Discovery of novel TAF1 DNA binding domains in TFIID promoter recognition and gene transcription

Elizabeth C Curran

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Edith Wang, Chair

Richard Gardner, Member

Chris Hague, Member

Program Authorized to Offer Degree:

Pharmacology

© Copyright 2017
Elizabeth C Curran

University of Washington

Abstract

Discovery of novel TAF1 DNA binding domains in
TFIID promoter recognition and gene transcription

Elizabeth C Curran

Chairs of Supervisory Committee:
Associate Professor Edith H. Wang
Professor Ning Zheng
Department of Pharmacology

The central dogma of biology describes the basic flow of genetic information starting from DNA, which is converted to message RNA, then into functional proteins. This classical view provides a framework for how a cell can respond to its environment through changes in gene expression. While advances in DNA sequencing technology has greatly increased our knowledge of how this can occur on the genomic level, there are still a number of unanswered questions as to how that genetic information is accessed by transcriptional regulators. Controlling when a particular gene is transcribed is vital for cell survival and proliferation. Seminal work from the late 1980s and 1990s demonstrated the transcription factor IID (TFIID) complex as playing a critical role in regulating RNA polymerase II (RNAPII) transcriptional initiation. From these studies, we know TFIID is the first member of the core transcriptional machinery to recognize and bind gene promoter regions located at transcriptional start sites. Initial findings identified five proteins within the TFIID complex that possess DNA binding activities (TBP, TAF1, TAF2, TAF6, and TAF9). With the exception of TBP, little is known about their DNA binding surfaces, even though recent reports show the TAF components of TFIID are essential for promoter specificity.

Presented here is the discovery of two DNA binding domains (DBDs) in TAF1, the largest TFIID subunit. This work identified and characterized two previously unknown DBDs, the winged helix (WH) located in the central DUF3591 domain and the evolutionarily conserved zinc knuckle (ZnK). The WH was exposed through structural studies of the TAF1/TAF7 heterodimer. ZnK was uncovered using bioinformatics as a strictly conserved motif in TAF1. Both DBDs play an imperative role in targeting TFIID to promoters of key cell cycle regulators. Understanding the critical binding regions of TAF proteins will further our knowledge of a fundamental biological process and give valuable insight into how transcription initiation can be regulated.

ACKNOWLEDGEMENTS

The first big thank you goes to Edith Wang for allowing me to explore a truly fascinating aspect of biology. I'm so grateful she was willing to take me on as a graduate student. I learned so much not only about transcription, but how to be a leader in the lab. To me, she is the definition of patience and kindness. Her generosity with her time and openness makes her an exceptionally gifted teacher and mentor. I feel so fortunate to have had such a wonderful experience in her lab. I will cherish the memories and can only hope to pass on my knowledge to future scientist as well as Edith.

Next I would like to acknowledge my co-mentor, Ning Zheng. His enthusiasm for science is unparalleled. During my interview with the Pharmacology department, my escort looked at my schedule and saw that I was interviewing with Ning; her response couldn't describe him any better, "Whenever, I feel down about my science I talk to Ning because his excitement is infectious." I felt this in all of our interactions and always left with the sense that I could accomplish anything. The most important thing that I have taken away from his mentorship has been how to engage people and to convey science in a way that highlights the most intriguing parts of a project. Ning has the remarkable talent to look at a given field and be able to identify the most important and "burning" questions. This astonishing ability to ask deep question often coincides with working on extremely challenging projects. His willingness to be apart of collaboration working to understand fundamental aspects of biology as a whole. It was a privilege to have his mentorship.

Also, thanks to everyone in the Zheng Lab for your help over the years. This project would not have been possible without all your technical aid. The diversity in their cultural backgrounds brought new insights into the way I approached science and was a warm and vibrant atmosphere to conduct research. A very special thank you to Hui Wang. Her work on crystalizing the TAF1/TAF7 complex initiated my thesis project. She has been such a joy to work with and provided not only scientific support but also personal support during the most difficult times working with TAF1. "Dr. Tom" Hinds was an incredible resource during my graduate work. His knowledge of classical biochemistry was extremely helpful.

In the pharmacology break room, people would sometimes jokingly referred to me as The Queen of the Undergrads because everyone knew the Wang lab always had a several undergrads helping out at any one time. They would then be anxiously awaiting an amusing anecdote describing the latest lab snafu. The stories seemed endless about what had become common sense to us. They ranged from not realizing a pipette needed a disposable tip to throwing two-weeks worth of growing and collecting cells to just dumping the lysates in the waste. These conversations always quickly turned to an affectionate recollection of our own early days at the bench. Being surrounded by young scientist provided a unique opportunity to mentor different personality types, or skill levels, all the while teaching me to become a better communicator. The best part was sharing in the motivation behind pursuing a scientific based education and being continually surprise in our new endeavors. They all helped create this body of work, whether it was making minipreps or listening in lab meeting. Thanks to Mike Allen, Hailee Arst, Matt Bernard, Carolyn Brager, Luyi Cheng, Tamara Dominguez, Curt Hauser, Rachael Jappert, Ariana Kamaliazad, Kyle Kilmer, Megan Kufeld, Elly Larson, Fred Lu, Jared Rahn, Jacob Straus, Jennifer Wong, Mega

There are several special thanks for members of the Pharmacology department. First, I would like to thank Rich Gardner for helping get over the many hurdles of grad school. In 2011, I reached out to a local PI, whose science sounded cool, for some advice. I was looking to go back school, and he encouraged me to apply. I would not be in the Pharmacology department without his great advice. A special acknowledgement goes out to Diane Schulsted and Debbie Bale for helping me stay on track and meeting all my deadlines. Diane's selflessness will be something that I carry with me as I continue my scientific career. She always was warm a sense of calm and helped to lift the spirits of everyone around her.

I would like to acknowledge my previous mentors. Thank you to those people who were influential in teaching me a solid scientific foundation. I started in Chris Coughlan lab as an undergraduate student eager to apply my newly acquired textbook knowledge and not fully realizing the overhelming detail that goes into basic lab experiments. I owe my sterile technique to Chris. Thanks for being an amazing mentor and a wonderful introduction to bench science. Thanks to Dave Brow for giving me my first "real" job. I still remember after my first successful in vitro splicing experiment, I made the comment about how I finally felt like a "real" scientists and with such ease he reminded that I already was a "real" scientist. That sense of accomplishment was

instrumental in making me realize I wanted to pursue a career in scientific research. Meghan Moledor in addition to being a fantastic scientific mentor; she was an amazing female role model instilling the importance of work life balance. I will remember her wisdom to never loose my sense of self and making time for the things I love outside of science, like cooking and family.

Thanks for my parents and grandparents who always encouraging me to follow my dreams. I am so appreciative of your continual support of my education over the years. My family has a very strong tradition of pursuing education, and everyone innately searches for knowledge. These values were instilled in me at a very young and gave rise to my curious nature. While my time in graduate school comes to an end, I will always be a life long learner.

Lastly, Kyle, my best friend and life partner throughout this crazy journey; there aren't enough words to express how fortunate I am to have you in my life. Science was only a small fraction of what I learned in grad school. He was there when I was redefining confidence after being surrounded by the world's leading scientist. Always reminding me why I started in the first place, and he helped me remember that spark and passion for science. Thanks for being there every step of the way and knowing I work on TAF1. While this has been some of the most challenging times for us, I wouldn't have wanted to do it without him by my side. I can't wait for all the wonderful adventures ahead.

TABLE OF CONTENTS

GLOSSARY

2xBromo	Double Bromo Domain
BLI	Biolayer interferometry
CD1	Cyclin D1
CD1P	Cyclin D1 Promoter
CPE	Core Promoter Elements
DBD	DNA Binding Domain
DCE	Downstream Core Element
DPE	Downstream Promoter Element
DS	Double Stranded
DUF	Domain of Unknown Function
EMSA	Electrophoretic Mobility Shift Assay
GTFs	General Transcription Factors
IMD	Initiator MTE DPE
Inr	Initiator
ML	Missing Link
MTE	Motif Ten
RNAP	RNA Polymerase
SCP	Super Core Promoter
SS	Single Stranded
TAFs	TBP Associated Factors
TBP	TATA Binding Protein
TF	Transcription Factor
TFIID	Transcription Factor IID
WH3A	Three point mutations in helix 8 of WHD
WHD	Winged Helix Domain
ZnK	Zinc Knuckle
ZnM	Two cysteine mutations in ZnK
ZnRK	Two charged mutations in ZnK

LIST OF FIGURES

CHAPTER1 FIGURES

CHAPTER 2 FIGURES

CHAPTER 3 FIGURES

CHAPTER 4 FIGURES

CHAPTER 5 FIGURES

CHAPTER 1:
BACKGROUND AND SIGNIFICANCE

RNA Polymerase II Transcription

Transcription is an essential biological process for all living organisms, and the first step in gene expression. Accessing the genetic code via transcription is a fundamental action required to sustain life. Many aspects of transcription are conserved across Archaea, Bacteria, and Eukarya. For instance, all cells convert their stable template DNA into transient messenger RNA (mRNA), and every example of transcription can be broken down into three major steps: initiation, elongation, and termination. Initiation involves identifying gene targets in response to cellular needs. Elongation or the synthesis of mRNA is carried out by a multi-subunit enzyme called an RNA polymerase in all living systems. As the RNA polymerase moves along a gene, it sequentially adds RNA bases corresponding to the DNA template creating a polyribonucleic acid chain. The polymerase can catalyze up to several hundred reactions per minute to rapidly generate mRNA. Termination is the last step and completes the transcriptional process, and the mRNA goes on to be translated into functional protein. Bacterial RNA polymerases can bind directly upstream of genes to initiate new mRNA production (Sekine et al., 2012). Conversely, eukaryotic RNA polymerases require additional co-factors to detect transcriptional start sites. The precise regulation of this activity is critical for normal proliferation, differentiation, and development.

The three main eukaryotic RNA polymerases (I, II, and III) were discovered in 1969 and categorized based on their transcriptional targets (Seifart and Sekeris, 1969). The distinctions were made by their sensitivity to α-amanitin, a small molecule transcriptional inhibitor isolated from death cap mushrooms (Kedinger et al., 1970). RNA polymerase I (RNAPI) is insensitive to α-amanitin, while RNAPII is highly sensitive and shows inhibition at 1 μg/mL. RNAPIII is slightly sensitive with inhibition occurring at 50 μg/mL (Zylber and Penman, 1971). RNAPI is unique in that it only transcribes a single gene, 45S ribosomal RNA (rRNA). Human cells have an average rDNA copy number around 200, and the transcriptional products makes up nearly 80% of all RNA in

the cell (Gibbons et al., 2014). RNAPIII transcribes an estimated 400-500 genes that generally consists of non-coding RNAs, such as 5S rRNA, tRNAs and splicing RNAs (Moqtaderi et al., 2010). RNAPII by far has the largest number and most diverse set of transcriptional targets with over 25,000 human genes (Sandelin et al., 2007). The majority of targets are genes encoding proteins, and the expression of specific gene subsets creates transcriptional patterns that can lead to cellular differentiation. Each polymerase has a specialized set of co-factors recruiting them to their target genes (Fig. 1-1; Geiger et al., 2010; Khatter et al, 2017; Buratowski,S,et al., 1989, Taylor et al., 2013). Elucidating the intricate mechanism of recruitment could help delineate normal versus aberrant transcription.

Recruiting RNAPII to transcriptional start sites is orchestrated by a set of General Transcription Factors (GTFs) known as TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, which form a pre-initiation complex (PIC) (Grunberg and Hahn, 2013; Thomas and Chiang, 2006). Sequential recruitment of GTFs to the specific core promoters, DNA sequences signifying RNAPII-dependence, is required for the assembly of a transcriptionally competent PIC. This step-wise process is essential for the proper positioning of RNAPII at specific locations before initiating gene expression (Fig. 1-1). The spatial and temporal regulation is incredibly important for maintaining and responding to cellular needs. A fully assembled PIC is 2.5 megadaltons in size; it's formation is nucleated by TFIID, which directly binds to promoter sequences (Buratowski et al., 1989, Van Dyke et al., 1988). TFIIA further stabilizes this interaction by relieving the tight association and reorienting the TATA binding protein (TBP) component of TFIID towards the core promoter. TFIIB associates with the promoter near the TBP binding site and is firmly established by TFIIF/Pol II recruitment (Zheng et al., 2016). Lastly, TFIIE/TFIIH associates resulting in a PIC poised for elongation. Their binding is essential for promoter clearance (Van Dyke et al. 1988; Orphanides et al., 1996). Advancements in sequencing technologies have given us a wealth of knowledge about the sites of PIC assembly and the mapping of transcription start sites. However, there are still a number of unanswered questions surrounding transcriptional initiation. In particular, how TFIID is able to distinguish core promoters remains a major obstacle in the field. In this work, we use a structural and bioinformatics approach to identify subdomains within TFIID responsible for detecting promoters and help build a foundation for how the complex functions in whole organisms.

TFIID Complex

Making up nearly half the overall size of the PIC, TFIID is a multi-subunit complex comprised of TBP and 14 TBP Associated Factors (TAFs) in humans (Fig 1-2, Pugh and Tjian, 1991). TFIID transcriptional activity was first discovered using classical chromatographic fractionation of mammalian cell nuclear extracts. The fraction containing TFIID was essential to recapitulate transcription *in vitro*. Its low abundance and complexity were major challenges in identifying the mammalian complex; obstacles that continue to be a limiting factor in its study. Researchers turned to model organisms, such as yeast and drosophila, to define the key components and functional role of TFIID in transcription initiation (Dynlacht et al, 1991; Goodrich and Tjian, 2010). A notable discovery included TBP's participation in transcription initiation for all three RNA polymerases. SL1 and TFIIIB are somewhat equivalent complexes to TFIID for RNAPI and RNAPIII, respectively (Knutson et al., 2014, Gibbons et al., 2014, Moqtaderi et al., 2010). The associated co-factors, in this case TFIID's TAFs, are responsible for RNAPII specificity. The eventual isolation and cloning of human TAFs revealed their diverse role in supporting activated transcription.

TBP was originally isolated from crude extracts and identified as being responsible for binding the TATA box sequence (Hoey et al., 1990). TBP is able to support basal levels of transcription; however, researchers elegantly showed recombinant TBP could not respond to activators. Rather co-factors found in partially purified TBP from cells had the capacity to robustly activate transcription upon activator stimulation (Dynlacht et al., 1991). This was the first instance showing a necessary role of TAFs in transcription regulation and that TFIID was potentially a complex. The development of an antibody against TBP was the linchpin that led to the breakthrough of identifying TAF proteins associated with TBP. TAFs were first named based on their molecular weight. However, the size of any one particular TAF was not consistent across species. Genetic screens and biochemically assays led to the complete identification of the TFIID complex subunits across many eukaryotic species and the adaptation of a common nomenclature TAF 1-15 (Tora, 2002).

Some TAFs are not exclusive to the TFIID complex; TAFs are also part of hTFTC (human TBP-free TAF_{II}-containing), hPCAF/GCN5 (p300/CBP associated factors), and hSTAGA (SPT3-TAF_{II}31-GCN5 acetyl transferase). All three complexes contain the histone acetyltransferase GCN5, histone fold TAFs (9, 10, and 12), and lack TBP (Thomas and Chiang, 2006). STAGA and TFTC also contain TAF5, and TFTC further contains the additional TAFs 4, 6, 7 (Wieczorek et al. 1998). The TAFs in these complexes are thought to be necessary for the structural integrity of the core, and the differences come from modular domains that attach via protein-protein interactions and lead to specialized physiological functions (Brand et al., 1999).

Since the discovery of TFIID, its structural features have been extensively examined using electron microscopy, which revealed the overall architecture of the complex. Surveying both the yeast and human forms of TFIID, EM images consistently depict TFIID consisting of three lobes positioned in a clamp shape (Grob et al., 2006; Cianfrocco et al., 2013). These studies also illustrate flexibility within the complex, a feature that may be required for promoter recognition but restricts high-resolution structural determination (Grob et al., 2006, Louder et al., 2016). Using immunolabeling, the stoichiometry and general location of TAFs have been determined (Fig. 1-2, Leurent et al., 2002; Leurent et al., 2004; Papai et al., 2009). It is now know that TAFs interact with p53, c-Jun, SP1, YY1, CFT, Rap1 and viral proteins including HIV-1 Tat and EA1 (Naar et al., 1998, Li et al., 2007; Lui et al., 2009; Papai et al., 2010). Cryo-EM studies show that activators can bind to different subunits in the TFIID complex indicating TFIID is capable of multi-signal integration (Lui et al., 2009). These interactions have been proposed to help recruit TFIID to genes for activation.

TFIID Core Promoter Binding

In addition to interacting with activators, TFIID directly binds RNAPII core promoters. This vital function is carried out by a small subset of TFIID subunits recognizing specific core promoter elements (CPEs) to orient the PIC at transcription start sites (Chalkley and Verrijzer, 1999; Shao et al., 2005, Kim et al., 1993). TFIID's ability to selectivity recognize promoter sequences was initially thought to be largely dictated by TBP binding to the TATA box. TBP was the first TF shown to bind core promoters and was sufficient support basal transcription (Killeen et al., 1992). It is found archaea, as well as eurkayotes, and plays a vital role in transcription initiation for all

systems (Thomm, 1996). However, growing evidence supports the idea that TBP alone is not sufficient to identify gene promoters, especially in higher organisms (Pugh and Tjian, 1991; Weis and Reinberg, 1997; Zhang et al. 2016). The majority of human protein-coding genes lack a TATA box sequence, yet TBP in TFIID is still able to engage TATA-less promoters (Seizl et al., 2011). Therefore, other PIC components must be involved in dictating DNA specificity. A recent study demonstrated that TAFs are imperative for sequence selectivity while TBP showed no sequence preference for a strong promoter over the inactive mutated form (Zhang et al., 2016; Juven-Gershon et al., 2006). These data expand upon previous DNA footprinting analyses that illustrated an extended region of protection near the transcriptional start site and downstream promoter sequences by TBP in complex with TAF1 and TAF2 (Chalkley and Verrijzer, 1999; Matangkasombut et al., 2004).

The specific combination of CPEs varies from gene to gene, and their unique arrangement may help to ensure genes are expressed at specific times in response to environmental queues. Most of these elements are recognized by TAFs to enhance the interaction between TFIID and promoter DNA (Thomas and Chiang, 2006). While there is no one universal sequence for all RNAPII promoters, there are seven major CPEs that have been identified in eukaryotes. The TATA box, one of the first identified CPEs, is found in bacterial promoters as well as eukaryotes. The other core promoter elements include: Initiator (Inr), motif ten element (MTE), downstream promoter element (DPE), downstream core element (DCE) and TFIIB recognition elements upstream and downstream ($BRE_{u\ and\ d}$). The Initiator (Inr) is positioned at the transcription start site of over 40% of human protein-encoding genes and is enriched at TATA-less promoters (Yang et al., 2007; Vo Ngoc et al., 2017). Motif ten element (MTE), downstream promoter element (DPE), and downstream core element (DCE) are located downstream of the transcription start site and thought to contribute to promoter recognition in the absence of a TATA box (Ohler et al., 2002; Thomas and Chiang, 2006; Lewis et al., 2005; Lee et al., 2005). TAF6 and TAF9 bind the DPE through conserved regions adjacent to their histone folds (Shao et al. 2005). TAF1, the largest subunit of TFIID, has been shown to bind the DCE, a discontinuous sequence found in viral promoters as well as the human beta globin gene (Lee et al., 2005). TAF1 also interacts with the Inr when in complex with TAF2 (Chalkley and Verrijzer, 1999). Interestingly, TAF2 alone shows a preference for $YNGAG^{A}/_{C}{}^{A}/_{Y}$ sequences, which has not been classified as a

CPE (Verrijer et al., 1994). However, this sequence does appear in the promoters of cell cycle regulators, such as cyclin B1 (Martin et al. 1999). The vast amount of information we have regarding promoter sequences is overshadowed by how little we know about the important TAF/DNA binding interfaces or their modes of interaction. Characterizing these protein-DNA interactions will be key to understanding how the entire TFIID complex mediates transcriptional initiation.

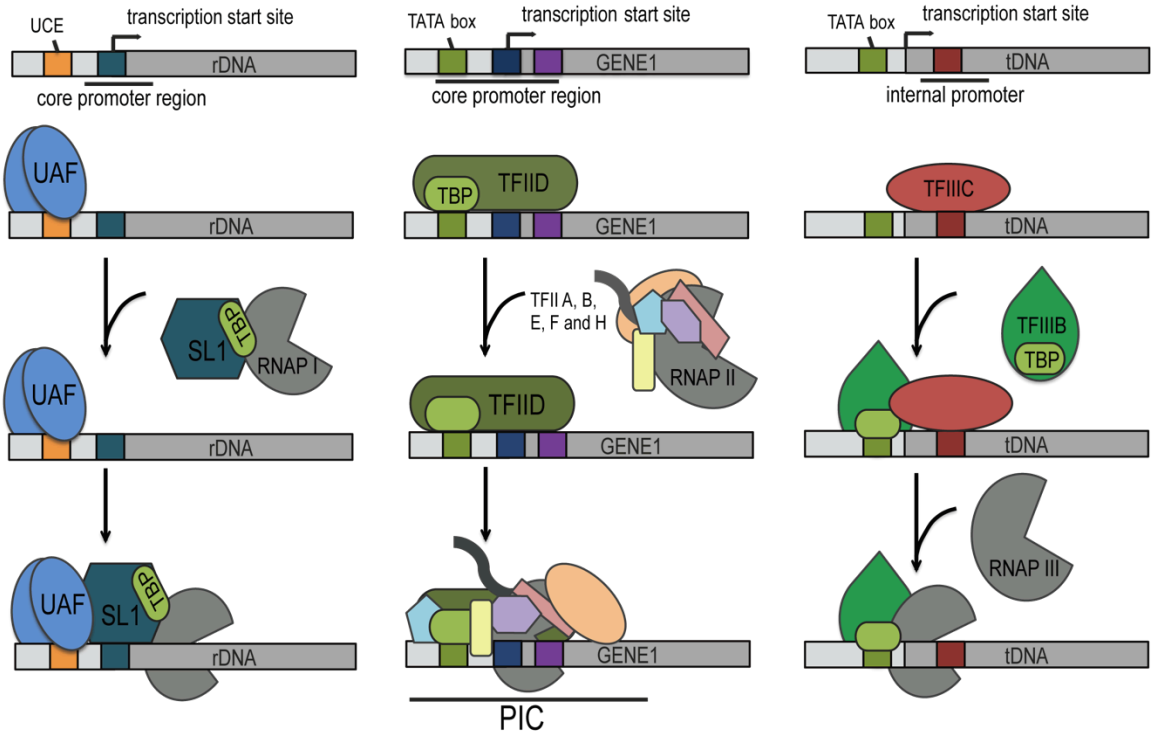
Extensive contact between TFIID and core promoters could serve several purposes. First, TFIID must recognize a variety of promoter sequences; therefore, multiple DNA binding subunits could increase the likelihood of discerning diverse promoters and thus be used universally for nearly all RNAPII targets. Secondly, multiple binding points could act cooperatively and play a role in enhancing TFIID promoter association. A longer dwelling time can increase the probability of proper PIC assembly. A genome wide promoter mapping study found over 10,000 TFIID binding sites across the human genome, and the majority of TFIID sites were located at actively transcribed genes (Kim et al. 2005). Additionally, stable TFIID promoter association may act as a place-maker for gene activation. A subpopulation of TFIID was found at inactive genes, which may represent genes poised for activation. Therefore, these genes would be primed for fast activation as well as sustained transcriptional activity. This idea is supported by data showing TFIID promoter association impacts the duration and amplitude of transcription (Pennington et al. 2003). Lastly, the coordination of multiple contact points could also increase selectivity to prevent aberrant transcription, which may be important in disease pathologies. For example, mutations in the DCE have been linked to β-thalassemia, (Lee et al. 2005), a blood disorder resulting from reduction in beta chain hemoglobin levels. The promoter mutations cause a decrease in gene expression, and consequently an imbalance in hemoglobin components (Thein, 2013). The result is a patient developing microcytic anemia (Cao and Galanello, 2010). Therefore, mapping the DNA-binding interface within TFIID will help us understand how these interactions contribute to transcription and potentially how they can be exploited for therapeutic purposes in disease models.

In fact, manipulating basal transcription factors has recently been highlighted as a potential target for cancer therapeutics (Villicana et al. 2014). Cancer cells have an extremely high transcriptional demand and could be much more sensitive to transcription

perturbations as a novel treatment option (Kotsantis et al. 2016). An overarching therapeutic goal is to directly impede transcription in cancer cells, which could effectively alter the apoptotic/anti-apoptotic balance, and tip the scale in favor of cell death. This line of therapy has the potential to be selectively efficacious because decreasing cancerous mRNA and disrupting the fine transcriptional balance could push cancers into apoptosis while having little detrimental effect of normal cells (Villicana et al. 2014, Brander et al., 2017). For example, mRNA of anti-apoptotic, oncogenes, and proliferation regulators, which are frequently overexpressed in cancer cells (Lee and Young, 2013), have relatively short half-lives and, therefore, decreasing their transcription could lead to faster turnover and rapid down-regulation versus normally expressed stable transcripts (Brander et al., 2017). This could result in cancer cells not being as protected and potentially more prone to apoptosis than health cells by subtly altering the transcriptional output. TFIIH is currently the only GTF targeted in cancer because it plays a dual role in transcription and DNA repair pathways (Compe and Egly, 2012). However, TFIID is an attractive target given its vital role in physiological functions such as cell cycle progression (Villicana et al. 2014). Additionally, dysregulation of TAFs has been observed in several types of cancer, both TAF1 and TAF2 are altered in ovarian cancer. Ovarian cancer cells have a high frequency of TAF1 mutations (Zhao et al, 2013) and often overexpress TAF2 (Ribeiro et al. 2014). While the mechanisms resulting from these changes are still unknown; however, it does emphasize their critical part in maintaining cellular fitness. Given TAF1/TAF2 role in promoter binding, fully understanding their DNA binding surfaces will be crucial for the design of potential drugs that modulate this activity without disrupting their other functions.

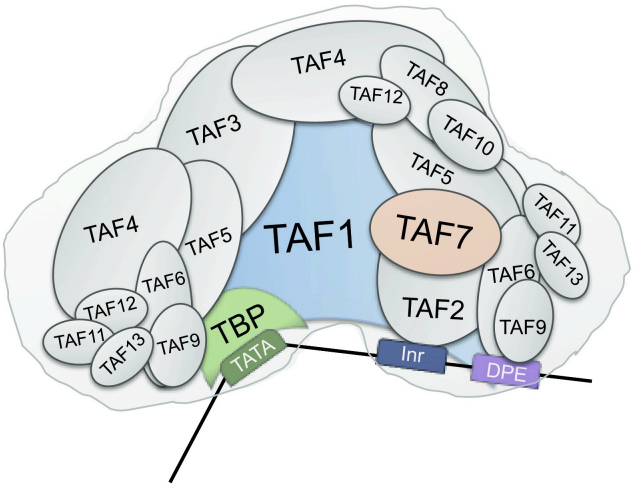
FIGURES

Figure 1 – 1: Eukaryotic Preinitiation Complex Assembly Pathway



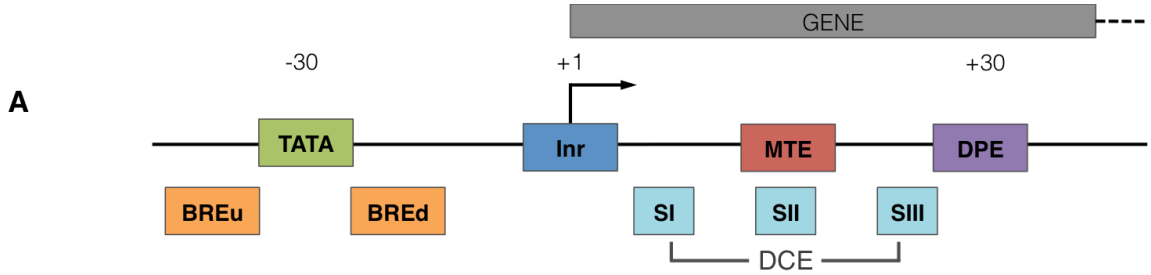
TBP is found in assembly machinery for all three RNAP. **A)** The targeting of RNAPI to rDNA is facilitated by UAF then recruited by SL1, the TBP containing component. **B)** RNAPII requires the most co-factors to form a transcriptionally active PIC. TFIID, a complex with TBP, binding instigates followed by TFIIA. Then TFIIB, TFIIF and RNAPII are recruited, and finally the arrival of TFIIE and TFIIH completes PIC formation. **C)** At tRNA promoters, TFIIIC binds and recruits TBP containing TFIIIB, which then allows RNAPIII binding. TFIIIC then leaves before transcription commences.

Figure 1 – 2: Schematic of TFIID Complex Bound to a Core Promoter



Organization of TFIID components with respect to core promoter elements. The stoichiometry and arrangement of TAFs as determined by immunolabeling in cryo-EM studies. Recognition of CPEs is based on biochemical analysis (figure adapted from Goodrich and Tjian, 2010).

Figure 1 – 3: Summary of Core Promoter Elements



B	Element	TATA Box	Initiator	**M**otif **T**en **E**lement	**D**ownstream **P**romoter **E**lement
	Sequence	TATAWAAR	YYANWYY	CSARCSSAAC	RGWYVT
C	Genome % Prevalence	17%	62%	8.5%	12%
D	Binding Protein	TBP	TAF1 / TAF2	Unknown	TAF6 / TAF9

A) CPE location with respect to transcriptional start site. The super core promoter (SCP) is optimized for transcriptional activity and contains the four promoter elements shown (Juven-Gershon et al., 2006). Features of the SCP depicted in **B)** Element name and consensus sequence, **C)** prevalence of core elements expressed as a percentage, and **D)** binding partners. Abbreviations are as followed A – Adenine, C – Cytosine, G – Guanine, T – Thymine, W – Weak, (AT), S – Strong (CG), R – puRine (AG), Y – pYrimidine (CT), V – not T (V comes after T and U) ACG, and N – any Nucleotide (adapted from Juven-Gershon and Kadonaga, 2010).

CHAPTER 2:
HUMAN TAF1 CONTAINS A STRUCTURALLY CONSERVED WINGED HELIX DNA BINDING DOMAIN

INTRODUCTION

Overview of TAF1

TAF1 is the largest subunit of TFIID, which can interact with TBP through an N-terminal TBP-binding sequence. TAF1 binding represses TBP's TATA box recognition until TFIIA competes off TAF1 association (Anandapadamanaban et al., 2013). TAF1 also directly interacts with TAF7, and RAP74, a subunit in TFIIF, for which the binding sites have been mapped. TAF7 is a dissociable regulator and modulates TAF1 enzymatic activity, including phosphorylation and histone acetyltransferase (Chiang and Roeder, 1995, Kloet et al., 2012). The interaction with RAP74 is thought to help recruit TFIIF and RNAPII to TSS (Ruppert and Tjian, 1995). TAF1 is an essential protein, and a gene deletion causes lethality throughout eukaryotes. Genetic screens and biochemical assays in yeast show all RNAPII genes are dependent on TAF1, with only a few exceptions (Irvin and Pugh, 2006; Warfield et al., 2017). Moreover, knockdown experiments in human cells demonstrate TAF1 augments apoptosis pathways (Kimura et al., 2008). The indispensable and wide-reaching role of TAF1 underscores the importance of discovering the mechanisms of its behavior.

TAF1 has also been reported to possess various biochemical activities including protein phosphorylation, histone acetylation, and acetylated histone tail recognition (Wassarman and Sauer, 2001). In human TAF1, these activities have been mapped to the two terminal kinase domains, a central domain of unknown function (DUF3591 domain), and two tandem bromodomains (2xBromo), respectively (Fig. 2-1A, Dikstein et al., 1996; Jacobson et al., 2000; Mizzen et al., 1996; Ruppert and Tjian, 1995; Wassarman and Sauer, 2001). Due to its large size and solubility issues, structural information for human TAF1 is limited to the approximate cryo-EM mapping of the subunit within TFIID (Cianfrocco et al., 2013; Papai et al., 2011). High-resolution crystal structures only exist for its small TBP-interacting domain and the tandem bromodomains

(Kim et al., 1993; Jacobson et al., 2000). The structural basis of all the other biochemical functions of TAF1 remains elusive.

In addition to structural features of TAF1, many aspects of TAF1 function remain unresolved including the mechanism for binding to core promoters. TAF1's DNA binding domain (DBD) was proposed to be located in its second half. The first group to suggest a DBD in TAF1 annotated an HMG-box based on sequence similarity (Sekiguchi et al., 1991). This postulation stems from this region's homology to other known high mobility group (HMG) proteins, a canonical DBD. HMG-boxes were first discovered in HMG proteins as being mediators of chromatin binding causing their slow or high mobility in native gel assays; now HMG-boxes are common among a wide variety of DNA binding proteins. This superfamily includes proteins involved with DNA replication, chromatin remodeling, and transcription regulation. HMG-boxes contain three alpha-helices in an irregular array that preferentially bind non-B type DNA or bent, kinked, and unwound DNA with high affinity (Malarkey and Churchill, 2012). TAF1's HMG-box has yet to be validated as a bona-fide DBD, and it has even been argued that TAF1's HMG-box lacks sufficient homology to be accurately categorized (Mencia and Struhl, 2001). In light of this, the search for TAF1's DBD remains open-ended. A structural approach gave the much-needed insight into this conundrum and definitively determines a DBD in TAF1. This chapter overviews the discernment of a winged helix domain in TAF1 discovered through structural similarities.

Winged-Helices

Helix-turn-helix (HTH) motifs are one of the most common double stranded DNA recognition domains (Aravind et al., 2003). As the name suggests, a basic motif consists of two helices bridged by a linker creating a turn. The helix following the turn is more often the recognition helix because it contacts DNA, usually in a sequence dependent manner (Struhl, 1989). The solving of the crystal structure of hepatocyte nuclear factor 3 (HNF-3), a transcription factor essential during development, in complex with DNA expanded the simplistic HTH model to include winged helix domains (WHD) (Clark et al.,1993). The HNF-3 fold contained three helices supported by three beta sheets with two loops or wings flanking either side. The third helix fits snuggly into the major groove of B-type DNA; the wings can act as additional sequence recognition

residues. An advantage of this shape is to create a wedge that can open-up DNA (Brennan, 1993, Harami et al., 2013). WHDs have become a wide spread DNA interacting domain in a variety of transcription factors; their role has grown to include the recognition of several types of DNA, physically generating torsion to initiate unwinding, and even mediate protein-protein interactions (Harami et al., 2013). Interestingly, WHDs are found in all three eukaryotic RNA polymerase transcription initiation machinery, including the RNAPII general transcription factors TFIIE and TFIIF (Gajiwala and Burley, 2000; Teichmann et al., 2012). The results presented in this chapter characterize a previously unannotated WHD in TAF1's central DUF domain, and is the first look at TAF1 DNA binding activity. It may also provide a foundation for further exploring TFIID's ability to recognize core promoters.

RESULTS

Overall TAF1-TAF7 Structure

The DUF3591 domain of TAF1 has been linked to the transcriptional regulation of a subset of essential genes, such as G1 cyclins and major histocompatibility (MHC) class I genes (Dunphy et al., 2000; Weissman et al., 1998), but has yet to be stably isolated in a functional form in solution. Initial attempts were made to solve the core structure of TAF1 with a construct including the DUF3591, and RAPiD (RAP74 interacting Domain) to the beginning of the 2xBromo domain. However, this protein was unstable and failed to produce crystals. A soluble form of human TAF1 DUF3591 domain could only be obtained by co-expressing it with full-length human TAF7. The TAF1-TAF7 complex was crystallized using an *in situ* proteolysis approach. Its structure was determined by single-wavelength anomalous diffraction with a platinum-derivatized crystal. The final model contains the majority of the purified TAF1 fragment (amino acids 609 – 1109) and the N-terminal fragment of TAF7 (amino acids 11 – 154) with two TAF1 loops and one TAF7 loop missing, none of which is conserved across eukaryotic species (Fig. 2-1A, B).

The TAF1-TAF7 crystal structure reveals several unpredicted structural domains, which assemble into a compact architecture with the overall shape of a pyramid (Fig. 2-2A). The apex of the pyramid is formed by a winged helix (WH) domain, which is

embedded in the middle of the TAF1 DUF3591 fragment. A wide triple barrel constructed by the interdigitation of both proteins acts as the base of the architecture. The triple barrel fold was first discovered in the crystal structure of the RAP74-RAP30 complex of TFIIF, which acts downstream of TFIID in PIC assembly (Fig. 2-3; Gaiser et al., 2000). The TAF1-TAF7 triple barrel can be superimposed on TFIIF with a root-mean-square deviation (RMSD) of 2.8 Å out of 158 aligned Cα atoms, suggesting that parts of the two complexes might share a common evolutionary origin. The functional importance of this domain will be discussed in detail in Chapter 4.

Additional α-helices and loop regions of TAF1 constitute the middle portion of the pyramid, holding the top WH domain and the bottom triple barrel together. At the primary sequence level, the TAF1 DUF3591 domain can be considered an α-helical polypeptide interrupted by the triple barrel-forming sequence and WH domain (Fig. 2-2B). The TAF7 N-terminal fragment contributes mainly to the construction of the triple barrel core (Fig. 2-2A). Its C-terminal region, however, assumes an independent coiled structure, interacting with both the triple barrel and TAF1 C-terminal α-helical regions (Fig. 2-2B). With the majority of its polypeptide located at one corner of the pyramid base, TAF7 inserts a long β-hairpin through the center of the complex and reaches the solvent on the other side. Together, TAF1 and TAF7 bury a total surface area of ~3900 $Å^2$, consistent with the formation of a stable complex in solution. Conservation of TAF1 and TAF7 are depicted in Figures 2-4 and 2-5 along with several notable features of the sequences.

Structural Aspects of TAF1 Winged Helix Domain

The WH fold consists of three α-helices packing against three β-strands (Fig. 2-6A) and is found in basal transcription factors of RNAPII, subunits of RNAP I and III, and many other transcriptional regulators (Gajiwala and Burley, 2000; Teichmann et al., 2012). These proteins use their WH domains to either recognize DNA or mediate protein-protein interactions. The structure of the TAF1-TAF7 complex unveils a WH domain within the TAF1 DUF3591 fragment, establishing for the first time its existence in TFIID. Previous mutational studies have shown that removal of three amino acids (Δ848-850) in the DUF3591 domain of TAF1 deprived it of the ability to support cell proliferation and transcribe cyclin D1, a cell cycle regulating protein (Hilton et al., 2005).

Intriguingly, this deletion mutation maps to the central region of the WH domain, indicating a vital role for the WH fold in supporting TAF1's function (Fig. 2-4).

A structural homology search identified the DNA-binding E2F4 transcription factor as highly structurally similar to TAF1 WH with an RMSD of 2.1 Å out of 59 aligned Cα atoms (Fig. 2-6A). Moreover, the 75-residue TAF1 WH domain contains 22 positively charged residues, several of which are strictly conserved among TAF1 orthologs (Fig. 2-4). With a predicted isoelectric point of 11.3, the TAF1 WH domain features several highly basic surface areas (Fig. 2-6B). The superposition of TAF1 WH with E2F4 bound to DNA revealed a potential mode of nucleic acid binding (Fig. 2-6C). The third α-helix of the TAF1 WH domain, α8, corresponds to a structural element commonly used by DNA-binding WH proteins, such as E2F4, to recognize the major groove of DNA (Fig 2-6C; Gajiwala and Burley, 2000; Zheng et al., 1999). The α8 helix is solvent exposed on top the TAF1-TAF7 complex and perhaps easily accessible to DNA. Collectively, these features elude to TAF1 WH domain conveying DNA binding function, which could contribute to the interactions of TAF1 with core promoter DNA elements in the context of TFIID or its subcomplexes (Chalkley and Verrijzer, 1999; Lee et al., 2005).

TAF1 WH Domain Binds Promoter DNA

To investigate the DNA-binding activity of TAF1 WH domain, we performed electrophoretic mobility shift assays (EMSA) using TAF1 DUF3591 domain in complex with TAF7, as the TAF1 DUF3591 domain alone is insoluble, and individually purified TAF7. Considering that TAF1 has been previously shown to directly and selectively effect the transcription rates of G1 cyclins such as cyclin D1, we carried out the binding assays with ^{32}P-labeled double-stranded DNA containing human cyclin D1 core promoter (positions -22 to +29, CD1P) sequence. Incubation of TAF1-TAF7 complex with the radiolabeled CD1P produced a complex with reduced mobility compared to free unbound CD1P DNA (Fig. 2-7). By contrast, no mobility-shifted complex was detected with TAF7 alone (Fig. 2-7) under these conditions. TAF7 was able to bind through an electrostatic charged Arg-rich region at sub-physiological salt conditions. Point mutations to these arginines abolished DNA binding (Fig 2-8C). A salt titration was preformed to further characterize the type of interaction TAF7 (Fig. 2-8 A). Increasing

the ionic strength on the binding conditions revealed an inverse relationship between TAF7 DNA binding and salt concentration (Fig 2-8A). The strong ionic dependence indicates TAF7 primary mode of binding is through non-specific charge-charge interactions. At 200 mM NaCl, the majority of TAF7 binding disappeared. Conversely, the TAF1/7 complex did not display similar sensitivity to salt (Fig 2-8B). TAF1/7 binding was maintained up to 300 mM NaCl demonstrating a more complex binding mode that is not purely electrostatic. To reduce any contributions of TAF7 to DNA binding in EMSA, standard binding reactions for TAF1/7 were conducted at 100 mM NaCl. The TAF7 Arg-rich motif forms a one-turn helix and packs directly against the TAF1 Gly-rich loop, which may further restrict the ability of TAF7 to access DNA when in complex with TAF1 (Fig 2-8D, E). These results indicate that the TAF1 component of the complex is responsible for DNA binding.

The superposition of TAF1 WHD with E2F4 WHD bound to DNA suggested that TAF1 WHD may interact with the major groove of DNA helix (Fig. 2-6C). A competition experiment was designed using increasing amounts of unlabeled double-stranded or single-stranded DNA to determine TAF1/7's preference for a type of DNA where. The competition assay clearly showed TAF1/7 binding is specific for double-stranded DNA (Fig. 2-9A). TAF1/7 binding persists in the presence of 50 fold molar excess of unlabeled single-stranded DNA, whereas 5 fold molar excess of double-stranded DNA dramatically decreased binding. This data confirms TAF1/7 binds helical DNA similar to other WHD.

The sequence conservation of the DUF domain and the similarity between the human and yeast structures of human and yeast TAF1/TAF7 complex (Bhattacharyaa et al., 2014) led us to ask whether the yeast ortholog was capable of binding DNA. We tested the ability of the yeast TAF1 core domain to bind double-stranded DNA using EMSA. Yeast TAF1/TAF7 was co-purified and corresponded to the portions of human TAF1/TAF7 used in the crystallographic studies. The complex was able to efficiently bind DNA (Fig. 2-9B). This was the first instance showing a cross-species comparison to identify a DBD within TAF1. The preservation of this DBD emphasizes usefulness of combining structural and conservation biology to elucidate the function of a particular domain.

To test the sequence preference of TAF1-TAF7 DNA-binding activity, we used additional DNA probes of similar lengths to CD1P including a randomized DNA sequence (Random) and the initiator and downstream region (positions -6 to -+38, IMD) of the super core promoter (SCP) in EMSA. Intriguingly, TAF1-TAF7 complex bound to all three DNA probes, but with a slight preference for the IMD promoter fragment (Fig. 2-9C, D). The SCP has been artificially engineered for maximal transcriptional activity by combining the optimal core promoter elements found in eukaryotes (Juven-Gershon et al., 2006). Accordingly, TFIID bound with higher affinity to the SCP compared to other natural core promoters (Juven-Gershon et al., 2006), which parallels our DNA binding EMSA results for the TAF1-TAF7 complex. Because purified full-length TAF1 alone does not bind selectively to specific DNA sequences (Chalkley and Verrijzer, 1999), we predict that its WH domain might support promoter recognition by making sequence-independent DNA contacts.

Superposition analysis of the TAF1 WH domain with transcription factor E2F4 showed that the third α-helix of the TAF1 WH domain (α8) resembles the DNA-binding element of E2F4 (Fig. 2-6C, 2-10A). The α8 helix contains several highly conserved positive charged residues (Arg or Lys) in a row (Fig 2-10B), forming a solvent-exposed basic surface that we hypothesized was responsible for DNA-binding within the WH domain (Fig. 2-6B). In agreement with our prediction, mutation of the conserved positive charged residues R864/K865 to alanine (WH2A) reduced binding (data not shown) while mutating R864/K865/K868 to alanine (WH3A) completely eliminated DNA-binding to CD1P (Fig. 2-10C). None of these mutations disrupted the assembly and biochemical behavior of the TAF1-TAF7 complex (data not shown). These data unveil a promoter-binding function, which is dictated by the TAF1 WH domain and requires the conserved basic residues of α8, a feature of canonical WH-DNA interactions. Intriguingly, a recent survey of uterine serous carcinomas by whole exome sequencing revealed the prevalence of missense mutations at R864 (Fig 2-10B), one of the three positively charged residues on α8 helix of the WH, further implicating the importance of this domain in TAF1 function (Zhao et al., 2013).

Second DNA binding domain within DUF3591

The work describing the WHD provided the first look at TAF1's promoter recognition capabilities. The significance was further validated, when our structure of TAF1 DUF domain bound to TAF7 was modeled into the cryo-EM structure of promoter bound TFIID (Louder et al., 2016). This model confirmed the WH can contact downstream promoter sequences (Louder et al., 2016). The fitting of the TAF1/TAF7 complex was essential for improving the resolution of the TFIID structure and further revealed a small secondary density of protein contacting the Inr, which Louder et al. attribute to TAF1. The orientation of the fitted TAF1/TAF7 complex suggested this density corresponded to a loop (aa 992-1075) that was removed during the limited proteolysis needed to stabilize the complex for crystallization; hence, this domain was named the missing link (ML). This region is only conserved among metazoans. (Fig. 2-11A), and secondary structure analysis predicts this region can form an α-helical bundle (Fig. 2-11A, B). The ML polypeptide was expressed and the purified protein was analyzed for DNA binding activity. Quantitative analysis of the EMSA data demonstrated ML binds with very low affinity, K_d = 25uM (Fig. 2-12A). This finding is consistent with our results for the winged helix mutant in complex with TAF7, which did not show DNA binding even though the ML domain was present (Fig. 2-10C). The protein concentrations used in the WH mutant binding assays were well-below the concentration required to detect any ML binding.

The ML region does show some preference for double stranded DNA. Competition assays show unlabeled double stranded DNA competed binding of labeled DNA at five-fold molar excess compared to the twenty-five-fold molar excess of unlabeled single stranded DNA required to have the same effect (Fig. 2-12B). To further explore the biophysical characteristics of ML binding, binding was examined under various salt conditions. Standard ML binding conditions were 75mM NaCl, to mimic the conditions of the cryo-EM analysis. ML binding was shown to be particularly sensitive to salt; as the concentration of salt increased, binding was severely impaired (Fig. 2-12C). At physiological salt conditions (150 mM NaCl), the majority of binding disappeared. These results suggest the binding can be attributed to weak electrostatic interactions and resembles that of TAF7. With this in mind, we predicted two sets of charged residues, a double arginine (RR Mutant) and double lysine (KK Mutant), may be involved

in this binding activity. These charged amino acids are mostly conserved among organisms containing ML and are predicted to be located on the same plane at the end of two helices. (Fig. 2-11 A, C). Mutating these residues resulted in a reduction in DNA binding activity (Fig. 2-13 A). Moreover, combing the two mutation sets in one protein caused a severe loss of binding capacity (RRKK Mutant, Fig 2-13 A). This indicates both sets of residues are critical for maximal binding. Biolayer interferometry (BLI) analysis using a shift in light to measure binding showed a reduction in steady state association by the ML mutants (Fig. 2-13 C, D, E). The shape of the curves, which depicts the kinetics of binding, were consistent between WT and mutants. This suggests these mutations may impact the conformation of the protein causing a change in the magnitude of the light shift rather than the binding kinetics. Biophysical analysis, such as dynamic light scattering, of the protein under varying salt conditions could determine the effects of salt on the shape of the molecule.

The relatively low affinity of this module to DNA is context dependent and predicts that it may only act as a stabilizing contact point rather than driving TAF1/promoter interactions. The density in the cryo-EM may be an artifact or another DNA binding module in TFIID. The overall advancements in cryo-EM have given us a general idea of the overall shape of TFIID and insights into the topological rearrangements that ensue upon promoter binding (Louder et al., 2016; Cianfrocco et al., 2013). Further refinements are necessary to definitively resolve all the crucial interfaces between the TAF subunits of TFIID and core promoter sequences during the dynamic structural reorganization evoked by promoter engagement.

CONCLUSION

Despite TAF1's important role in TFIID, structure-function studies of TAF1 have been elusive because of its unyielding size, complicated interactions, and a limited view of its activities. The structure depicted in this chapter offers the first glimpse of the human TAF1 central DUF3591 domain, which represents the largest conserved region of the protein. The crystal structure of the TAF1 DUF3591 domain in complex with TAF7 revealed an intricately organized multi-domain architecture, which was characterized by a heterodimeric triple barrel and a multivalent binding mode between the two proteins. Such a feature makes it possible for TAF7 to act as a dynamic regulator of TAF1 as

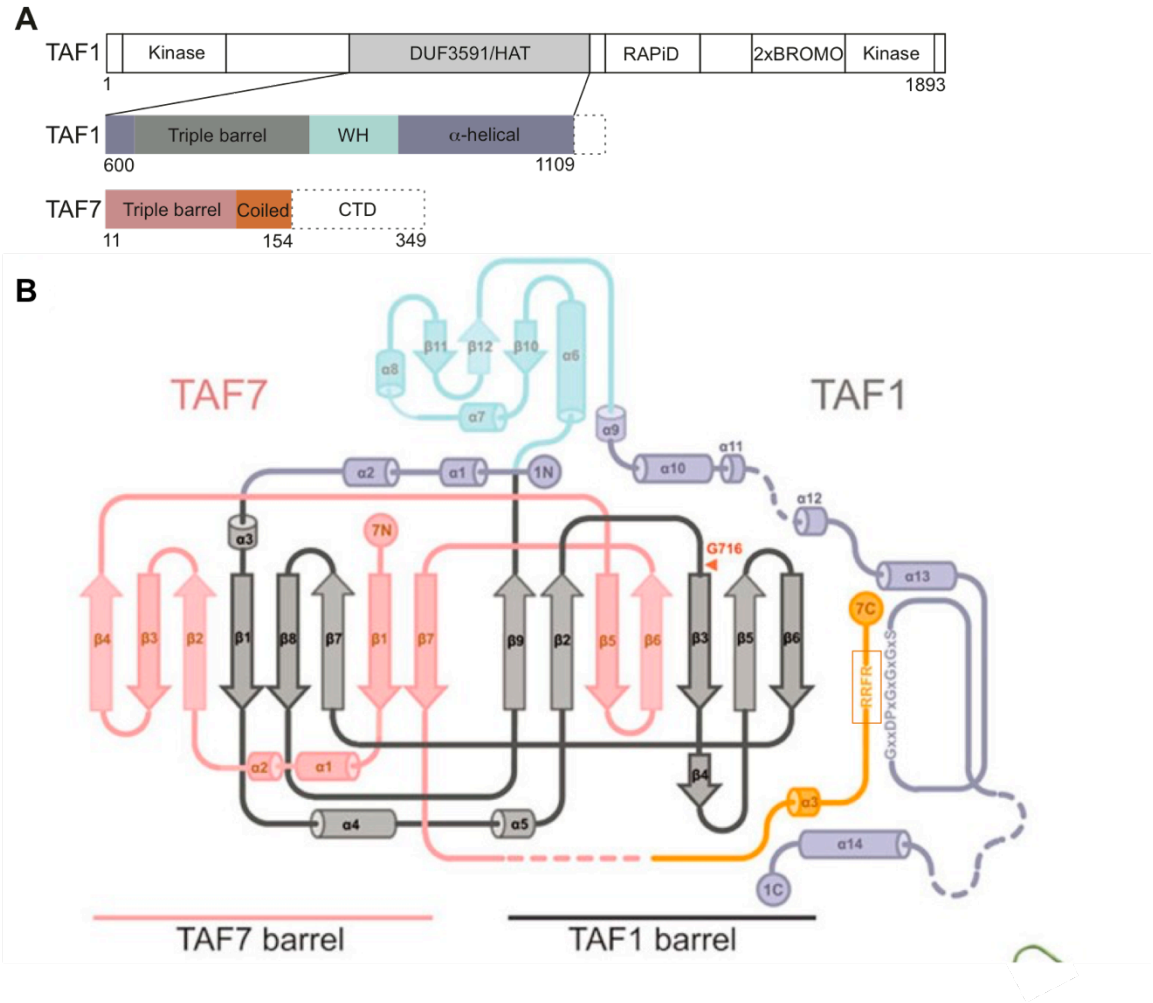
suggested by several studies (Devaiah et al., 2010; Gegonne et al., 2001; Kloet et al., 2012). Multivalency might also be essential for the large-degree of topological reorganization seen with TFIID upon promoter recognition (Cianfrocco et al., 2013).

Furthermore, this structural analysis identified the first TAF1 DBD by unraveling a previously unrecognized WH sub-domain within the core TAF1 DUF3591 domain. The WH domain was shown to possess a non-specific double-stranded DNA-binding activity. Although promoter DNA binding activity of TAF1 has been previously implicated with the cross-species TAF1-TAF2 subcomplex of TFIID (Chalkley and Verrijzer 1999), our structure-based approach revealed for the first time the structural basis of this activity. Moreover, the TAF1/TAF2 heterodimer selectivity for the Inr sequence indicated cooperative binding where one subunit in isolation may not selectively recognize promoter elements.

Based on these results, we hypothesize that TAF1, in the context of TFIID, likely relies on its promoter DNA binding activity to achieve and sustain close contacts with the core promoters of genes, including those required for cell cycle progression. Once TAF1 is recruited to the promoter region by either gene specific activator recognition or the assistance of other TFIID subunits, the WHD of TAF1 provides additional DNA-binding capacity with low sequence preference to maintain and reinforce the interaction of TFIID with core promoters. Upon arrival of additional regulatory events, TAF1 may become appropriately equipped to mediate the transcription of select genes via its proposed enzymatic activity. Despite the lack of sequence selectivity, the promoter DNA binding activity of TAF1 could secure its dwell time on core promoter regions during transcription initiation, adopting a poised state ready to function when necessary. Through its dynamic interaction with TAF1, the peripheral TFIID subunit TAF7 might differentially modulate the DNA binding activity of TAF1 at the promoters of different genes. By combining structural and biochemical analyses, these studies have established the role of TAF1 in promoter DNA binding and furnished the missing structural framework for delineating the function of the TAF1-TAF7 module within TFIID and for understanding the structural ramifications of TAF1 mutations found in human cancers (Zhao et al., 2013).

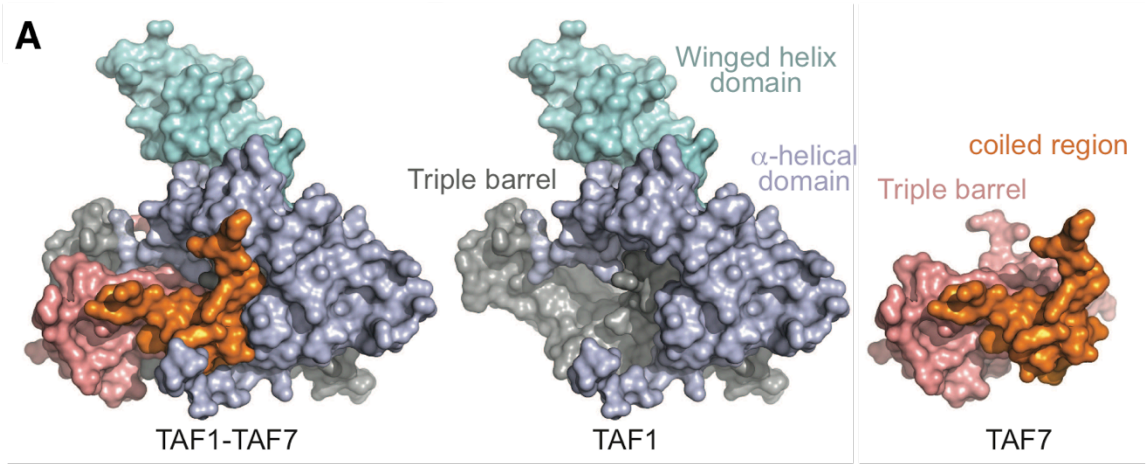
FIGURES

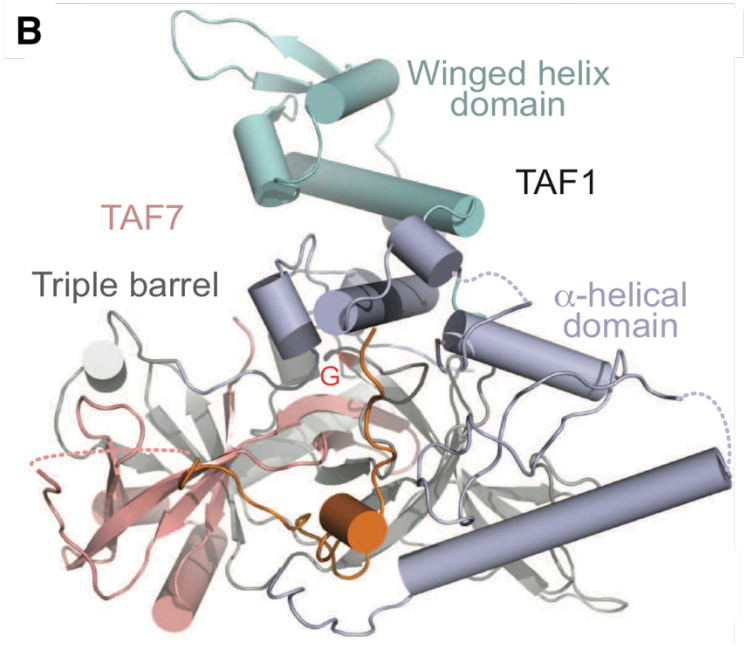
Figure 2 – 1: Graphical Representation of TAF1 / TAF7 Complex



Schematic of TAF1 / TAF7 complex. **A)** Domain organization of TAF1 and the TAF1 and TAF7 fragments crystallized by *in situ* proteolysis method. Different structural domains of the crystallized TAF1 and TAF7 fragments are labeled and colored. Regions outlined by dotted lines represent the proteolytically removed segments. **B)** Topology diagram of the TAF1-TAF7 triple barrel in the context of the entire crystallized complex. Dashed lines represent loop regions not visible in the crystal structure. The Gly-rich motif of TAF1 and the Arg-rich motif of TAF7 are highlighted.

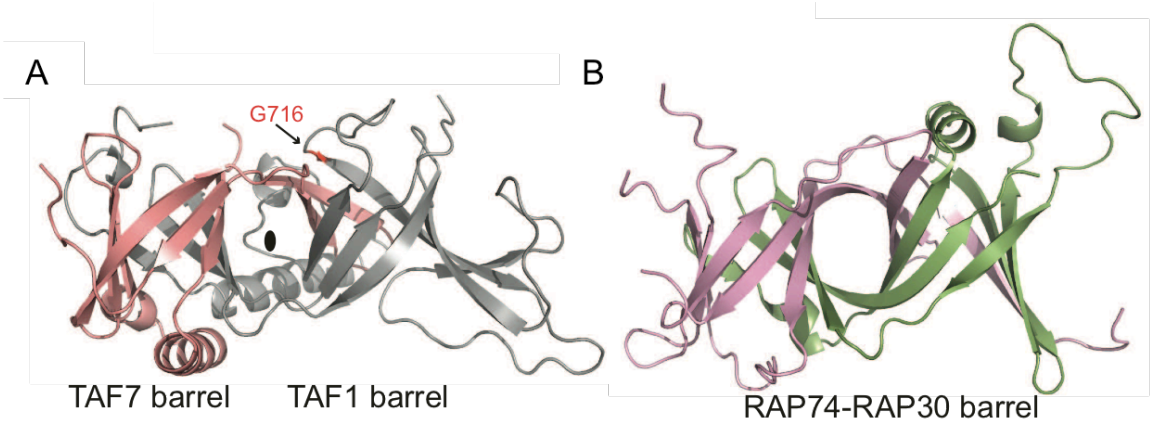
Figure 2 – 2: Crystal Structure of TAF1 / TAF7 in Complex





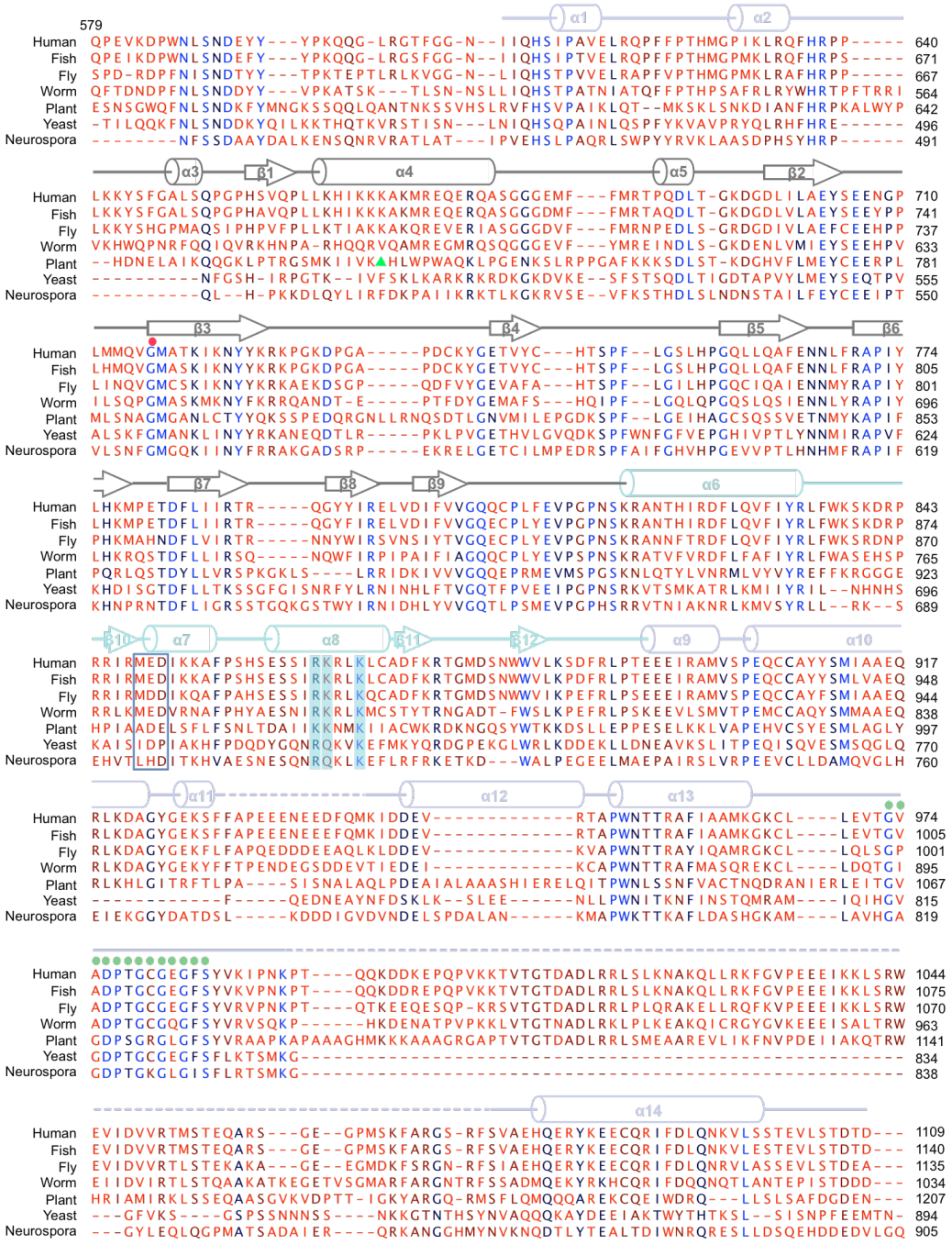
Overall structure of the TAF1-TAF7 complex. **A)** Surface representations of the TAF1-TAF7 complex with structural domains colored with the same scheme as shown in Fig. **2-1 A**. TAF1 and TAF7 are separately shown on the right from the same angle. **B)** Ribbon diagram of the TAF1-TAF7 complex structure in two orthogonal views. The TAF1 and TAF7 protein domains are colored with the same scheme as in Fig. **2-1 A**. The TAF1 G716 residue mutated in the *ts*13 hamster cell line is colored in red and indicated by a letter "G". Dashed lines represent loop regions that are not visible in the crystal structure.

Figure 2 – 3: Heterodimeric Triple Beta Barrel



A) Ribbon diagram of the TAF1-TAF7 triple barrel core. The pseudo two-fold symmetry axis relating the two distal barrels is indicated by the black oval. The residue G716, which is mutated in the *ts*13 hamster cell line, is colored in red and indicated. **B)** Ribbon diagram of the RAP74-RAP30 triple barrel.

Figure 2 – 4: Sequence Alignments of TAF1

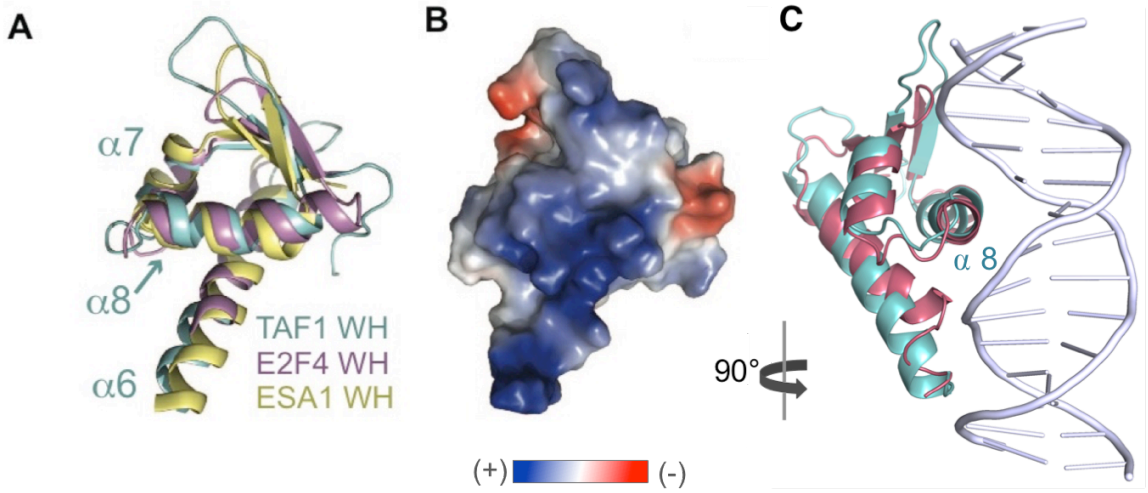


Sequence alignment of DUF3591 domains of representative eukaryotic TAF1 orthologs annotated with schematic of secondary structure depicted above. Green triangle represents an ubiquitin-like domain in plants whose sequence is not shown. Red dot indicates G716, which is mutated in the *ts*13 cell line. Green dots label the conserved Gly-rich motif. The residues deleted in Δ848-850 are boxed in blue. The basic residues mutated in winged helix are highlighted in cyan.

Figure 2 – 5: Sequence Alignments of TAF7

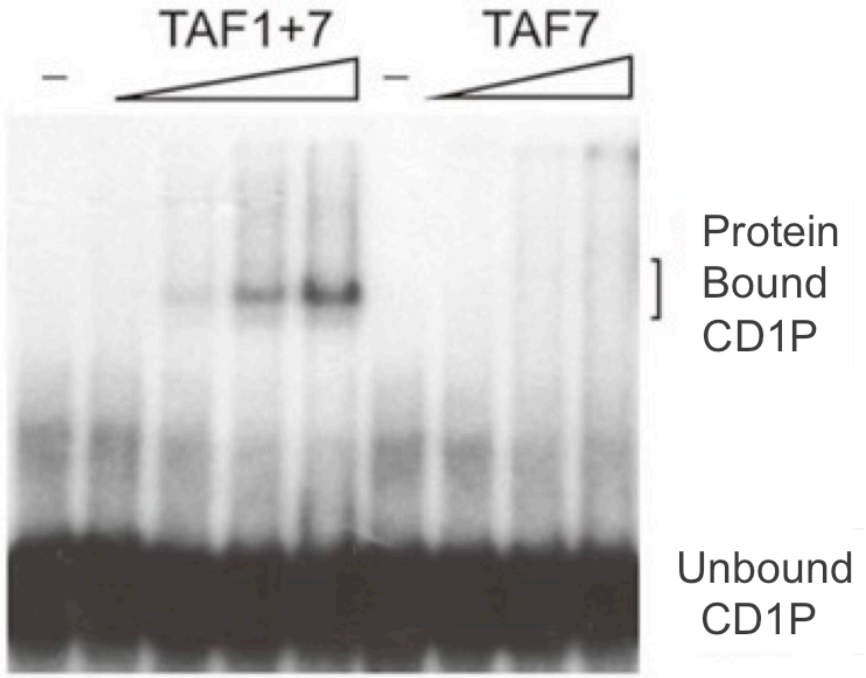
Sequence Alignments of TAF7: Sequence alignment of representative eukaryotic TAF7 orthologs annotated with secondary structures. Only the N-terminal region of TAF7, which is included in the crystal structure, is shown. The sequence of the TAF7 C-terminal domain is highly divergent among orthologs belonging to *Homo sapiens, Danio rerio, Caenorhabditis elegans, Drosophila melanogaster, Arabidopsis thaliana, Saccharomyces cerevisiae, & Neurospora crassa.* Dashed lines represent regions invisible in the electron density. Blue dots indicate the three salt-bridge-forming residues in the TAF7 β-hairpin inserted in the TAF1 barrel. Yellow dots label the Arg-rich motif. A previously mapped 19 amino acids region critical for TAF1 binding is marked by a black bar (Chiang & Roeder, 1993, Science 267:531)

Figure 2 – 6: Structural Features of TAF1's Winged Helix Domain



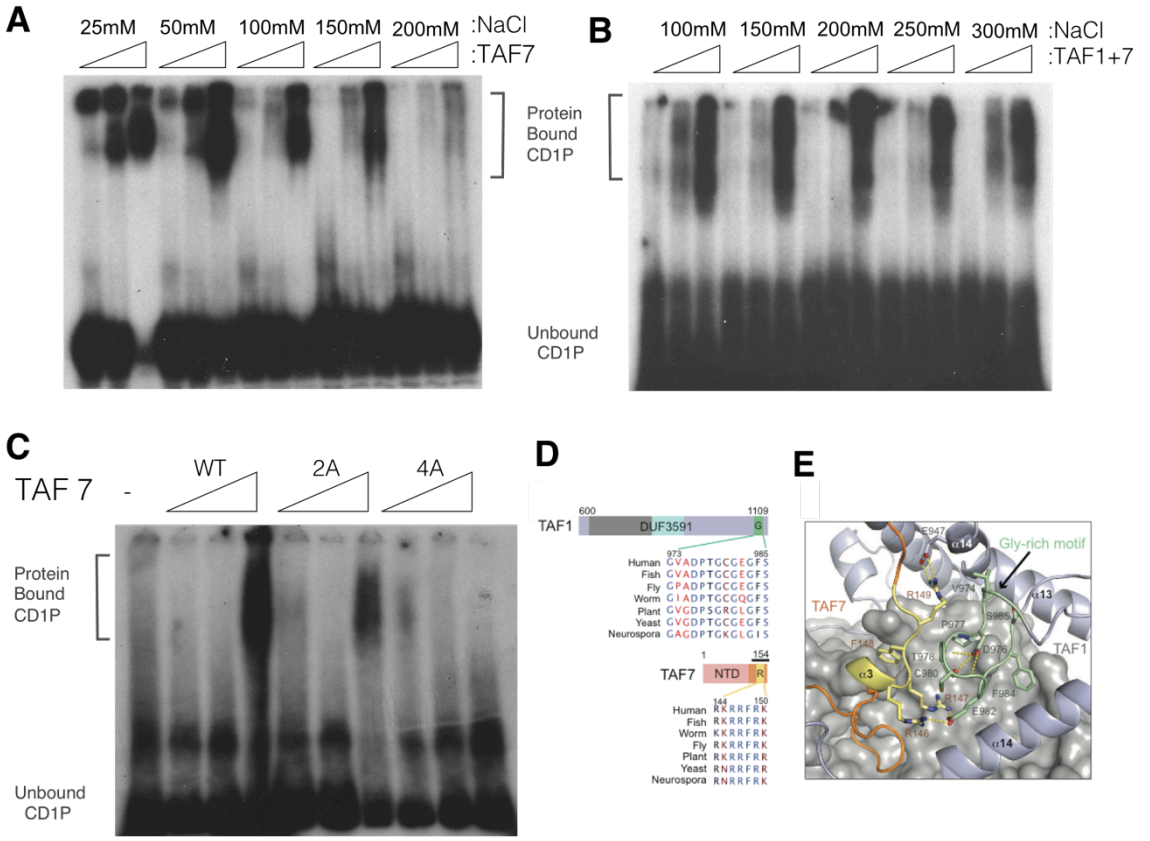
(A) Superposition of the WH domains of TAF1, the transcription factor E2F4, and the yeast HAT ESA1 protein. **B)** Electrostatic potential surface of the TAF1 WH domain viewed from the same angle as in **A** and 180° away. The surface colors are clamped between red ($-83.5kTe^{-1}$) and blue ($+83.5kTe^{-1}$). **C)** TAF1 WH aligned with E2F4 (pdb:1CF7) WHD bound to DNA.

Figure 2 – 7: TAF1's Winged Helix Domain Directly Binds DNA



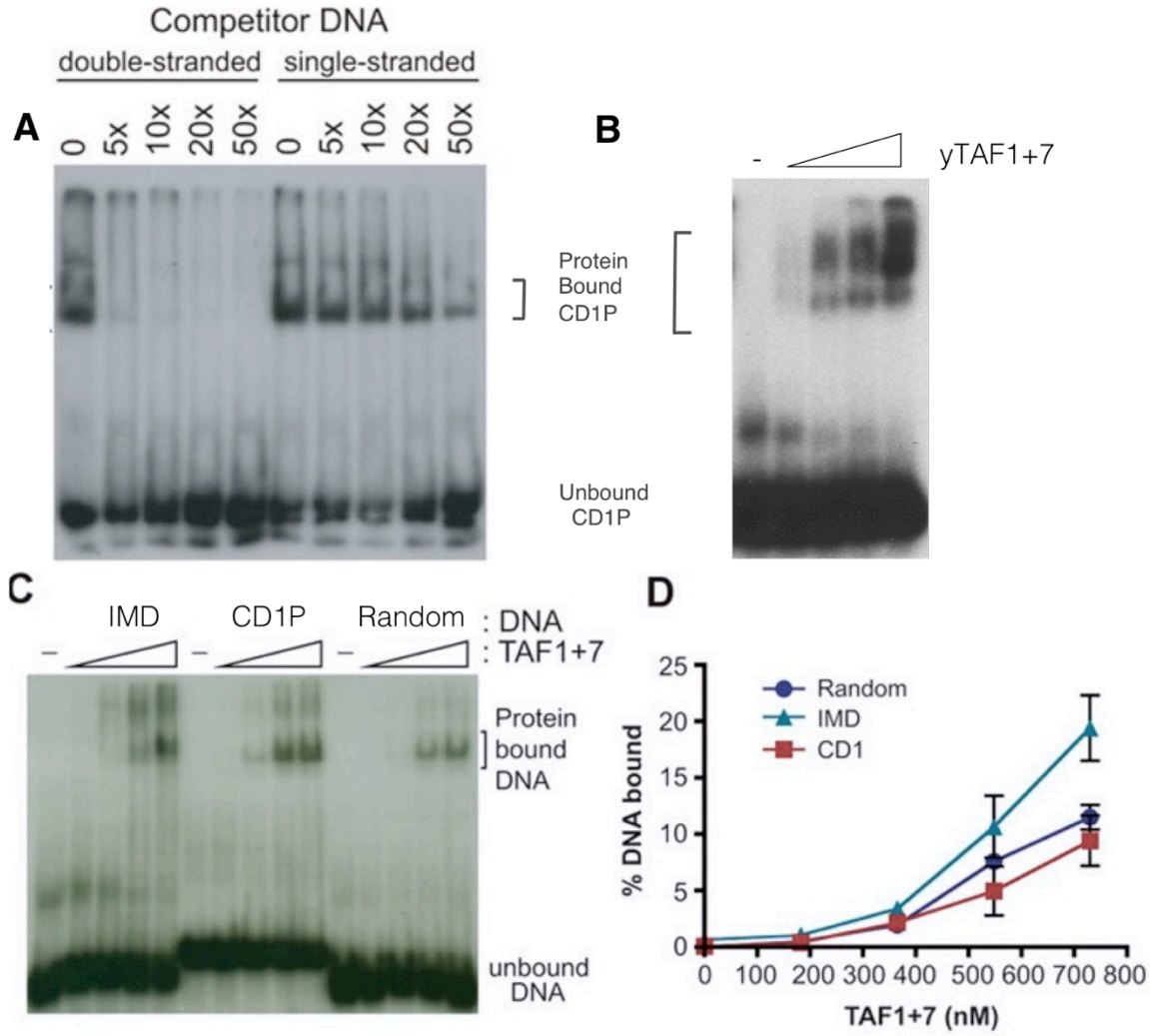
TAF1 displays distinct DNA-binding activity illustrated by a discrete band with slower mobility in EMSA. The EMSA was carried out with increasing concentrations of TAF1/TAF7 corresponding to the material used in Figure 2 – 1 A and full-length TAF7 purified from E. coli. Reactions were resolved on native 5% polyacrylamide before being subjected to autoradiography. The positions of protein-bound and -unbound ^{32}P-labeled cyclin D1 core promoter (CD1P) fragments are shown.

Figure 2 – 8: TAF7's Conserved Loop Can Interact with DNA



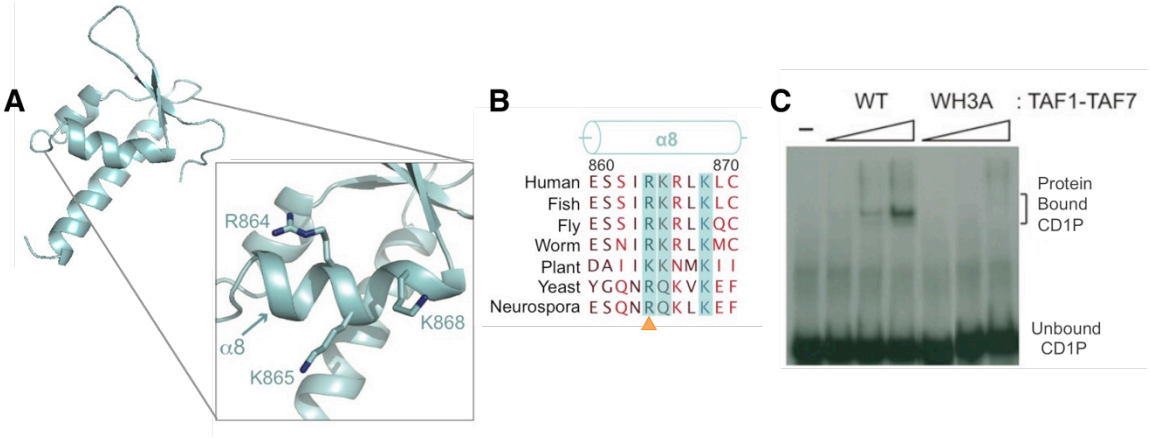
EMSA illustrating **A)** the dependence of TAF7/DNA binding on ionic conditions, while **B)** TAF1/TAF7 binding is insensitive to salt concentrations. TAF7 (196 nM, 393 nM, 787 nM) or TAF1/7 (134 nM, 266 nM, 533 nM) were titrated at each salt condition. As increasing amounts of NaCl were added to the binding reactions as labeled above, TAF1/TAF7 binding persists compared to the decreasing signal of TAF7 binding. **C)** Conserved positively charged residues within NTD of TAF7 (196 nM, 393 nM, 787 nM) mediate DNA binding at 100mM. All binding reactions were performed with full-length TAF7 purified from E. coli or TAF1/TAF7 purified from insect cells and resolved on native 5% PAG before visualization by autoradiography. Shown are the positions of protein-bound and -unbound ^{32}P-labeled cyclin D1 core promoter (CD1P) fragments. **D)** Schematic diagram of the crystallized TAF1 DUF3591 domain and TAF7 NTD with their conserved Gly-rich and Arg-rich sequence motif, respectively. The interacting TAF1 Gly-rich motif and TAF7 Arg-rich motif are colored pale green and bright yellow, and labeled "G" and "R", respectively. Black bar above TAF7 represents the 19-amino acid sequence, which has been documented as a critical TAF1-binding region. **E)** A close-up view of the interactions between the conserved TAF1 Gly-rich motif and TAF7 Arg-rich motif, which are colored in the same scheme as shown in **D**. Dashed lines indicate intermolecular hydrogen bonds and salt bridges. The triple barrel in the complex is shown by surface representation.

Figure 2 – 9: DUF3591 Binding is Specific to Double Stranded DNA and Conserved in Yeast



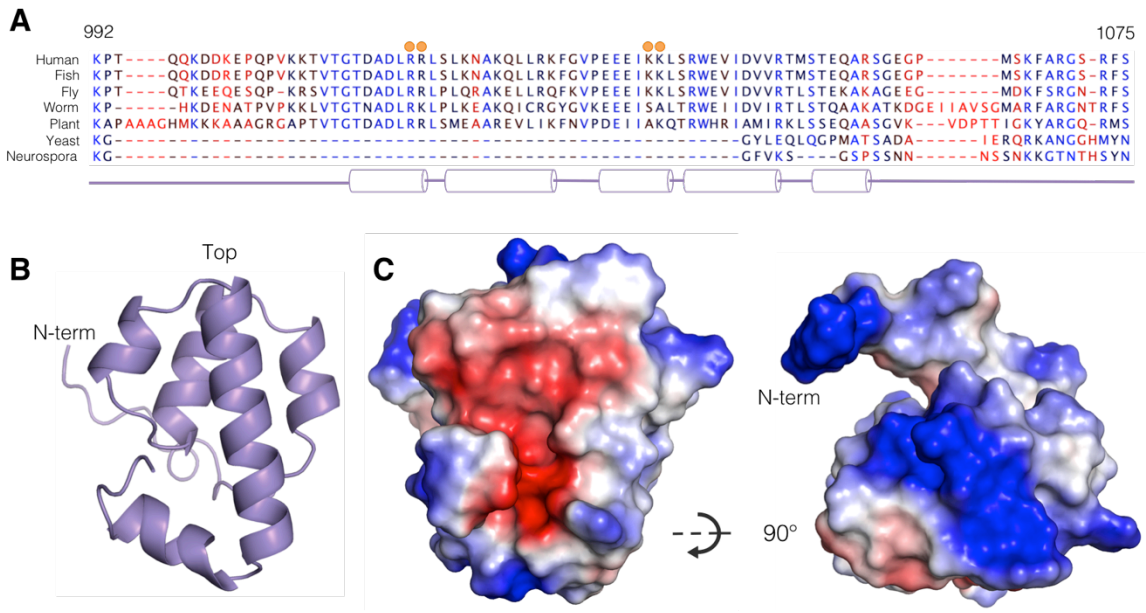
A) TAF1 binds to double-stranded DNA. TAF1-TAF7 (22.5 pmoles) was incubated with radiolabeled CD1P probes in the presence of the indicated molar excess amount of unlabeled double-stranded or single-stranded CD1P. **B)** WH binding is conserved as shown by EMSA using yeast fragments (133nM, 266nM, 666nM, 1.3 µM) corresponding to the human TAF1/TAF7 complex. Autoradiograph of a native 5% PAG shows the positions of protein-bound and -unbound ^{32}P-labeled CD1P fragments. **C)** WHD displays minimal sequence-specificity. Increasing concentrations of TAF1-TAF7 complex was subjected to EMSA using different radiolabeled DNA probes. IMD: initiator and downstream promoter region of super core promoter, CD1P: cyclin D1 core promoter, Random: random double-stranded DNA sequence. Migrations of protein-bound and -unbound DNA fragments are indicated. **D)** Quantitation of DNA binding to the DNA probes described in **C**. Regions containing the bound and unbound radiolabeled DNA probes were excised from the dried gel and quantified by liquid scintillation. The percentage of total counts that the shifted complex represented was calculated for each sample. The graphed results are the average from several independent experiments, IMD: *n* = 3; CD1: *n* = 4; Random: *n* = 3. Error bars represent SEM.

Figure 2 – 10: Charged Residues in α8 Helix of TAF1 were Essential for DNA Binding



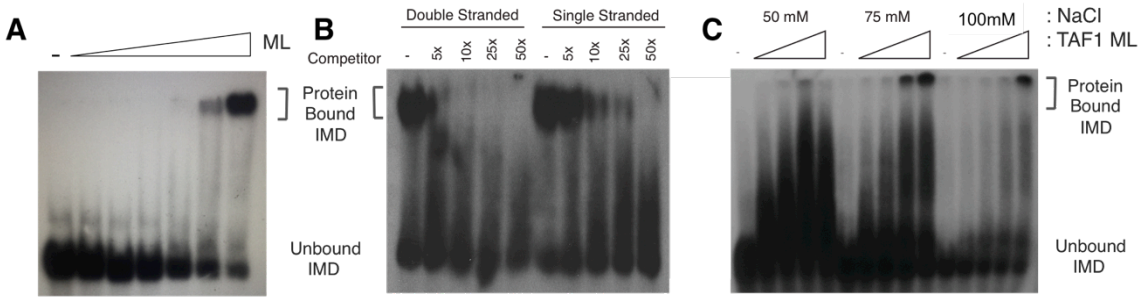
Identification of essential residues in WHD for DNA binding. **A)** A close-up view of the α8 helix of TAF1 with three solvent-exposed basic residues. **B)** Sequence alignment of α8 helix in the TAF1 WH domains. The three basic residues mutated in our studies are highlighted in cyan. Orange triangle indicates mutation identified in exome sequencing of uterine carcinoma. **C)** WH domain mutations abolish DNA-binding activity of TAF1. TAF1-TAF7 complexes assembled in insect cells with WT-TAF1 or TAF1-WH mutant (WH3A) were analyzed by EMSA. Protein-bound and -unbound ^{32}P-labeled CD1P fragments were detected by autoradiography after separation on 5% native PAG.

Figure 2 – 11: Second Putative DNA Binding Domain in DUF3591



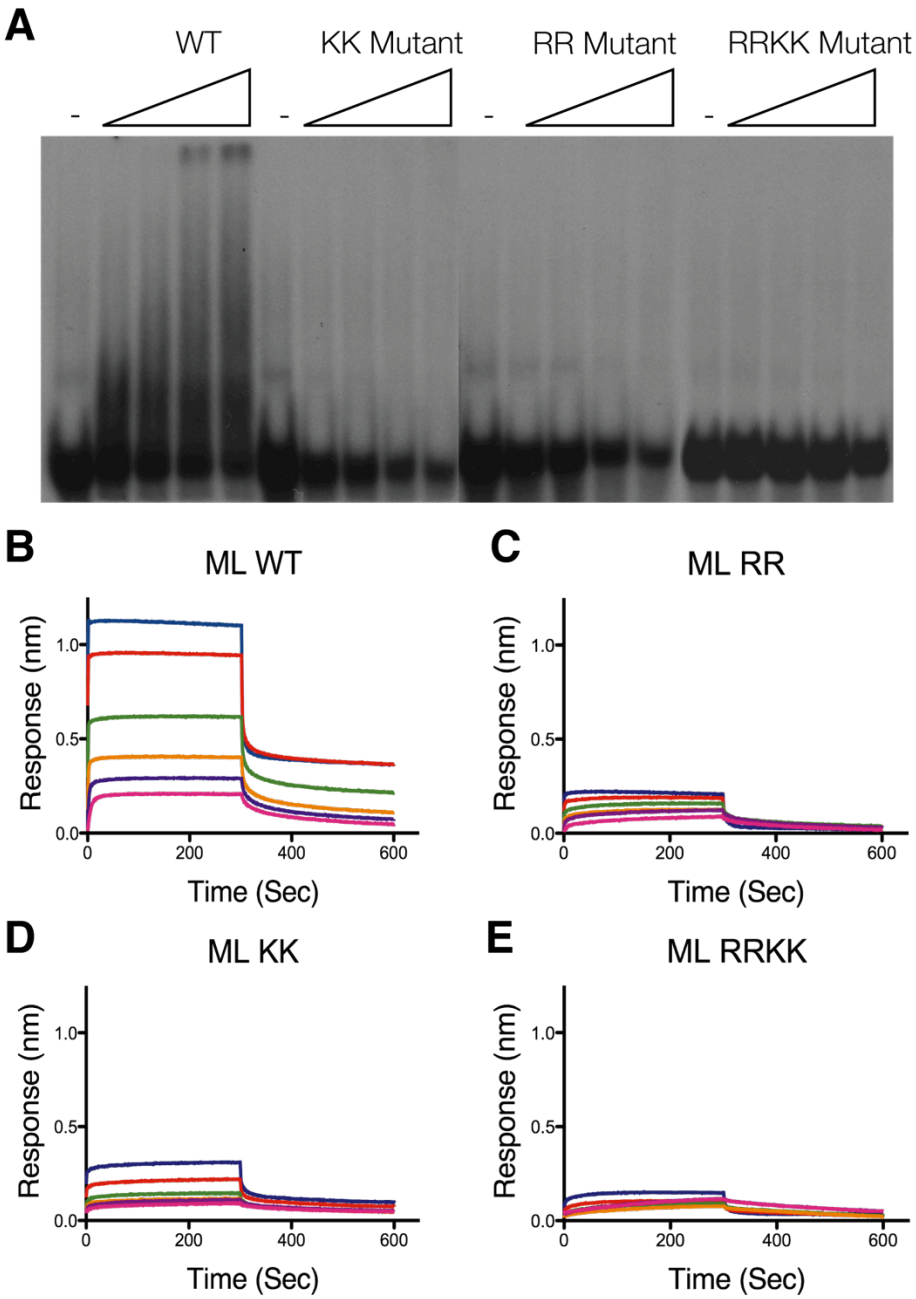
A) Alignment of the missing link (ML, aa 992-1075) region shows conservation among higher eukaryotes. Secondary structure depicted below the alignment. Orange circles indicate alanine mutations of charged residues. **B)** Tertiary structure prediction of ML using I-TASSER and **C)** electrostatic map as well as a 90 degree rotation to view the top of the helical bundle. The surface colors are clamped between red ($-61.6 kTe^{-1}$) and blue ($+61.6 kTe^{-1}$).

Figure 2 – 12: Low Affinity Binding of Missing Link



EMSA characterization of ML. **A)** illustrates binding as a function of increasing concentration up to 27uM of ML protein. **B)** ML (800 pmols) was incubated with 4 ng ^{32}P-IMD probe in the presence of the indicated molar excess amount of unlabeled double or single stranded IMD, and **C)** reflects the sensitivity of binding to salt concentrations. Binding reactions were performed with protein purified from E. coli and resolved on native 6% PAG before being subjected to autoradiography. The positions of protein-bound and -unbound ^{32}P-IMD core promoter fragments are shown.

Figure 2 – 13: DNA Binding of Missing Link Charge Mutants



Charged residues within ML contribute to DNA interactions. **A)** EMSA of ML WT, RR (R1017A/R1018A), KK (K1039A/K1040A), and double mutant on native 6% PAG to separate protein-bound and -unbound ^{32}P-labeled IMD core promoter fragments. Protein concentrations used in assays: 9 μM, 14 μM, 23 μM, and 32 μM. **B)** Bio-layer interferometry binding curves using biotinylated double-stranded IMD fragments described in A and the following ML protein concentrations: 35 μM, 12 μM, 3.8 μM, 1.29 μM, 432 nM, 142 nM**.** Raw data from a representative experiment was plotted with GraphPAD Prism.

CHAPTER 3:
TAF1 CONTAINS AN EVOLUTIONARILY CONSERVED ZINC KNUCKLE

INTRODUCTION

The lack of sequence specificity of the WHD suggests an additional *cis* or *trans* binding motif that assists in TFIID promoter selectivity. The literature indicates TAF1 may in fact contain multiple DBDs; this idea is supported by several pieces of evidence (Lewis et al., 2005; Louder et al., 2016). In addition to the structural data proposed by cryo-EM, TAF1 has been shown to bind two entirely distinct and separate promoter elements (i.e. Inr and DCE) implying multiple binding modules (Chalkley and Verrijzer, 1999, Lee et al., 2005). The DCE is a cluster of three non-contiguous sequences. Each was shown to be important for transcription and could crosslink to TAF1; the current model implies TAF1 directly binds the entire DCE, a region too large to be bound by the WH alone. Moreover, yeast TAF1 has been shown to contain a promoter-binding region near its C-terminus outside the WH (Mencia and Struhl, 2001). The study examined internal TAF1 deletions and uncovered a C-terminal region as being crucial for promoter binding. This region corresponds to the sequence N-terminally adjacent to the double bromodomain in the human ortholog, and yeast TAF1 lacking this region was unable to bind promoter DNA. By comparing this region to human TAF1, we identified an evolutionarily conserved zinc knuckle (ZnK) motif as a genuine DBD. Multiple DNA binding domains in TAF1 may convey plasticity to allow recognition of diverse promoter sequences and permit structural rearrangements during promoter engagement.

Zinc Knuckles

Zinc fingers (ZnF) are synonymous with DNA binding and are the second major nucleic acid binding motif after HTHs. ZnFs are a divergent superfamily able to bind DNA, RNA, proteins and even lipids (Gamsjaeger, R. et al. 2007). They are classified by the arrangement of their zinc coordinating residues (Krishna et al., 2003). Groups are divided based on the number and spacing of cysteine and histidine residues. Multiple zinc coordinating motifs are often found in succession creating a ‘finger’ with each

phalanx acting as a binding surface (Vandevenne et al., 2013). Zinc knuckles (ZnKs) are generally smaller forms of zinc fingers. The CCHC ZnK is commonly found in viral nucleocapsid proteins and is essential for genome packing (De Guzman et al., 1998; Abd El-Wahab et al., 2014). Tandem ZnKs bind to genomic RNA acting as a chaperone to guide structural rearrangement including RNA dimerization for replication and condensing RNA for packaging (Abd El-Wahab et al., 2014). Eukaryotic proteins containing ZnK domains have a variety of functions ranging from nucleotide processing to facilitating post-translational modifications (Armas and Calcaterra, 2013). This chapter reveals a previously uncharacterized ZnK within TAF1 and its ability to directly interact with DNA in a sequence dependent manner. The TAF1 ZnK is a novel and strictly conserved element, which may contribute to promoter binding of TFIID and gene expression.

RESULTS

Second DNA binding domain outside the TAF1 DUF3591 Domain

Our search for a promoter specific DNA binding module within TFIID lead to the discovery a second DNA binding motif in TAF1 outside of the central DUF domain. The first evidence of a second DBD in TAF1 was revealed in a fragment longer than the region used to solve the crystal structure (TAF1b, aa 576-1371 in Fig 3-1C). This TAF1b construct, while not stable enough for crystallization, was used in biochemical analysis. A set of complexes, wild type (TAF1b/TAF7) and winged-helix mutant (TAF1b WH3A/TAF7), were generated and assessed for DNA binding activity. Surprisingly, the WH3A mutant still bound DNA compared to the TAF1a 3A/TAF7 complex (Fig 3-1A). Additionally, the WT TAF1b had two bands at higher concentrations suggesting this complex has two binding modes. An altered stoichiometry could explain the presence of two bands; for example a ratio of one protein to two DNA probes versus the 1:1 ratio presumed for TAF1a and TAF1b WH3A. As a second validation, a smaller fragment of TAF1 without the winged helix was created (TAF1c, aa 992-1371, Fig 3-1C). The benefit of this construct is that it was soluble without TAF7; in an EMSA, this region of TAF1 efficiently bound DNA (Fig 3-1B). The concentrations used for this assay were below the level needed to see binding for the ML DBD. BLI was used as an orthogonal approach to assess binding properties. Biotinylated double-stranded DNA

oligonucleotide was loaded on streptavidin probes and incubated with different concentrations of purified TAF1c protein. Association and dissociation kinetics were monitored in real-time over two consecutive 5-minute periods, respectively (Fig 3-1 B). TAF1c displayed an affinity of 1.3 μM compared to the 34 μM affinity of the ML region. Taken together, this data begun to strengthen the notion TAF1 contains a second DBD outside DUF3591.

TAF1 contains an evolutionarily conserved zinc knuckle motif

TAF1 is an essential protein found in all eukaryotic organisms, and its preservation across these kingdoms signifies its biological importance (Noguchi et al. 1994, Wassarman and Sauer (2001). By using sequence information and mapping regions of high conservation, we identified portions of the protein that were selectively maintained throughout evolution. A sequence alignment performed on full-length TAF1 using eight species that span the eukaryotic kingdom revealed two domains characterized by strictly conserved residues, which are absent in other regions of the protein (Fig 3-2A). The DUF3591 domain is the largest conserved region of which the structural features were discussed in Chapter 2. Notably, in the *ts13* mutant cell line, this domain harbors a temperature-sensitive point mutation (G716D) that causes cells to arrest in late G1 phase of the cell cycle (Noguchi et al., 1994), which will be discussed in detail in Chapter 4 (EXPAND). In brief, the mutation is located in the triple barrel region at the interface between the barrel and the alpha helical domain (Fig. 2-2B). The *ts13* cell line has been used to assess function of TAF1 within the cell cycle (Ruppert et al., 1993, Suzuki et al., 1997; Hilton et al., 2005). This underscores the physiological significance of the core domain, and the power of using conservation to identify regions of importance within the TAF1 protein. The second region of strict conservation sits between DUF3591 and the 2x Bromo. A closer look at the amino acid sequence within this region revealed an invariable $Cx_2Cx_4Hx_6C$ motif (Fig 3-2B).

Bioinformatics analyses predicted this motif as a putative CCHC ZnK, a widely occurring domain commonly found in nucleic acid binding proteins (Darlix et al., 2002; Loughlin et al., 2011; Armas and Calcaterra, 2013). A large fraction of CCHC-domain containing proteins, which can be aligned with TAF1 ZnK, are involved in DNA binding, hinting that TAF1 ZnK also may interact with DNA (Fig 3-3A). In addition to the zinc

coordinating residues, several other amino acids are common among CCHC motifs: glycine following the second cysteine, glycine preceding the histidine, and proline after the third cysteine. The consistent location of these residues suggests that they may be important for proper domain folding. The spacing of the first two cysteines and histidine is strictly conserved; however, the spacing between the histidine and third cysteine can vary. Interestingly, TAF1 shares the same spacing of two proteins: I-factor from *Drosophila* and FAM90a in humans (Abad et al., 2013; Dawson et al., 1997; Bosch et al., 2007). I factor is a LINE-like transposable element and contains a ZnK in ORF1, which has been shown to bind DNA (Dawson et al., 1997). FAM90 is a protein family found in primates, thought to have arisen during multiple duplication and rearrangement events (Bosch et al., 2007). The ZnK in FAM90 has been proposed to function as a DNA binding domain. The commonality of DNA binding to this similarly spaced group of zinc knuckle proteins further suggests that TAF1 ZnK may interact with DNA.

A structural model of TAF1 ZnK was generated using I-TASSER (Zhang, 2008; Roy, et al., 2010; Yang et al., 2015) and shows a compact fold with the C-C-H-C side chains pointing towards the center, allowing for coordination of a zinc ion (Fig 3-3B). Our analysis estimated the probability of ligand binding for the C-C-H-C to be above 90% based on the COACH binding prediction. A structural homology search identified HIV-1 nucleocapsid protein and lin-28 as top hits. Both proteins are known to interact with nucleic acid, and solution structures of these proteins bound to RNA have been solved (Loughlin et al., 2011; De Guzman et al., 1998; Zeng et al., 2016). An alignment between these structures with the modeled TAF1 reveals substantial similarities in the zinc coordinating residues as well as overlapping charged residues (Fig 3-3B). Charged surface analysis on ZnK of TAF1 identified one side of the ZnK fold that is enriched with charged residues with positive electrostatic potential (Fig 3-3C). Positive electrostatic patches are a highly predictive hallmark of DNA binding proteins. Furthermore, this charged surface displays considerable similarity to the HIV-1 nucleic acid binding surface and contains two strictly conserved positively charged amino acids that may be critical for electrostatic interactions (Fig 3-2B). These outward facing positively charged residues produce an interface to presumably interact with negatively charged nucleic acids. TAF1 displays the largest electrostatic patches, which is common in DNA binding compared to RNA binding proteins (Jones et al., 2003, Shazman and Mandel-Gutfreund, 2008). The electrostatic surface more closely resembles that of HIV-1 than lin-28. This

may be due to the differences in the binding modes of the two proteins. The HIV-1 solution structure was solved with genomic viral RNA (De Guzman et al., 1998). The RNA took on a hairpin shape and the HIV-1 protein binds the double stranded RNA section. Conversely, the lin-28 structure bound to single stranded RNA (Loughlin et al., 2011). Taken together, these features signify TAF1 ZnK has the potential to function as a DNA binding domain and thus contribute to TFIID promoter recognition.

TAF1 Zinc Knuckle Binds DNA

Structure homology modeling predicts that TAF1 ZnK potentially interacts with nucleic acids. We performed electrophoretic mobility shift assays (EMSA) to determine if the ZnK domain directly binds DNA. The protein fragment used in this assay contained the annotated ZnK domain and flanking regions (ZnA, aa 1234-1375) so as to not exclude any potential interacting residues. Increasing concentrations of ZnA protein were incubated with three different ^{32}P-labeled double-strand DNA fragments then subjected to native polyacrylamide gel electrophoresis. The fragments represent an optimized promoter containing an Inr, MTE and DPE (IMD), an endogenous promoter, cyclin D1 (CD1P), and a random DNA sequence (Random). We observed that TAF1 ZnA bound to all three DNA fragments (Fig 3-4A). We followed up this analysis by using an orthogonal technique, bio-layer interferometry (BLI), for a more quantitative measurement of binding affinity. Biotinylated double-stranded IMD, CD1P and Random oligonucleotides were loaded on streptavidin probes and exposed to different concentrations of ZnA protein. The association and dissociation kinetics were measured in real-time, respectively (Fig 3-4B,C, D). Binding affinities were calculated based on steady state levels for each protein concentration. Dissociation of the protein from the probe indicated that the proteins were not irreversibly aggregating with DNA. ZnA binds significantly better to the optimized promoter (IMD) and CD1P DNA (288 ± 46 nM, and 411 ± 101 nM, respectively) over random DNA (1011 ± 228 nM). This data correlates with the strength of promoters with these elements. In a luciferase reporter assay, the IMD has the highest activity followed by the CD1 promoter (Fig 3-5). Interestingly, the null construct lacking a promoter sequence is still able to support a low level of transcription indicating general transcription factors are able to bind and initiate transcription from random sequences. The promoter sequences CD1P and IMD were 10 to 100 times more active than the null construct, respectively, which distinctly

demonstrates sequence does play an important role in transcriptional activation. The lack of strong specificity may in fact be a necessary feature of TAF1 given the diversity within RNAPII promoters. Overall, these studies provide a possible mechanism for how compromising the TAF1 ZnK leads to diminished transcription.

Further characterization of this region by BLI revealed some interesting features. As expected, the ZnA fragment preferentially binds to double stranded CD1P (411 nm) over the corresponding single stranded DNA (1359 nm) and RNA (836 nm) (Fig 3-6A,B,C). The affinities for all three probes types was similar to binding random DNA, which may signify TAF1 ZnK can bind several forms of nucleic acids. During active transcription, TAF1 ZnK may encounter single stranded DNA as the transcription bubble forms. Moreover, TAF1 ZnK could also have the opportunity to interact with newly synthesized pre-mRNA as it exits RNAPII. The binding to additional forms of nucleic acids could be mediated through electrostatic interactions with the negatively charged sugar backbone; structural studies such as NMR or crystallography could definitively show these interactions.

Zinc Knuckle is critical for DNA binding

To begin to delineate the ZnK DNA binding domain, TAF1 ZnA protein was incubated with IMD DNA fragment, and the mixture subsequently exposed to increasing concentrations of subtilisin protease. Digestion products were separated by SDS-PAGE, and the pattern of protein fragments compared to the apo-ZnA digestion pattern generated in the absence of IMD incubation. Protein regions bound to DNA were protected from proteolytic cleavage and led to three stabilized species (Fig 3-7A). As anticipated, at the higher concentrations of protease, ZnA is fully degraded, as binding is a dynamic process and the DNA is unable to permanently protect the protein from enzymatic degradation. The full-length construct (ZnA) was considerably stabilized by the presence of DNA along with two smaller fragments (ZnC and ZnD), which were subsequently N-terminally sequenced. Surprisingly, fragment ZnC began at the same amino acid as the intact ZnA protein (aa 1234). The reduction in size, therefore, must be due to cleavage of the C-terminus. The N-terminus of the second fragment (ZnD) was mapped to amino acid 1256.

To further define the domain, we analyzed ZnA sequence with EVfold, a program that mines evolutionary information to detect connections between residues in a protein and predicts a three-dimensional shape based on the co-conservation of amino acids (Marks et al., 2012). Strong connections are given high coupling scores and indicate the residues have a high probability of existing in the same three-dimensional space. Plotting the connections illustrate the distance between co-evolved residues and identify modular domains within proteins. The predicted structure for ZnK takes on a L shape; the ZnK fold hugs an α-helix and creates a cradle enriched in positive charges, presumably a DNA binding surface (Fig 3-8 A, B). This surface corresponds to the stabilized fragment in the limited proteolysis assay. The bound DNA protected this area from the protease leaving the C-termini accessible for cleavage. EVfold analysis reinforced the ZnA construct contains two smaller subdomains: the N-terminus from aa 1234-1313, which contains the ZnK, and a C-terminal domain spanning aa 1313-1375 (Fig 3-7B). It is very striking that the defining boundary between these two subdomains corresponds to the end of the yeast TAF1 homolog. An alignment of fungal TAF1 proteins reveals that sequence conservation ends three amino acids after the last zinc knuckle cysteine residue. Taken together, we mapped the potential minimal ZnK binding domain to be aa 1256-1303 (ZnD) (Fig 3-7C). The ZnD fragment resulted from N-terminal cleavage and removal of the predicted long helix resulting in a minimal compact fold.

We thus generated two constructs, ZnC (aa 1234-1303) and ZnD, and purified these protein fragments (Fig 3-7D). Both ZnC and ZnD are able to bind DNA as shown by EMSA (Fig 3-9A, 3-10A). However, further analysis was limited due to innate biophysical properties; the proteins were ill behaved in these assays. To overcome these technical challenges, ZnC and ZnD were analyzed by BLI as an orthogonal approach, which does not require electrophoretic forces to separate different species. Using IMD loaded probes, the K_d values for wild type ZnC (343 ± 42 nM) and ZnD (503 ± 67 nM) were similar to that obtained for ZnA (Fig 3-9B &10B), suggesting that ZnD represents the core domain of DNA binding. As expected, mutations to the strictly conserved cysteines (Zn_Mut) completely abolished binding activity (Fig 3-9,10C). Due to the lack of appreciable signal above background, this binding data could only be used for qualitative assessment. Next, we sought to determine the residues critical for DNA binding. Guided by our predicted structural model, we mutated two positively charged

residues between the histidine and third cysteine, which are located on the ZnK predicted DNA binding surface and are conserved across species (Fig 3-2B, 3-3E). Consistent with a critical role in binding DNA, mutations of these two positively charged residues either substantially compromised to abrogated the DNA binding activity for both ZnC and ZnD, respectively (Fig 3-9D &10D). The residual DNA binding activity of the ZnC_RK mutant is most likely attributable to several additional positively charged residues at the N-terminal region of the ZnC fragment. Together, these experiments demonstrate the capacity of TAF1 ZnK to bind to DNA and map two charged residues essential for this activity. The identification of an additional DNA binding module in human TAF1 implies TFIID employs a multi-level approach to engage DNA, and with this knowledge we will continue to progress our understanding of promoter recognition and transcriptional initiation.

CONCLUSION

CCHC zinc knuckles are found across the biological spectrum and are critical for a variety of physiological functions. The majority of ZnKs are involved in nucleotide processing including chaperoning, splicing, transcriptional activation and termination (Darlix et al., 2002; Loughlin et al., 2011; Armas and Calcaterra, 2013). The ZnK found in TAF1 has a unique spacing shared by only a small number of annotated ZnK proteins, I factor and FAM90a, both of which have been annotated as interacting with DNA (Abad et al., 1989; Dawson et al., 1997). Given the propensity for zinc knuckles to bind DNA and TFIID's role in transcription, we sought to determine if TAF1 ZnK interacts with promoter DNA. We found that TAF1 ZnK can directly bind to DNA with an affinity similar to other zinc knuckles (Dey et al., 2005; Vandevenne et al., 2013) and identified the residues important for this function.

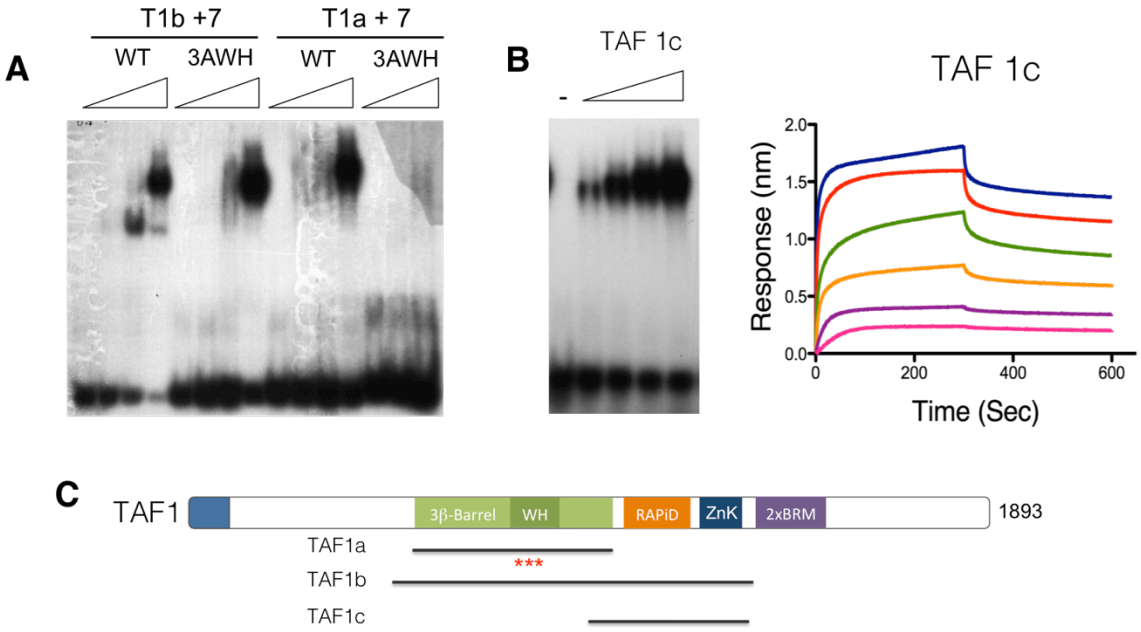
This chapter highlights the usefulness of leveraging a bioinformatics approach for discovering the function for uncharacterized protein domains. Conservation mapping and predictive modeling were critical for accurately defining the ZnK DBD. Progressive truncations validated the theoretical minimal DBD. Using structural homology comparison, two charged restudies were found to be important for DNA binding based on the similarity to other ZnK nucleic acid binding surfaces. Interestingly, the ZnK overlaps with the region that was first proposed as being involved in DNA binding. While

misclassified as an HMG-box, this region was important for TFIID binding to yeast promoters and now our data proves the DNA binding module responsible resides in this portion of TAF1.

The sequence specificity of the ZnK, while significant, is not absolute. This mirrors ChIP-seq studies that fail to identify any sequences enriched by isolating endogenous TAF1 (http://www.factorbook.org/human/chipseq/tf/TAF1/#). This data strengthens our data showing a marginal yet significant preference for promoter DNA over a random sequence. TAF1 ZnK is still able to bind random DNA with the considerable affinity of 1μM suggesting ZnK may bind with high affinity at some promoters and low affinity at other. This could be also argued as a caveat of using an *in vitro* system to understand the complexities of the physiological conditions. Follow-up experiments such as ChIP seq with mutant forms of TAF1 could help resolve this proviso and identify any sequences that preferentially bind with an intact ZnK domain over WHD. Likewise, structural studies would definitely establish the mechanism of binding to DNA and be used to clearly visual if the protein/DNA contacts could allow for flexibility in sequence recognition. Additionally, RNA-seq could also be used to validate TAF1 ZnK can bind RNA *in vivo*, and structural studies could show whether or not RNA could fit the model. Overall, the discovery of a second DBD with in TAF1 could explain how TFIID accommodates for the diversity among RNAPII gene promoters, which also do not contain any one single universal element. Therefore, multiple DBDs could be preferentially used at different promoters giving TFIID's plasticity to recognize nearly all RNAPII dependent genes.

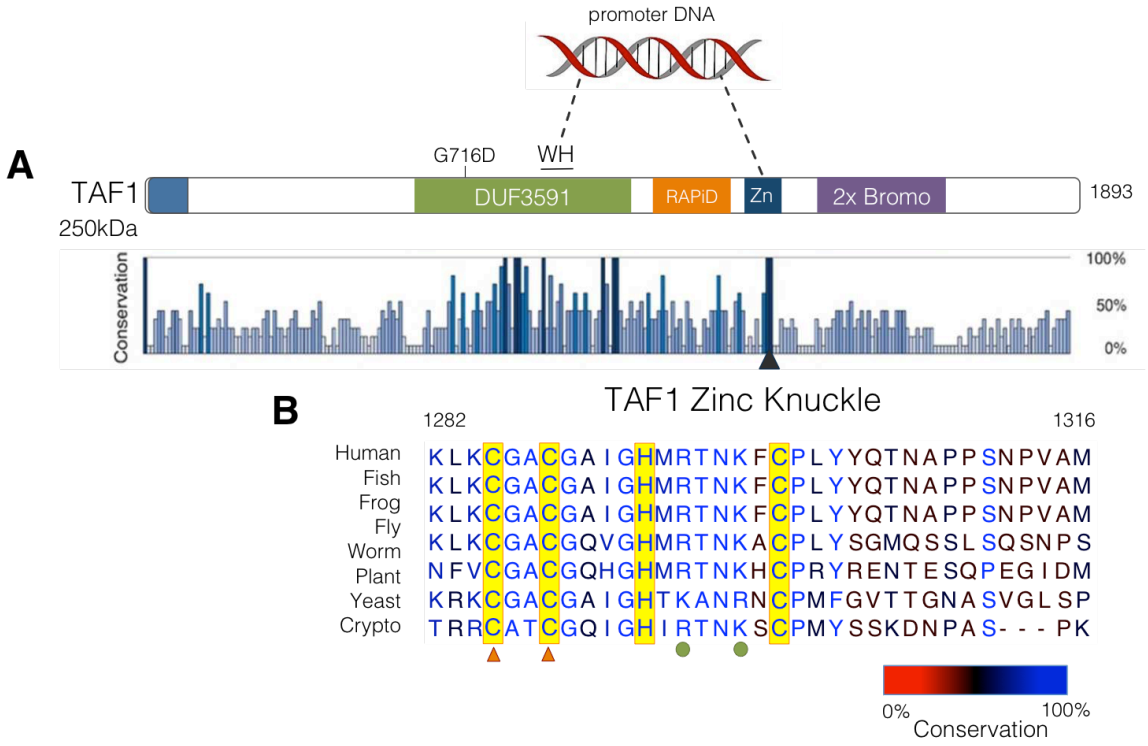
FIGURES

Figure 3 – 1: TAF1 Contains Multiple DNA Binding Domains



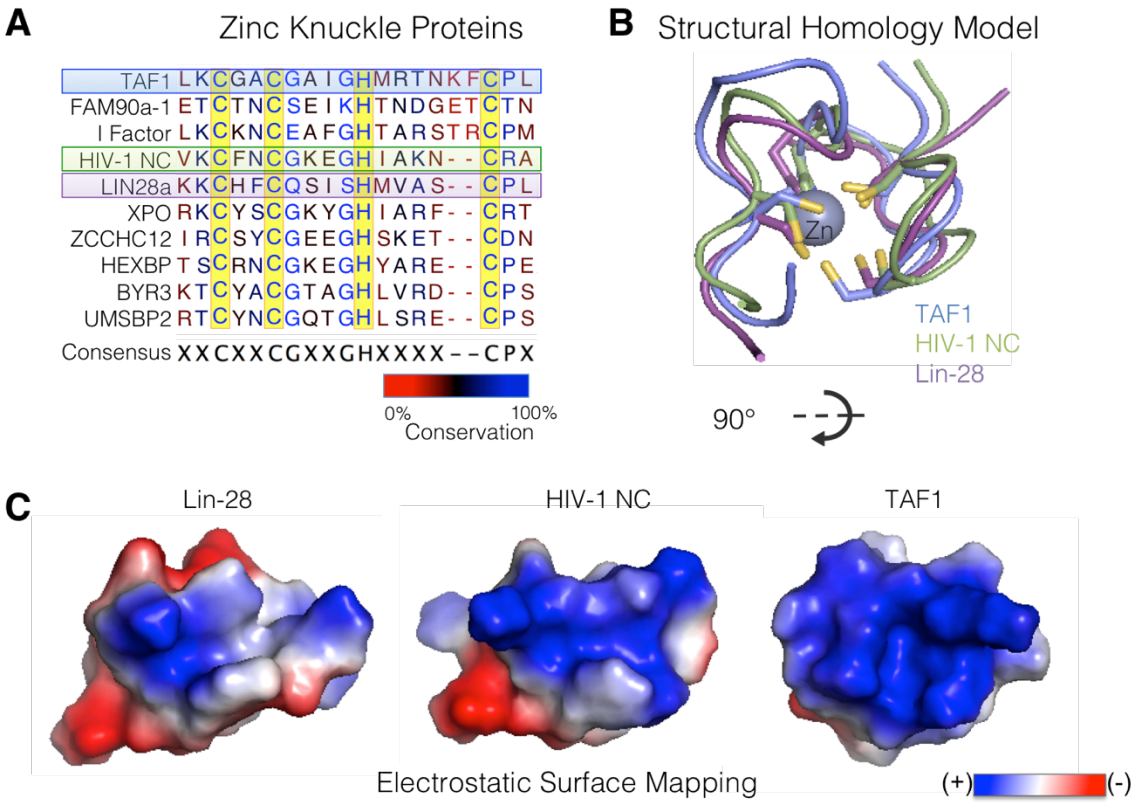
TAF1 DNA binding was retained in the absence of functional WH. **A)** EMSA were performed with two sets of complexes human, TAF1a (aa 600-1109) and TAF1b (aa 576-1375) co-purified with full length TAF7 from insect cells. WT and 3AWH forms of each TAF1-TAF7 were titrated, bound to ^{32}P-labeled IMD core promoter fragments and complexes were resolved on native 5% PAG and visualized using autoradiography. **B)** TAF1c (aa 992-1375) purified from E. coli was subjected to EMSA and resolved on native 5% PAG before autoradiography. The positions of protein-bound and -unbound ^{32}P-labeled IMD core promoter fragments are shown. Bio-layer interferometry binding curves using biotinylated double-stranded IMD fragments described in A and the following TAF1c protein concentrations: 30 μM, 10 μM, 3.3 μM, 1.1 μM, 370 nM, 120 nM. Raw data was plotted with GraphPAD Prism for one representative binding experiment. **C)** Linear Schematic of TAF1 constructs with asterisk indicating the location of WH3A mutations.

Figure 3 – 2: TAF1 Contains an Evolutionarily Conserved Zinc Knuckle



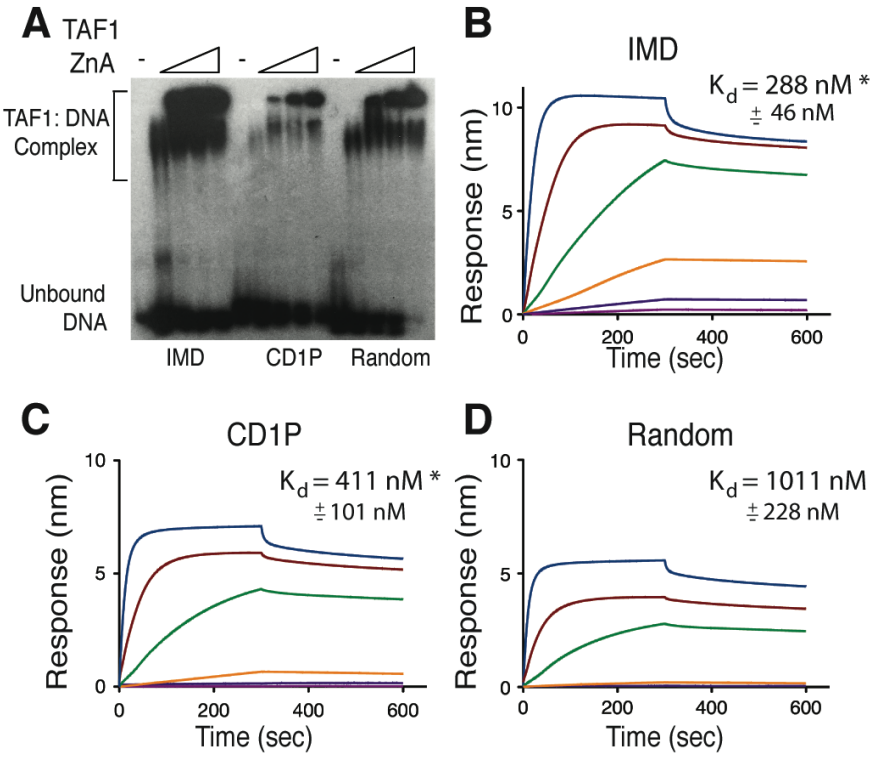
A) Linear schematic of full length human TAF1 with percent conservation in eukaryotes shown. The height and intensity of the blue bars reflect the strength of conservation. **B)** Annotated TAF1 zinc knuckle conservation alignment spanning vertebrates, insects, nematodes, plants, and fungi. A scale of red to blue letter coloring represents residue conservation with red denoting low conservation and blue highly conserved. The strictly conserved zinc knuckle cysteines and histidine are boxed in yellow. Conserved positive residues signified by the green dots. Triangles indicate cysteines subject to mutational analysis.

Figure 3 – 3: TAF1 ZnK Model Shares Biophysical Properties with other ZnKs



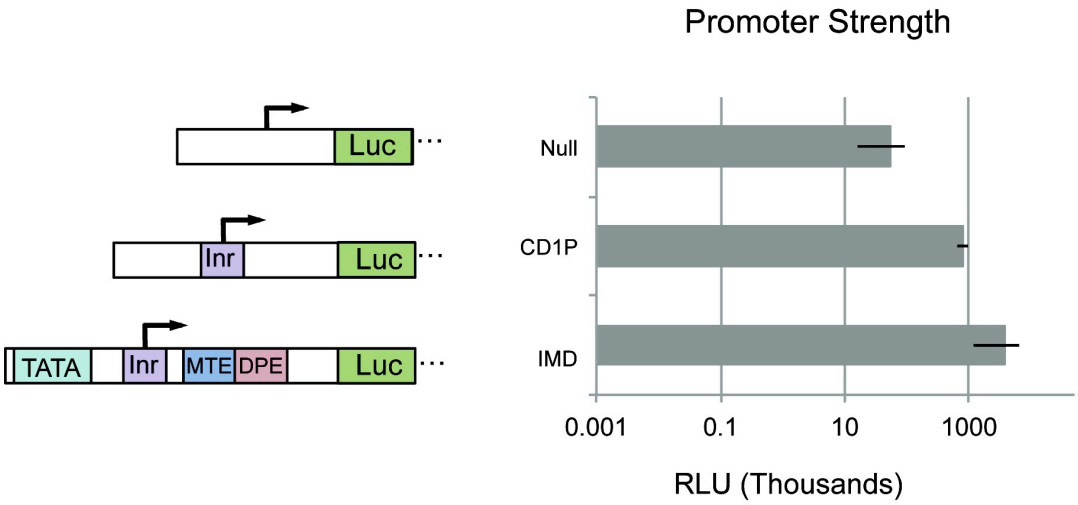
A) Sequence alignment of zinc knuckle containing proteins from eukaryotes and viruses. The scale of red to blue lettering indicates residue similarity with red being least similar and blue most similar. Dashes signify gaps and zinc knuckle cysteines and histidine are boxed in yellow. **B)** Alignment of TAF1 ZnK model (blue) from I-TASSER prediction analysis with known structure of ZnK of HIV-1 nucleocapsid protein (green) and lin28 (purple). **C)** Electrostatic surface map of lin28 (left), HIV-1 nucleocapsid protein nucleic acid binding surface (middle) and corresponding region of TAF1 ZnK (right).

Figure 3 – 4: TAF1 ZnK Preferentially Binds Promoter DNA



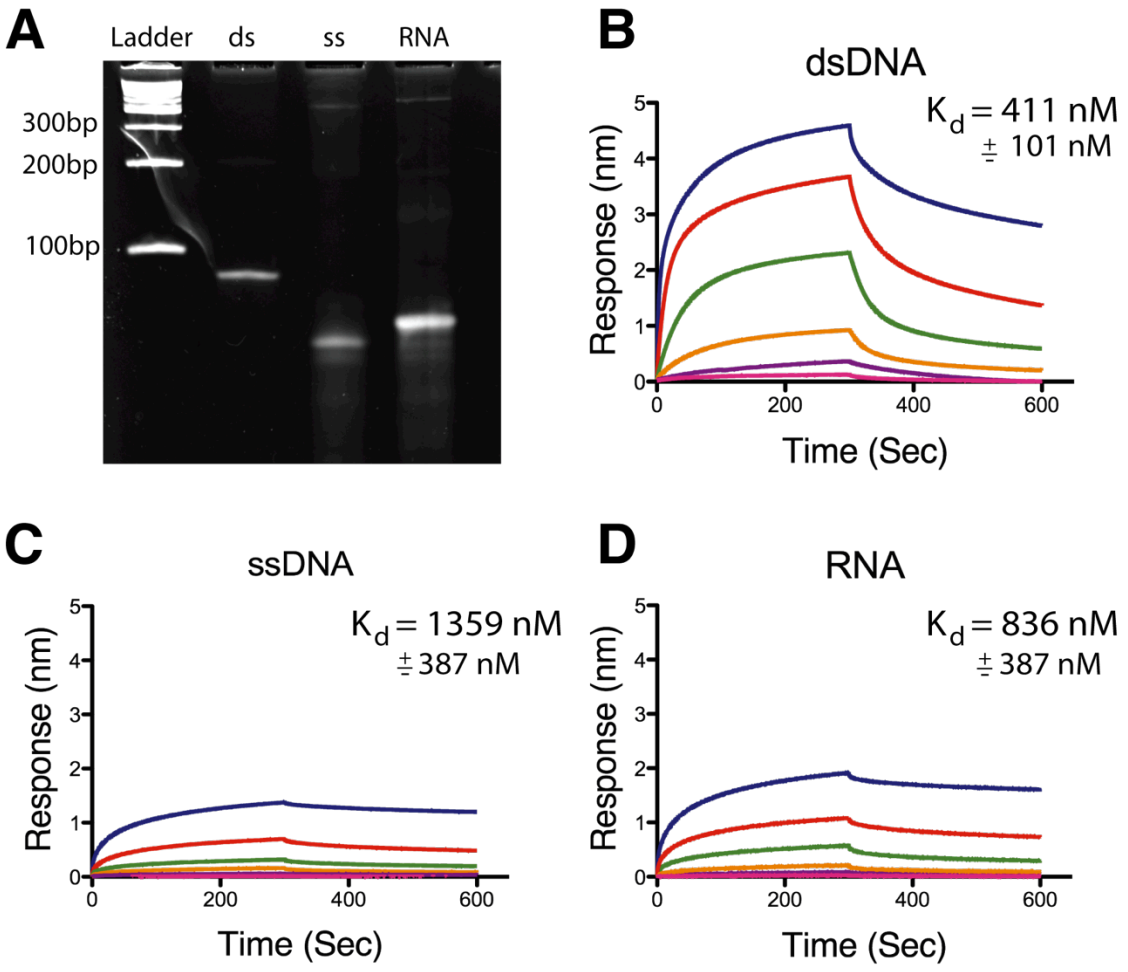
TAF1 zinc knuckle directly binds DNA. **A)** EMSA of TAF1 ZnA (aa 1234-1375) and three radiolabeled DNA fragments: IMD of super core promoter (position -6 to +38), cyclin D1 promoter (position -22 to +29, CD1P), and Random DNA sequence. Samples were separated on native 5% PAG and visualized using autoradiography. **B, C, D)** Bio-layer interferometry binding curves using biotinylated double-stranded DNA fragments described in A and the following ZnA protein concentrations: 3 μM, 1 μM, 333 nM, 111 nM, 37 nM, 12 nM. Raw data was plotted with GraphPAD Prism for one representative experiment.

Figure 3 – 5: Promoter Strength Assessment



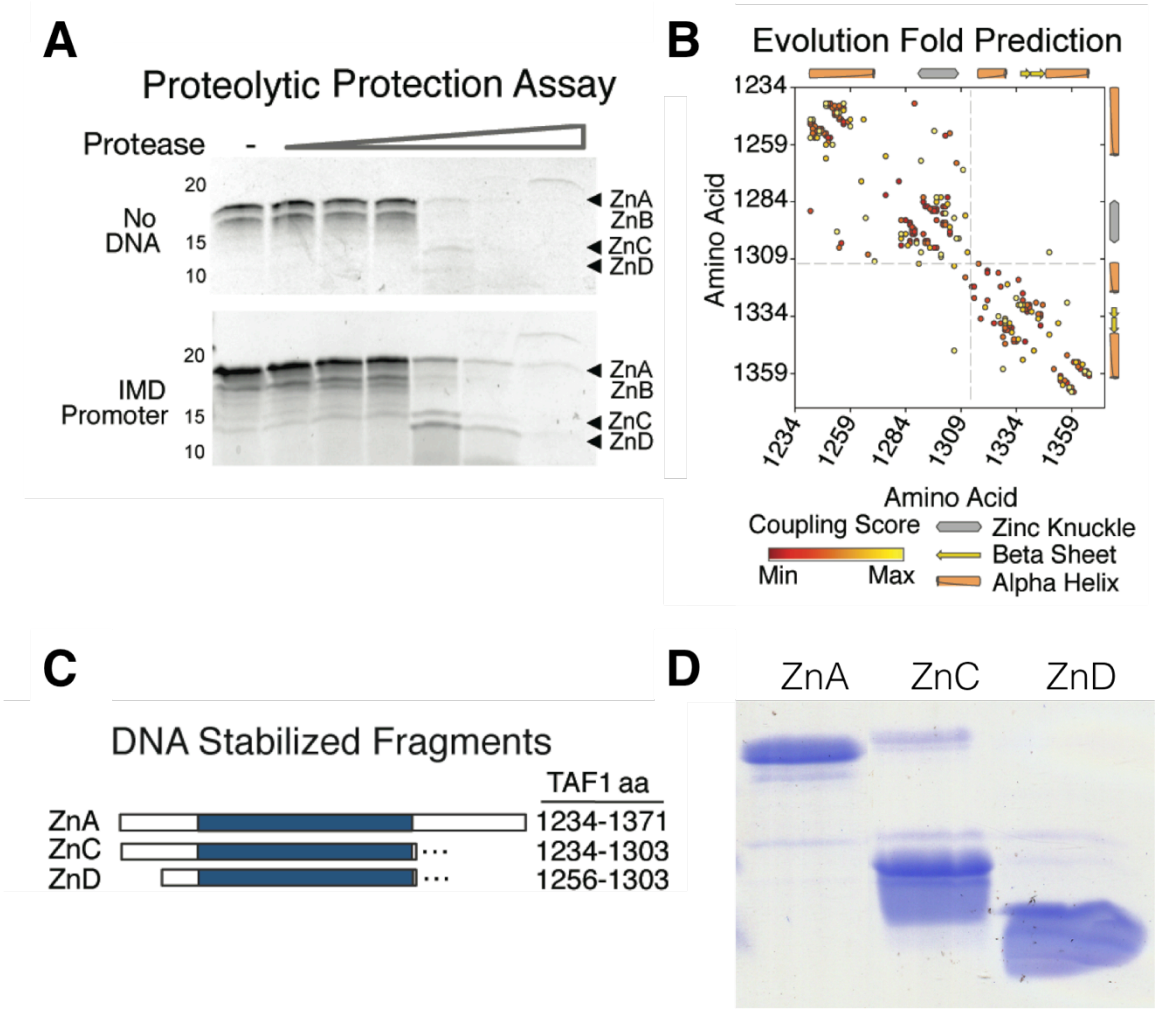
Luciferase reporter assay of pGL2 Basic (Null), cyclin D1 (CD1P), and IMD promoter driven reporter constructs. Signal expressed on a log scale in relative light units (n=4).

Figure 3 – 6: TAF1 Zinc Knuckle Preferentially Binds Double Stranded DNA



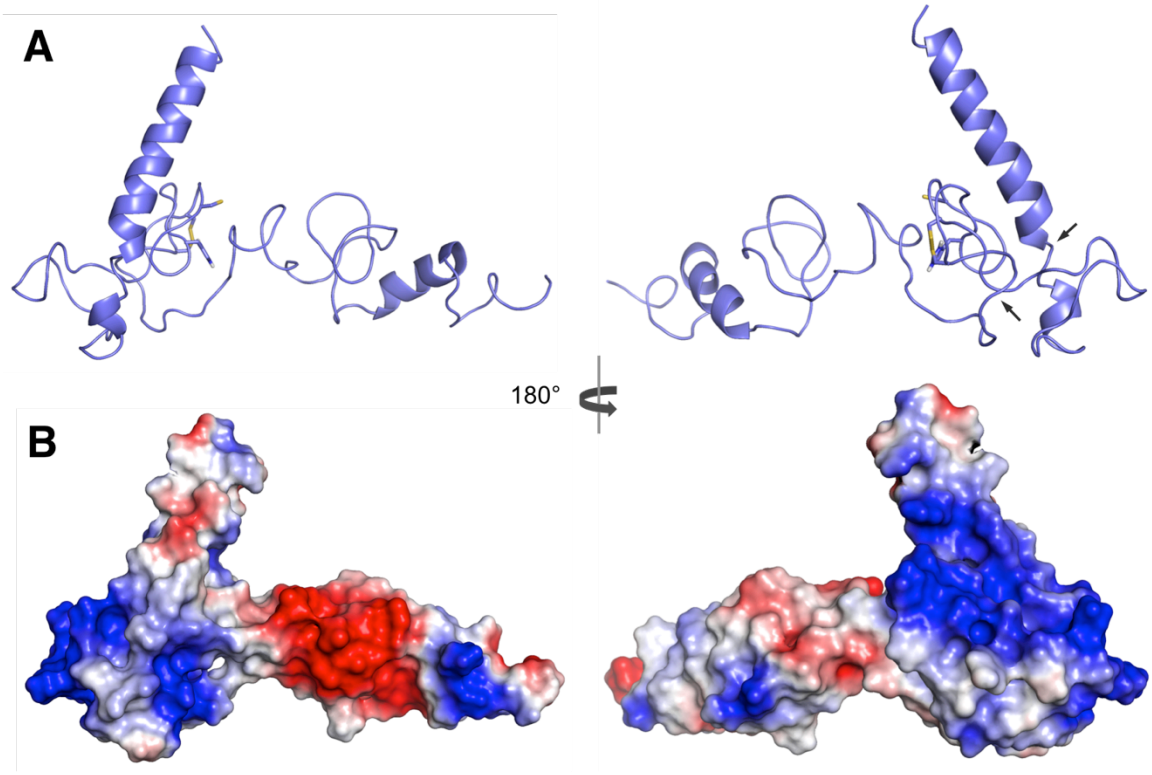
A) Nucleic acid probes separated on 12% Native PAG and stained with ethidium bromide. Bio-layer interferometry binding curves using biotinylated **B)** double-stranded DNA fragment with cyclin D1 promoter (position -22 to +29, dsDNA), **C)** single-stranded DNA with sense strand of cyclin D1 promoter (position -22 to +29, ssDNA), and **D)** single-stranded RNA fragment with sense strand of cyclin D1 promoter (position -22 to +29, RNA), at the following ZnA (aa 1234-1375) protein concentrations: 3 μM, 1 μM, 333 nM, 111 nM, 37 nM, 12 nM. Raw data was plotted with GraphPAD Prism for one representative experiment.

Figure 3 – 7: TAF1 ZnK Minimal Binding Domain



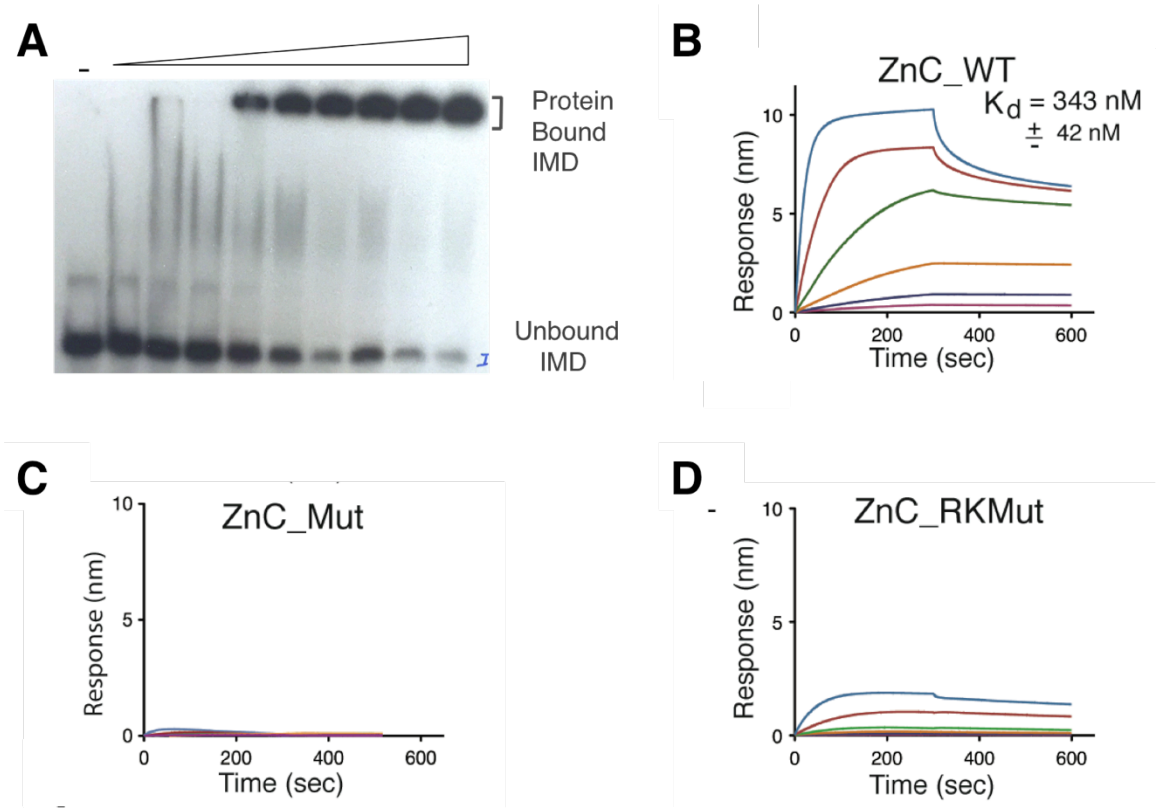
Determination of the Minimal DNA Binding Domain and Core Module of TAF1 Zinc Knuckle. **A)** TAF1 (aa 1234-1375) was incubated without (upper) and with (lower) IMD promoter DNA followed by digestion with increasing concentrations of protease. Digestion products were resolved by SDS-PAGE and detected by coomassie blue staining. Arrowheads indicate fragments stabilized by DNA. **B)** EVfold map of ZnA's top 100 interactions. The numbers correspond to amino acid residues of full-length TAF1. Outer axes display predicted secondary structure. **C)** Diagram of DNA stabilized regions of TAF1 with blue box indicating CCHC ZnK. **D)** SDS-PAGE of purified ZnA, ZnC and ZnD polypeptides using in binding assays.

Figure 3 – 8: TAF1 ZnA Structural Model



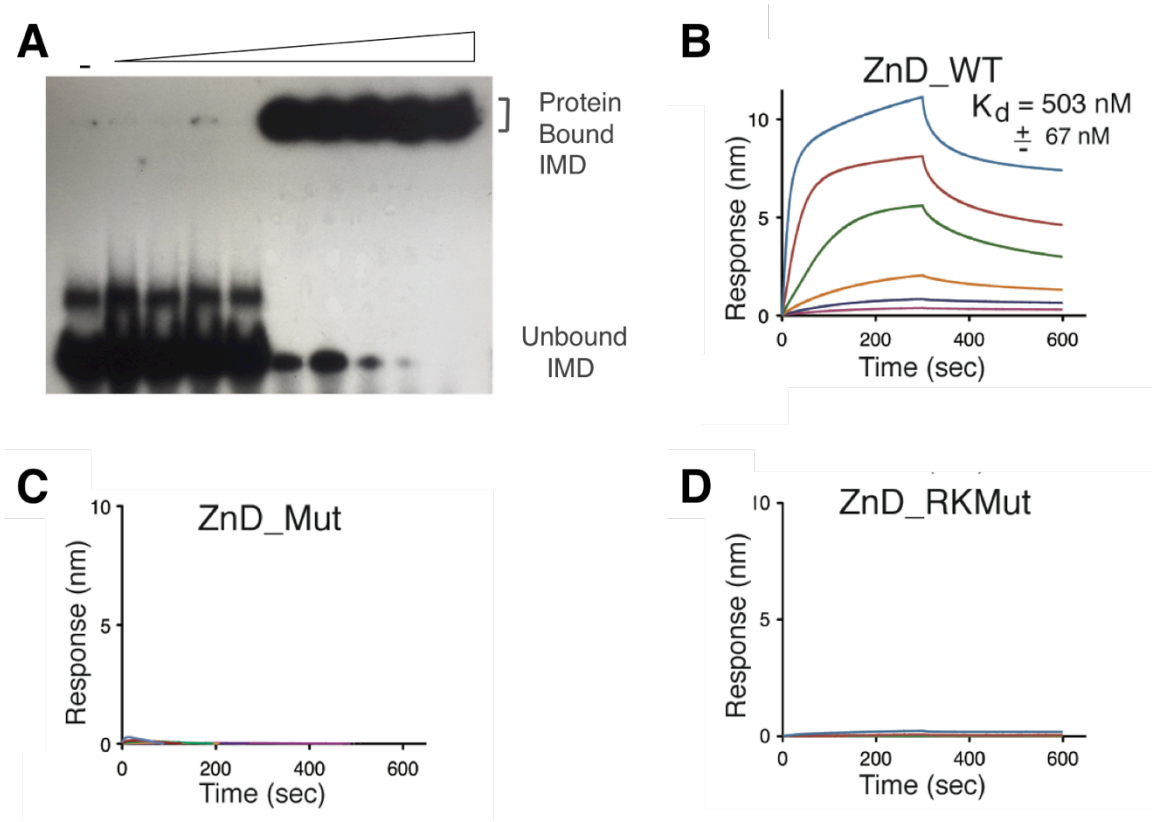
A) Ribbon representation of the predicted EVfold ZnA structure. **B)** Electrostatic map as well as a 180° rotation about a perpendicular axis. The surface colors are clamped between red ($-83.8 kTe^{-1}$) and blue ($+83.8 kTe^{-1}$). Arrows indicate protease cut sites.

Figure 3 – 9: TAF1 ZnC Binds DNA



A) EMSA of wild type ZnC.(protein shown in Fig 3-7 D). Increasing amounts of ZnC up to 1.6 µM were incubated with ^{32}P-labeled IMD core promoter fragments and were resolved on native 5% PAG. Bio-layer interferometry binding curves using biotinylated double-stranded DNA fragments with cyclin D1 promoter and **B)** ZnC wild type, **C)** ZnC cysteine mutant (C1285A and C1288A), **D)** ZnC charge mutant (R1295A and K1298A), Zn proteins titrated as follows: 3 µM, 1 µM, 333 nM, 111 nM, 37 nM, 12 nM. Raw data was plotted with GraphPAD Prism for one representative experiment.

Figure 3 – 10: The minimal TAF1 ZnD Fragment Binds DNA



A) EMSA of wild type ZnD.(protein shown in Fig 3-7 D). Increasing amounts of ZnD were incubated with ^{32}P-labeled IMD core promoter fragments and bound complexes were resolved on native 5% PAG. BLI DNA binding curves for **B)** ZnD wild type **C)** ZnD cysteine mutant (C1285A and C1288A), **D)** ZnD charge mutant (R1295A and K1298A). Zn proteins were titrated as follows: 3 µM, 1 µM, 333 nM, 111 nM, 37 nM, 12 nM. Raw data was plotted with GraphPAD Prism for one representative experiment.

CHAPTER 4:
PHYSIOLOGICAL CHARACTERIZATION OF TAF1 DNA BINDING DOMAINS

INTRODUCTION

Cell Cycle Regulation

The cell cycle is a universal process for all cells to grow and proliferation. This ordered progression is essential to create new progeny, establish cellular diversity through differentiation, and repair/replace damaged tissue. The cell cycle can be divided into two main parts: (1) the growth and development stage, called interphase, and (2) the divisional stage or mitosis (M) phase. M phase is the stage that actually separates the genetic material and cellular components to generate a new cell. However, cells spend the majority of their time in interphase, which can be broken down into three sub-phases: G1, S, and G2 (Fig. 4-1A). The G1 phase describes the basal state of the cell as it carries out normal metabolism and growth. The S or synthesis phase occurs after the cell has made the decision to divide during which DNA is replicated. The cell then moves into the G2 phase, where the cell continues to grow and produce the machinery necessary for cellular division. The transition from phase to phase is highly regulated, and checkpoints at the interface of each phase ensure the previous stage was fully completed before the progression on to the next step.

Cells initiate cellular division in response to mitogenic signals that trigger the progression from G1 to S-phase. Mitogens are chemical substances that induce signal transduction pathways and directly impact G1 checkpoint factors. The transition from G1 to S is particularly important as it commits the cell to complete a full cell division. Cyclins are key cell cycle regulatory molecules; there are four major types of cyclins (A, B, D, and E). The fluctuation in their expression patterns drives the movement through the cell cycle (Fig. 4-1B). The D type cyclins are expressed throughout the cell cycle and are essential for G1 progression. Cyclin E controls the switch from G1 into S phase. Cyclin A regulates the progression through S phase and finally cyclin B is crucial for mitosis. Cyclins bind CDKs (cyclin dependent kinases) to activate and target them to the nucleus. CDKs are a family of serine/threonine protein kinases whose phosphorylation

activity modulates the activity of their targets, including many transcription factors that impact gene expression. As cells move through G1, they experience a transcriptional wave driven by a positive-feedback loop stimulating gene activation involving a myriad of transcription factors and CDKs (Bertoli et al., 2013).

Cyclin D1 (CD1) senses growth factors and is imperative to progress through the G1 phase. During the G1-S transition, CD1 regulates the activity of CDK4 and CDK6. The interactions between CD1 and CDKs are being investigated as potential therapeutic targets (Alao, 2007, Musgrove et al., 2011). Overexpression of cyclin D1 has been shown in a variety of cancers including breast cancer and squamous cell carcinoma (Alao, 2007, Kim and Diehl, 2009). Additionally, compounds effecting the transcription of CCND1 gene would be an alternative approach to directly decrease CD1 levels or in combination to produce a multi-prong treatment plan (Musgrove et al., 2011).

TAF1's role in cell cycle

TAF1's original gene name was CCG1, cell cycle gene 1; it was discovered in a genetic screen to be important for the regulation of G1/S transition (Sekiguchi et al. 1991). Since then, the cloning and characterization of this gene revealed it as an essential member of the general transcription factor TFIID (Ruppert et al., 1993). These studies demonstrated TAF1 plays a prominent role in the expression of cyclins A, D, and E (Suzuki et al., 1997; Wang et al., 1997). TAF1's functional importance in cell cycle regulation is underscored by a single missense mutation (G716D) within the DUF3591 domain, which causes the temperature sensitive mutant hamster cell line *ts*13 to undergo a G1 cell cycle arrest (Hayashida et al., 1994; Hisatake et al., 1993; Ruppert et al., 1993). The G716 residue is strictly conserved among all TAF1 orthologs (Fig 2-4). This residue maps to the N-terminal end of the strand β3 and lies at the top of the TAF1 barrel (Fig 2-3). It is completely buried from the solvent by two α-helices in the α-helical domain, which in turn closely pack against the WH domain (Fig 2-2B). In this position, the G716D mutation will likely perturb the local structure of the TAF1 barrel and affect the topological configuration between the triple barrel and the WH domain at non-permissive temperature. We have also previously demonstrated that a deletion mutation (∆848-850) within TAF1's core domain renders the protein unable to complement the *ts*13 defects in both cell cycle progression and cyclin D1 transcription (Suzuki et al.,

1997; Hilton et al., 2005). The ts13 cell line can therefore be utilized to assess the functional importance of a particular residue or region of TAF1.

Our *in vitro* data leaves many unanswered questions about the physiological mechanisms of TAF1 WH and ZnK binding activities. Are both required for TFIID for promoter association? Will defects to the WHD or ZnK impact gene expression? This chapter explores the cellular function of the WHD and ZnK under physiological conditions. We concluded that both domains are vital for recruiting the TFIID complex to core promoters and expression of cell cycle genes *in vivo*. Furthermore, this data in combination with the *in vitro* analysis provides valuable information about TAF1's role in RNAPII transcriptional initiation.

RESULTS

TAF1 DBDs are Essential for Cell Proliferation

Given the *in vitro* biochemical data linking TAF1 WH and ZnK domains to DNA binding, we asked whether it plays a physiological function in cells. We used the *ts*13 cell proliferation assay to determine if the WHD or ZnK were required to complement the temperature-sensitive growth defect. *ts*13 cells were maintained at the permissive temperature of 33.5°C; when shifted to 39.5°C, the cells arrest and undergo apoptosis unless a functional copy of TAF1 is exogenously expressed (Schafer, 1998; Song et al., 2002). Cells transfected with wild-type TAF1 continue through the cell cycle and proliferate. Conversely, non-functional TAF1 results in a decrease in the number of viable cells. As shown in Fig. 4-2 A and B, expression of WT TAF1 reproducibly resulted in viable proliferation of *ts*13 cells at 39.5˚C, whereas the vector control completely failed to rescue the defects of the cell line. Remarkably, alanine mutation of R864/K865/K868 in the WH3A mutant severely compromised the ability of TAF1 to rescue *ts*13 cell proliferation, echoing its effects on the DNA-binding function of the TAF1 WHD.

Subsequently, we chose to mutate the first two cysteines in the ZnK domain of TAF1 to ensure the motif was fully compromised (Fig 3-2B). Changing the cysteines to alanines resulted in a ~60% reduction in cell viability, indicating a structurally intact ZnK is vital for full physiological function (Figure 4-2A, B). The level of cell survival was

comparable to point mutations in the WHD. Moreover, combining the ZnK and WH mutations (3AZnM) tended to further reduce cell viability compared to the single mutants, but the decrease was not statistically significant. No visible change in *ts*13 cell growth was observed at the permissive temperature 33.5˚C in the presence of WT-TAF1 or mutants (data not shown). Importantly, all TAF1 constructs were expressed at equivalent levels according to western blot analysis; therefore, the mutants do not cause the protein to become unstable and degraded (Fig 4-2C). Together, these data indicate that the normal function of TAF1 in transcription regulation depends on the activity of both the WHD and ZnK.

Zinc knuckle and winged helix of TAF1 are necessary for cyclin gene expression

TAF1 is known to be important for transcription of the cell cycle genes cyclin A, D1 and E. The failure of ZnM and WH3A to complement the growth defect in *ts13* cells signifies that these mutations possibly compromise a function of TAF1 important for cyclin gene expression. We determined that this functional defect is not due to an inability to incorporate into TFIID, but rather a deficiency in promoter recognition. We verified the TAF1 mutants can successfully integrate into the TFIID complex by expressing each TAF1 variant as an N-terminal HA-fusion protein (HA-TAF1), immunoprecipitating TFIID complexes with an antibody against the TAF4 subunit, and immunoblotting for exogenously expressed TAF1 variants using an anti-HA antibody. TBP was used as experimental control. We observed that all TAF1 variants incorporated at relatively equal levels compared to the wild-type protein (Fig 4-3A). Conversely, the TAF1 mutants failed to stimulate transcription from the cyclin D1 and A2 promoters.

By leveraging a luciferase reporter assay, a conventional method for studying gene expression, we asked whether exogenously expressed TAF1 variants affected cyclin promoter function. The luciferase gene was placed under the control of the human cyclin D1 or cyclin A2 promoter. TAF1 variants were co-transfected with each luciferase reporter construct, and the amount of luciferase activity, a measure of promoter activity, was assessed approximately 36 hours post-transfection. Reporter activity in the absence of exogenously expressed TAF1 was quantified to assess the endogenous transcriptional activity of the cyclin promoters. This signal was used to normalize across experiments and served as a control ensuring transcription was not

negatively impacted by expressing a single TFIID subunit. A nominal increase in TAF1 should enhance TFIID formation because under normal conditions TAF1 is in limited supply compared to other TFIID components (Wang et al., 2012). Indeed, the expression of wild type TAF1 significantly increased the level of luciferase activity from both promoters when compared to basal levels (Fig 4-3B). In contrast, TAF1 with mutations to either the WHD or ZnK failed to significantly stimulate the activity of the cyclin promoters above background levels. Similarly, the double mutant had no effect on cyclin promoter activity (Fig 4-3B). Hence, there was no statistical difference between all TAF1 mutants and background levels. We consistently saw the transcription levels with the double mutant equivalent to those observed with vector while the single mutants maintained a measurable level of transcription above vector suggesting an ability to maintain some low levels of residual activity. Collectively, these data suggest WHD and ZnK are critical for TFIID to participate in transcriptional initiation.

TFIID promoter occupancy depends on TAF1 DBD

Next, we determined if the transcriptional impact was due to a loss of promoter recognition at the cyclin D1 and A2 promoters. The WHD and ZnK variants did not associate with core promoters of cell cycle genes in chromatin immunoprecipitation experiments (Figure 4-4B). For these studies, different HA-TAF1 mutants were expressed in HEK293 cells and recovered TAF1-bound promoter fragments were quantified by qPCR. Cyclin D1 and cyclin A2 were chosen as target promoters because the proper expression of these genes is required for cell cycle progression; their promoter function also has been shown to be dependent on TAF1 activity. The specific promoter regions for qPCR amplification were selected based on ENCODE ChIP data for TAF1 (Fig 4-4A). Additionally, TAF1 promoter binding is considered synonymous with canonical TFIID binding, because TAF1 is unique to TFIID whereas other TAFs and TBP are found in additional auxiliary transcription regulatory complexes (Thomas and Chaing, 2006).

We observed that wild type TAF1 effectively bound to cyclin D1 and A2 promoter regions, while neither the ZnM nor WH3A mutant selectively enriched for these promoter fragments (Fig 4-4B). The double mutant (3AZnM) appeared to be even less efficient at associating with promoter sequences than either single domain mutant. The ChIP data

analysis was complicated by what appears to be a loss in DNA binding specificity by the mutant proteins. This was reflected in the amplification of non-specific products in mutant samples analogous to the background signal of control samples (Fig 4-4C). We observed that the target product signal decreased as the amount of non-specific signal increased, thereby artificially inflating the quantitation of desired targets in mutant samples. The overall conclusion is that WH3A and ZnM mutants consistently bind less efficiently to both cyclin genes, which indicates ZnK and WH can contribute equally to TFIID promoter recognition.

CONCLUSION

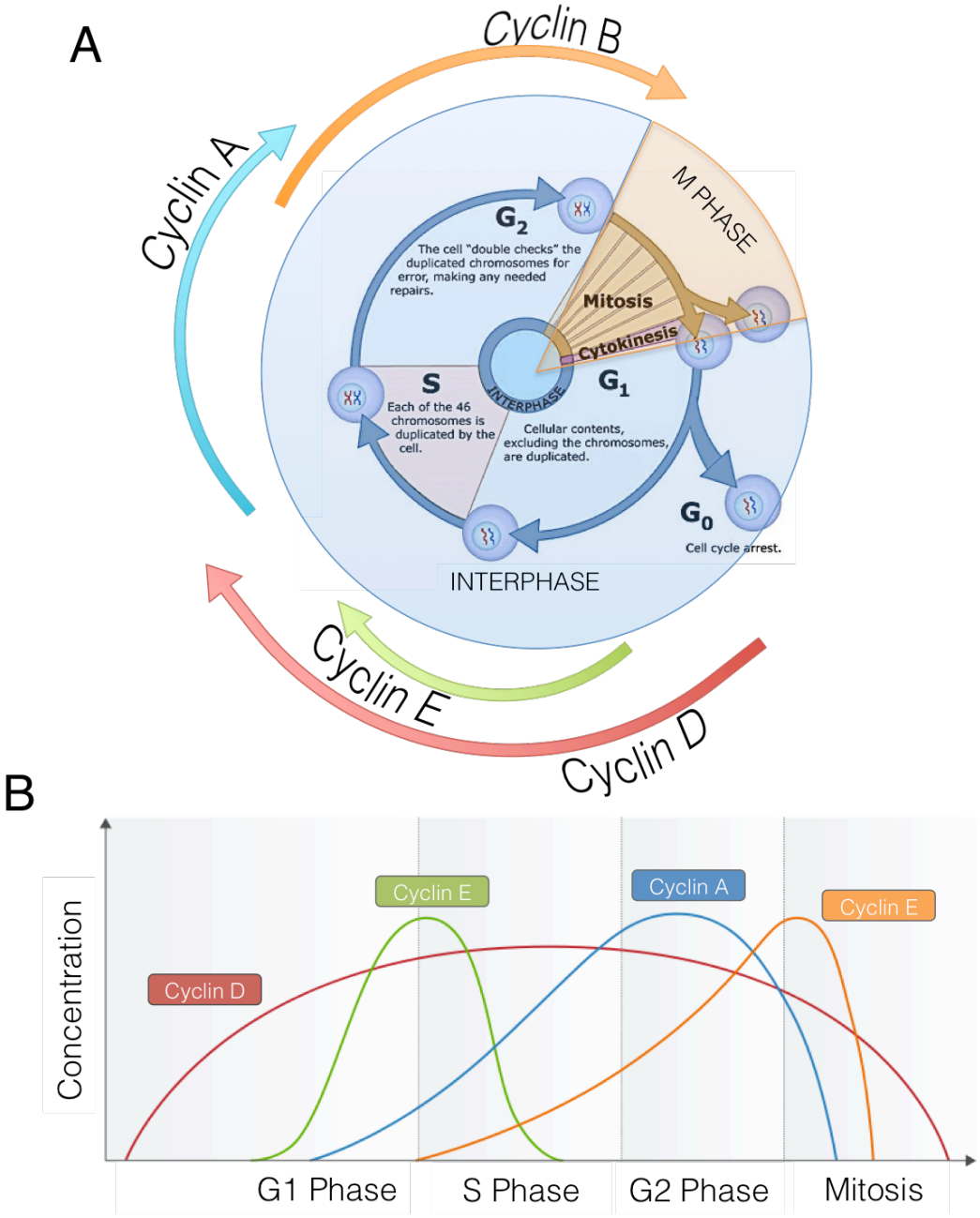
Cyclins are master regulators of the cell cycle. Their phasic expression mediates enzymatic activities required for growth and proliferation (Song et al., 2002). The impact of the TAF1 WHD and ZnK mutations on cyclin transcription is not due to an inability to incorporate into TFIID, implying these domains are not involved in TAF interactions or the core structure of TAF1, both are important for TFIID function. While this study focuses on the ability of WHD and ZnK to bind nucleic acids, our findings do not exclude the possibility they may interact with other regulatory proteins. The close proximity of the ZnK to the double bromodomain suggests the ZnK could play a role in facilitating the recognition of epigenetic markings, which could influence promoter association (Van Nuland et al., 2013). There is growing evidence that zinc fingers can play a dual role: supporting DNA binding and facilitating protein interactions with histones or transcriptional activators, such as those found in CBP and ZYMND11 (Song et al., 2002; Wen et al. 2014; Wang et al., 2014; Li et al., 2016). In this same fashion, TAF1's WHD may also interact with proteins; many WHD are molecular adaptors to facilitate protein-protein interaction (Aravind et al., 2004). For example, ESA1, a yeast histone acetyltransferase, contains a WHD thought to mediate protein-protein interactions, which is structurally similar to TAF1's WHD (Fig 2-6A). Further exploration will be required to determine if TAF1 DBDs is capable of these additional activities.

The work presented in this chapter further expands upon the biochemical analysis to characterize the WH and ZnK DBDs. Importantly, this work also established a strong functional correlation between promoter DNA binding and the normal function of TAF1 in transcription and cell cycle regulation and emphasize the challenge of

dissecting the mechanism of specific domains in a dynamic macromolecular complex. We further demonstrate the WHD and ZnK are important for cellular proliferation. Moreover, the effect of G716D mutation in the *ts*13 mutant further suggests that the intact structural coupling between the WHD and the triple barrel might also be crucial for TAF1 function. Interestingly, the double mutant (3AZnM) did not have a dominant negative phenotype; rather combining mutations only slightly but not significantly decreased cell proliferation. This implies both domains are equally important for the proper function of TAF1 in cells and do not have a synergistic effect. Mutations in either the WHD or ZnK lead to a reduction in TFIID association with cyclin D1 and A2 promoter sequences, as shown by a decrease in TAF1 promoter binding by ChIP and a correlative reduction in luciferase reporter activity assays. The discovery of these DNA binding modules in human TAF1 is a significant step forward for understanding transcriptional initiation. TAF1 DNA binding activity is a vital yet under studied aspect of transcription. Elucidating TAF1's role in core promoter recognition will have a profound impact on our perception of transcriptional regulation and how we may be able to manipulate this activity to benefit human health.

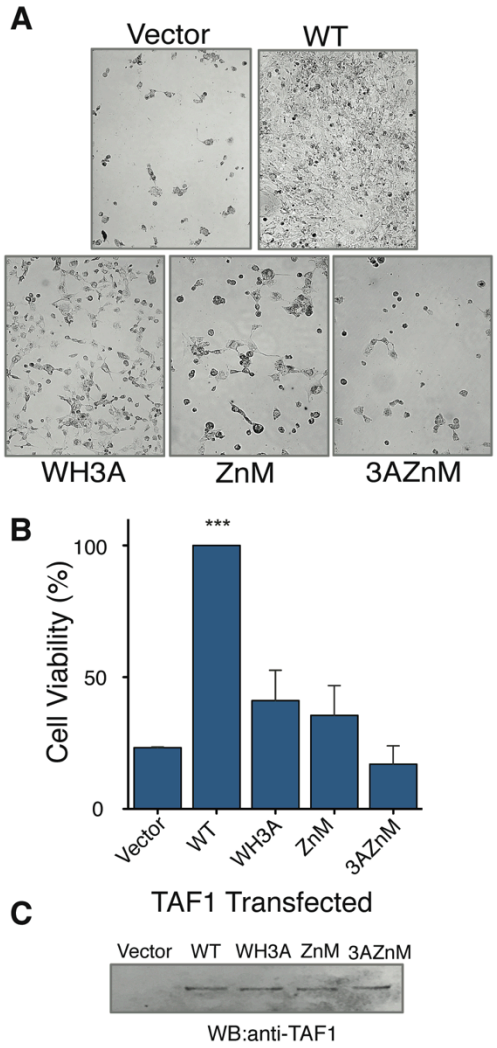
FIGURES

Figure 4 – 1: Cyclins are Master Regulators of Cell Cycle Progression



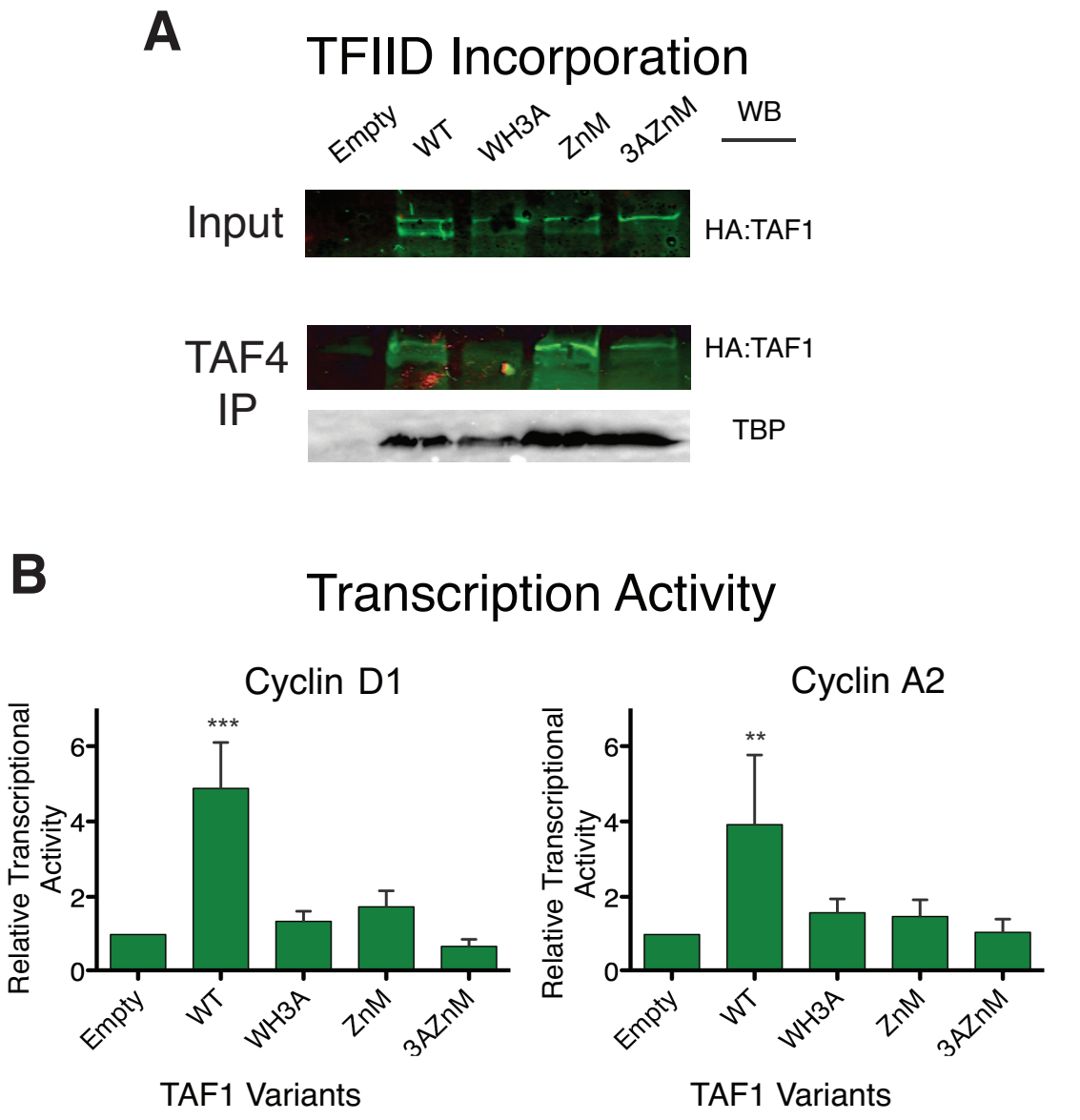
A) Circular illustration of the stages of the cell cycle divided into the two major intervals: interphase and M phase. Interphase is further broken down into G1, S, and G2 phases. The outer arrows depicts cyclin expression. **B)** A linear representation of cyclin expression showing the phasic levels throughout the cell cycle.

Figure 4 – 2: TAF1 Zinc Knuckle is Required for Cell Viability



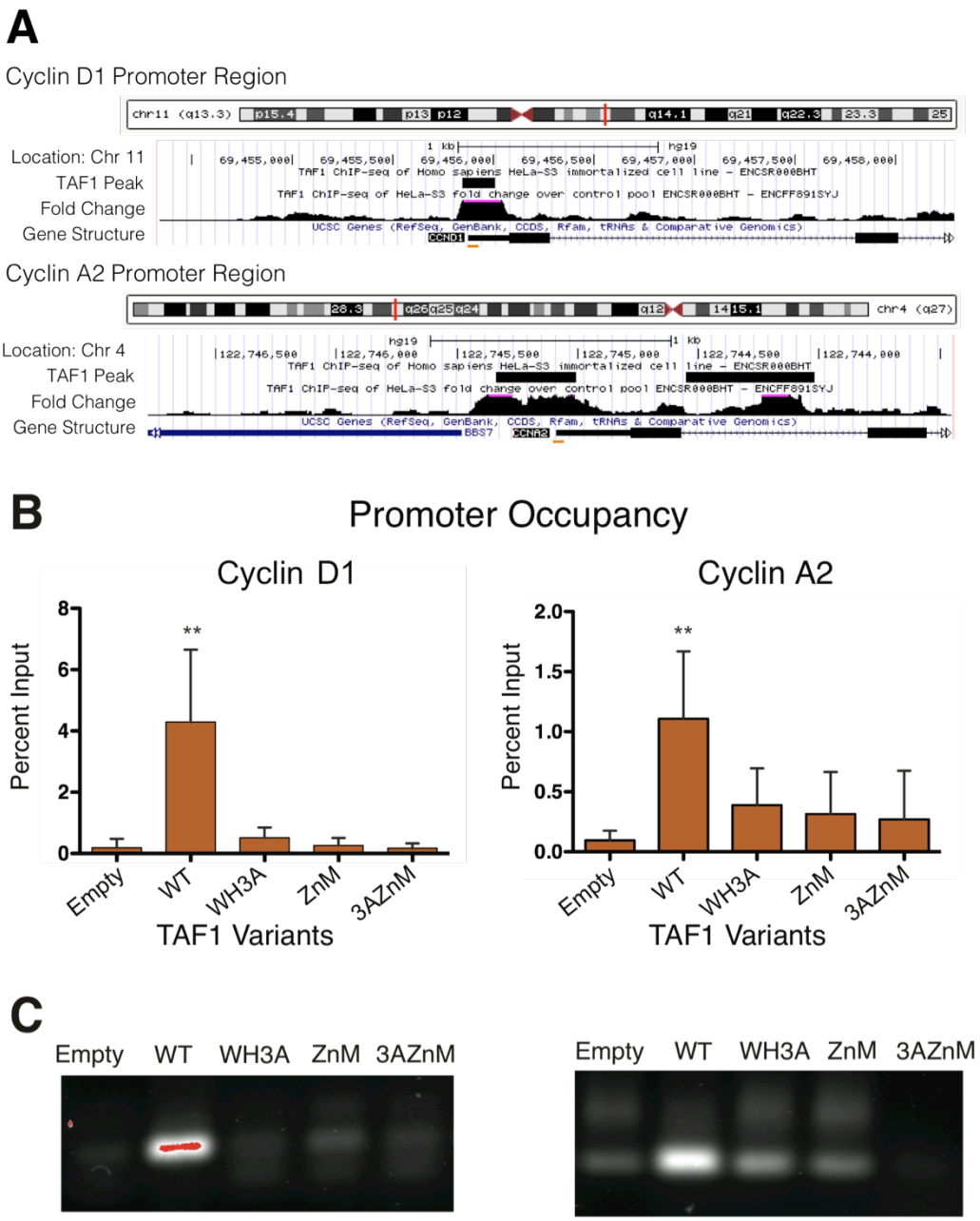
A) Phase contrast images of ts13 cells transfected with pCS2+ (vector), wild type TAF1 (WT), winged-helix mutant (WH3A R864/K865/K868), zinc knuckle mutant containing two cysteine to alanine mutations (ZnM), and double mutant (3AZnM) at 39°C for 72hrs. **B)** Quantitation of viable DAPI stained cells expressed as a percent relative to cells transfected with WT-TAF1, given a value of 100%. The bar graph depicts the average from 3 independent complementation experiments. Error bars represent standard deviation. Two-tailed analysis compared to vector with a 95% confidence, ***$p < 0.0001$ (Unpaired t test). **C)** Western blot of ts13 lysates expressing HA-tagged TAF1 proteins. Proteins were immunoprecipitated with anti-HA antibody and immunoblotted for TAF1 using TAF1 specific double bromodomain antibody.

Figure 4 – 3: Mutated TAF1 DBDs Incorporate into TFIID but Fail to Stimulate Transcription



A) Incorporation of TAF1 variant into TFIID. Input nuclear extracts were blotted to visualize starting levels of HA-TAF1 expression. TFIID complexes were isolated by immunoprecipitation using anti-TAF4 antibody and incorporated HA-TAF1 variants detected by anti-HA immunoblotting. Blots were striped and reprobed for TBP as a loading control. **B)** Luciferase assay of cyclin D1 and cyclin A2 promoter driven reporter constructs co-transfected with TAF1 variants. Luciferase activity was normalized for total protein and expressed relative to reporter activity without exogenous TAF1 (Empty), given a value of 1.0 (n=3). All error bars represent standard deviation. Two-tailed analysis compared to empty vector with a 95% confidence, **p ≤ 0.007, ***p = 0.0001(Unpaired t test).

Figure 4 – 4: TAF1 DBDs are Necessary for Promoter Occupancy



A) Encode Promoter Maps for cyclin D1 and A2 in HEK293 cells. Orange dash indicates the primer location. **B)** Chromatin immunoprecipitation of HA-TAF1 variants expressed in HEK293 cells followed by qPCR for cyclin D1 and cyclin A2 promoters (n=4). Recovery was expressed as percent of input. All error bars represent standard deviation. Two-tailed analysis compared to empty vector with a 95% confidence, **p ≤ 0.007, ***p = 0.0001(Unpaired t test). **C)** Products visualized on agarose gel and stained with EtBr show a decrease in specific product for TAF1 mutants.

CHAPTER 5:
CONCLUSION AND FUTURE WORK

TAF1's DBDs Role in General Transcription

TFIID promoter recognition is an essential step in transcriptional regulation, yet we are only beginning to untangle the complexities of its mechanism. Our data offer several sites of contact in TAF1 that can contribute to the stabilization of TFIID at RNAPII promoters during PIC formation. The dynamic rearrangement of TFIID, as it establishes a stable association, might take on several conformational states to properly position and load GTFs onto core promoters. As TFIID converts from one-state to another, different DNA binding modules may become exposed and potentially engage with DNA while the complex transitions from its canonical form to a configuration capable of supporting transcription. Alternatively, multiple DBDs may engage simultaneously to provide TFIID with several points of interaction to secure promoter association. Multiple DBDs could also convey onto TFIID the necessary flexibility to associate with a variety of core promoter sequences, such that all DNA binding domains may not be required at every promoter. With these possibilities in mind, we propose two models for how TAF1 contributes to TFIID promoter recognition (Fig 5-1), either simultaneous binding where the WH and ZnK equally contribute to DNA binding or sequential binding where one DNA binding domain engages during the initial stages followed by a second DNA binding event in the rearranged conformation.

Current data in the field seems to favor sequential binding as a mechanism for TFIID promoter recognition. Notably, TAF7 has been shown to dissociate from TFIID, which promotes transcription initiation (Chiang and Roeder, 1995; Lavigne et al., 1996). This dissociation event could release WH binding from promoter DNA allowing RNAPII engagement and polymerase escape. Moreover, as the transcription bubble forms, TFIID contacts the initiator, as well as downstream CPEs, which could further disrupt WH association. Yet, TFIID is the only GTF to remain at the core promoter after initiation indicating new contacts must be formed to retain TFIID and facilitate multiple rounds of transcription (Zawel et al., 1995). Interestingly, this model is consistent with

TAF1 regulating both the magnitude and duration of transcription (Pennington et al., 2013). Sequential binding is further supported by the recent discovery of two modes of transcriptional initiation: the first round and reinitiation (Zhang et al., 2015). The data revealed there are two TFIID promoter bound states: (1) a form that first recognizes DNA for entry into the initial transcription event and (2) a reorganized form able to support subsequent rounds of transcription. Structural studies affirm reorganization is required for a fully functional PIC (Louder et al., 2016). Overlaying the structure of promoter bound TFIID with other PIC structures has exposed potential steric hindrance between TAF1 and RNAPII near the transcriptional start site (Louder et al., 2016). This may explain why all the structural studies of PICs to date lack the TFIID complex (Plaschka et al., 2016). Future studies that use advanced high-resolution techniques will be needed to unravel the details of this elusive aspect of PIC assembly, as the current views may underestimate the flexibility and complexity of TFIID promoter recognition.

While our analysis of the cyclin genes shows both WH and ZnK domains as being equally important for promoter recognition, this may not be universal for all RNAPII genes. Variable dependence on TAF1 has been suggested in previous work (Hilton et al., 2003, Devaiah et al., 2010). This introduces the idea that TAF1 could play a passive supporting role in TFIID PIC assembly at some genes and a driving role presumed for cell cycle genes. Our preliminary data examining the c-Fos gene argues TAF1 function is not ubiquitous. Transcriptional reporter assays using the c-Fos promoter, which contains a TATA box, illustrated TAF1 overexpression does not universally stimulate all RNAPII gene promoters (Fig 5-2B). Moreover, ChIP data revealed WT TAF1 was not significantly enriched at the c-Fos promoter (Fig 5-2C). The TAF1 signal detected in the ENCODE experiments maybe tethered to promoters through TBP binding (Fig 5-2A). There are several techniques that could be employed to delineate any differences in DBD promoter dependence. A straightforward method would be to sequence the DNA associated with the isolated mutant TAF1 to find promoter sequences bound by one over the other domain. This could also give insight into the selectivity of each domain, which could then be used to identify specific genes or gene classes preferentially regulated by an individual TAF1 DBD.

While the focus of my work centered on the DNA binding abilities of the WHD and ZnK, there is still the possibility these domains are dual binding surfaces and can

bind to other biomolecules. It is well documented WHDs and ZnKs have the capacity to interact with proteins, as well as lipids, in addition to nucleic acids (Matthews et al., 2000, Teichmann et al., 2012). TAF1's WHD and ZnK DNA binding activity may not be mutually exclusive with other biomolecules. There are instances of ZnK domains having a dual function and interacting with a variety of biomolecules under different contexts. TAF1's DBDs may also possess multifunctional interacting motifs. For example, the ZnK could interact with histone proteins in the absences of a suitable DNA sequence. The negatively charged patch created between H2A/H2B has been shown to interact with the Sgf11 zinc finger (Morgan et al., 2016). The TAF1 ZnK could facilitate proper positioning of the 2xBromo towards acetylated histone tails. Intriguingly, TFIID can associate with silenced mitotic chromatin, acting as an inheritable epigenetic marker (Chen et al., 2002; Christova and Oelgeschlager, 2002; Segil et al., 1996), conceivably through protein-protein interactions. Carefully controlled competition assays or crosslinking studies could aid in defining any potential interactions. In any case, investigating additional WHD/ZnK binding partners would be a worthwhile avenue to pursue.

Future Applications

The discovery of TAF1's DBDs has the potential to be a valuable building block for a variety of research applications. The first interesting and perhaps far reaching future application could be a research tool that denotes active genes. Engineered DBDs with a fluorescent tag, modeled after TAF1 WHD or ZnK, could seek out RNAPII core promoters and then be monitored by microscopy. Since the WHD and ZnK regions are highly conserved, the TAF1 synthetic transcription marker might have the capability to recognize RNAPII genes in a variety of model systems. This approach could detect when are genes turned on and off in response to stimuli or as a result of developmental cues. These small promoter specific marks could be used to survey different cell lines, disease states, creating maps of transcription. By using sub-domains of a native general transcription factor, researchers may even be able to watch in real-time the activation and deactivation of genes.

Another application could be the engineering of transcriptional inhibitors. Work has already begun to identify the potential of expressing just the ZnK in cells and ask whether this small domain impacts transcription. The idea is that the ZnK could bind to

promoters preventing TFIID binding and inhibiting PIC assembly at transcriptional start sites. Proof of principle experiments are already underway; the expression of TAF1 ZnK alone appears to cause cell death whereas expression of DUF domain regions do not have the same effect (data not shown). Follow up experiments are needed to confirm the activation of apoptotic pathways, and several key obstacles will have to be overcome including controlling for protein expression and normalizing for gene expression to validate the effects.

Disease Implications

The mechanisms outlined in this work also provide insight into the pathologies of human disease states associated with TAF1 mutations by GWAS. The cBioPortal and catalog of somatic mutations in cancer (COSMIC) databases denote several mutations to the conserved CCHC residues of TAF1 including: C1285R, C1288R, H1293N, C1300W/R. Exome sequencing of uterine serous carcinomas also revealed missense mutations at R864 in the WHD. These findings reinforce the notion that these residues are crucial for proper TAF1 function in cells. The essential nature of TAF1 for cell viability leads to the belief that it is doubtful disease-associated mutations completely obliterate TAF1 function. Loss of function mutations can cause global transcriptional dysfunction and cell death. More likely, TAF1 function is altered in some subtle manner that results in aberrant transcription from select genes. TFIID's role in proliferation and cell survival makes it a particularly attractive target for novel drug design in cancer treatment (Ribeiro et al., 2014; Zhao et al., 2013).

In addition to understanding the consequences of mutations in the TAF1 protein, our data also has the potential to help explain how mutations that map to promoter regions could result in improper TAF1/TFIID binding and PIC positioning, causing altered gene expression (De Vooght et al., 2009). For example, mutations in the human beta-globin promoter, which contains a non-canonical TATA sequence, lead to beta thalassemia, resulting in reduced hemoglobin production (Cao and Galanello, 2010). One hypothesis suggests this blood disorder arises from a transcription mediated decrease in globin gene products, possibly because TAF1 and thereby TFIID cannot effectively bind to the mutated promoter (Lee et al., 2005). As genome sequencing becomes a standard practice in diagnostics, the functional effects of genetic variations in

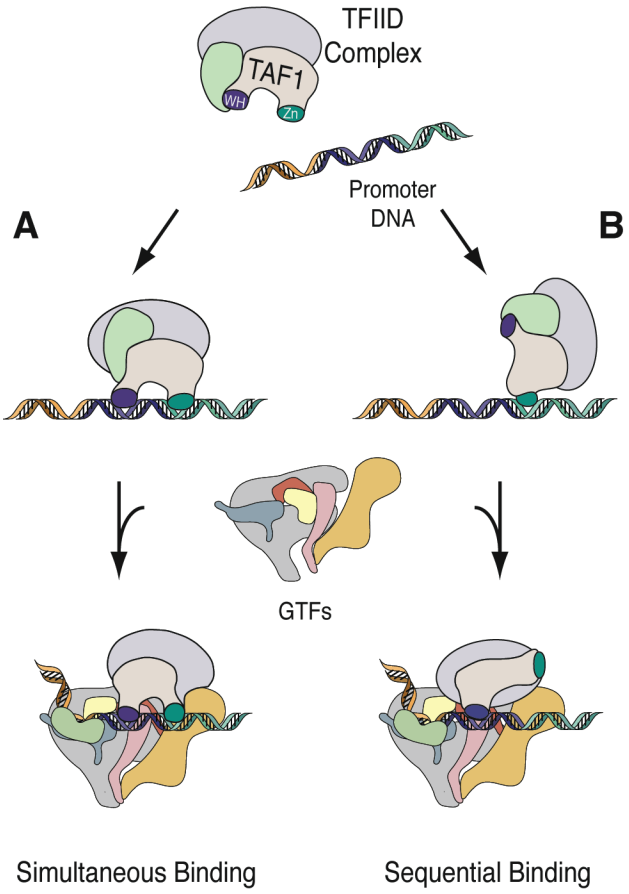
TAF1 or TFIID dependent promoters will become increasingly important for patient care and drug development.

Much of what we know about mammalian gene expression was discovered from experiments using viral promoters because they could be easily manipulated ex vivo and studied for transcriptional impacts. This knowledge also provides insight into how viruses can hijack the mammalian transcriptional machinery during infection. Therefore, developing chemical agents to augment transcription could combat viral pestilence. The hepatitis delta virus was used as a model to confirm the RNAPII PIC is recruited to viral promoters to activate transcription (Abrahem and Pelchat, 2008). With this in mind chemical agents could be designed to prevent such recruitment. A compound identified by Zhang et al., 2016 could inhibit new transcription while leaving actively transcribed genes unaffected. This concept is a potential starting point to hinder viral production.

In summary, we uncovered two critical DNA binding domains in TAF1, the largest subunit of the TFIID general transcription factor. The conserved zinc knuckle and winged helix were both essential for the *in vivo* function of TAF1. Mutations within this region do not disrupt TAF1 incorporation into TFIID, but instead interfere with its ability to associate with core promoters. The reduction in promoter association resulted in inefficient transcription initiation. We speculate that the two motifs in TAF1 provide greater flexibility within TFIID for promoter DNA sequence recognition.

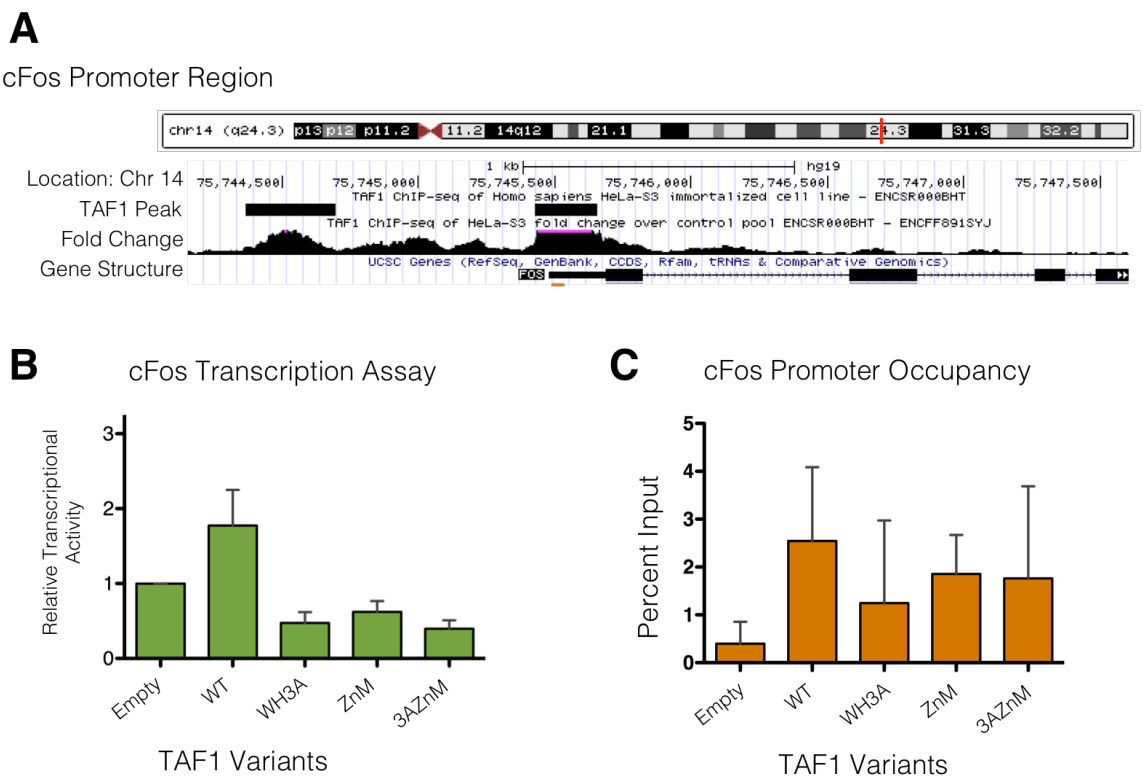
FIGURES

Figure 5 – 1: Model of TAF1's role in Promoter Recognition at Cell Cycle Genes



A) ZnK and WH Bind Simultaneously B) Sequential Binding of ZnK and WH

Figure 5 – 2: TAF1 Independent RNAPII Gene Expression



A) ENCODE Promoter Maps for c-Fos promoter region. Orange dash indicates the primer location. **B)** Luciferase assay of c-Fos promoter driven reporter construct co-transfected with TAF1 variants. Luciferase activity was normalized for total protein and expressed relative to reporter activity without exogenous TAF1 (Empty), given a value of 1.0 (n=3). **C)** Chromatin immunoprecipitation of HA-TAF1 variants expressed in HEK293 cells followed by qPCR for c-Fos promoter (n=3). Recovery was expressed as percent of input. All error bars represent standard deviation. Statistical analysis shows no significance between samples.

DISCLOSURES

Much of the data and text in these chapters are published in the following manuscripts:

Wang, H., Curran, E., Hinds, T., Wang, E., and Zheng, N. (2014). Crystal Structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. Cell Res. 24(12): 1433-1444.

Curran, E., Wang, H., Hinds, T., Wang, E., and Zheng, N. (2017). Zinc knuckle of TAF1 is a DNA binding module critical for TFIID promoter occupancy. *In Review*

REFERENCES

1. Abad, P., Vaury, C., Pelisson, A., Chaboissier, M., Busseau, I., and Bucheton, A. (1989). A Long Interspersed Repetitive Element—the I Factor of Drosophila teissieri—is Able to Transpose in Different Drosphila Species. PNAS. 86(22):8887-8891.

2. Abd El-Wahab, E., Smyth, R., Mailler, E., Bernacchi, S., Vivet-Boudou, V., Hijnen, M., Jossinet, F., Mak, J., Paillart, J., and Marquet, R. (2014). Specific recognition of the HIV-1 genomic RNA by the Gag precursor. *Nature Communications* 5, Article number: 4304

3. Abrahem, A and Pelchat, M. (2008). Formation of an RNA polymerase preinitiation complex on an RNA promoter derived from the hepatitis *delta* virus RNA genome. Nucleic Acids Res. 36(16): 5201–5211.

4. Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. *58*, 1948-1954.

5. Alao, J. (2007). The regulation of cyclin D1 degradation: roles in cancer development and the potential for therapeutic invention. Mol Cancer. 6: 24.

6. Anandapadamanaban, M., Andresen, C., Helander, S., Ohyama, Y., Siponen, M., Lundstrom, P., Kokubo, T., Ikura, M., Moche, M., and Sunnerhagen, M. (2013). High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation. Nat Struct Mol Biol 20(8): 1008-1014.

7. Aravind, L., Anantharaman, V., Balaji, S., Babu, M., and Iyer, L. (2004). The many faces of the helix-turn-helix domain: Transcription regulation and beyond. FEMS. 29:231-262.

8. Armas, P., and Calcaterra, N. (2013). Retroviral Zinc Knuckles in Eukaryotic Cellular Proteins. Nova Science Publishers. 51-80.

9. Bertoli, C., Skotheim, J., and Bruin, R. (2013). Control of cell cycle transcription during G1 and S phases. Nature Reviews Molecular Cell Biology. 14: 518-528.

10. Bhattacharyaa, S., Loua, X., Hwangc, P., Rajashankard, K., Wange, X., Gustafssonb, J. A., Fletterickc, R., Jacobsone, R., and Webb. P. (2014). Structural and functional insight into TAF1–TAF7, a subcomplex of transcription factor II D. PNAS. 111(25): 9103–9108.

11. Bosch, N., Cáceres, M., Cardone, M. F., Carreras, A., Ballana, E., Rocchi, M., Armengol, L., and Estivill, X. (2007). Characterization and evolution of the novel gene family *FAM90A* in primates originated by multiple duplication and rearrangement events. Hum Mol Genet. 16(21): 2572-2582.

12. Bradner, J., Hinisz, D., and Young, R. (2017). Transcriptional Addiction in Cancer. Cell. 168(4):629-643.

13. Brand, M., Leurent, C., Mallouh, V., Tora, L., and Schultz, P. (1999). Three-Dimensional Structures of the TAF_{II}-Containing Complexes TFIID and TFTC. *Science. 286(5447):2151-2153.*

14. Brennan, R. (1993). The winged-helix DNA-binding motif: Another helix-turn-helix takeoff. Cell. 74(5):773-776.

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

16. Cao, A., and Galanello, R., (2010). Beta-thalassemia. Genetics in Medicine, 12(2): 61-76.

17. CCP4 (1994). The CCP4 Suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr *D50*, 760-763.

18. Chalkley, G. and Verrijzer, C. (1999). DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. EMBO. J 18(17): 4835-4845.

19. Chen, D., Hinkley, C., Henry, R. W., and Huang, S. (2002). TBP Dynamics in Living Human Cells: Constitutive Association of TBP with Mitotic Chromosomes. Mol Biol Cell. 13(1): 276-284.

20. Chiang, C. M., and Roeder, R. G. (1995). Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. Science. 267:531-536.

21. Christova, R., and Oelgeschläger, T. (2002). Association of human TFIID-promoter complexes with silenced mitotic chromatin in vivo. Nat Cell Biol. 4(1):79-82.

22. Cianfrocco, M., Kassavetis, G., Grob, P., Fang, J., Juven-Gershon, T., Kadonaga, J., and Nogales, E. (2013). Human TFIID Binds to Core Promoter DNA in a Reorganized Strcutural State. Cell. 152(1-2): 120-131.

23. Clark, K., Halay, E., Lai, E., and Burley, S. (1993). Co-Crystal structure of the HNF-3/forkhead DNA-recognition motif resembles histone H5. Nature. 364 :412-420.

24. Compe, E. and Egly, J. M. (2012). TFIIH: when transcription met DNA repair. Nature Reviews Molecular Cell Biology. 13: 343-354.

25. Darlix, J., Lastra, M., Mely, Y., and Roques, B. (2002). Nucleocapsid Protein Chaperoning of Nucleic Acids at the Heart of HIV Structure, Assembly and cDNA Synthesis. HIV Sequence Compendium. 69-88.

26. Dawson, A., Hartwood, E., Paterson, T., and Finnegan, D. (1997). A LINE-like Transposable element in Drosphila, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids. EMBO. 16(14):4448-4455.

27. De Guzman, R., Wu, Z., Stalling, C., Pappalardo, L., Borer, P., and Summers, M. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. Science. 279: 384-388.

28. Devaiah, B.N., Lu, H., Gegonne, A., Sercan, Z., Zhang, H., Clifford, R.J., Lee, M.P., and Singer, D.S. (2010). Novel functions for TAF7, a regulator of TAF1-independent transcription. J Biol Chem. 285: 38772-38780.
29. De Vooght, K., van Wijk, R., and van Solinge, W. (2009). Management of Gene Promoter Mutations in Molecular Diagnostics. Clinical Chemistry. 55(4): 698-708.

30. Dey, A., York, D., Smalls-Mantey, A., and Summers, M. (2005) Composition and Sequence-Dependent Binding of RNA to the Nucleocapsid Protein of Moloney Murine Leukemia Virus. *Biochemistry,* 44, 3735-3744.

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

32. Dunphy, E., Johnson, T., Auerbach, S., and Wang, E. (2000). Requirement for TAF(II)250 acetyltransferase activity in cell cycle progression. Mol Cell Biol *20*, 1134-1139.

33. Dynlacht, B., Hoey, T., and Tjian, R. (1991). Isolation of Coactivators Associated with the TATA-Binding Protein That Mediate Transcriptional Activation. Cell. 66:563-576.

34. Gaiser, F., Tan, S., and Richmond, T.J. (2000). Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. J Mol Biol *302*, 1119-1127.

35. Gajiwala, K., and Burley S. (2000) Winged helix proteins. Curr. Op. in Structural Bio. 10: 110-116.

36. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M., and Mackay, J. P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. Trends in Biochemcial Sci. 32(2). 63-70.

37. Gegonne, A., Weissman, J., and Singer, D. (2001) TAFII55 binding to TAFII250 inhibits its acetyltransferase activity. PNAS. 98:22. 12432-12437.

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

39. Gibbons, J., Branco, A., Yu, S., and Lemos, B. (2014). Ribosomal DNA Copy Number is Coupled with Gene Expression Variation and Mitochondrial Abundance in Humans. Nature Communications. 5, 4820.

40. Goodrich, J. A. and R. Tjian (2010). Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. Nat Rev Genet 11(8): 549-558.

41. Grob, P., Cruse, M., Inouye, C., Peris, M., Penczek, P., Tjian, R., and Nogale, E. (2006). Cryo-Electron Microscopy Studies of Human TFIID: Conformational Breathing in the Integration of Gene Regulatory Cues. Structure. 14(3):511-520.

42. Grünberg, S., and Hahn, S. (2013). Structural insights into transcription initiation by RNA polymerase II. Trends Biochem Sci *38*, 603-611.

43. Harami, G., Gyimesi, M., and Kovacs, M. (2013). From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. Cell: Trends in Biochemical Sciences. 38(7): 364-371.

44. Hayashida, T., Sekiguchi, T., Noguchi, E., Sunamoto, H., Ohba, T., and Nishimoto, T. (1994). The CCG1/TAF_{II}250 gene is mutated in thermosensitive G1 mutants of the BHK21 cell line derived from golden hamster. Gene *141*, 267-270.

45. Hilton, T., Li, Y., Dunphy, E.L., and Wang, E. (2005). TAF1 histone acetyltransferase activity in Sp1 activation of the cyclin D1 promoter. Mol Cell Biol *25*, 4321-4332.

46. Hilton, T., and Wang, E. 2003. TFIID recruitment and Sp1 activation: dual function of TAF1 in cyclin D1 transcription. J. Biol. Chem. 15:12992-13002.

47. Hisatake, K., Hasegawa, S., Takada, R., Nakatani, Y., Horikoshi, M., and Roeder, R.G. (1993). The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. Nature *362*, 179-181.

48. Hoey, T., Dynlacht, D., Peterson, M., Pugh, B., and Tjian, R. (1990). Isolation and Characterization of the Drosophila Gene Encoding the TATA Box Binding Protein, TFIID. Cell. 61: 1179-1186.

49. Irvin, J., and Pugh, B. F. (2006). Genome-wide Transcriptional Dependence on TAF1 Functional Domains. JBC. 281, 6404-6412.

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

51. Jones, S., Shanahan, H., Berman, H., and Thornton, J. (2003). Using electrostatic potential to predict DNA-binding sites on DNA-binding proteins. Nucleic Acids Res. 31(24):7189-7198.

52. Juven-Gereshon, T., Cheng, S., and Kadonaga, J. (2006). Rational design of a super core promoter that enhances gene expression. Nat. Methods. 3(11):917-922.

53. Juven-Gershon, T. and Kadinaga, J. (2010). Regulation of Gene Expression via the Core Promoter and the Basal Transcription Machinery. Dev Biol. 339(2): 225-229.

54. Juven-Gershon, T., Hsu, J., and Kadonaga, J. (2006) Perspectives on the RNA polymerase II core promoter. Biochemical Society Transactions. 34(6). 1051-1054.

55. Kedinger, C., Gniazdowski, M., Mandel, J. Jr, Gissinger, F., and Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. Biochem Biophys Res Commun. 38(1):165-71.

56. Khatter, H., Vorlander, M., and Muller, C. (2017). RNA polymerase I and III: similar yet unique. Curr Op in Structural Biology. 47:88-94.

57. Killeen, M., Coulombe, B., and Greenblatt, J. (1992). Recombinant TBP, Transcription Factor IIB, and RAP30 Are Sufficient for Promoter Recognition by Mammalian RNA Polymerase II. JBC 267(14): 9463-9466.

58. Kim, J. K. and Diehl, J. A. (2009). Nuclear cyclin D1: An oncogenic driver in human cancer. J Cell Physiol. 220(2): 292-296.

59. Kim, T., Barrera, L., Zheng, M., Qu, C., Singer, M., Richmond, T. A., Wu, Y., Green, R., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature 436(7052): 876-880.

60. Kim, Y., Geiger, J., Hahn, S., and Sigler, P. (1993). Crystal structure of a yeast TBP/TATA-box complex. Nature. 365(6446):512-20.

61. Kimura, J., Nguyen, S., Liu, H., Taira, N., Miki, Y., and Yoshida, K. (2008). A functional genome-wide RNAi screen identifies TAF1 as a regulator for apoptosis in response to genotoxic stress. Nucleic Acids Res. 36(16): 5250-5259.

62. Kloet, S., Whiting, J., Gafken, P., Ranish, J., and Wang, E. (2012). Phosphorylation-Dependent Regulation of Cyclin D1 and Cyclin A Gene Transcription by TFIID Subunits TAF1 and TAF7. Molecular and Cellular Biology. 32(16): 3358-3369.

63. Knutson, B. A., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae RNA polymerase I Core Factor complex. Nat Struct Mol Biol 21(9): 810-816.

64. Kotsantis, P., Silva, L. M., Irmscher, S., Jones, R. M., Folkes, L., Gromak, N., and Petermann, E. (2016). Increased global transcription activity as a mechanism for

replication stress in cancer. Nature Comm. 7: 13087.

65. Krishna, S., Majumdar, I., and Grishin, N. (2003) Structural classification of zinc finger. Nucleic Acids Research. 31(2): 532-50.

66. Lavigne, A. C., Mengus, G., May, M. et al. (1996). Multiple interactions between hTAFII55 and other TFIID subunits. Requirements for formation of stable ternary complexes between hTAFII55 and the TATA-binding protein. JBC. 271: 19774-19780.

67. Lee, D., Gershenzon, N., Gupta, M., Ioshikhes, I., Reinberg, D., and Lewis, B. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Mol Cell Biol 25(21): 9674-9686.

68. Lee, T. I., and Young, R. (2013). Transcriptional Regulation and its Misregulation in Disease. Cell. 152(6): 1237-1251.

69. Leurent, C., Sanders, S., Demény, M., Garbett, K., Ruhlmann, C., Weil, P.A., Tora, L., and Schultz, P. (2004). Mapping key functional sites within yeast TFIID. EMBO J. 23(4): 719–727.

70. Leurent, C., Sanders, S., Ruhlmann, C., Mallouh, V., Weil, P.A. Kirschner, D., Tora, L., and Schultz, P. (2002). Mapping histone fold TAFs within yeast TFIID. EMBO J. 21(13): 3424–3433.

71. Lewis, B., Sims III, R., Lane, W., and Reinberg, D. (2005). Functional Characterization of Core Promoters Elements: DPE-Specific Transcription Requires the Protein Kinase CK2 and the PC4 Coactivator. Mol. Cell. 18:471-481.

72. Li, A., Piluso, L., Cai, X., Gadd, B., Ladurner, A., and Liu X. (2007). An Acetylation Switch in p53 Mediates Holo-TFIID Recruitment. *Mol Cell* 28, 498-421.

73. Li, Y. Y., Wen, H., Shi, X. B. and Li, H. T. (2015). Determination of Protein-DNA (ZMYND11-DNA) Interaction by a Label-Free Biolayer Interferometry Assay. bio-protocol. 5(4): e1402.

74. Liu, W. Coleman, P., Ma, E., Grob, P., Yang, J., Zhang, Y., Dailey, G., Nogales, E., an Tjian, R. (2009). Structures of three distinct activator–TFIID complexes. Genes Dev. 23(13): 1510–1521.

75. Longo, P., Kavran, J., Kim, M., and Leahy, D. (2013). Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol.* 529, 227–240.

76. Louder, R., He, Y., López-Blanco, J., Fang, J., Chacón, P., and Nogales, E. (2016). Structure of promoter-bound TFIID and insight into human PIC assembly. Nature. 531(7596): 604-609.

77. Loughlin, F., Gebert, L., Towbin, H., Brunschweiger, A., Hall, J., and Allain, F. (2011). Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. Nat.Struct.Mol.Biol. 19: 84-89.

78. Malarkey, C. S. and M. E. Churchill (2012). "The high mobility group box: the ultimate utility player of a cell." Trends Biochem Sci 37(12): 553-562.

79. Marks, D., Hopf, T., and Sander, C. (2012). Protein structure prediction from sequence variation. 30(11): 1072-1080.

80. Martin, J., Halenbeck, R., and Kaufmann, J. (1999). Human transcription factor hTAF(II)150 (CIF150) is involved in transcriptional regulation of cell cycle progression. Mol Cell Biol 19(8): 5548-5556.

81. Matangkasombut,O., Auty,R., Buratowski, S. (2004) Structure and Function of the TFIID Complex. *Advances in Protein Chemistry,* 67, 67-92.

82. Matthews, J., Kowalski, K., Liew, C. K., Sharpe, B., Fox, A., Crossley, M., and Mackay, J. P. (2000). A class of zinc fingers involved in protein-protein interactions. FEBS. 267(4): 1030-1038.

83. Mencia, M. and K. Struhl (2001). Region of yeast TAF 130 required for TFIID to associate with promoters. Mol Cell Biol 21(4): 1145-1154.

84. 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.*, et al.* (1996). The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell *87*, 1261-1270.

85. Morgan, M., Haj-Yahya, M., Ringel, A., Bandi, P., Brik, A., and Wolberger, C. (2016). Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science. 351(6274):725-728.

86. Moqtaderi, Z., Wang, J., Raha, D., White, R., Snyder, M., Weng, Z., and Struhl, K. (2010). Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nature Structural and Molecular Biology. 17, 635–640.

87. Musgrove, E., Caldon, C., Barraclough, J., Stone, A., and Sutherland, R. (2011). Cyclin D as a therapeutic target in cancer. Nature Review Cancer. 11(8):558-572.

88. Näär, A. M. et al. Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro. *Genes & Dev.* 12, 3020-3031 (1998).

89. Noguchi, E., Sekiguchi, T., Nohiro, Y., Hayashida, T., Hirose, E., Hayashi, N., and Nishimoto, T. (1994). Minimum essential region of CCG1/TAFII250 required for complementing the temperature-sensitive cell cycle mutants, tsBN462 and ts13 cells, of hamster BHK21 cells. Somat Cell Mol Genet 20(6): 505-513.

90. Ohler, U., Guo-chun Liao, G., Niemann, H., and Rubin, G. (2002). Computational analysis of core promoters in the *Drosophila* genome. Genome Biol. 3(12): research0087.1–87.12.
91. Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. Gens Dev. 10: 2657-2683.

92. Otwinowski, Z., and Minor, W., eds. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode (New York: Academic Press).

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

94. Papai, G., Tripathi, M., Ruhlmann, C., Werten, S., Crucifix, C., Weil, P.A., and Schultz, P. (2009). Mapping the initiator binding TAF2 subunit in the structure of hydrated yeast TFIID. Structure.17(3): 363–373.

95. Papai, G., Weil, P.A., and Schultz, P. (2011). New insights into the function of transcription factor TFIID from recent structural studies. Curr. Op. in Genetics & Development. 21(2): 219-224.

96. Pennington, K. L., Marr, S. K., Chirn, G. W., and Marr, M. T., 2nd (2013). Holo-TFIID controls the magnitude of a transcription burst and fine-tuning of transcription. Proc Natl Acad Sci U S A 110(19): 7678-7683.

97. Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J, Cramer P. (2016) Transcription initiation complex structures elucidate DNA opening. Nature, 533(7603),353-358.

98. Pugh, B. F., and Tjian, R. (1991). Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Gens Dev. 5: 1935-1945.

99. Ren, S., Jiang, Y., Yoon, H. R., Hong, S. W., Shin, D., Lee, S., Lee, D., Jin, M., Min, I., and Kim, S. (2014). Label-free Detection of the Transcription Initiation Factor Assembly and Specific Inhibition by Aptamers. Bull. Korean Chem. Soc. 35(5):1279-1284.

100. Ribeiro,J, Lovasco,L., Vanderhyden,B., and Freiman,R. (2014) Targeting TBP-Associated Factors in Ovarian Cancer. *Frontiers in Oncology*, 4, 45.

101. Roy, A., Kucukural, A., and Zhang, Y.(2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 5: 725-738.

102. Ruppert, S., Wang,E., and Tjian, R. (1993). Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature*, 362, 175-179.

103. Ruppert, S. and R. Tjian (1995). Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. Genes Dev 9(22): 2747-2755.

104. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Reviews Genetics* 8, 424-436.

105. Schafer, K. (1998). The Cell Cycle: A Review. Vet Pathol. 35, 461-478.

106. Segil, N., Guermah, M., Hoffmann, A., Roeder, R., and Heintz, N. (1996). Mitotic regulation of TFIID: inhibition of activator-dependent transcription and changes in subcellular localization. Genes & Dev. 10: 2389-2400.

107. Seifart, K., and Sekeris C. (1969). Alpha-amanitin, a specific inhibitor of transcription by mammalian RNA-polymerase. Z Naturforsch B. 24(12):1538-44.

108. Seizl M, Hartmann H, Hoeg F, Kurth F, Martin DE, Söding, J., and Cramer, P. (2011) A Conserved GA Element in TATA-Less RNA Polymerase II Promoters. PLoS ONE 6(11): e27595.

109. Sekiguchi, T., Y. Nohiro, Y. Nakamura, N. Hisamoto, and T. Nishimoto. (1991). The human *CCG1* gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein. Mol. Cell. Biol. 11: 3317–3325.

110. Sekine, S., Tagami, S., and Yokoyama, S. (2012). Structural basis of transcription by bacterial and eukaryotic RNA polymerase. Curr. Op in Strct Bio. 22(1):110-118.

111. Shao, H., Revach, M., Moshonov, S., Tzuman, Y., Gazit, K., Albeck, S., Unger, T., and Dikstein, R. (2005). Core promoter binding by histone-like TAF complexes. Mol Cell Biol 25(1): 206-219.

112. Shazman, S., and Mandel-Gutfreund, Y. (2008). Classifying RNA-Binding Proteins Based on Electrostatic Properties. PLoS Comput Biol. 4(8): e1000146.

113. Song, C., Keller, K., Chen, Y., Murata, K., and Stamatoyannopoulos, G. (2002). Transcription coactivator CBP has direct DNA binding activity and stimulates transcription factor DNA binding through small domains. *Biochem Biophys Res Commun*. 296(1): 118–124.

114. Struhl, K. (1989). Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. TIBS. 14: 137-140.

115. Suzuki-Yagawa, Y., Guermah, M., and Roeder, R. (1997). The ts13 mutation in the TAF(II)250 subunit (CCG1) of TFIID directly affects transcription of D-type cyclin genes in cells arrested in G1 at the nonpermissive temperature. Mol Cell Biol. 17(6): 3284–3294.

116. Taylor, N.M., Baudin, F., von Scheven, G., and Müller, C.W. (2013). RNA polymerase III-specific general transcription factor IIIC contains a heterodimer resembling TFIIF Rap30/Rap74. Nucleic Acids Res *41*, 9183-9196.

117. Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012). Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. Transcription. 3:1, 2-7.

118. Thein, S.L. (2013). The Molecular Basis of ß–Thalassemia. Cold Spring Harb Perspect Med. 3(5): a011700.

119. Thomas, M. and Chiang, C. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41(3): 105-178.

120. Thomm, M. (1996). Archaeal transcription factors and their role in transcription initiation. FEMS Microbiology Reviews 18:159-171.

121. 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.

122. Van Dyke, M. W., Roeder, R.G., and Sawadogo, M. (1988). Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. Science. 241:1335-1338.

123. Van Nuland, R., Schram, A., van Schaik, F., Jansen, P., Vermeulen, M., and Timmers, H. (2013). Multivalent Engagement of TFIID to Nucleosomes PLoS ONE. 8(9): e73495.

124. Vandevenne,M., Jacques,D.A., Artuz,C., Nguyen,C.D., Kwan,A.H.Y., Segal,D.J., Matthews,J.M., Crossley,M., Guss,J.M. and Mackay,J.P. (2013) New Insights into DNA Recognition by Zinc Fingers Revealed by Structural Analysis of the Oncoprotein ZNF217. *J Biol Chem.*, 288(15),10616-10627.

125. Verrijzer, C. P., Yokomori, K., Chen, J. L., and Tjian, R. (1994). Drosophila TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. Science 264(5161): 933-941.

126. Villicana, C., Cruz, G., and Zurita, M. (2014). The basal transcription machinery as a target for cancer therapy. Cancer Cell Int 14(1): 18.

127. Vo Ngoc, L., Cassidy, C.J., Huang, C.Y., Duttke, S.H., Kadonaga, J.T. (2017). The human initiator is a distinct and abundant element that is precisely positioned in focused core promoters. Genes Dev. 31(1):6-11.

128. Wassarman, D.A., and Sauer, F. (2001). TAF(II)250: a transcription toolbox. J Cell Sci *114*, 2895-2902.

129. Wang, E. and Tjian, R. (1994) Promoter-selective transcriptional defect in cell cycle mutant ts13 rescued by hTAFII250. *Science*, 263(5148),811-814.

130. Wang, E., Zou, S., & Tjian, R. (1997) $TAF_{II}250$-dependent transcription of cyclin A is directed by ATF activator proteins. *Genes Dev.* 11, 2658-2669.

131. Wang, H., Curran, E., Hinds, T., Wang, E., and Zheng, N. (2014). Crystal Structure of a TAF1-TAF7 complex in human transcription factor IID reveals a promoter binding module. Cell Res. 24(12): 1433-1444.

132. Wang, J., Qin,S., Li,F., Li,S., Zhang,W., Peng,J., Zhang,Z., Gong,Q., Wu,J. and Shi,Y. (2014) Crystal structure of human BS69 Bromo-ZnF-PWWP reveals its role in H3K36me3 nucleosome binding. *Cell Res.,* 24(7), 890-3.

133. Wang M, Weiss, M., Simonovic. M., Haetinger, G., Schrimpf, S., Hengartner, M., and von Mering, C. (2012). PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics* **11**, 492-500.

134. Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., and Hahn, S. (2017). Transcription of Nearly ALL Yeast RNA Polymerase II-Transcribed Genes is Dependent on Transcription Factor TFIID. Mol Cell. 68(1): 118-129.

135. Weis, L., and Reinberg, D. (1997). Accurate Positioning of RNA Polymerase II on a Natural TATA-Less Promoter Is Independent of TATA-Binding-Protein-Associated Factors and Initiator-Binding Proteins. Mol. Cell. Biol. 17(6):2973-2984.

136. Weissman, J.D., Brown, J.A., Howcroft, T.K., Hwang, J., Chawla, A., Roche, P.A., Schiltz, L., Nakatani, Y., and Singer, D.S. (1998). HIV-1 tat binds $TAF_{II}250$ and represses $TAF_{II}250$-dependent transcription of major histocompatibility class I genes. Proc Natl Acad Sci U S A *95*, 11601-11606.

137. Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J., Li, B., Barton, M., Li, W., Li, H., and Shi, X.(2014). ZMYND11 links histone H3.3 K36 trimethylation to transcription elongation and tumor suppression. Nature. 508(7495): 263–268.

138. Wieczorek, E., Brand, M., Jarq, X., and Tora, L. (1998). Function of TAFII-containing complex without TBP in transcription by RNA polymerase II. Nature. 393:187-191.

139. Yamamoto M., Yoshida M., Ono K., Fujita T., Ohtani-Fujita N., Sakai T., and Nikaido T. (1994) Effect of tumor suppressors on cell cycle-regulatory genes: RB suppresses $p34^{cdc2}$ expression and normal p53 suppresses cyclin A expression.Exp. Cell Res. 210:94–101.

140. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and

identification of DNA motifs enriched in human TATA-less core promoters. Gene 389(1): 52-65.

141. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: Protein structure and function prediction. Nature Methods, 12: 7-8.
142. Zawel, L., Kumar, K. P., and Reinberg, D. (1995). Recycling of the general transcription factors during RNA polymerase II transcription. Genes Dev. 9: 1479-1490. Link TFIID is the only GFT remains associated with the promoter.

143. Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., Liu, S., Su, Y., Guo, J., Huang, L., Wan, J., Wu, H., Qian, J., Cheng, X., Zhu, H., Ming, G., Jin, P., and Song, H. (2016). Lin28A binds active promoters and recruits Tet1 to regulate gene expression. *Mol Cell*. 61(1): 153–160.

144. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics, 9: 40.

145. Zhang, Z., Boskovic, Z., Hussain, M., Hu, W., Inouye, C., Kim, H., Abole, A. K., Doud, M., Lewis, T., Koehler, A., Schreiber, S., and Tjian, R. (2015). Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. eLife. 4:e07777.

146. Zhang, Z., English, B., Grimm, J., Kazane, S., Hu, W., Tsai, A., Inouye, C., You, C., Piehler, J., Schultz, P., Lavis, L., Revyakin, A., and 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.

147. Zhao,S., Choi,M., Overton,J., Bellone,S., Roque,D., Cocco,E., Guzzo,F., English,D., Varughese,J., Gasparrini,S., Bortolomai,I., Buza,N., Hui,P., Abu-Khalaf,M., Ravaggi,A., Bignotti,E., Bandiera,E., Romani,C., Todeschini,P., Tassi,R., Zanotti,L., Carrara,L., Pecorelli,S., Silasi,D., Ratner,E., Azodi,M., Schwartz,P., Rutherford,T., Stiegler,S., Mane,S., Boggon,T., Schlessinger,J., Lifton,R., and Santin,A. (2013) Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc Natl Acad Sci U S A,* 110(8), 2916-2921.

148. Zheng, N., Fraenkel, E., Pabo, C., and Pavletich, N. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. Genes Dev *13*, 666-674.

149. Zylber, E., and Penman, S. (1971). Products of RNA Polymerase in HeLa Cell Nuclei. PNAS. 68(11): 2861-2865.

APPENDIX I: MATERIALS AND METHODS

CHAPTER 2

Protein expression and purification of TAF1/TAF7 complexes

Human TAF1 (amino acids 600-1236 or 5-1375) was co-expressed with the human TAF7 full-length protein as a glutathione S-transferase (GST) fusion protein in High Five (Invitrogen) insect cells. The TAF1-TAF7 complex was purified by glutathione affinity chromatography using lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1.0 mM DTT) supplemented with protease inhibitors. The protein complex was further purified by anion exchange and size exclusion chromatography (GE Healthcare) after off-column cleavage by the tobacco etch virus (TEV) protease. The eluted complex was concentrated by ultrafiltration to 12 $mg \cdot ml^{-1}$ in buffer containing 25 mM Hepes, pH 7.5, 150 mM NaCl, 1.0 mM DTT. Trypsin (0.1% w/w) was added to the complex sample to trim off flexible regions during crystallization screening. The protein samples used for EMSA assays were purified the same way as described above except the final size exclusion was omitted.

Crystallization and data collection

The crystals of human TAF1-TAF7 complex were obtained at 4°C by the hanging-drop vapor diffusion method, using 1.2 µl protein complex sample mixed with an equal volume of reservoir solution containing 0.2 M potassium sodium tartrate tetrahydrate, pH 7.4, 20% w/v PEG3350. After being briefly equilibrated in the reservoir solution, the crystals were transferred to a cryoprotectant solution containing 25% w/v PEG3350, 0.2 M potassium sodium tartrate tetrahydrate, pH 7.4 and 10% glycerol, and then flash-frozen in liquid nitrogen. The human TAF1-TAF7 complex heavy atom derivative crystals were prepared by soaking the native crystals in a buffer containing 0.2 M potassium sodium tartrate tetrahydrate, pH 7.4, 20% w/v PEG3350 supplemented with 1.0 mM $K_2Pt(NO_2)_4$ for 12 hours, followed by a back soaking in the reservoir solution for an additional hour. The crystals were subsequently harvested the same way as the native ones. All data sets were collected at the BL8.2.1 and BL8.2.2 beamlines at the

Advanced Light Source of the Lawrence Berkeley National Laboratory. Native crystals ($P2_12_12_1$, a=83.2Å, b=94.5Å, c=101.8Å, a=β=ɣ=90˚) diffracted to 2.3Å resolution, and heavy atom derivative crystals ($P2_12_12_1$, a=83.1Å, b=94.5Å, c=102.2Å, a=β=ɣ=90˚) diffracted to 2.8Å resolution. X-ray diffraction data statistics are shown in Supplemental Table 1.

Structure determination

The human TAF1-TAF7 complex structure was determined by SAD (single-wavelength anomalous diffraction) phasing of a platinum (Pt) protein derivative. Reflection data were indexed, integrated and scaled with the HKL2000 package (Otwinowski and Minor, 1997). Experimentally phased map with a well-defined solvent boundary and readily interpretable electron density for protein was calculated with PHENIX (Adams et al., 2002). An initial model was established using PHENIX and refined against the native data set. The final model was manually built with the program COOT (CCP4, 1994). The complex structure was refined to an R_{factor}=18.6% and an R_{free}=23.4%. Crystallographic data statistics are shown in Supplemental Table 1. 99.81% of all residues of human TAF1-TAF7 complex are in the favored and allowed region of the Ramachandran plot. All structural figures were prepared using PyMOL (http://www.pymol.org/).

Electrophoretic mobility shift assays (EMSA)

TAF1-TAF7 complex (4-180 pmoles) purified from insect cells was incubated with 4 ng of ^{32}P 5'-end labeled DNA probes in 10 mM Hepes pH 7.9, 5 mM $MgCl_2$, 100 mM NaCl, 10% glycerol, 20 mM tetrasodium pyrophosphate, 0.2 mM dI:dC for 1 hour at 25°C. For gel electrophoresis, 6x loading buffer (20% Ficoll, 0.025% bromophenol blue) was added and binding reactions were loaded onto nondenaturing 5% polyacrylamide (37.5:1 acrylamide:bis) that was pre-run in 0.5x TBE at 100V for 30 minutes. Samples were resolved for 1.5 hours at 100V and shifted complexes detected by autoradiography. Competition assays using 22.5 pmoles of purified TAF1-TAF7 complex were performed as described above with the addition of either unlabeled cyclin D1 core promoter (CD1) double-stranded (20, 40, 80, 200 ng) or sense single-stranded (10, 20, 40, 100 ng) DNA to the binding reaction. DNA binding was quantified by excising the

protein bound and unbound DNA bands from the dried gels. The amount of radioactivity in each band was measured by liquid scintillation (Tri-Carb B2810 TR, PerkinElmer). The percent DNA bound was calculated and plotted using Prism (GraphPad Software). The sequences of DNA probes using for EMSA are provided in Appendix II.

TAF1 ML Protein Purification

For expression of His-SET3 tagged TAF1 proteins, cDNA of TAF1 (ML,992-1075) was PCR amplified from CS2+ containing full length TAF1 and cloned into pAL-SET3 using ligase independent cloning; site-directed mutagenesis was used to introduce RR and KK point mutations (Appendix II). TAF1 ML expression constructs were transformed into BL21* cells for protein expression. Starter cultures of 10 mL were diluted into 1L LB containing 30 ug/ml chloramphenicol, and cells were grown to an optical density of 0.8-1.0 at 600 nm and chilled on ice for 1 h. TAF1 expression was induced by addition of 0.4 mM IPTG (isopropyl--D-thiogalactopyranoside) for 18 h at 16 °C. Cells were harvested by centrifugation and lysed by pulse-sonication lysis in 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5mM imidazole, 5% glycerol, 0.05% Triton-X, 1 mM DTT supplemented with protease inhibitors. His-SET3-ML was purified using Ni-nitrilotriacetic acid (NTA)- agarose (Qiagen), washed with Buffer A (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM imidazole, 5% glycerol, 1 mM DTT) followed by a high salt wash (50 mM Tris-HCl, pH 8.0, 1 M NaCl, 5% glycerol, 1 mM DTT) and a final wash with Buffer A. Protein was eluted with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 200 mM imidazole, 5% glycerol, 1 mM DTT. Proteins were further purified by cation exchange (GE Healthcare) after off-column cleavage by the tobacco etch virus (TEV) protease.

Bio-layer Interferometry (BLI)

5'-biotinylated oligonucleotides were annealed to their complement (IDT, USA) in 1x binding buffer (10 mM HEPES pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) by heating at 72 ºC for 5 min and slow cooling to 25˚C. Double-stranded IMD oligo was purified by extraction from agarose after gel electrophoresis and ethanol precipitated. The 100nM of oligo was then loaded onto streptavidin-coated (SA) sensors and the interactions between the DNA coated probes and TAF1 proteins were measured using ForteBio Octet RED96 system (Pall Life Sciences). For loading, SA sensors were

prewetted in assay buffer (10 mM HEPES pH 7.9, 5 mM $MgCl_2$, 75 mM NaCl, 10% glycerol, 0.1% ovalbumin, 0.2 mM dI:dC) for 30 min then dipped into individual wells on 96-well plates (200 µL/well) containing biotinylated DNA in assay buffer (Ren et al., 2014; Li et al., 2015). Unbound DNA was removed by washing and exposed streptavidin blocked with biocytin (100nM). Purified ML was diluted in assay buffer to 35uM, 12uM, 3.8uM, 1.29 uM, 432nm, and 142nm. Association of the protein to the probe was observed for 300 s followed by dissociation in assay buffer for 300 s. Throughout loading, association and dissociation sensors were shaking (1,000 rpm) at a constant 30 °C to promote specific interaction and to reduce non-specific binding. A reference sensor (no biotinylated DNA loaded) was included and subjected to the same procedure to control for non-specific interactions. Sensors were regenerated (10 mM HEPES pH 7.9, 1M NaCl) between experiments and a new baseline established. Reference-subtracted data was used to calculate the equilibrium dissociation constant (K_d) by steady-state response analysis assuming 1:1 kinetics.

CHAPTER 3

Conservation Alignments and Protein Modeling

TAF1 sequences from *H. sapiens*, *D. rerio, X. tropicalis, D. melanogaster, C. elegans, A. thaliana, S. cerevisiae,* and *C. neoformans* acquired from the NCBI database and aligned using CLC Sequence View 7. Similarly, sequences for FAM90a-I, I factor, HIV1 nucleocapsid, lin28a, XPO, ZCCHC12, HEXBP, BYR3, and UMSBP2 were downloaded from the NCBI database and aligned to human TAF1 zinc knuckle. The tertiary structure for TAF1 ZnK was generated using Iterative Threading Assembly Refinement (I-TASSER; Zheng, 2008; Roy, et al., 2010; Yang et al., 2015); this analysis also determined closely related structures and predicted ligand-binding site. EV-Fold analysis was performed on TAF1 1234-1371 as described (Marks et al., 2012). The top 100 couplings were plotted and color-coded based on the strength of the interaction.

TAF1 ZnK Protein Purification

For expression of TAF1 fusion proteins, cDNA of TAF1 (TAF1c-992-1371, ZnA-1234-1371, ZnC-1234-1303, ZnD-1256-1303) was PCR amplified from CS2+ containing

full length TAF1 and cloned into pAL-GB1(for ZnK constructs) or pALSET3 (for TAF1c) using ligase independent cloning; site-directed mutagenesis was used to introduce point mutations into ZnC and ZnD constructs (Appendix II). TAF1 expression constructs were transformed into BL21* cells for protein expression. Starter cultures of 10 mL were diluted into 1L LB containing 30 ug/ml chloramphenicol, and cells were grown to an optical density of 0.8-1.0 at 600 nm and chilled on ice for 1 h. TAF1 expression was induced by addition of 0.2 mM IPTG (isopropyl--D-thiogalactopyranoside) for 18 h at 16 °C. Cells were harvested by centrifugation and lysed by pulse-sonication lysis in 50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5mM imidazole, 5% glycerol, 0.05% Triton-X, 1 mM DTT supplemented with protease inhibitors. His-GB1-TAF1 was purified using Ni-nitrilotriacetic acid (NTA)- agarose (Qiagen), washed with Buffer A (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM imidazole, 5% glycerol, 1 mM DTT) followed by a high salt wash (50 mM Tris-HCl, pH 8.0, 1 M NaCl, 5% glycerol, 1 mM DTT) and a final wash with Buffer A. Protein was eluted with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 200 mM imidazole, 5% glycerol, 1 mM DTT. For proteolytic digestion and electrophoretic mobility shift assays TAF1 proteins was further purified by cation exchange (GE Healthcare) after off-column cleavage by the tobacco etch virus (TEV) protease. Proteins for biolayer interferometry were further purified by size-exclusion chromatography (GE Healthcare) using 10 mM Hepes, pH 7.9, 150 mM NaCl, 1 mM DTT.

Electrophoretic Mobility Shift Assay (EMSA)

TAF1 ZnK (1234-1371) (7.6 – 61.3 pmoles) purified from *E. coli* was incubated with 4 ng of ^{32}P 5′-end labeled DNA in 10 mM Hepes pH 7.9, 5 mM $MgCl_2$, 100 mM NaCl, 10% glycerol, 20 mM tetrasodium pyrophosphate, 0.2 mM dI:dC for 1 h at 25 °C. For gel electrophoresis, 6x loading buffer (20% Ficoll, 0.025% bromophenol blue) was added and binding reactions were loaded onto nondenaturing 5% polyacrylamide (37.5:1 acrylamide:bis) pre-run in 0.5× TBE at 100 V for 30 min. Samples were resolved for 1.5 h at 100 V and shifted complexes detected by autoradiography. The sequences of DNAs used for EMSA are provided in Appendix II.

Bio-layer Interferometry (BLI)

5'-biotinylated oligonucleotides were annealed to their complement (IDT, USA) in 1x binding buffer (10 mM HEPES pH 7.9, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT) by heating at 72 ºC for 5 min and slow cooling to 25˚C. Double-stranded oligos were purified by extraction from agarose after gel electrophoresis and ethanol precipitated. Purified biotinylated double-stranded oligos were loaded onto streptavidin-coated (SA) sensors and the interactions between the DNA coated probes and TAF1 proteins were measured using ForteBio Octet RED96 system (Pall Life Sciences). For loading, SA sensors were prewetted in assay buffer (10 mM HEPES pH 7.9, 5 mM $MgCl_2$, 100 mM NaCl, 10% glycerol, 0.1% ovalbumin, 0.2 mM dI:dC) for 30 min then dipped into individual wells on 96-well plates (200 µL/well) containing biotinylated DNA (100nM) in assay buffer (Ren et al., 2014; Li et al., 2015). Unbound DNA was removed by washing and exposed streptavidin blocked with biocytin. ZnK proteins were diluted in assay buffer to 3uM, 1uM, 333nm 111nm, 37nm, and 12nm. Association of the protein to the probe was observed for 300 s followed by dissociation in assay buffer for 300 s. Throughout loading, association and dissociation sensors were shaking (1,000 rpm) at a constant 30 °C to promote specific interaction and to reduce non-specific binding. A reference sensor (no biotinylated DNA loaded) was included and subjected to the same procedure to control for non-specific interactions. Sensors were regenerated (10 mM HEPES pH 7.9, 1M NaCl) between experiments and a new baseline established. Reference-subtracted data was used to calculate the equilibrium dissociation constant (K_d) by steady-state response analysis assuming 1:1 kinetics.

DNA Protection Assay

TAF1 (aa 1234-1371) was exposed to limited proteolytic digestion in the presence and absence of DNA (IMD) to determine DNA-bound stabilized regions. TAF1 (625 pmol) was incubated with or without DNA (700 pmol) in 10 mM Hepes pH 7.9, 5 mM $MgCl_2$, 100 mM NaCl, 10% glycerol, 20 mM tetrasodium pyrophosphate, 0.2 mM dI:dC) for 1h at 25°C. Subtilisin (0.002, 0.006, 0.02, 0.06, 0.2, and 0.6 ug/ml) was added to each sample and incubated overnight on ice at 4 °C. Samples were divided in two and analyzed by as follows: SDS-PAGE, coomassie stained, and scanned for analysis;

SDS-PAGE, transferred to PVDF membrane, coomassie stained and stabilized fragments excised and N-terminally sequenced.

CHAPTER 4

ts13 Complementation Assay

Mammalian TAF1 expression plasmids contain N-terminal HA-tagged human TAF1 coding sequence inserted downstream of the CMV promoter in CS2+ vector (Wang and Tjian, 1993). Point mutations in TAF1 were introduced by site-directed mutagenesis and confirmed by DNA sequencing (Appendix II: List of Primers). *ts*13 cells were grown at 33.5 °C in Dulbecco's modified Eagles medium (Gibco) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and penicillin/streptomycin. For complementation assays, cells were seeded into 6-well dishes, grown overnight to 70%-80% confluency and transfected with 2 µg of CS2+ or TAF1 expression plasmids using polyethylenimine (PEI) transfection reagent (ratio 2.5:1 of PEI:DNA) according to previously described protocol (Longo et al., 2013). Transfected cells were maintained for an additional 18-24 h at 33.5 °C, then either harvested for Western Blot analysis or shifted to nonpermissive temperature of 39.5 °C. Number of DAPI positive viable cells was determined after 36-48 h at 39.5 °C, and the percentage relative to WT-TAF1 expression (given the value of 100%) was calculated.

TFIID Immunoprecipitation

For preparation of nuclear extracts, HEK293 cells were resuspended in buffer A (10 mM HEPES, pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT) and incubated on ice for 15 min. Cells were lysed by pushing through a 25-gauge needle 5 times. The crude nuclear pellet was isolated by centrifugation for 20 s at 12,000 g, resuspended in buffer C (20 mM HEPES, pH 7.9, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride, 0.5 mM DTT), and incubated for 30 min at 4°C. After centrifugation for 5 min at 12,000 g, the supernatant/nuclear extract was incubated with anti-TAF4 or anti-HA antibody overnight at 4°C. Ten microliters of protein A–Sepharose CL-4B (GE Healthcare) was added, and the samples were nutated for 2 h at

4°C. Isolated proteins were washed 5 times with Buffer C and analyzed by immunoblotting using antibodies against HA, TAF1 bromodomain (gift from X.Liu), or TBP (gift from R. Tjian).

Luciferase Reporter Assay

Human cyclin D1 luciferase construct was constructed by subcloning the EcoRI to PvuII fragment of pD1-G065 (kindly provided by Yue Xiong) into the SmaI site of pGL2-basic as previously described (Hilton et al., 2003). The human cyclin A2 luciferase construct has been described (Yamamoto et al., 1994). HEK293 cells were grown at 37°C in Dulbecco's modified Eagles medium (Gibco) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, and penicillin/streptomycin. For luciferase assays, cells were seeded into 24-well dishes, grown overnight to 60%-70% confluency and cotransfected with 100 ng of CS2+ or TAF1 expression constructs and 50 ng luciferase reporter construct using FuGene transfection reagent (ratio 2.5:1 of FuGene:DNA) according to the manufacturer's protocol (Roche). Transfected cells were maintained an additional 36 h at 37 °C, after which cells were harvested and lysed in 100 µL of 1x passive lysis buffer (Promega). Luciferase activity was measured according to manufacturer's protocol (Promega) on Lumat LB 9507 (Berthold Technologies) and normalized for total protein. Transcription activity was expressed relative to CS2+ vector transfected cells (set to 1).

Chromatin Immunoprecipitation (ChIP)

HEK 293 cells were seeded on 10-cm plates and transfected with CS2+ or TAF1 expression constructs and incubated for 24-36h at 37 °C. Before harvesting, cells were cross-linked with 400 ul of 37% formaldehyde and analyzed as previously described (Kloet et al., 2012). In brief, cells were lysed in 0.5 mL immunoprecipitation (IP) buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, and 0.5% Nonidet P-40) containing protease inhibitors and sonicated 8 times for 20 s at 40% amplitude (Branson Digital Sonifier)**.** Nuclear extract was incubated overnight at 4°C with either anti-HA (Roche) or mouse IgG (Abcam). Immunoprecipitated complexes were washed and eluted. Cross-links were reversed and DNA was isolated. Input DNA was purified from 100 ul extract. Fifty nanograms of input DNA and 2µL of purified TAF1

bound DNA was amplified with SsoFast EvaGreen Supermix (Bio-Rad) using primers spanning the promoters of cyclin D1 and cyclin A2 (Supplemental Table 1). Quantitative PCR was performed on the Applied Biosystems 7500 Real-Time PCR system and the data expressed as percent input.

APPENDIX II: SUPPLEMENTARY TABLES

List of Primers

Oligonucleotides		
Name	**Use**	**Nucleotide Sequence (5'-3' only sense strand listed)**
Cysmut	Mutagenesis	5'-ctgccctaaaactgaaagctggggcagctggtgccattggaca c-3'
RKmut	Mutagenesis	5'-tgccattggacacatggcgactaacgcattctgccccctctattat-3'
3AWH	Mutagenesis	5'-gacgacaggctagcgctctgcgctgac-3'
RRmut	Mutagenesis	5'-cagatgcagaccttgctgcctttccctgaaaaatg-3'
KKmut	Mutagenesis	5'-ctgaggaagagattgcagcgttgtcccgctgg-3'
Cyclin D1 -91 For	qPCR	5'-cgtcacacggactacagggg-3'
Cyclin D1 +29 Rev	qPCR	5'-cgctcggctctcgcttctgc-3'
Cyclin A2 -143 For	qPCR	5'-tccagcgggctgctcgctgc -3'
Cyclin A2 -41 Rev	qPCR	5'- ctcgagaccacgcagggccgagga -3'
c-Fos 839 For	qPCR	5'-aacccctcatcttgggggcccac-3'
c-Fos Rev	qPCR	5' ttctcagttgctcgctgcagatgc-3'
IMD fragment of Super Core Promoter (IMD)	DNA Binding	5'-gtcctcagtcgcgatcgaacactcgagccgagcagacgtgccta-3'
Cyclin D1 core promoter (CD1P)	DNA Binding	5'-aggggagttttgttgaagttgcaaagtcctggagcctccagagggctgtcg- 3'
CD1P transcript	RNA Binding	5'-aggggaguuuuguugaaguugcaaaguccuggagcc uccagagggcugucg- 3'
Random DNA	DNA Binding	5'-gaa tgtcgcgaaggcagcgttcgcgaagacagcaaagaag-3'

Sense strand mutagenesis primers, quantitative PCR primers sequences used in ChIP analysis, and DNA binding probe sequences.

Table 1: Summary Table of Binding Constants

Protein	Probe	K_d (nM)	Error (nM)
ZnA	IMD	288	46
	CDP1	411	101
	Random	1011	228
	ssDNA	1359	387
	RNA	836	387
ZnC WT	IMD	342	42
ZnC Mut	IMD	N/A	N/A
ZnC RK Mut	IMD	N/A	N/A
ZnD WT	IMD	503	67
ZnD Mut	IMD	N/A	N/A
ZnD RK Mut	IMD	N/A	N/A

VITA

Elizabeth Colette Curran was born and raised in Portage, Wisconsin. After graduating from high school, she attended the University of Denver, where she earned a Bachelor's of Science in Molecular Biology. An internship with Amgen between her junior and senior year started her on a career path focused on scientific discovery. She then continued her scientific career as a research associate at the University of Wisconsin-Madison in the Department of Biomolecular Chemistry. Eager to investigate research related to human health, she moved to Seattle to explore the biotechnology industry as a research scientist with Emergent Biosolutions. In 2012, she joined the University of Washington Department of Pharmacology as a graduate research associate and earned a Doctorate of Philosophy in 2017.