able to glean information regarding Taf2-Taf14 interaction through genetic analyses (Chapter 3), the fact that all of the mutant variants that displayed significant negative growth phenotypes also displayed a reduction in the ability of Taf2 to stably incorporate into the TFIID complex was disheartening. Ideally, the goal of my mutagenesis strategy was to generate a separation-of-function mutation whereby Taf2 stability in the TFIID complex is unperturbed but the biochemical function (i.e. DNA or transcription factor binding or ability to undergo a conformational change) of Taf2 and the TFIID complex is disrupted. I hypothesized that I was not able to identify any of these separation-of-function variants through my first mutagenesis screen because the majority of variants involved ala block mutations that

were quite large and potentially disruptive to local domain folding. To rectify this issue, I was intent on performing saturation 3ala (primarily) block mutagenesis, in the sense that every single residue would be mutated, on the N-terminus of Taf2 (Table 4.1). The N-terminus was chosen because a large number of the Ts or lethal variants I derived in my first screen localized to the N-terminal β-sheet motifs.

Through these analyses, I identified an additional 9 *taf2* variants that display a growth defect at the permissive temperature and are Ts at the non-permissive temperature as well as 6 *taf2* variants that are inviable (summarized in Table 4.1). Unfortunately, as I observed with my original screen, all of these variants were unable to stably associate with the TFIID complex as well as WT Taf2. In light of these results, I began contemplating what experiments one could do if s/he had a *taf2* tight temperature sensitive variant whose obvious molecular phenotype was a disruption in the ability to interact with the TFIID complex.

Answering this question requires an in-depth explanation to how I have been performing my co-IP assays, which I like to call competition coIPs. First, when I perform the coIP to test these variants, I use pseudodiploid strains expressing both an untagged WT Taf2 and the tagged Taf2 test variant. Furthermore, both versions are driven off the *ADH* promoter, which overexpresses Taf2 relative to its endogenous chromosomal locus (data not shown). Thus, the amount of both versions of Taf2 are likely in excess to what is required to associate with TFIID at normal physiological levels. In this instance, if the mutant version of Taf2 can associate with TFIID as well as the WT version of Taf2, coIP efficiencies will remain the same. However, if the mutant version does not associate with

Table 4.1 TAF2 SD mutants Round 2 - N-terminal Saturation Mutagenesis

AA Targeted	Sequence	Mutations	Growth Phenotype	CoIP Phenotype
29-31	RVA	AAA	WT Growth	Not Tested
32-34	HEK	AAA	WT Growth	Not Tested
35-37	ISL	AAA	WT Growth	Not Tested
38-40	DID	AAA	WT Growth	Not Tested
41-43	LAT	AAA	WT Growth	Not Tested
44-46	HCI	AAA	WT Growth	Not Tested
47-49	TGS	AAA	WT Growth	Not Tested
50-52	ATI	AAA	WT Growth	Not Tested
53-55	III	AAA	WT Growth	Not Tested
56-58	PLI	AAA	WT Growth	Not Tested
59-61	QNL	AAA	WT Growth	Not Tested
62-64	EYV	AAA	Slow Growth and Ts	Not Tested
65-67	TFD	AAA	Very Slow Growth and Ts	Failed to CoIP TFIID Subunits
68-70	CKE	AAA	WT Growth	Not Tested
71-73	MTI	AAA	Very Slow Growth and Ts	Failed to CoIP TFIID Subunits
74-76	KDV	AAA	WT Growth	Not Tested
77-80	LVEN	AAAA	WT Growth	Not Tested
105-108	LYSD	AAAA	WT Growth	Not Tested
109-112	NSIE	AAAA	WT Growth	Not Tested
113-116	QSHF	AAAA	WT Growth	Not Tested
117-119	LRS	AAA	WT Growth	Not Tested
121	F	A	WT Growth	Not Tested
128	P	A	WT Growth	Not Tested
144,146	IKI	AKA	WT Growth	Not Tested
154-155	LS	AA	WT Growth	Not Tested
156-176	Deletion	(GGG) insert	WT Growth	Not Tested
177-179	PIT	AAA	WT Growth	Not Tested
180-182	LQI	AAA	Inviable	Failed to CoIP TFIID Subunits
183-185	EYE	AAA	Very Slow Growth and Ts	Failed to CoIP TFIID Subunits
191-193	SGI	AAA	WT Growth	Not Tested
194-196	KFD	AAA	Inviable	Failed to CoIP TFIID Subunits
197-199	TVY	AAA	WT Growth	Not Tested
201-203	DKP	AAA	WT Growth	Not Tested
204-206	WLW	AAA	WT Growth	Not Tested
207-209	NVY	AAA	WT Growth	Not Tested
210-212	TSN	AAA	WT Growth	Not Tested
213-215	GEI	AAA	WT Growth	Not Tested
216-218	CSS	AAA	WT Growth	Not Tested
220-222	SYW	AAA	Near WT @ Permssive, Mild Ts	Not Tested

Table 4.1 continued TAF2 SD mutants Round 2 - N-terminal Saturation Mutagenesis

AA Targeted	Sequence	Mutations	Growth Phenotype	CoIP Phenotype
223-225	VPC	AAA	Near WT @ Permssive, Mild Ts	Not Tested
226-229	VDLL	AAA	Near WT @ Permssive, Mild Ts	Not Tested
232-234	KST	AAA	Near WT @ Permssive, Mild Ts	Not Tested
235-237	WEL	AAA	Inviable	Failed to CoIP TFIID Subunits
239-240	FS	AA	WT Growth	Not Tested
241-242	VP	AA	WT Growth	Not Tested
247-250	NIGT	AAAA	WT Growth	Not Tested
251-253	SKL	AAA	WT Growth	Not Tested
254-257	IGQN	AAAA	WT Growth	Not Tested
258-363	Deletion	(GGG) insert	Inviable	Failed to CoIP TFIID Subunits
364-365	KK	AA	Slow Growth and Ts	Failed to CoIP TFIID Subunits
366-368	KCI	AAA	WT Growth	Not Tested
374-376	PVA	AAA	WT Growth	Not Tested
377-379	PHH	AAA	WT Growth	Not Tested
380-382	IGW	AAA	Inviable	Failed to CoIP TFIID Subunits
383-387	AIGAF	AAAAA	Slow Growth and Ts	Failed to CoIP TFIID Subunits
393-432	Deletion	(GGG) insert	WT Growth	Not Tested
433-435	PIQ	AAA	WT Growth	Not Tested
436-438	IFT	AAA	WT Growth	Not Tested
446-448	ELT	AAA	WT Growth	Not Tested
449-450	VI	AA	WT Growth	Not Tested
451-453	NST	AAA	WT Growth	Not Tested
459-461	IID	AAA	WT Growth	Not Tested
462-464	FYS	AAA	Near WT @ Permssive, Mild Ts	Not Tested
465-467	KEF	AAA	Near WT @ Permssive, Mild Ts	Not Tested
468-470	GSY	AAA	WT Growth	Not Tested
471-473	PFT	AAA	Slow Growth and Ts	Failed to CoIP TFIID Subunits
474-476	CYS	AAA	Slow Growth and Ts	Failed to CoIP TFIID Subunits
477-479	MVF	AAA	Slow Growth and Ts	Failed to CoIP TFIID Subunits
480-482	LPT	AAA	Very Slow Growth and Ts	Failed to CoIP TFIID Subunits
483-485	APS	AAA	WT Growth	Not Tested
486-488	KHM	AAA	Near WT @ Permssive, Mild Ts	Not Tested
489-491	DFA	AAA	WT Growth	Not Tested
492-494	ALG	AAA	WT Growth	Not Tested
495-497	ICN	AAA	Inviable	Failed to CoIP TFIID Subunits
498-501	TRLL	AAAA	Near WT @ Permssive, Tight Ts	Not Tested
502-505	YPLE	AAAA	WT Growth	Not Tested
783-817	Deletion	no insert	Near WT @ Permssive, Tight Ts	Not Tested

** Note: All mutants displayed steady-state protein levels similar to WT.

the TFIID complex as well as WT Taf2, the mutant will display a reduction in coIP efficiency. This does not mean that the mutant variant cannot associate with TFIID. Indeed, in shuffled strains containing only a mutant taf2 variant that does not associate with TFIID as well as WT, I am able to observe co-precipitation of Tafs at levels greater than observed for that mutant variant in the "competition" coIP (data not shown).

Based on this rationale, it is my supposition that these taf2^{ts} variants can weakly but productively associate with TFIID at the permissive temperature, allowing them to grow at least to a certain extent. However, upon shift to the non-permissive temperature, Taf2 can no longer associate with the TFIID complex resulting in a loss of appropriate transcription regulation for the genes that require Taf2 function. In addition, in an ideal mutant, this loss of interaction between the taf2 variant and TFIID would likely manifest as an abrupt growth arrest when acutely shifted to the non-permissive growth temperature. As mentioned above, the problem with my initial random mutagnesis screen was that none of the mutants I generated could arrest growth in liquid culture following shift to the non-permissive temperature. I tested four of the mutant variants derived from my saturation mutagenesis screen of the Taf2 N-terminus for the ability to arrest following shift to the non-permissive temperature. All of these variants quickly arrested growth, including a variant with only 2ala block mutation (*taf2-ts148*; Figure 4.1).

To confirm my hypothesis that these taf2^{ts} variants dissociate from the TFIID complex following shift to the non-permissive temperature, coIP analyses of shuffled strains harboring these variants would need to be performed pre- and post-temperature shift. In lieu of these data, I believe this type of Taf2 mutant variant could be useful for two major lines of inquiry. First, upon disruption of Taf2 from the TFIID complex, what is the

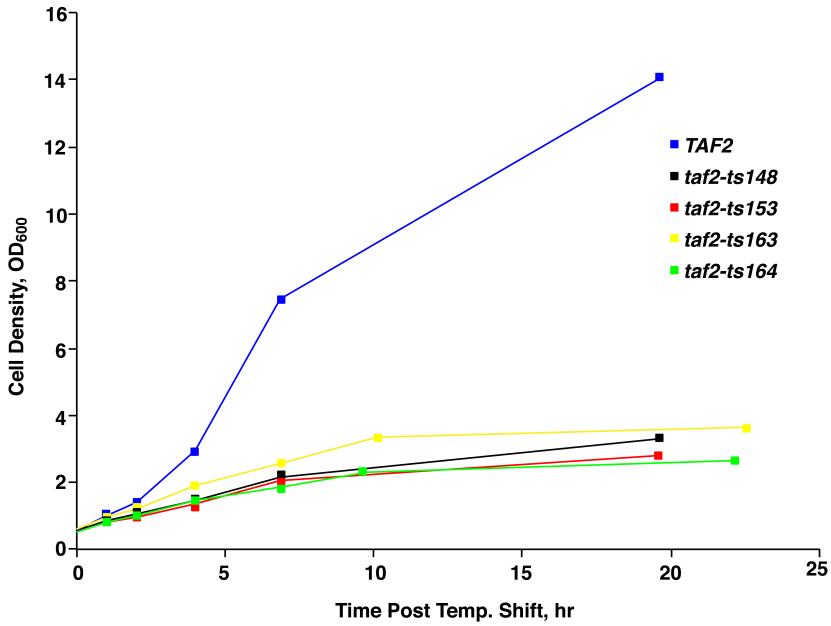


Figure 4.1: *taf2ts* Mutants Arrest Growth Upon Shift to the Non-permissive Temperature. *S. cerevisiae* harboring *taf2*^{ts} variants containing 2- or 3ala block mutations were grown at 25^{o} prior to abrupt shift to 37^{o}. Following the temperature shift, cell density (OD_{600}) was measured over time. Results are plotted above.

state of the rest of the TFIID complex? Both genetic and biochemical perturbations related to Taf1 demonstrate that a TFIID subcomplex can form containing all of the TFIID subunits sans Taf2, Taf8 and Taf14 (Papai et al., 2009; Singh et al., 2004). Would this subcomplex form in strains harboring these taf2^{ts} variants upon shift to the non-permissive temperature? If so, how would TFIID occupancy change on gene promoters? More importantly, how would gene transcription change?

With current methodologies, all of these questions could be answered in a rather straightforward manner. For example, ChIP-exo can be used to measure TFIID occupancy genome-wide by scoring Taf1 occupancy (Rhee and Pugh, 2012). In addition, ChiP-exo of Pol II can serve as a surrogate for active mRNA gene transcription. In the case that my hypothesis is not correct and shift to the non-permissive temperature does not alter Taf-TFIID association, I would employ the auxin inducible degron system to conditionally deplete Taf2 from the cell (Nishimura et al., 2009). Assuming the stability of the rest of the TFIID Tafs is not altered by targeted degradation of Taf2, the end result would likely be the

formation of a similar putative TFIID subcomplex. Ultimately, I would hope that this TFIID subcomplex could still effectively drive transcription for a subset of genes. In this case, these experiments would identify the group of genes that are reliant on Taf2 function. Similarities between these genes such as trans-activator binding sites, sequences of the core promoter or even the gene product could provide insights into the role of Taf2 in gene regulation in *S. cerevisiae*.

Dissection of the Putative Taf2 DNA Binding Domain

The structural model of TFIID bound to promoter DNA published by the Nogales lab may lead to novel insights into the function of Taf2 within TFIID (Louder et al., 2016). Using docking of endoplasmic reticulum aminopeptidase-1, a homologous structure to Taf2, Louder et al. was able to determine the structure of Taf2 to atomic resolution, including key evolutionarily conserved amino acids that are predicted to directly contact DNA. From a genetic perspective, these residues would serve as fertile ground for site-directed mutagenesis to determine if mutation of this putative DNA binding domain results in a growth defects in yeast.

While I was performing my original site-directed mutagenesis screen, I did identify one variant with mutations that correspond to this putative Taf2 DNA binding domain that also contained a growth defect, *taf2-m32* (ala block mutation of aa 775-780). Strains harboring *taf2-m32* displayed a mild slow growth phenotype at all temperatures tested. In addition, coIP analyses demonstrate that this variant can stably incorporate into the TFIID complex (Figure 4.2b). Based on my model generated in Chapter 3 (Figure 3.12),

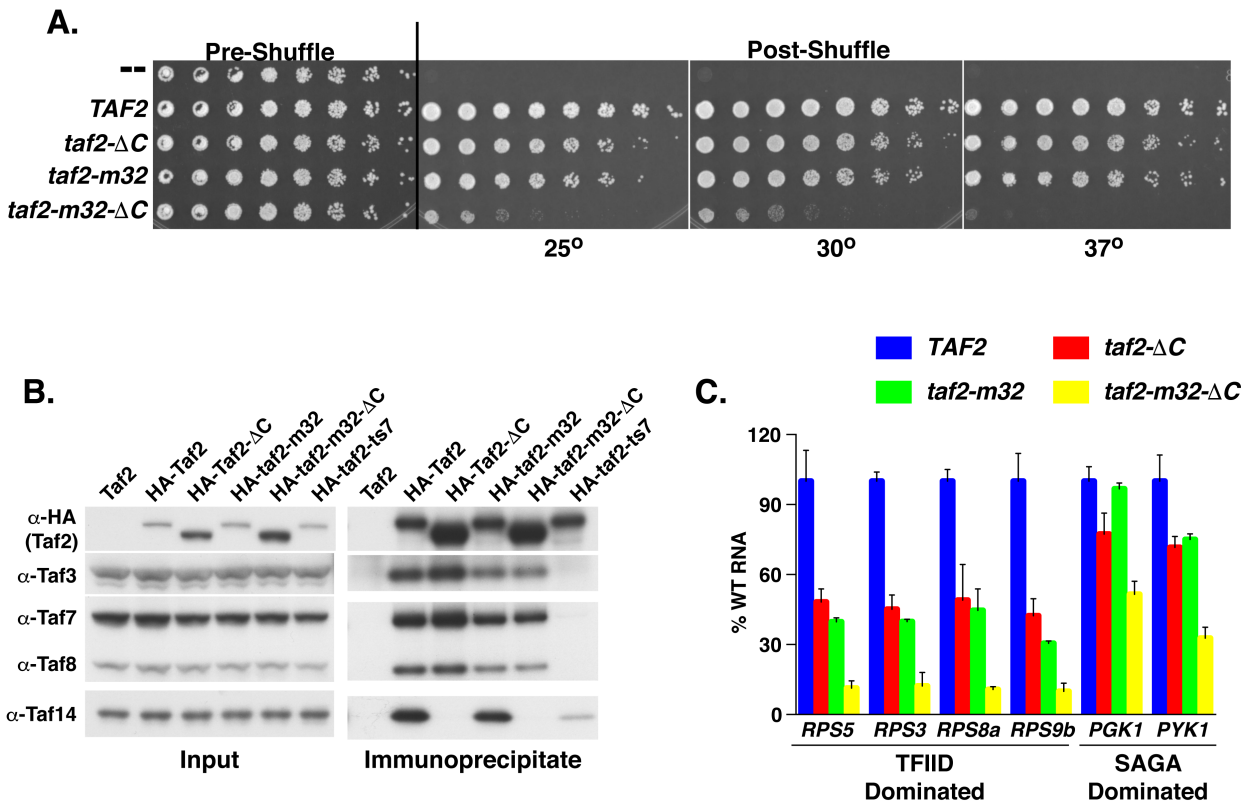


Figure 4.2: Growth, CoIP and RNA Analysis of the m32 + ΔC *taf2* Combination Mutant. ***A,B.*** Plasmid shuffle complementation and coIP analyses performed as described in Figure 3.2. ***C.*** Steady-state RNA measurements using qRT-PCR as described in Figure 3.4. Data generated from two biological replicates. Error bars represent standard deviation.

I proposed that Taf14 binding to active chromatin marks served as another contact between TFIID and gene promoters. Interestingly, cells containing Taf14-less TFIID also display a mild slow growth phenotype at all temperatures tested, although it is more apparent at lower temperatures. This led me to hypothesize that if I were to disrupt two putative contact points between TFIID and the promoter, by disrupting Taf14-TFIID stable association and by mutating the putative Taf2 DNA binding domain, then I would observe a synthetic-sick phenotype.

Indeed, when I combined the mutations in *taf2-m32* with the *taf2-ΔC* mutation (*taf2-m32-ΔC*), strains harboring this double mutant showed significantly reduced growth

at the permissive temperature and a Ts phenotype at the non-permissive temperature (Figure 4.2a). Immunoblot analysis of these strains demonstrated that taf2-m32-ΔC protein was made and stable at levels even greater than WT Taf2 but comparable to taf2-ΔC (Figure 4.2b). Importantly, this taf2-m32-ΔC variant could immunoprecipiate TFIID at levels comparable to WT Taf2 (immunoprecipitation of taf2-ts7 included as a reference for a taf2 mutant that fails to coIP TFIID efficiently). qRT-PCR analyses were performed with RNA extracted from cells grown at the permissive temperature and harboring *TAF2*, *taf2-m32*, *taf2-ΔC* or *taf2-m32-ΔC*. Consistent with data shown in Figure 3.11, RNA extracted from *taf2-ΔC* mutant cells shows a roughly 2-fold reduction in transcripts derived from the ribosomal protein-encoding genes (*RPS3*, *RPS5*, *RPS8A* and *RPS9B*) without a comparable reduction in the glycolytic genes (*PGK1* and *PYK1*). RNA extracted from *taf2-m32* mutant essentially shows the same steady-state RNA phenotype as *taf2-ΔC* mutant cells. However, *taf2-m32-ΔC* mutant cells display a ~10 fold reduction in steady-state RNA from the ribosomal protein-encoding genes. But, these cells also display a ~2-3-fold reduction in the glycolytic genes (Figure 4.3c). These data demonstrate that the mutations in *taf2-m32* do compromise Taf2 function, especially when combined with mutations that disrupt Taf14 stable association with TFIID.

It would also be informative to perform the same ChIP-exo analysis as described above to determine on which genes TFIID and Pol II occupancy is impacted. For these experiments, I think it would be imperative to use the auxin inducible degron system to accurately interpret the results. Specifically, cells would be pseudodiploid for a WT Taf2 copy that could be targeted for degradation and a second test variant of Taf2 (either WT, taf2-ΔC, taf2-m32 or taf2-m32-ΔC) that would be unaffected by inducible depletion. Under

normal growth conditions, all of the cells would grow similarly to WT because none of these mutant alleles are dominant. However, upon depletion of the WT copy, the molecular phenotype would reflect the test version of Taf2 without the complication of indirect effects associated with working with shuffled yeast strains (Sun et al., 2013, 2012). Again, these types of experiments could inform us as to which genes Taf2 function is critical for TFIID-dependent gene regulation.

While these *in vivo* molecular phenotypes are promising, the main hurdle to advancing this work is an assay that measures Taf2-DNA binding. Despite significant effort, I was only able to observe a small, inconsistent amount of Taf2-DNA binding when Taf2 was in vast excess (>100-fold) to DNA (data not shown). Alternatively, it is likely that Taf2 plays an important role in TFIID-DNA binding that I have not been able to observe with pure Taf2 alone. Therefore, in order to determine if the *taf2-m32* mutations impact DNA binding, I propose that TFIID-DNA enzymatic and chemical footprinting be performed with WT and taf2-m32 mutant forms of TFIID. The results of these footprinting analyses could be informative regarding the impact of a Taf2-DNA binding defective mutant on TFIID-DNA binding affinity or on the sequences protected from digestion. Either result would support the findings observed *in vivo* and further establish Taf2-DNA binding as a key accepted activity for TFIID-dependent transcription regulation.

The Role of Taf14 in Chromatin Transcription

The experiments detailed in Chapter III cemented the Taf2 C-terminus as a critical domain for the stable association of Taf14 with the TFIID complex. In addition, biochemical and structural studies have demonstrated that Taf14 can bind acetylated or

crotonylated histone marks (Andrews et al., 2016; Shanle et al., 2015). Together, these data led me to put forth a model (Figure 3.12) that predicts that yeast TFIID makes multiple promoter contacts through DNA and active histone marks, both of which are important for TFIID function.

To test this model, especially pertaining to the role of Taf14 in TFIID promoter binding, I would need to perform both binding studies and *in vitro* transcription with chromatinized templates. For *in vitro* binding studies, I would prepare end-biotinylated positioned mononucleosomal templates for immobilization on streptavidin beads (Gaykalova et al., 2009). This template would be designed to mimic the structure of the +1 nucleosome at a TFIID-dominated promoter. In addition, these mononucleosomes could be acetylated *in vitro* using purified SAGA and/or NuA4. I would then perform pull-down assays with WT or taf2-ΔC mutant TFIID, with or without TFIIA. If my model were correct, TFIID-template binding would be stimulated in the presence of acetylated histones. Furthermore, this stimulation would be blunted, if not eliminated in the absence of Taf14. These pull-down assays could also be supplemented with footprinting analysis to determine if stimulating binding to a nucleosomal template enhances or changes TFIID-DNA binding properties.

Similarly, I would perform *in vitro* transcription assays with purified *in vivo* assembled chromatin templates, using either of one of the methods developed by the Kornberg laboratory (Griesenbeck et al., 2003) or the Tsukiyama laboratory (Unnikrishnan et al., 2012). I would use these chromatinized templates for *in vitro* transcription with extracts (Woontner et al., 1991) that were immunodepleted for TFIID and then reconstituted with either WT TFIID or taf2-ΔC mutant TFIID (Sanders et al., 2002a). Again,

I would expect to see a reconstitution of activity with WT TFIID that would not be observed, at least to the same extent, with taf2-ΔC mutant TFIID. Similar results were observed by the Roeder lab with regard to Taf3 PHD finger-mutant TFIID and transcription from H3K4me3-modified chromatin templates (Lauberth et al., 2013).

These studies would support my model regarding the role Taf14 plays in TFIID-promoter association. In addition, these studies would help reinforce the idea that while yeast and metazoan TFIID have differences regarding promoter structure and chromatin binding domains, the general mechanisms through which TFIID engages the promoter are conserved.

Biochemical and Structural Analysis of TFIID Submodules

I spent a significant amount of time and energy during my dissertation research attempting to reconstitute TFIID using recombinant overexpression methods. While I made progress in replicating published data (Bieniossek et al., 2013), I was never able to fully reconstitute the complex. Nonetheless, I believe that my efforts have paved the way for an iterative expression and immunopurification strategy where each TFIID subunit is expressed in combination with other TFIID subunits in the hopes of forming TFIID submodules. Alternatively, I have identified a genetic reagent that could help drive formation of TFIID submodules in vivo and help to identify how yeast Tafs assemble in vivo. Using either method, the identification of these submodules could then be used to assemble the holoTFIID complex. In addition, submodules could be analyzed by cryoEM and/or X-ray crystallography for structural determination.

The multi-ORF overexpression system in *S. cerevisiae* was highly effective at expressing between one and six distinct TFIID subunits at the same time in the same cell. Using this methodology, the first complex I would attempt to reconstitute is the Taf2-Taf8-Taf10-Taf14 complex. When the Berger, Tora and Schultz labs identified the Taf2-Taf8-Taf10 complex, I had already transitioned away from reconstituting TFIID and was intent on dissecting the interactions between Taf2 and Taf14 (Trowitzsch et al., 2015). Given my knowledge of the interaction between Taf14 and Taf2, I strongly believe that I could form this Taf2-Taf8-Taf10-Taf14 complex. In addition, I hypothesize that this complex will solubilize Taf8 in a manner that heretofore I have been unable to do through my attempts to form Taf subcomplexes in *Sf9* cells. Through collaboration, the structures of these TFIID subcomplexes could be analyzed with cryoEM and computational docking methods to model the structure of these four subunits. New EM detectors have allowed structural biologists to resolve cryoEM structures to atomic resolution, revolutionizing the field of structural biology of large complexes (Callaway, 2015). Additionally, I would perform cryoEM analysis for every TFIID subcomplex that I generate, especially if I were able to iteratively build the TFIID complex one subunit or heterodimer at a time. This step-wise approach could be used to map protein interfaces, potential conformational changes upon protein-protein interactions and the locations of new proteins as they are added in each step of the TFIID complex building process.

An entirely different question is whether these TFIID subcomplexes exist in cells (other than the Taf2-Taf8-Taf10 complex). When I was attempting to reconstitute the 5-Taf TFIID core complex + the Taf8-Taf10 heterodimer, I hypothesized that extra yeast specific domains of Taf8 may be the cause of the aggregation phenotype I was observing

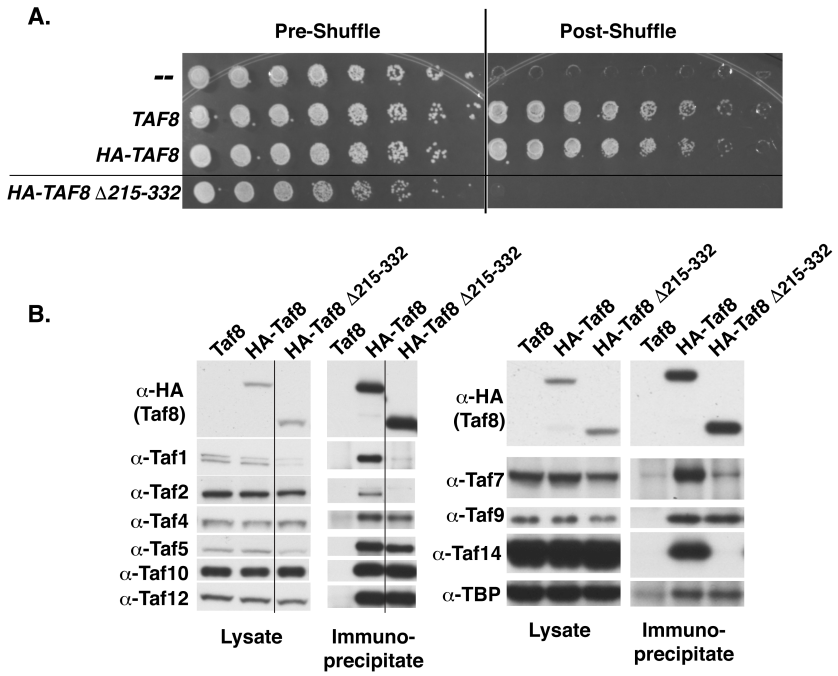


Figure 4.3: Identification of a *TAF8* Deletion Variant that Facilitate Formation of a TFIID Subcomplex *in vivo*. *A.* Plasmid shuffle analyses performed as described in Chapter III Methods and Figure 3.2a with the exception that the strain used is *taf8* null with a *URA3*-marked *TAF8* covering plasmid (YSLS58, (Gangloff et al., 2001b)). The expression plasmids encoding the test *TAF8* alleles were derived from p415ADH. HA=HA_3NLS. ***B.*** Co-IP analyses as described in Chapter III methods and in Figure 3.2b. For these experiments, either untagged or HA-tagged Taf8 and the HA-tagged Taf8 Δ215-332 variant were immunoprecipitated out of cell extracts rather than Taf2 forms.

(yeast Taf8 is 200aa larger than metazoan Taf8, see Figure 1.7). In an attempt to solubilize Taf8, I deleted some of the yeast specific *TAF8* sequence to determine if it was essential for viability. One of the variants I generated (*TAF8 Δ215-332*) displayed a mildly dominant negative phenotype (Figure 4.3a, left; compare growth to strain Pre-Shuffle). In addition, this variant could not support viability (Figure 4.3a, right). I then performed my customary steady-state immunoblotting and coIP analyses and found that this variant of Taf8 *could* efficiently coIP many of the TFIID subunits, but *failed* to coIP others (Figure 4.3b). Interestingly, the subunits Taf8 Δ215-332 could co-precipitate corresponded very well to the 5-Taf TFIID core (Taf4, Taf5, Taf6, Taf9, Taf12) with the addition of Taf10. However, the Taf8 Δ215-332 variant could not efficiently coIP Tafs 1, 2, 7 and 14.

Preliminary data suggests that Taf3, Taf11 and Taf13 also fail to coIP with this complex (data not shown). The idea that Tafs 3, 11 and 13 could not associate with this complex is in conflict with my *Sf9* cell co-expression analyses (Figure 2.4a). However, as I described above, the stoichiometry observed within the *Sf9* cell-generated 7Taf and 9Taf complexes did not reflect the stoichiometry observed for TFIID purified from *S. cerevisiae*. Thus, it is possible that association of Tafs 3, 11 and 13 with the overexpressed 5-Taf TFIID core is an artifact that does not reflect the true nature of TFIID. Indeed, through personal communication with the Berger and Schultz labs, I was informed that association of Taf3 with the 5-Taf core resulted in what they described as a dead-end complex, meaning it could not associate with any additional TFIID subunits.

Furthermore, this complex was also able to co-precipitate TBP. This result is interesting because it also conflicts with my *Sf9* cell co-expression data where TBP was not effectively precipitated when I tried to co-express every TFIID subunit at the same time (Figure 2.4b). In that TFIID co-expression experiment, I was able to see co-precipitation of the entire 5-Taf core + Taf8-10 by immunoblotting (Figure 2.4 and data not shown). While this result is perplexing, it made me consider the differences between TFIID and SAGA. In the Schultz and Berger model, the binding of Taf8 and Taf10 to the 5-Taf TFIID core results in a conformational change that would facilitate the association of the remaining TFIID subunits. What would then happen if the domains of Taf8 that were required for this conformational changes were not present or functional? Could SAGA subunits associate with the 5-Taf core in the absence of this conformational change resulting in TBP binding? At this point, this is all conjecture; however, the *TAF8 Δ215-332* mutant could be useful for exploring what the fate of the remaining TFIID subunits if the 5-Taf TFIID core cannot

undergo the Taf8-induced conformational change *in vivo* (Figure 4.4; model of TFIID assembly and theoretical submodule).

My results suggest that the subunits that Taf8 Δ215-332 cannot co-precipitate are Taf1, Taf2, Taf3, Taf7, Taf11 and Taf13. Could these subunits form a submodule in the absence of the TFIID 5-Taf core or do they just individually assemble onto 5-Taf core + Taf8-Taf10 complex? To address this question, I would engineer a strain to express two forms of Taf8: a conditional WT Taf8 variant that could be depleted using the auxin inducible degron system and the Taf8 Δ215-332 variant. Following depletion of the WT Taf8 protein, my hypothesis is that Taf8 Δ215-332 would block the conformational change of the 5-Taf TFIID core and, as a result, the rest of the TFIID subunits may be prevented from properly associating with the TFIID complex. I would then individually immunoprecipitate Tafs 1, 2, 3, 11 and 13 and immunoblot for all of these subunits to see if any form subcomplexes. Currently, we have a poor understanding of the interaction network and structure of this lobe of the TFIID complex (Taf1, Taf3, Taf11 and Taf13 all map to the A lobe of TFIID, Figure 1.6). These above proposed experiments could address this deficiency.

If immunopreciptiation of Taf1 (or any of the A lobe Tafs) were to still result in co-precipitation of the 5-Taf core in strains containing only Taf8 Δ215-332, then I would instead individually deplete every single TFIID subunit using the auxin inducible degron system (Nishimura et al., 2009). Following depletion, I would perform coIP analysis, forming immunoprecipitates using antibodies against several different TFIID subunits to determine which subcomplexes remain. This approach was employed, to a certain extent, by the Tjian laboratory where they used siRNA to individually knockdown many of the

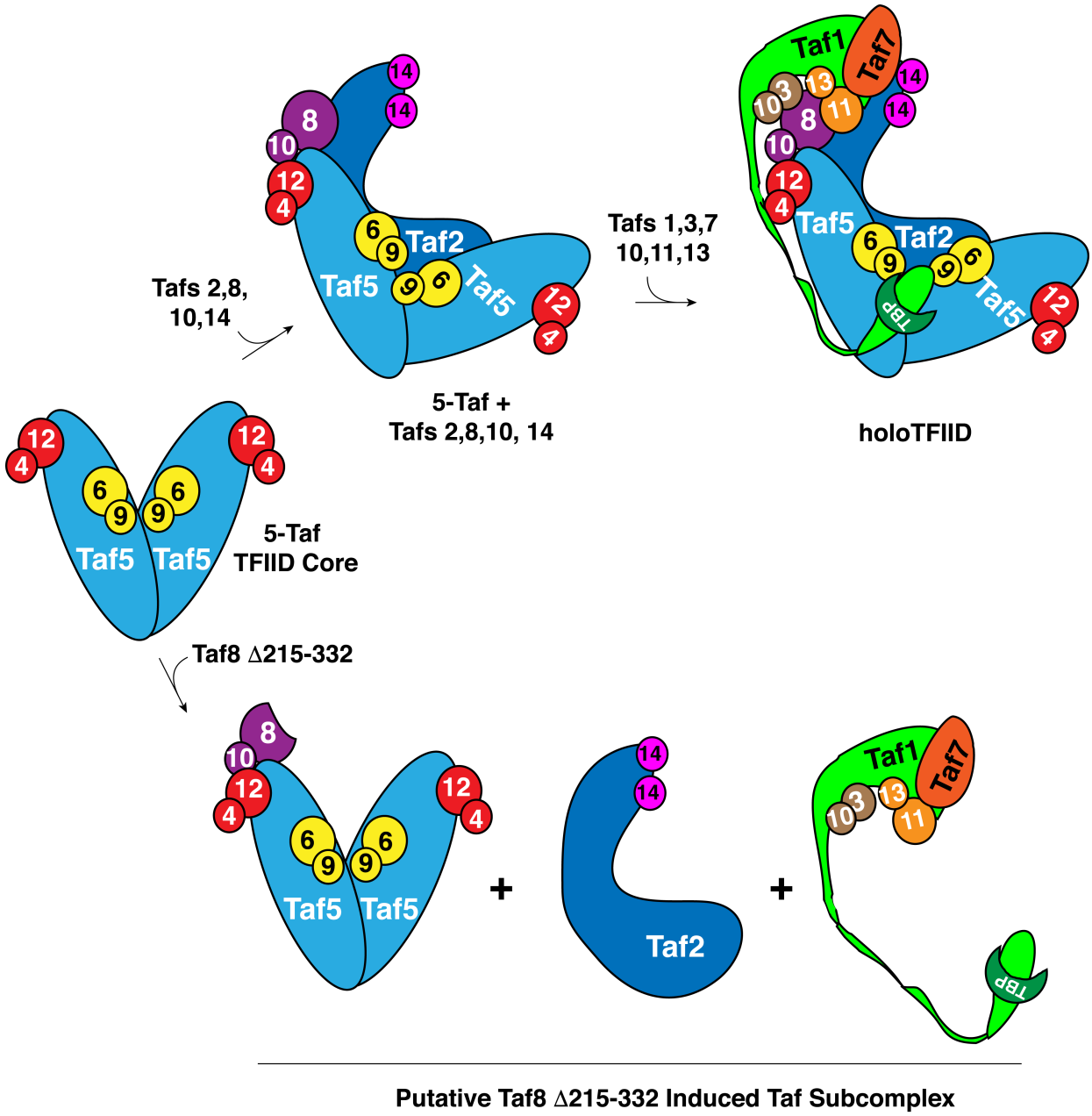


Figure 4.4: Model of TFIID Assembly Based on Submodules Revealed by Taf8 Δ215-332. Schematic of TFIID assembly starting with the 5-Taf TFIID core complex. Modular assembly proceeds through step-wise association TFIID submodules with the 5-Taf core. Binding of Tafs 2, 8, 10 and 14 to the 5-Taf core results in a conformational change that enables association of the rest of the TFIID subunits. In the presence of Taf8 Δ215-332, association of Taf2 and Taf14 are disrupted from binding to the 5-Taf core resulting in the accumulation of putative TFIID submodules.

TFIID subunits and then monitored steady-state protein levels of the other TFIID subunits (Wright et al., 2006). Through this knockdown paradigm, they were the first to suggest the existence of the 5-Taf TFIID core complex. However, these analyses were not performed on all of the TFIID subunits. In addition, following knockdown, Tjian and colleagues did not follow up by performing multiple distinct coIP reactions to identify the residual subcomplexes.

Taken together, I have described several distinct approaches that attack the problem of identifying stable subcomplexes within holoTFIID, if they exist. These subcomplexes could then be used to assemble the holoTFIID complex *in vitro*. If TFIID subcomplexes can be produced efficiently and in high abundance, this method could easily be used to make mutant forms of TFIID that would allow for in-depth biochemical and structural characterization of the TFIID complex.

REFERENCES

Adkins, M.W., Tyler, J.K. (2006) Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol Cell,* **21**, 405–416.

Albright, S.R., Tjian, R. (2000) TAFs revisited: more data reveal new twists and confirm old ideas. *Gene,* **242**, 1–13.

Allard, S., Utley, R.T., Savard, J., Clarke, A., Grant, P., Brandl, C.J., Pillus, L., Workman, J.L., Côté, J. (1999) NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J,* **18**, 5108–5119.

Allen, B.L., Taatjes, D.J. (2015) The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol,* **16**, 155–166.

Allison, L.A., Moyle, M., Shales, M., Ingles, C.J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell,* **42**, 599–610.

Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M., Ingles, C.J. (1988) The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol. Cell. Biol.* **8**, 321–329.

Almer, A., Hörz, W. (1986) Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the *PHO5/PHO3* locus in yeast. *EMBO J*, **5**, 2681–2687.

Almer, A., Rudolph, H., Hinnen, A., Hörz, W. (1986) Removal of positioned nucleosomes from the yeast *PHO5* promoter upon *PHO5* induction releases additional upstream activating DNA elements. *EMBO J*, **5**, 2689–2696.

Andrews, F.H., Shinsky, S.A., Shanle, E.K., Bridgers, J.B., Gest, A., Tsun, I.K., Krajewski, K., Shi, X., Strahl, B.D., Kutateladze, T.G. (2016) The Taf14 YEATS domain is a reader of histone crotonylation. *Nat Chem Biol* **12**, 396–398.

Arnold, J.J., Smidansky, E.D., Moustafa, I.M., Cameron, C.E. (2012) Human mitochondrial RNA polymerase: structure-function, mechanism and inhibition. *Biochim Biophys Acta*, **1819**, 948–960.

Auty, R., Steen, H., Myers, L.C., Persinger, J., Bartholomew, B., Gygi, S.P., Buratowski, S. (2004) Purification of active TFIID from Saccharomyces cerevisiae. Extensive promoter contacts and co-activator function. *J Biol Chem*, **279**, 49973–49981.

Awrey, D.E., Weilbaecher, R.G., Hemming, S.A., Orlicky, S.M., Kane, C.M., Edwards, A.M. (1997) Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem,* **272**, 14747–14754.,

Baek, H.J., Kang, Y.K., Roeder, R.G. (2006) Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly. *J Biol Chem*, **281**, 15172–15181.

Bai, Y., Perez, G.M., Beechem, J.M., Weil, P.A. (1997) Structure-function analysis of *TAF130*: identification and characterization of a high-affinity TATA-binding protein interaction domain in the N terminus of yeast TAF(II)130. *Mol Cell Biol*, **17**, 3081–3093.

Barbaric, S., Luckenbach, T., Schmid, A., Blaschke, D., Hörz, W., Korber, P. (2007) Redundancy of chromatin remodeling pathways for the induction of the yeast *PHO5* promoter *in vivo*. *J Biol Chem*, **282**, 27610–27621.

Barbaric, S., Münsterkötter, M., Goding, C., Hörz, W. (1998) Cooperative Pho2-Pho4 interactions at the *PHO5* promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. *Mol Cell Biol*, **18**, 2629–2639.

Barbarić, S., Münsterkötter, M., Svaren, J., Hörz, W. (1996) The homeodomain protein Pho2 and the basic-helix-loop-helix protein Pho4 bind DNA cooperatively at the yeast *PHO5* promoter. *Nucleic Acids Res*, **24**, 4479–4486.

Barbaric, S., Reinke, H., Hörz, W. (2003) Multiple mechanistically distinct functions of SAGA at the *PHO5* promoter. *Mol Cell Biol*, **23**, 3468–3476.

Barberis, A., Müller, C.W., Harrison, S.C., Ptashne, M. (1993) Delineation of two functional regions of transcription factor TFIIB. *Proc Natl Acad Sci U S A*, **90**, 5628–5632.

Barillà, D., Lee, B.A., Proudfoot, N.J. (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **98**, 445–450.

Basehoar, A.D., Zanton, S.J., Pugh, B.F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, **116**, 699–709.

Baumli, S., Hoeppner, S., Cramer, P. (2005) A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *J Biol Chem*, **280**, 18171–18178.

Bergman, L.W., Kramer, R.A. (1983) Modulation of chromatin structure associated with derepression of the acid phosphatase gene of *Saccharomyces cerevisiae*. *J Biol Chem*, **258**, 7223–7227.

Bergman, L.W., Stranathan, M.C., Preis, L.H. (1986) Structure of the transcriptionally repressed phosphate-repressible acid phosphatase gene (*PHO5*) of *Saccharomyces cerevisiae*. *Mol Cell Biol*, **6**, 38–46.

Bhattacharya, S., Lou, X., Hwang, P., Rajashankar, K.R., Wang, X., Gustafsson, J.-Å., Fletterick, R.J., Jacobson, R.H., Webb, P. (2014) Structural and functional insight into TAF1-TAF7, a subcomplex of transcription factor II D. *Proc Natl Acad Sci U S A*, **111**, 9103–9108.

Bhaumik, S.R., Raha, T., Aiello, D.P., Green, M.R. (2004) *In vivo* target of a transcriptional activator revealed by fluorescence resonance energy transfer. *Genes Dev*, **18**, 333–343.

Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O., Fleharty, B.E., Paulson, A., Allali-Hassani, A., Zhou, J.-Q., Mer, G., Grant, P.A., Workman, J.L., Zang, J., Min, J. (2011) Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J*, **30**, 2829–2842.

Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., Berger, I. (2013) The architecture of human general transcription factor TFIID core complex. *Nature,* **493**, 699–702.

Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A.C., Davidson, I., Moras, D. (1998) Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the *SPT3* family. *Cell*, **94**, 239–249.

Boeke, J.D., Trueheart, J., Natsoulis, G., Fink, G.R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol*, **154**, 164–175.

Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., Devys, D. (2014) The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev*, **28**, 1999–2012.

Borggrefe, T., Davis, R., Bareket-Samish, A., Kornberg, R.D., 2001. Quantitation of the RNA polymerase II transcription machinery in yeast. *J Biol Chem*, **276**, 47150–47153.

Börner, T., Aleynikova, A.Y., Zubo, Y.O., Kusnetsov, V.V. (2015) Chloroplast RNA polymerases: Role in chloroplast biogenesis. *Biochim Biophys Acta*, **1847**, 761–769.

Both, G.W., Furuichi, Y., Muthukrishnan, S., Shatkin, A.J. (1975) Ribosome binding to reovirus mRNA in protein synthesis requires 5' terminal 7-methylguanosine. *Cell*, **6**, 185–195.

Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., Boeke, J.D. (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, **14**, 115–132.

Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., Allis, C.D. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev*, **15**, 3286–3295.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., Allis, C.D. (1996) *Tetrahymena* histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, **84**, 843–851.

Bryant, G.O., Ptashne, M. (2003) Independent recruitment *in vivo* by Gal4 of two complexes required for transcription. *Mol Cell*, **11**, 1301–1309.

Buratowski, S., Hahn, S., Guarente, L., Sharp, P.A. (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell*, **56**, 549–561.

Buratowski, S., Sopta, M., Greenblatt, J., Sharp, P.A. (1991) RNA polymerase II-associated proteins are required for a DNA conformation change in the transcription initiation complex. *Proc Natl Acad Sci U S A*, **88**, 7509–7513.

Buratowski, S., Zhou, H. (1993) Functional domains of transcription factor TFIIB. Proc Natl Acad Sci U S A, **90**, 5633–5637.

Burgers, P.M. (1999) Overexpression of multisubunit replication factors in yeast. *Methods*, **18**, 349–355.

Burgess, R.R., Travers, A.A., Dunn, J.J., Bautz, E.K. (1969) Factor stimulating transcription by RNA polymerase. *Nature*, **221**, 43–46.

Burke, T.W., Kadonaga, J.T. (1997) The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes Dev*, **11**, 3020–3031.

Cairns, B.R., Henry, N.L., Kornberg, R.D. (1996) TFG/TAF30/ANC1, a component of the yeast SWI/SNF complex that is similar to the leukemogenic proteins ENL and AF-9. *Mol Cell Biol*, **16**, 3308–3316.

Callaway, E. (2015) The revolution will not be crystallized: a new method sweeps through structural biology. *Nature,* **525**, 172–174.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P., Workman, J.L. (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, **123**, 581–592.

Carter, R., Drouin, G. (2010) The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. *Mol Biol Evol*, **27**, 1035–1043.

Chalkley, G.E., Verrijzer, C.P. (1999) DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J*, **18**, 4835–4845.

Chen, J.L., Attardi, L.D., Verrijzer, C.P., Yokomori, K., Tjian, R. (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, **79**, 93–105.

Cheng, J.X., Floer, M., Ononaji, P., Bryant, G., Ptashne, M. (2002) Responses of four yeast genes to changes in the transcriptional machinery are determined by their promoters. **Curr Biol**, **12**, 1828–1832.

Cho, E.J., Takagi, T., Moore, C.R., Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, **11**, 3319–3326.

Choder, M., Bratosin, S., Aloni, Y. (1984) A direct analysis of transcribed minichromosomes: all transcribed SV40 minichromosomes have a nuclease-hypersensitive region within a nucleosome-free domain. *EMBO J*, **3**, 2929–2936.

Christianson, T.W., Sikorski, R.S., Dante, M., Shero, J.H., Hieter, P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.

Cianfrocco, M.A., Kassavetis, G.A., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J.T., Nogales, E. (2013) Human TFIID binds to core promoter DNA in a reorganized structural state. *Cell*, **152**, 120–131.

Cieniewicz, A.M., Moreland, L., Ringel, A.E., Mackintosh, S.G., Raman, A., Gilbert, T.M., Wolberger, C., Tackett, A.J., Taverna, S.D. (2014) The bromodomain of Gcn5 regulates site specificity of lysine acetylation on histone H3. *Mol Cell Proteomics*, **13**, 2896–2910.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., Zaret, K.S. (2002) Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*, **9**, 279–289.

Cirillo, L.A., McPherson, C.E., Bossard, P., Stevens, K., Cherian, S., Shim, E.Y., Clark, K.L., Burley, S.K., Zaret, K.S. (1998) Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J*, **17**, 244–254.

Conaway, J.W., Shilatifard, A., Dvir, A., Conaway, R.C. (2000) Control of elongation by RNA polymerase II. *Trends Biochem Sci*, **25**, 375–380.

Conaway, R.C., Conaway, J.W. (2013) The Mediator complex and transcription elongation. *Biochim Biophys Acta*, **1829**, 69–75.

Corden, J.L., Cadena, D.L., Ahearn, J.M., Dahmus, M.E. (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A*, **82**, 7934–7938.

Cramer, P., Bushnell, D.A., Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**, 1863–1876.

Crick, F. (1970) Central dogma of molecular biology. *Nature*, **227**, 561–563.

Dabrowski, S., Kur, J. (1998) Cloning and expression in *Escherichia coli* of the recombinant his-tagged DNA polymerases from *Pyrococcus furiosus* and *Pyrococcus woesei*. *Protein Expr Purif*, **14**, 131–138.

Dantonel, J.C., Murthy, K.G., Manley, J.L., Tora, L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature,* **389**, 399–402.

De Kloet, S.R., Andrean, B.A. (1976) Methylated nucleosides in polyadenylate-containing yeast messenger ribonucleic acid. *Biochim Biophys Acta*, **425**, 401–408.

Deato, M.D.E., Marr, M.T., Sottero, T., Inouye, C., Hu, P., Tjian, R. (2008) MyoD targets TAF3/TRF3 to activate myogenin transcription. *Mol Cell*, **32**, 96–105.

Deng, W., Roberts, S.G.E. (2006) Core promoter elements recognized by transcription factor IIB. *Biochem Soc Trans*, **34**, 1051–1053.

Dikstein, R., Ruppert, S., Tjian, R. (1996) $TAF_{II}250$ is a bipartite protein kinase that phosphorylates the base transcription factor RAP74. *Cell*, **84**, 781–790.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., Shilatifard, A., (2002) Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem, **277**, 28368–28371.

Dreyfus, M., Régnier, P. (2002) The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell*, **111**, 611–613.

Dudley, A.M., Rougeulle, C., Winston, F. (1999) The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step in vivo. *Genes Dev*, **13**, 2940–2945.

Duffy, J.J., Geiduschek, E.P. (1977) Purification of a positive regulatory subunit from phage *SP01*-modified RNA polymerase. *Nature*, **270**, 28–32.

Durso, R.J., Fisher, A.K., Albright-Frey, T.J., Reese, J.C. (2001) Analysis of TAF90 mutants displaying allele-specific and broad defects in transcription. *Mol Cell Biol*, **21**, 7331–7344.

Edwards, A.M., Darst, S.A., Feaver, W.J., Thompson, N.E., Burgess, R.R., Kornberg, R.D. (1990) Purification and lipid-layer crystallization of yeast RNA polymerase II. *Proc Natl Acad Sci U S A*, **87**, 2122–2126.

Edwards, A.M., Kane, C.M., Young, R.A., Kornberg, R.D. (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J Biol Chem*, **266**, 71–75.

Feaver, W.J., Henry, N.L., Bushnell, D.A., Sayre, M.H., Brickner, J.H., Gileadi, O., Kornberg, R.D. (1994) Yeast TFIIE. Cloning, expression, and homology to vertebrate proteins. *J Biol Chem*, **269**, 27549–27553.

Feaver, W.J., Henry, N.L., Wang, Z., Wu, X., Svejstrup, J.Q., Bushnell, D.A., Friedberg, E.C., Kornberg, R.D. (1997) Genes for *Tfb2*, *Tfb3*, and *Tfb4* subunits of yeast transcription/repair factor IIH. Homology to human cyclin-dependent kinase activating kinase and IIH subunits. *J Biol Chem*, **272**, 19319–19327.

Fishburn, J., Mohibullah, N., Hahn, S. (2005) Function of a eukaryotic transcription activator during the transcription cycle. *Mol Cell*, **18**, 369–378.

Fisher, R., Blumenthal, T. (1980) Analysis of RNA polymerase by trypsin cleavage. Evidence for a specific association between subunits sigma and beta involved in the closed to open complex transition. *J Biol Chem*, **255**, 11056–11062.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., Reinberg, D. (1991) The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. *Proc Natl Acad Sci U S A*, **88**, 9999–10003.

Flores, O., Lu, H., Reinberg, D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J Biol Chem*, **267**, 2786–2793.

Flores, O., Maldonado, E., Reinberg, D. (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J Biol Chem*, **264**, 8913–8921.

Forget, D., Langelier, M.-F., Thérien, C., Trinh, V., Coulombe, B., 2004. Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms. *Mol Cell Biol,* **24**, 1122–1131.

Fox, T.D., Losick, R., Pero, J. (1976) Regulatory gene 28 of bacteriophage *SPO1* codes for a phage-induced subunit of RNA polymerase. *J Mol Biol*, **101**, 427–433.

Fukuda, A., Yamauchi, J., Wu, S.Y., Chiang, C.M., Muramatsu, M., Hisatake, K. (2001) Reconstitution of recombinant TFIIH that can mediate activator-dependent transcription. *Genes Cells Devoted Mol Cell Mech*, **6**, 707–719.

Gangloff, Y.G., Pointud, J.C., Thuault, S., Carré, L., Romier, C., Muratoglu, S., Brand, M., Tora, L., Couderc, J.L., Davidson, I. (2001a). The TFIID components human TAF(II)140 and Drosophila BIP2 (TAF(II)155) are novel metazoan homologues of yeast TAF(II)47 containing a histone fold and a PHD finger. *Mol Cell Biol*, **21**, 5109–5121.

Gangloff, Y.G., Sanders, S.L., Romier, C., Kirschner, D., Weil, P.A., Tora, L., Davidson, I. (2001b) Histone folds mediate selective heterodimerization of yeast TAF(II)25 with TFIID components yTAF(II)47 and yTAF(II)65 and with SAGA component ySPT7. *Mol Cell Biol*, **21**, 1841–1853.

Garbett, K.A., Tripathi, M.K., Cencki, B., Layer, J.H., Weil, P.A. (2007) Yeast TFIID serves as a coactivator for Rap1p by direct protein-protein interaction. Mol Cell Biol, **27**, 297–311.

Gary, S.L., Burgers, M.J. (1995) Identification of the fifth subunit of *Saccharomyces cerevisiae* replication factor C. *Nucleic Acids Res*, **23**, 4986–4991.

Gaykalova, D.A., Kulaeva, O.I., Bondarenko, V.A., Studitsky, V.M. (2009) Preparation and analysis of uniquely positioned mononucleosomes. *Methods Mol Biol*, 523, 109–123.

Gazit, K., Moshonov, S., Elfakess, R., Sharon, M., Mengus, G., Davidson, I., Dikstein, R. (2009) TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes. *J Biol Chem*, **284**, 26286–26296.

Geiger, S.R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A.J.R., Cramer, P., (2010) RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. *Mol Cell*, **39**, 583–594.

Gerik, K.J., Gary, S.L., Burgers, P.M. (1997) Overproduction and affinity purification of *Saccharomyces cerevisiae* replication factor C. *J Biol Chem*, **272**, 1256–1262.

Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., Ponticelli, A.S. (2004) Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol Cell Biol*, **24**, 10975–10985.

Gietz, R.D., Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc*, **2**, 31–34.

Gileadi, O., Feaver, W.J., Kornberg, R.D. (1992) Cloning of a subunit of yeast RNA polymerase II transcription factor b and CTD kinase. *Science*, **257**, 1389–1392.

Gill, G., Pascal, E., Tseng, Z.H., Tjian, R. (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the $dTAF_{II}110$ component of the Drosophila TFIID complex and mediates transcriptional activation. *Proc Natl Acad Sci U S A*, **91**, 192–196.

Goldstein, A.L., McCusker, J.H. (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast*, **15**, 1541–1553.

Golub, E.I. (1988) “One minute” transformation of competent *E. coli* by plasmid DNA. *Nucleic Acids Res*, **16**, 1641.

Goodrich, J.A., Hoey, T., Thut, C.J., Admon, A., Tjian, R. (1993) *Drosophila* $TAF_{II}40$ interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell*, **75**, 519–530.

Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., Berger, S.L., Workman, J.L. (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev*, **11**, 1640–1650.

Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., Workman, J.L. (1999) Expanded lysine acetylation specificity of Gcn5 in native complexes. *J Biol Chem*, **274**, 5895–5900.

Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates, J.R., Workman, J.L., (1998) A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell*, **94**, 45–53.

Gregory, P.D., Schmid, A., Zavari, M., Lui, L., Berger, S.L., Hörz, W. (1998) Absence of Gcn5 HAT activity defines a novel state in the opening of chromatin at the *PHO5* promoter in yeast. *Mol Cell*, **1**, 495–505.

Gregory, P.D., Schmid, A., Zavari, M., Münsterkötter, M., Hörz, W. (1999) Chromatin remodelling at the PHO8 promoter requires SWI-SNF and SAGA at a step subsequent to activator binding. *EMBO J*, **18**, 6407–6414.

Griesenbeck, J., Boeger, H., Strattan, J.S., Kornberg, R.D. (2003) Affinity purification of specific chromatin segments from chromosomal loci in yeast. *Mol Cell Biol*, **23**, 9275–9282.

Grünberg, S., Henikoff, S., Hahn, S., Zentner, G.E. (2016) Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. *EMBO J*, doi:10.15252/embj.201695020

Gu, B., Eick, D., Bensaude, O., 2013. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II *in vivo*. *Nucleic Acids Res*, **41**, 1591–1603.

Guglielmi, B., Soutourina, J., Esnault, C., Werner, M. (2007) TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc Natl Acad Sci U S A*, **104**, 16062–16067.

Gulyas, K.D., Donahue, T.F. (1992) *SSL2*, a suppressor of a stem-loop mutation in the *HIS4* leader encodes the yeast homolog of human ERCC-3. *Cell,* **69**, 1031–1042.

Guzder, S.N., Sung, P., Bailly, V., Prakash, L., Prakash, S. (1994) RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature*, **369**, 578–581.

Haag, J.R., Pikaard, C.S. (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nat Rev Mol Cell Biol*, **12**, 483–492.

Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol*, **11**, 394–403.

Hahn, S., Young, E.T. (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics*, **189**, 705–736.

Haldenwang, W.G., Lang, N., Losick, R. (1981) A sporulation-induced sigma-like regulatory protein from *B. subtilis*. *Cell*, **23**, 615–624.

Haldenwang, W.G., Losick, R. (1980) Novel RNA polymerase sigma factor from *Bacillus subtilis*. *Proc Natl Acad Sci U S A*, **77**, 7000–7004.

Haldenwang, W.G., Losick, R. (1979) A modified RNA polymerase transcribes a cloned gene under sporulation control in *Bacillus subtilis*. *Nature*, **282**, 256–260.

Hanes, S.D. (2014) The Ess1 prolyl isomerase: traffic cop of the RNA polymerase II transcription cycle. *Biochim Biophys Acta*, **1839**, 316–333.

Haruki, H., Nishikawa, J., Laemmli, U.K. (2008) The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol Cell*, **31**, 925–932.

He, Y., Yan, C., Fang, J., Inouye, C., Tjian, R., Ivanov, I., Nogales, E. (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature*, **533**, 359–365.

Henry, N.L., Campbell, A.M., Feaver, W.J., Poon, D., Weil, P.A., Kornberg, R.D. (1994) TFIIF-TAF-RNA polymerase II connection. *Genes Dev*, **8**, 2868–2878.

Henry, N.L., Sayre, M.H., Kornberg, R.D. (1992) Purification and characterization of yeast RNA polymerase II general initiation factor g. *J Biol Chem*, **267**, 23388–23392.

Heyduk, T., Lee, J.C., Ebright, Y.W., Blatter, E.E., Zhou, Y., Ebright, R.H. (1993) CAP interacts with RNA polymerase in solution in the absence of promoter DNA. *Nature*, **364**, 548–549.

Higgins, D.R., Prakash, S., Reynolds, P., Polakowska, R., Weber, S., Prakash, L. (1983) Isolation and characterization of the *RAD3* gene of *Saccharomyces cerevisiae* and inviability of *rad3* deletion mutants. *Proc Natl Acad Sci U S A*, **80**, 5680–5684.

Hiller, M., Chen, X., Pringle, M.J., Suchorolski, M., Sancak, Y., Viswanathan, S., Bolival, B., Lin, T.-Y., Marino, S., Fuller, M.T. (2004) Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Dev*, **131**, 5297–5308.

Hisatake, K., Ohta, T., Takada, R., Guermah, M., Horikoshi, M., Nakatani, Y., Roeder, R.G. (1995) Evolutionary conservation of human TATA-binding-polypeptide-associated factors $TAF_{II}31$ and $TAF_{II}80$ and interactions of $TAF_{II}80$ with other TAFs and with general transcription factors. *Proc Natl Acad Sci U S A*, **92**, 8195–8199.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.

Hoey, T., Weinzierl, R.O., Gill, G., Chen, J.L., Dynlacht, B.D., Tjian, R. (1993) Molecular cloning and functional analysis of *Drosophila TAF110* reveal properties expected of coactivators. *Cell*, **72**, 247–260.

Holstege, F.C., van der Vliet, P.C., Timmers, H.T. (1996) Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J*, **15**, 1666–1677.

Hottiger, M.O. (2015) Nuclear ADP-Ribosylation and Its Role in Chromatin Plasticity, Cell Differentiation, and Epigenetics. *Annu Rev Biochem*, **84**, 227–263.

Hu, G., Schones, D.E., Cui, K., Ybarra, R., Northrup, D., Tang, Q., Gattinoni, L., Restifo, N.P., Huang, S., Zhao, K. (2011) Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. *Genome Res*, **21**, 1650–1658.

Huisinga, K.L., Pugh, B.F. (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol Cell*, **13**, 573–585.

Hwang, W.W., Venkatasubrahmanyam, S., Ianculescu, A.G., Tong, A., Boone, C., Madhani, H.D. (2003) A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell*, **11**, 261–266.

Imasaki, T., Calero, G., Cai, G., Tsai, K.-L., Yamada, K., Cardelli, F., Erdjument-Bromage, H., Tempst, P., Berger, I., Kornberg, G.L., Asturias, F.J., Kornberg, R.D., Takagi, Y. (2011) Architecture of the Mediator head module. *Nature*, **475**, 240–243.

Imbalzano, A.N., Zaret, K.S., Kingston, R.E. (1994) Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J Biol Chem*, **269**, 8280–8286.

Inostroza, J., Flores, O., Reinberg, D. (1991) Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. *J Biol Chem*, **266**, 9304–9308.

Inoue, H., Nojima, H., Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23–28.

International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931–945.

Jacobson, R.H., Ladurner, A.G., King, D.S., Tjian, R. (2000) Structure and function of a human $TAF_{II}250$ double bromodomain module. *Science*, **288**, 1422–1425.

Jakobovits, E.B., Bratosin, S., Aloni, Y. (1980) A nucleosome-free region in SV40 minichromosomes. *Nature*, **285**, 263–265.

John, S., Howe, L., Tafrov, S.T., Grant, P.A., Sternglanz, R., Workman, J.L. (2000) The something about silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAF(II)30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)-FACT complex. *Genes Dev*, **14**, 1196–1208.

Johnson, K.M., Carey, M. (2003) Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator. *Curr Biol*, **13**, 772–777.

Johnson, K.M., Wang, J., Smallwood, A., Arayata, C., Carey, M. (2002) TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes Dev*, **16**, 1852–1863.

Juven-Gershon, T., Cheng, S., Kadonaga, J.T. (2006) Rational design of a super core promoter that enhances gene expression. *Nat Methods*, **3**, 917–922.

Kabani, M., Michot, K., Boschiero, C., Werner, M. (2005) Anc1 interacts with the catalytic subunits of the general transcription factors TFIID and TFIIF, the chromatin remodeling complexes RSC and INO80, and the histone acetyltransferase complex NuA3. *Biochem Biophys Res Commun*, **332**, 398–403.

Kadonaga, J.T. (2012) Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol*, **1**, 40–51.

Kaffman, A., Herskowitz, I., Tjian, R., O'Shea, E.K. (1994) Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80-PHO85. *Science,* **263**, 1153–1156.

Kaffman, A., Rank, N.M., O'Neill, E.M., Huang, L.S., O'Shea, E.K. (1998) The receptor Msn5 exports the phosphorylated transcription factor Pho4 out of the nucleus. *Nature*, **396**, 482–486.

Kamenova, I., Warfield, L., Hahn, S. (2014) Mutations on the DNA binding surface of TBP discriminate between yeast TATA and TATA-less gene transcription. *Mol Cell Biol*, **34**, 2929–2943.

Kaufmann, J., Ahrens, K., Koop, R., Smale, S.T., Müller, R. (1998) CIF150, a human cofactor for transcription factor IID-dependent initiator function. *Mol Cell Biol*, **18**, 233–239.

Kaufmann, J., Smale, S.T. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev*, **8**, 821–829.

Kaufmann, J., Verrijzer, C.P., Shao, J., Smale, S.T. (1996) CIF, an essential cofactor for TFIID-dependent initiator function. *Genes Dev*, **10**, 873–886.

Kelleher, R.J., Flanagan, P.M., Kornberg, R.D. (1990) A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell*, **61**, 1209–1215.

Kelly, W.G., Dahmus, M.E., Hart, G.W. (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J Biol Chem*, **268**, 10416–10424.

Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J., Boone, C., Emili, A., Weissman, J.S., Hughes, T.R., Strahl, B.D., Grunstein, M., Greenblatt, J.F., Buratowski, S., Krogan, N.J. (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*, **123**, 593–605.

Keogh, M.-C., Podolny, V., Buratowski, S. (2003) Bur1 kinase is required for efficient transcription elongation by RNA polymerase II. *Mol Cell Biol*, **23**, 7005–7018.

Killeen, M., Coulombe, B., Greenblatt, J. (1992) Recombinant TBP, transcription factor IIB, and RAP30 are sufficient for promoter recognition by mammalian RNA polymerase II. *J Biol Chem*, **267**, 9463–9466.

Kim, T.K., Roeder, R.G., 1994. Involvement of the basic repeat domain of TATA-binding protein (TBP) in transcription by RNA polymerases I, II, and III. J. Biol. Chem. 269, 4891–4894.

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., Kornberg, R.D. (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, **77**, 599–608.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., Strahl, B.D. (2005) A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol*, **25**, 3305–3316.

Klebanow, E.R., Poon, D., Zhou, S., Weil, P.A. (1997) Cloning and characterization of an essential *Saccharomyces cerevisiae* gene, *TAF40*, which encodes yTAF$_{II}$40, an RNA polymerase II-specific TATA-binding protein-associated factor. *J Biol Chem*, **272**, 9436–9442.

Klebanow, E.R., Poon, D., Zhou, S., Weil, P.A. (1996) Isolation and characterization of *TAF25*, an essential yeast gene that encodes an RNA polymerase II-specific TATA-binding protein-associated factor. *J Biol Chem*, **271**, 13706–13715.

Knutson, B.A., Hahn, S. (2011) Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Mol Cell Biol*, **31**, 818–831.

Kobayashi, N., Boyer, T.G., Berk, A.J. (1995) A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol Cell Biol*, **15**, 6465–6473.

Kokubo, T., Swanson, M.J., Nishikawa, J.I., Hinnebusch, A.G., Nakatani, Y. (1998) The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein. *Mol Cell Biol*, **18**, 1003–1012.

Koleske, A.J., Young, R.A. (1994) An RNA polymerase II holoenzyme responsive to activators. *Nature*, **368**, 466–469.

Kolodziej, P.A., Woychik, N., Liao, S.M., Young, R.A. (1990) RNA polymerase II subunit composition, stoichiometry, and phosphorylation. *Mol Cell Biol*, **10**, 1915–1920.

Komarnitsky, P., Cho, E.J., Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, **14**, 2452–2460.

Korber, P., Barbaric, S. (2014) The yeast *PHO5* promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Res*, **42**, 10888–10902.

Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868–871.

Koschubs, T., Lorenzen, K., Baumli, S., Sandström, S., Heck, A.J.R., Cramer, P. (2010) Preparation and topology of the Mediator middle module. *Nucleic Acids Res*, **38**, 3186–3195.

Koschubs, T., Seizl, M., Larivière, L., Kurth, F., Baumli, S., Martin, D.E., Cramer, P. (2009) Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *EMBO J*, **28**, 69–80.

Kotani, T., Miyake, T., Tsukihashi, Y., Hinnebusch, A.G., Nakatani, Y., Kawaichi, M., Kokubo, T. (1998) Identification of highly conserved amino-terminal segments of dTAFII230 and yTAFII145 that are functionally interchangeable for inhibiting TBP-DNA interactions *in vitro* and in promoting yeast cell growth *in vivo*. *J Biol Chem*, **273**, 32254–32264.

Kremer, S.B., Kim, S., Jeon, J.O., Moustafa, Y.W., Chen, A., Zhao, J., Gross, D.S. (2012) Role of Mediator in regulating Pol II elongation and nucleosome displacement in *Saccharomyces cerevisiae*. *Genetics*, **191**, 95–106.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., Korber, P. (2016) Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*, **167**, 709–721.

Krogan, N.J., Dover, J., Khorrami, S., Greenblatt, J.F., Schneider, J., Johnston, M., Shilatifard, A., 2002. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem*, **277**, 10753–10755.

Krogan, N.J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O.W., Ding, H., Haw, R.A., Pootoolal, J., Tong, A., Canadien, V., Richards, D.P., Wu, X., Emili, A., Hughes, T.R., Buratowski, S., Greenblatt, J.F. (2003a) A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. *Mol Cell*, **12**, 1565–1576.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., Greenblatt, J. (2003b) Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol*, **23**, 4207–4218.

Kuhn, C.-D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., Cramer, P. (2007) Functional architecture of RNA polymerase I. *Cell*, **131**, 1260–1272.

Kushnirov, V.V. (2000) Rapid and reliable protein extraction from yeast. *Yeast*, **16**, 857–860.

Ladurner, A.G., Inouye, C., Jain, R., Tjian, R. (2003) Bromodomains mediate an acetyl-histone encoded antisilencing function at heterochromatin boundaries. *Mol Cell*, **11**, 365–376.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., Ebright, R.H. (1998) New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev*, **12**, 34–44.

Larivière, L., Geiger, S., Hoeppner, S., Röther, S., Strässer, K., Cramer, P. (2006) Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nat Struct Mol Biol*, *13*, 895–901.

Larivière, L., Plaschka, C., Seizl, M., Wenzeck, L., Kurth, F., Cramer, P. (2012) Structure of the Mediator head module. *Nature*, **492**, 448–451.

Larivière, L., Seizl, M., van Wageningen, S., Röther, S., van de Pasch, L., Feldmann, H., Strässer, K., Hahn, S., Holstege, F.C.P., Cramer, P. (2008) Structure-system correlation identifies a gene regulatory Mediator submodule. *Genes Dev*, **22**, 872–877.

Larschan, E., Winston, F. (2005) The S*accharomyces cerevisiae* Srb8-Srb11 complex functions with the SAGA complex during Gal4-activated transcription. *Mol Cell Biol*, **25**, 114–123.

Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., Roeder, R.G. (2013) H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell*, **152**, 1021–1036.

Lawit, S.J., O'Grady, K., Gurley, W.B., Czarnecka-Verner, E. (2007) Yeast two-hybrid map of *Arabidopsis* TFIID. *Plant Mol Biol*, **64**, 73–87.

Layer, J.H., Miller, S.G., Weil, P.A. (2010) Direct transactivator-transcription factor IID (TFIID) contacts drive yeast ribosomal protein gene transcription. *J Biol Chem*, **285**, 15489–15499.

Layer, J.H., Weil, P.A. (2013) Direct TFIIA-TFIID protein contacts drive budding yeast ribosomal protein gene transcription. *J Biol Chem*, **288**, 23273–23294.

Lee, J.M., Greenleaf, A.L. (1991) CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of *Saccharomyces cerevisiae*. *Gene Expr*, **1**, 149–167.

Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A., Kirschner, D.B., Tora, L., Schultz, P. (2002) Mapping histone fold TAFs within yeast TFIID. *EMBO J*, **21**, 3424–3433.

Leurent, C., Sanders, S.L., Demény, M.A., Garbett, K.A., Ruhlmann, C., Weil, P.A., Tora, L., Schultz, P. (2004) Mapping key functional sites within yeast TFIID. *EMBO J*, **23**, 719–727.

Lewis, B.A., Kim, T.K., Orkin, S.H. (2000) A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci U S A*, **97**, 7172–7177.

Li, B., Howe, L., Anderson, S., Yates, J.R., Workman, J.L. (2003) The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, **278**, 8897–8903.

Li, H., Zhang, Z., Wang, B., Zhang, J., Zhao, Y., Jin, Y. (2007) Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. *Mol Cell Biol*, **27**, 5296–5305.

Li, W., Notani, D., Rosenfeld, M.G. (2016) Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nat Rev Genet*, **17**, 207–223.

Li, X.Y., Bhaumik, S.R., Green, M.R. (2000) Distinct classes of yeast promoters revealed by differential TAF recruitment. *Science*, **288**, 1242–1244.

Li, X.-Y., Bhaumik, S.R., Zhu, X., Li, L., Shen, W.-C., Dixit, B.L., Green, M.R. (2002) Selective recruitment of TAFs by yeast upstream activating sequences. Implications for eukaryotic promoter structure. *Curr Biol*, **12**, 1240–1244.

Li, Y., Flanagan, P.M., Tschochner, H., Kornberg, R.D. (1994) RNA polymerase II initiation factor interactions and transcription start site selection. *Science*, **263**, 805–807.

Li, Y., Sabari, B.R., Panchenko, T., Wen, H., Zhao, D., Guan, H., Wan, L., Huang, H., Tang, Z., Zhao, Y., Roeder, R.G., Shi, X., Allis, C.D., Li, H. (2016) Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell*, **62**, 181–193.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y.R., Li, W., Li, H., Shi, X. (2014) AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell*, **159**, 558–571.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., Kadonaga, J.T. (2004) The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev*, **18**, 1606–1617.

Lim, M.K., Tang, V., Le Saux, A., Schüller, J., Bongards, C., Lehming, N. (2007) Gal11p dosage-compensates transcriptional activator deletions via Taf14p. *J Mol Biol,* **374**, 9–23.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C., Shilatifard, A. (2010) AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell*, **37**, 429–437.

Liu, W.-L., Coleman, R.A., Ma, E., Grob, P., Yang, J.L., Zhang, Y., Dailey, G., Nogales, E., Tjian, R. (2009) Structures of three distinct activator-TFIID complexes. *Genes Dev*, **23**, 1510–1521.

Lorch, Y., Kornberg, R.D. (2015) Chromatin-remodeling and the initiation of transcription. *Q Rev Biophys*, **48**, 465–470.

Louder, R.K., He, Y., López-Blanco, J.R., Fang, J., Chacón, P., Nogales, E. (2016) Structure of promoter-bound TFIID and model of human pre-initiation complex assembly. *Nature*, **531**, 604–609.

Lu, H., Zawel, L., Fisher, L., Egly, J.M., Reinberg, D. (1992) Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature*, **358**, 641–645.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251–260.

Lutzmann, M., Kunze, R., Buerer, A., Aebi, U., Hurt, E. (2002) Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *EMBO J*, **21**, 387–397.

Ma, D., Watanabe, H., Mermelstein, F., Admon, A., Oguri, K., Sun, X., Wada, T., Imai, T., Shiroya, T., Reinberg, D. (1993) Isolation of a cDNA encoding the largest subunit of TFIIA reveals functions important for activated transcription. *Genes Dev*, **7**, 2246–2257.

Maldonado, E., Ha, I., Cortes, P., Weis, L., Reinberg, D. (1990) Factors involved in specific transcription by mammalian RNA polymerase II: role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. *Mol Cell Biol*, **10**, 6335–6347.

Malecova, B., Dall'Agnese, A., Madaro, L., Gatto, S., Coutinho Toto, P., Albini, S., Ryan, T., Tora, L., Puri, P.L. (2016) TBP/TFIID-dependent activation of MyoD target genes in skeletal muscle cells. *eLife*, **5**. doi:10.7554/eLife.12534

Malik, S., Barrero, M.J., Jones, T. (2007) Identification of a regulator of transcription elongation as an accessory factor for the human Mediator coactivator. *Proc Natl Acad Sci U S A*, **104**, 6182–6187.

March, S.C., Parikh, I., Cuatrecasas, P. (1974) A simplified method for cyanogen bromide activation of agarose for affinity chromatography. *Anal Biochem*, **60**, 149–152.

Matangkasombut, O., Buratowski, R.M., Swilling, N.W., Buratowski, S. (2000). Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes Dev*, **14**, 951–962.

Maxon, M.E., Goodrich, J.A., Tjian, R. (1994) Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev*, **8**, 515–524.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., Bentley, D.L. (1997a) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, **11**, 3306–3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L. (1997b) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357–361.

Mencía, M., Moqtaderi, Z., Geisberg, J.V., Kuras, L., Struhl, K. (2002) Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Mol Cell*, **9**, 823–833.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., Shilatifard, A. (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A*, **98**, 12902–12907.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., Severinov, K. (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit

RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A*, **98**, 892–897.

Mizzen, C.A., Yang, X.J., Kokubo, T., Brownell, J.E., Bannister, A.J., Owen-Hughes, T., Workman, J., Wang, L., Berger, S.L., Kouzarides, T., Nakatani, Y., Allis, C.D. (1996) The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*, **87**, 1261–1270.

Moqtaderi, Z., Yale, J.D., Struhl, K., Buratowski, S. (1996) Yeast homologues of higher eukaryotic TFIID subunits. *Proc Natl Acad Sci U S A*, **93**, 14654–14658.

Mosley, A.L., Hunter, G.O., Sardiu, M.E., Smolle, M., Workman, J.L., Florens, L., Washburn, M.P. (2013). Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcriptional elongation. *Mol Cell Proteomics*, **12**, 1530–1538.

Mukundan, B., Ansari, A. (2011) Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. *J Biol Chem*, **286**, 37053–37057.

Mumberg, D., Müller, R., Funk, M. (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, **156**, 119–122.

Mumberg, D., Müller, R., Funk, M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res*, **22**, 5767–5768.

Murakami, K., Calero, G., Brown, C.R., Liu, X., Davis, R.E., Boeger, H., Kornberg, R.D. (2013) Formation and fate of a complete 31-protein RNA polymerase II transcription preinitiation complex. *J Biol Chem*, **288**, 6325–6332.

Murakami, K., Gibbons, B.J., Davis, R.E., Nagai, S., Liu, X., Robinson, P.J.J., Wu, T., Kaplan, C.D., Kornberg, R.D. (2012) Tfb6, a previously unidentified subunit of the general transcription factor TFIIH, facilitates dissociation of Ssl2 helicase after transcription initiation. *Proc Natl Acad Sci U S A*, **109**, 4816–4821.

Murakami, K., Tsai, K.-L., Kalisman, N., Bushnell, D.A., Asturias, F.J., Kornberg, R.D. (2015) Structure of an RNA polymerase II preinitiation complex. *Proc Natl Acad Sci U S A*, **112**, 13543–13548.

Murakami, K.S. (2015) Structural biology of bacterial RNA polymerase. *Biomolecules*, **5**, 848–864.

Muthukrishnan, S., Both, G.W., Furuichi, Y., Shatkin, A.J. (1975) 5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation. *Nature*, **255**, 33–37.

Myers, L.C., Leuther, K., Busnell, D.A., Gustafsson, C.M., Kornberg, R.D. (1997) Yeast RNA polymerase II transcription reconstituted with purified proteins. *Methods*, **12**, 212–216.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, **320**, 1344–1349.

Naumovski, L., Friedberg, E.C. (1983) A DNA repair gene required for the incision of damaged DNA is essential for viability in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **80**, 4818–4821.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., Kanemaki, M. (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods*, **6**, 917–922.

Nonet, M., Sweetser, D., Young, R.A. (1987) Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell*, **50**, 909–915.

Oelgeschläger, T., Chiang, C.M., Roeder, R.G. (1996) Topology and reorganization of a human TFIID-promoter complex. *Nature*, **382**, 735–738.

Ohkuma, Y., Roeder, R.G. (1994) Regulation of TFIIH ATPase and kinase activities by TFIIE during active initiation complex formation. *Nature*, **368**, 160–163.

Ohtsuki, K., Kasahara, K., Shirahige, K., Kokubo, T. (2010) Genome-wide localization analysis of a complete set of Tafs reveals a specific effect of the *taf1* mutation on Taf2 occupancy and provides indirect evidence for different TFIID conformations at different promoters. *Nucleic Acids Res*, **38**, 1805–1820.

Olins, D.E., Olins, A.L. (2003) Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol*, **4**, 809–814.

O'Neill, E.M., Kaffman, A., Jolly, E.R., O'Shea, E.K. (1996) Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. *Science*, **271**, 209–212.

Ozer, J., Bolden, A.H., Lieberman, P.M. (1996) Transcription factor IIA mutations show activator-specific defects and reveal a IIA function distinct from stimulation of TBP-DNA binding. *J Biol Chem*, **271**, 11182–11190.

Ozer, J., Mitsouras, K., Zerby, D., Carey, M., Lieberman, P.M. (1998) Transcription factor IIA derepresses TATA-binding protein (TBP)-associated factor inhibition of TBP-DNA binding. *J Biol Chem*, **273**, 14293–14300.

Papai, G., Tripathi, M.K., Ruhlmann, C., Layer, J.H., Weil, P.A., Schultz, P. (2010) TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature*, **465**, 956–960.

Papai, G., Tripathi, M.K., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., Schultz, P. (2009) Mapping the initiator binding Taf2 subunit in the structure of hydrated yeast TFIID. *Structure*, **17**, 363–373.

Pardee, T.S., Bangur, C.S., Ponticelli, A.S. (1998) The N-terminal region of yeast TFIIB contains two adjacent functional domains involved in stable RNA polymerase II binding and transcription start site selection. *J Biol Chem*, **273**, 17859–17864.

Park, E., Guzder, S.N., Koken, M.H., Jaspers-Dekker, I., Weeda, G., Hoeijmakers, J.H., Prakash, S., Prakash, L. (1992) *RAD25* (SSL2), the yeast homolog of the human xeroderma pigmentosum group B DNA repair gene, is essential for viability. *Proc Natl Acad Sci U S A*, **89**, 11416–11420.

Pinto, I., Ware, D.E., Hampsey, M. (1992) The yeast *SUA7* gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection *in vivo*. *Cell*, **68**, 977–988.

Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., Cramer, P. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature*, **533**, 353–358.

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., Petrotchenko, E.V., Borchers, C.H., Baumeister, W., Herzog, F., Villa, E., Cramer, P. (2015) Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, **518**, 376–380.

Poon, D., Bai, Y., Campbell, A.M., Bjorklund, S., Kim, Y.J., Zhou, S., Kornberg, R.D., Weil, P.A. (1995) Identification and characterization of a TFIID-like multiprotein complex from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*, **92**, 8224–8228.

Poon, D., Campbell, A.M., Bai, Y., Weil, P.A. (1994) Yeast Taf170 is encoded by *MOT1* and exists in a TATA box-binding protein (TBP)-TBP-associated factor complex distinct from transcription factor IID. *J Biol Chem*, **269**, 23135–23140.

Poon, D., Schroeder, S., Wang, C.K., Yamamoto, T., Horikoshi, M., Roeder, R.G., Weil, P.A., 1991. The conserved carboxy-terminal domain of *Saccharomyces cerevisiae* TFIID is sufficient to support normal cell growth. *Mol Cell Biol*, **11**, 4809–4821.

Poon, D., Weil, P.A. (1993) Immunopurification of yeast TATA-binding protein and associated factors. Presence of transcription factor IIIB transcriptional activity. *J Biol Chem*, **268**, 15325–15328.

Pugh, B.F., Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev*, **5**, 1935–1945.

Pugh, B.F., Tjian, R. (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell*, **61**, 1187–1197.

Qiu, H., Hu, C., Hinnebusch, A.G. (2009) Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell*, **33**, 752–762.

Rani, P.G., Ranish, J.A., Hahn, S. (2004) RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Mol Cell Biol*, **24**, 1709–1720.

Ranish, J.A., Hahn, S. (1991) The yeast general transcription factor TFIIA is composed of two polypeptide subunits. *J Biol Chem*, **266**, 19320–19327.

Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X., Eng, J., Aebersold, R. (2004) Identification of *TFB5*, a new component of general transcription and DNA repair factor IIH. *Nat Genet*, **36,** 707–713.

Ranish, J.A., Lane, W.S., Hahn, S. (1992) Isolation of two genes that encode subunits of the yeast transcription factor IIA. *Science*, **255**, 1127–1129.

Rappaport, J., Reinberg, D., Zandomeni, R., Weinmann, R. (1987) Purification and functional characterization of transcription factor SII from calf thymus. Role in RNA polymerase II elongation. *J Biol Chem*, **262**, 5227–5232.

Ray, B.L., White, C.I., Haber, J.E. (1991) The *TSM1* gene of *Saccharomyces cerevisiae* overlaps the *MAT* locus. *Curr Genet*, **20**, 25–31.

Reed, L.D. and Muench, H. (1938) A Simple Method of Estimating Fifty Percent Endpoints. *Am J Epidemiol,* **27**, 493-497.

Reese, J.C., Apone, L., Walker, S.S., Griffin, L.A., Green, M.R. (1994) Yeast $TAF_{II}S$ in a multisubunit complex required for activated transcription. *Nature*, **371**, 523–527.

Reinberg, D., Horikoshi, M., Roeder, R.G. (1987) Factors involved in specific transcription in mammalian RNA polymerase II. Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream of the initiation site. *J Biol Chem*, **262**, 3322–3330.

Reinberg, D., Roeder, R.G. (1987a) Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. *J Biol Chem*, **262**, 3310–3321.

Reinberg, D., Roeder, R.G. (1987b) Factors involved in specific transcription by mammalian RNA polymerase II. Transcription factor IIS stimulates elongation of RNA chains. *J Biol Chem*, **262**, 3331–3337.

Reines, D. (1992) Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J Biol Chem*, **267**, 3795–3800.

Reiss, D.J., Mobley, H.L.T. (2011) Determination of target sequence bound by PapX, repressor of bacterial motility, in flhD promoter using systematic evolution of ligands by exponential enrichment (SELEX) and high throughput sequencing. *J Biol Chem*, **286**, 44726–44738.

Rhee, H.S., Pugh, B.F. (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, **483**, 295–301.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Séraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*, **17**, 1030–1032.

Riva, M., Schäffner, A.R., Sentenac, A., Hartmann, G.R., Mustaev, A.A., Zaychikov, E.F., Grachev, M.A. (1987) Active site labeling of the RNA polymerases A, B, and C from yeast. *J Biol Chem*, **262**, 14377–14380.

Roeder, R.G., Rutter, W.J. (1969) Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, **224**, 234–237.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., Stewart, A.F., (2001) The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J*, **20**, 7137–7148.

Ruiz-García, A.B., Sendra, R., Pamblanco, M., Tordera, V., (1997) Gcn5p is involved in the acetylation of histone H3 in nucleosomes. *FEBS Lett*, **403**, 186–190.

Sakurai, H., Fukasawa, T. (1998) Functional correlation among Gal11, transcription factor (TF) IIE, and TFIIH in Saccharomyces cerevisiae. Gal11 and TFIIE cooperatively enhance TFIIH-mediated phosphorylation of RNA polymerase II carboxyl-terminal domain sequences. *J Biol Chem*, **273**, 9534–9538.

Sakurai, H., Kim, Y.J., Ohishi, T., Kornberg, R.D., Fukasawa, T. (1996) The yeast GAL11 protein binds to the transcription factor IIE through GAL11 regions essential for its *in vivo* function. *Proc Natl Acad Sci U S A*, **93**, 9488–9492.

Sanders, S.L., Garbett, K.A., Weil, P.A. (2002a) Molecular characterization of *Saccharomyces cerevisiae* TFIID. *Mol Cell Biol*, **22**, 6000–6013.

Sanders, S.L., Jennings, J., Canutescu, A., Link, A.J., Weil, P.A. (2002b) Proteomics of the eukaryotic transcription machinery: identification of proteins associated with components of yeast TFIID by multidimensional mass spectrometry. *Mol Cell Biol*, **22**, 4723–4738.

Sanders, S.L., Weil, P.A. (2000) Identification of two novel TAF subunits of the yeast *Saccharomyces cerevisiae* TFIID complex. *J Biol Chem*, **275**, 13895–13900.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., Kouzarides, T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature*, **419**, 407–411.

Saragosti, S., Moyne, G., Yaniv, M. (1980) Absence of nucleosomes in a fraction of SV40 chromatin between the origin of replication and the region coding for the late leader RNA. *Cell*, **20**, 65–73.

Sauer, F., Hansen, S.K., Tjian, R. (1995a) Multiple TAF_{II}s directing synergistic activation of transcription. *Science*, **270**, 1783–1788.

Sauer, F., Hansen, S.K., Tjian, R. (1995b) DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* bicoid. *Science*, **270**, 1825–1828.

Sawa, C., Nedea, E., Krogan, N., Wada, T., Handa, H., Greenblatt, J., Buratowski, S. (2004) Bromodomain factor 1 (Bdf1) is phosphorylated by protein kinase CK2. *Mol Cell Biol*, **24**, 4734–4742.

Scheer, E., Delbac, F., Tora, L., Moras, D., Romier, C. (2012) TFIID TAF6-TAF9 complex formation involves the HEAT repeat-containing C-terminal domain of TAF6 and is modulated by TAF5 protein. J Biol Chem, 287, 27580–27592.

Schmitt, M.E., Brown, T.A., Trumpower, B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res*, **18**, 3091–3092.

Schneider, K.R., Smith, R.L., O'Shea, E.K. (1994) Phosphate-regulated inactivation of the kinase PHO80-PHO85 by the CDK inhibitor PHO81. *Science*, **266**, 122–126.

Schulze, J.M., Kane, C.M., Ruiz-Manzano, A. (2010) The YEATS domain of Taf14 in *Saccharomyces cerevisiae* has a negative impact on cell growth. *Mol Genet Genomics*, **283**, 365–380.

Schulze, J.M., Wang, A.Y., Kobor, M.S. (2009) YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. *Biochem Cell Biol Biochim Biol Cell*, **87**, 65–75.

Schwabish, M.A., Struhl, K. (2004) Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol Cell Biol*, **24**, 10111–10117.

Schwer, B., Shuman, S. (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell*, **43**, 311–318.

Sekimizu, K., Nakanishi, Y., Mizuno, D., Natori, S. (1979) Purification and preparation of antibody to RNA polymerase II stimulatory factors from Ehrlich ascites tumor cells. *Biochemistry*, **18**, 1582–1588.

Selleck, W., Howley, R., Fang, Q., Podolny, V., Fried, M.G., Buratowski, S., Tan, S. (2001) A histone fold TAF octamer within the yeast TFIID transcriptional coactivator. *Nat Struct Biol*, **8**, 695–700.

Shanle, E.K., Andrews, F.H., Meriesh, H., McDaniel, S.L., Dronamraju, R., DiFiore, J.V., Jha, D., Wozniak, G.G., Bridgers, J.B., Kerschner, J.L., Krajewski, K., Martín, G.M., Morrison, A.J., Kutateladze, T.G., Strahl, B.D. (2015) Association of Taf14 with acetylated histone H3 directs gene transcription and the DNA damage response. *Genes Dev*, **29**, 1795–1800.

Shao, H., Revach, M., Moshonov, S., Tzuman, Y., Gazit, K., Albeck, S., Unger, T., Dikstein, R. (2005) Core promoter binding by histone-like TAF complexes. *Mol Cell Biol*, **25**, 206–219.

Shen, W.-C., Bhaumik, S.R., Causton, H.C., Simon, I., Zhu, X., Jennings, E.G., Wang, T.-H., Young, R.A., Green, M.R. (2003) Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly. *EMBO J*, **22**, 3395–3402.

Shilatifard, A. (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem*, **75**, 243–269.

Sikorski, R.S., Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*, **122**, 19–27.

Simon, M., Seraphin, B., Faye, G. (1986) *KIN28*, a yeast split gene coding for a putative protein kinase homologous to *CDC28*. *EMBO J*, **5**, 2697–2701.

Sims, R.J., Rojas, L.A., Beck, D., Bonasio, R., Schüller, R., Drury, W.J., Eick, D., Reinberg, D. (2011) The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science*, **332**, 99–103.

Singh, M.V., Bland, C.E., Weil, P.A. (2004) Molecular and genetic characterization of a Taf1p domain essential for yeast TFIID assembly. *Mol Cell Biol*, **24**, 4929–4942.

Sipiczki, M. (2000) Where does fission yeast sit on the tree of life? *Genome Biol*, **1**, REVIEWS1011.

Smale, S.T., Baltimore, D. (1989) The “initiator” as a transcription control element. *Cell*, **57**, 103–113.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.-W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., Allis, C.D. (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol Cell Biol*, **22**, 1298–1306.

Strahl, B.D., Ohba, R., Cook, R.G., Allis, C.D. (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc Natl Acad Sci U S A*, **96,** 14967–14972.

Sugiura, M., Okamoto, T., Takanami, M. (1970) RNA polymerase sigma-factor and the selection of initiation site. *Nature*, **225**, 598–600.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., Cramer, P., (2013) Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell*, **52**, 52–62.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., Cramer, P. (2012) Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res*, **22**, 1350–1359.

Sun, Z.-W., Allis, C.D. (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature*, **418**, 104–108.

Svaren, J., Schmitz, J., Hörz, W. (1994) The transactivation domain of Pho4 is required for nucleosome disruption at the *PHO5* promoter. *EMBO J*, **13**, 4856–4862.

Svejstrup, J.Q., Feaver, W.J., LaPointe, J., Kornberg, R.D. (1994) RNA polymerase transcription factor IIH holoenzyme from yeast. *J Biol Chem*, **269**, 28044–28048.

Sweetser, D., Nonet, M., Young, R.A. (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc Natl Acad Sci U S A*, **84**, 1192–1196.

Swinstead, E.E., Paakinaho, V., Presman, D.M., Hager, G.L. (2016) Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective: Multiple transcription factors can effect chromatin pioneer functions through dynamic interactions with ATP-dependent chromatin remodeling factors. *BioEssays News Rev Mol Cell Dev Biol*, **38**, 1150–1157.

Takagi, Y., Calero, G., Komori, H., Brown, J.A., Ehrensberger, A.H., Hudmon, A., Asturias, F., Kornberg, R.D. (2006) Head module control of mediator interactions. *Mol Cell*, **23**, 355–364.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A.S., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., Florens, L., Seidel, C.W., Lin, C., Smith, E.R., Shilatifard, A., Conaway, R.C., Conaway, J.W. (2011) Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *Cell*, **146**, 92–104.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., Lu, Z., Ye, Z., Zhu, Q., Wysocka, J., Ye, Y., Khochbin, S., Ren, B., Zhao, Y. (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016–1028.

Tanese, N., Pugh, B.F., Tjian, R. (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev*, **5**, 2212–2224.

Thomas, M.C., Chiang, C.-M. (2006) The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol*, **41**, 105–178.

Thompson, C.M., Koleske, A.J., Chao, D.M., Young, R.A. (1993) A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell*, **73**, 1361–1375.

Thut, C.J., Chen, J.L., Klemm, R., Tjian, R. (1995) p53 transcriptional activation mediated by coactivators $TAF_{II}40$ and $TAF_{II}60$. *Science*, **267**, 100–104.

Tora, L. (2002) A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. *Genes Dev*, **16**, 673–675.

Treich, I., Carles, C., Riva, M., Sentenac, A. (1992) *RPC10* encodes a new mini subunit shared by yeast nuclear RNA polymerases. *Gene Expr*, **2**, 31–37.

Trowitzsch, S., Viola, C., Scheer, E., Conic, S., Chavant, V., Fournier, M., Papai, G., Ebong, I.-O., Schaffitzel, C., Zou, J., Haffke, M., Rappsilber, J., Robinson, C.V., Schultz, P., Tora, L., Berger, I. (2015) Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat Commun*, **6**, 6011.

Unnikrishnan, A., Akiyoshi, B., Biggins, S., Tsukiyama, T. (2012) An efficient purification system for native minichromosome from Saccharomyces cerevisiae. *Methods Mol Biol*, **833**, 115–123.

Valay, J.G., Simon, M., Faye, G. (1993) The kin28 protein kinase is associated with a cyclin in *Saccharomyces cerevisiae*. *J Mol Biol*, **234**, 307–310.

Vannini, A., Cramer, P. (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell*, **45**, 439–446.

Venkatesh, S., Smolle, M., Li, H., Gogol, M.M., Saint, M., Kumar, S., Natarajan, K., Workman, J.L. (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*, **489**, 452–455.

Venter, U., Svaren, J., Schmitz, J., Schmid, A., Hörz, W. (1994) A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the *PHO5* promoter. *EMBO J*, **13**, 4848–4855.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., Timmers, H.T.M. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, **131**, 58–69.

Verrijzer, C.P., Chen, J.L., Yokomori, K., Tjian, R. (1995) Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell*, **81**, 1115–1125.

Verrijzer, C.P., Yokomori, K., Chen, J.L., Tjian, R. (1994) Drosophila $TAF_{II}150$: similarity to yeast gene *TSM-1* and specific binding to core promoter DNA. *Science*, **264**, 933–941.

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M., Gut, I., Kremmer, E., Andrau, J.-C., Eick, D. (2015) Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. *Transcription*, **6**, 91–101.

Wang, H., Curran, E.C., Hinds, T.R., Wang, E.H., Zheng, N. (2014) Crystal structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. *Cell Res*, **24**, 1433–1444.

Wang, L., Du, Y., Ward, J.M., Shimbo, T., Lackford, B., Zheng, X., Miao, Y., Zhou, B., Han, L., Fargo, D.C., Jothi, R., Williams, C.J., Wade, P.A., Hu, G. (2014) INO80 facilitates pluripotency gene activation in embryonic stem cell self-renewal, reprogramming, and blastocyst development. *Cell Stem Cell*, **14**, 575–591.

Wang, S.-S., Zhou, B.O., Zhou, J.-Q. (2011) Histone H3 lysine 4 hypermethylation prevents aberrant nucleosome remodeling at the *PHO5* promoter. *Mol Cell Biol*, **31**, 3171–3181.

Wei, C.M., Moss, B. (1975) Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci U S A*, **72**, 318–322.

Weil, P.A., Luse, D.S., Segall, J., Roeder, R.G. (1979) Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. *Cell*, **18**, 469–484.

Welch, M.D., Drubin, D.G. (1994) A nuclear protein with sequence similarity to proteins implicated in human acute leukemias is important for cellular morphogenesis and actin cytoskeletal function in *Saccharomyces cerevisiae*. *Mol Biol Cell*, **5**, 617–632.

Werten, S., Mitschler, A., Romier, C., Gangloff, Y.-G., Thuault, S., Davidson, I., Moras, D. (2002) Crystal structure of a subcomplex of human transcription factor TFIID formed by TATA binding protein-associated factors hTAF4 (hTAF(II)135) and hTAF12 (hTAF(II)20). *J Biol Chem*, **277**, 45502–45509.

Wery, M., Shematorova, E., Van Driessche, B., Vandenhaute, J., Thuriaux, P., Van Mullem, V. (2004) Members of the SAGA and Mediator complexes are partners of the transcription elongation factor TFIIS. *EMBO J*, **23**, 4232–4242.

Wigneshweraraj, S.R., Kuznedelov, K., Severinov, K., Buck, M. (2003) Multiple roles of the RNA polymerase beta subunit flap domain in sigma 54-dependent transcription. *J Biol Chem*, **278**, 3455–3465.

Wintzerith, M., Acker, J., Vicaire, S., Vigneron, M., Kedinger, C. (1992) Complete sequence of the human RNA polymerase II largest subunit. *Nucleic Acids Res*, **20**, 910.

Wood, A., Krogan, N.J., Dover, J., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J.F., Johnston, M., Shilatifard, A. (2003) Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol Cell*, **11**, 267–274.

Woontner, M., Wade, P.A., Bonner, J., Jaehning, J.A., 1991. Transcriptional activation in an improved whole-cell extract from *Saccharomyces cerevisiae*. *Mol Cell Biol*, **11**, 4555–4560.

Woychik, N.A., Lane, W.S., Young, R.A. (1991) Yeast RNA polymerase II subunit *RPB9* is essential for growth at temperature extremes. *J Biol Chem*, **266**, 19053–19055.

Woychik, N.A., McKune, K., Lane, W.S., Young, R.A. (1993) Yeast RNA polymerase II subunit *RPB11* is related to a subunit shared by RNA polymerase I and III. *Gene Expr*, **3**, 77–82.

Woychik, N.A., Young, R.A., 1989. RNA polymerase II subunit *RPB4* is essential for high- and low-temperature yeast cell growth. *Mol Cell Biol*, **9**, 2854–2859.

Wright, K.J., Marr, M.T., Tjian, R. (2006) TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A*, **103**, 12347–12352.

Yakovchuk, P., Gilman, B., Goodrich, J.A., Kugel, J.F. (2010) RNA polymerase II and TAFs undergo a slow isomerization after the polymerase is recruited to promoter-bound TFIID. *J Mol Biol*, **397**, 57–68.

Yang, J., Zhang, Y. (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*, **43**, W174-181.

Yatherajam, G., Zhang, L., Kraemer, S.M., Stargell, L.A. (2003) Protein-protein interaction map for yeast TFIID. *Nucleic Acids Res*, **31**, 1252–1260.

Yen, K., Vinayachandran, V., Batta, K., Koerber, R.T., Pugh, B.F. (2012) Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell*, **149**, 1461–1473.

Yokomori, K., Zeidler, M.P., Chen, J.L., Verrijzer, C.P., Mlodzik, M., Tjian, R. (1994) Drosophila TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes Dev*, **8**, 2313–2323.

Yoon, H., Miller, S.P., Pabich, E.K., Donahue, T.F. (1992) *SSL1*, a suppressor of a *HIS4* 5'-UTR stem-loop mutation, is essential for translation initiation and affects UV resistance in yeast. *Genes Dev*, **6**, 2463–2477.

Yoshikawa, K., Tanaka, T., Ida, Y., Furusawa, C., Hirasawa, T., Shimizu, H. (2011) Comprehensive phenotypic analysis of single-gene deletion and overexpression strains of *Saccharomyces cerevisiae*. *Yeast*, **28**, 349–361.

Zaborowska, J., Egloff, S., Murphy, S. (2016) The pol II CTD: new twists in the tail. *Nat Struct Mol Biol*, **23**, 771–777.

Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., Greenleaf, A.L. (1988) The C-terminal repeat domain of RNA polymerase II largest subunit is essential *in vivo* but is not required for accurate transcription initiation *in vitro*. *Proc Natl Acad Sci U S A*, **85**, 3698–3702.

Zentner, G.E., Kasinathan, S., Xin, B., Rohs, R., Henikoff, S. (2015) ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape *in vivo*. *Nat Commun*, **6**, 8733.

Zhang, G., Darst, S.A. (1998) Structure of the *Escherichia coli* RNA polymerase alpha subunit amino-terminal domain. *Science*, **281**, 262–266.

Zhang, Q., Zeng, L., Zhao, C., Ju, Y., Konuma, T., Zhou, M.-M. (2016) Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure*, **24**, 1606–1612.

Zhang, Z., Boskovic, Z., Hussain, M.M., Hu, W., Inouye, C., Kim, H.-J., Abole, A.K., Doud, M.K., Lewis, T.A., Koehler, A.N., Schreiber, S.L., Tjian, R. (2015) Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. *eLife,* **4**.

Zhang, Z., English, B.P., Grimm, J.B., Kazane, S.A., Hu, W., Tsai, A., Inouye, C., You, C., Piehler, J., Schultz, P.G., Lavis, L.D., Revyakin, A., Tjian, R. (2016) Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev*, **30**, 2106–2118.

Zhou, H., Spicuglia, S., Hsieh, J.J.-D., Mitsiou, D.J., Høiby, T., Veenstra, G.J.C., Korsmeyer, S.J., Stunnenberg, H.G. (2006) Uncleaved TFIIA is a substrate for taspase 1 and active in transcription. *Mol Cell Biol*, **26**, 2728–2735.