Determining *in vivo* transcription factor dynamics using formaldehyde-mediated techniques

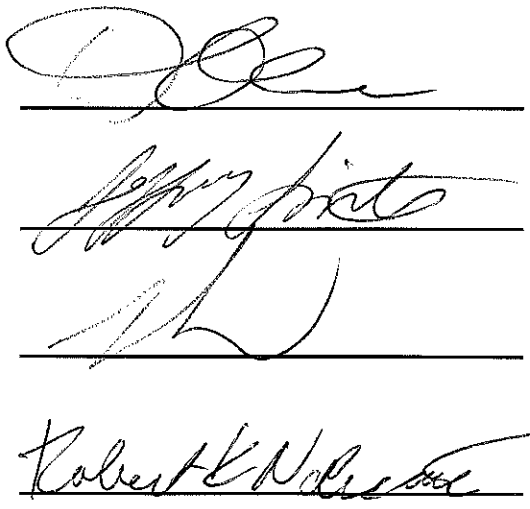
Elizabeth Hoffman
Queensbury, NY

B.S. State University of New York at Geneseo, 2011
M.S. University of Virginia, 2015

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Biochemistry & Molecular Genetics

University of Virginia
April, 2018



ABSTRACT

Transcription is a fundamentally important process for determining cell type identity, growth, and function. A central regulatory step occurs during the formation of the transcriptional preinitiation complex (PIC) at promoter sites. While the PIC consists mainly of a set of general transcription factors (GTFs) and RNA polymerase II (Pol II), there are many other factors and complexes that recruit the machinery and/or control chromatin structure at promoters to facilitate PIC assembly and activity. Although much is known about the binding locations of these factors, their *in vivo* dynamics remain largely unknown. Recent studies indicate that binding of factors *in vivo* is a dynamic process, but techniques for measuring chromatin interaction dynamics have in general been limited either by insufficient time resolution or by the inability to monitor binding to single copy genes. Our lab developed the crosslinking kinetics (CLK) assay, which measures binding dynamics based on the formaldehyde crosslinking-time dependent nature of the measured chromatin immunoprecipitation (ChIP) signal. We updated the crosslinking and quenching conditions to improve the quench efficiency and measured TATA-binding protein (TBP), a component of the PIC, at select promoters. We found that there is a wide range of formaldehyde crosslinking rates, and crosslinking time-dependent changes in ChIP signal can be described by factor-limited or crosslink-limited models. The residence time of TBP was ~2 minutes at the promoters measured, while the fractional promoter occupancy varied from ~0.05 to 0.7. We have also adapted the Anchor Away approach, a nuclear depletion technique, to estimate the stability of a chromatin-factor interaction by elucidating the off-rate and residence time as the nuclear factor is depleted. This data is approaching agreement with the CLK data at some loci, but in disagreement at others. A previous study developed the Competition ChIP approach, where a diploid strain with differentially tagged copies of the target allele, one under the control of an inducible promoter, can measure residence time of a factor at specific sites based on the exchange of tagged protein as

determined by ChIP. The CLK, Anchor Away, and Competition ChIP data for TBP at select loci are beginning to converge. By comparing binding dynamics measured with these three techniques, the *in vivo* binding dynamics of PIC components can be determined to understand PIC assembly pathways and transcriptional regulation at individual promoters. Additionally, we measured binding dynamics of the activator Gal4 at several of the *GAL* genes. A residence time of ~14 minutes is in agreement with the previous CLK study as well as previous results obtained using competition ChIP. Our collaborators have measured real-time transcriptional output at two *GAL* genes. Comparison of the results from different approaches will be valuable in correlating transcription factor dynamics with gene expression and better understanding the molecular mechanisms of complex assembly and regulator activity *in vivo*.

PUBLICATIONS

The following papers correspond to the material presented in the indicated chapters.

Chapter II

Hoffman EA, Zaidi H, Shetty SJ, Bekiranov S, Auble DT. An improved method for measuring chromatin-binding dynamics using time-dependent formaldehyde crosslinking. *Bio-Protocol*, *in press*.

Zaidi H*, **Hoffman EA***, Shetty SJ, Bekiranov S, Auble DT. Second-generation method for analysis of chromatin binding using formaldehyde crosslinking kinetics. *J. Biol. Chem.* **292**, 19338-19355, 2017.

Hoffman EA, Frey BL, Smith LM, Auble DT. Formaldehyde crosslinking: a tool for the study of chromatin complexes. *J. Biol. Chem.* 290(44): 26404-26411, 2015.

Viswanathan R, **Hoffman EA**, Shetty SJ, Berkiranov S, Auble DT. Analysis of chromatin binding dynamics using the crosslinking kinetics (CLK) method. *Methods* 70(2-3): 97-107, 2014.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AA	Anchor Away
bp	Base pair
CC	Competition ChIP
C_{FH}	Formaldehyde concentration used for crosslinking, mol
ChIP	Chromatin immunoprecipitation
CK2	Casein kinase 2
CLK	Crosslinking kinetics assay
C_{TF}	Nuclear concentration of transcription factor, mol
DBD	DNA binding domain
FCS	Fluorescence correlation spectroscopy
FLIP	Fluorescence loss in photobleaching
FRAP	Fluorescence recovery after photobleaching
GTF	General transcription factor
HA	Hemagglutinin
HAT	Histone acetyltransferase
k_a	On-rate of transcription factor, 1/s
k_d	Off-rate of transcription factor, 1/s
K_d	Dissociation constant, k_d/k_a, mol
kDa	Kilodalton
k_{xl}	Formaldehyde crosslinking rate, 1/mol s
LBD	Ligand binding domain
Mot1	Modifier of transcription 1
NC2	Negative cofactor 2
OE	Overexpression
PIC	Preinitiation complex

Pol II	RNA polymerase II
qPRC	Quantitative PCR (polymerase chain reaction)
SAGA	Spt-Ada-Gcn5 acetyltransferase
SCF	Skip1-Cullin-F-box protein
SMT	Single molecule tracking
S_{sat}	ChIP signal at saturation
SSR	Sum of squared residuals
TAF	TBP-associated factor
TCS	Temporal image correlation spectroscopy
θ_b	Fractional occupancy
TBP	TATA-binding protein
TF	Transcription factor
τ_{xl}	Crosslinking time, s
θ_{xl}	Fraction of TF crosslinked to binding site
$t_{1/2}$	Residence time of TF binding, s
UAS_G	Upstream activating sequence for *GAL* genes
WCE	Whole cell extract
WT	Wild type

ACKNOWLEDGEMENTS

Graduate school has been the hardest challenge I've faced so far, but it has also been incredibly rewarding and enlightening. I would like to thank my advisor, David Auble, for taking me on as a student and being so supportive throughout my graduate career. I have faced no shortage of challenges and David has done everything he can to help me succeed. I also appreciate his support in pursuing activities outside the lab, as I know advisors do not always encourage this. These opportunities have greatly enriched my time in graduate school, helped me grow as a person, and shaped what I want to pursue as a future career. David has been a great mentor, as a scientist as a person. He truly cares about the people in his lab and pushes us to always keep improving.

I would like to thank my committee members, Jeff Smith, Stefan Bekiranov, and Bob Nakamoto, for their support and help throughout graduate school. Jeff was always willing to answer my questions, especially about yeast genetics, and let me use their lab equipment. Stefan was extremely patient as I learned CLK modeling and offered insightful suggestions at our meetings. Bob let us use the quench-flow apparatus early on before we updated the CLK method.

I would like to thank the Auble lab members, especially Savera and Kristyna. Savera helped me improve my ChIP experiments and did many CLK experiments for me when I couldn't consistently get data and we have shared many conversations on science and life. I was especially glad she was a lab member when it was just the two of us working in lab. Although Kristyna has only been in the lab less than a year, she has become a great friend and a bright presence in the lab. I want to thank her for helping me with the microscopy data analysis and my programming questions. She has been a good person to practice my science communication with, as well as sometime to talk to about anything. I would also like to thank Jason for being my bay-mate, helping to keep us sane, and pushing me to keep improving my science. Even though Kasia and I were only in lab together for a few months, she was a great lab-mate that became a friend. Ramya

was very helpful when we wrote the first CLK methods section has been a wonderful resource recently for post-grad career options.

Thanks to the Charlottesville Flute Ensemble for warmly welcoming me into your group and giving me an artistic outlet and a way to keep playing flute.

I am so happy that I got involved in WIMS and CommUNIty C'ville. These groups are full of driven, wonderful people and they have allowed me to find my own new passions and been great resources for science policy and future career options. I have also gained more confidence in myself (finally) from my participation and for that, I am extremely grateful.

My time in graduate school would have been significantly less enjoyable and more stressful without my friends. Thanks to Ryan, who answered many, many questions, helped me improve protocols, and always focused the dissection scope when I couldn't find the needle. He has been a good friend outside of the lab as well and I'm glad to have someone else to go skating with. Thanks to Naomi for the baked goods and friendship. Jackie has always been incredibly supportive and kind and I am glad to have her as a friend. Thanks to Lindsey, who was my mentor when I first arrived at UVA and has remained a wonderful friend who gives great life advice over delicious meals. A special thank you to Kelly and Alexandra, my biggest support system in Charlottesville. Thank you for the encouragement, advice, and being such strong role models for me.

Although my childhood friends don't really know what I do in lab besides work with yeast, their enthusiasm and encouragement means so much to me. Thank you to Kait, Kate, Allison, Vincenza, and Tricia for always being there for me and growing into strong, inspiring women.

Thank you to my family for your support and encouragement, I hope I've made you proud. Thanks to my cousins for being even more excited than I am that I'll be the first PhD in the Child family. My parents instilled in me a love of science and are also wonderful human beings. I am grateful for their unending support, especially in the beginning of grad school.

Thanks to Ben for always being curious and encouraging about my research and Stacey for the support. Last, but not least, thank you to my boyfriend, Andrew. Even though he's 500 miles away, he does everything in his power to support me, whether that be helping me troubleshoot a procedure or sending me Cheetos during quals. I appreciate his patience, encouragement, and sacrifices.

CHAPTER I

Introduction

Overview of transcription

The central dogma of biology is that information in our genes flows from DNA to RNA to protein (Crick 1958). The process of RNA production from DNA, known as transcription, is mediated by RNA polymerase (Weiss and Gladstone 1959), which is required for gene expression across life forms (Struhl 1999, Ptashne 2005, Jun et al. 2011, Koster et al. 2015). In eukaryotes, there are three different polymerases, I, II, and III, which each mediate transcription of a specific type of RNA (Roeder and Rutter 1969) and associate with their own specific factors. However all three types of polymerase require some form of TBP and TFIIB (Vannini and Cramer 2012). RNA polymerase II (Pol II)-mediated transcription is the focus of the studies in the following chapters.

Gene expression is regulated by the specific interactions of transcription factors (TFs) with promoters and regulatory regions (Figure 1.1). These patterns can vary across cell types and in different cellular conditions and are intimately linked to chromatin structure (Kim et al. 2005, Consortium 2012, Rhee and Pugh 2012, Dowen et al. 2014). The first step of transcription is initiation, involving coordinated binding and activity of a variety of proteins to regulate specific genes. Activator and/or repressor proteins bind to specific DNA sequences to tag them for activity, with the end goal being recruitment of transcription machinery to the core promoter (Hahn 2004). These proteins find their binding sites by diffusion through nuclei (Hager et al. 2009), with transient chromatin interactions lasting milliseconds to several minutes (Dinant et al. 2009, Hager, McNally et al. 2009, Chen et al. 2014). Longer DNA binding times correlate with transcription output (Lickwar et al. 2012), so most of these interactions are transient, as the necessary components to start transcription have not been fully recruited yet. Another regulatory step in transcription is recruitment of co-activator and co-repressor complexes to DNA, usually either through the presence of activators/repressors or chromatin modifications. Chromatin

regulatory complexes can exchange histone variants or remove histones to open up DNA for transcription (Koster et al. 2015). Each eukaryotic nucleosome consists of two copies each of H3, H4, H2A, and H2B wrapping 147 base pairs of DNA (Luger et al. 1997). Chromatin is very stable *in vitro* and in heterochromatic regions and in Pol II transcribed genes (Svensson 2015) and can allow binding of activators/repressors while preventing binding of the larger Pol II machinery and preinitiation complex (PIC) (Struhl 1999), but nucleosomes near promoters tend to have higher turnover (Guillemette 2005, Zhang 2005, Jin 2007). Nucleosomes can prevent binding of the PIC components, but once bound, the PIC can prevent nucleosome deposition back onto that area of chromatin (Workman and Roeder 1987, Meisterernst et al. 1990, Sikorski and Buratowski 2009, Levine et al. 2014). Once a conducive chromatin environment has been established, the preinitiation complex can assemble on DNA, although this does not necessarily mean it will. The complex consists minimally of general transcription factors TFIIB, TFIID, TFIIF, TFIIE, and TFIIH, Pol II, and sometimes TFIIA. The assembly of the PIC was defined *in vitro* on TATA-containing promoters as a step-wise process (Buratowski et al. 1989), with TFIID (TBP) binding first, followed by stabilization with TFIIA, TFIIB binding, TFIIF/Pol II association, then TFIIE and TFIIH binding (Roeder 1996). As only ~15-20% of the yeast genome contains a TATA sequence (Basehoar et al. 2004), this assembly pathway likely varies *in vivo*. The characteristics and assembly of the PIC will be discussed in more detail in the next sections.

After the PIC has assembled, RNA polymerase II (Pol II) initiates RNA synthesis, escapes the promoter, and elongation ensues. Once the required machinery has been recruited to promoter DNA, enzymatic activity of TFIIH phosphorylates Pol II and initiates transcription by melting the promoter and changing the complex from closed to open, positioning the template DNA strain in the active site cleft of Pol II (Wang et al. 1992). It's important to note that only a small percentage of Pol II binding events lead to productive transcription (Stasevich et al. 2014) (Darzacq et al. 2007). Pausing by Pol II is a frequent event in higher-order eukaryotes and is highly regulated as well (Core et al. 2008, Price 2008, Fuda et al. 2009, Churchman and

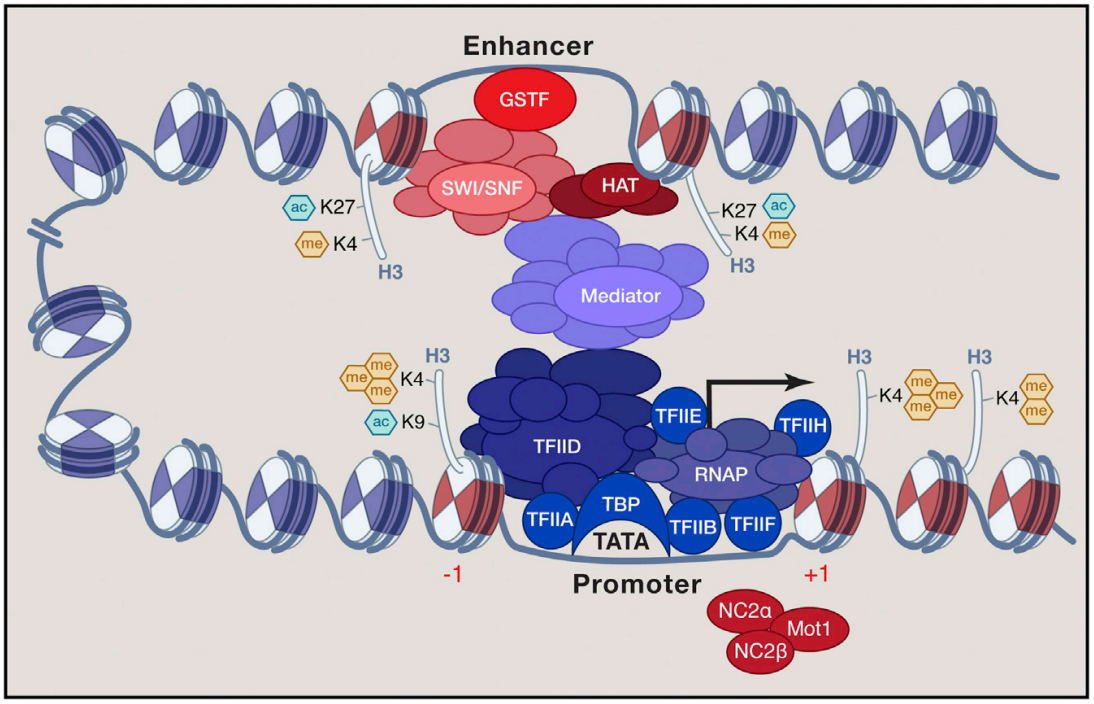


Figure 1.1. Regulation of transcription initiation; from Koster *et al.*, 2015. Overview of transcription. Activator/repressor proteins for specific genes (bright red; gene-specific transcription factors: GSTFs) bind DNA elements such as enhancers to recruit regulatory complexes including Mediator (light purple), histone acetyltransferases (maroon, HAT), and chromatin remodelers (pink, SWI/SNF). Chromatin structure is altered, allowing the preinitiation complex to bind (dark purples and blue; GTFs, RNA polymerase II – RNAP). TFIID/TBP binds the TATA-box in the promoter and recruit other general transcription factors (TFIIA/B/E/F/H) and RNA polymerase II. The Mot1 and NC2 proteins can remove TBP from the promoter. Nucleosomes are represented by the wheels wrapped in blue DNA; red colored histones indicate mobile proteins, while the blue histones are more stable.

Weissman 2011, Buckley et al. 2014). For transcription to initiate, the C-terminal domain of Pol II must be hyperphosphorylated at serine residues 2 and 5. These residues cycle through phosphorylation and dephosphorylation as transcription proceeds (Kim and Dahmus 1989, Buratowski and Sharp 1990). In yeast, Cdk7/Kin28 phosphorylates serine 5 to initiate transcription (Serizawa et al. 1993) and later recruits the guanylyltransferase enzyme for mRNA capping (Cho et al. 1997, Rodriguez et al. 2000, Pei et al. 2001). Typically, multiple short RNAs (3-10 bases) are synthesized before Pol II is fully engaged in initiation and transcribes full-length RNAs (Luse and Jacob 1987, Holstege et al. 1997). Once ~30 bases of RNA are synthesized, Pol II is thought to release the core promoter and enter the elongation stage of transcription (Hahn 2004). The general elongation factor TFIIS can be recruited to the PIC at select promoters *in vitro* by TBP, Mediator, and SAGA and can bind TFIIF and Pol II, potentially aiding in stabilization of the PIC (Kim et al. 2007). TFIIS aids in restarting arrested Pol II and promoting elongation and can be recruited to these stalled polymerases by the Ccr4-Not elongation factor complex (Dutta et al 2015). Other factors relating to RNA processing, synthesis, and export can be recruited to elongating Pol II (Bentley 2002, Hahn 2004). Serine 2 phosphorylation by C-terminal domain kinases including P-TEFb leads to recruitment of 3'-end processing factors (Komarnitsky et al. 2000, Ahn et al. 2004). In yeast, dephosphorylation of serine 5 is done by Ssu72 (Krishnamurthy et al. 2004), while serine 2 dephosphorylation is carried out by TFIIF-associated C-terminal domain phosphatase 1 (FCP1) in yeast (Archambault et al. 1998) and can move reinitiation along by transforming elongating Pol II back to initiating Pol II (Chambers et al. 1995). Following elongation, termination and transcription-coupled mRNA processing occur and the whole process can begin again.

Transcription factors locate their specific chromatin binding sites by diffusion through the nucleus (Misteli 2001, Kampmann 2005, Gorski et al. 2006), interacting transiently with non-specific sites until reaching a specific site to which the TF binds more stably and elicits a biological response (Misteli 2001, Phair et al. 2004, von Hippel 2007). The TF binding site

search process has been the subject of intense interest and is an area of active investigation (Slutsky and Mirny 2004, Halford 2009, Mirny et al. 2009, van Royen et al. 2011, Hammar et al. 2012, Wang et al. 2013, Chen et al. 2014). Dynamic TF binding correlates with transcriptional output and is consistent with the stochastic nature of gene expression itself (Kepler and Elston 2001, Raser and O'Shea 2005, Cai et al. 2006, Darzacq et al. 2007, Gorski et al. 2008, Larson et al. 2011). In developing molecular models for how transcription complexes assemble and function in cells, a key goal is to determine how TF-chromatin binding dynamics at particular promoters relate to the stochastic properties of RNA synthesis *in vivo* (Hager et al. 2009, Stasevich et al. 2014). Interestingly, some TF-chromatin interactions may be exceedingly long-lived, particularly in the context of the cell cycle or differentiated state (Chen et al. 2002, Giglia-Mari et al. 2009). A goal in measuring these dynamic parameters is to ultimately determine in detail how they correlate with transcriptional output, which is a largely unexplored area. In the Introduction sections that follow, the components of the preinitiation complex are discussed, as well as current understanding of their regulation, assembly, and binding dynamics to set up the experimental questions discussed in the subsequent chapters.

Components of the preinitiation complex

The preinitiation complex consists of general transcription factors (GTFs), including the TATA-binding protein (TBP) and TFIIA/B/D/E/F/H, and RNA polymerase II (Pol II) (Figures 1.1, 1.2) and, as the name suggests, assembles *in vitro* in order to start the transcriptional initiation step at specific sites. In both mammalian and yeast cells, the minimal factors required are TBP, TFIIB, TFIIF, TFIIE, and TFIIH, along with Pol II (Luse 2013). Evidence for these GTFs was found when purified Pol II and crude subcellular fractions from human cells were able to transcribe the adenovirus DNA *in vitro* (Weil et al. 1979) and the individual factors required for basal transcription were identified through multiple fractionation steps of nuclear extracts (Matsui et al. 1980, Samuels et al. 1982, Sawadogo and Roeder 1985, Reinberg and Roeder 1987,

Flores et al. 1989, Flores et al. 1992, Sayre et al. 1992, Ge et al. 1996). Further *in vitro* work was done to characterize each isolated factor and determine an *in vitro* assembly pathway (discussed in the next section). Many crystal structures of partial PICs have been experimentally determined, but a recent paper reported cryo-EM models with the previously elusive TFIIH subcomplex (Figure 1.2) (Schilbach et al. 2017).

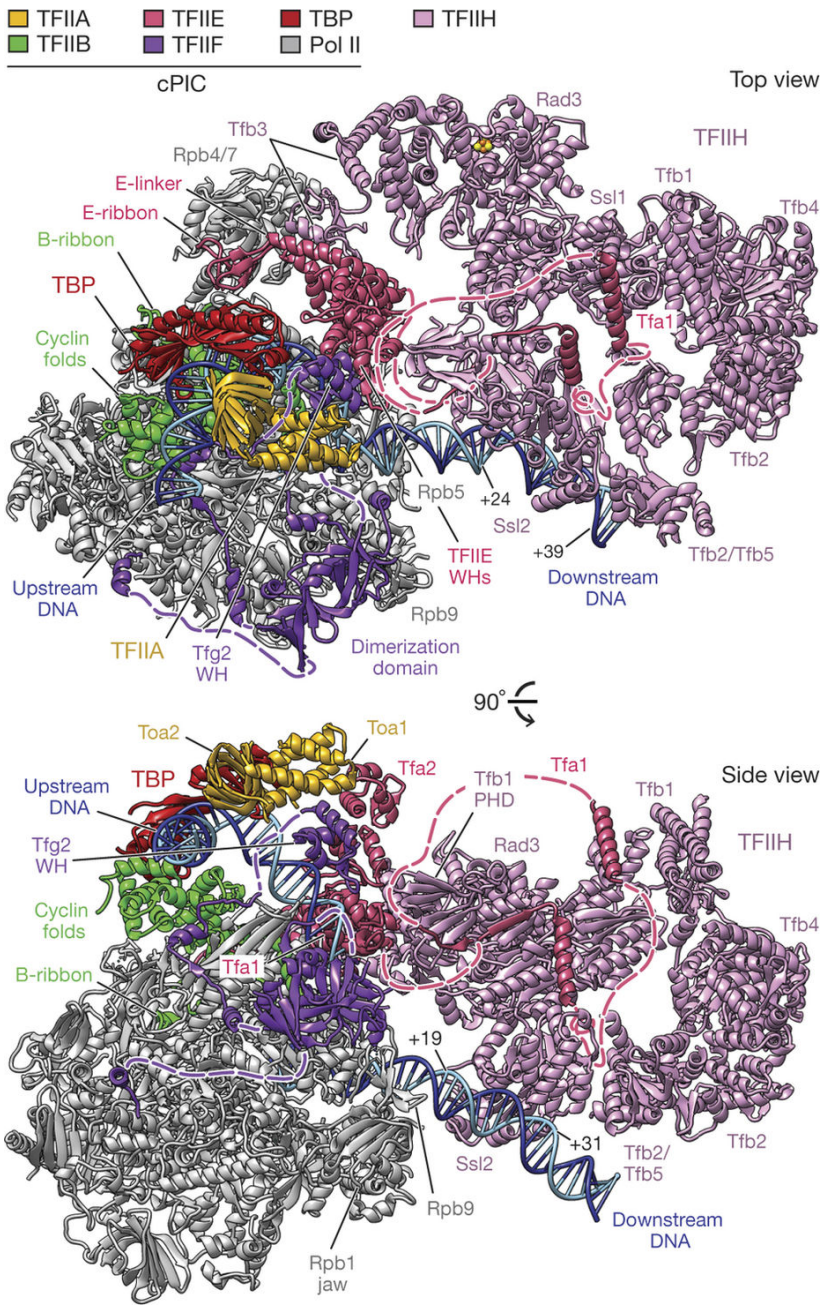


Figure 1.2. Two views of the cryo-EM structure of the yeast PIC; from Schilbach *et al.* 2017. Structures for all PIC components are shown. DNA (aqua and blue) is bound and bent by TBP (red), TFIIA (yellow, Toa1/Toa2), TFIIB (green), TFIIF (purple, only Tfg2 is shown and not Tfa1), Pol II (silver), TFIIE (pink, Tfa1/Tfa2), and TFIIH (light purple, Rad3/Ssl1/Ssl2/Tfb1/Tfb2/Tfb4/ Tfb5). Flexible linkers in TFIIE and TFIIF are indicated by dashed lines. The bottom image is a 90° rotation of the top image.

The PIC components in this section are introduced in the order they are thought to assemble *in vitro* on a TATA-containing promoter (Buratowski et al. 1989). This is one possible order of assembly, as discussed in the next section. The *in vivo* assembly of PIC components is unknown and could vary depending on a variety of factors, such as the promoter class; determining how the PIC assembles genome-wide is a major goal of our research and will be discussed further in future chapters. .

The first component of the PIC to assemble *in vitro* is TFIID, which can be minimally represented by TBP for PIC formation. TFIID consists of TBP and 13 TAFs (TBP-associated factors) (Papai et al. 2011). TBP binds the TATA box upstream of the transcription start site, which contains a consensus sequence TATA(A/T)A(A/T)(A/G) (Basehoar et al. 2004) ~30 base pairs upstream of the start site in mammalian cells (Ponjavic et al. 2006) and ~60-120 base pairs upstream in yeast; there can be multiple start sites within this region (Guarente 1987, Struhl 1987). Once TBP has bound the TATA box, it bends the DNA at approximately a 90° angle by inserting two phenylalanine pairs between the first and last di-nucleotides of the TATA box (Horikoshi et al. 1992, Kim et al. 1993, Kim et al. 1993, Delgadillo et al. 2009). At promoters that do not have a TATA box, TFIID is stabilized by TAFs binding to DNA around the promoter (Burke and Kadonaga 1997, Thomas and Chiang 2006) or binding to acetylated and methylated histone tails (Jacobson et al. 2000, Vermeulen et al. 2007). TFIID contains three lobes (A-C) in

the shape of a horseshoe, with TBP in the center lobe on the inside (Andel et al. 1999). TFIIB interacts with TBP and is able to bind to the DNA template as well as the catalytic subunit of Pol II (Sainsbury et al. 2013). TFIIB also plays a role in selecting the transcription start site (Pinto et al. 1994, Pardee et al. 1998, Hawkes and Roberts 1999) through its B-finger domain, which can work with Pol II/TFIIF for start site selection (Thomas and Chiang 2006).

TFIIF is a Pol II binding factor (Sopta et al. 1985) and is thought to help load Pol II into the PIC. Additionally, TFIIF is thought to help stabilize TFIIB in the PIC (Eichner et al. 2010, Fishburn and Hahn 2012, Zhang et al. 2016). Recent work found that TFIIF is not always associated with the PIC; TFIIF modified by casein kinase 2 (CK2) is present in the PIC, but phosphorylated TFIIF is not a stable PIC component (Cabart et al. 2011), which should be kept in mind when interpreting TFIIF dynamics in the future. It is possible that TFIIF helps load Pol II into the PIC, but once this has been accomplished it undergoes a conformational change and dissociates from the PIC.

One of the roles of TFIIE is to act as a loading factor for TFIIH (Maxon et al. 1994, Ohkuma et al. 1995), but it can help with promoter melting at supercoiled DNA templates without the help of TFIIH (Holstege et al. 1995). The corresponding factor in Archaea is called TFE, which is able to facilitate promoter melting without the help of a TFIIH-like subunit (Grohmann et al. 2011).

The GTF TFIIH has enzymatic activity, containing two helicases (Ssl2p and Rad3p in yeast, XPB and XPD in humans) and a kinase (Kin28p in yeast, Cdk7 in humans) (Luse 2013). One of the functions of this factor is to melt the promoter DNA, as Pol II is not able to separate the template and non-template DNA strands alone (Tirode et al. 1999). At supercoiled promoters, superhelical tension can cause spontaneous unwinding of the DNA and TFIIH is not required (Pan and Greenblatt 1994, Luse 2013). TFIIH interacts with the DNA downstream of the rest of the PIC complex, although the exact location is not well defined (Kim et al. 2000, He et al. 2013).

RNA polymerase II consists of 12 subunits (Rpb1-Rpb12) in both humans and yeast, with Rpb1 as the largest subunit by molecular mass (Young 1991). Five of the subunits (Rpb5, Rpb6, Rpb8, Rpb10. Rpb12) are part of all three polymerases and Rpb1, Rpb2, Rpb3, and Rpb11 have a PolI and Pol III homologous counterpart. Only the remaining three subunits (Rpb4, Rpb7, and Rpb9) and the C-terminal donation of Rpb1 are present in Pol II alone (Thomas and Chiang 2006). Bacterial and archaea only have one polymerase, but there is similarity between some of the Pol II subunits with the other polymerases, indicating a similar structure and mechanism of action (Elbright 2000). Pol II consists of four elements: the core, clamp, shelf, and jaw lobe. The core contains the regions of Rpb1 and Rpb2 that comprise the active center and all four elements form the cleft where DNA enters and the active site is located. The clamp is able to move in order to open and close the cleft (Hahn 2004). Structural studies indicate that double-stranded DNA is not inserted into the cleft and only single-stranded template DNA is fed into the cleft and reaches the active site (Bushnell and Kornberg 2003, Hahn 2004).

Potential preinitiation complex assembly pathways

Since the process of transcriptional regulation requires the coordinated action of many proteins, the order in which they assemble is important for understanding how regulators control initiation. There are two general schemes for assembly of transcription complexes on promoters: random and regulated (step-wise) assembly (Figure 1.3) (Hager et al. 2009). The hierarchical step-wise assembly pathway of the PIC components was defined *in vitro* via native gel electrophoresis and DNase I footprinting using a TATA-containing promoter (Buratowski et al. 1989). It was proposed that the first step in the pathway was recognition of the TATA element by TFIID, followed by addition of TFIIA on the side upstream of TFIID to stabilize the complex. Interestingly, TFIID binding to a cryptic TATA site was enhanced more than binding to the TATA element upon TFIIA addition, indicating that TFIIA may not be necessary at all promoters (Buratowski et al. 1989). TFIIB is thought to be the next factor to join the complex. When

purified Pol II was added to binding mixtures, no new complexes were generated unless TFIIB and TFIID were present, indicating the role of TFIIB in potentially bridging TFIID and Pol II in the complex. Addition of purified Pol II was necessary for formation of larger complexes, indicating Pol II as the next step in the complex (Buratowski et al. 1989). The authors note that there was another complex formed with Pol II addition, shown by a difference in mobility on the gel, but they did not know that TFIIF was the factor responsible for this (Flores et al. 1988). TFIIE is the next factor to join the complex. A loss of DNase I protection was observed when ATP/dATP was added to the solution, indicating ATP-dependent dissociation of TFIIE. The authors found that TFIIA could not necessarily be detected in the complexes formed after Pol II was added (Buratowski et al. 1989). This canonical pathway agrees with several other studies as well (Van Dyke et al. 1988, Sayre et al. 1992, Zawel and Reinberg 1992, Conaway and Conaway 1993, Roeder 1996). Later studies identified binding of the pre-formed Pol II-TFIIF complex after addition of TFIIB (Zawel et al. 1995, Leuther et al. 1996) and TFIIH binding as the last step to completing PIC assembly (Conaway and Conaway 1993, Zawel and Reinberg 1993, Zawel et al. 1995).

A recent study using an *in vitro* reconstituted system and cryo-electron microscopy recapitulated this assembly pathway on an optimized TATA-containing promoter using TBP and TFIIA/B/E/F expressed and purified from *E. coli* and Pol II and TFIIH from HeLa cell nuclear extract (He et al. 2013). Another *in vitro* study using a single-molecule system examined TFIIB dynamics and found that TFIIB was stably loaded onto TATA-containing promoters only after TBP, TFIIA, and Pol II/TFIIF were added (Zhang et al. 2016). Without Pol II/TFIIF, TFIIB bound transiently, indicating the necessity of Pol II/TFIIF binding after TFIIB (Zhang et al. 2016). The results from this study were in agreement with the proposed step-wise assembly of the PIC, at least up until Pol II/TFIIF addition. However, another recent study using the *HIS4* promoter and cryo-EM found an abundance of a PIC containing TBP, TFIIE, TFIIH, and DNA in yeast. When TFIIB and Pol II were added, the full PIC was generated (Murakami et al. 2013).

This is in contrast to the earlier defined canonical step-wise assembly, suggesting formation of a partial PIC, and resembles the scaffold complex proposed by the Hahn lab as a reinitiation intermediate (Yudkovsky et al. 2000). Another study at the heat shock genes also supported a partial assembly model, but TFIIH was found to be absent at some promoters, while TBP and TFIID/B/E/F were present in a complex (Zanton and Pugh 2006). Some genes acquired TFIIH downstream, indicating differential recruitment of this factor.

Additionally, a FRAP study examined the *in vivo* behavior of several PIC components by looking at TBP, TFIIB, TAF1, and the Pol II subunit Rpb1 (Sprouse et al. 2008). If factors were pre-assembled before binding to DNA, they would have identical recovery curves. However, this was not the case; TFIIB had the fastest recovery followed by TAF1, TBP, and Rpb1, so these factors may assemble individually. Other evidence suggests that components can arrive at a promoter on different time scales, but only occasionally form a full complex (Hager et al. 2009). This random assembly model is supported by another study using Pol I as well (Dundr et al. 2002).

The pathway leading to PIC assembly *in vitro* is rather well understood (Orphanides et al. 1996, Roeder 1996, Hahn 2004, Vannini and Cramer 2012); however, the original studies of PIC formation were done *in vitro* at genes with TATA sequences. Only about 20% of the yeast genes have a TATA sequence, leading to the questions of whether regulation is different at genes that do or do not contain a TATA consensus sequence and how PICs form on TATA-containing promoters *in vivo*. Recent studies have provided evidence for the stepwise assembly of the PIC (He et al. 2013) as well as alternative pathways where PICs are partially assembled and require addition of the remaining core components to initiate transcription (Thomas and Chiang 2006, Zanton and Pugh 2006). Many questions still remain regarding PIC assembly, which explains the motivation for our ChIP-based studies. In theory, the binding dynamics of different PIC components can be compared and an assembly pathway can be modeled. For example, if factor A had a long residence time and factor B had a shorter residence time, A could potentially bind

before B in a linear pathway; however there could be more complicated cases where A binds first and leaves before B binds. If A and B have approximately the same residence time, they could pre-assemble as a complex before binding. The effective dynamic measurements we make with the ChIP-based approaches would need to be combined with structural data, modeling, and genetic manipulation studies to rule out certain cases and to model assembly pathways.

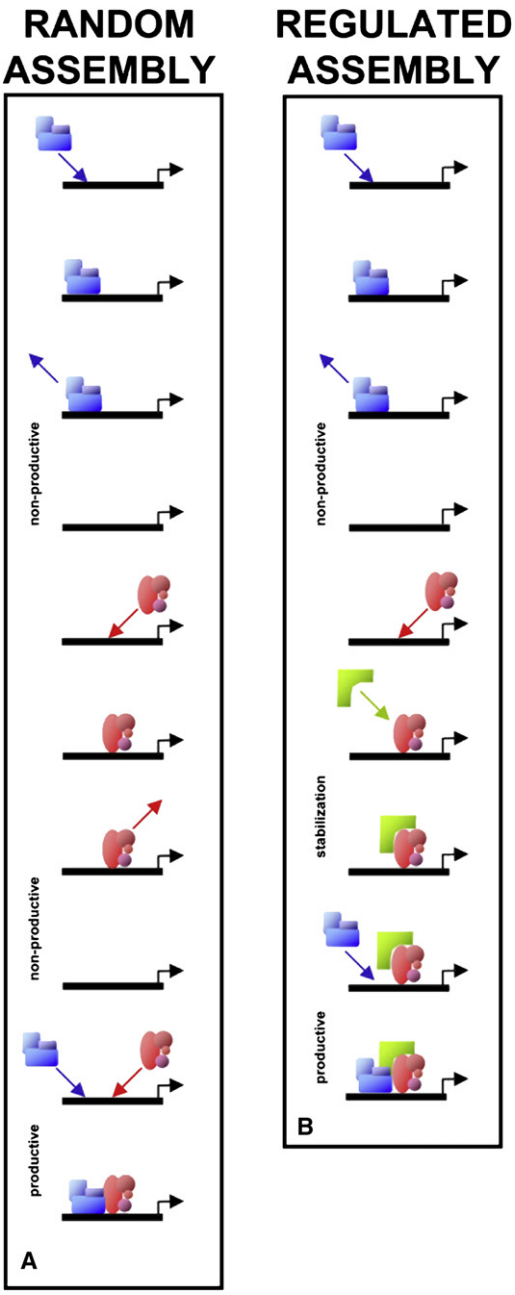


Figure 1.3. Potential complex assembly models; From Hager *et al.*, 2009. A) Subcomplexes (red and blue) dynamically associate and dissociate from a promoter (black line) upstream of the transcription initiation site (black arrow). Assembly of the full complex happens randomly when both subcomplexes happen to bind the promoter simultaneously. B) Binding of the green shape stabilizes the red subcomplex, allowing the blue subcomplex to stably bind and leads to full complex formation. Without the presence of the red or green subcomplexes, the blue subcomplex dissociates from the promoter.

Regulation of the preinitiation complex

Along with the general transcription factors and Pol II, transcription is dependent on a variety of other regulatory proteins, including chromatin remodelers, histone (de)acetylases, and the Mediator complex. ATP-dependent remodeling proteins, like SWI/SNF and CHD1, bind to chromatin modifications on genes to catalyze nucleosome movement or eviction so the PIC can be assembled (Hargreaves and Crabtree 2011). In human cells, the p300 histone acetyltransferase works with Mediator, a regulatory complex that interacts with activators and the GTFs, by acetylating itself and chromatin, leading to p300 dissociation followed by binding of TFIID to Mediator, allowing PIC formation (Black et al. 2006, Black et al. 2008). The SAGA complex, which is an H3 histone acetyltransferase and transcriptional coactivator (Grant et al. 1997, Baker and Grant 2007), is often recruited to the PIC.

The Mediator complex, which acts as a bridge between activators and GTFs, consists of over 20 subunits and was first discovered in yeast (Kim et al. 1994) as a global regulator of gene expression (Kelleher et al. 1990, Flanagan et al. 1991, Poss et al. 2013). Mediator is generally required for transcription and is an interface between DNA-bound transcription factors and the PIC (Borggrefe and Yue 2011). The complex consists of four main modules: Cdk8/kinase (Cdk8, CCNC, Med12, and Med13), head (Med6, Med8, Med11, Med17, Med18, Med19, Med20, and

Med22), middle (Med1, Med4, Med7, Med9, Med10, Med21, and Med31), and the tail (Med2, Med3, Med5, Med14, Med15, and Med16) (Allen and Taatjes 2015). Due to a large number of intrinsically disordered domains, the Mediator complex is structurally dynamic (Toth-Petroczy et al. 2008). Because Mediator is a structurally flexible complex and can alter its subunit composition, it may be able to differentially mediate gene expression depending on the specific gene (Poss et al. 2013). For example, the Med17 and Med21 genes are required for almost all protein-coding gene expression (Thompson and Young 1995, Holstege et al. 1998), while other subunits affect very specific sub-sets of genes. Part of this functional specificity is due to the fact that different transcription factors bind different Mediator subunits, leading to differing outcomes (Poss et al. 2013). The subunit Med15 interacts with Gal4 (Park et al. 2000), which explains why *GAL* gene expression is altered when the Med15-Gal4 interaction is disrupted (Sakurai et al. 1993).

Specific physical or functional interactions have been shown between the Mediator and each of the PIC components (Poss et al. 2013). Mediator coordinates TFIID binding to promoter DNA (Johnson et al. 2002) and plays a large role in recruitment of TFIIB, TFIID, and TFIIE to promoters, even in basal transcription (Baek et al. 2006). Mutations in Mediator subunits have caused defects in TFIID recruitment (Lim et al. 2007, Takahashi et al. 2009). The tail subunit Med15 has been implicated in stable binding of TFIIE and TFIIH (Sakurai and Fukasawa 1997, Badi and Barberis 2001). Mediator is also known to bind the unphosphorylated C-terminal domain of Pol II with high affinity (Myers et al. 1998, Näär et al. 2002) and can enhance TFIIH-mediated phosphorylation of this domain (Kim et al. 1994). The Med11 and Med19 subunits have both been shown to interact with TFIIH (Baidoobonso et al. 2007, Esnault et al. 2008). Mediator has also been found to remain at the promoter with some of the GTFs following transcriptional initiation, consistent with the formation of a stable scaffold that facilitates transcription reinitiation (Yudkovsky et al. 2000). Mediator can bind to either Pol II or the Cdk8 domain, but not both at the same time, and large structural changes occur in the complex depending on

whether Poll or Cdk8 is bound (Poss et al. 2013, Petrenko et al. 2016). The preference of Mediator for the kinase domain or Pol II can help regulate transcription initiation, as dissociation from Cdk8 is required for Pol II association with the PIC. This switch may occur generally due to intrinsic changes within the complex or by specific factors, such as the poly(ADP-ribose)-polymerase PARP-1 (Petrenko et al. 2016), which has been shown to mediate between the two Mediator forms at retinoic-inducible promoters (Pavri et al. 2005).

SAGA (Spt-Ada-Gcn5 acetyltransferase) is a coactivator complex that influences the chromatin landscape and recruits PIC components to initiate transcription (Koutelou et al. 2010). It is comprised of 20 subunits in yeast with several functional units including a Tra1 recruitment module (Grant et al. 1998, Brown et al. 2001), acetylation module (Gcn5, Ada2, Ada3) (Horiuchi et al. 1995, Sterner et al. 1999, Balasubramanian et al. 2002), and TBP interaction subcomplex (Spt3, Spt8) (Sterner et al. 1999, Bhaumik et al. 2004, Mohibullah and Hahn 2008). SAGA does not directly bind DNA, but is recruited to specific genes through Tra1 interaction with activators (Brown et al. 2001) and Gcn5 can bind acetylated H3 and H4 tails through its bromodomain (Owen et al. 2000). The acetylation module can then acetylate histone H3 to open up the chromatin and allow other factors and the PIC to assemble (Balasubramanian, Pray-Grant et al. 2002). The SAGA subunit Spt3 is required for TBP recruitment to promoters, but not for SAGA recruitment (Bhaumik and Green 2001, Larschan and Winston 2001), suggesting that SAGA binding to promoters contributes to TBP binding and not vice versa. Following transcription initiation, a partial SAGA complex without Spt8 stays with Pol II as it transcribes, acetylating and removing nucleosomes as Pol II moves through the gene (Govind et al. 2007). SAGA also plays a role in elongation, as the subunits Ubp8, Sus1, and Sgf11 deubiquitinate H2B, which leads to recruitment of Ctk1 and phosphorylation of serine 2 on the C-terminal domain of Pol II (Wyce et al. 2007).

In addition to large regulatory complexes that help recruit individual components of the PIC, there are also proteins that regulate TBP, such as the Swi2/Snf2 family member modifier of

transcription 1 (Mot1), which dissociates TBP in an ATP-dependent manner (Auble and Hahn 1993, Auble et al. 1994, Adamkewicz et al. 2000). Mot1 has been shown to both negatively and positively regulate transcription by removing TBP from DNA using ATP hydrolysis. TBP can either be removed from correct binding sites to repress transcription, or be removed from non-specific sites to allow for correct binding and promote transcription through subsequent PIC assembly (Madison and Winston 1997, Muldrow et al. 1999, Poorey et al. 2013, Zentner and Henikoff 2013). This activity is responsible for the dynamics of TBP genome-wide (Sprouse et al. 2008) and helps to regulate activated transcription (Auble 2009). In addition to its role in initiation, Mot1 has been shown to play a role in elongation (Poorey et al. 2010).

Another regulator of TBP is negative cofactor 2 (NC2), a factor that has been shown to block PIC assembly following TBP binding to DNA in vitro (Meisterernst and Roeder 1991) and it also allows TBP to relocate along the DNA (Schluesche et al. 2007). Similar to Mot1, NC2 is known as a global transcriptional repressor, but also has a role in gene activation (Prelich 1997, Geisberg et al. 2001, Dasgupta et al. 2002). NC2 can bind TBP and prevent other GTFs from binding and it can also move TBP away from core promoters by diffusing along the DNA (Auble 2009). Gene activation can potentially be aided by NC2 stabilizing weak TBP-DNA interactions (Cang and Prelich 2002) or by moving TBP along DNA to other promoters. Both Mot1 and NC2 have been found in high proportions at active promoters (Geisberg et al. 2002, Zanton and Pugh 2004, Huisinga and Pugh 2007) and can both be present at the same promoter at the same time (van Werven et al. 2008, Butryn et al. 2015), both working to limit transcription at activated genes (Auble 2009).

Only about 15-20% of the yeast genome contains a TATA box (Basehoar et al. 2004) but the “TATA-less” promoters still require TBP (Pugh and Tjian 1991). It was thought that these two types of promoters direct different pathways for recruitment of TBP and/or PIC assembly. The Pugh lab study suggests that genes under selective pressure that need to be more adaptable, such as genes implicated in response to different environmental stressors, are mainly TATA-

containing, while the housekeeping genes are generally TATA-less. Additionally, TATA-containing genes tend to undergo more regulation by chromatin remodelers compared to the TATA-less genes (Basehoar et al. 2004). Using a temperature sensitive strain for *taf1*, a subunit in TFIID, it was found that the majority of the TATA-containing promoters were TAF1-independent, while most of the TATA-less promoters were TAF1-dependent (Basehoar et al. 2004). A mutant strain for the SAGA subunit Spt3 showed the opposite results: TATA-containing genes were SAGA-dependent and TATA-less genes SAGA-independent (Basehoar et al. 2004). This is not entirely surprising since TFIID is needed to deliver TBP to TATA-less promoters, but other studies have shown that SAGA and TFIID can partially compensate for loss of the other at most genes, so there is some fluidity in the genes they target (Lee et al. 2000, Huisinga and Pugh 2004). While in general, the TATA-containing genes represent the highly regulated stress-response genes and the housekeeping genes are predominantly TATA-less, it's important to note that regulation of both sets of genes involves both TATA-containing and TATA-less genes, not only one or the other, as well as factors like TFIID and SAGA (Basehoar et al. 2004); the thinking in the field has been that these two classes are more "exclusive" instead of "dominant" (Taatjes 2017). A difference in nucleosome positioning and TBP turnover rates has been implicated in the two promoter classes (van Werven et al. 2009). However, work from the Hora and Hahn labs has revisited the question of TFIID v. SAGA regulation by measuring newly synthesized RNA in depletion strains to determine the dependence of different genes on each regulator. They found that both SAGA and TFIID are recruited genome-wide and are required for Pol II transcription (Baptista et al. 2017, Warfield et al. 2017). The class distinctions made are still important however, as they may reflect other regulatory process happening at distinct promoter classes, such as the influence of Mediator (Taatjes 2017). Under stress conditions there could also be more of a distinction in class regulation. While Baptista et al. (2017) looked at the transcriptional dependence of Pol II on TFIID under stress conditions using depletion strains, Warfield et al. (2017) did not look at SAGA dependence under these same stress conditions. Heat

shock lowered the dependence of transcription on TFIID by ~1.8 fold, but this change was approximately equal across all classes (traditional SAGA- v. TFIID-regulated, TATA consensus v. no TATA consensus) (Baptista et al. 2017). There may be more of a dependence on SAGA at the stress-response genes under stress conditions even though TFIID is still involved at these genes, but this remains to be tested.

Current understanding of preinitiation complex dynamics

Transcriptional regulation depends on the correct timing of specific factors binding to chromatin to control gene expression. As discussed in the previous section, transcription factor binding can be influenced by a variety of factors in response to environmental cues to adjust cellular processes accordingly. It's important to know how long these factors are bound to chromatin in order to understand how they assemble and work together to regulate transcription. Earlier *in vitro* studies suggested that some of the GTFs, including TBP, remained stably bound after transcription initiation (Zawel et al. 1995, Yudkovsky et al. 2000, Hahn 2004). However, recent work has found that TBP is actually highly dynamic *in vivo* (Sprouse et al. 2008, Auble 2009, van Werven et al. 2009, de Graaf et al. 2010, Zhang et al. 2016). This finding is not limited to TBP, as other factors have also been shown to transiently interact with chromatin *in vivo* (McNally 2000, Dion et al. 2007, Karpova et al. 2008, Hager et al. 2009, Lickwar et al. 2012, Morisaki et al. 2014) when some were previously observed to stably interact with DNA *in vitro* (Perlmann et al. 1990). It is important to develop methods to capture and measure dynamics of protein-chromatin interactions *in vivo* to answer biological questions regarding the stability and assembly of these factors and how that correlates to transcriptional output at individual loci.

There are two broad categories of methods for measuring *in vivo* chromatin binding dynamics: imaging approaches and chromatin immunoprecipitation (ChIP)-based approaches (Figure 1.4). Live cell imaging approaches include FRAP (fluorescence recovery after photobleaching), FLIP (fluorescence loss in photobleaching), FCS (fluorescence correlation

spectroscopy), TICS (temporal image correlation spectroscopy), and SMT (single molecule tracking) (Mueller et al. 2013); these techniques have been employed for measuring TF mobility utilizing fluorescent tags on TFs of interest (Becker 2002, de Graaf, Mousson et al. 2010). With the possibility of engineering transcription factors with fluorescent labels, FRAP studies were able to view the movement of labeled factors in live cells, providing evidence that binding events were highly dynamic (Hager et al. 2009). FRAP measures the rate of recovery of fluorescently tagged proteins into a photobleached area of the cell. The resulting recovery curves represent the ensemble dynamics of all tagged proteins rather than the binding behavior at a particular chromosomal locus. The measurements can be modeled to extract the average length of binding events, length of diffusion events, and the average of the on and off rates (Sprague et al. 2004). Since the implementation of FRAP, other fluorescent techniques have been employed, such as FCS, which measures fluorescence fluctuations as fluorescently tagged proteins move in and out of a sub-micron sized volume (Mueller et al. 2013). This technique is able to make measurements on the sub-millisecond scale, which is much faster than the scale detectable by FRAP (Chen et al. 2008). FRAP, FLIP, and SMT all have resolution on the millisecond scale (Mueller et al. 2013).

The FRAP technique can be used to distinguish different DNA binding rates, as proteins that do not bind DNA return to the laser-bleached spot quickly, but proteins that are part of a complex or bind DNA have a slower recovery; general assembly patterns can be inferred by comparing these measurements. The recovery rate of proteins can be examined under different experimental conditions, which was done for several of the general transcription factors (Sprouse et al. 2008). FRAP curves for Pol II, TBP, TFIIB, the TFIID component TAF1, and the Pol II large subunit Rpb1 were measured. The individual components all had different recovery rates, with TBP recovering in about 15 seconds, indicating the highly dynamic binding nature of this factor. Recovery rates were also measured with mutated *mot1*, which caused the TBP recovery time to decrease due to loss of the regulatory function of Mot1, and a chromatin-binding compromised

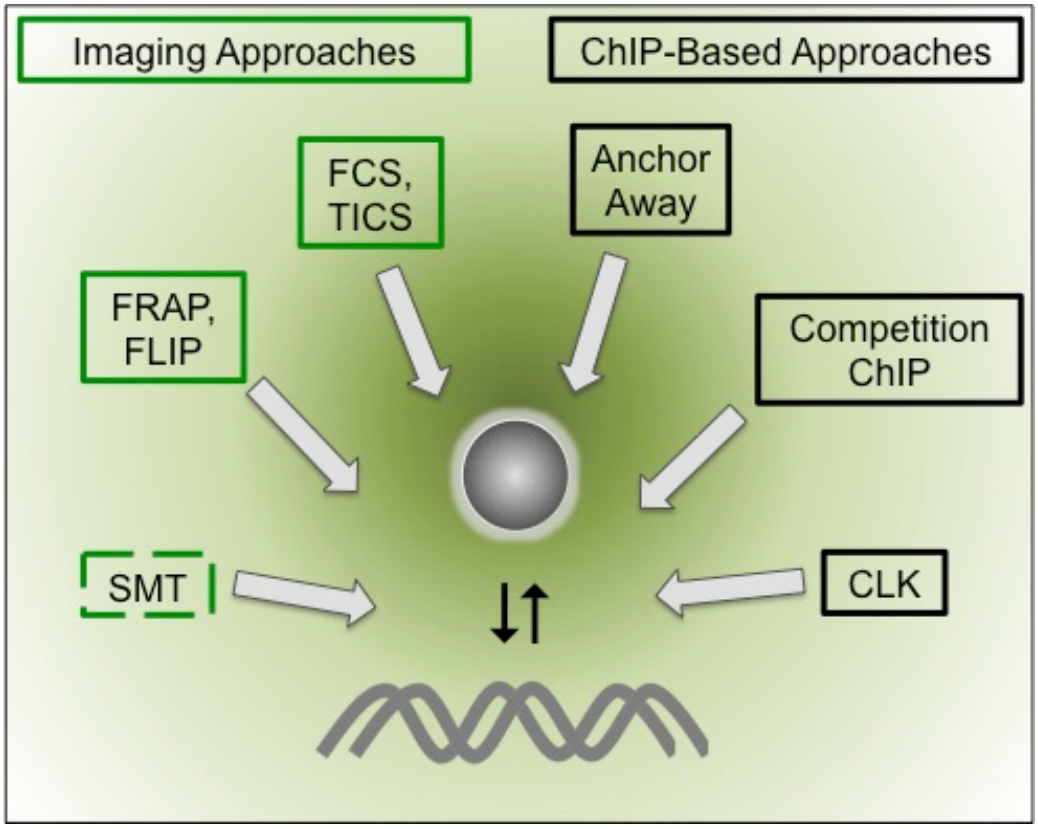


Figure 1.4. Techniques to measure chromatin binding dynamics. Overview of methods used for measuring chromatin-binding dynamics *in vivo*. The gray sphere represents a TF and its reversible interaction with DNA (gray helix) is shown by the black arrows. The various experimental approaches in use are shown by the names and acronyms arrayed around the TF and DNA. See text for explanation of each method and supporting references. The solid gray rectangle highlights SMT as an approach that monitors the behavior of single molecules, which is in contrast to the other methods (bracketed by dashed rectangles), which measure the ensemble behavior of populations of molecules.

TBP mutant V71R, which showed increased recovery time since V71R TBP does not bind DNA (Sprouse et al. 2008). By looking at several factors in the PIC in different mutant strains, the regulation of these factors can be determined based on the effect on binding dynamics when certain mutations are made. This can also provide insight into how the components of the PIC are recruited and stabilized *in vivo*.

FLIP is a technique similar to FRAP, but instead of bleaching an area and measuring the recovery of the fluorescent signal, an area is bleached repeatedly and images are taken between each bleaching step to measured the signal decay (Mueller et al. 2013). The FLIP technique can be combined with FRAP to quantify binding dynamics; two regions in a nucleus are compared simultaneously over time, but only one of them is bleached. This technique is especially useful for determining the residence time of a protein in subnuclear structures (van Royen et al. 2009). While both FRAP and FLIP offer a way to measure the ensemble dynamics of a factor with a fairly long residence time compared with other fluorescent techniques, the protein of interest must be expressed at fairly high levels to visualize, subpopulations of very transient binding interactions can be masked, and these techniques don't yet have the sub-millisecond resolution needed for some binding interactions (Mueller et al. 2013).

In FCS, the fluorescent signal of a certain volume is measured over time. This signal is then usually fit to a diffusion model with two distinct species: one fraction that is freely diffusing and the other fraction representing chromatin binding; this is also how FRAP data is fit (Mikuni et al. 2007). Slower moving factors remain correlated with the earlier measurements since they are not moving out of the defined volume, while more transient factors have less of a correlation (Mueller et al. 2013). TICS is similar to FCS, but looks at all pixels in an image sequence instead of a small volume. Both of these techniques can measure events as fast as a microsecond (Chen et al. 2008), but TICS is preferred when measuring longer binding events because photobleaching is more uniform, which makes it easier to correct for during analysis. Additionally, these techniques work best with low fluorescence, so factors present at low cellular concentrations are

ideal (Chen et al. 2008). However, cell movement is a limiting factor, as correlation is sensitive to small changes in movement of the cell (Mueller et al. 2013).

Unlike the previous fluorescent approaches, SMT is able to track single labeled proteins over time instead of measuring ensemble movements, which allows the full distribution of dynamics to be observed, as some sub-groups can be masked with ensemble measurements (Li and Xie 2011, Mueller et al. 2013). This technique works best with low concentrations of labeled proteins, and because there are no assumptions about protein dynamics in the analysis, this type of data can help inform the model selection for other techniques like FRAP and FCS (Mazza et al. 2012). SMT is best for binding dynamics that occur on the order of seconds (more transient interactions) (Mueller et al. 2013). A recent study utilized a modified version of SMT to quantify binding dynamics of TFIID, TFIIA, and TFIIB to in *in vitro* TATA-containing promoter. TFIIB promoter binding lasted only a few seconds in the absence of Pol II/TFIIF, but bound the DNA for minutes if they were present, giving insight into the assembly process as well as the effect that other transcription factors can have on binding dynamics of another factor (Zhang et al. 2016).

Recently, FRAP, FCS, and SMT results have converged to yield consistent parameters for a few TFs (Mazza et al. 2012, Mazza et al. 2013). The general agreement between these methods provides a critical framework for addressing unanswered questions in the field, particularly those requiring new methods, which will need to be appropriately validated. Under some circumstances it is possible to measure transcription factor binding to a particular native locus *in vivo* (Elf et al. 2007, Karpova et al. 2008) and several studies have used amplified gene arrays to capture transient factor binding (McNally 2000, Yao et al. 2006, Karpova et al. 2008, Rafalska-Metcalf et al. 2010), but it is unknown how this artificial system correlates with the true single copy gene binding dynamics (Mueller et al. 2013). Most imaging studies have inferred factor binding properties without identification of the specific chromatin sites to which the observed factor associates. This limitation has spurred work exploiting the localization precision of the ChIP assay to extract transcription factor binding dynamics information.

The ChIP-based techniques we use to extract transcription factor binding dynamics will be discussed in detail in Chapters II and III. All three methods measure ensemble dynamics at single loci on the time scale of minutes to an hour, with potential to measure dynamics on the seconds scale. As the field moves forward, it will be important to compare fluorescent and ChIP-based approaches to better understand the dynamics of transcription factors. Each method has limitations that another method may be able to compensate for or complement. These methods may also measure different aspects of binding, which will be important to figure out when consensus is not reached between methods. The more information available about binding dynamics of a factor, the more we will know about how it is regulated and works in the process of transcription *in vivo*. By coupling techniques that measure residence time with those that measure dynamics of transcription, we can better understand the molecular mechanisms of gene expression in cells.

Summary

Transcription is a complex and highly regulated process that requires the coordination of a myriad of proteins. One crucial step is establishment of a productive PIC to initiate transcription, including TBP, general transcription factors, and Pol II. The assembly pathway of the PIC *in vivo* is unclear but there is evidence for either a step-wise or random assembly and may differ depending on the gene class. Recent studies indicate that the binding of PIC components to chromatin *in vivo* is highly dynamic, unlike the stable interactions indicated by previous *in vitro* studies. There is a need in the field to develop techniques and measure the protein-chromatin binding dynamics *in vivo* at specific promoters to more fully understand how genes are regulated in real time, which this study addresses.

Scope of this study

While much is known about the biochemical properties of transcription factors, their *in vivo* binding dynamics with chromatin are unclear. Previous studies indicated that transcription factors are highly dynamic *in vivo*. The following chapters in this dissertation describe the development of three different ChIP-based methods to measure chromatin-binding dynamics at single loci: the crosslinking kinetics (CLK) assay, Anchor Away, and Competition ChIP. TBP binding dynamics were measured and compared using these three independent methods to understand the stability of this protein *in vivo* (Chapters II and III). An important aspect of a ChIP-based technique is formaldehyde crosslinking; there is still much unknown about the chemical effects of this molecule *in vivo*. Through improvement of the CLK assay from the previous version, the range of formaldehyde binding dynamics was further investigated, which has implications for many techniques using this reagent (Chapter II).

The *in vivo* binding dynamics for the inducible activator Gal4 are also unknown; Gal4 was thought to very stably bind to DNA but recent studies indicate this interaction is much faster than previously thought. Using CLK and Anchor Away, the binding dynamics of Gal4 were measured and compared to a microscopic approach that measures the kinetics of RNA synthesis (Chapter IV). These dynamics were also measured in strains with a weakened Gal4 binding site and disrupted regulatory complex. By comparing these dynamic measurements, important biological questions can be answered about how activators are regulated, what complexes they recruit or are recruited by, and how their dynamics contribute to transcriptional output.

Taken together, this study shows the dynamic nature of transcription factors and provides insight into how these dynamics are regulated *in vivo*. This opens the door for studying dynamics of other factors involved in transcription to better understand the molecular mechanisms of this highly regulated and complex system.

CHAPTER II

The crosslinking kinetics assay and formaldehyde crosslinking chemistry

Most of the data in this chapter were published in Zaidi *et al.*, 2017 and Hoffman *et al.*, 2015.

Formaldehyde crosslinking underpins many of the most commonly used experimental approaches in the chromatin field, especially in capturing site-specific protein–DNA interactions. Extending such assays to assess the stability and binding kinetics of protein–DNA interactions is more challenging, requiring absolute measurements with a relatively high degree of physical precision. We previously described an experimental framework called the crosslinking kinetics (CLK) assay, which uses time-dependent formaldehyde crosslinking data to extract kinetic parameters of chromatin binding. Many aspects of formaldehyde behavior in cells are unknown or undocumented, however, and could potentially affect CLK data analyses. In this chapter, we report biochemical results that better define the properties of formaldehyde crosslinking in budding yeast cells. These results have the potential to inform interpretations of "standard" chromatin assays, including chromatin immunoprecipitation. Moreover, the chemical complexity we uncovered resulted in the development of an improved method for measuring binding kinetics, which we refer to as CLKv2. Optimum conditions included an increased formaldehyde concentration and more robust glycine quench conditions. Notably, we observed that formaldehyde crosslinking rates can vary dramatically for different protein–DNA interactions *in vivo*. Some interactions were crosslinked much faster than the *in vivo* macromolecular interactions, making them suitable for kinetic analysis. For other interactions, we found the crosslinking reaction occurred on the same time scale or slower than binding dynamics; for these interactions, it was sometimes possible to compute the *in vivo* equilibrium-binding constant but not binding on- and off-rates. Selected DNA

interactions for components of the preinitiation complex displayed dynamic behavior on the minutes time scale with this updated method.

Introduction

Formaldehyde crosslinking and its utility in capturing protein-chromatin interactions

Prior to its use in the chromatin field, formaldehyde use had a long history in a number of fields, including vaccine production (Nencioni et al. 1991, Eckels and Putnak 2003) and histology (Werner et al. 2000). We are most interested in its use in chromatin immunoprecipitation approaches and protein-protein interaction studies to understand the location and abundance of transcription factor binding along DNA. Previous work (see below) has been done to describing formaldehyde reactivity toward proteins, DNA, and their constituent monomers. This information provides a basis for understanding how formaldehyde functions in widely used assays in the chromatin field, and conversely, highlights less well-understood aspects of formaldehyde behavior in cells. These issues are of significance for designing crosslinking-based studies as well as for properly interpreting the resulting data. The analysis of formaldehyde-fixed chromatin has provided fundamental insights into where and when regulatory factors associate with the DNA template *in vivo*, but it in general does not provide unambiguous information about chromatin binding kinetics or an absolute measure of fractional occupancy. A major goal of ongoing work is to understand kinetic and thermodynamic aspects of chromatin complex assembly at single copy loci *in vivo*. Development of experimental strategies to achieve these goals will require a deeper and more comprehensive understanding of the effects mediated by formaldehyde in cells. Figure 2.1 shows the main features of formaldehyde interactions with and trapping of macromolecular complexes in cells.

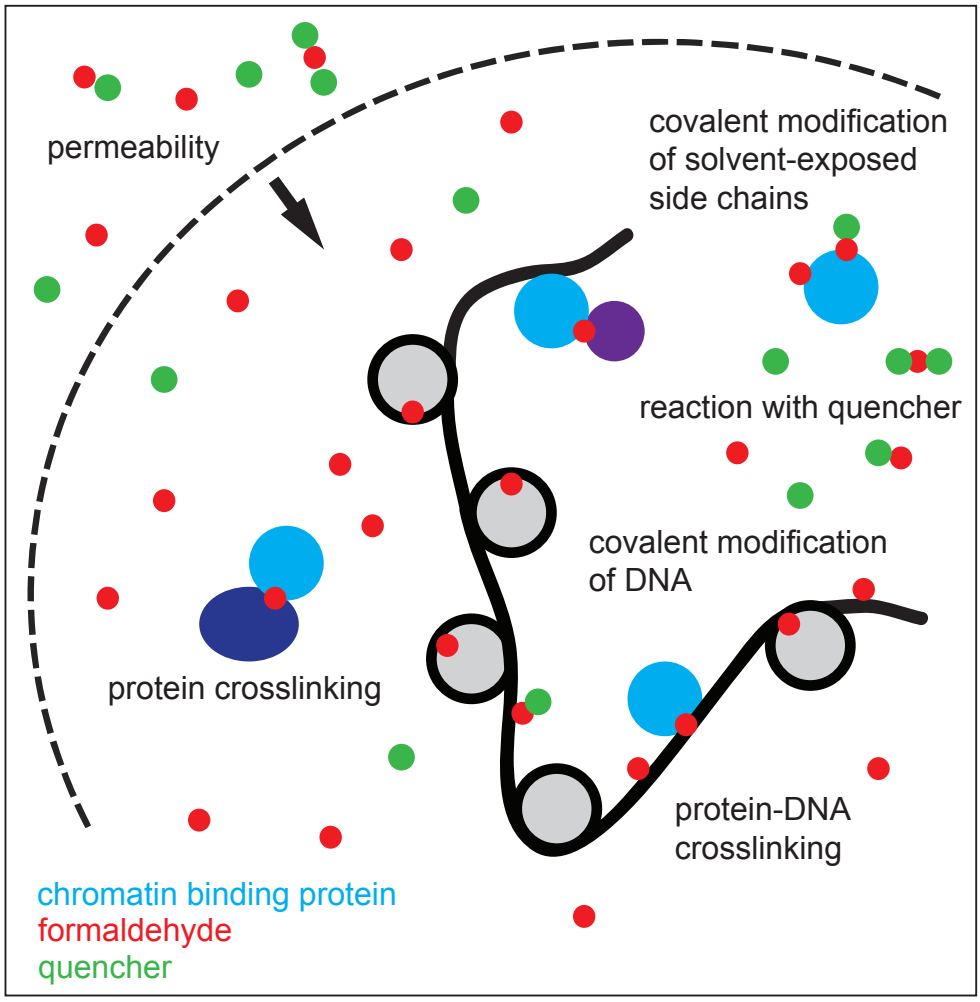


Figure 2.1. This cartoon depicts the main aspects of formaldehyde reactivity in cells. The dashed arc represents cell or nuclear membranes, which are thought to be highly permeable to formaldehyde (red circles). The thick black curved line represents DNA, shown assembled as nucleosomes (light gray circles). A chromatin interacting factor is schematized in cyan, with other partner proteins shown in blue and purple. Small molecules such as glycine and Tris that react with formaldehyde and can therefore quench reactivity with cellular constituents are shown as green circles. Formaldehyde can crosslink macromolecules together as well as modify exposed groups on macromolecules, forming a product species potentially stabilized by reactivity with a quencher. Quenchers are ordinarily added to the extracellular milieu and may exert their main effects outside the cell.

Basic chemistry of formaldehyde

Formaldehyde is the smallest aldehyde, an electrophilic molecule susceptible to chemical attack by a wide range of nucleophilic species of biological interest. The chemical complexity of formaldehyde-mediated reaction products was investigated more than 70 years ago (Fraenkel-Conrat et al. 1945). Initially using amino acids, and subsequently proteins and other substrates, it was shown that formaldehyde reacts *in vitro* with a wide range of functional groups, forming a complex array of products (Fraenkel-Conrat and Olcott 1948). It has been known since the 1940s that such products can include intramolecular and intermolecular crosslinked species, and that the reaction conditions (e.g. pH, temperature) can strongly influence the nature, yield, and half-life of chemical modifications (French and Edsall 1945). The concentration of formaldehyde used, incubation times, and other conditions can vary substantially among different applications employing formaldehyde fixation, yielding very different chemical products (reviewed in (Sutherland et al. 2008)).

Formaldehyde reacts with macromolecules in several steps (Figure 2.2). In the first step, a nucleophilic group on an amino acid or DNA base (for example) forms a covalent bond with formaldehyde, resulting in a methylol adduct which is then converted to a Schiff base. Methylols and Schiff bases can decompose rapidly (Metz et al. 2004) or may be stabilized in a second chemical step involving another functional group, often on another molecule, leading to formation of a methylene bridge (Feldman 1973). A methylene bridge might form between a solvent-exposed group on a macromolecule and a small molecule in solution such as glycine, which is frequently used as a formaldehyde quencher. Alternatively, and of most interest to biologists, is the formation of a covalent bond linking functional groups in two different macromolecules. The small size of formaldehyde dictates its linkage of groups that are ~2 angstroms apart, making it well suited for capture of interactions between macromolecules that are in close proximity (Solomon and Varshavsky 1985, Quievryn and Zhitkovich 2000).

(Commercial preparations of formaldehyde may also contain formaldehyde aggregates (Toews et al. (2008) whose reactivities and distance spanning capabilities are unclear.)

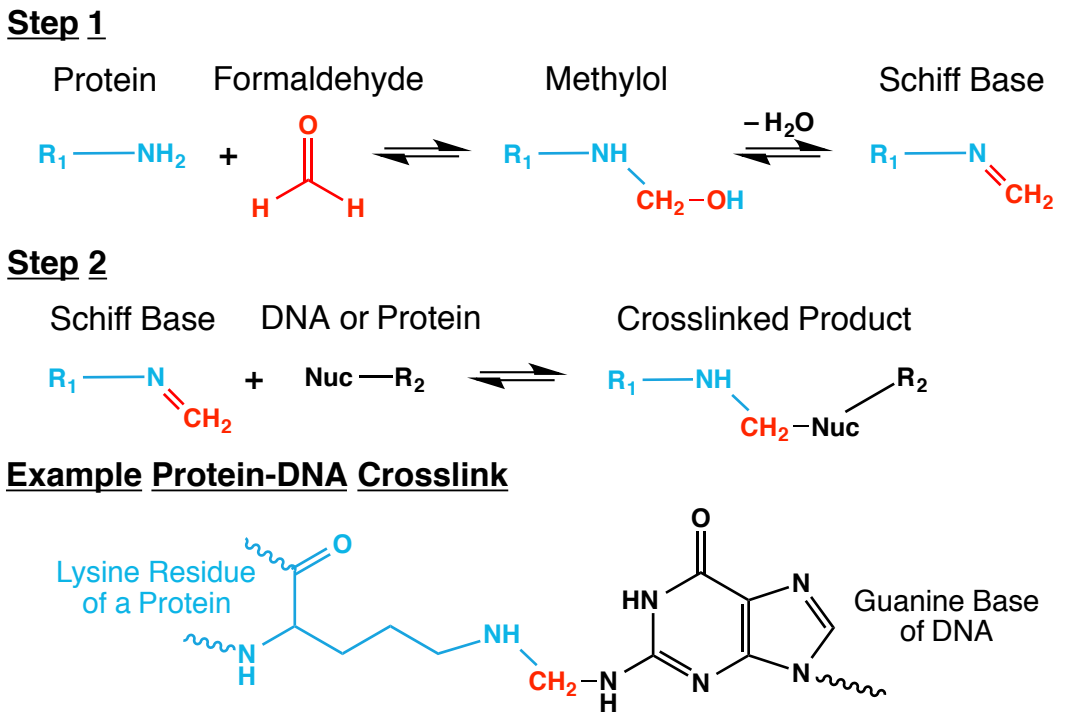


Figure 2.2. Formaldehyde crosslinking of biomolecules occurs in two steps. First, formaldehyde reacts with a relatively strong nucleophile, most commonly a lysine ε-amino group from a protein. This reaction forms a methylol intermediate that can lose water to yield a Schiff base (an imine). Second, the Schiff base reacts with another nucleophile, possibly an amino group of a DNA base, to generate a crosslinked product. This second nucleophile might be from DNA, another protein, the same protein as the first nucleophile, a quencher molecule, or other endogenous small molecule, and therefore a protein-DNA crosslink is only one of many possible products. All of the reactions in this two-step process are reversible, which is a key feature of formaldehyde crosslinking for chromatin capture. A specific example of a protein-DNA crosslink is shown. The atoms are color coded to match those of Figure 2.1: protein (cyan), formaldehyde (red), DNA (black).

Formaldehyde reactivity with proteins

As studies of formaldehyde reactivity became more sophisticated, it was found that conditions that more closely resemble those used for crosslinking components in cells yield a subset of the products identified in the earlier studies (Toews et al. 2008). Using model peptides, formaldehyde was found to react with N-terminal amino groups and side chains of cysteine, histidine, lysine, tryptophan and arginine (Metz, Kersten et al. 2004). Reaction products were in some cases influenced by the peptide sequence, yielding intramolecular crosslinks as well as linkages of the N-terminus and histidine, asparagine, glutamine, tryptophan, tyrosine and arginine residues to glycine molecules added to the reaction (Metz, Kersten et al. 2004). Despite the long incubation time (48 hours), adducts were not detected between glycine and peptide cysteine or lysine residues. Subsequent work employing model substrates along with formaldehyde concentrations and reaction times more in line with those used with cells identified a smaller subset of formaldehyde reaction products involving lysine, tryptophan and cysteine side chains as well as the peptide N-terminus (Toews et al. 2008). Such studies have often been motivated by interest in developing techniques for analysis of native protein complex subunit composition. As discussed more below, the rapid reactivity of formaldehyde with cellular components suggests that cells are highly permeable to formaldehyde, and the requirement for crosslinked groups to be closely apposed makes formaldehyde a good candidate for capturing macromolecular complexes *in vivo* containing specific but unstably bound subunits, which can then be analyzed by mass spectrometry (Klockenbusch et al. 2012).

In discussing the complexity of crosslinked complexes formed by incubation of cells with formaldehyde, it is important to distinguish between two types of complexity. The first is the chemical complexity arising from the variety of macromolecular functional groups that can potentially react with formaldehyde, and the second is complexity associated with the types and numbers of macromolecules crosslinked to each other. Although formaldehyde can potentially generate a great variety of chemically distinct products *in vitro*, the biologically relevant chemical

complexity is in all likelihood simpler under incubation conditions more typically used for analyses of macromolecular complexes *in vivo*. This is due to several factors, including a lowered effective formaldehyde concentration in cells compared to most model experiments *in vitro*, limiting the ability of formaldehyde to locate and interact with a functional group. While there is much greater macromolecular diversity in cells than in typical *in vitro* experiments, native macromolecules likely provide a smaller range of chemically reactive groups than model substrates used *in vitro*. As discussed below, N-terminal amino groups may be less available and side chains are less accessible to formaldehyde crosslinking due to protein tertiary structure in native proteins. These factors would decrease the proportion of potentially chemically reactive groups and allow for a smaller, less diverse set of chemical products *in vivo.* For instance, reactivity with native proteins is limited to those nucleophilic groups that are accessible to formaldehyde, and indeed, studies exploring differential formaldehyde reactivity have been used to provide insight into enzyme structure and catalytic function (Means and Feeney 1995). Solvent accessible lysine residues have been found to provide the most reactive functional groups in native proteins, and moreover, modification of native proteins by formaldehyde does not appear to alter tertiary structure very much (Toews et al. 2010). This is consistent with early work in the chromatin field that established that lysine residues are the predominant sites of formation of methylene bridges in histone complexes; such studies led to the suggestion as well that formaldehyde crosslinking does not in general perturb protein structure (Jackson 1978). The apparent preference of formaldehyde for accessible lysine residues may explain in part why formaldehyde has emerged as the crosslinker of choice for trapping protein-DNA complexes, as lysine residues are common mediators of interactions with DNA (Rohs et al. 2010). The differential reactivity of accessible groups on protein surfaces has also been explored to understand how formaldehyde fixation impacts epitope recognition by antibodies (Vani et al. 2006). Of note, the potential for formaldehyde to affect antibody recognition could possibly impact a wide range of experiments that require quantification of recovered fixed material by

immunoprecipitation. Conditions can often be worked out such that formaldehyde treatment does not adversely impact antibody recognition (Vasilescu et al. 2004, Klockenbusch and Kast 2010), but to our knowledge this has not been examined in great detail in the chromatin field. Importantly, the apparent predominance of a subset of reactive sites on macromolecules under typical experimental conditions is not to suggest that overall crosslinking complexity in cells is necessarily simple. Even though *in vivo* crosslinking is probably predominated by a subset of the chemical products observed *in vitro*, there is potential for macromolecules to become crosslinked together in multiple ways and in multiple combinations, forming larger daisy-chained structures that complicate *in vivo* crosslinking results. Indeed, there is some evidence that formaldehyde treatment of cells can result in higher order chromatin or nuclear structures whose formation may yield misleading interpretations of chromatin association data by trapping factors within dense crosslinked networks (Schmiedeberg et al. 2009, Gavrilov et al. 2015).

Formaldehyde reactivity with DNA

Formaldehyde reacts with amino and imino groups of DNA bases, and extensive studies have been performed to document kinetic and thermodynamic aspects of such reactions (Fraenkel-Conrat 1954, Haselkorn and Doty 1961, McGhee and von Hippel 1975, McGhee and von Hippel 1975). While formaldehyde reactivity with proteins does not appear to perturb protein tertiary structure, formaldehyde reactivity with DNA is notably different as covalent modification of DNA bases requires disruption of base pairing in duplex DNA, and in fact, formaldehyde was used in pioneering studies to probe DNA melting (Utiyama and Doty 1971, von Hippel and Wong 1971, McGhee and von Hippel 1976, Shikama and Miura 1976, McGhee and von Hippel 1977). Modified bases are thus precluded from base pairing and promote further DNA denaturation (Utiyama and Doty 1971). This likely occurs to some extent in stretches of naked DNA in cells treated with formaldehyde, although under typical conditions employed for *in vivo* studies, the recovered DNA is by and large suitable for enzymatic manipulation (Gavrilov

and Razin 2009). Formaldehyde modification of naked DNA *in vitro* may be more extensive (Wu et al. 2011). Conformational changes in DNA that promote formaldehyde reactivity have been referred to as DNA "breathing" or base flipping. Measurement of the rates of such spontaneous conformational changes is an active area of investigation (Yin et al. 2014) and it is unclear what specific DNA conformational changes are required to allow reaction of DNA bases with formaldehyde (i.e., full extrahelical extrusion of a DNA base may not be required). The rates of formaldehyde reactivity with naked DNA *in vitro* were found to be orders of magnitude below (faster than) diffusion-limited rates, although these studies make clear that reaction conditions can have large effects on reactivity and indeed, it was recognized early on that it would be difficult to extrapolate rates of reaction obtained in relatively simple *in vitro* systems to other more complex systems, let alone *in vivo* (McGhee and von Hippel 1975).

Capture of protein-DNA complexes with formaldehyde crosslinking

The early use of formaldehyde as a probe of macromolecular structure led to the discovery that formaldehyde can crosslink histones to DNA (Brutlag et al. 1969). Retrieval of the crosslinked complexes and analysis of the associated DNA then gave birth to the ChIP assay (Gilmour and Lis 1985, Solomon and Varshavsky 1985, Solomon et al. 1988), which has become ubiquitous in the chromatin field in an ever-expanding number of variations (Collas 2010). Although ChIP assays performed without crosslinking have proven valuable for analyses of stable chromatin complexes (Kasinathan et al. 2014), crosslinking has made it possible to identify interactions that would not otherwise withstand the isolation procedure. Given the utility of crosslinking and its critical role in establishing many of the principles underlying the current understanding of chromatin structure and function, a clear picture of formaldehyde chemistry is critical to ensure that any biases resulting from formaldehyde crosslinking are taken into account.

The ability of formaldehyde to crosslink amino acids to DNA bases has been examined systematically *in vitro*. In comparing the products of reactions containing lysine, cysteine,

histidine, or tryptophan with each of the four DNA bases, the highest yield of crosslinked product was obtained with lysine and deoxyguanosine (Lu et al. 2010), consistent with lysine being most reactive among residues in native proteins described above. Similar results were obtained using short peptides and trinucleotides (Lu et al. 2010). In the context of protein-DNA interactions, the first chemical step could involve reaction with either an amino acid side chain in a protein, the protein N-terminus, or an amino or imino group on a DNA base; importantly, however, the ε-amino group on the lysine side chain is a better nucleophile than are the amino/imino groups on DNA bases whose lone pair electrons are delocalized in the aromatic ring due to less steric hindrance. For this reason it seems reasonable to speculate that in most crosslinked protein-DNA complexes, a Schiff base is formed on a lysine residue first, followed by nucleophilic attack by the DNA base held in proximity to the side chain, resulting in a methylene bridge.

Interestingly, and in line with this idea, formaldehyde reactivity with DNA was stimulated substantially by adding amino acids or histones to an *in vitro* reaction, resulting in stable products that in some cases contained both DNA and the protein or amino acid (Siomin et al. 1973). The ~20-30-fold stimulation in the reaction rates observed in these early experiments by addition of glycine or lysine (for example) was striking; furthermore, formaldehyde crosslinking of proximal functional groups on specific, stable macromolecular complexes presumably can occur even faster owing to the constrained physical proximity of the reacting species (Jencks 1997). In addition to the ubiquity of lysine side chains in DNA-binding proteins (for interaction with the phosphate backbone), the DNA bases provide a high density of amino and imino groups along the length of the nucleic acid. These two features may contribute to the relatively higher yield of protein-DNA crosslinks compared to protein-protein crosslinks as measured by conjugation of chromatin regulatory complexes that interact indirectly with DNA (Zeng et al. 2006). It has been observed that for some transcriptional co-regulators, protein-protein crosslinks are not efficiently detected between factors that interact with chromatin indirectly when using 3C methods; this could be due to either inefficiencies in formaldehyde

crosslinking between proteins (due to non-optimal reactive side chain availability) or rarity of these protein-protein interactions (Gavrilov et al. 2015).

Several different crosslinking agents have been used in ChIP (Nowak et al. 2005) but all of the features described above including cell permeability, short spacer length, rapid reactivity- as well as reversibility (discussed below)- have led to formaldehyde becoming the crosslinking agent of choice for ChIP. This utility has been borne out by many genome-wide studies that have shown how profiles of crosslinked complexes capture transcription factor binding to physiologically significant DNA sites (for example, see (Ren et al. 2000, Harbison et al. 2004, Venters et al. 2011). Since DNA site-specific transcription factors can also bind to nonspecific sites (Elf et al. 2007, Hager et al. 2009, Mirny et al. 2009, Hammar et al. 2012), crosslinking of non-specifically bound proteins to DNA would be expected to occur and may account in part for binding events detected in genome-wide studies that cannot be readily explained physiologically. Non-DNA binding proteins are not crosslinked to chromatin (Solomon and Varshavsky 1985, Hall and Struhl 2002), and non-specifically bound factors are presumably bound to a multitude of disparate sites at low levels consistent with their relative occupancies (Buck and Lieb 2004, Struhl 2007). Analyses of chromatin binding by a series of mutants in the methyl CpG binding protein 2 gene led to the conclusion that there is a threshold interaction lifetime of about 5 seconds required for crosslinking (Schmiedeberg et al. 2009). However, it has been possible to ChIP transcription factors whose interactions with chromatin are known from imaging studies to be highly transient (~few seconds time scale) (Mazza et al. 2013, Chen et al. 2014, Stasevich et al. 2014, Zhang et al. 2016). Importantly, *in vivo* ChIP signals have been found to correlate with DNA binding specificities and affinities measured *in vitro* (Toth and Biggin 2000, Kaplan et al. 2011), supporting the use of formaldehyde for measuring chromatin binding interactions in cells with quantitative rigor and over a broad thermodynamic range. A better understanding of formaldehyde's effects in cells could potentially be obtained by biochemical studies of protein-DNA complex crosslinking *in vitro*. However, while the data are not extensive, it is noteworthy

that there are examples in which *in vitro* and *in vivo* binding behaviors differ when assessed using formaldehyde in one system or another (Solomon and Varshavsky 1985, Toth and Biggin 2000). A study by Solomon and Varshavsky (1985) found that histones could be crosslinked to DNA both *in vitro* and *in vivo* but *lac* repressor and α-protein could not be crosslinked to DNA *in vitro*. Additionally, pronase-digested SV40 chromosomes treated with formaldehyde *in vivo* were suggested to have crosslinks within digested DNA fragments, similar to *in vitro* treatment, but also between these fragments. Toth and Biggin (2000) were able to crosslink the Eve protein to DNA *in vitro* but not *in vivo*; however, they were able to crosslink another *Drosophila* protein, Zeste, to DNA both *in vitro* and *in vivo*.

Kinetics and stability of formaldehyde-mediated crosslinks

Formaldehyde crosslinking in chromatin studies typically employ relatively low formaldehyde concentrations (1%/360 mM or less). The relatively easy detection of protein-DNA complexes following incubation times of 30 min or less suggests that macromolecular crosslinking occurs relatively rapidly, as suggested by the earliest ChIP experiments (Solomon and Varshavsky 1985, Solomon, Larsen et al. 1988, Dedon et al. 1991). Relatively rapid formaldehyde reactivity in cells is also consistent with the ability to distinguish ChIP signals over short time intervals (seconds to minutes) (Hall and Struhl 2002, Katan-Khaykovich and Struhl 2002, Poorey et al. 2013). Formaldehyde crosslinks are quite stable *in vivo* compared to the durations of most crosslinking experiments, with crosslink half-lives of ~10-20 hours depending on the cell type and conditions (Quievryn and Zhitkovich 2000). In ChIP experiments, crosslinks are most often reversed by heat (Jackson 1978). The reversibility of formaldehyde crosslinking has been explored in some detail in an effort to recover proteins from fixed tissue and cell samples (Shi et al. 2013). The temperature and salt concentration dependence of the formaldehyde crosslink reversal rate has been established, revealing a crosslink half-life consistent with the estimate of crosslink half-life in cells (tens of hours at 37 °C) (Kennedy-

Darling and Smith 2014). That study also quantitatively showed the extent to which heat can increase the crosslink reversal rate. More such measurements on other aspects of crosslinking chemistry will be useful in developing a quantitative understanding of how the abundance of a particular crosslinked species obtained under some set of conditions (i.e. the effect of pH, quencher choice and concentration, and formaldehyde concentration) relates to dynamic aspects of complex assembly/disassembly and stability (Poorey et al. 2013).

Quenching and reversal of formaldehyde-mediated crosslinks

To limit formaldehyde reactivity to a particular time interval, unreacted formaldehyde is quenched with an excess of a small reactive molecule added to the reaction **(**Figure 2.3**)**. Quenching is important but not well understood (Sutherland et al. 2008). Glycine has been typically used as a sink for unreacted formaldehyde in ChIP (Kuo and Allis 1999, Nowak et al. 2005) as well as in approaches to map higher order chromatin structure (de Wit and de Laat 2012). The efficacy of glycine is improved by reduced pH, but detailed studies of the quenching reaction have not been reported (Sutherland et al. 2008). In principle, formaldehyde crosslinking could be quenched by reaction of the quencher with formaldehyde molecules in solution or reaction with formaldehyde conjugates on other molecules in the cell, if the quencher is readily cell permeable. As discussed above, formaldehyde-mediated glycine conjugates have been detected or inferred *in vitro*, although there was no evidence for such conjugates seen in proteins analyzed from formaldehyde-treated cells (Vasilescu et al. 2004, Sutherland et al. 2008, Poorey et al. 2013). On the other hand, evidence suggests that glycine-DNA conjugates are formed in an *in vitro* reaction (Wu et al. 2011). Despite the fact that glycine has been used routinely to quench crosslinking, Tris[3] is a more efficient quencher (Sutherland et al. 2008), which can be explained chemically by the ability of Tris to form a cyclic product upon reaction with formaldehyde (Wu et al. 2011) **(Figure 2. 3)**. However, at higher concentrations of Tris, which would likely be used

for quenching, Tris can also facilitate crosslink reversal (Kawashima et al. 2014) thereby potentially impacting the yield of crosslinked material.

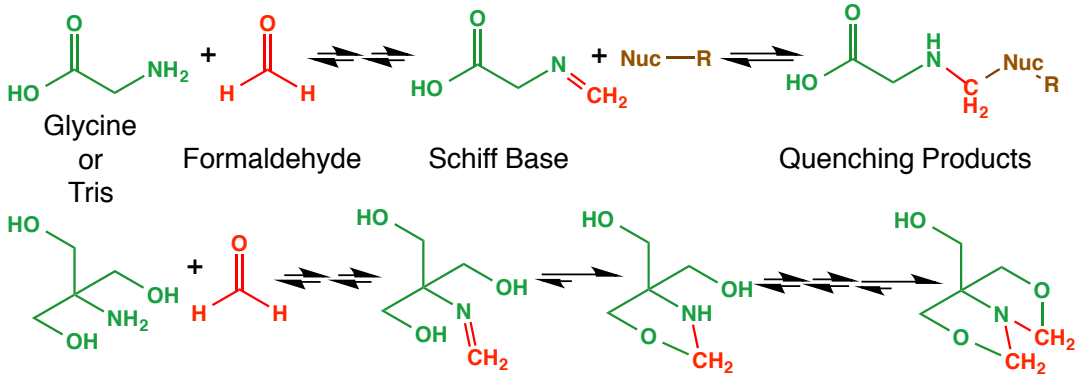


Figure 2.3. **Formaldehyde quenching reactions with glycine and Tris, the two most common quenchers.** The chemical reactions are analogous to those shown in Figure 2.2 with the amino group of glycine or Tris acting as the primary nucleophile. The Schiff base formed from glycine may or may not react with a second nucleophile, but regardless, the crosslinking between macromolecules has been quenched. The Tris molecule has readily available second nucleophiles (hydroxyl groups) that create stable intramolecular five-membered rings. It is also possible for Tris to react with two formaldehyde molecules, leading to the final product shown. The propensity for Tris to form these stable intramolecular products likely allows it to scavenge formaldehyde from other molecules and thereby facilitate crosslink reversal. The atoms are color coded: quencher (green), formaldehyde (red), miscellaneous nucleophile (brown).

The crosslinking kinetics (CLK) assay

Gene regulation is a complicated and highly regulated process involving the coordinated assembly of dozens of proteins on promoter DNA within the context of chromatin (Kim et al. 2005, Consortium 2012, Rhee and Pugh 2012, Dowen et al. 2014). As discussed in Chapter 1, *in vitro* studies have provided a structurally detailed paradigm for how the transcription preinitiation complex (PIC) is assembled and regulated (Horn et al. 2016, Coulon et al. 2013, Cramer 2014,

Hager, McNally et al. 2009, He, Fang et al. 2013, Luse 2013, Zawel and Reinberg 1992, Conaway and Conaway 1993, Roeder 1996), but less is known about the dynamic assembly of PICs *in vivo* or how transcription factors (TFs) contribute kinetically to PIC assembly or to the rate of the initiation of synthesis of individual RNAs. To develop molecular models for how these processes occur *in vivo*, estimates of on- and off-rates for TF binding to specific loci *in vivo* are required. In instances in which kinetic measurements cannot be made, biophysically rigorous estimates of site-specific *in vivo* affinity (as opposed to estimates of relative affinity) and fractional occupancy would be valuable.

Chromatin immunoprecipitation is quite possibly the most widely used assay for characterizing the interactions between TFs and specific sites on chromatin and typically uses formaldehyde to crosslink TFs to their chromatin sites (Solomon and Varshavsky 1985). While it is an undeniably powerful approach for determining transcription factor binding locations with high precision (Rhee and Pugh 2012), standard ChIP assays are static, non-kinetic measurements that do not provide unambiguous insight into the *in vivo* kinetics of these dynamic interactions. Several assays have expanded ChIP to attempt to capture these relationships. We previously developed a ChIP-based method, the crosslinking kinetics (CLK) assay, which exploits the time dependence of formaldehyde crosslinking to model chromatin-TF binding dynamics on a broad time scale and at individual loci (Poorey et al. 2013). The basic principle of the method is based on the following equation:

$$TF + chromatin \underset{k_d}{\overset{k_a}{\rightleftharpoons}} complex \xrightarrow[formaldehyde]{k_{xl}} crosslinked$$

A factor can bind and dissociate from chromatin to form a complex at the rate of k_a and k_d, respectively. When formaldehyde is added to the system, the TF-chromatin complex becomes crosslinked at a rate of k_{xl}. In this approach, cells are incubated with formaldehyde for various periods of time, unreacted formaldehyde is then quenched, and the extent of DNA site crosslinking of a TF of interest at each time point is quantified by ChIP. The time-dependent

increase in ChIP signal results from a combination of time-dependent formaldehyde reactivity and time-dependent binding of free TF molecules to unoccupied DNA sites in the cell population and eventually reaches saturation, assuming an excess of TF molecules, where all available chromatin binding sites have theoretically been occupied (Figure 2.4). To distinguish kinetic effects of crosslinking chemistry from kinetic effects of TF binding, measurements are made using congenic cells differing only in the concentration of TF and the data are fit using both sets of data simultaneously (Poorey et al. 2013, Viswanathan et al. 2014). A number of assumptions were made for the development of this model: binding occurs before crosslinking, crosslinking is irreversible, the unbound pool of the factor of interest is in excess and not depleted by formaldehyde, and formaldehyde is not limiting. It's important to note that no assumptions are made about the rate of crosslinking or TF binding dynamics.

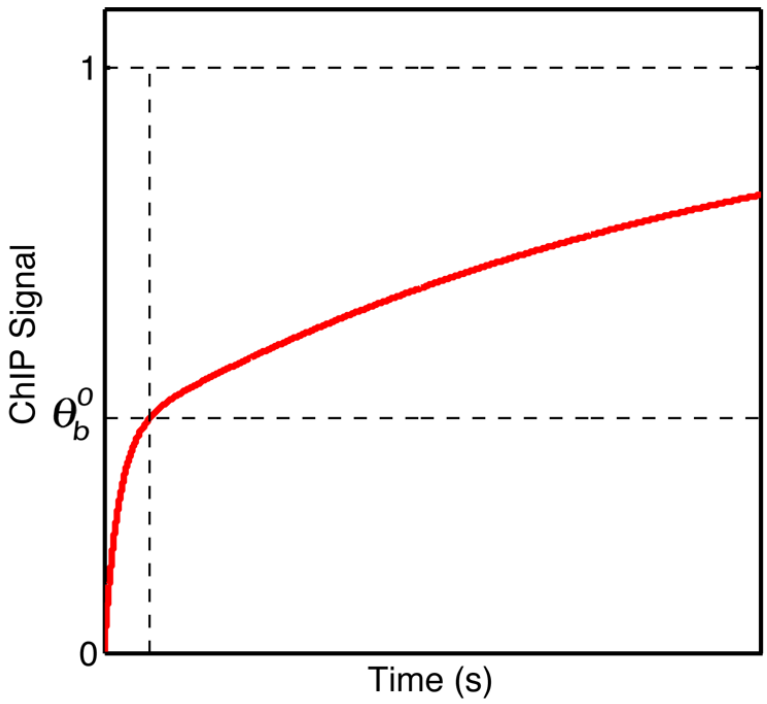


Figure 2.4. Simulation of the dependence of ChIP signal on formaldehyde incubation time with the CLK model. The curve is comprised of two parts: the initial steep rise of the ChIP signal is dependent on formaldehyde crosslinking time and the second shallower rise indicates the

increase in ChIP signal due to TF dynamics as factors find available binding sites on chromatin and are crosslinked. The bend where the first part of the curve transitions to the second part indicates the fractional occupancy of the TF at the specific locus without formaldehyde at steady-state conditions.

The CLK method modeling was derived from the concept that a TF can only be in three states over the course of the assay: bound to DNA but not crosslinked, unbound, or crosslinked to DNA. This is represented by the following formula, derived by Stefan Bekiranov and Kunal Poorey (Poorey et al. 2013), where $\theta_b(t)$ is the fraction of bound sites, $\theta_u(t)$ is the fraction of unbound sites, and $\theta_{xl}(t)$ is the fraction of crosslinked sites:

$$\theta_b(t) + \theta_u(t) + \theta_{xl}(t) = 1 \quad (1)$$

The rate of change in the fraction of sites bound by the TF is represented by the following, where r_{1f} is the association rate of the TF, r_{1b} is the dissociation rate, and r_2 is the crosslinking rate:

$$\frac{d\theta_b(t)}{dt} = r_{1f} - r_{1b} - r_2 \quad (2)$$

and the rate of change of TF crosslinked to binding sites is:

$$\frac{d\theta_{xl}(t)}{dt} = r_2 \quad (3)$$

If first-order kinetics are assumed, $r_{1f} = k_a C_{TF} \theta_u$, where k_a is the on-rate, C_{TF} is the concentration

of the factor in the nucleus, and θ_u is the fraction of unbound sites. The overall dissociation rate is $r_{1b} = k_d\theta_b$, where k_d is the off-rate and θ_b is the fraction of sites bound by the TF. If the overall rates are substituted into reaction (2), they yield the following:

$$\frac{d\theta_b(t)}{dt} = k_a C_{TF}\theta_u(t) - k_d\theta_b(t) - k_{xl}C_{FH}\theta_b(t) \quad (4)$$

$$\frac{d\theta_{xl}(t)}{dt} = k_{xl}C_{FH}\theta_b(t) \quad (5)$$

We assume that there are no crosslinked sites before formaldehyde addition (defined as t_0). Therefore, before crosslinking at time $t = 0$:

$$\theta_{xl}(0) = 0 \quad (6)$$

At time $t = 0$, there is no crosslinking and therefore $r_2 = 0$; equation (2) describes a TF binding to its *in vivo* DNA site in this case. When we set $r_2 = 0$ and $d\,\theta_b\,/dt = 0$ (steady-state before crosslinker is added), the equilibrium fraction of bound sites, θ^0_b, at time $t = 0$ is:

$$\theta_b^0 = \frac{k_a C_{TF}}{k_a C_{TF} + k_d} \quad (7)$$

Using equations (1) – (3) and (6) and (7), the fraction of sites with a TF crosslinked, $\theta_{xl}(t)$, can be determined as a function of time:

$$\theta_{xl}(t) = 1 - \frac{\tau_+ e^{-t/\tau_+} - \tau_- e^{-t/\tau_-}}{\tau_+ - \tau_-} + \frac{\theta_b^0 \tau_+ \tau_- k_{xl} C_{FH}}{\tau_+ - \tau_-}\left(e^{-t/\tau_+} - e^{-t/\tau_-}\right) \quad (8)$$

In this final equation, τ_+ and τ_- are two time constants/relaxation times and represent a short and

long crosslinking time required to reach steady state. Equation (8) is complicated and involves a number of parameters. Simpler approximates were developed for four cases to fit the data to non-linear models, which require good estimates of initial parameters. These cases represent different possibilities of crosslinking and TF-binding dynamic regimes and are illustrated in Figure 2.5. Initial estimates are fit to the approximate form of Eq. (8). For Cases 1 and 2, output parameters for k_{xl}, θ_b, and k_a are calculated; k_d and $t_{1/2}$ are extrapolated using $\theta_b = k_a C_{TF}/(k_a C_{TF}+k_d)$ and $t_{1/2} = \ln(2)/k_d$, respectively. For Cases 3 and 4, the dissociation constant ($K_d = k_d/k_a$) and θ_b are determined, but not the individual kinetic rates, since the steady-state saturation level cannot be accurately determined. The specific approximations and assumptions for each case used to model fit data are discussed further in the *Computation Modeling* section of Materials & Methods.

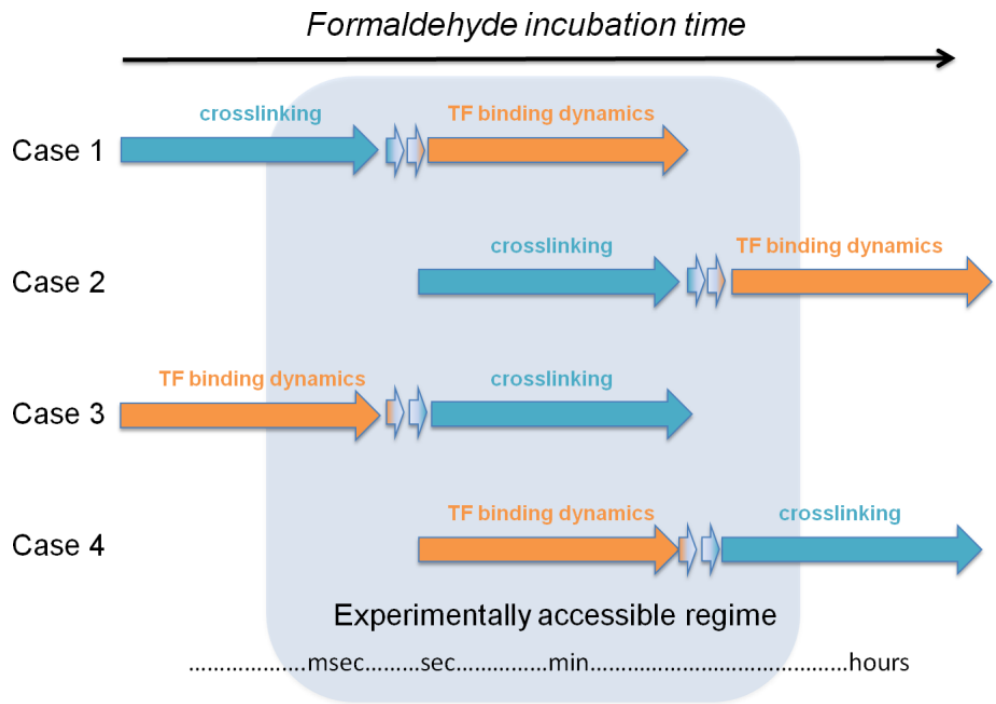


Figure 2.5. Schematic illustrating four possible cases in which crosslinking kinetics (blue arrows) and TF binding dynamics (orange arrows) contribute to the increase in ChIP signal with increasing formaldehyde incubation time (black arrow at top of figure). The observation that the ChIP signal increases quickly with short formaldehyde incubation times and

then more slowly with longer incubation times indicates that the two processes of crosslinking and TF-binding can be separated. However, reactions that occur too quickly (ms) and too slowly (>40 min) are outside the accessible regime with this method (indicated by light blue shading). Depending on the specific locus and factor examined, the crosslinking rate and TF dynamics can vary, yielding four possible cases for model fitting. Cases 1 and 2 result when formaldehyde crosslinking dynamics are faster than those of the TF-chromatin binding and are rate-limited by the on-rate of the TF. Crosslinking occurs rapidly (ms/sec time scale) in Case 1 followed by TF-binding dynamics on the order of seconds to ~30 min, while Case 2 crosslinking occurs at a slower rate and binding dynamics are even slower. Cases 3 and 4 result from more rapid TF binding dynamics than crosslinking rates and are referred to as "crosslinking-limited" later in the paper and the resulting fit can either be a single exponential or linear, depending on the time scale of the two components. In Case 3, TF-binding dynamics happen in a matter of seconds and most of the increase in ChIP signal depends on the crosslinking rate. The ChIP signal resulting from Case 4 is completely dependent on and limited by the crosslinking rate and results in a linear rise. So far, we have observed data that represents Cases 1, 3 and 4.

A challenge with the development of locus-specific kinetic assays such as CLK is that aspects of the effects of formaldehyde on cells largely remain a black box (Gavrilov et al. 2015), and validation of the extracted dynamic parameters is difficult because complementary approaches are still being developed and there are few "gold standard" interactions with convergent kinetic measurements obtained by different approaches. Support for the CLK approach was obtained by measurement of binding dynamics for two TFs with very different dynamic properties that had been assessed by live cell imaging (Robinett 1996, Karpova et al. 2008, Poorey et al. 2013). However, live cell imaging has its own technical challenges (Mueller,

Stasevich et al. 2013) and in most cases it is not possible to identify particular single copy chromatin sites of interaction by live cell imaging (Hager et al. 2009, Larson et al. 2009, Morisaki et al. 2014). An alternative approach is competition ChIP, an assay that measures the rate of turnover between an endogenous and inducible copy of a TF, which will be discussed in the next chapter. Recent work demonstrates that quantitative estimates of locus-specific binding kinetics can be obtained by modeling competition ChIP data, including the estimation of residence times much shorter than the time for full induction of the competitor TF (Zaidi et al. 2017). Importantly, comparison of CLK and competition ChIP data for TATA-binding protein (TBP) to a few specific loci shows that the time scales for chromatin interaction are similar as judged by the two methods, with residence times for promoter binding being in general on the order of several minutes (Zaidi et al. 2017).

Nonetheless, locus-specific TF-chromatin dynamics are just beginning to be explored, with only a small number of TFs and chromatin sites for which CLK, competition ChIP, and/or live cell imaging kinetic data are available. A key aspect of the CLK assay involves the trapping of bound species using formaldehyde. This chapter reports biochemical results that better define the chemical behavior of formaldehyde in yeast cells. An increased formaldehyde concentration led to more rapid crosslinking, which improved the time resolution and analytical ability of the assay to extract locus-specific binding kinetic information for some TFs. For other TFs, an increased formaldehyde concentration resulted in their depletion from the soluble pool, and in some cases rapid depletion. These observations emphasize the importance of optimizing the CLK approach for analysis of the dynamic behavior of a particular TF. This improved CLK method utilizes more rapid crosslinking and more efficient quenching in yeast cells, as well as improved computational methods for data analysis and improved approaches for distinguishing contributions of crosslinking rate and binding kinetics to the time-dependent increases in ChIP signal. These new conditions have resulted in a more robust method and the ability to model and analyze crosslinking kinetic data with more reliability and confidence.

Materials and Methods

The updated version of crosslinking kinetic (CLKv2) analysis is a modified ChIP procedure that yields kinetic measurements for transcription factor binding to specific loci by fitting ChIP data obtained from cells treated with formaldehyde for different periods of time. Before the samples can be collected for kinetic analysis, control experiments are completed to optimize experimental conditions; the schematic in Figure 1 outlines the general workflow for this process.

Yeast strain construction and growth conditions

For the CLKv2 assay, two strains are used for the analysis of a TF of interest: a wild type (WT) strain and an overexpression (OE) strain. The overexpression strain is isogenic to the WT strain other than driving levels of the TF that are modestly higher (~3-5-fold) over the WT levels. Kinetic analysis of the WT and OE strains in parallel (described below) highly constrains fits of the data by revealing the mass action contribution to the increase in ChIP signal over time. The OE strain can be engineered by introducing into cells an additional copy of the TF gene on a plasmid or by integrating into the genome under control of the native promoter or an appropriate heterologous promoter. If the TF functions as a stable biochemical entity with more than one type of subunit, the OE strain needs to be engineered to drive balanced expression of each subunit, as for example, was done for the analysis of TFIIE (Zaidi et al. 2017). For a detailed description of OE strain and plasmid construction, see Poorey *et al*., 2013, and Zaidi *et al*., 2017.

To generate strains for kinetic analysis using an OE plasmid, transformation of WT *S. cerevisiae* cells in the strain background of interest with either the OE plasmid construct or an empty vector carrying the same selection marker was carried out. Transformants were selected on appropriate agar plates and restreaked for single colonies. The WT and OE strains were

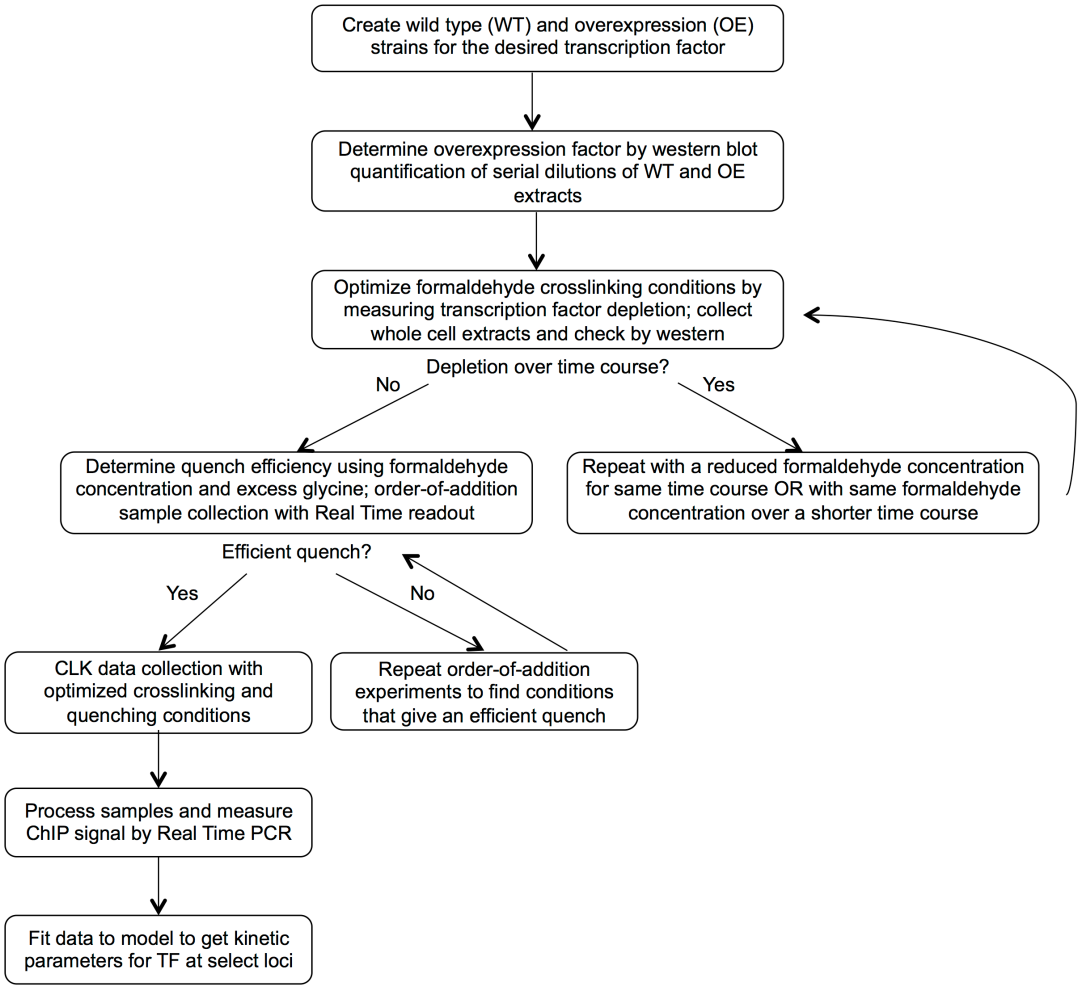


Figure 2.6. CLKv2 workflow. Control experiments are performed to optimize the assay conditions before collection of CLK data and fitting to the CLK model.

archived by storage at -80 °C using standard yeast glycerol stock methods (Amberg et al. 2005). Many of the *S. cerevisiae* strains used were described previously (Poorey et al. 2013). Other strains were newly developed for the work presented here and all are listed in Table 2.1; plasmids used for strain construction are listed in Table 2.2.

TBP ChIP was performed in two ways: (1) using a monoclonal antibody that recognizes untagged TBP, and (2) using an antibody that recognizes the epitope tag on TBP-myc. Chromatin-associated myc-tagged TBP was measured using the epitope-tagged strain YAD154. TBP ChIP using the monoclonal TBP antibody was performed in various strains as described below. YAD154 cells used for the TBP-myc ChIP experiments were grown in YPD overnight at 30°C and harvested at OD_{600} ~ 1. For other TBP ChIP experiments comparing strains with two different levels of TBP, AY146 (wild type TBP levels) and YSC018 (harboring a 2μ TBP overexpression plasmid) were obtained from the TBP shuffling strain YAD165 as described previously (Poorey et al. 2013). Cells were grown in synthetic medium without leucine plus 2% glucose overnight at 30°C. Culture volumes for each type of experiment are noted below and range from 100-450 ml depending on the experiment. When an OD_{600} of ~0.8 was reached, cells were pelleted and resuspended in an equivalent volume of YEP plus 2% glucose medium. They were grown at 30°C for approximately one hour until an OD_{600} of 1.0 was reached and cells were then formaldehyde crosslinked as described below. This regimen allowed cells to be initially grown under plasmid selection, but then transferred to YPD in order to standardize ChIP results which could otherwise be potentially influenced by effects of growth medium, and in addition, growth in YPD prior to crosslinking permitted direct comparison with previously published work (Consortium 2012, Rhee and Pugh 2012).

Table 2.1. Yeast strains used in this chapter.

Strain	Genotype	Reference or source
YPH499	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1*	Sikorski and Hieter, 1989
YTK539	*MATa his3-Δ1 leu2Δ0 met15Δ0 ura3Δ0 ace1Δ :: KAN TRP1:: pCap2-ACE1-tripleGFP-HIS3*	Karpova et al, 2008
YSC002	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ACE1-tripleGFP-URA3 pMW101 [ACE1-triple GFP HIS3 2μ]*	Poorey et al, 2013
YRV005	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 GAL4-TAP*	Ghaemmaghami et al., 2003
AY146	*MATa * spt15::natMX pRS425 [LEU 2μ]*	Poorey et al, 2013
YSC018	*MATa * spt15::natMX pSH223 [TBP LEU 2μ]*	Poorey et al, 2013
YTK260	*MAT a/α, HIS5/his3Δ1, leu2Δ0/leu2Δ0, ura3Δ0/ura3Δ0, met15Δ0 LYS2::pHIS3-lacI-GFP-NLS-NAT1, CU3::KAN- (LacO)256 , CU1::(LacO)256*	Poorey et al, 2013
YSC001	*MAT a/α, HIS5/his3Δ1, leu2Δ0/leu2Δ0, ura3Δ0/ura3Δ0, met15Δ0 LYS2::pHIS3-lacI-GFP-NLS-NAT1, CU3::KAN- (LacO)256 , CU1::(LacO)256 pSC001 [pHIS3-GFP-LacI URA3 2μ]*	Poorey et al, 2013
YAD154	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 SPT15-myc*	Poorey et al, 2013
AY151	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 TFA1-TAP pRS315 [LEU2 CEN ARS]*	This study
AY152	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 TFA1-TAP pTAF1-TAP [pRS315-TFA1-TAP TFA2 LEU2 CEN ARS]*	This study
YRV006	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 TFA1-TAP*	Ghaemmaghami et al., 2003
YBR049C	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 REB1-TAP*	Ghaemmaghami et al. 2003
YRV018	*MATa ade2–1 his3–11,15 leu2–3,112 trp1–1 ura3–1 can1–100 abf1::HIS3MX6 pRS415-ABF1-FLAG [LEU2}*	Miyake et al, 2004
YGR186W	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 TFG1-TAP*	Ghaemmaghami et al. 2003
ML307-1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 CAT8-13xMyc::kanmx*	Jeff Smith lab
YNL167C	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 SKO1-TAP*	Ghaemmaghami et al. 2003

Table 2.2. Plasmids used for strain construction in this chapter.

Plasmid name	Information	Reference or source
pMW101	*ACE1-triple GFP HIS3 2μ*	Poorey et al., 2013
pRS425	*LEU 2μ*	Christianson T.W., et al., 1992
pSH223	*TBP LEU 2μ*	Steve Hahn
pSC001	*pHIS3-GFP-LacI URA3 2μ*	Poorey et al., 2013
pRS315	*LEU2 CEN ARS*	Sikorski and Hieter, 1989
pTFA1-TAP	*TFA1-TAP TFA2 LEU2 CEN ARS*	This study

For TFIIE ChIP, strains were used with WT or elevated levels of the two TFIIE subunits Tfa1 and Tfa2 in which the Tfa1 subunit was TAP tagged. YRV006 (*TFA1*-TAP, Dharmacon) (Poorey et al. 2013) was transformed with an empty pRS315 vector (Sikorski and Hieter 1989) (AY151, WT TFIIE levels) or a pRS315-based plasmid carrying copies of *TFA1*-TAP and *TFA2* under control of their endogenous promoters (AY152, over-expressed levels of TFIIE). Cells were grown at 30°C overnight in synthetic medium without leucine and with 2% glucose. When an OD_{600} of ~0.8 was reached, cells were pelleted, resuspended in an equivalent volume of YPD, and grown at 30°C to an OD_{600} of 1.0 as described above. Strain construction for AceI (YTK539 and YSC002) and LacI (YTK260 and YSC001) as well as collection for ChIP was described previously (Poorey et al. 2013).

For western blotting, strains YGR186W, YBR049C, YRV018, and ML307-1 were grown overnight in YPD at 30°C to OD_{600} of 1; YRV005 was grown in YEP + 2% raffinose at 30°C overnight to OD_{600} 0.8, then 2% galactose was added and cells incubated to OD_{600} ~1.0. YTK539 cells were grown under conditions of copper induction as previously described (Poorey et al. 2013).

CLK data collection

Quenching and crosslinking conditions. Different crosslinking and quenching conditions were tested with the TBP-myc strain (YAD154) in order to explore the relationship between crosslinking rate and formaldehyde concentration, as well as quenching efficiency. In all experiments, cells were first grown in a 5 ml YPD culture overnight at 30°C then diluted in a larger YPD culture and grown at 30°C overnight to an OD_{600} of 1.0. To test the effect of 250 mM glycine, 100 ml cell cultures were incubated with 2.7 ml 37% formaldehyde (1% final, Fisher) followed by addition of 10 ml 2.5 M glycine (pH 6.3) at various times. To test the effect of 2.93 M glycine, 450 ml cultures were grown in YPD overnight at 30°C to OD_{600} of 1. Cells were then concentrated five-fold by centrifugation and resuspended in 90 ml YPD. The concentrated

cultures were then incubated with 2.7 ml 37% formaldehyde (1% final concentration) by addition of formaldehyde to the culture while rapidly mixed using a stir bar. At various times thereafter, 10 ml aliquots were removed and added to 440 mL glycine pH 5 contained in 450 ml Sorvall centrifuge bottles. Bottles were capped by hand as quickly as possible and vigorously shaken. Samples were washed and worked up as detailed below.

To test different formaldehyde concentrations and other quenching conditions, TBP-myc cells were grown as described above. To test formaldehyde concentrations at 1% or lower, in most cases the appropriate volume of 37% formaldehyde was added to a rapidly stirring 100 ml culture, and the reaction was then quenched after specific incubation times by addition of 3 M glycine or 3 M Tris-HCl, pH 8, to achieve the indicated final quencher concentration. For reactions in which formaldehyde was added to a final concentration greater than 1%, cells were concentrated five-fold in YPD as described above, 37% formaldehyde was added to achieve the indicated final concentration, and after particular incubation times, 10 ml aliquots were removed to centrifuge bottles or tubes containing 3 M glycine or Tris yielding the final concentration of the quencher indicated in the figure legends. Cell samples quenched in Tris were worked up and analyzed as described above except that the first TBS wash contained 120 mM Tris-HCl pH 8 rather than glycine.

Quenching reversal experiments. To determine the stability of crosslinked material in the presence of quencher, crosslinked cells were incubated in solution containing glycine or Tris for different periods of time prior to ChIP work-up. For Tris-quenched samples, 100 ml cultures of AY146 cells were grown overnight in synthetic media lacking leucine and containing 2% glucose at 30°C. Cells were then transferred to YPD at an OD_{600} of 0.8 and grown until reaching an OD_{600} of 1. Each sample was crosslinked by adding formaldehyde to 1% for 5 minutes and then quenched by adding 10 ml 2.5 M glycine to each 100ml culture. Cells were pelleted and resuspended in either 750 mM Tris-HCl pH 8, or TBS buffer (which contains 50 mM Tris-HCl,

pH 8 as described above) and incubated at room temperature for 10 or 30 minutes. Subsequent steps were carried out as described below.

To test crosslink stability in the presence of glycine, 250 ml replicate cultures of AY146 cells in synthetic media plus 2% glucose and without leucine were incubated overnight at 30°C, then resuspended in YPD and grown to an OD_{600} of 1 as described above. Three aliquots of 50 ml were taken from each culture and pelleted at room temperature. Each pellet was then resuspended in 10 ml YPD and transferred to a flask on a stir plate. Formaldehyde was then added to 5% final concentration to each sample and mixed at room temperature for 5 minutes. 10 ml from each sample were quenched in 440 ml 3 M glycine pH 5 at room temp for 0, 10, or 30 minutes. The zero minute sample was pelleted at 4°C immediately after quenching; the other time point samples were pelleted the same way after glycine incubation of 10 or 30 minutes. Following incubation of the crosslinked cells in glycine solution for the indicated times, the cells were processed for ChIP as described below.

Order-of-addition experiments. Order-of-addition experiments were performed to test quenching efficiency using the TBP-myc strain, YAD154. Replicate cultures of YAD154 cells (300 ml) were grown overnight at 30°C in YPD to an OD_{600} of 1.0, then concentrated by resuspension in 60 mL YPD. In each experiment, three 10 ml aliquots were collected in duplicate: (1) no formaldehyde control samples in which 3 M glycine pH 5 was added to 2.93 M final concentration, (2) samples in which 3 M glycine was added to 2.93 M final concentration before 5% formaldehyde addition for 8 minutes, and (3) 5% formaldehyde incubation for 8 minutes followed by addition of 3 M glycine pH 5 to 2.93 M final concentration. Following these treatments, cell samples were washed in 50 ml TBS plus 300 mM glycine pH 5 followed by washing in 50 ml TBS, both washes at 4^{o} C. Subsequent work-up for ChIP and Real Time PCR for TBP binding to the *URA1* locus were performed as described above.

Order-of-addition experiments for Gal4 with the previously published CLK conditions (15) were done in the same way as order-of-addition experiments described above, except different glycine and formaldehyde concentrations were used. For each sample set, three 100 ml YPH499 cultures were grown overnight at 30°C in YEP + 2% raffinose. When an OD_{600} of 0.8 was reached, each culture was induced with 2% galactose. At OD_{600} of 1.0, samples were collected in duplicate. The following experimental parameters were used: (1) 2.5 M glycine pH 6.3 was added to 250 mM final concentration, (2) 2.5 M glycine pH 6.3 was added to 250 mM final concentration for 5 minutes before addition of 1% formaldehyde for 8 minutes, and (3) 1% formaldehyde incubation for 8 minutes before addition of 2.5 M glycine pH 6.3 to 250 mM final concentration for a 5 minutes incubation. The subsequent steps were the same as above, except analysis was performed for interaction at the *GAL3* locus.

Micrococcal nuclease (MNase) digestion. The protocol was adapted from a published MNase ChIP-seq method (Wal and Pugh 2012). Five ml YPD primary cultures were grown overnight at 30°C. These were added to 495 ml YPD media the next day and cells were grown until the OD_{600} was ~0.8. Each culture was then split into three 150 ml aliquots. Cells in one aliquot were collected by centrifugation for 5 minutes at 4000 rpm in an eppendorf 5810R benchtop centrifuge. Cells in the second aliquot were crosslinked with 5% formaldehyde for 30 seconds then quenched with 2.93 M glycine pH 5 for one minute, and then the cells were pelleted by centrifugation. The third aliquot was mixed with 2.93 M glycine pH 5 for one minute and cells were collected by centrifugation. Each cell pellet was washed with 50 ml ice cold TBS and cells were collected by centrifugation as above. Samples were processed and digested as described (Wal and Pugh 2012) with the following exceptions. Aliquots were taken for the undigested (input) samples and digestions were performed by adding 500, 1000 or 2000 units of MNase (Worthington Biochemical) as indicated in the figure. Samples were then resolved on 2% agarose gels, stained with ethidium bromide, and imaged.

Whole cell extract preparation and western blotting. Two sets of datasets were collected for western blotting: (1) duplicate WT and OE strains to determine the OE factor and (2) duplicate WT strains to measure the TF depletion over the experimental time course; samples for (1) were prepared as whole cell extracts (WCE) and samples for (2) were prepared as whole cell extracts and chromatin. Collection for both datasets began with a 5 ml YPD primary culture grown overnight at 30°C. The next day, cultures for the OE factor were diluted into 50 ml of appropriate media (SC-Leu for TBP and TFIIE CLK strains) and grown until the OD_{600} was ~0.8. Cells were pelleted for 5 min at 4,000 rpm and 4°C in an Eppendorf 5810R benchtop centrifuge. Cell pellets were resuspended in 50 ml YPD and incubated at 30 °C with shaking for about one hour until the OD_{600} was ~1.0. Cells were spun again as before; WCE preparation is described below. For the depletion samples, the primary culture was diluted into 300 ml YPD and grown to an OD_{600} = 1.0, then concentrated five-fold as described above. Following removal of a zero minute (no formaldehyde) control, formaldehyde was added to 5% and cells were incubated for various times at room temperature as indicated in the figures and then 10 ml aliquots were quenched in 440 ml of 3 M glycine pH 5. Samples were spun down and then prepared as either chromatin or whole cell extracts; the Benoit's buffer (200 mM Tris-HCl (pH 8.0), 400 mM $(NH_4)_2SO_4$, 10 mM $MgCl_2$, 1 mM EDTA, 10% glycerol, 7 mM β-mercaptoethanol) lysis extraction protocol was employed for whole cell extracts (Consortium 2012, Rhee and Pugh 2012). Chromatin extracts for western blotting were prepared in the same way as chromatin was prepared for ChIP. The only difference for chromatin samples was the use of 300 mM glycine pH 5 in the first TBS wash instead of 250 mM glycine pH ~6.3 (Consortium 2012, Rhee and Pugh 2012). Both chromatin and WCE protein levels were quantified with Bradford protein dye (Bio-Rad) using bovine serum albumin as the standard. 8% or 10% denaturing protein gels were used to resolve 15 μg protein for each sample. Unless otherwise noted, before loading the gel, samples were incubated at 95°C for five minutes. This heating step was left out for unheated samples.

Coomassie staining or membrane transfer was performed following electrophoresis. For staining, the gel was incubated with Coomassie dye (Research Organics Inc) for one hour at room temperature with gentle shaking, followed by overnight destaining (40% methanol, 10% acetic acid) at room temperature. The gel was imaged with the FluorChemQ system (protein simple). For gel transfer, proteins were transferred to Immobilon P and western detection of particular protein species was performed using the antibodies listed in Table 2.3 and detection with Amersham ECL Prime (GE Healthcare). Quantification of bands on the blots was done using ImageJ software (NIH). The overexpression factor was calculated by dividing the quantified OE band by the corresponding WT band; the values were averaged for at least two biological replicate sets of samples to determine the overall OE factor. The bands for the depletion time points were normalized to the zero minute time point and/or a loading control. The effects of formaldehyde incubation time on soluble TF levels were estimated by averaging the effects observed in at least two biological replicate sets of samples.

Table 2.3 Antibodies used for western blotting to check TF depletion in chromatin and whole cell extracts.

Antibody	Specific binding	Company	Catalogue Num.	Secondary Antibody	Protein on W.B.
α-Protein A	Protein A (TAP tag)	Sigma Aldrich	P-3775	α-rabbit	Tfa1 (TFIIE), Tfg1 (TFIIF), Reb1, Gal4, Sko1
α-TBP, monoclonal	TBP C-terminus	Abcam	ab61411	α-mouse	TBP
α-FLAG	FLAG peptide sequence	Sigma Aldrich	F1804	α-mouse	Abf1
α-GFP	GFP tag	Thermo Fisher	A11122	α-rabbit	Ace1
α-myc	c-myc	Abcam	ab32	α-mouse	TBP, Cat8
α-Sir2	Sir2 N-terminus	Santa Cruz Biotech.	sC6666	α-goat	Sir2
α-Sua7	rabbit polyclonal antiserum		D.T.A	α-mouse	TFIIB
α-Rpb1 8WG16	C-terminal Rbp1	Covance	MMS-126R	α-mouse	Rbp1 (Pol II)

Collection of crosslinking time points and preparation of chromatin samples. We found that collection of eight crosslinking time points in a single experiment was manageable. A single eight time point experiment performed with optimized glycine quenching required nearly 4 liters of 3M glycine, which was made by adding 900.84 g glycine (Bio-Rad) to a total volume of 4 L water. The solution was gently heated on a hot plate to help the glycine dissolve. The pH of the resulting solution was then adjusted to 5 using a few milliliters of concentrated HCl (Fisher). The glycine was then aliquoted into eight 500 ml bottles, each of which contained 440 ml of the solution. The flask containing 90 ml cell culture was rapidly mixed with a stir bar, and 14 ml 37% formaldehyde (Fisher) was added to the culture (resulting in 5% final formaldehyde concentration) at time zero. 10 ml aliquots of culture were then removed from the flask using a Pipet Aid and immediately added to the aliquoted glycine solution. For each sample, bottles were immediately capped and vigorously shaken for a few seconds to ensure good mixing. All subsequent steps were performed at 4°C by keeping the samples on ice, and using buffers and centrifuges chilled to 4°C. Quenched cell samples were pelleted by centrifugation for 7 minutes at 5000 rpm in an SLA-3000 rotor and Sorvall RC 5B centrifuge. Cell pellets were resuspended in 50 ml TBS plus 300 mM glycine and transferred to 50 ml conical tubes. The tubes were centrifuged for 5 minutes at 4000 rpm in an eppendorf 5810R benchtop centrifuge. Cell pellets were then washed with 50 ml TBS (40 mM Tris-HCl pH 7.5, 300 mM NaCl) and spun as before. Each pellet was transferred to a FastPrep tube and cell pellets were stored at -80°C for later work-up or resuspended in 600 µl 140 mM ChIP lysis buffer (50 mM Hepes pH 7.5, 140 mM NaCl, 1% Triton X-100, 0.1% sodium deoxycholate) with protease inhibitors (Roche Complete Protease Inhibitor Cocktail Tablet OR 1.0 mM phenylmethylsulfonyl fluoride, 2.0 mM benzamidine, 2.0 mM pepstatin, 0.6 mM leupeptin, and 2.0 mg of chymostatin per ml of buffer) for bead beading.

Once pellets were resuspended in ChIP lysis buffer, acid-washed beads (Sigma) were added to just above the liquid line and samples were processed for 7 cycles of 45 sec on, 1 min off in a FastPrep machine (MP Biomedicals). Tube bottoms were punctured with an 18-guage

needle (BD PrecisionGlide) and placed in 13 x 100 mm glass tubes and the flow-through liquid recovered by centrifugation for 5 min at 3000 rpm in an eppendorf 5810R benchtop centrifuge. Each sample was briefly vortexed and transferred to a 1.5 ml eppendorf tube on ice. Samples were then sonicated with a Branson Sonifier 250 with microtip probe for 7 cycles of 5 pulses each with 30% output and 90% duty cycle. This was followed with a 5 min spin at 14000 rpm and 4°C in an eppendorf 5415C benchtop centrifuge. The supernatant was transferred to a new eppendorf tube. Following a second spin for 20 min at 14000 rpm, supernatants were collected and the protein was quantified by Bradford protein assay as described above.

ChIP and real time PCR. Chromatin immunoprecipitation was performed with 1 mg total protein for each sample. For each time point IP, mock, and total (input) samples were assayed. IP and mock sample volumes were adjusted to 500 µl with 140 mM ChIP lysis buffer with protease inhibitors added. For TBP ChIP, 2.5 µl of anti-TBP antibody (Cat# ab61411, Abcam) was used in the IP. For TBP-myc, 2.5 µl of anti-Myc antibody (Cat#ab32, Abcam) was used. For LacI and AceI, 5 µl of anti-GFP antibody (Cat# A11122 Life Technologies Inc) was added to samples. The IP and mock samples were inverted overnight at 4°C. Following overnight incubation, the IP and mock samples were then incubated with 40 µl Sepharose A Fast Flow 4 beads (GE Healthcare) for 2 hours at 4°C. Samples were washed twice with 1 ml of 140 mM ChIP lysis buffer, 500 mM ChIP lysis buffer (same as 140 mM ChIP lysis buffer but containing 500 mM NaCl), LiCl wash buffer (10 mM Tris pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA), and 1X TE (10 mM Tris-Cl pH 8.0, 1 mM EDTA). Two elutions of the bound material were performed by adding 75 µl elution buffer (50 mM Tris-HCl pH 8.0, 1% SDS, 10 mM EDTA) to each sample for 10 min at 65°C. The two elutes were combined and incubated overnight at 65°C along with the total samples, which consisted of 0.1 mg input chromatin protein combined with 150 µl elution buffer. The following day, samples were cleaned up using

the QiaQuick PCR cleanup kit (Qiagen) following the manufacturer's instructions and DNA was eluted with 50 µl DEPC water pre-warmed at 55°C.

ChIP for TFA1-TAP was performed as described above, except 40 µl of a 50% slurry of IgG Sepharose 6 Fast Flow beads (GE Healthcare) was added to the IP sample and 40 µl of a 50% slurry of Sepharose 6 Fast Flow beads (GE Healthcare) was used for the mock samples. An overnight IP was carried out at 4°C followed by washing the bead pellet the next day as described above.

To quantify the ChIP DNA, real time PCR was performed using appropriate primer sets (listed in Table 2.4), iQ SYBR Green Supermix and a MyiQ instrument (Bio-Rad). A standard curve consisting of four serial dilutions of the input was also run to determine ChIP levels for the IP and mock samples. The standard curve inputs were run in duplicate and all unknowns (IP, mock, total samples) were run in triplicate. The relative ChIP signal for each time point was calculated by subtracting the mock signal from the IP signal and then dividing by the total signal (5X input sample). The kinetic data reported here represent the average from at least two independent experiments for each strain and condition; this averaged data was plotted before moving onto the fitting procedure to check for consistencies between the two datasets.

Table 2.4. Oligonucleotides used for Real-Time PCR (5'-3') in this chapter.

Name	Sequence
ACT1-F	CAG CTT TTA GAT TTT TCA CGC TTA
ACT1-R	TTT TCG ATC TTG GGA AGA AAA A
LOS1-F	TTT GAG AAG TTG TCG GTA AGC A
LOS1-R	GCA TTC CTC GAT TTG ACT GG
URA1-F	AAG ATG CCC ATC ACC AAA AA
URA1-R	AAG AAT ACC GGT TCC CGA TG
CU3-F	TCT CGG CCT AGC TCA TCA GT
CU3-R	AAG ACA GAT CCA CGT CTT TGG
NTS2-F	GCA CCT GTC ACT TTG GAA AAA
NTS2-R	TCG CCG AGA AAA ACT TCA AT
HSC82-F	TCT TGA AAC GCT ACA GAA CCA A
HSC82-R	CAC CAG CCA TAT TTC AGA ATG A
U6-F	TTC GTC CAC TAT TTT CGG CTA
U6-R	GGG TTA CTT CGC GAA CAC AT
CUP1-F	AGA AGC AAA AAG AGC GAT GC
CUP1-R	GAC AAT CCA TAT TGC GTT GG

Statistical Analysis. Student t-tests (non-paired, two-tail, equal variance) were used to compare data sets represented on scatterplots and marked with an asterisk (*) for significance ($p < 0.05$).

Model fitting and data analysis

Poorey *et al*. (*Science*, 2013) developed the mass-action kinetic model used to analyze the data generated in that study and extract kinetic parameters. The procedure to fit the data to this model

was systematically studied and re-formulated in Zaidi *et al*. (2017) based on the experimental regimes seen in the experimental data. A general flowchart for fitting the data is below (Figure 2.7).

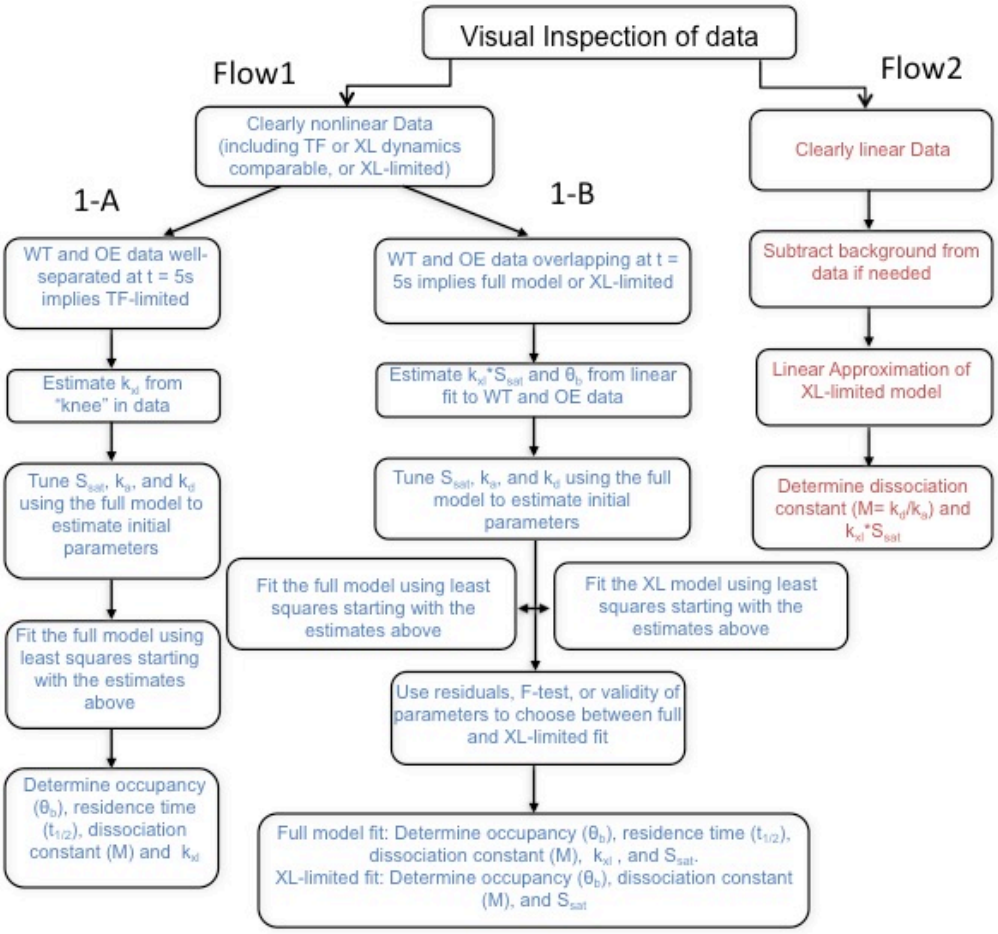


Figure 2.7. Flow chart for fitting CLKv2 data. After visual inspection of the data, the fitting procedure in either Flow1 arm (blue) or Flow2 arm (red) was followed for a given locus. Fitting procedure for non-linear data was further broken down into arm 1-A or 1-B. Arm 1-A represents the TF full-model fit, while 1-B could be crosslinking-limited or full model fit. Flow2 is the linear crosslinking-limited fit.

Computational Modeling. The Crosslinking Kinetics (CLK) model is described by Eqns. 7, 11, and 16 in Sec. 2.2 of Poorey et al. (Poorey et al. 2013) Supplementary Material. The model is characterized by the transcription factor association rate (k_a) and disassociation rate (k_d) of binding to chromatin, the formaldehyde-transcription factor crosslinking rate (k_{xl}), the saturation level of the ChIP signal (S_{sat}), the transcription factor concentration *in vivo* (C_{TF}), and the formaldehyde concentration (C_{FH}). The ChIP signal, S(t), is related to the *in vivo* fraction of a given binding site cross-linked by the TF across cells (θ_{xl}) by the relationship $\theta_{xl}(t) = S(t)/S_{sat}$, where S_{sat} is the saturation value of the ChIP signal. This scaling of the ChIP signal ensures that $\theta_{xl}(t)$ approaches 1 as crosslinking time goes to infinity, as required by the CLK model. Two physically interpretable parameter regimes of the model are the transcription-factor dynamics limited (TF-limited) regime where TF dynamics are much slower than crosslinking dynamics (i.e., $k_a*C_{TF} << k_{xl}*C_{FH}$ and $k_d << k_{xl}*C_{FH}$) and the crosslinking dynamics limited (XL-limited) regime where crosslinking dynamics are much slower than TF dynamics (i.e., $k_{xl}*C_{FH} << k_a*C_{TF}$ and $k_{xl}*C_{FH} << k_d$), as detailed in Sec. 2.3 of Poorey et al. (Poorey et al. 2013) Supplementary Material. Finally, for extremely slow crosslinking dynamics which occur on the timescale of the full range of crosslinking times or longer (i.e., $k_{xl}*C_{FH}*\theta_b*t_l << 1$ where t_l is the last crosslinking time point, which is usually 1200s), the CLK model predicts that $\theta_{xl}(t)$ will be a linear function of crosslinking time, $\theta_{xl} \sim k_{xl}*C_{FH}*\theta_b*t$. Notably, we observe TF-limited, XL-limited, and linear in crosslinking time CLK curves depending on the TF and locus examined.

The simulations presented in Fig. 2.16 show the expected CLK curves in the TF-limited, the XL-limited, and the linear regimes, while the schematic diagram shows the physical interpretation of the *in vivo* dynamics in these regimes. The hallmarks of the TF-limited model are a relatively fast exponential rise at time scales of less than ~100 seconds but often less than 5 seconds (first crosslinking time point in the experiment) followed by a slower exponential rise (see Fig. 2.16A). Notably, when the first relatively fast exponential rise is less than 5 seconds, we

observe a non-zero y-intercept in the WT and OE data with a clear separation between the WT and OE y-intercepts. When the rise in the first relatively fast exponential is ~100 seconds, we find a zero y-intercept, an initial fast exponential rise in the data followed by a slower exponential rise, hence forming what looks like a "knee" in the data around the transition from the fast to the slow exponential for both the WT and OE data. Interestingly, the y-intercept for very fast crosslinking or "knee" for modestly fast crosslinking in the WT data yields an excellent approximation of the in vivo occupancy, θ_b. The XL-limited model shows a single exponential rise with a zero y-intercept for the WT and OE data (see Fig. 2.16B). The linear model shows a near-zero y-intercept at t = 5 seconds, and no sign of saturation on the experimental time scale of 700 seconds to 1200 seconds (see Fig, 2.16C). Importantly, the two crosslinking dynamics limited models, XL-limited and linear, display relatively high sensitivity to formaldehyde concentration (as shown in Fig. 2.9B, C) while the TF-limited (which we also refer to as the "full model" (see Tables 2.6-2.9) for reasons described below) does not. While the full mathematical model presented in equations (11) and (16) in the Supplemental Material of Poorey et al (2013) can be used to fit and represent all of these parameter regimes, we use and refer to a "full model" fit for data that clearly show the double exponential behavior (i.e., relatively fast crosslinking rise followed by a second TF-dynamics limited rise with a relatively clear kink or knee in between the two). Moreover, in the case of XL-limited behavior, we use the single exponential XL-limited model shown in equation (21) of the Supplementary Materials of Poorey et al (2013), which is a highly accurate approximation of the "full" model (equations (11) and (16) in the Supplemental Material of Poorey et al (2013)) in the XL-limited parameter regime. Finally, for linear in crosslinking time data, we use the linear model shown in equation (22) of the Supplementary Materials of Poorey et al (2013), which is a highly accurate approximation of the "full" model in the very slow crosslinking dynamics parameter regime.

For data that showed negative curvature (i.e., TF-limited or XL-limited), we started by visually estimating S_{sat} to be close to the late time point over-expression ChIP signal. Hence, our

initial guess was normally S_{sat} between 1 and 5, except for LacI, where we started with S_{sat} ~10. In the case of data that visually showed TF-limited behavior (e.g., TBP at *ACT1*, *LOS1*, and *URA1*), we estimated the initial value for k_{xl} by looking at the time (τ_{xl}) around which the data showed a "knee." Setting $\ln[2]/k_{xl} \sim \tau_{xl}$ gives an estimate for k_{xl}. The y-intercept of a linear extrapolation of the late-time S(t) data points (i.e. linear extrapolation of the S(t) data points that are approximated by the second exponential) divided by S_{sat} gives an initial estimate for θ_b. The in vivo occupancy, θ_b, is expressed in terms of k_a and k_d as $\theta_b = k_a*C_{TF}/(k_a*C_{TF}+k_d)$. For a given θ_b, we can sweep over a wide range of k_a and S_{sat} values to see where the theoretical curves match with the WT and OE experimental data. Importantly, the overall on-rate, k_a*C_{TF}, dominates the rate at which the second exponential rises. With these starting estimates for the kinetic parameters, we run the NonLinearModelFit routine in Mathematica (Wolfram Research 2016) to fit the full model to the data using least squares. The fit reliably gives us k_a, k_d, and S_{sat} (equivalently, S_{sat}, $\theta_b = k_a*C_{TF}/(k_a*C_{TF}+k_d)$, and $t_{1/2}=\ln(2)/ k_d$).

For data that did not show TF-limited behavior (but still showed negative curvature, as opposed to a purely linear response, for example, TFIIE at *ACT1* and *URA1*), there were two possibilities: either the data was XL-limited (showing a single exponential), or the knee was not markedly visible by inspection because of the experimental time scales. We started by fitting a straight line to the short crosslinking time data to estimate $k_{xl}*S_{sat}$ and θ_b. With these estimates, we swept over a wide range of k_a, k_{xl}, and S_{sat} values to match the theoretical full model with the data. With these tuned estimates, we fit both the XL-limited model and the full model to the data, and determined which model yielded a better fit of the data by looking at the validity of parameters obtained, the sum of squared residuals (SSR), or by conducting an F test.

For data that fit the linear XL-limited model best (e.g., TBP at *NTS2* and *ACE1*), we subtracted the y-intercept (extrapolated ChIP signal at t = 0 second) from the data as background,

and fit a line to each of the WT and OE data using least squares. The overexpression factor is known (Table 2.5), so we could extract $k_{xl}*S_{sat}$ and θ_b from the two slopes.

Table 2.5. Estimate of nuclear protein concentrations for factors used in the chapter, based on nuclear volume from Jorgensen et al, 2007.

Factor	Concentration in the nucleus (μM)	Reference	Overexpression concentration (μM) (This study)
Ace1-GFP	1	Ghaemmaghami S et al, 2003 (4), Karpova et al, 2004 (8)	10
TBP	12	Borggrefe et al, 2001 (9)	38
LacI-GFP	1	Poorey et al, 2013 (3)	3.6
Tfa1-TAP	14.4	Borggrefe et al, 2001 (9)	26.5

For some loci it was not obvious if the data would fit the full model/TF-limited model or the linear XL-limited model (e.g., TBP at *HSC82* and *U6*). It was important to answer the question of the better fit because the two models have a different number of effective parameters: the full-model fit has four free parameters (S_{sat}, k_a, k_d, and k_{xl}), while the linear regime has only one: $S_{sat}*k_{xl}*\theta_b$. The sum of squared residuals (SSR) with the linear fit (with fewer degrees of freedom) was lower than the SSR with the full model fit (with more degrees of freedom); hence, the linear fit was chosen without the need to conduct an F-test comparing the two models. The full model fit gave worse SSR values because we were explicitly starting with estimates close to the TF-limited regime when fitting the full model, which lead the minimization of the difference between the model and data to a suboptimal, local minimum. Note that the SSR was calculated without normalizing the data using S_{sat} because the SSR scales with S_{sat} and S_{sat} is unknown in the linear fit case.

For TFIIE at *ACT1*, the final parameters from the XL-limited fit were unphysical ($\theta_b \sim 0$ and $K_d = k_d/k_a \sim 10^7$ mol); hence, the full model fit was chosen. An F-test was performed to choose the XL-limited fit for TFIIE at *LOS1* over the full model fit. For TFIIE at *URA1*, the parameter estimates from a full model fit satisfied XL-limited binding dynamic conditions. Therefore, the TFIIE data at *LOS1* and *URA1* were fit with the XL-limited model.

To estimate the errors associated with our output parameters, we ran our fitting procedure on simulated data for each locus. Specifically, we simulated the data at each locus with the mean value at each time point given by the theoretical fit and the variance given by the mean of the squared residuals. We simulated and fit the data at each locus for one thousand successful fitting iterations. The standard deviation in the simulated fit parameters was calculated on the log scale (see Figure 2.20 and 2.21), and was transformed back from the log scale to determine the lower and upper bounds on the error bars quoted in Tables 2.6-2.9. Error bars for k_{xl} could not be estimated in the case of TBP at *LOS1* since the fit parameters were TF-limited and fitting the full-model to the simulated data gave spurious values for k_{xl} in addition to failing often. Hence the error bars for k_a, C_{TF}, k_d and S_{sat} for TBP at *LOS1* were calculated by fitting the TF-limited model to the simulated *LOS1* data.

In addition to the parameter error analysis, we assessed the significance of each of the fits as well as the parameters derived from the fits by calculating their associated adjusted R-squared and p-values, respectively, as shown in Tables 2.10-2.12. Importantly, the estimation of p-values relies on an approximation that is equivalent to the model being linear in the parameters, which it is not. Nevertheless, we estimated parameter p-values as guides to their relative significance. We found all the fits to significantly explain the variance in the data—accounting for the number of fitting parameters—with the adjusted R-squared ranging from 0.93-0.99 for unambiguous fits (i.e., excluding full model fits of TBP at *SNR6* and *HSC82*). We also found all kinetic parameters were significantly different than 0 except the on-rate for LacI at the *lac* array and TFIIE fits, suggesting that these were the least significant parameters. In the case of TFIIE, we also estimated the 95% confidence intervals of the parameters, which take into account significance information via the linearized model assumptions. For a number of the kinetic parameters, the linearized model assumption produced a lower limit of the confidence interval that yielded an unphysical, negative value, which we set to zero. For linear fits, we performed linear regression analysis separately for the wild type and overexpressed TF data from which we derived linear

coefficients (i.e., beta = θ_b*k_{xl}*S_{sat}*C_{FH}) shown in Tables 2.10 and 2.11 along with each fit's adjusted R-squared value. We calculate the occupancy and association constants shown in Tables 2.7 and 2.8 using the wild type and overexpressed beta values shown in Tables 2.10 and 2.11.

Results

Determining updated experimental conditions

The CLK method relies on time-resolved formaldehyde crosslinking ChIP data to assess the kinetics and thermodynamics of TF-chromatin binding. The original CLK method (Poorey et al. 2013, Viswanathan et al. 2014) employed 1% formaldehyde (360 mM) and reactions were quenched with 250 mM glycine (Kuo and Allis 1999, Wu et al. 2011). Under these conditions, the concentration of glycine is sub-stoichiometric to the formaldehyde concentration as added, but crosslinking was performed by adding formaldehyde to cells in YPD medium, which is made from an amino acid-rich extract of yeast cells and as such, the concentration of unreacted formaldehyde that reaches cells under these conditions is unknown and is most likely well below the initial concentration. Order-of-addition experiments showed that 250 mM glycine could block crosslinking of the Gal4-promoter interaction (Poorey et al. 2013), but we noted in subsequent work that quenching may be variably efficient under these conditions (Hoffman et al. 2015). Indeed, we have noticed that for unknown reasons the quench efficiency can be variable from experiment to experiment for certain TFs (Fig. 2.8) (Sutherland et al. 2008, Zaidi et al. 2017). To better define time-dependent crosslinking behavior and the impact of different quenching conditions on the resulting ChIP signals, data were obtained using 1% (360 mM) formaldehyde and either 250 mM or 2.93 M glycine using the interaction between yeast TBP-myc and the *URA1* promoter as a model interaction. The high concentration of 2.93 M glycine used in this and subsequent experiments was the maximum achievable based on the solubility of glycine in aqueous solution (~3M) and subsequent dilution resulting from addition of a relatively small volume of concentrated yeast cell culture to the quenching solution (see Experimental

procedures). For this reason, we refer to this as the "max glycine" quench condition hereafter. As shown in Fig. 2.9A, the max glycine quench conditions resulted in lower ChIP signals at each time point compared to 250 mM glycine. These results demonstrate that the concentration of glycine used in the quench can have a significant effect on the magnitude of the ChIP signal, suggesting that more robust quenching of formaldehyde can be achieved with a higher concentration of glycine.

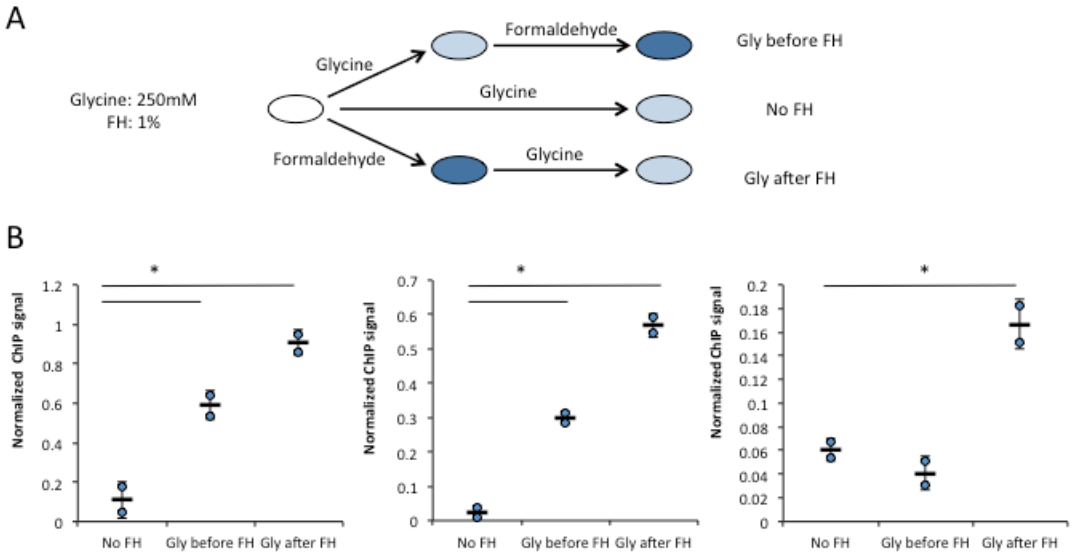


Figure 2.8. Variability observed in quench efficiency with original crosslinking and quenching conditions. A) Order of addition experiments identical to setup in Fig. 3, but with 1% formaldehyde and 250mM glycine, pH ~6.3 conditions. B) Plots of the real time PCR quantitation for three different sets of samples of Gal4 binding to the *GAL3* promoter. Experiments were done in duplicate with standard deviation shown.

In addition to lower signals at each time point obtained using max glycine conditions, some time-dependent datasets showed initial shallow slopes, which continuously increased until the curve reaches apparent linear behavior at longer times (Fig. 2.9). We refer to this as "positive

curvature". This type of behavior has several possible explanations (discussed below) but none are accounted for in the original CLK model. To better understand how glycine concentration affected the time course of formaldehyde crosslinking, experiments were performed to test both the dependence of the reaction on formaldehyde concentration and how ChIP data were affected using Tris, rather than glycine, to quench the reaction. Tris has been reported to be a robust quencher of formaldehyde reactivity (Sutherland et al. 2008). As shown in Fig. 2.9B, using max glycine quenching conditions, the ChIP signal depended on the formaldehyde concentration, as reaction with 4.7% formaldehyde increased the ChIP signal at each time point compared to reactions that employed 1% formaldehyde. A dependence on formaldehyde concentration was also seen in reactions using Tris as the quenching agent (Fig. 2.9C). However, in reactions that were quenched with Tris, the ChIP signals obtained for a given concentration of formaldehyde were reduced compared to the values obtained using glycine, and the resulting reaction progress curves showed positive curvature similar to reactions quenched with max glycine discussed above.

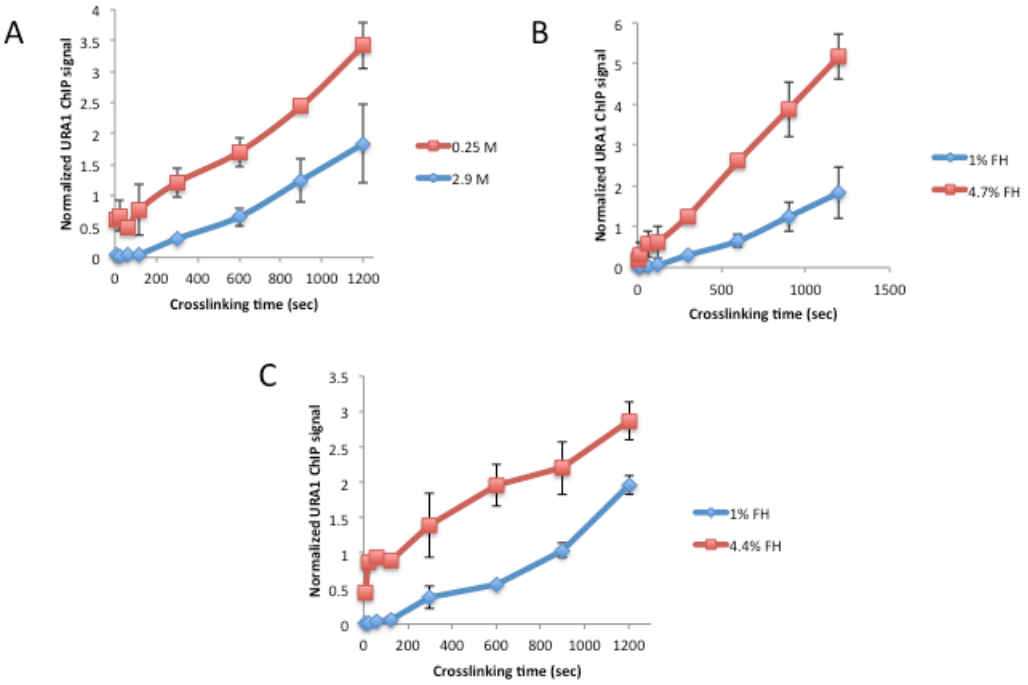


Figure 2.9. Effect of different formaldehyde and quench conditions on TBP-myc ChIP signal at the *URA1* locus. A) The TBP-myc strain was crosslinked for varying amounts of time with 1% (360 mM) formaldehyde followed by quenching with either low (0.25 M, red line) or high (2.9 M, blue line) glycine. Chromatin immunoprecipitation (ChIP) was performed followed by analysis with real time PCR at the *URA1* locus. Normalized ChIP signal is the IP signal minus mock signal divided by an input signal; values were determined from a standard curve. B) Similar to A, but TBP-myc expressing cells were crosslinked for varying amounts of time with 1% (blue line) or 4.7% (red line) formaldehyde and quenched with high (2.9 M) glycine. C) Similar to B, but 1% (blue line) or 4.4% (red line), formaldehyde was used for crosslinking and 600 mM Tris pH 8 for quenching. For each plot, two replicates were collected and the error bars represent the standard deviation. The data in all panels is from David Auble.

Although Tris is apparently a more efficient quencher than glycine, it also has the potential to reverse crosslinks (Shi et al. 2013, Kawashima et al. 2014). Crosslink reversal would be problematic for the CLK assay as it could lead to underestimates of ChIP signal, with potentially large percentage-wise effects on the modest levels of crosslinked material obtained after short crosslinking times. To test the potential for reversal with both Tris and glycine, samples were crosslinked, quenched, and incubated at room temperature for different periods of time in the quenching solution. As shown in Fig. 2.10A, incubation of cells in Tris-containing solution led to a loss of TBP ChIP signal over time. While the diminished ChIP signal from 10 to 30 minutes was only statistically significant with 50 mM Tris, the ChIP signal obtained with 750 mM Tris from 10 to 30 minutes also trended downward. In contrast, there was no detectable decrease in TBP ChIP signal over time when crosslinked cells were incubated in max glycine solution (Fig. 2.10B). Thus, although Tris is a robust quenching agent, we ruled out its use in the assay because it decreased the recovery of crosslinked complexes.

The results thus far led to implementation of two significant changes in the CLK methodology. First, to obtain the most accurate time resolved ChIP data, we employed the more robust quenching afforded by max glycine conditions, which lack the negative attributes of Tris as a quencher. Second, as the crosslinking rate is dependent on formaldehyde concentration, we employed 5% formaldehyde rather than 1% as used in previous work (Poorey et al. 2013) (and most ChIP experiments published to date). While 5% formaldehyde optimized the assay for analysis of several interactions in this study, it will be important to determine the optimal formaldehyde concentration for analysis of other types of interactions and in other cell types. We sought the highest feasible formaldehyde concentration for two reasons. First, experimentally, we wanted the ChIP signal to be minimally affected by noise. Second, since the overall crosslinking rate depends on the formaldehyde concentration, faster crosslinking would yield better time resolution between the crosslinking and binding dynamics timescales. To achieve the desired concentrations of reagents in the reactions and to obtain sufficient cellular material for analysis, cell cultures were concentrated by centrifugation, formaldehyde was added to the concentrated cell suspension, and then aliquots of cells were quenched by dilution in a much larger volume of glycine at high concentration. This approach also has the advantage that formaldehyde reactivity is reduced by dilution to 0.1% after glycine addition. Prior work showed that little crosslinking was detectable using 0.1% formaldehyde so dilution alone was expected to have a substantial impact on formaldehyde reactivity (Figure 2.11) (Viswanathan et al. 2014). In addition, the glycine quenching solution was adjusted to pH 5 which further improves the ability of glycine and formaldehyde to react (Sutherland et al. 2008). We refer to the experimental approach employing all of these modifications as CLKv2 (Fig. 2.12A) to distinguish it from the original CLK method.

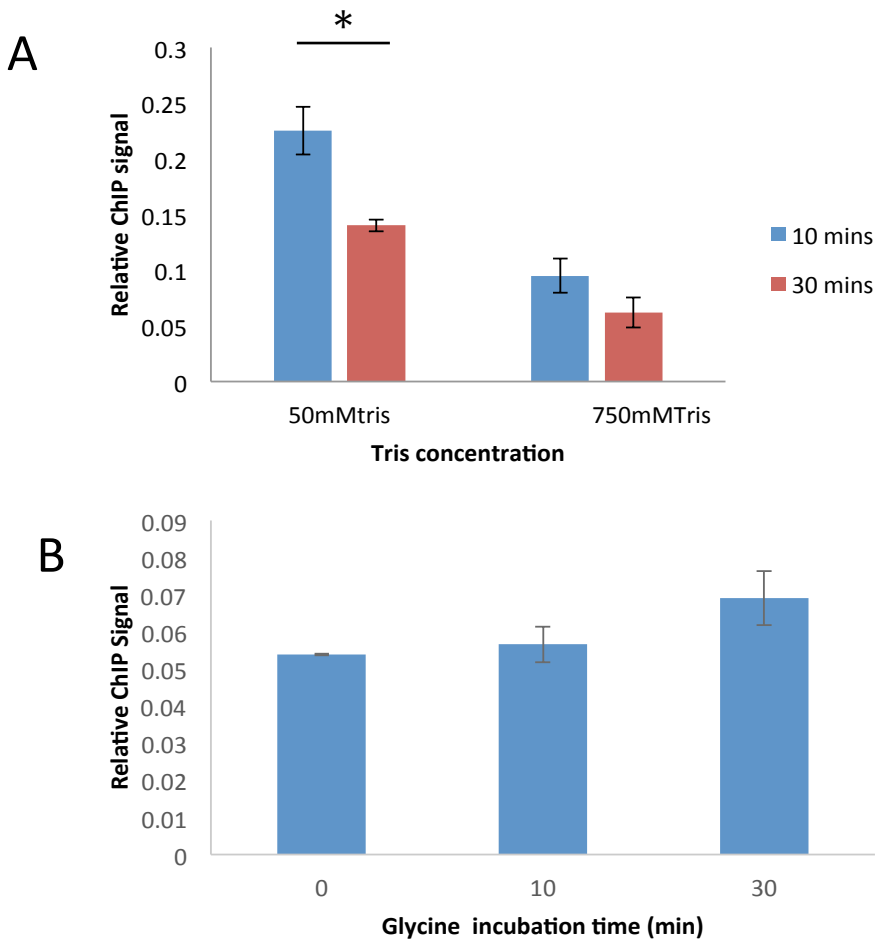


Figure 2.10. Tris, but not glycine, quenching reverses ChIP signal over time. A) Average TBP ChIP signal at the *URA1* locus in cells crosslinked with 1% formaldehyde, quenched with 250 mM glycine, and resuspended in either 50 mM or 750 mM Tris, pH 8. Samples were incubated at room temperature for 10 (blue circles) or 30 (red circles) minutes before processing. A student t-test was performed to determine statistical significance between conditions. B) Average ChIP signal of cells crosslinked with 5% formaldehyde and quenched with 2.93 M glycine pH 5. Samples were incubated in glycine quench solution for 0, 10 or 30 minutes before processing. All experiments were performed with two biological replicate samples and error bars represent the standard deviation. * $p < 0.05$. The data in both panels is from Savera Shetty.

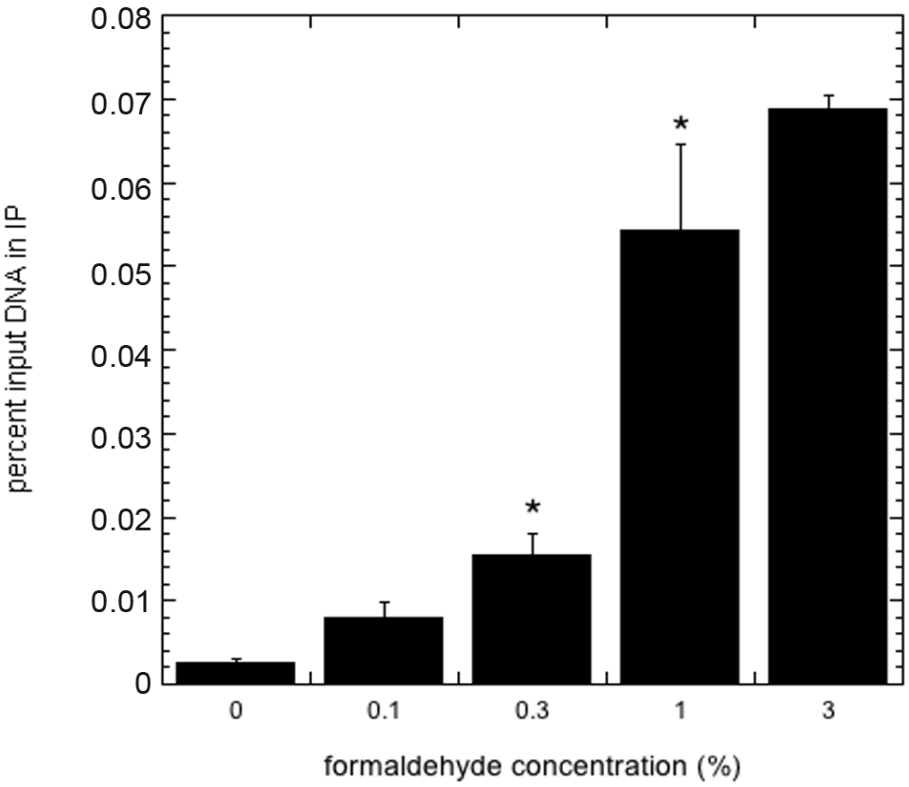


Figure 2.11. Dependence of ChIP signal on formaldehyde concentration. Wild-type YPH499 cells were incubated with the indicated concentrations of formaldehyde for ten minutes, then chromatin was isolated and ChIP was performed to quantify the association of TBP with the *URA1* promoter. **(A)** Percent of input DNA in the anti-TBP immunoprecipitates. Asterisks indicate sample comparisons with p values < 0.05: ChIP signal obtained with 0.3% formaldehyde versus no formaldehyde (p = 0.02, two-tailed Student's T-test)); 1% formaldehyde ChIP signal versus 0, 0.1% and 0.3% p values were all < 0.03. The difference between the 1% and 3% samples was not significant (p value = 0.187). Error bars represent the standard deviation obtained from two biological replicates. Data from David Auble.

As shown in Figs. 2.12B and C, order-of-addition experiments established that glycine was a very efficient quencher of formaldehyde reactivity when used in this way; the TBP-myc ChIP signal obtained in reactions in which formaldehyde was added first was ~28-fold higher than in reactions with no formaldehyde. In contrast, the ChIP signal obtained when glycine was added before formaldehyde was not statistically different from the background ChIP signal obtained with no formaldehyde at all ($p = 0.20$). Next, the use of 5% formaldehyde prompted us to evaluate how this higher level of formaldehyde might generally impact cellular constituents. As shown in Fig. 2.13A, protein yields were reduced in whole cell extracts prepared from cells treated with 5% formaldehyde for increasing periods of time. In contrast, there was no change in the yield of chromatin protein associated with extracts prepared as normally done for ChIP. In addition, there was little change in the pattern of protein bands or their relative intensities over a time course of formaldehyde incubation, indicating that the majority of proteins present in these chromatin extracts were not notably depleted or modified (Fig. 2.13B). This suggests that the reduced yield of protein in whole cell extracts was due to crosslinked cells being refractory to lysis by rapid agitation with glass beads, whereas soluble protein contents were more efficiently released in the chromatin extract preparation procedure, which utilizes a combination of glass bead agitation plus sonication. Protein samples are typically heated to facilitate their denaturation prior to electrophoresis, but formaldehyde crosslinks are also reversible by heat so we analyzed protein extracts on gels with and without heating. There was relatively little difference in overall protein banding pattern when chromatin extract proteins were analyzed following brief heating to facilitate protein denaturation versus unheated samples (Fig. 2.13B). Heating did reduce an indistinct smear of protein toward the top of the lanes of unheated samples, consistent with heat improving denaturation of the samples. Brief heating had a dramatic effect on the ability to detect TBP in extracts by western blotting (Fig. 2.13C). The formaldehyde crosslink reversal time is much longer than this brief heating period (Quievryn and Zhitkovich 2000), suggesting that heating in this experiment facilitated disruption of TF-protein complexes and protein

unfolding rather than crosslink reversal. In the case of TBP, it is likely that its association with TAFs (TATA-binding protein associated factors) and potentially other regulatory factors in extracts (Thomas and Chiang 2006) make detection of monomeric TBP difficult or impossible without heating. MNase titration experiments were performed to examine the overall chromatin state resulting from treatment with crosslinker and quencher. Compared to untreated cells, we observed no difference in the overall chromatin digestion pattern when cells were treated with 2.93 M glycine, pH5 with or without prior formaldehyde incubation (Fig. 2.13D). Thus, formaldehyde and glycine do not cause any detectable bulk structural changes in chromatin or affect chromatin accessibility.

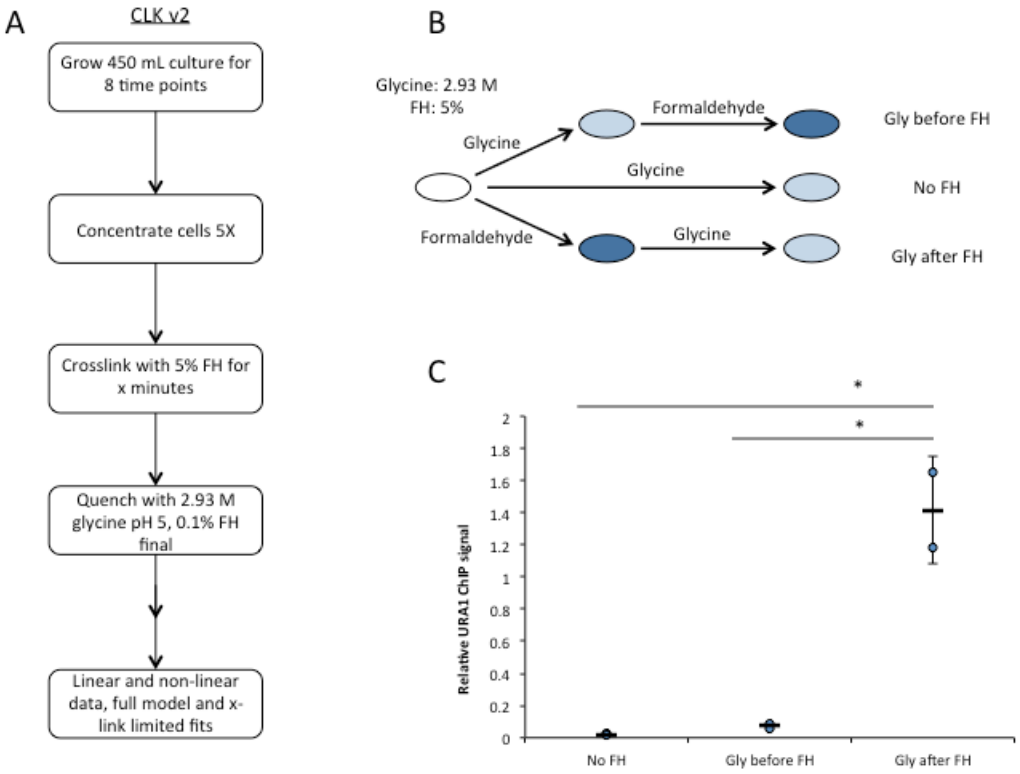


Figure 2.12. CLK v.2 quenching conditions and overview of the updated method. A) Flowchart of CLK v.2 method focusing on sample collection. B) Order of addition experiments to verify new excess glycine conditions are shown in the schematic. Three experiments were set up:

1) glycine alone added to samples, 2) glycine addition to samples then formaldehyde crosslinking, and 3) formaldehyde crosslinking followed by glycine quenching. For all samples, 5% formaldehyde and 2.93 M glycine pH 5 were used. C) Real time PCR read out from experiments done in B; data were obtained from two biological replicates and the error is the standard deviation. * $p < 0.05$. The data in panel C is from Savera Shetty.

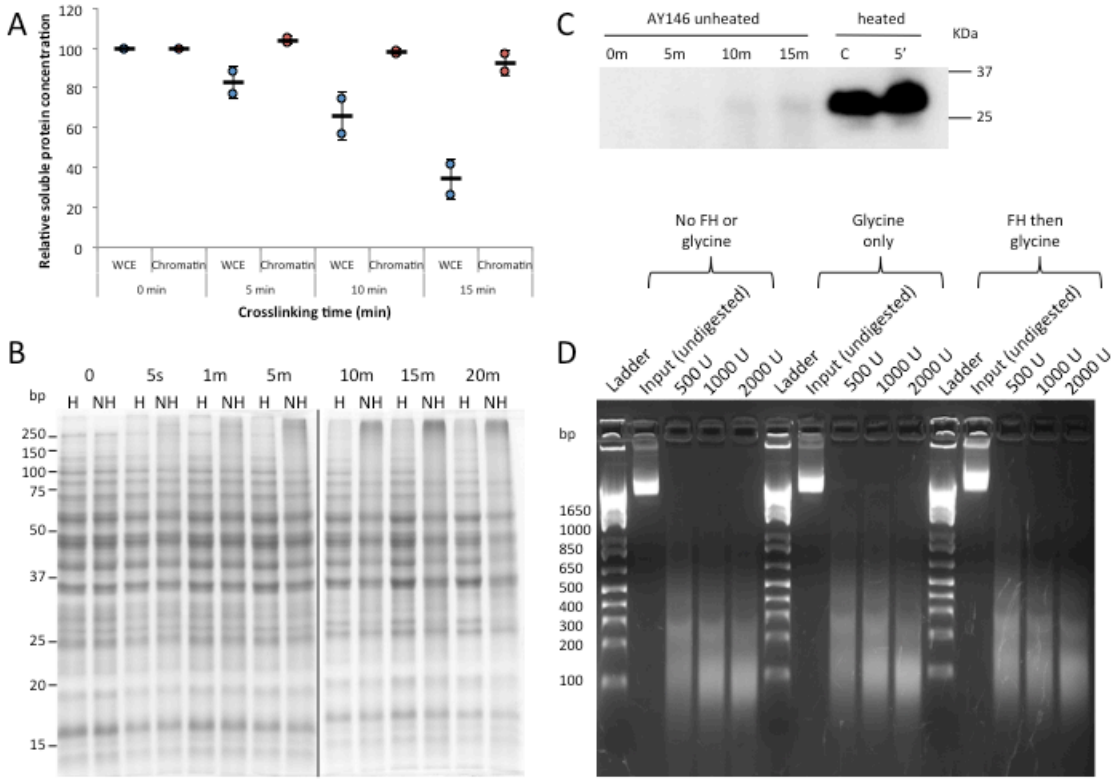


Figure 2.13. Effect of formaldehyde crosslinking on proteins. A) Relative concentration of protein in either whole cell extract (blue circles) or chromatin (red circles) samples crosslinked with 5% formaldehyde for varying amounts of time. Bradford assays were used to determine the concentration. Duplicates were used for each time point and samples were normalized to their respective zero time point. Error bars represent the standard deviation. B) Coomassie stained SDS-PAGE gel of AY146 whole cell extract samples from cells crosslinked for varying amounts of time with 5% formaldehyde. Fifteen microgram samples were heated (H) for 5 minutes at

95°C or not heated (NH) before loading. C) Samples from the AY146 strain were crosslinked for 0, 5, 10, or 15 minutes with 5% formaldehyde and either heated for five minutes at 95°C or unheated before loading into an SDS-PAGE gel. The western was probed with a TBP antibody and visualized with chemiluminescence. C (control) is recombinant TBP protein. D) MNase digestion of chromatin from wild type cells visualized on a 2% agarose gel. Cells were treated with no formaldehyde or glycine, 2.93 M glycine pH 5, or 5% formaldehyde for 30 seconds followed by 2.93 M glycine pH 5. The indicated amounts of MNase (units) were added to aliquots of cells and the resulting chromatin resolved by gel electrophoresis. The base pair lengths of molecular weight standards are denoted to the left of the image. The images in panels B and C are from Savera Shetty.

Examining suitable factors with new conditions

A key requirement for the CLK method is that the unbound pool of the TF being investigated is not depleted significantly by formaldehyde incubation (Poorey et al. 2013). This ensures that there are sufficient molecules available for interaction with unbound DNA sites and that the overall on-rate, which depends on the concentration of the free TF, does not change over the course of the reaction. To determine the effect of 5% formaldehyde on the soluble pools of particular TFs, western blots were performed using extracts obtained from cells treated with formaldehyde for various periods of time. Based on the results in Figs. 2.14B and C, a brief heating step was used prior to loading samples on the gels in order to accurately estimate the relative amount of soluble TF without reversing any crosslinks that had formed. Western blotting showed that 5% formaldehyde treatment resulted in depletion of some TFs and not others, and the rates of depletion among those that were depleted varied significantly (Fig. 2.14A-F, 2.15A-D). TBP, Gal4, and Ace1 were not significantly depleted in these experiments, whereas Reb1, Cat8, Abf1, TFIIB and Tfa1 were stable for ~10 min and then were depleted. In contrast, the largest subunit of RNA polymerase II, Rpb1, and the TFIIF subunit Tfg1 were rapidly depleted. Both

whole cell extract (WCE) and chromatin samples showed the same depletion patterns; chromatin samples were not prepared for Sir2 or Cat8 because the WCE samples showed depletion (Fig 2.14A, D). This indicates that some factors such as TBP and Gal4 are readily amenable to analysis by CLKv2. As shown below, others such as Tfa1 can be investigated as long as the crosslinking time course is confined to the period in which the levels of the factor are not depleted. Other factors such as Rpb1 and Tfg1 cannot be investigated at present using these conditions. However, it should be noted that in principle one could incorporate the TF depletion rate into the dynamic model.

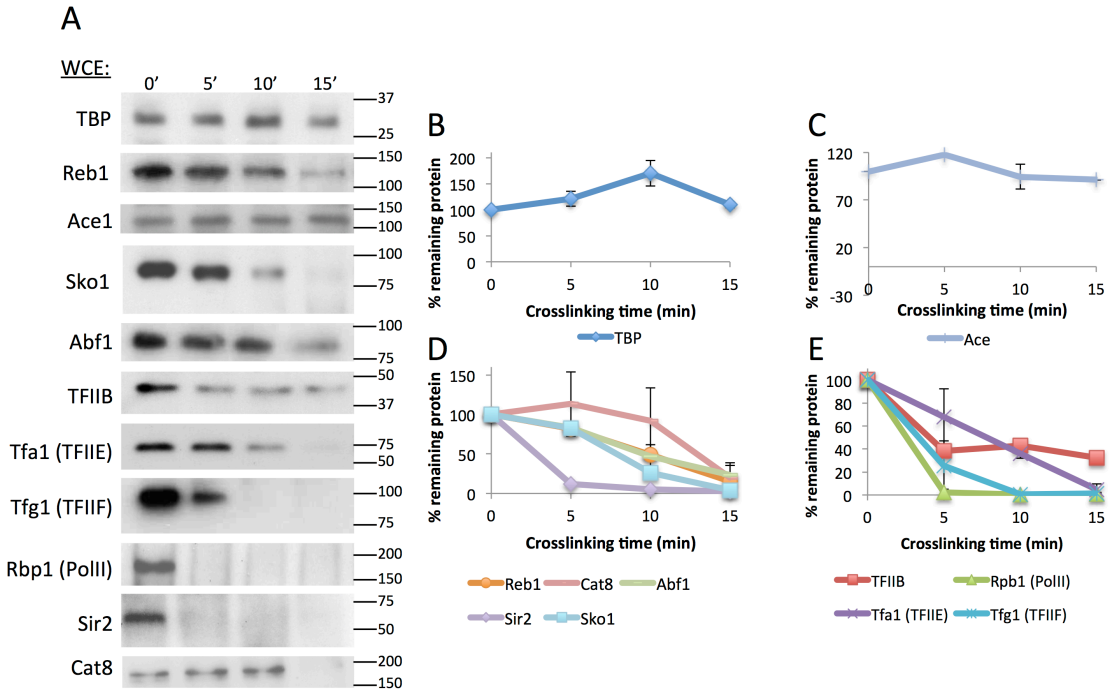


Figure 2.14. Protein levels in crosslinked whole cell extract samples over time. A) Western blots of whole cell extract (WCE) samples for all factors. Antibodies used are listed in Table 2.3 and molecular weight is denoted to the right in kDa. Samples were crosslinked with 5% formaldehyde for 0-15 minutes and quenched with excess glycine. B) Quantification of WCE western blot bands shown in A for TBP. Each sample was normalized to the 0 time point as a percentage. Two replicates were averaged for the plot and error bars represent standard deviation.

C) Same as B, except for Ace1. D) Same as B, except for transcription factors Reb1, Cat8, Abf1, Sir2, and Sko1. E) Same as B, except for preinitiation complex components TFIIB, Tfa1 (TFIIE), Tfg1 (TFIIE), and Rpb1 (RNA polymerase II). Independently performed Western blots using chromatin rather than WCE samples showed the same trends.

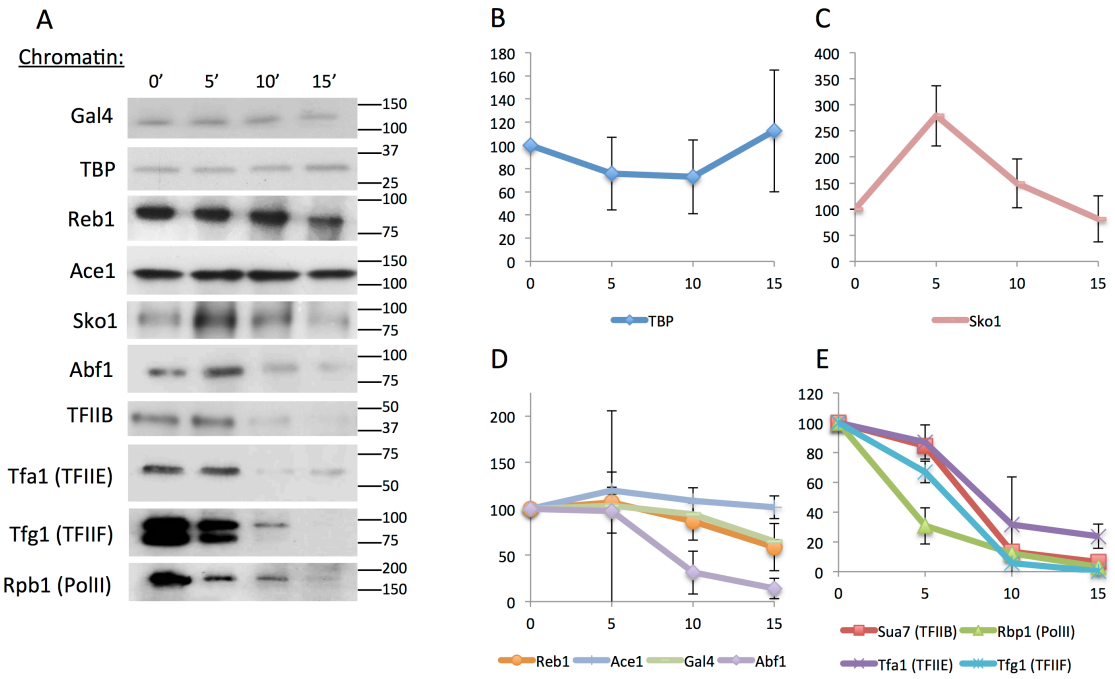


Figure 2.15. Protein levels in crosslinked chromatin samples over time. A) Western blots of chromatin samples for all factors except Sir2 and Cat8. Antibodies used are listed in Table 2.3 and molecular weight is denoted to the right in kDa. Samples were crosslinked with 5% formaldehyde for 0-15 minutes and quenched with excess glycine. B) Quantification of chromatin western blot bands shown in A for TBP. Each sample was normalized to the 0 time point as a percentage. Two replicates were averaged for the plot and error bars represent standard deviation. C) Same as B, except for Sko1. D) Same as B, except for transcription factors Reb1, Ace1, Gal4 and Abf1, Sir2. E) Same as B, except for preinitiation complex components TFIIB, Tfa1 (TFIIE), Tfg1 (TFIIE), and Rpb1 (RNA polymerase II). Data in figure A is from both Savera Shetty and Elizabeth Hoffman.

Model fits with updated conditions

To measure dynamics using the CLKv2 method, ChIP data for an interaction of interest are acquired in two different strains, each of which differ only in the concentration of the TF. One strain ("WT") expresses the TF of interest at wild type levels and the other ("OE" for "over-expression") typically harbors an additional copy of the TF gene which increases the TF concentration ~2-3-fold on average (see Table 2.5 for overexpression values). The CLK model contains as variables the on-rate for TF-chromatin binding (k_a), the off rate (k_d), and the formaldehyde crosslinking rate (k_{xl}); the fractional occupancy (θ_b) and residence time ($t_{1/2}$) are calculated from the variables and are not direct outputs of the fits. The saturation level of the ChIP signal (S_{sat}) is an additional parameter obtained from the fits, which we use to normalize the ChIP signal and, thereby, obtain an estimate of the fraction of crosslinked TF at a given site (θ_{xl}). The concentration of the TF in the nucleus (C_{TF}) and the formaldehyde concentration (C_{FH}) are experimentally measured quantities used in the fitting calculations (For reference, all variables are defined at the beginning of this dissertation under the List of Abbreviations section). The CLK model makes no assumptions about the relative rates of chromatin binding or crosslinking, and indeed it provides a framework sufficiently flexible to model a wide range of chemical and dynamic behavior (15, 16). Using the CLKv2 conditions, and as discussed in detail below, a wide range of behaviors were observed, including interactions with binding dynamics slower than crosslinking, comparable to crosslinking, or faster than crosslinking. In the binding dynamics-limited scenario (Fig. 2.16A, D), crosslinking is much faster than the on- and off-rates for chromatin binding. The hallmarks of the binding-dynamics limited behavior (referred to as "TF-limited") include two exponentials: a very steep exponential rise at short time scales (seconds), often manifesting as a non-zero y-intercept in the WT and OE data with a clear separation in the WT and OE y-intercepts, followed by a slower exponential rise. This clear separation in time scales makes it possible to extract binding dynamics, including the on- and off-rate (15). In

contrast, if the rate of crosslinking is slower than the time scale of TF binding dynamics, crosslinking-limited (referred to as "XL-limited") data show a single exponential rise with a zero y-intercept for the WT and OE data (Fig. 2.16B). The simulation in Fig. 2.16B and schematic in Fig. 2.16E show that for XL-limited interactions, the crosslinking time scale is slower than for the TF-limited case, and under these conditions TF binding and unbinding can occur prior to crosslinking. If the crosslinking rate is so slow (Fig. 2.16F) that its associated time is longer than the latest crosslinking time (usually 1200 seconds for this study), the ChIP signal rises linearly (or nearly linearly) as shown in Fig. 2.16C. In the linear version of the XL-limited model the theoretical curve shows a near-zero y-intercept, and no sign of saturation on the experimentally accessible time scale.

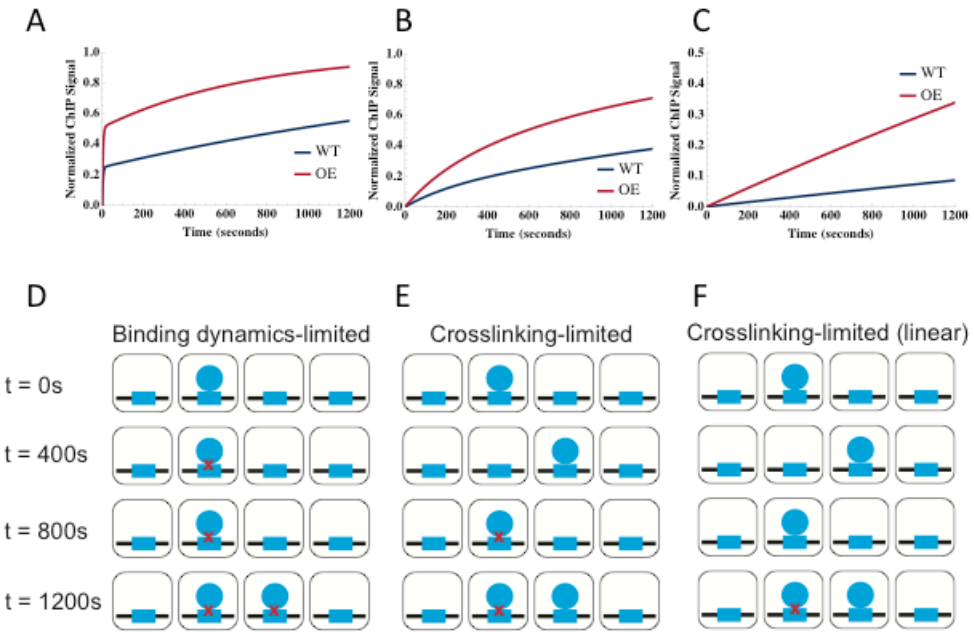


Figure 2.16. Overview of CLKv2 possible model fits. A-C) Simulations of CLKv2 fits. For each plot, blue represents the wild type strain and red is the overexpression strain. From left to right, the fits correspond to: binding dynamics (TF)-limited (A), crosslinking-limited (B), and linear crosslinking-limited behavior (C). D-F) Schematic of binding dynamics for each of the three CLKv2 cases with formaldehyde crosslinking over time: TF-limited (D), XL-limited (E),

and linear XL-limited (F). In each square cell, the TF (blue circles) binds to its binding site (blue rectangles); red x's represent crosslinking by formaldehyde. Crosslinking time increases as the panels progress from top to bottom. The simulations in A-C were made by Hussain Zaidi and D-F by David Auble.

Once crosslinking time-dependent data have been acquired, determining which scenario describes the data and fitting to the model is described in the flow chart in Fig. 2.7. The fitting procedures themselves are described in detail in the Methods section. We note that different sets of parameters are gained from each type of fit as shown in the schematic: TF-limited fits yield k_a, k_d, k_{xl}, and S_{sat} from which the dissociation constant, K_d, θ_b, and $t_{1/2}$ can be derived. However, the XL-limited fit only gives K_d, k_{xl} and S_{sat} from which θ_b can be derived, and the linear model provides K_d and $k_{xl}*S_{sat}$ from which θ_b can be derived.

Data were obtained for a number of TF-chromatin interactions using CLKv2. The interactions of TBP with the *LOS1*, *ACT1* and *URA1* promoters are shown in Fig. 2.17A-C. Applying the flow chart shown in Fig. 2.8 revealed that these interactions were well described by the TF-limited behavior (See *Computational Modeling* section under **Materials & Methods** for a definition of TF-limited behavior and the full model). At the *URA1* promoter, TBP-myc displayed both a linear ChIP signal with crosslinking time and sensitivity to formaldehyde concentration consistent with XL-limited dynamics, suggesting that although myc-tagged TBP complements growth, the myc tag had a relatively strong effect on crosslinking and possibly TBP binding dynamics. The data describing the interaction between LacI-GFP and an array of LacI sites is shown in Fig. 2.17D and was also well described by TF-limited behavior. The fractional occupancies of the three TBP loci ranged from 0.04-0.07, while the residence times were about 60-90 seconds (Table 2.6). Consistent with prior work (Poorey et al. 2013), this indicates that

these promoters are unoccupied by TBP most of the time, and that the TBP complexes that do form are not very long-lived. LacI fractional occupancy was lower still, but the complexes formed had half-lives of 1056 seconds (Table 2.8). This long lifetime is consistent with both prior CLK and live imaging data (Poorey et al. 2013). TBP binding to *NTS2* (the promoter for Pol I transcription) and Ace1 binding to *CUP1* were both best approximated by the linear model (Fig. 2.17E, F; Tables 2.7 & 2.8). The linear behavior of Ace1 CLK data using the CLKv2 conditions is consistent with rapid binding dynamics (Karpova et al. 2004, Poorey et al. 2013) being faster than the crosslinking rate. The high fractional occupancy of Ace1 at *CUP1* (0.83) is also consistent with prior observations (Karpova et al. 2004, Poorey et al. 2013). The fractional occupancy of TBP at *NTS2* (0.73) was much higher than TBP occupancies at the other promoters, consistent with the high transcriptional activity of the rDNA in cells in log phase growth in rich medium (Dammann et al. 1993). Tables 2.6 and 2.7 provide all the measured kinetic parameters along with their associated errors. Notably, error analysis derived from multiple fits of simulated data (see Materials and Methods) showed that most parameters were associated with a single well-defined distribution (Figs. 2.20 and 2.21).

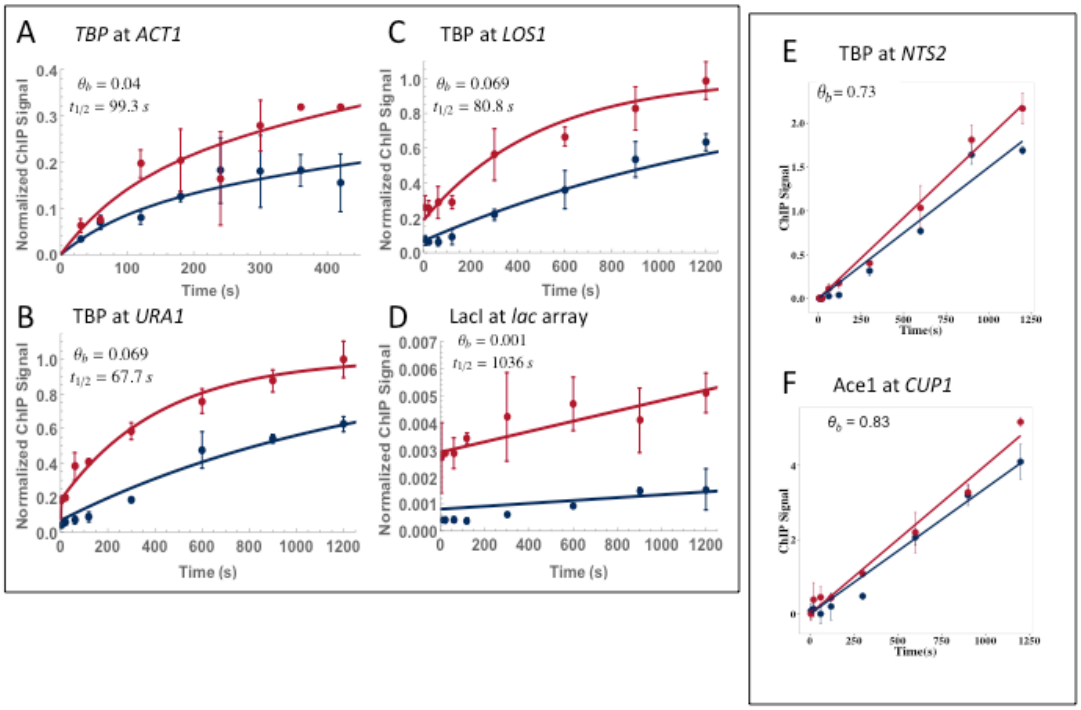


Figure 2.17. CLKv2 fits of data for TBP, LacI, and Ace1. A-D) All fits shown are TF-limited full model fits. The blue line is the wild type strain, while red has the factor overexpressed; overexpression factors are listed in Table 2.5. TBP is shown at *ACT1* (A), *URA1* (B), and *LOS1*(C); LacI is shown at a *lac* array (D). Occupancy (θ_b) and $t_{1/2}$ are denoted on the plots. E-F) Both fits are linear crosslink-limited. TBP is shown at *NTS2* (E) and AceI is shown at *CUP1* (F). Only occupancy (θ_b) is shown since residence time is not extracted with this fit. Each dataset resulted from two biological replicates and the average is shown. Error bars represent the standard deviation. Data in all panels was collected by Savera Shetty and fit by Hussain Zaidi and Stefan Bekiranov.

Datasets obtained for TBP binding to the *HSC82* and *SNR6* promoters were not obviously linear or non-linear; these ambiguous cases required a more rigorous selection process for the best fit (see flowchart, Fig. 2.18A, and Methods section for detailed explanation). These datasets were

fit with both the TF-limited and linear models and the sum of squared residuals (SSR) derived from the fits were compared for the appropriate fit (Fig. 2.18B, C). Both loci had a better fit with the linear model; the SSR for the TF-limited/linear models for *HSC82* and *SNR6* were 0.11/0.042 and 3.35/0.43, respectively. The occupancy of TBP at *HSC82* and *SNR6* was 0.57 and 0.73, respectively.

As mentioned earlier, it is possible to model kinetic behavior of TFs that are depleted by formaldehyde by focusing measurements on the formaldehyde incubation time period where levels remain stable. TFIIE was significantly depleted by about ten minutes (Fig. 2.14A, E), but the protein levels were not detectably changed through seven minutes of formaldehyde incubation (Fig. 2.19A). This allowed us to measure TFIIE interaction with the *ACT1*, *LOS1*, and *URA1* promoters (Fig. 2.19B-D, Table 2.9). TFIIE binding to *URA1* and *LOS1* was best described by a crosslinking-limited model, whereas a full model fit described binding to *ACT1*. Fractional occupancies were well below saturation for all three sites, and at *ACT1* we compute a residence time of about 6 minutes, which is within the bounds of an estimate of its 95% confidence interval (Figure 2.11) and on par with the time-scale for TBP interaction at this site.

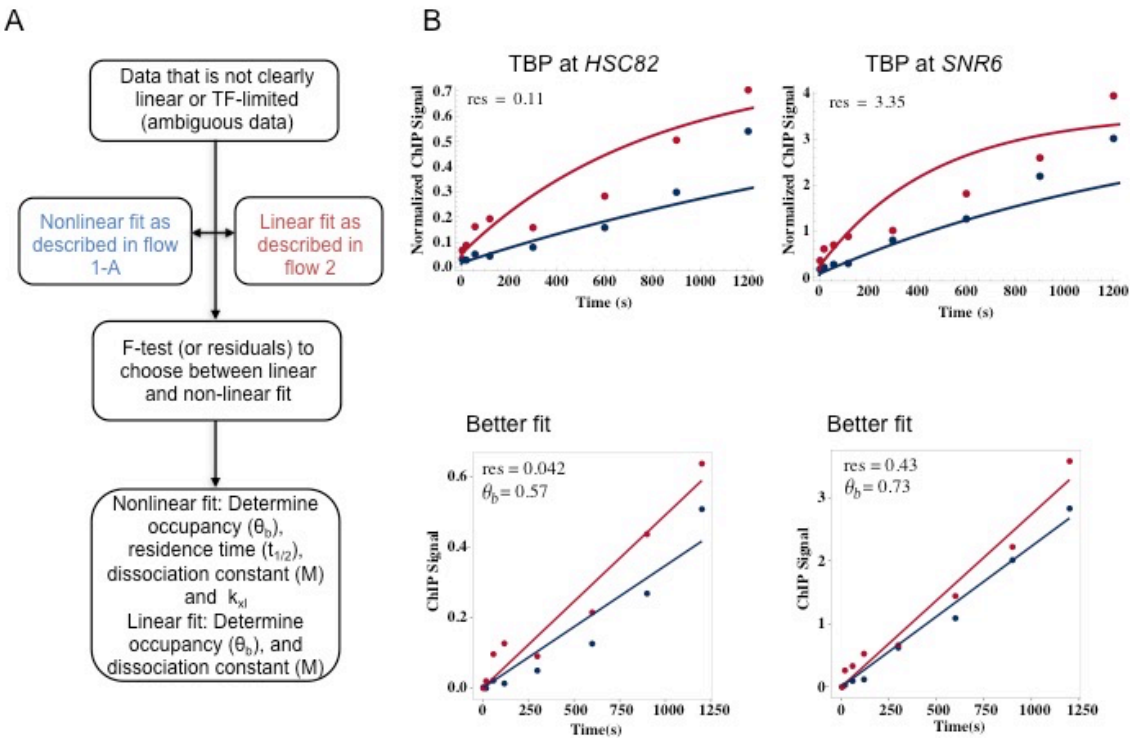


Figure 2.18. Resolution of ambiguous TBP fits. A) Flow chart to determine best fit for ambiguous data. Data was fit with both linear and full models and F-tests or sum of squared residuals (SSR) was then used to differentiate the best fit. B-C) TBP fits at *HSC82* (B) and *SNR6* (C) were fit with both full (top) and linear (bottom) models. SSR derived from the fits was used to find that both datasets were best represented with the linear fit; SSR is shown on all four plots and occupancy (θ_b) for the linear fit. Each dataset resulted from two biological replicates and the average is shown. Error bars represent the standard deviation. Data in panels B and C was collected by Savera Shetty and fit by Hussain Zaidi and Stefan Bekiranov.

Table 2.6. Measurements for TBP-binding dynamics at select promoters with a full model fit.

TBP-Full model	k_{xl} (1/mol s)[1]	τ_{xl} (s)[2]	k_a*C_{TF} (1/s)[3]	k_d (1/s)[4]	S_{sat}[5]	K (mol)[6]	$t_{1/2}$ (s)[7]	θ_b[8]
ACT1	0.14 (+31.2, -0.51)	2.76 (+45.05, -0.73)	3.21 (+1.3, -0.92) E-04	6.98 (+3.3, -2.2) E-03	1.35 (+0.35, -0.28)	2.61 (+1.0, -0.74) E-04	99.3 (+99.32, -56.51)	0.044 (+0.44, -0.016)
LOS1	N/A	N/A	6.31(+1.1, -0.93) E-04	8.58 (+2.3, -1.8) E-03	0.44 (+0.031, -0.029)	1.63 (+0.32, -0.27) E-04	80.8 (+21.45, -16.94)	0.069 (+0.012,0.01)
URA1	0.30 (+1129.02, -4.74)	1.29 (+19.22, -0.081)	7.59 (+1.0, -0.91)E-04	1.0 (+0.29, -0.23) E-02	0.71 (+0.036, -0.035)	1.62 (+0.36, -0.29) E-04	67.7 (+21.02, -16.2)	0.069 (+0.015, -0.01)

[1]Formaldehyde crosslinking rate
[2]Crosslinking time
[3]On-rate of transcription factor X nuclear concentration of factor
[4]Off-rate of transcription factor
[5]ChIP signal at saturation
[6]Dissociation constant, k_d/k_a
[7]Residence time of TF binding
[8]Occupancy
N/A: not applicable

Table 2.7. Measurements for TBP-binding dynamics at select promoters with a linear model fit.

TBP-Linear	$k_{xl}*S_{sat}$ (1/mol s)[1]	τ_{xl} (s)[2]	k_a*C_{TF} (1/s)[3]	k_d (1/s)[4]	S_{sat}[5]	K (mol)[6]	$t_{1/2}$ (s)[7]	θ_b[8]
NTS2	1.1 (+0.084, -0.078) E-03	N/A	N/A	N/A	N/A	4.51 (+2.21, -1.44) E-06	N/A	0.73 (+0.077, -0.07)
SNR6	1.7 (+0.12, -0.11) E-03	N/A	NA	NA	NA	4.46 (+2.0, -1.37) E-06	NA	0.73 (+0.074, -0.067)
HSC82	3.4 (+0.54, -0.47) E-04	N/A	NA	NA	NA	8.96 (+5,73, -3.51) E-06	NA	0.57 (+0.13, -0.1)

[1]Formaldehyde crosslinking rate
[2]Crosslinking time
[3]On-rate of transcription factor X nuclear concentration of factor
[4]Off-rate of transcription factor
[5]ChIP signal at saturation
[6]Dissociation constant, k_d/k_a
[7]Residence time of TF binding
[8]Occupancy
N/A: not applicable

Table 2.8. Measurements for Ace1 and LacI binding dynamics.

	k_{xl} (1/mol s)[1]	τ_{xl} (s)[2]	k_a*C_{TF} (1/s)[3]	k_d (1/s)[4]	S_{sat}[5]	K (mol)[6]	$t_{1/2}$ (s)[7]	θ_b[8]
Ace1 (linear) @ CUP1	2.28 (+0.1, -0.098) E-03	N/A	NA	NA	NA	2.07 (+1.14, -0.77) E-07	NA	0.83 (+0.055, -0.052)
Lac1 (full model) @ LacO	3323 (+5.26E09, -767979)	1.2E-04 (+0.0034, -5.0 E-07)	5.33 (+5.09, -2.37) E-07	6.57 (+1.51, -1.22)E-04	906.9 (+893.7, -473.1)	1.23 (+1.75, -0.79) E-03	1055.5 (+254.6, -206.1)	8.10 (+8.3, -3.8) E-04

[1]Formaldehyde crosslinking rate
[2]Crosslinking time
[3]On-rate of transcription factor X nuclear concentration of factor
[4]Off-rate of transcription factor
[5]ChIP signal at saturation
[6]Dissociation constant, k_d/k_a
[7]Residence time of TF binding
[8]Occupancy
N/A: not applicable

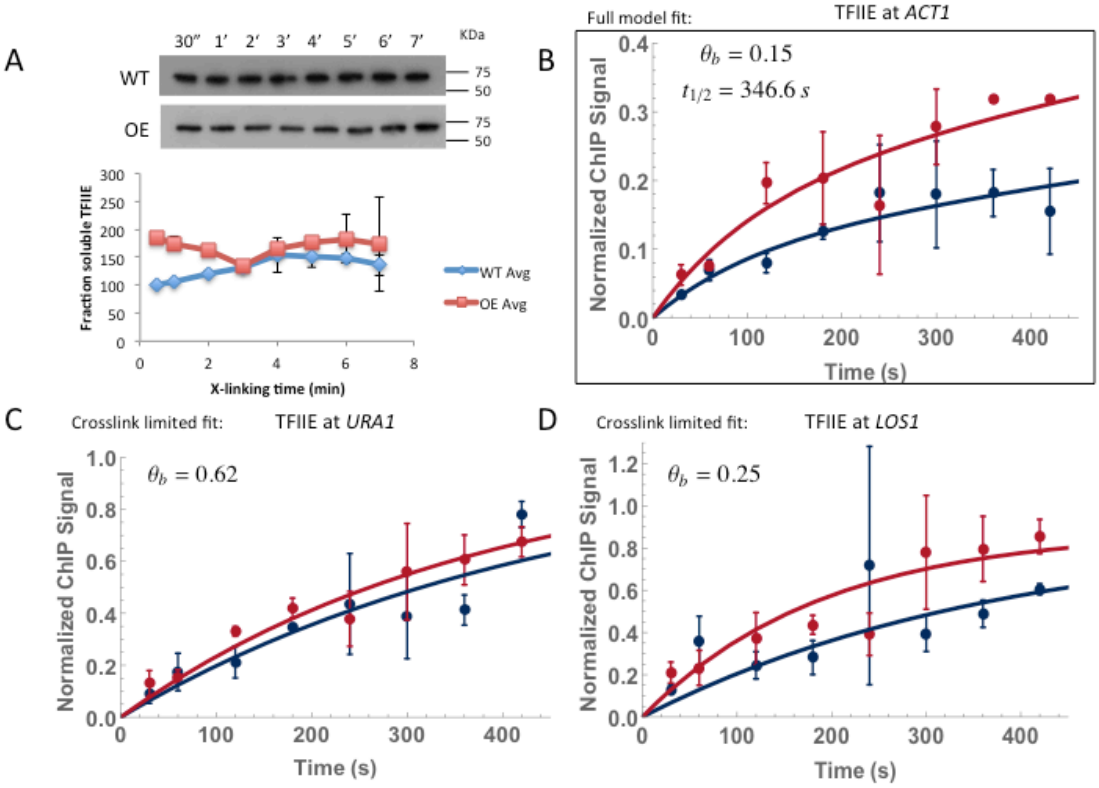


Figure 2.19. CLKv2 for TFIIE on a shorter experimental time scale. A) Western blot of Tfa1-TAP chromatin using an anti-TAP antibody. Samples were crosslinked for 30 seconds to seven minutes. Wild type and overexpression strains were both tested for depletion. Quantification of the signal was plotted below; two replicates were averaged and the standard deviation is shown as error bars. The wild type strain was normalized to its 30 second time point; the overexpression

strain was normalized to its 30 second time point and multiplied by the overexpression factor (Table 2.5). The overexpression factor was determined by running four 5% formaldehyde-crosslinked time points for the wild type and overexpression strains on the same gel and blotting for TAP tag (data not shown). Bands were quantified with ImageJ (NIH) and compared to determine overexpression. B) TFIIE at *ACT1* resulted in a full model fit; occupancy and residence time are denoted. C, D) TFIIE at both *URA1* and *LOS1* gave crosslink-limited fits; only the occupancy is shown. Each dataset resulted from two biological replicates and the average is shown. Error bars represent the standard deviation. The data in panels B-D was fit by Hussain Zaidi and Stefan Bekiranov.

Table 2.9. Measurements for TFIIE binding dynamics at select promoters.

TFIIE	**k_{xl} (1/mol s)[1]**	**τ_{xl} (s)[2]**	**k_a*C_{TF} (1/s)[3]**	**k_d (1/s)[4]**	**S_{sat}[5]**	**K (mol)[6]**	**$t_{1/2}$ (s)[7]**	**θ_b[8]**
Full model fit:								
ACT1	4.80E-03 (+1.3, -0.22)	80.2 (+992.3, -16.85)	3.5 (+9.4, -2.7) E-04	2.0 (+7.2, -2.2) E-03	3.11 (+3.21, -2.94)	7.80 (+14, -6.2) E-05	346.6 (+523.9, -156.8)	0.149 (+0.48, -0.01)
XL-limited fit:								
URA1	1.96 (+1.0, -0.69) E-03	206.5 (+76.52, -52.94)	N/A	N/A	2.71 (+2.71, -0.48)	7.17 (+18.0, -5.5) E-06	N/A	0.66 (+0.15, -0.12)
LOS1	4.3 (+5.4, -2.5) E-03	90.1 (+95.4, -44.2)	N/A	N/A	0.294 (+0.08, -0.05)	4.14 (+2.4, -1.0) E-05	N/A	0.31 (+0.14, -0.1)

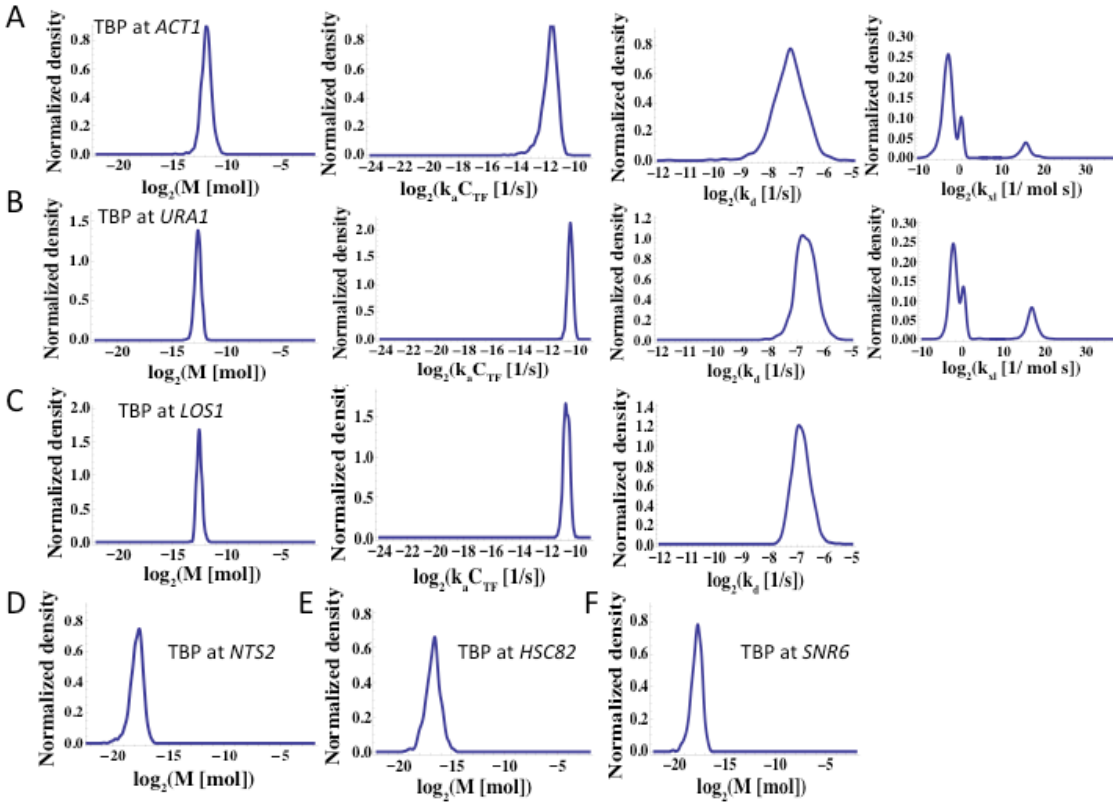


Figure 2.20. Error distribution generated from simulated data for TBP. A) Error distributions for TBP at *ACT1*. Plots are from left to right: M (this is K_d), k_a*C_{TF}, k_d, and k_{xl}. B) Error distributions for TBP at *URA1*. Plot order is identical to A. C) Error distributions for TBP at *LOS1*. Plots are in the same order as A, but analysis for k_{xl} was not done. D) Error distributions for TBP at *NTS2* for M. E) Error distributions for TBP at *HSC82* for M. F) Error distributions for TBP at *SNR6* for M. All plots were generated by Hussain Zaidi.

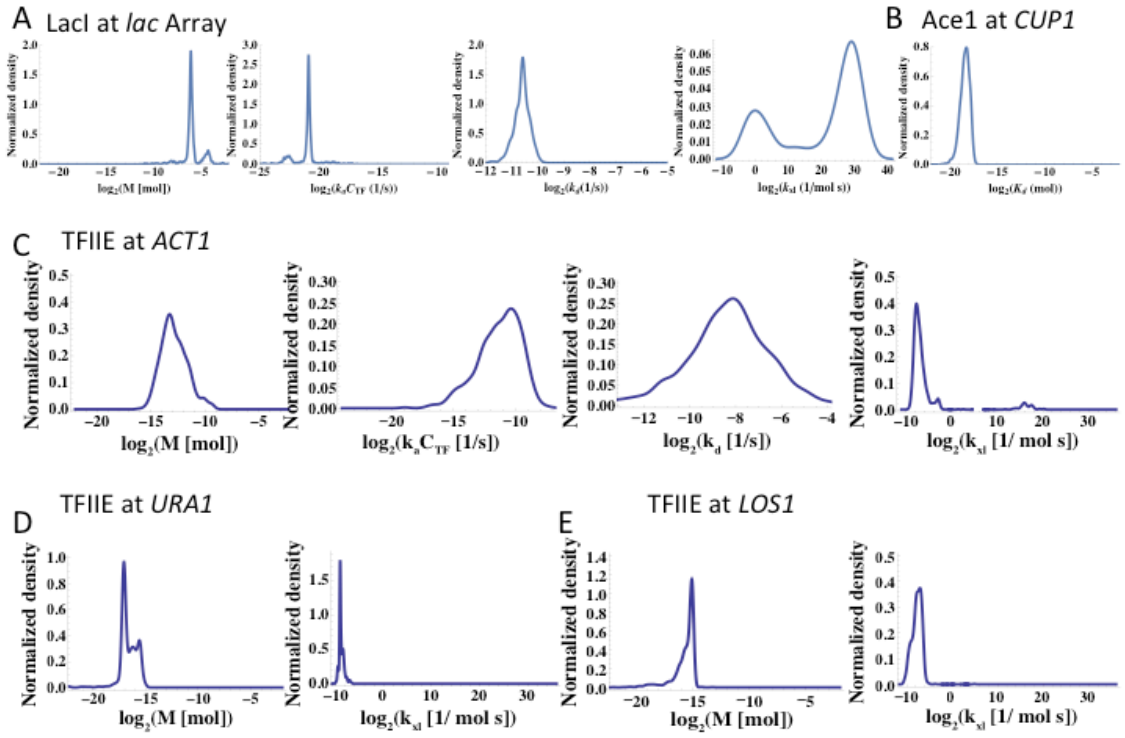


Figure 2.21. Error distribution generated from simulated data for all other CLKv2 factors. A) Error distributions for lacI at the *lac* array. Plots are from left to right: M (this is K_d), k_a*C_{TF}, k_d, and k_{xl}. B) Error distributions for Ace1 at *CUP1* for M. C) Error distributions for TFIIE at *ACT1*. Plot order is identical to A. D) Error distributions for TFIIE at *URA1*. Plots for M and k_{xl} are shown. E) Error distributions for TFIIE at *LOS1*. Plots for M and k_{xl} are shown. All plots were generated by Hussain Zaidi.

Table 2.10. Statistical analysis of TBP CLK model fits at select loci.

Full model						
Protein	**Locus**	**Parameter p-value:**				**Adjusted R^2**
		S	**k_a*C_{TF}**	**k_d**	**k_{xl}**	
TBP	*URA1*	6.04E-10	2.05E-05	8.8E-03	0.58	0.99
TBP	*ACT1*	3.38E-04	0.0170	0.034	0.44	0.98
TBP	*LOS1*	1.52E-08	1.38E-04	3.1E-03	<1E-07	0.99
TBP	*SNR6*	3.15E-04	0.87	0.91	1.00	0.90
TBP	*HSC82*	0.017	0.82	0.87	1.00	0.89
Linear						
Protein	**Locus**	**Wild Type Beta***	**Wild Type Adjusted R^2**	**Over-expression Beta***	**Over-expression Adjusted R^2**	
TBP	*SNR6*	1.45E-08	0.99	1.37E-07	0.98	
	HSC82	1.53E-05	0.93	1.58E-06	0.96	
	NTS2	7.30E-07	0.97	5.08E-09	0.99	

*Beta = θ_b*k_{xl}*S_{sat}*C_{FH}

Table 2.11. Statistical analysis of lacI and Ace1 model fits at select loci.

Full model										
Protein	**Locus**	**Parameter p-value:**				**Confidence interval:**				**Adjusted R^2**
		S	**k_a*C_{TF}**	**k_d**	**k_{xl}**	**S**	**k_a*C_{TF} (1/s)**	**k_d (1/s)**	**k_{xl} (1/M s)**	
LacI	*lac* array	0.99	0.99	1.93E-03	<1E-07	0.0, 163.50	0.0, 8.01E-05	2.94E-04, 1.02E-03	1460.3, 1460.3	0.97
Linear										
Protein	**Locus**	**Wild Type Beta***	**Wild Type Adjusted R^2**	**Over-expression Beta***	**Over-expression Adjusted R^2**					
Ace1	*CUP1*	7.24 E-08	0.99	2.85E-08	0.99					

*Beta = θ_b*k_{xl}*S_{sat}*C_{FH}

Table 2.12. Statistical analysis of TFIIE CLK model fits at select loci.

Full model										
Protein	**Locus**	**Parameter p-value:**				**Confidence interval:**				**Adjusted R^2**
		S	**k_a*C_{TF}**	**k_d**	**k_{xl}**	**S**	**k_a*C_{TF} (1/s)**	**k_d (1/s)**	**k_{xl} (1/M s)**	
TFIIE	**ACT1**	0.19	0.70	0.71	0.36	0.0, 1.83	0.0, 0.0031	0.0, 0.012	0.0, 0.014	0.97

XL-limited								
Protein	**Locus**	**Parameter p-value:**			**Confidence interval:**			**Adjusted R^2**
		S	**K_d**	**k_{xl}**	**S**	**K_d (1/M)**	**k_{xl} (1/M s)**	
TFIIE	**URA1**	0.012	0.36	0.14	0.25, 1.74	0.0, 1.71	0.0, 0.0045	0.97
TFIIE	**LOS1**	0.016	0.52	0.44	0.22, 1.78	0.0, 9.28	0.0, 0.016	0.93

Discussion

The CLK assay was conceived to provide biophysically rigorous on and off rates for TF binding to single copy loci in vivo (Poorey et al. 2013). We sought to develop an approach that would also be generally applicable and potentially scalable to genome-wide analysis. The biggest obstacle to implementation of this assay has been to develop general experimental conditions and a companion model that accurately account for the many effects occurring in cells that undergo formaldehyde crosslinking and to distinguish them from the contributions of binding kinetics to the time dependent change in ChIP signal. The new, updated version of the CLK assay extends our understanding of the effects of formaldehyde on yeast cells and uses our observations to both improve the CLK assay conditions and to improve the approach to data analysis. Formaldehyde crosslinking is ubiquitous in the chromatin field, so the results may contribute to the understanding and interpretation of ChIP and related types of experimental results in general as well.

Our results demonstrate improvement in formaldehyde quenching using a higher concentration of glycine than was used previously. The residual unquenched formaldehyde that

remains following addition of 250 mM glycine as commonly used and in the original CLK procedure likely inflated the ChIP signal values at short crosslinking times as the unquenched formaldehyde continued to capture complexes during the centrifugation step that follows quenching. However, despite this, the relative differences in ChIP signal change with time apparent in the original CLK data do capture the relative differences in binding dynamics validated by other methods. For example, the rapid rise in Ace1 ChIP signal with short crosslinking times observed originally is consistent with the known highly dynamic behavior of Ace1 binding to its sites in the *CUP1* promoter (Karpova et al. 2008), whereas the shallow slope and gradual approach to saturation seen with LacI time-dependent ChIP signals are consistent with its long residence time (Robinett 1996), which we confirmed by live cell imaging (Poorey et al. 2013). Remarkably, the residence times for TBP binding to particular promoters reported here are also broadly consistent with the residence times obtained with the original version of the CLK assay (Poorey et al. 2013). The results argue that TBP has residence times at these promoters on the order of one to several minute time scale. Thus, although the original CLK data was modeled assuming infinitely fast quenching, we nonetheless captured the relative time scale of dynamic behavior as validated by both live cell imaging and in this study using CLKv2.

Based on the results presented here, although Tris is highly effective in quenching unreacted formaldehyde, it is unsuited for use in this type of kinetic analysis due to its ability to reverse crosslinks. The crosslink reversal that we observed is consistent with a prior report (Kawashima et al. 2014) and is exacerbated by the relatively high concentration of Tris required to completely react with a relatively high concentration of added formaldehyde. We also show that time-dependent increases in ChIP signal can be affected by the concentration of formaldehyde. The use of a formaldehyde concentration that is as high as possible boosts the crosslinking rate, thereby extending the useful range of the assay. Although we employed 5% formaldehyde here, this may not be advisable or appropriate for analysis of other TFs or in other types of cells. The best formaldehyde concentration ought to be determined empirically by

choosing the concentration that yields the best separation between the crosslinking and binding dynamics time scales, and which does not impact overall recovery of soluble components or deplete the unbound TF in the soluble pool over the kinetic time course. Those factors that are stable constituents of multi-subunit complexes such as Rpb1 may be impossible to assess using this approach; what is observed by western blotting as their rapid depletion from extracts may be due to rapid crosslinking to other biologically relevant polypeptides with which they stoichiometrically co-associate.

Using the CLKv2 method, we find that crosslinking rates are highly variable and depend on the particular TF-DNA site of interaction (Tables 2.6-2.9). Prior to our measurement of formaldehyde crosslinking rates in vivo, crosslinking of ChIP complexes was generally thought to be rapid (Solomon and Varshavsky 1985, Solomon et al. 1988, Aparicio et al. 2005, Lu et al. 2010), and this was supported qualitatively by the differences in ChIP signals that were observed at closely spaced time points (Hoffman et al. 2015) and that highly transient interactions (residence times on the ~second scale) could nonetheless be captured by formaldehyde crosslinking in ChIP experiments (Hager et al. 2009, Poorey et al. 2013, Chen et al. 2014, Viswanathan et al. 2014). In addition, there is a global correlation between steady-state ChIP signals and in vitro binding affinity (Toth and Biggin 2000, Kaplan et al. 2011) consistent with the overall ChIP signal level not being merely proportional to the rate of capture by crosslinking. In vitro, the rate of formaldehyde reaction with DNA bases is relatively slow (McGhee and Von Hippel 1975), but reactivity could be greatly accelerated when DNA and amino acids were present together (Siomin et al. 1973). Interestingly, the rates of TBP crosslinking to the *URA1* and *ACT1* promoters calculated by CLKv2 (k_{xl}, Table 2.6) are in the same range as in vitro crosslinking rates obtained in reactions containing DNA and amino acids (Siomin et al. 1973). Experiments measuring formaldehyde reactivity with amino acids and proteins have shown that formaldehyde adducts tend to be mainly formed with cysteine, lysine, and tryptophan side chains as well as the N-terminal group of polypeptide chains (Metz et al. 2004, Toews et al. 2008). In

reactions containing both nucleic acids and protein/amino acid substrates, the most efficient crosslinking was found to occur between lysine and deoxyguanosine (Solomon and Varshavsky 1985, Roeder 1996, Poorey et al. 2013). We suggest that the wide range in crosslinking rates reported here reflects the variation in reactive chemical groups on the TF surface and their proximity and orientation to reactive groups on DNA bases at or near binding sites.

Although some factors of interest were eventually depleted from extracts following formaldehyde treatment, our results with the TFIIE subunit Tfa1 show that it is still possible to investigate them kinetically if the crosslinking time course is confined to a temporal window in which their overall levels are not affected by formaldehyde. A possible limitation in this approach is that a shorter time course may make it more difficult to determine the saturation level of the ChIP signal, an estimate for which is required for confident fitting of the data and accurate estimates of the parameters. An alternative approach for future work is to extend the current model to include the depletion of the TF of interest in the fitting. Conceptually, by quantifying the rate of TF depletion from Westerns such as those shown in Figs. 2.14 and 2.15, the decrease in the overall level of the TF with crosslinking time could be modeled and the level of the TF at different times included explicitly as a parameter during the analysis of the data.

In instances in which the crosslinking rate is much slower than TF-DNA binding, CLKv2 yields the fractional occupancy as well as the equilibrium binding constant. Although the residence time cannot be estimated from the data in these situations, the fractional occupancy and binding constant are useful parameters as they provide insight into the variation in site occupancy across the cell population, which could have implications for understanding the molecular basis of transcriptional noise (Stewart-Ornstein et al. 2012, Ravarani et al. 2016), as well as energetic barriers in the intracellular environment that reduce binding from in vitro values obtained using purified components. If the crosslinking rate can be determined, this can be used to set an upper limit for binding dynamics. For many biological systems, knowing whether binding is occurring

faster or equal to the second, minute or tens of minutes time scale would be valuable for developing dynamic models for the order of events underlying transcriptional responses.

Implications and Future Directions

One of the most important take-ways from the updated CLK method is the breadth of formaldehyde crosslinking times calculated *in vivo*, depending on the TF and locus. Given the pivotal role of formaldehyde in understanding chromatin biology and the continuing evolution of technologies exploiting its properties, it is critical that a deeper understanding is achieved of formaldehyde crosslinking as it occurs in cells. Although formaldehyde can mediate countless chemical reactions *in vitro*, the conditions used for crosslinking in cells suggest that, with respect to chromatin, the chemical complexity of macromolecule-containing reaction products is more limited, with reactions occurring mainly with solvent-exposed lysine residues and endo- and exocyclic amino groups on bases. Formaldehyde has a number of other properties that make it well suited for trapping macromolecular complexes in cells, including cell permeability and the temperature-dependent stability of methylene bridge containing adducts. Macromolecules that do not interact are in general not crosslinked together efficiently, and methylol/Schiff base intermediates are reversibly formed and appear to be inefficiently trapped by reaction with quenchers in cells. This explains why proteins and DNA isolated from formaldehyde-treated cells appear unmodified in general (Solomon and Varshavsky 1985). Within minutes of formaldehyde incubation, there is very little detectable free DNA (<10%) (Wu et al. 2011, Kennedy-Darling and Smith 2014), and crosslinking appears to occur uniformly along DNA as well (Solomon and Varshavsky 1985).

ChIP has been developed by empirical determination of seemingly optimal crosslinking conditions, with low recoveries occurring for either too little or too much crosslinking (Orlando 2000). If the formaldehyde concentration is too low or the incubation time is too short, not

enough crosslinked material will be produced. On the other hand, too high a concentration of formaldehyde or too long of incubation time also reduces recovery, presumably reflecting the formation of complexes that are insoluble or the masking of epitopes recognized by the antibody used for immunoprecipitation. We have observed little effect of formaldehyde incubation time on chromatin protein yield over a broad range of formaldehyde concentrations and incubation times (Viswanathan et al. 2014), but formaldehyde concentrations above 3% do impact yield even after moderate incubation times (Zaidi et al. 2017), suggesting the formation of such complexes. Given the dense concentration of macromolecules in the nucleus, it is plausible that formaldehyde may cause the formation of higher order networks of crosslinked chromatin (Gavrilov et al. 2015), as illustrated in Figure 2.22. This is an area worthy of additional investigation, particularly since it may explain nonspecific DNA crosslinking that occurs in ChIP or ChIP-related methods (Fan et al. 2008). Additionally, other artificial enrichment phenomena, such as localization of unrelated proteins, can occur with ChIP at highly expressed genes; these, too, warrant deeper investigation (Fan and Struhl 2009, Teytelman et al. 2013) to ensure that apparent ChIP signals in fact represent true association with the loci of interest. Since not all promoters are crosslinked quickly with formaldehyde, another option moving forward with the CLK method is to explore alternative crosslinkers, such as glyoxal (Richter et al. 2018). While the proportion of promoters genome wide that can be fit with the full model is unknown, a different crosslinker could potentially increase this proportion and therefore extract more kinetic parameters at the most sites possible. Glyoxal has been shown to crosslink faster than paraformaldehyde at a pH of 4-5 and can be quenched with ammonium chloride (Richter et al. 2018). This could potentially be used at a lower concentration than 5% formaldehyde to avoid formation of complexes of crosslinked chromatin.

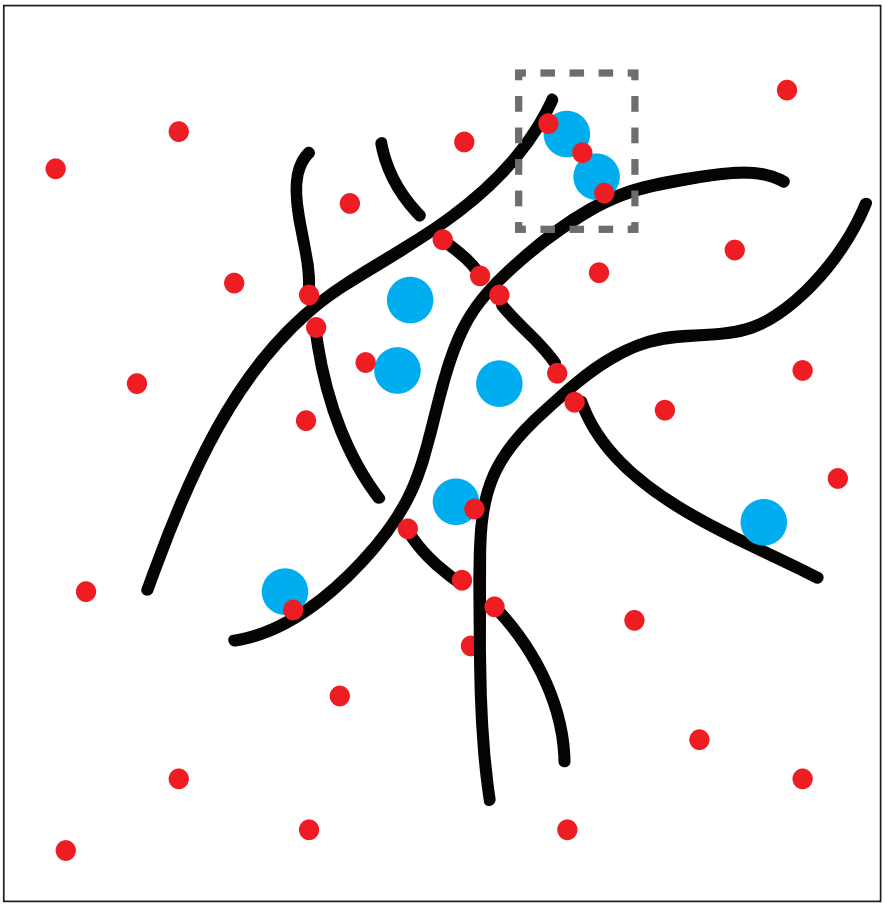


Figure 2.22. **Potential effects of formaldehyde in mediating formation of higher order chromatin structures.** The black wavy lines denote chromatin fibers, which may become a crosslinked meshwork in the presence of formaldehyde (red circles). The formation of these potentially confounding structures may or may not be mediated by physiologically relevant higher order interactions captured by crosslinking (dashed gray rectangle). Such a meshwork may define localized neighborhoods in the nucleus that trap proteins (cyan) that may or may not interact specifically with nearby DNA sequences in an unperturbed cell.

ChIP assays have been pivotal in establishing our current view of chromatin structure and function. As answers to deeper questions about chromatin binding dynamics and higher order structure are pursued, it is imperative that we understand more fully how procedures employed to

obtain snapshots of chromatin state may perturb the very properties being measured, particularly as variations in experimental design can yield different interpretations (Nacheva et al. 1989, Gavrilov et al. 2015). In this regard, an understanding of the behavior and properties of formaldehyde in the cell are important for determining the best methods for measuring dynamic interactions using crosslinking. A better understanding of formaldehyde crosslinking may in turn lead to better quantitative models (Voit et al. 2015). This will be especially important as we move forward with the CLK assay and analyze more TF components of the preinitiation complex, as well as subunits of the even larger SAGA and Mediator regulatory complexes. Understanding how crosslinking at higher concentrations influences our results is necessary to interpret the results of the amalgamated factors and compare the dynamics of assembling complexes in the larger context of transcription; this will be discussed further in the next chapters.

CHAPTER III

Complementary ChIP-based approaches

The molecular framework explaining the process of transcriptional factor recruitment and transcription initiation is based on an extensive body of work that has elucidated the structural organization of transcription complexes in exquisite detail and at ever increasing resolution along the chromatin template. Imaging studies have uncovered highly dynamic interactions between transcription factors and chromatin, many of which occur for mere seconds. A critical challenge is to determine how the static snapshots of transcription complexes defined biochemically relate to the highly dynamic and stochastic process as it occurs in living cells. While the transcription factor-chromatin binding dynamics were captured for several factors by the crosslinking kinetics (CLK) assay, it's important to verify these measurements with other independent methods. Two other such ChIP-based assays are Anchor Away (AA) and Competition ChIP (CC), which employ a nuclear depletion approach and an inducible system with different epitope tags, respectively. Our lab has used both of these approaches to measure binding dynamics for TBP at select promoters. For Anchor Away, the nuclear depletion rate needs to be measured to determine a bound for the fastest dynamics and as variable built into the modeling once it is developed. The nuclear depletion rate of TBP was calculated by imaging fixed cells treated with rapamycin for varying amounts of time; at 11 minutes, 50% of TBP is depleted from the nucleus. Using single exponential decay fits, the residence time for several loci was calculated and compared to CLK and CC TBP data. There is some agreement between the CC and CLK data at *ACT1, URA1,* and *LOS1*, but the Anchor Away data needs to be model fit before an accurate comparison can be made to come to a consensus for *in vivo* TBP binding dynamics. Additionally, a TFIIB CC strain was made for use in future experiments. By comparing the TBP kinetic measurements to other PIC components, as well as the chromatin landscape and features of each promoter site, the

relationship between PIC assembly kinetics and transcriptional initiation will be better understood.

Introduction

Anchor Away as a nuclear depletion technique to measure protein dissociation rates

Anchor Away (AA) is an approach that is frequently used for conditional nuclear protein depletion in *S. cerevisiae*, developed as an alternative approach to using temperature-sensitive mutant strains (Haruki et al. 2008). It is also a potentially valuable but underutilized molecular approach for measuring chromatin interaction dynamics. Nuclear depletion of a tagged protein is induced by "anchoring" it in the cytoplasm via association with a ribosomal protein; this process is induced by rapamycin-dependent heterodimerization of the human FKBP12 to the human mTOR FRB domain fused to the anchor and target proteins, respectively (Figure 3.1) (Chen et al. 1995, Belshaw et al. 1996, Haruki et al. 2008). Additionally, the *TOR1* allele in the yeast strain is mutated to *tor1-1* and the rapamycin-binding gene *FPR1* is deleted (*Δfpr1*) to avoid rapamycin toxicity (Haruki et al. 2008). To a first approximation, the decrease in ChIP signal over time that results from rapamycin-induced nuclear depletion is a measure of the off-rate of the factor for a particular site if the export rate is fast and dissociation of the factor from chromatin is rate-limiting. The AA technique has been used as a complementary approach to CC for measuring TBP binding dissociation, thereby assigning relative timescales for TBP unbinding at different classes of promoters (Grimaldi et al. 2014). Consistent with CC data, TBP binding to several Pol II promoters was found to be much more dynamic than TBP associated with the rDNA, which was markedly more stable (van Werven et al. 2009, Grimaldi et al. 2014, Zaidi et al. 2017). AA has a similar temporal window as CC; cells can respond rapidly to rapamycin, but nuclear depletion can take 20-30 minutes to achieve completion, as shown below with TBP (Haruki et al. 2008). However, this does not mean that only proteins with longer residence times and slower dissociation rates can be measured; see below for more details. To date, a wide variety of nuclear

proteins have been successfully depleted by the AA approach, supporting its potential value in making dynamic measurements (Haruki et al. 2008, Grimaldi et al. 2014, Baptista et al. 2017, Warfield et al. 2017). Only single exponential fits of datasets have been done that do not take the factor export rate into account, but by developing a more representative model with fewer assumptions, more accurate information can be extracted.

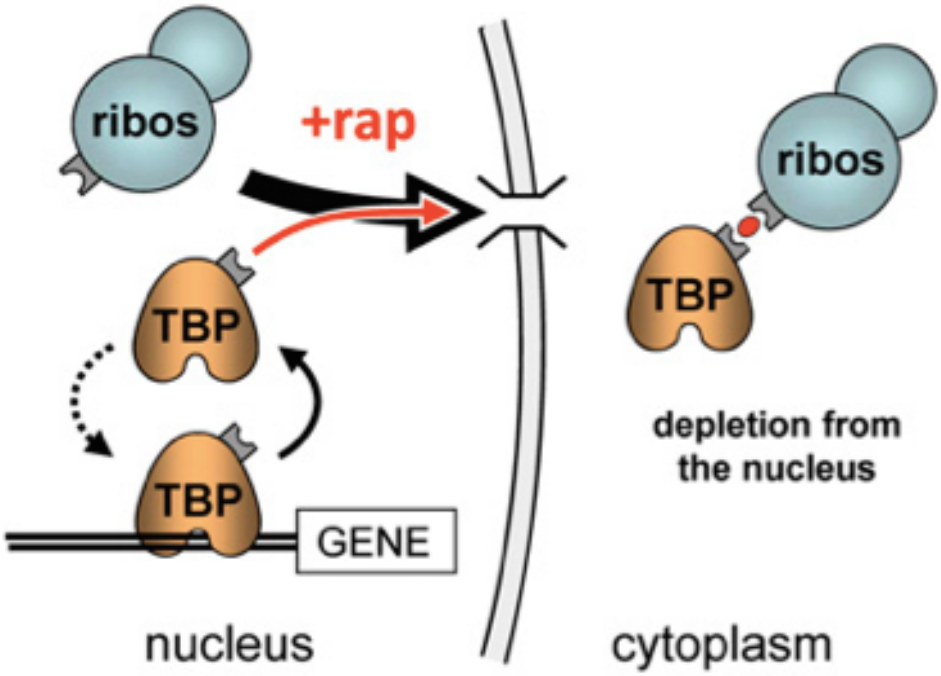


Figure 3.1. Cartoon of the Anchor Away method from Grimaldi *et al*., 2014. Each of the proteins involved (ribos, TBP) has an additional domain, represented by a gray crown shape. As TBP dissociates from DNA (double-stranded black line upstream of "GENE") and rapamycin (red circle) is added to the system, heterodimerization occurs between TBP and the blue ribosomal anchor protein. The ribosomal protein transports the target protein out of the nucleus into the cytoplasm, depleting it from the nucleus as long as rapamycin remains present in the system.

Our lab developed an approach for performing Anchor Away under conditions similar to those for the CLK assay. As in CLK, cultures are grown and then concentrated 5X to ensure an efficient quench. Rapamycin is added to the culture and at discrete time points, a treated culture aliquot is removed and crosslinked for five minutes to capture complexes at each time point of interest, followed by glycine quenching as with CLK. Since the culture is only crosslinked after removal from the larger culture, formaldehyde incubation time is reduced compared to CLK, which can actually be a benefit since increased time spent in a higher concentration of formaldehyde (like the 5% used in CLK) can potentially lead to amalgamation of insoluble or masked complexes (Hoffman et al. 2015). While this method has great potential, some caveats need to be kept in mind including the need to measure the depletion rate of the protein of interest to use as an input variable in the fitting process. This will be explored further in the Discussion. Even with these caveats, the Anchor Away method has potential to be a useful ChIP-based approach for measuring binding dynamics, especially in association with other independent approaches.

Competition ChIP measures turnover rates at single-copy binding sites

Competition ChIP (CC) (Nalley et al. 2006, Yu and Kodadek 2007) has been used to measure TF or histone turnover (Schermer et al. 2005, Dion et al. 2007, Jamai et al. 2007, Rufiange et al. 2007, Collins et al. 2009, van Werven et al. 2009, Lickwar et al. 2012). CC measurements are made using cells expressing two differentially tagged isoforms of the same DNA binding protein, one of which is constitutively expressed and the other is controlled by an inducible promoter such as *GAL1*, *CUP1*, or the Tet-on system (Figure 3.2). The different tags allow DNA association of each isoform to be measured by ChIP following induction of the competitor. Using the levels of the ChIP signals for the two isoforms measured at different times after induction, one can extract site-specific residence times rather than simple turnover ratios (Lickwar et al. 2013, Zaidi et al. 2017). The ability to model actual residence times is significant,

as it allows comparison with other dynamic measurements made at the same loci such as binding of other TFs, histone turnover, rates of RNA synthesis, and how these change in response to different conditions. An alternative approach utilizes metabolic labeling and has been particularly informative for understanding histone turnover (Deal et al. 2010).

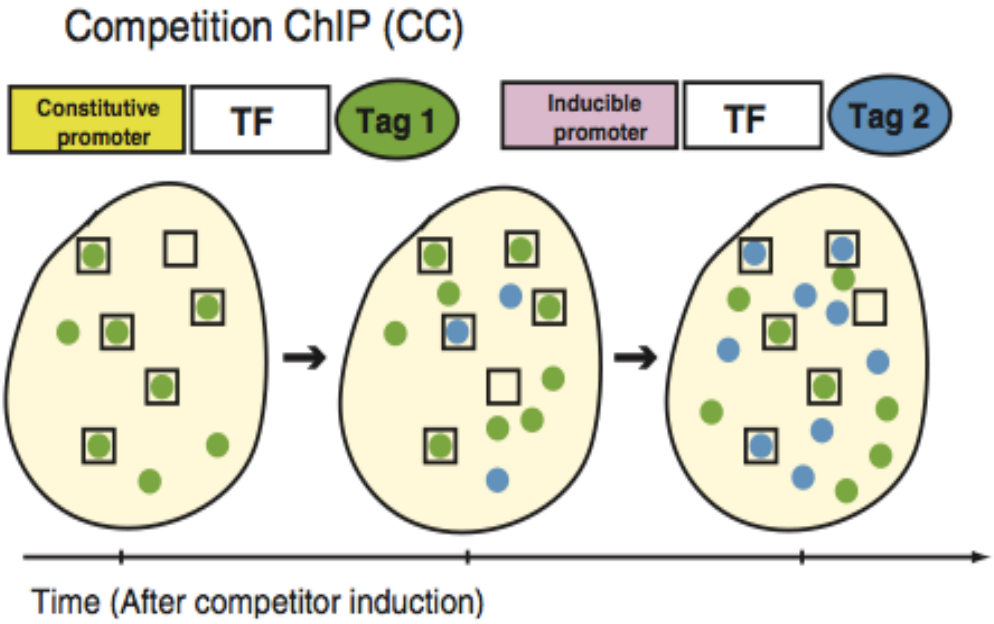


Figure 3.2. Schematic of the Competition ChIP assay from Mueller *et al.* 2013. A diploid strain is constructed with a differentially tagged allele (green and blue ovals) and with one allele under the control of an inducible promoter (pink rectangle). Inducer is added to the system and samples are collected from time zero. At the time when the inducer is added, all binding sites contain the transcription factor (TF) under control of a constitutive promoter with Tag 1; this is represented by the left image showing a nucleus and square binding sites with green circular TF-Tag1 either diffusing throughout the nucleus or bound to a specific site. As time post induction increases, some of these proteins dissociated and newly-synthesized Tag2-labeled TF can then bind DNA; this is shown in the center and right illustrations. As time increases, more TF-Tag2 is chromatin bound and eventually the two proteins will reach equilibrium (right panel).

The power of CC for understanding TF binding dynamics on a global scale is illustrated by analysis of the TF Rap1 (Lickwar et al. 2012). The TF dissociation rate was modeled as approximately equal to its site-specific turnover rate. Transcriptional activation correlated with longer residence times, and thus more stable Rap1 binding, while lower expression was associated with "treadmilling," a phenomenon in which Rap1 occupancy was comparable to occupancies at transcriptionally active promoters, but the Rap1 chromatin interactions had much shorter residence times. Treadmilling was proposed to be linked to unstable PIC formation and occurred at loci where nucleosomes and Rap1 competed for chromatin binding leading to shorter binding times. The treadmilling phenomenon illustrates the limitation of using relative occupancies to predict TF binding stability, as occupancy at an active promoter did not necessarily correlate with a more stable binding interaction (Lickwar et al. 2012). CC has also been used to study the turnover of promoter-bound TBP (van Werven et al. 2009, Grimaldi et al. 2014, Zaidi et al. 2017). TBP was most stable at Pol I promoters, but was much more rapidly turned over at Pol II promoters compared to both Pol I and III promoters, highlighting fundamental differences in promoter binding dynamics for the same factor at different sites and roughly correlating TBP complex stability with transcriptional activity. A limitation of CC results from the lag time required for induction of the competitor. For this reason, CC was thought to be appropriate for measuring TFs with residence times longer than ~500 seconds (Lickwar et al. 2013). However, this is no longer a limitation, as a model was developed that uses the rate of synthesis of the competitor to extract residence times as fast as 1.3 minutes (Figure 3.3) (Zaidi et al. 2017). This is a significant development since many complexes are short-lived (Hager et al. 2009); the dynamics of more interactions can now potentially be captured and quantified using Competition ChIP, although interactions on the second or sub-second scale will still be beyond the scope of CC to measure. An approach related to CC utilizes site-specific recombination to replace one epitope tag with another, circumventing potential problems with increases in TF dosage resulting from production of the competitor molecule (Verzijlbergen et al. 2010).

However, recombination kinetics and asynchrony likely limit this method to the analysis of chromatin binding events with lifetimes on the order of the cell cycle or longer.

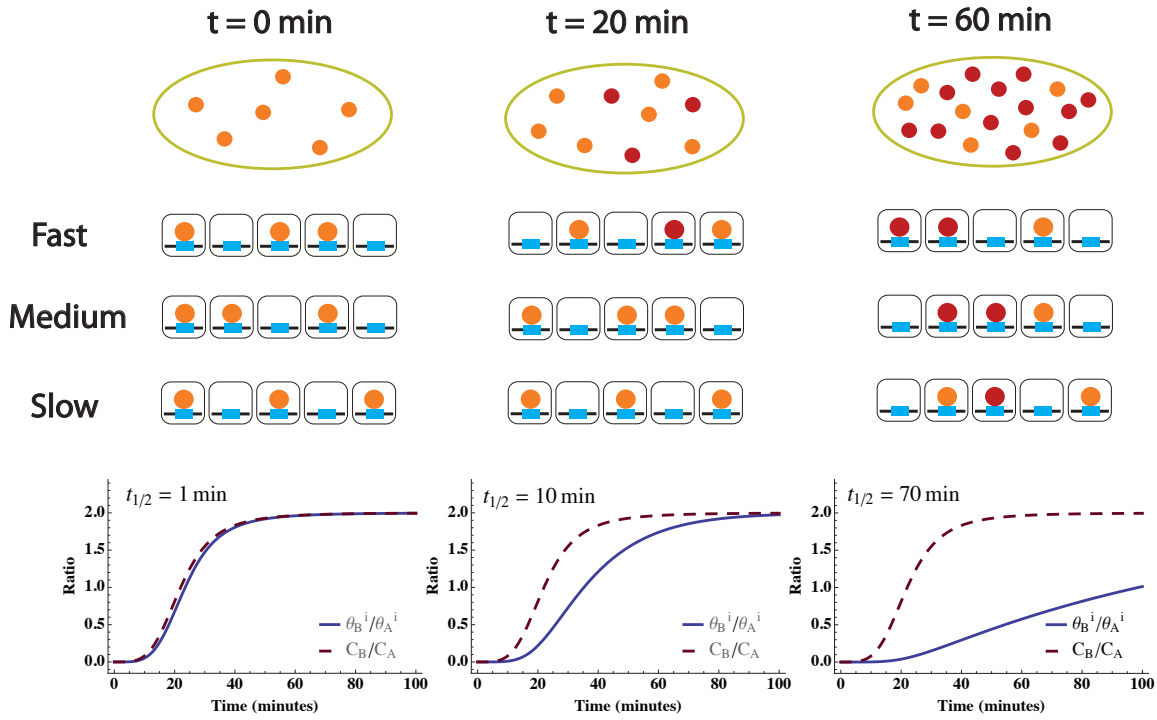


Figure 3.3. Schematic of Competition ChIP assay and resulting induction curves from Zaidi *et al.* 2017. Constitutively activated protein is shown as red circles and induced protein as orange circles. At the start of induction, only constitutively expressed protein is present, as seen in the left column: only orange circles are present inside the cell at the top. Three possibilities exist for the time frame of the residence time: fast, medium, and slow; all show only orange circles occupying binding the blue rectangular sites in each cell in the t = 0 min column. As time increases, the inducible protein is synthesized and begins to bind some DNA sites, seen in the "fast" row in the middle column. By 60 minutes, all potential residence time lengths have inducible protein associated with DNA, cycling with the constitutive protein. The induction curves for fast, medium, and slow binding dynamics are shown in the bottom row from left to right, respectively. Fast binding dynamics almost mimics the induction curve, while medium dynamics have a delayed induction curve (middle), and the slow binding dynamics have a

significantly lagging induction curve. The induction curve plots the ratio of protein with one tag (B) over the other (A) based on western blotting (dotted line labeled C_B/C_A in each plot). The locus-specific induction curve is plotted as occupancy of one tagged protein (B) over the other (A) and is shown as a solid blue line labeled θ_B/θ_A in each plot.

While each of these approaches is able to provide meaningful insight into the *in vivo* dynamics of transcriptional components, they are even more powerful in conjunction. By comparing the results from these three ChIP-based approaches, kinetic behavior can be measured with much more certainty and the process of transcriptional regulation modeled on a larger and more general scale. This chapter discusses dynamic values obtained for TBP using the Anchor Away and Competition ChIP methodologies and how the extracted TBP dynamics from all three ChIP-based approaches compare.

Materials & Methods

Nuclear depletion experiments and Anchor Away sample collection

Nuclear extract isolation and western blotting. Nuclear depletion with rapamycin was first tested by making nuclear extracts (Lue and Kornberg 1987, Ponticelli and Struhl 1990). First, a control experiment was performed to verify collection of nuclear extracts without cytoplasmic contamination. Duplicate primary YPD cultures were started for the TBP Anchor Away strain HHY154 (Haruki et al. 2008) and grown overnight at 30°C. Each culture was added to 2 L YPD the next day and grown at 30°C until the OD_{600} was approximately 5-8. Cells were then spun down in 1 L bottles for 9 minutes at 4,000 rpm and resuspended in 135 ml 50 mM Tris-Cl pH 7.5 (1 M Tris, adjust pH to 7.5 with concentrated HCl) with 30 mM DTT at room temperature with shaking for 15 minutes at 30°C. Cells were spun down for 5 minutes at 5,000 rpm and resuspended in 20 ml YPD/S (1 L: 10 g yeast extract, 20 g peptone, 20 g glucose, 182.17 g

sorbitol (1 M)) at room temperature. Two ml of 50 mM Tris-Cl pH 7.5 with 2X protease inhibitors (made from 200X pepstatin, 1000X leupeptin, 500X chymostatin, 100X PMSF, and 100X benzamidine), 1 mM DTT, and 30 mg zymolyase 100T was added to the cell suspension followed by an incubation at 30°C with gentle shaking. Spheroplast formation was checked by adding 10 µl cell suspension to 1 ml 1% SDS and reading the OD_{600} before and after zymolyase treatment. The starting OD is usually ~1.4 and drops to ~0.4-0.5 when approximately 75% of the cells are spheroplasts. When the OD_{600} of spheroplasted cells had dropped enough, digestion was halted by adding 270 ml YPD/S; cells were then pelleted for 9 minutes at 4,000 rpm. Cell pellets were resuspended in 550 ml YPD/S; the YPD/S was slowly added to prevent cell clumping. Cells then recovered by shaking at 30°C for 30 minutes, followed by pelleting at 4000 rpm for 9 minutes and resuspension in 270 ml ice cold YPD/S. The YPD/S wash was repeated once, then cells were spun for 9 minutes at 4,000 rpm at 4°C and resuspended in 270 ml 1 M sorbitol at 4°C. Resuspended cells were spun down again and pellets were gently resuspended in 135 ml Buffer A (18% polysucrose 400, 10 mM Tris-Cl pH 7.5, 20 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, 0.15 mM spermine, and 0.5 mM spermidine) with 3 mM DTT and 1X protease inhibitors (stocks were 200X pepstatin, 1000X leupeptin, 500X chymostatin, 100X PMSF, and 100X benzamidine). Spheroplasts were lysed by passage through a homogenizer twice at 1000 rpm. To remove unlysed cells and large debris, the lysate was spun at 5,400 rpm for 5 minutes at 4°C. The upper 2/3 clearer-colored supernatant was transferred to a new bottle and spun again two times; the pelleted material and some of the viscous glop above the pellet was tossed. A well-defined pellet was observed by the last spin. Lysates were then spun for 30 minutes at 13,000 rpm and 4°C to recover the nuclei. The pellets were resuspended in ~10 ml Buffer B (100 mM Tris-acetate pH 7.9, 50 mM potassium acetate, 10 mM $MgSO_4$, 20 % glycerol, and 2 mM EDTA) with 3 mM DTT and 1X protease inhibitors (same as Buffer A). Resuspension was done with a glass rod and dounce by hand. The nuclei were frozen at -80°C until ready to run a western blot.

Western blots were performed similarly to those previously described in Chapter II Materials & Methods: *Whole cell extract preparation and western blotting*. Two identical 12% denaturing SDS-PAGE gels were run with titrated amounts of the collected extracts. One to five µg of sample was mixed with an appropriate amount of 4X Laemmli buffer and heated at 95°C for 5 minutes before loading into the gel. Proteins were transferred overnight onto Immobulin P PVDF membrane and detected using either TBP or PGK1 antibodies (58C9, Abcam #ab61411; 22C5D, Abcam #ab113687, respectively).

Once nuclear extracts were obtained, samples were collected following incubation of cells with rapamycin for various periods of time. Nuclear extracts from these samples were used to check for nuclear depletion of the desired factor. Four duplicate cultures, 500 ml each, were grown the same way as described above. When an OD_{600} of ~5-8 was achieved, one 500 ml culture spun down as above for the zero minute time point. Rapamycin was then added to the remaining three cultures to a final concentration of 10 µg/ml; cultures were incubated on a stir plate at room temperature with rapid stirring. At 5, 10, and 15 minutes, one 500 ml culture was filtered to remove rapamycin and prevent further depletion, then the cells on the filter were resuspended in 33.75 ml 50 mM Tris-HCl pH 7.5 plus 30 mM DTT and incubated for 15 minutes at 30°C. Extracts were prepared as described in the previous paragraph, but resuspension amounts were quartered because only 500 ml cultures were collected per samples instead of 2 L. TBP and PGK1 proteins were detected on western blots, as described above.

Imaging of fixed cells to visualize nuclear depletion. A second way to determine the nuclear depletion of the protein of interest was performed by quantifying GFP-labeled protein in the nucleus and cytoplasm over time (Haruki et al. 2008). Two 25 ml SC + adenine + 2% glucose cultures were started and grown overnight at 30°C; one culture was started with a colony of the GFP-tagged Anchor Away strain, HHY209, and the other from a colony of the parental strain, HHY168. The OD_{600} of both cultures was checked in the morning; cultures were diluted if the

OD_{600} was above 0.4. When the OD_{600} was ~0.2, each culture was split into two 12.5 ml cultures. Rapamycin was added to a final concentration of 10 µg/ml to one culture and the other culture was left untreated as a control; both strains were treated in this way. Cultures were incubated in a 30°C shaker between collections. A 1ml aliquot was taken from each untreated culture as the zero time point. For the parental strain, a 1 ml aliquot was taken at the following times: 30 minutes (+ rapamycin culture), 75 minutes (both +/- rapamycin cultures). For the GFP-tagged strains, a 1 ml aliquot was taken at 5, 10, 15, 20, 25, 30, 45, 60, and 75 minutes for the culture with rapamycin added and at 75 minutes for the untreated culture. After an aliquot was removed, the cells were spun down 3 minutes at 4,000 rpm and the pellet was resuspended in 100% methanol pre-chilled at -20°C. Cells were fixed for 6 minutes at -20°C, then spun down at 4,000 rpm. Pellets were rehydrated in 1 ml PBS with 0.2% TWEEN20 and 20 ng/ml DAPI for 5 minutes at room temperature. Cells were centrifuged 3 minutes at 4,000 rpm; most of the supernatant was removed and pellets were resuspended in 4 µl PBS. The total resuspension volume was ~12 µl and more PBS was added to bring the volume up if needed. All 12 µl of cell resuspension was added to the center of a round, walled slide. A circle of 1% agarose (made with ddH_2O) was placed on top of the droplet and a slide was placed on top of the agar pad. A circular weight was placed on top of the cover slide to spread the fixed cells into a monolayer. Each slide was imaged using a Zeiss light microscope; Colibri LED lights at 365 and 470 nm were used to image DAPI and GFP channels, respectively. Transmitted light images were also taken for each sample. For each channel (GFP, DAPI, transmitted light), 11 z-stacks were taken for each image; 40-50 images were taken per sample. A second set of parental and GFP-tagged treated/untreated samples was collected, fixed, and imaged the next day.

Quantification of nuclear depletion using microscopy images. After multiple microscopy images were collected for each time point for two biological replicates, the nuclear and cytoplasmic GFP intensities were quantified through a series of steps in MATLAB developed by Tomáš Vičar with

help from Kristýna Kupková. Cell segmentation was done for the cell and nucleus. First, the transmitted light and DAPI images were preprocessed using median and Gaussian filters. Next, a set of features was extracted from each pixel in each stack of the transmitted light images; all 11 z-stack slices were used. These features were then fed to a machine learning classifier, the Random forest algorithm, which classified each pixel as a cell or not. Random forest was trained on 10 manually generated images, where the cell outline had been drawn by hand in ImageJ, and then with 20 images generated from the algorithm. The 20 images generated were manually fixed and rerun with Random forest to refine the algorithm. The algorithm was run for all transmitted light images followed by thresholding to remove small objects and generate the cellular masks. Next, all z-stacks in each DAPI image were adaptively thresholded and smaller objects were removed to generate the nuclear masks. Several morphological operations were applied to both cellular and nuclear masks to give the objects a more smoothed, biological shape. At this point, both cellular and nuclear masks were manually edited to remove any improperly identified objects. Once both masks were generated, cells that did not have a corresponding nucleus in the masks were removed, and vice versa. Masks were finalized after this step and saved as TIFF files.

Once cell segmentation was complete, GFP z-stacks were normalized before mask application. The mode of each stack was determined and subtracted from each pixel to eliminate background. For quantification of GFP intensity, the slice with the highest summed DAPI intensity under the nuclear mask for each cell was used to compute the mean nuclear and cellular GFP. To quantify the GFP intensity under the nuclear and cellular masks for each image, each mask was overlaid on the corresponding GFP image and the GFP intensity was computed. The average quantified pixel intensity for each cell and nucleus was imported into a table. The parental cellular and nuclear GFP intensities across the four collected time points was averaged into one value for background cellular fluorescence and one value for nuclear fluorescence. Each of these values was subtracted from the experimental TBP-GFP data to give final intensity values.

The experimental time course was plotted as a boxplot and fit with an exponential fit through the medians in R that gave a depletion value of τ.

Anchor Away sample collection. All Anchor Away strains used were purchased from EUROSCARF and are listed in Table 3.1; there overall workflow for collection of Anchor Away samples is outlined in Figure 3.4 below. Duplicate 5 ml primary cultures were started with one colony from the desired strain and incubated with shaking overnight at 30°C. All of each primary culture was added to 450 ml YPD and grown at 30°C until the OD_{600} was ~1; the cells were then spun down and concentrated 5X in 90 ml YPD. Similar to the CLKv2 sample collection procedure, 440 ml 3 M glycine pH 5 was prepared for each of the eight samples collected. A 10 ml aliquot was removed for the zero minute time points; it was crosslinked with 5% formaldehyde for 5 minutes at room temperature on a stir plate, then quenched with 440 ml 3 M glycine (2.93 M final quench concentration) and spun down for 7 minutes at 5,000 rpm at 4°C. Rapamycin was added to a final concentration of 10 μg/ml to the remaining 80 ml culture and mixed in with a stir bar. Aliquots of 10 ml were removed at desired time points and crosslinked with 5% formaldehyde for 5 minutes, as with the zero minute time point. Crosslinked cells were quenched and spun down, as above. All pellets were washed with cold TBS + 300 mM glycince and processed as chromatin samples. The sample preparation through real time data collection was identical to the protocol described in Chapter II for the CLKv2 assay. The normalized ChIP signal for each time point was calculated as a percentage of the zero minute ChIP signal and plotted as a decay curve in GraphPad Prism. In the future, data will be fit with a more complex model that takes the depletion time into account, described in the Discussion section later.

Table 3.1. Yeast strains used in this chapter.

Strain	Genotype	Reference or source
HHY154	*MATalpha tor1-1 fpr1::NAT RPL13A-2xFKB12::TRP1 TBP1-FRB::kanMX6*	Haruki et al, 2008
HHY168	*MATα tor1-1 fpr1::NAT RPL13A-2×FKBP12::TRP1*	Haruki et al, 2008
HHY209	*MATα tor1-1 fpr1::NAT RPL13A-2×FKBP12::TRP1 TBP1-FRB-GFP:KAN*	Haruki et al, 2008
YPH501	*MATa/MATα ura3-52/ura3-52 lys2-801_amber/lys2-801_amber ade2-101_ochre/ade2-101_ochre trp1-Δ63/trp1-Δ63 his3-Δ200/his3-Δ200 leu2-Δ1/leu2-Δ1*	Sikorski and Hieter, 1989
YMV070	*MATa, ura3-52 lys2-801^a ade2-101^o trp1-Δ63 his3-Δ200 leu2-Δ1 MOT1-13xMyc(HIS3) SPT16-HA(TRP1)*	True et al, 2016
YAD154	*MATa ura3-52 lys2-801^a ade2-101^o trp1-Δ63 his3-Δ200 leu2-Δ1 SPT15-myc*	Poorey et al, 2013

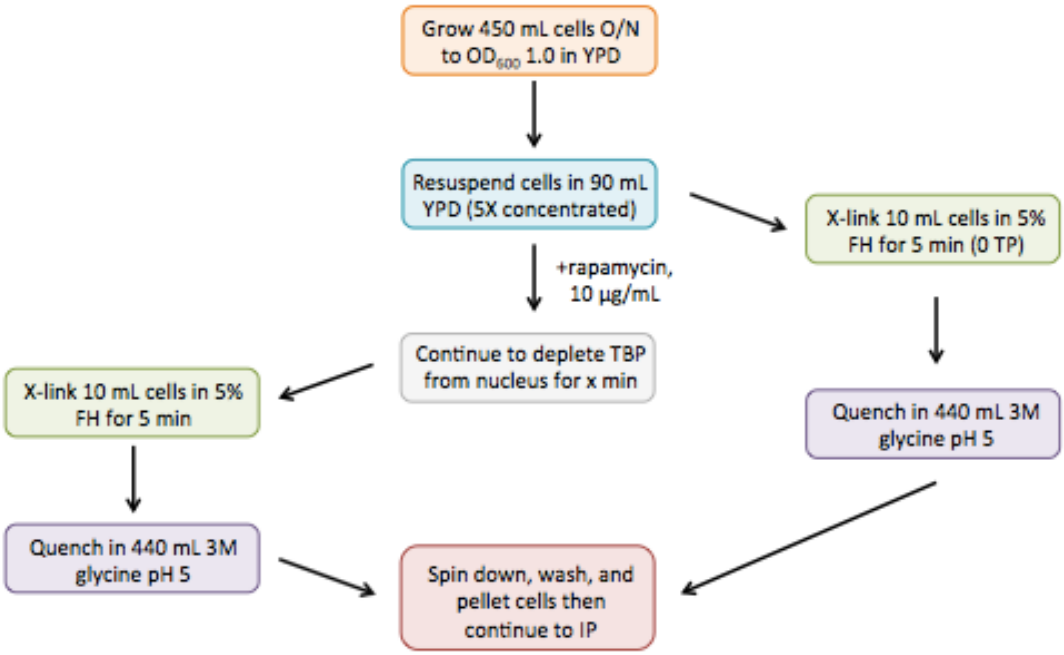


Figure 3.4. Workflow for Anchor Away sample collection. Samples are grown in rich media until the OD_{600} is ~1.0, at which point they are concentrated 5 times. A 10 ml aliquot is removed and crosslinked for 5 minutes in 5% formaldehyde, then quenched with 2.93 M glycine pH 5. The remaining culture is treated with rapamycin to induce nuclear depletion. Aliquots of 10 ml are removed at desired time points and crosslinked and quenched in the same way as the zero time

point. After quenching, cells are spun down and processed as described in Chapter II and outlined in Figure 2.6.

Construction of the TFIIB Competition ChIP strain

The following protocol was adapted from van Werven *et al.*, 2009 with Myc tagging adapted from Gauss *et al*, 2005. An overview of the strain construction is shown in Figure 3.5.

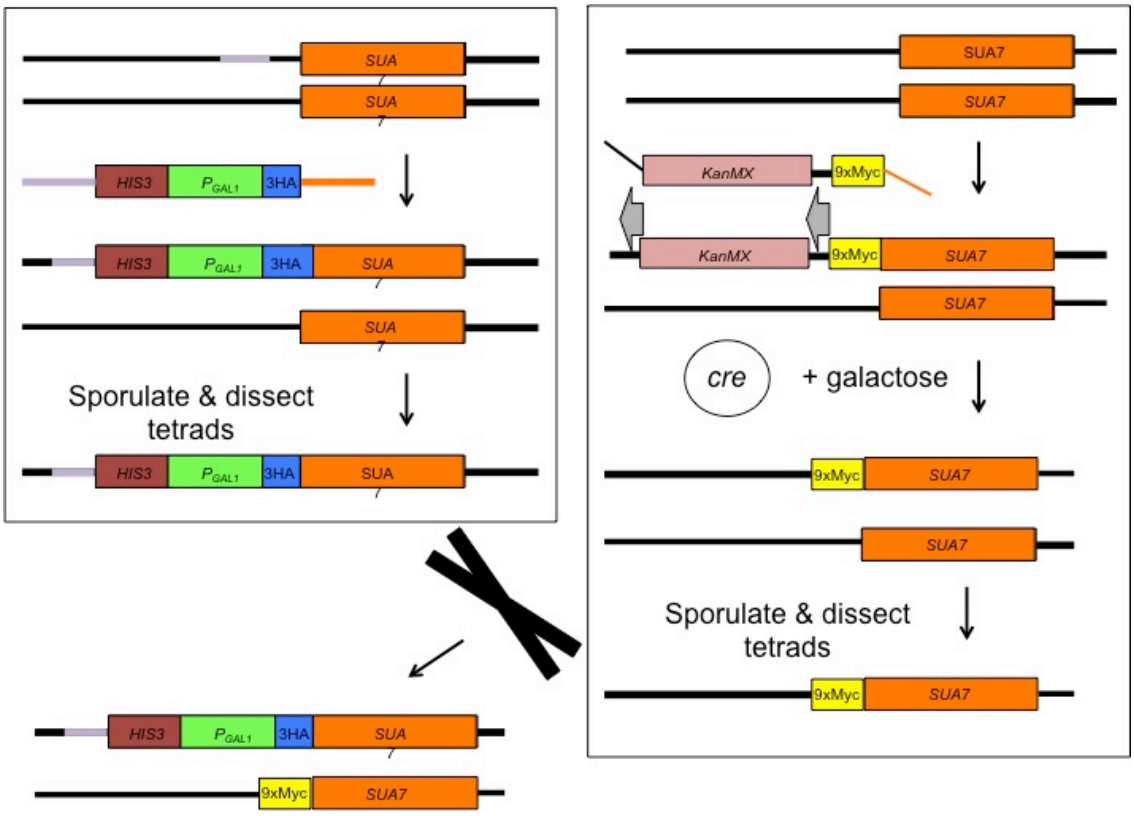


Figure 3.5. Overview of TFIIB Competition ChIP strain construction. Yeast TFIIB is encoded by *SUA7*. Two identical diploid strains are differentially tagged, one with HIS3-PGAL1-3HA (left side) and the other with KanMX-9xMyc (right side), where KanMX is flanked by loxP sites (gray arrows). A cassette containing each tag is amplified from a plasmid and transformed into the strain. The Myc strain is then transformed with a galactose controlled Cre recombinase plasmid and induced with galactose addition to excise the KanMX marker. Once tagged, both

strains are sporulated and dissected to yield a tagged haploid. The haloids are crossed to make the final Sua7 Competition ChIP strain.

Amplification of the inducible GAL1 promoter (P_{GAL1}) and N-terminal HA tag cassette. A cassette containing HISMX6-PGAL1-3HA was PCR amplified from the pFA6-HisMX6-PGAL1-3HA plasmid (Longtine et al. 1998) using the PGAL-3HA-F/PGAL-3HA-R primers (Table 3.2), Phusion HF polymerase (NEB), and an annealing temperature of 60°C; a total volume of 300 µl was used for the PCR reaction and the suggested NEB protocol was followed. A few microliters of the PCR amplicon with 10X loading dye was checked on a 1% agarose gel with ethidium bromide. After verification of the correct band size, the rest of the PCR reaction was precipitated with a phenol extraction. An equal volume of phenol:cholorphorm:isoamyl alcohol was added to the pooled amplicon, vortexed briefly, and spun for 5 minutes at 13,200 rpm. The top aqueous phase was transferred to a new tube followed by addition of an equal volume of isopropanol, brief vortexing, and incubation at room temperature for 15 minutes. The sample was then spun for 30 minutes at 13,200 rpm and the supernatant was removed post-spin. The pellet was washed with 800 µl 80% ethanol pre-chilled to -20°C and spun at 13,200 rpm for 10 minutes. The supernatant was removed and the pellet was dried in the hood for ~20 minutes, then resuspended in 20 µl 1X TE pH 8.

Table 3.2. Primers used to make the TFIIB Competition ChIP strain (5'-3').

Name	Sequence
PGAL-3HA-F	AAA AAA AGT GAA GAG AAT AAT CAT CAC TTA TAA AGA CAA CTT AAT AGA CGG AAT TCG AGC TCG TTT AAA C
PGAL-3HA-R	AGG ACC CCT TCT TCC TGC TCT TTT ATC TAT GCT CTC CCT AGT GCA CTG AGC AGC GTA ATC TG
3HA-Sua7_check-F	TAT ACT TTA ACG TCA AGG
3HA-Sua7_check-R	CTT CCT GCT CTT TTA TCT
HISMX6-Sua7_check-F	CTC GCC TTG ACT GCA CAT
N-lox-9Myc-F	TGT GGA CGA TCC AGT GAT AGA GAA GGG GAG AAG TAG ATA CGC AGA ATG TGC AGG TCG ACA ACC CTT AAT
N-lox-9Myc-R	ATT AGG ACC CCT TCT TCC TGC TCT TTT ATC TAT GCT CTC CCT AGT GCG GCC GCA TAG GCC ACT
pCORE-F	TGG TGT GGA CGA TCC AGT
pCORE-R	AGC ACA TAC AAC ATC CCC
Avitag-lox_check-R	CAC CGT TGT GAT CAT CAT TTG

Transformation of the N-terminal cassette directly upstream of SUA7. Following amplification of the desired fragment, a transformation of the cassette into the diploid wild type YPH501 strain was performed. The night before the transformation, a 100 ml YPD culture was started with a single colony and grown overnight at 30°C with shaking; the OD_{600} was checked in the morning, and the procedure began when the OD_{600} was approximately 1. The culture was spun dow 5 minutes at 4,000 rpm and 4°C. The pellet was then washed in 10 ml ddH_2O and spun as in the previous step. Another wash with 5 ml TE/LiOAc (1X TE, 1X LiOAc) was performed, followed by resuspension of cells in 1 ml TE/LiOAc and transfer to a 1.5 ml eppendorf tube on ice. A stock tube of 10 mg/ml ssDNA was boiled at 95°C for 5 minutes, then cooled on ice. Four transformation tubes were set up with 150 µl YPH501 cells and 15 µl boiled ssDNA (150 µg). Two tubes had 10 µl each of concentrated HISMX6-PGAL1-3HA PCR product added, one had 1 µl of a control plasmid such as pRS315 (Sikorski and Hieter 1989), and the last had nothing

added as a negative control. Tubes were flicked to mix; 900 µl TE/LiOAc/PEG (1X TE, 1X LiOAc, 40% PEG 4000) was added to each and tubes were immediately vortexed. Reactions were incubated at 42°C for 45 minutes then pelleted with a quick spin. Pellets were resuspended in 1 ml YPD and incubated in a shaker at 30°C for 3-4 hours, then quick spun again. Pellets were resuspended in 150 µl 1X TE and plated on SC-HIS plates followed by incubation at 30°C for 2-3 days. Alternatively, pellets can be resuspended in 150 µl 1X TE following heat shock at 42°C, plated onto YPD plates and incubated at 30°C overnight, then replica plated on SC-HIS plates and left to grow up at 30°C for two days. Once candidate colonies had grown up, they were restreaked onto SC-HIS plates and grown at 30°C for two days to get single colonies.

Verification of transformation candidates by PCR with gDNA and western blotting To check the candidate transformants, gDNA was made from single colonies grown in 10 ml YPD at 30°C overnight. Semi-saturated cultures were spun down for 5 minutes at 4,000 rpm and 4°C to pellet the cells, which were then resuspended in 500 µl ddH_2O and transferred to an eppendorf tube. Each tube was briefly spun down to pellet the cells and the supernatant was removed; pellets were then briefly vortexed to resuspend cells in the residual liquid. To lyse cells and extract DNA, 200 µl lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM Na_2EDTA), 200 µl of 25:24:1 phenol:chloroform:isoamyl alcohol, and 300 mg of acid washed beads (Fisher) were added to each cell suspension. Samples were vortexed for 3-4 minutes, followed by an addition of 200 µl 1X TE pH 8. Cells were spun down for 5 minutes at 13,200 rpm; the top aqueous phase was transferred to a new eppendorf tube and mixed with 1 ml 100% ethanol by inversion. Cells were centrifuged for 2 minutes at 13,200 rpm and the supernatant was discarded. Each pellet was resuspended in 400 µl 1X TE with 3 µl 10 mg/ml RNase A and incubated at 37°C for 5 minutes; 5.4 µl 7.5 M ammonium acetate and 1 ml 100% ethanol were added and tubes were inverted to mix. Samples were spun 2 minutes at 13,200 rpm and the

supernatant was removed. Pellets were air-dried at room temperature for ~20 minutes, then pellets were resuspended in 50 µl 1X TE.

Two sets of primers (3HA-Sua7_check-F/3HA-Sua7_check-R and HISMX6-Sua7_check-F/3HA-Sua7_check-R) were used to check the candidate gDNA. PCR reactions were performed using Phusion HF polymerase (NEB) with a 50°C annealing temperature. Resulting PCR amplicons were checked on a 1% agarose gel with ethidium bromide run at 100 V for ~60 minutes.

The HA tag was also verified by western blot, similar to the protocol described in Chapter II Materials & Methods: *Whole cell extract preparation and western blotting*. First, duplicate 5 ml YPD cultures were started for the tagged strain, diploid wild type YPH501 (negative control), and an HA-tagged positive control strain (YMV070); primary cultures were grown overnight in a shaker at 30°C. The next morning, 2.5 ml of the tagged strain primary culture was added to 100 ml YEP + 2% raffinose and the other 2.5 ml were added to 50 ml YPD; the control strains were each added to 50 ml YPD. All cultures were grown at 30°C. When the OD_{600} of all cultures except the tagged strain in raffinose reached 1, the cultures were spun down for 5 minutes at 4,000 rpm and 4°C. When the OD_{600} of the tagged strain culture in raffinose reached 0.8, it was spun down for 5 minutes at 4,000 rpm and 4°C, then resuspended in 100 ml YEP + 2% galactose and incubated at 30°C until the OD_{600} was ~1. The culture was then spun down for 5 minutes at 4,000 rpm at 4°C. All pellets were washed with 50 ml cold (at 4°C) TBS and processed as whole cell extracts (see Chapter II Materials & Methods referenced above). A 10% SDS-PAGE gel was run with 20 µg of each sample mixed with 4X Laemmli buffer and heated for 5 minutes at 95°C before loading. Once the gel was transferred onto PVDF membrane, an antibody to detect HA was used (12ca5, Roche cat# 11583816001).

Verification of correct insertion by sequencing. Using the gDNA from 1-2 verified candidates, PCR reactions were done using the HISMX6-Sua7_check-F/3HA-Sua7_check-R primers and the

Phusion HF polymerase with an annealing temperature of 50°C; 300 µl of reaction was set up per sequencing sample. The PCR product was pooled and then concentrated by adding 1/10th volume 3 M sodium acetate and 1 ml cold (-20°C) 100% ethanol. The sample was vortexed briefly then left at -20°C for 1 hour. The DNA was spun down for 30 minutes at 14,000 rpm and 4°C. The resulting pellet was washed with 800 µl 80% ethanol at -20°C and spun down for 10 minutes at 14,000 rpm and 4°C; the ethanol wash was repeated once. The supernatant was removed, the pellet was dried for ~20 minutes in the hood, and the pellet was resuspended in 50 µl 1X TE. All of the PCR reaction was run on a 1% agarose gel with ethidium bromide; the desired band was visualized with a GelDoc and UV box, then excised from the gel using a razor blade. The NuceloSpin Gel and DNA cleanup kit was used to clean up the gel slice and isolate the DNA fragments; the standard kit protocol was followed with final elution volume of 30 µl. Desired samples were sent to GeneWiz for Sanger sequencing and contained the following: ~60 ng DNA, 25 pmol PGAL1_U_F primer, and ddH_2O to 15 µl. Once sequence was received from GeneWiz, it was compared to expected sequence using BLAST (NIH).

Sporulation, dissection, and verification of tagged tetrads. Following sequencing and western blot confirmation of the correct tag, the diploid strain was sporulated using a protocol from the Amberg lab. Duplicate primary cultures of 3 ml YPD was grown overnight at 30°C. The next day, cells were spun down for 3 minutes at 3,000 rpm. Each pellet was washed twice with 1 ml ddH_2O and spun down as in the last step. Each pellet was resuspended in 1 ml sporulation media (0.25% yeast extract, 1.5% potassium acetate, 0.05% dextrose, 1X adenine/uracil/tryptophan/histidine/leucine/lysine). In a separate 15 ml tube, 500 µl of the resuspended cells was added to 2 ml sporulation media. Tubes were incubated at room temperature on a rotor for 3-5 days. About 10 µl of cells were put on a glass slide and checked via microscope for formation of tetrads. When at least 20% of the cells had formed tetrads, the sporulation culture was ready for dissection (Amberg et al. 2005). One culture was dissected at a

time; 100 µl sporulated cells were aliquoted into an eppendorf tube and 3.75 µl of 50 mg/ml 20T zymolyase (MP Biomedicals) was added to start cell wall digestion; cells were vortexted for ~1 second to mix the cells and enzyme. The digestion was done at room temperature for ~3.5 minutes; 1 ml cold (at 4°C) ddH_2O was slowly pipetted down the side of the tube to stop the digestion. Without disturbing the tube contents, 15 µl of digested culture was pipetted down the center of a YEP + 2% galactose plate and allowed to dry for a few minutes at room temperature. A dissection scope, provided by the Smith lab, was used to pick up and separate tetrad candidates. About 16 tetrads were picked per plate and at least two plates of dissections were performed. Plates were incubated at 30°C until tetrads grew up. Each tetrad was restreaked onto a YEP + 2% galactose plate to get single colonies; plates were incubated at 30°C for 2-3 days. A master plate was made with each tetrads streaked out onto a YEP + 2% galactose plate; master plates were grown at 30°C for 1-2 days, then replica plated onto relevant plates (YPD, YEP + 2% galactose, SC-Ura/-Leu/-Ade/-His/-Trp) and incubated at 30°C for 2 days to check for 2:2 segregation of tetrads. A mating test with tester strains 14a and 17α, from the Smith lab, was also done to test for a 2:2 segregation of tetrads. A line (or multiple lines if many tetrads checked) of each tester strain was made vertically down an SD plate; small cross-hatches were made on these lines of each tetrad tested. The SD plates were incubated at 30°C for 2-3 days then each plate was checked for absence or presence of growth.

Amplification and transformation of the loxP-kanMX-loxP-9xMyc N-terminal tag cassette. The second tag was inserted into a wild type YPH501 strain, identical to the one used for the first tagging. A cassette containing the 9xMyc tag, as well as a loxP-flanked *kanMX* marker, was amplified from the pOM20 plasmid (Gauss et al. 2005) using the N-lox-9Myc-F and N-lox-9Myc-R primers, the Phusion HF polymerase (NEB), and an annealing temperature of 68°C; this was done the same way as the first tag outlined above in *Amplification of the inducible PGAL1 promoter and N-terminal HA tag cassette*. Once the cassette was concentrated, it was transformed

into YPH501, as described above in *Transformation of the N-terminal cassette in front of SUA7*. Once candidate transformants were identified, they were checked via PCR with gDNA, as outlined earlier in *Verification of transformation candidates by PCR with gDNA and western blotting* (western blotting was not done until the *kanMX* marker was knocked out).

Inducible knockout of kanMX upstream of the 9x-Myc tag. The diploid strain with the integrated marker and tag was transformed with pSH47, a *PGAL1*-controlled Cre recombinase plasmid. The transformation was similar to the protocol described earlier in *Transformation of the N-terminal cassette in front of SUA7*, but 50 µl of yeast cells, 5 µl of ssDNA (50 µg), and 300 µl of TE/LiOAc/PEG were used. Instead of purified PCR amplicon, 1 µl of pSH47 was added to the transformation reaction and one tube was used as a negative control with no DNA added. Cells were plated on SC-URA plates and incubated for two days at 30°C. Candidates were streaked out on SC-URA plates to get single colonies after incubation at 30°C for two days. Once the Cre recombinase transformation was confirmed, the *kanMX* marker was knocked out. Two colonies were taken from one plate containing the labeled strain with Cre recombinase plasmid; each colony was added to 2 ml YEP + 2% galactose for 2 hours at 30°C. Dilutions of cells (1:100, 1:1000, and 1:10000) were plated on YPD plates, with a total volume of 200 µl cell dilution spread on each. Plates were incubated at 30°C for two days, then dilutions with less than 400 colonies grown up were replica plated onto YPD+G418 plates and incubated at 30°C for 1-2 days. The YPD and YPD+G418 plates were compared; colonies that grew on the YPD plate but not the YPD+G418 plates were streaked out on YPD plates and grown for two days at 30°C to get single colonies.

Verification of knockout candidates by PCR with gDNA and western blotting. Knockout of *kanMX* from the 9xMyc-tagged diploid strain was confirmed by PCR using gDNA, as described

earlier in *Verification of transformation candidates by PCR with gDNA and western blotting*; the HISMX6-Sua7_check-F/3HA-Sua7_check-R primers were used to check for loss of insert.

The 9xMyc tag was also confirmed by western blot, similar to the protocol described in Chapter II Materials & Methods: *Whole cell extract preparation and western blotting*. First, duplicate 5 ml YPD cultures were started for the tagged strain, diploid wild type YPH501 (negative control), and a Myc-tagged positive control strain (YAD154); primary cultures were grown overnight in a shaker at 30°C. The next morning, each of the primary cultures was added to 50 ml YPD and grown at 30°C. When the OD_{600} of each culture reached 1, the cells were spun down for 5 minutes at 4,000 rpm and 4°C. All pellets were then washed with 50 ml cold (at 4°C) TBS and processed as whole cell extracts (see Chapter II Materials & Methods referenced above). A 10% SDS-PAGE gel was run with 25 μg of each sample mixed with 4X Laemmli buffer and heated for 5 minutes at 95°C before loading. Once the gel was transferred onto PVDF membrane, an antibody to detect Myc was used following blocking (9E10, abcam #ab32).

Verification of correct insertion and knockout by sequencing. Sequencing was done for one or two of the 9xMyc-tagged candidates with confirmed knockout, as described earlier in *Verification of correct insertion by sequencing*. The pCORE-F and Avitag-lox_check-R primers were used for amplification of the desired fragment and the sequencing primer was pCORE-R.

Sporulation and dissection of tagged tetrads. The 9xMyc-tagged diploid strain was sporulated and dissected, as described above for the HA-tagged strain. However, the tetrads were dissected onto YPD plates. Presence of the Myc tag was verified by PCR with the HISMX6-Sua7_check-F/3HA-Sua7_check-R primers and N-lox-9Myc-F/N-lox-9Myc-R primers.

Haploid strain crossing to get differentially-tagged diploid. The *HISMX6-PGAL1*-3HA and 9xMyc *SUA7* N-terminally tagged diploids were crossed. One colony of each was streaked out

next to each other on a YPD plate and mixed together with ~15 µl ddH_2O; this was done for several haploid combinations. The mating reaction was incubated at 30°C for 4-6 hours. A wooden stick was streaked down the middle of the reaction and drawn down the center of another YPD plate. This plate was checked under the microscope for presence of diploids; ~16 diploids were picked per plate and incubated at 30°C for 2-3 days, then restreaked for single colonies. Once diploid colonies grew up, the mating type was checked using the mating test described earlier in *Sporulation, dissection, and verification of tagged tetrads*. Verified diploid strains were checked by PCR amplification of gDNA with HISMX6-Sua7_check-F/3HA-Sua7_check-R primers.

Spot test to compare growth of tagged strain to original wild type diploid. Primary cultures for two isolates of the tagged strain, YPH501 wild type diploid, and any controls were grown from single colonies in 5 ml YPD at 30°C overnight. The next day, the OD_{600} for each culture was measured; all cultures were diluted to the lowest OD_{600} measured, or 1 if the cultures are saturated. Using a 96-well plate, 200 µl of each sample was put into a well in the first column. To the next four columns, 180 µl ddH_2O was added. To make the serial dilutions, 20 µl from the first column was added to the second column using a multi-channel pipette (1:10). This was repeated for the next column: 20 µl from the second column was added to the third column (1:100). This was continued for the remaining columns (1:1,000 to 1:100,000). Once the dilution series was completed, 3 µl from each well in the first column was removed with a multi-channel pipette and plated on the far left side of a YPD plate in a vertical column; this was repeated for each column, plating columns left to right. Plating was repeated on a YEP + 2% galactose plate as well. Plates were left on the bench for a few minutes to dry, then incubated at 30°C for ~2 days.

Collection of induction curve and verification by western blot. Whole cell extract samples of the differentially tagged diploid strain were collected before and after induction with galactose.

Duplicate 5 ml YPD cultures were started with one colony each of the diploid strain and incubated with shaking overnight at 30°C. All of each primary culture was added to 450 ml YEP + 2% raffinose and grown at 30°C until the OD_{600} was ~0.6. The zero minute time point was removed; 50 ml were collected and spun down for 5 minutes at 4,000 rpm and 4°C. Another 50 ml aliquot was removed and 2% glucose was added; this negative control was grown for the length of the time course at 30°C then spun down. Raffinose was added to the remaining culture at a final concentration of 2%; this culture was incubated at 30°C. At desired time points, a 50 ml aliquot was removed from the large culture and spun down at 4,000 rpm/4°C for 5 minutes. All cell pellets were washed with cold TBS and processed as whole cell extracts, as outlined in Chapter II Materials & Methods: *Whole cell extract preparation and western blotting*. Two identical 10% SDS-PAGE gels were run and probed with either an HA or Myc antibody. The 3HA-Sua7 and 9xMyc-Sua7 bands were quantified with ImageJ (NIH) and plotted and fit in Excel.

Results

Detection of TBP in nuclear extracts

The AA nuclear depletion rate is essential for determining the relationship between rapamycin-induced changes in ChIP signal, chromatin binding off-rate, and rate of export from the nucleus. To measure the nuclear depletion rate, the first approach was to analyze nuclear extracts from the TBP Anchor Away strain, HYY154 (Haruki et al. 2008). First, it was determined if nuclear extracts could be isolated that were free from cytoplasmic contamination. Biological replicates were collected and verified by western blot with probes for TBP, as well as PGK1, a cytoplasmic marker protein (Kumar et al. 2002, Finnigan et al. 2016). Whole cell extracts from the TBP CLK WT strain AY146 were analyzed alongside the nuclear extracts to compare PGK1 levels and determine the extent of cytoplasmic contamination. Nuclear extracts were successfully isolated with minimal cytoplasmic contamination (<5%; compare lanes 1&3

and 2&4 in Figure 3.6B), shown in Figure 3.6A and B. Next, a time course was collected following rapamycin treatment to measure the nuclear level of TBP. An untreated 0 min time point sample was compared to samples obtained at 5, 10, and 15 minutes after rapamycin addition. The 5 minute sample is denoted as 4.5' or 6.5' depending on the replicate; filtering out rapamycin from cells took different amounts of time for the two samples. When these nuclear extracts were probed for TBP, depletion was not observed (Figure 3.6C, D). PGK1 levels in the nuclear extracts from the HHY154 strain time course were lower than those in the control HHY154 WCE samples, but there was still a very detectable level of PGK1, indicating that there were some cytoplasmic contaminants present (Figure 3.6E). When the nuclear depletion time course experiment was repeated, some depletion was observed, but not to the extent expected based on ChIP data and microscopy data (see below) (Figure 3.6G). There was not an observable difference in global protein levels between the samples as detected by Coomassie Blue staining, indicating that the inability to detect TBP depletion in the extracts was not due to differential sample loading (Figure 3.6F, I).

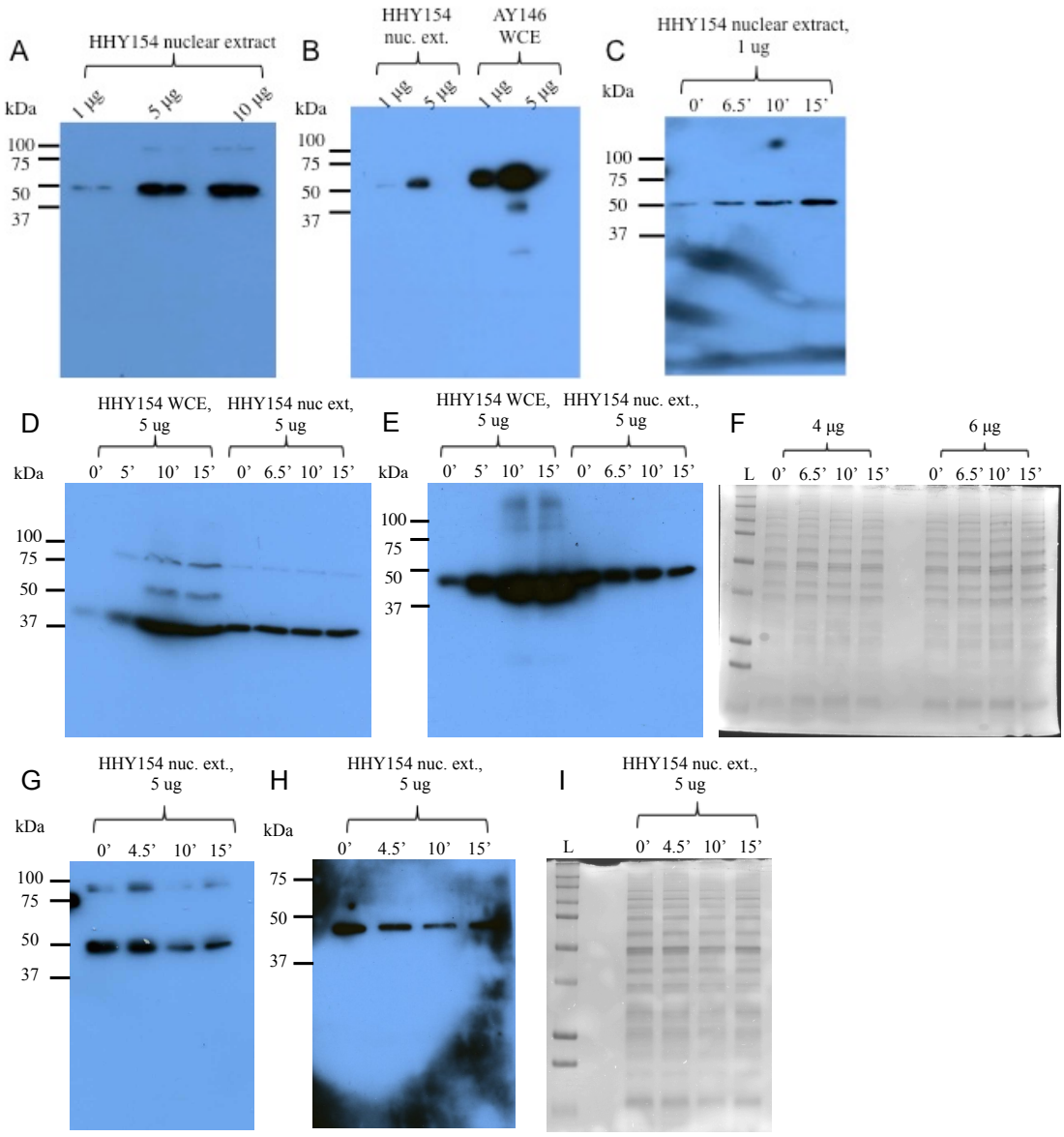


Figure 3.6. Western blots of nuclear depletion samples detecting protein of interest and cytoplasmic contamination. A) Western blot of TBP Anchor Away strain (HHY154) nuclear extracts probed with a TBP antibody and visualized with chemiluminescence. Increasing amounts of sample were loaded onto the SDS-PAGE gel after heating for 5 minutes. The expected TBP-FRB band is at 50 kDa. B) Similar to A, but only two amounts of nuclear extract (nuc. ext.) sample were analyzed. Matching amounts of control sample from the TBP CLK wild type strain AY146 whole cell extracts (WCE) were loaded for comparison. The blot was probed with PGK1 antibody and should detect a band at 45 kDa. C) Western blot of HHY154 time course samples; 1

μg of each sample loaded. The 0 minute sample was removed and the remaining culture was treated with 10 mg/ml rapamycin to induce nuclear depletion of TBP. The blot was incubated with a TBP antibody. D) Western blot of HHY154 whole cell extract and nuclear extract time course samples probed with a TBP antibody; 5 μg of each sample was loaded. E) The same as D, except a PGK1 antibody was used to test for cytoplasmic contamination. F) Commassie stain of the nuclear extract samples used in E and F; either 4 or 6 μg of each sample was loaded onto the gel. G) Similar to D, but the time course with rapamycin treatment was collected again and probed with a TBP-recognizing antibody on a western blot; 5 μg of each sample was loaded onto the gel. H) Identical to G, except samples were probed with a PGK1 antibody. I) Commassie stain of the time course samples used in G and H run on an SDS-PAGE denaturing gel. Molecular weights are denoted to the left of the western blots.

Nuclear depletion estimated via microscopy images

Since the nuclear depletion of TBP could not be accurately measured by isolating nuclear extracts, potentially due to some return to steady-state TBP levels once rapamycin was filtered from the cells, a different approach was taken using a TBP-FRB-GFP strain, HHY209, and its non-tagged parental strain, HHY168 (Haruki et al. 2008) with help from Tineke Lenstra in Dan Larson's lab at the NCI. Cultures for each strain were grown overnight. Prior to treatment with rapamycin, each culture was split in two; one culture was treated with rapamycin and the other was used as a non-treated control. Following removal of a zero time point control from the untreated cultures, rapamycin was added to the other cultures to induce nuclear depletion. Samples were taken at various time points up to 75 minutes, methanol fixed, and imaged using a light microscope (Figure 3.7). For each image, z-stacks were captured for DAPI, GFP, and transmitted light channels; multiple images were taken for each time point to have at least 1,000 cells to analyze.

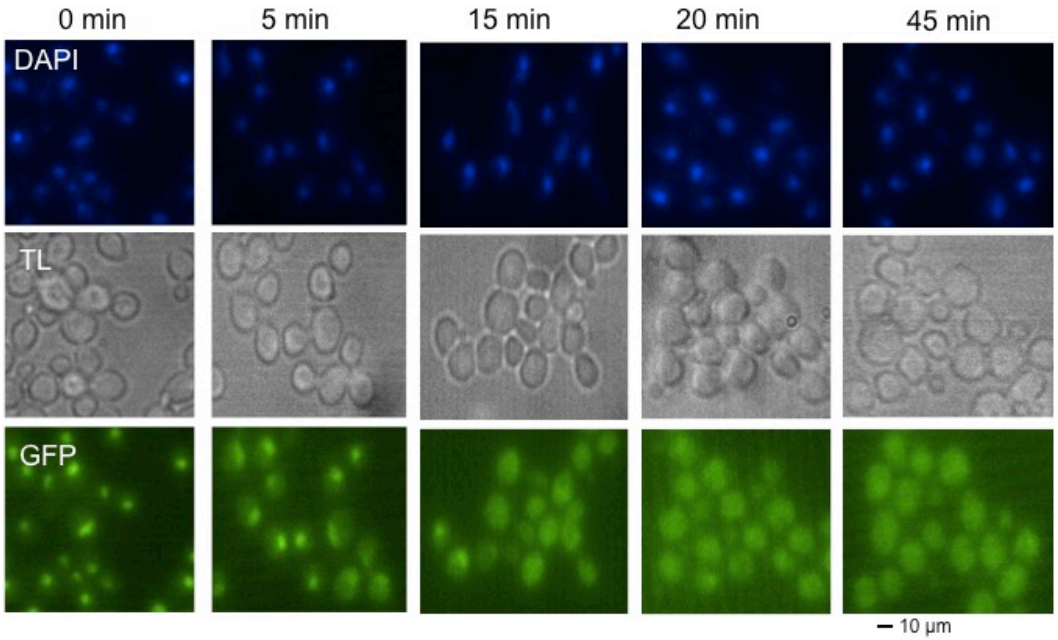


Figure 3.7. Images of methanol fixed cells in a rapamycin treatment time course. Rapamycin was added after the 0 minute sample was removed from culture. For each time point, images were taken for DAPI (blue), transmitted light (grey), and GFP (green) channels. The time course went out to 75 minutes with and without rapamycin (data not shown). The 75 minutes with rapamycin treatment images were very similar to the 45 and 60 minute images; the 75 minute time point without rapamycin looked identical to the zero time point. Punctate GFP indicated nuclear localization, while diffuse signal shows movement to the cytoplasm.

The nuclear and cytoplasmic intensities for each cell at each time point were calculated using MATLAB scripts provided by Tomáš Vičar with help from Kristýna Kupková. The cellular outlines were identified on the transmitted light images using a machine learning script and made into a mask (Figure 3.8A). The fluorescent DAPI channel was used to determine the nuclear outlines by semi-automated script (Figure3.8B); these outlines were made into the nuclear mask for each image. The cellular and nuclear outlines were identified well by the MATLAB script, but a manual editing step was added to eliminate any misdrawn outlines before the finalized masks

were generated. The final outlines were superimposed on the transmitted light images (Figure 3.8C), and then saved individually as mask images. The masks were overlaid onto the GFP z-slice with the highest DAPI intensity sum for each cell, which represents the middle of each cell and should offer the best representation of the nuclear TBP. The fluorescent intensity of GFP in the nucleus and cell was measured for each cell. Before the GFP intensity was measured, each z-stack used was normalized through mode subtraction to remove noise. Most of the GFP images had a crisscross pattern on the bottom, probably an artifact from the microscope laser. The regions affected by these marks were excluded from analysis. After all images had been measured, the GFP intensity of the non-tagged parental strain was calculated by averaging the nuclear and cellular intensities across all cells and time points. The average nuclear:cellular intensity was consistent across all parental time points, so averaging all time points to get a background autofluorescence nuclear and cytoplasmic value was not seen as an issue (Figure 3.9A). This calculated nuclear or cytoplasmic value was subtracted from each experimental cell to further normalize the data. The median nuclear:cytoplasmic TBP value for each time point was plotted as a boxplot; the best-fit exponential line through the medians gave a depletion time of 10.9 minutes (Figure 3.9B).

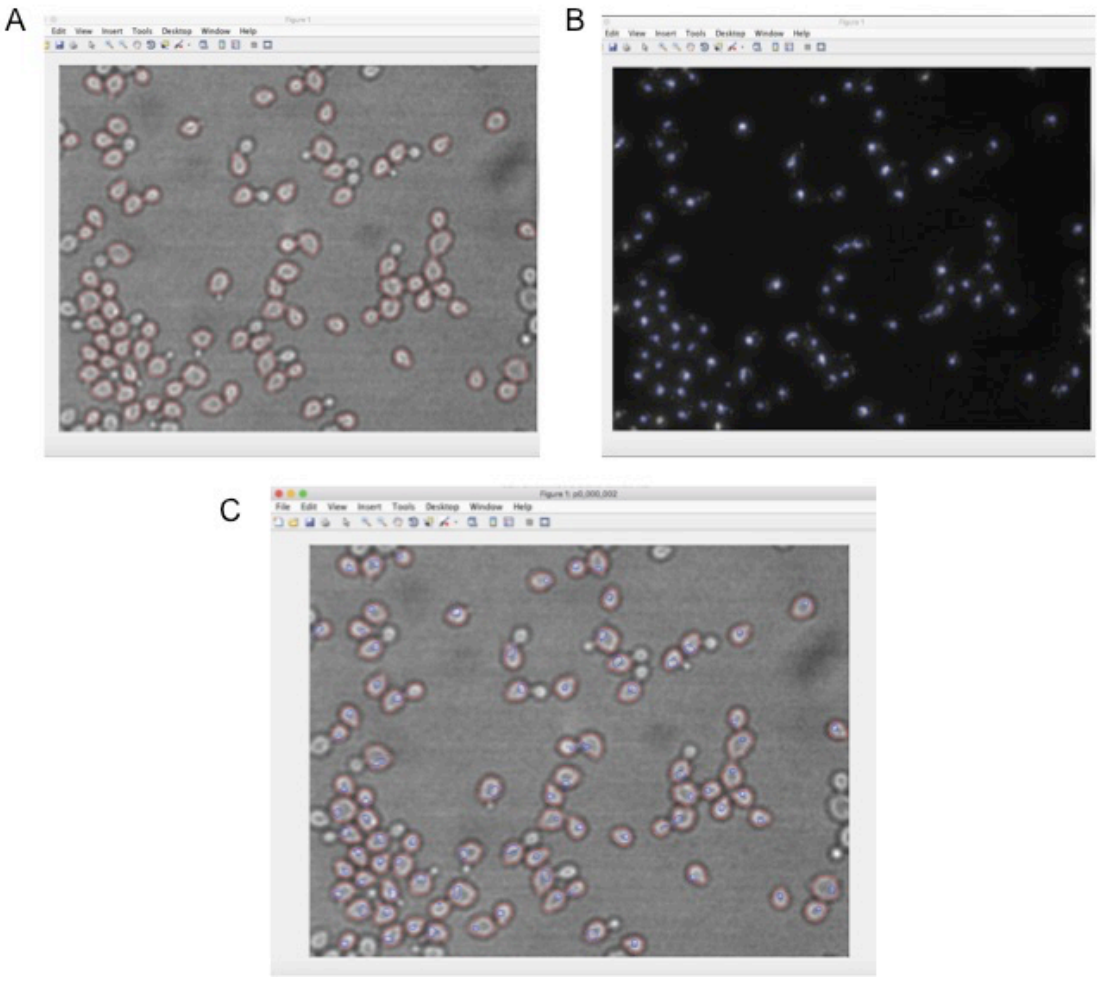


Figure 3.8. Outlines of nuclei and cells from two different channels were drawn to generate masks. A) Cellular outlines (red) on the transmitted light image used to generate them after machine-learning object identification. B) Nuclear outlines (blue) shown on the fluorescent DAPI image used to identify them. C) Cellular (blue) and nuclear (red) outlines after manual editing shown on the transmitted light image before finalized mask images were generated.

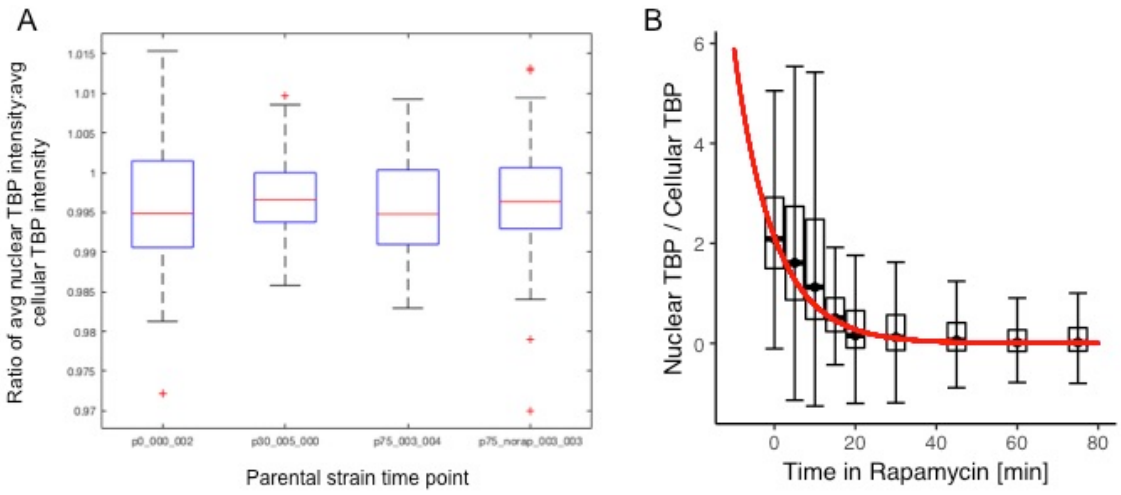


Figure 3.9. Boxplots of TBP nuclear depletion due to rapamycin addition. A) Boxplots of the median nuclear:cytoplasmic GFP intensity in the non-GFP-tagged parental Anchor Away strain at 0 minutes (no rapamycin), 30 minutes (rapamycin), 75 minutes (rapamycin), and 75 minutes (no rapamycin). The red line across each box represents the median intensity ratio, with the first and third quartiles shown by the bottom and top blue lines, respectively. The top and bottom shorter black lines represent the upper and lower values, with the red asterisks showing extreme upper or lower values. B) A boxplot of the rapamycin treated TBP-FRB-GFP Anchor Away strain. Rapamycin was added after the zero minute time point. The median GFP intensity ratio was calculated from the average of over 1,000 cells for each time point. The bottom and top lines of the box show the first and third quartiles, respectively; the "whiskered" shorted black lines show the upper and lower values. The data was modeled with an exponential fit through the medians to determine the depletion time (τ).

Anchor Away locus-specific data

Changes in TBP chromatin occupancy were measured by AA for several loci using the TBP AA strain, HYY154 (Haruki et al. 2008). Two biological replicate datasets were obtained and averaged for select loci to generate depletion curves (Figure 3.10). The majority of the data

sets are at or almost at zero percent depletion by the end of the 60-minute time course. The data were modeled as single-exponential decay processes and the residence times ($t_{½}$) and off-rates (k_d) were estimated from the fits (Table 3.3); these calculations are approximate since the depletion time is not taken into account and the modeling is a simplistic exponential fit. The approximate residence times ranged from ~6 to ~28 minutes, with *URA1* and *ACT1* on the faster side of around 7 minutes and *LOS1* on the slower side at ~28 minutes. No difference in apparent residence time was observed for consensus TATA-containing versus non-consensus promoters (Table 3.3) (Basehoar et al. 2004).

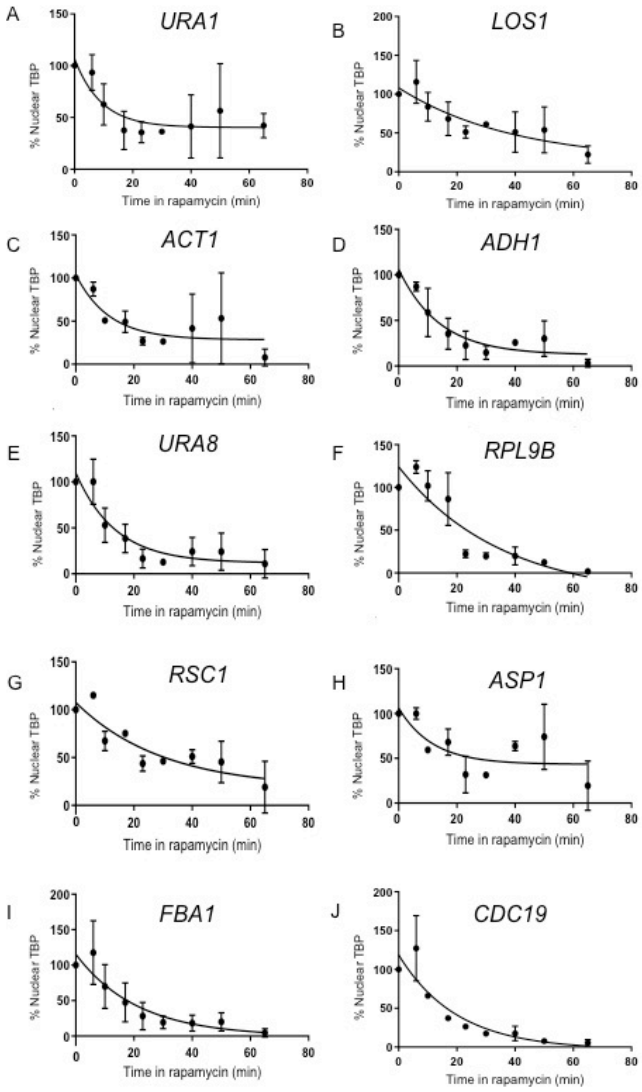


Figure 3.10. TBP Anchor Away data sets at select loci. A) Anchor Away data for TBP at *URA1*. The percentage of nuclear GFP is shown on the y-axis over the experimental time course (x-axis). A black circle indicates the average percent nuclear TBP of two replicates for each time point. The error bars show the standard deviation and the black line is a one-phase exponential decay curve fit. B) The same as A, but for TBP at *LOS1*. C) The same as A, but for TBP at *ACT1*. D) The same as A, but for TBP at *ADH1*. E) The same as A, but for TBP at *URA8*. F) The same as A, but for TBP at *RPL9B*. G) The same as A, but for TBP at *RSC1*. H) The same as A, but for TBP at *ASP1*. I) The same as A, but for TBP at *FBA1*. J) The same as A, but for TBP at *CDC19*.

Table 3.3. Apparent residence times and off-rates obtained from TBP AA data at select loci.

Locus	t½ (m)	k_d	Consensus TATA?
URA1	**6.4**	**0.11±0.05**	**Y**
LOS1	**27.4**	**0.025±0.022**	**N**
ASP1	**8.1**	**0.086±0.077**	**N**
RSC1	**20.7**	**0.034±0.026**	**N**
ACT1	**7.9**	**0.088±0.047**	**N**
RPL9B	**24.5**	**0.028±0.022**	**N**
URA8	**9.4**	**0.074±0.028**	**Y**
ADH1	**9.7**	**0.071±0.023**	**Y**
FBA1	**14.7**	**0.047±0.021**	**Y**
CDC19	**13.8**	**0.050±0.025**	**Y**

Competition ChIP strain construction

The Competition ChIP strain for TFIIB was constructed by first tagging one *SUA7* allele with HA and placing it under the control of the *PGAL1* promoter, a technique developed by the Pringle lab and utilized by the Timmers lab to construct a TBP Competition ChIP strain (Longtine et al. 1998, van Werven et al. 2009). A cassette with *PGAL1* and three HA tags, along with a *HIS3* auxotrophic marker was amplified from the pFA6-HisMX6-PGAL1-3HA plasmid (Figure 3.11A). This cassette was then transformed into the diploid wild type YPH501 strain, selected for on SC-HIS plates, and verified by PCR using two sets of primers (Figure 3.11B). Three candidates were found with the correct insert based on two sets of primers; one candidate is shown in Figure 3.11B; the other two candidates are from a separate analysis and are not shown.

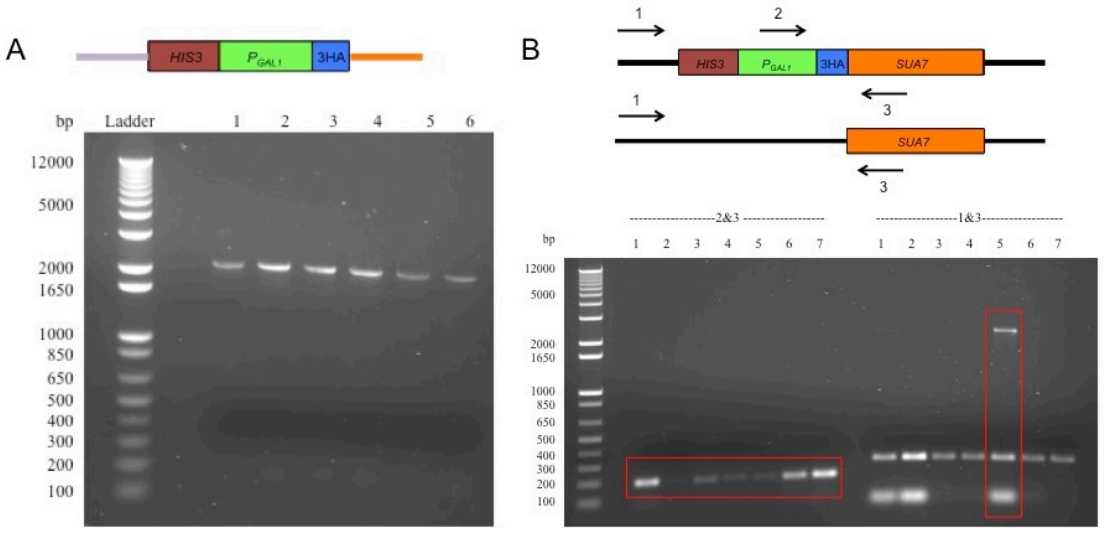


Figure 3.11 Amplification and verification of the HIS3-PGAL1-3HA cassette before and after insertion into diploid yeast strain YPH501. A) PCR amplification of the HIS3-PGAL1-3HA cassette (top) from the pFA6-HisMX6-PGAL1-3HA plasmid. Six identical PCR products were run on a 1% agarose gel with ethidium bromide staining and visualized with a GelDoc. A band at ~2100 bp was expected. The base pair length is denoted to the left of the 1 kb plus DNA ladder (NEB). B) PCR amplification of potential candidates following transformation of the

cassette into the diploid strain run on a 1% agarose gel with ethidium bromide. Two sets of primers were used; 1 and 3 are just outside the insertion sites and yield the WT and insert product, and 2 and 3 are inside the insertion site and only yield one insert product. Red boxes indicate candidates with the correct band sizes for each primer set. Schematics of the cassette or strain expected are shown above each gel image.

The next step was to check the newly made strain for galactose-induced expression of the HA-tagged allele. Whole cell extracts were collected for one of the candidate strains either grown in YEP media with raffinose followed by galactose induction for one hour, or YPD. Expression was measured by western blot with an HA-recognizing antibody (12ca5, Roche) (Figure 3.12). The molecular weight of Sua7-3HA is about 38 kDa; the correct band was observed in samples grown in galactose, but not raffinose or glucose, as expected.

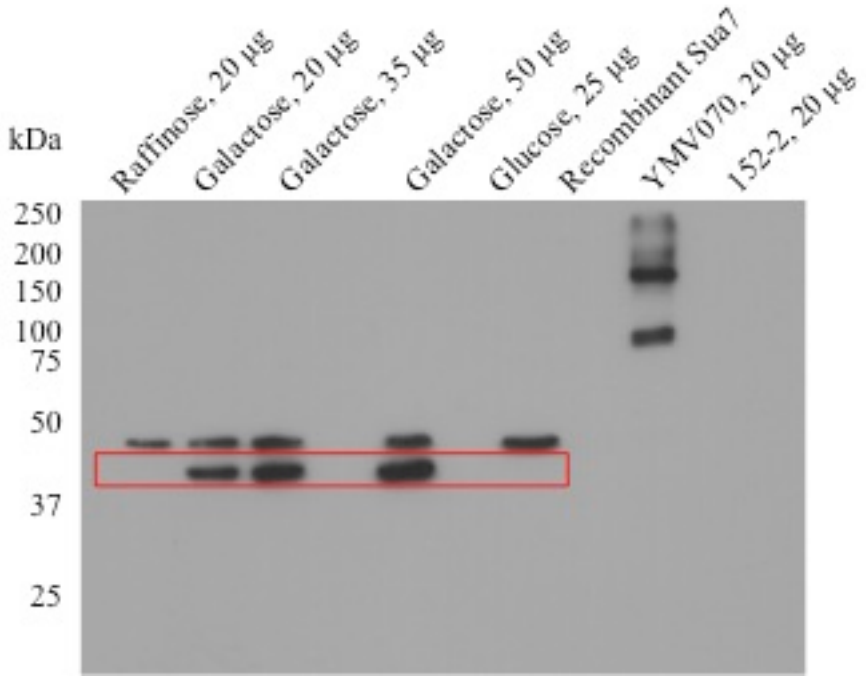


Figure 3.12. Western blot of diploid YPH501 strain with one allele of *SUA7* under the control of the *PGAL1* promoter labeled with a 3xHA tag. Samples were grown in rich media

with either raffinose, galactose, or glucose as the sugar source. The band at ~38 kDa, indicated by the red box, is the desired band showing induced Sua7-3HA. Recombinant Sua7 is purified protein not tagged with HA; YMV070 is an Spt16-HA strain (True et al. 2016) used as a positive control and 152-2 is a strain lacking an HA tag and was used as a negative control. Molecular weight is denoted to the left in kDa.

Following confirmation of the tagging of one *SUA7* allele, a tagged haploid segregant was obtained by sporulation and dissection (EH005). A selection of the growth test results obtained for the resulting tetrads is shown in Figure 3.13. Haploids with the correct tag grew on YEP+galactose plates and SC-HIS+galactose plates, but not on YPD plates. Once candidate haploid strains were identified, their mating type was determined with two tester strains, one MATalpha, and the other MATa, described earlier in Materials & Methods.

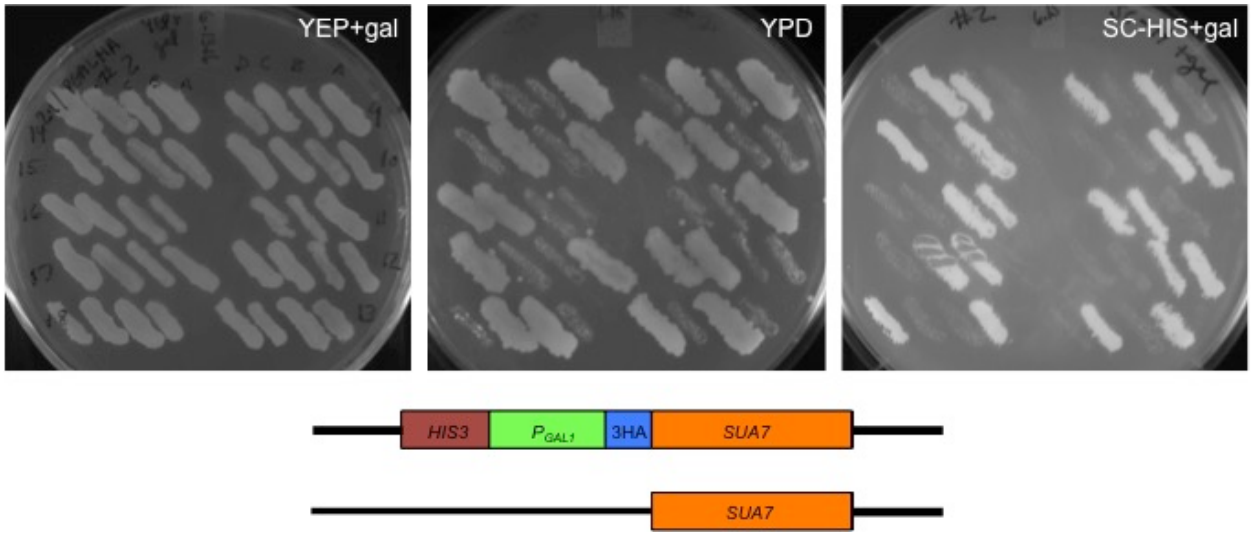


Figure 3.13. Growth tests for tetrads dissected from sporulated EH005 (pGAL1-HA-SUA7/SUA7). A selection of plates used to check for tetrad growth. From left to right, tetrads were grown on YEP+2% galactose, YPD, and SC-HIS+2% galactose plates. Each group of

diagonal streaks (3-4 per group, two groups per row) represents a dissected tetrad; 10 dissected tetrads are shown on each plate. Candidate tagged haploid strains were identified from this screen; the schematic below the images shows the new haploid strain.

The presence of the HA tag in the new candidate haploid strains was confirmed by western blotting and an immunoprecipitation using an HA-recognizing antibody (12ca5, Roche) (Figure 3.14). The HA tag was detected in all candidates tested by western. Compared to the control YPH499-1 haploid wild type strain, the HA-tagged candidate 2-17B had a much higher signal (~10x) at the *URA1* promoter; the *GAL1/10 ORF* signal was higher in 2-17B compared to that of 499-1, but it was ~3x lower than the *URA1* signal in the tagged strain. This IP was not repeated with another replicate, but based on the single replicate, HA can be detected by IP in this strain.

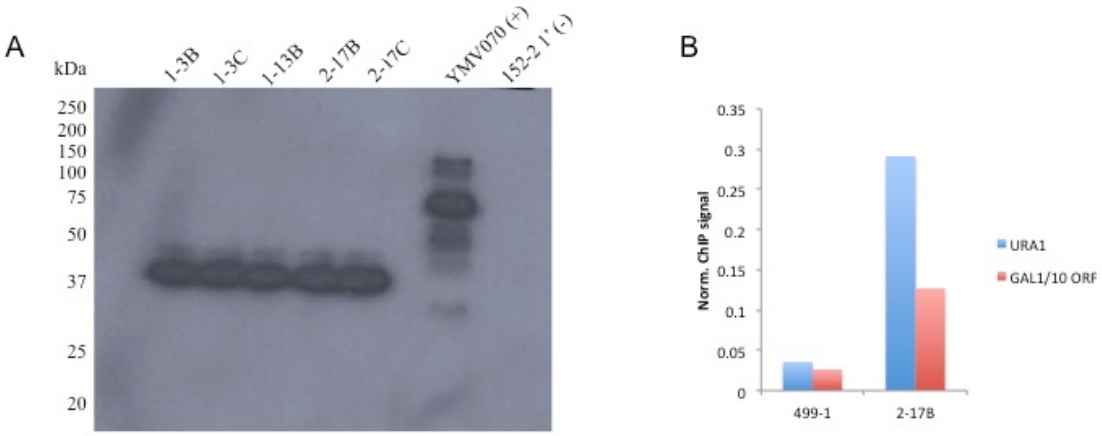


Figure 3.14. Verification of the HA tag in haploid candidate strains by western blotting and immunoprecipitation. A) Western blot of potential tagged haploid strains. Whole cell extracts for each aample were run on an SDS-PAGE gel and transferred onto a PVDV membrane before incubation with an HA-recognizing antibody. Haploid candidates are 1-3B/C, 1-13B, and 2-17B/C; YMV070 is an Spt16-HA strain for a positive control; 152-2 is a TFIIE CLK

overexpression strain that is not tagged with HA used as a negative control. A band is expected around 38 kDa for the Sua7-3HA protein. Molecular weight is denoted to the left of the western image. B) An IP was done with one sample to check for detection of the HA tag. The haploid wild type strain YPH499 was used as a control. The normalized ChIP signal was compared at the *URA1* promoter and *GAL1/10* open reading frame for both strains.

An attempt was made to tag the second *SUA7* allele in the HA-tagged diploid strain; however, this was ultimately unsuccessful. First, the pCORE approach was used (Storici and Resnick 2006). A *KANMX4-KIURA3* cassette amplified from the pCORE plasmid with 45-base pair homologous regions on either side of the cassette was integrated directly upstream of the *SUA7* start site. A short cassette containing the Avitag then replaced this insert. In the already HA-tagged diploid, the Avitag could not be successfully integrated for unknown reasons (data not shown). The pCORE method was also tried in a wild type YPH501 strain that was not tagged; again the pCORE replacement with Avitag was unsuccessful. An altered pCORE technique that supposedly works better for diploids was next tried (Storici and Resnick 2006). A modified CORE cassette with a galactose-inducible SceI enzyme and SceI cut site was amplified from the pGSKU plasmid and inserted upstream of *SUA7*. This cassette was designed to improve insertion efficiency of the Avitag at the correct site by making a double stranded cut at the SceI cute site (Storici and Resnick 2006). Again, the Avitag could not be successfully detected after CORE cassette replacement. The next attempted technique used untagged YPH501 and integrated a cassette from the pUG6-Myc-N-Avitag plasmid, which has a *KanMX* marker flanked on each side by a loxP site followed by an N-terminal Avitag (Werven and Timmers 2006), directly upstream of the *SUA7* start site. Once the cassette was integrated, a galactose-controlled Cre recombinase-containing plasmid (pSH47 (Güldener et al. 1996)) was transformed into the labeled strain. Cells were grown in galactose, activating Cre recombinase and excising *KANMX* by cutting both loxP

sites. The Avitag was successfully integrated by this approach, but biotinylated Sua7 could not be detected by Streptavidin IP followed by real time PCR or western blot (data not shown). Additionally, Sua7 levels were decreased compared to YPH501 when whole cell extracts were run on a western blot (data not shown). Ultimately, the second *SUA7* allele was successfully tagged by using a similar method, but with 9xMyc tags instead of the Avitag; the results are described below.

To tag the second *SUA7* allele with a 9xMyc tag, a cassette containing a KanMX marker flanked on each side by a *loxP* cut site followed by nine N-terminal Myc tags was amplified from the pOM20 plasmid (Gauss et al. 2005). The amplified cassette was 2083 base pairs in length (Figure 3.15A) and all candidates tested successfully integrated the tag-containing fragment (Figure 3.15B).

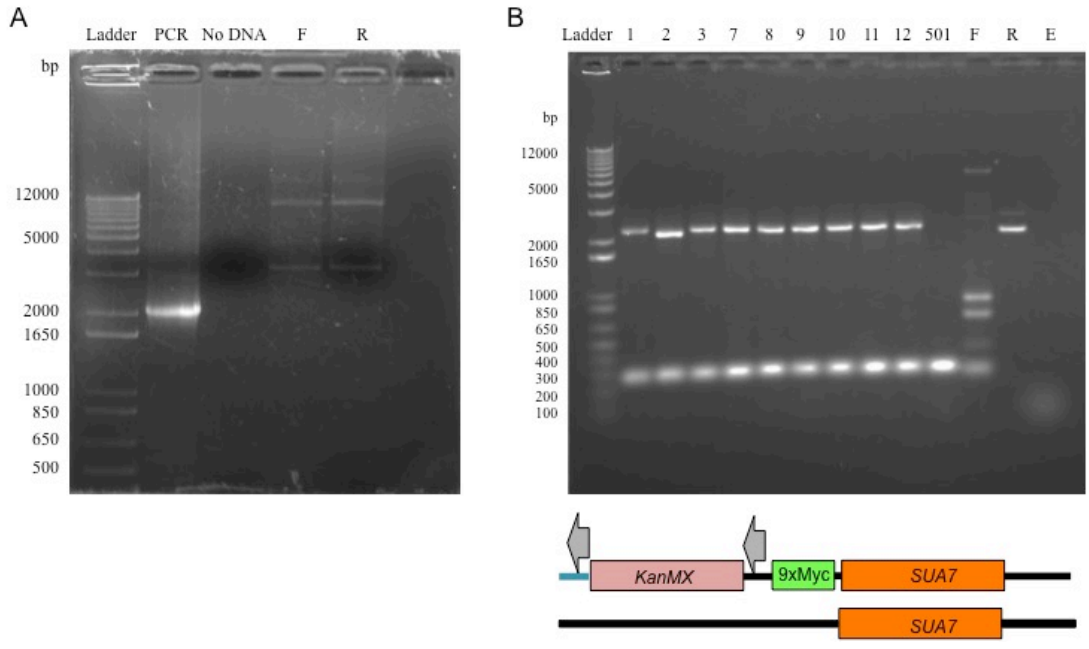


Figure 3.15. PCR amplification of Myc-tagged cassette from pOM20 and integration into YPH501. A) Agarose gel with ethidium bromide staining of the concentrated PCR amplified cassette from pOM20 (PCR). A no DNA negative control sample (no DNA) and forward (F) or reverse (R) primer only samples were also run on the gel. The expected 2083 base pair (bp) band

in the PCR land was excised and transformed into YPH501. B) PCR verification using primer set 1&3 from Figure 3.11B of potential candidates following transformation of the pOM20 amplified cassette into YPH501; candidates are numbered 1-3 and 7-12. PCR for YPH501 was done as a control, as well as PCR with only the forward (F) or reverse (R) primer and no DNA (E). The expected band size for the insert was 2268 base pairs and 284 base pairs for the wild type. Base pair lengths are shown to the left of both gel images. The newly tagged diploid strain with marker still present is represented as by the diagram under the gel image.

Following successful integration of the *KANMX-9xMyc* cassette, a plasmid containing the Cre recombinase was transformed into a newly tagged diploid strain and confirmed on an SC-URA plate. Cre recombinase is required to cut the loxP sites and excise *kanMX* (Sternberg and Hamilton 1981, Sauer and Henderson 1988, Orban et al. 1992, Gu et al. 1993). The Cre recombinase was then activated by growth in galactose. Dilutions of the galactose culture were plated onto YPD plates and replica plated on YPD+G418 plates. Comparison of the two plates indicated which candidates had lost the *KANMX* cassette. Overall, there was 51% efficiency for the Cre recombinase-induced knockout of *KANMX*. A selection of knockout candidates was checked by PCR and all had an insert of expected length (Figure 3.16).

Detection of the Myc tag in the diploid was confirmed by western blotting and visualized in two candidates using the 9E10 Myc antibody (Abcam) (Figure 3.17). Correct insertion and sequence of the 9xMyc cassette was verified by Sanger sequencing in one candidate strain. After tag confirmation, the strain was then sporulated and dissected after tetrad formation. Since there was no longer a marker tracking the presence of the 9xMyc tag, PCR was done with the same primer set used in the last two figures, as well as the original primer set that amplified the cassette from pOM20, to verify the presence or absence of the tag (Figure 3.18A, B). Tetrads 1, 3, 6, 10, and 14 segregated 2:2 for wild type:Myc tag, as expected. However, tetrads 2 and 4 were from a

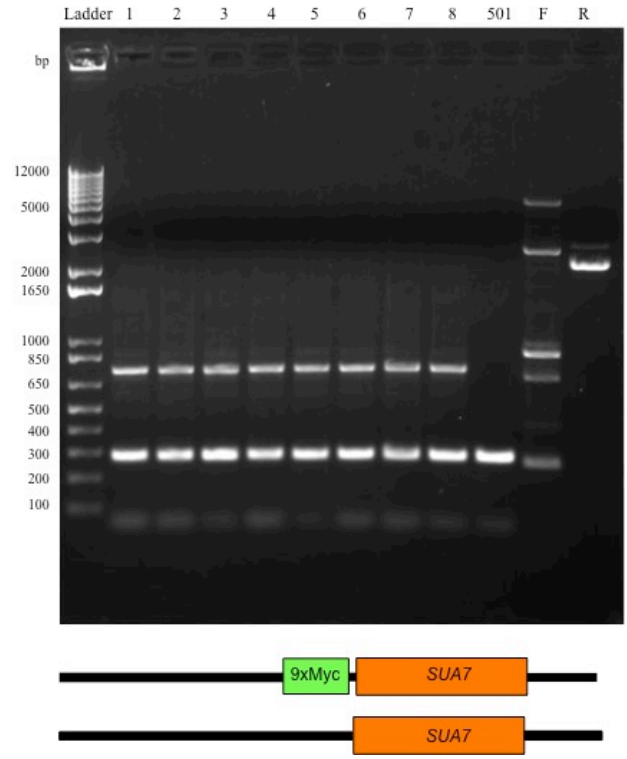


Figure 3.16. Agarose gel electrophoresis to check for knockout of the *KANMX* marker from a 9xMyc-tagged diploid strain. Eight potential candidates were checked by PCR with the same primer set used in Figure 3.15B. Two bands were expected at lengths of 761 for the knockout insert and 284 for the wild type. All candidates had both expected bands. PCR of the wild type diploid YPH501 and candidate 1 with only the forward (F) or reverse (R) primer were run as controls. The base pair length is denoted to the left of the image. A schematic of the resulting tagged strain is shown under the gel image.

different dissection and did not segregate 2:2, and thus were discarded as candidates. In all lanes showing the presence of the Myc tag in Figure 3.18A, there was a faint band at the wild type position. This was not contamination since this band wasn't present in the sample without DNA, but there may have been some low level wild type DNA still present somehow in these diploid samples.

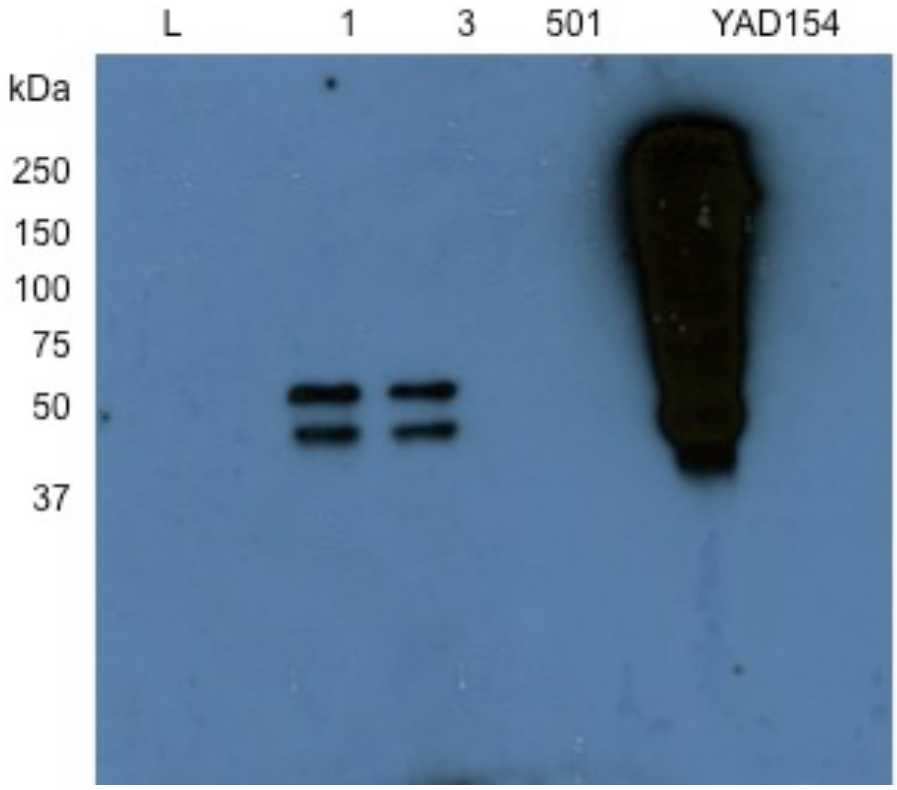


Figure 3.17. Western blot of extract from Myc-tagged diploid strain. Whole cell extracts of two candidate diploid strains, an untagged negative control (501), and a tagged positive control (YAD154) were run on an SDS-PAGE gel and probed with a Myc-recognizing antibody. A band was expected around 50 kDa for the 9xMyc-Sua7; Sua7 normally runs as a doublet with the top band representing the correct mass. Molecular weight is marked on the left side of the blot.

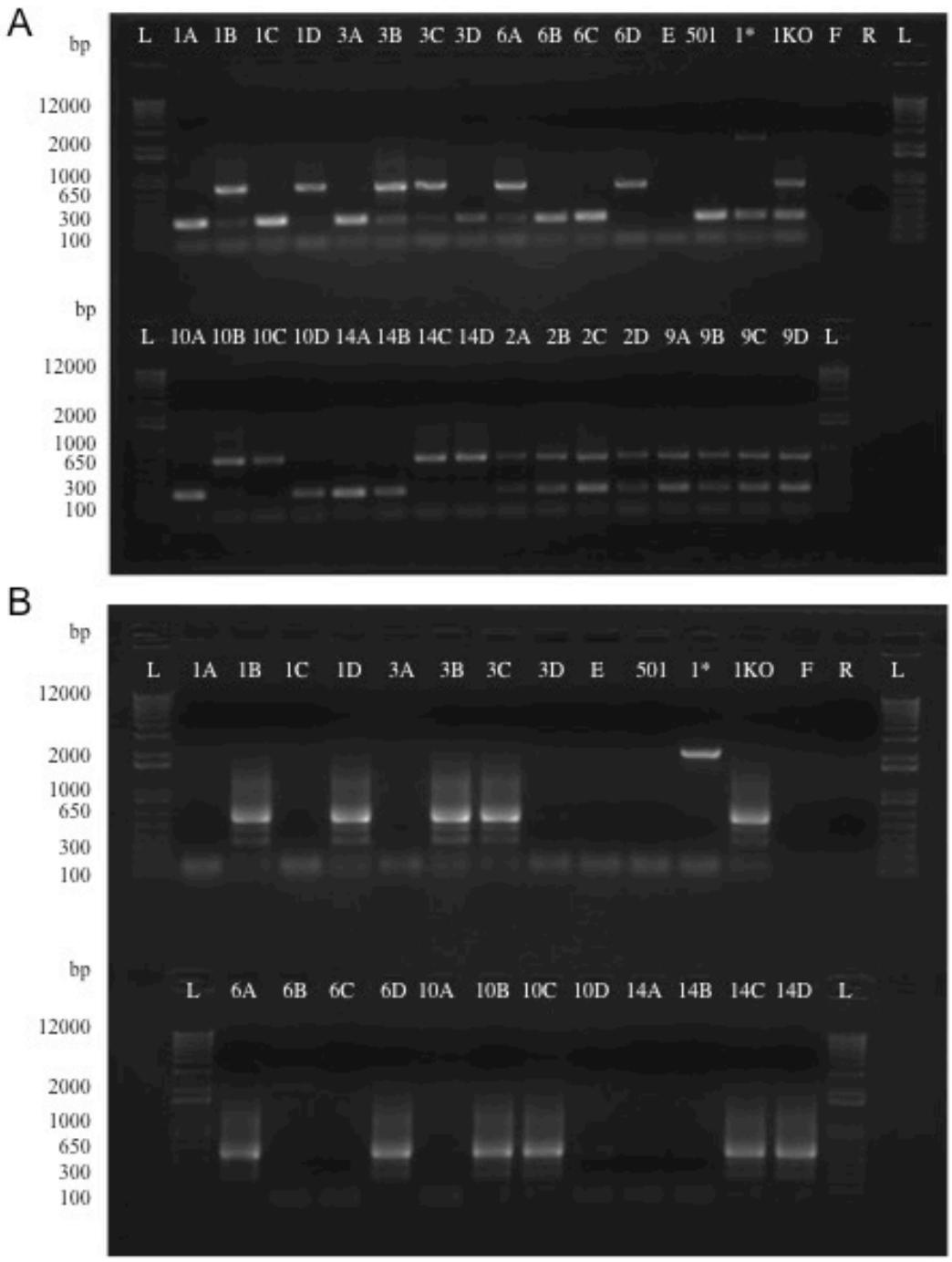


Figure 3.18. Gel electrophoresis of PCR reactions to check for presence of wild type or 9xMyc-tagged *SUA7* allele in dissected tetrads. A) PCR was performed using cells from several "full" tetrads using primers just outside the insert region (used in the last two figures; primers 1&3 in Figure 3.11). A band at 761 base pairs (bp) indicates a 9xMyc tag; a band at 284

bp is a wild type allele. E is a no DNA control; 1* is the 9xMyc-labeled diploid before *KANMX* was knocked out and yields band sizes of 2268 (insert) and 284 bp (WT); 1 KO is the 9xMyc-labeled diploid after *KANMX* was removed by Cre recombinase and yields band sizes of 761 and 284 bp; forward primer alone (F) and reverse primer alone (R) were also used as controls; L is the 1 kb plus DNA ladder. B) Similar to A, but using primers that should only detect the insert and not the wild type. A band at 576 base bp indicates a 9xMyc tag; no visible band is a wild type allele. 1* yields a 2083 bp band and 1 KO gives a 576 bp. Tetrads 2 and 4 were excluded from this gel because they didn't have a 2:2 segregation of wild type to Myc tag in the first PCR reaction in A.

After the 9xMyc tag was identified in haploid candidates, it was also successfully verified by western blotting for one candidate tetrad (Figure 3.19). Two of the tetrads showed a Myc tagged protein and the other two did not, indicating a wild type genotype, as expected. Next, a mating test was performed for tetrads 1, 3, 6, 10, and 14 using the 14a and 17α tester strains; haploids from all tetrads segregated 2:2 for MATa:MATalpha (data not shown). A 9xMyc-SUA7 haploid strain was crossed with a HIS3-PGAL1-3HA-SUA7 strain of the opposite mating type, yielding the diploid Competition ChIP strain. The mating type of several candidates was checked using the tester strains mentioned above as well as two new diploid strains; neither new diploid strain grew when crossed with either mating tester, as expected (Figure 3.20) (Strathern et al. 1979, Jensen et al. 1983, Goutte and Johnson 1988, Herskowitz 1989, Li et al. 1995, Johnson et al. 1998, Kim et al. 2000).

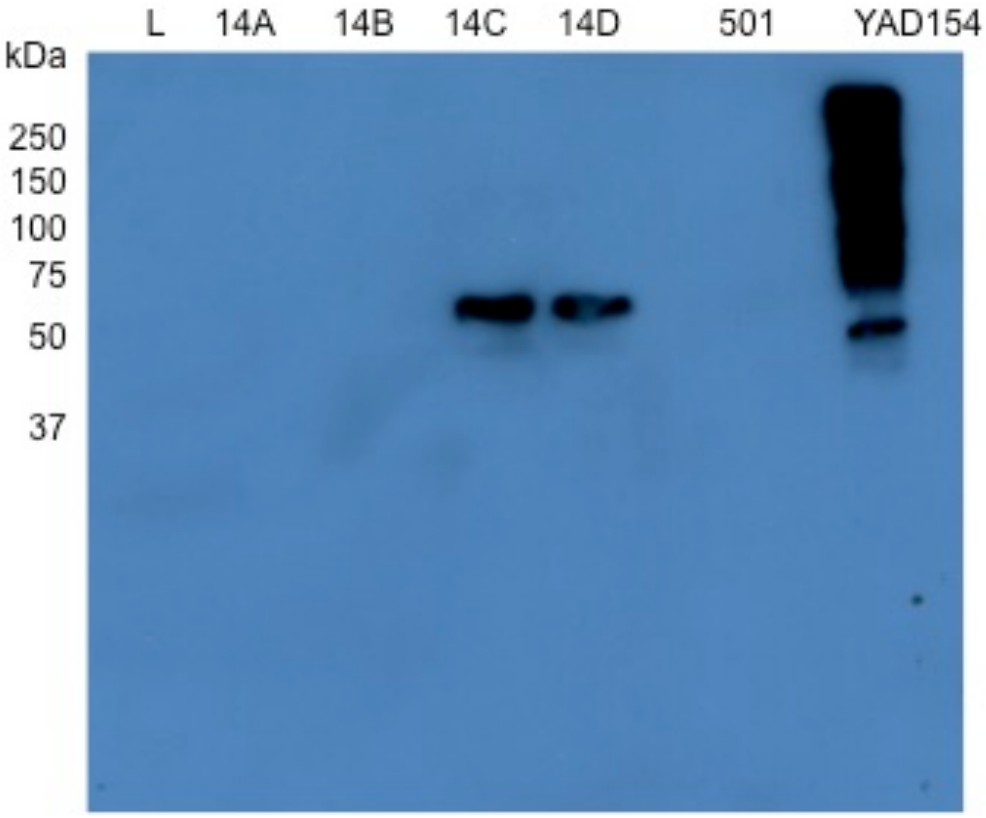


Figure 3.19. Western blot to check for Myc tagged Sua7 expression in haploid tetrad candidate samples. Whole cell extracts of four haploid strains from one tetrad, an untagged negative control (501), and a tagged positive control (YAD154) were run on an SDS-PAGE gel and probed with a Myc-recognizing antibody. A band was expected around 50 kDa for the 9xMyc-Sua7. Molecular weight is marked on the left side of the blot.

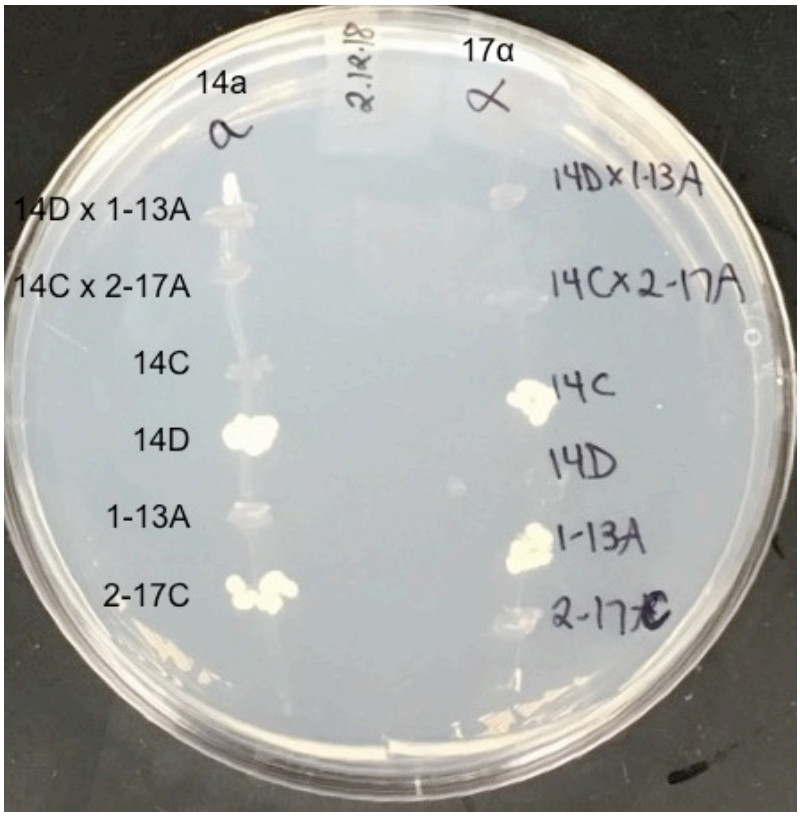


Figure 3.20. Mating type test for two differentially tagged Sua7 Competition ChIP strains. The 14a tester strain was streaked in a vertical line on the left side of the SD plate and 17α on the right. The new diploid strains, 14D x 1-13A and 14C x 2-17A, were crosshatched against the tester strains, followed by the haploids 14C, 14D, 1-13A, and 2-17C. A white outcrop of yeast growth indicates the strain is the opposite mating type of the tester it is crosshatched with.

To verify that both tag inserts were present in the two new diploid strains, PCR was performed with the same primers used in the previous figures (3.15B and 3.16) and that lay just outside the insertion site for both tags. Both expected bands were observed in both samples, as well as a faint wild type background band (Figure 3.21).

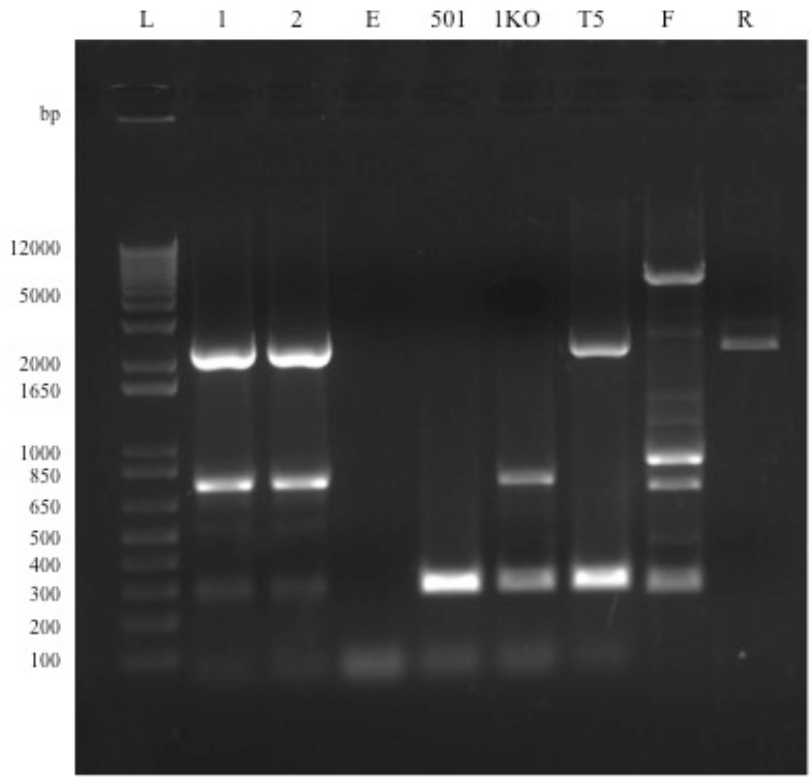


Figure 3.21. Gel electrophoresis of PCR to check for correct insert bands in the newly made TFIIB Competition ChIP strains (1, 2). Bands were expected at 761 bp for the 9xMyc tag and 2187 bp for the HIS3-PGAL1-3HA tag. E is a no DNA control; 501 is the wild type diploid strain and a 284 bp band is expected; 1KO is the 9xMyc-labeled diploid and bands at 761 and 284 bp are expected; T5 is the HA-labeled diploid and bands at 2187 and 284 bp are expected; forward primer (F) and reverse primer (R) only controls were run as well. Base pair lengths are denoted to the left of the gel image.

An estimate of the rate of competitor induction is critical for CC as the induction time is taken into account when fitting each locus to generate the residence time (Zaidi et al. 2017). Western blots were performed using samples obtained over a 90 minute time course after galactose addition and are shown in Figures 3.22A and C. Unexpectedly, the Sua7 Competition

ChIP strain had a slower induction time than the induction rate for TBP using a similar galactose-inducible promoter (van Werven et al. 2009) and it was not clear if the saturation level was reached. Therefore, induction was followed for a longer time course of 180 minutes (Figure 3.22B, D). Samples were run in duplicate; one blot was probed with an α-HA antibody and the other was probed with an α-Myc antibody (9E10, Abcam). The 9xMyc-Sua7 levels were relatively unchanged over the time course, as expected (Figure 3.22C, D; quantification not shown). Sua7-3HA was detectable at around 25 minutes after galactose addition and reached saturation by ~90 minutes (Figures 3.22A, B; Figure 3.23A); a 60 minute induction time was calculated from a Boltzmann fit (used to fit a sigmoidal curve) of the induction curve (Figure 3.23A, B).

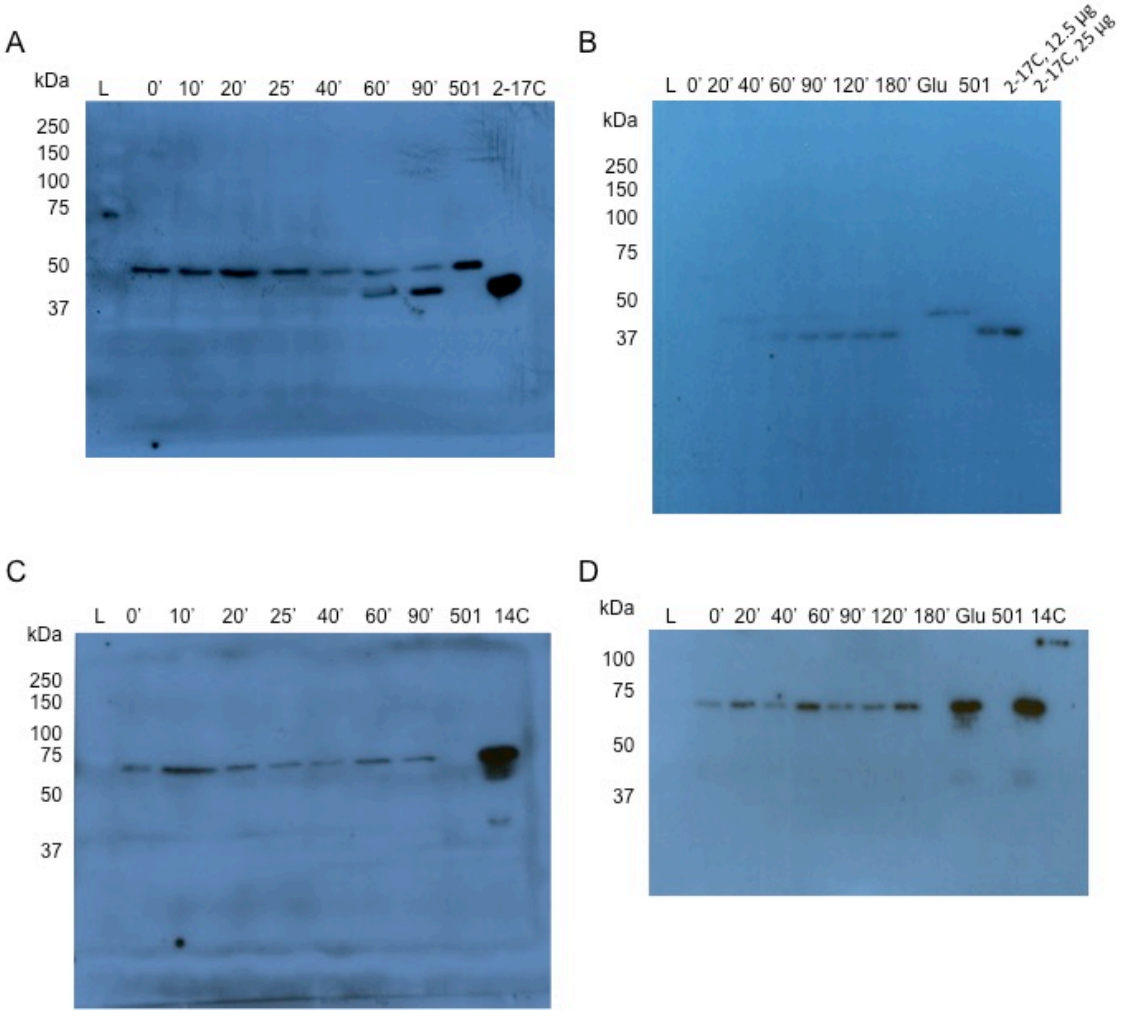


Figure 3.22. Western blots for two induction curves of a 9xMyc-SUA7/HIS3-PGAL1-3HA-SUA7 Competition ChIP strain. A) A time course following galactose induction of 3HA-Sua7 at time zero. Whole cell extracts for each time point and a positive (2-17C) and negative (501) control were run on an SDS-PAGE gel and transferred to PVDV membrane. A 12ca5 antibody to detect HA was used. 3HA-Sua7 runs at ~38 kDa. 501 is the YPH501 wild type diploid and 2-17C is a 3HA-Sua7 haploid. B) Similar to A, but for a longer induction time course. Two different amounts of 2-17C whole cell extract were loaded as a positive control. C) Identical to A, except a 9E10 Myc-detecting antibody was used and the 9xMyc-SUA7 haploid 14C was the positive control. D) Similar to D, but for the longer induction time course.

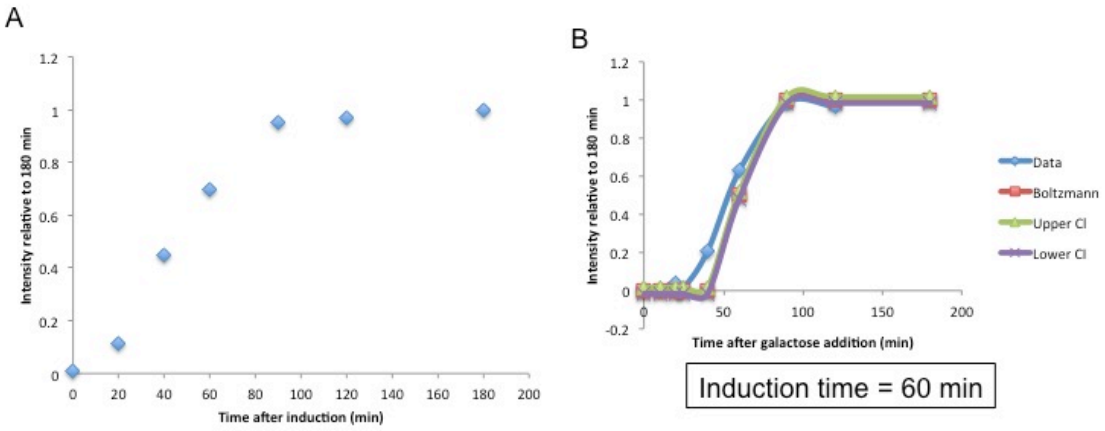


Figure 3.23. Time course for induction of Sua7 competitor in the Sua7 Competition ChIP strain. A) The induction curve for one replicate of the 180-minute time course. The intensity of each 3HA-Sua7 band from the western blot in Figure 3.22B was measured using ImageJ. Each value was normalized to the 180-minute sample intensity and plotted as a function of induction time in minutes. B) Boltzmann fit, used to fit S-shaped functions, of the induction curve from an average of three independent replicates (two replicates were 90 minute time courses). The blue line represents the data, the red line is the Boltzmann fit, and the green and purple lines are the

upper and lower boundaries, respectively. An induction time of 60 minutes was calculated from the fit in Excel.

The above results suggested that the strain used for construction of the Sua7 CC strain may unfortunately have a kinetic defect in galactose-induced growth. To test this, a spot test assay was performed to compare the growth of the Sua7 CC strain to the original diploid, YPH501. A different wild type strain, BY4741, was included for comparison as this is the strain background of the TBP CC strain. All strains grew similarly on glucose plates, although CC strain #2 grew slightly slower than the other three strains. However, there was a growth deficiency in YPH501 and both Competition ChIP strains compared to BY4741 on galactose medium (Figure 3.24) due to a *gal2* mutation, which is discussed in the next section.

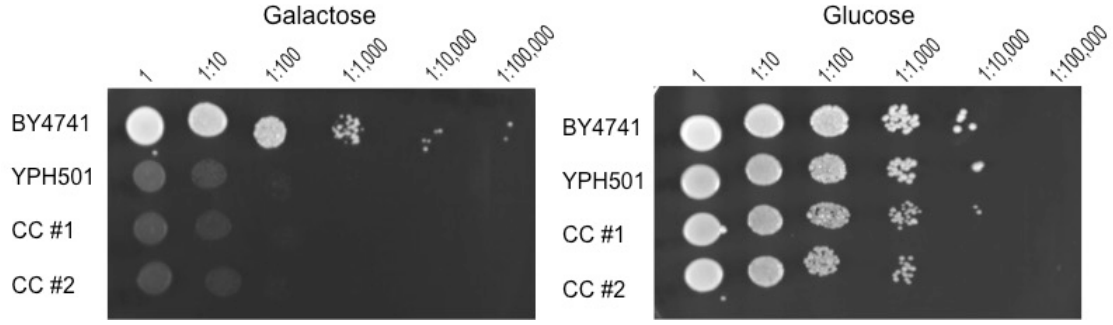


Figure 3.24. Spot tests on galactose and glucose plates for two wild type and two Sua7 Competition ChIP strains. Serial dilutions for each strain were made from 1 to 1:10,000, plated on two different sugar sources, and grown for 2 days at 30°C. CC is the Sua7 Competition ChIP strain.

Discussion

The Anchor Away assay

Use of the AA assay for rigorous kinetic analysis will require additional work, including a number of control experiments that will be required to properly interpret the data. Once key parameter is to determine the factor depletion rate (TBP in our case). Without knowing the rate of rapamycin-mediated depletion of TBP, it cannot be determined whether the measured off-rate is determined by the protein depletion from the nucleus or dictated by the rate of the protein's dissociation from its binding site, followed by export out of the nucleus. By measuring the global depletion rate of TBP from the nucleus, a limit can be established for the dynamic range of the assay. In our experiments, the depletion time of TBP was ~11 minutes, so residence times less than 11 minutes cannot be accurately measured because the protein cannot be moved from the nucleus that quickly. However, 11 minutes does not need to be the ultimate limit of the assay, as it's possible to model the nuclear depletion rate and therefore extend the kinetic range of the assay for measuring more rapid dynamics, similar conceptually to what was done with the induction rate to extend the dynamic range of the CC assay where the residence time can be extracted as the difference between the induction and experimentally measured curves (Zaidi et al. 2017). An attempt to measure the depletion of TBP from the nucleus was made by isolating nuclear extracts from cells treated with rapamycin for various periods of time. Unfortunately, depletion was not observed. We speculate that since the procedure to isolate nuclear extracts takes time, including generating spheroplasts and several washing and incubation steps in the absence of rapamycin, it's likely that the nuclear exported TBP re-normalized and returned to equilibrium before the extracts were obtained.

In contrast, the analysis of the TBP nuclear depletion rate was possible by microscopy using GFP-tagged TBP. For each image, a DAPI, GFP, and transmitted light channel were used. There are pretty standard automated ways to identify nuclei with the DAPI stain built into ImageJ (NIH), but using the transmitted light to find the cell outlines in an automated way proved a much

more difficult task because the difference in intensity between the cell wall and inside/outside of the cell is small. This difference is much larger and therefore boundaries are much easier to detect if a fluorescent protein is used. If a cell wall protein, such as Cwp1 or Cwp2 (Ram et al. 1998), was tagged with a fluorescent tag other than GFP, such as mCherry or mRuby 2 (Lee et al. 2013), detection of the cellular outline would have been significantly easier. Future strains imaged for nuclear depletion would benefit from the MATLAB pipeline developed by T. Vičar and K. Kupková and would include a second fluorescent tag for cell outline detection. The MATLAB script to identify the outline of cells employed machine learning to train the program to correctly identify these objects. A manual editing step was incorporated for final approval of defined cells, but the program did an astonishingly good job at correctly outlining cells. I attempted a segmentation approach using CellProfiler to identify cells, but it only succeeded about half the time (data not shown). Manual outlines in ImageJ to generate masks were also made, but this approach was extremely time intensive and measurements would have taken many months longer to complete. Some of the manually drawn masks were used to train the machine-learning algorithm to identify outlines. In the entire analysis, the most difficult part was the cellular segmentation. Once the cellular masks were made, the nuclear masks were easily made based on the DAPI signal and then GFP fluorescence intensity could be easily quantified.

The background subtraction was another important aspect of the analysis of TBP depletion using microscopy since a confocal microscope was used and therefore light could reflect from one z-stack into the surrounding stacks (Nomarski and Weill 1955, Barone-Nugent et al. 2002, Cody et al. 2005). The mode of each full stack (11 slices) was subtracted from each pixel in the stack to eliminate background. To analyze the GFP images, the z-stack with the brightest DAPI intensity was chosen for each cell in each image; the thinking was that this stack would represent the middle of the cell and would be the best place to measure TBP-GFP. Following average GFP intensity quantification of each cell and nucleus, the average GFP intensity across all cells in all time points for the parental untagged strain was calculated to

subtract the autofluorescence of the yeast cells (Billinton and Knight 2001). This script used for this analysis will be extremely useful in the future for identifying cells without the use of a fluorescent tag, and not just for future Anchor Away depletion analysis. While it takes a while to analyze the images, the extra tagging step is eliminated and may be a shorter analysis time in the long run compared to tagging a cell wall protein and then analyzing the images. The depletion time of GFP alone should also be measured as a control in the future. The rate of TBP trafficking out of the nucleus can only be as fast or slow as that of GFP and should also be taken into consideration.

Once the depletion rate for TBP in the Anchor Away strain was determined, TBP ChIP signals were measured at various loci in response to rapamycin. At most loci, TBP levels reached or approached zero by 60 minutes, which is consistent with prior work (Haruki, Nishikawa et al. 2008). Where TBP levels did not reach zero (binding to *URA1*, *ASP1*, and *ACT1*), the errors associated with the measurements were significant so the differences could be more apparent than real. Since the measured depletion time for 50% off the TBP molecules to leave the nucleus is ~11 minutes, almost all of the TBP should be depleted by about 22 minutes. The Anchor Away data sets show ~50% nuclear TBP depletion occurs anywhere from 10-30 minutes after rapamycin addition. Since the depletion rate is an average of all the TBP protein bound, there are DNA binding sites that can have a longer or shorter binding time, so TBP may take a shorter or longer time to be depleted from the nucleus across various loci. One of the goals of this assay is to determine the dynamic range of TBP at individual promoters. Future work could look at TBP binding genome wide to get a better measurement for this range than only looking at a few select promoters.

The apparent residence time and dissociation-rate for TBP bound to specific loci was calculated from the exponential decay curve fit; these are apparent rates because the actual rate could be impacted by factor depletion time. It's uncertain what behavior is actually captured by Anchor Away, as there are several possibilities: a protein complex binding to a chromatin site,

then dissociating as a complex; a protein complex binding to a chromatin site, then dissociating as individual components; a protein could be free-floating in the nucleus and transported into the cytoplasm; or a protein binding and dissociating several times before it's depleted from the nucleus. When comparing the residence times generated from Anchor Away and the CLKv2 method for TBP at three loci for which we have data using both methods, the Anchor Away values at *URA1* and *ACT1* were fairly close to one another (within a minute of each other), but the CLK values for these loci were 3-4 times faster than those measured for Anchor Away (Table 3.4). Stefan Bekiranov has worked out a kinetic model of TF-chromatin binding that can be used in future work to fit the locus-specific Anchor Away data and incorporate the nuclear depletion rate/nuclear concentration of the factor as measured by microscopy. The TBP Anchor Away has not yet been fit with this model because we have not had the bandwidth to write the scripts and troubleshoot the model. If the values are in agreement for TPB at *LOS1*, *URA1*, and *ACT1* after fitting with the TF-chromatin binding model, then these two assays may be measuring a similar kind of behavior. If there is still a discrepancy, more work will need to be done to look into why this difference remains and if a different activity is captured in the two experimental assays.

Table 3.4. List of Anchor Away approximated residence times and CLKv2 calculated residence times at select loci.

Locus	**Anchor Away $t_{1/2}$ (min)**	**CLKv2 $t_{1/2}$ (min)**
URA1	6.4	1.1 (+0.35, -0.28)
LOS1	27.4	1.35 (+0.35, -0.28)
ACT1	7.9	1.65 (+1.65, -0.95)

Competition ChIP as an independent ChIP-based approach to measure binding dynamics

As another ChIP-based assay to measure TF-chromatin dynamics, a Competition ChIP strain for TFIIB (Sua7) was successfully constructed. One *SUA7* allele is under control of the *PGAL1* promoter, and inducible using galactose addition. When the constructed CC strains were compared to the original diploid strain, YPH501, and a different background strain, BY4741, grown on galactose media, a growth deficiency was observed in YPH501 and the derived CC strains. This is most likely from the *gal2* mutation in the YPH501 background, which is not listed in current genotypic information for the strain and we discovered after the strain was made. Gal2 is a permease required for the utilization of galactose (Tschopp et al. 1986), and a permease deficiency would likely lead to the kinetic defect that we observe. The induction time measured for the TFIIB CC strains was ~60 minutes, much longer than that reported for a TBP CC strain (van Werven et al. 2009, Zaidi et al. 2017). Since curves generated from collection of CC samples are fit using the induction curve, this delayed induction is probably okay; however the time course would need to be extended. A plasmid containing *GAL2* can be introduced to the CC strain (Ulery et al. 1991), or wild type *GAL2* can be integrated into the genome to restore growth in galactose (Keeney et al. 1995). The TBP CC strain was made in the W303 background (van Werven et al. 2009). This strain background is derived from the *gal2* mutated SC288 and shares >85% similarity, but fortuitously has a wild type *GAL2* (Young and Court 2008).

TFIIB CC data can be obtained once the Gal2 deficiency is corrected. Libraries will be made of the ChIP samples for each TFIIB allele obtained from cells fixed at various time points after galactose addition. Using the rate of competitor induction, the data at each locus can then be fit and the residence time and k_d determined using the modeling approach that was used for TBP CC data (Zaidi et al. 2017) but updated to accommodate high throughput sequencing rather than array data. While TFIIB is somewhat depleted using 5% formaldehyde in the CLKv2 assay (Zaidi et al. 2017), which limits the useful range of CLKv2 for analysis of TFIIB dynamics. This may make CC a better assay for this factor. TFIIB CLKv2 could also be performed using a lower

concentration of formaldehyde to minimize depletion of Sua7 from the protein pool; pilot experiments would be required to determine the crosslinking behavior and whether the resulting data could be modeled to yield binding kinetic information of value.

No genome-scale, locus-specific binding dynamic information is available for TFIIB, although a recent microscopy study reported that TFIIB has a quick turnover rate and binds to *in vitro* promoters for ~30 seconds when other PIC components and Pol II are present (Zhang et al. 2016). Recent evidence supports the idea that TBP has a wide range of residence times *in* vivo, from less than 1.3 minutes to around 60 minutes (Zaidi et al. 2017). One goal of studying PIC component dynamics is to determine if the chromatin binding kinetics depend on the class of promoter or local chromatin environment. The Zaidi *et al.* study classified loci as either TFIID- or SAGA-dependent and TATA or TATA-less, based on distinctions made in an earlier study (Basehoar et al. 2004). There was no correlation between residence times based on either of these groupings. Additionally, there was no correlation based nucleosome positioning and only a slight correlation of higher nascent transcription rates (as defined in (Pelechano et al. 2010)) to shorter residence times. However, there was a correlation with the previously determined dynamics of the essential DNA-binding regulator Rap1 (Lickwar et al. 2012). It's somewhat surprising there was no difference between residence times of TATA and TATA-less genes, as TATA genes are thought of as more dynamic with a faster turnover (van Werven et al. 2009) since these genes are traditionally related to stress response and their expression is higher regulated (Basehoar et al. 2004, Huisinga and Pugh 2004). It's not surprising that there wasn't a different in residence time based on TFIID or SAGA regulation because in contrast to previous views, it has recently been established that almost all genes are dependent on both of these complexes (Baptista et al. 2017, Warfield et al. 2017). Instead of classifying a gene as regulated by TFIID or SAGA, genes were split into five categories based on mRNA synthesis changes in SAGA mutant strains (Baptista et al. 2017). When the TBP residence times generated from Zaidi *et al.* were split into these five groups, there was no difference observed between them (Figure 3.25). This suggests that the wide

range in binding dynamics for TBP cannot be ascribed to these different categories, but instead is due to some other features of the nuclear environment that haven't been related to binding dynamics yet. It will be interesting to see if any other PIC components have some correlation with these gene expression groupings. Taken together, these results show that there is more to the dynamics of TBP than just promoter DNA sequences; the chromatin architecture, binding environment, and dynamics of other factors will need to be considered.

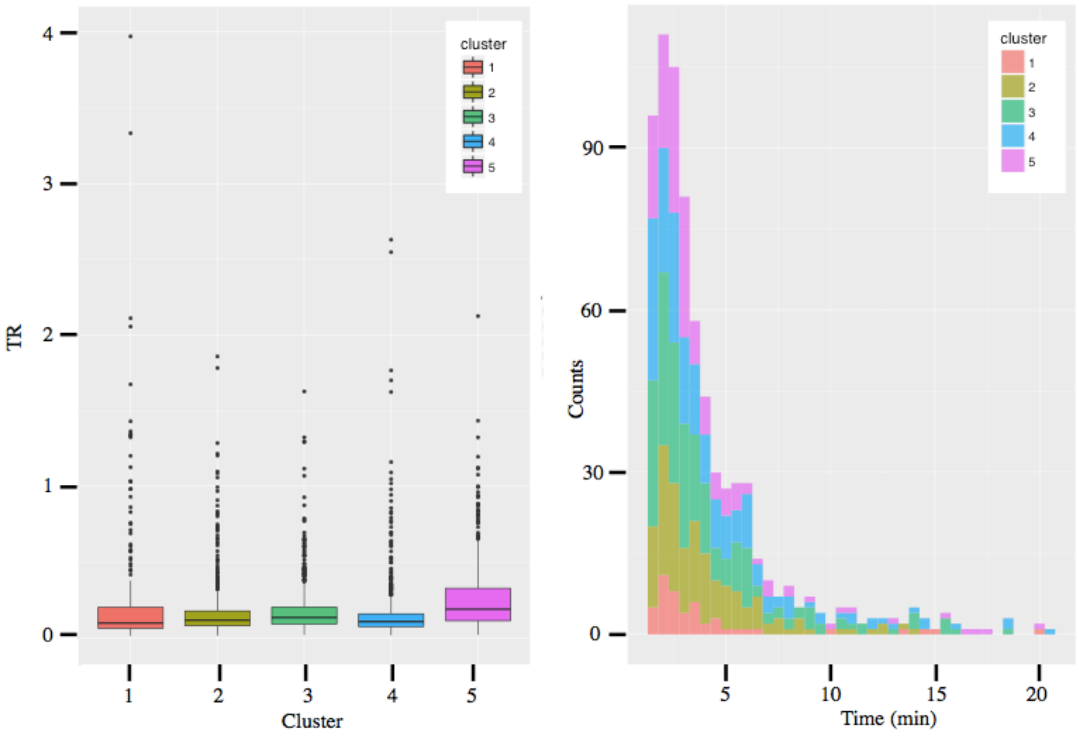


Figure 3.25. Plots of TBP residence times at RNAPII-dependent genes split into five groups based on the Baptista *et al.* 2017 classification. The left side shows a box plot of the median transcription rate (y-axis) across the five groups. The right plot shows an overlay of the residence time (x-axis) and the number of counts for each of the five groups. Corresponding colors are denoted in the legend in the right hand corner of each plot.

TBP binding kinetic data are available for all three ChIP approaches at three loci (*URA1, LOS1, ACT1*). The three sets of measurements are not in perfect concordance (see Anchor Away discussion above), but the CC and CLKv2 results are similar (Table 3.5). Since the fastest residence time that can be measured with CC is 1.3 minutes, it is unknown if the actual value is closer to one minute or possibly even much faster, which could be a big difference. Taking into account the error from CLKv2, the values for binding dynamics at the three loci as measured with CC and CLKv2 could be very close to one another. With improved Anchor Away fitting, the datasets for at least *ACT1* and *URA1* will hopefully converge. The large discrepancy between methods at the *LOS1* promoter may be due to rapamycin-mediated effects. The Anchor Away strain is a *tor1Δ* strains; in other *tor1Δ* strains, addition of rapamycin *in vivo* has been shown to enrich and sequester Los1-GFP in the cytoplasm in yeast (McCormick et al. 2015). While this should not directly affect transcription of the protein as initiated by TBP, rapamycin could play a part in altering dynamics at this gene. This may explain why there is not consensus between the binding dynamics measurements.

Table 3.5. Comparison of TBP residence times using the three independent ChIP-based approaches.

Locus	AA $t_{1/2}$ (min)	CC $t_{1/2}$ (min)	CLKv2 $t_{1/2}$ (min)
URA1	6.4	Too fast	1.1
LOS1	27.4	Too fast	1.3
ACT1	7.9	Too fast	1.7
ASP1	8.1	1.36	---
RSC1	20.7	20.63	---
RPL9B	24.5	Too fast	---
URA8	9.4	1.38	---
ADH1	9.7	2.92	---
FBA1	14.7	Too fast	---
CDC19	13.8	1.69	---

The ultimate goal of using several independent ChIP-based approaches is to compare the kinetic parameter values to come to a consensus on the actual *in vivo* dynamics; there are already CLK and Competition ChIP data sets for TBP, and TFIIB data are on the way. Eventually, the goal is to have data sets from all three methods for most of the PIC components (TBP, TFIIB, TFIIA, TFIIF, TFIIE). By looking at the individual components together, the *in vivo* dynamics and assembly of the PIC can be better understood. Comparison to nucleosome positioning and RNA synthesis rates will also aid in completing the overall picture of transcriptional regulation.

Future directions: the power of combined approaches

Even though the individual components involved in transcription are known, as well as much of their biochemistry, there are still fundamental unanswered questions about how these complexes assemble and are regulated *in vivo*. Measuring the binding dynamics of these transcription factors in cells can provide valuable information that can be used to answer some of these questions. An array of experimental approaches can provide information about *in vivo* chromatin binding dynamics. Independent live cell imaging modalities have converged so as to yield self-consistent measurements for a few factors (Mazza et al. 2012, Mazza et al. 2013). Cross-validation has also been done in comparing CC and AA relative dynamics (Grimaldi et al. 2014) and in comparing FRAP and the original CLK method (Poorey et al. 2013). It is essential that future studies of TF dynamics include different experimental approaches and attempt to reconcile discrepant dynamic observations when they arise. Among the thousands of TFs present in a diverse array of cell types and conditions, only a relatively small collection have been measured; it thus seems likely that the full spectrum of TF dynamic behavior is far from understood. Errors or naiveté in data interpretation using one method revealed by comparison with another could lead to improved tools for more accurately measuring dynamics as well as identifying new dynamic phenomena. Different experimental approaches provide access to kinetic behavior on specific and often quite different time scales. However, a holistic view of

chromatin binding behavior will require models that account for all of the observations. In general, imaging approaches have uncovered much more rapid dynamic behavior than can be observed using ChIP-based approaches. Such disconnected sets of observations may represent different types of dynamic behavior occurring simultaneously rather than being discordant observations (Karpova et al. 2008). Future studies that compare results obtained by different methods applied to the same TF-chromatin interactions will be valuable in sorting this out.

TFs interact with other factors and frequently assemble into higher order complexes, such as the PIC. The different molecular approaches for measuring chromatin-binding dynamics may in some circumstances provide complementary information about higher order assembly as well as chromatin binding (Figure 3.26). For example, as CC relies on the inducible replacement of one TF isoform with another, it might capture the dynamics of complex assembly if it is rate limiting, rather than chromatin binding per se. Likewise, CC may be sensitive to slow steps in localization of TFs to different nuclear sub-compartments. In contrast, as CLK measures DNA crosslinking directly, it provides specific information on DNA binding but no information about other steps, and albeit in the context of a formaldehyde-treated cell. AA most likely captures TF-chromatin dissociation, a process computed by CLK in the context of the on-rate and fractional occupancy.

The pathway for PIC assembly is well understood biochemically, but whether the *in vitro* assembly pathway is adequate to describe PIC assembly *in vivo* is not known, and there is some evidence that alternative pathways may exist (Zanton and Pugh 2006). Two models for multi-factor TF binding to chromatin have been suggested: random and regulated assembly. In the first, random collisions with the DNA template result in complex formation, whereas in the second a regulatory molecule stabilizes or destabilizes complex formation (Figure 3; (Hager et al. 2009)). Emerging experimental and computational approaches have recently provided new insight into how this can occur (Chen et al. 2014, Stasevich et al. 2014), but many interesting biological systems and regulatory scenarios remain to be explored.

Measurements of TF dynamics have required the development of new experimental approaches, which has been challenging. Technical innovation will undoubtedly continue, but it is now possible to begin to apply the array of validated tools to specific biological problems of increasing regulatory complexity. Combined experimental approaches provide a reality check on each other and can potentially reveal complementary aspects of dynamic behavior that no one method is capable of capturing. As the landscape of TF dynamic behavior becomes better defined, it will be particularly interesting to see how the behaviors of individual TFs impact the formation of PICs and other higher order complexes, as well as more detailed understanding of how TF dynamics impact RNA synthesis, both in individual cells and across cell populations.

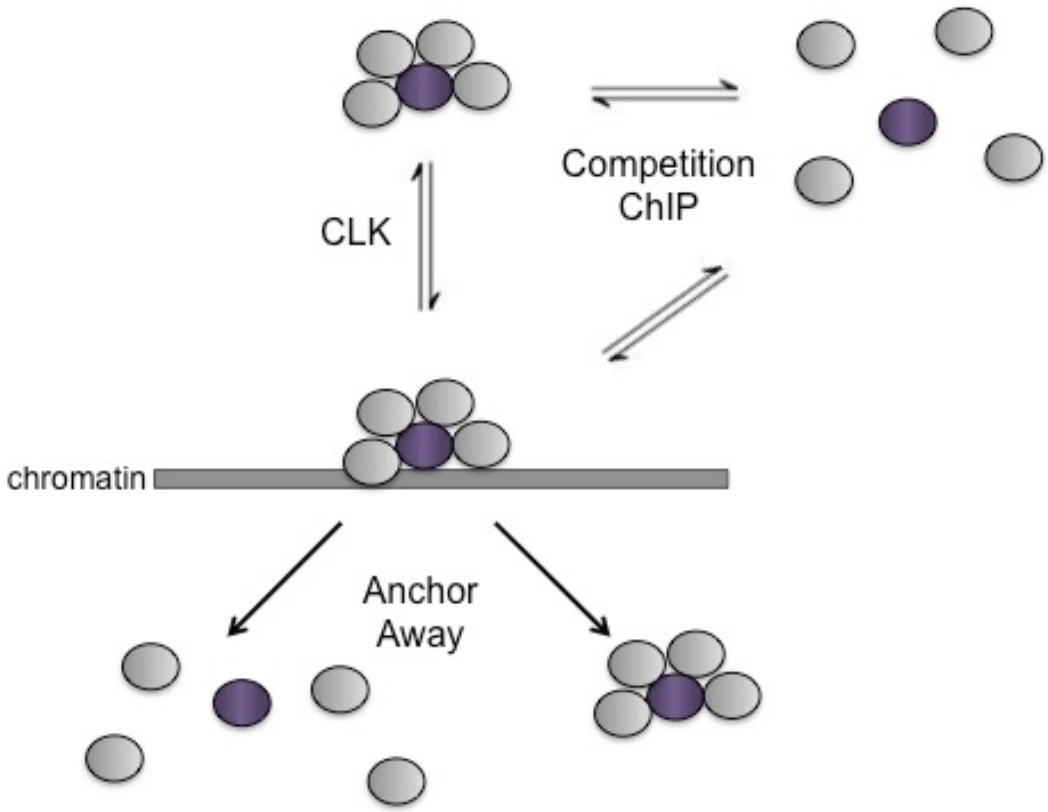


Figure 3.26. Higher order chromatin assembly and speculative contributions of different molecular methodologies to understanding it. The figure illustrates a TF of interest (purple) that associates with other factors to form a complex that associates with chromatin. The multi-

subunit-containing chromatin complex may assemble from individual subunits that come together individually and perhaps randomly at a chromatin site, multi-subunit complexes may first assemble and then associate with a chromatin site, or the process may be a combination of both. CC can capture the rate-limiting step of TF turnover, which may involve direct DNA binding by the competitor or assembly into a complex that associates with DNA. CLK appears to measure DNA binding directly by modeling crosslinking to DNA itself, whereas AA measures the DNA dissociation rate.

CHAPTER IV

Dynamics of the transcriptional activator, Gal4

Gal4 is a transcription factor required to activate the galactose inducible *GAL* genes. Under non-inducing conditions including glucose and raffinose, this set of genes is not transcribed, but upon growth in galactose-containing media, the activation domain of Gal4 is released from its interaction with Gal80. Upon Gal4 activation, regulatory complexes and the preinitiation complex components are recruited via Gal4's interaction with the upstream activating sequences (UAS) for each induced *GAL* gene and transcription can occur to allow growth in galactose. While this system has been studied in detail for many years, the *in vivo* dynamics of Gal4 at the individual *GAL* promoters is not known. Using the crosslinking kinetics assay (CLK), Anchor Away, and a live cell imaging technique developed by the Larson lab at NCI, Gal4 binding dynamics were measured at the *GAL3* promoter under several conditions, including wild type cells, cells with a weakened *GAL3* UAS_G, and *gal11Δ* strains, which is a subunit in Mediator that interacts with Gal4. In WT cells, Gal4 bound to *GAL3* for ~14 minutes based on CLK, ~3 minutes according to Anchor Away, and ~2 minutes with the live cell imaging approach. Gal4 residence time on the weakened *GAL3* UAS_G was much shorter when measured with CLK (<1 s) and live cell imaging (~1 minute) compared to the WT residence time. In *gal11Δ* cells, Gal4 residence time at *GAL3* was less than 1.1 second when measured by the CLK assay. These results indicate that interrupting the Gal4-Mediator interaction has a substantial effect on Gal4 binding dynamics. Kinetic data will eventually be acquired under all of these conditions and by all three methods to better understand methodological strengths and weaknesses and to arrive at consensus Gal4 binding dynamics parameters. Collectively, these results suggest that Gal4 binding to *GAL3* persists for at least several minutes in WT cells, even a modest mutation in the consensus UAS_G has a dramatic effect on Gal4 binding stability *in vivo*, and the Gal11-containing Mediator complex is required for stable binding

of Gal4 to an activated promoter. Since the *GAL* system has widespread use as an inducible model, an understanding of the *in vivo* dynamics of Gal4 and associated factors under different conditions will likely impact our general understanding of how transcriptional control occurs in vivo.

Introduction

Gal4 activation as a model system for measuring transcription dynamics

Transcription is a highly regulated system requiring a complex assortment of co-activators, suppressors, general transcription factors, and RNA polymerase. The *S. cerevisiae* Gal4 protein is a well-studied activator that regulates transcriptional expression of the *GAL* genes (reviewed in (Johnston 1987)). These include the structural (*GAL1, GAL10, GAL2,* and *GAL7*) (Bassel and Mortimer 1971) and regulatory (*GAL4, GAL80,* and *GAL3*) genes required for galactose uptake and processing (Figure 4.1) (Johnston 1987). A few additional genes (*MTH1, PCL10,* and *FUR4*) are also regulated by Gal4, but are required for adapting to growth on galactose and not for metabolism (Ren 2000).

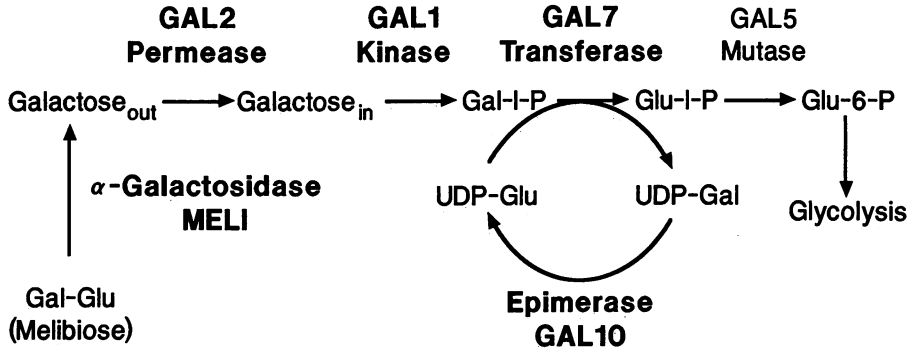


Figure 4.1. Galactose uptake and metabolism by the *GAL* genes; figure from Johnston, 1987. The Gal2 permease transports galactose into the cell, where it is transformed through a series of steps to glucose-6-phosphate, which can be used in glycolysis.

Gal4 is an 99.4 kDa protein composed of 881 amino acids with a cysteine-zinc binuclear cluster DNA binding domain on the N-terminal, a linker domain, a dimerization domain, and two activation domains (Johnston 1987, Lohr et al. 1995, Traven et al. 2006). One activation domain consists of residues 148-196 and the other is from 768-881 (Ma and Ptashne 1987, Hong et al. 2008); the C-terminal domain overlaps with the Gal80 binding domain (Johnston 1987, Johnston et al. 1987, Ma and Ptashne 1987).

Gal4 binds to an upstream activating sequence, UAS_G, which has the sequence CGG-N_{11}-CCG (Bram and Kornberg 1985, Giniger et al. 1985, Shimada and Fukasawa 1985, Bram et al. 1986, Keegan et al. 1986). Variants of this sequence are present in all *GAL* gene promoters in nucleosome free regions, but the number and affinity varies for each gene (Johnston 1987, Lohr, Venkov et al. 1995). The *GAL1*, *GAL10*, and *GAL7* genes are clustered on chromosome II (Douglas and Hawthorne 1964, Bassel and Mortimer 1971, St. John and Davis 1981), while *MEL1*, *GAL3*, *GAL4*, and *GAL80* reside on separate chromosomes (Johnston 1987). The *GAL1-GAL10* divergent promoter contains four binding sites, three of which are stronger and one weaker; only the two middle sites are essential for gene expression (West et al. 1984). *GAL7* and *GAL2* both have two binding sites in their promoters, while the remaining *GAL* genes have one binding site (Johnston 1987). One site is sufficient for gene expression, and in some cases having two binding sites increases expression (West et al. 1984, Giniger, Varnum et al. 1985, Lorch and Kornberg 1985). As Gal4 binding sites do not have the same sequences, the Gal4 binding affinity of each differs (Johnston 1987).

GAL gene expression is controlled by the sugar source. Under both inducing (galactose) and non-inducing conditions (raffinose), Gal4 is bound to the UAS_G, but is prevented from activating transcription by the Gal80 protein under non-inducing conditions (Figure 4.2). When cells are grown in glucose, Gal4 is actively repressed through the catabolite repression protein (CRP) and does not bind to a UAS_G (Johnston 1987). Gal80 is still bound to Gal4 to block the activation domain, and transcription is repressed because glucose can be used in the glycolytic

pathway without metabolic processing beforehand and therefore the *GAL* genes are not needed (Douglas and Hawthorne 1964, Adams 1972, Matsumoto et al. 1981, St. John and Davis 1981, Yocum and Johnston 1984). The Gal80 binding site on Gal4 overlaps the activation site, preventing recruitment of the required transcriptional machinery. When galactose is added, nuclear Gal3 is thought to remove Gal80 from Gal4, potentially through a transient interaction with Gal80 that causes a conformational change in Gal80, removing Gal80 from the nucleus and causing Gal4 activation (Lohr et al. 1995, Egriboz et al. 2011). There is also evidence that in the presence of galactose, Gal80 is degraded by the Skip1-Cullin-F-box protein (SCF) E3 ubiquitin ligase containing Mdm30, thus freeing Gal4 for transcriptional activation (Ang et al. 2012). Once Gal80 is removed, Gal4 is free to recruit co-activators and the transcriptional machinery to the promoter, including the SAGA and Mediator (see Chapter I for more information) complexes, and the general transcription factors (GTFs) TBP and TFIIB. Previous studies have shown that Gal4 directly targets SAGA and this interaction requires the Gal4 activation domain and Spt20 SAGA subunit (Bhaumik and Green 2001, Larschan and Winston 2001, Bryant and Ptashne 2003). When Gal4 is bound to DNA, its activation domain directly binds the SAGA Tra1 subunit (Bhaumik et al. 2004, Knutson and Hahn 2011); Gal4 recruitment of SAGA is independent of the Gcn5 histone acetyltransferase (HAT) subunit, indicating the complex works as a PIC scaffold and not a histone modifier (Bhaumik and Green 2001, Traven et al. 2006). Once SAGA binds, Spt3 can recruit TBP and the rest of the PIC. There are conflicting results on whether the Mediator regulatory complex is a direct target of Gal4 or if SAGA is required for its recruitment (Larschan and Winston 2001, Bryant and Ptashne 2003, Kuras et al. 2003, Bhaumik, Raha et al. 2004, Lemieux and Gaudreau 2004). There is evidence that Mediator acts upstream of Gal4 and that the degradation process is initiated through Snf1 signaling to Mediator (Ang et al. 2012). It appears that Gal4 and Mediator may directly interact, but the presence of SAGA at the UAS_G is required for stability and formation of a productive PIC (Traven et al. 2006). Taken together is it

clear that despite all the effort, more work is needed to figure out the *in vivo* molecular mechanisms of assembly and this open question provides motivation for the work in this Chapter.

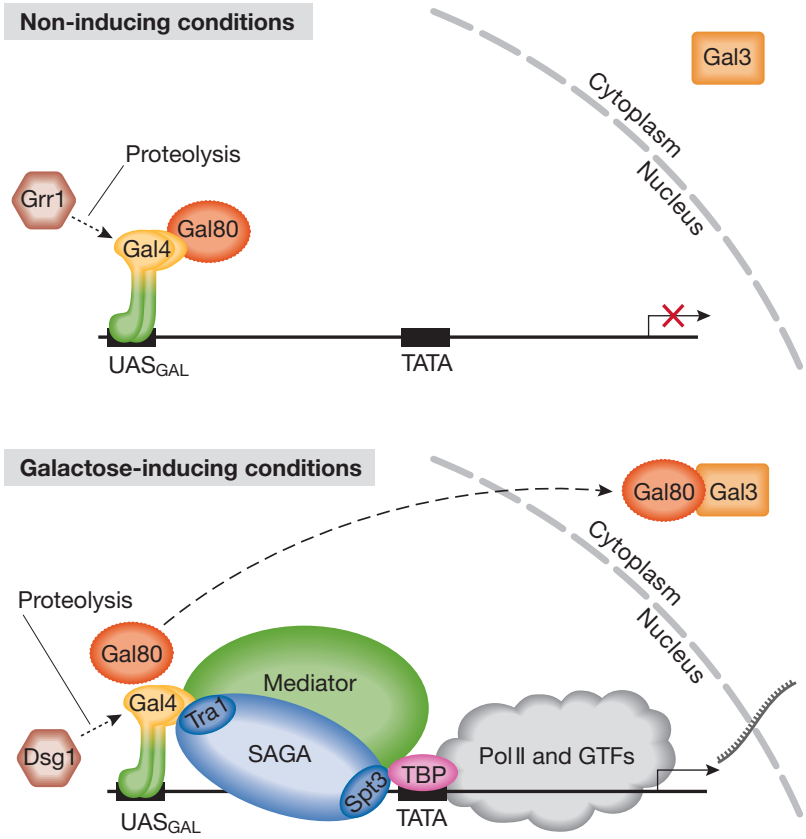


Figure 4.2. Schematic of the mechanism of Gal4 transcriptional activation; Figure from Traven *et al*, 2006. Under non-inducing conditions (top), such as raffinose, Gal80 binds to the activation domain of upstream activating sequence (UAS)-bound Gal4, blocking transcription from that site. The F-box protein Grr1 regulates Gal4 levels in order to limit accumulation in non-inducing conditions. When galactose is added to the system (bottom), transcription is initiated by removal of Gal80 binding to Gal4 by Gal3, which then transports Gal80 out of the nucleus. Gal4 can then bind to co-activator and regulatory complexes through its activation domain, recruit the core machinery, and initiate transcription. The F-box protein Dsg1 promotes turnover of transcriptionally active Gal4 under inducible conditions.

Gal4 binding dynamics

Although the Gal4 system is well studied, the dynamics of Gal4 binding to its regulatory sites are not well defined. We want to measure these binding dynamics to understand how this activator helps to regulate transcription and works in conjunction with other co-activator complexes. What is the dynamic range of Gal4 binding to the *GAL* genes? How does this correspond to RNA synthesis and the dynamics of transcription itself? How are the binding dynamics of Gal4 affected when complexes it interacts with are compromised? By measuring dynamics, we can better understand how Gal4 acts *in vivo* and answer these biological questions. Two studies investigated the binding of Gal4 binding to the transcriptionally divergent *GAL1* genes (Nalley et al. 2006, Collins et al. 2009) with differing conclusions. Nalley *et al*. used a ChIP-based competition assay that is not unlike CC. Fusion proteins with the ligand binding domain (LBD) of the oestrogen receptor-α (ER) were fused to the DNA binding domain (DBD) of Gal4, yielding a Myc-Gal4(DBD)-ER(LBD)-VP16-Flag fusion protein that was bound to Hsp90, provided in excess in the system. When β-oestradiol was added, the fusion protein interaction with Hsp90 is disrupted and the chimeric activator is free to compete for DNA binding with the endogenous Gal4 protein. Antibodies were used to measure binding of the different Gal4 species at the *GAL1/10* promoter by ChIP. The authors found that Gal4 was stably bound for upwards of an hour, and additionally, inhibition of proteasome activity had no effect on Gal4 binding (Nalley et al. 2006), which was surprising based on results from a previous study (Lipford et al. 2005). Collins *et al.* challenged these results. They performed an additional control using β-estradiol alone without competitor and found a four-fold increased of Gal4 bound to the *GAL1/10* promoter compared to in the absence of β-estradiol. It seems that β-estradiol was somehow changing the binding stability or association of Gal4 with the promoter, which confounds the results since both the binding of competitor and Gal4 are induced. They repeated the experiments using an alternative competitor, 4-hydroxytamoxifen (4HT) and found that within 15 minutes, 75% of the endogenous promoter-bound Gal4 had been replaced by the

competitor. Therefore, Gal4 binding was much more dynamic than was thought based on the prior study, with an persistence on the promoter of possible minutes. The first CLK paper measured a residence time of ~10 minutes for Gal4 at *GAL3* (Poorey et al. 2013), but with a better understanding of the CLK method, it became imperative to repeat the measurements using optimized conditions as described in Chapter II. Taken together, it's clear that there is no consensus on Gal4 binding dynamics. Knowing the time scale of the Gal4-chromatin interaction will give insight into important biological questions, especially when these measurements are combined with RNA synthesis data. For example, if Gal4 is highly stable, it may be able to initiate multiple rounds of transcription without rebinding to DNA. If instead the residence time is highly dynamic, perhaps Gal4 is degraded and needs to rebind for each round. Additionally, looking at the binding dynamics of Gal4 in different strain contexts can inform us how Gal4 is recruited and the effect that other co-activators have on its stability. One of the goals of our work is to quantify Gal4 binding dynamics to clear up this debate to better understand the biological implications of the dynamic behavior. By mutating regulatory complexes, such as SAGA and Mediator, the mechanism of activator-induced transcription and the effect on preinitiation complex stability and gene expression can be elucidated. These results are not just informative for the *GAL* system, but will provide further insight into how the process of transcriptional activation works.

In addition to the formaldehyde-mediated, ChIP-based techniques that we use to capture protein-DNA complexes, the Larson lab developed a complementary live-cell imaging approach that visualizes RNA synthesis in real time from galactose-induced promoters (Figure 4.3) (Larson et al. 2011). The method involves insertion of a cassette with 24 binding sites for the PP7 bacteriophage coat protein into the 5' untranslated region of the gene of interest. The fusion protein PP7-GFP is constitutively expressed and binds to the transcriptional stem loops, essentially GFP labeling the product. The initiation of pre-mRNA and elongation to termination of an RNA transcript can be visualized. By performing autocorrelation analysis on the time-

dependent fluorescent signal, one can extract transcription kinetics, including the dwell time of each RNA molecule at the transcription site and the number of RNA polymerase II molecules transcribing a particular site based on the fluorescence intensity (Larson et al. 2011). For these reasons, we sought to correlate the transcription kinetics measured for *GAL* genes in this system with our data from ChIP-based approaches. There is already published data for a strain made using this technique at *GAL1/10* (Lenstra et al. 2015) and Tineke Lenstra has collected data for Gal4 at *GAL3* and with a weakened *GAL3* UAS_G. This technique works well for Gal4 because it only binds to a small number of genes and is present at about 300 copies per cell (extrapolated from (Borggrefe et al. 2001, Jorgensen et al. 2007, Poorey et al. 2013). This would not be a viable way to measure TBP transcription dynamics as many promoters, but because of the nature of Gal4, this system offered an opportunity to cross-validate our kinetic data by a completely independent method.

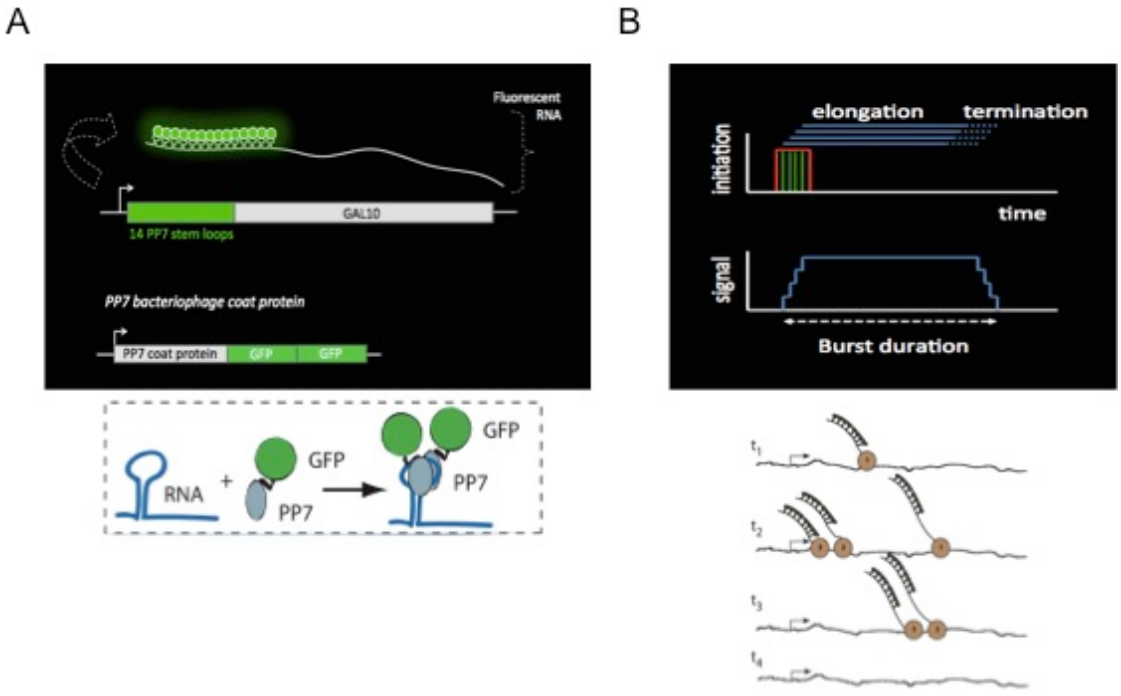


Figure 4.3. Live-cell imaging technique to visualize initiation and elongation of fluorescently labeled transcripts in yeast; adapted from Larson *et al.*, 2011 and from personal communication. A) Multiple PP7 binding sites are inserted upstream of the target gene. GFP-

labeled PP7 is constitutively expressed in the cell concurrently (top). When the stem loops are transcribed, the GFP-labeled PP7 binds them and fluorescently labels the transcript (bottom). B) Simulation of traces from fluorescently labeled transcripts. Initiation is represented by the vertical green lines as PP7 binds; elongation is shown by the horizontal blue lines and termination occurs when the signal drops (top). Multiple RNA polymerase II molecules can be transcribing at the same site; the intensity of the elongation signal can be correlated to the number of polymerases. The bottom panel shows an example of polymerases bound at four transcription sites.

Materials & Methods

CLKv2 data set collection

Yeast strains used in this chapter. Several sets of wild type and overexpression strains were used for CLK datasets and controls; all strains used are listed below in Table 4.1. To detect Gal4 by western blot, a Gal4-TAP-tagged strain, YRV005 (Ghaemmaghami et al. 2003) was used since we have had difficulty detecting Gal4 in the past due to its low cellular copy number (Poorey et al. 2013). To detect wild type Gal4, the wild type haploid YPH499 strain (Sikorski and Hieter 1989) was transformed with the empty *URA3*-marked pRS426 plasmid (Christianson et al. 1992) to yield strain AY156 or with a 2µ pSJ4 plasmid with an extra copy of Gal4 with its endogenous promoter (Johnston and Hopper 1982) to make the overexpression strain AY157 (YRV004) (Poorey et al. 2013).

A yeast strain with a weakened *GAL3* UAS_G was constructed by David Auble. The normal *GAL3* UAS_G (CGGTCCACTGTGTGCCG) was replaced with the weakest of the four *GAL1-10* UAS_G sequences (AGGAAGACTCTCCTCCG) (Bram et al. 1986) and transformed with the pRS426 2µ empty plasmid (Christianson et al. 1992) to make the WT CLK strain AY158; the strain with a weakened *GAL3* UAS_G was transformed with pSJ4 (Johnston and Hopper 1982) to make the CLK OE strain AY159.

CLK strains were also constructed with a deleted *GAL11*. The locus was interrupted with a *KanMX4* gene insert (Wach et al. 1994) from the pRS400a plasmid by PCR-mediated gene disruption (Brachmann et al. 1998) using the GAL11_KO-F/GAL11_KO_R primers (Figure 4.4). Knock-out candidates were selected on YPD+G418 plates and verified by PCR with gDNA using a primer set upstream of the *GAL11* ORF to detect the insert (GAL11_detect1/GAL11_detect2) and other set in the endogenous ORF (GAL11_detect1/GAL11_detect3); the gDNA preparation is described in Chapter III in Materials & Methods, *Verification of transformation candidates by PCR with gDNA and western blotting*. To make the WT CLK strain EH004, the knockout strain was transformed with the pRS426 plasmid (Christianson et al. 1992); the CLK OE strain EH003 was made by transforming the knockout strain with the pSJ4 plasmid (Johnston and Hopper 1982). A spot test was done to compare growth of the *gal11Δ* strain to the wild type YPH499 and WT and OE CLK strains; this protocol is described in Chapter III Materials & Methods *Spot test to compare growth of tagged strain to original wild type diploid*.

The TBP CLK strains AY146 (WT) and AY147 (OE) (Zaidi et al. 2017) used in Chapter II were also collected in this chapter, but grown in SC-LEU+2% raffinose and YEP+2% galactose media to measure TBP dynamics at *GAL3*.

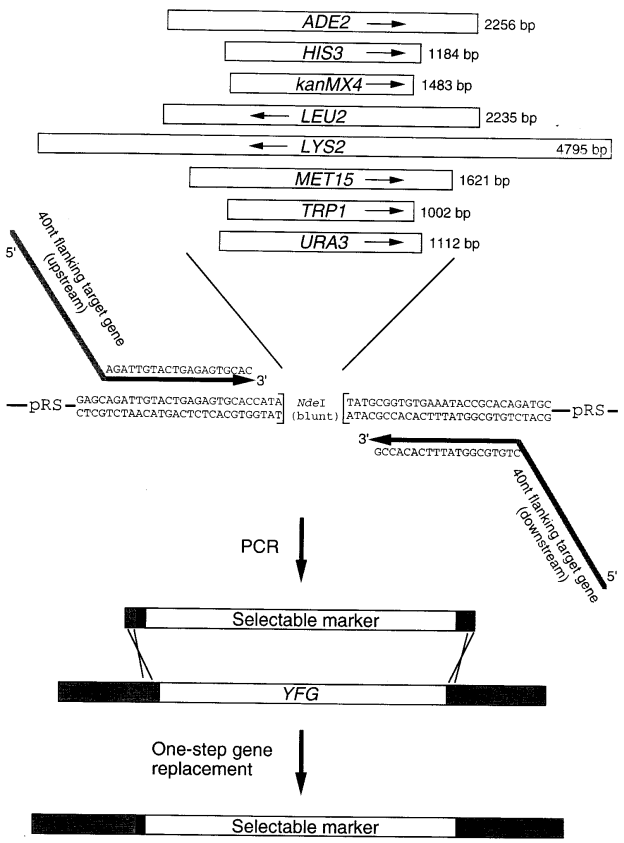


Figure 4.4. Diagram of PCR-mediated gene disruption; from Brachmann *et al.*, 1998. A variety of selectable markers were incorporated into a series of pRS400a plasmids at a blunt ended *NdeI* site to be used for gene disruption (Brachmann et al. 1998). A universal set of primers can be used to amplify any of these markers; 40 base pairs of homology upstream or downstream of the gene to be deleted (YFG in this example) are added to the 5' end of each primer. The gene-replacement marker is amplified, transformed in a yeast strain, and integrated into the desired locus by one-step gene replacement. The resulting strain has a knocked-out gene replaced by the selectable marker.

Table 4.1. Yeast strains used in this chapter.

Strain	Genotype	Reference or source
YRV005	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 GAL4-TAP*	Ghaemmaghami *et al.* 2003
YPH499	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1*	Sikorski and Hieter, 1989
AY156	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 pRS426 [URA3 2μ]*	This study
AY157	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 pSJ4 [GAL4 URA3 2μ]*	Poorey *et al*., 2013
AY158	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 pRS426 [URA3 2μ]*	This study
AY159	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 pSJ4 [GAL4 URA3 2μ]*	This study
EH004	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 Δgal11 :: KanMX4 pRS426 [URA3 2μ]*	This study
EH003	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 Δgal11 :: KanMX4 pSJ4 [GAL4 URA3 2μ]*	This study
AY146	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 spt15::natMX pRS426 [LEU 2μ]*	Poorey et al, 2013
AY147	*MATa ura3-52 lys2-801a ade2-101o trp1-Δ63 his3-Δ200 leu2-Δ1 spt15::natMX pSH223 [TBP LEU 2μ]*	Poorey et al, 2013

Western blotting to test for factor depletion. The same experimental crosslinking and quenching conditions described in Chapters II & III were used to collect Gal4 samples: 5% formaldehyde to crosslink and 2.93 M glycine pH 5 to quench. The Gal4-TAP strain YRV005 (Ghaemmaghami et al. 2003) was grown in duplicate 5 ml YPD overnight at 30°C; in the morning, all of each primary culture was added to 300 ml YEP + 2% galactose. When the OD_{600} was ~1, cells were spun down for 7 minutes at 5,000 rpm and 4°C, then 5X concentrated by resuspension in 60 ml YEP + 2% galactose. Time points were collected as described in Chapter II Materials & Methods *Whole cell extract preparation and western blotting*. Briefly, a 10 zero minute time point was removed from each culture and quenched in 440 ml 3 M glycine pH 5. Formaldehyde was added to 5% and incubated with the culture on a stir plate. At 5, 10, and 15 minutes, a 10 ml aliquot of crosslinked cells was removed from each culture and quenched with 3 M glycine pH 5. Quenched cells were spun down for 7 minutes at 5,000 rpm and 4°C and processed as chromatin, outlined in

Chapter II Materials and Methods *Collection of crosslinking time points and preparation of chromatin samples*. Whole cell extracts were not collected because we cannot detect Gal4-TAP in these samples. Chromatin samples were run on an SDS-PAGE gel and transferred to a PVDV membrane, as outlined previously in Chapter II Material & Methods *Whole cell extract preparation and western blotting*. The membranes were probed with an α-Protein A primary antibody (Sigma Aldirch, P-3775) to detect the TAP tag and an α-rabbit-HRP secondary antibody (GE Healthcare). The blots were transferred onto film, scanned, and saved as a TIFF before quantification with ImageJ (NIH). The intensity of the 5, 10, and 15 minute Gal4-TAP bands were normalized to the 0 minute band.

Collection of CLK samples. The Gal4 CLK time course was collected the same way as the TBP samples outlined in Chapter II Materials and Methods *Collection of crosslinking time points and preparation of chromatin samples*, with a few alterations. For each time course, 450 ml of cells were grown in SC-URA+2% raffinose (AY146 & AY147 grown in SC-LEU+2% raffinose) until the OD_{600} was ~0.8, at which point the culture was spun down and the cells were resuspended in 450 ml YEP+2% galactose for about an hour until the OD_{600} was approximately 1. A previous graduate student, Ramya Viswanathan, showed that 1 hr in galactose was more than sufficient to induce Gal4 binding to promoters (R. Viswanathan, Thesis, 2012). Each strain (WT and OE) was collected in duplicate and processed as chromatin. Samples were immunoprecipitated, as described in Chapter II for the CLKv2 assay, with Gal4-TA C-10 antibody (sc-1663x; Santa Cruz Biotechnology) for all strains except AY146 and AY147. IP for these two strains was done with a TBP antibody that recognized the C-terminus of the protein (ab61411, Abcam). Real time PCR was performed as described in Chapter II Materials & Methods using primers for the *GAL3* and *GAL1/10* promoters (see Table 4.2).

Table 4.2. Oligonucleotides used in this chapter (5'-3').

Name	Sequence
GAL11_KO-F	AAA GAT CAA GGA TTA AAA CGC TAT TTC TTT TAA ATC TGC TGA TTG TAC TGA GAG TGC ACC
GAL11_KO-R	GTA ACT TCA AAA GTA TCA AAA GTA TGG AAA CTT CAA ATG TCT GTG CGG TAT TTC ACA CCG
GAL11_detect1	TAC ATA TTC CCC CGC TGA TTG
GAL11_detect2	CCT ATT AAT TTC CCC TCG TCA
GAL11_detect3	TCT TTG CAG TAA TTG TTT GGG
GAL3-F (UAS)	CCG AAC ATG CTC CTT CAC TA
GAL3-R (UAS)	GCA TGG CGA TTT CAT TCT TT
GAL1-10-F	GGC ACA TCT GCG TTT CAG GA
GAL1-10-R	GTA CGG ATT AGA AGC CGC CG

Anchor Away strains and sample collection

The HHY183 Gal4 FRB-tagged strain and HHY168 parental control strain (Haruki et al. 2008) were purchased from EUROSCARF. The samples were collected as described in Chapter III Materials & Methods *Anchor Away sample collection*, except for the media used. The 450 ml cultures were grown in YEP+2% raffinose until the OD_{600} was ~0.8, then spun down at pellets were resuspended in YEP+2% galactose for ~1 hour. Samples were processed as chromatin followed by an IP, as described in Chapter II for the CLKv2 assay, with Gal4-TA C-10 antibody (sc-1663x; Santa Cruz Biotechnology). Real time was performed as described in the Chapter II Materials and Methods; strains were analyzed with the *GAL3* and *GAL1-10* ORF primers.

Results

Gal4 dynamics using the CLK assay

Cells were incubated with formaldehyde for varying amounts of time, quenched, immunoprecipitated with a Gal4 antibody, and analyzed by PCR. Two biological replicate were each collected by Savera Shetty and David Auble, then fit with the updated modeling developed by Hussain Zaidi and Stefan Bekiranov used in Zaidi *et al*., 2017. I updated the scripts to include error bars and performed error analysis with scripts provided by Hussain Zaidi. Gal4 binding to *GAL3* was best represented by the full model fit (Figure 4.5) and all possible parameters could be measured or calculated. However, Gal4 at *GAL1-10* was fit with the linear model (Figure 4.6) and therefore the kinetic information was more limited. Dynamic parameters for Gal4 binding to *GAL3* and *GAL1-10* promoter sequences are listed in Table 4.3 below. The residence time of Gal4 at the *GAL3* promoter was 843 seconds, which is in the same range as the original CLK method measurement (Poorey et al. 2013). The overexpression factor used to fit the original CLK data was 2.5, based on western blotting quantification (Poorey et al. 2013). Since a 2μ plasmid carries the overexpressed GAL4 gene, there is potential that the factor could be slightly different in our hands than when the overexpression factor was measured originally (Ludwig and Bruschi 1991), especially since Gal4 is difficult to quantify by western blotting. Given the difficulty in quantifying Gal4 protein levels and the uncertainty in the overexpression factor, the overexpression factor was allowed to float to find the best fit; a value of 5 was settled upon and used for all Gal4 CLK data analysis. The residence time of Gal4 at *GAL1-10* could not be measured since the data was fit with the linear model. A previous study using Gal4 Competition ChIP showed that 75% of Gal4 was competed from its binding site within 15 minutes of competitor induction (Collins et al. 2009), which suggests a residence time similar to what we measured for Gal4 at the *GAL3* locus. The fractional occupancy of Gal4 is similar at both promoters and agrees with the previously published CLK data for *GAL3* (Poorey et al. 2013). The dissociation constants (K_d) for Gal4 at *GAL3* and *GAL1-10* differ by about a factor of four.

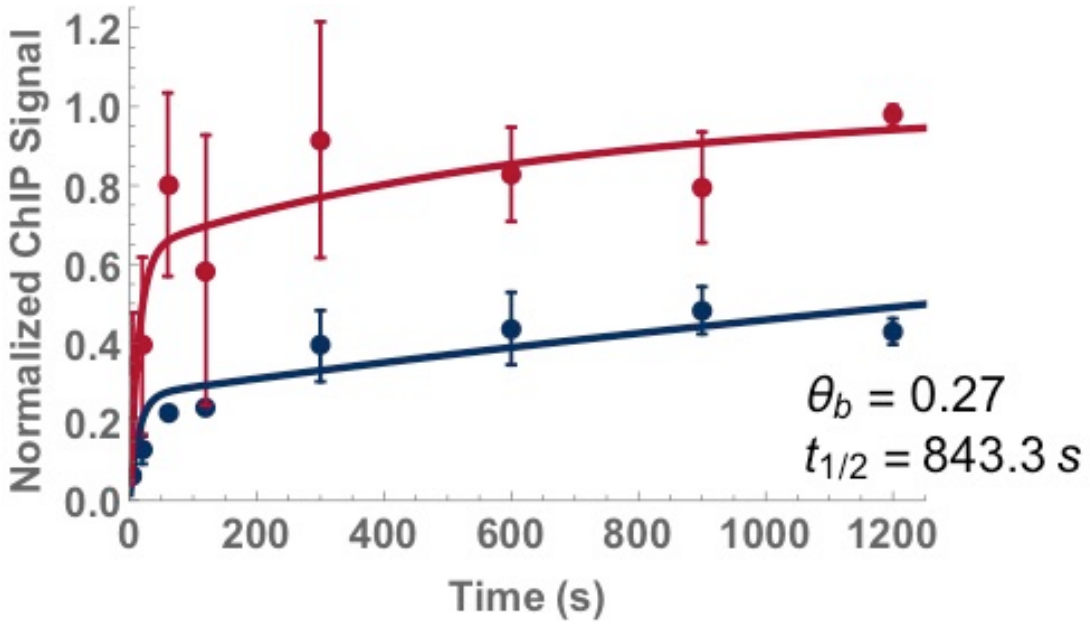


Figure 4.5. Model fit of Gal4 CLK data at *GAL3*. Samples were crosslinked with 5% formaldehyde for the amount of time indicated on the x-axis, then quenched and processed as chromatin samples. Immunoprecipitation with a Gal4 antibody and analysis by real time PCR at the *GAL3* promoter followed. Duplicate collections were averaged for the wild type (blue) and overexpression (red) strains and the data was fit with the CLK full model. A residence time of approximately 843 seconds and a fractional occupancy of 0.27 were calculated from the model fit. The ChIP signal was normalized to the 1200-second time point and error bars represent the standard deviation. This data was collected by Savera Shetty and fit by Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman.

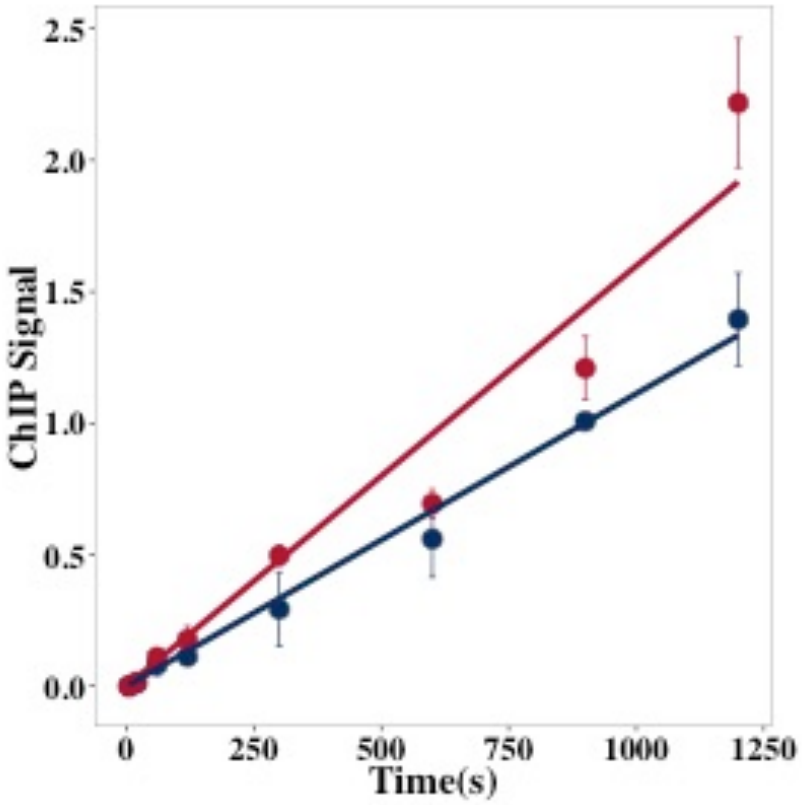


Figure 4.6. Model fit of Gal4 CLK data at *GAL1-10*. Samples were crosslinked with 5% formaldehyde for the amount of time indicated on the x-axis, then quenched and processed as chromatin samples. Immunoprecipitation with a Gal4 antibody and analysis by real time PCR at the *GAL3* promoter followed. Duplicate collections were averaged for the wild type (blue) and overexpression (red) strains and the data was fit with the CLK linear model. This data was collected by David Auble and fit by Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman.

Table 4.3. Estimated kinetic parameters for Gal4 at select loci.

	k_{xl} (1/mol s)	τ_{xl} (s)	k_a*C_{TF} (1/s)	k_d (1/s)	S_{sat}	K_d (k_d/k_a) (mol)	$t_{1/2}$ (s)	θ_b
Gal4 @ *GAL3*	0.0384 (+0.034, -0.026)	10.03	3.08E-04 (+0.0039, -0.0025)	8.22E-04 (+0.0039, -0.0025)	0.243 (+0.026, -0.025)	4.8E-07 (+9.03E-06, -7.47E-06)	843.25 (+61.98, -39.72)	0.27 (+0.059, -0.052)
Gal4 @ *GAL1-10*	($k_{xl}*S_{sat}$) 9.97E-04 (+7.04E-05, -6.58E-05)	386.38 (+27.38, -25.57)	Not estimated	Not estimated	Not estimated	1.10E-07 (+2.15E-07, -1.58E-07)	Not estimated	0.38 (+0.074, -0.066)

Gal4 dynamics using the Anchor Away ChIP-based approach

Similar to TBP, we want to use multiple ChIP-based approaches to measure and compare Gal4 binding dynamics. An Anchor Away strain, HHY183 (Haruki et al. 2008), with Gal4 tagged as the target protein, was treated with rapamycin. Cell culture aliquots were removed at various times, crosslinked with formaldehyde, and quenched in glycine. Chromatin samples were immunoprecipitated and real time qPCR was performed using primers to detect binding to the *GAL3* and *GAL1-10* promoter regions. Data were fit with a single exponential decay curve and apparent dissociation rates and residence times were estimated from the fits (Figure 4.7, Table 4.4). The residence time at *GAL3* was about 2.4 minutes and about 3.6 minutes at *GAL1-10*. The residence time at *GAL3* is shorter than that measured with CLK, but the Anchor Away data do not have associated estimates of error, and as discussed in Chapter III, may be modeled somewhat naively. The depletion of Gal4-FRB-GFP has not been measured yet because individual Gal4 molecules are difficult to image because of their low cell copy number (~300 Gal4/cell, calculated from (Poorey et al. 2013)). However, the residence time for Gal4 binding to *GAL3* is close to that measured by live cell imaging in the Larson lab; the value is slightly more than the 100 seconds burst duration. While burst duration and residence time are not the same thing, they could potentially be correlated; see the Discussion section.

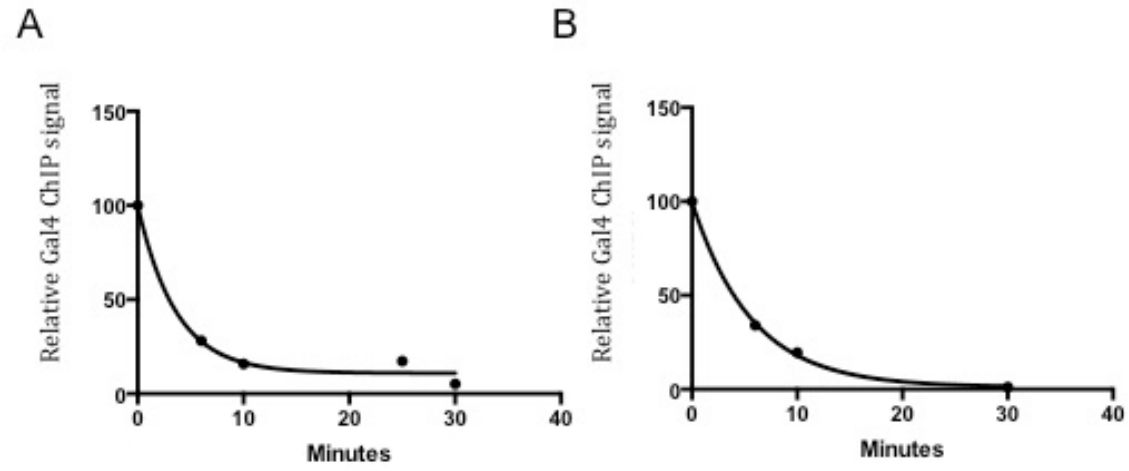


Figure 4.7. Anchor Away plots of Gal4 at two loci. A) Single exponential decay curve fit of Gal4 occupancy of the *GAL3* promoter. Rapamycin was added to cells and aliquots of cells were formaldehyde crosslinked and glycine quenched at the indicated time points. Samples were processed for ChIP and immunoprecipitated with a Gal4 antibody, followed by real time PCR analysis. The ChIP signal of each time point was normalized to the 0-minute sample. Each data set represents the average value for two biological replicate time courses. The x-axis indicates the minutes after rapamycin addition. B) Similar to A, but plot represents depletion of Gal4 at the *GAL1-10* promoter.

Table 4.4. Approximated Gal4 binding dynamics using the Anchor Away approach at select loci.

	k*1*	**k*1* error**	**$t_{1/2}$ (min)**
Gal4 @ *GAL3*	**0.29**	**0.044**	**2.42**
Gal4@ *GAL1/10*	**0.19**	**0.055**	**3.58**

Gal4 dynamics using the CLK assay with a mutated GAL3 UAS_G

There have not been any studies that examine the *in vivo* dynamics of Gal4 at different UAS_G sequences. The divergent *GAL1-10* locus has four binding sites (West et al. 1984), with different binding affinities for Gal4 (Bram et al. 1986). To determine how a modest change in affinity impacts Gal4 binding dynamics in vivo, the *GAL3* UAS_G was replaced with one of the weaker-affinity UAS_G sequences from *GAL10*, the "proximal low-affinity 3" site identified in Bram *et al.* If the binding affinity is weaker, then we expected the on-rate and fractional occupancy to be lower than with the wild type *GAL3* UAS_G; the dissociation rate and residence time may be similar, since the affinity for binding Gal4 is affected, not necessarily the ability to unbind from the site, but they may both be faster since Gal4 stability is weakened. Savera Shetty crosslinked cells with the weakened UAS_G for various amounts of time and quenched them in

glycine. Chromatin samples were immunoprecipitated and real time qPCR was performed using primers to detect binding to the *GAL3* promoter region. Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman fit the data (Figure 4.8, Table 4.5). We found that the crosslinking-limited model was the best fit for the data. The fractional occupancy was 0.027, about 10 times lower than that of Gal4 at the wild type *GAL3* UAS_G. The association and dissociation rates (k_a and k_d) could not be calculated because the crosslinking and TF-binding dynamic regimes timescales were very similar and could not be separated. However, the crosslinking-limited model was fit using the crosslinking rate (τ_{XL}) from the wild type data and k_a and k_d bounds were estimated keeping K_d fixed. The dissociation constant, K_d, is larger for the mutated strain compared to the wild type; theoretically, the association rate could be slower with a similar dissociation rate as the wild type to yield this number, or both rates could change accordingly. Additionally, the residence time couldn't be measured, but an upper bound of 1 second was found. This value is orders of magnitude faster than the residence time of Gal4 when bound to the wild type site.

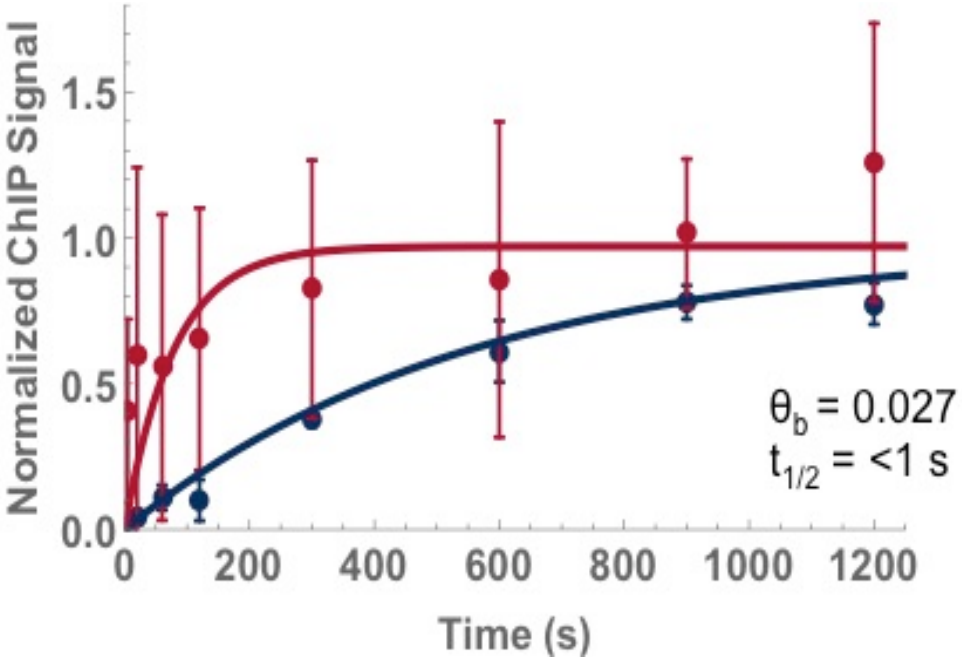


Figure 4.8. Model fit of Gal4 CLK data with mutated *GAL3* UAS_G at the *GAL3* promoter. Samples were crosslinked with 5% formaldehyde for the amount of time indicated on the x-axis, then quenched and processed as chromatin samples. Immunoprecipitation with a Gal4 antibody and analysis by real time PCR at the *GAL3* promoter followed. Two biological replicates were averaged for both the wild type (blue) and overexpression (red) strains and the data was fit with the CLK crosslink-limited model. A residence time could not be calculated, but the upper bound was estimated to be ~1 second. A fractional occupancy of 0.027 was calculated from the model fit. The ChIP signal was normalized to the 1200-second time point and error bars represent the standard deviation. This data was collected by Savera Shetty and fit by Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman.

Table 4.5. Estimated kinetic parameters for Gal4 with a mutated *GAL3* UAS_G at the *GAL3* promoter.

	k_{xl} (1/mol s)	τ_{xl} (s)	k_a*C_{TF} (1/s)	k_d (1/s)	S_{sat}	K_d (k_d/k_a) (mol)	$t_{1/2}$ (s)	θ_b
Gal 4, mutant *GAL3* UAS @ *GAL3*	**Not estimated**	**Not estimated**	**> 0.0187**	**> 0.672**	**0.078**	**6.46E-06**	**< 1**	**0.027**

Gal4 dynamics at wild type and mutated GAL3 UAS_G as measured through microscopy approach

While the previous chapter compared different ChIP-based approaches, we have a collaborator who uses microscopy to analyze Gal4-mediated dynamics of transcriptional dynamics at *GAL* genes. The lab of Dan Larson developed a method to visualize synthesis of individual RNA molecules at specific genes (Larson et al. 2011), such as *GAL3* or *GAL1-10*, which is described in more detail in the Introduction. When RNA is synthesized from a single specific promoter, GFP-labeled PP7 coat protein binds to the stem loops formed in the RNA. The synthesis of each RNA molecule can be imaged in live cells as a spot at the transcription site, and the fluorescence intensity correlated with the number of RNA polymerase II molecules present at

the site (Larson et al. 2011). In the Larson lab, Tineke Lenstra engineered the reporter strains and imaged RNA transcription kinetics induced by Gal4 binding the *GAL3* promoter and Gal4 binding to the same mutated *GAL3* promoter as we used above; preliminary data is shown in Figure 4.9. She used the live cell imaging approach to trace a specific transcription site over time in multiple cells. Bursting was observed in the wild type strain, since the data from these cells was not well approximated by a Poisson distribution. The Poisson distribution fits data that have independent events and since this distribution does not fit the wild type data, the wild type initiation events are dependent on the previous events. Bursting occurs when correlated initiation occurs: one event increases the chance of another event. After autocorrelation, the transcriptional burst duration of the wild type *GAL3* gene was about 100 seconds, while the burst duration in the mutated UAS_G –driven gene was about 70 seconds. Interestingly, the mutated strain seems to have lost the *GAL3* bursting, as the data is best modeled with a Poisson distribution, indicating independent events. There was no correlation between burst duration and Gal4 residence time at the WT and mutated *GAL3* promoters. The Gal4 residence time is about 8 times longer as measured with CLK; the residence time at the mutated locus is at least 70 times longer in the microscopy data than we infer from the CLK data. Potential explanations of this discrepancy follow in the Discussion.

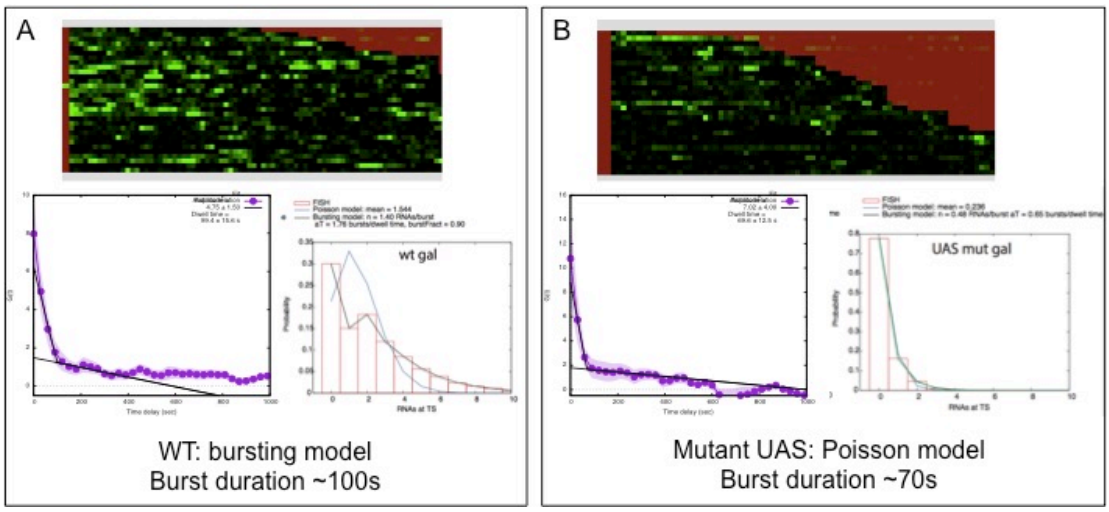


Figure 4.9. Single molecule fluorescence *in situ* hybridization data for Gal4 at the wild type *GAL3* and mutated promoters from Tineke Lenstra. A) Kinetic data from Gal4-mediated transcription at the wild type *GAL3* promoter. Live imaging traces for transcription sites are shown in the top. Each row represents a transcription site and the x-axis represents time imaged; around 30 cells were used for imaging. Red areas indicate traces that were not used for analysis. The bottom left shows the autocorrelations of the live cell data and the bottom right plots the number of RNA transcripts at each site fit with a Poisson (light blue) or bursting (dark blue) model. The calculated burst duration time of ~100 seconds is denoted. B) Identical to A, but for Gal4-mediated transcription at the weakened *GAL3* promoter. The data was best described by the Poisson model with a burst duration of ~70 seconds.

Gal4 dynamics using the CLK assay in a gal11Δ strain

We were interested in the effect of Mediator on Gal4 binding dynamics, as this has not been studied *in vivo* yet and will provide valuable information into how Gal4 recruitment and activation is regulated. It is not known if Mediator recruits Gal4, or vice versa. There is evidence that Mediator acts upstream of Gal4 and indirectly activates Gal4 (see Introduction), but Gal4 could also recruit Mediator once it is activated, like SAGA. Binding to the Mediator subunit Gal11 is essential for Gal4's interaction with Mediator (Jeong et al. 2001) and *gal11Δ* strains show decreased levels of Gal4 protein and Gal4-dependent transcription (Long et al. 1991). It is unclear what the effect of *GAL11* loss would have on Gal4 dynamics since SAGA is also present to directly bind Gal4 through Spt20 (Bhaumik and Green 2001, Larschan and Winston 2001, Bryant and Ptashne 2003) and there may be another subunit of Mediator that directly interacts with Gal4 (Jeong et al. 2001) such as Srb4 (Koh et al. 1998). First, a *GAL11* knockout strain was constructed through PCR-mediated gene disruption (Brachmann et al. 1998). The *GAL11* gene was replaced with the selectable *KanMX4* marker amplified from the pRS400a plasmid (Figure 4.10A, B). Integration was checked by two sets of PCR primers (Gal11_detect1/Gal11_detect2 &

Gal11_detect1/Gal11_detect3); the first set of primers is specific to the insert and the second used the same forward primer, which has homology to the insert for the last five base pairs, and a reverse primer specific to the *GAL11* ORF. It would have been better to use a second primer set that was completely outside the insert, or at least a forward primer outside the insert, to test for presence of the wild type band because lack of a band doesn't indicate that there isn't anything there. Once the strain was constructed, growth on YPD and YEP+2% galactose plates was assessed for the *gal11Δ* strain, CLK WT and OE *gal11Δ* strains, and the parental YPH499 strain (Figure 4.10C). All strains grew similarly on YPD, although the *gal11Δ* strains were slightly slower than YPH499. However, on galactose plates, the growth of the *gal11Δ* strains was compromised compared to the wild type YPH499. Next, cells were incubated with formaldehyde for varying amounts of time, quenched, immunoprecipitated with a Gal4 antibody, and analyzed by PCR. Two biological replicates were collected by Savera Shetty and the data was fit by Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman (Figure 4.11, Table 4.6). The data were best fit by the crosslink-limited model. However, the crosslinking and Gal4-binding time scales were so close that it was hard to separate them kinetically to derive definitive values. The crosslinking rate for Gal4 binding to the *GAL3* promoter discussed earlier was used to find upper bounds on the association (k_a) and dissociation (k_d) rates while keeping the dissociation constant (K_d) fixed. An upper bound of 1.1 seconds was estimated for the residence time and the fractional occupancy was 0.054. The residence time bound is over 800 times faster than the residence time of Gal4 at *GAL3* and the occupancy is also much lower, indicating that loss of the Gal11 subunit may destabilize Gal4 binding to the *GAL3* UAS_G.

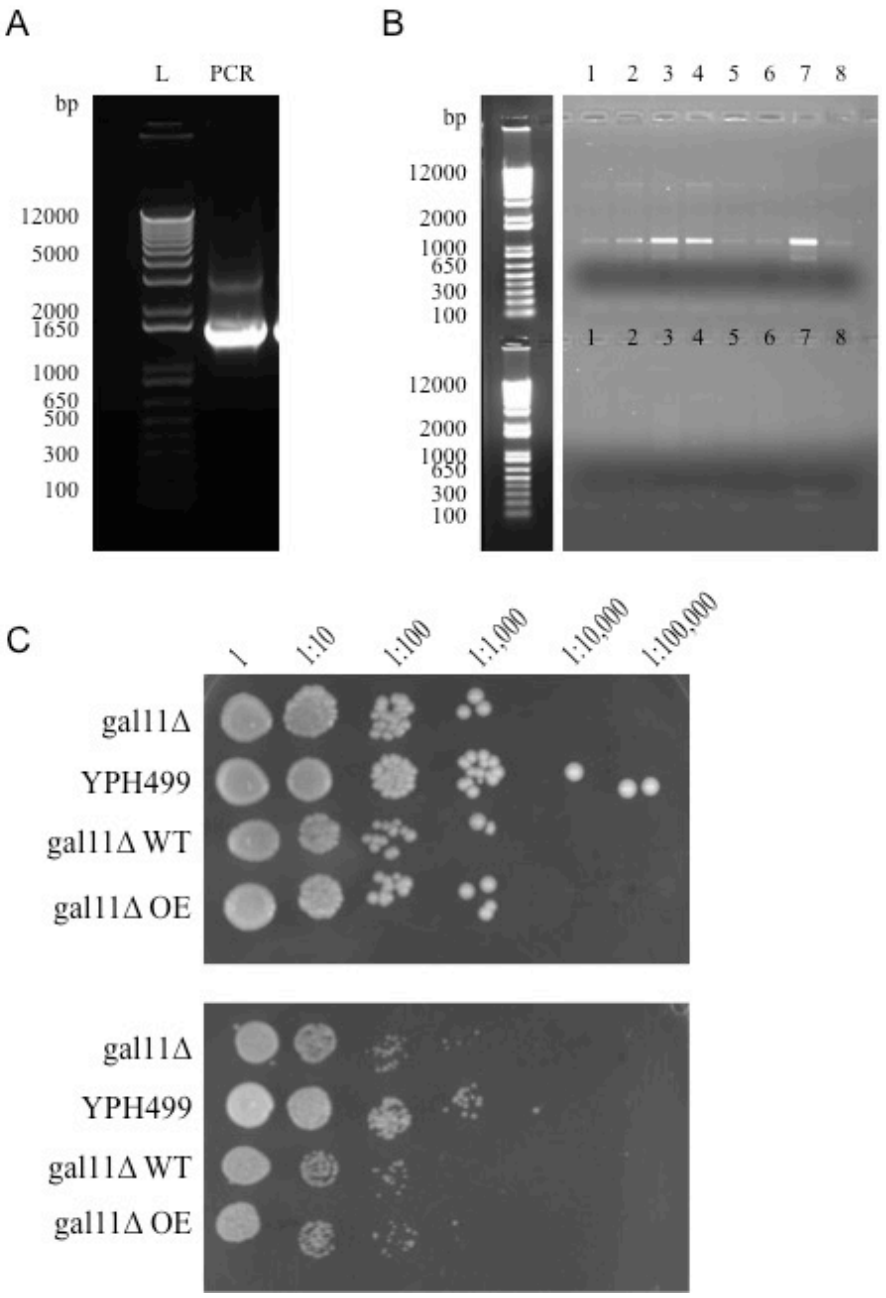


Figure 4.10. Construction of the *gal11Δ* strain through PCR-mediated gene replacement. A) The *KanMX4* selectable marker was amplified from the pRS400a plasmid by primers with added 40-bp homology upstream and downstream of the *GAL11* ORF. The PCR reaction was run on a 1% agarose gel with ethidium bromide staining to verify a 1610 base pair insert. The DNA base pair length is noted to the left of the gel image along the 1 kb-plus DNA ladder. B) After transformation of the *KanMX4* cassette into the haploid YHP499 strain, eight candidates (1-8)

were checked for insert presence by PCR of gDNA and run on a 1% agarose gel with ethidium bromide staining. The top row of samples was amplified using primers specific to the insert and should give a 1079 bp band if the insert is present. The bottom row of samples was amplified with the same forward primer, which overlaps with the insert by five base pairs, and a primer only present in the wild type *GAL11* ORF. No band should be seen for either the wild type or insert. C) Spot test of the *gal11Δ* strain, CLK WT and OE *gal11Δ* strains, and the wild type YPH499 strain. Dilutions of cell culture were plates on either YPD or YEP+2% galactose plates and grown at 30°C for two days before imaging.

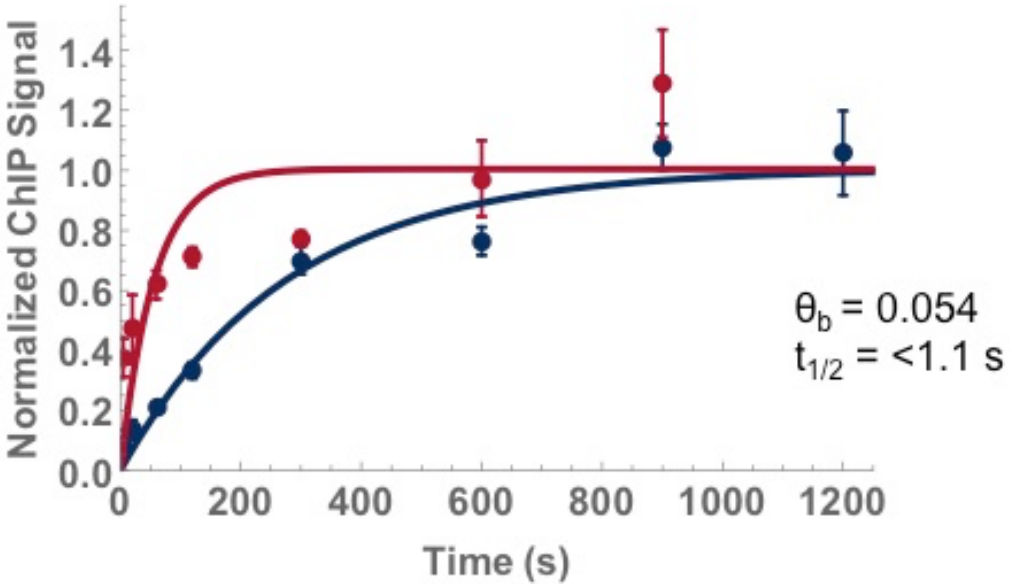


Figure 4.11. Model fit of Gal4 CLK data in the *gal11Δ* strain at the *GAL3* promoter. Samples were crosslinked with 5% formaldehyde for the amount of time indicated on the x-axis, then quenched and processed as chromatin samples. Immunoprecipitation with a Gal4 antibody and analysis by real time PCR at the *GAL3* promoter followed. Duplicate collections were averaged for the wild type (blue) and overexpression (red) strains and the data was fit with the CLK crosslink-limited model. A residence time could not be calculated, but the upper bound was estimated to be ~1.1 seconds. A fractional occupancy of 0.054 was calculated from the model fit.

The ChIP signal was normalized to the 1200-second time point and error bars represent the standard deviation. This data was collected by Savera Shetty and fit by Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman.

Table 4.6. Estimated kinetic parameters for Gal4 at the *GAL3* promoter in a *gal11Δ* strain.

	k_{xl} (1/mol s)	τ_{xl} (s)	k_a*C_{TF} (1/s)	k_d (1/s)	S_{sat}	K_d (k_d/k_a) (mol)	$t_{1/2}$ (s)	θ_b
Gal 4 @ *GAL3* in Δ*gal11* strain	Not estimated	Not estimated	> 0.037	> 0.654	0.0395	3.18E-06	< 1	0.054

TBP dynamics at GAL3 as measured using the CLK assay

TBP is thought to be recruited to promoters after Gal4 binding, as the PIC is needed for transcription initiation at *GAL* genes. We were interested in measuring the dynamics of TBP at these genes to determine if there was a difference in binding dynamics at these promoters compared to those we looked at in Chapter II to learn more about transcriptional regulation in the presence of an inducible activator. The TBP CLK WT and OE strains were grown in raffinose selective media and induced in galactose for an hour before sample collection. After immunoprecipitation and real time PCR, the data were fit with the full model and dynamic parameters were estimated (Figure 4.12, Table 4.7). We focused on the *GAL3* promoter, but other binding sites will be investigated in future work. The measured residence time for TBP at *GAL3* was 196 seconds, which is within a factor of two of the residence times of TBP at *ACT1*, *URA1*, and *LOS1* measured from full model fits in Chapter II. The fractional occupancy of TBP at *GAL3* was 0.05, which is in agreement with the CLKv2 data for TBP at *ACT1*, *URA1,* and *LOS1* promoters in Chapter II. The crosslinking rate, association rate, and dissociation rate are also all in agreement or within a factor of two compared to the Chapter II measurements, which were not generated after inducing an activating factor. Based on these data, it appears that TBP has similar

dynamics at all four promoters (*ACT1*, *LOS1*, *URA1*, *GAL3*). However, the sample size is very low and a better understanding will come from analysis of additional *GAL* loci and also more loci genome-wide. There is genome-wide data for TBP collected by Savera Shetty and fit by Stefan Bekiranov, but it has not been fully analyzed yet for comparison, as the background correction still needs to be worked out due to redistribution of TBP in the overexpression strain across the gene instead of mainly at the promoter.

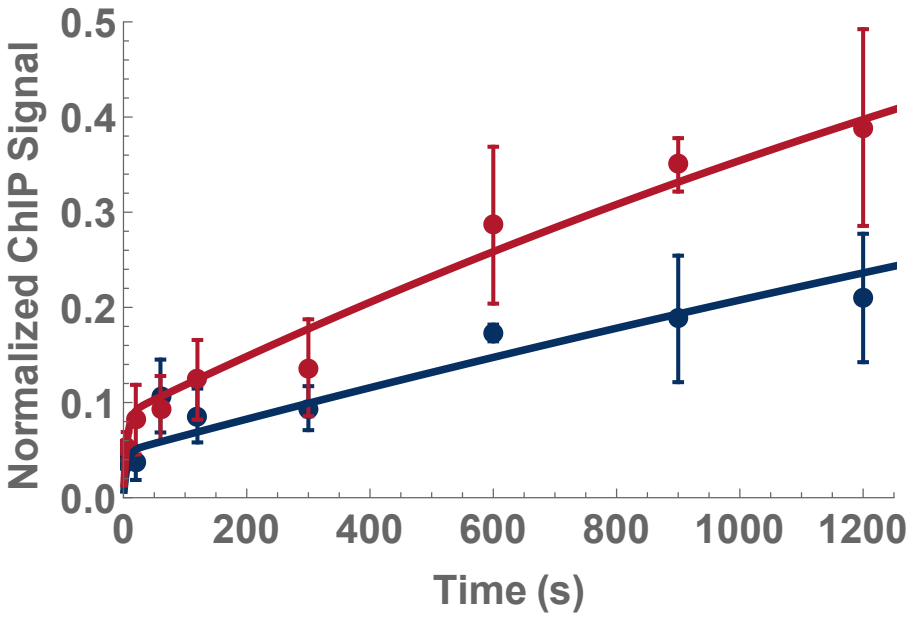


Figure 4.12. Model fit of CLK data for TBP binding to the *GAL3* promoter. The TBP CLK strains were grown in YEP+2% raffinose, then induced for 1 hour in YEP+2% galactose. Samples were crosslinked with 5% formaldehyde for the amount of time indicated on the x-axis, then quenched and processed as chromatin samples. Immunoprecipitation with a TBP antibody and analysis by real time PCR at the *GAL3* promoter followed. Duplicate collections were averaged for the wild type (blue) and overexpression (red) strains and the data was fit with the CLK full model. The residence time was 196 seconds and a fractional occupancy of 0.05 was calculated from the model fit. The ChIP signal was normalized to the 1200-second time point and

error bars represent the standard deviation. This data were obtained by Savera Shetty and fit by Hussain Zaidi, Stefan Bekiranov, and Elizabeth Hoffman.

Table 4.7. Estimated kinetic parameters for TBP at the *GAL3* promoter.

	k_{xl} (1/mol s)	τ_{xl} (s)	$k_a * C_{TF}$ (1/s)	k_d (1/s)	S_{sat}	K_d (k_d/k_a) (mol)	$t_{1/2}$ (s)	θ_b
TBP @ *GAL3*	0.097 (+0.99, -0.21)	3.97	1.87E-04 (+0.0014, -0.00091)	3.53E-03 (+0.0038, -0.0032)	0.566 (+0.45, -0.31)	3.4E-06 (+0.0014, -9.5E-04)	196.36 (+7.32, -6.15)	0.05 (+0.051, -0.36)

Discussion

Gal4 binding dynamics at GAL3 and GAL1-10 using ChIP-based techniques

The dynamics of Gal4 binding to *GAL3* were previously measured using the original CLK method (Poorey et al. 2013). Using the improved crosslinking/quenching conditions and fitting methods, we measured binding dynamics to this locus again and found values that were largely consistent with the earlier measurements. The residence time and fractional occupancy were similar in both sets of measurements; even though the original quench conditions were not as robust as in the improved method (Zaidi, Hoffman et al. 2017), it is remarkable that the original method yielded consistent kinetic values. Gal4 had a residence time of ~14 minutes at *GAL3 in vivo* based on the CLKv2 results. The measured residence time of Gal4 at *GAL3* measured from the new data is agreement with the Collins *et al.* competition ChIP study demonstrating that ~75% of bound Gal4 was replaced by the competitor in ~15 minutes (Collins et al. 2009), but in opposition to the earlier competition study claiming Gal4 was bound promoters for upwards of an hour (Nalley et al. 2006). These two studies investigated binding to the the *GAL1-10* binding sites and not *GAL3*, but the Gal4 residence time is similar to the residence time at *GAL3*. Since the CLKv2 data for Gal4 binding to the *GAL1-10* locus was best

fit with a linear model, the residence time could not be extracted, so no comparison can be made to the Nalley *et al.* and Collins *et al.* studies. Additionally, the dissociation constants of Gal4 at *GAL3* and *GAL1-10* reported here differs from previously measured *in vitro* rates by an order of magnitude (Giniger and Ptashne 1988, Parthun and Jaehning 1990), but is in agreement with the original CLK study (Poorey et al. 2013), suggesting a slower dissociation of Gal4 *in vivo* than *in vitro*. This makes sense, as the cellular milieu *in vivo* is more complex than an *in vitro* experiment and transcription factor binding could be influenced by more factors or environmental conditions that are not present *in vitro.*

The Anchor Away approach was also used to measure binding dynamics of Gal4 at *GAL3* and *GAL1-10*. The apparent residence time of Gal4 at both *GAL3* and *GAL1-10* promoters was ~4 times faster than that measured for Gal4-*GAL3* with CLKv2. The nuclear depletion rate of Gal4 has not been determined yet, either by using a fluorescently labeled protein and microscopy or by isolating rapamycin-treated nuclear extracts. Previous attempts to measure nuclear depletion by running nuclear extracts on westerns were unsuccessful and Gal4 is difficult to visualize by fluorescent tagging because of its low copy number. However, a collaborator has successfully labeled Gal4 with the HALO tag (Los et al. 2008, Stagge et al. 2013) and this could provide a method for measuring nuclear depletion in the future. Until this depletion rate is measured, we won't know the bound of binding dynamics that we can measure. A limit of this method is the time it takes for nuclear depletion to occur. If the off-rate is longer than the depletion rate, then we can measure the off-rate accurately. Measuring the depletion rate of the factor can help us determine if we are hitting a wall because depletion is slow and therefore figure out if the binding dynamics could be much faster than what was measured

Gal4 binding dynamics were also measured using CLKv2 for binding to the *GAL3* promoter while harboring a weakened *GAL3* UAS_G. The CLKv2 data were best fit with the crosslink-limited single exponential model, which cannot yield a residence time. However, an upper bound was approximated by changing the k_a and k_d values generated from a full-model fit

and keeping the K_d value (from the crosslink-limited fit) constant. The model fit indicates that the formaldehyde crosslinking and binding dynamic regimes are overlapping, which makes it extremely difficult to extract some parameters of interest. The data can be fit with the full model, but previous data collected with 1% formaldehyde illustrates the formaldehyde dependence of the Gal4-*GAL3* with the weakened UAS_G (data not shown). If there were no dependence on formaldehyde concentration, the data with 5% and 1% formaldehyde would overlap. However this is not the case; consistent with crosslinking being limited, the ChIP signals obtained using 1% formaldehyde decrease ~5-fold compared to the ChIP signals obtained using 5% formaldehyde. Therefore, the crosslinking-limited fit was used for fitting this dataset. Regardless, the fractional occupancy of Gal4 at the weakened *GAL3* was ~10 times less than at the wild type *GAL3* UAS_G, indicating that mutating the UAS_G reduces the occupancy of Gal4, consistent with the reduced affinity of the protein for its binding site. While the crosslinking rate could not be determined, there could also be a difference in crosslinking efficiency due to differing promoter sequences. There are six guanine residues in the wild type UAS_G, but only four in the weakened binding site. Deoxyguanosine crosslinking to lysine gave the highest yield in a previous study (Lu et al. 2010) and could potentially contribute to a lower crosslinking efficiency of Gal4 with the weakened *GAL3* promoter. The residence time is also drastically reduced for the Gal4-*GAL3* interaction in the mutant compared to wild type. It is surprising that this decrease is so large. When the central T-A base pair in the consensus sequence is removed, Gal4 binding affinity is about 25 times lower than for the WT binding site *in vitro* (Vashee et al. 1993), but the weakened UAS_G still contains the T-A pair. It's possible that Gal4 binding stability is more adversely affected *in vivo* than indicated by *in vitro* studies.

Gal4 binding dynamics at GAL3 using microscopy techniques

As discussed earlier in this chapter and in previous chapters, binding dynamics measurements are difficult to make and therefore convergence on consensus values using

multiple independent methods is critical. Tineke Lenstra and Dan Larson at NCI used a microscopic approach to the question of Gal4-mediated RNA synthesis dynamics. RNA is fluorescently labeled as it is transcribed and the intensity of the transcript can be correlated to the number of RNAPII molecules at that transcription site. The dwell time of the RNA can also be determined by measuring the length of transcriptional bursts. At *GAL3*, the transcript has a dwell time of ~100 seconds; our thought was that dwell time and residence would be similar, but this may not be the case. Gal4 may bind to promoters for multiple rounds of transcription, as suggested in the first CLK paper (Poorey et al. 2013), so multiple transcript dwell times would accumulate before Gal4 dissociates from the UAS_G. The data shows that Gal4-mediated transcription fits a bursting model (Lenstra et al. 2015), so multiple rounds of transcription may occur before a new Gal4 initiates transcription from that site. This approach was also done to measure the dwell time of transcripts at the weakened *GAL3* UAS_G that we used earlier in CLK experiments. The dwell time was found to be ~70 seconds, which is much longer than the residence time measured by CLK, but shorter than the wild type; this trend was also observed in the CLK data. Since we don't have an accurate measurement for the residence time of Gal4 at this weakened UAS_G, it's hard to say how close these measurements are. The error on the microscopy dwell time is 12.5 seconds, which still doesn't put it close to the upper bound of 1 second for the CLK data; the error on the CLK residence time bound is not known. Either way, the trend of shorter Gal4 binding to the weakened UAS_G is observed in both data sets. The microscopy data indicates a loss of bursting transcriptional dynamics, which presumably would also occur in the CLK data if Gal4 were only transiently bound to the promoter. Future studies to measure nascent RNA with GRO-seq (Core et al. 2008) or NET-seq (Churchman and Weissman 2011) in this strain with a weakened binding site will be very helpful.

Gal4 binding dynamics in a gal11Δ strain

Gal4 binds to subunits of both the SAGA and Mediator complexes. We hypothesized that in a strain where the *GAL11* subunit was knocked out, Gal4 binding would be more dynamic without *GAL11* present to stabilize it, but didn't know the extent of this decrease since SAGA is still present as a scaffold. When CLK was used to measure the kinetics at the *GAL3* locus in this mutated strain, the data was best fit with a crosslink-limited model with residence time with an upper bound of ~1 second and a very low fractional occupancy. We found the residence time for Gal4 binding to the weakened *GAL3* UAS_G was much shorter than Gal4 residence time at the wild type locus and about one-fifth as many sites were bound when Mediator was compromised. It appears that the Gal11-Gal4 interaction is important for Gal4 binding to the UAS_G sequence in vivo. There was also a growth defect on galactose plates in the knockout strain, which would agree with destabilization of Gal4 under conditions that require *GAL* gene expression; previous work has shown that *GAL* gene expression in a *gal1Δ* strain is about 10-15% of the observed levels in a wild type strain ((Jeong et al. 2001), David Auble, unpublished). Growth was only slightly compromised in the *gal1Δ* strains on glucose plates compared to the wild type strain, as expected (Shi et al. 1996), presumably because SAGA acts as a scaffold to stabilize Gal4 binding (Traven et al. 2006) and Gal4 potentially interacts with another subunit of Mediator (Koh et al. 1998). In the future, it would be interesting to determine what effect a knockout of the SAGA Spt20 subunit would have on Gal4 binding kinetics. There is conflicting evidence on the necessity of SAGA for Mediator recruitment; one study found it was necessary (Bhaumik et al. 2004), but several have show Mediator can be independently recruited (Larschan and Winston 2001, Bryant and Ptashne 2003, Lemieux and Gaudreau 2004). SAGA is thought to stabilize Mediator and is required for PIC formation (Traven et al. 2006); presumably the dynamics of Gal4 would be more unstable in a SAGA-compromised strain than in the Mediator mutated *gal11Δ* strain.

TBP dynamics in response to Gal4 recruitment

A goal of measuring binding dynamics is to integrate kinetic behavior into the larger scheme of transcriptional regulation. One approach for beginning to address how binding dynamics relate to transcriptional regulation is to determine how TBP-DNA binding dynamics vary at different promoter classes. The binding dynamics of TBP at the *GAL* genes was not reported in the genome-wide Competition ChIP study due to the nature of the induction strain requiring pGAL1. (van Werven et al. 2009, Zaidi et al. 2017), so we measured TBP binding to *GAL3*. We wanted to know if there was a different in binding dynamics at the inducible activator-driven promoter compared to the promoters we looked at in Chapter II. The residence time of TBP at *GAL3* was within a factor of two compared to the value measured for full model fit promoters in Chapter II, and the occupancy and crosslinking rates were approximately the same for all promoters. The dissociation constant differed by several orders of magnitude between the measurement for TBP binding at *GAL3* and *URA1*/*LOS1*/*ACT1*, and the k_a and k_d values are off by a factor of two or three, so there is some difference in binding efficiencies between the glucose and galactose induced promoters. Overall, the TBP binding dynamics are fairly similar at *GAL3* compared to earlier loci tested in Chapter II.

Conclusions and Future Directions

Although the *GAL* system has been extensively studied, only a few studies have investigated the *in vivo* dynamics of the Gal4 protein and prior to this work virtually nothing was known about the dynamic behavior of factors that physically or functionally interact with Gal4. Our study aimed to measure kinetic parameters using several independent techniques, including ChIP-based and microscopic approaches. Using the CLKv2 assay, we measured kinetic parameters for Gal4 binding to the *GAL3* and *GAL1-10* promoters, as well as to *GAL3* in a strain with a weakened *GAL3* UAS_G and in *gal1Δ* cells. Anchor Away measurements were made for Gal4 binding to *GAL3* and *GAL1-10*, which is an improvement over previously published work

that did not fit the measured decay in ChIP signal to any model. Eventually, this approach should be used to collect data for all strains and promoters that CLKv2 has been used for, including the weakened *GAL3* UAS_G and a *gal1Δ* strain. Our collaborators, Dan Larson and Tineke Lenstra, measured the transcriptional dwell time at the *GAL3* promoter in the wild type and weakened UAS_G strains. Comparing this dwell time to our ChIP-based dynamics data can provide important biological information about how transcription factor binding dynamics relate to transcriptional output. If a factor has a short residence time compared to the locus activation, this suggests the factor has a catalytic role at an early step. If the residence time of a factor is similar to the dwell time, this would indicate the factor needs to be bound to the promoter the entire time the gene is activated. Competition ChIP for Gal4 at *GAL1/10* was previously performed in a somewhat controversial study (Nalley et al. 2006) and repeated with a difference competitor a few years later with different results (Collins et al. 2009). We would like to construct a Gal4 Competition ChIP strain and measure the binding dynamics at all *GAL* genes by fitting with the updated model from Zaidi *et al.* (Zaidi et al. 2017). Since the strain constructed for TFIIB relies on induction of an alternatively tagged allele with the galactose promoter, a different approach will have to be used for induction, such as the Tet system (Gossen et al. 1995, Bellí et al. 1998), but the same differential tagging approach can be used. Iterations of this strain can be constructed to match the datasets collected by CLKv2, such as a weakened *GAL3* UAS_G and *gal1Δ* strain, in order to compare dynamic measurements across multiple independent approaches.

As mentioned earlier, it would also be interesting to measure Gal4 dynamics in a strain with a SAGA knockout, such as an *spt20Δ* strain to better understand the effect that SAGA has on Gal4 stability and regulation *in vivo*. While Gal4 directly interacts with Tra1 in the SAGA complex (Bhaumik et al. 2004), Tra1 is an essential protein and haploid knockout strains are inviable (Saleh et al. 1998). Spt20 is required for complex integrity (Grant et al. 1997, Sterner et al. 1999) and Tra1 recruitment to a *GAL1* UAS_G is significantly reduced in an *spt20Δ* strain (Bhaumik et al. 2004). If Gal4 binding dynamics in both Mediator and SAGA strains are

measured, this could answer some unsolved questions in the field. For example, does Gal4 directly interact with Mediator *in vivo*? Is SAGA required for Mediator contact with Gal4 *in vivo*? Are there different requirements at different sets of *GAL* genes? What is the order of recruitment/assembly of Mediator, SAGA, and Gal4 and does this order change at different genes? Comparing the dynamics in an *spt20Δ* strain to those in a *gal11Δ* strain can also provide insight into how much each complex contributes to Gal4 binding dynamics and if they provide redundant roles. If this data is also coupled with transcriptional data, more could be learned about the requirements for regulatory complexes in creating a productive promoter and how they contribute to transcriptional bursting. The effect of certain binding dynamics on transcriptional output can be an extremely helpful tool when using the *GAL* system to artificially induce a gene, as different mutants could be used to control or induce certain levels of gene expression or phenomena like bursting.

This dissertation has focused on ChIP-based methodologies, and when consensus Gal4 dynamic parameters are obtained, it will be possible to turn attention more fully from methodological and technical concerns to answering biological questions, using methods appropriate for the system and time-scale of the interactions being measured. These techniques yield measurements that are pushing the field forward and have the power to answer important biological questions about how activators assemble and regulate transcription, as well as how different complexes and parts of the transcriptional machinery are recruited. Combined with transcriptional data, we can greatly improve our knowledge of how the complicated process of transcription works in cells.

CHAPTER V

Future Directions

The previous chapters discuss development and application of three independent ChIP-based approaches for measuring protein-DNA binding dynamics. These measurements are important to make because of the biological questions that we can answer with the information gained. For example, questions about the order of assembly, stability contribution from different factors across promoters, and the effect of binding on RNA synthesis can be addressed. The following chapter will briefly discuss some limitations of the methods and then provide examples of a few biological questions and scenarios that can be addressed with the measurements obtained from these techniques.

Method Limitations

Going forward, there are more control experiments that ought to be done for the Anchor Away technique especially. So far, we have assumed that the measured dynamics represent the fraction of TFs that are bound to DNA and are then exported by the ribosomal protein fusion once they dissociate. However, there is an unbound pool of protein present that may or may not be depleted quickly. If this pool is depleted quickly (much faster than the off-rate of the bound fraction), then our underlying assumption is correct. However, if this pool is depleted more slowly or on the same time scale as the bound TFs, then the dynamics measured are a mix of the two pools and do not accurately represent the binding of the TF of interest. Figure 5.1 shows these two scenarios. One way to determine how quickly the unbound fraction is transported out of the nucleus is to tag the ribosomal anchor protein (Rpl13Ap) (Haruki et al. 2008) with GFP and measure its depletion rate in cells grown in rapamycin using microscopy, as was done for TBP in Chapter III. If Rpb13Ap-GFP is exported out of the nucleus much faster than the measured depletion rate of TBP (~11 minutes), this would indicate that the unbound protein is also leaving the nucleus quickly. A strain could also be engineered that labels both the target protein, such as

TBP, and the ribosomal protein with different fluorophores (ex. TBP-GFP and Rpl13Ap-mCherry) to image the cytoplasmic trafficking of both concurrently. While we would not be able to differentiate between chromatin-bound TBP or non chromatin-bound TBP, we could track co-localization of the proteins and check for presence of two groups: a collection of co-localized proteins that are exported quickly (for example, seconds) and a co-localized cohort that is exported more slowly (for example, minutes). If two groups are observed, this suggests that the unbound fraction is in fact getting exported more quickly than the fraction bound to chromatin. However, if a clear distinction in depletion times is not seen to indicate two different groups, this suggests that the fraction bound to chromatin and the non chromatin-bound fraction are exported at similar rates from the nucleus. The implication then is that the previously measured depletion time for TBP (~11 minutes) is measuring both fractions, that bound to DNA and that unbound and free-floating in the nucleus. The depletion time incorporated into the future modeling equations should represent the depletion of the bound fraction alone, and if this is not the case, the fact that this rate consists of both fractions should be kept in mind when interpreting data.

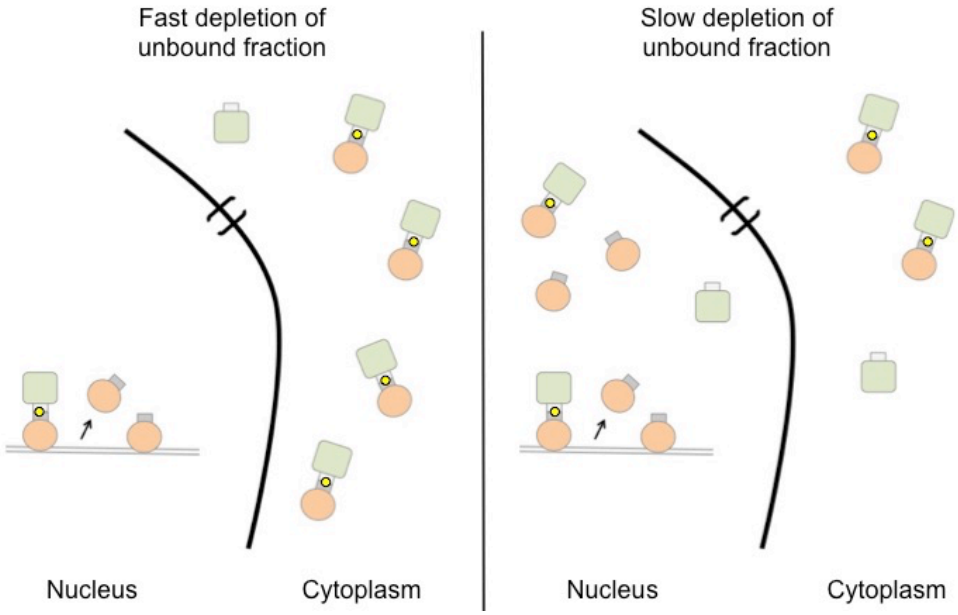


Figure 5.1. Schematic of two potential interpretations of the Anchor Away depletion. There are two populations in the Anchor Away experiments: chromatin bound and non-chromatin

bound (unbound) protein. The left side of the panel shows a fast depletion of the unbound (orange) protein by the cytoplasmic protein (green) when the two are heterodimerized with rapamycin (yellow). The unbound fraction quickly moves from the nucleus to the cytoplasm. The bound protein fraction remains on DNA (gray bar) until it is depleted. The right side of the panel shows a slower depletion of the unbound fraction; some of the unbound fraction remains in the nucleus and some is transported into the cytoplasm while TBP remains bounds to DNA.

The high concentration of formaldehyde is potentially a limiting variable when exploring dynamics of other factors, especially with CLKv2. As seen in previous chapters, this high concentration can deplete a given factor over the experimental time course, which can limit the factors and promoters we can extract all dynamic parameters from. There are ways to work around this by using a shorter time course or, in principle, incorporating the depletion rate into the modeling equations. For each new factor that is tested, the optimal formaldehyde concentration should first be determined. While 5% formaldehyde worked for TBP, a lower concentration of formaldehyde may be optimal for others. Additionally, glyoxal should be tested as a crosslinker. This molecule uses ammonium chloride instead of glycine for quenching (Richter et al. 2018), so a smaller volume of quencher may be sufficient, making experiments easier to conduct. Additionally, glyoxal is suggested to preserve cellular structure better than formaldehyde (Richter et al. 2018), which is an important advantage. Another suggestion is to resuspend cells in water instead of rich media when we're preparing to crosslink them in order to improve accessibility of formaldehyde to proteins without interference from media components. This is an interesting idea because if formaldehyde is not depleted by reactivity with constituents in the media, then a higher concentration will be available for crosslinking proteins to chromatin and the overall crosslinking rate of complexes of interest will be increased.

It's also important to consider the experimental formaldehyde concentration when interpreting CLKv2 data, especially when comparing occupancies of different factors. For example, TFIIE has different occupancies than TBP at *URA1* and *LOS1*; this could mean that TFIIE truly has a higher occupancy than TBP, which is a little surprising since TBP and TFIIE should be equally present at promoters as they're both components of the PIC, or this could be the occupancy of crosslinked complexes and not what we naively thought was the occupancy in the unperturbed state. The occupancy is the fraction of available promoters; once TBP binds and is crosslinked with formaldehyde, some of the sites may no longer be able to bind TFIIE (assuming TBP is required to bind first before TFIIE can bind); formaldehyde could cause base flipping or crosslink TBP or other necessary factors in ways that block TFIIE access. The number of available sites for TFIIE may be smaller than that for TBP but TFIIE could bind to a similar number as TBP; this could raise the measured fractional occupancy for TFIIE, but in cells without formaldehyde, the occupancies are actually similar.

What can we learn from these measurements?

While being able to measure *in vivo* dynamics of proteins is a significant technical advance, the most exciting outcome is what we can then learn about transcriptional regulation. Using the dynamics measured in combination with techniques like live-cell imaging or GRO-seq (see Chapter IV), the dynamics can be coupled with RNA synthesis to determine assembly and contribution of PIC components on transcriptional output.

Much is still unknown about the transcriptional machinery *in vivo*: are activators or Mediator recruited first and what are the dynamics of the interactions between these components? How does Mediator influence and stabilize PIC assembly? How does PIC assembly vary across genes? Figure 5.2 provides a simplified overview of some components of the transcriptional machinery and their potential dynamics. Activators can bind chromatin quickly or more slowly (see Chapter I; reviewed in Hager et al. 2009) and can also bind Mediator. Mediator interacts with

GTFs that assemble to form the PIC, either on a promoter with a TATA consensus sequence (shown) or on a non-consensus sequence. How these GTFs assemble and the dynamics of PIC complex assembly *in vivo* are some of the major questions we're interested in answering. A double arrow is shown between components in Figure 5.2 to indicate the unknown order of recruitment. By measuring dynamics of these individual components in wild type strains and also in strains compromised for another component, we can relate their kinetics to recruitment order and better understand the role of other components on stability of TF binding.

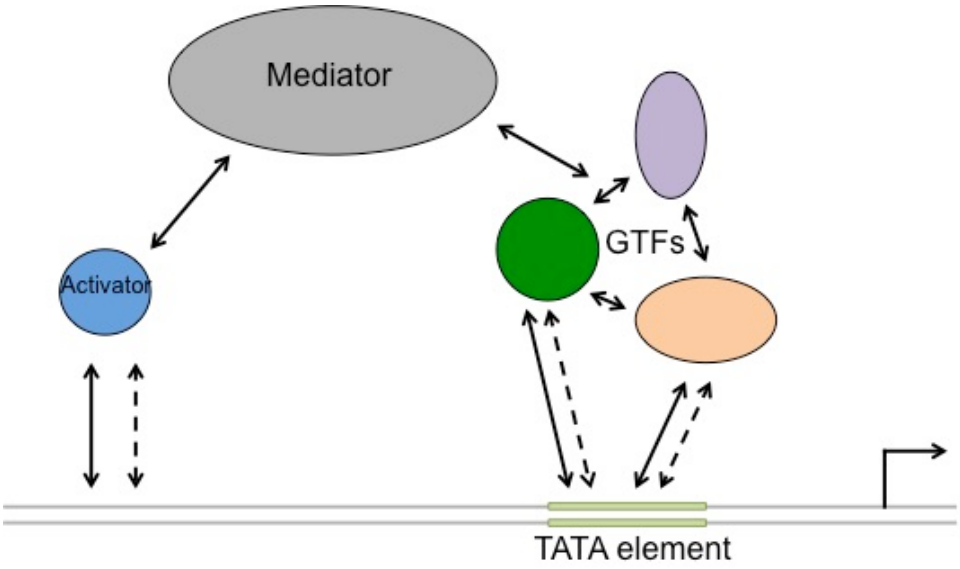


Figure 5.2. General schematic of proteins that comprise or contribute to formation of the PIC and their possible rates of association and dissociation. An activator protein, the Mediator complex, and multiple general transcription factors (GTFs) are required for PIC assembly on DNA. The order of their assembly on chromatin is unclear, and they could either assemble quickly (solid line) or more slowly (dotted line). Additionally, how the GTFs assemble to form the PIC is not know (indicated by double arrows). A TATA element is shown where the PIC binds, but it is not necessary for PIC formation or productive transcription.

One approach to determine the effect of Mediator on the stability of activators or PIC components would be to use a strain with a Mediator defect, such as the *gal11Δ* strain that was used in Chapter IV or an Anchor Away strain that depletes an essential Mediator subunit, and measure dynamics of key GTFs like TBP and TFIIE at particular activator controlled genes, such as the Gal4 dependent genes. If Mediator is required first and then recruits PIC components, it is expected that the occupancy of the GTFs would be much lower in the mutated strain and a much shorter residence time would be measured than in the wild type strain. However, this could indicate two things: Mediator could be 1) required to bind first, or 2) necessary for stabilization of the GTFs. Since Mediator is required for all Pol II transcription (see Chapter I), I would expect at least the residence time of GTFs to be faster in the compromised strain compared to the wild type time. The occupancy may also be lower in the mutant strain compared to wild type. At TATA consensus-containing genes, there may be more of an effect on residence time and/or occupancy (i.e. faster residence time, lower occupancy compared to TATA-less genes) since TBP and Mediator interact. Genes without a TATA-consensus may not be as affected due to stabilization of TFIID binding required for PIC formation. Additionally, dynamics of activators (such as Gal4) could be measured. If the residence time and/or occupancy are unchanged in the mutant strain, this would indicate that Mediator is recruited to DNA after activator binding. If residence time/occupancy are changed and found to be faster/lower, respectively, this would suggest that Mediator recruits the activator and it necessary for its stability. By measuring the dynamics of different parts of the assembly pathway from activator to GTFs, we can understand how each component affects the stabilization of others and generally the order in which they assemble genome-wide to determine how different gene classes are regulated.

Another example of what can be learned from these measurements is illustrated in Figure 5.3. With residence times of two factors, such as TBP and TFIIE, we can compare them to understand how they assemble in the PIC and develop models for how they contribute to transcription. If TBP and TFIIE have similar residence times, they could be binding DNA as a

pre-assembled complex or individually joining the PIC on a similar time scale. If live imaging of RNA synthesis was then done, the dwell time of RNA at a specific locus measured from the imaging could be compared to the residence time of TBP and TFIIE. Say for example, the residence times of TBP and TFIIE are about 100 seconds. If the dwell time of RNA is also about 100 seconds, then this could indicate that the PIC components remain until the synthesis of the RNA molecule is complete and they then dissociate and reassemble to start another round of transcription. If the RNA dwell time at the given locus is longer than the residence time, say 500 seconds, the PIC components do not need to be bound the entire length of synthesis and can bind, initiate transcription, then dissociate. If the dwell time is shorter than the residence time of the factors, 20 seconds for example, then the results would suggest that the PIC components stay bound through multiple rounds of transcription before dissociating, indicating a potential role for a stable PIC scaffold in bursting activity (see Chapter IV; Zenklusen et al. 2008). Correlating residence times of PIC components with the RNA synthesis time scale can provide critical information about how these factors regulate and influence transcriptional output. Comparing these values genome-wide would have the potential to yield a wealth of information about the macromolecular assembly processes and regulatory mechanisms used in vivo.

A future goal once dynamics have been measured for key factors in the PIC (TFIIA, TFIIB, TFIIE, TFIIF, TBP) is to work out the assembly pathways at specific genes in order to identify differences in steps that relate to regulation, or regulatory capacity. Working out an assembly pathway is difficult using only kinetic data because it is hard to determine linear assemblies, but can potentially be done with the aid of structural data, knockout data, and modeling. Figure 5.4 shows a few simple examples of how kinetic data can begin to conceptualize assembly pathways; combining all the PIC factors will be much more complicated. For example, if two factors, named A and B, have a residence time of 120 seconds and 20 seconds, respectively, it's possible that factor A can bind first followed by factor B. Linear

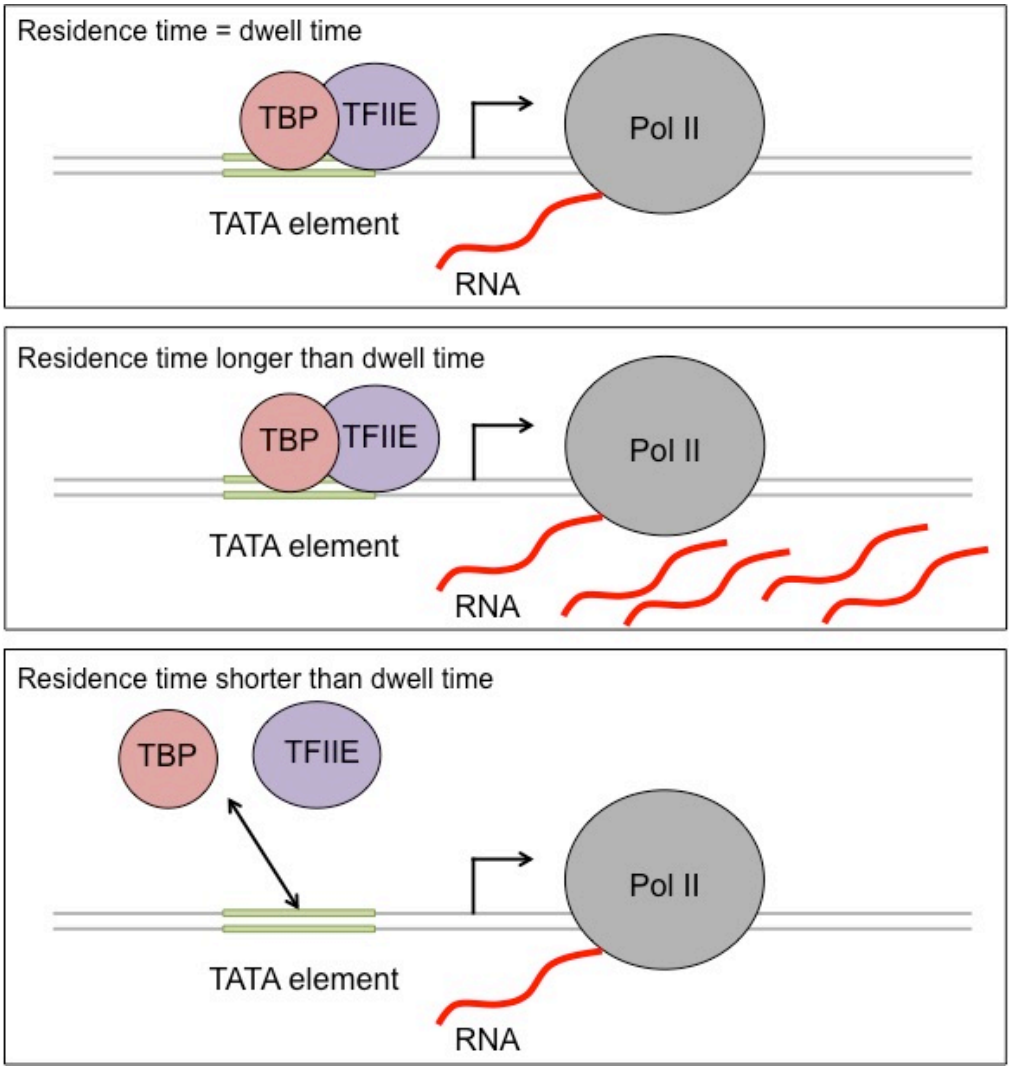


Figure 5.3. Examples of how correlation of factor residence time and RNA synthesis dwell time can provide important biological insight into transcriptional regulation. The top panel illustrates a situation where the residence times of TBP and TFIIE are approximately equal to the dwell time of Pol II; one RNA molecule is made for each round of TBP and TFIIE binding. The middle panel shows a simulation where TBP and TFIIE residence times are much longer than the dwell time of Pol II; multiple rounds of RNA synthesis can be carried out before the two factors unbind. The bottom panel presents an example where TBP and TFIIE bind briefly and have a residence time shorter than the Pol II dwell time. They do not need to be bound for RNA synthesis to complete and their binding produces one round of initiated synthesis.

pathways will be more difficult to prove, as there is another possibility that factor A binds first, dissociates, then factor B binds; this is still a linear pathway, but is not what residence times may suggest. If factors A and B form a pre-assembled complex before binding, we would expect their residence times to be approximately equal; however they may bind together but dissociate separately at different times. The dynamic measurements will provide a good starting place to build assembly pathways that will give insight into how initiation is regulated genome wide. Combined with RNA synthesis data, as described in Figure 5.3, the dynamics measured from individual PIC components and regulatory complexes can help us fully assemble a model of how the process of transcriptional regulation occurs in cells in real time across gene classes.

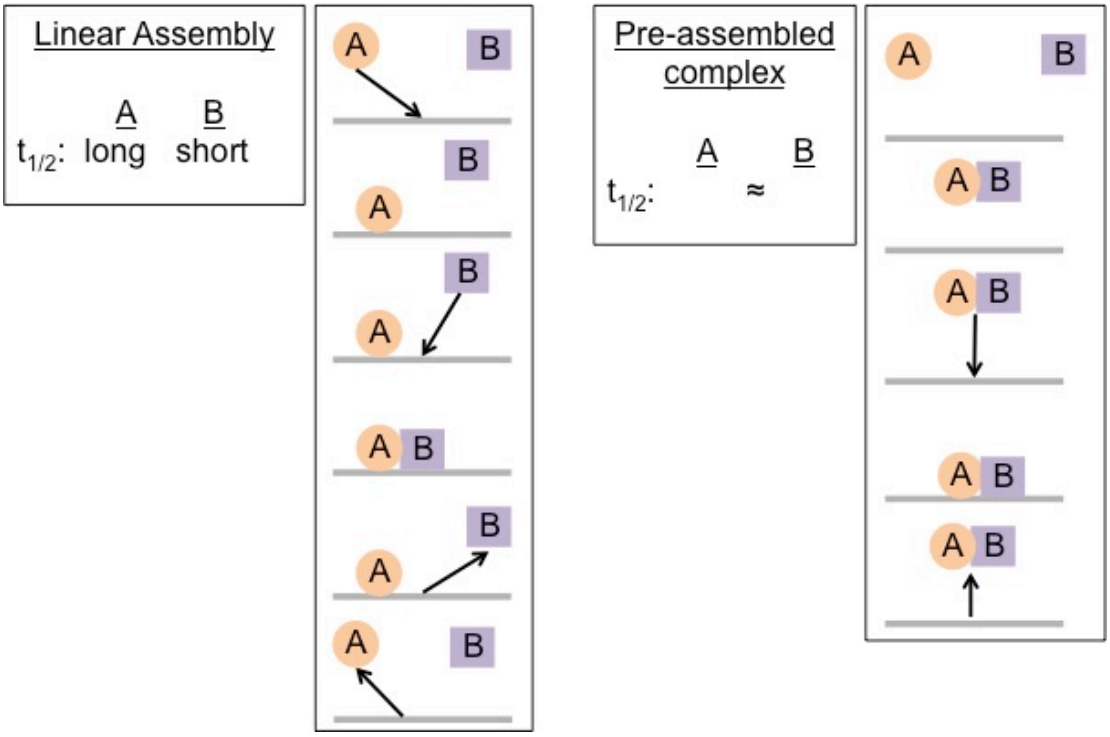


Figure 5.4. Schematic of potential assembly pathways directed by factor residence time. Two assembly orders are shown: a linear assembly on the left and a pre-assembled complex on the right. One possible linear pathway where factor A (orange) binds DNA (gray line) first followed by factor B (purple) could have a long residence time ($t_{½}$) for A and shorter time for B.

A pre-assembled complex could have a similar residence time for both factors, since both would bind together and could potentially dissociate together.

Appendix A

This appendix documents the sequences for the cassettes used in construction of the TFIIB Competition ChIP strain.

DNA sequence of cassette amplified from pFA6-HisMX6-PGAL1-3HA:

GAATTCGAGCTCgtttaaacTGGATGGCGGCGTTAGTATCGAATCGACAGCAGTATAGCG
ACCAGCATTCACATACGATTGACGCATGATATTACTTTCTGCGCACTTAACTTCGCAT
CTGGGCAGATGATGTCGAGGCGAAAAAAAATATAAATCACGCTAACATTTGATTAA
AATAGAACAACTACAATATAAAAAAACTATACAAATGACAAGTTCTTGAAAACAAG
AATCTTTTTATTGTCagtactTCACATCAAAACACCTTTGGTTGAGGGAACGTCATTGGT
GCCATTGCTAGAAATAGCTTCTCTTATGGCAACAGCCAAAGCCTTGAACGCACTCTC
ACTTCTGTGGTGATCGTTGAAACCTCTCAGACAATCAACATGCAAAGTAATTCTGGC
CGCCTCCGCGAAACTTTCCAAAAAGTGTGGAATCATTTCAGTGGATAAATCACCAAT
CATCTCTCTCTTCAATCCAAGGTCGATTACAGCAAATGGTCTATTAGATAAATCGACT
ACGGCACGTGATAGCGCCTCATCCAATGGTGCGAACCCAGTACCGAATCTTTTTACA
CCACGGACAGCACCCATTGCTTCTTTGAACGCTTGCCCTAATGCGATACCGCAATCTT
CGGTAGTATGGTGATCGTCAATGTGCAGGTCACCGATACATTCAACAATAAGAGACC
AACCAGAGTGTTTTGCCAACGCATGGATCATATGATCCAAAAAGCCAACACCTGTGT
GAATATCGATGACCTGTGACTGAGTAGCTTGGGAAGCTACATCGTCATCCTTCTTTGC
AGGAAGAATCGAATCTTTTATTTGAATATAACCACCATTCAGCGAAATAGCGATTTG
AATTTTAGTTTCATTAGTGATACGGGAGATAAACGCCTTGCGCTCCTGAACAGTTTGT
TTTTGCTTTTTTTGGGCTGGTTCTGccatggTTGTTTATGTTCGGATGTGATGTGAGAACT
GTATCCTAGCAAGATTTTAAAAGGAAGTATATGAAAGAAGAACCTCAGTGGCAAAT
CCTAACCTTTTATATTTCTCTACAGGGGCGCGGCGTGGGGACAATTCAACGCGTCTGT
GAGGGGAGCGTTTCCCTGCTCGCAGGTCTGCAGCGAGGAGCCGTAATTTTTGCTTCG
CGCCGTGCGGCCATCAAAATGTATGGATGCAAATGATTATACATGGGGATGTATGGG
CTAAATGTACGGGCGACAGTCACATCATGCCCCTGAGCTGCGCACGTCAAGACTGTC
AAGGAGGGTATTCTGGGCCTCCATGTCGCTGGCCGGGTGACCCGGCGGGGACGAGG
CAAGCTAAACagatctgtaaagagccccattatcttaGCCTAAAAAAACCTTCTCTTTGGAACTTTCA
GTAATACGCTTAACTGCTCATTGCTATATTGAAGTACGGATTAGAAGCCGCCGAGCG
GGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGC
GTTCCTGAAACGCAGATGTGCCTCGCGCCGCACTGCTCCGAACAATAAAGATTCTAC
AATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAACCTGGCCCCACAAA
CCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTT
AGCCTTATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGAT
ATATAAATGCAAAAACTGCATAACCACTTTAACTAATACTTTCAACATTTTCGGTTTG
TATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAATTGTTAATATACCTCT
ATACTTTAACGTCaaggagaaaaaacccggatctcaaaatgtctttaattaaCATCTTTTACCCATACGATGT
TCCTGACTATGCGGGCTATCCGTATGACGTCCCGGACTATGCAGGATCCTATCCATAT
GACGTTCCAGATTACGCTGCTCAGTGC

HA amino acid sequence:

YPYDVPDYAGYPYDVPDYAGSYPYDVPDYAAQC

This cassette was inserted at the transcription start site, but the first two endogenous amino acids of *SUA7* were deleted because they were both start codons.

DNA sequence of cassette amplified from pOM20 (includes *KanMX*):

atgtgcaggtcgacaacccttaatataacttcgtataatgtatgctatacgaagttattaggtctagagatctgtttagcttgccttgtccccgccg
ggtcacccggccagcgacatggaggcccagaataccctccttgacagtcttgacgtgcgcagctcaggggcatgatgtgactgtcgcccgt
acatttagcccatacatccccatgtataatcatttgcatccatacattttgatggccgcacggcgcgaagcaaaaattacggctcctcgctgcag
acctgcgagcagggaaacgctcccctcacagacgcgttgaattgtccccacgccgcgcccctgtagagaaatataaaaggttaggatttgc
cactgaggttcttctttcatatacttccttttaaaatcttgctaggatacagttctcacatcacatccgaacataaacaaccatgggtaaggaaaag
actcacgtttcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcga
caatctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggt
cagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccc
cggcaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcga
ttcctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgatttt
gatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataagcttttgccattctcaccggattcagtcgtcactcatggtg
atttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgc
catcctatggaactgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtt
tcatttgatgctcgatgagtttttctaatcagtactgacaataaaaagattcttgttttcaagaacttgtcatttgtatagttttttttatattgtagttgttct
attttaatcaaatgttagcgtgatttatatttttttttcgcctcgacatcatctgcccagatgcgaagttaagtgcgcagaaagtaatatcatgcgtca
atcgtatgtgaatgctggtcgctatactgctgtcgattcgatactaacgccgccatccagtgtcgaaaacgagctctcgagaaccttaatataa
cttcgtataatgtatgctatacgaagttattaggtgatatccgtacgctgcaggtcgactccggttctgctgctagtggtgaacaaaagttgatttc
tgaagaagatttgaacggtgaacaaaagctaatctccgaggaagacttgaacggtgaacaaaaattaatctcagaagaagacttgaacggat
cctctagaggtgaacaaaagttgatttctgaagaagatttgaacggtgaacaaaagctaatctccgaggaagacttgaacggtgaacaaaaat
taatctcagaagaagacttgaacggatcctctagaggtgaacaaaagttgatttctgaagaagatttgaacggtgaacaaaagctaatctccga
ggaagacttgaacggtgaacaaaaattaatctcagaagaagacttgaacggatccactagcactagtggcctatgcggccgc

This cassette was integrated into YPH501, followed by cre-mediated excision of *KanMX*. An extra ATG was placed in front of the cassette; the first two amino acids of *SUA7* were knocked out again since they are both start codons.

DNA sequence of 9xMyc-tag left after *KanMX* excision:

atgtgcaggtcgacaacccttaatataacttcgtataatgtatgctatacgaagttattaggtgatatccgtacgctgcaggtcgactccggttct
gctgctagtggtgaacaaaagttgatttctgaagaagatttgaacggtgaacaaaagctaatctccgaggaagacttgaacggtgaacaaaa
attaatctcagaagaagacttgaacggatcctctagaggtgaacaaaagttgatttctgaagaagatttgaacggtgaacaaaagctaatctcc
gaggaagacttgaacggtgaacaaaaattaatctcagaagaagacttgaacggatcctctagaggtgaacaaaagttgatttctgaagaagat
ttgaacggtgaacaaaagctaatctccgaggaagacttgaacggtgaacaaaaattaatctcagaagaagacttgaacggatccactagcac
tagtggcctatgcggccgc

9xMyc-tag Amino Acid sequence:

EQKLISEEDLNGEQKLISEEDLNGEQKLISEEDLNGSSRGEQKLISEEDLNGEQKLISEEDL
NGEQKLISEEDLNGSSRGEQKLISEEDLNGEQKLISEEDLNGEQKLISEEDL

Appendix B

This appendix documents the sequence of the *KanMX* cassette used to dirsupt *GAL11* to make the *gal11Δ* strain.

<u>*KanMX* cassette amplified from p400a:</u>

gattgtactgagagtgcaccatagggttaattaaggcgcgccagatctgtttagcttgcctcgtccccgccgggtcacccggccagcgacat
ggaggcccagaataccctccttgacagtcttgacgtgcgcagctcaggggcatgatgtgactgtcgcccgtacatttagcccatacatcccca
tgtataatcatttgcatccatacattttgatggccgcacggcgcgaagcaaaaattacggctcctcgctgcagacctgcgagcagggaaacgc
tcccctcacagacgcgttgaattgtccccacgccgcgcccctgtagagaaatataaaaggttaggatttgccactgaggttcttctttcatatact
tccttttaaaatcttgctaggatacagttctcacatcacatccgaacataaacaaccatgggtaaggaaaagactcacgtttcgaggccgcga
ttaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaa
gcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacgg
aatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggcaaaacagcattccagg
tattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttta
acagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctg
gcctgttgaacaagtctggaaagaaatgcataagcttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttattt
ttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcg
gtgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagttt
ttctaatcagtactgacaataaaaagattcttgttttcaagaacttgtcatttgtatagttttttatattgtagttgttctattttaatcaaatgttagcgtg
atttatatttttttttcgcctcgacatcatctgcccagatgcgaagttaagtgcgcagaaagtaatatcatgcgtcaatcgtatgtgaatgctggtcg
ctatactgctgtcgattcgatactaacgccgccatccagtgtcgaaaacgagctcgaattcatcgatgattatgcggtgtgaaataccgcacag

This cassette contains KanR and was inserted directly into the ORF of *GAL11*.

REFERENCES

Adamkewicz, J. I., C. G. Mueller, K. E. Hansen, W. A. Prud'homme and J. Thorner (2000). "Purification and enzymic properties of Mot1 ATPase, a regulator of basal transcription in the yeast Saccharomyces cerevisiae." Journal of Biological Chemistry **275**: 21158 –21168.

Adams, B. G. (1972). "Induction of galactokinase in Saccharomyces cerevisiae: kinetics of induction and glucose efects." Journal of Bacteriology **111**: 308-315.

Ahn, S. H., M. Kim and S. Buratowski (2004). "Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing." Molecular Cell **13**: 67 – 76.

Allen, B. L. and D. J. Taatjes (2015). "The Mediator complex: a central integrator of transcription." Nature **518**: 376-380.

Amberg, D. C., D. Burke and J. N. Strathern (2005). Methods in Yeast Genetics: A Cold Spring Harbor laboratory course manual. , Cold Spring Harbor Laboratory Press.

Andel, F., A. G. Ladurner, C. Inouye, R. Tjian and E. Nogales (1999). "Three-dimensional structure of the human TFIID-IIA-IIB complex." Science **286**: 2153-2156.

Ang, K., G. Ee, E. Ang, E. Koh, W. L. Siew, Y. M. Chan, S. Nur, Y. S. Tan and N. Lehming (2012). "Mediator Acts Upstream of the Transcriptional Activator Gal4." PLoS Biology **10**(3): e1001290.

Aparicio, O., J. V. Geisberg, E. Sekinger, A. Yang, Z. Moqtaderi and K. Struhl (2005). Chromatin Immunoprecipitation for Determining the Association of Proteins with Specific Genomic Sequences In Vivo. Current Protocols in Molecular Biology, John Wiley & Sons, Inc.

Archambault, J., G. Pan, G. K. Dahmus, M. Cartier, N. Marshall, S. Zhang, M. E. Dahmus and J. Greenblatt (1998). "FCP1, the RAP74- interacting subunit of a human protein phosphatase that dephos- phorylates the carboxyl-terminal domain of RNA polymerase IIO." Journal of Biological Chemistry **273**: 27593–27601.

Auble, D. T. (2009). "The dynamic personality of TATA-binding protein." Trends in Biochemical Sciences **34**(2): 49-52.

Auble, D. T. and S. Hahn (1993). "An ATP-dependent inhibitor of TBP binding to DNA." Genes & Development **7**: 844 – 856.

Auble, D. T., K. E. Hansen, C. G. Mueller, W. S. Lane, J. Thorner and S. Hahn (1994). "Mot1, a global repressor of RNA polymerase II transcrip- tion, inhibits TBP binding to DNA by an ATP-dependent mechanism." Genes & Development **8**: 1920 –1934.

Badi, L. and A. Barberis (2001). "Proteins that genetically interact with the Saccharomyces cerevisiae transcription factor Gal11p emphasize its role in the initiation-elongation transition." Molecular Genetics and Genomics **265**: 1076-1086.

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

Baidoobonso, S. M., B. W. Guidi and L. C. Myers (2007). "Med19(Rox3) regulates intermodule interactions in the Saccharomyces cerevisiae mediator complex." Journal of Biological Chemistry **282:**: 5551–5559.

Baker, S. P. and P. A. Grant (2007). "The SAGA continues: expanding the cellular role of a transcriptional co-activator complex." Oncogene **26**: 5329– 5340.

Balasubramanian, R., M. G. Pray-Grant, W. Selleck, P. A. Grant and S. Tan (2002). "Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation." Journal of Biological Chemistry **277**: 7989–7995.

Baptista, T., S. Grünberg, N. Minoungou, M. J. E.Koster, H. T. M. Timmers, S. Hahn, D. Devys and L. Tora (2017). "SAGA Is a General Cofactor for RNA Polymerase II Transcription." Molecular Cell **68**(1): 130-143.

Barone-Nugent, E. D., A. Barty and K. A. Nugent (2002). "Quantitative phase-amplitude microscopy I: optical microscopy." Journal of Microscopy **206**(3): 194-203.

Basehoar, A. D., S. J. Zanton and B. F. Pugh (2004). "Identification and Distinct Regulation of Yeast TATA Box-Containing Genes." Cell **116**(5): 699-709.

Bassel, J. and R. Mortimer (1971). "Genetic Order of the Galactose Structural Genes in Saccharomyces cerevisiae." Journal of Bacteriology **108**(1): 179-183.

Becker, M. (2002). "Dynamic behavior of transcription factors on a natural promoter in living cells." EMBO reports **3**(12): 1188-1194.

Bellí, G., E. Garí, L. Piedrafita, M. Aldea and E. Herrero (1998). "An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast." Nucleic Acids Research **26**(4): 942-947.

Belshaw, P. J., S. N. Ho, G. R. Crabtree and S. L. Schreiber (1996). "Controlling protein association and subcellular localization with a synthetic ligand that induces heterodimerization of proteins." Proceedings of the National Academy of Sciences **93**: 4604-4607.

Bentley, D. (2002). "The mRNA assembly line: transcription and processing machines in the same factory." Current Opinion in Cell Biology **14**(3): 336-342.

Bhaumik, S. R. and M. R. Green (2001). "SAGA is an essential in vivo target of the yeast acidic activator Gal4p." Genes & Development **15**: 1935–1945.

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

Billinton, N. and A. W. Knight (2001). "Seeing the Wood Through the Trees: A Review of Techniques for Distinguishing GFP from Endogenous Autofluorescence." Analytical Biochemistry **291**(2): 175-197.

Black, J. C., J. E. Choi, S. R. Lombardo and M. Carey (2006). "A mechanism for coordinating chromatin modification and preinitiation complex assembly." Molecular Cell **23**: 809–818.

Black, J. C., A. Mosley, T. Kitada, M. Washburn and M. Carey (2008). "The SIRT2 deacetylase regulates autoacetylation of p300." Molecular Cell **32**: 449–455.

Borggrefe, T., R. Davis, A. Bareket-Samish and R. D. Kornberg (2001). "Quantitation of the RNA Polymerase II Transcription Machinery in Yeast." Journal of Biological Chemistry **276**(50): 47150-47153.

Borggrefe, T. and X. Yue (2011). "Interactions between subunits of the Mediator complex with gene-specific transcription factors." Seminars in Cell and Developmental Biology **22**: 759–768.

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

Bram, R. J. and R. D. Kornberg (1985). "Specific protein binding to far upstream activating sequences in polymerase II promoters." Proceedings of the National Academy of Sciences **82**(1): 43-47.

Bram, R. J., N. F. Lue and R. D. Kornberg (1986). "A GAL family of upstream activating sequences in yeast: roles in both induction and repression of transcription." The EMBO Journal **5**(3): 603-608.

Brown, C. E., L. Howe, K. Sousa, S. C. Alley, M. J. Carrozza, S. Tan and J. L. Workman (2001). "Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit." Science **292**: 2333–2337.

Brutlag, D., C. Schlehuber and J. Bonner (1969). "Properties of formaldehyde-treated nucleohistone." Biochem. **8**: 3214-3218.

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

Buck, M. J. and J. D. Lieb (2004). "ChIP-chip: Considerations for the Design, Analysis, and Application of Genome-Wide Chromatin Immunoprecipitation Experiments." Genomics **83**: 349-360.

Buckley, M. S., H. Kwak, W. R. Zipfel and J. T. Lis (2014). "Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation." Genes & Development **28**: 14-19.

Buratowski, S., S. Hahn, L. Guarente and P. A. Sharp (1989). "Five Intermediate Complexes in Transcription Initiation by RNA Polymerase II " Cell **56**: 549-561.

Buratowski, S. and P. A. Sharp (1990). "Transcription initiation complexes and upstream activation with RNA polymerase II lacking the C- terminal domain of the largest subunit." Molecular & Cellular Biology **10**: 5562–5564.

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

Bushnell, D. A. and R. D. Kornberg (2003). "Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription." Proceedings of the National Academy of Sciences **100**(12): 6969-6973.

Butryn, A., J. M. Schuller, G. Stoehr, P. Runge-Wollmann, F. Förster, D. T. Auble and K. P. Hopfner (2015). "Structural basis for recognition and remodeling of the TBP:DNA:NC2 complex by Mot1." Elife **4**: e07432.

Cabart, P., A. Ujvari, M. Pal and D. S. Luse (2011). "Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes." Proceedings of the National Academy of Sciences **108**(38): 15786-15791.

Cai, L., N. Friedman and X. S. Xie (2006). "Stochastic protein expression in individual cells at the single molecule level." Nature **440**: 358-362.

Cang, Y. and G. Prelich (2002). "Direct stimulation of transcription by negative cofactor 2 (NC2) through TATA-binding protein (TBP)." Proceedings of the National Academy of Sciences **99**: 12727–12732.

Chambers, R. S., B. Q. Wang, Z. F. Burton and M. E. Dahmus (1995). "The activity of COOH-terminal domain phosphatase is regulated by a docking site on RNA polymerase II and by the general transcription factors IIF and IIB." Journal of Biological Chemistry **270**: 14962–14969.

Chen, D., C. S. Hinkley, R. W. Henry and S. Huang (2002). "TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes." Molecular Biology of the Cell **13**: 276-284.

Chen, H., E. R. Farkas and W. W. Webb (2008). "Chapter 1: In vivo applications of fluorescence correlation spectroscopy." Methods in Cell Biology **89**: 3–35.

Chen, J., Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian and Z. Liu (2014). "Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells." Cell **156**(6): 1274-1285.

Chen, J., Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian and Z. Liu (2014). "Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells." Cell **156**: 1274-1285.

Chen, J., X.-F. Zheng, E. J. Brown and S. L. Schreiber (1995). "Identification of an 11-kDa FKBP12-rapamycin-binding domain within the 289-kDa FKBP12- rapamycin-associated protein and characterization of a critical serine residue." Proceedings of the National Academy of Sciences **92**: 4947–4951.

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

Christianson, T. W., R. S. Sikorski, M. Dante, J. H. Shero and P. Hieter (1992). "Multifunctional yeast high-copy-number shuttle vectors." Gene **110**(1): 119-122.

Churchman, L. S. and J. S. Weissman (2011). "Nascent transcript sequencing visualizes transcription at nucleotide resolution." Nature **469**: 368-373.

Cody, S. H., S. D. Xiang, M. J. Layton, E. Handman, M. H. C. Lam, J. E. Layton, E. C. Nice and J. K. Heath (2005). "A simple method allowing DIC imaging in conjunction with confocal microscopy." Journal of Microscopy **217**(3): 265-274.

Collas, P. (2010). "The current state of chromatin immunoprecipitation." Mol. Biotechnol. **45**: 87-100.

Collins, G. A., J. R. Lipford, R. J. Deshaies and W. P. Tansey (2009). "Gal4 turnover and transcription activation." Nature **461**: E7.

Conaway, R. C. and J. W. Conaway (1993). "General Initiation Factors for RNA Polymerase II." Annual Review of Biochemistry **62**(1): 161-190.

Consortium, T. E. P. (2012). "An integrated encyclopedia of DNA elements in the human genome. ." Nature **489**(7414): 57-74.

Core, L. J., J. J. Waterfall and J. T. Lis (2008). "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters." Science **322**(5909): 1845-1848.

Coulon, A., C. C. Chow, R. H. Singer and D. R. Larson (2013). "Eukaryotic transcriptional dynamics: from single molecules to cell populations." Nature Reviews Genetics **14**(8): 572-584.

Cramer, P. (2014). "A Tale of Chromatin and Transcription in 100 Structures." Cell **159**(5): 985-994.

Crick, F. H. (1958). "On protein synthesis." Symposia of the Society for Experimental Biology **12**: 138– 163.

Dammann, R., R. Lucchini, T. Koller and J. M. Sogo (1993). "Chromatin structures and transcription of rDNA in yeastSaccharomyces cerevisiae." Nucleic Acids Research **21**(10): 2331-2338.

Darzacq, X., Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair and R. H. Singer (2007). "In vivo dynamics of RNA polymerase II transcription." Nature Structural & Molecular Biology **14**: 796–806.

Dasgupta, A., R. P. Darst, K. J. Martin, C. A. Afshari and D. T. Auble (2002). " Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms." Proceedings of the National Academy of Sciences **99**: 2666-2671.

de Graaf, P., F. Mousson, B. Geverts, E. Scheer, L. Tora, A. B. Houtsmuller and H. T. M. Timmers (2010). "Chromatin interaction of TATA-binding protein is dynamically regulated in human cells." Journal of Cell Science **123**(15): 2663-2671.

de Wit, E. and W. de Laat (2012). "A decade of 3C technologies: insights into nuclear organization." Genes & Dev. **26**: 11-24.

Deal, R. B., J. G. Henikoff and S. Henikoff (2010). "Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones." Science **328**: 1161–1164

Dedon, P. C., J. A. Soults, C. D. Allis and M. A. Gorovsky (1991). "Formaldehyde Cross-Linking and Immunoprecipitation Demonstrate Developmental Changes in H1 Association with Transcriptionally Active Genes." Mol. Cell. Biol. **11**: 1729-1733.

Delgadillo, R. F., J. E. Whittington, L. K. Parkhurst and L. J. Parkhurst (2009). "The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism." Biochemistry **48**: 1801–1809.

Dinant, C., M. S. Luijsterburg, T. Hofer, G. von Bornstaedt, W. Vermeulen, A. B. Houtsmuller and R. van Driel (2009). "Assembly of multiprotein complexes that control genome function." Journal of Cell Biology **185**: 21–26.

Dion, M. F., T. Kaplan, M. Kim, S. Buratowski, N. Friedman and O. J. Rando (2007). "Dynamics of replication-independent histone turnover in budding yeast." Science **315**: 1405–1408.

Douglas, H. C. and D. C. Hawthorne (1964). "Enzymatic expression and genetic linkage of genes controlling galactose utilization in Saccharomyces " Genetics **49**: 837-844.

Dowen, J. M., Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham, L. N. Zhang, A. S. Weintraub, J. Schuijers, T. I. Lee, K. Zhao and R. A. Young (2014). "Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes." Cell **159**(2): 374-387.

Dundr, M., U. Hoffmann-Rohrer, Q. Hu, I. Grummt, L. I. Rothblum, R. D. Phair and T. Misteli (2002). "A kinetic framework for a mammalian RNA polymerase in vivo." Science **298**: 1623–1626.

Dutta, A., Babbarwal, V., Fu, J., Brunke-Reese, D., Libert, D. M., Willis, J., and J. C. Reese (2015). "Ccr4-Not and TFIIS function cooperatively to rescue arrested RNA Polymerase II." Molecular and Cellular Biology **35**(11): 1915-1925.

Eckels, K. H. and R. Putnak (2003). "Formalin-inactivated Whole Virus and Recombinant Subunit Flavivirus Vaccines." Adv. in Virus Res. **61**: 395-418.

Egriboz, O., F. Jiang and J. E. Hopper (2011). "Rapid GAL gene switch of Saccharomyces cerevisiae depends on nuclear Gal3, not nucleocytoplasmic trafficking of Gal3 and Gal80." Genetics **189**: 825-836.

Eichner, J., H. T. Chen, L. Warfield and S. Hahn (2010). "Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex." The EMBO Journal **29**: 706-716.

Elbright, R. H. (2000). "RNA polymerase: structural similarites between bactieral RNA polymerase and eukaryotic RNA polymerase II." Journal of Molecular Biology **304**: 687-698.

Elf, J., G. W. Li and X. S. Xie (2007). "Probing Transcription Factor Dynamics at the Single-Molecule Level in a Living Cell." Science **316**(5828): 1191-1194.

Esnault, C., Y. Ghavi-Helm, S. Brun, J. Soutourina, N. Van Berkum, C. Boschiero, F. Holstege and M. Werner (2008). "Mediator-dependent recruitment of TFIIH modules in preinitiation complex." Molecular Cell **31**: 337-346.

Fan, X., N. Lamarre-Vincent, Q. Wang and K. Struhl (2008). "Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments." Nucleic Acids Res. **36**: e125.

Fan, X. and K. Struhl (2009). "Where does mediator bind in vivo?" PLoS ONE **4**: e5029.

Feldman, M. Y. (1973). "Reactions of nucleic acids and nucleoproteins with formaldehyde." Prog. Nucl. Acids Res. Mol. Biol. **13**: 1-49.

Finnigan, G. C., A. Duvalyan, E. N. Liao, A. Sargsyan and J. Thorner (2016). "Detection of protein–protein interactions at the septin collar in Saccharomyces cerevisiae using a tripartite split-GFP system." Molecular Biology of the Cell **27** (17): 2708-2725.

Fishburn, J. and S. Hahn (2012). "Architecture of the Yeast RNA Polymerase II Open Complex and Regulation of Activity by TFIIF." Molecular and Cellular Biology **32**(1): 12-25.

Flanagan, P. M., R. J. Kelleher, M. H. Sayre, H. Tschochner and R. D. Kornberg (1991). "A mediator required for activation of RNA polymerase II transcription in vitro." Nature **350**: 436-438.

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

Flores, O., E. Maldonado, Z. Burton, J. Greenblatt and D. Reinberg (1988). "Factors involved in specific transcription by mammalian RNA polymerase II. RNA polymerase II-associating protein 30 is an essential component of transcription factor IIF." Journal of Biological Chemistry **263**: 10812– 10816.

Flores, O., E. Maldonado and D. Reinberg (1989). "Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II." Journal of Biological Chemistry **264**: 8913–8921.

Fraenkel-Conrat, H. (1954). "Reaction of nucleic acid with formaldehyde." Biochim. Biophys. Acta **15**: 307-309.

Fraenkel-Conrat, H., M. Cooper and H. S. Olcott (1945). "The reaction of formaldehyde with proteins." J. Am. Chem. Soc. **67**: 950-954.

Fraenkel-Conrat, H. and H. S. Olcott (1948). "Reaction of Formaldehyde with Proteins. VI. Cross-linking of Amino Groups with Phenol, Imidazole, or Indole Groups." J. Biol. Chem. **174**: 827-843.

French, D. and J. T. Edsall (1945). "The Reactions of Formaldehyde with Amino Acids and Proteins." Adv. Protein Chem. **2**: 277-335.

Fuda, N. J., M. B. Ardehali and J. T. Lis (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." Nature **461**: 186-192.

Gauss, R., M. Trautwein, T. Sommer and A. Spang2 (2005). "New modules for the repeated internal and N-terminal epitope tagging of genes in Saccharomyces cerevisiae." Yeast **22**: 1-12.

Gavrilov, A. and S. V. Razin (2009). "Formaldehyde fixation of cells does not greatly reduce the ability to amplify cellular DNA." Anal. Biochem. **390**: 94-96.

Gavrilov, A., S. V. Razin and G. Cavalli (2015). "In vivo formaldehyde cross-linking: it is time for black box analysis." Briefings in Functional Genomics **14**(2): 163-165.

Ge, H., E. Martinez, C. M. Chiang and R. G. Roeder (1996). "Activator-dependent transcription by mammalian RNA polymerase II: in vitro reconstitution with general transcription factors and cofactors." Methods in Enzymology **274**: 57–71.

Geisberg, J. V., F. C. Holstege, R. A. Young and K. Struhl (2001). "Yeast NC2 associates with the RNA polymerase II preinitiation complex and selectively affects transcription in vivo." Molecular & Cellular Biology **21**: 2736–2742.

Geisberg, J. V., Z. Moqtaderi, L. Kuras and K. Struhl (2002). "Mot1 associates with transcriptionally active promoters and inhibits association of NC2 in Saccharomyces cerevisiae." Molecular & Cellular Biology **22**(23): 8122-8134.

Ghaemmaghami, S., W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea and J. S. Weissman (2003). "Global analysis of protein expression in yeast." Nature **425**(6959): 737-741.

Giglia-Mari, G., A. F. Theil, P. O. Mari, S. Mourgues, J. Nonnekens, L. O. Andrieux, J. de Wit, C. Miquel, N. Wijgers, A. Maas, M. Fousteri, J. H. Hoeijmakers and W. Vermeulen (2009). "Differentiation driven changes in the dynamic organization of Basal transcription initiation." PLoS Biology **7**: e1000220.

Gilmour, D. S. and J. T. Lis (1985). "In vivo interactions of RNA polymerase II with genes of Drosophila melanogaster." Mol. Cell. Biol. **5**: 2009-2018.

Giniger, E. and M. Ptashne (1988). "Cooperative DNA binding of the yeast transcriptional activator Gal4." Proceedings of the National Academy of Sciences **85**: 382-386.

Giniger, E., S. M. Varnum and M. Ptashne (1985). "Specific DNA Binding of GAL4, A Positive Regulatory Protein of Yeast " Cell **40**: 767-774.

Gorski, S. A., M. Dundr and T. Misteli (2006). "The road much traveled: trafficking in the cell nucleus." Current Opinion in Cell Biology **18**(3): 284-290.

Gorski, S. A., S. K. Snyder, S. John, I. Grummt and T. Misteli (2008). "Modulation of RNA polymerase assembly dynamics in transcriptional regulation." Molecular Cell **30**: 486-497.

Gossen, M., S. Freundlieb, G. Bender, G. Muller, W. Hillen and H. Bujard (1995). "Transcriptional activation by tetracyclines in mammalian cells." Science **268**(5218): 1766-1769.

Goutte, C. and A. D. Johnson (1988). "a1 Protein alters the dna binding specificity of α2 repressor." Cell **52**(6): 875-882.

Govind, C. K., F. Zhang, H. Qiu, K. Hofmeyer and A. G. Hinnebusch (2007). "Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions." Molecular Cell **25**: 31–42.

Grant, P. A., S. D., M. G. Pray-Grant, J. R. Yates and J. L. Workman (1998). "The ATM-related cofactor Tra1 is a component of the purified SAGA complex." Molecular Cell **2**: 863–867.

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

Grimaldi, Y., P. Ferrari and M. Strubin (2014). "Independent RNA polymerase II preinitiation complex dynamics and nucleosome turnover at promoter sites in vivo." Genome Research **24**: 117–124.

Grohmann, D., J. Nagy, A. Chakraborty, D. Klose, D. Fielden, R. H. Ebright, J. Michaelis and F. Werner (2011). "The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation." Molecular Cell **43**: 263-274.

Gu, H., Y.-R. Zou and K. Rajewsky (1993). "Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting." Cell **73**(6): 1155-1164.

Guarente, L. (1987). "Regulatory Proteins in Yeast." Annual Review of Genetics **21**: 425-452.

Guillemette, B., Bataille, A. R., Gévry, N., Adam, M., Blanchette, M., Robert. F., and L. Gaudreau (2005). "Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning." PLoS Biology **3**(12): e384.

Güldener, U., S. Heck, T. Fiedler, J. Beinhauer and J. H. Hegemann (1996). "A new efficient gene disruption cassette for repeated use in budding yeast." Nucleic Acids Research **24**: 2519–2524.

Hager, G. L., J. G. McNally and T. Misteli (2009). "Transcription Dynamics." Molecular Cell **35**(6): 741-753.

Hahn, S. (2004). "Structure and mechanism of the RNA polymerase II transcription machinery." Nature Structural & Molecular Biology **11**(5): 394-403.

Halford, S. E. (2009). "An end to 40 years of mistakes in DNA-protein association kinetics?" Biochemical Society Transactions **37**: 343-348.

Hall, D. B. and K. Struhl (2002). "The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein crosslinking in vivo." J. Biol. Chem. **277**: 46043-46050.

Hammar, P., P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg and J. Elf (2012). "The lac Repressor Displays Facilitated Diffusion in Living Cells." Science **336**(6088): 1595-1598.

Harbison, C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J.-B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel and R. A. Young (2004). "Transcriptional Regulatory Code of a Eukaryotic Genome." Nature **431**: 99-104.

Hargreaves, D. C. and G. R. Crabtree (2011). "ATP-dependent chromatin re- modeling: genetics, genomics and mechanisms." Cell Research **21**: 396–420.

Haruki, H., J. Nishikawa and U. K. Laemmli (2008). "The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes." Molecular Cell **31**: 925-932.

Haselkorn, R. and P. Doty (1961). "The reaction of formaldehyde with polynucleotides." J. Biol. Chem. **236**: 2738-2745.

Hawkes, N. A. and S. G. Roberts (1999). "The role of human TFIIB in transcription start site selection in vitro and in vivo." Journal of Biological Chemistry **274**: 14337-14343.

He, Y., J. Fang, D. J. Taatjes and E. Nogales (2013). "Structural visualization of key steps in human transcription initiation." Nature **495**(7442): 481-486.

Herskowitz, I. (1989). "A regulatory hierarchy for cell specialization in yeast." Nature **342**(6251): 749-757.

Hoffman, E. A., B. L. Frey, L. M. Smith and D. T. Auble (2015). "Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes." Journal of Biological Chemistry **290**(44): 26404-26411.

Holstege, F. C., U. Fiedler and H. T. Timmers (1997). "Three transitions in the RNA polymerase II transcription complex during initiation." The EMBO Journal **16**(24): 7468-7480.

Holstege, F. C., E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander and R. A. Young (1998). "Dissecting the regulatory circuitry of a eukaryotic genome." Cell **95**(5): 717-728.

Holstege, F. C. P., D. Tantin, M. Carey, P. C. van der Vliet and H. T. Timmers (1995). "The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA." The EMBO Journal **14**: 810-819

Hong, M., M. X. Fitzgerald, S. Harper, C. Luo, D. W. Speicher and R. Marmorstein (2008). "Structural basis for dimerization in DNA recognition by Gal4." Structure **16**: 1019-1026.

Horikoshi, M., C. Bertucciolo, R. Takada, J. Wang, T. Yamamoto and R. G. Roeder (1992). " Transcription factor TFIID induces DNA bending upon binding to the TATA element." Proceedings of the National Academy of Sciences **89**: 1060-1064.

Horiuchi, J., N. Silverman, G. A. Marcus and L. Guarente (1995). "ADA3, a putative transcriptional adaptor, consists of two separable domains and interacts with ADA2 and GCN5 in a trimeric complex." Molecular & Cellular Biology **15**: 1203–1209.

Horn, A. E., J. F. Kugel and J. A. Goodrich (2016). "Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity." Nucleic Acids Research: gkw321.

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

Huisinga, K. L. and B. F. Pugh (2007). "A TATA binding protein regulatory network that governs transcription complex assembly." Genome Biology **8**: R46.

Jackson, V. (1978). "Studies on Histone Organization in the Nucleosome Using Formaldehyde as a Reversible Cross-Linking Agent." Cell **15**: 945-954.

Jacobson, R. H., A. G. Ladurner, D. S. King and R. Tjian (2000). "Structure and function of a human TAFII250 double bromodomain module." Science **288**: 1422–1425.

Jamai, A., R. M. Imoberdorf and M. Strubin (2007). "Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication." Molecular Cell **25**: 345–355

Jencks, W. P. (1997). "From chemistry to biochemistry to catalysis to movement." Ann. Rev. Biochem. **66**: 1-18.

Jensen, R., J. George F. Sprague and I. Herskowwiz (1983). "Regulation of yeast mating-type interconversion: feedback control of HO gene expression by the mating-type locus." Proceedings of the National Academy of Sciences **80**: 3035-3039.

Jeong, C.-J., S.-H. Yang, Y. Xie, L. Zhang and S. A. J. T. Kodadek (2001). "Evidence That Gal11 Protein Is a Target of the Gal4 Activation Domain in the Mediator." Biochemistry **40**(31): 9421-9427.

Jin, C. and G. Felsenfield (2007). "Nucleosome stability mediated by histone variants H3.3 and H2A.Z." Genes & Development **21**(12): 1519-1529.

Johnson, K. M., J. Wang, A. Smallwood, C. Arayata and M. Carey (2002). "TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA." Genes & Development **16**(14): 1852-1863.

Johnson, P. R., R. Swanson, L. Rakhilina and M. Hochstrasser (1998). "Degradation Signal Masking by Heterodimerization of MATα2 and MATa1 Blocks Their Mutual Destruction by the Ubiquitin-Proteasome Pathway." Cell **94**(2): 217-227.

Johnston, M. (1987). "A Model Fungal Gene Regulatory Mechanism: the GAL Genes of Saccharomyces Cerevisiae." Microbiological Reviews **51**(4): 458-476.

Johnston, S. A. and J. E. Hopper (1982). "Isolation of the yeast regulatory gene GAL4 and analysis of its dosage effects on the galactose/melibiose regulon." Proceedings of the National Academy of Sciences **79**(22): 6971-6975.

Johnston, S. A., J. J. M. Salmeron and S. S. Dincher (1987). "Interaction of positive and negative regulatory proteins in the galactose-regulonofyeast." Cell **50**: 143-146.

Jorgensen, P., N. P. Edgington, B. L. Schneider, I. Rupes, M. Tyers and B. Futcher (2007). "The Size of the Nucleus Increases as Yeast Cells Grow." Molecular Biology of the Cell **18**(9): 3523-3532.

Jun, S. H., M. J. Reichlen, M. Tajiri and K. S. Murakami (2011). "Archaeal RNA polymerase and transcription regulation." Critical Reviews in Biochemistry and Molecular Biology **46**: 27–40.

Kampmann, M. (2005). "Facilitated diffusion in chromatin lattices: mechanistic diversity and regulatory potential." Molecular Microbiology **57**(4): 889-899.

Kaplan, T., X.-Y. Li, P. J. Sabo, S. Thomas, J. A. Stamatoyannopoulos, M. D. Biggin and M. B. Eisen (2011). "Quantitative Models of the Mechanisms That Control Genome-Wide Patterns of Transcription Factor Binding during Early Drosophila Development." PLoS Genetics **7**(2): e1001290.

Karpova, T. S., T. Y. Chen, B. L. Sprague and J. G. McNally (2004). "Dynamic interactions of a transcription factor with DNA are accelerated by a chromatin remodeller." EMBO reports **5**(11): 1064-1070.

Karpova, T. S., M. J. Kim, C. Spriet, K. Nalley, T. J. Stasevich, Z. Kherrouche, L. Heliot and J. G. McNally (2008). "Concurrent Fast and Slow Cycling of a Transcriptional Activator at an Endogenous Promoter." Science **319**(5862): 466-469.

Kasinathan, S., G. A. Orsi, G. E. Zentner, K. Ahmad and S. Henikoff (2014). "High-resolution mapping of transcription factor binding sites on native chromatin." Nat. Methods **11**: 203-209.

Katan-Khaykovich, Y. and K. Struhl (2002). "Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors." Genes & Dev. **16**: 743-752.

Kawashima, Y., Y. Kodera, A. Singh, M. Matsumoto and H. Matsumoto (2014). "Efficient extraction of proteins from formalin-fixed paraffin-embedded tissues requires higher concentration of tris(hydroxymethyl)aminomethane." Clinical Proteomics **11**(1): 4.

Keegan, L., G. Gill and M. Ptashne (1986). "Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein." Science **231**(4739): 699-704.

Keeney, J. B., K. B. Chapman, V. Lauermann, D. F. Voytas, S. U. Åstrom, U. v. Pawel-Rammingen, A. Bystrom and J. D. Boeke (1995). "Multiple Molecular Determinants for Retrotransposition in a Primer tRNA." Molecular & Cellular Biology **15**(1): 217-226.

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

Kennedy-Darling, J. and L. M. Smith (2014). "Measuring the Formaldehyde Protein-DNA Cross-Link Reversal Rate." Analyt. Chem. **86**: 5678-5681.

Kepler, T. B. and T. C. Elston (2001). "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations." Biophysical Journal **81**: 3116-3136.

Kim, B., Nesvizhskii, A. I., Rani, P. G., Hahn, S., Aebersold, R., and J. A. Ranish (2007). "The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes." Proceedings of the National Academy of Sciences **104**(41): 16068-16073.

Kim, J., E. Bortz, H. Zhong, T. Leeuw, E. Leberer, A. K. Vershon and J. P. Hirsch (2000). "Localization and Signaling of GβSubunit Ste4p Are Controlled by a-Factor Receptor and thea-Specific Protein Asg7p." Molecular & Cellular Biology **20**(23): 8826-8835.

Kim, J. L., D. B. Nikolov and S. K. Burley (1993). "Co-crystal structure of TBP recognizing the minor groove of a TATA element." Nature **365**(6446): 520-527.

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

Kim, T. K., R. H. Ebright and D. Reinberg (2000). "Mechanism of ATP-dependent promoter melting by transcription factor IIH." Science **288**: 1418- 1422.

Kim, W.-Y. and M. E. Dahmus (1989). "The major late promoter of adenovirus-2 is accurately transcribed by RNA polymerase IIO, IIA, and IIB." Journal of Biological Chemistry **264**: 3169–3176.

Kim, Y., J. H. Geiger, S. Hahn and P. B. Sigler (1993). "Crystal structure of a yeast TBP/TATA-box complex." Nature **365**: 512-520.

Kim, Y. J., S. Björklund, Y. Li, M. H. Sayre and R. D. Kornberg (1994). "A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II." Cell **77**(4): 599-608.

Klockenbusch, C. and J. Kast (2010). "Optimization of formaldehyde cross-linking for protein interaction analysis on non-tagged integrin beta1." J. Biomed. Biotechnol. **2010**: 927585.

Klockenbusch, C., J. E. O'Hara and J. Kast (2012). "Advancing formaldehyde cross-linking toward quantitative proteomic applications." Anal. Bioanal. Chem. **404**: 1057-1067.

Knutson, B. and S. Hahn (2011). "Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and Nua4 complexes." Molecular & Cellular Biology **31**: 818-831.

Koh, S. S., A. Z. Ansari, M. Ptashne and R. A. Young (1998). "An Activator Target in the RNA Polymerase II Holoenzyme." Molecular Cell **1**(6): 895-904.

Komarnitsky, P., E.-J. Cho and S. Buratowski (2000). "Differrent phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." Genes & Development **14**: 2452–2460.

Koster, M. J. E., B. Snel and H. T. M. Timmers (2015). "Genesis of Chromatin and Transcription Dynamics in the Origin of Species." Cell **161**(724-736).

Koutelou, E., C. L. Hirsch and S. Y. R. Dent (2010). "Multiple faces of the SAGA complex." Current Opinion in Cell Biology **22**: 374-382.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey (2004). "Ssu72 is an RNA polymerase II CTD phosphatase." Molecular Cell **14**: 387–394.

Kumar, A., S. Agarwal, J. A. Heyman, S. Matson, M. Heidtman, S. Piccirillo, L. Umansky, A. Drawid, R. Jansen, Y. Liu, K.-H. Cheung, P. Miller, M. Gerstein, G. S. Roeder and M. Snyder (2002). "Subcellular localization of the yeast proteome." Genes & Development **16**: 707–719.

Kuo, M.-H. and C. D. Allis (1999). "In Vivo Cross-Linking and Immunoprecipitation for Studying Dynamic Protein:DNA Associations in a Chromatin Environment." Methods **19**(3): 425-433.

Kuras, L., T. Borggrefe and R. D. Kornberg (2003). "Association of the Mediator complex with enhancers of active genes." Proceedings of the National Academy of Sciences **100**: 13887–13891.

Larschan, E. and F. Winston (2001). "The S. cerevisiae SAGA complex functions in vivo as a coactivator for transcriptional activation by Gal4." Genes & Development **15**: 1946–1956.

Larson, D. R., R. H. Singer and D. Zenklusen (2009). "A single molecule view of gene expression." Trends in Cell Biology **19**(11): 630-637.

Larson, D. R., D. Zenklusen, B. Wu, J. A. Chao and R. H. Singer (2011). "Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene." Science **332**(475): 475-478.

Lee, S., W. A. Lim and K. S. Thorn (2013). "Improved Blue, Green, and Red Fluorescent Protein Tagging Vectors for S. cerevisiae." PLoS ONE **8** (7): e67902.

Lee, T. I., H. C. Causton, F. C. Holstege, W. C. Shen, N. Hannett, E. G. Jennings, F. Winston, M. R. Green and R. A. Young (2000). "Redundant roles for the TFIID and SAGA complexes in global transcription." Nature **405**(701–704).

Lemieux, K. and L. Gaudreau (2004). "Targeting of Swi/Snf to the yeast GAL1 UAS G requires the Mediator, TAF IIs, and RNA polymerase II." The EMBO Journal **23**: 4040-4050.

Lenstra, T. L., A. Coulon, C. C. Chow and D. R. Larson (2015). "Single-molecule imaging reveals a switch between spurious and functional ncRNA transcription." Molecular Cell **60**(4): 597-610.

Leuther, K. K., D. A. Bushnell and R. D. Kornberg (1996). "Two-Dimensional Crystallography of TFIIB– and IIE–RNA Polymerase II Complexes: Implications for Start Site Selection and Initiation Complex Formation." Cell **85**(5): 773-779.

Levine, M., C. Cattoglio and R. Tjian (2014). "Looping back to leap forward: transcription enters a new era." Cell **157**: 13–25.

Li, G.-W. and X. S. Xie (2011). "Central dogma at the single-molecule level in living cells." Nature **475**(7356): 308-315.

Li, T., M. R. Stark, A. D. Johnson and C. Wolberger (1995). "Crystal Structure of the MATa1/MATα2 Homeodomain Heterodimer Bound to DNA." Science **270**(5234): 262-269.

Lickwar, C. R., F. Mueller, S. E. Hanlon, J. G. McNally and J. D. Lieb (2012). "Genome-wide protein-DNA binding dynamics suggest a clutch for transcription factor function." Nature **484**: 251–255.

Lickwar, C. R., F. Mueller and J. D. Lieb (2013). "Genome-wide measurement of protein-DNA binding dynamics using competition ChIP." Nature Protocols **8**(7): 1337-1353.

Lim, M. K., V. Tang, A. Le Saux, J. Schüller, C. Bongards and N. Lehming (2007). "Gal11p dosage-compensates transcriptional activator deletions via Taf14p." Journal of Molecular Biology **374**: 9-23.

Lipford, J. R., G. T. Smith, Y. Chi and R. J. Deshaies (2005). "A putative stimulatory role for activator turnover in gene expression." Nature **438**: 113–-116.

Lohr, D., P. Venkov and J. Zlatanova (1995). "Transcriptional regulation in the yeast GAL gene family: a complex genetic network." Federation of American Societies for Experimental Biology Journal **9**(9): 777-787.

Long, R. M., L. M. Mylin and J. E. Hopper (1991). "GALil (SPT13), a Transcriptional Regulator of Diverse Yeast Genes, Affects the Phosphorylation State of GAL4, a Highly Specific Transcriptional Activator." Molecular & Cellular Biology **11**(4): 2311-2314.

Longtine, M. S., A. M. III, D. J. Demarini, N. G. Shah, A. Wach, A. Brachat, P. Philippsen and J. R. Pringle (1998). "Additional Modules for Versatile and Economical PCR-based Gene Deletion and Modification in Saccharomyces cerevisiae." Yeast **14**: 953-961.

Lorch, Y. and R. D. Kornberg (1985). "A region flanking the GAL7 gene and a binding site for GAL4 protein as upstream activating sequences in yeast " Journal of Molecular Biology **186**: 821-824.

Los, G. V., L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood (2008). "HaloTag: a novel protein labeling technology for cell imaging and protein analysis." ACS Chemical Biology **3**(6): 373-382.

Lu, K., W. Ye, L. Zhou, L. B. Collins, X. Chen, A. Gold, L. M. Ball and J. A. Swenberg (2010). "Structural Characterization of Formaldehyde-Induced Cross-Links Between Amino Acids and Deoxynucleosides and Their Oligomers." Journal of the American Chemical Society **132**(10): 3388-3399.

Lu, K., W. Ye, L. Zhou, L. B. Collins, X. Chen, A. Gold, L. M. Ball and J. A. Swenberg (2010). "Structural Characterization of Formaldehyde-Induced Cross-links Between Amino Acids and Deoxynucleosides and Their Oligomers." J. Am. Chem. Soc. **132**: 3388-3399.

Ludwig, D. L. and C. V. Bruschi (1991). "The 2-micron plasmid as a nonselectable, stable, high copy number yeast vector." Plasmid **25**(2): 81-95.

Lue, N. F. and R. D. Kornberg (1987). "Accurate initiation at RNA polymerase II promoters in extracts from Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences **84**: 8839-8843.

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

Luse, D. S. (2013). "The RNA polymerase II preinitiation complex. Through what pathway is the complex assembled?" Transcription **5**(1): e27050.

Luse, D. S. and G. A. Jacob (1987). "Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter." Journal of Biological Chemistry **262**: 14990-14997.

Ma, J. and M. Ptashne (1987). "The carboxyl-terminal 30 amino acids of GAL4 are recognized by GAL80." Cell **50**: 137-142.

Ma, J. and M. Ptashne (1987). "Deletion analysis of GALA defines two transcriptional activating segments." Cell **48**: 847-853.

Madison, J. M. and F. Winston (1997). "Evidence that Spt3 functionally interacts with Mot1, TFIIA, and TATA-binding protein to confer pro- moter-specific transcriptional control in Saccharomyces cerevisiae." Molecular & Cellular Biology **17**: 287–295.

Matsui, T., J. Segall, P. A. Weil and R. G. Roeder (1980). "Multiple factors required for accurate initiation of transcription by purified RNA polymerase II." Journal of Biological Chemistry **255**: 11992–11996.

Matsumoto, K., A. Toh-e and Y. Oshima (1981). "Isolation and characterization of dominant mutations resistant to carbon catabolite repression of galactokinase synthesis in Saccharomyces cerevisiae." Molecular & Cellular Biology **1** 83-93.

Maxon, M. E., J. A. Goodrich and R. Tjian (1994). "Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance." Genes & Development **8**(5): 515-524.

Mazza, D., A. Abernathy, N. Golob, T. Morisaki and J. G. McNally (2012). "A benchmark for chromatin binding measurements in live cells." Nucleic Acids Research **40**(15): e119.

Mazza, D., F. Mueller, T. J. Stasevich and J. G. McNally (2013). "Convergence of chromatin binding estimates in live cells." Nature Methods **10**(8): 691-692.

McCormick, M. A., Delaney, J. R., Tsuchiya, M., Tsuchiyama, S., Shemorry, A., Sim, S., Chou, A. C.-Z., Ahmed, U., Carr, D., Murakami, C. J., *et al.* (2015). "A comprehensive analysis of

replicative lifespan in 4,698 single-gene deletion strains uncovers conserved mechanisms of aging." Cellular Metabolism **22**(5):895-906.

McGhee, J. D. and P. H. von Hippel (1975). "Formaldehyde as a probe of DNA structure. II. Reaction with endocyclic imino groups of DNA bases." Biochemistry. **14**: 1297-1303.

McGhee, J. D. and P. H. Von Hippel (1975). "Formaldehyde as a probe of DNA structure. I. Reaction with exocyclic amino groups of DNA bases." Biochemistry **14**(6): 1281-1296.

McGhee, J. D. and P. H. von Hippel (1976). "Froamldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA." Biochem. **16**: 3276-3293.

McGhee, J. D. and P. H. von Hippel (1977). "Formaldehyde as a probe of DNA structure. 3. Equilibrium denaturation of DNA and synthetic polynucleotides." Biochem. **16**: 3267-3276.

McNally, J. G. (2000). "The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells." Science **287**(5456): 1262-1265.

Means, G. E. and R. E. Feeney (1995). "Reductive Alkylation of Proteins." Analyt. Biochem. **224**: 1-16.

Meisterernst, M., M. Horikoshi and R. G. Roeder (1990). "Recombinant yeast TFIID, a general transcription factor, mediates activation by the gene-specific factor USF in a chromatin assembly assay." Proceedings of the National Academy of Sciences **87**: 9153–9157.

Meisterernst, M. and R. G. Roeder (1991). "Family of proteins that interact with TFIID and regulate promoter activity." Cell **67**: 557–567.

Metz, B., G. F. A. Kersten, P. Hoogerhout, H. F. Brugghe, H. A. M. Timmermans, A. de Jong, H. Meiring, J. ten Hove, W. E. Hennink, D. J. A. Crommelin and W. Jiskoot (2004). "Identification of formaldehyde-induced modifications in proteins: reactions with model peptides." Journal of Biological Chemistry **279**: 6235-6243.

Mikuni, S., M. Tamura and M. Kinjo (2007). "Analysis of intranuclear binding process of glucocorticoid receptor using fluorescence correlation spectroscopy." Federation of European Biochemical Societies Letters **581**: 389–393.

Mirny, L., M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith and A. Kosmrlj (2009). "How a protein searches for its site on DNA: the mechanism of facilitated diffusion." Journal of Physics A: Mathematical and Theoretical **42**(43): 434013.

Misteli, T. (2001). "Protein Dynamics: Implications for Nuclear Architecture and Gene Expression." Science **291**(5505): 843-847.

Mohibullah, N. and S. Hahn (2008). "Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3." Genes & Development **22**: 2994–3006.

Morisaki, T., W. G. Müller, N. Golob, D. Mazza and J. G. McNally (2014). "Single-molecule analysis of transcription factor binding at transcription sites in live cells." Nature Communications **5**.

Mueller, F., T. J. Stasevich, D. Mazza and J. G. McNally (2013). "Quantifying transcription factor kinetics: At work or at play?" Critical Reviews in Biochemistry and Molecular Biology **48**(5): 492-514.

Muldrow, T. A., A. M. Campbell, P. A. Weil and D. T. Auble (1999). "MOT1 can activate basal transcription in vitro by regulating the distribu- tion of TATA binding protein between promoter and nonpromoter sites." Molecular & Cellular Biology **19**: 2835–2845.

Murakami, K., G. Calero, C. R. Brown, X. Liu, R. E. Davis, H. Boeger and R. D. Kornberg (2013). "Formation and Fate of a Complete 31-Protein RNA Polymerase II Transcription Preinitiation Complex." Journal of Biological Chemistry **288**(9): 6325-6332.

Myers, L. C., C. M. Gustafsson, D. A. Bushnell, M. Lui, H. Erdjument-Bromage, P. Tempst and R. D. Kornberg (1998). "The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain." Genes & Development **12**(1): 45-54.

Näär, A. M., D. J. Taatjes, W. Zhai, E. Nogales and R. Tjian (2002). "Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation." Genes & Development **16**(11): 1339-1344.

Nacheva, G. A., D. Y. Guschin, O. V. Preobrazhenskaya, V. L. Karpov, K. K. Ebralidse and A. D. Mirzabekov (1989). "Change in the pattern of histone binding to DNA upon transcriptional activation." Cell **58**: 27-36.

Nalley, K., S. A. Johnston and T. Kodadek (2006). "Proteolytic turnover of the Gal4 transcription factor is not required for function in vivo." Nature **442**: 1054-1057.

Nencioni, L., G. Volpini, S. Peppoloni, M. Bugnoli, T. de Magistris, I. Marsili and R. Rappuoli (1991). "Properties of Pertussis Toxin Mutant PT-9K/129G after Formaldehyde Treatment." Infect.and Immunity. **59**: 625-630.

Nomarski, G. and A. R. Weill (1955). "Application à la métallographie desméthodes interférentielles à deux ondes polarisées." Revue de Metallurgie **2**: 121–128.

Nowak, D. E., B. Tian and A. R. Brasier (2005). "Two-step cross-linking method for identification of NF-KB gene network by chromatin immunoprecipitation." Biotechniques **39**: 715-725.

Ohkuma, Y., S. Hashimoto, C. K. Wang, M. Horikoshi and R. G. Roeder (1995). "Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-alpha." Molecular and Cellular Biology **15**(9): 4856-4866.

Orban, P. C., D. Chui and J. D. Marth (1992). "Tissue- and site-specific DNA recombination in transgenic mice." Proceedings of the National Academy of Sciences **89**: 6861-6865.

Orlando, V. (2000). "Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation." TIBS **25**: 99-104.

Orphanides, G., T. Lagrange and D. Reinberg (1996). "The general transcription factors of RNA polymerase II." Genes & Development **10**(21): 2657-2683.

Owen, D. J., P. Ornaghi, J. C. Yang, N. Lowe, P. R. Evans, P. Ballario, D. Neuhaus, P. Filetici and A. A. Travers (2000). "The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p." The EMBO Journal **19**: 6141–6149.

Pan, G. and J. Greenblatt (1994). "Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site." Journal of Biological Chemistry **269**: 30101-30104.

Papai, G., P. A. Weil and P. Schultz (2011). "New insights into the function of transcription factor TFIID from recent structural studies." Current Opinion in Genetics & Development **21**: 219–224.

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

Park, J. M., H.-S. Kim, S. J. Han, M.-S. Hwang, Y. C. Lee and Y.-J. Kim (2000). " In vivo requirement of activator- specific binding targets of mediator." Molecular & Cellular Biology **20**: 8709–8719.

Parthun, M. R. and J. A. Jaehning (1990). "Purification and Characterization of the Yeast Transcriptional Activator GAL4." Journal of Biological Chemistry **265**(1): 209-213.

Pavri, R., B. Lewis, T. K. Kim, F. J. Dilworth, H. Erdjument-Bromage, P. Tempst, G. de Murcia, R. Evans, P. Chambon and D. Reinberg (2005). "PARP-1 determines specificity in a retinoid signaling pathway via direct modulation of mediator." Molecular Cell **18**: 83–96.

Pei, Y., S. Hausmann, C. K. Ho, B. Schwer and S. Shuman (2001). "The length, phosphorylation state, and primary structure of the RNA polymerase II carboxy-terminal domain dictate interactions with mRNA capping enzymes." Journal of Biological Chemistry **276**: 28075–28082.

Pelechano, V., S. Chávez and J. E. Pérez-Ortín (2010). "A complete set of nascent transcription rates for yeast genes." PLoS ONE **5**: e15442.

Perlmann, T., P. Eriksson and O. Wrange (1990). "Quantitative analysis of the glucocorticoid receptor-DNA interaction at the mouse mammary tumor virus glucocorticoid response element." Journal of Biological Chemistry **265**: 17222–17229.

Petrenko, N., Y. Jin, K. H. Wong and K. Struhl (2016). "Mediator Undergoes a Compositional Change during Transcriptional Activation." Molecular Cell **64**: 1-12.

Phair, R. D., P. Scaffidi, C. Elbi, J. Vecerova, A. Dey, K. Ozato, D. T. Brown, G. Hager, M. Bustin and T. Misteli (2004). "Global Nature of Dynamic Protein-Chromatin Interactions In Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins." Molecular and Cellular Biology **24**(14): 6393-6402.

Pinto, I., W.-H. Wu, J. G. Na and M. Hampsey (1994). "Characterization of sua7 mutations defines a domain of TFIIB involved in transcription start site selection in yeast." Journal of Biological Chemistry **269**: 30569-30573.

Ponjavic, J., B. Lenhard, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki and A. Sandelin (2006). "Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters." Genome Biology **7**: R78.

Ponticelli, A. S. and K. Struhl (1990). "Analysis of Saccharomyces cerevisiae his3 Transcription In Vitro: Biochemical Support for Multiple Mechanisms of Transcription." Molecular & Cellular Biology: 2832-2839.

Poorey, K., R. O. Sprouse, M. N. Wells, R. Viswanathan, S. Bekiranov and D. T. Auble (2010). "RNA synthesis precision is regulated by preinitiation complex turnover." Genome Research **20**: 1679 –1688.

Poorey, K., R. Viswanathan, M. N. Carver, T. S. Karpova, S. M. Cirimotich, J. G. McNally, S. Bekiranov and D. T. Auble (2013). "Measuring Chromatin Interaction Dynamics on the Second Time Scale at Single-Copy Genes." Science **342**(6156): 369-372.

Poss, Z. C., C. C. Ebmeier and D. J. Taatjes (2013). "The Mediator complex and transcriptional regulation." Critical Reviews in Biochemistry and Molecular Biology **48**(6): 574-608.

Prelich, G. (1997). "Saccharomyces cerevisiae BUR6 encodes a DRAP1/ NC2a homolog that has both positive and negative roles in transcription in vivo." Molecular & Cellular Biology **17**: 2057–2065.

Price, D. H. (2008). "Poised polymerases: on your mark…get set...go! ." Molecular Cell **30**: 7-10.

Ptashne, M. (2005). "Regulation of transcription: from lambda to eukaryotes." Trends in Biochemical Sciences **30**: 275–279.

Pugh, B. F. and R. Tjian (1991). "Transcription from a TATA-less promoter requires a multisubunit TFIID complex." Genes & Development **5**: 1935-1945.

Quievryn, G. and A. Zhitkovich (2000). "Loss of DNA–protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function." Carcinogenesis **21**(8): 1573-1580.

Rafalska-Metcalf, I. U., S. L. Powers, L. M. Joo, G. LeRoy and S. M. Janicki (2010). "Single cell analysis of transcriptional activation dynamics." PLoS ONE **5**: e10272.

Ram, A. F. J., H. V. D. Ende and F. M. Klis (1998). "Green fluorescent protein-cell wall fusion proteins are covalently incorporated into the cell wall of Saccharomyces cerevisiae." FEMS Microbiology Letters **162**(2): 249-255.

Raser, J. M. and E. K. O'Shea (2005). "Noise in gene expression: origins, consequences, and control." Science **309**: 2010-2013.

Ravarani, C. N. J., G. Chalancon, M. Breker, N. S. de Groot and M. M. Babu (2016). "Affinity and competition for TBP are molecular determinants of gene expression noise." Nature Communications **7**: 10417.

Reinberg, D. and R. G. Roeder (1987). "Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE." Journal of Biological Chemistry **262**: 3310–3321.

Ren, B., F. Robert, J. J. Wyrick, O. Aparicio, E. G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C. J. Wilson, S. P. Bell and R. A. Young (2000). "Genome-Wide Location and Function of DNA Binding Proteins." Science **290**: 2306-2309.

Rhee, H. S. and B. F. Pugh (2012). "Genome-wide structure and organization of eukaryotic pre-initiation complexes." Nature **483**(7389): 295-301.

Richter, K. N., N. H. Revelo, K. J. Seitz, M. S. Helm, D. Sarkar, R. S. Saleeb, E. D'Este, J. Eberle, E. Wagner, C. Vogl, D. F. Lazaro, F. Richter, J. Coy‐Vergara, G. Coceano, E. S. Boyden, R. R. Duncan, S. W. Hell, M. A. Lauterbach, S. E. Lehnart, T. Moser, T. Outeiro, P. Rehling, B. Schwappach, I. Testa, B. Zapiec and S. O. Rizzoli (2018). "Glyoxal as an alternative fixative to formaldehyde in immunostaining and super‐resolution microscopy." The EMBO Journal **37**(6): 139-159.

Robinett, C. C. (1996). "In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition." The Journal of Cell Biology **135**(6): 1685-1700.

Rodriguez, C. R., E.-J. Cho, M.-C. Keogh, C. L. Moore, A. L. Greenleaf and S. Buratowski (2000). "Kin28, the TFIIH-associated carboxy- terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II." Molecular & Cellular Biology **20**: 104–112.

Roeder, R. (1996). "The role of general initiation factors in transcription by RNA polymerase II." Trends in Biochemical Sciences **21**(9): 327-335.

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

Rohs, R., X. Jin, S. M. West, R. Joshi, B. Honig and R. S. Mann (2010). "Origins of specificity in protein-DNA recognition." Ann. Rev. Biochem. **79**: 233-269.

Rufiange, A., P.-É. Jacques, W. Bhat, F. Robert and A. Nourani (2007). "Genome-wide replication-independent histone H3 exchange occurs predominantly at promoters and implicates H3 K56 acetylation and Asf1." Molecular Cell **27**: 393–405

Sainsbury, S., J. Niesser and P. Cramer (2013). "Structure and function of the initially transcribing RNA polymerase II–TFIIB complex." Nature **493**(7432): 437-440.

Sakurai, H. and T. Fukasawa (1997). "Yeast Gal11 and transcription factor IIE function through a common pathway in transcriptional regulation." Journal of Biological Chemistry **272**: 32663–32669.

Sakurai, H., H. Yoshiki and T. Fukasawa (1993). "Yeast GAL11 protein is a distinctive type transcription factor that enhances basal transcription in vitro." Proceedings of the National Academy of Sciences **90**: 8382-8386.

Saleh, A., D. Schieltz, N. Ting, S. B. McMahon‖, D. W. Litchfield, J. R. Y. III, S. P. Lees-Miller, M. D. Cole and C. J. Brand (1998). "Tra1p Is a Component of the Yeast Ada·Spt Transcriptional Regulatory Complexes." Journal of Biological Chemistry **273**: 26559-26565.

Samuels, M., A. Fire and P. A. Sharp (1982). "Separation and characterization of factors mediating accurate transcription by RNA polymerase II." Journal of Biological Chemistry **257**: 14419–14427.

Sauer, B. and N. Henderson (1988). "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1." Proceedings of the National Academy of Sciences **85**: 5166-5170.

Sawadogo, M. and R. G. Roeder (1985). "Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay." Proceedings of the National Academy of Sciences **82**: 4394–4398.

Sayre, M. H., H. Tschochner and R. D. Kornberg (1992). "Reconstitution of transcription with five purified initiation factors and RNA polymerase II from Saccharomyces cerevisiae." Journal of Biological Chemistry **267**: 23376 –23382.

Schermer, U. J., P. Korber and W. Hörz (2005). "Histones are incorporated in trans during reassembly of the yeast PHO5 promoter." Molecular Cell **Volume 19**(2): 279–285.

Schilbach, S., M. Hantsche, D. Tegunov, C. Dienemann, C. Wigge, H. Urlaub and P. Cramer (2017). "Structures of transcription pre-initiation complex with TFIIH and Mediator." Nature **551**(204-209).

Schluesche, P., G. Stelzer, E. Piaia, D. C. Lamb and M. Meisterernst (2007). "NC2 mobilizes TBP on core promoter TATA boxes." Nature Structural & Molecular Biology **14**(12): 1196-1201.

Schmiedeberg, L., P. Skene, A. Deaton and A. Bird (2009). "A Temporal Threshold for Formaldehyde Crosslinking and Fixation." PLoS ONE **4**: e4636.

Serizawa, H., J. W. Conaway and R. C. Conaway (1993). "Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription." Nature **363**: 371–374.

Shi, S.-R., C. R. Taylor, C. B. Fowler and J. T. Mason (2013). "Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies." PROTEOMICS - Clinical Applications **7**(3-4): 264-272.

Shi, S.-R., C. R. Taylor, C. B. Fowler and J. T. Mason (2013). "Complete solubilization of formalin-fixed, paraffin-embedded tissue may improve proteomic studies." Proteomics Clin. Appl. **7**: 264-272.

Shi, X., A. Finkelstein, A. J. Wolf, P. A. Wade, Z. F. Burton and J. A. Jaehning (1996). "Paf1p, an RNA Polymerase II-Associated Factor in Saccharomyces cerevisiae, May Have both Positive and Negative Roles in Transcription." Molecular & Cellular Biology **16**(2): 669-676.

Shikama, K. and K.-I. Miura (1976). "Equilibrium studies on the formaldehyde reaction with native DNA." Eur. J. Biochem. **63**: 39-46.

Shimada, H. and T. Fukasawa (1985). "Controlled transcription of the yeast regulatory gene GAL80." Gene **39**: 1-9

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

Sikorski, T. W. and S. Buratowski (2009). "The basal initiation machinery: beyond the general transcription factors." Current Opinion in Cell Biology **21**: 344–351.

Siomin, Y. A., V. V. Simonov and A. M. Poverenny (1973). "The reaction of formaldehyde with deoxynucleotides and DNA in the presence of amino acids and lysine-rich histone." Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis **331**(1): 27-32.

Slutsky, M. and L. A. Mirny (2004). "Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential." Biophysical Journal **87**: 4021-4035.

Solomon, M. J., P. L. Larsen and A. Varshavsky (1988). "Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene." Cell **53**(6): 937-947.

Solomon, M. J. and A. Varshavsky (1985). "Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures." Proceedings of the National Academy of Sciences **82**(19): 6470-6474.

Sopta, M., R. W. Carthew and J. Greenblatt (1985). "Isolation of three proteins that bind to mammalian RNA polymerase II." Journal of Biological Chemistry **260** 10353-10360

Sprague, B. L., R. L. Pego, D. A. Stavreva and J. G. McNally (2004). "Analysis of Binding Reactions by Fluorescence Recovery after Photobleaching." Biophysical Journal **86**(6): 3473-3495.

Sprouse, R. O., T. S. Karpova, F. Mueller, A. Dasgupta, J. G. McNally and D. T. Auble (2008). "Regulation of TATA-binding protein dynamics in living yeast cells." Proceedings of the National Academy of Sciences **105**(36): 13304-13308.

St. John, T. P. and R. W. Davis (1981). "The organization and transcription of the galactose gene cluster Saccharomyces." Journal of Molecular Biology **152**: 285-315.

Stagge, F., G. Y. Mitronova, V. N. Belov, C. A. Wurm and S. Jakobs (2013). "Snap-, CLIP- and Halo-Tag Labelling of Budding Yeast Cells." PLoS ONE **8**(10): e78745.

Stasevich, T. J., Y. Hayashi-Takanaka, Y. Sato, K. Maehara, Y. Ohkawa, K. Sakata-Sogawa, M. Tokunaga, T. Nagase, N. Nozaki, J. G. McNally and H. Kimura (2014). "Regulation of RNA polymerase II activation by histone acetylation in single living cells." Nature **516**(7530): 272-275.

Sternberg, N. and D. Hamilton (1981). "Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites." Journal of Molecular Biology **150**(4): 467-486.

Sterner, D. E., P. A. Grant, S. M. Roberts, L. J. Duggan, R. Belotserkovskaya, L. A. Pacella, F. Winston, J. L. Workman and S. L. Berger (1999). "Functional organization of the yeast SAGA complex: Distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction." Molecular & Cellular Biology **19**: 86–98.

Stewart-Ornstein, J., J. S. Weissman and H. El-Samad (2012). "Cellular Noise Regulons Underlie Fluctuations in Saccharomyces cerevisiae." Molecular Cell **45**(4): 483-493.

Storici, F. and M. A. Resnick (2006). "The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast." Methods in Enzymology **409**: 329-345.

Strathern, J. N., L. C. Blair and I. Herskowitz (1979). "Healing of mat mutations and control of mating type interconversion by the mating type locus in Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences **76**(7): 3425-3429.

Struhl, K. (1987). "Promoters, activator proteins, and the mechanism of transcriptional initiation in yeast." Cell **49**: 295-297.

Struhl, K. (1999). "Fundamentally different logic of gene regulation in eukaryotes and prokaryotes." Cell **98**: 1-4.

Struhl, K. (2007). "Interpreting chromatin immunoprecipitation experiments." Evaluating Techniques in Biochemical Research, Zuk, D. (ed.) **Cell Press, Cambridge, MA**: 29-33.

Sutherland, B. W., J. Toews and J. Kast (2008). "Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions." Journal of Mass Spectrometry **43**(6): 699-715.

Svensson, J. P, Shukla, M., Menendez-Benito, V., Norman-Axelsson, U., Audergon, P., Sinha, I., Tanny, J. C., Allshire, R. C. and K. Ekwall (2015). "A nucleosome turnover map reveals that the stability of histone H4 Lys20 methylation depends on histone recycling in transcribed chromatin." Genome Research **25**(6): 872-883.

Taatjes, D. J. (2017). "The Continuing SAGA of TFIID and RNA Polymerase II Transcription." Molecular Cell **68**(1): 1-2.

Takahashi, H., K. Kasahara and T. Kokubo (2009). "Saccharomyces cerevisiae Med9 comprises two functionally distinct domains that play different roles in transcriptional regulation." Genes to Cells **14**: 53–67.

Teytelman, L., D. M. Thurtle, J. Rine and A. van Oudenaarden (2013). "Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins." PNAS **110**: 18602-18607.

Thomas, M. C. and C.-M. Chiang (2006). "The General Transcription Machinery and General Cofactors." Critical Reviews in Biochemistry and Molecular Biology **41**(3): 105-178.

Thompson, C. M. and R. A. Young (1995). "General requirement for RNA polymerase II holoenzymes in vivo." Proceedings of the National Academy of Sciences **92**: 4587–4590.

Tirode, F., D. Busso, F. Coin and J. M. Egly (1999). "Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7." Molecular Cell **3**: 87-95.

Toews, J., J. C. Rogalski, T. J. Clark and J. Kast (2008). "Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions." Analytica Chimica Acta **618**(2): 168-183.

Toews, J., J. C. Rogalski and J. Kast (2010). "Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications." Analytica Chimica Acta **676**: 60-67.

Toth, J. and M. D. Biggin (2000). "The specificity of protein–DNA crosslinking by formaldehyde: in vitro and in Drosophila embryos." Nucleic Acids Research **28**: e4.

Toth-Petroczy, A., C. J. Oldfield, I. Simon, Y. Takagi, A. K. Dunker, V. N. Uversky and M. Fuxreiter (2008). "Malleable machines in transcription regulation: the mediator complex." PLoS Computational Biology **4**: e1000243.

Traven, A., B. Jelicic and M. Sopta (2006). "Yeast Gal4: a transcriptional paradigm revisited." EMBO reports **7**(5): 496-499.

True, J. D., J. J. Muldoon, M. N. Carver, K. Poorey, S. J. Shetty, S. Bekiranov and D. T. Auble (2016). "The Modifier of Transcription 1 (Mot1) ATPase and Spt16 Histone Chaperone Co-regulate Transcription through Preinitiation Complex Assembly and Nucleosome Organization." Journal of Biological Chemistry **291**(29): 15307-15319.

Tschopp, J. F., S. D. Emr, C. Field and R. Schekman (1986). "GAL2 codes for a membrane-bound subunit of the galactose permease in Saccharomyces cerevisiae." Journal of Bacteriology **166**(1): 313-318.

Ulery, T. L., D. A. Mangus and J. A. Jaehning (1991). "The yeast IMP1 gene is allelic to GAL2." Molecular & General Genetics **230**: 129-135.

Utiyama, H. and P. Doty (1971). "Kinetic Studies of Denaturation and Reaction with Formaldehyde on Polydeoxyribonucleotides." Biochem. **10**: 1254-1264.

Van Dyke, M. W., R. G. Roeder and M. Sawadogo (1988). "Physical Analysis of Transcription Preinitiation Complex Assembly on a Class II Gene Promoter." Science **241**(4871): 1335-1338.

van Royen, M. E., P. Farla, K. A. Mattern, B. Geverts, J. Trapman and A. B. Houtsmuller (2009). "Fluorescence recovery after photobleaching (FRAP) to study nuclear protein dynamics in living cells." Methods in Molecular Biology **464**: 363-385.

van Royen, M. E., A. Zotter, S. M. Ibrahim, B. Geverts and A. B. Houtsmuller (2011). "Nuclear proteins: finding and binding target sites in chromatin." Chromosome Reseach **19**: 83-98.

van Werven, F. J., H. van Bakel, H. A. van Teeffelen, A. F. Altelaar, M. G. Koerkamp, A. J. Heck, F. C. Holstege and H. T. Timmers (2008). "Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome." Genes & Development **22**: 2359-2369.

Vani, K., S. A. Bogen and S. R. Sompuram (2006). "A high throughput combinatorial library technique for identifying formalin-sensitive epitopes." J. Immunol. Methods **317**: 80-89.

Vannini, A. and P. Cramer (2012). "Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries." Molecular Cell **45**(4): 439-446.

Vashee, S., H. Xu, S. A. Johnston and T. Kodadek (1993). "How do "Zn2 cys6" proteins distinguish between similar upstream activation sites? Comparison of the DNA-binding specificity of the GAL4 protein in vitro and in vivo." Journal of Biological Chemistry **268**: 24699-24706.

Vasilescu, J., X. Guo and J. Kast (2004). "Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry." Proteomics **4**: 3845-3854.

Venters, B. J., S. Wachi, T. N. Mavrich, B. E. Andersen, P. Jena, A. J. Sinnamon, P. Jain, N. S. Rolleri, C. Jiang, C. Hemeryck-Walsh and B. F. Pugh (2011). "A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces." Mol. Cell **41**: 480-492.

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

Verzijlbergen, K. F., V. Menendez-Benito, T. v. Welsem, S. J. v. Deventer, D. L. Lindstrom, H. Ovaa, J. Neefjes, D. E. Gottschling and F. v. Leeuwen (2010). " Recombination-induced tag exchange to track old and new proteins." Proceedings of the National Academy of Sciences **107** 64–68

Viswanathan, R., E. A. Hoffman, S. J. Shetty, S. Bekiranov and D. T. Auble (2014). "Analysis of chromatin binding dynamics using the crosslinking kinetics (CLK) method." Methods **70**: 97-107.

Voit, E. O., H. A. Martens and S. W. Omholt (2015). "150 years of the mass action law." PLoS Comp. Biol. **11**: e1004012.

von Hippel, P. H. (2007). "From “Simple” DNA-Protein Interactions to the Macromolecular Machines of Gene Expression." Annual Review of Biophysics and Biomolecular Structure **36**(1): 79-105.

von Hippel, P. H. and K.-Y. Wong (1971). "Dynamic aspects of native DNA structure: Kinetics of the formaldehyde reaction with calf thymus DNA." J. Mol. Biol. **61**: 587-613.

Wach, A., A. Brachat, R. Pöhlmann and P. Philippsen (1994). "New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae." Yeast **10**: 1793-1808.

Wal, M. and B. F. Pugh (2012). Genome-Wide Mapping of Nucleosome Positions in Yeast Using High-Resolution MNase ChIP-Seq. Methods in Enzymology, Elsevier**:** 233-250.

Wang, F., S. Redding, I. J. Finkelstein, J. Gorman, D. R. Reichman and E. C. Greene (2013). "The promoter search mechanism of E. coli RNA polymerase is dominated by three-dimensional diffusion." Nature Structural & Molecular Biology **20**: 174-181.

Wang, W., M. Carey and J. D. Gralla (1992). "Polymerase II promoter activation: closed complex formation and ATP-driven start site opening." Science **255**: 450-453.

Warfield, L., S. Ramachandran, T. Baptista, D. Devys, L. Tora and S. Hahn (2017). "Transcription of Nearly All Yeast RNA Polymerase II- Transcribed Genes Is Dependent on Transcription Factor TFIID." Molecular Cell **68**: 1-12.

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

Weiss, S. and L. Gladstone (1959). "A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid." Journal of the American Chemical Society **81**: 4118–4119.

Werner, M., A. Chott, A. Fabiano and H. Battifora (2000). "Effect of formalin tissue fixation and processing on immunohistochemistry." Am. J. Surg. Pathol. **24**: 1016-1019.

Werven, F. J. v., H. A. A. M. v. Teeffelen, F. C. P. Holstege and H. T. M. Timmers (2009). "Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome." Nature Structural & Molecular Biology **16**(10): 1043-1049.

Werven, F. J. v. and H. T. M. Timmers (2006). "The use of biotin tagging in Saccharomyces cerevisiae improves the sensitivity of chromatin immunoprecipitation." Nucleic Acids Research **34**(4): e33.

West, R. W., R. R. Yocum and M. Ptashne (1984). "Saccharomyces cerevisiae GALJ-GALJO Divergent Promoter Region: Location and Function of the Upstream Activating Sequence UASG." Molecular & Cellular Biology **4**(2): 2467-2478.

Wolfram Research, I. (2016). Mathematica, Version 10.6. Champaign, IL

Workman, J. L. and R. G. Roeder (1987). "Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II." Cell **51**: 613–622.

Wu, C.-H., S. Chen, M. R. Shortreed, G. M. Kreitinger, Y. Yuan, B. L. Frey, Y. Zhang, S. Mirza, L. A. Cirillo, M. Olivier and L. M. Smith (2011). "Sequence-Specific Capture of Protein-DNA Complexes for Mass Spectrometric Protein Identification." PLoS ONE **6**(10): e26217.

Wu, C.-H., S. Chen, M. R. Shortreed, G. M. Kreitinger, Y. Yuan, B. L. Frey, Y. Zhang, S. Mirza, L. A. Cirillo, M. Olivier and L. M. Smith (2011). "Sequence-Specific Capture of Protein-DNA Complexes for Mass Spectrometric Protein Identification." PLoS ONE **6**: e26217.

Wyce, A., T. Xiao, K. A. Whelan, C. Kosman, W. Walter, D. Eick, T. R. Hughes, N. J. Krogan, B. D. Strahl and S. L. Berger (2007). "H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal

recruitment prior to removal by Ubp8 within a SAGA-related complex." Molecular Cell **27**: 275–288.

Yao, J., K. M. Munson, W. W. Webb and J. T. Lis (2006). "Dynamics of heat shock factor association with native gene loci in living cells." Nature **442**: 1050–1053.

Yin, Y., L. Yang, G. Zheng, C. Gu, C. Yi, C. He, Y. Q. Gao and X. S. Zhao (2014). "Dynamics of spontaneous flipping of a mismatched base in DNA duplex." PNAS **111**: 8043-8048.

Yocum, R. R. and M. Johnston (1984). "Molecular cloning of the GAL80 gene from Saccharomyces cerevisiae and characterization of a gal80 deletion " Gene **32**: 75-82.

Young, M. J. and D. A. Court (2008). "Effects of the S288c genetic background and common auxotrophic markers on mitochondrial DNA function in Saccharomyces cerevisiae." Yeast **25**: 903–912.

Young, R. A. (1991). "RNA polymerase II." Annual Review of Biochemistry **60**: 689– 715.

Yu, P. and T. Kodadek (2007). "Dynamics of the hypoxia-inducible factor- 1-vascular endothelial growth factor promoter complex." Journal of Biological Chemistry **282**: 35035–35045.

Yudkovsky, N., J. A. Ranish and S. Hahn (2000). "A transcription reinitiation intermediate that is stabilized by activator." Nature **408**: 225-229.

Zaidi, H., E. A. Hoffman, S. J. Shetty, S. Bekiranov and D. T. Auble (2017). "Second-generation method for analysis of chromatin binding with formaldehyde–cross-linking kinetics." Journal of Biological Chemistry **292**(47): 19338-19355.

Zaidi, H. A., D. T. Auble and S. Bekiranov (2017). "RNA synthesis is associated with multiple TBP-chromatin binding events." Scientific Reports **7**: 39631.

Zanton, S. J. and B. F. Pugh (2004). "Changes in genomewide occupancy of core transcriptional regulators during heat stress." Proceedings of the National Academy of Sciences **101**: 16843–16848.

Zanton, S. J. and B. F. Pugh (2006). "Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock." Genes & Development **20**(16): 2250-2265.

Zawel, L., K. P. Kumar and D. Reinberg (1995). "Recycling of the general transcription factors during RNA polymerase II transcription." Genes & Development **9**(12): 1479-1490.

Zawel, L. and D. Reinberg (1992). "Advances in RNA polymerase II transcription." Current Opinion in Cell Biology **4**(3): 488-495.

Zawel, L. and D. Reinberg (1993). "Initiation of transcription by RNA polymerase II: a multi-step process." Progress in Nucleic Acid Research and Molecular Biology **44**: 67-108.

Zeng, P.-Y., C. R. Vakoc, Z.-C. Chen, G. A. Blobel and S. L. Berger (2006). "In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation." Biotechniques **41**: 694-698.

Zenklusen, D., Larson, D. R., and R. H. Singer (2008). "Single-RNA counting reveals alternative modes of gene expression in yeast." Nature Structural & Molecular Biology **15**: 1263-1271.

Zentner, G. E. and S. Henikoff (2013). "Mot1 redistributes TBP from TATA-containing to TATA-less promoters." Molecular & Cellular Biology **33**: 4996 –5004.

Zhang, H., Roberts, D. N., and B. R. Cairns (2005). "Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for ativation through histone loss." Cell **123**(2): 219-31.

Zhang, Z., B. P. English, J. B. Grimm, S. A. Kazane, W. Hu, A. Tsai, C. Inouye, C. You, J. Piehler, P. G. Schultz, L. D. Lavis, A. Revyakin and R. Tjian (2016). "Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis." Genes & Development **30**(18): 2106-2118.